Diversity of extracellular proteins during the transition from the ‘proto-apicomplexan’ alveolates to the apicomplexan obligate parasites

THOMAS J. TEMPLETON¹,² * and ARNAB PAIN³,⁴

¹ Department of Protozoology, Institute of Tropical Medicine (NEKKEN), Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan
² Department of Microbiology and Immunology, Weill Cornell Medical College, New York 10021, USA
³ Pathogen Genomics Laboratory, Biological and Environmental Sciences and Engineering (BESE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Jeddah 23955-6900, Kingdom of Saudi Arabia
⁴ Global Station for Zoonosis Control, Global Institution for Collaborative Research and Education (GI-CoRE), Hokkaido University, N20 W10 Kita-ku, Sapporo 001-0020, Japan

(Received 22 May 2015; revised 12 August 2015; accepted 14 August 2015)

SUMMARY

The recent completion of high-coverage draft genome sequences for several alveolate protozoans – namely, the chromerids, Chromera velia and Vitrella brassicaformis; the Perkinsids Perkinsus marinus; the apicomplexans, Gregarina niphonbrodes, as well as high coverage transcriptome sequence information for several colpodellids, allows for new genome-scale comparisons across a rich landscape of apicomplexans and other alveolates. Genome annotations can now be used to help interpret fine ultrastructure and cell biology, and guide new studies to describe a variety of alveolate life strategies, such as symbiosis or free living, predation, and obligate intracellular parasitism, as well to provide foundations to dissect the evolutionary transitions between these niches. This review focuses on the attempt to identify extracellular proteins which might mediate the physical interface of cell–cell interactions within the above life strategies, aided by annotation of the repertoire of predicted surface and secreted proteins encoded within alveolate genomes. In particular, we discuss what descriptions of the predicted extracellular proteomes reveal regarding a hypothetical last common ancestor of a pre-apicomplexan alveolate – guided by ultrastructure, life strategies and phylogenetic relationships – in an attempt to understand the evolution of obligate parasitism in apicomplexans.

Key words: Apicomplexa, Chromera, Plasmodium, Colpodella, Vitrella, Alveolata, Toxoplasma, Perkinsus, Cryptosporidium, Gregarina.

INTRODUCTION

The alveolates are defined as a superphylum within a supergroup also containing Stramenopiles and Rhizaria (termed SAR, for Stramenopiles, Alveolata and Rhizaria; Adl et al. 2012), and are united by the presence of namesake sub-membranous flattened vesicles termed alveoli. The alveolates include the ciliates, such as Tetrahymena and Paramecium; dinoflagellates, including important pathogens of shellfish; Perkinsus, which represents a branch related to dinoflagellates; the chromerids, such as Chromera and Vitrella, and the closely related colpodellids; and the obligate parasites termed Apicomplexa, including the human pathogens, Toxoplasma and the malaria parasite Plasmodium. Apicomplexans encompass a spectrum of parasitic transmission strategies, such as the transmission via environmentally durable oocysts, including Cryptosporidium, Gregarina and the coccidians Toxoplasma and Eimeria; or via insect vectors, as for the mosquito-borne Plasmodium, and tick-transmitted Babesia and Theileria. Apicomplexans have evolved to specifically target a variety of host cells, such as gut epithelial cells in the instance of Cryptosporidium and Gregarina; gut epithelial recognition followed by disseminative infections having universal host cell invasion and development with regard to Toxoplasma; and life cycle stage-dependent host cell recognition in the insect-transmitted parasites, including pathology-relevant erythrocytic stages with respect to Plasmodium and Babesia, and lymphocytes in the case of Theileria.

Apicomplexans are named for their apical complex which imparts cell polarity for specific interactions with target cells, and provides a conduit for secretions from organelles, such as rhoptries and micronemes. However, this structure is not unique to apicomplexans. For example, the predatory alveolate and close cousin to Apicomplexa, Colpodella, as well as Perkinsus, have highly developed apical complexes and secretion systems...
Apicomplexans might be further defined and unified as having the additional hallmarks of obligate parasitism, and the capacity for gliding motility. This review seeks to illuminate, using manually curated extracellular proteome predictions derived from new whole genome and transcriptome annotations, the evolutionary leap between a hypothetical ‘proto-apicomplexan’, and a last common ancestor within the Apicomplexa.

Ultrastructure studies give a picture of the breadth of apical complex structures from apicomplexans and closely related alveolates such as Perkinsus, Colpodella and the chromerids. These structures and their component secretory organelles loosely range from apical polarity and secretion in Perkinsus (Coss et al. 2001); pseudo-conoid apical complexes, as in Chromera (Moore et al. 2008; Obornik et al. 2011) and Vitrella (Obornik et al. 2012); a specialized apical region capable of forming apparent tight junctions, as in Colpodella (Simpson and Patterson, 1996; Brugerolle, 2002); and the namesake apical regions of apicomplexans conferring apical adhesion and secretion, and mediation of gliding motility, invasion of target cells and tissue disruption (reviewed in Gubbels and Duraisingh, 2012). Alveolates have diverse secretory systems at their disposal, organelles whose composition and character differ based upon genera and the life cycle stage. Examples include the broad variety of trichocyst-like secretory organelles of ciliates which release protein and pigment cargos involved in predatory and defence mechanisms (reviewed in Lobban et al. 2007; Briguglio and Turkewitz, 2014), dense granules having diverse structures and phylogenetic distribution, and apical rhoptries and micronemes (reviewed in Gubbels and Duraisingh, 2012). Regulated secretion via calcium fluxes might be a common theme underpinning alveolates ranging from Paramecium to Plasmodium (reviewed in Vayssié et al. 2000; Plattner et al. 2012); in part utilizing conserved plant-like calcium-dependent protein kinases such as described for regulated exocytosis (Lourido et al. 2010) and parasite egress from host cells (McCoy et al. 2012) in Toxoplasma, and processes of Plasmodium (reviewed in Holder et al. 2012). The use of secretory organelles as taxonomic morphological markers is, it might be argued, still in its infancy as a phylogenetic tool; and to our knowledge no cellular localization studies have mapped secretory proteins to their resident organelles in alveolates other than apicomplexans. Indeed, thus far no candidate orthologues for, say, microneme or rhoptry proteins have been identified which are conserved in both pre-apicomplexans and apicomplexans, such as by bioinformatics screens or immunolocalization assays. For Apicomplexa there is a rich, albeit sometimes muddy literature describing constituent proteins of rhoptry, microneme and dense granule organelles. We will not attempt to review that literature here; rather, we will present examples to give an overview comparison of apicomplexan and pre-apicomplexan alveolate extracellular domains and domain architectures in proteins.

Based on ultrastructure and life strategy criteria, in addition to phylogenetic analyses, Colpodella might be considered to be a compelling candidate for the ‘last common ancestor’ of the Apicomplexa (Brugerolle, 2002; Leander et al. 2003). This protozoan is capable of specific recognition of prey, such as the kinetoplastid Bodo caudatus; adhering to it and forming a junction via its apical complex; and engulfing the target cell into its gullet-like compartment (Simpson and Patterson, 1996; Brugerolle, 2002). Colpodella is not parasitic, unlike its cousin, the intracellular pathogen, Perkinsus. The molecular and mechanical details underpinning the invasion strategy of bivalve haemocytes by Perkinsus are not known (Soudant et al. 2013) and warrant further study, particularly utilizing the recently released high coverage genome sequence information for this alveolate pathogen. The development of cell invasion or parasitism does not necessarily lie along the evolutionary pathway to the Apicomplexa, as these features have evolved multiple times in the alveolates, including Perkinsus (Soudant et al. 2013), and the endoparasite of fish epithelial tissue, the ciliate Ichthyophthirius multifiliis (Coyne et al. 2011). Within the Perkinsids different host targets exist, including the exquisite example of Parvilucifera prorocentri, which invades and develops within dinoflagellates (Hoppenrath and Leander, 2009). As detailed in the sections following, our descriptions of predicted extracellular proteins suggest that the endosymbiotic chromerids and predatory colpodellids are more akin to Apicomplexa than Perkinsus, in support of phylogenetic analyses (Moore et al. 2008; Woo et al. 2015).

