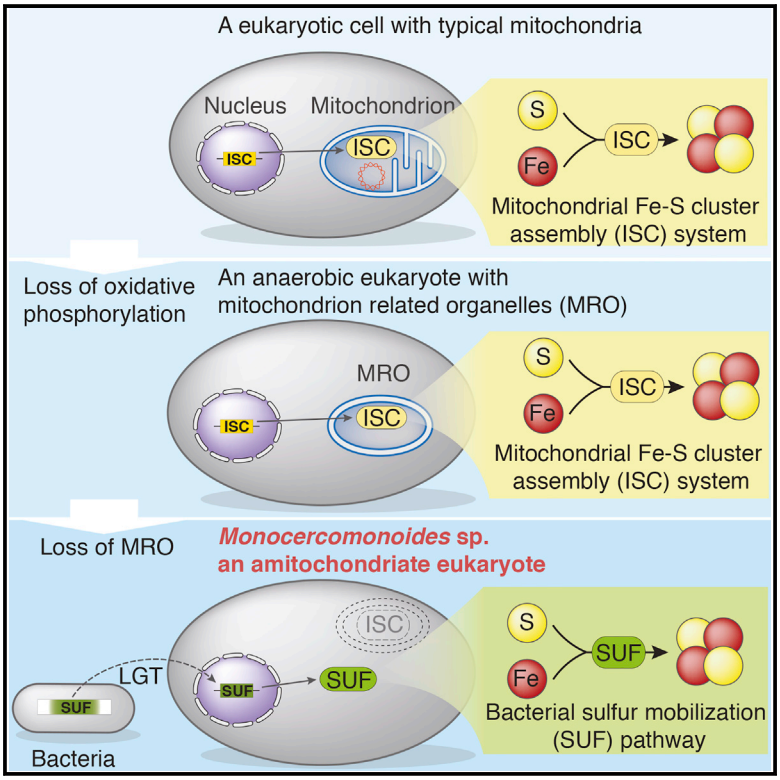
Article

Current Biology

# A Eukaryote without a Mitochondrial Organelle

## Graphical Abstract




## Highlights

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

## Authors

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

## Correspondence

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

## In Brief

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

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

CellPress

Current Biology
Article

# A Eukaryote without a Mitochondrial Organelle

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

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

## SUMMARY

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

## INTRODUCTION

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

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

## RESULTS AND DISCUSSION

### Genome Characteristics

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

**Table 1. Overview of Metamonada Genomes**

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

See also Tables S1 and S3.

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

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

## Absence of Mitochondrial Proteins

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

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

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

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

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

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


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

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

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

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

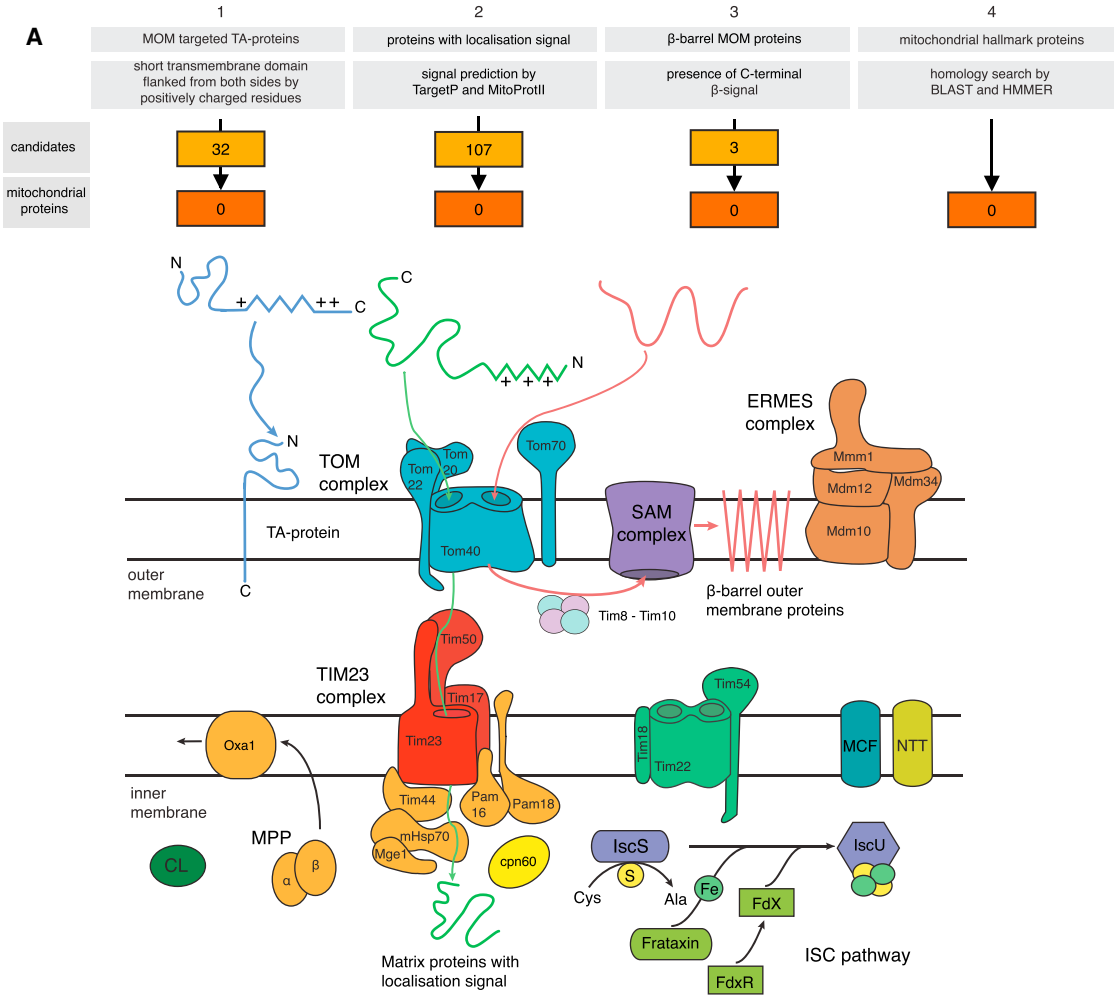



**Figure 2. Search Strategies for Proteins Functionally Related to the Mitochondrion in *Monocercomonoides***

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

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

See also Tables S4, S5, and S6.

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

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

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

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

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

## Energy Metabolism without a Mitochondrion

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

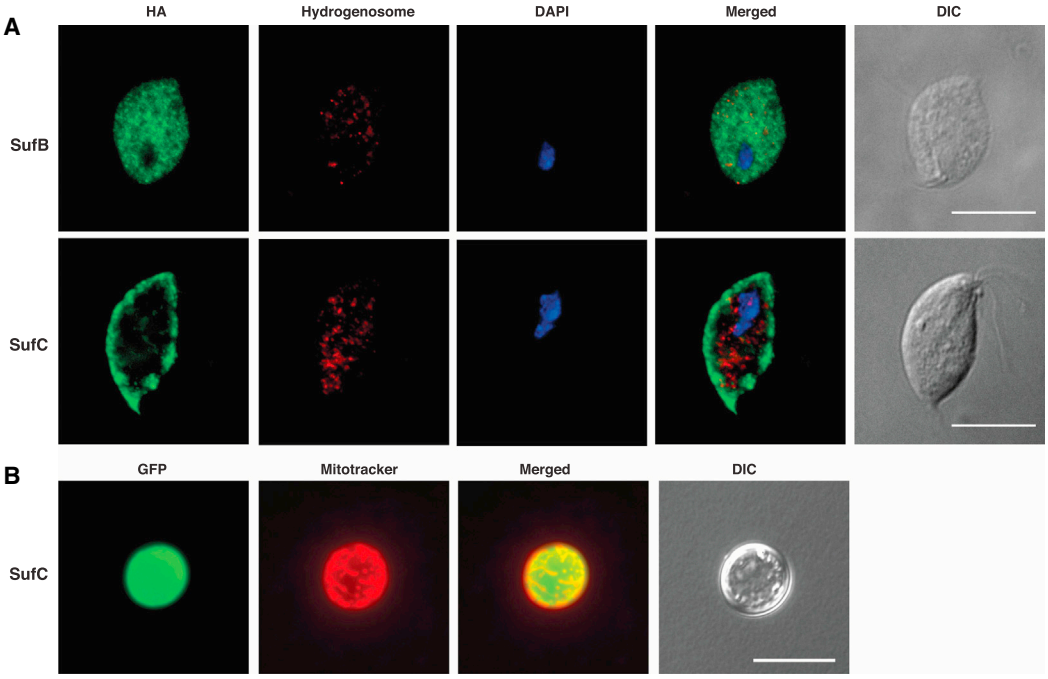
In addition to extended glycolysis, *Monocercomonoides* sp. contains a complete set of three genes for enzymes involved in arginine deiminase pathway—arginine deiminase, ornithine carbamoyltransferase, and carbamate kinase. This pathway may also be used for ATP production by arginine degradation as in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

