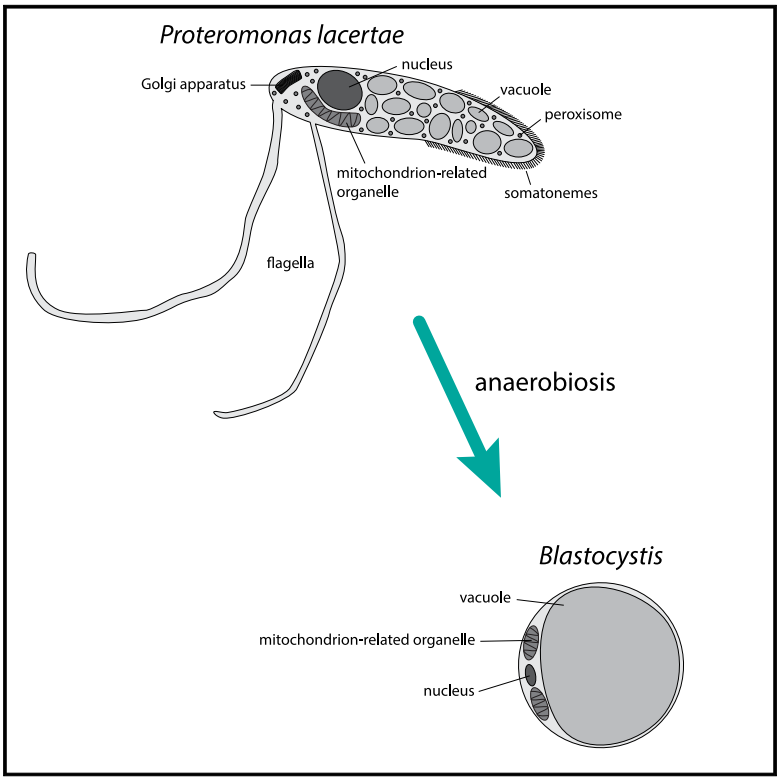
Article

Current Biology

Evolutionary analysis of cellular reduction and anaerobicity in the hyper-prevalent gut microbe *Blastocystis*

Graphical abstract



Authors

Kristína Záhonová, Ross S. Low, Christopher J. Warren, ..., Andrew P. Jackson, Joel B. Dacks, Anastasios D. Tsaousis

Correspondence

dacks@ualberta.ca (J.B.D.), a.tsaousis@kent.ac.uk (A.D.T.)

In brief

The microscopic eukaryote *Blastocystis* is a frequent member of the animal gut microbiome, with altered biology adapted to this niche. Záhonová et al. characterize, at the genomic and cellular level, these adaptations by studying *Proteromonas*, the closest relative of *Blastocystis*. This yields insights about their metabolism and organellar evolution.

Highlights

- A high-quality genome of *Proteromonas* was generated and compared with *Blastocystis*
- The most highly reduced peroxisome to date is reported
- The transition to specialized metabolism in *Blastocystis* is detailed
- A co-evolutionary mechanism for mitochondrion and peroxisome dynamics is proposed

Záhonová et al., 2023, Current Biology *33*, 2449–2464
June 19, 2023
https://doi.org/10.1016/j.cub.2023.05.025

CellPress

Current Biology

CellPress
OPEN ACCESS

Article

Evolutionary analysis of cellular reduction and anaerobicity in the hyper-prevalent gut microbe *Blastocystis*

Kristína Záhonová,[1,2,3,4,17] Ross S. Low,[5,6,17] Christopher J. Warren,[7,17] Diego Cantoni,[7,17] Emily K. Herman,[1,8] Lyto Yiangou,[7] Cláudia A. Ribeiro,[7] Yasinee Phanprasert,[1,9] Ian R. Brown,[7] Sonja Rueckert,[10,11] Nicola L. Baker,[7]

(Author list continued on next page)

[1]Division of Infectious Diseases, Department of Medicine, University of Alberta, 1-124 Clinical Sciences Building, 11350-83 Avenue, Edmonton T6G 2G3, Canada
[2]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Branišovská 1160/31, České Budějovice (Budweis) 370 05, Czech Republic
[3]Department of Parasitology, Faculty of Science, Charles University, BIOCEV, Průmyslová 595, Vestec 252 50, Czech Republic
[4]Life Science Research Centre, Department of Biology and Ecology, Faculty of Science, University of Ostrava, Chittussiho 10, Ostrava 710 00, Czech Republic
[5]Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK
[6]The Earlham Institute, Norwich Research Park, Norwich NR4 7UZ, UK
[7]Laboratory of Molecular & Evolutionary Parasitology, RAPID Group, School of Biosciences, University of Kent, Giles Lane, Stacey Building, Canterbury, Kent CT2 7NJ, UK
[8]Department of Agricultural, Food, and Nutritional Science, Faculty of Agricultural, Life, and Environmental Sciences, University of Alberta, 2-31 General Services Building, Edmonton, AB T6G 2H1, Canada
[9]School of Science, Mae Fah Luang Universit, 333 Moo 1, T. Tasud, Muang District, Chiang Rai 57100, Thailand
[10]School of Applied Sciences, Sighthill Campus, Room 3.B.36, Edinburgh EH11 4BN, Scotland
[11]Faculty of Biology, AG Eukaryotische Mikrobiologie, Universitätsstrasse 5, S05 R04 H83, Essen 45141, Germany
[12]Gut Microbiome Research Group, Mae Fah Luang University, 333 Moo 1, T. Tasud, Muang District, Chiang Rai 57100, Thailand

(Affiliations continued on next page)

SUMMARY

Blastocystis is the most prevalent microbial eukaryote in the human and animal gut, yet its role as commensal or parasite is still under debate. *Blastocystis* has clearly undergone evolutionary adaptation to the gut environment and possesses minimal cellular compartmentalization, reduced anaerobic mitochondria, no flagella, and no reported peroxisomes. To address this poorly understood evolutionary transition, we have taken a multi-disciplinary approach to characterize *Proteromonas lacertae*, the closest canonical stramenopile relative of *Blastocystis*. Genomic data reveal an abundance of unique genes in *P. lacertae* but also reductive evolution of the genomic complement in *Blastocystis*. Comparative genomic analysis sheds light on flagellar evolution, including 37 new candidate components implicated with mastigonemes, the stramenopile morphological hallmark. The *P. lacertae* membrane-trafficking system (MTS) complement is only slightly more canonical than that of *Blastocystis*, but notably, we identified that both organisms encode the complete enigmatic endocytic TSET complex, a first for the entire stramenopile lineage. Investigation also details the modulation of mitochondrial composition and metabolism in both *P. lacertae* and *Blastocystis*. Unexpectedly, we identify in *P. lacertae* the most reduced peroxisome-derived organelle reported to date, which leads us to speculate on a mechanism of constraint guiding the dynamics of peroxisome-mitochondrion reductive evolution on the path to anaerobiosis. Overall, these analyses provide a launching point to investigate organellar evolution and reveal in detail the evolutionary path that *Blastocystis* has taken from a canonical flagellated protist to the hyper-divergent and hyper-prevalent animal and human gut microbe.

INTRODUCTION

First described in humans by Émile Brumpt in 1912,[1] *Blastocystis* is possibly the most prevalent eukaryote colonizing the human gut. It is found in up to 100% of individuals in some populations,[2] and it is estimated that at least one out of every six humans worldwide could be carrying this organism[3]; its prevalence in some groups of animals could be much higher.[4]

Blastocystis is an obligate symbiont that has long been suspected of being a potential pathogen, but evidence for this is contradictory.[5] Indeed, it has been suggested recently that *Blastocystis* could, in fact, be a marker for a healthy human gut.[6] *Blastocystis* is also found in many other animal hosts, raising the potential for zoonotic transmission. One factor that may contribute to the uncertainty over both *Blastocystis* pathogenicity and importance of zoonotic transmission is its genetic

Jan Tachezy,[3] Emma L. Betts,[7,10] Eleni Gentekaki,[9,12] Mark van der Giezen,[13,14] C. Graham Clark,[15] Andrew P. Jackson,[5] Joel B. Dacks,[1,2,16,18,*] and Anastasios D. Tsaousis[7,19,20,21,*]

[13]Department of Chemistry, Bioscience and Environmental Engineering, University of Stavanger Richard Johnsens Gate 4, 4021 Stavanger, Norway
[14]Biosciences, University of Exeter, Stocker Road, Exeter EX4 4QD, UK
[15]Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK
[16]Centre for Life's Origin and Evolution, Division of Biosciences, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK
[17]These authors contributed equally
[18]Twitter: @DacksLab1
[19]Twitter: @TsaousisLab
[20]Twitter: @ADTsaousis
[21]Lead contact
*Correspondence: dacks@ualberta.ca (J.B.D.), a.tsaousis@kent.ac.uk (A.D.T.)
https://doi.org/10.1016/j.cub.2023.05.025

diversity. Despite sharing morphological identity, humans have been demonstrated to harbor at least 12 distinct variants, known as subtypes (STs), four of which are common.[5,7–9] These STs differ substantially from each other, with orthologous proteins sharing only 60% amino acid identity on average and genome comparisons revealing substantial differences in gene number.[10]

At the time of description, the taxonomic affinities of *Blastocystis* were unclear and remained so for over 80 years. In the intervening period, it was variously described as a yeast, a sporozoan (older taxonomic term for the apicomplexans), and a flagellate cyst, among others. This lack of clarity was due in great part to the absence of useful morphological characters: spheres 5–10 μm in diameter, *Blastocystis* can be said to resemble soap bubbles or frog-spawn. As a result, it was usually listed as *Incertae sedis* in taxonomic schemes. It was only when DNA sequences became available that definitive links to other organisms could be made. When its relatives were finally identified, they were completely unexpected: *Blastocystis* is a member of the Stramenopila,[11] the lineage containing diatoms, large multicellular kelps, and relatively much smaller, heterotrophic biflagellates. The latter are considered to be the prototypical stramenopile morphology.[12]

In order to understand the evolution of *Blastocystis* from a small biflagellated ancestor into today's non-flagellated intestinal symbiont, with the accompanying morphological simplification and metabolic adaptations, a suitable outgroup organism is needed. It has been clear since the first molecular identification of *Blastocystis* as a stramenopile that it is specifically related to the Slopalinida, a group of intestinal symbionts of various hosts comprising the Opalinidae and Proteromonadidae. The Slopalinida and *Blastocystis* together comprise the Opalinata.[13,14] One of the genera in the Slopalinida is *Proteromonas*, small biflagellate cells found commonly in the lower large bowel of amphibians and reptiles, and occasionally in rodents.[14] The best-studied species is *Proteromonas lacertae*, which has typical stramenopile appearance. It is 15–20 μm in length, with a pyriform cell body, two flagella (but without the characteristic mastigonemes[15]), but with somatonemes (hair-like structures) on the posterior half of the cell.[14,16,17] Not only is the external appearance nothing like *Blastocystis*, but the internal cell structure also differs dramatically. *P. lacertae* has a single nucleus in close contact with a single large mitochondrion. The flagellar rootlet (rhizoplast) passes through the Golgi apparatus and a groove in the nucleus, ending on the mitochondrion.[14,16] This results in all these organelles being located near the base of the flagellum and close to the apical end of the cell. Both organisms have mitochondria with tubular cristae and form cysts with a single nucleus. However, trophic forms of *Blastocystis* have multiple nuclei and many mitochondrion-related organelles (MROs) distributed around the periphery of the cell together with the other organelles, including the Golgi apparatus.[18]

Because *Proteromonas* is a member of the sister lineage to *Blastocystis*, structurally a more typical stramenopile, and yet is also a member of the anoxic animal intestinal microbiome, we have investigated the genome and cell biology of *P. lacertae* to better understand the reductive evolution and genomic peculiarities accompanying the transformation of *Blastocystis* into the highly divergent but dominant member of the protistan community in the human and animal gut microbiome today.

RESULTS

Microscopical investigations of *P. lacertae*

Since Brugerolle and Bardele,[17] there has not been a comprehensive investigation of *Proteromonas* using modern microscopy techniques. Thus, to better understand the cellular structure of *P. lacertae* as a comparator to *Blastocystis*, we started with microscopical analyses using a combination of fluorescence, transmission, and scanning electron microscopy. Unlike *Blastocystis* (Figure 1A), *P. lacertae* has an elongated cell shape that was confirmed using scanning electron microscopy (Figure 1B). This revealed the two classic heterokont types of flagella (but lacking mastigonemes), the cortical ridges on the cell body, and the somatonemes on the lower end of the cell, consistent with past reports.[17] Anecdotally, a more circular form was also observed (Figure 1Cii), which we speculate could be a cell going through the process of encystation. Transmission electron microscopy confirmed the presence in *P. lacertae* of the various typical eukaryotic organelles (nucleus, endoplasmic reticulum [ER], Golgi apparatus), including a single mitochondrion network surrounding the nucleus (Figures 1Bv and 1Ci). This was further confirmed using MitoTracker labeling and captured using fluorescence microscopy (Figure 1Bi). Given the highly divergent cellular structure of *Blastocystis* compared with the relatively canonical stramenopile form of *P. lacertae*, we decided to undertake an 'omics investigation to understand the cellular evolution of these two organisms.

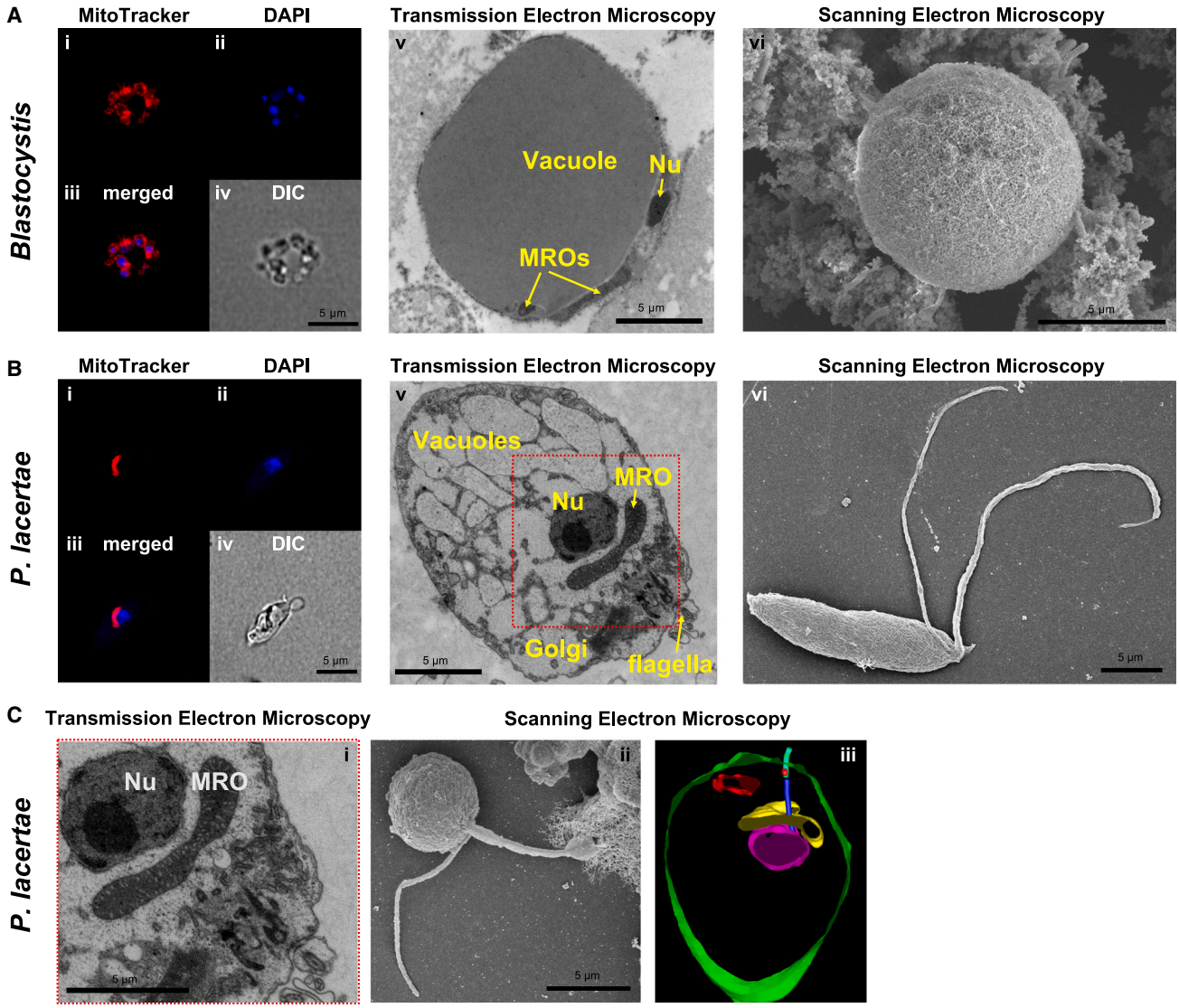


Figure 1. Assorted microscopical observations of *Blastocystis* and *P. lacertae* cells

Assorted microscopical observations of *Blastocystis* (A) and *P. lacertae* (B and C) cells.

(A and B) Confocal microscopy of cells stained with MitoTracker Deep Red (Ai and Bi) labeling the mitochondrion-related organelles (MROs), while DAPI staining (Aii and Bii) labels both the nuclei and the mitochondrial DNA within the MROs (A) or only the nucleus (B). (Aiii and Biii) Merge of the MitoTracker and DAPI staining. (Aiv and Biv) Differential interference contrast (DIC) of cells in the (Ai)–(Aiii) and (Bi)–(Biii) images. (Av and Bv) Transmission electron microscopy (TEM) demonstrating cell shape and labeling the nucleus (Nu) and MROs (M). (Avi and Bvi) Scanning electron microscopy (SEM) of cells showing the overall shape of the organism and the absence (A) or presence (B) of extracellular organelles such as the flagella, which are characteristic to other stramenopiles.

(C) (Ci) A higher magnification of the *P. lacertae* cell shown in (Bv), centering specifically to the relation between the nucleus and the MRO. (Cii) SEM of a *P. lacertae* cell showing a more circular overall shape of the organism, while maintaining the two characteristic flagella. (Ciii) Reconstructed cartoon of a *P. lacertae* cell, which is a result of 12 serial sections of 70 nm thickness visualized under TEM. Green color illustrates the periphery of the cell, purple the nucleus, yellow the mitochondrion, blue the flagella, and red the Golgi apparatus.

The *P. lacertae* genome is substantially larger than that of *Blastocystis*

The *P. lacertae* LA genome was assembled into a 52.3 Mb sequence consisting of 1,449 contigs with a maximum contig length of 864,525 bp and N50 value of 92,586 bp (Table 1). Genome annotation contains 35,706 gene models, of which 28,067 were supported by transcriptomic data. The gene set returned a BUSCO score of 85.6%, comparable with the best-annotated *Blastocystis* genomes[10,19,20] (Table 1). We also produced a transcriptome for the marine stramenopile *Cafeteria burkhardae* (formerly *roenbergensis*[21]) as a genuinely free-living outgroup in comparative analyses; if a gene presents a reciprocal best match in the *C. burkhardae* transcriptome, this provides a means of confirming a definite *Blastocystis* loss if that gene is absent in *Blastocystis* but present in *P. lacertae*. The final transcript set of *C. burkhardae* consisted of 28,952 transcripts after decontamination, with a BUSCO score of 70.4%. The *P. lacertae* genome sequence has higher gene density (684 genes/Mb) and thus

Table 1. Assembly and genome statistics for *P. lacertae* LA, *Blastocystis* ST1, *Blastocystis* ST4, and *Blastocystis* ST7

	Proteromonas lacertae LA	*Blastocystis* ST1	*Blastocystis* ST4	*Blastocystis* ST7
Scaffolds	1,449	580	1,301	54
Contigs	1,449	1,092	1,355	155
N50	92,586	36,659	29,524	296,810
Genome size (Mb)	52.25	16.40	12.92	18.82
Gene number	35,706	6,544	5,707	6,020
GC (%)	27.1	53.0	39.7	45.2
Average coverage	54.3	80.0	300.0	12.4
BUSCO (%)	85.55	85.08	81.81	78.79
Total gene length (Mb)	33.1	11.6	7.9	7.8
Total gene length (%)	63	70	61	41
Gene density (genes/Mb)	684	399	442	320
Intergenic length (Mb)	19.1	4.8	5	11

shorter average intergenic length (534.9 bp) compared with *Blastocystis*. No evidence was found for the widespread segmental duplications that are a distinctive feature of the *Blastocystis* genome.[10] Similarly, we did not observe the insertion of premature stop codons upon poly-adenylation of *P. lacertae* transcripts, as observed in *Blastocystis*.[22] Overall, the *P. lacertae* genome is roughly three to five times larger than a *Blastocystis* genome and contains around six times as many genes.