Physical interactions with foreign cells among alveolates include symbiotic or commensal relationships, predation and avoidance thereof, and parasitism. These interactions would be expected to be driving forces in the selection for specific extracellular proteins that underpin recognition, adherence and response to target cells. For example, Plasmodium sporozoites interact with a diverse set of host tissues – firstly salivary gland tissue in the mosquito, and subsequently tropism to hepatocytes in the liver – utilizing receptor-mediated recognition of target cells (reviewed in Sinnis and Coppi, 2007). Alveolate recognition of the environment also includes positive and negative taxis in response to gradients of nutrients, toxins, light and gravity (Eckert, 1972; Fenchel and Finlay, 1984; Francis and Hennessey, 1995; Hemmersbach et al. 1999; Selbach and Kuhlmann, 1999; Cadetti et al. 2000; reviewed in Echevarria et al. 2014), as well as avoidance interactions with predators (Knoll et al. 1991; Selbach and Kuhlmann, 1999).
functions in membrane fusion (Liu and Eimeria) transfer of metazoan ephrin proteins (Gerlo and have been proposed to have originated via lateral membranes by GPI moieties, including ciliate receptors described in a section below; tethering to such as in rhodopsins or signalling channels; single onment might either be anchored within the surface (GPI)-linked proteins composed of 6-cys domains family which encodes glycophosphatidyl inositol cycle, including hepatocyte and intraerythrocytic and appear to have roles in multiple stages of the life-Paramecium such as the mating type surface protein mtA in which appears to be mediated by surface proteins, as such extracellular proteins might have arisen by vertical inheritance, in many instances followed by lineage-specific divergence such that their origin is obscure; or by lateral transfer (reviewed in Aravind et al. 2012). The GPI-linked immobilization antigens of ciliates were the first alveolate cell surface proteins to be identified; and the characterization of agglutination by specific immune sera helped to formulate concepts of antigenic diversity, antigenic switching and allelic exclusion (reviewed in Caron and Meyer, 1989; Beale and Preer, 2008). These themes were of value later in describing the immune pressure driven antigenic diversity and switching in the surface proteins of the kinetoplastid and human pathogen, Trypanosoma brucei, as well as Plasmodium. Despite a wealth of literature, it remains unknown why ciliates devote themselves to great amplification of genes encoding surface coat immobilization antigens (e.g. the family exemplified by P. tetraurelia protein AAA61739-2), and appear to switch expression of these genes. Instances of gene amplification of predicted surface and secreted proteins are frequently repeated in the alveolates, and might be driven by multiple mechanisms. For example, ciliates possess extensive amplifications of genes encoding membrane attack complex perforin (MACPF)-like domains (e.g. protein family exemplified by TITHERM_01380980 in Tetrahymena) which may participate in lytic pore formation, as do the macrophage MACPF proteins of vertebrates, and mediate membrane traversal in apicomplexans such as Toxoplasma and Plasmodium (reviewed in Kafsack and Carruthers, 2012). It is not known why ciliates require large numbers of MACPF domain encoding genes – perhaps they serve as attack complexes in predation, or in defence from predators – and if the functions and pressures driving gene amplification are conserved or differ for the MACPF gene expansions observed in apicomplexans. Many other examples of extensive amplifications of genes encoding predicted extracellular alveolate proteins are described in the following sections.

ALVEOLATE PHYLOGENY, APICOMPLEXANS AND ‘PROTO-APICOMPLEXANS’

The phylogenetic tree depicted in Fig. 1 gives our current understanding of the relationships of the pre-apicomplexan alveolates with respect to the Apicomplexa, with the chromerids and colpodellids branching at the base of the apicomplexan clade (Janouškovec et al. 2015; Woo et al. 2013). The placements of specific genera within this general classification of Alveolata is an ongoing work,
Fig. 1. Phylogenetic tree of the alveolates, showing relationships with the Apicomplexa (generalized from Templeton et al. 2010; Janouškové et al. 2015; Woo et al. 2015). The apicomplexan clade is indicated by a grey box. The grouping indicated by the yellow box shows the relationship of the Chromera, Vitrella and the colpodellids; and as a sister clade to the Apicomplexa and thereby putative model ‘proto-apicomplexans’.

require solutions to many puzzles concerning phylogenetic relationships, such as the affinity of Perkinsus and Colponema with respect to chromerids and dinoflagellates, and the nature of the candidate last-common relatives leading to the Apicomplexa. The chromerids and colpodellids are the most closely related clades to the Apicomplexa, with Gregarina and Cryptosporidium serving as bookends on the apicomplexan side of the transition to obligate parasitism (Templeton et al. 2010). These relationships can then be used, coupled with new understandings of ultrastructures and life strategies (e.g. see Okamoto and Keeling, 2014; Portman and Slapeta, 2014), as a foundation for examining the transition from a hypothetical ‘proto-apicomplexan’ (indicated by the yellow shaded box in Fig. 1) to the phylum Apicomplexa, using the repertoires of predicted extracellular proteins. As a basis for this review, we have performed extensive and sensitive basic local alignment search tool (BLAST) screens of extracellular proteins in order to compare apicomplexans with other alveolates and to take advantage of new genome sequence information from the ciliates, Ichthyophthirius (Coyne et al. 2011), Oxytricha (Swart et al. 2013) and Stylonychia (Aeschlimann et al. 2014); the chromerids, Chromera velia and Vitrella brassicaformis (Woo et al. 2015); Perkinsus marinus; high coverage transcriptome sequence information for several colpodellids (Janouškové et al. 2015) and GenBank-deposited genome sequence information for the apicomplexans Gregarina niphandrodes, Cryptosporidium muris, Eimeria tenella and Hammondia. Genome sequence information is also available for Tetrahymena thermophila (Eisen et al. 2006), P. tetraurelia (Aury et al. 2006), Babesia bovis (Brayton et al. 2007) and other Babesia species (Cornillot et al. 2012; Jackson et al. 2014), Eimeria spp. (Heitlinger et al. 2014; Reid et al. 2014), Theileria parva and Theileria annulata (Gardner et al. 2005; Pain et al. 2005), Cryptosporidium parvum and Cryptosporidium hominis (Abramnsen et al. 2004; Xu et al. 2004) and numerous Plasmodium species (available and described at the online database resource, http://www.plasmodb.org).

**GLYCOXYLATION, MUCINS AND SUGAR-BINDING DOMAINS**

The fundamental cell–cell interaction of alveolates is probably recognition of carbohydrate residues on target cells mediated by cell surface proteins containing lectin-like domains (Robert et al. 2006; Wood-Charlson et al. 2006; Wootton et al. 2007; Martel, 2009). Such ligand and receptor interactions might be either simply adhesive, graded along the abundance of receptors and affinity of interaction with target molecules, or involve a signalling component such that the protozoan receives information regarding the target cell which triggers a response, such as a change in flagellar activity. Alveolates possess a great range of possible carbohydrate-binding domains that appear to either be specific to classes within alveolates, or have broader distribution within prokaryotes and eukaryotes. Some domains are well-described, such as the conserved domains lectin, ricin and chitin binding (see Table 1); whereas many lectin proteins were identified based upon experimental affinity for carbohydrates as, for example, in Cryptosporidium (Bhat et al. 2007; Bhalchandra et al. 2013). Examples of carbohydrate-binding receptors include recognition of erythrocyte surface sialic acid by the Plasmodium merozoite protein EBA-175 during invasion (reviewed in Gaur and Chitnis, 2011); recognition of sialic acid via an unrelated saccharide-binding module, as well as a gal-lectin domain, within MIC1 participating in host cell invasion by Toxoplasma (TGME49_291890 in Fig. 2A; Friedrich et al. 2010); and recognition of host carbohydrates mediated by an N-terminal domain within the microneme-secreted proteins MIC3 and MIC8 (CBL domain shown in TGME49_286740, Fig. 2A) during host cell invasion by Toxoplasma tachyzoites (Cérède et al. 2005). These carbohydrate-binding domains appear to be specific to 1 or more apicomplexan genera and are not found in proto-apicomplexans. Select alveolate proteins with predicted carbohydrate-binding activity are shown in Fig. 2A. It is probable that many alveolate carbohydrate-binding domains remain anonymous because they are not similar to known modules with such activity.

