Short communications

Author for correspondence:
Johannes T. Dessens
e-mail: johannes.dessens@lshtm.ac.uk

Plasmodium LCCL domain-containing modular proteins have their origins in the ancestral alveolate

Callum De Hoest-Thompson[1], Virginia Marugan-Hernandez[1] and Johannes T. Dessens[2]

[1]The Royal Veterinary College, Hawkshead Lane, North Mymms AL9 7TA, UK
[2]Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK

(iD) CDH-T, 0009-0001-5653-5165; VM-H, 0000-0003-1512-4682; JTD, 0000-0002-2070-6073

Plasmodium species encode a unique set of six modular proteins named LCCL lectin domain adhesive-like proteins (LAPs) that operate as a complex and that are essential for malaria parasite transmission from mosquito to vertebrate. LAPs possess complex architectures obtained through unique assemblies of conserved domains associated with lipid, protein and carbohydrate interactions, including the name-defining LCCL domain. Here, we assessed the prevalence of *Plasmodium* LAP orthologues across eukaryotic life. Our findings show orthologous conservation in all apicomplexans, with lineage-specific repertoires acquired through differential *lap* gene loss and duplication. Besides Apicomplexa, LAPs are found in their closest relatives: the photosynthetic chromerids, which encode the broadest repertoire including a novel membrane-bound LCCL protein. LAPs are notably absent from other alveolate lineages (dinoflagellates, perkinsids and ciliates), but are encoded by predatory colponemids, a sister group to the alveolates. These results reveal that the LAPs are much older than previously thought and pre-date not only the Apicomplexa but the Alveolata altogether.

1. Introduction

Malaria parasites encode a set of six conserved modular proteins, named LCCL lectin adhesive-like proteins (LAPs), that are essential for parasite transmission from mosquito to vertebrate [1–4]. The LAPs have amino-terminal endoplasmic reticulum signal peptides and possess modular architectures composed of a variety of domains implicated in lipid, protein and carbohydrate binding, including the typifying LCCL domain (IPR004043, named after the metazoan proteins *Limulus* clotting factor C, Coch-5b2 and Lgl1 where the domain was first identified [5]) (figure 1) [1,4,6]. *Plasmodium* LAPs operate as a protein complex and subcellularly co-localize [7–12], demonstrating a functional relatedness accompanying their structural similarities. However, the precise molecular mechanism by which the LAPs operate in *Plasmodium* remains an enigma.

The LAPs were previously reported to be confined to the phylum Apicomplexa and to exhibit conservation across apicomplexan lineages, albeit with possible differences in their repertoires [6,13]. Given the large expansion of genomic and transcriptomic sequence data for both apicomplexan and other organisms since these studies were carried out more than a decade ago, we decided to re-assess the LAPs in terms of their structure, repertoire and distribution across eukaryotic life. The results obtained confirm the presence of apicomplexan lineage-specific LAP repertoires but also reveal complete
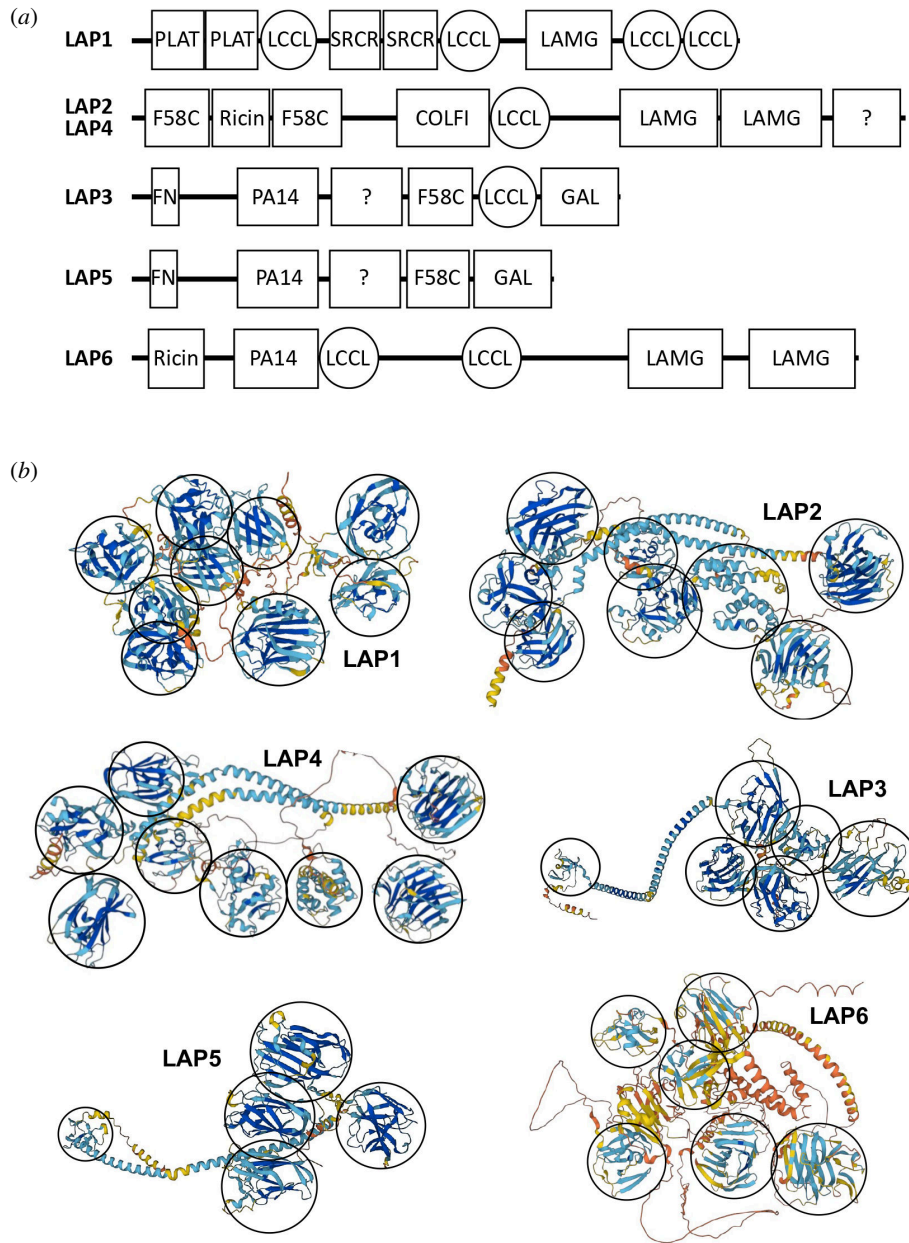
Figure 1. Predicted domain composition and organization of the LAPs. (*a*) Schematic linear depiction of LAP1-LAP6 with various conserved domains indicated: PLAT (polycystin-1, lipoxygenase, alpha-toxin, IPR001024); LCCL (*Limulus* coagulation factor C, Coch-5b2 and Lgl1, IPR004043); SRCR (scavenger receptor cysteine-rich, IPR036772); LAMG (laminin G superfamily, IPR013320, cl22861); F58C (coagulation factor V and VIII carboxy-terminal, IPR000421, related to discoidin); Ricin (ricin B lectin domain, IPR000772); COLFI (fibrillar collagen carboxy-terminal domain, SM00038, PF01410); FN (fibronectin type II domain, IPR000562); PA14 (protective antigen of anthrax toxin, IPR037524); GAL (galactose binding-like domain superfamily, IPR008979); ? (unspecified domains). (*b*) Predicted AlphaFold three-dimensional structure projections of *Plasmodium falciparum* LAP1-LAP6. Protein domains are encircled.

