U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: Harb OS, Boehme U, Crouch K, et al. Genomes. In: Walochnik J, Duchêne M, editors. Molecular Parasitology: Protozoan Parasites and their Molecules [Select Chapters]. Cham (CH): Springer; 2016 Oct.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/

Genomes

Omar S. Harb,[✉1] Ulrike Boehme,[2] Kathryn Crouch,[3] Olukemi O. Ifeonu,[4] David S. Roos,[1] Joana C Silva,[4] Fatima Silva-Franco,[5] Staffan Svärd,[6] Kyle Tretina,[4] and Gareth Weedall[7]

Abstract

In the last decade, the rise of affordable high-throughput sequencing technologies has led to rapid advances across the biological sciences. At the time of writing, annotated reference genomes are available within most clades of eukaryotic pathogens, and including un-annotated sequences over 550 genomes are available in total. This has greatly facilitated studies in many areas of parasitology. In addition, the volume of functional genomics data, including analysis of differential transcription and DNA-protein interactions, has increased exponentially. With this unprecedented increase in publicly available data, tools to search and compare datasets are also becoming ever more important. A number of database resources are available, and access to these has become fundamental for a majority of research groups. This chapter discusses the current state of genomics research for

Author Affiliations: 1 Department of Biology, University of Pennsylvania, Philadelphia USA Email: oharb@upenn.edu Tel: +1 215-746-7019. 2 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA Email: ucb@sanger.ac.uk. 3 Wellcome Trust Centre for Molecular Parasitology, B6-28 SGDB, 120 University Place, Glasgow, G12 8TA Email: kathryn.crouch@glasgow.ac.uk Tel: +44 141 330 3746. 4 University of Maryland School of Medicine, Insti-tute for Genome Sciences, BioPark II, Room 645, 801 W. Baltimore St., Baltimore, MD 21201 USA JS Email: jcsilva@som.umaryland.edu OI Email: KAbolude@som.umaryland.edu KT Email: KTretina@som.umaryland.edu Tel: +1 410-706-6721. 5 Institute of Integrative Biology, University of Liv-erpool, Liverpool, L69 7ZB, UK Email: F.Silva-Franco@liverpool.ac.uk. 6 Department of Cell and Molecular Biology, Uppsala University, BMC, Box 596, SE-75123, Uppsala, Sweden Email: staffan.svard@icm.uu.se Tel: +46 184714558. 7 Vector Biology Department, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK. Email: gareth.weedall@lstmed.ac.uk.

✉ Corresponding author.

This is an Accepted Manuscript of a book chapter published by Springer-Verlag Wien in Molecular Parasitology: Protozoan Parasites and their Molecules, available online: https://www.springer.com/gp/book/9783709114155.

a number of eukaryotic parasites, discussing the genome and functional genomics resources available at the time of writing and highlighting functionally important or unique aspects of the genome for each group. In addition publicly accessible database resources pertaining to eukaryotic parasites are also discussed.

Introduction

Arguably the field of genomics began when Friedrich Miescher first isolated DNA in 1869 [1], paving the way for the work of many scientists in understanding the role of this material in heredity [2], discovering its double-helical structure [3] and deciphering the genetic code [4]. However, the technological advance that the entire field of genomics rests on is sequencing [5–7]. The ability to read the genetic code is relatively new, having only been developed in the last 50 years. Sanger sequencing, which relies on dideoxy chain termination, remained the method of choice for several decades; however, early implementations of dideoxy chain termination methods were not well parallelized and analysis was initially a painstaking manual process. Later, data analysis was carried out computationally, but limited by the processing capacity of computers of the era. These factors combined to limit early sequencing to individual genes, small genomic fragments or the genomes of small viruses and organelles. The emergence of techniques such as fluorescence-based cycle sequencing and the polymerase chain reaction in addition to the increased use of computational power to automatically read and analyze results, allowed larger scale genome projects to be undertaken [8]. Indeed within a few years of this marriage of techniques and fields the first bacterial, protozoan, fungal, plant and animal genomes were sequenced [9–12] [13] [14]. Despite these advances, sequencing of whole genomes remained relatively costly and time consuming. As an example, sequencing the human genome took roughly 10 years at a price tag of 3 billion US dollars (https://www.genome.gov/11006943) [15].

The first forays into high-throughput analysis of sequence data came in the form of microarrays. A microarray consists of a panel of oligo-nucleotide probes bonded to a solid surface such as a glass slide. Hybridisation of nucleic acids from a specimen to individual probes is detected by the intensity of a fluorescent signal. This technique was the first to make querying of sequence polymorphisms, transcript expression levels and segmental duplications possible on a genomic level, and cheap enough to be widely available. In addition, microarrays forced the development of computational tools and techniques to handle data on a genomic scale. However, an important limitation of microarrays is the requirement for prior knowledge of the genome and the coincident inability to make *de novo* discoveries (i.e., one can query the presence of known SNPs, but not discover new SNPs). A large volume of functional genomics data has been obtained using microarray technologies, but with a small number of exceptions (such as diagnostics), microarrays have for the most part been superseded by next-generation sequencing technologies.

Two factors have been instrumental in enabling sequencing to be taken to the next level: continued growth of computer processing capacity following Moore's law [16] and the development of "next-generation" sequencing (NGS) methods (also known as second generation sequencing), which enable massively parallel sequencing of millions of fragments by synthesis [17, 18]. One of the major advantages of next generation sequencing is that it can be applied to a wide variety of methodologies including (*readers are directed to an excellent series of manuscripts http://www.nature.com/nrg/series/nextgeneration/index.html*) and unlike microarrays, does not require any prior knowledge of the sample:

- DNA sequencing: High-throughput technology makes sequencing for *de novo* assembly of new genomes ever more affordable. Comparison of resequenced isolates against a reference is a common technique for discovery of sequence polymorphisms, while analysis of coverage depth and mapping topology can reveal information about structural variations such as chromosomal translocations and segmental duplications.
- RNA sequencing: Sequencing of RNA can provide important information about gene structure such as the locations of UTRs and intron/exon boundaries, and the presence of alternative or *trans*- splice variants. Analysis of RNAseq coverage depth over a time course or under different experimental conditions reveals

information about transcription of genes under differing conditions and combination of this technique with ribosomal profiling enables identification of the translational status of the genome. Specialised sample preparation techniques enable the sequencing of non-coding RNA species such as those involved in the RNAi-mediated translational-silencing.

- Epigenomics: Chromatin immunoprecipitation (ChIP)-sequencing is a powerful technique that allows determination of the "footprint" of DNA-binding proteins. This can be used to examine promotor-binding sites, transcription, replication and repair mechanisms and factors such as histone-modification that can affect transcription. Other techniques are available, such as bisulfite sequencing which enables profiling of DNA methylation.
- Metagenomics: Sequencing of DNA extracted from samples that contain mixed populations of organisms can be used to survey populations in environmental samples (such as soil) or biological samples (such as gut microbiomes). Metagenomics techniques can be used to determine the makeup of populations and to survey how this changes over time or under different conditions. Metagenomics analysis is a fast-growing field in which the problems of analysis have not yet been solved.

It is not surprising that the dawn of large scale sequencing projects necessitated an expansion in the field of bioinformatics and data management. As high-throughput sequencing has become cheaper it has moved from being a specialist technique to a tool used daily in labs across the world. This has necessitated the development of user-friendly tools that can run on desktop machines, and thrust the field of bioinformatics into the foreground. The expansion of massively parallel sequencing has also led to a revolution in the teaching of biology, with computational techniques for management and analysis of genomic-scale datasets now being taught in many undergraduate courses. Data warehousing is also becoming a priority, with data repositories such as the National Center for Biotechnology Information (NCBI) [19] having to rethink both their submissions procedures and their approaches to storage.

Parasite genomics

The field of parasite genomics has benefited tremendously from the sequencing revolution. While only a handful of parasite genomes were sequenced by 2005, the number has exploded to over 550 genomes (http://genomesonline.org) [20] by 2015. This number reflects both annotated and unannotated genomes and will already be out of date by the time this chapter is in print. Besides the technological advances, this increase in sequences has been aided by a number of initiatives with parasitology components. These include projects supported by the Wellcome Trust Sanger Institute in the United Kingdom and a number of parasite specific genome sequencing white papers supported by the National Institute of Allergy and Infectious Diseases (NIAID) Genomic Centers for Infectious Diseases (GCID) in the United States. Together these centers have generated sequence, assemblies and annotation from many important human and veterinary parasites. All data from these projects are available via project specific websites (ie. GeneDB: http://genedb.org) [21] and/or through the International Nucleotide Sequence Databases (GenBank, EMBL Nucleotide Sequence Database, and the DNA Data Bank of Japan [22–24]).

General features of protozoan parasite genomes

Amoebae

The amoebae, single celled eukaryotes that shared a most recent common ancestor with humans after plants but before fungi, are from a sparsely sampled and little studied domain of the tree of life. As with most protists the best known are those that cause disease in humans, which of the amoebae are the *Entamoebae* and the *Acanthamoebae*. The *Entamoebae* are intestinal parasites or commensals of a wide range of animals in addition to humans. The *Acanthamoebae* are free-living amoebae of interest to humans primarily as opportunistic

pathogens. These two and the social amoebae such as *Dictyostelium* species, are the best studied amoebae and those for which there exist sequenced genome assemblies, [25–27].

Entamoebae

The described species of *Entamoeba* are generally obligate parasites or commensals. They have simple life cycles consisting of a vegetative stage, the trophozoite, which lives in the host's large intestine and feeds upon bacteria and a transmissible stage, the cyst, which allows survival outside the host and transmission to a new host. Possible exceptions to these rules include two species (*Entamoeba moshkovskii* and *Entamoeba bangladeshi*) that can survive outside of the host and may be primarily free-living organisms, and one species (*Entamoeba gingivalis*) that colonises the mouth and may have lost the ability to form cysts instead being transmitted directly in the trophozoite form.

The human pathogen *Entamoeba histolytica* is the most studied species of the genus. A draft genome assembly was first published in 2005, with subsequent updates, though it remains fragmented and chromosomes cannot be defined [27–29]. Unusual features of the *E. histolytica* genome include an unusual organisation of tRNA genes, which occur in arrays of sets of tRNA genes separated by repetitive intergenic DNA [30], and rRNA genes encoded on extrachromosomal circular DNA occurring in multiple copies per cell [31]. Two features of *Entamoebae* associated with their anaerobic environments are the loss of the function and genome of the mitochondrion, which occurs as a relict organelle, the mitosome, and the related lateral transfer of genes, many involved in anaerobic metabolic processes and apparently derived from anaerobic bacteria [32].

Genomic re-sequencing suggests little nucleotide diversity among *E. histolytica*, even among lineages derived from widely separated geographical locations [33]. In contrast, gene copy number variation appears to be extensive [33], which may be associated with the genomic plasticity observed among *E. histolytica* lineages [34]. Studies using tRNA repetitive intergenic DNA or SNP markers also suggest very little linkage disequilibrium among markers, which suggests extensive outcrossing among parasite lineages [33, 35, 36]. Genetic diversity in other *Entamoeba* species is largely unknown, apart from studies of the 18S ribosomal RNA gene, which indicate that some 'species' may in fact be species complexes [37].

Genomic data exist for four other species of *Entamoeba*: *E. nuttalli*, *E. dispar*, *E. moshkovskii* and *E. invadens*. For the first three of these, the data are available but no reference publication yet exists. Most closely related to *E. histolytica*, *Entamoeba nuttalli* is a pathogen of macaques [38–40]. *Entamoeba dispar* infects humans and is of primary interest as a relative of *E. histolytica* (only recently defined as a separate species) that appears to be non-pathogenic [41]. *Entamoeba moshkovskii* is of uncertain status as a parasite or a free-living organism and has recently been associated with disease in humans [42, 43]. *Entamoeba invadens*, a pathogen of reptiles, is of primary interest as a model species for the process of encystation (which cannot be induced in axenic *E. histolytica* cultures). The genome of *E. invadens* is considerably larger than that of *E. histolytica* [44]. Genomic data for a number of *E. histolytica* strains, from a range of geographical locations and associated with different disease manifestations, are available via AmoebaDB (http://AmoebaDB.org) [45] (Table 1.1).

Table 1.1. Genome datasets of amoebae available in AmoebaDB.

Species	Strain	Dataset	Sequencing platform	Reference
Entamoeba histolytica	HM-1:IMSS	De novo genome assembly	Sanger	[27, 28]
Entamoeba histolytica	HM-1:IMSS-A	De novo genome assembly	454, Illumina	
Entamoeba histolytica	HM-1:IMSS-B	De novo genome assembly	454, Illumina	
Entamoeba histolytica	HM-1:CA	De novo genome assembly	454, Illumina	
Entamoeba histolytica	HM-3:IMSS	De novo genome assembly	454, Illumina	
Entamoeba histolytica	KU27	De novo genome assembly	454, Illumina	

Table 1.1 continued from previous page.

Species	Strain	Dataset	Sequencing platform	Reference
Entamoeba histolytica	KU48	De novo genome assembly	454, Illumina	
Entamoeba histolytica	KU50	De novo genome assembly	454, Illumina	
Entamoeba histolytica	MS96-3382	De novo genome assembly	454, Illumina	
Entamoeba histolytica	DS4-868	De novo genome assembly	454, Illumina	
Entamoeba histolytica	Rahman	De novo genome assembly	454	
Entamoeba histolytica	HM-1:IMSS-A	Re-sequencing	SOLiD	[33]
Entamoeba histolytica	HM-1:IMSS-B	Re-sequencing	SOLiD	[33]
Entamoeba histolytica	Rahman	Re-sequencing	SOLiD	[33]
Entamoeba histolytica	2592100	Re-sequencing	SOLiD	[33]
Entamoeba histolytica	MS84-1373	Re-sequencing	SOLiD	[33]
Entamoeba histolytica	MS27-5030	Re-sequencing	SOLiD	[33]
Entamoeba histolytica	PVBM08B	Re-sequencing	SOLiD	[33]
Entamoeba histolytica	PVBM08F	Re-sequencing	SOLiD	[33]
Entamoeba histolytica	HK-9	Re-sequencing	SOLiD	[33]
Entamoeba histolytica	IULA:1092:1	Re-sequencing	SOLiD	[33]
Entamoeba nuttalli	P19	De novo genome assembly	Illumina	
Entamoeba dispar	SAW760	De novo genome assembly	Sanger	
Entamoeba moshkovskii	Laredo	De novo genome assembly	454	
Entamoeba invadens	IP1	De novo genome assembly	Sanger	[44]
Acanthamoeba castellani	Neff	De novo genome assembly	Sanger, 454, Illumina	[25]
Acanthamoeba castellanii	Ma	De novo genome assembly	Illumina	
Acanthamoeba mauritaniensis	1652	De novo genome assembly	Illumina	
Acanthamoeba quina	Vil3	De novo genome assembly	Illumina	
Acanthamoeba astronyxis		De novo genome assembly	Illumina	
Acanthamoeba palestinensis		De novo genome assembly	Illumina	
Acanthamoeba sp (T4b-type)		De novo genome assembly	Illumina	
Acanthamoeba triangularis	SH621	De novo genome assembly	Illumina	
Acanthamoeba sp Incertae sedis		De novo genome assembly	Illumina	
Acanthamoeba sp	Galka	De novo genome assembly	Illumina	
Acanthamoeba lugdunensis	L3a	De novo genome assembly	Illumina	
Acanthamoeba culbertsoni	A1	De novo genome assembly	Illumina	
Acanthamoeba rhysodes	Singh	De novo genome assembly	Illumina	
Acanthamoeba lenticulata	PD2S	De novo genome assembly	Illumina	

Acanthamoebae

The *Acanthamoebae* are of importance for human health as a cause of keratitis when they infect the eye,often via contaminated contact lenses [46]. More usually, they are free-living, soil-dwelling pathogens of bacteria.

A draft genome assembly of *Acanthamoeba castellanii* was published in 2013 [25]. The genome encodes large families of genes involved in cell signalling and environmental sensing, such as protein kinases [25]. As in the *Entamoebae* a proportion of genes appear to have been acquired by lateral gene transfer, though the number of such genes in *A. castellanii* is larger and a larger proportion appear to have been acquired from aerobic and free-living bacteria [25]. Interestingly, in contrast to *Entamoeba* genes, which contain few introns, *Acanthamoeba* genes are intron-rich [25]. Thirteen additional *Acanthamoeba* species' genome sequence assemblies, representing a geographically diverse range of species and strains, were recently made available via AmoebaDB (Dr. Andrew Jackson, University of Liverpool; Table 1.1).

Giardia

Giardia intestinalis, also known as *Giardia duodenalis* or *Giardia lamblia*, is a unicellular protozoan parasite that infects the upper intestinal tract of humans and animals [47]. The disease, giardiasis, manifests in humans as an acute diarrhea that can develop to a chronic diarrhea but the majority of infections remain asymptomatic [47]. Giardiasis has a global distribution with 280 million cases reported annually, with its impact being more pronounced in the developing world.

G. intestinalis is divided into eight morphologically identical genotypes or assemblages (A to H). Only assemblages A and B have been associated with human infections and they are further divided into sub-assemblages: AI, AII, AIII, BIII, and BIV[48]. Despite extensive efforts to associate specific assemblages to symptoms, conflicting results have been obtained and there is to date no clear correlation between assemblage and symptoms.

Giardia, like the other diplomonads, has two nuclei and each nucleus is diploid, resulting in a tetraploid genome [49]. *G. intestinalis* has 5 different linear chromosomes with TAGGG repeats [50]. The study of the genome structure and architecture in *Giardia* using pulsed-field gel electrophoresis (PFGE) revealed differences in size of individual chromosomes within and between *G. intestinalis* isolates [51]. The size differences were attributed to frequently recombining telomeric regions and differences in copy number of rDNA arrays [50]. Evidence of aneuploidy has been suggested in individual *Giardia* cells based on cytogenetic evidence [52], with the most common karyotype differing between different assemblage A and B isolates.

