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

THOMAS J. TEMPLETON[1,2]* *and* ARNAB PAIN[3,4]

[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 perkinsid *Perkinsus marinus*; the apicomplexan, *Gregarina niphandrodes*, 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 repertoires 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

* Corresponding author: Department of Protozoology, Institute of Tropical Medicine (NEKKEN), Nagasaki University, 1-12-4 Sakamoto Nagasaki 852-8523, Japan and Department of Microbiology and Immunology, Weill Cornell Medical College, New York 10021, USA. E-mail: tjt2001@med.cornell.edu

Parasitology (2016), **143**, 1–17.
doi:10.1017/S0031182015001213

(Brugerolle, 2002). 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; Oborník *et al*. 2011) and *Vitrella* (Oborník *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;

https://doi.org/10.1017/S0031182015001213 Published online by Cambridge University Press

Hamel *et al*. 2011). A G-protein-coupled receptor having 7-transmembrane domains was characterized in *Tetrahymena* and shown to be involved in chemoattraction (Lampert *et al*. 2011). Some alveolates, such as in *Nassula citrea*, have evolved eyespots, and even a primitive lens termed an *ocelloid*, as in the dinoflagellates, *Erythropsidinium* and *Nematodinium* (Gomez, 2008; Hayakawa *et al*. 2015; Gavelis *et al*. 2015). The dinoflagellate *Oxyrrhis marina* (Slamovits *et al*. 2011), as well as chromerids and colpodellids, possess amplified gene families of rhodopsin-like proteins, such as exemplified by Cvel_15171.t1 in *Chromera*, Vbra_19386.t1 in *Vitrella* and BE-2_cDNA_131008@a107687_15 in *Alphamonas edax*.

Self–self recognition during fertilization or ciliate conjugation is another physical cell–cell interaction which appears to be mediated by surface proteins, such as the mating type surface protein mtA in *Paramecium* (Sonneborn, 1938; Byrne, 1973; Singh *et al*. 2014). MtA is 1275 amino acids long (e.g. XP_001450586·1 in *Paramecium tetraurelia*) and has a structure of a N-terminal signal peptide, multiple transmembrane domains at the carboxyl terminus, and predicted extracellular cysteine-rich furin-like domains adjacent the transmembrane region. In *P. tetraurelia*, mtA appears to have 3 or more possible paralogues of similar architecture, although it is not known if these proteins participate in mating recognition or have other functions. Multi-transmembrane domain proteins with predicted extracellular furin-like domains are found within amplified families in other ciliates, such as *Tetrahymena* (e.g. TTHERM_01337410) and *Oxytricha*; but without functional studies it is not possible to speculate if some of these proteins participate in conjugation. Self-recognition proteins have also been described for gamete fertilization in *Plasmodium*; namely, members of an amplified gene family which encodes glycophosphatidyl inositol (GPI)-linked proteins composed of 6-cys domains (Van Dijk *et al*. 2010) and the HAP2 protein which functions in membrane fusion (Liu *et al*. 2008). The 6-cys domain proteins are related to the *Toxoplasma* and *Eimeria* GPI-linked surface coat SAG proteins, and have been proposed to have originated via lateral transfer of metazoan ephrin proteins (Gerloff *et al*. 2005; Arredondo *et al*. 2012; Reid *et al*. 2014). Ten members of the family are present in *Plasmodium* and appear to have roles in multiple stages of the life-cycle, including hepatocyte and intraerythrocytic stages (Ishino *et al*. 2005; Sanders *et al*. 2005; Annoura *et al*. 2014).

Surface proteins mediating recognition of the environment might either be anchored within the surface membrane by multiple transmembrane domains, such as in rhodopsins or signalling channels; single transmembrane domains, such as TRAP/MIC2 receptors described in a section below; tethering to membranes by GPI moieties, including ciliate immobilization antigens and the circumsporozoite coat protein of *Plasmodium* sporozoites; or via interaction with other membrane-anchored proteins, such as the *Plasmodium* gamete surface protein, P230 (termed Pfs230 in *Plasmodium falciparum*; Williamson *et al*. 1993). Globular domains within 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 TTHERM_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*. 2015). The placements of specific genera within this general classification of Alveolata is an ongoing work,

https://doi.org/10.1017/S0031182015001213 Published online by Cambridge University Press

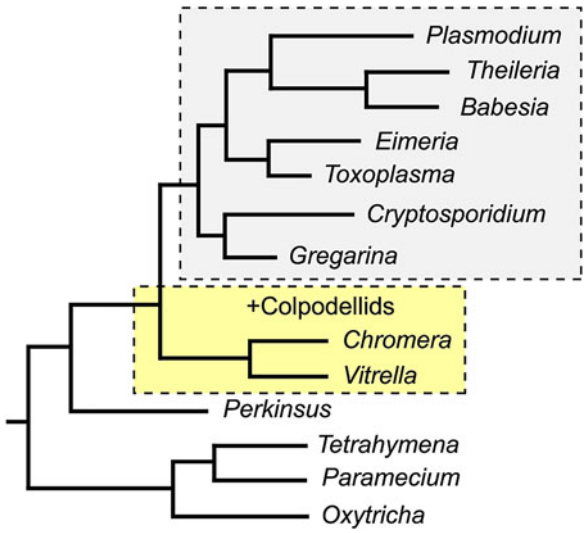

Fig. 1. Phylogenetic tree of the alveolates, showing relationships with the Apicomplexa (generalized from Templeton *et al*. 2010; Janouškovec *et al*. 2015; Woo *et al*. 2015). The apicomplexan clade is indicated by a grey box. The grouping indicated by the yellow box shows the relationship of the *Chromera, Vitrella* and the colpodellids; and as a sister clade to the Apicomplexa and thereby putative model 'proto-apicomplexans'.

requiring 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 (Janouškovec *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* (Abrahamsen *et al*. 2004; Xu *et al*. 2004) and numerous *Plasmodium* species (available and described at the online database resource, http://www.plasmodb.org).

