Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Crit Rev Biochem Mol Biol. Author manuscript; available in PMC 2019 October 25.

Published in final edited form as:

Crit Rev Biochem Mol Biol. 2017 June ; 52(3): 254–273. doi:10.1080/10409238.2017.1290043.

Genomics of Apicomplexan Parasites

Lakshmipuram S. Swapna[1], **John Parkinson**[1,2]

[1.]Program in Molecular Structure and Function, Hospital for Sick Children, Toronto, Ontario, Canada

[2.]Departments of Biochemistry, Molecular Genetics and Computer Science, University of Toronto, Toronto, Ontario, Canada

Abstract

The increasing prevalence of infections involving intracellular apicomplexan parasites such as *Plasmodium*, *Toxoplasma*, and *Cryptosporidium* (the causative agents of malaria, toxoplasmosis and cryptosporidiosis respectively) represent a significant global healthcare burden. Despite their significance, few treatments are available; a situation that is likely to deteriorate with the emergence of new resistant strains of parasites. To lay the foundation for programs of drug discovery and vaccine development, genome sequences for many of these organisms have been generated, together with large scale expression and proteomic datasets. Comparative analyses of these datasets are beginning to identify the molecular innovations supporting both conserved processes mediating fundamental roles in parasite survival and persistence, as well as lineage-specific adaptations associated with divergent life cycle strategies. The challenge is how best to exploit these data to derive insights into parasite virulence and identify those genes representing the most amenable targets.

In this review, we outline genomic datasets currently available for apicomplexans and discuss biological insights that have emerged as a consequence of their analysis. Of particular interest are systems-based resources, focusing on areas of metabolism and host invasion that are opening up opportunities for discovering new therapeutic targets.

Keywords

apicomplexan genomics; parasite genomics; systems-based approaches; metabolism; host-invasion; host cell modulation

Introduction

Apicomplexans are a group of single celled obligate eukaryotic intracellular parasites (Cavalier-Smith, 1993, Adl *et al.*, 2007), that form a major clade of the superphylum,

Correspondence should be addressed to: john.parkinson@utoronto.ca, swapna.seshadri@sickkids.ca, Mailing Address: Molecular Structure and Function, Hospital for Sick Children, Peter Gilgan Center for Research and Learning, 20th Floor, RM 20.9709, 686 Bay Street, Toronto, ON, M5G 0A4, Canada, Phone: +1 416 813 5746, FAX: +1 416 813 5022.

Declaration of interest:
The authors declare that they have no conflict of interest.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Alveolata. The prevalence of infections involving parasites such as *Plasmodium, Cryptosporidium* and *Toxoplasma* (the causative agents of malaria, cryptosporidiosis and toxoplasmosis) represents a huge burden on child health worldwide: malaria impacts over 200 million individuals and kills over 300000 children (WHO, 2016); *Cryptosporidium* infections are the 2nd leading cause of the 800,000 infant diarrheal associated fatalities (Kotloff *et al.*, 2013, Checkley *et al.*, 2015); while *Toxoplasma*, which is estimated to infect about 30% of the worldwide population (Flegr *et al.*, 2014), is the leading cause of infectious retinitis in children and is life-threatening in pregnancy and to the immunocompromised (Figure 1) (Garza-Leon and Garcia, 2012, Torgerson and Mastroiacovo, 2013, Schluter *et al.*, 2014). Other species, such as members of *Eimeria* and *Neospora*, have a significant impact on agriculture, causing huge losses in the poultry (McDonald and Shirley, 2009) and cattle (Goodswen *et al.*, 2013) industries, respectively. Despite their significance, treatment options are limited and the emergence of strains resistant to current drugs highlights an urgent need to develop new therapeutics (Hyde, 2007, Petersen *et al.*, 2011a, Doliwa *et al.*, 2013, Miotto *et al.*, 2013, Seeber and Steinfelder, 2016). To help drive such programs, technological advances in sequencing and other high-throughput approaches has enabled researchers to assemble extensive genomic, transcriptomic, and proteomic datasets for many of these organisms. The current challenge lies in integrating these complementary datasets to better understand the biological processes by which these parasites are able to infect their hosts and cause disease, with emphasis on identifying which proteins play essential roles and may be targeted for intervention strategies.

Apicomplexan parasites inhabit a wide range of environments, from marine to terrestrial, and also exhibit variation in host specificity. They possess complex life cycles, often involving multiple hosts and developmental stages (Figure 1b). These stages include: sporogony (invasive stage, with a single round of asexual reproduction), merogony (invasive stage, with multiple rounds of asexual reproduction) and gametogony (sexual reproduction). These stages may occur in the same organism and same tissue (monoxenic lifestyle) or in different organisms and different tissues (heteroxenic lifestyle).

Apicomplexans can be classified into three major clades based on their phylogenetic relationships and host specificities: Aconoidasida, Coccidia and a third lineage featuring Gregarina and Cryptosporidium species (Barta and Thompson, 2006, Adl *et al.*, 2007, Wasmuth *et al.*, 2009). The Aconoidasida include the Haemosporida (Plasmodium) and the Piroplasmida (Babesia and Theileria), and lead heteroxenous life cycles, alternating between an arthropod vector, in which the parasite undergoes sexual reproduction, and a vertebrate host supporting asexual propagation, typically in the circulatory system. On the other hand, Cryptosporidium species are restricted to the gastro-intestinal (GI) tract of animals while Coccidia feature members that are either fully, or partially restricted to the GI tract. For example, the coccidian *Eimeria tenella*, which causes coccidiosis in poultry, undergoes a monoxenic lifestyle, restricted to a single host in which it colonizes epithelial cells of the intestine. Other, so called ‘tissue-cyst forming’ coccidians such as *Sarcocystis hominis* and *Toxoplasma gondii* feature heteroxenous life cycles where gametogony occurs in the intestinal epithelium of the definitive host (e.g. cat for *T. gondii* and human for *S. hominis*) and merogony occurs in the tissues of intermediate hosts. Although individual lineages share distinctive life cycle strategies, each species has its own specialisation with respect to host

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

range, and even the tissue type that it infects (Cowper *et al.*, 2012, Woo *et al.*, 2015). For example *T. gondii* is thought to be capable of exploiting all warm blooded animals as intermediate hosts, while *S. hominis* relies on bovines as its intermediate host. Likewise, although all ~200 Plasmodium species infect erythrocytes, they show specialised adaptations; *P. vivax* only infects smaller (or) newer reticulocytes (Malleret *et al.*, 2015), whereas *P. falciparum* can infect all reticulocytes (White, 2009).

All apicomplexans are characterised by several unique ultrastructural features and organelles that provide essential functions for the parasite to complete its life cycle (Figure 1a). The inner membrane complex (IMC) is a highly specialised endomembrane system found directly beneath plasma membrane in all alveolates, made up of several flattened sacs called alveoli, providing shape and stability to the cell as well as enabling replication, motility and invasion (Gubbels and Duraisingh, 2012, Harding and Meissner, 2014). The unique secretory organelles found at the apical end of the parasite – the bar shaped micronemes, club shaped rhoptries (consisting of two subcompartments – rhoptry neck region and rhoptry bulb region), and dense granules – are instrumental for carrying out motility, invasion, and host modulation processes (Kemp *et al.*, 2013). Dense granules are known to be present in Coccidia and *Cryptosporidium* (Bonnin *et al.*, 1995). Although dense granule-like structures have been identified in *Theileria* (Shaw *et al.*, 1991), *Babesia* (Gohil *et al.*, 2010) and *Plasmodium* (Culvenor *et al.*, 1991), it is not clear whether they perform equivalent roles (Mercier *et al.*, 2005). Another unique aspect of apicomplexans is the apicoplast, which is a four-membraned organelle hosting several important metabolic pathways and essential for the survival of the parasite (Lim and McFadden, 2010, McFadden and Yeh, 2016). This organelle is absent in *Cryptosporidium* and *Gregarina*.

In 2002, the first genome of an apicomplexan, *P. falciparum,* was sequenced (Gardner *et al.*, 2002). Since then, the advent of next generation sequencing (NGS) technologies have resulted in increasingly impressive genomic and transcriptomic datasets (Goodwin *et al.*, 2016). Over the last decade, genomes from all major apicomplexan clades have been generated, along with their close relatives, the free living, non-parasitic chromerids (Abrahamsen *et al.*, 2004, Xu *et al.*, 2004, Gardner *et al.*, 2005, Pain *et al.*, 2005, Carlton *et al.*, 2008, Reid *et al.*, 2012, Walzer *et al.*, 2013, Reid *et al.*, 2014, Blazejewski *et al.*, 2015, Woo *et al.*, 2015, Lorenzi *et al.*, 2016b) (Figure 1d). Further, falling costs in sequencing is ushering in a new era of population genomics (Ellegren, 2014), in which tens or even hundreds of isolates may be sequenced to gain insights into evolutionary dynamics and epidemiology (Miotto *et al.*, 2013, Miotto *et al.*, 2015, Lorenzi *et al.*, 2016a).

Analysis of apicomplexan genomes shows that like other parasites, the genome sizes of apicomplexans are reduced in comparison to those of free-living eukaryotes (Kissinger and DeBarry, 2011). The largest apicomplexan genomes belong to the Coccidia clade, which vary from 50–130 million base pairs (Mbp), while those from other clades are much smaller (Haemosporidia - 20–25 Mbp; Piroplasmida - 8 Mbp; and *Cryptosporidium spp.* - 9 Mbp). Although the factors governing genome sizes in apicomplexans is unknown, it is likely that they are partially influenced by the lineage-specific losses (Woo *et al.*, 2015) observed during their evolution as well as occurrence of repetitive elements, which are prevalent in the genomes of *Eimeria spp.*, *P. falciparum* and *S. neurona* (Battistuzzi *et al.*, 2016). Genome

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

size is reflected in terms of the number of genes encoded, with *B. bovis* possessing the smallest complement (3781 genes) and *T. gondii* the largest (8920 genes) (Figure 1d). The reduced genomes of *Cryptosporidium spp.* appears to be related to their ability to salvage many nutrients from their host, reducing the need to support many complex biochemical pathways. Interestingly, the number of chromosomes associated with each group appears conserved: 4 in the Piroplasmids (Jones *et al.*, 1997, Pain *et al.*, 2005, Katzer *et al.*, 2011), 8 in *Cryptosporidium spp.* (Blunt *et al.*, 1997, Xu *et al.*, 2004), and 14 in Haemosporidia (Janse *et al.*, 1994, Gardner *et al.*, 2002, Pain *et al.*, 2008, Tachibana *et al.*, 2012). In Coccidia, genetic linkage maps are available only for *T. gondii*, identifying 14 chromosomes (Khan *et al.*, 2005).

Genomic resources available for apicomplexan parasites

As of October 2016, genomes of 18 apicomplexan species have been sequenced, with some species featuring genomes from many additional strains. Genome sequences have been deposited in major repositories such as those available at the Wellcome Trust Sanger Institute, National Centre for Biotechnology Information, and GeneDB. However, the most comprehensive resource for apicomplexan genomic datasets is EupathDB (Aurrecoechea *et al.*, 2013) - a central repository for storing information about eukaryotic pathogens. The database is subdivided according to apicomplexan parasite clades – CryptoDB (Cryptosporidum, Gregarina), ToxoDB (Coccidians), PlasmoDB, PiroplasmaDB (Babesia, Theileria), with data for chromerids also deposited in CryptoDB. This portal summarizes the available genomic resources for the organisms and serves as a single-point access for various datasets including: genomic (including nuclear and organellar sequences and population genomic data), transcriptomic (expressed sequence tags, microarrays, RNA-Seq and SAGE tags), proteomic and epigenomic (ChIP-Seq). Two other noteworthy resources for transcriptomic data include the Database of Apicomplexa Transcriptomics, an up-to-date portal hosting transcriptomic datasets of various kinds, including single-cell and isolate-based data (Jakalski *et al.*, 2015) and PartiGeneDB, which hosts EST datasets (Parkinson *et al.*, 2004). Proteomic datasets specific to *T. gondii* and *C. parvum* are also available in EpicDB (Madrid-Aliste *et al.*, 2009).

Aside from EupathDB, several apicomplexan-specific repositories dedicated to specialised aspects of apicomplexan biology are also available. The Liverpool Library of Apicomplexan Metabolic Pathways (LLAMP) is a web resource that provides draft metabolic reconstructions for eight species based on annotations provided by EuPathDB supplemented with additional literature evidence as well as sequence homology based predictions (Shanmugasundram *et al.*, 2013). The Malaria Parasite Metabolic Pathways (MPMP) is a manually curated web repository for functional genomics of *P. falciparum* (Ginsburg and Abdel-Haleem, 2016). Serving as a gold standard for metabolic annotations for *P. falciparum*, the MPMP database also provides information regarding gene expression data, protein localization, and drug-related data (e.g. links to the PubChem small molecule database (Kim *et al.*, 2016) and literature sources) for the parasite. Additional tools for predicting protein localization to various sub-cellular organelles have been developed, such as ApicoAP (Cilingir *et al.*, 2012) and HECTAR (Gschloessl *et al.*, 2008) for predicting apicoplast-targeting motifs, as well as the more generic iPSORT (Bannai *et al.*, 2002) for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

recognising signal peptide, mitochondrial target peptides, and chloroplast target peptides, and SignalP (Petersen *et al.*, 2011b) for predicting secreted proteins.

Gene complements provide insights into apicomplexan biology

Comparisons between gene complements across apicomplexans, highlight both core conserved functionality as well as lineage-specific innovations. Focusing on genes involved in housekeeping functions, such as translation, protein folding and cell cycle, each species features roughly equivalent sets (Table 1), with the higher numbers associated with *P. falciparum* and *T. gondii* likely associated with more extensive curation efforts. Surprisingly, some of the proteins involved in housekeeping function are reported to be significantly diverged enough from other eukaryotes to not be related by sequence similarity (Wasmuth *et al.*, 2009). In terms of metabolism and transport, gene numbers reflect lineage-specific distribution, with higher losses in *Cryptosporidium spp.* and Piroplasmids. Most significantly, almost 50% of the proteins within a species are of unknown function (Table 1), with 77% unknown for *T. parva* and 68% unknown for *E. tenella*, emphasising the pressing need to improve functional annotation efforts. Comprehensive sequence based comparisons suggest that most apicomplexan sequences are specific to the phylum with only 25% of apicomplexan sequences sharing significant similarity with sequences outside the phylum (Wasmuth *et al.*, 2009). Furthermore many genera-specific innovations within the Apicomplexa are associated with specialised pathways involved in host cell invasion and modification of host processes (Wasmuth *et al.*, 2009). More recently, a gene family analysis based on the proteomes of 26 alveolates, revealed that members of the Apicomplexa share ~2200 orthologous groups of genes with all other alveolates, while only 81 groups are associated with the emergence of parasitism. This suggests that the genome of free-living ancestor of apicomplexans already encoded much of the machinery that may have been adapted to support a parasitic life style. As with the earlier study, numerous lineage-specific gene gains and losses were identified, many associated with host-parasite interactions. For example, the divergence of the coccidian lineage was associated with 537 losses and 414 gains (Woo *et al.*, 2015). In a later section we discuss some of the functions associated with these changes.