The genomes of six *G. intestinalis* isolates, representing three different assemblages (A, B and E), are available to date [53–56]. The first genome to be sequenced was WB-C6 (assemblage A1), which has a haploid size of ~11.7 MB distributed over the five chromosome [55]. The compact genome contains few introns and promoters are short and AT rich. 6470 open reading frames (ORFs) were identified but only 4787 were later shown to be associated with transcription [57]. Genes are placed on both DNA strands and sometimes even overlapping. Reduction of components in metabolic pathways, DNA replication and transcription was also detected. Several genes had bacterial origin and are candidates of lateral gene transfer [55]. Variable surface proteins (VSPs) are involved in antigenic variation in *Giardia* and later analyses have shown that there are 186 unique VSP genes in the WB genome [53]. Chromosome-wide maps have been established by optical mapping of the WB genome [58]. The results resolved some misassemblies in the genome and indicated that the actual genome size of the WB isolate is 12.1 Mb, in close agreement with PFGE analyses. The major discrepancy was an underestimation of the size of chromosome 5, the largest of the *Giardia* chromosomes. Chromosome 5 contained an 819 kbp gap in the optical map, most likely rDNA repeats [58].

Shortly after publication of the WB genome the genome of the GS isolate (assemblage B) was sequenced using 454 technology [59]. However, the genome was highly fragmented with 2931 contigs. 4470 ORFs were identified and the genomes show 78% amino acid identity in protein coding regions. The repertoire of *vsp* genes was very different compared to the WB isolate but only 14 VSP genes were complete. The GS genome was later re-sequenced, resulting in 544 contigs and a much more complete repertoire of VSPs (275,[55]). Moreover, the GS

genome had a much higher level of allelic sequence heterozygosity (ASH) compared to WB (0.5% versus 0.01%). ASH was distributed differently into low and high ASH regions over the GS genomic contigs [59].

The third genome represents the first non-human isolate to be sequenced. The P15 isolate originates from a symptomatic pig (piglet no. 15) and belongs to assemblage E [54]. Assemblage E has been found to be more closely related to assemblage A than to assemblage B [48] and the identity of protein coding sequences was 90% between P15 and WB and 81% between P15 and GS [54], consistent with earlier results. Obtaining the sequence of three phylogenetically distinct *Giardia* groups (WB, P15 and GS) made it possible to assign lineage specificity to the genes identified in the three genomes. 91% of the genes (~4500 protein encoding genes) were found to be present in all three *Giardia* genomes (three-way orthologs) and 9% of genes are variable, most of which are members of four large gene families (the Variant-specific Surface Proteins (VSP), NEK Kinases, Protein 21.1 and High Cysteine Membrane Proteins (HCMP)). The highest number of isolate-specific genes (38) was found in the P15 isolate, followed by GS (31) and WB (5). The P15 and GS isolates shared 20 proteins to the exclusion of WB, with 13 of these found in a cluster of 20 kbp in the P15 genome[54]. Interestingly the ORFs in this genomic cluster are not expressed in any of the conditions tested. The chromosomal architecture in *Giardia* show core gene-rich stable regions with maintained gene order interspersed with non-syntenic regions harboring VSPs and other non-core genes. These regions often have a higher GC% and show nucleotide signatures that deviate from surrounding regions, in part due to the common occurrence of VSP and high-cysteine membrane protein (HCMP) genes that are more GC-rich than the genome on average. The level of ASH in the P15 isolate was lower than in the GS isolate, 0.0023% [54].

Three assemblage AII isolates have been sequenced (DH1, AS98 and AS175 [53, 56]. The amount of genetic diversity was characterized in relation to the genome of WB, the assemblage A reference genome. The analyses showed that the divergence between AI and AII is approximately 1 %, represented by ~100,000 single nucleotide polymorphisms (SNP) distributed over the chromosomes with enrichment in the variable genomic regions containing VSPs and HCMPs [56]. The level of ASH in two of the AII isolates (AS98 and AS175) was found to be 0.25–0.35 %, which is 25–30 fold higher than in the WB isolate and 10 fold higher than the assemblage AII isolate DH1 (0.037 %, [56]).

There is a need for further genomic analyses of *Giardia* genomes. The assemblage A (WB) and B (GS) reference genomes can be improved, which will facilitate reference-based genome mapping of data from clinical and environmental isolates. More isolates from the A and B assemblages should be sequenced so that all the genetic differences between the human infecting isolates can be identified. Genomic information from the remaining assemblages, C-D, F-H can reveal species-specific genomic features. Sequence data from other *Giardia* species like *Giardia muris* will be important for further studies of the evolution of *Giardia* biology and virulence. In addition to the underlying genomic sequence and annotation, a number of functional datasets are available for the GiardiaDB.

Cryptosporidium

Cryptosporidium are protozoan parasites with significant impact to the health of humans and livestock. They infect the intestinal and gastric epithelium of a variety of vertebrates, causing a disease known as cryptosporidiosis. Human cryptosporidiosis is responsible for diarrhea-induced death of young children in developing countries, and in immune-compromised adults it constitutes an acute, usually self-limiting, diarrheal illness that results in significant morbidity and sometimes death. A recent study found *Cryptosporidium* to be the second leading cause of moderate-to-severe diarrhea in developing countries, and diarrheal diseases to be the second leading cause of death among children under five globally [60].

There are no licensed vaccines against *Cryptosporidium* and the only FDA-approved drug (Nitazoxanide) is only effective in immunocompetent patients. Thus, the development of alternative therapeutic agents and vaccines against this disease is urgently required, and remains a high public health priority. The lack of a practical and

reproducible axenic *in vitro* culture system for *Cryptosporidium* is a major limitation to the development of specific anti-cryptosporidial vaccines [61, 62]. Advances in next-generation sequencing technologies and in genome assembly and annotation methodologies [63–66] have facilitated the generation of -omics data for *Cryptosporidium*, with genomics resources now available for multiple *Cryptosporidium* species (Table 1.2, [67]). These developments prompted a shift to *in silico* studies aiming to identify a wide pool of potential vaccine targets, to be further filtered according to properties common to antigens [68]. This approach is similar to reverse vaccinology studies that have led to licensed vaccines in other organisms [69, 70], and is particularly promising in organisms that, like *Cryptosporidium*, are difficult to cultivate continuously in the laboratory.

Apart from human, *Cryptosporidium* species infect other vertebrates including fish, birds and rodents, and some species are capable of zoonotic transmission [71, 72]. Some have a somewhat restricted host range, such as *Cryptosporidium hominis*, a human parasite that infects the small intestine, *Cryptosporidium muris*, a gastric parasite of rodents, and *Cryptosporidium baileyi*, an avian parasite. *Cryptosporidium parvum* and *Cryptosporidium meleagridis* have a wider host range and are known to infect both avian and mammalian species, including humans. *C. parvum* and *C. hominis* are considered class B agent of bioterrorism and are significant causes of gastrointestinal infections worldwide.

Table 1.2. *Cryptosporidium* species with completed or draft genomes.

Species	Number of draft genomes	Natural host range	Predilection site
C. hominis	8	Human, primates	Intestinal
C. parvum	8	Human, Bovine	Intestinal
C. meleagridis	1	Various vertebrates	Intestinal
C. baileyi	1	Birds	Respiratory
C. muris	1	Rodents	Gastric
C. sp. chipmunk LX-2015	1	Rodents, Human	Intestinal

Cryptosporidium genomic resources

Cryptosporidium genomes are compact, with >75% consisting of protein-coding sequences, have an average size of approximately 8.5 to 9.5 mega base pairs (Mbp), and each encode ~4000 genes (Table 1.2). *C. parvum* (isolate IOWA II) was the first species for which a genome was published [73]. The genome was found to be 9.1 Mbp in length, assembled into thirteen supercontigs. Pulsed-field gel electrophoresis studies had shown the nuclear-encoded genome to consist of 8 chromosomes, and therefore the assembly includes five unresolved gaps. About 5% of the 3,807 predicted protein-coding genes in this assembly contained introns, and the average gene length was 1,795 base pairs (bp). At about the same time the genome of *C. hominis* (isolate TU502) was published [74]. Since the two species were known to be closely related, with about 95-97% DNA sequence identity between them, the *C. hominis* genome was sequenced to a much lower depth of coverage. The primary goal was to identify differences relative to *C. parvum*, rather than reconstruct a gold-standard genome assembly. Consequently, this assembly is much more fragmented, with the likely 8 chromosomes split among 1,413 contigs, which are grouped into ~240 scaffolds.

There were some fundamental differences between the annotated gene sets in the two species. The average gene length of *C. hominis* was 1,360 bp, about 500 bp less than that of *C. parvum*, and about 5-20% of the *C. hominis* genes were predicted to contain introns, compared to 5% in *C. parvum* [73, 75]. In addition, only 60% of the *C. hominis* genome was estimated to be coding compared to 75% for *C. parvum*. These differences are remarkable for such closely related taxa and were thought to be due to erroneous gene models in *C. hominis* due to the high degree of genome fragmentation. To address these questions, the genome assembly for *C. hominis* has recently been re-sequenced, assembled and annotated, improving the assembly from draft to "nearly finished" form, with

preliminary data available in CryptoDB.org. This effort increased the average gene length by 500 bp, bringing it to 1,845 bp, in line with gene length in *C. parvum* (Table 1.2). The improved genome assembly consists of only 120 contigs, a ten-fold reduction in contig number relative to the original *C. hominis* assembly. The genome assembly is more comprehensive, with an additional 370 Kb sequence, also now comparable in length to that of *C. parvum*. Finally, there was a 25% increase in the predicted fraction of the genome that encodes for proteins. The now marked similarities between the re-annotated *C. hominis* gene set and that of *C. parvum* provide encouraging evidence that the predicted genes are a significant improvement over the original annotation, but validation of gene structures awaits community effort. *C. parvum* IOWA II was also recently re-annotated, based on full-length cDNA clone sequences and RNA-Seq data (Table 1.2, [76]).

Both *C. hominis* and *C. parvum* are intestinal parasites. *C. muris* (isolate RN66), the third species sequenced, was chosen for two primary reasons: its evolutionary distance to *C. hominis* and *C. parvum*, and the fact that it is a gastric species, which is rare among *Cryptosporidium* parasites. Currently, the field is rapidly expanding, with the genome sequence for several isolates of *C. parvum* and of *C. hominis* now available, as well as the genomes of other species (Table 1.3).

The availability of multiple isolate genomes per species allows analyses that can shed light into species evolution, including age and population structure, and will facilitate studies that address key questions of great translational impact, including the amino acid sequence variations in current candidate vaccine antigens, and the identification of genomic correlates of virulence whenever isolates with different pathogenic potential are available. In an effort to support research that addresses key questions in the evolution of the *Cryptosporidium* genus, and the discovery of parasite-encoded factors that control host specificity, *C. meleagridis* UKMEL1 was sequenced, a species which appears to lack host specificity and that is considerably more distantly related to *C. hominis* and *C. parvum* than they are to each other, but a closer relative to them that is *C. muris*. *C. baileyi* can complete its life cycle in embryonated chicken eggs, of critical importance for the establishment of an avian model system of cryptosporidiosis, and *C. baileyi* TAMU-09Q1 was sequenced to support its development of such a system. Determining the proportion of Cryptosporidium infections that are caused by human-specific parasites rather than by zoonotic infections remains a critical question in the field. Accordingly, the genome of a zoonotic infection by a *Cryptosporidium* species with origin in the chipmunk was conducted with the goal of identifying genotyping markers that differentiate among *Cryptosporidium* subtypes [77].

A major challenge for the generation of *Cryptosporidium* whole genome sequence data has been the need to propagate the parasites in vertebrate hosts, a step needed to generate DNA material in sufficient quantity and of the quality need for use in high-throughput sequencing applications. A novel method for preparing genomic *Cryptosporidium* DNA directly from human stool samples that satisfies the criteria these applications has now been developed [78]. The authors used this approach to generate five assemblies each for *C. parvum* and *C. hominis*. Finally, a new *C. hominis* (isolate UdeA01) also isolated from human stool has been sequenced independently [76].

All the genomics data described above is publicly available through CryptoDB [67]. This database also provides a platform to easily query the annotation and a variety of pre-computed analysis data (including homology information across taxa). Multiple aspects of the data can be easily visualized, including synteny, polymorphism and expression data. CryptoDB also contains *Cryptosporidi*um information other than genome sequences, including gene expression and proteomics data (Table 1.4).

Table 1.2. **Genome statistics for representative *Cryptosporidium* species.**

Species	Isolate	GenBank accession	Assembly length (bp)	No. contigs	Largest contig (bp)	No. protein-coding genes	Average gene length (bp)	Percent coding
C. hominis	TU502 (2004)	AAEL00000000	8,743,570	1413	90,444	3,886	1,360	60.4%
C. hominis	TU502_new (2014)	SUB482083	9,110,085	120	1,270,815	3,745	1,845	75.8%
C. parvum	Iowa	AAEE00000000	9,103,320	13	1,278,458	3,807	1,795	75.3%
C. parvum[a]	Iowa	AAEE00000000	9,103,320	13	1,278,458	3,865	1,783	75.7%
C. meleagridis	UKMEL1	SUB482042	8,973,224	57	732,862	4,326	1,861	89.7%
C. baileyi	TAMU-09Q1	SUB482078	8,502,994	153	702,637	3,700	1,776	77.3%
C. muris	RN66	AAZY02000000	9,245,250	84	1,324,930	3,934	1,780	79.2%

a 2015 re-annotation

Table 1.3. ***Cryptosporidium*** **genomes available in CryptoDB.**

Species	Isolate	Year	Sequencing Institution[a]	GenBank Accession	RNA-Seq SRA Accession	Assembly length (bp)	No. contigs	Largest contig (bp)
C. hominis	TU502	2004	VCU	AAEL01	-	8,743,570	1422	90,444
		2013		AAEL02	-	8,915,516	358	282,140
C. hominis	TU502_new	2014	IGS/Tufts	SUB482083	SRS566230	9,110,085	120	1,270,815
C. hominis	37999	2014	CDC	JRXJ01	-	9,054,010	78	1,029,232
C. hominis	UKH1	2014	IGS/Tufts	SUB482088	SRS566214	9,141,398	156	542,781
C. hominis	UKH3	2015	PHW	LJRW01	-	9,136,308	34	1,295,005
C. hominis	UKH4	2015	PHW	LKHI01	-	9,158,280	18	1,295,931
C. hominis	UKH5	2015	PHW	LKHJ01	-	9,179,731	18	1,281,265
C. parvum	Iowa II	2004	Univ. Minnesota	AAEE01	-	9,087,724	18	1,278,458
C. parvum	UKP2	2015	PHW	LKHK01	-	9,126,082	18	1,285,807
C. parvum	UKP3	2015	PHW	LKHL01	-	9,085,686	18	1,258,884
C. parvum	UKP4	2015	PHW	LKHM01	-	9,001,535	18	1,283,549
C. parvum	UKP5	2015	PHW	LKHN01	-	9,283,240	18	1,284,088
C. parvum	UKP6	2015	PHW	LKCK01	-	9,112,937	18	1,296,567
C. parvum	UKP7	2015	PHW	LKCL01	-	9,221,024	18	1,295,191
C. parvum	UKP8	2015	PHW	LKCJ01	-	9,203,314	18	1,288,507
C. meleagridis	UKMEL1	2014	IGS	SUB482042	-	8,973,224	57	732,862
C. baileyi	TAMU-09Q1	2014	Texas A&M	SUB482078	SRS566232	8,502,994	153	702,637
C. muris	RN66	2008	TIGR	AAZY02	SRS000463	9,238,736	97	1,182,920

Table 1.3 continued from previous page.

Species	Isolate	Year	Sequencing Institution[a]	GenBank Accession	RNA-Seq SRA Accession	Assembly length (bp)	No. contigs	Largest contig (bp)
C. sp.	chipmunk LX-2015	2015	CDC	JXRN01	-	9,509,783	853	478,353

a CDC: Division of Foodborne, Waterborne, and Environmental Diseases, Centers for Disease Control and Prevention; IGS: Institute for Genome Sciences; PHW: Public Health Wales (Microbiology); TIGR – The Institute for Genome Research; VCU: Virginia Commonwealth University.

Table 1.4. Other *Cryptosporidium* genomic resources available in CryptoDB.