GLYCOSYLATION, 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 (Céréde *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

https://doi.org/10.1017/S0031182015001213 Published online by Cambridge University Press

Table 1. Phylogenetic distributions for select alveolate extracellular domains[a]

Domain	Domain ID[b]	Ciliates	*P. mar.*[c]	*C.vel.*[c]	*V. brass.*[c]	*Colp.*[d]	*C. parv.*[c]	Coccidians	*Plasmodium*
FN3	cd00063								
IG_FLMN	cl19759					?			
IPT/TIG	cl15674								
Myxo disulfo repeat	cl11785[e] TIGR02232								
G8	pfam10162								
ChtBD1	cd00035					?			
WSC	cl02568								
HYR	cl03620[e]					?			
HINT	cd00081		Nuclear version?						
PAN-APPLE	cl00112								
MAM	cd06263					?			
CAO	pfam01179								
COWP	–[f]								
TOX1	–[g]					?			
Notch	cl02419								
RICIN	cd00161								
NEC	cl02436								
SR	cl02509								
Laminin G3/ Pentraxin	cl00102								
LCCL	cl02694	*Oxytricha* ?							
CCP (Sushi)	cl00043								
DISCOIDIN	Pfam00754								
TSP1	cl15278								
PA14	smart00758								
vWA	cl00057						*C. muris (1 gene?)*		
MacP F	cl02616								
TRP	cl18480[e]								
GCC2/GCC3	pfam07699								
EGF	cd00054								

[a] White boxes indicate the absence and orange-shaded boxes indicate the presence of the domain (rows) in the genome or transcriptome 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 transcriptome 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*. (2004*b*); 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*.'

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

https://doi.org/10.1017/S0031182015001213 Published online by Cambridge University Press

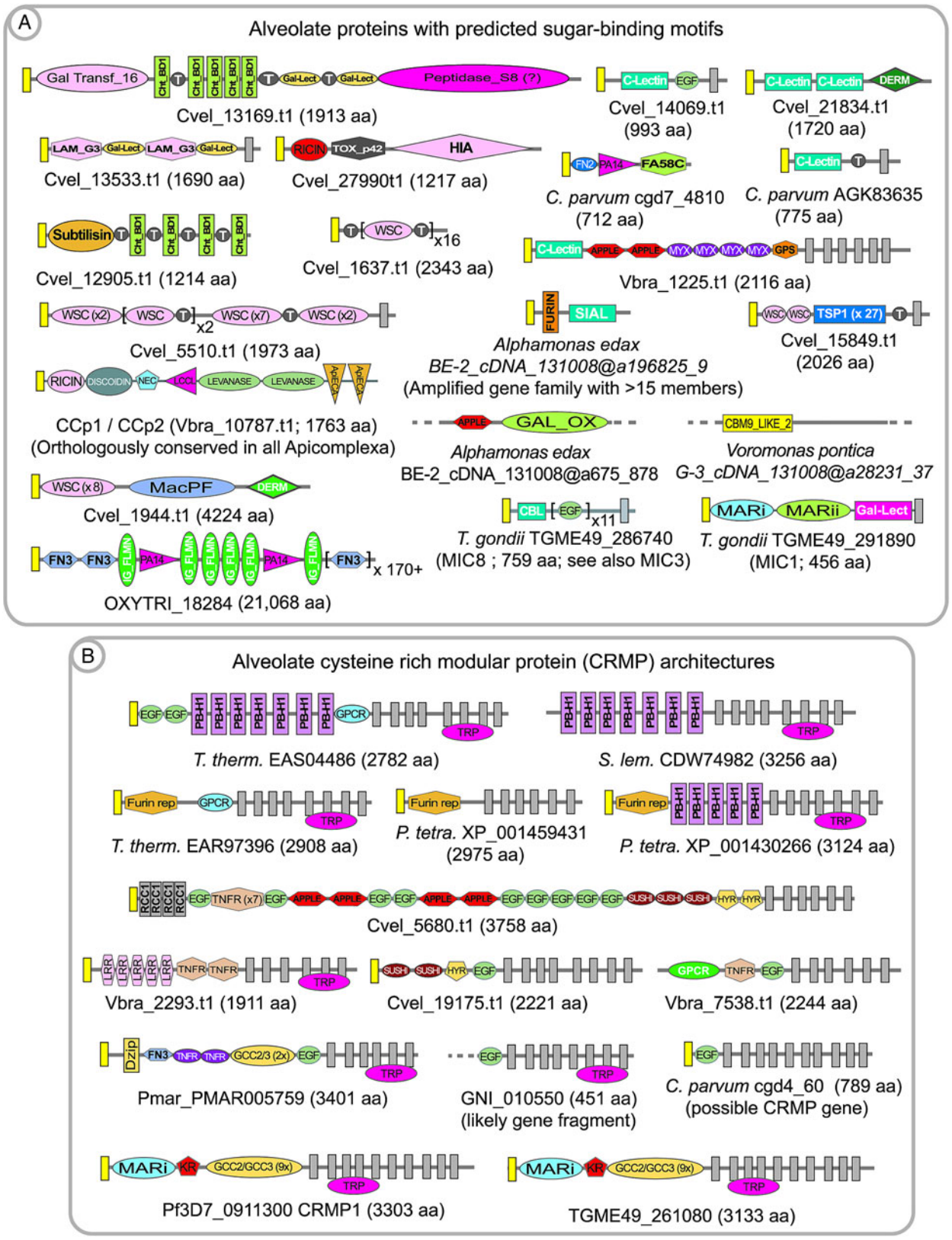

Fig. 2. Examples of predicted alveolate extracellular proteins having sugar-binding domains or mucin-like threonine repeat domains (A) and cysteine-rich modular protein (CRMP) architectures (B). Signal peptides are indicated by yellow boxes; and transmembrane domains are indicated by dark grey boxes. 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.

https://doi.org/10.1017/S0031182015001213 Published online by Cambridge University Press

(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*. 2004*b*; 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

https://doi.org/10.1017/S0031182015001213 Published online by Cambridge University Press

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*. 2004*a*; 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

https://doi.org/10.1017/S0031182015001213 Published online by Cambridge University Press