Blastocystis genomes are reduced in gene number and diversity relative to other stramenopiles

The observed divergence in coding content could be due to gene gain in *P. lacertae*, gene loss in *Blastocystis*, or a combination thereof. To explore this, OrthoMCL was used to cluster the predicted proteomes of *Blastocystis* ST4 and ST7, *P. lacertae*, *C. burkhardae*, and a selection of other stramenopiles. To prevent poor gene models from influencing the clustering analysis, only genes with transcript support in *P. lacertae* were included. This reduced the number of genes to 28,067, with a BUSCO score of 75.1%. This approach established the phylogenetic distribution of gene clusters, distinguishing those that were species-specific (i.e., gene gains) from widely conserved genes that were absent in specific species (i.e., gene losses). OrthoMCL assigned 122,818 predicted protein sequences from 10 genomes or transcriptomes into 24,945 orthogroups, with an additional 52,806 sequences that did not cluster (see Table S1).

Of these clusters, 2,363 containing 13,585 sequences were unique to *P. lacertae*. In addition, 6,932 *P. lacertae* sequences did not cluster and were considered also species-specific, based on the current sampling. This analysis shows that up to 73% of the high-confidence *P. lacertae* gene set possessing transcriptomic support is species-specific and, therefore, suggests that much of the genome size discrepancy comes from unique genes in *P. lacertae*.

Nonetheless, besides these gains, gene loss has also contributed to the divergence between *Blastocystis* and *P. lacertae*. Figure 2A compares the proportions of all cluster types in *Blastocystis* ST7 and *P. lacertae*. For each organism, the genes missing from their genomes relative to the other (described as "sister-lineage losses" and listed in Table S1, sheets D–F for *Blastocystis* and sheets G–I for *P. lacertae*) are expressed as a proportion of their current gene number to reflect the scale of gene loss relative to the ancestor. After excluding species-specific gains, 3,161 *P. lacertae* genes were found to be missing from the *Blastocystis* ST7 genome (i.e., combined entries in Tables S1D–S1F, shown as indigo segment in Figure 2A); this is 52.5% of the total *Blastocystis* gene number (n = 6,020), or 93.7% of conserved *Blastocystis* genes (n = 3,372). Hence, almost as many ancestral genes have been lost from *Blastocystis* as retained; conversely, only 545 conserved genes are missing from *P. lacertae*, which is 1.9% of its high-confidence gene set that we are using here (n = 28,067) or 3.8% of conserved *P. lacertae* genes (n = 14,482). Thus, since sharing an ancestor, *Blastocystis* has lost many more conserved genes than *P. lacertae*, but gained far fewer new genes, resulting in a much greater reduction relative to its ancestral state.

To explore the functional consequences of these distinct evolutionary histories, we examined the functional terms associated with gene losses, identifying KEGG ontology (KO) and gene ontology (GO) terms that are significantly over- or under-represented relative to their frequency across the whole genome (Table S2). Figure 2A shows significant KO terms for gene losses, alongside conserved genes for comparison. Although conserved gene sets are naturally enriched for terms associated with core cell function, such as "ribosome" (KO03011) and "transcription" (KO03021), such terms are under-represented among *Blastocystis* gene losses ($p < 0.01$). Conversely, sequences associated with "chromosome and associated proteins" (KO03036) and "cytoskeleton proteins" (KO04812) are over-represented among *Blastocystis* gene losses ($p < 0.0001$) but not *P. lacertae* losses. KO04812 is associated with 13 gene losses, including intraflagellar transport proteins, dyneins, and kinesins. This association of gene loss with the motile cytoskeleton is also reflected among over-represented GO terms, the most significant of which is "cilium" (GO: 0044441; $p < 0.0001$) and "ATP-dependent microtubule motor activity" (GO: 1990939; $p < 0.0001$) (Table S2B).

Consequently, while the obvious disparity in genome size is mainly due to considerable *P. lacertae* gene gains, there is a definite asymmetry between the species in gene loss. This reflects a substantial loss of conserved gene functions in *Blastocystis*, potentially in line with its more simplified morphology and biochemical adaptations to the anaerobic/gut environment.

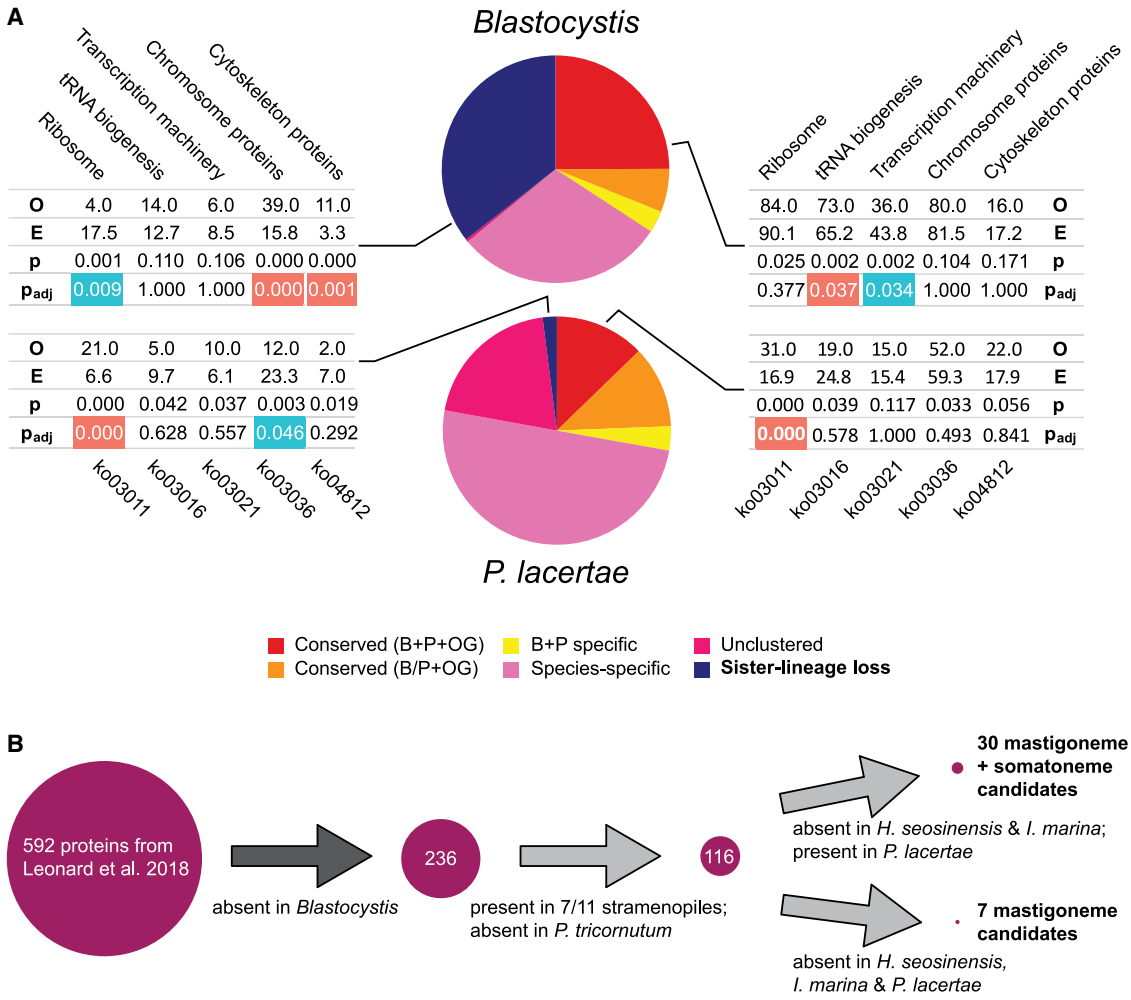


Figure 2. Comparative genomic analyses of orthogroups and flagellar proteins

(A) Comparison of gene clustering for *P. lacertae* and *Blastocystis*. Predicted gene sets for *P. lacertae* (P; n = 26,100) and *Blastocystis* ST7 (B; n = 6,020) were each clustered using OrthoMCL. The pie charts show the proportion of genes falling into six categories. Gene clusters that were present in both B and P, as well as stramenopile outgroups (OGs), were categorized as "conserved," as were clusters present in B or P (as appropriate) and OG. Clusters found in both B and P but not OG, as well as species-specific clusters and unclustered genes (assumed also to be species-specific), are also shown. These five categories cover all genes found in the genome. The sixth category, "sister-lineage loss," is shown in the same pie chart to emphasize the scale of gene loss relative to contemporary genome size. This category includes those genes assumed to be lost from the *P. lacertae* or *Blastocystis* ST7 genome since their lineage separation. For example, the *P. lacertae* genome contains 3,161 genes that are conserved in other stramenopiles but absent from *Blastocystis*, and so assumed to have been lost from *Blastocystis* after separating from *P. lacertae*. When combined with the contemporary *Blastocystis* gene set, these losses are 36% of all genes, and 51% when *Blastocystis*-specific genes are excluded. Five KEGG orthology (KO) terms that are significantly enriched among conserved genes (right) or sister-lineage losses (left) in each organism are tabulated besides the pie charts. For each KO term, a hypergeometric test assesses the significance of the difference between the observed (O) and expected (E) incidences, with a p value adjusted for multiple tests using Bonferroni correction. Terms that are over-represented relative to their genomic frequency are shaded red, while under-represented terms are shaded blue.

(B) Flow chart of stepwise homology searching for flagellar-associated proteins. This shows the datasets (circles with protein numbers), homology searching analyses (dark arrow), and filtering of results (light arrows).

See also Figure S1B and Tables S1, S2, and S4A–S4F.

Analysis of a large-scale flagellar protein dataset identifies candidate mastigoneme proteins and supports mastigoneme-somatoneme homology

Enrichment of cytoskeleton-related GO terms among *Blastocystis* gene losses points to evolutionary changes associated with motility. Although there is both morphological and genomic variation, the flagellum and its associated molecular machinery are conserved across stramenopiles.[23,24] Our scanning electron microscopy shows well-developed flagella in *P. lacertae* (Figure 1Bvi), but no indication of flagella in *Blastocystis* (Figure 1Avi). Indeed, though different morphological forms of *Blastocystis* have been described with varying degrees of confidence, no flagellated stage has ever been reported, suggesting that flagellar motility has also been lost in this lineage. Our analysis of 16 proteins found previously[25] to be constant in flagellated organisms but absent in all non-flagellated organisms, found no credible candidates in *Blastocystis* but presence of 13/16 in *P. lacertae* (Figures S1A and S2; Table S3A). These data further increased our confidence that

Blastocystis truly lacks flagella, speculatively as a result of the specialization of living in the gut and due to the fecal-oral transmission mechanism. Importantly, we concluded that *Blastocystis* can be used as a *de facto* negative control for downstream analyses of stramenopile flagellar evolution.

Stramenopiles were defined by the possession of tripartite hairs or mastigonemes (i.e., tinselation) on their posterior flagellum.[26] Despite this synapomorphy, there is a range of flagellar states within the group. The loss of flagella has taken place on at least two occasions, once in *Blastocystis* and once in pennate diatoms.[27] Moreover, there are a few taxa possessing flagella but lacking tinselation. The protein composition of the mastigonemes is poorly understood, with only three proteins having been localized to the structure.[28,29] Their exclusivity for this feature, and whether there are remaining components, is unclear. Notably, *P. lacertae* lacks tinselated flagella but its somatonemes have been proposed as homologous.[26] The *P. lacertae* genome thus provides a unique opportunity to identify candidate mastigoneme proteins and to assess this homology argument.

We performed a series of comparative genome analyses to identify a core set of flagellar- and mastigoneme-correlated proteins (Figure 2B; Tables S4A–S4F). From a set of 592 proteins previously used for investigating stramenopile flagellar evolution,[30] a reduced set of 236 was first identified using *Blastocystis* as a negative filter to remove proteins that have promiscuous or additional non-flagellar functions. This set was then searched against diverse stramenopile genomes or transcriptomes chosen to represent the canonical tinselated state (11 taxa), the non-tinselated state (*Incisomonas marina*[31] and *Halocafeteria seosinensis*[32]), non-flagellated state (*Blastocystis* and *Phaeodactylum tricornutum*), or somatonemal state (*P. lacertae*). We found 116 proteins in the majority of the flagellated taxa but absent in both *Blastocystis* and *P. tricornutum*, thus being more confidently flagella-associated. Of these proteins, 37 were found to be widely present but missing in *I. marina* and *H. seosinensis*,[32] hence representing mastigoneme-associated candidates that warrant molecular biological investigation. Notably, 30/37 were found to be present in *P. lacertae*, consistent with homology between somatonemes and mastigonemes.

Both *Blastocystis* and *P. lacertae* possess a conserved membrane-trafficking system

Several GTPases in the Rab family have flagella-associated function, including IFT27, Rab23, Rab28, and RABL2.[33] A comprehensive molecular evolutionary analysis of the Rab complements across *P. lacertae*, *Blastocystis*, and a selection of stramenopile genomes was undertaken. Altogether we identified 40 Rab sequences from *P. lacertae* and assigned these to specific Rab subfamilies (Figure S1B; Data S1), allowing us to deduce Rab complement in the last stramenopile common ancestor (LSCA) and providing context for presence and absence of components in *Blastocystis*. The LSCA is deduced not to have possessed five Rab proteins present in the last eukaryotic common ancestor (LECA)[34]: Rab4, Rab14, Rab20, Rab24, and Rab34 (Figure S1B). Notably, however, *P. lacertae* and other stramenopiles encode the flagella-associated RABs, i.e., Rab23, Rab28, RABL2, and IFT27, while these are not found in *Blastocystis*.[10]

The difference in the Rab complement between *Blastocystis* and *P. lacertae* raised questions regarding the conservation of other machinery in the membrane-trafficking system (MTS). Membrane-trafficking is critical for basic cellular function in eukaryotes and is important for pathogenic mechanisms in diverse protistan parasites, being responsible for the movement of cellular material between organelles, as well as import and export of material from the cell, and cell surface modulation[35] (inter alia).

Comparative genomics and phylogenetics were used to identify and classify the membrane-trafficking machinery of *P. lacertae*. Overall, the *P. lacertae* genome encodes a relatively complete MTS, similar to that of free-living eukaryotes (Figure S3; Table S3B). Notably, *P. lacertae* encodes the complete TSET complex and, using *P. lacertae* proteins as queries, we were able to find orthologs in *Blastocystis* (Data S2). This is the first instance of complete TSET complexes in stramenopiles, suggesting it was likely present in the LSCA. It also means that this complex can play a role in membrane-trafficking in *Blastocystis*, in contrast to previous reports.[10] Given its role as a primary modulator of clathrin mediated endocytosis in plants,[36] the presence of this complex has potentially important implications for understanding the cell biology of material uptake from the host in *Blastocystis*.

Furthermore, examination of the multi-subunit tethering complex complement identified three of the four proteins of the Dsl1 complex in *P. lacertae* but confirmed identification of only a single Dsl1 complex subunit in the majority of *Blastocystis* STs (Figure S3; Table S3B). In yeast, Dsl1 functions at the ER for vesicle tethering, but also peroxisome assembly.[37] Notably, paucity of Dsl1 complex components is correlated with the loss or modification of peroxisomes[38] and, consistent with this, peroxisomes have not been visualized in *Blastocystis* nor have any of the peroxin proteins (Pex), which are involved in peroxisome proliferation and assembly, been identified in the *Blastocystis* genomes.[10] The identification of Dsl1 machinery in *P. lacertae* raises the possibility that a peroxisomal organelle is present in this organism.

P. lacertae possesses the most rudimentary peroxisome ever reported

To test this possibility, we searched for Pex orthologs in the *P. lacertae* genome, identifying homologs for the peroxisomal membrane E3 ubiquitin ligase Pex10 and the farnesylated receptor of peroxisomal membrane proteins (PMPs) Pex19, as well as a possible homolog of the ubiquitin-conjugating protein Pex4 (Figure 3A; Table S3C). Notably, while Pex10 and 19 are considered to be unequivocal informatic markers of peroxisomes,[39] we did not identify any other Pex proteins. Consistent with the lack of Pex proteins comprising the peroxisomal targeting signals 1 (PTS1) and 2 (PTS2) recognition machinery in *P. lacertae* (Figure 3A; Table S3C), our searches for proteins bearing these targeting signals failed to identify any of the known peroxisome matrix proteins normally targeted by those methods. Although we did identify 126 and two proteins harboring PTS1 and PTS2 motifs, respectively (Table S5), these either had no known hits in the database (90 PTS1 and both PTS2) or were attributed to a variety of other cellular functions, and so we anticipate that the *P. lacertae* peroxisome does not function via proteins in its matrix. By contrast, examination of diverse additional stramenopiles revealed a relatively complete complement of peroxins. The

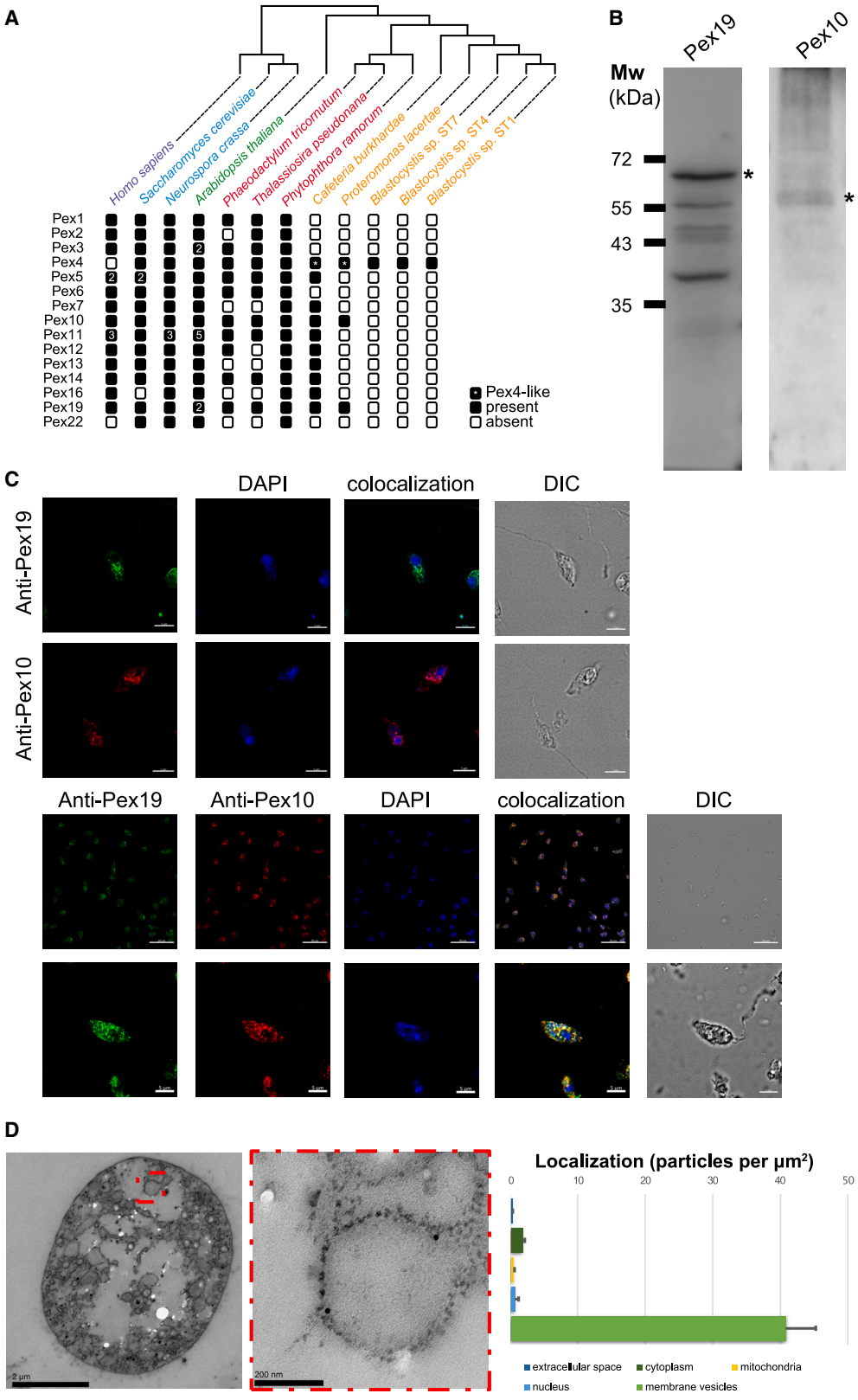


Figure 3. Bioinformatic and molecular cell biological evidence for a minimal peroxisome in *P. lacertae*

(A) Complement of Pex proteins in representative eukaryotes and selection of stramenopiles. Numbers in squares indicate counts of paralogous sequences.
(B) Immunoblotting of anti-Pex19 (rabbit) and anti-Pex10 (rat) antibodies showing a strong band at 68 and 59 kDa, respectively (marked by asterisks).

