

Des comètes aux humains

L'évolution du vivant à la lecture de son ADN

Des comètes aux humains

L'évolution du vivant à la lecture de son ADN

GUY-FRANCK RICHARD

 **edp sciences**

17, avenue du Hoggar – P.A. de Courtabœuf
BP 112, 91944 Les Ulis Cedex A

Cet ouvrage est la deuxième édition du livre « Des comètes à l'homme synthétique
– Une histoire de la vie à la lecture des génomes » publié en 2018,
ISBN : 978-2-7598-2214-0.

Composition et mise en pages : Flexedo
Illustration de couverture : peb & fox

Imprimé en France
ISBN (papier) : 978-2-7598-3514-0
ISBN (ebook) : 978-2-7598-3515-7

Tous droits de traduction, d'adaptation et de reproduction par tous procédés, réservés pour tous pays. La loi du 11 mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'article 41, d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective », et d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (alinéa 1^{er} de l'article 40). Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles 425 et suivants du code pénal.

© EDP Sciences, 2024

À Romain et Émily,
mes deux diploïdes hétérozygotes.

SOMMAIRE

<i>Prologue</i>	13
1. L'origine des premières molécules organiques	17
La soupe primordiale de Miller	19
Les critiques de l'expérience de Miller	21
Le Monde Fer-Soufre	23
L'origine extraterrestre de la vie.....	26
Philae a rendez-vous avec Tchoury.....	28
Bibliographie	30
2. Des premières molécules organiques aux premières molécules du vivant	31
Une définition du vivant	31
L'ARN porteur d'information	32
L'ARN catalyseur de réactions chimiques.....	35
Le monde à ARN.....	37
Des premières molécules organiques au premier ARN	39
Bibliographie	42
3. Du monde à ARN aux premières cellules	43
L'Arbre du Vivant.....	44
Les spécificités du monde eucaryote	45
Les premières cellules	48
L'origine possible des premières membranes cellulaires.....	50
À quoi ressemblaient les protocellules ?.....	52
Bibliographie.....	54
4. Du monde à ARN au monde à ADN	55
Les remarquables propriétés de l'ADN	55
De l'ARN aux protéines.....	61

... des protéines à l'ADN-U	65
... et de l'ADN-U à l'ADN-T	67
Le paradoxe d'Eigen.....	68
Le chaînon manquant de l'ARN à l'ADN ?.....	71
Bibliographie.....	72
5. L'énergie des premières cellules et la Grande Oxydation	73
L'énergie des premières cellules.....	74
L'arrivée des cyanobactéries.....	77
L'invention de la photosynthèse	78
L'invention de la respiration	79
La Grande Oxydation.....	80
L'ozone, protecteur du vivant.....	81
Bibliographie.....	83
6. Gènes et code génétique.....	85
Qu'est-ce qu'un gène ?	85
Le code génétique.....	87
La presque universalité du code génétique.....	91
L'intrigante question des ARNt synthétases.....	92
Bibliographie.....	96
7. Les premiers eucaryotes	97
La théorie endosymbiotique.....	97
Un scénario possible pour l'apparition des premiers eucaryotes.....	99
L'origine du noyau des eucaryotes	102
Endosymbioses secondaires.....	104
Des amibes endosymbiotiques.....	105
... et des levures endosymbiotiques.....	106
Quand l'évolution rétrécit les génomes	107
Les nombreuses inventions du monde eucaryote	108
Bibliographie.....	110
8. Des premiers eucaryotes aux premiers animaux.....	111
L'acquisition de la mobilité chez les eucaryotes.....	112
Les gènes en morceaux des eucaryotes	114
Des protozoaires aux métazoaires	117
Bibliographie.....	120
9. Des métazoaires à l'homme.....	121
Les fossiles, témoins de l'explosion précambrienne.....	125
Les horloges moléculaires	126
L'apparition de la symétrie bilatérale	129
Les premiers chordés	130
Des chordés aux vertébrés	132

Des vertébrés aux mammifères.....	136
L'origine virale du placenta.....	138
La souris, proche parent de l'homme.....	138
Bibliographie.....	140
10. Des algues aux arbres.....	143
Les embryophytes, premières plantes terrestres	143
Les trachéophytes et les premiers arbres.....	145
Les angiospermes ou plantes à fleurs	146
La biodiversité des forêts	150
Le génome de l'arabette, angiosperme modèle.....	151
Le vivant s'est complexifié par bonds successifs.....	152
Bibliographie.....	156
11. Un animal doué de raison	159
Une machine à fabriquer des images mentales.....	160
Le cerveau récompense ou punit.....	162
Le séquençage du génome humain.....	164
La quête des gènes qui font l'Humanité.....	169
Quelle part de l'intelligence est-elle héritable ?	172
Le métagénome humain	174
Bibliographie.....	177
12. L'homme moderne est un hybride.....	179
Mastiquer ou réfléchir, il faut choisir	181
On a – presque – tous en nous quelque chose de Néandertal.....	184
L'homme de Néandertal pratiquait une médecine rudimentaire.....	187
L'homme de Denisova.....	189
Les apports des gènes ancestraux aux génomes humains modernes	190
Les premières migrations de nos ancêtres humains.....	192
Bibliographie.....	194
13. L'histoire des migrations humaines	195
Le peuplement de l'Europe.....	196
Le peuplement de l'Asie	199
Le peuplement des Amériques	200
Et l'Afrique dans tout ça ?	202
Bibliographie.....	204
14. Lamarck, Darwin, évolution et sélection naturelle	205
L'origine des espèces	208
Les limites de la théorie darwinienne.....	212
La théorie neutraliste	216
Les grandes extinctions de masse	222
La sixième extinction.....	226

La phylogénie moléculaire et la théorie de la coalescence.....	228
Les limites de la phylogénie moléculaire.....	230
Bibliographie.....	231
15. Les forces évolutives qui façonnent les génomes	233
Les génomes hybrides, une généralité du vivant.....	234
Le sexe, un moyen simple de créer de la diversité.....	238
Le prix du sexe	242
Le cas fascinant des bdelloïdes	243
Les transferts horizontaux	246
Les duplications totales de génomes	249
Que deviennent les gènes dupliqués ?	250
Évolution rapide des génomes eucaryotes par chromothripsie	253
Diversité des mécanismes d'évolution des génomes.....	256
Bibliographie.....	260
16. La fabrique du vivant	261
La révolution de la biologie moléculaire	262
La naissance de la biologie de synthèse	265
Les premiers gènes synthétiques.....	267
Les premiers génomes synthétiques	268
Pourquoi synthétiser des génomes entiers ?	272
Vers la création de nouvelles formes du vivant	275
Bibliographie.....	281
17. La vie existe-t-elle ailleurs que sur Terre ?	283
Les exoplanètes	284
Une vie prébiotique dans notre système solaire ?.....	286
Une autre vie intelligente est-elle possible ?.....	288
Bibliographie.....	293
<i>Épilogue : L'humanité, d'hier à demain.....</i>	<i>295</i>
LUCA : le dernier ancêtre commun universel.....	295
LUCA a-t-il jamais existé ?.....	297
Le libéralisme du vivant	299
Demain, l'univers sans l'homme	307
Bibliographie.....	311
<i>Glossaire</i>	<i>313</i>

*Les noms suivis d'un astérisque
sont définis dans le glossaire à la fin de l'ouvrage.*

*Je tiens à remercier chaleureusement le Dr Alexandre Kaminski
pour sa lecture assidue et critique de la première version
de ce livre, ainsi que pour ses conseils avisés.*

Guy-Franck Richard

PROLOGUE

Rien en biologie n'a de sens, excepté à la lumière de l'évolution.

Theodosius Dobzhansky, 1973

Lorsque les Anciens entendaient le tonnerre et voyaient tomber la foudre, ils y voyaient la colère de Zeus, dieu du ciel et père des dieux de l'Olympe. Plus de vingt siècles plus tard, Benjamin Franklin, l'inventeur du paratonnerre, démontrera la nature électrique de la foudre. De nos jours, on sait qu'elle est une décharge électrique se produisant lorsque de l'électricité statique s'accumule entre deux nuages ou entre un nuage et la terre. Cette décharge est accompagnée par un puissant dégagement de chaleur. Le bruit du tonnerre vient de l'expansion rapide de la colonne d'air qui a été chauffée par le passage de la foudre. Nulle intervention divine donc dans ces phénomènes météorologiques très communs.

Plus près de nous, le mystère des pierres mouvantes de la Vallée de la Mort en Californie (*sailing stones* ou *sliding stones* en anglais) a été récemment élucidé. Des roches se déplaçaient sur le sable de la vallée sans intervention humaine ou animale apparente. Les pierres pouvaient peser jusqu'à plusieurs dizaines, voire plusieurs

centaines de kilogrammes et les déplacements laissaient sur le sable des sillons particulièrement longs et bien visibles. Découvert dans les années 1900, il fallut de nombreuses études et plus d'un siècle pour comprendre ce phénomène étrange. Finalement, en 2014, une équipe scientifique trouva l'explication précise, en utilisant un GPS fixé sur une des pierres, des films image par image et une station météo portative qui donnait la température et la vitesse locale des vents ainsi que leur direction. Cette étude démontra que lors du dégel partiel des larges plaques de glace accumulées au fond de la vallée pendant l'hiver, celles-ci étaient poussées par les vents sur une fine pellicule d'eau résultant de leur dégel. En glissant ainsi, les plaques de glace entraînaient à leur tour les pierres, parfois sur plusieurs dizaines de mètres. Ainsi, le mystère des pierres mouvantes de la Vallée de la Mort fut résolu par une approche scientifique adaptée à la dimension du problème.

L'histoire des sciences est remplie de nombreux exemples similaires. Dans le présent ouvrage, nous allons essayer de fournir des éléments de réponse à la question de l'origine et de l'évolution de la vie, depuis les molécules organiques primitives fabriquées dans les poussières interstellaires ou au fond des sources chaudes océaniques, jusqu'à l'homme. Des explications rationnelles reposant sur des faits scientifiquement démontrés seront apportées à chaque étape du lent processus évolutif qui a mené jusqu'à une forme de vie intelligente. Il y sera fait une revue d'anciens concepts revisités à la lumière des découvertes du XXI^e siècle dans les domaines de l'astronomie, de la chimie et de la biologie. En particulier, l'apport de la **génomique*** à la compréhension de l'évolution des organismes

Génomique : Domaine de la génétique qui étudie la composition, la structure et l'évolution des génomes. La génomique comparative regroupe l'ensemble des approches bioinformatiques qui permettent la comparaison des génomes entre eux, en vue de déterminer les liens de parenté et les vitesses d'évolution entre les espèces comparées.

vivants fut considérable et les progrès accomplis dans les domaines du séquençage de l'ADN et de l'analyse informatique de ces séquences ont permis de résoudre des questions anciennes et très prégnantes en **génétique***, permettant de trancher entre différents scénarios évolutifs possibles.

Génétique : Discipline de la biologie qui étudie les gènes et les règles de leur transmission héréditaire.

Les différents concepts qui sous-tendent le vivant vont être présentés progressivement, de façon à en simplifier la compréhension par un lectorat de non-spécialistes. Les différents acteurs du vivant seront présentés par ordre chronologique, en commençant par les premières molécules organiques, l'ARN, puis les protéines et l'ADN, les premières cellules et enfin les premiers organismes multicellulaires, jusqu'à l'homme. La théorie scientifique de la sélection naturelle, telle que proposée par Charles Darwin il y a 150 ans, sera présentée, expliquée et discutée en s'appuyant sur les données les plus récentes issues de la génomique. Nous verrons aussi que les êtres vivants sont imparfaits, que le milieu naturel ne sélectionne pas des individus optimisés, mais simplement les individus les plus performants à un moment donné de leur histoire. Il y a très peu d'inventions nouvelles dans le vivant, la plupart du temps des « bricolages » évolutifs vont créer la diversité génétique, physiologique et morphologique observée. Tout ce qui paraissait au final assez mystérieux pour les naturalistes du XIX^e siècle apparaît aujourd'hui clairement à la lecture des génomes, comme la lumière blanche décomposée par un prisme en ses différentes longueurs d'onde. Comme le disait Theodosius Dobzhansky (1900-1975), célèbre généticien ukrainien : « *Rien en biologie n'a de sens, excepté à la lumière de l'évolution* ». C'est cette lumière que nous allons essayer de jeter maintenant sur l'histoire de la vie. Enfin, nous aborderons le domaine de la biologie synthétique et de la modification dirigée des génomes et tout ce que ces approches

apportent à la fois comme espoir pour l'humanité dans sa lutte contre les maladies génétiques et comme immense défi de société dans ce siècle qui commence.

L'auteur est conscient que chaque chapitre pourrait faire l'objet d'un ouvrage à part entière, en regard des extraordinaires richesses et complexité des questions posées. Il a donc fallu faire des choix et certains de ces choix pourront paraître regrettables aux spécialistes des différents domaines abordés ici. Je m'en excuse par avance auprès d'eux. Il était bien sûr impossible de citer tous les travaux. Il a au contraire été choisi d'essayer de donner une vision synthétique de l'apparition et de l'évolution de la vie sur Terre, telles que nous les comprenons aujourd'hui. Bien évidemment, il reste encore des zones d'ombre. Bien évidemment, des chercheurs de tous horizons travaillent à les éclairer. Nul doute qu'ils y arriveront car la curiosité est le plus puissant moteur des progrès scientifiques et technologiques. Et l'homme, le plus curieux des animaux.

1

L'origine des premières molécules organiques

L'origine des premières molécules ayant participé à la formation des organismes vivants primordiaux a été longtemps débattue et le sera sans doute encore pendant longtemps. Néanmoins, des observations très récentes faites par la sonde *Rosetta* en analysant les matériaux trouvés sur la comète Tchoury apportent un éclairage nouveau sur cette question. Avant de détailler ces découvertes, il est bon de rappeler que les êtres vivants, à la différence des minéraux, sont tous composés des mêmes **molécules*** dites « organiques ». L'architecture de ces molécules est basée sur l'atome de carbone. Un exemple de molécule organique est le dioxyde de carbone (formule chimique CO_2), un composé très abondant sur Terre produit par la respiration des êtres vivants, animaux et végétaux. L'atome

Molécule : Une molécule est un assemblage de plusieurs éléments chimiques (ou atomes) dans une configuration précise. Lorsque l'assemblage est de grande taille, on parle alors de « macromolécule ».

de carbone est un élément relativement abondant dans l'univers, qui présente des propriétés chimiques intéressantes lui permettant de s'associer à un grand nombre d'autres atomes, dans plusieurs types d'architectures différentes. Nous n'entrerons pas ici dans le détail de ces propriétés chimiques, il suffit de savoir que toute la vie sur Terre s'est organisée autour de l'atome de carbone. On va parler de molécules carbonées, ou plus souvent lorsqu'il s'agit du vivant, de macromolécules carbonées, car elles sont généralement de très grande taille.

Pour prendre une image simple, on peut considérer le carbone comme constituant la charpente d'une construction. Sans lui, pas de structure pérenne possible. D'autres éléments viennent compléter cette charpente : toit, cheminée, portes et fenêtres par exemple. Ces autres éléments sont souvent également composés de carbone, mais pas exclusivement. En revanche, ils sont tous reliés à la charpente, directement ou indirectement, afin que l'édifice ainsi bâti soit intègre et fonctionnel. En plus du carbone, d'autres éléments chimiques interviennent dans la construction : hydrogène, oxygène, azote, soufre et phosphore sont les plus couramment trouvés associés au carbone. Le monde organique, c'est-à-dire le monde des êtres vivants, est non seulement formé de molécules carbonées, mais il produit également des molécules carbonées très spécifiques, qui sont en quelque sorte sa signature. Ainsi, les matériaux organiques dont sont formés tous les êtres vivants, sans exception, ainsi que les molécules fabriquées et excrétées par ces mêmes êtres vivants, s'articulent autour de l'atome de carbone. C'est en cela qu'ils sont spécifiques du vivant. On peut alors se demander où et quand ont été créés les premiers matériaux organiques, les premières briques de la vie sur Terre ?

C'est un biochimiste russe et un généticien britannique qui vont proposer de façon totalement indépendante dans les années 1920 que les premières molécules du vivant soient apparues dans ce qui sera appelé « la soupe primordiale ». Alexandre Oparine (1894-1980)

publia en 1924 *Proiskhozhdenie zhizni* (L'origine de la vie), qui fut traduit pour la première fois en anglais en 1938. De façon tout à fait remarquable, John Burdon Sanderson Haldane (1892-1964) publia en 1929 un article intitulé *The Origin of Life*, sans connaissance liminaire du travail d'Oparine. Ces deux ouvrages suggéraient que l'atmosphère de la Terre primitive était très pauvre en oxygène et qu'en présence d'un apport énergétique approprié tel que des éclairs ou une source d'ultraviolets, il était possible de former des molécules organiques simples, les premières « briques » de la vie sur Terre. Les océans primitifs auraient alors joué le rôle d'une immense soupe chaude **prébiotique*** contenant des monomères de ces molécules organiques, qui se seraient assemblés plus ou moins rapidement en polymères, menant ainsi aux premières cellules vivantes. Cette hypothèse d'une soupe prébiotique à l'origine de la vie sur Terre trouvera quelque trente années plus tard un soutien expérimental grâce à la célèbre expérience de Miller.

LA SOUPE PRIMORDIALE DE MILLER

Stanley Miller (1930-2007) publia en 1953 les résultats d'une expérience censée décrire une possible origine de la vie, qui allait faire couler de l'encre pendant plusieurs décennies et dont les résultats sont encore aujourd'hui le sujet de débats passionnés. Il remplit un ballon en verre scellé avec l'équivalent d'un verre d'eau et trois gaz : hydrogène (formule chimique H_2), méthane (CH_4) et ammoniaque (NH_3). Puis il porta l'eau à ébullition et fit parcourir la vapeur d'eau résultante par des décharges électriques pendant toute la semaine que dura son expérience (Figure 1). L'idée était de tenter de recréer dans ce ballon l'atmosphère d'une Terre primitive constamment parcourue par des éclairs et d'étudier le résultat des réactions chimiques résultantes. À la fin de la première journée, l'eau contenue dans le ballon était devenue rose et à la fin de la semaine avait pris une apparence trouble et une teinte rouge sombre. L'analyse des molécules présentes

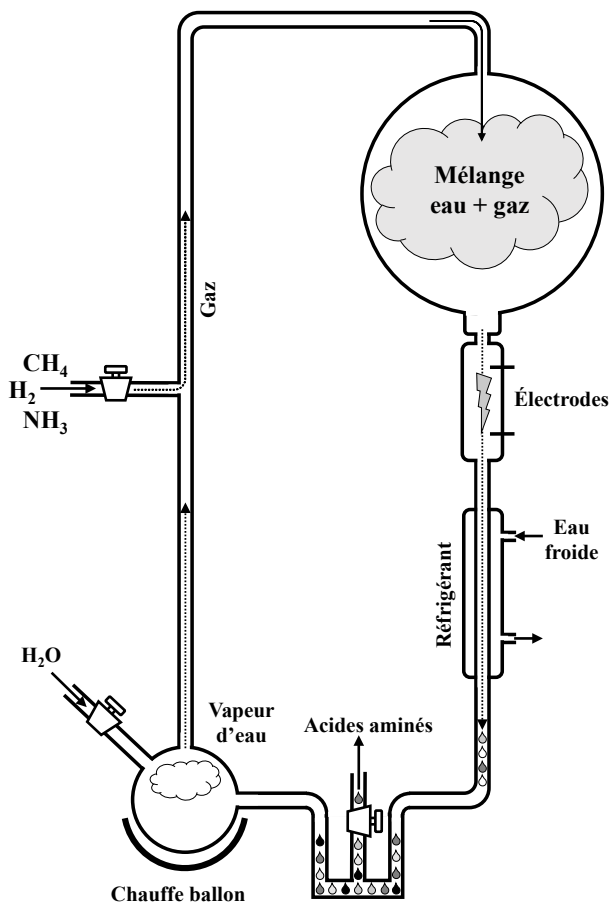


Figure 1 | L'expérience de Miller. Un ballon est rempli avec l'équivalent d'un verre d'eau et chauffé à ébullition (en bas à gauche). La vapeur d'eau circule et se mélange aux gaz dans le grand ballon (en haut à droite). Le mélange passe entre deux électrodes qui produisent une décharge électrique continue puis est refroidi, se condense et descend le long de la colonne vers le tube en U où il peut être collecté à intervalles réguliers. L'expérience dure une semaine en continu.

dans le ballon à la fin de la semaine identifia **cinq acides aminés***, des molécules organiques présentes chez tous les êtres vivants : l'acide aspartique, la glycine, l' α -alanine,

Acide aminé : Molécule organique simple, dont l'enchaînement précis forme les protéines, macromolécules essentielles du monde vivant.

la β -alanine et l'acide α -amino-n-butyrique, ainsi que la présence de deux autres acides aminés non identifiés. En tout, environ un milligramme d'acides aminés avait été synthétisé dans cette sorte de « soupe primordiale ». Ces résultats suggéraient que des « briques organiques élémentaires » pouvaient être assemblées dans des conditions que l'on supposait être celles de l'atmosphère primitive terrestre il y a environ 4 milliards d'années. Cette expérience, publiée dans la prestigieuse revue *Science*, eut un retentissement considérable dans la communauté scientifique et devint rapidement le sujet de nombreux débats, dont certains se poursuivent encore à l'heure actuelle.

LES CRITIQUES DE L'EXPÉRIENCE DE MILLER

Parmi celles-ci, la critique de la composition de l'atmosphère primordiale terrestre est la plus prégnante. En effet, l'hypothèse de Miller reposait sur une atmosphère riche en hydrogène gazeux, contenant du méthane et de l'ammoniaque, propice aux réactions pouvant conduire aux acides aminés observés à la fin de l'expérience. Or, il est très vraisemblable que la composition de l'atmosphère ait été très différente à cette époque, contenant en fait une faible concentration d'hydrogène gazeux. En effet, si l'atmosphère terrestre était composée principalement d'hydrogène et d'hélium lors de sa formation, l'allumage du soleil il y a 4,4 milliards d'années produisit un souffle suffisamment violent pour emporter les gaz les plus légers des planètes proches vers les planètes extérieures. L'hydrogène gazeux a donc disparu de l'atmosphère terrestre à cette époque, y laissant seulement les gaz les plus lourds, méthane (CH_4), ammoniaque (NH_3) et vapeur d'eau (H_2O). De plus, l'activité volcanique

violente aurait permis l'enrichissement en dioxyde de carbone (CO_2), en sulfure d'hydrogène (H_2S), voire en faibles quantités d'azote (N_2) et de monoxyde de carbone (CO). Cette composition serait beaucoup moins favorable à la formation des acides aminés observés par Miller. Mais l'atmosphère va encore changer de manière dramatique il y a environ 4 milliards d'années. Sous l'effet du refroidissement lent de la terre, la vapeur d'eau va se condenser sous forme d'eau liquide et des pluies torrentielles vont tomber à la surface de la planète entraînant deux conséquences majeures, la formation des océans et un changement atmosphérique radical. L'eau de pluie chargée de dioxyde de carbone (CO_2) va éroder les roches volcaniques et précipiter le calcium qu'elles contiennent pour former du carbonate de calcium (CaCO_3). Ce calcaire va alors se déposer au fond des jeunes océans pour former les toutes premières roches sédimentaires. L'atmosphère est maintenant très fortement appauvrie en dioxyde de carbone et en vapeur d'eau. Son composant majeur devient donc l'azote (N_2) et il le restera jusqu'à nos jours. La diminution du CO_2 atmosphérique aura pour résultat de réduire l'effet de serre et donc de faire baisser la température de surface pour permettre l'apparition des premières cellules, il y a environ 3,7 milliards d'années, comme nous le verrons plus loin.

Une autre critique de l'expérience de Miller repose sur l'utilisation de décharges électriques comme source d'énergie. Avec une atmosphère terrestre quasiment dépourvue d'oxygène gazeux (O_2) et *a fortiori* d'ozone (O_3), les rayonnements ultraviolets du soleil ne pouvaient pas être filtrés par la haute atmosphère comme ils le sont à l'heure actuelle par la couche d'ozone. L'énergie de ces rayons ultraviolets était donc suffisante pour activer la synthèse de molécules organiques à partir d'eau et de monoxyde de carbone atmosphérique, comme cela fut montré par la suite. Néanmoins, la présence d'importantes décharges électriques dans les nuages dégagés par les grandes éruptions volcaniques (telle celle du Eyjafjallajökull d'Islande au printemps 2010) suggère que de tels phénomènes devaient être

fréquents dans l'atmosphère terrestre primitive et donc pourraient avoir contribué – au moins en partie – à l'apparition de certaines molécules organiques.

La nature des molécules synthétisées lors de l'expérience de Miller est également sujet à critiques. En effet, dans les conditions de l'expérience, seuls des acides aminés ont été obtenus. Aucune autre molécule organique n'a été observée, en particulier aucune **purine*** ou **pyrimidine***, essentielle à la propagation de la vie telle qu'on la connaît (voir Chapitre 2).

Purines et pyrimidines : Molécules organiques simples qui constituent les bases de l'ADN et de l'ARN. Elles sont également directement impliquées dans nombre de réactions chimiques essentielles à la propagation de la vie.

Enfin, la concentration des molécules obtenues dans l'expérience de Miller était trop faible pour envisager leur interaction et la formation subséquente de molécules plus complexes à partir de ces briques organiques élémentaires. C'est d'ailleurs un des principaux arguments repris par les contempteurs de cette expérience. Pour l'ensemble de ces raisons, l'expérience tendant à prouver l'existence d'une « soupe primordiale » dans les conditions décrites par Stanley Miller a été régulièrement critiquée au fil des années et des théories alternatives ont été avancées afin d'expliquer l'origine de la vie sur Terre.

LE MONDE FER-SOUFRE

Günther Wächterhäuser (1938-), un chimiste allemand, a été l'un des plus virulents adversaires de la théorie des briques organiques élémentaires de Miller. Son principal argument repose sur le fait que dans une **solution aqueuse***, toute molécule organique simple va se détacher de son lieu de formation et s'en éloigner trop rapidement pour avoir le temps de réagir avec une autre molécule. En

d'autres termes, la concentration de molécules organiques capables de réagir entre elles sera trop faible pour aboutir à la formation de molécules plus complexes. Wächterhäuser propose donc que les réactions chimiques primitives ne se soient pas produites dans un monde à trois dimensions comme dans l'expérience de Miller, mais dans un monde à deux dimensions, à la surface de minéraux riches en fer et soufre, aboutissant dans la même réaction à la formation de molécules organiques et de pyrite (minéral composé de disulfure de fer, de formule FeS_2). Cette théorie présente plusieurs caractéristiques intéressantes. Tout d'abord, elle nécessite un environnement riche en sulfure d'hydrogène (H_2S), en dioxyde de carbone (CO_2) et en fer. On peut trouver ce genre de composition chimique à la résurgence de sources hydrothermales profondes, dans les régions de fort volcanisme qui recouvraient la terre primordiale. Dans cette théorie donc, pas besoin de spéculer sur la nature exacte de l'atmosphère terrestre primitive. D'autre part, cette réaction peut se produire spontanément, sans apport d'énergie extérieur, ce qui permet de s'affranchir des éclairs de Miller ou du rayonnement ultraviolet. Finalement, elle se déroule dans un espace à deux dimensions, ce qui augmente la probabilité d'interaction entre les molécules organiques.

Plus récemment, la synthèse expérimentale de molécules organiques **carbonylées***, contenant un cœur fer-soufre, dans des conditions de pression et de température élevées, suggère que ces molécules puissent être formées au fond des océans là où règnent les conditions nécessaires à leur formation, puis transportées vers la surface océanique où des conditions de température et de pression plus faibles permettraient la synthèse d'autres molécules organiques, comme proposé par Wächterhäuser (Figure 2). Ainsi, la théorie du

Carbonyle : Une molécule carbonylée contient un atome de carbone lié à un atome d'oxygène. De nombreuses molécules organiques sont carbonylées.

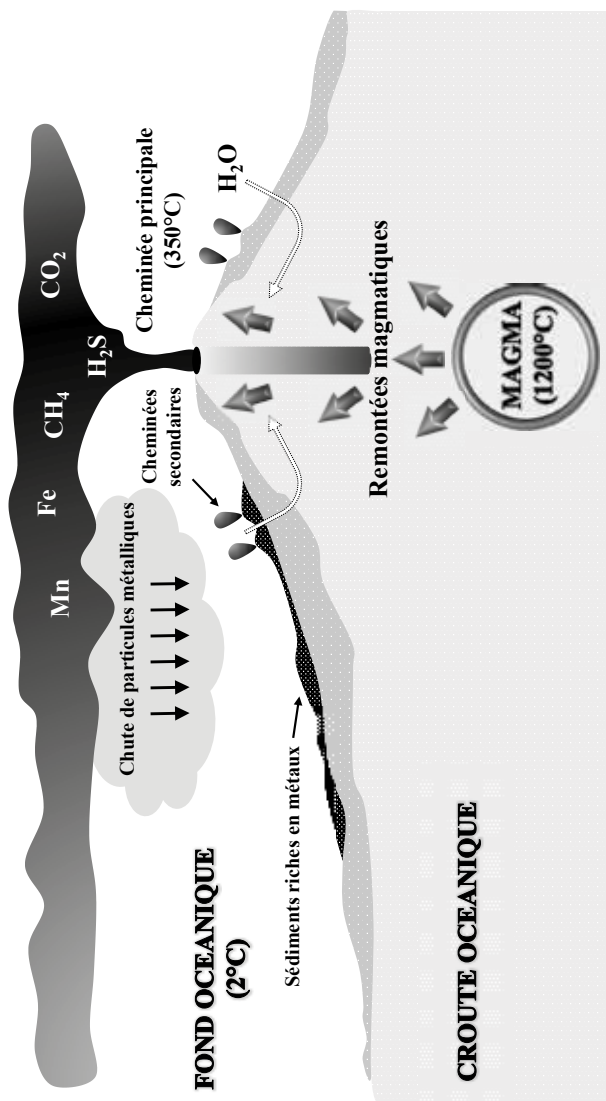


Figure 2 | L'origine volcanique sous-marine des molécules prébiotiques. Les remontées magmatiques provoquent une intense activité volcanique sous-marine qui relâche au fond des océans des molécules de sulfure d'hydrogène (H_2S), de gaz carbonique (CO_2), de méthane (CH_4), ainsi que des métaux comme le fer (Fe) ou le manganèse (Mn). Les particules métalliques éjectées sédimentent sur le plancher océanique, formant des surfaces rocheuses riches en fer et en soufre, propices aux réactions chimiques prébiotiques. Les infiltrations d'eau de mer (flèches en pointillé) permettent un brassage continu entre le soufre contenu dans l'eau et le sulfure d'hydrogène (H_2S) relargué par la cheminée volcanique.

monde fer-soufre reçoit un soutien certain de l'expérimentation et semble plus attractive que celle de la soupe primordiale. Néanmoins, d'intéressantes observations sur les météorites et plus récemment sur les comètes suggèrent qu'une partie des molécules organiques terrestres pourrait bien être en fait d'origine... extraterrestre.

L'ORIGINE EXTRATERRESTRE DE LA VIE

L'hypothèse de la panspermie, qui remonte à l'Antiquité, propose que la vie sur Terre soit d'origine extraterrestre. Avant toute chose, une clarification s'impose car sous le terme panspermie se trouve regroupé un certain nombre d'hypothèses plus ou moins farfelues. La radiopanspermie propose que des formes microscopiques d'organismes vivants, tels que des bactéries, puissent être propagées à travers l'espace, en provenance d'autres systèmes solaires, poussés par le flux de particules éjecté de l'atmosphère d'une étoile. Cette hypothèse se heurte à deux problèmes majeurs. Premièrement, il est impossible à un organisme vivant de survivre plus de quelques secondes dans l'espace sans être protégé des radiations solaires, ultraviolets et rayonnements cosmiques. Certains auteurs ont imaginé que de tels organismes pourraient voyager à l'intérieur d'astéroïdes, protégés ainsi d'un environnement délétère, c'est l'hypothèse de la lithopanspermie. Il faudrait en revanche imaginer que ces gros rochers aient voyagé entre les étoiles, s'affranchissant ainsi de la gravité de leur système stellaire originel. Deuxièmement, la radio – ou la lithopanspermie ne font que reculer le problème de la formation de la vie sur Terre à celui de la formation de la vie dans un autre système stellaire. D'autres hypothèses encore plus farfelues font intervenir le transport délibéré ou accidentel d'organismes vivants sur Terre par une civilisation extraterrestre infiniment plus évoluée que la nôtre (panspermie dirigée ou accidentelle). Insistons sur le fait que ces hypothèses ne reposent ni sur une expérimentation irréfutable, ni sur une argumentation scientifique solide et toutes font donc partie du domaine de la science-fiction.

La pseudopanspermie en revanche est plus attractive, car elle semble corroborée par l'observation scientifique. Cette hypothèse propose que les molécules organiques à l'origine des premiers êtres vivants aient été produites dans l'espace, dans des conditions totalement étrangères à celles que l'on peut trouver sur Terre. Nalin Chandra Wickramasinghe (1939-), mathématicien et astronome britannique, proposa dès 1974 que du formaldéhyde – une molécule organique carbonylée – créé dans les nuages interstellaires par des réactions chimiques particulières puisse se condenser sur des grains de **silicates*** (minéraux à base de silice, SiO_2) et de carbone dans la poussière gravitant autour des étoiles. L'étude des **chondrites*** carbonées tombées sur Terre ces cinquante dernières années a été particulièrement instructive, en particulier celle de la météorite de Murchison, du nom du village d'Australie près duquel elle s'est écrasée en 1969. Cette chondrite de grande taille a permis la récupération d'une centaine de kilogrammes de fragments dont le plus gros pesait sept kilos. L'analyse de plusieurs de ces fragments a montré la présence de cinq acides aminés relativement abondants : glycine, alanine, acide glutamique, valine et proline, ainsi que d'autres molécules organiques. Des analyses ultérieures ont également montré la présence de deux pyrimidines. Un des problèmes majeurs dans l'étude des chondrites est d'éviter les contaminations par des molécules organiques d'origine terrestre. Même dans le cas de la météorite de Murchison, considérée comme peu contaminée par les molécules du sol, un doute pouvait persister quant à la véritable origine extraterrestre des composés identifiés. C'est alors qu'il fallut se tourner vers l'espace.

Chondrites : Météorites non métalliques, divisées en plusieurs sous-groupes, dont les chondrites carbonées qui contiennent des proportions variables de carbone et d'oxygène. Les chondrites représentent environ 85 % des météorites qui s'écrasent sur Terre.

PHILAE A RENDEZ-VOUS AVEC TCHOURY

Le programme européen Rosetta visait à faire atterrir un petit appareil de mesure, nommé Philae, à la surface de la comète 67P/Churyomov-Gerasimenko (surnommée « Tchoury »), du nom des deux astronomes qui la découvrirent en 1969. Après dix années de voyage jusqu'à la comète, la mise en orbite de Rosetta eut lieu à l'été 2014 et l'atterrisseur Philae toucha le sol de Tchoury le 12 novembre 2014, devenant le premier engin spatial à atterrir sur une comète. Environ vingt minutes après que Philae ait touché la surface de la comète, les instruments de mesure de l'atterrisseur rapportaient la présence de seize molécules organiques, dont des molécules carbonylées et de nombreux composés azotés. Les analystes de Rosetta ont pu également analyser certaines des particules autour de Tchoury. De façon remarquable, les molécules organiques identifiées dans ces particules étaient de même nature que celles découvertes dans la météorite de Murchison. Ce résultat prouvait donc que les composés organiques des chondrites terrestres étaient bien de nature extraterrestre et non le résultat d'une contamination.

Un atterrissage mouvementé : Les dispositifs initialement prévus pour ancrer Philae sur le sol de Tchoury ne fonctionnèrent pas correctement. À cause de la faible gravité à la surface de la comète, l'atterrisseur rebondit une première fois jusqu'à une altitude de 1 km et après un vol plané de deux heures, rebondit une seconde fois avant d'être stoppé par une paroi, dans une position inclinée qui ne permettait pas l'éclairage complet de ses panneaux solaires. Philae remplit malgré tout la plupart de ses objectifs et put analyser la composition chimique de la surface de Tchoury.

Une étude européenne pilotée par un laboratoire niçois a réussi à reconstituer les conditions réactionnelles présentes au sein des comètes et des météorites : une température de -195 °C , une pression égale à 10 milliardièmes de la pression atmosphérique terrestre et une irradiation continue d'ultraviolets, en présence d'eau,

d'ammoniaque et de méthanol, un alcool simple. Le mélange réactionnel était ensuite lentement réchauffé jusqu'à la température de la pièce et les produits de la réaction analysés. La plupart des produits obtenus dans ces conditions expérimentales étaient des sucres, en particulier du ribose, un sucre composant essentiel de l'ARN, une des macromolécules les plus importantes du vivant, comme nous le verrons au chapitre suivant. Cette remarquable expérience confirme bien que des composés organiques complexes, futures briques des macromolécules du vivant, peuvent se former dans des conditions réactionnelles proches de celles existant au sein de comètes ou de météorites, renforçant encore l'hypothèse de la pseudopanspermie.

À RETENIR

- L'hypothèse de la « soupe primordiale » de Stanley Miller se heurte à plusieurs critiques, qui la rendent peu vraisemblable en tant que génératrice des premières molécules organiques.
 - Les analyses effectuées sur la comète Tchoury montrent que des molécules organiques impliquées dans la structure des organismes terrestres primordiaux existent bien dans l'espace interstellaire.
 - Ces molécules ont été créées dans des conditions de pression, de température et d'irradiation qui existaient lors de la formation du système solaire, il y a environ 4,6 milliards d'années.
 - La possibilité de générer des composés organiques sur des minéraux riches en fer et en soufre apparaît comme une autre voie, cette fois-ci d'origine terrestre, pour créer de la matière organique.
- Ainsi, il n'y aurait pas une origine unique à la vie sur Terre, mais au moins deux origines différentes, qui travaillant indépendamment – ou pas – l'une de l'autre ont contribué, pendant des millions d'années, à accumuler les premières molécules qui allaient devenir la base de tous les organismes vivants, des bactéries à l'homme.

BIBLIOGRAPHIE

L'expérience de Miller : Miller, *Science* 117 : 528-529 (1953).

La terre primordiale : Nisbet and Sleep, *Nature* 409: 1083-1091 (2001).

La météorite de Murchison : Kvenvolden *et al.*, *Nature* 228: 923-926 (1970).

Les molécules découvertes sur Tchoury : Goessmann *et al.*, *Science* 349: aab0689-1 (2015)

2

Des premières molécules organiques aux premières molécules du vivant

UNE DÉFINITION DU VIVANT

Avant d'aller plus loin, il est sans doute bon de définir les caractéristiques du vivant, de la même façon que nous avons précédemment défini la nature des molécules organiques qui le composent. Nous parlerons ici du « vivant » plutôt que de la « vie », afin de ne pas entrer dans un débat long et complexe, aux ramifications écologiques, sociologiques et philosophiques inévitables, sur la nature de la vie. Nous nous en tiendrons à une définition réductionniste qui, si elle est critiquable car éludant la complexité du problème pris dans son ensemble, veut servir de base à une réflexion centrée sur les organismes pris de façon individuelle plutôt que collective. Nous définirons donc la propriété principale du vivant par la capacité à se reproduire presque à l'identique, seule et unique propriété commune à tous les organismes. La reproduction est la seule et unique finalité de tout être vivant, c'est leur caractéristique intrinsèque, celle qui les définit comme tels. Nous savons que l'homme a pu s'affranchir en

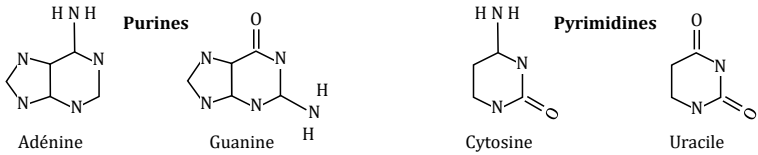
partie de cette fonction en programmant sa reproduction, grâce à la maîtrise biologique de son propre corps et à la maîtrise sociale de son écosystème. Mais les autres êtres vivants n'ont pas cette possibilité, se reproduire est leur seule et unique finalité.

Cette propriété fondamentale étant donc maintenant définie, on peut en déduire ses deux caractéristiques essentielles : la première nécessite qu'un organisme vivant soit porteur de l'information qui le définit et la deuxième qu'il soit capable de reproduire cette information *ad infinitum*, ou presque. Les molécules organiques simples du chapitre précédent ne possèdent aucune de ces deux caractéristiques. Elles ne font donc pas partie du vivant. Il est donc légitime de poser la question suivante : existe-t-il une molécule organique qui possède ces deux propriétés ?

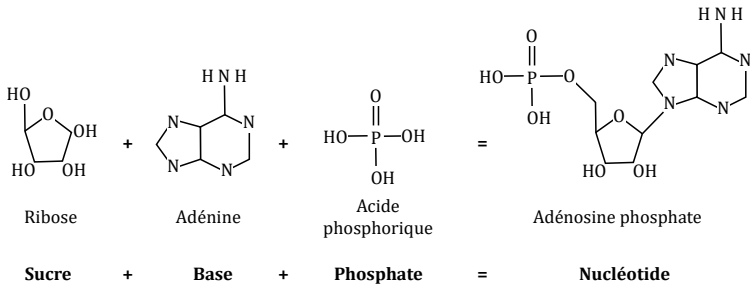
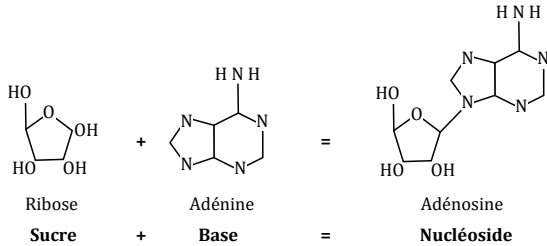
L'ARN PORTEUR D'INFORMATION

À cette étape de l'évolution, il est temps de présenter une des macromolécules les plus importantes du vivant : l'acide ribonucléique ou ARN. La structure de l'ARN est basée sur un squelette fait de sucres (le ribose, qui donne son nom à la molécule) reliés entre eux par une liaison particulière : la liaison phosphodiester. À ces sucres sont attachées les purines et les pyrimidines qui forment les bases de l'ARN. Les quatre bases sont l'adénine et la guanine (purines), la cytosine et l'uracile (pyrimidines), généralement abrégées A, G, C et U. Ces bases attachées au ribose sont appelées nucléosides (ou ribonucléosides). Ainsi, l'adénine attachée au ribose est appelée adénosine. Finalement, l'ajout d'un groupement phosphate transforme un nucléoside en **nucléotide*** (Figure 3).

Nucléotide : On appelle nucléotide une base liée à un sucre et un, deux ou trois groupements phosphate. Les quatre nucléotides de l'ARN (ou ribonucléotides) sont l'adénosine, la cytidine, la guanosine et l'uridine, dont les noms dérivent de la base correspondante.



Les quatre bases de l'ARN



H: Hydrogène O: Oxygène (les atomes de carbone ne sont pas représentés)	N: Azote P: Phosphore
---	--------------------------

Figure 3 | Bases de l'ARN, nucléosides et nucléotides.

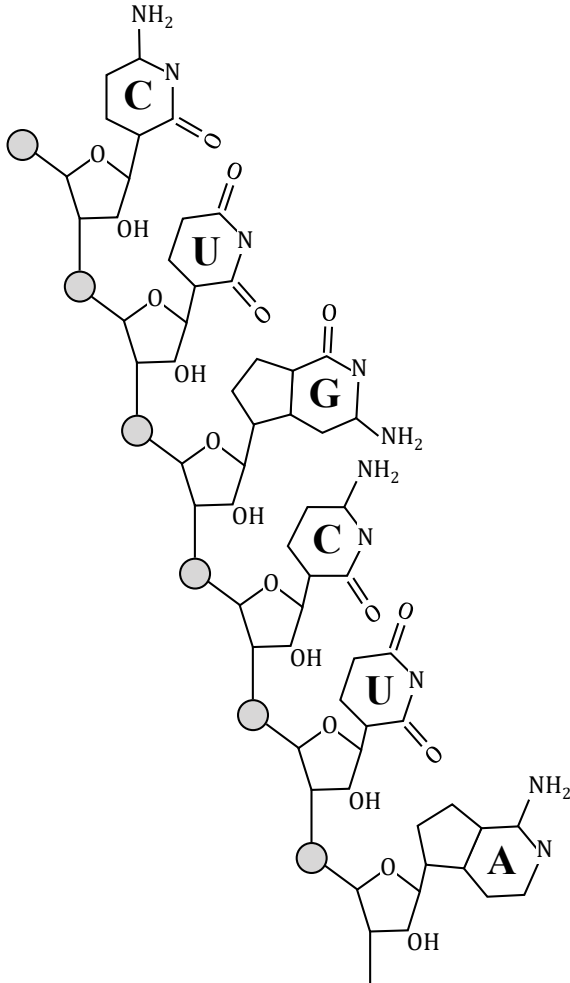


Figure 4 | La molécule d'ARN. Un enchaînement de six nucléotides est représenté. Les phosphores de la liaison phosphodiester sont représentés par des cercles gris. A = adénine, C = cytosine, G = guanine, U = uracile. La séquence des quatre bases représentées ici se lit de haut en bas : CUGCUA.

Une façon de se représenter l'ARN est d'imaginer une guirlande électrique portant des ampoules de quatre couleurs différentes. Les douilles (molécules de ribose) sont reliées entre elles par le fil électrique (liaisons phosphodiesters) et sur celles-ci sont fixées des ampoules colorées (les quatre bases) (Figure 4). La succession des bases le long de la molécule d'ARN détermine sa **séquence***. Celle-ci représente l'information génétique portée par la molécule d'ARN.

Séquence : On appelle séquence d'un ARN ou d'un ADN l'ordre précis de l'enchaînement des quatre bases le long de la molécule. On appelle séquence d'une protéine l'ordre précis de l'enchaînement des acides aminés le long de la molécule.

L'ARN CATALYSEUR DE RÉACTIONS CHIMIQUES

La guirlande formée par la molécule d'ARN est flexible, de sorte qu'elle peut s'enrouler sur elle-même, de façon à pouvoir prendre à peu près n'importe quelle conformation. La forme prise par la guirlande – que l'on appellera structure secondaire – dépend exclusivement de sa séquence. En effet, les bases peuvent effectuer des liaisons entre elles, une purine s'appariant avec une pyrimidine : l'adénine peut ainsi faire deux liaisons avec l'uracile quand la cytosine peut faire trois liaisons avec la guanine. De l'appariement des bases de l'ARN découle la structure de la molécule, les bases complémentaires étant appariées entre elles dans des régions dites « en tige-boucle » qui contiennent donc un brin double d'ARN, les autres bases non appariées formant des boucles contenant l'ARN sous forme de brin simple. La structure d'un ARN découlant de sa séquence, chaque ARN aura une structure unique ou presque, certaines séquences très proches pouvant partager des structures similaires (Figure 5).

La structure secondaire de l'ARN lui permet de se replier dans l'espace et de catalyser certaines réactions chimiques. Il est donc

découverte que certaines molécules d'ARN étaient capables de catalyser une réaction chimique révolutionnaire en biologie dans les années 1980. Thomas Cech, un chimiste américain qui obtint le prix Nobel de Chimie en 1989 avec Sydney Altman, montra qu'un ARN particulier venant d'un **protozoaire*** cilié – *Tetrahymena thermophila* – se comportait comme le catalyseur d'une réaction ajoutant une cytosine à une chaîne d'ARN de cinq cytosines, allongeant effectivement la chaîne d'une base. La réaction était catalysée par l'ARN, pouvait être itérative et réversible. Par la suite, il fut montré que d'autres molécules d'ARN pouvaient également jouer le rôle de catalyseurs. Ainsi, l'ARN est non seulement une molécule qui porte l'information de sa propre structure, mais cette structure est également capable de catalyser des réactions biochimiques complexes. Ces découvertes menèrent à proposer l'ARN comme première molécule du vivant.

LE MONDE À ARN

Jusqu'aux années 1980, le dogme central de la biologie postulait que le programme génétique de tous les organismes vivants était contenu dans leur ADN, un peu comme le code source d'un logiciel informatique définit ses caractéristiques et son mode de fonctionnement. L'ADN est une macromolécule organique qui ressemble de très près à l'ARN, à deux différences chimiques près, sur lesquelles nous reviendrons plus tard. Mais si l'ADN est le support de l'hérédité des organismes vivants – il répond donc bien à la première caractéristique du vivant, être porteur de l'information nécessaire à sa propre reproduction –, en revanche il ne possède pas naturellement la seconde propriété qui est la capacité à reproduire cette information. L'ARN, à la fois doté de la capacité à stocker l'information et de la possibilité de la reproduire par son activité catalytique, possède bien ces deux caractéristiques. C'est donc en toute logique que Walter Gilbert proposa en 1986 l'existence d'un ancestral « monde à ARN ». Cette hypothèse est séduisante pour plusieurs raisons, mais laisse ouvertes certaines questions.

Des molécules d'ARN trouvés dans des ciliés, des bactéries ou des levures sont encore à l'heure actuelle dotées d'activité catalytique et peuvent être considérées comme des reliques de ce monde à ARN. Certaines d'entre elles ont en partie perdu au cours de l'évolution leur capacité à se replier selon une structure secondaire bien précise et nécessitent la présence de protéines pour maintenir la structure catalytique en place. Dans le **ribosome***, l'énorme machinerie qui synthétise les protéines, c'est l'ARN qui catalyse la formation de la **liaison peptidique*** entre les acides aminés, les protéines servant principalement à maintenir l'ARN dans sa conformation active. Il est vraisemblable que le ribosome ancestral du monde à ARN ne contenait aucune protéine et que l'ARN seul était capable de se replier dans la structure catalytique requise.

Ribosome : Énorme machine macromoléculaire responsable de toute la synthèse des protéines dans une cellule. Chaque ribosome est constitué de quatre molécules d'ARN et de 79 protéines. Chaque cellule contient plusieurs milliers de ribosomes.

Le fait que l'ARN soit à la fois dépositaire de l'information nécessaire à sa propre structure et doté d'activité catalytique est certainement un argument de poids pour en faire une des toutes premières molécules du vivant. Expérimentalement, à partir d'une molécule d'ARN dotée d'une activité catalytique limitée à la synthèse de 14 bases, on peut sélectionner au cours d'étapes successives une molécule d'ARN capable de reproduire 95 bases d'une autre molécule d'ARN. De même, la capacité de l'ARN à catalyser la formation d'une liaison peptidique entre deux acides aminés permet d'expliquer facilement la transition du monde à ARN au monde des protéines, des catalyseurs bien plus diversifiés et bien plus efficaces que les ribozymes. Néanmoins, la théorie du monde à ARN, bien que très séduisante pour les raisons évoquées ci-dessus, nécessite d'expliquer l'origine de la première molécule d'ARN suffisamment longue pour catalyser une réaction chimique.

DES PREMIÈRES MOLÉCULES ORGANIQUES AU PREMIER ARN

Nous avons vu au chapitre précédent que de sérieux arguments scientifiques contredisaient le modèle de la soupe primordiale de Miller. Un de ces principaux arguments repose sur la faible concentration des molécules organiques ainsi formées, limitant leur possibilité d'interaction. C'est un argument un peu similaire qui est repris par les critiques du monde à ARN. Si on admet qu'à un certain moment dans l'histoire du vivant sont apparus les premières pyrimidines, puis les premiers ribonucléotides, comment ces molécules se sont-elles assemblées pour former le premier ARN ?

En solution aqueuse, la formation de polymères de ribonucléotides par condensation de monomères est impossible car elle entre en compétition avec leur **hydrolyse***.

Hydrolyse : Réaction chimique catalysée par une molécule d'eau qui détruit une liaison entre deux atomes.

En revanche, des monomères de ribonucléotides peuvent se polymériser à la surface d'une argile, la montmorillonite. Cette argile tire son nom de la ville de Montmorillon dans la Vienne, mais est également appelée terre de Sommières, nom éponyme d'une localité du Gard. On trouve également de la montmorillonite aux États-Unis (Utah, Wyoming, Texas) et le rover Spirit en a découvert sur la planète Mars. Cette argile est composée de silicate d'aluminium et de magnésium. Elle a la caractéristique de gonfler en présence d'eau, lui conférant des propriétés intéressantes de plasticité. Des polymères de ribonucléotides ont été obtenus par condensation d'un dérivé d'adénosine en présence de montmorillonite. Les auteurs ont observé la formation de trimères, tétramères et pentamères d'adénosine, en trois jours et à température ambiante. En revanche, la taille des polymères obtenus était trop courte pour former des molécules. Cette expérience encourageante fut suivie quelques années plus tard par la synthèse de polymères contenant 55 molécules d'adénosine sur des montmorillonites, par un processus itératif et dans des

conditions expérimentales sans doute assez proches de celles d'une terre primitive. Ainsi, l'hypothèse qu'un ARN primitif puisse être assemblé à partir des ribonucléotides le constituant est soutenue par une démonstration expérimentale, transférant cette hypothèse dans le domaine du possible. Néanmoins, dans ces expériences, la synthèse des ribonucléotides était limitée au même monomère (en l'occurrence une adénosine). La synthèse alternative des quatre ribonucléotides ne pouvait être obtenue que dans des conditions extrêmes, peu compatibles avec celles de la chimie prébiotique. Or, afin d'obtenir des ARNs capables de se replier selon une structure secondaire et donc de servir de ribozymes, il faut pouvoir incorporer les quatre ribonucléotides (Figure 5). L'hypothèse du monde à ARN fut dès lors remise en question, jusqu'à ce qu'une approche totalement différente la remette au centre du débat.

Il a récemment été découvert un peu fortuitement que la ligature de monomères, dimères ou trimères de ribonucléotides sur un ARN servant de matrice complémentaire était très efficace, dans des conditions réactionnelles compatibles avec une chimie prébiotique. Cette réaction est particulièrement active avec des trimères de nucléotides (appelés trinucleotides). Ainsi, on peut imaginer que ce type de réaction se soit produit dans un mélange contenant des monomères, des dimères et des trimères de ribonucléotides, permettant la reproduction d'un ARN contenant les quatre bases, ouvrant ainsi la voie à la formation de ribozymes, ARNs plus complexes, dotés d'activité enzymatique (Figure 6).

Par ce processus itératif, on peut facilement synthétiser des ARNs contenant les quatre ribonucléotides, afin d'obtenir des ribozymes. Ceux-ci vont alors catalyser leur propre synthèse par le même procédé. À partir de là, la voie est ouverte à la synthèse de nombreux ARNs dotés d'activité catalytique, dont les plus efficaces ont été sélectionnés car capables de se reproduire plus rapidement. La sélection naturelle commençait son œuvre ; elle allait la continuer avec l'apparition des premières cellules.

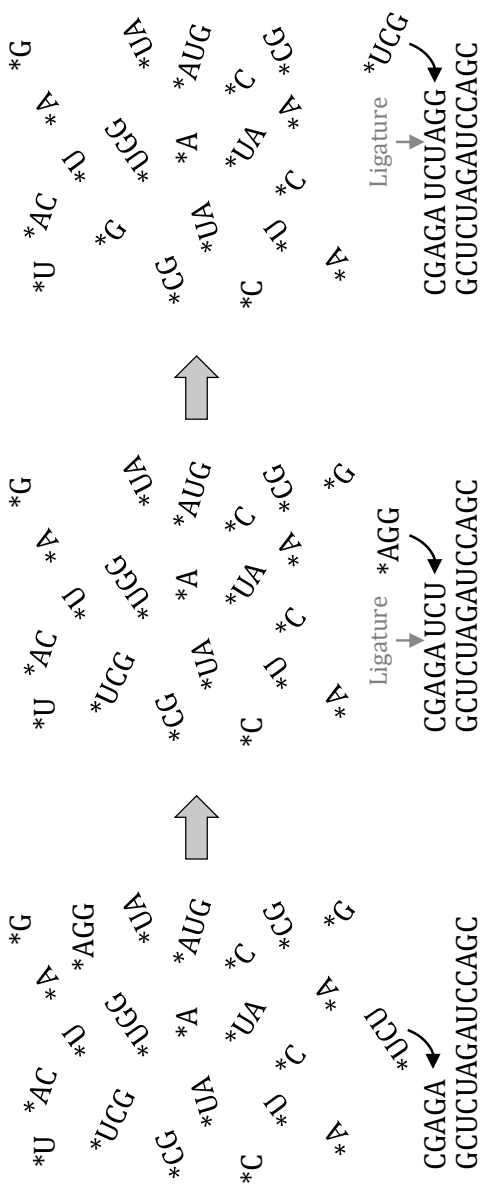


Figure 6 | Synthèse d'ARN par additions successives de trinucleotides. L'astérisque représente le phosphate qui permet la ligature.

À RETENIR

- Le principal attribut d'un être vivant est sa capacité à se reproduire – presque – à l'identique. C'est son unique finalité. Cette caractéristique devait déjà être présente dans les premières molécules du vivant.
- L'ARN est la seule molécule du vivant à la fois dotée d'activité catalytique et porteuse de sa propre information génétique. Cette remarquable propriété a mené à l'hypothèse du « Monde à ARN ».
- Les premiers ARNs furent peut-être synthétisés par addition de trinucleotides, avant l'apparition des premiers ribozymes, littéralement « enzymes ARN », dotés d'activités catalytiques permettant une plus grande diversité de réactions chimiques.

BIBLIOGRAPHIE

- La structure secondaire des ARN : Michel, Jacquier et Dujon, *Biochimie* 64 : 867-881 (1982).
- Le Monde à ARN : Gilbert, *Nature* 319: 618 (1986).
- La synthèse d'un ARN à partir de monomères et trimères de ribonucleotides : Prywes *et al.*, *eLife* 5: e17756 (2016).
- La synthèse enzymatique d'un ribozyme à partir de trinucleotides : Attawater *et al.*, *eLife* 7: e35255 (2018).

3

Du monde à ARN aux premières cellules

La cellule est la plus petite unité fonctionnelle d'un organisme vivant capable de se reproduire de façon autonome. Le mot « autonome » exclut de fait les virus, car ceux-ci ne peuvent se reproduire sans l'aide de la machinerie cellulaire de leur hôte. On distingue deux types de cellules : les **procaryotes*** qui ne contiennent pas de noyau et les **eucaryotes*** qui en possèdent un. C'est un zoologiste français, Édouard Chatton (1883-1947), chercheur à l'Institut Pasteur, qui propose en 1925 de classer les **protistes*** en deux groupes : les procaryotes et les eucaryotes. Néanmoins, les contours de ces deux domaines sont restés flous pendant plusieurs décennies. La première définition stricte des procaryotes et des eucaryotes a été proposée en 1962, par Roger Stanier (1916-1982) et Cornelius

Procaryote : Du latin pro « avant » et du grec karyon « noyau ».

Eucaryote : Du grec eu « bien » et karyon « noyau ». Eucaryote signifie donc littéralement « qui a un bon, un vrai noyau ».

Van Niel (1897-1985). Pour des raisons académiques et politiques, Roger Stanier quitta les États-Unis en 1971 à l'invitation d'Elie Wollman, pour prendre la direction de l'ancien laboratoire de Jacques Monod et François Jacob à l'Institut Pasteur. Il y étudia les **cyanobactéries*** jusqu'à sa disparition.

Avant cette époque, des confusions existaient sur la définition et la classification du monde bactérien. Une bactérie fut alors définie comme une entité cellulaire limitée par une paroi, contenant de l'ADN et de l'ARN, fabriquant sa propre énergie et se reproduisant par division. Cette définition était suffisante pour distinguer les bactéries des virus, mais insuffisante pour les distinguer des cellules contenant un noyau. Stanier et Niel proposèrent alors de définir un eucaryote comme une cellule contenant des sous-structures intracellulaires entourées d'une membrane individuelle faisant barrière entre ces structures et le reste de la cellule. Cette définition fera date et est d'ailleurs toujours valide à l'heure actuelle, même si l'on a tendance à définir une cellule eucaryote plus simplement comme une cellule contenant un noyau bien différencié contenant le matériel génétique sous forme d'ADN, alors que dans les cellules procaryotes celui-ci n'est pas contenu dans un compartiment spécifique. Les organismes procaryotes sont plus communément appelés bactéries et on les divise maintenant en deux règnes : les **archéobactéries*** (ou archées) et les **eubactéries***.

L'ARBRE DU VIVANT

Jusqu'en 1977, il était considéré que tous les êtres vivants appartenaient soit au monde procaryote soit aux eucaryotes. Carl Woese (1928-2012) et George Fox (1945-) proposèrent à cette date une nouvelle classification du vivant basée sur la séquence de l'ARN ribosomique. Les trois séquences eucaryotes disponibles à l'époque se ressemblaient plus entre elles et étaient clairement plus éloignées des séquences procaryotes. Mais de façon très claire, les séquences d'ARNs procaryotes formaient deux groupes bien distincts, aussi

éloignés l'un de l'autre que du groupe des eucaryotes. L'un de ces deux groupes contenait *Escherichia coli* ainsi que d'autres bactéries typiques, lactobacille, mycoplasme, streptocoque, vibrio, etc. Woese et Fox appelèrent ces bactéries *eubactéries* (bonnes, vraies bactéries). Le deuxième groupe bactérien contenait quatre bactéries anaérobies, utilisant le dioxyde de carbone (CO₂) comme source d'énergie. Ils notent que ce métabolisme est parfaitement adapté au type d'environnement qui devait exister sur Terre il y a 3-4 milliards d'années et nomme ces bactéries *archéobactéries* (bactéries primitives, originelles), appelées maintenant plus couramment *archées*. Cette découverte d'un troisième domaine du vivant allait profondément et durablement changer notre vision du monde. La robustesse de ce que l'on appelle encore de nos jours « l'Arbre du Vivant de Woese » a été maintes fois éprouvée par l'arrivée de nouvelles séquences d'ARN ribosomique. Certains aménagements mineurs y ont été apportés au cours du temps. Il existe des embranchements à l'intérieur des archées et la classification des eucaryotes est encore le sujet de nombreux débats. On les classait autrefois en quatre embranchements : plantes, animaux, champignons et protistes. Cette classification, essentiellement basée sur des caractères biologiques communs, n'a plus de sens à l'heure actuelle, alors que des centaines de séquences de génomes eucaryotes sont disponibles. Il est plus logique d'utiliser une classification moléculaire basée sur l'analyse et la comparaison de ces séquences entre elles. Nous reviendrons largement sur ces méthodes de comparaison et de classification quand nous aborderons les concepts liés à la **phylogénie*** moléculaire.

LES SPÉCIFICITÉS DU MONDE EUCARYOTE

Les eucaryotes, comme dit précédemment, doivent leur nom à l'existence d'un noyau qui englobe et protège l'information génétique, la séparant physiquement du reste de la cellule. Ce n'est pas leur seule particularité. Tous les organismes eucaryotes possèdent des sous-compartiments cellulaires appelés **mitochondries*** qui sont

responsables de la synthèse de l'énergie utilisée par la cellule. Les transferts d'électrons à travers la paroi mitochondriale déclenchent la synthèse du carburant cellulaire : l'adénosine triphosphate ou **ATP***. Il est remarquable de noter que cet ATP est également un des quatre nucléotides de l'ARN et on conçoit donc aisément qu'il a vraisemblablement été une des toutes premières sources d'énergie du monde à ARN.

Presque tous les organismes eucaryotes possèdent des mitochondries, à l'exception de ceux que l'on appelle parfois **archézoaires***. De plus, parmi les eucaryotes, les végétaux possèdent une deuxième machine à fabriquer de l'énergie : les **chloroplastes***. Ces organites intracellulaires sont responsables de la **photosynthèse***, réaction qui transforme l'énergie lumineuse en sucre et en oxygène. C'est grâce à la photosynthèse effectuée par les premières cellules végétales que l'atmosphère de la Terre primitive changera radicalement, comme nous le verrons dans un chapitre suivant. Alors que le monde bactérien est totalement dénué de noyau, de mitochondries et de chloroplastes, le monde eucaryote en est abondamment pourvu et on est en droit de se questionner sur l'origine de ces organites intracellulaires. Nous verrons un peu plus tard que l'origine du monde eucaryote est un des plus fascinants problèmes qui se pose aux biologistes de l'évolution et il n'est pas encore tout à fait résolu à l'heure actuelle.

Photosynthèse : Processus métabolique permettant aux algues, aux plantes et à certaines bactéries (cyanobactéries) de synthétiser de la matière organique grâce à la lumière du soleil.

Ainsi, l'Arbre du Vivant de Woese a été affiné par les données les plus récentes du séquençage de nombreux organismes eucaryotes. On peut en trouver une représentation très simplifiée sur la figure ci-dessous.

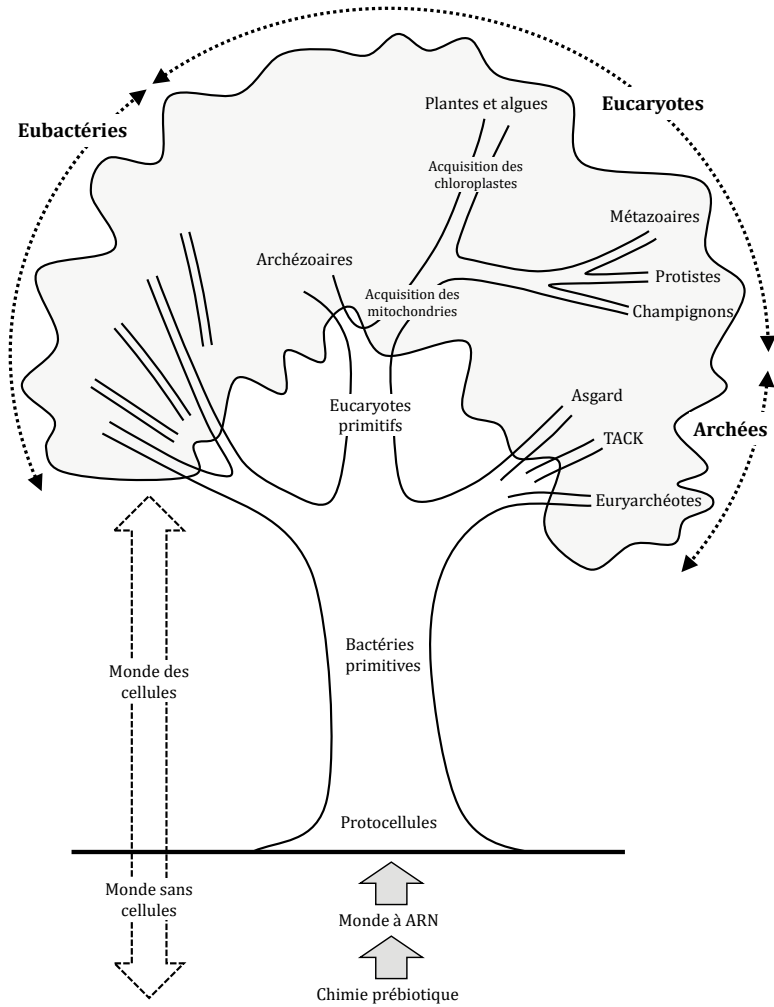


Figure 7 | L'Arbre du Vivant.

LES PREMIÈRES CELLULES

Pour passer d'un monde ouvert au monde fermé des cellules, il fallait cloisonner les réactions chimiques prébiotiques dans un espace confiné dans lequel l'acidité (représentée par le **pH***) et la concentration des différentes molécules et composés chimiques étaient mieux contrôlées. Il fallait donc inventer les membranes cellulaires. Chaque cellule devient alors une mini-usine dont la fonction unique est de se reproduire et de reproduire l'information génétique qu'elle contient. La membrane d'une cellule a une composition bien particulière, très différente de la composition chimique de l'ARN, de l'ADN ou des protéines. Elle est formée de lipides, en particulier de phospholipides, constitués d'une tête **hydrophile*** et d'une longue queue **hydrophobe***. De façon tout à fait remarquable, la tête des phospholipides de la membrane des archéobactéries est une molécule de glycérol-1 phosphate (G1P), alors que chez les eubactéries il s'agit de glycérol-3 phosphate (G3P), une molécule qui est l'exact reflet dans un miroir du G1P (Figure 8). Or, les enzymes impliquées dans la fabrication du G1P ou du G3P sont très spécialisées, celle qui fabrique l'un est incapable de fabriquer l'autre. À partir de cette observation, on peut émettre deux hypothèses. Soit la première cellule contenait les deux phospholipides et chacun des deux lignages procaryotes a par la suite conservé un seul des deux, soit la séparation des lignées bactériennes précède l'établissement des membranes et dans ce cas il faut imaginer des premières cellules... sans membrane.

pH : Le potentiel hydrogène ou pH est une mesure de l'acidité d'une solution.

Hydrophile : Composé ayant une affinité pour l'eau et tendance à s'y dissoudre. Le sucre alimentaire est hydrophile.

Hydrophobe : Composé n'ayant pas d'affinité pour l'eau et ne pouvant s'y dissoudre. Les huiles et les graisses sont hydrophobes.

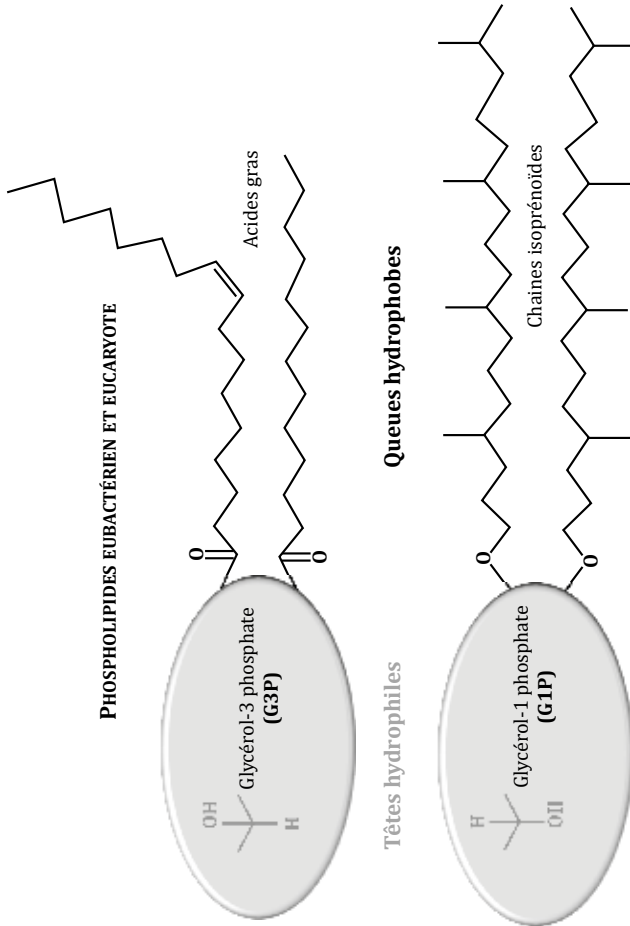


Figure 8 | Les phospholipides des membranes bactériennes et eucaryotes. Haut : structure des phospholipides eubactérien et eucaryote. La tête hydrophile est constituée de glycérol-3 phosphate (G3P), reliée aux acides gras hydrophobes. Bas : structure d'un phospholipide archéobactérien. La tête hydrophile est constituée de glycérol-1 phosphate, reliée aux chaînes hydrophobes.

L'ORIGINE POSSIBLE DES PREMIÈRES MEMBRANES CELLULAIRES

Plusieurs auteurs, dont Wächterhäuser, ont proposé que le glycérol phosphate sous ses deux formes (G1P et G3P) ait pu être à l'origine synthétisé chimiquement sur des surfaces de pyrite (FeS_2), grâce à la chimie fer-soufre dont nous avons précédemment discuté. Ces lipides resteraient fortement collés à la surface du minéral par leur nature fortement hydrophobe et s'accumuleraient donc à la surface, augmentant leur hydrophobicité. La diminution locale de la concentration des molécules d'eau favoriserait la condensation des acides aminés et des ribonucléotides en précurseurs de protéines et d'ARN (Figure 9, Étape 1). Wächterhäuser suggère ensuite que l'augmentation de la concentration de surface des lipides, particulièrement des lipides à longue chaîne hydrophobe, entraînerait la séparation de la phase hydrophobe qui se retrouverait flottant dans une phase hydrophile. Les molécules organiques emprisonnées dans ces particules lipidiques se retrouveraient alors isolées du milieu extérieur, permettant le maintien d'un pH neutre à peu près constant, favorable aux réactions biochimiques (Figure 9, Étape 2). Si la cellule primitive ainsi formée se retrouve dans un environnement liquide peu salé, elle aura tendance à se remplir d'eau et sa membrane finira par se rompre par choc osmotique.

Choc osmotique : Passage rapide d'un milieu riche en sels vers un milieu pauvre en sels, entraînant une entrée soudaine d'eau dans la cellule qui peut rompre l'intégrité de sa membrane et entraîner la mort cellulaire.

