Striated fibre assemblins localise in the feeding groove of the 'typical excavate' *Paratrimastix pyriformis*

Aaron A. Heiss[1,2], Yi-Kai Fang[1], C. Priscila Peña-Diaz[1], Justyna Zítek[1], Ji•í Novák[1], Adam Šmída[1], Marie Zelená[3], Ezra Švagr[1], Vladimír Hampl[1]

[1] Charles University, Faculty of Science, Department of Parasitology, BIOCEV, Vestec 252 50, Czech Republic
[2] Kyungpook National University, Kyungpook Institute of Oceanography, Buk-gu, Daegu 41566, Republic of Korea
[3] Institute of Molecular Genetics of the Czech Academy of Sciences, Laboratory of Cell Motility, Vídeňská 1083, Prague 142 20, Czech Republic

ABSTRACT

Spatial arrangement of the cytoskeleton in the cells of protists has been used for decades for taxonomy and phylogenetic inference at various levels. In contrast, the protein composition of non-microtubular structures is mostly unknown. Exceptions are system I fibres in algae, which are built of striated fiber assemblins (SFA). Interestingly, SFAs are also components of a range of other, dissimilar structures, playing a role in the cortex of ciliates, cells division in apicomplexans, and adhesion of the parasite *Giardia* to the intestine. In a broad bioinformatic survey, we show the presence of three ancestral eukaryotic paralogues of SFA, and note that they are present in all ‚typical excavates' – simple flagellates bearing a ventral feeding groove. In one representative, *Paratrimastix pyriformis*, we localised one of the SFA paralogues using specific antibodies and expansion microscopy. We show that it co-localises specifically with structures attached to the basal body of the posterior flagellum, namely the right microtubular root, composite fibre, and B-fibre. As the morphology of the ‚typical excavates' may be ancestral to eukaryotes, we speculate that the role in the development or function of this feeding apparatus may be ancestral to the protein.

INTRODUCTION

The eukaryotic cytoskeleton is a complex and compound system comprising microtubules, microfilaments, and various fibrous structures (Andersen et al. 1991). It is an important construct; in the context of unicellular eukaryotes, it can be the primary determinant of the morphology of an entire organism. It is sufficiently conserved as to have been recognised, early in its scientific history, as taxonomically significant at deep phylogenetic levels (e.g., Moestrup 1982), and it is still regarded both as phylogenetically predictive and as a topic for investigation based upon molecular phylogeny (e.g., Simpson 2003, Yubuki & Leander 2013, Heiss et al. 2018). In most protists, the core of the cytoskeleton is the flagellar apparatus, comprising the basal bodies and the microtubular and fibrous structures directly associated with them. Some of these structures, termed flagellar roots, can extend through the full length of the cell, often providing an armature stabilising the cell's shape (Andersen et al. 1991).

The bulk of attention in this regard has been focused on the microtubular components of the cytoskeleton (Moestrup 2000, Yubuki & Leander 2013). From an evolutionary point of view, microtubular roots are suspected to be homologous in different lineages if they attach to similar points on the basal bodies, appear in a similar arrangement both to the basal bodies and to each other,

and travel along similar paths through the cell (Yubuki & Leander 2013). However, in some cases at least, associations with non-microtubular structures are also regarded as significant and diagnostic (Simpson 2003, Heiss et al. 2018). It is easily inferrable that if microtubular structures are regarded as putatively homologous based on similar attachments, positioning, and paths through the cell, the addition of comparable fibrous structures raises the probability of homology considerably further (Simpson 2003). Some of the most taxonomically widespread of these components are termed System I and System II striated fibres (Melkonian 1980; Andersen et al. 1991).

System I fibres are of particular interest in that they frequently associate with microtubules; they have also been termed 'striated microtubule-associated fibres' (SMAFs: Lechtreck & Melkonian 1991, Lechtreck 1993). They were first identified under the electron microscope in green algae (history summarised in Moestrup 1982), and later characterised biochemically; the monomers out of which they are constructed have been termed striated fibre assemblins (SFAs) (Lechtreck & Melkonian 1991). Seemingly unrelated structures in a distantly related organism, the fibrous components of the microribbons in the adhesive disc of the vertebrate gut parasite *Giardia*, were likewise first described electron-microscopically (Holberton & Ward 1980). Similarities at the primary and secondary structure level between SFAs (Lechtreck & Melkonian 1991) and •-giardin, one of the monomers comprising giardial microribbons (Holberton et al. 1988), led to the two proteins being recognised as homologous (Lechtreck & Melkonian 1991, Weber et al. 1993). This assessment emphasised the presence of coiled-coil domains, which were thought to assist in the polymerisation of monomers. Given the vast evolutionary distance between green algae and *Giardia*, this suggested that SFAs might be found in other organisms as well. Meanwhile, the known diversity of SFAs expanded in a different direction as further study revealed that •-giardin was not the only assemblin homologue in *Giardia*, with the discovery and localisation first of •-giardin and then of the 'Striated-fibre Assemblin-Like Protein' SALP-1 (Palm et al. 2005).

Recent work has greatly extended the number and diversity of putative SFA genes, in terms both of taxonomic diversity and of the sizes and structures of this gene family within genomes. The first significant taxonomic expansion of known SFA-bearing lineages was a bioinformatic survey and subsequent immunological study that identified SFA homologues in a number of apicomplexan species, some of which had multiple paralogues (Lechtreck 2003), and formed a fibrous structure acting during cell division (Francia et al. 2012). A similar combination of approaches later documented SFAs in ciliates, a dinoflagellate, diatoms, and oomycetes (Harper et al. 2009). Very recently, ciliates were found to harbour an immense diversity of SFA genes. One study found, characterised, and localised some 30 paralogues in the model ciliate *Paramecium*, which were classified into five 'structural groups' collectively containing thirteen 'paralogue groups', each of which could be identified ultrastructurally, immunologically, developmentally, and phylogenetically (Nabi et al. 2019). Another study did the same for *Tetrahymena*, finding ten paralogues in that organism, and incidentally expanding the number in *Paramecium* to 72 (Soh et al. 2020). While the latter study focussed on ciliates, it also found sequences putatively identifiable as SFA genes in almost all other major lineages, with the notable exceptions of the 'supergroups' Amoebozoa and Obazoa. In addition, it found three clades of paralogues, separated at the deepest phylogenetic level, suggesting that the divergence of one SFA gene into three occurred in the common ancestor of all SFA-bearing lineages.

In spite of this, the localisation of SFAs in the cell is known for only a few lineages: green algae, *Giardia*, apicomplexans, ciliates, and oomycetes (Harper et al. 2009; Francia et al. 2012; Nabi et al. 2019; Soh et al. 2020). In this study, we add another organism to this list, the free-living flagellate *Paratrimastix pyriformis*. This organism is a member of Preaxostyla (Simpson 2003; Zhang et al. 2015), which is one of the lineages of Metamonada, a broad grouping of anaerobic and microaerophilic protists. Metamonada also includes *Giardia*, although it is a very distant relative of *Paratrimastix*. Unlike *Giardia*, *P. pyriformis* is a representative of so-called 'typical excavates', unicellular organisms bearing a ventral feeding groove supported by three microtubular roots attached to the posterior

basal body. Based on hypothetical rootings of the eukaryotic tree (Derelle et al. 2015; Adl et al. 2019; Cerón-Romero 2022) and proposed homologies between the microtubular roots of eukaryotes (Yubuki and Leander 2013), this cell plan may possibly be ancestral to all eukaryotes, i.e., present in last eukaryotic common ancestor (LECA). Should SFAs be associated with excavate groove structures, this could suggest points of homology between organisms with and those without feeding grooves. In particular, we examine SFAs in *P. pyriformis* by using antibodies raised against two of its SFA paralogues to localise them within the cytoskeleton. Our observations employ expansion microscopy, a technique that achieves greater light-microscopic resolution not by improving the optics but by enlarging the specimen (Gorilak et al. 2021).

MATERIALS AND METHODS

Bioinformatics

Data curation. A 'core' dataset was developed based on that of Lechtreck et al. (2003), which contained only well-attested SFA sequences, specifically including those sequences experimentally confirmed to be SFAs (*Spermatozopsis similis* SFA, GenBank AAB26641.1; *Giardia intestinalis* •-giardin, GenBank P15518.2; *Toxoplasma gondii* SFA2, ToxoBase TGG_7332; *Phytophthora ramorum*, JGI-81943). To provide taxonomic breadth, the core dataset was expanded to include additional sequences, using data from Harper et al. (2009) and carefully selected sequences from GenBank (NCBI: https://www.ncbi.nlm.nih.gov/). To assure the exclusion of spurious data, each core sequence was compared against all remaining core sequences using BLASTP (Altschul et al. 1991). All comparisons took into account alignment length, percent identity, and E-value. Further amino-acid sequences were obtained from a combination of (a) additional published datasets (Nabi et al. 2019, Soh et al. 2020), (b) the public databases GenBank and EukProt (Richter et al. 2022), and (c) genome and transcriptome projects for the oxymonads *Monocercomonoides exilis* (Treitli et al. 2021), *Blattamonas nauphoetae* (Novák et al. 2023), and *Streblomastix strix* (Treitli et al. 2019); the non-oxymoand preaxostylans *Paratrimastix pyriformis* (Novák et al. 2023) and *Trimastix marina* (Zhang et al. 2015); and the fornicates *Retortamonas caviae* and *R. dobelli* (Füssy et al. 2021), *Ergobibamus cyprinoides*, *Aduncisulcus paluster*, *Dysnectes brevis*, and *Chilomastix cuspidata* (Leger et al. 2017). All sequences were confirmed by reciprocal BLAST. When possible, partial sequences from one database were replaced with corresponding longer sequences from another, and only data from a single source was used for each sequence (i.e., sequences with nonoverlapping regions at opposite ends were not assembled into longer contigs). Sequences over ~450 bases long were manually trimmed to within ~150 residues of 'core' data, with methionine as the N-terminal residue when possible.

Sequence alignment. Initially, amino acid sequences were assembled into a single matrix and aligned using MUSCLE (Edgar 2004), followed by manual alignment and masking. Preliminary trees were generated from masked alignments using RAxML v8.2.12 (Stamatakis 2014), using the LG substitution model and the CAT model of site variation, with support evaluated using 200–500 rapid bootstraps. Two primary clades of paralogues (Group 1 and Group 2) were identified early in this process (see Results); these paralogues were divided into separate matrices and analysed independently. Sequences that were not clearly recognisable in the alignment as belonging to a specific paralogue were duplicated with one duplicate each aligned to each paralogue, and assessed using rapid-bootstrap phylogenies as described above. Long-branching sequences (defined as having branch lengths • 1.5 substitutions per site and being • 2× the length of the next-longest sequence), near-identical sequences, and members of overrepresented clades (predominantly ciliates) were pruned from all matrices. The final dataset contained 302 sequences and 262 sites, with 158 and 144 sequences in the Group 1 and Group 2 matrices, and with 247 and 242 sites, respectively.

Phylogeny. Trees were generated from all three data sets using RAxML, under the LG model with a gamma distribution of site rate variation and empirically determined base frequencies, with

the final topology in each case being the best of 100 parallel runs, and with support values taken from 1,000 bootstrap replicates using the same model. Analyses were repeated using IQ-TREE (Nguyen et al. 2015), with 100 parallel runs under the LG+F+G model to provide guide trees, and the LG+PMSF+C20+G model for final analyses, which included 200 bootstrap replicates. Each of the three final datasets was also subjected to Bayesian analysis using MrBayes v3.2.7 (Ronquist et al. 2012) to implement four parallel runs, each with eight chains, a heating parameter of 0.05, two swaps every two generations, and a 10% burnin. These were repeated using the same parameters (with exceptions noted below) with the RAxML tree as a starting tree, with five permutations of the tree applied to each chain, and heating parameters of 0.10 (Group 1) or 0.05 (Group 2 clade and combined data); the best tree from each of these analyses was used for a third Bayesian analysis, with three (Group 1-clade) or two (Group 2-clade and combined data) permutations per tree and heating parameters of 0.05 (Group 1-clade) or 0.10 (Group 1-clade and combined data). Since no runs converged after 107 generations, the run with the highest terminal likelihood was used for each analysis. As with RAxML analyses, Bayesian analyses used the LG substitution model with a gamma distribution of site variation and empirically determined base frequencies.

Structural analysis. Protein-domain prediction was carried out on all sequences using the Simple Modular Architecture Research Tool ('SMART') online platform (Letunic and Bork 2018). Results from these analyses were processed through an in-house pipeline that (a) extracted and compiled predictions of coiled-coil and low-complexity regions as well as Pfam predictions for 'SF-assemblin' and 'Baculo_PEP_C' (the second-most-common Pfam prediction), (b) aligned, trimmed and masked per-site prediction results according to the Group 1 or Group 2 alignments, as appropriate, and (c), for easier visualisation, mapped all predictions onto a single line (priority being given, in order, to coiled-coil domains, low-complexity domains, 'SF-assemblin' predictions, and 'Baculo_PEP_C' predictions) and collapsed lines by a factor of 7 (the number of residues required for one complete coil in a coiled-coil structure). Other reduction factors were also used. The factor-7 visualisations were aligned with their corresponding taxa in phylogenetic trees for visual comparison.