orthologous conservation of the LAPs in their photosynthetic chromerid relatives. Our surprising discovery of LAPs in colponemids shows the LAPs are much older than previously assumed.

2. Results

2.1. LAP structure and domain composition

The six LAPs that were first identified in *Plasmodium* have distinct domain compositions and topologies [14]. The exceptions to this are LAP2 and LAP4, which have identical domain composition and order and are, in effect, structural paralogues (figure 1). Similarly, LAP5 has the same domain topology and composition as LAP3 but lacks the LCCL domain and is included in the LAP family by virtue of its paralogous relationship with LAP3 (figure 1). Thus, the six LAPs represent, in fact, only four distinct structural types referred to in this paper as LAP1, LAP2/4, LAP3/5 and LAP6, respectively (figure 2).

During the analysis of LAP orthologues for the presence of conserved domains, it became apparent that recognized LAP domains are not identified in all orthologues. This probably reflects amino acid sequence divergence between orthologues within the structural constraints of maintaining functionality, allowing domains to elude detection in some orthologues. It
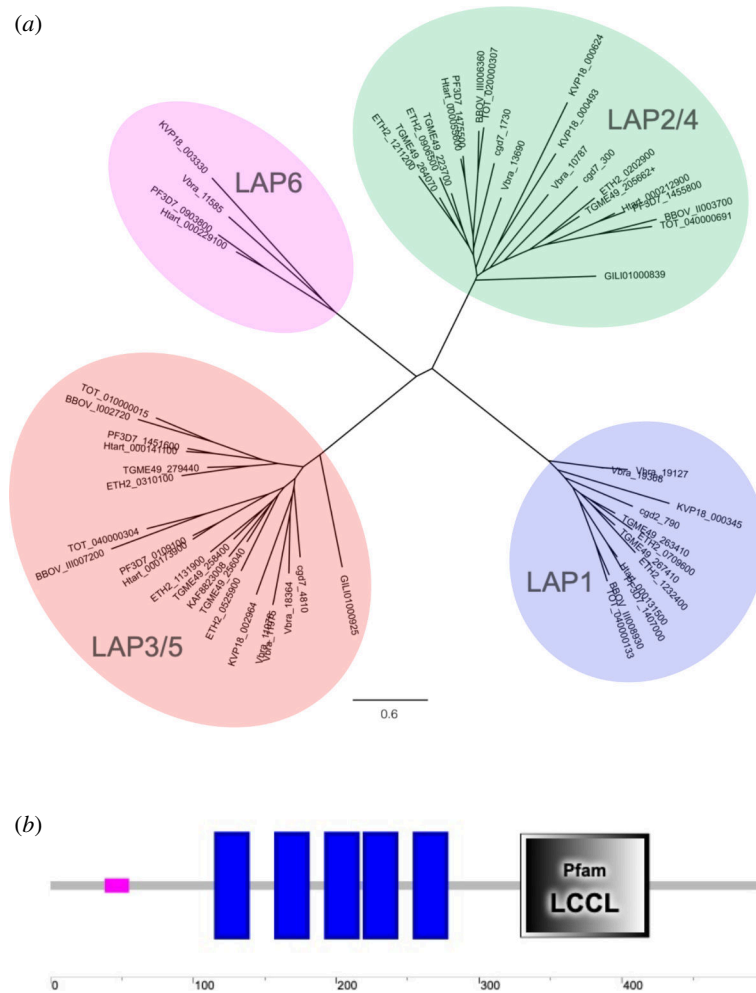
Figure 2. LAPs cluster by type. (*a*) Phylogeny generated from LAP1 type (blue), LAP2/4 type (green), LAP3/5 type (red) and LAP6 type (pink) amino acid sequences identified in apicomplexan, chromerid and colponemid lineages. See electronic supplementary material, table S1 for sequence details. Scale bar: substitutions per site. (*b*) Predicted structure of the membrane-bound LCCL-domain proteins found in chromerids. Blue rectangles represent predicted transmembrane helices (image generated in SMART).