(legend continued on next page)

Pseudofungi encode nearly all of the examined proteins, but even considering taxa outside of the Bigyra and *Aureococcus anophagefferens*, which appear to have degenerated their complement independently, we found an average of 14.8 of 17 examined proteins encoded, suggesting that the LSCA possessed a full peroxin set, consistent with previous data (Figures 3A and S4; Table S4G). However, we noted that within the Bigyra we could trace the losses of several peroxins upstream of *P. lacertae* and *Blastocystis*. Pex22 was not found in any of the bigyran taxa, while progressively smaller complements were seen within the opalozoan taxa examined (Figure S4; Table S4G).

Most notably, in model systems (e.g., yeast and mammals), Pex3 (Pex16) and Pex19 function together for the incorporation of PMPs such as Pex10 into the peroxisomal membrane. Although Pex10 and Pex19 were confidently reported in *P. lacertae* and *C. burkhardae*, we mapped the loss of Pex3 to the base of the Opalozoa within Stramenopila (Figure S4), with the caveat that Pex3 may display a low degree of conservation and its identification can be problematic using bioinformatics tools.[40] Nonetheless, examining the primary sequence conservation of the established Pex3 and Pex10 binding regions of Pex19 (Tables S6A and S6B), we found that the *P. lacertae* and *C. burkhardae* Pex3 binding regions of Pex19 are less conserved (9.5% and 12.2% average identity, respectively) compared with those orthologs from organisms possessing Pex3 (14.8%). By contrast, the Pex10 binding regions of Pex19 in *P. lacertae* and *C. burkhardae* are slightly better conserved (26.2% and 29.7% average identity, respectively) than when compared among Pex3-possessing taxa (22.5%). This is consistent with a degeneration of the Pex3- binding region but conservation of the Pex10-binding region in Pex19-possessing organisms that have lost Pex3. Overall, the *P. lacertae* Pex complement is minimal, but does suggest the presence of a peroxisome-derived organelle.

Because the *in silico* analysis suggested the possible presence of a minimal peroxisomal organelle, we used a multiphasic approach. Western blotting using two heterologous anti-Pex19 antibodies[41] revealed a cross-reacting band at ~68 kDa in *P. lacertae* protein extracts (Figure 3B) that corresponds to the predicted size of *P. lacertae* Pex19.

Pex3-Pex19 binding is the best-established mechanism of Pex19 membrane-association. However, an alternate mechanism has been proposed, whereby Pex19 is targeted to membranes via a C-terminal farnesylation.[42,43] Using the program GPS-Lipid,[44] we identified putative farnesylation motifs (CAAX) in the *P. lacertae* Pex19, which are conserved in the majority of the Pex19 orthologs (Table S6C), meaning that Pex19 retains the capacity to interact with PMPs.[42] Therefore, we used confocal microscopy to localize the binding locations of anti-Pex19 (Abcam) in *P. lacertae* cells, which revealed punctate localization within the cell and no co-localization with the nucleus (Figure 3C). To increase resolution of this localization, we employed immuno-electron microscopy using the same antibody. This mainly resulted in gold particles localized in the periphery of single-membrane-bound bodies, which is consistent with a peroxisomal localization (Figure 3D). To confirm these observations, we raised a Pex10 antibody against a specific peptide of the predicted protein. Western blotting and immunofluorescence microscopy demonstrated the specificity of the antibody (clear band at 59 kDa), a punctate localization, and co-localization with Pex19, consistent with the protein being present in the same organelle (Figures 3B and 3C; Video S1, 0:00–0:50 min). These data suggest the presence of a peroxisomal organelle in *P. lacertae*, but also raise further questions regarding its function.

The two hallmark activities of peroxisomes are β-oxidation of fatty acids (FAs) and metabolism of reactive oxygen species (ROS).[45] Therefore, we searched the predicted proteome and genome of *P. lacertae* for these enzymes and compared the same pathways of *C. burkhardae* and *Blastocystis*. Consistent with the absence of PTS-recognizing peroxins in *P. lacertae* and peroxisomes in *Blastocystis*,[10] we did not identify any of the peroxisomally targeted FA β-oxidation enzymes in these two species (Figure 4A). On the other hand, we reconstructed the full pathway in *C. burkhardae*. Surprisingly, the mitochondrial β-oxidation of FA is also incomplete in *P. lacertae* and *Blastocystis*, with 3-hydroxyacyl-CoA dehydrogenase, 3-ketoacyl-CoA thiolase, and trifunctional protein missing in the former and 3-ketoacyl-CoA thiolase and trifunctional protein missing in the latter, while mitochondrial β-oxidation is still operating in *C. burkhardae* (Figure 4A).

There are several ROS metabolizing enzymes known from different cell compartments. The best-known peroxisomal ROS metabolizing enzyme is catalase. However, catalase was shown to be missing in the *Blastocystis* genome,[10,19] and we further did not identify this protein in *P. lacertae* or *C. burkhardae*. Superoxide dismutase was identified in *P. lacertae* and *C. burkhardae*, yet both were predicted to be mitochondrion-localized. Peroxiredoxin, an additional enzyme able to reduce H_2O_2, is similarly predicted to operate only in the *C. burkhardae* mitochondria (Table S6D). PXMP2 and PXMP4 (Figure 4A), two additional peroxisomal proteins involved in ROS metabolism, are transmembrane and thus targeted to the peroxisome via the Pex19 system. Both are encoded in *C. burkhardae* and, notably, while we were unable to identify homologs of either in *Blastocystis*, we did identify a PXMP2 homolog in *P. lacertae*. A validated antibody was raised against the *P. lacertae* PXMP2 protein (Figures 4B and 4C) and showed co-localization with Pex10 in immunofluorescence microscopy (Figure 4D; Video S1, 0:51–1:40 min). Taken together, the evidence is consistent with a highly reduced peroxisome-derived organelle in *P. lacertae*, with the only known enzyme localized within being PXMP2, which is speculated to be involved in ROS metabolism.

(C) Cellular localization of Pex19 and Pex10 in *P. lacertae* cells by immunofluorescence. Rabbit anti-Pex19 antiserum or rat anti-Pex10 shows a discrete localization in the cells and co-localization of the two. DAPI stains the *P. lacertae* nucleus and mitochondrial DNA. Differential interference contrast (DIC) images show the cells used for immunofluorescence. Scale bar, 5 μm (rows 1, 2, and 4) and 20 μm (row 3).
(D) Localization of Pex19 in *P. lacertae* cell by immuno TEM shows compartmental localization. Densities of labeling in different compartments of *P. lacertae* cells suggest that Pex19 is mainly localized in membrane vesicles. Scale bar, 5 μm and 200 nm (insert).
See also Figure S4; Tables S3C, S4G, S5, and S6; and Video S1 (0:00–0:50).

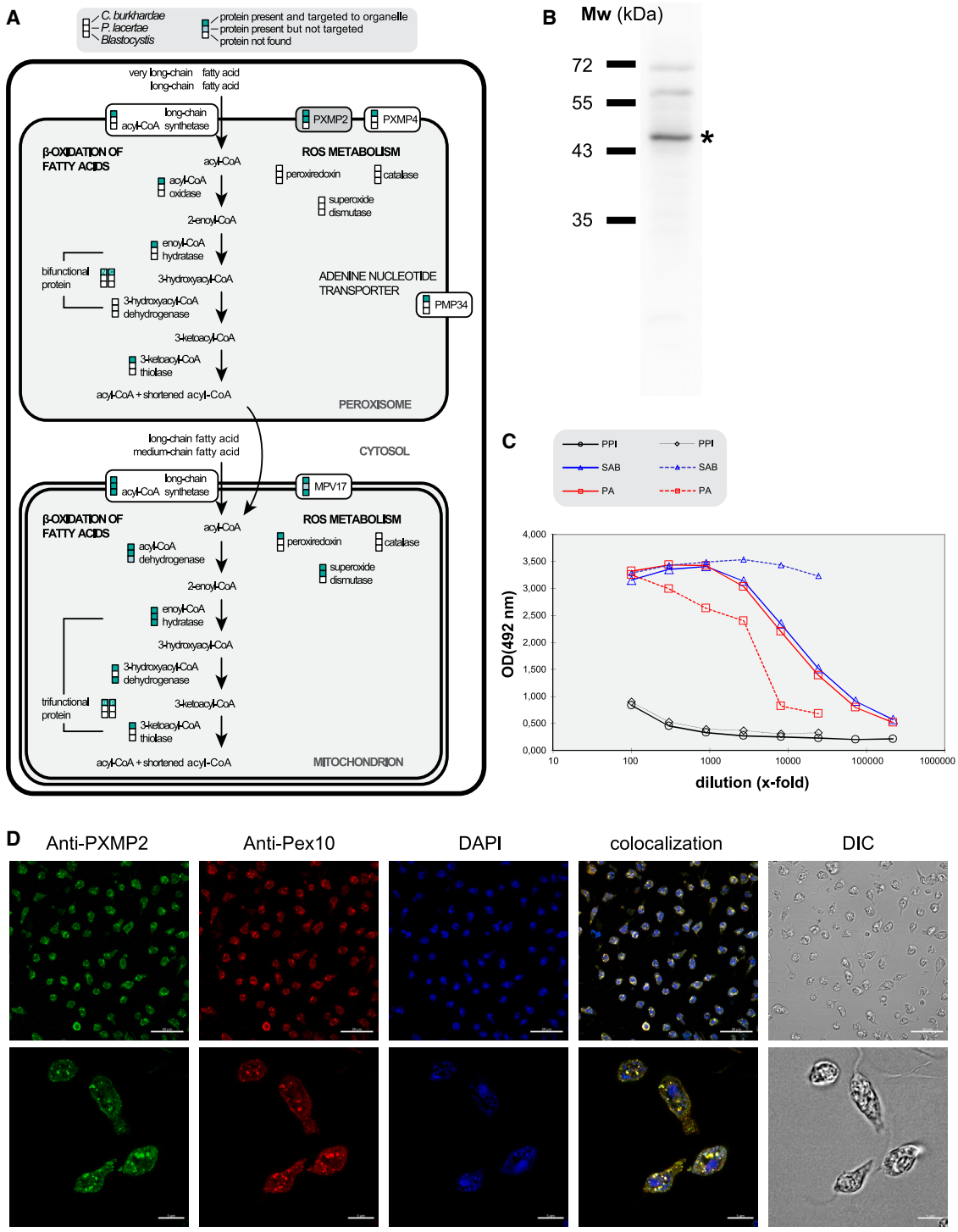


Figure 4. Bioinformatic and molecular cell biological evidence for a potential function in the minimal peroxisome in *P. lacertae*

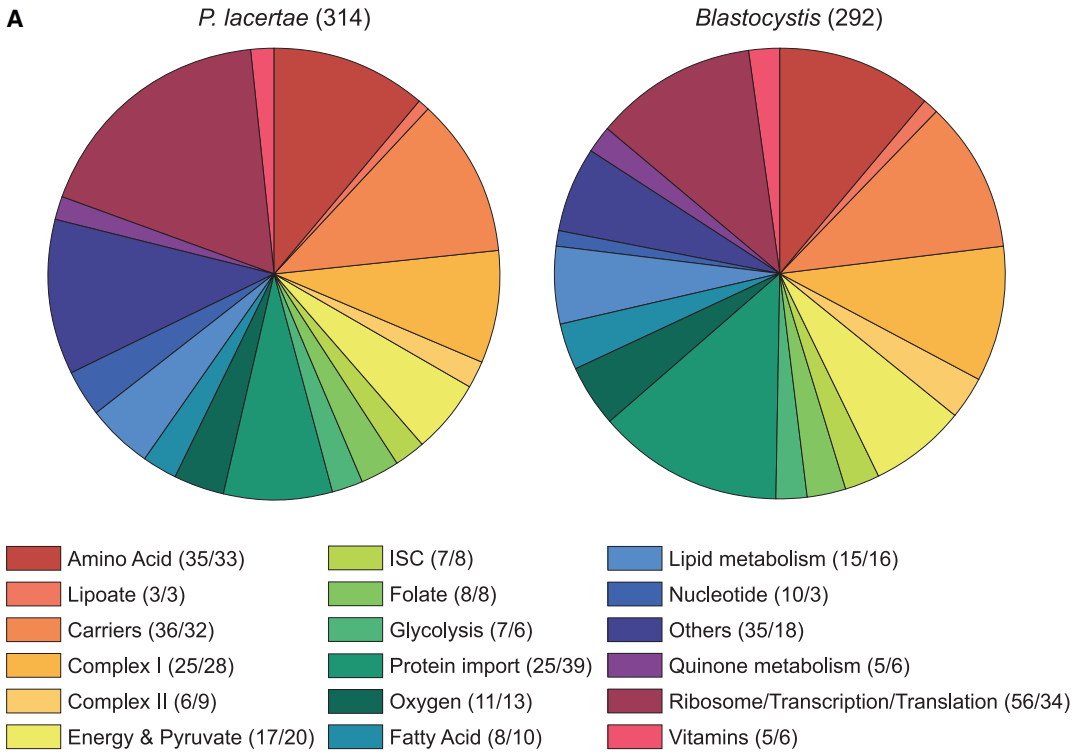
(A) Model of shared FA oxidation and ROS metabolism in the peroxisomes and mitochondria of Opalozoa.

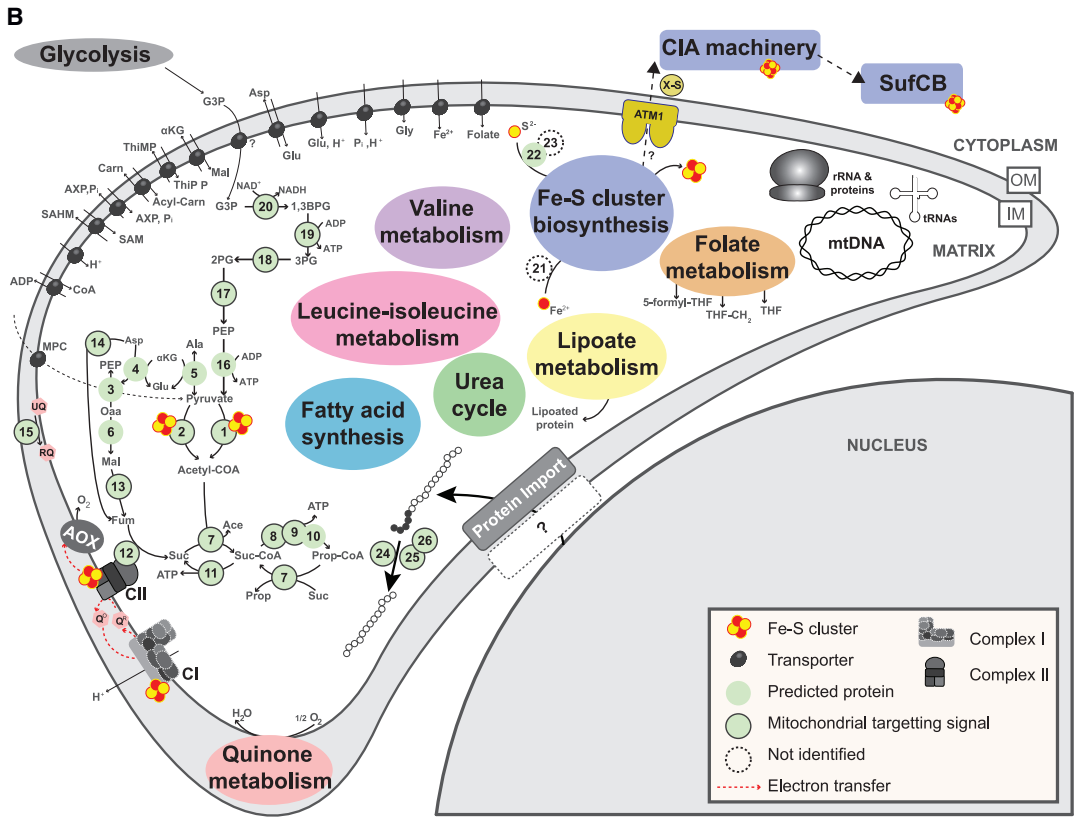
(B) Immunoblotting of anti-PXMP2 (rabbit) antibody showing a strong band at 49 kDa (marked by asterisk).

(C) ELISA curves demonstrating the specificity of the antibody against the peptide which against it was raised. The optical density OD (492 nm) can be correlated to antibody affinity or concentration versus the two animals tested. PPI, pre-immune serum; SAB, final bleed; PAs, purified antibodies.

(D) Cellular localization of PXMP2 and Pex10 in *P. lacertae* cells by immunofluorescence. Rabbit anti-PXMP2 antiserum shows a discrete localization and co-localization with the rat anti-Pex10 in both widefield and higher magnification images. DAPI stains the *P. lacertae* nucleus and mitochondrial DNA. Differential interference contrast (DIC) images show the cells used for immunofluorescence. Scale bar, 20 μm (row 1) and 5 μm (row 2).

See also Tables S3C, S4G, S5, and S6 and Video S1 (0:51–1:40).





(legend on next page)

It has been generally held that anaerobic and parasitic lineages have reduced peroxisomes. The recent description of the anaerobic peroxisomes in Archamoebae (*Entamoeba histolytica*, *Mastigamoeba balamuthi*, and *Pelomyxa schiedti*) are striking counter-examples and raise the possibility of alternate metabolic functions for the organelle in oxygen-shunning organisms.[46–48] Nonetheless, microaerophilic organisms such as *Trichomonas* and *Giardia* most prominently seem to have lost the organelles entirely, as apparently has *Blastocystis*. To our knowledge, the organelle present in *P. lacertae* is the most reduced, but putatively functional, peroxisomal organelle currently described and represents a tremendous opportunity to study a late intermediate stage in the evolutionary degeneration of this organelle in anaerobic lineages. Given the paucity of Pex proteins encoded in the *P. lacertae* genome but the presence of what appears to be a peroxisome-derived organelle, as well as the recent examples of peroxisomes in *Entamoeba* and *Toxoplasma*,[47,49] where the organelle was held not to be present, it may well be worthwhile re-examining some of the other organisms where the peroxisome has been presumed lost.