Data type	Description	Species	Reference
EST	EST library and predicted full length cDNA	C. parvum HNJ-1	[79]
EST	ESTs from Database of Expressed Sequence Tags (dbEST)	C. baileyi TAMU-09Q1, C. hominis TU502, C. meleagridis UKMEL1, C. muris RN66, C. parvum Iowa II	[80]
RT-PCR	Expression profiling of life cycle stages post-infection	C. parvum Iowa II	[81]
Microarray	Global gene expression in oocysts (environmental stage) and oocysts treated with UV	C. parvum Iowa II	[82]
RNA-Seq	Transcriptome of sporozoites and HTC-8 infection time course	C. parvum Iowa II	(Lippuner et al.)
RNA-Seq	Transcriptome in normal culture conditions	Chromera velia CCMP2878, Vitrella brassicaformis CCMP3155	[83]
Mass Spectrometry	Enriched cytoskeletal and membrane fractions	C. parvum Iowa II	[84]
Mass Spectrometry	Mitochondrial fraction proteomics	C. parvum Iowa II	(Putignani et al.)
Mass Spectrometry	Proteome of intact oocyst, oocyst wall and sporozoites by linear ion trap MS	C. parvum Iowa II	[85]
Mass Spectrometry	Proteome during sporozoite excystation	C. parvum ISSC162	[86]
Mass Spectrometry	Sporozoite peptides from 2D gel LC-MS/MS analysis	C. parvum Iowa II	[87]
SNPs	SNPs determined by aligning high throughput sequencing reads of C. parvum TU114 to the C. parvum reference genome	C. parvum TU114, C. parvum Iowa II	[75]

Piroplasms

Piroplasms are a vast group of poorly characterized Haemosporidia that are named after their pyriform (pear-shaped) structure visible during intracellular stages in the host erythrocytes. They are found in numerous mammals, birds, and reptiles, and are often transmitted by ixodid ticks after parasite replication in the tick gut [88]. While little is known about the life cycle of most piroplasms, well-described species of *Theileria* commonly infect mammalian host leukocytes, followed by a tick-infective stage in red blood cells (RBCs), while *Babesia* do not have a leukocyte-infective stage [89, 90]. Some *Babesia* species are known to infect humans (*B. microti, B. divergens, B. duncani*), where they cause a malaria-like disease [89]. The diseases caused by these parasites can lead to fevers and even death in equid and ruminant livestock species, all around the world. Consequently, most of the genomics resources developed for piroplasm research to date have focused on species that infect bovids (Table 1.5). Most of these resources are available through PiroplasmaDB (http://PiroplasmaDB.com).

The first piroplasm genomes were published in 2005, and consisted of *Theileria* species of domestic cattle and wild buffalo. *T. parva* causes a tremendous economic impact in eastern, central and southern Africa [90], while

T. annulata is distributed throughout much of southern Asia and southeast Europe [94]. Their genomes are small at ~8.3 Mbp in length, are AT-rich with GC-content of 33%, and contain ~4,000 nuclear protein-coding genes. These properties are similar to the genomes of other Piroplasmida that have been sequenced since (Table 1.6). Several genomic features were uncovered that are typical of other sequenced piroplasm genomes, such as the presence of telomeric multi-gene families, and several incomplete or absent biosynthetic pathways, implying a critical dependence on salvaging resources from their hosts [97]. These two piroplasms are unique, however, in their ability to transform host leukocytes to have cancer-like phenotype. This phenotype correlates with the expansion of two multi-gene families: the Subtelomere-encoded Variable Secreted Protein (SVSP) gene family and the *T. annulata* schizont AT-hook/*T. parva* Host Nucleus (TashAT/TpHN) gene families [98, 99]. Two other *Theileria* species have been sequenced, *T. orientalis* [98], an economically important pathogen of cattle in eastern Asia, and *T. equi* [95], which has a worldwide distribution and infects equids. These two genomes have many similar features, with the exception that the genome of *T. equi* is larger, mostly due to a significant increase in the number of species-specific genes, including antigen-encoding families such as the Equi Merozoite Antigen (EMA) family [95].

With a genome size of ~8.2Mpb, the *B. bovis* genome sequence revealed a genomic organization that is remarkably similar to *T. parva*, with extensive synteny and multiple, large multi-gene families potentially contributing to host immune evasion [106]. However, the smallest apicomplexan genome sequenced to date is *B. microti*, the principal agent of human babesiosis and a common pathogen transmitted by blood transfusions [89, 101]. With a genome size of 6.5Mbp, *B. microti* represent the closest record of a natural representation of an apicomplexan "core genome" and comparative genomics with this reduced genome could yield insights into the most essential gene products of apicomplexans that could make excellent chemotherapeutic targets. *B. microti* is also the only example of an apicomplexan with a circular mitochondrial genome[101].

One apicomplexan with somewhat unclear phylogenetic position is *Cytauxzoon felis*. While originally considered a separate genus, the existence of exo-erythrocytic forms, particularly schizonts, in macrophages/monocytes indicates that this parasite might be more appropriately considered in the family Theileriidae. *C. felis* is an emerging pathogen of domestic cats (*Felis catus*) in the southern United States, and as such its genome was sequenced in an effort to identify potential vaccine targets [93]. With a 9.1 Mbp genome, it has more protein coding genes in common with *T. parva* than it does with *B. bovis*, and was found to encode a gene that is syntenic with a block of genes around the *T. parva* antigen, and vaccine candidate, p67 [93].

There are currently no licensed vaccines against apicomplexans for use in humans, although the RTS,S malaria vaccine is close to licensure. With a few notable exceptions, such as coccidiosis (*Eimeria*), toxoplasmosis (*Toxoplasma*), and East Coast Fever (*Theileria parva*) vaccines, very few vaccines against piroplasms have been used on a commercial scale, which may be due, in part, to antigenic diversity in these parasites [107]. Genomic resources have also recently started to become available for some piroplasms (Table 1.7). These data are critical for identification of potential virulence genes, mapping recombination hotspots, and estimate genome-wide variation among various isolates, including vaccine strains [105]. One weakness of piroplasm whole-genome datasets is their reliance on *ab initio* gene predictors for the majority of their structural annotations (determining where exons start and end in the genome). Given the fact that these genomes are smaller, denser, and more AT-rich than most eukaryotes sequenced to date, these gene predictors may not be optimal for gene prediction in these genomes, and experimental evidence should be rigorously incorporated into genome re-annotation efforts in order to take full advantage of the genome sequences that are present for these apicomplexans. The coupling of whole-genome variation data with gene expression data is a powerful method to give insight into gene structure, variation and function, and will hopefully assist the design of better prophylaxis against piroplasm-mediated diseases.

Table 1.5. The first publication of available piroplasm whole genome sequences, and a few features of their genomes. All of these are available at PiroplasmaDB, with the exception of *B. divergens*.

Genus	Species	Strain(s)	Year Published	Reference	Hosts	Assembly Length (Mbp)	Genome %GC	# Nuclear, Protein-encoding genes
Babesia	bigemina	BOND*, PR, BbiS3P, JG29	2014	[91]	Bovids	13.8	51	4,457
Babesia	bovis	T2Bo	2007	[92]	Bovids	8.2	41.8	3,671
Babesia	divergens	1802A*, Rouen 1987	2014	[91]	Bovids	9.6	42	4,134
Babesia	microti	RI	2012	[89]	Rodents, Humans	6.5	36	3,513
Cytauxzoon	felis	Winnie	2013	[93]	Felids	9.1	31.8	4,323
Theileria	annulata	Ankara	2005	[94]	Bovids	8.4	32.5	3,792
Theileria	equi	WA	2012	[95]	Equids	11.6	39.5	5,330
Theileria	parva	Muguga	2005	[90]	Bovids	8.3	34.1	4,035
Theileria	orientalis	Shintoku	2012	[96]	Bovids	9	41.6	3,995

* = genomic statistics shown for this isolate; %GC = percentage GC content for the whole genome.

Table 1.6. Whole-genome data for several piroplasm species. These resources are not available at PiroplasmaDB, but can be found associated with their respective references.

Genus	Species	Strains	Year Published	Data Type	Reference
Babesia	bovis	C9.1	2014	WGS	[91]
Babesia	divergens	None Indicated	2014	WGS, draft assembly	[100]
Babesia	microti	R1, Gray	2013	Complete Genome Assembly	[101]
Babesia	bovis	T2Bo_Vir., T2Bo_Att., L17_Vir., L17_Att., T_Vir., T_Att.	2011	WGS	[102]
Theileria	parva	Marikebuni, Uganda, MugugaMarikebuni, MugugaUganda	2012	WGS, draft assemblies	[103]
Theileria	parva	ChitongoZ2, KateteB2, Kiambu Z464/C12, MandaliZ22H10, Entebbe, Nyakizu, Katumba, Buffalo LAWR, Buffalo Z5E5	2013	WGS	[104]
Theileria	parva	Muguga, Kiambu5, Serengeti-transformed	2015	WGS	[105]

Table 1.7. Gene expression data not found at PiroplasmaDB for several piroplasm species. Most expression data, including more EST data, is found at PiroplasmaDB for piroplasms.

Genus	Species	Strains	Year Published	Reference	Data Type
Babesia	bovis	T2Bo	2007	[108]	Microarray
Babesia	bovis	T2Bo	2013	[109]	RNAseq
Babesia	bovis	T2Bo_Vir., T2Bo_Att., L17_Vir., L17_Att., T_Vir., T_Att.	2013	[109]	Microarray, RNAseq
Babesia	bigemina	PR	2014	[91]	LC-MS
Cytauxzoon	felis	Winnie	2013	[93]	EST
Theileria	annulata	Ankara	2012	[110]	Microarray

Table 1.7 continued from previous page.

Genus	Species	Strains	Year Published	Reference	Data Type
Theileria	annulata	Ankara	2013	[111]	Microarray
Theileria	annulata	Ankara	2013	[112]	LC-MS/MS
Theileria	parva	Muguga	2005	[113]	MPSS

Plasmodium reference genomes

To date several complete reference genomes of *Plasmodium*, the aetiological agent of malaria, have been sequenced. Advances in technology have also led to the sequencing of many additional lab strains and clinical isolates. The first reference to be published in 2002 was *P. falciparum* 3D7 [12], the species responsible for the majority of human morbidity. Additional genomes of species that infect humans have been sequenced (*P. vivax* [114]) or are in the process of being sequenced and *analysed* (*P. malariae* and *P. ovale*). The simian-and human infecting *P. knowlesi* [115], the chimpanzee malaria *P. reichenowi* [116] and the simian malaria parasite *P. cynomolgi* [117] are also part of the reference genome collection. Draft genomes of three rodent malaria parasites that are widely used as model systems, *P. yoelii yoelii* [118], *P. chabaudi chabaudi* AS and *P. berghei* ANKA were initially sequenced and analysed in 2005 [119]. Due to the highly fragmented nature of these genomes, they were re-sequenced in 2014 [120]. Two avian malaria genomes, *P. relictum* and *P. gallinaceum* have been sequenced and are in the process of being analysed. They will provide a valuable missing link to understand the evolutionary context of human malaria. All of the published genomes mentioned above can be searched in PlasmoDB [121] and GeneDB [21].

The publication of *P. falciparum* 3D7 in 2002 was a major milestone [12]. It enabled the malaria community to systematically analyse the gene content and tailor their experiments based on genomic data. This is also shown by over 2000 citations of the genome paper since publication. After the initial publication, assembly and annotation of the *P. falciparum* 3D7 genome has been continuously improved over time. In 2011 a new *P. falciparum* 3D7 assembly (version 3) was made publicly available. This new version includes the correction of major mis-assemblies. The current genome version has a size of 23.3 Mb and encodes 5429 genes (Table 1.8). It is highly AT-rich with a GC-content of only 19.3%. The overall structure of *Plasmodium* genomes sequenced to date is very similar (Table 1.8). The nuclear genome consists of 14 chromosomes, the size ranges from 19Mb to 26Mb with a comparable number of genes. About three quarters of genes are conserved across all *Plasmodium* genomes, representing the core genome. *Plasmodium* genomes also exhibit a high degree of synteny.. The majority of the variation between *Plasmodium* species is found in the subtelomeric regions at the end of the chromosomes. In these regions, each of the *Plasmodium* species has a unique set of gene families that are often involved in immune evasion and virulence. The most important gene family in *P. falciparum* 3D7 is the VAR gene family that encodes the erythrocyte membrane protein 1 (PfEMP1). PfEMP1 plays a role in antigenic variation. Of around 60 gene family members, only one protein is expressed on the surface of infected red blood cells at a time. PfEMP1 can also bind to host endothelial receptors and therefore plays an important role in pathogenicity. Additional gene families include rifins and stevors. It has been recently shown that rifins are expressed on the surface of infected red blood cells where they mediate microvascular binding of infected red blood cells [122]. The function of stevors is unknown. Both, rifins and stevors belong to the PIR (Plasmodium interspersed repeats) superfamily. This superfamily is the only subtelomeric gene family found so far that is present in all of the *Plasmodium* species.

Closely related to *P falciparum* is the chimpanzee malaria parasite *P. reichenowi*. A comparative genomics analysis only showed minor differences between these two genomes. There is an almost complete co-linearity in the core areas of the genome. The organisation of var genes and other virulence-associated genes is also conserved. Differences were found in the reticulocyte-binding proteins, a gene family involved in invasion. These

genes encode ligands that are important for the recognition of host erythrocytes. Members of this gene family are located on chromosome 13, where two almost identical genes are present in *P. falciparum* (RH2a and RH2b). *P. reichenowi* lacks RH2a, but encodes a new reticulocyte-binding protein, RH7. The most significant difference between *P. reichenowi* and *P. falciparum* was found in the rifin and stevor multigene families. There are currently 463 rifins and 66 stevors annotated in *P. reichenowi*, while *P. falciparum* only encodes 185 rifins. The difference in this multigene family also explains the difference in the overall number of genes found in the nuclear genome (Table 1.8).

P. vivax is the major source of human malaria outside of Africa. In contrast to *P. falciparum* this species has a dormant stage in the human liver and can stay inactive for years. The nuclear genome of the Salvador I strain *P. vivax* has a size of 26.8 Mb and encodes 5433 genes (Table 1.8). With a GC-content of 42.3% *P. vivax* has the highest GC-content found so far in *Plasmodium*. Unique to *P. vivax* is an isochore structure. Chromosomes have AT-rich chromosome ends and internal-regions of high GC-content.

Closely related to *P. vivax* is the malaria parasite *P. knowlesi*. *P. knowlesi* is primarily a simian infecting malaria parasite, but has also been reported to cause natural infections in humans mainly in South East Asia. The nuclear genome has a size of 24.4 Mb, a GC-content of 38.6%, the number of protein-coding genes is 5290 (Table 1.8). There are two novel features in the *P. knowlesi* genome. The major variant gene families that are usually located in subtelomeres, are found in chromosome-internal regions dispersed on all 14 chromosomes. These regions are often also associated with intrachromosomal telomeric repeats. Another unusual feature unique to *P. knowlesi* is a phenomenon called molecular mimicry. KIR proteins that are part of the PIR superfamily contain stretches of sequences that are identical to the host proteins AHNAK and CD99, which has a critical immunoregulatory role in host T-cell function. It is speculated that these proteins might interfere with host recognition processes. Another important gene family is the SICAvar (schizont infected cell agglutination) gene family. SICAvars are expressed on the surface of infected erythrocytes and are the largest family of variable surface antigens in *P. knowlesi*.

Phylogenetically related to *P. knowlesi* and *P. vivax* is the simian malaria parasite *P. cynomolgi*. *P. cynomolgi* is used as a model organism for human *P. vivax* infections. Both share the ability to form a dormant liver stage. Strain B of *P. cynomolgi* has been sequenced and published in 2011 [117]. The genome has a size of 26.2 Mb and encodes 5722 genes (Table 1.8). Of those, around 90% have 1:1 orthologs to *P. vivax* and *P. knowlesi*. *P. cynomolgi* and *P. vivax* share a common isochore structure, while the presence of intrachromosomal telomeric repeats is common to *P. cynomolgi* and *P. knowlesi*. Comparative genome analysis found a number of copy-number variants in multigene families, e.g. in reticulocyte-binding proteins.

Of particular interest are the rodent malaria parasites, *P. berghei*, *P. chabaudi chabaudi* and *P. yoelii yoelii*. They are used as model organisms for experimental studies of human malaria. The genome size of the rodent malaria parasite genomes ranges from 18.8 Mb to 22.7 Mb (Table 1.8). The GC-content is around 22%. *P. yoelii yoelii* has the highest number of genes in the nuclear genome, mostly due to a large expansion of PIR genes (980). Gene synteny is conserved along the 14 chromosomes, with only one known synteny breakpoint. Analysis of gene families in the rodent-infective species reveals that the gene family is the PIR gene family. [120]. The second largest gene family encodes fam-a proteins. Fam-a proteins are exported to the infected red blood cell and are expanded in the rodent malaria parasites. All other *Plasmodium* genomes sequenced to date have only one fam-a family gene. The number ranges from 161 in *P. yoelii yoelii*, to 148 in *P. chabaudi chabaudi* and 74 in *P. berghei*.

Table 1.8. ***Plasmodium*** **reference genomes.**

	P. falciparum 3D7 (v3) [3]	*P. reichenowi* CDC (v1) [3]	*P. vivax* Sal1[1]	*P. knowlesi* H (v2)[3]	*P. cynomolgi* B [2]	*P. berghei* ANKA (v3)[3]	*P. chabaudi* AS (v3) [3]	*P. yoelii* yoelii 17X (v3) [3]
Genome size (Mb)	23.2	24	26.8	24.4	26.2	18.8	18.9	22.7
No. of chromosomes	14	14	14	14	14	14	14	14
G+C content (%)	19.3	19.2	42.3	38.6	40.4	22	23.6	21.5
No. of unassigned contigs	0	237	2745	148	1649	5	0	138
No. of genes [4]	5429	5736	5433	5290	5722	5034	5183	5948
% of genes with introns	54.1	55.9	52.1	54	75.8	52.4	53.5	59.8
No. of PIRs [5]	227	529	346	70	256	217	208	980
manually curated	yes	yes	no	yes	no	yes	yes	yes

(1) Carlton et al., Nature 455, 757-63 (2008)
(2) Tachibana et al., Nat Genet. 44, 1051-5 (2012)
(3) genome version from 1.10.2015
(4) including pseudogenes and partial genes, excluding non-coding RNA genes
(5) including pseudogenes and partial genes

Trypanosomatids

Trypanosomatids are a group of parasitic unicellular flagellate eukaryotes. Their range of hosts is diverse and includes humans and as well as a wide variety of species from both the animal and plant kingdoms. Trypanosomatids belong to the kinetoplastida, which is included in the phylum Euglenozoa, a branch that diverged early in the eukaryotic tree [123, 124]. While a number of Kinetoplastida are pathogenic parasites most are free-living organisms found in soils and aquatic habitats. The name Kinetoplastida derives from the presence of large amounts of mitochondrial DNA, visible by light microscopy as a dense mass known as the kinetoplast with its contained DNA referred to as kDNA. Trypanosomatids are obligate parasites that can be monoxenous or dixenous (usually an insect vector and other animal or plant [125]).