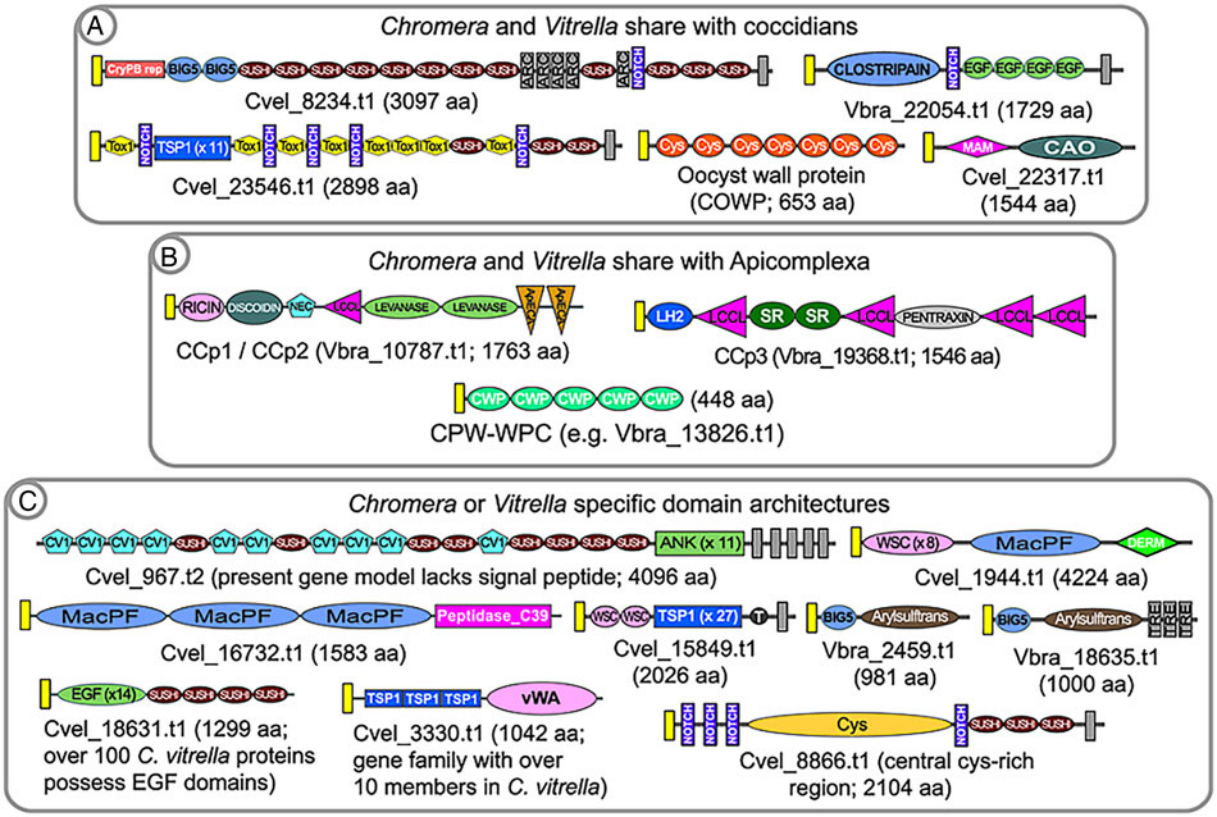

Fig. 3. Examples of predicted alveolate extracellular proteins that unite *Chromera* and *Vitrella* to coccidians (A) or Apicomplexa (B), to the exclusion of ciliates and dinoflagellates. Because of the fragmented nature of the colpodellid databases it is not possible to include them in these analyses; rather, they are discussed in the text. Examples of *Chromera* or *Vitrella*-specific domain architectures are shown in (C). Signal peptides are indicated by yellow boxes; and transmembrane domains are indicated by dark grey boxes. Domain abbreviations refer to either SMART (http://smart.embl-heidelberg.de/browse.shtml) or pfam (http://pfam.xfam.org). See also domain abbreviations and identifiers in Table 1.

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*,

https://doi.org/10.1017/S0031182015001213 Published online by Cambridge University Press

cgd6_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 (Table 1). However, it is important to obtain complete genome sequence information for one or more colpodellids, because 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 metazoans, and thereby may have arisen through lateral transfer (Templeton *et al*. 2004*b*; 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 SICAvar 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-

https://doi.org/10.1017/S0031182015001213 Published online by Cambridge University Press

Table 2. Examples of *Vitrella* and *Chromera* multi-gene families encoding predicted secreted proteins and component domains

Preliminary domain annotation	Number of genes[a]	Gene examples
Cysteine-rich modular protein family (CRMP)	>50	Vbra_20580.t1; Vbra_1972.t1; Vbra_13707.t1; Vbra_19335.t1; Vbra_2892.t1; Vbra_7649.t1; Cvel_5680.t1; Cvel_19175.t1
CAST domain protein family	10	Vbra_19068.t1; Vbra_18265.t1; Vbra_17728.t1; Vbra_19445.t1; Vbra_21566.t1; Cvel_3066.t1; Cvel_1096.t1
CCP (sushi) repeats	32	Vbra_14682.t1; Vbra_16381.t1; Vbra_18256.t1; Cvel_16357.t1; Cvel_17481.t1
Subtilisin; some examples have APPLE, EGF, vWA, and possibly MAM domains	24	Vbra_14867.t1; Vbra_10459.t1; Vbra_4853.t1; Vbra_16127.t1; Vbra_10049.t1; Vbra_14853.t1
COWP domains	32	Vbra_11165.t1; Vbra_9156.t1; Vbra_19321.t1; Cvel_26466.t1; Cvel_30099.t1
SCP domains	6	Vbra_13572.t1; Vbra_21604.t1; Vbra_18146.t1; Cvel_15549.t1; Cvel_19684.t1; Cvel_12695.t1
Arylsulfotransferase; some examples have APPLE or BIG5 domains	62	Vbra_5814.t1; Vbra_10419.t1; Vbra_17229.t1; Vbra_20167.t1; Vbra_310.t1; Vbra_643.t1
Insulinase	8	Vbra_11072.t1; Vbra_11062.t1; Vbra_13953.t1
APPLE domains; with a variety of companion domains, including glycohydrolases, and variable number of transmembrane domains	>30	Vbra_12089.t1; Vbra_12469.t1; Vbra_5849.t1; Vbra_17172.t1; Vbra_1188.t1; Vbra_2074.t1; Cvel_5441.t1; Cvel_2650.t1; Cvel_14497.t1
TSP1; many examples with vWA or a variety of other domains	>30	Vbra_3936.t1; Vbra_7675.t1; Vbra_13616.t1; Vbra_12027.t1;Vbra_5439.t1; Vbra_14095.t1
Glycohydrolase	9	Vbra_11140.t1; Vbra_3374.t1; Vbra_9561.t1
Ferritin; possible secreted iron-binding protein	6	Vbra_11299.t1; Vbra_2699.t1; Vbra_5212.t1; Vbra_10559.t1; Vbra_2704.t1
Fasciclin repeats	15	Vbra_6840.t1; Vbra_7979.t1; Vbra_14724.t1; Cvel_23968.t1; Cvel_5831.t1
CPW_WPC domains	13	Vbra_8795.t1; Vbra_13826.t1; Vbra_18088.t1; Cvel_23115.t1
EGF, with a variety of companion domains	>20	Vbra_11245.t1; Vbra_13334.t1; Vbra_13332.t1; Vbra_13334.t1; Cvel_20083.t; Cvel_167.t1

[a] Number of genes in *Vitrella* with a specific domain, or the number of paralogues within the relevant gene family. Some proteins have repeats of a specific domain, such as arrays of TSP1, Sushi or COWP domains.

