Anaeramoebae are a divergent lineage of eukaryotes that shed light on the transition from anaerobic mitochondria to hydrogenosomes

Courtney W Stairs, Petr Táborský, Eric D Salomaki, Martin Kolisko, Tomáš Pánek, Laura Eme, Miluše Hradilová, Čestmír Vlček, Jon Jerlström-Hultqvist, Andrew J Roger, et al.

TITLE

Anaeramoebae are a divergent lineage of eukaryotes that shed light on the transition from anaerobic mitochondria to hydrogenosomes

AUTHORS

Courtney W. Stairs[1], Petr Táborský[2], Eric D. Salomaki[3], Martin Kolisko[3], Tomáš Pánek[2], Laura Eme[4], Miluše Hradilová[5], Čestmír Vlček[5], Jon Jerlström-Hultqvist[6,7], Andrew J. Roger[7], and Ivan Čepička[2]

AFFILIATIONS

1. Department of Biology, Lund University, Sölvegatan 35, 22 362, Lund, Sweden
2. Department of Zoology, Faculty of Science, Charles University, Viničná 7, 128 44 Prague, Czech Republic
3. Institute of Parasitology, Biology Centre, Czech Acad. Sci., České Budějovice, 370 05, Czech Republic
4. Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, Orsay, France
5. Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Prague, Czech Republic
6. Current address: Department of Cell and Molecular Biology, BMC, Box 596, Uppsala Universitet, Sweden, SE-751 24 Uppsala, Sweden
7. Centre for Comparative Genomics and Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Canada

Lead contact: CWS
courtney.stairs@biol.lu.se
Corresponding authors: CWS, IC
courtney.stairs@biol.lu.se, ivan.cepicka@centrum.cz
Senior authors: AJR, IC

SUMMARY

Discoveries of diverse microbial eukaryotes and their inclusion in comprehensive phylogenomic analyses have crucially re-shaped the eukaryotic tree of life in the 21st century[1]. At the deepest level, eukaryotic diversity comprises 9-10 'supergroups'. One of these eukaryotic supergroups, the Metamonada, is important to our understanding of the evolutionary dynamics of eukaryotic cells including the remodeling of mitochondrial function. All metamonads thrive in low-oxygen environments and lack classical aerobic mitochondria instead possessing mitochondrion-related organelles (MROs) with metabolisms that are adapted to life without oxygen. Theses MROs lack an organellar genome, the Krebs cycle and oxidative phosphorylation[2] and often synthesize ATP by substrate-level phosphorylation coupled to the production of hydrogen[3,4]. The events that occurred during the transition from an oxygen-respiring mitochondrion to a functionally streamlined MROs early in metamonad evolution remain largely unknown. Here, we present the transcriptomes of two recently described, enigmatic anaerobic protists from the genus *Anaeramoeba*[5]. Using phylogenomic analysis, we found that these species represent a divergent, phylum-level lineage, in the tree of metamonads, forming a sister group of the Parabasalia, crucially reordering the deep branching order of the metamonad tree. Metabolic reconstructions of the *Anaeramoeba* MROs revealed many 'classical' mitochondrial features previously not seen in metamonads, including a disulfide relay import system, propionate production, and amino acid metabolism. Our findings suggest that the cenancestor of Metamonada likely had an MRO with more classical mitochondrial features than previously anticipated and demonstrate how discoveries of novel lineages of high taxonomic rank continue to transform our understanding of early eukaryote evolution.

HIGHLIGHTS 4 X 85 CHAR

- Deep transcriptomic sequencing of two new anaerobic metamonads
- The Anaeramoebae represents a new principal lineage of Metamonada
- Anaeramoebae are free-living relatives of the parasitic parabasalids
- Retention of mitochondrial features in mitochondrion-related organelles

ETOC BLURB 350 CHAR

Stairs et al. report a new phylum-level lineage of anaerobic eukaryotes named the Anaeramoebae. These free-living anaerobes represent a new branch on the tree of life sister to the parasitic parabasalids, including Trichomonas vaginalis. The Anaeramoebae have complex mitochondrial proteomes with features not found in their closest relatives.

RESULTS

Anaeramoebae reshuffle the phylogeny of Metamonada

To deepen our understanding of the phylogenetic placement and metabolic potential of the Anaeramoeba species we sequenced the transcriptomes of two species of *Anaeramoeba*[5] (Figure 1A, 1B). To place the *Anaeramoeba* species on the tree of eukaryotes, we performed a phylogenomic analysis with representatives from all major eukaryote lineages (Figure 1C, S1). We found that the two *Anaeramoeba* species form a monophyletic group (herein referred to as Anaeramoebae) within Metamonada sister to the clade of Parabasalia with full support in maximum likelihood (ML) analyses (Figure 1D, S1) and in Bayesian analysis (see[6]).

Metamonada is represented by three major lineages (phyla) - Preaxostyla (e.g., *Trimastix*, *Monocercomonoides*), Fornicata (e.g., *Carpediemonas*, *Chilomastix*, *Giardia*), and Parabasalia (e.g., *Trichomonas*, *Tritrichomonas*, *Trichonympha*)[7]. Recently, *Barthelona*, a genus of free-living flagellates, was shown to belong to Metamonada as well, forming a sister group to Fornicata[8]. To date, nearly all phylogenetic investigations recover a sister-group relationship between Parabasalia and Fornicata to the exclusion of Preaxostyla, including a recent analysis that placed barthelonids as sister to the Fornicata with 100% bootstrap support[8]. This is despite the fact that this relationship could potentially represent a long branch attraction (LBA) artefact, as species of Fornicata and Parabasalia form some of the longest branches in phylogenetic trees of eukaryotes[3,9–11]. In the analyses presented here, adding the *Anaeramoeba* species, which are amongst the shortest branches of any known metamonads, breaks the long stem branch leading to the Parabasalia group.

We tested whether alternative hypotheses could be rejected by the data using approximately unbiased (AU) topology tests. The topology constraining the monophyly of Parabasalia+Fornicata+*Barthelona*, was rejected (p=0.00311; Data S1) suggesting that this previously reported relationship is unlikely to be correct. Topologies constraining the monophyly of Parabasalia+Fornicata+*Anaeramoeba*+*Barthelona* (p=0.0263), Discoba+Metamonada+*Collodictyon*+*Malawimonas* (p=0.00263), and the Phylobayes consensus topology (*Malawimonas*+Fornicata+*Barthelona*+Preaxostyla; p=3.66 x 10^{-5}) were all rejected.

We hypothesize that the clade comprising Preaxostyla, *Barthelona*, and Fornicata recovered herein, is the result of improved taxon sampling and the use of a sophisticated substitution model for phylogenetic inference (*i.e.,* the LG+C60+F+G model, Figure 1) that together have offset an LBA artefact previously grouping Parabasalia with *Barthelona* and Fornicata, to the exclusion of Preaxostyla. Thus, the inclusion of Anaeramoebae in phylogenomic analyses fundamentally changes the known relationships within Metamonada[3,9–11].

Despite their relatedness, anaeramoebids and parabasalids have dissimilar morphologies

The close relationship between Parabasalia and Anaeramoebae is surprising because members of these two groups are quite dissimilar on a morphological and ultrastructural level[5]. Anaeramoebids are predominantly amoeboid organisms with cell structures and motility reminiscent

of certain members of the supergroup Amoebozoa[5]. Their cells contain an acentriolar centrosome, from which sparse, individual microtubules radiate. In contrast, most parabasalids are flagellates with intricate cytoskeletons composed of large microtubular and non-microtubular networks. Although several parabasalids (e.g., *Dientamoeba fragilis* and *Trichomonas vaginalis*) have amoeboid character, these species form terminal branches in the phylogenetic tree of Parabasalia and we assume that their ability to move using pseudopodia has arisen independently from that of anaeramoebids[12]. Two out of six known *Anaeramoeba* species, including *A. ignava*, have been shown to have a flagellate stage[5]. The flagellates display a unique morphology that is not readily comparable with the morphology of other flagellate protists, in particular Parabasalia[5]. Moreover, Parabasalia possess a conspicuous stacked Golgi apparatus, the so-called parabasal body[12]. In contrast, we failed to observe a "stacked" Golgi apparatus in anaeramoebids using several microscopical techniques[5]. Since the anaeramoebids are fundamentally distinct from the three main metamonad phyla[7], we propose classifying these species in a new phylum, Anaeramoebae[5].

Anaeramoebid MROs have the most complete protein import machinery among metamonads

Previous microscopic examinations presented evidence that anaeramoebids possess MROs[5] (Figure 1A, B insets). We therefore sought to reconstruct the mitochondrial proteome from the transcriptomic data. We identified 102 and 122 proteins that are predicted to function in the MROs of *A. ignava* and *A. flamelloides*, respectively, based on homology to known mitochondrial or MRO proteins and/or predicted mitochondrial localization (Data S1). We found no evidence of transcripts encoding proteins related to mitochondrial genome (mtDNA) maintenance or expression, suggesting that, like all other MROs from metamonads, anaeramoebid MROs lack mtDNA. Below, we focus our discussion on the pathways that have not been observed in metamonads before (Figure 2, blue and orange shaded circles).