Antibody Preparation

Antiserum generation. The reconstruction proteins for antiserum production were cloned and purified as described in a preprint manuscript (Peña-Diaz et al., 2023). The genes *PaPyr7259* (P7) and *PaPyr8006* (P8) were amplified from *Paratrimastix pyriformis* cDNA using specific primers (PaPyr7259-F ATCGCATATGAGCACCCCGCCGCCCAGCTCTCCTTCCCC; PaPyr7259-RP ATATCTCGAGATGGGTGACCAGGTGCAGGCCCTCCTGCAG; PaPyr8006-F ATCGCATATGCAAGCCGAGACTCCGATTCCCACTC; PaPyr8006-R GCTAGCGGCCGCATGGGTGATGACGTGGAGGCCGTCCTG). Amplicons were cloned into the bacterial expression vector pET30a using the primers' C-terminal His-tags. Proteins were expressed in *E. coli* Rosseta2 DE3 by autoinduction for no more than 18 hours. Bacterial lysates were prepared by resuspending the cell pellets in HNGB (50 mM HEPES-KOH, pH 8.0, 300 mM NaCl, 8 M urea, 10% (v/v) glycerol, 2 mM beta-mercaptoethanol) in the presence of a protease inhibitor cocktail (Roche, Basel, Switzerland). These mixtures were incubated on ice for 30 min, then loaded into a 35-ml standard pressure cell and broken on a G-M™ high pressure cell press homogenizer at 1,000 psi on a high ratio in three rounds. Whole-cell lysates were ultracentrifuged at 100,000 × g for 1 hour at 4° C in a SW40Ti rotor on a Beckman ultracentrifuge. Supernatants (clear lysate) were loaded in 20 ml columns packed with 5 ml of HisPur™ Ni-NTA resin (Thermo-Fischer, Waltham, Massachusetts, USA) equilibrated with HNGB. Columns were further washed with the same buffer, and proteins were eluted using 250 mM imidazole buffered in HNGB. Sample eluates were verified by SDS-PAGE and western blotting. The extracted proteins were then used to raise polyclonal antibodies in rats as previously described (Mach et al. 2013).

Affinity purification. Cyanogen-bromide-activated Sepharose® 4B (C9142, Sigma-Aldrich, St. Louis, USA) was used for affinity purification of anti-P7 and -P8 antisera. Beads were re-swelled

following the manufacturer's instructions in ice cold 1 mM HCl and washed with coupling buffer (0.1 M carbonate buffer, 0.5 M NaCl, pH 8.3). Purified *P. pyriformis* proteins were dialysed against coupling buffer and incubated with the beads overnight at 4° C. Beads were collected and washed with coupling buffer, and subsequently incubated in 0.1 M Tris-HCl pH 8.0 for 2 hours at room temperature. Beads were then washed for three cycles with Tris buffer (0.1 M Tris-HCl, 0.5 M NaCl, adjust pH to 8.0) followed by acetate buffer (0.1 M sodium acetate and 0.5 M NaCl, pH 4.0) at 4° C. The antisera were then incubated with beads overnight at 4° C, and the bound antibodies were eluted by 0.2 M glycine, pH 2.8.

Co-immunoprecipitation

Prior to immunoprecipitation, 2 L of well-grown *P. pyriformis* culture was filtered as previously described (Zítek et al. 2022). Briefly, a *P. pyriformis* culture on the exponential phase of growth was filtered, first through filter paper and subsequently a 3 µm polycarbonate filter membrane, to remove co-cultured bacteria. Filters were then washed twice with 10% (v / v) LB medium, and collected by centrifugation at 1,000 × *g* for 10 min at 4° C.

Immunoprecipitation was carried out using Dynabeads™ Protein G Immunoprecipitation Kit (Thermo-Fisher, Waltham, Massachusetts, USA) according to manufacturer's instructions. Briefly, *P. pyriformis* cell lysate was prepared using lysis buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM MgCl, and 0.1% (v / v) Triton X-100) supplemented with cOmplete EDTA-free protease inhibitors (Roche, Basel, Switzerland). To the pellet of 107 of *P. pyriformis* cells, 1 mL of lysis buffer was added and incubated with rotation for 1h at 4 °C. Cleared lysate was centrifuged for 5 min 1,000 × g at 4 °C, the supernatant was transferred to a new 1.5 mL tube and kept on ice. The magnetic beads-antibody complex was incubated with *P. pyriformis* cell lysate for 30 min with rotation at room temperature. For further western blot analysis the pulled-down proteins were eluted under denaturation conditions. The whole experiment was repeated using RIPA lysis buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% (v / v) Triton X-100, 0.1% (v / v) SDS, and 0.5% (v / v) sodium deoxycholate) supplemented with cOmplete EDTA-free protease inhibitor.

Expansion Microscopy

Cells were prepared according to the methods developed by Gorilak et al. (2021), with minor modifications.

Fixation. Cells were cultured and filtered as described above. Roughly 5×10^5 cells per specimen were resuspended in sucrose buffer (25 mM sucrose; 10 mM HEPES-KOH, pH 7.5; 2 mM $MgSO_4$) containing 4% formaldehyde and 4% acrylamide. Fixed cells were transferred directly to poly-L-lysine-coated coverslips, incubated overnight at 4° C, and then washed once with phosphate-buffered saline solution (PBS).

Gelation. Immediately prior to gelation, 0.5% N, N, N', N'-tetramethylenediamine and ammonium persulphate were mixed with a monomer solution containing 19% sodium acrylate, 10% acrylamide, and 0.1% N, N'-methylenebisacrylamide in PBS. This was then dispensed in 50-µl aliquots onto glass coverslips in a wet chamber, on which coverslips containing cells were placed, cell side down. These were incubated on ice for 5 min and then at 37° C for 30 min.

Denaturation. Specimens were incubated at room temperature for a few minutes in denaturation buffer (50 mM Tris, 200 mM NaCl, 200 mM sodium dodecyl sulphate, adjusted to pH 9.0) until coverslips detached from gels. Gels were then transferred to 1.5-µl centrifuge tubes with fresh denaturation buffer and incubated for 1 hr at 95° C. After this, specimens were washed at room temperature with PBS thrice for 20 min each.

Antibody staining. Specimens were blocked with 2% bovine serum albumin in PBS for • 1 hr at 4° C. For primary antibody staining, rat anti-P7 and anti-P8 sera, and guinea pig anti-tubulin (AA345, ABCD antibodies, Geneva, Switzerland), were diluted in 250 µl of blocking buffer in 1:50 and

1:100 titrations, respectively, and incubated overnight at 4°C. Gels were then washed thrice with PBS for 20 min at 4°C. Secondary antibodies (A11007, Alexa Fluor™ 594, goat anti-rat IgG; A11073, Alexa Fluor™ 488, goat anti-guinea pig IgG; Invitrogen, Waltham, Massachusetts, USA) were diluted 1:250 in blocking buffer, in which specimens were incubated overnight in the dark at 4° C, and afterward again washed thrice with PBS for 20 min at 4° C. Prepared gels were expanded by incubation in three changes of ultrapure water for 20 min at 4°C.

Imaging and image processing. Expanded gels were transferred to poly-L-lysine-coated glass-bottomed dishes. Specimens were imaged using a Nikon CSU-W1 spinning disk field scanning confocal microscope (Nikon Instruments Inc., New York, USA) with a 63× water immersion objective. Voxel sizes were calculated to be 56 × 56 × 150 nm. Images were processed using Image-J as implemented in the FIJI package (Schindelin et al., 2012).

RESULTS

Phylogenetic Analysis

Full analysis. To map the diversity of SF-assemblins across eukaryotes, we compiled a dataset of 302 SFA homologues from 128 taxa and subjected it to phylogenetic analysis. Two clades were immediately apparent, separated by one of the longest branches in the tree, and with near-maximal statistical support (Fig. S1). Each clade had its own characteristic indel, and as such, clade membership could be predicted for most sequences from the alignment alone.

One of these clades, corresponding to 'Group 1' in Soh et al. (2020), contained (a) all members of the green algal lineage, and therefore all green-algal SFAs studied; (b) all of the three previously studied SFA homologues in *Giardia*, namely •-giardin, •-giardin, and SALP-1; and (c) all of the previously identified apicomplexan and stramenopile SFA homologues. The other clade (corresponding to Group 2 in Soh et al. 2020) contained almost all of the ciliate sequences and about half of the fornicate sequences; all other 'kingdom-scale' lineages were far less well-represented. Nevertheless, all lineages except green algae were represented in this clade, as was the only centrohelid sequence identified in our search and curation process. As such, this centrohelid sequence represented the only lineage present in Group 2 that was not also in Group 1. Notably absent from either group were any of the lineages comprising the 'megagroup' Amorphea (Adl et al. 2012). To improve the phylogenetic signal within groups, we analysed each separately.

Group 1 analysis (Figure 1). With only three exceptions, all of the lineages present in Group 1 were monophyletic and received significant statistical support (Fig. 1). Two of those exceptions were the cryptists, of which *Palpitomonas* branched separately from the remainder (the latter being recovered as a well-supported clade on a very long branch), and the ciliates, which branched separately from not only the other alveolate lineages but the other members of the SAR clade as well. The third exception was Metamonada, for which Fornicata and Preaxostyla branched separately in all analyses, although they were both well supported individually (94% and 88% bootstrap support, respectively). The backbone of the Group 1 tree was unsupported.

Four lineages showed internal paralogy. The most obvious of these was the haptophytes, for which two well-supported paralogues were present, each of which contained most of the same taxa as the other, and both of which were themselves joined with significant support. The internal topologies of the two paralogues were largely consonant with one another, and each received generally strong statistical support.

Likewise, the non-ciliate alveolates were present in two well-supported clades, one with slightly fewer taxa and longer branches. In this case, though, internal topologies were not consistent, and received significant support only for closely related taxa. The larger of the two clades included all of the apicomplexan sequences previously identified as 'SFA1' or 'SFA3' homologues (Lechtreck 2003), while the other clade contained all sequences previously identified as 'SFA2'.

The fornicates include the diplomonad *Giardia intestinalis*, which is the only member of the group to have been the subject of any previous SFA-related study. Although *Giardia* was also present in Group 2 (see below), all three of the SFA paralogues previously studied in *Giardia* (•-giardin, •-giardin, and SALP-1) were present in Group 1. The two giardin paralogues also included other diplomonads (*G. muris*, *Spironucleus*, and *Trepomonas*), and each of the three also included a clade comprising the related *Retortamonas dobelli* and *R. caviae*, as well as *Dysnectes brevis*. All clades were well supported, except for the inclusion of *Dysnectes* in the •-giardin clade. Only one sequence per taxon was present for all other fornicates except *Aduncisulcus*, which had three that branched together, two with significant support. The backbone topology within the fornicate clade received no statistical support.

Finally, the preaxostylans showed a marked separation between the free-living basal members and the derived and generally symbiotic oxymonads. The basal group contained one sequence from *Trimastix marina* and two from *Paratrimastix pyriformis*; these three, and the oxymonad clade as a whole, had no supported branching order. The oxymonads grouped together on a long branch with maximal support. Two of the sequences, both from *Blattamonas*, had a statistically supported sister relationship. As with the basal group, though, the remaining sequences, which included one from *Streblomastix* and two from *Monocercomonoides*, had no supported branching order, either with one another or with the *Blattamonas* clade.

Group 2 analysis (Figure 2). Like in the Group 1 tree, the backbone of the Group 2 tree was unsupported (Fig. 2). Unlike in the Group 1 tree, support for individual 'kingdom-scale' lineages varied, with some strongly supported, some well below 50%, and several polyphyletic. Soh et al. (2020) noted two subclades (which they termed 'Group 2a' and 'Group 2b') which also were recovered in our Group 2 phylogeny (Fig. 2). Support values for major lineages within Group 2b were generally higher than those in clade Group 2a. Multiple preliminary analyses of Group 2a and b independently of one another showed no significant improvement in statistical support, either between or within groups (data not shown). The two subgroups were similar in size, with 76 and 68 taxa, respectively. Taxon overlap between them was substantial but not total; in particular, Group 2b contained a subset of the lineages found in Group 2a. The composition of each clade was dominated by fornicates and ciliates. Paralogy within the ciliates was extensive, with 72 paralogues from *Paramecium tetraurelia* identified by Soh et al. (2020), pruned down to 25 in our dataset; other species typically had 8–10 paralogues each. The majority of the sequences from fornicates were to be found in Group 2a, with *G. intestinalis* appearing in both groups. Outside of fornicates, other lineages also had paralogues within Group 2, although in lower numbers overall. The most prevalent of these were non-ciliate alveolates, with 12 sequences, discobids, with eight, and preaxostylans, with seven; of these, non-ciliate alveolates and preaxostylans were roughly equally represented Group 2a and Group 2b.

Structural Analysis

Secondary-structure prediction identified only two consistently appearing domain types: coiled-coil motifs and low-complexity regions. The former are thought to be characteristic of assemblin proteins (Weber et al. 1993), but in spite of this, consistent patterns of predicted coiled-coil domains were not present on the large scale, either of all sequences in our dataset or in any of the three major clades within it (Fig. 1, 2). However, patterns did appear for specific lineages, most strongly in the cryptists and the shorter-branching of the two homologues in the haptophytes (Fig. 1).

With only a few exceptions, all of them minor variations, cryptophytes had a single predicted coiled-coil domain starting at slightly more than one-fourth of the length of the predicted protein and running to about halfway through. The exceptions within the cryptophytes had longer predicted domains, extending only toward the N-terminus; the C-terminal ends of these domains were remarkably consistent. The non-cryptophyte cryptists provided another variation, with a second, smaller coiled-coil domain near the C-terminal end.

The most common pattern in the shorter-branching haptophyte clade consisted of two short coiled-coil domains at the N-terminal end of the protein, spaced apart by about the length of one of the domains. The two principal exceptions to this were to have a single long coiled-coil domain that spanned the entire length of the two seen in most haptophytes, or to have a third, short domain near the C-terminus (although one of these exceptions also lacked the first of the 'standard' two domains as well). This 'standard' haptophyte pattern was more evident in prymnesiophytes than in pavlovophytes. No pattern at all was apparent in the longer-branching homologue, either for haptophytes as a whole or for either of the group's principal subclades.

