Large, rapidly evolving gene families are at the forefront of host–parasite interactions in *Apicomplexa*

ADAM J REID*

Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK

(Received 29 May 2014; revised 15 August 2014; accepted 26 August 2014; first published online 26 September 2014)

SUMMARY

The *Apicomplexa* is a phylum of parasitic protozoa, which includes the malaria parasite *Plasmodium*, amongst other species that can devastate human and animal health. The past decade has seen the release of genome sequences for many of the most important apicomplexan species, providing an excellent basis for improving our understanding of their biology. One of the key features of each genome is a unique set of large, variant gene families. Although closely related species share the same families, even different types of malaria parasite have distinct families. In some species they tend to be found at the ends of chromosomes, which may facilitate aspects of gene expression regulation and generation of sequence diversity. In others they are scattered apparently randomly across chromosomes. For some families there is evidence they are involved in antigenic variation, immune regulation and immune evasion. For others there are no known functions. Even where function is unknown these families are most often predicted to be exposed to the host, contain much sequence diversity and evolve rapidly. Based on these properties it is clear that they are at the forefront of host–parasite interactions. In this review I compare and contrast the genomic context, gene structure, gene expression, protein localization and function of these families across different species.

Key words: gene families, *Apicomplexa*, host–parasite interactions, antigenic variation, immune evasion, *Plasmodium*, *Toxoplasma*, *Eimeria*, *Babesia*, *Theileria*.

INTRODUCTION

Apicomplexans are parasitic protozoa, which possess an apical complex, a unique set of organelles involved in invasion of host cells. The phylum includes the malaria parasite genus *Plasmodium*, *Toxoplasma gondii* – a parasite capable of infecting virtually all cell types in all warm-blooded animals – and a plethora of other parasites of humans and livestock: *Cryptosporidium*, *Eimeria*, *Neospora*, *Theileria* and *Babesia*. Thanks to improving technologies, large financial investment and much hard work over the past decade we have been furnished with annotated genome assemblies for the majority of important apicomplexans. Many of their genes are conserved across the phylum, representing the core genome of an apicomplexan parasite. However each genome has revealed unique genes and gene families, which should to tell us about the environmental niche of each parasite and its survival strategy. A gene family is a set of genes with similar sequences, domain structure and function that are thought to have a common ancestor.

By infecting the bodies and invading the cells of complex organisms, apicomplexan parasites endanger the life of their host and open themselves to attack from its immune system (Turner, 2002 #515).

To be successful the parasite must optimize its chances of passing into a new host. This requires the parasite to minimize damage to its host until it finds a new host, while preventing the immune system from killing it. Good strategies for achieving these goals will differ in different tissues and in different hosts. Thus, in each parasite, specialized genes should have evolved to satisfy these needs. Evidence points to the large, rapidly mutating, clade-specific gene families having a large role to play in these processes.

These gene families are sometimes referred to as contingency gene families. Contingency genes have been defined as those which allow enhanced phenotypic variation (Bayliss et al. 2001). These genes are found in large families and are thought to be subject to higher mutation rates than other genes, by processes such as gene conversion, recombination, duplication and deletion. They often reside in specialized regions of the genome, which may facilitate mutation and correct gene expression. Many examples encode surface antigens: membrane-bound proteins to which the host immune system develops antibodies. The classical example in Apicomplexa is the var family from *Plasmodium falciparum*, which is involved in antigenic variation (Guizetti and Scherf, 2013). However, other families encode products which are secreted into the host cell. These perform roles such as remodelling the host cell to favour the parasite (e.g. ROP kinases of the *Coccidia*, FIKK kinases in *P. falciparum* and SVSPs

* Corresponding author. Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. E-mail: ar11@sanger.ac.uk

Parasitology (2015), 142, S57–S70. © Cambridge University Press 2014. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.
in *Theileria* spp.). Each new group of apicomplexan parasites investigated reveals several unique multi-gene families. In this review I will discuss these large gene families in apicomplexan parasites: how genome sequencing has been used to reveal their full extent in each genome, their genomic context, functions and how those functions relate to different parasitic niches.

**IDENTIFYING GENE REPERTOIRES**

Whole genome sequencing has dramatically improved our understanding of large variant gene families, their extent, diversity and genomic context. However, it has remained challenging to accurately determine their full repertoires. While contingency gene families make it difficult for the human immune system to battle parasites, they also make it difficult for us to produce complete genome assemblies, frustrating our attempts to better understand them. This is because these families are inherently repetitive. The whole or part of one gene may look much like that of another gene. This is likely key to their functional role, but also makes it difficult to piece together the genome.

The *var* gene family of *P. falciparum*, involved in antigenic variation and sequestration, is a prime example. Only using a variety of genome sequencing technologies, as well as manual finishing work, over several years have we become fairly sure that the *P. falciparum* 3D7 reference genome sequence is complete and describes the full complement of *var* genes (M. Berriman, personal communication). The problem is that each new parasite strain sequenced has a largely distinct set of *var* genes, variously recombined and mutated. This makes it difficult to determine the set of *var* genes in each new strain using current technologies without laboriously generating a completely new genome assembly. Furthermore, repetitive parts of these genes are sometimes longer than sequencing reads, meaning that de novo assembly approaches will fail to resolve them. Instead, researchers have been exploring targeted de novo assembly strategies to explore *var* gene repertoires in worldwide strains of *P. falciparum* (Assefa, 2013). Similarly, it was found that the *T. gondii* genome sequence contained collapsed repeats of the *rop5* gene, a member of the *rop* family (Reese et al. 2011). A genetic linkage analysis had identified the locus of this gene as important for parasite virulence. Detailed analysis of the locus identified the collapsed repeat, which was found to vary in copy number between parasites with differing virulence. Without continued improvement of genome sequences, using a variety of technologies, it is likely that key features of parasite biology will be missed.

Although it can be difficult to define every gene in a repetitive gene family, straightforward analysis of even a draft assembly will highlight much of their repertoire. Each new genome assembly of sufficient phylogenetic distance from those we have will likely identify new such families. Furthermore, sequencing more members of well-studied genera will be useful in understanding the evolution of these families. Table 1 summarizes our knowledge of these large families (those with greater than 20 members) in apicomplexan species sequenced to date.

**GENOMIC CONTEXT**

Complete genome sequences have allowed us to analyse the genomic context of gene families in unprecedented detail. Genomic context appears to be important for regulation of gene expression and generation of diversity in large apicomplexan gene families. In *P. falciparum* the *var*, *rif* and *stevor* families cluster together, proximal to each telomere (Fig. 1). The *phist* and *fibh* families are also found in these subtelomeric regions but further towards the centromeres.

Subtelomeric location is common to contingency families in almost all species of *Plasmodium* examined. It has been suggested to play a role in regulating expression and generating diversity by promoting recombination (Scherf et al. 2008). The occasionally zoonotic monkey malaria parasite *Plasmodium knowlesi* is an exception. In *P. knowlesi*, both SICAvars and pirs are spread throughout the chromosomes (Fig. 1). However, despite their internal location, they are associated with telomere-like repeats, which might play a role in promoting recombination (Pain et al. 2008). *Var* genes (excluding the highly conserved *var1csa* and *var2csa*) can be classified into three types based on their promoter sequences: *upsA*, *upsB* and *upsC* (Lavstsen et al. 2003). These distinct promoters are associated with distinct contexts. *UpsB* *var* genes are subtelomeric and transcribed away from the telomere, *upsA* are subtelomeric, but transcribed towards the telomere. * UpsC* *var* genes are found in core chromosomal regions. The telomeress of *P. falciparum* chromosomes (Freitas-Júnior et al. 2000) and also *var* genes (Ralph et al. 2005) have been shown to localize in a small number of foci at the periphery of the nucleus. This also occurs in rodent malaria parasites and is thought to foster recombination, generating new variants (Figueiredo et al. 2000).