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).

ON THE ORIGIN OF GLIDING MOTILITY IN APICOMPLEXANS

Arguably the singular revolution in the transition to obligate parasitism in the Apicomplexa was the development of gliding motility as a means to facilitate tissue traversal and cell invasion. Alveolates use flagella for motility, typically as flagella pairs in the case of the dinoflagellates and chromerids; in rows of cilia, such as in the namesake ciliates; or the combination of cilia and single apical flagella, such as described in the elegant *Ileonema simplex*. Gliding motility, however, is unique to apicomplexans, although flagellar motility has been retained in the microgamete stages of *Plasmodium*. Apicomplexan gliding motility has been well described elsewhere (e.g. Daher and Soldati-Favre, 2009; Frénal *et al*. 2010; Jacot *et al*. 2014); and discussion of the intracellular components of the gliding motility molecular machinery, termed the glideosome, in proto-apicomplexans are described in Woo *et al*. (2015). Here we will describe the contribution of new genome sequence information to understanding the possible origin of the apicomplexan surface receptor involved in gliding motility, called the TRAP/MIC2 superfamily of transmembrane proteins (reviewed in Morahan *et al*. 2009).

TRAP/MIC2 family proteins link extracellular adhesion to interaction with the cytoplasmic actin and myosin motility apparatus. All TRAP/MIC2 proteins described to date possess an extracellular region containing one or more TSP1 domains and,

https://doi.org/10.1017/S0031182015001213 Published online by Cambridge University Press

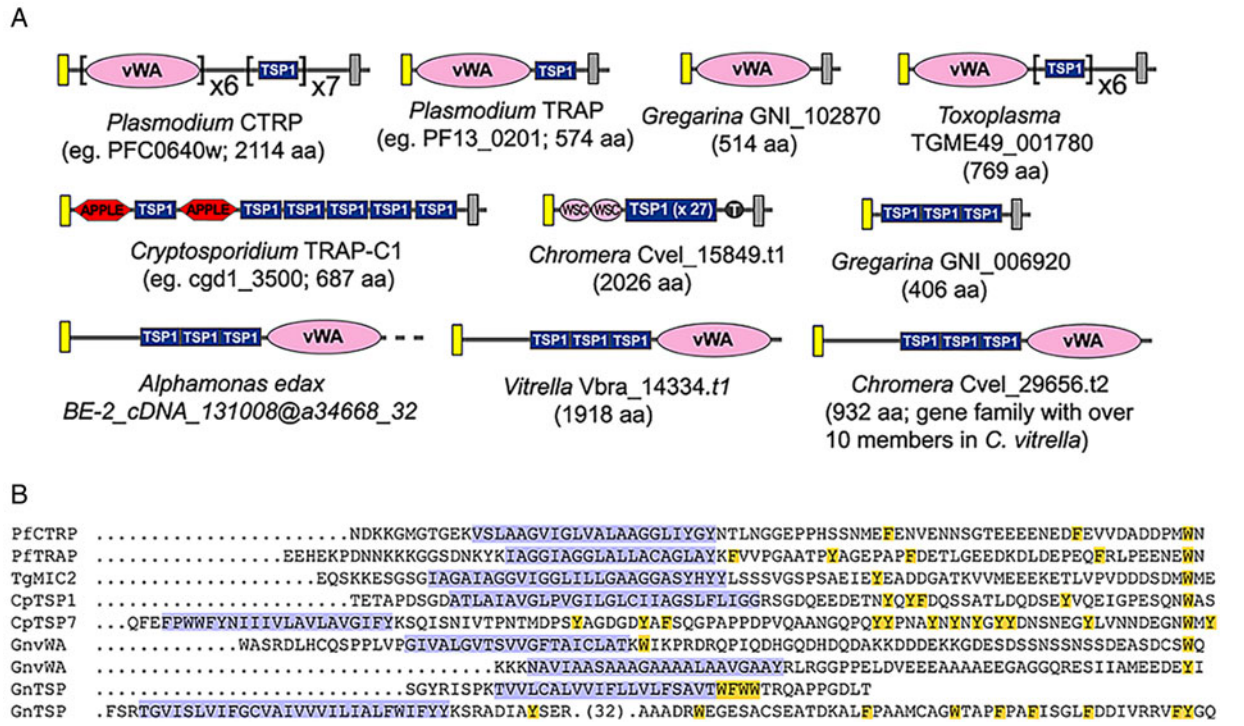

Fig. 4. Domain architectures of predicted apicomplexan TRAP proteins and representative TSP1 and vWA proteins from chromerids, colpodellids and *Gregarina* (A). Proteins are not drawn to scale; lengths in amino acids (aa) indicated. The gene for *Alphamonas edax BE-2*_cDNA_131008@a34668_32 appears to be incomplete at the 3′ end, indicated by a dashed line. Amino acid sequences of the cytoplasmic domains of TRAP/MIC2 proteins and candidates; the sequences are not aligned based upon amino acid similarities, but rather to show conserved features within the short, acidic cytoplasmic domain and conserved aromatic residues adjacent the C-terminus (B). Predicted transmembrane regions are highlighted in blue and aromatic residues highlighted in yellow. Gene IDs are as follows: PfCTRP (PF3D7_1133400), PfTRAP (PF3D7_1133400), TgMIC2 (TGME49_001780), CpTSP1 (cgd1_3500), CpTSP7 (cgd5_4470), GnvWA (GNI_102870 and GNI_030200), GnTSP (GNI_006920 and GNI_113530).

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 not 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

https://doi.org/10.1017/S0031182015001213 Published online by Cambridge University Press

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 multi-domain 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

by KAUST-faculty baseline funding and CRG grants from OCRF, KAUST and Global Station for Zoonosis Control, Global Institution for Collaborative Research and Education (GI-CoRE), Hokkaido University, Japan.

REFERENCES

Abrahamsen, M. S., Templeton, T. J., Enomoto, S., Abrahante, J. E., Zhu, G., Lancto, C. A., Deng, M., Liu, C., Widmer, G., Tzipori, S., Buck, G. A., Xu, P., Bankier, A. T., Dear, P. H., Konfortov, B. A., Spriggs, H. F., Iyer, L., Anantharaman, V., Aravind, L. and Kapur, V. (2004). Complete genome sequence of the apicomplexan *Cryptosporidium parvum*. *Science* **304**, 441–445.

