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.

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. 1a and b) (Treitli et al. 2018). Like all oxymonads,
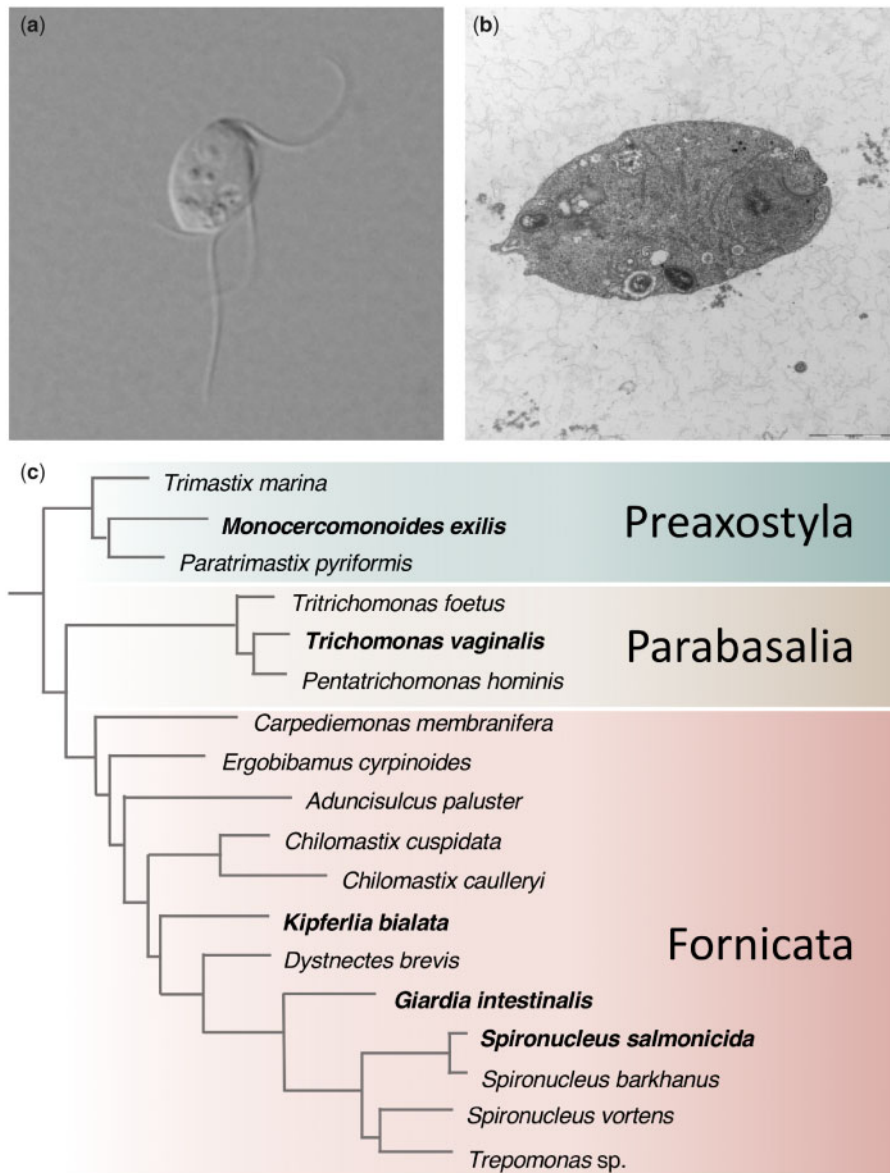
Open Access

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

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

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).

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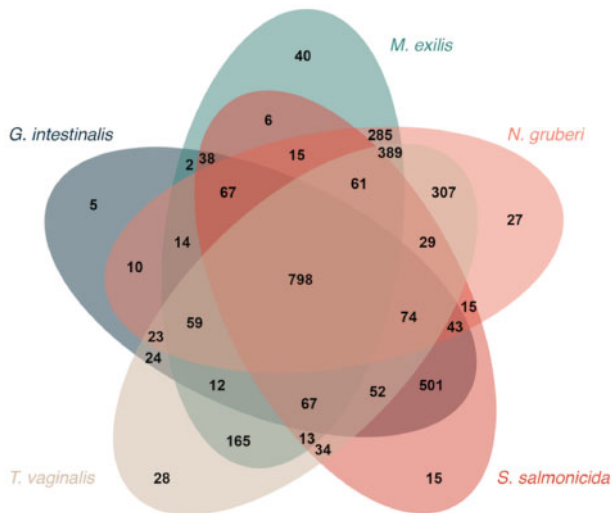
FIG. 2. Venn diagram of orthologous clusters shared and unique to *Monocercomonoides exilis*, other metamonads, and *Naegleria gruberi*.