Cette cellule primitive ne pourra donc survivre que dans un milieu suffisamment salé pour éviter cela. Les membranes de ces ancêtres de cellules (ou protocellules) contiendraient à l'origine autant de G1P que de G3P, les deux lipides étant fabriqués dans les mêmes proportions sur les surfaces de pyrite. Ces protocellules pouvaient facilement incorporer des lipides se trouvant dans le

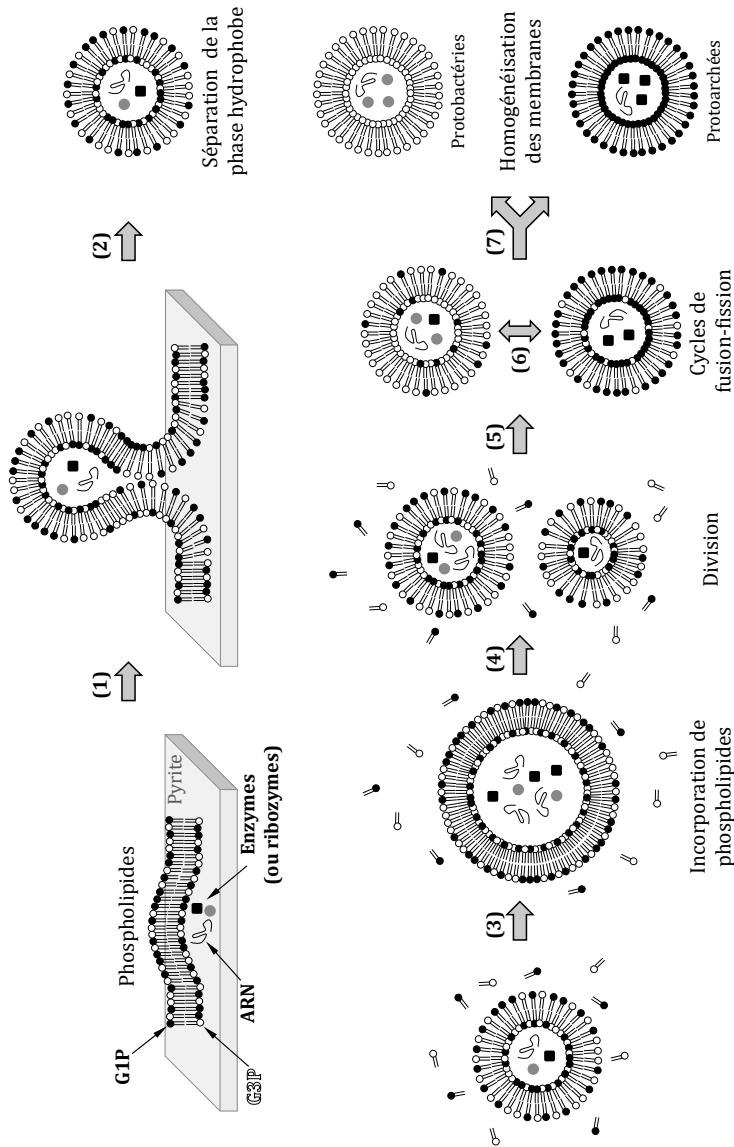


Figure 9 | La formation des membranes des premières cellules.

milieu environnant, comme cela a été montré dans des systèmes modèles (Figure 9, Étape 3). L'augmentation de taille des cellules était contrebalancée par leur division, sous l'effet de forces mécaniques internes ou externes (Figure 9, Étape 4). Dans les conditions de température élevée qui régnaient dans les sources chaudes océaniques, ces protocellules devaient subir des cycles de fusion-fission qui entraînaient petit à petit, la séparation des protocellules en deux sous-populations, l'une contenant essentiellement du G1P et l'autre du G3P, car la stabilité des membranes formées d'un seul des deux phospholipides (G1P ou G3P) est plus grande que celle de membranes formées par les deux, pour des raisons strictement physico-chimiques (Figure 9, Étapes 5-7). Cette plus grande stabilité membranaire aurait permis de sélectionner une enzyme capable de catalyser la formation de l'un ou l'autre type de phospholipide, fixant de façon définitive le G1P dans la protocellule archéobactérienne et le G3P dans la protocellule eubactérienne. L'autre effet direct des cycles de fusion-fission aurait été de mélanger l'information génétique (ARN ou ADN) entre les différentes protocellules, augmentant les combinaisons possibles tout en homogénéisant leur contenu. Ainsi, deux lignées de protocellules, l'une menant aux archées et l'autre aux eubactéries auraient été simultanément sélectionnées au cours de millions d'années de cette transition du monde prébiotique au monde vivant.

À QUOI RESSEMBLAIENT LES PROTOCELLULES ?

L'apparition des premières membranes marque la transition entre le monde prébiotique et le monde vivant. Même si de nombreuses questions restent ouvertes concernant cette transition, les progrès récents dans ce domaine ont permis d'affiner entre les différentes hypothèses qui existaient il y a encore quelques années et le scénario présenté ici est actuellement le plus vraisemblable. Certaines questions demeurent encore, telle l'apparition sélective des deux familles d'enzymes différentes pour synthétiser les deux types de membranes

du monde vivant : les membranes à G1P et celles à G3P. On peut aussi se demander si les premières cellules contenaient uniquement de l'ARN, de l'ARN et des protéines ou bien déjà de l'ADN ? S'il est délicat de répondre à cette question, il est vraisemblable que les premiers ribozymes aient été plus ou moins rapidement enclos dans des systèmes fermés par une membrane lipidique. Cette étape présente l'énorme avantage d'isoler partiellement le ribozyme du milieu extérieur et donc de contrôler les conditions de température, de salinité et de pH des protocellules à ARN. À partir de là, il est possible de voir apparaître un protoribosome qui va permettre la synthèse des premières protéines qui vont fournir les activités enzymatiques permettant de passer au monde à ADN, comme nous le verrons au chapitre suivant.

Nul doute que la sélection naturelle des premières cellules les plus aptes à se diviser rapidement en conservant leur matériel génétique a joué un rôle majeur à cette étape de l'évolution du vivant. L'arrivée de l'ADN, plus stable que l'ARN, va le remplacer en réduisant les erreurs faites lors de la reproduction de l'information génétique et en permettant d'autre part à l'ARN de jouer un autre rôle, celui de transfert de l'information génétique de l'ADN aux protéines, comme nous le verrons au chapitre 6.

À RETENIR

- Le monde vivant est constitué de trois domaines : eubactéries, archées et eucaryotes.
- Les membranes des eucaryotes et des eubactéries ont une composition lipidique similaire et ont donc vraisemblablement une origine commune. Celles des archées n'ont pas la même composition.
- Les premières cellules contenaient vraisemblablement des ARNs de type ribozyme, et devaient être capables de fabriquer ou d'assimiler les deux types de lipides.

BIBLIOGRAPHIE

- L'Arbre du Vivant de Woese : Woese and Fox, *Proceedings of the National Academy of Sciences USA* 74: 5088-5090 (1977).
- La synthèse des lipides et l'évolution des membranes : Pereto *et al.*, *Trends in Biochemical Sciences* 29: 469-477 (2004).
- Synthèse et croissance de modèles de protocellules : Mansy *et al.*, *Nature* 454: 122-126 (2008).

4

Du monde à ARN au monde à ADN

LES REMARQUABLES PROPRIÉTÉS DE L'ADN

À l'exception de certains **virus*** dont le matériel génétique est composé d'ARN, l'intégralité des organismes vivants actuels utilise l'ADN comme support de l'hérédité. Il est donc temps d'introduire la star de l'information génétique : l'ADN ou acide désoxyribonucléique. Il présente deux particularités chimiques et une particularité structurale importantes qui le différencient de son frère aîné l'ARN. Il contient un atome d'oxygène en moins sur le deuxième carbone du ribose, d'où le nom de *désoxyribose*. Cette modification chimique de son squelette sucré peut

Virus : Un virus est un organisme simple qui contient une information génétique, sous forme d'ADN ou d'ARN, mais qui est incapable de la reproduire lui-même. Il est obligé de pénétrer une bactérie, une cellule animale ou végétale afin d'utiliser la machinerie de reproduction de son hôte involontaire pour se reproduire. C'est donc un parasite intracellulaire obligatoire, existant dans tous les domaines du vivant, des bactéries aux plantes et aux animaux.

paraître minime, mais elle a un impact majeur sur la stabilité de la molécule d'ADN. En effet, l'oxygène en question dans l'ARN peut réagir naturellement avec une molécule d'eau pour hydrolyser la liaison phosphodiester entre deux riboses et couper la chaîne d'ARN.

Dans l'ADN, cet oxygène n'existe pas, donc la chaîne de désoxyribose ne peut pas s'hydrolyser spontanément en présence d'eau. La chaîne d'ADN est donc chimiquement plus stable que la chaîne d'ARN. La deuxième propriété de l'ADN est que l'une de ses bases, l'uracile, y est remplacée par la thymine. Les trois autres bases de l'ARN (adénine, cytosine et guanine) sont identiques (Figure 10).

Le remplacement de l'uracile dans l'ARN par la thymine dans l'ADN augmente la stabilité de la molécule. En effet, en solution aqueuse, la cytosine se transforme naturellement en uracile au cours du temps. En l'absence de système de détection et de réparation de ce phénomène naturel, des changements de séquence (cytosine -> uracile) apparaîtraient au cours du temps, menant à des **mutations***, des modifications permanentes du patrimoine génétique.

Mutation : Modification permanente, accidentelle ou provoquée du génome d'un organisme vivant. Chez les microorganismes, toutes les mutations se transmettent à la descendance. Chez les organismes sexués, une mutation ne pourra se transmettre à la descendance que si elle touche la lignée germinale (ovules ou spermatozoïdes).

Il existe une machinerie capable de remplacer un uracile par une cytosine afin de restaurer la bonne séquence, mais pour que ce système de réparation des cytosines puisse émerger au cours de l'évolution sans causer trop de dégâts, il fallait auparavant supprimer les uraciles existant naturellement dans la molécule, afin de détecter les uraciles dus uniquement à la modification des cytosines. C'est pour cette raison que la thymine a émergé au cours de l'évolution pour remplacer l'uracile dans la molécule d'ADN, et que ce mécanisme de réparation des cytosines désaminées n'a pu voir le jour que quand la thymine eut remplacé l'uracile (Figure 11).

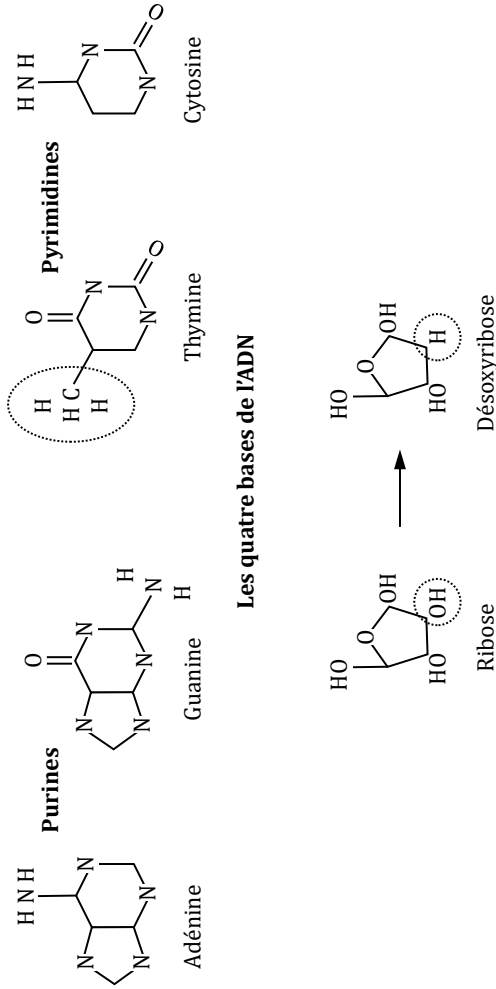


Figure 10 | Le désoxyribose et les quatre bases de l'ADN. Les quatre bases de l'ADN sont identiques à celles de l'ARN, à l'exception de l'uracile qui est remplacée par la thymine. La différence chimique entre les deux bases vient de la présence d'une fonction méthyle (CH₃) sur la thymine, absente de l'uracile (cercle en pointillé). Le ribose est transformé en désoxyribose par perte d'un oxygène (cercle en pointillé).

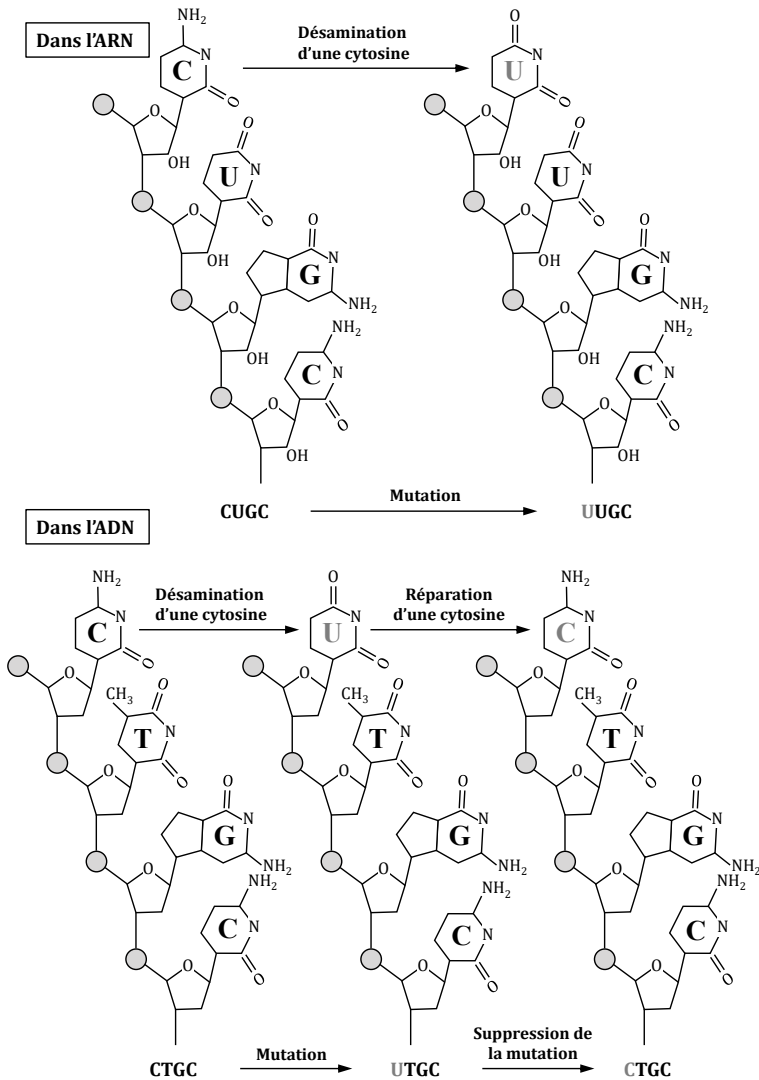


Figure 11 | Désamination des cytosines et réparation des uraciles. Haut : Dans l'ARN, si une cytosine s'hydrolyse spontanément au cours du temps et se transforme en uracile (en gris), la séquence change et une mutation apparaît dans la molécule. Bas : Dans l'ADN, si une cytosine se transforme en uracile, un mécanisme de réparation va reconnaître cette uracile (qui n'est pas une des quatre bases de l'ADN) et la réparer en cytosine pour supprimer la mutation et redonner la séquence correcte.

La troisième particularité de l'ADN est structurale. Nous avons vu que l'ARN pouvait adopter une forme en brin double en se repliant sur lui-même, les A s'appariant avec les U et les G avec les C, afin de former une structure secondaire importante pour l'activité catalytique des ribozymes (Figure 5). Dans la plupart des organismes vivants (à l'exception encore de certains virus), l'ADN est sous forme de brin double. Chaque base d'un brin d'ADN est appariée à sa base complémentaire sur le brin opposé. La structure de l'ADN a été élucidée en 1953 par l'Américain James Watson et l'Anglais Francis Crick, dans un article séminal qui leur valut en 1962 le prix Nobel de Physiologie ou Médecine, avec l'Américain Maurice Wilkins. Cette structure de l'ADN dite « en double hélice » a depuis fait le tour de la planète et est certainement à l'heure actuelle la structure la plus connue d'une macromolécule du vivant. Elle a inspiré de nombreux artistes dont Salvador Dalí, qui s'en est inspiré dans plusieurs tableaux surréalistes dont *Galacidalacidesoxyribonucleicacid* peint en 1963, qui se veut un hommage aux découvreurs de la structure de l'ADN et à sa femme, Gala. Cet arrangement très particulier des brins de l'ADN présente deux avantages, d'une part celui de réduire l'accessibilité des bases à des solvants qui pourraient affecter leur intégrité, mais surtout celui d'avoir une « sauvegarde » de l'information génétique. En effet, en plus de la désamination des cytosines abordée ci-dessus, d'autres modifications chimiques, telles que la méthylation des cytosines ou la désamination des guanines et des adénines, peuvent modifier la séquence de l'ADN, entraînant des mutations. Si cela se produit sur un des deux brins de la molécule, il est encore possible d'utiliser l'autre brin d'ADN comme matrice pour réparer la base modifiée, un peu comme un disque de sauvegarde permet de récupérer une information effacée par mégarde de son disque de travail.

Dans les tout premiers temps de la chimie prébiotique et des premières cellules, l'accumulation de mutations dans l'ARN permettait une évolution rapide des séquences afin de générer de la diversité génétique et augmenter les chances d'innovation. L'apparition des

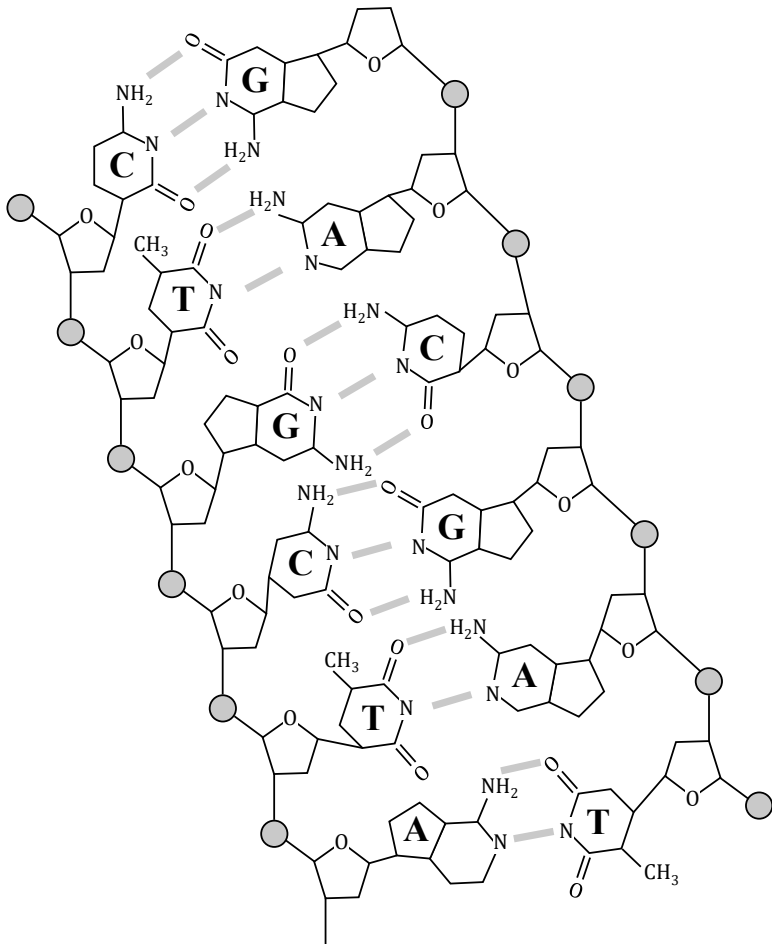


Figure 12 | La double hélice d'ADN. Les traits gris représentent les liaisons hydrogène qui relient les bases : A et T forment deux liaisons, G et C en forment trois, comme dans l'ARN.

premières cellules et le besoin de stabiliser l'information génétique pour réduire les chances de mutations délétères dans ces cellules ont permis de sélectionner une molécule plus stable que l'ARN. Cette plus grande stabilité chimique alliée à la présence d'un deuxième brin « de sauvegarde » donne à l'ADN une supériorité incontestable sur l'ARN comme support de l'information génétique. Il conservera ce rôle au cours de l'évolution des êtres vivants, jusqu'à nos jours.

DE L'ARN AUX PROTÉINES...

Pour passer de l'ARN à l'ADN, il faut inventer deux enzymes aux noms compliqués : une *ribonucléotide réductase* qui va convertir le ribose en désoxyribose et une *thymidylate synthase* qui va synthétiser la thymine spécifique de l'ADN à partir de l'uracile spécifique de l'ARN. Il est intéressant de noter que les trois classes de ribonucléotide réductases du monde vivant utilisent toutes les trois la chimie du soufre et un atome de fer (ou de cobalt), réminiscence de la chimie fer-soufre proposée pour la formation des premières molécules organiques.

Ces deux réactions sont à l'heure actuelle catalysées par des protéines et on peut douter qu'un ARN ancestral ait été doté des propriétés catalytiques complexes nécessaires pour les effectuer. On peut donc en conclure que le monde à ARN a donné naissance d'abord aux protéines, qui ont permis de créer les fonctions catalytiques complexes nécessaires à la transformation de l'ARN en l'ADN. Nous avons déjà vu rapidement que les protéines sont des macromolécules résultant de l'enchaînement d'acides aminés reliés entre eux par une liaison peptidique. Il existe vingt-deux acides aminés dans les protéines du vivant, classés en différentes catégories en fonction de leurs propriétés chimiques. Tous les acides aminés – ou presque – contiennent une partie générique commune qui porte une fonction acide (-COOH) chargée négativement et une fonction amine (-NH₂) chargée positivement dans des conditions physiologiques. Ils portent donc tous une charge négative et une charge positive, qui

se neutralisent mutuellement dans le milieu intracellulaire. La partie générique commune est liée à une chaîne latérale particulière qui confère à chaque acide aminé des propriétés chimiques légèrement différentes. Certains d'entre eux sont acides (chargés négativement), d'autres sont basiques (chargés positivement), d'autres sont hydrophobes ou hydrophiles, ce qui modifie leur interaction avec d'autres molécules, y compris d'autres protéines. D'autres encore portent un atome de soufre ou ont d'autres propriétés chimiques que l'on peut utiliser pour les classer. Notons que la sélénocystéine ne se retrouve que dans quelques dizaines seulement de protéines chez l'homme et que la pyrrolysine n'est présente que dans certaines archées (Figure 13).

Le nombre de séquences possibles pour une protéine de 100 acides aminés s'écrit 1 suivi de 130 zéros et dépasse très largement le nombre d'atomes dans l'univers. Il existe donc une infinité de séquences protéiques possibles dont seulement quelques dizaines de milliers sont utilisées dans le monde vivant. La multiplicité des séquences protéiques leur confère une diversité structurale bien plus importante que celle de l'ARN, permettant de générer de nouvelles activités catalytiques dont l'ARN était incapable. De plus, certaines protéines sont capables de changer de structure en fixant une molécule, ou sont capables de fixer plusieurs substrats dans des régions différentes de la protéine. Certaines sont strictement organisées dans l'espace en une structure secondaire qui détermine leur spécificité et leur fonction enzymatique, d'autres (la plus grande partie) sont partiellement ou totalement désordonnées, ce qui leur donne une plus grande spécificité de fonction. La protéine p53 humaine, qui joue un rôle central dans la signalisation des dommages de l'ADN, contient un grand domaine structuré et plusieurs domaines désordonnés pouvant adopter des conformations différentes en fonction des molécules qui vont s'y fixer. Cette souplesse dans la structure de la protéine permet à p53 d'interagir avec pas moins de 14 partenaires différents ! Un exploit dont serait bien incapable un ribozyme dont la structure secondaire

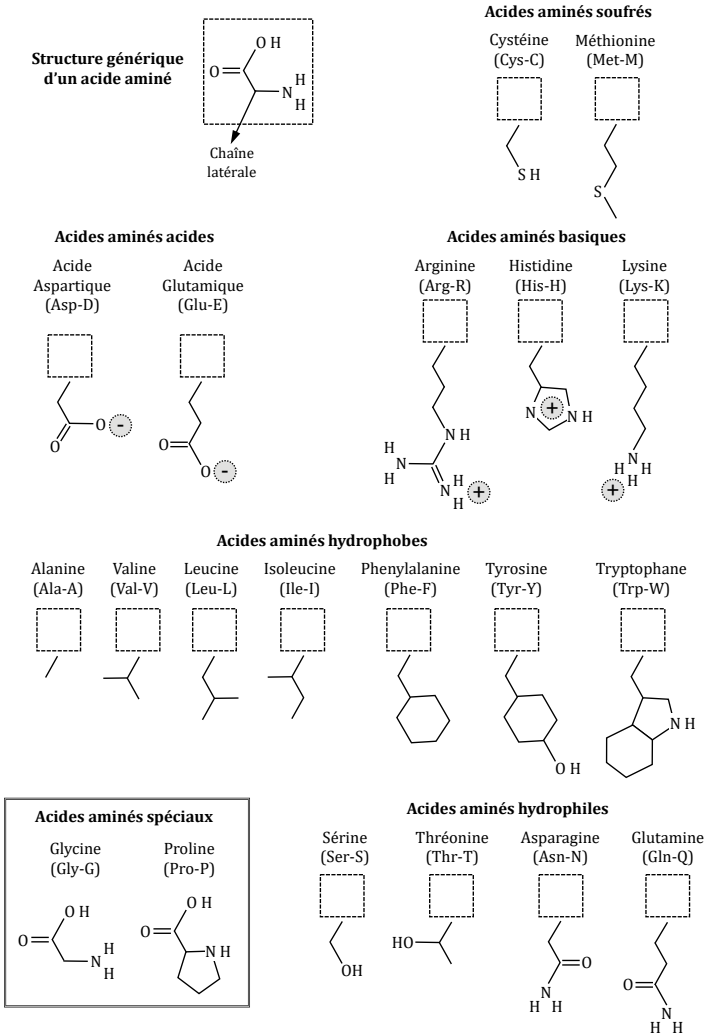


Figure 13 | Les acides aminés. La structure générique commune représentée en haut à gauche et encadrée en pointillés est reliée aux différentes chaînes latérales, classées par propriétés chimiques. La glycine et la proline ont des structures particulières. Le nom de chaque acide aminé est indiqué au-dessus de sa chaîne latérale. Les carrés en pointillés représentent la structure générique à l'endroit où elle s'attache à la chaîne latérale. Entre parenthèses, le code à trois lettres et celui à une lettre. Les acides aminés chargés sont indiqués par un (+) ou un (-).

est fixée par la séquence et limitée par les quatre ribonucléotides qui la composent. Ainsi, l'arrivée des premières protéines a donné au vivant de nouveaux outils, plus flexibles et plus diversifiés pour inventer de nouvelles fonctions, dont celles qui vont permettre l'arrivée de l'ADN.

La transition du monde ribonucléique (monde à ARN) à un monde ribonucléoprotéique (ARN et protéines) est assez simple à expliquer et nous en avons exploré les prémices au chapitre précédent en présentant le ribosome. C'est un biologiste roumain, George Emil Palade, qui découvrit à l'aide d'un microscope électronique dans les années 1950 des granules denses à l'intérieur des cellules. Pour cette découverte, il reçut en 1974 le prix Nobel de Physiologie ou Médecine, avec Albert Claude et Christian de Duve. Ces granules denses furent baptisés plus tard ribosomes et montrés comme étant responsables de toute la synthèse protéique des organismes vivants. Pour essayer de comprendre l'importance du ribosome dans le vivant, il faut savoir qu'une bactérie en cours de croissance rapide comme *Escherichia coli* qui vit dans le tube digestif humain, contient environ 70 000 ribosomes, leur masse représentant 40-50 % de la masse cellulaire totale. Un organisme vivant en train de se reproduire n'est en fait qu'une énorme machine à fabriquer des ribosomes, qui vont fabriquer les protéines nécessaires à fabriquer d'autres ribosomes, etc. Or, nous avons vu précédemment que l'activité catalytique qui permettait la formation de la liaison peptidique entre deux acides aminés était portée par l'ARN ribosomique, les protéines n'étant présentes que pour maintenir en place l'ARN dans la conformation structurale active. Cette propriété de l'ARN ribosomique suggère qu'un ribosome ancestral, préexistant aux premières protéines, était sans doute déjà capable de catalyser la formation de la liaison peptidique. Il apparaît ainsi très vraisemblable que ce ribosome ancestral, formé uniquement d'ARN, soit à l'origine des premières protéines. Une fois celles-ci assemblées, même de courte séquence, la diversité des structures obtenues a permis de créer rapidement de nouvelles fonctions,

à la fois de support structural pour l'ARN et de catalyse de nouvelles réactions chimiques, pour finalement fabriquer l'ADN.

... DES PROTÉINES À L'ADN-U

Le passage du monde ribonucléoprotéique (ARN et protéines) à un monde désoxyribonucléoprotéique (ADN, ARN et protéines) est plus complexe à expliquer. D'autant plus complexe que l'ARN est sous forme de brin simple alors que la plupart des molécules d'ADN sont sous forme de brins doubles complémentaires, ce qui demande en plus des deux activités enzymatiques décrites plus haut (ribonucléotide réductase et thymidylate synthase), une machinerie complexe pour reproduire la double hélice d'ADN à l'identique, d'une génération à l'autre. Cette reproduction d'une molécule d'ADN en une deuxième molécule identique à la première porte le nom de **réplication***. Elle repose sur l'ARN d'abord, qui assure l'initiation d'un des deux brins de l'ADN, puis sur les protéines qui assurent l'essentiel de la synthèse des deux brins. Un pan entier de la biologie moléculaire étudie la réplication de l'ADN, mécanisme très complexe et hautement régulé, sur l'efficacité duquel reposent la survie et la reproduction de tous les organismes vivants. Avant d'explorer les questions liées à la réplication de l'ADN, il est bon de revenir sur son invention.

Réplication : Mécanisme complexe et hautement régulé permettant de dupliquer une molécule d'ADN pour donner deux molécules identiques à la première. C'est de cette façon que se transmet l'information génétique d'une cellule mère à une cellule fille.

Deux enzymes sont donc essentielles pour passer du monde à ARN au monde à ADN. La ribonucléotide réductase est capable de transformer trois nucléotides, la guanosine, la cytidine et l'adénosine en leurs équivalents désoxyribonucléotidiques. Le quatrième nucléotide de l'ADN, la thymidine, est quant à lui fabriqué à partir de la

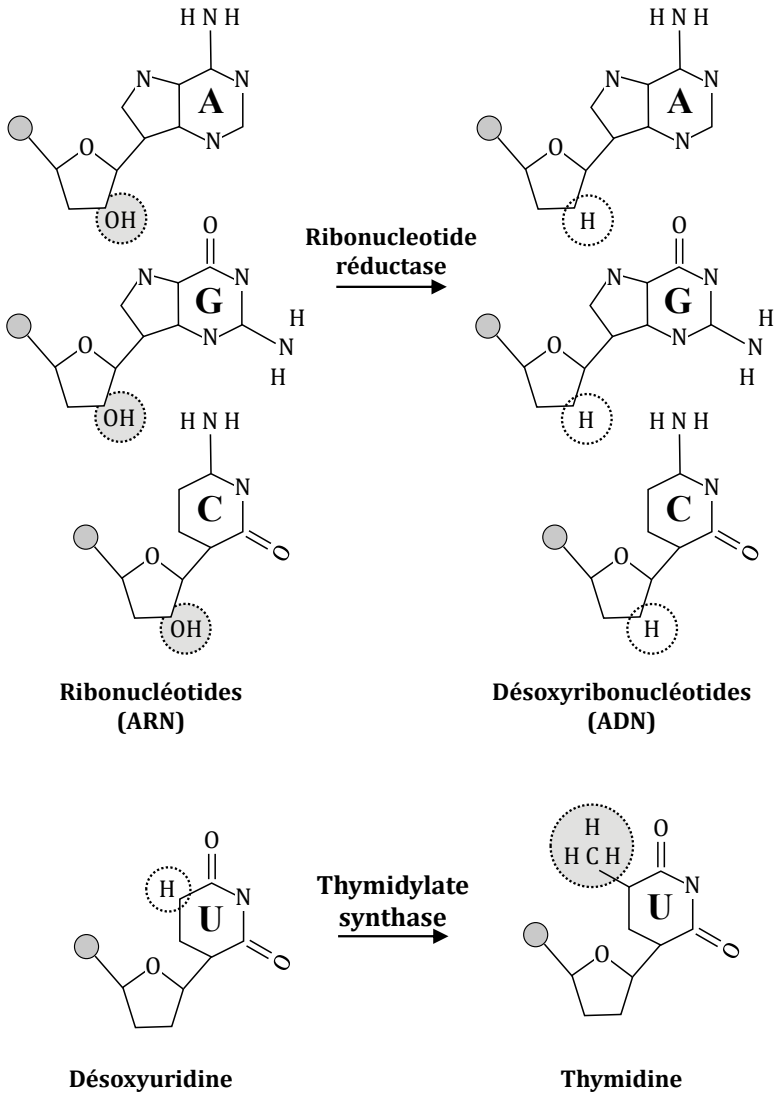


Figure 14 | De l'ARN à l'ADN. Les trois ribonucléotides, adénosine, cytidine et guanosine, peuvent être transformées en leurs équivalents déoxyribonucléotides par la ribonucléotide réductase. La thymidine est fabriquée à partir de la désoxyuridine, par la thymidylate synthase.

désoxyuracile et nécessite la thymidylate synthase (Figure 14). Pour cette raison, certains auteurs ont suggéré qu'un monde à ADN-U (pour uracile) aurait précédé le monde à ADN-T (pour thymine), tel que nous le connaissons aujourd'hui. La découverte que certains virus bactériens contiennent de l'ADN-U au lieu de l'ADN-T apporte un certain soutien à cette hypothèse.

... ET DE L'ADN-U À L'ADN-T

Il existe deux formes de thymidylate synthase, cette protéine capable de transformer l'uracile en thymine. Les deux thymidylate synthases se nomment ThyA et ThyX et leur structure ainsi que leur mécanisme d'action n'ont rien en commun. Environ 30 % du monde bactérien utilise ThyX alors que le reste du monde vivant utilise l'autre protéine, ThyA. Il existe même certains organismes qui utilisent les deux synthases. On pourrait donc supposer que la thymine – donc l'ADN – a été inventée deux fois au cours de l'évolution, dans deux lignées vivantes distinctes. On pourrait également se demander laquelle a précédé l'autre et surtout pourquoi le monde vivant a-t-il eu besoin de deux synthases pour passer à l'ADN-T ?

ThyX est une enzyme environ dix fois moins efficace que ThyA pour transformer l'uracile en thymine. De plus, le temps nécessaire à une bactérie pour répliquer entièrement son génome est beaucoup plus élevé pour les bactéries utilisant ThyX que pour celles utilisant ThyA. Les organismes possédant ThyX ont des génomes plus petits et contiennent en moyenne 40 % d'ADN en moins à répliquer. Ces observations soulèvent l'intéressante possibilité que les organismes ancestraux au contenu en ADN plus réduit, aient d'abord utilisé ThyX pour fabriquer la thymine. Puis, lorsque les organismes se sont complexifiés et que la quantité d'ADN à répliquer est devenue plus importante, la vitesse de réplication est devenue un facteur limitant pour la propagation rapide de ces microorganismes. Si une thymidylate synthase plus rapide est créée à ce moment-là, elle va être rapidement sélectionnée par les bactéries contenant plus d'information

génétique à répliquer. À ce moment-là, ThyX peut être perdue par ces bactéries, n'étant plus une protéine essentielle à la survie de l'organisme. Cette hypothèse est compatible avec l'observation que 70 % des bactéries thermophiles et hyperthermophiles (qui sont proposées comme étant les survivants des plus anciens organismes vivants) utilisent ThyX, bien que l'échantillonnage étudié soit un peu faible pour pouvoir conclure de façon définitive. ThyX serait donc la thymidylate synthase ancestrale et aurait été remplacée au cours de l'évolution par ThyA, une enzyme plus rapide, qui aurait été sélectionnée par les organismes devant répliquer de plus grandes quantités d'ADN.

Comme pour les premières molécules organiques, des solutions indépendantes au même problème ont été trouvées dans différentes lignées biologiques. C'est d'ailleurs une observation courante dans le monde vivant. Plusieurs solutions au même problème existent, mais lorsqu'une solution est trouvée par un organisme, elle est adoptée par l'ensemble des descendants de cet organisme, sans chercher une meilleure solution, qui serait plus rapide et moins coûteuse en termes d'énergie. Le vivant n'est pas économe, son unique souci est sa reproduction.

LE PARADOXE D'EIGEN

Maintenant que l'ADN-T – l'ADN « moderne » – existe grâce à l'invention de la thymine, il faut le répliquer. On pourrait se demander pourquoi ce rôle n'a pas été entièrement dévolu à l'ARN, déjà doté de l'activité catalytique nécessaire à sa propre synthèse. La vitesse de synthèse d'un ARN à partir d'un ribozyme est lente (quelques nucléotides par heure dans des conditions expérimentales optimales) et la fidélité de synthèse des ribozymes est faible (environ 1-4 % d'erreurs suivant les molécules utilisées). Avec un taux d'erreur de 1 %, il est statistiquement impossible de reproduire à l'identique et en grand nombre une molécule d'ARN de plus de 100 bases. Or le contenu génétique des organismes modernes est bien plus important, comme

nous le verrons plus tard. C'est le paradoxe présenté par Manfred Eigen dans un article fameux datant de 1971 : pour reproduire à l'identique une information de plus de 100 bases, il faut une molécule complexe dont la synthèse nécessite plus de 100 bases. À l'époque à laquelle Eigen décrit ce paradoxe, les capacités catalytiques de l'ARN ne sont pas encore connues. Il lui était donc impossible de savoir qu'une molécule d'ARN de courte taille était capable d'en synthétiser une autre. Néanmoins, le paradoxe reste entier : comment passer d'un monde dans lequel le taux d'erreur pour reproduire un ARN est de l'ordre de 1 %, au monde moderne, à des organismes dont les génomes contiennent des millions, des milliards de bases, sans perdre toute l'information nécessaire à cette transition ?

Il y a plusieurs façons de résoudre ce paradoxe. La première est de postuler que plusieurs molécules d'ARN ont coexisté et coopéré à l'intérieur des premières cellules vivantes. Appelons « G » un ARN capable de reproduire avec un taux d'erreur de 1 % toute molécule d'ARN, y compris lui-même. Appelons « E » un autre ARN incapable de reproduire une autre molécule mais capable de détecter 90 % des erreurs commises par G. Tant que G et E sont associés dans la même cellule, les deux molécules d'ARN sont capables de se propager et de se maintenir. Lors d'une division de cette cellule, les deux molécules vont ségréger de façon stochastique dans les cellules filles. Si G et E se trouvent séparées lors d'une division, elles seront incapables de survivre l'une sans l'autre et disparaîtront. Tant que G et E se trouvent dans une même cellule, elles pourront se reproduire et la taille de G pourra même augmenter sensiblement. On pourra objecter que cette hypothèse ne fait que reculer le paradoxe d'un ordre de grandeur. En effet, lorsque G aura atteint une taille de 1 000 bases, E ne sera plus capable de corriger toutes les erreurs commises par G et l'on atteindra à nouveau la limite de Eigen. À la différence près que l'on sera passé d'un monde de ribozymes mesurant une centaine de bases et capables uniquement d'activités catalytiques lentes et peu spécifiques, à un monde où l'augmentation de la taille de l'information génétique

permet de fabriquer des protéines, dont l'activité catalytique sera beaucoup plus rapide et fidèle.

Une enzyme de nature protéique est environ 10 000 fois plus rapide et au moins 100 fois plus fidèle qu'un ribozyme. Les enzymes qui répliquent l'ADN se nomment « ADN polymérases ». Celles de la levure de boulangerie *Saccharomyces cerevisiae* synthétisent environ 2 000 nucléotides par minute avec un taux d'erreur de 1 base sur 10 millions si l'activité de correction d'erreurs est fonctionnelle, ou de 1 base sur 10 000 si son activité de correction a été inactivée. Les performances des ADN polymérases d'organismes moins ancestraux que la levure sont aussi rapides et encore plus fidèles. Ces nombres sont à comparer avec le taux d'erreur de 1-4 % environ des ribozymes dotés d'une activité ADN polymérase. De façon générale, les ARNs sont des catalyseurs bien moins efficaces que les protéines. Un ribozyme mis en compétition avec une polymérase de nature protéique perdra la bataille de la vitesse et de la fidélité. On comprend bien que, dès lors qu'une polymérase protéique est apparue, même sous sa forme la plus simple c'est-à-dire sans activité correctrice d'erreurs, elle a dû rapidement être sélectionnée pour répliquer les molécules d'ARN (ou d'ADN). On a pu ainsi passer d'un monde où l'information génétique était de taille limitée – une centaine de bases contenues dans un ou plusieurs ARNs – à un monde dans lequel la taille de l'information génétique va pouvoir exploser pour atteindre plusieurs millions, voire milliards de bases, et dont toute l'information est portée par l'ADN.

Il faut quand même noter ici qu'il est possible de synthétiser et de sélectionner, *in vitro*, une molécule d'ADN capable d'activité catalytique. Toutefois, il n'existe pas naturellement dans la nature d'ADN catalytique, de *désoxyribozyme*, du moins nul n'en a décrit à ce jour. La raison en est très vraisemblablement que la flexibilité et l'efficacité des protéines pour assurer les fonctions catalytiques nécessaires à la vie ont fait obstacle à la sélection de molécules d'ADN pour des fonctions similaires, le cantonnant au rôle de support de l'information génétique qu'il occupe encore aujourd'hui.

LE CHAÎNON MANQUANT DE L'ARN À L'ADN ?

Les polymérases protéiques ont donc rapidement supplanté les ribozymes et ont permis l'émergence du monde à ADN. Mais justement, combien existe-t-il d'ADN polymérases ? Le monde vivant en contient sept types différents¹ qui appartiennent à deux grandes familles structurales : la famille Klenow² et la famille Pol β . Jusqu'en 2016, toutes les ADN polymérases appartenaient à une de ces deux familles structurales, toutes sauf celles du type PolD, dont on ne connaissait pas encore la structure. Celle-ci a été résolue par une équipe de l'Institut Pasteur qui a montré de façon remarquable qu'elle n'appartenait pas à une des deux grandes familles structurales déjà décrites, mais qu'elle ressemblait plus à une ARN polymérase, c'est-à-dire à une enzyme capable de synthétiser de l'ARN à partir d'ADN. Les polymérases de type PolD sont donc vraisemblablement le chaînon manquant qui a permis de passer du monde à ARN au monde à ADN. Les ancêtres de PolD devaient être capables de fabriquer de l'ARN et de l'ADN, puis la capacité à synthétiser l'ARN est devenue l'exclusivité d'autres polymérases et les PolD modernes ne répliquent plus que l'ADN.

Finalement, on peut se demander si les transformations qui nous ont emmenés d'un monde à ARN à un monde à ARN-protéines, puis à un monde à ARN-protéines-ADN, se sont déroulées successivement, telles que décrites ici dans un but de simplification, ou bien si certaines de ces innovations se sont produites en parallèle dans plusieurs lignées protocellulaires pour au final converger dans les cellules primitives qui ont donné naissance à tous les organismes vivants. Certains auteurs suggèrent même que l'ADN et les polymérases capables de le répliquer auraient pu apparaître d'abord dans des virus qui les auraient transmis au reste du monde vivant en infectant

1. On regroupe les ADN polymérases en sept types en fonction de leur séquence : PolA, PolB, PolC, PolD, PolX, PolY et les transcriptases réverses.

2. Du nom de Hans Klenow (1923-2009), biochimiste danois qui a été le premier à isoler et à étudier de façon précise la polymérase éponyme.

les premières cellules à ARN. Mais peut-on réellement dater l'apparition des toutes premières cellules capables de se diviser de façon autonome ?

À RETENIR

- L'ADN est une molécule chimiquement plus stable que l'ARN et sa structure en brin double en fait un meilleur support que l'ARN pour conserver et propager l'information génétique.
- Les enzymes qui transforment l'ARN en ADN sont des protéines et il est donc très vraisemblable que les protéines ont existé avant l'ADN.

BIBLIOGRAPHIE

- La structure de l'ADN : Watson et Crick, *Nature* 171: 737-738 (1953).
- Le remplacement des ribozymes par des enzymes protéiques : Jeffares *et al.*, *Journal of Molecular Evolution* 46: 18-36 (1998).
- Les virus, le monde à ARN et le monde à ADN : Forterre, *Biochimie* 87: 793-803 (2005).

5

L'énergie des premières cellules et la Grande Oxydation

On peut dater l'apparition des premières cellules vivantes grâce à des communautés bactériennes très particulières, les **stromatolithes***. Ces structures en feuillets ont été formées par des bactéries qui se sont déposées dans un environnement aquatique de faible profondeur. Ces multiples couches bactériennes sont parfois interrompues par des dépôts de **sédiments*** pour former des stromatolithes dits en « double couche », dont la forme discoïde en coussins ou en mamelons les distingue très clairement de dépôts sédimentaires normaux qui créent des structures laminaires plus classiques. La présence de stromatolithes est révélatrice de la présence de communautés microbiennes formant des **biofilms***, sans doute l'organisation la plus ancienne du vivant. Avec le temps, certains de ces stromatolithes peuvent se fossiliser et on peut retrouver leurs traces au milieu de sédiments vieux de plusieurs milliards d'années. Jusqu'en 2016, les plus vieux stromatolithes fossiles avaient été identifiés en Australie comme datant de 3,5 milliards d'années. Mais la même année, des stromatolithes

encore plus vieux étaient trouvés au sud-ouest du Groenland. Ces structures fossiles coniques, de 1-4 cm de hauteur, découvertes entre des dépôts de roches sédimentaires vieux de 3,7 milliards d'années, repoussent donc l'apparition des premières cellules vivantes de 200 millions d'années. C'est tout à fait remarquable, car cette nouvelle datation signifie que la vie sous forme cellulaire a mis « seulement » 600-700 millions d'années à émerger après la formation de la Terre.

Biofilm : Communauté de microbes (bactéries, champignons et algues unicellulaires) vivant parfois en symbiose, adhérant entre eux et à une surface au sein d'un milieu aqueux et produisant parfois une matrice biologique au rôle protecteur. C'est l'un des deux modes de vie des organismes unicellulaires, l'autre étant le mode « planctonique », défini par la flottaison libre au sein d'un milieu aqueux.

L'ÉNERGIE DES PREMIÈRES CELLULES

Comme nous l'avons vu, les premières cellules étaient des bactéries (archées et eubactéries), les cellules eucaryotes n'arrivant que plus tardivement dans l'histoire du vivant. À cette date, le taux d'oxygène atmosphérique était extrêmement bas, de l'ordre de 0,1 % de son niveau actuel. Toute vie **aérobie*** était donc impossible et les premières bactéries étaient donc forcément **anaérobies***. Mais alors comment fabriquaient-elles leur énergie ? Il est certain que les premières cellules vivantes étaient capables de synthétiser toutes les molécules dont elles avaient besoin pour vivre et pour se reproduire à partir de matière inorganique, par synthèse chimique.

Aérobie : Un organisme aérobie est un organisme capable de respirer, c'est-à-dire d'utiliser l'oxygène comme source d'énergie. On distingue les organismes à aérobie stricte (qui ne peuvent vivre sans oxygène) de ceux à aérobie facultative (qui peuvent utiliser d'autres sources d'énergie que l'oxygène, comme la fermentation). S'oppose à anaérobie.

Il existe une grande variété de bactéries modernes capables de chimiosynthèse anaérobie. On les trouve à des profondeurs marines que n'atteignent jamais les rayons du soleil, d'environ 250 mètres à 4 000 mètres, sur des fonds de sable, de grès, de calcaire, de basalte ou de granit. Certaines de ces espèces bactériennes sont thermophiles, elles ne peuvent vivre qu'à haute température (entre 50 °C et 70 °C), d'autres sont hyperthermophiles, ne vivant qu'à des températures supérieures à 80 °C. *Pyrolobus fumarii*, une archéobactérie hyperthermophile a été découverte aux abords d'une cheminée hydrothermale de la dorsale atlantique¹ et vit à 113 °C, une température à laquelle l'eau de mer serait sous forme de vapeur à la surface des océans². Par comparaison, aucun eucaryote n'est capable de survivre à une température supérieure à 60 °C. Il est possible que ces archéobactéries hyperthermophiles soient des survivants d'une époque lointaine où la vie est apparue dans des sources hydrothermales très chaudes au fond des océans. Il est clair que le monde bactérien est capable d'utiliser de nombreuses réactions chimiques d'**oxydo-réduction*** pour fabriquer de l'énergie.

Oxydo-réduction : Un échange d'électrons entre un réducteur et un oxydant est appelé réaction d'oxydo-réduction. Un réducteur est une molécule qui donne facilement ses électrons, comme le sulfure d'hydrogène (H₂S). Un oxydant est une molécule qui accepte facilement des électrons, comme l'oxygène atmosphérique (O₂). Ces réactions sont essentielles pour le métabolisme énergétique de tous les êtres vivants.

La réduction du soufre par l'hydrogène pour former du sulfure d'hydrogène (H₂S) est un des procédés énergétiques le plus couramment utilisés chez les bactéries. Quiconque s'est déjà promené dans les îles éoliennes au nord de la Sicile, particulièrement sur Vulcano,

1. Appelée aussi « fumeur noir ».

2. L'eau non salée bout à 100 °C à la surface (pression atmosphérique = 1 bar). À 4 000 mètres de profondeur, il faudrait chauffer l'eau à environ 400 °C pour la faire bouillir.

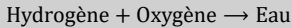
est familier avec l'odeur nauséabonde du sulfure d'hydrogène. Celle-ci est due à la présence de *Pyrodictium occultum*, une archée hyperthermophile vivant dans une source chaude marine de l'île. *P. occultum* a été le premier organisme vivant capable d'être cultivé à des températures supérieures à 100 °C et dont le métabolisme énergétique était basé sur la réduction du soufre en sulfure d'hydrogène. De nombreuses autres espèces de thermophiles et d'hyperthermophiles sont également capables de catalyser cette réaction d'oxydo-réduction. Le sulfure d'hydrogène n'est qu'un des nombreux composés soufrés capables d'être utilisés comme source d'énergie par des bactéries ou des archées, il en existe des dizaines d'autres, dans différents états d'oxydation. Le monoxyde de carbone (CO) peut être réduit puis transformé en méthane (CH₄) par différentes bactéries méthanogènes, à des températures supérieures à 100 °C. Différentes espèces d'*Archaeoglobus* fabriquent de l'énergie en oxydant de l'acide formique (CH₂O₂), de l'acide acétique (C₂H₄O₂) ou de l'acide lactique (C₃H₆O₃). *Archaeoglobus veneficus*, une archée anaérobie qui vit dans des fumeurs noirs de la dorsale atlantique à 3 500 mètres de profondeur, peut également métaboliser le glucose ou l'éthanol pour produire de l'énergie. Certaines bactéries anaérobies peuvent réduire le nitrate (NO₃⁻), mais également d'autres molécules inorganiques contenant de l'azote dans différents états d'oxydation. On peut également citer l'oxydation du soufre et du fer sur la pyrite par *Thiobacillus ferrooxidans*, qui porte bien son nom. Au total, il existe plus de 300 réactions pouvant se dérouler à des températures de plus de 100 °C et à des pressions plus ou moins élevées, utilisées par le monde bactérien pour produire de l'énergie, nombre d'entre elles se produisant en condition d'anaérobie.

Ainsi, les nombreux exemples d'utilisation de composés chimiques simples par des bactéries modernes vivant à des températures et à des pressions compatibles avec l'origine des premières cellules vivantes, suggère que la vie a effectivement pu commencer sur ces bases biochimiques, avant d'évoluer vers des systèmes plus complexes, utilisant

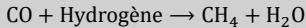
une énergie disponible en abondance et en permanence (ou presque) : l'énergie solaire.

Quelques réactions d'oxydo-réduction couramment utilisées par diverses cellules pour fabriquer de l'énergie :

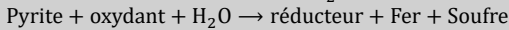
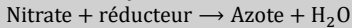
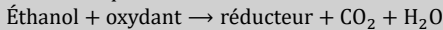
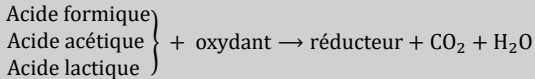
OXYDATION DE L'HYDROGÈNE



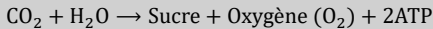
RÉDUCTION DU MONOXYDE DE CARBONE



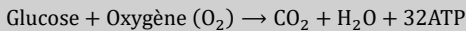
AUTRES RÉACTIONS D'OXYDO-RÉDUCTION



PHOTOSYNTHÈSE



RESPIRATION



L'ARRIVÉE DES CYANOBACTÉRIES

Les cyanobactéries vont profondément et durablement changer l'atmosphère terrestre, permettant l'apparition de nouvelles formes de vie dont le métabolisme reposera dorénavant sur la consommation d'oxygène. Elles sont parfois appelées algues bleues, mais c'est un abus de langage car les algues sont des organismes eucaryotes alors que les cyanobactéries sont des procaryotes. Néanmoins, ce surnom leur vient d'une coloration bleu-vert (ou vert en fonction des bactéries), due à la présence des pigments responsables de la photosynthèse. Celle-ci permet de transformer l'énergie solaire en énergie chimique utilisable par la cellule. Il existe une grande variété de cyanobactéries, ayant colonisé toutes les niches écologiques humides, océans, lacs d'eau douce, sources chaudes et tous types de sols jusqu'aux rochers

de l'Antarctique. Elles peuvent pousser sous forme planctonique ou bien en formant des biofilms, en fonction des espèces. Certaines d'entre elles sont capables de fixer l'azote atmosphérique (N_2) et de le transformer en nitrites (NO_2) ou en nitrates (NO_3), utilisés par les plantes pour leur croissance. On trouve donc souvent des cyanobactéries vivant en **symbiose*** dans des rizières où elles servent de fertilisant naturel. C'est sans doute un des groupes de microorganismes qui a eu le plus de succès au cours de l'évolution, en termes de diversité d'habitats et de niches écologiques.

Symbiose : Association constante, obligatoire et spécifique entre deux organismes vivants ne pouvant vivre l'un sans l'autre.

Les stromatolithes modernes étant formés par des cyanobactéries, on est en droit de penser que les stromatolithes vieux de 3,7 milliards d'années que nous avons évoqués au début du chapitre étaient donc constitués de cyanobactéries ancestrales.

L'INVENTION DE LA PHOTOSYNTHÈSE

Les enzymes de la photosynthèse sont contenues à l'intérieur des cyanobactéries, dans des sortes de sacs aplatis appelés **thylakoïdes***. La chlorophylle, mais également des pigments de la famille des phycobiliprotéines dont la composition peut varier d'une espèce à l'autre de cyanobactérie, vont transformer l'énergie lumineuse en énergie chimique. En présence de lumière, les enzymes de la photosynthèse vont transformer l'eau (H_2O) et le dioxyde de carbone (CO_2) en un sucre (comme le glucose, $C_6H_{12}O_6$), et en oxygène atmosphérique (O_2). L'invention de la photosynthèse par les cyanobactéries a donc lentement transformé l'atmosphère terrestre, en l'enrichissant graduellement en oxygène, transformation qui va s'avérer létale pour la plupart des organismes vivants. En effet, l'oxygène est un oxydant puissant et un poison violent pour les cellules qui ne savent pas l'utiliser. L'augmentation de la concentration en oxygène due à la photosynthèse a vraisemblablement tué la plupart des espèces vivantes à

l'époque, dans une première grande extinction de masse qui ne sera que la première d'une longue série. Les cellules vivant dans la croûte terrestre ou dans les sources chaudes soufrées au fond des océans ont survécu à cette augmentation d'oxygène atmosphérique. Leur descendance donnera les bactéries et les archées anaérobies modernes, qui ne peuvent vivre qu'en absence d'oxygène. Mais d'autres cellules se sont adaptées à ces nouvelles conditions de vie, car l'augmentation d'oxygène a été lente au début. C'est ainsi qu'est apparue la respiration, qui permet d'utiliser l'oxygène atmosphérique afin de produire de l'énergie.

L'INVENTION DE LA RESPIRATION

La respiration est la réaction inverse de la photosynthèse, puisqu'elle utilise le glucose ($C_6H_{12}O_6$) et l'oxygène atmosphérique (O_2) pour fabriquer de l'eau (H_2O), du dioxyde de carbone (CO_2) et de l'énergie sous forme d'adénosine triphosphate ou ATP. Au sein des cyanobactéries, les enzymes de la photosynthèse sont physiquement proches de celles de la chaîne respiratoire, ce qui facilite les transferts chimiques et la fabrication d'ATP. Comme dit précédemment, l'ATP est le carburant utilisé par l'ensemble du monde vivant aérobic. Il est constitué d'une molécule d'adénosine, un des quatre nucléotides de l'ARN, auquel sont liés trois phosphates. Chaque liaison entre deux phosphates est hautement énergétique et son hydrolyse libère l'énergie nécessaire à tous les processus cellulaires. L'hydrolyse de l'ATP le transforme en ADP (adénosine diphosphate) qui sera régénérée par la suite en ATP par la respiration. L'ATP est la source d'énergie essentielle des organismes aérobies et on considère que chaque jour le corps humain consomme et régénère l'équivalent de son propre poids en ATP. Le séquençage d'une quarantaine de cyanobactéries des trois grandes classes connues a clairement montré que seules les *Oxyphotobacteria* possédaient les enzymes de la photosynthèse. Les premières cyanobactéries étaient donc incapables d'utiliser la lumière pour fabriquer de l'oxygène, cette capacité n'apparaissant qu'après la

divergence des *Oxyphotobacteria* et des deux autres classes (Figure 15). En revanche, les enzymes de la respiration sont retrouvées chez toutes les cyanobactéries et ont été acquises indépendamment car toutes ne possèdent pas exactement les mêmes complexes respiratoires. Ceci confirme bien que l'ancêtre des cyanobactéries était incapable d'utiliser l'oxygène et que les trois classes acquièrent cette capacité de façon indépendante, après leur divergence.

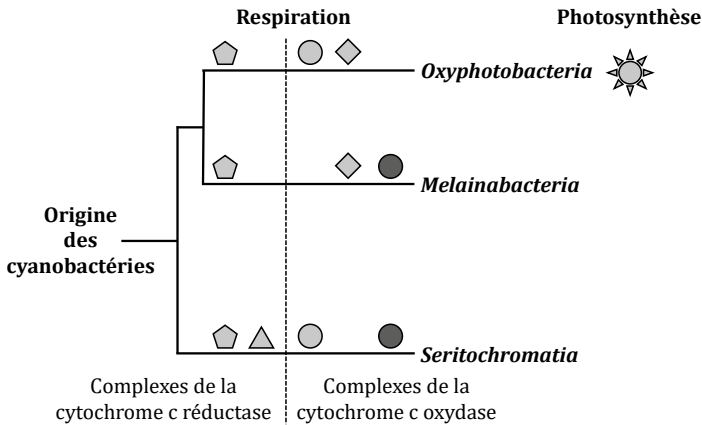


Figure 15 | Les complexes protéiques de la photosynthèse et de la respiration chez les cyanobactéries. L'arbre représente les trois classes de cyanobactéries : oxyphotobactéries, melainabactéries et seritochromaties. Les figures géométriques représentent les différents complexes de la respiration présents dans chaque classe. La photosynthèse n'a été trouvée que chez les oxyphotobactéries, alors que les enzymes de la respiration sont présentes dans les trois branches de l'arbre. Noter que les trois classes de cyanobactéries n'utilisent pas exactement les mêmes enzymes, indiquant que la respiration a été inventée indépendamment dans les trois branches.

LA GRANDE OXYDATION

Cette transformation radicale de l'atmosphère terrestre, appelée « Grande Oxydation » ou « Grande Oxygénation », s'est produite il y a environ 2,4 milliards d'années et a permis d'atteindre environ le niveau actuel d'oxygène terrestre. Cela eut pour effet direct de faire disparaître une grande partie des premiers organismes vivants et d'en sélectionner d'autres, capables non seulement de résister à l'oxygène

atmosphérique, mais de l'utiliser afin de fabriquer de l'énergie pour leur propre croissance. À partir de là, ces organismes se sont mis à croître rapidement dans tous les océans, diminuant grâce à la photosynthèse le niveau de dioxyde de carbone. L'oxygène réagissant avec le méthane atmosphérique (CH₄) va diminuer sa concentration, or le méthane est un très puissant gaz à effet de serre, bien plus que le CO₂. La croissance actuelle rapide des niveaux de CO₂ et de méthane due aux activités industrielles et agricoles humaines augmente l'effet de serre, faisant grimper la moyenne des températures depuis l'ère préindustrielle. À l'inverse, la diminution de ces deux gaz il y a 2 milliards d'années a entraîné un refroidissement massif du globe menant à une des plus importantes périodes de glaciation de son histoire, de -2,4 à -2,1 milliards d'années : la glaciation huronienne. Cette baisse des températures va permettre l'apparition de formes de vie **mésophiles***, telles que les cellules eucaryotes, qui ne peuvent vivre à trop haute température, à l'inverse des bactéries thermophiles. Pendant cette longue période, la concentration en oxygène atmosphérique va augmenter jusqu'à atteindre une valeur proche de sa valeur actuelle, soit environ 21 % des gaz contenus dans l'atmosphère terrestre. Puis, celle-ci va diminuer pour des raisons inconnues avant de remonter jusqu'aux valeurs actuelles, il y a environ 900 millions d'années. Cet oxygène va maintenant pouvoir être activé par le rayonnement ultraviolet en provenance du soleil pour former de l'ozone (O₃). Cette couche d'ozone en haute altitude va filtrer les rayons ultraviolets les plus nocifs, assurant la protection des organismes vivants à la surface du globe.

L'OZONE, PROTECTEUR DU VIVANT

On se souvient de la découverte, dans les années 1980, de la diminution de la couche d'ozone au-dessus de l'Antarctique, qui avait affolé à juste titre l'humanité tout entière. Cette destruction de l'ozone atmosphérique était provoquée par l'utilisation massive de chlorofluorocarbures, des gaz contenant du chlore, utilisés principalement

dans les bombes aérosols et dans l'industrie du froid car facilement compressibles et peu réactifs chimiquement. Le chlore détruisant l'ozone, ces gaz ont depuis été interdits dans l'industrie afin de nous protéger. La diminution ou la perte de la couche d'ozone aurait des conséquences dramatiques, non seulement pour l'humanité, mais pour la vie sur Terre. En effet, les rayonnements ultraviolets sont nocifs pour les cellules vivantes, provoquant des dommages dans l'ADN qui peuvent être irréversibles s'ils sont trop importants. La vie terrestre telle que nous la connaissons serait impossible si la couche d'ozone diminuait de façon significative sur l'ensemble du globe. Or, il y a environ deux milliards d'années, c'est l'inverse qui s'est produit. L'augmentation de l'oxygène atmosphérique a entraîné la formation d'ozone qui, formant une couche protectrice aux rayons ultraviolets, a permis à la vie telle que nous la connaissons actuellement de pouvoir naître, se développer à la surface de la Terre et évoluer jusqu'à l'époque actuelle.

Ainsi, on comprend mieux pourquoi l'invention de la photosynthèse par les cyanobactéries a été un des changements les plus révolutionnaires dans l'histoire de la vie sur Terre : (i) l'augmentation de l'oxygène atmosphérique a tué la plupart des cellules vivantes et sélectionné celles qui étaient capables d'utiliser cet oxygène nouvellement créé pour fabriquer de l'énergie ; (ii) la diminution du méthane et du dioxyde de carbone a entraîné un refroidissement rapide et une longue glaciation permettant l'apparition de formes de vie mésophiles ; (iii) finalement, la formation de la couche d'ozone a protégé les cellules vivantes des rayonnements ultraviolets mortels. Tout est maintenant en place pour permettre l'explosion de nouvelles formes de vie extrêmement diversifiées, des cellules eucaryotes primitives jusqu'aux plantes et aux animaux.

À RETENIR

- Les premières bactéries ont émergé seulement 600-700 millions d'années après la formation de la Terre.
- Celles-ci utilisaient des réactions chimiques d'oxydo-réduction pour fabriquer leur énergie.
- Les cyanobactéries ont été les premiers organismes vivants capables de photosynthèse, transformant l'énergie solaire en oxygène.
- La fabrication d'oxygène par les cyanobactéries a transformé l'atmosphère terrestre lors de la Grande Oxydation, il y a environ 2,4 milliards d'années.

BIBLIOGRAPHIE

L'origine de la respiration et de la photosynthèse chez les cyanobactéries : Soo, Hemp, Parks, Fischer and Hugenholtz, *Science* 355 : 1436-1440 (2017).

La Grande Oxydation : Lyons, Reinhard and Planavsky, *Nature* 506: 307-315 (2014).

6

Gènes et code génétique

Nous avons vu précédemment que les cellules eucaryotes possèdent un certain nombre de spécificités qui les distinguent nettement des eubactéries et des archées. Tout d'abord, leur noyau qui contient l'information génétique codée par l'ADN, mais le codage de cette information est lui-même très spécifique et nécessite maintenant quelques explications, à la fois pour comprendre l'origine du **code génétique*** et celle des cellules eucaryotes.

QU'EST-CE QU'UN GÈNE ?

L'ADN est constitué de quatre bases accrochées à un sucre, le désoxyribose, pour former quatre désoxyribonucléotides différents : l'adénosine, la cytidine, la guanosine et la thymidine. Nous savons également que les protéines sont les effecteurs indispensables de toute cellule vivante, grâce à la flexibilité de leurs structures et leur efficacité à catalyser des réactions enzymatiques. Chaque protéine va jouer un ou plusieurs rôles, remplir des fonctions précises au sein de la cellule : fonction structurale, fonction catalytique (enzyme),

transmission de signal biologique, etc. La séquence – et donc la structure – de chaque protéine est codée par l'ADN sous forme de gènes. Mais qu'est-ce qu'un **gène*** exactement ? On peut assez simplement le définir comme un segment d'ADN transcrit en ARN et possédant une fonction biologique. Notons qu'une partie des gènes effectuent leur fonction par l'intermédiaire d'un ARN, et une autre partie par l'intermédiaire d'une protéine, comme nous allons maintenant le décrire.

Gène : Segment d'ADN transcrit en ARN ayant une fonction dans la cellule.

Deux processus métaboliques hautement régulés vont se succéder pour transformer le message porté par la molécule d'ADN en ARN puis en protéines. Tout d'abord, la **transcription*** va transformer la séquence ADN en ARN. Ces ARN transcrits sont des copies conformes du gène dont ils sont issus, à la différence que la thymine y est remplacée par l'uracile et que le désoxyribose y est remplacé par le ribose (voir Chapitre 4). Ces ARNs ont plusieurs destins cellulaires possibles. Certains de ces ARNs jouent un rôle catalytique (comme l'ARN ribosomique), d'autres régulent l'expression de certains gènes, d'autres encore vont subir un deuxième processus métabolique, la **traduction***, qui va transformer la séquence portée par ces ARN – que l'on nomme messagers – en protéines. C'est la lecture de ces ARN messagers (ou ARNm) par le ribosome qui permet leur traduction en protéine, grâce au code génétique. L'information génétique d'une cellule vivante est donc contenue dans l'ensemble de ses gènes (le génome), mais l'ADN n'en est que le gardien, le dépositaire.

Le génome d'une bactérie comme *Escherichia coli*, une entérobactérie **commensale***

Commensal : Un organisme commensal est l'hôte d'un autre organisme avec lequel il cohabite sans lui causer de dommages. L'hôte nourrit partiellement l'organisme, mais à la différence d'un parasite, la présence d'un commensal n'est pas néfaste à l'hôte.

des mammifères, contient environ 4 200 gènes qui codent les protéines assurant la survie et la reproduction de cette bactérie. Le génome de la levure de boulangerie, un eucaryote unicellulaire, contient environ 6 500 gènes codant des protéines. L'hypothèse « un gène donne une protéine », qui était dans l'air du temps au début du xx^e siècle, fut confirmée expérimentalement par Georges Beadle et Edward Tatum dans un article historique publié en 1941. En étudiant un champignon filamenteux, *Neurospora crassa*, ils découvrirent que la déficience de certains mutants pour la synthèse de la vitamine B6 était liée à un seul gène. Ils en conclurent que ce gène devait coder une enzyme responsable d'une étape de la synthèse de cette vitamine. Ces résultats furent rapidement confirmés dans d'autres organismes et valurent à leurs auteurs de recevoir le prix Nobel de Physiologie ou Médecine en 1958. Le principe d'un gène codant une protéine est vrai dans le monde procaryote, mais trouve ses limites chez les eucaryotes, comme nous le verrons plus loin.

LE CODE GÉNÉTIQUE

Mais quelle est la nature du code génétique qui permet la traduction d'un ARN messager en protéine ? Les expériences qui ont mené à la découverte du code génétique se sont déroulées au début des années 1960 et sont parmi les plus astucieuses de la biologie moderne. Un peu de mathématiques pour commencer. Les protéines peuvent contenir 22 acides aminés différents. Parmi ceux-ci, la sélénocystéine et la pyrrolysine sont très rares et sont incorporées par des mécanismes spécifiques qui ne seront pas détaillés ici. Il reste donc 20 acides aminés à coder par les quatre bases de l'ADN. On comprend rapidement qu'une seule base ne peut pas coder un acide aminé, il manquerait 16 bases pour coder l'ensemble. Si deux bases codaient un acide aminé, il existerait 4^2 soit 16 variations possibles de deux bases, ce qui serait encore insuffisant pour coder 20 acides aminés. À partir de trois bases pour un acide aminé, on atteint un nombre de possibilités suffisantes ($4^3 = 64$) pour coder tous les acides aminés. De fait, certaines

expériences préliminaires suggéraient que le nombre de bases nécessaires était sans doute proche de six. Francis Crick, Sydney Brenner et leurs collaborateurs vont utiliser un élégant système pour démontrer que le code génétique lit les bases de l'ADN trois par trois. La bactérie *Escherichia coli* peut être infectée et tuée par un virus, le **phage*** T4. En utilisant des agents chimiques,

on peut obtenir des mutants de ce virus qui n'infectent plus la bactérie. Francis Crick avait réussi à obtenir un grand nombre de mutants avec l'acridine, un agent chimique

Phage (ou bactériophage) : Virus qui infecte les bactéries, en injectant son matériel génétique (ADN ou ARN) afin de le faire reproduire par la cellule bactérienne hôte.

qui se lie à l'ADN et qui provoque une addition ou une soustraction d'un nucléotide. Ces mutants du phage T4 n'étaient plus capables d'infecter les bactéries. En soumettant ces mutants (appelés *plus*) à un deuxième traitement par l'acridine, il était possible d'isoler d'autres mutants (appelés *moins*) qui supprimaient la première mutation, permettant au phage d'infecter à nouveau *Escherichia coli*. En combinant les différentes mutations obtenues, les règles suivantes purent être établies : (i) une mutation *plus* était supprimée par une mutation *moins* ; (ii) une mutation *moins* était supprimée par une mutation *plus* ; (iii) deux, quatre ou cinq mutations *plus* ne se supprimaient pas ; (iv) trois ou six mutations *plus* redonnaient des phages capables d'infecter les bactéries. L'analyse fine de ces mutants leur permettra de proposer que le code génétique lise trois bases à la fois et non six, ce qui s'avérera par la suite exact. Ces trois bases lues simultanément par le ribosome prirent le nom de « codon » (Figure 16). Au même moment, un autre groupe publiait la synthèse d'une protéine contenant uniquement l'acide aminé phénylalanine à partir d'un ARN ne contenant que des uraciles. Cette expérience démontrait donc que le codon UUU codait la phénylalanine et ouvrait la voie au décryptage du code génétique dans son entier.

Tableau 1 | Quasi-universalité du code génétique.

Première base du codon	Deuxième base du codon				Troisième base du codon
	U	C	A	G	
U	Phénylalanine (Phe)	Sérine (Ser)	Tyrosine (Tyr)	Cystéine (Cys)	U
	Phénylalanine (Phe)	Sérine (Ser)	Tyrosine (Tyr)	Cystéine (Cys)	C
	Leucine (Leu)	Sérine (Ser)	STOP	STOP (1)	A
	Leucine (Leu) (7)	Sérine (Ser)	STOP	Tryptophane (Trp)	G
C	Leucine (Leu) (4)	Proline (Pro)	Histidine (His)	Arginine (Arg)	U
	Leucine (Leu) (4)	Proline (Pro)	Histidine (His)	Arginine (Arg)	C
	Leucine (Leu) (4)	Proline (Pro)	Glutamine (Gln)	Arginine (Arg)	A
	Leucine (Leu) (4) (6)	Proline (Pro)	Glutamine (Gln)	Arginine (Arg)	G
	A	Isoleucine (Ile) (3)	Théonine (Thr)	Asparagine (Asn)	Sérine (Ser)
Isoleucine (Ile) (3)		Théonine (Thr)	Asparagine (Asn)	Sérine (Ser)	C
Isoleucine (Ile) (3)		Théonine (Thr)	Lysine (Lys)	Arginine (Arg) (2) (5)	A
Méthionine (Met)		Théonine (Thr)	Lysine (Lys)	Arginine (Arg) (2) (5)	G
G	Valine (Val)	Alanine (Ala)	Aspartate (Asp)	Glycine (Gly)	U
	Valine (Val)	Alanine (Ala)	Aspartate (Asp)	Glycine (Gly)	C
	Valine (Val)	Alanine (Ala)	Glutaminate (Glu)	Glycine (Gly)	A
	Valine (Val) (7)	Alanine (Ala)	Glutaminate (Glu)	Glycine (Gly)	G

Il existe 61 codons correspondant à un acide aminé et trois codons « non-sens » ou « stop », indiquant la fin de la protéine. Le code génétique est dégénéré, c'est-à-dire que plusieurs codons différents codent le même acide aminé, par exemple la leucine est codée par CUU, CUC, CUA et CUG. Cet exemple illustre aussi qu'un des intérêts d'un code à trois lettres est de limiter le rôle délétère des mutations. En effet, si une mutation se produit sur la troisième base du codon, elle ne produira pas de changement d'acide aminé dans la protéine car les quatre codons possibles commençant par CU codent tous les quatre le même acide aminé. Tous les acides aminés sont codés par au moins deux codons différents, à l'exception de la méthionine et du tryptophane, cette dégénérescence partielle donnant sa robustesse au code génétique.

LA PRESQUE UNIVERSALITÉ DU CODE GÉNÉTIQUE

Le code génétique est universel, il est le même dans tous les organismes vivants, ou presque. Il présente certaines variations en

Le tableau se lit de la façon suivante : pour le codon GCA, la première base est G, la deuxième C et la troisième A.

À l'intersection de ces trois bases, on trouve l'acide aminé Alanine (abrégié Ala). Notez que les trois autres codons commençant par GC codent également le même acide aminé, d'où la dégénérescence du code génétique.

