

Univerzita Karlova

Přírodovědecká fakulta

Doktorský studijní program: Parazitologie



Mgr. Vojtěch Žárský

*Analýza genomu volně žijící améby *Mastigamoeba balamuthi* a porovnání s patogenní amébou *Entamoeba histolytica**

*Analysis of the genome of a free-living amoeba *Mastigamoeba balamuthi* and its comparison with pathogenic *Entamoeba histolytica**

Disertační práce

Vedoucí práce: prof. RNDr. Jan Tachezy, Ph.D

Praha, 2020

Charles University

Faculty of Science

Doctoral study program: Parasitology



Mgr. Vojtěch Žárský

Analysis of the genome of a free-living amoeba *Mastigamoeba balamuthi* and its comparison with pathogenic *Entamoeba histolytica*

Analýza genomu volně žijící améby *Mastigamoeba balamuthi* a porovnání s patogenní amébou *Entamoeba histolytica*

Doctoral thesis

Supervisor: prof. RNDr. Jan Tachezy, Ph.D

Prague, 2020

Declaration of the author

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

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

Mgr. Vojtěch Žárský

Declaration of the thesis supervisor

Data presented in this thesis resulted from team collaboration at the Laboratory of biochemical and molecular parasitology and during projects with our partners. I declare that the involvement of Mgr. Vojtěch Žárský in this work was substantial and that he contributed significantly to obtain the results.

prof. RNDr. Jan Tachezy, Ph.D

Acknowledgments

Many thanks to my supervisor Jan Tachezy for his unwavering support, to my lab fellows and all other people who supported me during my PhD study. Special thanks to my wife Alina for her encouragement and help.

Table of Contents

Abstract.....	6
Abstrakt.....	7
1 Introduction.....	8
1.1 Free-living and parasitic eukaryotes.....	8
1.2 The genesis of the group Archamoebae.....	11
1.3 Mitochondrion-related organelles of Archamoebae and evolutionary adaptation to anaerobiosis...12	
1.4 Peroxisomes of Archamoebae.....	15
1.5 Diversity of free-living and parasitic Archamoebae.....	17
1.6 Life cycle and pathology of <i>Entamoeba histolytica</i>	19
2 Aims.....	22
3 List of publications and author contribution.....	23
4 Summary.....	25
5 References.....	29

Abstract

Examination and comparison of organisms have been tremendously important for the study of life's history on earth. The progress of our understanding of the genetic basis of heredity and the recent boom of sequencing technologies allows us to continue in this exciting field of research from the perspective of genes and genomes. In this work, I focus on the study of an anaerobic amoeba *Mastigamoeba balamuthi*, which is related to an important human pathogen *Entamoeba histolytica*. Comparative analysis allows us to draw some conclusions about the nature of the common ancestor of *Mastigamoeba* and *E. histolytica*, how it adapted to the anaerobic lifestyle, and about the way the *Entamoeba* lineage evolved to become a successful parasite. Surprisingly we also noticed that besides hydrogenosomes (hydrogen-producing organelles related to mitochondria), *M. balamuthi* also harbors peroxisomes – organelles thought to be absent in anaerobic organisms. This finding motivated us to inquire more about peroxisomes in other eukaryotic lineages. We found out that there is a reduced set of peroxisomal markers in certain *Entamoeba* species. Moreover, we showed that peroxisomes were independently lost in several lineages of parasitic helminths and a free-living tunicate *Oikopleura dioica*.

Abstrakt

Výzkum a srovnávání organismů byly nesmírně významné pro studium historie života na Zemi. Pokrok v porozumění genetickému základu dědičnosti a nebývalý rozvoj sekvenačních technologií nám umožňuje pokračovat na tomto vzrušujícím poli výzkumu z perspektivy genů a genomů. V této práci se zaměřuji na studium anaerobní améby *Mastigamoeba balamuthi*, která je příbuzná významnému lidskému parazitu *Entamoeba histolytica*. Srovnávací analýza nám umožnila učinit některé závěry o vlastnostech společného předka *M. balamuthi* a *E. histolytica*, jak se přizpůsobil životu bez kyslíku, a jak se linie *Entamoeba* vyvinula v úspěšného parazita. Zjistili jsme přitom překvapivě, že vedle hydrogenosomů (odvozené mitochondrie produkující vodík) má *M. balamuthi* také peroxisomy - organely o kterých se soudilo, že je anaerobní organismy ztratily. Tento objev nás podnítil ke studiu distribuce peroxisomů u dalších eukaryotických linií. Zjistili jsme, že i u některých druhů *Entamoeba* nalézáme redukovaný soubor proteinů typických pro peroxisomy. Dále jsme ukázali, že peroxisomy byly nezávisle ztraceny u některých skupin parazitických červů a u volně žijícího pláštěnce *Oikopleura dioica*.

1 Introduction

1.1 Free-living and parasitic eukaryotes

Organisms live within an intricate network of interactions with abiotic environment and with each other, with a variable degree of dependency and directionality. These range from interactions that span vast time and space (think about photosynthetic organisms of the past) to very immediate proximal interactions, for example, a prey being eaten by a predator. When the individuals of different species live in a tight long term relationship, we call such interaction symbiosis. For example, a wood degrading protist *Trichonympha* inhabiting termite gut provides the host with a critical ability to use wood as an energy source. However, if one player in symbiosis benefits at the expense of the other, the interaction is described as parasitism.

Parasites are ever-present in all lineages of life, including eukaryotes, that is organisms with a unique cell structure called the nucleus and other complex features. Modern phylogeny, which is primarily based on a comparison of the genomes of extant organisms, has helped us to resolve evolutionary relationships between many eukaryotic lineages (Burki et al. 2020). It turns out that a vast hidden diversity of eukaryotes is represented by one-celled eukaryotes – protists, and that parasitism is a popular life strategy that has been adopted numerous times across the diversity of eukaryotes (Figure 1).

Poulin and Randhawa define six optimal parasitic strategies towards which all the eukaryotic parasitic lineages evolve (Poulin and Randhawa 2015). Indeed we observe many convergent evolutionary patterns when looking at parasites of different origins, for example, loss of certain metabolic pathways, genomic streamlining, surface modification. Comparative studies of successful parasitic groups and their relatives have brought many exciting insights into the evolution of parasitism.

Certain specialized structures or mechanisms that are essential for parasite survival can be traced back to the free-living ancestor. We call such a shift in the function of a trait in the evolution an exaptation (formerly pre-adaptation) (S. J. Gould and Vrba 1982). For example, the apical complex that many apicomplexans use to establish contact with host cells and secrete effector molecules, as well as a specific kind of asexual fission called schizogony, were both found in free-living lineages related to apicomplexans (Janouškovec et al. 2015). Similarly, a study comparing a free-living kinetoplastid *Bodo saltans* with parasitic trypanosomatids revealed that *B. saltans* codes for surface glycoproteins that the authors call Bodonids, that share structural similarity to and may have the same ancestry as some

trypanosomatid cell surface proteins (*Trypanosoma cruzi* Dispersed Gene Family-1 and *Leishmania* Promastigote Surface Antigen) (Jackson et al. 2016).

So it turns out that the free-living ancestors of successful parasitic groups were possibly well equipped to acquire parasitic lifestyle and that we can observe radiation of closely related parasitic lineages of independent origin that can be strikingly similar in appearance. For example, it was recently shown that two parasitic lineages formerly classified as Apicomplexa (*Piridium* and *Platyproteum*) are both of an independent origin (Mathur et al. 2019). Similarly, phylogenetic analyses of kinetoplastids established that several lineages related to the parasitic trypanosomatids (*Trypanoplasma*, *Azumiobodo*, *Ichthyobodo*) have independently acquired parasitic lifestyle (Lukeš et al. 2014; Yazaki et al. 2017).

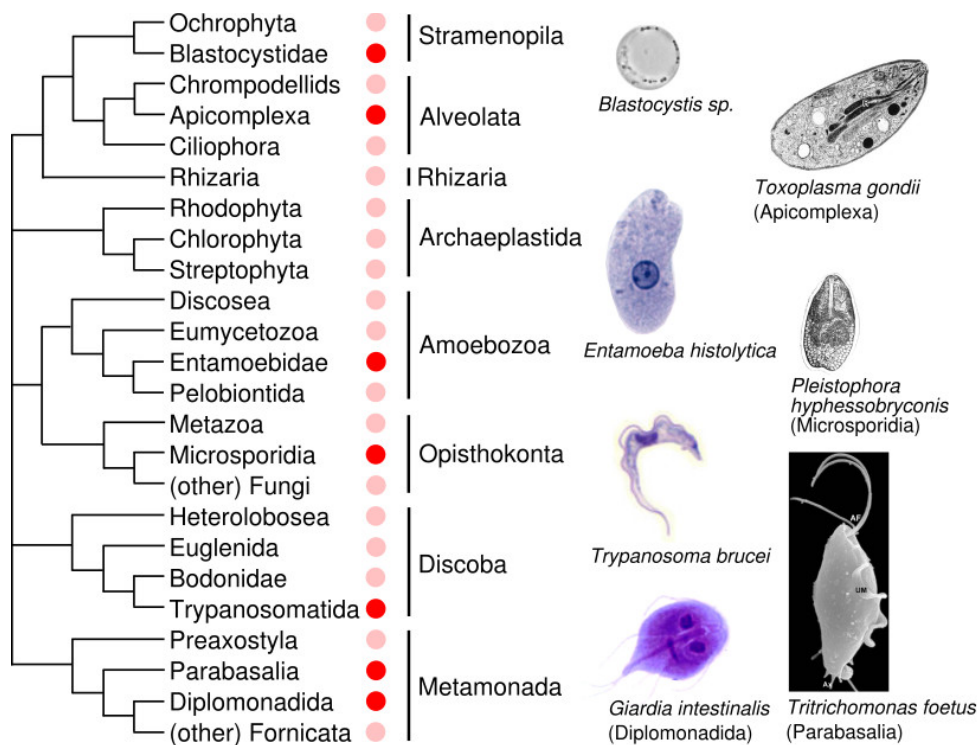


Figure 1. A simplified overview of eukaryotic phylogeny. Obligate parasitic groups with important human pathogens are labeled with a red circle. Groups that contain parasitic taxa are labeled with a pink circle. Photographs of parasitic protists are shown: *Blastocystis sp.* (Tan 2008), *Toxoplasma gondii* (Shaw and Tilney 1999), *Entamoeba histolytica* (S.J. Upton, www.k-state.edu), *Pleistophora hypheobryconis* (Lom and Corliss 1967), *Trypanosoma brucei* (www.cdc.gov), *Tritrichomonas foetus* (de Andrade Rosa, de Souza, and Benchimol 2013), *Giardia intestinalis* (www.nih.gov.jp). The eukaryotic phylogeny is based on (Burki et al. 2020).

It is usually assumed that acquisition of an obligate parasitic lifestyle is a one-way ticket because such transition requires a high level of specialization, and it is often accompanied by losses of essential synthetic pathways. Although probably rare, counterexamples do exist. There are several secondarily free-living trichomonads (*Monotrichomonas carabina*, *Ditrichomonas honigbergii*, *Honigbergiella sp.*, *Tetratrichomonas undula*, and *Pseudotrichomonas keilini*) (Yubuki et al. 2010). Also, a phylogenomics study shows that a free-living diplomonad *Trepomonas sp.* evolved from a parasitic ancestor, and this has been facilitated by the acquisition of essential genes (e.g., ribonucleotide reductase for the synthesis of deoxyribonucleotides from ribonucleotides) via lateral transfer from bacteria (Morin 2000; Xu et al. 2016).

1.2 The genesis of the group Archamoebae

A group of anaerobic amoebae that comprises parasitic *Entamoeba* and free-living *Mastigamoeba*, commonly called Archamoebae, has been initially classified by Thomas Cavalier-Smith into a group of amoeboid organisms called Sarcodina (Figure 2A) (Cavalier-Smith 1981). Later in the 80s, Cavalier-Smith proposed a hypothesis stating that some of the anaerobic eukaryotes belong to a (paraphyletic) group that has been called Archezoa and that this group is primarily amitochondriate, i. e. these lineages have split from other eukaryotes before the acquisition of the mitochondrion (Cavalier-Smith 1983). One of the proposed amitochondriate lineages was Archamoebae (Figure 2B).

Although the Archezoa hypothesis was later proven to be incorrect, it boosted interest in these understudied organisms, their evolution, and cell biology. Several lines of evidence appeared contesting the Archezoa hypothesis, and Archamoebae have played an essential part in this quest: As in other proposed Archezoa lineages, organelles of mitochondrial origin have been discovered; first mitosomes in *Entamoeba histolytica* (Tovar, Fischer, and Clark 1999) and later hydrogenosomes in *Mastigamoeba balamuthi* (Gill et al. 2007).

Furthermore, Archamoebae have been robustly placed within a large eukaryotic supergroup called Amoebozoa, while some of the other former Sarcodina lineages have been reassigned to other eukaryotic supergroups including Rhizaria, Heterolobosea, and Metazoa. It has also been shown that Archamoebae is possibly a sister group of Eumycetozoa, which are aerobic slime molds represented by e.g., *Dictyostelium discoideum*, an important model organism (Figure 2C) (Baptiste et al. 2002; Kang et al. 2017).

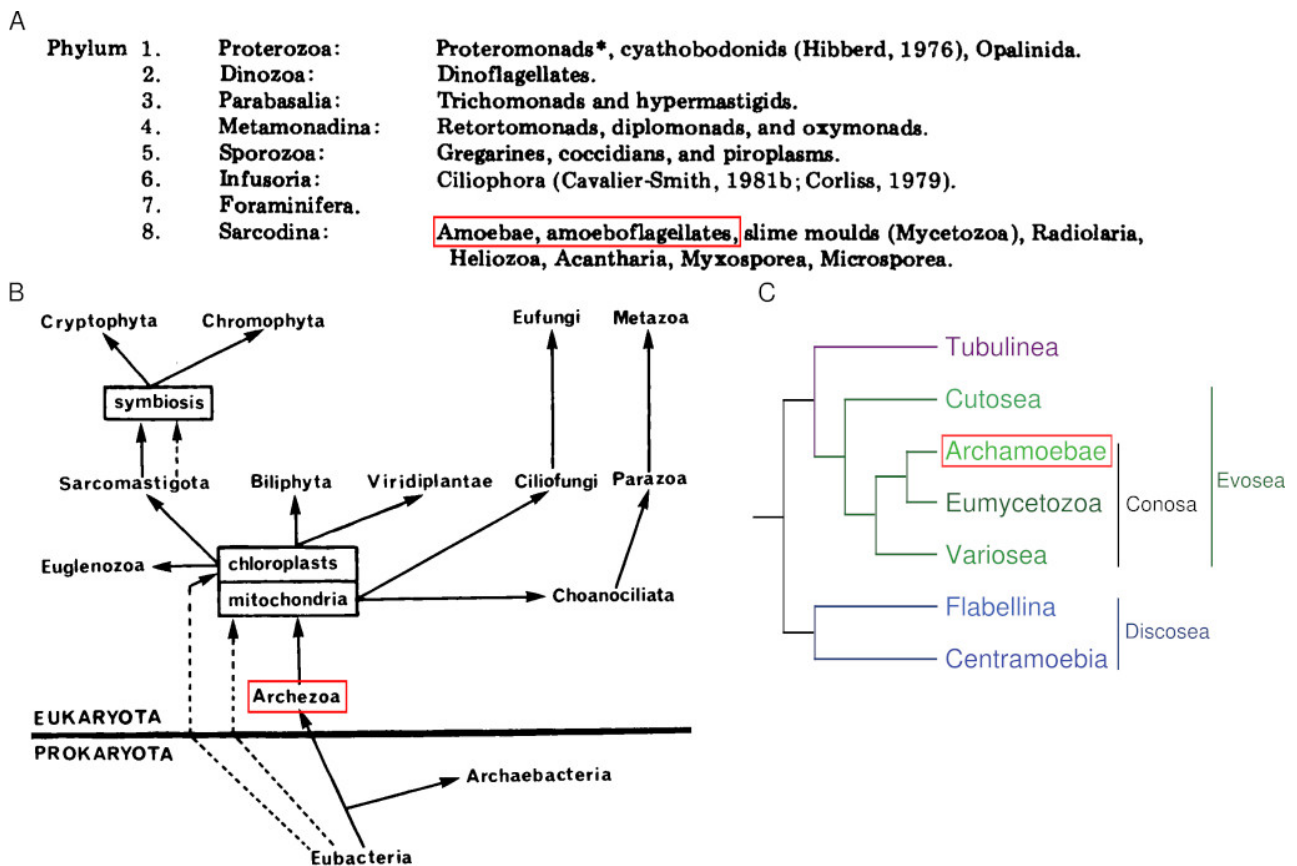


Figure 2. The genesis of the placement of the Archamoebae taxon. Groups that contain Archamoebae are marked by a red box. (A) Former group Sarcodina associated a variety of amoeboid organisms (Cavalier-Smith 1981). (B) Former group Archezoa was proposed to contain primarily amitochondriate eukaryotes (Parabasalia, Diplomonadida, Microsporidia, Archamoebae) (Cavalier-Smith 1983). (C) Multigene phylogenetic studies robustly place Archamoebae within the Amoebozoa group as a sister to Eumycetozoa (Bapteste et al. 2002; Kang et al. 2017).

1.3 Mitochondrion-related organelles of Archamoebae and evolutionary adaptation to anaerobiosis

It is generally accepted that eukaryotic mitochondria arose from endosymbiosis of an alphaproteobacterium and that the common ancestor of eukaryotes was capable of oxidative phosphorylation (Roger, Muñoz-Gómez, and Kamikawa 2017). However, scattered among the whole tree of eukaryotes, we find lineages adapted to live without oxygen (Stairs, Leger, and Roger 2015). This is foremost reflected by the change in the metabolism and structure of mitochondria. An

established term Mitochondrion-Related Organelles (MRO) describes all diversity of mitochondria of both aerobic and anaerobic organisms. Comparing MROs from different organisms and lineages, we observe a mosaic distribution of components and functions. However, we see many strikingly convergent tendencies in the evolution of the MROs, not only in similar reductions but also in the acquisition of specific genes possibly by lateral gene transfer (LGT) from other anaerobic eukaryotes or anaerobic bacteria.

MROs were previously divided into five categories (Müller et al. 2012): (i) Aerobic mitochondria (found in e.g., mammals, yeast, plants) have a genome and use electron transport chain (ETC) with oxygen as a terminal electron acceptor to produce energy. (ii) Anaerobic mitochondria (e.g., *Ascaris*, *Euglena*) have a genome and respire aerobically under high O₂ conditions but switch to an alternative terminal acceptor of electrons (e.g., fumarate) under low O₂ conditions. (iii) Hydrogen producing mitochondria (e.g., *Blastocystis*, *Nyctotherus*) have a genome but lack the complex V (ATP synthase) and rather synthesize ATP by anaerobic substrate-level phosphorylation producing H₂. (iv) Hydrogenosomes (e.g., *Trichomonas*, *Stygiella*, *Neocallimastix*, *M. balamuthi*) lack the organellar genome and ETC, and ATP is synthesized by substrate-level phosphorylation via anaerobic pyruvate metabolism and H₂ synthesis. (v) Mitosomes (e.g., *Giardia*, *Cryptosporidium*, *Entamoeba*) are the most reduced forms of MROs that neither produce ATP nor hydrogen. These are not definite categories, and there are MROs with mixed characteristics (Gawryluk et al. 2016). The most extreme case of MRO reduction is its complete loss that has been only recently described in an oxymonad *Monocercomonoides* (Karnkowska et al. 2016).

Archamoebae are obligate anaerobic amoebozoans, and this is also reflected by the presence of anaerobic MROs. Mitosomes and hydrogenosomes were described in *E. histolytica* (Tovar, Fischer, and Clark 1999) and *M. balamuthi* (Gill et al. 2007), respectively. Both organelles received considerable attention, and their known components are summarized in Figure 3. We can see a typical hydrogenosomal anaerobic metabolism in hydrogenosomes of *M. balamuthi*: Pyruvate:ferredoxin oxidoreductase oxidizes pyruvate to acetyl-CoA and hydrogenase transfers electrons from reduced ferredoxin to protons, forming molecular hydrogen. Substrate-level phosphorylation is then accomplished by the acetyl-CoA synthetase that forms acetate from acetyl-CoA, creating one ATP (Nývtová et al. 2015). Furthermore, Archamoebae acquired two unique functions not typical for other MROs: (i) The typical mitochondrial ISC system for iron-sulfur cluster synthesis has been replaced by a bacterial NIF system (composed of a cysteine desulfurase and a scaffold for iron-sulfur cluster assembly). In *M. balamuthi* two NIF systems are present - one cytosolic and another hydrogenosomal (Nývtová et al. 2013), while in *E. histolytica* the single NIF system has been shown to be localized in

the cytosol (Ali et al. 2004; Mi-ichi et al. 2009; Dolezal et al. 2010), although some results suggested its presence also in the mitosome (Maralikova et al. 2010). (ii) In *E. histolytica*, it has been shown that the sulfur activation pathway localizes to mitosomes and that it is essential for the synthesis of cholesteryl sulfate (Mi-ichi et al. 2015). Components of the sulfur activation pathway were also found in the hydrogenosomes of *M. balamuthi* (Nývltová et al. 2015).

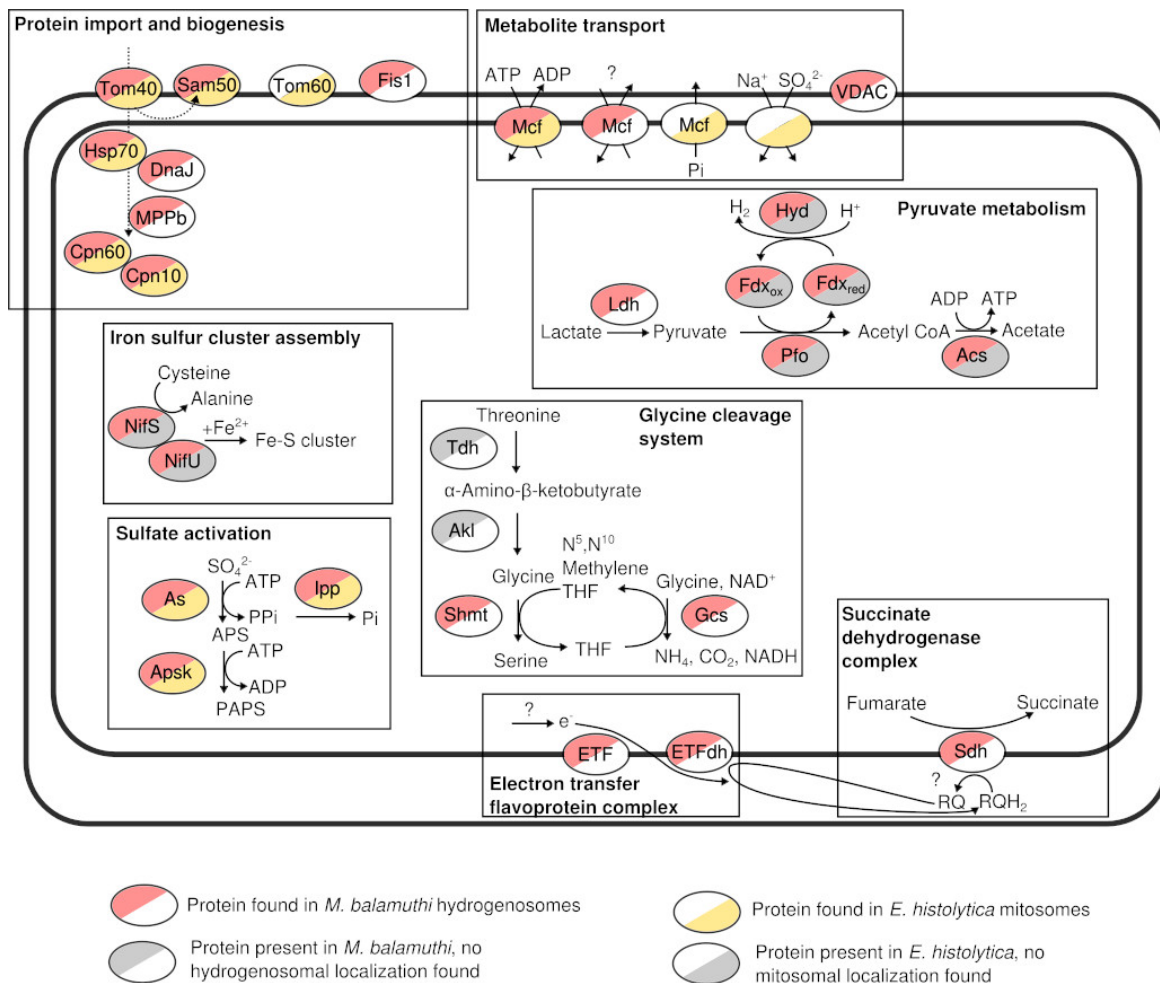


Figure 3. Comparison of *M. balamuthi* hydrogenosomes and *E. histolytica* mitosomes. Mcf, mitochondrial carrier family; Ldh, D-lactate dehydrogenase; Hyd, hydrogenase; Fdx, ferredoxin; PFO, Pyruvate:ferredoxin oxidoreductase; ACS, acetyl-CoA synthetase; AS, ATP sulfurylase; Apsk, adenosine 5'-phosphosulfate kinase; Ipp, inorganic pyrophosphatase; Shmt, serine hydroxymethyltransferase; GCS, glycine cleavage system; Sdh, succinate dehydrogenase; ETF, electron transfer flavoprotein; ETFdh, electron transferring flavoprotein dehydrogenase; RQ, rholoquinone.

Although living in low oxygen conditions, anaerobic eukaryotes still need to deal with oxidative and nitrosative stress. As with most other anaerobes, *E. histolytica* lacks glutathione metabolism and glutathione-dependent pathways (Fahey et al. 1984). Instead, thioredoxin system, superoxide dismutase, peroxiredoxins, and enzymes typical for anaerobic prokaryotes (flavodiiron protein, iron-sulfur flavoprotein, and rubrerythrin) aid detoxification of reactive oxygen species and reactive nitrogen species in *E. histolytica* (Loftus et al. 2005; Jeelani and Nozaki 2016). Interestingly an enzyme for the synthesis of rhodoquinone, a low electron potential quinol, has been found in *M. balamuthi* (Stairs et al. 2018).

1.4 Peroxisomes of Archamoebae

Peroxisomes were described in rat liver and a ciliate *Tetrahymena pyriformis* by Christian de Duve in the 60s as organelles that contain oxidases that reduce oxygen to hydrogen peroxide and large amounts of catalase which reduces hydrogen peroxide to water (De Duve 1969). Due to the presence of peroxisomes in major lineages of eukaryotes de Duve realized that peroxisomes are an ancient eukaryotic organelle. During that time, an unusual hydrogen-producing organelle of trichomonads was characterized, and it has been suggested that it may be specialized peroxisome (Cornford and Outka 1970). However, this has been refused by others (Lindmark, Müller, and Shio 1975), and finally, it has been shown that the hydrogenosomes are organelles related to mitochondria (Hrady et al. 2004). This illustrates certain level of ambiguity when it comes to peroxisomes, their evolutionary origins, and relationship to mitochondria and the endomembrane system, which remains until present-day (Gabaldón et al. 2006; Schlüter et al. 2006; Speijer 2017; Sugiura et al. 2017). With the advancement of genetic and biochemistry tools, molecular mechanisms of protein import into mitochondria and peroxisomes have been described in detail, and distinct highly specialized protein machineries for biogenesis and maintenance of each organelle were described (Pfanner and Neupert 1990; S. j. Gould et al. 1990). Modes of protein import and organelle biogenesis are unambiguous defining characteristics of each of the eukaryotic organelles. Peroxins, also called PEX proteins, are components of the protein import and division machinery of peroxisomes. While the functions of peroxisomes are very diverse (Pieuchot and Jedd 2012), a set of about 14 PEX proteins is highly conserved in peroxisomes of all eukaryotic lineages, and these PEX proteins are the most reliable markers to evaluate the presence of peroxisomes (Gabaldón et al. 2006; Schlüter et al. 2006).

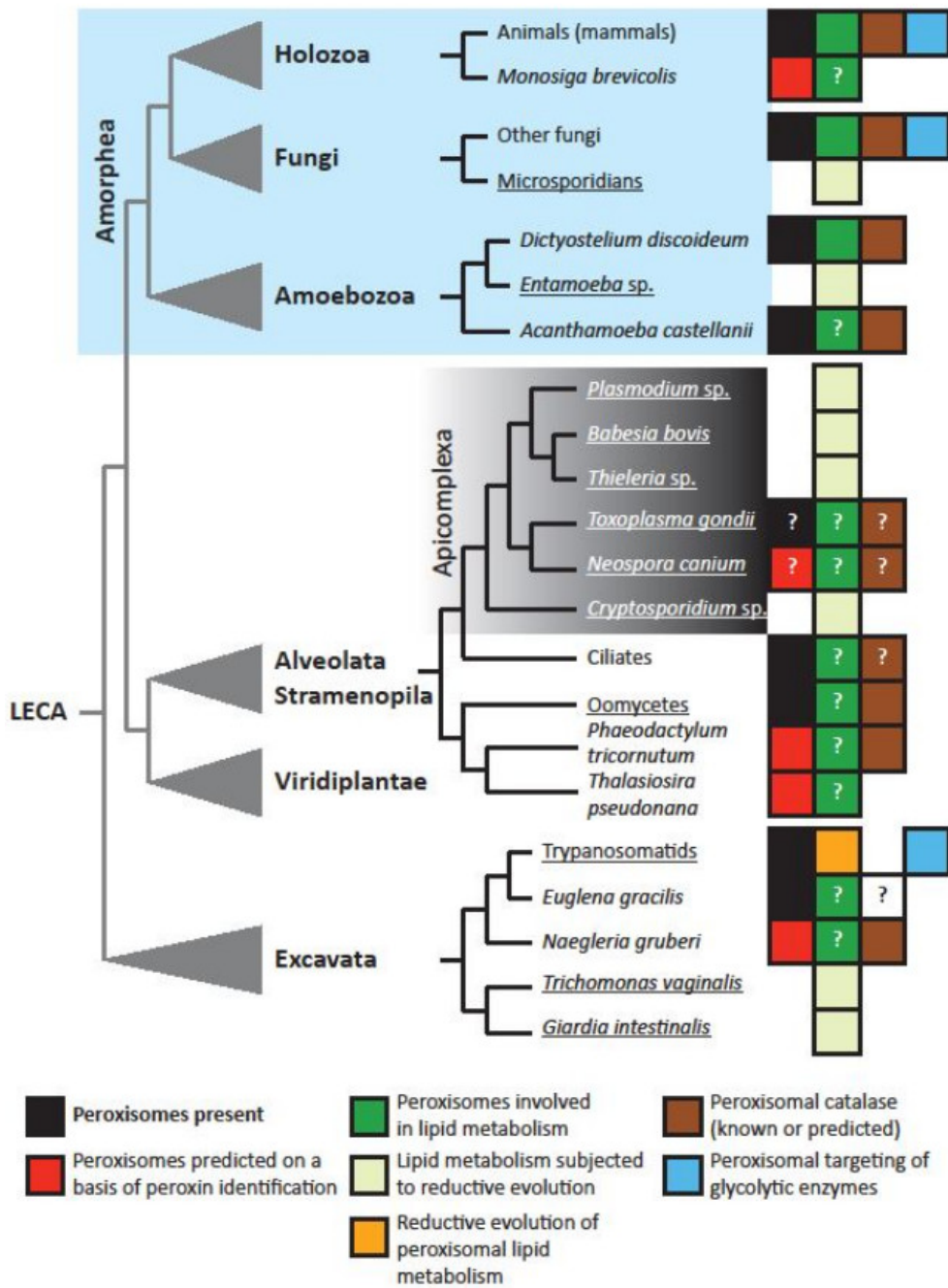


Figure 4. Functional diversity and distribution of peroxisomes in eukaryotes. The predicted presence or absence of peroxisomes among selected eukaryotes is shown. Known or predicted function of the organelles in lipid metabolism, ROS detoxification, and glycolysis are also summarized. White fields represent the absence of certain characteristics. Parasitic species are underlined. Picture reproduced from (Gabaldón, Ginger, and Michels 2016).

Advances in the sequencing of eukaryotic genomes allowed to perform large-scale comparisons of peroxisomal markers throughout eukaryotic diversity. It was observed that: peroxisomes are present in most lineages of aerobic eukaryotes except for certain apicomplexan parasites (e.g., *Plasmodium*, *Theileria*) and that peroxisomes seem to be absent in anaerobic eukaryotes including Archamoebae (Gabaldón 2010; Gabaldón, Ginger, and Michels 2016; Ludewig-Klingner et al. 2018) (Figure 4).

1.5 Diversity of free-living and parasitic Archamoebae

Archamoebae is a group of anaerobic amoeboid organisms with several genera: *Mastigamoeba* (Figure 5A-B) and *Mastigella* (Figure 5C) are free-living amoeboflagellates with a single flagellum. *Rhizomastix* (Figure 5D) are both free-living, and endobiotic amoeboflagellates with a single flagellum. *Pelomyxa* (Figure 5E) are giant free-living amoebae with many non-motile flagella. *Entamoeba* (Figure 5F), *Endolimax* (Figure 5G) and *Iodamoeba* are endobionts inhabiting mostly animal intestines that have entirely lost flagellum. Prokaryotic endosymbionts are common in Archamoebae.

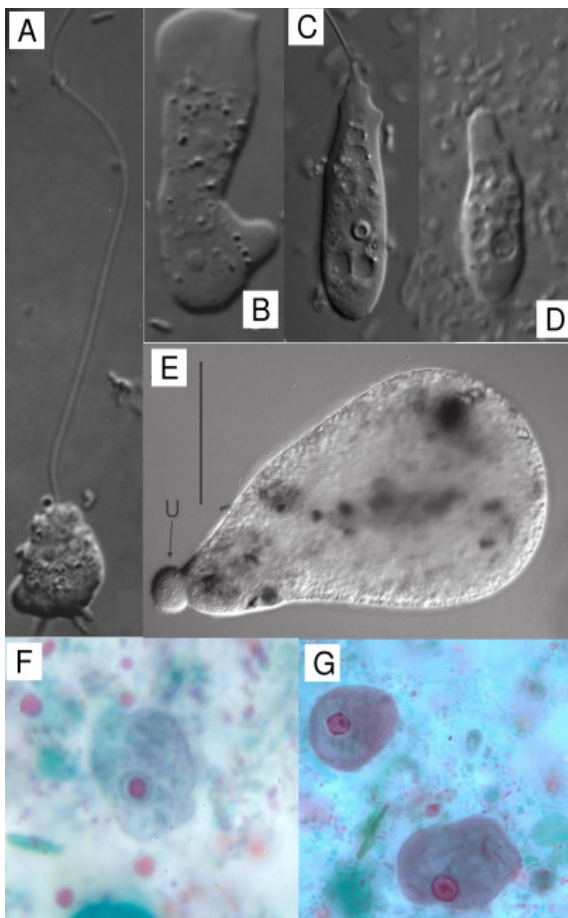


Figure 5. Selected representatives of Archamoebae genera. (A) *Mastigamoeba balamuthi* amoeboflagellate with an anterior flagellum and posterior pseudopodia (Pánek et al. 2016). (B) *M. balamuthi* amoeba (Pánek et al. 2016). (C) *Mastigella eilhardi* amoeboflagellate (Zadrobílková, Walker, and Čepička 2015). (D) *Rhizomastix bicornata* amoeboflagellate (Zadrobílková et al. 2016). (E) *Pelomyxa palustris* (Ptáčková et al. 2013). (F) *Endolimax nana* trichrome-stained trophozoite (www.cdc.gov). (G) *Entamoeba coli* trichrome-stained trophozoites (www.cdc.gov). DIC (A-E) and bright-field (F-G).

Efforts to unravel the phylogeny of Archamoebae brought many exciting results (Figure 6). Contrary to earlier phylogenetic studies based on SSU rRNA (Ptáčková et al. 2013; Zadrobílková, Walker, and Čepička 2015), a multigene analysis revealed that the endobiotic/parasitic Entamoebida form a basal lineage of Archamoebae (Pánek et al. 2016). Parasitic or endobiotic lifestyle evolved independently at least three times (*Entamoebidae*, *Endolimax+Iodamoebae*, and within *Rhizomastix*). Striking is the similarity in morphology and lifestyle of *Entamoebidae*, and *Endolimax*, as they evolved independently a very similar life cycle: infectious cyst is ingested, trophozoite excysts in the small intestine and travels to the large intestine where it multiplies and produces cysts. The common ancestor of Archamoebae was possibly a free-living anaerobic amoeboid flagellate with a single anterior cilium.

Mastigamoeba balamuthi (originally named *Phreatamoeba balamuthi*) has been isolated from a well in the Gambia, and an axenic culture has been established (Chavez, Balamuth, and Gong 1986). Three phenotypes were observed: (i) An amoeba that is typically multinucleate varying from 10 to 160 μm in length (Figure 5A), (ii) a smaller amoeboflagellate with a single anterior flagellum and a body length from 6 to 50 μm (Figure 5B), (iii) a cyst surrounded by a resistant wall with diameter from 9 to 18 μm . It is of particular interest because it is one of the closest free-living relatives of parasitic *Entamoeba* while it possibly retains ancestral characteristics of the group Archamoebae (Pánek et al. 2016).

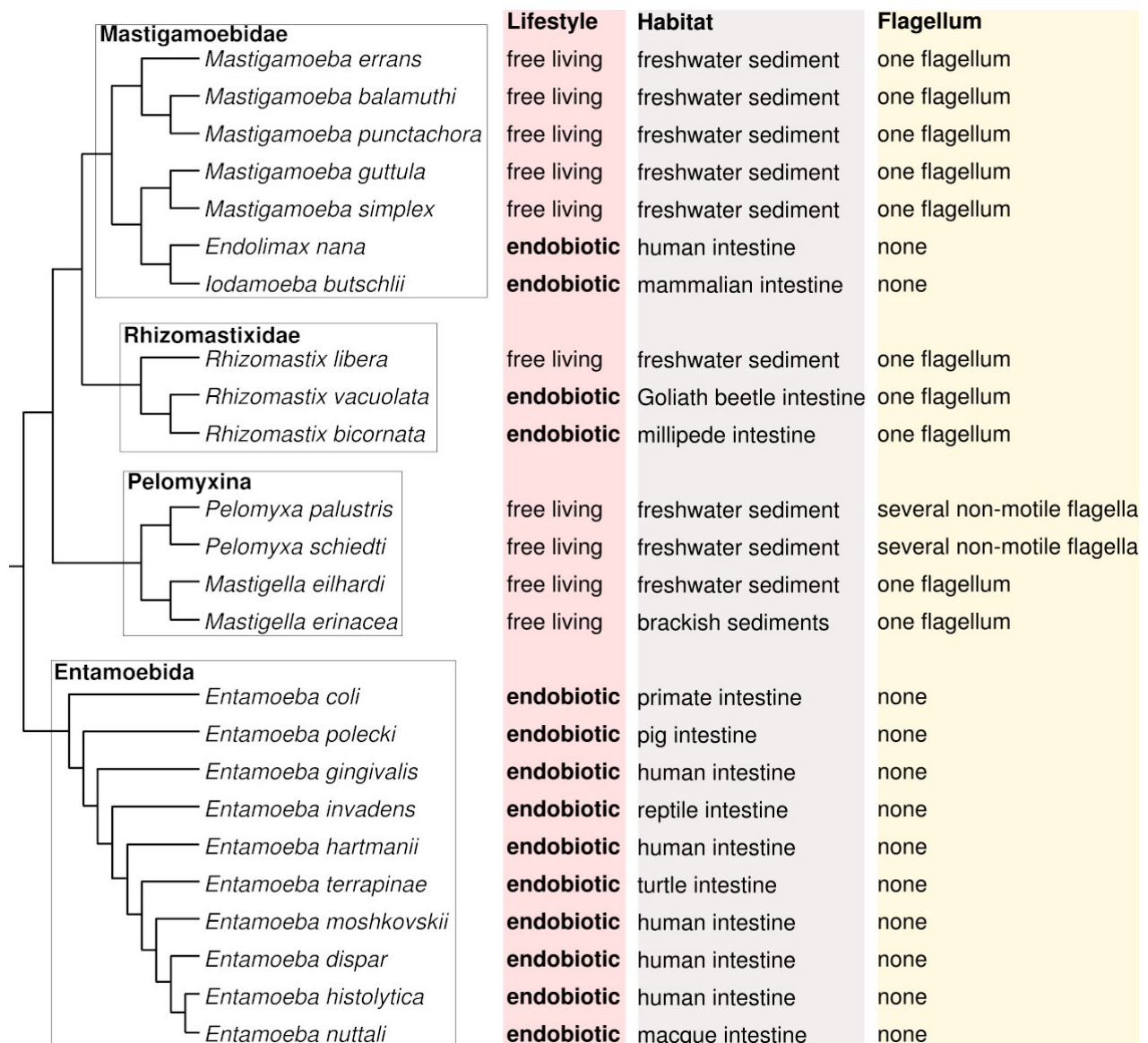


Figure 6. Evolution and diversity of selected Archamoebae. The relationships between major groups are based on (Pánek et al. 2016). The phylogeny of Entamoebida is based on (Stensvold et al. 2011).

1.6 Life cycle and pathology of *Entamoeba histolytica*

The human intestine is inhabited not only by bacteria but also by a number of protists (e.g., *Blastocystis*, *Endolimax*, *Iodamoeba*, *Entamoeba*, *Pentatrichomonas*, *Chilomastix*), most of which are asymptomatic or their pathogenesis is rather mild if any (Graczyk et al. 2005; Shah et al. 2012). However, the human pathogen *Entamoeba histolytica* is unusual and dangerous because, in addition to mild infections within the large intestine typical for all *Entamoeba* species, it can also develop life-

threatening invasive amoebiasis. During the life-cycle, a mature *E. histolytica* cyst infects humans via contaminated water, food, or hands. The cyst passes through the stomach and excysts in the small intestine. This process involves detachment of the amoeba inside the cyst from the cyst wall, formation of a crack, and active escape of the amoeba, which is also called a trophozoite (Yorke and Adams 1926). The trophozoite then migrates to the large intestine where it dwells at the intestinal wall, divides, and produces cysts which are then passed with the stool (Figure 7).

During the initial infection, trophozoites adhere to the colonic mucin layer. The disease is accompanied by mucin depletion, flattening of intestinal epithelial cells, and presence of neutrophils. *E. histolytica* secretory molecules disrupt tight junctions and intestinal ion transport, causing diarrhea (intraintestinal disease). Amoebic lesions in the intestinal epithelium can progress into colonic ulcers in the submucosa. The trophozoites may further enter the bloodstream, disseminate in the liver, causing an amoebic liver abscess (extraintestinal disease) (Marie and Petri 2014; Haque et al. 2003). Virulence factors and mechanisms necessary for the pathogenesis include galactose and N-acetyl-d-galactosamine (Gal/GalNAc)-specific lectin that mediates adherence of trophozoites to mucin and host cells (William A. Petri, Haque, and Mann 2002), trogocytosis i.e., ingestion of cell fragments (“nibbling”) (Ralston et al. 2014), secreted cysteine proteases (Serrano-Luna et al. 2013), hydrolases (Thibeaux et al. 2013) and transcription factors (Narayanasamy et al. 2018).

E. histolytica can be found worldwide, but it is more common in the tropics and sub-tropics, especially in areas with poor sanitation. About 10% of *E. histolytica* infections progress to an invasion of the epithelium, and the extraintestinal disease develops in less than 1% of cases. The invasive outcome of *E. histolytica* infection depends on a complex interaction of factors, which include: nutrition, immunity, intestinal mucin layer, and the intestinal microbiome (Haque et al. 2003; Ali, Clark, and Petri 2008).

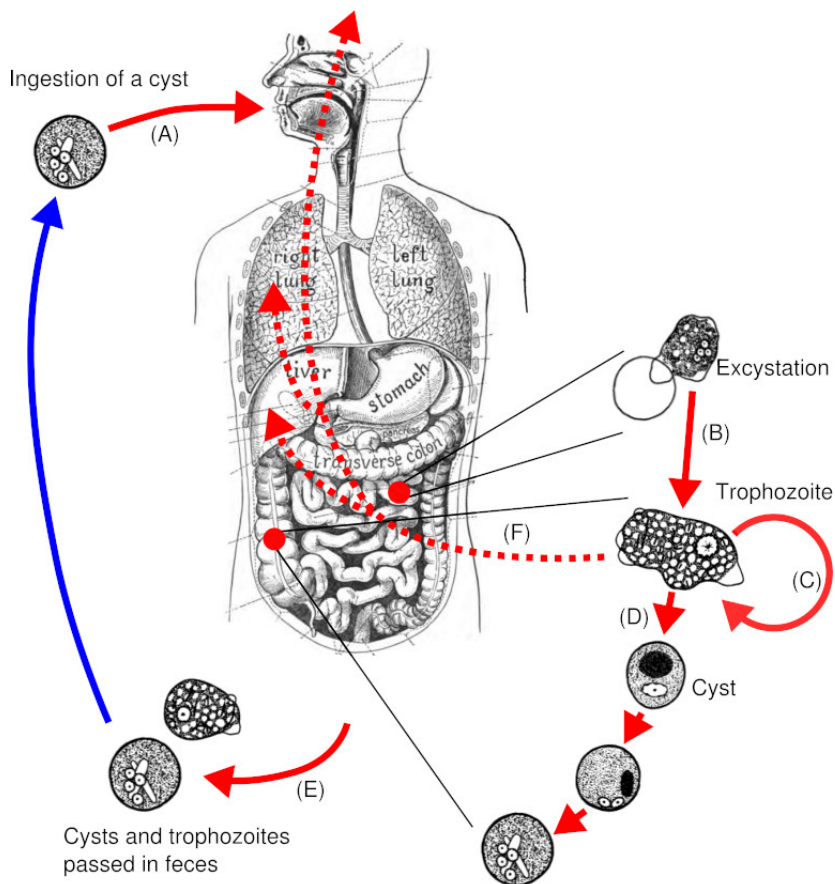


Figure 7. The life cycle of *Entamoeba histolytica*. (A) Humans are infected by ingesting a mature cyst with contaminated water or food; the cyst is then passed through the stomach. (B) In the small intestine, the amoeba sheds the cyst wall and migrates to the large intestine, (C) where it divides, and (D) forms cysts, (E) that are passed with feces. The cyst can survive in the outside environment for weeks. (F) In about 10% of cases, trophozoites can invade and cross intestinal mucosa to the submucosa and cause amoebic dysentery. In less than 1% of cases, the trophozoites reach the bloodstream and disseminate in various organs, particularly the liver, but also lungs and brain (extraintestinal disease). Drawings of *E. histolytica* were reproduced from (Yorke and Adams 1926).

2 Aims

- To analyze the genome of *M. balamuthi* and compare it with genomes of parasitic Entamoebidae.
- To describe adaptations of *M. balamuthi* to free-living anaerobic lifestyle.
- To identify putative ancestral and novel features that allowed the evolution of Entamoebidae into a successful parasitic group.
- To explore the distribution of peroxisomal markers in Archamoebae and other eukaryotes with respect to parasitism and anaerobic lifestyle.

3 List of publications and author contribution

Le, T., **Žárský, V.**, Nývltová, E., Rada, P., Harant, K., Vancová, M., Verner, Z., Hrdý, I. and Tachezy, J., 2020. Anaerobic peroxisomes in *Mastigamoeba balamuthi*. *Proceedings of the National Academy of Sciences*.

Comment: This work presents the first evidence of peroxisomes in an anaerobic organism and outlines possible functions of peroxisomes in *M. balamuthi*.

Author contribution: Discovery of peroxisomal markers in the genome of *M. balamuthi*. Expression of several peroxisomal markers for antibody production.

Nývltová, E., Šut'ák, R., **Žárský, V.**, Harant, K., Hrdý, I. and Tachezy, J., 2017. Lateral gene transfer of p-cresol- and indole-producing enzymes from environmental bacteria to *Mastigamoeba balamuthi*. *Environmental microbiology*, 19(3), pp.1091-1102.

Comment: *M. balamuthi* acquired the cresol synthesis pathway via lateral gene transfer from prokaryotes. It may use the bacteriostatic property of cresol to compete with other microbiota.

Author contribution: A phylogenetic analysis of 4-hydroxyphenylacetate decarboxylase and tryptophanase.

Žárský, V. and Tachezy, J., 2015. Evolutionary loss of peroxisomes – not limited to parasites. *Biology direct*, 10(1), p.74.

Comment: We show that peroxisomes were independently lost in several lineages of parasitic helminths and also in a free-living tunicate *Oikopleura dioica*.

Author contribution: Bioinformatics analysis of peroxisomal components in metazoans.

Žárský, V., Klimeš, V., Eliáš, M., Pačes, J., Vlček, Č., Nývltová, E., Hrdý, I., Barlow, L., Dacks, J., Hall, N., Roger, A., Tachezy, J., *Mastigamoeba balamuthi* genome and the nature of the free-living ancestor of *Entamoeba*. (unpublished manuscript)

Comment: In this manuscript, we present a draft genome sequence of *M. balamuthi* and compare it with genomes of Entamoebids and other amoebozoans. We show that some features crucial for the parasitism of *Entamoeba* were probably present in the common ancestor of *M. balamuthi* and *Entamoeba*, while others may have been acquired in the *Entamoeba* lineage. We also describe flagellar components and plant cell wall-degradation pathways of *M. balamuthi*.

Author contribution: Bioinformatics analysis of the *M. balamuthi* genome, detection of transposable elements, and comparative analyses of amoebozoan genomes. Reconstruction of metabolic pathways, flagellar components, nuclear pore complex, and cyst wall components. Analysis of genes possibly transferred from prokaryotes.

4 Summary

We sequenced and assembled the genome and transcriptome of the anaerobic amoeba *Mastigamoeba balamuthi* and compared these with the genomes of parasitic Entamoebidae and other amoebozoans (Figure 8). The draft genome sequence of *M. balamuthi* has 57.3 million base pairs (Mbp) (Entamoebidae have 20.8-40.9 Mbp) with an extremely high GC content of 61% (Entamoebidae have 23-30% GC). We used the actin gene to compare GC content also with species that don't have their genome sequenced (*Rhizomastix*, *Pelomyxa*, *Mastigella*), and we observed that within Archamoebae the GC content positively correlates with the free-living lifestyle. We predicted 16287 protein-coding genes (Entamoebidae have between 8300 and 12500 protein-coding genes) with an average of 3.4 introns per gene (less than one intron per gene in Entamoebidae). *M. balamuthi* has a rich set of transposable elements: DNA transposons (found in *E. moshkovskii* and *E. invadens*), LINE (long interspersed nuclear elements) retrotransposons (found in *E. histolytica* and *E. dispar*) and LTR (long terminal repeats) retrotransposons (Pritham, Feschotte, and Wessler 2005).

We reconstructed evolutionary histories of amoebozoan genes and focused on the main events in the evolution of Archamoebae. During the transition to anaerobiosis of the common ancestor of Archamoebae, mitochondrial genome and respiration complexes have been lost, except for the succinate dehydrogenase complex (Complex II) which was found in *M. balamuthi* hydrogenosomes. Also, the detoxification system for ROS (reactive oxygen species) and RNS (reactive nitrogen species) has been modified in a way that is typical for anaerobic eukaryotes: the thioredoxin system, peroxiredoxins, and superoxide dismutase are present, while the glutathione synthesis and glutathione-dependent pathways have been lost. Also, flavodiiron protein, iron-sulfur flavoprotein, and rubrerythrin, enzymes typical for anaerobic prokaryotes, have been gained via LGT.

We found a complete pathway for chitin synthesis and degradation, and cell wall-binding components that were so far found only in Entamoebidae (lectins Jessie1/2/3 and Jacob). We, however, did not find homolog of the *Entamoeba* Gal/GalNAc lectin, which is attached to the plasma membrane by a C-terminal transmembrane domain and tethers the chitin-binding lectin Jacob via protein-protein interaction (Chatterjee et al. 2009). Instead, we discovered that in *M. balamuthi*, the chitin-binding proteins (chitinase, Jessie, Jacob) have each a C-terminal transmembrane domain. Thus they are possibly directly anchored to the plasma membrane. It has been observed that during excystation *E. histolytica* trophozoite first detaches from the cyst wall and then actively escapes through a small opening (Yorke and Adams 1926), while *M. balamuthi* instead degrades its cyst (J. Tachezy personal

observation). Based on this, we propose that the indirect attachment of *Entamoeba* cyst wall proteins via Gal/GalNAc lectin allows for a quick release of the trophozoite from the cyst as soon as it reaches the small intestine.

M. balamuthi codes for an extensive repertoire of cysteine proteases, some of which are closely related to *E. histolytica* cysteine proteases essential for virulence and pathogenicity (Serrano-Luna et al. 2013). We also found a homolog of *E. histolytica* amoebapore (a cytolytic protein) in the genome of *M. balamuthi*. The *M. balamuthi* gene, however, resembles rather mammalian prosaposins, as it codes for a polyprotein composed of several saposin domains, each of which is homologous to the amoebapore (Zhai and Saier 2000). Several gene families of *E. histolytica* cell surface (Gal/GalNAc lectin, BspA, Ariel1, EhSTIRPS) were probably acquired in the Entamoebidae lineage, and we did not find any homologs in *M. balamuthi* genome (W. A. Petri et al. 1987; Davis et al. 2006; Mai and Samuelson 1998; MacFarlane and Singh 2007).

While Entamoebidae have completely lost flagellum, *M. balamuthi* forms flagellated cells. Because the genome sequencing projects of amoebozoans were mostly focused on species that lack flagella (*Acanthamoeba*, Dictyostelida, Entamoebidae), the description of genes associated with the amoebozoan flagellum has been missing. We found a rich set of proteins necessary for the assembly and function of the flagellum. We, however, did not find any genes for dyneins of the outer dynein arms, which is in accordance with ultrastructural observations that the outer dynein arms are absent in Archamoebae (Walker et al. 2001; Pánek et al. 2016).

We found a broad set of enzymes and metabolite transporters required for the degradation and utilization of polymers of the plant cell wall (cellulose, hemicellulose, and pectin) in the genome of *M. balamuthi*. This is a probable adaptation allowing to feed on plant debris in the anoxic sediments typical for *Mastigamoeba* and other free-living Archamoebae. It is also possible that these pathways were acquired by the common ancestor of Archamoebae and then secondarily lost in Entamoebidae.

We also discovered that *M. balamuthi* acquired a cresol synthesis pathway (cresol is an odorous phenolic compound) via lateral gene transfer from prokaryotes, and we showed that *M. balamuthi* produces cresol in concentrations that are bacteriostatic to non-cresol-producing bacteria (Nývltová et al. 2017). The acquisition of cresol synthesis may provide *M. balamuthi* a competitive advantage over other microbes in its habitat.

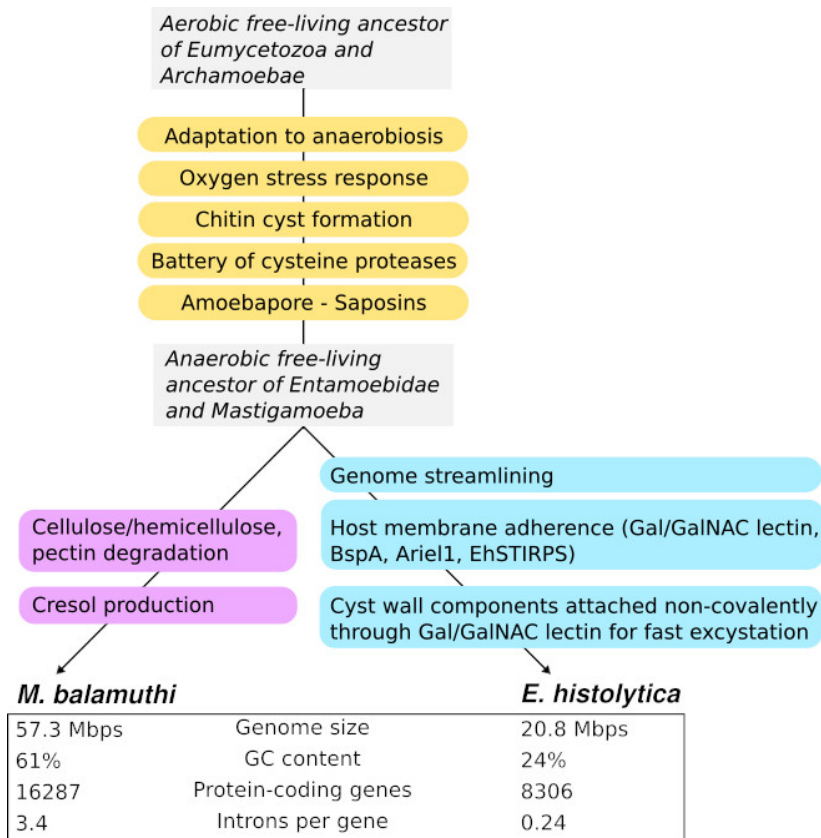


Figure 8. An overview of the preexisting features and adaptations during the evolution of *M. balamuthi* and *Entamoeba*.

During the analysis of the *M. balamuthi* genome, we found a complete set of conserved peroxins (PEX proteins) required for the peroxisomal biogenesis and protein import, which is surprising because no peroxisomes have been described in anaerobic organisms so far (Le et al. 2020). This encouraged us to inquire more about the distribution of peroxins in eukaryotic lineages (Figure 9). We found a reduced set of seven peroxins in the genomes of *E. histolytica*, *E. dispar*, and *E. moshkovskii*, while in *E. invadens*, we found only one peroxin - Pex19. We hypothesize that there are functional remnants of peroxisomes in some entamoebids (*E. histolytica*, *E. dispar*, *E. moshkovskii*) while in *E. invadens* peroxisomes were entirely lost (unpublished data). Independently we discovered Pex19 in the genome of *Monocercomonoides exilis*, which is an oxymonad that lacks both mitochondria and peroxisomes (Karnkowska et al. 2016). This suggests the possibility of an additional peroxisome-independent function of Pex19.

Furthermore, we discovered that parasitic roundworms of the order Trichocephalida and a large group of parasitic flatworms Neodermata (containing flukes, tapeworms, and monogeneans) had lost peroxisomes. Most surprisingly we found, that peroxisomes were lost in a free-living tunicate *Oikopleura dioica* that inhabits oxygen-rich surface sea waters (Žárský and Tachezy 2015).

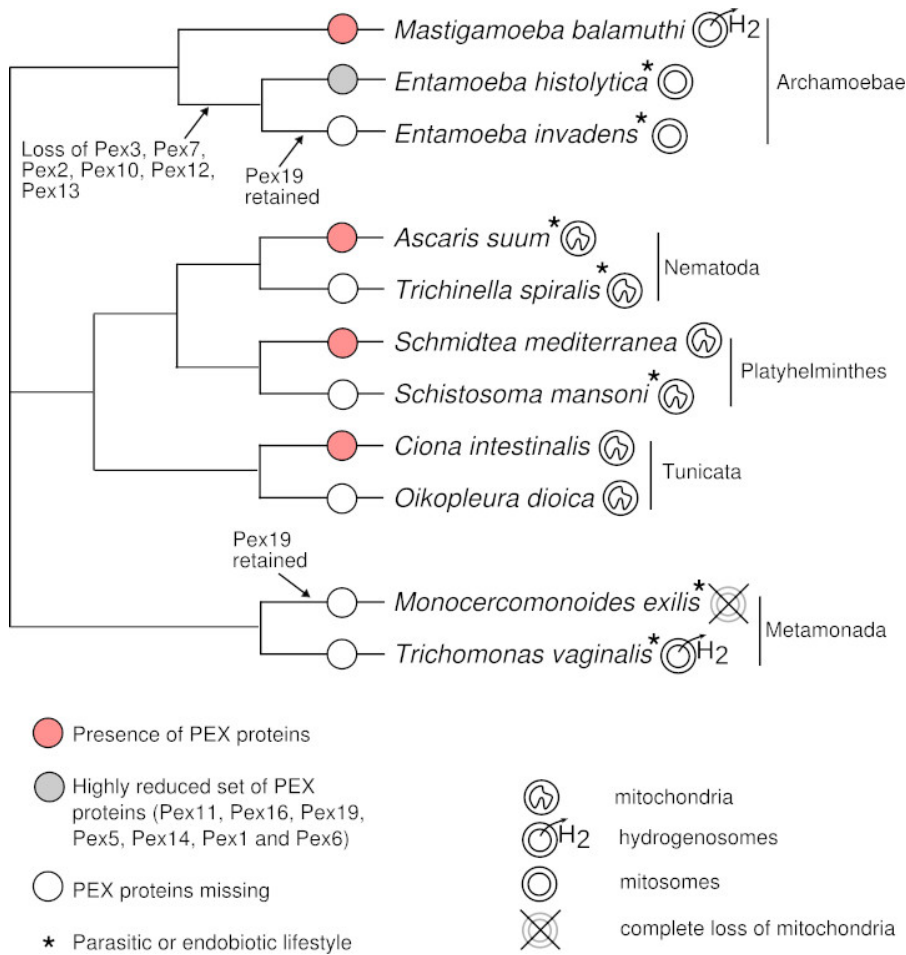


Figure 9. Novel insights into the distribution of peroxisomes in eukaryotes. Presence of mitochondria and mitochondrion-related organelles is highlighted.

5 References

- Ali, Ibne Karim M., C. Graham Clark, and William A. Petri. 2008. "Molecular epidemiology of amebiasis." *Infection, Genetics and Evolution* 8 (5): 698–707.
- Ali, Vahab, Yasuo Shigeta, Umechiyo Tokumoto, Yasuhiro Takahashi, and Tomoyoshi Nozaki. 2004. "An intestinal parasitic protist, *Entamoeba histolytica*, possesses a non-redundant nitrogen fixation-like system for iron-sulfur cluster assembly under anaerobic conditions." *Journal of Biological Chemistry* 279 (16): 16863–74.
- Andrade Rosa, Ivone de, Wanderley de Souza, and Marlene Benchimol. 2013. "High-resolution scanning electron microscopy of the cytoskeleton of *Tritrichomonas foetus*." *Journal of Structural Biology* 183 (3): 412–18.
- Bapteste, Eric, Henner Brinkmann, Jennifer A. Lee, Dorothy V. Moore, Christoph W. Sensen, Paul Gordon, Laure Duruflé, et al. 2002. "The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*." *Proceedings of the National Academy of Sciences* 99 (3): 1414–19.
- Burki, Fabien, Andrew J. Roger, Matthew W. Brown, and Alastair G. B. Simpson. 2020. "The new tree of eukaryotes." *Trends in Ecology & Evolution* 35 (1): 43–55.
- Cavalier-Smith, T. 1981. "Eukaryote kingdoms: seven or nine?" *Biosystems* 14 (3): 461–81.
- Cavalier-Smith, T. 1983. "A 6-kingdom classification and a unified phylogeny." *Intracellular Space as Oligogenetic Ecosystem*, 1027–34.
- Chatterjee, Anirban, Sudip K. Ghosh, Ken Jang, Esther Bullitt, Landon Moore, Phillips W. Robbins, and John Samuelson. 2009. "Evidence for a 'wattle and daub' model of the cyst wall of *Entamoeba*." *PLoS Pathogens* 5 (7).
- Chavez, Lawrence A., William Balamuth, and Thomas Gong. 1986. "A Light and Electron Microscopical study of a new, polymorphic free-living amoeba, *Phreatamoeba balamuthi* n. g., n. sp." *The Journal of Protozoology* 33 (3): 397–404.
- Cornford, M. E., and D. E. Outka. 1970. "Peroxisomes as organelles of terminal oxidation in trichomonads." *Journal of Cell Biology* 47.

- Davis, Paul H., Zhi Zhang, Minghe Chen, Xiaochun Zhang, Subhra Chakraborty, and Samuel L. Stanley. 2006. "Identification of a family of Bsp-A like surface proteins of *Entamoeba histolytica* with novel leucine rich repeats." *Molecular and Biochemical Parasitology* 145 (1): 111–16.
- De Duve, C. 1969. "The peroxisome: a new cytoplasmic organelle." *Proceedings of the Royal Society of London. Series B, Biological Sciences* 173 (1030): 71–83.
- Dolezal, Pavel, Michael J. Dagley, Maya Kono, Peter Wolyneec, Vladimir A. Likić, Jung Hock Foo, Miroslava Sedinová, et al. 2010. "The essentials of protein import in the degenerate mitochondrion of *Entamoeba histolytica*." *PLoS Pathogens* 6 (3).
- Fahey, R. C., G. L. Newton, B. Arrick, T. Overdank-Bogart, and S. B. Aley. 1984. "*Entamoeba histolytica*: a eukaryote without glutathione metabolism." *Science* 224 (4644): 70–72.
- Gabaldón, Toni. 2010. "Peroxisome diversity and evolution." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365 (1541): 765–73.
- Gabaldón, Toni, Michael L. Ginger, and Paul A. M. Michels. 2016. "Peroxisomes in parasitic protists." *Molecular and Biochemical Parasitology*, 209 (1): 35–45.
- Gabaldón, Toni, Berend Snel, Frank van Zimmeren, Wieger Hemrika, Henk Tabak, and Martijn A. Huynen. 2006. "Origin and evolution of the peroxisomal proteome." *Biology Direct* 1 (1): 8.
- Gawryluk, Ryan M. R., Ryoma Kamikawa, Courtney W. Stairs, Jeffrey D. Silberman, Matthew W. Brown, and Andrew J. Roger. 2016. "The earliest stages of mitochondrial adaptation to low oxygen revealed in a novel rhizarian." *Current Biology* 26 (20): 2729–38.
- Gill, Erin E., Sara Diaz-Triviño, Maria José Barberà, Jeffrey D. Silberman, Alexandra Stechmann, Daniel Gaston, Ivica Tamas, and Andrew J. Roger. 2007. "Novel mitochondrion-related organelles in the anaerobic amoeba *Mastigamoeba balamuthi*." *Molecular Microbiology* 66 (6): 1306–20.
- Gould, S.j., G.a. Keller, M. Schneider, S.h. Howell, L.j. Garrard, J.m. Goodman, B. Distel, H. Tabak, and S. Subramani. 1990. "Peroxisomal protein import is conserved between yeast, plants, insects and mammals." *The EMBO Journal* 9 (1): 85–90.
- Gould, Stephen Jay, and Elisabeth S. Vrba. 1982. "Exaptation—a missing term in the science of form." *Paleobiology* 8 (1): 4–15.

- Graczyk, Thaddeus K., Clive K. Shiff, Leena Tamang, Fair Munsaka, Anna M. Beitin, and William J. Moss. 2005. "The association of *Blastocystis hominis* and *Endolimax nana* with diarrheal stools in Zambian school-age children." *Parasitology Research* 98 (1): 38.
- Haque, Rashidul, Christopher D. Huston, Molly Hughes, Eric Houpt, and William A. Petri. 2003. "Amebiasis." *New England Journal of Medicine* 348 (16): 1565–73.
- Hrdy, Ivan, Robert P. Hirt, Pavel Dolezal, Lucie Bardonová, Peter G. Foster, Jan Tachezy, and T. Martin Embley. 2004. "Trichomonas hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I." *Nature* 432 (7017): 618–22.
- Jackson, Andrew P., Thomas D. Otto, Martin Aslett, Stuart D. Armstrong, Frederic Bringaud, Alexander Schlacht, Catherine Hartley, et al. 2016. "Kinetoplastid phylogenomics reveals the evolutionary innovations associated with the origins of parasitism." *Current Biology* 26 (2): 161–72.
- Janouškovec, Jan, Denis V. Tikhonenkov, Fabien Burki, Alexis T. Howe, Martin Kolísko, Alexander P. Mylnikov, and Patrick J. Keeling. 2015. "Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives." *Proceedings of the National Academy of Sciences* 112 (33): 10200–207.
- Jeelani, Ghulam, and Tomoyoshi Nozaki. 2016. "Entamoeba thiol-based redox metabolism: a potential target for drug development." *Molecular and Biochemical Parasitology, Redox Biology of Parasites: Targets for Therapy*, 206 (1): 39–45.
- Kang, Seungho, Alexander K. Tice, Frederick W. Spiegel, Jeffrey D. Silberman, Tomáš Pánek, Ivan Čepička, Martin Kostka, et al. 2017. "Between a pod and a hard test: the deep evolution of amoebae." *Molecular Biology and Evolution* 34 (9): 2258–70.
- Karnkowska, Anna, Vojtěch Vacek, Zuzana Zubáčová, Sebastian C. Treitli, Romana Petrželková, Laura Eme, Lukáš Novák, et al. 2016. "A eukaryote without a mitochondrial organelle." *Current Biology* 26 (10): 1274–84.
- Le, Tien, Vojtěch Žárský, Eva Nývltová, Petr Rada, Karel Harant, Marie Vancová, Zdeněk Verner, Ivan Hrdý, and Jan Tachezy. 2020. "Anaerobic peroxisomes in *Mastigamoeba balamuthi*." *Proceedings of the National Academy of Sciences* 117 (4): 2065–75.

- Lindmark, Donald G., Miklós Müller, and Helen Shio. 1975. "Hydrogenosomes in *Trichomonas vaginalis*." *The Journal of Parasitology* 61 (3): 552–54.
- Loftus, Brendan, Iain Anderson, Rob Davies, U. Cecilia M. Alsmark, John Samuelson, Paolo Amedeo, Paola Roncaglia, et al. 2005. "The genome of the protist parasite *Entamoeba histolytica*." *Nature* 433 (7028): 865–68.
- Lom, Jiri, and John O. Corliss. 1967. "Ultrastructural observations on the development of the microsporidian protozoon *Plistophora hyphessobryconis* Schaperclaus." *The Journal of Protozoology* 14 (1): 141–52.
- Ludewig-Klingner, Ann-Kathrin, Victoria Michael, Michael Jarek, Henner Brinkmann, and Jörn Petersen. 2018. "Distribution and evolution of peroxisomes in alveolates (Apicomplexa, Dinoflagellates, Ciliates)." *Genome Biology and Evolution* 10 (1): 1–13.
- Lukeš, Julius, Tomáš Skalický, Jiří Týč, Jan Votýpka, and Vyacheslav Yurchenko. 2014. "Evolution of parasitism in kinetoplastid flagellates." *Molecular and Biochemical Parasitology*, 195 (2): 115–22.
- MacFarlane, Ryan C., and Upinder Singh. 2007. "Identification of an *Entamoeba histolytica* serine-, threonine-, and isoleucine-rich protein with roles in adhesion and cytotoxicity." *Eukaryotic Cell* 6 (11): 2139–46.
- Mai, Zhiming, and John Samuelson. 1998. "A new gene family (ariel) encodes asparagine-rich *Entamoeba histolytica* antigens, which resemble the amebic vaccine candidate serine-rich *E. histolytica* protein." *Infection and Immunity* 66 (1): 353–55.
- Maralikova, Barbora, Vahab Ali, Kumiko Nakada-Tsukui, Tomoyoshi Nozaki, Mark Van Der Giezen, Katrin Henze, and Jorge Tovar. 2010. "Bacterial-type oxygen detoxification and iron-sulfur cluster assembly in amoebal relict mitochondria." *Cellular Microbiology* 12 (3): 331–42.
- Marie, Chelsea, and William A. Petri. 2014. "Regulation of virulence of *Entamoeba histolytica*." *Annual Review of Microbiology* 68 (1): 493–520.
- Mathur, Varsha, Martin Kolísko, Elisabeth Hehenberger, Nicholas A. T. Irwin, Brian S. Leander, Árni Kristmundsson, Mark A. Freeman, and Patrick J. Keeling. 2019. "Multiple independent origins of apicomplexan-like parasites." *Current Biology* 29 (17): 2936-2941.e5.

- Mi-ichi, Fumika, Tomofumi Miyamoto, Shouko Takao, Ghulam Jeelani, Tetsuo Hashimoto, Hiromitsu Hara, Tomoyoshi Nozaki, and Hiroki Yoshida. 2015. “*Entamoeba* mitochondria play an important role in encystation by association with cholesteryl sulfate synthesis.” *Proceedings of the National Academy of Sciences* 112 (22): E2884–90.
- Mi-ichi, Fumika, Mohammad Abu Yousuf, Kumiko Nakada-Tsukui, and Tomoyoshi Nozaki. 2009. “Mitochondria in *Entamoeba histolytica* contain a sulfate activation pathway.” *Proceedings of the National Academy of Sciences* 106 (51): 21731–36.
- Morin, Loïc. 2000. “Long branch attraction effects and the status of ‘basal eukaryotes’: phylogeny and structural analysis of the ribosomal RNA gene cluster of the free-living diplomonad *Trepomonas agilis*.” *Journal of Eukaryotic Microbiology* 47 (2): 167–77.
- Müller, Miklós, Marek Mentel, Jaap J. van Hellemond, Katrin Henze, Christian Woehle, Sven B. Gould, Re-Young Yu, Mark van der Giezen, Aloysius G. M. Tielens, and William F. Martin. 2012. “Biochemistry and evolution of anaerobic energy metabolism in eukaryotes.” *Microbiology and Molecular Biology Reviews* 76 (2): 444–95.
- Narayanasamy, Ravi Kumar, Carlos Alberto Castañón-Sánchez, Juan Pedro Luna-Arias, Guillermina García-Rivera, Bartolo Avendaño-Borromeo, María Luisa Labra-Barrios, Jesús Valdés, María Esther Herrera-Aguirre, and Esther Orozco. 2018. “The *Entamoeba histolytica* TBP and TRF1 transcription factors are GAAC-box binding proteins, which display differential gene expression under different stress stimuli and during the interaction with mammalian cells.” *Parasites & Vectors* 11 (1): 153.
- Nývltová, Eva, Courtney W. Stairs, Ivan Hrdý, Jakub Rídl, Jan Mach, Jan Pačes, Andrew J. Roger, and Jan Tachezy. 2015. “Lateral gene transfer and gene duplication played a key role in the evolution of *Mastigamoeba balamuthi* hydrogenosomes.” *Molecular Biology and Evolution* 32 (4): 1039–55.
- Nývltová, Eva, Robert Šuták, Karel Harant, Miroslava Šedinová, Ivan Hrdý, Jan Pačes, Čestmír Vlček, and Jan Tachezy. 2013. “NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of *Mastigamoeba balamuthi*.” *Proceedings of the National Academy of Sciences* 110 (18): 7371–76.

- Nývltová, Eva, Robert Šut'ák, Vojtěch Žárský, Karel Harant, Ivan Hrdý, and Jan Tachezy. 2017. "Lateral gene transfer of p-cresol- and indole-producing enzymes from environmental bacteria to *Mastigamoeba balamuthi*." *Environmental Microbiology* 19 (3): 1091–1102.
- Pánek, Tomáš, Eliška Zadrobílková, Giselle Walker, Matthew W. Brown, Eleni Gentekaki, Miluše Hroudová, Seungho Kang, et al. 2016. "First multigene analysis of Archamoebae (Amoebozoa: Conosa) robustly reveals its phylogeny and shows that Entamoebidae represents a deep lineage of the group." *Molecular Phylogenetics and Evolution* 98 (May): 41–51.
- Petri, W. A., R. D. Smith, P. H. Schlesinger, C. F. Murphy, and J. I. Ravdin. 1987. "Isolation of the galactose-binding lectin that mediates the in vitro adherence of *Entamoeba histolytica*." *The Journal of Clinical Investigation* 80 (5): 1238–44.
- Petri, William A., Rashidul Haque, and Barbara J. Mann. 2002. "The bittersweet interface of parasite and host: lectin-carbohydrate interactions during human invasion by the parasite *Entamoeba histolytica*." *Annual Review of Microbiology* 56 (1): 39–64.
- Pfanner, Nikolaus, and Walter Neupert. 1990. "The mitochondrial protein import apparatus." *Annual Review of Biochemistry* 59 (1): 331–53.
- Pieuchot, Laurent, and Gregory Jedd. 2012. "Peroxisome assembly and functional diversity in eukaryotic microorganisms." *Annual Review of Microbiology* 66 (1): 237–63.
- Poulin, Robert, and Haseeb S. Randhawa. 2015. "Evolution of parasitism along convergent lines: from ecology to genomics." *Parasitology* 142 (S1): S6–15.
- Pritham, Ellen J., Cédric Feschotte, and Susan R. Wessler. 2005. "Unexpected diversity and differential success of DNA transposons in four species of *Entamoeba* protozoans." *Molecular Biology and Evolution* 22 (9): 1751–63.
- Ptáčková, Eliška, Alexei Yu. Kostygov, Lyudmila V. Chistyakova, Lukáš Falteisek, Alexander O. Frolov, David J. Patterson, Giselle Walker, and Ivan Cepicka. 2013. "Evolution of Archamoebae: morphological and molecular evidence for pelobionts including *Rhizomastix*, *Entamoeba*, *Iodamoeba*, and *Endolimax*." *Protist* 164 (3): 380–410.
- Ralston, Katherine S., Michael D. Solga, Nicole M. Mackey-Lawrence, Somlata, Alok Bhattacharya, and William A. Petri. 2014. "Trogocytosis by *Entamoeba histolytica* contributes to cell killing and tissue invasion." *Nature* 508 (7497): 526–30.

- Roger, Andrew J., Sergio A. Muñoz-Gómez, and Ryoma Kamikawa. 2017. "The origin and diversification of mitochondria." *Current Biology* 27 (21): R1177–92.
- Schlüter, Agatha, Stéphane Fourcade, Raymond Ripp, Jean Louis Mandel, Olivier Poch, and Aurora Pujol. 2006. "The evolutionary origin of peroxisomes: an ER-peroxisome connection." *Molecular Biology and Evolution* 23 (4): 838–45.
- Serrano-Luna, Jesús, Carolina Piña-Vázquez, Magda Reyes-López, Guillermo Ortiz-Estrada, and Mireya de la Garza. 2013. "Proteases from *Entamoeba spp.* and pathogenic free-living amoebae as virulence factors." Review Article. *Journal of Tropical Medicine*. 2013.
- Shah, Mitanshu, Christopher Bryan Tan, Dhyan Rajan, Shadab Ahmed, Krishnaiyer Subramani, Kaleem Rizvon, and Paul Mustacchia. 2012. "*Blastocystis hominis* and *Endolimax nana* co-infection resulting in chronic diarrhea in an immunocompetent male." *Case Reports in Gastroenterology* 6 (2): 358–64.
- Shaw, Michael K., and Lewis G. Tilney. 1999. "Induction of an acrosomal process in *Toxoplasma gondii*: visualization of actin filaments in a protozoan parasite." *Proceedings of the National Academy of Sciences* 96 (16): 9095–99.
- Speijer, Dave. 2017. "Evolution of peroxisomes illustrates symbiogenesis." *BioEssays* 39 (9): 1700050.
- Stairs, Courtney W, Laura Eme, Sergio A Muñoz-Gómez, Alejandro Cohen, Graham Dellaire, Jennifer N Shepherd, James P Fawcett, and Andrew J Roger. 2018. "Microbial eukaryotes have adapted to hypoxia by horizontal acquisitions of a gene involved in rhodoquinone biosynthesis." Edited by John McCutcheon. *ELife* 7 (April): e34292.
- Stairs, Courtney W., Michelle M. Leger, and Andrew J. Roger. 2015. "Diversity and origins of anaerobic metabolism in mitochondria and related organelles." *Philosophical Transactions of the Royal Society B: Biological Sciences* 370 (1678): 20140326.
- Stensvold, C. Rune, Marianne Lebbad, Emma L. Victory, Jaco J. Verweij, Egbert Tannich, Mohammed Alfellani, Paulette Legarraga, and C. Graham Clark. 2011. "Increased sampling reveals novel lineages of *Entamoeba*: consequences of genetic diversity and host specificity for taxonomy and molecular detection." *Protist* 162 (3): 525–41.

- Sugiura, Ayumu, Sevan Mattie, Julien Prudent, and Heidi M. McBride. 2017. “Newly born peroxisomes are a hybrid of mitochondrial and ER-derived pre-peroxisomes.” *Nature* 542 (7640): 251–54.
- Tan, Kevin S. W. 2008. “New insights on classification, identification, and clinical relevance of *Blastocystis* spp.” *Clinical Microbiology Reviews* 21 (4): 639–65.
- Thibeaux, Roman, Christian Weber, Chung-Chau Hon, Marie-Agnès Dillies, Patrick Avé, Jean-Yves Coppée, Elisabeth Labruyère, and Nancy Guillén. 2013. “Identification of the virulence landscape essential for *Entamoeba histolytica* invasion of the human colon.” *PLoS Pathogens* 9 (12).
- Tovar, Jorge, Anke Fischer, and C. Graham Clark. 1999. “The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*.” *Molecular Microbiology* 32 (5): 1013–21.
- Walker, Giselle, Alastair G. B. Simpson, Virginia Edgcomb, Mitchell L. Sogin, and David J. Patterson. 2001. “Ultrastructural identities of *Mastigamoeba punctachora*, *Mastigamoeba simplex* and *Mastigella commutans* and assessment of hypotheses of relatedness of the pelobionts (protista).” *European Journal of Protistology* 37 (1): 25–49.
- Xu, Feifei, Jon Jerlström-Hultqvist, Martin Kolisko, Alastair G. B. Simpson, Andrew J. Roger, Staffan G. Svärd, and Jan O. Andersson. 2016. “On the reversibility of parasitism: adaptation to a free-living lifestyle via gene acquisitions in the diplomonad *Trepomonas* sp. pc1.” *BMC Biology* 14 (1): 62.
- Yazaki, Euki, Sohta A. Ishikawa, Keitaro Kume, Akira Kumagai, Takashi Kamaishi, Goro Tanifuji, Tetsuo Hashimoto, and Yuji Inagaki. 2017. “Global kinetoplastea phylogeny inferred from a large-scale multigene alignment including parasitic species for better understanding transitions from a free-living to a parasitic lifestyle.” *Genes & Genetic Systems* pp.16-00056.
- Yorke, Warrington, and A. R. D. Adams. 1926. “Observations on *Entamoeba histolytica*.” *Annals of Tropical Medicine & Parasitology* 20 (3): 279–302.
- Yubuki, Naoji, Vít Céza, Ivan Cepicka, Akinori Yabuki, Yuji Inagaki, Takeshi Nakayama, Isao Inouye, and Brian S. Leander. 2010. “Cryptic diversity of free-living parabasalids, *Pseudotrichomonas keilini* and *Lacusteria cypriaca* n. g., n. sp., as inferred from small subunit rDNA sequences.” *Journal of Eukaryotic Microbiology* 57 (6): 554–61.

- Zadrobílková, Eliška, Pavla Smejkalová, Giselle Walker, and Ivan Čepička. 2016. “Morphological and molecular diversity of the neglected genus *Rhizomastix* Alexeieff, 1911 (Amoebozoa: Archamoebae) with description of five new species.” *Journal of Eukaryotic Microbiology* 63 (2): 181–97.
- Zadrobílková, Eliška, Giselle Walker, and Ivan Čepička. 2015. “Morphological and molecular evidence support a close relationship between the free-living Archamoebae *Mastigella* and *Pelomyxa*.” *Protist* 166 (1): 14–41.
- Žárský, Vojtěch, and Jan Tachezy. 2015. “Evolutionary loss of peroxisomes – not limited to parasites.” *Biology Direct* 10 (1): 74.
- Zhai, Yufeng, and Milton H Saier. 2000. “The amoebapore superfamily.” *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* 1469 (2): 87–99.



Anaerobic peroxisomes in *Mastigamoeba balamuthi*

Tien Le^a, Vojtěch Žárský^a, Eva Nývltová^a, Petr Rada^a, Karel Harant^a, Marie Vancová^b, Zdeněk Verner^a, Ivan Hrdý^a, and Jan Tachezy^{a,1}

^aDepartment of Parasitology, Faculty of Science, BIOCEV, Charles University, 25242 Vestec, Czech Republic; and ^bInstitute of Parasitology, Biology Centre, Czech Academy of Sciences, 370 05 České Budějovice, Czech Republic

Edited by Tom M. Fenchel, University of Copenhagen, Helsingor, Denmark, and approved December 12, 2019 (received for review July 3, 2019)

The adaptation of eukaryotic cells to anaerobic conditions is reflected by substantial changes to mitochondrial metabolism and functional reduction. Hydrogenosomes belong among the most modified mitochondrial derivative and generate molecular hydrogen concomitant with ATP synthesis. The reduction of mitochondria is frequently associated with loss of peroxisomes, which compartmentalize pathways that generate reactive oxygen species (ROS) and thus protect against cellular damage. The biogenesis and function of peroxisomes are tightly coupled with mitochondria. These organelles share fission machinery components, oxidative metabolism pathways, ROS scavenging activities, and some metabolites. The loss of peroxisomes in eukaryotes with reduced mitochondria is thus not unexpected. Surprisingly, we identified peroxisomes in the anaerobic, hydrogenosome-bearing protist *Mastigamoeba balamuthi*. We found a conserved set of peroxin (Pex) proteins that are required for protein import, peroxisomal growth, and division. Key membrane-associated Pexs (*MbPex3*, *MbPex11*, and *MbPex14*) were visualized in numerous vesicles distinct from hydrogenosomes, the endoplasmic reticulum (ER), and Golgi complex. Proteomic analysis of cellular fractions and prediction of peroxisomal targeting signals (PTS1/PTS2) identified 51 putative peroxisomal matrix proteins. Expression of selected proteins in *Saccharomyces cerevisiae* revealed specific targeting to peroxisomes. The matrix proteins identified included components of acyl-CoA and carbohydrate metabolism and pyrimidine and CoA biosynthesis, whereas no components related to either β -oxidation or catalase were present. In conclusion, we identified a subclass of peroxisomes, named “anaerobic” peroxisomes that shift the current paradigm and turn attention to the reductive evolution of peroxisomes in anaerobic organisms.

peroxisome | *Mastigamoeba balamuthi* | mitochondria | anaerobiosis

Peroxisomes are single membrane-bound organelles present in nearly all eukaryotes (1). They are defined by a conserved set of proteins named peroxins (Pexs) that are required for peroxisomal biogenesis (2). Biochemically, most peroxisomes share the function of fatty acid oxidation, which generates hydrogen peroxide, and hydrogen peroxide-degrading enzymes, namely, catalase, to prevent cellular oxidative damage (3). Additional peroxisomal pathways are highly variable, indicating the remarkable versatility of peroxisomal functions (4). This is reflected by the various names of peroxisomal subtypes, such as glyoxysomes, glycosomes, and Woronin bodies (5). The need for the compartmentalization of enzymes that produce reactive oxygen species (ROS) has been proposed as a driving force behind the evolution of peroxisomes. Peroxisomes were most likely derived from the endoplasmic reticulum (ER) and originated endogenously within eukaryotes (6, 7), although they were long thought to be endosymbiotic remnants similar to mitochondria and plastids (8). The β -oxidation of fatty acids is considered to be the earliest pathway that resided in an ancestral peroxisomal proteome due to its widespread occurrence (6). Phylogenetic analysis suggested that this pathway was inherited from the endosymbiotic ancestor of mitochondria, duplicated and retargeted to peroxisomes to minimize the harmful effects of ROS (6, 7). However, a parallel role of

the ER in the evolution of peroxisomal β -oxidation has been proposed (9, 10).

In addition to fatty acid metabolism, there is a plethora of other metabolic, regulatory, and evolutionary links between peroxisomes and mitochondria (11, 12). For example, the glyoxylate cycle of peroxisomes of land plants, fungi, alveolates, and lower animals uses acetyl-CoA as a substrate to produce succinate that can be imported into mitochondria to replenish the tricarboxylic acid (TCA) cycle. Both peroxisomes and mitochondria divide by fission and share multiple components of the fission machinery (13). More recently, mitochondria appeared to be directly involved in peroxisome biogenesis as a source of peroxisomal membranes and membrane proteins (14). Because of this tight interplay between peroxisomes and mitochondria and the dependence of key peroxisomal oxidases on the presence of molecular oxygen, it is not surprising that peroxisomes are generally absent in anaerobic unicellular eukaryotes (protists) (15, 16). These eukaryotes harbor highly reduced anaerobic mitochondria, such as hydrogenosomes and mitosomes (17). Hydrogenosomes lack most mitochondrial metabolic functions, including β -oxidation. ATP is generated at the substrate level using an anaerobic pathway that catabolizes pyruvate or malate to acetate, CO₂, and molecular hydrogen. Anaerobic conditions apparently prevent the use of oxygen-dependent β -oxidation of fatty acid metabolism in both peroxisomes and mitochondria, as well as other oxygen-dependent catabolic

Significance

It is generally accepted that peroxisomes are absent in anaerobic eukaryotes. These organelles have evolved to compartmentalize oxidative pathways and prevent cellular oxidative damage, namely, the β -oxidation of fatty acids that utilizes molecular oxygen and produces hydrogen peroxide. Mitochondria possess a parallel β -oxidation pathway coupled with the respiratory chain. The adaptation of eukaryotes to anaerobiosis is reflected by the reduction of mitochondrial metabolism and, not surprisingly, concomitant loss of peroxisomes. The free-living anaerobic protist *Mastigamoeba balamuthi* is an organism that contradicts this paradigm. Although *Mastigamoeba* possesses hydrogenosomes, an anaerobic form of mitochondria, it also harbors peroxisomes. These organelles contain the archetypical Pexs that are required for peroxisomal biogenesis; however, they lack the hallmarks of peroxisomal metabolism, β -oxidation and catalase.

Author contributions: J.T. designed research; T.L., V.Ž., E.N., P.R., K.H., M.V., Z.V., I.H., and J.T. performed research; and J.T. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: MaxQuant results were uploaded to the PRIDE partner repository (<https://www.ebi.ac.uk/pride/archive/>) with the dataset identifier PXD014205. The sequences reported in this paper have been deposited in the online resource for community annotation of eukaryotes (ORCAE, <https://bioinformatics.psb.ugent.be/orcae/>) and are given in Dataset S1.

¹To whom correspondence may be addressed. Email: tachezy@natur.cuni.cz.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1909755117/-DCSupplemental>.

and biosynthetic pathways in anaerobic protists. As a consequence, peroxisomes may become dispensable for these organisms.

Mastigamoeba balamuthi is a free-living anaerobic amoeba of the Archamoebae lineage. The characteristics of the mitochondria of *M. balamuthi* conform to those of hydrogenosomes (18). The anaerobic character of *M. balamuthi* emphasizes the presence of a nitrogen-fixing (NIF) and oxygen-sensitive ϵ -bacterial FeS cluster assembly machinery that replaced the standard mitochondrial iron-sulfur cluster (ISC) machinery (19). Moreover, oxygen-sensitive enzymes such as hydrogenases are present in *M. balamuthi* hydrogenosomes and in the cytosol, producing hydrogen from both compartments (19). Archamoebae belongs to the group Conosa with the sister lineage Mycetozoa that inhabits aerobic niches and contains oxygen-respiring mitochondria and standard peroxisomes (20, 21). Thus, it is likely that hydrogenosomes in *M. balamuthi* evolved from aerobic mitochondria via secondary adaptation to anaerobic environments, and we expected that an anaerobic lifestyle would lead to the loss of peroxisomes, as observed in other anaerobic protists. To our surprise, we found that *M. balamuthi* retains a full set of Pexs to form peroxisomes, and we identified enzymes that are delivered to the peroxisomal matrix via typical peroxisomal targeting signals (PTSs) to catalyze fragmented metabolic pathways. However, the oxygen-dependent peroxisomal hallmarks, including β -oxidation of fatty acids and catalase, are absent. These organelles represent a peroxisomal subtype named anaerobic peroxisomes.

Results

Identification and Cell Localization of Pexs. Analysis of the *M. balamuthi* genome sequence using HHpred searches indicated the presence of genes encoding a set of 14 Pexs involved in the recognition of PTS1 and PTS2 (Pex5 and 7, respectively), protein docking and import (Pex13 and 14), receptor recycling (Pex1, 2, 6, 10, and 12), membrane protein targeting (Pex3, 16, and 19), and peroxisome fission (Pex11-1 and Pex11-2) (Fig. 1A and SI Appendix, Fig. S1 and Table S1). Most *M. balamuthi* Pexs (*MbPexs*) were colinear or slightly shorter than the corresponding orthologs, and all possessed the expected functional domains (SI Appendix, Fig. S2). The C-terminal domain of Pex14 is markedly short; however, its N-terminal domain contains characteristic phenylalanine residues that are indispensable for its interaction with Pex5 (22) (SI Appendix, Fig. S2). The functional domains were identified using HHpred searches with high confidence (probability values ranged from 84 to 100) (SI Appendix, Fig. S2), whereas the overall similarity of *MbPex* protein sequences to their human orthologs ranged between 14% for Pex3 and 57% for Pex7 (SI Appendix, Table S2). The correct identification of *MbPexs* containing domains that are present in various proteins with unrelated functions, such as ATPases associated with diverse cellular activities (AAA, Pex1, and Pex6), tetratricopeptide repeats (TPR and Pex5), WD40 repeats (Pex7), and proteins of the RING family (Pex2, Pex10, and Pex12), was supported by phylogenetic analysis using human, yeast, and *Dictyostelium discoideum* orthologs and selected outgroups (Fig. 1B). Phylogenetic analysis of the *MbPex11-1* and *MbPex11-2* paralogs revealed their relationships with mammalian Pex11 α and Pex11 γ , respectively (SI Appendix, Table S2 and Fig. S3). These proteins are involved in the recruitment of the peroxisomal membrane fission machinery, as are Fis1, Mff, and DRP1 (23, 24). Indeed, our searches identified genes for Fis1 (m51a1_g2061) and DRP1 (m51a1_g631) in the *M. balamuthi* genome, which supports the predicted function of *MbPex11-1* and *MbPex11-2*.

The *MbPex5* receptor contains a conserved Cys-13 residue that is required for ubiquitination by the E2 ubiquitin-conjugating enzyme Pex4 (25) and thus receptor recycling (SI Appendix, Fig. S2). In *M. balamuthi*, we identified neither Pex4 nor the Pex4-anchoring

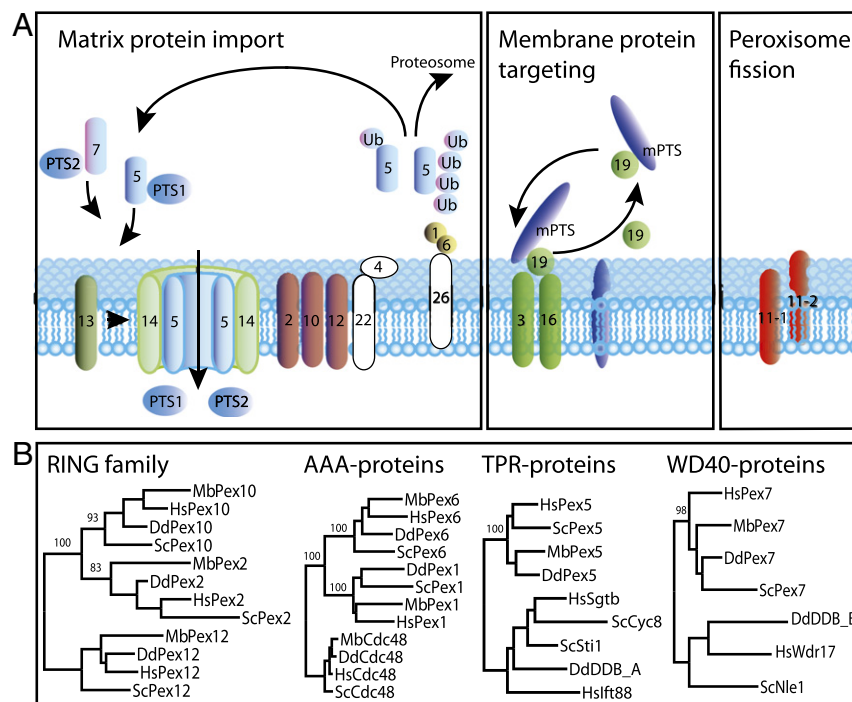


Fig. 1. (A) Scheme of the peroxisomal machinery identified in *M. balamuthi*. The white color indicates Pexs that were not identified in *M. balamuthi*. (B) Phylogenetic analysis of selected proteins was performed to discriminate peroxins (Pexs) and proteins that share common domains, including ATPases associated with diverse cellular activities (AAA, Pex1, and Pex6), tetratricopeptide repeats (TPR and Pex5), WD40 repeats (Pex7), and proteins of the RING family (Pex2, Pex10, and Pex12). Mb, *M. balamuthi*; Hs, *Homo sapiens*; Dd, *Dictyostelium discoideum*; Sc, *Saccharomyces cerevisiae*. Accession numbers are given in SI Appendix, Fig. S2 and Table S1.

protein Pex22 (Fig. 1A). Moreover, Pex26, which anchors the Pex1/6 complex, seems to be absent as well. However, Pex4, Pex22, and Pex26 are not generally distributed in eukaryotes (*SI Appendix, Fig. S1*). For example, Pex4 and Pex22 are absent in metazoans, and Pex22 is absent in diatoms, whereas metazoan Pex26 is replaced by Pex15 in *Saccharomyces cerevisiae* (26), and Pex26 is absent in kinetoplastids and diatoms (*SI Appendix, Fig. S1*) (27).

To validate our in silico predictions, we investigated the cellular localization of membrane-associated *MbPexs*. Confocal immunofluorescence microscopy (IF) using specific polyclonal antibodies (Abs) raised against *MbPex3*, *MbPex11-1*, and *MbPex14* labeled numerous small vesicles scattered within the *M. balamuthi* cytosol. Importantly, *MbPex14* colocalized in the same vesicles with *MbPex3* (Pearson's correlation coefficient [PCC] $r = 0.85$) as well as with *MbPex11* (PCC $r = 0.86$) (Fig. 2A). To further support the specific localization of Pexs and to estimate the size and shape of putative peroxisomes and their

relationships with other organelles, we subjected *M. balamuthi* to stimulated emission depletion (STED) and immunoelectron microscopy analysis (Fig. 2B and C). Dual-color STED imaging not only confirmed the colocalization of Pex3 and Pex14 signals but also enabled the observation of separate Pex3 and Pex14 signals that were localized in close proximity, suggesting a heterogeneous distribution of Pex3 and Pex14 across the organelles (Fig. 2B), as reported for Pexs in peroxisomes of human fibroblasts (28). Pex3/Pex14-labeled vesicles appeared mostly as circular, elliptical, or elongated structures. Their sizes ranged from a minimum of 80 nm (circular structures) to a maximum of 440 nm (elongated structures), and the average area of labeled vesicles was $0.013 \pm 0.009 \mu\text{m}^2$ (mean \pm SD, $n = 176$). Pex11-labeled vesicles were considerably smaller than hydrogenosomes visualized with antisuccinate dehydrogenase subunit B (SdhB) Ab (Fig. 2B). Neither hydrogenosomes nor ER structures and vesicles of the unstacked Golgi apparatus (29) colocalized with Pex11-1-labeled vesicles. Immunoelectron microscopy indicated

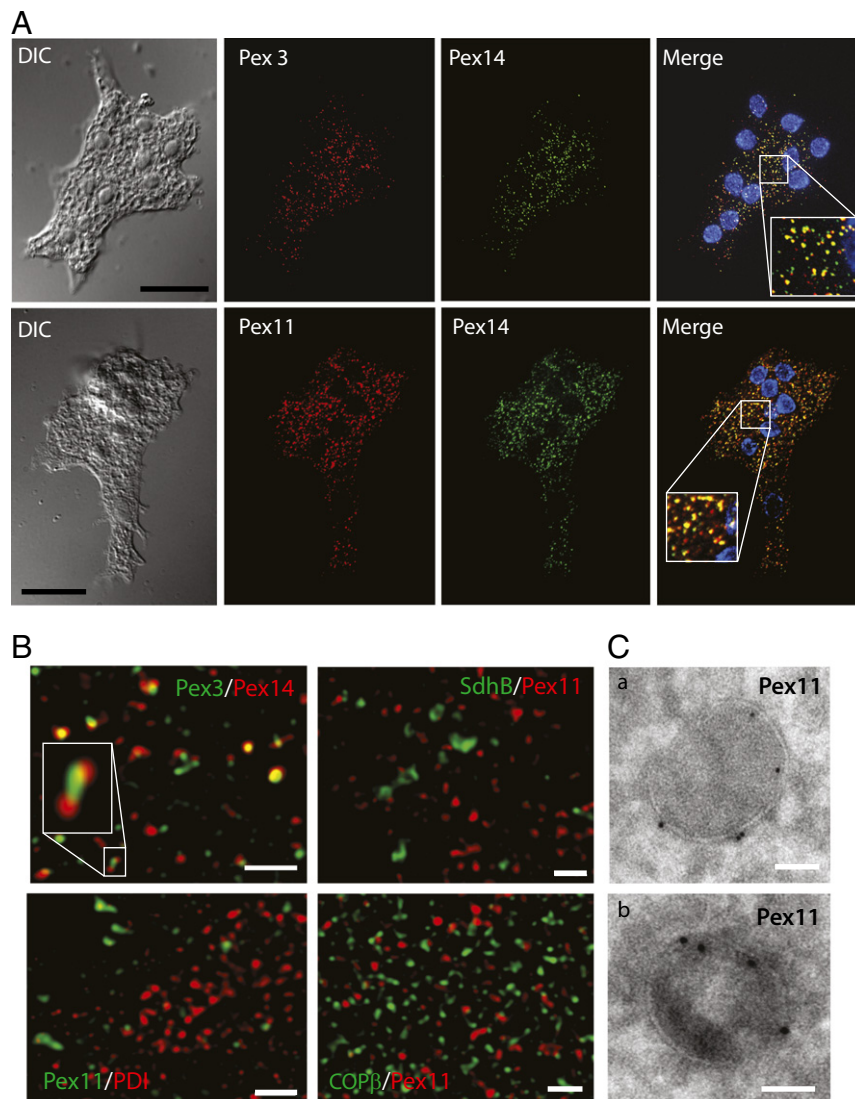


Fig. 2. Pexs colocalized in organelles distinct from hydrogenosomes, the Golgi apparatus, and the ER. (A) Confocal IF microscopy revealed colocalization of Pex3, Pex11-1, and Pex14. DIC, differential interference contrast. (Scale bar: 100 μm .) (B) STED microscopy analysis of Pex3, Pex11-1, Pex14, the hydrogenosomal marker succinate dehydrogenase (SdhB), the ER marker protein disulfide isomerase (PDI), and the Golgi marker COP β -subunit (COP β), which were visualized using specific rat or rabbit polyclonal Abs. (Scale bar: 500 nm.) (C) Immunoelectron microscopy detection of Pex11-1 on *M. balamuthi* cryosections (a and b) using a polyclonal anti-Pex11-1 Ab and protein A conjugated with 5 nm gold particles. (Scale bar: 50 nm.)

that Pex11-1 is associated with a single membrane that surrounds vesicles with a dark matrix (Fig. 2C and *SI Appendix*, Fig. S4). The average size of these vesicles was 143 ± 38 nm (mean \pm SD, $n = 20$) in diameter, which is within the range of sizes estimated by STED analysis. The number of putative peroxisomes in *M. balamuthi* was counted with Pex14 as a marker using confocal IF. We found 60 ± 20 peroxisomes per $100 \mu\text{m}^2$ (mean \pm SD, $n = 20$), which is comparable to the 50 to 119 glycosomes per $100 \mu\text{m}^2$ determined in *Trypanosoma brucei* (30, 31) but higher than the 15.6 peroxisomes per $100 \mu\text{m}^2$ found in human fibroblasts (32).

Finally, we selected MbPex14 to experimentally test its ability to target peroxisomes. Because *M. balamuthi* is not amenable to transformation, we used the yeast strain BY4742:POX1-EGFP expressing the integrated GFP-tagged peroxisomal marker protein acyl-CoA oxidase (Pox1), and this strain was transformed with a plasmid that allowed expression of MbPex14 fused with the fluorescent protein mCherry. Fluorescence microscopy of transformed cells cultivated with oleate to stimulate peroxisome formation showed that MbPex14 was incorporated into the Pox1-labeled peroxisomes. Western blot analysis of cellular fractions revealed that the majority of MbPex14 and Pox1 was present in the large granule fraction (LGF) (Fig. 3).

Matrix Proteins of *M. balamuthi* Peroxisomes. To identify the matrix components of the *M. balamuthi* peroxisomes, we initially searched for proteins with predicted PTS1 and PTS2 signals in the *M. balamuthi* genome. These searches revealed 664 proteins that possessed the predicted C-terminal PTS1 motif and 103 proteins with a putative PTS2 motif located within 100 amino acids (aa) of the N-terminal domain (*Dataset S1*). Next, we performed quantitative proteomic analysis of *M. balamuthi* subcellular fractions using the localization of organelle proteins by isotope tagging (LOPIT) method (33). Altogether, we identified 2,437 proteins in all fractions (*Dataset S1*). The distribution of peroxisomal proteins was traced using 5 membrane Pexs (*MbPex3*, *MbPex10*, *MbPex11-1*, *MbPex12*, and *MbPex14*) (Fig. 4). The peroxisomal markers had a higher relative abundance in fractions distinct from the marker proteins for hydrogenosomes and the ER (*SI Appendix*, Fig. S5 and *Dataset S1*). Next, we searched for proteins with quantitative distributions in cellular fractions similar to those determined for marker Pexs, which allowed the identification of 1,119 proteins, including additional Pexs: *MbPex2*, *MbPex5*, *MbPex11-1*, *MbPex13*, and *MbPex16*. Of these 1,119 proteins, 57 proteins contained a predicted PTS1/PTS2 motif (Fig. 4 and *Dataset S1*). Then, we removed 6 ribosomal proteins that are unlikely to reside in peroxisomes. The final dataset (*SI Appendix*, Table S3) contained 51 putative peroxisomal matrix proteins, of which 12 candidates were selected to validate their cell localization in yeast. Hydrogenosomal D-lactate dehydrogenase (D-LDH-M) and cytosolic NifU were used as controls for targeting specificity (18, 19). The proteins were expressed with an N-terminal mCherry tag in the yeast strain BY4742:POX1-EGFP. Fluorescence microscopy revealed that eight peroxisomal candidates (D-LDH; pantetheine-phosphate adenylyltransferase, PPAT; nudix hydrolase, Nudt; peroxisomal processing peptidase, PPP; inorganic pyrophosphatase 2, IPP-2; inositol dehydrogenase, IDH; tagatose-6-phosphate kinase, T6PK; and fumarylacetoacetate hydrolase, FAH) and the Pox1 marker colocalized together in yeast peroxisomes (Fig. 5 and *SI Appendix*, Table S3). As expected, NifU was localized to the cytosol, and D-LDH-M was localized to mitochondria. To investigate whether putative matrix proteins were translocated into the peroxisomal matrix, we performed a protein protection assay using D-LDH and PPAT. MbPex14 was used as a membrane marker. Both matrix proteins were fully membrane protected upon the treatment of LGF with proteinase K, whereas MbPex14 was partially cleaved (*SI Appendix*, Fig. S6).

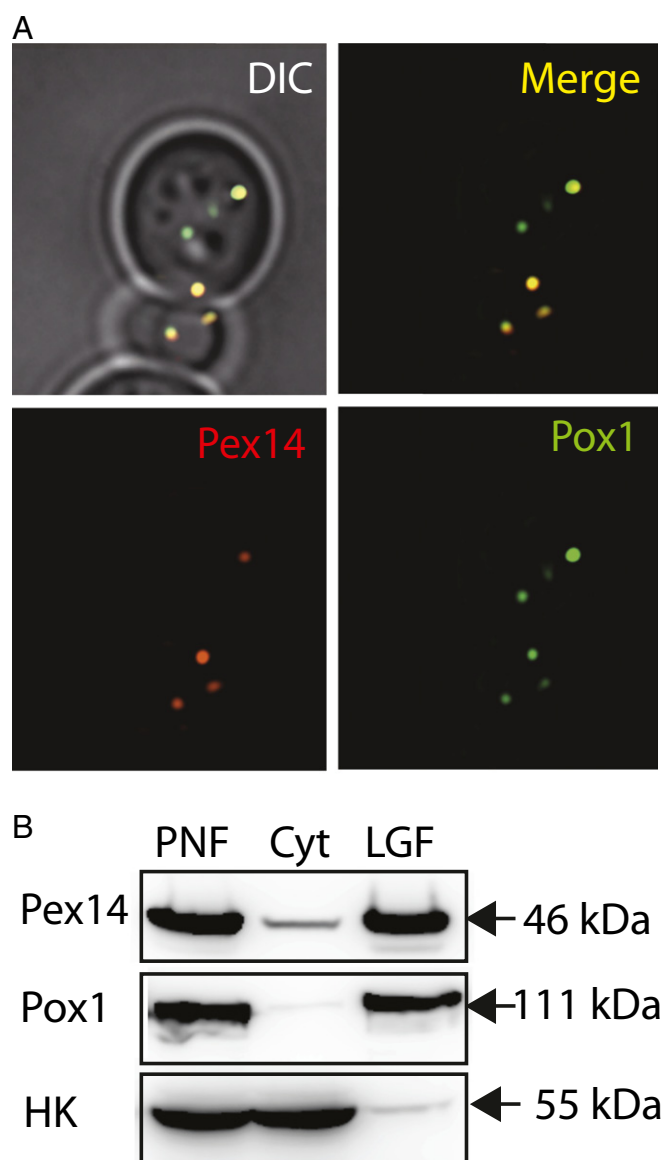


Fig. 3. MbPex14 is incorporated into *S. cerevisiae* peroxisomes. (A) Fluorescence microscopy of yeast expressing GFP-tagged Pox1 (peroxisomal marker) and mCherry-tagged Pex14 of *M. balamuthi*. DIC, differential interference contrast. Cells visualized at 1000 \times magnification. (B) Western blot analysis of the yeast postnuclear fraction (PNF), cytosol (Cyt), and large granule fraction (LGF). HK, hexokinase (cytosolic marker).

Although the ability of the tested *M. balamuthi* proteins to target yeast peroxisomes strongly supports their peroxisomal localization in *M. balamuthi*, we developed a specific polyclonal Ab against IPP-2 for its direct localization in *M. balamuthi*. IPP is particularly interesting because there are multiple paralogs of this enzyme with either N-terminal mitochondrial targeting sequences (IPP-1) (18) or a predicted C-terminal PTS1 (IPP-2, IPP-3, and IPP-4) (*SI Appendix*, Fig. S7). STED microscopy revealed that IPP-2 clearly colocalizes with Pex11-1-labeled vesicles (PCC $r = 0.53$). IPP-2 labeled smaller mostly round structures surrounded by the Pex11-1 signal (Fig. 6A). Elongated structures labeled with Pex11-1 were often associated with two separate blobs of IPP-2, which likely represent dividing peroxisomes. Immunoelectron microscopy revealed the presence of IPP-2 in the matrix of vesicles with a morphology similar to those in which we detected Pex11-1 (Figs. 2C and 6B and *SI Appendix*, Fig. S8). In double-labeling experiments,

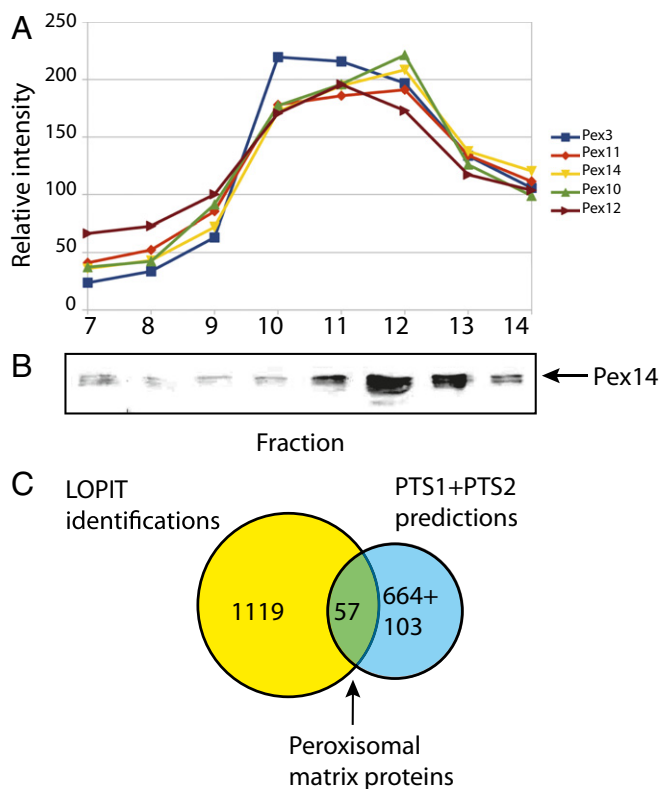


Fig. 4. Prediction of peroxisomal proteins using LOPIT and PTS1/PTS2 predictions. (A) Relative TMT intensities for marker Pexs in cellular fractions 7 to 14. (B) Western blot analysis of fractions 7 to 14 using an anti-Pex14 Ab. (C) Venn diagram depicting the intersection between the proteins identified by LOPIT and the proteins with a predicted PTS1/PTS2. Relative TMT intensities for hydrogenosomal and ER proteins are given in *SI Appendix, Fig. S3*.

both IPP-2 and Pex11-1 were observed in the matrix and at the membrane, respectively, within the same organelle (Fig. 6C).

Next, we tested whether the targeting of selected *M. balamuthi* candidates to peroxisomes was dependent on the predicted PTS1 signal. The proteins were expressed in BY4742:POX1-EGFP yeast with or without PTS1, and their presence was traced in the cytosolic fraction and LGF by Western blot. All proteins with truncated C-termini were significantly reduced or absent in the LGF (Fig. 7). The PTS1 motif of proteins that were successfully targeted to peroxisomes included two proteins (Nudt and PPP) with the classic C-terminal SKL tripeptide, whereas other proteins possessed its variants, which included ARL (α -LDH), AKL (IPP-2, IDH, and PPAT), SRL (T6PK), and ARM (FAH). Based on these experiments, the PTS1 consensus sequence of *M. balamuthi* peroxisomal proteins appeared to be [S/A][R/K][L/M], which corresponds well to the broader consensus sequence [S/A/P/C][K/R/H][L/M/I] established for peroxisomes of metazoans, plants, and glycosomes of kinetoplastids (34, 35).

The identification of PPP among peroxisomal proteins of *M. balamuthi* is noteworthy, as it is believed that PPP is present only in multicellular organisms (36). Our homology searches in available genomes of amoebozoans revealed the presence of PPP orthologs in mycetozoans such as *D. discoideum* and in *Acanthamoeba castellanii*; however, we did not find PPP in entamoebids (*SI Appendix, Fig. S9*). PPP was also detected in selected members of the eukaryotic supergroups SAR and Excavata. In the latter group, PPP was found in *Naegleria*, but it was absent in kinetoplastid glycosomes. Phylogenetic analysis revealed a monophyletic origin of PPPs, including *M. balamuthi*, other unicellular eukaryotes, and the known PPPs Tysnd1

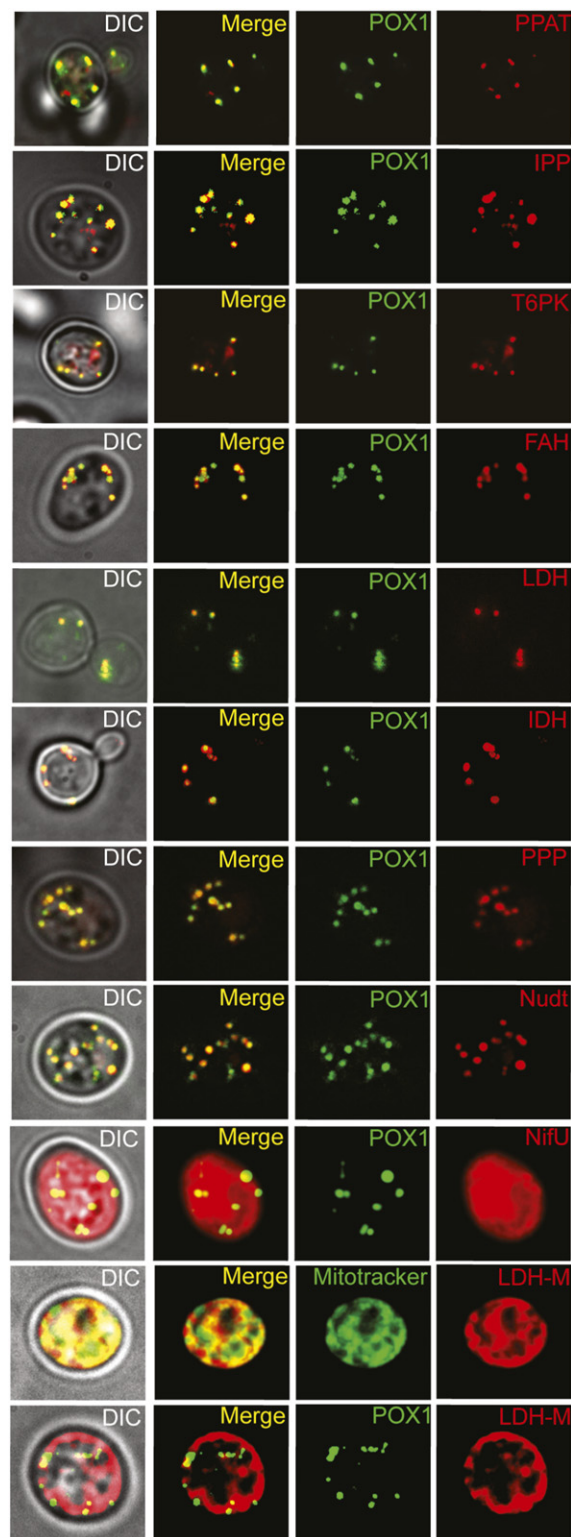


Fig. 5. Localization of *M. balamuthi* peroxisomal matrix candidates in *S. cerevisiae* peroxisomes. GFP-tagged Pox1 was used as a peroxisomal marker. mCherry-tagged *M. balamuthi* proteins: FAH, fumarylacetoacetate hydrolase; IPP, inorganic pyrophosphatase 2; Nudt, nudix hydrolase; T6PK, tagatose-6-phosphate kinase; LDH, lactate dehydrogenase; PPAT, pantetheine-phosphate adenylyltransferase; IDH, inositol dehydrogenase; PPP, peroxisomal processing peptidase; NifU, cytosolic marker; and LDH-M, hydrogenosomal paralog of lactate dehydrogenase. DIC, differential interference contrast. Cells visualized at 1000 \times magnification.

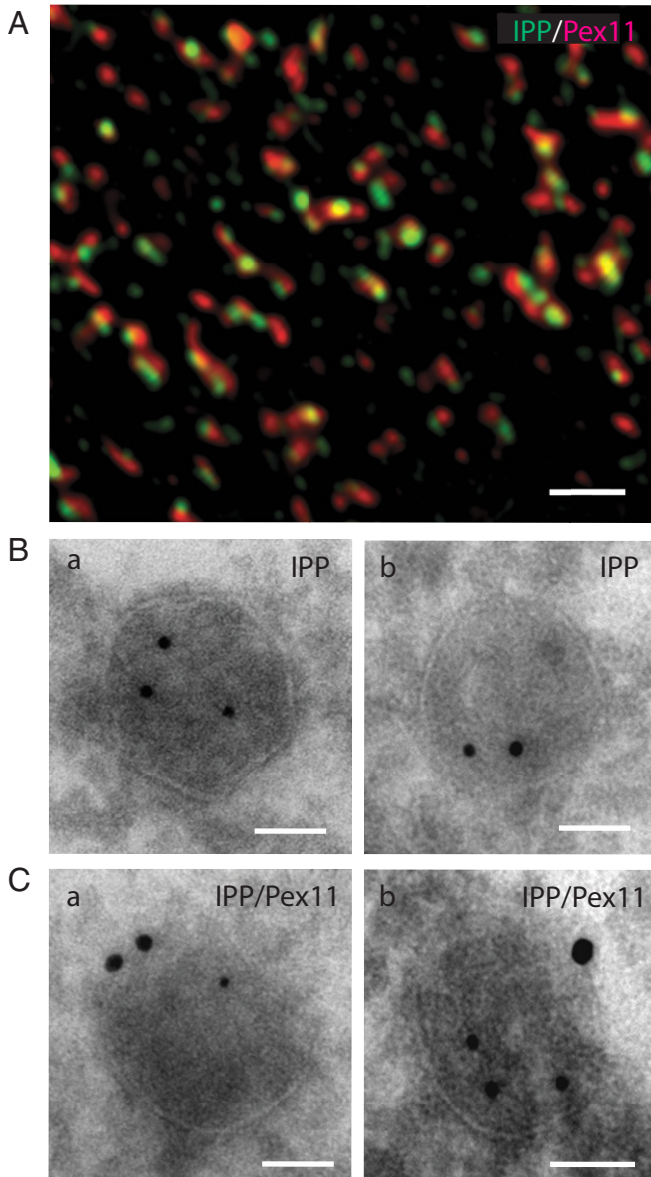


Fig. 6. Immunolocalization of the peroxisomal matrix protein IPP-2 in *M. balamuthi*. (A) STED microscopy using specific polyclonal Abs against IPP-2 (green) and Pex11-1 (red). (Scale bar: 500 nm.) (B) Immunoelectron microscopy detection of IPP-2 on *M. balamuthi* cryosections (a and b) using a rat polyclonal anti-IPP-2 Ab and anti-rat IgG conjugated with 10 nm gold particles. (C) Colocalization of IPP-2 and Pex11 (a and b). IPP-2 was detected using a rat polyclonal anti-IPP-2 Ab and anti-rat IgG conjugated to 10 nm gold particles, Pex11-1 was detected using a rabbit polyclonal anti-Pex11-1 Ab with secondary anti-rabbit IgG conjugated with 15 nm gold particles. (Scale bar: 50 nm.)

and Deg15 in metazoans and plants, respectively, that formed group IV (*SI Appendix*, Fig. S9). All members of this group possessed PTS1, mostly with the SKL tripeptide, including the *M. balamuthi* PPP.

Predicted Metabolic Pathways. Neither bioinformatic nor proteomic analyses identified any component of typical peroxisomal pathways associated with oxygen metabolism, such as β -oxidation or catalase (*Dataset S1* and *SI Appendix*, Table S4). However, we identified components of acetyl-CoA metabolism, pyrimidine biosynthesis, and carbohydrate metabolism (Fig. 8 and *SI Appendix*, Table S4). PPAT is a candidate that is potentially involved in

the regulation of peroxisomal CoA levels. This enzyme catalyzes the second-to-last step in CoA synthesis (Fig. 8 and *SI Appendix*, Fig. S10). The following step of CoA synthesis is catalyzed by

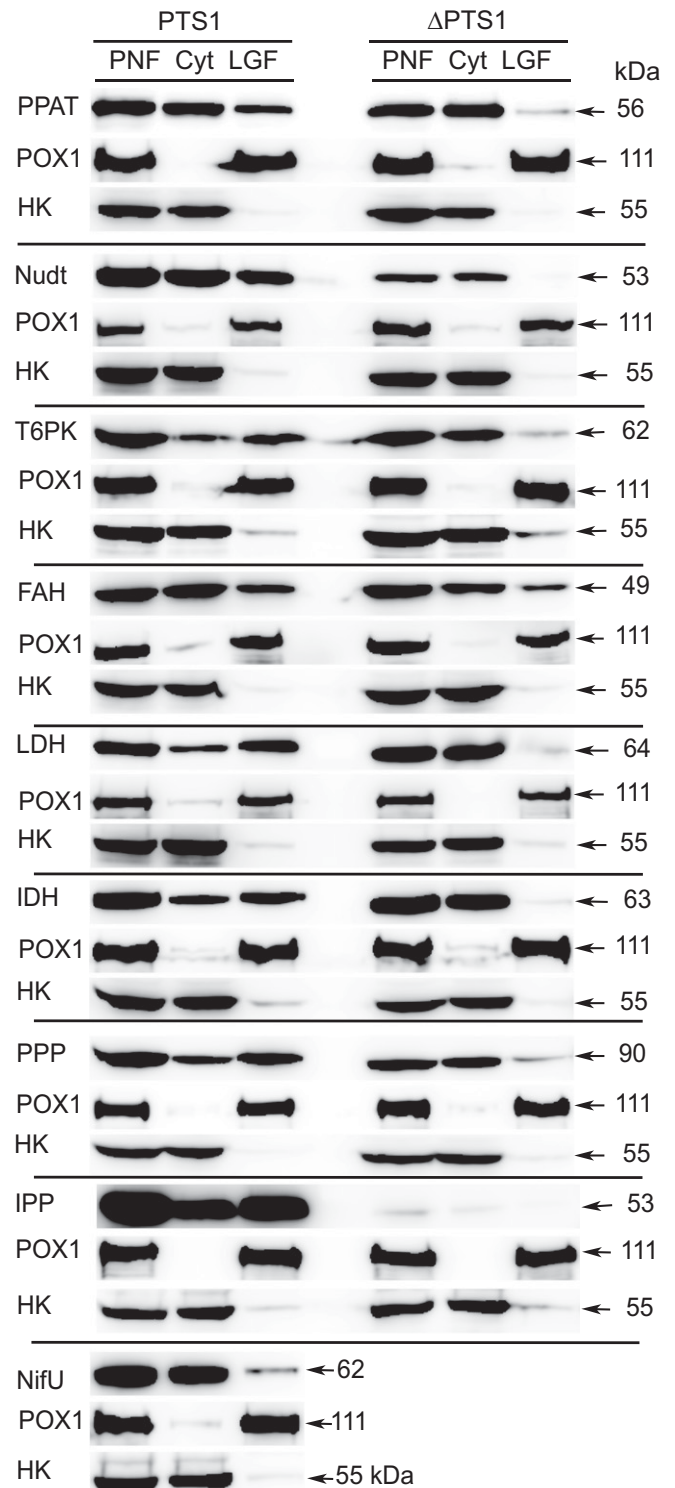


Fig. 7. PTS1-dependent import of *M. balamuthi* proteins into *S. cerevisiae* peroxisomes. *M. balamuthi* peroxisomal matrix proteins with the PTS1 signal or without the 13 C-terminal aa (Δ PTS1) were expressed in yeast. *M. balamuthi* proteins were detected using an anti-mCherry Ab, Pox1 was detected with an anti-GFP Ab, and hexokinase (HK) was used as a cytosolic marker and detected using anti-*S. cerevisiae* HK Ab.

dephospho-CoA kinase (DPCK), which is encoded in the genome (m51a1_g647); however, DPCK was not identified in the proteome. CoA and ATP are required for the activation of fatty acids by acyl-CoA synthetase (ACS). We identified two ACS paralogs with LOPIT support; however, only ACS-1 possessed a predicted PTS1 motif (*SI Appendix, Fig. S11*) (37). ACS is linked with IPP, which cleaves PPI to prevent the reverse reaction. Indeed, we identified two IPP paralogs (IPP-2 and IPP-3) in *M. balamuthi* peroxisomes (*SI Appendix, Table S3*). Pyrimidine biosynthesis includes the conversion of dihydroorotate to uridylylate. This pathway requires the activities of three enzymes: dihydroorotate dehydrogenase (DHODH), orotate phosphoribosyl transferase (OPRT), and orotidine-5'-monophosphate decarboxylase (OMPDC). In the *M. balamuthi* genome, we identified a single gene encoding a fused protein with PTS1 (ANL), which contains domains for all three enzymes, including an N-terminal DHODH domain, followed by OPRT and OMPDC (*SI Appendix, Fig. S12*). OPRT and OMPDC are often fused in eukaryotes to form a single peroxisomal protein, whereas DHODH functions as a separate protein, either in the cytosol (bacterial class 1) or in mitochondria (class 2) (38) (*SI Appendix, Fig. S12*). Therefore, we performed phylogenetic analysis of the *M. balamuthi* OMPDC-OPRT-DHODH domains. All three components appeared to be of bacterial origin, including DHODH, which clustered with class 1A orthologs present in bacteria, kinetoplastids, and diatoms (*SI Appendix, Fig. S13 A-C*). The activity of DHODH class 1A is dependent on fumarate as an electron acceptor that could be produced in *M. balamuthi* by peroxisomal FAH or imported from the cytosol, where fumarate is formed by fumarase,

which is encoded in the *M. balamuthi* genome (m51a1_g4783) (Fig. 8). OPRT activity required 5-phosphoribosyl-1-pyrophosphate (PRPP), which is formed from D-ribose phosphate by the activity of ribose-5-phosphate diphosphokinase. We predicted two alternative pathways for the formation of D-ribose phosphate via either glucose-6-phosphate dehydrogenase or the ADP ribose diphosphatase activity of Nudt. However, nudix hydrolases represent a large protein family with various functions, and the protein sequence analysis of *MbNudt* did not allow us to unequivocally predict its substrate specificity (*SI Appendix, Fig. S14*). Carbohydrate metabolism includes the putative D-LDH, IDH, and T6PK with a PTS1 and malate dehydrogenase (MDH) with a PTS2 (Fig. 8 and *SI Appendix, Table S3* and Figs. S15 and S16). Altogether, these results indicate the presence of peroxisomes in the anaerobic protist *M. balamuthi*.

Discussion

Free-living *M. balamuthi* is a eukaryote characterized as possessing both an anaerobic type mitochondria (hydrogenosomes) and peroxisomes. These “anaerobic” peroxisomes lack metabolic pathways generating ROS, such as β -oxidation, as well as ROS-detoxifying enzymes, including catalase. However, they contain components of acyl-CoA and carbohydrate metabolism and pyrimidine and CoA biosynthesis. *M. balamuthi* possesses a set of 14 Pexs required for the biogenesis of peroxisomes and machinery for the targeting of proteins to the peroxisome via PTS1/PTS2 and PPP.

In other eukaryotes with anaerobic forms of mitochondria, peroxisomes are believed to be absent (15, 16). Two proteins (ACS and Pex-4) were recently suggested to represent putative peroxisomal remnants in *Giardia intestinalis*, but their significance needs to be clarified (39). In addition to anaerobes, the absence of peroxisomes was noted in parasitic organisms with unconventional mitochondria such as *Plasmodium falciparum* and related species of the Aconoidasida group that live in aerobic niches (40, 41). However, during erythrocytic stages, *Plasmodium* relies mainly on cytosolic glycolysis, similar to anaerobes. In addition, the absence of peroxisomes has been predicted in several lineages of parasitic helminths (42), most of which adapted to oxygen-poor environments and can gradually replace the TCA cycle and aerobic phosphorylation with anaerobic energy metabolism (43). The identification of peroxisomes in anaerobic *M. balamuthi* is thus unprecedented, considering the correlation between the adaptation of mitochondria to anaerobiosis and the loss of peroxisomes.

Four peroxisomal proteins, Pex3, Pex10, Pex12, and Pex19, which are considered unequivocal in silico peroxisomal markers (2), and 10 other Pexs (Pex1, Pex2, Pex5, Pex6, Pex7, Pex11-1, Pex11-2, Pex14, and Pex16), were identified in the *M. balamuthi* genome. Ten of 14 predicted Pexs were identified by proteomics, which indicates that these proteins are available for peroxisome biogenesis. Targeting of *M. balamuthi* matrix proteins with PTS1 to peroxisomes was experimentally confirmed in *S. cerevisiae*. Notably, *M. balamuthi* peroxisomes seem to retain the PTS2 system, which is not as universally distributed as PTS1 (42, 44–46). As in other peroxisome-bearing organisms, ubiquitination-dependent recycling of PTS receptors likely occurs in anaerobic peroxisomes with contributions from the Pex2/10/12 RING-finger complex and the AAA-type ATPases Pex1/6. Although we did not identify a Pex4 ortholog in *M. balamuthi*, Pex4 could be functionally replaced by other enzymes with the required activity. For example, in humans, Pex4 is functionally replaced by the related ubiquitin-conjugating proteins Ubc5a/b/c (47). *M. balamuthi* possesses all three members of the machinery that target proteins to the peroxisomal membrane, *MbPex3*, *MbPex16*, and *MbPex19*. Pex19 is a soluble shuttling receptor that was not found in the LOPIT data; however, a single gene encoding a putative *MbPex19* was found in the *M. balamuthi*

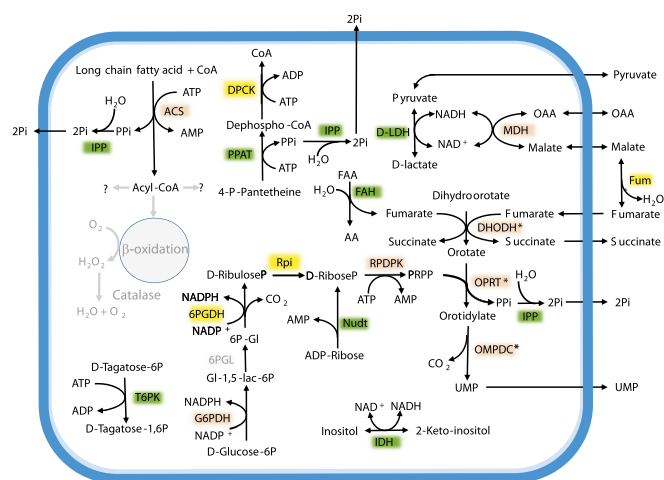


Fig. 8. Predicted metabolism of *M. balamuthi* anaerobic peroxisomes. The green box indicates enzymes that were localized in yeast peroxisomes; the pink box indicates proteins with a predicted PTS1/2 and support for peroxisomal localization by LOPIT (*SI Appendix, Table S5*); the yellow box indicates genes identified in the *M. balamuthi* genome; and gray indicates enzymes: ACS, Acyl-CoA synthetase; DPCK, dephospho-CoA kinase; FAH, fumarylacetoacetate hydrolase; Fum, fumarase; G6PDH, glucose-6-phosphate dehydrogenase; IDH, inositol dehydrogenase; IPP, inorganic pyrophosphatase; 6PGDH, 6-phosphogluconate dehydrogenase; DHODH, dihydroorotate dehydrogenase; OPRT, orotate phosphoribosyl transferase; OMPDC, orotidine-5'-monophosphate decarboxylase; MDH, malate dehydrogenase; Nudt, nudix hydrolase; 6-phosphogluconolactonase, 6PGL; 6-phosphogluconate dehydrogenase, 6PGDH; PPTAT, pantetheine-phosphate adenyltransferase; RDPK, ribose-5-phosphate diphosphokinase; ribose phosphate isomerase, Rpi; and T6PK, tagatose-6-phosphate kinase. Substrates are as follows: AA, acetoacetate; FAA, fumarylacetoacetate; PRPP, 5-phosphoribosyl-1-pyrophosphate; GI-1,5-lac-6P, D-glucono-1,5-lactone-6-phosphate; D-glucose-6P, OAA, oxalacetate; D-tagatose-6P, D-tagatose-6-phosphate; D-tagatose-1,6-bisphosphate; D-riboseP, D-ribose phosphate; and PPI, pyrophosphate. *Domains of the fusion enzyme DHODH-OPRT-OMPDC.

genome. Finally, *M. balamuthi* possesses two *MbPex11* paralogs that may facilitate the elongation of the peroxisomal membrane and the fission of the organelles (23, 24). Collectively, these findings show that *M. balamuthi* anaerobic peroxisomes contain a rather standard set of Pexs that are required for peroxisome biogenesis.

In contrast, most typical peroxisomal metabolic pathways, such as β -oxidation and hydrogen peroxide detoxification, are absent or highly reduced in *M. balamuthi*. We predicted the presence of ACS, which requires CoA as an obligate cofactor for the activation of medium- and long-chain fatty acids. This enzyme activates the substrate mainly for β -oxidation (37); however, because β -oxidation is absent in *M. balamuthi* and we did not find any other acyl-CoA-dependent degradation or biosynthetic pathways, the fate of acyl-CoA in anaerobic peroxisomes is currently unknown. Interestingly, anaerobic peroxisomes are likely involved in CoA synthesis via the PPAT and DPCK pathways. This is supported by intraperoxisomal localization of *MbPPAT*, which was predicted by LOPIT data and demonstrated by the protein protection assay, although the cell localization of DPCK in *M. balamuthi* remains unclear. In most eukaryotes, CoA synthesis is compartmentalized to the cytosol or plastids, and in mammalian cells, PPAT and DPCK form the fusion protein CoA synthase that is partially associated with the mitochondrial outer membrane (48, 49). The peroxisomal localization of CoA synthesis has not been reported thus far, although DPCK has been found in the *Arabidopsis* peroxisomal proteome (50).

There are several enzymes that could be involved in the maintenance of the redox balance of anaerobic peroxisomes. We can predict that MDH serves as a redox shuttle enzyme that oxidizes NADH, whereas the metabolites oxaloacetate and malate cross the peroxisomal membrane as they do in plant peroxisomes (51). In addition, there are two more enzymes that are known to mediate peroxisomal redox balance, NAD-dependent D-LDH and NADP-dependent glucose-6-phosphate dehydrogenase (G6PDH). D-lactate and malate exported from peroxisomes can be taken up by hydrogenosomes as substrates for hydrogenosomal D-LDH and MDH, respectively (18). G6PDH is an enzyme of the pentose phosphate pathway, which operates in the cytosol; however part of this pathway involving G6PDH was found in peroxisomes and glycosomes (52, 53).

Our data suggest that anaerobic peroxisomes possess a significant part of the de novo pyrimidine synthesis pathway. We found a unique fusion protein with three OMPDC-OPRT-DHODH domains that may convert dihydroorotate to uridylate. The DHODH domain corresponds to class 1A DHODH, which is typically a bacterial protein and is critical for the synthesis of pyrimidine under anaerobiosis using fumarate as an electron acceptor (54, 55). Most eukaryotes, including the aerobic amoeba *D. discoideum*, possess mitochondrial DHODH class 2 that provides electrons to ubiquinone (56, 57). Thus, the acquisition of class 1A DHODH by *M. balamuthi* is likely associated with the adaptation of this lineage to anaerobiosis. DHODH is known as a cytosolic or mitochondrial protein, whereas the next two steps converting orotate to uridylate (UMP) via the activities of OPRT and OMDC take place in peroxisomes or glycosomes (58). Interestingly, various combinations of OPRT and OMDC have been reported in fusion proteins from eukaryotes and bacteria (59). However, none of the known fusion proteins are fused with DHODH, as found in *M. balamuthi*, which may allow the conversion of dihydroorotate to uridylate in a single compartment.

Surprisingly, *M. balamuthi* peroxisomes contain a putative T6PK. This finding is supported by the LOPIT data, the presence of PTS1, and targeting to yeast peroxisomes. T6PK belongs to the ATP-dependent phosphosugar kinase (PfkB) family of mainly bacterial enzymes (60). In eukaryotes, T6PK was predicted to reside in the glycosome of *Leishmania major* (61), and

genes for T6PK are annotated in genomes of other *Leishmania* species, leptomonads and *Entamoeba* species.

Is there any relationship between anaerobic peroxisomes and hydrogenosomes? Analysis of peroxisomal matrix proteins revealed the presence of at least three enzymes (MDH, D-LDH, and IPP) with dual localization in peroxisomes and hydrogenosomes. Each enzyme seems to employ different mechanisms for dual protein targeting. Gene duplication and acquisition of different targeting signals are involved in the dual localization of D-LDH. It has been shown that *M. balamuthi* possesses two closely related copies of D-LDH (18). The first copy, D-LDH-M, is targeted to hydrogenosomes via N-terminal targeting sequence (NTS) (18). Here, we found that the second D-LDH copy possesses a C-terminal extension with a PTS1 signal, whereas the NTS is absent. When D-LDH was expressed with an N-terminal tag in yeast, the protein appeared in peroxisomes (this study), whereas previous localization of this D-LDH with a C-terminal tag that masked PTS1 resulted in cytosolic mislocalization (18). IPP is present in four copies (this work and ref. 18). IPP-1 is targeted to hydrogenosomes via NTS, whereas IPP-2, IPP-3, and IPP-4 possess PTS1. However, in contrast to D-LDH paralogs, the origin of hydrogenosomal and peroxisomal IPPs is different (18). IPP-1 is eukaryotic in origin and clusters together with the mitochondrial IPP in *Entamoeba histolytica*, whereas IPP-2 and -3 were most likely acquired by lateral gene transfer (LGT) from bacteria (18). Bioinformatic analysis of MDH predicted multiple signals within the N-terminal extension. There are three nearly identical copies with a hydrogenosomal NTS and a cleavage site for mitochondrial processing peptidase between the amino acid residues at positions 30 to 31 (18, 62). Peroxisomal localization of MDH is most likely driven by PTS2, which has been predicted at positions 11 to 19 within the NTS. Similar cotargeting to peroxisomes and mitochondria mediated by the bifunctional N-terminal domain was observed for the yeast catalase Cta1p (63).

The discovery of anaerobic peroxisomes in *M. balamuthi* provides a perspective for studies of the evolution of eukaryotes that live in the absence of oxygen. In the last two decades, great progress has been made in this direction. However, most studies have focused on investigations of anaerobic forms of mitochondria. Genomic and transcriptomic analyses of parasitic, endosymbiotic, and free-living protists revealed a functional continuum from strictly oxygen-dependent mitochondria via mitochondria of facultatively anaerobic organisms adapted to an anaerobic lifestyle to various extents to highly reduced mitochondria in parasitic species (64, 65). In contrast, peroxisomes were considered to be present in aerobes and absent in anaerobes based on few examples of anaerobic protists, such as *Trichomonas vaginalis*, *G. intestinalis*, and *E. histolytica*, in which peroxisomes have not been found (2, 66, 67). The discovery of peroxisomes in anaerobic *M. balamuthi* with hydrogenosomes and anaerobic energy metabolism may change this view and turn attention to the evolution of peroxisomes in anaerobic or facultatively anaerobic organisms. In particular, it would be interesting to investigate how peroxisomes and mitochondria coevolved, reflecting the presence of oxygen in the environment and reconsidering the role of oxygen and ROS in the evolution of peroxisomes.

Materials and Methods

Cell Cultivation. *M. balamuthi* (ATCC 30984) was maintained axenically in proteose peptone–yeast extract–glucose–cysteine (PYGC) medium at 24 °C (68). The *S. cerevisiae* strain BY4742:POX1-EGFP (69) was grown in a rich or selective medium as previously described (70). The formation of peroxisomes was stimulated using oleate medium containing 0.05% yeast extract and 0.1% oleic acid (71).

Bioinformatics. Genes encoding *M. balamuthi* Pexs were obtained by TBLAST searches of the draft *M. balamuthi* genome sequence available at the Orcae database (<https://bioinformatics.psb.ugent.be/orcae/>, refs. 19 and 72) using the orthologous sequences of *S. cerevisiae*. Searches for PTS1 in 14,841 predicted

proteins of *M. balamuthi* given in *SI Appendix, Table S3* were performed using PeroxisomeDB (<http://www.peroxisomedb.org/>) considering 12 C-terminal residues (73). PTS2 was locally predicted using the consensus sequence (R/K)-(L/I/V)-X5-(H/Q)-(L/A) (50, 74, 75). Transmembrane domains were predicted using TMHMM server v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>). Predicted peroxisomal proteins were annotated using the eggNOG database (<http://eggnogdb.embl.de/>).

Protein Expression and Antibodies. Recombinant *M. balamuthi* Pex14 and IPP-2 were produced in *Escherichia coli* BL21 Rosetta cells (Novagen) using the pET42b expression vector (Novagen). Partial sequences of Pex3 and Pex11-1 were produced using pETM11-SUMO2-GFP (76). Abs against Pex3, Pex11-1, Pex14, and IPP-2 were raised in rats and rabbits (Davids Biotechnologie GmbH). The other Abs included a rat polyclonal anti-PDI Ab (29), rat polyclonal anti-COPI β -subunit Ab (29), rat polyclonal anti-succinate dehydrogenase (Sdh) β -subunit Ab (19), rabbit polyclonal anti-hexokinase Ab (a kind gift from Doron Rapaport, University of Tübingen, Tübingen, Germany), and mouse monoclonal anti-GFP Ab (Santa Cruz Biotechnology, Inc.). Details of gene cloning and protein expression are given in *SI Appendix, Materials and Methods*.

Expression and Localization of Proteins in *S. cerevisiae*. Genes for *M. balamuthi* peroxisomal candidates were amplified and subcloned (*SI Appendix, Materials and Methods and Table S5*) into a modified pTVU100 vector (77) that allows protein expression in yeast with an N-terminal mCherry tag (pTVU100-N-mCherry). *S. cerevisiae* BY4742:POX1-EGFP (kindly provided by Zdena Palková, Charles University, Prague, Czech Republic) was transformed with the pTVU constructs using the lithium acetate/single-stranded carrier DNA/PEG method (78). Transformed strains were selected and maintained on uracil-free medium. Cells were incubated for 15 to 20 h in oleate medium prior to microscopy. For the investigation of PTS1-dependent protein targeting to yeast peroxisomes, the genes of interest were subcloned into the same vector without sequences coding for the 13 aa at the C-terminus.

Light Microscopy. For IF, *M. balamuthi* cells were processed as described (18). Images were obtained with a Leica SP8 confocal laser scanning microscope (Leica Microsystems). Images were deconvolved with Huygens Professional version 17.10 (Scientific Volume Imaging, The Netherlands, <https://svi.nl/HomePage>) and further processed using Fiji software (79). PCC was calculated using JACoP (80). The peroxisome number was determined in *M. balamuthi* cells using Pex14 as a marker. The number of labeled organelles and the corresponding cell area were determined per 20 cells using Icy software (81). To compensate for large variations in cell size, the number of peroxisomes was expressed per 100 μm^2 .

To observe GFP- and mCherry-tagged proteins in *S. cerevisiae* transformants, living yeast in synthetic medium were immobilized in 0.2% agarose on slides. Fluorescence signals of GFP, mCherry, and 4',6-diamidino-2-phenylindole (DAPI) were detected using a Nikon Eclipse TiE with an Andor iXon Ultra DU897 camera (Nikon Instruments, Inc.) and processed as described above.

STED microscopy was performed as described previously (82). Images were obtained using an Abberior STED 775 QUAD scanning microscope (Abberior Instruments GmbH) equipped with a Nikon CFI Plan Apo Lambda objective and deconvolved with Huygens Professional version software 17.04 using the classic maximum likelihood estimation algorithm. The secondary Abs were Abberior STAR 580 anti-rat or rabbit Abs and Abberior STAR 635p anti-rat or anti-rabbit Abs.

Transmission Immunoelectron Microscopy. Immunogold labeling of thawed cryosections (Tokuyasu technique) was used to detect Pex11-1 and IPP-2 as described (*SI Appendix, Materials and Methods*) (83). The sections were incubated for 30 min at room temperature with a rabbit polyclonal anti-PEX11-1 Ab, a polyclonal rat IPP-2 Ab, or a mixture of both Abs. For detection, we used various secondary probes conjugated to gold nanoparticles

of different sizes: protein A conjugated with 5 nm nanoparticles (University Medical Centre, Utrecht, The Netherlands), goat anti-rabbit IgG-15 nm (BBI Solutions), and goat anti-rat IgG conjugated with 10 nm nanoparticles (BBI Solutions). The samples were observed using a JEOL 1010 transmission electron microscope.

Fractionation of *M. balamuthi* Cells. Subcellular fractions of *Mastigamoeba* were obtained by differential and Percoll gradient centrifugation of the cell homogenate. Fifteen fractions (0.5 mL each) were collected, washed, and analyzed by Western blot using Ab against Pex14. Eight fractions with a Pex14 signal were submitted for quantitative mass spectrometry. Details of cellular fractionation are given in *SI Appendix, Materials and Methods*.

Isolation of the LGF from *S. cerevisiae*. The yeast LGF was obtained using previously described methods (84) with modifications given in *SI Appendix, Materials and Methods*.

LOPIT. Protein samples (50 μg) were processed and trypsin digested as described previously (85). Tandem mass tag (TMT) reagents were added to each sample according to the manufacturer's protocol (Thermo Scientific Pierce), and after 60 min, the reaction was stopped by the addition of 0.5% hydroxylamine. Peptides were separated using a C18 column (Kinetex 1.7 μm , EVO C18, 150 \times 2.1 mm). Thirty-four fractions were collected, and each of the 2 most distant fractions was pooled into 17 fractions as described previously (86). Data acquisition was performed as described previously using a Thermo Orbitrap Fusion (Q-OT- qIT, Thermo Fisher Scientific) (85).

Raw data were processed in Proteome Discoverer 2.1. (Thermo Fisher Scientific). Searches were performed against a local *M. balamuthi* protein database with 14,841 entries and a common contaminant database. To predict the cellular localization of peroxisomal proteins, marker proteins with the expected localization were used for the detection of fractions with maximum TMT values. The following markers were used—peroxisomes: Pex3, Pex10, Pex11, Pex12, and Pex14; hydrogenosomes: Tom40, voltage-dependent anion channel (VDAC), Cpn60, Cpn10, Sdh subunits A and C, ADP/ATP carrier, hydrogenosomal Nifs, and T- and H-protein of the glycine cleavage system (Gcs); and ER: Sec61, Sec22, reticulon, signal recognition particle receptor subunit α (SRPRA), calnexin, and PDI. To filter putative peroxisomal matrix proteins, we used the following criteria: the protein was detected by more than one peptide in at least one of two independent fractionations with a maximum TMT value within the range of peroxisomal markers; it possesses predicted PTS1 or PTS2 consensus sequences; and it lacked predicted transmembrane domains. More details are given in *SI Appendix, Materials and Methods*.

Data Availability. MaxQuant results were uploaded to the PRIDE partner repository (<https://www.ebi.ac.uk/pride/archive/>) with the dataset identifier PXD014205. The sequences reported in this paper have been deposited in the online resource for community annotation of eukaryotes (ORCAE, <https://bioinformatics.psb.ugent.be/orcae/>) and are given in *Dataset S1*.

ACKNOWLEDGMENTS. This work was supported by the Czech Science Foundation (16-06123S), Národní Program Udržitelnosti II (LQ1604) provided by the Ministry of Education, Youth, and Sport of the Czech Republic, Project CePaViP (CZ.02.1.01/0.0/0.0/16_019/0000759), provided by European Regional Development Fund, and the MICOBION project funded by European Union H2020 (810224). T.L. was supported by Charles University Grant Agency 170218. We thank the Imaging Methods Core Facility at BIOCEV for STED microscopy; the Core Facility of the Institute of Parasitology and Biology Centre of the Czech Academy of Sciences (České Budějovice, Czech Republic) for electron microscopy, both supported by the Czech-BiolImaging RI project LM2015062; the core facility OMICS Proteomics BIOCEV for MS analysis; and Michaela Marcinková and Eliška Kočířová for technical assistance.

1. T. Gabaldón, Peroxisome diversity and evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 765–773 (2010).
2. A. Schlüter *et al.*, The evolutionary origin of peroxisomes: An ER-peroxisome connection. *Mol. Biol. Evol.* **23**, 838–845 (2006).
3. C. De Duve, P. Baudhuin, Peroxisomes (microbodies and related particles). *Physiol. Rev.* **46**, 323–357 (1966).
4. J. J. Smith, J. D. Aitchison, Peroxisomes take shape. *Nat. Rev. Mol. Cell Biol.* **14**, 803–817 (2013).
5. L. Pieuchot, G. Jedd, Peroxisome assembly and functional diversity in eukaryotic microorganisms. *Annu. Rev. Microbiol.* **66**, 237–263 (2012).
6. K. Bolte, S. A. Rensing, U. G. Maier, The evolution of eukaryotic cells from the perspective of peroxisomes: Phylogenetic analyses of peroxisomal beta-oxidation enzymes support mitochondria-first models of eukaryotic cell evolution. *BioEssays* **37**, 195–203 (2015).

7. D. Speijer, Evolution of peroxisomes illustrates symbiogenesis. *BioEssays* **39**, 1700050 (2017).
8. C. de Duve, The origin of eukaryotes: A reappraisal. *Nat. Rev. Genet.* **8**, 395–403 (2007).
9. T. Gabaldón, A metabolic scenario for the evolutionary origin of peroxisomes from the endomembranous system. *Cell. Mol. Life Sci.* **71**, 2373–2376 (2014).
10. T. Gabaldón, Evolutionary considerations on the origin of peroxisomes from the endoplasmic reticulum, and their relationships with mitochondria. *Cell. Mol. Life Sci.* **71**, 2379–2382 (2014).
11. M. Schrader, J. Costello, L. F. Godinho, M. Islinger, Peroxisome-mitochondria interplay and disease. *J. Inher. Metab. Dis.* **38**, 681–702 (2015).
12. M. Fransen, M. Nordgren, B. Wang, O. Apanasets, Role of peroxisomes in ROS/RNS-metabolism: Implications for human disease. *Biochim. Biophys. Acta* **1822**, 1363–1373 (2012).

13. M. P. Menendez-Gutierrez, T. Roszer, M. Ricote, Biology and therapeutic applications of peroxisome proliferator-activated receptors. *Curr. Top. Med. Chem.* **12**, 548–584 (2012).
14. A. Sugiura, S. Mattie, J. Prudent, H. M. McBride, Newly born peroxisomes are a hybrid of mitochondrial and ER-derived pre-peroxisomes. *Nature* **542**, 251–254 (2017).
15. E. Gentekaki *et al.*, Extreme genome diversity in the hyper-prevalent parasitic eukaryote *Blastocystis*. *PLoS Biol.* **15**, e2003769 (2017).
16. T. Gabaldón, M. L. Ginger, P. A. M. Michels, Peroxisomes in parasitic protists. *Mol. Biochem. Parasitol.* **209**, 35–45 (2016).
17. T. M. Embley, W. Martin, Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630 (2006).
18. E. Nývltová *et al.*, Lateral gene transfer and gene duplication played a key role in the evolution of *Mastigamoeba balamuthi* hydrogenosomes. *Mol. Biol. Evol.* **32**, 1039–1055 (2015).
19. E. Nývltová *et al.*, NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of *Mastigamoeba balamuthi*. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7371–7376 (2013).
20. H. Hayashi, T. Suga, Some characteristics of peroxisomes in the slime mold, *Dictyostelium discoideum*. *J. Biochem.* **84**, 513–520 (1978).
21. T. Pánek *et al.*, First multigene analysis of *Archamoebae* (*Amoebozoa*: *Conosa*) robustly reveals its phylogeny and shows that *Entamoebidae* represents a deep lineage of the group. *Mol. Phylogenet. Evol.* **98**, 41–51 (2016).
22. C. Neufeld *et al.*, Structural basis for competitive interactions of Pex14 with the import receptors Pex5 and Pex19. *EMBO J.* **28**, 745–754 (2009).
23. J. Koch, C. Brocard, PEX11 proteins attract Mff and human Fis1 to coordinate peroxisomal fission. *J. Cell Sci.* **125**, 3813–3826 (2012).
24. H. Otera *et al.*, Mff is an essential factor for mitochondrial recruitment of Drp1 during mitochondrial fission in mammalian cells. *J. Cell Biol.* **191**, 1141–1158 (2010).
25. C. Williams, M. van den Berg, R. R. Sprenger, B. Distel, A conserved cysteine is essential for Pex4p-dependent ubiquitination of the peroxisomal import receptor Pex5p. *J. Biol. Chem.* **282**, 22534–22543 (2007).
26. S. Goto, S. Mano, C. Nakamori, M. Nishimura, *Arabidopsis* ABERRANT PEROXISOME MORPHOLOGY9 is a peroxin that recruits the PEX1-PEX6 complex to peroxisomes. *Plant Cell* **23**, 1573–1587 (2011).
27. V. C. Kalel, P. Mäser, M. Sattler, R. Erdmann, G. M. Popowicz, Come, sweet death: Targeting glycosomal protein import for antitrypanosomal drug development. *Curr. Opin. Microbiol.* **46**, 116–122 (2018).
28. S. Galiani *et al.*, Super-resolution microscopy reveals compartmentalization of peroxisomal membrane proteins. *J. Biol. Chem.* **291**, 16948–16962 (2016).
29. L. D. Barlow, E. Nývltová, M. Aguilar, J. Tachezy, J. B. Dacks, A sophisticated, differentiated Golgi in the ancestor of eukaryotes. *BMC Biol.* **16**, 27 (2018).
30. S. Bauer, J. C. Morris, M. T. Morris, Environmentally regulated glycosome protein composition in the African trypanosome. *Eukaryot. Cell* **12**, 1072–1079 (2013).
31. S. K. Banerjee, P. S. Kessler, L. Saveria, M. Parsons, Identification of trypanosomatid PEX19: Functional characterization reveals impact on cell growth and glycosome size and number. *Mol. Biochem. Parasitol.* **142**, 47–55 (2005).
32. C. Krause, H. Rosewich, A. Woehler, J. Gärtner, Functional analysis of PEX13 mutation in a Zellweger syndrome spectrum patient reveals novel homooligomerization of PEX13 and its role in human peroxisome biogenesis. *Hum. Mol. Genet.* **22**, 3844–3857 (2013).
33. T. P. Dunkley, R. Watson, J. L. Griffin, P. Dupree, K. S. Lilley, Localization of organelle proteins by isotope tagging (LOPIT). *Mol. Cell. Proteomics* **3**, 1128–1134 (2004).
34. R. Rucktäschel, W. Girzalsky, R. Erdmann, Protein import machineries of peroxisomes. *Biochim. Biophys. Acta* **1808**, 892–900 (2011).
35. J. Moyersoen, J. Choe, E. Fan, W. G. Hol, P. A. Michels, Biogenesis of peroxisomes and glycosomes: Trypanosomatid glycosome assembly is a promising new drug target. *FEMS Microbiol. Rev.* **28**, 603–643 (2004).
36. M. Helm *et al.*, Dual specificities of the glyoxysomal/peroxisomal processing protease Deg15 in higher plants. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11501–11506 (2007).
37. P. A. Watkins, J. M. Ellis, Peroxisomal acyl-CoA synthetases. *Biochim. Biophys. Acta* **1822**, 1411–1420 (2012).
38. T. Annoura, T. Nara, T. Makiuchi, T. Hashimoto, T. Aoki, The origin of dihydroorotate dehydrogenase genes of kinetoplastids, with special reference to their biological significance and adaptation to anaerobic, parasitic conditions. *J. Mol. Evol.* **60**, 113–127 (2005).
39. K. Acosta-Virgen *et al.*, *Giardia lamblia*: Identification of peroxisomal-like proteins. *Exp. Parasitol.* **191**, 36–43 (2018).
40. A. K. Ludewig-Klingner, V. Michael, M. Jarek, H. Brinkmann, J. Petersen, Distribution and evolution of peroxisomes in alveolates (*Apicomplexa*, *Dinoflagellates*, *Ciliates*). *Genome Biol. Evol.* **10**, 1–13 (2018).
41. A. B. Vaidya, M. W. Mather, Mitochondrial evolution and functions in malaria parasites. *Annu. Rev. Microbiol.* **63**, 249–267 (2009).
42. V. Žárský, J. Tachezy, Evolutionary loss of peroxisomes—Not limited to parasites. *Biol. Direct* **10**, 74 (2015).
43. M. Müller *et al.*, Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* **76**, 444–495 (2012).
44. L. L. Cross, H. T. Ebeed, A. Baker, Peroxisome biogenesis, protein targeting mechanisms and PEX gene functions in plants. *Biochim. Biophys. Acta* **1863**, 850–862 (2016).
45. N. H. Gonzalez *et al.*, A single peroxisomal targeting signal mediates matrix protein import in diatoms. *PLoS One* **6**, e25316 (2011).
46. A. M. Motley, E. H. Hettema, R. Ketting, R. Plasterk, H. F. Tabak, *Caenorhabditis elegans* has a single pathway to target matrix proteins to peroxisomes. *EMBO Rep.* **1**, 40–46 (2000).
47. C. P. Grou *et al.*, Members of the E2D (UbcH5) family mediate the ubiquitination of the conserved cysteine of Pex5p, the peroxisomal import receptor. *J. Biol. Chem.* **283**, 14190–14197 (2008).
48. S. Skrede, O. Halvorsen, Mitochondrial pantetheinophosphate adenyllyltransferase and dephospho-CoA kinase. *Eur. J. Biochem.* **131**, 57–63 (1983).
49. A. G. Tahiliani, J. R. Neely, Mitochondrial synthesis of coenzyme A is on the external surface. *J. Mol. Cell. Cardiol.* **19**, 1161–1167 (1987).
50. S. Reumann *et al.*, In-depth proteome analysis of *Arabidopsis* leaf peroxisomes combined with *in vivo* subcellular targeting verification indicates novel metabolic and regulatory functions of peroxisomes. *Plant Physiol.* **150**, 125–143 (2009).
51. I. Pracharoenwattana, J. E. Cornah, S. M. Smith, *Arabidopsis* peroxisomal malate dehydrogenase functions in β -oxidation but not in the glyoxylate cycle. *Plant J.* **50**, 381–390 (2007).
52. M. L. S. Güther, M. D. Urbaniak, A. Tavendale, A. Prescott, M. A. J. Ferguson, High-confidence glycosome proteome for procyclic form *Trypanosoma brucei* by epitope-tag organelle enrichment and SILAC proteomics. *J. Proteome Res.* **13**, 2796–2806 (2014).
53. V. D. Antonenkov, Dehydrogenases of the pentose phosphate pathway in rat liver peroxisomes. *Eur. J. Biochem.* **183**, 75–82 (1989).
54. T. Annoura, T. Nara, T. Makiuchi, T. Hashimoto, T. Aoki, The origin of dihydroorotate dehydrogenase genes of kinetoplastids, with special reference to their biological significance and adaptation to anaerobic, parasitic conditions. *J. Mol. Evol.* **60**, 113–127 (2005).
55. M. Nagy, F. Lacroute, D. Thomas, Divergent evolution of pyrimidine biosynthesis between anaerobic and aerobic yeasts. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8966–8970 (1992).
56. V. Hines, L. D. Keys, 3rd, M. Johnston, Purification and properties of the bovine liver mitochondrial dihydroorotate dehydrogenase. *J. Biol. Chem.* **261**, 11386–11392 (1986).
57. P. S. Andersen, P. J. G. Jansen, K. Hammer, Two different dihydroorotate dehydrogenases in *Lactococcus lactis*. *J. Bacteriol.* **176**, 3975–3982 (1994).
58. P. A. Michels, V. Hannaert, F. Bringaard, Metabolic aspects of glycosomes in trypanosomatidae—New data and views. *Parasitol. Today (Regul. Ed.)* **16**, 482–489 (2000).
59. T. Makiuchi, T. Nara, T. Annoura, T. Hashimoto, T. Aoki, Occurrence of multiple, independent gene fusion events for the fifth and sixth enzymes of pyrimidine biosynthesis in different eukaryotic groups. *Gene* **394**, 78–86 (2007).
60. R. Cabrera, J. Babul, V. Guixé, Ribokinase family evolution and the role of conserved residues at the active site of the PfkB subfamily representative, Pfk-2 from *Escherichia coli*. *Arch. Biochem. Biophys.* **502**, 23–30 (2010).
61. F. R. Opperdoes, J. P. Szikora, *In silico* prediction of the glycosomal enzymes of *Leishmania major* and trypanosomes. *Mol. Biochem. Parasitol.* **147**, 193–206 (2006).
62. E. E. Gill *et al.*, Novel mitochondrion-related organelles in the anaerobic amoeba *Mastigamoeba balamuthi*. *Mol. Microbiol.* **66**, 1306–1320 (2007).
63. V. Y. Petrova, D. Drescher, A. V. Kujumdzieva, M. J. Schmitt, Dual targeting of yeast catalase A to peroxisomes and mitochondria. *Biochem. J.* **380**, 393–400 (2004).
64. M. M. Leger, R. M. R. Gawryluk, M. W. Gray, A. J. Roger, Evidence for a hydroxymethyl-type anaerobic ATP generation pathway in *Acanthamoeba castellanii*. *PLoS One* **8**, e69532 (2013).
65. H. J. Santos, T. Makiuchi, T. Nozaki, Reinventing an organelle: The reduced mitochondrion in parasitic protists. *Trends Parasitol.* **34**, 1038–1055 (2018).
66. W. de Souza, A. Lanfredi-Rangel, L. Campanati, Contribution of microscopy to a better knowledge of the biology of *Giardia lamblia*. *Microsc. Microanal.* **10**, 513–527 (2004).
67. W. de Souza, Special organelles of some pathogenic protozoa. *Parasitol. Res.* **88**, 1013–1025 (2002).
68. L. A. Chávez, W. Balamuth, T. Gong, A light and electron microscopical study of a new, polymorphic free-living amoeba, *Phreatamoeba balamuthi* n. g., n. sp. *J. Protozool.* **33**, 397–404 (1986).
69. M. Čáp, L. Stěpánek, K. Harant, L. Váňová, Z. Palková, Cell differentiation within a yeast colony: Metabolic and regulatory parallels with a tumor-affected organism. *Mol. Cell* **46**, 436–448 (2012).
70. S. R. Green, C. M. Moehle, Media and culture of yeast. *Curr. Protoc. Cell Biol.* **Chapter 1**, 1.6.1–1.6.12 (2001).
71. M. T. McCammon, M. Veenhuis, S. B. Trapp, J. M. Goodman, Association of glyoxylate and beta-oxidation enzymes with peroxisomes of *Saccharomyces cerevisiae*. *J. Bacteriol.* **172**, 5816–5827 (1990).
72. L. Sterck, K. Billiau, T. Abeel, P. Rouzé, Y. Van de Peer, ORCAE: Online resource for community annotation of eukaryotes. *Nat. Methods* **9**, 1041 (2012).
73. G. Neuberger, S. Maurer-Stroh, B. Eisenhaber, A. Hartig, F. Eisenhaber, Motif refinement of the peroxisomal targeting signal 1 and evaluation of taxon-specific differences. *J. Mol. Biol.* **328**, 567–579 (2003).
74. P. B. Lazarow, The import receptor Pex7p and the PTS2 targeting sequence. *Biochim. Biophys. Acta* **1763**, 1599–1604 (2006).
75. C. R. Flynn, R. T. Mullen, R. N. Trelease, Mutational analyses of a type 2 peroxisomal targeting signal that is capable of directing oligomeric protein import into tobacco BY-2 glyoxysomes. *Plant J.* **16**, 709–720 (1998).
76. J. Mikolajczyk *et al.*, Small ubiquitin-related modifier (SUMO)-specific proteases: Profiling the specificities and activities of human SENPs. *J. Biol. Chem.* **282**, 26217–26224 (2007).

77. K. Malinská, J. Malinský, M. Opekarová, W. Tanner, Visualization of protein compartmentation within the plasma membrane of living yeast cells. *Mol. Biol. Cell* **14**, 4427–4436 (2003).
78. R. D. Gietz, R. H. Schiestl, Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 38–41 (2007).
79. J. Schindelin *et al.*, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
80. S. Bolte, F. P. Cordelières, A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).
81. F. de Chaumont *et al.*, Icy: An open bioimage informatics platform for extended reproducible research. *Nat. Methods* **9**, 690–696 (2012).
82. A. Makki *et al.*, Triplet-pore structure of a highly divergent TOM complex of hydrogenosomes in *Trichomonas vaginalis*. *PLoS Biol.* **17**, e3000098 (2019).
83. I. Kaurov *et al.*, The diverged trypanosome MICOS complex as a hub for mitochondrial cristae shaping and protein import. *Curr. Biol.* **28**, 3393–3407.e5 (2018).
84. J. J. Smith *et al.*, Transcriptome profiling to identify genes involved in peroxisome assembly and function. *J. Cell Biol.* **158**, 259–271 (2002).
85. J. Štáfková *et al.*, Dynamic secretome of *Trichomonas vaginalis*: Case study of β -amylases. *Mol. Cell. Proteomics* **17**, 304–320 (2018).
86. Y. Wang *et al.*, Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* **11**, 2019–2026 (2011).

Lateral gene transfer of *p*-cresol- and indole-producing enzymes from environmental bacteria to *Mastigamoeba balamuthi*

Eva Nývltová, Robert Šut'ák, Vojtěch Žárský, Karel Harant, Ivan Hrdý and Jan Tachezy*
Department of Parasitology, Charles University in Prague, Faculty of Science, Prague, Czech Republic.

Summary

p-Cresol and indole are volatile biologically active products of the bacterial degradation of tyrosine and tryptophan respectively. They are typically produced by bacteria in animal intestines, soil and various sediments. Here, we demonstrate that the free-living eukaryote *Mastigamoeba balamuthi* and its pathogenic relative *Entamoeba histolytica* produce significant amounts of indole via tryptophanase activity. Unexpectedly, *M. balamuthi* also produces *p*-cresol in concentrations that are bacteriostatic to non-*p*-cresol-producing bacteria. The ability of *M. balamuthi* to produce *p*-cresol, which has not previously been observed in any eukaryotic microbe, was gained due to the lateral acquisition of a bacterial gene for 4-hydroxyphenylacetate decarboxylase (HPAD). In bacteria, the genes for HPAD and the S-adenosylmethionine-dependent activating enzyme (AE) are present in a common operon. In *M. balamuthi*, HPAD displays a unique fusion with the AE that suggests the operon-mediated transfer of genes from a bacterial donor. We also clarified that the tyrosine-to-4-hydroxyphenylacetate conversion proceeds via the Ehrlich pathway. The acquisition of the bacterial HPAD gene may provide *M. balamuthi* a competitive advantage over other microflora in its native habitat.

Introduction

The aromatic amino acids tyrosine and tryptophan are, in addition to being utilized as a carbon source, important precursors of the biologically active derivatives *p*-cresol

and indole respectively (Lacoste, 1961; Spoelstra, 1978). These volatile odorous compounds are typically produced via microbial activity in the intestines of animals, but they can also be produced in other microbial habitats, such as soil environments and in marine and estuarine sediments (Updegraff, 1949; Spoelstra, 1978). Indole, an important signaling molecule, is produced by many species of gram-negative and gram-positive bacteria, and its production is dependent on the enzyme tryptophanase (Newton *et al.*, 1965). *p*-Cresol acts microbiostatically by causing an increase in membrane fluidity and, consequently, leakage of intracellular constituents (Blum and Speece, 1991; Heipieper *et al.*, 1991; McDonnell and Russell, 1999). Due to its toxicity, the ability to produce *p*-cresol in the intestinal tract is limited to a restricted set of anaerobic bacteria that are able to tolerate high *p*-cresol concentrations (Hafiz and Oakley, 1976). There are numerous reports describing *p*-cresol production by nonspecific intestinal microflora (Scheline, 1968; Spoelstra, 1978) and by bacteria in marine sediments (Updegraff, 1949). However, the ability to produce *p*-cresol has been unambiguously demonstrated in only a few organisms, including the anaerobic sediment inhabitant *Clostridium scatologenes* (Scheline, 1968; Elsdén *et al.*, 1976), the gut pathogen *Clostridium difficile* and a ruminal strain of *Lactobacillus* sp. (Yokoyama and Carlson, 1981).

Direct evidence that *p*-cresol is produced from tyrosine was provided by Spoelstra using radiolabelled substrates (Spoelstra, 1978). However, the complete metabolic pathway has not been systematically studied, and only the last step, the decarboxylation of 4-hydroxyphenylacetate (4-HPA) to *p*-cresol (Elsden *et al.*, 1976), is understood in detail. This reaction is catalysed by the oxygen-sensitive enzyme 4-HPA decarboxylase (HPAD), which belongs to a group of glycyl radical enzymes (GREs) (Craciun and Balskus, 2012). GREs utilize radicals provided by cofactors such as S-adenosylmethionine (SAM) (Lanz and Booker, 2012). The SAM-dependent activation is catalysed by the specific [4Fe4S]-activating enzyme (AE) (Selvaraj *et al.*, 2013). Unlike other GREs, the activity of the HPAD of *C. difficile* requires, in addition to the GRE (HpdB) and AE (HpdA), a third component, the HpdC subunit; however,

Received 27 July, 2016; revised 10 November, 2016; accepted 22 November, 2016. *For correspondence. E-mail tachezy@natur.cuni.cz; Tel. +420 325 874 144; Fax +420 224 919 704.

the exact role of the HpdC subunit is unknown (Andrei *et al.*, 2004).

The production of indole is rare in eukaryotic microbes, and the production of *p*-cresol has not been observed. There is only a single report in which two species of anaerobic protists, *Trichomonas vaginalis* and *Tritrichomonas foetus*, were shown to produce indole (Lloyd *et al.*, 1991). In addition, a tryptophanase gene was identified in another anaerobic protist, *Entamoeba histolytica* (Loftus *et al.*, 2005); however, the production of indole by this parasite has not been studied.

Here, we investigated the volatile metabolic compounds produced by the free-living protist *Mastigamoeba balamuthi* (formerly *Phreatamoeba balamuthi*), which inhabits anoxic and hypoxic environments (Chavez *et al.*, 1986), and its parasitic relative *E. histolytica*. Our study reveals that both Archamoebae produce indole from tryptophan and that the tryptophanase gene was likely acquired by their common ancestor. Unexpectedly, *M. balamuthi* also produces a significant amount of *p*-cresol. Using a combination of enzymatic assays and genomic searches, we traced the pathway responsible for the production of *p*-cresol from tyrosine and identified the unique *M. balamuthi* HPAD (*MbHPAD*).

Results

Analysis of volatile metabolic end products

Initially, we compared the production of volatile metabolic end products between *M. balamuthi* and *E. histolytica*, with a focus on the expected production of indole. The cells were grown in standard media for 72 h, at which point cells formed an approximately 90% confluent monolayer (the growth curve for *M. balamuthi* is given in Supporting Information Fig. S1). The metabolic end products were then analysed using GC-MS. Atypical result is shown in Fig. 1. A peak corresponding to indole was detected in the *M. balamuthi* and *E. histolytica* samples. Unexpectedly, one of the most prominent peaks preceding that of indole in the *M. balamuthi* samples corresponded to *p*-cresol. No *p*-cresol production was observed in *E. histolytica*. The other detected compounds mainly included primary alcohols with increasing numbers of carbon atoms, from ethanol to 1-octanol, as well as esters (Fig. 1).

Next, we tested whether indole and cresol production were dependent on the availability of tryptophan and tyrosine, respectively, in the culture medium. *Mastigamoeba balamuthi* and *E. histolytica* were grown for 72 h in media that contained 1–3 mM tyrosine or tryptophan, and then,

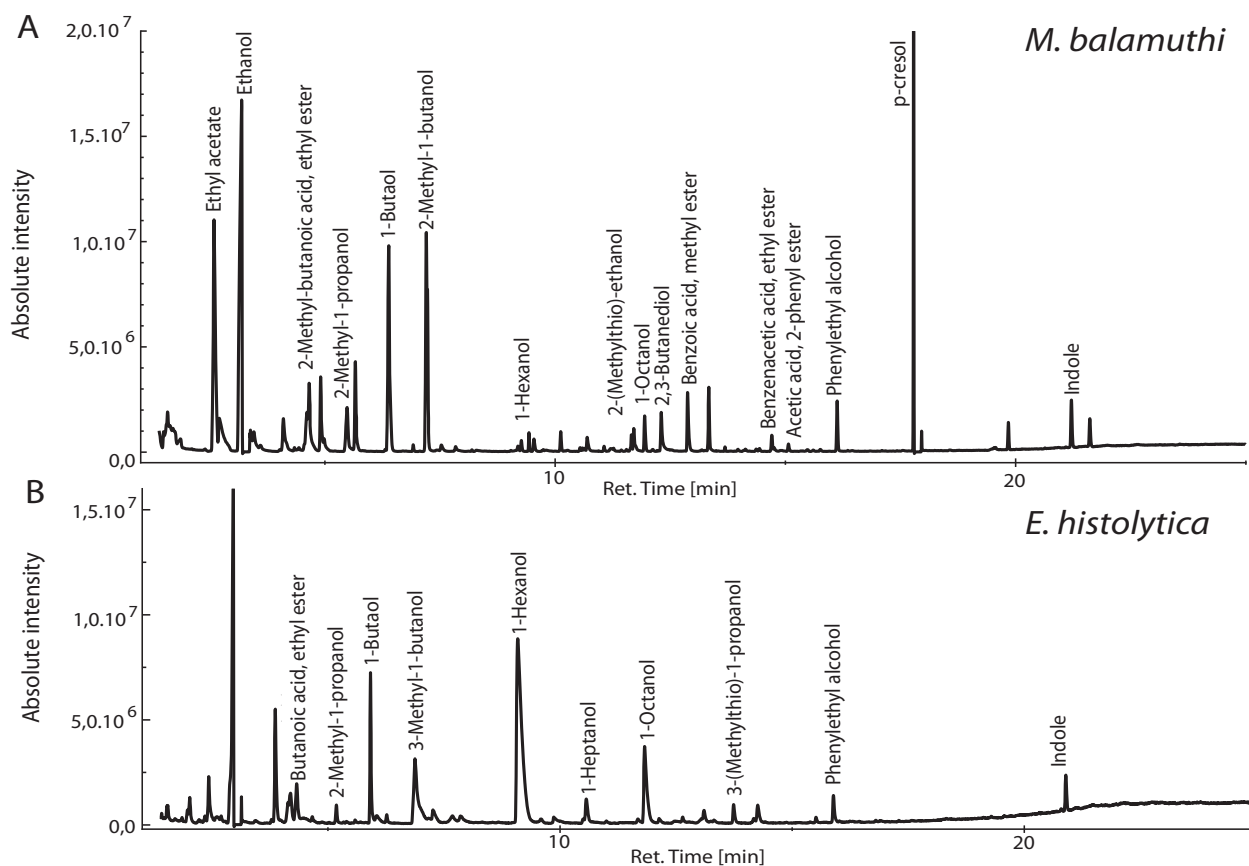


Fig. 1. Representative profiles of volatile metabolic end products determined using SPME-GS-MS analysis.

Table 1. Tryptophan- and tyrosine-dependent production of indole and *p*-cresol, respectively, by *M. balamuthi* and *E. histolytica*.

	<i>M. balamuthi</i>		<i>E. histolytica</i>	
	Concentration ^a (μM)	Production ^a (μg mg ⁻¹ protein/3 days) ± SD (<i>n</i>) ^b	Concentration ^a (μM)	Production ^a (μg mg ⁻¹ protein/3 days) ± SD (<i>n</i>) ^b
Indole				
1 mM Trp	42.0 ± 7.5 (8)	4.3 ± 1.5 (8)	6.8 ± 0.8 (6)	0.7 ± 0.1 (6)
2 mM Trp	80.8 ± 9.1 (8)	11.9 ± 3.5 (8)	8.5 ± 0.5 (6)	1.0 ± 0.1 (6)
3 mM Trp	110.1 ± 12.1 (8)	12.3 ± 3.33 (8)	8.6 ± 0.5 (6)	1.1 ± 0.1 (6)
<i>p</i> -cresol				
1 mM Tyr	152.3 ± 24.3 (8)	23.6 ± 8.2 (8)	ND	ND
1.5 mM Tyr	670.2 ± 45.3 (8)	71.3 ± 18.3 (8)	ND	ND
2 mM Tyr	871.7 ± 58.3 (8)	97.6 ± 11.0 (8)	ND	ND
3 mM Tyr	1124.4 ± 298.3 (8)	141.7 ± 37.8 (8)	ND	ND

a. Concentration of indole or *p*-cresol in the culture medium (total volume 10 ml) after 3 day cultivation of *M. balamuthi* or *E. histolytica*. The final number of *M. balamuthi*, and *E. histolytica* varied in individual experiments between 1 and 1.5 × 10⁶ cells, and 0.9 and 1 × 10⁶ cells respectively. The protein amount was determined by Lowry method and production was expressed as μg of product per mg of protein per 3 days of cultivation.

b. SD (*n*), Standard deviation (number of independent experiments). ND, not determined.

the amounts of produced indole and *p*-cresol were determined (Table 1). *Mastigamoeba balamuthi* produced approximately 6-times more indole than *E. histolytica* at 1 mM tryptophan, and indole production increased twofold to 2.5-fold with increasing tryptophan concentration. A slight tryptophan-dependent increase in indole was also observed in *E. histolytica*. The production of *p*-cresol by *M. balamuthi* increased linearly with increasing tyrosine concentration. The highest concentration of *p*-cresol measured in the medium was 1.12 mM (Table 1).

These results reveal that free-living *M. balamuthi* produces indole and *p*-cresol and that tryptophan and tyrosine, respectively, stimulate the production of these metabolic end-products, whereas the parasitic *E. histolytica* produces indole but not *p*-cresol.

Conversion of tyrosine to *p*-cresol

The unexpected discovery of *p*-cresol production by *M. balamuthi* prompted us to investigate the pathway of *p*-cresol production. Initially, we assayed the activity of HPAD, which is known to catalyse the decarboxylation of 4-HPA to *p*-cresol in *C. difficile* (Selmer and Andrei, 2001). Indeed, HPAD activity was detected in the *M. balamuthi* lysate using 4-HPA as a substrate. Importantly, the activity was dependent on the addition of SAM, which is consistent with the mechanisms of GREs (Table 2).

In yeast, the catabolism of amino acids to the corresponding alcohols/acids includes three steps: transamination, decarboxylation of the α-keto acid to an aldehyde and reduction/oxidation of the aldehyde to the corresponding alcohol/acid (Hazelwood *et al.*, 2008). Therefore, we first assessed the participation of tyrosine aminotransferase (TAT) in *p*-cresol production by *M. balamuthi*. The formation

of *p*-cresol was dependent on the addition of tyrosine and α-ketoglutarate (Table 2), consistent with the activity of TAT in converting L-tyrosine to 4-hydroxyphenylpyruvate (4-HPP) using α-ketoglutarate as an ammonia acceptor (Sivaraman and Kirsch, 2006) (Table 2). Next, we proved that 4-HPP, a product of L-tyrosine transamination, supports *p*-cresol production in the presence of SAM (Table 2). To provide evidence that 4-HPP is converted to 4-HPA through 4-hydroxyphenylacetaldehyde (4-HPAL), we measured 4-HPP decarboxylase (HPPDC) activity and the NAD-dependent activity of *p*-HPAL dehydrogenase in the *M. balamuthi* cell lysate, obtaining values of 421 ± 6 nmol min⁻¹.(mg protein)⁻¹ (*n* = 12) and 368 ± 59 nmol min⁻¹.(mg protein)⁻¹ (*n* = 10) respectively.

Collectively, these experiments indicate that tyrosine is converted to *p*-cresol in four consecutive steps that require

Table 2. Enzymatic activities involved in *p*-cresol production.

Substrate	<i>p</i> -cresol production (nmol min ⁻¹ mg) ± SD (<i>n</i>)
TAT activity	
Tyrosine	0 (10)
Tyrosine + SAM	7.4 ± 0.9 (10)
Tyrosine + α-ktg	9.3 ± 1.2 (10)
Tyrosine + α-ktg+ SAM	322.2 ± 6.9 (10)
Tyrosine + α-ktg+ SAM+NAD ⁺	421.4 ± 6.2 (6)
HPPDC activity	
4-HPP	9.1 ± 0.9 (10)
4-HPP + SAM	332.4 ± 5.1 (10)
4-HPP+SAM+NAD ⁺	503.1 ± 11.2 (6)
HPAD activity	
4-HPA	8.2 ± 2.8 (10)
4-HPA + SAM	292.1 ± 14.7 (10)

SD, standard deviation; *n*, number of independent experiments; α-ktg, α-ketoglutarate.

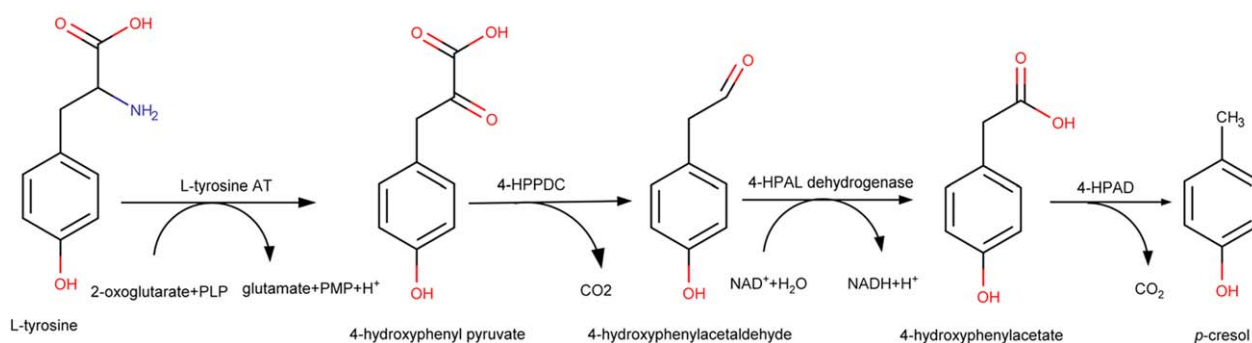


Fig. 2. Scheme of the conversion of tyrosine to *p*-cresol in *M. balamuthi*.

the activities of TAT, HPPDC, 4-HPAL dehydrogenase and HPAD (Fig. 2).

Unique primary structure of MbHPAD. Homology searches of the *M. balamuthi* genome sequence using the HpdB component of *C. difficile* HPAD as the query identified three genes with significant similarity. Two of the genes, with 28% similarity to HpdB, correspond to pyruvate formate lyase (PFL) (Stairs *et al.*, 2011; Nývltová *et al.*, 2015). We considered the third gene, which has the highest similarity (41%) to HpdB, a putative *M. balamuthi* HPAD. Surprisingly, the protein alignment revealed that MbHPAD is a fusion protein, with the N-terminal domain corresponding to HpdB and the C-terminal domain homologous to the AE (HpdA in *C. difficile*) (Fig. 3 and Supporting Information Fig. S2A). The protein coding sequence of MbHPAD predicted from the genome sequence was confirmed by RT PCR. Neither the BLAST

searching strategies nor the searches using the sensitive profile hidden Markov model-based tool HMMER (Finn *et al.*, 2011) revealed the presence of HpdC in the *M. balamuthi* genome.

The unusual primary structure of MbHPAD and the absence of an HpdC ortholog prompted us to seek direct evidence that MbHPAD catalyses *p*-cresol production. Thus, MbHPAD was heterologously expressed under anaerobic conditions in *Escherichia coli* (Supporting Information Fig. S3A). The incubation of the recombinant protein in the reaction mixture with 4-HPA led to the production of *p*-cresol (Supporting Information Fig. S3B). The specific activity of the recombinant MbHPAD was 0.8 ± 0.1 nmol min⁻¹ mg and increased to 6.9 ± 0.4 nmol min⁻¹ mg in the presence of SAM. These results confirm that the MbHPAD fusion protein catalyses SAM-dependent *p*-cresol formation.

MbHPAD phylogeny. Phylogenetic analysis of HPAD was performed using 193 HpdB homologs, including sequences of other glycy radical enzymes (Fig. 4 and Supporting Information Fig. S2B). MbHPAD clustered with high statistical support within the branch of HPADs present in *Clostridium* species, the actinobacterium *Olsenella uli* and sequences found in the metagenome of estuarine sediments (Baker *et al.*, 2015). Distinct branches were formed by other GREs. The MbHPAD sequence topology together with the observation that *M. balamuthi* appears to be the only eukaryote possessing HPAD indicates that MbHPAD was acquired by lateral gene transfer (LGT) from bacteria.

HPPDC and TAT. A single gene for the putative HPPDC was identified in the *M. balamuthi* genome (MbHPPDC), with a typical HH motif and the thiamine diphosphate (ThDP) binding fingerprint of ThDP-dependent carboxylases (Supporting Information Fig. S4A) (Andrews and McLeish, 2012). Enzymes in this family can utilize various 2-keto acid substrates, including 4-HPP (Supporting Information Fig. S4A) (Andrews and McLeish, 2012). Thus, we produced a recombinant MbHPPDC and confirmed that the enzyme catalyses the conversion of 4-HPP to 4-HPAL

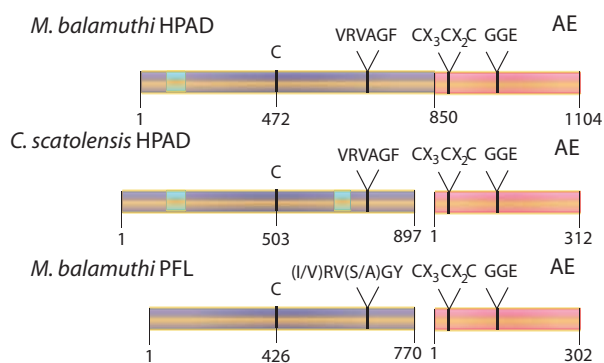


Fig. 3. Domain structure of MbHPAD compared with that of HPAD from *C. scatologenes* and PFL from *M. balamuthi*. The conserved C503 likely harbors a thyl radical. The VRVAGF motif includes a radical storage site at glycine 873. The CX₃CX₂C motif coordinates the [4Fe4S] cluster. The GGE motif binds SAM (Nicolet and Drennan, 2004). Boxes represent the two peptide sequences (Q121-K167 and N672-E700) surrounding the radical-binding domain of *C. scatologenes* HPAD that are absent in pyruvate formate-lyases (PFLs). AE, activating enzyme. Amino acid residue numbers are based on the *C. scatologenes* HPAD sequence (Selmer and Andrei, 2001; Feliks *et al.*, 2013).

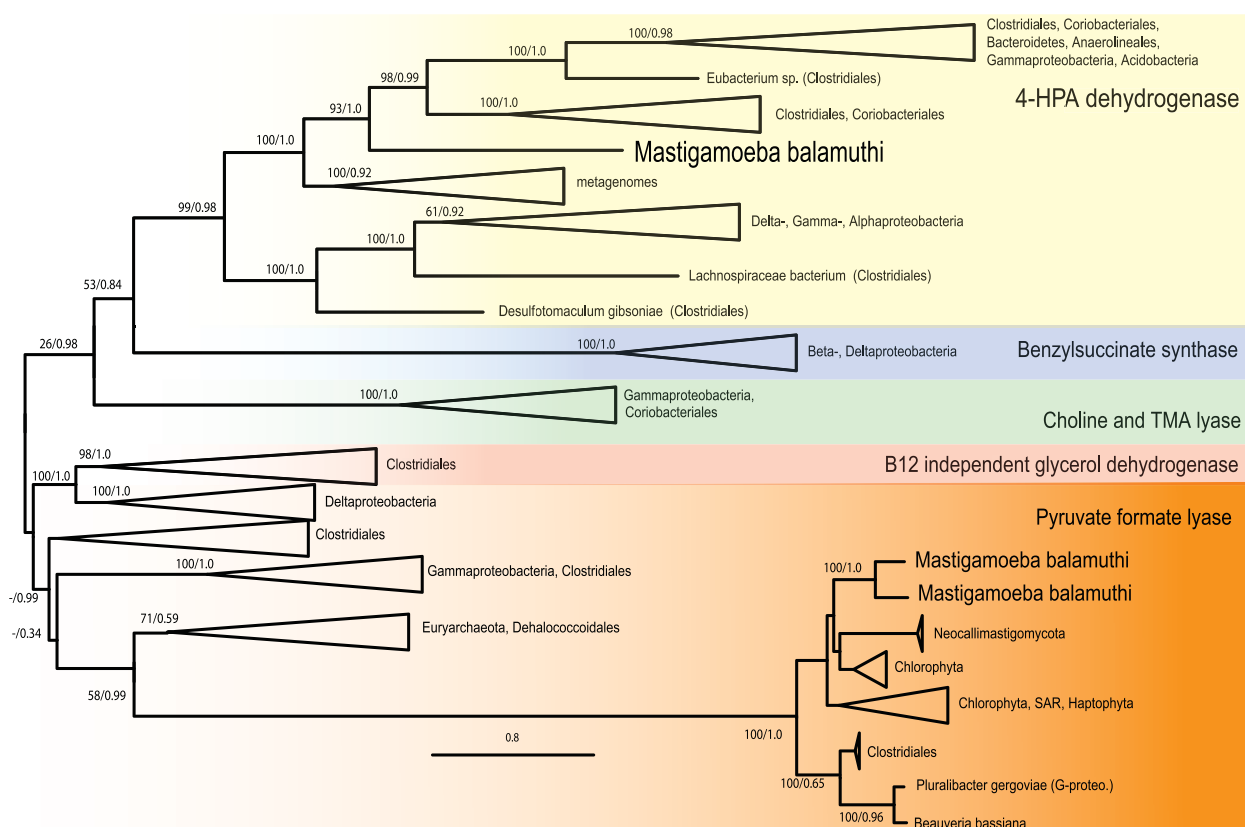


Fig. 4. Phylogeny of HPAD.

A tree for HPAD (193 taxa and 655 sites) was built using Bayesian inference (BI). Bootstrap support (BP) and posterior probability (PP) values were calculated for each branch using RAxML and PhyloBayes respectively.

(Supporting Information Fig. S5). Searches for *Mb*HPPDC orthologs and their phylogenetic analysis revealed the presence of a putative HPPDC in only a few species from three eukaryotic lineages: Amoebozoa, Apicomplexa and Fungi (Supporting Information Fig. S4B). These eukaryotic genes appear to be of polyphyletic origin; however, the branching order has relatively low statistical support for clarifying this issue.

In the *M. balamuthi* genome, we identified three genes with conserved residues characteristic of the AT family I of proteins that represent candidates for enzymes with TAT activity (Supporting Information Table S1, Supporting Information Fig. S6).

Production of p-cresol by M. balamuthi suppresses the growth of Citrobacter freundii. Cresol is known for its bacteriostatic effect. Therefore, we examined whether the *p*-cresol production by *M. balamuthi* provides advantages against bacteria present in the same environment. First, we tested the sensitivity of *M. balamuthi* to *p*-cresol. The amoebae were treated with 1–15 mM *p*-cresol for 72 h, and the viability of cells was then assessed by cultivation in *p*-cresol-free medium. *Mastigamoeba balamuthi* was able to tolerate up to 10 mM *p*-cresol. Next, we evaluated the

sensitivity of *C. freundii*, a facultative anaerobic bacterium, to *p*-cresol. After 12 h of cultivation, the number of bacterial cells in the presence of 1 and 2 mM *p*-cresol were 27% and 8%, respectively, of the number of cells grown without *p*-cresol (Fig. 5). Finally, we cultivated *M. balamuthi* in medium with 1–3 mM tyrosine for 48 h, followed by co-cultivation with *C. freundii* for 12 h. Under these conditions, we observed that *C. freundii* growth slowed with increasing concentrations of tyrosine in the medium and that growth was reduced to approximately 38% at 3 mM tyrosine (corresponding to a production of over 1 mM *p*-cresol, Table 1). These results indicate that *M. balamuthi* produces *p*-cresol in concentrations that affect the growth of *C. freundii*.

Tryptophanase phylogeny. Searching the *M. balamuthi* genome using the *E. histolytica* tryptophanase gene as a query revealed two *M. balamuthi* genes with 56% and 57% protein sequence similarity to the *E. histolytica* gene. In a phylogenetic reconstruction using a dataset with 122 tryptophanase orthologs, *M. balamuthi* tryptophanase formed a statistically well-supported branch with *Entamoeba* species as a sister group and *Blastocystis hominis*, a member of the *Stramenopila* group (Fig. 6 and Supporting

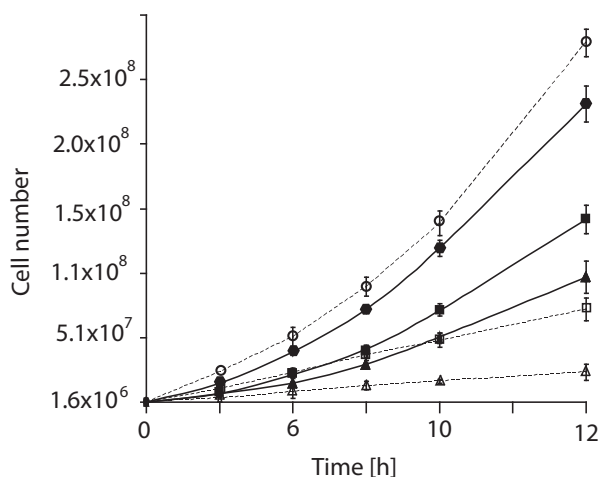


Fig. 5. Effect of *p*-cresol produced by *M. balamuthi* on the growth of *C. freundii*.

M. balamuthi was cultivated in PYGC medium with 1 mM (●), 2 mM (■) or 3 mM (▲) tyrosine for 48 h to form a monolayer, and *C. freundii* was then added and co-cultivated for 12 h (full lines). In the control experiments (dashed lines), *C. freundii* was cultivated in PYGC medium without (○) or with 1 mM (□) or 2 mM (△) *p*-cresol. Addition of 1–3 mM tyrosine to the medium without *M. balamuthi* had no effect on *C. freundii* growth. Error bars represent standard deviations, $n = 6$.

Information Fig. S7). Interestingly, the clade of fungi belonging to *Sordariomycota* appeared at a basal position to the *Archamoebae/Blastocystis* branch. The *Archamoebae/Blastocystis* and fungal clades grouped, with high statistical support, with the bacterial clade consisting of *Fusobacteria* and several genera of *firmicutes*. In addition, we found tryptophanase sequences in two members of the *Alveolata* group (*Vitrella* and *Phaeodactylum*), with filamentous thermophilic bacteria (*Caldilinea* and *Anaerolinea*) as the nearest neighbors, and *Dictyostelium purpureum*, which, together with *T. vaginalis*, was interleaved with members of the *Bacteroidetes* phylum. The topology of the tryptophanase phylogeny suggests that the tryptophanase gene was present in a common ancestor of *M. balamuthi* and *Entamoeba* species. However, the unusual relationships with other eukaryotic sequences (*Blastocystis* and fungi) do not permit an unequivocal explanation of their origin.

Discussion

Our investigations demonstrate that the free-living anaerobic protist *M. balamuthi* can produce the aromatic compound *p*-cresol from tyrosine, a feature that has not previously been observed in any eukaryotic microbe. The pathway includes four consecutive steps that require the activities of TAT, HPPDC, 4-HPAL dehydrogenase and HPAD (Fig. 2). The origin of *MbHPAD* exemplifies the acquisition of genetically linked genes via LGT from

bacteria to eukaryotes. *MbHPAD* appears to be an ortholog of the enzyme that catalyses *p*-cresol production in *Clostridia*; however, in contrast to bacterial HPAD, *MbHPAD* displays a unique fusion with the AE. The closely related pathogen *E. histolytica* lacks this pathway; however, both *Archamoebae* produce indole by tryptophanase activity. *p*-Cresol production and tolerance to this toxic compound have primarily been investigated for the gut pathogen *C. difficile*. Although *p*-cresol is known as a disinfectant, this anaerobic bacterium has been shown to withstand up to 38 mM *p*-cresol *in vitro* (Hafiz and Oakley, 1976). Most other microbes are considerably more sensitive to *p*-cresol toxicity. For example, *p*-cresol inhibits the growth of various environmental bacteria at concentrations of 0.25–2.4 mM (Blum and Speece, 1991). In our experiments, the growth *C. freundii* was decreased to approximately 27% in the presence of 1 mM *p*-cresol. Similarly, 1.4 mM *p*-cresol restricts the growth of the free living protist *Tetrahymena pyriformis* (Schultz *et al.*, 1996). It has been proposed that *p*-cresol production provides *C. difficile* with a competitive advantage against other gut microbes (Dawson *et al.*, 2008). Free-living *M. balamuthi* seems to use a similar strategy in its anaerobic sediment environment as that of *C. difficile* in the gut by producing *p*-cresol in concentrations that affect other microorganisms, such as *C. freundii*, and tolerating this toxin at concentrations of up to 10 mM. The growth of *M. balamuthi* is rather slow, with a generation time of approximately 12 h, and could be even slower in nature. The generation time of most bacteria ranges from 12 min to approximately 60 min. Consequently, too many fast-growing bacteria can scavenge available nutrients and produce metabolic end-products in toxic concentrations for slowly growing cells. Thus, production of *p*-cresol might be an important factor for *M. balamuthi* to restrict the bacterial growth and alter microenvironment to be more favourable for proliferation of this protist. *Mastigamoeba balamuthi* most likely acquired this potential ‘weapon’ against cohabiting bacteria from *p*-cresol-producing anaerobic bacteria via LGT, as indicated by the phylogenetic analysis of HPAD. In addition to *C. scatologenes*, the only defined species of free-living bacteria in which *p*-cresol production has been studied (Scheline, 1968; Elsdon *et al.*, 1976), we identified several novel putative HPAD orthologs in environmental samples from estuarine sediments that clustered with *MbHPAD* (Baker *et al.*, 2015). The results of this analysis are consistent with previous observations of *p*-cresol production by unspecified bacteria in marine sediments (Updegraff, 1949). Therefore, the acquisition of the HPAD gene by *M. balamuthi* from its bacterial cohabitants in anaerobic niches is the most plausible explanation.

Although the reason for *p*-cresol production is likely related to its toxicity, it is difficult to predict the specific role of indole production in *M. balamuthi* and *E. histolytica*

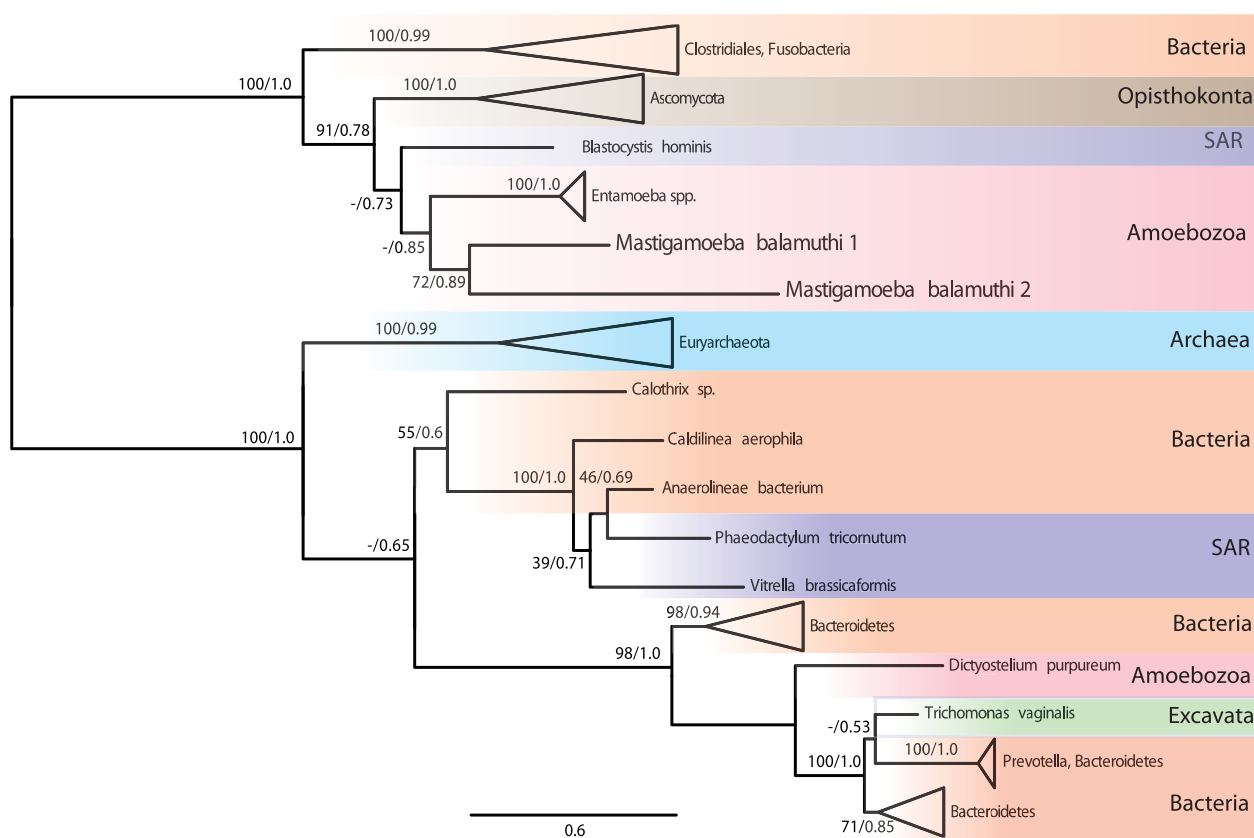


Fig. 6. Phylogeny of tryptophanase.

A tree for tryptophanase (122 taxa and 359 sites) was built using Bayesian inference (BI). Bootstrap support (BP) and posterior probability (PP) values were calculated for each branch using RAxML and PhyloBayes respectively.

because indole could serve as a signaling molecule in multiple biological functions (Kim and Park, 2015). For example, indole produced by enteric bacteria is an important factor for cross-talk between the microbiota and the gut wall and can serve as a beneficial signal for the modulation of the mucosal barrier, inflammatory responses and the function of enteroendocrine L cells (Bansal *et al.*, 2010;Chimerel *et al.*, 2014). In this respect, comparing indole production by pathogenic *E. histolytica* and non-pathogenic *E. dispar* could be of interest, as both organisms possess a gene for tryptophanase.

We show here that *p*-cresol production in *M. balamuthi* proceeds in part via the Ehrlich pathway, which mediates the catabolism of amino acids to corresponding 'fusel' alcohols and acids, including 4-HPA (Hazelwood *et al.*, 2008). However, in contrast to yeast, which excrete these fusel products, *M. balamuthi* utilizes 4-HPA as a substrate for 4-HPAD to produce *p*-cresol. Information about the specific intermediates of the conversion of 4-HPP to *p*-cresol and the corresponding enzymes is scarce. Anaerobic microbes in pig feces have been shown to produce *p*-cresol from 4-HPP and 4-HPA (Spoelstra, 1978). The involvement of tyrosine transamination in *p*-cresol formation has been

suggested in *Clostridia* (Mathus *et al.*, 1995). We demonstrated that *p*-cresol production in *M. balamuthi* is dependent on tyrosine and α -ketoglutarate, which are a substrate and a co-substrate of TAT in the production of 4-HPP and glutamate respectively. The conversion of 4-HPP to 4-HPA by the oxygen-dependent activity of *p*-HPP oxidase has been described in the aerobic soil bacterium *Arthrobacter* sp. (Blakley, 1977). However, the participation of 4-HPP oxidase in *p*-cresol formation in anaerobes is unlikely, as HPAD activity is highly sensitive to oxygen. Indeed, we demonstrated that 4-HPP is first converted to 4-HPAL by the activity of HPPDC, and 4-HPAL is then converted to *p*-HPA by the activity of the NAD-dependent 4-HPAL dehydrogenase.

Because the activity of HPAD and other GREs is dependent on the presence of the AE, a successful bacteria-to-eukaryote gene transfer relies on the acquisition of both genes, which is more likely if the genes are genetically linked (Andersson and Roger, 2002). Indeed, the components required for HPAD activity in *C. difficile* are arranged in the HPAD operon (*hpdB/A/C*). Interestingly, in *M. balamuthi*, we found a single ORF for the fused HPAD-AE protein, which had an arrangement corresponding to the

hpdB/A genes in the bacterial operon but without the *hpdC* protein, which was either lost or not present in the donor bacterium. In protists, the fusion of genes acquired via LGT from bacterial operons is not unprecedented. Andersson and Roger (2002) demonstrated the fusion of the small and large subunits of glutamate synthase in diplomonads, and fusion of two subunits of the iron-sulfur cluster assembly machinery SufCB was observed in *Blastocystis*, *Pygmsuia* and *Stygiella* (Tsaousis *et al.*, 2012; Stairs *et al.*, 2014; Leger *et al.*, 2016). The fusion of PFL with the AE has been noted in the diatom *Thalassiosira pseudonana* (Stairs *et al.*, 2011). These findings, including the demonstration of the MbHPAD-AE fusion protein, strongly support the occurrence of operon-mediated gene transfer from an ancestral bacterial donor to evolving eukaryotes.

Whereas LGT is a generally accepted phenomenon that contributes to eukaryotic genome evolution, our understanding of the mechanisms by which bacterial genes are transferred to eukaryotic cells is rather limited. An important factor is the opportunity to encounter bacterial DNA (Keeling and Palmer, 2008). Indeed, phagotrophs such as *Mastigamoeba* and *Entamoeba* species engulf and digest bacteria, which likely enhances exposure to bacterial DNA. Moreover, an intimate association between *Mastigamoeba aspera* and ectobiotic bacteria has been observed (Siemensma, 2016), which may further increase bacterial DNA exposure. Interestingly, some bacterial species can import DNA into eukaryotic cells via a type IV secretion system and integrate this DNA into the recipient genome (Lacroix and Citovsky, 2016). However, the extent to which this mechanism contributed to bacteria-to-eukaryote LGT during the course of evolution is not known.

The involvement of LGT in the evolution of parasitic *E. histolytica* has been investigated on the whole-genome level (Clark *et al.*, 2007; Alsmark *et al.*, 2009). Sixty-eight cases of LGT from various prokaryotic lineages have been predicted, particularly from gut bacteria to *E. histolytica*. An attractive explanation of this bias is that the host gut bacteria share the same ecological niche with *Entamoeba*, significantly contributing to this gene transfer (Hirt *et al.*, 2015). However, because the environment of an animal's gut is rather anaerobic and it is likely that the free-living *Entamoeba* ancestor also occupied anaerobic niches (Edgcomb *et al.*, 2002), in some cases, the observed bias maybe more of a reflection of 'ancestral' LGT within an anaerobic ecosystem than a 'recent' adaptation of the *Entamoeba* lineage to parasitism. For example, a previous phylogenetic analysis of tryptophanase revealed that *E. histolytica* clusters with maximum support with *Fusobacterium* (a pathogen found in oral cavities) (Clark *et al.*, 2007; Alsmark *et al.*, 2009). However, when we used enriched taxon sampling that included the tryptophanase of *M. balamuthi*, the tryptophanase of *E. histolytica* clustered with those of *M. balamuthi*, *Blastocystis* and *Sordariomycetes*

fungi, which does not support the acquisition of the gene from the human pathogen but instead suggests acquisition from environmental Firmicutes.

Collectively, a comparison of free-living species with their parasitic relatives and extensive sampling of prokaryotes in the same environments is essential to distinguish between 'recent' LGTs associated with adaptations to parasitism and 'ancient' LGTs that might predispose free-living protists to parasitism.

Experimental procedures

Cell cultivation

The *M. balamuthi* (strain ATCC 30984) was maintained axenically as previously described (Chavez *et al.*, 1986). For indole and *p*-cresol detection, the cells were grown for 72 h in PYGC medium containing 1–3 mM tryptophan or tyrosine. *Entamoeba histolytica* (strain HM-1:IMSS) was maintained as described (Diamond *et al.*, 1995) or in Y1-S medium containing 1–3 mM tryptophan. *Citrobacter freundii* was maintained as described (Zubacova *et al.*, 2013). The cell density of *M. balamuthi* was determined using a hemocytometer. *Entamoeba histolytica* cells were counted using a cell counter (ZTM Series Coulter Counter, Beckman Coulter). The density of *C. freundii* cells was determined spectrophotometrically at an optical density of 600 nm, and the number of cells was calculated according to the following formula $OD_{600} 1.0 = 8 \times 10^8 \text{ cells ml}^{-1}$.

p-cresol susceptibility assays

Mastigamoeba cells were incubated in the presence of 1, 5, 8, 10 or 15 mM *p*-cresol for 72 h. After incubation, the medium was replaced with fresh medium without *p*-cresol. The viability of the cells was evaluated microscopically after 5 days. To test the *p*-cresol susceptibility of *C. freundii*, the cells were incubated with 0.5, 1 and 2 mM *p*-cresol for 12 h and the cell density was determined at different time intervals as indicated. For the co-cultivation experiments, *M. balamuthi* was grown in medium with 1–3 mM tyrosine until it formed a confluent cell layer (approximately 140,000 cells). Then, 160,000 cells ml^{-1} *C. freundii* were inoculated and co-cultivated with *M. balamuthi* for 12 h. Bacterial growth was measured spectrophotometrically at a wavelength of 600 nm (OD_{600}). As a control, *C. freundii* was grown in PYGC medium containing 1–3 mM tyrosine without *M. balamuthi*.

Determination of volatile metabolic end-products by gas chromatography coupled with mass spectrometry (GC-MS)

Indole, *p*-cresol and other volatile metabolic end products were measured using GC-MS as described previously (Sutak *et al.*, 2012). Samples were prepared using headspace solid-phase microextraction using 1 ml of cell-free medium from cultures in exponential growth phase stirred in a closed vial at 50°C. Carboxen/PDMS fibres (Supelco, Bellefonte, PA, USA) were incubated in the sample headspace for 20 min and then inserted into the GC injection port using splitless-mode

injection at 250°C. GC-MS analyses were performed on a GC-MS-QP2010 Plus system (Shimadzu, Kyoto, Japan) equipped with a Supelcowax polyethylene glycol capillary column (Supelco), with helium as the carrier gas at a constant linear flow rate of 40 cm s⁻¹. The mass spectrometer was operated in scan mode in an *m/z* range of 35–350.

Determination of *p*-cresol and indole production in culture media using HPLC

Mastigamoeba balamuthi (6000 cells) or *E. histolytica* (3000 cells) was inoculated into 10 ml of PYGC or YI-S medium, respectively, and grown for 3 days. The cells were harvested by centrifugation, the cell density was determined using a hemocytometer, and the total protein amount determined by the Lowry method. A 300- μ l aliquot of the culture medium (supernatant) was collected and *p*-cresol and indole were extracted into 300 μ l of 80% acetonitrile with 20% methanol. The mixture was incubated for 30 min at room temperature, and the precipitated proteins were removed by centrifugation for 15 min at 13,000 $\times g$. Similarly, for the determination of enzymatic activities, 300 μ l of the sample, following anaerobic incubation with the substrates, was extracted into 300 μ l of 80% acetonitrile with 20% methanol. The mixture was incubated for 30 min at room temperature, and the precipitated proteins were removed by centrifugation for 15 min at 13,000 $\times g$. *p*-Cresol and indole were measured using HPLC with fluorescence detection (270 nm excitation/340 nm emission wavelengths) as described previously (Doerner *et al.*, 2009) using a Dionex Ultimate 3000 system. Samples were injected into a Dionex Acclaim RSLC 120 C18 column that was equilibrated with buffer containing 10% acetonitrile and 50 mM ammonium acetate. A linear gradient (0%–60%) of acetonitrile was established, with a flow rate of 500 μ l min⁻¹ for 10 min. Data were processed in Chromeleon 7.0 software, and the amounts of *p*-cresol and indole were determined based on the corresponding external calibration curves.

Production of recombinant enzymes

The coding genes for *MbHPAD* and *MbHPPDC* (accession numbers KX356607 and KX352254 respectively) were amplified from cDNA by PCR using primers with NdeI and HindIII restriction sites (Supporting Information Table S1) and cloned into pET42b vectors (Novagen). *Escherichia coli* BL21 (DE3) GroES/EL (Novagen) transformed with *MbHPAD* was grown anaerobically in 2 l of LBG medium with 0.2% glucose that was enriched with 0.5 mM Mohr's salt, 50 mM NaNO₃ and 0.5 mM sodium sulfide at 37°C to OD₆₀₀ = 0.8. *Escherichia coli* BL21 (DE3) GroES/EL (Novagen) transformed with *MbHPPDC* was grown aerobically in 3 l of LB medium at 37°C to an OD₆₀₀ = 0.8. Then, 0.1 mM IPTG was added, and the cells were incubated at 23°C for 3 h to OD₆₀₀ = 1.2. The cells were then harvested via centrifugation and resuspended in 20 ml of lysis buffer (50 mM HEPES, 20 mM NaCl, 10 mM imidazole, 10% glycerol, 5 mM MgCl₂, 5 mM DTT and 5 mM cysteine [pH 8.0]) with 10 μ g ml⁻¹ DNase, 5 μ g ml⁻¹ lysozyme and 0.2 mM protease inhibitors (TLCK and leupeptin). The cells were sonicated, and the cell debris was removed by ultracentrifugation at 150,000 $\times g$ for 35 min at 4°C. The

recombinant proteins were purified by affinity chromatography with an Ni-NTA agarose column under native conditions according to the manufacturer's protocol (Qiagen).

Enzyme assays

The activity of HPAD was determined as described previously (Andrei *et al.*, 2004). The assay mixture contained 1 ml of 50 mM Tris-HCl buffer (pH 7.4), 1 mM dithiothreitol, 1 mM pyridoxal phosphate, 1 mM MgCl₂, 2 mM 4-HPA, 2 mM SAM (Sigma) and an aliquot of cellular lysate or recombinant *MbHPAD*. The mixture was incubated under anaerobic conditions for 20, 40 and 60 min. The reaction was stopped by extracting the mixture into 80% acetonitrile with 20% methanol for 30 min at room temperature. This mixture was centrifuged at 13,000 $\times g$ for 15 min, and the extracted *p*-cresol was measured using HPLC as described above. The activities of TAT and 4-HPPDC were determined based on *p*-cresol production in reaction mixtures containing aliquots of the cellular fractions in 50 mM Tris-KCl buffer (pH 7.2), 1 mM dithiothreitol, 1 mM pyridoxal phosphate, 1 mM MgCl₂, 2 mM 4-HPA, 2 mM SAM (Sigma) and either (i) 2 mM tyrosine with 2 mM α -ketoglutarate, to measure TAT activity, or (ii) 2 mM 4-HPP (Sigma), to determine HPPDC activity. The concentrations of the *p*-cresol produced were determined using HPLC as described above.

Alternatively, the activity of HPPDC was determined based on 4-HPP consumption (Blakley, 1977). The activity of 4-HPAL dehydrogenase was determined based on NADH oxidation using 4-HPA as the substrate (Blakley, 1977).

The production of 4-HPAL by recombinant *MbHPPDC* was determined in a reaction mixture that contained 50 mM Tris-KCl buffer (pH 7.2), 2 mM 4-HPP (Sigma), 1 mM MgCl₂ and 1 mM thiamine diphosphate. The mixture was incubated for 45 min at 25°C, and the formation of 4-HPAL was determined using HPLC as described previously (Hazen *et al.*, 1996). Briefly, samples were injected into a Dionex Acclaim RSLC 120 C18 column that was equilibrated with buffer A (5% methanol, 0.1% trifluoroacetic acid [pH 2.5]). A linear gradient (0%–35%) of buffer B (90% methanol, 0.1% trifluoroacetic acid [pH 2.5]) was applied following a 35% isocratic flow of buffer B for 3 min. The flow rate was 300 μ l min⁻¹.

Phylogenetic analysis

BLAST searches for homologs of *MbHPAD*, *MbHPPDC*, TAT and tryptophanase were performed using the NCBI non-redundant protein database, and the resulting datasets were manually edited to remove redundant sequences, incomplete sequences and distant homologs. Sequences were aligned with Mafft (options: max iterate 1000, local pair) (Katoh and Standley, 2013). Alignments were trimmed with BMGE (BLOSUM30 matrix, block size of 5) (Criscuolo and Gribaldo, 2010). Maximum likelihood trees were built using RAxML with an LG amino acid substitution model with gamma correction for the heterogeneity of evolutionary rates among sites, and the bootstrap support was computed using 500 rapid bootstrap replicates (Stamatakis, 2006). Bayesian inference was conducted using PhyloBayes running two chains under the CAT model. For each chain, 30,000 generations were run, and

trees were sampled every 10 generations. A burn-in of 3,000 generations was discarded (Lartillot *et al.*, 2009). The draft *M. balamuthi* genome sequence is available at <http://www.ebi.ac.uk/ena/data/view/CBKX00000000>.

Acknowledgment

This work was supported by the Czech Grant Foundation (P305/11/1061), LQ1604 NPU II provided by MEYS, CZ.1.05/1.1.00/02.0109, and BIOCEV provided by ERDF and MEYS.

References

- Alsmark, U.C., Sicheritz-Ponten, T., Foster, P.G., Hirt, R.P., and Embley, T.M. (2009) Horizontal gene transfer in eukaryotic parasites: a case study of *Entamoeba histolytica* and *Trichomonas vaginalis*. *Methods Mol Biol* **532**: 489–500.
- Andersson, J.O., and Roger, A.J. (2002) Evolutionary analyses of the small subunit of glutamate synthase: gene order conservation, gene fusions, and prokaryote-to-eukaryote lateral gene transfers. *Eukaryot. Cell* **1**: 304–310.
- Andrei, P.I., Pierik, A.J., Zauner, S., Andrei-Selmer, L.C., and Selmer, T. (2004) Subunit composition of the glycy radical enzyme p-hydroxyphenylacetate decarboxylase. A small subunit, HpdC, is essential for catalytic activity. *Eur J Biochem* **271**: 2225–2230.
- Andrews, F.H., and McLeish, M.J. (2012) Substrate specificity in thiamin diphosphate-dependent decarboxylases. *Bioorg Chem* **43**: 26–36.
- Baker, B.J., Lazar, C.S., Teske, A.P., and Dick, G.J. (2015) Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome* **3**: 14.
- Bansal, T., Alaniz, R.C., Wood, T.K., and Jayaraman, A. (2010) The bacterial signal indole increases epithelial-cell tight-junction resistance and attenuates indicators of inflammation. *Proc Natl Acad Sci USA* **107**: 228–233.
- Blakley, E.R. (1977) The catabolism of L-tyrosine by an *Arthrobacter* sp. *Can J Microbiol* **23**: 1128–1139.
- Blum, D.J., and Speece, R.E. (1991) A database of chemical toxicity to environmental bacteria and its use in interspecies comparisons and correlations. *Res J Water Pollut Contr Fed* **63**: 198–207.
- Chavez, L.A., Balamuth, W., and Gong, T. (1986) A light and electron microscopical study of a new, polymorphic free-living amoeba, *Phreatamoeba balamuthi* n. g., n. sp. *J Protozool* **33**: 397–404.
- Chimerel, C., Emery, E., Summers, D.K., Keyser, U., Gribble, F.M., and Reimann, F. (2014) Bacterial metabolite indole modulates incretin secretion from intestinal enteroendocrine L cells. *Cell Rep* **9**: 1202–1208.
- Clark, C.G., Alsmark, U.C., Tazreiter, M., Saito-Nakano, Y., Ali, V., Marion, S., *et al.* (2007) Structure and content of the *Entamoeba histolytica* genome. *Adv Parasitol* **65**: 51–190.
- Craciun, S., and Balskus, E.P. (2012) Microbial conversion of choline to trimethylamine requires a glycy radical enzyme. *Proc Natl Acad Sci USA* **109**: 21307–21312.
- Crisuolo, A., and Gribaldo, S. (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* **10**: 210.
- Dawson, L.F., Stabler, R.A., and Wren, B.W. (2008) Assessing the role of p-cresol tolerance in *Clostridium difficile*. *J Med Microbiol* **57**: 745–749.
- Diamond, L.S., Clark, C.G., and Cunnick, C.C. (1995) YI-S, a casein-free medium for axenic cultivation of *Entamoeba histolytica*, related *Entamoeba*, *Giardia intestinalis* and *Trichomonas vaginalis*. *J Eukaryot Microbiol* **42**: 277–278.
- Doerner, K.C., Cook, K.L., and Mason, B.P. (2009) 3-Methylindole production is regulated in *Clostridium scatologenes* ATCC 25775. *Lett Appl Microbiol* **48**: 125–132.
- Edgcomb, V.P., Simpson, A.G., Zettler, L.A., Nerad, T.A., Patterson, D.J., Holder, M.E., and Sogin, M.L. (2002) Pelobionts are degenerate protists: insights from molecules and morphology. *Mol Biol Evol* **19**: 978–982.
- Elsden, S.R., Hilton, M.G., and Waller, J.M. (1976) The end products of the metabolism of aromatic amino acids by Clostridia. *Arch Microbiol* **107**: 283–288.
- Feliks, M., Martins, B.M., and Ullmann, G.M. (2013) Catalytic mechanism of the glycy radical enzyme 4-hydroxyphenylacetate decarboxylase from continuum electrostatic and QC/MM calculations. *J Am Chem Soc* **135**: 14574–14585.
- Finn, R.D., Clements, J., and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29–W37.
- Hafiz, S., and Oakley, C.L. (1976) *Clostridium difficile*: isolation and characteristics. *J Med Microbiol* **9**: 129–136.
- Hazelwood, L.A., Daran, J.M., van Maris, A.J., Pronk, J.T., and Dickinson, J.R. (2008) The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism. *Appl Environ Microbiol* **74**: 2259–2266.
- Hazen, S.L., Hsu, F.F., and Heinecke, J.W. (1996) p-Hydroxyphenylacetaldehyde is the major product of L-tyrosine oxidation by activated human phagocytes. A chloride-dependent mechanism for the conversion of free amino acids into reactive aldehydes by myeloperoxidase. *J Biol Chem* **271**: 1861–1867.
- Heipieper, H.J., Keweloh, H., and Rehm, H.J. (1991) Influence of phenols on growth and membrane permeability of free and immobilized *Escherichia coli*. *Appl Environ Microbiol* **57**: 1213–1217.
- Hirt, R.P., Alsmark, C., and Embley, T.M. (2015) Lateral gene transfers and the origins of the eukaryote proteome: a view from microbial parasites. *Curr Opin Microbiol* **23**: 155–162.
- 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.
- Keeling, P.J., and Palmer, J.D. (2008) Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* **9**: 605–618.
- Kim, J., and Park, W. (2015) Indole: a signaling molecule or a mere metabolic byproduct that alters bacterial physiology at a high concentration? *J Microbiol* **53**: 421–428.
- Lacoste, A.M. (1961) Dégradation du tryptophane par les bactéries de la panse de ruminants. *CR Acad Sci* **252**: 1233–1241.
- Lacroix, B., and Citovsky, V. (2016) Transfer of DNA from bacteria to eukaryotes. *mbio* **7**: e00863-16.
- Lanz, N.D., and Booker, S.J. (2012) Identification and function of auxiliary iron-sulfur clusters in radical SAM enzymes. *Biochim Biophys Acta* **1824**: 1196–1212.

- Lartillot, N., Lepage, T., and Blanquart, S. (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**: 2286–2288.
- Leger, M.M., Eme, L., Hug, L.A., and Roger, A.J. (2016) Novel hydrogenosomes in the microaerophilic jakobid *Stygiella incarcerata*. *Mol Biol Evol* **33**: 2318–2336.
- Lloyd, D., Lauritsen, F.R., and Degn, H. (1991) The parasitic flagellates *Trichomonas vaginalis* and *Tritrichomonas foetus* produce indole and dimethyl disulphide: direct characterization by membrane inlet tandem mass spectrometry. *J Gen Microbiol* **137**: 1743–1747.
- Loftus, B., Anderson, I., Davies, R., Alsmark, U.C., Samuelson, J., Amedeo, P., et al. (2005) The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**: 865–868.
- Mathus, T.L., Townsend, D.E., and Miller, K.W. (1995) Anaerobic biogenesis of phenol and p-cresol from L-tyrosine. *Fuel* **74**: 1505–1508.
- McDonnell, G., and Russell, A.D. (1999) Antiseptics and disinfectants: activity, action, and resistance. *Clin Microbiol Rev* **12**: 147–179.
- Newton, W.A., Morino, Y., and Snell, E.E. (1965) Properties of crystalline tryptophanase. *J Biol Chem* **240**: 1211–1218.
- Nicolet, Y., and Drennan, C.L. (2004) AdoMet radical proteins—from structure to evolution—alignment of divergent protein sequences reveals strong secondary structure element conservation. *Nucleic Acids Res* **32**: 4015–4025.
- Nyvtova, E., Stairs, C.W., Hrdy, I., Ridl, J., Mach, J., Paces, J., et al. (2015) Lateral gene transfer and gene duplication played a key role in the evolution of *Mastigamoeba balamuthi* hydrogenosomes. *Mol Biol Evol* **32**: 1039–1055.
- Scheline, R.R. (1968) Metabolism of phenolic acids by the rat intestinal microflora. *Acta Pharmacol Toxicol (Copenh)* **26**: 189–205.
- Schultz, T.W., Bryant, S.E., and Kissel, T.S. (1996) Toxicological assessment in *Tetrahymena* of intermediates in aerobic microbial transformation of toluene and p-xylene. *Bull Environ Contam Toxicol* **56**: 129–134.
- Selmer, T., and Andrei, P.I. (2001) p-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glycy radical enzyme catalysing the formation of p-cresol. *Eur J Biochem* **268**: 1363–1372.
- Selvaraj, B., Pierik, A.J., Bill, E., and Martins, B.M. (2013) 4-Hydroxyphenylacetate decarboxylase activating enzyme catalyses a classical S-adenosylmethionine reductive cleavage reaction. *J Biol Inorg Chem* **18**: 633–643.
- Siemensma, F.J. (2016). *Microworld, World of Amoeboid Organisms*. Kortenhoef, the Netherlands: World-wide electronic publication. URL <http://www.arcella.nl>
- Sivaraman, S., and Kirsch, J.F. (2006) The narrow substrate specificity of human tyrosine aminotransferase—the enzyme deficient in tyrosinemia type II. *FEBS J* **273**: 1920–1929.
- Spoelstra, S.F. (1978) Degradation of tyrosine in anaerobically stored piggery wastes and in pig feces. *Appl Environ Microbiol* **36**: 631–638.
- Stairs, C.W., Roger, A.J., and Hampl, V. (2011) Eukaryotic pyruvate formate lyase and its activating enzyme were acquired laterally from a Firmicute. *Mol Biol Evol* **28**: 2087–2099.
- Stairs, C.W., Eme, L., Brown, M.W., Mutsaers, C., Susko, E., Dellaire, G., et al. (2014) A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist *Pygsuia*. *Curr Biol* **24**: 1176–1186.
- Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Sutak, R., Hrdy, I., Dolezal, P., Cabala, R., Sedinova, M., Lewin, J., et al. (2012) Secondary alcohol dehydrogenase catalyzes the reduction of exogenous acetone to 2-propanol in *Trichomonas vaginalis*. *FEBS J* **279**: 2768–2780.
- Tsaousis, A.D., Ollagnier de, C.S., Gentekaki, E., Long, S., Gaston, D., Stechmann, A., et al. (2012) Evolution of Fe/S cluster biogenesis in the anaerobic parasite *Blastocystis*. *Proc Natl Acad Sci USA* **109**: 10426–10431.
- Updegraff, D.M. (1949) The production of phenol and para-cresol by marine bacteria. *J Bacteriol* **57**: 555–564.
- Yokoyama, M.T., and Carlson, J.R. (1981) Production of skatole and para-cresol by a rumen *Lactobacillus* sp. *Appl Environ Microbiol* **41**: 71–76.
- Zubacova, Z., Novak, L., Bublikova, J., Vacek, V., Fousek, J., Ridl, J., et al. (2013) The mitochondrion-like organelle of *Trimastix pyriformis* contains the complete glycine cleavage system. *PLoS One* **8**: e55417.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. *In vitro* growth of *M. balamuthi*. 6000 cells were added to 50 ml of PYGC medium in culture tissue flask and cultivated at 24°C. The cell density of *M. balamuthi* was counted using a hemacytometer.

Fig. S2. Analysis of the HPAD protein sequence. A) Protein sequence alignment of 4-hydroxyphenylacetate decarboxylase (HPAD), pyruvate formate-lyase (PFL) and the corresponding activating enzyme (AE). *M. balamuthi*HPAD (KX356607); *C. difficile*, *Clostridium difficile* HPAD (WP_021358752), AE (CBE01729); *C. scatologenes* HPAD (WP_029954956), AE (WP_029163541); *A. burkinensis*, *Anaeroarcus burkinensis* HPAD (WP_027938505), AE (WP_027937636); *M. balamuthi*, PFL1 (KX398023), AE1 (ADM47482); *M. balamuthi*, PFL2 (KX398024), AE2 (m51a1_g12196); *Ch. reinhardtii*, *Chlamydomonas reinhardtii* HPAD (CAF04129), AE (AAW32935); *N. frontalis*, *Neocallimastix frontalis*, HPAD (AAS06904), AE (AAS06905); *E. coli*, *Escherichia coli* HPAD (WP_000184850), AE (AJO82692); *C. ultunense*, *Clostridium ultunense* (WP_040354202); *T. pseudonana*, *Thalassiosira pseudonana*. Black arrows indicate C503 and G875 that likely harbor thyl and glycy radical respectively. Black dots indicate a conserved motif around G875 that interacts with the AE. Amino acid residues involved in the binding of 4-hydroxyphenylacetate are indicated by diamonds (S344, G345, C503, E505, H536 and E637)(Feliks et al., 2013; Selmer and Andrei, 2001). Amino acid residue numbers are based on those of the *C. scatologenes* HPAD sequence. Conserved SAM binding sites are indicated by triangles, white arrows indicate radical-activating patterns, including the CX₃CX₂C motif for the [4Fe4S] cluster of HPAD and PFL AEs (Nicolet and Drennan, 2004). The N-terminal methionine of AEs is indicated in bold, and stars (*) indicate the site of the stop codon. HPAD, 4-hydroxyphenylacetate decarboxylase; PFL, pyruvate formate-lyase; AE, activating enzyme. B) Phylogeny of HPAD. A tree for HPAD (193 taxa

and 655 sites) was built using Bayesian inference (BI). Bootstrap support (BP) and posterior probability (PP) values were calculated for each branch using RAXML and PhyloBayes respectively.

Fig. S3. Recombinant *MbHPAD* catalyses *p*-cresol production. (A) *MbHPAD* was heterologously expressed in *E. coli* with C-terminal 6xHis tag. The tagged *MbHPAD* was isolated from the cell lysate by affinity chromatography using an Ni-NTA agarose column. Lane 1, and 2; SDS-PAGE of molecular weight markers, and a fraction with isolated *MbHPAD* respectively. Lane 3, immunoblot analysis of the same fraction using anti 6xHis tag antibody. *MbHPAD* appeared as a 130 kDa band (calculated molecular weight is 121 kDa). Identity of *MbHPAD* was confirmed by mass spectrometry. Proteins bands of lower sizes are cleavage products. (B) HPLC analysis of *p*-cresol production catalysed by recombinant *MbHPAD*. The reaction mixture contained 1 ml of 50 mM Tris-HCl buffer (pH 7.4) 1 mM DTT, 1 mM MgCl₂, 2 mM 4-HPA, 2 mM SAM (Sigma), and the recombinant *MbHPAD*. The mixture was incubated under anaerobic conditions for 60 min. The reaction was stopped by the extraction of the mixture into 80% acetonitrile with 20% methanol for 30 min. at room temperature. The mixture was centrifuged at 13,000 × *g* for 15 min, and the extracted *p*-cresol was measured using HPLC with fluorescence detection (270 nm excitation/340 nm emission wavelengths) using a Dionex Ultimate 3000 system. Samples were injected into a Dionex Acclaim RSLC 120 C18 column that was equilibrated with a buffer containing 10% acetonitrile and 50 mM ammonium acetate. A linear gradient (0%–60%) of acetonitrile was with a flow rate 500 μl min⁻¹ for 10 min. Data were processed in Chromeleon 7.0 software, and the amounts of *p*-cresol and indole were determined based on the corresponding external calibration curves. (a) Positive control: 100 ng of *p*-cresol in the reaction mixture without substrate and the recombinant *MbHPAD*. (b) The reaction mixture with substrate and without *MbHPAD*. (c) The reaction mixture with substrate and with 1.8 mg or (d) 0.8 mg *MbHPAD*. The bar represents 25 × 10⁶ counts as detected by a fluorescence detector.

Fig. S4. Analysis of the HPPDC protein sequence. A) Alignment of the *M. balamuthi* 4-hydroxyphenylpyruvate decarboxylase (*MbHPPDC*) against selected eukaryotic and prokaryotic thiamine diphosphate-dependent decarboxylases. AbPPDC, *Azospirillum brasilense* phenylpyruvate decarboxylase; ScPPDC, *Saccharomyces cerevisiae* phenylpyruvate decarboxylase ARO10; EclPDC, *Enterobacter cloacae* indolpyruvate decarboxylase; LlkDcA, *Lactococcus lactis* branch chain 2-keto acid decarboxylase; ScPDC, *S. cerevisiae* pyruvate decarboxylase Pdc1p; ZmPDC, *Zymomonas mobilis* pyruvate decarboxylase. Core residues of the HH-motif in decarboxylases are marked with dots. Thiamine diphosphate-binding motifs are marked with diamonds. Residues suggested to influence substrate specificity are marked with arrows (Andrews and McLeish, 2012; Versees et al., 2007). Enzymes in this family can utilize various 2-keto acid substrates, such as pyruvate, which is preferentially metabolized by pyruvate decarboxylase (PDC), and substrates with an aromatic moiety, such as phenylpyruvate

and 4-HPP, which are substrates of phenylpyruvate decarboxylase (Kneen et al., 2011). The substrate specificity is influenced mainly by the amino acid composition of the hydrophobic pocket at the active site. Whereas aromatic residues are involved in the binding of pyruvate, these residues are replaced by aliphatic residues to accommodate large substrates including 4-HPP. Consistent with this, *MbHPPDC* possesses Met303 instead of the Y290 found in *Zymomonas mobilis* PDC or the F292 of *S. cerevisiae* PDC (Andrews and McLeish, 2012).

Fig. S5. Production of 4-hydroxyphenylacetaldehyde (4-HPAL) catalysed by recombinant *MbHPPDC*. (A) *MbHPPDC* was heterologously expressed in *E. coli* with C-terminal 6xHis tag. The tagged *MbHPPDC* was isolated from the cell lysate by affinity chromatography using an Ni-NTA agarose column. Lane 1, SDS-PAGE of a fraction with isolated *MbHPPDC*. Calculated molecular weight is 61 kDa. Lane 2, immunoblot analysis of the same fraction using anti 6xHis tag antibody. (B) HPLC analysis of 4-hydroxyphenylacetaldehyde (4-HPAL) production catalysed by recombinant *MbHPPDC*. The reaction mixture contained 50 mM Tris-KCl buffer (pH 7.2) 2 mM 4-HPP (Sigma), 1 mM MgCl₂, and 1 mM thiamine diphosphate. The mixture was incubated for 45 min at 25°C and the formation of 4-HPAL was determined using HPLC as described previously (Hazen et al., 1996). Briefly, samples were injected into a Dionex Acclaim RSLC 120 C18 column that was equilibrated with buffer A (5% methanol, 0.1% trifluoroacetic acid [pH 2.5]). A linear gradient (0%–35%) of buffer B (90% methanol, 0.1% trifluoroacetic acid [pH 2.5]) was applied following a 35% isocratic flow of buffer B for 3 min. Flow rate was 300 μl min⁻¹. The bar represents 25 × 10⁶ counts as detected by a fluorescence detector.

Fig. S6. Alignment of putative *M. balamuthi* family I aminotransferases (ATs) against selected eukaryotic and prokaryotic orthologs. Tyrosine transamination can be catalysed by the activity of aspartate aminotransferase (AAT) utilizing a broad range of substrates, including tyrosine, and by aromatic aminotransferases (AroATs), or specific TATs, all members of the PLP-dependent transaminase family I (Jensen and Gu, 1996). MbAT-1 I_α, *M. balamuthi* AT-1, subfamily family I_α (KX352251); AcaAT, *Acanthamoeba castellanii* aspartate AT (AAT)(XP_004340638); RnAAT, *Rattus norvegicus* AAT I_α, (P13221); EcaAT I_α, *Escherichia coli* AAT I_α (P00509); EcTAT I_α, *Escherichia coli* tyrosine aminotransferase I_α (P04693); ScHAT subfamily I_β, *Saccharomyces cerevisiae* histidinol-phosphate AT (P07172); MbAT-2 I_γ, *M. balamuthi* AT-2 subfamily I_γ (KX352252); MbAT-3 I_γ, *M. balamuthi* AT-3 subfamily I_γ (KX352253); RnTAT I_γ, *Rattus norvegicus* TAT I_γ (P04694); TcTAT I_γ, *Trypanosoma cruzi* TAT I_γ (EAN99619). Invariant residues of PLP-dependent AT (), invariant residues of family I AT () (Jensen and Gu, 1996).

Fig. S7. Tryptophanase phylogeny. A tree for tryptophanase (122 taxa and 359 sites) was built using Bayesian inference (BI). Bootstrap support (BP) and posterior probability (PP) values were calculated for each branch using RAXML and PhyloBayes respectively.

Table S1. List of primers.

Table S2. List of accession numbers.

RESEARCH

Open Access



Evolutionary loss of peroxisomes – not limited to parasites

Vojtěch Žárský and Jan Tachezy*

Abstract

Background: Peroxisomes are ubiquitous eukaryotic organelles that compartmentalize a variety of metabolic pathways that are primarily related to the oxidative metabolism of lipids and the detoxification of reactive oxygen species. The importance of peroxisomes is underscored by serious human diseases, which are caused by disorders in peroxisomal functions. Some eukaryotic lineages, however, lost peroxisomes. These organisms are mainly anaerobic protists and some parasitic lineages including *Plasmodium* and parasitic platyhelminths. Here we performed a systematic *in-silico* analysis of peroxisomal markers among metazoans to assess presence of peroxisomes and peroxisomal enzymes.

Results: Our analyses reveal an obvious loss of peroxisomes in all tested flukes, tapeworms, and parasitic roundworms of the order Trichocephalida. Intriguingly, peroxisomal markers are absent from the genome of the free-living tunicate *Oikopleura dioica*, which inhabits oxygen-containing niches of sea waters. We further map the presence and predicted subcellular localization of putative peroxisomal enzymes, showing that in organisms without the peroxisomal markers the set of these enzymes is highly reduced and none of them contains a predicted peroxisomal targeting signal.

Conclusions: We have shown that several lineages of metazoans independently lost peroxisomes and that the loss of peroxisomes was not exclusively associated with adaptation to anaerobic habitats and a parasitic lifestyle. Although the reason for the loss of peroxisomes from *O. dioica* is unclear, organisms lacking peroxisomes, including the free-living *O. dioica*, share certain typical r-selected traits: high fecundity, limited ontogenesis and relatively low complexity of the gene content. We hypothesize that peroxisomes are generally the first compartment to be lost during evolutionary reductions of the eukaryotic cell.

Reviewers: This article was reviewed by Michael Gray and Nick Lane.

Keywords: Peroxisome, Reduction, Parasitic helminths, *Oikopleura dioica*

Open Peer Review

Reviewed by Michael Gray and Nick Lane. For the full reviews, please go to the Reviewers' comments section.

Background

Peroxisomes are single membrane-bound organelles that proliferate by fission, although it has been shown that peroxisomes can emerge *de novo* from the endoplasmic reticulum [1]. Peroxisomes participate in a variety of metabolic functions, such as the reactive oxygen species detoxification, long-chain fatty acid beta-oxidation, plasmalogen synthesis, amino acid degradation, and purine

metabolism [2]. The diversity of peroxisomal functions is well exemplified by atypical peroxisomes, termed glycosomes, which compartmentalize the first seven enzymes of glycolysis and which are indispensable for the survival of *Trypanosoma brucei*, the causative agent of sleeping sickness [3]. Other types of peroxisomes with highly specialized roles have been described, such as glyoxysomes in plants and Woronin bodies in filamentous ascomycetes [4].

A unique group of proteins referred to as peroxins (Pexs) is required for peroxisome biogenesis and protein import. Peroxins mediate the post-translational import of folded proteins bound to cofactors or even of protein complexes [5]. Enzymes destined for the peroxisomal matrix are recognized by the specific cytosolic receptors

* Correspondence: tachezy@natur.cuni.cz
Department of Parasitology, Faculty of Science, Charles University in Prague,
Viničná 7, 128 44 Prague, Czech Republic



Pex5 and Pex7. Pex5 recognizes the peroxisomal targeting signal 1 (PTS1), which is composed of a canonical Ser-Lys-Leu tripeptide at the extreme C-terminus with common deviations of the canonical sequence [6]. Some other proteins carry a nonapeptide motif near the N-terminus termed PTS2, which is recognized by Pex7 [7].

A protein (cargo) carrying the PTS1 sequence is first recognized by soluble Pex5, which then interacts with the peroxisomal membrane proteins Pex14 and Pex13, which leads to a translocation into the peroxisomal lumen [8–10]. Pex5 is then either monoubiquitinated by Pex10 and Pex12 E3 ubiquitin ligases or polyubiquitinated by the Pex2 E3 ubiquitin ligase [11, 12]. The monoubiquitinated Pex5 is recycled to the cytoplasm by Pex1 and Pex6, both of which carry two ATPase associated with diverse cellular activities (AAA) [13] and which in mammals are recruited to the membrane by the Pex26 protein [14]. In *Saccharomyces cerevisiae*, the function of Pex26 is carried out by an unrelated protein Pex15, which is specific to yeast [15]. Alternatively, polyubiquitinated Pex5 is degraded by the proteasome.

Hydrophobic proteins targeted to the peroxisomal membrane are typically recognized by the cytosolic receptor Pex19, which binds to the peroxisomal membrane proteins Pex3 and Pex16 [16, 17]. Subsequently, the cargo protein is inserted into the peroxisomal membrane. Alternatively, some peroxisomal membrane proteins are first inserted into the ER and are then transported to the peroxisomal membrane via a process that depends on Pex19 and Pex3 [18]. In some organisms (e.g. in *Saccharomyces cerevisiae*) Pex16 is absent [19].

As the enzymatic content of peroxisomes is known to vary considerably among species or even different tissues, the components of the peroxisomal protein import system are the most reliable peroxisomal markers. A core set of at least 13 peroxins is common to the main eukaryotic lineages (Pex1,2,3,5,6,7,10,11,12,13,14,16,19, Fig. 1a); thus, these proteins were probably present in the last common ancestor of eukaryotes [20, 21].

A novel trafficking route between mitochondria and peroxisomes, which is mediated by the mitochondria-derived vesicles (MDV) was recently described [22]. Specific subset of MDVs that contain mitochondrial-anchored protein ligase (MAPL) was shown to fuse with a subpopulation of peroxisomes. However the function of MDVs in regard to peroxisomes is unknown and none of the discovered components of MDVs can be considered to be a specific peroxisomal marker. Thus, we didn't include these components to our dataset.

The essential role of peroxisomes is underlined by the increasing list of diseases that are associated with disorders of peroxisome biogenesis or even the dysfunction of a single peroxisomal enzyme [23]. However, some unicellular

eukaryotes are able to live without peroxisomes. These eukaryotes includes parasitic protists that live in oxygen-poor environments, such as *Giardia intestinalis*, *Entamoeba histolytica* and *Trichomonas vaginalis*, and the loss of peroxisomes has also been described in intracellular parasites of the Apicomplexa (e.g., *Plasmodium falciparum*, *Cryptosporidium parvum*) and Microsporidia (e.g., *Encephalitozoon cuniculi*) [4, 21]. However peroxisomes were identified in an apicomplexan *Toxoplasma gondii* [24]. More recently, the loss of many genes associated with peroxisomes was observed in the genomes of certain parasitic helminths, including tapeworms and flukes and a loss of peroxisomes in those lineages was suggested [25, 26]. Here, we executed a large-scale analysis of metazoan genomes to assess the presence of the systems required for the biogenesis and metabolic function of peroxisomes. Our results revealed the unexpected loss of peroxisomal functions and peroxisomes in metazoans including not only parasitic species but also, surprisingly, an aerobic free-living organism.

Results and discussion

Loss of peroxins and peroxisomes

To assess the presence of peroxisomal components, we collected an extensive dataset of the predicted proteomes of 111 metazoans based on completed or advanced genome sequencing projects (see Additional file 1: Table S1). We then assessed the presence or absence of 14 peroxins conserved in metazoans [21] (see Additional file 2: Table S2) by assigning the protein sequences to the evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG) database of orthologous groups, which was constructed from representative metazoan sequences [27]. For homology analysis, we used a highly sensitive profile-hidden Markov model (HMMER) search algorithm [28].

We found a broad set of peroxins in most metazoans (Fig. 1a). However, not all of these metazoans contained a complete set of peroxins. Peroxins such as Pex26 appeared to be repeatedly lost during evolution, as we observed the lack of this protein in nematodes and other invertebrates. We also observed the loss of the PTS2-binding protein Pex7 in all nematodes which is in agreement with lack of the PTS2 pathway in *Caenorhabditis elegans* and other lineages [29–31]. Some other peroxins might have diverged beyond recognition by sequence analysis, as was proposed for a hypothetical *Trypanosoma brucei* Pex3 [32].