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. 3a). This accounts for more than 30% of the *M. exilis* introns analyzed, a proportion similar to other eukaryotes (fig. 3b). The other two root positions give similar estimates (supplementary fig. S3a 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. 3a). 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. 3a and supplementary fig. S3a 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
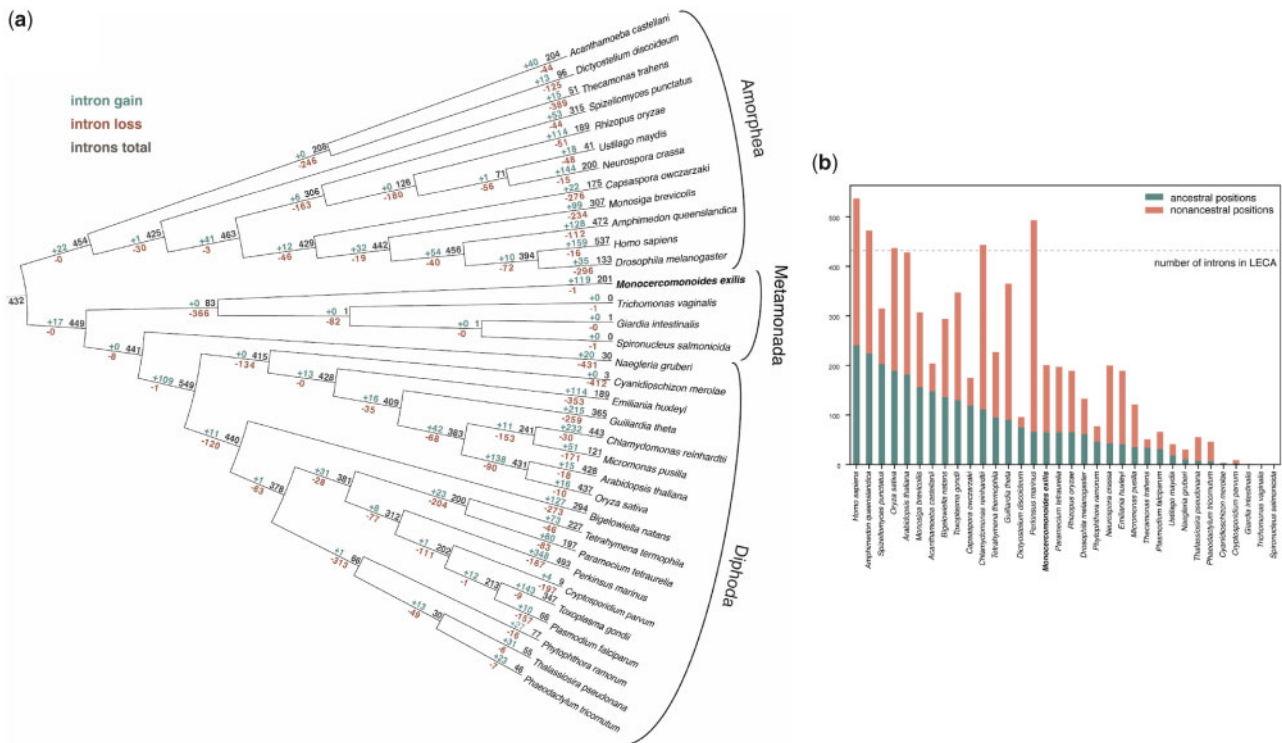
FIG. 3. Intron gains and losses along the eukaryote phylogeny. (*a*) Intron gains and losses along the eukaryote phylogeny as reconstructed by Dollo parsimony. The numbers are derived from an analysis of 3,546 intron positions in a reference set of 100 groups of orthologous genes of 34 phylogenetically 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 multi-vesicular 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
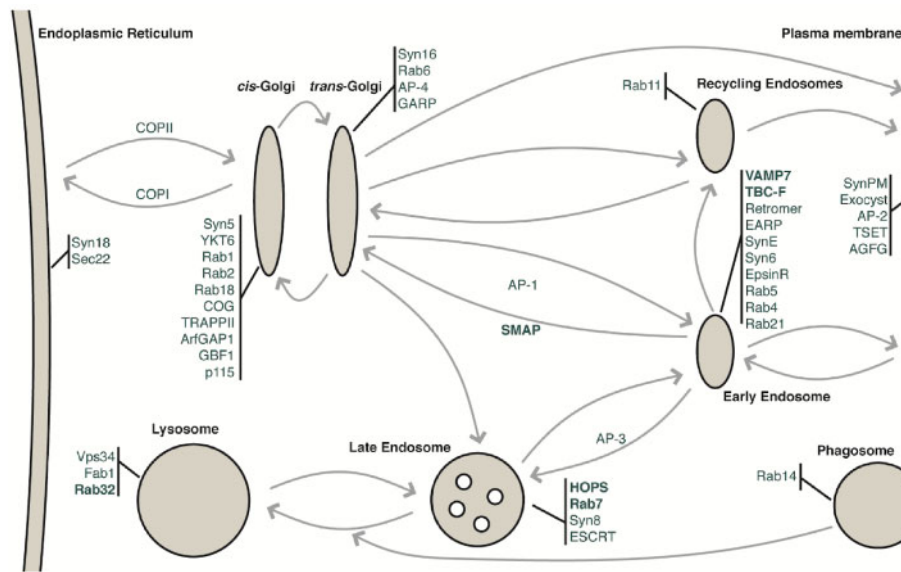