Microscopy

Notably, both main groups of SFA are present in a broad diversity of 'typical excavates', i.e., flagellates with the excavate ventral feeding groove (Simpson 2003). We investigated the localisation of this protein within the 'typical excavate' cytoskeleton, which may be the ancestral form for most or all eukaryotic lineages (Yubuki and Leander 2013). We chose *Paratrimastix pyriformis*, which is easily cultured and relatively large for a 'typical excavate', as our model organism, and raised antibodies against four of its SFA paralogues: PaPyr1826, PaPyr3315, PaPyr7259 and PaPyr8006. The specificity of the antibodies was tested on western blots and by co-immunoprecipitation (co-IP). The former two antibodies were not specific and were not used further. Antibodies against the two Group 1 paralogues PaPyr7259 (hereafter referred to as 'P7') and PaPyr8006 (hereafter referred to as 'P8') recognised bands of expected sizes in Western blots. Co-IP (Table S1-2) clearly showed that both antibodies precipitated the target protein but also the other paralogue. This was not affected by the presence of RIPA buffer preventing the formation of protein complexes, which indicates that the precipitation is caused by a cross-reaction of the P7 antibody to P8 protein and vice versa (Table S3). Co-IP experiments detected a number of common contaminants, such as ribosomal proteins, but consistently also detected alpha-tubulin and, in the case of P8, also beta-tubulin.

These proteins were localised by immunofluorescence on expanded cells of *P. pyriformis*. The cells were co-stained with anti-tubulin antibody to visualise the complete tubulin cytoskeleton of the cell (Fig. 3, 4). The anti-tubulin antibody heavily stained flagellar axonemes, and flagella and basal bodies (BBs) were identifiable based on their positions relative both to one another and to previously documented (O'Kelly et al. 1999) microtubular (MT) structures. Our expansion microscopy preparations were sufficient to distinguish individual MTs' paths within cells, albeit not always to quantify numbers of MTs when they were close together.

However, cells were preserved with varying degrees of distortion. In general, the better-preserved cells were less well-stained, and vice-versa, thereby constraining us to interpret structures from moderately distorted cells. Flagella were frequently broken off, the point of breakage varying across the entire length of the flagellum but generally proximal to the cell (e.g., Fig. 3). Within cells, MTs could almost always be identified with previously described structures, but many such MTs deviated from published accounts in their distal orientations and connections, either individually or in groups. In addition, some MTs appeared exceptionally long, curling around and constrained only by the outline of the cell; there were also very short fragments of MTs visible in some cells. Our description is based on the positions, orientations, and connections most often seen in the less-disrupted cells, with reference to previously published descriptions. Also apparent in our preparations was a substantial amount of particulate signal, most easily attributable to non-specific labelling or monomeric proteins. The random distribution of these signals occasionally gave the erroneous impression of colocalisation, which impression was dispelled through the comparison of multiple cells in varying orientations.

Near the BBs, the ventrally projecting anterior flagellum was associated with an anterior microtubular root (AR), which arced over the anterior side of the BB in a broad leftward and posterior curve (Fig. 4A, B). It gave rise to a conspicuous dorsal fan comprising a radiating array of ~20 microtubules (MTs) (Fig. 4A, B). These were usually evenly spaced, but there were some cases in

which one or two gaps interrupted otherwise regular spacing; gaps were not found in consistent locations between cells. The dorsal fan followed the outline of the anterior part of the cell, initially projecting dorsally and slightly leftward (Fig. 3), then curving posteriorly and leftward (Fig. 3, 4A), and finally posteriorly and rightward (Fig. 4B–D), abutting and terminating at the outer edge of the posterior right root (see below), about halfway down the cell's anteroposterior length (Fig. 4E). The fan thus presented as a broadly ribbonlike helix, curving ~¾ around the cell from its origin.

The left and right posterior roots ('LR' and 'RR', respectively) each initially comprised tightly spaced parallel MTs (Fig. 4B). A single intermediate MT, the singlet root (SR), originated near the RR, angled to the left, and joined the LR (Fig. 4B, C). This junction was close to the SR's point of origin, such that in some cells the SR was difficult to perceive as an independent entity at all (Fig. 3). Posterior to this point, the LR and SR ran along the left ventral margin of the cell, being joined about one-third of the length of the cell by the AR (Fig. 3, 4B, C).

Not far from its point of origin, the RR split into two distinct but close subunits, the inner and outer RRs (iRR and oRR, respectively) (Fig. 4C). A second split, within the iRR, was evident closely posterior to this. While the MTs of the oRR stayed closely together, those of both parts of the iRR were spaced one to two MT diameters apart, not much less than the space comprising each split. This often made both splits difficult to discern from the tubulin stain alone. The middle part of the iRR contained about four MTs and remained a discernable unit to the posterior end of the cell. Individual MTs frayed off from the inner (medial) part at more-or-less regular intervals, forming a widely spaced set of inner RR MTs (iRMTs) spanning the distance diagonally between the iRR, on the one hand, and the LR and its associated structures, and on the other (Fig. 3). The first MT to split off from the inner part of the iRR, corresponding to MT 'y' in Simpson et al. (2000), travelled leftward to join the LR, SR, and AR (Fig. 3, 4D); posterior to that point, these MTs appeared as a single entity (hereafter, the 'compound root', 'CR') (Fig. 3, 4E, F). Other MTs in this subroot of the iRR either abutted and terminated at the CR or, more posteriorly, passed close by the CR to terminate independently beyond it. The numbers of each of these varied from cell to cell, with some having no abutting MTs at all and others having several; but all cells had 9–10 iRMTs in total (including 'y'), and all except 'y' terminated at or close by the CR. At the posterior end of the cell, the ~4 MTs of the 'outer' or 'middle' part of the iRR remained spaced roughly one MT apart from each other, this root eventually curving leftward and then anteriorly for a small distance. The oRR, in contrast, remained either in parallel to the CR or at a close angle to it (Fig. 3, 4).

Consistent with our co-IP results, P7 and P8 ABs bound to the same structures (the signal from P8 is shown in Figs. 3 and 4). One of the most conspicuous signals was a colocalisation with the MTs of the iRR and iRMTs along their whole length (Fig. 3). Thus, which part of the RR a MT was eventually part of could be determined not far from its point of origin. No distinction could be observed between the 'middle' and 'inner' parts of the iRR. No other MTs were tagged along their entire length. However, while the LR and SR were untagged along their anterior length, they acquired an SFA signal immediately anterior to their being joined by 'y'. Posterior to that point, the SFA signal appeared only on the LR+SR MTs, and, when MT identity could be determined, not on 'y' (Fig. 3; Fig. 4G). While all MTs in the iRR bore a very strong P7/P8 signal, when balanced such that the tubulin and SFA signals were equal, the tubulin signal was much stronger in 'y' than in any of the other MTs in the iRR (Fig. 3).

Non-root structures were also labelled by our SFA ABs. One very strong signal, tagged equally strongly by both tubulin and SFA ABs, was to an anteroposteriorly oriented root-like structure located in the proximity of the cell's epipodium (which was not itself visible in our preparation), which we termed the epipodial complex ('EC'). This was 2.5–3 µm long, and often sufficiently close to the LR as to appear part of it. In many cells, however, the EP diverged ventrally from the LR at a narrow angle, most often projecting slightly ventrally from the cell. One other, much more subtle, signal was evident at the anterior of the cell, a ~300–500-nm-wide structure that appeared separate from any MTs (Fig. 4B, C). This signal was possible but difficult to identify in cells

with their anteroposterior axes close to the plane of optical section, but could be made out distinctly in cells oriented perpendicularly to the plane of section. It ran from the junction of the RR and the posterior basal body ('PBB') diagonally posteriorly and rightward, to terminate about a third of the way through the cell at the oRR. This corresponds positionally to the B-fibre as described by O'Kelly et al. (1999).

Two non-MT SFA-tagged structures appeared at the posterior of the cell, one on either face of the oRR. One of these appeared close but anterior to the closing of the groove, a little distance away from the dorsal side of the oRR (i.e., the side of the root facing the cytoplasm) (Fig. 4E). The other was on the opposite (outside or plasma membrane) side of the oRR, closer to it, and only at the posterior end of the cell, following the oRR to its termination (Fig. 4G). These structures are in the same positions, and of comparable shapes, to the composite fibre and groove-terminating structure (respectively) described in *P. eleionoma* by Simpson et al. (2000).

DISCUSSION

Phylogenetic Diversity of SFAs

Our work does not add drastically to the spectrum of lineages shown in previous studies to have SFA homologues (most recently Nabi et al. 2019, Soh et al. 2020), but it does fill in some needed detail. We confirm the conclusion of Soh et al. 2020 that eukaryotic SFAs form three major clades (Group 1, Group 2a, and Group 2b), and based on their distribution, we infer that all three were present in their common ancestor (Figure 5). We note that, at the deepest scale of eukaryotic diversity, there is (yet) no evidence of SFAs in any members of Amorphea; all major lineages of Diaphoretickes have at least one paralogue, but the presence of all three is not consistently spread across lineages. However, all three clades of SFAs appear in Discoba, Fornicata, Preaxostyla, and Malawimonadida (collectively, the excavate lineages), as well as Ancyromonadida, each of which has generally been supposed to be closer to the eukaryotic root than Amorphea and Diaphoretickes (Derelle et al. 2015; Brown et al. 2018; Adl et al. 2019).

There is an intriguing exception to this pattern: we found no SFA homologues in Hemimastigophora, which may be either deeply branching within, or sister to, Diaphoretickes (Lax et al. 2019). Even more intriguingly, hemimastigophorans have quite ciliate-like architectures (Foissner et al. 1988), while in contrast, ciliates have a staggering amount of within- and between-lineage paralogy of SFA genes. This implies that the similarity in cytoskeletal architecture between these two phylogenetically distant lineages is truly convergent. It should be interesting to find out what proteins constitute the fibres in hemimastigophorans that are analogous to those in ciliates.

It is possible, of course, that hemimastigophorans actually do have SFAs, and that other lineages likewise have them, but that their amino-acid sequences have diverged to the point that they are not recognisable by similarity searches. This would be in keeping with the common wisdom that intermediate-filament proteins in general are more conserved at the secondary and tertiary structure levels than they are at the primary (Harper et al. 2009). We doubt that this common wisdom is accurate, however; in fact, our bioinformatic analysis suggests that the opposite is the case. The proteins that we identified purely on the basis of amino-acid sequence similarity were effective at marking structural features in a previously uninvestigated organism, showing that primary structure is actually recognisably conserved. On the other hand, our structural analysis finds no consistent patterns in the number, size, or location of the coiled-coil domains thought to be characteristic of SFAs. In fact, we find that an expanded survey of ciliate lineages provides more exceptions to than confirmations of the patterns predicted to be characteristic of the various clades within ciliates, in spite of us using some of the same tools as the studies predicting these patterns (Nabi et al. 2019; Soh et al. 2020). Since we found no structural domains other than coiled-coil motifs with any consistency in our SFA data, we must conclude that secondary-structure similarities are not as relevant as has

previously been supposed. Of course, beyond our localisation work, the SFAs that we have identified have been completely uncharacterised, even at the level of banding patterns. Further work on these proteins may illuminate more relevant secondary and higher-order structural homologies across different lineages.

Expansion Microscopy of *Paratrimastix pyriformis*

The fact all three groups of SFAs (1, 2a and 2b) are present in the genomes of all investigated 'typical excavates' (Simpson 2003) led us to the idea of localising these proteins in one of their representatives, *Paratrimastix pyriformis*. However, our use of tubulin ABs to colocalise with SFAs produced noteworthy results on its own. The ultrastructure of *P. pyriformis* was previously studied by O'Kelly et al. (1999), who performed an analysis of two strains using serial-section transmission electron microscopy (TEM). One of these strains ('RCP' = ATCC 50562) is the same as the one used in our work. Additional context can be provided by TEM studies of the related 'typical excavates' *Trimastix marina* (Zhang et al. 2015) and *P. eleionoma* (Simpson et al. 2000; note that this organism was at the time called "*Trimastix marina*", and is not the same as the organism described by Zhang et al. 2015), and, to a lesser extent, the oxymonad *Monocercomonoides hausmanni* (Simpson et al. 2002).

Although membranous structures were not possible to visualise directly under expansion microscopy (ExM), a silhouette of the cell was usually visible from AB-stained monomers in the cytoplasm or from non-specific staining, and the overall architecture of the cytoskeleton was apparent from the tubulin stain (Fig. 3, 4). Besides the overall shape of the cell, we note that the feeding groove was also readily identifiable, as it is framed by the coherent posterior roots (specifically, the CR and oRR) and spanned by the iRMTs (Fig. 3, 4). The feeding groove is the eponymous structure of excavates (Simpson 2003), and hypotheses of microtubular root homology are based in part upon the roots' relation to the groove. In particular, the LR and RR are given the 'universal' labels 'R1' and 'R2', respectively, and based on its position, the AR is given the 'universal' label 'R3'. These 'universal' terms are used to ease comparison between eukaryotic lineages (Moestrup 2000; Simpson 2003; Yubuki and Leander 2013).

Also apparent was that many of our cells were significantly damaged, in some cases with individual MTs separated from their roots and coiling loosely, giving the impression of a broken spring in a clock. This cautions us against interpreting the overall architecture of the cytoskeleton shown in our micrographs too directly. However, the colocalisation of SFA ABs to tubulin in damaged cells was consistent with that of better-preserved cells, thereby suggesting that the two proteins bound very tightly together. Regardless, the variability of preservation also suggests that ExM may be further optimised.

The Microtubular Roots of *Paratrimastix pyriformis*

One of the primary microtubular structures characteristic of 'typical excavate' architecture is a posterior right root (RR or R2) that splits prominently, relatively close to its point of origin, into two subroots with roughly comparable numbers of MTs (Simpson 2003). We observed two splits in the RR, however, neither very wide (Fig. 4D, E). This lack of prominence was due in part to the MTs of the middle part spacing out somewhat from one another, and those from the inner or leftmost part separating more widely toward the left side of the cell, to fray apart independently as iRMTs. If we are to distinguish an 'inner' from an 'outer' RR, then, the distinction must be made between the two rightmost subroots, which in addition to the subtle split between them, took different paths at the posterior end of the cell, and were most clearly delineated by the iRR being stained, and the oRR being unstained, by SFA ABs (Fig. 4C–G). This distinction is sufficiently subtle as not to have been apparent in the TEM work of O'Kelly et al. (1999), who regarded the iRMTs as "an indistinct inner subroot" (p. 155). They did not illustrate the posterior end of the cell, and their text only stated that the LR (which we take as the CR) and the oRR "reunited at the posterior end of the ventral groove" (p. 155). Our work, even though of a lower resolution and of damaged cells, nevertheless suggests

consistently that the reunion is more complex than that: there are in fact three groups of MTs (the CR, iRR, and oRR), and although the oRR and CR do closely follow the same path, they do not merge, and the iRR deviates significantly from the end of that path (Fig. 4G).