Subtelomeric gene families are also a feature of *Theileria* spp. but not of Babesia or any *Coccidia* so far examined (Fig. 1). A much greater sampling of the *Apicomplexa* will be required to understand just how common this arrangement is and whether it has evolved multiple times in *Apicomplexa*.

Relative orientation of genes is important for expression of the *vesI* gene family in *Babesia bovis*. *VesI/smORF* units comprise a handful of genes,
Large gene families of the Apicomplexa

Generally with at least one ves1a, one ves1b and a smORF (Brayton et al. 2007). Ves1a and ves1β together encode VESA proteins, involved in antigenic variation and cytoadherence (Allred et al. 2000; O’Connor and Allred, 2000). The function of smORF is not known. The ves1 family is one of two known amongst apicomplexans, which shows true properties of antigenic variation, with only one variant expressed at a time. Transcription from the functional locus of active transcription (LAT) is thought to require a ves1a and a ves1β in divergent orientation, separated by a bidirectional promoter (Al-Khedery and Allred, 2006). Ves1/smORF units are found scattered throughout Babesia chromosomes, not in subtelomeres or in tandem arrays (Fig. 1).

The srs and sag genes families, which encode unrelated, but structurally similar, surface antigens in Coccidia are found spread throughout chromosomes, mostly in tandem arrays (Fig. 1b). These arrays tend to contain genes, which are more similar to each other than other clusters, and tend to share domain subfamily architectures (Reid et al. 2012; Wasmuth et al. 2012). This is probably maintained through local gene duplication as well as gene conversion (Reid et al. 2012).

Tandem arrays of genes can easily arise through gene duplication but the maintenance of this pattern may have advantages for promoting gene conversion. In some cases genes within a tandem array may be more similar to each other than those in other arrays (e.g. Coccidia srs/sag). In other cases they are not e.g. Plasmodium pirs (Otto et al.) and Theileria ShI sub-telomeric genes (Weir et al. 2010), suggesting recombination between arrays. Similarity within loci might be due to either recent divergence, through copying or a continued process of local gene conversion. Some genes families, present in tandem arrays, may preserve their orientation, all being transcribed towards centromere or telomere, in order that they can line up with homologous arrays from other chromosomes or other parts of the same chromosome for recombination.

Currently it is clear that subtelomere context is not a common feature of large apicomplexan gene families. Indeed there is no evidence for it in Coccidia. In the haemosporidia (Plasmodium) and piroplasms (Babesia, Theileria) it may have been lost or gained multiple times.

PROTEIN DOMAIN STRUCTURE

The most common feature of large apicomplexan gene family sequences is a signal peptide: a hydrophobic N-terminal sequence of 10–20 amino acids. This is an indication that such genes may be directly involved in host-parasite interactions as they can be exported from the parasite cell via the secretory pathway. Many of these gene families appear to be ultimately bound to membrane, using either transmembrane domains or glycophasiditidylinositol (GPI) anchors. In general, extracellular proteins tend to be rich in disulphide bonds to improve stability, and indeed apicomplexans are no exception (Fass, 2012). Despite these similarities, gene families in different groups of parasites share no detectable homology and in some cases can be shown to derive from different ancestral families.

The var gene family exhibits a complex protein domain architecture. Plasmodium falciparum var genes are composed of Duffy binding-like (DBL), CIDR and C2 domains (Gardner et al. 2002). DBL domains have six subfamilies and CIDR two. There is a great diversity of domain architectures in the var family, although the N-termini are usually composed of a DBLa domain followed by a CIDR domain, suggesting this structure is key to their function (Gardner et al. 2002). DBL domains are also found in several other genes of P. falciparum, which are involved in binding host cells and are required for invasion by the merozoite, implying an ancestral role for this domain in host cell binding (Tolia et al. 2005). The C-terminus includes a transmembrane domain, which tethers the proteins to the erythrocyte surface.

Other protein families have much simpler architectures. Plasmodium falciparum rifS have an N-terminal signal peptide, a PEXEL motif and either one or two transmembrane domains depending on the subfamily (Bultrini et al. 2009). PEXEL motifs are required for export of proteins to the red blood cell surface (Hiller et al. 2004; Marti et al. 2004). Based on remote sequence similarity, conservation of intronic sequence motifs and predicted protein secondary structure, it has been proposed that rif genes are related to the pirs, a family present in all other species so far sequenced (Janssen et al. 2004).

These families are an example of those which have diverged almost beyond recognition, and which may help to provide insights into evolution of gene families in malaria.