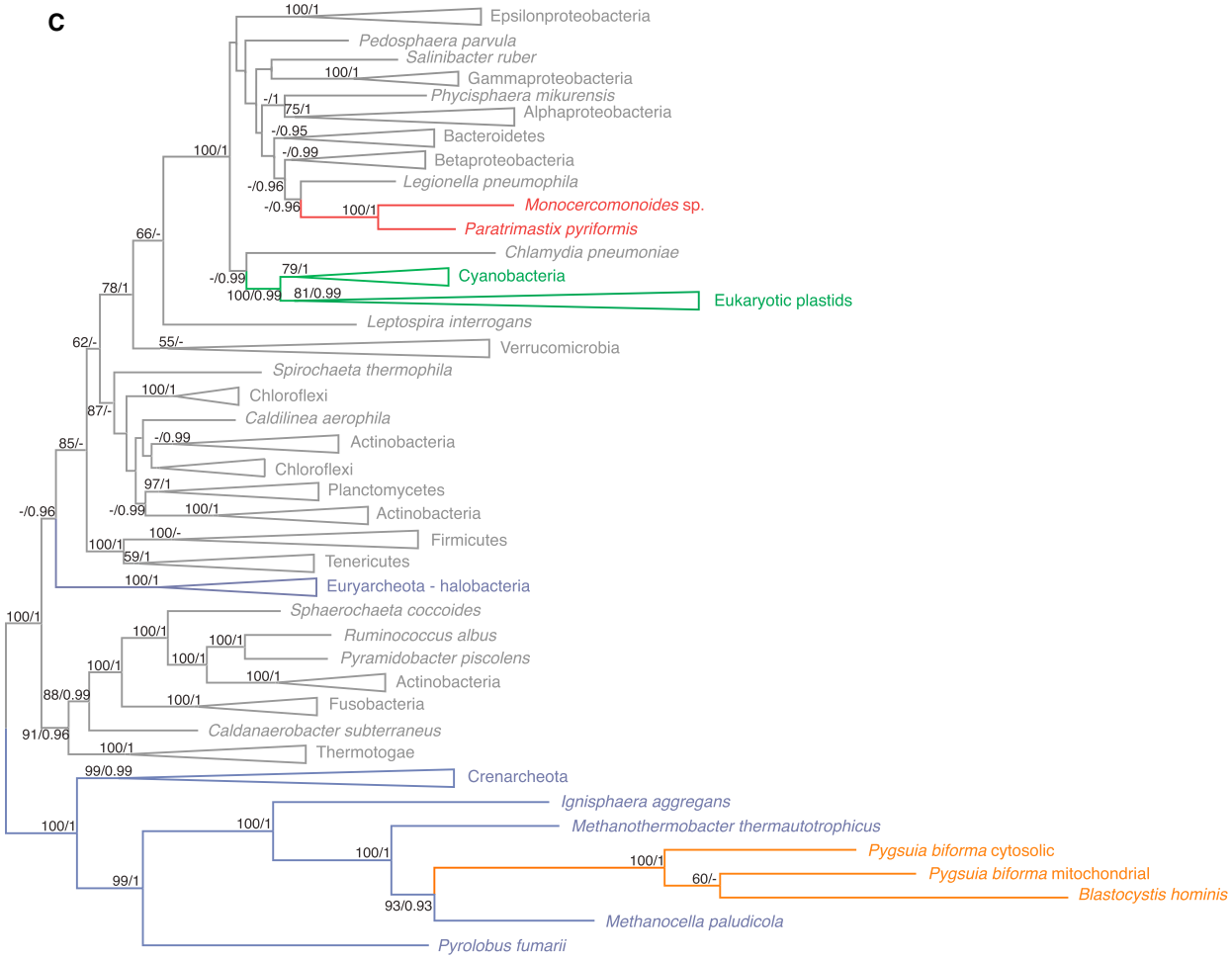
## Fe-S Cluster Assembly without a Mitochondrion

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

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

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




*(legend on next page)*

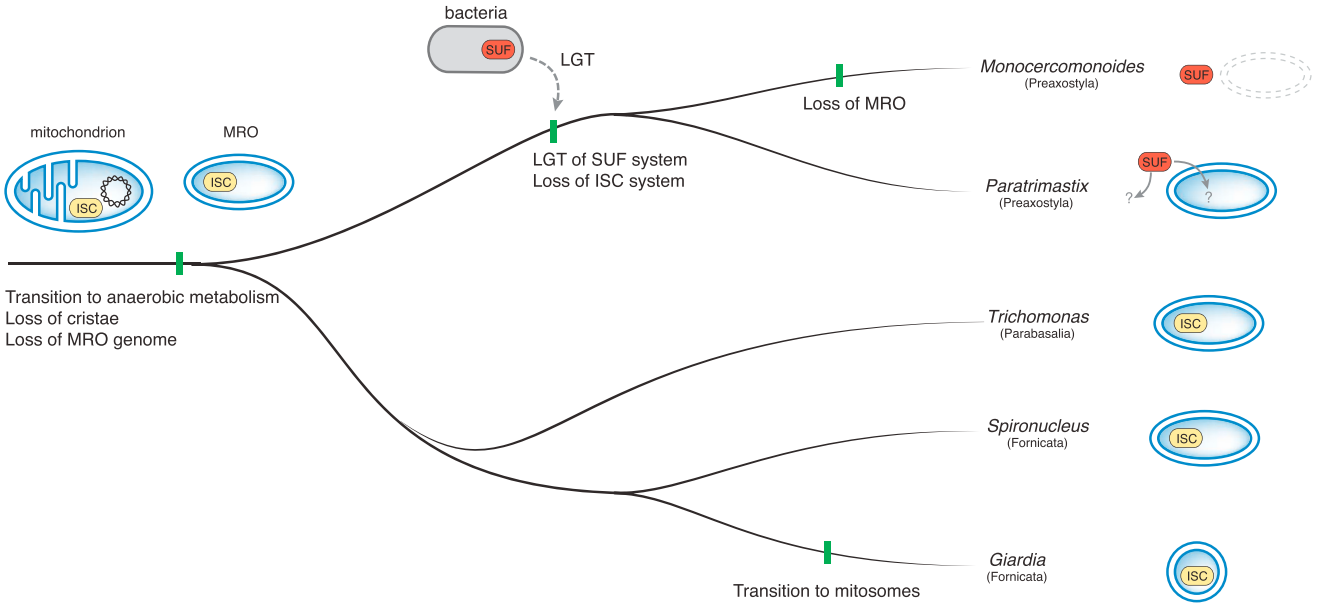



**Figure 4. Reductive Evolution of Mitochondria in Metamonads**

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

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

## Conclusions

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

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

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

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

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

See also Figures S1–S3.