Jumbled genomes - missing synteny, abundance of low complexity regions

Comparisons of gene orderings reveal blocks of synteny between homologous genes in closely related apicomplexans. Notably, such relationships do not extend across the entire phylum (DeBarry and Kissinger, 2011) although one study did find that gene composition (not ordering) was significantly conserved between *Cryptosporidium* and *Plasmodium* species (Mazurie *et al.*, 2013). Further, syntenic relationships are typically restricted within a lineage, one exception occurring within the Aconoidasida where ~1,300 orthologs shared between Plasmodium and Theileria parasites, are distributed across 435 microsyntenic regions (Pain *et al.*, 2005). Within a lineage, different clades show varying degrees of synteny (i.e. blocks of collinear genes). For example, *T. gondii* shares synteny with other tissue-cyst forming coccidians (Reid *et al.*, 2012, Walzer *et al.*, 2013), albeit more restricted with *Sarcocystis neurona* (Blazejewski *et al.*, 2015), but not with the non tissue-cyst forming coccidians *Eimeria spp.* (Lorenzi *et al.*, 2016b). Within a genus, species show significant

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

synteny, in proportion to the phylogenetic separation between the species (Carlton *et al.*, 2008, Reid *et al.*, 2014). Interestingly sites of syntenic breaks within a genus, both at telomeres and other regions, appear enriched for species-specific genes and multi-copy gene families and depleted for core genes (DeBarry and Kissinger, 2011). Many of the families found at these sites are implicated in host-parasite interactions, which is elaborated below (Reid, 2015).

Contributing to this lack of synteny, as well as the large variations in genome sizes, is the presence of repetitive sequence, both within and between coding regions (Battistuzzi *et al.*, 2016). For *S. neurona*, the presence of repetitive sequence has resulted in the doubling of its genome size relative to other coccidians (Reid *et al.*, 2014). Repetitive elements include simple repeats (duplication of simple sets of nucleotides), tandem repeats (duplications of complex sets of nucleotides – typically found at telomeres and centromeres of chromosomes) and transposable elements (TEs, sequences which can change position within a genome) (Kapitonov and Jurka, 2008). TEs are classed into two types: Class I (retrotransposons), which operate via a copy-and-paste mechanism and include Long- and Short-Interspersed Nuclear Elements (LINEs and SINEs respectively), and Class II (DNA transposons), which operate via a cut-and-paste mechanism. The abundance and distribution of each class of repeat varies between genomes. For example, in *P. falciparum*, simple repeats predominate, representing ~14% of its genome, while for *E. tenella*, simple repeats and LINEs make up ~13% and 11.5% of its genome respectively and for *S. neurona*, its ~130 Mbp genome features a high proportion of DNA transposons (11.5%), LINEs (7.7%) and simple repeats (6%) (Blazejewski *et al.*, 2015). In terms of distribution, for *Eimeria. spp.*, around 50% of the genes are found in repeat-rich regions, resulting in proteins enriched in homopolymeric amino acid repeats (Reid *et al.*, 2014). In *S. neurona*, only Class I transposons are distributed in exons and Class II transposons are almost exclusively located in introns and intergenic regions. Intriguingly, although evolutionary analyses suggests these transposons are no longer active, their continued maintenance within the *S. neurona* genome, relative to other coccidians, suggests they may play some functional role (Blazejewski *et al.*, 2015). In *P. falciparum*, 35% of the genes encode homopolymeric repeats, most of which are asparagine-rich, spread across most protein families (Singh *et al.*, 2004). Although the functional role of these repeats is unclear, studies in *Eimeria* and *P. falciparum* suggest they are located mostly in loop regions and away from sites involved in domain-domain interactions and active sites and are therefore unlikely to interfere with protein structure or function.

Invasion machinery: Apicomplexan-specific conserved mechanism of host invasion

Host cell invasion is a rapid and complex process that relies on an orchestrated cascade of interactions between invading parasite and host cell. To orchestrate these processes, apicomplexans have evolved hundreds of specialised invasion proteins, many transported into exocytic organelles (micronemes, rhoptries and dense granules) that occupy up to one third of the cell volume (Figure 1a). The discharge of these apical organelles marks the beginning of host cell penetration and occurs in a tightly coordinated program (Sibley, 2011,

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Sharma and Chitnis, 2013). First, parasite surface receptors (e.g. SAG1-related sequences (*SRS*) in *Toxoplasma* / 6-Cysteine s48/45 family of proteins in *Plasmodium* (Arredondo *et al.*, 2012)) initiate host cell recognition and attachment. This is followed by secretion of the microneme proteins that strengthen host cell attachment and play a major role in the formation of the so-called "moving junction" which forms a specific interface to facilitate invasion. Formation of this junction relies on a set of conserved parasite proteins – rhoptry neck protein 2 (RON2) and apical membrane antigen 1 (AMA1) (Lamarque *et al.*, 2011, Tonkin *et al.*, 2011). In *Plasmodium*, only the canonical RON2-AMA1 interaction has been characterised (Srinivasan *et al.*, 2011) whereas three additional paralogs of AMA1 and two paralogs of RON2 are reported in *Toxoplasma,* expressed in a stage-specific manner (Poukchanski *et al.*, 2013, Lamarque *et al.*, 2014). Structural analysis of the different RON2-AMA1 pairs identify a pair with substantially divergent structure and an atypical mechanism, revealing molecular diversity at parasite host-cell interface, likely relevant for stage-specific changes (Parker *et al.*, 2016). Next, the parasite discharges the contents of the rhoptries, which are released into the host cell (Kemp *et al.*, 2013). A subset of rhoptry neck proteins (RONs) are critical for forming the moving junction (Proellocks *et al.*, 2010), while other rhoptry proteins (ROPs) (Counihan *et al.*, 2013), together with dense granule (GRA) proteins (Mercier *et al.*, 2005, Mercier and Cesbron-Delauw, 2015), interact with host cell targets, manipulating pathways to protect the (now) intracellular parasite against clearance (*Toxoplasma*), or extensively remodelling the host cell by altering its mechanical properties (*Plasmodium* in RBCs) (Tiburcio *et al.*, 2015). Many rhoptry proteins are secreted into the host cell during the process of invasion (Lim *et al.*, 2012) whereas several dense granule proteins are secreted after PVM formation. Providing the parasite with the motile force to gain entry to the host cell is the inner-membrane complex (IMC). The IMC acts as a scaffold through which the glideosome complex (GAP40/GAP45/GAP50) exploits actomyosin (actin/MyoA/MLC) to provide the driving force for anterograde movement (Kono *et al.*, 2013, Bargieri *et al.*, 2014, Harding and Meissner, 2014, Tardieux and Baum, 2016). However, recent studies show that host invasion can occur even when gliding motility is blocked, suggesting a relook at the existing mechanism (or) the presence of alternate mechanisms (Andenmatten *et al.*, 2013, Meissner *et al.*, 2013). Although the host cell invasion process is largely conserved across apicomplexans, details vary among the different species and different stages. *Theileria* has a distinct mechanism and can enter the host cell in any orientation (unlike the others which enter via the apical complex), independent of parasite and host cell actin, using a zippering mechanism (Shaw, 2003). *Plasmodium* and *Toxoplasma* modulate the host cell cytoskeleton, whereas Cryptosporidium recruits host cell sugar transporters in addition to modulating the actin cytoskeleton (Chen *et al.*, 2005, O'Hara and Chen, 2011). Further, after the parasite enters host cell, virulent strains of *Toxoplasma* use proteins secreted from rhoptries for modulating the host cell (mice) IRG resistance GTPases to protect them from clearance (Khaminets *et al.*, 2010), whereas these factors do not protect the mice against *Plasmodium* (Liesenfeld *et al.*, 2011).

Genome comparisons reveal complex evolutionary patterns associated with the emergence of these invasion related proteins. For example, amongst the 30 IMC proteins identified to date is a common core of highly conserved proteins, representing the recruitment and subsequent diversification of ancient eukaryotic proteins, supplemented with many lineage-

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

specific proteins (Table 1). These latter genes, which include the *Plasmodium* specific PF3D7_1345600 (MAL13P1.228), may provide the IMC with additional lineage-specific roles such as scaffolding during gametocytogenesis (Kono *et al.*, 2012). Similarly, several attachment proteins secreted by the micronemes and rhoptries are also well conserved across the phylum, including the microneme protein, AMA1, and its interaction complex (RON2/4/5), responsible for binding the parasite to the host cell surface (Boucher and Bosch, 2015), while others such as the *Plasmodium*-specific – Duffy binding proteins which bind to Duffy antigen on erythrocyte surface (eg. PvDBP1 and PvDBP2 from *P. vivax*) and reticulocyte binding proteins which enable selective invasion of reticulocytes (RBP; eg. PvRBP1, PvRBP2 from *P. vivax)*, are restricted to distinct lineages. Even when conserved, microneme proteins can exhibit distinct ligand binding specificities even between closely related species (Carruthers and Tomley, 2008). For example, the *T. gondii* and *N. caninum* orthologs of MIC1 (TgMIC1 and NcMIC1) bind sialic acid and glycosaminoglycans respectively; while, unlike TgMIC4, NcMIC4 can bind lactose.

In contrast to proteins mediating cell attachment and host cell entry, those proteins involved in manipulating host pathways tend to be less well conserved, with many families of proteins specific to individual lineages, reflecting divergent life cycle strategies (Kemp *et al.*, 2013). For example, as the parasite enters the cell, the release of rhoptry and dense granule proteins results in the formation and subsequent modification of the parasitophorous vacuole. *Babesia* and *Theileria*, which lack this compartment, also lack many of these genes (Lingelbach and Joiner, 1998). Cryptosporidium is enclosed in a unique parasitophorous vacuole –like structure, thought to derive through extension of the host cell microvillus membrane (Clode *et al.*, 2015). This separates the parasite from both the host cell cytoplasm and gut environment, which may explain the relatively low number of host associated genes. In the next section we provide further details on these genes and pathways.

Host cell modulation: Lineage- and species- specific gene families modulating host-specific adaptations

Once inside the host cell, further suites of protein effectors are secreted into the host cytosol to interact with host cell targets, manipulating pathways and optimizing nutrient acquisition (Gubbels and Duraisingh, 2012). In *Plasmodium*, secreted proteins can be divided into 2 categories: a) those characterised by a PEXEL (plasmodium export element) or HTS (host-targeting signal) motif (Hiller *et al.*, 2004, Marti *et al.*, 2004) at the N-terminus, cleaved by plasmepsin V, for dense granule targeting (Russo *et al.*, 2010) b) PEXEL negative exported proteins, which share an internal transmembrane segment (Marti and Spielmann, 2013, de Koning-Ward *et al.*, 2016). The exported proteins are translocated across PVM into the host cell via the Protein Translocon of Exported Proteins (PTEX) complex (de Koning-Ward *et al.*, 2009, Elsworth *et al.*, 2014) in *P. falciparum*, exporting hundreds of proteins across PVM into the host cell (Bullen *et al.*, 2012). Although a PEXEL-like sequence motif has been characterised for dense granule targeting in *Toxoplasma*, it is not involved in protein export into the host cell (Hsiao *et al.*, 2013). In *Toxoplasma,* dense granule proteins GRA17 and GRA23 have been identified to mediate movement of small molecules across PVM (Gold *et al.*, 2015). An in-silico comparative analysis of protein secretion and effectors has identified

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

PEXEL-like motifs in *Babesia* and *Cryptosporidium* and plasmepsin V orthologues in all lineages, whereas translocon components are restricted to *Plasmodium* (Pelle *et al.*, 2015), supporting the experimental results. Clustering the set of predicted secreted proteins using TribeMCL identifies 331 families (Pelle *et al.*, 2015). These effectors tend to be associated with large, lineage-specific, variant gene families, which are sequentially diverse, rapidly evolving, and responsible for species-specific lifestyles (Table 1; (Reid, 2015)). Aside from the presence of signal peptides, little homology is observed between these families. Amongst these families are those involved in sequestration and antigenic variation (e.g. var - *Plasmodium* and ves - *Babesia*), others are involved in modulating the host immune response (e.g. Sag1-related sequences (SRS) – tissue cyst forming coccidia), and several others in modulating the host cell (ROP kinases - Coccidia, FIKK - *Plasmodium* and SVSP - *Theileria*) (Ward *et al.*, 2004, Schneider and Mercereau-Puijalon, 2005, Schmuckli-Maurer *et al.*, 2009, Sibley *et al.*, 2009, Lim *et al.*, 2012, Wei *et al.*, 2013), in order to protect the parasite and enable its growth. For example, ROP5 and ROP18 modulate host immunity related GTPases (Behnke *et al.*, 2012), while ROP16 modifies host cell signalling and ROP38 can inhibit host cell transcription (Kemp *et al.*, 2013). Even where families may be conserved, their complements can vary dramatically across species and even strains, for example the coccidians, *N. caninum* and *S. neurona* feature 227 and 23 members of the SRS proteins respectively, while *T. gondii* strains Me49 and GT1 feature 109 and 90 SRS proteins respectively (Wasmuth *et al.*, 2012). Similarly, the genomes of *T. parva and T. annulata* encode 85 and 51 SVSP proteins respectively.

Differences in the complements of these proteins relate to distinct life cycle requirements. For example, SRS proteins are composed of signature 6-cysteine domains, previously classified into one of 8 families (Wasmuth *et al.*, 2012). *S. neurona* which features a reduced complement of SRS proteins relative to other tissue-cyst forming coccidians, nonetheless feature an expansion of SRS proteins with family 2 domains, previously associated with cyst wall integrity (Tomita *et al.*, 2013). This raises the possibility that the emergence of this family was critical for the formation of cysts enabling the transition from a monoxenic to a heteroxenic life cycle. The subsequent expansion of SRS proteins through tandem duplication, particularly those composed of two domains featuring family 7 and family 8 and implicated in host immune modulation, may have subsequently allowed the parasite to broaden its host range. Coccidians also display significant variation in ROP kinases (*E. tenella* – 27; *S. neurona* - 15; *N. caninum* - 44; and *T. gondii* - 55; (Talevich and Kannan, 2013, Blazejewski *et al.*, 2015)). Further, while the tissue-cyst forming coccidians share many ROP kinase orthologs, most members in *E. tenella* lack orthologs in the other species, suggesting unique functions, perhaps associated with its purely enteric lifestyle (Blazejewski *et al.*, 2015).

The mechanisms underlying the expansions of many of these families vary across lineages and can be related to their genomic context. By promoting recombination, the localization of *var, rif* and *stevor* multigene families to subtelomeric regions is thought to be responsible for the generation of the genetic diversity driving antigenic variation in *Plasmodium* (Scherf *et al.*, 1998). Interestingly, for *P. knowlesi*, whereas multigene families are not associated with subtelomeric regions, the pir and *SICAvar* gene families, involved in antigenic variation, are nevertheless associated with telomere-like repeats suggested to play a role in recombination

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Pain *et al.*, 2008). On the other hand, the coccidian families of surface antigens, *srs* and *sag* are scattered throughout the genome, mostly within tandem arrays that likely arose through gene duplication and subsequent gene conversion (Reid *et al.*, 2012, Wasmuth *et al.*, 2012). Notably the genomes of *Cryptosporidium spp.* do not encode any large gene families, consistent with studies indicating a lack of antigenic variation (Widmer and Sullivan, 2012).