Alveolate interactions with host cells might conversely involve purposeful display of carbohydrate...
residues on the parasite membrane surface, in order to engage host lectins. Such ‘mucin-like’ proteins are typically highly decorated with O-linked glycosylation, such as within large stretches of threonine and serine residues. One of the first apicomplexan mucin proteins to be identified was the Cryptosporidium sporozoite surface protein gp900 (e.g. AAC98153), which is proposed to be involved in host cell invasion (Barnes et al. 1998). Gp900 is composed of cysteine-rich domains and a lengthy array of threonine residues. Subsequent annotation of the C. parvum genome revealed a large repertoire of predicted mucin proteins.

Table 1. Phylogenetic distributions for select alveolate extracellular domains

| Domain | Domain ID | Ciliates | P. mar. | C. vel. | V. brass. | Colp. | C. parv. | Coccidians | Plasmodium |
|--------|-----------|----------|---------|---------|-----------|-------|---------|------------|------------|
| FN3    | cd00063   |          |         |         |           |       |         |            |            |
| IG, FLMN | c19759 |          |         |         |           |       |         |            |            |
| IPT/TIG | c15674   |          |         |         |           |       |         |            |            |
| Myxo disulfo repeat | c11785 |          |         |         |           |       |         |            |            |
| GS     | pfam10162 |          |         |         |           |       |         |            |            |
| ChBD1  | cld0035   |          |         |         |           |       |         |            |            |
| WSC    | cld02568  |          |         |         |           |       |         |            |            |
| HYR    | cld03620  |          |         |         |           |       |         |            |            |
| HINT   | cld0081   |          | Nuclear version? | | | | | | |
| PAN-APPLE | cld0112 |          |         |         |           |       |         |            |            |
| MAM    | cld06263  |          |         |         |           |       |         |            |            |
| CAO    | pfam01179 |          |         |         |           |       |         |            |            |
| COWP   | -         |          |         |         |           |       |         |            |            |
| TOX1   | -         |          |         |         |           |       |         |            |            |
| Notch  | cld02419  |          |         |         |           |       |         |            |            |
| RICIN  | cld00161  |          |         |         |           |       |         |            |            |
| NEC    | cld02436  |          |         |         |           |       |         |            |            |
| SR     | cld02509  |          |         |         |           |       |         |            |            |
| LCCl   | cld02694  | Oxytricha ? |          |         |           |       |         |            |            |
| CCP (Sushi) | cld0043 |          |         |         |           |       |         |            |            |
| DISCOIDIN | Pfam00754 |          |         |         |           |       |         |            |            |
| TSP1   | cld05278  |          |         |         |           |       |         |            |            |
| PA14   | smart00758 |          |         |         |           |       |         |            |            |
| sWA    | cld0057   |          |         |         |           |       |         |            |            |
| MacPF | cld02616  |          |         |         |           |       |         |            |            |
| TRP    | cld18480  |          |         |         |           |       |         |            |            |
| GCC2/GCC3 | pfam07699 |          |         |         |           |       |         |            |            |
| EGF    | cld00054  |          |         |         |           |       |         |            |            |

a White boxes indicate the absence and orange-shaded boxes indicate the presence of the domain (rows) in the genome or transcript sequence information for the relevant group (columns). Boxes shaded with grey and having a question mark indicate that the domain was not found, but the absence is qualified by the fact that the genome sequence information is incomplete for the relevant group.

b Domain accession identifiers. Domain information can be retrieved at the NCBI Conserved Domain website: http://www.ncbi.nlm.nih.gov/cdd.

c Species abbreviations: P. mar., Perkinsus marinus; C. vel., Chromera velia; V. brass., Vitrella brassicaformis; and C. parv., Cryptosporidium parvum.

d General colpodellid grouping which includes Colpodella angusta, Colpodella sp BE-6, Alphamonas edax and Voromonas pontica. Domains were surveyed by local tblastn screening of transcript sequence information (databases described in Janouškovec et al. 2015) using chromerid and apicomplexan domain queries. The databases are incomplete and thus negative results are provisionally indicated by grey and a question mark, rather than white shading. Moreover, positive hits were not necessary for all organisms; for example, the HINT domain was only observed in V. pontica.

e At the time of publication this accession identifier was valid, but the relevant entry could not be retrieved at the NCBI Conserved Domain website: http://www.ncbi.nlm.nih.gov/cdd.

f Cysteine-rich domain found in Cryptosporidium oocyst wall proteins (COWP) and coccidians (Spano et al. 1997).

g Domain described in the Supplemental material for Templeton et al. (2004b); specifically, ‘Domain typically with 6 cysteines, seen thus far mainly in animals with a few occurrences in plants. It is found in the sea anemone toxin metridin and fused to animal metal proteases, plant prolyl hydroxylases and is vastly expanded in the genome of C. elegans.’
Fig. 2. Examples of predicted alveolate extracellular proteins having sugar-binding domains or mucin-like threonine repeat domains (A) and cysteine-rich modular protein (CRMP) architectures (B). Signal peptides are indicated by yellow boxes; and transmembrane domains are indicated by dark grey boxes. Colpodellids also possess CRMP proteins, but are not shown here due to the fragmentation of the transcriptome databases, and because they do not differ remarkably to chromerid examples. Domain abbreviations refer to either SMART (http://smart.embl-heidelberg.de/browse.shtml) or pfam (http://pfam.xfam.org) conserved domains. See also domain abbreviations and identifiers in Table 1.
(Abrahamsen et al. 2004), which are predominantly species-specific; in that for the most part they are not conserved in the repertoire of predicted mucins encoded within the genome of the Cryptosporidium parasite of gastric tissue, *C. muris* (Templeton, 2008). A mucin in Toxoplasma gondii, termed CST1 (TGME49_064660), is composed of a large repeat of SAG-related sequence (SRS) domains plus a threonine-rich array similar to Gp900, and is thought to be highly modified with N-acetyl-galactosamine (Tomita et al. 2013). CST1 was recently found to be crucial for the integrity of tissue cyst walls, with the threonine-rich region playing a critical role. Annotation of chromerids and colpodellids also reveals mucin proteins (for example Cvel_819.t1, Cvel_541.t1, and Colpodella_angusta_Spi-2-cDNA_ca@a28207_52), as well as a conserved O-linked glycosylation machinery which was first described in coccidians (Templeton et al. 2004b; Walker et al. 2010). Our rough annotation of *P. marinus* indicates that it has perhaps an order of magnitude more genes which encode predicted mucins than *Chromera*, with potentially over 500 mucin genes within several families; based upon the features of predicted secretion, the presence of threonine repeats, and transmembrane or GPI-anchor domains. A *Perkinsus* mucin protein family has a conserved cysteine at the C-terminal residue, which possibly confers association with the surface membrane via fatty acylation of the cysteine residue (e.g. pmar_ XP_002783417-1). Annotation of *Chromera, Vitrella* and colpodellids also revealed numerous proteins with predicted sugar-binding domains interspersed with threonine-rich repeats, suggesting that the proteins participate in polymerization to form intra- and inter-molecular matrices based upon sugar-binding motifs and sugar moieties (Fig. 2A). Parenthetically, regarding the stabilization of protein matrices, it has also been proposed that peroxidase-mediated cross-linking of di-tyrosine residues contribute to the integrity of coccidian oocyst walls (Mai et al. 2011).

**TWO COMPONENT SENSORY TRANSDUCTION HISTIDINE KINASE**

Little is known regarding signal transduction across alveolate surface membranes in response to external environmental information, either in cell–cell or cell–nutrient interactions. The complexity of such signalling might distinguish apicomplexans and non-apicomplexans, if the former are considered to reside in relatively defined environments within hosts, and thus have a lesser requirement to respond to changing external environments during free-living life cycle stages. This hypothesis, however, must be reconciled with the apparent complexity of environmental recognition that arises during transformation between stages, changes in tissue localization within hosts, and transmission between hosts, such as during completion of the life cycle of *Plasmodium*.