Adl, S. M., Simpson, A. G., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E. Le Gall, L., Lynn, D. H., McManus, H., Mitchell, E. A., Mozley-Stanridge, S. E., Parfrey, L. W., Pawlowski, J., Rueckert, S., Shadwick, R. S., Schoch, C. L., Smirnov, A. and Spiegel, F. W. (2012). The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology* **59**, 429–493.

al-Khedery, B., Barnwell, J. W. and Galinski, M. R. (1999). Antigenic variation in malaria: a 3′ genomic alteration associated with the expression of a *P. knowlesi* variant antigen. *Molecular Cell* **3**, 131–141.

Aeschlimann, S. H., Jönnsson, F., Postberg, J., Stover, N. A., Petera, R. L., Lipps, H. J., Nowacki, M. and Swart, E. C. (2014). The draft assembly of the radically organized *Stylonychia lemnae* macronuclear genome. *Genome Biology and Evolution* **6**, 1707–1723.

Annoura, T., van Schaijk, B. C., Ploemen, I. H., Sajid, M., Lin, J. W., Vos, M. W., Dinmohamed, A. G., Inaoka, D. K., Rijpma, S. R., van Gemert, G. J., Chevalley-Maurel, S., Kiełbasa, S. M., Scheltinga, F., Franke-Fayard, B., Klop, O., Hermsen, C. C., Kita, K., Gego, A., Franetich, J. F., Mazier, D., Hoffman, S. L., Janse, C. J., Sauerwein, R. W. and Khan, S. M. (2014). Two *Plasmodium* 6-Cys family-related proteins have distinct and critical roles in liver-stage development. *The FASEB Journal* **28**, 2158–2170.

Aravind, L., Anantharaman, V., Zhang, D., de Souza, R. and Iyer, L. M. (2012). Gene flow and biological conflict systems in the origin and evolution of eukaryotes. *Frontiers in Cellular and Infection Microbiology* **2**, 89.

Arredondo, S. A., Cai, M., Takayama, Y., MacDonald, N. J., Anderson, D. E., Aravind, L., Clore, G. M. and Miller, L. H. (2012). Structure of the *Plasmodium* 6-cysteine s48/45 domain. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 6692–6697.

Aury, J. M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B. M., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Câmara, F., Duharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A. M., Kissmehl, R., Klotz, C., Koll, F., Le Mouël, A., Lepère, G., Malinsky, S., Nowacki, M., Nowak, J. K., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Bétermier, M., Weissenbach, J., Scarpelli, C., Schächter, V., Sperling, L., Meyer, E., Cohen, J. and Wincker, P. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**, 171–178.

Barnes, D. A., Bonnin, A., Huang, J.-X., Gousset, L., Wu, J., Gut, J., Doyle, P., Dubremetz, J.-F., Ward, H. and Petersen, C. (1998). A novel multi-domain mucin-like glycoprotein of *Cryptosporidium parvum* mediates invasion. *Molecular and Biochemical Parasitology* **96**, 93–110.

Beale, G. H. and Preer, J. R., Jr. (2008). Paramecium *Genetics and Epigenetics*. CRC Press, Boca Raton, FL, USA.

Bhalchandra, S., Ludington, J., Coppens, I. and Ward, H. D. (2013). Identification and characterization of *Cryptosporidium parvum* Clec, a novel C-type lectin domain-containing mucin-like glycoprotein. *Infection and Immunity* **81**, 3356–3365.

Bhat, N., Joe, A., PereiraPerrin, M. and Ward, H. D. (2007). *Cryptosporidium* p30, a galactose/*N*-acetylgalactosamine-specific lectin, mediates infection *in vitro*. *Journal of Biological Chemistry* **282**, 34877–34887.

Brayton, K. A., Lau, A. O., Herndon, D. R., Hannick, L., Kappmeyer, L. S., Berens, S. J., Bidwell, S. L., Brown, W. C., Crabtree, J., Fadrosh, D., Feldblum, T., Forberger, H. A., Haas, B. J., Howell, J. M., Khouri, H., Koo, H., Mann, D. J., Norimine, J., Paulsen, I. T., Radune, D., Ren, Q., Smith, R. K. Jr, Suarez, C. E., White, O., Wortman, J. R., Knowles, D. P. Jr, McElwain, T. F. and Nene, V. M. (2007). Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa. *PLoS Pathogens* **3**, 1401–1413.

Briguglio, J. S. and Turkewitz, A. P. (2014). *Tetrahymena thermophila*: a divergent perspective on membrane traffic. *Journal of Experimental Zoology* **332B**, 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 Syngen 4 of *Paramecium aurelia*. *Genetics* **74**, 63–80.

Cadetti, L., Marroni, F., Marangoni, R., Kuhlmann, H.-W., Gioffré, D. and Colombetti, 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. *Annual Review of Microbiology* **43**, 23–42.

Carter, V., Shimuzu, 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.

Céréde, O., Dubremetz, J. F., Soête, M., Deslée, D., Vial, H., Bout, D. and Lebrun, M. (2005). Synergistic role of micronemal proteins in *Toxoplasma gondii* virulence. *Journal of Experimental Medicine* **201**, 453–463.

Chattrejee, A., Banerjee, S., Steffen, M., O'Conner, R. M., Ward, H. D., Robbins, P. W. and Samuelson, J. (2010). Evidence for mucin-like glycoproteins that tether sporozoites of *Cryptosporidium parvum* to the inner surface of the oocyst wall. *Eukaryotic Cell* **9**, 84–96.

Cornillot, E., Hadj-Kaddour, K., Dassouli, A., Noel, B., Ranwez, V., Vacherie, B., Augagneur, Y., Brès, V., Duclos, A., Randazzo, S., Carcy, B., Debierre-Grockiego, F., Delbecq, S., Moubri-Ménage, K., Shams-Eldin, H., Usmani-Brown, S., Bringaud, F., Wincker, P., Vivarès, C. P., Schwarz, R. T., Schetters, T. P., Krause, P. J., Gorenflot, A., Berry, V., Barbe, V. and Ben Mamoun, C. (2012). Sequencing of the smallest apicomplexan genome from 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 *Perkinsus* sp. Isolated from the Baltic clam *Macoma balthica*. *Journal of Eukaryotic Microbiology* **48**, 38–51.

Coyne, R. S., Hannick, L., Shanmugam, D., Hostetler, J. B., Brami, D., Joardar, V. S., Johnson, J., Radune, D., Singh, I., Badger, J. H., Kumar, U., Saier, M., Wang, Y., Cai, H., Gu, J., Mather, M. W., Vaidya, A. B., Wilkes, D. E., Rajagopalan, 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 glideosome function in apicomplexans. *Current Opinion in Microbiology* **12**, 408–414.