In eukaryotes, most mitochondrion- or MRO-targeted proteins are transcribed from the nuclear genome, translated in the cytoplasm and imported into mitochondrial compartments by the translocase of the outer mitochondrial membrane (TOM) and translocase of the inner mitochondrial membrane (TIM) following recognition of N-terminal or internal targeting signals. We observed that anaeramoebids have a similar complement of TOM and TIM components previously described in *Trichomonas vaginalis*[13] (Figure 2A, 3).

We uncovered components of the intermembrane space (IMS) disulfide relay system that has never been identified in any metamonad. In model eukaryotes, the disulfide relay system is necessary for the import and folding of some proteins destined for the intermembrane space[14]. Mia40 is an IMS oxidoreductase that, together with the sulfhydryl oxidase Erv1, generates disulfide bonds into its substrate proteins[14,15]. Under aerobic conditions, yeast Mia40-Erv1 transfers electrons from substrate proteins to oxygen or to cytochrome *c*. However, under anaerobic conditions, electrons can be transferred to fumarate via the IMS-localized fumarate reductase Osm1[16]. In yeast, the sub-cellular localization of Osm1 is mediated by alternative translation start sites that generate protein products with ER or mitochondrial targeting signals[17], intrinsic features of the protein[16] or redox state of the

cell[16]. We identified homologues of the relay system (Mia40, Erv1, and Osm1), a cytoplasmic class II fumarate hydratase (FH), and two putative Mia40 substrates: the IMS protease Atp23 and Tim10 (Figure 2A). In yeast, Atp23 processes the Atp6 subunit of ATP synthase[18] and controls the protein turnover of other IMS proteins linked to lipid metabolism such as Ups1[19]. Using the *Anaeramoeba* proteins as queries, we were able to detect Atp23-like proteins in the parabasalids *Trichomonas vaginalis* and *Tritrichomonas foetus* (Data S1). Given that parabasalids and anaeramoebids do not encode any components of ATP synthase or Ups1, Atp23 likely has another substrate. Future phylogenomic profiling initiatives or cell biological investigations should consider querying the reduced MRO proteomes of parabasalids and anaeramoebids to identify other Atp23 substrates that are conserved across eukaryotes.

The targeting of proteins to the mitochondrial matrix is often dependent on the recognition of a positively charged N-terminal targeting signal (NTS). Many *Anaeramoeba* MRO matrix proteins with complete N-termini (*A. ignava*: 35/72 [48.6%] and *A. flamelloides* 51/86 [59.3%]) have NTS with a net positive charge (Data S1). This is in contrast to observations in *Trichomonas*, where only 10% (39/359) experimentally localised hydrogenosomal proteins have detectable NTSs and they do not have a net positive charge[20]. Similar observations have been made in *Giardia*[13]. In *Trichomonas*, this suggests that the majority of hydrogenosomal proteins are imported via cryptic targeting signals[20]. This loss of NTS-dependent import in *Trichomonas* hydrogenosomes could derive from loss of the respiratory chain-driven membrane potential that is critical for the import of positively charged NTSs[20]. Our observation of positively-charged NTS-dependent protein import in *Anaeramoeba* MROs – organelles that are not predicted to have a membrane potential – has an number of implications: (i) the positively-charged NTS-dependent import of proteins can exist in MROs that do not have a respiratory chain, such as *Anaeramoeba* species; (ii) *Anaeramoeba* MROs might have a membrane potential that is generated independent of a respiratory chain[21]; and/or (iii) the loss of positively-charged NTS-dependent import likely occurred independently along the Parabasalia and Fornicata lineages. Future experimental investigations of the proteome of *Anaeramoeba* MROs are needed to confirm that positively-charged NTS are a major feature of MRO proteins.

Anaeramoebids encode hydrogenosomal and cytoplasmic hydrogen metabolism

The hydrogenosomes of *Trichomonas* oxidize pyruvate via pyruvate:ferredoxin oxidoreductase (PFO) to generate acetyl-CoA and reduced ferredoxin (Fd^-). Electrons from Fd^- and NADH are used by a trimeric, electron confucating [FeFe]-hydrogenase (HYDA, NUOE and NUOF). The proper assembly of the HYDA subunit requires three MRO-localized maturase proteins (HYDE-G). We identified homologues of PFO, HYDA, NUOE, NUOF and the maturase proteins in both *A. ignava* and *A. flamelloides* which are predicted to function in the MRO (Data S1; Figure 2B). We identified two paralogues of HYDA corresponding to the Group A and Group B hydrogenases[22] (Figure 2C, Figure S2, STAR methods). We identified numerous copies of Group A hydrogenases in both anaeramoebid species, predicted to function in the MRO or cytoplasm, however, the Group B hydrogenases are predicted to be exclusively cytoplasmic.

Some of the putatively cytoplasmic anaeramoebid Group A HYDA proteins have N-terminal extensions that encode WD40 domains that have never been reported on HYDA proteins until this study (Figure 3, Figure S2). WD40 domains are abundant in most eukaryotic proteomes and are often considered interaction hubs for protein-protein interaction networks[23]. Since WD40 domain-containing proteins often interact with other WD40 domain-containing proteins[23], we manually investigated every predicted protein that contains a WD40 domain (IPR020472) to identify potential interaction partners. We failed to detect any WD40-domain containing proteins that contain an additional domain related to electron transfer reactions (e.g., butyryl-CoA dehydrogenase, carbon monoxide dehydrogenase).

Without biochemical characterization, we cannot be confident about the direction of the reactions catalyzed by the putative HYDA enzymes. We suspect that 'canonical' MRO-localized Group A HYDA functions as part of a trimeric, electron-confurcating, H_2-evolving hydrogenase as presumed in *Trichomonas*[24,25]. The function of the remaining Group A HYDA proteins or the Group B HYDA proteins in eukaryotes has not been explored *in vitro*. It has been proposed that the cytoplasmic Group A HYDA proteins of *Trichomonas* might function as H_2:NADP+ oxidoreductase consuming H_2 under certain conditions[26]. Therefore, the anaeramoebid cytoplasmic Group A or Group B HYDA proteins may oxidize H_2 produced by the MRO a feature of which might be ancestral to the Parabasalia+Anaeramoebae group, or even the entire Metamonada lineage.

NADH and Fd coupled electron confurcation that is catalyzed by the multimeric hydrogenases (e.g., NUOE-NUOF-HYDA) is only thermodynamically favourable at low partial pressures of hydrogen[27]. Therefore, H_2-evolving bacteria and eukaryotes often grow syntrophically with other organisms who consume[28–30] H_2. Intracellular H_2-cycling has even been documented within a single bacterial cell[31]. In syntrophic eukaryotic systems the H_2-evolving organelle is often spatially associated with the H_2-consumer[28–30]. We observed a similar spatial organization between the MROs and proposed prokaryotic symbionts of both *Anaeramoeba* species in transmission electron micrographs[5]. This suggests that the symbionts and MROs might be engaging in metabolic exchange (e.g., H_2) and will be pursuing the characterization of this relationship in the future.

Anaeramoebid MROs have more ATP-producing pathways than Trichomonas vaginalis *hydrogenosomes*

In *Trichomonas vaginalis*, the acetyl-CoA generated from the PFO reaction is converted to acetate via acetate:succinate CoA transferase (ASCT subtype 1C) to generate succinyl-CoA which can serve as a substrate for the Krebs cycle enzyme succinyl-CoA synthetase (SCS), ultimately synthesizing ATP[32]. We detected proteins related to hydrogenosomal-type ATP production in both *Anaeramoeba* species (Figure 3). Unlike all other metamonads studied to date, the anaeramoebids also encode the methylmalonyl-CoA pathway for the synthesis of ATP. In aerobic mitochondria and the mitochondrial compartment of some anaerobes[33], succinyl-CoA is converted to propionyl-CoA with the concomitant production of ATP by the actions of methylmalonyl-CoA epimerase (MME), methylmalonyl-CoA mutase (MMM), and propionyl-CoA carboxylase (PCC). The CoA moiety of propionyl-CoA is transferred to succinate by ASCT 1B to generate propionate and succinyl-CoA,

respectively. We also detected homologues of 3-hydroxybutyrate (3HB) dehydrogenase for the conversion of 3HB to methylmalonylate semialdehyde synthase (MMSSyn). This represents the first evidence for propionate synthesis in Metamonada. We hypothesize that the propionate and acetate produced by the MRO could be key metabolites supporting the syntrophic interaction between *Anaeramoeba* and its symbionts as seen in other eukaryote:prokaryote syntrophies[30].

Anaeramoebid MROs have versatile amino acid interconversion pathways

The mitochondria of model eukaryotes are responsible for the biosynthesis and interconversion of amino acids[34]. A common feature retained in MROs in free-living protists[3,35], is the mitochondrial glycine cleavage system (GCS; Figure 2E) necessary for the generation of methylene tetrahydrofolate (CH_2-THF) from glycine for serine biosynthesis. We identified a putatively MRO-localized GCS and serine hydroxymethyltransferase (SHMT) in both anaeramoebid species. We also detected proteins involved in proline and glutamate interconversion and threonine synthesis (Figure 2E) that are only sparsely distributed across other metamonads (Figure 3), suggesting that anaeramoebid MROs retain the most ancestral repertoire of amino acid biosynthesis capabilities among metamonads.