One example of a possible alveolate signalling system, which is known only from annotation work and has not been pursued at the lab bench, is the observation that ciliates, *Perkinsus* and the chromerids possess large families of predicted two component sensory transduction histidine kinases (e.g. Cvel_8519.t1 in *Chromera* and Pmar_PMAR009211 in *Perkinsus*). The colpodellid transcriptome libraries also possess a broad range of 2 component sensory transducers, but these must be approached with caution due to possible bacterial contamination within the databases. In alveolates, these proteins possess multiple transmembrane domains, typically clustered at the N-terminus; a PAS domain (Pfam PF00989) in some versions; a histidine kinase domain; and a C-terminal response receiver domain. The proteins appear to lack signal peptides, although the N-terminal transmembrane domain might function as a transfer sequence. In prokaryotes, the two component systems are integrated in the bacteria membrane and transduce a variety of environmental signals, but in alveolates their cellular localization and function has not been determined. In protozoans, the two component receptors are not exclusive to alveolates and are also found in stramenopiles, fungi and plants; and thus their origin in alveolates might have arisen through vertical inheritance. These receptors are absent in Apicomplexa; perhaps because their function, albeit unknown in alveolates, became vestigial following commitment to obligate parasitism.

**CYSTEINE-RICH MODULAR PROTEIN (CRMP)**

Annotation of predicted extracellular proteins within the whole genome sequence information for *Plasmodium* revealed an amplified gene family with 4 members, each encoded protein having a structural theme of large arrays of a cysteine-rich modules in the extracellular domain; multiple transmembrane spanning domains; and a large, low complexity predicted cytoplasmic domain (Thompson et al. 2007; Douradinha et al. 2011). In addition to the cysteine-rich modules a single EGF-like domain, and in some versions an additional kringle domain were also present, leading to the name cysteine-rich modular protein (CRMP) for the family (e.g. PF3D7_0911300, PF3D7_1475400, PF3D7_0718300 and PF3D7_1208200 in *P. falciparum*). The recent description of a sialic-acid-binding module in some *Toxoplasma* microneme proteins (Friedrich et al. 2010) allows identification of a similar domain in the N-terminal region of apicomplexan CRMPs. Gene knockout studies in *Plasmodium* demonstrated that the genes are essential for transmission to mosquitoes, and the protein products appear to function in the transmission stages (Thompson et al. 2007; Douradinha et al. 2011). CRMPs are now known to
be present in all apicomplexans, with the exception of Cryptosporidium, as well as Perkinsus and are amplified in large families in the chromerids, colpodellids and ciliates (Fig. 2B). They also have a broader distribution in protozoa, such as stramenopiles, suggesting their presence in the last common ancestor of alveolates. The multi-transmembrane region is conserved across the phylogenetic distribution and often shows similarity to an ion channel termed the transient receptor potential (TRP) domain (reviewed in Venkatachalam and Montell, 2007). Thus a reasonable hypothesis is that the CRMP proteins participate in an environmental sensing role, with extracellular recognition of ligands and signalling across the membrane.

In chromerids, CRMP proteins are highly amplified, with approximately 40 members in Chromera. The fragmented nature of the colpodellid transcriptome libraries precludes an estimation of the extent of the gene amplifications, but they appear to have similar domain structures to examples in the chromerids. The ciliates Tetrahymena, Oxytricha and Stylonychia also have a highly amplified representation of CRMP proteins, with up to 100 genes within each genome. Thus, a great reduction in the number and variety of CRMP proteins accompanied the transition to the apicomplexan clade, with a complete loss of the genes in Cryptosporidium. This is perhaps in accordance with a role of CRMP proteins in recognition and response to the extracellular environment; one hypothesis being that the obligate parasitic apicomplexans might encounter a relatively defined environment, and thus do not require a broad repertoire of CRMP proteins. Perkinsus also appears to have a reduced number of CRMP proteins, and thus a correlation of CRMP proteins with life strategies would need to take into account the parasitic and free-living components of the Perkinsus life cycle; but also indicates a correlation with parasitism and loss of crmp genes.

CAST MULTI-DOMAIN PROTEIN
Numerous alveolates possess members of an amplified gene family, termed CAST multi-domain protein, which has not been functionally characterized. These giant proteins, ranging from several thousand amino acids in length, have architectures consisting of a large repeated array of cysteine-rich modules; a single transmembrane domain having conserved features; and a large (>150 kDa) presumed cytoplasmic domain having a low complexity, predicted coiled-coil character. The protein probably originated prior to the divergence of the alveolate lineage, since it is also found in stramenopiles and choanoflagellates. In the ciliate Oxytricha, the gene is highly amplified, with perhaps over 100 members (e.g. OXYTRI_15408); whereas in Toxoplasma there appear to be less than 5 genes encoding predicted CAST multi-domain proteins (e.g. TGME49_207480 and TGME49_253930), which are typically annotated as ‘GCC2 and GCC3 domain-containing proteins.’ Across the alveolates, apparent gene losses have shaped the phylogenetic distribution of the gene, and within the alveolates representatives of this protein are present in the ciliate Oxytricha, but not in Tetrahymena and Paramecium; in Chromera (e.g. Cvel_3066.t1), Vitrella and colpodellids, but absent in Perkinsus; in the coccidians, Eimeria and Toxoplasma; and absent in other apicomplexans such as Cryptosporidium, Theileria, Babesia and Plasmodium. The conserved sequence surrounding the transmembrane region suggests a conservation of a juxta-membrane function, such as interaction with the membrane or signalling. The cytoplasmic domain of the CAST multi-domain protein, which includes the namesake (and perhaps erroneously ascribed) CAST domain, appears to be conserved and is large (over 1500 aa), low complexity and with possibly with a coiled-coil structure. Toxoplasma is the obvious experimental organism in which to determine the cellular localization and function of the CAST multi-domain proteins.

OOCYST WALL PROTEIN
Cryptosporidium oocysts can be obtained in abundance and high purity following the experimental infection of a calf. Thus, Cryptosporidium is an excellent system in which one can study coccidian cyst structure (Spano et al. 1997; Chattrejee et al. 2010; Samuelson et al. 2013). An oocyst wall protein, termed OWP or Cryptosporidium oocyst wall protein (COWP), was purified from oocyst wall extracts and its protein sequence determined (Spano et al. 1997). The COWP protein is composed of repeats of variations of a highly cysteine-rich module. With the advent of whole genome sequence information it is now known that COWP genes are amplified in Cryptosporidium, which has 9 genes, and are also amplified in all cyst-forming coccidians and Gregarina (Fig. 3A; Templeton et al. 2004a; Templeton et al. 2010). OWP modules are also found in genes amplified in the chromerids (e.g. Vbra_11165.t1 in Vitrella and Cvel_20950.t1 in Chromera; Woo et al. 2015) and colpodellids, and thus, this component of the structure of the oocyst predates the specialization to Apicomplexa. OWP genes are not found in Perkinsus, dinoflagellates and ciliates, and thus their origin possibly occurred in the last common ancestor of the chromerids and apicomplexans. It has not been addressed if specific COWP genes are orthologously shared in the chromerids and colpodellids; nor if it is known if genes are vertically inherited as orthologues in apicomplexans, thus indicating possible conserved functions within the oocyst wall structure. The genes
are differentially amplified in Chromera versus Vitrella, and this might indicate that differing architectures underpin differing oocyst wall characters in the 2 closely related protozoans. Vitrella possesses as many as 30 OWP genes, whereas Cryptosporidium possesses 9 genes, emphasizing possible structural differences related to the number of encoded genes. Apicomplexans which do not have an externally shed oocyst stage have lost OWP genes; such as Plasmodium, Babesia and Theileria. Perkinsus lacks OWP genes and also does not possess a durable cyst stage; with only one report describing an apparently abundant ‘cell wall’ protein which is probably unrelated to cyst walls (Montes et al. 2002). The observation that OWP proteins are present in proto-apicomplexans provides markers with which to describe the great diversity in structures of inner and outer cell walls in the alveolates.