Another consistent observation of our preparations was that the RR comprised no more than 20 MTs in total: the 'y' MT, 9–10 iRMTs, ~4 iRR MTs, and ~5 oRR MTs. In contrast, O'Kelly et al. (1999) describe ~30 MTs in the whole RR, and their micrographs show 40. Importantly, their micrographs are of the same strain as the one that we observed. This difference could have a number of explanations. It is possible that the RR originally comprised 40 MTs in our cultures, but that some MTs degraded in the course of our preparation. We suspect this to be unlikely, however: were that to be the case, we would expect to find varying numbers of MTs in the RR. In contrast, we were able to confirm ~20 MTs in seven cells, and our observations of over 40 others were more consistent with 20 than with 40 MTs. We can only suppose that, through differing culture conditions over the nearly quarter-century between our respective studies, the internal morphology of *P. pyriformis* has changed. To our knowledge, there have not been studies investigating this possibility. If it is realistic, it may be desirable to revisit the morphology of previously studied organisms. Regardless, our findings differ from those of O'Kelly et al. (1999) only in the number of MTs in the RR, and not in the overall arrangement of MTs or other structures.

Our observations of the multiple fates of MTs in the RR indicate that, rather than a simple bipartite split, there are actually four subunits, as listed above. The separation of these is emphasised by the presumedly different amount of SFA binding to MTs in different parts. Specifically, the 'y' MT was tagged strongly; the iRMTs and the iRR were tagged so strongly as effectively to suppress the tubulin signal; and the oRR MTs were untagged entirely. The distinction between the iRMTs and the iRR, while subtle, was the most anterior separation, and the fates of the MTs of each group (as described above) were quite different. While it could be argued that there are in fact five subunits, with a distinction to be made between those iRMTs that abut the CR and those that pass it by, we suspect that this difference is artefactual. Specifically, the number of bypassing MTs was greater in more-distrupted cells, while the number of iRMTs in total was constant.

The complexity of division of the RR in *P. pyriformis* is taken further by *P. eleionoma*, which, besides a 'y' MT, an iRR, iRMTs, and an oRR, also has oRMTs to the right of all the others (Simpson et al. 2000). As our order of listing implies, however, the order of these subparts is different, in that the iRMTs lie between the iRR and oRR and not between the iRR and the 'y' MT. Furthermore, the iRR 'reassembles' from MTs that are indistinguishable from the iRMTs anterior to its formation. In addition, the inside part of the LR also frays apart in *P. eleionoma*, forming a substantial part of the groove floor. Finally, unlike in *P. pyriformis*, where the iRMTs all either abut or pass closely by the CR, and all terminate at a sharp angle to it, in *P. eleionoma* all MTs in both the LR and RR either persist to the posterior end of the cell or terminate while parallel to the 'assembled' parts of the roots. We note that these observations are emphatically inconsistent with the hypothesis of Simpson et al. (2000), promoted as well by Zhang et al. (2015), that 'reassembly' of iRMTs into an iRR is a feature shared by both *P. pyriformis* and *P. eleionoma*. In particular, 'reassembly' is not possible with iRMTs terminating at a steep angle to the CR; and if the second split is homologous in both species (i.e., it existed in their common ancestor), the iRRs of one organism correspond to the iRMTs of the other, and vice versa. Meanwhile, sufficient genetic data from *P. eleionoma* was not available to include in our bioinformatic survey, but since both *P. pyriformis* and *T. marina* have SFA paralogues in both Groups 1 and 2, we expect that *P. eleionoma* does as well. It will be interesting to learn which MTs SFA binds to in *P. eleionoma*.

Trimastix marina presents yet a third pattern of posterior roots. One could say that it is simpler than either species of *Paratrimastix* in that it comprises only three parts: an oRR retaining a tightly spaced arrangement of parallel MTs throughout the root's length, widely spaced MTs apparently arising from the oRR (confusingly, termed 'oRMTs' in Zhang et al. 2015, but in the same position and with the same origin as the 'iRMTs' in O'Kelly et al. 1999), and an iRR that arises directly

from the split (Zhang et al. 2015). Neither an array of spaced oRMTs (in the sense of those in *P. eleionoma*) nor a 'y' MT are described or shown by Zhang et al. (2015), thereby suggesting the 'y' MT as a synapomorphy for *Paratrimastix* in general, and the oRMTs as a synapomorphy for *P. eleionoma* in particular. Our bioinformatic survey only obtained a single copy of the Group 1 paralogue of SFA from *T. marina*; however, only ~60% as much data for *T. marina* was available as for *P. pyriformis*, and it is quite possible that additional paralogues went undetected in *T. marina*. Regardless, the use of even a single paralogue in *T. marina* will assist in refining hypotheses of homology, especially given the cross-sensitivity of our P7 and P8 ABs.

Independent Fibrous Structures in *Paratrimastix* and Other Protists

Strictly speaking, exact colocalisation between MTs and SFA ABs was only found in the RR. These ABs also bound to the LR, but only just anterior to its being joined by the 'y' MT from the RR (Fig. 3). The impression is made of the SFA signal 'transferring' from MT 'y' to the SR+LR combination. This suggests that SFA takes on an independent structural role here, in which it reinforces this junction. Testing such an hypothesis will be difficult, but given the known amino acid sequence of SFA, may be possible. It will certainly be interesting to see whether SFA performs the same 'transfer' between MTs in *P. eleionoma*.

The epipodial complex, a long, thin structure either running parallel to or nearby the LR, presents a puzzle, as the epipodium itself is not identifiable from fluorescence microscopy alone. The epipodium is shown to be accompanied by the excavate C fibre by O'Kelly et al. (1999). If the C fibre is composed of (or contains) SFA, this would be consistent with our observations of a structure containing both SFA and tubulin. While O'Kelly et al. (1999) describe the C fibre in *P. pyriformis* as nonstriated, their micrographs only show it to be amorphous in cross-section. Its spacing from the LR would be compatible with a striated structure between it and the LR MTs. Meanwhile, *P. eleionoma*'s C-fibre is described as striated (Simpson et al. 2000), although with striations of 15–20 nm periodicity, much smaller than the 28 nm measured by Lechtreck and Melkonian (1991) in their initial investigation of SFA. At present, then, the exact identity of the EC, which almost certainly does contain SFA, is best regarded as unresolved.

More certain is the proposition that the 'composite fibre' and 'groove-terminating sheet' seen at the posterior end of the cell in *P. eleionoma* (Simpson et al. 2000) and *T. marina* (Zhang et al. 2015) comprises or contains SFA in *P. pyriformis* (Fig. 3, 4E, G). These structures were not previously documented in *P. pyriformis* by O'Kelly et al. (1999), but their survey did not depict the posterior end of the groove, and described it in only cursorily. However, the composite fibre is documented in *P. eleionoma* (Simpson et al. 2000), appears in the same location as the structure we observed in *P. pyriformis*, and is reported to have a 30-nm periodicity of striation, which is within the realm of comparison to the 28-nm periodicity of SFA. This is unlike the period of striation in *P. eleionoma*'s 'groove-terminal sheet' (GTS), reported as 15–20 nm (Simpson et al. 2000). Notably, this is the same period of striation as that of the C-fibre, raising the possibility that the GTS and C-fibre contain the same paralogue of SFA.

Even more certain is the proposition that the B fibre is composed of SFA. The structure that we identify as the B-fibre is in the same position, makes the same connections, and has the same shape as the B-fibre observed in the same strain of *P. pyriformis* under TEM (O'Kelly et al. 1999); and like the composite fibre, the B-fibre's 30-nm striations are comparable to the 28-nm periodicity described for SFA. Importantly, unlike the other anti-SFA-tagged structures in *P. pyriformis*, the B-fibre forms a structure that is mostly independent of MTs. In other words, the proteins making up a B-fibre should not all colocalise with tubulin. The putative B-fibre in our preparations is likewise visible by an SFA signal at a clear distance from any tagged tubulins (Fig. 4B, C).

Both the B-fibre and the composite fibre are thought to be characteristic of 'typical excavates' (Simpson 2003), and homologues of these structures are not currently known in other organisms (Heiss et al. 2018). Their homology among excavates remains conjectural, supported only by their

similar position and appearance under TEM. The proposition of homology would be made far stronger if it were shown that B-fibres and composite fibres in all 'typical excavates' are constructed from the same materials. Our localisation of SFAs in these fibres offers SFAs as a possible tool by which hypotheses of homology may be tested. In other words, if all of these fibres are composed of SFAs, they are more probably actually homologous.

But that homology may extend further. The narrower root of chlorophyte algae is thought to be homologous with the excavate RR (Moestrup 2000; Yubuki and Leander 2013). System I fibres, which are built out of SFA (Lechtreck and Melkonian 1991) and which were first described in green algae (Melkonian and Preisig 1984), have a geometric relationship to these roots that is not dissimilar to that of the B fibre to the RR in excavates. In other words, it is possible that the B fibre is not unique to excavates, being both compositionally and structurally homologous in organisms with cell architectures and phylogenetic relationships far removed from those of 'typical excavates'.

Paralogy within the Striated Fibre Assemblin Gene Family

Our observations are of structures tagged by ABs raised against two Group 1 paralogues of SFA (P7 and P8). We find that both ABs tag the same structures, in spite of the substantial sequence differences between these paralogues (evident in their branch lengths in Fig. 1). However, this need not mean that both proteins are found in the same structures, since only their recognition sites need be similar. Given the varying associations between MTs and the structures tagged by our ABs, for instance, it is possible that one paralogue is found in the more-intimately MT-bound material accompanying the iRR and iRMTs, and the other in structures maintaining an identity and shape at least in part independent of MTs, such as the B-fibre. Another candidate dichotomy is offered by, on the one hand, the 30-nm periodicity of striations reported for the B-fibre and composite fibre, and on the other, the 15–20-nm periodicity reported for the C-fibre and GTS (Simpson et al. 2000). It could be that one Group 1 SFA paralogue has a periodicity of 28–30 nm and the other of 15–20 nm. This could be investigated through in vitro characterisation of each protein. The proteins' tendency to self-polymerise (data not shown) suggests the feasibility of such an investigation. Additional confirmation of this would result from more-specific ABs to each paralogue.

Regardless, while other 'typical excavates' have putatively homologous structures, most do not have multiple paralogues within the Group 1 clade of SFAs. Furthermore, the high-to-maximal support values for each excavate clade within our phylogeny suggests that those lineages that do have multiple paralogues evolved those paralogues separately from one another. Any differences in localisation between P7 and P8 are therefore likely the result of subfunctionalisation, the ancestral preaxostylan most probably having a single Group 1 SFA that served all of the purposes seen in *P. pyriformis*.

Contrasting with this situation is that of the larger-scale paralogy within the larger SFA gene family. In particular, it may be of note that, while one or another paralogue is unknown from several lineages within the 'mega-group' Diaphoretickes, all 'typical excavate' lineages possess SFA paralogues in Groups 1, 2a, and 2b (Fig. 5; S1). There are several other fibrous structures in the 'typical excavate' cell architecture that remain uncharacterised. In particular, the I fibre has a markedly periodic substructure (Simpson 2003), and it may be that this structure contains Group 2 SFAs. Should this be the case, and should it be consistent across different excavate lineages, the homology of this conspicuous fibre would be much more probable. We note that Group 2 paralogues have to date only been characterised in ciliates (Nabi et al. 2019; Soh et al. 2020). This offers localisation of Group 2 paralogues in both *P. pyriformis* and other organisms as a prime goal for future AB studies.

While SFAs are known as structural proteins, we might propose nonstructural functions for them as well. In almost all flagellates, and in all known 'typical excavates', cells do not divide symmetrically (Beech et al. 1991). Rather, one daughter cell always receives the parental anterior flagellum (AF), and the other the parental posterior flagellum (PF). Each of the parent cell's roots and

other flagellum-related structures are passed on with its respective flagellum. The daughter cell receiving the parental PF maintains that flagellum as a PF, and all of its roots remain posterior roots; while in the other daughter cell, the AF transforms into a PF, and existing anterior roots transform into posterior roots, which remain posterior roots through all subsequent divisions. This means that PF-related structures are permanent, outlasting the independent lifetime of individual cells as the roots are inherited between generations. Significantly, with the possible exception of the EC, all of the structures tagged by our Group 1 SFA ABs are associated with the PF's basal body. We might conjecture that the presence of SFAs on the iRR and iRMTs could provide a signal to maintain them intact while AF-related structures are broken down or modified, or that SFA stabilises the posterior-system MTs directly, or both. This is of course speculation, and any role of SFAs in maintaining cytoskeletal structures across generations should be investigated through (e.g.) knockdown experiments.

Conclusions & Future Directions

In this work we localise striated fiber assemblins (SFAs), known from multiple seemingly unrelated structures — the adhesive disc of *Giardia* (Holberton et al. 1988), the kinetodesmal fibers of ciliates (Nabi et al. 2019; Soh et al. 2020), the cell division apparatus of apicomplexans (Lechtreck 2003), and the mastigonts of green algae (Lechtreck and Melkonian 1991) — in the cell of the 'typical excavate' *Paratrimastix pyriformis*. We find that the SFAs we investigated target only structures associated with the posterior flagellum, but in numerous places, and with differing relationships to microtubules (MTs). These range from colocalisation along the entire length of the MTs of the inner part of the right root, to structures associated with but separate from MTs, such as the composite fibre at the posterior end of the cell, to structures attached to MTs only at their edges, maintaining their form otherwise completely independently, such as the B-fibre. All of these parts of the cytoskeleton contribute to forming the armature of the ventral feeding groove. The 'typical excavate' morphology based upon this is found amongst organisms comprising three lineages of uncertain relationships to one another, but all of which are thought to be close to the root of the eukaryote tree (Brown et al. 2018; Heiss et al. 2018). As such, this morphology may be characteristic of the situation in the last eukaryotic common ancestor (LECA). If so, this suggests that the original function of the SFA protein family in eukaryotes might have some relation to the ontogeny or functionality of the excavate feeding apparatus. This hypothesis should be confirmed by localisation in other 'typical excavate' taxa.