Variations les plus communes du code génétique :

- (1) Dans toutes les mitochondries, le codon UGA code un tryptophane.
- (2) Dans les mitochondries de vertébrés, les codons AGA et AGG sont des codons STOP.
- (3) Dans les mitochondries de vertébrés, les codons AUU, AUC et AUA sont des codons méthionine et initient la traduction de l'ARN messager. Dans les mitochondries de levure, seul le codon AUA code une méthionine.
- (4) Dans les mitochondries de levure, les codons CUU, CUC, CUA et CUG codent une thréonine.
- (5) Dans les mitochondries d'invertébrés, les codons AGA et AGG codent la sérine.
- (6) Chez la levure *Candida albicans* et les levures qui en sont proches, le codon CUG code une sérine.
- (7) Chez les bactéries et les archées, les codons GUG et UUG sont souvent des codons méthionine.

particulier dans les mitochondries où le codon UGA, qui est normalement un des trois codons « stop », est traduit en tryptophane. De plus, spécifiquement dans les mitochondries des vertébrés, certains codons arginine et isoleucine sont traduits différemment, alors que dans les mitochondries de levure, ce sont certains codons isoleucine et leucine qui sont traduits différemment. Chez certaines levures comme *Candida albicans*, le codon CUG code une sérine plutôt qu'une leucine comme dans le reste du monde vivant. Il existe aussi des variations dans la nature du codon qui initie le début de la traduction protéique, mais pris dans son ensemble, le code génétique est remarquablement homogène et universel (Tableau 1). Cela suggère fortement qu'il n'a été inventé qu'une fois au cours de l'évolution et donc que son invention remonte aux protocellules eubactériennes et archéobactériennes. Le code génétique a sans doute été utilisé sous une forme ancestrale par l'ARN ribosomique qui était déjà capable de catalyser la formation de la liaison peptidique bien avant l'existence du ribosome, comme nous l'avons vu précédemment. Une fois le code fixé sur trois bases, il s'est propagé tel quel dans tout le monde vivant, aux quelques rares exceptions mineures mentionnées ci-dessus, qui correspondent très vraisemblablement à des modifications ultérieures du code. Il a donc sans doute précédé l'apparition de l'ADN, à une époque lointaine où l'ARN servait à la fois de support de l'information génétique et de catalyseur de réactions chimiques. Puis, l'arrivée de l'ADN a cantonné l'ARN au rôle de messenger de cette information et dans certains cas très particuliers de catalyse, rôles qu'il occupe encore aujourd'hui.

L'INTRIGANTE QUESTION DES ARNt SYNTHÉTASES

Il faut ici mentionner que le passage de l'ARN messenger à la protéine, le décodage du message, ne se fait pas directement, mais par l'intermédiaire d'un petit ARN appelé ARN de transfert (ARNt). Chaque acide aminé est chargé sur un ARNt spécifique grâce à une enzyme appelée **ARNt synthétase***. Chaque ARNt chargé de son

acide aminé va ensuite reconnaître les trois nucléotides de son codon grâce aux trois nucléotides complémentaires de l'ARNt qui forment l'anticodon. Cette reconnaissance codon-anticodon se fait au cœur du ribosome et l'acide aminé chargé par l'ARNt est incorporé à la chaîne protéique en cours d'élongation.

Il existe deux grandes familles d'ARNt synthétases, celles dites de classe I et celles de classe II. La structure des synthétases de classe I est très différente de celle des synthétases de classe II. Les deux familles ne proviennent pas d'un ancêtre commun, montrant à nouveau que l'évolution a trouvé deux moyens différents de parvenir au même résultat. Comme pour la thymidylate synthase qui permet de passer du monde à ADN-U au monde à ADN-T, on pourrait penser que l'ARNt synthétase a été inventée deux fois, dans deux branches différentes du vivant et que les descendants modernes de ces cellules ancestrales contiennent l'une ou l'autre synthétase. Il n'en est rien car chaque ARNt synthétase est spécifique d'un et d'un seul acide aminé. Chacune des deux classes contient 10 ARNt synthétases différentes capables chacune de charger un et un seul acide aminé sur son ARNt correspondant. Il existe donc 10 synthétases de classe I et 10 autres de classe II, qui ensemble chargent les 20 acides aminés sur leurs ARNt respectifs¹. Chaque cellule vivante contient donc un mélange de synthétases de classe I et de classe II. De plus, l'analyse détaillée des séquences des synthétases découvertes dans les trois grands domaines du vivant ne permet pas de définir une règle précise quant à leur acquisition au cours de l'évolution. Par exemple, l'ARNt synthétase chargeant la leucine chez les eucaryotes est plus proche de celle que l'on trouve chez les archées que de celle retrouvée chez les eubactéries, de même pour la synthétase chargeant la phénylalanine. En revanche, la synthétase chargeant la valine chez les eucaryotes est plus proche des synthétases eubactériennes que de celles du monde des archées.

1. Il existe en fait 21 synthétases, car la lysine peut être chargée par une synthétase de classe I ou de classe II, en fonction des espèces. Comme souvent en biologie, chaque règle a ses exceptions.

Enfin, pour d'autres synthétases, comme celle chargeant l'acide aspartique, il est tout simplement impossible de retracer l'histoire évolutive en comparant les séquences des protéines de chaque organisme. Alors, quel scénario évolutif pour les ARNt synthétases, ces protéines essentielles du monde vivant ?

On a vu précédemment que l'ARN était capable de catalyser la formation de la liaison peptidique, bien avant l'arrivée du ribosome. La traduction de l'ARN en protéine est donc vraisemblablement un mécanisme très ancien, remontant au monde à ARN. Il est donc plausible que les premières synthétases aient été des molécules d'ARN capables d'activer et de charger les acides aminés sur leur ARNt respectif, des sortes de « ribozymes synthétases ». L'analyse des séquences protéiques suggère l'existence de trois synthétases protéiques ancestrales (leucine, isoleucine et valine) qui auraient remplacé au cours du temps les ribozymes synthétases. Puis, le succès de ces enzymes – plus rapides et plus efficaces que leur équivalent ribozyme – aurait permis par duplication successive leur propagation dans les premières protocellules. L'existence de deux familles structurellement bien distinctes suggère que les synthétases protéiques ont bien été inventées deux fois au cours de l'évolution, mais que les protocellules se sont échangé ces enzymes par un mécanisme de **transfert horizontal*** qui sera détaillé plus loin. Dans le cas des ARNt synthétases, le transfert semble s'être fait des protocellules archées vers les protocellules eubactéries et dans ce sens uniquement. Cette observation un peu étonnante suggère qu'à cette époque lointaine les eubactéries primitives aient pu être plus perméables à des transferts de gènes, peut-être en raison de la composition différente de leurs membranes lipidiques (voir chapitre 3), ou d'un niveau de structuration cellulaire plus primitif permettant une plus grande perméabilité membranaire que celle des protocellules archéobactériennes. Ces transferts de synthétases protéiques se sont forcément produits après l'apparition de la traduction protéique. Les enzymes ancestrales de classe I (leucine, isoleucine et valine) et de classe II (glycine et lysine) furent inventées avant la divergence des

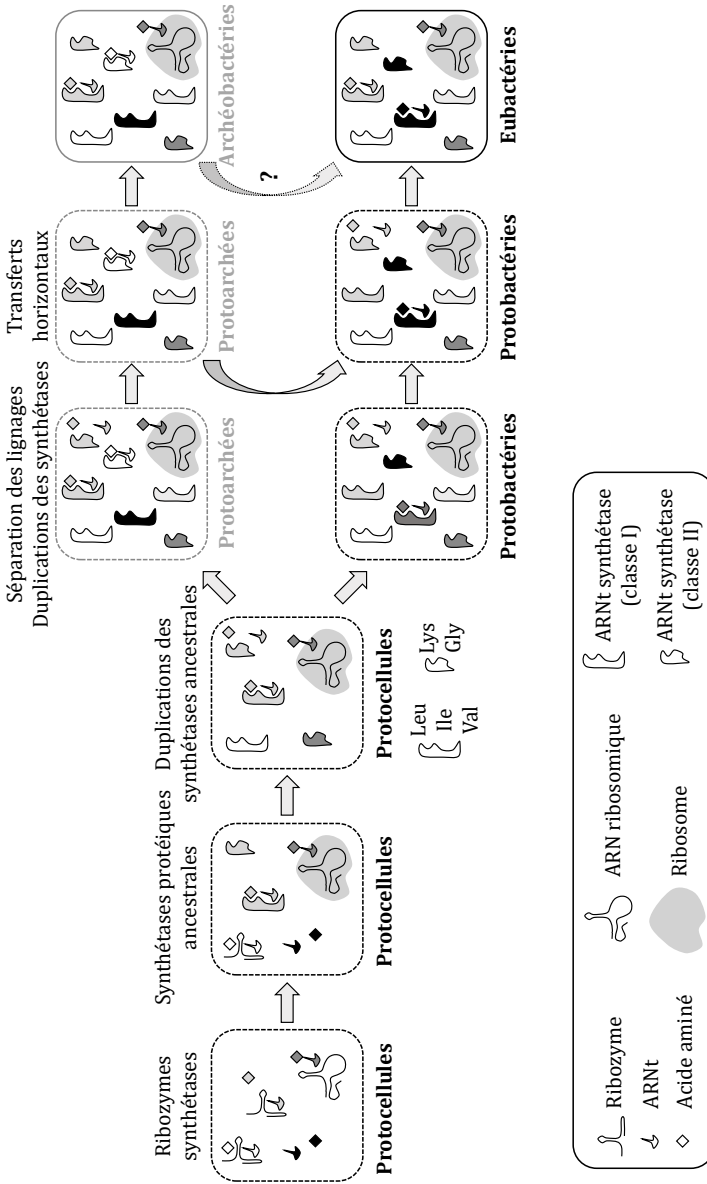


Figure 17 | La possible évolution des ARNT synthétases.

eubactéries et des archées, les synthétases suivantes passant également d'un organisme à un autre par transfert horizontal au cours de millions d'années d'évolution et au gré des rencontres – physiques – entre microorganismes primitifs, au sein de biofilms ancestraux jusqu'à ce que la machinerie de traduction et le code génétique soient définitivement fixés (Figure 17).

À RETENIR

- La transcription est le mécanisme qui transforme une séquence d'ADN en ARN.
- La traduction est le mécanisme qui transforme le message porté par un ARN en protéine, par l'intermédiaire du code génétique.
- À de rares exceptions, le code génétique est universel. Il a donc été fixé dans sa forme actuelle avant la divergence entre les procaryotes et les eucaryotes.

BIBLIOGRAPHIE

Un gène code une protéine : Beadle et Tatum, *Proceedings of the National Academy of Sciences USA* 27: 499-506 (1941).

L'évolution des ARNt synthétases : Woese, Olsen, Ibba et Söll, *Microbiology and Molecular Biology Reviews* 64 : 202-236 (2000).

7

Les premiers eucaryotes

LA THÉORIE ENDOSYMBIOTIQUE

Une découverte récente a permis de clarifier l'origine de l'ancêtre commun de toutes les cellules eucaryotes. Une équipe de chercheurs suédois a mis en évidence une nouvelle lignée d'archéobactéries dont le génome contient de nombreux gènes eucaryotes. Ces archées ont été nommées Loki, Thor, Odin et Heimdall, en référence aux dieux de la mythologie nordique et appartiennent maintenant à l'embranchement Asgard, du nom de la mythique demeure des dieux (Figure 7). Elles sont trouvées tout autour de la planète dans des sources géothermales de la dorsale atlantique, du parc de Yellowstone ou de l'île japonaise de Taketomi. La séquence de leur génome a montré la présence de nombreux gènes que l'on ne trouve que chez les eucaryotes, codant entre autres des protéines impliquées dans le trafic intracellulaire, une spécificité eucaryote impliquant l'invagination de la membrane cellulaire et permettant la **phagocytose***. La comparaison de ces gènes

Phagocytose : Propriété de certaines cellules de capturer et d'ingérer des particules ou des micro-organismes afin de les détruire.

avec les gènes eucaryotes correspondants montre clairement une relation parentale ancestrale. C'est donc l'ancêtre de l'une des bactéries de cet embranchement qui a donné naissance à la lignée eucaryote.

L'origine des organites intracellulaires, mitochondries et chloroplastes, est restée confidentielle jusqu'à ce que Lynn Margulis (1938-2011), une zoologiste américaine, publie en 1967 une étude qui allait faire date dans l'histoire de la biologie. Ce travail, refusé par plusieurs journaux scientifiques très connus, fut finalement publié dans le *Journal of Theoretical Biology* et présentait les cellules eucaryotes comme des **symbiotes***

de plusieurs bactéries. Cette idée n'était pas nouvelle en soi, elle avait été proposée par un biologiste russe, Konstantin Mereschowski (1855-1921) en 1905 dans un article traduit en anglais en 1999. Dans cet article absolument remarquable pour l'époque, il discrédite l'idée d'alors que les chloroplastes

Symbiotes (ou symbiontes) :

Organismes symbiotiques vivant soit en association mutualiste (les deux organismes retirent un intérêt à l'association), soit en association parasitaire (l'un des deux organismes en retire un avantage tandis que l'autre paye un coût).

aient pu être des organes particuliers des cellules végétales et propose que ce soient des symbiotes de cyanobactéries¹. Quelques années plus tard, un autre biologiste russe, Boris Kozo-Polyansky (1891-1957) généralisait l'idée que tous les organismes eucaryotes descendaient de symbiotes. Lynn Margulis était parfaitement au courant de ces travaux et a même passé une partie de son existence à promouvoir le travail de Kozo-Polyansky et à faire traduire son livre du russe à l'anglais. Néanmoins, c'est le nom de Lynn Margulis qui reste étroitement associé à la théorie endosymbiotique, malgré ses positions

1. "According to this theory, the plant cell is nothing other than an animal cell with cyanophytes that have invaded it; as a consequence the plant kingdom is to be derived from the animal kingdom". Mereschowski, *European Journal of Phycology* 34: 287-295 (1999).

violemment antidarwinistes qui lui attirèrent de nombreuses inimitiés parmi ses pairs.

UN SCÉNARIO POSSIBLE POUR L'APPARITION DES PREMIERS EUCARYOTES

Bien avant la Grande Oxydation, à l'époque où les premiers procaryotes étaient exclusivement anaérobies, eut lieu une lente, mais sensible, augmentation de l'oxygène atmosphérique due à la lyse de la vapeur d'eau présente dans l'atmosphère en hydrogène et en oxygène, sous l'action du rayonnement solaire. Cette augmentation de l'oxygène devint un danger pour les cellules vivant proches ou au contact de l'atmosphère, car il altère et dégrade l'ARN, l'ADN et les protéines par oxydation. Les porphyrines sont des molécules organiques de structure cyclique associées à un atome de fer ou de magnésium, qui peuvent capter et transporter l'oxygène atmosphérique. Les cellules capables de fabriquer les premières porphyrines furent alors protégées de cette oxydation. Les deux porphyrines les plus connues sont la chlorophylle des organismes photosynthétiques et l'hème qui entre dans la composition de l'hémoglobine et des cytochromes, des enzymes de la chaîne respiratoire mitochondriale. Les cellules contenant de la porphyrine vont donc devenir plus résistantes à la lente augmentation de l'oxygène atmosphérique car elles seront capables de le capter afin de se protéger de l'oxydation. Les procaryotes qui dépendaient de réactions chimiques anaérobies pour la fabrication de leur ATP vont pouvoir maintenant utiliser les porphyrines dans des chaînes respiratoires de cytochromes, permettant la fabrication d'ATP par consommation d'oxygène. L'un de ces procaryotes sera phagocyté par une archée de la famille Asgard, lui apportant les protéines de la chaîne respiratoire qui lui permettront de métaboliser à son tour l'oxygène atmosphérique. On sait par l'analyse des génomes bactériens modernes, que c'est un ancêtre de *Rickettsia prowazekii*, une alphaprotéobactérie parasite intracellulaire obligatoire, qui fut ingéré par une archée Asgard, donnant ainsi

naissance aux mitochondries et aux premières cellules eucaryotes primitives, les *protoeucaryotes*.

D'autres bactéries à porphyrines vont pouvoir absorber la lumière et l'utiliser pour développer un système photosynthétique. La photosynthèse permettant de fabriquer un peu d'ATP, ces cyanobactéries ancestrales vont rapidement obtenir un avantage de croissance par rapport aux autres. Elles vont alors utiliser la photosynthèse pour fabriquer de l'énergie tout en rejetant de l'oxygène dans l'atmosphère. C'est cet oxygène qui mènera à la Grande Oxydation il y a 2,4 à 2,1 milliards d'années et qui tuera toutes les cellules incapables de le métaboliser. Finalement, un des *protoeucaryotes* phagocyttera une cyanobactérie ancestrale pour donner naissance à une cellule eucaryote photosynthétique capable de respiration, une *protoalgue*, ancêtre des cellules végétales modernes (Figure 18).

Il y a des arguments scientifiques solides en faveur de la théorie endosymbiotique. Les chloroplastes et les mitochondries contiennent de l'ADN. Cet ADN code une quarantaine de gènes dans les mitochondries animales, une soixantaine de gènes dans les mitochondries végétales et une centaine de gènes dans les chloroplastes. Les gènes présents dans ces deux types d'organelles ne sont pas suffisants à assurer leur reproduction ni leur fonction. Ils nécessitent les gènes de leurs cellules hôtes, ce sont donc bien des symbiotes obligatoires. Le séquençage de ces génomes a montré que l'ARN ribosomique chloroplastique était proche de celui des cyanobactéries modernes. Les chloroplastes possèdent également des thylakoïdes qui contiennent les pigments responsables de la photosynthèse, dont l'organisation est similaire à ceux des cyanobactéries. Mais l'argument le plus fort vient de la séquence récente des génomes bactériens, qui montre de façon incontestable que le génome mitochondrial des eucaryotes modernes est apparenté au génome de *Rickettsia prowazekii*, une alphaprotéobactérie, et que le génome nucléaire vient d'une bactérie Asgard. Ces arguments de similarité de séquences sont des arguments puissants en faveur de la théorie endosymbiotique.

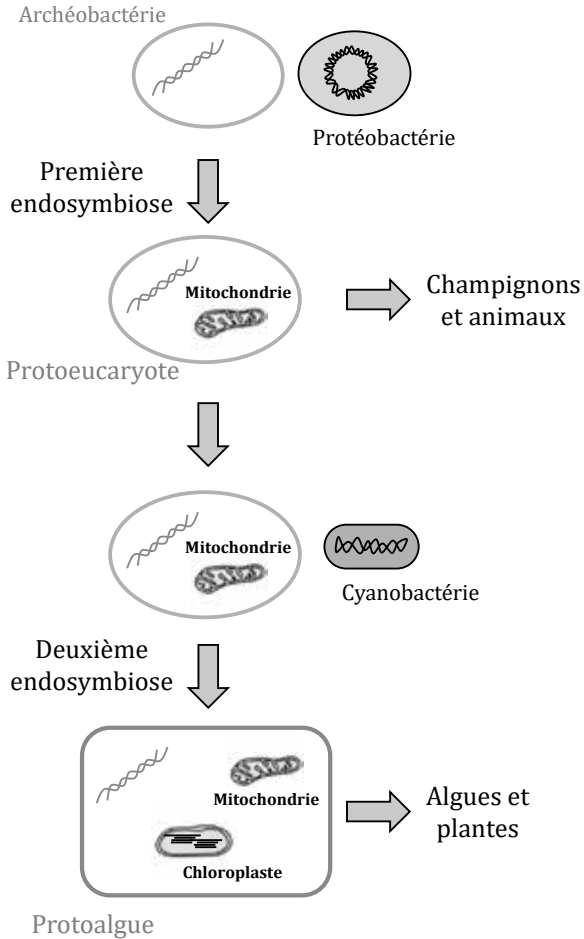


Figure 18 | La formation des premières cellules eucaryotes.

L'ORIGINE DU NOYAU DES EUCARYOTES

L'origine de la membrane nucléaire eucaryote, séparant le **cytoplasme*** du noyau contenant les chromosomes, n'est pas encore tranchée. Sa composition lipidique est de même nature que celle des membranes cellulaires eucaryotes et eubactériennes et contient donc du G3P (voir Chapitre 3). Deux hypothèses différentes peuvent être avancées pour expliquer l'origine du noyau. La première postule que le noyau ait été formé en premier

Cytoplasme : Dans les cellules eucaryotes, c'est le milieu intracellulaire qui contient le noyau, les mitochondries et les chloroplastes. On y trouve les enzymes du métabolisme ainsi que les ribosomes responsables de la traduction des ARN messagers en protéines.

et que l'énergie soit arrivée par la suite. Cette hypothèse repose sur la théorie endosymbiotique et propose que le noyau résulte de l'ingestion d'une archée par une eubactérie. L'archée serait devenue le noyau cellulaire et les gènes de l'eubactérie auraient en partie été transférés dans ce noyau et en partie perdus. Par la suite, une deuxième endosymbiose d'une alphaprotéobactérie aurait donné la première cellule eucaryote dotée de mitochondries. Dans cette hypothèse, la composition de la membrane nucléaire aurait été modifiée pour contenir du G3P au lieu du G1P archéobactérien (Figure 19). Dans la deuxième hypothèse, une archée ancestrale aurait d'abord phagocyté une alphaprotéobactérie, lui amenant une nouvelle source d'énergie sous forme de mitochondrie, permettant l'augmentation du volume cellulaire et le développement de structures complexes et de systèmes membranaires sophistiqués qui mèneront à la formation du noyau. C'est l'hypothèse « énergie en premier » (Figure 19). Cette dernière hypothèse nécessiterait néanmoins la transformation des membranes nucléaires et cellulaires contenant du G1P en membranes à G3P, peut-être par transferts horizontaux de gènes venant d'eubactéries.

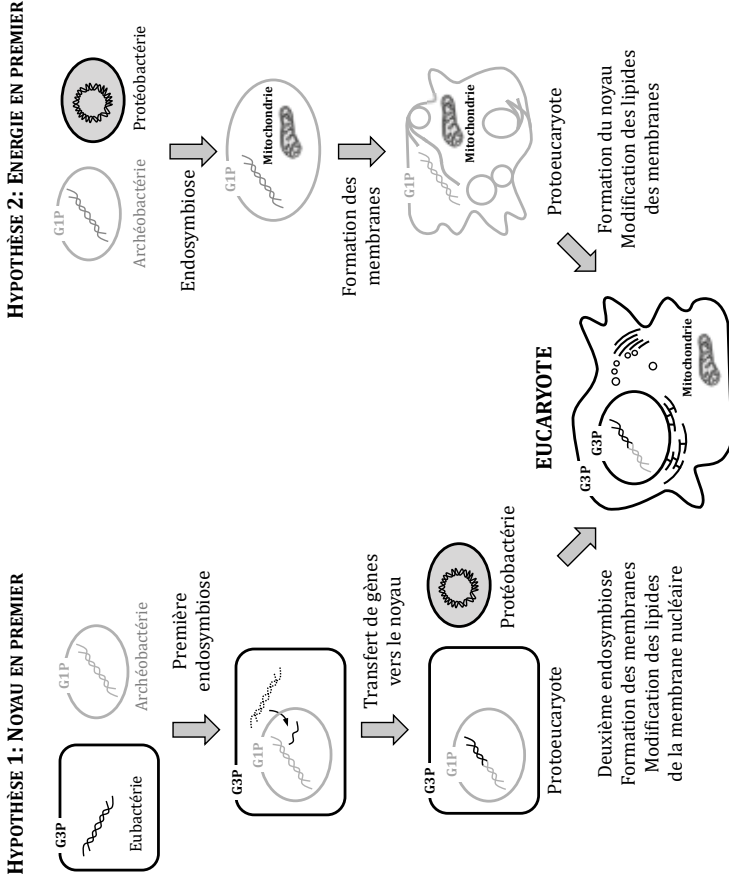


Figure 19 | L'origine du noyau des eucaryotes. Les membranes en noir contiennent du G3P, les membranes en gris contiennent du G1P.

Il faut noter que certains eucaryotes ont un noyau, mais ne possèdent pas de mitochondries. On les appelle archézoaires (Figure 7) et ils contiennent d'autres organelles leur permettant de fabriquer de l'énergie (hydrogénosome et mitosome). Leur existence suggère que l'acquisition du noyau ait été précoce dans cette branche des eucaryotes et se soit produite avant l'acquisition des mitochondries. Néanmoins, les organelles trouvées chez ces eucaryotes semblent au moins en partie dériver de mitochondries. Ces observations ne permettent donc pas de trancher définitivement entre les deux hypothèses proposées pour l'origine du noyau.

Notons que ces événements de symbioses multiples sont présentés ici dans un ordre chronologique. Il est possible qu'ils se soient passés plus ou moins simultanément, à l'échelle des temps considérés. Des *protoeucaryotes* de différents types ont pu exister en même temps et se retrouver en compétition dans des niches écologiques très proches. La sélection naturelle a éliminé les organismes les moins aptes à survivre et les survivants ont expérimenté d'autres symbioses, qui elles-mêmes ont été soumises à sélection, ainsi de suite jusqu'à ce que les premiers eucaryotes ayant à la fois un génome suffisamment stable et une production d'énergie suffisamment efficace pour prendre le dessus sur leurs compétiteurs puissent émerger.

ENDOSYMBIOSES SECONDAIRES

Les diatomées présentent un cas intéressant d'endosymbiose secondaire. Elles font partie des chromoalvéolés, des eucaryotes ancestraux vivant dans les océans. Ce sont des microalgues unicellulaires photosynthétiques, vivant à faible profondeur. Elles sont responsables de la fabrication d'un quart de l'oxygène terrestre et de 40 % des 45-50 milliards de tonnes de composés organiques produits annuellement dans les océans. Leur importance dans le recyclage global du carbone terrestre est comparable à l'ensemble de toutes les forêts tropicales. Les diatomées possèdent un frustule, une sorte de paroi rigide formée de silice hydratée, qui leur sert à filtrer les particules qui

les traversent et apportent à l'algue une rigidité protectrice contre les prédateurs. La fabrication du frustule des diatomées contrôle le recyclage de toute la silice contenue dans les océans, car chaque atome de silicium y circulant est incorporé en moyenne 40 fois dans le frustule des diatomées avant de finalement sédimenter sur les fonds océaniques. Le génome d'une diatomée, *Thalassiosira pseudonana*, a été séquencé et assemblé en 2004. Il contient 24 chromosomes et environ 11 000 gènes codant des protéines. La comparaison de ces gènes avec ceux des autres génomes séquencés montre clairement que les diatomées sont les descendantes d'une cellule eucaryote ayant ingéré une autre cellule eucaryote, une algue rouge (ou rhodophyte). Le génome mitochondrial de l'algue a été perdu alors qu'une partie des gènes nucléaires a été transférée dans le noyau de la cellule hôte. Le chloroplaste rhodophyte a été conservé, mais est maintenant entouré de deux membranes indépendantes : la membrane interne est celle du chloroplaste de l'algue rouge et la membrane externe est ce qu'il reste de la membrane de la cellule ingérée. Ainsi, l'ingestion d'une cellule eucaryote par un autre eucaryote a conduit à ce que l'on appelle une endosymbiose secondaire.

DES AMIBES ENDOSYMBIOTIQUES...

Une expérience intrigante va amener de l'eau au moulin de l'endosymbiose. En 1966, un chercheur américain, Kwang Jeon, voit une de ses cultures d'amibes, des eucaryotes unicellulaires, infectée par une bactérie très virulente. Les amibes poussent plus lentement, leur taille est plus petite et elles sont plus fragiles. Les bactéries elles-mêmes sont incapables de se diviser dans aucun milieu de culture testé, prouvant que ce sont des parasites obligatoires de l'amibe. Jeon continue à cultiver ses amibes infectées pendant cinq années, puis se livre à des expériences de microdissection de noyaux d'amibes afin de découvrir si l'infection bactérienne a changé quelque chose au génome eucaryote. À sa grande surprise, la transplantation d'un noyau d'une amibe infectée dans le cytoplasme d'une amibe non infectée n'est

pas viable, alors que l'expérience inverse donne une cellule parfaitement capable de se diviser. De toutes les combinaisons de noyaux et de cytoplasmes testés, seule celle impliquant la transplantation d'un noyau infecté dans une cellule non infectée est létale. Jeon en conclut que le noyau d'une amibe infectée avait besoin de la présence des bactéries dans le cytoplasme pour permettre la survie de la cellule. Les deux organismes étaient devenus des symbiotes obligatoires. Ce résultat était la première mise en évidence expérimentale de la possibilité de créer, dans des conditions de laboratoire contrôlées, un symbiote entre un eucaryote et une eubactérie.

... ET DES LEVURES ENDOSYMBIOTIQUES

Beaucoup plus récemment, des équipes californiennes ont réussi à créer en laboratoire des symbioses entre une bactérie, *Escherichia coli*, et la levure de boulangerie, *Saccharomyces cerevisiae*, un eucaryote modèle en génétique. L'expérience est fascinante. Le génome de la bactérie est modifié pour la rendre incapable de synthétiser la thiamine (ou vitamine B1), indispensable à sa croissance. Si on ne fournit pas de thiamine dans le milieu de culture, la bactérie meurt. En parallèle, la chaîne respiratoire mitochondriale de la levure est mutée. Ces cellules modifiées sont incapables de pousser dans un milieu qui les oblige à respirer, par exemple en présence de glycérol comme seul nutriment. En revanche, elles sont parfaitement capables de synthétiser la thiamine. On a donc d'un côté des bactéries qui ne peuvent pousser qu'en présence de thiamine et d'un autre côté des levures incapables de pousser sur un milieu contenant du glycérol. Dans des conditions expérimentales particulières, qui favorisent la fusion entre les membranes bactériennes et fongiques, les auteurs de cette étude ont été capables de sélectionner des cellules qui sont capables de pousser dans un milieu sans thiamine et en présence uniquement de glycérol comme seule source d'énergie. L'analyse par microscopie des levures obtenues montre clairement qu'elles contiennent de façon stable dans leur cytoplasme un petit nombre de bactéries vivant en symbiose avec

elles. Ces dernières apportent aux levures la chaîne respiratoire qui leur permet d'utiliser le glycérol pour leur croissance, en échange de la thiamine qu'elles sont incapables de synthétiser. Un parfait exemple d'endosymbiose reproduit de façon spectaculaire en laboratoire !

QUAND L'ÉVOLUTION RÉTRÉCIT LES GÉNOMES

Certains organismes eucaryotes ne possèdent pas de mitochondries. C'est le cas des archézoaires, tels *Trichomonas vaginalis*, un pathogène sexuellement transmissible, colonisant les muqueuses du tractus urogénital humain. Cet archézoaire flagellé fait partie des excavates, un lignage encore mal caractérisé, mais proche des eucaryotes ancestraux. La séquence de son génome a révélé un nombre élevé de gènes pour un organisme primitif (environ 60 000, 2-3 fois plus que le génome humain), mais cela est très certainement surestimé, en raison principalement du très grand nombre d'éléments répétés qui rendent difficile l'assemblage précis de ce génome. *Trichomonas vaginalis* est un organisme microaérophile, c'est-à-dire qu'il ne peut survivre que dans un environnement pauvre en oxygène, d'où sa niche écologique. Afin de contrer les effets néfastes de l'oxygène, il utilise une grande variété de gènes codant des protéines aux propriétés antioxydantes. Mais parmi les trouvailles remarquables de ce génome atypique, ont été découverts les gènes impliqués dans le métabolisme de l'hydrogénosome. Il s'agit d'un organite intracellulaire qui permet la synthèse d'ATP par la fermentation de métabolites cellulaires. Cette synthèse est associée à un dégagement d'hydrogène, d'où son nom. L'hydrogénosome, comme la mitochondrie, provient également d'une endosymbiose bactérienne. En revanche, il ne contient pas de gènes, à l'inverse des mitochondries. La plupart des gènes codant des protéines chargées de fonctions mitochondriales sont absents du génome de ce protiste, montrant que l'hydrogénosome a subi depuis l'endosymbiose primitive une évolution réductrice qui aboutit à une sorte de mitochondrie simplifiée, capable uniquement d'assurer une fonction énergétique très primitive.

Un autre exemple d'eucaryote unicellulaire ayant subi une simplification drastique de son génome est trouvé chez les microsporidies, des parasites intracellulaires obligatoires d'autres cellules eucaryotes, très proches des champignons et des levures comme *Saccharomyces cerevisiae*. Le génome de l'une d'entre elles, *Encephalitozoon cuniculi*, fut en 2001 l'un des tout premiers génomes eucaryotes séquencés. Il contient 11 chromosomes dont la taille totale (2,9 millions de nucléotides) en fait l'un des plus petits génomes eucaryotes connus. Cette taille réduite est due non seulement à un faible nombre de gènes (environ seulement 2 000), mais également à leur taille, 15 % plus courte que les protéines **homologues*** chez la levure *Saccharomyces cerevisiae*, ainsi qu'à la faible distance séparant les gènes. *Encephalitozoon cuniculi* possède également une mitochondrie minimum, mais différente de l'hydrogénosome de *Trichomonas vaginalis*, car incapable de fabriquer de l'hydrogène. Cette mitochondrie minimum – appelée mitosome – protège la cellule contre le stress oxydatif et participe au métabolisme cellulaire.

Homologie : Deux gènes (ou protéines) sont homologues quand leurs séquences partagent un certain degré d'identité, indiquant qu'ils ont un ancêtre commun.

LES NOMBREUSES INVENTIONS DU MONDE EUCARYOTE

Nous venons de voir deux exemples d'évolution réductrice chez les eucaryotes, l'un chez les excavates, l'autre chez les microsporidies. Dans les deux cas, les génomes mitochondriaux ont été considérablement réduits par rapport à leur ancêtre eucaryote et ne contiennent plus qu'un nombre limité de gènes pouvant assurer une partie seulement des fonctions mitochondriales. De plus, chez *Encephalitozoon cuniculi*, le génome nucléaire a également été réduit, caractéristique habituelle des organismes parasites, qui détournent la machinerie de la cellule hôte à leur propre profit, leur permettant

de se débarrasser des gènes correspondants. Étant donné le nombre de gènes perdus, ces génomes, bien qu'informatifs, ne peuvent pas rendre compte du contenu génétique de l'ancêtre commun eucaryote. À cette fin, le génome d'un autre protiste, *Naegleria gruberi*, a été séquencé et publié en 2010. Il s'agit d'un excavate unicellulaire, comme *Trichomonas vaginalis*, capable de vivre sous sa forme amibe ou de différencier deux flagelles lui permettant de se déplacer plus rapidement vers la nourriture. Son habitat naturel est l'eau douce ainsi que les sols humides et sa répartition est mondiale. Son génome couvre 41 millions de nucléotides et contient 12 chromosomes. Il peut se reproduire en se divisant clonalement ou bien par reproduction sexuée, bien que celle-ci n'ait jamais été observée en laboratoire. Sa séquence prédit la présence de 15 727 gènes codant des protéines (seulement 25 % de moins que les vertébrés). *Naegleria gruberi* contient des mitochondries dotées d'une chaîne respiratoire conventionnelle. De plus, il possède également un métabolisme anaérobie similaire à ce que l'on trouve chez d'autres excavates comme *Trichomonas vaginalis*, lui permettant de vivre dans des conditions d'hypoxie temporaire comme celle existant dans les environnements boueux qui sont son habitat naturel. La comparaison des gènes de *Naegleria gruberi* montre que 60 % d'entre eux ont un homologue dans le monde procaryote et donc prédatent le monde eucaryote. Parmi ceux-ci, environ 16 % sont de nouveaux gènes correspondant à du « bricolage évolutif », c'est-à-dire à la réutilisation de domaines protéiques procaryotes, créant de nouvelles protéines par « copier-coller » successifs. Les 40 % de gènes sans homologue sont présents uniquement chez les eucaryotes et correspondent donc à l'ensemble minimum de gènes ayant été inventés dans ce règne et permettant de décrire un organisme eucaryote. Le génome de *Naegleria gruberi* a donc permis de définir, dans un organisme n'ayant pas subi d'évolution réductrice de son génome, l'ensemble des gènes présents dans l'ancêtre commun de tous les eucaryotes et montre que la transition du monde procaryote aux cellules à noyau fut accompagnée d'un grand

nombre d'inventions évolutives, inventions dont la plupart seront conservées pendant deux milliards d'années, jusqu'à nos jours, dans tout le lignage eucaryote. Le passage du monde procaryote au monde eucaryote porte donc la signature de milliers de nouveaux gènes, de milliers de nouvelles fonctions, qui n'existaient pas chez les ancêtres bactériens et qui sont la spécificité des cellules à noyau. Aucune autre branche évolutive ultérieure n'a créé autant de nouveautés génétiques que les cellules eucaryotes.

À RETENIR

- Les eucaryotes sont le résultat d'une endosymbiose entre une archée et une eubactérie.
- L'endosymbiose d'une cyanobactérie par un eucaryote ancestral a donné naissance à la lignée des algues et des plantes.
- L'arrivée des eucaryotes est concomitante à la création de milliers de nouvelles fonctions cellulaires, absentes du monde procaryote, et à l'origine de leur succès évolutif.

BIBLIOGRAPHIE

- La théorie endosymbiotique : Martin et Kowallik, *European journal of Phycology* 34: 287-295 (1999).
- L'origine du noyau : López-García et Moreira, *BioEssays* 28: 525-533 (2006).
- Les amibes endosymbiotiques : Jeon, *Science* 176: 1122-1123 (1972).
- Les levures endosymbiotiques : Mehta, Supekova, Chen, Pestonjamas et al., *Proceedings of the National Academy of Sciences USA*: 115: 11796-11801 (2018).

8

Des premiers eucaryotes aux premiers animaux

L'arrivée des premières cellules eucaryotes amène son lot d'innovations évolutives, permettant la complexification des formes et l'adaptation à tous les milieux, du fond des océans au sommet des plus hautes montagnes, et de l'équateur aux pôles. Les premiers eucaryotes unicellulaires, protistes et algues, sont nés dans le milieu marin, vraisemblablement dans des zones relativement calmes où eubactéries et archées partageaient la même niche écologique, rendant possibles les endosymbioses. Il s'agissait d'organismes autonomes capables de fabriquer leur matière organique à partir des ressources minérales de leur environnement. À partir de ces eucaryotes unicellulaires relativement simples, vont se développer des organismes ayant perdu la capacité d'utiliser exclusivement des ressources minérales et devant se nourrir d'autres êtres vivants. Ce sont les premiers animaux unicellulaires, ou protozoaires. Dans le même temps, ces animaux vont développer un flagelle leur permettant d'augmenter leur mobilité et de faciliter leur déplacement vers leurs futures proies. Puis, vont se

développer des organismes plus complexes, assemblages multicellulaires de différents tissus spécialisés qui vont mener aux animaux multicellulaires, ou métazoaires, dont nous faisons partie. Cette évolution va se dérouler en plusieurs étapes et s'étaler sur quelques centaines de millions d'années.

L'ACQUISITION DE LA MOBILITÉ CHEZ LES EUCARYOTES

Les premiers eucaryotes étaient donc des organismes unicellulaires, à l'origine peu mobiles. Cette mobilité leur fut apportée par l'apparition des flagelles ou des cils que l'on retrouve encore à l'heure actuelle chez certains d'entre eux (paramécies, certaines algues) et chez les **métazoaires*** (cellules épithéliales, spermatozoïdes). Elle a permis aux premiers animaux unicellulaires, les protozoaires, de se déplacer pour chercher leur nourriture et chasser leurs proies bactériennes. L'étude des gènes formant le flagelle eucaryote montre que la tubuline qui le compose est similaire à une protéine du **cytosquelette*** bactérien appelée FtsZ, suggérant une origine génétique procaryote au flagelle eucaryote. Néanmoins, le flagelle

Métazoaires : Animaux multicellulaires, à l'opposé des protozoaires, animaux unicellulaires.

Cytosquelette : Le cytosquelette est l'ensemble des fibres protéiques polymérisées qui organisent l'architecture de toutes les cellules vivantes. Il existe des différences importantes entre le cytosquelette eucaryote et procaryote.

eucaryote est complètement différent du flagelle bactérien, tant par sa structure que par le nombre et le type de protéines qui sont impliquées dans son architecture. Le corps basal du flagelle montre une organisation des fibres de tubuline très caractéristique, formant un cercle de neuf triplets de microtubules.

Ce corps basal, conservé chez certaines cellules eucaryotes ayant perdu leur flagelle, va évoluer afin de devenir le centre organisateur

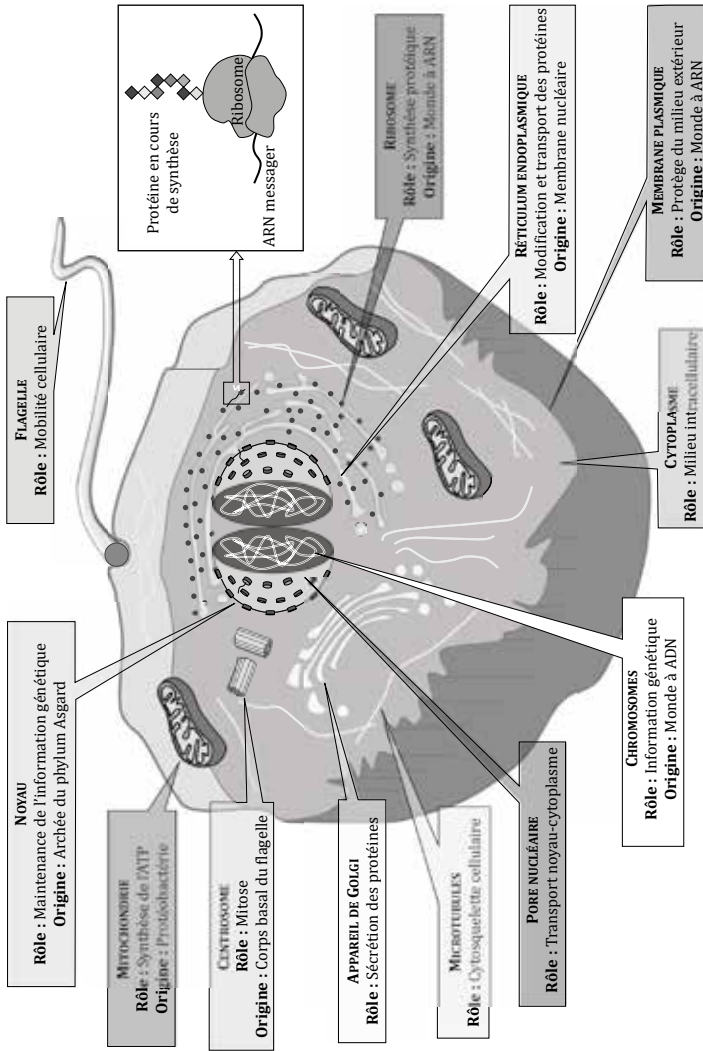


Figure 20 | Schéma d'une cellule eucaryote animale. L'origine des organites est mentionnée quand elle est claire. L'encart à droite montre le grossissement d'un ribosome en cours de traduction.

des microtubules (ou centrosome), qui va permettre la bonne séparation des chromosomes dans chaque cellule fille lors de la division cellulaire, que l'on nomme mitose chez les eucaryotes (Figure 20).

Le séquençage complet du génome d'une algue unicellulaire doté de deux flagelles, *Chlamydomonas reinhardtii*, a confirmé l'origine des cils eucaryotes. Le génome de cette algue contient 15 143 gènes, dont 195 sont impliqués dans la formation des flagelles. La comparaison de ces gènes avec ceux des animaux et des plantes montre clairement que les flagelles de *Chlamydomonas* ont la même origine que les cils des animaux et qu'ils ont été perdus chez les plantes supérieures. Le corps basal de ces flagelles est bien l'homologue fonctionnel du centrosome, avec son architecture caractéristique en neuf triplets de tubuline. Cette étude confirme que les eucaryotes ont inventé leur propre flagelle et que chez certains d'entre eux, la base de ce flagelle a servi de support à l'organisation et à la bonne séparation des chromosomes pendant la mitose.

LES GÈNES EN MORCEAUX DES EUCARYOTES

Il semble vraisemblable que les animaux multicellulaires sont apparus il y a environ 1 milliard d'années, avant l'explosion des nombreuses espèces de l'époque précambrienne, que nous verrons un peu plus tard. Ils sont le résultat de coopération cellulaire entre eucaryotes primitifs. Les eucaryotes unicellulaires les plus proches des métazoaires sont les choanoflagellés, de forme ovoïde, possédant un long cil entouré d'un collier de microvillosités caractéristique et pouvant vivre en colonies de quelques dizaines à quelques centaines d'individus. Depuis le XIX^e siècle, des similarités morphologiques frappantes entre les choanoflagellés et les éponges avaient été observées. Les éponges (ou spongiaires) sont les plus primitifs des métazoaires. Ces animaux vivent accrochés à leur substrat et se nourrissent de bactéries, de débris organiques et d'algues unicellulaires en filtrant l'eau qui les contient. En 2008, le génome du choanoflagellé *Monosiga brevicollis* a été entièrement séquencé par un consortium

de scientifiques américains. Son génome est de la taille de celui d'un champignon et contient environ 9 200 gènes, à peine moins de la moitié des gènes du génome humain. La comparaison des autres génomes eucaryotes avec celui de *M. brevicollis* montre clairement qu'il existe un ancêtre commun direct entre les métazoaires et les choanoflagellés. La structure des gènes de *M. brevicollis* est très similaire à celle des métazoaires, contenant beaucoup d'**introns***, quoique souvent de petite taille en comparaison.

Un intron est une séquence présente dans l'ADN, qui se retrouve transcrite dans l'ARN, mais qui sera excisée par un mécanisme extrêmement précis afin de former l'ARN messager mature qui sera traduit par la suite en protéine par le ribosome. Les séquences se retrouvant dans l'ARNm sont appelées **exons*** (Figure 21A). Les gènes procaryotes ne contiennent que très rarement des introns et les quelques cas observés correspondent toujours à des ribozymes, reliques du monde à ARN. En revanche, tous les génomes eucaryotes en contiennent, en nombre variable. Celui de la levure *Saccharomyces cerevisiae* n'en possède qu'une poignée, alors que tous les gènes humains ou presque en contiennent plusieurs. Le mécanisme d'excision des introns et de raboutage (ou épissage) des exons est un processus complexe et hautement régulé qui nécessite une machinerie conservée dans tout le monde eucaryote. Certains ARN messagers peuvent subir des épissages alternatifs, en fonction des conditions de vie cellulaire, des types de tissus ou d'autres facteurs d'origine endogène ou exogène. L'épissage alternatif permet ainsi à un seul gène de coder plusieurs protéines (Figure 21B). Le génome humain contient environ 20 000 gènes, mais on estime que ces gènes codent vraisemblablement 100 000 protéines, peut-être le double, grâce à l'épissage alternatif. On voit donc que la structure morcelée des gènes eucaryotes est beaucoup plus complexe que celle des gènes procaryotes. Contrairement à l'aphorisme souvent attribué à Jacques Monod, ce qui est vrai pour les bactéries ne l'est pas pour l'éléphant et ceci se confirmera pour d'autres mécanismes du vivant. Pour en revenir à

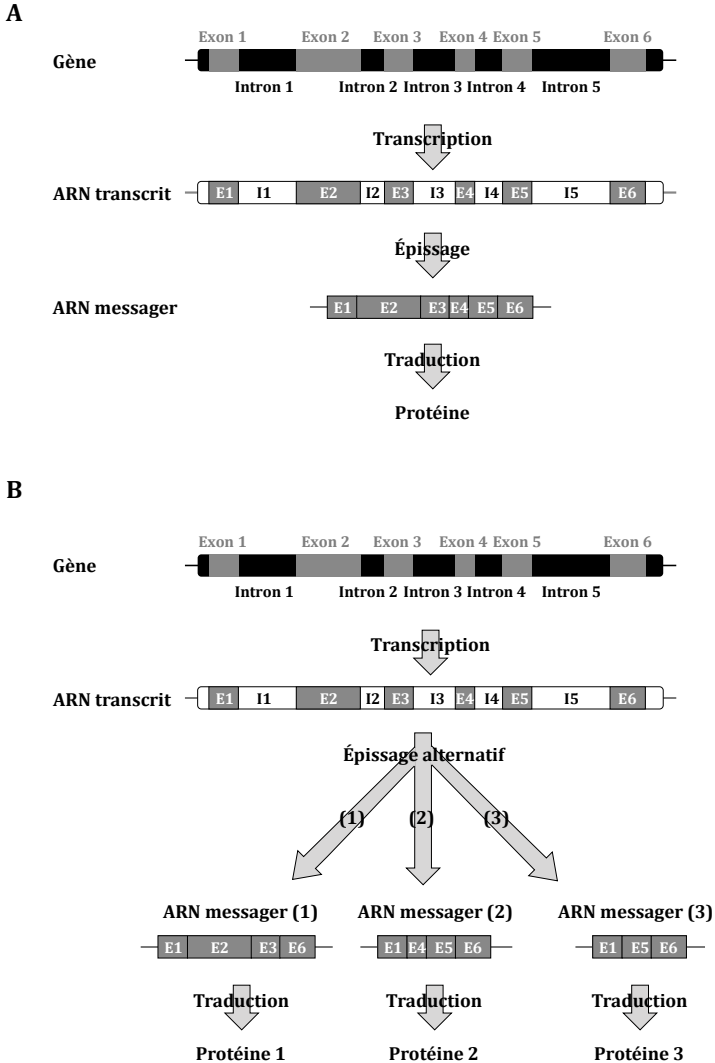


Figure 21 | Épissage et épissage alternatif. A : Épissage. Un gène formé par six exons et cinq introns est transcrit en ARN, qui va subir l'excision des introns et l'épissage des exons pour donner l'ARN messager. **B : Épissage alternatif.** Le même gène peut subir trois épissages alternatifs (par exemple dans trois tissus différents) pour donner trois ARNm contenant des exons différents qui seront traduits en trois protéines sensiblement différentes.

notre choanoflagellé, son génome possède donc un grand nombre d'introns, une caractéristique typique des métazoaires. Mais il y a d'autres conclusions importantes qui furent tirées de son séquençage.

DES PROTOZOAIRES AUX MÉTAZOAIRES

De façon remarquable, *M. brevicollis* possède 78 gènes que l'on ne retrouve que chez les animaux, dont un grand nombre sont impliqués dans la signalisation cellulaire et les processus d'adhésion qui sont à l'origine de la multicellularité. Parmi ceux-ci, l'invention du collagène a été déterminante pour l'apparition des métazoaires. Le collagène est la protéine la plus abondante du règne animal. Elle est fibreuse, inextensible et résistante à la traction, formant la matrice structurelle de tous les organismes animaux. Le génome de *M. brevicollis* contient plusieurs gènes codant des protéines à collagène, ainsi que d'autres codant des protéines impliquées dans la matrice extracellulaire, matrice qui va déterminer les différentes structures des formes animales. D'autre part, ce choanoflagellé contient également des gènes d'intégrines responsables chez les métazoaires de l'interaction avec les fibres de collagène, ainsi que cinq gènes contenant des domaines d'immunoglobulines, protéines initialement impliquées dans l'adhésion cellulaire avant d'avoir été détournées de leur fonction première pour devenir la base du système immunitaire des vertébrés. Enfin, le génome de *M. brevicollis* contient plusieurs gènes spécifiques des animaux, codant des protéines régulant la transcription, dont la protéine p53, centrale à tous les mécanismes de signalisation cellulaire, dont des formes mutantes sont fréquemment trouvées associées à plusieurs cancers chez l'homme.

Le séquençage récent d'autres animaux unicellulaires a permis de confirmer ces observations et d'affiner les modèles évolutifs du monde animal. *Creolimax fragrantissima* et *Capsaspora owczarzaki* sont deux amibes, des animaux unicellulaires aux cycles de vie très différents de *M. brevicollis*. La première se fixe à un endroit en formant un cyste qui va subir plusieurs cycles de réplication de son noyau sans division

cellulaire, pour former un cénocyte, une cellule géante contenant plusieurs noyaux à sa périphérie. Celui-ci va finir par éclater en relâchant dans le milieu extérieur autant d'amibes qui vont recommencer ce cycle. *C. owczarzaki* a un cycle de vie encore très différent, passant d'une vie unicellulaire sous forme de cyste à la multicellularité en produisant une matrice extracellulaire agrégeant plusieurs cellules sans que celles-ci ne se touchent directement. Le séquençage de ces deux espèces d'amibes, en plus de celui de *M. brevicollis*, a permis d'affiner le modèle évolutif ayant mené aux métazoaires. Il a montré que tous les animaux unicellulaires – et donc leur ancêtre commun – contenaient déjà les gènes codant les protéines d'adhésion cellulaire, ceux codant les protéines de signalisation permettant les échanges d'information entre cellules, ainsi que les gènes codant les facteurs de transcription, régulant finement l'expression de nombreux gènes. Ces gènes, existant chez les protozoaires ont été réutilisés de façon différente et adaptée aux besoins spécifiques des métazoaires. L'étude des modes de vie des protozoaires suggère que l'ancêtre commun unicellulaire des métazoaires se nourrissait de bactéries, pratiquait la reproduction sexuée et était capable de différencier des types cellulaires spécialisés.

À l'aune de ces découvertes dues au séquençage des génomes d'animaux unicellulaires, on peut donc maintenant essayer de retracer un scénario possible des protozoaires aux métazoaires. L'ancêtre des métazoaires était une cellule animale capable de se reproduire elle-même de façon clonale, et en présence de stimuli environnementaux (lumière, nourriture, température, etc.) de différencier certains états cellulaires, par exemple de changer de sexe pour permettre la reproduction sexuée, ou bien de fabriquer un flagelle pour se déplacer (Figure 22A). Ces différents types cellulaires se retrouvent intégrés dans un *protométazoaire*, mais toujours dépendants des stimuli environnementaux. Dans le même temps, des mécanismes additionnels de régulation de la transcription et de communication intercellulaire se mettent en place. Une matrice extracellulaire permet de

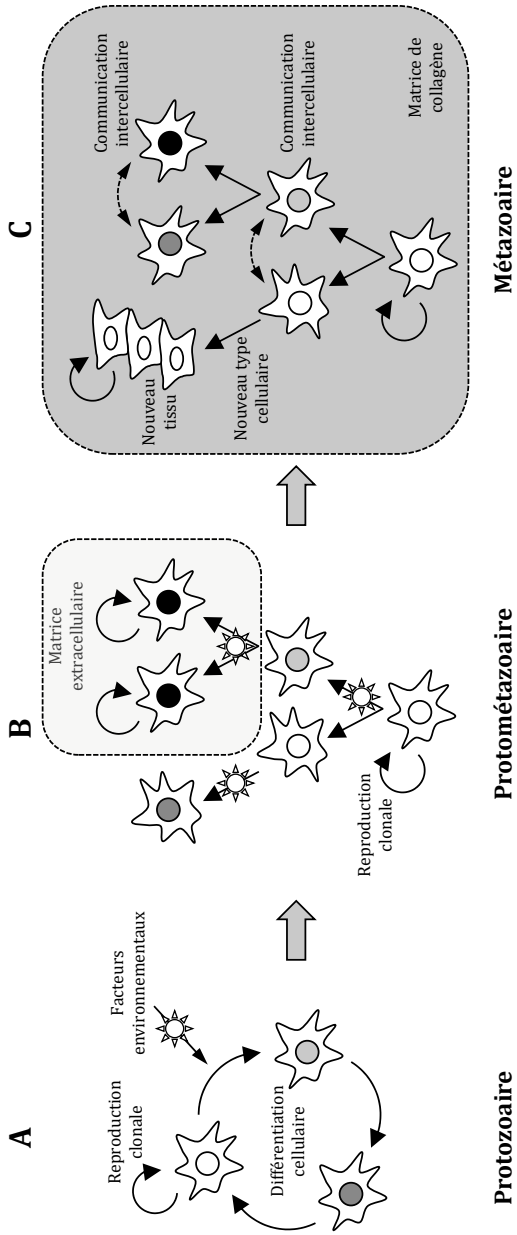


Figure 22 | L'origine des métazoaires.

maintenir ensemble ce réseau de cellules, mais la taille de l'organisme est encore limitée, sans doute quelques dizaines à quelques centaines de cellules (Figure 22B). Dans la dernière étape, la différenciation cellulaire devient indépendante des stimuli environnementaux, et ne dépend plus que de programmes développementaux établis et maintenus par communication intercellulaire. De nouveaux types cellulaires vont apparaître lors du développement, qui vont former des tissus aux destins spécifiques, maintenus ensemble par des fibres de collagène. Certains vont assurer une fonction nutritive, une fonction motrice, une fonction excrétrice ou d'autres fonctions tissulaires spécifiques (Figure 22C). Le premier métazoaire est né. Originaire des océans, il lui faudra encore plusieurs dizaines de millions d'années avant de pouvoir gagner la terre ferme.

À RETENIR

- Les animaux ont la nécessité de se nourrir d'autres êtres vivants, bactéries, champignons, plantes ou autres animaux. Leur matrice extracellulaire est principalement formée de collagène.
- Les animaux multicellulaires (métazoaires) ont évolué à partir d'animaux unicellulaires (protozoaires).
- Les génomes des protozoaires contenaient déjà l'ensemble des gènes des futurs métazoaires.
- Les gènes des eucaryotes sont morcelés par des introns qui augmentent la taille et la complexité des génomes eucaryotes.

BIBLIOGRAPHIE

- Le génome de *Chlamydomonas reinhardtii* : Merchant, Prochnik, Vallon, Haris *et al.*, *Science* 318: 245-251 (2007).
- Le génome de *Monosiga brevicollis* : King, Westbrook, Young, Kuo *et al.*, *Nature* 451: 783-788 (2008).
- L'origine des métazoaires : Sebé-Pedrós, Degnan et Ruiz-Trillo, *Nature Reviews Genetics* 18: 498-512 (2017).

9

Des métazoaires à l'homme

L'homme moderne, ou *Homo sapiens*, est un animal bilatérien, deutérostomien, chordé, vertébré, tétrapode, mammifère, thérien, primate (Figure 23). Cet apparent charabia représente en fait un complexe système de classification du monde vivant en différents rangs **taxonomiques*** correspondant à des propriétés morphologiques ou développementales. Les bases de cette classification furent posées par le grand naturaliste suédois, Carl von Linné au XVIII^e siècle. Les principes de cette classification systématique des êtres vivants évoluèrent avec le temps au siècle suivant, sous l'impulsion des découvertes de Charles Darwin sur la sélection naturelle et sur lesquelles nous reviendrons plus tard. Au XX^e siècle et surtout au XXI^e, le séquençage des génomes

Taxonomie : Domaine de la biologie qui s'intéresse à la classification des êtres vivants, selon des critères morphologiques, physiologiques ou embryologiques. Un taxon est l'ensemble des êtres vivants partageant les mêmes caractéristiques taxonomiques.

complets a permis de classer les êtres vivants dans une **phylogénie*** qui décrit les relations entre eux d'après la proximité plus ou moins grande des séquences de leurs gènes, un critère plus discriminant et moins sujet à controverse que leur morphologie ou de leur physiologie. Nous verrons un peu plus tard comment sont établies ces phylogénies dites « moléculaires », car utilisant la séquence des gènes.

Phylogénie : Étude des liens de parenté entre êtres vivants, entre populations et entre espèces.

Homo sapiens, comme tous les animaux, est un eucaryote (Figure 23). Un de nos lointains ancêtres était un animal unicellulaire qui vivait dans le milieu marin, il y a environ 1,5 milliard d'années, avant de devenir un métazoaire, puis un animal bilatérien. Ceux-ci présentent une symétrie bilatérale qui définit à la fois un axe antéro-postérieur (avant-arrière) et un axe dorso-ventral (dos-ventre), permettant d'orienter l'animal dans l'espace afin de favoriser ses déplacements dans une seule direction. Cela permet à l'animal de trouver plus facilement sa nourriture et de développer une spécialisation de la tête et de la queue. À partir de là vient la séparation des animaux en deux grandes branches, ceux dont la bouche se forme avant l'anus, appelés protostomiens, et qui ont donné naissance aux lignées des insectes et des vers, et les deutérostomiens, ceux dont l'anus se forme en premier, leur nom signifiant « bouche en second » en grec. Les poissons, les oiseaux et les mammifères, dont *Homo sapiens*, sont des deutérostomiens. Chez ces derniers arrive une invention importante, celle de la chorde, sorte de baguette rigide sur laquelle repose le système nerveux central primitif, et qui va se différencier sur le dos de l'animal, donnant son nom à la lignée des chordés.

Cette chorde sera remplacée chez les vertébrés par la colonne vertébrale, sur laquelle va se fixer un squelette osseux ou cartilagineux interne qui représente la structure autour de laquelle est construit le corps de l'animal. Les vertébrés apparaissent il y a environ 450 millions

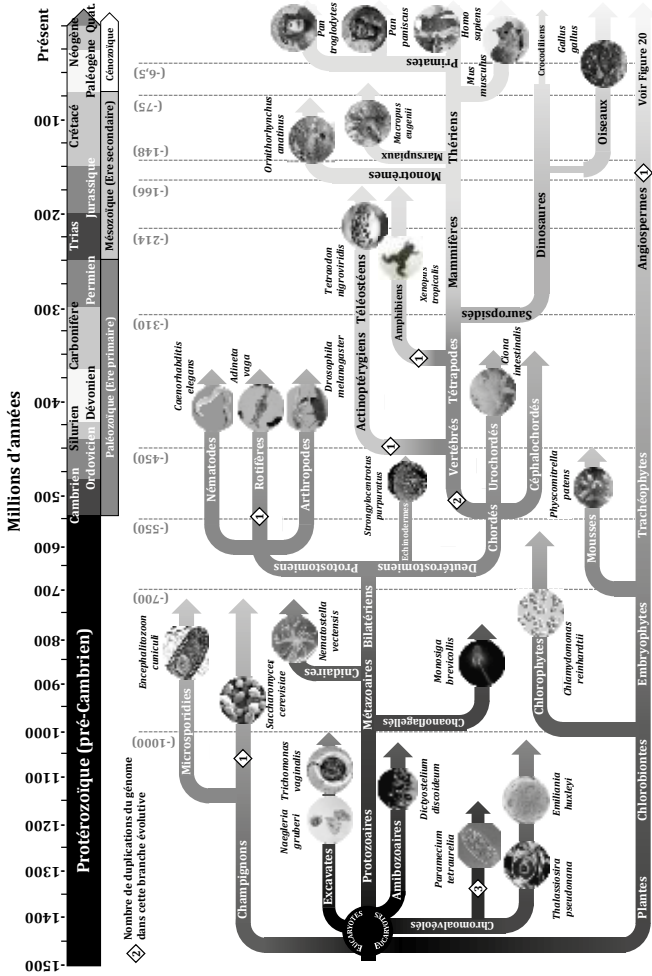


Figure 23 | Arbre du vivant des premiers eucaryotes jusqu'à nos jours. La flèche représente les ères et périodes géologiques. Les différentes branches de l'arbre des eucaryotes sont représentées, les flèches indiquant que les branches continuent jusqu'à l'époque actuelle. Les espèces représentées ont leur génome séquencé. Les dates de divergence des différentes branches sont indiquées par des traits verticaux pointillés gris. Quat. : quaternaire.

d'années et sont un formidable succès évolutif car on les trouve dans toutes les niches écologiques à la surface de la Terre ou dans les océans. Certains de ces vertébrés vont sortir du milieu marin pour gagner la terre ferme. Ce sont les tétrapodes, qui possèdent deux paires de membres et respirent à l'aide d'un poumon. Les mammifères, dont nous faisons partie, sont des tétrapodes qui nourrissent leurs nouveau-nés à l'aide d'une glande cutanée produisant du lait, d'où le nom de cette lignée dérive. Certains de ces mammifères sont des thériens, c'est-à-dire qu'ils portent leur progéniture dans un utérus pendant tout ou partie de leur développement. Chez les marsupiaux, la fin du développement de l'embryon se fait en dehors de l'utérus, dans une poche ventrale, nommée *marsupium*. Les marsupiaux regroupent entre autres les kangourous, koalas, wallabies, opossums et wombats, la plupart d'entre eux vivant en Australie, Tasmanie et Nouvelle-Guinée, bien que certaines espèces d'opossums vivent sur le continent américain. Chez les mammifères placentaires (ou euthériens, donc possédant un « vrai utérus »), le développement se fait entièrement à l'intérieur de l'utérus, grâce à l'expansion du placenta, tissu nourricier qui permet de subvenir aux besoins de l'embryon jusqu'à la naissance. Nous reviendrons sur l'origine du placenta, qui a été « inventé » au moins deux fois dans la lignée des mammifères thériens. Finalement, au sein des mammifères placentaires, les primates regroupent tous les singes et bien sûr les **hominidés***, dont les humains. Le cheminement a été long depuis les premiers métazoaires, sans doute guère plus que quelques dizaines de cellules coopérant au sein d'un organisme contenant des tissus très simples, jusqu'aux nombreuses spécialisations des vertébrés modernes. Nous allons maintenant voir en détail l'arrivée de ces différentes innovations évolutives.

Hominidé : Les hominidés ou grands singes, comprennent les bonobos, les chimpanzés, les gorilles, les orangs-outans, les humains, ainsi que plusieurs lignées disparues dont certaines étaient des ancêtres de l'Homme.

LES FOSSILES, TÉMOINS DE L'EXPLOSION PRÉCAMBRIENNE

Les paléontologues du XVIII^e siècle avaient déjà observé la présence étonnante de très nombreux fossiles dans des couches géologiques datant du Cambrien (– 541 à – 485 millions d'années), la première période de l'ère primaire. Les fossiles sont les restes minéralisés d'un animal ou d'une plante, conservés à l'intérieur d'une couche sédimentaire. Ce processus, long et rare, résulte d'un dépôt de l'animal ou de la plante sur un fond marin, qui va être recouvert au fil des années par des sédiments. Lors de la décomposition de l'animal, les tissus vivants sont progressivement remplacés par des substances minérales qui font une sorte d'empreinte de l'animal. Cette empreinte va se retrouver piégée au sein d'une roche sédimentaire jusqu'à sa découverte accidentelle ou lors de fouilles paléontologiques. Les parties molles des animaux se décomposant plus rapidement que les parties calcifiées (coquille, squelette), on ne retrouve en général que l'empreinte des parties dures des animaux, la fossilisation des parties molles ne pouvant se faire que dans des conditions sédimentaires exceptionnelles.

Échelles des temps géologiques : Les temps géologiques sont divisés en quatre éons (Hadéen, Archéen, Protérozoïque et Phanérozoïque). Le Phanérozoïque (« éon de la vie visible » en grec) est subdivisé en trois ères. Le Paléozoïque en est la première ère, les deux autres étant le Mésozoïque et le Cénozoïque (voir Figure 23).

Les fossiles découverts dans les roches sédimentaires datant du Cambrien sont très diversifiés. On y trouve de nombreux **phylums*** tels les célèbres trilobites (anciens arthropodes marins), des petits escargots, des échinodermes et des crustacés. La plupart des phylums

Phylum : Le phylum (ou embranchement) est le rang de classification des espèces vivantes juste en dessous du règne. Les eucaryotes sont divisés en cinq règnes : protozoaires, chromistes, champignons, plantes et métazoaires. Il existe 34 phylums différents au sein du règne des métazoaires.

modernes des métazoaires sont représentés dans ces couches fossilisées, suggérant une explosion des espèces animales à cette époque. Pendant près de deux siècles perdura cette idée d'une « explosion cambrienne » du vivant, qui aurait eu lieu il y a environ 500 millions d'années. L'analyse de gènes provenant de différents phylums de métazoaires modernes permit il y a une vingtaine d'années de réviser cette théorie. En comparant les séquences de huit gènes chez plusieurs métazoaires, il a été montré que l'ancêtre commun à tous ces organismes vivait il y a 1-1,2 milliard d'années, une époque bien plus reculée que les 500 millions d'années suggérés par les restes fossilisés des roches sédimentaires du Cambrien. Il est donc clair que les fossiles trouvés dans les sédiments du Cambrien correspondent à des phylums animaux qui existaient depuis des centaines de millions d'années quand ils ont été fossilisés. Comment se fait-il alors que l'on n'ait pas retrouvé – pour le moment – de restes de ces animaux ou de leurs ancêtres dans des couches sédimentaires précambriennes ? Les fossiles représentent essentiellement les parties rigides, osseuses, des animaux. Il est très rare que les parties molles d'un organisme arrivent à se fossiliser plus rapidement qu'elles ne sont naturellement décomposées. Il est possible que des ancêtres à corps mous de la faune cambrienne aient existé depuis des centaines de millions d'années, mais n'aient pas été fossilisés. La fossilisation étant elle-même un phénomène naturel rare, il est aussi possible que des fossiles précambriens existent, mais n'aient pas encore été retrouvés. Les fossiles pouvant eux-mêmes être détruits par les mouvements tectoniques des roches sédimentaires, il est également possible que, si de tels fossiles aient jamais existé, ils n'aient pu parvenir intacts jusqu'à nous. Des fouilles approfondies dans des couches sédimentaires vieilles d'un milliard d'années, dans plusieurs régions du globe seraient nécessaires pour éclaircir définitivement cette question.

LES HORLOGES MOLÉCULAIRES

Une méthode plus moderne de datation des espèces se base sur les séquences d'ADN. On a longtemps utilisé pour comparer les génomes

entre eux la séquence de l'ADN ribosomique par commodité ; celui-ci étant présent en grand nombre de copies dans les cellules, il est relativement facile d'en obtenir une séquence à peu près complète. On dispose maintenant de nombreuses séquences de génomes entiers, ce qui rend les calculs plus précis, mais montre aussi que tous les gènes n'évoluent pas à la même vitesse, c'est-à-dire que leur horloge moléculaire est différente. Avec le temps, des mutations apparaissent dans le génome des êtres vivants. Ces mutations sont le principal moteur de l'évolution du vivant. Certaines de ces mutations sont délétères pour l'organisme et vont donc être éliminées, mais d'autres vont présenter un avantage ou bien être simplement neutres et vont éventuellement se fixer dans le génome et dans la population. La vitesse de **fixation*** de mutations dans un organisme vivant en fonction du temps détermine la vitesse de l'horloge moléculaire dans cet organisme. Cette horloge permet de dater l'époque de la divergence entre deux lignées du vivant, en comparant les séquences des gènes **orthologues***, c'est-à-dire des gènes qui ont la même origine évolutive. Par exemple, les gènes codant la tubuline ont la même origine chez tous les animaux, la tubuline humaine descendant de la tubuline des premiers métazoaires. Ce sont donc des gènes orthologues. Lorsque l'on connaît le nombre de mutations fixées dans plusieurs gènes orthologues entre deux espèces

Fixation : Une mutation dans un gène va vers la fixation lorsqu'elle se propage au cours des générations au détriment d'une autre mutation dans le même gène qui, elle, va tendre à disparaître. Une mutation fixée finit par envahir complètement la population, de façon à ce que sa fréquence atteigne 100 % dans cette population.

Gènes orthologues : Deux gènes homologues appartenant à deux organismes différents sont dits orthologues s'ils proviennent d'un gène unique ayant appartenu à l'ancêtre commun des deux organismes avant leur **spéciation***.

données, il suffit de connaître le temps qui sépare l'apparition de ces deux espèces pour calculer la vitesse de l'horloge moléculaire, habituellement exprimée en nombre de mutations par millions d'années. Une fois déterminée la vitesse de l'horloge moléculaire dans une branche du vivant, on peut dater l'apparition d'une nouvelle espèce dans cette branche par rapport aux espèces déjà caractérisées. En revanche, on ne peut calculer d'horloge moléculaire que pour des espèces dont on possède des fossiles dont l'apparition peut être assez précisément datée.

L'utilisation d'une horloge moléculaire est très efficace pour dater l'âge de la divergence entre deux phylums, sous deux réserves. Premièrement, si la divergence des deux espèces est trop lointaine, l'horloge n'est plus linéaire car apparaissent des phénomènes de **saturation*** des mutations. Si le même nucléotide a muté plus d'une fois, on ne peut plus calculer l'époque de la divergence entre les deux séquences. On ne peut donc appliquer cette méthode qu'à des organismes dont la divergence évolutive est relativement récente et dont on possède des fossiles, ce qui limite quelque peu son application et nous ramène aux limitations des fossiles eux-mêmes. Deuxièmement, à des distances trop courtes, l'horloge n'est plus fiable car les mutations observées entre deux gènes orthologues représentent deux versions possibles (et viables) du même gène, aucune des deux versions n'ayant encore été fixée dans la population. Pour toutes ces raisons, il peut donc exister un décalage temporel important entre la datation d'une série de fossiles et l'estimation de l'âge de sa divergence avec une autre espèce, calculé à partir de l'horloge moléculaire. Dans la suite de ce livre, la distinction

Saturation : Une position dans deux gènes orthologues est dite saturée quand le nucléotide correspondant a subi plus d'une mutation depuis l'ancêtre commun de ces deux gènes. Il n'est donc pas possible de déterminer l'âge de leur divergence en utilisant une horloge moléculaire.

sera donc faite entre les deux modes de calcul. Nous appellerons une datation « DFC » lorsque son âge aura été déterminé à partir du Dernier Fossile Connue, ou bien « HM » lorsqu'elle aura été déterminée à partir d'une Horloge Moléculaire. En appliquant cette règle aux métazoaires, on dira donc qu'ils sont apparus sur Terre il y a 1,2-1 milliard d'années HM ou bien 500 millions d'années DFC.

L'APPARITION DE LA SYMÉTRIE BILATÉRALE

Chez les animaux bilatériens, la formation de deux axes orthogonaux (antéro-postérieur et dorso-ventral) permet d'orienter l'organisme lors de son développement. Cette orientation différencie certaines parties qui seront plus tard spécialisées dans des fonctions précises, comme la bouche ou l'anus. Les bilatériens représentent 90 % des métazoaires existants. Cette réussite évolutive est mise sur le compte de l'avantage que représente l'orientation du corps, permettant de se déplacer de façon plus rapide et plus efficace vers les sources de nourriture. Les animaux non bilatériens sont regroupés dans plusieurs phylums, dont le plus important est celui des cnidaires. Il comprend les anémones de mer, les coraux, les hydres (ou polypes) et les méduses. La plupart des cnidaires montrent une symétrie radiale et non bilatérale. Néanmoins, les anémones de mer présentent une surprenante symétrie bilatérale. Chez les animaux bilatériens, la mise en place des deux axes au cours du développement est sous le contrôle des gènes *Hox*, une famille de gènes **paralogues*** qui détermine la formation de l'axe antéro-postérieur, ainsi que du gène *decapentaplegic (dpp)* qui contrôle la formation de l'axe dorso-ventral. Chez l'homme, il existe une quarantaine de gènes *Hox* regroupés sur quatre chromosomes différents. L'analyse des gènes de l'anémone de mer *Nematostella vectensis* a montré que cet animal possédait cinq

Gènes paralogues : Deux gènes appartenant au même organisme sont dits paralogues s'ils proviennent d'un événement de duplication d'un gène dans cet organisme.

gènes *Hox* ainsi que le gène *dpp*. Cette découverte démontre que les gènes impliqués dans le bilatéralisme ne sont pas une invention des animaux bilatériens, mais qu'ils étaient déjà présents dans l'ancêtre commun des cnidaires et des bilatériens (Figure 23). Cette découverte n'est pas sans rappeler l'observation du chapitre précédent montrant que les gènes à l'origine du développement des métazoaires étaient déjà présents dans l'ancêtre unicellulaire.