Transcriptional regulation in apicomplexa: the AP2 gene family

Although apicomplexans possess RNA-polymerase associated factors and basal transcription factors, they lack many specific transcription factors (TFs) observed in other eukaryotes (Coulson *et al.*, 2004, Callebaut *et al.*, 2005). Instead, to support complex developmental cycles, potentially featuring multiple hosts, apicomplexans rely on a phylum-specific expansion of novel TFs that feature a version of the AP2 (Apetala2)-integrase DNA binding domain (Balaji *et al.*, 2005). Although AP2 TFs are abundant in all apicomplexans, only 4 are shared across most apicomplexan clades, with the rest representing lineage-specific expansions (Table 1; (Woo *et al.*, 2015)). Tissue-cyst coccidians have the largest complement with 53 AP2 TFs, followed by *Plasmodium* species (~25) and piroplasmids (~20); monoxenic species feature smaller complements (*Cryptosporidium spp.* – 9; *Eimeria tenella* – 15). The lower incidence of AP2 TFs in *Cryptosporidium spp.* has been suggested to be offset by the presence of other TFs (Oberstaller *et al.*, 2014). Studies in *P. falciparum* have revealed that AP2 TFs have unique binding preferences; possessing high affinity primary binding motifs as well as secondary binding motifs (De Silva *et al.*, 2008, Campbell *et al.*, 2010). The ability to bind multiple, distinct motifs significantly increases the potential complexity of the transcriptional regulatory networks governed by this family. Several studies are unravelling the roles of AP2 TFs different in aspects of their life cycle and development (Painter *et al.*, 2011), including chromosome biology (Flueck *et al.*, 2010), commitment to gametocytogenesis (Kafsack *et al.*, 2014, Sinha *et al.*, 2014), normal morphogenesis inside mosquito (Kaneko *et al.*, 2015) in *Plasmodium*, and parasite virulence, host invasion (Walker *et al.*, 2013), and tissue cyst development (Radke *et al.*, 2013) in *Toxoplasma*.

Apicomplexan metabolism: Lineage-specific adaptations to host environment

It is well-established that variations in metabolic potential help govern a pathogen's ability to colonize, persist and establish virulence within infected hosts (McKinney *et al.*, 2000, Song *et al.*, 2013, Xia *et al.*, 2013). With the availability of genome sequences, there is mounting interest in the use of genome-scale metabolic reconstructions to identify critical pathways required for growth. Such reconstructions rely on accurately annotating enzymes from sequence data alone, for which several tools are available (Claudel-Renard *et al.*, 2003, Arakaki *et al.*, 2006, Moriya *et al.*, 2007, Hung *et al.*, 2010). However, initial reconstructions based on these methods alone are typically incomplete, with many reactions, even in essential pathways, missing, due to an inability to detect the gene responsible. For apicomplexans, these 'pathway holes' constitute a significant component; in some reconstructions accounting for 40% of the reactions (Pinney *et al.*, 2007). Consequently,

Author Manuscript

methods have been developed to fill these gaps through experimental- and literature-based evidence (Schomburg *et al.*, 2013, Shanmugasundram *et al.*, 2013), comparative genomics approaches (Lee and Sonnhammer, 2003, Chen and Vitkup, 2006, Suhre, 2007, Zhou *et al.*, 2008), as well as integrative approaches (Dale *et al.*, 2010, Hung and Parkinson, 2011, Benedict *et al.*, 2014).

A comparative picture of the metabolic potential of different apicomplexan species is provided in Figure 2. Of the 391 enzyme categories (ECs) predicted from genomes of 18 apicomplexan species using an integrative annotation strategy (Hung and Parkinson, 2011), only 147 are shared across all clades (Figure 2a). Among the major clades, *Cryptosporidium spp.* and Piroplasmida have the most streamlined metabolism, with 204 and 213 enzymes (as defined by distinct enzyme commission (EC) numbers) respectively, whereas *Plasmodium* and Coccidia have a much larger metabolic repertoire, comprising 267 and 366 enzymes respectively, with Coccidia featuring 85 unique enzymes. *Cryptosporidium* and Piroplasmida appear to have lost the most enzymes relative to other apicomplexans (52 and 19 respectively), reflective of streamlined metabolisms with many nutrients scavenged from their hosts (Mogi and Kita, 2010). Species within a lineage largely share the same metabolic repertoire; for example, *C. muris* (199 enzymes), *C. parvum* (181 enzymes), *C. hominis* (175 enzymes) share 170 common enzymes; *B. bovis* (204 enzymes), *T. parva* (203 enzymes), *T. annulata* (203 enzymes) share 200 common enzymes; *P. knowlesi* and *P. vivax* share an identical metabolic complement (265 enzymes), with minor variations from other Plasmodium species (261 to 267 enzymes); the tissue-cyst coccidians *S. neurona* (305 enzymes), *N. caninum* (356 enzymes), *T. gondii* (354 enzymes) and *H. hammondi* (354 enzymes) share a common complement of 300 enzymes (254 with *E. tenella*). Most of the clade-specific enzymes are associated with core pathways (Figure 2b). For example, the 8 *Plasmodium*-specific enzymes are associated with arginine-proline metabolism and thiamine biosynthesis, while the 11 *Cryptosporidium*-specific enzymes are found in pyrimidine, amino sugar, and starch and sucrose metabolism, reflecting a previous analysis that showed different lineages may have acquired different sets of enzymes to perform similar core metabolic functions (Hung and Parkinson, 2011). As noted above, coccidians have the largest complement of metabolic enzymes, many of which are clade-specific and associated with core pathways dominated by carbohydrate and amino acid metabolism. This suggests that coccidians have evolved a diverse metabolic repertoire to adapt to multiple environments including, for example, enzymes driving gluconeogenesis that provides carbohydrate reserves to allow oocysts to sporulate outside the host (Ginger, 2006).

While most metabolic functionality occurs in the cytoplasm, several pathways are partitioned to mitochondria and the apicoplast, a plastid of secondary endosymbiotic origin (Sheiner *et al.*, 2013, McFadden and Yeh, 2016). Amongst the enzymes in the apicoplast are those involved in type II fatty acid biosynthesis (Shears *et al.*, 2015), isoprenoid synthesis (providing cofactors for electron transport chain and glycoprotein synthesis), heme biosynthesis and the formation of iron-sulfur clusters. Recently, isoprenoid precursor biosynthesis by apicoplast has been shown to be essential for normal gametocytogenesis in *P. falciparum* (Wiley *et al.*, 2015). Further, this has also been identified to be the only essential role during Plasmodium erythrocyte stages, with type II fatty acid biosynthesis pathway being dispensable. However, this pathway appears to be essential during

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Plasmodium liver stages and in *Toxoplasma* (Sheiner *et al.*, 2013), suggesting that species-specific and host cell-type specific differences exist. While less than 50 genes (mainly involved in transcription and translation), are encoded by the plastid's own genome (Lim and McFadden, 2010), the remaining derive from nuclear encoded genes (predictions of apicoplast targeted nuclear-encoded genes is ~8–10% for *Plasmodium* and *Theilera* and only ~1% for *Babesia* (Sato, 2011)) and are targeted to the organelle through specific signal peptides. The apicoplast forms a tight physical association with mitochondria, attributed to metabolic dependencies; isoprenoid precursors generated by the apicoplast form the basis of ubiquinones driving the electron transport chain in mitochondria, while acetyl CoA, generated by mitochondria, is exploited as a major carbon source for fatty acid synthesis in the apicoplast (Sheiner *et al.*, 2013). Furthermore, the apicoplast and mitochondrion share components of the haem biosynthetic pathway, which commences in the mitochondrion and proceeds in the apicoplast, before completing in the mitochondrion (Koreny *et al.*, 2013). With highly reduced mitochondria (mitosomes), *C. parvum* and *C. hominis* lack a working tricarboxylic acid (TCA) cycle and rely instead on glycolysis for energy production (Henriquez *et al.*, 2005). Interestingly, the mitosome of the related parasite *C. muris* appears less reduced and features a functional TCA cycle (Henriquez *et al.*, 2005, Mogi and Kita, 2010). Further, lacking an apicoplast, *Cryptosporidium spp.* also rely on type I fatty acid metabolic pathways, present in species forming oocysts shed into the environment (*Cryptosporidium*, *Toxoplasma*, *Eimeria*), for *de novo* biosynthesis. *Theileria spp.* also show a reduced dependence on apicoplast and higher dependence of host for many substrates, again lacking the enzymes required for type II fatty acid biosynthesis (Gardner *et al.*, 2005).

Due to their essential role in growth and survival, many metabolic pathways have been targeted for drug development, including shikimate pathway (McConkey *et al.*, 2004), fatty acid metabolism (Goodman and McFadden, 2008, Shears *et al.*, 2015), and isoprenoid biosynthesis (Moreno and Li, 2008). However, key to these strategies is identifying those enzymes that mediate the most critical roles in these pathways. In the next section we describe how modeling has contributed to an improved understanding of metabolic function in the Apicomplexa.

Modeling metabolism in the apicomplexa

With the increasing availability of high quality metabolic reconstructions, a variety of modeling approaches have been developed to gain insights into the roles of individual enzymes and pathways in parasite growth. Among the more robust approaches are constraint based models such as flux balance analysis (FBA) which rely on stoichiometric relationships in reactions rather than explicit kinetic data (Figure 3b; (Lee *et al.*, 2006, Raman and Chandra, 2009)). FBA operates by calculating sets of fluxes across a metabolic network that optimizes a specific objective (e.g. maximizing growth rate of the parasite; (Hung and Parkinson, 2011)). FBA has so far been applied to four apicomplexans: *C. hominis*, *P. falciparum*, *T. gondii,* and *S. neurona*. The first was for *C. hominis* and comprised an analysis of 231 genes involved in 540 reactions (Vanee *et al.*, 2010). The model was integrated with proteomics data from the sporozoite and oocyst stages, to predict the importance of the differential expression of high- and low affinity hexokinases in oocysts (associated with glucose-limited environments outside the host), and sporozoites (associated

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

with the glucose-rich environment within the host) respectively. For *P. falciparum,* three models have been generated (Huthmacher *et al.*, 2010, Plata *et al.*, 2010, Fang *et al.*, 2014). The first featured a compartmentalized network of 366 genes and 1001 reactions; FBA predictions showed high correlation with gene knockout data and drug inhibition assays, and predicted 40 novel drug targets that lacked significant sequence similarity with human sequences (Plata *et al.*, 2010). Four of these are associated with shikimate biosynthesis, three with ubiquinone metabolism, and one with nicotinamide metabolism (nicotinate nucleotide adenylyltransferase). The last enzyme was experimentally validated to be a potentially effective and druggable target using drug inhibition experiments. In a separate model of 376 genes and 1019 reactions, Fang and colleagues used time-dependent gene expression, to explore stage-specific growth across the intraerythrocytic development cycle (IDC) of the parasite, helping link specific metabolites to corresponding physiological functions, such as the likely role of coenzyme A in late-IDC DNA replication and cell division (Fang *et al.*, 2014). The model also captures the contribution of different energy producing pathways to ATP production in the IDC, with the bulk generated from glycolysis.

The first metabolic model for *T. gondii* was a curated version consisting of 382 genes involving 571 reactions, separated across 5 subcellular compartments (Song *et al.*, 2013). Applying reaction constraints based on gene expression data, FBA revealed that strain-specific differences in growth rates are driven by differing capacities for energy production, highlighting the potential of metabolism to impact the parasite's virulence. Further, the model predicted strain-specific differences in drug susceptibilities, which were subsequently validated through drug inhibition studies. A later study, involving the semi-automated curation of 527 genes involved in 867 reactions, predicted two enzymes involved in acetyl CoA biosynthesis, ATP-citrate lyase and acetyl-CoA synthase, to be functionally redundant, with their simultaneous knockout to be lethal; a finding that was also confirmed experimentally (Tymoshenko *et al.*, 2015). Lastly, a model of *S. neurona* comprising 330 genes and 536 reactions, predicted the parasite to be more sensitive to *in silico* knockouts of enzymes from glycolysis and TCA cycle (Blazejewski *et al.*, 2015). However, the presence of alpha-glucosidase, suggested *S. neurona* may exploit fructose and sucrose as alternate energy sources to glucose, offering the potential for the parasite to rapidly adapt to new hosts. These studies demonstrate the potential of modeling to capture the metabolic complexity of apicomplexan life cycles and drive the generation of new testable hypotheses.

Systems biology for the apicomplexa: Beyond metabolism

Moving beyond genomic analyses, with the recognition that genes and their protein products do not operate in isolation but form parts of complex biological systems, there has been increasing interest in applying systems-based approaches to the study of parasite processes (Figure 3). With well characterized pathway relationships, metabolism has naturally been at the forefront of such studies. However advances in transcriptomics and proteomics are opening up new opportunities for systems analyses based on co-expression, phosphorylation and co-elution profiles.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Co-expression networks

Extensive transcriptional datasets are now available for several apicomplexans (Figure 1c) and provide unique opportunities to gain functional insights on the basis of co-expression profiles (Figure 3; (Le Roch *et al.*, 2003)). In brief, for each gene a profile of expression is created from the multiple transcriptome datasets that have been generated. Pairwise comparisons between these profiles for each pair of genes then yields a matrix of correlation scores (e.g. Pearson correlation coefficient; (Stuart *et al.*, 2003)). This matrix can then be represented as a network in which nodes (genes) are connected by edges if they exhibit a correlation score above a specified threshold (Figure 3a). The nodes in the resulting network can then be clustered, on the basis of shared patterns of interactions (Zhang and Horvath, 2005, Horvath and Dong, 2008), to define groups or modules of co-expressed genes representing functionally related genes (e.g. members of the same pathway). Applied to datasets examining the impact of 20 chemical compounds on gene expression in *P. falciparum* (Hu *et al.*, 2010), a co-expression network approach was used to identify many functionally related genes sharing similar expression profiles, suggesting shared regulatory mechanisms. For example, 31 of 42 proteins predicted to be part of the invasion process were experimentally observed to be localized to apical organelles (20 proteins), parasite periphery (4 proteins) or IMC (7 proteins). Further, on the basis of the function of their neighbours within the co-expression network, three proteins: PF3D7_1345600 (MAL13P1.228), PF3D7_1460600 (PF14_0578) and PF3D7_0522600 (PFE1130w) were identified as novel members of the IMC. More recently, the analysis of 2 expression datasets associated with *T. gondii* encompassing 22 time points revealed two distinct sub-networks of invasion related genes (Blazejewski *et al.*, 2015). The first composed of genes encoding micronemal proteins, which drive host-cell attachment and formation of the moving junction; the second composed of genes encoding rhoptry proteins, largely associated with modulation of host pathways. The dense connections within these networks illustrate the tight co-ordination in timing of expression associated with these genes.