cannot be ruled out, however, that certain domains in certain orthologues may have lost functionality altogether and are retained solely as a linker to maintain the tertiary and quaternary structures of the LAPs, although this seems less likely given their length. The examination of new LAP orthologues also allowed us to identify further, hitherto undetected, modules in the LAP family (figure 1). For example, in LAP1, we identified a second PLAT domain (polycystin-1, lipoxygenase, alpha-toxin, IPR001024) at the amino terminus (figure 1*a*). We also identified a second F58C domain (coagulation factor V and VIII carboxy-terminal, IPR000421), related to discoidin, at the amino terminus of LAP2/4, as well as two tandem LAMG (laminin G) domains in the carboxy-terminal portion of LAP2/4 (figure 1*a*). The latter domain is structurally related to the concanavalin A-like lectin/glucanase domain superfamily (IPR013320) as well as the LamG superfamily (cl22861) and is found in a variety of animal proteins including pentraxins, laminin G, glycosyl hydrolases and plant lectins. A galactose-binding-like domain (IPR008979, GAL) was newly identified in the carboxy-terminal portion of LAP3/5, and we also discovered a second LAMG domain in the carboxy-terminal portion of LAP6 (figure 1*a*).

Examination of the predicted three-dimensional structures of the LAPs in the AlphaFold Protein Structure Database allowed us to identify more precisely the number and relative positions of discrete protein domains in each of the LAPs (figure 1*b*). Integration of this data with our conserved domain searches revealed the presence of two additional, unspecified protein domains in LAP3/5 and LAP2/4, respectively (figure 1*a*). Protein structure comparison by distance matrix alignment (DALI) of resolved three-dimensional structures of conserved protein domains against the AlphaFold-generated three-dimensional structure predictions of the *Plasmodium falciparum* proteome did not identify additional LCCL domain-containing proteins, but confirmed the presence of LCCL (PDB: 1JBI), F58C (PDB: 3NNG), ricin B (PDB: 4P6A), fibronectin type II (FN) (PDB: 7PRK), protective antigen of anthrax toxin (PA14) (PDB: 4GQ7), LAMG (PDB: 6OAI) and scavenger receptor cysteine-rich (SRCR) (PDB: 5JFB) domains in corresponding LAPs (figure 1*a*). Collectively, these results show that structural conservation closely mirrors sequence conservation in this group of proteins.

Some LAP regions were occasionally recognized as distinct conserved domains by different search algorithms/databases or in different orthologues. For example, the F58C domain overlapped with the GAL domain, as well as with a LamNT domain (laminin N-terminal domain, SM000136). Similarly, the COLFI (fibrillar collagen carboxy-terminal domain, SM00038, PF01410) domain overlapped with a FBG domain (fibrinogen-related domain, SM000186). These observations could reflect a common structural architecture of the domains in question, achieved through common ancestry or convergent evolution. Indeed, the

F58C domain forms part of the galactose-binding-like domain superfamily (IPR008979), indicating that it has sugar-binding properties. Furthermore, DALI searches with galectin (PDB: 7XFA) and pea lectin (PDB: 1BQP) identified domains overlapping with the LAMG and GAL domains, indicating these all share a structural scaffold. Overall, the majority of LAP modules appear to possess carbohydrate-binding features.

2.2. LAP distribution and repertoires

A range of LAP orthologues were identified in genomes of the main apicomplexan lineages Hematozoa, Coccidia, Cryptosporidia and Gregarinia (electronic supplementary material, table S1). LAP amino acid sequences clustered according to type by phylogenetic methods (figure 2a), confirming the correct assignment of type to the various LAP sequences. Distinct LAP repertoires were identified in different apicomplexan lineages (table 1). At least one LAP orthologue of each of types LAP1, LAP2/4 and LAP3/5 was identifiable in all lineages examined, while LAP6 was the least prevalent type and is missing in coccidian and cryptosporidian lineages (table 1). LAP6 is also notably absent in the tick-transmitted Piroplasmida (represented by *Theileria orientalis* and *Babesia bovis*), a notable difference from its closely related insect-transmitted relatives (Haemosporida) (table 1). The two families within the Haemosporida lineage: *Plasmodiidae* (represented by *P. falciparum*) and *Haemoproteidae* (represented by *Haemoproteus tartakovskyi*), have identical LAP repertoires, as is true for the two families in the coccidian lineage: *Eimeriidae* (represented by *Eimeria tenella*) and *Sarcocystidae* (represented by *Toxoplasma gondii*) (table 1). We were unable to evaluate LAP repertoires in the apicomplexan lineages Marosporida and Nephromycida [15] owing to incomplete genome coverage. Nonetheless, we identified a LAP3 in *Cardiosporidium cionae* (Nephromycida) (electronic supplementary material, table S1), as well as putative LAP1, LAP2/4 and LAP3-encoding transcripts in *Nephromyces* (Nephromycida) and *Rhytidocystis* (Marosporida) species (not shown), indicating that LAPs are present in all apicomplexan lineages to date.