LES PREMIERS CHORDÉS

L'ancêtre commun de tous les chordés vivait il y a plus de 550 millions d'années et a donné trois sous-phylums : les céphalochordés, les urochordés et les vertébrés. Les céphalochordés sont des animaux rudimentaires ressemblant vaguement aux poissons. Le plus répandu est l'amphioxus, un petit animal marin de quelques centimètres de long au corps allongé. Les urochordés (appelés aussi tuniciers) sont des ascidies, des animaux en forme d'outre, qui filtrent le milieu marin pour se nourrir et qui possèdent une tunique cuticulaire produite par leur épiderme. Ce sont les plus proches cousins des vertébrés. Le génome d'un urochordé, une ascidie, la cione *Ciona intestinalis*, fut entièrement séquencé en 2002, nous apportant des informations essentielles sur les déterminants spécifiques du phylum des chordés. Le génome de la cione contient 150-160 millions de bases, soit 20 fois moins que le génome humain. Il existe une certaine incertitude sur sa taille exacte à cause de la présence d'environ 10 % de séquences d'ADN répété que l'on ne peut pas précisément attribuer à un chromosome ou à un autre. Au total, environ 16 000 gènes distincts ont été identifiés, du même ordre de grandeur que d'autres invertébrés comme la mouche du vinaigre (*Drosophila melanogaster*). De façon générale, il a été remarqué que les familles de gènes paralogues, fréquentes chez tous les vertébrés, étaient souvent représentées par un gène unique chez la cione, suggérant fortement que la formation de cette famille par duplications successives s'est produite après la divergence entre urochordés et vertébrés. Son génome contient

la plupart des gènes responsables du fonctionnement du **système endocrinien*** des vertébrés, dont le système thyroïdien, mais pas les récepteurs aux hormones sexuelles qui semblent donc être une innovation des vertébrés.

Système endocrinien : Ensemble des organes impliqués dans la synthèse et la sécrétion des hormones, qui sont les messagers de l'organisme.

On y trouve également les enzymes responsables de l'**apoptose***, un mécanisme de mort cellulaire programmée qui permet d'éliminer les cellules malades ou abîmées, mécanisme qui est souvent dérégulé dans les cancers, aboutissant à la prolifération incontrôlée des cellules tumorales. Le génome de cet urochordé contient également les gènes responsables de la formation des muscles striés et du muscle cardiaque et le cœur de la cione a une origine génétique similaire à celle du cœur des vertébrés. Plusieurs gènes du système nerveux central bien connus chez les vertébrés sont également présents et participent à la formation d'un cerveau primitif d'environ 300 cellules et d'une corde de 40 cellules. Neuf gènes *Hox* ont pu être identifiés, qui se trouvent tous regroupés sur le même chromosome. En revanche, pas d'hémoglobine chez la cione, alors qu'elle est présente chez tous les vertébrés, mais un autre transporteur d'oxygène, l'hémocyanine, également présent chez les arthropodes et les mollusques. Finalement, la découverte la plus extraordinaire est celle d'un gène codant une cellulose synthase et de plusieurs codant des protéines essentielles pour la synthèse de la paroi cellulaire des plantes. Une fois écartée l'éventualité d'une contamination par de l'ADN d'un possible symbiote de la cione, il a bien fallu accepter la possibilité que non seulement cette ascidie soit capable de digérer la cellulose, mais qu'elle soit également capable d'en synthétiser, ce qu'aucun autre animal n'est capable de faire ! Il s'agit donc d'un cas remarquable de transfert horizontal – que nous avons déjà évoqué – au cours duquel un gène de plante serait passé

chez les urochordés. La question de savoir si cette cellulose synthase est encore fonctionnelle chez la cione reste néanmoins posée.

DES CHORDÉS AUX VERTÉBRÉS

La séquence du premier être vivant vertébré, le poisson *Tetraodon nigroviridis*, a été publiée en 2004. Elle a été réalisée en grande partie au Génoscope d'Évry, sous la direction de Jean Weissenbach. Le tétrodon est un petit poisson originaire du Japon (où il se nomme *Midori fugu*), mesurant une dizaine de centimètres. C'est un vertébré actinoptérygien, un poisson à nageoires rayonnantes, dont les ancêtres ont divergé des sarcoptérygiens (les poissons à nageoires lisses) il y a environ 450 millions d'années (Figure 23). Le génome du tétrodon contient 342 millions de bases, soit 10 % du génome humain, pour un nombre de gènes légèrement supérieur. Il porte très peu d'ADN répété en comparaison de celui des mammifères. C'est donc un génome compact, ce qui a facilité son séquençage. Plusieurs découvertes sur l'évolution des vertébrés ont été faites grâce à l'analyse de son génome. Tout d'abord les gènes codant des protéines ont évolué plus rapidement chez ces poissons que chez les mammifères. Leur horloge moléculaire n'est donc pas la même que celle des mammifères, elle est plus rapide pour une raison encore inconnue. Mais la découverte la plus étonnante de cette étude fut la mise en évidence d'une duplication totale du génome d'un ancêtre du tétrodon. C'est en comparant la **synténie*** – l'ordre des gènes le long des chromosomes – du poisson et de l'homme que les chercheurs ont découvert cette duplication. Si des gènes orthologues, dans deux organismes différents, sont arrangés dans le même ordre le long des chromosomes de deux espèces, on dit que la synténie est conservée pour ces gènes, entre ces deux espèces. On sait depuis longtemps que des espèces proches présentent une meilleure conservation de la synténie que des espèces éloignées. En effet, avec le

Synténie : Ordre local des gènes le long d'un chromosome ou d'un segment chromosomique.

temps, l'ordre de gènes a tendance à se modifier, suite à des cassures et à des réarrangements chromosomiques.

En comparant la conservation de la synténie entre le génome humain et celui du tétrodon, Jean Weissenbach et ses collaborateurs ont découvert qu'à chaque région du génome humain correspondaient deux régions du génome du poisson, localisées sur deux chromosomes différents. Ils en conclurent donc que l'ancêtre du tétrodon avait entièrement dupliqué son génome après la divergence avec l'ancêtre des mammifères (Figure 24). Cette duplication fut suivie de la perte d'un des deux gènes dupliqués, afin de rétablir le bon nombre de gènes dans l'organisme. Des duplications totales de génomes ont été découvertes chez d'autres organismes, certaines espèces ayant même subi plusieurs duplications totales. Ces événements, rares au regard des distances évolutives, mais finalement présents dans de nombreux phylums eucaryotes, sont un des moteurs de l'évolution des espèces vivantes. De même que les événements d'hybridation sur lesquels nous reviendrons plus tard, ils jouent un rôle fondamental dans la complexification du vivant, rôle resté insoupçonné des généticiens jusqu'à très récemment.

La séquence d'un autre vertébré que le tétrodon fut publiée la même année, celle d'un oiseau, le coq bankiva ou coq doré (*Gallus gallus*), originaire du Sud-Est asiatique et dont les principales espèces de poules de basse-cour descendent, depuis sa domestication par l'homme il y a au moins 7 400 ans. Le génome du coq est presque trois fois plus petit que le génome humain (environ un milliard de nucléotides) et contient 20 000-23 000 gènes codants des protéines, un nombre similaire à celui des mammifères. La taille réduite de son génome s'explique en partie par sa grande pauvreté en éléments répétés, en particulier en **pseudogènes*** et en

Pseudogène : Gène inactif suite à l'accumulation de mutations. Un pseudogène dérive souvent d'un gène dupliqué dont l'une des deux copies a par la suite perdu sa fonction.

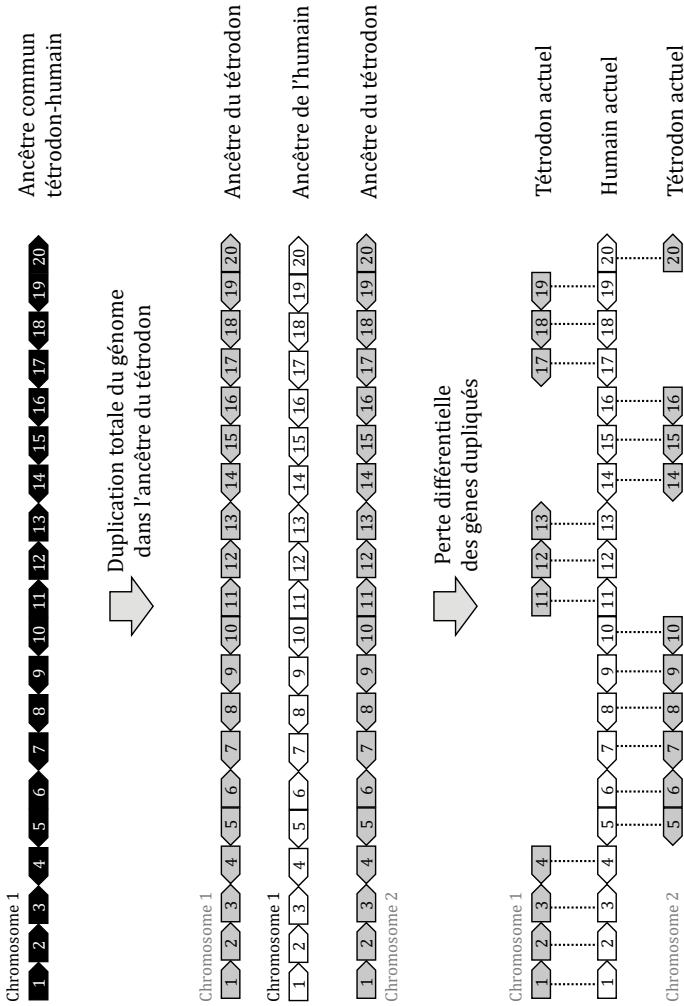


Figure 24 | Duplication totale du génome du tétrédon et comparaison de la synténie avec *Homo sapiens*. Les flèches représentent les gènes, transcrits dans un sens ou dans l'autre. Les gènes humains sont en blanc, les gènes du tétrédon sont en gris. Les gènes de l'ancêtre commun homme-tétrédon sont en noir.

rétrotransposons*, très fréquents chez les mammifères. Seulement 51 gènes ou pseudogènes formés par rétrotransposition ont été identifiés dans ce génome, en comparaison des quelque 15 000 trouvés chez les mammifères.

Rétrotransposon : Séquence particulière d'ADN capable de se dupliquer en se déplaçant de façon autonome dans un génome.

La relativement petite taille de ce génome s'explique également par le faible nombre de duplications de grands segments d'ADN chromosomique en comparaison des primates. Le génome du coq doré est donc plus compact et moins répété que celui des mammifères. Environ 60 % des gènes ont un homologue avec un gène humain, les autres correspondent à de nouveaux gènes, ou à des gènes perdus chez les mammifères et dont certains ont été amplifiés dans la lignée des oiseaux en familles multigéniques. Un exemple remarquable est celui des gènes codant les kératines aviennes. Cette protéine fibreuse est le constituant principal des poils, plumes, cornes, becs et ongles de nombreux animaux. Il existe deux formes de kératine : la forme α présente chez les mammifères et la kératine β présente chez les reptiles et les oiseaux. Les deux protéines ne se ressemblent pas. Le génome de *Gallus gallus* contient pas moins de 150 gènes de kératine β , dont une trentaine sont trouvés répétés à la queue leu leu sur le même chromosome. À l'inverse, certains gènes ont disparu du génome de l'oiseau, comme ceux codant les récepteurs de l'organe voméro-nasal (impliqué dans la détection des **phéromones***), ou de façon peu surprenante les gènes codant les protéines du lait maternel ou ceux codant les protéines de la salive et de l'émail dentaire.

Phéromone : Substance chimique fabriquée et excrétée par la plupart des animaux, servant à transmettre des informations entre les individus et jouant un rôle particulièrement important dans l'attraction sexuelle.

Toutes ces fonctions ayant été perdues dans la lignée des oiseaux, il est normal de ne plus y trouver les gènes correspondants. De façon plus surprenante, une famille de gènes (*OR5U1/OR5BF1*) codant des récepteurs olfactifs est trouvée en aussi grand nombre (238 gènes) chez le coq doré que chez l'homme, allant à l'encontre de la croyance couramment répandue que l'odorat des oiseaux serait peu développé. Finalement, la comparaison des chromosomes aviens avec ceux de l'homme et de la souris a permis de reconstituer le génome putatif de leur ancêtre commun, qui devait posséder 20 paires de chromosomes plus une paire de chromosomes sexuels. Des dizaines de réarrangements successifs au cours des 310 millions d'années d'évolution (HM) ont finalement mené à 19 paires de chromosomes murins, 22 paires de chromosomes humains et 32 paires de chromosomes aviens. Ainsi, le séquençage du génome du coq doré nous a éclairés non seulement sur la présence ou l'absence des gènes spécifiques de cette lignée, mais également sur les spécificités des génomes de mammifères. Comme bien souvent en génomique comparative, en apprendre plus sur une branche évolutive permet également de jeter un éclairage nouveau sur les autres branches du vivant.

Il faut noter que tous les oiseaux actuels sont les descendants de l'archéopteryx, un petit dinosaure volant. Le génome des oiseaux, de même que celui des crocodiliens, ces autres survivants de l'extinction des dinosaures, est donc le plus proche du génome des dinosaures éteints à la fin du crétacé (Figure 23).

DES VERTÉBRÉS AUX MAMMIFÈRES

L'apparition des mammifères remonte à – 220 millions d'années (DFC), au cours du Trias. On distingue les mammifères thériens, c'est-à-dire ceux dont les petits passent au moins une partie de leur développement dans un utérus, des mammifères monotrèmes qui pondent des œufs et dont l'exemple le plus populaire est l'ornithorynque ou *Ornithorhynchus anatinus*. Cet animal étonnant vit exclusivement en Australie orientale et en Tasmanie. Son bec de canard,

ses ergots venimeux et son mode de reproduction en font un proche parent des oiseaux et des reptiles, alors que sa fourrure, sa queue de castor, ses pattes de loutre et la lactation de ses petits sont des traits typiques des mammifères. Les mammifères ont divergé des **sauropsidés*** il y a environ 315 millions d'années HM, alors que les monotrèmes ont divergé des thériens il y a « seulement » 166 millions d'années HM. La séquence du génome de l'ornithorynque a été publiée en 2004 et sa comparaison avec le génome de la poule et les génomes de mammifères a révélé des signatures propres aux deux lignées.

Sauropsidés : Groupe d'animaux regroupant les oiseaux, les reptiles et les dinosaures (Figure 23).

L'ornithorynque possède 18 527 gènes codant des protéines, un nombre similaire aux autres mammifères, répartis le long de 52 chromosomes et couvrant 1,84 milliard de nucléotides. La moitié de ce génome est constituée d'ADN répété, une caractéristique des génomes de mammifères. De nombreux gènes impliqués dans la perception olfactive sont dupliqués, en particulier dans la famille V1R et V2R, qui contiennent chacune respectivement 700 et 950 paralogues dérivant de duplications propres aux monotrèmes. Ce répertoire de gènes olfactifs est plus proche de celui des mammifères que de celui des reptiles. Le génome de l'ornithorynque contient une copie du gène codant la vitellogénine, une protéine formant une partie des réserves énergétiques du jaune d'œuf, qui est absente du génome des mammifères, mais présente en trois copies dans le génome des oiseaux. Le venin de l'ornithorynque contient au moins 19 substances différentes, dont plusieurs protéines β -défensines. Les gènes codant ces protéines ont été dupliqués chez l'animal à partir de gènes dont la fonction était à l'origine différente. La même observation a été faite pour d'autres protéines de la même famille chez les reptiles, pointant là une intéressante convergence évolutive entre les deux lignées cousines. L'ornithorynque présente donc un mélange de certaines signatures propres aux mammifères et d'autres propres aux oiseaux et

aux reptiles, prouvant que la dualité de son histoire évolutive telle que décrite dans sa morphologie et dans sa physiologie est bien inscrite dans son patrimoine génétique.

L'ORIGINE VIRALE DU PLACENTA

Parmi les mammifères thériens, on distingue les marsupiaux (koalas, kangourous) des euthériens (ou placentaires), comme la souris ou l'homme. Chez ces derniers, le développement de l'embryon se fait entièrement dans l'utérus et le placenta y est beaucoup plus élaboré que chez les marsupiaux. Le placenta est l'organe qui connecte physiquement l'embryon à la paroi de l'utérus. Il est responsable de l'apport en nutriment et en oxygène et de l'évacuation des déchets et du dioxyde de carbone. Deux gènes humains sont spécifiquement exprimés dans le placenta, *syncytin-1* et *syncytin-2*. Ces gènes dérivent tous les deux du gène codant la protéine d'enveloppe d'un virus ancestral qui a infecté la lignée des primates il y a 25-40 millions d'années (HM). Le génome de la souris contient deux gènes homologues, *syncytin-A* et *syncytin-B*, qui dérivent également d'une infection virale dans la lignée murine il y a environ 20 millions d'années (HM), mais qui ne sont pas orthologues des gènes humains. L'inactivation de *syncytin-A* chez la souris est associée à un défaut dans l'établissement de l'architecture syncytiale, menant à une déficience de vascularisation aboutissant à terme à la mort de l'embryon. Le placenta a donc été inventé au moins deux fois indépendamment dans les deux lignées de mammifères, par le même mécanisme d'infection virale. Cet exemple remarquable montre comment la capture d'un gène viral dans une lignée peut mener à la création d'une nouvelle fonction essentielle conservée dans tous les descendants de cette lignée.

LA SOURIS, PROCHE PARENT DE L'HOMME

Le séquençage du génome de la souris suivit de peu celui du génome humain et permit des comparaisons entre proches

cousins mammifères placentaires. L'ancêtre commun de la souris *Mus musculus* et de l'homme vivait il y a environ 75 millions d'années (HM), une époque à laquelle les grands dinosaures régnaient sur la Terre (Figure 23). C'était vraisemblablement un petit mammifère, ce qui lui permit d'échapper aux prédateurs dinosauriens et de survivre à leur extinction. L'explosion démographique des mammifères suit de près l'extinction des dinosaures et, si ceux-ci n'avaient pas disparu, il est probable que les mammifères n'auraient pas eu le succès écologique qui fut le leur. Le génome de la souris est 14 % plus petit que le génome humain (2,4 milliards de nucléotides au lieu de 2,9 milliards pour *Homo sapiens*) et contient autant de gènes que le génome humain (environ 22 000) dont 99 % ont un homologue chez l'homme. Les protéines codées par ces gènes se ressemblent à 80 %. L'ensemble des gènes de la souris n'est donc pas très différent de l'ensemble des gènes humains et il n'y a pas eu d'invention évolutive majeure depuis la séparation de ces deux lignées.

Le génome de la souris contient, comme le génome humain, plusieurs milliers de pseudogènes, très fréquents chez les mammifères et qui montrent clairement que lorsqu'un gène se duplique, la pression de sélection sur ce gène se relâche, comme nous le verrons plus en détail dans les chapitres suivants. L'ordre des gènes le long des chromosomes est également très conservé, puisque 90 % des gènes de chaque génome se trouvent dans une région de synténie. Certains chromosomes sont très bien conservés, par exemple, le chromosome 20 humain qui se retrouve entièrement sur un bras du chromosome 2 de la souris ou le chromosome 17 humain représentant une large fraction du chromosome 11 murin, mais d'autres chromosomes montrent un plus grand nombre de réarrangements. On trouve 217 blocs conservés de synténie entre les deux génomes, correspondant donc à autant de réarrangements chromosomiques entre l'homme et la souris. Parmi les découvertes les plus intrigantes, les grandes duplications de segments chromosomiques retrouvées chez l'homme (voir chapitre 15) sont apparemment absentes du

génomé murin. En revanche, celui-ci accumule les mutations neutres ainsi que les petites **délétions*** deux fois plus rapidement en moyenne que le génome humain. Cette plus grande fréquence de délétions peut expliquer – au moins en partie – la plus petite taille du génome de la souris.

Délétion : Mutation qui enlève des nucléotides à une séquence d'ADN. Les délétions peuvent être de faible taille (de deux à quelques nucléotides) ou bien représenter plusieurs milliers de nucléotides.

À RETENIR

- Les fossiles sont les restes minéralisés d'anciens animaux ou plantes et permettent de dater la présence de ces espèces dans les couches géologiques.
- Les horloges moléculaires, basées sur la comparaison des séquences d'ADN, permettent de dater l'apparition d'une nouvelle espèce par rapport à une autre déjà caractérisée dans le même phylum.
- L'apparition de la symétrie bilatérale a permis d'optimiser l'orientation dans l'espace des métazoaires et le développement de la bouche et de l'anus aux deux extrémités de l'organisme.
- L'ancêtre du tétrodon a subi une duplication totale de son génome, mise en évidence par rapport à la comparaison de la synténie avec le génome humain.
- Les gènes permettant le développement du placenta chez les mammifères euthériens sont d'origine virale.

BIBLIOGRAPHIE

- L'explosion cambrienne : Erwin, Laflamme, Tweedt, Sperling *et al.*, *Science* 334: 1091-1097 (2011).
- La symétrie bilatérale : Finnerty, Pang, Burton, Paulson and Martindale, *Science* 304: 1335-1337 (2004).

Le génome du tétrodon : Jaillon, Aury, Brunet, Petit *et al.*, *Nature* 431: 946-957 (2004).

L'origine virale du placenta : Dupressoir, Vernochet, Bawa, Harper *et al.*, *Proceedings of the National Academy of Sciences USA* 106: 12127-12132 (2009).

10

Des algues aux arbres

Alors que les premiers animaux unicellulaires complexifiaient leurs génomes et développaient de nouvelles régulations génétiques qui les mèneraient à la multicellularité, les premières cellules végétales naissaient dans le milieu marin, de la symbiose entre une cellule eucaryote et une cyanobactérie (voir Chapitre 7). Ces protoalgues allaient encore parcourir un long chemin évolutif avant de sortir des océans et de se répandre à la surface de la terre, d'abord comme des mousses, puis des arbustes, pour former enfin les premières forêts il y a environ 400 millions d'années.

LES EMBRYOPHYTES, PREMIÈRES PLANTES TERRESTRES

Les premières cellules eucaryotes à avoir phagocyté une cyanobactérie ancestrale ont donné naissance aux premières algues unicellulaires, comme vu précédemment. L'analyse des séquences d'ADN chloroplastique de plusieurs espèces végétales modernes montre une seule origine commune à toutes les plantes modernes : glaucophytes, algues rouges (ou rhodophytes), algues vertes et embryophytes, ces

dernières regroupant toutes les plantes terrestres. Les comparaisons de séquences entre différents phylums de plantes et de champignons montrent que la divergence des algues vertes et des plantes terrestres a eu lieu il y a sans doute 1 milliard d'années HM. Cette datation doit néanmoins être prise avec prudence car en l'absence de fossiles de plantes ou de champignons datant de cette époque, c'est l'horloge moléculaire des vertébrés qui a été utilisée pour cette étude. En utilisant la même méthode de calcul, il a été estimé que la divergence entre les animaux, les plantes et les champignons datait d'environ 1,5-1,6 milliard d'années, c'est-à-dire 400-500 millions d'années après l'apparition des premières cellules eucaryotes.

Il est vraisemblable que les premiers eucaryotes terrestres soient le résultat d'une association symbiotique entre un champignon et un eucaryote photosynthétique. Cette hypothèse se base sur la découverte de symbiotes associés aux plus anciens fossiles connus de champignons, vieux de 460 millions d'années. À l'heure actuelle, on trouve encore de telles associations dans les lichens, formes symbiotiques d'un champignon filamenteux et d'une algue (ou d'une cyanobactérie). Puis apparurent les mousses, qui sont les plus ancestrales des plantes terrestres. Leur structure est simple, elles ne possèdent pas de racines, seulement des rhizoïdes qui leur permettent de s'accrocher à leur substrat. Les feuilles contiennent un nombre limité de couches cellulaires et elles ne possèdent aucun des tissus conducteurs acheminant la sève que l'on trouve chez les plantes vasculaires. On date la divergence des mousses et des plantes vasculaires à environ 700 millions d'années, c'est-à-dire pendant la période précambrienne à laquelle s'est produite l'explosion de nouvelles formes de vie chez les métazoaires.

Le séquençage du génome d'une mousse, *Physcomitrella patens*, a révélé des propriétés surprenantes chez ce représentant des plantes terrestres ancestrales. Tout d'abord, son génome est hautement répétitif, plus de la moitié étant composé d'éléments répétés de type rétrotransposons. L'horloge moléculaire montre que ces éléments capables de se dupliquer d'un endroit à l'autre du génome, à la manière de

certain virus, montrent deux pics d'activité relativement récents, le premier il y a 4-6 millions d'années, le second encore plus récemment il y a seulement un million d'années. L'étude des gènes paralogues suggère que le génome de *P. patens* a subi deux évènements de duplication totale de son génome, le premier il y a 40-48 millions d'années et le second il y a 27-35 millions d'années. Le génome actuel de cette mousse a donc considérablement évolué depuis que son ancêtre a vu le jour il y a environ 700 millions d'années. Ce qui nous remémore que les génomes des espèces que nous séquençons à l'heure actuelle, y compris le nôtre, ne sont que les lointains descendants des génomes des premiers individus de l'espèce. Finalement, de façon remarquable, plusieurs centaines de gènes de *P. patens* sont présents dans des régions de synténie avec d'autres plantes terrestres ayant divergé il y a au moins 500 millions d'années. Ces gènes sont impliqués dans l'organisation spécifique cellulaire et tissulaire de ces plantes et montrent un niveau élevé de co-régulation, suggérant que cette organisation est importante pour le bon développement de toutes ces plantes.

LES TRACHÉOPHYTES ET LES PREMIERS ARBRES

Le développement des plantes vasculaires, ou trachéophytes, fut dépendante de deux innovations biologiques essentielles : la lignine et la vascularisation. La lignine est un polymère entrant dans la composition du bois, de même que la cellulose. Elle apporte à la plante rigidité et imperméabilité et permet aux plantes de s'élever du sol en résistant à la gravité. Avant la lignine, le monde végétal terrestre était contraint de rester au ras du sol. Après la lignine, les premières fougères et les premiers arbustes purent voir le jour. La deuxième grande innovation fut la mise en place d'un système vasculaire de transport de la sève par les vaisseaux du **phloème*** et du

Phloème : Les vaisseaux du phloème contiennent la sève élaborée, constituée d'eau et des sucres fabriqués au cours de la photosynthèse dans la partie supérieure de la plante.

xylème*. Celle-ci fut essentielle dans le développement de plantes de taille croissante en transportant les nutriments sous forme de sève élaborée et de sève brute, cette dernière circulant dans les trachéides du xylème qui ont donné leur nom aux trachéophytes.

Xylème : Le xylème transporte la sève brute (eau et molécules organiques) en provenance des racines vers les feuilles, les fleurs et les fruits. Les vaisseaux du xylème sont formés de trachéides, successions de cellules mortes, qui donnent leur nom aux trachéophytes.

La photosynthèse étant essentielle à la survie et à la pousse des plantes vasculaires, la compétition pour la lumière explique sans doute la taille croissante des arbres à partir de cette époque. Les premières forêts étaient présentes sur Terre il y a environ 400 millions d'années et la production de grandes quantités de matières organiques sous forme de bois lignifié permit la sélection de champignons capable de dégrader et de recycler cette matière organique, résultant en une association étroite entre végétaux à lignine et champignons, encore présente dans nos forêts modernes. La formation des premières racines a dû fortement contribuer à l'érosion des sols et à leur aération mécanique, permettant une plus grande absorption du dioxyde de carbone atmosphérique (CO₂), entraînant sa diminution et réduisant sensiblement la température à la surface de la Terre.

LES ANGIOSPERMES OU PLANTES À FLEURS

Les plantes à fleurs sont les plantes les plus répandues à la surface de la Terre à l'heure actuelle. Elles représentent un élément central du cycle du carbone, de l'oxygène et de l'eau et sont donc essentielles à la stabilité de la biosphère et du climat. Elles fournissent aux sociétés humaines de quoi se nourrir et se vêtir, ainsi que des matériaux de construction et certaines molécules médicamenteuses. Les angiospermes se répartissent en deux grands groupes :

les **monocotylédones*** et les **eudicotylédones***. Les angiospermes représentent plus de 350 000 espèces de plantes, dont 75 % sont des eudicotylédones.

Monocotylédone : Plante dont l'embryon possède un seul cotylédon ou feuille primordiale.

Eudicotylédone : Plante dont l'embryon possède deux cotylédons.

Parmi les monocotylédones, on trouve essentiellement des graminées alimentaires comme le riz, le sorgho, le blé, l'orge, le maïs ou le seigle, mais également l'ananas, la banane et le palmier à huile. Les eudicotylédones comprennent des espèces aussi diversifiées que le raisin et la fraise, des arbres fruitiers comme le pommier, le poirier et le pêcher, ou bien le cacao et le café, ou encore le soja et l'eucalyptus. Une étude récente publiée conjointement par deux équipes de l'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) a permis de reconstruire le génome de l'ancêtre commun des eudicotylédones, des monocotylédones et des angiospermes. Les génomes de 34 espèces d'angiospermes (13 monocotylédones, 20 dicotylédones et 1 angiosperme ancestral, *Amborella trichopoda*) ont été comparés entre eux et avec celui de l'épicéa, d'une mousse et de *Chlamydomonas reinhardtii*, l'algue verte dont nous avons déjà parlé. Ce travail remarquable a permis de découvrir que l'ancêtre commun de toutes les angiospermes possédait 15 chromosomes et environ 23 000 gènes, un nombre similaire à celui de l'espèce humaine, dont 10 263 sont spécifiques des plantes à fleurs et correspondent à des processus biologiques propres à cette lignée, tels que la pollinisation par interaction entre pollen et pistil ou le développement de la fleur. Plusieurs événements de duplication ou de tripllication totale de génomes se sont produits dans les deux lignées d'angiospermes et ont été suivis par des pertes massives de gènes et des fusions de chromosomes, conservant un nombre stable de gènes et de chromosomes tout au long de leur évolution. Nous

verrons dans un prochain chapitre que ces duplications de génomes suivies de pertes de gènes, qui remodelent rapidement les génomes et que nous avons déjà évoquées, sont fréquentes dans le monde eucaryote, en particulier chez les plantes. De telles duplications se sont produites à de nombreuses reprises dans le monde végétal, à l'origine des angiospermes, chez toutes les graminées et dans de nombreuses branches des eudicotylédones. La banane, par exemple, a subi trois duplications totales de son génome en plus de celle commune aux angiospermes, le soja en a subi deux, ainsi que le cacao, la papaye, le chou et l'arabette (Figure 25).

Le génome du riz et celui du raisin ont été identifiés comme ceux évoluant le plus lentement et ont donc été proposés comme génomes de référence pour les monocotylédones et les eudicotylédones, respectivement. Cette étude permet enfin d'inférer plus précisément l'âge des différentes lignées étudiées. L'ensemble des travaux précédents suggéraient que les angiospermes étaient apparues il y a 120 à 170 millions d'années HM, à la transition Jurassique-Crétacé, ou bien il y a 130 millions d'années DFC, en pleine période Crétacé. L'étude des deux laboratoires de l'INRAE permet de reculer l'âge d'apparition des angiospermes à 214 millions d'années HM, à la fin du Trias, prouvant leur émergence bien avant l'âge de leur dernier fossile connu, comme il a été montré pour les métazoaires. Les eudicotylédones ont quant à eux divergé des monocotylédones il y a 99 à 150 millions d'années si l'on utilise une horloge moléculaire lente, ou il y a 167 à 250 millions d'années en utilisant une horloge plus rapide. On voit donc qu'à l'époque où émergent les premières angiospermes, les vertébrés marchaient déjà sur terre depuis plus de 200 millions d'années et il faudra attendre la dernière centaine de millions d'années qui nous sépare du temps présent pour que les plantes à fleurs envahissent toute la planète et se diversifient dans la flore que nous connaissons bien et qui nous nourrit, nous habille et nous protège des intempéries depuis des millénaires.

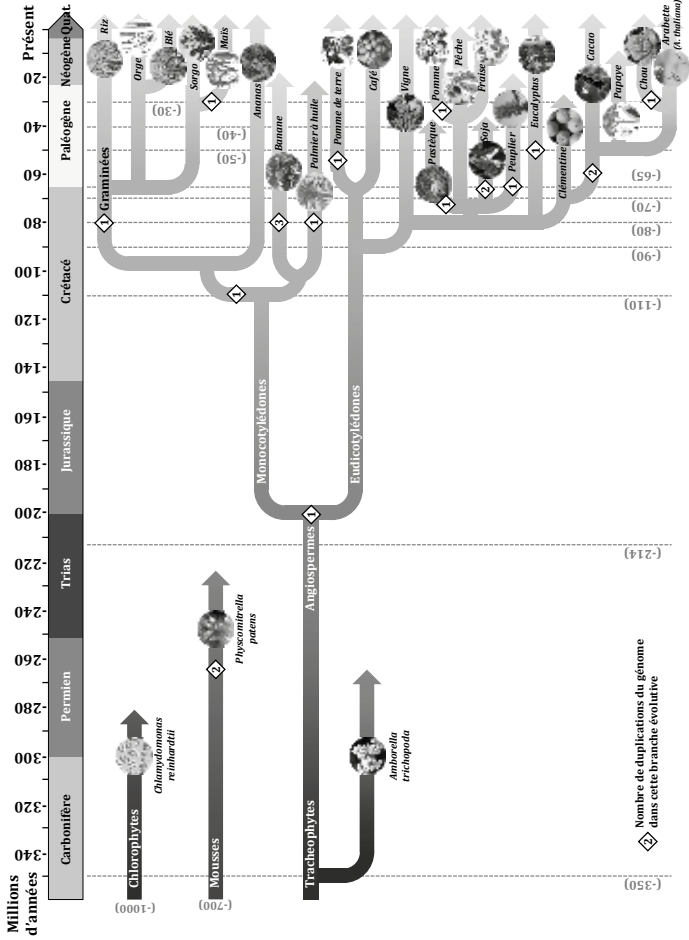


Figure 25 | Arbre des angiospermes. Les dates de divergence des chlorophytes et des mousses par rapport aux ancêtres des trachéophytes sont indiquées à la racine de leurs flèches respectives (- 1000 et - 700). Les traits pointillés gris verticaux correspondent aux dates de divergence des différentes lignées.

LA BIODIVERSITÉ DES FORÊTS

Les premières forêts de trachéophytes sont apparues il y a environ 400 millions d'années, peu de temps après les premiers vertébrés. Elles sont un élément essentiel de l'écosystème terrestre, nous procurant bois, nourriture, mais fournissant également une part non négligeable de l'oxygène atmosphérique par la photosynthèse. Par le même processus, les arbres absorbent le dioxyde de carbone, réduisant l'effet de serre et régulent donc en partie la température de notre planète. La déforestation intensive est à l'origine de l'érosion des sols et des effondrements de terrain qu'elle occasionne avec les dégâts considérables que l'on connaît en matériels et en vies humaines. Connaître et protéger la biodiversité des forêts est donc essentiel à la survie de l'humanité, mais également à celle de toutes les autres formes de vie terrestre. Une étude récente a tenté de déterminer le nombre exact d'espèces d'arbres sur terre en utilisant une base de données répertoriant 38 millions d'arbres. Leur estimation indique qu'il existe environ 73 300 espèces d'arbres différentes, dont plus de 9 000 restent à découvrir. De façon attendue, ce sont les forêts tropicales qui contiennent le plus d'espèces différentes, environ le tiers de toutes les espèces connues. C'est en Amérique du Sud que l'on trouve le plus d'espèces rares (8 200), puis en Eurasie (6 100) et en Afrique (3 900). La plupart des forêts sont dominées par un très petit nombre d'espèces, mais contiennent 30-40 % d'espèces rares, sur tous les continents. Finalement, la comparaison des espèces présentes sur les cinq continents (Afrique, Océanie, Eurasie, Amériques du Sud et du Nord) montre que moins de 0,1 % des espèces leur sont communes. La plupart des espèces sont donc spécifiques d'un continent, les deux continents partageant le plus d'espèces étant les deux continents américains, physiquement reliés depuis des millions d'années. Il y a donc encore beaucoup à découvrir avant d'obtenir la représentation la plus fidèle possible de la biodiversité des arbres de nos forêts, sur tous les continents.

LE GÉNOME DE L'ARABETTE, ANGIOSPERME MODÈLE

La séquence du génome d'une plante modèle, *Arabidopsis thaliana*, utilisée depuis des décennies par les généticiens, a permis d'étudier les particularités génétiques d'une angiosperme modèle. Le génome de cette arabette contient seulement cinq chromosomes couvrant 125 millions de nucléotides, 25 fois moins que les génomes de primates. La séquence prédit 25 000 gènes regroupés en 11 000 familles différentes. Environ 30 % de ces gènes ne montrent aucune homologie avec des gènes de champignons ou d'animaux et semblent être spécifiques des plantes. Parmi ceux-ci, 80-90 % des protéines impliquées dans la régulation de la transcription semblent avoir été inventées dans la lignée végétale. Les protéines jouant un rôle dans le transport de sels ou de petites molécules à travers la membrane sont également différentes de celles des animaux. Certaines protéines du cytosquelette animal sont absentes du génome de la plante et sont remplacées par d'autres, en cohérence avec les particularités d'organisation et de composition des cellules végétales. Certains gènes de développement ne ressemblent à aucun gène animal ou fongique, mais viennent des cyanobactéries et sont donc le résultat de transferts entre un génome cyanobactérien ancestral et le génome de l'hôte *protoeucaryote* lors de l'endosymbiose qui a mené à la lignée végétale (voir Chapitre 7). Les gènes impliqués dans l'apoptose, c'est-à-dire la mort cellulaire programmée qui joue un rôle de régulation de la croissance cellulaire chez les animaux et dont le dérèglement est la cause de cancers, sont absents du génome de l'arabette. De toute évidence, les plantes ne possédant pas de structure morphologique délimitée et finie, n'ont pas besoin de réguler leur croissance cellulaire, elles peuvent croître indéfiniment, ou presque. Finalement, le génome de l'arabette contient 24 duplications de larges segments d'ADN chromosomique, semblables à celles observées chez les primates. Ces **duplications segmentaires*** sont une des raisons du niveau élevé de redondance génique observé chez cette plante. Nous verrons dans un prochain chapitre que les duplications partielles de génomes sont un

trait caractéristique des génomes eucaryotes et particulièrement des génomes de plantes.

La compréhension fine de l'évolution de la flore terrestre se heurte au même problème que la compréhension de l'évolution des animaux. Les plus anciens fossiles d'embryophytes sont vieux de 476 millions d'années, alors que la phylogénie moléculaire suggère que les embryophytes ont divergé des algues vertes il y a environ 1 milliard d'années et que les trachéophytes ont divergé des mousses il y a environ 700 millions d'années, c'est-à-dire plusieurs centaines de millions d'années avant l'âge de leurs premiers fossiles. Ces derniers sont largement biaisés vers les plantes vasculaires et attestent de leur présence pendant la période Silurienne, époque à laquelle les premiers animaux amphibiens sortent des eaux. Ces fossiles sont tous localisés dans les mêmes régions géographiques, ce qui ne donne hélas qu'un faible aperçu de la diversité présente à l'origine des embryophytes. Si les datations moléculaires sont exactes, cela signifie que des centaines de millions d'années de l'évolution des plantes terrestres nous sont encore inconnues. Pour des raisons anthropocentriques, l'homme a consacré beaucoup de temps et d'énergie à séquencer des génomes animaux et des génomes d'hommes modernes et ancestraux. Pour savoir où l'on va, il faut comprendre d'où l'on vient. Cette approche a permis d'appréhender avec beaucoup plus de pertinence qu'auparavant l'origine de l'humanité. Il faut espérer que l'étude de l'évolution des plantes, par le séquençage systématique de génomes d'espèces modernes dérivant d'espèces ancestrales, permette prochainement de pouvoir enfin approcher à un niveau de résolution similaire les grands mécanismes qui ont modelé l'évolution de nos très lointains cousins les végétaux.

LE VIVANT S'EST COMPLEXIFIÉ PAR BONDS SUCCESSIFS

Arrivés à ce stade de notre voyage dans le passé, grâce aux séquences des génomes, des êtres vivants qui nous ont précédés, on peut essayer de résumer de façon synthétique les principales innovations du vivant

depuis la formation de la Terre il y a environ 4,6 milliards d'années (Figure 26). À cette époque, la terre était le lieu d'un volcanisme très actif et deux séries de bombardements météoritiques ensemencèrent notre planète des toutes premières molécules organiques qui allaient constituer les briques originelles du vivant, sucres, acides aminés et bases. L'atmosphère ne contient pas encore d'oxygène, mais est composée d'hydrogène, de gaz carbonique, de méthane, d'ammoniaque et de vapeur d'eau, car il fait beaucoup trop chaud pour que l'eau soit sous forme liquide. Sous l'effet de son lent refroidissement, la vapeur d'eau finit par se condenser et tomber sous forme de déluge qui, emportant une partie du dioxyde de carbone atmosphérique, va contribuer à l'accélération du refroidissement en diminuant l'effet de serre. Il s'écoule seulement 200 millions d'années avant que la température ne permette les réactions chimiques qui vont créer le Monde à ARN. Mais la vie n'existe pas encore, seule des molécules d'ARN, des ribozymes dotés d'activité catalytique se répliquent, embryonnaires prémices de génomes en devenir.

Il est remarquable qu'il n'ait fallu qu'environ 600 millions d'années pour sortir du Monde à ARN et voir apparaître les premières bactéries. C'est à cette époque que se fixe le code génétique et que les protéines commencent à remplacer les ARN en tant que catalyseurs de réactions biochimiques. Puis, l'ADN plus stable est inventé et remplace l'ARN comme support de l'hérédité. Ce dernier se trouve alors cantonné à un rôle de transfert de l'information génétique. Il deviendra information structurelle, régulateur transcriptionnel ou messager qui sera traduit en protéine. Le vivant marque alors une pause, comme s'il était satisfait de toutes ces nouveautés et cherchait une nouvelle inspiration. Il faudra attendre plus d'un milliard d'années jusqu'à l'invention de la photosynthèse qui conduira à la Grande Oxydation, et permettra 300 millions d'années plus tard l'apparition des premiers organismes eucaryotes, grands consommateurs d'oxygène. L'atmosphère de la terre a considérablement évolué. Elle est maintenant composée principalement d'azote

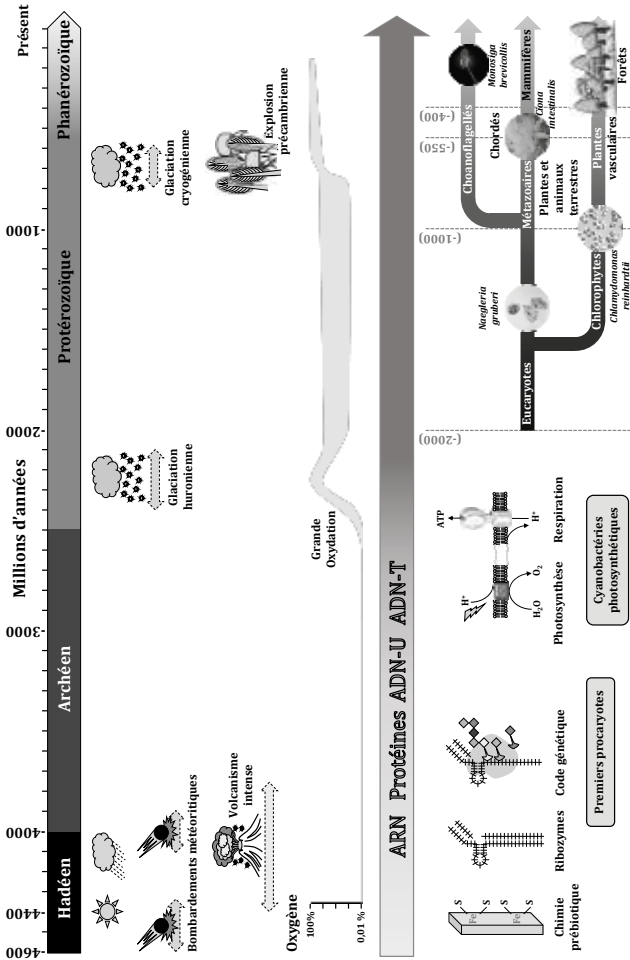


Figure 26 | L'histoire du vivant depuis la formation de la Terre. Haut : le soleil représente le moment où notre étoile s'est allumée. Le nuage correspond à la condensation de la vapeur d'eau sous forme de pluies. Les périodes des bombardements météoritiques primitif et tardif sont indiquées, de même que la période de volcanisme actif et deux périodes de glaciation. Milieu : La fourchette des niveaux d'oxygène atmosphériques est représentée, en quantités relatives par rapport à la période actuelle, fixée comme référence à 100 %. Bas : les périodes présumées d'apparition des différentes molécules du vivant et les avancées majeures du vivant sont indiquées.

et d'oxygène, avec un peu de vapeur d'eau et de gaz carbonique. L'ozone est en cours de formation sous l'influence du rayonnement ultra-violet. Notre planète ressemble vraisemblablement à une Mars recouverte d'océans dans lesquels vivent des bactéries et des archées. Sous l'influence des vents, des courants et de la pluie, certaines d'entre elles ont peut-être gagné la terre ferme et ont colonisé des environnements aquatiques non marins, lacs et rivières. D'autres ont trouvé un habitat au fond des sources chaudes océaniques où elles ont prospéré.

Puis vient une période entre 1,8 et 0,8 milliard d'années avant le temps présent, que les paléontologues nomment « le milliard ennuyeux », pendant laquelle le taux d'oxygène redescend à environ 1 % de sa valeur actuelle. Les formes de vie animales et végétales évoluent lentement vers des formes multicellulaires, métazoaires pour les unes, mousses pour les autres, avant l'explosion précambrienne caractérisée par une débauche d'innovations biologiques qui mèneront aux lignées dont les êtres vivants actuels descendent tous. Les êtres vivants vont alors sortir du milieu marin et gagner la terre ferme. Les premières fougères et les premiers arbustes, puis les premiers vertébrés terrestres, vont peupler le monde. Celui-ci change de couleur et le vert apparaît dans le paysage. Le taux d'oxygène remonte, entraînant une explosion des formes de vie eucaryotes qui se nourrissent de celui-ci. Nous sommes 300 millions d'années avant le temps présent, le décor est planté pour que les mammifères entrent en scène. Nous verrons plus loin que sans la chute d'une certaine météorite ils pourraient bien ne jamais avoir évolué vers d'autres formes animales plus complexes que celles des petits rongeurs qu'ils étaient à cette époque reculée.

La vie a en apparence progressé par « bonds » successifs, de lentes périodes de stagnation étant suivies de formidables inventions. Le séquençage récent de nombreux organismes eucaryotes, lointains descendants d'organismes primitifs, montre bien comment l'étude des génomes permet de répondre à de grandes questions

sur l'évolution des organismes vivants. Cette branche extrêmement puissante de la génétique, que l'on appelle génomique comparative, a un impact sans précédent sur la génétique moderne, permettant depuis environ deux décennies d'affiner notre compréhension des mécanismes fondamentaux ayant mené des premières cellules aux animaux et aux plantes. Il faut néanmoins garder à l'esprit que ce qui est apparent peut cacher une réalité plus complexe, car notre compréhension de ces époques reculées est encore bien imparfaite. Les organismes vivants modernes dont nous décryptons aujourd'hui la séquence génomique ne sont que les lointains descendants des organismes ancestraux que nous essayons de décrire. Des millions de générations les séparent de leurs ancêtres et des changements conséquents de leur génome ont pu se produire depuis cette époque reculée, pouvant fausser – au moins partiellement – certaines de nos interprétations modernes.

À RETENIR

- Les premiers eucaryotes terrestres étaient vraisemblablement des lichens, symbiotes entre un champignon et une algue.
- La lignine et les vaisseaux du phloème et du xylème permettent aux plantes de s'élever du sol en devenant des fougères, des arbustes et des arbres.
- les angiospermes ou plantes à fleurs sont apparus il y a seulement 214 millions d'années, mais regroupent maintenant la grande majorité des plantes terrestres.
- Tous les angiospermes ont subi une ou plusieurs duplications de leur génome au cours de leur évolution.

BIBLIOGRAPHIE

L'origine des plantes terrestres : Tisserant, Malbreil, Kuo, Kohler *et al.*, *Proceedings of the National Academy of Sciences USA* 110: 20117-20122 (2013).

L'origine des angiospermes : Murat, Armero, Pont, Klopp et Salse,
Nature Genetics 49 : 490 (2017).

Les différentes espèces d'arbres sur terre : Gatti, Reich, Gamarra,
Crowther *et al.*, *Proceedings of the National Academy of Sciences*
USA 119: e2115329119 (2022).

11

Un animal doué de raison

Dans son roman au titre éponyme paru en 1967, l'écrivain français Robert Merle (1908-2004) pose la question de la condition humaine et de ce qui sépare l'homme de l'animal, en particulier d'un autre mammifère très intelligent, le dauphin. Le roman se déroule à l'époque de la guerre froide, durant laquelle une équipe de scientifiques américains apprend l'anglais à des dauphins et arrive à les faire communiquer dans cette langue. Seize ans plus tard, l'écrivain américain David Brin publie le second roman de science-fiction de son cycle *Élévation : Marée stellaire* (*Startide Rising*). Dans un futur peut-être pas si lointain, l'humanité a appris à « élever » des espèces animales à une intelligence supérieure. Après les chimpanzés, c'est le tour des dauphins d'être transformés par la génétique du futur en néo-dauphins, dotés de capacités d'abstraction égales à celles des hommes ainsi que de la parole. On pourrait également citer *La Ferme des animaux* de George Orwell (1903-1950), dans lequel les animaux devenus intelligents prennent le pouvoir et chassent les hommes de la ferme où ils vivent, ou bien encore *La Planète de singes* publié

en 1963 par l'écrivain français Pierre Boule (1912-1994), best-seller de science-fiction dans lequel les chimpanzés et les gorilles du futur devenus suffisamment intelligents ont réduit les humains en esclavage. Les exemples littéraires dans lesquels les animaux s'élèvent jusqu'à atteindre l'intelligence humaine ne manquent pas. On est alors en droit de se demander quelle différence biologique fondamentale différencie l'homme de l'animal, d'où vient à l'homme cette formidable capacité d'abstraction pour laquelle les animaux sont – *a priori* – moins compétents ? La réponse est évidente pour tout collégien un peu dégourdi : c'est bien le développement exceptionnel de son cerveau qui donne à l'homme la puissance de sa pensée abstraite et la profondeur de sa réflexion cognitive.

UNE MACHINE À FABRIQUER DES IMAGES MENTALES

L'homme appartient à l'ordre des primates, dont le nom donné par le naturaliste suédois Carl von Linné – du latin *primas*, « qui occupe la première place » – suggère la prééminence dans l'échelle évolutive. Nous partageons cet ordre avec nos proches cousins, les grands singes (chimpanzés, bonobos, gorilles, orangs-outans), mais également avec les petits singes et les lémuriens. Qu'en est-il donc du cerveau de nos cousins primates ? Historiquement, plusieurs auteurs comparèrent le poids net des encéphales de plusieurs animaux pour arriver à la conclusion que celui de l'homme n'était ni le plus gros, car nous sommes largement battus par les cétacés et les grands mammifères comme l'éléphant, ni le plus important rapporté au poids du corps, car nous sommes battus par de petits mammifères comme les furets. Le poids de l'encéphale n'évolue donc pas de façon linéaire en fonction du poids ou du volume des espèces concernées. En revanche, il existe une corrélation assez grossière entre le poids du cerveau et la surface du corps de l'espèce considérée, suggérant une relation entre la taille de l'encéphale et sa relation au monde extérieur à travers ses multiples activités sensorielles. Le poids du cerveau humain se trouve quant à lui légèrement au-dessus de cette droite de corrélation. Mais

quelle partie du cerveau est plus spécifiquement impliquée dans cet avantage intellectuel qui a permis à l'homme d'occuper tous les continents, sous toutes les latitudes ?

Le plan d'ensemble de l'encéphale est conservé des poissons jusqu'à nous. Il comprend vers l'avant les deux hémisphères cérébraux avec le bulbe olfactif et les ganglions de la base qui contrôlent le mouvement, puis en arrière le thalamus qui sert de relais aux informations sensorielles et l'hypothalamus qui régule les sécrétions hormonales du système endocrinien par l'intermédiaire de l'hypophyse. En arrière se trouve le cerveau moyen, épaissi localement pour former le « toit » optique, servant à la vision, mais également chez les mammifères à l'audition. Enfin, le cerveau postérieur dont la face dorsale se différencie en cervelet, contribue à la coordination, à la synchronisation et à la précision des gestes. Les faces ventrale et latérale donnent le bulbe rachidien qui connecte le cerveau à la moelle épinière et qui contrôle le rythme cardiaque, la respiration et la pression artérielle. Cette architecture est commune à tous les vertébrés (Figure 27). Certaines régions du cerveau vont se développer plus ou moins dans certaines branches du vivant. Par exemple, la fonction olfactive très importante pour les poissons a entraîné un développement significatif de la région antérieure de leur cerveau. Chez les primates en général et chez l'homme en particulier, c'est la partie dorsale des hémisphères cérébraux, appelée néocortex, qui a subi une incroyable expansion au point d'occuper quasiment tout l'espace crânien. Le rapport de taille entre le néocortex et la masse globale de l'encéphale est trois fois supérieur chez l'homme à celui du chimpanzé et 150 fois supérieur à celui d'un petit mammifère insectivore comme le hérisson. Le volume de la boîte crânienne étant limité par des contraintes physiques évidentes, le néocortex a augmenté sa surface en se pliant et se repliant en d'innombrables circonvolutions caractéristiques de notre cerveau. C'est le néocortex qui fabrique les images mentales, c'est le centre de l'abstraction du cerveau et de sa puissance cognitive, le centre de l'imagination qui fournit à l'homme à la fois un

formidable outil créatif et la possibilité de s’imaginer lui-même dans des situations alternatives, ouvrant la voie à d’infinis fantasmes et aux créations les plus audacieuses.

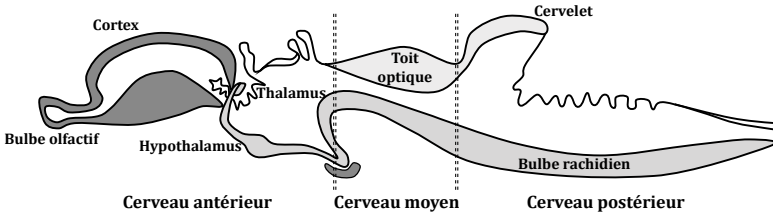


Figure 27 | Plan d’ensemble du cerveau commun à tous les vertébrés.

LE CERVEAU RÉCOMPENSE OU PUNIT

En dehors du développement exceptionnel du néocortex, le cerveau humain n’est pas fondamentalement différent de celui de ses cousins les grands singes, en particulier pour tout ce qui concerne les réactions comportementales à un stimulus externe. Tous les mammifères possèdent des régions du cerveau associées à ce que l’on nomme système de récompense (ou système de renforcement). Les neurones appartenant à ce système déclenchent la sécrétion de divers **neurotransmetteurs***, tels que la dopamine et des opioïdes et cannabinoïdes endogènes, qui vont renforcer la motivation de l’individu à rechercher à nouveau la situation ou le comportement qui a provoqué cette sécrétion (contact affectif ou sexuel, absorption de nourriture, etc.). En l’absence de la possibilité d’obtenir une récompense, on obtient une punition (ou renforcement négatif), telle que la douleur ou la peur. C’est ce système de renforcement qui génère les célèbres réflexes conditionnels étudiés par Ivan Pavlov (1849-1936) sur les chiens et qui lui vaudront le prix Nobel de Physiologie ou Médecine en 1904. Le comportement social de tous

Neurotransmetteur : Composé chimique libéré par les neurones, ayant une action sur d’autres neurones.

les mammifères repose donc sur le désir d'activer (ou de réactiver si la situation gratifiante s'est déjà produite) le circuit de récompense, aussi souvent que possible. C'est une caractéristique essentielle à la survie de l'individu (absorption de nourriture ou apprentissage du danger) ou de l'espèce (reproduction).

Les principales réactions à l'impossibilité de réaliser une action gratifiante ont été étudiées en détail par un célèbre neurobiologiste français, Henri Laborit (1914-1995), qui a obtenu en 1957 le prix Lasker pour ses travaux sur le comportement humain. Il existe plusieurs réponses neurologiques à l'impossibilité d'activer le système de renforcement. L'agressivité est la réaction la plus simple, la plus fondamentale pour échapper à la punition. Chez le chimpanzé, ce comportement se caractérise toujours par l'agression d'un dominant sur un dominé. Une hiérarchie sociale est donc mise en place très tôt afin de résoudre les conflits résultant de l'inhibition du système de récompense. Dans les sociétés humaines, l'agressivité peut mener au conflit et à la guerre. Mais Laborit insiste sur le fait que l'agressivité humaine est aussi un moyen de résoudre l'angoisse créée par la structuration hiérarchique de la société et ses multiples interdits et tabous, qui finissent par inhiber les comportements gratifiants. Il faut noter que nos autres cousins, les bonobos, présentent des comportements moins agressifs que les chimpanzés (ou que l'homme), ce qui suggère que l'inhibition de leur système de récompense n'active pas exactement la même réponse.

Chez tous les mammifères, la notion de territoire est importante et l'empiétement sur le territoire d'un autre individu (ou groupe d'individus) active la plupart du temps une réponse agressive. Chez l'homme, le territoire s'étend à la notion de propriété. Ce comportement, appris depuis l'enfance, provoque une dépendance, une accoutumance qui entraîne une gratification par l'objet possédé (maison, voiture, bijoux, personne aimée) et une réaction agressive lorsqu'une tierce personne essaie de nous en déposséder.

La deuxième réponse possible déclenchée par l'impossibilité d'obtenir une récompense est la fuite. Quand deux mammifères entrent

en compétition pour le même territoire ou la même femelle, la première réponse est souvent agressive, puis l'un des deux animaux prend le dessus et l'autre prend la fuite. Si la fuite n'est pas possible (par exemple si les deux animaux sont en cage), la confrontation ne peut se finir que par la mort ou par la soumission du vaincu. Dans nos sociétés humaines modernes, la fuite n'est pas toujours possible. Les confrontations dans le milieu professionnel ou dans la cellule familiale ne peuvent la plupart du temps pas être évitées. Quand la fuite devient impossible, des rapports de dominance vont s'établir. Ces rapports peuvent mener à des situations extrêmes dans lesquelles le dominé voyant son système de récompense systématiquement inhibé va plonger dans la dépression, voire dans la toxicomanie, dans laquelle la dépendance à des drogues va mimer l'effet de la récompense sur son système nerveux.

Si les comportements humains et animaux dépendent en grande partie des mêmes circuits neuronaux, au moins chez les mammifères, le cerveau humain possède des spécificités cognitives et surtout imaginatives qui surpassent celles de nos plus proches cousins. On peut alors se demander quelle est la base génétique de cette spécificité ? Peut-on utiliser la génomique comparative, qui a permis par le passé de répondre à des questions fondamentales sur l'évolution des organismes vivants, afin de répondre à cette question ? En d'autres termes, le secret de la pensée et du comportement humain se trouvant inscrit dans notre génome, peut-on le percer à jour en séquençant entièrement celui-ci ?

LE SÉQUENÇAGE DU GÉNOME HUMAIN

Le projet « Génome humain » fut initié en 1990, par un consortium international comprenant des laboratoires américains, anglais, japonais, allemands, chinois et le Centre National de Séquençage français. Cette séquence fut publiée en 2004, trois ans après la publication d'une première séquence partielle et 14 ans après le début du projet. Ce génome – complété à 99 % – recouvre 2,85 milliards de

nucléotides, c'est-à-dire 18 fois plus que celui d'un chordé comme la cione et 69 fois plus que celui d'un choanoflagellé. Le génome humain contient 19 599 gènes identifiés, auxquels vont s'ajouter quelques-uns non encore identifiés, ce qui porte ce nombre aux alentours de 20 000-25 000 gènes, plus probablement dans la partie basse de cette fourchette. On remarque que ce nombre est du même ordre de grandeur que le nombre de gènes de la cione et à peine deux fois plus grand que celui d'un choanoflagellé. À nombre de gènes similaires, comment se fait-il que le génome humain contienne plusieurs dizaines de fois plus d'ADN que ces organismes ? Il y a plusieurs raisons à cela.

Tout d'abord, la plupart des gènes humains sont morcelés, c'est-à-dire que leur séquence est interrompue par un ou plusieurs introns (Figure 21). Les gènes humains contiennent 9 exons en moyenne, mais l'ensemble de ces exons couvre seulement 2 % du génome ! Une partie des 98 % restants correspond donc à des introns, mais une partie seulement, car la taille importante du génome humain tient aussi à la quantité d'éléments transposables qu'il contient. Nous avons vu précédemment qu'en comparaison du génome humain le génome du coq était pauvre en rétrotransposons. Ceux-ci sont une classe particulière d'éléments transposables, des morceaux d'ADN de taille modérée (quelques centaines à quelques milliers de nucléotides) capable de se propager, de « sauter », d'un endroit à un autre du génome, sur le même chromosome ou sur un autre chromosome. Il existe différentes familles d'éléments transposables – appelons-les plus simplement **transposons*** – et on les trouve dans tous les organismes vivants, bactéries, archées et

Transposon : Séquence d'ADN capable de se déplacer de façon autonome dans un génome. Il existe deux grandes familles de transposons, les transposons à ADN qui se déplacent par un mécanisme de « couper-coller » et les rétrotransposons qui sont d'abord transcrits en ARN avant leur transposition par un mécanisme de « copier-coller ».

eucaryotes. Ces bouts d'ADN qui se propagent dans un génome en se transposant d'un endroit à un autre, peuvent accidentellement sauter dans un gène essentiel et l'inactiver. Si c'est un gène essentiel à sa survie, la cellule va mourir.

Il est aussi possible que l'élément saute dans un endroit proche d'un gène et qu'il change l'expression de celui-ci, modifiant le programme génétique de la cellule qui le contient. Cela peut être bénéfique (ou pas) à la cellule ou à l'organisme. Dans d'autres cas, plus rares, un transposon va être à l'origine de la formation d'un nouveau gène ou d'un nouvel exon. Le génome humain contient plus de 3 millions de transposons, qui représentent près de la moitié de l'ADN total du génome. Plus d'un million d'entre eux appartiennent à une famille bien particulière que l'on appelle éléments *Alu**. Ces transposons *Alu* dérivent d'un petit ARN associé à un complexe protéique impliqué dans la maturation de certaines protéines, l'ARN 7SL. Une délétion de 155 nucléotides à l'intérieur de ce petit ARN, suivie d'une duplication, a donné naissance à une séquence *Alu* ancestrale, datant d'après l'émergence des mammifères. Il y a eu plusieurs vagues d'expansions des séquences *Alu* dans les génomes de mammifères (Figure 28). Une première amplification eut lieu il y a 55 millions d'années, au moment de l'émergence des primates, puis une seconde amplification de la sous-famille *Alu-Y* se produisit il y a 25 millions d'années. Finalement, il y a environ 5 millions d'années, c'est-à-dire après la séparation de la lignée menant aux chimpanzés et de la nôtre, deux petites amplifications de sous-familles jeunes se produisirent, une chez les hominidés et une deuxième dans la lignée humaine.

Lors de la première expansion des *Alu*, on estime qu'une nouvelle insertion d'élément se produisait à chaque nouvelle naissance. À l'heure actuelle, certains *Alu* sont encore actifs et il y a environ une nouvelle insertion toutes les 200 naissances chez l'homme. La raison pour laquelle ces éléments sont 200 fois moins actifs maintenant qu'à leur origine n'est pas totalement comprise, mais ce sont des

composants importants des génomes de primates en général et du génome humain en particulier, autant d'un point de vue quantitatif que pour leur impact sur le remodelage et l'évolution de ces génomes.

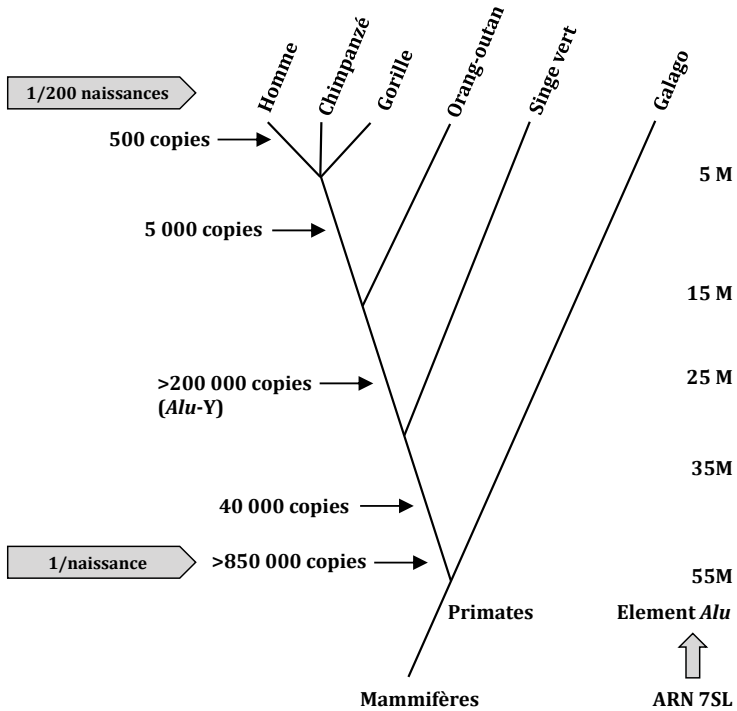


Figure 28 | Invasions des éléments Alu dans les génomes de primates. Les branches de l'arbre des primates sont représentées, les flèches sur la gauche indiquant les époques des différentes invasions d'éléments Alu, ainsi que leur nombre estimé lors de chaque invasion. Les dates indiquées sur la droite sont en millions d'années.

Finalement, la troisième cause de la taille importante du génome humain par rapport à ses ancêtres métazoaires ou chordés est la quantité importante de duplications segmentaires, ou plus généralement de *Copy Number Variations*¹ (CNV*), qu'il contient. Un CNV est un morceau d'ADN dont la taille peut varier de cinquante

1. Variations du nombre de copies.

à plusieurs dizaines de milliers de nucléotides, répété plusieurs fois dans le génome, de façon plus ou moins espacée le long des chromosomes. Environ 273 millions de nucléotides, soit presque 10 % du génome humain, correspondent à des CNVs. Le nombre de chacun de ces segments d'ADN répété varie en fonction des individus et des lignages. En effet, de la même façon que pour les séquences *Alu*, certains CNVs sont communs à tous les hominidés, d'autres sont spécifiques d'une branche seulement, ce qui permet de reconstruire partiellement l'histoire évolutive de ces séquences répétées.

Au final, si on retirait tous les transposons (1,3 milliard de nucléotides), tous les introns (900 millions de nucléotides) et tous les CNVs du génome humain (300 millions de nucléotides), cela aurait pour effet de réduire sa taille à 500 millions de nucléotides, à peine trois fois le génome de son ancêtre chordé vieux de 550 millions d'années.

La séquence du génome humain a été publiée comme un premier « brouillon » en 2001, qui couvrait 2,69 milliards de bases (environ 88 % du génome complet). De nombreuses régions contenant des éléments répétés n'avaient pu être correctement assemblées, particulièrement les régions des **centromères*** qui s'attachent aux microtubules du centrosome lors de la mitose, permettant la bonne ségrégation des chromosomes. Ceux-ci sont particulièrement grands chez les mammifères et leur nature hautement répétée les rendait impossibles à séquencer avec les techniques de l'époque. Trois ans plus tard, en 2004, était publiée la séquence dite « finale » du génome humain. Celle-ci couvrait alors 2,85 milliards de bases (environ 93 % de la séquence totale), mais contenait encore 341 « trous » correspondant aux centromères et à certaines séquences répétées impossibles à séquencer. Il a finalement fallu attendre 2022 et un changement drastique des technologies de séquençage utilisées pour obtenir la première séquence complète du génome humain, contenant les centromères et toutes les séquences répétées. Celle-ci couvre 3,055 milliards de bases et prédit 19 969 gènes codant des protéines. Il est vraisemblable que les prochaines années vont voir une floraison de génomes eucaryotes

séquencés avec la même technologie, permettant une étude beaucoup plus fine des propriétés de chacun et certainement des découvertes inattendues.

LA QUÊTE DES GÈNES QUI FONT L'HUMANITÉ

Comme nous l'avons vu, la puissance de la génomique dépend du nombre de séquences de génomes complets – ou quasi complets – disponibles. De plus, leur position dans l'échelle évolutive est importante. Plus les organismes sont éloignés, moins l'analyse sera fine, le seul moyen d'augmenter la précision des résultats étant de disposer de séquences d'individus relativement proches. Déterminer quels gènes définissent la spécificité de l'homme par rapport aux autres grands singes uniquement grâce à la séquence de son génome est impossible. C'est pour cette raison qu'en parallèle du projet de séquençage du génome humain, plusieurs projets de séquençage de génomes de primates ont vu le jour. Les deux cousins les plus proches de l'homme sont le chimpanzé (*Pan troglodytes*) et le bonobo (*Pan paniscus*). Ce sont donc sur eux que se sont concentrés les tout premiers programmes de séquençage de primates.

L'ancêtre commun à ces trois hominidés existait il y a environ 4,5-6,5 millions d'années et les chimpanzés se sont séparés des bonobos il y a seulement un million d'années. Ceux-ci vivent essentiellement au sud du fleuve Congo alors qu'on trouve des chimpanzés au nord du Congo et dans toute l'Afrique équatoriale. Ces derniers partagent certains comportements avec les êtres humains, les mâles montrant leur agressivité pour assurer leur domination ou pour obtenir un rapport sexuel, les individus coopérant pour défendre leur habitat et pour attaquer des groupes étrangers au leur. À l'inverse, les bonobos passent une bonne partie de leur existence à jouer, montrent peu d'esprit de compétition pour la domination, ne forment pas d'alliance dans le but d'agresser d'autres groupes et montrent une intense activité sexuelle, pas nécessairement associée à un souci de reproduction et fréquemment avec des individus du

même sexe. Ainsi, les chimpanzés comme les bonobos présentent des traits comportementaux très différents, certains d'entre eux communs avec l'espèce humaine. Cela suggère que l'ancêtre de ces trois hominidés devait posséder une mosaïque de traits communs aux trois lignages. Le génome du chimpanzé contient 3 milliards de nucléotides, celui du bonobo 2,7 milliards, deux valeurs proches du génome humain (3,055 milliards). Les gènes de bonobo sont en moyenne identiques à 99,6 % aux gènes de chimpanzé et à 98,7 % aux gènes humains. Les protéines codées par ces gènes sont donc quasiment identiques. Comme dans le génome humain, les transposons représentent environ la moitié du génome des deux singes, même si les sites d'insertion et le nombre de copies de chaque sous-famille varie légèrement entre les trois hominidés. Aucune indication d'un échange de gènes entre chimpanzés et bonobos n'a pu être détectée, suggérant que le Congo a représenté une barrière naturelle isolant les deux espèces de grands singes et empêchant les croisements interspécifiques. Comme la divergence des humains avec l'ancêtre des deux singes est plus ancienne que la divergence entre les deux singes, on s'attend à ce que les gènes des chimpanzés et ceux des bonobos soient aussi éloignés des gènes humains et évidemment plus proches entre eux. Or, on trouve ce que l'on appelle des événements de séparation incomplète des lignages, car 1,6 % du génome humain est plus proche du bonobo que du chimpanzé et 1,7 % est plus proche du chimpanzé que du bonobo. Cela signifie que les gènes de l'ancêtre commun n'ont pas tous évolué à la même vitesse après la séparation des trois lignages. Au total, 25 % de tous les gènes humains contiennent au moins une région qui présente un événement de séparation incomplète des lignages, les gènes codant des protéines membranaires ou impliquées dans l'adhésion cellulaire contenant les régions les plus longues. Ces gènes étant maintenant identifiés, leur étude devrait aider à la détermination des facteurs génétiques responsables des particularités comportementales des hominidés (Figure 29).

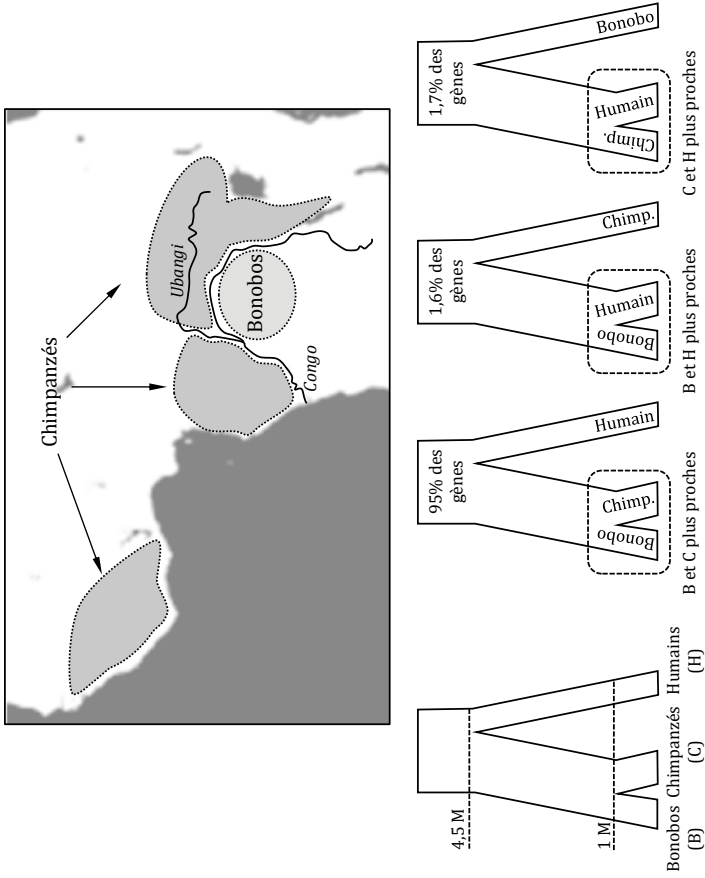


Figure 29 | Séparation incomplète de lignages entre humains, chimpanzés et bonobos. Haut : régions d'Afrique où vivent les chimpanzés et les bonobos. Bas : dates estimées de séparation des trois lignages (en millions d'années) et flux de gènes dans chacun après séparation.

QUELLE PART DE L'INTELLIGENCE EST-ELLE HÉRITABLE ?