CHROMERIDS AND COLPODELLIDS AS COCCIDIANS

The conservation of the OWP in chromerids, colpodellids and coccidians, but their absence in Perkinsus and ciliates, is congruent with the known taxonomic affinity of chromerids and colpodellids with the Apicomplexa. Annotation of the predicted proteome of the chromerids revealed numerous additional predicted extracellular proteins having complex, multi-domain architectures which are shared with coccidians (Woo et al. 2015). Examples include the large, multi-domain protein TRAP-C2, first described in Cryptosporidium (perhaps erroneously named and not a TRAP family protein; Spano et al. 1998); a protein with a fusion of a MAM domain and a copper amine oxidase; and a transmembrane protein containing clostripain, notch and EGF domains (Fig. 3A). A first hypothesis might be that these proteins are involved in formation of the coccidian external oocyst, since they are not present in other apicomplexans, such as Plasmodium and Babesia, which lack external cyst stages; nor are they conserved in ciliates or Perkinsus. Thus, as suggested from phylogenetic trees, the last common ancestor of the apicomplexan lineage was coccidian-like; in that it possessed an environmentally-durable cyst stage. Two proteins, one with a HINT domain (e.g. C. parvum, cgd7_5290; Gregarina, GNI_039770; and Chromera, Cvel_10247.t1) and another encoding Fringe + Galactose transferase (e.g. C. parvum,
cgdf6.1450; and Chromera, Cvel_3306.t1), group the chromerids with Cryptosporidium or Gregarina to the exclusion of the coccidians and other Apicomplexa, thus providing support for placing Gregarina and Cryptosporidium at the base of the apicomplexan clade.

For colpodellids, it remains difficult to identify predicted orthologues of multi-domain extracellular proteins because the transcriptome databases are fragmented. For example, we have identified many copper amine oxidases in the colpodellid databases, but no fusions with a MAM domain, as described above. Another example is the presence of possible fragments, but no full-length TRAP-C2 orthologues. For this reason, it is of value to annotate colpodellid transcriptomes for the presence or absence of component extracellular domains, rather than survey for orthologues of large, complex multi-domain proteins. Here again the chromerids share extracellular domains with coccidians, to the exclusion of colpodellids; for example, the MAM domain as described above; a clostripain domain, found fused to Notch and EGF repeats in 1 coccidian and chromerid protein; and the TOX1 domain found fused to Notch and EGF repeats in 1 coccidian. Domain proteins. Here again the chromerids share extracellular domains, rather than component domains common to chromerids and colpodellids; for example, the transcriptome information might not have sufficiently high coverage, particularly across life cycle stages, for discussions of negative data.

CHROMERIDS AND COLPODELLIDS AS APICOMPLEXANS

Annotation of the P. falciparum genome revealed a family of proteins, termed CCP or LAP, having a rich multi-domain architecture of predicted sugar and lipid-binding domains (Pradel et al. 2004; Raine et al. 2007; Carter et al. 2008). Studies in P. falciparum and the rodent malaria parasite, Plasmodium berghei, indicate that the proteins function in sexual stage parasites, and gene disruption studies indicate a probable manifestation of phenotype in the mosquito midgut ookinete stage. Recent whole genome information indicates that the CCP/LAP genes are conserved as homologues not only across Apicomplexa, including Cryptosporidium and Gregarina, but also in the chromerids, Chromera and Vitrella (see Fig. 3 – figure supplement 4 in Woo et al. 2015). The colpodellids also possess the component domains of CCP/LAP proteins, although the fragmentation of the transcriptome sequence information makes determination of possible orthologous conservation of the multi-domain architectures. All members of the CCP/LAP family, as well as their component domains, are absent in ciliates and Perkinsus. Thus, any hypotheses of their function in Apicomplexa must also consider their function to be ancient, with orthologues present in the chromerids. The CCP/LAP proteins are predicted to be targeted to the crystalloid of Plasmodium ookinetes (Carter et al. 2008), in addition to extracellular secretion, and thus might serve as markers to determine if a similar organelle is present in all apicomplexans and chromerids. The cysteine-rich CPW_WPC domain family additionally present in all apicomplexans, with the exception of Cryptosporidium; is amplified in the chromerids and colpodellids; and is absent in Perkinsus and the ciliates. This protein family thus represents another marker with which to investigate conserved structures uniting proto-apicomplexans and apicomplexans.

The phylogenetic distribution of the component domains of predicted extracellular multi-domain proteins also group the chromerids with either ciliates, ciliates plus Perkinsus, coccidians, apicomplexans or all alveolates (Table 1). For example, component domains of the multi-domain CCP/LAP proteins (namely, ricin, NEC, SR, LCCL, as well as other domains) also have a phylogenetic distribution uniting the chromerids with Apicomplexa, to the exclusion of Perkinsus and the ciliates. Many of the extracellular domains common to chromerids and Apicomplexa are also found in metazoa, and thereby may have arisen through lateral transfer (Templeton et al. 2004b; Aravind et al. 2012). Table 2 illustrates the variety of domains and multi-domain architecture expansions, using chromerids as examples.

Not shown in Tables 1 and 2 are the numerous alveolate extracellular domains and proteins which appear to have been ‘invented’ de novo, in that they are genera- or species-specific. Some of these are discussed elsewhere in this review, such as presumptive saccharide-binding domains, and examples of the numerous highly amplified, anonymous protein families in the ciliates and dinoflagellates. Within P. falciparum examples of lineage-specific domains and proteins include the Duffy binding-like domain within PfEMP1 and EBA-175-like proteins which confer cytoadhesion of infected erythrocytes and recognition of erythrocyte during invasion, respectively; and the SURFIN, RIF and STEVOR proteins, which have unknown functions (Dzikowski et al. 2006; Frech and Chen, 2013). Parasite-encoded erythrocyte surface proteins also show species-specificity; for example, the SICAvir proteins found in Plasmodium knowlesi and other primate malaria parasites (al-Khedery et al. 1999; Frech and Chen, 2013; Lapp et al. 2013). Other genera-specific examples of domains and proteins are the highly amplified families of secreted FAINT domain proteins in Theileria (Pain et al. 2005) and the VESA erythrocyte surface antigens in Babesia (O’Conner et al. 1997; Jackson et al. 2014). In coccidians, the SAG and SRS proteins are perhaps the best examples of ‘inventions’ conferring host interactions, in this instance likely having an origin via lateral gene transfer, as described in a previous section. Such highly evolved lineage-
specific proteins may have conferred new host interactions which allowed exploitation by the parasite, followed by selection by functional and host immune response pressures which drove their diversification and amplification (for reviews see, e.g., Templeton, 2009; Mackinnon and Marsh, 2010; Smith et al., 2013; Jackson et al., 2014; Smith, 2014).
in many instances, one or multiple vWA domains. Additional hallmarks of TRAP/MIC2 proteins are a single transmembrane domain, in some instances with a juxta-membrane rhomboid protease cleavage site; a short, charged cytoplasmic domain; and C-terminal region aromatic residues which are thought to interact with cytoplasmic components of the motility apparatus. Figure 4A shows the variety of domain architectures from predicted TRAP/MIC2 members across the Apicomplexa. The chromerids, which appear to lack gliding motility, possess numerous predicted extracellular proteins harbouring TSP1 and vWA domains; thus the presence of these domains in the alveolates does not correlate with gliding motility. Indeed, Vitrella has an expansion of over 30 proteins harbouring TSP1 domains. Vitrella, Chromera and the colpodellid A. edax all possess proteins with 3 TSP1 domains followed by a C-terminal vWA domain; and the proteins appear to have an orthologous relationship based upon the presence of additional conservation throughout the sequence to the N-terminal side of the TSP1 domains (Fig. 4A). Thus, this TSP1 plus vWA domain architecture probably has a conserved function in the colpodellids and chromerids. However, none of these proteins appear to possess the additional hallmark TRAP/MIC2 features; namely, a transmembrane domain followed by a short, charged cytoplasmic domain having aromatic residues (qualifying here that gene models may not have been precisely determined for the chromerids and colpodellids).