Other pathways

Oxygen and reactive oxygen species can oxidize DNA, lipids, and proteins, therefore, most eukaryotes have strategies to prevent and repair oxidative damage in all compartments of the cell, including mitochondria and MROs. In the hydrogenosomes of *Trichomonas vaginalis*, oxygen is metabolized directly to water via flavodiiron protein (FDP), NADH oxidase (NO), or by the concerted action of nitrate reductase (NR), superoxide dismutase (SOD), and rubrerythrin (RBR) (Figure 2F).

We identified homologues of NO, NR, SOD, and RBR in the *Anaeramoeba* species (Figure 2F) and two FDPs predicted to function in the cytoplasm (FDP-Rb) and MRO (Fd-FDP) with distinct evolutionary histories and domain organization (see Figure S3A). We noted that the MRO-targeted FDP (Fd-FDP) had an N-terminal 2Fe-2S ferredoxin domain (IPR036010) in addition to the core FDP domains (IPR036866 and IPR029039) (Figure S3A). The *T. vaginalis* homologue of this enzyme (XP_001583562.1), that lacks a Fd domain, uses electrons derived from pyruvate or NADH via the Fd electron carrier for oxygen detoxification[36]. Therefore, it is possible that the Anaeramoebid Fd-FDP fusion proteins interact directly with NADH- (NUOE, NUOF) or pyruvate- (PFO) producing enzymes to reduce oxygen where the Fd acts *in cis* rather than *in trans*.

To detoxify hydrogen peroxide, *Trichomonas vaginalis* hydrogenosomes can use the peroxidase OsmC/Ohr together with the H and L-protein components of the GCS[37] (Figure 2F). We identified OsmC/Ohr-like proteins in *A. ignava* one of which was predicted to function in the MRO (Data S1). We also identified OsmC/Ohr homologues in the transcriptomes of other metamonads, however these sequences too fragmented to predict their subcellular localization. In phylogenetic analyses, the metamonad OsmC/Ohr proteins were not monophyletic suggesting these proteins might have distinct evolutionary histories (Figure S3B). Whether the *A. ignava* OsmC/Ohr-like protein

functions with the GCS proteins cannot be determined at present. If the *Anaeramoeba* and *Trichomonas* proteins do have distinct evolutionary origins, it suggests that this oxygen stress response pathway has convergently evolved in two related lineages.

Nearly all mitochondria and related organelles studied to date function in the biosynthesis of Fe-S clusters. In most organelles, these cofactors are synthesized by the iron-sulphur cluster (ISC) system. We identified all the major components of the FeS system in both species of *Anaeramoeba* (Figure 2G, Data S1). We found numerous mitochondrial carrier family proteins (SLC25A) that are predicted to transport nucleotides, S-adenosylmethionine, folate, iron, and carnitine (Data S1).

DISCUSSION

Our phylogenomic analysis shows that the Anaeramoebae represents an anciently diverged principal lineage of the eukaryotic supergroup Metamonada. The Anaeramoebae form a sister group of the Parabasalia and, as such, represents the long-sought free-living relatives of this predominantly parasitic group. The inclusion of Anaeramoebae in phylogenomic analyses leads to a crucial rearrangement of the high-order relationships within Metamonada, thereby forming two fundamental groups within Metamonada: Anaeramoebae+Parabasalia clade and Fornicata+*Barthelona*+Preaxyostyla. The discovery and robust phylogenetic placement of the Anaeramoebae within the tree of metamonads has important implications for evolutionary and comparative genomic investigations of the evolutionary adaptation to host and anaerobic environments that occurred in the parabasalid lineage. For example, genomes of Anaeramoebae could be surveyed for proteins homologous to those involved in host:pathogen interactions in parabasalids to investigate how and when molecular systems for pathogenesis originated in this lineage.

Our analyses predict the Anaeramoebae possess hydrogen-producing MROs that have the most ‘mitochondrion-like’ proteome of all metamonads and utilize positively-charged NTSs for protein import. These MROs represent important ‘missing links’ between typical aerobic mitochondria of other eukaryote supergroups and the anaerobic MROs of the Metamonada. Collectively, these data allow us to speculate that the last common ancestor of Metamonada was a phagocytosing heterotroph that resided in low-oxygen environments, engaged in syntrophic interactions with prokaryotes (through e.g., hydrogen exchange), and possessed MROs with more canonical mitochondrial features than previously anticipated.

ACKNOWLEDGEMENTS

IC, PT, and TP were supported by the Czech Science Foundation (project 18-18699S). CWS is supported by the Swedish Research Council (Vetenskapsrådet starting grant 2020-05071). Work carried out in the Roger lab was supported by Foundation grant FRN-142349, awarded to A.J.R. by the Canadian Institutes of Health Research. LE is supported by funding from the European Research Council (ERC StG 803151). EDS was supported by the MSCA-IF SMART

(CZ.02.2.69/0.0/0.0/20_079/0017809). MK was supported by Czech Science Foundation (project 18-28103S). MK was supported by Czech Science Foundation (project 18-28103S). This project was also supported by ERD fund “Centre for Research of Pathogenicity and Virulence of Parasites” (no. CZ.02.1.01/0.0/0.0/16_019/0000759).The authors would like to thank Diogo Meireles for supplying OsmC/Ohr protein family alignments[38]. The authors would like to thank the IT4Innovations National Super Computer Center, Technical University of Ostrava, Ostrava, Czech Republic for access to the HPC resources (projects Open-18-30 and Open-20-18). This article is dedicated to the memory of Prof. Thomas Cavalier-Smith, a scientist and educator who inspired protistologists worldwide.

AUTHOR CONTRIBUTIONS

Conceptualization: CWS, TP, PT, MK, AJR, IC

Methodology CWS, ES, LE, MK, TP, PT

Software: CWS, ES, LE, MK, TP, PT

Formal Analysis CWS, ES, LE, MK, TP, PT, AJR, IC

Investigation CWS, ES, LE, MK, PT, JJH, MH, CV

Resources TP, JJH, AJR, IC

Data Curation CWS, TP, ES, LE, MK, PT, AJR, IC

Writing – Original Draft CWS, LE, MK, TP, AJR, IC

Writing -Review & Editing CWS, ES, LE, MK, TP, AJR, IC

Visualization CWS, ES

Supervision CWS, MK, AJR, IC

Project Administration CWS, MK, AJR, IC

Funding Acquisition AJR, IC

DECLARATION OF INTERESTS

The authors do not have any competing interests to declare.

FIGURE LEGENDS

Figure 1: Anaeramoebae are the closest free-living relatives of parabasalids. Differential interference contrast micrographs of living amoebae of *Anaeramoeba flamelloides* (A) and A. *ignava* (B). Insets - Transmission electron micrographs of the cells of Anaeramoeba spp. showing MROs (M) and prokaryotic symbionts (P); scale bar = 10 μm; inset scale bar = 0.5 μm. C, D. Phylogenomic analysis inferred from 95 taxa, 155 proteins and 37169 sites. Bipartition support values derived from 400 non-parametric bootstraps (LG+C60+F+G model) are mapped onto the maximum likelihood tree

estimated with IQTREE under the LG+C60+F+G model of evolution. Support values greater than or equal to 99 or greater than 80 are shown in closed and open circles, respectively. All bipartitions were recovered with posterior probability=1.0 in Bayesian analyses with CAT+GTR, except the monophyly of Metamonada indicated with *. Tree files for all analyses are available[6], see Figure S1 for uncollapsed tree)