Deng, M., Templeton, T. J., London, N. R., Bauer, C., Schroeder, A. A. and Abrahamsen, M. S. (2002). *Cryptosporidium parvum* genes containing thrombospondin type 1 domains. *Infection and Immunity* **70**, 6987–6995.

Douradinha, B., Ausustijn, K. D., Moore, S. G., Ramesar, J., Mota, M. M., Waters, A. P., Janse, C. J. and Thompson, J. (2011). *Plasmodium* cysteine 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.

Echevarria, M. L., Wolfe, G. V., Strom, S. L. and Taylor, A. R. (2014). Connecting alveolate cell biology with trophic ecology in the marine plankton using the ciliate *Favella* as a model. *FEMS Microbiology Ecology* **90**, 18–38.

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

Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q., Amedeo, P., Jones, K. M., Tallon, L. J., Delcher, A. L., Salzberg, S. L., Silva, J. C., Haas, B. J., Majoros, W. H., Farzad, M., Carlton, J. M., Smith, R. K. Jr, Garg, J., Pearlman, R. E., Karrer, K. M., Sun, L., Manning, G., Elde, N. C., Turkewitz, A. P., Asai, D. J., Wilkes, D. E., Wang, Y., Cai, H., Collins, K., Stewart, B. A., Lee, S. R., Wilamowska, K., Weinberg, Z., Ruzzo, W. L., Wloga, D., Gaertig, J., Frankel, J., Tsao, C. C., Gorovsky, M. A., Keeling, P. J., Waller, R. F., Patron, N. J., Cherry, J. M., Stover, N. A., Krieger, C. J., del

Toro, C., Ryder, H. F., Williamson, S. C., Barbeau, R. A., Hamilton, E. P. and Orias, E. (2006). Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biology* **4**, e286.

Fenchel, T. and Finlay, B. J. (1984). Geotaxis in the ciliated protozoon *Loxodes*. *Journal of Experimental Biology* **110**, 17–33.

Francis, J. T. and Hennessey, T. M. (1995). Chemorepellents in *Paramecium* and *Tetrahymena*. *Journal of Eukaryotic Microbiology* **42**, 78–83.

Frech, C. and Chen, N. (2013). Variant surface antigens of malaria parasites: functional and evolutionary insights from comparative gene family classification and analysis. *BMC Genomics* **14**, 427.

Frénal, K., Polonais, V., Marq, J.-B., Stratmann, R., Limenitakis, J. and Soldati-Favre, D. (2010). Functional dissection of the apicomplexan glideosome molecular architecture. *Cell Host and Microbe* **8**, 343–357.

Friedrich, N., Santos, J. M., Liu, Y., Palma, A. S., Leon, E., Saoros, S., Kiso, M., Blackman, M. J., Matthews, S., Feizi, T. and Soldati-Favre, D. (2010). Members of a novel protein family containing microneme adhesive repeat domains act as sialic acid-binding lectins during host cell invasion by apicomplexan parasites. *Journal of Biological Chemistry* **285**, 2064–2076.

Gardner, M. J., Bishop, R., Shah, T., de Villiers, E. P., Carlton, J. M., Hall, N., Ren, Q., Paulsen, I. T., Pain, A., Berriman, M., Wilson, R. J., Sato, S., Ralph, S. A., Mann, D. J., Xiong, Z., Shallom, S. J., Weidman, J., Jiang, L., Lynn, J., Weaver, B., Shoaibi, A., Domingo, A. R., Wasawo, D., Crabtree, J., Wortman, J. R., Haas, B., Angiuoli, S. V., Creasy, T. H., Lu, C., Suh, B., Silva, J. C., Utterback, T. R., Feldblyum, T. V., Pertea, M., Allen, J., Nierman, W. C., Taracha, E. L., Salzberg, S. L., White, O. R., Fitzhugh, H. A., Morzaria, S., Venter, J. C., Fraser, C. M. and Nene, V. (2005). Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. *Science* **309**, 134–137.

Gaur, D. and Chitnis, C. E. (2011). Molecular interactions and signalling mechanisms during erythrocyte invasion by malaria parasites. *Current Opinion in Microbiology* **14**, 422–428.

Gavelis, G. S., Hayakawa, S., White, R. A., Gojobori, T., Suttle, C. A., Keeling, P. J. and Leander, B. S. (2015). Eye-like ocelloids are built from different endosymbiotically acquired components. *Nature* **523**, 204–207.

Gerloff, D. L., Creasey, A., Maslau, S. and Carter, R. (2005). Structural models for the protein family characterized by gamete surface proteins Pfs230 of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13598–13603.

Gomez, F. (2008). Erythropsidinium (Gymnodiniales, Dinophyceae) in the Pacific Ocean, a unique dinoflagellate with an ocelloid and a piston. *European Journal of Protistology* **44**, 291–298.

Gubbels, M. J. and Duraisingh, M. T. (2012). Evolution of apicomplexan secretory organelles. *International Journal for Parasitology* **42**, 1071–1081.

Hamel, A., Fisch, C., Combettes, L., Dupuis-Williams, P. and Baroud, C. N. (2011). Transitions between three swimming gaits in *Paramecium* escape. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 7290–7295.

Hayakawa, S., Takaku, Y., Hwang, J. S., Horiguchi, T., Suga, H., Gehring, W., Ikeo, K. and Gojobori, T. (2015). Function and evolutionary origin of unicellular camera-type eye structure. *PLoS ONE* **10**, e0118415.

Heitlinger, E., Spork, S., Lucius, R. and Dieterich, C. (2014). The genome of *Eimeria falciformis*—reduction and specialization in a single host apicomplexan parasite. *BMC Genomics* **15**, 696.

Hemmersbach, R., Volkmann, D. and Hader, D. P. (1999). Graviorientation in protists and plants. *Journal of Plant Physiology* **154**, 1–15.

Holder, A. A., Ridzuan, M. A. and Green, J. L. (2012). Calcium-dependent protein kinase 1 and calcium fluxes in the malaria parasite. *Microbes and Infection* **14**, 825–830.

Hoppenrath, M. and Leander, B. S. (2009). Molecular phylogeny of *Parvilucifera prorocentri* (Alveolata, Myzozoa): insights into Perkinsid character evolution. *Journal of Eukaryotic Microbiology* **56**, 251–256.

Ishino, T., Chinzei, Y. and Yuda, M. (2005). Two proteins with 6-cys motifs are required for malarial parasites to commit to infection of the hepatocyte. *Molecular Microbiology* **58**, 1264–1275.

Jacot, D., Frénal, K., Marq, J. B., Sharma, P. and Soldati-Favre, D. (2014). Assessment of phosphorylation in *Toxoplasma* glideosome assembly and function. *Cellular Microbiology* **16**, 1518–1532.