Beyond the Apicomplexa, LAP orthologues were identified in their closest living relatives: the chromerids (table 1; electronic supplementary material, table S1). The latter constitute a group of coral-associated photosynthetic algae whose plastids share many similarities with the apicoplast including four limiting membranes [16,17]. The Chromerida are considered to share a most recent common ancestor with the Apicomplexa, a view that is strongly supported by phylogenomics [18–21]. The two known species of chromerids, *Chromera velia* and *Vitrella brassicaformis*, together with their predatory relatives the colpodellids, form the chrompodellid lineage [16]. We identified putative LAP-encoding transcripts in the colpodellids (not shown), indicating that the LAPs are encoded across the Chrompodellida.

The analysis of the *V. brassicaformis* genome revealed that the chromerids possess the most extensive LAP repertoire encoding all LAP types (table 1; electronic supplementary material, table S1) plus an additional two LCCL-domain proteins not found in the Apicomplexa, each possessing a single LCCL domain downstream of five predicted transmembrane helices (figure 2b). Assuming that the latter interacts with other LAP members, as do LAPs 1–6 in *Plasmodium* [9,22], this protein could serve to anchor the LAP complex to cellular membranes in the chromerid lineage. Other membrane-bound proteins interacting with the LAP complex, such as NAD(P) transhydrogenase [9], may have made this protein obsolete in apicomplexan lineages like *Plasmodium*. In any event, the LAP repertoire found in chromerids indicates that the most recent common ancestor of the Apicomplexa and Chromerida already possessed genes for all the different LAP types, which subsequently spread to extant lineages by vertical transfer followed by lineage-specific gene duplication or loss.

Apicomplexa and chrompodellids—together with dinoflagellates, perkinsids and ciliates—form the Alveolata supergroup [23]. We did not find LAP orthologues in the latter three alveolate lineages. We could, however, identify intact transcripts encoding full-length LAP3 and LAP2/4 proteins in predatory colponemids (figure 2a; electronic supplementary material, table S1). The availability of only transcriptome shotgun assemblies (TSA) sequence data for this organism precludes a precise determination of its LAP repertoire. AlphaFold predictions of the three-dimensional structures of *Colponema vietnamica* LAP3 and LAP2/4 proteins revealed a very similar modular architecture to their apicomplexan counterparts (electronic supplementary material, figure S1), supporting structural conservation of the LAPs throughout evolution. The Colponemida were recently shown by phylogenomics to form a sister group to all other alveolates [24], and they thus represent an older lineage in which LAPs are present. We did not find LAP orthologues in other members of the TSAR (Telonemia, Stramenopila, Alveolata and Rhizaria) supergroup [25], or in any other eukaryotic lineages for that matter, indicating that the LAPs first evolved in the last common ancestor of Alveolata and Colponemida. Interestingly, the AlphaFold-predicted *C. vietnamica* LAP3 structure revealed two additional, unspecified domains of some 60 amino acids each on either side of the GAL domain (electronic supplementary material, figure S1), suggesting that the LAP3/5 predecessor was more complex than its current apicomplexan equivalents and providing further evidence for domain losses during its evolution.

2.3. LAP evolution

To shine further light on the evolution of the LAPs, we carried out phylogenetic reconstructions of each of the LAP types separately. Very similar trees were obtained using Bayesian inference and maximum likelihood methods (figure 3; electronic supplementary material, figure S2). LAP1-, LAP2/4- and LAP3/5-type proteins clustered broadly according to established apicomplexan groupings (figure 3). The LAP1 tree revealed a chromerid-specific gene duplication event that has given rise to the two LAP1 paralogues in this lineage (figure 3a). Moreover, an independent Coccidia-specific *lap1* gene duplication (figure 3a; arrow) has given rise to the two LAP1 paralogues in the *Eimeriidae* and *Sarcosystidae* (table 1). In contrast to LAP1, the LAP2/4 phylogeny revealed two distinct radiations that have likely descended from ancestral LAP2/4 paralogues (figure 3b). A more recent gene duplication event specific to the coccidian lineage was revealed for one of these paralogues (figure 3b, black arrow),

Table 1. Minimum LAP repertoire in apicomplexan and chromerid lineages.

lineage		LAP type					
		LAP1	LAP2/4	LAP3	LAP5	LAP6	other[a]
Hematozoa	Haemosporida	1	2	1	1	1	0
	Piroplasmida	1	2	1	1	0	0
Coccidia	*Eimeriidae*	2	3	2	1	0	0
	Sarcocystidae	2	3	2	1	0	0
Cryptosporidia		1	2	0	1	0	0
Gregarinia		1	2	1	0	1	0
Chromerida		2	2	1	2	1	2

[a]LCCL domain-containing protein with five transmembrane helices.

giving rise to the extra (third) LAP2/4 protein in the Coccidia compared with Hematozoa, Cryptosporidia or Gregarinia (table 1).