Broad coverage genome sequence information has recently become available for the apicomplexan, G. niphandrodes, in which gliding motility is well described. Gregarines possess exquisite drapery-like longitudinal surface structures termed epicytic folds, which are proposed to be involved in gliding motility (reviewed in Valigurová et al. 2013). We were unable to identify clear homologues of TRAP/MIC2 proteins in the G. niphandrodes; however, the parasite does possess numerous genes encoding proteins having single vWA domains, including examples with signal peptides, C-terminal transmembrane domains and C-terminal aromatic residues (Fig. 4A and B). The gene predictions for G. niphandrodes appear to be preliminary and require validation, but the number may exceed 20 such TRAP-like proteins. The cousin of gregarines, Cryptosporidium, has multiple predicted TRAP/MIC2 proteins, although this protozoan lacks extracellular examples of vWA domains; rather, the
Cryptosporidium predicted TRAP/MIC2 proteins are composed of TSP1 and apple domains (Deng et al. 2002). One Cryptosporidium TSP1 domain protein, termed TRAP-C2, has a large array of TSP1 domains, plus Notch, TOX1 and CCP/Sushi domains, and a C-terminal transmembrane domain. However, this protein does not have TRAP/MIC2 features within the predicted cytoplasmic domain; namely, a charged character and C-terminal aromatic residues. TRAP-C2 is now known to be conserved as predicted orthologues in coccidians, gregarines, as well as chromerids (Fig. 3A; in Chromera, Cvel_23546.t1). The G. niphandrodes version differs in that the predicted cytoplasmic domain is charged and possesses C-terminal aromatic residues, and thus might be investigated as a candidate TRAP protein. The Cryptosporidium protein GP900, discussed above, has been implicated in cell invasion and is composed of extracellular arrays of a genera-specific, cysteine-rich domain; a single transmembrane domain; and a short, charged cytoplasmic domain with aromatic residues reminiscent of TRAP/MIC2 proteins. Tissue culture is unavailable for G. niphandrodes and C. parvum and thus limits genetic manipulation; however, a newly developed mouse model and gene manipulation method for C. parvum (Vinayak et al. 2015) shows great promise and might be used to characterize the function of potential TRAP/MIC2 proteins. Alternatively, the ability of candidate proteins to complement TRAP/MIC2 proteins might be tested in another system, such as Toxoplasma. If GP900 functions as a TRAP/MIC2 protein, despite its lack of TSP1 or vWA domains, then this would indicate that the prototypic features of a gliding motility receptor might be the TRAP/MIC2-like cytoplasmic domain. The TRAP/MIC2 architectural paradigm works well in identifying apicomplexan receptor candidates for mediating gliding motility, but possible proto-apicomplexan precursors to such proteins remain obscure since clear orthologous relationships are not apparent.

What can be said about the chromerids and colpodellids, as representative proto-apicomplexans, with respect to the innovation of gliding motility? One-to-one orthologous relationships of the intracellular components of glideosome proteins have not been conclusively identified (Woo et al. 2015), but related protein expansions have been observed for GAP40, GAP50, GAPM and ISP proteins. These sequence similarities did not extend to the ciliates, which supports the phylogenetic relationship of chromerids and apicomplexans. Greater understanding of the origin of gliding motility may come from refining proteomic and molecular studies to characterize candidate proteins, as well as obtaining ultrastructural, proteomic and whole genome sequence and transcriptome information for more proto-apicomplexan organisms. Describing the evolution of TRAP/MIC2 proteins awaits a better understanding of the function of gregarine and Cryptosporidium predicted receptor proteins.

Concluding remarks

Recently derived whole genome sequence information for the chromerids, C. velia and V. brassicaformis, and high coverage transcriptome information for colpodellids, supports their phylogenetic relationship with the Apicomplexa, and allows annotation with the goal of describing the molecular hallmarks of transition of a free-living alveolate to obligate parasitism in the Apicomplexa. The annotations described herein support the hypothesis that chromerids are more closely related to Apicomplexa than are the alveolates Perkinsus, dinoflagellates and ciliates, and thus far serve as the closest and best-described ‘outgroup’ in which we can study the transition to parasitism in the Apicomplexa. However, the chromerids also possess highly amplified families of predicted external sensory proteins uniting them with the dinoflagellates and ciliates. The great reduction or complete loss of orthologues for these families within Apicomplexa suggests, as one hypothesis, that obligate parasitism reduces the requirement for response to interacting with and interpreting the unpredictable and highly variable external environment. Conservation of numerous predicted extracellular proteins, such as the OWP domain-containing oocyst wall proteins, as well as complex multidomain proteins, between the chromerids and coccidians suggest that structural aspects of the cyst stage are conserved; that is, the chromerids can be viewed as ‘model coccidians’ rather than grouping with dinoflagellates or ciliates. Other conserved extracellular proteins, such as the LCCL and CPW_WPC domain containing proteins and numerous extracellular domains also group the chromerids with all apicomplexans. The chromerids have provided few clues towards understanding the development of gliding motility in the apicomplexans, although some glideosome proteins appear to have origins prior to the transition to Apicomplexa. Further functional and systems biology studies, such as in the gregarines and proto-apicomplexans, are required to unravel the steps which occurred in the evolution of gliding motility in the apicomplexans.

Acknowledgements

The authors would like to thank Bruce Taylor and Richard Culleton for their critical readings of the paper.

Financial Support

T.J. Templeton would like to acknowledge the generous support of a Visiting Professorship at the Institute of Tropical Medicine (NEKKEN), Nagasaki University, Japan. Research in A. Pain’s research group is supported
Briguglio, J. S. and Turkewitz, A. P. (2014). Tetrathyridia thermophila: a divergent perspective on membrane traffic. Journal of Experimental Zoology 322B. 500–516.

Brugerolle, G. (2002). Colpodella vorax: ultrastructure, predation, life-cycle, mitosis, and phylogenetic relationships. European Journal of Protistology 38. 1113–125.

Byrne, B. C. (1973). Mutational analysis of mating type inheritance in Syngon 4 of Paramecium aurelia. Genetics 74. 63–80.

Cadetti, L., Marroni, F., Marangoni, R., Kuhlmann, H.-W., Glocer, D. and Colomboletti, G. (2000). Phototaxis in the ciliated protozoan Ophryoglena flava: dose-effect curves and action spectrum determination. Journal of Photochemistry and Photobiology B: Biology 57. 41–50.

Caron, F. and Meyer, E. (1989). Molecular basis of surface antigen variation in paramecia. Journal of Structural Biology 43. 23–42.

Carter, V., Shimizu, S., Arai, M. and Dessens, J. T. (2008). PbsR is synthesized in macrogametocytes and involved in formation of the malaria crystalloids. Molecular Microbiology 68. 1560–1569.

Casée, O., Dub, Vaidya, A. B., Wilkes, D. E., Rajaith, V., Bouts, D. and Lebrun, M. (2005). Synergistic role of macromolecular proteins in Toxoplasma gondii virulence. Journal of Experimental Medicine 201. 453–463.

Chatterjee, A., Bannejee, S., Steffen, M., Conner, R. M., Ward, H. D., Robinson, P. W. and Bitoon, J. (2012). Evidence for mucin-like glycoproteins that tether sporozoites of Cryptosporidium parum to the inner surface of the oocyst wall. Eukaryotic Cell 9. 84–96.

Cornillot, E., Hadj-Kaddour, K., Dossouli, A., Noel, B., Ranzew, V., Vacherie, B., Augagneur, Y., Bres, V., Duclos, A., Randazzo, S., Carrey, B., Dehier-Groocke, F., Delong, S., Moubray, K., Shams-El-Hin, S., Usman, Brown, S., Brincaud, P., Wincker, P., Vivarès, C. P., Schwarz, R. T., Schetters, T. P., Krause, P. J., Gorenflo, A., Berry, V., Barbe, V. and Ben Mamoun, C. (2012). Sequencing of the smallest apicomplexan genome identifies the human pathogen Babesia microti. Nucleic Acids Research 40. 9102–9114.

Coss, C. A., Robledo, J. A. F. and Vasta, G. R. (2001). Fine structure of clonally propagated in vitro life stages of a Perakisus sp. Isolated from the Baltic clam Macoma balthica. Journal of Eukaryotic Microbiology 48. 38–51.