At the same time, we confirmed and expanded the earlier finding (Soh et al. 2020) of three paralogous clades of SFAs common to 'typical excavates' and many members of Diaphoretickes (Adl et al. 2012, 2019). In particular, the separation between Groups 1 and 2 (the latter including as subclades Groups 2a and 2b) is well-supported, the branch between the clades is substantially long, and the sequences of the proteins in each clade follow visibly different patterns when combined in the same alignment. While not as strongly separated, members of Groups 2a and 2b are still unambiguously assignable to only one paralogue. This is consistent with all three paralogues being present in the common ancestor to all lineages containing them. To date, there has been no taxonomic overlap between those lineages in which Group 1 and Group 2 paralogues have been investigated (Soh et al. 2020). As such, investigation of the 'missing' clade in organisms already characterised should be enlightening.

It should be noted that studies of the eukaryotic cytoskeleton have tended to focus on the region of the cell proximal to the basal bodies (O'Kelly et al. 1999 is one such example). Indeed, it is this region that has been the focus of hypotheses of morphological homologies across unicellular eukaryotes (Sleigh 1988; Andersen et al. 1991; Moestrup 2000; Yubuki and Leander 2013). Whether such rearrangements as we have confirmed in *Paratrimastix* are found in other unicellular eukaryotes is largely unknown, and should be surveyed in the future. We note that ExM offers a venue in which to accomplish this, as well as to make other general cytoskeletal observations, with far less time and

expense than traditional serial-section TEM work. This is not to say that ExM should be seen as a replacement for TEM studies, as it can only show AB-labelled structures, and each methodology is prone to different artefacts. Rather, we see ExM and TEM as potentially complementary approaches to address the same topics. In particular, the identity of the non-MT components of the cytoskeleton, which would be shown more conclusively by immunological TEM, would be impractical to determine without a probable target (in this case, SFA) identified by other means, such as ExM. Thus, our conjectures of 'typical excavate' fibres containing SFAs should be confirmed using immunological TEM in *P. pyriformis*, and they offer a target to be explored using that method in other organisms.

Overall, our work offers a tantalising glimpse at putative homologies across much of the tree of eukaryotes. A combination of deeper investigation of unsampled SFA paralogues within already-examined lineages and wider surveys of hitherto uninvestigated lineages will build on this, promising to establish a structural basis for comparison of eukaryote morphology approaching the same level of informativeness as that of microtubules. We expect future work, including immunolocalisation under transmission electron microscopy, to offer great insights into the common architecture of the eukaryotic cell.

ACKNOWLEDGEMENTS

We would like to thank to Sebastian Cristian Treilti and Zoltán Füssy for help with bioinformatic searches through the genomic and transcriptomic datasets. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 771592) and by supported by Czech Science Foundation project nr. 23-07277S.

REFERENCES

Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, le Gall L, Lynn DH, McManus H, Mitchell EAD, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick RS, Schoch CL, Smirnov A, Spiegel FW (2012): The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **57:** 429–493 (doi: 10.1111/j.1550-7408.2012.00644.x).

Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F, Cárdenas P, •epi•ka I, Chistyakova L, del Campo J, Dunthorn M, Edvardsen B, Eglit Y, Guillou L, Hampl V, Heiss AA, Hoppenrath M, James TY, Karpov S, Kim E, Kolisko M, Kudryavtsev A, Lahr DJG, Lara E, le Gall L, Lynn DH, Mann DG, Massana i Molera R, Mitchell EAD, Morrow C, Park JS, Pawlowski JW, Powell MJ, Richter DJ, Rueckert S, Shadwick L, Shimano S, Spiegel FW, Torruella i Cortes G, Youssef N, Zlatogursky V, Zhang Q (2019): Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.*, **66:** 4–119 (doi: 10.1111/jeu.12691).

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990): Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410 (doi: 10.1016/S0022-2836(05)80360-2).

Andersen RA, Barr DJS, Lynn DH, Melkonian M, Moestrup Ø, Sleigh MA (1991): Terminology and nomenclature of the cytoskeletal elements associated with the flagellar/cillary apparatus in protists. Protoplasma **164:** 1–8.

Beech PL, Heimann K, Melkonian M (1991): Development of the flagellar apparatus during the cell cycle in unicellular algae. *Protoplasma* **164:** 70–90.

Brown MW, Heiss AA, Kamikawa R, Inagaki Y, Yabuki A, Tice AK, Shiratori T, Ishida K, Hashimoto T, Simpson AGB, Roger AJ (2018): Phylogenomics places orphan protistan lineages as deep

sisters to Amorphea and identifies a novel candidate super-group. *Genome Biol. & Evol.* **10:** 427–433 (doi: 10.1093/gbe/evy014).

Cerón-Romero MA, Fonseca MM, de Oliveira Martins L, Posada D, Katz LA (2022): Phylogenomic analyses of 2,768 genes in 158 lineages support a root of the eukaryotic tree of life between opisthokonts and all other lineages. *Genome Biol. Evol.* **14:** evac119 (doi: 10.1093/gbe/evac119).

Derelle R, Torruella G, Klimeš V, Brinkmann H, Kim E, Vl•ek •, Lang BF, Eliáš M (2015): Bacterial proteins pinpoint a single eukaryotic root. *Proc. Nat'l. Acad. Sci. USA* **112:** E693–E699 (doi: 10.1073/pnas.1420657112).

Edgar RC (2004): MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **32:** 1792–1797 (doi: 10.1093/nar/gkh340).

Foissner W, Blatterer H, Foissner I (1988): The Hemimastigophora (*Hemimastix amphikineta* nov. gen., nov. spec.), a new protistan phylum from Gondwanian soils. *Europ. J. Protistol.* **23:** 361–383 (doi: 0932-4739/88/0023-0361).

Francia ME, Jordan CN, Patel JD, Sheiner L, Demerly JL, Fellows JD, Cruz de Leon J, Morrissette NS, Dubremetz JF, Striepen B (2012): Cell division in apicomplexan parasites is organized by a homolog of the striated rootlet fiber of algal flagella. *PLoS Biol.* **10:** e1001444 (doi: 10.1371/journal.pbio.1001444).

Füssy Z, Vinopalová M, Treitli SC, Pánek T, Smejkalová P, •epi•ka I, Doležal P, Hampl V (2021): Retortamonads from vertebrate hosts share features of anaerobic metabolism and pre-adaptations to parasitism with diplomonads. *Parasitol. Int.* **82:** 102308 (doi: 10.1016/j.parint.2021.102308).

Gorilak P, Pružincová M, Vachova H, Olšinová M, Cernohorska MS, Varga V (2021): Expansion microscopy facilitates quantitative super-resolution studies of cytoskeletal structures in kinetoplastid parasites. *Open Biol.* **11:** 210131 (doi: 10.1098/rsob.210131).

Harper JDI, Thuet J, Lechtreck KF, Hardham AR (2009): Proteins related to green algal striated fiber assemblin are present in stramenopiles and alveolates. *Protoplasma* **236:** 97–101 (doi: 10.1007/s00709-009-0041-z).

Heiss AA, Kolisko M, Ekelund F, Brown MW, Roger AJ, Simpson AGB (2018): Combined morphological and phylogenomic reexamination of malawimonads, a critical taxon for inferring the evolutionary history of eukaryotes. *R. Soc. Open Sci.* **5:** 171707 (doi: 10.1098/rsos.171707).

Holberton DV, Ward AP (1980): Isolation of the cytoskeleton from *Giardia*. Tubulin and a low molecular weight protein associated with microribbon structures. *J. Cell Sci.* **47:** 139–166.

Holberton D, Baker DA, Marshall J (1988): Segmented •-helical coiled-coil structure of the protein giardin from the *Giardia* cytoskeleton. *J. Mol. Biol.* **204:** 789–795.

Lax G, Eglit Y, Eme L, Bertrand EM, Roger AJ, Simpson AGB (2018): Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. *Nature* **564:** 410–414 (doi: 10.1038/s41586-018-0708-8).

Lechtreck KF (2003): Striated fiber assemblin in apicomplexan parasites. *Mol. Biochem. Parasitol.* **128:** 95–99 (doi: 10.1016/S0166-6851(03)00038-0).

Lechtreck KF, Melkonian M (1991): Striated microtubule-associated fibers: Identification of assemblin, a novel 34-kD protein that forms paracrystals of 2-nm filaments in vitro. *J. Cell Biol.* **3:** 705–716.

Leger MM, Kolisko M, Kamikawa R, Stairs CW, Kume K, •epi•ka I, Silberman JD, Andersson JO, Xu F, Yabuki A, Eme L, Zhang Q, Takishita K, Inagaki Y, Simpson AGB, Hashimoto T, Roger AJ (2017): Organelles that illuminate the origins of Trichomonas hydrogenosomes and Giardia mitosomes. *Nat. Ecol. Evol.* **1:** 0092 (doi: 10.1038/s41559-017-0092).

Letunic I, Bork P (2018): 20 years of the SMART protein domain annotation resource. *Nuc. Acids Res.* **46:** D493–D496.

Mach J, Poliak P, Matušková A, Žárský V, Janata J, Lukeš J, Tachezy J (2013): An advanced system of the mitochondrial processing peptidase and core protein family in *Trypanosoma brucei* and

multiple origins of the core I subunit in eukaryotes. *Genome Biol. Evol.* **5:** 860–875 (doi: 10.1093/gbe/evt056).

Melkonian M (1980): Ultrastructural aspects of basal body associated fibrous structures in green algae: A critical review. *BioSystems* **12:** 85–104.

Melkonian M, Preisig HR (1984): Ultrastructure of the flagellar apparatus in the green flagellate *Spermatozopsis similis*. *Pl. Syst. Evol.* **146:** 145–162.

Moestrup Ø (1982): Flagellar structure on algae: A review with new observations particularly on the Chrysophyceae, Phaeophyceae (Fucophyceae), Euglenophyceae and *Reckertia*. *Phycologia* **21:** 425–528.

Moestrup Ø (2000): The flagellate cytoskeleton: Introduction of a general terminology for microtubular flagellar roots in protists. Pages 69–94 in Green JC, Leadbeater BSC (eds.): *The Flagellates: Unity, Diversity and Evolution*. London: Taylor & Francis.

Nabi A, Yano J, Valentine MS, Picariello T, Van Houten JL (2019): SF-Assemblin genes in *Paramecium*: Phylogeny and phenotypes of RNAi silencing on the ciliary-striated rootlets and surface organization. *Cilia* **8:** 2 (doi: 10.1186/s13630-019-0062-y).

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015): IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32:** 268–274 (doi: 10.1093/molbev/msu300).

Novák LVF, Treitli SC, Pyrih J, Hałakuc P, Pipaliya SV, Vacek V, Brzo• O, Soukal P, Eme L, Dacks JB, Karnkowska A, Eliáš M, Hampl V (2023): Genomics of Preaxostyla flagellates illuminates the path towards the loss of mitochondria. *PLoS Genet.* **19:** e1011050 (doi: 10.1371/journal.pgen.1011050).

O'Kelly CJ, Farmer MA, Nerad TA (1999): Ultrastructure of *Trimastix pyriformis* (Klebs) and similarities of *Trimastix* species with retortamonads and jakobids. *Protist* **150:** 149–162.

Palm D, Weiland M, McArthur AG, Winiecka-Krusnell J, Cipriano MJ, Birkeland SR, Pacocha SE, Davids B, Gillin F, Linder E, Svärd S (2005): Developmental changes in the adhesive disk during *Giardia* differentiation. *Mil. Biochem. Parasitol.* **141:** 199–207 (doi: 10.1016/j.molbiopara.2005.03.005).

Peña-Diaz P, Braymer JJ, Vacek V, Zelená M, Lometto S, Hrdý I, Treitli SC, Hochberg G, Py B, Lill R, Hampl V (2023): Characterisation of the SUF FeS cluster machinery in the amitochondriate eukaryote *Monocercomonoides exilis*. bioRxiv preprint (doi: 10.1101/2023.03.30.534840).

Sleigh MA (1988): Flagellar root maps allow speculative comparisons of root patterns and of their ontogeny. *Biosystems* **21:** 277–282.

Richter DJ, Berney C, Strassert JFH, Poh YP, Herman EK, Muñoz-Gómez SA, Wideman JG, Burki F, de Vargas C (2022): EukProt: A database of genome-scale predicted proteins across the diversity of eukaryotes. *Peer Comm. J.* **2:** e56 (doi: 10.24072/pcjournal.173).

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012): MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61:** 539–542 (doi: 10.1093/sysbio/sys029).

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012). Fiji — An open source platform for biological image analysis. *Nat. Methods* **9:** 10.1038/nmeth.2019 (doi: 10.1038/nmeth.2019).

Simpson AGB (2003): Cytoskeletal organisation, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). *Int. J. Syst. Evol. Microbiol.* **53:** 1759–1777.

Simpson AGB, Bernard C, Patterson DJ (2000): The ultrastructure of *Trimastix marina* Kent, 1880 (Eukaryota), an excavate flagellate. *Eur. J. Protistol.* **36:** 229–252.

Simpson AGB, Radek R, Dacks JB, O'Kelly CJ (2002): How oxymonads lost their groove: An ultrastructural comparison of *Monocercomonoides* and excavate taxa. *J. Eukaryot. Microbiol.* **49:** 239–248.

Soh AWJ, van Dam TJP, Stemm-Wolf AJ, Pham AT, Morgan GP, O'Toole ET, Pearson CG (2020): Ciliary force-responsive striated fibers promote basal body connections and cortical interactions. *J. Cell Biol.* **219:** e201904091 (doi: 10.1083/jcb.201904091).

Stamatakis A (2014): RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312–1313 (doi: 10.1093/bioinformatics/btu033).