The LAP3/5 phylogeny showed LAP5 and LAP3 proteins separately clustering in Hematozoa and Coccidia (figure 3c), indicating they have descended from ancestral *lap5* and *lap3* genes, respectively, instead of having been derived from more recent gene duplications followed by loss/gain of a LCCL module. Indeed, this view is supported by the presence of both LAP3 and LAP5-encoding genes in chromerids (table 1). A chromerid-specific gene duplication appears to have given rise to the two LAP5 paralogues (Vbra_11975 and Vbra_11976) in this lineage (figure 3c). Their corresponding genes are tandemly located on the genome, pointing to a recent duplication event. A *lap3* gene duplication event can be identified in the Coccidia (figure 3c, black arrow) giving rise to the two LAP3 paralogues in this lineage. Based on available genomic sequence data, gregarine and cryptosporidian lineages have lost their *lap5* and *lap3* genes, respectively (figure 3c; table 1).

The two distinct ancestral origins of the LAP2/4 proteins are conserved across chromerids and apicomplexan lineages (figure 3b), suggesting that the two paralogues of this protein could have distinctive roles. Indeed, while *Plasmodium* LAP2 and LAP4 have the same loss-of-function phenotype [2,4], probably because they operate as part of the same LAP complex [7,9,22], it was shown in *P. berghei* that their specific interactions are different: LAP2 binds strongly to LAP3, whereas LAP4 interacts strongly with LAP5 [9,22]. This observation combined with the apparent descendants of LAP3 and LAP5 from ancestral *lap3* and *lap5* genes (figure 3c) indicates that these specific pairings could be ancestrally conserved.

Even though we did not find any LAP orthologues in the two largest alveolate lineages: dinoflagellates and ciliates, we identified a number of 'LAP-like' proteins in some dinoflagellates (*Symbiodiniaceae* and *Amoebophryaceae*), possessing different combinations of LAP modules including Ricin, F58C, SRCR and LCCL (electronic supplementary material, figure S3). It is tempting to speculate that these could be the products of rearrangements and modifications of the *lap* genes that were present in the ancestral dinoflagellate. Interestingly, a gene product identified in *Effrenium voratum* (CAJ1387072) contains, besides F58C and SRCR domains, two CPW-WPC domains (electronic supplementary material, figure S3). The CPW-WPC domain (IPR006387) is characterized by six conserved cysteine residues and six well-conserved aromatic sites specific to proteins from apicomplexans and chromerids [26]. In *Plasmodium*, CPW-WPC domain-containing proteins were recently shown to interact with the LAP complex with which they share a subcellular localization [9]. This is to our knowledge the first example of a protein that combines CPW-WPC domains with those found in LAPs, further supporting a functional link between these modules.

3. Discussion

Our studies presented here show that the LAPs are evolutionarily conserved from colponemids through to recent apicomplexans like *Plasmodium*. Colponemids have little in common with apicomplexans in terms of their biology: Apicomplexans are endosymbionts of animals with often parasitic relationships that possess an apical complex and gliding motility. In contrast, colponemids are free-living biflagellate eukaryovores that possess extrusive organelles (trichocysts and toxicysts) for active predation by phagocytosis, and their habitats appear restricted to freshwater environments [24]. The transition from phagocytosis to myzocytosis (tube feeding) is thought to have instigated the evolution of the apical complex (including conoid and secretory organelles) in the Myzozoa (i.e. alveolates excluding Ciliophora) [24,27]. The fact that colponemids do not feed by myzocytosis thus indicates that the LAPs pre-date the apical complex. In addition, plastid organelles appear to be ancestrally absent from colponemids [24,28], implying that the LAPs also pre-date the acquisition of the plastid through secondary endosymbiosis with a red alga [29]. Our results furthermore indicate that the LAPs, and possibly the molecular processes in which they are involved, have been lost in dinoflagellate, perkinsid and ciliate lineages.