Coyne, R. S., Hannon, L., Shannumag, D., Hostetler, J. B., Brami, D., Jobard, V. S., Johnson, D., Radune, S., Singh, I., Badger, J. H., Kumar, U., Saier, M., Wang, Y., Gu, J., Mistry, M. W., Joshi, D., Wilkes, D. E., Rajaith, V., Asai, D. J., Pearson, C. G., Findly, R. C., Dickerson, H. W., Wu, M., Martens, C., Van de Peer, Y., Roos, D. S., Cassidy-Hanley, D. M. and Clark, T. G. (2011). Comparative genomics of the pathogenic ciliate Ichthyophthirius multifiliis, its free-living relatives and a host species provide insights into adoption of a parasitic lifestyle and prospects for disease control. Genome Biology 12. R100.

Daher, D. W. and Soldati-Favre, D. (2009). Mechanisms controlling glicose function in apicomplexa. Current Opinion in Microbiology 12. 463–464.

Deng, M., Templeton, T. J., London, N. R., Bauer, C., Schroeder, A. A. and Abrahamson, M. S. (2002). Cryptosporidium parum genome genes containing thiomponent type 1 domains. Infection and Immunity 70. 6987–6995.

Douradinha, B., Ausstijn, K. D., Moore, S. G., Ramesar, J., Mata, M. M., Waters, A. P., Janse, C. J. and Thompson, J. (2011). Plasmodium cyaenine repeat modular proteins 3 and 4 are essential for malaria parasite transmission from the mosquito to the host. Malaria Journal 10. 71.

Dzikowski, R., Templeton, T. J. and Deitsch, K. (2006). Variant antigen gene expression in malaria. Cellular Microbiology 8. 1371–1381.

Eckert, R. (1972). Bioelectric control of ciliae activity. Science 176. 473–481.

Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortmann, J. R., Badger, J. H., Ren, Q., Amedeo, P., Jones, K. M., Tallon, I. J., Delcher, A. L., Salzberg, S. L., Silva, J. C., Haas, B. J., Majoros, W. H., Frazar, M., Carlton, J. M., Smith, R. K. J., Garg, J., Pearlman, E. R., Karrer, K. M., Sun, L., Manning, G., Elde, N. C., Turkeiwit, A. P., Asai, D. J., Wilkes, D. E., Wang, Y., Cai, H., Collins, K., Stewart, A. B., Lee, S. R., Willmam, K., Weinberg, Z., Ruzzo, W. L., Wiga, D., Gaertig, J., Frankel, J., Tsao, K. C., Gorovyc, M. A., Keeling, P. J., Waller, R. F., Patton, N. J., Cherry, J. M., Stover, N. A., Krieger, C. J., de
Extracellular proteins and alveolate evolution

Babesia reveals a history of genomic innovation underlying host-parasite interaction. Nature Chemical Biology 42, 7113–7113.

Janouškovec, J., Tikhonenkov, D. V., Burki, F., Howe, A. T., Koliško, M., Mylnikov, A. P. and Keeling, P. J. (2015). Factors mediating plant dependency and the origins of parasitism in apicomplexans and their close relatives. Proceedings of the National Academy of Sciences of the United States of America Online Early Edition 112, 10200–10207.

Kafšák, B. F. C. and Carruthers, V. B. (2012). Apicomplexan perforin-like proteins. Communicative and Integrative Biology 3, 18–23.

Knoell, G., Haacke-Bell, B. and Plattner, H. (1991). Local trichocyst exosorption provides an efficient escape mechanism for Paramecium cells. European Journal of Protistology 27, 381–385.

Lampert, T. J., Coleman, K. D. and Hennessey, T. M. (2011). A knockout mutation of a constitutive GPCR in Tetrahymena decreases both G-protein activity and the origins of parasitism in apicomplexans and their close relatives. Proceedings of the National Academy of Sciences of the United States of America Online Early Edition 112, 10200–10207.

Lapp, S. A., Korir-Morrisson, C., Jiang, J., Bai, Y., Corredor, V. and Galinski, M. R. (2013). Spleen-dependent regulation of antigen variation in malaria parasites. Plasmodium knowlesi SCAVr expression profiles in spleen in Plasmodium berghei infection. Journal of Proteomics 102, e84653.

Leander, B. S., Kuvardina, O. N., Aleshin, V. V., Mylnikov, A. P. and Keeling, P. J. (2003). Molecular phylogeny and surface morphologies of Colpodella edax (Alveolata): insights into the phagotrophic ances- trophic complex. Journal of Eukaryotic Microbiology 50, 334–340.

Liu, Y., Tewari, R., Ninj, J., Blagborough, A. M., Garhov, S., Pei, J., Grishin, N. V., Steele, R. E., Sinden, R. E., Snell, W. J. and Billiker, O. (2008). The conserved plant gene family HAP2 after functions at partitioning of fusogenic membranes in Chlamydomonas and Plasmodium. Genes and Development 22, 1051–1063.

Lobban, C. S., Hallam, S. J., Mukherjee, P. and Petrich, J. W. (2007). Photostability and multifunctionality of hypericin-like pigments in heterotrich ciliates: a phlogistic perrenial? Photochemistry and Photobiology 83, 1074–1094.

Lourido, S., Shuman, J., Zhang, C., Shokat, K. M., Hui, R. and Sibley, L. D. (2010). Calcium-dependent protein kinase 1 is an essential regulator of exocytosis in Toxoplasma. Nature 465, 359–363.

Mackinnon, M. J. and Marsh, K. (2010). The selection landscape of malaria parasites. Science 328, 866–871.

Mai, K., Smith, N. C., Feng, Z.-P., Katrih, M., Slapeta, J., Slapatová, I., Wallach, M. G., Davies, M. J., Zhang, X., Norton, R. S. and Belli, S. L. (2011). Peroxidase catalyzes cross-linking of an intrinsically unstructured protein via dye-dyrosine bonds in the oocyst wall of the apicomplexan parasite, Eimeria maxima. International Journal for Parasitology 41, 1157–1166. http://dx.doi.org/10.1016/j.ijpara.2011.07.019.

Martel, C. M. (2009). Conceptual bases for prey biorecognition and feeding selectivity in the microplanktivore marine phagophag Otysryxis marina. Microbial Ecology 57, 589–597.

McCoy, J. M., Whitehead, L., van Dooren, G. G. and Tonkin, C. J. (2012). TheCDPK regulon is required for Toxoplasma gondii host cell invasion. Journal of Immunology 188, 4510–4519.

Montes, J. F., Durfort, M., Llado, A. and Garcia-Valero, J. (2002). Characterization and immunolocalization of a main proteinaceous component of the cell wall of the protozoan parasite Perkinsus marinus. Parasitology 124, 477–486. http://dx.doi.org/10.1017/S0031182002005924.

Moore, R. B., Obornik, M., Janouškovec, J., Chudinský, T., Vancová, M., Vancová, L., Ford, S. W., Davies, N. W., Bolch, C. J. S., Heimann, K., Slapeta, J., Hoegh-Guldberg, O., Logsdon, J. M. and Carter, D. A. (2008). A photosynthetic alveolate closely related to apicomplexan parasites. Nature 451, 959–963.

Morahan, B. J., Wang, L. and Coppell, R. L. (2009). No TRAP, no invasion. Trends in Parasitology 25, 77–84.

Obornik, M., Vancová, M., Lai, D-H., Janouškovec, J., Keeling, P. J. and Lukš, J. (2011). Morphology and ultrastructure of multiple life cycle stages of the photosynthetic relative of Apicomplexa, Chlamydomonas vohrei. Protist 162, 115–130.

Obornik, M., Modry, D., Lukš, M., Cernotíková-Stibníčová, E., Cihlář, J., Tesárová, M., Kotábová, E., Vancová, M., Prášil, O. and Lukš, J. (2012). Morphology, microanatomy and life cycle of Vitella basi- sicaformis n. sp., n. gen., a novel Chlorella from the Great Barrier Reef. Protist 163, 306–323.

O’Connor, R. M., Lane, T. J., Stoup, S. E. and Allred, D. R. (1997). Characterization of a variant erythrocyte surface antigen (VESA1) expressed by Babesia microti during antigenic variation. Molecular and Biochemical Parasitology 89, 259–270.