Treitli SC, Kolisko M, Husník F, Keeling PJ, Hampl V (2019): Revealing the metabolic capacity of *Streblomastix strix* and its bacterial symbionts using single-cell metagenomics. *Proc. Nat'l Acad. Sci. USA* **116:** 19675–19684 (doi: 10.1073/pnas.1910793116).

Treitli SC, Peña-Diaz P, Hałakuc P, Karnkowska A, Hampl V (2021): High quality genome assembly of the amitochondriate eukaryote *Monocercomonoides exilis*. *Microb. Genom.* **7:** 000745 (doi: 10.1099/mgen.0.000745).

Weber K, Geisler N, Plessmann U, Bremerich A, Lechtreck KF, Melkonian M (1993): SF-assemblin, the structural protein of the 2-nm filaments from striated microtubule associated fibers of algal flagellar roots, forms a segmented coiled coil. *J. Cell Biol.* **121:** 837–845.

Yubuki N, Leander BS (2013): Evolution of microtubule organizing centers across the tree of eukaryotes. *Plant J.* **75:** 230–244 (doi: 10.1111/tpj.12145).

Zhang Q, Táborský P, Silberman JD, Pánek T, •epi•ka I, Simpson AGB (2015): Marine isolates of *Trimastix marina* form a plesiomorphic deep-branching lineage within Preaxostyla, separate from other known trimastigids (*Paratrimastix* n. gen.). *Protist* **166:** 468–491.

Zítek J, Füssy Z, Treitli SC, Peña-Diaz P, Vaitová Z, Zavadska D, Harant K, Hampl V (2022): Reduced mitochondria provide an essential function for the cytosolic methionine cycle. *Curr. Biol.* **32:** 5057–5068.e5 (doi: 10.1016/j.cub.2022.10.028).

FIGURE LEGENDS

Figure 1: Phylogenetic tree of Group 1 paralogues of striated fibre assemblin. Tree was generated from maximum-likelihood search using GTR+gamma model as implemented in RAxML; topology is from the best of 100 starting random-taxon-addition parsimony trees, and support values at nodes are from 1000 nonparametric bootstrap replicates. Coloured background highlights previously characterised lineages; shaded sequence names indicate previously investigated proteins. Red arrows indicate homologues used in present study. Figures at right are mappings of coiled-coil domains onto sequence, with each character indicating seven amino acid sites (the number of sites generating a single coil); '@' indicates location of probable coiled-coil domains, '-' indicates sequence with no coiled-coil domain prediction; absence of symbol indicates sequence not available.

Figure 2: Phylogenetic tree of Group 2 paralogues of striated fibre assemblin. Tree was generated from maximum-likelihood search using GTR+gamma model as implemented in RAxML; topology is from the best of 100 starting random-taxon-addition parsimony trees, and support values at nodes are from 1000 nonparametric bootstrap replicates. Upper (purple) region indicates Group 2b paralogues; lower (salmon) region indicates Group 2a. Ciliates, the only lineage in which Group 2 paralogues have been characterised, highlighted in yellow. Shading with numbers indicates 'structural groups' defined by Nabi et al. 2019. Coloured sequence names indicate previously characterised proteins. Figures at right are mappings of coiled-coil domains onto sequence, with each character indicating seven amino acid sites (the number of sites generating a single coil); '@' indicates location of probable coiled-coil domains, '-' indicates sequence with no coiled-coil domain prediction; absence of symbol indicates sequence not available.

Figure 3: Expansion microcopy of whole cell of *Paratrimastix pyriformis*. Cell is viewed from ventral-right side, stained with antibodies to structural proteins. Note that all flagella except

posterior flagellum have broken off close to their insertion points. Top: striated fibre assemblin paralogue *PaPyr8006*; middle: tubulin; bottom: merged images. Key to annotations: 'AF' = anterior flagellum; 'AR' = anterior root; 'B' = B-fibre; 'CR' = compound root; 'EC' = epipodial complex; 'fan' = dorsal fan; 'iRMT' = inner right root microtubules; 'iRR' = inner right root; 'LF' = left flagellum; 'PF' = posterior flagellum; 'oRR' = outer right root; 'RF' = right flagellum; 'SR+EC+LR' = compound root comprising singlet root, EC, left root; 'y' = 'y' microtubule in iRMT. Scale bar ~ 2 µm (approximated based upon diameter of flagellar axonemes and size of cell).

Figure 4: Optical slices from expansion microscopy of cell of *Paratrimastix pyriformis*. Cell is viewed in cross-section, with dorsal to top of page and cell's left to viewer's right. Slices are spaced ~2–4 µm apart. Left panels: striated fibre assemblin paralogue *PaPyr8006*; middle panels: tubulin; right panels: merged images. Key to annotations: 'ABB' = anterior basal body; 'AF' = anterior flagellum; 'AR' = anterior root; 'B' = B-fibre; 'CF' = composite fibre; 'CR' = compound root; 'EC' = epipodial complex; 'fan' = dorsal fan; 'GTS' = groove-terminating sheet 'iRMT' = inner right root microtubules; 'iRMTs' = inner right-root microtubules; 'iRR' = inner right root; 'LBB' = left basal body; 'LF' = left flagellum; 'LR' = left root; 'PBB' = posterior basal body; 'PF' = posterior flagellum; 'oRR' = outer right root; 'RBB' = right basal body; 'RF' = right flagellum; 'RR' = right root; 'SR' = singlet root; 'SR+EC+LR' = compound root comprising singlet root, EC, left root; 'y' = 'y' microtubule in iRMT. Arrow indicates initial split between iRR and iRMTs. Scale bar ~ 2 µm (approximated based upon diameter of flagellar axonemes and size of cell).

Figure 5: Summary trees showing known distribution of paralogues of striated fibre assemblin in major lineages of eukaryotes. Black lines indicate lineages known to possess at least one copy of paralogue represented by tree; grey lines indicate no known sequence. Light coloured background indicates eukaryotic 'megagroups' that may be paraphyletic; dark coloured arcs at edges of tree indicate eukaryotic 'supergroups' that are expected to be holophyletic.

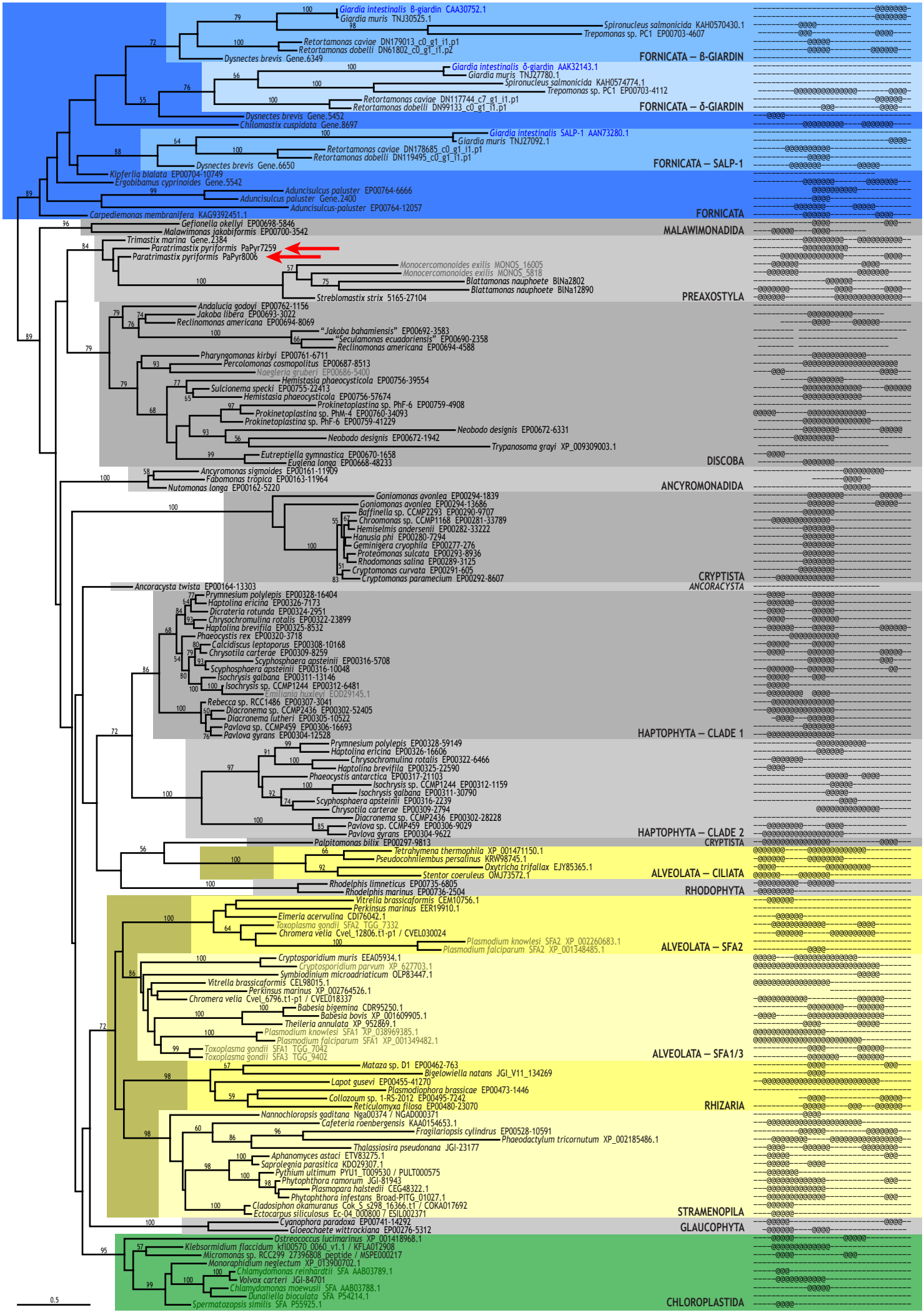
SUPPLEMENTARY DATA

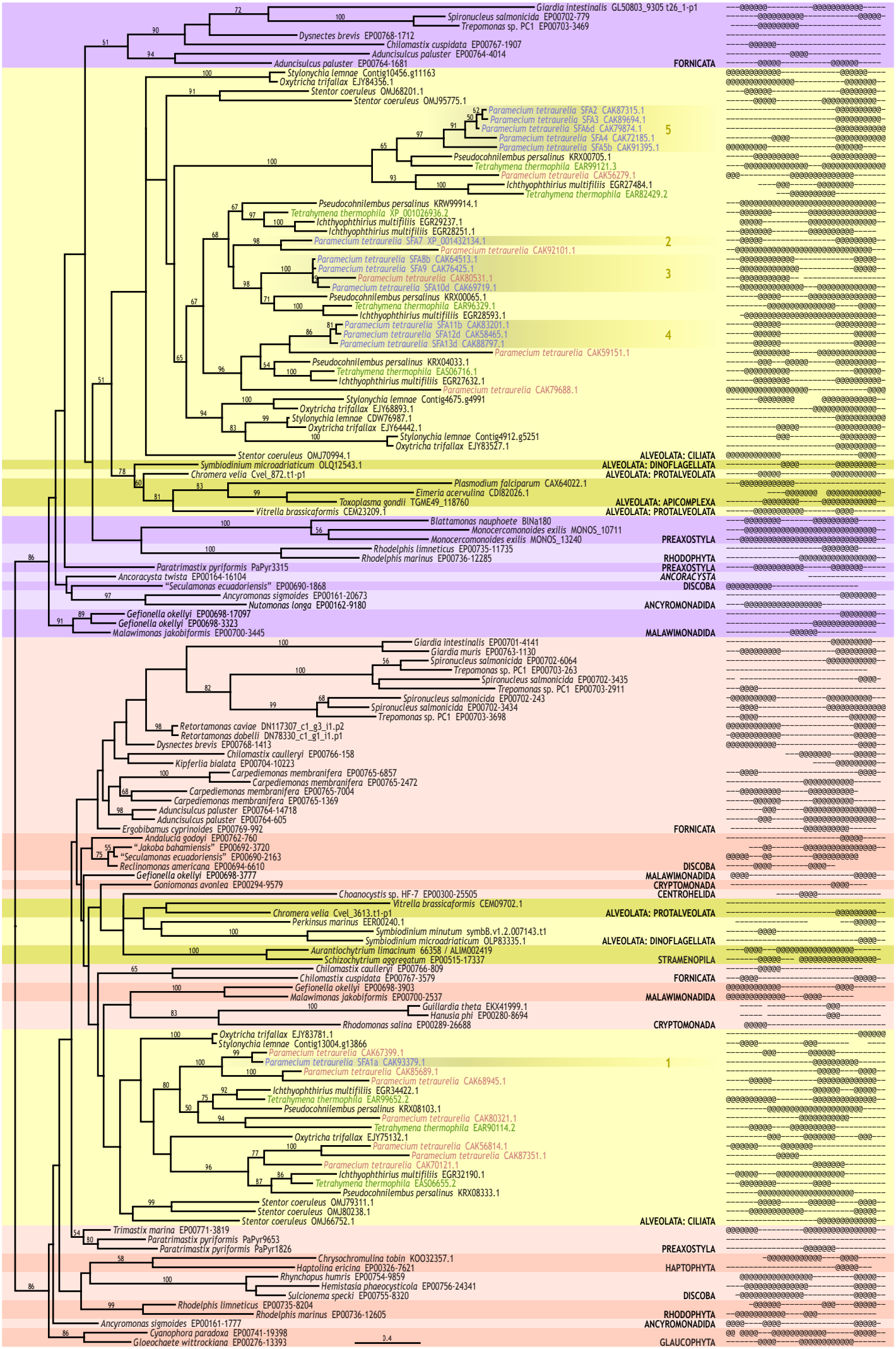
Figure S1: Phylogeny of combined dataset containing all paralogues of striated fibre assemblin. Tree was generated from maximum-likelihood search using GTR+gamma model as implemented in RAxML; topology is from the best of 100 starting random-taxon-addition parsimony trees, and support values at nodes are from 1000 nonparametric bootstrap replicates.