Définir un ensemble de gènes déterminant l'intelligence humaine n'est pas chose aisée et a été le sujet de débats passionnés depuis des décennies. Mais de quelle intelligence parle-t-on ? Afin de simplifier la question, il sera ici fait référence uniquement aux capacités cognitives qui permettent de résoudre des problèmes de logique simple, dans ce que l'on nomme habituellement « test d'intelligence » ou « test de QI ». Le séquençage partiel ou total de milliers de génomes humains et leur analyse massive, associée aux résultats de ce genre de tests a permis de faire des progrès considérables dans ce domaine. Une étude de 2017 portant sur plus de 78 000 individus a réussi à mettre en évidence l'existence de 52 gènes influençant plus ou moins l'intelligence humaine. De même que les gènes ne sont pas strictement identiques entre deux espèces différentes, ils ne le sont pas entre deux individus de la même espèce. Ces différences peuvent porter sur une ou plusieurs bases, correspondre à de courtes insertions ou délétions à l'intérieur d'un gène, ou encore à d'autres modifications qui vont altérer sa séquence et éventuellement sa fonction. On nomme les différentes versions d'un même gène au sein d'une population des **allèles***. Quand ces différences portent sur un seul nucléotide, on les appelle alors des SNP (pour *Single-Nucleotide Polymorphism*²). Leur analyse fine a permis de détecter 336 SNPs présents dans 52 gènes, dont une douzaine avait déjà été identifiée comme étant associée à l'intelligence. Mais en quelle proportion ces 336 SNPs influencent-ils réellement l'intelligence cognitive ?

Allèle : Un gène peut présenter de légères différences dans sa séquence, qui peuvent plus ou moins modifier sa fonction. On appelle « allèles » les différentes formes d'un même gène.

La notion d'héritabilité en génétique humaine est complexe. Elle est définie par la proportion des différences observées entre individus qui

2. Polymorphisme d'un seul nucléotide.

peuvent être attribuées à des différences génétiques entre ces mêmes individus. Cette variabilité peut être quantifiée par des tests de QI ou par le niveau d'étude, qui est fortement corrélé à l'intelligence cognitive. En fin de compte, quelle proportion de la variation entre tests de QI (ou entre niveaux d'étude) au sein d'une population peut être attribuée à des gènes ? Jusqu'à récemment, la réponse était que seulement 1 % de ces variations pouvaient être attribuées à la génétique. Ou dit plus simplement, que 99 % de l'intelligence cognitive ne reposait pas sur des gènes, mais sur d'autres facteurs familiaux, sociaux ou environnementaux. Des progrès considérables ont été réalisés dans ce domaine en augmentant la taille des populations étudiées et en utilisant des scores cumulés reflétant les infimes variations portées par chacun des allèles impliqués dans l'intelligence cognitive. Car pris individuellement, aucun de ces gènes ne montre d'effet dramatique (dans un sens ou dans un autre) sur l'intelligence, mais ce sont bien les interactions entre les produits des différents allèles de ces gènes qui sont importantes. On se trouve là en présence d'un caractère – l'intelligence cognitive – qui n'est pas déterminé par un allèle unique, comme peuvent l'être la couleur des yeux ou le groupe sanguin, mais par les produits de dizaines ou de centaines de gènes ! En appliquant cette méthode à des cohortes de centaines de milliers de participants, on augmente terriblement la puissance de ce type d'analyse. À l'heure actuelle, on estime qu'environ 10 % de la variabilité cognitive entre individus est attribuable à des différences génétiques entre eux. Il est vraisemblable que l'augmentation de la taille des populations testées et l'amélioration des méthodes d'analyse permettront d'augmenter cette valeur, mais nul ne sait réellement quelle sera la limite atteinte, bien que des études sur des vrais jumeaux suggèrent que l'héritabilité maximum d'un trait complexe comme l'intelligence serait de l'ordre de 50 %. Ce qui reviendrait à dire que les 50 % restants seraient dus à des facteurs autres, familiaux, sociaux ou environnementaux.

Notons au passage que tous ces gènes ont des homologues chez nos cousins chimpanzés et bonobos. Définir ce qui détermine les

capacités cognitives et imaginatives des hommes par rapport aux autres primates s'avère donc une tâche ardue si l'on considère le nombre d'associations d'allèles possibles chez un individu. En plus de ces différences alléliques, on trouve de grandes variations du nombre de copies de certains gènes entre les hominidés, ces différences étant dues aux CNVs et aux duplications segmentaires. Une partie des recherches se dirige actuellement dans cette direction-là, espérant aider à la compréhension de ce qui fait de nous des hommes et nous différencie de nos plus proches cousins.

LE MÉTAGÉNOME HUMAIN

Nous avons vu l'importance des symbioses dans l'évolution des espèces vivantes. Les cellules eucaryotes sont le résultat de symbioses et de nombreux organismes végétaux ou animaux vivent en symbiose avec un ou plusieurs autres organismes ou microorganismes. C'est également notre cas, car l'être humain est un symbiote. Il vit en association étroite avec des communautés bactériennes et fongiques présentes à la surface de son épiderme, de son tractus génital et de son tube gastro-intestinal. C'est dans ce long tube digestif qui court de la bouche à l'anus que vivent en symbiose avec nous la plupart des microorganismes, essentiellement des bactéries, mais également des levures et des eucaryotes unicellulaires. On estime que 10 000-100 000 milliards de microorganismes vivent dans le tube digestif – dix fois plus que le nombre de cellules qui constituent un individu adulte – la plupart d'entre eux dans le côlon. La totalité des microorganismes symbiotiques d'un individu porte le nom de **microbiote*** et l'ensemble de leurs gènes est appelé microbiome, bien qu'en pratique ces termes soient souvent employés de façon synonyme. Les études portant sur le microbiote ont commencé sérieusement au début du XXI^e siècle, comme le prolongement logique de l'achèvement du programme de séquençage du génome humain. On sait à la fois beaucoup et encore bien peu de choses sur le rôle du microbiote dans la physiologie d'un individu. Il détermine en partie le métabolisme

des nutriments, en particulier de certaines graisses, jouant ainsi une fonction dans l'obésité. L'assimilation de certains médicaments y est aussi liée, de même que le renouvellement des cellules de l'intestin. Il est en interaction directe avec le système immunitaire, contribuant à certaines maladies telles que la maladie de Crohn – une inflammation chronique du tube digestif – ou l'asthme. Mais un des effets les plus remarquables et les moins compris est son rôle dans le comportement locomoteur (et peut-être cognitif ou psychiatrique) par la production de molécules actives sur le système nerveux central. Pour toutes ces raisons, en Europe, en Asie et aux États-Unis ont démarré des programmes de recherche très ambitieux ayant pour but de percer à jour tous les secrets du microbiote. Les résultats des premières études, portant sur des échantillons provenant de 124 individus européens, ont produit 576 milliards de nucléotides de séquence, soit 200 fois la taille du génome humain. Plus de 99 % des gènes identifiés sont d'origine bactérienne et chaque individu contient environ 160 espèces bactériennes différentes. Ce travail a permis de mettre en évidence l'existence d'un microbiote commun à tous les échantillons, contenant des fonctions importantes pour la dégradation des sucres, la synthèse des acides gras ainsi que celle des acides aminés et des vitamines indispensables à notre survie. L'étude fine de ces 124 **métagénomés*** permettra aussi de mieux comprendre les interactions génétiques entre le microbiome et son génome humain hôte.

Métagénome : Le métagénome humain est l'ensemble des gènes humains, bactériens, fongiques et viraux vivant dans l'écosystème représenté par un être humain. La métagénomique est la science qui étudie les génomes d'organismes présents dans un échantillon environnemental donné.

De nombreuses espèces de levures font également partie du microbiote intestinal humain. Des levures comme *Saccharomyces cerevisiae*, *Candida albicans*, *Candida glabrata* et des espèces de *Penicillium*, mais

également des basidiomycètes³, *Malassezia*, *Rhodotorula* et *Ustilago*, sont retrouvés dans des proportions variables entre individus en bonne santé et individus atteints de maladies chroniques de l'intestin. Les quantités relatives de ces différentes levures varient en fonction de l'état de santé du système digestif. Par exemple, *Saccharomyces cerevisiae* est plus souvent retrouvée dans des muqueuses intestinales en bonne santé, alors que *Candida glabrata* est plus souvent retrouvée associée à la maladie de Crohn. De façon remarquable, le rapport entre levures et basidiomycètes est plus élevé chez les individus en bonne santé et plus faible chez les malades atteints d'inflammations intestinales chroniques. De plus, certaines de ces levures varient en proportion de la même façon que certaines bactéries du microbiote, suggérant que des interactions existent entre levures et bactéries au sein de leur environnement intestinal, interactions dont on ignore à peu près tout à l'heure actuelle.

Dans ces conditions, on comprend mieux que la pharmacogénomique, qui consisterait à administrer des médicaments à un individu en tenant compte de son patrimoine génétique afin d'obtenir une efficacité maximum et des effets secondaires minimums, ne peut pas se contenter de la séquence du génome de l'individu à guérir, mais doit aussi prendre en compte celle du microbiote. Comprenons bien que nous sommes encore à des années d'une approche pragmatique et rationnelle de la pharmacogénomique. Nous sommes à peu près dans la situation d'un explorateur débouchant d'une forêt équatoriale, armé de sa seule machette et découvrant l'océan qu'il doit traverser pour rentrer chez lui ! La tâche est colossale, mais pas insurmontable. Elle prendra du temps, beaucoup de ressources et ne sera vraisemblablement pas achevée avant la prochaine génération. Mais il n'y a techniquement rien d'impossible et il est probable que le déchiffrage

3. Champignons ou moisissures se reproduisant par différenciation de cellules spécialisées, les basides, qui contiennent les cellules germinales ou spores. Les « champignons à chapeau » cueillis dans les bois à l'automne sont des basidiomycètes et c'est leur appareil reproducteur que l'on consomme.

et l'interprétation du métagénome humain, c'est-à-dire du génome et du microbiome associé, de façon individuelle, transformeront de façon profonde et irréversible la recherche biomédicale ainsi que notre façon de pratiquer la médecine.

À RETENIR

- Le cerveau est le centre du système de récompense et de punition qui régule les comportements de tous les mammifères.
- Le séquençage du génome humain a montré que notre génome contenait 5-10 fois moins de gènes qu'anticipé, que ces gènes étaient morcelés par de nombreux introns et que les séquences répétées (comme les transposons) représentaient plus de 50 % de la séquence totale du génome.
- Les humains présentent des caractères de comportement social proches à la fois des chimpanzés et des bonobos, nos plus proches cousins.
- Le microbiote est l'ensemble des microorganismes vivant en symbiose avec un être humain et qui régule l'ensemble de son métabolisme.

BIBLIOGRAPHIE

Le cerveau : Changeux, L'homme neuronal, Pluriel Sciences, Hachette (1983).

Les éléments *Alu* dans les génomes de primates : Batzer et Deininger, *Nature Reviews Genetics* 3 : 370-380 (2002).

La comparaison des génomes humain, bonobo et chimpanzé : Prüfer, Munch, Hellmann, Akagi *et al.*, *Nature* 486: 527-531.

12

L'homme moderne est un hybride

L'Afrique est le berceau de l'humanité. Cette théorie scientifique, qui n'est restée pendant longtemps qu'une hypothèse, repose sur plusieurs découvertes qui n'ont été confirmées que relativement récemment. L'hypothèse de « l'East Side Story » postulait que la formation du **Grand Rift*** en Afrique de l'Est il y a une dizaine de millions d'années avait séparé les populations de grands singes en deux lignées. Alors qu'à l'ouest du rift, les populations simiennes auraient continué à profiter

Grand Rift : La vallée du Grand Rift est un ensemble géologique de failles de 6 000 km de long du nord au sud, s'étirant entre le nord du Liban et l'Afrique australe, en passant par la mer Rouge et les grands lacs africains. Il a longtemps été pensé par certains paléontologues que les premiers Homo sapiens avaient dû voir le jour à l'est du rift africain, dans une région de savanes propice à l'acquisition de la station debout, alors que les grands singes vivaient à l'ouest du rift dans une région de forêts denses, plus propices à leur anatomie et à leur mode de vie.

d'abondantes précipitations propices à une végétation boisée, la mise en place d'un climat plus sec à l'est du rift aurait été accompagnée de la formation de savanes faiblement arborées, moins propices à un mode de vie arboricole. La présence de nombreux fossiles d'australopithèques, les plus proches cousins des humains, dans la vallée du Grand Rift, aurait longtemps conforté cette hypothèse. Néanmoins, la découverte de restes d'australopithèques dans les années 1990 au Tchad, bien au nord du Grand Rift, par l'équipe de Michel Brunet, la remet en question.

La possibilité que les premiers *Homo sapiens* aient vu le jour à l'est du rift, puis aient envahi le reste du continent à partir de là restait en revanche posée. Cette hypothèse, basée sur des fossiles en mauvais état, fut confortée par la découverte en Éthiopie de crânes humains partiellement fossilisés. L'état remarquable de conservation de ces restes humains montre clairement des caractéristiques morphologiques d'*Homo sapiens*, même si certains caractères ancestraux demeurent. La **datation radiométrique*** est assez précise pour leur donner un âge : 195 000 ans pour l'un, 160 000 ans pour l'autre (Figure 30). Jusqu'en 2017, le plus ancien *Homo sapiens* connu vivait donc en Éthiopie et avait traversé la mer Rouge environ 60 000 ans auparavant pour peupler le monde. En juin 2017, une découverte étonnante allait bouleverser cette histoire. Dans le Jebel Irhoud marocain, des restes fossilisés de cinq *Homo sapiens* (trois adultes, un adolescent et un enfant d'environ 8 ans) sont découverts par une équipe franco-allemande. La datation du site de cette découverte donne un âge de 315 000 ans (\pm 34 000 ans) à ces fossiles, repoussant l'âge précédemment établi pour *Homo sapiens* de plus de 100 000 ans ! De plus, cette découverte prouve que notre ancêtre n'a pas uniquement vu le jour dans la partie orientale du continent africain, à l'est de la vallée du Grand Rift, mais qu'il vivait aussi dans une autre région d'Afrique, relançant le débat sur son origine, ou plutôt sur ses origines.

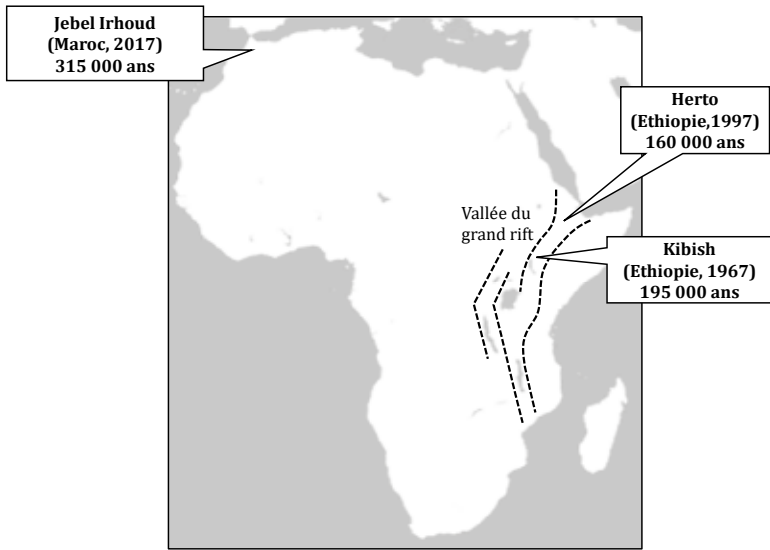


Figure 30 | Localisation, date de la découverte et datation radiométrique des restes *Homo sapiens* en Afrique.

MASTIQUER OU RÉFLÉCHIR, IL FAUT CHOISIR

La découverte de nombreux fossiles d'humains ancestraux depuis le XIX^e siècle a aidé à mettre en place les histoires évolutives possibles de la lignée humaine. Le plus vieux représentant connu du **genre*** *Homo* est l'*Homo habilis*, qui signifie en latin « homme manuel, habile de ses mains ». Celui-ci vivait en Afrique il y a 1,5-2,5 millions d'années. Il avait déjà acquis la bipédie, taillait ses propres outils à partir de pierres pour obtenir des éclats tranchants, même si cette caractéristique semble avoir été partagée par certaines espèces n'appartenant pas au genre *Homo*, comme les australopithèques qui l'ont précédé.

Genre : Rang de classification taxonomique qui vient juste au-dessus de celui de l'espèce. Pour l'homme, le nom de genre est *Homo* et le nom d'espèce *sapiens*.

L'*Homo habilis* était principalement végétarien, mais occasionnellement charognard. Il mesurait 1,20-1,50 mètre pour un poids de 30-40 kilogrammes et sa capacité crânienne était légèrement supérieure à celle du chimpanzé (550-800 cm³). L'*Homo erectus*, qui signifie littéralement en latin « homme dressé », vivait quant à lui en Afrique, en Asie centrale et orientale, entre deux millions d'années et 100 000 ans avant le présent. Il maîtrisait la station debout, mesurait entre 1,50 et 1,65 mètre, pesait de 45 à 55 kilogrammes et sa capacité crânienne était nettement plus importante (850-1 100 cm³). Cette nette augmentation du volume du cerveau entre ces deux humains ancestraux est encore à l'heure actuelle le sujet de nombreux débats passionnés entre anthropologues. Chez les humains modernes que nous sommes, le cerveau consomme 60 % de l'énergie d'un nourrisson au repos et chez l'adulte c'est 25 % de l'énergie qui est consommée par cet organe, contre seulement 8 % chez un grand singe. Mais les humains ingèrent à peu près la même quantité de calories que des mammifères de corpulence comparable. Une augmentation de la taille du cerveau a donc dû s'accompagner de la réduction de volume d'un autre organe, à consommation calorique constante. Il a donc été suggéré que le tube digestif humain a diminué de taille de façon concomitante au changement d'un régime essentiellement végétarien avec des apports réduits en viande à une alimentation plus riche en viande. En effet, celle-ci contenant plus de protéines que les végétaux, elle apporte une alimentation plus énergétique à poids égal. La consommation régulière de viande a pu rendre possible la réduction graduelle de la taille du système digestif, en particulier celle de l'intestin, permettant en contrepartie l'accroissement du cerveau sans besoin d'améliorer le métabolisme énergétique.

Homo erectus est essentiellement un chasseur de petits animaux, mais pas exclusivement, car il peut lui arriver de chasser de gros mammifères comme l'éléphant. Il est également consommateur de fruits et de racines. Il domestique le feu au moins 400 000 ans avant notre ère, ce qui lui donne la possibilité de fabriquer des outils

plus performants, de cuire ses aliments, de se réchauffer et de tenir à distance les prédateurs. L'utilisation du feu dans l'alimentation a profondément transformé ces sociétés primitives. La cuisson de la viande a considérablement réduit le temps nécessaire à l'alimentation carnée, la viande cuite se mastiquant plus facilement, le collagène animal étant en partie détruit par la cuisson. Celle des légumes, en particulier des tubercules riches en amidon, a également raccourci le temps consacré aux repas, les légumes cuits étant plus faciles à assimiler et se digérant mieux car les longs polymères de sucres végétaux sont cassés en petites molécules plus faciles à assimiler. Ce temps de repas économisé a alors pu être employé à la confection de meilleurs outils, plus adaptés aux besoins, ainsi qu'au développement de nouvelles idées, de nouveaux concepts, prémices des futures inventions de l'*Homo sapiens* (Tableau 2).

Le fait que les dents d'*H. erectus* soient de plus petite taille que celles de ses ancêtres suggère que la mastication était moins importante pour lui, ce qui n'est pas le cas lorsque l'on doit se nourrir de viande ou de végétaux crus. Cette hypothèse présuppose que le feu ait été domestiqué et utilisé pour la cuisson des aliments il y a au moins 1,5 million d'années. Or, les plus anciennes traces de foyers ayant servi pour l'alimentation ont été retrouvées en Israël datées de 800 000 ans, soit au moins 700 000 ans après l'augmentation du volume de la boîte crânienne chez *H. erectus*. Il est donc vraisemblable que celle-ci soit liée au changement global de régime alimentaire plutôt qu'à une maîtrise du feu dans la préparation des repas de ces humains ancestraux.

Homo erectus est un infatigable voyageur. Des outils de pierre et des restes d'animaux ont été découverts dans la province de Kalinga, au nord des Philippines. Plus de 400 os et la presque totalité d'un squelette de *Rhinoceros philippinensis* proprement découpé à l'aide d'outils en pierre témoignent de la présence d'une communauté humaine ancestrale. Ces grands voyageurs – très vraisemblablement des *Homo erectus* – seraient sortis d'Afrique et auraient colonisé cette région du

monde il y a 700 000 ans d'après la datation. Ils seraient passés par l'île de Bornéo et Palawan à l'ouest ou bien par la Chine et Taïwan au nord pour atteindre finalement l'île de Luzon aux Philippines. Bien que le niveau des océans ait été plus bas à cette époque, il est vraisemblable (même si pas formellement prouvé) que ces humains ancestraux aient traversé l'une ou l'autre étendue d'eau sur des embarcations de fortune, pour coloniser ces îles bien avant l'arrivée des néandertaliens et des humains modernes. Des restes d'*Homo erectus* retrouvés sur l'île de Java et en Chine confirment que celui-ci a bien voyagé jusqu'en Indonésie, en faisant le premier humain à sortir d'Afrique.

ON A – PRESQUE – TOUS EN NOUS QUELQUE CHOSE DE NÉANDERTAL

Il n'est pas complètement clair si *Homo erectus* est un ancêtre direct de l'homme de Néandertal et de l'*Homo sapiens*, ou bien si des espèces intermédiaires comme *Homo heidelbergensis*, *Homo mauritanicus* ou *Homo rhodesiensis* se sont intercalées. Cette question est encore matière à nombreux débats chez les paléontologues.

L'homme de Néandertal vivait en Europe et en Asie occidentale et centrale entre 400 000 et 28 000 ans avant l'époque présente. Il est de grande taille (1,65 mètre pour les hommes) et de corpulence robuste (90 kilogrammes). Il est, par de nombreux caractères, très proche d'*Homo sapiens*. Il fabrique des armes et des outils très sophistiqués et c'est un grand chasseur. Il a la possibilité physiologique de posséder un langage articulé et la complexité de ses outils suggère des capacités cognitives importantes compatibles avec l'usage de la parole. Les premières sépultures connues sont néandertaliennes et remontent à 100 000 ans. Des gravures sur des os ou des pierres, l'utilisation d'ocre et des collections de plumes ou de minéraux rares prouvent l'existence d'une culture artistique, que l'on croyait autrefois l'apanage exclusif des *Homo sapiens* (Tableau 2).

On en sait beaucoup plus à l'heure actuelle sur l'homme de Néandertal grâce au séquençage de son génome. Séquençer de

Tableau 2 | Principales caractéristiques des hommes archaïques.

	<i>Homo habilis</i>	<i>Homo erectus</i>	<i>Homo neanderthalensis</i>
Époque	1,5-2,5 millions d'années	100 000 ans-2 millions d'années	400 000-28 000 ans
Région	Afrique	Afrique, Asie centrale et orientale	Europe et Asie occidentale
Taille	1,20-1,50 m	1,50-1,65 m	1,65 m (pour les hommes)
Poids	30-40 kg	45-55 kg	90 kg (pour les hommes)
Capacité crânienne	550-800 cm ³	850-1 100 cm ³	1 500 cm ³ *
Régime alimentaire	Végétarien (charognard)	Omnivore	Omnivore
Chasseur	Non	Petit gibier, gros gibier occasionnel	Gros gibier, pêcheur
Outils	Pierre taillée	Pierre taillée	Pierre, os, outils emmanchés
Maîtrise du feu	Non	Oui	Oui
Cuisson des aliments	Non	Peu probable	Oui
Sépultures	Non	Non	Oui
Culture artistique	Non	Non	Oui
Culture médicale	Non	Non	Oui
Langage articulé	?	?	Possible

* Capacité crânienne légèrement supérieure à celle de l'homme moderne (1 400 cm³).

l'ADN ancien conservé dans des fragments d'os est à l'heure actuelle possible, bien que techniquement difficile. L'ADN s'hydrolyse au cours du temps et se modifie chimiquement. Une molécule d'ADN du Pléistocène tardif, époque à laquelle vivaient les derniers Néandertaliens il y a 40 000 ans, est dégradée en petits morceaux de moins de 200 nucléotides. De plus, la désamination des cytosines en uraciles (voir Chapitre 4) entraîne des transitions de C vers T dans cet ADN, modifiant sa séquence. Cette limite physique due à l'hydrolyse de l'ADN rend impossible avec les techniques actuelles le séquençage d'ADN vieux de plus de quelques centaines de milliers d'années, le record actuel ayant été atteint avec le séquençage des restes d'un mammouth vieux d'un peu plus d'un million d'années congelé dans le permafrost sibérien. Accéder à la séquence de génomes de dinosaures morts il y a plusieurs dizaines de millions d'années relève donc entièrement du domaine de la science-fiction hollywoodienne. De plus, la manipulation d'ADN ancien requiert de travailler dans des conditions de propreté absolue afin d'éviter les contaminations par l'ADN moderne, infiniment plus abondant et présent dans les cellules épithéliales de la peau, la salive, les poils, les cheveux, etc., des expérimentateurs. Toutes ces raisons font que la détermination de la séquence d'un organisme vivant ancien est complexe et nécessite une maîtrise parfaite de ces différents paramètres.

Le premier *Homo neanderthalensis* a été séquençé à partir des restes de trois os découverts dans une grotte à Vindija, en Croatie, datés de 38 000 à 44 000 ans. À partir de 400 milligrammes de poudre d'os, l'équipe de Svante Pääbo, un chercheur suédois travaillant en Allemagne et spécialiste du séquençage d'ADN ancien, annonce en 2010 que 5,3 milliards de nucléotides appartenant à l'homme de Néandertal ont été séquençés et partiellement assemblés. Cette publication dans la prestigieuse revue américaine *Science* fait grand bruit car elle s'accompagne d'une découverte majeure qui remet complètement en cause la vision alors acceptée de l'évolution

humaine. Tout d'abord, cette séquence permet d'estimer la divergence entre *Homo sapiens* et *Homo neanderthalensis* à 825 000 ans. Cet âge a depuis été ramené à 550 000 ans par des estimations plus précises obtenues grâce à d'autres séquences de Néandertaliens. En comparant le génome néandertalien à cinq génomes humains d'origines géographiques diverses, il apparaît que certaines régions de ce génome sont plus proches des individus non africains que de ceux d'origine africaine. Une étude plus poussée de ces régions montre qu'elles correspondent à des transferts de gènes entre l'homme de Néandertal et l'*Homo sapiens*, gènes qui ont été par la suite fixés dans la population humaine et conservés jusqu'à nos jours. Ce transfert est daté d'il y a 50 000 à 80 000 ans. Ces résultats amènent à la conclusion que les deux populations humaines se sont mélangées et ont eu une descendance fertile, avant leur expansion vers l'Asie et vers l'Europe. La taille de la **population effective*** découlant de ces croisements, à l'origine de la colonisation de toute notre planète, est estimée à seulement un millier d'individus. Un tel scénario est compatible avec les restes archéologiques qui montrent que les hommes modernes apparaissent au Proche-Orient il y a 40 000-60 000 ans alors que les Néandertaliens y vivaient déjà. Les humains modernes hors d'Afrique possèdent donc des gènes d'origine néandertalienne. Le génome de chacun d'entre nous (à part les Africains modernes qui sont des *Homo sapiens* « purs ») contient environ 1-4 % de gènes provenant de notre lointain ancêtre de Néandertal, et si on met bout à bout tous ces morceaux d'ADN on arrive à reconstituer 35-45 % du génome d'*Homo neanderthalensis*. Cette étonnante découverte a valu à Svante Pääbo d'obtenir le prix Nobel de Médecine ou Physiologie en 2022.

L'HOMME DE NÉANDERTAL PRATIQUAIT UNE MÉDECINE RUDIMENTAIRE

Mais le séquençage peut également permettre d'éclairer la **paléo-écologie*** et donner de précieuses informations sur la culture et le

Paléocéologie : Science qui étudie les relations des êtres vivants ancestraux avec leur environnement.

mode de vie de nos ancêtres disparus, grâce à l'ADN prélevé sur les dents des anciens humains. La plaque dentaire est un biofilm formé de bactéries, de restes alimentaires et de protéines salivaires.

Si cette plaque n'est pas enlevée régulièrement, elle se minéralise sous l'action du calcium contenu dans la salive et se transforme en tartre. Les restes de plusieurs hommes de Néandertal ont été découverts en Espagne et en Belgique, leurs mâchoires contenant du tartre fossilisé. L'ADN contenu dans ce tartre a été extrait et séquencé. L'analyse des séquences a montré que les hommes vivant en Belgique mangeaient du rhinocéros laineux et du mouton, espèces vivant dans le nord de l'Europe à cette époque, ainsi que du coprin cendré, un champignon. Le régime alimentaire des hommes d'Espagne était plutôt végétarien et basé sur la consommation de *Schizophyllum commune*, un champignon poussant sur le bois mort et généralement considéré comme non comestible, de pignons de pin et de mousse, montrant que les hommes de Néandertal adaptaient leur régime alimentaire aux ressources disponibles localement. De façon remarquable, des séquences de peuplier (qui contient naturellement de l'aspirine) et de *Penicillium rubens* (un champignon fabriquant la pénicilline) furent également découvertes chez un des individus vivant en Espagne, suggérant que les propriétés médicamenteuses de ces produits lui étaient connues et qu'il les utilisait vraisemblablement pour soigner un abcès dentaire. De plus, cette étude mit en évidence de nombreuses séquences de bactéries pathogènes de la cavité buccale et impliquées dans la formation de caries, ainsi que d'une bactérie commensale de la bouche, *Methanobrevibacter oralis*, ce qui permit de reconstituer son génome ancestral. L'horloge moléculaire détermina que la divergence entre cette bactérie ancestrale et son équivalent moderne datait de 112 000 à 143 000 ans, un âge beaucoup plus récent que la divergence entre l'homme moderne et

le Néandertal (450-750 milliers d'années). Cette découverte montre que des transferts de microbes de la cavité buccale se sont produits entre les deux espèces d'hommes, longtemps après la séparation des deux lignages, prouvant l'existence d'interactions intimes prédatant l'époque de leur hybridation (50 000-80 000 ans).

L'HOMME DE DENISOVA

L'équipe de Svante Pääbo va déclencher un deuxième coup de tonnerre quelques mois plus tard, en publiant la séquence d'un homme archaïque dont les restes furent découverts dans une grotte des monts de l'Altaï, au sud de la Sibérie. Cet « homme de Denisova » du nom de la grotte où ces restes ont été trouvés, est un cousin de l'homme de Néandertal, les deux lignées ayant divergé il y a environ 400 000 ans, 150 000 ans après la divergence de leur ancêtre commun avec l'*Homo sapiens*, permettant ainsi d'affiner les données précédemment obtenues avec l'homme de Néandertal. En comparant cette séquence avec celle d'hommes modernes d'origines géographiques et ethniques différentes, il est apparu que le génome des individus originaires de Mélanésie contenait 4-6 % d'ADN de Dénisovien, alors que le génome des autres hommes modernes n'en contenait pas. Ceci suggère que l'homme de Denisova s'est mélangé aux populations d'*Homo sapiens* qui vivaient en Asie du Sud-Est, populations qui ont plus tard colonisé la Mélanésie et qui portent encore aujourd'hui dans leur génome la trace de cette hybridation ancestrale. Des études plus poussées réalisées en comparant deux génomes néandertaliens d'origines différentes, le génome dénisovien et 25 génomes d'hommes modernes, montrent de nombreux cas d'**introgressions*** entre les différentes populations archaïques. Il existe également des traces de transfert d'ADN entre Dénisoviens et une autre population d'hommes encore non identifiés, mais qui se seraient mélangés aux Dénisoviens il y a environ un million d'années et qui pourraient être des *Homo erectus*.

Introgression : Transfert interspécifique d'ADN par hybridation, entre deux espèces distinctes, mais suffisamment proches pour donner une descendance fertile. Les introgressions peuvent être de très grande taille, contrairement aux transferts horizontaux qui impliquent généralement un seul ou un très petit nombre de gènes. Le terme était originellement utilisé en génétique des plantes, jusqu'à ce que l'on découvre de nombreux cas d'introgressions en dehors du règne végétal.

Plus récemment, un autre fragment humain de la même grotte de Denisova, vieux d'environ 50 000 ans a été séquencé. Il s'agissait d'une jeune fille d'environ 13 ans à la date de son décès. La séquence de l'ADN mitochondrial, hérité essentiellement de la mère, montre que cette jeune dénisovienne était la fille d'une mère néandertalienne et d'un père dénisovien. Cette remarquable découverte prouve que ces deux populations d'humains ancestraux se sont non seulement partagé le même habitat mais qu'ils se sont également reproduits entre eux.

LES APPORTS DES GÈNES ANCESTRAUX AUX GÉNOMES HUMAINS MODERNES

Le génome de tous les humains hors d'Afrique contient donc 1-4 % d'ADN néandertalien et les mélanésien ont reçu en plus des gènes de l'homme de Denisova. On peut alors se demander dans quelle mesure la présence de ces gènes – ou plutôt des allèles de ces gènes – ancestraux a contribué aux différences entre êtres humains modernes. Il était connu depuis plusieurs années que les Européens étaient plus résistants à certaines infections bactériennes et virales que les Africains, à cause d'une réponse immunitaire différente. Une étude d'une équipe de l'Institut Pasteur dirigée par Lluís Quintana-Murci a montré en 2016 que cette différence était due à la présence de certains gènes néandertaliens dans le génome des Européens, conférant à leur système immunitaire un avantage sélectif qui s'est propagé jusqu'à aujourd'hui. Mais notre héritage néandertalien peut

aussi présenter un désavantage pour certaines maladies infectieuses. Une région de 50 000 bases sur le chromosome 3 humain est associée à un risque 60 % plus élevé de déclencher une forme grave de COVID-19. Cette séquence d'ADN est présente chez la moitié de la population d'Asie du Sud, chez 16 % des Européens et 9 % des Américains.

Un autre exemple est la présence d'un morceau d'ADN de 200 000 bases d'origine néandertalienne sur le chromosome 3 humain qui contient le gène *HYAL2*, impliqué dans la réponse cellulaire aux rayons ultra-violets. L'expression de ce gène diminue fortement après l'exposition aux UVs, entraînant de fréquents coups de soleil, mais permettant aussi de fabriquer plus efficacement la Vitamine D. Ce morceau d'ADN néandertalien est présent à une fréquence élevée (> 50 %) dans les populations d'Asie de l'Est mais presque absent du reste du monde, montrant que l'allèle néandertalien a été positivement sélectionné au cours de l'évolution dans les populations d'Asie de l'Est.

Mais *Homo neanderthalensis* n'est pas le seul à avoir apporté sa contribution au génome humain moderne. Le plateau tibétain est un des endroits les plus inhospitaliers au monde, en raison de la faible concentration en oxygène (seulement 40 % de celle du niveau de la mer), du climat et des faibles ressources alimentaires. Néanmoins, les Tibétains ont réussi à s'adapter à ces conditions extrêmes, en partie grâce à une plus faible mortalité infantile chez les femmes d'origine tibétaine, comparée aux grossesses des autres femmes essayant d'enfanter à haute altitude. La comparaison de leurs génomes avec ceux des autres populations humaines modernes et ancestrales a mis en évidence la présence d'un allèle particulier du gène *EPAS1* dans les populations tibétaines. Ce gène régule la concentration d'hémoglobine dans le sang à haute altitude. La séquence de ce gène chez les Tibétains ne peut s'expliquer que par son origine dénisovienne, le seul autre génome possédant cette séquence. Celle-ci est différente dans les autres populations humaines, en particulier chez les Han de la

Chine toute proche. *H. denisova* a donc apporté ce gène dans l'ancêtre des populations humaines modernes et il a été sélectionné au cours de l'évolution, car procurant un avantage sélectif aux populations tibétaines vivant à très haute altitude.

LES PREMIÈRES MIGRATIONS DE NOS ANCÊTRES HUMAINS

En résumé, quatre lignées d'hommes ancestraux ont quitté l'Afrique au cours du dernier million d'années : *Homo erectus* qui a colonisé une partie de l'Asie, de l'Indonésie et des Philippines, *Homo neanderthalensis* qui a colonisé le nord de l'Asie, le Moyen-Orient et l'Europe, *Homo denisova* qui a colonisé le nord de l'Asie, et *Homo sapiens*. Celui-ci est le dernier à sortir d'Afrique, il y a environ 60 000 ans. La route utilisée a sans doute été celle qui longe le Sud de la péninsule arabe (Figure 31, route A), mais il est aussi possible que les migrants aient traversé la mer Rouge au niveau de l'actuel désert du Sinaï (Figure 31, route B). La première hybridation, entre *H. neanderthalensis* et *H. sapiens*, se produit au Moyen-Orient et donne naissance à l'homme génétiquement moderne. La deuxième hybridation, entre le premier homme hybride et l'homme de Denisova se produit en Mélanésie et va engendrer la population mélanésienne actuelle, qui vit entre la Nouvelle-Guinée et les îles Fidji.

Ainsi, la lecture de ces génomes ancestraux permet d'apporter un nouvel éclairage à l'origine des hommes modernes et montre qu'à la différence des modèles communément admis, au moins trois espèces – sinon quatre – d'hommes ancestraux se sont mélangées, à différentes époques préhistoriques, pour donner naissance aux populations humaines actuelles. Ces découvertes fondamentales montrent que les soi-disant barrières inter-espèces peuvent être aisément franchies et remettent complètement en cause la notion, la définition même d'espèce, si chère à Darwin et aux premiers naturalistes.

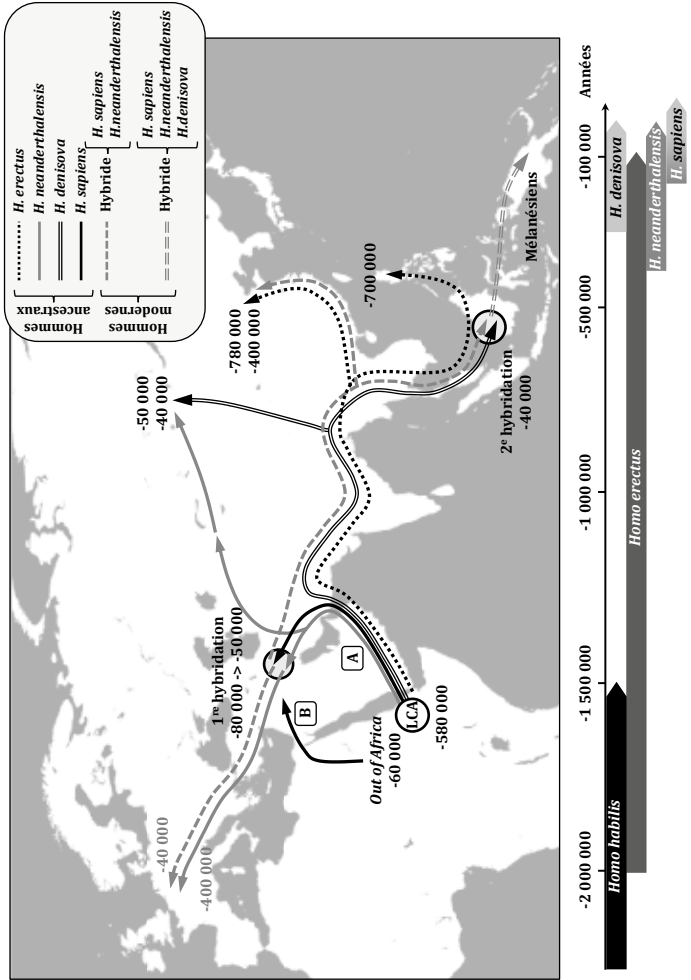


Figure 31 | Hybridations et évolution des différentes lignées humaines dans le temps et dans l'espace. LCA : Last Common Ancestor (dernier ancêtre commun de *H. neanderthalensis*, *H. denisova* et *H. sapiens*), estimé à -580 000 ans. Les deux routes possibles pour la sortie d'Afrique sont notées A et B (voir texte). Les dates sont données par rapport à l'époque actuelle.

À RETENIR

- Le passage d'une alimentation essentiellement végétarienne à un régime plus riche en viande a permis l'augmentation de taille du cerveau chez *Homo erectus*.
- *Homo sapiens* est sorti d'Afrique il y a environ 60 000 ans pour conquérir le monde.
- Le génome des *Homo sapiens* modernes contient 1-4 % d'ADN d'*Homo neanderthalensis*, preuve que nos lointains ancêtres se sont reproduits avec les hommes de Néandertal il y a environ 40 000 ans.
- Les humains résultant de ces croisements se sont eux-mêmes reproduits avec *Homo denisova* et leurs descendants ont peuplé la Mélanésie.

BIBLIOGRAPHIE

- Augmentation de taille du cerveau et régime alimentaire : Gibbons, *Science* 316: 1558-1560 (2007).
- L'origine pan-Africaine d'*Homo sapiens* : Hublin, Ben-Ncer, Bailey, Freidline *et al.*, *Nature* 546: 289-292 (2017).
- L'homme de Denisova : Reich, Gree, Kircher, Krause *et al.*, *Nature* 468: 1053-1060 (2010).
- Les gènes néandertaliens chez l'humain moderne : Reilly, Tjahjadi, Miller, Akey *et al.*, *Current Biology* 32: R970R983 (2022).

13

L'histoire des migrations humaines

Nous avons vu qu'au moins quatre espèces d'hommes ancestraux, *H. erectus*, *H. neanderthalensis*, *H. denisova* et *H. sapiens* étaient sortis d'Afrique à différentes époques pour explorer le monde. Notre espèce l'a fait deux fois. Une première vague de migration est partie d'Afrique il y a environ 100 000 ans. On suppose qu'elle a pris « la route du sud », qui longe l'extrême sud de la péninsule arabique, puis l'Inde, le Vietnam et la Thaïlande, pour atteindre **Sahul***, le plateau continental de terres émergées qui correspond à l'Australie, la Tasmanie et l'actuelle Nouvelle-Guinée, qui à l'époque, était reliées. On retrouve des restes d'*Homo sapiens* tout le long de cette route, datés d'entre 63 000 et 95 000 ans. De plus, un des génomes néandertaliens découvert dans la grotte de Denisova contient de l'ADN d'une population d'*Homo sapiens* plus ancienne que celle qui est sortie d'Afrique il y a 60 000 ans. Ces découvertes pointent vers une première sortie d'Afrique de notre ancêtre 40 000-50 000 ans avant celle qui mènera finalement à la colonisation de toute la planète. On ne retrouve pas l'ADN *Homo sapiens* de cette première sortie chez les humains actuels,

preuve que cette celle-ci sera finalement un échec, au moins d'un point de vue génétique car ces gènes ne se sont pas propagés et seront remplacés par ceux des humains qui viendront plus tard.

La deuxième sortie d'Afrique d'*Homo sapiens* a eu lieu il y a 60 000 ans, soit par la route du sud, soit par ce qui est maintenant le désert du Sinaï (Figure 31). Comme vu précédemment, ces humains se sont reproduits une première fois avec les néandertaliens qui vivaient au Moyen-Orient et leur descendance a migré vers l'Europe et vers l'Asie. Ils arrivent en Europe il y a environ 40 000 ans et très rapidement remplacent la population d'hommes de Néandertal qui vivaient là depuis 400 000 ans. On ne connaît pas la cause de ce remplacement rapide, mais plusieurs hypothèses ont été avancées. *Homo sapiens* a peut-être apporté une maladie infectieuse contre laquelle *Homo neanderthalensis* n'avait aucune (ou peu de) protection, et celle-ci les aurait rapidement décimés. Il est possible que des conflits pour les ressources locales aient débouché sur l'éradication des néandertaliens par les nouveaux arrivants, ou bien que la compétition pour la nourriture ait finalement été gagnée par *Homo sapiens*, sans que cela ait nécessairement débouché sur des affrontements violents. Les nouveaux arrivants étaient juste mieux adaptables à ce nouvel environnement que leurs prédécesseurs. À l'est, les descendants d'*Homo sapiens* et d'*Homo neanderthalensis* arrivèrent en Chine et jusqu'au plateau continental de Sunda, qui à l'époque regroupait les îles indonésiennes de Sumatra, Java, Bali et Bornéo. Là, on trouve les traces d'une deuxième hybridation avec l'homme de Denisova, qui donnera naissance à la population humaine qui vit maintenant en Mélanésie (Figure 31). L'Asie et l'Europe sont maintenant habitées par les *Homo sapiens* « génétiquement modernes » qui finiront par coloniser en quelques milliers d'années le reste de la planète.

LE PEUPEMENT DE L'EUROPE

L'Europe est de loin le continent qui a été le plus étudié, d'un point de vue historique, archéologique mais également génétique. Alors qu'il

avait fallu 13 ans pour séquencer le tout premier génome humain, grâce aux progrès à peine croyables des technologies de séquençage depuis le début du siècle, il est devenu possible de séquencer rapidement des centaines de génomes et d'utiliser ces données pour affiner l'histoire des grandes migrations à l'aube de l'humanité. Grâce à ça, on peut retracer le peuplement de l'Europe par les êtres humains en quatre grandes étapes (Figure 32). Les premiers hommes modernes, descendants des croisements entre *Homo sapiens* et *Homo neanderthalensis* arrivent en Europe depuis le Moyen-Orient il y a environ 43 000 ans. Ils y rencontrent les hommes de Néandertal qui vivent là depuis 400 000 ans. Ces derniers disparaissent, rapidement remplacés par les *H. sapiens*. Puis vient une période de glaciation, qui s'étend de 31 000 ans à 16 000 ans avant l'époque présente. Les glaces descendent très bas, recouvrant la majeure partie de l'Europe du Nord. Le gibier se fait plus rare, la végétation se rabougrit. Les humains se regroupent dans deux régions refuges, l'une centrée sur le sud de la France et le nord de l'Espagne, l'autre autour de la mer Noire et des régions méditerranéennes. La moitié des territoires européens sont dépeuplés et la population humaine est fortement réduite. À la fin de cette période glaciaire, une deuxième expansion se produit à partir de ces zones refuges. Le troisième peuplement européen est celui des agriculteurs venus d'Anatolie. L'agriculture, cette révolution du néolithique, va profondément transformer les sociétés humaines. L'abondance de nourriture qui en découle permet aux populations humaines de croître plus rapidement. Les premiers villages vont pouvoir se former, autour des champs cultivés, l'élevage va prendre son essor avec la domestication des espèces animales. Ces premiers regroupements de population vont entraîner l'apparition des premières épidémies, qui étaient inconnues des humains qui vivaient auparavant en petites communautés familiales. Mais cette époque est définitivement révoquée et les agriculteurs venus de l'est vont supplanter les chasseurs-cueilleurs européens, dont les gènes vont rapidement remplacer ceux de leurs prédécesseurs. Ces agriculteurs arrivent dans la péninsule

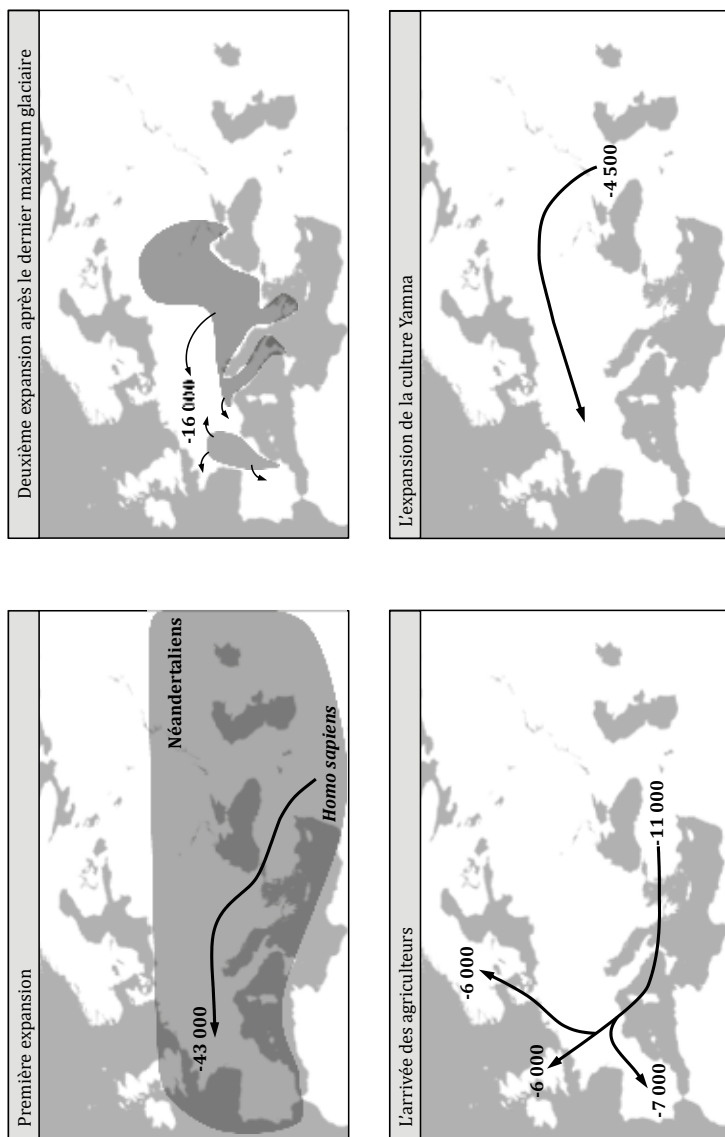


Figure 32 | Le peuplement de l'Europe en quatre étapes. Les dates sont données par rapport à l'époque actuelle.

ibérique il y a 7 000 ans et atteignent l'Angleterre et la Scandinavie il y a 6 000 ans.

La quatrième vague de peuplement de l'Europe commence il y a 4 500 ans avec la migration du peuple Yamnaya originaire d'Europe Centrale, entre la mer Caspienne et la mer Noire, dans la région de la Volga. Vers 3 000 avant notre ère, pendant l'âge du bronze, les populations fermières du Néolithique en Europe de l'Est ont été rapidement remplacées par la culture Yamna. De cette fusion naîtra une nouvelle culture s'étendant du Nord-Est de la France à la Russie en passant par la Scandinavie méridionale, modifiant profondément la structure économique et sociale de cette immense région et donnant naissance à ce qui est appelé la culture de la céramique cordée. L'analyse des séquences génomiques d'Européens ancestraux confirme et complète parfaitement ce scénario migratoire originellement basé sur des preuves archéologiques. De plus, il correspond également à la propagation des langues indo-européennes venues de cette région à la même époque. Le peuplement de l'Europe est maintenant terminé et la séquence des génomes européens modernes reflète et confirme ce scénario.

LE PEUPEMENT DE L'ASIE

L'Asie a subi plusieurs vagues de peuplement. La première sortie des *Homo sapiens* d'Afrique atteignit l'Asie du Sud-Est et Sahul il y a 65 000 ans. Celle-ci ne donnera pas de descendance génétique dans les populations modernes. Le petit nombre de génomes ancestraux asiatiques séquencés à ce jour ne permet pas de reconstituer avec certitude l'histoire du peuplement de l'Asie. Néanmoins, on sait qu'une deuxième vague de peuplement a eu lieu beaucoup plus tard. On a retrouvé dans un génome vieux de 45 000 ans autant d'origine eurasiennes que d'Asie de l'Est, alors qu'un génome provenant de Sibérie et vieux de 36 000 ans était plus proche des Eurasiens. Ces observations suggèrent que les Asiatiques de l'Est et les Eurasiens ont divergé après cette date, il y a donc 36 000-45 000 ans.

Une des vagues suivantes de peuplement date de l'expansion du peuple Yamnaya vers l'est, qui s'est déroulée à la même époque que le peuplement de l'Europe, il y a 4 500 ans. Cette population sera remplacée en Asie Centrale par un autre peuple issu d'Europe de l'Est, la culture de Sintashta, capable d'élever et de dresser des chevaux et de construire des chariots de combat légers. Ce peuple va émigrer à la fin de l'âge du bronze, vers 2 000 avant notre ère, vers l'Asie pour donner naissance aux cultures d'Andronovo, qui vont s'étaler de la mer Noire à l'Altaï et à l'Oural et mélanger leurs gènes à ceux des Asiatiques de l'Est (Figure 33). Celles-ci seront graduellement remplacées par d'autres cultures à partir de la fin de l'âge du bronze (1 500 avant J.-C.) et au cours de l'âge du fer. À nouveau, les données de séquençage de génomes ancestraux valident les preuves archéologiques en permettant de suivre les flux de gènes d'une population à une autre. On voit ainsi la puissance de la génomique moderne, basée sur les séquences de génomes humains plus ou moins anciens, en synergie avec l'histoire ou l'archéologie, jouer un rôle déterminant dans l'interprétation définitive de données historiques ou préhistoriques.

LE PEUPEMENT DES AMÉRIQUES

La séquence de nombreux génomes a établi que les Américains natifs (avant la colonisation européenne) étaient d'origine asiatique. La séparation des ancêtres des Américains et de la population asiatique dont ils descendent s'est produite il y a environ 23 000 ans. Cette population est restée isolée en Sibérie ou près du détroit de Behring jusqu'à ce que les glaces qui recouvraient le nord de l'Amérique fondent, livrant un passage vers l'est, il y a 13 000-15 000 ans. Ces premiers Américains vont descendre le long de la côte ouest pour finalement atteindre le sud du continent. Lors de cette migration, une scission se fait au sein de cette population, une partie d'entre elle migrant vers l'est il y a 13 000 ans (Figure 33).

La présence de gènes d'origine mélanésienne dans le génome d'un Brésilien natif suggère un possible apport par une route alternative à

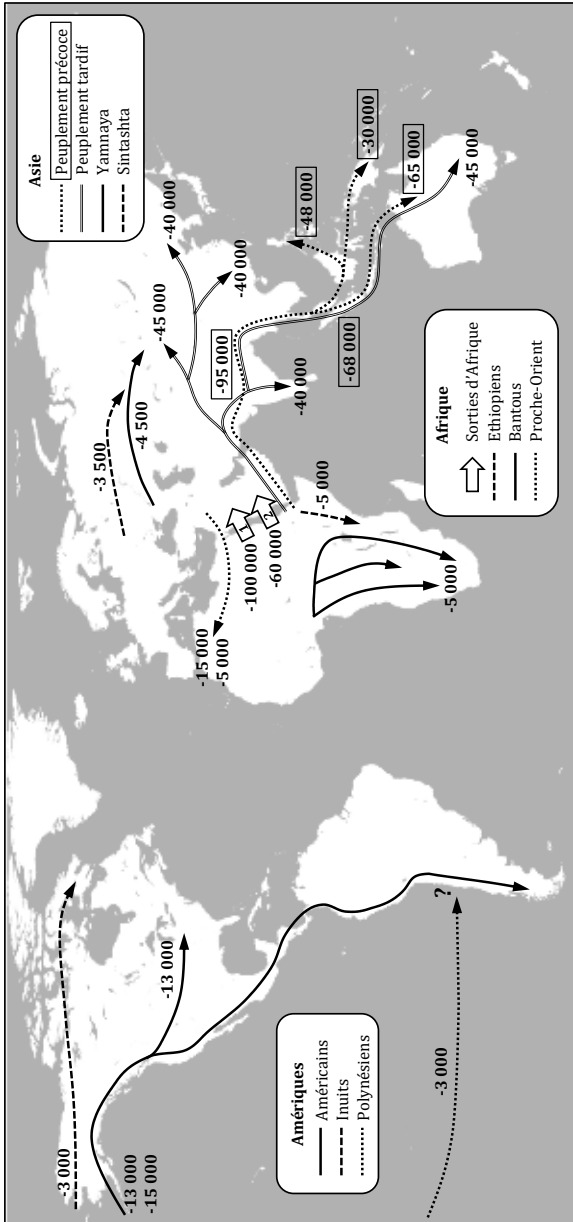


Figure 33 | Le peuplement de la Terre au fil des migrations humaines. Les dates sont données par rapport à l'époque actuelle.

travers l'océan Pacifique, peut-être par un contact précolombien avec des populations polynésiennes. Mais cela reste à confirmer.

Finalement, les Inuits du nord de l'Amérique ont une origine différente des autres américains. Ils ont émigré de Sibérie il y a 3 000-4 000 ans, à travers le détroit de Behring et ont remplacé la population maintenant éteinte d'Esquimaux qui avaient suivi le même chemin 1 000 ans plus tôt.

ET L'AFRIQUE DANS TOUT ÇA ?

De façon peu surprenante, le continent africain recèle la plus grande diversité génétique de la planète. Comme vu précédemment, au moins quatre populations d'hommes ancestraux sont sorties d'Afrique, entre 800 000 et 60 000 ans, *Homo erectus*, *H. neanderthalensis*, *H. denisova* et *H. sapiens*. Mais à l'intérieur du continent africain, des mouvements de peuples eurent également lieu. Le plus important d'entre eux est la dispersion des Bantous d'Afrique de l'Ouest qui a répandu les techniques agricoles dans toute l'Afrique sub-saharienne, il y a 4 000-5 000 ans. D'autres mouvements de peuples africains ont eu lieu ces derniers milliers d'années. Les populations d'éleveurs du sud Soudan se sont déplacées vers l'est il y a environ 7 000 ans et les fermiers-éleveurs éthiopiens ont migré vers le Kenya et la Tanzanie il y a 5 000 ans. On remarque également des flux de gènes en provenance du Proche-Orient dans des individus vieux de 5 000 à 15 000 ans au Maroc, suggérant des échanges entre une population asiatique et cette région d'Afrique (Figure 32).

Finalement, on ne peut que tristement constater que le continent africain, bien qu'il ait engendré toute l'humanité, est celui dont la génétique nous est la plus mal connue. Ceci est dû à plusieurs facteurs. Le premier et non des moindres découle du fait que les chercheurs qui étudient ces questions sont essentiellement d'origine non africaine et que leurs intérêts sont proches de leur origine culturelle et ethnique. D'autre part, les ressources financières pour ce genre de projet sont plus faciles à obtenir pour étudier sa propre culture que pour une

culture étrangère, d'autant plus éloignée qu'elle est Africaine. De surcroît, les restes d'humains ancestraux africains en suffisamment bon état pour être séquencés sont rares. Le climat chaud et humide du continent ne se prête pas à la bonne conservation du matériel génétique, qui se dégrade moins vite dans les climats secs et froids. Alors que le reste de l'humanité descend d'un petit nombre d'ancêtres communs et montre relativement peu de diversité génétique, celle des peuples africains est beaucoup plus importante et nécessiterait de séquencer des dizaines d'individus ancestraux à différents endroits du continent afin d'obtenir la vision globale la plus exacte possible. L'effort scientifique et financier pour combler cette lacune est donc énorme, il ne reste qu'à espérer que le pas pourra être franchi dans un avenir proche.

Après avoir tant parlé d'évolution et de migrations de populations, il est maintenant temps de remonter aux sources des théories évolutionnistes et d'essayer d'explorer l'héritage de Darwin à la lumière de la génomique du XXI^e siècle.

À RETENIR

- *Homo sapiens* est sorti deux fois d'Afrique, il y environ 100 000 ans et 60 000 ans, mais seule la deuxième migration lui a permis de coloniser toute la planète.
- L'Europe a été peuplée par quatre vagues migratoires successives, qui se sont succédées sur environ 40 000 ans.
- Le peuplement de l'Asie est complexe et encore mal caractérisé mais s'est effectué en plusieurs vagues après le peuplement primitif qui n'a pas donné de descendance moderne.
- Les ancêtres des natifs américains sont des peuples d'origine asiatique venus de Sibérie, qui ont franchi le détroit de Behring il y a environ 15 000 ans.
- Un effort de recherche considérable doit être entrepris pour mieux caractériser la diversité génétique des peuples africains ancestraux.

BIBLIOGRAPHIE

- La première sortie d'Afrique d'*Homo sapiens* : Rabett, *Nature Ecology and Evolution* 2: 212-219 (2018).
- Le peuplement de l'Asie : Bae, Douka, Petraglia, *Science* 358: 1269 (2017).
- Le peuplement du monde : Nielsen, Akey, Jakobson, Pritchard *et al.*, *Nature* 541: 302-310 (2017).

14

Lamarck, Darwin, évolution et sélection naturelle

Ce chapitre va remettre en cause un certain nombre de croyances convenues concernant l'évolution des espèces. En effet, on peut affirmer – à l'aune des connaissances en génétique du ^{xxi}e siècle – que Darwin avait à la fois raison... et partiellement tort.

Charles Darwin (1809-1882) était un naturaliste anglais dont le travail bouleversa profondément notre compréhension du monde vivant et dont l'œuvre principale, *L'Origine des espèces*, eut un formidable impact sur la biologie, impact qui continue à se propager à l'heure actuelle sous la forme des différents courants de pensée néodarwiniste. La première édition de cet ouvrage fut publiée en 1859 sous le titre original *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*¹. On attribue souvent à Darwin dans les médias mal informés

1. *Sur l'origine des espèces par le moyen de la sélection naturelle, ou la préservation des races favorisées dans la lutte pour la vie.*

ou au zinc du café des sports la « théorie de l'évolution ». C'est tout simplement faux. L'idée que les espèces animales évoluent au cours du temps en se transformant et que ces transformations se propagent d'une génération à la suivante a été proposée un siècle avant Darwin par un Français, Jean-Baptiste de Lamarck (1744-1829). Dans la pensée occidentale du XVIII^e siècle, l'Univers, la Terre et tous les êtres vivants qui l'habitent ont été créés par Dieu pendant les six jours de la Création. Il est donc inimaginable que les êtres humains, placés au sommet de cette Création puissent avoir un quelconque rapport avec les animaux.

Lamarck se pose en digne successeur du célèbre naturaliste suédois Carl von Linné. C'est un classificateur du vivant, en particulier des animaux. À cette époque, ceux-ci sont classés en fonction de leurs caractères morphologiques ou physiologiques. Au sommet de l'échelle des animaux se trouvent les mammifères, puis les oiseaux, puis les reptiles, puis les poissons. Ces quatre classes forment les vertébrés, qui sont au-dessus des invertébrés. Lamarck se spécialise dans la zoologie de ces derniers et il l'enseigne au Museum d'Histoire naturelle de Paris entre 1794 et 1807. Dans sa *Philosophie zoologique*, publiée en 1809, il présente sa nouvelle classification des invertébrés. Pour lui, les animaux se complexifient graduellement, sur un temps extrêmement long, qui ne permet pas de rendre ces transformations observables à l'échelle d'une (ou de plusieurs) vie(s) humaine(s). Pour cette raison, les êtres vivants nous apparaissent immuables. L'apport de Lamarck à la pensée scientifique de son époque est immense. C'est lui qui crée le mot « biologie » à partir du grec *bios*, vie, et *logos* « science », faisant ainsi – enfin – entrer la biologie dans le domaine des sciences au même titre que la physique, la chimie ou les mathématiques. Il soutient que cette complexification des êtres vivants s'applique également aux êtres humains, hypothèse iconoclaste et Ô combien hérétique pour l'époque ! En réalité, *Homo sapiens* ne descend pas d'un singe, mais les singes actuels et les humains ont un ancêtre commun. Il est donc plus correct de dire que l'homme et les singes modernes

sont de proches cousins du monde animal. Cette théorie du *transformisme* inspira d'ailleurs profondément la pensée scientifique et sociale du siècle des Lumières. Elle trouvera néanmoins une opposition de la part de certains naturalistes, qui interprétant à tort la pensée Lamarckienne, y verront une volonté propre des êtres vivants à se transformer. Finalement, c'est Darwin qui fera triompher le *transformisme* (sans jamais le nommer comme ça) en lui donnant une explication scientifique rationnelle, basée sur la théorie de la sélection naturelle des espèces les plus adaptées. À partir de là, le transformisme va quitter le cercle restreint des naturalistes et des académiciens pour gagner d'abord les salons parisiens puis le reste de la société à partir de la seconde moitié du XIX^e siècle.

Avant d'explorer dans le détail l'œuvre de Darwin et sa filiation moderne, il est bon de faire une digression rapide afin d'explicitier la notion de théorie. En français comme en anglais, le mot théorie a plusieurs définitions. Son tout premier sens est « ensemble organisé de principes, de règles, de lois scientifiques visant à décrire et à expliquer un ensemble de faits ». On comprend donc bien que la théorie – et tout particulièrement la théorie scientifique car c'est bien de sciences que l'on parle – n'a rien de spéculatif, mais repose au contraire sur des lois et des règles établies par l'approche scientifique. On peut ici en donner une définition non ambiguë. Une théorie scientifique est l'explication d'un phénomène naturel, testée et confirmée de façon reproductible par des expérimentateurs indépendants utilisant un protocole expérimental prédéfini. Une seconde définition du mot « théorie » trouvée dans le même dictionnaire est « connaissance purement spéculative ». C'est hélas l'acceptation plus couramment utilisée dans le langage vernaculaire et la raison de confusions profondes dans les tentatives de vulgarisation ou de contempation de l'œuvre de Darwin. L'emploi du mot « théorie » dans un sens purement spéculatif doit être qualifié d'hypothèse, « une proposition visant à fournir une explication vraisemblable d'un ensemble de faits, qui doit être soumise au contrôle de

l'expérience ou vérifiée dans ses conséquences ». On comprend donc bien la différence entre une théorie et une hypothèse. Une théorie est une explication testée et confirmée par plusieurs expérimentateurs utilisant la méthode scientifique, alors qu'une hypothèse doit encore être soumise à l'expérimentation pour être validée (ou invalidée). C'est bien dans sa première définition que nous avons utilisé le mot « théorie » jusqu'à présent et que nous l'utiliserons jusqu'à la fin de cet ouvrage.

L'ORIGINE DES ESPÈCES

Charles Darwin était un extraordinaire observateur de la nature. Le voyage d'exploration initialement géologique et cartographique du *HMS Beagle*, prévu pour durer deux ans, en pris finalement cinq. Au cours de ce voyage, Darwin passa la plupart de son temps à visiter les terres accostées et à prendre des notes sur les espèces vivantes rencontrées. Ce sont ces notes qui alimenteront sa réflexion et qui l'amèneront aux conclusions qui sont détaillées dans *On the Origin of Species*. On peut essayer de résumer la pensée de Darwin aux principes suivants :

- (i) Les espèces animales et végétales domestiques montrent une grande variabilité morphologique et physiologique. La nature fournit les variations successives de ces espèces, mais c'est l'homme qui sélectionne celles qui lui sont utiles. Cette sélection « *permet à l'agriculteur non seulement de modifier le caractère de son troupeau, mais de le transformer entièrement*². »
- (ii) Dans la nature, la variabilité au sein d'une espèce est à l'origine des différentes variétés de cette espèce. Ce sont les espèces dominantes, c'est-à-dire les plus florissantes, qui produisent le plus grand nombre de variétés. Ainsi, les formes dominantes ont tendance à produire encore plus de formes qui vont devenir dominantes.

2. Darwin, *L'Origine des espèces*, GF Flammarion (1992). Tous les passages cités entre guillemets dans ce chapitre sont tirés de cet ouvrage.

- (iii) Les ressources et l'espace disponibles sur Terre étant limités, tous les êtres vivants – y compris l'homme – sont engagés dans une lutte pour l'existence. Les individus ayant survécu vont pouvoir transmettre à leur descendance les caractères ayant assuré leur survie, leur donnant à leur tour une plus grande chance de survie. Darwin donne à ce principe « *en vertu duquel une variation si insignifiante qu'elle soit se conserve et se perpétue, si elle est utile, le nom de sélection naturelle, pour indiquer les rapports de cette sélection avec celle que l'homme peut accomplir.* »
- (iv) Darwin fait la distinction entre variations avantageuses, variations nuisibles et variations neutres, ces dernières d'après lui n'affectant pas la sélection naturelle. C'est l'accumulation des variations avantageuses qui va éventuellement permettre aux individus de remporter la lutte pour leur survie. La sélection sexuelle est un cas particulier de la sélection naturelle, dans laquelle le mâle dominant va se reproduire au détriment des autres mâles. Dans ce cas, ce sera le mâle le plus vigoureux qui pourra se reproduire et transmettre sa vigueur à ses descendants.
- (v) L'accumulation des variations avantageuses au cours des générations va finir par modifier la morphologie de l'organisme, telle que l'on aura affaire à une nouvelle *variété*. Les nouvelles variétés obtenues vont continuer à accumuler des variations avantageuses qui vont entraîner l'apparition de caractères de plus en plus divergents. Finalement, au bout de milliers de générations, une ou plusieurs variétés auront tellement divergé qu'elles seront devenues de nouvelles *espèces*.

Darwin représente l'accumulation des variations reliant les individus les uns aux autres sous forme de plusieurs arbres, à la manière d'arbres généalogiques dont certaines branches s'éteignent et d'autres se ramifient. Cette façon de représenter l'évolution sous forme d'arbres est toujours utilisée à l'heure actuelle bien que peu représentative de la réalité biologique. Pour Darwin, comme pour les zoologistes du ^{xx}e siècle, les espèces sont définies par la

barrière reproductive. Deux animaux appartiennent à des espèces différentes s'ils ne peuvent pas se reproduire entre eux ou bien si leur descendance est stérile, comme le mulet, produit d'un âne et d'une jument. Mais les données les plus récentes de séquençage des génomes révèlent une image très différente du vivant. Nous avons vu précédemment que les hommes modernes sont des hybrides de deux (ou trois) espèces d'hommes archaïques, dont la descendance a peuplé le monde. De nombreux exemples d'**hybridations interspécifiques*** existent dans le monde animal comme dans le monde végétal, comme nous le verrons au chapitre suivant. La vision actuelle de l'évolution des espèces ressemble plutôt à un arbre dont certaines branches séparées (des espèces différentes) sont reliées par des hybridations qui permettent le transfert de gènes d'une branche à l'autre. L'arbre darwinien est devenu un réseau d'interactions génétiques plus ou moins finement maillé en fonction des branches du vivant (Figure 34).

Hybridation interspécifique : Accouplement entre espèces différentes dont la descendance est fertile.

Avant d'étudier les limites de la théorie darwinienne et l'apport de la génétique moderne à cette théorie, il est bon de rappeler que lorsqu'il écrit *On the Origin of Species*, le naturaliste anglais n'a pas la compréhension de la nature des variations qui s'accumulent chez les êtres vivants, ces variations que les généticiens qui le suivront nommeront plus tard *mutations*. La génétique elle-même n'existe pas encore car ses premières lois seront écrites quelques années plus tard par un moine augustin d'origine tchèque encore inconnu, Johann Gregor Mendel (1822-1884), qui deviendra le père de la génétique, c'est-à-dire de la science qui étudie les gènes et leur transmission héréditaire. En fait, il est possible que Darwin ait eu connaissance des expériences de Mendel, car deux ouvrages en allemand qui y faisaient brièvement référence ont été retrouvés dans sa

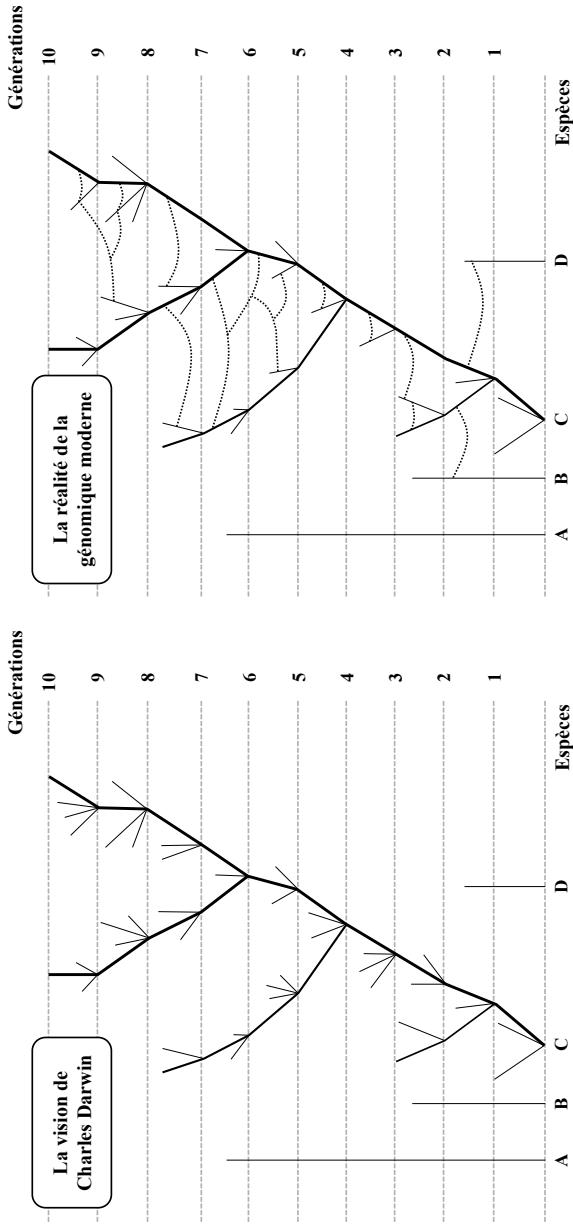


Figure 34 | L'évolution des espèces. Gauche : d'après Charles Darwin. Droite : D'après les données récentes de séquençage des génomes.

bibliothèque³. Néanmoins, il semblerait qu'il n'ait pas compris le lien qui existait entre ses propres observations et celles de Mendel et que les caractères héréditaires décrits par le moine tchèque correspondaient aux variations qu'il avait observées entre différentes variétés d'êtres vivants. Il faudra attendre l'aube du xx^e siècle pour concilier les travaux des deux scientifiques et en faire ce que l'on appelle généralement la théorie synthétique de l'évolution, qui jettera les bases du *néodarwinisme*.

LES LIMITES DE LA THÉORIE DARWINIENNE

La découverte des lois de la génétique et surtout celle de la nature moléculaire des gènes vont transformer radicalement et durablement la biologie, qui va passer d'une description et d'une classification précise des êtres vivants par les naturalistes du xix^e siècle à une compréhension mécanistique très fine des mécanismes à l'origine des modifications du vivant par les généticiens moléculaires du xx^e siècle. Le néodarwinisme mettra un nom sur les *variations* proposées par Darwin, qui sont les mutations pouvant toucher un ou plusieurs gènes portés par la molécule d'ADN. On fera le lien entre le **génotype*** d'un individu, c'est-à-dire l'ensemble des allèles de son génome et son **phénotype***, qui est l'ensemble des caractères observables d'un individu et qui représente l'expression du génotype.

Génotype : Ensemble de tous les allèles des gènes d'un être vivant.

Phénotype : Ensemble des caractères observables d'un être vivant. Le phénotype résulte de l'expression du génotype.