Okamoto, N. and Keeling, P. J. (2014). The 3D structure of the apical complex and the 3D associated complex with the flagellar apparatus revealed by serial TEM tomography in Plasmollama pacifica, a distant relative of the Apicomplexa. PLoS ONE 9, e84653.
Pain, A., Renaud, H., Berriman, M., Murphy, L., Yeats, C. A., Weir, W., Kerhornou, A., Aslett, M., Bishop, R., Bouchier, C., Cochet, M., Coulson, R. M., Cronin, A., de Villiers, E. P., Fraser, A., Fosker, N., Gardner, M., Gehle, A., Griffiths-Jones, S., Harris, D. E., Katzer, F., Larke, N., Lord, A., Maser, P., McKellar, S., Mooney, M., Morton, F., Nene, V., O’Neill, S., Price, C., Quail, M. A., Robbins, W. K., Rawlings, N. D., Rutter, S., Saunders, D., Seeger, K., Shah, T., Squares, R., Squares, S., Stott, N., Sung, K. J., Walkden, B., Wang, K. L., Watanabe, R., Watanabe, T., Watson, N., Weir, W., Kerhornou, A., Aslett, M., Bishop, R., Masen, P., Stover, N. A., Jones, T. A., Eddy, S. R., Herrick, G. A., Doak, T. G., Wilson, R. K., Mardis, E. R. and Landweber, L. F. (2013). The Oxysticha trisulis macronucleus genome: a complex eukaryotic genome with 16,000 tiny chromosomes. PLoS Biology 11, e1001473.

Templeton, T. J. (2008). The surface protein repertoires of Cryptosporidium spp. and other apicomplexans. In Giardia and Cryptosporidium: From Molecules to Disease (ed. Ortega-Piernas, G., Cacciò, S., Fayer, R., Mank, T. G., Smith, H. V. and Thompson, R. C. A.), pp. 369–381. CABI Publishing. Wallingford, England.

Templeton, T. J. (2009). The varieties of gene amplification, diversification and hypervariability in the human malaria parasite, Plasmodium falciparum. Mol. Microbiol. 66, 109–116.

Templeton, T. J., Lancot, C. A., Guldurghiev, V., Liu, C., London, N. R., Hadsall, C., Rahman, S. S., Anderson, K. G., Doerig, C., Gruber, A., Ivens, A. C., Parkinson, J., Rajandream, M. A., Shirley, M. W., Wang, K. L., Berriman, M., Tomley, F. M. and Pain, A. (2014). Genomic analysis of the causative agents of coccidiosis in domestic chickens. The Journal of Experimental Biology 217, 1467–1485.

Portman, N. and Slapeta, J. (2014). The flagellar contribution to the apical complex: a new tool for the eukaryotic Swiss Army knife? Trends in Parasitology 30, 58–54.

Pradel, G., Hayton, K., Aravinid, L., Iyer, L. M., Abrahamson, M. S., Bonavitz, A. A., Mejia, C. and Templeton, T. J. (2004). A multidomain adhesion protein family expressed in Plasmodium falciparum is essential for transmission to the mosquito. Journal of Experimental Medicine 199, 1533–1544.

Raine, J. D., Ecker, A., Mendoza, J., Tewari, R., Stanway, R. R. and Sinden, R. E. (2007). Female inheritance of malarial lap genes is essential for mosquito transmission. PLoS Pathogens 3, e30.

Björnsson, K., Pearson, A. H., Rantanen, K., Browne, H. P., Bryant, J., Dunn, M., Hung, S. S., Kawahara, F., Miranda-Saaavedra, D., Malas, T. B., Mourier, T., Naghra, H., Nair, M., Otto, T. D., Rawlings, N. D., Rivailler, P., Sanchez-Flores, A., Sanders, M., Subramanian, C., Tay, Y. L., Woo, Y. W., Wu, X., Badger, J. G., Doerig, C., Gruber, A., Ivens, A. C., Parkinson, J., Rajandream, M. A., Shirley, M. W., Wan, K. L., Berriman, M., Tomley, F. M. and Pain, A. (2014). Genomic analysis of the causative agents of coccidiosis in domestic chickens. Genome Research 24, 1476–1485.

Robert, E. C., Zuhov, M. V., Martin-Cereceda, M., Novarino, G. and Woetstyn, E. C. (2007). Cell surface lectin-binding glycoconjugates on marine planktonic protists. FEMS Microbiology Letters 265, 202–207.

Samuelson, J., Bushkin, G. G., Chatterjee, A. and Robbins, P. W. (2013). Strategies to control the structural components of cyst and oocyst walls. Eukaryotic Cell 12, 1578–1587.

Sanders, P. R., Gilson, P. R., Cantin, G. T., Greenbaum, D. C., Nebl, T., Carucci, D. J., McConville, M. J., Schofield, L., Hodder, A. N., Yates, J. R., III and Crabb, B. S. (2005). Distinct protein classes including novel surface antigens in red blood cell stages of Plasmodium falciparum. Journal of Biological Chemistry 280, 40169–40176.

Selbach, M. and Kuhlmann, H. W. (1999). Structure, fluorescent properties and function presented in phototaxis of the stigma apparatus in the ciliate Uhlamysdomon muenomyoeus. The Journal of Experimental Biology 202, 919–922.

Simpson, A. G. G. and Patterson, D. J. (1996). Ultrastructure and identification of the predatory flagellate Colpodella pugnax Cieckowsk (Apicomplexa) with a description of Colpodella turris sp. nov. and a review of the genus. Systematic Parasitology 33, 187–198.

Singh, D. P., Saudemont, B., Guglielmi, G., Aariz, O., Goût, J. F., Prajer, M., Potechkin, A., Przybys, E., Aubusson-Fleury, A., Bhullar, S., Bouhouke, K., Huhllier-Akakpo, M., Tanty, V., Bluegeon, C., Alberti, A., Labadie, K., Aury, J. M., Sperling, L., Duband, J., M. and Meyer, E. (2014). Genome-defense small RNAs exapted for epigenetic mating-type inheritance. Nature 509, 447–452.

Sinnis, P. and Coppi, A. (2007). A bacterial proteorhodopsin proton pump in a phylogenetically diverse bacterial proteorhodopsin. nature 452, 41–44.

Soudant, P., Cha, F.-L. E. and Volety, A. (2013). Host-parasite interactions: marine bivalve molluscs and protozoan parasites, Perkinseus species. Journal of Invertebrate Pathology 114, 196–216.

Smith, J. D. (2014). The role of FIE1MP1 adhesion domain classification in Plasmodium falciparum pathogenesis research. Molecular and Biochemical Parasitology 195, 82–87.
Janouškovec, J., Katris, N. J., Mast, F. D., Miranda-Saavedra, D., Mourier, T., Naeem, R., Nair, M., Panigrahi, A. K., Rawlings, N. D., Padron-Regalado, E., Ramaprasad, A., Samad, N., Tomčala, A., Wilkes, J., Neafsey, D. E., Doerig, C., Bowler, C., Keeling, P. J., Roos, D. S., Dacks, J. B., Templeton, T. J., Waller, R. F., Lukeš, J., Oborník, M. and Pain, A. (2015). Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. eLife 4:e06974. doi: 10.7554/eLife.06974.

Wood-Charlson, E. M., Hollingsworth, L. L., Krupp, D. A. and Weis, V. M. (2006). Lectin/glycan interactions play a role in recognition in a coral/dinoflagellate symbiosis. Cellular Microbiology 8, 1985–1993.

Wood-Charlson, E. M., Hollingsworth, L. L., Krupp, D. A. and Weis, V. M. (2006). Lectin/glycan interactions play a role in recognition in a coral/dinoflagellate symbiosis. Cellular Microbiology 8, 1985–1993.

Wootton, E. C., Zubkov, M. V., Jones, D. H., Jones, R. H., Martel, C. M., Thornton, C. A. and Roberts, E. C. (2007). Biochemical prey recognition by planktonic protists. Environmental Microbiology 9, 216–222.

Xu, P., Widmer, G., Wang, Y., Ozaki, L. S., Alves, J. M., Serrano, M. G., Puiu, D., Manque, P., Akiyoshi, D., Mackey, A. J., Pearson, W. R., Dear, P. H., Bankier, A. T., Peterson, D. L., Abrahamsen, M. S., Kapur, V., Tzipori, S. and Buck, G. A. (2004). The genome of Cryptosporidium hominis. Nature 431, 1107–1112.