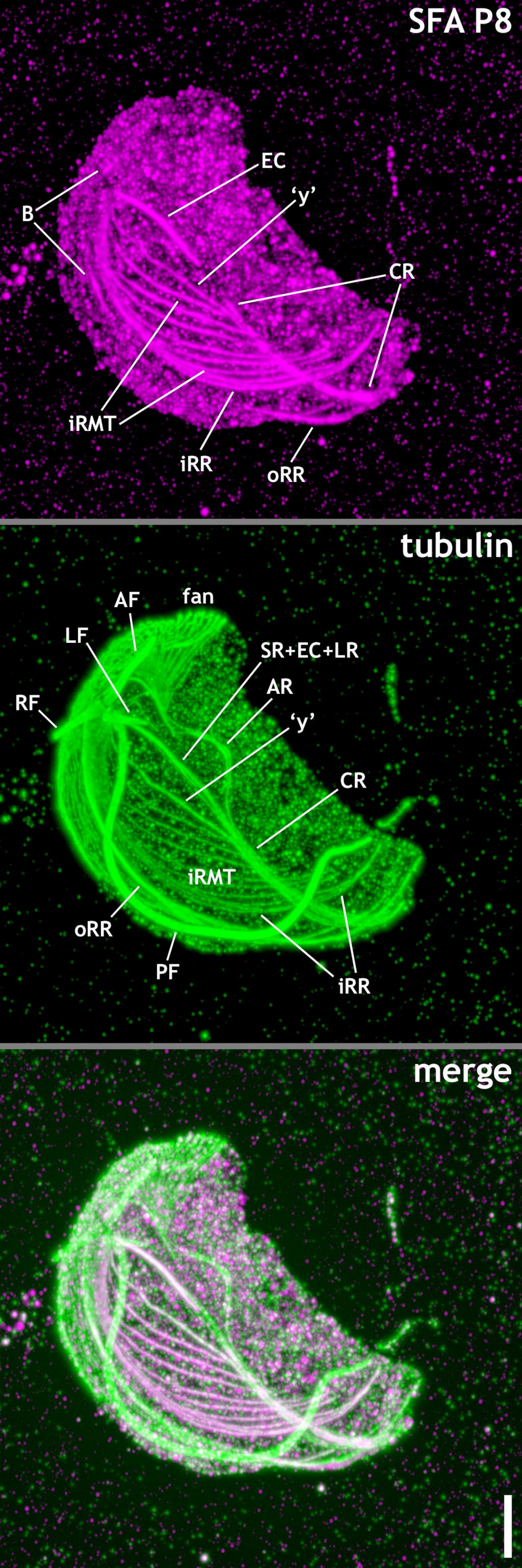
Table S1: Results of co-immunoprecipitation with P7 antibody. List of proteins identified after co-immuniprecipitation of sample with fold change relative to negative control (beads without antibody). Dash "-" indicates cases where protein was not identified in negative control. Striated fibre assemblins are shown in red.

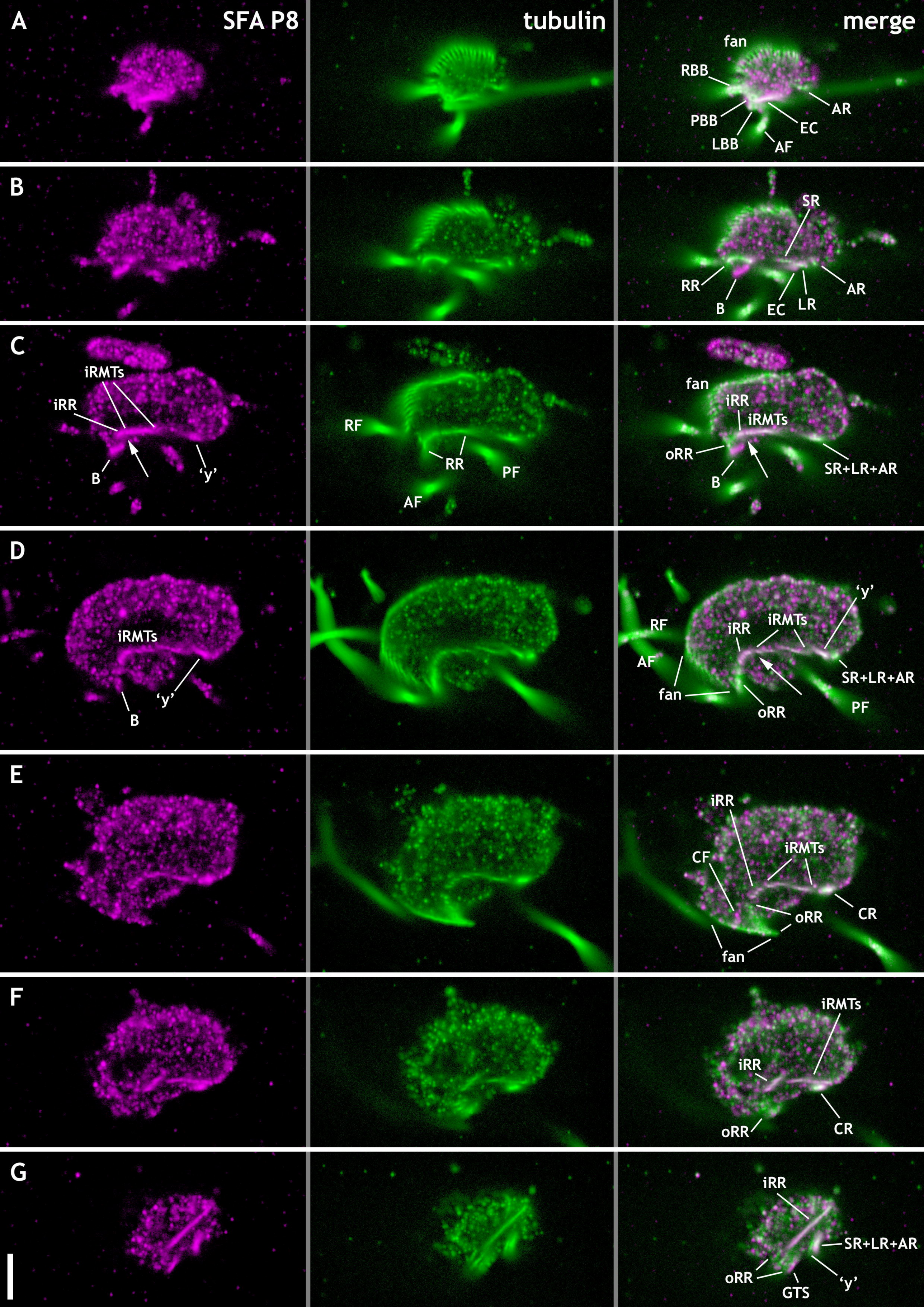
Table S2: Results of co-immunoprecipitation with P8 antibody. List of proteins identified after co-immuniprecipitation of sample with fold change relative to negative control (beads without antibody). Dash "-" indicates cases where protein was not identified in negative control. Striated fibre assemblins are shown in red.

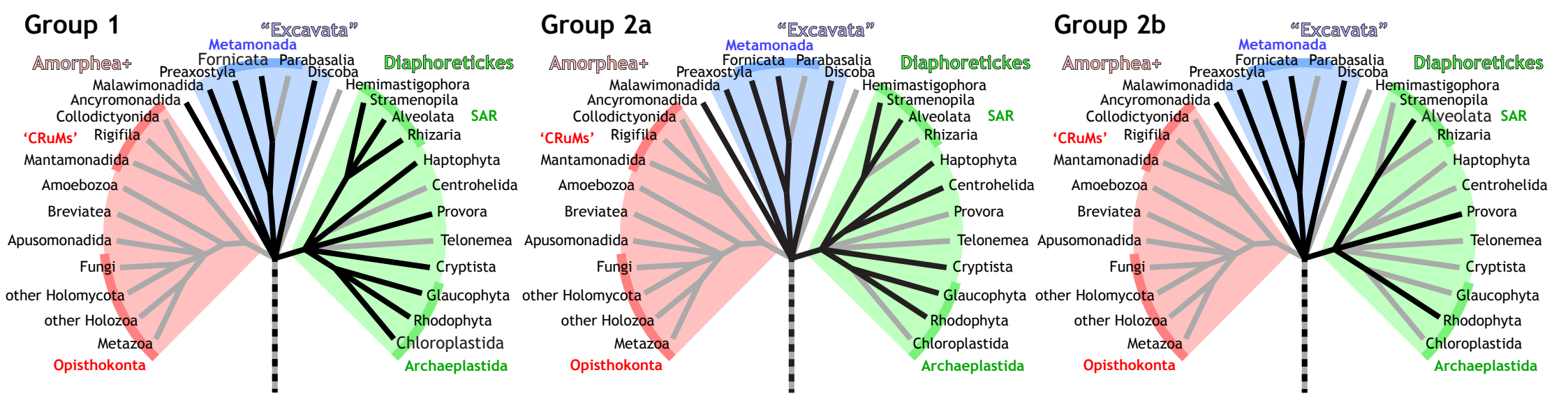
Table S3: Results of co-immunoprecipitation with P7 and P8 antibodies in RIPA buffer. List of proteins identified after co-immuniprecipitation of sample with fold change relative to negative control (beads without antibody). Dash "-" indicates cases where protein was not identified in negative control. Striated fibre assemblins are shown in red.


Giardia intestinalis B-giardin CAA30752.1
Giardia muris TNJ30525.1
Spironucleus salmonicida KAH0570430.1
Trepomonas sp. PC1 EP00703-4607
Retortamonas caviae DN179013_c0_g1_i1.p1
Retortamonas dobelli DN61802_c0_g1_i1.p2
Dysnectes brevis Gene.6349
FORNICATA — B-GIARDIN
Giardia intestinalis δ-giardin AAK32143.1
Giardia muris TNJ27780.1
Spironucleus salmonicida KAH0574774.1
Trepomonas sp. PC1 EP00703-4112
Retortamonas caviae DN117744_c7_g1_i1.p1
Retortamonas dobelli DN99133_c0_g1_i1.p1
FORNICATA — δ-GIARDIN
Dysnectes brevis Gene.5452
Chilomastix cuspidata Gene.8697
Giardia intestinalis SALP-1 AAN73280.1
Giardia muris TNJ27092.1
Retortamonas caviae DN178685_c0_g1_i1.p1
Retortamonas dobelli DN119495_c0_g1_i1.p1
Dysnectes brevis Gene.6650
FORNICATA — SALP-1
Kipferlia bialata EP00704-10749
Ergobibamus cyprinoides Gene.5542
Aduncisulcus paluster EP00764-6666
Aduncisulcus paluster Gene.2400
Aduncisulcus-paluster EP00764-12057
Carpediemonas membranifera KAG9392451.1
FORNICATA
Gefionella okellyi EP00698-5846
Malawimonas jakobiformis EP00700-3542
MALAWIMONADIDA
Trimastix marina Gene.2384
Paratrimastix pyriformis PaPyr7259
Paratrimastix pyriformis PaPyr8006
Monocercomonoides exilis MONOS_16005
Monocercomonoides exilis MONOS_5818
Blattamonas nauphoete BlNa2802
Blattamonas nauphoete BlNa12890
Streblomastix strix 5165-27104
PREAXOSTYLA
Andalucia godoyi EP00762-1156
Jakoba libera EP00693-3022
Reclinomonas americana EP00694-8069
"Jakoba bahamiensis" EP00692-3583
"Seculamonas ecuadoriensis" EP00690-2358
Reclinomonas americana EP00694-4588
Pharyngomonas kirbyi EP00761-6711
Percolomonas cosmopolitus EP00687-8513
Naegleria gruberi EP00686-5400
Hemistasia phaeocysticola EP00756-39554
Sulcionema specki EP00755-22413
Hemistasia phaeocysticola EP00756-57674
Prokinetoplastina sp. PhF-6 EP00759-4908
Prokinetoplastina sp. PhM-4 EP00760-34093
Prokinetoplastina sp. PhF-6 EP00759-41229
Neobodo designis EP00672-6331
Neobodo designis EP00672-1942
Trypanosoma grayi XP_009309003.1
Eutreptiella gymnastica EP00670-1658
Euglena longa EP00668-48233
DISCOBA
Ancyromonas sigmoides EP00161-11909
Fabomonas tropica EP00163-11964
Nutomonas longa EP00162-5220
ANCYROMONADIDA
Goniomonas avonlea EP00294-1839
Goniomonas avonlea EP00294-13686
Baffinella sp. CCMP2293 EP00290-9707
Chroomonas sp. CCMP1168 EP00281-33789
Hemiselmis andersenii EP00282-33222
Hanusia phi EP00280-7294
Geminigera cryophila EP00277-276
Proteomonas sulcata EP00293-8936
Rhodomonas salina EP00289-3125
Cryptomonas curvata EP00291-605
Cryptomonas paramecium EP00292-8607
CRYPTISTA
Ancoracysta twista EP00164-13303
ANCORACYSTA
Prymnesium polylepis EP00328-16404
Haptolina ericina EP00326-7173
Dicrateria rotunda EP00324-2951
Chrysochromulina rotalis EP00322-23899
Haptolina brevifila EP00325-8532
Phaeocystis rex EP00320-3718
Calcidiscus leptoporus EP00308-10168
Chrysotila carterae EP00309-8259
Scyphosphaera apsteinii EP00316-5708
Scyphosphaera apsteinii EP00316-10048
Isochrysis galbana EP00311-13146
Isochrysis sp. CCMP1244 EP00312-6481
Emiliania huxleyi EOD29145.1
Rebecca sp. RCC1486 EP00307-3041
Diacronema sp. CCMP2436 EP00302-52405
Diacronema lutheri EP00305-10522
Pavlova sp. CCMP459 EP00306-16693
Pavlova gyrans EP00304-12528
HAPTOPHYTA — CLADE 1
Prymnesium polylepis EP00328-59149
Haptolina ericina EP00326-16606
Chrysochromulina rotalis EP00322-6466
Haptolina brevifila EP00325-22590
Phaeocystis antarctica EP00317-21103
Isochrysis sp. CCMP1244 EP00312-1159
Isochrysis galbana EP00311-30790
Scyphosphaera apsteinii EP00316-2239
Chrysotila carterae EP00309-2794
Diacronema sp. CCMP2436 EP00302-28228
Pavlova sp. CCMP459 EP00306-9029
Pavlova gyrans EP00304-9622
HAPTOPHYTA — CLADE 2
Palpitomonas bilix EP00297-9813
CRYPTISTA
Tetrahymena thermophila XP_001471150.1
Pseudocohnilembus persalinus KRW98745.1
Oxytricha trifallax EJY85365.1
Stentor coeruleus OMJ73572.1
ALVEOLATA — CILIATA
Rhodelphis limneticus EP00735-6805
Rhodelphis marinus EP00736-2504
RHODOPHYTA
Vitrella brassicaformis CEM10756.1
Perkinsus marinus EER19910.1
Eimeria acervulina CDI76042.1
Toxoplasma gondii SFA2 TGG_7332
Chromera velia Cvel_12806.t1-p1 / CVEL030024
Plasmodium knowlesi SFA2 XP_002260683.1
Plasmodium falciparum SFA2 XP_001348485.1
ALVEOLATA — SFA2
Cryptosporidium muris EEA05934.1
Cryptosporidium parvum XP_627703.1
Symbiodinium microadriaticum OLP83447.1
Vitrella brassicaformis CEL98015.1
Perkinsus marinus XP_002764526.1
Chromera velia Cvel_6796.t1-p1 / CVEL018337
Babesia bigemina CDR95250.1
Babesia bovis XP_001609905.1
Theileria annulata XP_952869.1
Plasmodium knowlesi SFA1 XP_038969385.1
Plasmodium falciparum SFA1 XP_001349482.1
Toxoplasma gondii SFA1 TGG_7042
Toxoplasma gondii SFA3 TGG_9402
ALVEOLATA — SFA1/3
Mataza sp. D1 EP00462-763
Bigelowiella natans JGI_V11_134269
Lapot gusevi EP00455-41270
Plasmodiophora brassicae EP00473-1446
Collozoum sp. 1-RS-2012 EP00495-7242
Reticulomyxa filosa EP00480-23070
RHIZARIA
Nannochloropsis gaditana Nga00374 / NGAD000371
Cafeteria roenbergensis KAA0154653.1
Fragilariopsis cylindrus EP00528-10591
Phaeodactylum tricornutum XP_002185486.1
Thalassiosira pseudonana JGI-23177
Aphanomyces astaci ETV83275.1
Saprolegnia parasitica KDO29307.1
Pythium ultimum PYU1_T009530 / PULT000575
Phytophthora ramorum JGI-81943
Plasmopara halstedii CEG48322.1
Phytophthora infestans Broad-PITG_01027.1
Cladosiphon okamuranus Cok_S_s298_16366.t1 / COKA017692
Ectocarpus siliculosus Ec-04 D00800 7 ESIL002371
STRAMENOPILA
Cyanophora paradoxa EP00741-14292
Gloeochaete wittrockiana EP00276-5312
GLAUCOPHYTA
Ostreococcus lucimarinus XP_001418968.1
Klebsormidium flaccidum kfl00570_0060_v1.1 / KFLA0T2908
Micromonas sp. RCC299 27396808_peptide / MSPE000217
Monoraphidium neglectum XP_013900702.1
Chlamydomonas reinhardtii SFA AAB03789.1
Volvox carteri JGI-84701
Chlamydomonas moewusii SFA AAB03788.1
Dunaliella bioculata SFA P54214.1
Spermatozopsis similis SFA P55925.1
CHLOROPLASTIDA
0.5