The structural conservation of the LAPs despite their complex architectures points to conservation of their core functions. However, it is hard to identify a process that is shared between the biologically diverse colponemids, apicomplexans and chrompodellids, as well as being potentially lost in dinoflagellates, perkinsids and ciliates. In *Plasmodium*, where the LAPs
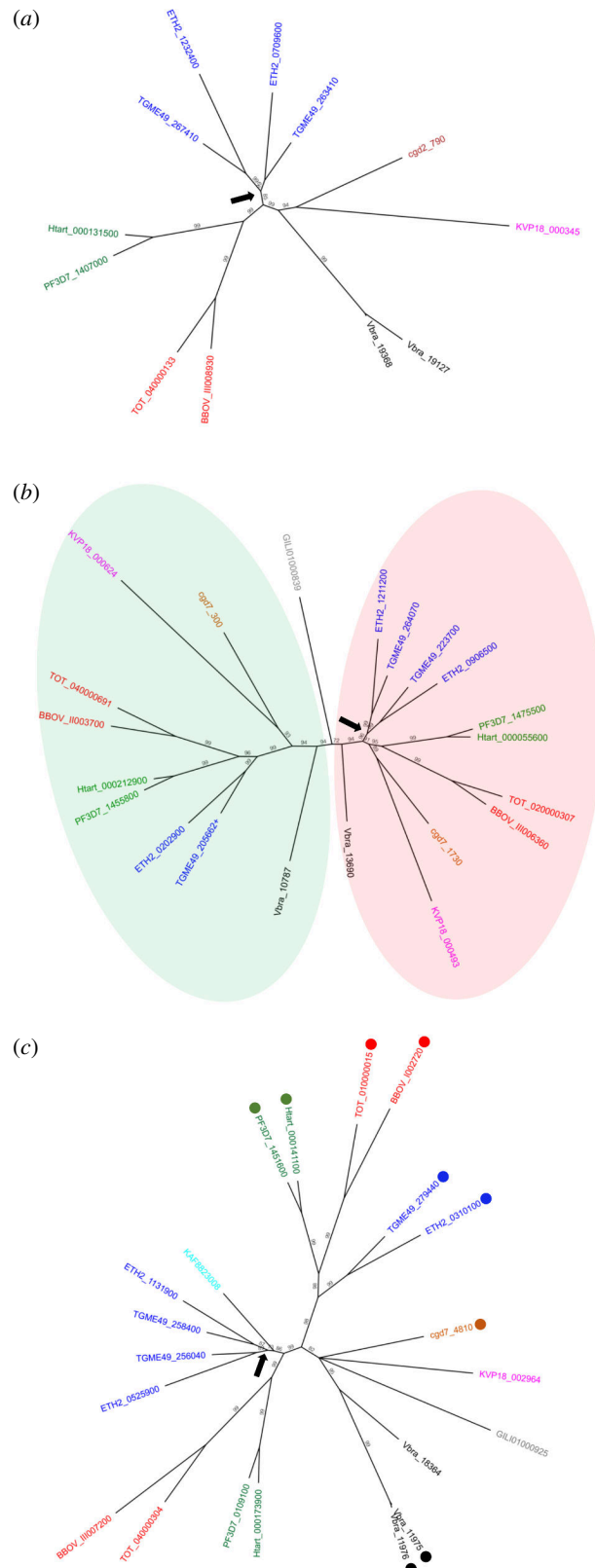
Figure 3. Evolution of LAP types. Bayesian inference-generated phylogenetic trees of LAP1 type (*a*), LAP2/4 type (*b*) and LAP3/5 type (*c*) amino acid sequences from chromerid (black), colponemid (grey), gregarine (pink), cryptosporidium (maroon), haemosporidian (green), piroplasmid (red), coccidian (blue) and nephromycid (cyan) lineages. LAP5 sequences in (*c*) are indicated with dots. Black arrows in (*a–c*) indicate gene duplication events in the coccidian lineage. In (*b*), differenty coloured backgrounds denote two distinct paralogous groups of LAP2/4 proteins. Numbers indicate branch support values.

have been best studied, the proteins have been located in the parasitophorous vacuole (PV) and host cell cytoplasm of *P. falciparum* gametocytes [4,30] and in the crystalloid—a multivesicular putative secretory organelle—in *P. berghei* ookinetes [10–12]. Moreover, genetic ablation of LAP expression abolished formation of crystalloids and blocked sporozoite development (*P. berghei*) or egress (*P. falciparum*) in the oocyst [2–4,11]. In *Cryptosporidium parvum*, LAP2 (Cpa135) was localized in apical secretory organelles of sporozoites as well as in the PV of sporozoite-infected host cells [31]. These combined observations point to a potential role of the LAPs in a vesicle-mediated exocytic pathway. Secretory organelles form part of both canonical and archetypal apical complexes found in Apicomplexa and related alveolate lineages, respectively [32], and the possible homology

between them is of great interest. Determining the subcellular localization, for example, using specific antibodies, of *Plasmodium* LAP orthologues in deeper lineages such as colponemids could shed important new light on this.

4. Material and methods

LAP orthologue searches were conducted using *Plasmodium* LAPs as protein queries in BlastP searches of nonredundant protein databases, or TblastN searches of whole genome shotgun (WGS) and TSA databases via the National Centre for Biotechnology Information (NCBI) (https://ncbi.nlm.nih.gov) and the Eukaryotic Pathogen, Vector and Host Informatics Resource (VEuPathDB) (https://veupathdb.org) [33].

Domain searches were conducted on protein queries with the NCBI CD-Search tool [34] and Conserved Domain Database (CDD) [35], with InterProScan of the InterPro family of databases [36] (https://www.ebi.ac.uk/interpro) and with Simple Modular Architecture Research Tool (SMART) (https://smart.embl.de) [37]. Conserved protein modules were assigned when detected in at least one orthologue.

Predicted three-dimensional structures of *P. falciparum* LAPs were retrieved via the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk) [38]. A comparison of query protein structures against a *P. falciparum* subset of the AlphaFold database was conducted using the AF-DB search option of the DALI server (http://ekhidna2.biocenter.helsinki.fi/dali) [39]. AlphaFold predicted three-dimensional structures of *C. vietnamica* LAP3 and LAP2/4 were generated in AlphaFold Colab (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb).

Phylogenetic analyses were conducted with Geneious Prime software. Phylogenies of combined LAP types were done with MUSCLE multiple alignment and FastTree tree building. Phylogenetic reconstruction of LAPs by type was carried out with Clustal Omega multiple alignment [40] followed by Bayesian inference (MrBayes 3.3.2) [41] or maximum likelihood (PhyML 3.3) [42] methods using Jones–Taylor–Thornton substitution models.