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

## EXPERIMENTAL PROCEDURES

### Genome and Transcriptome Sequencing

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

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

### Genome Annotation and Gene Searching

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

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

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

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

### Phylogenetic Analyses

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

### Subcellular Localization Prediction

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

### Mitochondrial Protein Searching Using a Mitominer-Based Database

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

### FISH

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

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

### Heterologous Protein Expression and Microscopy in *Trichomonas vaginalis*

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

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

### *Saccharomyces cerevisiae* Heterologous Expression System

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

## ACCESSION NUMBERS

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

## SUPPLEMENTAL INFORMATION

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

## AUTHOR CONTRIBUTIONS

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

## ACKNOWLEDGMENTS

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

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

## REFERENCES

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

19. Tsaousis, A.D., Kunji, E.R.S., Goldberg, A.V., Lucocq, J.M., Hirt, R.P., and Embley, T.M. (2008). A novel route for ATP acquisition by the remnant mitochondria of Encephalitozoon cuniculi. Nature *453*, 553–556.
20. Lill, R. (2009). Function and biogenesis of iron-sulphur proteins. Nature *460*, 831–838.
21. Tsaousis, A.D., Gentekaki, E., Eme, L., Gaston, D., and Roger, A.J. (2014). Evolution of the cytosolic iron-sulfur cluster assembly machinery in Blastocystis species and other microbial eukaryotes. Eukaryot. Cell *13*, 143–153.
22. Tian, H.-F., Feng, J.-M., and Wen, J.-F. (2012). The evolution of cardiolipin biosynthesis and maturation pathways and its implications for the evolution of eukaryotes. BMC Evol. Biol. *12*, 32.
23. Wideman, J.G., Gawryluk, R.M.R., Gray, M.W., and Dacks, J.B. (2013). The ancient and widespread nature of the ER-mitochondria encounter structure. Mol. Biol. Evol. *30*, 2044–2049.
24. Yarlett, N., Lindmark, D.G., Goldberg, B., Moharrami, M.A., and Bacchi, C.J. (1994). Subcellular localization of the enzymes of the arginine dihydrolase pathway in Trichomonas vaginalis and Tritrichomonas foetus. J. Eukaryot. Microbiol. *41*, 554–559.
25. Yousuf, M.A., Mi-ichi, F., Nakada-Tsukui, K., and Nozaki, T. (2010). Localization and targeting of an unusual pyridine nucleotide transhydrogenase in Entamoeba histolytica. Eukaryot. Cell *9*, 926–933.
26. Smith, A.C., Blackshaw, J.A., and Robinson, A.J. (2012). MitoMiner: a data warehouse for mitochondrial proteomics data. Nucleic Acids Res. *40*, D1160–D1167.
27. Lucattini, R., Likic, V.A., and Lithgow, T. (2004). Bacterial proteins predisposed for targeting to mitochondria. Mol. Biol. Evol. *21*, 652–658.
28. Denic, V. (2012). A portrait of the GET pathway as a surprisingly complicated young man. Trends Biochem. Sci. *37*, 411–417.
29. Borgese, N., Brambillasca, S., and Colombo, S. (2007). How tails guide tail-anchored proteins to their destinations. Curr. Opin. Cell Biol. *19*, 368–375.
30. Mowbrey, K., and Dacks, J.B. (2009). Evolution and diversity of the Golgi body. FEBS Lett. *583*, 3738–3745.
31. Liapounova, N.A., Hampl, V., Gordon, P.M.K., Sensen, C.W., Gedamu, L., and Dacks, J.B. (2006). Reconstructing the mosaic glycolytic pathway of the anaerobic eukaryote Monocercomonoides. Eukaryot. Cell *5*, 2138–2146.
32. Lloyd, D., Ralphs, J.R., and Harris, J.C. (2002). Giardia intestinalis, a eukaryote without hydrogenosomes, produces hydrogen. Microbiology *148*, 727–733.
33. Nixon, J.E.J., Field, J., McArthur, A.G., Sogin, M.L., Yarlett, N., Loftus, B.J., and Samuelson, J. (2003). Iron-dependent hydrogenases of Entamoeba histolytica and Giardia lamblia: activity of the recombinant entamoebic enzyme and evidence for lateral gene transfer. Biol. Bull. *204*, 1–9.
34. van der Giezen, M., and Tovar, J. (2005). Degenerate mitochondria. EMBO Rep. *6*, 525–530.
35. Müller, M., Mentel, M., van Hellemond, J.J., Henze, K., Woehle, C., Gould, S.B., Yu, R.-Y., van der Giezen, M., Tielens, A.G.M., and Martin, W.F. (2012). Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiol. Mol. Biol. Rev. *76*, 444–495.
36. Makiuchi, T., and Nozaki, T. (2014). Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa. Biochimie *100*, 3–17.
37. Schofield, P.J., Edwards, M.R., Matthews, J., and Wilson, J.R. (1992). The pathway of arginine catabolism in Giardia intestinalis. Mol. Biochem. Parasitol. *51*, 29–36.
38. Yarlett, N., Martinez, M.P., Moharrami, M.A., and Tachezy, J. (1996). The contribution of the arginine dihydrolase pathway to energy metabolism by Trichomonas vaginalis. Mol. Biochem. Parasitol. *78*, 117–125.
39. Stairs, C.W., Eme, L., Brown, M.W., Mutsaers, C., Susko, E., Dellaire, G., Soanes, D.M., van der Giezen, M., and Roger, A.J. (2014). A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist Pygsuia. Curr. Biol. *24*, 1176–1186.
40. van der Giezen, M., Cox, S., and Tovar, J. (2004). The iron-sulfur cluster assembly genes iscS and iscU of Entamoeba histolytica were acquired by horizontal gene transfer. BMC Evol. Biol. *4*, 7.
41. Maralikova, B., Ali, V., Nakada-Tsukui, K., Nozaki, T., van der Giezen, M., Henze, K., and Tovar, J. (2010). Bacterial-type oxygen detoxification and iron-sulfur cluster assembly in amoebal relict mitochondria. Cell. Microbiol. *12*, 331–342.
42. Mi-ichi, F., Abu Yousuf, M., Nakada-Tsukui, K., and Nozaki, T. (2009). Mitosomes in Entamoeba histolytica contain a sulfate activation pathway. Proc. Natl. Acad. Sci. USA *106*, 21731–21736.
43. Nývltová, E., Šuták, R., Harant, K., Šedinová, M., Hrdy, I., Paces, J., Vlček, Č., and Tachezy, J. (2013). NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of Mastigamoeba balamuthi. Proc. Natl. Acad. Sci. USA *110*, 7371–7376.
44. Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M., and Barras, F. (2003). Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase. J. Biol. Chem. *278*, 38352–38359.
45. Riboldi, G.P., de Oliveira, J.S., and Frazzon, J. (2011). Enterococcus faecalis SufU scaffold protein enhances SufS desulfurase activity by acquiring sulfur from its cysteine-153. Biochim. Biophys. Acta *1814*, 1910–1918.
46. Chahal, H.K., and Outten, F.W. (2012). Separate FeS scaffold and carrier functions for $SufB_2C_2$ and SufA during in vitro maturation of [2Fe2S] Fdx. J. Inorg. Biochem. *116*, 126–134.
47. Tsaousis, A.D., Ollagnier de Choudens, S., Gentekaki, E., Long, S., Gaston, D., Stechmann, A., Vinella, D., Py, B., Fontecave, M., Barras, F., et al. (2012). Evolution of Fe/S cluster biogenesis in the anaerobic parasite Blastocystis. Proc. Natl. Acad. Sci. USA *109*, 10426–10431.
48. Maguire, F., and Richards, T.A. (2014). Organelle evolution: a mosaic of 'mitochondrial' functions. Curr. Biol. *24*, R518–R520.
49. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.
50. Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics *19* (*Suppl 2*), ii215–ii225.
51. Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics *7*, 62.
52. Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. *31*, 5654–5666.
53. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. *9*, R7.
54. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
55. Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. *33*, D501–D504.
56. Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29-37.
57. Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Res. *40*, D290–D301.
58. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. *35*, W182-5.

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

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

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

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

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

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

**Current Biology, Volume *26***

## Supplemental Information

# A Eukaryote without a Mitochondrial Organelle

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

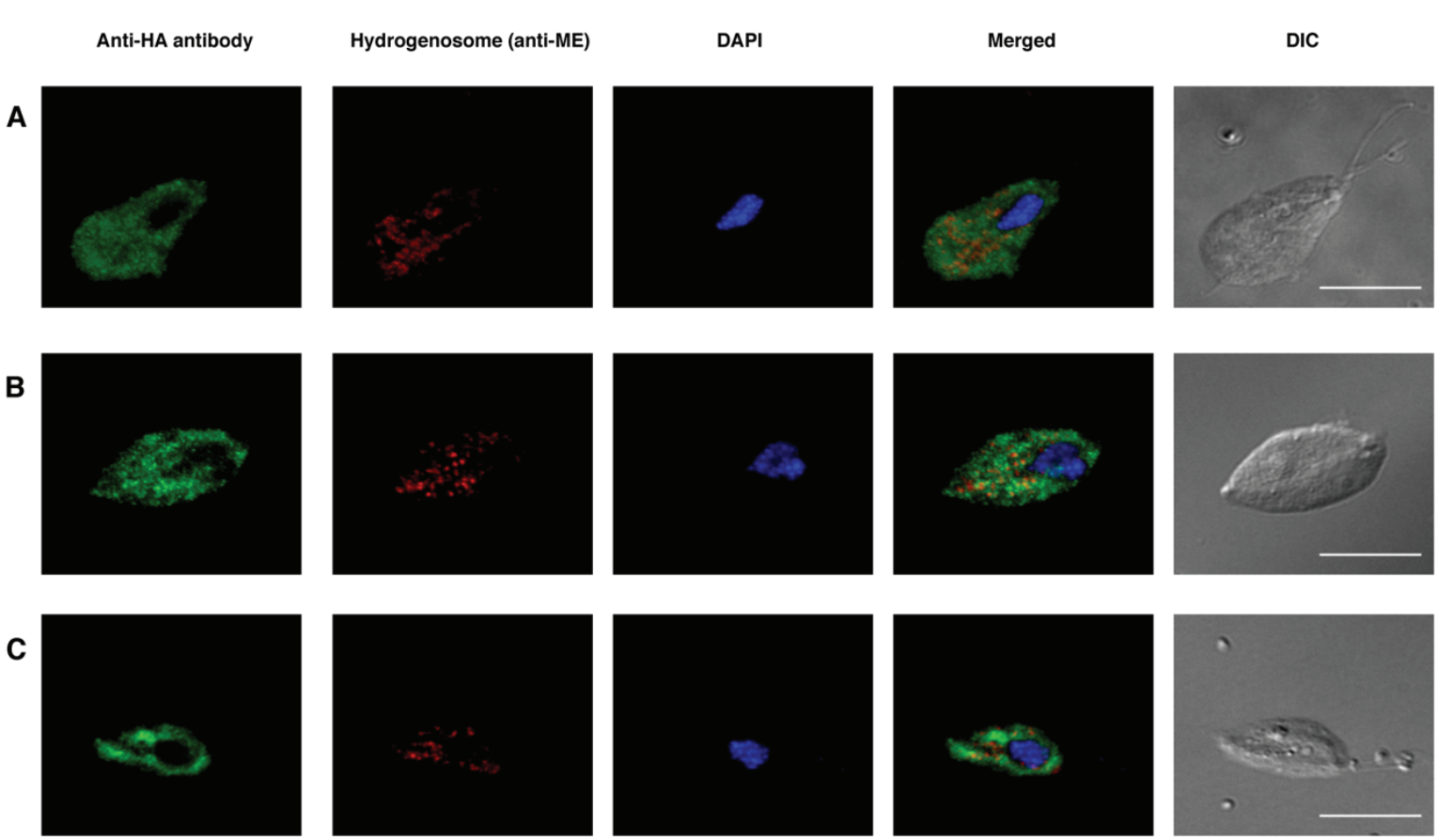



**Figure S1 (related to Figure 3). Heterologous localisation of *Monocercomonoides* sp. pyruvate-ferredoxin oxidoreductase (PFOR) in *Trichomonas vaginalis*. Related to Experimental Procedures.**

(A) PFOR1. (B) PFOR2. (C) PFOR3. Indicated proteins were HA-tagged at the C-terminus and expressed in *T. vaginalis* (green). The hydrogenosome was stained using an anti-Malic enzyme (ME) antibody (red) and the nucleus was stained using DAPI (blue). Scale bar, 10 μm.