Giardia intestinalis GL50803_9305 t26_1-p1
Spironucleus salmonicida EP00702-779
Trepomonas sp. PC1 EP00703-3469
Dysnectes brevis EP00768-1712
Chilomastix cuspidata EP00767-1907
Aduncisulcus paluster EP00764-4014
Aduncisulcus paluster EP00764-1681
FORNICATA
Stylonychia lemnae Contig10456.g11163
Oxytricha trifallax EJY84356.1
Stentor coeruleus OMJ68201.1
Stentor coeruleus OMJ95775.1
Paramecium tetraurelia SFA2 CAK87315.1
Paramecium tetraurelia SFA3 CAK89694.1
Paramecium tetraurelia SFA6d CAK79874.1
Paramecium tetraurelia SFA4 CAK72185.1
Paramecium tetraurelia SFA5b CAK91395.1
5
Pseudocohnilembus persalinus KRX00705.1
Tetrahymena thermophila EAR99121.3
Paramecium tetraurelia CAK56279.1
Ichthyophthirius multifiliis EGR27484.1
Tetrahymena thermophila EAR82429.2
Pseudocohnilembus persalinus KRW99914.1
Tetrahymena thermophila XP_001026936.2
Ichthyophthirius multifiliis EGR29237.1
Ichthyophthirius multifiliis EGR28251.1
Paramecium tetraurelia SFA7 XP_001432134.1
Paramecium tetraurelia CAK92101.1
2
Paramecium tetraurelia SFA8b CAK64513.1
Paramecium tetraurelia SFA9 CAK76425.1
Paramecium tetraurelia CAK80531.1
Paramecium tetraurelia SFA10d CAK69719.1
3
Pseudocohnilembus persalinus KRX00065.1
Tetrahymena thermophila EAR96329.1
Ichthyophthirius multifiliis EGR28593.1
Paramecium tetraurelia SFA11b CAK83201.1
Paramecium tetraurelia SFA12d CAK58465.1
Paramecium tetraurelia SFA13d CAK88797.1
4
Paramecium tetraurelia CAK59151.1
Pseudocohnilembus persalinus KRX04033.1
Tetrahymena thermophila EAS06716.1
Ichthyophthirius multifiliis EGR27632.1
Paramecium tetraurelia CAK79688.1
Stylonychia lemnae Contig4675.g4991
Oxytricha trifallax EJY68893.1
Stylonychia lemnae CDW76987.1
Oxytricha trifallax EJY64442.1
Stylonychia lemnae Contig4912.g5251
Oxytricha trifallax EJY83527.1
Stentor coeruleus OMJ70994.1
ALVEOLATA: CILIATA
Symbiodinium microadriaticum OLQ12543.1
ALVEOLATA: DINOFLAGELLATA
Chromera velia Cvel_872.t1-p1
ALVEOLATA: PROTALVEOLATA
Plasmodium falciparum CAX64022.1
Eimeria acervulina CDI82026.1
Toxoplasma gondii TGME49_118760
ALVEOLATA: APICOMPLEXA
Vitrella brassicaformis CEM23209.1
ALVEOLATA: PROTALVEOLATA
Blattamonas nauphoete BlNa180
Monocercomonoides exilis MONOS_10711
Monocercomonoides exilis MONOS_13240
PREAXOSTYLA
Rhodelphis limneticus EP00735-11735
Rhodelphis marinus EP00736-12285
RHODOPHYTA
Paratrimastix pyriformis PaPyr3315
PREAXOSTYLA
Ancoracysta twista EP00164-16104
ANCORACYSTA
"Seculamonas ecuadoriensis" EP00690-1868
DISCOBA
Ancyromonas sigmoides EP00161-20673
Nutomonas longa EP00162-9180
ANCYROMONADIDA
Gefionella okellyi EP00698-17097
Gefionella okellyi EP00698-3323
Malawimonas jakobiformis EP00700-3445
MALAWIMONADIDA
Giardia intestinalis EP00701-4141
Giardia muris EP00763-1130
Spironucleus salmonicida EP00702-6064
Trepomonas sp. PC1 EP00703-263
Spironucleus salmonicida EP00702-3435
Trepomonas sp. PC1 EP00703-2911
Spironucleus salmonicida EP00702-243
Spironucleus salmonicida EP00702-3434
Trepomonas sp. PC1 EP00703-3698
Retortamonas caviae DN117307_c1_g3_i1.p2
Retortamonas dobelli DN78330_c1_g1_i1.p1
Dysnectes brevis EP00768-1413
Chilomastix caulleryi EP00766-158
Kipferlia bialata EP00704-10223
Carpediemonas membranifera EP00765-6857
Carpediemonas membranifera EP00765-2472
Carpediemonas membranifera EP00765-7004
Carpediemonas membranifera EP00765-1369
Aduncisulcus paluster EP00764-14718
Aduncisulcus paluster EP00764-605
Ergobibamus cyprinoides EP00769-992
FORNICATA
Andalucia godoyi EP00762-760
"Jakoba bahamiensis" EP00692-3720
"Seculamonas ecuadoriensis" EP00690-2163
Reclinomonas americana EP00694-6610
DISCOBA
Gefionella okellyi EP00698-3777
MALAWIMONADIDA
Goniomonas avonlea EP00294-9579
CRYPTOMONADA
Choanocystis sp. HF-7 EP00300-25505
CENTROHELIDA
Vitrella brassicaformis CEM09702.1
Chromera velia Cvel_3613.t1-p1
ALVEOLATA: PROTALVEOLATA
Perkinsus marinus EER00240.1
Symbiodinium minutum symbB.v1.2.007143.t1
Symbiodinium microadriaticum OLP83335.1
ALVEOLATA: DINOFLAGELLATA
Aurantiochytrium limacinum 66358 / ALIM002419
Schizochytrium aggregatum EP00515-17337
STRAMENOPILA
Chilomastix caulleryi EP00766-809
Chilomastix cuspidata EP00767-3579
FORNICATA
Gefionella okellyi EP00698-3903
Malawimonas jakobiformis EP00700-2537
MALAWIMONADIDA
Guillardia theta EKX41999.1
Hanusia phi EP00280-8694
Rhodomonas salina EP00289-26688
CRYPTOMONADA
Oxytricha trifallax EJY83781.1
Stylonychia lemnae Contig13004.g13866
Paramecium tetraurelia CAK67399.1
Paramecium tetraurelia SFA1a CAK93379.1
1
Paramecium tetraurelia CAK85689.1
Paramecium tetraurelia CAK68945.1
Ichthyophthirius multifiliis EGR34422.1
Tetrahymena thermophila EAR99652.2
Pseudocohnilembus persalinus KRX08103.1
Paramecium tetraurelia CAK80321.1
Tetrahymena thermophila EAR90114.2
Oxytricha trifallax EJY75132.1
Paramecium tetraurelia CAK56814.1
Paramecium tetraurelia CAK87351.1
Paramecium tetraurelia CAK70121.1
Ichthyophthirius multifiliis EGR32190.1
Tetrahymena thermophila EAS06655.2
Pseudocohnilembus persalinus KRX08333.1
Stentor coeruleus OMJ79311.1
Stentor coeruleus OMJ80238.1
Stentor coeruleus OMJ66752.1
ALVEOLATA: CILIATA
Trimastix marina EP00771-3819
Paratrimastix pyriformis PaPyr9653
Paratrimastix pyriformis PaPyr1826
PREAXOSTYLA
Chrysochromulina tobin KOO32357.1
Haptolina ericina EP00326-7621
HAPTOPHYTA
Rhynchopus humris EP00754-9859
Hemistasia phaeocysticola EP00756-24341
Sulcionema specki EP00755-8320
DISCOBA
Rhodelphis limneticus EP00735-8204
Rhodelphis marinus EP00736-12605
RHODOPHYTA
Ancyromonas sigmoides EP00161-1777
ANCYROMONADIDA
Cyanophora paradoxa EP00741-19398
Gloeochaete wittrockiana EP00276-13393
GLAUCOPHYTA
0.4


SFA P8
EC
'y'
B
CR
iRMT
iRR
oRR
tubulin
AF
fan
LF
SR+EC+LR
AR
RF
'y'
CR
iRMT
oRR
PF
iRR
merge


A
SFA P8
tubulin
merge
fan
RBB
PBB
LBB
EC
AF
AR
B
SR
RR
B
EC
LR
AR
C
iRMTs
iRR
B
'y'
RF
RR
PF
AF
fan
iRR
iRMTs
oRR
B
SR+LR+AR
D
iRMTs
B
'y'
RF
AF
fan
iRR
iRMTs
'y'
SR+LR+AR
oRR
PF
E
iRR
CF
iRMTs
oRR
CR
fan
F
iRMTs
iRR
oRR
CR
G
iRR
SR+LR+AR
oRR
GTS
'y'


Group 1
"Excavata"
Metamonada
Fornicata
Parabasalia
Preaxostyla
Discoba
Amorphea+
Diaphoretickes
Malawimonadida
Ancyromonadida
Collodictyonida
'CRuMs'
Rigifila
Mantamonadida
Amoebozoa
Breviatea
Apusomonadida
Fungi
other Holomycota
other Holozoa
Metazoa
Opisthokonta
Hemimastigophora
Stramenopila
Alveolata
SAR
Rhizaria
Haptophyta
Centrohelida
Provora
Telonemea
Cryptista
Glaucophyta
Rhodophyta
Chloroplastida
Archaeplastida
Group 2a
"Excavata"
Metamonada
Fornicata
Parabasalia
Preaxostyla
Discoba
Amorphea+
Diaphoretickes
Malawimonadida
Ancyromonadida
Collodictyonida
'CRuMs'
Rigifila
Mantamonadida
Amoebozoa
Breviatea
Apusomonadida
Fungi
other Holomycota
other Holozoa
Metazoa
Opisthokonta
Hemimastigophora
Stramenopila
Alveolata
SAR
Rhizaria
Haptophyta
Centrohelida
Provora
Telonemea
Cryptista
Glaucophyta
Rhodophyta
Chloroplastida
Archaeplastida
Group 2b
"Excavata"
Metamonada
Fornicata
Parabasalia
Preaxostyla
Discoba
Amorphea+
Diaphoretickes
Malawimonadida
Ancyromonadida
Collodictyonida
'CRuMs'
Rigifila
Mantamonadida
Amoebozoa
Breviatea
Apusomonadida
Fungi
other Holomycota
other Holozoa
Metazoa
Opisthokonta
Hemimastigophora
Stramenopila
Alveolata
SAR
Rhizaria
Haptophyta
Centrohelida
Provora
Telonemea
Cryptista
Glaucophyta
Rhodophyta
Chloroplastida
Archaeplastida