References

1. Claudianos C, Dessens JT, Trueman HE, Arai M, Mendoza J, Butcher GA, Crompton T, Sinden RE. 2002 A malaria scavenger receptor-like protein essential for parasite development. *Mol. Microbiol.* **45**, 1473–1484. (doi:10.1046/j.1365-2958.2002.03118.x)

2. Raine JD, Ecker A, Mendoza J, Tewari R, Stanway RR, Sinden RE. 2007 Female inheritance of malarial lap genes is essential for mosquito transmission. *PLoS Pathog.* **3**, e30. (doi:10.1371/journal.ppat.0030030)

3. Saeed S, Tremp AZ, Dessens JT. 2015 Biogenesis of the crystalloid organelle in *Plasmodium* involves microtubule-dependent vesicle transport and assembly. *Int. J. Parasitol.* **45**, 537–547. (doi:10.1016/j.ijpara.2015.03.002)

4. Pradel G, Hayton K, Aravind L, Iyer LM, Abrahamsen MS, Bonawitz A, Mejia C, Templeton TJ. 2004 A multidomain adhesion protein family expressed in *Plasmodium falciparum* is essential for transmission to the mosquito. *J. Exp. Med.* **199**, 1533–1544. (doi:10.1084/jem.20031274)

5. Trexler M, Bányai L, Patthy L. 2000 The LCCL module. *Eur. J. Biochem.* **267**, 5751–5757. (doi:10.1046/j.1432-1327.2000.01641.x)

6. Dessens JT, Sinden RE, Claudianos C. 2004 LCCL proteins of apicomplexan parasites. *Trends Parasitol.* **20**, 102–108. (doi:10.1016/j.pt.2004.01.002)

7. Simon N, Scholz SM, Moreira CK, Templeton TJ, Kuehn A, Dude MA, Pradel G. 2009 Sexual stage adhesion proteins form multi-protein complexes in the malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* **284**, 14537–14546. (doi:10.1074/jbc.M808472200)

8. Pradel G, Wagner C, Mejia C, Templeton TJ. 2006 *Plasmodium falciparum*: co-dependent expression and co-localization of the PfCCp multi-adhesion domain proteins. *Exp. Parasitol.* **112**, 263–268. (doi:10.1016/j.exppara.2005.11.010)

9. Tremp AZ, Saeed S, Sharma V, Lasonder E, Dessens JT. 2020 *Plasmodium berghei* LAPs form an extended protein complex that facilitates crystalloid targeting and biogenesis. *J. Proteomics* **227**, 103925. (doi:10.1016/j.jprot.2020.103925)

10. Saeed S, Carter V, Tremp AZ, Dessens JT. 2010 *Plasmodium berghei* crystalloids contain multiple LCCL proteins. *Mol. Biochem. Parasitol.* **170**, 49–53. (doi:10.1016/j.molbiopara.2009.11.008)

11. Carter V, Shimizu S, Arai M, Dessens JT. 2008 PbSR is synthesized in macrogametocytes and involved in formation of the malaria crystalloids. *Mol. Microbiol.* **68**, 1560–1569. (doi:10.1111/j.1365-2958.2008.06254.x)

12. Saeed S, Carter V, Tremp AZ, Dessens JT. 2013 Translational repression controls temporal expression of the plasmodium berghei LCCL protein complex. *Mol. Biochem. Parasitol.* **189**, 38–42, (doi:10.1016/j.molbiopara.2013.04.006)

13. Lavazec C, Moreira CK, Mair GR, Waters AP, Janse CJ, Templeton TJ. 2009 Analysis of mutant *Plasmodium berghei* parasites lacking expression of multiple PbCCp genes. *Mol. Biochem. Parasitol.* **163**, 1–7. (doi:10.1016/j.molbiopara.2008.09.002)

14. Dessens JT, Saeed S, Tremp AZ, Carter V. 2011 Malaria crystalloids: specialized structures for parasite transmission? *Trends Parasitol.* **27**, 106–110. (doi:10.1016/j.pt.2010.12.004)

15. Tremp AZ, Saeed S, Dessens JT. 2023 NAD(P) transhydrogenase isoform distribution provides insight into apicomplexan evolution. *Front. Ecol. Evol.* **11**, fevo.2023.1216385. (doi:10.3389/fevo.2023.1216385)

16. Janoušlkovec J, Tikhonenkov DV, Burki F, Howe AT, Kolísko M, Mylnikov AP, Keeling PJ. 2015 Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proc. Natl. Acad. Sci. USA* **112**, 10200–10207. (doi:10.1073/pnas.1423790112)

17. Moore RB *et al.* 2008 A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**, 959–963. (doi:10.1038/nature06635)

18. Janoušlkovec J, Paskerova GG, Miroliubova TS, Mikhailov KV, Birley T, Aleoshin VV, Simdyanov TG. 2019 Apicomplexan-like parasites are polyphyletic and widely but selectively dependent on cryptic plastid organelles. *Elife* **8**, e49662. (doi:10.7554/eLife.49662)

19. Mathur V, Kolísko M, Hehenberger E, Irwin NAT, Leander BS, Kristmundsson Á, Freeman MA, Keeling PJ. 2019 Multiple independent origins of apicomplexan-like parasites. *Curr. Biol.* **29**, 2936–2941.(doi:10.1016/j.cub.2019.07.019)