Coccidian surface antigen gene families also use relatively simple domain structures. Toxoplasma srs genes usually contain two SAG domains (Jung et al. 2004), while the short, consistent lengths of Eimeria sag genes suggest a single domain structure (Reid, 2014 #517). Both families have signal peptides and GPI anchor addition sites. Furthermore, the individual domain sequences in each family both tend to contain six cysteines and are of similar lengths (200–300 amino acids). However, these common properties appear to be convergent, as the families have distinct origins. While Toxoplasma srs genes have been proposed to derive from metazoan ephrins by lateral transfer (Arredondo et al. 2012). Eimeria sag genes meanwhile are thought to be related to CAP domains (Reid, 2014 #517). CAP domains are widely used across the tree of life.
| Family | Subfamily | Species | Number per genome | Genomic location | Gene expression | Cellular localization of product | Function(s) |
|--------|-----------|---------|-------------------|------------------|----------------|---------------------------------|------------|
| var    |           | Plasmodium falciparum | ~ 60 (Gardner et al. 2002) | Subtelomeric/ intrachromosomal (Gardner et al. 2002) | Single transcript variant in cloned parasites, exc. varlesai (Kyes et al. 2007). Single protein expressed on surface (Chen et al. 1998; Scherf et al. 1998) | MC, RBC surface (Kyes et al. 2001) | Cytoadherence and antigenic variation (Kyes et al. 2001), immunomodulation (D’Ombraint et al. 2007) |
| SICAv | var       | Plasmodium knowlesi | 192 (Pain et al. 2008) | Not specific (Pain et al. 2008) | Multiple (protein) (Howard et al. 1983) | RBC surface (Howard et al. 1983) | Antigenic variation, virulence (Barnwell et al. 1983) |
| Pe-fam-a (PeTRAG) | var      | Plasmodium vivax | 36 (Tachibana et al. 2012) | Subtelomeric (Carlton et al. 2008) | Two or three variants per cell in trophozoites (Kaviratne et al. 2002) | RBC surface (Di Girolamo et al. 2008) | – |
| Pe-fam-e (RAD) | var      | Plasmodium vivax | 44 (Tachibana et al. 2012) | Mostly internal arrays on chromosomes 5 (Carlton et al. 2008) | – | – | – |
| rif/stevor /pir | rif | Plasmodium falciparum | 163 (Logan-Klumpler et al. 2012) | Subtelomeric (Gardner et al. 2002) | Multiple variants expressed in asexual in vitro parasite population (Boulad et al. 2003) | MC, PV, RBC surface (Kyes et al. 1999), A-type RBC (Khattab and Klinkert, 2006), A-type merozoite apex, B-type PV (Petter et al. 2007) | – |
| stevor |      | Plasmodium falciparum | 30 (Logan-Klumpler et al. 2012) | Subtelomeric (Cheng et al. 1998) | Two or three variants per cell in trophozoites (Kaviratne et al. 2002) | MC, PV, RBC surface (Ni et al. 2009), merozoite apex (ref) | – |
| pir |      | Plasmodium berghei | 100 (Otto et al.) | Subtelomeric (Hall et al. 2005) | – | RBC surface (Di Girolamo et al. 2008) | – |
|       |      | Plasmodium yoelii | 583 (Otto et al.) | Subtelomeric (Carlton et al. 2002) | 1–3 variants per cell (Cunningham et al. 2009) | RBC surface (Cunningham et al. 2005) | – |
| Plasmodium chabaudi |      | 194 (Otto et al.) | Subtelomeric (Hall et al. 2005) | Multiple expressed in population (Lawton et al. 2012) | – | RBC surface (Jansen et al. 2004) | Putative role in virulence (Spence et al. 2013) |
| Plasmodium vivax |      | 346 (Carlton et al. 2008) | Subtelomeric (del Portillo et al. 2001) | Multiple variants per cell (Fernandez-Becereta et al. 2005) | – | RBC surface (del Portillo et al. 2001) | – |
| Plasmodium knowlesi |      | 67 (Pain et al. 2008) | Not specific | – | – | – | Putative role in molecular mimicry (Pain et al. 2008) |
| Plasmodium cynomolgi |      | 265 (Tachibana et al. 2012) | V′ir-like (254) subtelomeric, kir-like (11) internal (Tachibana et al. 2012) | – | – | – | – |
| Genus | Species | Accession | Location | Activity | Reference |
|-------|---------|-----------|----------|----------|-----------|
| **Phist** | *Plasmodium falciparum* | 71 (Sargeant et al. 2006) | Subtelomeric (Sargeant et al. 2006) | – | Modulation of erythrocyte cytoskeleton (RESA) (Pei et al. 2007) |
| Plasmodium vivax | | 39 (Sargeant et al. 2006) | Subtelomeric (Sargeant et al. 2006) | – | – |
| Plasmodium cynomolgi | | 21 (Tachibana et al. 2012) | Subtelomeric (Tachibana et al. 2012) | – | – |
| **fbb** | *Plasmodium falciparum* | 20 (Schneider and Mercereau-Puijalon, 2005) | Subtelomeric (Schneider and Mercereau-Puijalon, 2005) | Multiple protein variants in blood stage (Nunes et al. 2007) Inner surface of RBC membrane or within the parasite (Nunes et al. 2007) | Phosphorylation of host membrane skeleton proteins altering RBC mechanical properties (Nunes et al. 2010) |
| **RMP-fam-a** | *Plasmodium berghei* | 23 (Otto et al.) | Subtelomeric (Hall et al. 2005) | – | – |
| | *Plasmodium yoelii* | 94 (Otto et al.) | Subtelomeric (Carlton et al. 2002) | – | – |
| | *Plasmodium chabaudi* | 132 (Otto et al.) | Subtelomeric, one internal cluster (chr. 13) (Otto et al.) | – | – |
| RMP-fam-b | *Plasmodium berghei* | 34 (Otto et al.) | Subtelomeric (Hall et al. 2005) | – | – |
| | *Plasmodium yoelii* | 48 (Otto et al.) | Subtelomeric (Carlton et al. 2002) | – | – |
| | *Plasmodium chabaudi* | 26 (Otto et al.) | Subtelomeric (Hall et al. 2005) | – | – |
| RMP-fam-c | *Plasmodium yoelii* | 22 (Otto et al.) | Subtelomeric (Carlton et al. 2002) | – | – |
| **srs** | *Toxoplasma gondii* | 104 (Reid et al. 2012) | Intrachromosomal, in tandem arrays of various size (Reid et al. 2012) | Multiple variants in tachyzoites (Tomavo, 1996) | Parasite surface (Tomavo, 1996) Cytoadherence (prior to invasion) (Jung et al. 2004), immune evasion (Kim and Boothroyd, 2005). |
| | *Neospora caninum* | 227 (Reid et al. 2012) | Intrachromosomal, in tandem arrays of various size (Reid et al. 2012) | Multiple variants in tachyzoites (Reid et al. 2012) | Parasite surface (Howe et al. 1998) – |
| **susa** | *Toxoplasma gondii* | 26 (Reid et al. 2012) | Intrachromosomal, with three tandem arrays (Reid et al. 2012) | Multiple variants in tachyzoites (Reid et al. 2012), Expressed in bradyzoites (Pollard et al. 2008) | Parasite surface (Pollard et al. 2008) – |
| | *Neospora caninum* | 38 (Reid et al. 2012) | Intrachromosomal, with three tandem arrays (Reid et al. 2012) | None in tachyzoites (Reid et al. 2012) | – |
| **sag** | *Eimeria spp.* | 16–172 (Reid, 2014 #517) | Intrachromosomal, with few, large tandem arrays (Reid, 2014 #517) | Multiple variants (Tabares et al. 2004) | Parasite surface (Tabares et al. 2004) – |
| **efl** | *Eimeria spp.* | 18–50 (Reid, 2014 #517) | Maybe in tandem (Reid, 2014 #517) | Multiple variants across different life stages (Reid, 2014 #517) | – |
| Family Subfamily | Species | Number per genome | Genomic location | Gene expression | Cellular localization of product | Function(s) |
|-----------------|---------|-------------------|------------------|----------------|---------------------------------|-------------|
| *esf2*          | *Eimeria* spp. | 2–23 (Reid, 2014 #517) | Tandemly arrayed (Reid, 2014 #517) | Multiple variants across different life stages (Reid, 2014 #517) | – | – |
| *rokh*          | *Toxoplasma gondii* | 55 (Talevich and Kannan, 2013) | Intrachromosomal, some small arrays (Reese et al. 2011) | Multiple variants in tachyzoites (Reid et al. 2012) | ROP18 localizes to host cytosol/ PVM (Taylor et al. 2006), ROP16 localizes to host nucleus (Saeij et al. 2007) | ROP18 protects PVM from attack by host IRGs (Fentress et al. 2010), ROP16 modifies host cell signalling (Saeij et al. 2007). |
| *Neospora caninum* | 44 (Talevich and Kannan, 2013) | Intrachromosomal (Reid et al. 2012) | Multiple variants in tachyzoites (Reid et al. 2012) | – | – |
| *Eimeria* spp. | 28 (Talevich and Kannan, 2013) | Intrachromosomal (Reid, 2014 #517) | Multiple variants across different life stages (Reid, 2014 #517) | – | – |
| *ves* | *ves1* | *Babesia bovis* | 72 ves1alpha, 43 ves1beta (Brayton et al. 2007) | Single transcript transcribed (Zupanska, 2009 #516) | RBC surface (O’Connor et al. 1997) | Cytoadherence (O’Connor and Allred, 2000), antigenic variation (Allred et al. 2000) |
| *Babesia bigemina* | 74 ves1a, 80 ves1b (Jackson et al. 2014) | – | – | – | – |
| *Babesia divergens* | 202 ves1 (Jackson et al. 2014) | Subtelomeric tandem arrays (Jackson et al. 2014) | – | – | – |
| *ves2* | *Babesia bigemina* | 116 (Jackson et al. 2014) | – | – | Secreted? (Jackson et al. 2014) | – |
| *Babesia divergens* | 95 ves2a, 62 ves2b (Jackson et al. 2014) | – | – | Secreted? (Jackson et al. 2014) | – |
| *smORF* | *Babesia bovis* | 44 (Brayton et al. 2007) | Intrachromosomal, associated with ves1 genes (Brayton et al. 2007) | Multiple variants transcribed (Brayton et al. 2007) | Secreted? (Jackson et al. 2014) | Analogous to ves2? (Jackson et al. 2014) |
| *bm* | *Babesia microti* | 24 (Cornillot et al. 2012) | Subtelomeric and intrachromosomal (Cornillot et al. 2012) | Single variant protein (Homer et al. 2000) | – | – |
| *Tpr/Tar* | *Theileria parva* | 39 (Gardner et al. 2005) | Single tandem array of 28, 11 dispersed copies (Gardner et al. 2005) | – | Predicted integral membrane proteins (Weir et al. 2010) | – |
| *Theileria annulata* | 93 (Weir et al. 2010) | Dispersed in small arrays (Pain et al. 2005) | – | Predicted integral membrane proteins (Weir et al. 2010) | – |
| *svp* | *Theileria parva* | 85 (Gardner et al. 2005) | Subtelomeric tandem arrays (Gardner et al. 2005) | All genes expressed in macroschizont (Schmuckli-Maurer et al. 2009) | Predicted secreted into host cell (Weir et al. 2010) | – |
| *Theileria annulata* | 51 (Weir et al. 2010) | Subtelomeric tandem arrays (Pain et al. 2005) | All genes expressed in macroschizont (Pain et al. 2005) | Predicted secreted into host cell (Weir et al. 2010) | – |
Theileria parasites invade and transform host leucocytes (reviewed in Dobbelaere and Kuenzi, 2004). It is thought that they transform the host cell using families of secreted proteins. On such example is the SVSP family, which encode a signal peptide and a Frequently Associated IN Theileria (FAINT) domain. FAINT domains are common to a range of Theileria subtelomeric genes from different families, all predicted to be secreted (Pain et al. 2005). It has been speculated that this low-complexity glutamine- and proline-rich domain may be difficult for vertebrate immune systems to recognize and therefore could provide some anonymity for the parasite (Gardner et al. 2005).