Blastocystis achieves a comparable metabolism to *P. lacertae*, but with reduced redundancy

Given this minimal peroxisomal complement and the integrally linked nature of this organelle with the mitochondria, we next investigated metabolic pathways with a particular focus on those that are modulated by these two unusual compartments. The comparison showed that *Blastocystis* STs have retained largely similar metabolic capabilities to both *P. lacertae* and *C. burkhardae*, with 291 pathways shared between all *Blastocystis* STs, *P. lacertae*, and *C. burkhardae* (Tables S3D and S6D). The greatest discrepancy comes from the overall number of genes that mapped from each genome. There is a difference of 347 genes between *Blastocystis* ST1 and *P. lacertae*, which appears to be made up of redundant KO terms, albeit with a notable difference in the aspartate biosynthetic pathway. *Blastocystis* encodes the fewest genes of these pathways, which may suggest that *Blastocystis* STs have lost complexity from conserved metabolic pathways without compromising capacity. Despite the difference in genome sizes and the numbers of sequences mapped to KEGG between *Blastocystis* STs and *P. lacertae*, they contain remarkably similar repertoires of pathways. If there is a fundamental difference between them, it is with respect to the number of genes involved in each pathway, the "gene richness" of metabolism. *Blastocystis* seemingly achieves a near comparable metabolic capacity to *P. lacertae*, but with substantially fewer genes (Figure 5A).

The most striking aspect of the metabolic comparison was found in the glycolytic pathway. It was previously shown that stramenopiles partitioned the second half of glycolysis in the mitochondrion.[50] We therefore checked the genome for the enzymes encoding glycolysis in *P. lacertae* to assess the presence or absence of possible mitochondrion-targeted enzymes. Similar to *Blastocystis*,[50] it seems that *P. lacertae* has replaced some ATP-utilizing enzymes for pyrophosphate utilizing ones. *P. lacertae* encodes all ten glycolytic enzymes. As is the case for all other studied stramenopiles,[50] glycolysis is branched and the enzymes for the second half of glycolysis (the C3 part) are located in both the cytosol and mitochondrion. *Blastocystis*, however, has lost the cytosolic branch and solely relies on the mitochondrial C3 branch.[50] Pyruvate kinase does not have a mitochondrial targeting signal in *Blastocystis*, but *Blastocystis* uses the pyrophosphate utilizing alternative to pyruvate kinase, phosphoenolpyruvate synthase (pyruvate dikinase), which does have a mitochondrial targeting signal. *P. lacertae* also seems to use phosphoenolpyruvate synthase, but this enzyme does not seem to have a recognizable mitochondrial targeting signal (Figure 5B) although transcriptomic data suggest a short amino-terminal extension, which could function as targeting signal. Targeting to MROs has been shown to be non-canonical.[51] It is currently not clear how the last step from phosphoenol pyruvate to pyruvate proceeds in *P. lacertae* if the preceding steps are mitochondrial but the last one is not. Unlike *Blastocystis*, but like all other stramenopiles, *P. lacertae* contains a putative mitochondrial pyruvate carrier (MPC).

P. lacertae and *Blastocystis* have comparable MROs and associated aerobic metabolism

Blastocystis has attracted attention for its MROs, which have been studied as an example of an intermediate stage between

Figure 5. Proposed metabolic map and functions of *P. lacertae* mitochondrion-related organelle (MRO) based on the genome predictions

(A) *In silico* predictions of mitochondrial proteins encoded by the *P. lacertae* versus *Blastocystis* (ST1) genomes. Each pie chart shows the percentage of proteins with a given function. Known mitochondrial proteins have been operationally divided into various groups, based on previous categorization in the *Blastocystis* NandII genome.[10] Numbers in parentheses refer to the number of proteins found in each pathway (*P. lacertae*/*Blastocystis*).

(B) Assorted metabolic features of the MRO's role in energy generation and amino acid and lipid metabolism. Protein descriptions (numbers) are outlined below: (1) pyruvate-ferredoxin/flavodoxin oxidoreductase (PFOR); (2) pyruvate-$NADP^+$ oxidoreductase (PNO); (3) phosphoenolpyruvate carboxykinase (ATP); (4) aspartate aminotransferase; (5) alanine aminotransferase; (6) malate dehydrogenase; (7) acetate:succinate CoA transferase; (8) methylmalonyl-CoA mutase; (9) methylmalonyl-CoA epimerase; (10) propionyl-CoA carboxylase alpha subunit; (11) succinyl-CoA synthetase; (12) succinate dehydrogenase subunit 5; (13) fumarase; (14) aspartate ammonia lyase; (15) rhodoquinone biosynthesis enzyme RquA; (16) pyruvate kinase; (17) enolase; (18) phosphoglycerate mutase (PGM); (19) phosphoglycerate kinase (PGK); (20) glyceraldehyde 3-phosphate dehydrogenase (GAPDH); (21) frataxin; (22) cysteine desulphurase (IscS); (23) cysteine desulphurase activator (Isd11); (24) mitochondrial intermediate peptidase (MIPEP); (25) mitochondrial metalloendopeptidase (OMA1); (26) CYM1, MOP112, Cym1p K06972 uncharacterized protein, pitrilysin metallopeptidase 1. Ace, acetate; ACP, acyl carrier protein; aKG, alpha-ketoglutarate; BCD, branched-chain amino acid degradation; 1,3BPG, 1,3-biphosphoglycerate; CI, complex I; CII, complex II; Carn, carnitine; CDP-DAG, cytidine diphosphate diacylglycerol; CL, cardiolipin; DHAP, dihydroxyacetone phosphate; DHOro, dihydroorotate; Fd, ferredoxin; Fum, fumarate; Gly3P, glycerol-3-phosphate; G3P, glyceraldehyde-3-phosphate; Mal, malate; MMC, methyl-malonyl-CoA; Nd(p), NAD(P); mtDNA, mitochondrial DNA; Oaa, oxaloacetate; Oro, orotate; PA, phosphatidic acid; PE, phosphatidylethanolamine; PEP, phosphoenol pyruvate; PI, phosphatidylinositol; 2PG, 2-phosphoglycerate; 3PG, 3-phosphoglycerate; Prop, propionate; PS, phosphatidylserine; QO/R, quinone/quinol, oxidized or reduced; RQ, rhodoquinone; SAHC, S-adenosylhomocysteine; SAM, S-adenosylmethionine; Suc, succinate; THF, tetrahydrofolate; ThiMP, thiamine monophosphate; ThiPP, thiamine pyrophosphate; UQ, ubiquinone; standard amino acid abbreviations are used.

See also Tables S3D, S5, and S6.

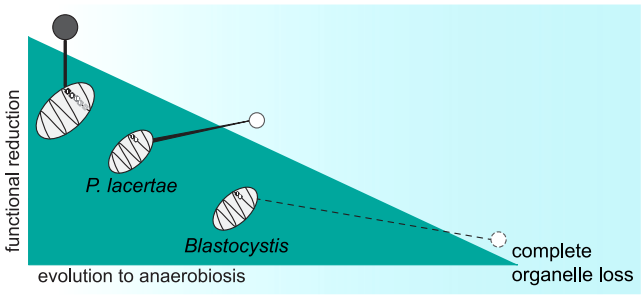


Figure 6. Guiding evolutionary dynamics of peroxisomal and mitochondria organelle reduction

The cartoon illustrates the proposed mechanism of evolutionary contingency that, due to shared pathways of lipid metabolism and defense against reactive oxygen species, whichever organelle starts to degenerate first places a constraint on the reductive evolution of the other until such time that these metabolic requirements become lifted through parasitism or full anaerobiosis. In the case of *P. lacertae*/*Blastocystis* the degeneration of the peroxisomal organelle manifests as a more conserved MRO. In other lineages the opposite may be the case. The span of organelle reduction is shown for peroxisomes (above the line) and MROs (below the line) with the line linking them representing their shared metabolic burden. I–V, respiratory complexes 1–5.

classical mitochondria and mitochondrial remnants.[52,53] Although the functions of aerobic (canonical) mitochondria are very well known, the functions of, and distinction among, anaerobic mitochondria, MROs, hydrogenosomes, and mitosomes is still blurred.[54] Our *in silico* predictions demonstrate that *P. lacertae* mitochondrial protein composition is similar to that of *Blastocystis* ST1, with 314 and 292 predicted proteins, respectively (Figure 5A). Despite the morphological differences between the two organelles (Figure 1), both MROs seem to have similar functions. Notably, the major distinction between the two organelles relates to the protein composition of the mitochondrial protein import machinery and proteins involved in organellar transcription and translation (Figure 5A; Table S3D). *P. lacertae* shows a reduced mitochondrial protein import machinery compared with *Blastocystis*, lacking proteins predicted to localize in the outer membrane of the organelle (e.g., Tom40, Tom70). By contrast, *P. lacertae* encodes almost double the number of proteins involved in mitochondrial transcription and translation when compared with *Blastocystis* (Figure 5A; Table S3D). Although the *Blastocystis* and *P. lacertae* mitochondrial genome complements have the identical number of protein coding genes, those of *P. lacertae* do encode more tRNAs and, notably, they do have distinctly different genomic organization (circular versus linear).[55] Whether the increased complement of transcriptional/translation machinery in *P. lacertae* reflects a requirement due to the linear mitochondrial genome organization is a matter for future molecular characterization.

Biochemically, anaerobic energy metabolism is the most striking difference between *P. lacertae* and *Blastocystis* MROs. The *P. lacertae* genome does not encode the [FeFe]-Hydrogenase and its maturase (HydE) that were shown to be present and localized in the *Blastocystis* organelle.[53] As the genes are also absent in *C. burkhardae*, the gene acquisitions parsimoniously took place in the *Blastocystis* lineage. Notably, attempts to show activity of this protein in *Blastocystis* have been unsuccessful, possibly due to incomplete machinery for maturation of the enzyme (it is lacking HydG and HydF[10]). In addition, *in silico* predictions have revealed that the *Blastocystis* ST1 genome encodes multiple pathways for the decarboxylation of pyruvate into acetyl-CoA and CO_2,[10,53] including the aerobic pyruvate dehydrogenase complex (PDC) and the anaerobic pyruvate:ferredoxin oxidoreductase (PFOR) and pyruvate:$NADP^+$ oxidoreductase (PNO).[10] Despite the absence of [FeFe]-Hydrogenase, *P. lacertae* encodes both enzymes for anaerobic decarboxylation (PFOR and PNO), with both having predicted mitochondrial targeting signals (Table S3D), and lacks all the genes coding for the aerobic complement (PDC and the pyruvate dehydrogenase kinases [PDK] 2/3/4 present in *Blastocystis*). *In vitro*, *P. lacertae* does not require inoculation into pre-reduced medium, in contrast to *Blastocystis* axenic culture (STAR Methods), implying that *P. lacertae* is more oxygen-tolerant. The PDC absence, while maintaining an anaerobic means for pyruvate decarboxylation, is puzzling and requires further investigation.

Like *Blastocystis* MROs, *P. lacertae* mitochondria harbor components of complex I and complex II of the electron transport chain, along with proteins involved in the (anaerobic) quinone metabolism, including the rhodoquinone biosynthesis enzyme RQUA[56] and alternative oxidase (AOX), which have been previously shown to associate with both complexes[57] (Figure 5B). These organelles also contain pathways for amino acid metabolism, cofactor/vitamin metabolism (folate, B5, B12, steroid, and lipoate), FA biosynthesis, and an incomplete tricarboxylic acid cycle, as well as maintenance of a mitochondrial genome (Figure 5). Like *Blastocystis*, *P. lacertae* encodes proteins of ISC assembly and export (e.g., ATM1) to support Fe-S assembly in the cytosol (CIA machinery) (Figure 5). In addition to the components of this machinery, *P. lacertae* encodes a fused sulfur mobilization protein (SufCB) that in *Blastocystis* was shown to bind to Fe-S clusters (e.g., [4Fe-4S]) and was expressed under oxygen stress and localized in its cytoplasm.[58]

In summary, the *P. lacertae* MRO seems to have a patchy conservation of anaerobic metabolism, while maintaining similar or more reduced functions when compared with the *Blastocystis* organelles. Together with the peroxisomal data, *P. lacertae* seems to have a more aerobically inclined metabolism, and thus could represent an intermediate stage between the microaerophilic stage and the more advanced anaerobic metabolism that is found in *Blastocystis*.

DISCUSSION

P. lacertae is a morphologically typical stramenopile, the genome of which offers a better comparison to the derived condition of *Blastocystis* than previous stramenopile genomes. It is the most closely related stramenopile to *Blastocystis* sequenced to date and is also adapted for life in the gut. In providing some indication of the ancestral state, it shows how the *Blastocystis* genome is genuinely small by stramenopile standards.[10] The reduced size is due to both the loss of specific cellular functions, such as motility and peroxisomes, as well as a profound genome-wide streamlining that caused as many genes to be lost as were retained in the *Blastocystis* lineage since divergence from the last common ancestor with *Proteromonas*. Genomic reduction in the *Blastocystis* genomes has influenced almost all aspects of cellular physiology, but, in

most cases, this has led to a simplification and not total loss of metabolic function.

Our data also contrast the relatively complex and conserved complement of mitochondrial metabolism genes to the almost completely reduced shared pathways in the peroxisomes. Co-evolved degeneration of these two organelles in the transition to anaerobiosis is seen convergently across multiple eukaryotic lineages and yet rules guiding this dynamic, if any, remain unclear. Because some inferred losses of peroxisomal proteins (e.g., Pex22 and Pex3) appear to have predated the move to anaerobiosis in the Opalozoa, our results raise the intriguing possibility that the earlier degeneration of peroxisomes could have acted as a brake on pathway loss in the mitochondria. As other lineages have more degenerate MROs, but more complete peroxisomal pathways (e.g., *M. balamuthi*[46] and *E. histolytica*[47]), the more general speculative evolutionary mechanism (Figure 6) would be that, for peroxisomes and mitochondria, the organelle that begins degeneration first results in negative selective pressure on pathway loss in the other. Once full adaptation to anaerobiosis has been achieved, then degeneration of both organelles proceeds. Whether this is simply a matter of contingency in this anaerobic lineage or reveals an evolutionary constraint in the transition to anaerobiosis remains to be investigated once many more lineages have been sampled.

Overall, the evolution of *Blastocystis* genomes has been characterized by a progressive, but pervasive genome-wide streamlining, with general loss of redundancy and punctuated by the loss of systems associated with cellular organelles, such as flagella and peroxisomes. This resulted in a lack of genomic versatility that mirrors the developmental and ecological uniformity we see in contemporary *Blastocystis*. This streamlining process is consistent with adaptation to a restricted niche within the host gut.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Culturing conditions for *P. lacertae* and *C. burkhardae*
- METHOD DETAILS
  - 'Omics sequencing, assembly and initial analysis
  - Genomics analysis
  - Informatic analyses of cellular systems
  - Microscopy
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.05.025.

ACKNOWLEDGMENTS

The authors wish to thank Andrew Roger (Dalhousie University), Marek Eliáš (University of Ostrava), Alastair Simpson (Dalhousie University), and Ida van der Klei (Rijksuniversiteit Groningen) for helpful discussion and long-term collaboration. This work was supported by Vanier Canada Graduate Scholarship (to E.K.H.), Gordon and Betty Moore Foundation (GBMF9327 to S.R.), Norwegian Research Council (301170 to M.v.d.G.), Leverhulme Trust Project Grant (RPG-2014-005 to A.P.J.), Interreg-2-seas H4DC grant (to A.D.T.), and Natural Sciences and Engineering Research Council of Canada (RES0043758 and RES0046091 to J.B.D.). This work was additionally funded by a Royal Society International Exchanges grant (2015/R1-IE150049, jointly to A.D.T. and J.B.D.). Computational resources were provided by the e-INFRA CZ project (ID: 90140), supported by the Ministry of Education, Youth and Sports of the Czech Republic. We would like to thank Sue Vaughan and Timm Mohr from Oxford Brookes University for assisting us with the reconstruction of the three-dimensional (3D) model of the *Proteromonas* cell using the images acquired from the transmission electron microscope.

AUTHOR CONTRIBUTIONS

Conceptualization, M.v.d.G., C.G.C., A.P.J., J.B.D., and A.D.T.; data curation, A.P.J., J.B.D., and A.D.T.; formal analysis, investigation, and results interpretation, K.Z., R.S.L., C.J.W., D.C., E.K.H., L.Y., C.A.R., Y.P., I.R.B., S.R., N.L.B., J.T., E.L.B., M.v.d.G., A.P.J., J.B.D., and A.D.T.; funding acquisition, M.v.d.G., C.G.C., and A.P.J.; project administration, A.P.J., J.B.D., and A.D.T.; supervision, S.R., E.G., A.P.J., J.B.D., and A.D.T.; visualization, K.Z., D.C., S.R., and A.D.T.; roles/writing – original draft, K.Z., R.S.L., A.P.J., J.B.D., and A.D.T.; writing – review & editing, all authors. All the authors have seen and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 16, 2022
Revised: March 22, 2023
Accepted: May 11, 2023
Published: June 1, 2023

REFERENCES

1. Brumpt, E. (1912). *Blastocystis hominis* n sp. et formes voisines. Bull. la Société Pathol. Exot. *5*, 725–730.
2. El Safadi, D., Gaayeb, L., Meloni, D., Cian, A., Poirier, P., Wawrzyniak, I., Delbac, F., Dabboussi, F., Delhaes, L., Seck, M., et al. (2014). Children of Senegal River Basin show the highest prevalence of *Blastocystis* sp. ever observed worldwide. BMC Infect. Dis. *14*, 164.
3. Scanlan, P.D., and Stensvold, C.R. (2013). *Blastocystis*: getting to grips with our guileful guest. Trends Parasitol. *29*, 523–529.
4. Betts, E.L., Gentekaki, E., and Tsaousis, A.D. (2020). Exploring micro-eukaryotic diversity in the gut: co-occurrence of *Blastocystis* subtypes and other protists in zoo animals. Front. Microbiol. *11*, 288.
5. Stensvold, C.R., and Clark, C.G. (2016). Current status of *Blastocystis*: a personal view. Parasitol. Int. *65*, 763–771.
6. Andersen, L.O., and Stensvold, C.R. (2016). *Blastocystis* in health and disease: are we moving from a clinical to a public health perspective? J. Clin. Microbiol. *54*, 524–528.
7. Jinatham, V., Maxamhud, S., Popluechai, S., Tsaousis, A.D., and Gentekaki, E. (2021). *Blastocystis* one health approach in a rural community of northern Thailand: prevalence, subtypes and novel transmission routes. Front. Microbiol. *12*, 746340.
8. Osorio-Pulgarin, M.I., Higuera, A., Beltran-Álzate, J.C., Sánchez-Jiménez, M., and Ramírez, J.D. (2021). Epidemiological and molecular characterization of *Blastocystis* infection in children attending daycare centers in Medellín, Colombia. Biology (Basel) *10*, 669.