Fig. 4. Results of bioinformatic analysis of membrane-trafficking machinery in *Monocercomonoides exilis*. The presence of membrane-trafficking compartments and pathways is hypothesized as shown, based on the complement of trafficking machinery identified, and the function of their homologs in model 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. 6a 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. 6a). 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. 6a, 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,
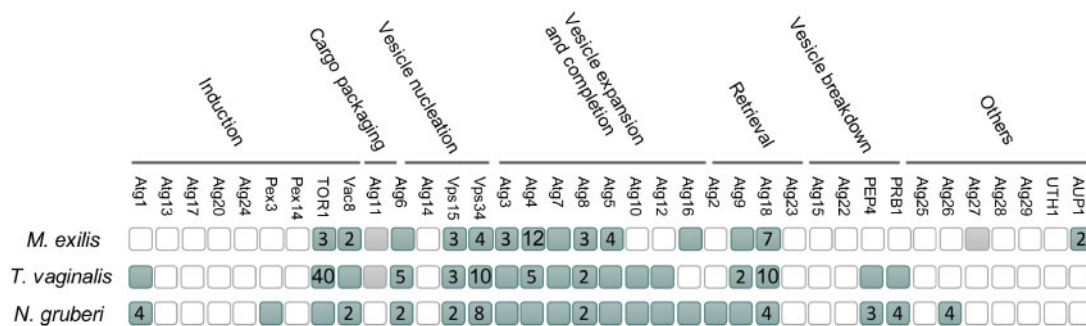
Induction					Cargo packaging				Vesicle nucleation					Vesicle expansion and completion								Retrieval				Vesicle breakdown				Others						
Atg1	Atg13	Atg17	Atg20	Atg24	Pex3	Pex14	TOR1	Vac8	Atg11	Atg6	Atg14	Vps15	Vps34	Atg3	Atg4	Atg7	Atg8	Atg5	Atg10	Atg12	Atg16	Atg2	Atg9	Atg18	Atg23	Atg15	Atg22	PEP4	PRB1	Atg25	Atg26	Atg27	Atg28	Atg29	UTH1	AUP1
M. exilis							3	2		■	3	4	3	12		3	4							7												2
T. vaginalis	■						40		■	5		3	10	5	■	2	■	■	■			2	10				■									
N. gruberi	4					■		2		2		2	8	■	■	2	■	■	■	■	■	4			3	4		4									