Phospho-proteome networks

Protein phosphorylation is one of the most ubiquitously used post-translational mechanisms for regulation inside a cell. Recent advances in phosphopeptide enrichment and mass-spectrometric techniques have made it possible to study protein phosphorylation from a global perspective (Figure 3c; (Villen and Gygi, 2008)). A comparative study of both intracellular and extracellular forms of the invasive stages of *P. falciparum* (schizont) and *T. gondii* (tachyzoite), revealed 5,000 and 10,000 new phosphorylation sites respectively, including, unexpectedly, tyrosine residues (Treeck *et al.*, 2011). The study further revealed that many parasite proteins are phosphorylated after secretion into the host cell, indicating novel routes for regulation of host pathways. More recent phospho-proteome studies in *T. gondi* and *P. falciparum*, focusing on calcium dependent protein kinase 3 and protein kinase G, are also beginning to expand on their respective roles in parasite egress by identifying novel phosphorylation sites in protein targets (Brochet *et al.*, 2014, Treeck *et al.*, 2014, Alam *et al.*, 2015). Phospho-proteome analyses add an additional dimension to apicomplexan biology, allowing researchers to examine the impact of post-translational modifications on stage-specific development as well as host-parasite interactions, and offering additional routes for therapeutic intervention.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Protein-protein interaction networks

The generation of large-scale protein-protein interaction datasets are proving revolutionary for understanding the organization of complex biological processes (Butland *et al.*, 2005, Krogan *et al.*, 2006, Hu *et al.*, 2009). In addition to aiding annotation efforts, such networks may be usefully exploited to identify proteins mediating critical roles in organization of complexes highlighting their potential for therapeutic intervention. To date, *P. falciparum* is the only apicomplexan for which protein-interaction data has been generated on a genome scale (LaCount *et al.*, 2005). Applying a yeast-two hybrid approach, a network of ~2800 interactions between 1267 proteins was generated. Due to challenges in the expression of *P. falciparum* genes in a heterologous system, attempts have been made to improve on the quality of this initial dataset through integration of additional functional datasets (Hu *et al.*, 2010, Ramaprasad *et al.*, 2012); however such studies tend to be limited to functional, rather than physical interactions. Instead, an alternative strategy based on protein co-elution offers promise for deriving information on physical associations (Figure 3d). Avoiding technical challenges that arise with approaches such as tandem affinity purification, co-elution profiling has been successfully applied to soluble human complexes, (Havugimana *et al.*, 2012), complexes conserved across 9 metazoans (Wan *et al.*, 2015) and, significantly, the kinetoplastid, *Trypanosoma brucei* (Gazestani *et al.*, 2016). These studies provide a valuable framework to generate similar protein-protein interaction networks for apicomplexans.

From genomes to populations

With falling costs in genome sequencing, attention is now focusing on examining strain level differences to gain insights into the emergence of genetic variation that impacts virulence and the emergence of resistance (Ellegren, 2014). An initial study of 16 geographically diverse isolates of *P. falciparum* revealed that while genes encoding housekeeping functions such as metabolic enzymes exhibited little variation, those encoding surface functions such as cytoadherence and antigenic variation displayed a rich diversity in sequence (Volkman *et al.*, 2007). A recent study of 182 clinical isolates of *P. vivax* also revealed that the genes exhibiting the most variation are antigenic and involved in immune evasion, and additionally revealed a richer diversity than *P. falciparum* indicating larger and/or more stable population (Hupalo *et al.*, 2016). More concerning, this study also revealed signals of recent evolution in response to antimalarial drug exposure. Similar concerns arose in a targeted analysis focusing on artemisinin resistance across 825 *P. falciparum* isolates (Miotto *et al.*, 2013). In addition to revealing an unusual population structure associated with isolates from Western Cambodia, a hub of artemisinin resistance, artemisinin resistance was associated with multiple SNPs in *kelch13*, resulting in slow parasite clearance (Ariey *et al.*, 2014). Genome-wide association studies also showed that nonsynonymous polymorphisms in genes encoding ferredoxin, apicoplast ribosomal protein S10, multidrug resistance protein 2, and the chloroquine resistance transporter were associated with artemisinin resistance and may act as predisposing factors, allowing the emergence of *kelch13* variations, and thus serving as risk markers for new resistance-causing mutations (Miotto *et al.*, 2015).

Population genomics studies have also been applied to *T. gondii* isolates to reveal 6 major clades based on specific gene markers, with three major types (designated type I, type II and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

type III) dominating in North America and Europe (Su *et al.*, 2012). Analysis of SNP distributions across 10 isolates, further reveal the presence of haploblocks indicating the significant influence of recombination and admixture on the global population structure of *T. gondii* (Minot *et al.*, 2012). Recently, a larger scale analysis of 64 *T. gondii* isolates, was able to recapitulate previously defined haplogroups (Lorenzi *et al.*, 2016b). Through the application of chromosome painting in which chromosomal segments are coloured according to ancestry (Yahara *et al.*, 2013), this study also revealed the extent to which shared inheritance of haploblocks shapes population structure. Further examination of these haploblocks revealed conserved regions enriched for specific secretory pathogenesis determinants (proteins involved in parasite-host interactions, host invasion and modulation), which undergo tandem amplifications and diversification, likely influencing host range and pathogenicity. Future population based studies offer the potential to associate phenotypic traits with specific genetic loci based on shared patterns of haploblock inheritance.

Conclusions and Future perspectives

The availability and analysis of apicomplexan genomes and related datasets provides unprecedented views of their biology and emergence as a successful parasitic phylum. Apicomplexans are hypothesised to have evolved through the endosymbiosis involving a red algal and an alveolate ancestor (Janouskovec *et al.*, 2010). This event likely introduced vast genetic diversity into alveolate gene pool, driving the emergence of new species with diverse metabolic pathways and distinct life styles: parasitic apicomplexans, free-living ciliates and photosynthetic chromerids and dinoflagellates. Comparative genomic analyses suggest that the lineage-specific loss of components important for free-living lifestyle (metabolic pathways, endomembrane trafficking, flagellar apparatus for motility) happened progressively within the apicomplexan lineage (Woo *et al.*, 2015). At the same time, genes encoding proteins driving parasite-specific processes such as invasion and host modulation, were either repurposed from pre-existing components associated with the free-living ancestor (e.g. components of the acto-myosin complex or inner membrane complex) or later acquired during adaptation to specific host niches (e.g. AP2 transcription factors or host-modulation machineries) (Janouskovec and Keeling, 2016).

From a practical perspective, genome analyses are proving essential to our understanding of the emergence of drug resistance as well as in the identification of novel drug targets. Key to this endeavour are systems-based approaches, based for example on the analysis of protein-interaction or metabolic networks that facilitate the elucidation of protein function in the context of the complex and/or pathway in which it operates. Although systems-based approaches provide important insights that cannot be gained from a reductionist approach, they come with the caveat that they incorporate automated functional annotations, which are likely to be erroneous especially for proteins with no known homologues. This underscores the need for community-wide efforts to generate and integrate functional datasets on various aspects of apicomplexan biology with the systems-based models, in order to improve the predictive validity of these models. We also anticipate greater emphasis on population-based studies that move beyond understanding the factors that govern population structure to reveal the mechanisms that allow these parasites to acquire resistance to therapeutics, in addition to virulence factors that drive pathogenesis.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Acknowledgements

The authors gratefully acknowledge Nirvana Nursimulu for helpful suggestions.

References

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, 441–5. [PubMed: 15044751]

Adl SM, Leander BS, Simpson AG, Archibald JM, Anderson OR, Bass D, Bowser SS, Brugerolle G, Farmer MA, Karpov S, Kolisko M, Lane CE, Lodge DJ, Mann DG, Meisterfeld R, Mendoza L, Moestrup O, Mozley-Standridge SE, Smirnov AV & Spiegel F, 2007 Diversity, nomenclature, and taxonomy of protists. Syst Biol, 56, 684–9. [PubMed: 17661235]

Alam MM, Solyakov L, Bottrill AR, Flueck C, Siddiqui FA, Singh S, Mistry S, Viskaduraki M, Lee K, Hopp CS, Chitnis CE, Doerig C, Moon RW, Green JL, Holder AA, Baker DA & Tobin AB, 2015 Phosphoproteomics reveals malaria parasite Protein Kinase G as a signalling hub regulating egress and invasion. Nat Commun, 6, 7285. [PubMed: 26149123]

Andenmatten N, Egarter S, Jackson AJ, Jullien N, Herman JP & Meissner M, 2013 Conditional genome engineering in Toxoplasma gondii uncovers alternative invasion mechanisms. Nat Methods, 10, 125–7. [PubMed: 23263690]

Arakaki AK, Tian W & Skolnick J, 2006 High precision multi-genome scale reannotation of enzyme function by EFICAz. BMC Genomics, 7, 315. [PubMed: 17166279]

Ariey F, Witkowski B, Amaratunga C, Beghain J, Langlois AC, Khim N, Kim S, Duru V, Bouchier C, Ma L, Lim P, Leang R, Duong S, Sreng S, Suon S, Chuor CM, Bout DM, Menard S, Rogers WO, Genton B, Fandeur T, Miotto O, Ringwald P, Le Bras J, Berry A, Barale JC, Fairhurst RM, Benoit-Vical F, Mercereau-Puijalon O & Menard D, 2014 A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature, 505, 50–5. [PubMed: 24352242]

Arredondo SA, Cai M, Takayama Y, Macdonald NJ, Anderson DE, Aravind L, Clore GM & Miller LH, 2012 Structure of the Plasmodium 6-cysteine s48/45 domain. Proc Natl Acad Sci U S A, 109, 6692–7. [PubMed: 22493233]

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 Jr., Wang H & Warrenfeltz S, 2013 EuPathDB: the eukaryotic pathogen database. Nucleic Acids Res, 41, D684–91. [PubMed: 23175615]

Balaji S, Babu MM, Iyer LM & Aravind L, 2005 Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. Nucleic Acids Res, 33, 3994–4006. [PubMed: 16040597]

Bannai H, Tamada Y, Maruyama O, Nakai K & Miyano S, 2002 Extensive feature detection of N-terminal protein sorting signals. Bioinformatics, 18, 298–305. [PubMed: 11847077]

Bargieri D, Lagal V, Andenmatten N, Tardieux I, Meissner M & Menard R, 2014 Host cell invasion by apicomplexan parasites: the junction conundrum. PLoS Pathog, 10, e1004273. [PubMed: 25232721]

Barta JR & Thompson RC, 2006 What is Cryptosporidium? Reappraising its biology and phylogenetic affinities. Trends Parasitol, 22, 463–8. [PubMed: 16904941]

Battistuzzi FU, Schneider KA, Spencer MK, Fisher D, Chaudhry S & Escalante AA, 2016 Profiles of low complexity regions in Apicomplexa. BMC Evol Biol, 16, 47. [PubMed: 26923229]

Behnke MS, Fentress SJ, Mashayekhi M, Li LX, Taylor GA & Sibley LD, 2012 The polymorphic pseudokinase ROP5 controls virulence in Toxoplasma gondii by regulating the active kinase ROP18. PLoS Pathog, 8, e1002992. [PubMed: 23144612]

Benedict MN, Mundy MB, Henry CS, Chia N & Price ND, 2014 Likelihood-based gene annotations for gap filling and quality assessment in genome-scale metabolic models. PLoS Comput Biol, 10, e1003882. [PubMed: 25329157]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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.

Blunt DS, Khramtsov NV, Upton SJ & Montelone BA, 1997 Molecular karyotype analysis of Cryptosporidium parvum: evidence for eight chromosomes and a low-molecular-size molecule. Clin Diagn Lab Immunol, 4, 11–3. [PubMed: 9008274]

Bonnin A, Gut J, Dubremetz JF, Nelson RG & Camerlynck P, 1995 Monoclonal antibodies identify a subset of dense granules in Cryptosporidium parvum zoites and gamonts. J Eukaryot Microbiol, 42, 395–401. [PubMed: 7620464]

Boucher LE & Bosch J, 2015 The apicomplexan glideosome and adhesins - Structures and function. J Struct Biol, 190, 93–114. [PubMed: 25764948]

Brochet M, Collins MO, Smith TK, Thompson E, Sebastian S, Volkmann K, Schwach F, Chappell L, Gomes AR, Berriman M, Rayner JC, Baker DA, Choudhary J & Billker O, 2014 Phosphoinositide metabolism links cGMP-dependent protein kinase G to essential Ca(2)(+) signals at key decision points in the life cycle of malaria parasites. PLoS Biol, 12, e1001806. [PubMed: 24594931]

Bullen HE, Charnaud SC, Kalanon M, Riglar DT, Dekiwadia C, Kangwanrangsan N, Torii M, Tsuboi T, Baum J, Ralph SA, Cowman AF, De Koning-Ward TF, Crabb BS & Gilson PR, 2012 Biosynthesis, localization, and macromolecular arrangement of the Plasmodium falciparum translocon of exported proteins (PTEX). J Biol Chem, 287, 7871–84. [PubMed: 22253438]

Butland G, Peregrin-Alvarez JM, Li J, Yang W, Yang X, Canadien V, Starostine A, Richards D, Beattie B, Krogan N, Davey M, Parkinson J, Greenblatt J & Emili A, 2005 Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature, 433, 531–7. [PubMed: 15690043]

Callebaut I, Prat K, Meurice E, Mornon JP & Tomavo S, 2005 Prediction of the general transcription factors associated with RNA polymerase II in Plasmodium falciparum: conserved features and differences relative to other eukaryotes. BMC Genomics, 6, 100. [PubMed: 16042788]

Campbell TL, De Silva EK, Olszewski KL, Elemento O & Llinas M, 2010 Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. PLoS Pathog, 6, e1001165. [PubMed: 21060817]

Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P, Cheng Q, Coulson RM, Crabb BS, Del Portillo HA, Essien K, Feldblyum TV, Fernandez-Becerra C, Gilson PR, Gueye AH, Guo X, Kang'a S, Kooij TW, Korsinczky M, Meyer EV, 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, 757–63. [PubMed: 18843361]

Carruthers VB & Tomley FM, 2008 Microneme proteins in apicomplexans. Subcell Biochem, 47, 33–45. [PubMed: 18512339]

Cavalier-Smith T, 1993 Kingdom protozoa and its 18 phyla. Microbiol Rev, 57, 953–94. [PubMed: 8302218]

Checkley W, White AC Jr., Jaganath D, Arrowood MJ, Chalmers RM, Chen XM, Fayer R, Griffiths JK, Guerrant RL, Hedstrom L, Huston CD, Kotloff KL, Kang G, Mead JR, Miller M, Petri WA Jr., Priest JW, Roos DS, Striepen B, Thompson RC, Ward HD, Van Voorhis WA, Xiao L, Zhu G & Houpt ER, 2015 A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for cryptosporidium. Lancet Infect Dis, 15, 85–94. [PubMed: 25278220]

Chen L & Vitkup D, 2006 Predicting genes for orphan metabolic activities using phylogenetic profiles. Genome Biol, 7, R17. [PubMed: 16507154]

Chen XM, O'hara SP, Huang BQ, Splinter PL, Nelson JB & Larusso NF, 2005 Localized glucose and water influx facilitates Cryptosporidium parvum cellular invasion by means of modulation of host-cell membrane protrusion. Proc Natl Acad Sci U S A, 102, 6338–43. [PubMed: 15851691]