Trypanosomatid Genomes

The nuclear genome of trypanosomatids has some unusual characteristics when compared with other eukaryotic genomes. Their genome is organized in polycistronic transcriptional units (PTUs) and the production of individual mRNAs from PTUs requires trans-splicing of a splice leader (SL) sequence [126]. PTUs are well conserved and exhibit a high degree of synteny between species. The kDNA has an unusual physical structure, being arranged in circles of DNA that are interlocked in a chain-mail like network. These mitochondrial mRNAs require post-processing in the form of insertion and deletion of uridines before being translated into proteins, a process known as RNA editing [127, 128]. Other peculiarities of trypanosomatid genomes include the almost complete lack of introns, kinetoplastid-specific histone modifications and histone variants, unique origins of replication in some genera, a special DNA base (Base J) [129], and the transcription of protein-coding genes by RNA pol I in African trypanosomes, a behavior unique among eukaryotes [130]. Although none of these unusual features seem to be exclusive of trypanosomatids and are also present, at least in some basic form, in

other free-living kinetoplastids, they may be related to the development of parasitism in trypanosomatids [124, 131].

Regulation of Gene Expression in Polycistronic Transcriptional Units

One of the most striking characteristics of trypanosomatid genomes is the organization of their protein-coding genes into long polycistronic transcriptional units (PTUs) that contain tens to hundreds of genes in the same orientation. Individual mRNAs are produced from the precursor mRNA by the 5' trans-splicing of a capped mini-exon or splice leader sequence (SL), followed by the polyadenylation of the 3' end. The 5' trans-splicing is linked to the polyadenylation of the upstream gene. Gene order within PTUs is highly conserved among trypanosomatids and the main differences are usually in the regions between the PTUs and at the ends of the chromosomes [126, 132].

The genes included in a PTU are functionally unrelated and can be expressed at different times of the cell cycle or in different life stages. Nonetheless, each PTU is transcribed from a single transcriptional start site (TSS), severely limiting the amount of regulation that could be provided by the induction or repression of promoters. In some cases correlation between the location of a gene in a PTU and its expression level has been described. For example, in *T. brucei*, genes downregulated after heat shock tend to be closer to the transcriptions start site (TSS), while upregulated genes tend to be more distal. Also, the position of the genes along the PTUs correlates with gene regulation during the different cell cycle stages. However, most of the genes do not seem to be ordered depending on their transcriptional regulation [123, 126, 133].

In most organisms, the start of transcription is a fundamental step in the regulation of gene expression. In trypanosomatids this layer is constrained, but a swift and specific regulation of gene expression is still needed. Dixenous species like *T. brucei* or *L. major*, have complex life cycles that require fast and extensive changes in morphology and metabolism. These changes depend, ultimately, on changes in gene expression. For example, the parasite has to quickly adapt to differences in temperature, energy sources and host immune system [130, 133]. Besides the regulation at the start of transcription, it is possible to modulate other steps in the transcription and translation process. Additional levels of control include transcriptional elongation, mRNA processing (trans-splicing and polyadenylation), export from the nucleus, mRNA degradation (in the cytoplasm and nucleus), translation (start and elongation) and protein degradation [126, 132].

Both mRNA processing and the control of the mRNA stability are important regulatory steps in trypanosomatids. The stability of the mRNAs depends on elements present in the 3' UTRs, for instance, duplicated genes in tandem arrays can be differentially regulated due to differences in their 3' UTRs. In *T. brucei,* the range of half-lives of mature mRNAs is very diverse and is also determined by the life-cyle stage. In addition, the half-life of a mRNA not only depends on the stability of the mature mRNA but also on the rates of destruction of the precursor mRNA. If a mRNA undergoes a late or delayed polyadenylation it is more susceptible to being degraded, even before finishing maturation [132, 134].

Trypanosomatids contain a large number of RNA binding proteins (RBPs) that likely regulate expression levels by binding to regulatory elements in the 3' UTRs of the mRNAs. The amount of RBPs is high compared with the number mRNAs. Consequently the current hypothesis proposes the binding of multiple RBPs to each 3' UTR, which would compete or cooperate dynamically with other RBPs. The mix of RBPs would determine the stability of the mRNA and could also modulate the translation process [132, 135]. The expression of protein-coding genes can also be regulated at the translational level. In ribosome profiling studies it has been shown that there is a wide range in the density of ribosomes associated to mRNAs, with differences between life stages. In addition, trypanosome mRNAs can contain upstream open reading frames in their 5' UTRs, which decrease the translation of the main ORF [132, 136, 137].

Multi-Copy Families of Surface Proteins

Genome reduction is frequent in parasites with functions that are essential for a free-living organism becoming obsolete inside a host. Surprisingly, compared with other single cell parasitic eukaryotes, trypanosomatid genomes do not appear to be specially reduced in size or function. On the contrary, in the evolution of parasitism in trypanosomatids the gain of new competences seems to have been more important than the loss of functions [134]. One example of this gain of functions is the presence of large multi-copy families that encode surface proteins. These families are specific to trypanosomatids and usually have a non-random distribution in the genome. A number of them have been implicated in pathogenesis and defence against the host immune system, such as the Major Surface Protease (MSP) family of metalloproteases involved in pathogenesis and conserved in all trypanosomatids. Other well-known examples are the Variant Surface Glycoprotein (VSG) and procyclin in *T. brucei*, delta-amastin and Promastigote Surface Antigen (PSA) in *Leishmania* and trans-sialidases in *T. cruzi* [134, 138, 139].

Epigenetic regulation

In eukaryotes, nuclear DNA is organized into a complex of DNA and proteins known as chromatin. The nucleosome is the basic unit of the chromatin, providing a sevenfold condensation. It comprises an octamer made of 2 copies of each of the core histones (H2A, H2B, H3 and H4) around which approximately 147 bp of DNA are wrapped. In addition, there is a histone (H1) in the DNA region between two nucleosomes that helps stabilize the chromatin. The chromatin is folded into a 30nm chromatin fiber that can be further compacted, up to the level of the distinct chromosomes that can be visualized during the eukaryotic mitosis [126, 129]. Although the nucleosomes are still the basic unit of chromatin in trypanosomatids, their histones are divergent from those found in yeast and vertebrates. DNA in trypanosomatids is not condensed into the 30nm chromatin fiber nor do chromosomes condense during mitosis. However, some differences in the level of condensation between life-cycle stages have been described [126, 130].

Mechanisms that influence the structure of chromatin have been implicated in the regulation of gene expression. In trypanosomatids, as in other eukaryotes, specific modifications of the N-terminal tails of histones, or the presence of histone variants correlate with regions of active or repressed transcription. As of yet, no conserved sequences have been identified in the transcription start sites (TSSs) of the PTUs. It has been proposed that TSSs could be determined by chromatin structure rather than the presence of conserved sequence motifs. Some of the histone modifications described in trypanosomatids are common in eukaryotes, but there are also some modifications and histone variations specific to trypanosomatids, such as H3V and H4V (probable markers of transcription termination sites) [129, 140].

Mitochondrial Genome: Architecture and RNA Editing

The kDNA is made up of circles of DNA that are interlocked in a chain-mail like network and are of two types: maxicircles and minicircles. Maxicircles store information for classical mitochondrial genes and proteins, but their transcripts require RNA editing, the insertion or deletion of uridines, before being translated. Minicircles encode guide RNAs (gRNA), that act as templates during the editing process. Unlike other eukaryotes, mitochondrial tRNAs are found in the nuclear genome and require specific target sequences to be transported into the mitochondria [123, 128]. The mitochondrial genome contains a few dozens of maxicircles, with identical sequence and a size of 20-40 kb, and thousands of minicircles. Minicircles differ in sequence content but their size is species specific and uniform, usually between 0.5-10 kb. Maxicircles are concatenated together and simultaneously interlinked with the minicircle network. The DNA network and associated proteins are organized in a dense disc visible by light microscopy. While all kinetoplastids contain maxi and minicircles, the concatenated network is unique to trypanosomatids [124, 141]. During the RNA editing uridines are inserted or deleted from mitochondrial mRNAs fixing errors in the sequence and restoring a viable coding sequence. The

sequences to be used as templates are stored in the gRNAs (50-60nt). These encode only a small portion of the information needed to repair a mRNA, therefore multiple gRNA are required to edit each mRNA. RNA editing is catalysed by the RNA editing core complex or editosome. Several modules can combine to build different versions of the editosome, each with different specificities [127, 142].

Base J

Base J, or the modification of thymine to beta-D-glucosyl_hydroxymethyluracil, is enriched at the ends of PTUs, at potential transcription terminal sites (TTSs) and in repetitive DNA elements, such as the telomeric repeats [123, 129].

Transposable Elements

The two main classes of transposable element are DNA transposons and RNA retrotransposons. DNA transposons move by "cut and paste" and depend on a DNA intermediate, while RNA retrotransposons use a "copy and paste" strategy, with a RNA intermediate. DNA transposons have not been found in trypanosomatid genomes, butRNA retrotransposons have been shown to be present. For example, several classes of potentially active retrotransposons have been identified in *T. brucei* and *T. cruzi*, some of which could be involved in the regulation of gene expression, such as SIDER2, which localizes to the 3'UTRs of mRNAs and affects its stability [126, 143, 144].

Sequenced Genomes

The first trypanosomatids sequenced, were *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major,* the causative agents of Sleeping Sickness, Chagas disease and Leishmaniasis in humans [145–147]. Since then, the genomes of other medically relevant trypanosomes have been published. *Leishmania* species that have been sequenced include *L. donovani* [148], *L. infantum*, *L. brasiliensis* [149], *L. mexicana* [138], *L. panamensis* [150], *L. peruviana* [151], *L. amazonensis* [152]. *Trypanosoma* species include *T. rangeli* [153]. Apart from the reference genomes, multiple strains and hundreds of isolates have been sequenced and are available in the databases [TriTrypDB, NCBI]. The range of published genomes has expanded to other dixenous species and includes parasites of reptiles (*Trypanosoma* grayi [154] and *Leishmania tarentolae* [155]), parasites of livestock (*T. evansi* [156]) or parasites of plants (*Phytomonas serpens*, *Phytomonas spp*. [157, 158]). In addition, the genomes of a few monoxenous trypanosomatids have been published (*Leptomonas seymouri* [159] and *Lotmaria passim* [160]). Some of theses species harbour symbiotic bacteria and have been used as a model to study the evolution of organelles (*Crithidia acanthocephali*, *Herpetomonas muscarum*, *Strigomonas oncopelti*, *Strigomonas galati* and *Strigomonas culicis*, *Angomonas desouzai* and *Angomonas deanei)* [161]. Additional genomes are available pre-publication in the genome databases (TriTrypDB, NCBI) include *Endotrypanum monterogeii*, *Leptomonas pyrrhocoris*, *Crithidia fasciculata;* the *Leishmanias L aethiopica*, *L. tropica*, *L. gerbilli*, *Leishmania enriettii*, *L. turanica,* and the *Trypanosomas T. congolense* and *T. vivax*.

Toxoplasma and related organisms

Toxoplasma gondii is a member of the tissue cyst-forming coccidian parasites, which include *Neospora caninum*, *Hammondia hammondia* and *Sarcocystis spp*. among others [162–165]. Of these *T. gondii* appears to be the most widely distributed both geographically and by host diversity, and is able to infect virtually any warm-blooded animal. While the diversity of *T. gondii* is restricted to three clonal lineages in Europe and North America, isolates from the southern hemisphere exhibit much wider genetic variability [166]. Amazingly, while *T. gondii* can infect a wide variety of warm-blooded organisms it can only undergo sexual recombination in Felidae. Cats shed infective sporozoites containing environmentally resistant cysts, which can be transmitted orally to other organisms such as rodents [165]. Following oral infection, sporozoites cross the small intestine and can infect a

variety of cells where they undergo a developmental switch to fast growing tachyzoites [166]. Tachyzoites replicate through a process called endodyogeny where two daughter cells are formed within a mother cell by a combination of *de novo* building of cytoskeletal and secretory components, replication and segregation of mother cell components (i.e. nucleus, mitochondria and apicoplast) and recycling of mother cell components [167, 168]. Pressure from the host immune system forces tachyzoites to undergo another developmental change into bradyzoites [169]. These semi-quiescent cells form clusters called tissue cysts that settle in brain and/or muscle tissue where they may remain for the life of the host, althoughreactivation of bradyzoites can occur in immunocompromised individuals. Bradyzoites also serve as a reservoir of transmission if an infected host is eaten by another animal. Interestingly, the tissue cyst tropism varies markedly between hosts. The fast replicating tachyzoite stage is often asymptomatic, but can cause acute morbidity or mortality in immunocompromised individuals. Placental transmission is known to cause foetal mortality or serious congenital defects.

T. gondii contains a ~65 Mb nuclear genome comprising 14 chromosomes [170–172], a 35 Kb apicoplast genome [173] and a mitochondrial genome. *T. gondii* genomic scale data such as expressed sequenced tags, sequenced BAC clones and whole genome shotgun sequencing were first made available through ToxoDB beginning in 2001 [174]. Since then, additional genomic scale data have been generated including genome sequence and transcriptomic data from a large scale population sequencing project [172]. The genome of the closely related *H. hammondia* and *N. caninum* are ~65 Mb and ~62 Mb in size, respectively, and not surprisingly also comprise 14 chromosomes each (Table 1.9) [171, 172, 175]. The genome of the more divergent *S. neurona* is almost twice the size of those previously described at ~130Mb, while a GC content of roughly 53% is common across this group (Table 1.9) [176]. A high degree of genomic synteny is observed between *T. gondii, H. hammondia* and *N. caninum.* This level of synteny is not maintained with between this group and *S. neurona*. [171, 172, 176].

Apicomplexan parasites in general have evolved secretory systems that transport effector molecules into their host cells. These have a range of functions, including modification the intracellular environment, promotion of immune evasion and modulation of host-cell transcription [177]. Most information about secretory effectors in coccidian parasites comes from *T. gondii* where numerous studies have defined dense granule [178], rhoptry [179], microneme [180] and SAG1 related sequences (SRS) proteins [181]. Comparative genomic analysis revealed that one of the primary features differentiating both different species of coccidian parasite and different strains of *T. gondii* is sequence diversity and copy number variation (CNV) at secretory effector loci. [171, 172, 175]. A comparison of 62 isolates of *T. gondii* and one isolate of *H. hammondia* showed that secretory effectors are often found in genomic regions exhibiting tandem amplification [172]. A comparison of reference isolates from the 16 major *Toxoplasma* haplogroups showed that all possess a repertoire of secretory effectors with most diversity occurring in rhoptry and SRS genes. Further comparison of secretory effectors between *T. gondii, H. hammondia* and *N. caninum* revealed additional diversity and a *T. gondii* specific family (*Tg*FAMs) of effectors, which may be important for host range and definitive host preferences [172]. Interestingly, a number of the *Tg*FAMs are clustered in telomeric regions and contain a variable region, which may implicate them in immune evasion [172, 182] but they also may play a role in during sexual development since many are expressed in the cat and in oocysts [183].

Table 1.9. Basic genome statistics for *T. gondii* and related organisms.

	*Toxoplasma gondii**	*Hammondia hammondi* H.H.34	*Neospora caninum* Liverpool	*Sarcocystis neurona***
Genome size (Mb)	63	65	62	128
No. of chromosomes	14	14	14	ND
No. of genes	8707	8176	7266	7140

Table 1.9 continued from previous page.

	*Toxoplasma gondii**	*Hammondia hammondi* H.H.34	*Neospora caninum* Liverpool	*Sarcocystis neurona***
% of genes with introns	76	76	77	81

* Average statistics from three strains: ME49, VEG and GT1
** Average statistics from two strains: SN1 and SN3
ND = not determined

Data integration and accessibility

Several databases exist that provide online and free access to parasite genomes, annotation and functional data (Table 2.0). The National Institutes of Allergy and Infectious Diseases (NIAID) in the United States initiated established bioinformatics resource centers (BRCs) in 2004 whose goal is to provide the global pathogen research community with free and online tools to mine genomic and functional genomic data, and additional data-types essential for pathogen surveillance and control [184]. The BRCs included one specifically tasked with providing support for the eukaryotic pathogen scientific community (EuPathDB, initially known as ApiDB) [185]. Now in its third five-year funding cycle, EuPathDB incorporates data from over 240 parasitic and evolutionarily related organisms spanning multiple phyla such as the Amoebozoa, Apicomplexa, Euglenozoa, Metamonada, Sarcomastigophora and numerous fungal phyla. Data includes genome sequence, structural and functional annotation, functional data covering the omics landscape including transcriptomic, proteomic and metabolomics. Most current database content can be accessed here: http://eupathdb.org/eupathdb/eupathGenome.jsp

Data within EuPathDB and its component sites are searchable via an intuitive graphical user interface that allows the development of complex *in silico* experiments to support hypothesis driven experiments. Data types include the underlying genomic sequences and annotations (close to 250 genomes represented), transcript level data (SAGE-tag, EST, microarray and RNA sequence data), protein expression data (including quantitative), epigenomic data (ChIP-chip and ChIP-seq), population-level (SNP) and isolate data, and host response data (antibody array). In addition, genomic analyses provide the ability to search for gene features, subcellular localization, motifs (InterPro and user defined), function (Enzyme commission annotation and GO terms) and evolutionary relationships based on gene orthology. Detailed tutorials and usage instructions are available through publications and online tutorials and exercises [121, 186]. A number of YouTube tutorials are available: https://www.youtube.com/user/EuPathDB/. EuPathDB resources provide Community annotation and curation via user comments (including images, files, PubMed records, etc) can be added to records in EuPathDB sites (Comments become immediately visible and searchable). A Graphical search system allows building complex searches in a step-wise manner that can be saved, modified and shared. An example strategy can be seem in figure 1 and accessed online by following this link: http://plasmodb.org/plasmo/im.do?s=df42a71ae3acbb1e. Browsing capability through a genome browser integrating genomes, annotation, analyses and functional data. Column and results analysis tools are also available to generate word cloud graphics, histograms, and GO term and pathways enrichment analyses.