9. Khaled, S., Gantois, N., Ly, A.T., Senghor, S., Even, G., Dautel, E., Dejager, R., Sawant, M., Baydoun, M., Benamrouz-Vanneste, S., et al. (2020). Prevalence and subtype distribution of *Blastocystis* sp. in senegalese school children. Microorganisms *8*, 1408.
10. Gentekaki, E., Curtis, B.A., Stairs, C.W., Klimeš, V., Eliáš, M., Salas-Leiva, D.E., Herman, E.K., Eme, L., Arias, M.C., Henrissat, B., et al. (2017). Extreme genome diversity in the hyper-prevalent parasitic eukaryote *Blastocystis*. PLoS Biol. *15*, e2003769.
11. Silberman, J.D., Sogin, M.L., Leipe, D.D., and Clark, C.G. (1996). Human parasite finds taxonomic home. Nature *380*, 398.
12. Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., et al. (2019). Revisions to the classification, nomenclature, and diversity of eukaryotes. J. Eukaryot. Microbiol. *66*, 4–119.
13. Cavalier-Smith, T., and Chao, E.E.Y. (2006). Phylogeny and megasystematics of phagotrophic heterokonts (kingdom Chromista). J. Mol. Evol. *62*, 388–420.
14. Kostka, M. (2017). Opalinata. In Handbook of the Protists, J.M. Archibald, A.G.B. Simpson, and C.H. Slamovits, eds. (Springer International Publishing), pp. 543–565.
15. Brugerolle, G., and Mignot, J.-P. (1990). 14h. Phylum Zoomastigina, class Proteromonadida. In Handbook of Protoctista: The Structure, Cultivation, Habitats and Life Histories of the Eukaryotic Microorganisms and Their Descendants Exclusive of Animals, Plants and Fungi: A Guide to the Algae, Ciliates, Foraminifera, Sporozoa, Water Molds, Slime Mold., L. Margulis, and J.O. Corliss, eds. (Jones and Bartlett Publishers), pp. 246–251.
16. Brugerolle, G., and Joyon, L. (1975). Étude cytologique ultrastructurale des genres *Proteromonas* et Karotomorpha (Zoomoastigophorea, Proteromonadida Grassé, 1952). Protistologica *11*, 531–546.
17. Brugerolle, G., and Bardele, C.F. (1988). Cortical cytoskeleton of the flagellate *Proteromonas lacertae*: interrelation between microtubules, membrane and somatonemes. Protoplasma *142*, 46–54.
18. Zierdt, C.H. (1991). *Blastocystis hominis*–past and future. Clin. Microbiol. Rev. *4*, 61–79.
19. Denoeud, F., Roussel, M., Noel, B., Wawrzyniak, I., Da Silva, C., Diogon, M., Viscogliosi, E., Brochier-Armanet, C., Couloux, A., Poulain, J., et al. (2011). Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parasite. Genome Biol. *12*, R29.
20. Wawrzyniak, I., Courtine, D., Osman, M., Hubans-Pierlot, C., Cian, A., Nourrisson, C., Chabe, M., Poirier, P., Bart, A., Polonais, V., et al. (2015). Draft genome sequence of the intestinal parasite *Blastocystis* subtype 4-isolate WR1. Genomics Data *4*, 22–23.
21. Schoenle, A., Hohlfeld, M., Rosse, M., Filz, P., Wylezich, C., Nitsche, F., and Arndt, H. (2020). Global comparison of bicosoecid *Cafeteria*-like flagellates from the deep ocean and surface waters, with reorganization of the family Cafeteriaceae. Eur. J. Protistol. *73*, 125665.
22. Klimeš, V., Gentekaki, E., Roger, A.J., and Eliáš, M. (2014). A large number of nuclear genes in the human parasite *Blastocystis* require mRNA polyadenylation to create functional termination codons. Genome Biol. Evol. *6*, 1956–1961.
23. Nanjappa, D., Sanges, R., Ferrante, M.I., and Zingone, A. (2017). Diatom flagellar genes and their expression during sexual reproduction in *Leptocylindrus danicus*. BMC Genomics *18*, 813.
24. Van Dam, T.J.P., Townsend, M.J., Turk, M., Schlessinger, A., Sali, A., Field, M.C., and Huynen, M.A. (2013). Evolution of modular intraflagellar transport from a coatomer-like progenitor. Proc. Natl. Acad. Sci. USA *110*, 6943–6948.
25. Judelson, H.S., Shrivastava, J., and Manson, J. (2012). Decay of genes encoding the oomycete flagellar proteome in the downy mildew *Hyaloperonospora arabidopsidis*. PLoS One *7*, e47624.
26. Patterson, D.J. (1989). Stramenopiles: chromophytes from a protistan perspective. In The Chromophyte Algae: Problems and Perspectives, Systematics Association Special Volume, No. 38, J.C. Green, B.S.C. Leadbeater, and W.L. Diver, eds. (Clarendon Press), pp. 357–379.
27. Montresor, M., Vitale, L., D'Alelio, D., and Ferrante, M.I. (2016). Sex in marine planktonic diatoms: insights and challenges. Perspect. Phycol. *3*, 61–75.
28. Hee, W.Y., Blackman, L.M., and Hardham, A.R. (2019). Characterisation of stramenopile-specific mastigoneme proteins in *Phytophthora parasitica*. Protoplasma *256*, 521–535.
29. Blackman, L.M., Arikawa, M., Yamada, S., Suzaki, T., and Hardham, A.R. (2011). Identification of a mastigoneme protein from *Phytophthora nicotianae*. Protist *162*, 100–114.
30. Leonard, G., Labarre, A., Milner, D.S., Monier, A., Soanes, D., Wideman, J.G., Maguire, F., Stevens, S., Sain, D., Grau-Bové, X., et al. (2018). Comparative genomic analysis of the 'pseudofungus' *Hyphochytrium catenoides*. Open Biol. *8*, 170184.
31. Cavalier-Smith, T., and Scoble, J.M. (2013). Phylogeny of Heterokonta: *Incisomonas marina*, a uniciliate gliding opalozoan related to Solenicola (Nanomonadea), and evidence that Actinophryida evolved from raphidophytes. Eur. J. Protistol. *49*, 328–353.
32. Park, J.S., Cho, B.C., and Simpson, A.G.B. (2006). *Halocafeteria seosinensis* gen. et sp. nov. (Bicosoecida), a halophilic bacterivorous nanoflagellate isolated from a solar saltern. Extremophiles *10*, 493–504.
33. Blacque, O.E., Scheidel, N., and Kuhns, S. (2018). Rab GTPases in cilium formation and function. Small GTPases *9*, 76–94.
34. Elias, M., Brighouse, A., Gabernet-Castello, C., Field, M.C., and Dacks, J.B. (2012). Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. J. Cell Sci. *125*, 2500–2508.
35. Klinger, C.M., Ramirez-Macias, I., Herman, E.K., Turkewitz, A.P., Field, M.C., and Dacks, J.B. (2016). Resolving the homology—function relationship through comparative genomics of membrane-trafficking machinery and parasite cell biology. Mol. Biochem. Parasitol. *209*, 88–103.
36. Johnson, A., Dahhan, D.A., Gnyliukh, N., Kaufmann, W.A., Zheden, V., Costanzo, T., Mahou, P., Hrtyan, M., Wang, J., Aguilera-Servin, J., et al. (2021). The TPLATE complex mediates membrane bending during plant clathrin–mediated endocytosis. Proc. Natl. Acad. Sci. USA *118*, e2113046118.
37. Perry, R.J., Mast, F.D., and Rachubinski, R.A. (2009). Endoplasmic reticulum-associated secretory proteins Sec20p, Sec39p, and Dsl1p are involved in peroxisome biogenesis. Eukaryot. Cell *8*, 830–843.
38. Klinger, C.M., Klute, M.J., and Dacks, J.B. (2013). Comparative genomic analysis of multi-subunit tethering complexes demonstrates an ancient pan-eukaryotic complement and sculpting in Apicomplexa. PLoS One *8*, e76278.
39. Schlüter, A., Fourcade, S., Ripp, R., Mandel, J.L., Poch, O., and Pujol, A. (2006). The evolutionary origin of peroxisomes: an ER-peroxisome connection. Mol. Biol. Evol. *23*, 838–845.
40. Kalel, V.C., Li, M., Gaussmann, S., Delhommel, F., Schäfer, A.B., Tippler, B., Jung, M., Maier, R., Oeljeklaus, S., Schliebs, W., et al. (2019). Evolutionary divergent PEX3 is essential for glycosome biogenesis and survival of trypanosomatid parasites. Biochim. Biophys. Acta Mol. Cell Res. *1866*, 118520.
41. McDonnell, M.M., Burkhart, S.E., Stoddard, J.M., Wright, Z.J., Strader, L.C., and Bartel, B. (2016). The early-acting peroxin PEX19 is redundantly encoded, farnesylated, and essential for viability in *Arabidopsis thaliana*. PLoS One *11*, e0148335.
42. Emmanouilidis, L., Schütz, U., Tripsianes, K., Madl, T., Radke, J., Rucktäschel, R., Wilmanns, M., Schliebs, W., Erdmann, R., and Sattler, M. (2017). Allosteric modulation of peroxisomal membrane protein recognition by farnesylation of the peroxisomal import receptor PEX19. Nat. Commun. *8*, 14635.
43. Lyschik, S., Lauer, A.A., Roth, T., Janitschke, D., Hollander, M., Will, T., Hartmann, T., Kopito, R.R., Helms, V., Grimm, M.O.W., et al. (2022). PEX19 coordinates neutral lipid storage in cells in a peroxisome-independent fashion. Front. Cell Dev. Biol. *10*, 859052.

44. Xie, Y., Zheng, Y., Li, H., Luo, X., He, Z., Cao, S., Shi, Y., Zhao, Q., Xue, Y., Zuo, Z., et al. (2016). GPS-Lipid: a robust tool for the prediction of multiple lipid modification sites. Sci. Rep. *6*, 28249.

45. Gabaldón, T. (2010). Peroxisome diversity and evolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. *365*, 765–773.

46. Le, T., Žárský, V., Nývltová, E., Rada, P., Harant, K., Vancová, M., Verner, Z., Hrdý, I., and Tachezy, J. (2020). Anaerobic peroxisomes in *Mastigamoeba balamuthi*. Proc. Natl. Acad. Sci. USA *117*, 2065–2075.

47. Verner, Z., Žárský, V., Le, T., Narayanasamy, R.K., Rada, P., Rozbeský, D., Makki, A., Belišová, D., Hrdý, I., Vancová, M., et al. (2021). Anaerobic peroxisomes in *Entamoeba histolytica* metabolize myo-inositol. PLoS Pathog. *17*, e1010041.

48. Záhonová, K., Treitli, S.C., Le, T., Škodová-Sveráková, I., Hanousková, P., Čepička, I., Tachezy, J., and Hampl, V. (2022). Anaerobic derivates of mitochondria and peroxisomes in the free-living amoeba *Pelomyxa schiedti* revealed by single-cell genomics. BMC Biol. *20*, 56.

49. Moog, D., Przyborski, J.M., and Maier, U.G. (2017). Genomic and proteomic evidence for the presence of a peroxisome in the apicomplexan parasite *Toxoplasma gondii* and other coccidia. Genome Biol. Evol. *9*, 3108–3121.

50. Río Bártulos, C.R., Rogers, M.B., Williams, T.A., Gentekaki, E., Brinkmann, H., Cerff, R., Liaud, M.F., Hehl, A.B., Yarlett, N.R., Gruber, A., et al. (2018). Mitochondrial glycolysis in a major lineage of eukaryotes. Genome Biol. Evol. *10*, 2310–2325.

51. Garg, S., Stölting, J., Zimorski, V., Rada, P., Tachezy, J., Martin, W.F., and Gould, S.B. (2015). Conservation of transit peptide-independent protein import into the mitochondrial and hydrogenosomal matrix. Genome Biol. Evol. *7*, 2716–2726.

52. Pérez-Brocal, V., and Clark, C.G. (2008). Analysis of two genomes from the mitochondrion-like organelle of the intestinal parasite *Blastocystis*: complete sequences, gene content, and genome organization. Mol. Biol. Evol. *25*, 2475–2482.

53. Stechmann, A., Hamblin, K., Pérez-Brocal, V., Gaston, D., Richmond, G.S.S., van der Giezen, M., Clark, C.G., and Roger, A.J. (2008). Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. Curr. Biol. *18*, 580–585.

54. Richardson, E., Zerr, K., Tsaousis, A., Dorrell, R.G., and Dacks, J.B. (2015). Evolutionary cell biology: functional insight from "endless forms most beautiful.". Mol. Biol. Cell *26*, 4532–4538.

55. Pérez-Brocal, V., Shahar-Golan, R., and Clark, C.G. (2010). A linear molecule with two large inverted repeats: the mitochondrial genome of the stramenopile *Proteromonas lacertae*. Genome Biol. Evol. *2*, 257–266.

56. Stairs, C.W., Eme, L., Muñoz-Gómez, S.A., Cohen, A., Dellaire, G., Shepherd, J.N., Fawcett, J.P., and Roger, A.J. (2018). Microbial eukaryotes have adapted to hypoxia by horizontal acquisitions of a gene involved in rhodoquinone biosynthesis. eLife *7*, e34292.

57. Tsaousis, A.D., Hamblin, K.A., Elliott, C.R., Young, L., Rosell-Hidalgo, A., Gourlay, C.W., Moore, A.L., and van der Giezen, M. (2018). The human gut colonizer *Blastocystis* respires using complex II and alternative oxidase to buffer transient oxygen fluctuations in the gut. Front. Cell. Infect. Microbiol. *8*, 371.

58. Tsaousis, A.D., Ollagnier de Choudens, S.O., Gentekaki, E., Long, S., Gaston, D., Stechmann, A., Vinella, D., Py, B., Fontecave, M., Barras, F., et al. (2012). Evolution of Fe/S cluster biogenesis in the anaerobic parasite *Blastocystis*. Proc. Natl. Acad. Sci. USA *109*, 10426–10431.

59. Mruk, D.D., and Cheng, C.Y. (2011). Enhanced chemiluminescence (ECL) for routine immunoblotting: an inexpensive alternative to commercially available kits. Spermatogenesis *1*, 121–122.

60. Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). Augustus: a web server for gene finding in eukaryotes. Nucleic Acids Res. *32*, W309–W312.

61. Korf, I. (2004). Gene finding in novel genomes. BMC Bioinformatics *5*, 59.

62. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

63. Zdobnov, E.M., and Apweiler, R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics *17*, 847–848.

64. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.

65. Bendtsen, J.D., Nielsen, H., Von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. *340*, 783–795.

66. Pejaver, V., Hsu, W.L., Xin, F., Dunker, A.K., Uversky, V.N., and Radivojac, P. (2014). The structural and functional signatures of proteins that undergo multiple events of post-translational modification. Protein Sci. *23*, 1077–1093.

67. Griffiths-Jones, S. (2005). Annotating non-coding RNAs with Rfam. Curr. Protoc. Bioinform. *33*, 12.5.1–12.5.12.

68. Smit, A., and Hubley, R. (2010). RepeatModeler Open-1.0. http://www.repeatmasker.org.

69. Xu, Z., and Wang, H. (2007). LTR-FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. *35*, W265–W268.

70. Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674–3676.

71. Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-seq. Bioinformatics *25*, 1105–1111.

72. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. *8*, 1494–1512.

73. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.

74. Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189.

75. Guindon, S., Delsuc, F., Dufayard, J.F., and Gascuel, O. (2009). Estimating maximum likelihood phylogenies with PhyML. Methods Mol. Biol. *537*, 113–137.

76. Barlow, L.D., Maciejowski, W., More, K., Terry, K., Vargová, R., Záhonová, K., and Dacks, J.B. (2023). Comparative genomics for evolutionary cell biology using AMOEBAE: understanding the Golgi and beyond. In Golgi: Methods and Protocols, Y. Wang, V.V. Lupashin, and T.R. Graham, eds. (Springer), pp. 431–452.

77. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

78. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

79. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

80. Eddy, S.R. (2011). Accelerated profile HMM searches. PLoS Comput. Biol. *7*, e1002195.

81. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

82. Maddison, W.P., and Maddison, D.R. (2015). Mesquite: a modular system for evolutionary analysis. https://www.mesquiteproject.org/.

83. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics *27*, 1164–1165.

84. Lartillot, N., and Philippe, H. (2006). Computing Bayes factors using thermodynamic integration. Syst. Biol. *55*, 195–207.

85. Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. *21*, 1095–1109.

86. Lartillot, N., Brinkmann, H., and Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol. Biol. *7*, S4.

87. Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics *19*, 1572–1574.

88. Huelsenbeck, J.P., Ronquist, F., Nielsen, R., and Bollback, J.P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. Science *294*, 2310–2314.

89. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

90. Almagro Armenteros, J.J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., and Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. Life Sci. Alliance *2*, e201900429.

91. Kume, K., Amagasa, T., Hashimoto, T., and Kitagawa, H. (2018). NommPred: prediction of mitochondrial and mitochondrion-related organelle proteins of nonmodel organisms. Evol. Bioinform. Online *14*. 1176934318819835.

92. Kulda, J. (1973). Axenic cultivation of *Proteromonas lacertae-viridis* (Grassi, 1879). J. Protozool. *20*, 536–537.

93. Clark, C.G., and Diamond, L.S. (2002). Methods for cultivation of luminal parasitic protists of clinical importance. Clin. Microbiol. Rev. *15*, 329–341.

94. Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., Li, Z., Chen, Y., Mu, D., and Fan, W. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. Preprint at arXiv. https://doi.org/10.48550/arXiv.1308.2012.

95. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. *40*, D109–D114.

96. Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. *428*, 726–731.

97. Wang, H.-C., Minh, B.Q., Susko, E., and Roger, A.J. (2018). Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst. Biol. *67*, 216–235.

98. Hirst, J., Schlacht, A., Norcott, J.P., Traynor, D., Bloomfield, G., Antrobus, R., Kay, R.R., Dacks, J.B., and Robinson, M.S. (2014). Characterization of TSET, an ancient and widespread membrane trafficking complex. eLife *3*, e02866.

99. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

100. Lametschwandtner, G., Brocard, C., Fransen, M., Van Veldhoven, P., Berger, J., and Hartig, A. (1998). The difference in recognition of terminal tripeptides as peroxisomal targeting signal 1 between yeast and human is due to different affinities of their receptor Pex5p to the cognate signal and to residues adjacent to it. J. Biol. Chem. *273*, 33635–33643.

101. Petriv, O.I., Tang, L., Titorenko, V.I., and Rachubinski, R.A. (2004). A new definition for the consensus sequence of the peroxisome targeting signal type 2. J. Mol. Biol. *341*, 119–134.

102. Snyder, W.B., Faber, K.N., Wenzel, T.J., Koller, A., Lüers, G.H., Rangell, L., Keller, G.A., and Subramani, S. (1999). Pex19p interacts with Pex3p and Pex10p and is essential for peroxisome biogenesis in *Pichia pastoris*. Mol. Biol. Cell *10*, 1745–1761.

103. Betts, E.L., Gentekaki, E., Thomasz, A., Breakell, V., Carpenter, A.I., and Tsaousis, A.D. (2018). Genetic diversity of *Blastocystis* in non-primate animals. Parasitology *145*, 1228–1234.