Cilingir G, Broschat SL & Lau AO, 2012 ApicoAP: the first computational model for identifying apicoplast-targeted proteins in multiple species of Apicomplexa. PLoS One, 7, e36598. [PubMed: 22574192]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Claudel-Renard C, Chevalet C, Faraut T & Kahn D, 2003 Enzyme-specific profiles for genome annotation: PRIAM. Nucleic Acids Res, 31, 6633–9. [PubMed: 14602924]

Clode PL, Koh WH & Thompson RC, 2015 Life without a Host Cell: What is Cryptosporidium? Trends Parasitol, 31, 614–24. [PubMed: 26440789]

Coulson RM, Hall N & Ouzounis CA, 2004 Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum. Genome Res, 14, 1548–54. [PubMed: 15256513]

Counihan NA, Kalanon M, Coppel RL & De Koning-Ward TF, 2013 Plasmodium rhoptry proteins: why order is important. Trends Parasitol, 29, 228–36. [PubMed: 23570755]

Cowper B, Matthews S & Tomley F, 2012 The molecular basis for the distinct host and tissue tropisms of coccidian parasites. Mol Biochem Parasitol, 186, 1–10. [PubMed: 22982139]

Culvenor JG, Day KP & Anders RF, 1991 Plasmodium falciparum ring-infected erythrocyte surface antigen is released from merozoite dense granules after erythrocyte invasion. Infect Immun, 59, 1183–7. [PubMed: 1997422]

Dale JM, Popescu L & Karp PD, 2010 Machine learning methods for metabolic pathway prediction. BMC Bioinformatics, 11, 15. [PubMed: 20064214]

De Koning-Ward TF, Dixon MW, Tilley L & Gilson PR, 2016 Plasmodium species: master renovators of their host cells. Nat Rev Microbiol, 14, 494–507. [PubMed: 27374802]

De Koning-Ward TF, Gilson PR, Boddey JA, Rug M, Smith BJ, Papenfuss AT, Sanders PR, Lundie RJ, Maier AG, Cowman AF & Crabb BS, 2009 A newly discovered protein export machine in malaria parasites. Nature, 459, 945–9. [PubMed: 19536257]

De Silva EK, Gehrke AR, Olszewski K, Leon I, Chahal JS, Bulyk ML & Llinas M, 2008 Specific DNA-binding by apicomplexan AP2 transcription factors. Proc Natl Acad Sci U S A, 105, 8393–8. [PubMed: 18541913]

Debarry JD & Kissinger JC, 2011 Jumbled genomes: missing Apicomplexan synteny. Mol Biol Evol, 28, 2855–71. [PubMed: 21504890]

Doliwa C, Escotte-Binet S, Aubert D, Sauvage V, Velard F, Schmid A & Villena I, 2013 Sulfadiazine resistance in Toxoplasma gondii: no involvement of overexpression or polymorphisms in genes of therapeutic targets and ABC transporters. Parasite, 20, 19. [PubMed: 23707894]

Ellegren H, 2014 Genome sequencing and population genomics in non-model organisms. Trends Ecol Evol, 29, 51–63. [PubMed: 24139972]

Elsworth B, Matthews K, Nie CQ, Kalanon M, Charnaud SC, Sanders PR, Chisholm SA, Counihan NA, Shaw PJ, Pino P, Chan JA, Azevedo MF, Rogerson SJ, Beeson JG, Crabb BS, Gilson PR & De Koning-Ward TF, 2014 PTEX is an essential nexus for protein export in malaria parasites. Nature, 511, 587–91. [PubMed: 25043043]

Fang X, Reifman J & Wallqvist A, 2014 Modeling metabolism and stage-specific growth of Plasmodium falciparum HB3 during the intraerythrocytic developmental cycle. Mol Biosyst, 10, 2526–37. [PubMed: 25001103]

Flegr J, Prandota J, Sovickova M & Israili ZH, 2014 Toxoplasmosis--a global threat. Correlation of latent toxoplasmosis with specific disease burden in a set of 88 countries. PLoS One, 9, e90203. [PubMed: 24662942]

Flueck C, Bartfai R, Niederwieser I, Witmer K, Alako BT, Moes S, Bozdech Z, Jenoe P, Stunnenberg HG & Voss TS, 2010 A major role for the Plasmodium falciparum ApiAP2 protein PfSIP2 in chromosome end biology. PLoS Pathog, 6, e1000784. [PubMed: 20195509]

Gardner MJ, Bishop R, Shah T, De Villiers EP, Carlton JM, Hall N, Ren Q, Paulsen IT, Pain A, Berriman M, Wilson RJ, 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 EL, 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, 134–7. [PubMed: 15994558]

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 MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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. [PubMed: 12368864]

Garza-Leon M & Garcia LA, 2012 Ocular toxoplasmosis: clinical characteristics in pediatric patients. Ocul Immunol Inflamm, 20, 130–8. [PubMed: 22409567]

Gazestani VH, Nikpour N, Mehta V, Najafabadi HS, Moshiri H, Jardim A & Salavati R, 2016 A Protein Complex Map of Trypanosoma brucei. PLoS Negl Trop Dis, 10, e0004533. [PubMed: 26991453]

Ginger ML, 2006 Niche metabolism in parasitic protozoa. Philos Trans R Soc Lond B Biol Sci, 361, 101–18. [PubMed: 16553311]

Ginsburg H & Abdel-Haleem AM, 2016 Malaria Parasite Metabolic Pathways (MPMP) Upgraded with Targeted Chemical Compounds. Trends Parasitol, 32, 7–9. [PubMed: 26530861]

Gohil S, Kats LM, Sturm A & Cooke BM, 2010 Recent insights into alteration of red blood cells by Babesia bovis: moovin' forward. Trends Parasitol, 26, 591–9. [PubMed: 20598944]

Gold DA, Kaplan AD, Lis A, Bett GC, Rosowski EE, Cirelli KM, Bougdour A, Sidik SM, Beck JR, Lourido S, Egea PF, Bradley PJ, Hakimi MA, Rasmusson RL & Saeij JP, 2015 The Toxoplasma Dense Granule Proteins GRA17 and GRA23 Mediate the Movement of Small Molecules between the Host and the Parasitophorous Vacuole. Cell Host Microbe, 17, 642–52. [PubMed: 25974303]

Goodman CD & Mcfadden GI, 2008 Fatty acid synthesis in protozoan parasites: unusual pathways and novel drug targets. Curr Pharm Des, 14, 901–16. [PubMed: 18473839]

Goodswen SJ, Kennedy PJ & Ellis JT, 2013 A review of the infection, genetics, and evolution of Neospora caninum: from the past to the present. Infect Genet Evol, 13, 133–50. [PubMed: 22985682]

Goodwin S, Mcpherson JD & Mccombie WR, 2016 Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet, 17, 333–51. [PubMed: 27184599]

Gschloessl B, Guermeur Y & Cock JM, 2008 HECTAR: a method to predict subcellular targeting in heterokonts. BMC Bioinformatics, 9, 393. [PubMed: 18811941]

Gubbels MJ & Duraisingh MT, 2012 Evolution of apicomplexan secretory organelles. Int J Parasitol, 42, 1071–81. [PubMed: 23068912]

Harding CR & Meissner M, 2014 The inner membrane complex through development of Toxoplasma gondii and Plasmodium. Cell Microbiol, 16, 632–41. [PubMed: 24612102]

Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S, Babu M, Craig SA, Hu P, Wan C, Vlasblom J, Dar VU, Bezginov A, Clark GW, Wu GC, Wodak SJ, Tillier ER, Paccanaro A, Marcotte EM & Emili A, 2012 A census of human soluble protein complexes. Cell, 150, 1068–81. [PubMed: 22939629]

Henriquez FL, Richards TA, Roberts F, Mcleod R & Roberts CW, 2005 The unusual mitochondrial compartment of Cryptosporidium parvum. Trends Parasitol, 21, 68–74. [PubMed: 15664529]

Hiller NL, Bhattacharjee S, Van Ooij C, Liolios K, Harrison T, Lopez-Estrano C & Haldar K, 2004 A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science, 306, 1934–7. [PubMed: 15591203]

Horvath S & Dong J, 2008 Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol, 4, e1000117. [PubMed: 18704157]

Hsiao CH, Luisa Hiller N, Haldar K & Knoll LJ, 2013 A HT/PEXEL motif in Toxoplasma dense granule proteins is a signal for protein cleavage but not export into the host cell. Traffic, 14, 519–31. [PubMed: 23356236]

Hu G, Cabrera A, Kono M, Mok S, Chaal BK, Haase S, Engelberg K, Cheemadan S, Spielmann T, Preiser PR, Gilberger TW & Bozdech Z, 2010 Transcriptional profiling of growth perturbations of the human malaria parasite Plasmodium falciparum. Nat Biotechnol, 28, 91–8. [PubMed: 20037583]

Hu P, Janga SC, Babu M, Diaz-Mejia JJ, Butland G, Yang W, Pogoutse O, Guo X, Phanse S, Wong P, Chandran S, Christopoulos C, Nazarians-Armavil A, Nasseri NK, Musso G, Ali M, Nazemof N, Eroukova V, Golshani A, Paccanaro A, Greenblatt JF, Moreno-Hagelsieb G & Emili A, 2009 Global functional atlas of Escherichia coli encompassing previously uncharacterized proteins. PLoS Biol, 7, e96. [PubMed: 19402753]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Hung SS & Parkinson J, 2011 Post-genomics resources and tools for studying apicomplexan metabolism. Trends Parasitol, 27, 131–40. [PubMed: 21145790]

Hung SS, Wasmuth J, Sanford C & Parkinson J, 2010 DETECT--a density estimation tool for enzyme classification and its application to Plasmodium falciparum. Bioinformatics, 26, 1690–8. [PubMed: 20513663]

Hupalo DN, Luo Z, Melnikov A, Sutton PL, Rogov P, Escalante A, Vallejo AF, Herrera S, Arevalo-Herrera M, Fan Q, Wang Y, Cui L, Lucas CM, Durand S, Sanchez JF, Baldeviano GC, Lescano AG, Laman M, Barnadas C, Barry A, Mueller I, Kazura JW, Eapen A, Kanagaraj D, Valecha N, Ferreira MU, Roobsoong W, Nguitragool W, Sattabonkot J, Gamboa D, Kosek M, Vinetz JM, Gonzalez-Ceron L, Birren BW, Neafsey DE & Carlton JM, 2016 Population genomics studies identify signatures of global dispersal and drug resistance in Plasmodium vivax. Nat Genet, 48, 953–8. [PubMed: 27348298]

Huthmacher C, Hoppe A, Bulik S & Holzhutter HG, 2010 Antimalarial drug targets in Plasmodium falciparum predicted by stage-specific metabolic network analysis. BMC Syst Biol, 4, 120. [PubMed: 20807400]

Hyde JE, 2007 Targeting purine and pyrimidine metabolism in human apicomplexan parasites. Curr Drug Targets, 8, 31–47. [PubMed: 17266529]

Jakalski M, Wakaguri H, Kischka TG, Nishikawa Y, Kawazu S, Matsubayashi M, Kawahara F, Tsuji N, Cao S, Sunaga F, Xuan X, Okubo K, Igarashi I, Tuda J, Mongan AE, Eshita Y, Maeda R, Makalowski W, Suzuki Y & Yamagishi J, 2015 DB-AT: a 2015 update to the Full-parasites database brings a multitude of new transcriptomic data for apicomplexan parasites. Nucleic Acids Res, 43, D631–6. [PubMed: 25414358]

Janouskovec J, Horak A, Obornik M, Lukes J & Keeling PJ, 2010 A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc Natl Acad Sci U S A, 107, 10949–54. [PubMed: 20534454]

Janouskovec J & Keeling PJ, 2016 Evolution: Causality and the Origin of Parasitism. Curr Biol, 26, R174–7. [PubMed: 26906491]

Janse CJ, Carlton JM, Walliker D & Waters AP, 1994 Conserved location of genes on polymorphic chromosomes of four species of malaria parasites. Mol Biochem Parasitol, 68, 285–96. [PubMed: 7739674]

Jones SH, Lew AE, Jorgensen WK & Barker SC, 1997 Babesia bovis: genome size, number of chromosomes and telomeric probe hybridisation. Int J Parasitol, 27, 1569–73. [PubMed: 9467743]

Kafsack BF, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, Williams AE, Drought LG, Kwiatkowski DP, Baker DA, Cortes A & Llinas M, 2014 A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature, 507, 248–52. [PubMed: 24572369]

Kaneko I, Iwanaga S, Kato T, Kobayashi I & Yuda M, 2015 Genome-Wide Identification of the Target Genes of AP2-O, a Plasmodium AP2-Family Transcription Factor. PLoS Pathog, 11, e1004905. [PubMed: 26018192]

Kapitonov VV & Jurka J, 2008 A universal classification of eukaryotic transposable elements implemented in Repbase. Nat Rev Genet, 9, 411–2; author reply 414. [PubMed: 18421312]

Katzer F, Lizundia R, Ngugi D, Blake D & Mckeever D, 2011 Construction of a genetic map for Theileria parva: identification of hotspots of recombination. Int J Parasitol, 41, 669–75. [PubMed: 21310160]

Kemp LE, Yamamoto M & Soldati-Favre D, 2013 Subversion of host cellular functions by the apicomplexan parasites. FEMS Microbiol Rev, 37, 607–31. [PubMed: 23186105]

Khaminets A, Hunn JP, Konen-Waisman S, Zhao YO, Preukschat D, Coers J, Boyle JP, Ong YC, Boothroyd JC, Reichmann G & Howard JC, 2010 Coordinated loading of IRG resistance GTPases on to the Toxoplasma gondii parasitophorous vacuole. Cell Microbiol, 12, 939–61. [PubMed: 20109161]

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, 2980–92. [PubMed: 15911631]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J & Bryant SH, 2016 PubChem Substance and Compound databases. Nucleic Acids Res, 44, D1202–13. [PubMed: 26400175]

Kissinger JC & Debarry J, 2011 Genome cartography: charting the apicomplexan genome. Trends Parasitol, 27, 345–54. [PubMed: 21764378]

Kono M, Herrmann S, Loughran NB, Cabrera A, Engelberg K, Lehmann C, Sinha D, Prinz B, Ruch U, Heussler V, Spielmann T, Parkinson J & Gilberger TW, 2012 Evolution and architecture of the inner membrane complex in asexual and sexual stages of the malaria parasite. Mol Biol Evol, 29, 2113–32. [PubMed: 22389454]

Kono M, Prusty D, Parkinson J & Gilberger TW, 2013 A membranous system in the spotlight: The apicomplexan Inner Membrane Complex and its adaptation to an endoparasitic life style. Frontiers in Bioscience, In Press.