Domain structures in these families suggest a diverse array of structures and of phylogenetic origins for surface proteins and other families. This is likely to be a result of the requirement for constant innovation in the parasites’ battle with the host but also of differing requirements related to the niche of the parasite and its strategy for transmission.

**Expression Patterns and Protein Localization**

Gene families involved in antigenic variation sensu stricto are expressed monoallelically, with only one family member localized to the parasite or host cell surface at a time (Borst, 1991). Although this is observed for var and VESA families, it is not what has been observed for most apicomplexan large gene families. Frequently, where measurements have been made, multiple members are expressed. Unlike switching of vsg in trypanosomes, switching to different genes in apicomplexans does not involve genomic rearrangement. Although many contingency gene families encode protein localized to the parasite or host cell surface, some families localize to the host cytosol or nucleus (e.g. ropk; Fig. 2).

Correct expression of *P. falciparum* var genes involves the suppression of all but one gene and switching between individual genes; control occurs at the level of transcription initiation (Scherf et al. 2008). Expression is the highest early after cell invasion, presumably because the protein must quickly be expressed on the surface of the invaded red cell. Control of transcription is thought to require genetic elements such as the intron (Deitsch et al. 2001) and the ups promoter region (Voss et al. 2006). An ApiAP2 transcription factor is known to bind to the intron, which acts as a bidirectional promoter and from which non-coding RNA is transcribed (Zhang et al. 2011). *Plasmodium falciparum* subtelomeres are characteristically marked by the heterochromatic epigenetic mark H3K9me3 (Salcedo-Amaya et al. 2009). During transcription the active var gene is marked by more highly acetylated histone H4 and
also H3K4me3 and H3K4me2 (Freitas-Junior et al. 2005; Lopez-Rubio et al. 2007). When not expressed it loses H3K4me3, but retains H3K4me2. This is thought to act as memory so that the same var gene is activated in the next invaded red cell (Lopez-Rubio et al. 2007). Conversely, methylation of H3K36 represses var gene expression (Jiang et al. 2013).

The mechanism for controlling monoallelic expression of Babesia ves1 genes is less well understood. Expressed genes are found in a tandem configuration and are expressed from LATs by a bidirectional promoter. It is thought that chromatin remodelling is required for transcription suggesting there may be some commonalities with control of var expression (Huang et al. 2013). However, it is not clear that chromatin remodelling plays a direct role in monoallelic expression.

For all other apicomplexan families multiple members appear to be expressed at once. A small number of P. knowlesi SICAvars are expressed by the parasite at the same time (Barnwell et al. 1983). Interestingly none are expressed in splenectomized monkeys, suggesting that they are regulated in response to the host immune system (Barnwell et al. 1983). Such a response has also been observed for pir genes in several Plasmodium species. In P. chabaudi it was shown that parasites, which are serially passaged between mice, avoiding the mosquito...
and liver stages, express a reduced set of pirs and RMP-fam-as and that this is associated with an increase in virulence (Spence et al. 2013). Whereas perhaps 10–20% of the pir repertoire (20–40 genes) is expressed in serially blood-passaged parasites (Lawton et al. 2012), mosquito transmitted parasites express more than half of their repertoire (>100 genes), quite the opposite of what we would expect from genes involved in antigenic variation sensu stricto (Spence et al. 2013). Similar results have been shown for Plasmodium yoelii, with B or T cell deficient mice showing differences in expression of these genes (Cunningham et al. 2005). Change in expression of the srs family of surface antigens in T. gondii is known to be important for evading the host immune response (Tomavo, 1996). Around half its repertoire of srs genes is expressed during the rapidly growing tachyzoite stage and multiple protein products have been localized to the parasite cell surface (Tomavo, 2001; Reid et al. 2012).

Fig. 2. Localization of gene family products. The localization of gene products is shown for intracellular and extracellular parasites, where known or hypothesized. Where multiple copies have been localized this is indicated by a bundle of circles. Abbreviations: PVM, parasitophorous vacuole membrane. n.b. The PVM is not present in Theileria infected leucocytes and is destroyed quickly after invasion by Babesia (Asada, 2012 #518). Evidence: TgROP18 localizes to host cytosol/PVM (Taylor et al. 2006), TgROP16 localizes to host nucleus (Saeij et al. 2007), TpSVSP predicted to localize to host cytosol (Weir et al. 2010), PIRF/Stevor localize to parasite apex (Petter et al. 2007; Blythe et al. 2008) and RBC surface (Kyes et al. 1999), PvVIR14 and PvVIR10 are exported to the membrane of iRBC whereas PvVIR17 remains inside the parasite (Bernabeu et al. 2012), rodent PIR family localize on or close to the surface of the RBC (Cunningham et al. 2009), TgSRS multiple gene products have been localized to the parasite surface (Tomavo, 1996), EntSAG multiple gene products localized to the parasite surface (Tabares et al. 2004), PfFIKks localize to Maurer’s cleft/host membrane cytoskeleton and within parasite (Nunes et al. 2010), PiEMP1 (var family) localizes to the host cell surface (Kyes et al. 2001), PkSICA (SICAvar family) localizes to the host cell surface (Howard et al. 1983), BbVES (ves1 family) localizes to the host cell surface (O’Connor et al. 1997).

and liver stages, express a reduced set of pirs and RMP-fam-as and that this is associated with an increase in virulence (Spence et al. 2013). Whereas perhaps 10–20% of the pir repertoire (20–40 genes) is expressed in serially blood-passaged parasites (Lawton et al. 2012), mosquito transmitted parasites express more than half of their repertoire (>100 genes), quite the opposite of what we would expect from genes involved in antigenic variation sensu stricto (Spence et al. 2013). Similar results have been shown for Plasmodium yoelii, with B or T cell deficient mice showing differences in expression of these genes (Cunningham et al. 2005). Change in expression of the srs family of surface antigens in T. gondii is known to be important for evading the host immune response (Tomavo, 1996). Around half its repertoire of srs genes is expressed during the rapidly growing tachyzoite stage and multiple protein products have been localized to the parasite cell surface (Tomavo, 2001; Reid et al. 2012). During switch to the encysted, bradyzoite form, there is a switch to a largely non-overlapping set of srs genes (Jung et al. 2004). The regulatory mechanism controlling this is unknown.

Several families are known to localize within the host cell and have been shown to adapt the host cell to function as a better home for the parasite. Fikh kinases are expressed in the blood stage of P. falciparum, with different members expressed at different stages of development. Their products have been shown to locate to within the parasite itself, the Maurer’s cleft in the host erythrocyte and to the inner side of the erythrocyte membrane (Nunes et al. 2007). Transcriptional data indicates that unlike Shi1 sub-telomeric genes, SVSP family genes of T. annulata are expressed in a stage-specific manner by the macroschizont and that the majority of the gene products contribute to the secretome (Weir et al. 2010). These are predicted to localize to the host cytosol. No members of large gene families in Theileria are thought to localize to the parasite or host membranes.