104. Tsaousis, A.D., Gentekaki, E., Eme, L., Gaston, D., and Roger, A.J. (2014). Evolution of the cytosolic iron-sulfur cluster assembly machinery in *Blastocystis* species and other microbial eukaryotes. Eukaryot. Cell *13*, 143–153.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
PXMP2 anti-rabbit antibody	this paper	N/A
Pex10 anti-rat antibody	this paper	N/A
Pex19 anti-rabbit antibody	Abcam	Catalog number: ab137072
MitoTracker Red CMXRos	Invitrogen	Catalog number: M7512
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Invitrogen	Catalog number: A-11008
Goat anti-Rat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Invitrogen	Catalog number: A-11006
Goat anti-Rat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 594	Invitrogen	Catalog number: A-11007
VECTASHIELD Antifade Mounting Medium	2Scientific	Catalog number: H-1000-10
Chemicals, peptides, and recombinant proteins		
DNeasy mini kit	Qiagen	Catalog number: 69504
RNeasy kit	Qiagen	Catalog number: 74104
glutaradehyde	Sigma-Aldrich	Catalog number: G5882
sodium cacodylate	Sigma-Aldrich	Catalog number: CO250
low melting point agarose	Millipore	Catalog number: 2070-OP
Alcain blue-0.1 % acetic acid dye	Sigma-Aldrich	Catalog number: 05500
Osmium tetroxide (OsO_4)	Sigma-Aldrich	Catalog number: 201030
propylene oxide	Sigma-Aldrich	Catalog number: 110205
Low viscosity Resin	Sigma-Aldrich	Catalog number: 900149
VH1 hardener	agar scientific	Catalog number: AGR1375
VH2 hardener	agar scientific	Catalog number: AGR1376
LV accelerator	agar scientific	Catalog number: AGR1381
uranyl acetate	agar scientific	Catalog number: AFR1000
potassium hydroxide	Sigma-Aldrich	Catalog number: 4.80864
phosphate buffered saline	Sigma-Aldrich	Catalog number: P4417
LR white resin	agar Scientific	Catalog number: AGR1281
gelatine	agar Scientific	Catalog number: AGG29209
bovine serum albumin	Sigma-Aldrich	Catalog number: A8654
Tween 20	Sigma-Aldrich	Catalog number: P1379
Reynold's lead citrate	As previously published	https://marclab.org/tools/protocols/reynolds-lead-citrate-stain-for-grids/
adult bovine serum	Gibco	Catalog number: 16170078
$MgCl_2$	Sigma-Aldrich	Catalog number: M8266
Tris-HCl	Sigma-Aldrich	Catalog number: PHG0002
EDTA free protease inhibitor	Sigma-Aldrich	Catalog number: 11873580001
Laemmli Protein Sample Buffer	Biorad	Catalog number: 1610737
NaCl	Sigma-Aldrich	Catalog number: S9888
Trizma	Sigma-Aldrich	Catalog number: T8524
ECL reagents	As previously published[59]	N/A
Deposited data		
Proteromonas lacertae genome	Genbank: BioProject PRJNA386230; Biosample SAMN06926116	Raw reads and assembly available at NCBI: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA386230

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Proteromonas lacertae transcriptome	Genbank: BioProject PRJEB60805; Biosample SAMEA112818123	Raw reads available at NCBI: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB60805
Cafeteria burkhardae transcriptome	Genbank: BioProject PRJEB54677; Biosample SAMEA110341195	Raw reads and assembly available at NCBI: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB54677
Software and algorithms		
AUGUSTUS v2.4	Stanke et al.[60]	RRID: SCR_008417; http://bioinf.uni-greifswald.de/augustus/
SNAP	Korf[61]	RRID: SCR_007936; http://snap.genomics.org.cn
BLAST	Altschul et al.[62]	RRID: SCR_004870; https://blast.ncbi.nlm.nih.gov/Blast.cgi
InterProScan v5.21-60.0	Zdobnov and Apweiler[63]	RRID: SCR_005829; http://www.ebi.ac.uk/Tools/pfa/iprscan/
TMHMM v2.0c	Krogh[64]	RRID: SCR_014935; https://services.healthtech.dtu.dk/services/TMHMM-2.0/
SignalP v4.1	Bendtsen et al.[65]	RRID: SCR_015644; https://services.healthtech.dtu.dk/services/SignalP-5.0/
ModPred	Pejaver et al.[66]	http://montana.informatics.indiana.edu/ModPred/
RfamScan v1.1.1	Griffiths-Jones[67]	RRID: SCR_007891; http://rfam.xfam.org/
RepeatModeler v1.0.4	Smit and Hubley[68]	RRID: SCR_015027; http://www.repeatmasker.org/RepeatModeler/
LTRfinder v1.0.5	Xu and Wang[69]	https://github.com/xzhub/LTR_Finder
BLAST2GO v4.1	Conesa et al.[70]	RRID: SCR_005828; http://www.blast2go.com/b2ghome
TopHat v2.1.1	Trapnell et al.[71]	RRID: SCR_013035; http://ccb.jhu.edu/software/tophat/index.shtml
Trinity v2.1.1	Haas et al.[72]	RRID: SCR_013048; http://trinityrnaseq.sourceforge.net/
BUSCO v1.1.1	Simão et al.[73]	RRID: SCR_015008; https://busco.ezlab.org/
OrthoMCL v2.0.9	Li et al.[74]	RRID: SCR_007839; http://www.orthomcl.org/cgi-bin/OrthoMclWeb.cgi
PhyML	Guindon et al.[75]	RRID: SCR_014629; http://www.atgc-montpellier.fr/phyml/
AMOEBAE	Barlow et al.[76]	https://github.com/laelbarlow/amoebae
MAFFT v7.458	Katoh and Standley[77]	RRID: SCR_011811; https://mafft.cbrc.jp/alignment/software/
trimAl v1.4	Capella-Gutiérrez et al.[78]	RRID: SCR_017334; http://trimal.cgenomics.org/
IQ-TREE v1.6.12	Nguyen et al.[79]	RRID: SCR_017254
HMMER v3.1b1	Eddy[80]	RRID: SCR_005305; http://hmmer.org
MUSCLE v3.8.31	Edgar[81]	RRID: SCR_011812; http://www.ebi.ac.uk/Tools/msa/muscle/
Mesquite v3.03	Maddison and Maddison[82]	RRID: SCR_017994; https://www.mesquiteproject.org/
ProtTest v3.4	Darriba et al.[83]	RRID: SCR_014628; http://darwin.uvigo.es/software/prottest_server.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PhyloBayes v3.3	Lartillot and Philippe[84,85] and Lartillot et al.[86]	RRID: SCR_006402; https://github.com/bayesiancook/pbmpi
MrBayes v3.2.2	Ronquist and Huelsenbeck[87] and Huelsenbeck et al.[88]	RRID: SCR_012067; https://nbisweden.github.io/MrBayes/
RAxML v8.1.3	Stamatakis[89]	RRID: SCR_006086; https://github.com/stamatak/standard-RAxML
TargetP-2.0	Almagro Armenteros et al.[90]	RRID: SCR_019022; https://services.healthtech.dtu.dk/services/TargetP-2.0/
NommPred	Kume et al.[91]	https://gitlab.com/kkei/NommPred

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Anastasios D. Tsaousis (A.Tsaousis@kent.ac.uk).

Materials availability

This study has generated two unique antibodies that can be obtained from the lead contact. Antibodies will be made available on request, but we may require a payment and/or a completed Materials Transfer Agreement if there is potential for commercial application.

Data and code availability

- Raw reads and assemblies have been deposited at NCBI and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Culturing conditions for *P. lacertae* and *C. burkhardae*

Proteromonas lacertae LA cultures[92] were grown axenically in LYI-S-2[93] with 15% adult bovine serum (Sigma Aldrich) at 22 °C in sterile borosilicate glass tubes and passaged every 2-3 weeks. *Cafeteria burkhardae* ATCC 50561 was maintained in artificial seawater for protozoa (ASWP) at 4 °C or at room temperature in light-shielded T-25 culture flasks.

METHOD DETAILS

'Omics sequencing, assembly and initial analysis

P. lacertae *and* C. burkhardae *nucleic acid extraction*

P. lacertae genomic DNA was extracted using DNeasy mini kit (Qiagen) according to the manufacturer's protocol. Twenty μg of DNA of high molecular weight was submitted for sequencing by the University of Liverpool's Centre for Genomic Research. Libraries were prepared by shearing DNA to approximately 10 kb fragments. Sequencing was done primarily on a single SMRT cell.

Whole RNA from *P. lacertae* and *C. burkhardae* was extracted using a RNeasy kit (Qiagen) according to the manufacturer's protocol. RNA samples were pooled and processed by the University of Liverpool's Centre for Genomic Research using polyA selection according to the manufacturer's protocol.

P. lacertae *genome and transcriptome sequencing and assembly*

Genome sequencing was done on a PacBio RSII sequencer with nine SMRT cells and assembled using SMRT Portal software (HGAP 3). The RNA samples were used to produce three Illumina RNASeq libraries from enriched RNA using the strand-specific ScriptSeq kit. Paired-end sequencing (2x125 bp) was carried out on one lane using Illumina HiSeq platform.

Assembly parameters were default except for genome size, which was set to 35 Mb, based on kmer size estimation as previously described,[94] and minimum seed read length, which was increased to 10,000 bp.

Gene calling was done using AUGUSTUS v2.4[60] and SNAP[61] utilizing a training set of 188 genes. The final gene set was manually curated. Gene annotations were assigned based on homology (BLASTx[62]), InterProScan v5.21-60.0,[63] TMHMM v2.0c,[64] SignalP v4.1,[65] ModPred,[66] RfamScan v1.1.1,[67] RepeatModeler v1.0.4,[68] LTRfinder v1.0.5,[69] and Kyoto Encyclopedia of Genes and Genomes (KEGG)[95] annotations were assigned to protein sequences by GhostKoala[96] and BLAST2GO v4.1[70] and were mapped

to KEGG pathways using KEGG Mapper – Reconstruct Pathway. To aid annotation, the transcriptome was mapped to the contigs using TopHat v2.1.1.[71]

C. burkhardae *transcriptome sequencing and assembly*

Library preparation and sequencing was performed as for *P. lacertae*. Reads were assembled using Trinity v2.1.1.[72] Bacterial contamination was removed using a PCA of kmer frequencies to filter sequences with >98% sequence identity with a known bacterial sequence.

The initial transcriptome contained 40,858 unique transcripts. Although polyA selected, bacterial transcripts from the xenic cell culture were frequent contaminants in the assembly. To remove prokaryotic transcripts, we examined kmer frequencies to identify a *Cafeteria*-specific signature. A positive transcript set included sequences with a top BLAST hit to a stramenopile, with >40% sequence identity. A negative transcript set included transcripts >95% sequence identity to a bacterial sequence. Kmer frequencies were calculated for these and all other transcripts and analysed using Principal Component Analysis (PCA). It was possible to clearly distinguish the positive and negative control groups, and thereby assign all remaining transcripts as eukaryotic or prokaryotic accordingly (Figure S5). 11,633 transcripts that clustered with the negative control group were removed after manual checking of their BLAST affinities. The final transcriptome contains 28,952 transcripts, which returned a BUSCO score of 70.40%.

For each of our datasets, completeness was calculated from the BUSCO v1.1.1[73] using eukaryota_obd9 dataset.

Genomics analysis

Comparative analysis of the *P. lacertae* and *Blastocystis* gene sets was carried out using OrthoMCL, to arrange genes into orthologous clusters that could then be examined for evolutionary conservation and loss. To establish whether genes were gained, lost, or conserved in each genome, we clustered them with various stramenopile outgroups, including the *C. burhardae* transcriptome produced here. We included genome assemblies with low scaffold and contig counts, high N50 values and high BUSCO scores for completeness: *Blastocystis* sp. ST4 WR1, *Blastocystis* sp. ST7 Singapore isolate B, *Pythium ultimum* DAOM BR144, *Phytophthora sojae* strain P6497, *Saprolegnia diclina* VS20, *Ectocarpus siliculosus* strain Ec32, *Thalassiosira pseudonana* CCMP1335, and *Schizochytrium* sp. transcriptome. All genomes and transcriptomes were downloaded from NCBI Genome. OrthoMCL v2.0.9[74] was used with an E-value threshold of 1e-5 for all-v-all BLAST to generate clusters of homologous genes. We acknowledge the inherent potential for false positives and negatives when using RBH and high-throughput methods. However, given the dataset size phylogenetic analysis of all proteins was untenable. In cases of specific cellular system analyses, phylogenetics was used as described below.

A cluster was considered 'conserved' if they contained at least one sequence from *P. lacertae*, at least one *Blastocystis* subtype, and at least one outgroup (including *C. burkhardae*). A cluster was 'species-specific' if it only contained sequences from a single genome (except *Blastocystis* where it may contain representative sequences from both subtypes). Clusters were considered to represent losses from the *Blastocystis* genomes if they were absent from all *Blastocystis* genomes, but present in both *P. lacertae* and at least one other stramenopile.

To confirm the *P. lacertae* species-specific clusters and rule out contamination several points were considered. 18,425 (89.8%) of the putative *P. lacertae*-specific genes are contiguous with conserved stramenopile genes. Furthermore, the remaining putative species-specific genes not physically linked to conserved loci have codon usage or predicted amino acid composition not significantly different to conserved genes (t-test, $p>0.3$), indicating that they are authentic coding sequences. Finally, these species-specific genes do not have close affinity to other organisms, which would be indicative of contamination; instead 78.8% have no homology with known proteins using BLASTp and, among those genes that do display homology, average sequence identity is low (30.1%) and does not exceed 87%. Thus, it is likely that the high proportion of *P. lacertae*-specific genes reflects the poor genomic sampling of stramenopiles to date.

After separating gene clusters in this way, KEGG orthology (KO) terms were associated with genes using the KEGGmapper tool. The observed incidence of each KO term among *P. lacertae* or *Blastocystis* conserved genes and gene losses respectively was compared to the expected incidence given its frequency in the whole genome, and a hypergeometric test with Bonferroni correction was applied to identify KO terms with significant under- or over-representation. To identify Gene Ontology (GO) terms that were significantly enriched, human homologues for *P. lacertae* or *Blastocystis* genes in the conserved and loss categories were identified using BLASTp, and these were compared to the background human Gene Ontology using Fishers exact test with Benjamini correction using GOnet.

Informatic analyses of cellular systems

Flagellar complement

Judelson et al.[25] previously surveyed the distribution of >1,000 motility-associated proteins in diverse eukaryotes. This identified 16 proteins that were conserved among all flagellated eukaryotes. Human homologues of these proteins (except for the transmembrane O-methyltransferase LRTOMT, which we found to be absent from trypanosomes) were used as queries in BLASTp searches to *Blastocystis* and other organisms: BBS4 (Bardet-Biedl syndrome 4 protein; NP_149017.2); BBS5 (Bardet-Biedl syndrome 5 protein; NP_689597.1); TTC8 (Tetratricopeptide repeat domain 8; NP_938051.1); CFAP20 (Cilia- and flagella-associated protein 20; NP_037374.1); DAW1 (Dynein assembly factor with WDR repeat domains 1; NP_849143.1); DYNC1I2 (Cytoplasmic dynein 1 intermediate chain 2; NP_0012587.1); DRC3 (Dynein regulatory complex subunit 3; NP_001123563.1); CLUAP1 (Clusterin-associated protein 1; NP_0013173.1); AGBL3 (Cytosolic carboxypeptidase 3; XP_0168676.1); KIF3C (Kinesin-like protein KIF3C; NP_002245.4);

IFT22 (Intraflagellar transport protein 22; NP_073614.1); IFT52 (Intraflagellar transport protein 52; NP_057088.2); IFT57 (Intraflagellar transport protein 57; NP_060480.1); IFT88 (Intraflagellar transport protein 88; NP_001340501.1); SPAG6 (Sperm-associated antigen 6; NP_036575.1); and RIBC2 (RIB43A-like with coiled-coils protein 2; NP_056468.1).

Gene sets were obtained from five flagellated organisms of diverse affinity [*Homo sapiens* (GRCh38), *Trypanosoma brucei* strain TREU927 (TryBru_Apr2005_chr11), *Tetrahymena thermophila* (JCVI-TTA1-2.2), *Chlamydomonas reinhardtii* (v5.5), and *Naegleria gruberi* strain NEG-M (v1.0)), three unflagellated organisms (*Schizosaccharomyces pombe* (ASM294v2), *Ostreococcus lucimarinus* (ASM9206v1), and *Hyaloperonospora arabidopsidis* (HyaAraEmoy2_2.0)], as well as the *P. lacertae* genome and *C. burkhardae* transcriptome produced in this study, and finally, the *Blastocystis hominis* strain Singapore B (sub-type 7) genome (ASM15166v1).

A reciprocal best match by BLASTp between the human protein query and the subject protein in the non-human genome was required to confirm that an ortholog was 'present' in the latter. Homologous sequences that were not best matches in reverse comparisons were considered to be non-orthologues and the query protein was recorded as 'absent'. KIF3C, a kinesin-like protein that is found in multiple copies in the human genome, among others (Figure S1A), was found in the three non-flagellated organisms and is a microtubule-based anterograde translocator with multiple functions, possibly unrelated to motility. BLASTp matches to other flagellar proteins were found in *Blastocystis*, (e.g. to cytosolic dynein DYNC1I2, dynein regulatory complex subunit 3, DRC3, and intraflagellar transport protein 22, IFT22) but these were not reciprocal best hits and, in fact, did not fully align with the query. This is detailed in Table S3A; for example, a *Blastocystis* protein was homologous to DRC3 but only to a 72-amino acid span, rather than to the >500 amino acids typical of true orthologues in flagellated organisms. Therefore, we conclude that these partial hits are caused by homologous domains shared by otherwise unrelated proteins.

For the 13 *P. lacertae* and 9 *C. burkhardae* proteins identified, phylogenetic analyses (Figure S2) showed that the genetic distances to *P. lacertae* or *C. burkhardae* homologues are consistent with orthologues in other flagellated species. To visualise the orthology between *P. lacertae* or *C. burkhardae* proteins and matches from other flagellated organisms, a Maximum-Likelihood phylogeny was estimated from an alignment of each query protein. The phylogeny was estimated with PhyML[75] after automated model selection using the Akaike Information Criterion. Default settings for tree-searching were employed, and 100 non-parametric bootstrap replicates were applied to assess node robustness. Given that we used human protein sequences to initially screen the *Blastocystis* gene set, and that *Blastocystis* and human are separated by a large phylogenetic distance, we note that the absence of conserved flagellar protein genes in *Blastocystis* is not changed when *Proteromonas*/*Cafeteria* orthologues are used as search queries instead. Cross-checking their gene names in Table S1 (highlighted in yellow shading) shows that no *Blastocystis* orthologues were identified by OrthoMCL, although paralogues for KIF3C are present, as noted above.

We used a large-scale dataset of flagellar associated proteins generated previously[30] for their investigation of stramenopile flagella (Tables S4A–S4F). As this dataset included 592 proteins that had been implicated as acting in the flagellum, but were not necessarily exclusive to the organelle, we instituted a series of bioinformatic filters aimed at identifying strong candidates for exclusive action in our structures of interest. The dataset was first searched against the *Blastocystis* genome as a negative control. Any proteins present in *Blastocystis* were removed as likely acting in other cellular processes. The resulting 236 protein dataset was then searched against a curated dataset of stramenopile genomes and transcriptomes. All 236 proteins were searched *via* the AMOEBAE bioinformatic workflow[76] that incorporates BLAST analysis against predicted proteins and nucleotide contigs, as well as HMMer analyses, and applies a reciprocal best hit e-value cut-off.

To increase the selection for proteins likely acting exclusively at flagella, we filtered the resulting dataset to identify proteins present in 7 of the 11 organisms possessing tinselated flagella, to account for possible false negatives in the genomes and/or transcriptomes, but importantly absent in the flagellum-lacking *P. tricornutum*. This yielded 116 candidates, including 37 previously annotated as flagellar associated (dynein, kinesin, IFT, radial spoke, flagellar) and 54 conserved unknown proteins. To identify candidate mastigoneme proteins, these 116 proteins were filtered to identify those missing in the combined *Halocafeteria seosinensis* transcriptome and genomic datasets. Of the 37 such proteins, only 7 were missing in all three organisms lacking tinselated flagella (*P. lacertae*, *H. seosinensis*,[32] and *Incisomonas marina*[31]), with the large remainder being present in *P. lacertae*, but missing from the other two.