Koreny L, Obornik M & Lukes J, 2013 Make it, take it, or leave it: heme metabolism of parasites. PLoS Pathog, 9, e1003088. [PubMed: 23349629]

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, Acacio 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, 209–22. [PubMed: 23680352]

Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MH, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A & Greenblatt JF, 2006 Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature, 440, 637–43. [PubMed: 16554755]

Lacount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, Schoenfeld LW, Ota I, Sahasrabudhe S, Kurschner C, Fields S & Hughes RE, 2005 A protein interaction network of the malaria parasite Plasmodium falciparum. Nature, 438, 103–7. [PubMed: 16267556]

Lamarque M, Besteiro S, Papoin J, Roques M, Vulliez-Le Normand B, Morlon-Guyot J, Dubremetz JF, Fauquenoy S, Tomavo S, Faber BW, Kocken CH, Thomas AW, Boulanger MJ, Bentley GA & Lebrun M, 2011 The RON2-AMA1 interaction is a critical step in moving junction-dependent invasion by apicomplexan parasites. PLoS Pathog, 7, e1001276. [PubMed: 21347343]

Lamarque MH, Roques M, Kong-Hap M, Tonkin ML, Rugarabamu G, Marq JB, Penarete-Vargas DM, Boulanger MJ, Soldati-Favre D & Lebrun M, 2014 Plasticity and redundancy among AMA-RON pairs ensure host cell entry of Toxoplasma parasites. Nat Commun, 5, 4098. [PubMed: 24934579]

Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ & Winzeler EA, 2003 Discovery of gene function by expression profiling of the malaria parasite life cycle. Science, 301, 1503–8. [PubMed: 12893887]

Lee JM, Gianchandani EP & Papin JA, 2006 Flux balance analysis in the era of metabolomics. Brief Bioinform, 7, 140–50. [PubMed: 16772264]

Lee JM & Sonnhammer EL, 2003 Genomic gene clustering analysis of pathways in eukaryotes. Genome Res, 13, 875–82. [PubMed: 12695325]

Liesenfeld O, Parvanova I, Zerrahn J, Han SJ, Heinrich F, Munoz M, Kaiser F, Aebischer T, Buch T, Waisman A, Reichmann G, Utermohlen O, Von Stebut E, Von Loewenich FD, Bogdan C, Specht S, Saeftel M, Hoerauf A, Mota MM, Konen-Waisman S, Kaufmann SH & Howard JC, 2011 The IFN-gamma-inducible GTPase, Irga6, protects mice against Toxoplasma gondii but not against Plasmodium berghei and some other intracellular pathogens. PLoS One, 6, e20568. [PubMed: 21698150]

Lim DC, Cooke BM, Doerig C & Saeij JP, 2012 Toxoplasma and Plasmodium protein kinases: roles in invasion and host cell remodelling. Int J Parasitol, 42, 21–32. [PubMed: 22154850]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Lim L & Mcfadden GI, 2010 The evolution, metabolism and functions of the apicoplast. Philos Trans R Soc Lond B Biol Sci, 365, 749–63. [PubMed: 20124342]

Lingelbach K & Joiner KA, 1998 The parasitophorous vacuole membrane surrounding Plasmodium and Toxoplasma: an unusual compartment in infected cells. J Cell Sci, 111 (Pt 11), 1467–75. [PubMed: 9580555]

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, Darde ML, Diaz-Miranda MA, Dubey JP, Fritz HM, Gennari SM, Gregory BD, Kim K, Saeij JP, Su C, White MW, Zhu XQ, Howe DK, Rosenthal BM, Grigg ME, Parkinson J, Liu L, Kissinger JC, Roos DS & David Sibley L, 2016a Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic Toxoplasma gondii genomes. Nat Commun, 7, 10147. [PubMed: 26738725]

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, Darde ML, Diaz-Miranda MA, Dubey JP, Fritz HM, Gennari SM, Gregory BD, Kim K, Saeij JP, Su C, White MW, Zhu XQ, Howe DK, Rosenthal BM, Grigg ME, Parkinson J, Liu L, Kissinger JC, Roos DS & Sibley LD, 2016b Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic Toxoplasma gondii genomes. Nat Commun, 7, 10147. [PubMed: 26738725]

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. [PubMed: 19159464]

Malleret B, Li A, Zhang R, Tan KS, Suwanarusk R, Claser C, Cho JS, Koh EG, Chu CS, Pukrittayakamee S, Ng ML, Ginhoux F, Ng LG, Lim CT, Nosten F, Snounou G, Renia L & Russell B, 2015 Plasmodium vivax: restricted tropism and rapid remodeling of CD71-positive reticulocytes. Blood, 125, 1314–24. [PubMed: 25414440]

Marti M, Good RT, Rug M, Knuepfer E & Cowman AF, 2004 Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science, 306, 1930–3. [PubMed: 15591202]

Marti M & Spielmann T, 2013 Protein export in malaria parasites: many membranes to cross. Curr Opin Microbiol, 16, 445–51. [PubMed: 23725671]

Martinsen ES, Perkins SL & Schall JJ, 2008 A three-genome phylogeny of malaria parasites (Plasmodium and closely related genera): evolution of life-history traits and host switches. Mol Phylogenet Evol, 47, 261–73. [PubMed: 18248741]

Mazurie AJ, Alves JM, Ozaki LS, Zhou S, Schwartz DC & Buck GA, 2013 Comparative genomics of cryptosporidium. Int J Genomics, 2013, 832756. [PubMed: 23738321]

Mcconkey GA, Pinney JW, Westhead DR, Plueckhahn K, Fitzpatrick TB, Macheroux P & Kappes B, 2004 Annotating the Plasmodium genome and the enigma of the shikimate pathway. Trends Parasitol, 20, 60–5. [PubMed: 14747018]

Mcdonald V & Shirley MW, 2009 Past and future: vaccination against Eimeria. Parasitology, 136, 1477–89. [PubMed: 19523251]

Mcfadden GI & Yeh E, 2016 The apicoplast: now you see it, now you don't. Int J Parasitol.

Mckinney JD, Honer Zu Bentrup K, Munoz-Elias EJ, Miczak A, Chen B, Chan WT, Swenson D, Sacchettini JC, Jacobs WR Jr. & Russell DG, 2000 Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature, 406, 735–8. [PubMed: 10963599]

Meissner M, Ferguson DJ & Frischknecht F, 2013 Invasion factors of apicomplexan parasites: essential or redundant? Curr Opin Microbiol, 16, 438–44. [PubMed: 23727286]

Mercier C, Adjogble KD, Daubener W & Delauw MF, 2005 Dense granules: are they key organelles to help understand the parasitophorous vacuole of all apicomplexa parasites? Int J Parasitol, 35, 829–49. [PubMed: 15978597]

Mercier C & Cesbron-Delauw MF, 2015 Toxoplasma secretory granules: one population or more? Trends Parasitol, 31, 60–71. [PubMed: 25599584]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Minot S, Melo MB, Li F, Lu D, Niedelman W, Levine SS & Saeij JP, 2012 Admixture and recombination among Toxoplasma gondii lineages explain global genome diversity. Proc Natl Acad Sci U S A, 109, 13458–63. [PubMed: 22847430]

Miotto O, Almagro-Garcia J, Manske M, Macinnis B, Campino S, Rockett KA, Amaratunga C, Lim P, Suon S, Sreng S, Anderson JM, Duong S, Nguon C, Chuor CM, Saunders D, Se Y, Lon C, Fukuda MM, Amenga-Etego L, Hodgson AV, Asoala V, Imwong M, Takala-Harrison S, Nosten F, Su XZ, Ringwald P, Ariey F, Dolecek C, Hien TT, Boni MF, Thai CQ, Amambua-Ngwa A, Conway DJ, Djimde AA, Doumbo OK, Zongo I, Ouedraogo JB, Alcock D, Drury E, Auburn S, Koch O, Sanders M, Hubbart C, Maslen G, Ruano-Rubio V, Jyothi D, Miles A, O'brien J, Gamble C, Oyola SO, Rayner JC, Newbold CI, Berriman M, Spencer CC, Mcvean G, Day NP, White NJ, Bethell D, Dondorp AM, Plowe CV, Fairhurst RM & Kwiatkowski DP, 2013 Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. Nat Genet, 45, 648–55. [PubMed: 23624527]

Miotto O, Amato R, Ashley EA, Macinnis B, Almagro-Garcia J, Amaratunga C, Lim P, Mead D, Oyola SO, Dhorda M, Imwong M, Woodrow C, Manske M, Stalker J, Drury E, Campino S, Amenga-Etego L, Thanh TN, Tran HT, Ringwald P, Bethell D, Nosten F, Phyo AP, Pukrittayakamee S, Chotivanich K, Chuor CM, Nguon C, Suon S, Sreng S, Newton PN, Mayxay M, Khanthavong M, Hongvanthong B, Htut Y, Han KT, Kyaw MP, Faiz MA, Fanello CI, Onyamboko M, Mokuolu OA, Jacob CG, Takala-Harrison S, Plowe CV, Day NP, Dondorp AM, Spencer CC, Mcvean G, Fairhurst RM, White NJ & Kwiatkowski DP, 2015 Genetic architecture of artemisinin-resistant Plasmodium falciparum. Nat Genet, 47, 226–34. [PubMed: 25599401]

Mogi T & Kita K, 2010 Diversity in mitochondrial metabolic pathways in parasitic protists Plasmodium and Cryptosporidium. Parasitol Int, 59, 305–12. [PubMed: 20433942]

Moreno SN & Li ZH, 2008 Anti-infectives targeting the isoprenoid pathway of Toxoplasma gondii. Expert Opin Ther Targets, 12, 253–63. [PubMed: 18269336]

Moriya Y, Itoh M, Okuda S, Yoshizawa AC & Kanehisa M, 2007 KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res, 35, W182–5. [PubMed: 17526522]

O'hara SP & Chen XM, 2011 The cell biology of cryptosporidium infection. Microbes Infect, 13, 721–30. [PubMed: 21458585]

Oberstaller J, Pumpalova Y, Schieler A, Llinas M & Kissinger JC, 2014 The Cryptosporidium parvum ApiAP2 gene family: insights into the evolution of apicomplexan AP2 regulatory systems. Nucleic Acids Res, 42, 8271–84. [PubMed: 24957599]

Pain A, Bohme 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 CM, Rajandream MA, Kocken CH, Thomas AW, Newbold CI, Barrell BG & Berriman M, 2008 The genome of the simian and human malaria parasite Plasmodium knowlesi. Nature, 455, 799–803. [PubMed: 18843368]

Pain A, Renauld H, Berriman M, Murphy L, Yeats CA, Weir W, Kerhornou A, Aslett M, Bishop R, Bouchier C, Cochet M, Coulson RM, 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 DA, Langsley G, Rajandream MA, 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, 131–3. [PubMed: 15994557]

Painter HJ, Campbell TL & Llinas M, 2011 The Apicomplexan AP2 family: integral factors regulating Plasmodium development. Mol Biochem Parasitol, 176, 1–7. [PubMed: 21126543]

Parker ML, Penarete-Vargas DM, Hamilton PT, Guerin A, Dubey JP, Perlman SJ, Spano F, Lebrun M & Boulanger MJ, 2016 Dissecting the interface between apicomplexan parasite and host cell:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Insights from a divergent AMA-RON2 pair. Proc Natl Acad Sci U S A, 113, 398–403. [PubMed: 26712012]

Parkinson J, Anthony A, Wasmuth J, Schmid R, Hedley A & Blaxter M, 2004 PartiGene--constructing partial genomes. Bioinformatics, 20, 1398–404. [PubMed: 14988115]

Pelle KG, Jiang RH, Mantel PY, Xiao YP, Hjelmqvist D, Gallego-Lopez GM, A OTL, Kang BH, Allred DR & Marti M, 2015 Shared elements of host-targeting pathways among apicomplexan parasites of differing lifestyles. Cell Microbiol, 17, 1618–39. [PubMed: 25996544]

Petersen I, Eastman R & Lanzer M, 2011a Drug-resistant malaria: molecular mechanisms and implications for public health. FEBS Lett, 585, 1551–62. [PubMed: 21530510]

Petersen TN, Brunak S, Von Heijne G & Nielsen H, 2011b SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods, 8, 785–6. [PubMed: 21959131]

Pinney JW, Papp B, Hyland C, Wambua L, Westhead DR & Mcconkey GA, 2007 Metabolic reconstruction and analysis for parasite genomes. Trends Parasitol, 23, 548–54. [PubMed: 17950669]

Plata G, Hsiao TL, Olszewski KL, Llinas M & Vitkup D, 2010 Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network. Mol Syst Biol, 6, 408. [PubMed: 20823846]

Poukchanski A, Fritz HM, Tonkin ML, Treeck M, Boulanger MJ & Boothroyd JC, 2013 Toxoplasma gondii sporozoites invade host cells using two novel paralogues of RON2 and AMA1. PLoS One, 8, e70637. [PubMed: 23940612]

Proellocks NI, Coppel RL & Waller KL, 2010 Dissecting the apicomplexan rhoptry neck proteins. Trends Parasitol, 26, 297–304. [PubMed: 20347614]

Radke JB, Lucas O, De Silva EK, Ma Y, Sullivan WJ Jr., Weiss LM, Llinas M & White MW, 2013 ApiAP2 transcription factor restricts development of the Toxoplasma tissue cyst. Proc Natl Acad Sci U S A, 110, 6871–6. [PubMed: 23572590]

Raman K & Chandra N, 2009 Flux balance analysis of biological systems: applications and challenges. Brief Bioinform, 10, 435–49. [PubMed: 19287049]

Ramaprasad A, Pain A & Ravasi T, 2012 Defining the protein interaction network of human malaria parasite Plasmodium falciparum. Genomics, 99, 69–75. [PubMed: 22178265]

Reid AJ, 2015 Large, rapidly evolving gene families are at the forefront of host-parasite interactions in Apicomplexa. Parasitology, 142 Suppl 1, S57–70. [PubMed: 25257746]

Reid AJ, Blake DP, Ansari HR, Billington K, Browne HP, Bryant J, Dunn M, Hung SS, Kawahara F, Miranda-Saavedra D, Malas TB, Mourier T, Naghra H, Nair M, Otto TD, Rawlings ND, Rivailler P, Sanchez-Flores A, Sanders M, Subramaniam C, Tay YL, Woo Y, Wu X, Barrell B, Dear PH, Doerig C, Gruber A, Ivens AC, Parkinson J, Rajandream MA, Shirley MW, Wan KL, Berriman M, Tomley FM & Pain A, 2014 Genomic analysis of the causative agents of coccidiosis in domestic chickens. Genome Res, 24, 1676–85. [PubMed: 25015382]

Reid AJ, Vermont SJ, Cotton JA, Harris D, Hill-Cawthorne GA, Konen-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. [PubMed: 22457617]

Russo I, Babbitt S, Muralidharan V, Butler T, Oksman A & Goldberg DE, 2010 Plasmepsin V licenses Plasmodium proteins for export into the host erythrocyte. Nature, 463, 632–6. [PubMed: 20130644]

Sato S, 2011 The apicomplexan plastid and its evolution. Cell Mol Life Sci, 68, 1285–96. [PubMed: 21380560]

Scherf A, Hernandez-Rivas R, Buffet P, Bottius E, Benatar C, Pouvelle B, Gysin J & Lanzer M, 1998 Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in Plasmodium falciparum. EMBO J, 17, 5418–26. [PubMed: 9736619]

Schluter D, Daubener W, Schares G, Gross U, Pleyer U & Luder C, 2014 Animals are key to human toxoplasmosis. Int J Med Microbiol, 304, 917–29. [PubMed: 25240467]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Schmuckli-Maurer J, Casanova C, Schmied S, Affentranger S, Parvanova I, Kang'a S, Nene V, Katzer F, Mckeever D, Muller J, Bishop R, Pain A & Dobbelaere DA, 2009 Expression analysis of the Theileria parva subtelomere-encoded variable secreted protein gene family. PLoS One, 4, e4839. [PubMed: 19325907]

Schneider AG & Mercereau-Puijalon O, 2005 A new Apicomplexa-specific protein kinase family: multiple members in Plasmodium falciparum, all with an export signature. BMC Genomics, 6, 30. [PubMed: 15752424]

Schomburg I, Chang A, Placzek S, Sohngen C, Rother M, Lang M, Munaretto C, Ulas S, Stelzer M, Grote A, Scheer M & Schomburg D, 2013 BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res, 41, D764–72. [PubMed: 23203881]

Seeber F & Steinfelder S, 2016 Recent advances in understanding apicomplexan parasites. F1000Res, 5.