Genomic data can give us clues about whether gene products are secreted, exported, membrane-bound
or GPI-anchored. However, experimental approaches are required to confirm this and provide greater detail. There is great difficulty in localizing multiple members of protein families due to the requirement for specific antibodies or genetic manipulation of multiple, similar, genes. Improvements in locating the product of each family member over time would allow greater insight into a family’s function.

**FUNCTIONAL ROLES**

It is generally assumed that large variant gene families have a role in antigenic variation, this being the most straightforward explanation for maintaining such a large repertoire of related genes. However, it is clear that most large gene families in Apicomplexa are not involved in antigenic variation as it is often understood and many may not be involved in antigenic variation at all (Table 1). It has been proposed that gene families, which are involved in antigenic variation, might have a distinct primary function. The primary functions of var and/or antigenic variation because it was recently exposed to the host immune system, not only to provide a variety of epitopes. Other Plasmodium species do not have the var genes, or a family, which might obviously replace them, except perhaps for SICAVar in *P. knowlesi*. However, *P. chabaudi* does sequester (Brugat et al. 2014) and shows antigenic variation (Branan et al. 1994; Phillips et al. 1997), suggesting these functions can be achieved by genes with very different properties. It is possible that immune evasion could be mediated by expression of all or most of the repertoire of a gene family. Many similar proteins, expressed at relatively low levels, might prevent induction of an effective immune response, resulting in immune evasion without mutually exclusive expression (Cunningham et al. 2005, 2009). Ves1 genes in *Babesia bovis* are implicated in antigenic variation and cytoadherence and, like *Plasmodium*, this species can sequester in the brain and cause neurological damage (Sondgeroth et al. 2013).

*Pir* is the only large gene family conserved across all Plasmodium species examined. It has been shown that *pir* and *RMP-fam-a* expression (Table 1) is modulated in response to the host immune system (Cunningham et al. 2005; Spence et al. 2013). In particular, expression of a smaller repertoire of *P. chabaudi* *pir* is associated with higher parasitemia and more damage to the host. These genes may have a role in attracting the immune system to a sufficient degree that it is prevented from doing excessive harm to the host.

A-type RIFINs and STEVORs have been shown to locate to the apical end of the invading merozoite stage of *P. falciparum*. It has been suggested that they provide a shield from immune attack against the more conserved proteins involved in invasion, through binding to erythrocyte surface molecules (Petter et al. 2007; Blythe et al. 2008). There is evidence for molecular mimicry amongst the related *pir* family in *P. knowlesi* (Pain et al. 2008). Seven *P. knowlesi* *pir* contain regions of up to 36 amino acids, which perfectly match the immune system protein CD99 from the natural host *Macaca mulatta*. This would suggest that these *pirs* may compete with host CD99 for binding and could interfere with the host immune response.

Within large variant gene families individual members are able to mutate quite freely due to redundancy of function and reduced selective pressure. They may therefore have a great deal of opportunity to develop new functions. Despite the great differences in var gene repertoire, even between different strains of *P. falciparum*, a small number are have been under strong positive selection to maintain their structure and function (Rowe et al. 2002). *Var2csa* is able to bind chondroitin sulphate A on the placenta and its expression is associated with placental malaria in women pregnant for the first time (Salanti et al. 2004).

DBLa domains are found in var genes and in genes involved in invasion in *P. falciparum*. The srs gene family takes a similar multi-functional role in Toxoplasma. Host cell invasion by apicomplexans is preceded by attachment and reorientation (Carruthers and Boothroyd, 2007). In Toxoplasma, the same gene family involved in immune evasion is also involved in the first of these processes (Jung et al. 2004). SRS57 has been shown to act in binding host cells (Dzierzinski et al. 2000), while SRS34A has a role in attracting an immune response. SRS29B, SRS34A and SRS29C are specific to the rapidly growing tachyzoite stage and are strongly antigenic (Lekutis et al. 2001). Along with many other srs genes these are turned off during the switch to bradyzoites. The tachyzoite may attract the immune system, or the immune system may prompt the parasite to switch forms (Bohne et al. 1993). This may have a role in limiting parasite proliferation and also distracting attention from the bradyzoite, which can then hide in tissue and await transmission. Perhaps this family might have started out with a cell adhesion function, which showed antigenic variation because it was exposed. It could then have evolved a function in immune evasion switching which allowed the quiescent form to hide in tissues.

The intracellular stage of the bovine parasite *Theileria*, equivalent to the malaria blood stage, is not known to express any molecules on the surface of the invaded cell, a leucocyte (Boulter and Hall, 1999). The largest gene family in *Theileria*, Tar/Tpr,
is predicted to encode mostly integral membrane proteins; however, no known function has been attributed to it (Weir et al. 2010). The next two largest variant gene families of \textit{Theileria} appear to encode largely secreted proteins (Pain et al. 2005). It has been proposed that SVSPs have a role in transforming the host cell. They possess nuclear localization signals and it is thought they might localize to the host nucleus much like products of the smaller \textit{TashAT} family (Shiels et al. 2004). Alternatively they might provide an immunological smokescreen effect by distracting immune cells from more conserved surface-bound antigens (Weir et al. 2010).

There are large kinase gene families from \textit{Plasmodium} and \textit{Coccidia}, which are also thought to be involved in modifying the host cell. \textit{Plasmodium falciparum} has a unique expansion of the Apicomplexan-specific FIKK family of kinases (Nunes et al. 2007). There is evidence that some members are involved in phosphorylation of host membrane skeleton proteins and alter the mechanical properties of the red blood cell (Nunes et al. 2010). ROP kinases and inactivated pseudokinases are stored in club-shaped organelles known as rhoptries, which introduce their cargo to the host cell upon invasion. In \textit{T. gondii}, ROP16 localizes to host nucleus where it activates the JAK-STAT pathway and modulates host gene expression (Saeij et al. 2007). TgROP18 localizes to the PVM (Taylor et al. 2006) and is able to prevent immune-related GTPases (IRGs) from attacking this membrane and destroying the parasite (Fentress et al. 2010). This family seems to provide a set of exquisite tools for modifying the host cell.

It has been proposed that \textit{Cryptosporidium} species do not display any form of antigenic variation (Singh et al. 2005). Interestingly, they also have a paucity of large variant gene families. Only the mucin-like glycoproteins show a large expansion. These have been proposed to function in tethering the sporozoite stage to the oocyst wall. However, it is not clear why 30 genes would be required to perform this task. Why \textit{Cryptosporidium} should be able to survive without the wealth of large gene families present in other apicomplexans is currently unclear.

Thus, a variety of roles have already been uncovered for large apicomplexan gene families. Most families however have no known function and await characterization (Table 1). Small research communities are assembling around some of these families, but many remain essentially unstudied.

**CONCLUSIONS**

The \textit{P. falciparum var} gene family is the most intensely studied large gene family in the \textit{Apicomplexa}. It is involved in the best-understood mechanisms of antigenic variation and sequestration in this phylum. While other species exhibit antigenic variation and sequestration they use gene families with different properties and evolutionary histories. Many gene families are concerned with manipulating the host immune response or modifying the host cell rather than antigenic variation or sequestration. This review has highlighted some of the common themes and important contrasts that exist amongst large variant gene families in apicomplexan species. These may be useful in directing research to understand how apicomplexans interact with their hosts.

**ACKNOWLEDGEMENTS**

I would like to thank Chris Newbold and Magdalena Zarowiecki for critical reading of the manuscript.

**FINANCIAL SUPPORT**

This work was supported by the Wellcome Trust.