It is worth reiterating that the list of candidate proteins for flagellum and mastigonemes that we have generated is intentionally non-exhaustive, as it excludes any proteins that may have redundant cellular functions or localizations, and could have been repurposed and thus retained in taxa that have lost the traits of interest. For example, only 236 of the 592 flagellar associated proteins reported previously,[30] were retained since their presence in the flagellum-lacking taxa strongly suggests promiscuous or redundant function, but this in no way invalidates their flagellar functions. Similarly, the three known mastigoneme-associated proteins were rejected by our filters due to their presence in *H. seosinensis*. This notwithstanding, our analysis has generated a list of 37 candidate proteins for investigation as to their involvement in mastigonemes. Moreover, the fact that 30/37 such candidates are present in *P. lacterae* but absent in the other two nontinselated taxa is consistent with the hypothesized homology between somatonemes and mastigonemes, the first such molecular evidence brought to bear on this argument.

Membrane-trafficking proteins

Rabs identified previously in *Phytophthora sojae*[34] served as queries in BLASTp and tBLASTn searches against *P. lacertae* predicted proteins and transcriptome, respectively. All identified hits above E-value threshold of 1e-04 were subjected to reverse BLAST against home-built database of GTPases and NCBI nonredundant database. Only those (Figure S1B) that recovered Rab as their best blast hit in reverse search in at least one of the databases were added to the dataset from a previous analysis.[34] Rabs were aligned using MAFFT v7.458[77] under L-INS-I strategy with a maximum of 1,000 iterations and poorly aligned positions were removed with trimAl v1.4[78] using -gt 0.5 option. Maximum-Likelihood trees (Data S1) were inferred using the LG+C20+F+G model, the

posterior mean site frequency method,[97] and LG+F+G guide tree in the IQ-TREE v1.6.12.[79] The strategy of rapid bootstrapping with a "thorough" maximum likelihood search with 1,000 bootstrap replicates was employed.

Homology searching was used to identify membrane-trafficking (Figure S3) and peroxisome biogenesis genes in *P. lacertae*. Functionally characterized genes from human and yeast were used as query sequences in BLASTp searches, and sequences that were retrieved with an E-value less than 0.05 were then used to search the genome of the query organism. The query sequence or a clear orthologue must be identified as the top hit with an E-value less than 0.05 for the candidate to be considered a true homologue. In cases where a homologue could not be identified by BLASTp, the genome was searched using tBLASTn. Additionally, if a homologue was previously identified in *Blastocystis* sp. or another close relative but not in *P. lacertae*, this sequence was used as a BLASTp or tBLASTn query. HMMER v3.1b1[80] searches were performed to identify highly divergent TSET homologues using sequences listed previously.[98]

For highly paralogous protein families, orthology was confirmed by phylogenetics (Data S2). Sequences were aligned using MUSCLE v3.8.31,[81] visualized using Mesquite v3.03,[82] and then manually masked and trimmed to include only homologous positions. Masked alignments are available upon request. ProtTest v3.4[83] was used to determine the best-fit model of sequence evolution. PhyloBayes v3.3[84–86] and MrBayes v3.2.2[87,88] were used for Bayesian inference of phylogeny, and RAxML v8.1.3[89] was used for Maximum-Likelihood analyses. PhyloBayes was run until the largest discrepancy observed across all bipartitions was less than 0.11 and at least 100 sampling points were achieved, MrBayes was used to search treespace for a minimum of one million MCMC generations, sampling every 1,000 generations, until the average standard deviation of the split frequencies of two independent runs (with two chains each) was less than 0.01. Consensus trees were generated using a burn-in value of 25%, well above the likelihood plateau in each case. RAxML was run with 100 pseudoreplicates.

Peroxisomal proteins

Peroxins from *Thalassiosira pseudonana* and *Phytophthora sojae* from KEGG pathway (04146) served as queries in tBLASTn searches against transcriptomic assembly of *C. burkhardae*. Identified *C. burkhardae* sequences then served as queries for search in *P. lacertae* and *Blastocystis* predicted proteins and/or genomes. Protein domains of found proteins were identified by InterProScan[63] implemented in Geneious Prime v2020.2.5.[99] *Homo sapiens* and *Phytophthora ramorum* Pex sequences (downloaded from the NCBI protein database; last accessed 16.8.2022) were used as queries in the searches in selection of stramenopiles genomes and transcriptomes by AMOEBAE workflow.[76]

Peroxisomal targeting signals 1 (PTS1) and 2 (PTS2) were defined as [SAC]-[KRHQ]-[LM]$[100] and ^x{2,22}-[R]-[LIVQ]-xx-[LIVQH]-[LSGA]-x-[HQ]-[LA],[101] respectively. Proteins bearing PTS1 or PTS2 were identified in *P. lacertae* by an in-house python script (https://github.com/kikinocka/ngs/blob/master/py_scripts/pts_search.py) and following BLASTp searches against NCBI nonredundant database. Mitochondrial predictions were assessed by TargetP-2.0[90] and NommPred[91] in mitochondrial, MRO, and stramenopile-specific settings.

Pex19 sequences from selected eukaryotes were aligned using built-in aligner in Geneious Prime v2020.2.5. Regions corresponding to Pex3 and Pex10 binding domains in *Pichia pastoris*[102] were extracted from the alignment and percentage identities were directly identified as percentage of residues that were identical.

Mitochondrial and metabolic predictions

The mitochondrial complement and metabolic enzymes of *P. lacertae* were annotated by use of reciprocal BLAST searches. Briefly, the mitochondrial proteomes of *H. sapiens*, *Mus musculus*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae*, *Trypanosoma brucei*, *Giardia lamblia* and *Blastocystis* ST1 were used as queries for BLAST searches against the *P. lacertae* proteome. Putative candidate orthologues were then used as queries in reciprocal BLAST searches against the whole proteome of the respective organism. A comparison script was used to compare the two outputs for a true hit. The positive matches from each of these were compiled and the consensus annotation was assigned.

Genes of interest were further validated using BLAST and alignment tools to check annotations were correct. Using this approach 1,100 genes were identified as potentially mitochondrial. A list of known anaerobic enzymes from microbial anaerobic eukaryotes was individually curated (length and position of Pfam domains) and used to construct predicted metabolic pathways.

Microscopy

Antibody production

For Pex10 (PIPex10_16-30-2111130-KHL-MBS; H-CNN KIS RKR KHV DDD M) and PXMP2 (Plac_PXMP2-2010476-KLH-MBS; C+KLIERNKSFDKRRSF) peptides were designed by Eurogentec (Belgium) based on the predicted proteins, and antibodies were generated using the Mini Speedy Program (28 days) for two Rats and one Rabbit respectively.

Medium and final bleeds were tested using Elisa (1 animal/plate with 2 Ag per plate) following the manufacturer's protocol. PXMP2 antisera from one animal was also purified.

We utilised two different heterologous antibodies for the Pex19 experiments:

1. *H. sapiens* anti-Pex19 rabbit polyclonal antibody (ProteinTech Europe; 14713–1-AP).
2. *A. thaliana* anti-Pex19 rabbit polyclonal antibody[41] kindly provided by Prof. Bonnie Bartel.

Transmission and immuno-electron microscopy

Three tubes of *P. lacertae* culture (4–10 days post passage) were pelleted at 800 x g for 10 minutes at room temperature. The supernatant was discarded, and each pellet was resuspended in 2 ml of 2.5 % glutaraldehyde (Sigma Aldrich) in 100 mM sodium cacodylate (Sigma Aldrich) buffer pH 7.2 and left to fix for two hours at room temperature. Following fixation, the sample was pelleted at 1,900 x g for two minutes and washed twice in cacodylate buffer for 10 minutes to remove the fixative. Once pelleted, the buffer was discarded and the pellet was resuspended in 500 μl of the same buffer, and subsequently 50 μl was transferred into another tube and warmed in a 55 °C water bath for five minutes. 50 μl of 3% low melting point agarose was added to the cells, and using a glass pipette, the mixture was quickly transferred into previously made gaskets (plastic cut and sandwiched between two glass slides to allow the gel to form a thin layer which were clamped together) and stored at 4 °C for 5–10 minutes until the gel had set. Once removed from the fridge, the gel was cut into very thin pieces and transferred into a drop of Alcain blue-0.1 % acetic acid dye Alcain blue, after which it was gently removed from the dye using a bent toothpick and placed in 3 ml of buffer to remove excess dye. Using a glass pipette, the buffer was removed, and care was taken not to remove gel fragments. 1–1.5 ml OsO_4 (made up from 1 ml 4% OsO_4, 1 ml milli-Q water and 2 ml 200 mM cacodylate buffer) was added and the sample was left at room temperature for 1 hour. Following this step, OsO_4 was discarded, and the sample was washed once for 10 minutes in 50% ethanol and was left overnight in 70% ethanol at 4 °C.

The following day, the sample was washed once in 90% ethanol and then three times in 100% ethanol. Following this wash step, the ethanol was discarded, and the fragments were washed twice in 3 ml propylene oxide. This was removed and 50% propylene oxide/50% low viscosity (LV) resin (12 g LV resin, 4 g VH1 hardener, 9 g VH2 hardener and 0.63 g LV accelerator) was added and left for 30 minutes at room temperature. 50/50 mix was removed and 100% LV resin mixture was added and left for 90 minutes. Following this, 10-12 fragments were transferred into fresh LV and left for another 90 minutes. Using a Pasteur pipette, 6 ml LV resin was put in a small mould and fragments were placed a small distance from the edge of the mould and gently pushed to the bottom using a bent toothpick. They were then placed in a 60 °C oven for 20-24 hours in preparation for sectioning. To section, the mould was cut where the cells were most concentrated (following light microscopy) and were superglued onto blank resin capsules where it was filed down, and the edges of the capsule were cut away using a glass knife to leave the mould raised. The knife has a boat at the back which was filled with milli-Q water. The automated knife was used to cut very thin (few microns thick) slices from the block, which would stack in the water. To expand them, they were exposed to chloroform vapours. Once enough had been collected, roughly seven slices were attached to a slot grid coated in plastic.

To stain, a rectangle of dental wax with labelled columns was covered in milli-Q water and then sealed in parafilm. In each of the columns, a drop of 4.5% uranyl acetate was placed at the top of each, then below that a drop of milli-Q water, then another below that. The slot grid was placed on the uranyl acetate to stain for 45 minutes, then washed gently under milli-Q then placed on the drop of milli-Q, and repeated. It was dried using filter paper. On another grid of wax, which was also wrapped in parafilm, two drops of milli-Q were placed below a drop lead acetate (in this container, the space was filled with potassium hydroxide). The slot grid was placed in the lead for 7 minutes and transferred to the first, then second milli-Q drop, dried on filter paper, and left for a short while underneath a light (while being kept in the air by forceps).

For immuno-electron microscopy (IEM), aspirated cultures of *P. lacertae* were fixed for 1 hour in freshly prepared phosphate buffered saline (PBS) solution containing 4% formaldehyde and were then washed several times with PBS. IEM samples were suspended in LR white resin (Agar Scientific). Resin permeation was aided by placing the samples in a vacuum for 2 minutes. The resin was then aspirated and replaced with fresh resin and the samples transferred into gelatine (Agar Scientific) capsules and hardened for 15 hours in a pre-warmed 60 °C oven. The hardened blocks were then polished and subsequently sectioned by ultra-microtome at a thickness of 70 μm, then placed on gold EM grids with approximately five sections per grid. Immuno-staining of the IEM grids was performed in humidifying chambers. Blocking of the samples was achieved via a 1-hour incubation in 2% bovine serum albumin in PBS with 0.05% Tween 20. Primary antibody binding was performed by 15-hour incubations with the Pex19 antibody, at three dilutions (1:10, 1:20 and 1:50) at 8 °C. The IEM grids were subsequently incubated for 30 minutes at room temperature, with the corresponding gold-conjugated secondary antibodies. Counter-staining was achieved via a 15-minute incubation with 4.5% uranyl acetate in PBS and a 2-minute incubation in Reynold's lead citrate.

Both TEM and IEM grids were imaged in a Jeol 1230 Transmission Electron Microscope operated at 80 kV and images were captured with a Gatan One view digital camera.

Scanning electron microscopy (SEM)

Specimens of *Blastocystis* and *P. lacertae* were prepared for scanning electron microscopy (SEM) from cultures (*Blastocystis*: xenic culture from Betts et al.[103]; *P. lacertae*: axenic culture in LYI-S-2 medium + adult bovine serum). Specimens were deposited with a pipette from the culture tubes into hand-made baskets [top end of a 1,000 μl pipette tip fixed with silicon to a 5 μm polycarbonate membrane filter (Millipore Corp.)] and placed in 12-well culture plates filled with PBS. A piece of Whatman No. 1 filter paper was mounted on the lid of the well plates and saturated with 4% (w/v) OsO_4. The lid was closed on the well plate and the specimens were fixed by OsO_4 vapours for 30 minutes in the dark. Five drops of 4% (w/v) OsO_4 were added directly to the basket and the specimens were fixed for an additional 30 minutes. The filters were washed with water and dehydrated with a graded series of ethanol. Filters were critical point dried with CO_2, mounted on stubs, sputter coated with 5 nm of platinum, and viewed using a scanning electron microscope Hitachi S-4300 (Hitachi, Tokyo, Japan).

Immunofluorescence microscopy

P. lacertae cultures were transferred in 15 ml tubes and pelleted at 800 x g for 8 minutes. Media was discarded and cells were incubated for 20 minutes with 2 nM of MitoTracker Red CMXRos (Invitrogen; optional) and then fixed with 2% formaldehyde for 20 minutes followed by permeabilization with 0.1% Triton-X in 1× PBS for 10 minutes. Cells were then aliquoted on poly-L-lysine (Sigma-Aldrich) cotted slides and left for 2 hours. After blocking for 1 hour in 5% skimmed milk in 1× PBS, the cells were probed with the rabbit anti-Pex19 (1:200) and/or rabbit anti-PXMP2 (1:400) antisera and/or rat anti-Pex10 (1:200) or rat anti-*Bh*SufCB (1:100[104]). Secondary Alexa Fluor 488-conjugated goat anti-rabbit IgG (H-L), Alexa Fluor 488-conjugated chicken anti-rat IgG, and Alexa Fluor 594-conjugated donkey anti-rat IgG (Molecular Probes) were used at a dilution of 1:1,000. Cells were mounted with DAPI-containing anti-fade mounting reagent (Vectashield) and observed using Airyscan imaging mode with the laser scanning Zeiss LSM 880 confocal microscope. Images were collected using Zeiss Zen Black software for confocal microscope and processed with ImageJ. For 3D image processing, Zen Blue software was used.

Protein extraction and western blotting

Two fully grown 15 ml tubes of *P. lacertae* culture were pelleted in 15 ml tubes at 1,000 x g for 10 minutes. The supernatants were discarded, the pellets were merged and resuspended in 8 ml of PBS, with 1 ml DNAase/RNAase mix (50 mM $MgCl_2$, 0.5 M Tris-HCl pH 7.0) and 10 μl of EDTA free protease inhibitor; the whole solution was kept on ice throughout. The mixture was then pelleted again at 1,000 x g for 10 minutes. Supernatant was discarded, and the pellet was resuspended in 500 μl PBS/10% DNAase/RNAase mix and 10 μl protease inhibitor was added. The sample was transferred to a 1.5 ml tube and was centrifuged at 4,500 x g for 2 minutes. The supernatant was discarded, and the pellet was resuspended in 400 μl 4× Laemmli Protein Sample Buffer (Biorad) and passed through a fine syringe. The sample was boiled on a 95 °C for 5 minutes. Following boiling, the sample was pelleted at 12,000 x g for 3 minutes and stored at -20 °C until used.

Total protein extract from *P. lacertae* suspended in 4× Laemmli Protein Sample Buffer (Biorad) were heated at 95 °C for 5 minutes, and 15 μl was loaded onto a 12% SDS-PAGE gel for protein separation. For immunoblotting, proteins were electrophoretically transferred into a 0.2 μm PVDF membrane. Membranes were blocked with Blocking Buffer (5% (w/v) skimmed milk in 1× TBS-T Buffer [8.76 g NaCl, 6.06 g Tris, 0.1% (v/v) Tween 20, pH 7.6]) for 1 hour at room temperature. Dilutions of anti-Pex10 and anti-PXMP2 serum (1:200 and 1:500), and anti-Pex19 and anti-PXMP2 purified antibody (1:200) were prepared in Blocking Buffer and incubated with membrane overnight at 4 °C. After washing the membranes three times with 1× TBS-T buffer, the appropriate HRP-conjugated secondary antibody (anti-rabbit for PXMP2 and Pex19, and anti-rat for Pex10) were added at a 1:10,000 dilution. After another round of washing with 1× TBS-T, membrane was developed using ECL reagents as previously described.[59]

QUANTIFICATION AND STATISTICAL ANALYSIS

Average % identity for Pex3 and Pex10 binding domains of Pex19 were calculated by AVERAGE function in Microsoft Excel. For *P. lacertae* and *C. burkhardae*, only % identities with the rest of the organisms and excluding each other were considered. For the remaining organisms, % identities among them were considered.

For IEM, each generated image was divided in grids using the ImageJ (v1.51) software followed by manual quantification of the gold particles per μm^2. Plots were generated using Microsoft Excel for Mac (v16.72). Each plot represents the average of gold particles per μm^2 per compartment, with error bars indicating the calculated standard deviation (SD).