Shanmugasundram A, Gonzalez-Galarza FF, Wastling JM, Vasieva O & Jones AR, 2013 Library of Apicomplexan Metabolic Pathways: a manually curated database for metabolic pathways of apicomplexan parasites. Nucleic Acids Res, 41, D706–13. [PubMed: 23193253]

Sharma P & Chitnis CE, 2013 Key molecular events during host cell invasion by Apicomplexan pathogens. Curr Opin Microbiol, 16, 432–7. [PubMed: 23895827]

Shaw MK, 2003 Cell invasion by Theileria sporozoites. Trends Parasitol, 19, 2–6. [PubMed: 12488213]

Shaw MK, Tilney LG & Musoke AJ, 1991 The entry of Theileria parva sporozoites into bovine lymphocytes: evidence for MHC class I involvement. J Cell Biol, 113, 87–101. [PubMed: 1901066]

Shears MJ, Botte CY & Mcfadden GI, 2015 Fatty acid metabolism in the Plasmodium apicoplast: Drugs, doubts and knockouts. Mol Biochem Parasitol, 199, 34–50. [PubMed: 25841762]

Sheiner L, Vaidya AB & Mcfadden GI, 2013 The metabolic roles of the endosymbiotic organelles of Toxoplasma and Plasmodium spp. Curr Opin Microbiol, 16, 452–8. [PubMed: 23927894]

Sibley LD, 2011 Invasion and intracellular survival by protozoan parasites. Immunol Rev, 240, 72–91. [PubMed: 21349087]

Sibley LD, Qiu W, Fentress S, Taylor SJ, Khan A & Hui R, 2009 Forward genetics in Toxoplasma gondii reveals a family of rhoptry kinases that mediates pathogenesis. Eukaryot Cell, 8, 1085–93. [PubMed: 19465561]

Singh GP, Chandra BR, Bhattacharya A, Akhouri RR, Singh SK & Sharma A, 2004 Hyper-expansion of asparagines correlates with an abundance of proteins with prion-like domains in Plasmodium falciparum. Mol Biochem Parasitol, 137, 307–19. [PubMed: 15383301]

Sinha A, Hughes KR, Modrzynska KK, Otto TD, Pfander C, Dickens NJ, Religa AA, Bushell E, Graham AL, Cameron R, Kafsack BF, Williams AE, Llinas M, Berriman M, Billker O & Waters AP, 2014 A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium. Nature, 507, 253–7. [PubMed: 24572359]

Song C, Chiasson MA, Nursimulu N, Hung SS, Wasmuth J, Grigg ME & Parkinson J, 2013 Metabolic reconstruction identifies strain-specific regulation of virulence in Toxoplasma gondii. Mol Syst Biol, 9, 708. [PubMed: 24247825]

Srinivasan P, Beatty WL, Diouf A, Herrera R, Ambroggio X, Moch JK, Tyler JS, Narum DL, Pierce SK, Boothroyd JC, Haynes JD & Miller LH, 2011 Binding of Plasmodium merozoite proteins RON2 and AMA1 triggers commitment to invasion. Proc Natl Acad Sci U S A, 108, 13275–80. [PubMed: 21788485]

Stuart JM, Segal E, Koller D & Kim SK, 2003 A gene-coexpression network for global discovery of conserved genetic modules. Science, 302, 249–55. [PubMed: 12934013]

Su C, Khan A, Zhou P, Majumdar D, Ajzenberg D, Darde ML, Zhu XQ, Ajioka JW, Rosenthal BM, Dubey JP & Sibley LD, 2012 Globally diverse Toxoplasma gondii isolates comprise six major clades originating from a small number of distinct ancestral lineages. Proc Natl Acad Sci U S A, 109, 5844–9. [PubMed: 22431627]

Suhre K, 2007 Inference of gene function based on gene fusion events: the rosetta-stone method. Methods Mol Biol, 396, 31–41. [PubMed: 18025684]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Tachibana S, Sullivan SA, Kawai S, Nakamura S, Kim HR, Goto N, Arisue N, Palacpac NM, 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, 1051–5. [PubMed: 22863735]

Talevich E & Kannan N, 2013 Structural and evolutionary adaptation of rhoptry kinases and pseudokinases, a family of coccidian virulence factors. BMC Evol Biol, 13, 117. [PubMed: 23742205]

Tardieux I & Baum J, 2016 Reassessing the mechanics of parasite motility and host-cell invasion. J Cell Biol, 214, 507–15. [PubMed: 27573462]

Tiburcio M, Sauerwein R, Lavazec C & Alano P, 2015 Erythrocyte remodeling by Plasmodium falciparum gametocytes in the human host interplay. Trends Parasitol, 31, 270–8. [PubMed: 25824624]

Tomita T, Bzik DJ, Ma YF, Fox BA, Markillie LM, Taylor RC, Kim K & Weiss LM, 2013 The Toxoplasma gondii cyst wall protein CST1 is critical for cyst wall integrity and promotes bradyzoite persistence. PLoS Pathog, 9, e1003823. [PubMed: 24385904]

Tonkin ML, Roques M, Lamarque MH, Pugniere M, Douguet D, Crawford J, Lebrun M & Boulanger MJ, 2011 Host cell invasion by apicomplexan parasites: insights from the co-structure of AMA1 with a RON2 peptide. Science, 333, 463–7. [PubMed: 21778402]

Torgerson PR & Mastroiacovo P, 2013 The global burden of congenital toxoplasmosis: a systematic review. Bull World Health Organ, 91, 501–8. [PubMed: 23825877]

Treeck M, Sanders JL, Elias JE & Boothroyd JC, 2011 The phosphoproteomes of Plasmodium falciparum and Toxoplasma gondii reveal unusual adaptations within and beyond the parasites' boundaries. Cell Host Microbe, 10, 410–9. [PubMed: 22018241]

Treeck M, Sanders JL, Gaji RY, Lafavers KA, Child MA, Arrizabalaga G, Elias JE & Boothroyd JC, 2014 The calcium-dependent protein kinase 3 of toxoplasma influences basal calcium levels and functions beyond egress as revealed by quantitative phosphoproteome analysis. PLoS Pathog, 10, e1004197. [PubMed: 24945436]

Tymoshenko S, Oppenheim RD, Agren R, Nielsen J, Soldati-Favre D & Hatzimanikatis V, 2015 Metabolic Needs and Capabilities of Toxoplasma gondii through Combined Computational and Experimental Analysis. PLoS Comput Biol, 11, e1004261. [PubMed: 26001086]

Vanee N, Roberts SB, Fong SS, Manque P & Buck GA, 2010 A genome-scale metabolic model of Cryptosporidium hominis. Chem Biodivers, 7, 1026–39. [PubMed: 20491062]

Villen J & Gygi SP, 2008 The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc, 3, 1630–8. [PubMed: 18833199]

Volkman SK, Sabeti PC, Decaprio D, Neafsey DE, Schaffner SF, Milner DA Jr., Daily JP, Sarr O, Ndiaye D, Ndir O, Mboup S, Duraisingh MT, Lukens A, Derr A, Stange-Thomann N, Waggoner S, Onofrio R, Ziaugra L, Mauceli E, Gnerre S, Jaffe DB, Zainoun J, Wiegand RC, Birren BW, Hartl DL, Galagan JE, Lander ES & Wirth DF, 2007 A genome-wide map of diversity in Plasmodium falciparum. Nat Genet, 39, 113–9. [PubMed: 17159979]

Walker R, Gissot M, Huot L, Alayi TD, Hot D, Marot G, Schaeffer-Reiss C, Van Dorsselaer A, Kim K & Tomavo S, 2013 Toxoplasma transcription factor TgAP2XI-5 regulates the expression of genes involved in parasite virulence and host invasion. J Biol Chem, 288, 31127–38. [PubMed: 24025328]

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. Proc Natl Acad Sci U S A, 110, 7446–51. [PubMed: 23589877]

Wan C, Borgeson B, Phanse S, Tu F, Drew K, Clark G, Xiong X, Kagan O, Kwan J, Bezginov A, Chessman K, Pal S, Cromar G, Papoulas O, Ni Z, Boutz DR, Stoilova S, Havugimana PC, Guo X, Malty RH, Sarov M, Greenblatt J, Babu M, Derry WB, Tillier ER, Wallingford JB, Parkinson J, Marcotte EM & Emili A, 2015 Panorama of ancient metazoan macromolecular complexes. Nature, 525, 339–44. [PubMed: 26344197]

Ward P, Equinet L, Packer J & Doerig C, 2004 Protein kinases of the human malaria parasite Plasmodium falciparum: the kinome of a divergent eukaryote. BMC Genomics, 5, 79. [PubMed: 15479470]

Wasmuth J, Daub J, Peregrin-Alvarez JM, Finney CA & Parkinson J, 2009 The origins of apicomplexan sequence innovation. Genome Res, 19, 1202–13. [PubMed: 19363216]

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, 3.

Wei F, Wang W & Liu Q, 2013 Protein kinases of Toxoplasma gondii: functions and drug targets. Parasitol Res, 112, 2121–9. [PubMed: 23681193]

White NJ, 2009 Malaria Edinburgh: Saunders Ltd.

Who, 2016 World Malaria Report. World Health Organization, 1–186.

Widmer G & Sullivan S, 2012 Genomics and population biology of Cryptosporidium species. Parasite Immunol, 34, 61–71. [PubMed: 21595702]

Wiley JD, Merino EF, Krai PM, Mclean KJ, Tripathi AK, Vega-Rodriguez J, Jacobs-Lorena M, Klemba M & Cassera MB, 2015 Isoprenoid precursor biosynthesis is the essential metabolic role of the apicoplast during gametocytogenesis in Plasmodium falciparum. Eukaryot Cell, 14, 128–39. [PubMed: 25446055]

Woo YH, Ansari H, Otto TD, Klinger CM, Kolisko M, Michalek J, Saxena A, Shanmugam D, Tayyrov A, Veluchamy A, Ali S, Bernal A, Del Campo J, Cihlar J, Flegontov P, Gornik SG, Hajduskova E, Horak A, Janouskovec J, Katris NJ, Mast FD, Miranda-Saavedra D, Mourier T, Naeem R, Nair M, Panigrahi AK, Rawlings ND, Padron-Regalado E, Ramaprasad A, Samad N, Tomcala A, Wilkes J, Neafsey DE, Doerig C, Bowler C, Keeling PJ, Roos DS, Dacks JB, Templeton TJ, Waller RF, Lukes J, Obornik M & Pain A, 2015 Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. Elife, 4, e06974. [PubMed: 26175406]

Xia Q, Muraoka WT, Shen Z, Sahin O, Wang H, Wu Z, Liu P & Zhang Q, 2013 Adaptive mechanisms of Campylobacter jejuni to erythromycin treatment. BMC Microbiol, 13, 133. [PubMed: 23767761]

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, 1107–12. [PubMed: 15510150]

Yahara K, Furuta Y, Oshima K, Yoshida M, Azuma T, Hattori M, Uchiyama I & Kobayashi I, 2013 Chromosome painting in silico in a bacterial species reveals fine population structure. Mol Biol Evol, 30, 1454–64. [PubMed: 23505045]

Zhang B & Horvath S, 2005 A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol, 4, Article17.

Zhou Y, Ramachandran V, Kumar KA, Westenberger S, Refour P, Zhou B, Li F, Young JA, Chen K, Plouffe D, Henson K, Nussenzweig V, Carlton J, Vinetz JM, Duraisingh MT & Winzeler EA, 2008 Evidence-based annotation of the malaria parasite's genome using comparative expression profiling. PLoS One, 3, e1570. [PubMed: 18270564]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

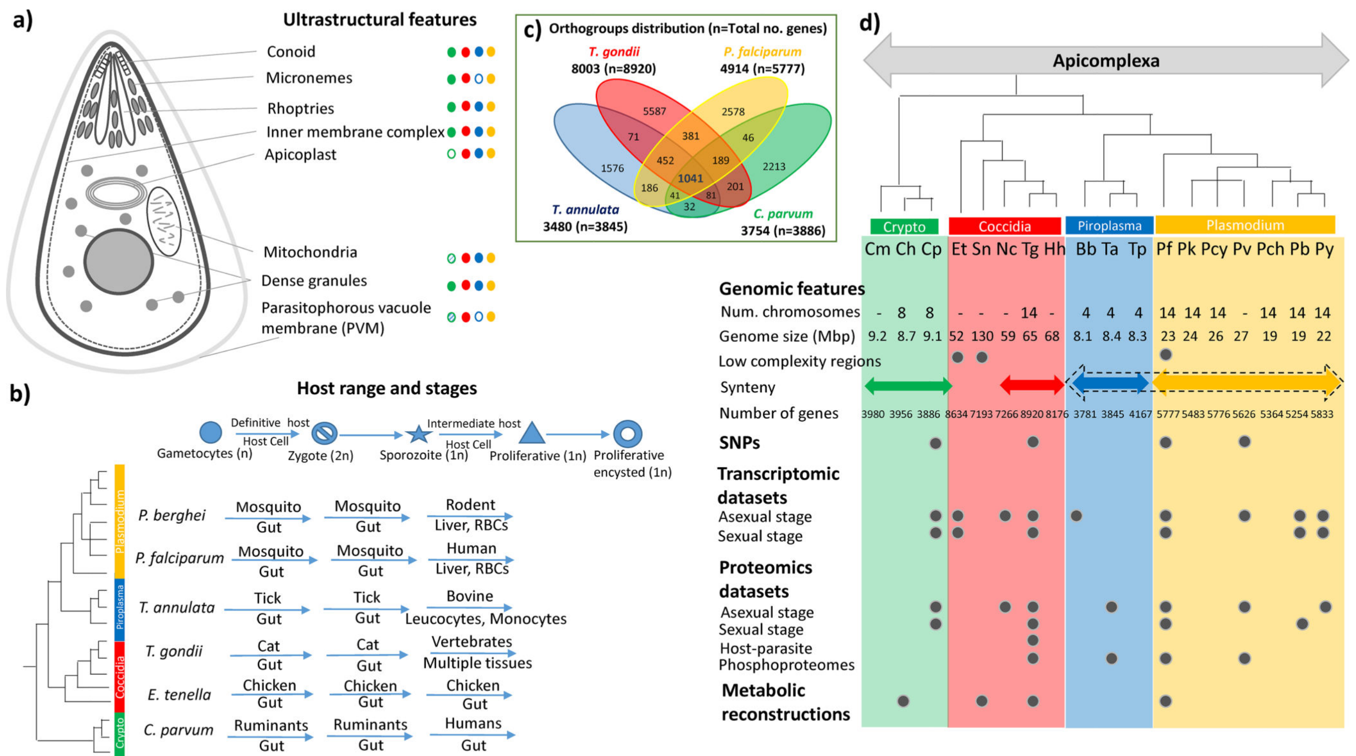

Figure 1: Ultrastructure, Life Cycle, and Genome Structure of Apicomplexa –

The figure highlights details of ultrastructural details unique to apicomplexa, along with their distribution in major clades (Green – Cryptosporidia, Red – Coccidia, Blue – Piroplasmida, Yellow – Plasmodium). Unusual organelles are shown as shaded ellipses – extracytoplasmic PVM (1) and reduced mitochondria-like organelles (2) in Cryptosporidium species. The various life cycle stages are listed along with the host and tissue range common for specific members of each clade in the quadrant below. A venn diagram representing the distribution of orthologous groups among four representative apicomplexans of the major clades is shown. On the right hand side, various genome sequence features as well as details of available functional datasets for sequenced genomes are listed. The species tree for the apicomplexan organisms used in this analysis is shown (based on http://tolweb.org, (3), (4)). The apicomplexan species listed are: Cm – *Cryptosporidum muris*, Ch – *Cryptosporidium hominis*, Cp – *Cryptosporidium parvum*, Et – *Eimeria tenella*, Sn – *Sarcocystis neurona*, Nc – *Neospora caninum*, Tg – *Toxoplasma gondii*, Hh – *Hammodia hammondi*, Bb – *Babesia bovis*, Ta – *Theileria annulata*, Tp – *Theileria parva*, Pf – *Plasmodium falciparum*, Pk – *Plasmodium knowlesi*, Pcy – *Plasmodium cynomolgi*, Pv – *Plasmodium vivax*, Pch – *Plasmodium chabaudi*, Pb – *Plasmodium berghei*, Py – *Plasmodium yoelii*.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