20. Muñoz-Gómez SA, Durnin K, Eme L, Paight C, Lane CE, Saffo MB, Slamovits CH. 2019 Nephromyces represents a diverse and novel lineage of the apicomplexa that has retained apicoplasts. *Genome Biol. Evol.* **11**, 2727–2740. (doi:10.1093/gbe/evz155)

21. Salomaki ED, Terpis KX, Rueckert S, Kotyk M, Varadínová ZK, Čepička I, Lane CE, Kolisko M. 2021 Gregarine single-cell transcriptomics reveals differential mitochondrial remodeling and adaptation in apicomplexans. *BMC Biol.* **19**, 77. (doi:10.1186/s12915-021-01007-2)

22. Tremp AZ, Sharma V, Carter V, Lasonder E, Dessens JT. 2017 LCCL protein complex formation in plasmodium is critically dependent on LAP1. *Mol. Biochem. Parasitol.* **214**, 87–90, (doi:10.1016/j.molbiopara.2017.04.005)

23. Adl SM *et al.* 2019 Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.* **66**, 4–119. (doi:10.1111/jeu.12691)

24. Tikhonenkov DV, Strassert JFH, Janoušlkovec J, Mylnikov AP, Aleoshin VV, Burki F, Keeling PJ. 2020 Predatory colponemids are the sister group to all other alveolates. *Mol. Phylogenet. Evol.* **149**, 106839, (doi:10.1016/j.ympev.2020.106839)

25. Burki F, Roger AJ, Brown MW, Simpson AGB. 2020 The new tree of eukaryotes. *Trends Ecol. Evol. (Amst.)* **35**, 43–55, (doi:10.1016/j.tree.2019.08.008)

26. Rao PN, Santos JM, Pain A, Templeton TJ, Mair GR. 2016 Translational repression of the cpw-wpc gene family in the malaria parasite plasmodium. *Parasitol. Int.* **65**, 463–471, (doi:10.1016/j.parint.2016.06.007)

27. Okamoto N, Keeling PJ. 2014 A comparative overview of the flagellar apparatus of dinoflagellates, perkinsids, and colpodellids. *Microorganisms* **2**, 73–91. (doi:10.3390/microorganisms2010073)

28. Janoušlkovec J, Tikhonenkov DV, Mikhailov KV, Simdyanov TG, Aleoshin VV, Mylnikov AP, Keeling PJ. 2013 Colponemids represent multiple ancient alveolate lineages. *Curr. Biol.* **23**, 2546–2552, (doi:10.1016/j.cub.2013.10.062)

29. Janouskovec J, Horák A, Oborník M, Lukes J, Keeling PJ. 2010 A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10949–10954. (doi:10.1073/pnas.1003335107)

30. Delrieu I, Waller CC, Mota MM, Grainger M, Langhorne J, Holder AA. 2002 PSLAP, a protein with multiple adhesive motifs, is expressed in *Plasmodium falciparum* gametocytes. *Mol. Biochem. Parasitol.* **121**, 11–20. (doi:10.1016/s0166-6851(02)00016-6)

31. Tosini F, Agnoli A, Mele R, Gomez Morales MA, Pozio E. 2004 A new modular protein of *Cryptosporidium parvum*, with ricin B and LCCL domains, expressed in the sporozoite invasive stage. *Mol. Biochem. Parasitol.* **134**, 137–147. (doi:10.1016/j.molbiopara.2003.11.014)

32. Gubbels MJ, Duraisingh MT. 2012 Evolution of apicomplexan secretory organelles. *Int. J. Parasitol.* **42**, 1071–1081, (doi:10.1016/j.ijpara.2012.09.009)

33. Amos B *et al.* 2022 VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Res.* **50**, D898–D911. (doi:10.1093/nar/gkab929)

34. Marchler-Bauer A, Bryant SH. 2004 CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* **32**, W327–31. (doi:10.1093/nar/gkh454)

35. Wang J *et al.* 2023 The conserved domain database in 2023. *Nucleic Acids Res.* **51**, D384–D388. (doi:10.1093/nar/gkac1096)

36. Paysan-Lafosse T *et al.* 2023 InterPro in 2022. *Nucleic Acids Res.* **51**, D418–D427. (doi:10.1093/nar/gkac993)

37. Letunic I, Khedkar S, Bork P. 2021 SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* **49**, D458–D460. (doi:10.1093/nar/gkaa937)

38. Jumper J *et al.* 2021 Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. (doi:10.1038/s41586-021-03819-2)

39. Holm L. 2020 Using Dali for protein structure comparison. *Methods Mol. Biol.* **2112**, 29–42. (doi:10.1007/978-1-0716-0270-6_3)

40. Sievers F *et al.* 2011 Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539. (doi:10.1038/msb.2011.75)

41. Huelsenbeck JP, Ronquist F. 2001 MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755. (doi:10.1093/bioinformatics/17.8.754)

42. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321. (doi:10.1093/sysbio/syq010)

43. De Hoest-Thompson C, Marugán-Hernandez V, Dessens J. 2024 Data from: *Plasmodium* LCCL domain-containing modular proteins have their origins in the ancestral alveolate. Figshare (doi:10.6084/m9.figshare.c.7227102)