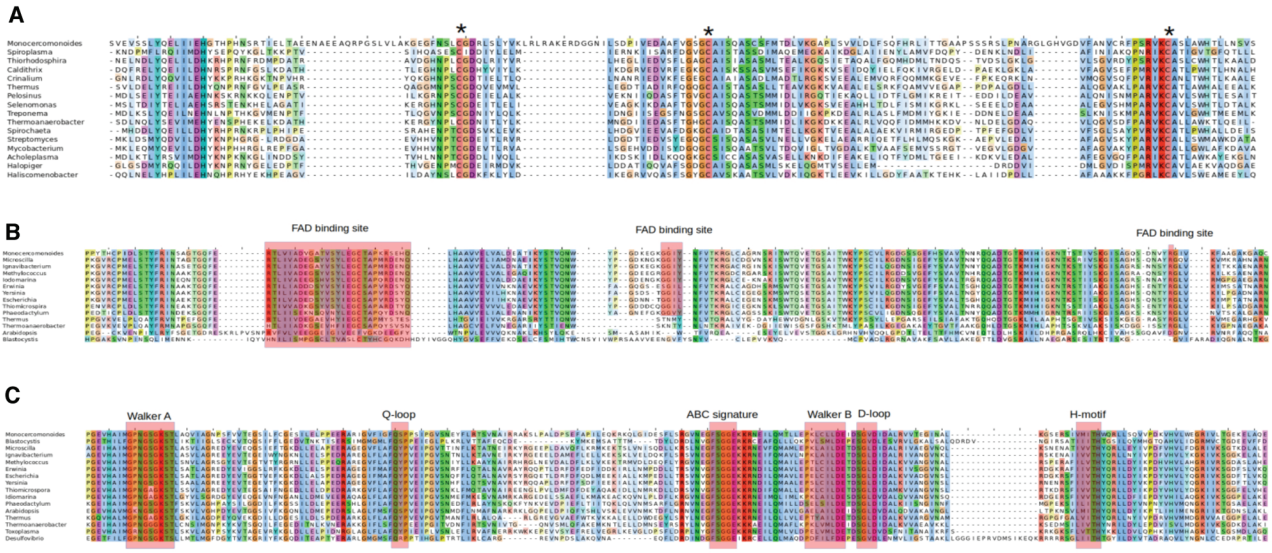



**Figure S2 (related to Figure 3). Alignments of catalytic sites in SUF system proteins of *Monocercomonoides* sp.** (A) SufU, all three conserved catalytic cysteins are present (indicated by stars) suggesting SufU can fulfil its role as an enhancer of cystein desulfurase SufS activity. (B) SufB, all three FAD binding sites are present (highlighted in red); they are present in most bacterial sequences but absent from those of *Pygsuia biforma* and *Blastocystis hominis*. (C) SufC, catalytic domains highlighted in red, all six important domains required for functioning as ATPase are present and well conserved.

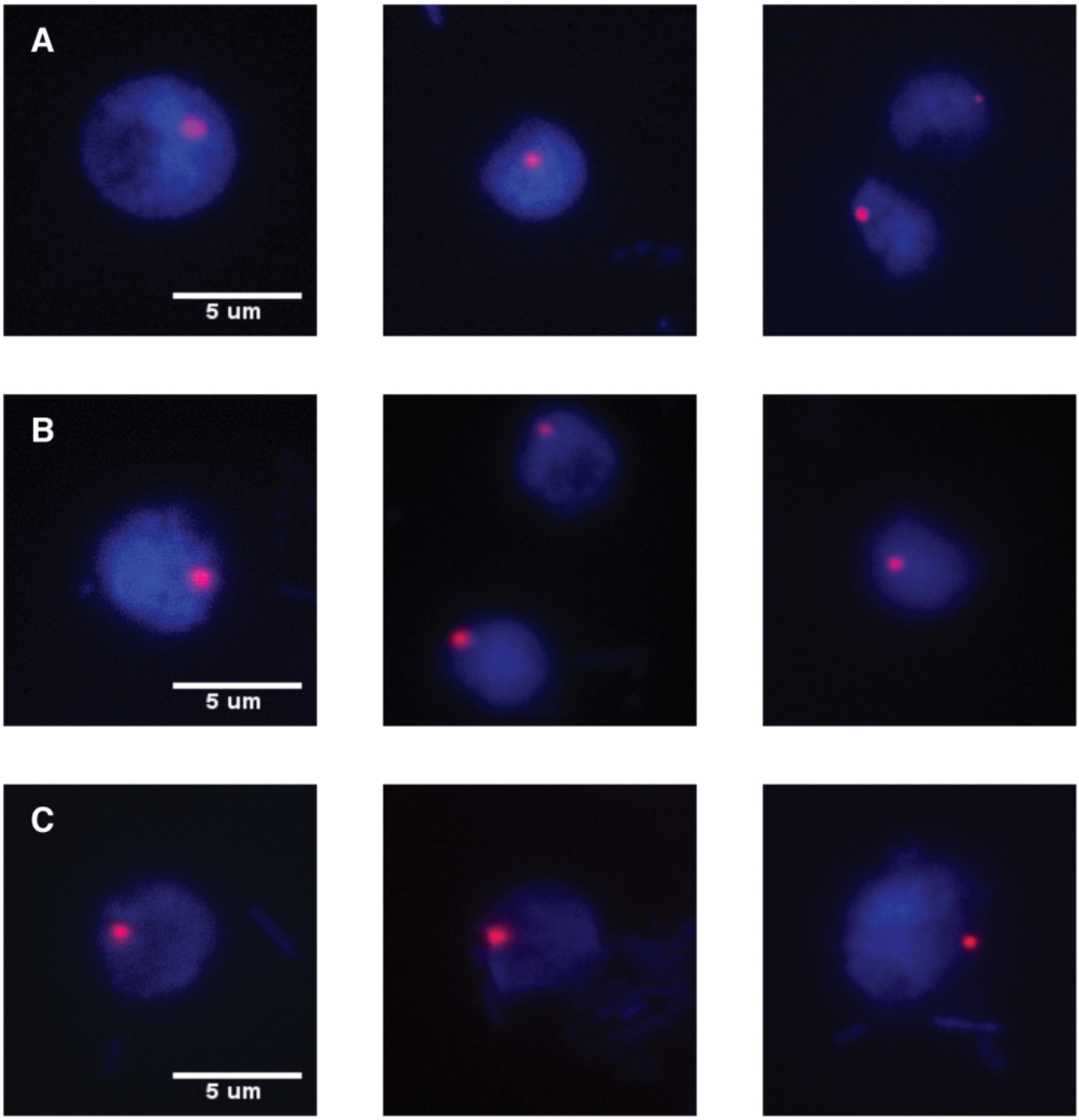



**Figure S3 (Related to Figure 3). The nuclear localisation of the SUF genes in *Monocercomonides* sp. detected by fluorescent in situ hybridization (FISH).**
FISH experiments with Suf gene probes ((A) *sufB*, (B) *sufSU* and (C) *sufC*) and DAPI staining show that genes of the SUF system reside in the genome of *Monocercomonoides* sp. and are not bacterial contamination. Sensitivity of FISH-TSA method allowed visualization of single-copy genes. Signals from single-copy gene probes indicate haploidy of *Monocercomonoides* sp. nuclei. Over 50 nuclei were examined in each experiment.

**Supplemental Experimental Procedures**

**Culture**

*Monocercomonoides* sp. PA203 strain was isolated from an individual of *Chinchilla laniger* by prof. Jaroslav Kulda in 1993 and is deposited in culture collection of the Department of Parasitology at Charles University in Prague. The agnotobiotic culture with bacteria, but no other eukaryote, was maintained by serial transfer every 2–4 days in modified TYSGM-9 medium [S1] at 37°C. A clonal lineage of *Monocercomonoides* sp., later used for genome and transcriptome sequencing, was prepared by serial dilution method.

**Contamination filtration**

Before DNA/RNA isolation, the largest portion of bacterial contamination was removed by filtration through a filter paper by gravity flow. The filtrate containing *Monocercomonoides* sp. cells was then filtered through a 3-μm pore polycarbonate filter (Whatman International Ltd., Maidstone, UK). Most bacteria appeared in the flow-through, while the trophozoites of *Monocercomonoides* sp. remained in the medium above the filter. The suspension of trophozoites was washed by continuous addition of fresh medium (approximately two volumes of the original culture) and finally transferred to a clean vessel. Filtration and wash steps were sometimes accelerated by application of a partial vacuum.

**Genome Sequencing and assembly**

Several shotgun and paired-end fragment libraries were prepared according the Rapid Library Preparation Method and Paired End Rapid Library Preparation Method - 3 kb Span protocols developed by Roche. The isolated genomic DNA (see Experimental Procedures) was sheared using either nebulization (for the shotgun library) or the Digilab Hydroshear device set up for 20 cycles at a calibrated speed 12 with the standard shearing assembly (for the paired-end library).

Shotgun fragment libraries with the average fragment length of 500 to 1200 bases were subsequently ligated to adaptors and amplified by emulsion PCR on beads using CPB ratio 7 (DNA copy per bead). The enriched beads were recovered from the emulsion, applied to a large region version of PicoTiter Plate and run four times on the 454 GS FLX+ sequencer using XL+ chemistry to generate over 2.5 Gb of sequencing data.