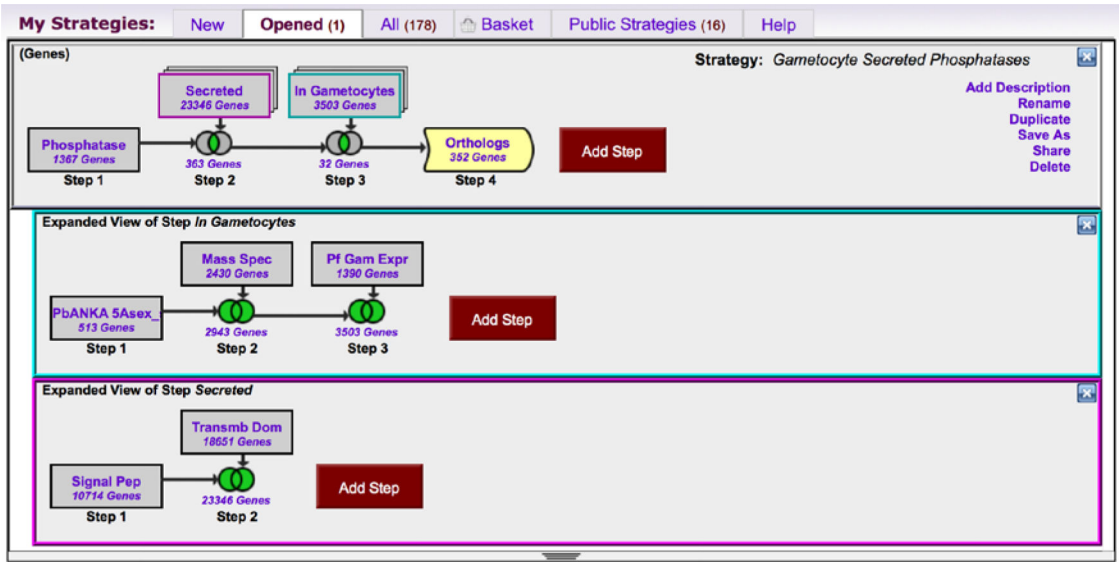

Fig. 1. Screen shot from PlasmoDB depicting a search strategy that identifies putative phosphatases that are secreted and expressed in gametocytes based on proteomics, RNA sequence and microarray experiments. Search strategies are constructed by adding steps that query underlying data. Step 1, identifies all putative phosphatases based on a text search. Step 2, identifies any of the genes in step 1 that also have a secretory signal peptide, at least one transmembrane domain or both (see expanded view of "Secreted"). Step 3 identifies any genes in step 2 that have evidence of expression based on data from three experiments in *P. falciparum* (See expanded view "Gametocytes") [120]Florens:2002bf, Silvestrini:2010io}. Step form transforms the results in step 3 to all orthologs in PlasmoDB.

Table 2.0. Online resources for genomic scale data.

Resource Name	Acronym	Content and functionality	Web address (URL)
National Center for Biotechnology Information	NCBI	Data repository and search capability (International Nucleotide Sequence Database Collaboration)	http://www.ncbi.nlm.nih.gov
The European Bioinformatics Institute	EMBL-EBI	Data repository and search capability (International Nucleotide Sequence Database Collaboration)	http://www.ebi.ac.uk
DNA Data Bank of Japan	DDBJ	Data repository and search capability (International Nucleotide Sequence Database Collaboration)	http://www.ddbj.nig.ac.jp/
Ensembl Protists	EnsemblProtists	Part of the larger Ensembl genomes which is a joint European Bioinformatics Institute and the Wellcome Trust Sanger Institute project providing Ensembl tools, data visualization, data minning and comparative analysis	http://protists.ensembl.org/
GeneDB	GeneDB	Core part of the Sanger Institute's Pathogen Genomics initiative. Provides early access to the latest sequence data and annotation/curation. In addition, the site includes some basic search functionality and genome browsing.	http://www.genedb.org/
The Eukaryotic Pathogen Databases	EuPathDB	One of four National Institutes of Allergy and Infectious Diseases Bioinformatic Centers. Provides integrated search capabilities of genomes and functional data dedicated to eukaryotic pathogens (and related organisms). Includes AmoebaDB, FungiDB, GiardiaDB, MicrosporidiaDB, PiroplasmaDB, PlasmoDB, ToxoDB, TrichDB, TriTrypDB, OrthoMCL and HostDB.	http://EuPathDB.org http://amoebadb.org http://cryptodb.org http://fungidb.org http://microsporidiadb.org

Table 2.0 continued from previous page.

Resource Name	Acronym	Content and functionality	Web address (URL)
			http://piroplasmadb.org http://plasmodb.org http://toxodb.org http://trichdb.org http://tritrypdb.org http://orthomcl.org http://hostdb.org

References

1. Dahm R (2005) Friedrich Miescher and the discovery of DNA. Dev Biol 278:274–288. doi: 10.1016/j.ydbio.2004.11.028 PubMed PMID: 15680349.
2. Avery OT, MacLeod CM, McCarty M (1944) STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. J Exp Med 79:137–158. PubMed PMID: 19871359.
3. Watson JD, Crick F (1953) Molecular structure of nucleic acids. Nature
4. Nirenberg M, Leder P, Bernfield M, Brimacombe R, Trupin J, Rottman F, O'Neal C (1965) RNA codewords and protein synthesis, VII. On the general nature of the RNA code. Proc Natl Acad Sci USA 53:1161–1168. PubMed PMID: 5330357.
5. Gilbert W, Maxam A (1973) The nucleotide sequence of the lac operator. Proc Natl Acad Sci USA 70:3581–3584. PubMed PMID: 4587255.
6. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467. PubMed PMID: 271968.
7. WU R (1972) Nucleotide sequence analysis of DNA. Nature
8. Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, Cocuzza AJ, Jensen MA, Baumeister K (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science 238:336–341. PubMed PMID: 2443975.
9. Fleischmann R, Adams M, White O, Clayton R, Kirkness E, Kerlavage A, Bult C, Tomb J, Dougherty B, Merrick J, al E (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269:496–512. doi: 10.1126/science.7542800 PubMed PMID: 7542800.
10. Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, Volckaert G, Ysebaert M (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature 260:500–507. doi: 10.1038/260500a0 PubMed PMID: 1264203.
11. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes. Science 274:546–563-7. PubMed PMID: 8849441.
12. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan M-S, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DMA, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B (2002)

Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419:498–511. doi: 10.1038/nature01097 PubMed PMID: 12368864.
13. The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant : Arabidopsis thaliana : Article : Nature. Nature 408:796–815. doi: 10.1038/35048692 PubMed PMID: 11130711.
14. The C nbsp elegans Sequencing Consortium. (1998) Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. Science 282:2012–2018. doi: 10.1126/science.282.5396.2012 PubMed PMID: 9851916.
15. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC (2001) Initial sequencing and analysis of the human genome. Nature
16. Moore GE (1998) Cramming more components onto integrated circuits.
17. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380. doi: 10.1038/nature03959 PubMed PMID: 16056220.
18. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309:1728–1732. doi: 10.1126/science.1117389 PubMed PMID: 16081699.
19. Kodama Y, Shumway M, Leinonen R, International Nucleotide Sequence Database Collaboration. (2012) The Sequence Read Archive: explosive growth of sequencing data. Nucleic Acids Res 40:D54–6. doi: 10.1093/nar/gkr854 PubMed PMID: 22009675.
20. Reddy TBK, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos EA, Kyrpides NC (2015) The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. Nucleic Acids Res 43:D1099–D1106. doi: 10.1093/nar/gku950 PubMed PMID: 25348402.
21. Logan-Klumpler FJ, De Silva N, Boehme U, Rogers MB, Velarde G, McQuillan JA, Carver T, Aslett M, Olsen C, Subramanian S, Phan I, Farris C, Mitra S, Ramasamy G, Wang H, Tivey A, Jackson A, Houston R, Parkhill J, Holden M, Harb OS, Brunk BP, Myler PJ, Roos D, Carrington M, Smith DF, Hertz-Fowler C, Berriman M (2012) GeneDB--an annotation database for pathogens. Nucleic Acids Res 40:D98–D108. doi: 10.1093/nar/gkr1032 PubMed PMID: 22116062.
22. Kulikova T, Akhtar R, Aldebert P, Althorpe N, Andersson M, Baldwin A, Bates K, Bhattacharyya S, Bower L, Browne P, Castro M, Cochrane G, Duggan K, Eberhardt R, Faruque N, Hoad G, Kanz C, Lee C, Leinonen R, Lin Q, Lombard V, Lopez R, Lorenc D, McWilliam H, Mukherjee G, Nardone F, Pastor MPG, Plaister S, Sobhany S, Stoehr P, Vaughan R, Wu D, Zhu W, Apweiler R EMBL Nucleotide Sequence Database in 2006. Available at: nar.oxfordjournals.org
23. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW GenBank. Available at: nar.oxfordjournals.org
24. Tateno Y, Fukami-Kobayashi K, Miyazaki S, Sugawara H, Gojobori T (1998) DNA Data Bank of Japan at work on genome sequence data. Nucleic Acids Res 26:16–20. PubMed PMID: 9399792.
25. Clarke M, Lohan AJ, Liu B, Lagkouvardos I, Roy S, Zafar N, Bertelli C, Schilde C, Kianianmomeni A, Bürglin TR, Frech C, Turcotte B, Kopec KO, Synnott JM, Choo C, Paponov I, Finkler A, Tan CSH, Hutchins AP, Weinmeier T, Rattei T, Chu JS, Gimenez G, Irimia M, Rigden DJ, Fitzpatrick DA, Lorenzo-Morales J, Bateman A, Chiu C-H, Tang P, Hegemann P, Fromm H, Raoult D, Greub G, Miranda-Saavedra D, Chen N, Nash P, Ginger ML, Horn M, Schaap P, Caler L, Loftus BJ (2013) Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. Genome Biol 14:R11. doi: 10.1186/gb-2013-14-2-r11 PubMed PMID: 23375108.

26. Eichinger L, Pachebat JA, Glöckner G, Rajandream M-A, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P, Fey P, Pilcher K, Chen G, Saunders D, Sodergren E, Davis P, Kerhornou A, Nie X, Hall N, Anjard C, Hemphill L, Bason N, Farbrother P, Desany B, Just E, Morio T, Rost R, Churcher C, Cooper J, Haydock S, van Driessche N, Cronin A, Goodhead I, Muzny D, Mourier T, Pain A, Lu M, Harper D, Lindsay R, Hauser H, James K, Quiles M, Babu MM, Saito T, Buchrieser C, Wardroper A, Felder M, Thangavelu M, Johnson D, Knights A, Loulseged H, Mungall K, Oliver K, Price C, Quail MA, Urushihara H, Hernandez J, Rabbinowitsch E, Steffen D, Sanders M, Ma J, Kohara Y, Sharp S, Simmonds M, Spiegler S, Tivey A, Sugano S, White B, Walker D, Woodward J, Winckler T, Tanaka Y, Shaulsky G, Schleicher M, Weinstock G, Rosenthal A, Cox EC, Chisholm RL, Gibbs R, Loomis WF, Platzer M, Kay RR, Williams J, Dear PH, Noegel AA, Barrell B, Kuspa A (2005) The genome of the social amoeba Dictyostelium discoideum. Nature 435:43–57. doi: 10.1038/nature03481 PubMed PMID: 15875012.
27. Lorenzi HA, Puiu D, Miller JR, Brinkac LM, Amedeo P, Hall N, Caler EV (2010) New assembly, reannotation and analysis of the Entamoeba histolytica genome reveal new genomic features and protein content information. PLoS Negl Trop Dis 4:e716. doi: 10.1371/journal.pntd.0000716 PubMed PMID: 20559563.
28. Loftus BJ, Hall N (2005) Entamoeba: still more to be learned from the genome. Trends in Parasitology 21:453. doi: 10.1016/j.pt.2005.08.007 PubMed PMID: 16099723.
29. Clark CG, Alsmark U, Tazreiter M (2007) Structure and content of the Entamoeba histolytica genome. Advances in …. doi: 10.1016/S0065-308X(07)65002-7
30. Clark CG, Ali IKM, Zaki M, Loftus BJ, Hall N (2006) Unique organisation of tRNA genes in Entamoeba histolytica. Mol Biochem Parasitol 146:24–29. doi: 10.1016/j.molbiopara.2005.10.013 PubMed PMID: 16307803.
31. BHATTACHARYA S, BHATTACHARYA A, DIAMOND LS, SOLDO AT (1989) Circular DNA of Entamoeba histolytica Encodes Ribosomal RNA. J Protozool 36:455–458. doi: 10.1111/j.1550-7408.1989.tb01080.x PubMed PMID: 2553935.
32. Rosenthal B, Mai Z, Caplivski D, Ghosh S, la Vega de H, Graf T, Samuelson J (1997) Evidence for the bacterial origin of genes encoding fermentation enzymes of the amitochondriate protozoan parasite Entamoeba histolytica. J Bacteriol 179:3736–3745. PubMed PMID: 9171424.
33. Weedall GD, Clark CG, Koldkjaer P, Kay S, Bruchhaus I, Tannich E, Paterson S, Hall N (2012) Genomic diversity of the human intestinal parasite Entamoeba histolytica. Genome Biol 13:R38. doi: 10.1186/gb-2012-13-5-r38 PubMed PMID: 22630046.
34. Willhoeft U, Tannich E (1999) The electrophoretic karyotype of Entamoeba histolytica. Mol Biochem Parasitol 99:41–53. doi: 10.1016/S0166-6851(98)00178-9 PubMed PMID: 10215023.
35. Gilchrist CA, Ali IKM, Kabir M, Alam F, Scherbakova S, Ferlanti E, Weedall GD, Hall N, Haque R, Petri WA, Caler E (2012) A Multilocus Sequence Typing System (MLST) reveals a high level of diversity and a genetic component to Entamoeba histolytica virulence. BMC microbiology 12:1. doi: 10.1186/1471-2180-12-151 PubMed PMID: 22221383.
36. Zaki M, Reddy SG, Jackson TFHG, Ravdin JI, Clark CG (2003) Genotyping of Entamoeba species in South Africa: diversity, stability, and transmission patterns within families. J Infect Dis 187:1860–1869. doi: 10.1086/375349 PubMed PMID: 12792862.
37. Stensvold CR, Lebbad M, Victory EL, Verweij JJ, Tannich E, Alfellani M, Legarraga P, Clark CG (2011) Increased Sampling Reveals Novel Lineages of Entamoeba: Consequences of Genetic Diversity and Host Specificity for Taxonomy and Molecular Detection. Protist 162:525–541. doi: 10.1016/j.protis.2010.11.002 PubMed PMID: 21295520.
38. Tachibana H, Yanagi T, Pandey K, Cheng X-J, Kobayashi S, Sherchand JB, Kanbara H (2007) An Entamoeba sp. strain isolated from rhesus monkey is virulent but genetically different from Entamoeba histolytica☆. Mol Biochem Parasitol 153:107–114. doi: 10.1016/j.molbiopara.2007.02.006 PubMed PMID: 17403547.