Ainsi, les variations darwiniennes (mutations du génome) seront liées à la morphologie des variétés observées par Darwin (phénotype). Les arbres dressés par la phylogénie moléculaire reprendront

3. Lettre de l'éditeur Andrew Sclater, dans *The Correspondence of Charles Darwin*, University Library Cambridge (2000).

l'idée originelle des arbres darwiniens. Finalement, la théorie de la sélection naturelle, pierre fondatrice de l'édifice darwinien, sera confirmée par d'innombrables expériences dans d'innombrables laboratoires à travers le monde et n'a jamais pu être mise en défaut jusqu'à présent. Une des expériences les plus élégantes qui démontre de façon limpide comment la sélection naturelle joue sur des mutations survenant de façon aléatoire dans une population, est celle de Salvador Luria (1912-1991), microbiologiste italo-américain et Max Delbrück (1906-1981), biophysicien germano-américain, qui obtinrent le prix Nobel de Physiologie ou Médecine en 1969. Cette expérience, si simple que n'importe quel étudiant de premier cycle universitaire peut la mettre en œuvre, utilise la propriété d'un phage à infecter et à tuer une population bactérienne. Les auteurs montrent que les mutations présentes dans la population bactérienne et qui les rendent résistantes au phage sont aléatoires et préexistantes à l'addition du phage. Ce n'est donc pas l'agent sélectif (le phage) qui induit l'apparition des mutations. Notons que cette expérience peut être reproduite à l'identique avec une population bactérienne et un antibiotique, expliquant l'apparition dans une population bactérienne de bactéries résistantes aux antibiotiques couramment utilisés. Ce que Darwin avait proposé pour les organismes complexes était vrai pour les bactéries, consacrant définitivement la sélection naturelle et le néodarwinisme triomphant.

Néanmoins, le naturaliste notait lui-même que sa théorie se heurtait à un certain nombre d'observations inattendues qu'il avait du mal à expliquer.

- (i) L'apparition soudaine de groupes entiers d'espèces, comme en témoigne l'abondance des fossiles du Cambrien, est pour lui un problème : « *Si des espèces nombreuses appartenant aux mêmes genres ou aux mêmes familles étaient réellement apparues tout à coup, ce fait anéantirait la théorie de l'évolution par la sélection naturelle.* » En effet, en agrément avec Lamarck, il considèrerait que la sélection d'un ensemble de formes descendantes d'un ancêtre commun

devait être fort long et forcément incompatible avec une explosion d'espèces dans un temps court, même à l'échelle géologique.

- (ii) L'absence de formes intermédiaires n'était pas compatible avec la théorie, qui postule que les améliorations sont incrémentales et non saltatoires. *Natura non facit saltum*, la nature ne fait pas de sauts, une application du principe de continuité qui suggère que les choses naturelles évoluent lentement, graduellement. L'absence de certaines formes intermédiaires dans les organismes vivants a mené plus tard à l'idée populaire du « chaînon manquant » dans l'évolution. Darwin contourne ce problème en mettant en cause l'insuffisance des archives géologiques qui ne contiennent pas les fossiles de tous les organismes ayant vécu à toutes les époques passées. Nous verrons qu'il y a d'autres explications à cette absence de formes intermédiaires qui n'est en fait qu'apparente.
- (iii) La dispersion géographique d'espèces très similaires pose également un problème à Darwin, qui ne peut pas concevoir – comme les autres savants de son époque – la dérive des continents et la possibilité qu'à une époque reculée toutes les terres émergées aient pu faire partie d'un immense et unique ensemble. Dans la réflexion scientifique du XIX^e siècle, encore imprégnée de pensée judéo-chrétienne, les continents sont immuables et ont été créés tels qu'on peut encore les observer aujourd'hui. Même si certains scientifiques avaient précédemment noté la complémentarité des côtes africaine et américaine, il faudra attendre le début du XX^e siècle pour que le météorologue et astronome Alfred Wegener (1880-1930) invente le terme « dérive des continents » (*Kontinentalverschiebung*), développe cette hypothèse à partir de données géologiques et géophysiques, et la popularise. Il faudra encore quelques décennies pour que naisse la théorie de la tectonique des plaques, qui est maintenant le modèle couramment accepté de formation et de dérive des plaques continentales. Les continents tels que nous les connaissons sont loin d'avoir toujours eu le tracé actuel. Il y a 1,83 milliards d'années existaient plusieurs continents dont un super continent,

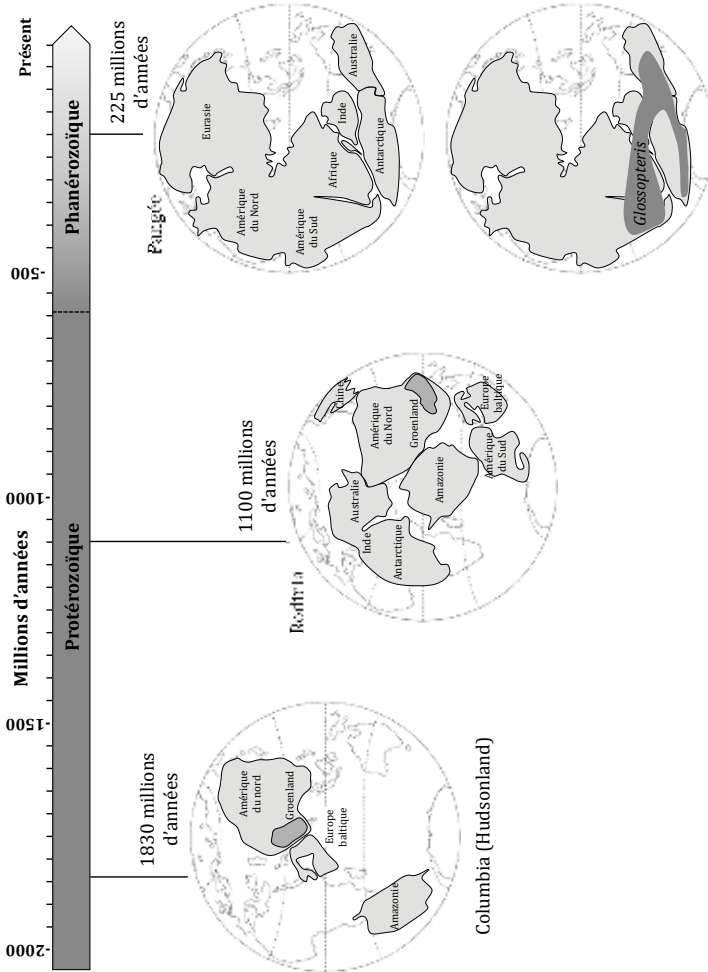


Figure 35 | La dérive des continents. La géographie des supercontinents à trois époques différentes est représentée superposée à une partie de la mappemonde moderne. L'aire géographique où des fossiles de *Glossopteris* ont été retrouvés est indiquée.

Colombia (ou Hudsonland), qui regroupait l'Amérique du Nord, le Groenland, le nord de l'Eurasie et la Scandinavie, et l'Amazonie. Le mouvement de ces différentes plaques au cours de centaines de millions d'années suivantes allait donner naissance à un autre supercontinent appelé Rodinia. Finalement, il y a environ 300 millions d'années, peu de temps avant les premiers dinosaures, émergea un autre supercontinent, la Pangée, qui allait donner naissance à toutes les terres actuelles (Figure 35).

La découverte qu'un genre de plantes disparues, *Glossopteris*, était retrouvé sous forme de fossiles en Inde, en Amérique du Sud, en Afrique, dans l'Antarctique et en Australie prouvait que ces continents n'en faisaient qu'un à l'époque où poussait cette plante (Figure 35). C'était la confirmation que l'hypothèse de Wegener était correcte et celle-ci fut par la suite confirmée par des données géographiques, géologiques et tectoniques. Si Darwin avait eu suffisamment confiance en sa théorie, au lieu de tenir cet argument comme une limite à son modèle, il serait arrivé à la conclusion logique que des espèces apparemment isolées géographiquement avaient dû à une époque lointaine partager le même espace. En d'autres termes, il aurait pu découvrir la dérive des continents avec un demi-siècle d'avance ! Mais outre ces questions soulevées avec lucidité par Darwin lui-même, le néodarwinisme allait être confronté à un nouvel obstacle de taille, avec la théorie de l'évolution moléculaire par dérive aléatoire appelée encore « théorie neutraliste de l'évolution ».

LA THÉORIE NEUTRALISTE

Motoo Kimura (1924-1994), théoricien de l'évolution japonais, publie en 1968 une étude qui allait jeter un énorme pavé dans la mare néodarwiniste. Les techniques de séquençage de l'ADN n'avaient pas encore été inventées, mais on savait déjà déterminer la séquence de petits bouts de protéines. Kimura compare les séquences protéiques des hémoglobines de plusieurs mammifères et rapporte les différences observées à la durée de temps écoulée depuis

la divergence entre ces espèces. Ses calculs le conduisent au résultat inattendu que la fixation d'une mutation dans une population se produit en moyenne tous les deux ans, ce qui serait bien sûr un intolérable fardeau génétique pour tout être vivant si ne serait-ce qu'une partie de ces mutations était délétère. Il en conclut donc que la grande majorité de ces mutations sont neutres, c'est-à-dire qu'elles ne modifient pas la séquence protéique. Cela est possible car la dégénérescence du code génétique fait qu'environ les trois quarts des remplacements d'une base par une autre sur la troisième base du codon ne provoquent pas de changement d'acide aminé dans la protéine correspondante et sont donc neutres (Tableau 1). Grâce au séquençage de génomes entiers, il a pu être vérifié que les mutations dans les gènes codant des protéines sont le plus souvent des mutations sur la troisième base du codon, confirmant la neutralité de la plupart des mutations fixées dans une population.

Mais une mutation ne va pas tomber systématiquement dans une région codante. En effet, les gènes eucaryotes sont morcelés par la présence de nombreux introns. Une mutation dans un intron sera la plupart du temps neutre, car elle ne sera pas traduite en protéine. Comme vu précédemment, seulement environ 2 % du génome humain est couvert par des exons qui vont être traduits en protéines. Une mutation se produisant dans les 98 % d'ADN restants sera donc neutre. Darwin, avec une remarquable intuition, avait supposé l'existence de « variations neutres » qui n'affectaient pas la sélection naturelle, mais dans son modèle ces variations devaient être minoritaires pour expliquer l'effet de la sélection naturelle. La découverte que la grande majorité des mutations sont neutres va poser un sérieux problème aux évolutionnistes, car la théorie neutraliste semble s'opposer à celle de la sélection naturelle. Si on les a effectivement souvent opposées, ces deux théories sont non seulement parfaitement compatibles, mais peuvent être facilement réconciliées, comme nous allons le voir.

Mais si la sélection naturelle n'est pas le moteur principal de l'évolution, selon les neutralistes, quel est-il donc ? Dans ce modèle, c'est

la **dérive génétique*** qui explique la diversité. La dérive est l'ensemble des phénomènes aléatoires qui vont mener à la modification de la fréquence d'un allèle donné dans une population (par exemple les yeux bleus), par échantillonnage trop faible de cet allèle dans la population.

Dérive génétique : Évolution des différents allèles des gènes d'une population en fonction d'événements stochastiques indépendants de l'apparition de nouvelles mutations ou de sélection naturelle. Une épidémie, une catastrophe naturelle, ou encore des taux de reproduction différents entre les individus sont des moteurs de la dérive génétique.

Concrètement, la dérive génétique s'exerce d'autant plus fortement que la taille de la population est réduite. En effet, dans une population de faible importance, la transmission des allèles n'est plus aléatoire et dépend du taux de reproduction de chaque individu. Si un individu portant un allèle rare se reproduit beaucoup plus que des individus portant un allèle fréquent, l'allèle rare va se disséminer dans la population alors que l'allèle fréquent va tendre à disparaître. En revanche, si la population est suffisamment importante, la fréquence de chaque allèle va rester constante au sein de la population, au cours des générations. Pour être valide, la théorie neutraliste nécessite :

- un taux de mutation constant ;
- une taille de population également constante et pas trop élevée ;
- que les allèles perdus par la dérive soient à peu près en nombre égal aux nouveaux allèles produits par mutation. C'est ce qu'on appelle l'équilibre mutation-dérive.

Il n'est donc pas évident que la théorie neutraliste puisse s'appliquer à toutes les époques et dans toutes les populations d'êtres vivants, la taille de ces populations n'étant pas forcément constante aux échelles de temps géologiques considérés. De plus, il n'est pas certain que le taux de mutation soit lui-même constant dans le temps,

en particulier à des époques lointaines où la concentration d'oxygène était différente. En effet, le taux atmosphérique d'oxygène va influencer sur les propriétés oxydatives des cellules en modifiant la quantité de réactifs oxygénés libres, réactifs qui sont fortement mutagènes et cause de mort cellulaire par stress oxydatif. L'augmentation d'oxygène atmosphérique va donc à la fois augmenter la mortalité, donc réduire la taille des populations (augmentant l'effet de dérive) et en même temps augmenter le taux de mutations chez les individus survivants.

Bien qu'en apparence contradictoire, la théorie neutraliste est parfaitement compatible avec la sélection naturelle, qu'elle complète. On peut essayer de faire une synthèse de ces deux théories de la façon suivante. La plupart des mutations accumulées au cours de l'évolution humaine se produisent dans des régions du génome où elles sont neutres. Sur des populations de petite taille, l'effet de dérive est donc important. La plupart des mutations qui se produisent dans des régions codantes (2 % du génome) sont délétères et donc soumises à une forte contrainte. Elles subissent une sélection purificatrice qui tend à les éliminer du génome. Ce sont les « variations nuisibles » de Darwin, qui contre-sélectionnent les gènes concernés en éliminant les individus les portant. Finalement, les mutations qui vont modifier la séquence des protéines en changeant leur fonction, celles que le naturaliste présente comme des « variations avantageuses », sont très rares et spécifiques de certains gènes qui subissent une très forte pression de sélection pour évoluer rapidement. Un des exemples les mieux étudiés est celui de la réponse immunitaire à une infection. Dans une population exposée à de nombreux pathogènes, les mutations survenant dans le système immunitaire vont permettre de répondre plus rapidement et plus efficacement à une infection. Un autre exemple que nous avons évoqué plus tôt est celui de l'adaptation à l'altitude des populations tibétaines grâce à la sélection d'un gène de dénisovien. Il s'agit encore de sélection positive d'un allèle favorable dans une population soumise à un stress important. Mais

ce type de sélection positive darwinienne est très rare et sous-tend des mécanismes biologiques bien précis.

En résumé, la plupart des mutations dans des régions non codantes sont neutres. La plupart des mutations dans des régions codantes sont délétères et sont donc éliminées par la sélection naturelle, la majorité des autres sont neutres. Une toute petite fraction des mutations dans des régions codantes très spécifiques va amener effectivement un avantage sélectif, mais ce n'est pas le principal mécanisme intervenant dans la formation de nouvelles espèces. La Figure 36 décrit ces différents mécanismes. À partir d'une population initiale que l'on suppose homogène par souci de simplification, des mutations aléatoires vont se produire dans chaque organisme. Certaines de ces mutations vont être délétères à plus ou moins long terme (ronds noirs), d'autres vont être neutres (ronds gris) et certaines vont procurer un avantage sélectif dans certaines conditions de vie bien particulières (carrés). La sélection purificatrice (à gauche) va éliminer au fil des générations les mutations délétères, alors que la sélection positive (à droite) va permettre aux organismes ayant reçu une mutation avantageuse de se reproduire mieux et donc de se propager dans la population. La dérive génétique (au milieu) va agir aléatoirement sur les différents allèles, en éliminant certains et en amplifiant d'autres de façon stochastique. L'effet conjugué de ces trois forces au fil des générations va mener à la structure de la population finale : les mutations délétères ont été éliminées, certaines mutations neutres ont été conservées et une mutation avantageuse a été sélectionnée. Aucune de ces mutations n'est fixée dans la population finale car aucune ne l'a complètement envahie. Cette population va donc continuer à évoluer en fonction des forces qui vont s'exercer sur elle (flèche pointillée). Il faut noter qu'une seule des nombreuses populations finales possibles a été représentée.

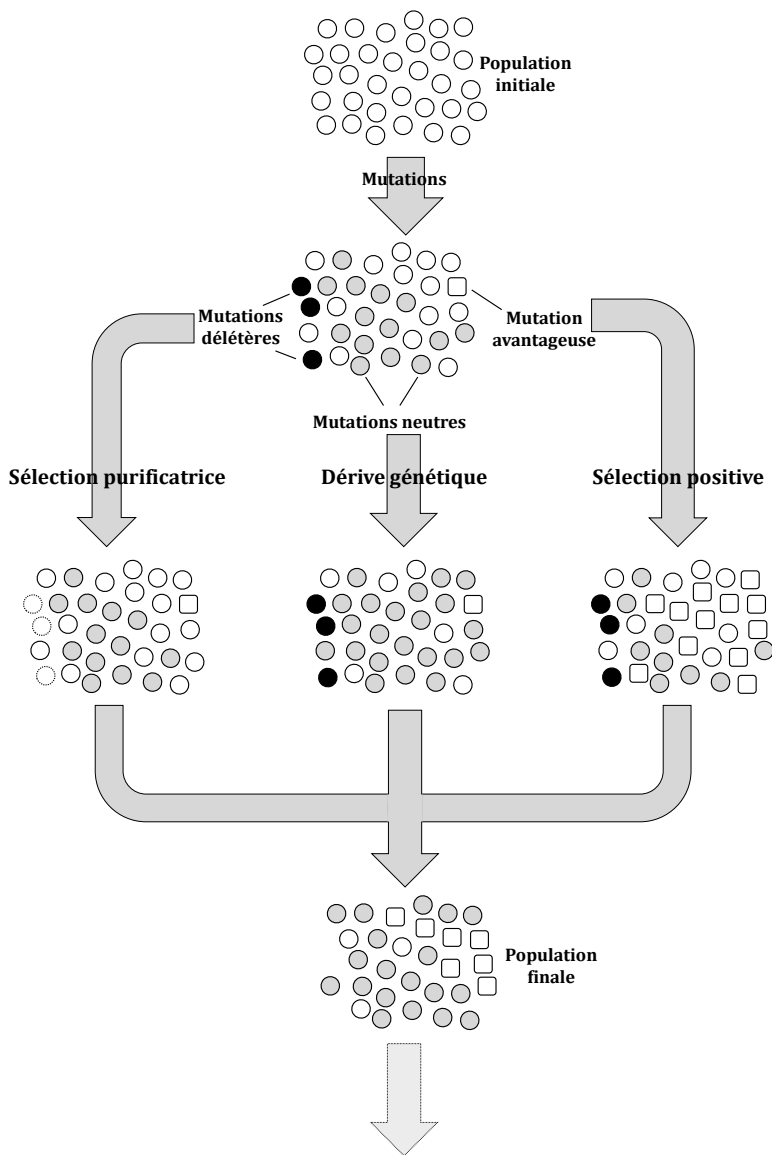


Figure 36 | Sélection purificatrice, dérive génétique et sélection positive.

LES GRANDES EXTINCTIONS DE MASSE

Si la plupart des mutations ne présentent aucun avantage ou désavantage sélectif, on peut se demander alors comment elles se propagent. Il y a plusieurs mécanismes connus, qui peuvent créer ce qu'on appelle un **goulet d'étranglement*** et vont augmenter ou diminuer la fréquence de certains allèles dans une population, par dérive génétique. Un navire faisant naufrage sur une île solitaire et loin des routes maritimes, l'émigration massive d'une population vers un nouveau continent, ou bien une catastrophe naturelle tuant aléatoirement la majeure partie d'un peuple sont autant d'exemples de dérive génétique qui vont faire disparaître

Goulet d'étranglement : Réduction drastique du nombre d'allèles présents dans une population suite à l'élimination d'une grande partie de celle-ci. Appelé *population bottleneck* en anglais. Le terme anglais est souvent utilisé, même dans des textes français. En français, on utilise aussi l'appellation de goulot d'étranglement, qui rappelle le goulot de la bouteille (*bottleneck*).

certaines gènes en augmentant la fréquence d'autres gènes, au sein d'une population soumise à cette dérive. Les grandes extinctions de masse font partie de ces événements catastrophiques qui modelèrent la structure des populations animales et végétales au cours des éons.

La première extinction de masse a sans doute été celle de la Grande Oxydation survenue il y a 2,4 milliards d'années, qui a dû tuer 99 % des organismes vivant à cette époque. Nous n'avons hélas aucune trace de cette époque lointaine et nous n'en aurons sans doute jamais car les cellules anaérobies qui y vivaient n'ont pas laissé de fossiles. En revanche, il existe de nombreuses traces fossilisées d'organismes ayant disparu plus récemment et l'étude de leur répartition dans les différentes couches sédimentaires a amené à la conclusion que les êtres vivants à la surface du globe ont traversé cinq extinctions massives d'espèces depuis le Cambrien, il y a 500 millions d'années, c'est-à-dire la période d'apparition des premiers chordés. D'après la comparaison

entre espèces fossiles marines disparues, les espèces vivantes disparaissent à un rythme inférieur ou égal à huit espèces par million d'années. La vitesse réelle de disparition de toutes les espèces vivantes est difficilement extrapolable à partir des seuls fossiles marins mais on se base sur cette observation pour comparer le rythme de disparition en fonction des époques. Celui-ci a décru régulièrement jusqu'à atteindre deux espèces par million d'années au début de notre ère. En période d'extinction de masse, ce niveau de disparition peut atteindre vingt espèces par million d'années.

La première de ces cinq extinctions eut lieu pendant l'Ordovicien (– 445 millions d'années) et entraîna la disparition de la moitié des genres et de 85 % des espèces vivantes. C'est la deuxième plus large extinction en nombre d'espèces. Elle fut provoquée par une grande glaciation qui entraîna des bouleversements profonds dans les écosystèmes (Figure 37).

La seconde extinction se produisit pendant le Dévonien (– 370 millions d'années), tuant 47 % des genres animaux. Les causes exactes de cette extinction sont encore débattues, mais il semblerait que des changements climatiques successifs l'aient provoquée.

La troisième extinction fut la plus terrible. Elle se produisit à la fin du Permien (– 250 millions d'années) et faillit faire entièrement disparaître la vie sur Terre. Les océans perdirent 80-96 % des espèces, dont les fameux trilobites. Sur Terre, la flore dominante des *Glossopteris*, des sortes de fougères à graines évoquées un peu plus tôt, fut entièrement remplacée, huit ordres d'insectes disparurent ainsi que les deux tiers des tétrapodes. Son origine est une gigantesque éruption de lave basaltique qui a formé les *trapps*⁴ de Sibérie, faisant remonter à la surface des millions de kilomètres cubes de lave pendant environ un million d'années, changeant ainsi de façon dramatique l'écosystème terrestre dans son ensemble.

4. Le mot *trapp* vient du suédois qui signifie « escalier », en référence aux paysages à l'aspect de larges marches de cette région.

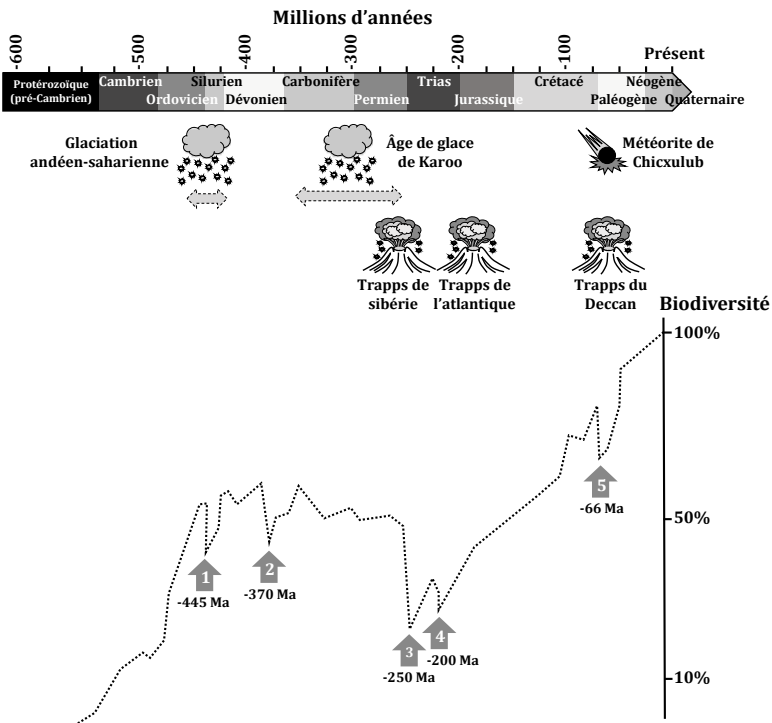


Figure 37 | Les grandes extinctions de masse. Sont indiquées les deux périodes de refroidissement, les grandes éruptions volcaniques et l'impact de la météorite de Chicxulub. Les cinq grandes extinctions sont : (1) extinction de l'Ordovicien ; (2) extinction du Dévonien ; (3) extinction du Permien ; (4) extinction du Trias ; (5) extinction du Crétacé, avec leurs âges respectifs (Ma : millions d'années). Augmentation et perte de biodiversité sont représentées en pourcentage par rapport à l'époque actuelle.

La quatrième extinction est celle du Trias (– 200 millions d'années) qui voit disparaître 40 % des genres d'animaux marins et les derniers des grands amphibiens. Elle fut provoquée par les gigantesques écoulements de lave d'une région magmatique semblable à celle des *trapps* sibériens, la province magmatique centre Atlantique qui se situait au centre de l'ancien continent de la Pangée et dont l'éruption a entraîné la séparation de l'Afrique et de l'Amérique et formé l'océan Atlantique. Cette quatrième extinction laisse le champ

libre aux dinosaures qui vont alors régner sur la faune terrestre, dans presque toutes les niches écologiques, pendant 134 millions d'années.

La cinquième grande extinction eut lieu au Crétacé (– 66 millions d'années), tuant 39 % des genres animaux, dont la plupart des dinosaures, à l'exception de ceux qui vont donner naissance aux crocodiles et aux oiseaux. Il y a deux origines possibles à cette catastrophe, les deux n'étant d'ailleurs pas exclusives. La formation des trapps du Deccan, à l'ouest de l'Inde, date de cette époque. Le relargage massif de gaz volcaniques comme le dioxyde de soufre aurait fait baisser la température globale de quelques degrés en réfléchissant les rayons du soleil, déclenchant la fin prématurée des dinosaures. La deuxième hypothèse repose sur la découverte d'un cratère de 180 kilomètres de diamètre et de 20 kilomètres de profondeur, sur la péninsule du Yucatán au Mexique. Le cratère de Chicxulub a été provoqué par l'impact d'une météorite de 10 kilomètres de diamètre, il y a 66 millions d'années. L'effet d'un tel impact nous est difficile à imaginer. Son énergie représente un milliard de fois celle dégagée par les bombes atomiques d'Hiroshima et Nagasaki. Les matériaux enfouis à des kilomètres sous terre sont projetés presque instantanément en altitude, à des chaleurs incandescentes, augmentant de plusieurs centaines de degrés la température de l'atmosphère à leur contact. Les matériaux les plus lourds (comme le quartz) retombent rapidement sur terre provoquant de gigantesques incendies, alors que les cendres plus légères forment un immense nuage dans la haute atmosphère qui enveloppe la planète entière, la privant des rayonnements solaires pendant plusieurs années. Les espèces végétales dépendantes de la photosynthèse meurent les premières, suivies par les herbivores, puis par les carnivores. C'est l'ensemble de la chaîne alimentaire qui s'effondre. Depuis la découverte de l'impact du cratère de Chicxulub, ce scénario était considéré comme le plus vraisemblable. Néanmoins, certaines études suggèrent que la quantité de débris projetés dans l'atmosphère était trop faible pour provoquer un hiver de cette ampleur. En revanche, les gaz soufrés relargués par l'impact sont

plus vraisemblablement à l'origine de l'extinction. Ces gaz formant des aérosols soufrés dans la stratosphère, entre 10 et 50 kilomètres d'altitude, bloquent les rayons du soleil, abaissent la température globale et empêchent la photosynthèse. La température moyenne de surface aurait baissé de 27 °C trois ans après l'impact, l'amenant en dessous de zéro pendant les trente années suivantes. Dans certaines régions continentales, cette diminution a pu atteindre 40-50 °C, alors que la température des océans diminuait dans le même temps d'une quinzaine de degrés. Les dinosaures seraient donc morts de froid avant de mourir de faim ! Il est également possible que le volcanisme géant du Deccan ait pu être déclenché par l'impact de la météorite de Chicxulub, infligeant aux pauvres dinosaures une double peine qui leur sera fatale. En dehors des restes – parfois monumentaux – fossilisés de leur squelette, on sait peu de choses de ses créatures géantes du Crétacé. Une des rares certitudes est que leur corps était recouvert de plumes et que les dinosaures volants comme l'archéoptéryx sont les ancêtres de nos oiseaux modernes. À cette spectaculaire extinction n'ont survécu que les petits mammifères, qui avec la disparition des dinosaures, vont maintenant pouvoir envahir la Terre et se diversifier.

LA SIXIÈME EXTINCTION

La dernière extinction, celle de l'Holocène a débuté il y a environ 13 000 ans, quand l'homme a commencé à coloniser la Terre. La surpopulation menant à la surexploitation des ressources naturelles, l'industrialisation et la pollution qui en résulte font disparaître les espèces vivantes à un rythme beaucoup plus rapide qu'au cours d'aucune des extinctions précédentes. Même si, en réduction globale de la biodiversité, nous sommes pour l'instant très loin d'avoir atteint le niveau des précédentes extinctions, la vitesse de réduction dramatique des espèces est alarmante. Environ un tiers des populations de tous les vertébrés terrestres ont diminué en taille de 15 % à 30 %. De nombreuses espèces qui n'étaient pas considérées en danger sont en

train de le devenir. Plus de 80 % des espèces de mammifères terrestres ont perdu 60 % de leur population en Afrique, en Asie et en Australie, cette réduction étant à peine moins importante (40 %) en Europe. Le lion (*Panthera leo*) qui vivait autrefois dans la plupart de l'Afrique, au sud de l'Europe et du Moyen-Orient jusqu'au nord-ouest de l'Inde est maintenant confiné dans quelques populations clairsemées d'Afrique sub-saharienne et dans le sanctuaire faunique de la forêt de Gir en Inde. À l'exception de l'Amérique du Sud et des plus hautes latitudes de l'Amérique du Nord, les populations de mammifères ont diminué en moyenne de 70 % partout dans le monde. Cette annihilation biologique va rapidement avoir des conséquences écologiques, économiques et sociales dramatiques, remettant en question l'existence même de l'*Homo sapiens* à la surface de la Terre. Il reste peu de temps, au plus deux ou trois décennies, pour freiner puis stopper cet assaut humain sans précédent sur la biodiversité, si nous voulons survivre nous-mêmes en temps qu'espèce.

Les extinctions massives d'espèces vivantes ont toujours été suivies d'un goulet d'étranglement, une réduction drastique des allèles présents dans une population suite à l'élimination d'une grande partie de celle-ci. La réduction de taille de la population va augmenter les effets de dérive génétique vus précédemment et diminuer le rôle de la sélection naturelle. Lors d'un cataclysme de très grande ampleur, les survivants sont choisis aléatoirement et non en fonction de leur génotype. Si les individus survivants portent des allèles légèrement désavantageux, la dérive va les amplifier au sein de la population. Un cas particulier de goulet d'étranglement est observé lorsqu'une petite partie d'une population part coloniser un autre espace géographique. C'est ce que l'on appelle un effet fondateur. Les membres fondateurs de la colonie portent un nombre limité d'allèles par rapport à la population dont ils sont originaires. Ces allèles vont éventuellement se fixer dans la nouvelle population par dérive.

Ainsi, les six extinctions massives d'espèces vivantes du passé ont créé des goulets d'étranglement à l'échelle de la planète, entraînant

une réduction dramatique de la diversité génétique – donc des chances moindres pour une espèce de survivre à la sélection darwinienne –, tout en favorisant la dissémination d'allèles rares par dérive génétique. La biodiversité terrestre actuelle est donc le résultat de plusieurs cycles « extinction-goulet-dérive » alternant avec de très longues périodes de sélection naturelle éliminant les allèles délétères. C'est ce dynamisme qui a abouti à la structuration génétique des espèces vivantes modernes. Si l'humanité ne met pas un frein rapide à l'extinction massive dont elle est responsable, celle-ci devrait déboucher d'ici quelques décennies sur un nouveau goulet d'étranglement auquel l'*Homo sapiens* lui-même n'est pas certain de survivre. Les espèces vivantes qui en sortiront repeupleront la Terre en quelques millions d'années et seront possiblement assez différentes de celles que nous côtoyons à l'heure actuelle.

LA PHYLOGÉNIE MOLÉCULAIRE ET LA THÉORIE DE LA COALESCENCE

Charles Darwin voyait le monde vivant comme les feuilles d'un arbre reliées à des branches, elles-mêmes reliées à des branches plus épaisses, puis au tronc (Figure 34). Cette vision généalogique du monde vivant a conduit à la théorie moderne de la coalescence, qui est un modèle rétrospectif d'évolution des populations. Dans ce modèle, la comparaison des allèles d'un ou de plusieurs gènes d'individus différents au sein d'une même espèce doit permettre de reconstruire l'histoire évolutive de ces allèles et de remonter ainsi jusqu'au dernier ancêtre commun des individus de cette population, *Last Common Ancestor* en anglais. Par extension, la théorie de la coalescence permettrait – notez l'usage du conditionnel – de remonter dans le temps jusqu'au dernier ancêtre commun unique, LUCA en anglais, « *Last Universal Common Ancestor* », appelé parfois le cénancêtre ou progénote, qui serait la structure cellulaire primitive vivant il y a 3,7 milliards d'années et dont la descendance aurait donné naissance à tous les êtres vivants. Donc, par définition, toutes les approches

mathématiques se basant sur la théorie de la coalescence ne peuvent mener qu'à un seul ancêtre commun dans chaque phylum.

Comme vu précédemment, la description du monde vivant repose depuis le XVIII^e siècle sur la classification taxonomique des êtres qui le composent. Cette taxonomie fut longtemps basée sur la présence ou l'absence de caractères morphologiques ou phénotypiques pour classer une espèce dans un **taxon*** ou dans un autre. C'est Willi Hennig (1913-1976), entomologiste allemand, qui posera dans les années 1950 les bases de la phylogénie moderne. Il propose alors de représenter la généalogie des espèces vivantes sous forme d'un arbre dont chaque branche est un **clade***, ou **groupe monophylétique*** contenant tous les descendants d'un ancêtre commun. L'approche cladistique a remplacé la classification taxonomique, car elle ne repose pas sur la présence de caractères communs qui peuvent découler d'une convergence évolutive, mais seulement sur l'analyse et la comparaison des séquences d'ADN entre elles.

Taxon : Ensemble des êtres vivants partageant les mêmes caractéristiques taxonomiques. Celles-ci peuvent reposer sur la morphologie, la physiologie, l'embryologie, la génétique ou les comportements culturels ou sociaux.

Groupe monophylétique (ou clade) : Groupe qui contient l'ancêtre commun et tous ses descendants.

Les premières approches algorithmiques de la phylogénie ont été développées à la même époque, lorsque les premiers ordinateurs – qui fonctionnaient à l'époque avec des cartes perforées – sont devenus accessibles à la communauté scientifique académique. Des méthodes de regroupement (*clustering* en anglais) ont été alors inventées pour déduire la phylogénie des espèces considérées. Les premiers à développer ce type d'algorithme furent Peter Sneath (1923-2011), un microbiologiste anglais et Robert Sokal (1926-2012), un entomologiste austro-américain. Les premiers travaux se basaient sur les groupes

sanguins des populations humaines, puis avec l'avènement du séquençage des protéines, sur les premières séquences protéiques. Après l'invention du séquençage de l'ADN, la phylogénie moléculaire a d'abord utilisé des séquences partielles de gènes comme l'ARN ribosomique, puis des gènes entiers et enfin, depuis le milieu des années 1990, des génomes entiers, d'abord de bactéries puis d'eucaryotes simples et enfin de vertébrés et de plantes. Des approches différentes, reposant sur des modèles mathématiques distincts, ont mené à de nombreux algorithmes capables de déterminer un arbre. De fait, une phylogénie est rarement unique et l'utilisation de méthodes de calcul différentes mène la plupart du temps à des phylogénies différentes. C'est maintenant un principe bien établi dans la communauté scientifique, mais la possibilité d'obtenir des arbres différents avec des modèles mathématiques différents a profondément troublé les premiers phylogénistes. Ce n'est pas le but de cet ouvrage de décrire en détail ces différentes méthodes mathématiques, mais il faut garder à l'esprit que différents modèles existent, en fonction des questions posées, des séquences à sa disposition et de l'éloignement des espèces considérées.

LES LIMITES DE LA PHYLOGÉNIE MOLÉCULAIRE

La puissance des approches modernes de phylogénie moléculaire n'est plus à démontrer. Celles-ci ont permis de résoudre les liens de parenté entre certains phylums et de reconstruire l'histoire évolutive la plus probable de branches entières du vivant. La phylogénie moléculaire appliquée au séquençage d'abord de gènes, puis de génomes entiers, a prouvé que les anciennes classifications étaient souvent fausses et ont permis de dresser des arbres donnant une représentation plus juste du vivant. Elle a néanmoins ses limites. Elle ne peut prendre en compte que les groupes monophylétiques. Si des transferts massifs de gènes, des introgressions de larges segments d'ADN ou des hybridations interspécifiques se produisent entre deux clades, ces événements vont masquer en partie ou totalement les relations de phylogénie. Les arbres obtenus seront alors partiellement faux. De

fait, comme nous allons le découvrir dans le chapitre suivant, toutes ces exceptions sont le plus souvent des règles dans le monde vivant, remettant en cause non seulement la phylogénie cladistique mais la notion même d'espèce.

À RETENIR

- Charles Darwin est le père de la théorie scientifique de la sélection naturelle. Celle-ci explique l'évolution des espèces vivantes et leur transformation au cours du temps, telles que proposées par Jean-Baptiste de Lamarck au siècle précédent.
- Il existe trois sortes de mutations, les mutations délétères qui sont rapidement éliminées des populations, les mutations avantageuses qui peuvent être positivement sélectionnées dans des conditions propices et les mutations neutres, qui sont les plus fréquentes.
- Les mutations neutres se fixent dans les populations par dérive génétique, un ensemble de mécanismes aléatoires qui éliminent certains allèles d'une population. La sélection naturelle n'agit pas sur les mutations neutres.
- Les extinctions massives d'espèces sont des phénomènes stochastiques qui provoquent une forte dérive génétique.
- La classification moderne des espèces vivantes repose sur la phylogénie moléculaire, qui compare les séquences des génomes afin d'en déduire les liens de parenté entre espèces.

BIBLIOGRAPHIE

- Philosophie zoologique : Jean-Baptiste de Lamarck, Flammarion ISBN : 978-2-0807-0707-9 (1994).
- L'origine des espèces : Charles Darwin, Flammarion ISBN : 978-2-0812-2107-9 (1992).
- La théorie neutraliste : Kimura, *Nature* 217 : 624-626 (1968).
- Les extinctions massives d'espèces : Raup et Sepkoski, *Science* 215 : 1501-1502 (1982).

15

Les forces évolutives qui façonnent les génomes

Depuis le milieu du xx^e siècle, il est commun de définir une espèce comme un ensemble d'individus qui sont capables de se croiser entre eux et dont la descendance est viable et non stérile. Darwin aborde le problème de la distinction entre espèces et variétés dans *On the Origin of Species*. La stérilité de certains hybrides est connue depuis longtemps. Le mulet par exemple, descendant d'un âne et d'une jument, ainsi que la mule, son équivalent femelle, sont stériles. Mais il semble que ni le naturaliste anglais, ni ses confrères, n'aient rapproché cette donnée de la notion d'espèce. Il discute d'ailleurs de cette notion et de celle de variété dans le deuxième chapitre de son œuvre : « *On comprendra, [...] que, selon moi, on a, dans un but de commodité, appliqué arbitrairement le terme espèce à certains individus qui se ressemblent de très près, et que ce terme ne diffère pas essentiellement du terme variété, donné à des formes moins distinctes et plus variables.* » Pour Darwin, comme pour les scientifiques de cette époque, la notion d'espèce est donc basée sur des caractères morphologiques proches

– pouvant entraîner des classifications différentes d'un naturaliste à un autre – et non sur les propriétés reproductives intrinsèques des êtres vivants. On y retrouve la notion de taxon chère aux classificateurs du vivant, discutée précédemment.

Nous avons vu que l'homme moderne était le descendant de croisements entre au moins trois espèces d'hommes ancestraux. Cela surprend généralement les étudiants à qui l'on présente ces résultats. Ne leur a-t-on pas appris que la descendance de deux espèces différentes devait être stérile ? Nous allons voir que le cas d'*Homo sapiens* est loin d'être unique et que les hybridations entre espèces différentes (interspécifiques) sont communément répandues dans la nature, nécessitant sans doute une remise en question sérieuse de la définition d'espèce et de notre façon de dessiner les arbres phylogénétiques.

LES GÉNOMES HYBRIDES, UNE GÉNÉRALITÉ DU VIVANT

Le botaniste américain Edgar Shannon Anderson (1897-1969) publie en 1949 un livre qui décrit des hybridations interspécifiques entre des espèces différentes de fleurs et les différentes combinaisons de génotypes possibles entre les descendants de ces hybridations sur plusieurs générations. Cet ouvrage va avoir un impact certain en botanique et les chercheurs de ce domaine vont rapidement admettre l'idée que ces échanges de gènes entre espèces différentes pouvaient mener à de nouvelles espèces. Les zoologistes, en revanche, ont gardé pendant très longtemps une vision très conservatrice de la notion d'espèce, toujours basée sur l'isolement reproducteur. C'est d'autant plus étonnant que de nouvelles espèces d'animaux hybrides non stériles étaient déjà connues : le ligre, croisement d'un lion et d'une tigresse, dont la femelle peut se reproduire avec un tigre ou avec un lion ; le grolar ou pizzly, croisement d'un ours blanc et d'un grizzly, est parfaitement fertile, tout comme le Canard des Hawaï, un hybride entre le Canard colvert et le Canard de Laysan. Cette vision très conservatrice des zoologistes s'est récemment heurtée de plein fouet à la réalité des hybridations ancestrales de la lignée humaine.

Tant que ce genre d'événement se passait chez les autres, on pouvait sans doute fermer pudiquement les yeux sans trop se poser de questions, mais la découverte de gènes néandertaliens et dénisoviens chez nos ancêtres annonçait le retour brutal à la réalité : les hybridations interspécifiques existent dans toutes les branches du vivant, y compris dans la nôtre.

Chez les levures, on trouve de nombreux hybrides naturels. Les *Saccharomyces* comprennent plusieurs espèces de levure, dont le groupe monophylétique des *Saccharomyces* « *sensu stricto* », qui regroupe *Saccharomyces cerevisiae* (la levure de boulangerie), *Saccharomyces paradoxus*, *Saccharomyces mikatae*, *Saccharomyces kudriavzevii* et *Saccharomyces bayanus*. Ces levures, bien que morphologiquement très proches entre elles, présentent beaucoup de variabilité génétique. La levure *Saccharomyces pastorianus*, utilisée pour le brassage de nombreuses bières de type *lager* qui fermentent à basse température, est un hybride complexe. La séquence de son génome a montré qu'elle était le résultat d'un croisement ancestral entre deux espèces de levures, *Saccharomyces uvarum* et une levure inconnue, suivi d'un transfert de gènes depuis une souche viticole de *Saccharomyces cerevisiae* et enfin une deuxième hybridation avec une autre souche de *Saccharomyces cerevisiae* (Figure 38). Mais les hybrides d'espèces de levures ne se limitent pas aux *Saccharomyces* et l'on trouve des dizaines d'hybrides naturels chez les **ascomycètes***.

Ascomycètes : Champignons qui forment des spores, les cellules germinales des champignons, protégées par un asque qui donne son nom au clade.

Comme Anderson l'avait compris, les hybridations interspécifiques sont extrêmement fréquentes chez les plantes. Mais le champion toutes catégories confondues est sans doute le blé, dont le génome hexaploïde est le résultat de deux hybridations successives. La séquence de *Triticum aestivum* n'a été achevée qu'en 2014, en partie en

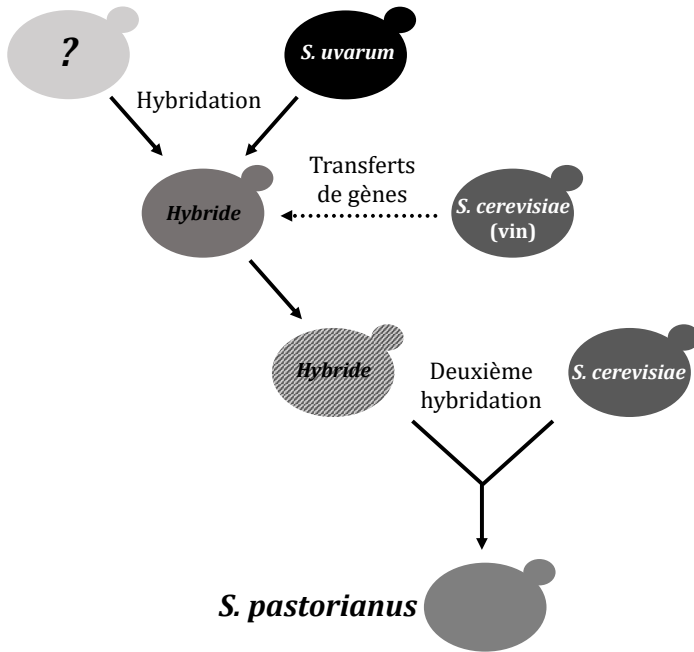


Figure 38 | Hybridations interspécifiques de levures.

raison de la nature hautement répétée de son génome. Celui-ci couvre 17 milliards de nucléotides – plus de six fois le génome humain – et contient environ 124 000 gènes, cinq à six fois plus que chez l’homme. Ce génome monumental est le résultat d’une première hybridation de deux génomes ancestraux, ayant eu lieu il y a environ 800 000 ans, suivie d’une deuxième hybridation avec un troisième génome il y a 400 000 ans. L’ancêtre, *Triticum urartu*, avant la première hybridation contenait 7 paires de chromosomes, le premier hybride en contenait 14 et le blé moderne résultant de la deuxième hybridation, *Triticum aestivum*, en contient 21 paires, soit 42 chromosomes au total (Figure 39). Ces deux hybridations se sont produites de façon indépendante de toute activité humaine ancestrale car l’agriculture a débuté il y a seulement 10 000 ans, au néolithique.

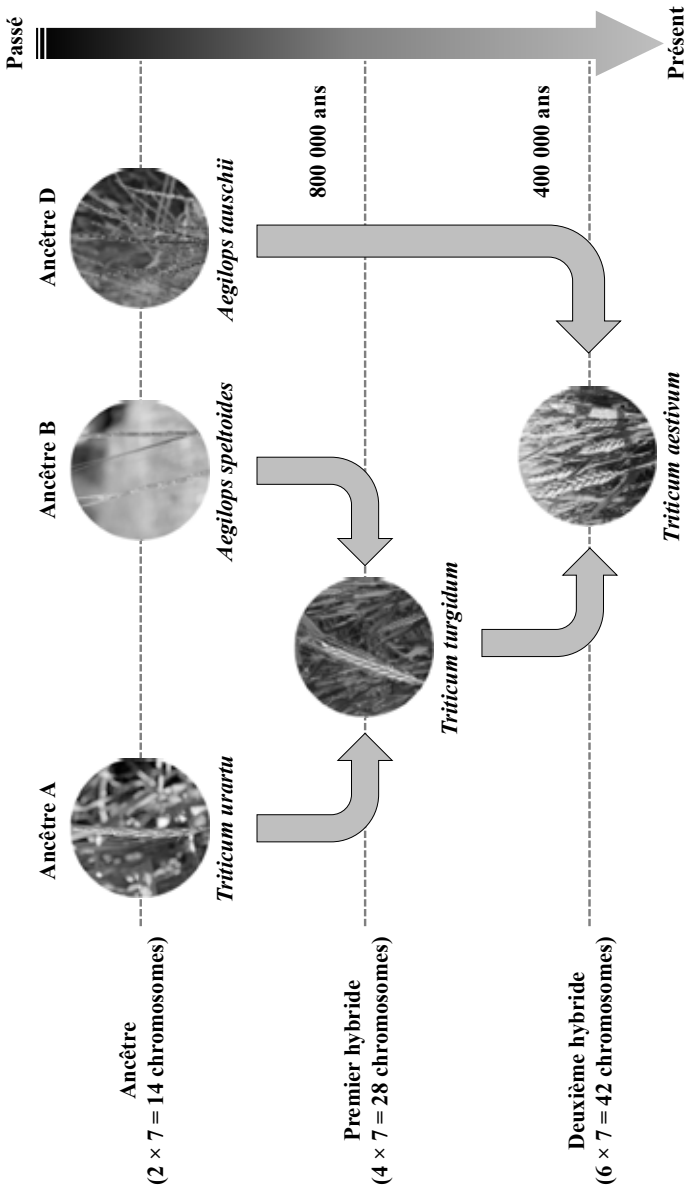


Figure 39 | Le génome hexaploïde du blé moderne.

LE SEXE, UN MOYEN SIMPLE DE CRÉER DE LA DIVERSITÉ

Avant l'invention du sexe, le monde vivant ne contenait que des bactéries, des archées et des virus. La différenciation en deux sexes apparaît avec les premières cellules eucaryotes. L'ancêtre commun à tous les eucaryotes était vraisemblablement **haploïde***, c'est-à-dire qu'il possédait un seul exemplaire de chaque chromosome. À partir de cet eucaryote primitif s'est développé un système de **syngamie*** permettant la fusion des noyaux haploïdes de deux cellules eucaryotes. La protéine responsable de la fusion des **gamètes*** mâle et femelle a été identifiée chez les plantes et les animaux. Il s'agit de la protéine Hap2, qui présente une homologie de séquence, de structure et de fonction avec une protéine retrouvée chez certains virus, permettant à l'enveloppe virale de fusionner avec la membrane de la cellule hôte. L'acquisition de cette protéine virale par les premières cellules eucaryotes a donc permis leur fusion pour former des cellules **diploïdes***, contenant deux lots de chromosomes. Ceci démontre qu'un gène essentiel à la reproduction sexuée a été transmis d'un virus à l'ancêtre de tous les eucaryotes, il y a probablement environ 1,5 milliard d'années.

Haploïde : Cellule ne possédant qu'un seul exemplaire de chaque chromosome. Les bactéries ainsi que les cellules germinales des animaux (spermatozoïdes et ovules) sont des exemples de cellules haploïdes.

Diploïde : Cellule possédant deux exemplaires de chaque chromosome. La plupart des animaux et des plantes sont diploïdes.

Syngamie : Fusion de deux cellules haploïdes de type sexuel opposé pour former une cellule diploïde, contenant deux exemplaires de chaque chromosome, l'un paternel et l'autre maternel.

Gamète : Cellule haploïde sexuée mature, qui peut fusionner avec un gamète de type sexuel opposé pour former un individu mature. Chez les mammifères, les gamètes mâles sont les spermatozoïdes et les gamètes femelles sont les ovules.

Ce nouvel eucaryote maintenant diploïde possède une deuxième copie de chaque chromosome. On appelle les chromosomes portant

le même lot de gènes et issus de deux cellules parentales différentes, des *chromosomes homologues*. Chaque gène est donc présent en deux copies, ou deux allèles. Si les deux allèles sont identiques, on dira que le gène est présent à l'état **homozygote***. Si les deux allèles sont différents, ce qui est le cas le plus fréquent, on dira que le gène est présent à l'état **hétérozygote***.

Homozygote : Deux gènes sont à l'état homozygote si les séquences nucléotidiques des deux allèles du gène portés par les deux chromosomes homologues sont strictement identiques.

Hétérozygote : Deux gènes sont à l'état hétérozygote si les deux allèles du gène sont différents, c'est-à-dire si les deux séquences nucléotidiques ne sont pas strictement identiques. C'est la situation la plus courante chez les individus sexués.

L'état diploïde est très avantageux pour un être vivant. En effet, une bactérie ne possédant qu'un seul exemplaire de chaque gène, toute mutation dans un gène essentiel à sa survie entraîne la mort de la cellule ayant reçu la mutation. Chez un individu diploïde, une mutation qui inactive un gène essentiel n'a normalement pas d'effet sur la cellule tant que la deuxième copie du gène – appelée copie sauvage – est intacte. On dit alors que l'allèle sauvage est *dominant* sur l'allèle muté, qui est lui-même *récessif* par rapport à l'allèle sauvage. La relation génétique de dominance-récessivité est donc apparue avec la syngamie, ancêtre de la reproduction sexuée. La possibilité de faire varier un des deux allèles sans mettre en danger la survie de la cellule a procuré un avantage évolutif certain, permettant de diversifier les gènes afin de créer de la nouveauté. Le formidable succès évolutif du monde eucaryote, caractérisé par l'explosion de la biodiversité animale et végétale, est lié au sexe. Mais la syngamie n'est que la moitié de la reproduction sexuée, car maintenant que les deux exemplaires de chaque allèle sont dans la même cellule, il faut les séparer pour créer la diversité. La séparation des allèles va s'effectuer lors d'un processus

complexe et hautement régulé appelé **méiose***, qui permet à une cellule diploïde de donner naissance à quatre cellules (ou gamètes) haploïdes. Chaque chromosome homologue est tout d'abord répliqué, pour donner deux exemplaires de chaque chromosome, soit quatre *chromatides-sœurs*, qui sont donc identiques deux à deux. Ces quatre chromatides vont ensuite échanger une partie de leurs allèles par **recombinaison homologue***, puis les quatre chromatides vont se séparer pour donner naissance à quatre cellules haploïdes contenant chacune une des quatre chromatides-sœurs (Figure 40).

Méiose : Mécanisme de division cellulaire des cellules germinales qui va donner quatre gamètes haploïdes à partir d'une cellule diploïde.

Recombinaison homologue : Échange de nucléotides entre deux segments d'ADN identiques ou très similaires. Elle peut se produire en réponse à une cassure accidentelle de l'ADN (rayons X) ou bien à la suite de cassures programmées comme lors de la méiose.

Par syngamie, les gamètes ainsi obtenus vont pouvoir fusionner pour reformer une cellule diploïde. Ainsi, les cycles successifs de syngamie-méiose permettent de brasser les allèles, de les mélanger afin de créer de la diversité génétique (Figure 40). Prenons l'exemple de deux mutations qui, prises individuellement, n'apportent pas d'avantages sélectifs, mais qui présentes dans la même cellule en amènent un. Dans une population asexuée, il faut que les deux mutations se produisent indépendamment dans la même cellule. En revanche, dans une population sexuée, l'échange et le brassage des allèles vont permettre rapidement aux deux mutations de se retrouver dans la même cellule. La méiose est sans doute apparue après la syngamie, peut-être dans un cyste eucaryote ancestral, caractérisé par la formation d'une enveloppe protectrice autour de la cellule, lui permettant de résister au froid, au gel et à la dessiccation. La cellule se trouve alors dans une sorte d'animation suspendue, où ses fonctions métaboliques sont réduites au minimum en attendant des

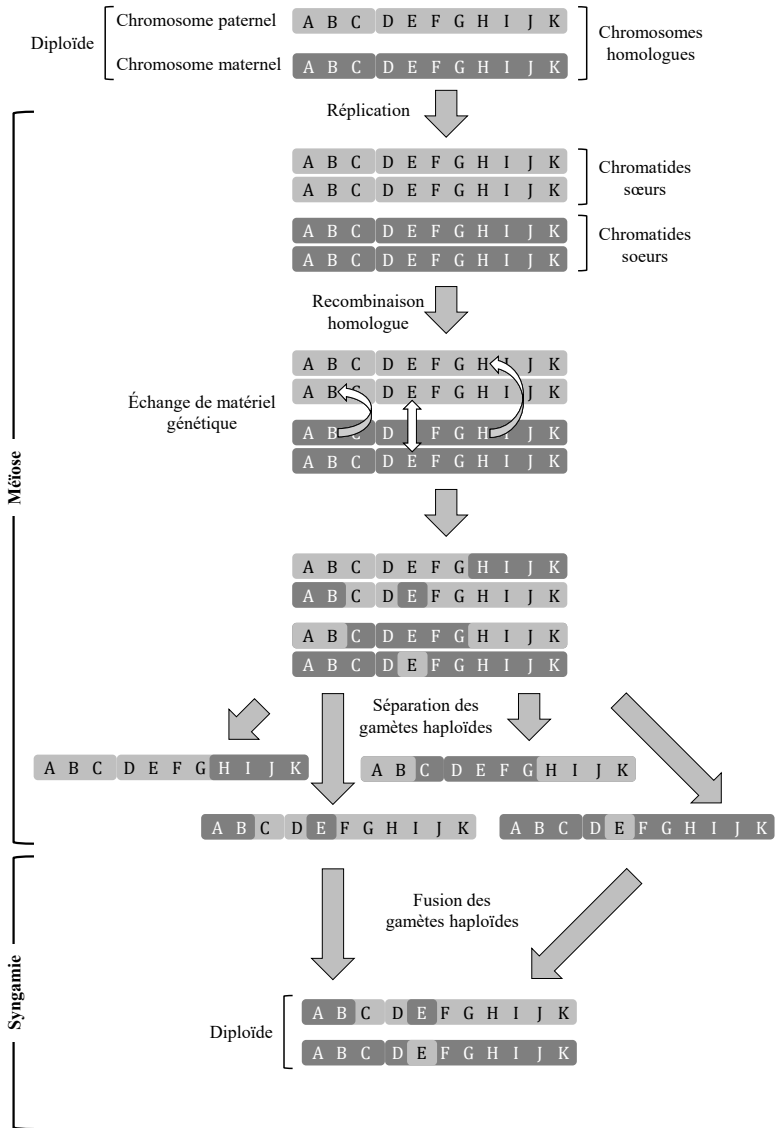


Figure 40 | La méiose.

jours meilleurs. La réplication chromosomique ainsi que la machinerie de recombinaison homologue existaient déjà dans le monde bactérien. L'invention de la méiose a simplement nécessité de réguler ce mécanisme en adaptant le cycle cellulaire d'une division asexuée à une reproduction sexuée. En termes de créativité, c'est loin d'être une aussi grande révolution que le passage du monde procaryote au monde eucaryote, mais l'invention du sexe va donner un avantage considérable aux organismes le pratiquant, même s'il a un coût énergétique certain.

LE PRIX DU SEXE

Le coût du sexe n'est pas négligeable. Au sein d'une population sexuée qui contient autant de mâles que de femelles, la moitié des individus nécessite l'autre moitié pour se reproduire. Au sein d'une population asexuée, chaque individu peut se reproduire seul. En d'autres termes, les populations sexuées doivent payer le coût des mâles dont la plupart sont inutiles, puisque seules les femelles sont absolument essentielles à la survie de l'espèce. Une population sexuée va donc grossir de façon linéaire, alors qu'une population asexuée va grossir de façon exponentielle. À court terme, le prix à payer pour la sexualité est énorme car la moitié de la population est fabriquée en pure perte. Mais en réalité, pas complètement. En effet, l'avantage évolutif du sexe va se faire sentir sur le long terme. L'accumulation des mutations possiblement délétères sera neutralisée par le brassage des allèles résultant de la méiose, un allèle déficient sur un chromosome étant compensé par un allèle fonctionnel sur l'autre chromosome. Une population asexuée devra faire face à la même accumulation de mutations mais sans pouvoir les contrebalancer avec un deuxième allèle fonctionnel. Cette accumulation de mutations délétères mènera sur le long terme à l'élimination de la population, suivant le modèle dit de « la roue à cliquet de Muller », où chaque mutation s'ajoute à la précédente sans pouvoir être compensée par l'arrivée d'un allèle fonctionnel par reproduction sexuée. John Maynard Smith (1920-2004),

célèbre généticien britannique spécialiste de l'évolution, a proposé que l'avantage du sexe pouvait être mieux appréhendé en étudiant les organismes se reproduisant par **parthénogenèse***. Si une mutation empêchant la méiose apparaît dans une femelle, celle-ci va se transmettre à toute la population descendante et non plus à la moitié seulement comme dans le cas d'une reproduction sexuée.

Parthénogenèse : Reproduction d'un individu adulte à partir d'un gamète femelle non fécondé.

Cette mutation devrait donc se répandre et se fixer rapidement dans la population, la parthénogenèse remplaçant à terme la reproduction sexuée. Or, cela n'est pas et de loin, le mode de reproduction préféré des animaux. En fait, de nombreux phylums contiennent à la fois des animaux sexués proches cousins d'autres parthénogénétiques, suggérant que le gain de la reproduction non sexuée n'est pas suffisant pour l'imposer sur le long terme à tout un phylum. La plupart des taxons animaux (et végétaux) préfèrent donc payer le coût nécessaire à entretenir des mâles inutiles, plutôt que de s'affranchir définitivement du sexe. Néanmoins, comme souvent en biologie, il existe au moins une exception à cette règle, avec le cas très particulier des rotifères bdelloïdes.

LE CAS FASCINANT DES BDELLOÏDES

Les rotifères sont de petits animaux protostomiens, qui ont donc divergé de la branche menant à l'homme il y a environ 700 millions d'années. Les bdelloïdes sont des animalcules mesurant un dixième de millimètre et vivant dans les mousses ou dans la terre humide. Cette lignée animale est vieille d'au moins 60 millions d'années et montre une grande diversité avec plus de 400 espèces connues. La physiologie d'*A. vaga* est remarquable car cet animal montre une extraordinaire résistance à la dessiccation, à la congélation et aux rayonnements ionisants qui cassent l'ADN et sont normalement létaux pour tous les

êtres vivants. Ces observations suggèrent qu'*A. vaga* possède de très efficaces mécanismes de réparation de ses chromosomes. Ces propriétés en font un organisme d'intérêt pour la recherche spatiale, entre autres. Ce sont des animaux asexués, ne produisant pas de mâle ou d'hermaphrodites et se reproduisant donc exclusivement par parthénogenèse. On n'a jamais observé de méiose chez cet animal, en faisant un cas extrêmement rare d'animal asexué. Le séquençage récent du génome de l'un d'entre eux, *Adineta vaga*, a permis de révéler sa structure unique. *A. vaga* possède un génome long de 244 millions de nucléotides et contenant 32 378 gènes, environ 50 % de plus que le génome humain. Ce génome diploïde est constitué de six paires de chromosomes, qui se ressemblent deux à deux. Ils proviennent donc d'une duplication ancestrale du génome de même nature que celles que nous avons précédemment décrites (Figure 41).

Mise à part cette duplication ancestrale, les chromosomes d'*A. vaga* ont deux caractéristiques intrigantes. Tout d'abord certaines régions sont parfaitement identiques entre chromosomes homologues, elles sont homozygotes. Cela indique un taux de recombinaison élevé entre ces régions, impliquant des échanges d'allèles fréquents entre chromosomes homologues. En l'absence supposée (car jamais observée) de recombinaison méiotique, cette observation est surprenante et suggère que ces fréquents échanges génétiques sur de grandes distances se produisent en dehors de la méiose, ce qui n'a jamais été observé chez d'autres eucaryotes. De plus, les transferts de gènes étrangers sont nombreux, représentant environ 8% du contenu en gènes de l'animal. Ces transferts de gènes interespèces, appelés transferts horizontaux (voir paragraphe suivant), très fréquents dans le monde bactérien, sont normalement rares chez les eucaryotes. *A. vaga* montre le taux le plus élevé de transferts horizontaux de tout le monde animal, particulièrement vers les extrémités des chromosomes.

À la suite de ces étonnantes découvertes, des questions restent en suspens. La présence de larges régions chromosomiques identiques

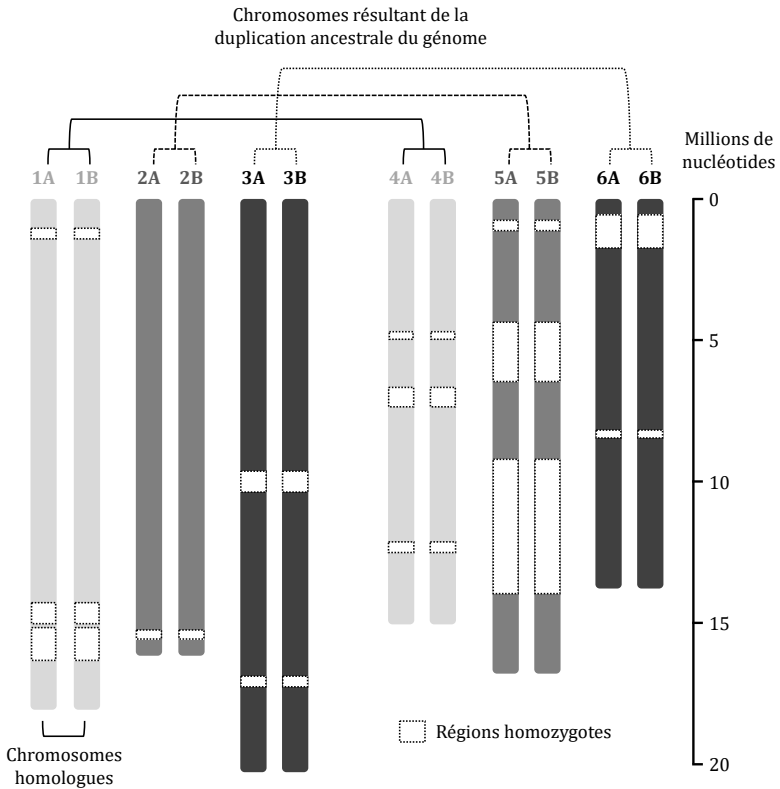


Figure 41 | Le génome d'*Adineta vaga*.

prouve que la recombinaison homologue est très active dans cet animal. Il pourrait s'agir de recombinaison méiotique qui ne donnerait naissance qu'à des individus femelles, expliquant pourquoi on n'a jamais observé de mâles dans cette lignée. On peut aussi penser que ces échanges de matériel génétique se produisent lors du cycle de vie caractéristique l'animal, alternant des périodes de dessiccations intenses qui provoquent des cassures de l'ADN, avec des périodes de vie en milieu humide permettant la réparation de ces cassures, entraînant de fréquents évènements de recombinaison homologue. La fréquence très élevée des transferts horizontaux pose aussi l'intrigante

question de l'efficacité des mécanismes de recombinaison. Nul doute que les recherches en cours sur les rotifères bdelloïdes permettront de lever un coin du voile sur les étonnantes capacités de ces étranges animaux.

LES TRANSFERTS HORIZONTAUX

À l'inverse du passage d'une information génétique d'un individu à sa propre descendance (transfert « vertical »), le transfert horizontal fait passer un gène (ou un groupe de gènes) d'un organisme vivant à un autre organisme vivant, physiquement proche dans l'espace mais n'ayant pas de lien de parenté directe avec le premier. Les transferts horizontaux de gènes entre bactéries sont connus depuis le milieu du XX^e siècle et ont été initialement décrits au Japon comme des transferts de la résistance à certains antibiotiques entre des souches de *Shigella* et d'*Escherichia coli*. Depuis cette époque, il a été montré que le transfert horizontal est un mécanisme très fréquent et très important, jouant un rôle capital dans la dynamique des génomes bactériens. Ceci pose d'ailleurs un problème certain pour la reconstruction des arbres phylogénétiques bactériens, comme dans l'exemple décrit plus tôt des ARNt synthétases. Des cas plus rares mais bien documentés de transferts horizontaux de gènes entre bactéries et levures ou entre levures ont été également décrits. Les transferts horizontaux vers un organisme sexué sont plus rares pour une raison simple, les gènes ainsi transférés doivent intégrer le génome contenu dans les gamètes pour être transmis à la descendance. Ainsi, la sexualité est un frein naturel aux transferts horizontaux. Il existe néanmoins quelques cas remarquables de transferts de gènes entre bactéries ou champignons et métazoaires.

Wolbachia pipientis est une bactérie symbiotique vivant à l'intérieur des cellules de plusieurs **arthropodes*** dont 20 % des espèces d'insectes,

Arthropodes : Invertébrés qui comprennent – entre autres – les arachnides, les insectes et les crustacés.

ainsi que dans certains **nématodes***. La comparaison du génome de *Wolbachia* avec celui de plusieurs insectes et vers a montré la présence d'une longue séquence bactérienne, contenant 44 des 45 gènes de *Wolbachia*, intégrés dans le génome d'une drosophile (mouche du vinaigre), hôte naturel de cette bactérie symbiotique. La presque totalité du génome bactérien avait donc été transférée dans celui de la mouche. Parmi toutes les espèces dont le génome fut passé au crible, un nématode, un moustique, une tique, trois espèces de guêpes et cinq espèces de drosophiles contenaient des fragments plus ou moins importants de l'ADN de *Wolbachia* intégrés dans leur génome. Cette découverte ébranla très sérieusement le dogme que les transferts horizontaux entre bactéries et eucaryotes étaient extrêmement rares et restreints à des cas isolés.

Nématodes : vers ronds, dont le plus connu est *Caenorhabditis elegans*, un organisme modèle en génétique.

Un autre exemple tout aussi étonnant concerne le transfert de gènes de levure vers certaines espèces de pucerons. Le puceron vert du pois est de couleur vert pomme ou rose, ce polymorphisme coloré leur permettant de se camoufler de leurs prédateurs. Ces couleurs viennent de caroténoïdes, des pigments que l'on trouve dans certaines bactéries, dans des champignons et dans les plantes où ils jouent un rôle important dans la photosynthèse. Les animaux ne synthétisent pas de caroténoïdes, mais ont besoin d'en absorber pour fabriquer de la vitamine A. L'étude de la séquence du génome de cette espèce de pucerons a montré la présence de sept gènes impliqués dans la synthèse de caroténoïdes, présents à trois endroits différents du génome de l'insecte. La comparaison de ces gènes avec les séquences présentes dans les bases de données a prouvé que ces gènes ne ressemblaient à aucun gène d'origine animale, végétale ou bactérienne, mais qu'ils ne pouvaient provenir que d'une espèce fongique. Ces gènes ont donc été acquis par l'ancêtre des pucerons verts du pois par un transfert horizontal depuis une levure vivant dans la même niche écologique

que le puceron. Puis, ce gène ancestral s'est dupliqué plusieurs fois dans la lignée menant au puceron moderne. Étant donné l'importance des caroténoïdes pour les animaux, il serait surprenant que ce type de transfert horizontal ne se soit produit qu'une seule fois au cours de l'évolution.

Un autre exemple de transfert horizontal massif vient des rotifères bdelloïdes précédemment évoqués. Le génome d'*Adineta vaga* contient plusieurs gènes d'origine étrangère, dans des régions proches des **télomères***. Ces gènes sont trouvés dans de grands segments d'ADN de plusieurs dizaines de milliers de nucléotides. Les gènes qui s'y trouvent sont des gènes de métabolisme, clairement d'origine bactérienne, fongique ou végétale, certains d'entre eux étant interrompus par des introns alors que leurs homologues bactériens ne le sont pas. On trouve des gènes codant des enzymes du métabolisme des sucres ou des

Télomère : Chaque chromosome linéaire contient deux télomères, situés aux extrémités des chromosomes et constitués d'ADN répété et de protéines. Les télomères servent à protéger les extrémités chromosomiques d'une dégradation et sont donc essentiels à l'intégrité des chromosomes. Le vieillissement cellulaire est en partie lié à leur raccourcissement.

acides aminés, ou encore jouant un rôle dans l'oxydo-réduction. Ces gènes sont souvent trouvés en association avec des transposons ou des restes d'éléments transposables, posant la question du rôle possible de ces éléments dans le transfert horizontal de la bactérie à l'hôte métabolite. Ces transferts massifs d'ADN étranger se produisent préférentiellement dans les régions télomériques pour des raisons qui ne sont pas encore bien comprises, mais qui ont sans doute à voir avec l'accessibilité de l'ADN par rapport à d'autres régions plus centrales des chromosomes.

Il est vraisemblable que d'autres cas de transferts horizontaux, entre des organismes vivants aussi éloignés que des bactéries et des animaux ou des plantes, seront découverts dans les années à venir,

quand les séquences de génomes de bonne qualité seront disponibles en plus grand nombre pour permettre de plus fines comparaisons. En effet, la puissance de la génomique comparative se mesure autant à l'aune du nombre de génomes disponibles qu'à la qualité de ces séquences. Si l'on dispose de peu de génomes, ou de génomes dont la séquence est de piètre qualité (erreurs de séquençage, mauvais assemblage des séquences dû à des régions répétées, etc.), il sera difficile de faire des comparaisons fiables et d'en tirer des conclusions claires. Heureusement pour le généticien génomiste, le nombre de génomes disponibles de bonne qualité augmente de façon exponentielle et l'apport de cette discipline à notre compréhension du vivant dans son ensemble ne fera que croître dans les prochaines années.

LES DUPLICATIONS TOTALES DE GÉNOMES

Lorsque le génome de la levure de boulangerie *Saccharomyces cerevisiae* fut totalement séquencé, en 1996, il apparut assez rapidement que de grandes régions de ce génome étaient dupliquées. L'analyse précise de ces régions permit d'arriver à la conclusion que le génome de cette levure était le résultat de la duplication totale du génome de son ancêtre, suivi de la perte de nombreux gènes dupliqués au cours des millions d'années d'évolution ayant suivi cette duplication. Suite à cette découverte, de nombreux autres épisodes de duplications totales de génomes eucaryotes furent rapportés.

Les vertébrés ont subi deux duplications totales de leur génome. Une troisième duplication aura lieu dans la lignée menant aux amphibiens. Nous avons vu précédemment que le génome de *Tetraodon nigroviridis* portait également la trace d'une duplication ancestrale de son génome, spécifique de la lignée des poissons téléostéens. Le cilié *Paramecium tetraurelia* exhibe fièrement trois duplications totales successives de son génome, portant son nombre de gènes total à environ 40 000, deux fois le nombre de gènes humains ! Comme vu précédemment, la plupart des plantes cultivées dont la séquence est connue, le maïs, la tomate, la pomme de terre, le soja, le coton, ont

subi au moins une duplication totale de leur génome dans un passé plus ou moins lointain. Lorsque le génome de la plante modèle, l'arabette *Arabidopsis thaliana*, fut séquencé, on y trouva la trace de ce qui pouvait être interprété à l'époque comme quatre duplications totales de son génome s'étant produites dans un intervalle de 100 millions d'années. Plus récemment, le séquençage du génome de la vigne cultivée, *Vitis vinifera*, a jeté un nouvel éclairage sur l'évolution de celui de l'arabette. Le génome de *Vitis vinifera* totalise 487 millions de nucléotides et contient 30 434 gènes codant des protéines. Sa comparaison avec celui d'*Arabidopsis thaliana* a montré que l'ancêtre dicotylédone de ces deux plantes possédait un génome hexaploïde résultant de deux hybridations successives. Deux duplications totales du génome se sont par la suite produites dans la branche menant à l'arabette, donnant la fausse impression de quatre duplications au total. Les duplications totales de génomes sont résumées sur la Figure 23 pour les animaux et les champignons, et sur la Figure 25 pour les plantes. On comprend bien avec à ces exemples que la richesse de la génomique comparative tient autant au nombre d'espèces séquencées dont on dispose qu'à leur proximité évolutive. Les détails évolutifs les plus fins ne seront visibles qu'en comparant des espèces très proches.