One 3 kb paired-end library was amplified by emulsion PCR using CPB ratio 1 and run three times on large region version of PicoTiter Plate on the 454 GS FLX+ sequencer using XL+ chemistry resulting in 1 Gb of sequencing data. GS Run Processor 2.8 (Roche) was used for standard image and signal processing in all cases.

We generated 2.4 Gb of the shotgun sequences and 1 Gb of the 3 kb paired end sequences, respectively, together in 8.5 millions of reads. The Newbler v2.8 assembler (Roche) was used to generate final genome sequence assembly from 454 single and pair reads (MIRA assembler was also tested but resulted slightly worse assembly). In the assembly there could be seen 9 different 16-23S rRNAs sequences of the bacterial origin. The final assembly was filtered with bacterial scaffolds, which could be easily recognized as large scaffolds homologous to known bacterial genomes (Genomic data filtering). The final assembly consist of 2,095 scaffolds spanning 74.74 Mb of the genome. Average coverage of the scaffolds was 35x and these contained 1.1 millions of both mapped paired end reads. The N50 was 71.4 kb which means that 50 % of the entire assembly is contained in scaffolds larger than 71.4 kb. The average GC content of these scaffolds is 36.8%.

**Genomic data filtering**

Since *Monocercomonoides* sp. grows with a mixture of bacterial species, a significant portion of assembled scaffolds was presumed to be bacterial in origin. Scaffolds were assessed as bacterial if they had high nucleotide sequence identity to known bacterial genomes using BLASTN again the non-redundant database available through NCBI. A total of 2,021 scaffolds longer than 2 kb showed high sequence similarity to nine bacterial genomes (from *Aeromonas hydrophila* subsp. *hydrophila* ATCC 7966, *Bacteroides fragilis* NCTC 9343, *Citrobacter koseri* ATCC BAA-895, *Clostridium saccharolyticum* WM1, *Enterobacteriaceae bacterium* strain FGI 57, *Eubacterium limosum* KIST612, *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586, *Haemophilus influenzae* R2866, and *Parabacteroides distasonis* ATCC 8503). The predicted GC content of *Monocercomonoides* sp. (36.8%) was also used to distinguish bacterial sequences (average of 47.7%).

The large amount of bacterial and repetitive sequences complicated the sequencing and assembly of the *Monocercomonoides* sp. genome. As a consequence, 12,439 contigs (both *Monocercomonoides* sp. and bacterial) of less than 2 kb in length were excluded from the final genome assembly and were treated as a separate set of data. BLASTN searches indicated that 3,343 of these smaller fragments had significant similarity to portions of the main *Monocercomonoides* sp. genome assembly. All contigs were also used for targeted searches for mitochondrial hallmarks and other genes of interest.

After gene prediction and annotation (see protein-coding gene finding and annotation) five more scaffolds were additionally removed from the final *Monocercomonoides* sp. assembly due to their high similarity to bacterial sequences. On the other hand, two smaller contigs containing genes discussed in the paper were added

to the genome assembly. As a result, the assembly submitted to GenBank contains 2,095 scaffolds (total length 74.72 Mb; N50=71.4 kb).

**Estimating genome completeness**

We used Conserved Eukaryotic Genes Mapping Approach (CEGMA) [S2] to estimate the number of conserved eukaryotic genes in the *Monocercomonoides* sp. genome assembly. The pipeline identified 157 out of 248 genes (63.3 %) from the predefined set of conserved eukaryotic genes (CEGs). Since this was a surprisingly low percentage, we applied the same CEGMA methodology to other divergent species. Similar numbers were seen in high quality complete genome sequences from *Plasmodium falciparum* (75.0%), *Giardia intestinalis* (46.4%) [S3] and *Trichomonas vaginalis* (68.5%) [S4]. This observation could be explained by the divergent nature of these genomes relative to the sequences present in the pipeline. In divergent genomes, the success of detection of a gene is related to the gene sequence conservation; that is, more conserved genes are detected more frequently. Consequently, the fraction of mapped CEGs is an underestimate of the genome sequence completeness. However, conserved genes (a highly conserved fraction of CEGMA CEGs defined as Group 4 in the CEGMA dataset [S3]) are easier to identify. For instance, a high proportion of Group 4 CEGs were identified in *Plasmodium falciparum* (96.6%), *G. intestinalis* (67.7%) and *T. vaginalis* (86.1%). Similarly, 81.5% of Group 4 CEGs were identified in the *Monocercomonoides* sp. suggesting that this genome is divergent and not necessarily incomplete.

The CEGMA pipeline performs *ab initio* gene prediction without any training set and transcriptomic data. This likely leads to poorly-predicted gene models compared to those obtained using a suite of methods for gene prediction we applied for our genomic data discussed in section Protein-coding gene finding and annotation. To validate the absence of genes not detected by the CEGMA pipeline we performed a hidden Markov model (HMM) search of missing CEGMA CEGs (a total of 91) in predicted *Monocercomonoides* sp. gene models using HMMer [S5]. After manual investigation of these cases we determined that (i) 41 genes were in fact in the genome but not identified in the initial CEGMA search; (ii) 28 genes represented mitochondrial proteins which we propose are missing from *Monocercomonoides* sp.; and (iii) 24 genes not directly related to the mitochondrion are genuinely absent in the *Monocercomonoides* sp. genome (Table S1). If consider the 41 CEGs detected by HMM, 80% (198) of CEGs were detected, and if we omit also the 28 KOGs that are exclusively associated with mitochondria 90% of CEGs from CEGMA were detected in *Monocercomonoides* sp.

We used also BUSCO [S6] pipeline to estimate genome completeness. We identified 34% of BUSCOs using its own implemented gene prediction pipeline, the same procedure was able to identify 40% of BUSCOs for *G. intestinalis* genome and 98.8% for yeast genome. We used also predicted proteins of *Monocercomonoides* sp. and *G. intestinalis*. This analyses resulted in 73% and 50% of BUSCOs respectively.

In the course of preparing a supermatrix for phylogenomic analyses (described in section Phylogenetic analyses) we identified for *Monocercomonoides* sp. 155 of the 163 (95%) eukaryotic genes used for the analyses.

We were also able to recover the complete set (77) of conserved families of cytosolic eukaryotic ribosomal proteins [S7] (Table S3), with single exception of L41e.

**cDNA library construction and transcriptome sequencing**

Total RNA was isolated using TRIzol Reagent (Life Technologies) from $16 \cdot 10^7$ cells that were (i) filtered using the procedure described above or (ii) unfiltered to avoid stressing the cells and thus altering transcript expression. For the construction of a 454 sequencing library, messenger RNA (mRNA) was isolated from total RNA (from filtered cells) using the Dynabeads mRNA purification kit (Life Technologies) and complementary DNA (cDNA) was prepared using the Smarter PCR cDNA synthesis kit (Clontech) with 19 cycles of cDNA amplification. A sequencing library was prepared using the GS FLX Titanium Rapid Library Preparation Kit and the fragment library was titrated and amplified by emulsion PCR and sequenced using the 454 technology on a GS-FLX Titanium PicoTiter Plate. A total of 508, 593 reads were generated resulting in 9,773 contigs. Transcriptome assembly of the 454 data was performed by Newbler v2.6 with default parameters (40 bp overlap and 90% identity).

For an Illumina sequencing library, 22 µg of total RNA from unfiltered cultures was sent to the Beijing Genomics Institute (BGI)-Shenzhen, Shenzhen, China (http://www.genomics.cn/index.php) and sequenced by using the HiSeq™ 2000 platform according to the manufacturer's instructions (Illumina, San Diego, CA). In brief, mRNA was isolated from total RNA using Sera-mag Magnetic Oligo (dT) Beads (Illumina) and fragmented into smaller pieces. cDNA was synthesized using the SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen, Camarillo, CA) with random hexamer (N6) primers (Illumina). cDNA fragments of size $200 \pm 25$ bp were selected and 15 rounds of PCR amplification were performed to enrich the purified cDNA template using PCR Primer PE 1.0 and PE 2.0 (Illumina) with Phusion DNA Polymerase. The cDNA library was sequenced on a PE flow cell using Illumina Genome Analyzer HiSeq 2000.

Illumina HiSeq 2000 sequencing resulted in 51,880,922 raw reads. Reads that did not pass the Illumina

built-in Failed-Chastity filter (chastity threshold 0.6) were removed. Furthermore, all reads with adaptor contamination (i.e. represented sequences from a multiplexed sample) were discarded. Low quality reads with more than 5% ambiguous sequences ("N"s) were removed. Finally, reads with more than 20% Q < 20 bases were also removed. After filtration the remaining 48,774,804 reads were assembled by Trinity [S8] using default parameters into 91,319 contigs (with the mean length of 384 bp). A total of 54, 998 Trinity-predicted non-redundant 'unigenes' were selected (with the mean length of 575 bp).