**REFERENCES**

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

Al-Khedery, B. and Allred, D. R. (2006). Antigenic variation in Babesia bovis occurs through segmental gene conversion of the ves multigene family, within a bidirectional locus of active transcription. \textit{Molecular Microbiology} 59, 402–414.

Allred, D. R., Carlton, J. M., Satcher, R. L., Long, J. A., Brown, W. C., Patterson, P. E., O’Connor, R. M. and Stroup, S. E. (2000). The ves multigene family of \textit{B. bovis} encodes components of rapid antigenic variation at the infected erythrocyte surface. \textit{Molecular Cell} 5, 153–162.

Arredondo, S. A., Cai, M., Takeyama, 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.

Assefa, S. (2013). \textit{De Novo Assembly of the Var Multi-Gene Family in Clinical Samples of Plasmodium Falciparum}. PhD thesis, University of Cambridge, UK.

Aurrøechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gairia, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Miller, J. A., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross, C., Stoeckert, C. J., Jr., Treatman, C. and Wang, H. (2009). PlasmoDB: a functional genomic database for malaria parasites. \textit{Nucleic Acids Research} 37, D539–543.

Barnwell, J. W., Howard, R. J., Coon, H. G. and Miller, L. H. (1983). Splenic requirement for antigenic variation and expression of the variant antigen on the erythrocyte membrane in cloned Plasmodium knowlesi malarial. \textit{Infection and Immunity} 40, 985–994.

Bayliss, C. D., Field, D. and Moxon, E. R. (2001). The simple sequence contingency loci of \textit{Haemophilus influenzae} and \textit{Neisseria meningitidis}. \textit{Journal of Clinical Investigation} 107, 657–662.

Bernabeu, M., Lopez, F. J., Ferrer, M., Martin-Jaular, L., Razaname, A., Corradin, G., Maier, A. G., Del Portillo, H. A. and Fernandez-Becerra, C. (2012). Functional analysis of Plasmodium vivax VIR proteins reveals different subcellular localizations and cytotoxidence to the ICAM-1 endothelial receptor. \textit{Cellular Microbiology} 14, 386–400.

Blythe, J. E., Yam, X. Y., Kuss, C., Bozdech, Z., Holder, A. A., Marsh, K., Langborne, J. and Preiser, P. R. (2008). \textit{Plasmodium falciparum} STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. \textit{Infection and Immunity} 76, 329–336.
Bowe, N., Heesemann, J. and Gross, U. (1993). Induction of bradyzoite-specific Toxoplasma gondii antigens in gamma interferon-treated mouse macrophages. *Infection and Immunity* 61, 1141–1145.

Borst, P. (1991). Molecular genetics of antigenic variation. *Immunity* 12, 239–33.

Boulter, N. and Hall, R. (1999). Immunity and vaccine development in the bovine theilerioses. *Advances in Parasitology* 44, 41–97.

Bozdech, Z., Linias, M., Pulliam, B. L., Wong, E. D., Zhu, J. and DeRisi, J. L. (2003). The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Comput Biol* 1, E5.

Brannan, L. R., Turner, C. M. and Phillips, R. S. (1994). Malaria parasites undergo antigenic variation at high rates in vivo. *Proceedings of the Royal Society B: Biological Sciences* 256, 71–75.

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., Fadros, D., Feldblum, T., Forberger, H. A., Haas, B. J., Howell, J. M., Khouri, H., Koo, H., Mann, D. J., Norimine, J., Paulson, I. T., Radune, D., Ren, Q., Smith, R. R., Jr., Sueng, C. E., White, O., Wortman, J. R., Knowles, D. P., Jr., McElwain, T. F. and Nene, V. M. (2007). Genome sequence of Babesia boliviensis and comparative analysis of apicomplexan hemoprotozoa. *PLoS Pathog* 3, 1401–1413.

Brugat, T., Cunningham, D., Sodenkamp, J., Coomes, S., Wilson, M., Spencer, W., Thompson, C. S. and Langhorne, J. (2014). Sequestration and histopathology in Plasmodium chabaudi malaria are influenced by the immune response in an organ-specific manner. *Cellular Microbiology* 16, 687–700.

Bultrini, E., Brick, K., Mukherjee, S., Zhang, Y., Silvestrini, F., Alano, P. and Pizzi, E. (2009). Revisiting the Plasmodium falciparum RIFIN family: from comparative genomics to 3D-model prediction. *RNA* 10, 392–405.

Carlton, J. M., Ingoli, S. V., Sub, B. B., Koool, T. W., Pertea, M., Silva, J. C., Ermolaeva, M. D., Allen, E. E., Selengut, J. D., Koo, H. L., Petrie, S. H., Wirth, D., Muka, B., Dzyierszinski, F., Voss, T. S., Maier, A. G., Pearce, J. A., Deitsch, K. W., Calderwood, M. S. and Wellems, T. E. (2004). The strategies of the Plasmodium falciparum yir yir family which potentially encode variant antigens. *Molecular and Biochemical Parasitology* 138, 33–44.

Crucrues, V. and Boothroyd, J. C. (2007). Pulling together: Plasmodium falciparum erythrocyte membrane protein-1 specifically suppresses early production of host interferon-gamma. *Cell Host and Microbe* 2, 130–138.

Dzierzinski, F., Mortuaire, M., Cesbron-Delauv, M. F. and Tomova, S. (2009). Telomeric heterochromatin propagation and histone acetylation of the malarial redomin parasite Plasmodium yoelii yoelii genes in telomeric chromosome clusters of *Plasmodium yoelii yoelii*. *Nature* 455, 757–760.

Dutta, S., Hagblom, P. and Wahlgren, M. (2005). Developmental regulation of the glucocerebrosidase, a N-linked glycoprotein that tether sporozoites of Plasmodium yoelii yoelii and its role in host cell adhesion and drastically reduces virulence in mice. *Cell Microbiology* 7, 574–585.

Fass, D. (2012). Sialylbinding in protein biochemistry. *Annual Review of Biophysics* 41, 63–79.

Fentress, S. J., Behke, M. S., Dunay, I. R., Mashayekhi, M., Rommereim, L. M., Fox, B. A., Bizik, D. J., Taylor, G. A., Turk, B. E., Lichti, C. F., Townsend, R. R., Qiu, W., Hui, R., Beatty, W. L. and Sibley, L. D. (2010). Phosphorylation of immunity-related GTPases by a Toxoplasma gondii-secreted kinase promotes macropage survival and virulence. *Cell Host and Microbe* 8, 484–495.

Fernandez-Becerra, C., Pein, O., de Oliveira, T. R., Yamamoto, M. M., Cassola, A. C., Rocha, C., Soares, I. S., de Braganca Pereira, C. A. and del Portillo, H. A. (2005). Variant proteins of Plasmodium falciparum are not clonally expressed in natural infections. *Molecular Microbiology* 58, 648–658.

Figueiredo, L. M., Pirrit, L. A. and Scherf, A. (2000). Genomic organisation and chromatin structure of Plasmodium falciparum chromosomes. *Molecular and Biochemical Parasitology* 106, 169–174.

Freitas-Junior, L. H., Bottiu, E., Pirrit, L. A., Deitsch, K. W., Scheidig, C., Guinet, F., Nehrbs, U., Wellemes, T. E. and Scherf, A. (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* 407, 1021–1026.

Freitas-Junior, L. H., Hernandez-Rivas, R., Ralph, S. A., Montiel-Condado, D., Ruvalcaba-Salazar, O. K., Rojas-Meza, A. P., Mancio-Silva, L., Leal-Silvestre, R. J., Gontijo, A. M., Pearce, J. A., Deitsch, K. W., Calderwood, M. S. and Wellems, T. E. and Scherf, A. (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* 407, 1021–1026.