39. Tachibana H, Yanagi T, Lama C, Pandey K, Feng M, Kobayashi S, Sherchand JB (2013) Prevalence of Entamoeba nuttalli infection in wild rhesus macaques in Nepal and characterization of the parasite isolates. Parasitol Int 62:230–235. doi: 10.1016/j.parint.2013.01.004 PubMed PMID: 23370534.
40. Feng M, Cai J, Min X, Fu Y, Xu Q, Tachibana H, Cheng X (2013) Prevalence and genetic diversity of Entamoeba species infecting macaques in southwest China. Parasitol Res 112:1529–1536. doi: 10.1007/s00436-013-3299-1 PubMed PMID: 23354942.
41. Diamond LS, Clark CG (1993) A redescription of Entamoeba histolytica Schaudinn, 1903 (Emended Walker, 1911) separating it from Entamoeba dispar Brumpt, 1925. J Eukaryot Microbiol 40:340–344. PubMed PMID: 8508172.
42. Ali IKM, Hossain MB, Roy S, Ayeh-Kumi PF, Petri WA Jr, Haque R, Clark CG (2003) Entamoeba moshkovskiiInfections in Children in Bangladesh. Emerging Infect Dis 9:580–584. doi: 10.3201/eid0905.020548 PubMed PMID: 12737742.
43. Shimokawa C, Kabir M, Taniuchi M, Mondal D, Kobayashi S, Ali IKM, Sobuz SU, Senba M, Houpt E, Haque R, Petri WA, Hamano S (2012) Entamoeba moshkovskii Is Associated With Diarrhea in Infants and Causes Diarrhea and Colitis in Mice. J Infect Dis 206:744–751. doi: 10.1093/infdis/jis414 PubMed PMID: 22723640.
44. Ehrenkaufer GM, Weedall GD, Williams D, Lorenzi HA, Caler E, Hall N, Singh U (2013) The genome and transcriptome of the enteric parasite Entamoeba invadens, a model for encystation. Genome Biol 14:R77. doi: 10.1186/gb-2013-14-7-r77 PubMed PMID: 23889909.
45. Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Caler EV, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Iodice J, Kissinger JC, Kraemer ET, Li W, Nayak V, Pennington C, Pinney DF, Pitts B, Roos DS, Srinivasamoorthy G, Stoeckert CJ, Treatman C, Wang H (2011) AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species. Nucleic Acids Res 39:D612–9. doi: 10.1093/nar/gkq1006 PubMed PMID: 20974635.
46. Lorenzo-Morales J, Martín-Navarro CM, López-Arencibia A, Arnalich-Montiel F, Piñero JE, Valladares B (2013) Acanthamoeba keratitis: an emerging disease gathering importance worldwide? Trends in Parasitology 29:181–187. doi: 10.1016/j.pt.2013.01.006 PubMed PMID: 23433689.
47. Ankarklev J, Jerlström-Hultqvist J, Ringqvist E, Troell K, Svard SG (2010) Behind the smile: cell biology and disease mechanisms of Giardia species. Nature Reviews Microbiology 8:413–422. doi: 10.1038/nrmicro2317 PubMed PMID: 20400969.
48. Cacciò SM, Ryan U (2008) Molecular epidemiology of giardiasis. Mol Biochem Parasitol 160:75–80. doi: 10.1016/j.molbiopara.2008.04.006 PubMed PMID: 18501440.
49. Bernander R, Palm JED, Svard SG (2001) Genome ploidy in different stages of the Giardia lamblia life cycle. Cell Microbiol 3:55–62. doi: 10.1046/j.1462-5822.2001.00094.x PubMed PMID: 11207620.
50. Adam RD (2001) Biology of Giardia lamblia. Clin Microbiol Rev 14:447–475. doi: 10.1128/CMR.14.3.447-475.2001 PubMed PMID: 11432808.
51. Adam RD, Nash TE, Wellems TE (1988) The Giardia lamblia trophozoite contains sets of closely related chromosomes. Nucleic Acids Res 16:4555–4567. PubMed PMID: 2837738.
52. Tůmová P, Hofštetrová K, Nohýnková E, Hovorka O, Král J (2006) Cytogenetic evidence for diversity of two nuclei within a single diplomonad cell of Giardia. Chromosoma 116:65–78. doi: 10.1007/s00412-006-0082-4 PubMed PMID: 17086421.
53. Adam RD, Dahlstrom EW, Martens CA, Bruno DP, Barbian KD, Ricklefs SM, Hernandez MM, Narla NP, Patel RB, Porcella SF, Nash TE (2013) Genome Sequencing of Giardia lamblia Genotypes A2 and B Isolates (DH and GS) and Comparative Analysis with the Genomes of Genotypes A1 and E (WB and Pig). Genome Biol Evol 5:2498–2511. doi: 10.1093/gbe/evt197 PubMed PMID: 24307482.
54. Jerlström-Hultqvist J, Franzén O, Ankarklev J, Xu F, Nohýnková E, Andersson JO, Svard SG, Andersson B (2010) Genome analysis and comparative genomics of a Giardia intestinalis assemblage E isolate. BMC Genomics 11:543. doi: 10.1186/1471-2164-11-543 PubMed PMID: 20929575.
55. Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, Davids BJ, Dawson SC, Elmendorf HG, Hehl AB, Holder ME, Huse SM, Kim UU, Lasek-

Nesselquist E, Manning G, Nigam A, Nixon JEJ, Palm D, Passamaneck NE, Prabhu A, Reich CI, Reiner DS, Samuelson J, Svard SG, Sogin ML (2007) Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. Science 317:1921–1926. doi: 10.1126/science.1143837 PubMed PMID: 17901334.

56. Ankarklev J, Franzén O, Peirasmaki D, Jerlström-Hultqvist J, Lebbad M, Andersson J, Andersson B, Svard SG (2015) Comparative genomic analyses of freshly isolated Giardia intestinalis assemblage A isolates. BMC Genomics 16:413. doi: 10.1186/s12864-015-1893-6 PubMed PMID: 26017011.
57. Birkeland SR, Preheim SP, Davids BJ, Cipriano MJ, Palm D, Reiner DS, Svard SG, Gillin FD, McArthur AG (2010) Transcriptome analyses of the Giardia lamblia life cycle. Mol Biochem Parasitol 174:62–65. doi: 10.1016/j.molbiopara.2010.05.010 PubMed PMID: 20570699.
58. Perry DA, Morrison HG, Adam RD (2011) Optical map of the genotype A1 WB C6 Giardia lamblia genome isolate. Mol Biochem Parasitol 180:112–114. doi: 10.1016/j.molbiopara.2011.07.008 PubMed PMID: 21835210.
59. Franzén O, Jerlström-Hultqvist J, Castro E, Sherwood E, Ankarklev J, Reiner DS, Palm D, Andersson JO, Andersson B, Svard SG (2009) Draft Genome Sequencing of Giardia intestinalis Assemblage B Isolate GS: Is Human Giardiasis Caused by Two Different Species? PLoS Pathog 5:e1000560. doi: 10.1371/journal.ppat.1000560 PubMed PMID: 19696920.
60. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omore R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acácio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM (2013) Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet 382:209–222. doi: 10.1016/S0140-6736(13)60844-2 PubMed PMID: 23680352.
61. Arrowood MJ (2002) In Vitro Cultivation of Cryptosporidium Species. Clin Microbiol Rev 15:390–400. doi: 10.1128/CMR.15.3.390-400.2002 PubMed PMID: 12097247.
62. Karanis P, Aldeyarbi HM (2011) Evolution of Cryptosporidium in vitro culture. Int J Parasitol 41:1231–1242. doi: 10.1016/j.ijpara.2011.08.001 PubMed PMID: 21889507.
63. Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE (2011) Landscape of next-generation sequencing technologies. Anal Chem 83:4327–4341. doi: 10.1021/ac2010857 PubMed PMID: 21612267.
64. Nagarajan N, Pop M (2013) Sequence assembly demystified. Nat Rev Genet 14:157–167. doi: 10.1038/nrg3367 PubMed PMID: 23358380.
65. Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet 12:671–682. doi: 10.1038/nrg3068 PubMed PMID: 21897427.
66. Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. Nature 13:329–342. doi: 10.1038/nrg3174
67. Heiges M, Wang H, Robinson E, Aurrecoechea C, Gao X, Kaluskar N, Rhodes P, Wang S, He C-Z, Su Y, Miller J, Kraemer E, Kissinger J (2006) CryptoDB: a Cryptosporidium bioinformatics resource update. Nucleic Acids Res 34:D419–22. doi: 10.1093/nar/gkj078 PubMed PMID: 16381902.
68. Manque PA, Tenjo F, Woehlbier U, Lara AM, Serrano MG, Xu P, Alves JM, Smeltz RB, Conrad DH, Buck GA (2011) Identification and immunological characterization of three potential vaccinogens against Cryptosporidium species. Clin Vaccine Immunol 18:1796–1802. doi: 10.1128/CVI.05197-11 PubMed PMID: 21918117.
69. Donati C, Rappuoli R (2013) Reverse vaccinology in the 21st century: improvements over the original design. Ann N Y Acad Sci 1285:115–132. doi: 10.1111/nyas.12046 PubMed PMID: 23527566.
70. Kelly DF, Rappuoli R (2005) Reverse vaccinology and vaccines for serogroup B Neisseria meningitidis. Adv Exp Med Biol 568:217–223. doi: 10.1007/0-387-25342-4_15 PubMed PMID: 16107075.
71. Xiao L, Herd RP (1994) Epidemiology of equine Cryptosporidium and Giardia infections. Equine Veterinary Journal. doi: 10.1111/j.2042-3306.1994.tb04323.x/pdf

72. Bouzid M, Hunter PR, Chalmers RM, Tyler KM Cryptosporidium Pathogenicity and Virulence. Available at: cmr.asm.org
73. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S, Buck GA, Xu P, Bankier AT, Dear PH, Konfortov BA, Spriggs HF, Iyer L, Anantharaman V, Aravind L, Kapur V (2004) Complete genome sequence of the apicomplexan, Cryptosporidium parvum. Science 304:441–445. doi: 10.1126/science.1094786 PubMed PMID: 15044751.
74. Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, Puiu D, Manque P, Akiyoshi D, Mackey AJ, Pearson WR, Dear PH, Bankier AT, Peterson DL, Abrahamsen MS, Kapur V, Tzipori S, Buck GA (2004) The genome of Cryptosporidium hominis. Nature 431:1107–1112. doi: 10.1038/nature02977 PubMed PMID: 15510150.
75. Widmer G, Lee Y, Hunt P, Martinelli A, Tolkoff M, Bodi K (2012) Comparative genome analysis of two Cryptosporidium parvum isolates with different host range. Infect Genet Evol 12:1213–1221. doi: 10.1016/j.meegid.2012.03.027 PubMed PMID: 22522000.
76. Isaza JP, Galván AL, Polanco V, Huang B, Matveyev AV, Serrano MG, Manque P, Buck GA, Alzate JF (2015) Revisiting the reference genomes of human pathogenic Cryptosporidium species: reannotation of C. parvum Iowa and a new C. hominis reference. Sci Rep 5:16324. doi: 10.1038/srep16324 PubMed PMID: 26549794.
77. Guo Y, Cebelinski E, Matusevich C, Alderisio KA, Lebbad M, McEvoy J, Roellig DM, Yang C, Feng Y, Xiao L (2015) Subtyping novel zoonotic pathogen Cryptosporidium chipmunk genotype I. J Clin Microbiol 53:1648–1654. doi: 10.1128/JCM.03436-14 PubMed PMID: 25762767.
78. Hadfield SJ, Pachebat JA, Swain MT, Robinson G, Cameron SJ, Alexander J, Hegarty MJ, Elwin K, Chalmers RM (2015) Generation of whole genome sequences of new Cryptosporidium hominis and Cryptosporidium parvum isolates directly from stool samples. BMC Genomics 16:650. doi: 10.1186/s12864-015-1805-9 PubMed PMID: 26318339.
79. Yamagishi J, Wakaguri H, Sugano S, Kawano S, Fujisaki K, Sugimoto C, Watanabe J, Suzuki Y, Kimata I, Xuan X (2011) Construction and analysis of full-length cDNA library of Cryptosporidium parvum. Parasitol Int 60:199–202. doi: 10.1016/j.parint.2011.03.001 PubMed PMID: 21397714.
80. Boguski MS, Lowe T, Tolstoshev CM (1993) dbEST—database for "expressed sequence tags." Nature genetics
81. Mauzy MJ, Enomoto S, Lancto CA, Abrahamsen MS, Ms R (2012) The Cryptosporidium parvum transcriptome during in vitro development. PLoS ONE 7:3.
82. Zhang H, Guo F, Zhou H, Zhu G (2012) Transcriptome analysis reveals unique metabolic features in the Cryptosporidium parvum Oocysts associated with environmental survival and stresses. BMC Genomics 13:647. doi: 10.1186/1471-2164-13-647 PubMed PMID: 23171372.
83. Woo YH, Ansari H, Otto TD, Klinger CM, Kolisko M (2015) Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. eLife
84. Madrid-Aliste CJ, Dybas JM, Angeletti RH, Weiss LM, Kim K, Simon I, Fiser A (2009) EPIC-DB: a proteomics database for studying Apicomplexan organisms. BMC Genomics 10:38. doi: 10.1186/1471-2164-10-38 PubMed PMID: 19159464.
85. Truong Q, Ferrari BC (2006) Quantitative and qualitative comparisons of Cryptosporidium faecal purification procedures for the isolation of oocysts suitable for proteomic analysis. Int J Parasitol 36:811–819. doi: 10.1016/j.ijpara.2006.02.023 PubMed PMID: 16696982.
86. Snelling WJ, Lin Q, Moore JE, Millar BC, Tosini F, Pozio E, Dooley JSG, Lowery CJ (2007) Proteomics analysis and protein expression during sporozoite excystation of Cryptosporidium parvum (Coccidia, Apicomplexa). Mol Cell Proteomics 6:346–355. doi: 10.1074/mcp.M600372-MCP200 PubMed PMID: 17124246.
87. Sanderson SJ, Xia D, Prieto H, Yates J, Heiges M, Kissinger JC, Bromley E, Lal K, Sinden RE, Tomley F, Wastling JM (2008) Determining the protein repertoire of Cryptosporidium parvum sporozoites. Proteomics 8:1398–1414. doi: 10.1002/pmic.200700804 PubMed PMID: 18306179.

88. MANWELL RD (1964) THE GENUS DACTYLOSOMA. J Protozool 11:526–530. PubMed PMID: 14231179.
89. Cornillot E, Hadj-Kaddour K, Dassouli A, Noel B, Ranwez V, Vacherie B, Augagneur Y, Brès V, Duclos A, Randazzo S, Carcy B, Debierre-Grockiego F, Delbecq S, Moubri-Ménage K, Shams-Eldin H, Usmani-Brown S, Bringaud F, Wincker P, Vivarès CP, Schwarz RT, Schetters TP, Krause PJ, Gorenflot A, Berry V, Barbe V, Ben Mamoun C (2012) Sequencing of the smallest Apicomplexan genome from the human pathogen Babesia microti. Nucleic Acids Res 40:9102–9114. doi: 10.1093/nar/gks700 PubMed PMID: 22833609.
90. Gardner MJ, Bishop R, Shah T, de Villiers EP, Carlton JM, Hall N, Ren Q, Paulsen IT, Pain A, Berriman M, Wilson RJM, Sato S, Ralph SA, Mann DJ, Xiong Z, Shallom SJ, Weidman J, Jiang L, Lynn J, Weaver B, Shoaibi A, Domingo AR, Wasawo D, Crabtree J, Wortman JR, Haas B, Angiuoli SV, Creasy TH, Lu C, Suh B, Silva JC, Utterback TR, Feldblyum TV, Pertea M, Allen J, Nierman WC, Taracha ELN, Salzberg SL, White OR, Fitzhugh HA, Morzaria S, Venter JC, Fraser CM, Nene V (2005) Genome sequence of Theileria parva, a bovine pathogen that transforms lymphocytes. Science 309:134–137. doi: 10.1126/science.1110439 PubMed PMID: 15994558.
91. Jackson AP, Otto TD, Darby A, Ramaprasad A, Xia D, Echaide IE, Farber M, Gahlot S, Gamble J, Gupta D, Gupta Y, Jackson L, Malandrin L, Malas TB, Moussa E, Nair M, Reid AJ, Sanders M, Sharma J, Tracey A, Quail MA, Weir W, Wastling JM, Hall N, Willadsen P, Lingelbach K, Shiels B, Tait A, Berriman M, Allred DR, Pain A (2014) The evolutionary dynamics of variant antigen genes in Babesia reveal a history of genomic innovation underlying host-parasite interaction. Nucleic Acids Res 42:7113–7131. doi: 10.1093/nar/gku322 PubMed PMID: 24799432.
92. Brayton KA, Lau AOT, Herndon DR, Hannick L, Kappmeyer LS, Berens SJ, Bidwell SL, Brown WC, Crabtree J, Fadrosh D, Feldblum T, Forberger HA, Haas BJ, Howell JM, Khouri H, Koo H, Mann DJ, Norimine J, Paulsen IT, Radune D, Ren Q, Smith RK, Suarez CE, White O, Wortman JR, Knowles DP, McElwain TF, Nene VM (2007) Genome sequence of Babesia bovis and comparative analysis of apicomplexan hemoprotozoa. PLoS Pathog 3:1401–1413. doi: 10.1371/journal.ppat.0030148 PubMed PMID: 17953480.
93. Tarigo JL, Scholl EH, McK Bird D, Brown CC, Cohn LA, Dean GA, Levy MG, Doolan DL, Trieu A, Nordone SK, Felgner PL, Vigil A, Birkenheuer AJ (2013) A novel candidate vaccine for cytauxzoonosis inferred from comparative apicomplexan genomics. PLoS ONE 8:e71233. doi: 10.1371/journal.pone.0071233 PubMed PMID: 23977000.
94. Pain A, Renauld H, Berriman M, Murphy L, Yeats CA, Weir W, Kerhornou A, Aslett M, Bishop R, Bouchier C, Cochet M, Coulson RMR, Cronin A, de Villiers EP, Fraser A, Fosker N, Gardner M, Goble A, Griffiths-Jones S, Harris DE, Katzer F, Larke N, Lord A, Maser P, McKellar S, Mooney P, Morton F, Nene V, O'Neil S, Price C, Quail MA, Rabbinowitsch E, Rawlings ND, Rutter S, Saunders D, Seeger K, Shah T, Squares R, Squares S, Tivey A, Walker AR, Woodward J, Dobbelaere DAE, Langsley G, Rajandream M-A, McKeever D, Shiels B, Tait A, Barrell B, Hall N (2005) Genome of the host-cell transforming parasite Theileria annulata compared with T. parva. Science 309:131–133. doi: 10.1126/science.1110418 PubMed PMID: 15994557.
95. Kappmeyer LS, Thiagarajan M, Herndon DR, Ramsay JD, Caler E, Djikeng A, Gillespie JJ, Lau AO, Roalson EH, Silva JC, Silva MG, Suarez CE, Ueti MW, Nene VM, Mealey RH, Knowles DP, Brayton KA (2012) Comparative genomic analysis and phylogenetic position of Theileria equi. BMC Genomics 13:603. doi: 10.1186/1471-2164-13-603 PubMed PMID: 23137308.
96. Hayashida K, Hara Y, Abe T, Yamasaki C, Toyoda A, Kosuge T, Suzuki Y, Sato Y, Kawashima S, Katayama T, Wakaguri H, Inoue N, Homma K, Tada-Umezaki M, Yagi Y, Fujii Y, Habara T, Kanehisa M, Watanabe H, Ito K, Gojobori T, Sugawara H, Imanishi T, Weir W, Gardner M, Pain A, Shiels B, Hattori M, Nene V, Sugimoto C (2012) Comparative genome analysis of three eukaryotic parasites with differing abilities to transform leukocytes reveals key mediators of Theileria-induced leukocyte transformation. MBio 3:e00204–12. doi: 10.1128/mBio.00204-12 PubMed PMID: 22951932.