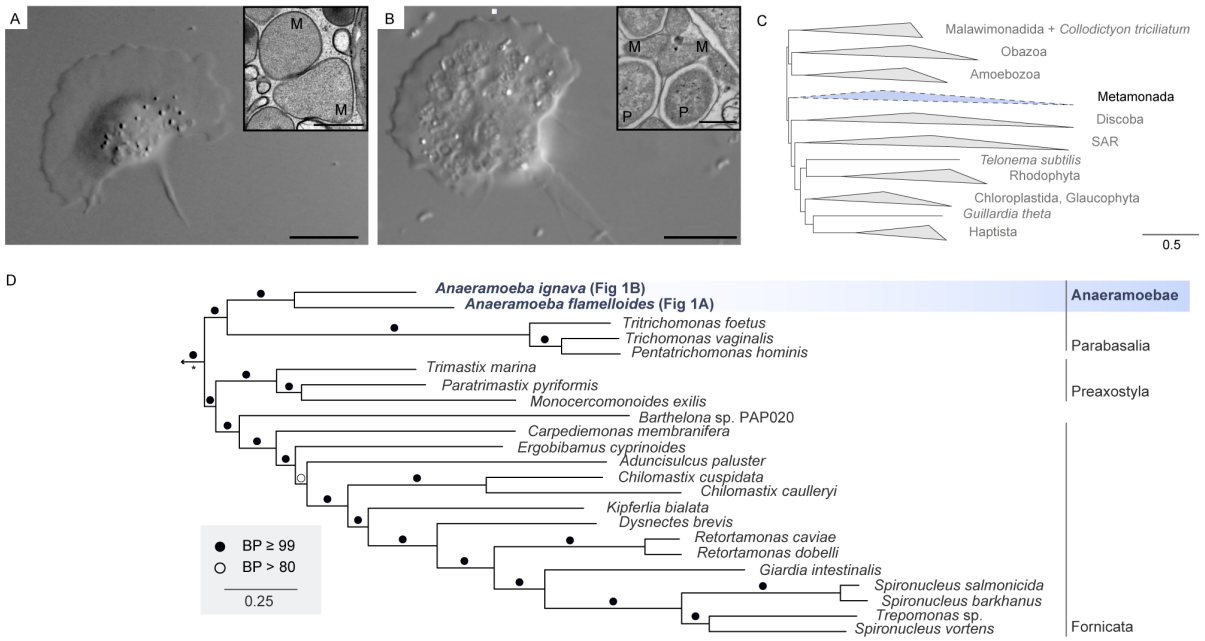


Figure 2 -- Selected metabolic pathways related to protein import and folding (A), pyruvate and hydrogen metabolism (B, C), energy metabolism (D), amino acid metabolism (E), oxidative stress response (F), and Fe-S cluster biosynthesis (G). Proteins that have previously not been identified in metamonads are shown in blue and orange representing putative MRO or cytoplasmic localization, respectively. Protein domain architectures of hydrogenase proteins are shown: 'WD40', WD-40 domain; 'G', NADH dehydrogenase subunit G (IPR019574); 'Lg', large hydrogenase subunit (IPR004108); 'Sm', small hydrogenase subunit (IPR004108); 'Fld', flavodoxin domain (IPR001094); 'FAD', FAD-binding domain (IPR017927). Detailed list of chaperones (crane) and proteases (scissors) can be found in Data S1. Numbered abbreviations: 1, pyruvate:ferredoxin oxidoreductase; 2, NADH:ubiquinone oxidoreductase E and F; 3 [FeFe]-hydrogenase: Group A (3a), WD40-GroupA (3w) and Group B (3b); 4, [FeFe]-hydrogenase E-G; 5, Acetate:succinate CoA transferase (ASCT) type 1C; 6 ASCT 1B; 7, succinyl-CoA synthetase; 8, Methylmalonyl (MM)-CoA epimerase; 9. MM-CoA mutase; 10, Propanoyl-CoA (Prop-CoA) carboxylase; 11, 3-hydroxybutyrate (3HB) dehydrogenase; 12, MM-semialdehyde (MMSA) dehydrogenase; 13, glycine (gly) cleavage system H,P,L,T proteins; 14, serine (ser) hydroxymethyltransferase; 15, Threonine (Thr) biosynthesis: Thr dehydrogenase (a) and Thr synthase (b); 16, D-pyrroline-5-carboxylate (DPC) dehydrogenase (a) Pyrroline-5-carboxylate reductase (b); NADHox, NADH oxidase; Fd-FDP, Ferredoxin-flavodiiron protein; NR, nitrate reductase; SOD, superoxide dismutase; RBR, rubrerythrin; OsmC, peroxidase; Yfh1, Frataxin;

IscS, Isd11, Cysteine desulphurase; IscU, IscA Fe-S scaffold proteins; Arh1, ferredoxin reductase; Ace, acetate; Suc, succinate; Prop, propionate; Pro, proline; Glu, glutamate; Fd, ferredoxin; SAM, sorting and assembly machinery, IMS, intermembrane space.

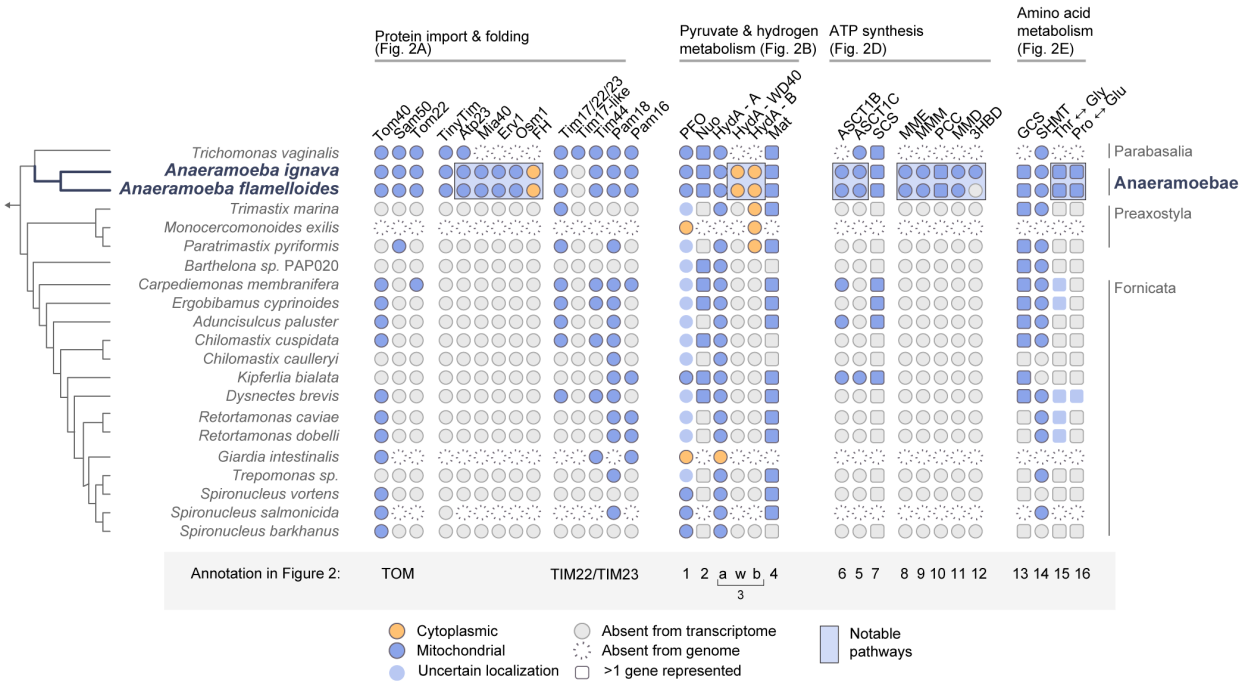
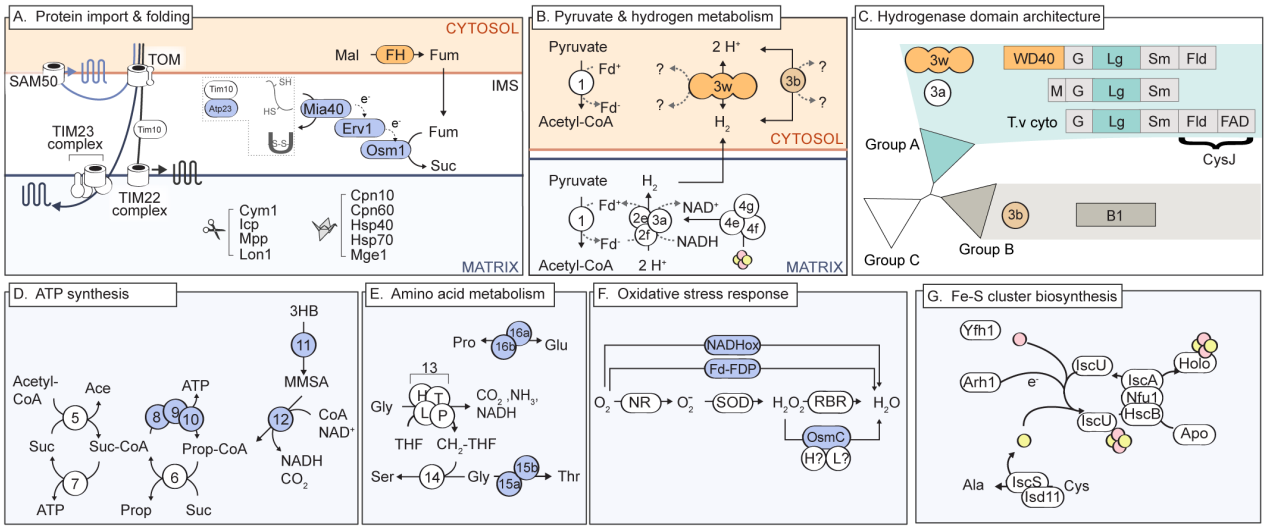


Figure 3 -- Distribution of various MRO functions across Metamonada. Publicly available genome and transcriptome projects were surveyed for genes encoding various MRO proteins. Those proteins predicted to function in the cytoplasm or MRO are shown in orange and blue, respectively. For Hydrogen metabolism, some organisms have multiple HYDA genes with protein products predicted to localize in the cytoplasm or MRO, therefore blue circles indicate that at least one protein is predicted to function in the MRO. 'WD40' denote whether a HYDA protein with a N-terminal WD40 domains was identified. If a gene encoding the protein was not found in a transcriptome or genome it is denoted with a hashed or grey icon, respectively. For visualization purposes, pathways or enzymes with more than one protein are shown with squares. Abbreviations and numbers are as shown in Figure 2. See Data S1 for sequence data for all metamonad proteins indicated. We failed to detect the *Trichomonas*-specific TIM17-like subunit[39], suggesting that this protein likely evolved specifically

along the parabasalid lineage.



DATA S1 LEGEND

A: Putative mitochondrial proteins identified in *Anaeramoeba* species and their associated annotations. *A. ignava* and *A. flamelloides* proteins are shown in blue and green, respectively. For each protein, a variety of annotations were applied orthologous group (bestOG), TargetP, MitoFates and Deeploc mitochondrial localization statistics, and assigned interproscan (IPR) domains. Predicted protein sequence and transcript sequences are provided. When applicable, references to figures 2 and 3 are provided. Submitochondrial localizations are as indicated: Outer Mitochondrial Membrane (OMM), Intermembrane space (IMS), Innermitochondrial membrane (IMM) and Mitochondrial Matrix (MM)

B: N-terminal targeting signal summary *of A. ignava* and *A. flamelloides* proteins with complete N-termini. *A. ignava* and *A. flamelloides* proteins are shown in blue and green, respectively. Proteins predicted to have complete N-termini and predicted NTS based on TargetP2 (orange) or mitofates (yellow) are shown with the predicted net charge at pH 5.5, 7.4 and 8.0

C: BUSCO summary statistics of transcriptome and genome data from representative metamonads. Please see BUSCO documentation for more information.

D: Distribution of proteins related to protein import into the MRO in Metamonada: For *A. ignava* and *A. flamelloides*, only 1 representative sequence is provided. See Sheet A_Mitochondrial_proteins for full list of sequences. For all other sequences, accession numbers of sequences were retrieved from Leger et al.[3]. When more than one sequence was detected, only the longest sequences is listed. See Leger et al. [3] for a complete list. '-' or '?' indicates the gene could not be identified in a genome or transcriptome respectively.

E: Phylogenetic distribution of metabolic proteins across Metamonada: For *A. ignava* and *A. flamelloides*, only 1 representative sequence is provided. See Sheet A_Mitochondrial_proteins for full list of sequences. For all other sequences, accession numbers of sequences were retrieved from Leger et al. When more than one sequence was detected, only the longest sequences is listed. See

Leger et al. for a complete list. '-' or '?' indicates the gene could not be identified in a genome or transcriptome respectively.

F: Mitochondrial carrier family proteins identified in Anaeramoeba species: Due to difficulty determining one-to-one orthologues between the two species, putative mitochondrial carrier family proteins from *A. ignava* and *A. flamelloides* are displayed separately**.**

G: Alignment statistics and model parameters for phylogenetic analyses.

H: Topology test results for each protein. The hypothesis was considered rejected for pvalues less than 0.05.

I: Transcriptome assembly statistics. Total number of reads before and have decontamination and number of coding sequences and protein sequences predicted by Transdecoder for *A. ignava* (blue) and A. f*lamelloides* (green)

J: Prokaryotic contaminants. Sequencing reads from A. ignava (blue) and A. flamelloides (green) were mapped to the following suspected contaminating prokaryotic genomes.

STAR METHODS

KEY RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Bacterial and virus strains		
Biological samples		
Chemicals, peptides, and recombinant proteins		
Artificial Seawater		
Cerophyll		
TRIzol reagent	ThermoFisher Scientific (Invitrogen/Life Technologies)	15596-026

Dynabeads Oligo(dT)25 kit	ThermoFisher Scientific (Invitrogen/Life Technologies)	61002
Ribo-Zero Gold Kit (Epicentre, Illumina). – WHICH KIT? HUMAN?		
Critical commercial assays		
"NEXTflex RNA-Seq Kit" Which one: https://perkinelmer-appliedgenomics.com/home/products/library-preparation-kits/illumina-rna-library-prep-kits/	PerkinElmer (BIOO Scientific)	
Deposited data		
Anaeramoeba ignava (transcriptome)	NCBI Bioproject PRJNA756164 Biosample SAMN20857051	Assembly available at figshare [6] figshare.com/articles/dataset/Anaeramoebae_are_deeply-branching_eukaryotes_that_clarify_the_transition_from_anaerobic_mitochondria_to_hydrogenosomes/12205517
Anaeramoeba flamelloides (transcriptome)	NCBI Bioproject PRJNA756164 Biosample SAMN20857050	Assembly available at figshare[6] figshare.com/articles/dataset/Anaeramoebae_are_deeply-branching_eukaryotes_that_clarify_the_transition_from_anaerobic_mitochondria_to_hydrogenosomes/12205517
Phylogenetic trees and alignments	FigShare	https://figshare.com/articles/dataset/Anaeramoebae_are_deeply-branching_eukaryotes_that_clarify_the_transition_from_anaerobic_mitochondria_to_hydrogenosomes/12205517
Experimental models: Cell lines		
Experimental models: Organisms/strains		
Anaeramoeba ignava strain BMAN	Táborskýa et al.[5]	
Anaeramoeba flamelloides strain BUSSELTON2	Táborskýa et al.[5]	

Oligonucleotides		
Recombinant DNA		
Software and algorithms		
trimmomatic v0.32	Bolger et al.[40]	https://github.com/timflutre/trimmomatic
trinityrnaseq_r20140717	Haas et al.[41]	https://github.com/trinityrnaseq/trinityrnaseq/
DeconSeq v0.1	Schmieder and Edwards[42]	http://deconseq.sourceforge.net/)
Transdecoder v5.5	Github	https://github.com/TransDecoder/TransDecoder/ https://transdecoder.g
emapper-1.0.3, (database 4.5.1)	Huerta-Cepas et al. [43]	https://github.com/eggnogdb/eggnog-mapper
BLAST 2.8.0+	Altschul et al., 1990[44]	https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=Download
ETE3	Huerta-Cepas et al.,[45]	http://etetoolkit.org/documentation/ete-view/
InterProScan v 5.22-61.0	Jones et al.[46]	https://www.ebi.ac.uk/interpro/search/sequence-search
TargetP 1.0	Emanuelsson et al.,[47]	http://www.cbs.dtu.dk/services/TargetP-1.1/index.php
TargetP 2.0	Almagro Armenteros et al.[48]	http://www.cbs.dtu.dk/services/TargetP/
MitoFates 1.2	Fukasawa et al.,[49]	http://mitf.cbrc.jp/MitoFates/cgi-bin/top.cgi
DeepLoc 1.0	Almagro Armenteros et al.,[50]	http://www.cbs.dtu.dk/services/DeepLoc/
Isoelectric Point Calculator v 2.0	Kozlowski [51]	http://isoelectric.org/
BUSCOv 4.0.4 (eukaryota_odb10)	Seppey et al.[52]	https://busco.ezlab.org/
mafft v5	Katoh[53,54]	https://mafft.cbrc.jp/alignment/software/

BMGE v1.12	Criscuolo and Gribaldo[55]	https://bioweb.pasteur.fr/packages/pack@BMGE@1.12
RAxML v8	Stamatakis[56]	https://cme.h-its.org/exelixis/web/software/raxml/
PREQUAL	Whelan et al. [57]	https://github.com/simonwhelan/prequal
DIVVIER	Ali et al. [58]	https://github.com/simonwhelan/Divvier
trimAL	Capella-Gutierrez et al.,[59]	http://trimal.cgenomics.org/
IQ-TREE v1.6, v2.0	Nguyen et al., [60]	http://www.iqtree.org/
PhyloBayes MPI v1.5	Lartillot et al., [61]	https://github.com/bayesiancook/pbmpi
HydDB v1	Søndergaard et al.,[62]	https://services.birc.au.dk/hyddb/
hmmer 3.2.1 (hmmscan)	Eddy[63]	hmmer.org
Other		

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, courtney.stairs@biol.lu.se

Materials availability

This study did not generate new unique reagents.

Data and code availability

RNA sequencing reads are deposited on the short-read archive (SRA) with Bioproject number PRJNA756164 for *Anaeramoeba ignava* (Biosample: SAMN20857051) and *Anaeramoeba flamelloides* (SAMN20857050). Assemblies have been deposited at FigShare and are publicly available as of the date of publication[6]. DOIs are listed in the key resources table. 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 Anaeramoeba ignava and Anaeramoeba flamelloides

Strain BUSSELTON2 of *Anaeramoeba flamelloides* and strain BMAN of *A. ignava* were maintained in monoeukaryotic culture with unidentified prokaryotes in ATCC medium 1525 at room temperature and were sub-cultured once a week[5].

METHOD DETAILS

RNA isolation and sequencing

RNA was isolated from 80 (BUSSELTON2) or 360 (BMAN) ml of high-density culture, using TRIzol Reagent (ThermoFisher Scientific) following the manufacturer´s protocol. Briefly, *Anaeramoeba flamelloides* cells were grown in 25-cm^2 culture flasks sealed with 40 mL of culture medium. For RNA extraction, liquid medium was removed by decanting from each flask and 5 mL of TRIzol the culture ('adherent cells' protocol). *Anaeramoeba ignava* cells were grown in 36 10 mL plastic tubes. Cells were collected by centrifugation (800 x g for 10 min) at room temperature. For RNA extraction, a total of 15 mL of TRIzol was used to resuspend and combine the pellets from the 36 tubes ('suspension cells' protocol). Messenger RNA was purified by multiple steps (3 steps of mRNA selection for BMAN, 2 steps for BUSSELTON2) using Dynabeads Oligo(dT)25 kit (Life Technologies) and in the case of BUSSELTON2 ribodepletion was performed by Ribo-Zero Gold Kit (Epicentre, Illumina). Illumina compatible RNAseq libraries were prepared using the NEXTflex RNA-Seq Kit (BIOO Scientific). 150 bp paired end reads were sequenced on the MiSeq platform and quality-filtered by the Genomics and Bioinformatics Core Facility, IMG Academy of Sciences of the Czech Republic (BUSSELTON2) and the Genomics Core Facility, EMBL, Heidelberg (BMAN).

Read decontamination, transcriptome assembly and annotation

Read statistics before and after decontamination are shown in Data S1. Adaptor trimming was performed using trimmomatic v0.32 using recommended settings (SLIDINGWINDOW:10:25 MINLEN:50)[40]. Initial assemblies were performed using Trinity (trinityrnaseq_r20140717[41]) using default parameters. The top 100 most abundant transcripts were extracted and used as queries against the *nt* database on GenBank to identify potential contaminants using a cut-off of 90% sequence identity. Genomes of those organisms most similar to the trinity transcripts were downloaded (Data S1) and used to decontaminate the raw reads using DeconSeq[42] (http://deconseq.sourceforge.net/). 'Cleaned' reads were reassembled using Trinity with default parameters. Transdecoder v5.5 was used to predict the open-reading frames and protein sequences for all transcripts (https://github.com/TransDecoder/TransDecoder/). Putative protein sequences were annotated using eggNOG-mapper (emapper-1.0.3, emapper DB: 4.5.1)[43]. Each protein sequence was queried against the *nr* database using BLAST 2.8.0+[44] and the taxonomic provenance of the top hit was assigned with ete[45]. The completeness measures assigned by TransDecder for each open

reading frame was manually investigated to determine whether the sequence was full-length, 5′-partial, 3′-partial or internal (5′- and 3′-partial) using BLAST again the nr database, see Data S1. We manually corrected some protein models predicted by TransDecoder that extended into the 5' untranslated region of the transcripts based on homology to other sequences in GenBank. Interpro (IPR) domains were annotated using interproscan[46] with the iprdomain setting. Subcellular localization was predicted using TargetP[47], TargetP2[48], MitoFates[49], and DeepLoc[50] and summarized in Data S1. Net charge of putative N-terminal targeting signals was calculated at pH 5.5, 7.4 and 8.0 using Isoelectric Point Calculator[51] (v. 2.0) For each prediction software, we arbitrarily assigned a confidence point of one for any prediction with the probability of mitochondrial localization greater than 0.5 and a point of 0.5 if the probability of mitochondrial localization was less than 0.5 but the software still predicted mitochondrial localization. Every protein with one confidence point were manually investigated for mitochondrial provenance. Putative MRO proteins from other free-living anaerobic protists (*Pygsuia biforma*[21], *Brevimastigamonas motovehiculus*[64]) were used as queries against the protein datasets of each transcriptome to identify similar homologues (-evalue 0.001). Transcriptome completeness was assessed using BUSCO[52] v 4.0.4 (eukaryota_odb10 (eukaryota, 2019-11-20)). We recovered 61% and 47% of near-universal single-copy orthologues for each species, respectively, which is comparable to the BUSCO coverage for complete genomes and deeply sequenced transcriptomes of other metamonads (Data S1).

Phylogenomic Dataset Assembly

To infer the phylogenetic placement of the two *Anaeramoeba* species, we updated a previously published phylogenomic dataset of 155 proteins[3] by adding orthologues from the Anaeramoebae, and three recently published Metamonada, *Retortamonas dobelli*[65], *Retortamonas* cf. *caviae*[65], and *Barthelona* sp.[8]. For each of the 155 proteins, we used the *Arabidopsis thaliana* or *Homo sapiens* homologue as a query against the predicted proteins of each *Anaeramoeba* species (e-value cut-off = 10e-10) using BLASTP [44] and retained up to five hits. Those potential homologues were then used as queries for a reciprocal BLAST search against a dataset containing the initial set of identified orthologues, but also their known deep paralogues (i.e., corresponding to duplications predating the last eukaryotic common ancestor, such as EF1-L and EF1-α). *Anaeramoeba* sequences were retained for subsequent analyses if their reciprocal best BLAST hit belonged to the orthologous clade of interest. These were added to the existing orthologue dataset, which was aligned using mafft-linsi[53,54]. Ambiguously aligned positions were trimmed using BMGE v1.12 (default settings, except from -m BLOSUM62 and -g 0.3;[55]). Preliminary maximum likelihood individual protein trees were computed using RAxML v8 (with the PROTGAMMALG model, branch support was estimated using 100 rapid bootstraps)[56]. For the three recently published metamonads, all previously identified orthologs were used as queries against the predicted proteins for each species (e-value cut-off = 10e-5) using BLASTP [44] and retained up to five hits per query. These were added to the existing dataset, filtered for sequencing errors and non-homologous sites using PREQUAL[57], aligned using mafft-ginsi[53,54], alignment uncertainty and errors were filtered using DIVVIER[58], and the filtered alignments

were trimmed of sites comprised of >99% gaps using trimAL (-gt 0.01)[59]. Preliminary maximum likelihood individual protein trees were estimated using IQ-TREE under the LG+C20+F+G model, and branch support was estimated using 100 RAxML rapid bootstraps inferred using the LG4X model. Each tree was then manually inspected and a single sequence corresponding to the correct orthologue was selected for each species. In case of in-paralogues specific to the newly added taxa, the sequence with the best coverage of the alignment or the shortest branching sequence was selected. Single protein trees were recomputed as previously described, and well-supported bipartitions (bootstrap support > 70) were further investigated to identify potential paralogues, contaminants, or lateral- or endosymbiotic gene transfers, which were then removed from the datasets. Final datasets were re-aligned and trimmed as described for the recently published Metamonada species and concatenated into a single large alignment (supermatrix containing 95 sequences and 53637 sites). Unaligned individual protein sequences have been deposited in a Figshare[6].

Phylogenomic tree reconstruction

Maximum likelihood (ML) phylogenies were inferred using IQ-TREE v1.6[60] under the LG+C60+F+G mixture model. Statistical confidence was derived from 1000 non-parametric bootstraps using the site-heterogeneous LG+C60+F+G with posterior mean site frequencies (PMSF) using the LG+C60+F+G ML tree as guide tree [66]. The supermatrix was also subjected to a Bayesian analysis with PhyloBayes MPI v1.5[61] under the CAT + GTR + G model (default settings). Four independent Markov chain Monte Carlo chains were run for ~6000 generations. Two chains converged to similar posterior distributions of topologies (maxdiff = 0.16; burn-in = 25%), Three of the four chains favoured different topologies where Metamonada was not monophyletic. Instead two separate clades comprised of parabasalids+Anaeramoeba and *Barthelona*+Fornicata+Preaxostyla branched sequentially in the eukaryote tree. The fourth chain recovered Metamonada as monophyletic with Parabasalia+*Anaeramoeba* branching sister to *Barthelona*+ fornicates+preaxostylids. In all chains, *Anaeramoeba* species grouped with parabasalids.

Investigation of select phylogenetic relationships

Since the addition of *Anaeramoeba* fundamentally altered the relationships of Metamonada in the ML tree, by recovering Preaxostyla+Fornicata rather than Parabasalia+Fornicata, we conducted AU tests to investigate support for this, and other relationships between excavates. We inferred ML trees using IQ-TREE under the LG+C60+F+G model, constraining the monophyly of Parabasalia+Fornicata+*Barthelona*, Parabasalia+Fornicata+*Barthelona*+*Anaeramoeba,* C*ollodictyon*+*Malawimonas*+Discoba+Metamonada, and Discoba+Metamonada, as well as the Bayesian consensus topology (that recovered paraphyletic metamonads) using the -g option. The AU test was implemented in IQ-TREE on our ML topology, the individual constrained trees, and 100

distinct local topologies saved during ML analysis of our concatenated dataset (-wt option in IQ-TREE). Newick formatted constraint topologies are provided (figshare.com/s/fa207b47f454b7129277).

Phylogenetic analysis of genes related to hydrogen production and oxidative stress

The initial dataset of *Anaeramoeba* hydrogenase proteins was built by retrieving any sequence containing the large hydrogenase domain (IPR004108) based on the InterProScan annotations. These proteins were subclassified as Group A or Group B using HydDB[62] and separated for the following analyses. For the Group A and Group B sequences, we first extracted the hydrogenase domain using hmmscan[63] (hmmer.org) and the Pfam hmm profile PF02906.12 (Fe_hyd_lg_C) or TIGR04105 from TIGR, using default settings, respectively. These sequences were used as queries against the non-redundant database (*nr*) and a local database of metamonad sequences[3] using BLAST[44] (e-value < 1e-10; -max_target_seqs 1500). Sequences from these BLAST results were combined and reduced using cd-hit[67] (-c 0.8). The retrieved Group A and Group B sequences were subclassified as above with HydDB[62] and only the regions corresponding to the large hydrogenase domain was extracted as above. For initial datasets, sequences were aligned using mafft-linsi[53] and trimmed using BMGE[55] (v1.12; -h 0.7, -m BLOSUM30) and the phylogeny was generated using FastTree2[68]. Trees were manually inspected to identify those clades that were distantly related to the sequences of interest that could be removed to reduce the size of the dataset. Final datasets were aligned using mafft-linsi and trimmed with BMGE as above. Phylogenies were inferred using IQTREE[69] v2.0 under the best scoring model of evolution decided by ModelFinder (supplemented with the C-series mixture models with -mset LG+C20,LG+C10,LG+C60,LG+C30,LG+C40,LG+C50) and 1000 ultrafast bootstraps (-bb 1000). The resulting maximum likelihood tree was used as a guide tree to compute 100 non-parametric bootstraps using the PMSF framework[66] implemented in IQTREE[69]. Model parameters and alignment features can be found in Data S1. Constrained estimation of topologies conditional on enforcing the monophyly of eukaryotes were performed in IQTREE as described above (Data S1).

Phylogenetic analyses demonstrated that the *Anaeramoeba* Group A sequences branched in a large clade composed of eukaryotic and prokaryotic sequences (i.e., *Thermotogales*) similar to previous analyses[70–72] (Figure S2). We conducted approximately unbiased (AU) tests to determine if alternative topologies could be rejected by these data. Topologies constraining the monophyly of eukaryotes, or Parabasalia+*Anaeramoeba* were not rejected (Data S1). However, the AU-test rejected the topology constraining the monophyly of Metamonada (p=0.00874, Data S1) suggesting that the metamonad hydrogenases might not share a most recent common ancestor or that other eukaryotic taxa received a metamonad-like hydrogenase after the split of Anaeramoeba+Parabasalia and Preaxostyla+*Barthelona*+Fornicata. Group B sequences were only found in Preaxostyla, Amoebozoa, and the anaeramoebids indicating they are less widespread among eukaryotes compared to the Group A paralogue. In phylogenetic analyses, the *Anaeramoeba* Group B sequences

formed a clade with *Entamoeba* and *Mastigamoeba* sequences; however, this relationship was poorly supported (BP=13; Figure S3).

For FDP and OsmC analyses we added the *Anaeramoeba* sequences to datasets presented in Stairs et al. (2019)[73] and Meireles et al. (2017)[38], respectively. Datasets were aligned with mafft and trimmed with BMGE, as above. Phylogenies were inferred using IQTREE[69] v2.0 under the best scoring model of evolution decided by ModelFinder (supplemented with the C-series mixture models with -mset LG+C20,LG+C10,LG+C60,LG+C30,LG+C40,LG+C50) and 1000 ultrafast bootstraps (-bb 1000). The resulting maximum likelihood tree was used as a guide tree to compute 100 non-parametric bootstraps using the PMSF framework[66] implemented in IQTREE[69]. Model parameters and alignment features can be found in Data S1. Constrained estimation of topologies conditional on enforcing the monophyly of eukaryotes were performed in IQTREE as described above (Data S1).

ADDITIONAL RESOURCES

Transcriptome assembly, gene prediction, protein models and all raw data for phylogenetic analyses are available at Figshare[6].

REFERENCES

1. Burki, F., Roger, A.J., Brown, M.W., and Simpson, A.G.B. (2020). The New Tree of Eukaryotes. Trends in Ecology & Evolution *35*, 43–55.

2. Stairs, C.W., Leger, M.M., and Roger, A.J. (2015). Diversity and origins of anaerobic metabolism in mitochondria and related organelles. Phil. Trans. R. Soc. B *370*, 20140326.

3. Leger, M.M., Kolisko, M., Kamikawa, R., Stairs, C.W., Kume, K., Čepička, I., Silberman, J.D., Andersson, J.O., Xu, F., Yabuki, A., et al. (2017). Organelles that illuminate the origins of Trichomonas hydrogenosomes and Giardia mitosomes. Nat Ecol Evol *1*, 0092.

4. Jerlström-Hultqvist, J., Einarsson, E., Xu, F., Hjort, K., Ek, B., Steinhauf, D., Hultenby, K., Bergquist, J., Andersson, J.O., and Svärd, S.G. (2013). Hydrogenosomes in the diplomonad Spironucleus salmonicida. Nature Communications *4*, 2493.

5. Táborský, P., Pánek, T., and Čepička, I. (2017). Anaeramoebidae fam. nov., a Novel Lineage of Anaerobic Amoebae and Amoeboflagellates of Uncertain Phylogenetic Position. Protist *168*, 495–526.

6. Stairs, Courtney (2021). Anaeramoebae are deeply-branching eukaryotes that clarify the transition from anaerobic mitochondria to hydrogenosomes. 136685713 Bytes. https://figshare.com/articles/dataset/Anaeramoebae_are_deeply-branching_eukaryotes_that_clarify_the_transition_from_anaerobic_mitochondria_to_hydrogenosomes/12205517.

7. 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.

8. Yazaki, E., Kume, K., Shiratori, T., Eglit, Y., Tanifuji, G., Harada, R., Simpson, A.G.B., Ishida, K.-I., Hashimoto, T., and Inagaki, Y. (2020). Barthelonids represent a deep-branching metamonad clade with mitochondrion-related organelles predicted to generate no ATP. Proc Biol Sci *287*, 20201538.

9. Takishita, K., Kolisko, M., Komatsuzaki, H., Yabuki, A., Inagaki, Y., Čepička, I., Smejkalová, P., Silberman, J.D., Hashimoto, T., Roger, A.J., et al. (2012). Multigene Phylogenies of Diverse Carpediemonas-like Organisms Identify the Closest Relatives of 'Amitochondriate' Diplomonads and Retortamonads. Protist *163*, 344–355.

10. Heiss, A.A., Kolisko, M., Ekelund, F., Brown, M.W., Roger, A.J., and Simpson, A.G.B. (2018). Combined morphological and phylogenomic re-examination of malawimonads, a critical taxon for inferring the evolutionary history of eukaryotes. R. Soc. open sci. *5*, 171707.

11. Hampl, V., Hug, L., Leigh, J.W., Dacks, J.B., Lang, B.F., Simpson, A.G.B., and Roger, A.J. (2009). Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups." Proceedings of the National Academy of Sciences *106*, 3859–3864.

12. Čepička, I., Dolan, M.F., and Gile, G.H. (2016). Parabasalia. In Handbook of the Protists, J. M. Archibald, A. G. B. Simpson, C. H. Slamovits, L. Margulis, M. Melkonian, D. J. Chapman, and J. O. Corliss, eds. (Springer International Publishing), pp. 1–44.

13. Doležal, P., Makki, A., and Dyall, S.D. (2019). Protein Import into Hydrogenosomes and Mitosomes. In Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes Microbiology Monographs., J. Tachezy, ed. (Springer International Publishing), pp. 31–84.

14. Backes, S., and Herrmann, J.M. (2017). Protein Translocation into the Intermembrane Space and Matrix of Mitochondria: Mechanisms and Driving Forces. Front. Mol. Biosci. *4*, 83.

15. Reinhardt, C., Arena, G., Nedara, K., Edwards, R., Brenner, C., Tokatlidis, K., and Modjtahedi, N. (2020). AIF meets the CHCHD4/Mia40-dependent mitochondrial import pathway. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease *1866*, 165746.

16. Neal, S.E., Dabir, D.V., Wijaya, J., Boon, C., and Koehler, C.M. (2017). Osm1 facilitates the transfer of electrons from Erv1 to fumarate in the redox-regulated import pathway in the mitochondrial intermembrane space. MBoC *28*, 2773–2785.

17. Williams, C.C., Jan, C.H., and Weissman, J.S. (2014). Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. Science *346*, 748–751.

18. Osman, C., Wilmes, C., Tatsuta, T., and Langer, T. (2007). Prohibitins Interact Genetically with Atp23, a Novel Processing Peptidase and Chaperone for the F_1F_O-ATP Synthase. MBoC *18*, 627–635.

19. Potting, C., Wilmes, C., Engmann, T., Osman, C., and Langer, T. (2010). Regulation of mitochondrial phospholipids by Ups1/PRELI-like proteins depends on proteolysis and Mdm35. The EMBO Journal *29*, 2888–2898.

20. 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.

21. Stairs, C.W., Eme, L., Brown, M.W., Mutsaers, C., Susko, E., Dellaire, G., Soanes, D.M., van der Giezen, M., and Roger, A.J. (2014). A SUF Fe-S Cluster Biogenesis System in the Mitochondrion-Related Organelles of the Anaerobic Protist Pygsuia. Current Biology *24*, 1176–1186.