Jackson, A. P., Otto, T. D., Darby, A., Ramaprasad, A., Xia, D., Echaide, I. E., Farber, M., Gahlot, S., Gamble, J., Gupta, D., Gupta, Y., Jackson, L., Malandrin, L., Malas, T. B., Moussa, E., Nair, M., Reid, A. J., Sanders, M., Sharma, J., Tracey, A., Quail, M. A., Weir, W., Wastling, J. M., Hall, N., Willadsen, P., Lingelbach, K., Shiels, B., Tait, A., Berriman, M., Allred, D. R. and Pain, A. (2014). The evolutionary dynamics of antigen genes in *Babesia* reveal a history of genomic innovation underlying host-parasite interaction. *Nucleic Acids Research* **42**, 7113–7131.

Janouškovec, J., Tikhonenkov, D. V., Burki, F., Howe, A. T., Kolísko, M., Mylnikov, A. P. and Keeling, P. J. (2015). Factors mediating plastid 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.

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

Knoll, G., Haacke-Bell, B. and Plattner, H. (1991). Local trichocyst exocytosis 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 chemoattraction. *PLoS ONE* **6**, e28022.

Lapp, S. A., Korir-Morrison, C., Jiang, J., Bai, Y., Corredor, V. and Galinski, M. R. (2013). Spleen-dependent regulation of antigenic variation in malaria parasites: *Plasmodium knowlesi* SICAvar expression profiles in splenic and asplenic hosts. *PLoS ONE* **8**, e78014.

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

Liu, Y., Tewari, R., Ning, J., Blagborough, A. M., Garbom, S., Pei, J., Grishin, N. V., Steele, R. E., Sinden, R. E., Snell, W. J. and Billker, O. (2008). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. *Genes and Development* **22**, 1051–1068.

Lobban, C. S., Hallam, S. J., Mukherjee, P. and Petrich, J. W. (2007). Photophysics and multifunctionality of hypercin-like pigments in heterotrich ciliates: a phylogenetic persepective. *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., Katrib, M., Slapeta, J., Slapetova, I., Wallach, M. G., Luxford, C., Davies, M. J., Zhang, X., Norton, R. S. and Belli, S. I. (2011). Peroxidase catalyzed cross-linking of an intrinsically unstructured protein via dityrosine bonds in the oocyst wall of the apicomplexan parasite, *Eimeria maxima*. *International Journal for Parasitology* **41**, 1157–1164.

Martel, C. M. (2009). Conceptual bases for prey biorecognition and feeding selectivity in the microplanktonic marine phagotroph *Oxyrrhis marina*. *Microbial Ecology* **57**, 589–597.

McCoy, J. M., Whitehead, L., van Dooren, G. G. and Tonkin, C. J. (2012). TgCDPK3 regulates calcium-dependent egress of *Toxoplasma gondii* from host cells. *PLoS Pathogens* **8**, e1003066.

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 atlanticus*. *Parasitology* **1224**, 477–484.

Moore, R. B., Oborník, M., Janouškovec, J., Chrudimský, T., Vancová, M., Green, D. H., Wright, 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 Coppel, R. L. (2009). No TRAP, no invasion. *Trends in Parasitology* **25**, 77–84.

Oborník, M., Vancová, M., Lai, D-H., Janouškovec, J., Keeling, P. J. and Lukeš, J. (2011). Morphology and ultrastructure of multiple life cycle stages of the photosynthetic relative of Apicomplexa, *Chromera velia*. *Protist* **162**, 115–130.

Oborník, M., Modrý, D., Lukeš, M., Cernotíková-Stříbrná, E., Cihlář, J., Tesařová, M., Kotabová, E., Vancová, M., Prášil, O. and Lukeš, J. (2012). Morphology, ultrastructure and life cycle of *Vitrella brassicaformis* n. sp., n. gen., a novel Chromerid from the Great Barrier Reef. *Protist* **163**, 306–323.

O'Conner, R. M., Lane, T. J., Stroup, S. E. and Allred, D. R. (1997). Characterization of a variant erythrocyte surface antigen (VESA1) expressed by *Babesia bovis* 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 association with the flagellar apparatus revealed by serial TEM tomography in *Psammosa pacifica*, a distant relative of the Apicomplexa. *PLoS ONE* **9**, e84653.

Pain, A., Renauld, 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., Goble, A., Griffiths-Jones, S., Harris, D. E., Katzer, F., Larke, N., Lord, A., Maser, P., McKellar, S., Mooney, P., Morton, F., Nene, V., O'Neil, S., Price, C., Quail, M. A., Rabbinowitsch, E., Rawlings, N. D., Rutter, S., Saunders, D., Seeger, K., Shah, T., Squares, R., Squares, S., Tivey, A., Walker, A.R., Woodward, J., Dobbelaere, D. A., Langsley, G., Rajandream, M. A., McKeever, D., Shiels, B., Tait, A., Barrell, B. and Hall, N. (2005). Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*. *Science* **309**, 131–139.

Plattner, H., Sehring, I. M., Mohamed, I. K., Miranda, K., De Souza, W., Billington, R., Genazzani, A. and Ladenburger, E. M. (2012). Calcium signalling in closely related protozoan groups (Alveolata): non-parasitic ciliates (*Paramecium, Tetrahymena*) *vs* parasitic Apicomplexa (*Plasmodium, Toxoplasma*). *Cell Calcium* **51**, 351–382.

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**, 58–64.

Pradel, G., Hayton, K., Aravind, L., Iyer, L. M., Abrahamsen, M. S., Bonawitz, 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.

Reid, A. J., Blake, D. P., Ansari, H. R., Billington, K., Browne, H. P., Bryant, J., Dunn, M., Hung, S. S., Kawahara, F., Miranda-Saavedra, D., Malas, T. B., Mourier, T., Naghra, H., Nair, M., Otto, T. D., Rawlings, N. D., Rivailler, P., Sanchez-Flores, A., Sanders, M., Subramaniam, C., Tay, Y. L., Woo, Y., Wu, X., Barrell, B., Dear, P. H., 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**, 1676–1685.

Robert, E. C., Zubhov, M. V., Martin-Cereceda, M., Novarino, G. and Wootton, E. C. (2006). 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 discover 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 merozoite surface antigens in raft-like membranes of *Plasmodium falciparum*. *Journal of Biological Chemistry* **280**, 40169–40176.

Selbach, M. and Kuhlmann, H. W. (1999). Structure, fluorescent properties and proposed function in phototaxis of the stigma apparatus in the ciliate *Chlamydodon mnemosyne*. *The Journal of Experimental Biology* **202**, 919–927.