97. Chaudhary K, Roos DS (2005) Protozoan genomics for drug discovery. Nat Biotechnol 23:1089–1091. doi: 10.1038/nbt0905-1089 PubMed PMID: 16151400.
98. Hayashida K, Hara Y, Abe T, Yamasaki C, Toyoda A, Kosuge T, Suzuki Y, Sato Y, Kawashima S, Katayama T, Wakaguri H, Inoue N, Homma K, Tada-Umezaki M, Yagi Y, Fujii Y, Habara T, Kanehisa M, Watanabe H, Ito K, Gojobori T, Sugawara H, Imanishi T, Weir W, Gardner M, Pain A, Shiels B, Hattori M, Nene V, Sugimoto C (2012) Comparative Genome Analysis of Three Eukaryotic Parasites with Differing Abilities To Transform Leukocytes Reveals Key Mediators of Theileria-Induced Leukocyte Transformation. MBio 3:e00204-12–e00204-12. doi: 10.1128/mBio.00204-12 PubMed PMID: 22951932.
99. Tretina K, Gotia HT, Mann DJ, Silva JC (2015) Theileria-transformed bovine leukocytes have cancer hallmarks. Trends in Parasitology 31:306–314. doi: 10.1016/j.pt.2015.04.001 PubMed PMID: 25951781.
100. Cuesta I, González LM, Estrada K, Grande R, Zaballos A, Lobo CA, Barrera J, Sanchez-Flores A, Montero E (2014) High-Quality Draft Genome Sequence of Babesia divergens, the Etiological Agent of Cattle and Human Babesiosis. Genome Announc 2:e01194-14–e01194-14. doi: 10.1128/genomeA.01194-14 PubMed PMID: 25395649.
101. Cornillot E, Dassouli A, Garg A, Pachikara N, Randazzo S, Depoix D, Carcy B, Delbecq S, Frutos R, Silva JC, Sutton R, Krause PJ, Mamoun CB (2013) Whole genome mapping and re-organization of the nuclear and mitochondrial genomes of Babesia microti isolates. PLoS ONE 8:e72657. doi: 10.1371/journal.pone.0072657 PubMed PMID: 24023759.
102. Lau AO, Kalyanaraman A, Echaide I, Palmer GH, Bock R, Pedroni MJ, Rameshkumar M, Ferreira MB, Fletcher TI, McElwain TF (2011) Attenuation of virulence in an apicomplexan hemoparasite results in reduced genome diversity at the population level. BMC Genomics 12:410. doi: 10.1186/1471-2164-12-410 PubMed PMID: 21838895.
103. Henson S, Bishop RP, Morzaria S, Spooner PR, Pelle R, Poveda L, Ebeling M, Küng E, Certa U, Daubenberger CA, Qi W (2012) High-resolution genotyping and mapping of recombination and gene conversion in the protozoan Theileria parva using whole genome sequencing. BMC Genomics 13:503. doi: 10.1186/1471-2164-13-503 PubMed PMID: 22998600.
104. Hayashida K, Abe T, Weir W, Nakao R, Ito K, Kajino K, Suzuki Y, Jongejan F, Geysen D, Sugimoto C (2013) Whole-genome sequencing of Theileria parva strains provides insight into parasite migration and diversification in the African continent. DNA Res 20:209–220. doi: 10.1093/dnares/dst003 PubMed PMID: 23404454.
105. Norling M, Bishop RP, Pelle R, Qi W, Henson S, Drábek EF, Tretina K, Odongo D, Mwaura S, Njoroge T, Bongcam-Rudloff E, Daubenberger CA, Silva JC (2015) The genomes of three stocks comprising the most widely utilized live sporozoite Theileria parva vaccine exhibit very different degrees and patterns of sequence divergence. BMC Genomics 16:729. doi: 10.1186/s12864-015-1910-9 PubMed PMID: 26403690.
106. Brayton KA, Lau AOT, Herndon DR, Hannick L, Kappmeyer LS, Berens SJ, Bidwell SL, Brown WC, Crabtree J, Fadrosh D, Feldblum T, Forberger HA, Haas BJ, Howell JM, Khouri H, Koo H, Mann DJ, Norimine J, Paulsen IT, Radune D, Ren Q, Smith RK, Suarez CE, White O, Wortman JR, Knowles DP, McElwain TF, Nene VM (2007) Genome Sequence of Babesia bovis and Comparative Analysis of Apicomplexan Hemoprotozoa. PLoS Pathog 3:e148. doi: 10.1371/journal.ppat.0030148
107. Cornelissen AW, Schetters TP (1996) Vaccines against protozoal diseases of veterinary importance. FEMS Immunol Med Microbiol 15:61–72. PubMed PMID: 8880130.
108. Lau A, Tibbals DL, McElwain TF (2007) Babesia bovis: the development of an expression oligonucleotide microarray. Exp. Parasitol.
109. Pedroni MJ, Sondgeroth KS, Gallego-Lopez GM, Echaide I, Lau AOT (2013) Comparative transcriptome analysis of geographically distinct virulent and attenuated Babesia bovis strains reveals similar gene expression changes through attenuation. BMC Genomics 14:763. doi: 10.1186/1471-2164-14-763 PubMed PMID: 24195453.
110. Durrani Z, Weir W, Pillai S, Kinnaird J, Shiels B (2012) Modulation of activation-associated host cell gene expression by the apicomplexan parasite Theileria annulata. Cell Microbiol 14:1434–1454. doi: 10.1111/j.1462-5822.2012.01809.x PubMed PMID: 22533473.

111. Kinnaird JH, Weir W, Durrani Z, Pillai SS, Baird M, Shiels BR (2013) A Bovine Lymphosarcoma Cell Line Infected with Theileria annulata Exhibits an Irreversible Reconfiguration of Host Cell Gene Expression. PLoS ONE 8:e66833. doi: 10.1371/journal.pone.0066833 PubMed PMID: 23840536.
112. Witschi M, Xia D, Sanderson S, Baumgartner M, Wastling JM, Dobbelaere DAE (2013) Proteomic analysis of the Theileria annulata schizont. Int J Parasitol 43:173–180. doi: 10.1016/j.ijpara.2012.10.017 PubMed PMID: 23178997.
113. Bishop R, Shah T, Pelle R, Hoyle D, Pearson T, Haines L, Brass A, Hulme H, Graham SP, Taracha ELN, Kanga S, Lu C, Hass B, Wortman J, White O, Gardner MJ, Nene V, de Villiers EP (2005) Analysis of the transcriptome of the protozoan Theileria parva using MPSS reveals that the majority of genes are transcriptionally active in the schizont stage. Nucleic Acids Res 33:5503–5511. doi: 10.1093/nar/gki818 PubMed PMID: 16186131.
114. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P, Cheng Q, Coulson RMR, Crabb BS, del Portillo HA, Essien K, Feldblyum TV, Fernandez-Becerra C, Gilson PR, Gueye AH, Guo X, Kanga S, Kooij TWA, Korsinczky M, Meyer EVS, Nene V, Paulsen I, White O, Ralph SA, Ren Q, Sargeant TJ, Salzberg SL, Stoeckert CJ, Sullivan SA, Yamamoto MM, Hoffman SL, Wortman JR, Gardner MJ, Galinski MR, Barnwell JW, Fraser-Liggett CM (2008) Comparative genomics of the neglected human malaria parasite Plasmodium vivax. Nature 455:757–763. doi: 10.1038/nature07327 PubMed PMID: 18843361.
115. Pain A, Böhme U, Berry AE, Mungall K, Finn RD, Jackson AP, Mourier T, Mistry J, Pasini EM, Aslett MA, Balasubrammaniam S, Borgwardt K, Brooks K, Carret C, Carver TJ, Cherevach I, Chillingworth T, Clark TG, Galinski MR, Hall N, Harper D, Harris D, Hauser H, Ivens A, Janssen CS, Keane T, Larke N, Lapp S, Marti M, Moule S, Meyer IM, Ormond D, Peters N, Sanders M, Sanders S, Sargeant TJ, Simmonds M, Smith F, Squares R, Thurston S, Tivey AR, Walker D, White B, Zuiderwijk E, Churcher C, Quail MA, Cowman AF, Turner CMR, Rajandream M-A, Kocken CHM, Thomas AW, Newbold CI, Barrell BG, Berriman M (2008) The genome of the simian and human malaria parasite Plasmodium knowlesi. Nature 455:799–803. doi: 10.1038/nature07306 PubMed PMID: 18843368.
116. Otto TD, Rayner JC, Böhme U, Pain A, Spottiswoode N, Sanders M, Quail M, Ollomo B, Renaud F, Thomas AW, Prugnolle F, Conway DJ, Newbold C, Berriman M (2014) Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. Nature Communications 5:4754. doi: 10.1038/ncomms5754
117. Tachibana S-I, Sullivan SA, Kawai S, Nakamura S, Kim HR, Goto N, Arisue N, Palacpac NMQ, Honma H, Yagi M, Tougan T, Katakai Y, Kaneko O, Mita T, Kita K, Yasutomi Y, Sutton PL, Shakhbatyan R, Horii T, Yasunaga T, Barnwell JW, Escalante AA, Carlton JM, Tanabe K (2012) Plasmodium cynomolgi genome sequences provide insight into Plasmodium vivax and the monkey malaria clade. Nat Genet 44:1051–1055. doi: 10.1038/ng.2375 PubMed PMID: 22863735.
118. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL, Peterson JD, Pop M, Kosack DS, Shumway MF, Bidwell SL, Shallom SJ, van Aken SE, Riedmuller SB, Feldblyum TV, Cho JK, Quackenbush J, Sedegah M, Shoaibi A, Cummings LM, Florens L, Yates JR, Raine JD, Sinden RE, Harris MA, Cunningham DA, Preiser PR, Bergman LW, Vaidya AB, van Lin LH, Janse CJ, Waters AP, Smith HO, White OR, Salzberg SL, Venter JC, Fraser CM, Hoffman SL, Gardner MJ, Carucci DJ (2002) Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. Nature 419:512–519. doi: 10.1038/nature01099 PubMed PMID: 12368865.
119. Hall N, Karras M, Raine JD, Carlton JM, Kooij TWA, Berriman M, Florens L, Janssen CS, Pain A, Christophides GK, James K, Rutherford K, Harris B, Harris D, Churcher C, Quail MA, Ormond D, Doggett J, Trueman HE, Mendoza J, Bidwell SL, Rajandream M-A, Carucci DJ, Yates JR, Kafatos FC, Janse CJ, Barrell B, Turner CMR, Waters AP, Sinden RE (2005) A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. Science 307:82–86. doi: 10.1126/science.1103717 PubMed PMID: 15637271.
120. Otto TD, Böhme U, Jackson AP, Hunt M, Franke-Fayard B, Hoeijmakers WAM, Religa AA, Robertson L, Sanders M, Ogun SA, Cunningham D, Erhart A, Billker O, Khan SM, Stunnenberg HG, Langhorne J,

Holder AA, Waters AP, Newbold CI, Pain A, Berriman M, Janse CJ (2014) A comprehensive evaluation of rodent malaria parasite genomes and gene expression. BMC Biology 12:86. doi: 10.1186/s12915-014-0086-0 PubMed PMID: 25359557.

121. Harb OS, Roos DS (2015) The Eukaryotic Pathogen Databases: a functional genomic resource integrating data from human and veterinary parasites. Methods Mol Biol 1201:1–18. doi: 10.1007/978-1-4939-1438-8_1 PubMed PMID: 25388105.
122. Goel S, Palmkvist M, Moll K, Joannin N, Lara P, R Akhouri R, Moradi N, Öjemalm K, Westman M, Angeletti D, Kjellin H, Lehtiö J, Blixt O, Ideström L, Gahmberg CG, Storry JR, Hult AK, Olsson ML, Heijne von G, Nilsson I, Wahlgren M (2015) RIFINs are adhesins implicated in severe Plasmodium falciparum malaria. Nat Med -. doi: 10.1038/nm.3812
123. Campbell DA, Thomas S, Sturm NR (2003) Transcription in kinetoplastid protozoa: why be normal? Microbes Infect 5:1231–1240. PubMed PMID: 14623019.
124. Simpson AGB, Stevens JR, Lukeš J (2006) The evolution and diversity of kinetoplastid flagellates. Trends in Parasitology 22:168–174. doi: 10.1016/j.pt.2006.02.006 PubMed PMID: 16504583.
125. Votýpka J, d'Avila-Levy CM, Grellier P, Maslov DA, Lukeš J, Yurchenko V (2015) New Approaches to Systematics of Trypanosomatidae: Criteria for Taxonomic (Re)description. Trends in Parasitology 31:460–469. doi: 10.1016/j.pt.2015.06.015 PubMed PMID: 26433249.
126. Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG, Figueroa-Angulo EE (2010) Gene expression in trypanosomatid parasites. J Biomed Biotechnol 2010:525241–15. doi: 10.1155/2010/525241 PubMed PMID: 20169133.
127. Aphasizhev R, Aphasizheva I (2014) Mitochondrial RNA editing in trypanosomes: small RNAs in control. Biochimie 100:125–131. doi: 10.1016/j.biochi.2014.01.003 PubMed PMID: 24440637.
128. Lukeš J, Guilbride DL, Votýpka J, Zíková A, Benne R, Englund PT (2002) Kinetoplast DNA network: evolution of an improbable structure. Eukaryotic Cell 1:495–502. doi: 10.1128/EC.1.4.495-502.2002 PubMed PMID: 12455998.
129. Maree JP, Patterton H-G (2014) The epigenome of Trypanosoma brucei: a regulatory interface to an unconventional transcriptional machine. Biochim Biophys Acta 1839:743–750. doi: 10.1016/j.bbagrm.2014.05.028 PubMed PMID: 24942804.
130. Daniels J-P, Gull K, Wickstead B (2010) Cell biology of the trypanosome genome. Microbiol Mol Biol Rev 74:552–569. doi: 10.1128/MMBR.00024-10 PubMed PMID: 21119017.
131. Lukeš J, Skalický T, Týč J, Votýpka J, Yurchenko V (2014) Evolution of parasitism in kinetoplastid flagellates. Mol Biochem Parasitol 195:115–122. doi: 10.1016/j.molbiopara.2014.05.007 PubMed PMID: 24893339.
132. Clayton CE (2014) Networks of gene expression regulation in Trypanosoma brucei. Mol Biochem Parasitol 195:96–106. doi: 10.1016/j.molbiopara.2014.06.005 PubMed PMID: 24995711.
133. Kelly S, Kramer S, Schwede A, Maini PK, Gull K, Carrington M (2012) Genome organization is a major component of gene expression control in response to stress and during the cell division cycle in trypanosomes. Open Biol 2:120033–120033. doi: 10.1098/rsob.120033 PubMed PMID: 22724062.
134. Jackson AP (2015) Genome evolution in trypanosomatid parasites. Parasitology 142 Suppl 1:S40–56. doi: 10.1017/S0031182014000894 PubMed PMID: 25068268.
135. Clayton C (2013) The regulation of trypanosome gene expression by RNA-binding proteins. PLoS Pathog 9:e1003680. doi: 10.1371/journal.ppat.1003680 PubMed PMID: 24244152.
136. Vasquez J-J, Hon C-C, Vanselow JT, Schlosser A, Siegel TN (2014) Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages. Nucleic Acids Res 42:3623–3637. doi: 10.1093/nar/gkt1386 PubMed PMID: 24442674.
137. Jensen BC, Ramasamy G, Vasconcelos EJR, Ingolia NT, Myler PJ, Parsons M (2014) Extensive stage-regulation of translation revealed by ribosome profiling of Trypanosoma brucei. BMC Genomics 15:911. doi: 10.1186/1471-2164-15-911 PubMed PMID: 25331479.
138. Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, Harris D, Her Y, Herzyk P, Imamura H, Otto TD, Sanders M, Seeger K, Dujardin J-C, Berriman M, Smith DF, Hertz-Fowler C, Mottram JC

(2011) Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. Genome Res 21:2129–2142. doi: 10.1101/gr.122945.111 PubMed PMID: 22038252.

139. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran A-N, Wortman JR, Alsmark UCM, Angiuoli S, Anupama A, Badger J, Bringaud F, Cadag E, Carlton JM, Cerqueira GC, Creasy T, Delcher AL, Djikeng A, Embley TM, Hauser C, Ivens AC, Kummerfeld SK, Pereira-Leal JB, Nilsson D, Peterson J, Salzberg SL, Shallom J, Silva JC, Sundaram J, Westenberger S, White O, Melville SE, Donelson JE, Andersson B, Stuart KD, Hall N (2005) Comparative genomics of trypanosomatid parasitic protozoa. Science 309:404–409. doi: 10.1126/science.1112181 PubMed PMID: 16020724.
140. Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, Wang X, Dewell S, Cross GAM (2009) Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. Genes Dev 23:1063–1076. doi: 10.1101/gad.1790409 PubMed PMID: 19369410.
141. Marande W, Lukeš J, Burger G (2005) Unique mitochondrial genome structure in diplonemids, the sister group of kinetoplastids. Eukaryotic Cell 4:1137–1146. doi: 10.1128/EC.4.6.1137-1146.2005 PubMed PMID: 15947205.
142. Liu B, Liu Y, Motyka SA, Agbo EEC, Englund PT (2005) Fellowship of the rings: the replication of kinetoplast DNA. Trends in Parasitology 21:363–369. doi: 10.1016/j.pt.2005.06.008 PubMed PMID: 15967722.
143. Bringaud F, Ghedin E, Blandin G, Bartholomeu DC, Caler E, Levin MJ, Baltz T, El-Sayed NM (2006) Evolution of non-LTR retrotransposons in the trypanosomatid genomes: Leishmania major has lost the active elements. Mol Biochem Parasitol 145:158–170. doi: 10.1016/j.molbiopara.2005.09.017 PubMed PMID: 16257065.
144. Bringaud F, Ghedin E, El-Sayed NMA, Papadopoulou B (2008) Role of transposable elements in trypanosomatids. Microbes Infect 10:575–581. doi: 10.1016/j.micinf.2008.02.009 PubMed PMID: 18467144.
145. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Böhme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UCM, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth T-J, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DMA, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream M-A, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CMR, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM (2005) The genome of the African trypanosome Trypanosoma brucei. Science 309:416–422. doi: 10.1126/science.1112642 PubMed PMID: 16020726.
146. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream M-A, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchettin G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RMR, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble A, Goffeau A, Harris D, Hertz-Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N, Litvin L, Lord A, Louie T, Marra M, Masuy D, Matthews K, Michaeli S, Mottram JC, Müller-Auer S, Munden H, Nelson S, Norbertczak H, Oliver K, O'Neil S, Pentony M, Pohl TM, Price C, Purnelle B, Quail MA, Rabbinowitsch E, Reinhardt R, Rieger M, Rinta J, Robben J, Robertson L, Ruiz JC, Rutter S, Saunders D, Schäfer M, Schein J, Schwartz DC, Seeger K, Seyler A, Sharp S, Shin H, Sivam D, Squares R, Squares S, Tosato V, Vogt C, Volckaert G, Wambutt R, Warren T, Wedler H, Woodward J, Zhou S, Zimmermann W, Smith DF, Blackwell JM, Stuart KD, Barrell B, Myler PJ (2005) The

genome of the kinetoplastid parasite, Leishmania major. Science 309:436–442. doi: 10.1126/science.1112680 PubMed PMID: 16020728.

147. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran A-N, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B (2005) The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science 309:409–415. doi: 10.1126/science.1112631 PubMed PMID: 16020725.
148. Downing T, Imamura H, Decuypere S, Clark TG, Coombs GH, Cotton JA, Hilley JD, de Doncker S, Maes I, Mottram JC, Quail MA, Rijal S, Sanders M, Schönian G, Stark O, Sundar S, Vanaerschot M, Hertz-Fowler C, Dujardin J-C, Berriman M (2011) Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance. Genome Res 21:2143–2156. doi: 10.1101/gr.123430.111 PubMed PMID: 22038251.
149. Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, Quail MA, Peters N, Adlem E, Tivey A, Aslett M, Kerhornou A, Ivens A, Fraser A, Rajandream M-A, Carver T, Norbertczak H, Chillingworth T, Hance Z, Jagels K, Moule S, Ormond D, Rutter S, Squares R, Whitehead S, Rabbinowitsch E, Arrowsmith C, White B, Thurston S, Bringaud F, Baldauf SL, Faulconbridge A, Jeffares D, Depledge DP, Oyola SO, Hilley JD, Brito LO, Tosi LRO, Barrell B, Cruz AK, Mottram JC, Smith DF, Berriman M (2007) Comparative genomic analysis of three Leishmania species that cause diverse human disease. Nat Genet 39:839–847. doi: 10.1038/ng2053 PubMed PMID: 17572675.
150. Llanes A, Restrepo CM, Del Vecchio G, Anguizola FJ, Lleonart R (2015) The genome of Leishmania panamensis: insights into genomics of the L. (Viannia) subgenus. Sci Rep 5:8550. doi: 10.1038/srep08550 PubMed PMID: 25707621.
151. Valdivia HO, Reis-Cunha JL, Rodrigues-Luiz GF, Baptista RP, Baldeviano GC, Gerbasi RV, Dobson DE, Pratlong F, Bastien P, Lescano AG, Beverley SM, Bartholomeu DC (2015) Comparative genomic analysis of Leishmania (Viannia) peruviana and Leishmania (Viannia) braziliensis. BMC Genomics 16:715. doi: 10.1186/s12864-015-1928-z PubMed PMID: 26384787.
152. Real F, Vidal RO, Carazzolle MF, Mondego JMC, Costa GGL, Herai RH, Würtele M, de Carvalho LM, Carmona e Ferreira R, Mortara RA, Barbiéri CL, Mieczkowski P, da Silveira JF, Briones MRDS, Pereira GAG, Bahia D (2013) The genome sequence of Leishmania (Leishmania) amazonensis: functional annotation and extended analysis of gene models. DNA Res 20:567–581. doi: 10.1093/dnares/dst031 PubMed PMID: 23857904.
153. Stoco PH, Wagner G, Talavera-Lopez C, Gerber A, Zaha A, Thompson CE, Bartholomeu DC, Lückemeyer DD, Bahia D, Loreto E, Prestes EB, Lima FM, Rodrigues-Luiz G, Vallejo GA, Filho JFDS, Schenkman S, Monteiro KM, Tyler KM, de Almeida LGP, Ortiz MF, Chiurillo MA, de Moraes MH, Cunha O de L, Mendonça-Neto R, Silva R, Teixeira SMR, Murta SMF, Sincero TCM, Mendes TA de O, Urmenyi TP, Silva VG, DaRocha WD, Andersson B, Romanha AJ, Steindel M, de Vasconcelos ATR, Grisard EC (2014) Genome of the avirulent human-infective trypanosome--Trypanosoma rangeli. PLoS Negl Trop Dis 8:e3176. doi: 10.1371/journal.pntd.0003176 PubMed PMID: 25233456.
154. Kelly S, Ivens A, Manna PT, Gibson W, Field MC (2014) A draft genome for the African crocodilian trypanosome Trypanosoma grayi. Sci Data 1:140024. doi: 10.1038/sdata.2014.24 PubMed PMID: 25977781.
155. Raymond F, Boisvert S, Roy G, Ritt J-F, Légaré D, Isnard A, Stanke M, Olivier M, Tremblay MJ, Papadopoulou B, Ouellette M, Corbeil J (2012) Genome sequencing of the lizard parasite Leishmania

tarentolae reveals loss of genes associated to the intracellular stage of human pathogenic species. Nucleic Acids Res 40:1131–1147. doi: 10.1093/nar/gkr834 PubMed PMID: 21998295.
156. Carnes J, Anupama A, Balmer O, Jackson A, Lewis M, Brown R, Cestari I, Desquesnes M, Gendrin C, Hertz-Fowler C, Imamura H, Ivens A, Kořený L, Lai D-H, Macleod A, McDermott SM, Merritt C, Monnerat S, Moon W, Myler P, Phan I, Ramasamy G, Sivam D, Lun Z-R, Lukeš J, Stuart K, Schnaufer A (2015) Genome and phylogenetic analyses of Trypanosoma evansi reveal extensive similarity to T. brucei and multiple independent origins for dyskinetoplasty. PLoS Negl Trop Dis 9:e3404. doi: 10.1371/journal.pntd.0003404 PubMed PMID: 25568942.
157. Kořený L, Sobotka R, Kovářová J, Gnipová A, Flegontov P, Horváth A, Oborník M, Ayala FJ, Lukeš J (2012) Aerobic kinetoplastid flagellate Phytomonas does not require heme for viability. Proceedings of the National Academy of Sciences 109:3808–3813. doi: 10.1073/pnas.1201089109
158. Porcel BM, Denoeud F, Opperdoes F, Noel B, Madoui M-A, Hammarton TC, Field MC, Da Silva C, Couloux A, Poulain J, Katinka M, Jabbari K, Aury J-M, Campbell DA, Cintron R, Dickens NJ, Docampo R, Sturm NR, Koumandou VL, Fabre S, Flegontov P, Lukeš J, Michaeli S, Mottram JC, Szöőr B, Zilberstein D, Bringaud F, Wincker P, Dollet M (2014) The streamlined genome of Phytomonas spp. relative to human pathogenic kinetoplastids reveals a parasite tailored for plants. PLoS Genetics 10:e1004007. doi: 10.1371/journal.pgen.1004007 PubMed PMID: 24516393.
159. Kraeva N, Butenko A, Hlaváčová J, Kostygov A, Myškova J, Grybchuk D, Leštinová T, Votýpka J, Volf P, Opperdoes F, Flegontov P, Lukeš J, Yurchenko V (2015) Leptomonas seymouri: Adaptations to the Dixenous Life Cycle Analyzed by Genome Sequencing, Transcriptome Profiling and Co-infection with Leishmania donovani. PLoS Pathog 11:e1005127. doi: 10.1371/journal.ppat.1005127 PubMed PMID: 26317207.
160. Runckel C, DeRisi J, Flenniken ML (2014) A draft genome of the honey bee trypanosomatid parasite Crithidia mellificae. PLoS ONE 9:e95057. doi: 10.1371/journal.pone.0095057 PubMed PMID: 24743507.
161. Alves JMP, Klein CC, da Silva FM, Costa-Martins AG, Serrano MG, Buck GA, Vasconcelos ATR, Sagot M-F, Teixeira MMG, Motta MCM, Camargo EP (2013) Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers. BMC Evol Biol 13:190. doi: 10.1186/1471-2148-13-190 PubMed PMID: 24015778.
162. Dubey JP, Ferguson DJP (2014) Life Cycle of Hammondia hammondi(Apicomplexa: Sarcocystidae) in Cats. J Eukaryot Microbiol 62:346–352. doi: 10.1111/jeu.12188 PubMed PMID: 25312612.
163. Dubey JP, Lindsay DS (1996) A review of Neospora caninum and neosporosis. Vet Parasitol. doi: 10.1016/S0304-4017(96)01035-7
164. Levine ND (1986) The taxonomy of Sarcocystis (Protozoa, Apicomplexa) species. J Parasitol 72:372–382. doi: 10.2307/3281676 PubMed PMID: 3091802.
165. Dubey JP, Lindsay DS, Speer CA (1998) Structures of Toxoplasma gondii tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. Clin Microbiol Rev 11:267–299. doi: 10.1016/S0001-706X(97)00656-6 PubMed PMID: 9564564.
166. Sibley LD, Ajioka JW (2008) Population structure of Toxoplasma gondii: clonal expansion driven by infrequent recombination and selective sweeps. Annu Rev Microbiol 62:329–351. doi: 10.1146/annurev.micro.62.081307.162925 PubMed PMID: 18544039.
167. Francia ME, Striepen B (2014) Cell division in apicomplexan parasites. Nature Reviews Microbiology. doi: 10.1038/nrmicro3184
168. Ouologuem DT, Roos DS (2014) Dynamics of the Toxoplasma gondii inner membrane complex. J Cell Sci 127:3320–3330. doi: 10.1242/jcs.147736 PubMed PMID: 24928899.
169. Miller CM, Boulter NR, Ikin RJ, Smith NC (2009) The immunobiology of the innate response to Toxoplasma gondii. Int J Parasitol 39:23–39. doi: 10.1016/j.ijpara.2008.08.002 PubMed PMID: 18775432.
170. Khan A, Taylor S, Su C, Mackey AJ, Boyle J, Cole R, Glover D, Tang K, Paulsen IT, Berriman M, Boothroyd JC, Pfefferkorn ER, Dubey JP, Ajioka JW, Roos DS, Wootton JC, Sibley LD (2005) Composite genome map

and recombination parameters derived from three archetypal lineages of Toxoplasma gondii. Nucleic Acids Res 33:2980–2992. doi: 10.1093/nar/gki604 PubMed PMID: 15911631.

171. Reid AJ, Vermont SJ, Cotton JA, Harris D, Hill-Cawthorne GA, Könen-Waisman S, Latham SM, Mourier T, Norton R, Quail MA, Sanders M, Shanmugam D, Sohal A, Wasmuth JD, Brunk B, Grigg ME, Howard JC, Parkinson J, Roos DS, Trees AJ, Berriman M, Pain A, Wastling JM (2012) Comparative genomics of the apicomplexan parasites Toxoplasma gondii and Neospora caninum: Coccidia differing in host range and transmission strategy. PLoS Pathog 8:e1002567. doi: 10.1371/journal.ppat.1002567 PubMed PMID: 22457617.
172. Lorenzi H, Khan A, Behnke MS, Namasivayam S, Swapna LS, Hadjithomas M, Karamycheva S, Pinney D, Brunk BP, Ajioka JW, Ajzenberg D, Boothroyd JC, Boyle JP, Dardé ML, Diaz-Miranda MA, Dubey JP, Fritz HM, Gennari SM, Gregory BD, Kim K, Saeij JPJ, Su C, White MW, Zhu X-Q, Howe DK, Rosenthal BM, Grigg ME, Parkinson J, Liu L, Kissinger JC, Roos DS, David Sibley L (2016) Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic Toxoplasma gondii genomes. Nature Communications 7:10147. doi: 10.1038/ncomms10147
173. Köhler S, Delwiche CF, Denny PW, Tilney LG (1997) A plastid of probable green algal origin in Apicomplexan parasites. Science. doi: 10.1126/science.275.5305.1485
174. Kissinger JC, Gajria B, Li L, Paulsen IT, Roos DS (2003) ToxoDB: accessing the Toxoplasma gondii genome. Nucleic Acids Res 31:234–236. doi: 10.1093/nar/gkg072 PubMed PMID: 12519989.
175. Walzer KA, Adomako-Ankomah Y, Dam RA, Herrmann DC, Schares G, Dubey JP, Boyle JP (2013) Hammondia hammondi, an avirulent relative of Toxoplasma gondii, has functional orthologs of known T. gondii virulence genes. Proceedings of the National Academy of Sciences 110:7446–7451. doi: 10.1073/pnas.1304322110
176. Blazejewski T, Nursimulu N, Pszenny V, Dangoudoubiyam S, Namasivayam S, Chiasson MA, Chessman K, Tonkin M, Swapna LS, Hung SS, Bridgers J, Ricklefs SM, Boulanger MJ, Dubey JP, Porcella SF, Kissinger JC, Howe DK, Grigg ME, Parkinson J (2015) Systems-based analysis of the Sarcocystis neurona genome identifies pathways that contribute to a heteroxenous life cycle. MBio 6:e02445–14. doi: 10.1128/mBio.02445-14 PubMed PMID: 25670772.
177. Hakimi M-A, Bougdour A (2015) Toxoplasma's ways of manipulating the host transcriptome via secreted effectors. Curr Opin Microbiol 26:24–31. doi: 10.1016/j.mib.2015.04.003 PubMed PMID: 25912924.
178. Mercier C, Cesbron-Delauw M-F (2015) Toxoplasma secretory granules: one population or more? Trends in Parasitology 31:60–71. doi: 10.1016/j.pt.2014.12.002 PubMed PMID: 25599584.
179. Boothroyd JC, Dubremetz JF (2008) Kiss and spit: the dual roles of Toxoplasma rhoptries. Nature Reviews Microbiology 6:79–88. doi: 10.1038/nrmicro1800 PubMed PMID: 18059289.
180. Carruthers VB, Tomley FM (2008) Microneme proteins in apicomplexans. Subcell Biochem 47:33–45. PubMed PMID: 18512339.
181. Wasmuth JD, Pszenny V, Haile S, Jansen EM, Gast AT, Sher A, Boyle JP, Boulanger MJ, Parkinson J, Grigg ME (2012) Integrated bioinformatic and targeted deletion analyses of the SRS gene superfamily identify SRS29C as a negative regulator of Toxoplasma virulence. MBio. doi: 10.1128/mBio.00321-12
182. Dalmasso MC, Carmona SJ, Angel SO, Agüero F (2014) Characterization of Toxoplasma gondii subtelomeric-like regions: identification of a long-range compositional bias that is also associated with gene-poor regions. BMC Genomics 15:1. doi: 10.1186/1471-2164-15-21 PubMed PMID: 24382143.
183. Behnke MS, Zhang TP, Dubey JP, Sibley LD (2014) Toxoplasma gondii merozoite gene expression analysis with comparison to the life cycle discloses a unique expression state during enteric development. BMC Genomics 15:350. doi: 10.1186/1471-2164-15-350 PubMed PMID: 24885521.
184. Greene JM, Collins F, Lefkowitz EJ, Roos D, Scheuermann RH, Sobral B, Stevens R, White O, Di Francesco V (2007) National Institute of Allergy and Infectious Diseases bioinformatics resource centers: new assets for pathogen informatics. Infect Immun 75:3212–3219. doi: 10.1128/IAI.00105-07 PubMed PMID: 17420237.
185. Aurrecoechea C, Brestelli J, Brunk BP, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer ET, Li W, Miller JA, Nayak V, Pennington C, Pinney DF,

Roos DS, Ross C, Srinivasamoorthy G, Stoeckert CJ, Thibodeau R, Treatman C, Wang H (2010) EuPathDB: a portal to eukaryotic pathogen databases. Nucleic Acids Res 38:D415–9. doi: 10.1093/nar/gkp941 PubMed PMID: 19914931.

186. Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Cade S, Doherty R, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Hu S, Iodice J, Kissinger JC, Kraemer ET, Li W, Pinney DF, Pitts B, Roos DS, Srinivasamoorthy G, Stoeckert CJ, Wang H, Warrenfeltz S (2013) EuPathDB: the eukaryotic pathogen database. Nucleic Acids Res 41:D684–91. doi: 10.1093/nar/gks1113 PubMed PMID: 23175615.