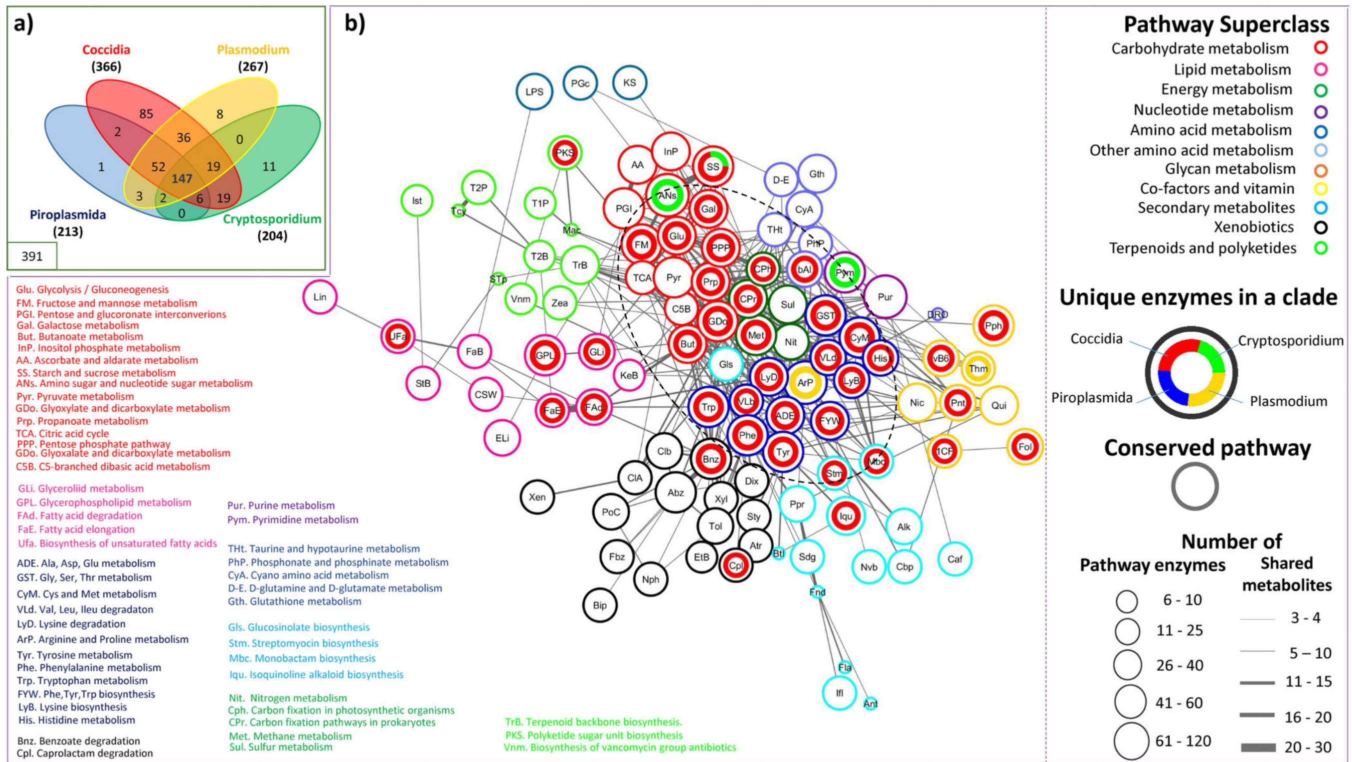

Figure 2: Metabolic potential in apicomplexan clades –

The venn diagram represents the distribution of enzymes common to all apicomplexans, shared between various clades, and unique to each clade. The distribution of conserved pathways and pathways with clade-specific enzymes (2 in a clade) is represented as a network. The network represents KEGG pathways (nodes) connected by number of shared metabolites, with pathways belonging to a superclass (same border colours) grouped together wherever possible. Each node is represented as a circos chart depicting the number of unique enzymes present in each major clade. Conserved pathways are indicated as empty circles. The core of the network, enclosed in a dashed circle, mainly encompasses pathways from amino acid, carbohydrate, energy, and nucleotide metabolism, with quite a few conserved pathways, as well as several clade-specific pathways, especially from coccidia. The abbreviated pathway names are expanded for those in the core, and the pathways with unique enzymes, in the periphery.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

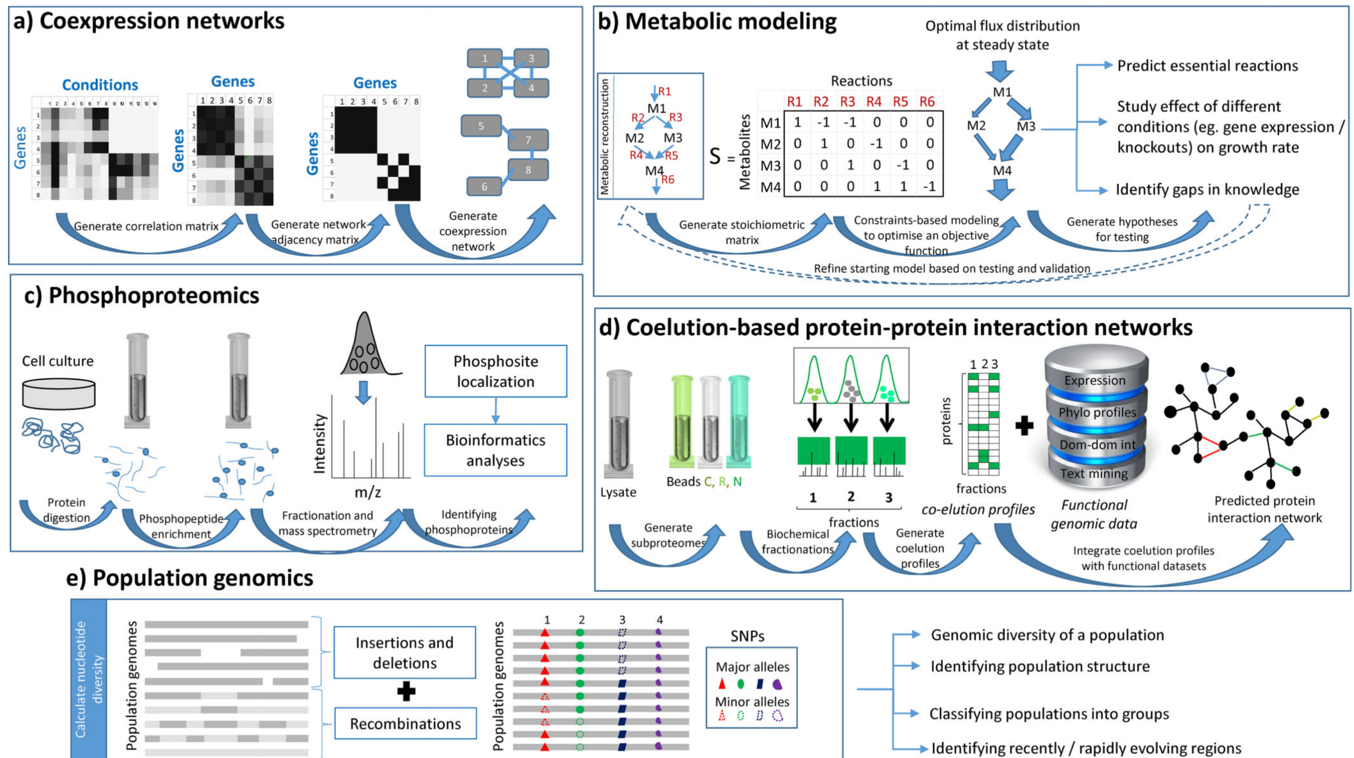

Figure 3: High-throughput post-genomic approaches –

The figure highlights the conceptual framework behind five post-genomic approaches utilised to study protein function in a systems-based context – a) Coexpression networks b) Metabolic modeling c) Phosphoproteomics d) Coelution-based protein-protein interaction networks e) Population genomics

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1:

Distribution of apicomplexan proteins in various biological processes

Category	GOSlim Term / Source	Crypto		Coccidia				Piroplasmida			Plasmodium		
		Cmur	Cpar	Eten	Ncan	Tgon	Hham	Bbov	Tpar	Tann	Pfal	Pber	Pviv
Proteome size		3980	3886	8634	7266	8920	8176	3781	3845	4167	5777	5254	5626
Housekeeping													
cellular protein modification process	GO:0006464	163	144	193	236	245	185	88	102	87	265	149	228
translation	GO:0006412	120	118	163	195	207	164	197	237	175	315	233	266
biosynthetic process	GO:0009058	202	183	248	216	327	265	178	240	172	495	274	356
catabolic process	GO:0009056	66	60	77	89	94	71	54	60	55	148	72	111
protein folding	GO:0006457	38	34	36	52	50	36	37	47	35	100	60	65
nucleobase-containing compound catabolic process	GO:0034655	35	32	41	53	56	32	31	31	30	49	42	45
cell cycle	GO:0007049	18	16	10	15	17	12	8	13	9	72	12	32
response to stress	GO:0006950	43	39	48	69	94	64	31	46	34	111	47	89
homeostatic process	GO:0042592	25	20	20	35	34	30	18	16	15	47	29	38
Metabolism													
cellular nitrogen compound metabolic process	GO:0034641	209	161	236	183	298	234	192	233	177	492	290	399
DNA metabolic process	GO:0006259	92	87	100	108	128	94	77	96	83	155	98	134
small molecule metabolic process	GO:0044281	93	80	120	148	155	105	88	100	81	245	145	193
lipid metabolic process	GO:0006629	43	33	50	74	84	63	28	34	24	138	53	77
carbohydrate metabolic process	GO:0005975	41	43	78	82	91	78	27	29	25	64	39	60
cellular amino acid metabolic process	GO:0006520	35	38	64	88	93	75	46	52	47	92	69	80
tRNA metabolic process	GO:0006399	49	45	55	62	65	57	52	55	54	84	59	68
Signaling													
signal transduction	GO:0007165	70	55	79	88	98	81	45	77	40	97	64	77
Transport													
transport	GO:0006810	154	125	145	190	210	130	102	149	111	344	168	200
transmembrane transport	GO:0055085	68	62	80	98	111	98	49	66	70	78	62	64
vesicle-mediated transport	GO:0016192	47	41	48	58	59	44	38	37	40	95	58	69
Invasion													

Category	GOSlim Term / Source	Crypto		Coccidia				Piroplasmida			Plasmodium		
		Cmur	Cpar	Eten	Ncan	Tgon	Hham	Bbov	Tpar	Tann	Pfal	Pber	Pviv
Inner membrane complex	(1)*, (2)#, (3)@, EupathDB	11*	12*	9*	11*	41*#	22	-	14*	12*	28@	21	20
Rhoptry neck proteins	(4)*, EupathDB	1*	1*	3*	7*	11	10	4*	4*	4*	7	7	7
Micronemal proteins	(5)*, (6)#, EupathDB	-	3*	8*	5	25	25	0	0	1	17#	10*	2
Host associated processes													
ROPK	(7)	0	0	27	44	55	0	0	0	0	0	0	0
Non-ROPK rhoptry proteins	(6)*, EupathDB	0	0	2	-	23*	23	2	0	3	22*	11	10
FIKK	EupathDB	0	0	0	0	1	1	0	0	0	21	1	1
SVSP	(8)	0	0	0	0	0	0	0	85	51	0	0	0
SRS	(8)*, EupathDB	0	0	0	227*	111	95	0	0	0	0	0	0
SAG	EupathDB	0	0	87	0	0	0	0	0	0	0	0	0
Dense granule proteins	(9)*, EupathDB	1*	1*	1	0	16	15	0	0	0	0	0	0
Var	(8)	0	0	0	0	0	0	0	0	0	60	0	0
|Rifin	EupathDB	0	0	0	0	0	0	0	0	0	185	0	0
stevor	EupathDB	0	0	0	0	0	0	0	0	0	42	0	0
Pir	(8)	0	0	0	0	0	0	0	0	0	0	100	346
PHIST	(8)*, EupathDB	0	0	0	0	0	0	0	0	0	79	3	39*
ves	(8)	0	0	0	0	0	0	72	0	0	0	0	0
tpr/tar	(8)	0	0	0	0	0	0	0	39	93	0	0	0
Transcription factors													
AP2	(6)	-	9	15	53	53	-	20	19	18	26	25	27
Unknown function													
Hypothetical proteins	EupathDB	2148	1478	5850	4096	4285	3966	1846	2983	2113	2112	1905	2320

The table shows the distribution of number of proteins for various species according to different categories. The top 20 GOSlim terms were grouped into functionally related categories, such as Housekeeping, Metabolism, Transport, and Signaling. The numbers for *P. falciparum* may have some annotation bias associated with them. The details of proteins involved in apicomplexan-specific processes such as invasion and host-parasite interactions were sourced from specific literature sources wherever available (indicated as *), and from EupathDB otherwise. An estimate of number of proteins of currently unknown function was also obtained from EupathDB.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript