Charles University, Faculty of Science

Univerzita Karlova, Přírodovědecká fakulta

PhD study program: Parasitology

Doktorský studijní program: Parazitologie

Mgr. Lukáš V. F. Novák

Genomics of Preaxostyla flagellates

Genomika bičíkovců skupiny Preaxostyla

Ph.D. Thesis

Thesis supervisor: doc. Mgr. Vladimír Hampl, Ph.D.

Prague, 2019

Declaration of the author

I declare that I elaborated this thesis independently. I also proclaim that the literary sources were cited properly and neither this work nor the substantial part of it have been used to reach the same or any other academic degree.

Prohlašuji, že jsem tuto práci zpracoval samostatně a že jsem uvedl všechny použité zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

Mgr. Lukáš V. F. Novák

Declaration of the thesis supervisor

Data presented in this thesis resulted from team collaboration at the Evolutionary protistology group and from cooperation with other associates.

I declare that the involvement of Mgr. Lukáš V. F. Novák in this work was substantial and that he contributed significantly to obtain the results.

Doc. Mgr. Vladimír Hampl, Ph.D.

Thesis supervisor

Acknowledgements

I would like to express my gratitude to my supervisor Vladimír Hampl for introducing me to the world of protistology and guiding me through this work, to my mentors Zuzana Vaitová and Anna Karnkowska for teaching me so much about laboratory work and bioinformatics, and to Andrew Roger for kindly hosting me in his laboratory in Canada.

Many thanks also to my high school biology teacher Petr Šíma for kindling my interest in the microscopic world, to my astronomy instructor Jindřiška Majorová whom I owe so much for my science soft skills, to my parents Kateřina Nováková and Vítězslav Novák for raising me in an atmosphere of curiosity and love for nature, to my partner and spouse Anna Novák Vanclová for joining me on this journey and supporting me unwaveringly, and to all my colleagues and friends for their help and care.

Table of contents

Abstract

Protists inhabiting oxygen-depleted environments have evolved various adaptation to thrive in their niches, including modified mitochondria to various degrees adapted to anaerobiosis. The most radically altered forms of these organelles (Mitochondria-Related Organelles, MROs) have completely lost their genomes and other defining features of canonical aerobic mitochondria. Anaerobic protists are often found as endobionts (parasites, mutualists, etc.) of larger organisms. The endobiotic lifestyle combined with anaerobiosis poses another source of evolutionary pressure forcing unique adaptations in the endobionts. Here we present new insights into the adaptations of an anaerobic protistan phylum Preaxostyla, especially with regard to the reductive evolution of mitochondria, which, uniquely among all known eukaryotes, led to a complete loss of the organelle in the oxymonad *Monocercomonoides exilis*.

We have obtained *M. exilis* genomic assembly of good quality and completeness, as well as genomic and transcriptomic data of varying quality and completeness from 9 other Preaxostyla species. Based on extensive, thorough gene searches and functional gene annotation on these datasets, as well as phylogenetic analyses and protein localization experiments, we conclude: 1) *M. exilis* has completely lost the mitochondrion. This was likely facilitated by a replacement of the mitochondrial system for iron-sulfur (Fe-S) cluster assembly (ISC) with an unrelated SUF system of bacterial origin, which was employed for function in the cytosol; 2) Despite the loss of mitochondria, *M. exilis* displays no major reduction in genomic or cellular complexity compared to other anaerobic protists endowed with MROs; 3) The SUF system for Fe-S clusters assembly is present in all studied Preaxostyla and was likely gained in a single lateral gene transfer event from bacteria into a common ancestor of extant Preaxostyla. No studied member of Preaxostyla has the mitochondrial ISC system; 4) The ATP-producing arginine deiminase (ADI) pathway is present in most studied Metamonada including Preaxostyla and likely represents an ancestral feature of Metamonada. Distribution and phylogeny of the 3 ADI pathway genes among eukaryotes is consistent with presence of the pathway already in the last eukaryotic common ancestor (LECA) and their evolutionary history was shaped by frequent losses and lateral gene transfers.

Abstrakt

Protista obývající prostředí chudá na kyslík si vyvinula řadu specifických adaptací, včetně modifikovaných mitochondrií, do různé míry přizpůsobených životu bez kyslíku. Nejradikálněji pozměněné typy těchto organel nazývaných MRO (Mitochondria-Related Organelles) zcela ztratily mitochondriální genom i další znaky definující kanonické aerobní mitochondrie. Anaerobní protista jsou často endobionty (tzn. parazity, mutualisty atp.) větších organizmů. Vedle anaerobního prostředí představuje endobiotický způsob života další zdroj selekčního tlaku, vyžadující unikátní adaptace včetně např. redukovaných biosyntetických schopností, nebo modifikovaného souboru povrchových proteinů. V této práci představujeme nové poznatky o adaptacích anaerobních protists kmene Preaxostyla, se zvláštním ohledem na reduktivní evoluci mitochondrií, která vedla k unikátní úplné ztrátě mitochondrie u zástupce této skupiny, druhu *Monocercomonoides exilis*.

Získali jsme genomovou sekvenci *M. exilis* dobré kvality a úplnosti i genomová a transkriptomová data z 9 dalších devíti zástupců Preaxostyla. Na základě těchto dat, pečlivých funkčních anotací genů, fylogenetických analýz a experimentální lokalizace proteinů vyvozujeme tyto závěry: 1) *M. exilis* zcela ztratil mitochondrii. Tato ztráta byla pravděpodobně umožněna nahrazením mitochondriálního systému (ISC) pro syntézu železosirných center nepříbuzným systémem SUF, který byl laterálně přenesen z bakterií do předka *M. exilis* a byl zapojen do metabolismu v cytosolu. 2) I přes ztrátu mitochondrie *M. exilis* nejeví žádné známky redukce genomové ani buněčné komplexity ve srovnání s jinými protisty s redukovanými mitochondriemi. 3) SUF systém pro syntézu železosirných center je přítomen ve všech studovaných zástupcích skupiny Preaxostyla a žádný z nich nemá mitochondriální systém ISC. 4) Arginindeiminázová (ADI) metabolická dráha produkující ATP je přítomna ve většině studovaných zástupců Metamonada a Preaxostyla a pravděpodobně představuje původní znak celé skupiny Metamonada. Distribuce a evoluční historie 3 genů ADI dráhy mezi eukaryoty je konzistentní s přítomností této dráhy u posledního společného předka eukaryot a jejich evoluční historie byla zřejmě formována častými ztrátami a laterálními přenosy.

1. Introduction

The scope of the introductory chapters is mostly limited to protistan lineages without mitochondrial genomes for the sake of brevity and relevance to the presented original work. Taxonomic names are given in their most recent generally accepted form (Adl et al. 2018) regardless of how they occur in the cited literature, unless explicitly stated otherwise. The term "anaerobic" is used to describe both microaerophilic and truly anaerobic organisms, as the true status of many of the discussed examples is not fully understood.

1.1. Historical overview

This work concerns various topics related to reductive evolution and eventual loss of mitochondria and evolution of anaerobic lifestyle and symbiosis in protists. It is therefore appropriate to open it with a brief recapitulation of the history of research and scientific thinking on these topics (Fig. 1). The rise and fall of the Archezoa hypothesis are discussed first, as this period of protistological past is crucial for our current understanding of distribution and evolution of mitochondria in eukaryotes, as well as for directing scientific interest towards anaerobically living protists not only as pathogens, but also as organisms with fascinating evolutionary histories which may illuminate the evolution of eukaryotes as a whole. The progress of genomic and transcriptomic research on anaerobic protists is discussed in order to show the presented work as a part of certain tradition of thinking and continuation of many previous efforts, as well as to introduce various concepts discussed further.

1.1.1. The Archezoa hypothesis

In the year 1983, Thomas Cavalier-Smith formulated an idea which ended up dramatically shaping protistological research for the following two decades (Cavalier-Smith 1983 in Cavalier-Smith 1987). The so called Archezoa hypothesis quickly collected supporting evidence and became a generally accepted textbook truth in the 1980s, attracting scientific interest and funding to research of previously understudied anaerobic organisms in the 1990s, leading to its own downfall and ultimate rejection under the weight of newly gathered evidence on the brink of the 21st century. The Archezoa hypothesis claimed that certain taxa of microbial eukaryotes, which were not known to harbor mitochondria, indeed completely lack these organelles, because they supposedly diverged from the rest of the eukaryotic diversity before the symbiogenetic event which resulted in the establishment of mitochondria. These taxa were grouped together in a subkingdom Archezoa, hence the name of the hypothesis. As originally

defined, subkingdom Archezoa contained 4 taxa: Microsporidia, Archamoebea, Parabasalia, and Metamonada (Fornicata and Oxymonadida).

The Archezoa hypothesis gained support from a number of early single-gene phylogenetic reconstructions of the eukaryotic tree of life. An analysis of the small ribosomal subunit rRNA of the microsporidian *Vairimorpha necatrix* placed this organism as sister to all other included eukaryotes already in 1987 (Vossbrinck et al. 1987). No other "Archezoa" sequences were present in this analysis. Later analysis of the same gene, now with broader sampling of eukaryotes, retrieved *Giardia intestinalis* (Fornicata) as the sister clade to all other eukaryotes, followed immediately by the branches of *Vairimorpha* (Microsporidia), *Trichomonas vaginalis* (Parabasalia), and the rest of eukaryotes (Kamaishi et al. 1996). The protein phylogeny of translation elongation factor EF-1α showed a branching order of 1st *Glugea plecoglossi* (Microsporidia), 2nd *G. intestinalis* (Fornicata), 3rd *Entamoeba histolytica* (Archamoebea), and 4th the rest of eukaryotes (Kamaishi et al. 1996). Multiple other single gene phylogenetic analyses performed in the 1980s and 1990s supported this general pattern, consistent with the Archezoa hypothesis, in which Microsporidia, Fornicata, Parabasalia, and Archamoebea branched together, although in various different orders, as a paraphylum at the base of the eukaryotic tree of life (Klenk et al. 1995; Viscogliosi et al. 1993; Yamamoto et al. 1997). The supposed primitive state of Microsporidia was also suggested to be supported by their prokaryotic-like ribosomes with the 16S and 23S rRNA, rather than 18S and 28S rRNA, and the 5.8S rRNA within 23S rRNA (Vossbrinck and Woese 1986). It was later shown (e.g. Philippe and Germot 2000) that these early single-gene all-eukaryotic phylogenetic studies were heavily influenced by the long branch attraction artefact. This means that the long branches of the supposed Archezoa taxa were artificially attracted to the long branch of the prokaryotic outgroup (Felsenstein 1978).

Certain cracks in the Archezoa hypothesis were known to the scientific community already at the time of its formulation. E. M. Cheissin in 1965 reported observation of "ovoid pellicular bodies" in *G. intestinalis* and suggested that these might be mitochondria changed by the oxygen-poor conditions (Cheissin 1965 in Lloyd and Harris 2002). Mitochondria-like structures were reported also in Microsporidia (reviewed in Vávra 1976). Jiří Čerkasov and colleagues in 1978 noted biochemical similarities between the hydrogenosomes of parabasalids and mitochondria (Čerkasov et al. 1978). Cavalier-Smith himself acknowledged the possibility of secondary amitochondriality, in which organisms originally endowed with aerobic mitochondria would lose these organelles as an adaptation to life in low-oxygen environments

(Cavalier-Smith 1987). Examples of such secondary amitochondriates included rumen dwelling Ciliophora and Fungi, all known to be closely related to mitochondriate taxa. This line of evidence convinced Cavalier-Smith to reformulate his hypothesis in 1991 by excluding Parabasalia from Archezoa after it was shown that several anaerobic ciliates have organelles resembling parabasalid hydrogenosomes, but with cristae-like structures which link them to mitochondria (Cavalier-Smith 1991). Further evidence of the mitochondrial origin of (at least some) hydrogenosomes came from physiological (Biagini et al. 1997) and molecular (Akhmanova et al. 1998) studies on anaerobic Ciliophora, as well as additional ultrastructural studies on anaerobic ciliates and the anaerobic chytridiomycete *Neocallimastix frontalis* (Benchimol, Durand, and Aquino Almeida 1997). These and other results led Martin Embley and colleagues to suggest that the evolutionary transition from aerobic mitochondria to hydrogenosomes is "fairly easy" and it might be also the case in parabasalids (Embley et al. 2003).

Serious challenge to the Archezoa hypothesis emerged in the middle of the 1990s, when multiple groups started reporting findings of mitochondria-derived genes in various supposed members of Archezoa (reviewed in Keeling 1998). Using PCR amplification with degenerate primers, David Horner and colleagues discovered a partial gene coding for the mitochondrial protein Cpn60 in *T. vaginalis* (Parabasalia) in 1996 (Horner et al. 1996) and Agnès Germot and colleagues discovered the mitochondrial 70-kDa heat shock protein (Hsp70) in *T. vaginalis* (Germot, Philippe, and Le Guyader 1996) and *Nosema locustae* (Microsporidia) (Germot, Philippe, and Le Guyader 1997). Andrew Roger, Graham Clark, and colleagues used a similar method to discover the mitochondrial chaperonin 60 (Cpn60) in *E. histolytica* (Archamoebea) (Clark and Roger 1995) and *T. vaginalis* (Parabasalia) (Roger, Clark, and Doolittle 1996) and later discovered this gene also in *G. intestinalis* (Fornicata) (Roger et al. 1998). Phylogenetic analyses of the genes supported their mitochondrial origin (Horner et al. 1996).

Another blow to the Archezoa hypothesis came with new development in the field of molecular phylogenetics. Phylogenetic analyses of ribosomal genes from the Archamoebea *E. histolytica*, *Mastigamoeba balamuthi* (Hinkle et al. 1994), and *Pelomyxa* sp. (Morin and Mignot 1995 in Keeling 1998) have shown their position among other mitochondriate eukaryotes. This led Cavalier-Smith to remove Archamoebea from Archezoa and group them with other amoeboid eukaryotes in the taxon Amoebozoa (Cavalier-Smith 1998). This position of Archamoebea was later supported by further, more sophisticated, phylogenetic and phylogenomic analyses (e.g. Fahrni et al. 2003; Tekle et al. 2016).

Multiple unique features were suggested to unite Microsporidia with the mitochondria-bearing Fungi (e.g. Vivares et al. 1996; Flegel and Pasharawipas 1995) and this relationship was also further supported by molecular phylogenetics of tubulins (Keeling and Doolittle 1996; Edlund et al. 1996). Therefore, Microsporidia were removed from Archezoa by Cavalier Smith as well (Cavalier-Smith 1998). These changes, together with the previous removal of Parabasalia, left only Metamonada (at that time including Fornicata and Oxymonadida) as the single putative primarily amitochondriate eukaryotic taxon and focused the interest of the protistological community towards the best studied member of Metamonada – *Giardia intestinalis*, turning it into the last pillar of the Archezoa hypothesis.

The non-archezoan nature of Archamoebea and Microsporidia was later corroborated by the discovery of a novel organelle of mitochondrial origin, the mitosome. Jorge Tovar and colleagues discovered the mitosome in 1999 when they investigated expression of the previously reported Cpn60 gene in *E. histolytica* (Tovar, Fischer, and Clark 1999) and the same observation was published the same year also by Zhiming Mai and colleagues (Mai et al. 1999). Both teams found out that the protein product of the Cpn60 gene is targeted to a previously undetected vesicular compartment in the cell of *E. histolytica*. Furthermore, rescue experiments have shown that the organellar import mechanism involved is compatible with the mitochondria-targeting signals of aerobic eukaryotes.

Three years later, Bryony Williams and colleagues reported finding of a similar organelle in the microsporidian *Trachipleistophora hominis* (B. A. P. Williams et al. 2002) (after the presence of such organelle was hypothesized based on findings of putatively mitochondrial genes in the genome of another microsporidian, *Encephalitozoon cuniculi*; Katinka et al. 2001). Their approach included immunological staining of the Hsp70 protein in both light and electron microscopy, which allowed them to observe that the target organelles are surrounded by a double membrane, a feature typical for mitochondria and their derivatives. They noted the similarities between the organelles of *T. hominis* and *E. histolytica* and suggested that such relictual mitochondria might be one day discovered also in *G. intestinalis*.

This came true only a year later when Jorge Tovar and colleagues not only reported the discovery of a mitosome in *G. intestinalis*, but also identified its function – synthesis of iron-sulfur clusters by the typically mitochondrial ISC system (Tovar et al. 2003).

All this evidence against the Archezoa hypothesis led to its gradual abandonment during the first years of the 21st century. Martin Embley and colleagues suggested in 2003 that “all

eukaryotes might contain an organelle of mitochondrial ancestry" and Cavalier-Smith himself completely rejected the idea of Archezoa the same year in an article (Cavalier-Smith 2003) where he also presented a new pan-eukaryotic phylogeny and classification, which already broadly resembled our current idea about the evolution of eukaryotes (Adl et al. 2018). In this classification, Metamonada were expanded to include not only Fornicata and Oxymonadida like in previous two decades, but also the entire Parabasalia and the genus *Trimastix* (current genera *Trimastix* and *Paratrimastix*), which had been shown to be a sister clade to Oxymonadida two years earlier by Joel Dacks and colleagues (Dacks et al. 2001).

Discovery of the relationship between oxymonads and *Trimastix* has shown that neither oxymonads, the least studied group of the supposed Archezoa, are primitively amitochondriate, because species of the genus *Trimastix* were known to contain hydrogenosome-like organelles of likely mitochondrial origin. Joel Dacks and colleagues suggested that the taxon containing *Trimastix* and Oxymonadida, now known as Preaxostyla, may be a useful system for investigations into the reductive evolution of mitochondria (Dacks et al. 2001).

1.1.2. The age of genomics and transcriptomics

Genomics and transcriptomics, i.e. the study of the entire information content of DNA and mRNA molecules in the cell, are powerful tools for understanding non-model organisms, especially in cases where unavailability of axenic culturing and/or genetic manipulation techniques precludes biochemical investigations and complicates functional and protein localization experiments.

1.1.2.1. Microsporidia

Microsporidia are a group of obligate intracellular parasites of substantial medical and economic importance, which also in some cases have extremely small genome sizes compared to other eukaryotes, even on the scale of single megabases (e.g. Biderre et al. 1994; Peyretaillade et al. 1998; Keeling and Slamovits 2004). Therefore, it is no surprise that a member of Microsporidia became the first eukaryote devoid of aerobic mitochondria to have its genomic sequence published. Michaël Katinka and colleagues sequenced, assembled, and annotated the genome of the mammal parasite *Encephalitozoon cuniculi* and published their results in 2001 (Katinka et al. 2001). The study used the genomic data for greater understanding of multiple aspects of *E. cuniculi* biology, ranging from the structure of the genome itself, to cell biology and also metabolic and physiological features associated with the parasitic lifestyle.

What is of special interest to this thesis is the authors' attempt to reconstruct the features of a supposed mitochondrial organelle of *E. cuniculi* based on the predicted set of putatively mitochondrion-targeted proteins encoded in the nuclear genome. Similar analysis was also performed on the second member of Microsporidia with a published genome, the honeybee pathogen *Nosema ceranae* in 2009 (Cornman et al. 2009). Now, 10 years later, there are 35 available genome assemblies of Microsporidia deposited in the specialized database MicrosporidiaDB (https://microsporidiadb.org).

1.1.2.2. *Cryptosporidium*

Genome assemblies of two species of *Cryptosporidium*, an apicomplexan parasite infecting human gut and known not to have an aerobic mitochondrion, were published in 2004: *C. parvum* by Mitchell Abrahamsen, Thomas Templeton, and colleagues (Abrahamsen et al. 2004), and *C. hominis* by Ping Xu, Giovanni Widmer, and colleagues (P. Xu et al. 2004). *Cryptosporidium parvum* had been shown to have a reduced mitochondrial organelle (Riordan et al. 2003), possibly involved in the synthesis of Fe-S clusters (LaGier et al. 2003), only a year before and the genomic studies on both species have identified gene complements consistent with these observations.

Based on the predicted proteins putatively localized in the mitochondrion, the authors have reconstructed a picture of an organelle devoid of the TCA cycle and oxidative phosphorylation. Other reported findings are related to lineage-specific features (lack of apicoplast, presence of apical complex) and to the parasitic way of life (limited biosynthetic capabilities, novel types of cell-surface and secreted proteins). The simultaneous availability of genomic data from 2 *Cryptosporidium* species allowed the authors to make comparisons between the genomes and conclude that they are strikingly similar in both composition and structure, which contrasts with significant phenotypic differences observed in the 2 organisms (P. Xu et al. 2004). Another 12 genome assemblies of the genus *Cryptosporidium* have been published since then and are deposited in the specialized database CryptoDB (https://cryptodb.org).

1.1.2.3. Archamoebea

The first archamoebid (and the first amoebozoan) to have its genome sequenced and published was *Entamoeba histolytica*, a human gut symbiont and occasional pathogen. Brendan Loftus and colleagues used the genomic data for functional annotation of various cellular systems potentially involved in the symbiotic lifestyle and pathogenesis, including virulence factors like

lectins, proteases, and amoebapores, as well as vesicle trafficking, signal transduction, and oxygen detoxification mechanisms (Loftus et al. 2005). *E. histolytica* had been previously shown to possess a reduced mitochondrial organelle and the genomic data supported the lack of an aerobic mitochondrion. However, the authors did not discuss other insights into the nature of the organelle. They also conducted a phylogenetic screen for laterally transferred genes in the *E. histolytica* genome and identified 96 LGT candidates with a major impact in energy metabolism catabolizing amino acids and carbohydrates. As of now, there are genomic assemblies from 8 different *E. histolytica* strains, as well as 4 other *Entamoeba* species, deposited in the database AmoebaDB (https://amoebadb.org).

Mastigamoeba balamuthi is a free-living relative of *E. histolytica*. In 2007 Erin Gill and colleagues performed a transcriptomic survey of this archamoebid in order to understand metabolic capabilities of its MRO and, by comparing it to the mitosome of *E. histolytica*, shed light on the evolution of mitochondria in Archamoebea (Gill et al. 2007). The predicted MRO proteome suggested a much more complex organelle than the mitosome of *E. histolytica* – probably a hydrogenosome with a role in energy and amino acid metabolism. Genomic assembly of *M. balamuthi*, was published in 2013 (Nývltová et al. 2013). Eva Nývltová and colleagues sequenced the genome in order to explore cellular localization of an unusual iron-sulfur cluster assembly system of *M. balamuthi*. The genomic data was later used to investigate the evolutionary history of various hydrogenosomal proteins (Nývltová et al. 2015).

1.1.2.4. Parabasalia

The first genome assembly of the best studied parabasalid, the human sexually transmitted pathogen *Trichomonas vaginalis*, was published in 2007 by Jane Carlton and colleagues (Carlton et al. 2007). Genome of this organism is surprisingly large and the reason for this became one of the questions investigated within the genomic project. The authors identified a high number of repetitive sequences with low polymorphism. This, together with an analysis of gene families age distribution, led them to conclude that *T. vaginalis* genome recently underwent a period of increased duplication including several large-scale duplications. They also reported and discussed annotation of a large number of various cellular systems, often with focus on pathogenesis (including surface virulence factors discussed in a separate publication; Hirt et al. 2007). Among these is an extensively annotated amino acid metabolism, which was hypothesized to have an important role in energy metabolism. Interestingly, many of the amino

acid metabolism-related genes were shown to have likely originated via lateral gene transfer from bacteria, which is reminiscent of the *E. histolytica* genome discussed above.

The authors also identified 138 genes whose protein products are putatively targeted into the hydrogenosome, the reduced mitochondrion of *T. vaginalis*. Annotation of these genes not only supported previously reported functions of the hydrogenosome-like ATP production via an anaerobic pathway involving hydrogenases and iron-sulfur cluster biosynthesis, but also suggested a possible novel role of hydrogenosome in the amino acid metabolism due to the presence of components of incomplete glycine cleavage system (GCS) and serine hydroxymethyl transferase (SHMT). Identification of multiple copies of hydrogenosome-targeted ferredoxin gene in the *T. vaginalis* genome allowed the authors to explain mechanisms behind the clinically important resistance to 5-nitroimidazole drug metronidazole.

Ten years later, in 2017, *Tritrichomonas foetus*, a parasite of significant veterinary and economic importance, became the only other parabasalid to have its genome sequenced and published to this day (Benchimol et al. 2017).

There are also transcriptomic assemblies available for 2 other members of Parabasalia: the human bowel parasite *Dientamoeba fragilis* and the poultry pathogen *Histomonas meleagridis*. The accompanying publications focused mostly on various pathogenesis-related systems of these organisms (Klodnicki, McDougald, and Beckstead 2013; Barratt et al. 2015; Mazumdar et al. 2017). The hydrogenosomes of both protists were predicted to have protein complements and metabolic capabilities similar to those of *T. vaginalis.* The arginine deiminase pathway was predicted to be complete in both.

1.1.2.5. Fornicata

A project designed to sequence a complete genome of *Giardia intestinalis* and directed by Mitchell Sogin started already in 1998 (Adam 2000) when it was partially justified by the possibility of *G. intestinalis* being a true archezoan, possibly the most primitive eukaryote alive (Knight 2004). However, the results were not published until 2007 (Morrison et al. 2007), when the Archezoa hypothesis has already been disproved, as discussed above. Minimalism and evolutionary reduction in genome structure, information machinery, cellular structure, and metabolism became the main theme of the *G. intestinalis* genomic publication by Hilary Morrison and colleagues (Morrison et al. 2007). These systems are simplified compared not

only to aerobic model eukaryotes, but also to *E. histolytica* and *T. vaginalis* which exhibit comparable anaerobic and endobiotic lifestyles.

Reconstruction of the amino acid, purine, pyrimidine, and lipid metabolism implies a very limited biosynthetic capabilities of *G. intestinalis*. The authors identified circa 100 LGT candidates, many of which are putatively involved in energy metabolism. This is reminiscent of the situation in *E. histolytica* and *T. vaginalis*. Moreover, several of the LGT candidates were shown to have a close phylogenetic affinity to their homologs from these 2 organisms specifically. Regarding the mitochondrial organelle of *G. intestinalis*, the mitosome, the authors identified several components of the mitochondrial protein import and maturation machinery, as well as the iron-sulfur cluster assembly system. No novel functions of the mitosome were suggested based on the genomic project. The subsequent genomic projects performed on other isolates of *G. intestinalis* representing different assemblages (possibly species) of the parasite were focused mostly on establishing phylogenetic relationships between the assemblages, as well as understanding cellular mechanisms underlying biological and clinical differences between them (Franzén et al. 2009; Jerlström-Hultqvist et al. 2010; Adam et al. 2013; Pollo et al. 2018). As of now, there are 7 genomic assemblies from various isolates of *G. intestinalis* deposited in the database GiardiaDB (https://giardiadb.org/).

The second diplomonad lineage to have its genome sequenced and published was *Spironucleus salmonicida,* an economically important parasite causing systemic infections in salmonid fish. Feifei Xu and colleagues focused on comparing *S. salmonicida* genomic data to the previously published genome of *G. intestinalis* in order to illuminate the adaptations behind different lifestyles of the two relatively closely related diplomonads (F. Xu et al. 2014). From this comparison, almost every cellular system of *S. salmonicida* comes out as more elaborate and complicated than in *G. intestinalis*. *S. salmonicida* was predicted to be able to utilize 5 more carbohydrates in glycolysis than *G. intestinalis*.

Regarding amino acid metabolism, *Spironucleus salmonicida* can catabolize not only arginine (thanks to the arginine deiminase pathway) like *G. intestinalis*, but also tryptophan for ATP production. Amino acid biosynthetic capabilities were also shown to be greater than in *G. intestinalis*, including the capacity to synthesize selenocysteine, which is not present in either *G. intestinalis* or *T. vaginalis*. *Spironucleus salmonicida* had been previously shown to harbor hydrogenosomes (Jerlström-Hultqvist et al. 2013) and the genomic project supported this observation, as well as the probable dual localization of the ATP-producing pyruvate metabolism in both the hydrogenosome and cytosol. *Spironucleus salmonicida* was shown to

have a greater complement of metabolite transporters and proteins involved in the oxidative stress response than *G. intestinalis* and *T. vaginalis*, likely making it better adapted to its less predictable environment. Among the *S. salmonicida* systems newly identified thanks to the genomic project were the encystation machinery, hinting at a previously unobserved capability of cyst formation, and novel groups of cysteine-rich membrane proteins, possibly analogous to the *G. intestinalis* variant-specific surface proteins.

G. intestinalis and *S. salmonicida* are both parasitic and highly derived diplomonads. Comparison between their respective genomes can bring insights into the nuances of their different lifestyles, but is insufficient to illuminate the larger evolutionary stories, e.g. the reduction of mitochondria or adaptations to anaerobiosis and parasitism. For this reason, Goro Tanifuji and colleagues sequenced and published the genome of *Kipferlia bialata* (Tanifuji et al. 2018), a free-living relative of diplomonads belonging to a paraphyletic assemblage called *Carpediemonas*-like organisms (CLOs), which is thought to be a better representative of the ancestral state of Fornicata. In a global analysis of the *K. bialata* genome coding capacity and a 3-way comparison with the 2 diplomonads, *K. bialata* showed the greatest diversity of protein functions represented by number of unique orthologous groups. This suggest a general reduction in cellular functions on the evolutionary path from the free-living CLOs to the parasitic diplomonads.

Hydrogenosome of *K. bialata* was predicted to possess a greater number of proteins, as well as metabolic functions, than the MROs of both *G. intestinalis* and *S. salmonicida*, being able not only to produce ATP via a *T. vaginalis*-like hydrogenosomal pathway, but also to catabolize the amino acid glycine via glycine cleavage system. In addition to the hydrogenosomal pyruvate-dependent ATP production, *K. bialata* was shown to also have a cytosolic pathway depending on a different enzyme catalyzing the last step. One notable exception to the generally higher richness of *K. bialata* protein complement compared to diplomonads are surface-localized proteins, which likely expanded in diplomonads as an adaptation to the parasitic lifestyle, as they mediate interactions with the host, including immune system evasion. Based on these observations, the authors proposed an evolutionary scenario in which the transition from the free-living CLOs to the parasitic diplomonads was partially facilitated by an expansion of the surface proteins, while the resulting reliance on host-supplied metabolites allowed diplomonads to lose many other functions and streamline their genomes.

Another provisional genome assembly of a CLO was produced for *Carpediemonas frisia* in a metagenomic project focused on elucidating the syntrophic interspecies interactions between

this protist and a community of prokaryotes (Hamann et al. 2017). Based on a metabolic network map inferred from their metagenomic data, Emmo Hamann and colleagues deduced that *C. frisia* stimulates growth of the associated prokaryotic community via excretion of partially digested nutrients and molecular hydrogen, while *C. frisia* itself is benefiting by having the waste hydrogen removed from environment, and also by an increased availability of prokaryotic prey. The energy metabolism of *C. frisia* was predicted to take place exclusively in cytosol.

CLOs also became the primary focus of an extensive and ambitious comparative transcriptomic project led by Michelle Leger, Martin Kolísko, and Ryoma Kamikawa published in 2017 (Leger et al. 2017). Transcriptomic assemblies of 5 CLOs (*Ergobibamus cyprinoides*, *Carpediemonas*-like NY0171, *Dysnectes brevis*, *Carpediemonas membranifera*, and *Kipferlia bialata*), 2 retortamonads (*Chilomastix cuspidate*, *Chilomastix caulleryi*), and 1 trimastigid (*Trimastix marina* PCT) were made available within the scope of this project. The authors used this data to reconstruct metabolic functions of the MROs of these organisms and compare them to other, previously characterized, metamonad MROs. Based on this comparison, they were able to reconstruct a hypothetical evolutionary scenario detailing various aspects of the MROs evolution.

1.1.2.6. Heterolobosea

Sawyeria marylandensis is a microaerophilic free-living amoeba belonging to Psalteriomonadidae, Heterolobosea. Maria José Barberà, Iñaki Ruiz-Trillo, and colleagues combined a transcriptome sequencing approach with a PCR survey and ultrastructural observations to characterize its MRO and published their results in 2010 (Barberà et al. 2010). The reconstructed MRO-localized energy metabolism resembles the situation in hydrogenosomes of *T. vaginalis*. The MRO was also predicted to host several other metabolic pathways not known from other hydrogenosomes, mostly related to amino acid metabolism. Components of a canonical ISC machinery for iron-sulfur cluster biogenesis was found in the transcriptomic data and predicted to be localized in the MRO.

1.1.2.7. Rhizaria

Mikrocytos mackini is an economically important causative agent of the Denman island disease in oysters (Hine et al. 2001). Its phylogenetic position used to be a total mystery even after its SSU rDNA was sequenced. Moreover, there was no evidence of a mitochondrion in the

M. mackini cell and it was not known, what kind of mitochondrion, if any, is present. Fabien Burki and colleagues used a transcriptomic approach to answer these two questions and published their results, together with a transcriptomic assembly, in 2013 (Burki et al. 2013). The obtained transcriptomic data allowed the authors to perform a robust phylogenomic analysis, which placed *M. mackini* in Rhizaria and showed an extreme rate of molecular evolution in the organism, explaining previous failures to place it using less sophisticated phylogenetic approaches. The search for mitochondrial genes revealed only 4 candidates, all involved in the ISC machinery for iron-sulfur cluster assembly. This implies that *M. mackini* likely possess a highly derived and reduced MRO analogous to the mitosome of *G. intestinalis*.

1.1.2.8. Breviatea

Pygsuia biforma is a recently described amoeboid flagellate living freely in hypoxic estuarine sediments and lacking an aerobic mitochondrion. Phylogenetic analyses of SSU rDNA sequences have placed *P. biforma* in Breviatea, a deep-branching eukaryotic group of then uncertain phylogenetic position. The transcriptomic assembly of *P. biforma* was first used by Matthew Brown and colleagues in 2013 for a large scale phylogenomic analysis (Brown et al. 2013). This project placed Breviatea together with Apusomonadida and Opisthokonta in a new eukaryotic clade named Obazoa. The authors also discovered and annotated a nearly complete integrin-mediated adhesion complex in the *P. biforma* transcriptome and discussed the implications of this discovery for understanding of the evolution of multicellularity in the related Metazoa lineage.

The transcriptomic data was also used by Courtney Stairs and colleagues to reconstruct the proteome and functions of the *P. biforma* MRO (Stairs et al. 2014). The organelle was predicted to have a number of unique features including cardiolipin, amino acid, rhodoquinone, and fatty acid metabolism, as well as remnants of the TCA cycle and respiratory chain, making it a very complex version of hydrogenosome. The most surprising feature of the MRO is however the complete lack of the typical mitochondrial ISC system for iron-sulfur cluster assembly, which was apparently replaced by a laterally transferred bacterial SUF system, also localized in this organelle.

1.1.2.9. Stramenopiles

Cantina marsupialis is a free-living flagellate isolated from anoxic sediments, which represents an independent deep-branching lineage of Stramenopiles (Yubuki et al. 2015). Fumiya Noguchi

and colleagues produced a transcriptomic assembly of *C. marsupialis* and searched it for putatively mitochondrial proteins (Noguchi et al. 2015). The reconstructed MRO metabolism, published in 2015, is more complex than that of *T. vaginalis* hydrogenosome. Aside from pyruvate-decarboxylating energy metabolism and ISC system, the MRO also harbors multiple amino acid metabolism-related pathways including a complete glycine cleavage system, an incomplete TCA cycle, and all the subunits of respiratory chain complex II.

1.1.2.10. Jakobida

Stygiella incarcerata is an anaerobic free-living flagellate nested within the aerobic Jakobida. The MRO of *S. incarcerata* became the focus of a transcriptomic project by Michelle Leger and colleagues published in 2016 (Leger et al. 2016). They identified a typical hydrogenosomal *T. vaginalis*-like energy metabolism located in the MRO, as well as the ISC machinery. The organelle was however predicted to harbor a more diverse set of other metabolic pathways than the hydrogenosome of *T. vaginalis*, including e.g. complete glycine cleavage system, branched-chain amino acid degradation, and cardiolipin biosynthesis. In this respect, the MRO of *S. incarcerata* seems to resemble MROs characterized in other free-living anaerobes like *Pygsuia biforma* or *Cantina marsupialis*, rather than those of parasites.

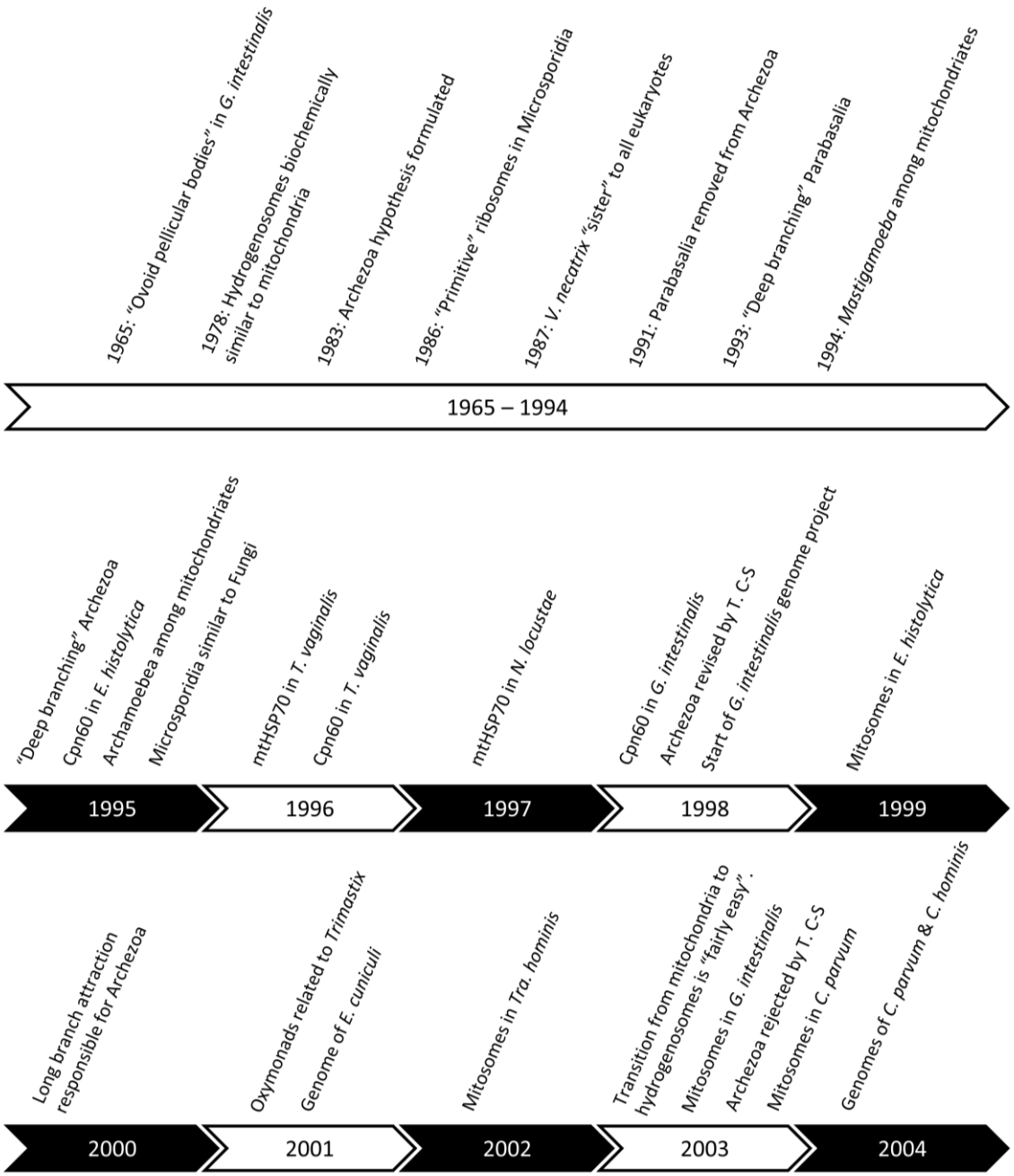


Figure 1a: Timeline of publications discussed in the historical overview (1965 – 2004).

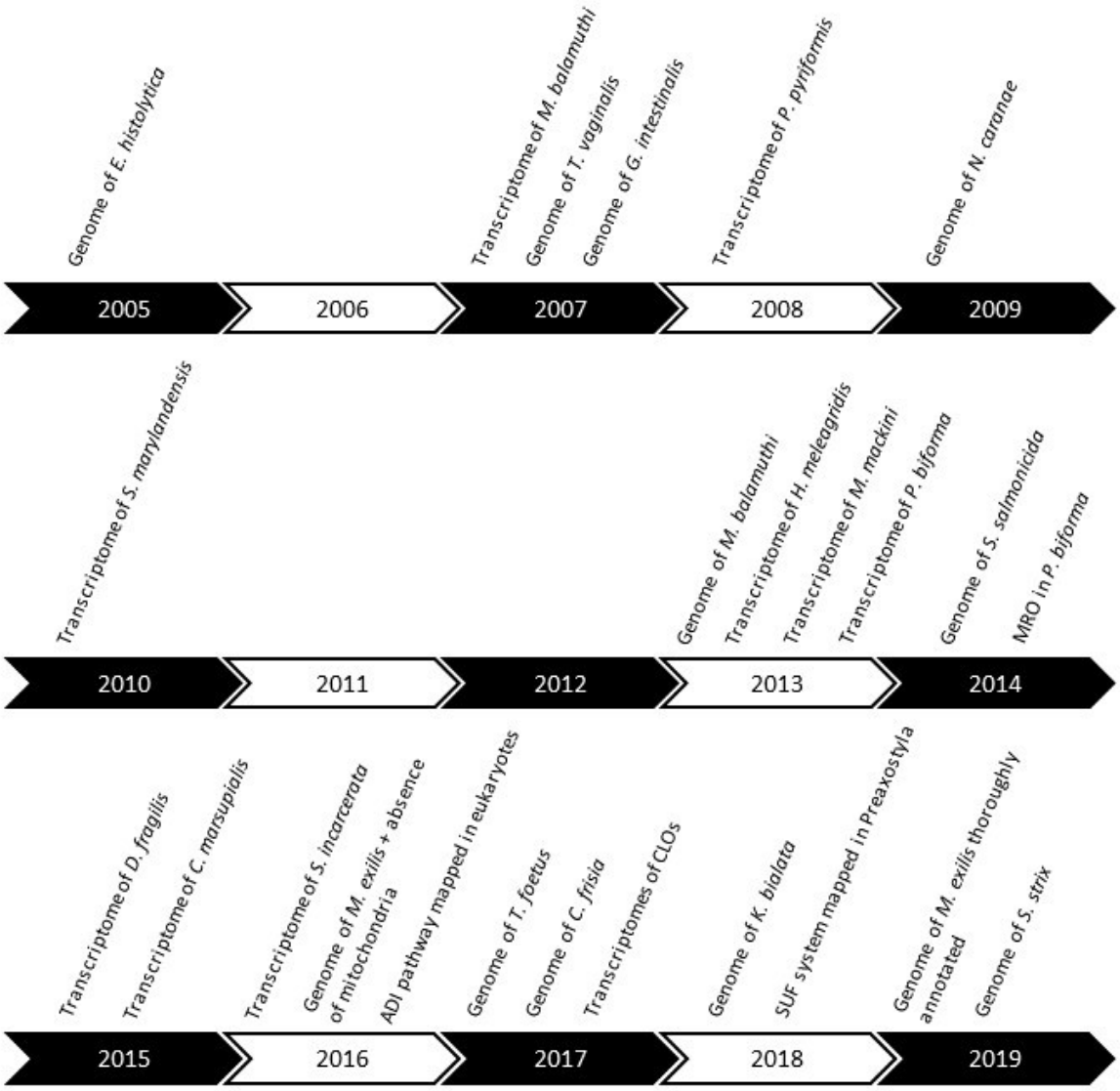


Figure 1b: Timeline of publications discussed in the historical overview (2005 – 2019).

1.2. Selected cellular systems related to anaerobiosis and endobiosis in Metamonada

Mitochondrion-related organelles, energy metabolism, iron-sulfur clusters assembly, amino acid metabolism, and plasma membrane transporters are all significantly altered in various anaerobic and endobiotic protists. Following chapters discuss these and other systems in representatives of Metamonada, the deep eukaryotic group to which Preaxostyla, the topic of the original work presented here, belong. *Trichomonas vaginalis*, the most studied and best understood metamonad, is used as a baseline for introduction of the discussed systems.

1.2.1. Mitochondrion-related organelles

Anaerobic derivatives of mitochondria are traditionally divided into 4 classes (M. Müller et al. 2012). Anaerobic mitochondria and hydrogen-producing mitochondria retain their genomes, relictual cristae, and a rich complement of electron transport chain proteins. Hydrogenosomes and mitosomes have no genome and usually no remnants of cristae. Members of Metamonada are known to contain hydrogenosomes and mitosomes. However, the distinction between these classes is being blurred by recent discoveries of organelles which do not fit perfectly into either of these categories, e.g. the reduced hydrogenosomes described in *Dysnectes brevis* (Leger et al. 2017).

1.2.1.1. Hydrogenosomes of *Trichomonas vaginalis*

Arguably the best understood MROs in protists are the hydrogenosomes of *T. vaginalis* and other parabasalids (reviewed in Hrdý, Tachezy, and Müller 2019). Metabolic capabilities of these organelles have been studied by biochemical methods since the 1970s and the publishing of *T. vaginalis* genome assembly in 2007 brought support for previous insights and allowed more detailed studies of hydrogenosomal protein complement and physiological functions.

1.2.1.1.1. Energy metabolism

The signature metabolic pathway of hydrogenosomes is the extended glycolysis, which produces acetate, CO_2, and H_2 from pyruvate, while performing substrate level phosphorylation of ADP to ATP (Lindmark and Müller 1973). The pathway uses pyruvate as a substrate, which is either directly imported into the hydrogenosome from cytosol or produced within the hydrogenosome by decarboxylation of imported malate through activity of malic enzyme (ME), the most abundant hydrogenosomal protein (Hrdý, Tachezy, and Müller 2019).

Pyruvate is decarboxylated to acetyl coenzyme A and CO_2 in a reaction catalyzed by pyruvate:ferredoxin oxidoreductase (PFO) (K. Williams, Lowe, and Leadlay 1987). Carbon dioxide is then passively excreted from the hydrogenosome. Both the decarboxylation of malate by ME and pyruvate by PFO are oxidative processes which release electrons. The final fate of the electrons from both reactions lies in the reduction of protons to molecular hydrogen (which is then passively excreted from the hydrogenosome), through the activity of [FeFe] hydrogenases (HydA) (Payne, Chapman, and Cammack 1993). However, the paths in which the electrons reach their destination are different.

Electrons released by PFO are immediately transferred to a small electron transfer protein ferredoxin (Fdx) (K. Williams, Lowe, and Leadlay 1987) and then to HydA. The Fdx is reoxidized by HydA and can be again used to carry another electron from PFO. The electrons released by ME are first carried by reduced pyridine nucleotide coenzyme NADH (Drmota 1996) and then transferred to HydA in cooperation with two subunits of the mitochondrial respiratory complex I (NuoE, NuoF), possibly by a process called electron bifurcation, where the transfer of the electron from NADH is driven by a simultaneous transfer of another electron from ferredoxin (Li et al. 2008; Hrdý et al. 2004; M. Müller et al. 2012).

Decarboxylation of pyruvate results in acetyl-CoA, which includes a high energy thioester bond. Energy of this bond is harvested in a subsequent step, composed of two reactions. First, the CoA moiety is transferred to succinate by acetate:succinate CoA transferase (ASCT) (van Grinsven et al. 2008), releasing acetate, which is then excreted from the hydrogenosome. The resulting succinyl-CoA then serves as a substrate for the only present enzyme of the TCA cycle, the succinyl-CoA synthetase (SCS) (Jenkins et al. 1991), which cleaves it into succinate and CoA, while concurrently forming ATP by substrate-level phosphorylation (Steinbüchel and Müller 1986).

The [FeFe] hydrogenase requires hydrogenase maturases (HydE, HydF, and HydG) for proper function and indeed, all the 3 proteins are found in *T. vaginalis* hydrogenosomes (Pütz et al. 2006). Adenylate kinase (ADK) is another protein involved in energy metabolism which is localized in the hydrogenosomes of *T. vaginalis*. This enzyme interconverts ATP, AMP, and ADP, keeping the cellular energy homeostasis.

1.2.1.1.2. Protection against reactive oxygen species and oxygen

Several of the hydrogenosomal energy metabolism enzymes are sensitive to deactivation by molecular oxygen (Lindmark and Müller 1973). Moreover, dangerous reactive oxygen species (ROS) can be created from oxygen by activity of the hydrogenosomal enzymes (Docampo, Moreno, and Mason 1987). Hydrogenosomes contain multiple enzymes dedicated to minimizing this threat by reducing oxygen and ROS to less harmful compounds.

Molecular oxygen can be reduced to water by a class A flavodiiron protein (FDP), which gains the electrons needed for the reaction from reduced Fdx. Oxygen can be reduced also by iron-sulfur flavoproteins (ISF). However, the end product is hydrogen peroxide, not water. ISF can gain electrons from both reduced Fdx and NADH (Smutná et al. 2009).

Superoxide radicals are converted to hydrogen peroxide by the superoxide dismutase (SOD) (Viscogliosi et al. 1998). Interestingly, the most common enzyme able to remove hydrogen peroxide, catalase, has not been found in *T. vaginalis*. The organism likely uses several alternative pathways with peroxidase activity for hydrogen peroxide detoxification.

The simplest one is a single enzyme rubrerythrin (RBR) with a diiron center and a rubredoxin domain. RBR likely uses electrons from the reduced NADH (Pütz et al. 2005). The peroxiredoxin system consists of 3 enzymes: thioredoxin peroxidase (TrxP), thioredoxin (Trx), and thioredoxin reductase (TrxR), which gains electrons from reduces NADPH (Coombs et al. 2004). The most unique system of *T. vaginalis* hydrogenosomes with peroxidase activity is the OsmC protein working in cooperation with 2 components of the incomplete glycine cleavage system: proteins H and L, which transfer electrons from NADH (Nývltová et al. 2016).

1.2.1.1.3. Amino acid metabolism

In its complete form, the glycine cleavage system (GCS) consists of 5 proteins: H, L, T, and 2 subunits of P. This (in eukaryotes) exclusively mitochondrial metabolic pathway is functionally connected with serine hydroxymethyltransferase (SHMT) and together they: 1) interconvert the amino acids serine and glycine, and 2) provide methyl residues into the one carbon pool metabolism by folate, and subsequently to multiple biosynthetic pathways including the biosynthesis of amino acids and nucleobases (reviewed in Douce et al. 2001).

Hydrogenosomes of *T. vaginalis* contain SHMT, which likely serves its canonical purpose of interconverting between serine and glycine (Mukherjee, Sievers, et al. 2006), and 2 components of GCS: H and L. These alone are unable to perform the activity of GCS and were shown to be

involved in ROS detoxification instead, as discussed above (Mukherjee, Brown, et al. 2006; Nývltová et al. 2016).

T. vaginalis hydrogenosomes have also been shown to contain arginine deiminase (ADI) catalyzing conversion of arginine to citrulline, the first reaction of the arginine deiminase pathway (Morada et al. 2011). Citrulline is further converted to ornithine and carbamoyl phosphate by cytosolic ornithine transcarbamylase (OTC). Resulting ornithine is used in polyamine biosynthesis, while carbamoyl phosphate is catabolized by cytosolic carbamate kinase (CK) while ATP is produced. These 2 other enzymes are cytosolic (Yarlett et al. 1994).

1.2.1.1.4. Iron-sulfur cluster assembly

The only physiological function which unites hydrogenosomes with aerobic mitochondria and other types of MROs including most of mitosomes is the synthesis of iron-sulfur (FeS) clusters via the ISC machinery (reviewed in Lill et al. 2015). Hydrogenosomes of *T. vaginalis* have been shown to be able to perform this function and multiple proteins of the ISC machinery have been identified (Tachezy, Sanchez, and Muller 2001; Sutak et al. 2004).

Cysteine desulfurase IscS provides sulfur by converting cysteine to alanine, while frataxin (Fxn) is responsible for delivery of iron. Sulfur and iron are then combined on the scaffold protein IscU/IscA to form the cluster. NFU and Ind1 are likely involved in cluster transfer to apoproteins. Interestingly, no components of the system responsible for FeS cluster export to the cytosol (e.g. the ABC transporter Atm1) were found in *T. vaginalis*.

1.2.1.1.5. Protein import and maturation

Outer hydrogenosomal membrane contains 4 beta-barrel proteins: Tom40, Sam50, Hmp35, and Hmp36. Tom40 forms the pore which mediates translocation of proteins across the hydrogenosomal outer membrane and is associated with 3 tail-anchored proteins forming the translocase of the outer membrane (TOM) complex (Tom22, Tom36, Tom46), some of which may serve as TOM receptors. Sam50, which is also physically associated with the TOM complex, facilitates insertion of beta-barrel proteins into the outer membrane. Hmp35 and Hmp36, which are unique to *T. vaginalis*, have unknown functions (Rada et al. 2011; Makki et al. 2019).

Family of small Tims, which work in the intermembrane space and help with transporting complicated substrates, are represented by a single type in *T. vaginalis*: Tim8/9/10/13 (Rada et al. 2011).

Proteins are translocated through the inner hydrogenosomal membrane thanks to a protein homologous to the Tim17/22/23 family (Rada et al. 2011). Energy required for translocation of proteins across the inner hydrogenosomal membrane is provided by the "presequence translocase-associated motor" PAM complex, physically associated with the Tim17/22/23 family protein. A complete set of PAM complex components (compared with *Saccharomyces cerevisiae*) was discovered in *T. vaginalis*: Hsp70 (with its partners GrpE and Jac1), Tim44, Pam18, and Pam16 (Rada et al. 2011; Schneider et al. 2011).

After import, the proteins are further processed inside the hydrogenosome. The N-terminal presequence, which serves for hydrogenosome targeting, is cleaved off by the dimeric mitochondrial processing peptidase (MPP α and β). The proteins are then helped to reach their final conformations by a complex of chaperon proteins Cpn60 and Cpn10.

1.2.1.1.6. Transmembrane transporters

It was suggested that Tom40 and Hmp35 might be responsible for metabolite and ion transport across the outer membrane of the hydrogenosome (Rada et al. 2011). Transport of various metabolites across the inner membrane of mitochondria is usually facilitated by the mitochondrial carrier (MC) family proteins. There were only 5 MCs found in the hydrogenosome of *T. vaginalis* and all of them are likely belonging to the AAC type responsible for transporting ATP and ADP (Rada et al. 2011; Dyall et al. 2000).

1.2.1.2. Hydrogenosomes in free-living fornicates

Fornicata is a lineage of metamonads which includes two monophyletic endobiotic taxa Retortamonadida and Diplomonadida (although both groups also have free-living members) and a paraphyletic assemblage of free-living *Carpediemonas*-like organisms (CLOs). Features shared by CLOs hint at the ancestral state of Fornicata as a whole.

Deep-branching CLOs *Carpediemonas membranifera*, *Ergobibamus cyprinoides*, *Aduncisulcus paluster*, and *Kipferlia bialata* are predicted to have a similar protein complement and corresponding metabolic capabilities of their MROs (Fig. 2), which resemble hydrogenosomes of *T. vaginalis*. The hydrogenosomal energy metabolic pathways consist of

PFO, ASCT, SCS, Fdx, HydA (and maturases HydE, -F, -G), NuoE, and NuoF (Leger et al. 2017).

Interestingly, the transcriptomes of these organisms also show presence of acetyl-CoA synthetase (ACS), enzyme which performs the energy harvesting reaction analogous to ASCT and SCS activity in a single step, converting acetyl-CoA to acetate while concomitantly generating ATP without the intermediacy of succinate (Sanchez and Müller 1996). Michelle Leger, Martin Kolísko, Ryoma Kamikawa, and colleagues (Leger et al. 2017) suggest that ACS constitute a second branch of the pyruvate-catabolizing energy metabolism, which does not work in the hydrogenosome, but in the cytosol. This is supported by a close phylogenetic relationship between the CLOs ACS sequences and the sequence from *G. intestinalis*, and the fact that ACS is located in the cytosol in most studied organisms which possess this enzyme (Tielens et al. 2010).

Iron-sulfur cluster assembly in CLOs is of the canonical mitochondrial type, just as in *T. vaginalis*. Two signature proteins of the ISC system: IscS and IscU were found in all the studied CLOs and just like in *T. vaginalis*, no components of the FeS cluster export machinery are present.

One important difference between the hydrogenosomes of CLOs and *T. vaginalis* is the presence of a complete glycine cleavage system (GCS), together with serine hydroxymethyl transferase (SHMT), in the hydrogenosomes of CLOs. This suggests that these organelles are capable of interconverting serine and glycine, as well as catabolizing them into one carbon units useful for other metabolic pathways. Activity of GCS might be coupled with the production of H_2 by HydA, as GCS requires supply of NAD+, which could be produced by NuoE/NuoF complex, while transferring electrons to HydA (Hampson, Barron, and Olson 1983).

The hydrogenosomal protein import and maturation machinery in CLOs is very similar to that in *T. vaginalis*. Both Tom40 and Sam50 translocases of the outer membrane are present, as well as a single member of the small Tims family, Tim8/9/10/13, and the inner membrane translocase Tim23, with the associated proteins Pam18, Pam16, Tim44, and Hsp70. Both subunits of the mitochondrial processing peptidase (MPP) and the chaperon Cpn60 were also found (Leger et al. 2017).

1.2.1.3. Reduced hydrogenosomes of *Dysnectes brevis*

Dysnectes brevis is a *Carpediemonas*-like organism and the closest known free-living relative of diplomonads (Takishita et al. 2012; Yubuki, Zadrobílková, and Čepička 2017). This phylogenetic position makes it crucial for understanding the evolutionary origin of diplomonads, their endobiotic lifestyle, unique cellular architecture, and MROs.

Based on the protein complement predicted from transcriptomic data, the MRO of *D. brevis* is typical for a CLOs in protein import machinery, ISC, GCS, and the hydrogen producing pathway consisting of HydA, -E, -F, -G, Fdx, NuoE, and NuoF. However, the hydrogenosomal pyruvate metabolism enzymes ASCT and SCS are missing, while the putatively cytosolic ACS is retained (Leger et al. 2017).

This paints a picture of a mitochondrial organelle which is still capable of hydrogen production (hence a hydrogenosome) but without any ATP production. Michelle Leger, Martin Kolísko, Ryoma Kamikawa, and colleagues hypothesize that the reason for keeping the hydrogen producing pathway intact is the presence of GCS in the organelle, which requires an electron sink for its function (Leger et al. 2017).

Such a reduced hydrogenosome might represent an intermediate step between the canonical, more plesiomorphic, hydrogenosomes of other CLOs and the mitosome of *G. intestinalis*, which has lost any role in energy metabolism, as well as GCS and the hydrogen producing machinery. The loss of GCS and subsequently the hydrogen producing machinery in diplomonads might have been facilitated by their endobiotic lifestyle, which guarantees a steady supply of metabolites from the host.

1.2.1.4. Divergent hydrogenosomes of *Spironucleus salmonicida*

The MRO of the salmonid-infecting diplomonad *S. salmonicida* is a unique type of hydrogenosome (Jerlström-Hultqvist et al. 2013). The protein import and maturation machinery consists of Tom40, Pam18, Hsp70, GrpE, Jac1, and Cpn60. The ISC iron-sulfur cluster assembly machinery consists of Frataxin, IscS, IscU, and NFU. The hydrogen producing pathway is retained, consisting of HydA, -E, -F, -G, and Fdx. No NuoE and NuoF, which transfer electrons from NADH to HydA, were found. This may be connected to the fact that GCS, which requires reoxidation of NADH by the NuoE/NuoF complex, is also absent from *S. salmonicida*. The only amino acid metabolism related enzyme localized to the hydrogenosome is SHMT. Also, no ASCT and SCS of the pyruvate metabolism were found.

However, even in absence of GCS and ASCT/SCS, there is still an important role for the hydrogen producing pathway. *S. salmonicida* has 2 sets of ACS genes. One (ACS1) is closely related to the homologs from *G. intestinalis* and CLOs and clearly of the same evolutionary origin, dating back to a common ancestor of all Fornicata. Second (ACS2) is unique to *S. salmonicida* and is only distantly related to other eukaryotic homologs. These phylogenetic differences correspond to a different subcellular localization of the protein products of the ACS genes. ACS1 is localized in the cytosol, like its homologs from other Fornicata, but ACS2 is uniquely targeted to the MRO.

Jon Jerlström-Hultqvist and colleagues hypothesize that *S. salmonicida* has two parallel pathways consisting of PFO and ACS, one in cytosol and one in the hydrogenosome (Jerlström-Hultqvist et al. 2013). The hydrogenosomal pathway is releasing electrons which are then transferred by Fdx to HydA and eventually carried away by hydrogen. The ACS2 genes were probably gained by lateral gene transfer from a bacterial lineage into an ancestor of *S. salmonicida*.

1.2.1.5. Mitosomes of *Giardia intestinalis*

Mitosomes are the simplest, most reduced, forms of mitochondria. The mitosomes of *G. intestinalis* have no role in energy, nor amino acid metabolism, and do not produce hydrogen. The pyruvate-dependent energy metabolism consisting of PFO, ACS, and HydA is located in the cytosol (no hydrogenase maturases were found). The glycine cleavage system is completely missing, as well as SHMT.

The only known physiological function of the *G. intestinalis* mitosomes is the synthesis of iron sulfur clusters by the ISC machinery consisting of IscS, IscA, IscU, NFU, Grx5, and BolA. No frataxin or transmembrane export components were found (Jedelský et al. 2011). It is also unclear, where the mitosomal ISC machinery gets the ATP and reducing equivalents it requires, similarly as in *Dysnectes brevis*.

The protein import and maturation machinery of *G. intestinalis* mitosome consists of Tom40, Pam16, Pam18, Tim44, Hsp70, Cpn60, and interestingly only the β subunit of MPP (Šmíd et al. 2008). The single subunit MPP of *G. intestinalis* was experimentally confirmed to function as a monomer without the help of MPP α. This state possibly coevolved together with the shorter mitosome-targeting presequences of *G. intestinalis* proteins.

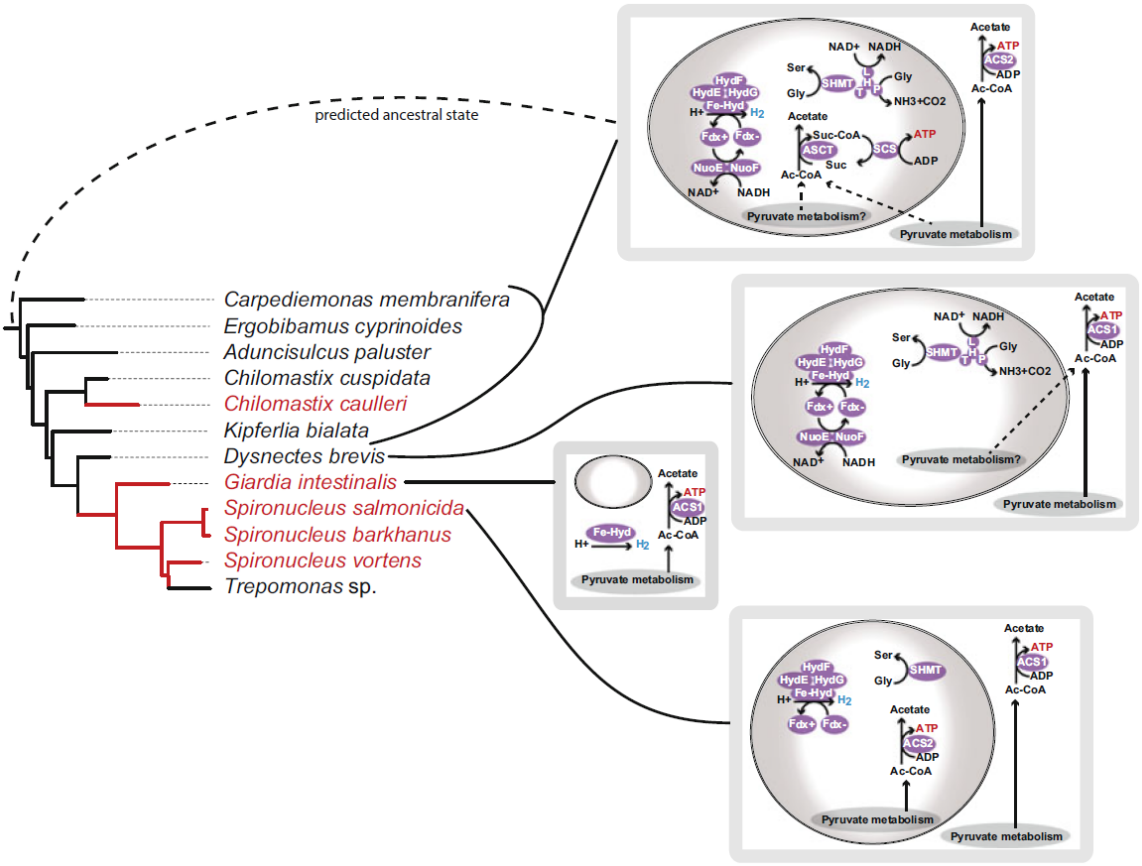


Figure 2: Putative metabolism of MROs in Fornicata. Free-living organisms are shown in black and endobiotic organisms are shown in red. (Leger et al. 2019)

1.2.1.6. Selected examples of MROs outside Metamonada

Multiple free-living anaerobic protists are predicted to have hydrogenosomes with broader metabolic capabilities than observed in *T. vaginalis*. MROs of *Sawyeria marylandensis* (Heterolobosea) (Barberà et al. 2010), *Cantina marsupialis* (Stramenopiles) (Noguchi et al. 2015), and *Stygiella incarcerata* (Jakobida) (Leger et al. 2016) have, apart from the canonical hydrogenosomal energy metabolism (by PFO, ASCT, SCS, and HydA), ISC machinery, and SHMT, also complete GCS, branched-chain amino acid degradation pathway, and alanine and aspartate aminotransferases. A similar complement of amino acid metabolism-related enzymes can be found in the hydrogenosome of *Pygsuia biforma* (Breviatea) (Stairs et al. 2014), which also contains cardiolipin, folate, and phosphatidylethanolamine biosynthesis pathways.

Michelle Leger and colleagues (Leger et al. 2016) suggest that similarities between the metabolic capabilities of the MROs of these distantly related organisms reflect their shared lifestyle of free-living anaerobic protists (endobionts like *T. vaginalis* do not need these pathways thanks to the supply of metabolites from their hosts) and analogous recent

evolutionary origins from aerobic ancestors (in contrast to e.g. free-living CLOs nested within exclusively anaerobic metamonads; Leger et al. 2016).

Pygsuia biforma is one of very few eukaryotes to have lost their ISC system for iron-sulfur (FeS) cluster biosynthesis, which is usually located in mitochondria and MROs and provides FeS clusters for the entire cell (except plastid). The ISC machinery in *P. biforma* was likely replaced by a laterally transferred fusion protein of SufB and SufC, components of the "sulfur mobilization" (SUF) machinery, which synthesizes FeS clusters in eukaryotic plastids and some lineages of prokaryotes. The SufBC gene is found in Pygsuia in 2 copies and their protein products have different cellular localization. One is localized in the cytosol and the other in the hydrogenosome (Stairs et al. 2014).

The *P. biforma* SufBC is closely related to the same gene in *Blastocystis hominis* (Stramenopiles), a human gut symbiont, which however has retained the canonical ISC machinery in its hydrogen producing mitochondrion and employs the SufBC protein exclusively in cytosol (Tsaousis et al. 2012). Phylogenetic analysis and architecture of the SufBC genes suggest that separate SufB and SufC genes were acquired laterally from a Methanomicrobiales archaeon by an ancestor of either *B. hominis* or *P. biforma*, then the fusion occurred and the fused SufBC gene was again laterally transferred to the other eukaryotic lineage.

Mastigamoeba balamuthi (Archamoebea) is a free-living amoeba from low oxygen freshwater environments. The extended glycolysis of *M. balamuthi* has similar architecture to that of *S. salmonicida*, with 2 parallel pathways consisting of PFO, ACS, and HydA in cytosol and in hydrogenosome (Gill et al. 2007). However, in the case of *M. balamuthi*, both cytosolic and hydrogenosomal ACS genes have the same evolutionary history, probably resulting from a duplication of a laterally transferred prokaryotic gene (Nývltová et al. 2015). The hydrogenosome of *M. balamuthi* also contains SHMT and GCS and, uniquely, sulfate activation pathway. This pathway, consisting of 3 enzymes, converts inorganic sulfate into 3′-phosphoadenosine-5′-phosphosulfate (PAPS) which further serves as a sulfuryl donor for various organic molecules (Patron, Durnford, and Kopriva 2008).

The ISC system was completely lost in *M. balamuthi* and replaced by a simple prokaryotic "nitrogen fixation" NIF system consisting of 2 enzymes: cysteine desulfurase NifS and scaffold protein NifU (Nývltová et al. 2013). The two proteins were shown to have a dual localization

to cytosol and the hydrogenosome. NIF system of *M. balamuthi* was probably acquired by lateral gene transfer from epsilon-proteobacteria.

Entamoeba histolytica is and endobiotic relative of *M. balamuthi* with highly reduced MROs which were the first mitosomes to be described. However, unlike other mitosomes of e.g. *G. intestinalis* or Microsporidia, the mitosomes of *E. histolytica* likely do not contain any pathway involved in FeS cluster biosynthesis. Not only is the ISC pathway lost like in *M. balamuthi*, but even the NIF system, sharing the same evolutionary origin as in *M. balamuthi*, is likely localized exclusively in cytosol. The only remaining function of *E. histolytica* mitosomes seems to be sulfate activation i.e. the production of PAPS (Mi-ichi et al. 2009), which was shown to be important for sulfolipid synthesis and cell proliferation (Mi-ichi et al. 2011).

1.2.2. Amino acid metabolism

1.2.2.1. Amino acid metabolism of *Trichomonas vaginalis*

The hypothetical metabolic map of *T. vaginalis* reconstructed based on the genomic assembly shows a potentially important role of amino acids in ATP production (Carlton et al. 2007). Multiple amino acids including serine, tryptophan, alanine, threonine, and methionine can be degraded to pyruvate or α-ketobutyrate, which both can enter the hydrogenosomal energy metabolism. This has been suggested for *T. vaginalis* already in 1995 based on biochemical experiments (Zuo, Lockwood, and Coombs 1995) and later similarly for *E. histolytica* based on reconstructed metabolic pathways (Anderson and Loftus 2005). Other way in which amino acids can contribute to energy metabolism in *T. vaginalis* is the arginine deiminase (ADI) pathway (discussed above) which converts arginine to NH_3, CO_2, and ornithine, while producing ATP. The ornithine produced by the ADI pathway can be later converted to either putrescine or proline. The energy output of the ADI pathway in *T. vaginalis* was shown to be around 10% of that from glucose metabolism (Yarlett et al. 1996).

Cysteine can be *de novo* synthesized from glycolysis intermediates, thanks to cysteine synthase (CS), and an LGT candidate (Westrop et al. 2006). Cysteine is believed to be the most important reducing agent involved in protection against ROS in *T. vaginalis* (S. Müller et al. 2003). Multiple aminotransferases were identified which likely work in amino acid degradation, but possibly may catalyze reversed reactions as well, potentially synthesizing glutamate, aspartate, alanine, glutamine, and glycine. Serine, glycine, and threonine can be interconverted thanks to serine hydroxymethyltransferase (SHMT) and threonine aldolase (TA). Interestingly, out of the

39 enzymes suggested to be involved in amino acid metabolism in *T. vaginalis*, 10 are LGT candidates. Arginine deiminase (ADI) and serine hydroxymethyltransferase (SHMT) are localized in the hydrogenosome.

Trichomonas vaginalis amino acid metabolism was a subject of recent liquid chromatography–mass spectrometry (LC-MS) studies in comparison to its relative *Tritrichomonas foetus* and under glucose restriction. Metabolic capabilities of the 2 parabasalids were shown to have major differences especially in cysteine and methionine metabolism (Westrop et al. 2017), although they generally have very similar metabolic profiles. Experiments with *T. vaginalis* under glucose restricted conditions show activation of the (ATP producing) ADI pathway and suppression of the (ATP consuming) methionine metabolic pathway, further highlighting the role of amino acid metabolism in energy economy of the organism (Huang et al. 2019).

1.2.2.2. Amino acid metabolism of *Spironucleus* spp.

The hypothetical amino acid metabolism of *Spironucleus salmonicida* containing 18 enzymes was reconstructed based on a genomic project (F. Xu et al. 2014). Similarly to *T. vaginalis*, it shows an important role of amino acids in ATP generation both via a complete ADI pathway, and conversion of tryptophan and serine to pyruvate. Both putrescine and proline can be generated from ornithine, one of the end products of the ADI pathway. Serine and glycine can be interconverted thanks to SHMT and serine can be used to synthesize cysteine in a two-step pathway consisting of serine O-acetyltransferase (SAT) and cysteine synthase (CS).

Interestingly and unlike *T. vaginalis* or *G. intestinalis*, *S. salmonicida* is capable of synthesis and incorporation of selenocysteine. Seryl-tRNA is phosphorylated by O-phosphoseryl tRNA kinase (PSTK) and selenocysteinyl-tRNA is later synthesized from it and selenophosphate by the action of O-phosphoseryl-tRNA(Sec) selenium transferase (SEPSECS). Selenocysteine-specific elongation factor (SelB), which is necessary for insertion of selenocysteine into proteins, is also present in *S. salmonicida*. Selenophosphate can be liberated from selenocysteine by a selenophosphate synthetase (SEPHS)-NifS fusion protein, while yielding alanine.

The amino acid metabolism was also investigated by biochemical methods in a related species *Spironucleus vortens* (Millet et al. 2011). Alanine and aspartate were shown to be produced by the organism, while lysine, arginine, leucine, and cysteine were consumed. However, the

experiments failed to measure any influence of the availability of these amino acids on growth rates of *S. vortens*.

1.2.2.3. Amino acid metabolism of *Giardia intestinalis*

Amino acids are an important energy source for *G. intestinalis*. Aspartate can be converted to malate via oxaloacetate by the activity of aspartate transaminase (AAT). The ADI is also present. Unlike *T. vaginalis* or *S. salmonicida*, *G. intestinalis* does not convert ornithine (end product of ADI pathway) to putrescine or proline but exports it in exchange for extracellular arginine (Schofield et al. 1995). The ADI pathway in *G. intestinalis* is predicted to produce up to 8 times more ATP than the glucose metabolism, making it the most productive energy producing pathway in the organism (Schofield et al. 1992). The pathway also plays an important role in *G. intestinalis* pathogenesis via extracellular arginine depletion (Stadelmann et al. 2013).

Two amino acids were detected to be produced by *G. intestinalis*: valine and alanine (Paget et al. 1993). Alanine is produced from pyruvate by alanine aminotransferase (ALAT) as a by-product of energy metabolism, especially under strictly anaerobic conditions (Paget et al. 1990). Regulation of alanine concentration in the cell is used by *G. intestinalis* for coping with osmotic changes in its environment (Schofield et al. 1995). The presumed biosynthetic pathway producing valine is not known in *G. intestinalis*. Cysteine is not synthesized *de novo* by *G. intestinalis* (Lujan and Nash 1994), although it plays an important role in protection against oxidative stress (Gillin and Diamond 1981). Neither glycine cleavage system or SHMT are present in *G. intestinalis*.

1.2.3. Transporters across plasma membrane

Transition from free-living to endobiotic lifestyle in protists is often accompanied by reduction in metabolic functions, especially biosynthetic pathways, as many metabolites can be scavenged from the stable and predictable environment of the host (reviewed in Dean et al. 2014). This can be illustrated e.g. by the comparatively low number of amino acids being *de novo* synthesized by the parasitic metamonads as discussed above. The reduction of biosynthetic capabilities emphasizes the importance of various plasma membrane-localized transporter proteins which facilitate scavenging of nutrients from outside the cell.

Paul Dean and colleagues reported a general pattern in the nutrient transporters repertoire they observed in a broad sample of parasitic protists (Dean et al. 2014). The number of transporter families is reduced, possibly thanks to the predictability of the host environment and the

selective pressure favoring genome minimalism and reduced antigenicity in parasites. On the other hand, the remaining transporter families are functionally diversified, gaining a broader substrate range, which might be connected with a large number of individual paralogues. Other category of transporters which is expanded in parasites are various exporters specific for toxic antimicrobial compounds.

In *T. vaginalis*, the largest known group of surface transporters is the ATP-binding cassette (ABC) superfamily with 98 paralogues (Kay et al. 2012). These proteins, powered by hydrolysis of ATP, serve almost exclusively as exporters in eukaryotes and are often involved in export of harmful compounds from the cell (multidrug resistance proteins – MRP; reviewed in Sodani et al. 2012). Another large group of ATP-consuming surface transporters in *T. vaginalis* are the P-type ATPases with 33 paralogues (Shah et al. 2002). These ion-transporting pumps and phospholipid translocases can be localized either on plasma membrane, or in the endoplasmic reticulum.

Among the secondary transporters (uniporters, symporters, and antiporters) in *T. vaginalis* are very common members of the multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) flippase superfamily with 47 paralogues – exporters of various organic compounds, often antimicrobial drugs as in the multi antimicrobial extrusion (MATE) family. Another common secondary transporter group in *T. vaginalis* is the major facilitator superfamily (MFS) (Marger and Saier 1993) with 57 paralogues. MFS has a broad range of functions including import of sugars (sugar porters – SP and GPH:cation symporters – GPH) and amino acids or peptides (proton-dependent oligopeptide transporters – POT), but also export of drugs and other (Pao et al. 1998).

The third most abundant group of secondary transporters in *T. vaginalis* is the amino acid auxin permease (AAAP) family with 40 paralogues. These proteins responsible for amino acid import are known to have a dramatically different numbers of paralogues in different parasitic protists, which likely reflects differences in their mode of nutrition (Bouvier et al. 2004).

Nucleosides and nucleobases can be imported by the equilibrative nucleoside transporter (ENT) family (reviewed in Hyde et al. 2001) with 9 paralogues in *T. vaginalis*. The imported nucleosides and nucleobases can be intracellularly converted to nucleotides and further used in metabolism. Choline-transporter-like (CTL) proteins, responsible for import of choline, a precursor of phospholipids, were not found in *T. vaginalis*.

Other transporters found in *T. vaginalis* in multiple paralogues are e.g. cation-chloride cotransporters (CCC) of the SLC12 family (reviewed in Arroyo, Kahle, and Gamba 2013),

drug/metabolite transporter (DMT) superfamily (Jack, Yang, and H. Saier 2001), divalent Anion:Na+ Symporters (DASS) of the SLC13 family (Pajor 2014), etc.

1.3. Preaxostyla

The phylum Preaxostyla (reviewed in Vladimír Hampl 2017) was created in 2001 when a close relationship between 2 seemingly very different groups of protists, Oxymonadida and the genus *Trimastix* (now known to be paraphyletic; Zhang et al. 2015), was uncovered thanks to SSU rRNA phylogeny (Dacks et al. 2001). This relationship was later corroborated by further phylogenetic and phylogenomic studies (Vladimír Hampl et al. 2009; Zhang et al. 2015). Besides molecular phylogeny, there is also an ultrastructural synapomorphy which defines Preaxostyla: a paracrystalline lattice-like cytoskeleton element which connects two components of the excavate-like mastigont (Simpson 2003) of trimastigids and forms part of a lineage-specific cytoskeletal structure called preaxostyle in oxymonads.

Preaxostyla is nested within Metamonada, one of the deepest eukaryotic lineages with an uncertain position in the eukaryotic tree of life (Derelle et al. 2015; Brown et al. 2018). Metamonada is currently known to be composed of 3 lineages: Fornicata, Parabasalia, and Preaxostyla, where Preaxostyla is sister to a clade containing Fornicata and Parabasalia (Vladimír Hampl et al. 2005; Vladimír Hampl et al. 2009).

The inner relationships within Preaxostyla based on molecular phylogeny and ultrastructural observations are as follows (Zhang et al. 2015; Treitli et al. 2018). Preaxostyla split into 3 major monophyletic lineages: Trimastigidae and Paratrimastigidae (previously grouped in a single genus *Trimastix*) and Oxymonadida. There are 3 described species of Trimastigidae (*Trimastix marina*, *T. inaequalis*, *T. elaverinus*), 2 described species of Paratrimastigidae (*Paratrimastix pyriformis*, *P. eleionoma*), and circa 140 described species of Oxymonadida, which split into one clade consisting of the families Polymastigidae and Streblomastigidae and another, weakly supported, clade consisting of the families Pyrsonymphidae, Saccinobaculidae, Oxymonadidae, and a genus *Opisthomitus* of uncertain affinity (reviewed in Vladimír Hampl 2017).

Most insights into cell biology and metabolism of Preaxostyla that we have come from studies on *Trimastix marina* (Trimastigidae), *Paratrimastix pyriformis* (Paratrimastigidae), *Monocercomonoides exilis*, and *Streblomastix strix* (Oxymonadida). There are multiple radical differences between Trimastigidae and Paratrimastigidae on one side and Oxymonadida on the

other. What unites them, apart from the common evolutionary history and the one cytoskeletal synapomorphy discussed above, is their anaerobic lifestyle and the absence of peroxisomes.

1.3.1. Morphology and ultrastructure of Preaxostyla

Trimastigidae and Paratrimastigidae are free-living bacteriovorous flagellates inhabiting sea and freshwater oxygen-poor habitats. Their cells have a typical "excavate" morphology and ultrastructure (Zhang et al. 2015) with usually 4 flagella: 1 anterior, 2 lateral, and 1 posterior, which passes through a feeding groove on the ventral side of the cell. The posterior flagellum has two vanes, with thickened margins in Paratrimastigidae. The cytoskeleton organization of Trimastigidae and Paratrimastigidae resembles that in other "excavate" taxa with the exception of the lattice-like structure connecting the R2 microtubular root with a thin sheet formed by the I fiber. Golgi apparatus is well developed, and hydrogenosome-like MROs are present. Mitosis is open.

Oxymonadida are morphologically and ultrastructurally very diversified. However, the cellular organization of the genera *Monocercomonoides* and *Blattamonas* is understood to be close to the ancestral state of the group (Radek 1994). The cell is oval or pyriform, 5 – 15 μm long. No feeding groove is present. There are 4 flagella, 3 anterior and 1 posterior. Basal bodies of the 4 flagella are organized in two spatially separated pairs. The 2 pairs of basal bodies are connected intracellularly by a unique cytoskeletal structure, the preaxostyle consisting of two layers, one of which is homologous to the lattice-like structure in Trimastigidae and Paratrimastigidae. The other layer consists of short, tightly packed microtubules homologous to the R2 microtubular root. A subset of the preaxostylar microtubules is elongated along the anterior-posterior axis of the cell and forms the axostyle. This structure is not homologous to the axostyle of Parabasalia. The axostyle of Polymastigidae and Streblomastigidae is a stable structure, while it is contractile in Pyrsonymphidae, Saccinobaculidae, and Oxymonadidae and takes part in locomotion of the organisms. No peroxisomes, stacked Golgi apparatus, or MROs are present, with a single possible exception of *Saccinobaculus doroaxostylus* where conspicuous organelles superficially resembling mitochondria were observed (Carpenter, Waller, and Keeling 2008).

1.3.2. Symbioses of Oxymonadida

All known oxymonads are endobionts, inhabiting digestive tract of tetrapods and arthropods. Members of the genus *Monocercomonoides* are known from a broad range of hosts including

mammals, reptiles, millipedes and insects (Treitli et al. 2018). All other known oxymonads were isolated from the digestive tract of insects, most often termites and wood-eating cockroaches. The nature of symbiosis between oxymonads and their hosts is not known in most cases, but there is no indication of pathogenicity. Some members of Pyrsonymphidae, Saccinobaculidae, and Oxymonadidae are suggested to have a mutualistic relationship with their termite and cockroach hosts in a way similar to the "hypermastigid" Parabasalia: large sizes of their cells allow the protists to phagocytize and digest pieces of wood ingested by the host. The host is suggested to benefit by gaining access to products of microbial digestion of cellulose by the protists and associated bacteria (reviewed in Brune and Ohkuma 2011).

Oxymonads are often observed to have association with bacteria both extracellularly and intracellularly and even inside the nucleus. The nature of these associations is unclear. However, presence of a specialized attachment structures formed by the protists, presumably to facilitate the association with epibionts, suggests a tight, well-coordinated symbiosis at least in some cases (Leander and Keeling 2004). Extreme form of such association is observed in the genus *Streblomastix* (Streblomastigidae) where almost entire surface of the cell, consisting of elongated thin lobes of cytoplasm, is covered in epibiotic bacteria of multiple types (Noda et al. 2006; Treitli et al. 2019).

1.3.3. Glycolysis in Preaxostyla

Many anaerobically living eukaryotes use a modified glycolytic pathway which incorporates pyrophosphate (PPi)-dependent instead of ATP-dependent enzymes in certain steps specifically PPi-phosphofructokinase (PFP) instead of phosphofructokinase (PFK) in the 3rd step and pyruvate-phosphate dikinase (PPDK) instead of pyruvate kinase (PK) in the 10th and last step. This alternative form of glycolysis can produce 3 ATP molecules per 1 glucose, in comparison to only 2 in the "classical" pathway known from aerobes (reviewed in Mertens 1993).

Three transcriptomic studies published in 2006 identified multiple alternative glycolytic enzymes in members of Preaxostyla (Slamovits and Keeling 2006a; Liapounova et al. 2006; Stechmann et al. 2006). The last step of glycolysis in *P. pyriformis* and *M. exilis* can be catalyzed by both PK and PPDK and the PPDK gene was identified also in *Streblomastix strix*. The 3rd step of glycolysis in *M. exilis* is catalyzed by the alternative PFP. Other 2 alternative glycolytic enzymes were also identified in *M. exilis*: fructose-bisphosphate aldolase (FBA) class II, type B and cofactor independent phosphoglycerate mutase (iPGM). This shows the

glycolytic pathway in Preaxostyla as an evolutionary mosaic composed of enzymes of different origins, likely acquired in adaptation to the anaerobic lifestyle (Fig. 3).

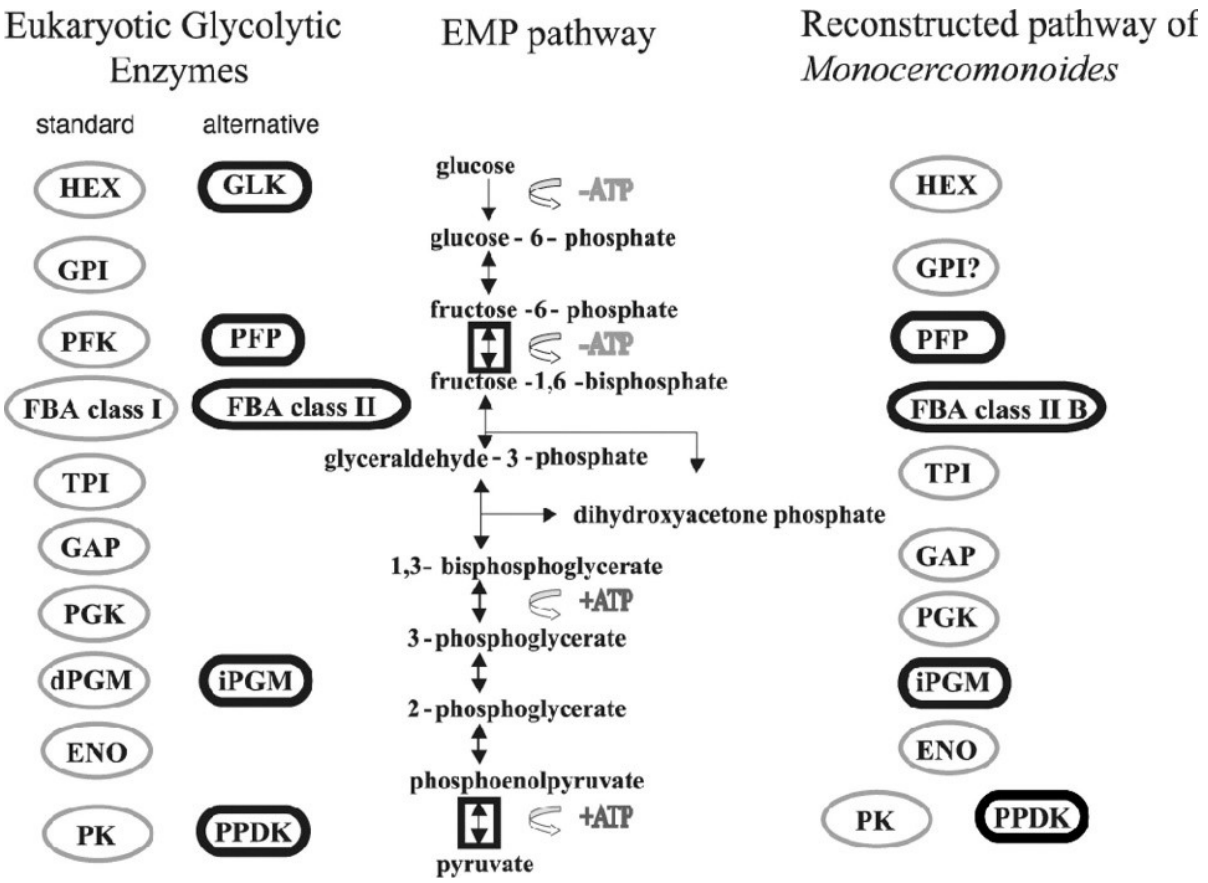


Figure 3: The reconstructed pathway of *Monocercomonoides exilis*. (Liapounova et al. 2006)

1.3.4. Proteases of *Monocercomonoides exilis*

Joel Dacks and colleagues in 2008 used transcriptomic approach to identify and characterize cathepsin B cysteine proteases in *M. exilis* (Dacks et al. 2008). The study was motivated by need to characterize this group of enzymes in a broader phylogenetic sample of eukaryotes outside traditional models like humans or trypanosomatids and by a recent discovery of 2 cathepsin B cysteine protease genes in another oxymonad, *S. strix* (Slamovits and Keeling 2006b). There were 11 different homologues of cathepsin B identified in the transcripts of *S. strix* and interestingly, no cathepsin L which are known to be often more numerous in other eukaryotes. The cathepsins are likely involved in excystation but may also have a role in lysosomal degradation and signaling.

1.3.5. Molecular genetics of Oxymonadida

Patrick Keeling and Brian Leander demonstrated in 2003 that *S. strix* uses a non-canonical genetic code where the canonical stop-codons TAA and TAG are repurposed to encode glutamine, while TGA is the only used stop-codon (Keeling and Leander 2003). The same modification of genetic code was later shown by Audrey De Koning and colleagues (De Koning et al. 2008) also in another oxymonad, an unidentified member of Polymastigidae, possibly *Monocercomonoides globus*, while *Saccinobaculus ambloaxostylus* (Saccinobaculidae) was shown to use the canonical genetic code (De Koning et al. 2008). This suggests the origin of the non-canonical genetic code in an ancestor of the lineage unifying Polymastigidae and Streblomastigidae.

Streblomastix strix was shown by Slamovits and Keeling in 2005 to have a relatively high density of spliceosomal introns (Slamovits and Keeling 2006b) compared to other studied members of Metamonada (e.g. Vaňáčová et al. 2005). Moreover, the position of more than half of the identified introns is conserved with regard to other eukaryotes, hinting at an ancient origin. This suggests that spliceosomal introns were abundant in the last common ancestor of Metamonada and were subsequently dramatically reduced in lineages leading to *T. vaginalis* and *G. intestinalis*.

1.3.6. MRO of *Paratrimastix pyriformis*

Potential mitochondrial organelles in *Paratrimastix pyriformis* (prev. *Trimastix pyriformis, T. convexa*) were reported based on ultrastructural studies (Brugerolle and Patterson 1997; O'Kelly, Farmer, and Nerad 1999). They are 0.5 to 1 μm in diameter, spherical or dumbbell-shaped, surrounded by a double membrane, and often associated with endoplasmic reticulum. However, the inability to culture *P. pyriformis* (or any other Preaxostyla) axenically precluded biochemical studies which would discern the metabolic roles of these organelles.

A transcriptomic survey by Vladimír Hampl and colleagues in 2008 uncovered 19 possibly mitochondrial or hydrogenosomal genes in *Paratrimastix pyriformis* (Paratrimastigidae) (Vladimír Hampl et al. 2008). Among these were energy metabolism-related components of the anaerobic extended glycolysis: PFO, [FeFe] hydrogenase (HydA), and 2 of the 3 hydrogenase maturases (HydE, HydG), as well as aconitase, a mitochondrial enzyme of the TCA cycle. Also 3 genes coding for proteins involved in mitochondrial protein import and maturation were detected: Tom40, Cpn60, and MPP subunit α. Among other potentially MRO-localized proteins

were all components of the glycine cleavage system (GCS H, L, T, P1, and P2), pyridine nucleotide transhydrogenase (PNT) and lipoyltransferase – enzymes involved in metabolism of cofactors, and 3 mitochondrial carrier (MC) family proteins likely responsible for translocation of metabolites across the MRO inner membrane.

The set of potentially MRO-localized proteins of *P. pyriformis* was broadened in a subsequent transcriptomic study by Zuzana Zubáčová and colleagues (including author of this thesis) published in 2013 (Zubáčová et al. 2013). The 3rd hydrogenase maturase was found (HydF) as well as new components of the protein import and maturation machinery (Sam50, Tim17, and Pam18) and new members of mitochondrial carrier (MC) family and enzymes potentially involved in amino acid metabolism (SHMT, OTC). Experimental evidence was presented for the MRO localization of GCS H protein, making the glycine cleavage system the only known function of the *P. pyriformis* MRO.

2. Aims

- To generate the genomic assembly of oxymonad *Monocercomonoides exilis*, perform automatic and manual prediction and annotation of protein-coding genes. To evaluate the completeness and quality of the genomic assembly by annotating various cellular systems. To thoroughly search for genes of mitochondrial origin.
- To search for genes coding for enzymes of the arginine deiminase (ADI) pathway in a broad sample of eukaryotes *in silico*. To explore the evolutionary history of the pathway in eukaryotes by molecular-phylogenetic methods.
- To search for genes coding for components of iron-sulfur cluster assembly systems in a broad sample of Preaxostyla. To explore the evolutionary history of these systems in Preaxostyla by molecular-phylogenetic methods.
- To annotate various cellular systems of *Monocercomonoides exilis* based on the previously obtained genomic assembly. To explore how these systems might have contributed to the loss of mitochondria, or how they responded to it, by comparing the results to other studied anaerobic protists.

3. List of publications and author contribution

- Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, et al. **A Eukaryote without a Mitochondrial Organelle.** Curr Biol. Elsevier; **2016**; 26:1274–84. https://doi.org/10.1016/j.cub.2016.03.053.

  Author: Manual gene prediction and annotation (amino acid metabolism). Search for mitochondrial genes.

- Novák L, Zubáčová Z, Karnkowska A, Kolisko M, Hroudová M, Stairs CW, et al. **Arginine deiminase pathway enzymes: evolutionary history in metamonads and other eukaryotes.** BMC Evol Biol. BioMed Central; **2016**; 16:197. https://doi.org/10.1186/s12862-016-0771-4.

  Author: Preparation of *Paratrimastix pyriformis* genomic data, *in silico* searches and phylogenetic analyses, writing the manuscript.

- Vacek V, Novák LVF, Treitli SC, Táborský P, Čepička I, Kolísko M, et al. **Fe-S Cluster Assembly in Oxymonads and Related Protists.** Mol Biol Evol. **2018**; https://doi.org/10.1093/molbev/msy168.

  Author: Discovery of SUF genes in Preaxostyla. Preparation of *Paratrimastix pyriformis* genomic data, *in silico* searches.

- Karnkowska A, Treitli SC, Brzoň O, Novák L, Vacek V, Soukal P, et al. **The oxymonad genome displays canonical eukaryotic complexity in the absence of a mitochondrion.** Mol Biol Evol. **2019**; https://doi.org/10.1093/molbev/msz147.

  Author: Manual gene annotation, cellular systems reconstruction (amino acid metabolism, selenium utilization machinery, calcium homeostasis regulation).

4. Summary

4.1. Genome of *Monocercomonoides exilis* and the loss of mitochondria

A draft genome assembly of *Monocercomonoides exilis* was produced using the 454 whole-genome shotgun sequencing technology (Karnkowska et al. 2016). The 75 Mb genome was assembled into 2,095 scaffolds at circa 35× coverage. A set of 16,629 protein-coding genes was predicted, with circa 1.9 introns per gene. Completeness of the genome assembly was estimated by multiple ways. An independently acquired transcriptome was mapped to the genome with 96.9% success. The Core Eukaryotic Genes Mapping Approach (CEGMA) (Parra, Bradnam, and Korf 2007) estimated the genome completeness at 90% level. Also, 77 out of 78 families of eukaryotic ribosomal proteins (Lecompte et al. 2002) were identified in the genome. These analyses show that the genome assembly is nearly complete.

4.1.1. The search for mitochondria

No trace of mitochondrial DNA was found in the assembly. Genes of mitochondrial origin were searched in the assembly by BLAST and HMMER tools using manually selected queries, as well as the MitoMiner database reference set (Smith, Blackshaw, and Robinson 2012) and no were found. TargetP, MitoProt, TMHMM (Emanuelsson et al. 2007), and other tools were used to search for candidate genes with sequence motives characteristic for mitochondrial proteins. The resulting candidates were manually inspected, and no were found to be genuinely mitochondrial proteins.

Similarly to mitochondria, no Golgi apparatus has ever been reported in *Monocercomonoides*. However, numerous Golgi-associated proteins were identified in the genome assembly. This suggests that the Golgi-specific functions are conserved in *Monocercomonoides*, only in different ultrastructural architecture to other eukaryotes. The same, however, cannot be said about mitochondria, as no mitochondria-associated proteins were found.

Taken together, these thorough and exhaustive analyses of the *Monocercomonoides* genome indicate that the mitochondrion, i.e. the subcellular compartment as well as its numerous functions, was indeed completely lost in *Monocercomonoides*. This makes *Monocercomonoides exilis* the first known eukaryotic organism completely devoid of mitochondria. The absence of mitochondria most likely represents a secondary loss, as *Paratrimastix pyriformis*, a close relative of *Monocercomonoides* (Dacks et al. 2001), was shown to have a reduced mitochondrial organelle resembling hydrogenosomes (Vladimír Hampl et al. 2008; Zubáčová et al. 2013).

4.1.2. Life without mitochondria

Extended glycolysis of *M. exilis* was shown to be analogous to that in mitosome-bearing protists *Giardia intestinalis* or *Entamoeba histolytica*, consisting of cytosolically localized PFO, ACS, and HydA. No hydrogenase maturases were found, just like in the protists with mitosomes. Besides the sugar-catabolizing extended glycolysis, *M. exilis* is likely able to generate ATP also from amino acids, thanks to a complete ADI pathway.

No genes of the mitochondrial ISC system responsible for iron-sulfur clusters assembly were found in *M. exilis*. Instead, components of the prokaryotic SUF system were identified: SufB, SufC, and fused SufD, SufS and SufU. Presence of the SUF system in *M. exilis* was first identified by the author of this thesis. Presence of spliceosomal introns together with fluorescence *in situ* hybridization (FISH) experiments and a presence of related genes in *Paratrimastix pyriformis* (data discussed in Supplement I) confirm that these SUF components are indeed encoded by the protists genome and do not represent a prokaryotic contamination. Phylogenetic analyses of the genes do not support their close relationship to their plastidal homologues in phototrophic eukaryotes, nor the genes in *Pygsuia biforma* or *Blastocystis hominis* and rather suggest their origin in an independent LGT event from and unknown group of Bacteria.

These observations can be explained by a following evolutionary scenario: A common ancestor of *Monocercomonoides exilis* and *Paratrimastix pyriformis* possessed a type of mitochondrion or MRO with the functional ISC system and in addition to this, gained the SUF system by lateral gene transfer from bacteria. The SUF system genes were incorporated into the eukaryotic nuclear genome and their protein products were recruited for FeS cluster production in the cytosol. This resulted in two redundant sources of FeS clusters in the cell, one in the mitochondrion and other in the cytosol, easing the selective pressure for retention of the mitochondrion. The lineage leading to *M. exilis* experienced no other pressure for mitochondrion retention (possibly thanks to its endobiotic lifestyle characteristic for oxymonads) and so the mitochondrion was gradually reduced and eventually lost completely (Fig. 4).

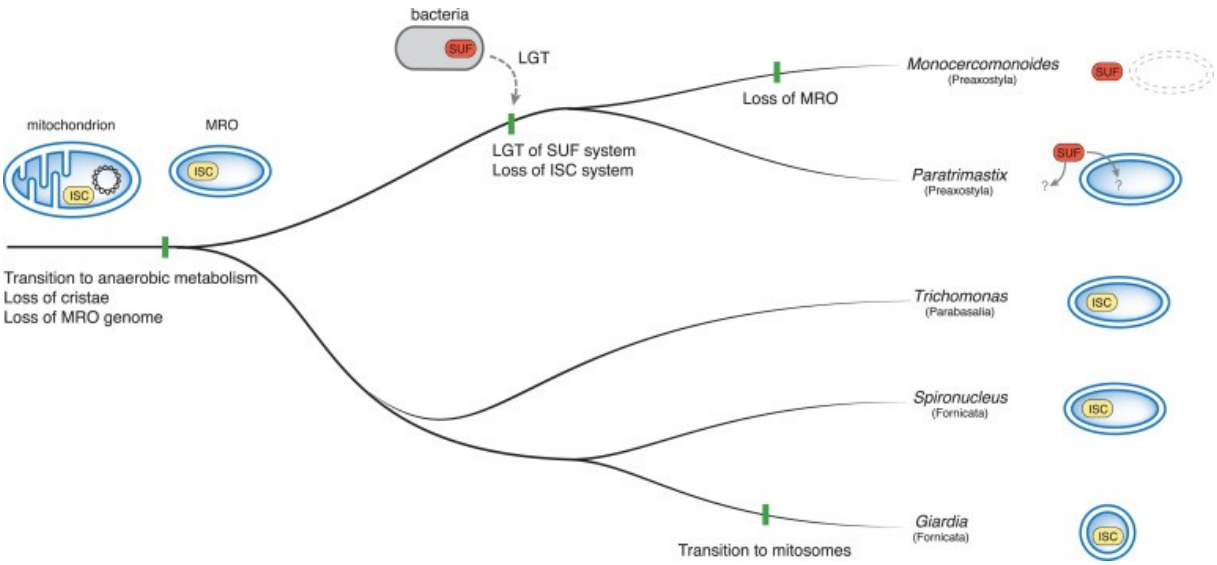


Figure 4: Reductive evolution of mitochondria in metamonads. (Karnkowska et al. 2016)

4.1.3. Deep dive into the *Monocercomonoides exilis* genome

A large number of various cellular systems in *M. exilis* was later annotated and reconstructed based on the genome assembly in order to better understand the evolutionary history and cell biology of the organism (Karnkowska et al. 2019).

4.1.3.1. Genome maintenance, cytoskeleton, and endomembrane system

All or nearly all conserved components of the various machineries responsible for genome maintenance and gene expression were found in *M. exilis*. These include e.g. histones, origin recognition complex, replication machinery, topoisomerase, various DNA repair pathways, general transcription factors, translation initiation factors, etc. All of these systems are comparable in complexity to mitochondriate organisms and some of them are more complex than in other studied metamonads.

Complement of cytoskeleton-related genes in *M. exilis* is similar to that in other studied metamonads, including some well-known metamonad-specific losses like myosin-based motility and the dynactin complex. On the other hand, it has a more complete set of actin-related proteins than other metamonads.

Endomembrane system seems to be similar to other metamonads based on the repertoire of associated genes identified in the genome. Interestingly, among the endomembrane system-related proteins identified in *M. exilis* are Vps13 and Vps39, which are involved in the

interaction of mitochondria with other membranous structures in cells of other organisms. Their presence in the amitochondriate *M. exilis* poses a unique opportunity to study their function.

Monocercomonoides exilis contains 3 genes which are involved in mitochondrial dynamics in other organisms, including 2 dynamin-related proteins (DRP), which it shares with other metamonads, and surprisingly also MSTO1 protein, not present in either *T. vaginalis* or *G. intestinalis*. These proteins had other, non-mitochondrion related, functions, reported in other organisms.

4.1.3.2. Metabolism

An expanded repertoire of genes involved in oxygen stress response was identified in *M. exilis*, including rubrerythrin, nitroreductase, flavodiiron protein, catalase, and hemerythrin. Many of these genes were acquired by LGT from bacteria. This enlarged set of oxygen stress response-related genes may reflect the loss of mitochondrion, or it may be a more general response to the lifestyle of the organism, comparable to other anaerobic protists.

Based on the hypothetical metabolic map of the amino acid metabolism reconstructed by the author of this thesis (Fig. 5), *M. exilis* may be able to synthesize at least alanine, serine, cysteine, and selenocysteinyl-tRNA, and possibly glutamate and glutamine as well. Amino acids can be also used to generate ATP, either via the ADI pathway from arginine, or via catabolism of tryptophan, cysteine, serine, threonine, and methionine. The activity of tryptophanase, a rare enzyme among eukaryotes, can generate indole, which may play a role in the interaction of *M. exilis* with its host. Neither glycine cleavage system (GCS), nor SHMT, known to be localized in MROs of some anaerobic protists, are present. As first suggested by the author of this thesis, the loss of the GCS (this system is present in the mitochondriate relative of *M. exilis* – *P. pyriformis*; Zubáčová et al. 2013) may be intimately connected to the loss of the mitochondrion itself, representing either a prerequisite for the organelle loss, or a result of it. The reconstructed amino acid metabolism of *M. exilis* is more complex than in mitosome-bearing protists like *G. intestinalis*, *E. histolytica*, or *C. parvum*, yet less complex than in the hydrogenosome-bearing *T. vaginalis*.

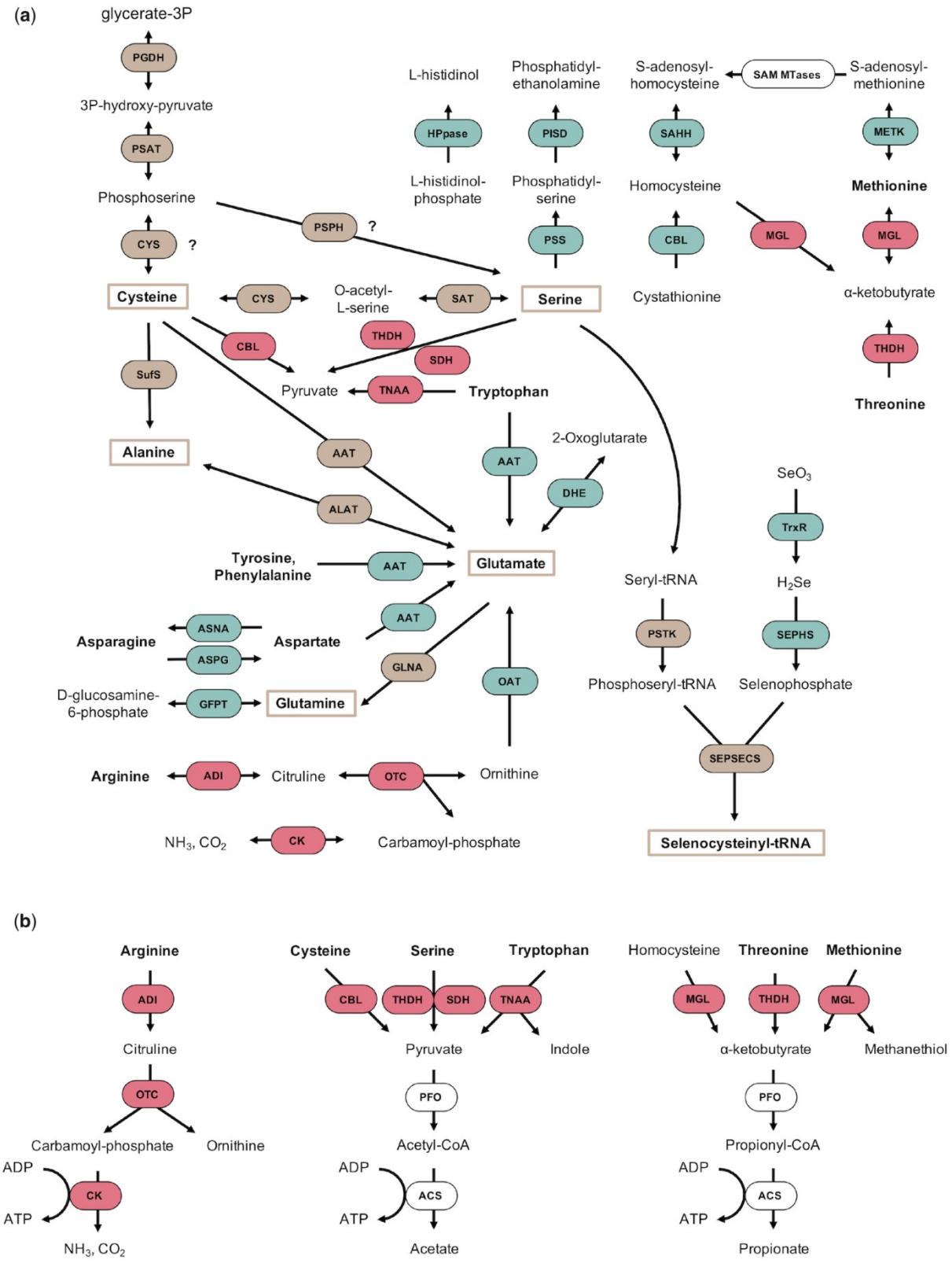


Figure 5: Putative amino acid related biochemical pathways in *Monocercomonoides exilis*. (a) Amino acid metabolism. (b) Reactions putatively involved in ATP production by amino acids catabolism. Brown color indicates enzymes possibly involved in amino acid biosynthesis pathways. Red color indicates enzymes possibly involved in ATP production. (Karnkowska et al. 2019)

Fatty acid metabolism is present but reduced in a way similar to *G. intestinalis* and *E. histolytica*: shorter chain fatty acid biosynthesis and fatty acid degradation pathways are missing. Only the non-oxidative phase of the pentose-phosphate pathway (PPP) is present. This has also been reported in *E. histolytica*. NADPH, usually synthesized by the oxidative phase of PPP, may be synthesized by NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPN), an LGT candidate in *M. exilis* genome.

Purines or pyrimidines cannot be *de novo* synthesized, nor catabolized by *M. exilis*, and the organism relies on salvage of external nucleotides from the environment of the host. Unlike other metamonads, *M. exilis* does not rely on the salvage of deoxyribonucleotides, because it can convert ribonucleotides to deoxyribonucleotides thanks to ribonucleoside-triphosphate reductase.

Autophagy machinery is reduced similarly as in *T. vaginalis*. Specifically, the autophagy related 1 complex is almost entirely missing. However, *M. exilis* is likely capable of generating autophagosomes. No mitophagy-specific components were identified. They were likely lost with the loss of the mitochondrion.

In summary, practically none of the studied cellular systems are significantly altered relative to mitochondriate eukaryotes. Many of the systems are either similar in complexity, or even more complex than in other metamonads, especially *G. intestinalis*. The only significant changes connected with the loss of mitochondria seem to be the replacement of the mitochondrial ISC system for FeS cluster assembly with the bacterial SUF system, the loss of the mitochondrial glycine cleavage system, and the loss of mitophagy-specific proteins.

4.2. Distribution and evolutionary history of the ADI pathway in Eukaryota

Complete ADI pathway, which was previously known among eukaryotes only from members of Parabasalia and Fornicata, was discovered in *Monocercomonoides exilis* (Preaxostyla). This implies a question whether the pathway was present already in a common ancestor of all Metamonada, and whether it can be found also elsewhere in eukaryotes. Publicly available gene databases of NCBI, JGI, and Marine Microbial Eukaryote Transcriptome Sequencing Project, as well as unpublished genomic and transcriptomic data from multiple laboratories were searched for homologues of the ADI pathway genes in order to sample as broad eukaryotic diversity as possible (Novák et al. 2016).

The individual genes composing the pathway (ADI, OTC, and CK) were found in all major lineages of eukaryotes and the complete set, suggesting a functional pathway, was identified not only in most studied metamonads, but also in *Harpagon schusteri* (Heterolobosea, Discoba), *Mastigamoeba balamuthi* (Archamoebea, Amoebozoa), *Pygsuia biforma* (Breviatea, Obazoa), and *Chlorella variabilis* and *Coccomyxa subellipsoidea* (Chlorophyta, Archaeplastida) (Fig. 6). Interestingly, all these organisms are either anaerobes, microaerophiles, or aerobes tolerant to prolonged periods of anaerobic conditions (Kessler 1974; Chávez, Balamuth, and Gong 1986; Cavalier-Smith 2003; Pánek et al. 2012; Stairs et al. 2014; Atteia et al. 2013; Lepère et al. 2016).

4.2.1. Phylogenetic analyses of the ADI pathway genes in Eukaryota

To investigate the evolutionary histories of the individual enzymes in eukaryotes, we preformed single-gene phylogenetic analyses of all 3 using RAxML (Stamatakis 2014) and IQ-TREE (Nguyen et al. 2015) tools. The eukaryotic sequences of ADI formed a single, moderately supported clade. The eukaryotic sequences of OTC split into 3 well-supported clades branching on different positions between prokaryotic sequences. One of these clades contains sequences from all the organisms supposedly harboring the complete pathway excluding the 2 algae. A sequence originating from the archaeon *Lokiarchaeum* sp., which is known to be one of the closest prokaryotic relatives of eukaryotes (Spang et al. 2015), is also included in this clade on an unsupported position among the eukaryotes. The eukaryotic sequences of CK split into multiple clades. One of these clades includes sequences from Parabasalia, Fornicata, *Mastigamoeba balamuthi*, and *Pygsuia biforma*, while the sequences from Preaxostyla and *Chlorella variabilis* plus *Coccomyxa subellipsoidea* branch separately on different positions among prokaryotes.

In summary, the sequences originating from the organisms with the complete ADI pathway tend to branch together in single-gene phylogenetic trees, which suggests that they had a common evolutionary history, possibly being inherited vertically from a common ancestor, which, in case of these particular lineages, would very likely be also the common ancestor of all eukaryotes.

In other words, the complete ADI pathway might have been present already in the LECA, the last eukaryotic common ancestor. To test this, we performed series of phylogenetic analyses on the 3 genes in concatenation. As predicted by the hypothesis of vertical inheritance from a common ancestor, the phylogenetic tree constructed from a concatenation of the genes, which

1) come from organisms with a complete set and 2) branch together in the single-gene trees, showed a well-supported eukaryotic clade branching inside Archaea.

These results show that the ADI pathway is clearly ancestral in Metamonada and possibly also in Eukaryota. History of the individual enzymes in some organisms is, however, less clear. The CK genes of Preaxostyla are likely a result of LGT from prokaryotes, which replaced the original metamonad CK. The OTC and CK from *Chl. variabilis* and *C. subellipsoidea* probably have different origin than the genes in metamonads, *H. schusteri*, *P. biforma*, and *M. balamuthi*. These conclusions are also supported by the approximately unbiased (AU) (Shimodaira 2002) and expected likelihood weight (ELW) (Strimmer and Rambaut 2002) statistical tests of topology performed on the datasets.

4.2.2. Subcellular localization of the ADI pathway in Preaxostyla

Subcellular localization of the ADI pathway proteins in Preaxostyla was investigated by heterologous expression experiments. HA-tagged OTC and CK genes from *Paratrimastix pyriformis* and ADI, OTC, and CK genes from *Monocercomonoides exilis* were expressed in *T. vaginalis* and the localization of their products in the cell was determined using fluorescence microscopy. All the proteins were shown to have a cytosolic localization, meaning they do not have any signal sequence recognizable by the *T. vaginalis* hydrogenosomal protein import machinery.

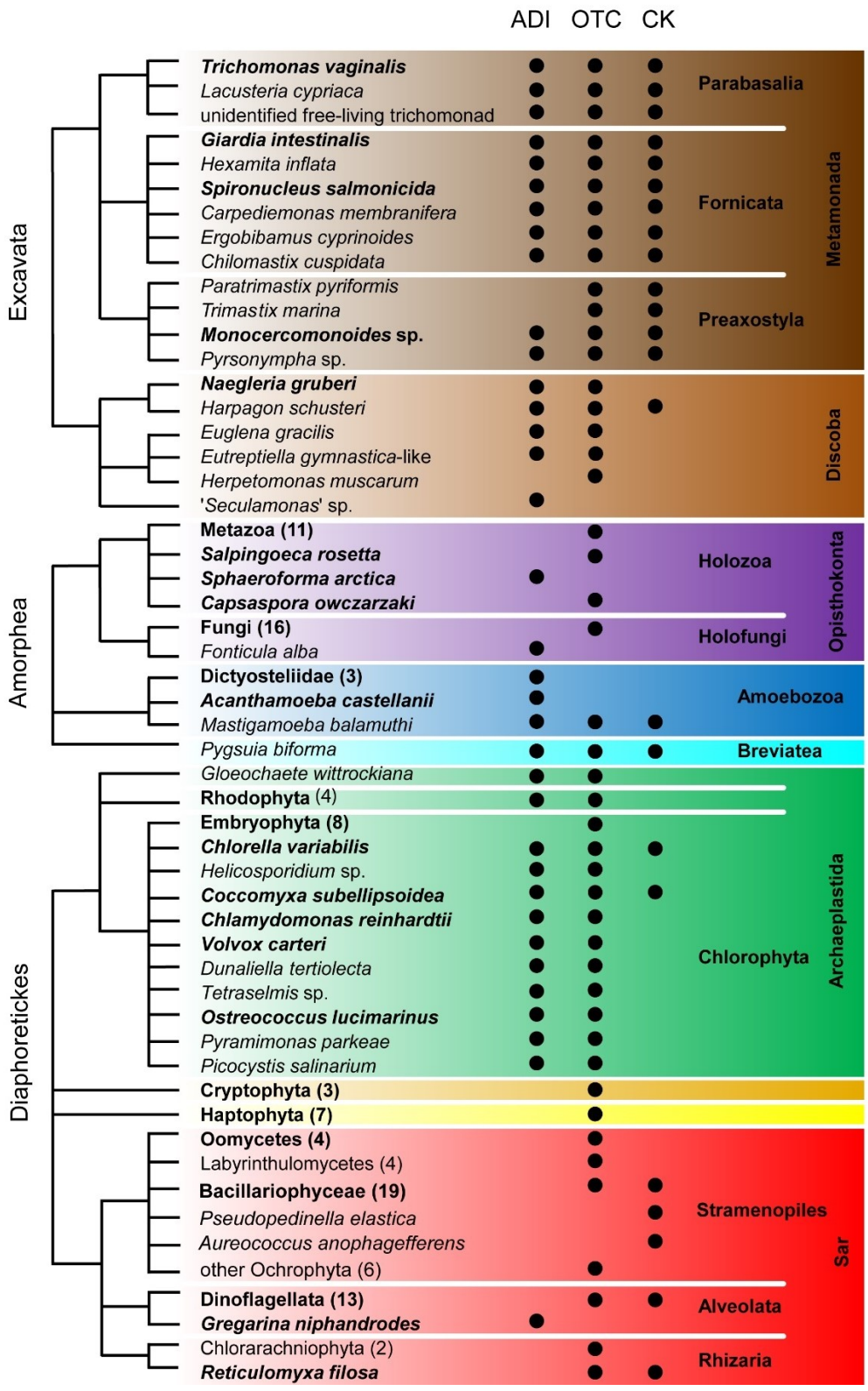


Figure 6: Distribution of enzymes of the arginine deiminase pathway across eukaryotic diversity. Arginine deiminase (ADI), ornithine transcarbamylase (OTC), carbamate kinase (CK). Taxon names in boldface indicate lineages containing at least one representative with a sequenced nuclear genome. Numbers in brackets indicate number of sequences from the given taxon included in our analyses. (Novák et al. 2016)

4.3. Distribution and evolutionary history of the SUF pathway in Preaxostyla

Lateral gene transfer of the genes coding for the SUF pathway from prokaryotes into an ancestor of *Monocercomonoides exilis* was a crucial moment for the eventual loss of mitochondria. It is therefore important to understand the evolutionary history of this pathway in a broader phylogenetic context. Genomic and transcriptomic data from 10 species of Preaxostyla: 7 species of oxymonads, 2 species of trimastigids, and 1 paratrimastigid were searched for components of 4 systems responsible for FeS cluster assembly in eukaryotes: CIA, ISC, SUF, and NIF (Vacek et al. 2018).

4.3.1. FeS cluster assembly systems in Preaxostyla

No components of the mitochondrial ISC pathway, nor the NIF pathway known from Archamoebea (Nývltová et al. 2013), were found in any of the Preaxostyla. On the other hand, there were identified components of two other iron-sulfur cluster assembly systems in all the organisms. The cytosolic CIA pathway is a known in all studied eukaryotes, but it is unable to synthesize FeS clusters *de novo* on its own and needs cooperation of another pathway: ISC, and presumably also NIF or SUF. Apparently, the SUF system plays this role in all the studied members of Preaxostyla.

The SUF system in Preaxostyla consists of 3 genes: SufB, SufC, and often fused SufD + SufS + SufU. It is possible that the SufDSU fusion is present in all Preaxostyla, especially because it was identified in all 4 organisms for which a genomic assembly is available (*Paratrimastix pyriformis*, *Blattamonas nauphoetae*, *Monocercomonoides exilis*, and *Streblomastix strix*), while the non-fused genes come from, inherently incomplete, transcriptomic assemblies (Fig. 7).

4.3.2. Evolutionary history of the SUF system in Preaxostyla

Phylogenetic analysis was performed on a concatenated dataset of the SufB, SufC, SufD, and SufS genes using IQ-TREE (Nguyen et al. 2015). The resulting tree has shown a single, well-supported clade of all the sequences originating from Preaxostyla. This strongly suggests an origin of the SUF system in a common ancestor of the entire Preaxostyla. The phylogenetic analysis was unable to pinpoint a specific prokaryotic lineage from which the genes were transferred. However, the composition of the Preaxostyla SUF system, as well as the conspicuous fusion, hint at a possible origin in Firmicutes, Thermotogae, Spirochaetes, Proteobacteria, or Chloroflexi groups of Bacteria.

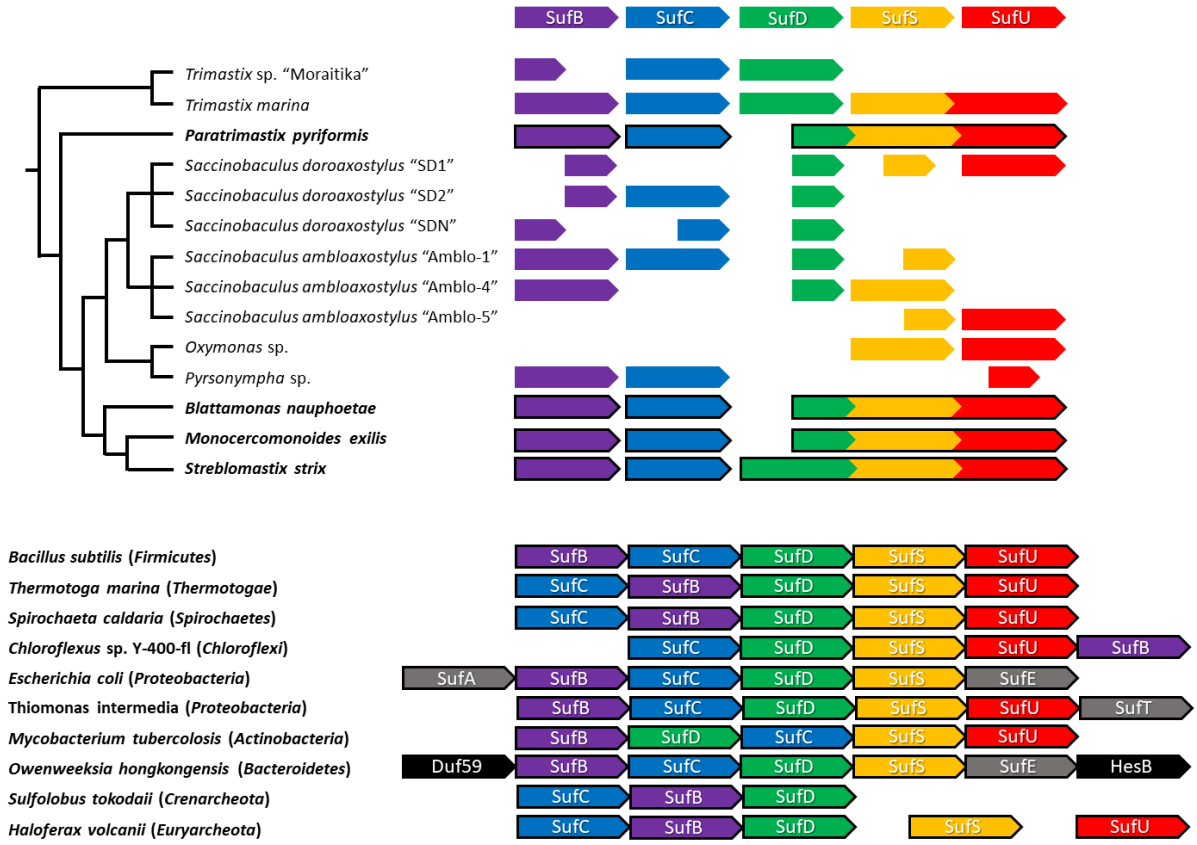


Figure 7: Inventory of SUF proteins in Preaxostyla. The scheme shows SUF genes/transcripts identified in the members of Preaxostyla. For organisms in bold, genomic data were investigated, in others transcriptomic or single cell transcriptomic data sets were used. Completeness of a gene/transcript is indicated by the length of the arrow. Gene fusions are marked by fused arrows. At the bottom are given schematics of typical SUF gene operons in selected representatives of prokaryotic groups. (Vacek et al. 2018)

4.4. Conclusions

The original work presented in this thesis has resulted in publishing a good quality genome assembly of *Monocercomonoides exilis,* which was used to provide a compelling evidence of a complete loss of mitochondria in *M. exilis* and to annotate a large number of cellular systems in this organism. Following studies investigated distribution and evolutionary histories of 2 pathways which were crucial for evolution of *M. exilis* and its adaptation to the amitochondriate, anaerobic, endobiotic lifestyle: the ADI pathway and the SUF system. This work has brought *M. exilis* (and whole Preaxostyla) into spotlight of protistological research and emphasized the importance of understudied protistan lineages for our understanding of the evolutionary history of eukaryotes. Further research is needed in order to understand *M. exilis* adaptations in a broader phylogenetic context (preliminary results discussed in supplement I) and to establish new experimental procedures for study of this organism and its relatives.

5. List of abbreviations

1P5CD	1-pyrroline-5-carboxylate dehydrogenase
AAAP	Transmembrane amino acid transporter
AAT	Aspartate aminotransferase
ABC	ATP-binding cassette transporter
AcnB	Aconitase
ACS	Acetyl-CoA synthetase
ACY	Aminoacylase
ADI	Arginine deiminase
ADK	Adenylate kinase
ADP	Adenosine diphosphate
ADSL	Adenylosuccinate lyase
ADSS	Adenylosuccinate synthase
ALAT	Alanine transaminase
AS	Asparagine synthase
ASCT	Acetate:succinate CoA transferase
ASNA	Asparagine synthetase
ASPG	L-asparaginase
Atm1	ABC transporter Atm1
ATP	Adenosine triphosphate
AU	Approximately unbiased test
BCAT	Branched-chain amino acid aminotransferase
BLAST	Basic local alignment search tool
BolA	DNA-binding transcriptional regulator BolA
CaCA	Ca^{2+}:Cation Antiporter
CBL	Cystathionine beta-lyase

CCC	Cation-chloride cotransporter
CDF	Cation diffusion facilitator
cDNA	Complementary DNA
CEGMA	Core eukaryotic genes mapping approach
CIA	Cytosolic iron-sulfur cluster assembly
CK	Carbamate kinase
CLO	*Carpediemonas*-like organism
CoA	Coenzyme A
Cpn10	10 kDa chaperonin
Cpn60	60 kDa chaperonin
CTL	Choline-transporter-like protein
CYS	Cysteine synthase
CysJ	NADPH-dependent sulfite reductase
DASS	Divalent Anion:Na^+ Symporter
DCAM	Adenosylmethionine decarboxylase
DCDA	Diaminopimelate decarboxylase
DCOR	Ornithine decarboxylase
DHE	Glutamate dehydrogenase
DHFR	Dihydrofolate reductase
DNA	Deoxyribonucleic acid
DRP	Dynamin-related protein
EF	Elongation factor
EMP	Embden–Meyerhof–Parnas pathway
ENO	Enolase
ENT	Equilibrative nucleoside transporter
FBA	Fructose-bisphosphate aldolase

Fd	Ferredoxin
FDP	Flavodiiron protein
FeFe	Iron-iron
FeS	Iron-sulfur
FISH	Fluorescence *in situ* hybridization
Fld-Fd	Flavodoxin-ferredoxin fusion protein
FolC	Folylpolyglutamate synthase
FolD	Bifunctional methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase
FTCD	Glutamate formimidoyltransferase
Fxn	Frataxin
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GAPN	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase
GCAT	Glycine C-acetyltransferase
GCS	Glycine cleavage system
GCSH	Glycine cleavage system protein H
GCSL	Glycine cleavage system protein L
GCSP1	Glycine cleavage system protein P1
GCSP2	Glycine cleavage system protein P2
GCST	Glycine cleavage system protein T
gDNA	Genomic DNA
GFPT	Glutamine-fructose-6-phosphate aminotransferase
GGH	Gamma-glutamyl hydrolase
GLNA	Glutamine synthetase
GPH	Glycoside-pentoside-hexuronide:cation symporter
GPI	Glucose-6-phosphate isomerase

GrpE	GroP-like gene E
Grx5	Monothiol glutaredoxin-5
HA	Human influenza hemagglutinin
HMMER	Biosequence analysis using profile hidden Markov models
HMP	Hydrogenosomal membrane protein
HSP	Heat shock protein
HUTI	Imidazolonepropionase
HXK	Hexokinase
HydA	[FeFe] hydrogenase
HydE	Hydrogenase maturase HydE
HydF	Hydrogenase maturase HydF
HydG	Hydrogenase maturase HydG
Ind1	Iron-sulfur protein IND1
iPGM	2,3-bisphosphoglycerate independent phosphoglycerate mutase
ISC	Iron sulfur cluster
ISF	Iron-sulfur flavoproteins
Jac1	J-type co-chaperone JAC1
JGI	Joint Genome Institute
LAM	Lysine 2,3-aminomutase
LC-MS	Liquid chromatography–mass spectrometry
LECA	Last eukaryotic common ancestor
LGT	Lateral gene transfer
LipB	Lipoyltransferase
MATE	Multi antimicrobial extrusion
MC	Membrane carrier
ME	Malic enzyme

MetH	Methionine synthase
METK	S-adenosylmethionine synthetase
MFS	Major facilitator superfamily transporter
MGL	Methionine gamma-lyase
MPPa	Mitochondrial-processing peptidase alpha subunit
MPPb	Mitochondrial-processing peptidase beta subunit
MRO	Mitochondrion-related organelles
mRNA	Messenger RNA
MRP	Multidrug resistance-associated protein
MSTO1	Misato homolog 1
MTHFD1	Formate-tetrahydrofolate ligase
MTHFR	Methylenetetrahydrofolate reductase
mtHSP70	Heat shock protein 70, mitochondrial
NAD	Nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NFU	Iron-sulfur cluster scaffold-like protein NFU
NIF	Nitrogen fixation
NuoE	NADH-quinone oxidoreductase subunit E
NuoF	NADH-quinone oxidoreductase subunit F
NuoG	NADH-quinone oxidoreductase subunit G
OAT	Ornithine aminotransferase
OCD	Ornithine cyclodeaminase
OsmC	Osmotically inducible protein C

OTC	Ornithine transcarbamylase
P5CR	Pyrroline-5-carboxylate reductase
PAM	Presequence translocase-associated motor
Pam16	Presequence translocase-associated motor 16 kDa subunit
Pam18	Presequence translocase-associated motor 18 kDa subunit
PAPS	3′-phosphoadenosine-5′-phosphosulfate
pATPase	P-type ATPase
PCR	Polymerase chain reaction
PEPCK	Phosphoenolpyruvate carboxykinase
PepD	Aminoacyl-histidine dipeptidase
PFK	Phosphofructokinase
PFO	Pyruvate:ferredoxin oxidoreductase
PFP	Pyrophosphate:fructose 6-phosphate phosphotransferase
PGDH	Phosphoglycerate dehydrogenase
PGK	Phosphoglycerate kinase
PGM	Phosphoglucomutase
PISD	Phosphatidylserine decarboxylase
PK	Pyruvate kinase
PMM	Phosphomannomutase
PotE	b(0,+)-type amino acid transporter
PPDK	Pyruvate phosphate dikinase
PRAC	Proline racemase
PSAT	Phosphoserine transaminase
PSS	Phosphatidylserine synthase
PSTK	O-phosphoseryl-tRNASec kinase
RAxML	Randomized axelerated maximum likelihood

RBR	Rubrerythrin
RFC	Reduced folate carrier
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rRNA	Ribosomal RNA
SAHH	S-adenosyl-L-homocysteine hydrolase
SAM	Sorting assembly machinery
Sam50	Sorting assembly machinery 50 kDa subunit
SAT	Serine acetyltransferase
SBP2	SECIS-binding protein 2
SCS	Succinyl-CoA synthetase
SDH	L-serine ammonia-lyase
SECIS	Selenocysteine insertion sequence
SelB	Selenocysteine-specific elongation factor
SEPHS	Selenide, water dikinase / selenophosphate synthase
SEPSECS	O-phosphoseryl-tRNA(Sec) selenium transferase
SHMT	Serine hydroxymethyltransferase
SOD	Superoxide dismutase
SP	Sugar porter
SpeB	Agmatinase
SpeE	Spermidine synthase
SPNS	Sphingolipid transporter
SSU	Small subunit
SUF	Sulphur utilization factor
SufB	Iron-sulfur cluster assembly protein SufB
SufC	Iron-sulfur cluster assembly protein SufC

SufD	Iron-sulfur cluster assembly protein SufD
SufDS	Iron-sulfur cluster assembly fusion protein SufD-SufS
SufDSU	Iron-sulfur cluster assembly fusion protein SufD-SufS-SufU
SufS	Iron-sulfur cluster assembly protein SufS
SufU	Iron-sulfur cluster assembly protein SufU
TA	Threonine aldolase
TauE	Sulfite exporter TauE/SafE family protein
TCA	Tricarboxylic acid
TDH	Threonine dehydrogenase
THDH	Threonine ammonia-lyase
THRC	Threonine synthase
Tim14	Inner membrane translocase subunit 14 kDa subunit
Tim17	Inner membrane translocase subunit 17 kDa subunit
TMHMM	Hidden Markov model for predicting transmembrane helices
TNAA	Tryptophanase
TOM	Translocase of outer mitochondrial membrane
Tom40	Translocase of outer mitochondrial membrane 40 kDa subunit
TPI	Triosephosphate isomerase
TRK	K^+ Transporter
Trx	Thioredoxin
TrxP	Thioredoxin peroxidase
TrxR	Thioredoxin reductase
TYDC	Tyrosine decarboxylase

6. References

Abrahamsen, Mitchell S, Thomas J Templeton, Shinichiro Enomoto, Juan E Abrahante, Guan Zhu, Cheryl A Lancto, Mingqi Deng, et al. 2004. "Complete Genome Sequence of the Apicomplexan, Cryptosporidium Parvum." *Science (New York, N.Y.)* 304 (5669): 441–45. https://doi.org/10.1126/science.1094786.

Adam, Rodney D., Eric W. Dahlstrom, Craig A. Martens, Daniel P. Bruno, Kent D. Barbian, Stacy M. Ricklefs, Matthew M. Hernandez, et al. 2013. "Genome Sequencing of Giardia Lamblia Genotypes A2 and B Isolates (DH and GS) and Comparative Analysis with the Genomes of Genotypes A1 and E (WB and Pig)." *Genome Biology and Evolution* 5 (12): 2498–2511. https://doi.org/10.1093/gbe/evt197.

Adam, Rodney D. 2000. "The Giardia Lamblia Genome." *International Journal for Parasitology* 30 (4): 475–84. https://doi.org/10.1016/S0020-7519(99)00191-5.

Adl, Sina M., David Bass, Christopher E. Lane, Julius Lukeš, Conrad L. Schoch, Alexey Smirnov, Sabine Agatha, et al. 2018. "Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes." *Journal of Eukaryotic Microbiology* 66 (1): jeu.12691. https://doi.org/10.1111/jeu.12691.

Akhmanova, Anna, Frank Voncken, Theo van Alen, Angela van Hoek, Brigitte Boxma, Godfried Vogels, Marten Veenhuis, and Johannes H.P. Hackstein. 1998. "A Hydrogenosome with a Genome." *Nature* 396 (6711): 527–28. https://doi.org/10.1038/25023.

Anderson, Iain J, and Brendan J Loftus. 2005. "Entamoeba Histolytica: Observations on Metabolism Based on the Genome Sequence." *Experimental Parasitology* 110 (3): 173–77. https://doi.org/10.1016/j.exppara.2005.03.010.

Arroyo, Juan Pablo, Kristopher T. Kahle, and Gerardo Gamba. 2013. "The SLC12 Family of Electroneutral Cation-Coupled Chloride Cotransporters." *Molecular Aspects of Medicine* 34 (2–3): 288–98. https://doi.org/10.1016/J.MAM.2012.05.002.

Atteia, Ariane, Robert Van Lis, Aloysius G M Tielens, and William F. Martin. 2013. "Anaerobic Energy Metabolism in Unicellular Photosynthetic Eukaryotes." *Biochimica et Biophysica Acta - Bioenergetics*. https://doi.org/10.1016/j.bbabio.2012.08.002.

Barberà, Maria José, Iñaki Ruiz-Trillo, Julia Y A Tufts, Amandine Bery, Jeffrey D Silberman,

and Andrew J Roger. 2010. “Sawyeria Marylandensis (Heterolobosea) Has a Hydrogenosome with Novel Metabolic Properties.” *Eukaryotic Cell* 9 (12): 1913–24. https://doi.org/10.1128/EC.00122-10.

Barratt, Joel L.N., Maisie Cao, Damien J. Stark, and John T. Ellis. 2015. “The Transcriptome Sequence of Dientamoeba Fragilis Offers New Biological Insights on Its Metabolism, Kinome, Degradome and Potential Mechanisms of Pathogenicity.” *Protist* 166 (4): 389–408. https://doi.org/10.1016/J.PROTIS.2015.06.002.

Benchimol, Marlene, Luiz G P de Almeida, Ana Tereza Vasconcelos, Ivone de Andrade Rosa, Maurício Reis Bogo, Luiza Wilges Kist, and Wanderley de Souza. 2017. “Draft Genome Sequence of Tritrichomonas Foetus Strain K.” *Genome Announcements* 5 (16). https://doi.org/10.1128/genomeA.00195-17.

Benchimol, Marlene, Roger Durand, and João Carlos Aquino Almeida. 1997. “A Double Membrane Surrounds the Hydrogenosomes of the Anaerobic Fungus Neocallimastix Frontalis.” *FEMS Microbiology Letters* 154 (2): 277–82. https://doi.org/10.1111/j.1574-6968.1997.tb12656.x.

Biagini, Giancarlo A., Anthony J. Hayes, Marc T.E. Suller, Carole Winters, Bland J. Finlay, and David Lloyd. 1997. “Hydrogenosomes of Metopus Contortus Physiologically Resemble Mitochondria.” *Microbiology* 143 (5): 1623–29. https://doi.org/10.1099/00221287-143-5-1623.

Biderre, C., M. Pages, G. Metenier, D. David, J. Bata, G. Prensier, and C. P. Vivares. 1994. “On Small Genomes in Eukaryotic Organisms: Molecular Karyotypes of Two Microsporidian Species (Protozoa) Parasites of Vertebrates.” *Comptes Rendus de l’Academie Des Sciences - Serie III* 317 (5): 399–404. http://www.ncbi.nlm.nih.gov/pubmed/7994619.

Bouvier, León A., Ariel M. Silber, Camila Galvão Lopes, Gaspar E. Canepa, Mariana R. Miranda, Renata R. Tonelli, Walter Colli, Maria Júlia M. Alves, and Claudio A. Pereira. 2004. “Post Genomic Analysis of Permeases from the Amino Acid/Auxin Family in Protozoan Parasites.” *Biochemical and Biophysical Research Communications* 321 (3): 547–56. https://doi.org/10.1016/J.BBRC.2004.07.002.

Brown, Matthew W, Aaron A Heiss, Ryoma Kamikawa, Yuji Inagaki, Akinori Yabuki, Alexander K Tice, Takashi Shiratori, et al. 2018. “Phylogenomics Places Orphan Protistan

Lineages in a Novel Eukaryotic Super-Group." *Genome Biology and Evolution* 10 (2): 427. https://doi.org/10.1093/GBE/EVY014.

Brown, Matthew W, Susan C Sharpe, Jeffrey D Silberman, Aaron A Heiss, B Franz Lang, Alastair G B Simpson, and Andrew J Roger. 2013. "Phylogenomics Demonstrates That Breviate Flagellates Are Related to Opisthokonts and Apusomonads." *Proceedings. Biological Sciences / The Royal Society* 280 (1769): 20131755. https://doi.org/10.1098/rspb.2013.1755.

Brugerolle, Guy, and David Patterson. 1997. "Ultrastructure of Trimastix Convexa Hollande, an Amitochondriate Anaerobic Flagellate with a Previously Undescribed Organization." *European Journal of Protistology* 33 (2): 121–30. https://doi.org/10.1016/S0932-4739(97)80029-6.

Brune, Andreas, and Moriya Ohkuma. 2011. "Role of the Termite Gut Microbiota in Symbiotic Digestion." In *Biology of Termites: A Modern Synthesis*, 439–75. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-3977-4_16.

Burki, Fabien, Nicolas Corradi, Roberto Sierra, Jan Pawlowski, Gary R. Meyer, Cathryn L. Abbott, and Patrick J. Keeling. 2013. "Phylogenomics of the Intracellular Parasite Mikrocytos Mackini Reveals Evidence for a Mitosome in Rhizaria." *Current Biology* 23 (16): 1541–47. https://doi.org/10.1016/j.cub.2013.06.033.

Carlton, Jane M., Yves Van De Peer, Robert P. Hirt, Joana C. Silva, Arthur L. Delcher, Michael Schatz, Qi Zhao, et al. 2007. "Draft Genome Sequence of the Sexually Transmitted Pathogen Trichomonas Vaginalis." *Science (New York, N.Y.)* 315 (5809): 207–12. https://doi.org/10.1126/science.1138294.

Carpenter, Kevin J, Ross F Waller, and Patrick J Keeling. 2008. "Surface Morphology of Saccinobaculus (Oxymonadida): Implications for Character Evolution and Function in Oxymonads." *Protist* 159 (2): 209–21. https://doi.org/10.1016/j.protis.2007.09.002.

Cavalier-Smith, Thomas. 1983. "Endocytobiology II." In *"Intracellular Space as Oligogenetic Ecosystem,"* 265–79. de Gruyter Berlin.

———. 1987. "Eukaryotes with No Mitochondria." *Nature* 326 (6111): 332–33. https://doi.org/10.1038/326332a0.

———. 1991. "Archamoebae: The Ancestral Eukaryotes?" *BioSystems* 25 (1–2): 25–38.

https://doi.org/10.1016/0303-2647(91)90010-I.

———. 1998. “A Revised Six-Kingdom System of Life.” *Biological Reviews of the Cambridge Philosophical Society* 73 (3): 203–66. http://www.ncbi.nlm.nih.gov/pubmed/9809012.

———. 2003. “The Excavate Protozoan Phyla Metamonada Grassé Emend. (Anaeromonadea, Parabasalia, Carpediemonas, Eopharyngia) and Loukozoa Emend. (Jakobea, Malawimonas): Their Evolutionary Affinities and New Higher Taxa.” *International Journal of Systematic and Evolutionary Microbiology* 53 (6): 1741–58. https://doi.org/10.1099/ijs.0.02548-0.

Čerkasov, J, A Čerkasovová, J Kulda, and D Vilhelmová. 1978. “Respiration of Hydrogenosomes of Tritrichomonas Foetus. I. ADP-Dependent Oxidation of Malate and Pyruvate.” *The Journal of Biological Chemistry* 253 (4): 1207–14. http://www.ncbi.nlm.nih.gov/pubmed/624725.

Chávez, L A, W Balamuth, and T Gong. 1986. “A Light and Electron Microscopical Study of a New, Polymorphic Free-Living Amoeba, Phreatamoeba Balamuthi n. g., n. Sp.” *The Journal of Protozoology* 33 (3): 397–404. http://www.ncbi.nlm.nih.gov/pubmed/3746722.

Cheissin, E M. 1965. “Ultrastructure of Lamblia Duodenalis 2. The Locomotory Apparatus, Axial Rod and Other Organelles.” *Arch. Protistenkd* 108: 8–18.

Clark, C G, and A J Roger. 1995. “Direct Evidence for Secondary Loss of Mitochondria in Entamoeba Histolytica.” *Proceedings of the National Academy of Sciences of the United States of America* 92 (14): 6518–21. https://doi.org/10.1073/pnas.92.14.6518.

Coombs, Graham H., Gareth D. Westrop, Pavel Suchan, Gabriela Puzova, Robert P. Hirt, T. Martin Embley, Jeremy C. Mottram, and Sylke Müller. 2004. “The Amitochondriate Eukaryote Trichomonas Vaginalis Contains a Divergent Thioredoxin-Linked Peroxiredoxin Antioxidant System.” *Journal of Biological Chemistry* 279 (7): 5249–56. https://doi.org/10.1074/jbc.M304359200.

Cornman, R. Scott, Yan Ping Chen, Michael C. Schatz, Craig Street, Yan Zhao, Brian Desany, Michael Egholm, et al. 2009. “Genomic Analyses of the Microsporidian Nosema Ceranae, an Emergent Pathogen of Honey Bees.” *PLoS Pathogens* 5 (6). https://doi.org/10.1371/JOURNAL.PPAT.1000466.

Dacks, Joel B., Teklu Kuru, Natalia A. Liapounova, and Lashitew Gedamu. 2008.

"Phylogenetic and Primary Sequence Characterization of Cathepsin B Cysteine Proteases from the Oxymonad Flagellate Monocercomonoides." *Journal of Eukaryotic Microbiology* 55 (1): 9–17. https://doi.org/10.1111/j.1550-7408.2007.00296.x.

Dacks, Joel B., Jeffrey D. Silberman, Alastair G. B. Simpson, Shigeharu Moriya, Toshiaki Kudo, Moriya Ohkuma, and Rosemary J. Redfield. 2001. "Oxymonads Are Closely Related to the Excavate Taxon Trimastix." *Mol. Biol. Evol.* 18 (6): 1034–44. http://mbe.oxfordjournals.org/cgi/content/abstract/18/6/1034.

Dean, Paul, Peter Major, Sirintra Nakjang, Robert P. Hirt, and T. Martin Embley. 2014. "Transport Proteins of Parasitic Protists and Their Role in Nutrient Salvage." *Frontiers in Plant Science* 5 (April): 153. https://doi.org/10.3389/fpls.2014.00153.

Derelle, Romain, Guifré Torruella, Vladimír Klimeš, Henner Brinkmann, Eunsoo Kim, Čestmír Vlček, B Franz Lang, and Marek Eliáš. 2015. "Bacterial Proteins Pinpoint a Single Eukaryotic Root." *Proceedings of the National Academy of Sciences of the United States of America* 112 (7): E693-9. https://doi.org/10.1073/pnas.1420657112.

Docampo, R, S N Moreno, and R P Mason. 1987. "Free Radical Intermediates in the Reaction of Pyruvate:Ferredoxin Oxidoreductase in Tritrichomonas Foetus Hydrogenosomes." *The Journal of Biological Chemistry* 262 (26): 12417–20. http://www.ncbi.nlm.nih.gov/pubmed/3040744.

Douce, Roland, Jacques Bourguignon, Michel Neuburger, and Fabrice Rébeillé. 2001. "The Glycine Decarboxylase System: A Fascinating Complex." *Trends in Plant Science* 6 (4): 167–76. https://doi.org/10.1016/S1360-1385(01)01892-1.

Drmota, T. 1996. "Iron-Ascorbate Cleavable Malic Enzyme from Hydrogenosomes of Trichomonas Vaginalis: Purification and Characterization." *Molecular and Biochemical Parasitology* 83 (2): 221–34. https://doi.org/10.1016/S0166-6851(96)02777-6.

Dyall, S. D., C. M. Koehler, M. G. Delgadillo-Correa, P. J. Bradley, E. Plumper, D. Leuenberger, C. W. Turck, and P. J. Johnson. 2000. "Presence of a Member of the Mitochondrial Carrier Family in Hydrogenosomes: Conservation of Membrane-Targeting Pathways between Hydrogenosomes and Mitochondria." *Molecular and Cellular Biology* 20 (7): 2488–97. https://doi.org/10.1128/mcb.20.7.2488-2497.2000.

Edlund, Thomas D., Jing Li, Govinda S. Visvesvara, Michael H. Vodkin, Gerald L. McLaughlin, and Santosh K. Katiyar. 1996. "Phylogenetic Analysis of β-Tubulin

Sequences from Amitochondrial Protozoa." *Molecular Phylogenetics and Evolution* 5 (2): 359–67. https://doi.org/10.1006/MPEV.1996.0031.

Emanuelsson, Olof, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. 2007. "Locating Proteins in the Cell Using TargetP, SignalP and Related Tools." *Nature Protocols* 2: 953–71. https://doi.org/10.1038/nprot.2007.131.

Embley, Martin, Mark van der Giezen, David S. Horner, Patricia L. Dyal, and Peter Foster. 2003. "Mitochondria and Hydrogenosomes Are Two Forms of the Same Fundamental Organelle." *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358 (1429): 191–203. https://doi.org/10.1098/rstb.2002.1190.

Fahrni, J. F., Ignacio Bolivar, Cédric Berney, Elena Nassonova, Alexey Smirnov, and Jan Pawlowski. 2003. "Phylogeny of Lobose Amoebae Based on Actin and Small-Subunit Ribosomal RNA Genes." *Molecular Biology and Evolution* 20 (11): 1881–86. https://doi.org/10.1093/molbev/msg201.

Felsenstein, J. 1978. "Cases in Which Parsimony or Compatibility Methods Will Be Positively Misleading." *Systematic Biology* 27 (4): 401–10. https://doi.org/10.1093/sysbio/27.4.401.

Flegel, T. W., and Tirasak Pasharawipas. 1995. "A Proposal for Typical Eukaryotic Meiosis in Microsporidians." *Canadian Journal of Microbiology* 41 (1): 1–11. https://doi.org/10.1139/m95-001.

Franzén, Oscar, Jon Jerlström-Hultqvist, Elsie Castro, Ellen Sherwood, Johan Ankarklev, David S. Reiner, Daniel Palm, Jan O. Andersson, Björn Andersson, and Staffan G. Svärd. 2009. "Draft Genome Sequencing of Giardia Intestinalis Assemblage B Isolate GS: Is Human Giardiasis Caused by Two Different Species?" *PLoS Pathogens* 5 (8): e1000560. https://doi.org/10.1371/journal.ppat.1000560.

Germot, Agnès, Hervé Philippe, and Hervé Le Guyader. 1996. "Presence of a Mitochondrial-Type 70-KDa Heat Shock Protein in Trichomonas Vaginalis Suggests a Very Early Mitochondrial Endosymbiosis in Eukaryotes." *Proceedings of the National Academy of Sciences of the United States of America* 93 (25): 14614–17. http://www.pnas.org/cgi/content/abstract/93/25/14614.

———. 1997. "Evidence for Loss of Mitochondria in Microsporidia from a Mitochondrial-Type HSP70 in Nosema Locustae." *Molecular and Biochemical Parasitology* 87 (2): 159–68. https://doi.org/10.1016/S0166-6851(97)00064-9.

Gill, Erin E, Sara Diaz-Triviño, Maria José Barberà, Jeffrey D Silberman, Alexandra Stechmann, Daniel Gaston, Ivica Tamas, and Andrew J Roger. 2007. "Novel Mitochondrion-Related Organelles in the Anaerobic Amoeba Mastigamoeba Balamuthi." *Molecular Microbiology* 66 (6): 1306–20. https://doi.org/10.1111/j.1365-2958.2007.05979.x.

Gillin, Frances D., and Louis S. Diamond. 1981. "Entamoeba Histolytica and Giardia Lamblia: Effects of Cysteine and Oxygen Tension on Trophozoite Attachment to Glass and Survival in Culture Media." *Experimental Parasitology* 52 (1): 9–17. https://doi.org/10.1016/0014-4894(81)90055-2.

Grinsven, Koen W. A. van, Silke Rosnowsky, Susanne W. H. van Weelden, Simone Pütz, Mark van der Giezen, William Martin, Jaap J. van Hellemond, Aloysius G. M. Tielens, and Katrin Henze. 2008. "Acetate:Succinate CoA-Transferase in the Hydrogenosomes of *Trichomonas Vaginalis*." *Journal of Biological Chemistry* 283 (3): 1411–18. https://doi.org/10.1074/jbc.M702528200.

Hamann, Emmo, Halina E Tegetmeyer, Dietmar Riedel, Sten Littmann, Soeren Ahmerkamp, Jianwei Chen, Philipp F Hach, and Marc Strous. 2017. "Syntrophic Linkage between Predatory Carpediemonas and Specific Prokaryotic Populations." *The ISME Journal* 11 (5): 1205–17. https://doi.org/10.1038/ismej.2016.197.

Hampl, Vladimír. 2017. "Preaxostyla." In *Handbook of the Protists*, 1139–74. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-28149-0_8.

Hampl, Vladimír, David S. Horner, Patricia Dyal, Jaroslav Kulda, Jaroslav Flegr, Peter G. Foster, and T. Martin Embley. 2005. "Inference of the Phylogenetic Position of Oxymonads Based on Nine Genes: Support for Metamonada and Excavata." *Molecular Biology and Evolution* 22 (12): 2508–18. https://doi.org/10.1093/molbev/msi245.

Hampl, Vladimír, Laura Hug, Jessica W Leigh, Joel B Dacks, B Franz Lang, Alastair G B Simpson, and Andrew J Roger. 2009. "Phylogenomic Analyses Support the Monophyly of Excavata and Resolve Relationships among Eukaryotic 'Supergroups'." *Proceedings of the National Academy of Sciences of the United States of America* 106 (10): 3859–64. https://doi.org/10.1073/pnas.0807880106.

Hampl, Vladimír, Jeffrey D Silberman, Alexandra Stechmann, Sara Diaz-Triviño, Patricia J Johnson, and Andrew J Roger. 2008. "Genetic Evidence for a Mitochondriate Ancestry in

the ‘Amitochondriate’ Flagellate Trimastix Pyriformis.” *PLoS ONE* 3 (1): 9. https://doi.org/10.1371/journal.pone.0001383.

Hampson, RK, LL Barron, and MS Olson. 1983. “Regulation of the Glycine Cleavage System in Isolated Rat Liver Mitochondria.” *The Journal of Biological Chemistry* 258 (5): 2993–99. https://europepmc.org/abstract/med/6402507.

Hine, PM, SM Bower, GR Meyer, N Cochennec-Laureau, and FCJ Berthe. 2001. “Ultrastructure of Mikrocytos Mackini, the Cause of Denman Island Disease in Oysters Crassostrea Spp. and Ostrea Spp. in British Columbia, Canada.” *Diseases of Aquatic Organisms* 45 (3): 215–27. https://doi.org/10.3354/dao045215.

Hinkle, G, D D Leipe, T A Nerad, and M L Sogin. 1994. “The Unusually Long Small Subunit Ribosomal RNA of Phreatamoeba Balamuthi.” *Nucleic Acids Research* 22 (3): 465–69. https://doi.org/10.1093/nar/22.3.465.

Hirt, R.P., C.J. Noel, T. Sicheritz-Ponten, J. Tachezy, and P-L. Fiori. 2007. “Trichomonas Vaginalis Surface Proteins: A View from the Genome.” *Trends in Parasitology* 23 (11): 540–47. https://doi.org/10.1016/j.pt.2007.08.020.

Horner, D. S., R. P. Hirt, S. Kilvington, D. Lloyd, and T. M. Embley. 1996. “Molecular Data Suggest an Early Acquisition of the Mitochondrion Endosymbiont.” *Proceedings of the Royal Society B: Biological Sciences* 263 (1373): 1053–59. https://doi.org/10.1098/rspb.1996.0155.

Hrdý, Ivan, Robert P Hirt, Pavel Dolezal, Lucie Bardonová, Peter G Foster, Jan Tachezy, and T Martin Embley. 2004. “Trichomonas Hydrogenosomes Contain the NADH Dehydrogenase Module of Mitochondrial Complex I.” *Nature* 432: 618–22. https://doi.org/10.1038/nature03149.

Hrdý, Ivan, Jan Tachezy, and Miklós Müller. 2019. “Metabolism of Trichomonad Hydrogenosomes.” In , 127–58. Springer, Cham. https://doi.org/10.1007/978-3-030-17941-0_6.

Huang, Kuo-Yang, Seow-Chin Ong, Chih-Ching Wu, Chia-Wei Hsu, Hsin-Chung Lin, Yi-Kai Fang, Wei-Hung Cheng, Po-Jung Huang, Cheng-Hsun Chiu, and Petrus Tang. 2019. “Metabolic Reprogramming of Hydrogenosomal Amino Acids in Trichomonas Vaginalis under Glucose Restriction.” *Journal of Microbiology, Immunology and Infection* 52 (4): 630–37. https://doi.org/10.1016/J.JMII.2017.10.005.

Hyde, Ralph J., Carol E. Cass, James D. Young, and James D. Stephen A. Baldwin. 2001. "The ENT Family of Eukaryote Nucleoside and Nucleobase Transporters: Recent Advances in the Investigation of Structure/Function Relationships and the Identification of Novel Isoforms." *Molecular Membrane Biology* 18 (1): 53–63. https://doi.org/10.1080/09687680118799.

Jack, Donald L., Nelson M. Yang, and Milton H. Saier. 2001. "The Drug/Metabolite Transporter Superfamily." *European Journal of Biochemistry* 268 (13): 3620–39. https://doi.org/10.1046/j.1432-1327.2001.02265.x.

Jedelský, Petr L, Pavel Doležal, Petr Rada, Jan Pyrih, Ondřej Smíd, Ivan Hrdý, Miroslava Sedinová, et al. 2011. "The Minimal Proteome in the Reduced Mitochondrion of the Parasitic Protist Giardia Intestinalis." *PloS One* 6 (2): e17285. https://doi.org/10.1371/journal.pone.0017285.

Jenkins, T M, T E Gorrell, M Müller, and P D Weitzman. 1991. "Hydrogenosomal Succinate Thiokinase in Tritrichomonas Foetus and Trichomonas Vaginalis." *Biochemical and Biophysical Research Communications* 179 (2): 892–96. https://doi.org/10.1016/0006-291x(91)91902-o.

Jerlström-Hultqvist, Jon, Elin Einarsson, Feifei Xu, Karin Hjort, Bo Ek, Daniel Steinhauf, Kjell Hultenby, Jonas Bergquist, Jan O. Andersson, and Staffan G. Svärd. 2013. "Hydrogenosomes in the Diplomonad Spironucleus Salmonicida." *Nature Communications* 4 (1): 2493. https://doi.org/10.1038/ncomms3493.

Jerlström-Hultqvist, Jon, Oscar Franzén, Johan Ankarklev, Feifei Xu, Eva Nohýnková, Jan O Andersson, Staffan G Svärd, and Björn Andersson. 2010. "Genome Analysis and Comparative Genomics of a Giardia Intestinalis Assemblage E Isolate." *BMC Genomics* 11 (October): 543. https://doi.org/10.1186/1471-2164-11-543.

Kamaishi, Takashi, Tetsuo Hashimoto, Yoshihiro Nakamura, Fuminori Nakamura, Shigenori Murata, Norihiro Okada, Ken-ichi Okamoto, Makoto Shimizu, and Masami Hasegawa. 1996. "Protein Phylogeny of Translation Elongation Factor EF-1α Suggests Microsporidians Are Extremely Ancient Eukaryotes." *Journal of Molecular Evolution* 42 (2): 257–63. https://doi.org/10.1007/BF02198852.

Karnkowska, Anna, Sebastian C Treitli, Ondřej Brzoň, Lukáš Novák, Vojtěch Vacek, Petr Soukal, Lael D Barlow, et al. 2019. "The Oxymonad Genome Displays Canonical

Eukaryotic Complexity in the Absence of a Mitochondrion." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msz147.

Karnkowska, Anna, Vojtěch Vacek, Zuzana Zubáčová, Sebastian C Treitli, Romana Petrželková, Laura Eme, Lukáš Novák, et al. 2016. "A Eukaryote without a Mitochondrial Organelle." *Current Biology : CB* 26 (10): 1274–84. https://doi.org/10.1016/j.cub.2016.03.053.

Katinka, M D, S Duprat, E Cornillot, G Méténier, F Thomarat, G Prensier, V Barbe, et al. 2001. "Genome Sequence and Gene Compaction of the Eukaryote Parasite Encephalitozoon Cuniculi." *Nature* 414 (6862): 450–53. https://doi.org/10.1038/35106579.

Kay, Christopher, Katharine D. Woodward, Karen Lawler, Tim J. Self, Sabrina D. Dyall, and Ian D. Kerr. 2012. "The ATP-Binding Cassette Proteins of the Deep-Branching Protozoan Parasite Trichomonas Vaginalis." *PLoS Neglected Tropical Diseases* 6 (6). https://doi.org/10.1371/JOURNAL.PNTD.0001693.

Keeling, Patrick J. 1998. "A Kingdom's Progress: Archezoa and the Origin of Eukaryotes." *BioEssays* 20 (1): 87–95. https://doi.org/10.1002/(SICI)1521-1878(199801)20:1<87::AID-BIES12>3.0.CO;2-4.

Keeling, Patrick J., and W. F. Doolittle. 1996. "Alpha-Tubulin from Early-Diverging Eukaryotic Lineages and the Evolution of the Tubulin Family." *Molecular Biology and Evolution* 13 (10): 1297–1305. https://doi.org/10.1093/oxfordjournals.molbev.a025576.

Keeling, Patrick J., and Brian S. Leander. 2003. "Characterisation of a Non-Canonical Genetic Code in the Oxymonad Streblomastix Strix." *Journal of Molecular Biology* 326 (5): 1337–49. https://doi.org/10.1016/S0022-2836(03)00057-3.

Keeling, Patrick J, and Claudio H Slamovits. 2004. "Simplicity and Complexity of Microsporidian Genomes." *Eukaryotic Cell* 3 (6): 1363–69. https://doi.org/10.1128/EC.3.6.1363-1369.2004.

Kessler, E. 1974. "Hydrogenase, Photoreduction, and Anaerobic Growth." *Botanical Monographs Oxford* 10: 456–73.

Klenk, Hans-Peter, Wolfram Zilllg, Martin Lanzendorfer, Bernd Grampp, and Peter Palm. 1995. "Location of Protist Lineages in a Phylogenetic Tree Inferred from Sequences of DNA-Dependent RNA Polymerases." *Archiv Für Protistenkunde* 145 (3–4): 221–30.

https://doi.org/10.1016/S0003-9365(11)80317-9.

Klodnicki, M. E., L. R. McDougald, and R. B. Beckstead. 2013. " A Genomic Analysis of Histomonas Meleagridis Through Sequencing of a CDNA Library ." *Journal of Parasitology* 99 (2): 264–69. https://doi.org/10.1645/ge-3256.1.

Knight, Jonathan. 2004. "Not so Special, after All?" *Nature* 429 (6989): 236–37. https://doi.org/10.1038/429236a.

Koning, Audrey P. De, Geoffrey P. Noble, Aaron A. Heiss, Jensen Wong, and Patrick J. Keeling. 2008. "Environmental PCR Survey to Determine the Distribution of a Non-Canonical Genetic Code in Uncultivable Oxymonads." *Environmental Microbiology* 10 (1): 65–74. https://doi.org/10.1111/j.1462-2920.2007.01430.x.

LaGier, M. J., Jan Tachezy, Frantisek Stejskal, Katerina Kutisova, and Janet S. Keithly. 2003. "Mitochondrial-Type Iron-Sulfur Cluster Biosynthesis Genes (IscS and IscU) in the Apicomplexan Cryptosporidium Parvum." *Microbiology* 149 (12): 3519–30. https://doi.org/10.1099/mic.0.26365-0.

Leander, Brian S., and Patrick J. Keeling. 2004. "Symbiotic Innovation in the Oxymonad Streblomastix Strix." *The Journal of Eukaryotic Microbiology* 51 (3): 291–300. https://doi.org/10.1111/j.1550-7408.2004.tb00569.x.

Lecompte, Odile, Raymond Ripp, Jean Claude Thierry, Dino Moras, and Olivier Poch. 2002. "Comparative Analysis of Ribosomal Proteins in Complete Genomes: An Example of Reductive Evolution at the Domain Scale." *Nucleic Acids Research* 30 (24): 5382–90. https://doi.org/10.1093/nar/gkf693.

Leger, Michelle M., Martin Kolisko, Ryoma Kamikawa, Courtney W. Stairs, Keitaro Kume, Ivan Čepička, Jeffrey D. Silberman, et al. 2017. "Organelles That Illuminate the Origins of Trichomonas Hydrogenosomes and Giardia Mitosomes." *Nature Ecology and Evolution* 1 (4): 0092. https://doi.org/10.1038/s41559-017-0092.

Leger, Michelle M., Martin Kolísko, Courtney W. Stairs, and Alastair G. B. Simpson. 2019. "Mitochondrion-Related Organelles in Free-Living Protists." In , 287–308. Springer, Cham. https://doi.org/10.1007/978-3-030-17941-0_12.

Leger, Michelle M, Laura Eme, Laura A Hug, and Andrew J Roger. 2016. "Novel Hydrogenosomes in the Microaerophilic Jakobid Stygiella Incarcerata." *Molecular*

Biology and Evolution 33 (9): 2318–36. https://doi.org/10.1093/molbev/msw103.

Lepère, Cécile, Isabelle Domaizon, Mylène Hugoni, Agnès Vellet, and Didier Debroas. 2016. "Diversity and Dynamics of Active Small Microbial Eukaryotes in the Anoxic Zone of a Freshwater Meromictic Lake (Pavin, France)." *Frontiers in Microbiology* 7: 130. https://doi.org/10.3389/fmicb.2016.00130.

Li, Fuli, Julia Hinderberger, Henning Seedorf, Jin Zhang, Wolfgang Buckel, and Rudolf K. Thauer. 2008. "Coupled Ferredoxin and Crotonyl Coenzyme A (CoA) Reduction with NADH Catalyzed by the Butyryl-CoA Dehydrogenase/Etf Complex from Clostridium Kluyveri." *Journal of Bacteriology* 190 (3): 843–50. https://doi.org/10.1128/JB.01417-07.

Liapounova, Natalia A, Vladimír Hampl, Paul M K Gordon, Christoph W Sensen, Lashitew Gedamu, and Joel B Dacks. 2006. "Reconstructing the Mosaic Glycolytic Pathway of the Anaerobic Eukaryote Monocercomonoides." *Eukaryotic Cell* 5 (12): 2138–46. https://doi.org/10.1128/EC.00258-06.

Lill, Roland, Rafal Dutkiewicz, Sven A. Freibert, Torsten Heidenreich, Judita Mascarenhas, Daili J. Netz, Viktoria D. Paul, et al. 2015. "The Role of Mitochondria and the CIA Machinery in the Maturation of Cytosolic and Nuclear Iron-Sulfur Proteins." *European Journal of Cell Biology* 94 (7–9): 280–91. https://doi.org/10.1016/j.ejcb.2015.05.002.

Lindmark, D G, and M Müller. 1973. "Hydrogenosome, a Cytoplasmic Organelle of the Anaerobic Flagellate Tritrichomonas Foetus, and Its Role in Pyruvate Metabolism." *The Journal of Biological Chemistry* 248 (22): 7724–28. http://www.ncbi.nlm.nih.gov/pubmed/4750424.

Lloyd, David, and Janine C Harris. 2002. "Giardia: Highly Evolved Parasite or Early Branching Eukaryote?" *Trends in Microbiology* 10 (3): 122–27. https://doi.org/10.1016/S0966-842X(02)02306-5.

Loftus, Brendan, Iain Anderson, Rob Davies, U. Cecilia M. Alsmark, John Samuelson, Paolo Amedeo, Paola Roncaglia, et al. 2005. "The Genome of the Protist Parasite Entamoeba Histolytica." *Nature* 433 (7028): 865–68. https://doi.org/10.1038/nature03291.

Lujan, Hugo D., and Theodore E. Nash. 1994. "The Uptake and Metabolism of Cysteine by Giardia Lamblia Trophozoites." *The Journal of Eukaryotic Microbiology* 41 (2): 169–75. https://doi.org/10.1111/j.1550-7408.1994.tb01491.x.

Mai, Zhiming, Sudip Ghosh, Marta Frisardi, Ben Rosenthal, Rick Rogers, and John Samuelson. 1999. " Hsp60 Is Targeted to a Cryptic Mitochondrion-Derived Organelle ('Crypton') in the Microaerophilic Protozoan Parasite Entamoeba Histolytica ." *Molecular and Cellular Biology* 19 (3): 2198–2205. https://doi.org/10.1128/mcb.19.3.2198.

Makki, Abhijith, Petr Rada, Vojtěch Žárský, Sami Kereïche, Lubomír Kováčik, Marian Novotný, Tobias Jores, Doron Rapaport, and Jan Tachezy. 2019. "Triplet-Pore Structure of a Highly Divergent TOM Complex of Hydrogenosomes in Trichomonas Vaginalis." *PLoS Biology* 17 (1): e3000098. https://doi.org/10.1371/journal.pbio.3000098.

Marger, Michael D., and Milton H. Saier. 1993. "A Major Superfamily of Transmembrane Facilitators That Catalyse Uniport, Symport and Antiport." *Trends in Biochemical Sciences* 18 (1): 13–20. https://doi.org/10.1016/0968-0004(93)90081-W.

Mazumdar, Rounik, Lukas Endler, Andreas Monoyios, Michael Hess, and Ivana Bilic. 2017. "Establishment of a de Novo Reference Transcriptome of Histomonas Meleagridis Reveals Basic Insights About Biological Functions and Potential Pathogenic Mechanisms of the Parasite." *Protist* 168 (6): 663–85. https://doi.org/10.1016/J.PROTIS.2017.09.004.

Mertens, E. 1993. "ATP versus Pyrophosphate: Glycolysis Revisited in Parasitic Protists." *Parasitology Today* 9 (4): 122–26. https://doi.org/10.1016/0169-4758(93)90169-G.

Mi-ichi, Fumika, Takashi Makiuchi, Atsushi Furukawa, Dan Sato, and Tomoyoshi Nozaki. 2011. "Sulfate Activation in Mitosomes Plays an Important Role in the Proliferation of Entamoeba Histolytica." *PLoS Neglected Tropical Diseases* 5 (8): e1263. https://doi.org/10.1371/journal.pntd.0001263.

Mi-ichi, Fumika, Mohammad Abu Yousuf, Kumiko Nakada-Tsukui, and Tomoyoshi Nozaki. 2009. "Mitosomes in Entamoeba Histolytica Contain a Sulfate Activation Pathway." *Proceedings of the National Academy of Sciences of the United States of America* 106 (51): 21731–36. https://doi.org/10.1073/pnas.0907106106.

Millet, Coralie O M, David Lloyd, Michael P. Coogan, Joanna Rumsey, and Joanne Cable. 2011. "Carbohydrate and Amino Acid Metabolism of Spironucleus Vortens." *Experimental Parasitology* 129 (1): 17–26. https://doi.org/10.1016/j.exppara.2011.05.025.

Morada, Mary, Ondřej Šmíd, Vladimír Hampl, Robert Sutak, Brian Lam, Paola Rappelli, Daniele Dessì, Pier L Fiori, Jan Tachezy, and Nigel Yarlett. 2011. "Hydrogenosome-

Localization of Arginine Deiminase in Trichomonas Vaginalis." *Molecular and Biochemical Parasitology* 176 (1): 51–54. https://doi.org/10.1016/j.molbiopara.2010.10.004.

Morin, L, and J P Mignot. 1995. "Are Archamoebae True Archezoa? The Phylogenetic Position of Pelomyxa Sp., as Inferred from Large Subunit Ribosomal RNA Sequencing." *Eur J Protistol* 31: 402.

Morrison, Hilary G, Andrew G McArthur, Frances D Gillin, Stephen B Aley, Rodney D Adam, Gary J Olsen, Aaron A Best, et al. 2007. "Genomic Minimalism in the Early Diverging Intestinal Parasite Giardia Lamblia." *Science (New York, N.Y.)* 317 (5846): 1921–26. https://doi.org/10.1126/science.1143837.

Mukherjee, Mandira, Mark T Brown, Andrew G McArthur, and Patricia J Johnson. 2006. "Proteins of the Glycine Decarboxylase Complex in the Hydrogenosome of Trichomonas Vaginalis." *Eukaryotic Cell* 5 (12): 2062–71. https://doi.org/10.1128/EC.00205-06.

Mukherjee, Mandira, Stuart A. Sievers, Mark T. Brown, and Patricia J. Johnson. 2006. "Identification and Biochemical Characterization of Serine Hydroxymethyl Transferase in the Hydrogenosome of Trichomonas Vaginalis." *Eukaryotic Cell* 5 (12): 2072–78. https://doi.org/10.1128/EC.00249-06.

Müller, Miklós, Marek Mentel, Jaap J van Hellemond, Katrin Henze, Christian Woehle, Sven B Gould, Re-Young Yu, Mark van der Giezen, Aloysius G M Tielens, and William F Martin. 2012. "Biochemistry and Evolution of Anaerobic Energy Metabolism in Eukaryotes." *Microbiology and Molecular Biology Reviews : MMBR* 76 (2): 444–95. https://doi.org/10.1128/MMBR.05024-11.

Müller, Sylke, Eva Liebau, Rolf D Walter, and R Luise Krauth-Siegel. 2003. "Thiol-Based Redox Metabolism of Protozoan Parasites." *Trends in Parasitology* 19 (7): 320–28. http://www.ncbi.nlm.nih.gov/pubmed/12855383.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–74. https://doi.org/10.1093/molbev/msu300.

Noda, Satoko, Tetsushi Inoue, Yuichi Hongoh, Miho Kawai, Christine A Nalepa, Charunee Vongkaluang, Toshiaki Kudo, and Moriya Ohkuma. 2006. "Identification and

Characterization of Ectosymbionts of Distinct Lineages in Bacteroidales Attached to Flagellated Protists in the Gut of Termites and a Wood-Feeding Cockroach." *Environmental Microbiology* 8 (1): 11–20. https://doi.org/10.1111/j.1462-2920.2005.00860.x.

Noguchi, Fumiya, Shigeru Shimamura, Takuro Nakayama, Euki Yazaki, Akinori Yabuki, Tetsuo Hashimoto, Yuji Inagaki, Katsunori Fujikura, and Kiyotaka Takishita. 2015. "Metabolic Capacity of Mitochondrion-Related Organelles in the Free-Living Anaerobic Stramenopile Cantina Marsupialis." *Protist* 166 (5): 534–50. https://doi.org/10.1016/j.protis.2015.08.002.

Novák, Lukáš, Zuzana Zubáčová, Anna Karnkowska, Martin Kolisko, Miluše Hroudová, Courtney W. Stairs, Alastair G. B. Simpson, et al. 2016. "Arginine Deiminase Pathway Enzymes: Evolutionary History in Metamonads and Other Eukaryotes." *BMC Evolutionary Biology* 16 (1): 197. https://doi.org/10.1186/s12862-016-0771-4.

Nývltová, Eva, Tamara Smutná, Jan Tachezy, and Ivan Hrdý. 2016. "OsmC and Incomplete Glycine Decarboxylase Complex Mediate Reductive Detoxification of Peroxides in Hydrogenosomes of Trichomonas Vaginalis." *Molecular and Biochemical Parasitology* 206 (1–2): 29–38. https://doi.org/10.1016/j.molbiopara.2016.01.006.

Nývltová, Eva, Courtney W. Stairs, Ivan Hrdý, Jakub Rídl, Jan Mach, Jan Paqes, Andrew J. Roger, and Jan Tachezy. 2015. "Lateral Gene Transfer and Gene Duplication Played a Key Role in the Evolution of Mastigamoeba Balamuthi Hydrogenosomes." *Molecular Biology and Evolution* 32 (4): 1039–55. https://doi.org/10.1093/molbev/msu408.

Nývltová, Eva, Robert Šuták, Karel Harant, Miroslava Šedinová, Ivan Hrdy, Jan Paces, Čestmír Vlček, and Jan Tachezy. 2013. "NIF-Type Iron-Sulfur Cluster Assembly System Is Duplicated and Distributed in the Mitochondria and Cytosol of Mastigamoeba Balamuthi." *Proceedings of the National Academy of Sciences of the United States of America* 110 (18): 7371–76. https://doi.org/10.1073/pnas.1219590110.

O'Kelly, C J, M A Farmer, and T A Nerad. 1999. "Ultrastructure of Trimastix Pyriformis (Klebs) Bernard et Al.: Similarities of Trimastix Species with Retortamonad and Jakobid Flagellates." *Protist* 150 (2): 149–62. https://doi.org/10.1016/S1434-4610(99)70018-0.

Paget, Timothy A., Mary L. Kelly, Edward L. Jarroll, Donald G. Lindmark, and David Lloyd. 1993. "The Effects of Oxygen on Fermentation in Giardia Lamblia." *Molecular and*

Biochemical Parasitology 57 (1): 65–71. https://doi.org/10.1016/0166-6851(93)90244-R.

Paget, Timothy A., Michael H. Raynor, Donald W.E. Shipp, and David Lloyd. 1990. "Giardia Lamblia Produces Alanine Anaerobically but Not in the Presence of Oxygen." *Molecular and Biochemical Parasitology* 42 (1): 63–67. https://doi.org/10.1016/0166-6851(90)90113-Z.

Pajor, Ana M. 2014. "Sodium-Coupled Dicarboxylate and Citrate Transporters from the SLC13 Family." *Pflugers Archiv European Journal of Physiology* 466 (1): 119–30. https://doi.org/10.1007/s00424-013-1369-y.

Pánek, Tomáš, Jeffrey D. Silberman, Naoji Yubuki, Brian S. Leander, and Ivan Cepicka. 2012. "Diversity, Evolution and Molecular Systematics of the Psalteriomonadidae, the Main Lineage of Anaerobic/Microaerophilic Heteroloboseans (Excavata: Discoba)." *Protist* 163 (6): 807–31. http://www.sciencedirect.com/science/article/pii/S1434461011001209.

Pao, S S, I T Paulsen, M H Saier, and Jr. 1998. "Major Facilitator Superfamily." *Microbiology and Molecular Biology Reviews : MMBR* 62 (1): 1–34. http://www.ncbi.nlm.nih.gov/pubmed/9529885.

Parra, Genis, Keith Bradnam, and Ian Korf. 2007. "CEGMA: A Pipeline to Accurately Annotate Core Genes in Eukaryotic Genomes." *Bioinformatics* 23 (9): 1061–67. https://doi.org/10.1093/bioinformatics/btm071.

Patron, Nicola J, Dion G Durnford, and Stanislav Kopriva. 2008. "Sulfate Assimilation in Eukaryotes: Fusions, Relocations and Lateral Transfers." *BMC Evolutionary Biology* 8 (1): 39. https://doi.org/10.1186/1471-2148-8-39.

Payne, M J, A Chapman, and R Cammack. 1993. "Evidence for an [Fe]-Type Hydrogenase in the Parasitic Protozoan Trichomonas Vaginalis." *FEBS Letters* 317 (1–2): 101–4. https://doi.org/10.1016/0014-5793(93)81500-y.

Peyretaillade, E, C Biderre, P Peyret, F Duffieux, G Méténier, M Gouy, B Michot, and C P Vivarès. 1998. "Microsporidian Encephalitozoon Cuniculi, a Unicellular Eukaryote with an Unusual Chromosomal Dispersion of Ribosomal Genes and a LSU RRNA Reduced to the Universal Core." *Nucleic Acids Research* 26 (15): 3513–20. https://doi.org/10.1093/nar/26.15.3513.

Philippe, Hervé, and Agnès Germot. 2000. "Phylogeny of Eukaryotes Based on Ribosomal

RNA: Long-Branch Attraction and Models of Sequence Evolution." *Molecular Biology and Evolution* 17 (5): 830–34. https://doi.org/10.1093/oxfordjournals.molbev.a026362.

Pollo, Stephen M. J., Sarah J. Reiling, Janneke Wit, Matthew Workentine, Rebecca A. Guy, G. William Batoff, Janet Yee, Brent R. Dixon, and James D. Wasmuth. 2018. "MinION Re-Sequencing of Giardia Genomes and de Novo Assembly of a New Giardia Isolate." *BioRxiv*, 343541. https://doi.org/10.1101/343541.

Pütz, Simone, Pavel Dolezal, Gabriel Gelius-Dietrich, Lenka Bohacova, Jan Tachezy, and Katrin Henze. 2006. "Fe-Hydrogenase Maturases in the Hydrogenosomes of Trichomonas Vaginalis." *Eukaryotic Cell* 5 (3): 579–86. https://doi.org/10.1128/EC.5.3.579-586.2006.

Pütz, Simone, Gabriel Gelius-Dietrich, Markus Piotrowski, and Katrin Henze. 2005. "Rubrerythrin and Peroxiredoxin: Two Novel Putative Peroxidases in the Hydrogenosomes of the Microaerophilic Protozoon Trichomonas Vaginalis." *Molecular and Biochemical Parasitology* 142 (2): 212–23. https://doi.org/10.1016/j.molbiopara.2005.04.003.

Rada, Petr, Pavel Doležal, Petr L. Jedelský, Dejan Bursac, Andrew J. Perry, Miroslava Šedinová, Kateřina Smíšková, et al. 2011. "The Core Components of Organelle Biogenesis and Membrane Transport in the Hydrogenosomes of Trichomonas Vaginalis." *PLoS ONE* 6 (9): e24428. https://doi.org/10.1371/journal.pone.0024428.

Radek, Renate. 1994. "Monocercomonoides Termitis n. Sp., an Oxymonad from the Lower Termite Kalotermes Sinaicus." *Archiv Für Protistenkunde* 144 (4): 373–82. https://doi.org/10.1016/S0003-9365(11)80240-X.

Riordan, Christina E., Jeffrey G. Ault, Susan G. Langreth, and Janet S. Keithly. 2003. "Cryptosporidium Parvum Cpn60 Targets a Relict Organelle." *Current Genetics* 44 (3): 138–47. https://doi.org/10.1007/s00294-003-0432-1.

Roger, Andrew J., C. Graham Clark, and W. Ford Doolittle. 1996. "A Possible Mitochondrial Gene in the Early-Branching Amitochondriate Protist Trichomonas Vaginalis." *Proceedings of the National Academy of Sciences of the United States of America* 93 (25): 14618–22. https://doi.org/10.1073/pnas.93.25.14618.

Roger, Andrew J., Staffan G. Svärd, Jorge Tovar, C. Graham Clark, Michael W. Smith, Frances D. Gillin, and Mitchell L. Sogin. 1998. "A Mitochondrial-like Chaperonin 60 Gene in Giardia Lamblia: Evidence That Diplomonads Once Harbored an Endosymbiont Related

to the Progenitor of Mitochondria." *Proceedings of the National Academy of Sciences of the United States of America* 95 (1): 229–34. https://doi.org/10.1073/pnas.95.1.229.

Sanchez, Lidya B., and Miklós Müller. 1996. "Purification and Characterization of the Acetate Forming Enzyme, Acetyl-CoA Synthetase (ADP-Forming) from the Amitochondriate Protist, *Giardia Lamblia*." *FEBS Letters* 378 (3): 240–44. https://doi.org/10.1016/0014-5793(95)01463-2.

Schneider, Rachel E., Mark T. Brown, April M. Shiflett, Sabrina D. Dyall, Richard D. Hayes, Yongming Xie, Joseph A. Loo, and Patricia J. Johnson. 2011. "The Trichomonas Vaginalis Hydrogenosome Proteome Is Highly Reduced Relative to Mitochondria, yet Complex Compared with Mitosomes." *International Journal for Parasitology* 41 (13–14): 1421–34. https://doi.org/10.1016/j.ijpara.2011.10.001.

Schofield, Philip J, M R Edwards, G Grossman, and E A Tutticci. 1995. "Amino Acid Exchange Activity of the Alanine Transporter of Giardia Intestinalis." *Experimental Parasitology* 80 (1): 124–32. https://doi.org/10.1006/expr.1995.1014.

Schofield, Philip J, Michael R Edwards, Jacqueline Matthews, and Justine R Wilson. 1992. "The Pathway of Arginine Catabolism in Giardia Intestinalis." *Molecular and Biochemical Parasitology* 51 (1): 29–36. https://doi.org/10.1016/0166-6851(92)90197-R.

Shah, Preetam H, Jonathan K Stiles, Richard W Finley, William B Lushbaugh, and John C Meade. 2002. "Trichomonas Vaginalis: Characterization of a Family of P-Type ATPase Genes." *Parasitology International* 51 (1): 41–51. http://www.ncbi.nlm.nih.gov/pubmed/11880226.

Shimodaira, Hidetoshi. 2002. "An Approximately Unbiased Test of Phylogenetic Tree Selection." *Systematic Biology* 51 (3): 492–508. https://doi.org/10.1080/10635150290069913.

Simpson, Alastair G B. 2003. "Cytoskeletal Organization, Phylogenetic Affinities and Systematics in the Contentious Taxon Excavata (Eukaryota)." *International Journal of Systematic and Evolutionary Microbiology* 53 (6): 1759–77. https://doi.org/10.1099/ijs.0.02578-0.

Slamovits, Claudio H, and Patrick J Keeling. 2006a. "Pyruvate-Phosphate Dikinase of Oxymonads and Parabasalia and the Evolution of Pyrophosphate-Dependent Glycolysis in Anaerobic Eukaryotes." *Eukaryotic Cell* 5 (1): 148. https://doi.org/10.1128/EC.5.1.148-

154.2006.

———. 2006b. “A High Density of Ancient Spliceosomal Introns in Oxymonad Excavates.” *BMC Evolutionary Biology* 6 (April): 34. https://doi.org/10.1186/1471-2148-6-34.

Šmíd, Ondřej, Anna Matušková, Simon R. Harris, Tomáš Kučera, Marián Novotný, Lenka Horváthová, Ivan Hrdý, et al. 2008. “Reductive Evolution of the Mitochondrial Processing Peptidases of the Unicellular Parasites Trichomonas Vaginalis and Giardia Intestinalis.” *PLoS Pathogens* 4 (12): e1000243. https://doi.org/10.1371/journal.ppat.1000243.

Smith, Anthony C., James A. Blackshaw, and Alan J. Robinson. 2012. “MitoMiner: A Data Warehouse for Mitochondrial Proteomics Data.” *Nucleic Acids Research* 40 (D1): D1160-7. https://doi.org/10.1093/nar/gkr1101.

Smutná, Tamara, Vera L. Gonçalves, Lígia M. Saraiva, Jan Tachezy, Miguel Teixeira, and Ivan Hrdý. 2009. “Flavodiiron Protein from *Trichomonas Vaginalis* Hydrogenosomes: The Terminal Oxygen Reductase.” *Eukaryotic Cell* 8 (1): 47–55. https://doi.org/10.1128/EC.00276-08.

Sodani, Kamlesh, Atish Patel, Rishil J Kathawala, and Zhe-Sheng Chen. 2012. “Multidrug Resistance Associated Proteins in Multidrug Resistance.” *Chinese Journal of Cancer* 31 (2): 58–72. https://doi.org/10.5732/cjc.011.10329.

Spang, Anja, Jimmy H Saw, Steffen L Jørgensen, Katarzyna Zaremba-Niedzwiedzka, Joran Martijn, Anders E Lind, Roel van Eijk, Christa Schleper, Lionel Guy, and Thijs J G Ettema. 2015. “Complex Archaea That Bridge the Gap between Prokaryotes and Eukaryotes.” *Nature* 521 (7551): 173–79. https://doi.org/10.1038/nature14447.

Stadelmann, Britta, Kurt Hanevik, Mattias K Andersson, Oystein Bruserud, and Staffan G Svärd. 2013. “The Role of Arginine and Arginine-Metabolizing Enzymes during Giardia - Host Cell Interactions in Vitro.” *BMC Microbiology* 13 (1): 256. https://doi.org/10.1186/1471-2180-13-256.

Stairs, Courtney W, Laura Eme, Matthew W Brown, Cornelis Mutsaers, Edward Susko, Graham Dellaire, Darren M Soanes, Mark van der Giezen, and Andrew J Roger. 2014. “A SUF Fe-S Cluster Biogenesis System in the Mitochondrion-Related Organelles of the Anaerobic Protist Pygsuia.” *Current Biology : CB* 24 (11): 1176–86. https://doi.org/10.1016/j.cub.2014.04.033.

Stamatakis, Alexandros. 2014. “RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies.” *Bioinformatics (Oxford, England)* 30 (9): 1312–13. https://doi.org/10.1093/bioinformatics/btu033.

Stechmann, Alexandra, Manuela Baumgartner, Jeffrey D Silberman, and Andrew J Roger. 2006. “The Glycolytic Pathway of Trimastix Pyriformis Is an Evolutionary Mosaic.” *BMC Evolutionary Biology* 6 (January): 101. https://doi.org/10.1186/1471-2148-6-101.

Steinbüchel, Alexander, and Miklós Müller. 1986. “Anaerobic Pyruvate Metabolism of Tritrichomonas Foetus and Trichomonas Vaginalis Hydrogenosomes.” *Molecular and Biochemical Parasitology* 20 (1): 57–65. https://doi.org/10.1016/0166-6851(86)90142-8.

Strimmer, Korbinian, and Andrew Rambaut. 2002. “Inferring Confidence Sets of Possibly Misspecified Gene Trees.” *Proceedings. Biological Sciences / The Royal Society* 269: 137–42. https://doi.org/10.1098/rspb.2001.1862.

Sutak, Robert, Pavel Dolezal, Heather L. Fiumera, Ivan Hrdy, Andrew Dancist, Maria Delgadillo-Correa, Patricia J. Johnson, Miklós Müller, and Jan Tachezy. 2004. “Mitochondrial-Type Assembly of FeS Centers in the Hydrogenosomes of the Amitochondriate Eukaryote Trichomonas Vaginalis.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (28): 10368–73. https://doi.org/10.1073/pnas.0401319101.

Tachezy, Jan, Lidya B. Sanchez, and Miklos Muller. 2001. “Mitochondrial Type Iron-Sulfur Cluster Assembly in the Amitochondriate Eukaryotes Trichomonas Vaginalis and Giardia Intestinalis, as Indicated by the Phylogeny of IscS.” *Mol. Biol. Evol.* 18 (10): 1919–28. http://mbe.oxfordjournals.org/cgi/content/abstract/18/10/1919.

Takishita, Kiyotaka, Martin Kolisko, Hiroshi Komatsuzaki, Akinori Yabuki, Yuji Inagaki, Ivan Cepicka, Pavla Smejkalová, et al. 2012. “Multigene Phylogenies of Diverse Carpediemonas-like Organisms Identify the Closest Relatives of ‘Amitochondriate’ Diplomonads and Retortamonads.” *Protist* 163 (3): 344–55. http://www.sciencedirect.com/science/article/pii/S1434461012000181.

Tanifuji, Goro, Shun Takabayashi, Keitaro Kume, Mizue Takagi, Takuro Nakayama, Ryoma Kamikawa, Yuji Inagaki, and Tetsuo Hashimoto. 2018. “The Draft Genome of Kipferlia Bialata Reveals Reductive Genome Evolution in Fornicate Parasites.” *PLoS ONE* 13 (3). https://doi.org/10.1371/JOURNAL.PONE.0194487.

Tekle, Yonas I., O. Roger Anderson, Laura A. Katz, Xyrus X. Maurer-Alcalá, Mario Alberto Cerón Romero, and Robert Molestina. 2016. "Phylogenomics of 'Discosea': A New Molecular Phylogenetic Perspective on Amoebozoa with Flat Body Forms." *Molecular Phylogenetics and Evolution* 99 (June): 144–54. https://doi.org/10.1016/J.YMPEV.2016.03.029.

Tielens, Aloysius G.M., Koen W.A. van Grinsven, Katrin Henze, Jaap J. van Hellemond, and William Martin. 2010. "Acetate Formation in the Energy Metabolism of Parasitic Helminths and Protists." *International Journal for Parasitology* 40 (4): 387–97. https://doi.org/10.1016/J.IJPARA.2009.12.006.

Tovar, Jorge, Anke Fischer, and C. Graham Clark. 1999. "The Mitosome, a Novel Organelle Related to Mitochondria in the Amitochondrial Parasite Entamoeba Histolytica." *Molecular Microbiology* 32 (5): 1013–21. https://doi.org/10.1046/j.1365-2958.1999.01414.x.

Tovar, Jorge, Gloria León-Avila, Lidya B Sánchez, Robert Sutak, Jan Tachezy, Mark van der Giezen, Manuel Hernández, Miklós Müller, and John M. Lucocq. 2003. "Mitochondrial Remnant Organelles of Giardia Function in Iron-Sulphur Protein Maturation." *Nature* 426 (6963): 172–76. https://doi.org/10.1038/nature01945.

Treitli, Sebastian C., Martin Kolisko, Filip Husník, Patrick J. Keeling, and Vladimír Hampl. 2019. " Revealing the Metabolic Capacity of Streblomastix Strix and Its Bacterial Symbionts Using Single-Cell Metagenomics ." *Proceedings of the National Academy of Sciences* 116 (39): 19675–84. https://doi.org/10.1073/pnas.1910793116.

Treitli, Sebastian C., Michael Kotyk, Naoji Yubuki, Eliška Jirounková, Jitka Vlasáková, Pavla Smejkalová, Petr Šípek, Ivan Čepička, and Vladimír Hampl. 2018. "Molecular and Morphological Diversity of the Oxymonad Genera Monocercomonoides and Blattamonas Gen. Nov." *Protist* 169 (5): 744–83. https://doi.org/10.1016/j.protis.2018.06.005.

Tsaousis, Anastasios D, Sandrine Ollagnier de Choudens, Eleni Gentekaki, Shaojun Long, Daniel Gaston, Alexandra Stechmann, Daniel Vinella, et al. 2012. "Evolution of Fe/S Cluster Biogenesis in the Anaerobic Parasite Blastocystis." *Proceedings of the National Academy of Sciences of the United States of America* 109 (26): 10426–31. https://doi.org/10.1073/pnas.1116067109.

Vacek, Vojtěch, Lukáš V F Novák, Sebastian C Treitli, Petr Táborský, Ivan Čepička, Martin

Kolísko, Patrick J Keeling, Vladimír Hampl, and Iñaki Ruiz-Trillo. 2018. “Fe-S Cluster Assembly in Oxymonads and Related Protists.” *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msy168.

Vaňáčová, Štěpánka, Weihong Yan, Jane M Carlton, and Patricia J Johnson. 2005. “Spliceosomal Introns in the Deep-Branching Eukaryote Trichomonas Vaginalis.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (12): 4430–35. https://doi.org/10.1073/pnas.0407500102.

Vávra, Jiří. 1976. “Structure of the Microsporidia.” In *Biology of the Microsporidia*, 1–84. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4684-3114-8_1.

Viscogliosi, Eric, Pilar Delgado-Viscogliosi, Delphine Gerbod, Manuel Dauchez, Sylvie Gratepanche, Alain J.P Alix, and Daniel Dive. 1998. “Cloning and Expression of an Iron-Containing Superoxide Dismutase in the Parasitic Protist, *Trichomonas Vaginalis*.” *FEMS Microbiology Letters* 161 (1): 115–23. https://doi.org/10.1111/j.1574-6968.1998.tb12936.x.

Viscogliosi, Eric, Herve Philippe, Anne Baroin, Roland Perasso, and Guy Brugerolle. 1993. “Phylogeny of Trichomonads Based On Partial Sequences of Large Subunit Rrna and On Cladistic Analysis of Morphological Data.” *Journal of Eukaryotic Microbiology* 40 (4): 411–21. https://doi.org/10.1111/j.1550-7408.1993.tb04935.x.

Vivares, C., C. Biderre, F. Duffieux, E. Peyretaillade, P. Peyret, C. Metenier, and M. Pages. 1996. “Chromosomal Localization of Five Genes in Encephalitozoon Cuniculi (Microsporidia).” *The Journal of Eukaryotic Microbiology* 43 (5): 97S-97S. https://doi.org/10.1111/j.1550-7408.1996.tb05021.x.

Vossbrinck, Charles R., J. V. Maddox, S. Friedman, B. A. Debrunner-Vossbrinck, and C. R. Woese. 1987. “Ribosomal RNA Sequence Suggests Microsporidia Are Extremely Ancient Eukaryotes.” *Nature* 326 (6111): 411–14. https://doi.org/10.1038/326411a0.

Vossbrinck, Charles R., and Carl R. Woese. 1986. “Eukaryotic Ribosomes That Lack a 5.8S RNA.” *Nature* 320 (6059): 287–88. https://doi.org/10.1038/320287a0.

Westrop, Gareth D., Gordon Goodall, Jeremy C. Mottram, and Graham H. Coombs. 2006. “Cysteine Biosynthesis in *Trichomonas Vaginalis* Involves Cysteine Synthase Utilizing *O* -Phosphoserine.” *Journal of Biological Chemistry* 281 (35): 25062–75. https://doi.org/10.1074/jbc.M600688200.

Westrop, Gareth D., Lijie Wang, Gavin J. Blackburn, Tong Zhang, Liang Zheng, David G. Watson, and Graham H. Coombs. 2017. "Metabolomic Profiling and Stable Isotope Labelling of Trichomonas Vaginalis and Tritrichomonas Foetus Reveal Major Differences in Amino Acid Metabolism Including the Production of 2-Hydroxyisocaproic Acid, Cystathionine and S-Methylcysteine." *PLOS ONE* 12 (12): e0189072. https://doi.org/10.1371/journal.pone.0189072.

Williams, Bryony A.P., Robert P Hirt, John M Lucocq, and T Martin Embley. 2002. "A Mitochondrial Remnant in the Microsporidian Trachipleistophora Hominis." *Nature* 418 (6900): 865–69. https://doi.org/10.1038/nature00949.

Williams, K, P N Lowe, and P F Leadlay. 1987. "Purification and Characterization of Pyruvate: Ferredoxin Oxidoreductase from the Anaerobic Protozoon Trichomonas Vaginalis." *The Biochemical Journal* 246 (2): 529–36. https://doi.org/10.1042/bj2460529.

Xu, Feifei, Jon Jerlström-Hultqvist, Elin Einarsson, Ásgeir Ástvaldsson, Staffan G. Svärd, and Jan O. Andersson. 2014. "The Genome of Spironucleus Salmonicida Highlights a Fish Pathogen Adapted to Fluctuating Environments." *PLoS Genetics* 10 (2): e1004053. https://doi.org/10.1371/journal.pgen.1004053.

Xu, Ping, Giovanni Widmer, Yingping Wang, Lulz S Ozaki, Joao M Alves, Myrna G Serrano, Daniela Pulu, et al. 2004. "The Genome of Cryptosporidium Hominis." *Nature* 431 (7012): 1107–12. https://doi.org/10.1038/nature02977.

Yamamoto, Ayako, Tetsuo Hashimoto, Emiko Asaga, Masami Hasegawa, and Nobuichi Goto. 1997. "Phylogenetic Position of the Mitochondrion-Lacking Protozoan Trichomonas Tenax, Based on Amino Acid Sequences of Elongation Factors 1α and 2." *Journal of Molecular Evolution* 44 (1): 98–105. https://doi.org/10.1007/PL00006127.

Yarlett, Nigel, Donald G Lindmark, Burt Goldberg, M Ali Moharrami, and Cyrus J Bacchi. 1994. "Subcellular Localization of the Enzymes of the Arginine Dihydrolase Pathway in Trichomonas Vaginalis and Tritrichomonas Foetus." *Journal of Eukaryotic Microbiology* 41 (6): 554–59. https://doi.org/10.1111/j.1550-7408.1994.tb01516.x.

Yarlett, Nigel, Martha P. Martinez, M. Ali Moharrami, and Jan Tachezy. 1996. "The Contribution of the Arginine Dihydrolase Pathway to Energy Metabolism by Trichomonas Vaginalis." *Molecular and Biochemical Parasitology* 78 (1–2): 117–25. https://doi.org/10.1016/S0166-6851(96)02616-3.

Yubuki, Naoji, Tomáš Pánek, Akinori Yabuki, Ivan Čepička, Kiyotaka Takishita, Yuji Inagaki, and Brian S. Leander. 2015. "Morphological Identities of Two Different Marine Stramenopile Environmental Sequence Clades: *Bicosoeca Kenaiensis* (Hilliard, 1971) and *Cantina Marsupialis* (Larsen and Patterson, 1990) Gen. Nov., Comb. Nov." *Journal of Eukaryotic Microbiology* 62 (4): 532–42. https://doi.org/10.1111/jeu.12207.

Yubuki, Naoji, Eliška Zadrobílková, and Ivan Čepička. 2017. "Ultrastructure and Molecular Phylogeny of *Iotanema Spirale* Gen. Nov. et Sp. Nov., a New Lineage of Endobiotic Fornicata with Strikingly Simplified Ultrastructure." *Journal of Eukaryotic Microbiology* 64 (4): 422–33. https://doi.org/10.1111/jeu.12376.

Zhang, Qianqian, Petr Táborský, Jeffrey D Silberman, Tomáš Pánek, Ivan Čepička, and Alastair G B Simpson. 2015. "Marine Isolates of Trimastix Marina Form a Plesiomorphic Deep-Branching Lineage within Preaxostyla, Separate from Other Known Trimastigids (Paratrimastix n. Gen.)." *Protist* 166 (4): 468–91. https://doi.org/10.1016/j.protis.2015.07.003.

Zubáčová, Zuzana, Lukáš Novák, Jitka Bublíková, Vojtěch Vacek, Jan Fousek, Jakub Rídl, Jan Tachezy, Pavel Doležal, Čestmír Vlček, and Vladimír Hampl. 2013. "The Mitochondrion-Like Organelle of Trimastix Pyriformis Contains the Complete Glycine Cleavage System." *PLoS ONE* 8 (3): e55417. https://doi.org/10.1371/journal.pone.0055417.

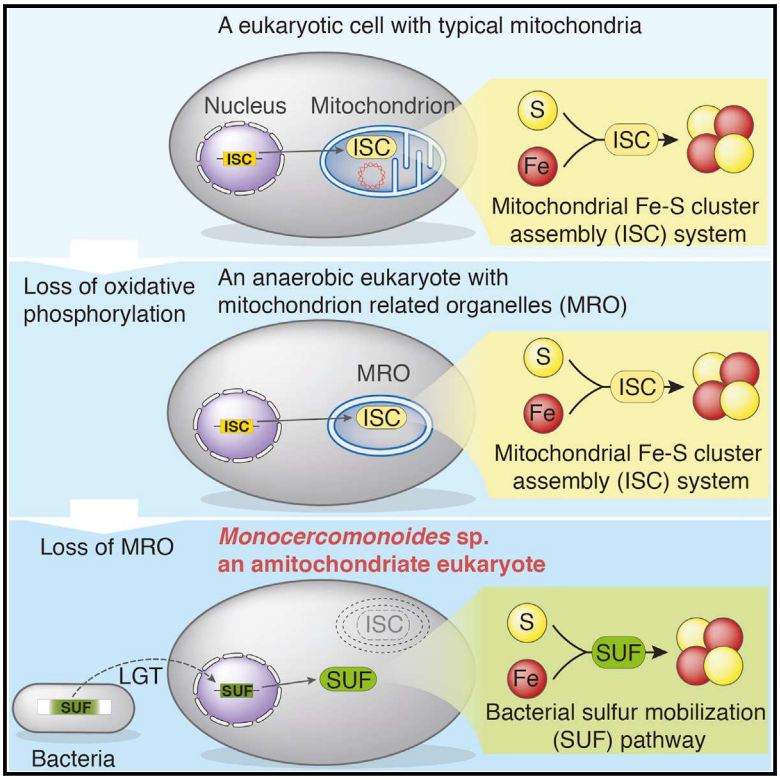
Zuo, X., B. C. Lockwood, and G. H. Coombs. 1995. "Uptake of Amino Acids by the Parasitic, Flagellated Protist Trichomonas Vaginalis." *Microbiology* 141 (10): 2637–42. https://doi.org/10.1099/13500872-141-10-2637.

Article

Current Biology

A Eukaryote without a Mitochondrial Organelle

Graphical Abstract



Highlights

- *Monocercomonoides* sp. is a eukaryotic microorganism with no mitochondria
- The complete absence of mitochondria is a secondary loss, not an ancestral feature
- The essential mitochondrial ISC pathway was replaced by a bacterial SUF system

Authors

Anna Karnkowska, Vojtěch Vacek, Zuzana Zubáčová, ..., Joel B. Dacks, Čestmír Vlček, Vladimír Hampl

Correspondence

ankarn@biol.uw.edu.pl (A.K.), vlada@natur.cuni.cz (V.H.)

In Brief

Karnkowska et al. overturn the paradigm that eukaryotes must have mitochondria. Their genomic investigation of the anaerobic microbial eukaryote *Monocercomonoides* sp. reveals a complete lack of mitochondrial organelle and functions including Fe-S cluster synthesis, which is carried out in the cytosol by a laterally acquired bacterial pathway.

Karnkowska et al., 2016, Current Biology *26*, 1274–1284
May 23, 2016
http://dx.doi.org/10.1016/j.cub.2016.03.053

CellPress

Current Biology
Article

A Eukaryote without a Mitochondrial Organelle

Anna Karnkowska,[1,2,7,*] Vojtěch Vacek,[1] Zuzana Zubáčová,[1] Sebastian C. Treitli,[1] Romana Petrželková,[3] Laura Eme,[4] Lukáš Novák,[1] Vojtěch Žárský,[1] Lael D. Barlow,[5] Emily K. Herman,[5] Petr Soukal,[1] Miluše Hroudová,[6] Pavel Doležal,[1] Courtney W. Stairs,[4] Andrew J. Roger,[4] Marek Eliáš,[3] Joel B. Dacks,[5] Čestmír Vlček,[6] and Vladimír Hampl[1,*]

[1]Department of Parasitology, Charles University in Prague, Prague 12843, Czech Republic
[2]Department of Molecular Phylogenetics and Evolution, University of Warsaw, Warsaw 00478, Poland
[3]Department of Biology and Ecology, University of Ostrava, Ostrava 710 00, Czech Republic
[4]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS B3H 4R2, Canada
[5]Department of Cell Biology, University of Alberta, Edmonton, AB T6G 2H7, Canada
[6]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague 14220, Czech Republic
[7]Present address: Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
*Correspondence: ankarn@biol.uw.edu.pl (A.K.), vlada@natur.cuni.cz (V.H.)
http://dx.doi.org/10.1016/j.cub.2016.03.053

SUMMARY

The presence of mitochondria and related organelles in every studied eukaryote supports the view that mitochondria are essential cellular components. Here, we report the genome sequence of a microbial eukaryote, the oxymonad *Monocercomonoides* sp., which revealed that this organism lacks all hallmark mitochondrial proteins. Crucially, the mitochondrial iron-sulfur cluster assembly pathway, thought to be conserved in virtually all eukaryotic cells, has been replaced by a cytosolic sulfur mobilization system (SUF) acquired by lateral gene transfer from bacteria. In the context of eukaryotic phylogeny, our data suggest that *Monocercomonoides* is not primitively amitochondrial but has lost the mitochondrion secondarily. This is the first example of a eukaryote lacking any form of a mitochondrion, demonstrating that this organelle is not absolutely essential for the viability of a eukaryotic cell.

INTRODUCTION

Mitochondria are organelles that arose through the endosymbiotic integration of an α-proteobacterial endosymbiont into the proto-eukaryote host cell. During the course of eukaryotic evolution, the genome and proteome of the mitochondrial compartment have been significantly modified, and many functions have been gained, lost, or relocated [1]. In extreme cases, the derivatives of mitochondria in anaerobic protists had become so modified that they had been overlooked [2] or not recognized as homologous to the mitochondrion [3]. Indeed, in the 1980s, the Archezoa hypothesis [4] proposed that some microbial eukaryotes primitively lacked mitochondria, peroxisomes, stacked Golgi apparatus, spliceosomal introns, and sexual reproduction. However, over the following decade, double-membraned organelles were identified in all investigated putative Archezoa. The final nail in the coffin of the Archezoa hypothesis was the demonstration that these organelles all contain some mitochondrial marker proteins, such as those involved in the iron-sulfur cluster (ISC) Fe-S clusters biogenesis system, translocases, maturases, and/or molecular chaperones known to facilitate the import of proteins into mitochondria. It is now widely accepted that mitochondria or mitochondrion-related organelles (MROs) are essential compartments in all contemporary eukaryotes and that mitochondrial endosymbiosis took place before radiation of all extant eukaryotes [5].

Metamonada, originally part of the Archezoa, are now classified as one of the main clades of the eukaryotic "super-group" Excavata [6] and are comprised of microaerophilic or anaerobic unicellular eukaryotes that are often specialized parasites or symbionts. Detailed cell and molecular biological studies, including genome sequencing, have been undertaken only for three parasitic species from two metamonad lineages—*Giardia intestinalis* [7] and *Spironucleus salmonicida* [8] (Fornicata) and *Trichomonas vaginalis* [9] (Parabasalia), which have provided important information regarding the functions of their MROs. The third lineage of metamonads, Preaxostyla, contains the basal paraphyletic free-living trimastigids and the derived endobiotic oxymonads [10]. The presence of mitochondrial homologs has been convincingly demonstrated in *Paratrimastix* (formerly *Trimastix*) *pyriformis*, although the biochemical functions of these organelles are largely unknown [11]. Endobiotic oxymonads belong to the least-studied former Archezoa. Here, we describe the first complete genome sequence analysis of an oxymonad, *Monocercomonoides* sp. PA203. We find that although this organism is a standard eukaryotic cell in other respects, it completely lacks any traces of a mitochondrion.

RESULTS AND DISCUSSION

Genome Characteristics

Using the 454 whole-genome shotgun sequencing methodology, we generated a draft genome sequence of the oxymonad *Monocercomonoides* sp. PA203, assembled into 2,095 scaffolds at ~35× coverage (see Experimental Procedures). The estimated size of the genome (~75 Mb) and the number of predicted protein-coding genes (16,629) is intermediate between what is found in diplomonads and *T. vaginalis* (Table 1). Almost 67% of predicted protein-coding genes contain introns (~1.9 introns per gene on average; Table 1). The assembly contains genes encoding tRNAs for all 20 amino acids, and ~50 ribosomal

Table 1. Overview of Metamonada Genomes

Taxa	Size (Mbp)	Guanine-Cytosine Content (%)	Protein-Coding Loci	Repetitive Regions	No. of Introns
Monocercomonoides sp. PA 203	~75	36.8	16,629	~38%	32,328
Trichomonas vaginalis isolate G3 [9]	~160	32.7	~60,000	~65%	65
Giardia intestinalis WB-C6 [7]	~11.7	49	6,480	9%	4
Spironucleus salmonicida ATCC 50377 [8]	12.9	33.4	8,076	5.2%	3

See also Tables S1 and S3.

DNA units were identified on small contigs outside the main assembly (see Supplemental Experimental Procedures). To estimate completeness of the genome sequence, we performed transcriptome mapping, in which 96.9% of transcripts mapped to the genome (see Supplemental Experimental Procedures), and checked the representation of core eukaryotic genes. Using the Core Eukaryotic Genes Mapping Approach (CEGMA) [12], we recovered 63.3% of core eukaryotic genes, a greater fraction than in the *G. intestinalis* genome (46.6%). However, when we excluded genes encoding mitochondrial proteins from the CEGMA dataset and used manually curated *Monocercomonoides* sp. gene models, the percentage of recovered genes increased to 90% (Table S1). For another set of 163 conserved eukaryotic genes used for phylogenomic analyses, the percentage of recovered genes exceeded 95% (Table S2). As the last measure of completeness, we identified 77 out of 78 conserved families of cytosolic eukaryotic ribosomal proteins [13] (Table S3), with the single exception of L41e, which is very short, difficult to detect, and has not been identified in other Metamonada genomes. Phylogenomic analysis (Figure 1) confirmed the relationship of *Monocercomonoides* sp. to *P. pyriformis* and other Metamonada and demonstrated that the *Monocercomonoides* lineage forms a much shorter branch relative to parabasalids and diplomonads. All these measures suggest that the assembled *Monocercomonoides* sp. genome sequence is nearly complete and its encoded proteins are, on average, less divergent than those of *G. intestinalis* and *T. vaginalis*.

With the first oxymonad genome sequence in hand, we focused our attention on one of the most puzzling aspects of their biology—the elusive nature of their mitochondrion.

Absence of Mitochondrial Proteins

No genes that are typically encoded on mitochondrial genomes (mtDNA) of other eukaryotes were found among the assembled scaffolds, suggesting that, like other metamonads, *Monocercomonoides* sp. lacks mtDNA. Next, we searched for homologs of nuclear genome-encoded proteins typically associated with mitochondria or MROs in other eukaryotes. The homologous core of the protein import machinery is regarded as strong evidence for the common origin of all mitochondria [14, 15]. As such, the presence of components of the translocases of the outer membrane (TOM) and inner membrane (TIM), sorting and assembly machinery (SAM) complex, and mitochondrial molecular chaperones (Hsp70 and Cpn60) in hydrogenosomes, mitosomes, and other MROs demonstrates that these organelles are related to mitochondria [16, 17]. While we were able to identify homologs of cytosolic chaperonins in the *Monocercomonoides* sp. genomic sequence, we were unable to identify homologs of any component of the mitochondrial import machinery (Figure 2A; Experimental Procedures; Tables S3 and S4).

All MROs, with the exception of the *G. intestinalis* mitosome [18], are known to export or import ATP and other metabolites typically using transporters from the mitochondrial carrier family (MCF) or, in mitosomes of the microsporidian *Encephalitozoon cuniculi* [19], by the bacterial-type (NTT-like) nucleotide transporters. We did not identify in the *Monocercomonoides* sp. genome any homologs of genes encoding known mitochondrial metabolite transport proteins (Figure 2A; Table S4).

Fe-S clusters are essential biological cofactors associated with many different proteins and are therefore synthesized de novo in every organism across the tree of life [20]. In eukaryotes, this is done mostly by the mitochondrial ISC assembly system and the cytosolic iron-sulfur assembly (CIA) system [21]. Analyses of the *Monocercomonoides* sp. genome revealed the presence of a CIA system but a complete lack of components of the ISC system (Figure 2A; Table S3; Experimental Procedures).

We could not identify either of two possible enzymes involved in the synthesis of cardiolipin, a phospholipid specific for energy-transducing membranes [22]. The majority of eukaryotes synthesize cardiolipins, and the process is localized to mitochondria, but a complete lack of cardiolipin has been experimentally shown for *G. intestinalis*, *T. vaginalis*, and *E. cuniculi* [22]. Furthermore, we could not identify any component of the endoplasmic reticulum (ER)-mitochondria encounter structure (ERMES; Figure 2A) [23].

We identified only two orthologs of the set of proteins predicted to localize to the mitochondrion-related compartment of the closely related *P. pyriformis* [11]: aspartate/ornithine carbamoyltransferase family protein and pyridine nucleotide transhydrogenase. Neither protein has an exclusively mitochondrial localization in eukaryotes [24, 25], and the *Monocercomonoides* sp. orthologs do not contain predicted mitochondrial targeting sequences.

To complement the targeted homology-based searches, we also performed an extensive search for putative homologs of known mitochondrial proteins using a pipeline based on the Mitominer database [26], which was enriched with identified mitochondrial proteins of diverse anaerobic eukaryotes with MROs (Experimental Procedures). The search recovered 76 *Monocercomonoides* sp. proteins as candidates for functions in a putative mitochondrion (Figure 2B; Table S5). Similarly to *G. intestinalis*, *T. vaginalis*, and *E. histolytica*, used as controls, the selected candidates were mainly proteins that are obviously not mitochondrial (e.g., histones) or for which the annotation is too general (e.g. "kinase domain-containing protein"), indicating that the specificity of the pipeline in organisms with

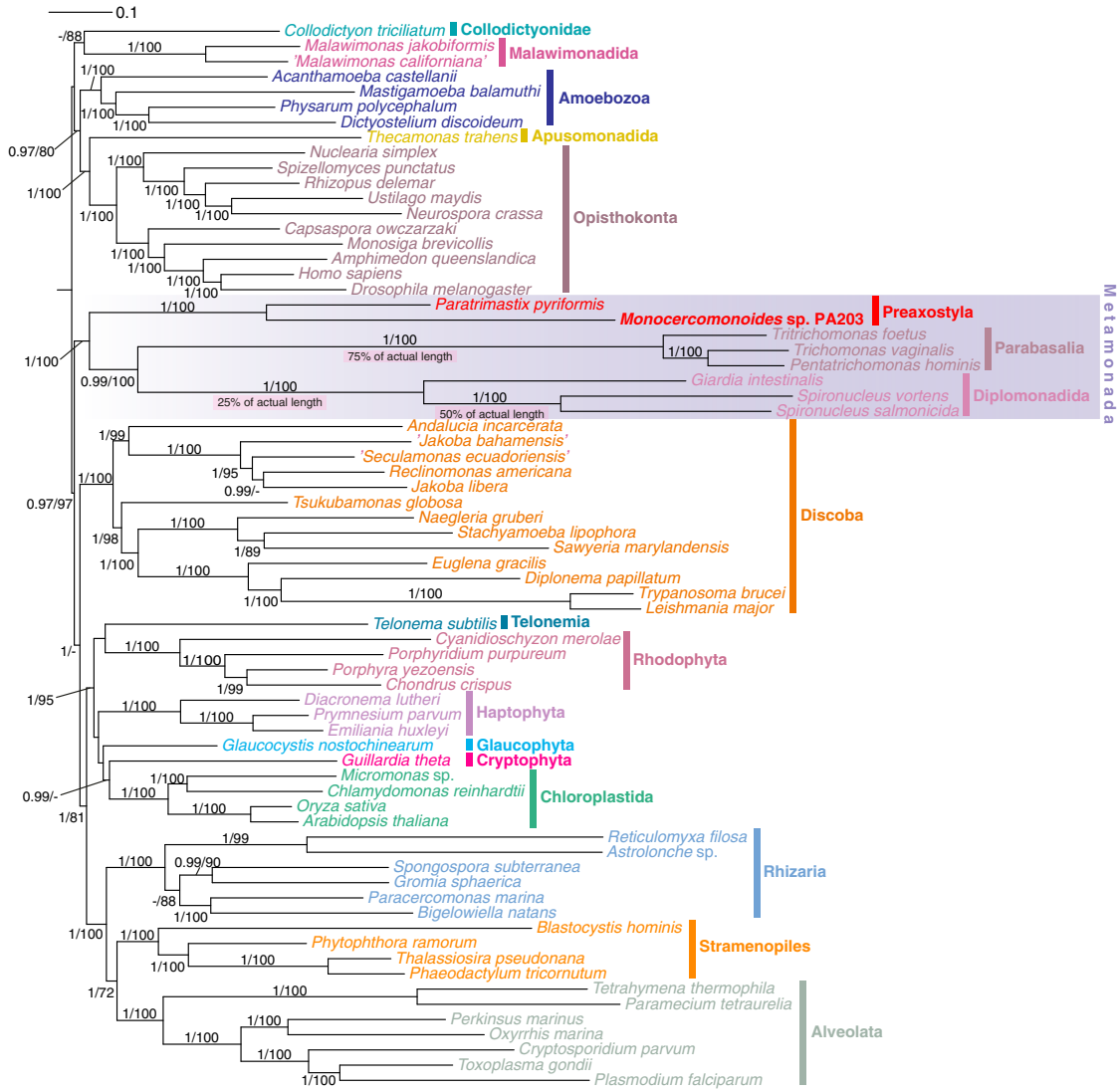


Figure 1. Unrooted Phylogeny of Eukaryotes Inferred from a 163-Protein Supermatrix

The tree displayed was inferred using PhyloBayes (CAT + Poisson substitution model). A maximum-likelihood (ML) tree inferred from the same supermatrix using RAxML (not shown) was very similar to the PhyloBayes tree, with the topological differences in the poorly resolved area comprising Chloroplastida, Cryptophyta, Glaucophyta, and Haptophyta, and in the position of Metamonada, in the ML tree placed sister (with strong bootstrap support) to Discoba. The branch support values shown are posterior probabilities (>0.95) from the PhyloBayes analysis and bootstrap values (>50%) from the ML analysis. Three branches are shown shortened to the indicated percentage of their actual length to fit them on the page. See also Table S2.

divergent mitochondrion is low. However, unlike all other control organisms, in which the search always recovered at least a few mitochondrial hallmark proteins, the set of 76 *Monocercomonoides* sp. candidates did not contain any such proteins. Only 11 of the *Monocercomonoides* candidates fall in the GO category "metabolism," but they do not assemble any obvious metabolic pathway. In summary, this approach (Table S5) failed to reveal any credible set of mitochondrial protein in *Monocercomonoides* sp.

As an alternative to homology searches, we have also attempted to identify mitochondrial proteins by searching for several types of signature sequences. The matrix proteins of mitochondria and MROs are expected to contain conserved N-terminal targeting signals needed for the targeted import into MROs [14]. We performed in silico prediction of mitochondrial targeting signals in the predicted *Monocercomonoides* sp. proteome and identified 107 candidate proteins (Figure 2A; Experimental Procedures; Table S6A). The presence of a predicted targeting signal by itself does not prove the targeting, as such amino acid sequences can also appear at random [27]. Functional annotation revealed that a majority of proteins recovered by this search fall into the Kyoto Encyclopedia of Genes and Genomes (KEGG) category "genetic information processing." Given the absence of a mitochondrial genome, or organellar translation machinery, it is unlikely that these proteins function in an MRO. Only eight candidates were assigned to the KEGG category "metabolism," and they are part of several different metabolic pathways. Finally, only three proteins were predicted to have a mitochondrial

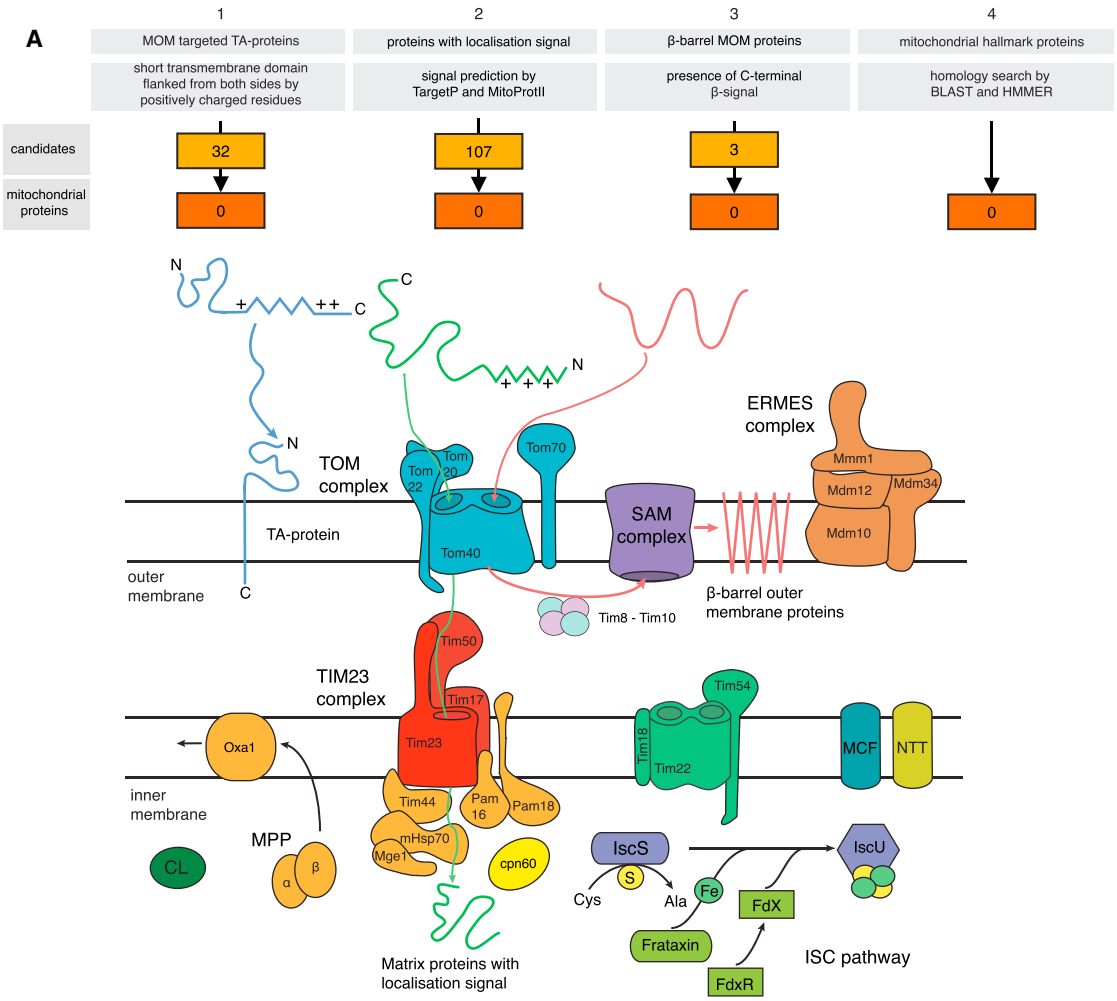


Figure 2. Search Strategies for Proteins Functionally Related to the Mitochondrion in *Monocercomonoides*

(A) Search strategies for mitochondrial proteins and for protein-localization signatures in a canonical eukaryotic cell (details are given in Supplemental Experimental Procedures): (1) mitochondrial outer membrane (MOM)-targeted tail-anchored (TA) proteins (Table S6B), (2) proteins with a mitochondrial targeting signal (Table S6A), (3) β-barrel MOM proteins, (4) 41 mitochondrial hallmarks proteins (Table S4), components of TOM and TIM translocases, cpn60, ERMES complex, ISC pathway components, cardiolipin synthase (CL).

(B) Semiautomatic pipeline for retrieving homologs of mitochondrial proteins from proteomes. We used a custom database for homology searching of mitochondrial proteins in the predicted proteomes of *Monocercomonoides* sp., *Giardia intestinalis*, *Entamoeba histolytica*, *Trichomonas vaginalis*, *Blastocystis* sp. subtype 7, and *Saccharomyces cerevisiae* (Table S5).

See also Tables S4, S5, and S6.

targeting signal and homology to a Mitominer protein (hydrolase-like family protein MONOS_10795, cytosolic TCP-1/cpn60 chaperonin family protein MONOS_13132, and ribonuclease Z MONOS_6181). This also suggests that both pipelines failed to recover specific sets of mitochondrial proteins but instead detected only low-specificity "noise."

The outer mitochondrial membranes accommodate two special classes of proteins, β-barrel and tail-anchored (TA) proteins, which are devoid of the N-terminal targeting signals and instead use specific C-terminal signals [28, 29]. We have identified 32 candidates for TA proteins in the predicted proteome, several of which appeared to be ER-targeted proteins. None of these had the hallmark characteristics of proteins targeted to the mitochondrial outer membrane (Figure 2A; Experimental Procedures; Table S6B). We also failed to identify any credible candidates for β-barrel outer membrane proteins (BOMPs) (Figure 2A; Experimental Procedures).

In summary, our comprehensive examination of the *Monocercomonoides* sp. genome based on homology searches and searches for specific N-terminal and C-terminal signals failed to recover proteins typically associated with MROs, including mitochondrial translocases, metabolite transporters and the ISC system for Fe-S cluster synthesis, ERMES, and enzymes responsible for cardiolipin synthesis.

In order to verify that our inability to find any reliable mitochondrial proteins is not caused by possible unprecedented divergence of *Monocercomonoides* sp. proteins or a failure of our methods, we searched for hallmark proteins of another cellular system, so far not observed in *Monocercomonoides* sp.—the Golgi complex. In this case, using homology-based searches, we detected numerous Golgi-associated proteins, including components of the COPI, AP-1, AP-3, AP-4, COG, GARP, TRAPPI, and Retromer complexes and Rab GTPases regulating transport to and from the Golgi (Table S3). This suggests the presence of Golgi-like compartments in oxymonads [30], despite the absence of a cytologically discernible Golgi apparatus.

The specific absence of mitochondria-associated proteins in *Monocercomonoides* sp. implies the legitimate absence of a mitochondrial compartment. If so, then how does the *Monocercomonoides* cell function without this organelle?

Energy Metabolism without a Mitochondrion

In order to compare the metabolism of *Monocercomonoides* sp. with anaerobic protists retaining mitochondrial compartments, we performed manual annotation of proteins of core pathways of energy metabolism normally associated with the presence and function of a MRO. As with many other organisms with secondarily reduced mitochondria, the *Monocercomonoides* sp. genome does not encode any enzymes for aerobic energy generation (e.g., TCA cycle or electron transport chain proteins). We did identify a complete set of glycolytic enzymes, including the alternative enzymes for anaerobic glycolysis [31], as well as the anaerobic fermentation enzymes pyruvate:ferredoxin oxidoreductase (PFOR) and [FeFe]-hydrogenases (Table S3). [FeFe]-hydrogenase maturases were absent, which is not unprecedented as they are also missing from *G. intestinalis* and *E. histolytica*, anaerobic parasites that are both capable of cytosolic H_2 production [32, 33]. Neither PFOR nor [FeFe]-hydrogenase has a predicted mitochondrial targeting sequence, and heterologous expression in *T. vaginalis* suggests a cytosolic localization of PFOR (Figure S1). In summary, *Monocercomonoides* sp. glucose metabolism appears to produce ATP via substrate-level phosphorylation steps in an extended glycolysis pathway, and the reduced co-factors are re-oxidized by fermentation, ultimately producing acetate and ethanol, or by [FeFe]-hydrogenase producing hydrogen gas. The situation in *Monocercomonoides* sp. is virtually identical to *G. intestinalis* and *E. histolytica*, which independently reduced their mitochondria to mitosomes and all the ATP production occurs in the cytosol [34–36].

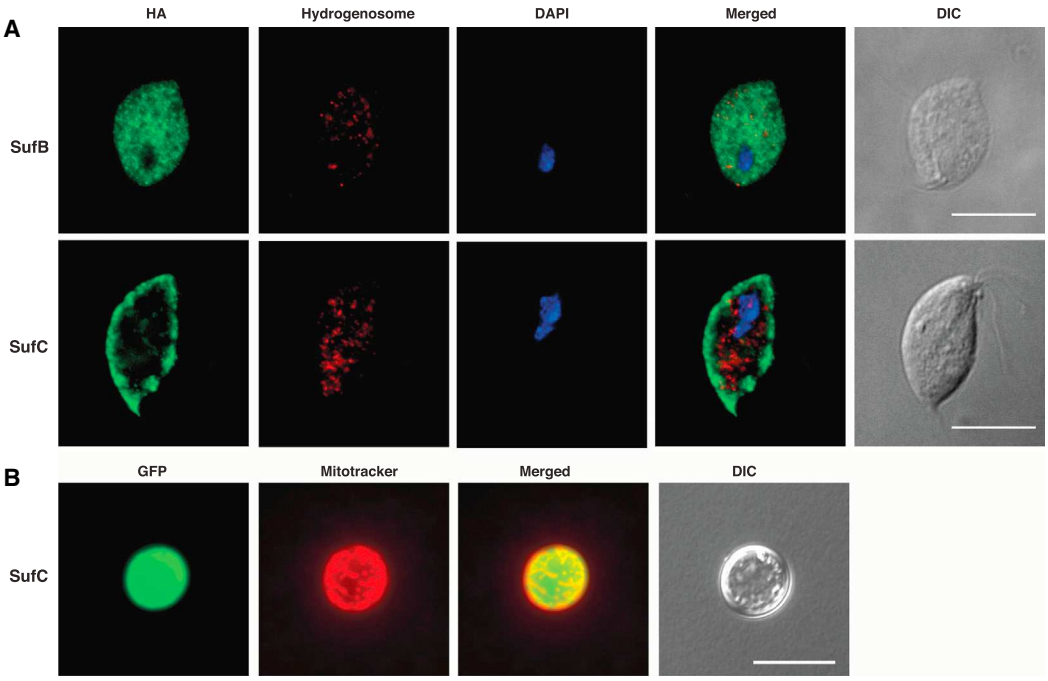
In addition to extended glycolysis, *Monocercomonoides* sp. contains a complete set of three genes for enzymes involved in arginine deiminase pathway—arginine deiminase, ornithine carbamoyltransferase, and carbamate kinase. This pathway may also be used for ATP production by arginine degradation as in *T. vaginalis* and *G. intestinalis* [37, 38]. In *G. intestinalis*, this pathway produces eight times more ATP than sugar metabolism.

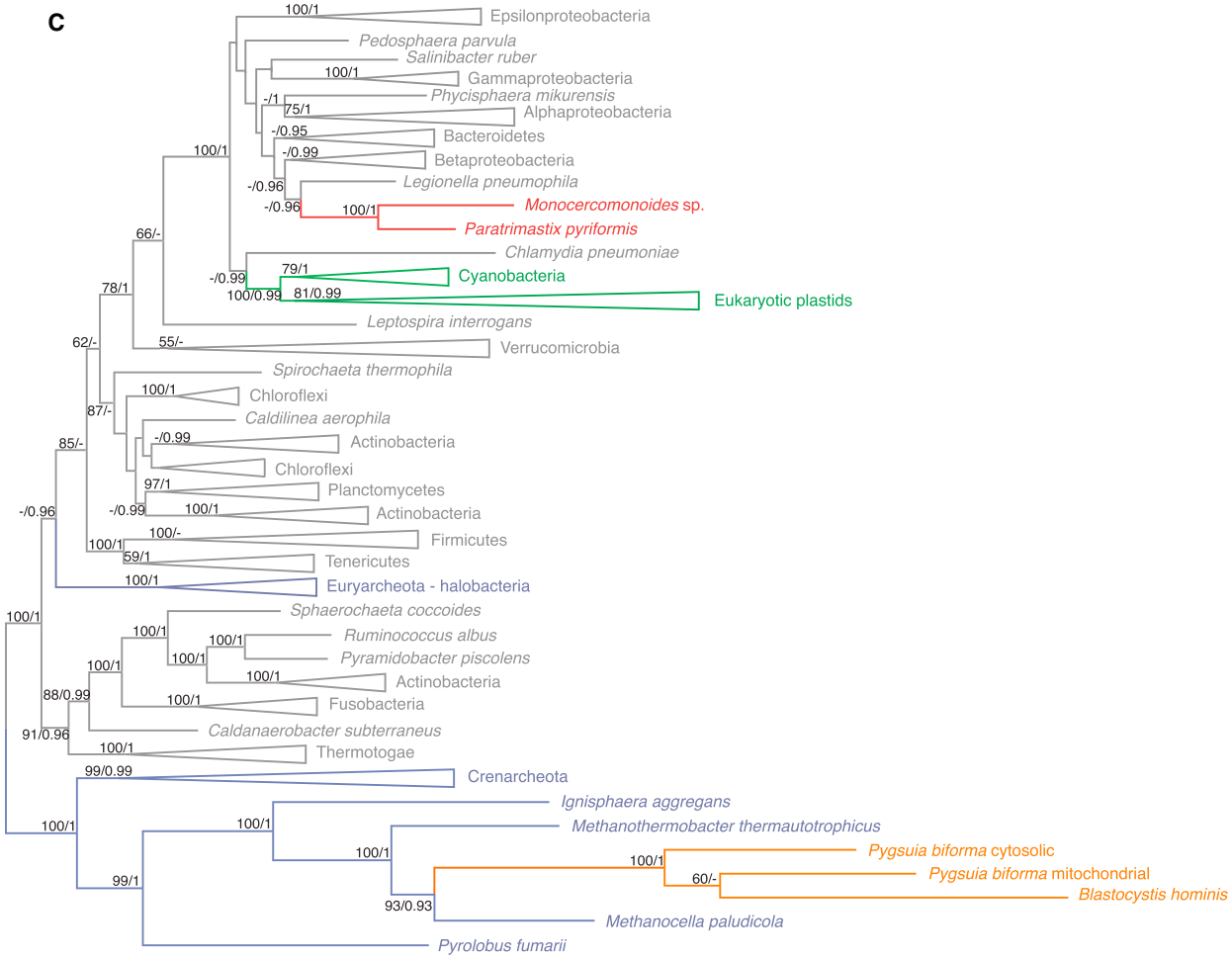
Fe-S Cluster Assembly without a Mitochondrion

Every eukaryotic cell contains a CIA machinery, which assists the final stages of the assembly of Fe-S clusters in proteins functioning in the eukaryotic cytosol and nucleus. Eight proteins were shown to be involved in the CIA pathway in yeast and humans: Cfd1, NUBP1 (Nbp35), NARFL (Nar1), CIAO1 (Cia1), Dre2, Tah18, Cia2, and MMS19. Four of them (i.e., Nbp35, Nar1, Cia1, and Cia2) [21] are conserved among eukaryotes and also present in the *Monocercomonoides* sp. genome (Table S3). We did not identify Cfd1 and MMS19, which are missing from many other eukaryotes, and Dre2 and Tah18, which are missing from the anaerobic protists containing MROs (including *E. histolytica*, *Mastigamoeba balamuthi*, *T. vaginalis*, *G. intestinalis*, and *Blastocystis* sp.) [21].

Despite the presence of the CIA pathway, it is commonly suggested that mitochondria and related organelles are essential to eukaryotic cells because the mitochondrial ISC system plays a critical role in the initial phase of the formation of cytosolic Fe-S clusters [20]. Although the ISC system is a near-universally conserved pathway in eukaryotes and seems to be the unifying feature of mitochondria and related organelles, genes encoding proteins of the mitochondrial ISC pathway have not been detected in the *Monocercomonoides* sp. genome. The functional replacement of the ISC system has been reported for only two lineages, *Pygsuia biforma* (Breviatea) and Archamoebae. A methanoarcheal sulfur mobilization (SUF) system [39] or a bacterial nitrogen fixation (NIF) [40] has apparently replaced the ISC system in the *P. biforma* and the Archamoebae lineages, respectively. Conflicting data exist on the localization of the NIF system in *E. histolytica* [41, 42]; however, in *M. balamuthi*, the NIF system localizes in the cytosol and the MRO [43].

The major issue remains: how does *Monocercomonoides* sp. form Fe-S clusters? Unexpectedly, we identified genes encoding four subunits of the SUF system: SufB, SufC, and fused SufS and SufU (Table S3). SufS is a "two-component" cysteine desulfurase, and its activity might be enhanced by SufE or SufU [44, 45]. In *Monocercomonoides* sp., SufS is fused with SufU, which is a unique feature. SufB and SufC can form a scaffold complex in prokaryotes, and SufB2C2 complex is capable of binding and transferring 4Fe-4S clusters to a recipient apoprotein [46]. All identified SUF system proteins apparently retain all important catalytic sites (Figure S2) and may perform de novo Fe-S clusters biogenesis by themselves or in concert with the CIA machinery. The SUF system for Fe-S cluster synthesis is found in plastids, bacteria, and archaea and has also been found in two microbial eukaryotes *P. biforma* [39] and *Blastocystis* sp. [47]. The





(legend on next page)

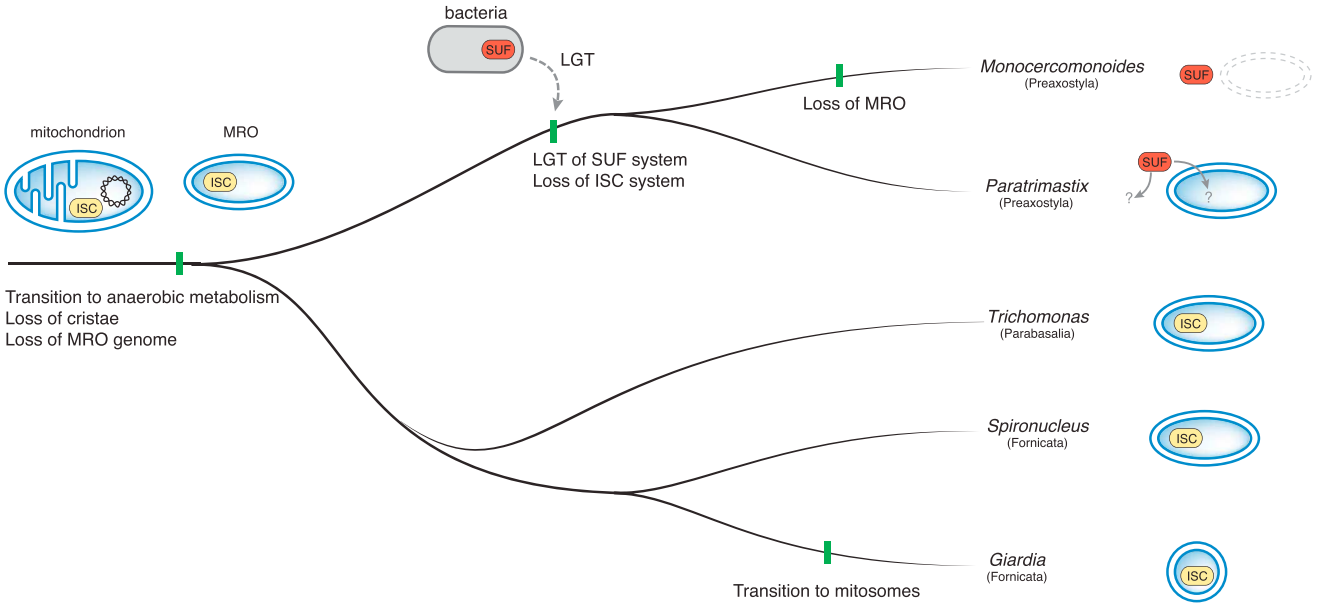


Figure 4. Reductive Evolution of Mitochondria in Metamonads

Transition to an anaerobic lifestyle occurred in a common ancestor of metamonads and was followed by reduction of mitochondria to MROs, accompanied by the loss of cristae and genome, and the transition to anaerobic metabolism. The ISC pathway for Fe-S cluster synthesis was present in a metamonad common ancestor. Further reduction to a mitosome took place in the *Giardia intestinalis* lineage. We propose that in the common ancestor of *Paratrimastix pyriformis* and *Monocercomonoides*, a SUF system acquired through LGT from bacteria substituted the MRO-localized ISC system. Subsequently, the MRO was lost completely in the lineage leading to *Monocercomonoides* sp. Localization of the SUF pathway in *P. pyriformis* is unknown.

presence of spliceosomal introns in the putative SufC and SufSU of *Monocercomonoides* confirms that these proteins are not prokaryotic contamination. Furthermore, fluorescence in situ hybridization (FISH) with *sufB* and *sufC* gene probes demonstrated their presence in the *Monocercomonoides* sp. nucleus (Figure S3). Importantly, homologs of these proteins were detected in the *P. pyriformis* genome, the closest sequenced relative to *Monocercomonoides*. The SUF system components of both *Monocercomonoides* sp. and *P. pyriformis* do not contain recognizable mitochondrial targeting signals, and our experiments with heterologous expression of *Monocercomonoides* sp. SufB and SufC proteins in *T. vaginalis* (Figure 3A) and SufC protein in yeast (Figure 3B) support a cytosolic localization. Phylogenetic analyses indicate that this SUF system was acquired by an ancestor of *Monocercomonoides* and *Paratrimastix* by lateral gene transfer (LGT) from bacteria independently of all other SUF-containing eukaryotes (Figure 3C). We propose that the acquisition of a cytosolic SUF system made the ancestral ISC system in the mitochondrion dispensable, which led to its loss and, in the *Monocercomonoides* lineage, to the complete loss of MROs (Figure 4).

Conclusions

Mitochondria and related organelles are currently considered to be indispensable components of eukaryotic cells. The genome sequence of *Monocercomonoides* sp. reported here suggests that this is not the case. Despite extensive searches, no mitochondrial marker proteins such as membrane protein translocases and metabolite transporters were identified. Crucially, the mitochondrion-specific ISC pathway for Fe-S cluster biogenesis is absent and apparently was replaced by a bacterial SUF system that functions in the cytosol. On the other hand, genes encoding other features once thought to be absent from these divergent eukaryotic cells, i.e., the Golgi body, were readily identifiable. The genome also contains genes for essential cytosolic pathways of energy metabolism, although we did observe examples of metabolic streamlining characteristic of other anaerobic or microaerophilic eukaryotes.

Figure 3. Heterologous Expression of *Monocercomonoides* sp. SUF System Proteins and Phylogeny of Concatenated SufB, SufC, and SufS Homologs

(A) Heterologous expression of *Monocercomonoides* sp. SufB and SufC proteins in *Trichomonas vaginalis*. *Monocercomonoides* sp. proteins with a C-terminal HA tag were expressed in *T. vaginalis* and visualized by an anti-HA antibody (green). The signal of the anti-HA antibody does not co-localize with hydrogenosomes stained using an anti-malic enzyme antibody (red). The nucleus was stained using DAPI (blue). Scale bar, 10 μm.

(B) Heterologous expression of *Monocercomonoides* sp. SufC protein in *Saccharomyces cerevisiae*. *Monocercomonoides* sp. proteins tagged with GFP were expressed in *S. cerevisiae* (green). The GFP signal does not co-localize with the yeast mitochondria stained by Mitotracker (red). Scale bar, 10 μm.

(C) Unrooted ML tree of concatenated SufB, SufC, and SufS sequences. Bootstrap support values above 50 and posterior probabilities greater than 0.75 are shown. *Monocercomonoides* sp. and *Paratrimastix pyriformis* are shown in red, eukaryotic plastids and cyanobacteria in green, *Blastocystis* sp. and *Pygsuia biforma* in orange, bacteria in gray, and archaea in blue.

See also Figures S1–S3.

Reduction of mitochondria is known from various eukaryotic lineages adapted to anaerobic lifestyle [48]. Mitosomes in *Giardia*, *Entamoeba*, and Microsporidia represent the most extreme cases of mitochondrial reduction known to date, and yet they still contain recognizable mitochondrial protein translocases and usually an ISC system. The specific absence of all these mitochondrial proteins in the genome of *Monocercomonoides* sp. indicates that this eukaryote has dispensed with the mitochondrial compartment completely. In principle, we cannot exclude the possibility that a mitochondrion exists in *Monocercomonoides* sp. whose protein composition has been altered entirely. However, such a hypothetical organelle could not be recognized as a mitochondrion homolog by any available means. Without any positive evidence for the latter scenario, we suggest that the complete absence of mitochondrial markers and pathways points to the bona fide absence of the organelle. Because all known oxymonads are obligate animal symbionts, and mitochondrial homologs are present in the close free-living sister lineage *Paratrimastix*, the absence of mitochondrion in *Monocercomonoides* sp. must be secondary. We hypothesize that the acquisition of the SUF system predated the loss of the mitochondrial ISC system in the common ancestor of Preaxostyla and allowed for the complete loss of the organelle in *Monocercomonoides* sp. lineage, the first known truly secondarily amitochondriate eukaryote.

EXPERIMENTAL PROCEDURES

Genome and Transcriptome Sequencing

All experiments were performed on the *Monocercomonoides* sp. PA203 strain. The culture (2 L with a cell density of approximately 4 × 10^5 cells/mL) was filtered to remove most of the bacteria before isolation of DNA (culturing and filtration details in Supplemental Experimental Procedures). DNA was isolated using DNeasy Blood and Tissue Kit (QIAGEN). Total genomic DNA was sequenced using a Genome Sequencer 454 GS FLX+ with XL+ reagents. A total of seven sequencing runs were performed, including four shotgun runs on libraries with the average fragment length of 500 to 800 and three runs on a 3-kb paired-end library. Two RNA sequencing (RNA-seq) experiments were performed using 454 and Illumina sequencing platforms. Details of sequencing are given in Supplemental Experimental Procedures.

Roche's assembler Newbler v.2.6 was used to generate a genome sequence assembly from 454 single and pair end reads. The final assembly consisted of 2,095 scaffolds spanning almost 75 Mb of the genome. The N50 scaffold size is 71.4 kb. Transcriptome assembly of the 454 data was performed by Newbler v.2.8 with default parameters, and Illumina-generated transcriptomic data were assembled using Trinity [49] (details in Supplemental Experimental Procedures). The CEGMA [12] was used to estimate the number of conserved eukaryotic genes in the *Monocercomonoides* sp. genome assembly (Table S1) and presence of cytosolic ribosomal eukaryotic proteins as an additional measure of completeness (Table S3).

Genome Annotation and Gene Searching

For the structural annotation, Augustus v.2.7 [50, 51], PASA2 [52], and EVM [53] were used. Gene models of particular interest were manually evaluated with the help of RNA-seq data or considering conservation with homologs (details in Supplemental Experimental Procedures).

Functional annotation was assigned to genes by similarity searches of predicted proteins using BLASTP [54] against the NCBI non-redundant protein database [55] and HMMER3 [56] searches of domain hits in the Pfam protein families database [57]. Additional annotation was performed using the KEGG automatic annotation server [58]. Annotation files are available at the web page http://www.protistologie.cz/hampllab/data.html.

tRNA genes were predicted with tRNAscan-SE [59]; rDNA sequences were not present in the original main assembly, but they were identified in contigs not assembled into scaffolds and added to the main assembly.

The *Monocercomonoides* sp. genome database was searched using the TBLASTN [54] algorithm, and *Monocercomonoides* proteome database and six-frame translation of the genomic sequence were searched using the BLASTP [54] algorithm or the profile hidden Markov model (HMM) searching method *phmmer* from the HMMER3 [56] package. We used a wide range of queries described in Supplemental Experimental Procedures.

Phylogenetic Analyses

We performed a number of maximum-likelihood and Bayesian phylogenetic analyses: (1) phylogenomic analyses of eukaryotes based on 163 genes and 70 taxa; (2) phylogenetic analyses of genes for SUF pathway enzymes; and (3) individual gene trees to support functional annotation of genes (details in Supplemental Experimental Procedures).

Subcellular Localization Prediction

Subcellular localization prediction for the *Monocercomonoides* sp. proteome was performed using TargetP v.1.1 [60] and MitoProt II v.1.101 [61]. TA proteins were identified and analyzed based on presence of a transmembrane domain (TMD) of moderate hydrophobicity flanked by positively charged residues [29, 62] (details in Supplemental Experimental Procedures). BOMPs were identified based on the presence of a conserved C-terminal β-signal, using a previously described pipeline [63].

Mitochondrial Protein Searching Using a Mitominer-Based Database

We prepared a custom database of mitochondrial proteins to search for genes encoding proteins with putative mitochondrial localization. The custom database was based on the MitoMiner database [26] reference set containing 12,925 proteins from 11 eukaryotic mitochondrial proteomes, which was enriched by known or predicted MRO-localized proteins of *E. histolytica*, *G. intestinalis*, *P. biforma*, *S. salmonicida*, *T. vaginalis*, and *P. pyriformis*. Homologs of proteins from this database were searched in the predicted proteome of *Monocercomonoides* sp. and in the predicted proteomes of *Blastocystis* sp., *E. histolytica*, *G. intestinalis*, *S. cerevisiae*, and *T. vaginalis*, which were used as control datasets. While searching the control datasets, the proteins of the searched organism were removed from the custom database. In the last step, only those candidates were kept whose first hit in the NCBI database [55] contained a predictable mitochondrial targeting signal (score > 0.5 in TargetP v.1.1 [60] and MitoProt II v.1.101 [61]). Further details are given in Supplemental Experimental Procedures.

FISH

We performed FISH experiments with labeled probes to determine whether the genes for SUF system proteins physically reside in the *Monocercomonoides* sp. genome or represent bacterial contamination. Details on preparation of labeled probes are given in Supplemental Experimental Procedures.

One liter of *Monocercomonoides* sp. culture was filtered to remove bacteria, and the cells were pelleted by centrifugation for 10 min at 2,000 × *g* at 4°C. FISH with digoxigenin-labeled probes was performed according to a previously described procedure [64] omitting the colchicine procedure. Cell nuclei and the probes were denatured under a coverslip in a single step in 50 μL of 50% formamide in 2 × SSC at 70°C for 5 min. Preparations were observed using an IX81 microscope (Olympus) equipped with an IX2-UCB camera. Images were processed using Cell software (Olympus) and ImageJ 1.42q.

Heterologous Protein Expression and Microscopy in *Trichomonas vaginalis*

The *T. vaginalis* transfection system was used to assess subcellular localization of SufB, SufC, and PFOR proteins. *Monocercomonoides* sp. cDNA preparation was performed as described for transcriptome sequencing (Supplemental Experimental Procedure). Constructs with the hemagglutinin (HA) tag fused to the 3′ end of the coding sequences of the studied genes were prepared and expressed in *T. vaginalis*, an anaerobic protist related to *Monocercomonoides* sp. and bearing a hydrogenosome (details are given in Supplemental Experimental Procedures). *Monocercomonoides* sp. proteins

expressed in *T. vaginalis* cells were visualized using standard techniques [14] (details are given in Supplemental Experimental Procedures).

Saccharomyces cerevisiae Heterologous Expression System

This expression system was used to confirm the results from the *T. vaginalis* expression system for SufC protein. The procedure was analogous to the one described in [11]. Details are given in Supplemental Experimental Procedures.

ACCESSION NUMBERS

Sequence data for the genome reads (experiment number SRX1470187), the 454 transcriptome reads sequenced using the 454 platform (experiment number SRX1453820), and the Illumina transcriptome reads sequenced using the Illumina platform (experiment number SRX1453675) have been deposited to the NCBI Sequence Read Archive under accession number SRA: SRP066769. The accession number for the *Monocercomonoides* sp. PA203 genome reported in this paper is GenBank: LSRY00000000. The accession number for the 454 transcriptome project reported in this paper is GenBank: GEEG00000000. The accession number for the Illumina transcriptome project reported in this paper is GenBank: GEEL00000000. The versions described in this paper are versions LSRY01000000, GEEG01000000, and GEEL01000000. Further additional information on the genome analysis can be found at http://www.protistologie.cz/hampllab/data.html.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2016.03.053.

AUTHOR CONTRIBUTIONS

The project was conceived in the laboratory of V.H. with the contribution of J.B.D. Genome and 454 transcriptome sequencing was performed by the Laboratory of Genomics and Bioinformatics. A.K. and V.H. coordinated the project. Z.Z. isolated genomic DNA. V.V. and V.H. isolated RNA. M.H. prepared sequencing libraries. Č.V. and A.K. assembled data. A.K. curated data, analyzed genomic and transcriptomic data, and conducted gene prediction and automatic functional annotation. A.K., V.V., S.C.T., R.P., L.N., V.Ž., L.D.B., E.K.H., M.E., and V.H. performed manual annotation. A.K., V.V., P.D., C.W.S., and V.H. performed mitochondrial gene searching. Z.Z. performed FISH experiments. V.V., Z.Z., and S.C.T. performed immunolocalization experiments. A.K., V.V., R.P., L.D.B., E.K.H., P.S., and L.E. performed phylogenetic analyses. A.K., V.V., Z.Z., S.C.T., and R.P. prepared figures. A.K. and V.H. wrote the manuscript in collaboration with A.J.R., M.E., and J.B.D., and all authors edited and approved the manuscript.

ACKNOWLEDGMENTS

A.K. and V.H. were supported by the Ministry of Education, Youth and Sports of CR within the National Sustainability Program II (Project BIOCEV-FAR) LQ1604 and by the project "BIOCEV" (CZ.1.05/1.1.00/02.0109). V.H. and sequencing were supported by Czech Science foundation project P506-12-1010. Z.Z. and localization experiments were funded by Czech Science foundation project 510 13-22333P. E.K.H. was supported by a Vanier Canada Graduate Scholarship and an Alberta Innovates – Health Solutions Graduate Studentship. The work of L.E., C.W.S., and A.J.R. was supported by a regional partnerships program grant (62809) from the Canadian Institute of Health Research and the Nova Scotia Health Research Foundation. The work of L.D.B., E.K.H., and J.B.D. was supported by an NSERC Discovery grant and an Alberta Innovates Technology Futures New Faculty Award to J.B.D. R.P. and M.E. were supported by Czech Science foundation project 15-16406S.

Received: December 23, 2015
Revised: March 5, 2016
Accepted: March 23, 2016
Published: May 12, 2016

REFERENCES

1. Huynen, M.A., Duarte, I., and Szklarczyk, R. (2013). Loss, replacement and gain of proteins at the origin of the mitochondria. Biochim. Biophys. Acta *1827*, 224–231.
2. Tovar, J., León-Avila, G., Sánchez, L.B., Sutak, R., Tachezy, J., van der Giezen, M., Hernández, M., Müller, M., and Lucocq, J.M. (2003). Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. Nature *426*, 172–176.
3. Lindmark, D.G., and Müller, M. (1973). Hydrogenosome, a cytoplasmic organelle of the anaerobic flagellate Tritrichomonas foetus, and its role in pyruvate metabolism. J. Biol. Chem. *248*, 7724–7728.
4. Cavalier-Smith, T. (1987). Eukaryotes with no mitochondria. Nature *326*, 332–333.
5. Gray, M.W. (2012). Mitochondrial evolution. Cold Spring Harb. Perspect. Biol. *4*, a011403.
6. Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., et al. (2012). The revised classification of eukaryotes. J. Eukaryot. Microbiol. *59*, 429–493.
7. Morrison, H.G., McArthur, A.G., Gillin, F.D., Aley, S.B., Adam, R.D., Olsen, G.J., Best, A.A., Cande, W.Z., Chen, F., Cipriano, M.J., et al. (2007). Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. Science *317*, 1921–1926.
8. Xu, F., Jerlström-Hultqvist, J., Einarsson, E., Astvaldsson, A., Svärd, S.G., and Andersson, J.O. (2014). The genome of Spironucleus salmonicida highlights a fish pathogen adapted to fluctuating environments. PLoS Genet. *10*, e1004053.
9. Carlton, J.M., Hirt, R.P., Silva, J.C., Delcher, A.L., Schatz, M., Zhao, Q., Wortman, J.R., Bidwell, S.L., Alsmark, U.C.M., Besteiro, S., et al. (2007). Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. Science *315*, 207–212.
10. Zhang, Q., Táborský, P., Silberman, J.D., Pánek, T., Čepička, I., and Simpson, A.G.B. (2015). Marine isolates of Trimastix marina form a plesiomorphic deep-branching lineage within Preaxostyla, separate from other known Trimastigids (Paratrimastix n. gen.). Protist *166*, 468–491.
11. Zubáčová, Z., Novák, L., Bublíková, J., Vacek, V., Fousek, J., Rídl, J., Tachezy, J., Doležal, P., Vlček, C., and Hampl, V. (2013). The mitochondrion-like organelle of Trimastix pyriformis contains the complete glycine cleavage system. PLoS ONE *8*, e55417.
12. Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics *23*, 1061–1067.
13. Lecompte, O., Ripp, R., Thierry, J.C., Moras, D., and Poch, O. (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. Nucleic Acids Res. *30*, 5382–5390.
14. Dolezal, P., Likic, V., Tachezy, J., and Lithgow, T. (2006). Evolution of the molecular machines for protein import into mitochondria. Science *313*, 314–318.
15. Zarsky, V., Tachezy, J., and Dolezal, P. (2012). Tom40 is likely common to all mitochondria. Curr. Biol. *22*, R479–R481, author reply R481–R482.
16. Dolezal, P., Smíd, O., Rada, P., Zubácová, Z., Bursać, D., Suták, R., Nebesárová, J., Lithgow, T., and Tachezy, J. (2005). Giardia mitosomes and trichomonad hydrogenosomes share a common mode of protein targeting. Proc. Natl. Acad. Sci. USA *102*, 10924–10929.
17. Burri, L., Williams, B.A.P., Bursac, D., Lithgow, T., and Keeling, P.J. (2006). Microsporidian mitosomes retain elements of the general mitochondrial targeting system. Proc. Natl. Acad. Sci. USA *103*, 15916–15920.
18. Jedelský, P.L., Doležal, P., Rada, P., Pyrih, J., Smíd, O., Hrdý, I., Sedinová, M., Marcinčiková, M., Voleman, L., Perry, A.J., et al. (2011). The minimal proteome in the reduced mitochondrion of the parasitic protist Giardia intestinalis. PLoS ONE *6*, e17285.

19. Tsaousis, A.D., Kunji, E.R.S., Goldberg, A.V., Lucocq, J.M., Hirt, R.P., and Embley, T.M. (2008). A novel route for ATP acquisition by the remnant mitochondria of Encephalitozoon cuniculi. Nature *453*, 553–556.
20. Lill, R. (2009). Function and biogenesis of iron-sulphur proteins. Nature *460*, 831–838.
21. 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.
22. Tian, H.-F., Feng, J.-M., and Wen, J.-F. (2012). The evolution of cardiolipin biosynthesis and maturation pathways and its implications for the evolution of eukaryotes. BMC Evol. Biol. *12*, 32.
23. Wideman, J.G., Gawryluk, R.M.R., Gray, M.W., and Dacks, J.B. (2013). The ancient and widespread nature of the ER-mitochondria encounter structure. Mol. Biol. Evol. *30*, 2044–2049.
24. Yarlett, N., Lindmark, D.G., Goldberg, B., Moharrami, M.A., and Bacchi, C.J. (1994). Subcellular localization of the enzymes of the arginine dihydrolase pathway in Trichomonas vaginalis and Tritrichomonas foetus. J. Eukaryot. Microbiol. *41*, 554–559.
25. Yousuf, M.A., Mi-ichi, F., Nakada-Tsukui, K., and Nozaki, T. (2010). Localization and targeting of an unusual pyridine nucleotide transhydrogenase in Entamoeba histolytica. Eukaryot. Cell *9*, 926–933.
26. Smith, A.C., Blackshaw, J.A., and Robinson, A.J. (2012). MitoMiner: a data warehouse for mitochondrial proteomics data. Nucleic Acids Res. *40*, D1160–D1167.
27. Lucattini, R., Likic, V.A., and Lithgow, T. (2004). Bacterial proteins predisposed for targeting to mitochondria. Mol. Biol. Evol. *21*, 652–658.
28. Denic, V. (2012). A portrait of the GET pathway as a surprisingly complicated young man. Trends Biochem. Sci. *37*, 411–417.
29. Borgese, N., Brambillasca, S., and Colombo, S. (2007). How tails guide tail-anchored proteins to their destinations. Curr. Opin. Cell Biol. *19*, 368–375.
30. Mowbrey, K., and Dacks, J.B. (2009). Evolution and diversity of the Golgi body. FEBS Lett. *583*, 3738–3745.
31. Liapounova, N.A., Hampl, V., Gordon, P.M.K., Sensen, C.W., Gedamu, L., and Dacks, J.B. (2006). Reconstructing the mosaic glycolytic pathway of the anaerobic eukaryote Monocercomonoides. Eukaryot. Cell *5*, 2138–2146.
32. Lloyd, D., Ralphs, J.R., and Harris, J.C. (2002). Giardia intestinalis, a eukaryote without hydrogenosomes, produces hydrogen. Microbiology *148*, 727–733.
33. Nixon, J.E.J., Field, J., McArthur, A.G., Sogin, M.L., Yarlett, N., Loftus, B.J., and Samuelson, J. (2003). Iron-dependent hydrogenases of Entamoeba histolytica and Giardia lamblia: activity of the recombinant entamoebic enzyme and evidence for lateral gene transfer. Biol. Bull. *204*, 1–9.
34. van der Giezen, M., and Tovar, J. (2005). Degenerate mitochondria. EMBO Rep. *6*, 525–530.
35. Müller, M., Mentel, M., van Hellemond, J.J., Henze, K., Woehle, C., Gould, S.B., Yu, R.-Y., van der Giezen, M., 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.
36. Makiuchi, T., and Nozaki, T. (2014). Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa. Biochimie *100*, 3–17.
37. Schofield, P.J., Edwards, M.R., Matthews, J., and Wilson, J.R. (1992). The pathway of arginine catabolism in Giardia intestinalis. Mol. Biochem. Parasitol. *51*, 29–36.
38. Yarlett, N., Martinez, M.P., Moharrami, M.A., and Tachezy, J. (1996). The contribution of the arginine dihydrolase pathway to energy metabolism by Trichomonas vaginalis. Mol. Biochem. Parasitol. *78*, 117–125.
39. 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. Curr. Biol. *24*, 1176–1186.
40. van der Giezen, M., Cox, S., and Tovar, J. (2004). The iron-sulfur cluster assembly genes iscS and iscU of Entamoeba histolytica were acquired by horizontal gene transfer. BMC Evol. Biol. *4*, 7.
41. Maralikova, B., Ali, V., Nakada-Tsukui, K., Nozaki, T., van der Giezen, M., Henze, K., and Tovar, J. (2010). Bacterial-type oxygen detoxification and iron-sulfur cluster assembly in amoebal relict mitochondria. Cell. Microbiol. *12*, 331–342.
42. Mi-ichi, F., Abu Yousuf, M., Nakada-Tsukui, K., and Nozaki, T. (2009). Mitosomes in Entamoeba histolytica contain a sulfate activation pathway. Proc. Natl. Acad. Sci. USA *106*, 21731–21736.
43. Nývltová, E., Šuták, R., Harant, K., Šedinová, M., Hrdy, I., Paces, J., Vlček, Č., and Tachezy, J. (2013). NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of Mastigamoeba balamuthi. Proc. Natl. Acad. Sci. USA *110*, 7371–7376.
44. Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M., and Barras, F. (2003). Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase. J. Biol. Chem. *278*, 38352–38359.
45. Riboldi, G.P., de Oliveira, J.S., and Frazzon, J. (2011). Enterococcus faecalis SufU scaffold protein enhances SufS desulfurase activity by acquiring sulfur from its cysteine-153. Biochim. Biophys. Acta *1814*, 1910–1918.
46. Chahal, H.K., and Outten, F.W. (2012). Separate FeS scaffold and carrier functions for $SufB_2C_2$ and SufA during in vitro maturation of [2Fe2S] Fdx. J. Inorg. Biochem. *116*, 126–134.
47. Tsaousis, A.D., Ollagnier de Choudens, S., 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.
48. Maguire, F., and Richards, T.A. (2014). Organelle evolution: a mosaic of 'mitochondrial' functions. Curr. Biol. *24*, R518–R520.
49. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.
50. Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics *19* (*Suppl 2*), ii215–ii225.
51. Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics *7*, 62.
52. Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. *31*, 5654–5666.
53. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. *9*, R7.
54. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
55. Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. *33*, D501–D504.
56. Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29-37.
57. Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Res. *40*, D290–D301.
58. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. *35*, W182-5.

59. Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964.

60. Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. *2*, 953–971.

61. Claros, M.G., and Vincens, P. (1996). Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur. J. Biochem. *241*, 779–786.

62. Borgese, N., Colombo, S., and Pedrazzini, E. (2003). The tale of tail-anchored proteins: coming from the cytosol and looking for a membrane. J. Cell Biol. *161*, 1013–1019.

63. Imai, K., Fujita, N., Gromiha, M.M., and Horton, P. (2011). Eukaryote-wide sequence analysis of mitochondrial β-barrel outer membrane proteins. BMC Genomics *12*, 79.

64. Zubáčová, Z., Krylov, V., and Tachezy, J. (2011). Fluorescence in situ hybridization (FISH) mapping of single copy genes on Trichomonas vaginalis chromosomes. Mol. Biochem. Parasitol. *176*, 135–137.

Novák et al. BMC Evolutionary Biology (2016) 16:197
DOI 10.1186/s12862-016-0771-4

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Arginine deiminase pathway enzymes: evolutionary history in metamonads and other eukaryotes

Lukáš Novák[1], Zuzana Zubáčová[1], Anna Karnkowska[1,5], Martin Kolisko[2,5], Miluše Hroudová[3], Courtney W. Stairs[2], Alastair G. B. Simpson[4], Patrick J. Keeling[5], Andrew J. Roger[2], Ivan Čepička[6] and Vladimír Hampl[1*]

Abstract

Background: Multiple prokaryotic lineages use the arginine deiminase (ADI) pathway for anaerobic energy production by arginine degradation. The distribution of this pathway among eukaryotes has been thought to be very limited, with only two specialized groups living in low oxygen environments (Parabasalia and Diplomonadida) known to possess the complete set of all three enzymes. We have performed an extensive survey of available sequence data in order to map the distribution of these enzymes among eukaryotes and to reconstruct their phylogenies.

Results: We have found genes for the complete pathway in almost all examined representatives of Metamonada, the anaerobic protist group that includes parabasalids and diplomonads. Phylogenetic analyses indicate the presence of the complete pathway in the last common ancestor of metamonads and heterologous transformation experiments suggest its cytosolic localization in the metamonad ancestor. Outside Metamonada, the complete pathway occurs rarely, nevertheless, it was found in representatives of most major eukaryotic clades.

Conclusions: Phylogenetic relationships of complete pathways are consistent with the presence of the Archaea-derived ADI pathway in the last common ancestor of all eukaryotes, although other evolutionary scenarios remain possible. The presence of the incomplete set of enzymes is relatively common among eukaryotes and it may be related to the fact that these enzymes are involved in other cellular processes, such as the ornithine-urea cycle. Single protein phylogenies suggest that the evolutionary history of all three enzymes has been shaped by frequent gene losses and horizontal transfers, which may sometimes be connected with their diverse roles in cellular metabolism.

Keywords: Arginine deiminase, Ornithine transcarbamylase, Carbamate kinase, Phylogeny, Metamonada, Preaxostyla, Protists

Background

The arginine deiminase pathway (ADI pathway, syn.: arginine dihydrolase pathway) catalyzes a conversion of arginine to ornithine, ammonium, and carbon dioxide, while generating ATP from ADP and phosphate. The enzymes involved in the three steps of the pathway are arginine deiminase (ADI, EC 3.5.3.6), ornithine transcarbamylase (OTC, EC 2.1.3.3), and carbamate kinase (CK, EC 2.7.2.2). The first reaction, catalyzed by ADI, is the deamination of arginine to yield citrulline and NH_4^+. OTC then catalyzes the conversion of citrulline and inorganic phosphate into carbamoyl-phosphate and ornithine. Finally, CK catalyzes the hydrolysis of carbamoyl phosphate to form CO_2 and NH_4^+, while the phosphate group is used to regenerate ATP from ADP.

The ADI pathway is widely distributed among bacteria, where it is often a major means of energy production [1]. However, the ammonium produced by this pathway has also been implicated in protecting some bacteria from the harmful effects of acidic environments [2, 3]. The pathway has been also described in Archaea [4]. This pathway has only been characterized in a few species of anaerobic eukaryotes namely the parabasalids *Trichomonas vaginalis* [5] and *Tritrichomonas foetus* [6], and the diplomonads

* Correspondence: vlada@natur.cuni.cz
[1]Department of Parasitology, Charles University, Faculty of Science, Prague, Czech Republic
Full list of author information is available at the end of the article

Giardia intestinalis [7], *Hexamita inflata* [8], and *Spironucleus salmonicida* [9]. All these species belong to Metamonada (Excavata), a clade of anaerobic protists with substantially modified mitochondria designated as hydrogenosomes or mitosomes. Metamonada consists of three lineages – Fornicata (e.g., *Giardia* and *Spironucleus*), Parabasalia (e.g., *Trichomonas* and *Tritrichomonas*), and finally Preaxostyla (*Trimastix*, *Paratrimastix* [10], and oxymonads) [11]. Currently, there is no information about the ADI pathway in Preaxostyla.

In *Trichomonas vaginalis* the ADI pathway generates up to 10 % of the energy produced by glucose fermentation [12]. OTC and CK were shown to be cytosolic, while ADI was described as membrane-associated in both *Trichomonas vaginalis* and *Tritrichomonas foetus* [6]. The ADI of *Trichomonas vaginalis* was later shown to be localized in hydrogenosomes and an *in situ* pH buffering function has been proposed [13]. The ADI pathway of *Giardia intestinalis* is completely cytosolic and produces up to 8 times more ATP than sugar metabolism [7]. Besides this energy-producing function, it has been proposed that the enzymes play an important role in the pathogenesis of *Giardia intestinalis* and *Trichomonas vaginalis*. The protists secrete ADI and OTC from their cells causing arginine depletion thus reducing the ability of the infected tissue to produce antimicrobial nitric oxide [14, 15]. Other known effects of these parasite's ADI pathway enzymes include growth arrest of intestinal epithelial cells [16], inhibition of T cell proliferation [15], and alteration of the phenotype and cytokine production of dendritic cells [17]. Another diplomonad with a characterized ADI pathway, the free-living *Hexamita inflata*, inhabits environments with varying levels of dissolved oxygen. It has been suggested that the ADI pathway may contribute to the metabolic flexibility of this organism, producing a significant amount of ATP under oxygen-limited conditions, while glycolysis is the main energy source under oxic or microoxic conditions, however the oxygen relationship might be incidental or secondary [8].

Of the three enzymes, only ADI itself is considered to be specific to the ADI pathway. CK has an additional role in purine and nitrogen metabolism and OTC may catalyze synthesis of citrulline as a nitrogen reservoir in plants [18] or be a part of ornithine-urea cycle in animals, diatoms and dinoflagellates [19, 20]. Therefore, the presence of ADI in organisms where no ADI pathway is known is intriguing and deserves further investigation. For example, within the chlorophytes the ADI gene was found in three species of *Chlorella* [21, 22] and *Chlamydomonas reinhardtii* [23] And ADI activity has been reported in multiple species of Chlorodendrophyceae, Trebouxiophyceae, Chlorophyceae, and Ulvophyceae [24], that is, in all classes of the “crown group” of Chlorophyta [25].

The first known sequence of a eukaryotic ADI, from *Giardia intestinalis*, showed no specific relationship to any bacterial or archaeal clade [26]. Later analyses included sequences from *Trichomonas vaginalis*, *Spironucleus vortens*, *Sp. barkhanus*, and *Sp. salmonicida* (Metamonada), *Euglena gracilis* and '*Seculamonas*' sp. (Discoba), *Chlamydomonas reinhardtii* and *Chlorella* sp. (core Chlorophyta [27]), and *Mastigamoeba balamuthi* and *Dictyostelium discoideum* (Amoebozoa). All the eukaryotic sequences formed a well-supported clade related to Archaea, consistent with a single origin of ADI in the eukaryotic domain [9, 13].

Due to its involvement in other pathways, it is not surprising that OTC is more widespread among eukaryotes compared to the other ADI pathway enzymes. The phylogenetic analysis of OTC by Zúñiga et al. [26] recovered two distinct eukaryotic clades branching in different positions among bacteria, one comprising sequences from Embryophyta and other composed of metazoan and fungal sequences. The only eukaryote outside these two clades was *Giardia intestinalis*, which was also the only one with a characterized ADI pathway. The sequence from *Giardia intestinalis* branched among bacterial sequences without close relationship to any other eukaryotic clade. Later analyses demonstrated that *Spironucleus salmonicida* and *Trichomonas vaginalis* OTC sequences formed a well-supported clade with *Giardia intestinalis* [9], suggesting the existence of a third independent group of eukaryotic OTCs present in Metamonada and potentially involved in the ADI pathway. The same analysis also showed two stramenopile sequences branching clearly inside the Metazoa-Fungi group.

Sequences of CK from *Giardia intestinalis*, *Hexamita* sp., and *Trichomonas vaginalis* formed a relatively well-supported clade not closely related to any bacterial or archaeal sequences [26]. The monophyly of eukaryotic CKs was later questioned after adding sequences from *Spironucleus salmonicida* and *Carpediemonas membranifera*, with the *Trichomonas vaginalis* sequence branching separately from other eukaryotes, although statistical support for this topology was very low [9].

In summary, the complete set of ADI pathway enzymes has been found in representatives of two out of three major lineages of Metamonada: Parabasalia and Fornicata. All the metamonad enzymes appear to be closely related to each other. This raises several questions about the evolutionary history of the pathway among eukaryotes. Is it present also in the third and least investigated lineage of metamonads, Preaxostyla? Was it present in the common ancestor of the Metamonada? Do representatives of other eukaryotic lineages possess the ADI pathway as well? If so, do all the eukaryotic enzymes involved in the ADI pathway originate from the same source or do they represent independent acquisitions?

Here, we take advantage of the recent progress in genome and transcriptome sequencing of less studied protists to perform an up-to-date survey and

phylogenetic analysis of ADIs, OTCs, and CKs. This survey focuses on elucidating the evolutionary history of the arginine deiminase pathway in eukaryotes, with special emphasis on Metamonada. In addition to phylogenetic studies, we determine the subcellular localization of these enzymes in two members of Preaxostyla, *Paratrimastix pyriformis* and oxymonad *Monocercomonoides* sp. PA203.

Results

Distribution of ADI, OTC, and CK across eukaryotes

Our survey revealed the presence of ADI, OTC, and CK in the three main eukaryotic clades defined by Adl et al., 2012 [28] (Fig. 1). The first and presumably the most specific enzyme of the pathway, i.e. without any role outside the ADI pathway reported so far, is ADI itself. This was found in 40 taxa, as shown on the schematic tree in

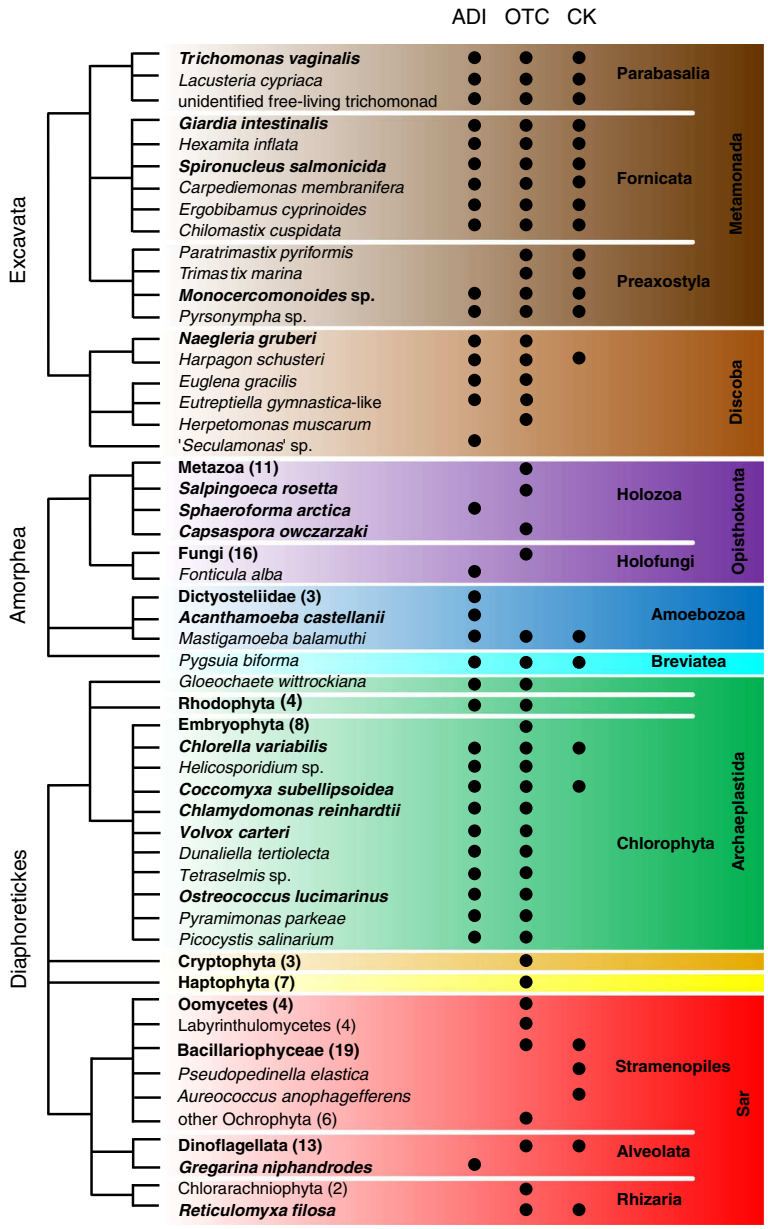


Fig. 1 Distribution of enzymes of the arginine deiminase pathway across eukaryotic diversity. Arginine deiminase (ADI), ornithine transcarbamylase (OTC), carbamate kinase (CK). Taxon names in boldface indicate lineages containing at least one representative with a sequenced nuclear genome. Numbers in brackets indicate number of sequences from the given taxon included in our analyses. Colored rectangles indicate major eukaryotic groups as follows: *dark brown* – Metamonada; *light brown* – Discoba; *violet* – Opisthokonta; *blue* – Amoebozoa; *cyan* – Breviatea, *green* – Archaeplastida; *orange* – Cryptophyta; *yellow* – Haptophyta; *red* – SAR. Excavata, Amorphea, and Diaphoretickes are names of the 3 putative largest clades of eukaryotes as proposed in Adl et al., 2012; relationships between them are not resolved

Fig. 1, and these taxa represent most eukaryotic supergroups (highlighted by colored backgrounds). Of these, 16 species (most metamonads, *Harpagon*, *Mastigamoeba*, *Pygsuia*, *Chlorella*, and *Coccomyxa*) encoded all three enzymes, while the other species encoded only one or two enzymes. ADI was not detected in any representative of the clades Metazoa, Fungi, Embryophyta, Cryptophyta, and Haptophyta, nor in Sar [28], with the single questionable exception of *Gregarina niphandrodes* (see below). OTC was the most widespread enzyme, being found in 131 taxa including the major multicellular groups of Metazoa, Fungi, and Embryophyta. CK was detected in all the investigated metamonads, multiple Bacillariophyceae, Dinoflagellata and 8 other species. Please note that the given numbers do not represent the actual quantity of eukaryotic species with the particular gene since several groups, e.g. Metazoa, Bacillariophyceae, are represented by only a limited number of randomly selected sequences.

Phylogenetic analyses

Arginine deiminase

Compared to the previous analyses we present a more robust analysis including 40 eukaryotic species (Fig. 1). The phylogenetic tree (Fig. 2) shows two clearly separated (RAxML bootstrap support/IQ-TREE bootstrap support: 100 %/100 %) clans of ADIs, one comprising all bacteria and one isolated eukaryote, *Gregarina niphandrodes*, in a highly nested but poorly resolved position, and the second composed of clans of Archaea plus a few Bacteria (64 %/97 %) and Eukaryota (51 %/81 %). The topology within the eukaryotic branch is poorly resolved overall, however, a few clades of lower-than-supergroup rank were recovered with strong support (i.e. bootstrap support > 80 %). These are Parabasalia, Diplomonadida, Oxymonadida, Chlorophyta, and Dictyosteliida.

Ornithine transcarbamylase

Our analysis included sequences from selected representatives of Metazoa, Fungi, and Embryophyta and 110 sequences from 103 other eukaryotes (Fig. 1). Several bacterial sequences of aspartate transcarbamylase, a protein closely related to OTC, were included to provide an outgroup for rooting the tree. Our analysis of OTC phylogeny (Additional file 1) supports the existence of three large groups and two separately-branching eukaryotic OTCs.

The first large clade is strongly supported (100 %/100 %) and contains Metazoa, Fungi, Oomycota, Bacillariophyceae (i.e. diatoms), a few lineages of other Stramenopiles (*Nannochloropsis*, *Vaucheria*, *Ectocarpus*, *Heterosigma* and *Ochromonas*), the holozoan *Capsaspora*, and one single excavate, the trypanosomatid *Herpetomonas muscarum*, which branches among Fungi.

The second eukaryotic group, already indicated in the analysis by Andersson et al. [9], is well supported (83 %/100 %) and includes all Metamonada, *Harpagon shusteri*, *Naegleria gruberi*, *Reticulomyxa filosa*, *Pygsuia biforma*, *Mastigamoeba balamuthi*, and many representatives of autotrophic groups, namely dinoflagellates, cryptophytes, euglenophytes, chlorarachniophytes, and stramenopiles like *Aureococcus*, *Aureoumbra*, *Pelagococcus*, *Pelagomonas*, and *Pseudopedinella*. The monophyly of Parabasalia is well supported. A sequence from a recently described archaeon *Lokiarchaeum* sp. is also included in this group, however at an unsupported position.

The third group is composed of euglenophytes, green algae with green plants, red algae, and haptophytes, with haptophytes branching inside red algae. A clade of mostly Desulfobacteraceae bacterial sequences branches inside this group of eukaryotic sequences.

The only two eukaryotes outside these three large clades are the choanoflagellate *Salpingoeca rosetta* (sequence obtained from the genome), which branches as sister to Microgenomates bacterium (78 %/100 %), and the rhizarian *Paulinella chromatophora* (red star in Additional file 1) inside Cyanobacteria with good statistical support (100 %/100 %). Since the *Paulinella* sequence originates from the genome of the chromatophore, not the *Paulinella* nucleus, it actually represents a cyanobacterial OTC.

Carbamate kinase

We have included 47 sequences from 44 eukaryotic species in our analysis (Fig. 1). Our tree (Additional file 2) shows eukaryotes falling into several separate clusters. One of two substantial groups is an unsupported clan of Fornicata, Parabasalia, *Harpagon schusteri*, *Pygsuia biforma*, and *Mastigamoeba balamuthi*. A well-supported Preaxostyla clade (96 %/100 %) branches at a different place among bacteria, as a sister group to Hadesarchaea archaeon and Anaerolineae bacterium (96 %/100 %). The second large eukaryotic clan (100 %/100 %) is composed of all the dinoflagellate sequences, as well as sequences from diatoms, *Pedinella* and *Aureococcus*. Dinoflagellata form a well-supported group within this clan. Three sequences from diatoms do not branch together with other ochrophytes (the photosynthetic Stramenopiles), and instead form a separate well-supported clan (100 %/100 %) among bacteria. This may represent a second form of the enzyme, since *Thalassiosira pseudonana* appears in both diatom groups. The only two known CKs from green plants (*Chlorella* and *Coccomyxa*) branch together (97 %/99 %) but separated from other eukaryotes. The *Reticulomyxa* CK sequence is also isolated from the rest of eukaryotes.

Concatenation

We also performed a phylogenetic analysis of a concatenation of all three enzymes. In the first step, we have

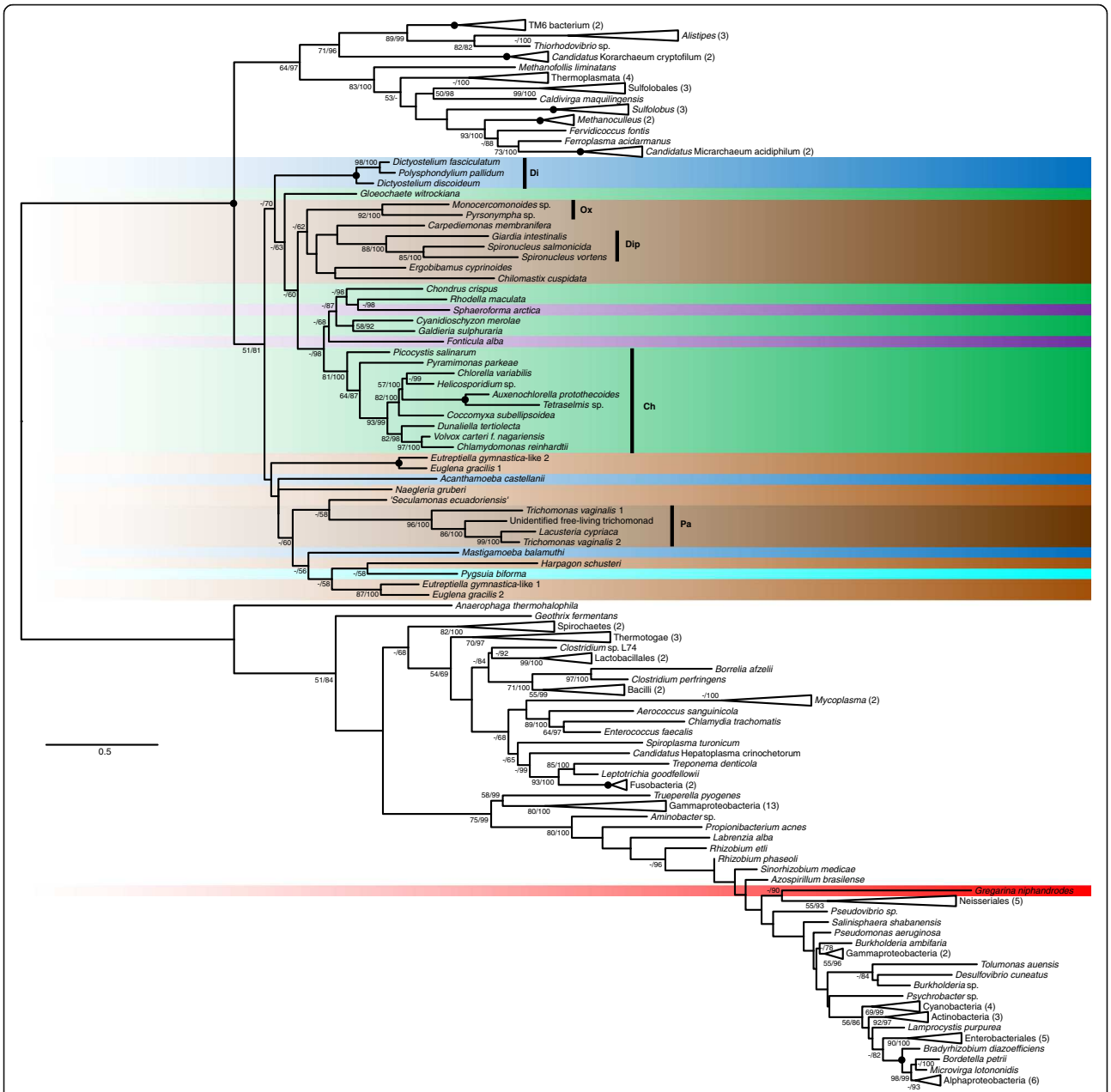


Fig. 2 Phylogenetic tree of ADI sequences. The tree based on a 257 positions long protein alignment of 152 sequences was constructed in RAxML using the LG4X + Γ model of substitution. Eukaryotic taxa are highlighted in different colors according to the major group they belong to. The color code is the same as in Fig. 1. The values at nodes represent RAxML bootstrap support/IQ-TREE bootstrap support. Only values above 50 % are shown. *Black* circles indicate support of 100 %/100 %. Vertical *black* bars indicate well-supported eukaryotic clades: Pa – Parabasalia; Di – Dictyosteliida; Ch – Chlorophyta; Ox – Oxymonadida; Dip – Diplomonadida. Species with multiple sequences included: *Euglena gracilis* 1 – GI 109790819; *Euglena gracilis* 2 – GI 109784514; *Eutreptiella gymnastica*-like 1 – CAMPEP 0200414012; *Eutreptiella gymnastica*-like 2 – CAMPEP 0200422928; *Trichomonas vaginalis* 1 – TVAG 183850; *Trichomonas vaginalis* 2 – TVAG 344520. The tree is unrooted

prepared an alignment supermatrix in which we have included all eukaryotes and representatives of prokaryotes that contain a complete set of the three enzymes, and may use the ADI pathway. In order to detect potential incongruities between gene partitions caused by lateral gene transfer we have performed a phylogenetic analyses of the individual gene partitions from this supermatrix. Based on these gene trees (Additional file 3) we removed taxon-gene sequences that branched with bootstrap support higher than 50 % within a clan of sequences outside its own domain (e. g. eukaryotic sequence outside Eukaryota) from the concatenated alignment – namely CKs from

Monocercomonoides sp., *Pyrsonympha* sp., *Chlorella variabilis*, and *Coccomyxa subelipsoidea*. We also removed OTCs from *Chlorella variabilis*, and *Coccomyxa subelipsoidea* because in the large single gene tree (Additional file 1) they branch within a clade which is sister to Chlorobi Bacteria with 100 % IQ-TREE bootstrap support.

The analysis performed on the alignment after removal of these sequences (Additional file 4) revealed a strong bipartition (100 %/100 %) grouping Eukaryota and Archaea to the exclusion of Bacteria and within this part of the tree the eukaryotes formed a well-supported (100 %/100 %) clan sister to the archaeon *Candidatus* Korarchaeum cryptofilum. In order to recover the relationships within the Eukaryota–Archaea group without the disturbing long branch of Bacteria we repeated the analysis without the bacterial sequences (Fig. 3). In this unrooted tree Eukaryota are grouped with *Candidatus* Korarchaeum cryptofilum to the exclusion of the rest of Archaea with high support (97 %/98 %). We also performed a Eukaryota-only analysis of the concatenated dataset for the purpose of hypotheses testing (Additional file 5).

Hypotheses testing

We used approximately unbiased (AU) and expected likelihood weight (ELW) tests to assess whether the inferred phylogenies are in a significant conflict with the monophyly of eukaryotes, metamonads and with the expected eukaryotic species tree. The results are summarized in Table 1. The AU tests rejected monophyly of metamonads in the OTC and CK trees and monophyly of expected eukaryotic phylogeny in concatenation. The ELW tests rejected the monophyly of metamonads in the OTC tree, the monophyly of both clades in the CK and monophyly of metamonads and the expected eukaryotic phylogeny in the concatenation tree.

Localization of ADI pathway enzymes in Preaxostyla

Another aim of this work was to infer the subcellular localization of ADI pathway enzymes in members of the poorly studied Preaxostyla clade. Genomic and transcriptomic projects have revealed the presence of all three enzymes in *Monocercomonoides* sp. and *Pyrsonympha* sp., while only OTC and CK enzymes were detected in *Trimastix marina* and *Paratrimastix pyriformis*. We have chosen *Monocercomonoides* sp. PA203 and *Paratrimastix pyriformis* for further study.

We investigated the presence of mitochondrion-targeting signals in the enzymes of interest (Additional file 6) using the signal prediction software TargetP 1.1 [29] and Mitoprot II v1.101 [30]. TargetP did not predict any targeting signals. Mitoprot II predicted a single mitochondrion-targeting signal, for the OTC sequence of *Paratrimastix pyriformis*.

To validate the results of mitochondrion-targeting signal prediction we used the *Trichomonas vaginalis* T1

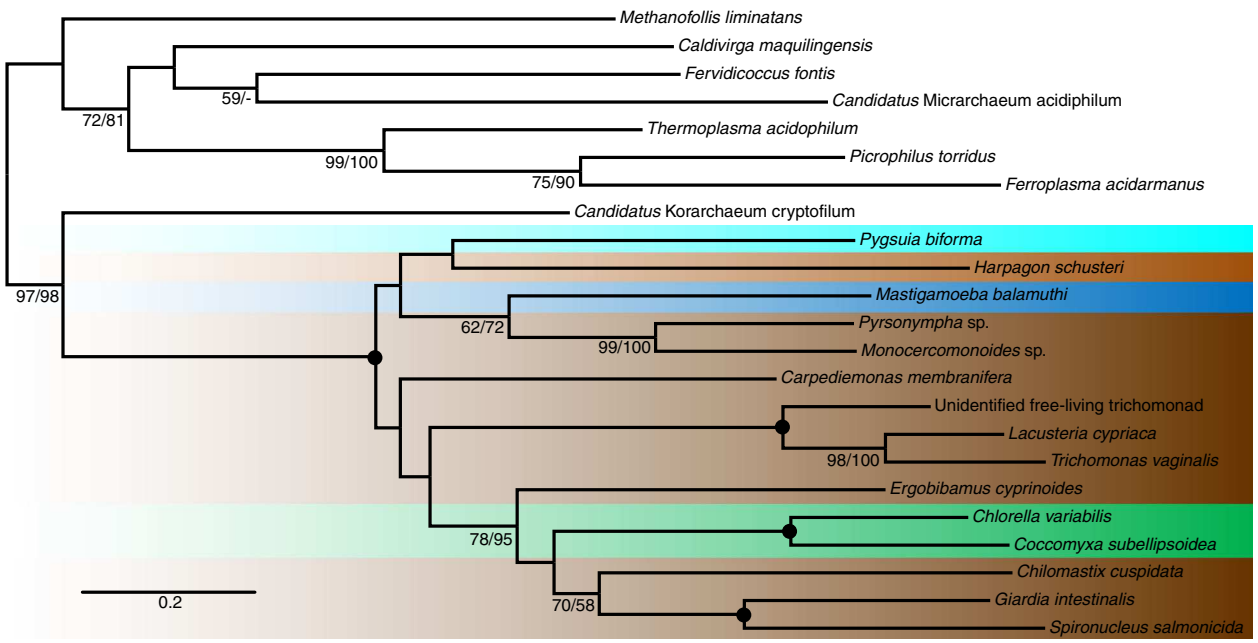


Fig. 3 Phylogenetic tree of concatenated ADI, OTC, and CK sequences with Bacteria removed. The tree based on a 750 positions long protein alignment of 23 sequences was constructed in RAxML using LG4X + Γ model. Eukaryotic taxa are highlighted in different colors according to the major group they belong to. The color code is the same as in Fig. 1. The values at nodes represent RAxML bootstrap support/IQ-TREE bootstrap support. Only values above 50 % are shown. *Black* circles indicate support 100 %/100 %. The tree is unrooted

Table 1 Results of approximately unbiased (AU) and expected likelihood weights (ELW) tests

Data set – hypothesis	AU test	ELW test
ADI – Metamonada monophyly.	0.64	0.10
OTC – Eukaryota monophyly	0.64	0.5
OTC – Metamonada monophyly	0	0
CK – Eukaryota monophyly	0.12	0
CK – Metamonada monophyly	0	0
Euk. Conc. – Metamonada monophyly	0.17	0
Euk. Conc. – expected euk. phylogeny	0	0

The tests were performed for 4 sets of taxa – ADI: ADI dataset (as in Fig. 2), OTC: OTC dataset (as in Additional file 1), CK: CK dataset (as in Additional file 2), Euk. Conc.: concatenation dataset without prokaryotic sequences (as in Additional file 5)

heterologous expression system, with the assumption that an undetected mitochondrion-targeting signal may nonetheless be recognized by the *Trichomonas* hydrogenosomal import machinery. We transfected *Trichomonas vaginalis* cells with plasmids containing HA-tagged OTC and CK from *Paratrimastix pyriformis* and ADI, OTC, and CK from *Monocercomonoides* sp. In all cases fluorescence microscopy showed that the heterologously expressed proteins do not co-localize with the signal from the hydrogenosomal marker protein (malic enzyme), but instead formed a diffuse pattern all over the cell (Fig. 4). This demonstrates that the inserted proteins are not recognized as hydrogenosomal-import targets in *Trichomonas vaginalis*. The results of these experiments are consistent with the fact that most ADI pathway enzymes in eukaryotes are localized in the cytosol.

Discussion

ADI pathway enzymes are widespread in eukaryotes

Before this study, only two closely related lineages of eukaryotes had been conclusively shown to possess a complete ADI pathway. These were Parabasalia and Diplomonadida, both being members of Metamonada, a subgroup of Excavata. Our survey has shown the presence of all three enzymes in ten other eukaryotic species. Among these are other members of Metamonada – including free-living members of Fornicata related to the predominantly parasitic diplomonads (*Ergobibamus cyprinoides*, *Chilomastix cuspidata, Carpediemonas membranifera*), and members of the third metamonad lineage, Preaxostyla (*Monocercomonoides* sp. and *Pyrsonympha* sp.). The ADI pathway was also identified in non-metamonads including the heterolobosean *Harpagon schusteri*, the amoebozoan *Mastigamoeba balamuthi*, the breviate *Pygsuia biforma*, and the green algae *Chlorella variabilis* and *Coccomyxa subelipsoidea*. Further functional studies are needed to determine whether these enzymes function within an ADI pathway in these species. It is possible that the possession of the complete pathway is connected with their anaerobic lifestyle since most of these organisms are anaerobes, microaerophiles or aerobes able to live for long periods under anaerobic conditions [31–37].

Many investigated eukaryotes possessed incomplete sets of ADI pathway enzymes. The presence of OTC or CK on their own is not surprising, as they are known to be involved in other biochemical processes including the ornithine-urea cycle or purine biosynthesis. The presence of ADI on its own was unexpected, yet we identified ADI in a broad spectrum of eukaryotic lineages without complete pathway. The apparent absence of OTC or CK may be due to the incompleteness of transcriptome or genome data, nevertheless, our observation suggests that ADI may also function outside the context of the ADI pathway in some eukaryotes.

Phylogenetic histories of enzymes

None of the enzyme phylogenies is completely consistent with the expected species relationships. In single gene trees, eukaryotes are always dispersed in multiple clades, suggesting complicated evolutionary histories. The backbone topologies were generally weakly supported, and many of these incongruences are probably the result of low phylogenetic signal. Nevertheless, some conflicts with species phylogeny are better supported and some were confirmed by phylogenetic tests. These can potentially be attributed to lateral gene transfers (LGTs, also known as horizontal gene transfers – HGTs) or endosymbiotic gene transfers (EGTs). The sister relationship of Preaxostyla and *Spirochaeta* in the CK tree, and the position of *Gregarina* within Bacteria in the ADI tree are two such examples of potential LGT, albeit the latter may also represent a contamination. Since haptophytes are known to harbor secondary plastids of potentially red algal origin [38], the position of haptophytes within a red algal clade in the OTC analysis, might represent a potential EGT. Conversely, many moderately and robustly supported eukaryotic clades are taxonomically reasonable, indicating the important role of vertical inheritance.

The taxon sampling in the concatenation analyses was lower, because the analyses included only those taxa that may utilize the arginine deiminase pathway and not those that use individual enzymes for other purposes. The resolution of the concatenation tree was higher than the individual gene trees and strongly supported the monophyly of eukaryotes (99 %/100 %) and their close relationship to Archaea. The clade of eukaryotes branched with Archaea (100 %/100 %) as a sister to *Candidatus* Korarcheum cryptophylum. Increased support of these nodes should partly be ascribed to the lower number of taxa but it also suggests that the phylogenetic signal regarding these deep nodes for this narrow set of taxa and after exclusion of obvious cases of LGT is largely congruent.

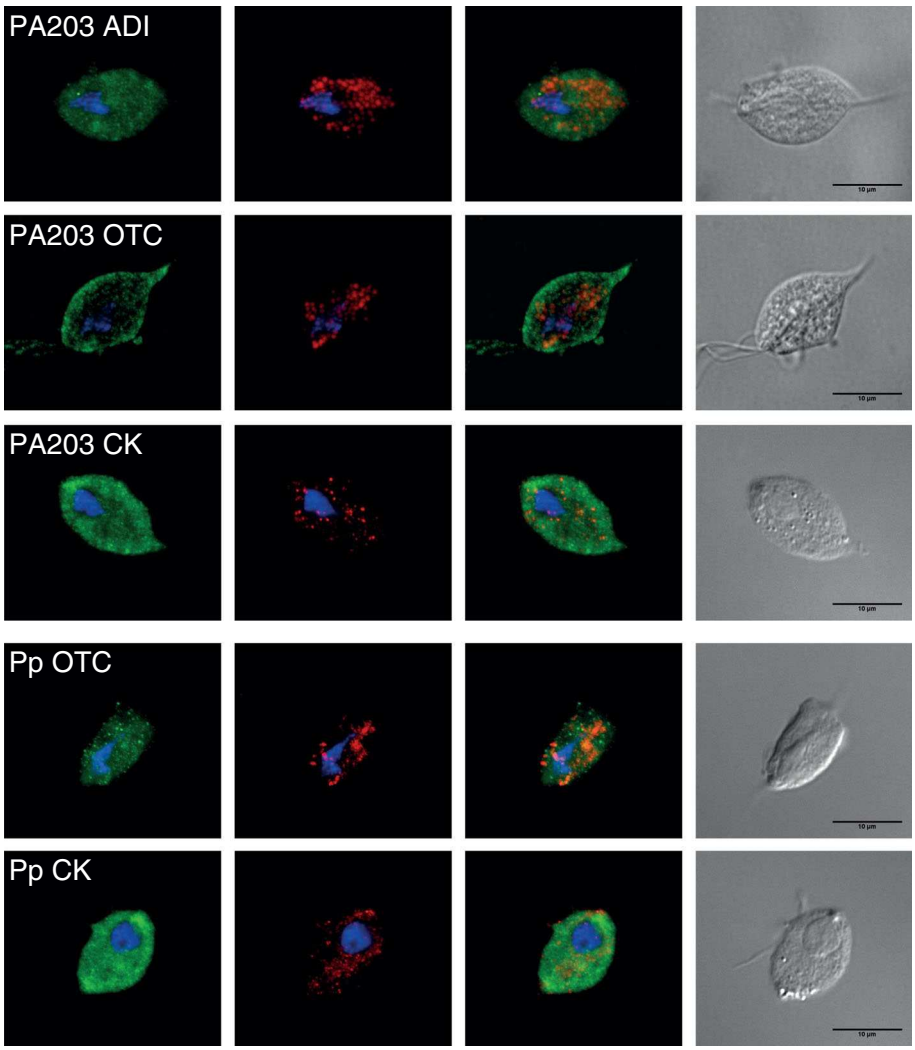


Fig. 4 Localisation of *Monocercomonoides* and *Paratrimastix* enzymes in *Trichomonas vaginalis* cells. Immunofluorescence micrographs of *Trichomonas vaginalis*, in which the HA-tagged versions of enzymes were expressed. *Green* signal from anti-HA antibody does not localize to hydrogenosomes of *Trichomonas vaginalis*, which are marked by *red* anti-malic enzyme antibody. *Blue* signal indicates DAPI-stained nuclei. Abbreviations: PA203 – *Monocercomonoides* sp. PA203; Pp – *Paratrimastix pyriformis*; ADI – arginine deiminase; OTC – ornithine transcarbamylase; CK – carbamate kinase

ADI pathway is ancestral in Metamonada

The presence of the complete ADI pathway is widespread in metamonads, protists that specifically inhabit low-oxygen environments. In most of the phylogenies, metamonad taxa branch close to each other, but they never form an exclusive clade. In CK trees there is a well-supported relationship between Preaxostyla (*Paratrimastix*, *Monocercomonoides,* and *Pyrsonympha*) sequences and a prokaryotic clan of Hadesarchaea archaeon and Anaerolineae bacterium. Metamonada do not appear monophyletic even in the concatenation trees and their monophyly was rejected by ELW test. Taking together all this information we propose that the complete arginine deiminase pathway was present already in the common ancestor of Metamonada and was vertically inherited by the extant metamonad lineages, with a few exceptions. The exceptions are the putative losses of ADI in *Trimastix marina* and *Paratrimastix pyriformis* and the putative replacement by a bacterial CK in the Preaxostyla. It is also possible that some enzymes or the whole pathway were laterally transferred from metamonads to other eukaryotes, which would disrupt the monophyly of Metamonada on trees.

Our localization experiments suggest that all enzymes in both *Paratrimastix pyriformis* and *Monocercomonoides* sp. are localized in the cytosol, like the enzymes in *Giardia intestinalis* but unlike the ADI in *Trichomonas vaginalis*. We therefore conclude that cytosolic localization of the pathway is an ancestral trait of all metamonads. It should be noted, however, that the cytosolic localization of ADI in

Monocercomonoides sp. may not be informative, since this protist does not contain mitochondrion [39]. *Paratrimastix pyriformis* harbors an organelle similar to the hydrogenosome, but no ADI-coding gene has been found in its transcriptome to test its localization.

ADI pathways in other eukaryotes

Besides Metamonada, five other species contain a complete ADI pathway and all branch within the eukaryotic clade on the concatenation tree (Fig. 3). These represent four different supergroups of eukaryotes; only *Harpagon* is from the same supergroup as Metamonada (supergroup Excavata), and even then it belongs to a different sub-branch (Discoba). Moreover, Excavata are likely not monophyletic at all [40] and the root of eukaryotes may be situated inside the group. *Harpagon*, *Mastigamoeba*, and *Pygsuia* ADI pathway enzymes branch close to Metamonada in all three gene trees, and so it is very probable that the pathway in these three species was derived from the same source as the pathway in Metamonada.

The situation in *Chlorella variabilis* and *Coccomyxa subelipsoidea* is less clear. These organisms are the only green algae known to contain all three ADI pathway enzymes. Their ADIs and OTCs branch together with other green algae and plants in the individual gene trees (while no other Plantae have CK), supporting the presence of the complete set in the last common ancestor of Chlorophyta. However, Chlorophyta did not branch in a common eukaryotic clade with Metamonada, *Harpagon*, *Mastigamoeba* and *Pygsuia* in OTC and CK phylogenies, suggesting that these two enzymes in Chlorophyta may have independent origins. In the concatenation tree, *Chlorella variabilis* and *Coccomyxa subelipsoidea* represented only by ADI sequences branch together with other eukaryotes. Presence of the ADI pathway in the last common ancestor of Chlorophyta would be consistent with the ADI enzymatic activity previously reported from members of Chlorodendrophyceae, Trebouxiophyceae, Chlorophyceae, and Ulvophyceae [24]. However, the function of the other enzymes in a typical ADI pathway is questionable, since the OTCs from *Chlorella autotrophica*, *Chlorella saccharophila* (Trebouxiophyceae), and *Dunaliella tertiolecta* (Chlorophyceae) were found to have no measurable activity in the direction of the ADI pathway, i.e. conversion of citrulline to ornithine [24]. It is therefore possible that the two Chlorophyta species with all three enzymes nonetheless do not use the ADI pathway.

Origin of the eukaryotic ADI pathway

The simplest explanation of the fact that the complete sets of ADI pathway enzymes from several eukaryotic lineages are related is that they are inherited from their common ancestors. The taxonomic composition of the eukaryotic clade in the concatenation tree is so broad that their common ancestor must have been either the last eukaryotic common ancestor (LECA) or its close descendant. This assumption is reasonable even if we would not consider *Chlorella* and *Coccomyxa*. An alternative explanation for the close relationships of ADI pathways would be that the genes were acquired more recently by one eukaryotic lineage (perhaps Metamonada, where it is most common), and then spread from this lineage into others via eukaryote-to-eukaryote lateral gene transfers.

Based on our data we are unable to decide which alternative is more likely. Vertical inheritance of the ADI pathway from LECA would be consistent with the sisterhood of the eukaryotic clade and the archaeon *Candidatus* Korarchaeum cryptofilum in the concatenation tree (Fig. 3, Additional file 4), since recent studies indicate Korarchaeota are indeed closely related to the eukaryotes [41]. Moreover, OTC sequences from *Lokiarchaeum* sp., which is the closest known relative to eukaryotes [42], are related to the Metamonada-containing eukaryotic clade. CK sequence of this archaeon branches outside eukaryotes sister to a proteobacterium *Desulphobacula toluolica* (97 %/ 100 %), but nodes separating these two from Metamonada did not receive strong support. ADI sequence from *Lokiarcheum* is not available, and so this organism was not included in the concatenation analysis. We must also take into account that the position of the root of the eukaryotic tree is still an unresolved question [40, 43, 44]. If the root falls outside Amorphea + Discoba (position of Metamonada relative to the root is not known), then our results do not necessarily uncover the condition of LECA.

The later acquisition of the pathway by Metamonada and its spread to unrelated eukaryotes by eukaryote-to-eukaryote transfer is supported by the fact that the concatenation trees are incongruent with the expected relationships of taxa.

Conclusions

Our broad survey of the arginine deiminase pathway enzymes has shown that they are present in representatives of all major lineages of eukaryotes. Sixteen protists (most metamonads, *Harpagon*, *Mastigamoeba*, *Pygsuia*, *Chlorella*, and *Coccomyxa*) contain the complete set of the three enzymes, while other organisms contain incomplete sets. The enzyme ADI is present in several species without the complete arginine deiminase pathway, suggesting its involvement in other cellular processes. The topology of individual gene trees is generally not very well supported and particularly in OTC and CK trees the eukaryotic enzymes form multiple clearly unrelated clades consisting of mixtures of eukaryotic supergroups. This indicates that multiple prokaryote-to-eukaryote and eukaryote-to-eukaryote LGT events took place in the history of these enzymes. It is possible that in some groups the enzyme acquisition was

connected with its involvement in novel biochemical processes like the ornithine-urea cycle.

Based on the presence of the complete pathway in most metamonads and based on the phylogenetic affinity of metamonad enzymes, we conclude that the ancestor of metamonads already possessed this pathway. The concatenation analyses suggest that eukaryotes with the complete ADI pathway, including Metamonads, *Harpagon*, *Mastigamoeba* and *Pygsuia* (and possibly *Chlorella* and *Coccomyxa*), may have acquired the genes from a single, archaeon-related source. One intriguing possibility is that the acquisition of the pathway may date back as deep as to LECA, but other scenarios involving LGT events are also plausible. To resolve the last issue, it will be necessary to obtain data from a more diverse set of prokaryotes and eukaryotes, especially those branching close to the root of the eukaryotic tree and close to the root of major eukaryotic lineages.

Methods

Obtaining sequences

The majority of eukaryotic sequences included in the survey were obtained from the NCBI database (Release 68), JGI database [45], or Marine Microbial Eukaryote Transcriptome Sequencing Project [46]. Initial searches were performed using BLASTp and tBLASTn algorithms [47] with *Giardia intestinalis, Trichomonas vaginalis*, and several bacterial sequences as queries. The searches of public databases were then repeated several times while restricted to a particular major eukaryotic lineage (e.g. Cryptophyta, Alveolata) and with a phylogenetically closest available sequence as a query. All the eukaryotic sequences with E-value lower than 10^{-3} were downloaded and used in subsequent analyses.

The prokaryotic sequences included in the survey were retrieved from the NCBI database using the same query as in the search for eukaryotic sequences. The search was also repeated several times with varying taxonomic restrictions to ensure that all the bacterial and archaeal phyla containing the particular enzyme are represented in the analysis. We used all archaeal sequences with E-value lower than 10^{-3} and a limited number of bacterial sequences with E-value lower than 10^{-3} and annotated as the protein of interest. It is important to note that the set of bacterial sequences used in our analyses is not exhaustive and therefore we do not infer any evolutionary hypotheses about Bacteria in this study.

In order to mitigate the risk of missing prokaryotic data influencing the relationships between eukaryotic groups we enriched our datasets with the closest prokaryotic homologs to each of the eukaryotic sequences by searching the NCBI database using BLASTp with each eukaryotic sequence as a query and downloaded the prokaryotic sequence with the lowest e-value from each search.

We investigated whether those eukaryotic sequences which were not branching within eukaryotic clades represent *bona fide* eukaryotic sequences or contamination of the data sets. Nucleotide sequences obtained from transcriptomic data were checked for similarity with sequences deposited in NCBI and those that were identical or very similar to bacterial genomes (bit score higher than 200 and similarity along the entire length of the sequence) were excluded. In the case of sequences obtained from genomes, the entire gene content of the contiguous sequence scaffold was used as a query for BLAST search of the NCBI database in order to identify any known sequences with high sequence similarity along the entire length of the sequence, indicating possible contamination. Furthermore, candidate sequences and neighboring genes were investigated for the presence of introns and the origin and annotation of surrounding genes. These steps should identify some sequences originating from contamination, however, others could still remain due to the lack of data from the source of the contamination or incorrect assembly of genomic data resulting in chimerical sequences.

The sequences downloaded from public databases were combined with sequences extracted from genomic and transcriptomic projects performed in the laboratories of co-authors. Brief information on the generation of these data sets is given below. Details of the *Monocercomonoides* sp. PA203 genome and transcriptome project are given in Karnkowska et al. [39]. Details of the *Paratrimastix pyriformis* transcriptome project are given in Zubáčová et al. [48]. Partial cDNA sequences corresponding to *Paratrimastix pyriformis* OTC and CK obtained in the transcriptome project were completed at their 5` ends by RACE using FirstChoice RLM-RACE kit (Life Technologies, AM1700). Amplifications by PCR were carried out using Takara Hot-Start ExTaq DNA Polymerase (Takara, RR006A) in 50 μl reactions. Outer 5` RLM-RACE PCR was done using the 5` RACE outer primer supplied in the kit and the following 5` RACE gene-specific outer primers: TpOTCout: CCAGCAGGAAGAGAAGGAGG and TpCKout: GCTTGCCGTAGTTGATGATG. Inner 5` RLM-RACE PCR was done using the 5` RACE inner primer supplied in the kit and the following 5` RACE gene-specific inner primers: TpOTCinn: AAGAGCTCGTGATCTGGAAG and TpCKinn: GCCAGAGGCGATGACAATGA (here and elsewhere, all primers reported in the 5' to 3' direction). The following touchdown program was used for each of the two PCRs: 95 °C (5 min), 15 cycles of 95 °C (1 min), 60 °C to 45 °C (35 s.) and 72 °C (2 min), 20 cycles of 95 °C (1 min), 45 °C (35 s.) and 72 °C (2 min), then a final polymerization step at 72 °C for 6 min. PCR products were cloned into pGEM-T Easy plasmid vector (Promega, A1360) and sequenced.

Sequences from *Lacusteria cypriaca* (strain LAI), an unidentified free-living trichomonad (strain LAGOS2D) and

Pyrsonympha sp. were mined from RNA-seq data sets generated using the llumina MiSeq sequencing platform. Sequences from *Trimastix marina*, *Carpediemonas membranifera*, *Chilomastix cuspidata*, *Mastigamoeba balamuthi*, and *Pygsuia biforma* were mined from RNA-seq data sets generated using the Illumina HiSeq sequencing platform. Sequences from *Harpagon schusteri* were mined from RNA-seq data obtained using the 454 sequencing platform and sequences of *Ergobibamus cyprinoides* were mined from data sets generated by combination of Sanger and 454 sequencing platforms. The assembled sequences were submitted to GenBank under accession numbers KT883858-KT883885.

Phylogenetic analyses

Inferred amino acid sequences were aligned using MAFFT version 7 [49] and the resulting alignments were manually trimmed. Highly variable misaligned sections of several eukaryotic sequences, possibly results of sequencing errors, were removed from the alignment manually. The concatenated alignment was constructed from the single gene alignments using SequenceMatrix [50]. The final alignments can be downloaded from our web page: http://protistologie.cz/hampllab/NovakData.zip [51]. Phylogenetic inference was performed using substitution models suggested by ProtTest 2.4 server [52] – LG4X + Γ model. Maximum Likelihood trees were inferred using RAxML-HPC2 version 8 available on CIPRES [53], with 10 starting trees and also using IQ-TREE v1.4.2 [54] under the LG + C20 + F + G4 model for the single gene trees and LG + C40 + F + G4 model for concatenated datasets. The model that best fits the data was determined by IQ-TREE according to the Bayesian information criterion (BIC). The LG matrix was combined to an amino acid class frequency mixture model with 20 (for single gene trees) and 40 (for concatenated datasets) frequency component profiles. Statistical support for branches was assessed by multiparametric bootstrapping (1000 replicates) in RAxML and by the ultrafast bootstrap approximation (UFboot) with 1000 replicates in IQ-TREE.

Topology tests

Phylogenetic hypotheses were tested by an approximately unbiased (AU) test [55] and expected likelihood weights (ELW) method [56] implemented in IQ-TREE 1.4.2 [54]. For all datasets we tested whether their ML phylogeny is in significant conflict with the monophyly of Metamonada. In the case of OTC and CK, we also tested whether the phylogeny is in significant conflict with the monophyly of eukaryotes. In the case of the concatenated dataset, we tested whether their ML phylogeny is in significant conflict with the monophyly of Metamonada as well as whether the relationships within the eukaryotic clade (Additional file 5) significantly conflicts with a user-defined expected species tree (Additional file 7). For the latter two tests we used a concatenated dataset without prokaryotic sequences to eliminate the influence of relationships among prokaryotes and position of the eukaryotic root. The ADI sequence attributed to *Gregarina niphandrodes* was excluded from all tests including ADI data, as this clearly represents contamination or very recent LGT.

To perform AU and ELW tests, a set of 1003 topologies was created, containing the unconstrained ML topology inferred by RAxML, 1000 topologies inferred by RAxML during bootstrapping, and the best trees inferred by RAxML under the selected constraints (eukaryotic monophyly, Metamonada monophyly or eukaryotic phylogeny). Site likelihoods for topologies were calculated by IQ-TREE using the LG + C20 + F + G4 model. The sets of site likelihoods were then compared using the AU test in IQ-TREE, with 10 000 replicates.

Cloning of ADI pathway genes

ADI pathway genes were amplified from *Paratrimastix* and *Monocercomonoides* cDNAs by PCR. *Paratrimastix pyriformis* and *Monocercomonoides* sp. PA203 cDNAs were prepared from 100 mL of bacterized culture and 1000 mL of culture filtered according to Hampl et al. [57], respectively. Isolations of total RNA were performed using TRI Reagent RNA Isolation Reagent (Sigma, T9424). Extractions of eukaryotic mRNA from total RNA were done using a Dynabeads mRNA Purification Kit (Life Technologies, 61006). The SMARTer PCR cDNA Synthesis Kit (Clontech, 634925) was used for cDNA synthesis following by cDNA amplification with an Advantage 2 PCR Kit (Clontech, 639207) using 21 cycles (*Paratrimastix*) and 19 cycles (*Monocercomonoides*) of amplification.

The following primers were used for amplifications of full-length cDNAs of the ADI pathway genes of *Paratrimastix* and *Monocercomonoides* (restriction sites NdeI, VspI, and BamHI are in bold): Tp OTC-F (TAA**CATATG**CCTCGCCACCTTACCAAGAT), Tp OTC-R (TAA**GGATCC**GTCAAGGAGGGGCTGGCCCA), Tp CK-F (TAA**CATATG**CGTATCCTCATCGCTCTCG), Tp CK-R (TAA**GGATCC**GGCGACAATGTGGGTACCAG), PA203 ADI-F (TAA**CATATG**ATGCAAGATATTCACGTTCC), PA203 ADI-R (TAA**GGATCC**CTGATTTCCCAGAGATGCTA), PA203 OTC-F (ATC**ATTAAT**ATGTCCGCTCCCGTTAGACA), PA203 OTC-R (TAA**GGATCC**CTCAATGGTCATTTTCTTGT), PA203 CK-F (CACTTCACATTA**CATATG**GTGAGAATTTTAATTGCTC), PA203 CK-R (CGTATGGGTA**GGATCC**TGGAACAATGTGAGTTCCTT). For transfection of *Trichomonas vaginalis,* the genes were cloned into TagVag2 plasmid vector [58] using restriction digestion and ligation, or directly using the In-fusion HD Cloning Kit (Clontech, 639648) in the case of *Monocercomonoides* CK. Lab-grown chemically

competent *Escherichia coli* XL1 cells were used for transformations with ligation mixtures, whereas Stellar competent cells (Clontech, 636763) were used for transformation with the in-fusion reactions. Bacterial clones were checked by colony PCR for the presence of the plasmids followed by sequencing of isolated plasmids.

Selectable transfection of *Trichomonas vaginalis*

Despite extensive efforts, we did not achieve either stable or transient transfection of *Paratrimastix pyriformis* and *Monocercomonoides* sp. PA203 with plasmid vectors specifically prepared for those two organisms (data not shown). Therefore, the *Trichomonas vaginalis* heterologous expression system was used to infer the subcellular localizations of the *Paratrimastix* and *Monocercomonoides* enzymes. Versions of ADI pathway genes with a C-terminal 2xHA-tag were electroporated into *Trichomonas* cells according to the protocol described by Sutak et al. [59]. Briefly, 250 mL of *Trichomonas vaginalis* T1 culture (strain kindly provided by Michaela Marcinčiková, Dept. of Parasitology, Charles University) was used for two electroporations performed for each of the genes. Cells were electroporated with 30 μg of TagVag2 plasmid isolated using the Wizard Plus Midipreps DNA Purification System (Promega, A7640). The exponential protocol (350 V, 975 μF, ∞ Ω, 4 mm cuvette) of the GenePulser Xcell Electroporation System (Biorad, 165–2660) was used for each transfection. Trichomonads were selected with 200 μg/ml of G418 (ZellBio, G-418-5) for at least five passages. Expression of the proteins was analyzed by Western blotting of cell homogenates (data not shown) and immunofluorescence microscopy with antibody.

Immunofluorescence microscopy

ADI pathway proteins of *Paratrimastix* and *Monocercomonoides* were identified in *Trichomonas* cells using an anti-HA rat monoclonal antibody (Roche, 11867423001). An antibody raised against malic enzyme, a hydrogenosomal marker in *Trichomonas vaginalis* [60], was used for double-labeling (antibody kindly provided by prof. Jan Tachezy, Dept. of Parasitology, Charles University). Alexa Fluor-488 goat anti-rat (green) and Alexa Fluor-594 goat anti-rabbit (red) (Life Technologies, A-11006 and A-11037) were used as secondary antibodies. Immunostaining was done according to Sagolla et al. [61] on superfrost microscopic slides coated with poly-L-lysine (Sigma, P8920). Preparations were counterstained with DAPI in Vectashield mounting medium (Vector Laboratories, H – 1200) and observed using a IX81 fluorescent microscope (Olympus) equipped with an IX2-UCB camera. Images were processed using Cell^R software (Olympus) and ImageJ 1.42q.

Additional files

Additional file 1: Phylogenetic tree of OTC sequences. The tree based on a 242 positions long protein alignment of 444 sequences was constructed in RAxML using the LG4X + Γ model of substitution. Eukaryotic taxa are highlighted in different colors according to the major group they belong to. The color code is the same as in Fig. 1. The values at nodes represent RAxML bootstrap support/IQ-TREE bootstrap support. Only values above 50 % are shown. Black circles indicate support of 100 %/100 %. Species with multiple sequences included: *Alexandrium tamarense* 1 – CAMPEP 0186340278; *Alexandrium tamarense* 2 – CAMPEP 0186191854; *Alexandrium tamarense* 3 – CAMPEP 0186247540; *Durinskia baltica* 1 – CAMPEP 0200033980; *Durinskia baltica* 2 – CAMPEP 0200081736; *Euglena gracilis* 1 – c20598 g1 i1; *Euglena gracilis* 2 – c34673 g1 i6; *Eutreptiella gymnastica*-like 1 – CAMPEP 0200420840; *Eutreptiella gymnastica*-like 2 – CAMPEP 0200409666; *Karenia brevis* 1 – CAMPEP 0188881430; *Karenia brevis* 2 – CAMPEP 0188950444; *Karlodinium micrum* 1 – CAMPEP 0200795676; *Karlodinium micrum* 2 – CAMPEP 0200767534. The tree is rooted with sequences of bacterial aspartate carbamoyltransferase (ATC; EC 2.1.3.2). (PDF 506 kb)

Additional file 2: Phylogenetic tree of CK sequences. The tree based on a 251 positions long protein alignment of 256 sequences was constructed in RAxML using the LG4X+ Γ model of substitution. Eukaryotic taxa are highlighted in different colors according to the major group they belong to. The color code is the same as in Fig. 1. The values at nodes represent RAxML bootstrap support/IQ-TREE bootstrap support. Only values above 50 % are shown. Black circles indicate support of 100 %/100 %. Vertical black bars indicate well-supported eukaryotic clades: Ch – Chlorophyta; Din – Dinoflagellata; Dip – Diplomonadida; Pa – Parabasalia; Pr – Preaxostyla. Species with multiple sequences included: *Giardia intestinalis* 1 – GSB 16453; *Giardia intestinalis* 2 – GL50803 16453; *Thalassiosira pseudonana* 1 – GI 223995860; *Thalassiosira pseudonana* 2 – GI 224000745; *Trichomonas vaginalis* 1 – TVAG 420500; *Trichomonas vaginalis* 2 – TVAG 261970; *Trichomonas vaginalis* 3 – TVAG 420510. The tree is unrooted. (PDF 482 kb)

Additional file 3: Phylogenetic trees of gene partitions used for concatenation. The values at nodes represent maximum likelihood bootstrap percentages. Eukaryota highlighted in grey. Red taxon names indicate removed sequences. Positions of the particular genes in the alignment: ADI 0–257, OTC 258–499, CK 500–750. The trees are unrooted. (PDF 270 kb)

Additional file 4: Phylogenetic tree of concatenated ADI, OTC, and CK sequences. The tree based on a 750 positions long protein alignment of 67 sequences was constructed in RAxML using LG4X + Γ model. Eukaryotic taxa are highlighted in different colors according to the major group they belong to. The color code is the same as in Fig. 1. The values at nodes represent RAxML bootstrap support/IQ-TREE bootstrap support. Only values above 50 % are shown. Black circles indicate support 100 %/100 %. The tree is unrooted. (PDF 272 kb)

Additional file 5: Phylogenetic tree of concatenated ADI, OTC, and CK sequences with Bacteria and Archaea removed. The tree based on a 750 positions long protein alignment of 15 sequences was constructed in RAxML using LG4X + Γ model. The tree is unrooted. (PDF 2 kb)

Additional file 6: The probability of mitochondrial localization of ADI pathway enzymes in *Monocercomonoides* sp. PA203 and *Paratrimastix pyriformis* as predicted by TargetP and MitoProt II. (DOCX 12 kb)

Additional file 7: Topology of the expected species tree of eukaryotes. The tree is unrooted. (PDF 2 kb)

Acknowledgements

Authors would like to thank to Tomáš Pánek, Vít Céza, and Vojtěch Vacek, who provided RNA or cDNA for sequencing of *Lacusteria cypriaca*, the unidentified free-living trichomonad, *Harpagon shusteri,* and *Monocercomonoides* sp. PA203, as well as Qianqian Zhang and Aaron Heiss, who prepared the *Trimastix marina* RNAseq data.

Funding

Z. Z. and localization experiments were funded from Czech Science foundation project 13-22333P. A. K. and V. H. were supported by the project BIOCEV – Biotechnology and Biomedicine Centre of the Academy of Sciences and Charles University (CZ.1.05/1.1.00/02.0109) from the European Regional Development

Fund. I. Č., V. H. and sequencing was supported by Czech Science foundation projects P506-12-1010 and GAP506/11/1317. M. K. and P. J. K. were supported by a grant from the Tula Foundation to the Centre for Microbial Diversity and Evolution. C. W. S. was supported by Natural sciences and engineering research council of Canada - Alexander Graham Bell Canadian graduate scholarship. P. J. K., A. J. R., and A. G. B. S. are Senior Fellows of the Canadian Institute for Advanced Research. Funding bodies did not play any role in the design of the study, collection, analysis, interpretation of data and in writing the manuscript.

Availability of data and materials
Sequences were deposited in GenBank under accession numbers KT883858-KT883885.

Authors' contributions
VH and LN designed the research. ZZ conducted the experiments. AK, MK, MH, CWS, AGBS, PJK, AJR, and IC took part in obtaining the sequences. LN and AK conducted the phylogenetic analyses. AK conducted the topology tests. LN and VH wrote the manuscript. All authors read, edited, and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Author details
[1]Department of Parasitology, Charles University, Faculty of Science, Prague, Czech Republic. [2]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Canada. [3]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic. [4]Department of Biology, Dalhousie University, Halifax, Canada. [5]Department of Botany, University of British Columbia, Vancouver, Canada. [6]Department of Zoology, Charles University, Faculty of Science, Prague, Czech Republic.

Received: 27 December 2015 Accepted: 28 September 2016
Published online: 06 October 2016

References
1. Cunin R, Glansdorff N, Piérard A, Stalon V. Biosynthesis and metabolism of arginine in bacteria. Microbiol Rev. 1986;50:314–52.
2. Marquis RE, Bender GR, Murray DR, Wong A. Arginine deiminase system and bacterial adaptation to acid environments. Appl Envir Microbiol. 1987;53:198–200.
3. Casiano-Colón A, Marquis RE. Role of the arginine deiminase system in protecting oral bacteria and an enzymatic basis for acid tolerance. Appl Environ Microbiol. 1988;54:1318–24.
4. Ruepp A, Soppa J. Fermentative arginine degradation in Halobacterium salinarium (formerly Halobacterium halobium): genes, gene products, and transcripts of the arcRACB gene cluster. J Bacteriol. 1996;178:4942–7.
5. Linstead D, Cranshaw MA. The pathway of arginine catabolism in the parasitic flagellate Trichomonas vaginalis. Mol Biochem Parasitol. 1983;8:241–52.
6. Yarlett N, Lindmark DG, Goldberg B, Moharrami M, Bacchi CJ. Subcellular Localization of the Enzymes of the Arginine Dihydrolase Pathway in Trichomonas vaginalis and Tritrichomonas foetus. J Eukaryot Microbiol. 1994;41:554–9.
7. Schofield PJ, Edwards MR, Matthews J, Wilson JR. The pathway of arginine catabolism in Giardia intestinalis. Mol Biochem Parasitol. 1992;51:29–36.
8. Biagini GA, Yarlett N, Ball GE, Billetz AC, Lindmark DG, Martinez MP, Lloyd D, Edwards MR. Bacterial-like energy metabolism in the amitochondriate protozoon Hexamita inflata. Mol Biochem Parasitol. 2003;128:11–9.
9. Andersson JO, Sjögren AM, Horner DS, Murphy CA, Dyal PL, Svärd SG, Logsdon JM, Ragan MA, Hirt RP, Roger AJ. A genomic survey of the fish parasite Spironucleus salmonicida indicates genomic plasticity among diplomonads and significant lateral gene transfer in eukaryote genome evolution. BMC Genomics. 2007;8:51.
10. Zhang Q, Táborský P, Silberman JD, Pánek T, Čepička I, Simpson AGB. Marine Isolates of Trimastix marina Form a Plesiomorphic Deep-branching Lineage within Preaxostyla, Separate from Other Known Trimastigids (Paratrimastix n. gen.). Protist. 2015;166:468–91.
11. Simpson AGB. Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). Int J Syst Evol Microbiol. 2003;53(Pt 6):1759–77.
12. Yarlett N, Martinez MP, Ali Moharrami M, Tachezy J. The contribution of the arginine dihydrolase pathway to energy metabolism by Trichomonas vaginalis. Mol Biochem Parasitol. 1996;78:117–25.
13. Morada M, Smid O, Hampl V, Sutak R, Lam B, Rappelli P, Dessì D, Fiori PL, Tachezy J, Yarlett N. Hydrogenosome-localization of arginine deiminase in Trichomonas vaginalis. Mol Biochem Parasitol. 2011;176:51–4.
14. Ringqvist E, Palm JED, Skarin H, Hehl AB, Weiland M, Davids BJ, Reiner DS, Griffiths WJ, Eckmann L, Gillin FD, Svärd SG. Release of metabolic enzymes by Giardia in response to interaction with intestinal epithelial cells. Mol Biochem Parasitol. 2008;159:85–91.
15. Stadelmann B, Hanevik K, Andersson MK, Bruserud O, Svärd SG. The role of arginine and arginine-metabolizing enzymes during Giardia - host cell interactions in vitro. BMC Microbiol. 2013;13:256.
16. Stadelmann B, Merino MC, Persson L, Svärd SG. Arginine consumption by the intestinal parasite Giardia intestinalis reduces proliferation of intestinal epithelial cells. PLoS One. 2012;7, e45325.
17. Banik S, Renner Viveros P, Seeber F, Klotz C, Ignatius R, Aebischer T. Giardia duodenalis arginine deiminase modulates the phenotype and cytokine secretion of human dendritic cells by depletion of arginine and formation of ammonia. Infect Immun. 2013;81:2309–17.
18. Schubert KR. Products of Biological Nitrogen Fixation in Higher Plants: Synthesis, Transport, and Metabolism. Annu Rev Plant Physiol. 1986;37:539–74.
19. Allen AE, Dupont CL, Oborník M, Horák A, Nunes-Nesi A, McCrow JP, Zheng H, Johnson DA, Hu H, Fernie AR, Bowler C. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature. 2011;473:203–7.
20. Dagenais-Bellefeuille S, Morse D. Putting the N in dinoflagellates. Front Microbiol. 2013;4:369.
21. Shafer J, Thompson JF. Arginine desimidase in Chlorella. Phytochemistry. 1968;7:391–9.
22. Laliberte G, Hellebust JA. Arginine utilization by Chlorella autotrophica and Chlorella saccharophila. Physiol Plant. 1990;79:57–64.
23. Sussenbach JS, Strijkert PJ. Arginine Metabolism in Chlamydomonas reinhardi. On the Regulation of the Arginine Biosynthesis. Eur J Biochem. 1969;8:403–7.
24. Laliberte G, Hellebust JA. The phylogenetic significance of the distribution of arginine deiminase and arginase in the Chlorophyta. Phycologia. 1991;30:145–50.
25. Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, De Clerck O. Phylogeny and Molecular Evolution of the Green Algae. CRC Crit Rev Plant Sci. 2012;31:1–46.
26. Zúñiga M, Pérez G, González-Candelas F. Evolution of arginine deiminase (ADI) pathway genes. Mol Phylogenet Evol. 2002;25:429–44.
27. Fucikova K, Leliaert F, Cooper ED, Skaloud P, D'hondt S, De Clerck O, Gurgel F, Lewis LA, Lewis PO, Lopez-Bautista J, Delwiche CF, Verbruggen H. New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. Front Ecol Evol. 2014;2:63.
28. Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, le Gall L, Lynn DH, McManus H, Mitchell EAD, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick L, Schoch CL, Smirnov A, Spiegel FW. The revised classification of eukaryotes. J Eukaryot Microbiol. 2012;59:429–514.
29. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc. 2007;2:953–71.
30. Claros MG, Vincens P. Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem. 1996;241:779–86.
31. Kessler E. Hydrogenase, photoreduction, and anaerobic growth. Bot Monogr Oxford. 1974;10:456–73.
32. Chávez LA, Balamuth W, Gong T. A light and electron microscopical study of a new, polymorphic free-living amoeba, Phreatamoeba balamuthi n. g., n. sp. J Protozool. 1986;33:397–404.
33. Cavalier-Smith T. The excavate protozoan phyla Metamonada Grasse emend. (Anaeromonadea, Parabasalia, Carpediemonas, Eopharyngia) and Loukozoa emend. (Jakobea, Malawimonas): their evolutionary affinities and new higher taxa. Int J Syst Evol Microbiol. 2003;53:1741–58.

34. Pánek T, Silberman JD, Yubuki N, Leander BS, Cepicka I. Diversity, Evolution and Molecular Systematics of the Psalteriomonadidae, the Main Lineage of Anaerobic/Microaerophilic Heteroloboseans (Excavata: Discoba). Protist. 2012;163:807–31.
35. Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, Soanes DM, van der Giezen M, Roger AJ. A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist Pygsuia. Curr Biol. 2014;24:1176–86.
36. Atteia A, van Lis R, Tielens AGM, Martin WF. Anaerobic energy metabolism in unicellular photosynthetic eukaryotes. Biochim Biophys Acta. 2013;1827: 210–23.
37. Lepère C, Domaizon I, Hugoni M, Vellet A, Debroas D. Diversity and Dynamics of Active Small Microbial Eukaryotes in the Anoxic Zone of a Freshwater Meromictic Lake (Pavin, France). Front Microbiol. 2016;7:130.
38. Andersen RA. Biology and systematics of heterokont and haptophyte algae. Am J Bot. 2004;91:1508–22.
39. Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Žárský V, Barlow LD, Herman EK, Soukal P, Hroudová M, Doležal P, Stairs CW, Roger AJ, Eliáš M, Dacks JB, Vlček Č, Hampl V, Huynen MA, Duarte I, Szklarczyk R, Tovar J, León-Avila G, Sánchez LB, Sutak R, Tachezy J, van der Giezen M, Hernández M, Müller M, et al. A Eukaryote without a Mitochondrial Organelle. Curr Biol. 2016;26:1274–84.
40. Derelle R, Torruella G, Klimeš V, Brinkmann H, Kim E, Vlček Č, Lang BF, Eliáš M. Bacterial proteins pinpoint a single eukaryotic root. Proc Natl Acad Sci. 2015;201420657.
41. Williams TA, Foster PG, Cox CJ, Embley TM. An archaeal origin of eukaryotes supports only two primary domains of life. Nature. 2013;504:231–6.
42. Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature. 2015;521:173–9.
43. Burki F. The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harb Perspect Biol. 2014;6:a016147.
44. He D, Fiz-Palacios O, Fu C-J, Fehling J, Tsai C-C, Baldauf SL. An alternative root for the eukaryote tree of life. Curr Biol. 2014;24:465–70.
45. Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, Kuo A, Minovitsky S, Nikitin R, Ohm RA, Otillar R, Poliakov A, Ratnere I, Riley R, Smirnova T, Rokhsar D, Dubchak I. The genome portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res. 2012;40(Database issue):D26–32.
46. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B, Bidle KD, Cameron CT, Campbell L, Caron DA, Cattolico RA, Collier JL, Coyne K, Davy SK, Deschamps P, Dyhrman ST, Edvardsen B, Gates RD, Gobler CJ, Greenwood SJ, Guida SM, Jacobi JL, Jakobsen KS, James ER, Jenkins B, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. PLoS Biol. 2014;12:e1001889.
47. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997;3389–3402.
48. Zubáčová Z, Novák L, Bublíková J, Vacek V, Fousek J, Rídl J, Tachezy J, Doležal P, Vlček Č, Hampl V. The Mitochondrion-Like Organelle of Trimastix pyriformis Contains the Complete Glycine Cleavage System. PLoS One. 2013;8, e55417.
49. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
50. Vaidya G, Lohman DJ, Meier R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 2011;27:171–80.
51. NovakData.zip [http://protistologie.cz/hampllab/NovakData.zip]. Accessed 4 Oct 2016.
52. Abascal F, Zardoya R, Posada D. ProtTest: Selection of best-fit models of protein evolution. Bioinformatics. 2005;21:2104–5.
53. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
54. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015;32:268–74.
55. Shimodaira H. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 2002;51:492–508.
56. Strimmer K, Rambaut A. Inferring confidence sets of possibly misspecified gene trees. Proc Biol Sci. 2002;269:137–42.
57. Hampl V. Inference of the Phylogenetic Position of Oxymonads Based on Nine Genes: Support for Metamonada and Excavata. Mol Biol Evol. 2005;22: 2508–18.
58. Hrdy I, Hirt RP, Dolezal P, Bardonová L, Foster PG, Tachezy J, Embley TM. Trichomonas hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. Nature. 2004;432:618–22.
59. Sutak R, Dolezal P, Fiumera HL, Hrdy I, Dancis A, Delgadillo-Correa M, Johnson PJ, Müller M, Tachezy J. Mitochondrial-type assembly of FeS centers in the hydrogenosomes of the amitochondriate eukaryote Trichomonas vaginalis. Proc Natl Acad Sci U S A. 2004;101:10368–73.
60. Drmota T. Iron-ascorbate cleavable malic enzyme from hydrogenosomes of Trichomonas vaginalis: purification and characterization. Mol Biochem Parasitol. 1996;83:221–34.
61. Sagolla MS, Dawson SC, Mancuso JJ, Cande WZ. Three-dimensional analysis of mitosis and cytokinesis in the binucleate parasite Giardia intestinalis. J Cell Sci. 2006;119:4889–900.

Fe–S Cluster Assembly in Oxymonads and Related Protists

Vojtěch Vacek,[1] Lukáš V.F. Novák,[1] Sebastian C. Treitli,[1] Petr Táborský,[2] Ivan Čepička,[2] Martin Kolísko,[3,4] Patrick J. Keeling,[4] and Vladimír Hampl*,[1]

[1]Department of Parasitology, Faculty of Science, Charles University, BIOCEV, Vestec, Czech Republic

[2]Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic

[3]Institute of Parasitology, Biology Centre, Czech Academy of Science, České Budějovice, Czech Republic

[4]Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

***Corresponding author:** E-mail: vlada@natur.cuni.cz.

Associate editor: Iñaki Ruiz-Trillo

Abstract

The oxymonad *Monocercomonoides exilis* was recently reported to be the first eukaryote that has completely lost the mitochondrial compartment. It was proposed that an important prerequisite for such a radical evolutionary step was the acquisition of the SUF Fe–S cluster assembly pathway from prokaryotes, making the mitochondrial ISC pathway dispensable. We have investigated genomic and transcriptomic data from six oxymonad species and their relatives, composing the group Preaxostyla (Metamonada, Excavata), for the presence and absence of enzymes involved in Fe–S cluster biosynthesis. None possesses enzymes of mitochondrial ISC pathway and all apparently possess the SUF pathway, composed of SufB, C, D, S, and U proteins, altogether suggesting that the transition from ISC to SUF preceded their last common ancestor. Interestingly, we observed that SufDSU were fused in all three oxymonad genomes, and in the genome of *Paratrimastix pyriformis*. The donor of the SUF genes is not clear from phylogenetic analyses, but the enzyme composition of the pathway and the presence of SufDSU fusion suggests Firmicutes, Thermotogae, Spirochaetes, Proteobacteria, or Chloroflexi as donors. The inventory of the downstream CIA pathway enzymes is consistent with that of closely related species that retain ISC, indicating that the switch from ISC to SUF did not markedly affect the downstream process of maturation of cytosolic and nuclear Fe–S proteins.

Key words: **Preaxostyla, SUF, amitochondriate, CIA, oxymonads.**

Letter

Iron–sulfur clusters are small inorganic prosthetic groups, which are among the most ancient and versatile cofactors. Their main function is mediating electron transport, which makes them a key part of many important processes such as photosynthesis, respiration, DNA replication and repair, and regulation of gene expression (Rudolf et al. 2006; Fuss et al. 2015; Paul and Lill 2015).

There are three pathways for the Fe–S clusters synthesis known in prokaryotes—NIF (nitrogen fixation), ISC (iron sulfur cluster), and SUF (sulfur utilization factor). The basic process of the Fe–S cluster biogenesis is similar in all three (Roche et al. 2013·). Sulfur (S^{2-}) is provided by cysteine desulfurase (NifS, IscS, SufS). The source of iron (Fe^{2+}) is unclear, however, for the mitochondrial ISC pathway frataxin is expected to be the provider (Pastore and Puccio 2013; Yoon et al. 2015). The sulfur and iron ions are first combined into a cluster on a scaffold protein (NifU, IscU, SufB–SufD complex), from which the cluster is transferred onto an apoprotein.

In eukaryotic cells, three compartments have distinct pathways for Fe–S cluster synthesis. Mitochondria typically use the ISC pathway, which was inherited from the alphaproteobacterial endosymbiont (Tachezy et al. 2001; Braymer and Lill 2017). This holds also for most mitochondrion-related organelles including the mitosomes of *Giardia intestinalis* (Tovar et al. 2003) and microsporidia (Katinka et al. 2001; Goldberg et al 2008), and hydrogenosomes of *Trichomonas vaginalis* (Sutak et al. 2004). Exceptions to this rule are found in mitochondrion-related organelles of *Pygsuia biforma*, *Mastigamoeba balamuthi*, and *Entamoeba histolytica* (Ali et al. 2004; van der Giezen et al. 2004; Mi-ichi et al. 2009; Nyvltova et al. 2013; Stairs et al. 2014), which contain SUF, NIF, or possibly none of these pathways, respectively. Eukaryotic plastids contain the SUF pathway, which was inherited from the cyanobacterial ancestor (Balk and Pilon 2011).

In the eukaryotic cytosol, the Fe–S cluster-containing proteins are formed by a cytosolic iron–sulfur cluster assembly (CIA), which is also responsible for maturation of nuclear Fe–S proteins. In yeast and human, the pathway contains at least eleven essential proteins (Sharma et al. 2010; Netz et al. 2014; Lill et al. 2015). CIA is unique to eukaryotes and most of its components do not have prokaryotic homologs, apart from Nbp35 (Boyd et al. 2009) and Cia2 (Tsaousis et al. 2014). It has been experimentally shown that the mitochondrial ISC pathway is necessary for the function of the CIA, probably because it synthesizes and transports an uncharacterized sulfur containing precursor to the cytosol (Kispal et al. 1999; Gerber et al. 2004; Biederbick et al. 2006; Pondarré et al. 2006). Dependency on ISC is interpreted as a major reason for the retention of mitochondrion-related organelles in anaerobic eukaryotes (Williams et al. 2002). Maturation of the cytosolic

Fe–S proteins by the CIA pathway starts with the formation of [4Fe–4S] cluster on the Cfd1-Nbp35 scaffold (Hausmann et al. 2005; Netz et al. 2012), transfer of electrons from NADPH is mediated by Dre2 and diflavin reductase Tah18 is required for this process (Zhang et al. 2008; Netz et al. 2010). The [4Fe–4S] cluster from Cfd1-Nbp35 is transferred to a target apoprotein by Nar1 and the late-acting CIA components Cia1, Cia2, and Met18, which form the so-called CIA-targeting complex (Lill et al. 2015).

A unique combination of the Fe–S cluster assembly enzymes has been found in a flagellate *Monocercomonoides exilis* (strains PA203; Treitli et al. 2018) from the group of Oxymonadida (Karnkowska et al. 2016). *Monocercomonoides exilis* contains the CIA pathway, however, the ISC pathway is absent together along with all other mitochondrial proteins. As there is no microscopic evidence for the existence of a mitochondrion, this is interpreted as showing the mitochondrion has been lost altogether (Karnkowska et al. 2016), which makes this oxymonad unique among eukaryotes. Instead of ISC, *M. exilis* contains SUF pathway, which is slightly reduced compared with bacterial pathways, containing only three proteins—SufB, SufC, and SufSU. SufSU represents a fusion of SufS (cysteine desulfurase) and SufU (an enhancer of SufS in prokaryotes; Albrecht et al. 2010, 2011; Riboldi et al. 2011; Karnkowska et al. 2016).

The *M. exilis* SUF proteins are not specifically related to plastid homologues, or homologues from any other microbial eukaryotes, but rather to enzymes found in eubacteria, and to homologues in the transcriptome of *Paratrimastix pyriformis*, a sister taxon of oxymonads (Zhang et al. 2015; Karnkowska et al. 2016). It has been proposed that the pathway was acquired by horizontal gene transfer (HGT) from a eubacterium in the common ancestor of *Monocercomonoides* and *Paratrimastix*. This apparently preceded the loss of mitochondria in *M. exilis*, because *P. pyriformis* retains a mitochondrion-related organelle (Hampl et al. 2008; Zubáčová et al. 2013). Localization of the SUF pathway in *P. pyriformis* is unknown and in *M. exilis* heterologous localizations in *Saccharomyces cerevisiae* and *T. vaginalis* suggest cytosolic localization (Karnkowska et al. 2016).

Oxymonads and *P. pyriformis* are classified into the group Preaxostyla within phylum Metamonada (supergroup Excavata; Hampl et al. 2009; Adl et al. 2012). All representatives of two other Metamonada lineages (Parabasalia and Fornicata) contain the ancestral ISC pathway and no genes for enzymes in the SUF pathway have been observed in these lineages, indicating that this modification of Fe–S cluster assembly is specific to Preaxostyla. This may have served as a precondition for the mitochondrial loss in oxymonads. To further reveal the evolutionary history of this unique evolutionary switch, we investigated genomes and transcriptomes of 16 members of Preaxostyla for presence of genes and/or transcripts involved in Fe–S cluster synthesis.

Results and Discussion

The following Preaxostyla data sets were investigated in this study: genomic assemblies of *M. exilis* strain PA203 (Karnkowska et al. 2016), *Blattamonas nauphoetae* strain NAU3, and *P. pyriformis* strain ATCC 50935, single cell genome assembly of *Streblomastix strix*, three single cell transcriptome assemblies of *Saccinobaculus doroaxostylus* (SD1, SD2, SDN), three single cell transcriptome assemblies of *Saccinobaculus ambloaxostylus* (Amblo-1, Amblo-5, Amblo-5), one single cell transcriptome assembly of *Oxymonas* sp., two single cell transcriptome assemblies of *Streblomastix* sp. (Streblo-1, Streblo-4), one single cell transcriptome assembly of *Pyrsonympha* sp., transcriptome assembly of *Trimastix marina* strain PCT (Leger et al. 2017), and transcriptome assembly of trimastigid "MORAITIKA". Quality and coverage varied extremely between data sets, which was probably the main reason for the lack of some genes/transcripts especially in the single cell genomes and transcriptomes.

The data were searched for all known components of ISC, SUF, CIA, and NIF pathways. In all examined data sets, we were unable to identify any gene involved in the ISC and NIF pathways, but we identified genes or gene fragments for proteins involved in the SUF and CIA pathways in all examined organisms with exception of "Streblo-4" (figs. 1 and 2). Mitochondrial targeting peptides were not predicted in any complete SUF proteins (supplementary table S1, Supplementary Material online), indicating cytosolic localization of these proteins.

Contaminants among the CIA proteins were not expected, because they have no close prokaryotic homologues. To filter prokaryotic contamination of SUF genes/transcripts and to exclude those that were severely truncated, sequences with >70% nucleotide similarity to prokaryotic sequences in NCBI and also sequences shorter than 65 aa were not included in further analyses. Phylogenetic trees were then constructed for individual SUF components (supplementary figs. S1–S4, Supplementary Material online). Every protein tree resolved a major Preaxostyla clade (shown in green in supplementary figs. S1–S4, Supplementary Material online). Sequences derived from the genomic assemblies were present only in this major clade, suggesting that this clade represents *bona fide* Preaxostyla genes. Species composition of this clade in every protein tree suggests that it was acquired prior to the last common ancestor of Preaxostyla.

Some sequences branched robustly (ML bootstrap >90%) outside this major clade (shown in red in supplementary figs. S1–S4, Supplementary Material online). Origin of these sequences is unclear, they may result from prokaryotic contaminations or more recent lineage specific HGT. For this reason, these sequences were not included in the concatenation analysis. Such outlying SUF proteins were especially common in the data set of *Oxymonas* sp. (supplementary figs. S1–S4, Supplementary Material online). These were all unique to *Oxymonas* and no closely related orthologues were found in any other oxymonad, suggesting that this data set contains high level of prokaryotic contamination. In contrast, we identified six closely related SufS genes from *S. doroaxostylus* isolate SDN, *Streblomastix* sp. (Streblo-1), *S. ambloaxostylus* (Amblo-5), and trimastigid "MORAITIKA." They formed a well-supported clade (ML bootstrap 100) that was deeply nested in a clade containing mixture of Archaea and Eubacteria (supplementary fig. S4, Supplementary Material online). These

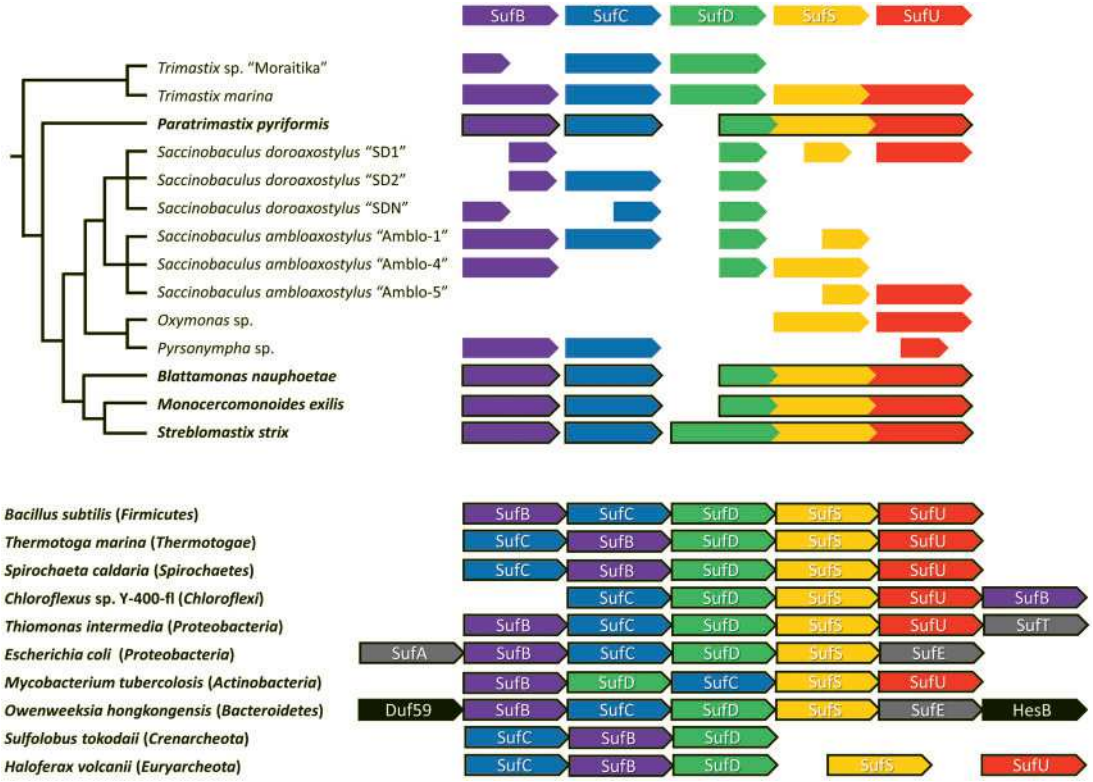


Fig. 1. Inventory of SUF proteins in Preaxostyla. The scheme shows SUF genes/transcripts identified in the members of Preaxostyla. The relationship within this groups is indicated by the tree. For organisms in bold, genomic data were investigated, in others transcriptomic or single cell transcriptomic data sets were used. Completeness of a gene/transcript is indicated by the length of the arrow. The order of Preaxostyla genes does not reflect their order in the genome. Gene fusions are marked by fused arrows. At the bottom are given schemes of typical SUF gene operons in representatives of prokaryotic groups, see supplementary figure S5, Supplementary Material online for broader prokaryotic representation.

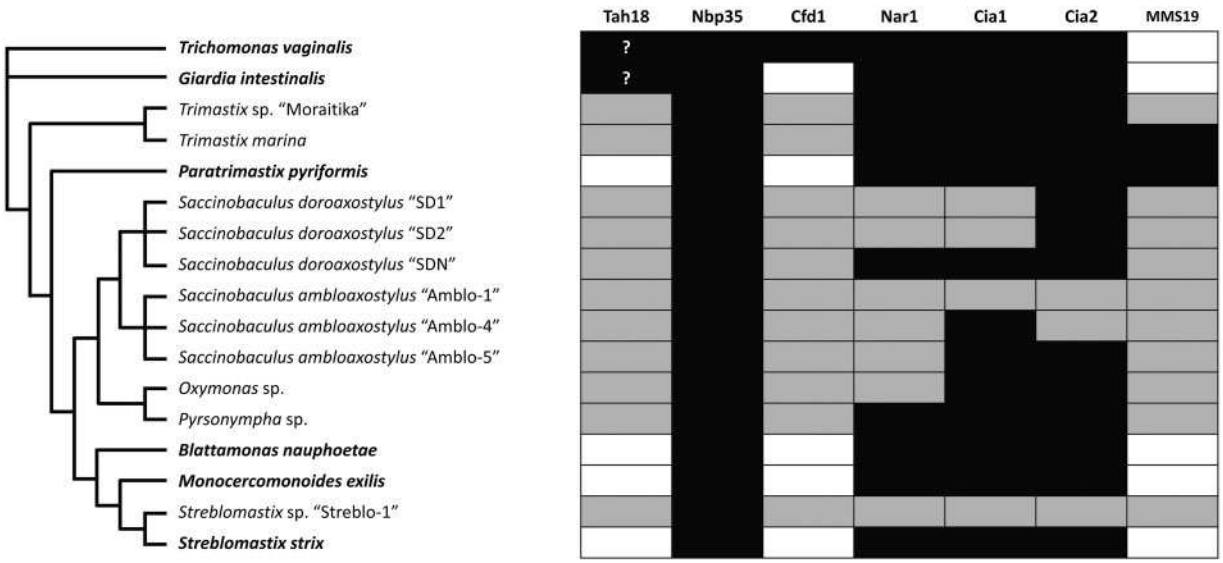


Fig. 2. Inventory of CIA proteins in Preaxostyla. Scheme shows the presence (black) or absence (white/gray) of CIA genes/transcripts in Preaxostyla with reference to Metamonada represented by *G. intestinalis* and *T. vaginalis*. White/grey shading indicates that the gene was not identified in available genome/transcriptome, respectively. The gene inventory of *T. vaginalis* and *G. intestinalis* was taken from Pyrih et al. 2016. Question marks indicate uncertain orthology to Tah18.

sequences may represent contaminants from closely related prokaryotes, but it is also possible given the fact that these genes do not share a high level of identity and were found in a larger number of species, that they represent a second and independent acquisition of SufS by a subset of Preaxostyla. It should be noted, however, that these homologues were identified only in transcriptomes with overall low quality and high level of contamination, and most critically that

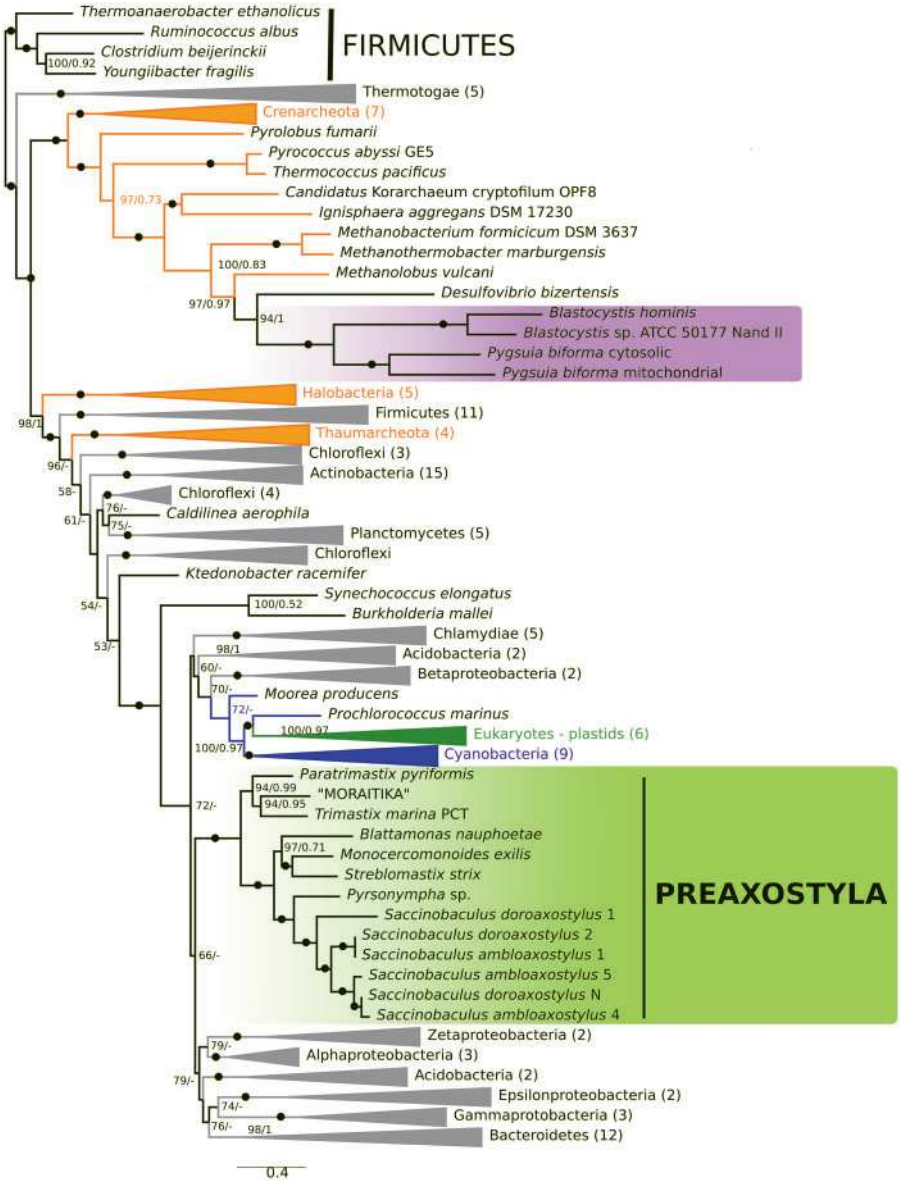


Fig. 3. Phylogenetic analysis of concatenated SufB, C, D, and S proteins. The topology of the tree was calculated by ML in IQ-TREE using partition-specific models. Numbers at nodes represent statistical support in regular ML bootstraps/Bayesian posterior probabilities. The support 99/0.99 and higher is indicated by filled circles, values <50 and 0.5 are not shown.

the sequences putatively ascribed to *Streblomastix* do not contain evidence of the alternative genetic code typical for this oxymonad (Keeling and Leander 2003).

A concatenated tree of SufB, SufC, SufD, and SufS was constructed including sequences from "green clades" with reliable Preaxostyla origin (fig. 3). In this tree, all Preaxostyla formed a single and well-supported clade (RaxML/MrBayes support 100/1) with an internal topology consistent with the relationships among Preaxostyla (Treitli et al. 2018). This result strongly suggests that the whole pathway originated in their common ancestor and was inherited vertically since then.

Three proteins of the pathway (SufD, S, and U) are fused (SufDSU) in *M. exilis*, *B. nauphoetae*, *S. strix*, and *P. pyriformis*. The homology of the N-terminal part of this protein with SufD has not been recognized previously (Karnkowska et al. 2016). Similar fusion is also present in the transcriptome of *T. marina*, where we identified a fusion of SufSU but SufD is on an independent contig. No fusions were evident in other data sets, but this may be a result of the fragmented nature of the data, making fusions hard to detect. A scheme of the detected SUF genes/transcripts and their fusions is shown in figure 1 (see supplementary fig. S5, Supplementary Material online for broader representation of prokaryotes). The phylogeny of SUFs (fig. 3) was unable to resolve the position of Preaxostyla sequences within prokaryotes. However, the arrangement of SUF operons in sequenced prokaryotic genomes suggests Firmicutes, Thermotogae, Spirochaetes, Proteobacteria, or Chloroflexi as probable donors. Operons in these groups are consistent with the SUF pathway composition as well as the SufDSU gene order found in Preaxostyla fusion genes.

The inventory of the CIA genes in Preaxostyla is in general very similar to that of other metamonads, including Nbp35,

Cia1, Nar1, and Cia2 (Pyrih et al. 2016; fig. 2). Preaxostyla lack proteins associated with the mitochondrion (Erv1 and Atm1), and proteins Dre2 and Tah18. Their absence is not surprising as the Dre2 is often missing in anaerobes (Basu et al. 2014; Tsaousis et al. 2014) and neither Erv1 nor Atm1 was found in other metamonads (Pyrih et al. 2016). The primary function of Tah18 is to provide electrons for Dre2 (Netz et al. 2010), so in the absence of Dre2 it is reasonable that Tah18 was probably lost as well. We were not able to identify MMS19 in any of the studied oxymonads, but protein containing N-terminal MMS19 domain was present in *P. pyriformis* (fig. 2). The conserved inventory of CIA proteins in Preaxostyla contrasts with the major switch of the upstream Fe–S cluster assembly pathway in this group and indicates functional robustness of the CIA pathway.

Materials and Methods

For single cell transcriptomes (*S. doroaxostylus*, *S. ambloaxostylus*, *Oxymonas* sp., *Streblomastix* spp., and *Pyrsonympha*), cells were manually picked by micropipette, washed 1–2 times, and then deposited directly into single cell lysis buffer and frozen in −80 °C freezer. Single cell cDNA was then amplified following Picelli et al. (2014) and Kolisko et al. (2014) protocols. Illumina Nextera XT protocol was used for sequencing library construction. Transcriptomes were assembled by Trinity 2.0.6 (Grabherr et al. 2011) and for quality trimming trimmomatic0.32 (Bolger et al. 2014) with default settings was used.

The trimastigid "MORAITIKA" was maintained as a monoeukaryotic polyxenic culture in the ATCC medium 1525 at room temperature. Total RNA was isolated from 300 ml of culture using TRI Reagent (Sigma). Isolated RNA was purified by Qiagen RNeasy Mini Kit (Qiagen) and RNase-Free DNase Set (Qiagen) according to the manufacturer's protocol. Total RNA was sent to EMBL where the libraries were prepared. Contigs were assembled by Trinity 2014-04-13p1 (Grabherr et al. 2011), quality trimming was done by fastx version 0.0.13 (fastq_quality_filter -Q33 -q 20 -p 70), contigs shorter than 200 nt were discarded.

The single cell of *S. strix* was manually picked by micropipette from gut content of the termite *Zootermopsis angusticolis*, three times washed in Trager U media, then DNA was isolated and the whole genome was amplified using illustra Single Cell GenomiPhi DNA Amplification Kit (GE Healthcare) according to the manufacturer's protocol. The amplified DNA was purified using Agencourt AMPure XP (Beckman Coulter), and sequencing libraries were prepared using Illumina TruSeq DNA PCR-Free (Illumina) for HiSeq 2500 or with Ligation Sequencing Kit 1D (Oxford Nanopore Technologies) for Oxford Nanopore sequencing. Draft genome was assembled as a hybrid assembly using SPAdes 3.10.0 (Bankevich et al. 2012; Antipov et al. 2016). Binning of the assembled data and separation of the eukaryotic genome from bacterial sequences was done using tetranucleotide frequencies using tetraESOM method (Dick et al. 2009), together with blast analysis of the assembled data.

Blattamonas nauphoetae strain NAU3 was grown as a monoeukaryotic polyxenic culture in modified TYSGM media (Diamond 1982) without gastric mucin. The genomic DNA was sequenced using Illumina MiSeq (coverage 62x) and Oxford Nanopore MinIon (coverage 2x) technology and assembled using the SPAdes 3.7.1 (Bankevich et al. 2012; Antipov et al. 2016) followed by scaffolding with SSPACE basic V2 (Boetzer et al. 2011) using the Illumina mate-pair reads.

Paratrimastix pyriformis was grown in a monoeukaryotic polyxenic culture on rye grass cerophyll infusion (Sonneborn's Paramecium medium, ATCC #802) at room temperature. The genomic DNA was isolated using DNeasy Blood & Tissue Kit (Qiagen). The *P. pyriformis* draft genome sequence was sequenced and assembled from raw genomic reads produced by 454, Illumina HiSeq (coverage 894x), and PacBio (coverage 11x) sequencing technologies using SPAdes 3.11.1 (Bankevich et al. 2012; Antipov et al. 2016) assembly toolkit. Automatic gene prediction for the *S. strix*, *B. nauphoetae* and *P. pyriformis* draft genomes was done using Augustus 3.2.3 (Stanke and Waack 2003).

Nucleotide data sets predicted proteins were searched by TBLASTN algorithms; six-frame translations of the transcriptomes and predicted proteomes were searched by BLASTP. Proteins not identified by BLAST were searched for by HMMER. In the BLAST searches, we have used full gene inventories of *M. exilis*, *Escherichia coli*, and *Bacillus subtilis* for the SUF pathway, *S. cerevisiae*, *T. vaginalis*, *G. intestinalis*, and *E. coli* for the ISC pathway, *Azotobacter vinelandii*, *E. histolytica*, and *M. balamuthii* for the NIF pathway and *S. cerevisiae* and human for the CIA pathway. HMMER searches were performed by HMMER 3.1b2 (Eddy 2011) using curated Pfam and custom created models for the aforementioned genes. Sequences of SUF and CIA pathway genes/proteins retrieved from unpublished data sets were deposited in GenBank under accession numbers MH608120–MH608208.

Sequences were aligned by MAFFT v. 7.222 (Katoh et al. 2002) and trimmed with BMGE 1.12 software (Criscuolo and Gribaldo 2010) using blosum30 matrix. Gene fragments originating from different assemblies of the same species (namely SufC—SDN_lcl|TR24651|c0_g1_i1 and SDN_lcl|TR10532|c0_g1_i1, SufS—SD1_lcl|TR11348|c0_g2_i1 and SD1_lcl|TR17340|c0_g1_i1) were concatenated to increase phylogenetic resolution. Alignments are available upon request. Phylogenetic trees were constructed by IQ-TREE v 1.6.1 (Nguyen et al. 2015) using the best fitting models according to Bayesian information criterion predicted by ModelFinder (Kalyaanamoorthy et al. 2017)—LG4M for SufS and SufD, C20 for SufC, and EX_EHO for SufD. For analysis of the concatenated alignment, gene-partition-specific models given above were used. Bayesian analysis was performed for concatenated data set using MrBayes 3.2 (Ronquist et al. 2012) with two runs, each of four chains of 10 mil. generations, sampling frequency 500 generations and uniform WAG+gamma+covarion model for the whole concatenate. The value of the average standard deviation of split frequencies did not drop below the 0.01; however, both chains have shown stable plateau of likelihood values after 1×10^6

generations. Tree from first 1×10^6 generations were discarded as burn-in before consensus tree calculation.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by the Czech Science Foundation (project 15-16406S to V.H.), by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 771592 to VH) the Grant Agency of Charles University (Project 1584314 to V.V.), the Ministry of Education, Youth and Sports of CR within the National Sustainability Program II (Project BIOCEV-FAR) LQ1604, by the project "BIOCEV" (CZ.1.05/1.1.00/02.0109), by the Centre for research of pathogenicity and virulence of parasites reg. nr.: CZ.02.1.01/0.0/0.0/16_019/0000759, by the Natural Sciences and Engineering Council of Canada (project RGPIN-2014-03994 to P.J.K.), by a grant from the Tula Foundation to the UBC Centre for Microbial Diversity and Evolution, by Purkyne Fellowship to M.K. (Czech Academy of Sciences). Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures."

References

Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, et al. 2012. The revised classification of eukaryotes. *J Eukaryot Microbiol*. 59(5):429–493.

Albrecht AG, Netz DJA, Miethke M, Pierik AJ, Burghaus O, Peuckert F, Lill R, Marahiel MA. 2010. SufU is an essential iron–sulfur cluster scaffold protein in *Bacillus subtilis*. *J Bacteriol*. 192(6):1643–1651.

Albrecht AG, Peuckert F, Landmann H, Miethke M, Seubert A, Marahiel MA. 2011. Mechanistic characterization of sulfur transfer from cysteine desulfurase SufS to the iron–sulfur scaffold SufU in *Bacillus subtilis*. *FEBS Lett*. 585(3):465–470.

Ali V, Shigeta Y, Tokumoto U, Takahashi Y, Nozaki T. 2004. An intestinal parasitic protist, *Entamoeba histolytica*, possesses a non-redundant nitrogen fixation-like system for iron–sulfur cluster assembly under anaerobic conditions. *J Biol Chem*. 279(16):16863–16874.

Antipov D, Korobeynikov A, McLean JS, Pevzner PA. 2016. HybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* 32(7):1009–1015.

Balk J, Pilon M. 2011. Ancient and essential: the assembly of iron–sulfur clusters in plants. *Trends Plant Sci*. 16(4):218–226.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 19(5):455–477.

Basu S, Netz DJ, Haindrich AC, Herlerth N, Lagny TJ, Pierik AJ, Lill R, Lukeš J. 2014. Cytosolic iron–sulphur protein assembly is functionally conserved and essential in procyclic and bloodstream Trypanosoma brucei. *Mol Microbiol*. 93(5):897–910.

Biederbick A, Stehling O, Rosser R, Niggemeyer B, Nakai Y, Elsasser H-P, Lill R. 2006. Role of human mitochondrial Nfs1 in cytosolic iron–sulfur protein biogenesis and iron regulation. *Mol Cell Biol*. 26(15):5675–5687.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4):578–579.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

Boyd JM, Drevland RM, Downs DM, Graham DE. 2009. Archaeal ApbC/Nbp35 homologs function as iron–sulfur cluster carrier proteins. *J Bacteriol*. 191(5):1490–1497.

Braymer JJ, Lill R. 2017. Iron–sulfur cluster biogenesis and trafficking in mitochondria. *J Biol Chem*. 292(31):12754–12763.

Criscuolo A, 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(1):210.

Diamond LS. 1982. A new liquid medium for xenic cultivation of *Entamoeba histolytica* and other lumen-dwelling protozoa. *J Parasitol*. 68(5):958–959.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield JF. 2009. Community-wide analysis of microbial genome sequence signatures. *Genome Biol*. 10(8):R85.

Eddy SR. 2011. Accelerated profile HMM searches. Pearson WR, editor. *PLoS Comput Biol*. 7(10):e1002195.

Fuss JO, Tsai CL, Ishida JP, Tainer JA. 2015. Emerging critical roles of Fe–S clusters in DNA replication and repair. *Biochim Biophys Acta-Mol Cell Res*. 1853(6):1253–1271.

Gerber J, Neumann K, Prohl C, Muhlenhoff U, Lill R. 2004. The yeast scaffold proteins Isu1p and Isu2p are required inside mitochondria for maturation of cytosolic Fe/S proteins. *Mol Cell Biol*. 24(11):4848–4857.

van der Giezen M, Cox S, Tovar J. 2004. The iron–sulfur cluster assembly genes iscS and iscU of *Entamoeba histolytica* were acquired by horizontal gene transfer. *BMC Evol Biol*. 4:7.

Goldberg AV, Molik S, Tsaousis AD, Neumann K, Kuhnke G, Delbac F, Vivares CP, Hirt RP, Lill R, Embley TM. 2008. Localization and functionality of microsporidian iron–sulphur cluster assembly proteins. *Nature* 452(7187):624–628.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 29(7):644–652.

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proc Natl Acad Sci USA*. 106(10):3859–3864.

Hampl V, Silberman JD, Stechmann A, Diaz-Triviño S, Johnson PJ, Roger AJ. 2008. Genetic evidence for a mitochondriate ancestry in the "amitochondriate" flagellate *Trimastix pyriformis*. *PLoS One* 3(1):e1383.

Hausmann A, Netz DJA, Balk J, Pierik AJ, Muhlenhoff U, Lill R. 2005. The eukaryotic P loop NTPase Nbp35: an essential component of the cytosolic and nuclear iron–sulfur protein assembly machinery. *Proc Natl Acad Sci USA*. 102(9):3266–3271.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14(6):587–589.

Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Žárský V, Barlow LD, Herman EK, et al. 2016. A eukaryote without a mitochondrial organelle. *Curr Biol*. 26(10):1274–1284.

Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, et al. 2001. Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. *Nature* 414(6862):450–453.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30(14):3059–3066.

Keeling PJ, Leander BS. 2003. Characterisation of a Non-canonical genetic code in the oxymonad *Streblomastix strix*. *J Mol Biol*. 326(5):1337–1349.

Kispal G, Csere P, Prohl C, Lill R. 1999. The mitochondrial proteins Atm1p and Nfs1p are essential for biogenesis of cytosolic Fe/S proteins. *EMBO J*. 18(14):3981–3989.

Kolisko M, Boscaro V, Burki F, Lynn DH, Keeling PJ. 2014. Single-cell transcriptomics for microbial eukaryotes. *Curr Biol*. 24(22): R1081–R1082.

Leger MM, Kolisko M, Kamikawa R, Stairs CW, Kume K, Čepička I, Silberman JD, Andersson JO, Xu F, Yabuki A, et al. 2017. Organelles that illuminate the origins of Trichomonas hydrogenosomes and Giardia mitosomes. *Nat Ecol Evol*. 1(4):0092.

Lill R, Dutkiewicz R, Freibert SA, Heidenreich T, Mascarenhas J, Netz DJ, Paul VD, Pierik AJ, Richter N, Stümpfig M, et al. 2015. The role of mitochondria and the CIA machinery in the maturation of cytosolic and nuclear iron–sulfur proteins. *Eur J Cell Biol*. 94(7–9):280–291.

Mi-ichi F, Yousuf MA, Nakada-Tsukui K, Nozaki T. 2009. Mitosomes in *Entamoeba histolytica* contain a sulfate activation pathway. *Proc Natl Acad Sci USA*. 106(51):21731–21736.

Netz DJA, Mascarenhas J, Stehling O, Pierik AJ, Lill R. 2014. Maturation of cytosolic and nuclear iron–sulfur proteins. *Trends Cell Biol*. 24(5):303–312.

Netz DJA, Pierik AJ, Stümpfig M, Bill E, Sharma AK, Pallesen LJ, Walden WE, Lill R. 2012. A bridging [4Fe–4S] cluster and nucleotide binding are essential for function of the Cfd1-Nbp35 complex as a scaffold in iron–sulfur protein maturation. *J Biol Chem*. 287(15):12365–12378.

Netz DJA, Stümpfig M, Doré C, Mühlenhoff U, Pierik AJ, Lill R. 2010. Tah18 transfers electrons to Dre2 in cytosolic iron–sulfur protein biogenesis. *Nat Chem Biol*. 6(10):758–765.

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.

Nyvltova E, Sutak R, Harant K, Sedinova M, Hrdy I, Paces J, Vlcek C, Tachezy J. 2013. NIF-type iron–sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of *Mastigamoeba balamuthi*. *Proc Natl Acad Sci USA*. 110(18): 7371–7376.

Pastore A, Puccio H. 2013. Frataxin: a protein in search for a function. *J Neurochem*. 126:43–52.

Paul VD, Lill R. 2015. Biogenesis of cytosolic and nuclear iron–sulfur proteins and their role in genome stability. *Biochim Biophys Acta-Mol Cell Res*. 1853(6):1528–1539.

Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*. 9(1):171–181.

Pondarré C, Antiochos BB, Campagna DR, Clarke SL, Greer EL, Deck KM, McDonald A, Han AP, Medlock A, Kutok JL, et al. 2006. The mitochondrial ATP-binding cassette transporter Abcb7 is essential in mice and participates in cytosolic iron–sulfur cluster biogenesis. *Hum Mol Genet*. 15(6):953–964.

Pyrih J, Pyrihová E, Kolísko M, Stojanovová D, Basu S, Harant K, Haindrich AC, Doležal P, Lukeš J, Roger A, et al. 2016. Minimal cytosolic iron–sulfur cluster assembly machinery of *Giardia intestinalis* is partially associated with mitosomes. *Mol Microbiol*. 102(4):701–714.

Riboldi GP, Larson TJ, Frazzon J. 2011. Enterococcus faecalis sufCDSUB complements *Escherichia coli* sufABCDSE. *FEMS Microbiol Lett*. 320(1):15–24.

Roche B, Aussel L, Ezraty B, Mandin P, Py B, Barras F. 2013. Reprint of: iron/sulfur proteins biogenesis in prokaryotes: formation, regulation and diversity. *Biochim Biophys Acta-Bioenergy*. 1827(8–9):923–937.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 61(3):539–542.

Rudolf J, Makrantoni V, Ingledew WJ, Stark MJR, White MF. 2006. The DNA repair helicases XPD and FancJ have essential iron–sulfur domains. *Mol Cell*. 23(6):801–808.

Sharma AK, Pallesen LJ, Spang RJ, Walden WE. 2010. Cytosolic iron–sulfur cluster assembly (CIA) system: factors, mechanism, and relevance to cellular iron regulation. *J Biol Chem*. 285(35):26745–26751.

Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, Soanes DM, Van Der Giezen M, Roger AJ. 2014. A SUF Fe–S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist Pygsuia. *Curr Biol*. 24(11):1176–1186.

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19(Suppl. 2):ii215–ii225.

Sutak R, Dolezal P, Fiumera HL, Hrdy I, Dancis A, Delgadillo-Correa M, Johnson PJ, Muller M, Tachezy J. 2004. Mitochondrial-type assembly of FeS centers in the hydrogenosomes of the amitochondriate eukaryote *Trichomonas vaginalis*. *Proc Natl Acad Sci USA*. 101(28):10368–10373.

Tachezy J, Sánchez LB, Müller M. 2001. Mitochondrial type iron–sulfur cluster assembly in the amitochondriate eukaryotes *Trichomonas vaginalis* and *Giardia intestinalis*, as indicated by the phylogeny of IscS. *Mol Biol Evol*. 18(10):1919–1928.

Tovar J, León-Avila G, Sánchez LB, Sutak R, Tachezy J, Van Der Giezen M, Hernández M, Müller M, Lucocq JM. 2003. Mitochondrial remnant organelles of Giardia function in iron–sulphur protein maturation. *Nature* 426(6963):172–176.

Treitli SC, Kotyk M, Yubuki N, Jirounková E, Vlasáková J, Smejkalová P, Šípek P, Čepička I, Hampl V. 2018. Molecular and morphological diversity of the Oxymonad Genera Monocercomonoides and Blattamonas gen. nov. *Protist* 169(5):744–783.

Tsaousis AD, Gentekaki E, Eme L, Gaston D, Roger AJ. 2014. Evolution of the cytosolic iron–sulfur cluster assembly machinery in Blastocystis species and other microbial eukaryotes. *Eukaryot Cell*. 13(1):143–153.

Williams BAP, Hirt RP, Lucocq JM, Embley TM. 2002. A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature* 418(6900):865–869.

Yoon H, Knight SAB, Pandey A, Pain J, Turkarslan S, Pain D, Dancis A. 2015. Turning *Saccharomyces cerevisiae* into a frataxin-independent organism. *PLoS Genet*. 11(5):e1005135.

Zhang Q, Táborský P, Silberman JD, Pánek T, Čepička I, Simpson AGB. 2015. Marine isolates of *Trimastix marina* form a plesiomorphic deep-branching lineage within preaxostyla, separate from other known trimastigids (*Paratrimastix* n. gen.). *Protist* 166(4):468–491.

Zhang Y, Lyver ER, Nakamaru-Ogiso E, Yoon H, Amutha B, Lee D-W, Bi E, Ohnishi T, Daldal F, Pain D, et al. 2008. Dre2, a conserved eukaryotic Fe/S cluster protein, functions in cytosolic Fe/S protein biogenesis. *Mol Cell Biol*. 28(18):5569–5582.

Zubáčová Z, Novák L, Bublíková J, Vacek V, Fousek J, Rídl J, Tachezy J, Doležal P, Vlček Č, Hampl V. 2013. The mitochondrion-like organelle of *Trimastix pyriformis* contains the complete glycine cleavage system. *PLoS One* 8(3):e55417.

The Oxymonad Genome Displays Canonical Eukaryotic Complexity in the Absence of a Mitochondrion

Anna Karnkowska,*[,1,2] Sebastian C. Treitli,[1] Ondřej Brzoň,[1] Lukáš Novák,[1] Vojtěch Vacek,[1] Petr Soukal,[1] Lael D. Barlow,[3] Emily K. Herman,[3] Shweta V. Pipaliya,[3] Tomáš Pánek,[4] David Žihala,[4] Romana Petrželková,[4] Anzhelika Butenko,[4] Laura Eme,[5,6] Courtney W. Stairs,[5,6] Andrew J. Roger,[5] Marek Eliáš,[4,7] Joel B. Dacks,[3] and Vladimír Hampl*[,1]

[1]Department of Parasitology, BIOCEV, Faculty of Science, Charles University, Vestec, Czech Republic

[2]Department of Molecular Phylogenetics and Evolution, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, Warsaw, Poland

[3]Division of Infectious Disease, Department of Medicine, University of Alberta, Edmonton, Canada

[4]Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

[5]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Canada

[6]Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

[7]Institute of Environmental Technologies, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

***Corresponding authors:** E-mails: ankarn@biol.uw.edu.pl; vlada@natur.cuni.cz.

Associate editor: Fabia Ursula Battistuzzi

Abstract

The discovery that the protist *Monocercomonoides exilis* completely lacks mitochondria demonstrates that these organelles are not absolutely essential to eukaryotic cells. However, the degree to which the metabolism and cellular systems of this organism have adapted to the loss of mitochondria is unknown. Here, we report an extensive analysis of the *M. exilis* genome to address this question. Unexpectedly, we find that *M. exilis* genome structure and content is similar in complexity to other eukaryotes and less "reduced" than genomes of some other protists from the Metamonada group to which it belongs. Furthermore, the predicted cytoskeletal systems, the organization of endomembrane systems, and biosynthetic pathways also display canonical eukaryotic complexity. The only apparent preadaptation that permitted the loss of mitochondria was the acquisition of the SUF system for Fe–S cluster assembly and the loss of glycine cleavage system. Changes in other systems, including in amino acid metabolism and oxidative stress response, were coincident with the loss of mitochondria but are likely adaptations to the microaerophilic and endobiotic niche rather than the mitochondrial loss per se. Apart from the lack of mitochondria and peroxisomes, we show that *M. exilis* is a fully elaborated eukaryotic cell that is a promising model system in which eukaryotic cell biology can be investigated in the absence of mitochondria.

Key words: **amitochondrial eukaryote, cell biology, Monocercomonoides, oxymonads, protist genomics.**

Article

Introduction

Mitochondria are core features of the eukaryotic cell. In addition to their signature role in ATP generation, they are integrated in diverse cellular processes including the biosynthesis and catabolism of amino acids, lipids, and carbohydrates, environmental stress tolerance, and the regulation of cell death. Despite the many independent transformations of the mitochondria into metabolically reduced and modified versions present in anaerobic organisms (Roger et al. 2017), mitochondria or mitochondrion-related organelles (MROs) were considered indispensable due to their essential core function(s) such as the biosynthesis of Fe–S clusters (Williams et al. 2002; Gray 2012; Lill et al. 2012).

However, the discovery of the first truly amitochondriate eukaryote, *Monocercomonoides* sp. PA203 (Karnkowska et al. 2016) showed that the outright loss of mitochondria is possible. This organism, now classified as *Monocercomonoides exilis* (Treitli et al. 2018), remains the only deeply inspected amitochondriate eukaryote, although the same status may hold true for its relatives, based on the limited cytological data from other oxymonads (Hampl 2017). Importantly, the ancestor of this lineage must have possessed a mitochondrial organelle, given the well-documented presence of MROs in relatives of oxymonads (Zubáčová et al. 2013; Leger et al. 2017). By studying *M. exilis*, we can determine how mitochondrial loss affects the rest of the cell and affords a unique opportunity to examine cellular systems that are normally integrated with mitochondria in a context where the organelle is absent.

Monocercomonoides exilis is a bacterivorous tetraflagellate living as a putative commensal in the intestine of caviomorph rodents (fig. 1*a* and *b*) (Treitli et al. 2018). Like all oxymonads,

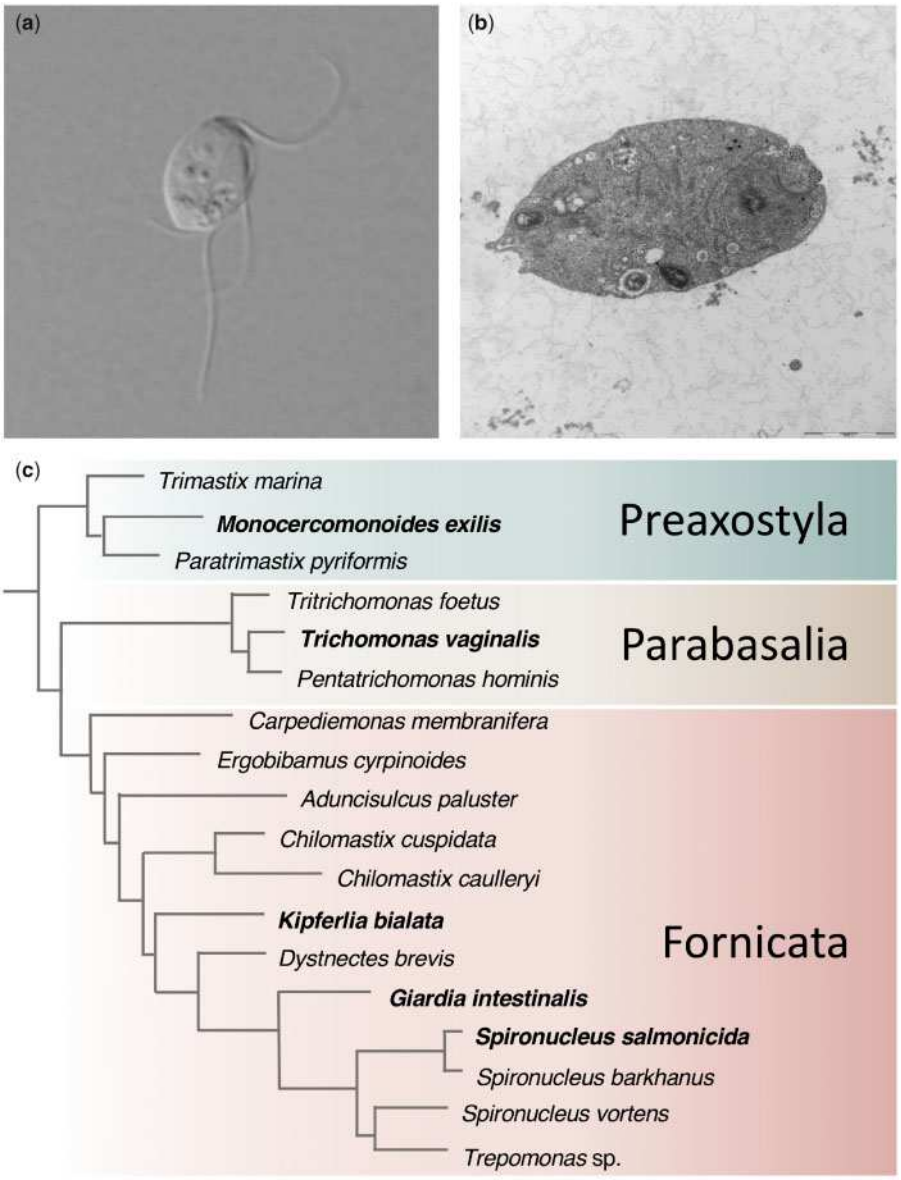


Fig. 1. The overall morphology of *Monocercomonoides exilis* and phylogeny of Metamonada. (*a*) A living cell of *M. exilis* PA203 under differential interference contrast (DIC). (*b*) TEM micrograph of *M. exilis* PA203 (credit Naoji Yubuki). (*c*) Relationships within Metamonada inferred from a phylogenomic data set (Leger et al. 2017); organisms with sequenced genomes are in bold.

M. exilis has a single long microtubular rodlike axostyle that originates from the nuclear region and is connected to the basal bodies by a characteristic fiber (i.e., the preaxostyle) consisting of a sheet of microtubules and a nonmicrotubular layer. Electron microscopic imaging of *M. exilis* showed that it lacks any conspicuous Golgi apparatus and mitochondria (Treitli et al. 2018).

Monocercomonoides exilis is a representative of a broader group of endobiotic protists called the oxymonads, which together with the free-living trimastigids, constitute the clade Preaxostyla, one of the three principal lineages of Metamonada (Leger et al. 2017; Adl et al. 2019) (fig. 1*c*). Metamonada comprise solely anaerobic/microaerophilic unicellular organisms with a diverse array of MRO types (Leger et al. 2017). Many metamonads are also parasites of agricultural or medical importance, three of which have become subject of in-depth genomic investigations: the parabasalid *Trichomonas vaginalis* (Carlton et al. 2007), the diplomonads *Giardia intestinalis* (Morrison et al. 2007), and *Spironucleus salmonicida* (Xu et al. 2014). A draft genome sequence has been reported for a free-living representative of the Fornicata, *Kipferlia bialata* (Tanifuji et al. 2018). Our *M. exilis* genome project has complemented this sampling by targeting a nonparasitic endobiont and the first representative of Preaxostyla. However, the initial genomic analysis of *M. exilis* was tightly focused on demonstrating mitochondrial absence (Karnkowska et al. 2016). Here, we present an in-depth analysis of the *M. exilis* draft genome sequence that addresses the genomic and cellular impact of mitochondrial loss in the context of metamonad evolution.

Results and Discussion

Focused Ion Beam Scanning Electron Microscope (FIB-SEM) Tomography of the Cell

To supplement the genomic analyses, corroborate the absence of mitochondria and Golgi stacks, and address several other predictions from the genomic information, we probed the *M. exilis* cell architecture by FIB-SEM tomography. We sectioned major parts of two cells fixed by two different protocols (supplementary videos S1 and S2, Supplementary Material online). The data obtained are consistent with previous transmission electron microscopy (TEM) investigation (Treitli et al. 2018). Importantly, although we acknowledge that the resolution of the microscopy could still allow for undetected highly reduced mitosomes, we did not observe any conspicuous mitochondria or MROs in this systematic examination of the *M. exilis* cells.

Genome and Predicted Proteome Features

The draft genome of *M. exilis* (Karnkowska et al. 2016, BioProject: PRJNA304271) is assembled into 2,095 scaffolds with an estimated genome size of ~75 Mb and an average GC content 36.8% (with coding regions and intergenic regions represented by 41.3% and 29% GC, respectively; table 1). By mapping sequencing reads onto the consensus genome assembly, we observed 5,150 of potential single nucleotide polymorphisms (SNPs), with the average SNP density of ~0.04 per kb. The vast majority of alternative bases have a frequency <20% (supplementary fig. S1, Supplementary Material online) suggesting that *M. exilis* cells are monoploid and most polymorphisms represent sequencing errors. This is consistent with previous fluorescence in situ hybridization (FISH) results that revealed a single signal for SUF genes in most nuclei (Karnkowska et al. 2016).

Extensive sequence diversity in terminal telomeric repeats has been found across eukaryotes where the most common, and likely ancestral, repeat type is TTAGGG (Fulnecková et al. 2013). In *M. exilis*, we identified this TTAGGG repeat element in 13 telomeric regions with at least 5 telomeric repeats at the beginning/end of the scaffold. FISH analyses with probes against telomeric repeats support the sequencing results demonstrating an average of 13 telomeric puncta in *M. exilis* nuclei (supplementary fig. S2, Supplementary Material online), suggesting the presence of 6 or 7 chromosomes. The length of telomeric regions estimated by the terminal restriction fragment method (Kimura et al. 2010) varied from 300 bases to 9 kb with a mean telomeric length of ~2.1 kb (supplementary fig. S2, Supplementary Material online).

Manual curation of the previously reported *M. exilis* genome annotation (Karnkowska et al. 2016) led to many changes including corrections of gene models and addition of new models for genes missed by the automated method originally employed. In total, 831 genes were manually curated in this study, which, together with previously curated genes (Karnkowska et al. 2016), yields a total of 1,172 manually curated genes in the current annotation release (supplementary table S1, Supplementary Material online). Three scaffolds—01876, 01882, and 01991—were recognized as probable contaminants and removed from the new version of assembly. The revised number of protein-coding genes in the *M. exilis* genome is 16,768. Homology-based approaches assigned putative functions to 6,476 (39%) *M. exilis* protein-coding genes, including 2,753 genes with domain annotations. This percentage is comparable with other metamonads, ranging from 15% of functionally annotated genes for *T. vaginalis* G3 (TrichDB Release 35) to 45% for *G. intestinalis* assemblage BGS (GiardiaDB Release 35). The annotated genome assembly and predicted genes for *M. exilis* will be available in the next release of GiardiaDB (https://giardiadb.org; last accessed 30 June, 2019).

Table 1. Summary of the *Monocercomonoides exilis* Genome Sequence Data.

Feature	Value
Genome	
Size of assembly (bp)	74,712,536
G + C content (%)	36.8
No. of scaffolds	2,092
N50 scaffold size (bp)	71,440
Protein-coding genes	
No. of predicted genes	16,768
No. of genes with introns	11,124
Mean gene length (bp)	2,703.8
Gene G + C content (%)	41.3
Mean length of intergenic regions (bp)	870.5
Intergenic G + C content (%)	29
Introns	
No. of predicted introns	31,693
Average no. of introns per gene	1.9
Intron G + C content (%)	25.2
Mean intron length (bp)	124.3
Noncoding RNA genes	
No. of predicted tRNA genes	153
No. of predicted 18S-5.8S-28S rDNA units	~50

The predicted proteins encoded by *M. exilis*, other metamonads and the heterolobosean *Naegleria gruberi* were clustered to define putative groups of orthologs (orthogroups) (fig. 2). Of the 2,031 orthogroups represented in *M. exilis*, the highest number (1,688, i.e., 83%) is shared with *N. gruberi*, which was previously suggested to be overrepresented, relative to other eukaryotes, in ancestral eukaryotic proteins (i.e., proteins that were present in the last eukaryotic common ancestor [LECA]) (Fritz-Laylin et al. 2010). The degree of orthogroup overlap was lower with *T. vaginalis* (1,564, i.e., 77%) and even more limited with diplomonads (1,057 for *G. intestinalis* and 1,065 for *S. salmonicida*, i.e., 52%). This pattern suggests that *M. exilis* has lost fewer ancestral eukaryotic proteins than other metamonads. Therefore, despite the absence of mitochondria, the proteome of *M. exilis* is likely more representative of the proteome of ancestral metamonads than that of either diplomonads or parabasalids.

The largest gene family in the *M. exilis* genome encodes protein tyrosine kinases (supplementary table S2, Supplementary Material online). The vast majority (320 out of 332 predicted tyrosine kinases) belong to the diverse tyrosine kinase-like group (supplementary table S3, Supplementary Material online). Although this group is also

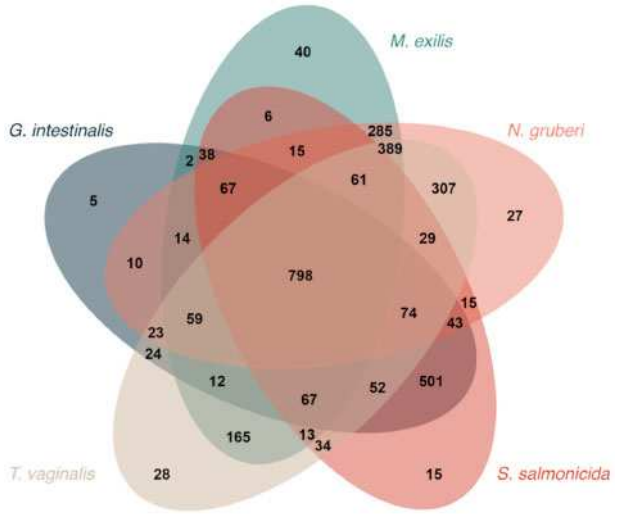


FIG. 2. Venn diagram of orthologous clusters shared and unique to *Monocercomonoides exilis*, other metamonads, and *Naegleria gruberi*.

expanded in *T. vaginalis* (Carlton et al. 2007), neither classical tyrosine kinases (TK group) nor members of the related tyrosine kinase-like group were identified in *G. intestinalis* (Manning et al. 2011), and only one occurs in *S. salmonicida* (Xu et al. 2014). Other abundant families in *M. exilis* include the Ras superfamily GTPases, cysteine proteases, and thioredoxins (see below) (supplementary table S2, Supplementary Material online).

Intron Gain and Loss in the *M. exilis* Lineage

Parasitic metamonad genomes sequenced so far are characterized by a scarcity of introns. Only 6 *cis*-spliced introns and 5 unusual split *trans*-spliced introns were found in *G. intestinalis* (Kamikawa et al. 2011; Franzén et al. 2013), 4 *cis*-spliced introns in *S. salmonicida* (Xu et al. 2014; Roy 2017), and 65 in *T. vaginalis* (Carlton et al. 2007). In contrast, we previously reported over 32,000 introns in the genome of *M. exilis* (Karnkowska et al. 2016). Sequencing of the free-living metamonad *K. bialata* revealed >120,000 introns, the highest number noted in metamonads so far (Tanifuji et al. 2018). We were unable to compare the *M. exilis* and *K. bialata* genomes because we completed our analyses prior to the release of the *Kipferlia* data.

With additional manual curation of gene models, the current estimate of the number of spliceosomal introns in *M. exilis* genome is 31,693, with an average number of 1.9 and 0.8 introns per gene and per kb of coding sequence, respectively. The high intron density is consistent with the previous report of introns in the oxymonad *Streblomastix strix* (Slamovits and Keeling 2006) and comparable with other free-living protists (e.g., *Dictyostelium discoideum*) (Eichinger et al. 2005) and well within the range exhibited by conventional eukaryotic genomes (Rogozin et al. 2012; Irimia and Roy 2014). The ubiquity of canonical GT-AG and lack of AT-AC boundaries indicates that only the major (U2-dependent) introns are present, which is consistent with the absence of minor (U12-dependent) spliceosome components (supplementary table S1, Supplementary Material online). The large number of introns in this organism increases the energetic cost of gene expression (Lynch and Marinov 2015). Clearly, sufficient ATP is produced in *M. exilis* by substrate-level phosphorylation to meet these costs in the absence of aerobically respiring mitochondria (Hampl et al. 2019).

To understand the origin and evolution of introns in *M. exilis*, we performed an analysis of the relative contribution of retention of ancestral introns and insertion of lineage-specific introns in the *M. exilis* genome. We analyzed introns in a set of 100 conserved eukaryotic genes with well-established orthologs in 34 reference species across representative lineages of eukaryotes (fig. 3). In total, these genes and species comprised 3,546 intron positions, 201 of them represented in *M. exilis* homologs. We used Dollo parsimony to reconstruct intron gains and losses along the eukaryote phylogeny using three alternative root positions. With one of the root positions, the LECA is inferred to have harbored 432 introns in this gene set, 65 of which have been retained in *M. exilis* (fig. 3*a*). This accounts for more than 30% of the *M. exilis* introns analyzed, a proportion similar to other eukaryotes (fig. 3*b*). The other two root positions give similar estimates (supplementary fig. S3*a* and *b*, Supplementary Material online). The absence of the remaining ancestral introns from the *M. exilis* genome could be explained by massive intron loss along the stem lineage of Metamonada, that is, before the split of the lineages leading to *M. exilis* on one side and the diplomonads plus *T. vaginalis* clade on the other (fig. 3*a*). However, it is to be noted that the recently sequenced genome of the diplomonad relative *K. bialata* is reported to include >120,000 introns (Tanifuji et al. 2018), so it is possible that many of these losses are in fact specific for Preaxostyla or oxymonads. Likewise, although our analysis suggests substantial acquisition of new introns in the *M. exilis* lineage (fig. 3*a* and supplementary fig. S3*a* and *b*, Supplementary Material online), the new data from *K. bialata* make it likely that many of these gains are more ancient (having occurred already in the metamonad stem lineage).

Genome Maintenance and Expression in *M. exilis*

Given the extraordinary absence of mitochondrial organelles from *M. exilis*, we examined genes encoding components of other cellular systems to assess whether they were similarly reduced or unusual.

We first investigated the systems responsible for maintenance and expression of the *M. exilis* nuclear genome. We identified all expected universally conserved genes encoding nucleus-functioning proteins (Iyer et al. 2008). For example, *M. exilis* encodes all four core histones (H2A, seven variants; H2B, three variants; H3; two variants; and H4, one variant) as well as the linker histone H1 (supplementary table S1, Supplementary Material online).

All essential components involved in DNA unwinding, primer synthesis, and DNA replication were also present in the *M. exilis* genome (supplementary table S1, Supplementary Material online). The origin recognition complex of *G. intestinalis* and *T. vaginalis* each contain ORC1 and ORC4, whereas *S. salmonicida* relies on a CDC6 complex. In *M. exilis*, we were only able to identify ORC1. In terms of replication machinery, most metamonads do not encode replication protein A

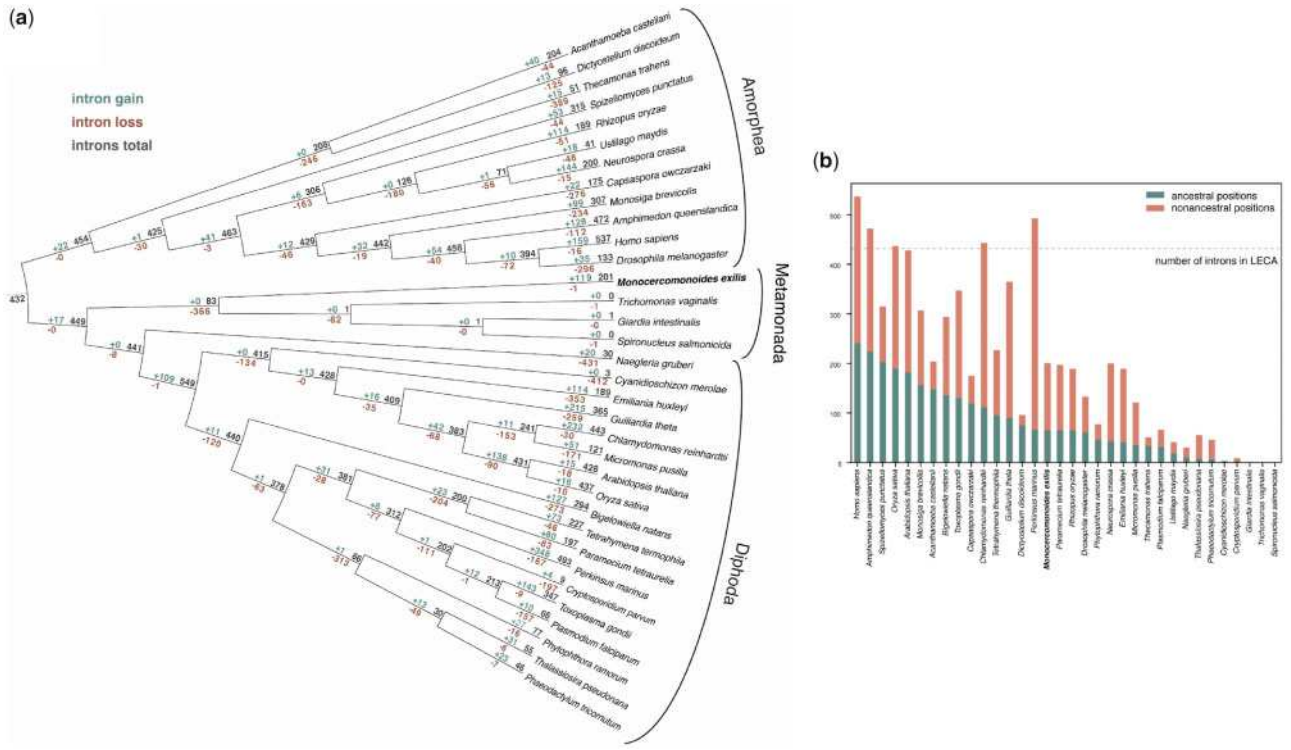


FIG. 3. Intron gains and losses along the eukaryote phylogeny. (*a*) Intron gains and losses along the eukaryote phylogeny as reconstructed by Dollo parsimony. The numbers are derived from an analysis of 3,546 intron positions in a reference set of 100 groups of orthologous genes of 34 phylogenetically diverse species. Root of the eukaryote phylogeny was considered between Amorphea and the remaining eukaryotes included in the analysis. (*b*) Numbers of ancestral (i.e., inherited from the LECA) and nonancestral (i.e., lineage-specific) introns in different eukaryotes. Derived from the analysis described in (a).

heterotrimeric complex or have a drastically reduced complex consisting of only one protein in the case of diplomonads (Morrison et al. 2007; Xu et al. 2014). Surprisingly, we identified all subunits of the replication protein A heterotrimeric complex in *M. exilis* (supplementary table S4, Supplementary Material online). In addition, similar to eukaryotes in general (Forterre et al. 2007), *M. exilis* employs two different types of topoisomerase I, Topo IB and Topo III, whereas parabasalids and fornicates have retained only the latter type (supplementary table S4, Supplementary Material online).

We similarly identified most of the components of various DNA repair pathways including base excision repair, nucleotide excision repair, and mismatch repair pathways (Costa et al. 2003; Kunkel and Erie 2005; Almeida and Sobol 2007; Fukui 2010) (supplementary tables S1 and S4, Supplementary Material online). The mismatch repair pathway appears to be complete in all other sequenced metamonads, whereas the base excision repair and nucleotide excision repair pathways are most complete in *M. exilis*, especially when compared with *G. intestinalis* or *S. salmonicida* (Marchat et al. 2011) (supplementary table S4, Supplementary Material online). The nonhomologous end joining pathway involved in repairing double-strand breaks is missing in *M. exilis*, similarly to other metamonads (Carlton et al. 2007; Morrison et al. 2007). However, the homologous recombination repair pathway for double-strand breaks repair is encoded in the *M. exilis* genome (supplementary table S4, Supplementary Material online) and looks more complete in *M. exilis* than in other metamonads.

We also investigated the complement of general transcription factors in *M. exilis* and identified subunits of all general transcription factors known to be highly conserved among eukaryotes (Orphanides et al. 1996; Latchman 1997; de Mendoza et al. 2013) (supplementary table S1, Supplementary Material online). Notably, *M. exilis* possesses both subunits of TFIIA, the primary function of which is stabilization of the preinitiation complex and assistance in the binding of TBP to the TATA box in promoters (Tang et al. 1996). The presence of TFIIA is in agreement with the presence of TATA-like motifs in ~52% of *M. exilis* promoter regions (3,374/6,509 genes with predicted UTR) (supplementary fig. S4, Supplementary Material online). In contrast, *T. vaginalis* lacks TFIIA, and an M3 motif has replaced the TATA box in this lineage (Smith et al. 2011).

Regarding the translation machinery, we identified 30 proteins in the *M. exilis* genome annotated as eukaryotic initiation factors or their associated factors (supplementary table S1, Supplementary Material online). This set is nearly complete compared with the mammalian translation machinery and it is almost identical to the sets of eukaryotic initiation factors present in the genomes of *T. vaginalis* and *G. intestinalis* (Kanehisa et al. 2014) (supplementary table S1, Supplementary Material online). Of note is the presence in *M. exilis* of enzymes responsible for the formation of

diphthamide (supplementary table S1, Supplementary Material online), a modified histidine residue present in archaeal and eukaryotic elongation factor 2 and important for its proper function (Su et al. 2013). The retention of this modification in *M. exilis* contrasts with the situation in parabasalids recently shown to have lost diphthamide biosynthesis genes (Narrowe et al. 2018).

Actin and Tubulin Cytoskeleton

The extensive cytoskeletal apparatus, including the hallmark oxymonad axostyle, is one of the better-described fascinating aspects of the cellular architecture of *Monocercomonoides* (Radek 1994; Treitli et al. 2018). Like many other cellular systems of *Monocercomonoides*, the actin and tubulin cytoskeletons of diplomonads and parabasalids depart in various ways from the general picture seen in other eukaryotes. The metamonad actin cytoskeleton is reduced and modified (supplementary tables S1 and S5, Supplementary Material online), with the highest reduction in *G. intestinalis* (Morrison et al. 2007; Paredez et al. 2014). *Monocercomonoides exilis* shares some of the unusual modifications with other metamonads, as it, for example, lacks proteins containing the myosin head domain, a trait that has so far only been reported from some rhodophytes, *T. vaginalis*, and diplomonads (Sebé-Pedrós et al. 2014; Brawley et al. 2017). On the other hand, it stands out by possessing a more complete set of actin family proteins (including actin-related proteins; ARPs) than other metamonads. Specifically, it has retained ARP4 and ARP6, important nuclear ARPs serving in several chromatin-remodeling complexes (Oma and Harata 2011), and also a gene for the actin-binding protein villin. However, its actual repertoire of proteins associated with the actin cytoskeleton is less complex than that of *T. vaginalis* (Kollmar et al. 2012) (supplementary tables S1 and S5, Supplementary Material online).

Monocercomonoides exilis encodes a conventional set of tubulins that is similar to those found in other metamonad organisms. All metamonads analyzed contain at least one complete gene for alpha-, beta-, gamma-, delta-, and epsilon-tubulin. The sixth ancestral eukaryotic paralog, zeta-tubulin, has frequently been lost during eukaryote evolution (Findeisen et al. 2014) and is also missing from metamonads including *M. exilis* (supplementary tables S1 and S5, Supplementary Material online). All metamonads also contain multiple members of both groups of motor proteins associated with microtubules, that is, kinesins (primarily mediating plus end-directed transport) and dyneins (mediating minus end-directed transport and flagellar motility). Our comprehensive phylogenetic analysis (supplementary fig. S5, Supplementary Material online) showed that the family of kinesins is well represented in metamonads. Of the 17 previously defined kinesin families with wide taxonomic distribution (Wickstead et al. 2010), only 3 are missing from metamonads as a whole, albeit others may be only patchily distributed in the group (supplementary table S5, Supplementary Material online).

Dyneins are large multi-subunit complexes consisting of one or more dynein heavy chains (DHC) and a variable number of intermediate chains, light intermediate chains, and light chains. Interestingly, eukaryotes use just a single dynein complex (called cytoplasmic dynein 1) for nearly all cytoplasmic minus end-directed transport (Roberts et al. 2013). In accordance with the previous reports for *G. intestinalis* and *T. vaginalis* (Wickstead and Gull 2012), *M. exilis* lacks two important components of the cytoplasmic dynein 1 complex, namely the specific intermediate (DYNC1I1) and light intermediate (DYNC1LI) chains, while keeping the heavy chain (DHC1) that constitutes the center of the complex. In eukaryotes, the presence of the cytoplasmic dynein 1 is coupled to the presence of the dynactin complex, a large multisubunit protein complex that enhances the motor processivity and acts as an adapter between the motor complex and the cargo. *Trichomonas vaginalis* and *G. intestinalis* have been reported as rare examples of eukaryotes lacking the dynactin complex in the presence of cytoplasmic dynein 1 (Hammesfahr and Kollmar 2012). Here, we show that all subunits specific to the dynactin complex are also missing from *M. exilis*, suggesting the absence of the complex from metamonads in general. The set of the axonemal dyneins is nearly complete in metamonads including *M. exilis* (supplementary fig. S6, Supplementary Material online). We are the first to report the dynein intermediate chain WDR34, a specific component of the intraflagellar transport dynein in metamonads.

The conservation of microtubule-dependent chromosome separation across extant eukaryotes strongly suggests that this feature was present in LECA. Microtubules and chromatids are connected by the kinetochore, a multiprotein structure that is assembled on centromeric chromatin. Based on comparative studies, orthologs of 70 kinetochore proteins have been identified in various eukaryotes suggesting that LECA had a complex kinetochore structure (van Hooff et al. 2017). However, the metamonads that have been studied to date have closed or semiopen mitosis (Ribeiro et al. 2002; Sagolla et al. 2006) and their kinetochores are divergent and degenerated in comparison to kinetochores of model organisms such as human or yeast (van Hooff et al. 2017). Of the 70 kinetochore orthologs, we have identified 15 in *M. exilis*, a comparable number to those identified in *G. intestinalis* (16), and slightly fewer than in *T. vaginalis* (27) (supplementary table S6, Supplementary Material online). Although the kinetochore is reduced, all investigated metamonads possess the most conserved components including Skp1, Plk, Aurora, or CenA suggesting the presence of a functional kinetochore in these species.

Overall, our analyses suggest that except for dispensing with myosin-based motility and the dynactin complex—traits shared by all metamonads for which genome data is available to date—*M. exilis* has a relatively canonical complement of cytoskeletal proteins.

Standard and Unconventional Aspects of the Endomembrane System

The endomembrane system is a critical interface between an organism and its extracellular environment, and it underpins host-parasite interactions in many microbial eukaryotes. The *M. exilis* endomembrane system noticeably lacks any reported morphologically recognizable Golgi bodies (Radek 1994;

Treitli et al. 2018) prompting suggestions of organelle absence similar to mitochondria. From our FIB-SEM data (supplementary videos S1 and S2, supplementary fig. S7, Supplementary Material online), we note that the cells contain a well-developed endoplasmic reticulum (ER), which sometimes forms stacks superficially resembling Golgi. However, these ER structures can clearly be distinguished by the ribosomes attached to their surfaces. Golgi stacks were not observed. On the other hand, our previous genomic analyses identified 84 genes that serve as indicators of Golgi presence, starkly contrasting with the absence of mitochondrial hallmark proteins (Karnkowska et al. 2016). To better characterize the endomembrane system of *M. exilis*, we have expanded our genomic analysis of membrane-trafficking machinery.

Monocercomonoides exilis has a relatively canonical complement of endomembrane system proteins (fig. 4) encoding most of the basic eukaryotic set (Koumandou et al. 2007). It shows neither extensive reduction nor expansion of this set as observed in the *G. intestinalis* (Morrison et al. 2007) and *T. vaginalis* (Carlton et al. 2007) genomes, respectively. For several protein membrane-trafficking complexes, we observed multiple versions of some components, but the lack of others. The retromer complex transports internalized plasma membrane receptors from endosomes to the trans-Golgi network (Seaman 2004). Although the cargo-recognition subcomplex was identified in *M. exilis*, neither membrane-deforming sorting nexin proteins nor the conventional cargo protein, Vps10, could be found. This is unusual, but not unprecedented, in eukaryotes (Koumandou et al. 2011). We found an expanded set of components for ESCRT II, III and IIIa subcomplexes, but a lack of all but Vps23 of the ESCRT I subunits. ESCRT complexes are best known for their role in protein degradation at the multivesicular body (MVB) and functional MVBs have been identified in *Tetrahymena* which also possesses only Vps23 as its ESCRT I (Leung et al. 2008; Cole et al. 2015). The genomic data predict that MVBs should exist in *M. exilis*, and indeed candidate MVBs—that is, single-membrane bound small compartments with internal vesicles—were frequently observed in the FIB-SEM images (supplementary fig. S7). The presence of this organelle could be significant as these compartments are the source of exosomes which are implicated in host-endobiotic interactions (Schorey et al. 2015). Finally, we observed at least two sets of all components of the HOPS complex that acts at the late endosome, but we were unable to identify any of the subunits that are specific to the CORVET complex that acts upstream at the early endosomes. Intriguingly, the Vps39, a HOPS-specific component has been recently shown to function in vacuole-mitochondria contact sites (vCLAMP), with Tom40 as the direct binding partner on mitochondria (González Montoro et al. 2018). This highlights the potential to use *M. exilis* to disentangle nonmitochondrial functions of this protein without the indirect effect on mitochondria. Similarly, Vps13, has been proposed to be present in several membrane contact sites, including endosome–mitochondrion contacts (Park et al. 2016) and proven to influence mitochondrial morphology in human cells (Yeshaw et al. 2019). Four Vps13 paralogs are encoded in the *M. exilis* genome.

We observed multiple paralogs of subunits in some endosomal-associated complexes potentially indicating diversified endolysosomal pathways. Some of the paralog expansions were small, such as the adaptor protein complexes, Rab11, and endosomal Qa-SNAREs together with their interacting SM proteins. Other complexes were more extensively expanded, including four paralogs of Syn6, EpsinR, Vps34, and TBC-F, five SMAP paralogs, and eight VAMP7 R-SNAREs. It is particularly striking that, despite encoding a single copy of the Rab7-specific GEFs Mon1 and CCZ1 (supplementary fig. S8 and supplementary table S1, Supplementary Material online) like other eukaryotes, *M. exilis* has an expanded set of Rab7 paralogs, including nine "conventional" Rab7 paralogs and a clade of nine additional very divergent Rab7-like (Rab7L) paralogs not found in other eukaryotes so far (supplementary fig. S9, Supplementary Material online). Some Rab7L loci are apparently nonfunctional (with coding sequences disrupted by mutations), indicating birth-and-death evolution of this gene group. Hence, we speculate that the Rab7L clade is involved in a novel, rapidly evolving endocytic process in *M. exilis*. Consistent with the observed diversified complement of endolysosomal membrane-trafficking machinery, we also noted that the cytoplasm contains numerous vesicles with electron lucent matrix some of them containing food particles (putative phagosomes), and others resembling endosomes of various shapes and sizes (supplementary fig. S7, Supplementary Material online). The conspicuous dark round globules observed in supplementary video S1, Supplementary Material online, are very likely glycogen granules observed also under classical TEM (Treitli et al. 2018).

Expanded Set of Proteolytic Enzymes

Proteases are important virulence factors for parasites and are known to degrade the host's extracellular matrix during the invasion (Sajid and McKerrow 2002). We identified 122 protease homologs, divided into 4 catalytic classes (cysteine, metallo, serine, and threonine) and 14 families according to Merops protease classification (Rawlings et al. 2008) (supplementary table S7, Supplementary Material online). The expansion of cysteine proteases is consistent with the expanded complement of Cathepsin B cysteine proteases previously observed in *M. exilis* (Dacks et al. 2008). We confirmed that the *M. exilis* genome encodes 44 Cathepsin B paralogs but no Cathepsin L genes (supplementary fig. S10). The high number of cysteine proteases is surprising because *M. exilis* is considered a commensal rather than a parasite. The large number of cysteine proteases previously reported for parasitic metamonads are often thought to be involved in tissue destruction or host defense (Carlton et al. 2007; Xu et al. 2014).

Salvage of Nucleotides from the Gut Environment

Gut symbionts often lose the ability to biosynthesize cellular building blocks like nucleotides, and *M. exilis* appears to be no exception. *Monocercomonoides exilis* lacks enzymes for de novo synthesis and catabolism of purines or pyrimidines

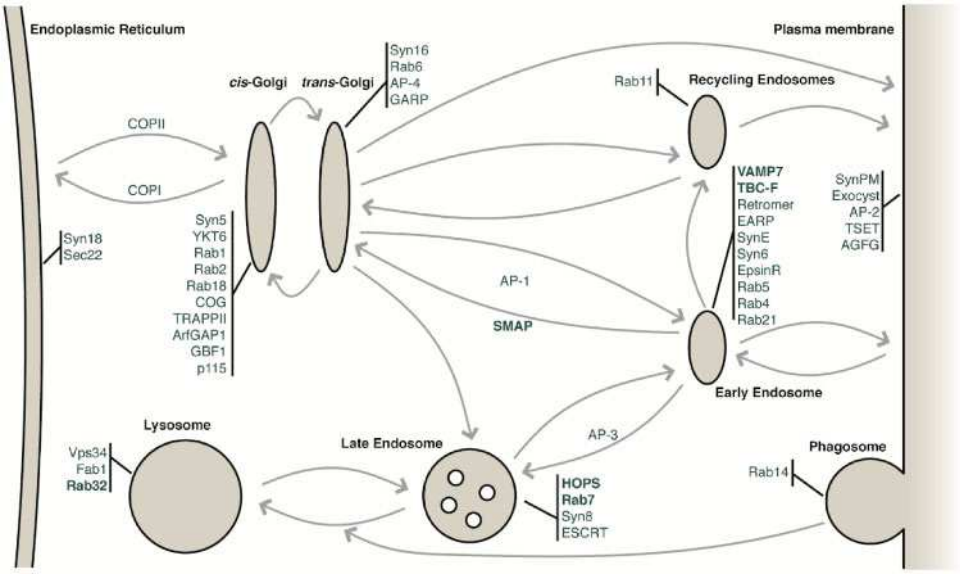


Fig. 4. Results of bioinformatic analysis of membrane-trafficking machinery in *Monocercomonoides exilis*. The presence of membrane-trafficking compartments and pathways is hypothesized as shown, based on the complement of trafficking machinery identified, and the function of their homologs in model systems. Selected membrane-trafficking proteins and protein complexes identified in the genome of *Monocercomonoides exilis* are shown. Several genes for membrane-trafficking proteins appear to have undergone lineage-specific duplications, and these are shown in bold font.

(supplementary table S1 and supplementary fig. S11, Supplementary Material online). This implies that the cell depends on external sources of these compounds that are incorporated into the nucleotide pool by salvage pathways in a manner similar to *T. vaginalis*, *G. intestinalis*, and several trypanosomatids (de Koning et al. 2005) (supplementary table S8, Supplementary Material online). However, there are some notable differences to other metamonads in the set of enzymes that may be used for salvaging of nucleotides (Aldritt et al. 1985; Munagala and Wang 2002; Munagala and Wang 2003). Probably the most crucial difference is that *M. exilis*, unlike *T. vaginalis* and *G. intestinalis*, does not rely on the salvage of deoxyribonucleotides (Wang and Cheng 1984; Baum et al. 1989), as it can convert ribonucleotides to deoxyribonucleotides by the action of ribonucleoside-triphosphate reductase.

Is the Absence of Mitochondrion Reflected in Modifications of Any Cellular System?

Mitochondria are tightly integrated into various systems/pathways in typical eukaryotic cells (Roger et al. 2017). It is therefore of interest to investigate how these systems are affected by the loss of mitochondria in *M. exilis*. Changes related to mitochondrial loss can be divided into three categories: 1) preadaptations, which subsequently made mitochondria dispensable, 2) functions lost concomitantly with mitochondria, and 3) postadaptations that evolved to compensate for the absence of mitochondria. Only specific changes in *M. exilis* not present in other Metamonada should be considered, but one should always keep in mind that even *M. exilis*-specific features may reflect adaptation to anaerobiosis or endobiotic lifestyle with no direct link to mitochondrial loss.

Like many anaerobic protists, *M. exilis* cannot synthesize ATP by oxidative phosphorylation; instead, the ATP is synthesized via glycolysis in the cytosol (Karnkowska et al. 2016). Coupled with the loss of oxidative phosphorylation, *M. exilis* does not encode genes for any of the tricarboxylic acid cycle enzymes (Karnkowska et al. 2016).

In some anaerobic organisms, glycolysis-derived pyruvate is oxidized to acetyl-CoA by pyruvate:ferredoxin oxidoreductase with the concomitant reduction of ferredoxin, which, in turn, serves as an electron donor for hydrogen evolution via an [FeFe]-hydrogenase (HYD). In *G. intestinalis* and *Entamoeba histolytica*, the resulting acetyl-CoA can be fermented to ethanol, catalyzed by the bifunctional aldehyde/alcohol dehydrogenase E (ADHE), or is converted to acetate by acetyl-CoA synthetase (ADP-forming) (Ginger et al. 2010). The latter reaction produces one molecule of ATP. We identified homologs of PFO, HYD, ADHE, and acetyl-CoA synthetase in the *M. exilis* genome (fig. 5 and supplementary table S1, Supplementary Material online). Acetate may be further fermented to aldehyde and ethanol by aldehyde dehydrogenase and ADHE suggesting that ethanol may be the final fermentation product in *M. exilis*.

As we reported previously, *M. exilis* possesses a complete arginine deiminase pathway that enables it to produce ATP by conversion of arginine to ornithine, NH_3, and CO_2 (Novák et al. 2016). Further analyses of its genome suggest that *M. exilis* can generate ATP by metabolizing other amino acids, including tryptophan, cysteine, serine, threonine, and methionine (fig. 6*a* and *b*, Supplementary Material online), as was previously reported in other protists (Anderson and Loftus 2005). One notable aspect of the amino acid catabolism in *M. exilis* is the presence of tryptophanase, an enzyme which occurs rarely in eukaryotes and has been found so far only

Fig. 5. Carbon and energy metabolism in *Monocercomonoides exilis*. Glucose metabolism (brown), pyruvate metabolism (red), and pentose-phosphate metabolism (green). Abbreviations and Enzyme Commission numbers are given in supplementary table S1, Supplementary Material online.

in anaerobic protists *T. vaginalis*, *Tritrichomonas foetus*, *Mastigamoeba balamuthi*, *Blastocystis* spp., *Pygsuia biforma*, and *E. histolytica* (Eme et al. 2017). Among the products of tryptophan degradation by tryptophanase is indole, a signaling molecule important, for example, for interactions between mammalian host and enteric bacteria, and indeed, *Ma. balamuthi* was shown to produce significant amounts of indole (Nývltová et al. 2017). The pentose-phosphate pathway (PPP) is integrated with the main metabolic energy generating pathways. PPP is involved in the generation of NADPH and pentose sugars and has an oxidative and a nonoxidative phase. We were unable to find homologs of the enzymes for

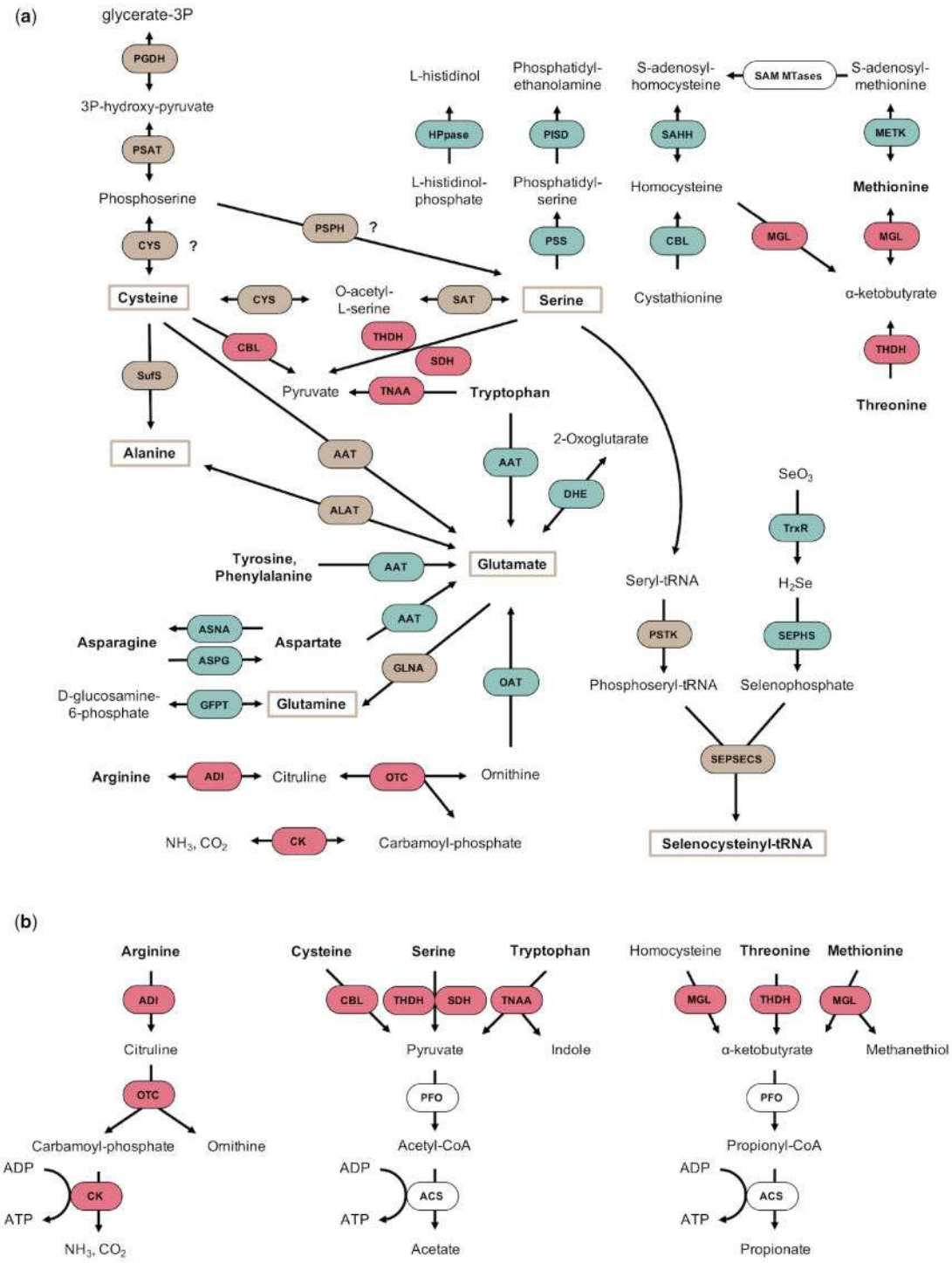


Fig. 6. Putative amino acid related biochemical pathways in *Monocercomonoides exilis*. (*a*) Amino acid metabolism. (*b*) Reactions putatively involved in ATP production by amino acids catabolism. Abbreviations and Enzyme Commission numbers are given in supplementary table S1, Supplementary Material online. Brown color indicates enzymes and products of putative amino acid biosynthesis pathways. Red color indicates enzymes putatively involved in ATP production by amino acids catabolism. Question marks indicate alternative pathways for cysteine and serine biosynthesis. "SAM MTases" stands for various S-adenosyl-methionine-dependent methyltransferases.

the oxidative phase but identified those for the nonoxidative one (fig. 5 and supplementary table S1, Supplementary Material online). We propose that the oxidative phase is likely absent in *M. exilis*, which is not unusual, as a truncated PPP has also been observed in *E. histolytica* (Loftus et al. 2005). The dehydrogenase reactions of the oxidative PPP are considered as one of the primary cellular sources of NADPH; therefore, NADPH must be synthesized via an alternative route in *M. exilis*. One such NADPH-generating enzyme might be the putative NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPN) identified in the *M. exilis* genome (supplementary table S1, Supplementary Material online). It was

proposed that GAPN plays an important role in NADPH generation in bacteria and archaea (Spaans et al. 2015). Interestingly, the *M. exilis* GAPN was identified as a lateral gene transfer (LGT) candidate (supplementary table S9, Supplementary Material online). In summary, the NAD(P)H and ATP generation pathways in *M. exilis* resemble those reported for mitosome-bearing anaerobes, and so there is no indication that they were affected by the complete loss of mitochondria.

An essential function of mitochondria and MROs in eukaryotes is the synthesis of Fe–S proteins. The ISC pathway, otherwise considered to be an essential house-keeping pathway in eukaryotes, was functionally replaced in all examined Preaxostyla by the SUF pathway (Vacek et al. 2018). We have argued previously that this replacement was a preadaptation for the subsequent loss of mitochondria in this lineage (Karnkowska et al. 2016). This unprecedented event apparently did not influence the number of Fe–S proteins in the cell, as we identified 70 candidates for such proteins in *M. exilis*, with no essential Fe–S protein missing (supplementary table S10, Supplementary Material online). With the exception of xanthine dehydrogenase containing a 2Fe–2S cluster, all other proteins, for which the type of cluster can be estimated, appear to have 4Fe–4S clusters. Similarly, 4Fe–4S clusters are more abundant than 2Fe–2S clusters in other anaerobes (Andreini et al. 2017). The presence of the full array of essential Fe–S proteins in *M. exilis* suggests that its cytosolic SUF and CIA systems are fully capable of satisfying cellular Fe–S cluster needs.

In most eukaryotes, fatty acid metabolism is integrated between the mitochondria, ER and peroxisomes. *Monocercomonoides exilis* possesses all the proteins necessary for the synthesis of diacylglycerol and for the interconversion of phosphatidylcholine, phosphotidylethanolamine, and phosphatidylserine from phosphatidate. It also possesses a suite of putatively ER-localized fatty acid biosynthesis proteins for very long fatty acid elongation by using malonyl-CoA (supplementary table S1, Supplementary Material online). However, we were unable to identify components for shorter chain fatty acid biosynthesis or fatty acid degradation pathways. Reduction of the fatty acid synthesis complex is also known from other microaerophilic protists such as *G. intestinalis* (Morrison et al. 2007) and *E. histolytica* (Loftus et al. 2005), both mitosome-possessing gut parasites. Given the lack of some lipid biosynthetic and degradation pathways, and the lack of any MRO in *M. exilis*, we searched for evidence of peroxisomes, an organelle that has long been predicted as also absent due based on microscopic evidence. Loss of peroxisomes (and peroxins) has been confirmed in several groups across the tree of eukaryotes and is often associated with the reduction of mitochondria (Žárský and Tachezy 2015; Gabaldón et al. 2016). We searched the *M. exilis* genome for peroxin homologs, but only Pex19, a cytosolic receptor for proteins targeted to the peroxisomal membrane, was found. This suggests that *M. exilis* lacks peroxisomes and that retention of Pex19 reflects a peroxisome-independent function of the protein, possibly associated with the ER (Yamamoto and Sakisaka 2018). This result is also consistent with our failure to identify any of the ER-localized Dsl1 complex subunits, as losses of peroxisomes and Dsl1 subunits are correlated (Klinger et al. 2013). As the peroxisomes are lost in many anaerobes, their absence cannot be attributed to the loss of mitochondria.

A second key role of both mitochondria and peroxisomes is the oxidative stress response. Lack of oxygen-dependent mitochondria and their reduction to MROs reduce the impact of the organelle on the production of reactive species in anaerobic and microaerophilic protists. However, many of them are transiently exposed to oxygen and have evolved a variety of strategies to cope with oxygen stress. Intracellular proteins and low-molecular-weight thiols are the main cellular antioxidants present in anaerobic protists (Müller et al. 2003). In the *M. exilis* genome, we identified superoxide dismutase responsible for the radical anion ($O^{2-\cdot}$) detoxification to O_2 and H_2O_2. We also found candidates for catalase and peroxiredoxins, which are involved in reduction of H_2O_2 to O_2 and H_2O, and hybrid cluster protein and rubrerythrin, which decompose H_2O_2 to H_2O (supplementary fig. S12 and supplementary table S1, Supplementary Material online). Peroxiredoxins must be recharged by reduction in reaction with thioredoxin, which also have been identified in the *M. exilis* genome (supplementary fig. S12 and supplementary table S1, Supplementary Material online). In other metamonads such as *G. intestinalis* or *S. salmonicida*, the main nonprotein thiol is cysteine (Brown et al. 1993; Stairs et al. 2019); the putative ability of *M. exilis* to synthesize cysteine suggests that cysteine might be also the main nonprotein thiol in this organism.

Our analysis of the *M. exilis* genome revealed an expanded repertoire of genes involved in oxygen stress response, mainly acquired by LGT from bacteria (supplementary table S9, Supplementary Material online). *Monocercomonoides exilis* genome encodes homologs of not only ruberythrin, nitroreductase, and flavodiiron protein but also rare among eukaryotic microaerophiles, of catalase, and hemerythrin, an enzyme involved in the protection of Fe–S cluster-containing proteins from oxidative damage in microaerophilic bacteria (Kendall et al. 2014) (supplementary table S9, Supplementary Material online). This enlarged set of proteins involved in oxygen stress response might be related to the complete loss of mitochondria. However, as many microaerophilic/anaerobic protists are also known to possess an expanded set of oxygen stress response proteins, this feature of *M. exilis* may instead just be reflective of its ecological niche.

Amino acid biosynthesis is another canonical mitochondrial function. *Monocercomonoides exilis* seems to be able to synthesize at least alanine, serine, cysteine, and selenocysteine, and, assuming availability of 2-oxoglutarate, also glutamate and glutamine (relevant biosynthetic pathways are highlighted in brown in fig. 6*a*). The crucial first step seems to be the synthesis of serine from a glycolysis intermediate 3-phosphoglycerate by a pathway consisting of three reactions. A gene encoding the enzyme catalyzing the third reaction, phosphoserine phosphatase, was not conclusively identified in *M. exilis* genome, but a possible candidate is the protein MONOS_5832 which is similar to the phosphoserine

phosphatase recently characterized in *E. histolytica* (Kumari et al. 2019). Alternatively, it has been shown that the conversion of phosphoserine to cysteine can be catalyzed by cysteine synthase in *T. vaginalis* (Westrop et al. 2006); this might also be the case for *M. exilis*. The reconstructed amino acid metabolic network (fig. 6*a*, supplementary table S1, Supplementary Material online) of *M. exilis* is more complex than those reported for *G. intestinalis* (Morrison et al. 2007), *E. histolytica* (Loftus et al. 2005), and *Cryptosporidium parvum* (Abrahamsen et al. 2004), but less complex than the amino acid metabolism of *T. vaginalis* (Carlton et al. 2007). *Monocercomonoides exilis* also lacks the glycine cleavage system (GCS) and serine hydroxymethyltransferase (SHMT), which are both present in its close relative *Paratrimastix pyriformis*, where they localize into the MRO (Hampl et al. 2008; Zubáčová et al. 2013). Related to amino acid metabolism and translation is the finding that the *M. exilis* genome encodes components of selenium utilization machinery, including enzymes responsible for the synthesis of selenocysteinyl-tRNA and the translation factor SelB required for selenocysteine incorporation into proteins during translation (supplementary table S1, Supplementary Material online). *Trichomonas vaginalis* and *G. intestinalis* do not utilize selenium, but certain *Spironucleus* species possess selenoproteins and selenocysteine biosynthesis machinery, with predicted roles in oxygen defense (Stairs et al. 2019). The latter proteins are related to the proteins we identified in *M. exilis* (Roxström-Lindquist et al. 2010). Out of the 49 genes encoding the 34 enzymes putatively involved in amino acid metabolism discussed above, 17 are of prokaryotic origin (supplementary table S9, Supplementary Material online) and were likely acquired via LGT. For comparison, the reconstructed amino acid metabolism of *T. vaginalis* contains 36 enzymes, 9 of which were identified as LGT candidates (Carlton et al. 2007). In summary, only the absences of GCS (a strictly mitochondrial complex) and SHMT are directly related to the absence of mitochondria. The loss of these enzymes, which might have accompanied the transition to an endobiotic lifestyle, removed another essential function from the MRO of the *M. exilis* ancestor, preadapting it for loss of the organelle.

In model systems, the mitochondrion is involved in the regulation of calcium homeostasis in the cell. The calcium flux is regulated by opening of Ca^{2+} channels on the cytoplasmic membrane and by pumping of Ca^{2+} into extracellular space and into the internal Ca^{2+} stores. The ER, mitochondria, and other endomembrane vesicles function as these stores (Contreras et al. 2010; García-Sancho 2014). In *M. exilis*, we identified five paralogs of the plasma membrane calcium-cation exchangers which are responsible for the regulation of the intracellular Ca^{2+} concentration (Yu and Choi 1997). Additionally, we identified seven paralogs of P-type Ca^{2+}-ATPases which also transport Ca^{2+} ions across the plasma membrane (Schatzmann 1966) and the ER membrane (Vandecaetsbeek et al. 2011). For comparison, *T. vaginalis* possesses four and six of these respective paralogs (not shown). This suggests that there were no obvious changes in the inventory of Ca^{2+} transporters associated with the loss of mitochondria.

The final systems we specifically examined were related to autophagy and cell death. While analyzing the membrane-trafficking system, we noted the presence of seven homologs of Rab32, including representatives of both main ancestral paralogs in this family (Rab32A and Rab32B; Elias et al. 2012). In mammalian cells, Rab32 proteins are associated with specialized lysosome-derived compartments, ER, mitochondria, and autophagosomes. We hypothesized that some aspects of the extended endolysosomal machinery described above, especially the multiple paralogs of Rab32, in the absence of mitochondria, could be related to the autophagosomal machinery. We therefore further examined the autophagosomal machinery encoded in the *M. exilis* genome (fig. 7 and discussed below).

Autophagy, the process by which large cellular compartments and cytosolic complexes are degraded, involves the mitochondria, as well as the ER, via the regulation of calcium, reactive oxygen-species, and physical association (Gomez-Suaga et al. 2017). Approximately 30 proteins, found broadly conserved across eukaryotes, are involved in the initiation, formation, and function of the autophagosomes (Gomez-Suaga et al. 2017). Interestingly, the AuTophaGy related 1 complex (Atg1; mammalian ULKs 1, 2, and 3) is almost entirely missing in *M. exilis*, *T. vaginalis*, and *N. gruberi*. The exceptions are a single divergent Atg11 homolog in *M. exilis* and Atg1 homologs in the other two protists. As this complex plays a role in the early steps of autophagosome formation, its absence suggests an alternative mechanism for membrane nucleation, as many other core autophagy proteins are present in these organisms. Phagosomal membrane nucleation and elongation downstream of Atg1 occurs by activation of the class III phosphatidylinositol 3-kinase (PtdIns3K) complex, which appears to be present in *M. exilis*. Membrane expansion is mediated by the Atg8 and Atg12 ubiquitinlike conjugation systems. Although the Atg8 complex is present, we did not identify components of the Atg12 Ubl conjugation system in *M. exilis*. As Atg12 is found in both *T. vaginalis* and *N. gruberi*, this pathway may be in the process of being lost in *M. exilis*, suggesting that the Atg8 complex is capable of membrane elongation alone. Indeed, several proteins known to support the function of Atg8, but not considered core autophagy machinery, are present in *M. exilis*, including sequestosome-1 (p62), which binds to Atg8 to facilitate degradation of ubiquitinated proteins, and HOG1, which enhances the stability of Atg8. Retrieval of proteins involved in autophagosome formation is mediated by the Atg9-Atg18 complex, both of which are present in *M. exilis*. However, the Atg18-interacting protein Atg2, which modulates lipid droplet size, is not found in *M. exilis* or *T. vaginalis*. The genomic complement identified suggests that *M. exilis* should be capable of generating autophagosomes, despite some canonical components not being identified. Consistent with this, we observed membranes that resembled the beginning structures of autophagosomes, the Omegasome (supplementary fig. S7 and supplementary videos S1 and S2, Supplementary Material online). However, confirming the identity of this structure,

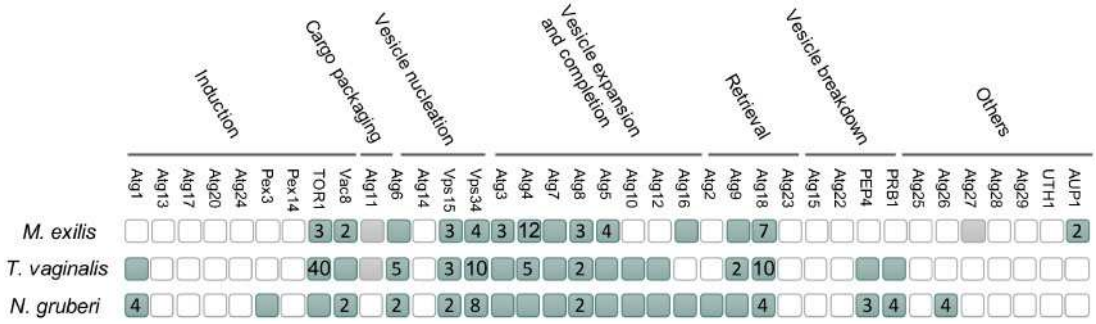


Fig. 7. Autophagy proteins in *Monocercomonoides exilis*, *Trichomonas vaginalis*, and *Naegleria gruberi*. Homologs of autophagy machinery identified by BlastP and pHMMER. Filled squares indicate presence of the component, whereas numbers indicate multiple paralogs. Missing squares indicate that the component could not be identified using these methods. Gray squares indicate a putative homolog whose identity could not be confirmed by reverse BLAST. Categories defined as in Duszenko et al. (2011).

and of the putative MVBs mentioned above, must await development of a transfection system in *M. exilis* for tagging and localization of the relevant molecular components. As might be expected, we could not find any orthologs of the proteins specifically involved in mitophagy (i.e., SLT2, UTH1, PTC6, FUNDC1, BNIP3, and BNIP3L) or pexophagy (i.e., PINK1, PARK2, PEX3, and PEX14) encoded in the *M. exilis* genome. Although a comprehensive analysis of the mitophagy-specific machinery, with deep sampling of eukaryotic genomes, has not been performed, analyses have shown that at least some of the machinery is conserved across the breadth of eukaryotes (Wu et al. 2017). Therefore, the loss of mitophagy proteins represents an example of a system likely lost together with mitochondria.

As the final step in the search for proteins involved in specific adaptations to the loss of mitochondria, we carefully inspected the list of genes that are unique to *M. exilis*, as recovered by Orthofinder (fig. 2). We could ascribe a function to 1,126 of these genes; another 196 genes probably function in trans- or retro-position or are derived from integrated viruses (supplementary table S11, Supplementary Material online). We have not found any example of gene(s) clearly related to the loss of mitochondrial function.

In summary, we have previously suggested that the significant reorganization of Fe–S cluster assembly machinery in *M. exilis* represented the key preadaptation for the loss of mitochondria (Karnkowska et al. 2016) and here we add that another such preadaptation could have been the loss of GCS and SHMT, which took place after the split from the *P. pyriformis* lineage. We have not found any clear case of function lost concomitantly with mitochondrion besides mitophagy. All other cellular systems remained relatively canonical or appear unusual probably due to involvement of host–endobiont interactions and no clear postadaptations to the amitochondriate cell organization were revealed.

Proteins Mediating Mitochondrial Dynamics Are Present in *M. exilis*

Mitochondria are organelles that constantly undergo repeated fission and fusion in order to maintain their number and quality. These dynamics are coordinated with the fundamental functions of mitochondria (Santos et al. 2018). Even reduced mitochondria, such as MROs, undergo dynamics, and many aspects of this process are shared with conventional mitochondria. Given the absence of the mitochondrial organelle in *M. exilis*, proteins involved in mitochondrial fission and fusion are expected to be absent, too. However, many of them are involved in other cellular processes than mitochondria dynamics. We searched *M. exilis* proteome for homologs of proteins annotated into Gene Ontology (GO) categories: mitochondrion localization (GO:0051646; any process in which a mitochondrion or mitochondria are transported to, and/or maintained in, a specific location within the cell) and mitochondrial organization (GO:0007005). Out of 24 identified candidates, only three appeared to be related to mitochondrial dynamics (supplementary table S12, Supplementary Material online). Two of these are dynamin-related protein (DRP) Dnm1 paralogs in *M. exilis* that are closely related to each other and fall phylogenetically to a broader subgroup of DRPs (class A; fig. 8) that includes proteins involved in mitochondrial (and peroxisomal) fission and proteins that seem to have been independently recruited to serve in various parts of the endomembrane system (Praefcke and McMahon 2004; Purkanti and Thattai 2015). In the absence of a mitochondrion, the two *M. exilis* DRPs are predicted to have a role in the dynamics of the endomembrane system, perhaps in endocytosis (in analogy to "true" dynamins and other endocytic DRPs that evolved independently in multiple eukaryotic lineages [Purkanti and Thattai 2015]). Indeed, the single DRP of *G. intestinalis* (also a member of the class A dynamins) seems to be involved in endocytosis (Zumthor et al. 2016), whereas its role in the mitosome dynamics remains unsettled (Rout et al. 2016; Voleman et al. 2017). On the other hand, at least one of the eight DRPs present in *T. vaginalis* contributes to the fission of the hydrogenosomes (Wexler-Cohen et al. 2014), indicating that MRO division in metamonads ancestrally depends on the dynamin family. It is interesting to note that MRO-possessing relatives of *M. exilis*—the trimastigids *Trimastix marina* and *P. pyriformis*—possess two different forms of class A DRPs, one apparently orthologous to the *M. exilis* DRPs and the other without an *M. exilis* counterpart (fig. 8). Since the last common ancestor of trimastigids was probably also the ancestor of oxymonads (Zhang et al. 2015; Leger et al. 2017), the

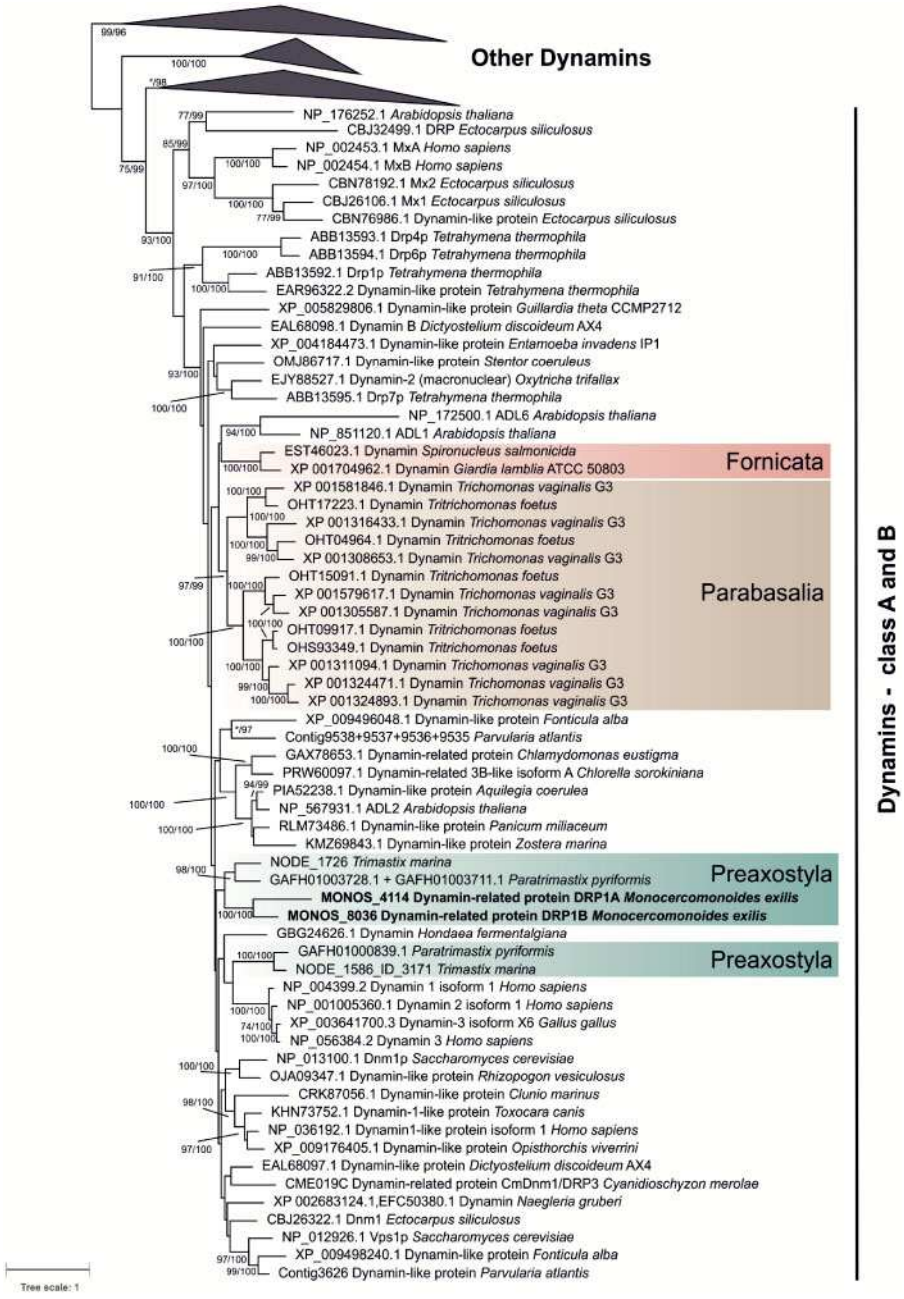


Fig. 8. Phylogenetic analysis of dynamin family showing the position of metamonad dynamins. Clades of Opa1, Mgm1, and Dynamin class C (labeled Other Dynamins) are collapsed since they do not include any metamonad dynamins. Topology is based on phylogenetic tree computed by ML method in RAxML version 8.2.11 (500 rapid bootstraps, PROTGAMMALG4X model). Branch supports were assessed by RAxML rapid bootstraps (500 replicates, only values >70 are shown) and IQ-Tree ultrafast bootstraps (5,000, only values >95 are shown). IQ-Tree 1.5.5 was run under LG+R8 model (based on model test). The final alignment contains 176 sequences and 548 amino acid positions.

absence of this second DRP form in *M. exilis* must be due to secondary loss. It is tempting to speculate that this second DRP form is involved in MRO fission in trimastigids and its absence in *M. exilis* reflects the loss of MROs.

Surprisingly, and in contrast to parabasalids and diplomonads, *M. exilis* also encodes an ortholog of MSTO1 (*misato*) protein. The MSTO1 function has been studied in humans; the protein was shown to be involved in the regulation of mitochondrial distribution and morphology and was proposed to be required for mitochondrial fusion and mitochondrial network formation (Gal et al. 2017). However, in *Drosophila melanogaster* the protein has a non-mitochondrial role, controlling the generation of mitotic microtubules by stabilizing the TCP-1 tubulin chaperone complex (Palumbo et al. 2015). The MSTO1 of *M. exilis* may have a similar nonmitochondrial function.

Conclusions

In this article, we have performed a far more extensive study of the genome of *M. exilis* in comparison to Karnkowska et

al. (2016) and have characterized many additional cellular systems and functions, including 831 new manually curated genes. Our goal was to determine how the loss of mitochondria impacts eukaryotic cell complexity. To our surprise, none of the newly examined cellular systems, including the cytoskeleton, kinetochore, membrane trafficking, autophagy, oxidative stress response, calcium flux, energy, amino acid, lipid, sugar, and nucleotide metabolism, are significantly altered relative to eukaryotes with mitochondria. The large number of introns, elaborate cytoskeleton, endomembrane system, and biosynthetic pathways, all requiring ATP, suggest that *M. exilis* is capable of providing enough ATP without a specialized ATP generating organelle. As such, *M. exilis* serves as living evidence that complex amitochondriate eukaryotic cell is a viable cell type and that such complex cells could have evolved during the evolutionary process of eukaryogenesis before the acquisition of the mitochondrion (Hampl et al. 2019).

From a cell biological perspective, *M. exilis* constitutes a unique experimental model in which to study mitochondrion-integrated systems. With the complete absence of the organelle, such systems can now be interrogated by knock-down or deletion experiments, without the confounding effects of mitochondrion-related functions. Autophagy is one example of such a system, but the same can be said for the functions of dynamins, membrane contact site proteins Vps13 and Vps39, *misato* protein, calcium flux regulatory proteins, Fe–S cluster assembly and more. The comprehensive and high-quality curated predicted proteome provided here, along with the biological insights into various cellular functions of *M. exilis*, should facilitate such future investigations.

Materials and Methods

FIB-SEM Tomography

Soft pellets of cell cultures were fixed in 2.5% glutaraldehyde and 1% formaldehyde in 0.1 M cacodylate buffer for one hour at room temperature (RT), postfixed by reduced 2% (for cell 1) or 1% (for cell 2) osmium tetroxide and 1.5% $K_3(FeCN)_6$ in 0.1 M cacodylate buffer for 1 h on ice (cell 1) or 30 min at RT (cell 2). The cell 1 was further incubated in thiocarbohydrazide for 20 min at RT and in 2% nonreduced osmium tetroxide (in ddH_2O) for 30 min at RT, the cell 2 was incubated in 1% nonreduced osmium tetroxide (in 0.1 M cacodylate buffer) for 30 min at RT. The cells were contrasted by incubating 1 h (cell 1) or 30 min (cell 2) in 1% uranyl acetate at RT, dehydrated in ethanol (cell 2) or ethanol/acetone series (cell 1) and embedded in EPON hard. Serial pictures were taken with FEI Helios NanoLab G3 UC–FIB-SEM microscope with Through lens, In Column and Mirror detectors (TLD, ICD, MD). Raw data were processed in Amira 6 software. First view through the cell in the video is made up of pictures taken with the microscope, whereas the others are calculated subsequently in Amira software.

Ploidy Estimation from *M. exilis* Sequencing Data

The genomic DNA (gDNA) of a clonal culture of *M. exilis* previously sequenced using a Genome Sequencer 454 GS FLX+ at ~35× coverage (Karnkowska et al. 2016) was used to the ploidy estimation. Genomic sequencing reads subjected to linker and quality trimming were mapped onto the previously assembled genome using CLC Genomics Workbench v. 9.5.2 with the following parameters: mismatch cost, 2; insertion cost, 3; deletion cost, 3; length fraction, 0.96; and similarity fraction, 0.96. Duplicate read removal and local realignment were performed using the same software. The resulting read mapping was used as an input for SNP calling with Platypus v. 0.8.1 with a minimum read coverage cut-off of 3 (Rimmer et al. 2014). For ploidy inference, allele frequency distribution at biallelic SNP loci in *M. exilis* was compared with the theoretical distributions in organisms with different ploidy levels (Yoshida et al. 2013).

Protein-Coding Gene Annotation

The previously reported set of gene models predicted by a combination of automated algorithms and manual curation (Karnkowska et al. 2016) was subjected to additional refinement concerning gene categories specifically targeted in the present study. This included incorporation of newly created models for previously missed genes and modification of existing models by changing exon–intron boundaries (sometimes resulting in gene model splitting/fusion) as suggested by transcriptomic evidence and/or sequence conservation within respective gene families. In addition, nine models were removed, since it turned out that the respective scaffolds (scaffold01876, scaffold01882, and scaffold01991) are most likely bacterial contaminants (based on high sequence similarity at the nucleotide level to bacterial genomes).

The automatic functional annotation was performed by similarity searches using BLAST (*e*-value $=< 1e^{-20}$) against the NCBI nr protein database and HMMER (http://hmmer.org/; last accessed 30 June, 2019) searches of domain hits from the PFAM protein family database (Finn et al. 2014). Additional annotation was performed using the KEGG Automatic Annotation Server (Moriya et al. 2007) which compares predicted genes to the manually curated KEGG Genes database (Kanehisa et al. 2014). Gene product names were assigned based on significant BlastP and domain matches. For cases, where there was no significant BLAST or domain hit, the gene was automatically assigned as a "hypothetical protein." GFF3 format was used for storing the annotation information. A set of 1,137 genes of interest was manually curated (supplementary table S1, Supplementary Material online). A locus tag identifier in the format MONOS_XXXXX was assigned to each predicted gene. Approximately 60% of the gene models remained as hypothetical proteins.

Ortholog clustering of translated proteins from annotated draft genome of *M. exilis* was performed with Orthofinder (Emms and Kelly 2015) using predicted proteomes from *T. vaginalis*, *G. intestinalis*, *S. salmonicida*, and *N. gruberi*.

Tyrosine kinases annotation was performed based on homology searches with kinase database (http://kinase.com; last accessed 30 June, 2019). For analysis of proteases the MEROPS database (Rawlings et al. 2016) was used to carry out a BlastP search of all *M. exilis* predicted proteins. Four hundred and

forty-three proteins with an *e*-value $=< 1e^{-10}$ were further analyzed and 122 were checked against MEROPS and confirmed with PFAM. For prediction of Fe–S-cluster-containing proteins the MetalPredator (Valasatava et al. 2016) was used. MetalPredator predicted that proteome of *M. exilis* contains about 54 [Fe–S] proteins. Another [Fe–S] proteins were predicted using BlastP (Altschul et al. 1997) against custom database of experimentally confirmed [Fe–S] proteins from *Escherichia coli* and *Saccharomyces cerevisiae* followed by the reciprocal BLAST against the NCBI nr database. Results were searched against InterPro database (Finn et al. 2017) to confirm presence of [Fe–S] cluster binding motif. TATA-like motif (A/G)TATTT(T/C/G) was searched in the genome assembly of *M. exilis* with the DREME algorithm (Bailey 2011). TATA-like motif was searched among annotated genes with predicted 5′ UTR in the region located 45 nucleotides upstream of the transcription start site.

Gene Searching and Identification

As queries for gene searching, published proteins from various organisms were used, most often from *Arabidopsis thaliana* from www.phytozome.net, *Dictyostelium discoideum* AX4 from NCBI, *G. intestinalis* WB from GiardiaDB.org, *Homo sapiens* from NCBI, *N. gruberi* v1.0 from genome.jgi-psf.org, *P. pyriformis* from NCBI, *T. vaginalis* G3 from TrichDB.org, *Trypanosoma brucei* TREU927 from eupathdb.org, and *Saccharomyces cerevisiae* RM11-1a from www.broad.mit.edu and S288C from NCBI. *Monocercomonoides exilis* hits were BLASTed back against the genome of the query protein and against NCBI nr database.

For identification of rapidly evolving proteins that may be difficult to detect with BLAST, more sensitive searches of the predicted *M. exilis* proteome were carried out using HMMER3.1 package (http://hmmer.org/; last accessed 30 June, 2019). Query HMMs were prepared using the hmmbuild program and input alignments of reference sequences, usually adopted from the Pfam database (seed alignments defined for the families of interest). Positive hits were evaluated as possible orthologs of the query proteins by blasting them against the NCBI protein sequence database.

We performed phylogenetic analyses and generated individual gene trees to support annotation process. Sequences were aligned using MAFFT (Katoh and Standley 2013) or MUSCLE (Edgar 2004), visually inspected and manually edited whenever necessary, and eventually trimmed with BMGE (Criscuolo and Gribaldo 2010) or manually. Maximum likelihood (ML) phylogenetic analyses were performed using one or more methods: RAxML 8.0.23 (Stamatakis 2014), IQ-TREE 1.3.11.1 (Nguyen et al. 2015), Phylobayes v4.1 (Lartillot et al. 2009), or MrBAYES v3.2.2 (Ronquist et al. 2012).

Intron Analyses

The history of intron gains and losses was studied for 100 groups of orthologous genes conserved in *M. exilis* and 33 additional representatives of different phylogenetic lineages of eukaryotes. The genes analyzed were a subset of 163 groups of orthologous genes used in a previously published phylogenomic analysis (Karnkowska et al. 2016). For the intron analyses, we excluded all species with only transcriptomic data available and two more species (*Reticulomyxa filosa* and *Chondrus crispus*) with poor representation of genes in the original data set. For *Acanthamoeba castellanii* and *S. salmonicida* 8 and 21 orthologs, respectively, missing in the original data set were identified by reciprocal blast searches of databases at NCBI and added to the alignments. Finally, groups of orthologs that lacked genes from more than one of the 34 species retained in the analysis were excluded. Information on the exon–intron structure of the genes for most species was obtained from the respective gene records in the GenBank or RefSeq databases at NCBI. For *Bigelowiella natans* and *Phytophthora ramorum*, the information was extracted from GFF files downloaded from the respective genome databases at the Joint Genome Institute (http://jgi.doe.gov/; last accessed 30 June, 2019). For five genes from *A. castellanii*, the respective models were not available in any database, so their exon–intron structure was reconstructed manually. MAFFT v7.271 with auto option (Katoh and Standley 2014) was used to align sets of orthologous protein sequences for subsequent intron mapping and definition of homologous intron positions. These analyses were performed using the Malin software (Csuros 2008). To restrict the analysis to confidently homologized introns, we filtered a total of 5,711 intron positions by keeping only those that were flanked in the protein sequence alignment by four nongap amino acid positions on both sides and that exhibit conservation of a particular amino acid in at least a half of the protein sequences aligned (i.e., "Minimum unambiguous characters at a site" was set to 17). This setting left 3,546 positions for further analyses. We then used the intron table created by Malin and a custom Python script to define ancestral introns, that is, those presumably occurring in the LECA. These were defined by intron positions represented in at least one species of Amorphea and at least one species from the remaining eukaryotic groups included in the analysis (i.e., assuming the position of the root of the eukaryote phylogeny as depicted in fig. 3*a*). The proportion of ancestral introns to the total number of introns was then plotted for each species (fig. 3*b*). Reconstruction of intron gain and loss was done in Malin using Dollo parsimony and three different species trees, using the unrooted topology as defined by our phylogenetic analysis reported previously and assuming three alternative placements of the root of the eukaryote phylogeny: between Amorphea and the remaining eukaryotes included in the analysis (fig. 3*a*); between Amorphea + Metamonada and the remaining eukaryotes included in the analysis (supplementary fig. S4*a*, Supplementary Material online); and between Metamonada and the remaining eukaryotes included in the analysis (supplementary fig. S4*b*, Supplementary Material online).

LGT Pipeline

In order to retrieve putative homologs of *M. exilis* proteins, the 16,629 predicted protein sequences (i.e., the version of the *M. exilis* proteome reported in Karnkowska et al. [2016]) were used as BlastP queries against the *nr* database at the NCBI (*e*-value cut-off: 10^{-10} and a maximum of 1,000 hits). Only data

sets containing at least four sequences were kept (4,733). Probable *M. exilis* in-paralogs (and their homologs retrieved by BlastP) were assembled in a single data set using an in-house Perl script that gathered data sets containing at least 50% of identical sequences. After this step, only 2,146 data sets remained. Since we were interested in LGT specifically from prokaryotes, we discarded data sets containing no prokaryotic homologs. This resulted in set of 824 protein data sets that were further considered for phylogenetic analyses.

A round of preliminary phylogenetic analyses was carried out in order to reduce unnecessary sequence redundancy in a reproducible fashion, and decrease computational time, thus allowing more rigorous downstream analyses. For this, each protein data set was aligned using the "MAFFT" algorithm (default parameters) from the MAFFT package v6.903 (Katoh et al. 2002). Regions of doubtful homology between sites were removed from the alignments using BMGE with default parameters, except for the substitution matrix, which was set to BLOSUM40 ("-m BLOSUM40"), and the gap threshold to 40% ("-g 0.4") (Criscuolo and Gribaldo 2010). At this step, we discarded alignments for which <80 sites were kept after trimming (154 alignments).

Preliminary phylogenetic trees were reconstructed for the remaining 670 protein alignments using FastTree 2.1.4 (with default parameters) (Price et al. 2010). An in-house Perl/Bioperl script was then used to parse these trees and automatically remove unnecessary sequence redundancy in order to reduce the size of each tree. Our method identifies sequences from closely related organisms (i.e., belonging to the same genus) that form a monophyletic clade and keeps only one representative per clade, except for *M. exilis*, for which all (in-) paralogs were kept.

A second round of phylogenetic analyses was performed using more thorough methods. Reduced protein data sets were realigned using the MAFFT-L-INS-i method of the MAFFT package, and then trimmed with BMGE (settings as previously described). ML trees were computed using RAxML 8.0.23 (Stamatakis 2014) with the LG4X model (Le et al. 2012) and statistical support was obtained from 100 rapid bootstrapping (rBS) replicates. Alignments and trees are available upon request.

An automated pipeline was developed, in-house, using Perl/Bioperl to parse phylogenetic trees and screen for LGT candidates using the following criteria: 1) *M. exilis* must branch within a clade containing no other eukaryotes with a few defined exceptions (fornicates, parabasalids, oxymonads, *P. pyriformis*, *N. gruberi*, *E. histolytica*, *Ma. balamuthi*, and *Blastocystis* sp.). These exceptions were allowed because we were interested not only in LGTs specific to *M. exilis* but also in cases of more ancient LGTs (e.g., at the various internal branches of excavate phylogeny). Similarly, we were interested in identifying genes of prokaryotic origin that are shared by *M. exilis* and other anaerobic protists. This clade containing prokaryotes and *M. exilis* (and possibly some of the allowed lineages) must have been supported by a bipartition with the rBS > 70%.

The resulting 174 candidate cases of LGTs were examined by eye taking into account all information contained in each BLAST result, alignment, sequence domain composition (see below), and phylogeny to exclude as many false positives as possible. Seventy-one genes were eventually removed from the list, leaving 103 *M. exilis* genes likely acquired by LGT (supplementary table S9, Supplementary Material online). Protein functional domains for each homolog in a given data set were identified using the HMMER 3 package (http://hmmer.org/; last accessed 30 June, 2019) against the PFAM 26.0 database (Punta et al. 2012) and were mapped onto phylogenetic trees with the ETE2 Python toolkit (Huerta-Cepas et al. 2010).

Fluorescence In Situ Hybridization

Unlabeled telomeric probes were generated using the primer dimer extension method described in Ijdo et al. (1991), but we used PrimeSTAR Max DNA polymerase (Clontech, R045A) instead of Taq polymerase for the PCR step. The purified PCR products were labeled with digoxigenin-11-dUTP, alkali stabile (Roche, 11093088910) using the DecaLabel DNA Labeling Kit (Thermo Scientific, K0621). The labeled probes were purified using columns from the QIAquick Gel Extraction Kit (Qiagen, 28704) and eluted into the final volume of 50 µl.

One liter of *M. exilis* culture was filtered to remove bacteria and the cells were pelleted by centrifugation for 10 min at 1,200 × g at 4 °C. FISH with digoxigenin-labeled probes was performed according to the previously described procedure (Zubáčová et al. 2011) except that the culture was not treated with colchicine and the stringency washes were performed at 45 °C. For probe detection, we used DyLight 594 Labeled Anti-Digoxigenin antibody (Vector Laboratories, DI-7594). Preparations were observed using an IX81 microscope (Olympus) equipped with an IX2-UCB camera. Images were processed using Cell-R software (Olympus) and Image J 1.42q. The number of signals from each nucleus was manually counted and the average number of signals was estimated from at least 50 nuclei.

Southern Blot Analysis of Telomeres

A Southern blot was performed with *M. exilis* gDNA isolated from 3,000 ml of filtered cell culture using DNeasy Blood & Tissue Kit (Qiagen, 69504). Estimation of the average telomere length was based on the method described in Kimura et al. (2010), with slight modifications. Briefly, 5 µg of gDNA was digested overnight in total volume 200 µl containing five units of the *Hin*fI (NEB, R0155S) and five units of *Rsa*I (NEB, R0167S) restriction enzymes. DNA was purified by ethanol precipitation and resuspended in 15 µl of nuclease-free water. After restriction enzyme treatment, the gDNA samples were run on a 0.8% agarose gel at 80 V for 5 h. The DNA was transferred onto a Hybond-N membrane (GE Healthcare) using vacuum blotting. The Southern blot hybridization was performed using the same probe which was used in the FISH procedure. Probe detection was done using the DIG-High Prime DNA Labelling and Detection Starter Kit II (Roche, 11585614910) according to the manufacturer's instructions. Digital images of hybridization signals were obtained using the ImageQuant LAS 4000 (GE Healthcare Life Sciences).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

The work in the V.H. lab was supported by the Czech Science Foundation project 15-16406S. A.K. was supported by the Polish Ministry of Science and Higher Education scholarship for outstanding young researchers. Work in the J.B.D. lab was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (RES0021028). J.B.D. is the Canada Research Chair (Tier II) in Evolutionary Cell Biology. Work in the lab of M.E. was supported by the Czech Science Foundation project 18-18699S. Work in the lab of A.J.R. was supported by a transitional operating grant, MOP-142349, from the Canadian Institutes of Health Research. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 771592). We acknowledge MEYS CR for funding within the National Sustainability Program II (project BIOCEV-FAR) LQ1604, project "BIOCEV" (CZ.1.05/1.1.00/02.0109), and the Centre for research of pathogenicity and virulence of parasites reg. nr.: CZ.02.1.01/0.0/0.0/16019/0000759. Finally, we acknowledge Markéta Dalecká and Adam Schröfel from the Imaging Methods Core Facility at BIOCEV, institution supported by the Czech-BioImaging large RI projects (LM2015062 and CZ.02.1.01/0.0/0.0/16_013/0001775, funded by MEYS CR) for their support with obtaining FIB-SEM data presented in this article.

References

Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S. 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304(5669):441–445.

Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F. et al. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. *J Eukaryot Microbiol*. 66(1):4–119.

Aldritt SM, Tien P, Wang CC. 1985. Pyrimidine salvage in *Giardia lamblia*. *J Exp Med*. 161(3):437–445.

Almeida KH, Sobol RW. 2007. A unified view of base excision repair: lesion-dependent protein complexes regulated by post-translational modification. *DNA Repair (Amst)*. 6(6):695–711.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25(17):3389–3402.

Anderson IJ, Loftus BJ. 2005. *Entamoeba histolytica*: observations on metabolism based on the genome sequence. *Exp Parasitol*. 110(3):173–177.

Andreini C, Rosato A, Banci L. 2017. The relationship between environmental dioxygen and iron-sulfur proteins explored at the genome level. *PLoS One* 12(1):e0171279.

Bailey TL. 2011. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* 27(12):1653–1659.

Baum KF, Berens RL, Marr JJ, Harrington JA, Spector T. 1989. Purine deoxynucleoside salvage in Giardia lamblia. *J Biol Chem*. 264(35):21087–21090.

Brawley SH, Blouin NA, Ficko-Blean E, Wheeler GL, Lohr M, Goodson HV, Jenkins JW, Blaby-Haas CE, Helliwell KE, Chan CX, et al. 2017. Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). *Proc Natl Acad Sci U S A*. 114(31):E6361–E6370.

Brown DM, Upcroft JA, Upcroft P. 1993. Cysteine is the major low-molecular weight thiol in *Giardia duodenalis*. *Mol Biochem Parasitol*. 61(1):155–158.

Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, Wortman JR, Bidwell SL, Alsmark UCM, Besteiro S, et al. 2007. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science* 315(5809):207–212.

Cole ES, Giddings TH, Ozzello C, Winey M, O'Toole E, Orias J, Hamilton E, Guerrier S, Ballard A, Aronstein T. 2015. Membrane dynamics at the nuclear exchange junction during early mating (one to four hours) in the ciliate *Tetrahymena thermophila*. *Eukaryotic Cell* 14(2):116–127.

Contreras L, Drago I, Zampese E, Pozzan T. 2010. Mitochondria: the calcium connection. *Biochim Biophys Acta* 1797(6-7):607–618.

Costa RMA, Chiganças V, Galhardo RDS, Carvalho H, Menck CFM. 2003. The eukaryotic nucleotide excision repair pathway. *Biochimie* 85(11):1083–1099.

Criscuolo A, 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.

Csuros M. 2008. Malin: maximum likelihood analysis of intron evolution in eukaryotes. *Bioinformatics* 24(13):1538–1539.

Dacks JB, Kuru T, Liapounova N, Gedamu L. 2008. Phylogenetic and primary sequence characterization of cathepsin B cysteine proteases from the oxymonad flagellate *Monocercomonoides*. *J Eukaryot Microbiol*. 55(1):9–17.

de Koning HP, Bridges DJ, Burchmore R. 2005. Purine and pyrimidine transport in pathogenic protozoa: from biology to therapy. *FEMS Microbiol Rev*. 29(5):987–1020.

de Mendoza A, Sebe-Pedros A, Sestak MS, Matejcic M, Torruella G, Domazet-Loso T, Ruiz-Trillo I. 2013. Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proc Natl Acad Sci U S A*. 110(50):E4858–E4866.

Duszenko M, Ginger ML, Brennand A, Gualdrón-López M, Colombo MI, Coombs GH, Coppens I, Jayabalasingham B, Langsley G, de Castro SL, et al. 2011. Autophagy in protists. *Autophagy* 7(2):127–158.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5):1792–1797.

Eichinger L, Pachebat JA, Glöckner G, Rajandream M-A, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, et al. 2005. The genome of the social amoeba *Dictyostelium discoideum*. *Nature* 435(7038):43–57.

Elias M, Brighouse A, Gabernet-Castello C, Field MC, Dacks JB. 2012. Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. *J Cell Sci*. 125(Pt 10):2500–2508.

Eme L, Gentekaki E, Curtis B, Archibald JM, Roger AJ. 2017. Lateral gene transfer in the adaptation of the anaerobic parasite *Blastocystis* to the gut. *Curr Biol*. 27(6):807–820.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. 16:157.

Findeisen P, Mühlhausen S, Dempewolf S, Hertzog J, Zietlow A, Carlomagno T, Kollmar M. 2014. Six subgroups and extensive recent duplications characterize the evolution of the eukaryotic tubulin protein family. *Genome Biol Evol*. 6(9):2274–2288.

Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi Z, El-Gebali S, Fraser M, et al. 2017. InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res*. 45(D1):D190–199.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. 2014. Pfam: the protein families database. *Nucleic Acids Res*. 42(Database issue):D222–D230.

Forterre P, Gribaldo S, Gadelle D, Serre MC. 2007. Origin and evolution of DNA topoisomerases. *Biochimie* 89(4):427–446.

Franzén O, Jerlström-Hultqvist J, Einarsson E, Ankarklev J, Ferella M, Andersson B, Svärd SG. 2013. Transcriptome profiling of *Giardia intestinalis* using strand-specific RNA-Seq. *PLoS Comput Biol*. 9(3):e1003000.

Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, et al. 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* 140(5):631–642.

Fukui K. 2010. DNA mismatch repair in eukaryotes and bacteria. *J Nucleic Acids*. 2010:1–6.

Fulnecková J, Sevcíková T, Fajkus J, Lukesová A, Lukes M, Vlcek C, Lang BF, Kim E, Eliás M, Sykorová E. 2013. A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biol Evol*. 5(3):468–483.

Gabaldón T, Ginger ML, Michels P. 2016. Peroxisomes in parasitic protists. *Mol Biochem Parasitol*. 209(1-2):35–45.

Gal A, Balicza P, Weaver D, Naghdi S, Joseph SK, Várnai P, Gyuris T, Horváth A, Nagy L, Seifert EL, et al. 2017. MSTO1 is a cytoplasmic pro-mitochondrial fusion protein, whose mutation induces myopathy and ataxia in humans. *EMBO Mol Med*. 9(7):967–984.

García-Sancho J. 2014. The coupling of plasma membrane calcium entry to calcium uptake by endoplasmic reticulum and mitochondria. *J. Physiol*. 2:261–268.

Ginger ML, Fritz-Laylin LK, Fulton C, Cande WZ, Dawson SC. 2010. Intermediary metabolism in protists: a sequence-based view of facultative anaerobic metabolism in evolutionarily diverse eukaryotes. *Protist* 161(5):642–671.

Gomez-Suaga P, Paillusson S, Miller C. 2017. ER-mitochondria signaling regulates autophagy. *Autophagy* 13(7):1250–1251.

González Montoro A, Auffarth K, Hönscher C, Bohnert M, Becker T, Warscheid B, Reggiori F, van der Laan M, Fröhlich F, Ungermann C. 2018. Vps39 interacts with Tom40 to establish one of two functionally distinct vacuole-mitochondria contact sites. *Dev Cell* 45(5):621–636.e7.

Gray MW. 2012. Mitochondrial evolution. *Cold Spring Harb Perspect Biol*. 4(9):a011403.

Hammesfahr B, Kollmar M. 2012. Evolution of the eukaryotic dynactin complex, the activator of cytoplasmic dynein. *BMC Evol Biol*. 12:95.

Hampl V. 2017. Preaxostyla. In: Archibald J, Simpson A, Slamovits C, editors. Handbook of the Protists. Cham (Switzerland): Springer International Publishing. p. 1139–1174.

Hampl V, Čepička I, Eliáš M. 2019. Was the mitochondrion necessary to start eukaryogenesis? *Trends Microbiol*. 27(2):96–104.

Hampl V, Silberman JD, Stechmann A, Diaz-Triviño S, Johnson PJ, Roger AJ. 2008. Genetic evidence for a mitochondriate ancestry in the "amitochondriate" flagellate *Trimastix pyriformis*. *PLoS One* 3(1):e1383.

Huerta-Cepas J, Dopazo J, Gabaldón T. 2010. ETE: a python Environment for Tree Exploration. *BMC Bioinformatics* 11:24.

Ijdo JW, Wells RA, Baldini A, Reeders ST. 1991. Improved telomere detection using a telomere repeat probe (TTAGGG)n generated by PCR. *Nucleic Acids Res*. 19(17):4780.

Irimia M, Roy SW. 2014. Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb Perspect Biol*. 6(6):a016071.

Iyer LM, Anantharaman V, Wolf MY, Aravind L. 2008. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol*. 38(1):1–31.

Kamikawa R, Inagaki Y, Roger AJ, Hashimoto T. 2011. Splintrons in *Giardia intestinalis*: spliceosomal introns in a split form. *Commun Integr Biol*. 4(4):454–456.

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*. 42(Database issue):D199–D205.

Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Žárský V, Barlow LD, Herman EK, et al. 2016. A eukaryote without a mitochondrial organelle. *Curr Biol*. 26(10):1274–1284.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30(14):3059–3066.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.

Katoh K, Standley DM. 2014. MAFFT: iterative refinement and additional methods. *Methods Mol Biol*. 1079:131–146.

Kendall JJ, Barrero-Tobon AM, Hendrixson DR, Kelly DJ. 2014. Hemerythrins in the microaerophilic bacterium *Campylobacter jejuni* help protect key iron-sulphur cluster enzymes from oxidative damage. *Environ Microbiol*. 16(4):1105–1121.

Kimura M, Stone RC, Hunt SC, Skurnick J, Lu X, Cao X, Harley CB, Aviv A. 2010. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. *Nat Protoc*. 5(9):1596–1607.

Klinger CM, Klute MJ, Dacks JB. 2013. Comparative genomic analysis of multi-subunit tethering complexes demonstrates an ancient pan-eukaryotic complement and sculpting in *Apicomplexa*. *PLoS One* 8(9):e76278.

Kollmar M, Lbik D, Enge S. 2012. Evolution of the eukaryotic ARP2/3 activators of the WASP family: WASP, WAVE, WASH, and WHAMM, and the proposed new family members WAWH and WAML. *BMC Res Notes* 5:88.

Koumandou VL, Dacks JB, Coulson RMR, Field MC. 2007. Control systems for membrane fusion in the ancestral eukaryote; evolution of tethering complexes and SM proteins. *BMC Evol Biol*. 7:1–17.

Koumandou VL, Klute MJ, Herman EK, Nunez-Miguel R, Dacks JB, Field MC. 2011. Evolutionary reconstruction of the retromer complex and its function in *Trypanosoma brucei*. *J Cell Sci*. 124(Pt 9):1496–1509.

Kumari P, Babuta M, Bhattacharya A, Gourinath S. 2019. Structural and functional characterisation of phosphoserine phosphatase, that plays critical role in the oxidative stress response in the parasite *Entamoeba histolytica*. *J Struct Biol*. 206(2):254-266.

Kunkel TA, Erie DA. 2005. DNA mismatch repair. *Annu Rev Biochem*. 74:681–710.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25(17):2286–2288.

Latchman DS. 1997. Transcription factors: an overview. *Int J Biochem Cell Biol*. 29(12):1305–1312.

Le SQ, Dang CC, Gascuel O. 2012. Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Mol Biol Evol*. 29(10):2921–2936.

Leger MM, Kolisko M, Kamikawa R, Stairs CW, Kume K, Čepička I, Silberman JD, Andersson JO, Xu F, Yabuki A, et al. 2017. Organelles that illuminate the origins of *Trichomonas* hydrogenosomes and *Giardia* mitosomes. *Nat Ecol Evol*. 1(4):0092.

Leung KF, Dacks JB, Field MC. 2008. Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage. *Traffic* 9(10):1698–1716.

Lill R, Hoffmann B, Molik S, Pierik AJ, Rietzschel N, Stehling O, Uzarska MA, Webert H, Wilbrecht C, Mühlenhoff U. 2012. The role of mitochondria in cellular iron-sulfur protein biogenesis and iron metabolism. *Biochim Biophys Acta* 1823(9):1491–1508.

Loftus B, Anderson I, Davies R, Alsmark UCM, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ, et al. 2005. The genome of the protist parasite *Entamoeba histolytica*. *Nature* 433(7028):865–868.

Lynch M, Marinov GK. 2015. The bioenergetic costs of a gene. *Proc Natl Acad Sci U S A*. 112(51):15690–15695.

Manning G, Reiner DS, Lauwaet T, Dacre M, Smith A, Zhai Y, Svard S, Gillin FD. 2011. The minimal kinome of *Giardia lamblia* illuminates early kinase evolution and unique parasite biology. *Genome Biol*. 12(7):R66.

Marchat LA, López-Camarillo C, Orozco E, López-Casamichana M. 2011. DNA repair in pathogenic eukaryotic cells: insights from comparative genomics of parasitic protozoan. In: Chen C, editor. Selected

topics in DNA repair. London, United Kingdom: INTECH Open Access Publisher.
Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35(Web Server issue):W182–W185.
Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, et al. 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* 317(5846):1921–1926.
Müller S, Liebau E, Walter RD, Krauth-Siegel RL. 2003. Thiol-based redox metabolism of protozoan parasites. *Trends Parasitol.* 19(7):320–328.
Munagala N, Wang CC. 2002. The pivotal role of guanine phosphoribosyltransferase in purine salvage by *Giardia lamblia*. *Mol Microbiol.* 44(4):1073–1079.
Munagala NR, Wang CC. 2003. Adenosine is the primary precursor of all purine nucleotides in *Trichomonas vaginalis*. *Mol Biochem Parasitol.* 127(2):143–149.
Narrowe AB, Spang A, Stairs CW, Caceres EF, Baker BJ, Miller CS, Ettema T. 2018. Complex evolutionary history of translation elongation factor 2 and diphthamide biosynthesis in Archaea and parabasalids. *Genome Biol Evol.* 10(9):2380–2393.
Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
Novák L, Zubáčová Z, Karnkowska A, Kolisko M, Hroudová M, Stairs CW, Simpson AGB, Keeling PJ, Roger AJ, Čepička I, et al. 2016. Arginine deiminase pathway enzymes: evolutionary history in metamonads and other eukaryotes. *BMC Evol Biol.* 16:197.
Nývltová E, Šuťák R, Žárský V, Harant K, Hrdý I, Tachezy J. 2017. Lateral gene transfer of *p*-cresol- and indole-producing enzymes from environmental bacteria to *Mastigamoeba balamuthi*. *Environ Microbiol.* 19(3):1091–1102.
Oma Y, Harata M. 2011. Actin-related proteins localized in the nucleus: from discovery to novel roles in nuclear organization. *Nucleus* 2(1):38–46.
Orphanides G, Lagrange T, Reinberg D. 1996. The general transcription factors of RNA polymerase II. *Genes Dev.* 10(21):2657–2683.
Palumbo V, Pellacani C, Heesom KJ, Rogala KB, Deane CM, Mottier-Pavie V, Gatti M, Bonaccorsi S, Wakefield JG. 2015. Misato controls mitotic microtubule generation by stabilizing the TCP-1 tubulin chaperone complex. *Curr Biol.* 25(13):1777–1783.
Paredez AR, Nayeri A, Xu JW, Krtkova J, Cande WZ. 2014. Identification of obscure yet conserved actin-associated proteins in *Giardia lamblia*. *Eukaryotic Cell* 13(6):776–784.
Park J-S, Thorsness MK, Policastro R, McGoldrick LL, Hollingsworth NM, Thorsness PE, Neiman AM. 2016. Yeast Vps13 promotes mitochondrial function and is localized at membrane contact sites. *Mol Biol Cell* 27(15):2435–2449.
Praefcke GJK, McMahon HT. 2004. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nat Rev Mol Cell Biol.* 5(2):133–147.
Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):e9490.
Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40(Database issue):D290–D301.
Purkanti R, Thattai M. 2015. Ancient dynamin segments capture early stages of host–mitochondrial integration. *Proc Natl Acad Sci U S A.* 112(9):2800–2805.
Radek R. 1994. *Monocercomonoides termitis* n. sp., an oxymonad from the lower termite *Kalotermes sinaicus*. *Arch Protistenkd.* 144(4):373–382.
Rawlings ND, Barrett AJ, Finn R. 2016. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 44(D1):D343–D350.
Rawlings ND, Morton FR, Kok CY, Kong J, Barrett AJ. 2008. MEROPS: the peptidase database. *Nucleic Acids Res.* 36(Database issue):D320–D325.
Ribeiro KC, Mariante RM, Coutinho LL, Benchimol M. 2002. Nucleus behavior during the closed mitosis of *Tritrichomonas foetus*. *Biol Cell* 94(4-5):289–301.
Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Wilkie AOM, McVean G, Lunter G. 2014. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet.* 46(8):912–918.
Roberts AJ, Kon T, Knight PJ, Sutoh K, Burgess SA. 2013. Functions and mechanics of dynein motor proteins. *Nat Rev Mol Cell Biol.* 14(11):713–726.
Roger A, Muñoz-Gómez SA, Kamikawa R. 2017. The origin and diversification of mitochondria. *Curr Biol.* 27(21):R1177–R1192.
Rogozin IB, Carmel L, Csuros M, Koonin EV. 2012. Origin and evolution of spliceosomal introns. *Biol Direct.* 7:11.
Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61(3):539–542.
Rout S, Zumthor JP, Schraner EM, Faso C, Hehl AB. 2016. An interactome-centered protein discovery approach reveals novel components involved in mitosome function and homeostasis in *Giardia lamblia*. *PLoS Pathog.* 12(12):e1006036.
Roxström-Lindquist K, Jerlström-Hultqvist J, Jørgensen A, Troell K, Svärd SG, Andersson JO. 2010. Large genomic differences between the morphologically indistinguishable diplomonads *Spironucleus barkhanus* and *Spironucleus salmonicida*. *BMC Genomics.* 11:258.
Roy SW. 2017. Transcriptomic analysis of diplomonad parasites reveals a trans-spliced intron in a helicase gene in *Giardia*. *PeerJ* 5:e2861.
Sagolla MS, Dawson SC, Mancuso JJ, Cande WZ. 2006. Three-dimensional analysis of mitosis and cytokinesis in the binucleate parasite *Giardia intestinalis*. *J Cell Sci.* 119(Pt 23):4889–4900.
Sajid M, McKerrow JH. 2002. Cysteine proteases of parasitic organisms. *Mol Biochem Parasitol.* 120(1):1–21.
Santos HJ, Makiuchi T, Nozaki T. 2018. Reinventing an organelle: the reduced mitochondrion in parasitic protists. *Trends Parasitol.* 34(12):1038–1055.
Schatzmann HJ. 1966. ATP-dependent Ca^{++}-extrusion from human red cells. *Experientia* 22(6):364–365.
Schorey JS, Cheng Y, Singh PP, Smith VL. 2015. Exosomes and other extracellular vesicles in host-pathogen interactions. *EMBO Rep.* 16(1):24–43.
Seaman MNJ. 2004. Cargo-selective endosomal sorting for retrieval to the Golgi requires retromer. *J Cell Biol.* 165(1):111–122.
Sebé-Pedrós A, Grau-Bové X, Richards TA, Ruiz-Trillo I. 2014. Evolution and classification of myosins, a paneukaryotic whole-genome approach. *Genome Biol Evol.* 6(2):290–305.
Slamovits CH, Keeling PJ. 2006. A high density of ancient spliceosomal introns in oxymonad excavates. *BMC Evol Biol.* 6:34.
Smith AJ, Chudnovsky L, Simoes-Barbosa A, Delgadillo-Correa MG, Jonsson ZO, Wohlschlegel JA, Johnson PJ. 2011. Novel core promoter elements and a cognate transcription factor in the divergent unicellular eukaryote *Trichomonas vaginalis*. *Mol Cell Biol.* 31(7):1444–1458.
Spaans SK, Weusthuis RA, van der Oost J, Kengen SWM. 2015. NADPH-generating systems in bacteria and archaea. *Front Microbiol.* 6:742.
Stairs CW, Kokla A, Ástvaldsson Á, Jerlström-Hultqvist J, Svärd S, Ettema T. 2019. Oxygen induces the expression of invasion and stress response genes in the anaerobic salmon parasite *Spironucleus salmonicida*. *BMC Biol.* 17(1):19.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
Su X, Lin Z, Lin H. 2013. The biosynthesis and biological function of diphthamide. *Crit Rev Biochem Mol Biol.* 48(6):515–521.
Tang H, Sun X, Reinberg D, Ebright RH. 1996. Protein–protein interactions in eukaryotic transcription initiation: structure of the preinitiation complex. *Proc Natl Acad Sci U S A.* 93(3):1119–1124.
Tanifuji G, Takabayashi S, Kume K, Takagi M, Nakayama T, Kamikawa R, Inagaki Y, Hashimoto T. 2018. The draft genome of *Kipferlia bialata*

reveals reductive genome evolution in fornicate parasites. *PLoS One* 13(3):e0194487.
Treitli SC, Kotyk M, Yubuki N, Jirounková E, Vlasáková J, Smejkalová P, Šípek P, Čepička I, Hampl V. 2018. Molecular and morphological diversity of the oxymonad genera *Monocercomonoides* and *Blattamonas* gen. nov. *Protist* 169(5):744–783.
Vacek V, Novák LVF, Treitli SC, Táborský P, Čepička I, Kolísko M, Keeling PJ, Hampl V. 2018. Fe–S cluster assembly in oxymonads and related protists. *Mol Biol Evol*. 35(11):2712–2718.
Valasatava Y, Rosato A, Banci L, Andreini C. 2016. MetalPredator: a web server to predict iron–sulfur cluster binding proteomes. *Bioinformatics* 32(18):2850–2852.
van Hooff JJ, Tromer E, van Wijk LM, Snel B, Kops GJ. 2017. Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO Rep*. 18(9):1559–1571.
Vandecaetsbeek I, Vangheluwe P, Raeymaekers L, Wuytack F, Vanoevelen J. 2011. The Ca^{2+} pumps of the endoplasmic reticulum and Golgi apparatus. *Cold Spring Harb Perspect Biol*. 3(5):pii: a004184.
Voleman L, Najdrová V, Ástvaldsson Á, Tůmová P, Einarsson E, Švindrych Z, Hagen GM, Tachezy J, Svärd SG, Doležal P. 2017. *Giardia intestinalis* mitosomes undergo synchronized fission but not fusion and are constitutively associated with the endoplasmic reticulum. *BMC Biol*. 15(1):27.
Wang CC, Cheng HW. 1984. Salvage of pyrimidine nucleosides by *Trichomonas vaginalis*. *Mol Biochem Parasitol*. 10(2):171–184.
Westrop GD, Goodall G, Mottram JC, Coombs GH. 2006. Cysteine biosynthesis in *Trichomonas vaginalis* involves cysteine synthase utilizing O-phosphoserine. *J Biol Chem*. 281(35):25062–25075.
Wexler-Cohen Y, Stevens GC, Barnoy E, van der Bliek AM, Johnson PJ. 2014. A dynamin-related protein contributes to *Trichomonas vaginalis* hydrogenosomal fission. *FASEB J*. 28(3):1113–1121.
Wickstead B, Gull K. 2012. Evolutionary biology of dyneins. In: Dyneins. Cambridge, Massachusetts US: Elsevier. p. 88–121.
Wickstead B, Gull K, Richards TA. 2010. Patterns of kinesin evolution reveal a complex ancestral eukaryote with a multifunctional cytoskeleton. *BMC Evol Biol*. 10:110.
Williams BAP, Hirt RP, Lucocq JM, Embley TM. 2002. A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature* 418(6900):865–869.
Wu X, Wu F-H, Wu Q, Zhang S, Chen S, Sima M. 2017. Phylogenetic and molecular evolutionary analysis of mitophagy receptors under hypoxic conditions. *Front Physiol*. 8:539.
Xu F, Jerlström-Hultqvist J, Einarsson E, Astvaldsson A, Svärd SG, Andersson JO. 2014. The genome of *Spironucleus salmonicida* highlights a fish pathogen adapted to fluctuating environments. *PLoS Genet*. 10(2):e1004053.
Yamamoto Y, Sakisaka T. 2018. The peroxisome biogenesis factors post-translationally target reticulon homology domain-containing proteins to the endoplasmic reticulum membrane. *Sci Rep*. 8(1):2322.
Yeshaw WM, van der Zwaag M, Pinto F, Lahaye LL, Faber AI, Gómez-Sánchez R, Dolga AM, Poland C, Monaco AP, van IJzendoorn SC, et al. 2019. Human VPS13A is associated with multiple organelles and influences mitochondrial morphology and lipid droplet motility. *Elife* 8:pii: e43561.
Yoshida K, Schuenemann VJ, Cano LM, Pais M, Mishra B, Sharma R, Lanz C, Martin FN, Kamoun S, Krause J, et al. 2013. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *Elife* 28(2):e00731.
Yu SP, Choi DW. 1997. Na^{+}-Ca^{2+} exchange currents in cortical neurons: concomitant forward and reverse operation and effect of glutamate. *Eur J Neurosci*. 9(6):1273–1281.
Žárský V, Tachezy J. 2015. Evolutionary loss of peroxisomes—not limited to parasites. *Biol Direct* 10:74.
Zhang Q, Táborský P, Silberman JD, Pánek T, Čepička I, Simpson A. 2015. Marine isolates of *Trimastix marina* form a plesiomorphic deep-branching lineage within Preaxostyla, separate from other known trimastigids (*Paratrimastix* n. gen.). *Protist* 166(4):468–491.
Zubáčová Z, Krylov V, Tachezy J. 2011. Fluorescence in situ hybridization (FISH) mapping of single copy genes on *Trichomonas vaginalis* chromosomes. *Mol Biochem Parasitol*. 176(2):135–137.
Zubáčová Z, Novák L, Bublíková J, Vacek V, Fousek J, Rídl J, Tachezy J, Doležal P, Vlček C, Hampl V. 2013. The mitochondrion-like organelle of *Trimastix pyriformis* contains the complete glycine cleavage system. *PLoS One* 8(3):e55417.
Zumthor JP, Cernikova L, Rout S, Kaech A, Faso C, Hehl AB. 2016. Static clathrin assemblies at the peripheral vacuole—plasma membrane interface of the parasitic protozoan *Giardia lamblia*. *PLoS Pathog*. 12(7):e1005756.

Supplement: Table of contents

1. Supplement I: Genomics of *Paratrimastix pyriformis* (work in progress)

Here is included an excerpt from a comparative genomic project in progress, which includes only work by the author of this thesis, i. e. assembly and annotation of the *Paratrimastix pyriformis* genome. The data and analyses shown here are therefore necessarily incomplete. This work was performed in coordination and cooperation with Sebastian C. Treitli, M.Sc.

1.1. Introduction

So far there has been only one species of eukaryotes known to science which has completely lost its mitochondria, the chinchilla gut symbiont *Monocercomonoides exilis* (Oxymonadida, Preaxostyla, Metamonada). The amitochondriate status of *M. exilis* has been thoroughly corroborated by a genomic project which failed to identify any mitochondrion-associated genes, while showing multiple other eukaryotic cellular systems to be completely represented (Karnkowska et al. 2016).

Following exhaustive functional annotation of various cellular system has shown, that the loss of mitochondria had little to no impact on the overall functioning of the *M. exilis* cell, which exhibits a conventional eukaryotic complexity comparable to other anaerobic protists (Karnkowska et al. 2019). Existence of such organism implies that mitochondria are not necessary for thriving of complex eukaryotic organisms, which informs our thinking about the origin of eukaryotes (Hampl, Čepička, and Eliáš 2019).

Oxymonadida contains, besides *M. exilis*, approximately 140 species of morphologically divergent and diverse flagellates exclusively inhabiting digestive tract of metazoans, none of which was clearly shown to possess a mitochondrion (Hampl 2017). It is therefore possible, that the entire Oxymonadida is an amitochondriate taxon. Sister to Oxymonadida are Paratrimastigidae, and sister to these two taxa are Trimastigidae, both of which include free-living, anaerobic, bacteriovorous flagellates with typical “excavate” morphology and ultrastructure (Treitli et al. 2018). *Paratrimastix pyriformis* (Paratrimastigidae) has been shown to possess a mitochondrion-related organelle (MRO) possibly resembling a hydrogenosome (Hampl et al. 2008; Zubáčová et al. 2013).

Transcriptomic and genomic assemblies of 5 members of Preaxostyla are available to us: *Trimastix marina* (Trimastigidae), *Paratrimastix pyriformis* (Paratrimastigidae), *Monocercomonoides exilis*, *Blattamonas nauphoetae*, and *Streblomastix strix* (Oxymonadida).

The transcriptome of *T. marina* (Leger et al. 2017), genome of *M. exilis* (Karnkowska et al. 2016), and single-cell metagenome of *S. strix* (Treitli et al. 2019) have been already published. The genomes of *P. pyriformis* and *B. nauphoetae* (data not shown here) are reported here for the first time. Comparative functional annotation of these datasets is used to map out events of the reductive evolution of mitochondria in Preaxostyla and to illuminate the adaptations connected to amitochondriality, anaerobiosis, and endobiosis in this group of protists.

1.2. Methods

1.2.1. *Paratrimastix pyriformis* cultivation and harvesting

Monoeukaryotic, xenic culture of *P. pyriformis* (strain RCP-MX, ATCC 50935) was maintained in the Sonneborn's *Paramecium* medium ATCC 802 at room temperature by serial transfer of 1 ml of well-grown culture (approximately $5x10^5$ cells in 1 ml) into a 15 ml test tube containing 10 ml of fresh, bacterized medium every 7 days. The medium was bacterized 24 hours before the transfer. Large volumes of the culture needed for DNA extraction were obtained by serial transfer of a well-grown culture into increasingly larger containers with a fresh bacterized medium over two weeks, resulting in 2 l of well-grown culture per 10 ml of the initial inoculum.

A two-step filtration of the culture was used in order to increase the concentration of *P. pyriformis* cells and reduce the bacterial contamination. First, the culture was filtered through the Macherey-Nagel MN 617 ¼ Folded Filter Paper and the flow-through was collected. Second, the flow-through was filtered through the Whatman Nuclepore Track-Etched Membrane (pore size: 3 μm) and washed twice with 3% LB medium. After the last wash, several milliliters of the liquid remaining above the filter and containing the concentrated *P. pyriformis* cells were collected. The liquid containing the cells was constantly manually stirred using plastic transfer pipettes during the filtration and washing in order to fasten the process.

1.2.2. RNA and DNA extraction and sequencing

The RNA for Illumina transcriptome sequencing was isolated from 10 l of a well-grown culture of *P. pyriformis* using the TRI reagent procedure (Chomczynski and Mackey 1995). The eukaryotic mRNA was purified based on the poly A residues using the Dynabeads mRNA Purification Kit for mRNA Purification from Total RNA preps (Thermo Fisher Scientific). The cDNA was synthesized by the SMARTer PCR cDNA Synthesis Kit (Takara Bio Group).

Sequencing of the cDNA was performed using the Illumina HiSeq 2000 sequencer at the company BGI Tech Solutions Co., Ltd.

The gDNA samples for PacBio, Illumina HiSeq, and Illumina MiSeq sequencing were each isolated from approximately 15 l of a well-grown culture of *P. pyriformis* using the Qiagen DNeasy Blood & Tissue Kit. The isolated gDNA was further ethanol-precipitated (using sodium acetate for neutralization) in order to increase the concentration and remove contaminants. The final concentration of gDNA used for PacBio sequencing (measured using the NanoDrop ND-1000 UV-Vis spectrophotometer) was 107 ng/µl in the volume of 5 µl. The PacBio sequencing was performed using the PacBio RSII sequencer at the company SEQme s.r.o. The Illumina sequencing was performed using the Illumina HiSeq and MiSeq sequencers at the Institute of Molecular Genetics of the ASCR, v. v. i.

The gDNA for MinION sequencing was isolated from 7 l of a well-grown culture of *P. pyriformis* using the Qiagen MagAttract HMW DNA Kit. The isolated gDNA was further ethanol-precipitated (using sodium acetate for neutralization) in order to increase the concentration and remove contaminants. The final concentration of gDNA (measured using the Promega Quantus Fluorometer) was 40 ng/µl in the volume of 300 µl.

Libraries for MinION sequencing were prepared using the Oxford Nanopore Technologies SQK-LSK108 Ligation Sequencing KiW 1D. Sequencing was performed using the Oxford Nanopore Technologies 9.4 MinION Flowcells. Adapter sequences on the ends of the raw reads were trimmed using Porechop v0.2.3 (https://github.com/rrwick/Porechop).

1.2.3. Genome assembly and processing

The raw reads produced by the MinION and PacBio runs were assembled using Canu v1.7.1 assembler (Koren et al. 2017). The prokaryotic contamination was removed from the draft genomic assembly first semi-automatically using tetranucleotide frequencies based on the Emergent Self-Organizing Maps (ESOM) pipeline (Haddad et al. 2009) and later manually using BLASTn and BLASTp searches against the NCBI databases. After the decontamination, the first round of correcting (or "polishing") was performed using Nanopolish with the MinION-acquired raw reads (Loman, Quick, and Simpson 2015). The assembly was improved (or "scaffolded") using Rascaf (RnA-seq SCAFfolder) (Song, Shankar, and Florea 2016) with the raw RNA-seq reads in order to get longer and more continuous scaffolds and more complete

gene models. The resulting scaffolds were again corrected, now using 12 runs of Pilon v1.21 with the Illumina HiSeq and MiSeq raw reads (Walker et al. 2014).

1.2.4. Gene prediction

The raw transcriptomic reads were aligned to the genomic assembly using the Burrows-Wheeler Aligner BWA-MEM algorithm (Li and Durbin 2009) and inspected in the Integrative Genomics Viewer (IGV) (Robinson et al. 2011) in order to manually identify protein-coding gene sequences and their intron/exon boundaries. In total, 708 protein-coding gene models were manually prepared and stored in the Generic Feature Format 3 (GFF3, http://gmod.org/wiki/GFF3) text files.

The manually prepared gene models were used to train the Augustus v3.2.3 automatic gene predictor (Stanke et al. 2004). The raw transcriptomic reads were aligned to the genomic assembly using Bowtie 2 (Langmead and Salzberg 2012) and TopHat2 (Kim et al. 2013) aligners. The Augustus script bam2hints was used to prepare intron hints from this alignment. Gene models were automatically predicted using trained Augustus, incorporating the intron hints. Second set of gene models was produced using the Program to Assemble Spliced Alignments (PASA) pipeline v2.0.2 (Haas et al. 2003).

Repetitive elements in the genomic assembly were identified using RepeatModeler v1.0.11 (Tarailo-Graovac and Chen 2009). Only those models which were identified as members of known repeat families were kept. The final gene models were produced using the EVidenceModeler (EVM) v1.1.1 (Haas et al. 2008), combining the gene models produced by Augustus and PASA, and the models of repetitive sequences produced by RepeatModeler. Completeness of the predicted gene set was estimated using the Benchmarking Universal Single-Copy Orthologs (BUSCO) v3 with the eukaryota_odb9 dataset (Simão et al. 2015).

1.2.5. Gene annotation

Automatic gene annotation was performed on the whole set of predicted genes using KAAS search (Moriya et al. 2007) against the KEGG database (Kanehisa et al. 2014), InterProScan (Zdobnov and Apweiler 2001), and BLAST search against the NCBI nr database. Manual search and annotation were performed by searching the predicted proteome using BLAST and HMMER (Finn, Clements, and Eddy 2011) tools with manually selected genes of interest from other organisms. The candidate genes were then validated by reciprocal BLAST against NCBI nr database and eventually by TMHMM Server v. 2.0 (Krogh et al. 2001) in case

transmembrane helices were expected. Gene models of the confirmed annotated genes were manually validated and improved if necessary. The resulting validated and annotated gene models were deposited in a Generic Feature Format 3 (GFF3, http://gmod.org/wiki/GFF3) text file. A list of all manually annotated genes can be downloaded here:
https://lukasnovak.eu/phdthesis/

1.3. Results and discussion

1.3.1. Genome characteristics

Paratrimastix pyriformis genome was assembled into 650 contigs spanning 56,722,987 base pairs, with N50 = 268,802 base pairs and N75 = 113,394 base pairs and 2137 N's. The largest contig is 1,707,755 base pairs long. The overall guanine-cytosine content of the assembly is 60.92%. The manual and automatic gene prediction resulted in 13,533 predicted protein-coding genes. From these, 4,605 genes were assigned a functional annotation in the automatic annotation step. BUSCO v3 eukaryota_odb9 completeness of the predicted proteome is 82.1%.

1.3.2. MRO and energy metabolism

All the previously identified (Zubáčová et al. 2013) components of the *P. pyriformis* mitochondrion protein import and maturation machinery were found in the predicted proteome: the β-barrel outer membrane translocases Tom40 and Sam50, the inner membrane translocase Tim17 and its associated protein Pam18, the α subunit of the mitochondrial processing peptidase (MPP) and the chaperone protein Cpn60. In addition to these, new components were identified: the β subunit of MPP, and Cpn10, the co-chaperone of Cpn60. The mitochondrial carrier (MC) proteins were again recovered in number of 4.

All components of the glycine cleavage system (GCS), previously shown to be localized in the MRO of *P. pyriformis* (Zubáčová et al. 2013), were confirmed: GCS-H, -L, -T, -P1, and -P2, as well as 3 paralogues of [FeFe]hydrogenase and 3 hydrogenase (HydA) maturases: HydE, HydF, HydG.

In addition to these previously reported "simple" hydrogenases, we also discovered 3 paralogues of [FeFe]hydrogenase fused to an N-terminal NADH-quinone oxidoreductase (NuoG) and a C-terminal NADPH-dependent sulfite reductase (CysJ). Similar "fused" hydrogenases have been previously reported in *Trichomonas vaginalis* (Tachezy and Doležal 2007), *Pygsuia biforma* (Stairs et al. 2014), and *Stygiella incarcerate* (Leger et al. 2016) and

were hypothesized to possibly catalyze NAD(P)-dependent formation of H_2 (Tachezy and Doležal 2007).

Hydrogenases in anaerobic protists often provide electron sink for the 1st reaction of the extended glycolysis catalyzed by pyruvate:ferredoxin oxidoreductase (PFO) and indeed, there were 5 homologues of PFO identified in the *P. pyriformis* genome.

Ferredoxin (Fd), found in 2 copies, may serve as an intermediate between PFO and HydA. In addition to Fd, 1 copy of a unique flavodoxin-ferredoxin fusion protein (Fld-Fd) was identified in the genome.

The second reaction of the extended glycolysis can be either catalyzed by a 2-enzyme system ASCT+SCS like in *Trichomonas vaginalis*, or by a single enzyme acetyl-CoA synthetase (ACS), e.g. in *Giardia intestinalis*. There were 2 paralogues of ACS found in the genome and no ASCT or SCS.

The glycolytic pathway was previously reported to include several alternative enzymes in *P. pyriformis*, however, no enzyme catalyzing the 3rd step of glycolysis was found (Stechmann et al. 2006). Here, we report identification of both the ATP-dependent phosphofructokinase (PFK) and the pyrophosphate-dependent pyrophosphate:fructose 6-phosphate phosphotransferase (PFP), both in 1 copy, in the genome, finalizing the mapping of the *P. pyriformis* glycolytic pathway.

Other identified enzymes involved in sugar metabolism are: NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPN, 3 paralogues), phosphoglucomutase (PGM, 2 paralogues), phosphomannomutase (PMM), and phosphoenolpyruvate carboxykinase (PEPCK).

Based on the predicted proteome, previous localization experiments (Zubáčová et al. 2013), and comparison with other metamonads (Leger et al. 2017), we propose a following hypothetical arrangement of the mitochondrial and energy metabolism in *P. pyriformis* (Fig. S1). Energy metabolism takes place completely in the cytosol and consists of glycolysis, employing both canonical and alternative (pyrophosphate-dependent) enzymes, and extended glycolytic steps catalyzed by PFO and ACS.

The MRO contains a complete GCS which catabolizes glycine into CO_2 and NH_3, while methylating tetrahydrofolate. The electron sink necessary for GCS function is provided by MRO-localized hydrogenase(s), cooperating with MRO-localized hydrogenase maturases. The

electron transport mechanism between GCS and hydrogenases is unclear, as no components of the mitochondrial respiratory complex 1 (NuoE, NuoF) were found. One possible candidate for the intermediate is the Fld-Fd fusion protein. Similar arrangement has been suggested for *Dysnectes brevis* (Fornicata). However, *D. brevis* has NuoE and NuoF (Leger et al. 2017). Alternatively, this process may be performed by the newly reported "fused" hydrogenases, as suggested for *T. vaginalis* (Tachezy and Doležal 2007). Further protein localization experiments, and biochemical experiments are needed to support or refute these hypotheses.

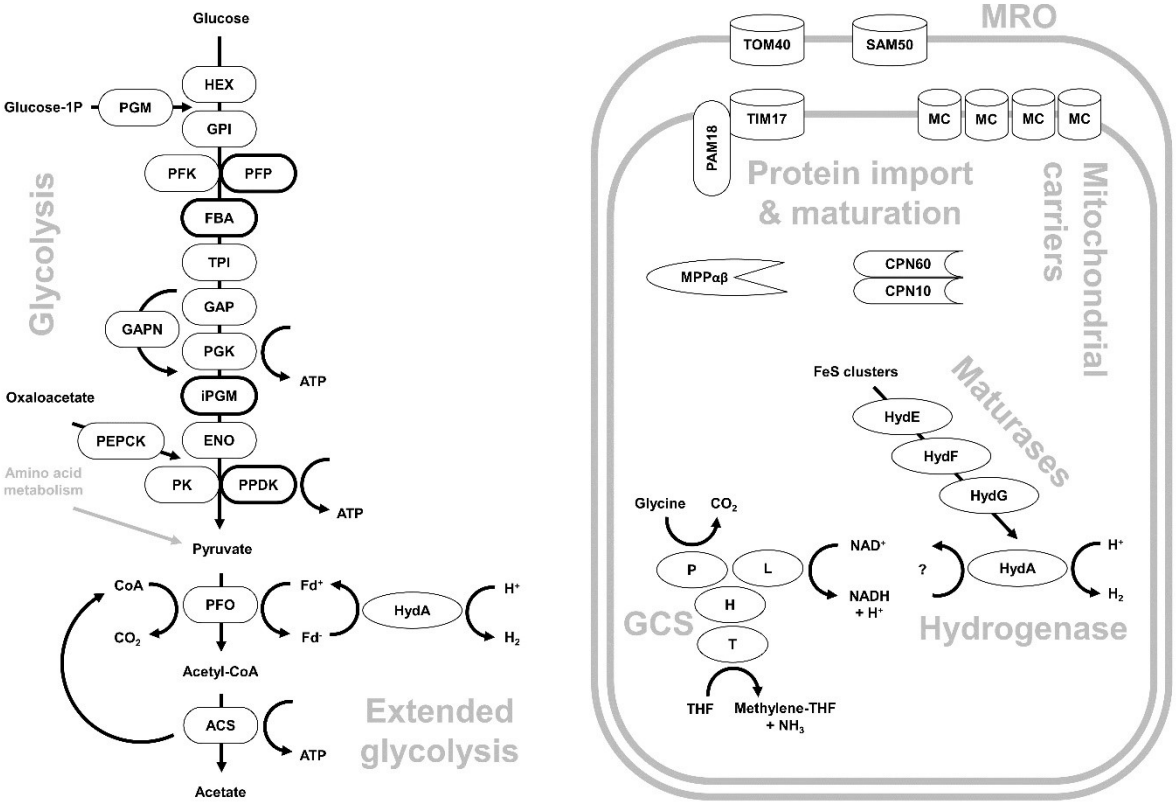


Figure S1: Hypothetical energy metabolism and MRO functions in *Paratrimastix pyriformis*. Reactions of glycolysis simplified. Bold outline indicates alternative glycolytic enzymes. For discussion of individual glycolytic enzymes and comparison with *M. exilis* see (Liapounova et al. 2006) or page 40. Question mark indicates an unknown intermediate between GCS and HydA responsible for presumed electron transfer.

1.3.3. Amino acid metabolism

We have identified 42 enzymes putatively involved in amino acid metabolism and 7 more enzymes putatively involved in folate metabolism and folate cycle, which are closely connected to the amino acid metabolism. This is higher number than in *Trichomonas vaginalis*, which has the most complex amino acid metabolism predicted among studied metamonads, containing 39 enzymes (Carlton et al. 2007).

We reconstructed a hypothetical metabolic map of the amino acid metabolism based on the metabolic maps in the KEGG database (Kanehisa et al. 2014) and catalytic activities of enzymes reported from other metamonads. The reconstructed metabolism (Fig. S2) shows a possibility of de novo biosynthesis for at least 8 protein-building amino acids: cysteine, alanine, serine, glycine, threonine, glutamate, glutamine, and selenocysteine. Methionine can be regenerated thanks to the methionine cycle. The selenocysteine biosynthesis pathway is notable, as capacity to synthesize this amino acid has been so far reported only in *Spironucleus salmonicida* among all metamonads (Xu et al. 2014).

Unlike many other metamonads (Novák et al. 2016), *P. pyriformis* is unable to utilize arginine for ATP production, as the enzyme catalyzing the first reaction of the arginine deiminase pathway – the arginine deiminase (ADI) itself is missing. Other amino acids can still be used in energy metabolism: cysteine, serine, and tryptophan can be converted to pyruvate, and methionine to α-keto-butyrate, which can both enter the extended glycolysis (Anderson and Loftus 2005).

The presence of a complete glycine cleavage system (GCS) may be connected to the presence of complete folate metabolism, folate cycle, and methionine cycle. The methyl residue liberated from glycine by the activity of GCS enters the connected folate and methionine cycles and can be later utilized in a multitude of metabolic pathways requiring one-carbon units (reviewed in Ducker and Rabinowitz 2017). The transsulfuration pathway, often associated with the folate and methionine cycles, was not found in *P. pyriformis*.

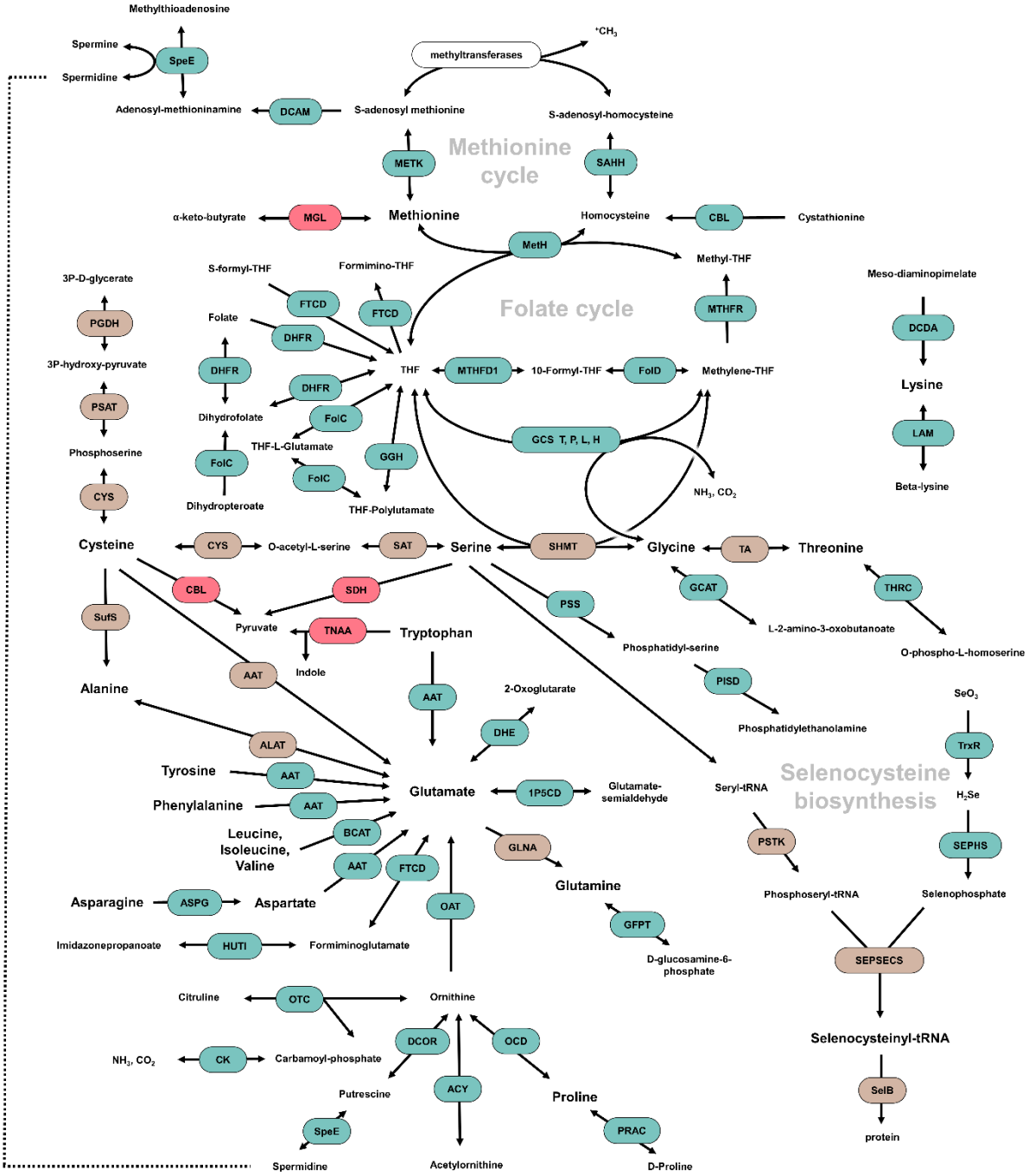


Figure S2: Hypothetical map of amino acid metabolism in *Paratrimastix pyriformis*. Brown color indicates enzymes possibly involved in amino acid biosynthesis pathways. Red color indicates enzymes possibly involved in ATP production. For comparison with *M. exilis* see (Karnkowska et al. 2019) or page 50.

1.3.4. Transporters across plasma membrane

Just like in *Trichomonas vaginalis*, the most numerous group of putative plasma membrane transporters (Tab. S1) identified in *P. pyriformis* is the ATP-binding cassette (ABC) superfamily. The multidrug resistance proteins (MRP), known for their role in export of harmful

compounds out of the cell, are represented by 9 paralogues. The ion transporting P-type ATPases have 16 paralogues and the phospholipid-transporting P-type ATPases have 3.

Another type of transporters responsible for extrusion of harmful compounds is the multi antimicrobial extrusion (MATE) family of the MOP flippase superfamily. There are 4 MATE paralogues in *P. pyriformis*.

The major facilitator superfamily (MFS) in *P. pyriformis* counts 7 uncharacterized members, 1 glycoside-pentoside-hexuronide:cation symporter (GPH), 2 sphingolipid transporters (SPNS), 1 reduced folate carrier (RFC) responsible for uptake of folate cofactors (reviewed in Matherly and Hou 2008), and interestingly no sugar porters (SP).

Amino acids can be imported by the transmembrane amino acid transporters (AAAP), represented by 7 genes in *P. pyriformis*, or the b(0,+)-type amino acid transporter (PotE) with a single paralogue.

The equilibrative nucleoside transporter (ENT) family responsible for transporting nucleosides and nucleobases is represented by 1 paralogue in *P. pyriformis*, just like the choline-transporter-like protein (CTL) which imports choline, a precursor of phospholipids.

The inorganic ions transporters in *P. pyriformis* are represented by 4 paralogues of the cation-chloride cotransporter (CCC), 2 paralogues of the K^+ transporter (TRK), 3 paralogues of the Ca^{2+}:cation antiporter (CaCA), 3 paralogues of the cation diffusion facilitator (CDF), 4 paralogues of the divalent anion:Na^+ symporter (DASS), and 2 paralogues of the sulfite exporter TauE/SafE family (TauE).

Name	**Abbrev.**	**# paralogues**
Ion transporting P-type ATPases	pATPase	16
Multidrug resistance proteins	MRP	9
Transmembrane amino acid transporter	AAAP	7
Major facilitator superfamily (uncharacterized)	MFS	7
Cation-chloride cotransporter	CCC	4
Divalent anion:Na+ symporter	DASS	4
Multi antimicrobial extrusion family	MATE	4
Ca^{2+}:cation antiporter	CaCA	3
Cation diffusion facilitator	CDF	3
Phospholipid-transporting P-type ATPases	pATPase	3

Sulfite exporter TauE/SafE family	TauE	2
K+ transporter	TRK	2
Choline-transporter-like protein	CTL	1
Equilibrative nucleoside transporter	ENT	1
Glycoside-pentoside-hexuronide:cation symporter	GPH	1
b(0,+)-type amino acid transporter	PotE	1
Reduced folate carrier	RFC	1
Sphingolipid transporter	SPNS	1

Table S1: Transporters across plasma membrane identified in *Paratrimastix pyriformis*. For comparison with *Trichomonas vaginalis* see (Dean et al. 2014) or page 36 and further.

1.4. Preliminary conclusions

Using Oxford Nanopore MinION, PacBio, and Illumina MiSeq and HiSeq sequencing technologies, we have produced a good quality, 56.7 mega base pairs long genome assembly of *Paratrimastix pyriformis* (Preaxostyla) and predicted 13,533 protein-coding genes. We have manually reconstructed the putative proteome of the mitochondrion-related organelle, which we hypothesize to functionally resemble the organelle of *Dysnectes brevis* (Fornicata) in the absence of ATP generation and presence of the glycine cleavage system and the hydrogen-producing machinery. The energy metabolism, consisting of glycolysis and its extension via PFO and ACS enzymes, is predicted to be localized in the cytosol. The reconstructed hypothetical amino acid metabolism is more complex than in *Trichomonas vaginalis* (Parabasalia) in number of enzymes as well as the ability to synthesize selenocysteine and provide one-carbon units to the cell metabolism via the glycine cleavage system and SHMT. On the other hand, *P. pyriformis* is unable to generate energy via the arginine deiminase pathway. Complement of putative plasma membrane transporters is less numerous than in *Trichomonas vaginalis*, which may reflect differences between the parasitic versus free-living lifestyles. Results of these and further analyses will be later compared to analogous results generated for 4 other Preaxostyla species in order to map the emergence and loss of various cellular systems and functions during the evolutionary history of this group of anaerobic protists.

1.5. References

Anderson, Iain J., and Brendan J. Loftus. 2005. "Entamoeba Histolytica: Observations on Metabolism Based on the Genome Sequence." *Experimental Parasitology* 110 (3 SPEC. ISS.): 173–77. https://doi.org/10.1016/j.exppara.2005.03.010.

Carlton, Jane M, Robert P Hirt, Joana C Silva, Arthur L Delcher, Michael Schatz, Qi Zhao, Jennifer R Wortman, et al. 2007. "Draft Genome Sequence of the Sexually Transmitted Pathogen Trichomonas Vaginalis." *Science (New York, N.Y.)* 315 (5809): 207–12. https://doi.org/10.1126/science.1132894.

Chomczynski, P., and K. Mackey. 1995. "Modification of the TRI Reagent(TM) Procedure for Isolation of RNA from Polysaccharide- and Proteoglycan-Rich Sources." *BioTechniques* 19 (6): 942–45. http://www.ncbi.nlm.nih.gov/pubmed/8747660.

Dean, Paul, Peter Major, Sirintra Nakjang, Robert P Hirt, and T Martin Embley. 2014. "Transport Proteins of Parasitic Protists and Their Role in Nutrient Salvage." *Frontiers in Plant Science* 5 (January): 153. https://doi.org/10.3389/fpls.2014.00153.

Ducker, Gregory S, and Joshua D Rabinowitz. 2017. "Cell Metabolism Review One-Carbon Metabolism in Health and Disease." *Cell Metabolism* 25 (1): 27–42. https://doi.org/10.1016/j.cmet.2016.08.009.

Finn, Robert D., Jody Clements, and Sean R. Eddy. 2011. "HMMER Web Server: Interactive Sequence Similarity Searching." *Nucleic Acids Research* 39 (SUPPL. 2): W29-37. https://doi.org/10.1093/nar/gkr367.

Haas, Brian J., Arthur L. Delcher, Stephen M. Mount, Jennifer R. Wortman, Roger K. Smith, Linda I. Hannick, Rama Maiti, et al. 2003. "Improving the Arabidopsis Genome Annotation Using Maximal Transcript Alignment Assemblies." *Nucleic Acids Research* 31 (19): 5654–66. https://doi.org/10.1093/nar/gkg770.

Haas, Brian J, Steven L Salzberg, Wei Zhu, Mihaela Pertea, Jonathan E Allen, Joshua Orvis, Owen White, C. Robin Robin, and Jennifer R Wortman. 2008. "Automated Eukaryotic Gene Structure Annotation Using EVidenceModeler and the Program to Assemble Spliced Alignments." *Genome Biology* 9 (1): R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Haddad, Isam, Karsten Hiller, Eliane Frimmersdorf, Beatrice Benkert, Dietmar Schomburg, and Dieter Jahn. 2009. "An Emergent Self-Organizing Map Based Analysis Pipeline for

Comparative Metabolome Studies." *In Silico Biology* 9 (4): 163–78. http://www.ncbi.nlm.nih.gov/pubmed/20109147.

Hampl, Vladimír. 2017. "Preaxostyla." In *Handbook of the Protists*, 1139–74. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-28149-0_8.

Hampl, Vladimír, Ivan Čepička, and Marek Eliáš. 2019. "Was the Mitochondrion Necessary to Start Eukaryogenesis?" *Trends in Microbiology* 27 (2): 96–104. https://doi.org/10.1016/j.tim.2018.10.005.

Hampl, Vladimír, Jeffrey D Silberman, Alexandra Stechmann, Sara Diaz-Triviño, Patricia J Johnson, and Andrew J Roger. 2008. "Genetic Evidence for a Mitochondriate Ancestry in the 'Amitochondriate' Flagellate Trimastix Pyriformis." *PLoS ONE* 3 (1): 9. https://doi.org/10.1371/journal.pone.0001383.

Kanehisa, Minoru, Susumu Goto, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. 2014. "Data, Information, Knowledge and Principle: Back to Metabolism in KEGG." *Nucleic Acids Research* 42 (D1): D199-205. https://doi.org/10.1093/nar/gkt1076.

Karnkowska, Anna, Sebastian C Treitli, Ondřej Brzoň, Lukáš Novák, Vojtěch Vacek, Petr Soukal, Lael D Barlow, et al. 2019. "The Oxymonad Genome Displays Canonical Eukaryotic Complexity in the Absence of a Mitochondrion." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msz147.

Karnkowska, Anna, Vojtěch Vacek, Zuzana Zubáčová, Sebastian C. Treitli, Romana Petrželková, Laura Eme, Lukáš Novák, et al. 2016. "A Eukaryote without a Mitochondrial Organelle." *Current Biology* 26 (10): 1274–84. https://doi.org/10.1016/j.cub.2016.03.053.

Kim, Daehwan, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L Salzberg. 2013. "TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions." *Genome Biology* 14 (4): R36. https://doi.org/10.1186/gb-2013-14-4-r36.

Koren, Sergey, Brian P Walenz, Konstantin Berlin, Jason R Miller, Nicholas H Bergman, and Adam M Phillippy. 2017. "Canu: Scalable and Accurate Long-Read Assembly via Adaptive k-Mer Weighting and Repeat Separation." *Genome Research* 27 (5): 722–36. https://doi.org/10.1101/gr.215087.116.

Krogh, A, B Larsson, G von Heijne, and E L Sonnhammer. 2001. "Predicting Transmembrane

Protein Topology with a Hidden Markov Model: Application to Complete Genomes." *Journal of Molecular Biology* 305 (3): 567–80. https://doi.org/10.1006/jmbi.2000.4315.

Langmead, Ben, and Steven L Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Leger, Michelle M., Martin Kolisko, Ryoma Kamikawa, Courtney W. Stairs, Keitaro Kume, Ivan Čepička, Jeffrey D. Silberman, et al. 2017. "Organelles That Illuminate the Origins of Trichomonas Hydrogenosomes and Giardia Mitosomes." *Nature Ecology and Evolution* 1 (4): 0092. https://doi.org/10.1038/s41559-017-0092.

Leger, Michelle M, Laura Eme, Laura A Hug, and Andrew J Roger. 2016. "Novel Hydrogenosomes in the Microaerophilic Jakobid Stygiella Incarcerata." *Molecular Biology and Evolution* 33 (9): 2318–36. https://doi.org/10.1093/molbev/msw103.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform." *Bioinformatics* 25 (14): 1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Liapounova, Natalia A, Vladimir Hampl, Paul M K Gordon, Christoph W Sensen, Lashitew Gedamu, and Joel B Dacks. 2006. "Reconstructing the Mosaic Glycolytic Pathway of the Anaerobic Eukaryote Monocercomonoides." *Eukaryotic Cell* 5 (12): 2138–46. https://doi.org/10.1128/EC.00258-06.

Loman, Nicholas J, Joshua Quick, and Jared T Simpson. 2015. "A Complete Bacterial Genome Assembled de Novo Using Only Nanopore Sequencing Data." *Nature Methods* 12 (8): 733–35. https://doi.org/10.1038/nmeth.3444.

Matherly, Larry H., and Zhanjun Hou. 2008. "Chapter 5 Structure and Function of the Reduced Folate Carrier. A Paradigm of a Major Facilitator Superfamily Mammalian Nutrient Transporter." *Vitamins and Hormones* 79: 145–84. https://doi.org/10.1016/S0083-6729(08)00405-6.

Moriya, Yuki, Masumi Itoh, Shujiro Okuda, Akiyasu C. Yoshizawa, and Minoru Kanehisa. 2007. "KAAS: An Automatic Genome Annotation and Pathway Reconstruction Server." *Nucleic Acids Research* 35 (SUPPL.2): W182–85. https://doi.org/10.1093/nar/gkm321.

Novák, Lukáš, Zuzana Zubáčová, Anna Karnkowska, Martin Kolisko, Miluše Hroudová, Courtney W. Stairs, Alastair G.B. Simpson, et al. 2016. "Arginine Deiminase Pathway

Enzymes: Evolutionary History in Metamonads and Other Eukaryotes." *BMC Evolutionary Biology*. https://doi.org/10.1186/s12862-016-0771-4.

Robinson, James T, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S Lander, Gad Getz, and Jill P Mesirov. 2011. "Integrative Genomics Viewer." *Nature Biotechnology* 29 (1): 24–26. https://doi.org/10.1038/nbt.1754.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs." *Bioinformatics* 31 (19): 3210–12. https://doi.org/10.1093/bioinformatics/btv351.

Song, Li, Dhruv S. Shankar, and Liliana Florea. 2016. "Rascaf: Improving Genome Assembly with RNA Sequencing Data." *The Plant Genome* 9 (3): 0. https://doi.org/10.3835/plantgenome2016.03.0027.

Stairs, Courtney W, Laura Eme, Matthew W Brown, Cornelis Mutsaers, Edward Susko, Graham Dellaire, Darren M Soanes, Mark van der Giezen, and Andrew J Roger. 2014. "A SUF Fe-S Cluster Biogenesis System in the Mitochondrion-Related Organelles of the Anaerobic Protist Pygsuia." *Current Biology : CB* 24 (11): 1176–86. https://doi.org/10.1016/j.cub.2014.04.033.

Stanke, Mario, Rasmus Steinkamp, Stephan Waack, and Burkhard Morgenstern. 2004. "AUGUSTUS: A Web Server for Gene Finding in Eukaryotes." *Nucleic Acids Research* 32 (Web Server issue): W309. https://doi.org/10.1093/NAR/GKH379.

Stechmann, Alexandra, Manuela Baumgartner, Jeffrey D Silberman, and Andrew J Roger. 2006. "The Glycolytic Pathway of Trimastix Pyriformis Is an Evolutionary Mosaic." *BMC Evolutionary Biology* 6 (January): 101. https://doi.org/10.1186/1471-2148-6-101.

Tachezy, Jan, and Pavel Doležal. 2007. "Iron–Sulfur Proteins and Iron–Sulfur Cluster Assembly in Organisms with Hydrogenosomes and Mitosomes." In *Origin of Mitochondria and Hydrogenosomes*, 105–33. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-38502-8_6.

Tarailo-Graovac, Maja, and Nansheng Chen. 2009. "Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences." In *Current Protocols in Bioinformatics*, 25:4.10.1-4.10.14. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471250953.bi0410s25.

Treitli, Sebastian C., Martin Kolisko, Filip Husník, Patrick J. Keeling, and Vladimír Hampl. 2019. " Revealing the Metabolic Capacity of Streblomastix Strix and Its Bacterial Symbionts Using Single-Cell Metagenomics ." *Proceedings of the National Academy of Sciences* 116 (39): 19675–84. https://doi.org/10.1073/pnas.1910793116.

Treitli, Sebastian C., Michael Kotyk, Naoji Yubuki, Eliška Jirounková, Jitka Vlasáková, Pavla Smejkalová, Petr Šípek, Ivan Čepička, and Vladimír Hampl. 2018. "Molecular and Morphological Diversity of the Oxymonad Genera Monocercomonoides and Blattamonas Gen. Nov." *Protist* 169 (5): 744–83. https://doi.org/10.1016/j.protis.2018.06.005.

Walker, Bruce J., Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A. Cuomo, et al. 2014. "Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement." Edited by Junwen Wang. *PLoS ONE* 9 (11): e112963. https://doi.org/10.1371/journal.pone.0112963.

Xu, Feifei, Jon Jerlström-Hultqvist, Elin Einarsson, Ásgeir Ástvaldsson, Staffan G. Svärd, and Jan O. Andersson. 2014. "The Genome of Spironucleus Salmonicida Highlights a Fish Pathogen Adapted to Fluctuating Environments." *PLoS Genetics* 10 (2): e1004053. https://doi.org/10.1371/journal.pgen.1004053.

Zdobnov, E. M., and R. Apweiler. 2001. "InterProScan - An Integration Platform for the Signature-Recognition Methods in InterPro." *Bioinformatics* 17 (9): 847–48. https://doi.org/10.1093/bioinformatics/17.9.847.

Zubáčová, Zuzana, Lukáš Novák, Jitka Bublíková, Vojtěch Vacek, Jan Fousek, Jakub Rídl, Jan Tachezy, Pavel Doležal, Čestmír Vlček, and Vladimír Hampl. 2013. "The Mitochondrion-Like Organelle of Trimastix Pyriformis Contains the Complete Glycine Cleavage System." *PLoS ONE* 8 (3): e55417. https://doi.org/10.1371/journal.pone.0055417.

2. Supplement II: Outreach

Author of this thesis has written and published 2 popular-science articles covering 2 of the scientific papers included in this thesis during his Ph.D. study.

The article "Living without mitochondria: the downfall of one textbook truth" covers the paper "A Eukaryote without a Mitochondrial Organelle" and was published in 2016 on the popular-science website TheScienceBreaker. It can be accessed here:

https://thesciencebreaker.org/breaks/evolution-behaviour/living-without-mitochondria-the-downfall-of-one-textbook-truth.

The article "Fool's gold inside us, microbes, and jumping genes" covers the paper "Fe S Cluster Assembly in Oxymonads and Related Protists" and was published in 2018 on the website of the International society for evolutionary protistology. It can be accessed here:

https://www.isep-protists.com/blog/fool-s-gold-inside-us-microbes-and-jumping-genes-guest-post-by-lukas-novak.

2.1. Living without mitochondria: the downfall of one textbook truth

It was the greatest leap in evolution since the emergence of life on Earth. So-called eukaryotic cells, the building blocks of all multicellular organisms like you and me, animals, plants, fungi, and also a whole zoo of single-celled protists, evolved from a common ancestor more than a billion years ago. This ancestor resembled current-day prokaryotes, i.e. bacteria and archaea. These organisms populated all imaginable habitats and developed a plethora of wildly diverse means of obtaining their energy. However, the simple structure of their cells limited their size and complexity of interactions. With time, our ancestor developed a series of innovations that are nowadays defining characteristics of the eukaryotic cell. How this process went through, what was its driving force, and in what order those particular steps took place still remains a question provoking lively discussions among scientists. However, regardless of the details, among the most crucial steps in forming a eukaryote was undoubtedly the acquisition of a mitochondrion.

Today, mitochondria are small organelles within our cells that form complex and ever-changing networks. The sausage-shaped mitochondria constantly split, merge, and move around in order to provide all cellular regions with a life-giving chemical energy. This energy, produced on the inner surface of mitochondria, is produced by reactions between the nutrients we get from food

and the oxygen we breathe in. Cells can produce energy also without the help of mitochondria and many organisms which live in oxygen-poor environment indeed do so, but the mitochondrial energy production is more than ten times more efficient. Mitochondria have also many other roles including production of various compounds, which will be important later in our story.

Once an independent bacterium, the ancestor of mitochondria somehow found itself inside another cell, survived against all odds, and developed a relationship of interdependence with its new host. Importantly, this interaction added a new trick to the repertoire of our ancestors: the above-mentioned unprecedentedly efficient way of producing energy which enabled an explosion of new, more complex and diverse life forms. To better understand this evolutionary process, scientists were searching for a missing link: a eukaryote lacking mitochondria. The search came up empty. All hot candidates, mostly parasitic protists, were shown to possess a simplified mitochondrion. By studying their mitochondria, researchers found out that no matter how small and simplified the organelle is, it always retains one function: the synthesis of iron-sulfur clusters. These inorganic compounds are essential components of many vital proteins and no cell can live without them. More and more examples accumulated over time supporting two textbook truths: first, there are no ancient amitochondriate eukaryotes living today and second, eukaryotes can never completely lose mitochondria, particularly because they need the iron-sulfur clusters that mitochondria produce.

This is where our work comes in. Among all those false alarms of seemingly amitochondriate protists, one candidate group remained unexplored: Oxymonads. These anaerobic protists are largely neglected as they cause no health issues and have no economic potential whatsoever. Most oxymonads live in the gut of termites, while others, like our research subject *Monocercomonoides*, inhabit a wider range of hosts - we isolated our specimen from feces of a chinchilla. To date, nothing in the oxymonad cell has been found that resembles a mitochondrion although it is clear that their ancestors must have had one because it is present in their closest free-living relatives. We decided to investigate this case further. By using various experimental methods, we searched for traces of mitochondria and failed. Not only the physical organelle seemed to be missing but also all the proteins that are usually associated with it.

To be absolutely sure, we sequenced all the genetic material of *Monocercomonoides* in a futile search for a trace of mitochondria. We found everything we would expect from such an organism, but all the mitochondrion-associated genes were missing, including the genes for

producing the iron-sulfur clusters. Could *Monocercomonoides* survive without these molecules? Not so fast! This creature had a surprise for us: a completely different kind of iron-sulfur cluster production system, apparently recently borrowed from bacteria by a process called horizontal gene transfer. So *Monocercomonoides* freed themselves from their dependence on mitochondria by finding a new way to produce the essential iron sulfur clusters on their own. This is the first report of a eukaryote that completely lost its mitochondria.

2.2. Fool's gold inside us, microbes, and jumping genes

The mineral pyrite, iron sulfide gemstone also known as fool's gold for its gold-like appearance, used to be a favorite material for alchemists in their futile struggles to create precious metals. Since antiquity it is also combined with silver in the so-called marcasite jewelry, popular especially in the 19th century, and until today pyrite is used for production of sulfur dioxide for various industrial applications. Whether you are alchemist, Victorian lady, chemical engineer, or anybody else for that matter, you always have a bit of pyrite with you, or rather inside you. The iron-sulfur (FeS) clusters are molecules composed of iron and sulfur atoms arranged in a pattern resembling pyrite crystal structure, which function as cofactors – small molecular "plug-ins" – in a vital group of proteins involved in such important tasks as electron transfer or DNA repair. FeS clusters are found in all living cells and are indispensable for life as we know it. Their ubiquity and importance even led to formulation of a hypothesis saying that pyrite and similar minerals played a crucial role in the very origin of life.

In eukaryotic cells – building blocks of animals, plants, fungi, and microbial protists – FeS clusters are typically produced by two different metabolic pathways. One set of enzymes (ISC) works in mitochondria. The other (CIA), localized in cytoplasm, provides FeS clusters to all the other parts of the cell including the nucleus. The cytoplasmic pathway doesn't work on its own, but depends on a, yet unidentified, product of the mitochondrial one. Mitochondria are therefore usually essential for the cell and even their most simplified forms tend to have at least this function intact. And so, production of FeS clusters became a central question for our team after we described the first known eukaryote completely devoid of mitochondria, a chinchilla gut-inhabiting protist *Monocercomonoides*. How can the cytoplasmic pathway build FeS clusters if mitochondria, together with their pathway, were lost? It turns out the answer is lateral gene transfer – sharing of genetic material between unrelated organisms. The ancestors of *Monocercomonoides* gained another pathway for FeS cluster synthesis, called SUF, from bacteria and recruited it for work instead of the lost mitochondrial one.

We investigated the evolutionary history of this gene transfer. When did it happen? What other organisms share it? And how did the bacterial SUF pathway change in its new home? We sampled a broad diversity of *Monocercomonoides*' relatives constituting a group called Preaxostyla. All of them are single-celled microbes which shun oxygen just like *Monocercomonoides*, but that's where the similarity ends. For example, *Streblomastix* resembles a bundle of snuggly packed symbiotic bacteria, only held together by a network of thin lobes – the actual protist cell. Another one, *Saccinobaculus*, looks like a bag with a snake inside, which constantly wriggles around. The "snake" is in fact a structure of cellular skeleton that helps the cell to move. Preaxostyla are wonderfully weird creatures indeed! We sequenced 10 species, chosen to cover all major lineages, and found the genes for the SUF pathway in each of them, even those species which, unlike *Monocercomonoides*, still retain mitochondria. Also, none of the organisms harbored the mitochondrial ISC pathway.

All the 5 genes constituting the SUF pathway in Preaxostyla show the same evolutionary history. That means they must have come in one gene transfer event from bacteria, which happened before all the species split – more than 100 million years in the past. This happened before the mitochondrion vanished, possibly representing the final nail in its coffin. When the mitochondrial pathway was replaced with a new substitute, the microbes simply lost any remaining motivation for keeping the costly organelle and got rid of it. We also found out that 3 of the 5 genes are fused together in Preaxostyla, likely producing a large chimeric protein, a situation not observed in bacteria.

Our findings are most interesting from the evolutionary point of view. They show another strong evidence of lateral gene transfer having a dramatic effect on eukaryotes, a notion which is still controversial. However, they might also have a broader impact in the future. The interesting fusion of 3 genes may indicate that the protein products of these particular genes may be closely cooperating, providing a hint on the general functioning of the SUF pathway. Also, remember that still mysterious connection between mitochondrial and cytoplasmic pathways for FeS cluster production in most eukaryotes including humans? Well, now we have a system where both pathways are in cytoplasm and no mitochondrion is involved. Cross-examination of these different arrangements might help us finally identify the elusive link. Maybe one day, microscopic inhabitants of animal guts will help us uncover the secrets of the gems inside us.