Gajria, B., Bahl, A., Bestelli, J., Dommer, J., Fischer, S., Gao, X., Heiges, M., Iodice, J., Kissinger, J. C., Mackey, A. J., Pinney, D. F., Roos, D. S., Steetcork, C. J. Jr., Wang, H. and Brunk, B. P. (2008). ToxoDB: an integrated Toxoplasma gondii database resource. *Nucleic Acids Research* 36, D553–D556.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shollom, S. J., Sub, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, W. M., Vaidya, A. B. and Martin, D. M. and D. M. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* 419, 498–502.

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., Shollom, S. J., Weidman, J., Jiang, L., Lynn, J., Weaver, B., Shoab, A., Domingo, A. R., Wirth, D., Crabtree, J., Wortman, J. R., Haas, B., Angiuoli, S. V., Creasy, T. H., Lu, C., Sub, B. (2005). Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. *Science* 314, 134–137.

Guizetti, J. and Scherf, A. (2013). Silence, activate, poise and switch! Mechanisms of antigenic variation in *Plasmodium falciparum*. *Cellular Microbiology* 15, 718–726.
Hall, N., Karras, M., Raine, J. D., Carlson, J. M., Kooij, T. W., Berriman, M., Florens, L., Janssen, C. S., Pain, A., Christophides, G. K., James, K., Rutherford, K., Harris, B., Harris, D., Churcher, C., Quail, M. A., Ormond, D., Doggett, J., Truemauer, F., Mendoza, J., Bidwell, S. L., Rajendraam, M. A., Carucci, D. J., Yates, J. R., 3rd, Kafatos, F. C., Janse, C. J., Barrell, B., Turner, C. M., Waters, A. P. and Sinn, R. E. (2005). A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. Science 307, 82–86.

Hillier, S., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C. and Haldar, K. (2004). A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science 306, 1934–1937.

Homer, M. J., Bruinsma, E. S., Lodds, M. J., Mor, M. H., Telford, S., Krajewski, S., Horrocks, P. and Newbold, C. (2006). Maurer’s clefts in Toxoplasma surface proteins. International Journal for Parasitology 36, 362–368.

Yang, R. J., Barrett, J. S. and Kao, V. (1983). Antigenic variation of Plasmodium knowlesi malaria: identification of the variant antigen on infected erythrocytes. Proceedings of the National Academy of the Sciences of the United States of America 80, 4129–4133.

Howe, D. K., Crawford, A. C., Lindsay, D. and Sibley, L. D. (1998). The protein and antigenic variants of Neospora caninum tachyzoites are homologous to the family of surface antigens of Toxoplasma gondii. Infection and Immunology 66, 5322–5328.

Huang, Y., Xiao, Y. P. and Allred, D. R. (2013). Unusual chromatin structure associated with monoparous transcription of the Babesia bovis ribosomal multigene family. International Journal for Parasitology 43, 163–172.

Jackson, A. P., Otto, T. D., Darby, A., Ramasubramaniam, S., Xia, D., Echaeid, I. E., Farber, M., Gahlot, S., Gamble, J., Gupta, D., Gupta, J., Jackson, L., Malaria, L. M., Malas, T. B., Moussa, E., Nair, M., Reid, A. J., Sanders, M., Sharma, J., Tracey, A., Quail, M. A., Weis, W., Waistling, J. M., Hall, N. W., Lengelbach, P., Siehrs, B., Tait, A., Berriman, M., Allred, D. R. et al. (2014). The evolutionary dynamics of variant antigen genes in Babesia reveal a history of genomic innovation underlying host-parasite interaction. Nucleic Acids Research 42, 7112–7121.

Janssen, C. S., Phillips, R. S., Turner, C. M. and Barrett, M. P. (2004). Plasmodium interspersed repeats: the major multigene superfamily of Toxoplasma surface proteins. Nucleic Acids Research 32, 5712–5720.

Jiang, L., Mu, J., Zhang, Q., Ni, T., Sinivasan, P., Rayavar, K., Weir, W., Waistling, J. M., Hall, N., Wu, Y., Kim, Y., Reif, A. J., Sanders, M., Sharma, J., Tracey, A., Quail, M. A., Weis, W., Waistling, J. M., Hall, N. W., Lengelbach, P., Siehrs, B., Tait, A., Berriman, M., Allred, D. R. et al. (2014). The evolutionary dynamics of variant antigen genes in Babesia reveal a history of genomic innovation underlying host-parasite interaction. Nucleic Acids Research 42, 7112–7121.

Kaviratne, M., Khan, S. M., Jarra, W. and Preiser, P. R. (2002). Large gene families of the Apicomplexa. International Journal for Parasitology 32, 1665–1681.

Kley, L., Ribeiro, J. M., Scherf, A., Aravind, L., Zhu, J., Zhao, K. and Wei, G., Jing, Q., Wakabayashi, Y., Bansal, A., Luo, Y., Yang, W., Turner, L., Lavstsen, T., Theander, T. G., Peng, W., Jiang, L., Mu, J., Zhang, Q., Ni, T., Srinivasan, P., Rayavar, K., Newbold, C. I., Pain, A., Berriman, M. and Janse, C. J. (2004). A comprehensive evaluation of rodent malaria parasite genomes and gene-expression under review.

Klevecz, R., 3D7, 1934–1937.

Koehl, M. A., Brunk, B. P., Myler, P. J., Roos, D., Carrington, M., Smith, D. F., Hoeijmakers, W. A. M., Religa, A. A., Robertson, L., Sanders, M., Ogun, S. A., Cunningham, D., Erhart, A., Billker, O., Khan, S. M., Stunnenberg, H. G., Langhorne, J., Holder, A. A., Waters, A. P., Newbold, C., Pain, A., Berriman, M. and Janse, C. J. (2014). A comprehensive evaluation of rodent malaria parasite genomes and gene-expression under review.

Koike, T., Murata, T., Takenaka, M., Ikeya, M., Konno, H. and Tomida, S. (2007). Identification of highly polymorphic family of glycosylphosphatidylinositol-anchored surface antigens with evidence of developmental regulation in Toxoplasma gondii. Infection and Immunology 76, 103–110.

Koike, T., Murata, T., Takenaka, M., Ikeya, M., Konno, H. and Tomida, S. (2007). Identification of highly polymorphic family of glycosylphosphatidylinositol-anchored surface antigens with evidence of developmental regulation in Toxoplasma gondii. Infection and Immunology 76, 103–110.

Koike, T., Murata, T., Takenaka, M., Ikeya, M., Konno, H. and Tomida, S. (2007). Identification of highly polymorphic family of glycosylphosphatidylinositol-anchored surface antigens with evidence of developmental regulation in Toxoplasma gondii. Infection and Immunology 76, 103–110.

Koike, T., Murata, T., Takenaka, M., Ikeya, M., Konno, H. and Tomida, S. (2007). Identification of highly polymorphic family of glycosylphosphatidylinositol-anchored surface antigens with evidence of developmental regulation in Toxoplasma gondii. Infection and Immunology 76, 103–110.

Koike, T., Murata, T., Takenaka, M., Ikeya, M., Konno, H. and Tomida, S. (2007). Identification of highly polymorphic family of glycosylphosphatidylinositol-anchored surface antigens with evidence of developmental regulation in Toxoplasma gondii. Infection and Immunology 76, 103–110.

Koike, T., Murata, T., Takenaka, M., Ikeya, M., Konno, H. and Tomida, S. (2007). Identification of highly polymorphic family of glycosylphosphatidylinositol-anchored surface antigens with evidence of developmental regulation in Toxoplasma gondii. Infection and Immunology 76, 103–110.

Koike, T., Murata, T., Takenaka, M., Ikeya, M., Konno, H. and Tomida, S. (2007). Identification of highly polymorphic family of glycosylphosphatidylinositol-anchored surface antigens with evidence of developmental regulation in Toxoplasma gondii. Infection and Immunology 76, 103–110.