22. Greening, C., Biswas, A., Carere, C.R., Jackson, C.J., Taylor, M.C., Stott, M.B., Cook, G.M., and Morales, S.E. (2016). Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. ISME J *10*, 761–777.

23. Stirnimann, C.U., Petsalaki, E., Russell, R.B., and Müller, C.W. (2010). WD40 proteins propel cellular networks. Trends in Biochemical Sciences *35*, 565–574.

24. Müller, M., Mentel, M., Hellemond, J.J. van, Henze, K., Woehle, C., Gould, S.B., Yu, R.-Y., Giezen, M. van der, Tielens, A.G.M., and Martin, W.F. (2012). Biochemistry and Evolution of Anaerobic Energy Metabolism in Eukaryotes. Microbiol. Mol. Biol. Rev. *76*, 444–495.

25. Hrdý, I., Hirt, R.P., Doležal, P., Bardonová, L., Foster, P.G., Tachezy, J., and Martin Embley, T. (2004). Trichomonas hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. Nature *432*, 618–622.

26. Sutak, R., Hrdý, I., Doležal, P., Cabala, R., Sedinová, M., Lewin, J., Harant, K., Müller, M., and Tachezy, J. (2012). Secondary alcohol dehydrogenase catalyzes the reduction of exogenous acetone to 2-propanol in Trichomonas vaginalis: Secondary alcohol dehydrogenase in T. vaginalis. FEBS Journal *279*, 2768–2780.

27. Stams, A.J.M., and Plugge, C.M. (2009). Electron transfer in syntrophic communities of anaerobic bacteria and archaea. Nat Rev Microbiol *7*, 568–577.

28. Beinart, R.A., Beaudoin, D.J., Bernhard, J.M., and Edgcomb, V.P. (2018). Insights into the metabolic functioning of a multipartner ciliate symbiosis from oxygen-depleted sediments. Mol Ecol *27*, 1794–1807.

29. Finlay, B.J., Esteban, G., Clarke, K.J., Williams, A.G., Embley, T.M., and Hirt, R.P. (1994). Some rumen ciliates have endosymbiotic methanogens. FEMS Microbiology Letters *117*, 157–161.

30. Hamann, E., Gruber-Vodicka, H., Kleiner, M., Tegetmeyer, H.E., Riedel, D., Littmann, S., Chen, J., Milucka, J., Viehweger, B., Becker, K.W., et al. (2016). Environmental Breviatea harbour mutualistic Arcobacter epibionts. Nature *534*, 254–258.

31. Wiechmann, A., Ciurus, S., Oswald, F., Seiler, V.N., and Müller, V. (2020). It does not always take two to tango: "Syntrophy" via hydrogen cycling in one bacterial cell. ISME J *14*, 1561–1570.

32. van Grinsven, K.W.A., Rosnowsky, S., van Weelden, S.W.H., Pütz, S., van der Giezen, M., Martin, W., van Hellemond, J.J., Tielens, A.G.M., and Henze, K. (2008). Acetate:Succinate CoA-transferase in the Hydrogenosomes of Trichomonas vaginalis. Journal of Biological Chemistry *283*, 1411–1418.

33. Pietrzak, S., and Saz, H. (1981). Succinate decarboxylation to propionate and the associated phosphorylation in Fasciola hepatica and Spirometra mansonoides. Molecular and Biochemical Parasitology *3*, 61–70.

34. Spinelli, J.B., and Haigis, M.C. (2018). The multifaceted contributions of mitochondria to cellular metabolism. Nat Cell Biol *20*, 745–754.

35. Zubáčová, Z., Novák, L., Bublíková, J., Vacek, V., Fousek, J., Rídl, J., Tachezy, J., Doležal, P., Vlček, Č., and Hampl, V. (2013). The Mitochondrion-Like Organelle of Trimastix pyriformis Contains the Complete Glycine Cleavage System. PLoS ONE *8*, e55417.

36. Smutná, T., Gonçalves, V.L., Saraiva, L.M., Tachezy, J., Teixeira, M., and Hrdý, I. (2009). Flavodiiron Protein from *Trichomonas vaginalis* Hydrogenosomes: the Terminal Oxygen Reductase. Eukaryot Cell *8*, 47–55.

37. Nývltová, E., Smutná, T., Tachezy, J., and Hrdý, I. (2016). OsmC and incomplete glycine decarboxylase complex mediate reductive detoxification of peroxides in hydrogenosomes of Trichomonas vaginalis. Molecular and Biochemical Parasitology *206*, 29–38.

38. Meireles, D.A., Domingos, R.M., Gaiarsa, J.W., Ragnoni, E.G., Bannitz-Fernandes, R., da Silva Neto, J.F., de Souza, R.F., and Netto, L.E.S. (2017). Functional and evolutionary characterization of Ohr proteins in eukaryotes reveals many active homologs among pathogenic fungi. Redox Biology *12*, 600–609.

39. Rada, P., Doležal, P., Jedelský, P.L., Bursac, D., Perry, A.J., Šedinová, M., Smíšková, K., Novotný, M., Beltrán, N.C., Hrdý, I., et al. (2011). The Core Components of Organelle Biogenesis and Membrane Transport in the Hydrogenosomes of Trichomonas vaginalis. PLoS ONE *6*, e24428.

40. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

41. 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.

42. Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One *6*, e17288.

43. Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol *34*, 2115–2122.

44. 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.

45. Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol Biol Evol *33*, 1635–1638.

46. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240.

47. Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. Journal of Molecular Biology *300*, 1005–1016.

48. 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.

49. Fukasawa, Y., Tsuji, J., Fu, S.-C., Tomii, K., Horton, P., and Imai, K. (2015). MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites. Mol Cell Proteomics *14*, 1113–1126.

50. Almagro Armenteros, J.J., Sønderby, C.K., Sønderby, S.K., Nielsen, H., and Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics *33*, 3387–3395.

51. Kozlowski, L.P. (2016). IPC – Isoelectric Point Calculator. Biol Direct *11*, 55.

52. Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods Mol Biol *1962*, 227–245.

53. Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research *30*, 3059–3066.

54. Katoh, K. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research *33*, 511–518.

55. Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol *10*, 210.

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

57. Whelan, S., Irisarri, I., and Burki, F. (2018). PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences. Bioinformatics *34*, 3929–3930.

58. Ali, R.H., Bogusz, M., and Whelan, S. (2019). Identifying Clusters of High Confidence Homologies in Multiple Sequence Alignments. Molecular Biology and Evolution *36*, 2340–2351.

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

60. 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. Molecular Biology and Evolution *32*, 268–274.

61. Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI: Phylogenetic Reconstruction with Infinite Mixtures of Profiles in a Parallel Environment. Systematic Biology *62*, 611–615.

62. Søndergaard, D., Pedersen, C.N.S., and Greening, C. (2016). HydDB: A web tool for hydrogenase classification and analysis. Sci Rep *6*, 34212.

63. Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS Comput Biol *7*, e1002195.

64. Gawryluk, R.M.R., Kamikawa, R., Stairs, C.W., Silberman, J.D., Brown, M.W., and Roger, A.J. (2016). The Earliest Stages of Mitochondrial Adaptation to Low Oxygen Revealed in a Novel Rhizarian. Current Biology *26*, 2729–2738.

65. Füssy, Z., Vinopalová, M., Treitli, S.C., Pánek, T., Smejkalová, P., Čepička, I., Doležal, P., and Hampl, V. (2021). Retortamonads from vertebrate hosts share features of anaerobic metabolism and pre-adaptations to parasitism with diplomonads. Parasitology International *82*, 102308.

66. 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. Systematic Biology *67*, 216–235.

67. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150–3152.

68. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE *5*, e9490.

69. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution *37*, 1530–1534.

70. Nývltová, E., Stairs, C.W., Hrdý, I., Rídl, J., Mach, J., Pačes, J., Roger, A.J., and Tachezy, J. (2015). Lateral Gene Transfer and Gene Duplication Played a Key Role in the Evolution of Mastigamoeba balamuthi Hydrogenosomes. Molecular Biology and Evolution *32*, 1039–1055.

71. Leger, M.M., Eme, L., Stairs, C.W., and Roger, A.J. (2018). Demystifying Eukaryote Lateral Gene Transfer (Response to Martin 2017 DOI: 10.1002/bies.201700115). BioEssays *40*, 1700242.

72. Stairs, C.W., Dharamshi, J.E., Tamarit, D., Eme, L., Jørgensen, S.L., Spang, A., and Ettema, T.J.G. (2020). Chlamydial contribution to anaerobic metabolism during eukaryotic evolution. Sci. Adv. *6*, eabb7258.

73. Stairs, C.W., Kokla, A., Ástvaldsson, Á., Jerlström-Hultqvist, J., Svärd, S., and Ettema, T.J.G. (2019). Oxygen induces the expression of invasion and stress response genes in the anaerobic salmon parasite Spironucleus salmonicida. BMC Biol *17*, 19.

74. Lamien-Meda, A., and Leitsch, D. (2020). Identification of the NADH-oxidase gene in Trichomonas vaginalis. Parasitol Res *119*, 683–686.

788