**Protein-coding gene finding and annotation**

Pipelines for *ab initio* gene prediction perform better if a training set of at least 200 known gene models is provided [S9]. Unfortunately, we did not have previous data from *Monocercomonoides* sp. that would be suitable for a training set and there are no genomes sequenced from close relatives available.
We used the following procedure for constructing a training set:

1. Prediction of conserved eukaryotic genes by the CEGMA pipeline (discussed above) [S2].
2. PASA2 assembly of transcripts [S10]. PASA strictly aligns transcript sequences to the genome (using gmap and blat) and assembles the aligned sequences into transcripts. In this step, a set of 237 assembled transcripts with complete structure (start and stop codon), only one possible start codon (accurate 5'end), at least one intron and reasonable BLASTP [S11] hits against the NCBI database were selected.
3. CEGMA and PASA sets of models were merged and redundant genes (with 70% identity at the amino-acid level) were removed. The remaining protein sequences were used as queries to search against the NCBI nr protein database to identify and remove low-quality models resulting in 446 gene models. This set was used as an initial training set for the first round of prediction performed by Augustus v2.7 [S9, S12].
4. Full-length PASA ORFs were compared with the gene models in the training set, and only those gene models which overlapped with PASA ORFs (excluding those which had more than one possible start codon in the first exon) were kept. Additionally, we manually searched these gene sets for specific genes of interest involved in energy, nucleotide and amino acid metabolism and genes for members of the Ras superfamily of GTPases. Selected gene models were manually curated and added to the training set. Genes with introns – indicating eukaryotic provenance and more useful for training – were retained resulting in 433 gene models in the final training set.
5. Gene models were predicted using the *ab initio* predictor Augustus v2.7 trained with the training set of models prepared in the previous step and with "hints" from the transcriptomic data.
6. Evidence modeller (EVM) [S13] is a software package that predicts the weighted consensus gene structure by combining information from gene predictions and protein and transcript alignments. We used EVM to predict gene structure from the Augustus and PASA predicted assemblies.
7. PASA was used to improve gene-model structure and add untranslated regions (UTR) to the draft gene set. Two cycles of annotation comparison and annotation updates were performed in order to maximize the incorporation of transcript alignments into gene structures.

This gene prediction pipeline resulted in 16,751 gene models used for automatic functional annotation.

Illumina RNA-seq reads (94.7%) were mapped on the genome assembly using TopHat2 [S14] and visualized in Integrative Genomics Viewer (IGV) [S15]. 97% of the gene models have transcriptome coverage. A large part of the unmapped 5.3% of transcriptome reads probably represents bacterial contamination in the sequenced transcriptome. We used bacterial genomes of identified major contaminants of the genomic data to clean the reads. We removed reads which were mapped by Bowtie2 [S16] to 9 bacterial genomes identified as main contaminants in the genomic data (see Genome data filtering). We assembled cleaned reads using Trinity [S8] and mapped them by GMAP [S17] to the final assembly (96.9%) and to all contigs from the assembled genomic data (97.6%). We annotated transcripts not mapped to the genome by similarity searches using BLAST (e-value =< $1e^{-20}$) against NCBI nucleotide and nr protein databases. We were able to assign 2282 sequences as bacterial (2152), plant (117) and animal (including human) (13) contaminations. Remaining 408 transcripts (0.4% of all transcripts) did not have any reasonable annotation.

The automatic functional annotation was performed by similarity searches using BLAST (e-value=<$1e^{-20}$) against NCBI nr protein database and HMMER [S18] searches of domain hits from Pfam protein motif database [S19]. Additional annotation was performed using the KEGG Automatic Annotation Server [S20] which compares predicted genes to the manually curated KEGG Genes database [S21]. 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 locus tag identifier in the format MONOS_XXXXX was assigned to each predicted gene. Approximately 60% of the gene models remained as unannotated. GFF3 and FASTA files are available at the web page: http://www.protistologie.cz/hampllab/data.html.

In addition, 6-frame translation of all scaffolds was performed and ORFs longer than 70 amino acids were used for automatic functional annotation and as an additional database searched for proteins of interest.

**RNA-coding gene finding and annotation**

All twenty tRNA synthetases and 153 tRNA genes (nine with introns) were identified in *Monocercomonoides* sp. as predicted by tRNAscan-SE [S22] distributed on 132 different scaffolds.

Ribosomal DNA (rDNA) sequences were identified in 22 contigs that were not assembled into scaffolds, as it was also observed for the *T. vaginalis* genome assembly [S4]. We suspect that the scattered assembly of the locus across multiple contigs is likely due to sequence heterogeneity among multiple copies of the rDNA operon. The coverage of the consensus rDNA operon sequence was very high (~2,000 x) compared to the contigs assembled into the main assembly (35x). From these data we estimate that the copy number of the rDNA locus is approximately 50.

**Analysis of repetitive sequences**

RepeatMasker v3.3.0 [S23] was used to detect repetitive regions in the genome. We used the RepbaseUpdate database [S24] containing sequences representing repetitive DNA from different eukaryotic species. However, due to the divergent nature of the *Monocercomonoides* sp. genome compared to the eukaryotes present in the RepbaseUpdate database, RepeatMasker failed to detect all observed repeat segments. To improve repeat detection, we created a custom library of 1,158 repetitive elements by performing *de novo* repeat identification using RepeatScout [S25]. To ensure that conserved protein families were not flagged as repetitive elements, BLAST2GO [S26] was used to determine similarity between the putative repeats and any known proteins. After this step, the library of repeats contained 1,000 repeat sequence masking 38.55% of the genome. Importantly, we did not perform this masking until after gene models were predicted. Analysis of an overlap between predicted repeats and ORFs (described above) enabled to detect proteins encoded by some well-known transposable elements (TEs), like reverse transcriptase or phage integrase.

**Gene searching**

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* from GiardiaDB.org*, Homo sapiens* from NCBI, *Naegleria gruberi* v1.0 from genome.jgi-psf.org, *Paratrimastix pyriformis* from NCBI, *T. vaginalis* G3 from TrichDB.org, *Trypanosoma brucei* TREU927 release 6.0 from eupathdb.org and *Saccharomyces cerevisiae* RM11-1a from www.broad.mit.edu and S288C from NCBI*,* as well as *Escherichia coli* and *Bacillus subtilis* for SUF machinery components and *Azotobacter vinelandii* for NIF machinery components. *Monocercomonoides* hits were blasted back against the genome of the query protein and against NCBI. Mitochondrial marker proteins (components of the mitochondrial protein import machinery and mitochondrial carrier family proteins) were searched in *Monocercomonoides* sp. proteome and in contigs not included in the assembly by a library of (HMMs) using HMMER [S18]. The identified sequences were used as queries in HHpred [S27] searches. Additionally using Bowtie2 [S16] we mapped raw reads from the *Monocercomonoides* sp. to the most conserved mitochondrial genes from *P. pyriformis* and *G. intestinalis* to exclude possibility that those genes were not identified because they are not present in the assembly.

**Mitominer searching**

To exhaustively search for any possible mitochondrial protein we prepared a custom mitochondrial protein sequence database. We combined the MitoMiner database Reference Set [S28] (12,925 proteins from 11 mitochondrial proteomes) with MROs proteins of *Entamoeba histolytica*, *G. intestinalis*, *Pygsuia biforma, Spironucleus salmonicida, T. vaginalis*, and *P. pyriformis* excluding ribosomal proteins (as no MRO-targeted ribosomal proteins were expected to be present in *Monocercomonoides* sp. in the absence of a mitochondrial genome). Redundant homologues (90% similarity threshold) were removed to decrease the database size. The resulting non-redundant database (myMitoMiner) contained 4,869 proteins. We performed a reciprocal best BLAST hit analysis using this database against the *Monocercomonoides* sp. predicted protein models with an e-value threshold of 0.001 and identified 1,859 candidates. This high number was partially the result of presence of false positives (i.e. cytosolic proteins) in the myMitoMiner database and low stringency of the search parameters, which retrieved many cytosolic paralogues of mitochondrial proteins. To filter this set of candidate proteins, we performed homology searches (BLASTP) of *Monocercomonoides* sp. candidates against the NCBI nr database and collected the top hit. If the first hit retrieved from nr had a predicted mitochondrial targeting sequence (> 0.5 by TargetP v1.1 [S29] or MitoProt II v1.101 [S30]) the *Monocercomonoides* sp. query was investigated further. Only 76 proteins passed all criteria and were assigned KEGG and GO categories using KEGG Automatic Annotation Server [S20] and Interproscan [S31]. Many of the candidates have clearly defined functions, which are not related to mitochondria (for example histones or Arf family proteins). The few proteins annotated to the GO category "Metabolism" do not appear to function in similar pathways (Table S5). Moreover, any of the candidates is exclusively localised to mitochondria in other eukaryotes. Presented results suggest that the candidates recovered by the myMitoMiner pipeline, i.e. the most probable mitochondrial proteins in

*Monocercomonoides* sp., are false positives. This raises the question if the pipeline is able to recover true mitochondrial proteins.

To validate the methods outlined above, we applied the same approach to recover mitochondrion-targeted proteins in several organisms (modifying the custom and NCBI nr databases to exclude them while searching for mitochondrial proteins). We chose five organisms with different types of MROs (more and less reduced) and mitochondria. As an examples of organisms with highly reduced MROs and divergent set of proteins we chose *G. intestinalis* and *E. histolytica*. Only 30 proteins were experimentally shown to localise to the mitosome of *G. intestinalis* and 139 and 95 proteins were detected in its MRO fractions of *G. intestinalis* and *E. histolytica*, respectively, by a proteomic analysis [S32, S33, S34]. *T. vaginalis* contains less reduced MRO - hydrogenosome, with 569 proteins detected in proteomic studies [S35]. *Blastocystis* sp. subtype 7 is bearing MRO with 365 proteins predicted to localise to its MRO [S36]. Finally we used *S. cerevisiae,* which has a well-characterized mitochondrial proteome. We applied the same methodology to these organisms. In the case of *G. intestinalis* our search identified 45 candidates for mitochondrial proteins, of which two have been experimentally localised to the MRO of *G. intestinalis* (Table S5B). For *E. histolytica* our search identified 55 candidates, of which three were detected in mitosomal proteome (Table S5C). Although those are low numbers, it demonstrates that it is possible to identify mitochondrial proteins from a divergent organism. We detected more candidates for less reduced *T. vaginalis* (202 candidates, 25 detected in proteomic studies) (Table S5D) and *Blastocystis* sp. (241 candidates, 97 of them predicted as localised to the MRO) (Table S5E). Application of the pipeline on the dataset of *S. cerevisiae* proved more fruitful yielding 223 candidates, of which 187 were mitochondrial according to experimental evidence (GFP and/or Mass-Spec) GO classifications (Table S5F). These tests demonstrate that the pipeline is able to detect a large fraction of typical mitochondrial proteins, including at least some of the most divergent ones.

**Search for mitochondrial signature sequences**

*Targeting signals*. In model organisms such as animals, plants, and fungi, some mitochondrial proteins (localised to the matrix, inner membrane or inter-membrane space) are encoded on the nuclear genome and possess an N-terminal sequence that will direct the protein post-translationally to the mitochondria. To detect these mitochondrial targeting signals (MTS), we used two publicly available prediction software tools TargetP v1.1 [S29] and MitoProt II v1.101 [S30]. We detected 107 proteins (with annotation) that had a putative MTS score greater than 0.5 from either program (Table S6A).

*Tail anchored proteins.* Some mitochondrial outer membrane (MOM), endoplasmic reticulum (ER) and peroxisomal proteins are so-called tail-anchored (TA) proteins [S37]. TA proteins are typically associated with the cytoplasmic face of the phospholipid bilayer via a single segment of hydrophobic amino acids localised less than 30 amino acids from the C-terminus. MOM proteins in particular are thought to possess a transmembrane domain (TMD) that is flanked by positively charged residues [S38]. To identify putative MOM candidate proteins, we used TMHMM v2.0 [S39] software tools to identify a total of 32 proteins with one transmembrane helix within 31 residues of the C-terminus (Table S6B). Manual investigation of these 32 proteins did not reveal any obvious mitochondrial proteins. Instead, the candidates with the shortest TMD and the highest charge flanking TMD were mainly well-known ER-associated proteins like Syntaxin 16 or mannose-binding endoplasmic reticulum-Golgi intermediate compartment lectin.

*β-Barrel proteins.* In mitochondria, seven subclasses of β-Barrel outer membrane proteins (MBOMPs) have been identified: Tom40, Sam50, VDAC, Mdm10, ATOM, Tac40 and MBOMP30 [S40]. MBOMPs typically contain a conserved C-terminal β-signal which promotes insertion to the MOM [S41]. We used a previously established bioinformatics pipeline for *de novo* identification of MBOMPs [S42]. This pipeline searches for a $P_oxGh_yxH_yxH_y$ motif ($P_o$, non-negatively charged polar residue; G, glycine; $H_y$, large hydrophobic residue; $h_y$, hydrophobic residue including Ala and Cys; and x, any residue) in the C-terminus. A total of 330 *Monocercomonoides* sp. proteins were flagged as possessing a C-terminal β-signal. Additional filters employed by the pipeline [S42] reduced this data set to three sequences. Two of these proteins (MONOS_12978 and MONOS_5184) are annotated as ribosomal proteins while the remaining candidate (MONOS_11898) is unannotated and only 97 amino acids long. As most MBOMP contain a barrel of at least 150 residues, it is unlikely that MONOS_11898 is a genuine MBOMP [S43].

**Phylogenetic analyses**

*Supermatrix for a phylogenomic analysis*. To investigate the phylogenetic position of *Monocercomonoides* sp. in the eukaryotic tree of life, we added the corresponding *Monocercomonoides* sp. and *P. pyriformis* [S44] orthologues to a previously published dataset of 163 conserved eukaryotic proteins [S45] (kindly provided by Dr. Martin Kolisko, University of British Columbia). New sequences were added to the pre-existing unmasked alignments using the “mafft–add" (align full length sequences) function of MAFFT [S46] and alignments were trimmed using the Gblocks server (http://molevol.cmima.csic.es/castresana/Gblocks_server.html) with the least stringent conditions for trimming. For each gene, a maximum likelihood (ML) tree was computed with rapid

bootstraps under the LG model of evolution with a gamma distribution for rate across sites in RAxML [S47] (v.8.1.11; CIPRES Science Gateway, http://www.phylo.org/sub_sections/portal/). The resulting trees were manually inspected to (i) detect possible contaminations or wrongly assigned orthologues and (ii) identify and retain the least divergent *Monocercomonoides* sp. and/or *P. pyriformis* copy in the cases of lineage-specific paralogues (in-paralogues). The final alignments were concatenated using FASconCAT [S48] and sequences of 20 taxa with a high proportion of missing data or closely related to other taxa in the dataset were removed. After all of the filtering steps, the resulting supermatrix contained 70 taxa, 163 genes, and 44,100 aligned amino acid positions. We identified a total of 155 out of these 163 genes in the *Monocercomonoides* sp. genome (see Table S2). The human homologues of the eight genes not identified in *Monocercomonoides* sp. include six mitochondrial proteins, a peroxisomal protein, and the cytosolic protein 5-oxoprolinase.

*SUF analyses data set preparation*. To investigate the phylogenetic history of the SUF proteins from *Monocercomonoides* sp., we retrieved representative prokaryotic homologues of SufB, SufC and SufS from the *nr* database at the NCBI. We combined this prokaryotic dataset with (i) eukaryotic representatives from chloroplast-bearing eukaryotes that have a chloroplast-localised SUF system, (ii) the SufC and SufB homologues from *Pygsuia biforma and Blastocystis* sp. subtype 7 and (iii) the SUF sequences from *Monocercomonoides* sp. and *P. pyriformis* identified here yielding a final dataset of 155 taxa.

*Phylogenetic reconstruction.* For phylogenomic analyses of the supermatrix of 163 conserved proteins the PhyloBayes tree was calculated using PhyloBayes [S49, S50] (v.3.3f, CAT-Poisson model). Bootstrap support values (200 replicates) from ML analysis were mapped on the PhyloBayes tree. For the analyses of the concatenated alignment of SUF proteins (see above), bootstrap support values (100 replicates) were mapped on the best-scoring ML tree from twenty independent ML tree using RAxML [S47] (v.8.0.23, PROTGAMMALG4X model). Bayesian inference posterior probabilities were calculated using PhyloBayes [S49, S50] (v.3.3f, CAT-GTR model).

**Fluorescence *in situ* hybridization (FISH)**

Unlabelled probes were amplified from *Monocercomonoides* genomic DNA using gene specific primers (5'ATCATAAAACAAATTTCTACAAAAAAGCACG3' and 5'CAGCACATCTTTTGCAAAGCC3' for *sufB*, 5'AGACGAAGATGACGATGAAGCA3' and 5'AGCCAATCACCGAAGATTGTCT3' for *sufC* and 5'ATGTGCACAGAGGAGTCCATTT3' and 5'GTCCTTCTCGATTCTGTCAGCA3' for *sufSU*). PCR products generated by PrimeSTAR Max DNA polymerase (Clontech, R045A) were cloned into the pJET1.2 blunt cloning vector (Thermo Scientific, K1231) and re-amplified from the plasmids to get 1 µg of input DNA for probe labelling. Purified PCR products were labelled by digoxigenin-11-dUTP, alkali stabile (Roche, 11093088910) using the DecaLabel DNA Labeling Kit (Thermo Scientific, K0621). Labelled probes were purified by the QIAquick Gel Extraction Kit (Qiagen, 28704) and eluted into the final volume of 50 µl.

***Trichomonas vaginalis* heterologous expression system**

The coding sequences of SUF system genes *sufB*, *sufC* and three copies of *pfor* were amplified from *Monocercomonoides* sp. cDNA by PCR with the following specific primers: PFOR1 (5‘CACTTCACATTACATATGACTGACAAAGAAATTGATT3‘ and 5‘CGTATGGGTAGGATCCAAATCCTTGTTCGGCCCAC3‘); PFOR2 (5‘ATCATTAATATGTCTCAGAATAAGGTA3' and 5‘TAAGGATCCCTTTGGATTAAACACGGCA3‘); PFOR3 (5‘TAACATATGATGTCTTCTGAAAATCAA3‘ and 5‘TAAGGATCCTTTTGCTGGTTGAGCAGC3‘); SufB (5'CATGATTAATATGACTGCTTCATCTAAACCTTCA3‘ and 5‘TGACGGATCCACCAACCGAGCCTTCCAAAAC3‘); SufC (5‘CATGGCATATGATGCAAACTCAAAAGCCCCTT3‘ and 5‘TGACGGATCCAATCTTCACAACTCCCTCTGC3‘).

The products were cloned into the TagVag2 vector. Only the gene for PFOR1 was cloned into plasmid directly using the In-fusion HD Cloning kit (Clontech, 639648). Lab-prepared chemically competent *Escherichia coli* XL1 cells were used for transformations with ligation mixtures, whereas Stellar competent cells (Clontech, 636763) were used for transformation with the in-fusion reactions. Bacterial clones were checked by colony PCR for the presence of insert-containing plasmids. Plasmids were purified from positive clones using the Wizard Plus Midipreps DNA Purification System (Promega, A7640), checked by sequencing the insert region, and electroporated into *T. vaginalis* T1 cells as previously described [S51].

Electroporated cells were selected with 200µg/ml of G418 (ZellBio GmbH, G-418-5) through at least five passages. Expression of the proteins was analysed by Western blotting of cell homogenates (data not shown) and immunofluorescence.

**Immunofluorescent microscopy**

*Monocercomonoides* sp. proteins expressed in *T. vaginalis* cells were visualised using an anti-HA rat monoclonal antibody (Roche, 11867423001). An antibody raised against the hydrogenosomal marker malic enzyme of *T. vaginalis* (kindly provided by prof. Jan Tachezy, Dept. of Parasitology, Charles University in Prague) was used for double-labelling. Alexa Fluor® 488 goat anti-rat (green) and Alexa Fluor® 594 goat anti-rabbit (red) (Life Technologies, A-11006 and A-11037) were used as secondary antibodies. Immunostaining was performed on Superfrost microscopic slides coated with poly-L-lysine (Sigma, P8920). Preparations were counterstained with DAPI in Vectashield mounting medium (Vector Laboratories, H - 1200).

Specimens were observed and images processed with the same devices and software as used for FISH experiments.

***Saccharomyces cerevisiae* heterologous expression system**

The coding sequence of *sufC* was amplified by PCR amplified from *Monocercomonoides* cDNA using PrimeStar MAX DNA polymerase (Clontech, R045A) and the following primers (5`CCATCCATACTCTAGAATGCAAACTCAAAAGCCCC3' and 5`CGGTATCGATAAGCTTAATCTTCACAACTCCCTCTG3'). The PCR products were cloned by in-fusion cloning (In-Fusion® HD Cloning Kit (Clontech, 638909) into the pUG35 vector with C-terminal GFP.

The wild type *S. cerevisiae* strain YPH499 (ATCC number: 204679) was used for transformation. Yeasts were grown on plates with YPD agar medium [S44] at 30°C. Transformation of the yeasts with 2 µg of plasmid DNA was performed using the previously described LiAc/SS-DNA/PEG method [S52]. Transformants were selected on synthetic drop-out medium without uracil [S44] at 30°C. Expression of *Monocercomonoides* GFP-tagged proteins in yeasts was analysed 3 days after transformation. Mitochondria were labelled with MitoTracker Red CMXRos dye (Life Technologies, M7512).

**Supplemental References**


S1. Diamond, L. S. (1982). A new liquid medium for xenic cultivation of Entamoeba histolytica and other lumen-dwelling protozoa. J. Parasitol. *68*, 958–959.

S2. Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics *23*, 1061–1067.

S3. Parra, G., Bradnam, K., Ning, Z., Keane, T., and Korf, I. (2009). Assessing the gene space in draft genomes. Nucleic Acids Res. *37*, 289–297.

S4. Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C. M., Besteiro, S., et al. (2007). Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. Science *315*, 207–212.

S5. Eddy, S. R. (2011). Accelerated Profile HMM Searches. PLoS Comput. Biol. *7*, e1002195.

S6. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V, and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *19*, 3210–3212.

S7. Lecompte, O., Ripp, R., Thierry, J. C., Moras, D., and Poch, O. (2002). Comparative analysis of ribosomal proteins in complete genomes: An example of reductive evolution at the domain scale. Nucleic Acids Res. *30*, 5382–5390.

S8. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

S9. Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics *7*, 62.

S10. Haas, B. J. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. *31*, 5654–5666.

S11. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

S12. Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics *19*, ii215–ii225.

S13. Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., and Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. *9*, R7.

S14. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

S15. Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.
S16. Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
S17. Wu, T. D., and Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics *21*, 1859–1875.
S18. Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29–W37.
S19. Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Res. *40*, D290–D301.
S20. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. *35*, W182–W185.
S21. Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. *42*, D199–D205.
S22. Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res. *25*, 0955–0964.
S23. Smit AFA, Hubley R, and Green P RepeatMasker Open-3.0 1996-2010. Available at: http://www.repeatmasker.org.
S24. Jurka, J., Kapitonov, V. V, Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. *110*, 462–467.
S25. Price, A. L., Jones, N. C., and Pevzner, P. a (2005). De novo identification of repeat families in large genomes. Bioinformatics *21 Suppl 1*, i351–i358.
S26. Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674–3676.
S27. Söding, J., Biegert, A., and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. *33*, W244–W248.
S28. Smith, A. C., Blackshaw, J. a, and Robinson, A. J. (2012). MitoMiner: a data warehouse for mitochondrial proteomics data. Nucleic Acids Res. *40*, D1160–D1167.
S29. Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. *2*, 953–971.
S30. Claros, M. G., and Vincens, P. (1996). Computational Method to Predict Mitochondrially Imported Proteins and their Targeting Sequences. Eur. J. Biochem. *241*, 779–786.
S31. Zdobnov, E. M., and Apweiler, R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics *17*, 847–848.
S32. Jedelský, P. L., Doležal, P., Rada, P., Pyrih, J., Šmíd, O., Hrdý, I., Šedinová, M., Marcinčiková, M., Voleman, L., Perry, A. J., et al. (2011). The Minimal Proteome in the Reduced Mitochondrion of the Parasitic Protist Giardia intestinalis. PLoS One *6*, e17285.
S33. Mi-ichi, F., Abu Yousuf, M., Nakada-Tsukui, K., and Nozaki, T. (2009). Mitosomes in Entamoeba histolytica contain a sulfate activation pathway. Proc. Natl. Acad. Sci. U. S. A. *106*, 21731–21736.
S34. Martincová, E., Voleman, L., Pyrih, J., Žárský, V., Vondráčková, P., Kolísko, M., Tachezy, J., and Doležal, P. (2015). Probing the Biology of Giardia intestinalis Mitosomes Using In Vivo Enzymatic Tagging. Mol. Cell. Biol. *35*, 2864–2874
S35. Schneider, R. E., Brown, M. T., Shiflett, A. M., Dyall, S. D., Hayes, R. D., Xie, Y., Loo, J. a, and Johnson, P. J. (2011). The Trichomonas vaginalis hydrogenosome proteome is highly reduced relative to mitochondria, yet complex compared with mitosomes. Int. J. Parasitol. *41*, 1421–1434.
S36. Denoeud, F., Roussel, M., Noel, B., Wawrzyniak, I., Da Silva, C., Diogon, M., Viscogliosi, E., Brochier-Armanet, C., Couloux, A., Poulain, J., et al. (2011). Genome sequence of the stramenopile Blastocystis, a human anaerobic parasite. Genome Biol. *12*, R29.
S37. Borgese, N., Colombo, S., and Pedrazzini, E. (2003). The tale of tail-anchored proteins: coming from the cytosol and looking for a membrane. J. Cell Biol. *161*, 1013–1019.
S38. Borgese, N., Brambillasca, S., and Colombo, S. (2007). How tails guide tail-anchored proteins to their destinations. Curr. Opin. Cell Biol. *19*, 368–375.
S39. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.
S40. Santos, H. J., Imai, K., Makiuchi, T., Tomii, K., Horton, P., Nozawa, A., Ibrahim, M., Tozawa, Y., and Nozaki, T. (2015). A Novel Mitosomal β-Barrel Outer Membrane Protein in Entamoeba. Sci. Rep. *5*, 8545.
S41. Kutik, S., Stojanovski, D., Becker, L., Becker, T., Meinecke, M., Krüger, V., Prinz, C., Meisinger, C., Guiard, B., Wagner, R., et al. (2008). Dissecting membrane insertion of mitochondrial beta-barrel proteins. Cell *132*, 1011–1024.

S42. Imai, K., Fujita, N., Gromiha, M. M., and Horton, P. (2011). Eukaryote-wide sequence analysis of mitochondrial β-barrel outer membrane proteins. BMC Genomics *12*, 79.
S43. Tusnády, G. E., Dosztányi, Z., and Simon, I. (2005). PDB_TM: Selection and membrane localization of transmembrane proteins in the protein data bank. Nucleic Acids Res. *33*, 275–278.
S44. Zubáčová, Z., Novák, L., Bublíková, J., Vacek, V., Fousek, J., Rídl, J., Tachezy, J., Doležal, P., Vlček, C., and Hampl, V. (2013). The mitochondrion-like organelle of Trimastix pyriformis contains the complete glycine cleavage system. PLoS One *8*, e55417.
S45. Kamikawa, R., Kolisko, M., Nishimura, Y., Yabuki, A., Brown, M. W., Ishikawa, S. A., Ishida, K., Roger, A. J., Hashimoto, T., and Inagaki, Y. (2014). Gene content evolution in Discobid mitochondria deduced from the phylogenetic position and complete mitochondrial genome of Tsukubamonas globosa. Genome Biol. Evol. *6*, 306–315.
S46. Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.
S47. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.
S48. Kück, P., and Meusemann, K. (2010). FASconCAT: Convenient handling of data matrices. Mol. Phylogenet. Evol. *56*, 1115–1118.
S49. Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286–2288.
S50. Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst. Biol. *62*, 611–615.
S51. Sutak, R., Dolezal, P., Fiumera, H. L., Hrdy, I., Dancis, A., Delgadillo-Correa, M., Johnson, P. J., Müller, M., and Tachezy, J. (2004). Mitochondrial-type assembly of FeS centers in the hydrogenosomes of the amitochondriate eukaryote Trichomonas vaginalis. Proc. Natl. Acad. Sci. U. S. A. *101*, 10368–10373.
S52. Gietz, R. D., Schiestl, R. H., Willems, A. R., and Woods, R. A. (1995). Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. Yeast *11*, 355–360.