Koike, T., Murata, T., Takenaka, M., Ikeya, M., Konno, H. and Tomida, S. (2007). Identification of highly polymorphic family of glycosylphosphatidylinositol-anchored surface antigens with evidence of developmental regulation in Toxoplasma gondii. Infection and Immunology 76, 103–110.
Ralph, S. A., Bischof, E., Mattei, D., Sismeiro, O., Dillies, M. A., Guigon, G., Coppee, J. Y., David, P. H. and Scherf, A. (2005). Transcriptome analysis of antigenic variation in Plasmodium falciparum – var silencing is not dependent on antisense RNA. *Genome Biology* 6, 189.

Reese, M. L., Zeiner, G. M., Saeij, J. P., Boothroyd, J. C. and Boyle, J. P. (2011). Polymorphic family of injected pseudokinases is paramount in Toxoplasma virulence. *Proceedings of the National Academy of Sciences of the United States of America* 108, 9625–9630.

Reid, A. J., Vermont, S. J., Cotton, J. A., Harris, D., Hill-Cawthorne, G., Konen-Waisman, S., Latham, S., Mourié, T., Norton, R., Quall, M. A., Sanders, M., Shanmugam, D., Sølak, A., Wasmuth, J. D., Brunk, B., Grigg, M. E., Howard, J. C., Parkinson, J., Roos, D. S., Trees, A. J., Berriman, M., Pain, A. and Wastling, J. M. (2012). Comparative genomics of the apicomplexan parasites *Toxoplasma gondii* and *Neospora caninum*: Coccidia differing in host range and transmission strategy. *PLoS Pathogens* 8, e1002567.

Reid, A. J., Blake, D. P., Anami, H. R., Billington, K., Browne, H. P., Breyer, J. M., Pasman, M., Hung, S. S., Kawahara, F., Miranda-Saavedra, D., Malas, T., Mourié, 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., Wu, X., Barrell, B., Bear, P. H., Dorfer, C., Gruber, A., Ivens, A. C., Parkinson, J., Rajandream, M. A., Shirley, M. W., Wan, K. I., Berriman, M., Tomley, F. M. and Pain, A. (2014). Genomic analysis of the causative agents of coccidiosis in domestic chickens. *Genome Res.* doi: 10.1101/gr.169553.113.

Rowe, J. A., Moulds, J. M., Newbold, C. I. and Miller, L. H. (1999). A var gene promoter controls allelic exclusion of virulence genes in Plasmodium falciparum. *The Journal of Experimental Medicine* 190, 1197–1201.

Rowe, J. A., Kyes, S. A., Rogerson, S. J., Babiker, H. A. and Raza, A. (2002). Identification of a conserved Plasmodium falciparum var gene unique to malaria in pregnancy. *Journal of Infectious Diseases* 185, 1207–1211.

Saej, J. P., Coller, S., Boyle, J. P., Jerome, M. E., White, M. W. and Boothroyd, J. C. (2007). Toxoplasma co-opts host gene expression by injection of a polymorphic kinase homologue. *Nature* 445, 324–327.

Salanti, A., Dahlback, M., Turner, L., Nielsen, M. A., Barford, L., Magistrado, P., Jensen, A. T., Lavstsen, T., Ofori, M. F., Marsh, K., Rowley, J. A., Kyes, S. A., Rogerson, S. J., Babiker, H. A. and Raza, A. (2005). Evidence for the involvement of a conserved Plasmodium falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature* 388, 292–295.

Saej, J. P., Khin, M. L., Pan, W. and Scherf, A. (2011). *Theileria annulata* DNA binding protein localized to the host cell nucleus alters the phenotype of a bovine macrophage cell line. *Eukaryotic cell* 3, 495–503.

Singh, I., Theodos, C. and Tzipori, S. (2005). Recombinant proteins of *Cryptosporidium parvum* induce proliferation of mesenteric lymph node cells in infected mice. *Infection and Immunity* 73, 5245–5248.

Soudroyer, K. S., Melwain, T. F., Allen, A. J., Chen, A. V. and Lau, A. O. (2013). Loss of neurovirulence is associated with reduction of cerebral capillary sequestration during acute Babesia bovis infection. *Parasites and Vectors* 6, 181.

Spence, P. J., Jarra, L., Levy, P., Reid, A. J., Chappell, L., Brurat, T., Sanders, M., Berriman, M. and Langhorne, J. (2013). Vector transmission regulates immune control of Plasmodium virulence. *Nature* 498, 228–231.

Tabares, E., Ferguson, D., Clark, J., Soon, P. E. W., Pain, A. L. and Tomley, F. M. (2004). *Eimeria tenella* sporozoites and merozoites differentially express glycosylphosphatidylinositol-anchored variant surface proteins. *Molecular and Biochemical Parasitology* 135, 123–132.

Tachibana, S., Sullivan, S. A., Kawai, S., Nakamura, S., Kim, H. R., Goto, N., Arisue, N., Palacpac, N. M., Honna, H., Yagi, M., Tougan, T., Kayakata, Y., Kaneko, O., Mita, T., Kita, K., Yasutomi, Y., Sutton, P. L., Shakhbutyan, R., Horii, T., Yasuaga, T., Barowell, J. W., Excoffier, A. C., Carlton, J. M. and Tanabe, K. (2012). *Plasmodium cynomolgi* genome sequences provide insight into Plasmodium vivax and the monkey malaria clade. *Nature Genetics* 44, 1051–1055.

Talevich, E. and Kannan, N. (2013). Structural and evolutionary adaptation of rhoptry kinases and pseudokinases, a family of coccidian virulence factors. *BMC Evolutionary Biology* 13, 117.

Taylor, S., Barragan, A., Su, C., Fux, B., Fentress, S. J., Tang, K., Beatty, W. L., Haji, H. E., Jerome, M., Behnke, M. S., White, M., Wootton, J. C. and Sibley, L. D. (2006). A secreted serine/threonine kinase determines virulence in the eukaryotic pathogen *Toxoplasma gondii*. *Science* 314, 1776–1780.

Tolia, N. H., Enemark, E. J., Sim, B. K. and Joshua-Tor, L. (2005). Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite *Plasmodium falciparum*. *Cell* 122, 183–193.

Tomavo, S. (1996). The major surface proteins of *Toxoplasma gondii*: structures and functions. *Current Topics in Microbiology and Immunology* 219, 45–54.

Tomavo, S. (2001). The differential expression of multiple isoenzyme forms during stage conversion of *Toxoplasma gondii*: an adaptive developmental strategy. *International Journal for Parasitology* 31, 1023–1031.

Voss, T. S., Healer, J., Marty, A. J., Duffy, M. F., Thompson, J. K., Beeson, J. G., Reeder, J. C., Grabb, B. S. and Cowman, A. F. (2006). A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. *Nature* 439, 1004–1008.

Wasmuth, J. D., Pszenny, V., Haile, S., Jansen, E. M., Gast, A. T., Sher, A., Boyle, J. P., Boulanger, M. J., Parkinson, J. and Grigg, M. E. (2012). Integrated bioinformatic and targeted deletion analyses of the SRS gene superfamily identify SRS295C as a negative regulator of *Toxoplasma* virulence. *MBio* 3, e00321-12.

Weir, W., Karagenc, T., Baird, M., Tait, A. and Shiels, B. R. (2010). Evolution and diversity of secretome genes in the apicomplexan parasite *Theileria annulata*. *BMC Genomics* 11, 42.

Zheng, Q., Huang, Y., Zhang, Y., Fang, X., Claeys, A., D:uchateau, M., Namane, A., Lopez-Rubio, J. J., Pan, W. and Scherf, A. (2011). A critical role of perinuclear filamentous actin in spatial repositioning and mutually exclusive expression of virulence genes in malaria parasites. *Cell Host and Microbe* 10, 451–463.