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
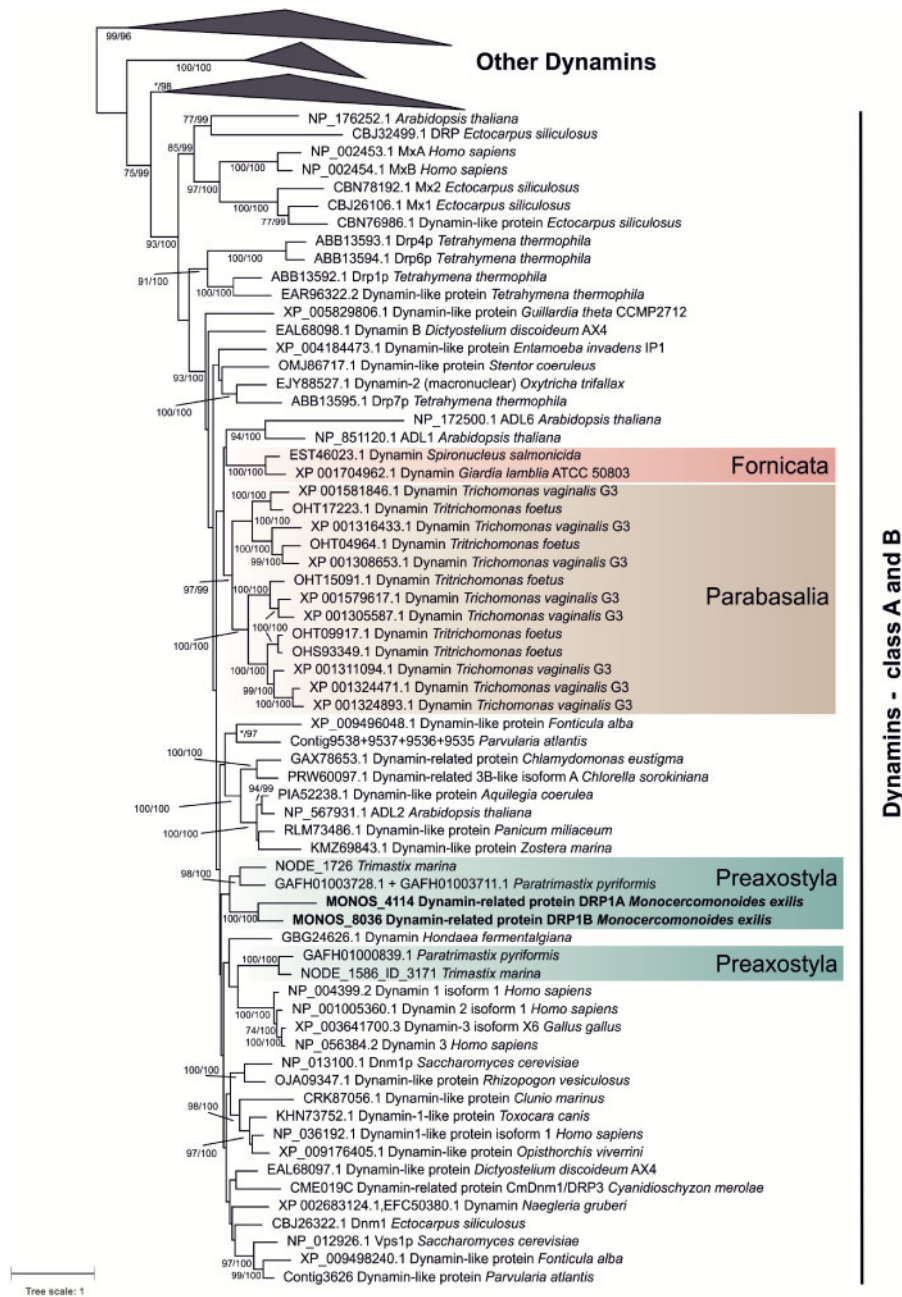
Fig. 8. Phylogenetic analysis of dynamin family showing the position of metamonad dynamins. Clades of Opa1, Mgm1, and Dynamin class C (labeled Other Dynamins) are collapsed since they do not include any metamonad dynamins. Topology is based on phylogenetic tree computed by ML method in RAxML version 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$_2$O) 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. 3a). The proportion of ancestral introns to the total number of introns was then plotted for each species (fig. 3b). 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. 3a); between Amorphea + Metamonada and the remaining eukaryotes included in the analysis (supplementary fig. S4a, Supplementary Material online); and between Metamonada and the remaining eukaryotes included in the analysis (supplementary fig. S4b, 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

Acknowledgments

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, Eliáš 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 posttranslationally 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.