QUE DEVIENNENT LES GÈNES DUPLIQUÉS ?

Il est techniquement difficile de différencier une hybridation entre deux espèces très proches d'une duplication totale du génome, si l'on ne dispose pas d'une bonne référence. Il est possible que certains génomes dans lesquels on pense qu'une duplication totale ancestrale se soit produite, soient en fait des hybrides de deux espèces proches. Après la duplication, l'organisme se retrouve avec deux jeux identiques des mêmes données. Que vont donc devenir les deux copies de chacun des gènes dupliqués ? La pression de sélection sur chaque gène se trouve considérablement relâchée et en théorie l'une des deux copies pourrait disparaître sans que cela soit fatal à l'organisme. C'est en partie ce qu'il se passe dans certaines espèces. Après la duplication

totale du génome, de nombreux gènes sont rapidement perdus, afin de revenir à une taille de génome comparable à ce qu'elle était avant la duplication. Une étude des gènes perdus après les duplications successives du génome de l'arabette montre une surreprésentation de gènes impliqués dans la transcription, dans la traduction et la dégradation des protéines, ainsi que des gènes dont le produit a une activité de liaison à l'ADN. Les gènes qui se retrouvent sous-représentés sont ceux codant des protéines mitochondriales ou chloroplastiques, ainsi que les gènes impliqués dans la réparation des dommages à l'ADN.

Certains gènes dupliqués peuvent théoriquement évoluer en accumulant des mutations qui vont modifier la fonction de l'une des deux copies pour créer une nouvelle fonction. C'est la **néofonctionnalisation*** de la deuxième copie. Même s'ils sont possibles, les cas de sélection positive menant à une néofonctionnalisation sont en pratique extrêmement rares, comme prédit par la théorie neutraliste.

Néofonctionnalisation : Création d'une nouvelle fonction par accumulation de mutations dans un gène dupliqué, par sélection darwinienne positive.

Un des rares exemples connus concerne la protéine antigèle des poissons vivant dans les eaux glaciales de l'Antarctique. La lycode *Lycodichthys dearborni* ressemble à une anguille trapue d'une vingtaine de centimètres de long et fabrique une protéine antigèle qui lui permet de survivre à des températures auxquelles son sang devrait geler. Le génome de cette lycode contient deux gènes paralogues issus d'une duplication, *SAS-A* et *SAS-B*. Ces deux gènes codent une enzyme impliquée dans la biosynthèse d'acide sialique. Indépendamment de son activité enzymatique, cette protéine possède une faible affinité pour la glace. Le gène *SAS-B* a été dupliqué sur un autre chromosome (*SAS-B'*) et cette duplication accompagnée d'une grande délétion enlevant la plus grande partie du début du gène, ne laissant intact que le dernier exon sur les six présents initialement. Ce gène

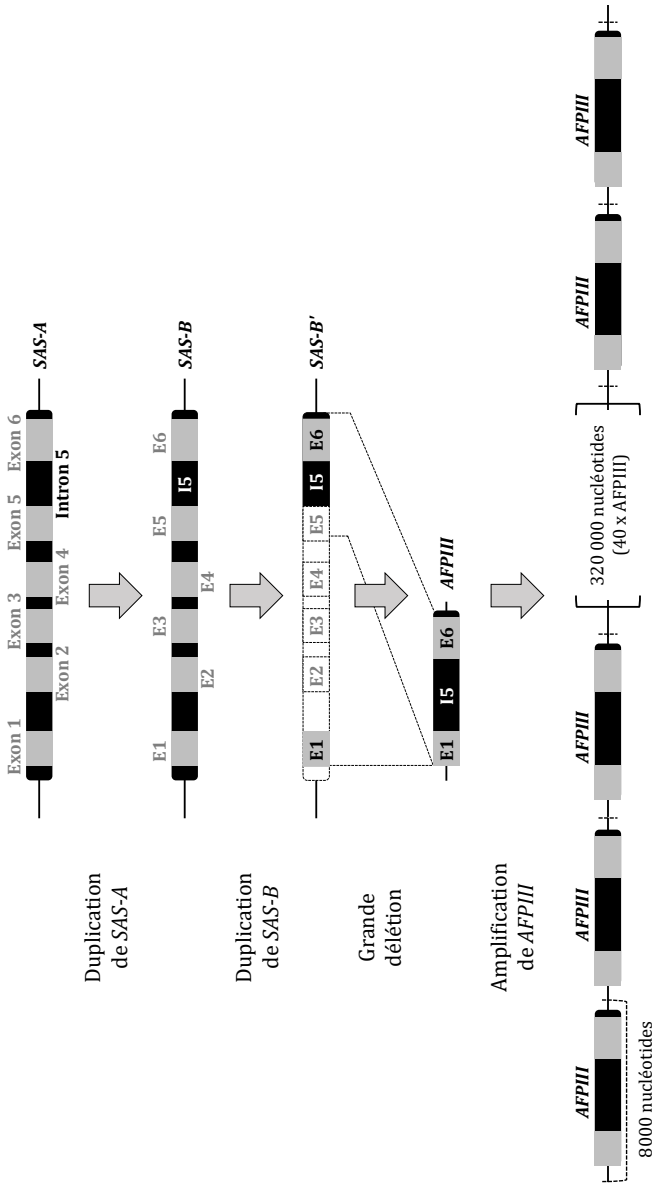


Figure 42 | Néofonctionnalisation de la protéine antigène de la lycode.

dupliqué et raccourci (appelé *AFPIII*) code une protéine plus courte qui se fixe de façon plus efficace que *SAS-B* aux cristaux de glace en empêchant totalement leur croissance. *AFPIII* a par la suite été amplifié en une quarantaine de copies répétées en tandem le long du chromosome, conférant à la lycode sa résistance physiologique aux grands froids et lui permettant de nager dans les eaux glaciales de l’océan Austral (Figure 42).

Une autre possibilité pour un gène dupliqué est la **subfonctionnalisation***, qui consiste à utiliser la deuxième copie pour la même fonction, mais dans un compartiment cellulaire différent (mitochondrie au lieu de noyau par exemple), ou à un moment différent de la vie de la cellule ou de l’organisme. Il n’y a pas dans ce cas de sélection de nouvelle fonction, mais une spécialisation de la copie dupliquée. Un des exemples les plus connus est celui de la duplication du gène *CYC1* chez la levure. Ce gène code le cytochrome C de la chaîne respiratoire mitochondriale. Il possède un paralogue, *CYC7*, qui lui est identique à 84 %, mais qui est seulement exprimé lorsque la concentration en oxygène est faible et que les cellules sont en condition d’hypoxie, alors que *CYC1* est exprimé quand le niveau d’oxygène disponible est normal.

Subfonctionnalisation : Modification d’un gène dupliqué par accumulation de mutations sélectionnées positivement, pour adapter sa fonction à un sous-compartiment cellulaire différent (mitochondrie, noyau, etc.) ou à un moment différent de la vie de la cellule ou de l’organisme.

ÉVOLUTION RAPIDE DES GÉNOMES EUCARYOTES PAR CHROMOTHRIPSIE

Les génomes des cellules cancéreuses montrent très souvent de larges réarrangements chromosomiques qui les différencient nettement de ceux des cellules normales. Ces réarrangements massifs peuvent avoir lieu lorsque les mécanismes de surveillance de la stabilité du génome

sont déficients ou inopérants pour quelque raison que ce soit. Quand ces mécanismes de surveillance sont inactivés, l'apparition accidentelle de cassures dans l'ADN, par suite de causes exogènes (rayons X, produits chimiques) ou endogènes (erreurs pendant la réplication de l'ADN), ne va pas être correctement détectée par la cellule et ne sera pas réparée, menant à des changements structuraux complexes des chromosomes touchés. En retour, ces changements complexes vont finir par dérégler la machinerie de contrôle de la division cellulaire et les cellules affectées vont se diviser de façon anarchique jusqu'à envahir l'organisme sous forme d'une tumeur, menant finalement à la mort de l'individu si elle n'est pas repérée et retirée à temps. Certains cancers ne forment pas de tumeur solide qui pourrait faire l'objet d'une ablation chirurgicale, par exemple les cancers du sang qui touchent les globules blancs. Les récentes avancées technologiques dans le domaine du séquençage ont permis de cartographier très précisément les réarrangements chromosomiques observés dans des tumeurs. Dans certaines d'entre elles, un ou plusieurs chromosomes étaient pulvérisés en plusieurs dizaines – voire plusieurs centaines – de morceaux et réassemblés de façon anarchique et apparemment aléatoire, fusionnés avec d'autres morceaux de chromosomes. Ce phénomène, initialement découvert dans une leucémie lymphoïde chronique, a depuis été observé dans de multiples cancers touchant des organes divers, prouvant que ces réarrangements massifs étaient chose fréquente dans l'espèce humaine. Il a été baptisé **chromothripsie***, de la racine grecque *chromo* que l'on retrouve dans chromosome et *thripsis* qui signifie « couper en morceaux ». Les chromosomes d'un génome soumis à une chromothripsie sont réarrangés de façon tellement spectaculaire qu'ils ne sont plus reconnaissables. On a ici affaire non pas à une modification incrémentale d'un génome mais clairement à un saut dramatique dans l'inconnu. S'il était possible de cloner une de ces cellules pour en faire un être vivant entier, il est vraisemblable que l'être en question ressemblerait bien peu à un *Homo sapiens* (en admettant que l'embryon soit viable) !

De grands réarrangements de génomes entiers ont été également décrits chez la levure *Saccharomyces cerevisiae*. Des souches de cette levure ont été modifiées pour être partiellement déficientes dans la synthèse des protéines, ce qui confère aux cellules un très net ralentissement de leur croissance. Après une centaine de générations, des mutants apparaissent qui poussent normalement et qui envahissent la culture des individus moins performants. On pourrait penser que ces cellules ont reçu une mutation qui leur permet de compenser le défaut initial, une application directe de la théorie darwinienne sur la sélection positive des variations avantageuses. Or il n'en est rien. La séquence de leur génome a montré que ces mutants portent toujours l'allèle déficient initial, mais que le gène déficient a été amplifié de façon dramatique plusieurs dizaines de fois dans le génome, compensant ainsi par le nombre de copies l'inefficacité de la protéine synthétisée. Mieux, au cours des centaines de générations suivantes, le nombre de copies du gène déficient augmente encore, créant des chromosomes géants. Ces amplifications sont parfois portées par le chromosome original, ou se présentent sous forme d'amplifications circulaires extra-chromosomiques ou d'autres fois encore vont créer des **néochromosomes*** en se réarrangeant avec un morceau d'un autre chromosome. Le mécanisme générant ces chromosomes géants est similaire à celui générant des chromothripsies et implique la réparation inadéquate de cassures de l'ADN.

Néochromosome : Chromosome créé par réarrangement génétique dans un génome donné, pouvant impliquer une duplication totale ou partielle de tout ou partie d'un ou plusieurs chromosomes préexistants.

Les génomes résultant de ces événements de remodelage ne ressemblent en rien aux génomes initiaux. L'évolution a fait un bond conceptuel majeur en inventant *de novo* une nouvelle structure génomique – souvent d'ailleurs plusieurs structures dans le même temps, ou presque. De tels bonds évolutifs importants ont été proposés par

deux paléontologues américains, Stephen Jay Gould (1941-2002) et Niles Eldredge (1943-), dans leur théorie des équilibres ponctués publiée en 1972. Leur modèle réfute la thèse de l'évolution darwinienne par accumulation de mutations incrémentales, mais suggère que les bonds qualitatifs sont effectués lors de modifications dramatiques regroupées sur une période de temps bien définie et limitée dans le temps (à l'échelle des temps géologiques). À l'heure actuelle, de telles transformations spectaculaires restant extrêmement rares, il est difficile de mesurer l'impact de tels mécanismes sur l'évolution de la vie terrestre. En revanche, il est vraisemblable que le séquençage massif d'un grand nombre d'espèces vivantes devrait nous en apprendre plus sur le rôle possible de ces mécanismes dans la création et l'évolution de nouvelles espèces.

DIVERSITÉ DES MÉCANISMES D'ÉVOLUTION DES GÉNOMES

En plus des mécanismes proprement darwiniens, impliquant l'accumulation incrémentale des mutations favorables et éliminant les mutations défavorables, et la dérive génétique qui de façon aléatoire élimine certains gènes et en garde d'autres, il existe donc plusieurs mécanismes à l'origine de la modification et de l'évolution des génomes. Les hybridations interspécifiques créent des individus hybrides qui après plusieurs générations vont apporter de la diversité en introduisant dans un des deux génomes des introgressions d'ADN de l'autre génome. L'exemple le plus parlant est celui de l'hybridation entre *Homo sapiens* et *Homo neanderthalensis* précédemment décrite. Les duplications totales de génomes, à l'œuvre dans toutes les branches du vivant, vont mener à la duplication d'un très grand nombre de gènes, dont certains vont permettre l'émergence de nouvelles fonctions, par neo – ou sub-fonctionalisation. Les duplications segmentaires de grands fragments d'ADN par un mécanisme impliquant la recombinaison homologue entre séquences répétées vont également introduire rapidement de la nouveauté. Les gènes résultant de ces duplications vont pouvoir également se spécialiser.

Les transferts horizontaux, que l'on croyait à tort restreints aux procaryotes, sont également un moteur puissant d'évolution des génomes eucaryotes, même si la reproduction sexuée érige une solide barrière à ce type d'évènements. On peut finalement citer la création de gène par duplication d'un transposon ou par transposition d'un morceau de gène ou d'un exon. Cette transposition pourra être accompagnée de la perte d'un ou plusieurs introns et pourra donner naissance à un nouveau gène, qui pourra dégénérer en pseudogène, dont les génomes de mammifères sont particulièrement remplis (Figure 43).

Compte tenu des nombreuses hybridations interspécifiques présentes dans toutes les branches du vivant, des transferts horizontaux, des invasions d'éléments transposables et des duplications totales ou partielles de génomes, il est très difficile de prédire le nombre de gènes d'un génome eucaryote en fonction de sa taille – et encore moins l'inverse. Néanmoins, on observe que le nombre de gènes d'une espèce évolue grossièrement en fonction du logarithme de la taille de son génome, même si cette corrélation est loin d'être parfaite et si de nombreuses espèces (en particulier de plantes) échappent à cette règle (Figure 44).

Tous les mécanismes d'évolution des génomes décrits ci-dessus sont non-darwiniens, en ce sens qu'ils impliquent des changements non incrémentaux mais saltatoires. Ce sont des moteurs puissants de l'évolution et la séquence des génomes modernes révèle l'ubiquité de leur empreinte dans toutes les branches du vivant. La nature ne fabrique pas des arbres, comme Darwin l'avait supposé et comme les phylogénistes du xx^e siècle les ont dessinés. La transformation et l'évolution des êtres vivants ressemblent plus à un réseau qu'à une forêt (Figure 34). Au sein de ce réseau, on trouve à la fois des transmissions verticales de mutations qui vont être soumises à la dérive génétique et à la sélection darwinienne et des transmissions horizontales, interspécifiques, de gènes, de larges fragments chromosomiques contenant plusieurs gènes, voire de génomes entiers. Les biologistes du xxi^e siècle doivent maintenant intégrer ces différents

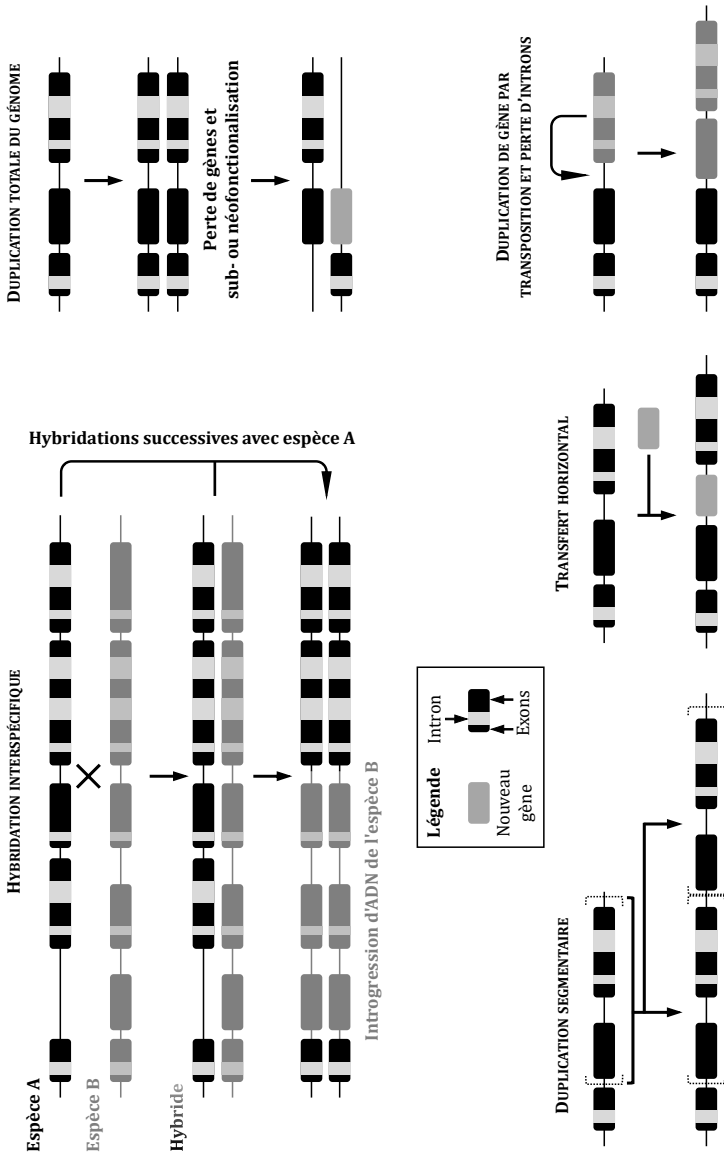


Figure 43 | Les mécanismes d'évolution rapide des génomes.

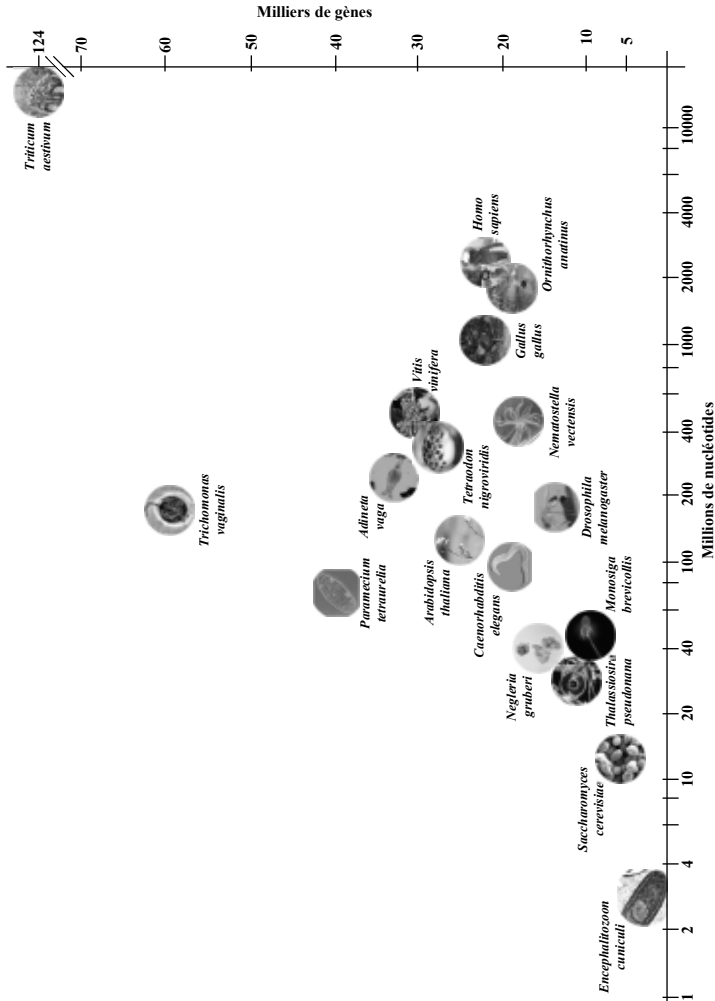


Figure 44 | Nombre de gènes en fonction des tailles des génomes eucaryotes. L'axe des abscisses représentant la taille des génomes est en échelle logarithmique.

mécanismes dans un modèle synthétique qui permette enfin de représenter le vivant et son évolution de la façon la plus exacte possible. L'apport des mathématiques à cette complexe problématique devrait être déterminant pour sa résolution.

À RETENIR

- La reproduction sexuée a été inventée à l'origine du monde eucaryote, suite à la capture d'un gène viral permettant la fusion des membranes cellulaires.
- Malgré le coût de l'entretien des mâles pour une espèce, le sexe permet de mélanger rapidement les allèles grâce à la méiose, augmentant rapidement la diversité génétique.
- Les hybridations interspécifiques et d'autres mécanismes non-darwiniens permettent de générer de la diversité en créant rapidement de nombreux nouveaux gènes.

BIBLIOGRAPHIE

- Les hybridations interspécifiques chez les levures : Nguyen, Legras, Neuvéglise et Gaillardin, *PLoS ONE* 6 : e25821 (2011).
- Néofonctionalisation chez la lycode : Deng, Cheng, Ye, He et Chen, *Proceedings of the National Academy of Sciences USA* 107 : 21593-21598 (2010).
- Le génome d'*Adineta vaga* : Simion, Narayan, Houtain, Derzelle *et al.*, *Science Advances* 7: eabg4216 (2021).

16

La fabrique du vivant

Les êtres vivants ne sont pas parfaits. Comme le disait François Jacob (1920-2013), lauréat du prix Nobel de Physiologie ou Médecine en 1965, « l'évolution bricole ». Le vivant réutilise ce qui marche pour créer de nouvelles fonctions. La néofonctionnalisation et la subfonctionnalisation des gènes dupliqués, que nous avons précédemment évoquées, en sont un exemple. Le développement du cerveau des primates, greffant un néocortex chargé des opérations cognitives sur le cerveau ancestral reptilien, a relégué celui-ci du rôle central qu'il occupait précédemment à un rôle subalterne de maintien de l'**homéostasie*** et de contrôle des activités émotionnelles. Un bricolage majeur de l'évolution ! Grâce à notre prodigieuse imagination et aux outils moléculaires développés au siècle dernier permettant

Homéostasie : Chez un être vivant, processus de régulation par lequel l'organisme maintient les différentes constantes de son milieu intérieur (température, pression osmotique, oxygénation, glycémie, acidité, etc.) dans les limites nécessaires à sa survie.

de manipuler l'ADN, nous pouvons maintenant construire, *de novo* ou presque, des organismes plus efficaces et plus optimisés que ceux qui ont été sélectionnés au cours de milliards d'années de bricolages évolutifs. La facilité de synthèse de courtes séquences d'ADN, leur assemblage par des techniques de biologie moléculaire et leur introduction dans des organismes vivants permet non seulement de manipuler les génomes mais d'en créer de nouveaux avec une facilité déconcertante et difficilement imaginable pour le profane. Évidemment, certains génomes se prêtent plus à ces manipulations que d'autres, non seulement pour des raisons techniques mais également éthiques. Parmi les génomes les plus faciles à manipuler, vient en tête celui de la levure de boulangerie, *Saccharomyces cerevisiae*, que nous avons déjà présenté comme un des plus puissants modèles de la génétique moléculaire moderne. C'est dans cette levure que va naître le premier gène synthétique, puis ce qu'il est convenu d'appeler maintenant la « génomique synthétique », cette discipline à la frontière entre génétique et ingénierie, qui découle du changement d'échelle et de la rationalisation des techniques de biologie moléculaire inventées une génération plus tôt.

LA RÉVOLUTION DE LA BIOLOGIE MOLÉCULAIRE

La biologie moléculaire est née au milieu des années 1970, suite à la découverte des systèmes de restriction enzymatique bactériens. Les bactéries possèdent un système efficace pour se défendre contre une invasion d'ADN étranger, par exemple lors d'une infection par un phage. Elles modifient leur ADN chromosomique en ajoutant un groupement méthyle (CH_3) sur certaines bases, grâce à une enzyme spécifique, une méthyltransferase qui va reconnaître une séquence longue de quelques nucléotides. À ce système de modification de l'ADN par méthylation est associée une **endonucléase*** qui reconnaît la même séquence non méthylée. Cette endonucléase va restreindre l'invasion d'ADN étranger en coupant tout ADN qui n'est pas méthylé sur cette séquence particulière. Lorsque l'ADN étranger entre dans

la cellule, par exemple lors de l'infection par un phage, cet ADN n'est pas méthylé, il est donc reconnu par la bactérie hôte comme de l'ADN étranger et l'endonucléase bactérienne va le couper à tous les sites qui ne sont pas méthylés, entraînant sa disparition et l'arrêt de l'infection (Figure 45).

Endonucléase : Enzyme qui coupe l'ADN (nucléase) à l'intérieur (endo) de la molécule. Il existe plusieurs familles d'endonucléases qui reconnaissent des sites de coupure de longueurs et de séquences différentes.

La première endonucléase de restriction fut purifiée en 1970 par Thomas Kelly et Hamilton Smith à partir de la bactérie *Haemophilus influenzae*. Celle-ci reconnaît un site de six bases dont la séquence est GTYUAC. Dans la nomenclature de l'ADN, Y est utilisé pour désigner une pyrimidine, c'est-à-dire C ou T, U désigne une purine, c'est-à-dire A ou G. La séquence de reconnaissance de l'endonucléase peut donc se lire : GT(C ou T)(A ou G)AC. Il existe plusieurs systèmes de restriction-modification et à l'heure actuelle des dizaines d'endonucléases de restriction sont purifiées et disponibles commercialement. Cette découverte marqua le début de l'ère de la biologie moléculaire, cette branche de la biologie qui allait utiliser des enzymes découvertes dans des bactéries ou des virus afin de manipuler l'ADN, le couper, le modifier, le coller et le propager dans d'autres cellules par clonage, donnant naissance à ce que l'on a appelé dans les années 1980 « le génie génétique ». À la découverte des enzymes de restriction et de modification, allait s'ajouter en 1983 l'invention d'une méthode d'amplification de l'ADN par Kary Mullis (1944-2019), l'amplification en chaîne par polymérase (*Polymerase Chain Reaction* ou PCR en anglais). Cette technique permet d'obtenir des milliards de molécules d'un fragment d'ADN particulier, à partir d'une seule molécule, en seulement quelques heures et à un coût dérisoire. La PCR a complètement révolutionné la biologie moléculaire en permettant d'amplifier et de manipuler des séquences

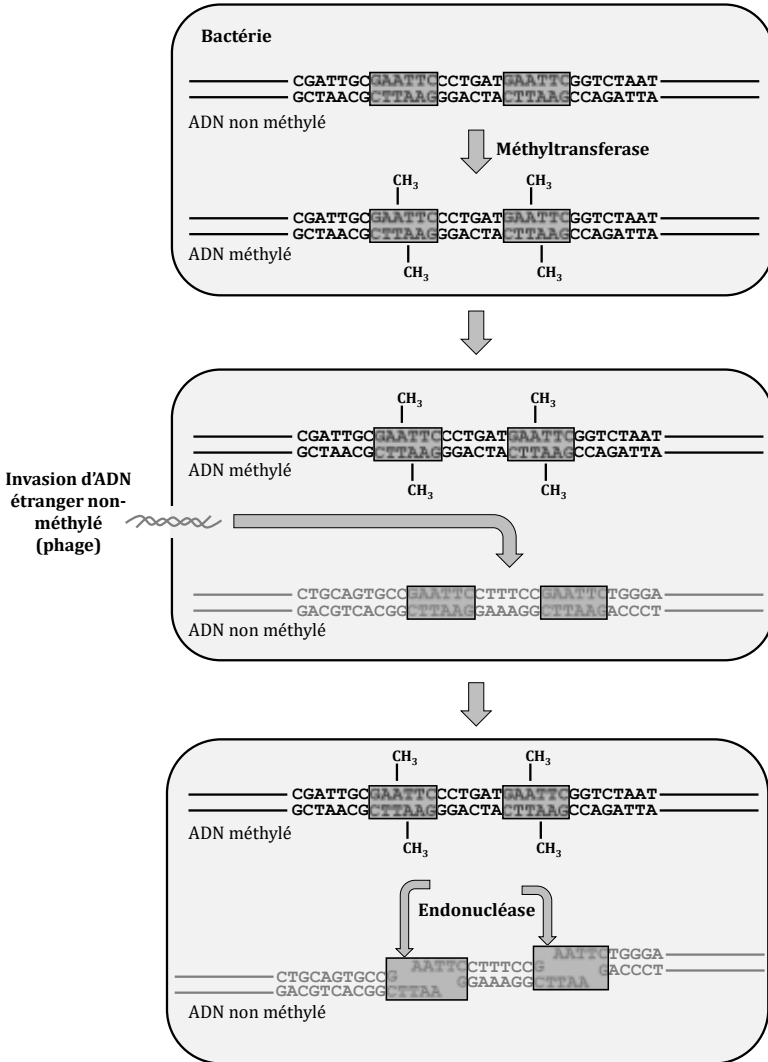


Figure 45 | Le système bactérien de restriction et modification.

d'ADN sans avoir besoin de passer par des étapes parfois fastidieuses d'amplification dans des bactéries.

LA NAISSANCE DE LA BIOLOGIE DE SYNTHÈSE

La biologie synthétique, ou biologie de synthèse, consiste à utiliser des principes d'ingénierie appliqués à la biologie moléculaire afin de créer, de modifier ou de manipuler le génome d'un être vivant dans le but de lui apporter de nouvelles caractéristiques génétiques dont il était auparavant dépourvu. La biologie synthétique n'est que le prolongement logique et rationalisé de la biologie moléculaire. Le concept en est né à la fin des années 1990 au célèbre MIT, le Massachusetts Institute of Technology, d'interactions entre biologistes et informaticiens. La biologie synthétique a pour but d'assembler de façon rationnelle des molécules d'ADN aux fonctions spécifiques, utilisées comme des briques élémentaires afin de construire des systèmes biologiques complexes. Elle entend se distinguer du génie génétique en ce qu'elle prône la construction de systèmes biologiques prévisibles et robustes à la façon dont un ingénieur construirait un système mécanique ou robotique. En pratique, elle utilise les mêmes outils que ceux du génie génétique, les outils développés par la biologie moléculaire des années 1970-1980.

On peut distinguer deux courants forts dans la biologie synthétique moderne. Une part importante des recherches actuelles porte sur le « recâblage » de voies métaboliques particulières, dans des microorganismes simples (bactéries ou levures) afin de faire fabriquer par ces organismes des molécules d'intérêt biomédical ou biotechnologique. Un bon exemple est la fabrication d'artémisinine, un médicament anti-paludisme, dont la découverte par Tu Youyou (1930-), chimiste chinoise, lui valut le prix Nobel de Physiologie ou Médecine en 2015. L'artémisinine est normalement extraite des feuilles d'une plante, *Artemisia annua* ou armoise annuelle, utilisée depuis des siècles en médecine chinoise. L'extraction de l'artémisinine est longue et coûteuse, ce qui pose un problème pour le traitement sur

le long terme des populations affectées par le paludisme. En 2006, une équipe de l'université Berkeley en Californie a réussi à modifier une souche de la levure *Saccharomyces cerevisiae* à l'aide d'outils de biologie synthétique, afin qu'elle fabrique un précurseur de l'artémisinine. La souche synthétique ainsi modifiée était capable de produire ce précurseur avec un titre suffisamment élevé pour envisager de remplacer tout ou partie de la production mondiale d'artémisinine extraite et produite de façon conventionnelle. En 2013, Sanofi lança la production industrielle d'artémisinine synthétique dans *S. cerevisiae*, dans le but de produire environ un tiers des besoins annuels mondiaux. D'autres compagnies, depuis des start-up ambitieuses jusqu'à de grosses industries, utilisent également des organismes génétiquement et rationnellement modifiés par la biologie synthétique, afin de produire des composés utiles pour l'agroalimentaire ou la médecine.

Le deuxième courant de la biologie synthétique moderne est celui qui consiste à redessiner et à construire des génomes synthétiques. Le généticien de la fin du xx^e siècle se trouvait dans la position d'un bricoleur chargé de comprendre le fonctionnement d'une voiture qu'il voyait pour la première fois de sa vie. Le généticien-bricoleur faisait alors ce qui lui paraissait le plus logique : démonter une à une chaque pièce de la voiture pour essayer d'en comprendre l'utilité. Certaines fonctions sont plus faciles que d'autres à appréhender. Le démontage d'un essuie-glace fournit rapidement une explication à sa fonction si on essaie de faire rouler la voiture sous la pluie. D'autres sont moins évidentes. Le démontage du vide-poches n'empêche pas la voiture de rouler, c'est un élément dispensable, non essentiel à la fonction première du véhicule. Pendant environ un demi-siècle, le généticien-bricoleur a donc retiré une à une, parfois par deux ou trois, le plus de pièces possible de sa voiture. Dans certaines voitures assez simples, de nombreuses pièces ont ainsi pu être identifiées et les fonctions essentielles répertoriées. Le généticien-bricoleur, muni de ces informations, se retrouve maintenant au début du xxi^e siècle

avec toutes les pièces à sa disposition pour construire un véhicule, de même marque et de même modèle que celui qu'il a appris à démonter pièce par pièce au cours de sa vie précédente. C'est avec cette idée en tête que les premiers projets de construction de génomes synthétiques ont débuté : sommes-nous capables, à partir de briques élémentaires dont la fonction biologique est connue, de fabriquer un organisme vivant ressemblant précisément à celui dont les briques sont issues ?

LES PREMIERS GÈNES SYNTHÉTIQUES

Les tout premiers gènes synthétiques ont été assemblés à la fin des années 1980. La technologie permettant la synthèse chimique d'**oligonucléotides*** avait suffisamment progressé pour permettre la synthèse de longs fragments d'ADN qui pouvaient être appariés entre eux puis ligaturés afin de créer une séquence de plus grande taille. Parmi les premiers gènes synthétiques à avoir été fabriqués, on trouve un inhibiteur d'une protéine kinase et une endonucléase de levure.

Oligonucléotide : Un oligonucléotide est un court morceau d'ADN (en général 20-100 nucléotides) synthétisé chimiquement. De nombreuses compagnies de biotechnologie fournissent de nos jours des oligonucléotides à un prix dérisoire (environ 10 centimes d'euros par nucléotide, soit 200 euros pour un gène de levure de taille moyenne).

Une protéine kinase est une protéine qui ajoute un groupement phosphate sur une autre protéine afin de modifier son activité, sa localisation ou son interaction avec d'autres protéines. Le génome humain code plus de 500 protéines kinases différentes. Le gène humain *PKI* code un inhibiteur d'une protéine kinase qui se fixe à sa cible avec une très haute affinité. La protéine PKI avait été purifiée depuis des années et sa séquence en acides aminés avait été déterminée, mais le gène codant cette protéine était encore inconnu. En 1987, une équipe américaine utilisa une série de 14 oligonucléotides longs de 17-25 bases afin de fabriquer *de novo* la séquence codant

les 31 premiers acides aminés de PKI, contenant l'essentiel de l'activité de cet inhibiteur. Une séquence mutante, codant une protéine différant de deux acides aminés, fut également construite. Les deux ADN synthétiques furent exprimés dans la bactérie *Escherichia coli* afin de purifier les protéines correspondantes, puis dans des cellules humaines, afin d'étudier leur fonction précise.

Dans le même temps, en France, à l'Institut Pasteur, l'équipe de Bernard Dujon construisait le premier gène synthétique venant de la mitochondrie de la levure *Saccharomyces cerevisiae*, codant une endonucléase très spécifique, nommée I-Sce I. Cette protéine reconnaît et coupe une séquence d'ADN longue de 18 nucléotides, beaucoup plus longue que les séquences reconnues par les endonucléases du système de restriction-modification évoquées plus tôt. Pour cette raison, I-Sce I et les autres protéines de la même famille qui furent découvertes plus tard furent baptisées « méganucléases ». Le gène codant I-Sce I étant d'origine mitochondriale, il suivait donc les règles de ce code génétique particulier (voir Chapitre 6). Il fallait donc le modifier pour l'adapter au code universel. La synthèse du gène synthétique fut réalisée à partir de huit oligonucléotides de synthèse mesurant 62-137 nucléotides de long, en cinq étapes successives qui aboutirent à la construction d'un gène synthétique de 754 nucléotides qui fut exprimé avec succès dans nombre de bactéries, levures, plantes et animaux, jusqu'aux cellules humaines.

Ces travaux pionniers marquent symboliquement les débuts de la génétique synthétique, qui allait devenir quelques années plus tard la génomique synthétique, avec la synthèse du tout premier génome bactérien, puis du tout premier génome eucaryote.

LES PREMIERS GÉNOMES SYNTHÉTIQUES

Au début du XXI^e siècle apparaît l'idée qu'il est possible de fabriquer non seulement des gènes synthétiques mais des génomes entiers synthétiques. John Craig Venter (1946-), entrepreneur américain dans le domaine des biotechnologies, fut l'un des pionniers du séquençage

des génomes, d'abord bactériens, puis humain. À la tête du J. Craig Venter Institute et avec l'aide de Hamilton Smith, une équipe emmenée par Daniel Gibson réussira la synthèse du premier génome procaryote et son transfert dans une cellule bactérienne. C'est le génome de *Mycoplasma genitalium*, une bactérie parasite des cellules épithéliales du tractus uro-génital humain qui sera entièrement synthétisé à partir de briques élémentaires mesurant 5 000-7 000 nucléotides et finalement assemblé dans la levure *S. cerevisiae* sous la forme d'un chromosome de plus de 582 000 nucléotides. L'assemblage du chromosome entier eut lieu initialement dans une levure pour des raisons pratiques, mais le but ultime était de transférer ce génome synthétique dans une bactérie receveuse afin de créer la première cellule vivante dotée d'un génome entièrement fabriqué par l'homme. Utilisant une technique de transfert de chromosome bactérien entier développée un peu plus tôt par une jeune chercheuse française, Carole Lartigue, dans l'équipe de Craig Venter, le génome synthétique entier d'un autre mycoplasme parasite de ruminants, *Mycoplasma mycoides*, sera transféré dans une bactérie receveuse d'une espèce proche, *Mycoplasma capricolum*. Le génome synthétique de plus d'un million de nucléotides de *M. mycoides* va remplacer le génome naturel de *M. capricolum* et la cellule le contenant sera capable de se diviser et montrera la même morphologie qu'une bactérie de l'espèce *M. mycoides*. Le premier organisme vivant utilisant un génome entièrement fabriqué par l'homme venait de voir le jour. Mais ce n'était que le début de l'aventure de la génomique synthétique.

Dans le même temps que l'équipe de Craig Venter assemblait le premier chromosome bactérien, un projet encore plus ambitieux voyait le jour aux États-Unis. Sous l'impulsion et la direction de Jef Boeke, à l'université Johns Hopkins de Baltimore, un consortium international composé de laboratoires américains, français, indiens et écossais réussit la synthèse et l'assemblage du premier chromosome d'un organisme eucaryote, le chromosome III de la levure de boulangerie. *S. cerevisiae* possède 16 chromosomes. Le chromosome III,

comportant les gènes régulant sa sexualité, contient 315 000 nucléotides et environ 200 gènes. Les deux premiers co-auteurs de cette publication majeure, un chercheur indien, Narayana Annaluru et une chercheuse française, Héloïse Muller, ont utilisé des outils similaires à ceux précédemment imaginés par l'équipe de Craig Venter, en y ajoutant toutefois quelques raffinements. Les transposons ont été supprimés du chromosome synthétique, ainsi que les ARNs de transfert qui ont été intégrés sur un autre chromosome. Tous les codons stop de séquence TAG ont été remplacés par des codons TAA, permettant de réutiliser TAG pour coder un autre acide aminé (voir ci-dessous). Mais le changement peut-être le plus fascinant est l'introduction entre chaque gène non essentiel de sites de reconnaissance pour la recombinase bactérienne *Cre*. L'expression transitoire de cette enzyme va générer des cassures chromosomiques multiples dans le chromosome synthétique. Ces bouts de chromosome vont être réparés par la cellule et réassemblés pour reformer un nouveau chromosome, en recollant entre eux les sites *Cre*. Les gènes essentiels, ne pouvant être perdus sans que la cellule ne meure, seront conservés mais les autres gènes pourront être perdus. Le nouvel assemblage peut se faire en recollant à l'envers certains fragments d'ADN, car les sites *Cre* étant symétriques, ils peuvent être recollés dans les deux sens. Le nouveau chromosome reconstitué aura perdu certains de ses gènes non essentiels et aura mélangé d'autres gènes, créant autant de nouvelles versions de lui-même ! Ce système, appelé *SCRaMbLE*¹ appliqué à l'ensemble du génome synthétique permettra à terme de déterminer expérimentalement le nombre de gènes minimum nécessaires à la vie et à la reproduction d'une cellule eucaryote et devrait apporter des éléments de réponse essentiels sur le rôle de la syntenie dans la structure, l'expression et la fonction de ce même génome (Figure 46).

1. Pour *Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution* (Rearrangement et modification de chromosomes synthétiques par évolution contrôlée par LoxP)

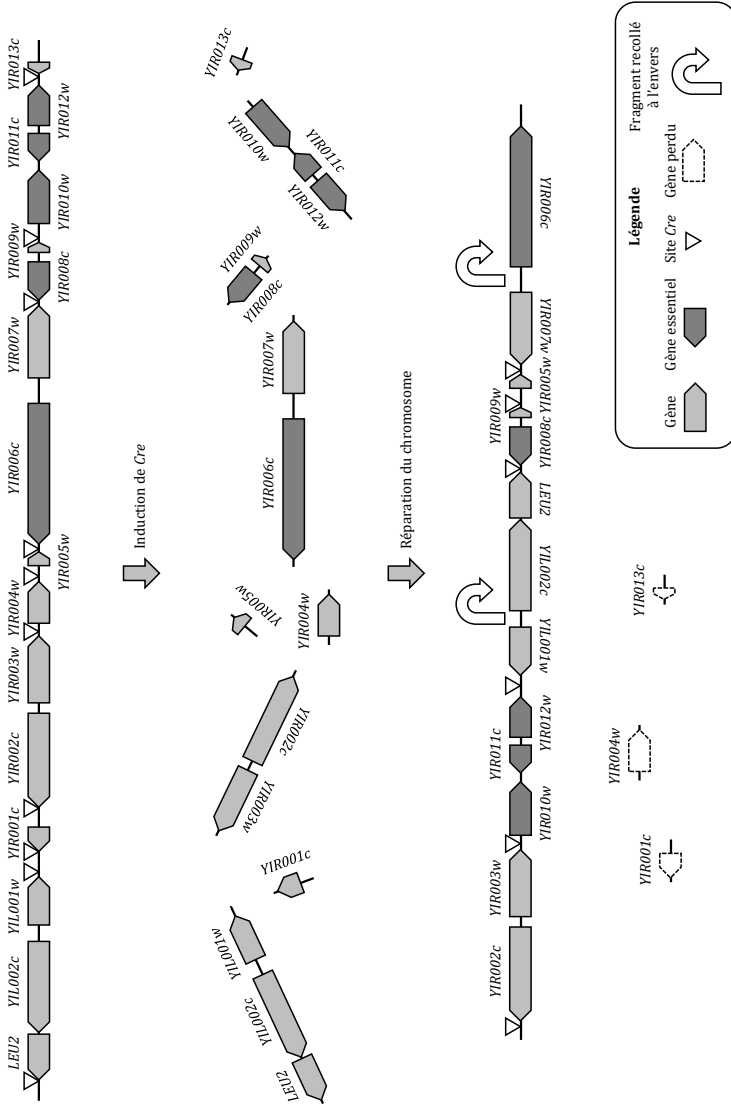


Figure 46 | Le système SCRaMBLE de la levure synthétique.

À l'heure actuelle, l'ensemble des autres chromosomes synthétiques de la levure est en cours de fabrication par le même consortium et une levure dont les 16 chromosomes ont été synthétisés et assemblés par l'homme devrait prochainement voir le jour. Les chromosomes de cette levure auront été redessinés et assemblés de façon rationnelle, en suivant des règles précises, définies dès le départ du projet et pouvant être appliquées à la construction d'autres chromosomes synthétiques, d'autres organismes eucaryotes.

POURQUOI SYNTHÉTISER DES GÉNOMES ENTIERS ?

On peut questionner l'intérêt de telles approches, chronophages et dispendieuses, en dehors de la compréhension fine des mécanismes régulant l'évolution et l'expression des génomes, qui ne semble intéresser que le monde académique. Ce serait se méprendre sur la finalité de ce type de recherches, particulièrement quand elles sont financées par des industriels ou par des consortiums public-privé. Le mycoplasme et la levure n'étaient que des échauffements. Sous l'impulsion de Georges Church, un généticien renommé de l'université Harvard à Boston et de Jef Boeke, le projet de synthétiser d'abord un chromosome puis un génome humain et de le transférer dans une cellule est en train de voir le jour. Cette perspective ouvre des horizons qui enflamment l'imagination. On sait comment déterminer la séquence de l'ensemble des gènes d'un homme mort, en utilisant l'approche précédemment décrite utilisée pour les hommes ancestraux. Pourrait-on donc ramener à la vie Albert Einstein, Léonard de Vinci ou Aristote ? Ou cette grand-mère que l'on n'a jamais connue ? Bien évidemment, il n'en est rien. En admettant que Georges Church et ses collaborateurs soient capables de réussir l'exploit de synthétiser un génome humain et de le transférer dans une cellule receveuse, cette cellule ne serait pas capable de reconstituer un être humain entier. La science d'aujourd'hui n'est pas encore capable de cet exploit, même s'il est envisageable qu'elle le soit un jour. Alors ressusciter Einstein, mythe ou possibilité ? La meilleure réponse que l'on puisse donner

à l'heure actuelle à cette question – sans tenir compte d'aucune des contraintes éthiques qui ne manqueraient pas de se poser – est la suivante.

On sait qu'un individu est plus que la somme des gènes qui le composent, en raison de modifications spécifiques de certains gènes, acquises avant et pendant le développement de l'embryon, qui vont réguler l'expression de ces gènes. Ces modifications sont dites « épigénétiques », car elles se superposent au patrimoine génétique en ajoutant une couche régulatrice supplémentaire. À l'heure actuelle, nous en sommes au balbutiement du déchiffrement de l'épigénome, mais il n'est pas impossible d'imaginer que d'ici quelques années on sera capable de le séquencer aussi facilement que le génome lui-même. Mais le problème n'est pas pour autant résolu, car un individu est plus que la somme de son génome et de son épigénome. On commence à peine à réaliser le rôle crucial du microbiote dans les comportements humains et ses interactions avec le système nerveux central (voir Chapitre 11). Un être humain est donc la somme de son génome, de son épigénome et de son microbiome. Mais il est également plus que ça. Comment quantifier et reproduire le rôle de l'éducation chez un être humain ? Comment reproduire le milieu familial dans lequel il a grandi et qui a modelé sa pensée ? Les neuropsychologues sont encore loin d'appréhender ce niveau supplémentaire de complexité. Ainsi, on peut affirmer sans crainte de se tromper qu'en admettant que l'on soit capable un jour, techniquement, de ramener un clone d'Einstein à la vie, on ne pourra jamais ramener Albert Einstein lui-même, ni à l'heure actuelle ni dans un futur éloigné.

En revanche, d'autres applications de la biologie synthétique sont envisageables et même sérieusement envisagées à la fois par le monde académique et par des compagnies pharmaceutiques ou agroalimentaires. En dehors des méganucléases, il existe d'autres endonucléases très spécifiques et plus faciles d'utilisation, capables de reconnaître et de couper un chromosome à un endroit précis, de façon plus ou

moins spécifique. La découverte récente et très médiatisée du système bactérien CRISPR-Cas9 par une chercheuse française, Emmanuelle Charpentier et une chercheuse américaine, Jennifer Doudna, et pour laquelle elles ont reçu le prix Nobel de Médecine ou Physiologie en 2020, a contribué à la formidable expansion des approches de modification des génomes à but thérapeutique. Ce système utilisé à l'origine par les bactéries fonctionne comme une endonucléase très spécifique capable de reconnaître une séquence unique dans un génome ciblé et de la couper. Utilisé par les bactéries pour se défendre des invasions de phages, il est maintenant employé dans tous les laboratoires au monde pour modifier les génomes eucaryotes à des endroits ciblés. Le séquençage rapide des génomes, la précision et la simplicité des endonucléases de type CRISPR-Cas9 et les techniques d'assemblage de chromosomes vont placer la génomique synthétique au centre des intérêts humains du *xxi*^e siècle de la même façon que la maîtrise de l'atome avait placé la physique nucléaire au centre de ceux du *xx*^e siècle. On ne doit pas redouter les développements futurs qui naîtront de ces recherches et qui seront très certainement fascinants pour l'espèce humaine. En revanche, il est évident qu'il faut disposer de garde-fous – au sens figuré comme au sens propre – afin d'éviter les problèmes éthiques qui pourraient se poser et entraver non seulement l'avancée de ces incroyables technologies mais également jeter un opprobre qui pourrait être fatal sur ce domaine encore balbutiant des sciences biologiques de synthèse.

Alors, si la fabrication, *de novo*, d'un homme synthétique paraît peu vraisemblable dans les décennies à venir, à la fois pour des raisons techniques et éthiques, il paraît en revanche à peu près certain que les hommes de demain seront partiellement synthétiques, certains de leurs allèles déficients remplacés par d'autres fonctionnels, permettant de s'affranchir de maladies génétiques graves. C'est le pari fait par des compagnies de biotechnologies dans le monde entier, une des entreprises pionnières dans ce domaine étant Collectis, créée par André Choulika à l'origine comme une start-up de l'Institut Pasteur

et qui fait maintenant partie des leaders mondiaux de l'ingénierie des génomes. Mais Cellectis est loin d'être la seule sur ce marché hautement compétitif et à très forte valeur ajoutée, avec en ligne de mire la guérison de certains types de cancers. Ce type de technologie a un coût encore très élevé, qui doit nous conduire à une réflexion profonde quant à son impact potentiel sur la société et aux différences de traitements qui seront disponibles pour ceux qui auront les moyens d'y mettre le prix. La démocratisation de ces soins génétiques n'est pas pour demain et il serait anormal et dangereux qu'une partie de notre société en soit exclue pour des raisons pécuniaires.

VERS LA CRÉATION DE NOUVELLES FORMES DU VIVANT

Mais nous n'en sommes qu'au début du futur. Le plus fascinant reste à accomplir. Notre généticien-bricoleur sait maintenant comment construire une voiture avec tous les éléments de base qui la constituent. Mais est-il capable de modifier ces mêmes éléments de façon profonde et irréversible, afin de créer de nouvelles voitures aux propriétés différentes ? C'est dans cette optique que certains ont envisagé de modifier des organismes vivants afin de leur faire fabriquer de nouvelles protéines, des protéines n'existant pas à l'état naturel car contenant un acide aminé supplémentaire.

L'addition d'un nouvel acide aminé requiert trois outils génétiques : (i) un codon disponible pour coder l'acide aminé ; (ii) un ARN de transfert dont l'anticodon est complémentaire de ce codon disponible ; (iii) une ARNt synthétase capable de charger le nouvel acide aminé sur l'ARNt. Une équipe californienne a donc introduit un codon stop TAG à l'intérieur d'un gène de résistance à un antibiotique, dans la bactérie *Escherichia coli*. La souche ainsi modifiée n'est pas résistante à l'antibiotique car l'ARNm contient ce codon stop qui interrompt prématurément sa traduction. Puis, ils ont introduit dans la même cellule un couple ARNt-tyrosine/ARNt synthétase-tyrosine provenant d'une archéobactérie, *Methanococcus jannaschii*.

Ce couple de **transgènes*** est capable de charger et d'incorporer dans une protéine un nouvel acide aminé modifié, la *O*-méthyl-tyrosine, une tyrosine modifiée par un groupement méthyl, qu'ils ont nommée tyrosine-X. Celle-ci est naturellement absente des protéines naturelles. Lorsque la souche bactérienne contenant les transgènes se divise en présence de tyrosine-X, elle montre une résistance à l'antibiotique utilisé dans le milieu de culture, car la tyrosine-X est incorporée au niveau du codon stop et donc la traduction du gène de résistance ne s'arrête pas. Cette résistance disparaît quand on retire l'acide aminé modifié du milieu. Par une méthode de détection physique des protéines, les auteurs de cette étude ont pu montrer que la tyrosine-X se trouvait effectivement bien incorporée dans la protéine conférant la résistance à l'antibiotique (Figure 47). Le même type d'expérience a depuis été reproduit dans la levure de boulangerie et plus récemment dans la souris avec un autre couple ARNt/ARNt synthétase.

Transgène : Gène étranger introduit dans le génome d'un organisme receveur plus ou moins génétiquement éloigné du génome de l'organisme donneur, par des techniques de biologie moléculaire.

Dans le même esprit, une autre équipe américaine a entrepris de modifier de façon rationnelle et dirigée une ARNt synthétase-glutamine d'*Escherichia coli* afin de lui faire charger un acide glutamique au lieu d'une glutamine. Après plusieurs modifications de la synthétase afin d'améliorer son affinité pour l'acide glutamique, l'ARNt glutamine lui-même a été modifié. Finalement, un couple hybride a été obtenu, entre une ARNt synthétase glutamine modifiée capable de charger un acide glutamique sur un ARNt glutamine modifié pour le recevoir. Une cellule exprimant ces deux gènes synthétiques serait donc capable d'incorporer un acide glutamique au lieu d'une glutamine à toutes les positions correspondant à l'ARNt glutamine modifié.

Ces changements étaient limités jusqu'à récemment à un petit nombre de gènes modifiés dans chaque organisme. Un premier pas

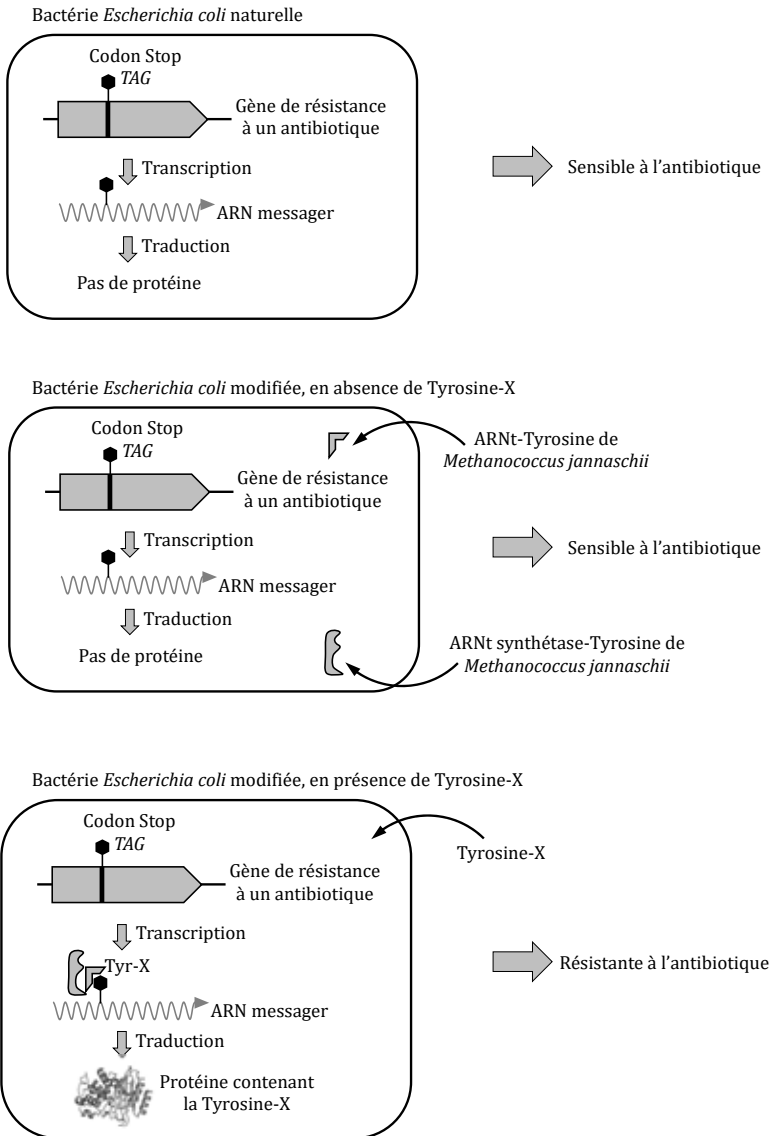


Figure 47 | La synthèse de nouvelles protéines contenant un 21^e acide aminé.

vers une nouvelle forme de vie complètement « recodée » a été fait par l'équipe de Georges Church, qui a entièrement modifié le génome de la bactérie *Escherichia coli* pour remplacer tous les codons stop TAG, par le codon stop TAA. Les 314 codons TAG à la fin d'autant de gènes bactériens n'existent donc plus dans cette bactérie de synthèse recodée. Il est maintenant possible de réintroduire le codon TAG dans certains gènes choisis et de l'utiliser pour incorporer un 21^e acide aminé, fabriquant ainsi un organisme capable de synthétiser de nouvelles protéines, des protéines n'ayant jamais existé dans aucun être vivant connu.

Ces problèmes de synthèse de nouvelles protéines ayant été en partie résolus, certains se sont logiquement tournés vers la création de nouvelles formes de support de l'hérédité, en particulier de nouvelles formes d'ADN. Sous la direction de Philippe Marlière (1955-), des chercheurs français et allemands ont réussi le tour de force de remplacer toutes les thymines contenues dans l'ADN d'*Escherichia coli* par une autre base, un analogue proche de la thymine, la chlorouracile. Dans un système de cultures en continu, des quantités contrôlées de cette nouvelle base sont régulièrement ajoutées au milieu de culture standard. Au bout de 1 000 générations (environ 170 jours de culture en continu), les bactéries sont capables de pousser dans un milieu contenant uniquement la nouvelle base et cette propriété est héréditaire, montrant qu'elle est bien portée par le chromosome bactérien. L'analyse de la composition de l'ADN de ces bactéries montre que plus de 90 % de cet ADN contient cette base, seules des traces de thymine étant encore détectables.

Dans le même esprit, une équipe américaine a réussi à propager dans *Escherichia coli* un morceau d'ADN circulaire contenant deux nucléotides modifiés (d5SICS et dNaM) à la place de la guanine et de son complément, la cytosine. L'ADN se réplique à l'intérieur de la bactérie avec presque la même efficacité qu'un ADN témoin. Cette molécule contient en plus des quatre bases habituelles (A, T, C et G) les deux nucléotides supplémentaires d5SICS et dNaM, ce qui en fait la première molécule contenant six bases différentes à se répliquer et à se propager dans un être vivant.

Mais le vivant n'a pas attendu la génétique de synthèse pour inventer de nouveaux nucléotides portant des bases différentes des quatre bases canoniques A, T, C et G. Certains phages contiennent dans leur ADN un cinquième nucléotide, la 2-amino adénine qui remplace l'adénine. Cette adénine modifiée (Z) s'apparie avec la thymine en formant trois liaisons hydrogène au lieu de deux, comme c'est le cas dans le couple A:T (Figure 12). Ainsi le génome de ces phages contient deux couples de bases, G:C et Z:T, qui forment trois liaisons entre elles. L'ADN qui en résulte montre donc une plus grande stabilité thermique. Jusqu'à très récemment, la voie de biosynthèse de Z était inconnue, de même que la manière dont cette base modifiée était reconnue et répliquée dans la cellule bactérienne hôte. Cette question a été élucidée par l'équipe d'Alexandre Kaminski à l'Institut Pasteur. La synthèse de Z utilise à la fois une enzyme codée par le génome du phage (PurZ) et une autre codée par le génome bactérien hôte (PurB). La répllication de cette base modifiée est assurée par une polymérase codée par le génome du phage (DpoZ). Comme vu précédemment, les bactéries disposent de systèmes de défense efficaces pour se prémunir de l'invasion de phages, entre autres le système de restriction-modification (Figure 45). Les phages dont l'ADN contient Z sont résistants aux endonucléases qui reconnaissent l'ADN non modifié, ce qui leur donne un avantage pour réussir à envahir efficacement leur hôte. De façon intéressante, Z a été retrouvé en 1969 dans les fragments d'une météorite tombée dans l'antarctique, suggérant que son existence remonte à la formation des toutes premières molécules organiques de la chimie prébiotique (voir Chapitre 1).

En admettant que nous ayons un jour les compétences scientifiques et technologiques nécessaires à la création *de novo* d'un être humain synthétique, il n'apparaît pas envisageable de le faire pour des raisons éthiques évidentes. En revanche, ces problèmes seraient moins prégnants pour la création d'un animal synthétique, que l'on pourrait nommer synthézoaire, du grec *sinthetos* qui signifie « composé, formé de » et *zoon* qui signifie « animal ». Cette nouvelle espèce ne serait d'ailleurs pas nécessairement issue d'une seule espèce naturelle, mais

pourrait parfaitement être le résultat d'un assemblage rationnel et déterministe de gènes provenant de plus d'une espèce, à condition qu'elles ne soient pas trop éloignées. Ce synthézoaire serait sans nul doute plus performant que la (ou les) espèce(s) dont il serait issu. En revanche, la survie du synthézoaire à une pression de sélection relativement forte reste posée. Dépourvu de « l'ADN inutile » dont sont encombrés la plupart des génomes animaux, le synthézoaire serait-il capable de s'adapter rapidement à un environnement changeant ? Rien n'est moins sûr. Il serait étonnant qu'un tel synthézoaire ne voie pas le jour dans un laboratoire avant le milieu de ce siècle. Son étude devrait permettre de répondre à certaines des questions que nous venons de poser.

Les possibilités de réécriture du vivant ouvertes par les expériences décrites ci-dessus donnent le vertige tant elles sont quasi infinies. Notre généticien-bricoleur va pouvoir arrêter de bricoler pour se mettre – enfin ! – à bâtir. De nouvelles formes de vie vont pouvoir être créées, infiniment plus rapidement que la nature n'aurait jamais pu le faire par le lent processus de la sélection naturelle. Les construire devrait nous donner suffisamment de recul et nous apporter suffisamment de connaissances pour compléter notre compréhension totale de l'évolution du vivant, depuis les premières molécules organiques jusqu'à nous. À moins que d'ici là, la découverte d'autres formes de vie extraterrestres ne nous ait fourni les clés de cette compréhension.

À RETENIR

- La biologie synthétique vise à modifier les génomes à l'aide d'outils moléculaires et d'approches rationalisées.
- Les premiers génomes synthétiques entièrement fabriqués et assemblés en laboratoire ont été ceux de la bactérie *Mycoplasma genitalium* et de la levure *Saccharomyces cerevisiae*.
- La génomique synthétique permet de modifier les génomes des êtres vivants afin de leur faire fabriquer de nouvelles protéines ou de les doter de nouvelles fonctions n'existant pas dans la nature.

BIBLIOGRAPHIE

Les bricolages évolutifs : Jacob, *Science* 196 : 1161-1166 (1977).

Le premier chromosome synthétique de levure : Annaluru, Muller, Mitchell, Ramalingam *et al.*, *Science* 344 : 55-58 (2014).

L'ADN ZTCG : Sleiman, Garcia, Lagune, Loc'h et al. *Science* 372 : 516-520 (2021).

17

La vie existe-t-elle ailleurs que sur Terre ?

Herbert George Wells (1866-1946), célèbre pour ses engagements politiques et ses romans de science-fiction, décrivait en 1895 une machine à explorer le temps dans le roman éponyme qui fut son premier succès littéraire. Depuis Albert Einstein, nous savons que l'univers est composé d'au moins quatre dimensions, mais que s'il est simple de se déplacer dans trois d'entre elles à une vitesse modérée, aucun être vivant n'a la capacité de se déplacer à son gré dans la quatrième, le temps. Le voyage dans le futur est théoriquement possible, pour un spationaute qui se déplacerait à une vitesse relativiste (au moins 10 % de la vitesse de la lumière, c'est-à-dire 30 000 kilomètres par seconde). Pour un tel voyageur, le temps ralentirait à l'intérieur de son vaisseau spatial alors qu'il s'écoulerait toujours à la même vitesse pour un observateur resté sur Terre. Ainsi, le spationaute vieillirait plus lentement, ce qui lui permettrait de se retrouver dans un temps futur par rapport à son âge réel à la fin de son voyage. Relativité du temps. Visiter le futur est donc théoriquement possible,

si la science était capable de fabriquer un engin spatial capable de se déplacer à une vitesse relativiste, mais ce serait une visite sans retour, car le voyage dans le passé est, lui, impossible. Nous ne saurons donc jamais précisément ce qu'il s'est passé sur Terre il y a environ 4,3 milliards d'années, lorsque les premières molécules organiques firent leur apparition. Ni il y a 3,7 milliards d'années, lorsque les premières cellules se formèrent. Ni il y a 2 milliards d'années lorsque les premiers eucaryotes apparurent par endosymbiose. Nous en sommes réduits à des hypothèses dont certaines – telle celle du Monde à ARN – sont heureusement fortement étayées par l'expérimentation et font office de théories scientifiques. En revanche, il reste un espoir, celui de voyager dans un autre passé, le passé d'une planète plus jeune que la Terre et sur laquelle des événements biochimiques similaires seraient en train de se produire en ce moment même. La découverte d'une telle planète ouvrirait de formidables perspectives pour comprendre l'histoire de la nôtre.

LES EXOPLANÈTES

Une exoplanète est une planète située en dehors du système solaire. Il existe plusieurs moyens de détecter de tels objets, mais l'un des plus efficaces est la méthode du transit. Si une planète passe devant son étoile, la variation d'intensité lumineuse du soleil va diminuer pendant tout le temps du passage. En observant des milliers d'étoiles pendant plusieurs années, on peut détecter de telles variations lumineuses périodiques, qui sont le signe de l'existence d'une possible exoplanète, dont la réalité doit être par la suite confirmée par d'autres méthodes. Mais la variation lumineuse est très faible et nécessite des instruments de mesure très performants. Pour un observateur distant, le passage de la Terre devant le soleil fait diminuer sa luminosité d'environ 0,01 % pendant 12 heures tous les 365 jours. Le télescope spatial *Kepler* a été lancé par la NASA en 2009 dans le but de rechercher par la méthode du transit les exoplanètes présentes dans une partie de notre galaxie, la Voie lactée. De 2009 à 2013, *Kepler* enregistra les

variations lumineuses de 150 000 étoiles toutes les 30 minutes. L'étude statistique fine des résultats obtenus pour 42 000 étoiles du type de notre soleil montre que 5,7 % de ces systèmes stellaires possèdent une planète de la taille de notre Terre orbitant avec une période de 200-400 jours (Figure 48). Extrapolant ce résultat à l'ensemble des étoiles de notre galaxie, on trouve le nombre de 10 milliards de planètes semblables à la Terre. En considérant de façon très prudente que seulement 5 % de ces planètes se trouvent dans la **zone habitable***, on atteint le nombre astronomique – c'est le cas de le dire ! – de 500 millions de planètes habitables semblables à la Terre dans notre galaxie ! Sachant que l'univers contient 100-200 milliards de galaxies, le nombre de planètes semblables à la nôtre et qui pourraient abriter des êtres vivants sous une forme ou sous une autre se compte donc en milliards de milliards.

Zone habitable : La zone habitable est définie comme la région de l'espace où les conditions sont favorables à l'apparition de la vie telle que nous la connaissons, en pratique ni trop près ni trop loin du soleil.

Les premières molécules prébiotiques ont pris naissance au sein des comètes aux origines de notre système solaire (voir Chapitre 1). Il est donc plus que vraisemblable – les mêmes causes créant les mêmes effets – qu'elles aient également pris forme dans des milliards d'autres systèmes solaires au moment de leur formation. La vie existerait donc partout dans l'univers, c'est une quasi-certitude, mais sous quelle forme ? Il est possible qu'elle s'articule comme sur Terre autour de l'atome de carbone, mais ce n'est pas la seule possibilité. Le silicium a été proposé (dans des œuvres de science-fiction) car c'est un élément très abondant, beaucoup plus que le carbone et il est doté de propriétés chimiques similaires à celui-ci, comme la possibilité de former quatre liaisons avec d'autres atomes.

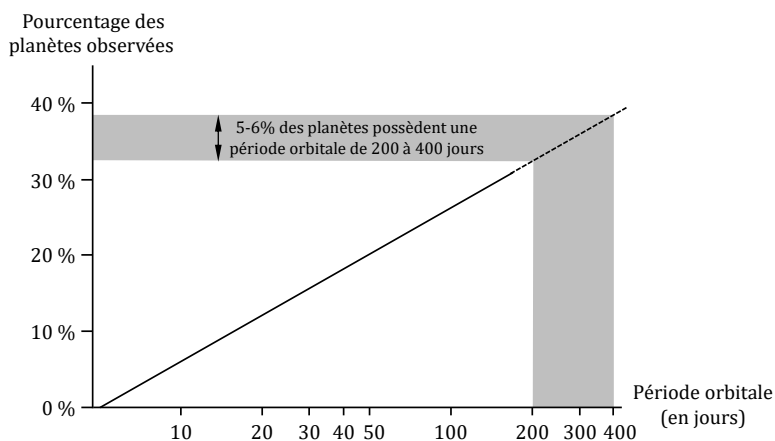
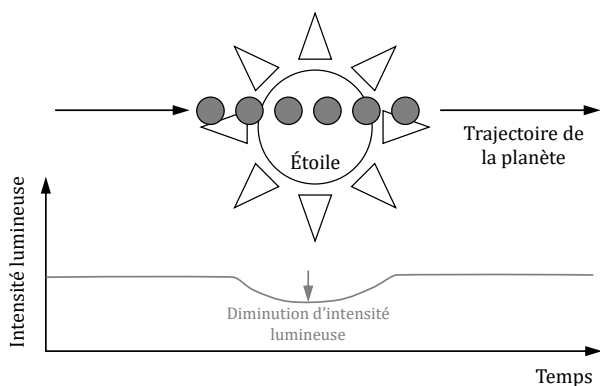


Figure 48 | La recherche d'exoplanètes. Haut : Méthode du transit. Bas : Estimation du nombre de planètes ayant une période orbitale de 200-400 jours terrestres.

UNE VIE PRÉBIOTIQUE DANS NOTRE SYSTÈME SOLAIRE ?

La mission spatiale Cassini-Huygens, développée par la NASA, l'Agence spatiale européenne (ESA) et l'Agence spatiale italienne (ASI), avait pour mission d'explorer la planète Saturne et ses lunes, en particulier Titan. Lancée en 1997, la sonde est restée en orbite autour

de Saturne de 2004 à septembre 2017, fournissant une incroyable moisson de données sur la planète géante, ses anneaux et ses lunes, en particulier Phœbé, Titan et Encelade. Cette dernière a un diamètre d'environ 500 kilomètres et possède plusieurs caractéristiques étonnantes mises en évidence par Cassini. La surface d'Encelade est recouverte de neige et de glace d'une épaisseur de 30-40 kilomètres. Des éruptions de particules de taille nanométrique¹ composées de silicates (SiO_2) entourés de glace ont lieu régulièrement à sa surface, les projetant dans l'espace et formant l'un des anneaux de Saturne, l'anneau appelé « E ». L'existence de ces grains est expliquée par la concentration importante de silicates au cœur d'Encelade dans des conditions de chaleur et de pression élevées (> 90 °C, 100-500 bars) et à pH fortement alcalin. À environ 200 kilomètres du noyau d'Encelade, se trouve un océan d'eau salée maintenue sous forme liquide grâce à la chaleur émanant du cœur du satellite. La température et la pression y sont plus faibles et les nanograins de silicates peuvent se former et être entraînés jusqu'à la surface où ils sont finalement éjectés, formant un anneau tout autour de la planète géante. Ces remarquables observations montrent qu'Encelade possède au moins trois des caractéristiques importantes à l'apparition de la vie : de l'eau sous forme liquide, une source de chaleur permettant les réactions prébiotiques et des particules de silicates à la surface desquelles peuvent se produire ces réactions, comme proposé dans la théorie de la pseudopanspermie (voir Chapitre 1). L'excitante possibilité existe donc que des molécules organiques plus ou moins complexes puissent être découvertes sur Encelade. Pour répondre à cette intrigante question, il va maintenant falloir attendre une mission d'exploration plus approfondie qui, avec les niveaux actuels de financement des agences spatiales, reste très hypothétique.

Plus récemment, la mission spatiale *Dawn* de la NASA a permis de survoler à basse altitude (385 kilomètres) la planète naine Cérés, le plus massif des corps en orbite dans la ceinture d'astéroïdes située

1. Un nanomètre est un milliardième de mètre, soit un millionième de millimètre.

entre Mars et Jupiter. *Dawn* a détecté dans une région bien définie à la surface de Cérès la présence de matériaux organiques contenant des groupements méthyle (CH_3) et méthylène (CH_2). Il est possible que ces composés organiques proviennent de l'impact d'une météorite, mais les auteurs de cette étude trouvent cette hypothèse peu crédible car de tels composés auraient sans doute été détruits par la chaleur de l'impact. Il est donc vraisemblable qu'ils aient été fabriqués au sein d'une chimie prébiotique propre à Cérès, qui pourrait nous éclairer sur les premières étapes de celle s'étant déroulée sur Terre il y a plus de 4 milliards d'années.

S'il est donc quasiment certain que des molécules carbonées prébiotiques, voire des cellules vivantes primitives, existent ailleurs que sur notre Terre, que ce soit sur un des satellites de Saturne ou dans un autre système stellaire de notre galaxie, en revanche la présence d'une forme vivante intelligente est bien moins évidente et bien moins certaine.

UNE AUTRE VIE INTELLIGENTE EST-ELLE POSSIBLE ?