Current Biology, Volume *33*

Supplemental Information

Evolutionary analysis of cellular reduction and anaerobicity in the hyper-prevalent gut microbe *Blastocystis*

Kristína Záhonová, Ross S. Low, Christopher J. Warren, Diego Cantoni, Emily K. Herman, Lyto Yiangou, Cláudia A. Ribeiro, Yasinee Phanprasert, Ian R. Brown, Sonja Rueckert, Nicola L. Baker, Jan Tachezy, Emma L. Betts, Eleni Gentekaki, Mark van der Giezen, C. Graham Clark, Andrew P. Jackson, Joel B. Dacks, and Anastasios D. Tsaousis

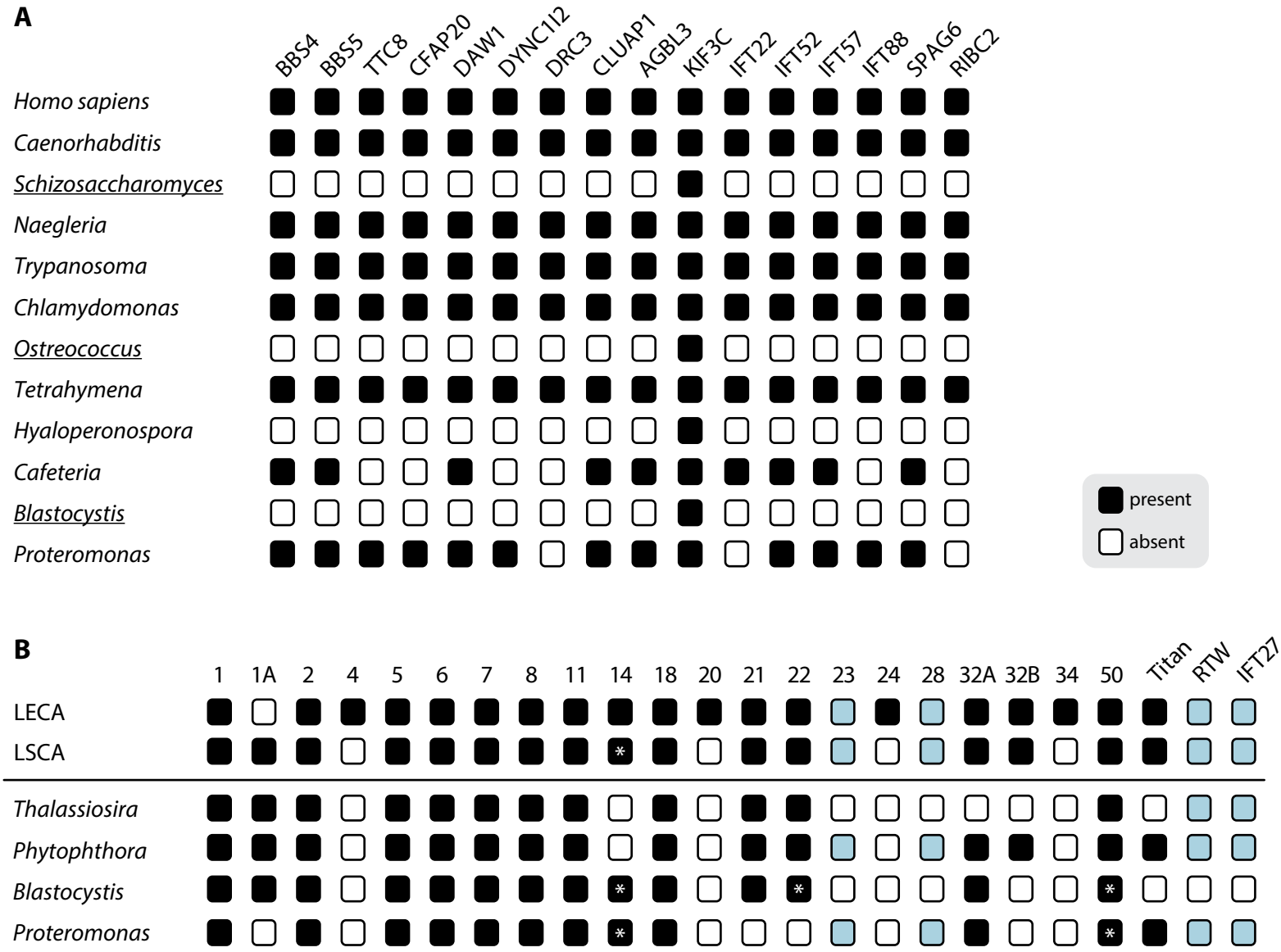


Figure S1. Presence/absence of flagellar (A) and Rab (B) proteins in P. lacerate and selected species, Related to Figure 2B, STAR methods, Data S1. (**A**) Presence or absence of 16 proteins conserved among all flagellated eukaryotes according to [S1]. These were identified in five flagellated organisms of diverse affinity (see Methods) as well as the Blastocystis ST7 genome (ASM15166v1) and the P. lacerate genome and C. burkhardae transcriptome produced in this study. A reciprocal best match by BLASTp between the human protein query and the subject protein in the non-human genome was required to confirm that an orthologue was 'present' in the latter. Non-flagellated organisms are underlined. (**B**) Repertoire of Rab GTPases in P. lacerate, and the deduced Rab set in the last stramenopile common ancestor (LSCA). Rab identity in studied species was confirmed by phylogenetic analysis. Asterisks denote Rabs that could not be unambiguously assigned. Rabs involved in flagellar function are highlighted by pale-blue. The Rab complement of LSCA was compared with that of the last eukaryotic common ancestor (LECA) [S2].

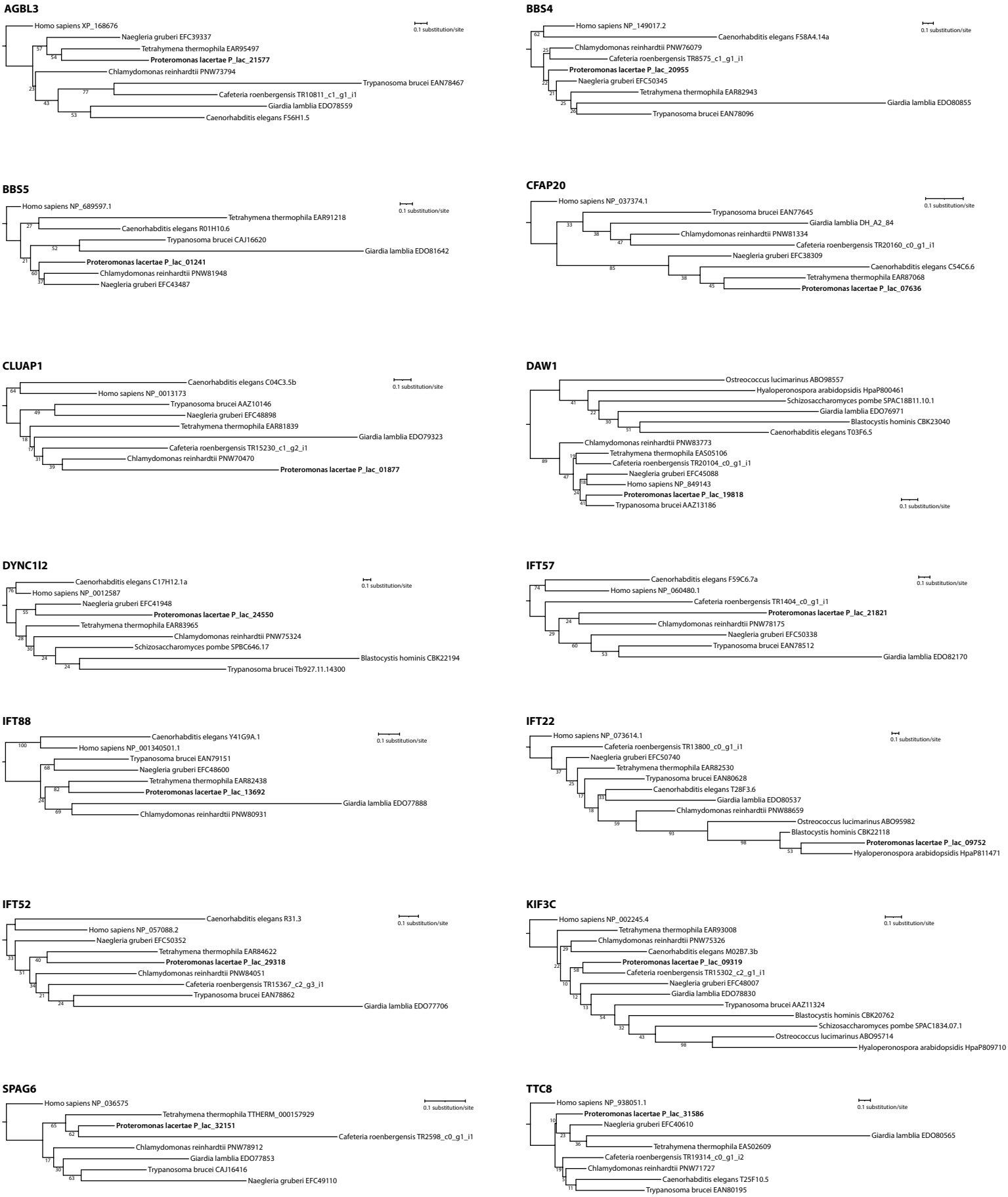


Figure S2. Phylogenetic trees of flagellar proteins, Related to STAR methods. The Maximum-Likelihood phylogenetic trees were estimated from amino acid sequence alignments under a LG+F+Γ model with 100 bootstrap replicates using RAxML [S3], except for thee CFAP21 tree, which was estimated from a nucleotide alignment using a GTR+Γ model. The trees are drawn as phylograms for convenience but are arbitrarily rooted.

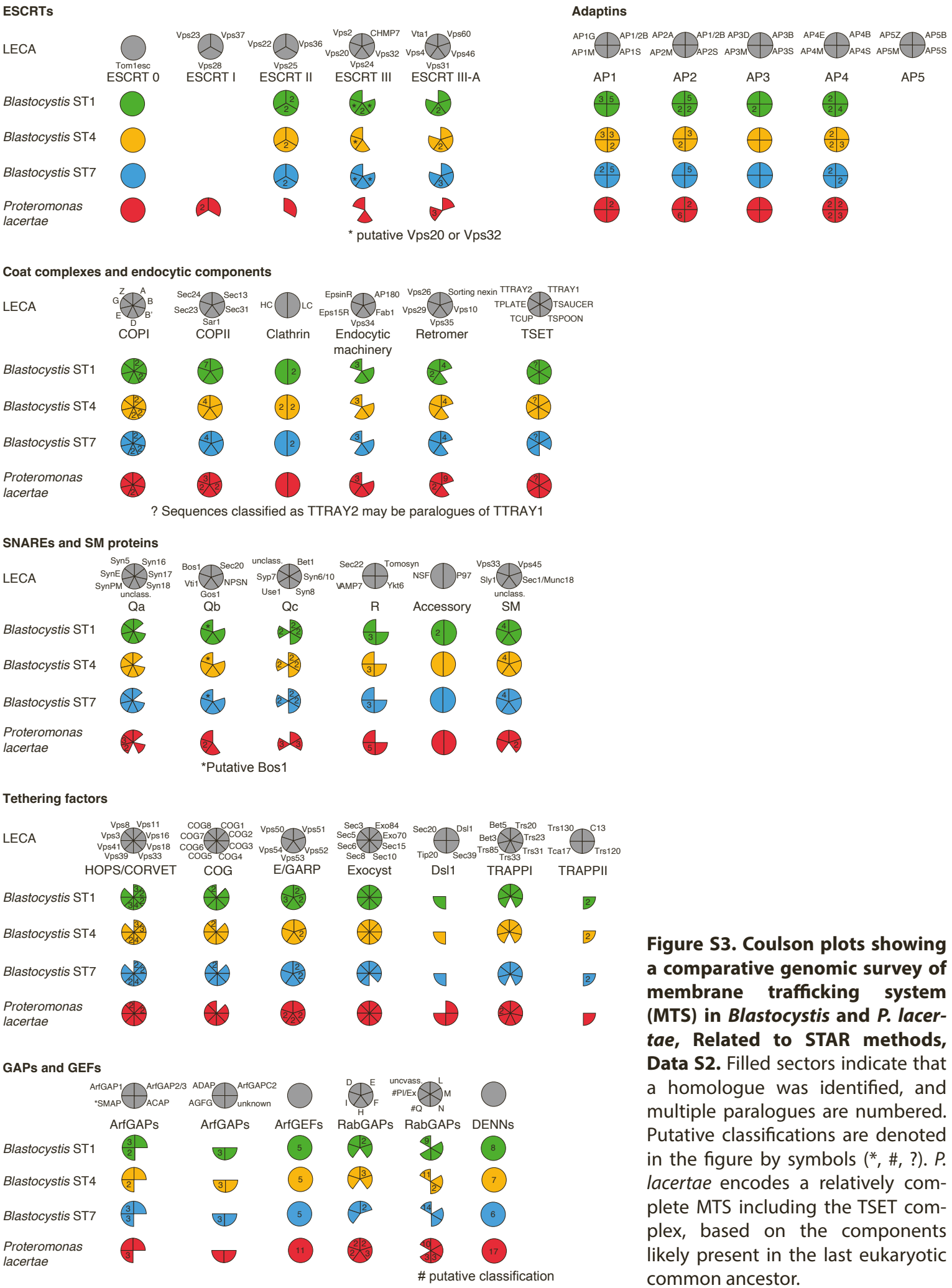


Figure S3. Coulson plots showing a comparative genomic survey of membrane trafficking system (MTS) in *Blastocystis* and *P. lacertae*, Related to STAR methods, Data S2. Filled sectors indicate that a homologue was identified, and multiple paralogues are numbered. Putative classifications are denoted in the figure by symbols (*, #, ?). *P. lacertae* encodes a relatively complete MTS including the TSET complex, based on the components likely present in the last eukaryotic common ancestor.

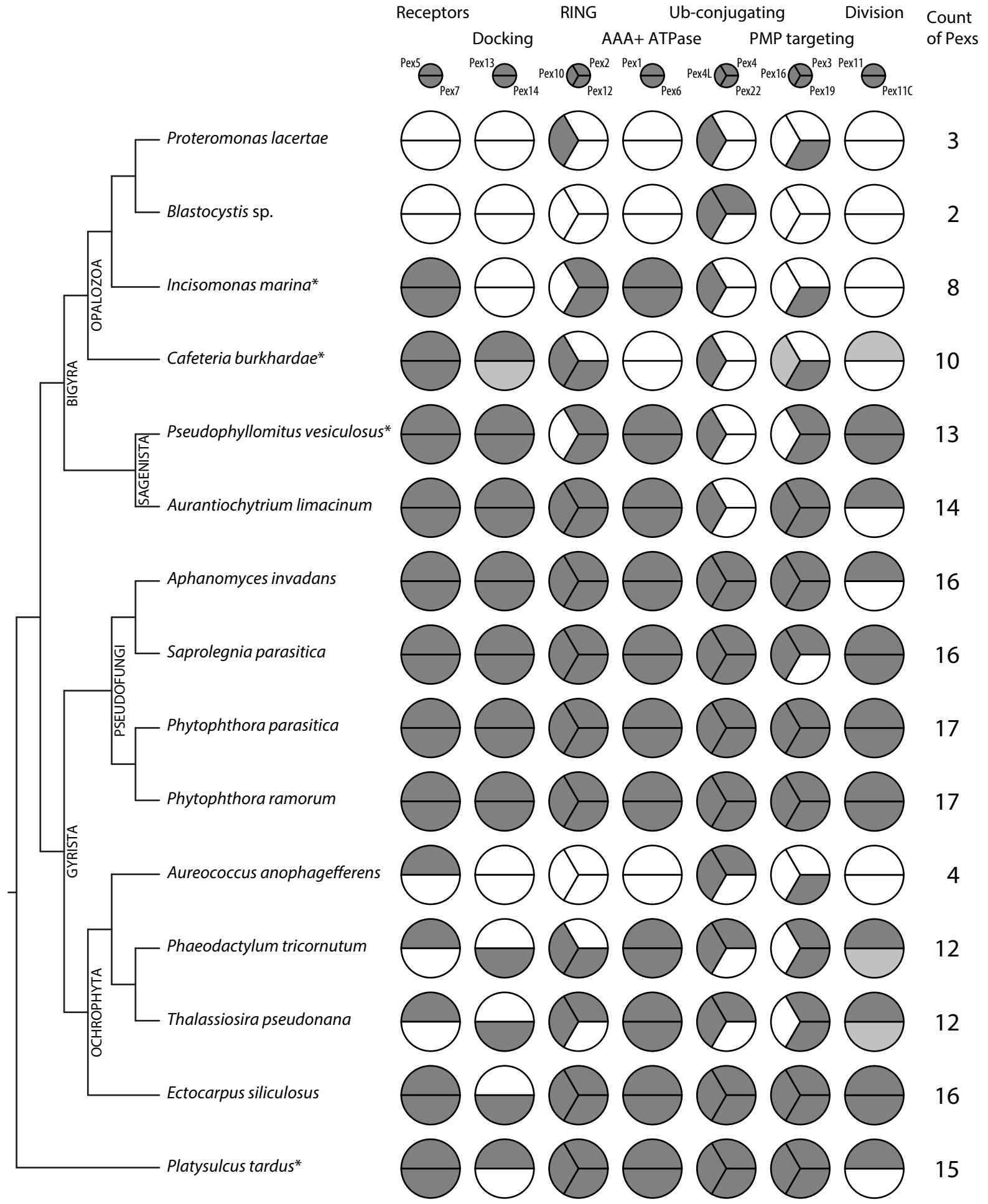


Figure S4. Comparative analysis of peroxisomal complement in stramenopiles, Related to Figure 3A. Pattern is consistent with progressive degeneration of peroxisomal complement in Bigyra, contrasted with conservation in other stramenopile lineages. Presence and absence of proteins is shown by dark grey and white color, respectively. Light grey color shows proteins that were found by manual searches. Asterisks mark species for which only transcriptomes were available.

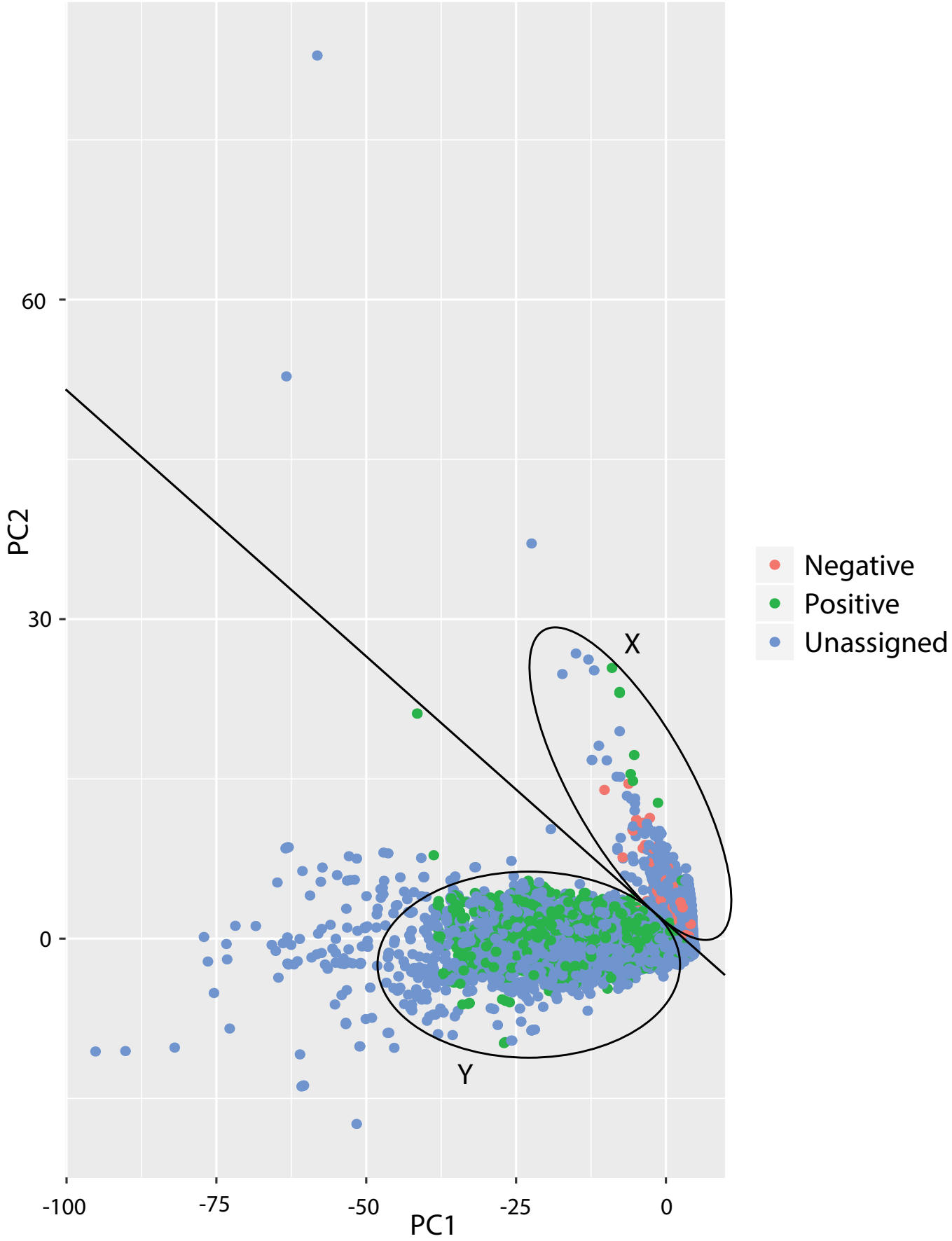


Figure S5. Principal component analysis for kmer frequencies of all transcripts, Related to STAR methods. Positive and negative control groups were clearly distinguished by the analysis, and thus the remaining transcripts were assigned as eukaryotic or prokaryotic.

Supplemental references

S1 Judelson, H.S., Shrivastava, J., and Manson, J. (2012). Decay of genes encoding the oomycete flagellar proteome in the downy mildew *Hyaloperonospora arabidopsidis*. PLoS One *7*, e47624.

S2 Elias, M., Brighouse, A., Gabernet-Castello, C., Field, M.C., and Dacks, J.B. (2012). Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. J. Cell Sci. *125*, 2500–2508.

S3 Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.