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

Singh, D. P., Saudemont, B., Guglielmi, G., Arnaiz, O., Goût, J. F., Prajer, M., Potekhin, A., Przybòs, E., Aubusson-Fleury, A., Bhullar, S., Bouhouche, K., Lhuillier-Akakpo, M., Tanty, V., Blugeon, C., Alberti, A., Labadie, K., Aury, J. M., Sperling, L., Duharcourt, S. 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 long and winding road: the *Plasmodium* sporozoite's journey in the mammalian host. *Parasitology International* **56**, 171–178.

Slamovits, C. H., Okamoto, N., Burri, L., James, E. R. and Keeling, P. J. (2011). A bacterial proteorhodopsin proton pump in marine eukaryotes. *Nature Communications* **8**, 183.

Sonneborn, T. M. (1938). Mating types in *Paramecium aurelia*: diverse conditions for mating in different stocks; occurrence, number, and interrelations of the types. *Proceedings of the American Philosophical Society* **79**, 411–434.

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

Smith, J. D. (2014). The role of PfEMP1 adhesion domain classification in *Plasmodium falciparum* pathogenesis research. *Molecular and Biochemical Parasitology* **195**, 82–87.

Smith, J. D., Rowe, J. A., Higgins, M. K. and Lavstsen, T. (2013). Malaria's deadly grip: cytoadhesion of *Plasmodium falciparum*-infected erythrocytes. *Cellular Microbiology* **15**, 1976–1983.

Spano, F., Puri, C., Ranucci, L., Putignani, L. and Crisanti, A. (1997). Cloning of the entire COWP gene of *Cryptosporidium parvum* and ultrastructural localization of the protein during sexual parasite development. *Parasitology* **114**, 427–437.

Spano, F., Putignani, L., Naitza, S., Puri, C., Wright, S. and Crisanti, A. (1998). Molecular cloning and expression analysis of a *Cryptosporidium parvum* gene encoding a new member of the thrombospondin family. *Molecular and Biochemical Parasitology* **92**, 147–162.

Swart, E. C., Bracht, J. R., Magrini, V., Minx, P., Chen, X., Zhou, Y., Khurana, J. S., Goldman, A. D., Nowacki, M., Schotanus, K., Jung, S., Fulton, R. S., Ly, A., McGrath, S., Haub, K., Wiggins, J. L., Storton, D., Matese, J. C., Parsons, L., Chang, W. J., Bowen, M. S., 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 *Oxytricha trifallax* macronuclear 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-Pierres, 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*. *Molecular and Biochemical Parasitology* **166**, 109–116.

Templeton, T. J., Lancto, C. A., Vigdorovich, V., Liu, C., London, N. R., Hadsall, K. Z. and Abrahamsen, M. S. (2004*a*). The *Cryptosporidium* oocyst wall protein is a member of a multigene family and has a homolog in *Toxoplasma*. *Infection and Immunity* **72**, 980–987.

Templeton, T. J., Iyer, L. M., Anantharaman, V., Enomoto, S., Abrahante, J. E., Subramanian, G. M., Hoffman, S. L., Abrahamsen, M. S. and Aravind, L. (2004*b*). Comparative analysis of Apicomplexa and genomic diversity of eukaryotes. *Genome Research* **14**, 1686–1695.

Templeton, T. J., Enomoto, S., Chen, W.-J., Huang, C.-G., Lancto, C. A., Abrahamsen, M. S. and Zhu, G. (2010). A genome-sequence survey for *Ascogregarina taiwanensis* supports evolutionary affiliation but metabolic diversity between a gregarine and *Cryptosporidium*. *Molecular and Biological Evolution* **27**, 235–248.

Thompson, J., Fernandez-Reyes, D., Sharling, L., Moore, S. G., Eling, W. M., Keyes, S. A., Newbold, C. I., Kafatos, F. C., Janse, C. J. and Waters, A. P. (2007). *Plasmodium* cysteine repeat modular proteins 1–4: complex proteins with roles throughout the malaria parasite life cycle. *Cellular Microbiology* **9**, 1466–1480.

Tomita, T., Bzik, D. J., Ma, Y. F., Fox, B. A., Markillie, L. M., Taylor, R. C., Kim, K. and Weiss, L. M. (2013). The *Toxoplasma gondii* cyst wall protein CST1 is critical for cyst wall integrity and promotes bradyzoite persistence. *PLoS Pathogens* **9**, e1003823.

Valigurová, A., Vaškovičová, N., Musilová, N. and Schrevel, J. (2013). The enigma of eugregarine epicytic folds: where gliding motility originates? *Frontiers in Zoology* **10**, 57.

Van Dijk, M. R., van Schaijk, B. C., Khan, S. M., van Dooren, M. W., Ramesar, J., Kaczanowski, S., van Genert, G. J., Kroeze, H., Stunnenberg, H. G., Eling, W. M., Sauerwein, R. W., Waters, A. P. and Janse, C. J. (2010). Three members of the 6-cys protein family of *Plasmodium* play a role in gamete fertility. *PLoS Pathogens* **6**, e1000853.

Vayssié, L., Skouri, F., Sperling, L. and Cohen, J. (2000). Molecular genetics of regulated secretion in *Paramecium*. *Biochemie* **82**, 269–288.

Venkatachalam, K. and Montell, C. (2007). TRP channels. *Annual Review Biochemistry* **76**, 387–417.

Vinayak, S., Pawlowic, M. C., Sateriale, A., Brooks, C. F., Studstill, C. J., Bar-Peled, Y., Cipriano, M. J. and Streipen, B. (2015). Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*. *Nature* **5223**, 477–480.

Walker, R. A., Slapetova, I., Slapeta, J., Miller, C. M. and Smith, N. C. (2010). The glycosylation pathway of *Eimeria tenella* is upregulated during gametocyte development and may play a role in oocyst wall formation. *Eukaryotic Cell* **9**, 127–135.

Williamson, K. C., Criscio, M. D. and Kaslow, D. C. (1993). Cloning and expression of the gene for *Plasmodium falciparum* transmission-blocking target antigen, Pfs230. *Molecular and Biochemical Parasitology* **58**, 355–358.

Woo, Y. H., Ansari, H., Otto, T. D., Klinger, C. M., Kolisko, M., Michálek, J., Saxena, A., Shanmugam, D., Tayyrov, A., Veluchamy, A., Ali, S., Bernal, A., del Campo, J., Cihlář, J., Flegontov, P., Gornik, S. G., Hajdušková, E., Horák, A.,

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.

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 protozoa. *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.