Si l'on considère que les propriétés chimiques intrinsèques des molécules prébiotiques sont les mêmes partout dans notre galaxie, il est très vraisemblable que la chimie qui les accompagne ait produit les mêmes résultats. On peut imaginer que le Monde à ARN soit présent (ou aie été présent) sur des millions d'autres planètes et possiblement des cellules de type bactéries ou archées. Mais quelles sont les chances d'y trouver d'autres êtres vivants organisés et intelligents ? C'est la question très sérieuse que se sont posée de nombreux hommes de science depuis bientôt un siècle.

L'apparition d'un néocortex accru permettant la synthèse d'images mentales et le développement d'un raisonnement cognitif par l'utilisation de symboles et d'un langage articulé ne s'est produite sur Terre que dans le phylum des primates. Or, il est impossible en science de tirer des conclusions à partir d'un événement unique. Certains auteurs, comme Jacques Monod, suggèrent que cet événement

unique ne pouvait être que le fruit du hasard, comme un numéro sort à la roulette d'un casino : « *L'univers n'était pas gros de la vie, ni la biosphère de l'homme. Notre numéro est sorti au jeu de Monte-Carlo. Quoi d'étonnant à ce que, tel celui qui vient d'y gagner un milliard, nous éprouvions l'étrangeté de notre condition ?* ». Cette étrangeté de la condition humaine paraîtrait beaucoup moins étrange si la preuve pouvait être apportée que la vie existe effectivement dans d'autres systèmes stellaires. Il est vraisemblable que nous aurons un jour cette preuve. Restera à déterminer si cette vie a eu le temps et l'opportunité de développer une intelligence cognitive comparable à la nôtre. La question est cruciale car de même que l'on ne peut tirer de conclusion générale sur un événement en apparence unique, découvrir que la vie sur une exoplanète aurait également évolué vers une forme de vie intelligente rendrait très improbable l'intervention du hasard dans les deux cas. La fréquence d'un événement très rare multipliée par la fréquence d'un événement très rare est presque nulle. Il faudrait alors se rendre à l'évidence que l'apparition d'une intelligence cognitive est contingente à l'apparition de la vie elle-même.

Il est très difficile d'estimer le nombre de planètes qui pourraient abriter une vie intelligente capable de communiquer avec nous. Frank Drake (1830-), astronome américain, essaya de rationaliser cette question. Il proposa en 1961 l'équation suivante qui depuis porte son nom :

$$N = R^* \times f_{\text{planète}} \times n_{\text{étoile}} \times f_{\text{vie}} \times f_{\text{intelligente}} \times f_{\text{communiquer}} \times L$$

où

N : nombre de planètes abritant une vie intelligente capable de communiquer avec nous

R* : Étoiles formées annuellement

$f_{\text{planète}}$: Étoiles autour desquelles gravite au moins une planète

$n_{\text{étoile}}$: Nombre de planètes propices à la vie par étoile

f_{vie} : Nombre de ces planètes où apparaît effectivement la vie

$f_{\text{intelligente}}$: Nombre de celles-ci où apparaît effectivement une vie intelligente

$f_{\text{communiquer}}$: Nombre de celles-ci hébergeant une vie intelligente capable de communiquer

L : Longévité d'une civilisation, en années

En choisissant des valeurs pour les différents paramètres qui lui paraissaient probables à l'époque, il arrive à $N = 10$ civilisations capables de communiquer dans la Voie lactée. Si les premiers paramètres de l'équation peuvent être déterminés par l'observation, les suivants (f_p , f_i et f_c) sont plus compliqués à estimer et varient de plusieurs ordres de grandeur en fonction des auteurs. Nous avons vu plus tôt qu'il y avait environ 500 millions de planètes habitables dans notre galaxie. En remplaçant les premiers termes de l'équation de Drake par cette valeur, celle-ci devient :

$$N = 500.10^6 \times f_{vie} \times f_{intelligente} \times f_{communiquer} \times L$$

Si l'on considère, sans être trop optimiste, qu'il y a 0,1 % de chance qu'une forme de vie se développe sur une planète et 0,01 % de chance pour que celle-ci devienne intelligente, on obtient :

$$N = 50 \times f_{communiquer} \times L$$

La durée de vie d'une civilisation capable de communiquer est difficile à estimer. La nôtre ne l'est que depuis la fabrication du premier radiotélescope il y a environ 90 ans. À la vitesse à laquelle nous épuisons nos ressources, il est sans doute réaliste d'estimer que notre civilisation ne durera pas plus de 500 ans s'il nous est impossible de coloniser d'autres planètes. Le dernier terme de l'équation de Drake est sans doute le plus complexe à déterminer. Quelles sont les chances qu'une autre civilisation intelligente ait envie de communiquer avec nous ? Dans sa célèbre trilogie de science-fiction², Liu Cixin postule que notre galaxie est remplie d'êtres vivants aussi intelligents que nous mais qu'aucun d'entre eux ne souhaite faire connaître sa présence pour éviter d'entrer en conflit avec les autres. Dans ce cas, le terme $f_{communiquer}$ de l'équation deviendrait nul et nous ne saurons jamais s'il existe une autre vie intelligente dans notre galaxie (ou ailleurs). Néanmoins, il est vraisemblable d'imaginer que d'autres êtres intelligents seraient aussi curieux que nous de savoir s'ils sont seuls dans l'univers. On peut donc raisonnablement penser qu'au

2. Le problème à trois corps, La forêt sombre, La mort immortelle.

moins 1 % des civilisations suffisamment avancées auront le désir de se faire connaître. La valeur de N deviendrait alors :

$$N = 50 \times 0,01 \times 500 = 250$$

Il y aurait donc environ 250 civilisations intelligentes désireuses de communiquer avec les autres dans notre seule galaxie. Mais alors où sont-elles ? Cette question a été posée dans les années 1950 par le célèbre physicien italien Enrico Fermi, célèbre pour ses travaux sur la radioactivité, lors de discussions portant sur l'existence possible de civilisations extraterrestres. On peut résumer ce qu'il est convenu d'appeler le paradoxe de Fermi de la façon suivante : « Si d'autres civilisations intelligentes existent dans la galaxie, pourquoi n'en trouve-t-on pas de traces ? ». Ce paradoxe en est resté un en apparence pendant longtemps, et de nombreux auteurs ont contribué à lui trouver diverses explications. Dans l'état de nos connaissances actuelles, une réflexion portant sur le paradoxe de Fermi pourrait nous amener aux remarques suivantes :

- (i) La civilisation humaine n'envoie régulièrement des signaux dans l'espace que depuis environ un siècle.
- (ii) La civilisation humaine n'a pas les moyens technologiques de voyager entre les systèmes stellaires à une vitesse relativiste qui lui permette de rejoindre les étoiles les plus proches en moins d'une génération. Il lui faudra vraisemblablement plusieurs siècles avant d'en arriver à ce stade de développement, en supposant que cela soit possible avec les ressources à notre disposition.
- (iii) La démographie explosive de l'espèce humaine et son mode de consommation des ressources naturelles rendent peu vraisemblable sa survie, au niveau scientifique et technologique actuel, pendant encore des siècles. Sans même tenir compte de la catastrophe climatique en cours, d'une guerre nucléaire ou de l'impact d'un météore de taille à déclencher l'extinction de notre propre espèce, l'utilisation grandissante de nos ressources naturelles forcera l'humanité à entrer sous peu dans une période peu propice aux investissements nécessaires à des explorations extrasolaires.

(iv) En admettant qu'*Homo sapiens* survive en tant qu'espèce capable d'envoyer des messages dans l'espace pendant encore mille ans (ce qui paraît peu vraisemblable), la période au cours de laquelle nous aurons été capables de communiquer avec d'éventuelles autres civilisations intelligentes correspondrait à environ 24 millièmes de pourcent de l'histoire de notre système solaire et à 8 millièmes de pourcent de l'histoire de la Voie lactée. Cent ou mille civilisations extraterrestres de notre niveau technologique ont eu le temps d'apparaître et de disparaître avant nous et autant après nous. Les chances pour que nous arrivions, dans le même intervalle de temps de quelques siècles, à entrer en contact avec l'une d'entre elles sont statistiquement proches de zéro. D'autant plus si elles sont éloignées de la Terre de plus de 1000 années-lumière, le temps que mettrait un signal radio pour nous atteindre. Ce qui ne doit pas nous empêcher de continuer à essayer. C'est d'ailleurs le but du programme SETI*, qui vise à détecter d'éventuels signaux radio envoyés, volontairement ou non, par une civilisation extraterrestre.

SETI : Search for Extra-Terrestrial Intelligence. Une initiative américaine des années 1960 dont le but est d'essayer de détecter des signaux intentionnels ou non, émis par une civilisation extraterrestre.

En conclusion de ce dernier chapitre, s'il est presque certain que la vie doit exister – sous une forme ou sous une autre – sur des centaines de millions de planètes semblables à la Terre, il est bien moins sûr que ces formes vivantes aient abouti à une espèce dotée de facultés cognitives comparables aux nôtres. Quant aux chances que cette espèce ait atteint un niveau technologique similaire dans un espace-temps proche, qui nous permette de communiquer, elles sont quasiment nulles. *Homo sapiens* ne doit donc pas se sentir seul dans l'univers infini, mais à la fois comme le lointain héritier d'un ancêtre inconnu et comme le père d'une descendance qu'il ne rencontrera vraisemblablement jamais.

À RETENIR

- Notre galaxie doit contenir environ 500 millions de planètes semblables à notre Terre, dont 250 abritent peut-être une vie intelligente.
- Dans notre système solaire, il est possible qu'Encelade ou Cères contiennent des molécules de la chimie prébiotique.

BIBLIOGRAPHIE

La recherche d'exoplanètes : Petigura, Howard et Marcy *Proceedings of the National Academy of Sciences USA* 110 : 19273-19278 (2013).
L'origine du vivant : Monod, *Le hasard et la nécessité*, Essais, Éditions du Seuil (1970).

ÉPILOGUE : L'HUMANITÉ, D'HIER À DEMAIN

La vie est ou n'est pas. Elle porte en elle son unique justification. La seule fonction d'un organisme vivant est de se reproduire à l'identique. Ou presque. Ce « presque » est la raison pour laquelle ces lignes peuvent être écrites aujourd'hui, car si la reproduction du vivant se faisait strictement à l'identique, aucun changement n'aurait pu être possible et nos ancêtres seraient restés à l'état de molécules d'ARN sans avenir.

Certaines questions concernant l'origine et l'évolution de la vie ont été résolues depuis une vingtaine d'années grâce au séquençage et à l'analyse des génomes, d'autres restent ouvertes et nous en avons soulevé quelques-unes dans ce livre. L'une des plus fascinantes, aussi l'une des plus complexes à résoudre, est celle du (ou des) premier(s) organisme(s) cellulaire(s) vivant(s).

LUCA : LE DERNIER ANCÊTRE COMMUN UNIVERSEL

Comme vu précédemment, la théorie de la coalescence postule qu'il a existé à un moment donné dans l'histoire du vivant, il y a environ 3,7 milliards d'années, une cellule procaryote primitive qui a donné naissance à tous les autres organismes vivants. C'est ce qu'on appelle

en anglais *Last Universal Common Ancestor* ou LUCA. Se référant à ce que nous avons étudié au cours des différents chapitres de cet ouvrage, on peut avancer l'hypothèse que ce progénote (ou cénancêtre) devait être une bactérie anaérobie, dont l'information génétique était codée sous forme d'ARN ou peut-être d'ADN-U et qui possédant le code génétique savait donc synthétiser des protéines. Une équipe de chercheurs allemands a comparé plus de six millions de gènes provenant de 1 847 eubactéries et 134 archées et appartenant à presque 300 000 familles protéiques, afin de rechercher les gènes communs à toutes ces espèces, qui pourraient donc avoir été présents dans la première cellule vivante. Ils ont réussi à identifier 355 gènes répondant à ce critère. L'image que ces gènes donnent de LUCA est celle d'un organisme anaérobie strict, vivant dans des sources hydrothermales chaudes, capable d'utiliser l'hydrogène gazeux (H_2) comme donneur d'électron et le dioxyde de carbone (CO_2) comme accepteur final d'électron et comme source de carbone pour les synthèses organiques. Les protéines de ce métabolisme carboné contiennent des centres fer-soufre, reliques vraisemblables d'un métabolisme organique ancestral lié à un environnement riche en métaux. LUCA est capable d'utiliser le méthane environnant comme source de méthylation des ARNt et des ARN ribosomiques. Il possède un ribosome primitif et sept ARNt synthétases, il est donc capable de synthèse protéique. De façon inattendue, il possède déjà une thymidylate synthase (ThyX), suggérant que son génome était déjà composé d'ADN-T. De façon encore plus surprenante, le génome reconstruit contient également sept gènes impliqués dans le métabolisme de l'oxygène, alors qu'il est clair que l'ancêtre commun devait exister bien avant la Grande Oxydation et donc dans des conditions dans lesquelles il n'y avait pas d'oxygène atmosphérique métabolisable. Entre autres, la présence surprenante de l'enzyme superoxyde dismutase, présente chez presque tous les organismes aérobies pourrait suggérer que l'ancêtre commun devait déjà se protéger de l'oxygène, plus d'un milliard d'années avant l'augmentation notable de celui-ci dans l'atmosphère ! Ce résultat

très inattendu peut s'expliquer par la présence d'enzymes ancestrales dévolues initialement à d'autres réactions que celles des enzymes antioxydantes modernes, qui auraient été détournées de leur activité initiale lorsque l'augmentation du niveau d'oxygène atmosphérique a commencé à avoir des effets délétères sur les organismes anaérobies ancestraux. Alternativement, ce résultat met peut-être aussi en évidence les limites des reconstructions phylogénétiques, particulièrement dans le monde procaryote où les transferts horizontaux de gènes sont monnaie courante.

LUCA A-T-IL JAMAIS EXISTÉ ?

S'il est possible que LUCA ait existé en tant que tel, comme Jacques Monod le faisait remarquer, « *Nous n'avons, à l'heure actuelle, pas le droit d'affirmer, ni celui de nier que la vie soit apparue une seule fois sur la Terre, et que, par conséquent, avant qu'elle ne fût, ses chances d'être étaient quasi nulles.* » Les arbres phylogénétiques et la théorie de la coalescence ne peuvent mener qu'à LUCA, puisque par essence ils ont été construits pour cela. Néanmoins, la vie a pu naître à plus d'un endroit de façons indépendantes puis évoluer sous forme de réseau d'interactions entre différentes protocellules échangeant continuellement de l'information génétique et d'où finirent par émerger des lignées ancestrales qui nous semblent, en apparence, provenir d'un ancêtre commun, mais dont l'origine plurielle nous est masquée par l'abondance des transferts horizontaux de matériel génétique. La découverte au début du ^{xxi}e siècle de virus géants apporte un argument de poids en faveur d'une apparition multiple de la vie. Ces virus sont des parasites obligatoires d'amibes et on les trouve dans toutes les régions du monde. Les génomes de ces virus sont remarquablement grands pour des organismes de ce type, le plus grand, celui des Pandoravirus, mesure 2,8 millions de nucléotides, soit environ la moitié du génome de la bactérie modèle *Escherichia coli*. De façon remarquable, l'analyse de leurs gènes montre que seulement 7 % d'entre eux ont un homologue dans le reste du monde vivant. De

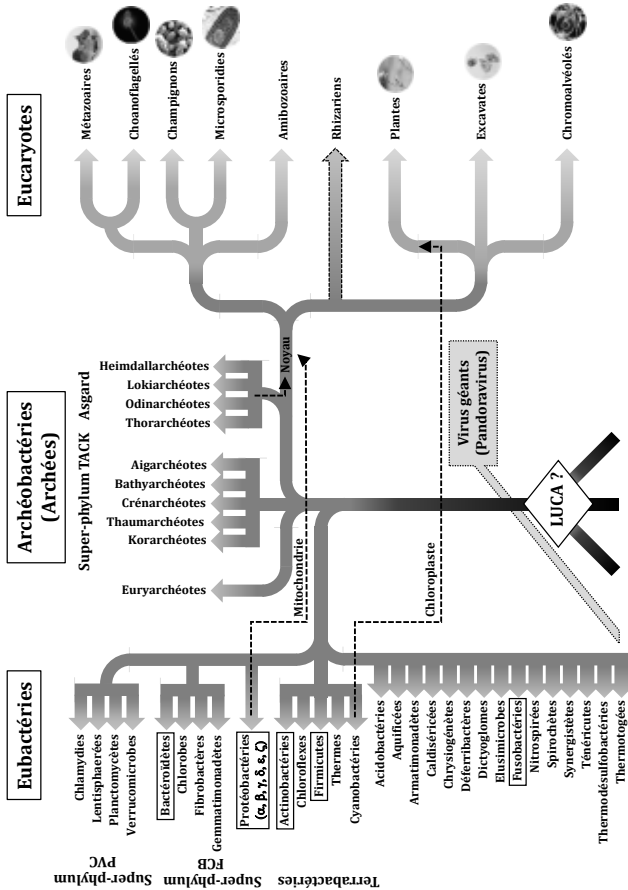


Figure 49 | Un arbre du vivant des phylums eubactériens et archéobactériens, ainsi que des principaux lignages eucaryotes. Le point de branchement des rhizariens (en pointillé) n'est pas précisément connu. Il est possible qu'il ait existé un ancêtre commun universel à tous les êtres vivants (LUCA), de même qu'il est possible qu'un quatrième règne du vivant, celui des virus géants, ait évolué parallèlement aux trois autres avec quelques échanges de gènes par transfert horizontal. Les bactéries les plus fréquemment retrouvées dans les microbiomes d'individus en bonne santé sont encadrées.

nos jours, lorsqu'un nouveau génome est séquencé, environ 15-20 % de ses gènes n'ont pas d'homologues dans le monde vivant et correspondent à de nouveaux gènes spécifiques de cette espèce. Dans le cas des virus géants, ce sont donc 93 % des gènes qui sont spécifiques de cette lignée, une observation unique à ce jour ! De plus, la comparaison des gènes codant les ADN polymérases impliquées dans la répllication du génome montre que celles des virus géants forment une branche à part dans l'arbre du vivant, bien distincte des trois règnes classiques, suggérant que ces virus pourraient correspondre à un quatrième règne, qui aurait évolué en quasi-indépendance des trois autres et expliquant le très petit nombre de protéines en commun avec eux. Seules des études détaillées de cette nouvelle famille de virus permettront de tirer au clair ces remarquables observations. En attendant de clarifier ces observations, les arbres phylogénétiques sont toujours dessinés en faisant remonter l'ancêtre de tous les êtres vivants à un unique et seul organisme (Figure 49).

LE LIBÉRALISME DU VIVANT

La doctrine libérale appliquée au vivant

Le libéralisme, en tant que doctrine politique, est né dans le giron libertaire du siècle des Lumières et a grandi pendant le début du siècle suivant, sous l'influence de penseurs tels qu'Adam Smith (1723-1790). Le libéralisme économique dans sa forme mondialisée et en l'absence de régulation mène à la disparition des plus petits acteurs de l'économie, rapidement avalés par de plus gros prédateurs. En moins d'une génération humaine, des empires financiers ou industriels ayant dévoré la plupart de leurs concurrents directs se sont formés dans les domaines de l'informatique, des télécommunications, de l'aviation civile, de la pharmaceutique, de la chimie, et demain en Europe dans ceux des transports ou de la production et de la distribution d'eau ou d'énergie. La libéralisation non régulée de l'économie aboutit à court ou moyen terme à la disparition des plus faibles au profit des plus forts. Quelques super prédateurs, dont il n'est pas la peine de rappeler

les noms, se partagent le monde économique, dévorant rapidement tout ce qui est digne d'être dévoré par eux, le seul mode efficace de survie consistant à rester suffisamment petit, tels nos ancêtres mammifères face à la toute-puissance des prédateurs dinosauriens.

L'être humain du XXI^e siècle est au centre de la doctrine libérale. Celle-ci lui confère deux droits fondamentaux et inaliénables : le droit à la propriété et le droit à la liberté, cette dernière étant le plus simplement définie comme tout ce qui ne nuit pas à autrui. Nous avons vu que le droit à la propriété imprégnait fortement le cerveau animal, en particulier celui des mammifères. Cette propriété se manifeste par le contrôle d'un territoire sur lequel une meute, une tribu, ou une société s'est établie. Chez les animaux non humains, il se caractérise également par l'appartenance pour les femelles au mâle dominant, seul apte à se reproduire et donc à propager ses gènes sur le territoire de la meute. Chez l'homme, la notion de couple implique que chacun des deux membres de ce couple « appartienne » à l'autre. L'irruption d'une tierce personne empiétant sur la propriété d'un des deux membres du couple va déclencher quasiment systématiquement une réaction d'agression envers l'intrus. Cela reste vrai que le modèle social soit basé sur le couple monogame ou sur une cellule familiale plus étendue, comme dans les sociétés polygames. Ainsi, malgré ses extraordinaires capacités cognitives et son très haut degré scientifique et technologique, la femme et l'homme – son inutile moitié – fonctionnent toujours sur le mode libéral de la propriété privée des moyens de reproduction. Mais qu'en est-il réellement de leur liberté de procréer ?

Nos gènes sont égoïstes

Comme l'écrivait le biologiste britannique Richard Dawkins il y a bientôt 50 ans, nous ne sommes que des machines égoïstes, des machines constituées d'une colonie de gènes dont le seul but est de se reproduire sans se soucier des conséquences de cette reproduction. Ce ne sont pas les individus qui sont sélectionnés, ce sont les gènes qui se transmettent ou non à la descendance. Cet égoïsme

génétique ne nuit en théorie pas à autrui, la reproduction de l'individu A n'empêchant pas celle de B. Au moins dans un monde aux ressources suffisantes pour permettre à chacun de se reproduire de façon égoïste. C'était le cas pour notre planète jusqu'à la moitié du ^{xx}^e siècle, mais depuis plusieurs décennies la donne reproductive a changé. La diffusion des pratiques d'aseptie dans la médecine et dans la vie de tous les jours, la découverte des antibiotiques, l'utilisation massive des vaccins et les politiques actives de veille médicale de la petite enfance ont permis de faire reculer la mortalité infantile d'un facteur sans commune mesure dans toute l'histoire de l'humanité. Celle-ci est tombée au cours du ^{xx}^e siècle de 150 enfants pour mille naissances en France à seulement 3,5 pour mille. La réduction de la mortalité infantile s'est accompagnée d'un allongement de la durée de vie, pour les mêmes raisons. Si la reproduction de l'espèce humaine continuait au même rythme, notre population – qui était seulement d'un milliard et demi au début du ^{xx}^e siècle – atteindrait le nombre incroyable de 16,5 milliards d'individus en 2100 ! (Figure 50)

Ainsi, dans un monde dans lequel la population humaine s'accroît de façon exponentielle, chaque être humain qui s'ajoute empiète forcément sur la capacité d'un autre être humain à faire de même sur le long terme, génération après génération. Grâce aux immenses progrès de la médecine du ^{xx}^e siècle la mortalité – en particulier infantile – a diminué de façon drastique dans toutes les sociétés humaines, avec pour résultat l'explosion démographique actuelle et ses effets sur la biodiversité et le futur de notre planète. L'égoïsme de nos gènes nous pousse donc à continuer à nous reproduire de façon effrénée et en apparence sans se soucier des lendemains. Mais cela est-il réellement le cas ?

On pourrait avoir l'impression que le vivant en général et nos sociétés humaines en particulier fonctionnent sur un mode libéral, reposant sur la propriété privée et la liberté de se reproduire. On pourrait même postuler – en jouant les provocateurs – que celui-ci s'est progressivement imposé comme la doctrine politique dominante sur

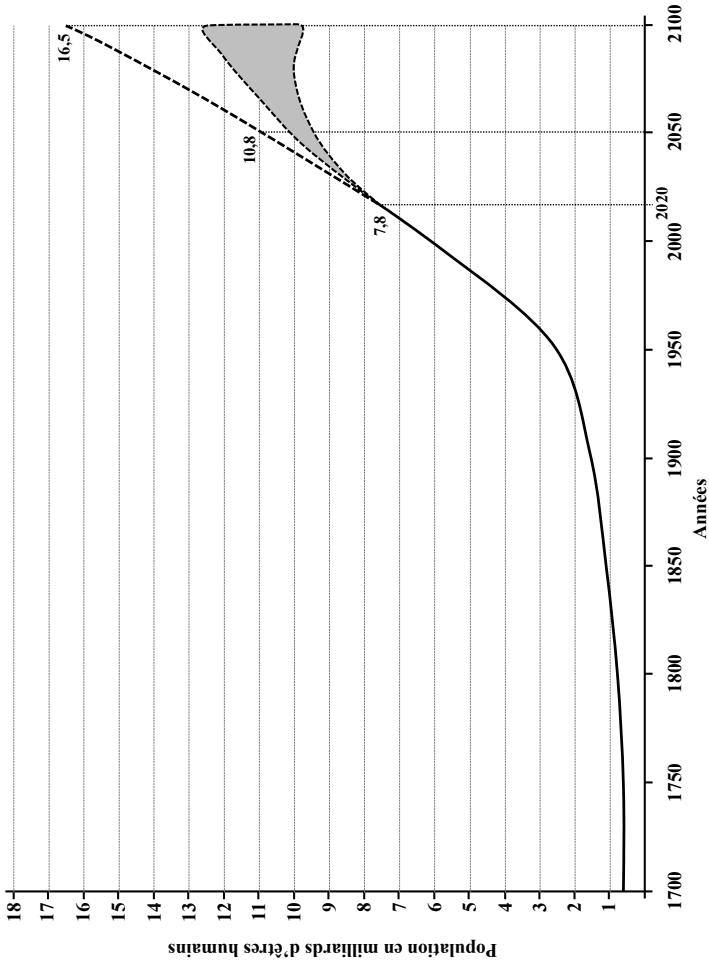


Figure 50 | L'accroissement de la population humaine. La valeur haute est indiquée ainsi que la zone correspondant à la probabilité de 80 % (en gris).

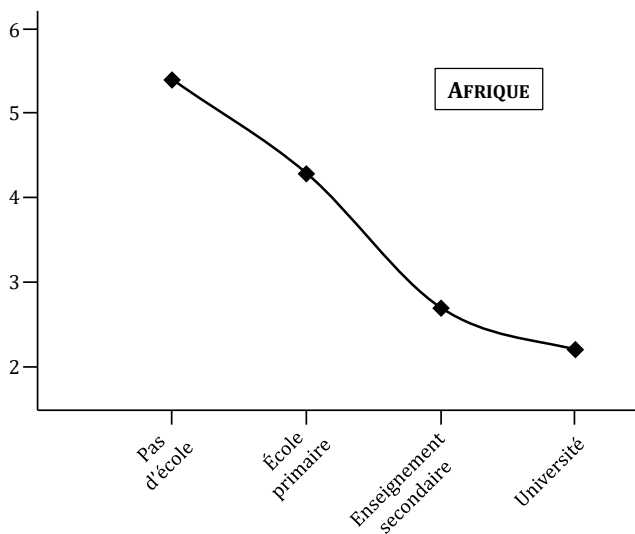
notre Terre du XXI^e siècle car, au fond, cette doctrine est inscrite dans nos gènes. Ce déterminisme génétique de la société libérale n'irait pas nécessairement à l'encontre du matérialisme historique de Marx et Engels, dans lequel la lutte des classes est héritée de l'histoire de chaque société humaine, comme les gènes le sont de chaque parent. Mais force nous est de constater que ce ne sont pas les individus dominants qui se reproduisent le plus.

Les individus les plus instruits se reproduisent moins

En partant du postulat que les individus dominants sont aussi les individus les plus diplômés, c'est même l'inverse qui se produit. En effet, il existe une corrélation très nette entre le niveau d'études et l'âge du premier enfant, les femmes les plus diplômées se reproduisant plus tardivement et engendrant un nombre moyen d'enfants plus faible que les femmes moins diplômées. Cela est vrai non seulement en France, en Europe et dans la plupart des autres pays développés, mais également dans les pays en voie de développement. Par exemple, en Afrique, les femmes n'ayant fait aucune étude ont en moyenne 5,4 enfants, celles qui sont allées à l'école primaire ont en moyenne 4,3 enfants, finir un cycle secondaire entraîne une chute remarquable à 2,7 enfants par femme et enfin pour celles qui sont allées à l'université, ce nombre descend à 2,2 enfants. Cette observation est également vraie dans notre pays, avec toutefois des variations moins marquées entre les différents niveaux sociaux (Figure 51). Le corollaire de cette remarquable observation est qu'un premier levier pour réduire les naissances dans les pays en voie de développement serait d'élever le niveau global d'étude de la population en général et des femmes en particulier.

Les gènes des individus dominants se transmettent donc moins, une observation qui, placée sous la loupe du néodarwinisme, pourrait suggérer que les individus avec un plus haut niveau d'études seraient contre-sélectionnés au sein d'une société. Ce n'est évidemment pas le cas.

Nombre d'enfants
par femme



Nombre d'enfants
par femme

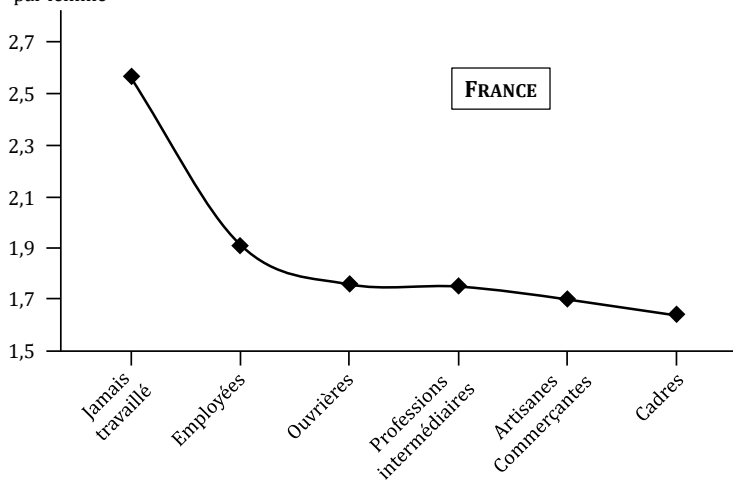


Figure 51 | Nombre moyen d'enfants par femme en fonction du niveau d'étude.

La liberté de se reproduire modère fortement la natalité

De tous les êtres vivants, seul *Homo sapiens* a la possibilité de s'affranchir des contraintes liées à sa biologie pour assurer la reproduction de son espèce. Celle-ci est entièrement dépendante de la volonté des femmes. La femelle de l'espèce *Homo sapiens* peut s'accoupler quand elle le souhaite, sans tenir compte de périodes spécifiques au cours de l'année. De fait, les êtres humains s'accouplent des centaines, voire des milliers de fois plus souvent au cours d'une vie que nécessaire pour assurer leur reproduction. Et cela d'autant plus fréquemment que les femmes n'ont pas à craindre de tomber enceintes à chaque fois si elles ont un accès facile à la contraception. Les relations sexuelles font partie des actions gratifiantes qui activent le système de renforcement présenté plus tôt (Chapitre 11). La recherche de cette récompense dirige une partie des activités sociales humaines. La possibilité pour les femmes de maîtriser médicalement leur corps grâce à la contraception et à l'avortement et le soutien social à la petite enfance sous la forme de crèches, d'écoles maternelles, d'aide à la garde d'enfants, permettent dans les sociétés qui les implémentent de réduire efficacement la natalité. La mise en place avec succès de planning familiaux dans divers pays au cours du xx^e siècle l'a prouvé. En Italie, dans les années 1960 quand la contraception et l'avortement étaient illégaux, le taux de natalité du nord du pays était inférieur à celui du sud. Cette différence s'est largement atténuée lorsque la loi les autorisa. En Iran, la différence de natalité entre les zones rurales et urbaines a disparu lorsqu'un planning familial fut mis en place dans les années 1990. D'autres exemples de régulation de la natalité ont été mis en place avec succès en Chine, en Corée du Sud, en Thaïlande au Sri Lanka et en Colombie. Le deuxième levier de contrôle de la natalité humain apparaît donc clairement ici. Il est moins complexe et coûteux à mettre en place dans un premier temps qu'une éducation pour tous – et surtout pour toutes – particulièrement compliquée dans des pays très étendus à forte population rurale.

Quand on leur laisse le choix et la liberté de procréer, les femmes *Homo sapiens* font donc celui de faire moins d'enfants. La liberté, un des deux piliers du libéralisme, va ainsi à l'encontre de l'égoïsme de nos gènes qui devraient nous pousser à les propager le plus rapidement et le plus efficacement possible. Les êtres humains apparaissent donc obéir à un schéma doctrinal libéral sur le plan individuel, érigeant à la foi la propriété et la liberté en dogmes indépassables. En revanche, *Homo sapiens* en tant qu'espèce se comporte de façon remarquablement opposée, la liberté de procréer étant liée à une diminution de la propagation des gènes des individus à qui l'on offre cette liberté.

Le libéralisme n'est pas néodarwinien

En conclusion, s'il est réel que le libéralisme imprègne très fortement presque toutes les sociétés du XXI^e siècle, sa conséquence va à l'encontre de la théorie néodarwinienne qui sélectionnerait positivement les individus les plus aptes à se reproduire. D'une part la liberté laissée aux femmes de procréer ou non est systématiquement associée à une baisse de la natalité, d'autre part celles-là mêmes qui sont les plus aptes à diriger la société propagent moins leurs gènes que les autres. La collision de ces deux réalités laisse augurer un XXI^e siècle mouvementé au cours duquel des affrontements de peuples et de politiques expansionnistes vont essayer de résoudre dans la violence les conflits d'un monde de plus en plus dichotomique. De fait, nous en voyons déjà les prémices dévastatrices tout autour de nous. Il n'est de salut que dans un aménagement des sociétés humaines pour contraindre la démographie et freiner la consommation effrénée de nos ressources afin d'assurer la survie de notre propre espèce.

À cette fin, nous disposons de deux leviers. Premièrement, la mise en place d'un planning familial à l'échelle mondiale géré par les différentes nations paraît inévitable à très court terme pour éviter de se déchirer dans des guerres sans fin pour le contrôle de ressources de plus en plus restreintes. Deuxièmement, de façon concomitante, donner à toutes les femmes un niveau d'instruction suffisant pour les extraire d'un rôle purement reproductif et les faire participer

pleinement à l'animation de toutes les couches sociales, apparaît comme indispensable. Cela est particulièrement important dans les pays en voie de développement où la croissance démographique est beaucoup plus élevée que dans les autres. Ces deux mesures sont à prendre de toute urgence pour faire entrer *Homo sapiens* dans une ère dans laquelle l'intérêt général et la survie de l'espèce dans son ensemble deviendrait supérieur aux intérêts de petits groupes d'États-nations se livrant une guerre sans merci pour le contrôle de ressources en voie d'épuisement. C'est le grand défi – l'unique peut-être – de ce XXI^e siècle et il lointin d'être certain, hélas, que notre espèce saura le relever avec succès.

DEMAIN, L'UNIVERS SANS L'HOMME

Le modèle cosmologique actuel le plus probable est que l'univers soit en perpétuelle expansion depuis sa création initiale, pendant l'Ère Primordiale, le premier des cinq âges de l'univers tels que proposés par Gregory P. Laughlin et Fred C. Adams, deux astrophysiciens américains, dans leur livre éponyme. Nous vivons maintenant dans le deuxième âge, l'Ère Stellifère, pendant lequel les étoiles et les galaxies se forment. D'ici 6 milliards d'années, la Voie lactée fusionnera avec la galaxie la plus proche de la nôtre, Andromède, pour n'en former plus qu'une. Environ à la même époque, notre soleil aura fini de brûler ses réserves d'hydrogène et d'hélium et se transformera en naine blanche, après être devenu une géante rouge dont la couronne aura dépassé l'orbite de Vénus. La Terre ne sera plus et depuis longtemps qu'un désert calciné ne pouvant plus abriter aucune forme de vie. Les étoiles de notre Voie lactée sont aussi éloignées les unes des autres que des grains de sable espacés de plusieurs kilomètres. Néanmoins, la fusion de notre galaxie avec celle d'Andromède entraînera de dramatiques changements dans la course des étoiles. Leur rapprochement entraînera des modifications de trajectoire et de vitesse, dues aux interactions de leurs champs gravitationnels. Les planètes orbitant autour d'elles seront chassées de leur orbite pour errer, solitaires,

dans l'espace interstellaire. Ce pourrait être le destin de la Terre et des autres planètes de notre système solaire d'ici 100 000 milliards d'années.

Dans l'Ère de Dégénérescence, à partir d'un million de milliards d'années¹, les étoiles ne pourront plus se former. Celles qui mourront seront transformées en naines blanches ou en étoiles à neutrons, en fonction de leur masse. Les dernières étoiles supermassives finiront de brûler leur hydrogène. Certaines d'entre elles seront transformées en trous noirs, qui commenceront à attirer à eux les restes d'étoiles les plus proches, sous l'effet de la gravité. Les protons commenceront à se désintégrer, se transformant en positrons en émettant de l'énergie. Le temps qu'il faudra pour voir disparaître les protons est plus long que l'âge actuel de l'univers mais il n'en restera plus un seul aux alentours de 10 000 milliards de milliards de milliards de milliards d'années². Les neutrons aussi disparaîtront, se transformant en d'autres particules, proton, électron et antineutrino. À la fin de cette ère, il n'en restera plus un seul. Les composants primordiaux des noyaux atomiques – protons et neutrons – disparaissant, la matière elle-même, telle que nous la connaissons, cessera d'exister.

Puis viendra l'Ère des trous noirs, au cours de laquelle l'univers contiendra seulement ces objets hypermassifs, cadavres d'étoiles géantes. Sous l'effet du **rayonnement de Hawking***, ceux-ci s'évaporeront lentement, en émettant temporairement des photons, seules sources lumineuses dans un univers mourant. Arrivés en fin de vie, ce qu'il restera d'eux explosera, libérant en une seconde l'énergie d'un milliard de fois la bombe d'Hiroshima, essentiellement sous forme de rayons gamma,

Rayonnement de Hawking : L'évaporation des trous noirs a été proposée par l'astrophysicien Stephen Hawking (1942-2018) en 1975 et le rayonnement qui y est associé a reçu son nom en hommage à cette découverte.

1. 10^{15} années en notation scientifique, ou bien 1 suivi de 15 zéros.

2. 10^{40} en notation scientifique, ou bien 1 suivi de 40 zéros.

visibles à plusieurs années-lumière de distance. Après un temps plus grand que celui déjà écoulé³, tous les trous noirs se seront évaporés et l'univers entrera dans son dernier âge, l'Ère Sombre.

Dans cet âge, il ne restera plus que des particules de faible énergie se déplaçant dans l'espace, formant lorsqu'elles se rencontreront d'éphémères particules instables. L'énergie des derniers photons sera incroyablement faible et leur longueur d'onde tellement grande qu'ils seront invisibles. L'univers tendra vers le froid absolu. Ce qu'il se passera par la suite est entièrement spéculatif. Les lois de la physique macromoléculaire telles que nous les connaissons devraient être remplacées par celles de la physique quantique. L'hypothèse de la mort thermique qui est présentée ici est la plus vraisemblable en fonction des connaissances actuelles, mais il existe des modèles alternatifs.

Si l'énergie de l'univers est plus élevée que celle que nous lui attribuons, un effet de transition de phase pourrait se produire. Il faut comprendre ici que l'univers est constitué essentiellement de vide et l'énergie du vide est considérée comme nulle par rapport à l'énergie de l'univers. Il n'existe pas de moyen à l'heure actuelle de savoir si c'est vraiment le cas. Mais si le vide a une énergie plus élevée que supposée, il pourrait se produire un phénomène de transition de phase, le passage de notre univers de son niveau d'énergie actuel à un niveau d'énergie plus faible, comme un ballon descend une pente. L'univers que nous connaissons disparaîtrait pour laisser la place à un nouvel univers, dans lequel les lois de la physique et de la chimie seraient différentes. Cette transition se propagerait à la vitesse de la lumière, de telle sorte qu'il n'y aurait aucun moyen de l'anticiper. Il n'y aurait de toute façon aucune échappatoire et toute créature vivante se trouvant sur le passage de la transition de phase mourrait instantanément. Nous ne nous en rendrions même pas compte ! Nous sommes à peu près certains que cela ne s'est pas produit depuis la naissance de l'univers, dans la portion que nous pouvons observer.

3. 10^{100} années en notation scientifique, ou bien 1 suivi de 100 zéros.

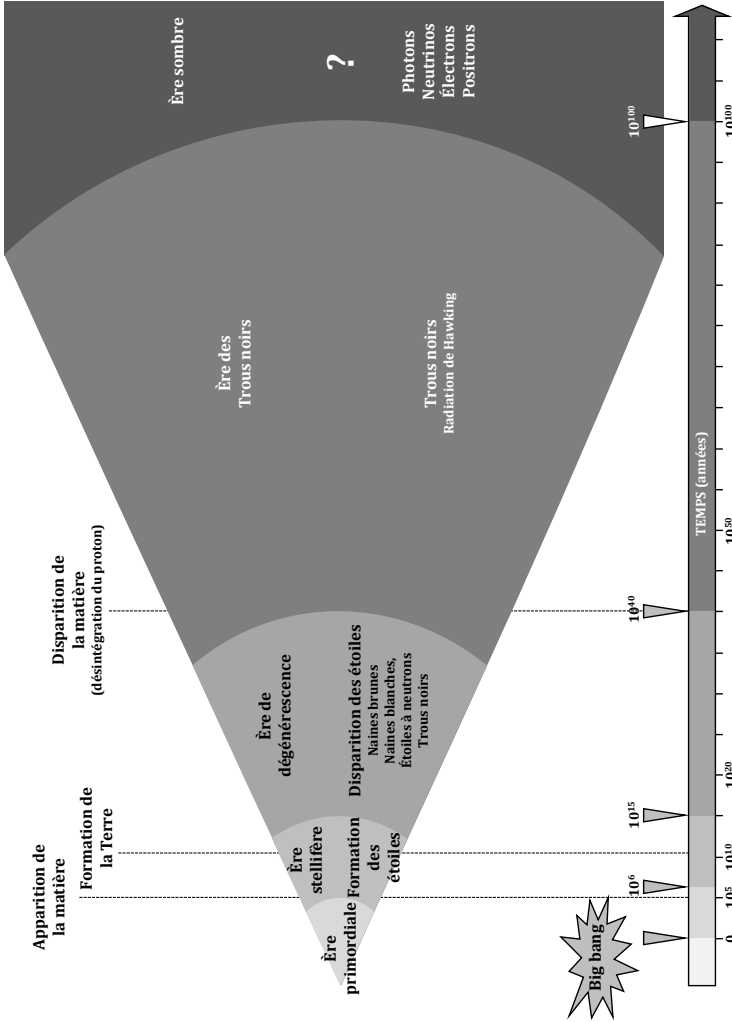


Figure 52 | Le futur de l'univers. L'échelle temporelle est en puissance de 10. Un milliard d'années = 10^9 années.

En revanche, cela pourrait se produire dans un avenir plus ou moins lointain.

Il existe encore d'autres hypothèses plus exotiques, comme celle du « Grand Déchirement », mais personne ne peut réellement dire ce qu'il adviendra du temps et de l'espace. Quelle que soit cette finalité, la vie sous toutes les formes que nous connaissons aura disparu depuis longtemps. L'évolution de l'univers est bien plus lente mais bien plus inéluctable aussi que ne l'a été celle de la vie sur Terre et si nous sommes là aujourd'hui pour écrire ces quelques lignes ou pour les lire, n'oublions jamais de rester immensément humbles face à l'incertitude immanente de ce qui nous échappe encore, mais que nous espérons comprendre un jour.

À RETENIR

- La population humaine croît à un rythme insoutenable sur le long terme pour les ressources de notre planète.
- Il existe deux leviers pour maîtriser cette croissance : la régulation des naissances par le planning familial étendu à tous les pays du monde et l'accès à l'éducation pour toutes les femmes.

BIBLIOGRAPHIE

- Dawkins, *The Selfish Gene*, Oxford University Press (1976).
- Crist, Mora et Engelman, The interaction of human population, food production and biodiversity protection, *Science* 356 : 260-264 (2017).
- Adams et Laughlin, *The Five Ages of the Universe*, Free Press Publishers (1999).

GLOSSAIRE

Acide aminé : Molécule organique simple. Leur enchaînement précis forme les protéines, macromolécules essentielles du monde vivant. On peut classer les acides aminés en fonction de leurs caractéristiques chimiques : l'acide aspartique et l'acide glutamique sont des acides aminés acides ; la lysine, l'arginine et la pyrrolysine sont basiques. Les acides aminés hydrophiles, comme la sérine, la thréonine, l'asparagine et la glutamine, sont faciles à dissoudre dans l'eau, alors que les acides aminés hydrophobes (glycine, alanine, valine, leucine, isoleucine, phénylalanine, tyrosine et tryptophane) sont difficiles, voire impossibles à dissoudre dans l'eau. La cystéine, la méthionine et la sélénocystéine contiennent un atome de soufre alors qu'histidine, proline, tryptophane, tyrosine et phénylalanine sont des molécules partiellement cycliques.

Aérobic : Un organisme aérobic est un organisme capable de respirer, c'est-à-dire d'utiliser l'oxygène comme source d'énergie. On distingue les organismes à aérobic stricte (qui ne peuvent vivre sans oxygène) de ceux à aérobic facultative (qui peuvent utiliser d'autres sources d'énergie que l'oxygène, comme la fermentation). S'oppose à anaérobic.

Allèle : Un gène peut présenter de légères différences dans sa séquence, qui peuvent plus ou moins modifier sa fonction. On appelle « allèles » les différentes formes d'un même gène. Un individu porte souvent deux allèles différents du même gène, un sur chaque chromosome homologue.

Alu : Élément transposable dont le nom vient du fait que les transposons de cette famille contiennent tous une séquence particulière d'ADN, le site de reconnaissance pour l'enzyme *Alu I*, qui est AGCT. Les éléments *Alu* sont classés en différentes familles en fonction de leur séquence.

Anaérobie : Un organisme anaérobie est un organisme incapable d'utiliser l'oxygène atmosphérique comme source d'énergie. S'oppose à aérobie.

Apoptose : Mort cellulaire programmée. C'est l'une des voies possibles de la mort cellulaire chez les animaux.

Archéobactérie (ou archée) : Organisme unicellulaire procaryote. Un des trois embranchements du vivant. Les deux autres sont les eubactéries et les eucaryotes.

Archézoaire : Eucaryote qui ne possède pas de mitochondrie ou qui possède des mitochondries très dégénérées. Les différentes lignées d'archézoaires sont toutes très anciennes et remontent à l'origine des cellules eucaryotes.

ARNt synthétase : Protéine dont la fonction est de charger un acide aminé particulier sur son ARNt accepteur. Il existe une ARNt synthétase pour chaque acide aminé, mais il en existe en fait 21 car la lysine peut être chargée par une synthétase de classe I ou de classe II, en fonction des espèces.

Arthropodes : Invertébrés qui comprennent – entre autres – les arachnides, les insectes et les crustacés.

Ascomycètes : Champignons qui forment des spores, les cellules germinales des champignons, protégées par une asque qui donne son nom au clade.

ATP : Adénosine triphosphate. L'ATP est le carburant utilisé par l'ensemble du monde vivant aérobie. Il est constitué d'une adénosine,

une des quatre bases de l'ARN, à laquelle sont liés trois groupements phosphates. Un phosphate est formé d'un atome de phosphore (P) lié à quatre atomes d'oxygène (O). Chaque liaison phosphodiester entre deux nucléotides d'une chaîne d'ADN ou d'ARN implique un groupement phosphate.

Biofilm : Communauté de microbes (bactéries, champignons et algues unicellulaires) vivant parfois en symbiose, adhérant entre eux et à une surface au sein d'un milieu aqueux et produisant parfois une matrice biologique au rôle protecteur. C'est l'un des deux modes de vie des organismes unicellulaires, l'autre étant le mode « planctonique », défini par la flottaison libre au sein d'un milieu aqueux.

Carbonyle : Une molécule carbonylée contient un groupement carbonyle, c'est-à-dire un atome de carbone lié à un atome d'oxygène par une double liaison covalente. De nombreuses molécules organiques sont carbonylées.

Catalyseur : Un catalyseur est une molécule qui augmente la vitesse d'une réaction sans être consommée par elle, ce qui lui permet d'être réutilisé plusieurs fois. En biologie, lorsqu'une protéine joue le rôle de catalyseur on l'appelle enzyme, lorsqu'il s'agit d'un ARN on le nomme ribozyme.

Centromère : Région chromosomique sur laquelle se fixent les microtubules et qui permet la bonne séparation des chromosomes lors de la mitose dans les cellules eucaryotes.

Chloroplaste : Organite cellulaire présent dans les cellules eucaryotes végétales et responsable de la photosynthèse.

Chondrites : Météorites non métalliques, divisées en plusieurs sous-groupes dont les chondrites carbonées qui contiennent des proportions variables de carbone et d'oxygène. Les chondrites représentent environ 85 % des météorites qui s'écrasent sur Terre.

Chromothripsie : Mécanisme par lequel des chromosomes sont fragmentés en dizaines, voire en centaines de fragments et réassemblés. Ce phénomène est très courant dans certaines tumeurs et on en

ignore l'origine et le mécanisme exact. De la racine grecque *chromo* que l'on retrouve dans chromosome et *thripsis* qui signifie « couper en morceaux ».

Clade : voir Groupe monophylétique.

CNV (*Copy Number Variation*) : Morceau d'ADN dont la taille peut varier de cinquante à plusieurs dizaines de milliers de nucléotides, répété plusieurs fois dans le génome humain, de façon plus ou moins espacée le long des chromosomes.

Code génétique : Ensemble de règles permettant de traduire la séquence codée dans un ARN messager par le ribosome pour fabriquer une protéine.

Commensal : Un organisme commensal est l'hôte d'un autre organisme avec lequel il cohabite sans lui causer de dommages. L'hôte nourrit partiellement l'organisme, mais à la différence d'un parasite, la présence d'un commensal n'est pas néfaste à l'hôte.

Cyanobactérie : Bactérie marine vivant sous forme coloniale et formant des filaments gluants. Certaines cyanobactéries sont capables d'effectuer la photosynthèse.

Cytoplasme : Dans les cellules eucaryotes, c'est le milieu intracellulaire qui contient le noyau, les mitochondries et les chloroplastes. On y trouve les enzymes du métabolisme ainsi que les ribosomes responsables de la traduction des ARN messagers en protéines.

Cytosquelette : Le cytosquelette est l'ensemble des fibres protéiques polymérisées qui organisent l'architecture de toutes les cellules vivantes. Il existe des différences importantes entre le cytosquelette eucaryote et procaryote.

Datation radiométrique : Un isotope du potassium (^{40}K) se désintègre avec le temps en argon (^{40}Ar). Si une roche contient suffisamment de potassium au moment de sa formation (une roche volcanique par exemple), le rapport potassium/argon permet de dater très précisément l'âge de cette roche. La période du ^{40}K étant de 1,25 milliard d'années, cette technique peut être utilisée pour dater avec précision l'âge de roches couvrant la quasi-totalité des

ères géologiques. En comparaison, la période du ^{14}C est beaucoup plus courte (5 730 ans) et ne permet donc de dater que des échantillons relativement récents (à l'échelle géologique).

Délétion : Mutation qui enlève des nucléotides à une séquence d'ADN. Les délétions peuvent être de faible taille (de deux à quelques nucléotides) ou bien représenter plusieurs milliers de paires de bases.

Dérive génétique : Évolution des différents allèles des gènes d'une population en fonction d'événements stochastiques indépendants de l'apparition de nouvelles mutations ou de sélection naturelle. Une épidémie, une catastrophe naturelle, ou encore des taux de reproduction différents entre les individus sont des moteurs de la dérive génétique.

Diploïde : Cellule possédant deux exemplaires de chaque chromosome. La plupart des animaux et des plantes sont diploïdes.

Duplication segmentaire : Duplication d'un segment de chromosome de quelques milliers à quelques millions de nucléotides, mécanisme d'évolution des génomes fréquent chez les eucaryotes.

Endonucléase : Protéine (parfois associée à un ARN) capable de cliver la liaison phosphodiester entre deux nucléotides. Il existe des endonucléases qui clivent l'ARN et d'autres spécifiques de l'ADN.

Eubactérie : Organisme unicellulaire procaryote. La plupart des bactéries commensales et pathogènes sont des eubactéries. C'est un des trois embranchements du vivant, les deux autres étant les archées et les eucaryotes.

Eucaryote : Du grec *eu* « bien » et *karyon* « noyau ». Eucaryote signifie donc littéralement « qui a un bon, un vrai noyau ». Les eucaryotes sont un des trois embranchements du vivant, avec les archées et les eubactéries.

Eudicotylédone : Plante dont l'embryon possède deux cotylédons.

Exons : Voir Introns.

Fixation : Une mutation (un allèle) dans un gène va vers la fixation lorsqu'elle se propage dans une population au cours des

générations au détriment d'une autre mutation (un autre allèle) dans le même gène qui, elle, va tendre à disparaître. Une mutation fixée finit par envahir complètement la population, de façon à ce que la fréquence de l'allèle portant cette mutation soit 100 %.

Gamète : Cellule haploïde sexuée mature, qui peut fusionner avec un gamète de type sexuel opposé pour former un individu mature. Chez les mammifères, les gamètes mâles sont les spermatozoïdes et les gamètes femelles sont les ovules.

Gène : Fragment d'ADN dont l'ARN transcrit a une fonction biologique.

Génétique : Discipline de la biologie qui étudie les gènes et les règles de leur transmission héréditaire.

Génomique : Domaine de la génétique qui étudie la composition, la structure et l'évolution des génomes. La génomique comparative regroupe l'ensemble des approches bioinformatiques qui permettent la comparaison des génomes entre eux, en vue de déterminer les liens de parenté et les vitesses d'évolution entre les espèces comparées.

Génotype : Ensemble de tous les allèles des gènes d'un être vivant.

Genre : Rang de classification taxonomique qui vient juste au-dessus de celui de l'espèce. Pour l'homme, le nom de genre est *Homo* et le nom d'espèce *sapiens*.

Goulet d'étranglement : Réduction drastique du nombre d'allèles présents dans une population suite à l'élimination d'une grande partie de celle-ci. Appelé *population bottleneck* en anglais. Le terme anglais est souvent utilisé, même dans des textes français. En français, on utilise aussi l'appellation de goulot d'étranglement, qui rappelle le goulot de la bouteille (*bottleneck*).

Grand Rift : La vallée du Grand Rift est un ensemble géologique de failles de 6 000 kilomètres de long du nord au sud, s'étirant entre le nord du Liban et l'Afrique australe, en passant par la mer Rouge et les Grands Lacs africains. Il a longtemps été pensé par certains paléontologues que les premiers *Homo sapiens* avaient dû voir le jour à l'est du rift africain, dans une région de savanes propice

à l'acquisition de la station debout, alors que les grands singes vivaient à l'ouest du rift dans une région de forêts denses, plus propices à leur anatomie et à leur mode de vie.

Groupe monophylétique (ou clade) : Groupe qui contient l'ancêtre commun et tous ses descendants.

Haploïde : Cellule ne possédant qu'un seul exemplaire de chaque chromosome. Les bactéries ainsi que les cellules germinales des animaux (spermatozoïdes et ovules) sont des exemples de cellules haploïdes.

Hétérozygote : Deux gènes sont à l'état hétérozygote si les deux allèles du gène sont différents, c'est-à-dire si les deux séquences nucléotidiques ne sont pas strictement identiques. C'est la situation la plus courante chez les individus sexués.

Homéostasie : Chez un être vivant, processus de régulation par lequel l'organisme maintient les différentes constantes de son milieu intérieur (température, pression osmotique, oxygénation, glycémie, acidité, etc.) dans les limites nécessaires à sa survie.

Hominidé : Les hominidés ou grands singes, comprennent les bonobos, les chimpanzés, les gorilles, les orangs-outans, les humains, ainsi que plusieurs lignées disparues dont certaines étaient des ancêtres de l'Homme.

Homologie : Deux gènes (ou protéines) sont homologues quand leurs séquences partagent un certain degré d'identité. L'homologie est une notion fondamentale dans la spéciation et dans la construction d'arbres phylogénétiques. Deux gènes homologues appartenant à deux organismes différents sont dits *orthologues* s'ils proviennent d'un gène unique ayant appartenu à l'ancêtre commun des deux organismes avant leur spéciation. Deux gènes homologues appartenant au même organisme sont dits *paralogues* s'ils proviennent d'un événement de duplication de gène propre à cet organisme.

Homozygote : Deux gènes sont à l'état homozygote si les séquences nucléotidiques des deux allèles du gène portés par les deux chromosomes homologues sont strictement identiques.

Hybridation inter-spécifique : Accouplement entre espèces différentes dont la descendance est fertile.

Hydrolyse : Réaction chimique catalysée par une molécule d'eau, qui détruit une liaison entre deux atomes.

Hydrophile : Composé ayant une affinité pour l'eau et tendance à s'y dissoudre. Le sucre alimentaire est hydrophile.

Hydrophobe : Composé n'ayant pas d'affinité pour l'eau et ne pouvant s'y dissoudre. Les huiles et les graisses sont hydrophobes.

Introgession : Transfert interspécifique d'ADN par hybridation, entre deux espèces distinctes mais suffisamment proches pour donner une descendance fertile. Les introgressions peuvent être de très grande taille, contrairement aux transferts horizontaux qui impliquent généralement un seul ou un très petit nombre de gènes. Le terme était originellement utilisé en génétique des plantes, jusqu'à ce que l'on découvre de nombreux cas d'introgressions en dehors du règne végétal.

Introns : Séquences présentes dans l'ARN lorsqu'il est transcrit à partir de l'ADN, puis excisées afin de former l'ARN messager mature qui sera traduit en protéine. Un exon est une séquence d'ADN se retrouvant dans l'ARN messager mature. L'excision des introns et le raboutage des exons est un processus complexe et hautement régulé qui nécessite une machinerie conservée dans tout le monde eucaryote. Les gènes des eucaryotes unicellulaires tels que levures et champignons contiennent relativement peu d'introns, alors que ceux des métazoaires en contiennent un grand nombre (9-10 introns par gène humain en moyenne). Les organismes procaryotes ne possèdent pas d'introns similaires, ce qui explique en partie la plus petite taille de leur génome.

Liaison peptidique : Liaison chimique stable entre deux acides aminés. Les peptides sont des chaînes d'acides aminés reliés entre eux par des liaisons peptidiques.

Méiose : Mécanisme de division cellulaire des cellules germinales qui va donner quatre gamètes haploïdes à partir d'une cellule diploïde.

Mésophile : Un organisme mésophile peut croître uniquement dans des conditions de température modérée, de l'ordre de 20-40 °C.

Métagénome : Le métagénome humain est l'ensemble des gènes humains, bactériens, fongiques et viraux vivant dans l'écosystème représenté par un être humain. La métagénomique est la science qui étudie les génomes d'organismes présents dans un échantillon environnemental donné.

Métazoaires : Animaux multicellulaires, à l'opposé des protozoaires (ou protistes), unicellulaires.

Microbiote : Ensemble des microorganismes vivant en symbiose avec l'homme. L'ensemble des séquences de leurs génomes se nomme le microbiome.

Mitochondrie : Organite cellulaire présent dans la plupart des cellules eucaryotes, dont le rôle est d'assurer la fabrication d'énergie sous forme d'ATP, par utilisation de l'oxygène atmosphérique lors de la respiration cellulaire.

Molécule : Une molécule est un assemblage de plusieurs éléments chimiques (ou atomes) dans une configuration précise. Lorsque l'assemblage est de grande taille, on parle alors de « macromolécule ».

Monocotylédone : Plante dont l'embryon possède un seul cotylédon ou feuille primordiale.

Mutation : Modification permanente, accidentelle ou provoquée du génome d'un organisme vivant. Chez les microorganismes, toutes les mutations se transmettent à la descendance. Chez les organismes sexués, une mutation ne pourra se transmettre à la descendance que si elle touche la lignée germinale (ovules ou spermatozoïdes).

Nématodes : vers ronds, dont le plus connu est *Caenorhabditis elegans*, un organisme modèle en génétique.

Néochromosome : Chromosome créé par réarrangement génétique dans un génome donné, pouvant impliquer une duplication totale ou partielle de tout ou partie d'un ou plusieurs chromosomes préexistants.

Néofonctionnalisation : Création d'une nouvelle fonction par accumulation de mutations dans un gène dupliqué, par sélection darwinienne positive.

Neurotransmetteur : Composé chimique libéré par les neurones, ayant une action sur d'autres neurones.

Nucléotide : On appelle nucléotide une base liée à un sucre et un, deux ou trois groupements phosphate. Les quatre nucléotides de l'ARN (ou ribonucléotides) sont l'adénosine, la cytidine, la guanosine et l'uridine, dont les noms dérivent de la base correspondante. L'ADN contient des désoxyribonucléotides, une base liée à un désoxyribose et un phosphate : désoxyadénosine, désoxycytidine, désoxyguanosine et désoxythymidine (ou simplement thymidine car il n'existe pas de thymidine dans l'ARN). Par souci de simplification, on nomme souvent les nucléotides de l'ARN/ADN par le nom de leur base : adénine, cytosine, guanine et uracile/thymine.

Oligonucléotide : Un oligonucléotide est un court morceau d'ADN ou d'ARN (en général 20-100 nucléotides) synthétisé chimiquement. De nombreuses compagnies de biotechnologie fournissent de nos jours des oligonucléotides à un prix dérisoire (environ 10 centimes d'euros par nucléotide, soit 200 euros pour un gène de levure de taille moyenne).

Orthologues : Deux gènes homologues appartenant à deux organismes différents sont dits orthologues s'ils proviennent d'un gène unique ayant appartenu à l'ancêtre commun des deux organismes avant leur spéciation. Voir Homologie.

Oxydo-réduction : Un réducteur est une molécule qui donne facilement ses électrons. Un oxydant est une molécule qui accepte facilement des électrons. Un échange d'électrons entre un réducteur et un oxydant est appelé réaction d'oxydo-réduction. Ces réactions sont essentielles pour le métabolisme énergétique de tous les êtres vivants. Le sulfure d'hydrogène (H_2S) est un réducteur alors que l'oxygène atmosphérique (O_2) est un oxydant.

Paléoécologie : Science qui étudie les relations des êtres vivants ancestraux avec leur environnement.

Paralogues : Deux gènes appartenant au même organisme sont dits paralogues s'ils proviennent d'un événement de duplication d'un gène dans cet organisme. Voir homologie.

Parthénogenèse : Reproduction d'un individu adulte à partir d'un gamète femelle non fécondé.

pH : Le potentiel hydrogène ou pH est une mesure de la quantité d'ions hydrogène (H^+ ou protons) dans une solution. Le pH d'une solution va définir son acidité : une solution de $pH < 7$ est acide, une solution de $pH > 7$ est basique (ou alcaline). À $pH = 7$, une solution est neutre.

Phage (ou bactériophage) : Virus qui infecte les bactéries, en injectant son matériel génétique (ADN ou ARN) afin de le faire reproduire par la cellule bactérienne hôte.

Phagocytose : Propriété de certaines cellules de capturer et d'ingérer des particules ou des micro-organismes afin de les détruire.

Phénotype : Ensemble des caractères observables d'un être vivant. Le phénotype résulte de l'expression du génotype.

Phéromone : Substance chimique fabriquée et excrétée par la plupart des animaux, servant à transmettre des informations entre les individus et jouant un rôle particulièrement important dans l'attraction sexuelle.

Phloème : Les vaisseaux du phloème contiennent la sève élaborée, constituée d'eau et des sucres fabriqués au cours de la photosynthèse dans la partie supérieure de la plante.

Photosynthèse : Processus métabolique permettant aux algues, aux plantes et à certaines bactéries (cyanobactéries) de synthétiser de la matière organique grâce à la lumière du soleil.

Phylogénie : Étude des liens de parenté entre êtres vivants, entre populations et entre espèces. La phylogénie moléculaire se base exclusivement sur des critères d'évolution de séquences d'ADN au cours du temps pour définir les liens de parenté. Du grec *phylon* « tribu, clan » et *genetikós* « origine ou source ».

Phylum : Le phylum (ou embranchement) est le rang de classification des espèces vivantes juste en dessous du *règne*. Le dernier consensus de classification du monde vivant considère qu'il existe deux super-règnes, les procaryotes et les eucaryotes, ces derniers étant divisés en cinq règnes : protozoaires, chromistes, champignons, plantes et animaux. Au sein du règne des animaux, il existe 34 phylums différents, ce nombre pouvant légèrement varier en fonction des auteurs. Le super-règne des procaryotes est subdivisé en deux règnes, archéobactéries et eubactéries, qui comprennent ensemble 32 phylums (trois phylums d'archées et 29 phylums d'eubactéries).

Population effective : La population effective reflète une population idéale dans laquelle les échanges génétiques se produisent dans des conditions peu réalistes mais avec la même dynamique que dans la population réelle. C'est une simplification de la réalité qui sous-estime généralement la taille de la population réelle considérée.

Prébiotique : La chimie prébiotique étudie la formation des molécules organiques qui précéderent l'apparition des premières cellules. À la chimie prébiotique succède la biochimie, qui étudie les réactions chimiques se déroulant au sein des organismes vivants.

Procaryote : Du latin *pro* « avant » et du grec *karyon* « noyau ».

Protistes : Eucaryotes unicellulaires qui ne sont ni des algues, ni des levures, ni des champignons. La classification de Chatton en 1925 était plus large et comprenait certaines bactéries.

Protozoaire : Eucaryote unicellulaire. Il existe des formes libres comme la paramécie ou l'amibe, ainsi que des formes parasitaires comme le trypanosome, responsable de la maladie du sommeil.

Pseudogène : Gène inactif suite à l'accumulation de mutations. Un pseudogène dérive souvent d'un gène dupliqué dont l'une des deux copies a par la suite perdu sa fonction. Cette duplication peut être due à une duplication totale du génome, à une duplication segmentaire ou à un mécanisme de rétrotransposition (voir Transposon). Les pseudogènes sont très fréquents dans les génomes de mammifères.

Purines (pyrimidines) : Molécules organiques simples qui constituent les bases de l'ADN et de l'ARN. Elles sont également directement impliquées dans nombre de réactions chimiques essentielles à la propagation de la vie. Adénine (A) et guanine (G) sont des purines. Uracile (U), thymine (T) et cytosine (C) sont des pyrimidines. Les guanines et les cytosines forment entre elles des liaisons plus fortes que les adénines et les uraciles. Les molécules d'ARN ou d'ADN qui contiennent plus de bases G ou C que de bases A, U ou T seront donc plus stables.

Pyrimidines : voir Purines.

Rayonnement de Hawking : L'évaporation des trous noirs a été proposée par l'astrophysicien Stephen Hawking (1942-) en 1975 et le rayonnement qui y est associé a reçu son nom en hommage à cette découverte.

Recombinaison homologue : Échange de nucléotides entre deux segments d'ADN identiques ou très similaires. Elle peut se produire en réponse à une cassure accidentelle de l'ADN (rayons X) ou bien à la suite de cassures programmées comme lors de la méiose.

Réplication : Mécanisme complexe et hautement régulé permettant de dupliquer une molécule d'ADN pour donner deux molécules identiques à la première. C'est de cette façon que se transmet l'information génétique d'une cellule mère à une cellule fille.

Rétrotransposon : Séquence particulière d'ADN capable de se dupliquer en se déplaçant de façon autonome dans un génome. Voir Transposon.

Ribosome : L'ARN ribosomique est un constituant du ribosome, une énorme machine macromoléculaire responsable de toute la synthèse des protéines dans une cellule. La structure du ribosome repose sur quatre molécules d'ARN, que l'on nomme en fonction de leur coefficient de sédimentation : 5S, 5,8S, 18S et 28S.

Sahul : Appellation du plateau continental de terres émergées qui correspond à l'heure actuelle à l'Australie, la Tasmanie et la Nouvelle-Guinée. Avant la fin du dernier maximum glaciaire

(– 16 000 ans), le niveau des océans était inférieur de 125 mètres à son niveau actuel et ces territoires n'en faisaient qu'un.

Saturation : Une position dans deux gènes orthologues est dite saturée quand le nucléotide correspondant a subi plus d'une mutation depuis l'ancêtre commun de ces deux gènes et qu'il n'est donc pas possible de déterminer l'âge de leur divergence en utilisant une horloge moléculaire.

Sauropsidés : Groupe d'animaux descendant d'un ancêtre commun et regroupant les oiseaux, les reptiles et les dinosaures.

Sédiment : Ensemble de particules amenées par le vent ou l'eau et se déposant sous l'effet de la gravité. Leur accumulation au cours des ères géologiques forme des couches sédimentaires rocheuses que l'on peut précisément dater en fonction de leur contenu en carbone 14, un des isotopes radioactifs du carbone. La période du carbone 14 est de 5 730 ans, ce qui signifie que sa quantité diminue de moitié tous les 5 730 ans. En mesurant précisément la quantité de carbone 14 contenu dans une roche sédimentaire, on peut déterminer son âge et donc l'âge des fossiles qu'elle contient.

Séquence : On appelle séquence d'un ARN ou d'un ADN l'ordre précis de l'enchaînement des quatre bases (des quatre nucléotides) le long de la molécule. On appelle séquence d'une protéine l'ordre précis de l'enchaînement des acides aminés le long de la molécule.

SETI : Search for Extra-Terrestrial Intelligence. Une initiative américaine des années 1960 dont le but est d'essayer de détecter des signaux intentionnels ou non, émis par une civilisation extraterrestre.

Silicates : Importante famille de minéraux à base de silice (SiO_2) qui constitue plus de 90 % de la croûte terrestre et que l'on trouve agrégés en grains de poussière autour de certaines étoiles, dont notre Soleil.

Solution aqueuse : Solution dans laquelle les molécules d'eau sont très largement majoritaires par rapport aux autres molécules. L'eau est le principal composant des organismes vivants.

Spéciation : Processus évolutif menant à l'apparition et à la sélection de nouvelles espèces vivantes.

Stromatolithe : Structure en feuillets formée par des bactéries qui se sont déposées dans un environnement aquatique de faible profondeur. Ces multiples couches bactériennes sont parfois interrompues par des dépôts de sédiments. La présence de stromatolithes est révélatrice de la présence de communautés microbiennes formant des biofilms.

Subfonctionnalisation : Modification d'un gène dupliqué par accumulation de mutations sélectionnées positivement, pour adapter sa fonction à un sous-compartiment cellulaire différent (mitochondrie, noyau, etc.) ou à un moment différent de la vie de la cellule ou de l'organisme.

Symbiose : Association constante, obligatoire et spécifique entre deux organismes vivants ne pouvant vivre l'un sans l'autre.

Symbiotes (ou symbiontes) : Organismes symbiotiques vivant soit en association mutualiste (les deux organismes retirent un intérêt à l'association), soit en association parasitaire (l'un des deux organismes en retire un avantage tandis que l'autre paye un coût).

Syngamie : Fusion de deux cellules *haploïdes* de type sexuel opposé pour former une cellule *diploïde*, contenant deux exemplaires de chaque chromosome, l'un paternel et l'autre maternel.

Synténie : Ordre local des gènes le long d'un chromosome ou d'un segment chromosomique.

Système endocrinien : Ensemble des organes impliqués dans la synthèse et la sécrétion des hormones, qui sont les messagers de l'organisme.

Taxon : Ensemble des êtres vivants partageant les mêmes caractéristiques taxonomiques. Celles-ci peuvent reposer sur la morphologie, la physiologie, l'embryologie, la génétique ou les comportements culturels ou sociaux.

Taxonomie : Domaine de la biologie qui s'intéresse à la classification des êtres vivants, selon des critères morphologiques, physiologiques ou embryologiques.

Télomère : Chaque chromosome linéaire contient deux télomères, situés aux extrémités des chromosomes et constitués d'ADN répété et de protéines. Les télomères servent à protéger les extrémités chromosomiques d'une dégradation et sont donc essentiels à l'intégrité des chromosomes. Leur raccourcissement est lié au vieillissement cellulaire.

Thylakoïde : Sorte de sac aplati contenant les enzymes nécessaires à la photosynthèse, que l'on trouve dans les cyanobactéries et les chloroplastes.

Traduction : Mécanisme moléculaire transformant la séquence portée par un ARN messager en protéine, par l'intermédiaire du code génétique.

Transcription : Mécanisme moléculaire transformant une séquence ADN en ARN.

Transfert horizontal : Transfert d'un ou plusieurs gènes d'un organisme vivant à un autre sans relation de descendance entre les deux organismes. Ce type de transfert génétique nécessite la présence des deux organismes au sein de la même niche écologique, voire une relation symbiotique entre eux. Cette transmission est différente du transfert vertical de gènes, tels que ceux transmis à une descendance. Les transferts horizontaux sont extrêmement fréquents dans le monde bactérien.

Transgène : Gène étranger introduit dans le génome d'un organisme receveur plus ou moins génétiquement éloigné du génome de l'organisme donneur, par des techniques de biologie moléculaire.

Transposon : Appelé aussi élément transposable. Séquence d'ADN capable de se déplacer de façon autonome dans un génome. La découverte en a été faite dans les années 1950 en étudiant la couleur des grains de maïs, par Barbara McClintock, qui reçut le prix Nobel de Physiologie ou Médecine en 1983 pour cette découverte. Il existe deux grandes familles de transposons, les transposons à ADN qui se déplacent par un mécanisme de « couper-coller » et les rétrotransposons qui sont d'abord transcrits en ARN avant leur

transposition par un mécanisme de « copier-coller ». Les rétro-transposons peuvent aussi dupliquer d'autres gènes qu'eux-mêmes et sont une source importante de création de pseudogènes.

Virus : Un virus est un organisme simple qui contient une information génétique, sous forme d'ADN ou d'ARN, mais qui est incapable de la reproduire lui-même. Il est obligé de pénétrer une bactérie, une cellule animale ou végétale afin d'utiliser la machinerie de reproduction de son hôte involontaire pour se reproduire. C'est donc un parasite intracellulaire obligatoire, existant dans tous les domaines du vivant, des bactéries aux plantes et aux animaux.

Xylème : Le xylème transporte la sève brute (eau et molécules organiques) en provenance des racines vers les feuilles, les fleurs et les fruits. Les vaisseaux du xylème sont formés de trachéides, successions de cellules mortes, qui donnent leur nom aux trachéophytes.

Zone habitable : La zone habitable est définie comme la région de l'espace où les conditions sont favorables à l'apparition de la vie telle que nous la connaissons, en pratique ni trop près ni trop loin du soleil.