The extreme case of the reduction of the peroxisomal components is the complete loss of peroxisomes. It has been established experimentally that the loss of certain peroxins that are essential for the peroxisome biogenesis, such as Pex3 and Pex19, leads to the complete loss of peroxisomes [1]. Thus, we interpreted absence of these

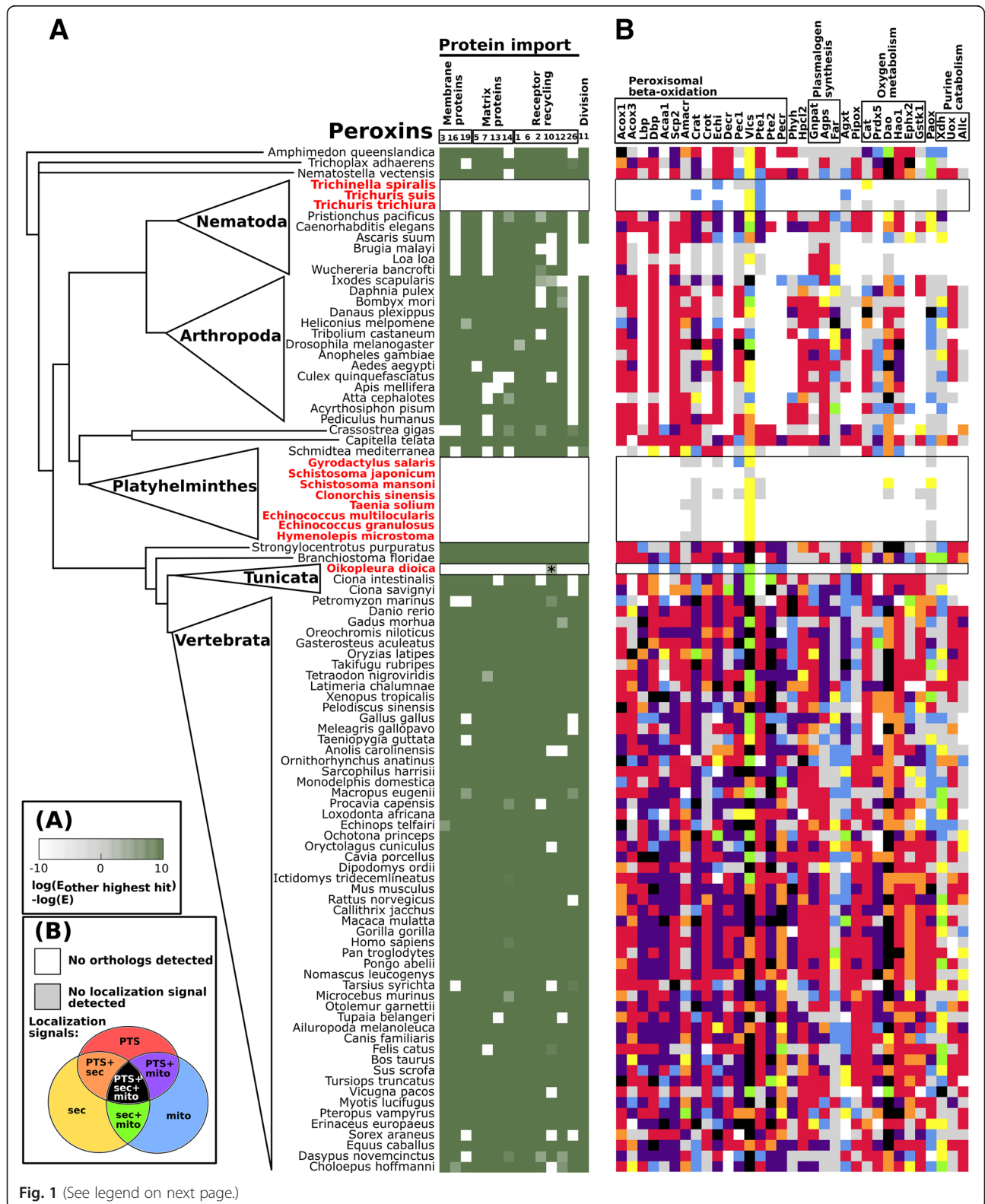


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Distribution of peroxins (A), and possible peroxisomal enzymes (B) in metazoan genomes. Organisms lacking peroxins are highlighted in red.

a Identification of peroxins. The green squares indicate identification with high confidence. The identification of Pex10 in the *Oikopleura dioica* genome with lower confidence (marked by an asterisk) was rejected as a false-positive based on phylogenetic analysis (Figure S1). **b** Identification of putative peroxisomal enzymes and enzymes of mitochondrial beta-oxidation. Predicted subcellular localization is indicated in the following colors: grey, no targeting signal detected; red, peroxisomal targeting sequence type 1 or 2 (pts); blue, mitochondrial localization signal (mito); and yellow, secretory pathway signal peptide (sec). Combinations of several different localization signals are shown as follows: violet, pts and mito; orange, pts and sec; green, sec and mito; and black, pts, mito and sec. (Acox1 - Acyl-CoA oxidase1, Acox3 - Acyl-CoA oxidase 3, Lbp - L-bifunctional protein, Dbp - D-bifunctional protein, Acaa1 - Peroxisomal beta-ketothiolase 1, Scp2 - Peroxisomal beta-ketothiolase 2, Amacr - Alpha-methylacyl-CoA racemase, Crat - Carnitine acetyltransferase, Crot - Carnitine octanoyltransferase, Ech1 - Enoyl Coenzyme A hydratase 1, Decr - Peroxisomal 2,4-dienoyl-CoA reductase 2, Pec1 - Peroxisomal 3,2-trans-enoyl-CoA isomerase, Vlcs - Very-long-chain acyl-CoA synthetase, Pte1 - Acyl-CoA thioesterase 2, Pte2 - Acyl-CoA thioesterase 1B, Phyh - Phytanoyl-CoA 2-hydroxylase, Hpcl2 - 2-Hydroxyphytanoyl-CoA lyase, Gnpat - Dihydroxyacetone phosphate acyltransferase, Agps - Alkyldihydroxyacetone phosphate synthase, Far - Fatty acyl-CoA reductase 2, Agxt - Alanine:glyoxylate aminotransferase, Pipox - Peroxisomal sarcosine oxidase/L-pipecolate oxidase, Cat - Catalase, Prdx5 - Peroxiredoxin V, Dao - d-amino acid oxidase, Hao1 - Hydroxyacid oxidase 1, Ephx2 - Epoxide hydrolase, Gstk1 - Glutathione S-transferase class Kappa, Paox - N1-acetylspermine/spermidine oxidase, Xdh - Xanthine dehydrogenase, Uox - Uricase, Allc - Allantoicase)

peroxins as evidence for the loss of peroxisomes per se in the examined organisms. We did not identify any peroxins in the genome of several lineages of parasitic helminths: flukes (*Schistosoma japonicum*, *S. mansoni*, and *Clonorchis sinensis*), tapeworms (*Taenia solium*, *Echinococcus multilocularis*, *E. granulosus*, and *Hymenolepis microstoma*) and a monogenean (*Gyrodactylus salaris*) as was described previously [25, 26]; all of these species belong to the Neodermata group of parasitic flatworms. Furthermore, we observed the independent loss of peroxins in the parasitic roundworms of the order *Trichocephalidae*, including *Trichinella spiralis* and the whipworms *Trichuris trichiura* and *Trichuris suis*. Most surprisingly, in addition to parasites, we did not identify any convincing orthologs of peroxins in the free living tunicate *Oikopleura dioica*. This organism belongs to the class *Appendicularia*, occupies the pelagic zone of the world's oceans, and exhibits a high rate of oxygen consumption [33] as well as several peculiar features: (i) its genome is extremely small (approximately 75 Mb), (ii) it has a short generation time (as short as 24 hours [34]), and (iii) adults of this species have a simplified body structure that resembles the tadpole-like larvae of most tunicates [35, 36]. Interestingly, the related marine tunicates *Ciona intestinalis* and *Ciona savignyi* of class *Ascidacea* harbor a regular set of peroxins. Our analysis revealed only one possible hit of Pex10, based on its zinc-finger domain, in *O. dioica*. However, the predicted protein lacks a typical N-terminal hydrophobic domain, and phylogenetic analysis of the zinc finger domain revealed that it is likely a false-positive hit as it forms a monophyletic clade with mammalian NHLRC1 proteins, while the Pex10 sequences form a distinct clade (Additional file 3: Figure S1).

Peroxisomal pathways

To further corroborate our findings, we predicted the presence of peroxisomal matrix-resident enzymes based

on predicted PTS1 and PTS2 motifs [7, 37]. For the detection of the signals we used PTS1/2 amino acid motifs according to the PSORT II software [38] and we also included sequences of mammalian PTS1, which do not comply with these motifs. We focused on the components of main peroxisomal functional pathways, including oxygen metabolism, beta-oxidation, plasmalogen biosynthesis, amino acid metabolism, and purine catabolism [2] (Fig. 1b, Additional file 4: Table S3). We further predicted mitochondrial and secretory signal peptides using TargetP software [39].

In the metazoans lacking Pexs identified above, most of the conserved peroxisomal matrix enzymes have been lost, and the few enzymes that are retained contained no apparent peroxisomal localization signals (Fig. 1b). Pathways that are exclusively found in organisms harboring peroxisomes include peroxisomal beta-oxidation (acyl-CoA oxidase1/Acox1, L-bifunctional protein/Lbp, peroxisomal beta-ketothiolase 1/Acaa1, and peroxisomal 2,4-dienoyl-CoA reductase 2/Decr) and plasmalogen synthesis (fatty acyl-CoA reductase 1/Far1, dihydroxyacetone phosphate acyltransferase/Gnpat and alkyldihydroxyacetone phosphate synthase/Agps). Interestingly, a gene coding for the typical peroxisomal enzyme catalase was detected in the genome of *Trichinella spiralis*; the predicted protein, however, lacks a PTS signal but has an N-terminal signal peptide instead.

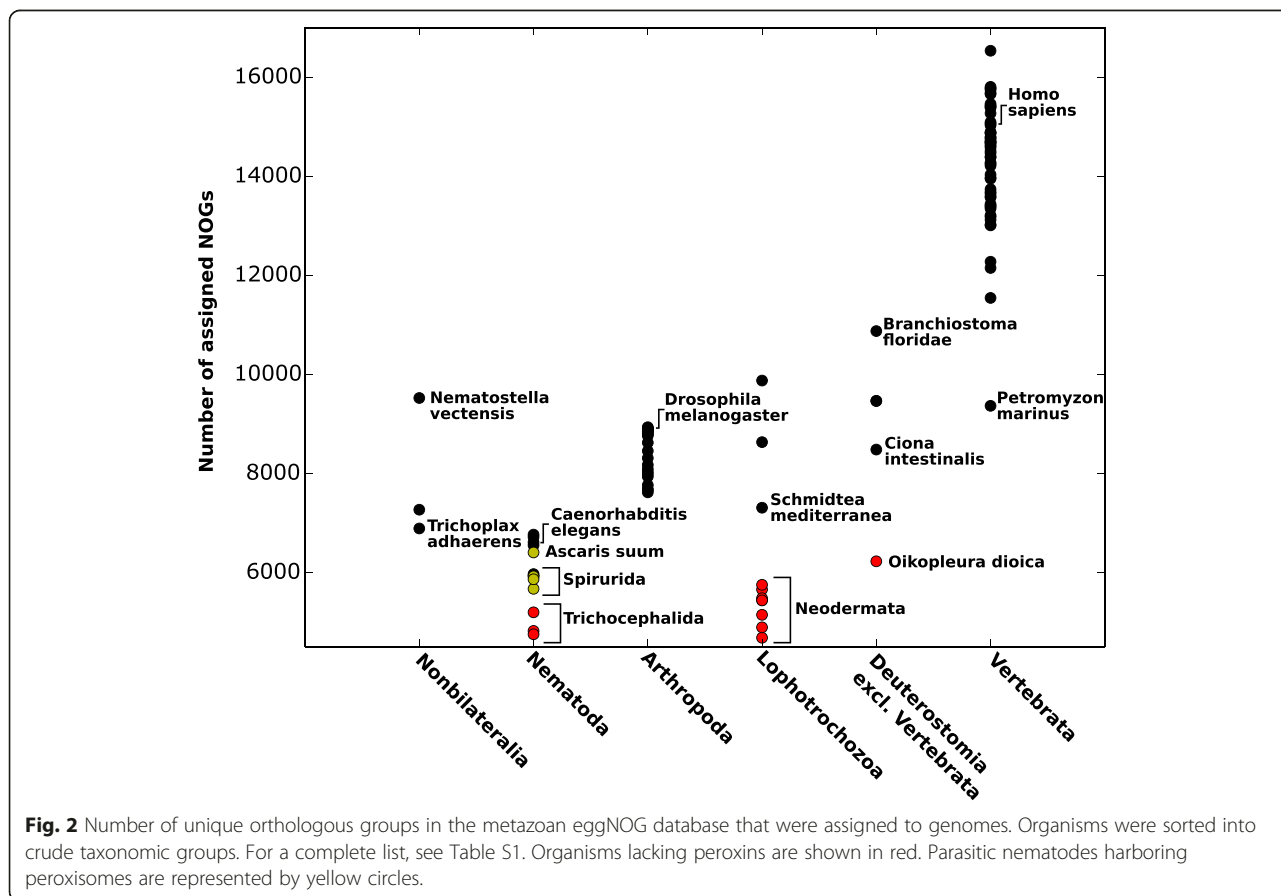
The prediction of targeting signals in the putative peroxisomal proteins revealed that enzymes typically possess either peroxisomal or mitochondrial localization signals or both, indicating a metabolic interdependence between peroxisomes and mitochondria and, possibly, the dual targeting of some enzymes (Fig. 1b). This is a well known phenomenon exemplified by e.g. the human isoform of carnitine acetyltransferase (Crat), which is important for both the mitochondrial and peroxisomal beta-oxidation in mammalian cells, and contains a mitochondrial localization signal at the N-terminus and a typical PTS1 sequence at the C-terminus [40].

Loss of peroxisomes and the genome complexity

The identified metazoan lineages lacking peroxisomes have very different phylogenetic positions and lifestyles: *Trichinella* and *Trichuris* develop in vertebrate cells and/or host tissues; schistosomes and tapeworms first develop in intermediate hosts and then reside in the bloodstream and the intestine of their definitive hosts, respectively; *Gyrodactylus salaris* is an ectoparasite of fish; and *Oikopleura* is a marine filter feeder. However, in ecological terms, all of these species share certain typical r-selected traits: high fecundity, early maturation, and simplified ontogenesis. To assess the relative complexity of their genomes and to evaluate the loss of peroxisomes in the genomic context, we quantified unique orthologous groups of proteins that were detected in the predicted proteomes (Fig. 2; Additional file 1: Table S1). It is apparent that lineages lacking peroxisomes contain a markedly reduced repertoire of conserved metazoan orthologs compared to their relatives harboring peroxisomes. Therefore, the loss of peroxisomes clearly accompanied the reductive evolution of the respective metazoan lineages, as reflected by genome shrinkage and gene loss.

Conclusions

In addition to the known loss of peroxisomes in parasitic flatworms (Neodermata) we observed the loss of peroxisomes in parasitic roundworms of the order Trichocephalida and most intriguingly in a free-living tunicate *Oikopleura dioica*. The loss of peroxisomes from Neodermata and Trichocephalida likely reflects their adaptation to an obligate parasitic lifestyle, which is specifically associated with the adaptation to anaerobiosis of some of the life stages and/or the reduced need for lipid synthesis and turnover. However, not all parasitic worms have lost peroxisomes, as we detected the presence of peroxisomal markers in the genomes of some helminths, such as *Ascaris suum*. The presence of peroxisomes in these organisms can be explained by aerobic metabolism during L1 and L2 larval stages of *Ascaris*, which live in an external environment in eggs, and in L3 larvae that migrate from the intestines to the lungs via the liver after egg ingestion. Only L4 larvae and adults live in the oxygen-poor environment of intestines and utilize anaerobic metabolism. In this context, the absence of peroxisomal markers in *Oikopleura dioica*, a free-living tunicate that inhabits oxygen-containing niches of sea waters and that exhibits a high rate of oxygen



consumption [33], is highly unexpected. Similar to other aerobes, *O. dioica* possesses classical mitochondria that utilize oxygen and oxidize fatty acids for ATP synthesis. Why might this organism have lost peroxisomes? We hypothesize that the loss of peroxisomes is associated with the reduction of genomic content and generally with a shift towards r-selected traits, which is typical for parasitic organisms. However, there might be other advantages of peroxisomal loss, e.g., rendering the organisms resistant to xenobiotics that become activated in the peroxisomal lumen in response to frequent redox reactions [41].

Comparative genomics shows that most of the typical cell compartments were present in the common eukaryotic ancestor [42]. However there are known cases where these compartments were radically functionally and/or structurally reduced, as the divergent anaerobic mitochondria of *Giardia intestinalis* called mitosomes [43], fragmented ER of *Entamoeba histolytica* [44], or absence of structured Golgi in several protist lineages [45, 46]. Still in all these cases the organelle and the apparatus for its biogenesis are to various extents conserved and retained. On the contrary it has been shown that peroxisomes have been lost on several occasions, not only in parasitic and anaerobic lineages. We thus further hypothesize that peroxisomes are generally the first conserved eukaryotic compartment to be lost during reductive evolution of the cell.

As the loss of peroxisomes was repeatedly observed in various metazoan lineages, it is likely that new lineages lacking peroxisomes will emerge and that new genomic and functional data will provide additional information to reveal the circumstances in which selection pressure resulted in the loss of peroxins and, consequently, peroxisomes.

Methods

Orthologous group assignment

Orthologous groups of metazoan proteins were extracted from the eggNOG database, ver. 3 [27]. Predicted metazoan protein sequences were searched against the HMM profiles created from the alignments of eggNOG groups using HMMER software [28]. Hits displaying an e-value greater than $1e^{-3}$ were regarded as false positives. The hits were scored based on their e-value (Ehit) and the next highest hit e-value (Enext) as follows: $\text{Score} = \log(\text{Enext}) - \log(\text{Ehit})$. Thus, only the best hits displayed a positive score.

Subcellular localization prediction

N-terminal targeting sequences were predicted using TargetP software with default settings [39]. The peroxisomal targeting signals were predicted using the following regular expressions: PTS1: '[SAC][KRHN][LIM]'

or 'ASL', 'THL', 'LKL', 'SKV', 'KKL', 'SQL', or 'PRL' at the C-terminus; PTS2: '[RK][LVIQ]..[LVIHQ][LSGAK]..[HQ][LAF]' within the first 100 amino acids [2, 38].

Phylogenetic analysis

Homologs of *O. dioica* GSOIDT00013970001 were retrieved from the SwissProt database using BLAST [47]. Sequences were aligned using MAFFT [48], trimmed and the phylogeny was constructed with Phylml [49] using the WAG substitution matrix.

Reviewers' comments

Reviewer's report 1 (Michael Gray, Dalhousie University, Canada)

Summary: This manuscript describes the results of a comprehensive survey to detect peroxisome gene markers (encoding peroxins and other characteristic peroxisomal proteins) in complete or substantially complete animal genome sequences. The results extend previous results that suggested the apparent absence of peroxisomes in certain parasitic (anaerobic) helminths. Unexpectedly, the authors also report the absence of peroxisomal markers from the genome of a free-living (and oxygen-consuming) tunicate. This study is straightforward in its execution and in the results obtained. The authors employed rigorous methods of gene identification, including HMM profile searches, to ensure that peroxisomal markers being sought were not missed. The results are clear-cut: Fig. 1 is especially effective in summarizing the main message. The authors point out that peroxisomal loss in animals correlates with a gene content of relatively low complexity (Fig. 2), from which they argue that peroxisomes may be the first compartment to be lost during evolutionary down-sizing of the eukaryotic cell.

Author's response: We thank the reviewer for his appreciation of our work.

Recommendations: I have no substantive criticisms of the work, which overall is novel and will be of interest to a wide readership. The results emphasize the importance of taking a comprehensive and rigorous approach to questions of presence/absence of genes/organelles before drawing firm evolutionary conclusions. Extension of this work to other groups of eukaryotes, as sufficient genome data become available, should certainly be done: fungi are an obvious first choice, but eukaryotic microbes (protists) should also be examined by the same approach.

Author's response: We agree with the reviewer that our analysis of peroxisomal markers should be extended to a wider range of eukaryotes in the future. In fact, we performed the preliminary analysis of peroxisomal markers across major eukaryotic lineages, however, we didn't find any novel losses of peroxisomes except these

in anaerobic protists, some apicomplexans and microsporidians that are mentioned in the Background section. Nevertheless, we can expect more variation in individual peroxins (losses and gains) as well as more functional diversity in various lineages of unicellular eukaryotes than in metazoans. This has been shown for example in the case of mitochondria that displayed considerably higher level of diversity in unicellular eukaryotes in comparison to metazoans. From this point of view, described losses of peroxisomes within metazoans including *Oikopleura dioica* are really unexpected and thus we would prefer the paper to be focused on the analysis of metazoans to keep the publication more straightforward.

Minor issues: While reading through the manuscript, I did encounter a few grammatical issues that I flagged in the attached PDF. In particular, the term 'r-selected traits' should be defined for the benefit of readers who will be unfamiliar with the concept.

Author's response: We edited the manuscript according to the reviewer's recommendations. We also explained the concept of 'r-selected traits'.

Reviewer's report 2 (Nick Lane, UCL, United Kingdom)

Summary: This is an interesting and short phylogenetic paper, which shows that peroxisomes have almost certainly been lost from various lines of metazoans over evolutionary time, not only in parasitic flukes, tapeworms and roundworms, but also in some free-living tunicates that inhabit oxygenated waters. The findings extend and strengthen earlier observations of loss of peroxisomes. These results are of interest in themselves, but could probably benefit from a little more discussion about their evolutionary significance.

Author's response: We thank the reviewer for his appreciation of our work.

Recommendations: The paper is clear and crisp, with little extraneous discussion. I would be happy to see it published more or less as it is. In my view, the paper would be stronger if it had included an ultrastructural study - I appreciate that it is hard to show the absence of structures by EM, but while the genetic dataset is fairly convincing evidence (and even if not, certainly worth publishing) the paper would be stronger still if it had included immunogold-labeling in the groups with or without peroxisomes, ER and so on. As Shakespeare said, give me the ocular proof. But I appreciate that this is not easily done, and I reiterate that the paper could be published as it is.

Author's response: While preparing the paper we faced an obvious principal problem as proving a nonexistence of anything is usually difficult and to some extent dubious. We were considering the electron microscopy to confirm the absence of peroxisomes, which are traditionally

visualized by detecting the catalase activity using 3,3' diaminobenzidine tetrahydrochloride (DAB). However e.g. in *Plasmodium falciparum* peroxidase active organelles were discovered although there is a well supported consensus that there are no peroxisomes present and interpretation of such results is difficult [50]. Thus, we decided to rely on the genomic data which shows a comprehensive set of coding genes. In organisms with predicted loss of peroxisomes we carefully checked for possible peroxin homologs in the predicted protein sequences, genome sequences translated in all 6 frames and in the transcriptomic data.

Recommendations: I would have appreciated a little more discussion of the evolutionary significance of the findings. For example, de Duve long argued that peroxisomes had an endosymbiotic origin, albeit with declining evidence over the last decade. I wondered if the authors have anything to say about the ease of loss of peroxisomes in relation to hydrogenosomes or mitosomes as organelles. How do they disappear? Is their loss linked with changes in the ER? These are all questions not directly addressed in the paper, and so perhaps not relevant, but some short discussion of these issues would not go amiss. I find it interesting that it seems to be so easy to lose peroxisomes but far less easy to lose extra membranes from around secondary or tertiary chloroplasts, or to lose mitochondria completely, or indeed ER. Why the peroxisomes and maybe Golgi? Does it relate to nuclear membrane dissolution during mitosis or meiosis?

Author's response: The origin of peroxisomes is a matter of discussion since their discovery by Christian de Duve. He proposed an endosymbiotic origin of the organelle [51, 52] that is supported mainly by a post-translational import of proteins to the peroxisomes, and the biogenesis of new peroxisomes by fission of pre-existing peroxisomes, the features known for endosymbiotic organelles such as mitochondria. Later it was, however, discovered that peroxisomes can arise de novo from the endoplasmic reticulum by knock-in of essential peroxin gene into peroxisome-lacking mammalian cells and yeast mutants and that peroxisomes are formed de novo also in normal cells [1, 53]. Some of the components of the peroxisomal protein import machinery are also homologous to the ERAD (Endoplasmic-reticulum-associated protein degradation) pathway of the endoplasmic reticulum [20]. These observations support an alternative hypothesis of endogenous origin of peroxisomes.

Our finding of relatively common evolutionary loss of peroxisomes is in a contrast with primary endosymbiotic organelles (mitochondria, chloroplasts) which are known to undergo reductions but to our knowledge there are no known cases of their complete loss. What is a reason for such a difference? Based on Blobel's idea of membranes

heredity [54], we can speculate that ER represent a source of membranes for peroxisome formation, which allows high dynamics of peroxisomes in respect of their size and number, *de novo* formation and eventually loss upon environmental conditions. However, mitochondrial membranes cannot arise *de novo* and cells may have mechanisms to prevent such losses. The other reason might be that none of peroxisomal functions is really essential, whereas mitochondria possess iron-sulfur cluster assembly (ISC) machinery that seems to be indispensable for all viable cells (of course we can discuss exceptions when ISC machinery is replaced by other systems). We do not think that easier loss of peroxisomes is related to nuclear membrane dissolution during mitosis or meiosis because e.g. Trichomonad mitosis is the closed mitosis without dissolution of the nuclear membrane, but they do not have peroxisomes.

Recommendations: Fig. 2 is interesting but seems a bit minimal in terms of identifying specific groups. For example, where is *C. intestinalis* and other groups mentioned in the text? I appreciate that this information is available in Table S1, but that is a frustrating format to compare information. A few more labels on Fig. 2 showing key comparators mentioned in the text would be valuable for those who are interested but don't want to spend a long time wading through the SI.

Author's response: We added labels in Fig. 2 for selected model species and organisms related to the proposed lineages missing peroxisomes.

Minor issues: I was uncertain about whether the presence/absence data for peroxins etc. are based on proteomes (as stated on page 5) or whether some data are based on full genome sequences.

Author's response: Our analysis was based exclusively on the *in silico* predicted proteomes from genomic data. When the loss of peroxisomes was suspected we further checked the available genomic and transcriptomic data. We clarified this issue at pages 5 and 7 in the text.

Minor issues: Many parasites have different phases of their life cycle, and I imagine (though don't know much about it) that their proteomes could differ significantly at different stages. It wouldn't be surprising to find that some parasites have peroxisomes at some stages of their life cycle but not others. This was not discussed at all, or at least I didn't notice it; and is not clear to me from the data.

Author's response: It is likely that in the parasitic helminths with peroxisomes the significance of peroxisomal functions, and the size and quantity of peroxisomes will be altered during the life cycle as this is well described in the case glycosomes of trypanosomatids [55]. However we are not aware of any such data in the case of analyzed parasitic helminths. Moreover, our data are based on *in silico* predictions of proteins, not on proteomic studies that could be affected by variations in proteomes.

Minor issues: I wondered if there are any known non-canonical pathways of peroxisome targeting that do not involve peroxins. Plainly these are missing, and I accept the conclusion that the peroxisomes have most likely been lost, given that other aspects of peroxisome metabolism are also missing. Even so, these details could at least be mentioned in the discussion.

Author's response: Indeed a novel trafficking route between mitochondria and peroxisomes has been described. We mention this pathway in the Background section.

Minor issues: How does *Oikopleura dioica* survive in oxygenated waters without peroxisomes? This is genuinely an interesting finding, and it might be that r-selection is indeed sufficient to explain the loss. It would be good to know how large population sizes (N_e) tend to be, and developmental time compared with related tunicates such as *C. intestinalis*, which do have peroxisomes. To a degree this probably correlates with NOG data in Fig. 2, but not entirely. I also wondered whether *O. dioica* had other mechanisms of oxygen detoxification, such as an alternative oxidase or uncoupling proteins. This is probably reasonably easy to check in their proteomic data and would be worth commenting on.

Author's response: The differences between *Ciona intestinalis* (Ascidacea) and *Oikopleura dioica* (Appendicularia) are truly striking. The adults of *C. intestinalis* are up to 15 cm long sessile sea squirts and their life cycle takes about 2 months. On the other hand the adults of *O. dioica* are pelagic tunicates of size between 0.5 and 1mm that resemble the tadpole-like larva of ascidiaceans with a generation time as short as 24 hours [34]. Comparison of two distinct haplotypes of *O. dioica* revealed a high estimate of population mutation rate, which is consistent with large effective population size and/or high mutation rate per generation [56].

Interestingly, our searches for alternative oxidases (AOX) revealed homologs of AOX in the genomes of *C. intestinalis* and *C. savignyi*. Counterintuitively we did not identify AOX in the genome of *O. dioica*. Furthermore, unlike most metazoans, *O. dioica* possesses only a single uncoupling protein (UCP4).

Additional files

Additional file 1: Table S1. Table of genomes used in the analysis with the count of assigned orthologous groups (NOGs). (XLS 59 kb)

Additional file 2: Table S2. Table of peroxins identified. (XLS 677 kb)

Additional file 3: Figure S1. Phylogeny of Pex10 and other zinc-finger domain containing proteins showing, that *O. dioica* GSOIDT00013970001 sequence isn't monophyletic with eukaryotic Pex10 sequences. Bootstrap supports are shown. (PDF 351 kb)

Additional file 4: Table S3. Table of putative peroxisomal enzymes and the prediction of their localization. (XLS 7288 kb)

Abbreviations

PTS: Peroxisomal targeting signa; Pex: Peroxin; ER: Endoplasmic reticulum; L1-4: Larval stages; ATP: Adenosine triphosphate; Acox1: Acyl-CoA oxidase 1; Acox3: Acyl-CoA oxidase 3; Lbp: L-bifunctional protein; Dbp: D-bifunctional protein; Acaa1: Peroxisomal beta-ketothiolase 1; Scp2: Peroxisomal beta-ketothiolase 2; Amacr: Alpha-methylacyl-CoA racemase; Crat: Carnitine acetyltransferase; Crot: Carnitine octanoyltransferase; Ech1: Enoyl Coenzyme A hydratase 1; Decr: Peroxisomal 2,4-dienoyl-CoA reductase 2; Pec1: Peroxisomal 3,2-trans-enoil-CoA isomerase; Vlcs: Very-long-chain acyl-CoA synthetase; Pte1: Acyl-CoA thioesterase 2; Pte2: Acyl-CoA thioesterase 1B; Phyh: Phytanoyl-CoA 2-hydroxylase; Hplcl2: 2-Hydroxyphytanoyl-CoA lyase; Gnpat: Dihydroxyacetone phosphate acyltransferase; Agps: Alkyldihydroxyacetone phosphate synthase; Far: Fatty acyl-CoA reductase 2; Agxt: Alanine:glyoxylate aminotransferase; Pipox: Peroxisomal sarcosine oxidase/L-pipecolate oxidase; Cat: Catalase; Prdx5: Peroxiredoxin V; Dao: d-amino acid oxidase; Hao1: Hydroxyacid oxidase 1; Ephx2: Epoxide hydrolase; Gstk1: Glutathione S-transferase class Kappa; Paox: N1-acetylspermine/spermidine oxidase; Xdh: Xanthine dehydrogenase; Uox: Uricase; Allc: Allantoicase.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

VZ conceived the original idea, designed the study, and carried out the orthologous group assignment, subcellular localization prediction, and phylogeny analysis. JT participated in the study design, coordination, writing and revising the manuscript. Both authors read and approved the manuscript.

Acknowledgements

This work was supported by the Czech Grant Foundation (P305/11/1061) and the Biotechnology and Biomedicine Center of the Academy of Sciences and Charles University (BIOCEV) (CZ.1.05/1.1.00/02.0109) from the European Regional Development Fund. V. Z. was supported by Charles University in Prague (GAUK 573112). We thank Ivan Hrdy and Viktor Zarsky for helpful comments and discussions.

Received: 26 October 2015 Accepted: 14 December 2015

Published online: 23 December 2015

References

- Hoepfner D, Schildknecht D, Braakman I, Philippsen P, Tabak HF. Contribution of the endoplasmic reticulum to peroxisome formation. *Cell*. 2005;122:85–95.
- Wanders RJA, Waterham HR. Biochemistry of mammalian peroxisomes revisited. *Annu Rev Biochem*. 2006;75:295–332.
- Opperdoes FR, Borst P. Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei*: The glycosome. *FEBS Lett*. 1977;80:360–4.
- Gabaldón T. Peroxisome diversity and evolution. *Philos Trans R Soc Lond B Biol Sci*. 2010;365:765–73.
- Rucktäschel R, Girzalsky W, Erdmann R. Protein import machineries of peroxisomes. *Biochim Biophys Acta BBA - Biomembr*. 1808;2011:892–900.
- der Leij IV, Franse MM, Elgersma Y, Distel B, Tabak HF. PAS10 is a tetratricopeptide-repeat protein that is essential for the import of most matrix proteins into peroxisomes of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci*. 1993;90:11782–6.
- Marzoch M, Erdmann R, Veenhuis M, Kunau WH. PAS7 encodes a novel yeast member of the WD-40 protein family essential for import of 3-oxoacyl-CoA thiolase, a PTS2-containing protein, into peroxisomes. *EMBO J*. 1994;13:4908–18.
- Albertini M, Rehling P, Erdmann R, Girzalsky W, Kiel JAKW, Veenhuis M, et al. Pex14p, a Peroxisomal membrane protein binding both receptors of the two PTS-dependent import pathways. *Cell*. 1997;89:83–92.
- Barnett P. The peroxisomal membrane protein Pex13p shows a novel mode of SH3 interaction. *EMBO J*. 2000;19:6382–91.
- Meinecke M, Cizmowski C, Schliebs W, Krüger V, Beck S, Wagner R, et al. The peroxisomal importomer constitutes a large and highly dynamic pore. *Nat Cell Biol*. 2010;12:273–7.
- Chang C-C, Warren DS, Sacksteder KA, Gould SJ. Pex12 interacts with Pex5 and Pex10 and acts downstream of receptor docking in peroxisomal matrix protein import. *J Cell Biol*. 1999;147:761–74.
- Platta HW, Magraoui FE, Bäumer BE, Schlee D, Girzalsky W, Erdmann R. Pex2 and Pex12 function as protein-ubiquitin ligases in peroxisomal protein import. *Mol Cell Biol*. 2009;29:5505–16.
- Thoms S, Erdmann R. Peroxisomal matrix protein receptor ubiquitination and recycling. *Biochim Biophys Acta BBA - Mol Cell Res*. 1763;2006:1620–8.
- Matsumoto N, Tamura S, Fujiki Y. The pathogenic peroxin Pex26p recruits the Pex1p–Pex6p AAA ATPase complexes to peroxisomes. *Nat Cell Biol*. 2003;5:454–60.
- Birschmann I, Stroobants AK, Van Den Den BM, Schäfer A, Rosenkranz K, Kunau W-H, et al. Pex15p of *Saccharomyces cerevisiae* provides a molecular basis for recruitment of the AAA peroxin Pex6p to peroxisomal membranes. *Mol Biol Cell*. 2003;14:2226–36.
- Fang Y, Morrell JC, Jones JM, Gould SJ. PEX3 functions as a PEX19 docking factor in the import of class I peroxisomal membrane proteins. *J Cell Biol*. 2004;164:863–75.
- Sacksteder KA, Jones JM, South ST, Li X, Liu Y, Gould SJ. Pex19 binds multiple peroxisomal membrane proteins, is predominantly cytoplasmic, and is required for peroxisome membrane synthesis. *J Cell Biol*. 2000;148:931–44.
- van der Zand A, Braakman I, Tabak HF. Peroxisomal membrane proteins insert into the endoplasmic reticulum. *Mol Biol Cell*. 2010;21:2057–65.
- Hetteema EH, Girzalsky W, van den Berg M, Erdmann R, Distel B. *Saccharomyces cerevisiae* Pex3p and Pex19p are required for proper localization and stability of peroxisomal membrane proteins. *EMBO J*. 2000;19:223–33.
- Gabaldón T, Snel B, Van Zimmeren F, Hemrika W, Tabak H, Huynen MA. Origin and evolution of the peroxisomal proteome. *Biol Direct*. 2006;1:8.
- Schlüter A, Fourcade S, Ripp R, Mandel JL, Poch O, Pujol A. The evolutionary origin of peroxisomes: An ER-peroxisome connection. *Mol Biol Evol*. 2006;23:838–45.
- Neuspiel M, Schauss AC, Braschi E, Zunino R, Rippstein P, Rachubinski RA, et al. Cargo-selected transport from the mitochondria to peroxisomes is mediated by vesicular carriers. *Curr Biol*. 2008;18:102–8.
- Schrader M, Fahimi HD. The peroxisome: still a mysterious organelle. *Histochem Cell Biol*. 2008;129:421–40.
- Kaasch AJ, Joiner KA. Targeting and subcellular localization of *Toxoplasma gondii* catalase: Identification of peroxisomes in an apicomplexan parasite. *J Biol Chem*. 2000;275:1112–8.
- Hahn C, Fromm B, Bachmann L. Comparative genomics of flatworms (Platyhelminthes) reveals shared genomic features of ecto- and endoparasitic Neodermata. *Genome Biol Evol*. 2014;6:1105–17.
- Tsai U, Zarowiecki M, Holroyd N, Garcarrubio A, Sanchez-Flores A, Brooks KL, et al. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*. 2013;496:57–63.
- Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, et al. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res*. 2012;40:D284–9.
- Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39(Web Server issue):W29–37.
- Faust JE, Verma A, Peng C, McNew JA. An inventory of peroxisomal proteins and pathways in *Drosophila melanogaster*. *Traffic*. 2012;13:1378–92.
- Gonzalez NH, Felsner G, Schramm FD, Klingl A, Maier U-G, Bolte K. A single peroxisomal targeting signal mediates matrix protein import in diatoms. *PLoS One*. 2011;6:e25316.
- Motley AM, Hetteema EH, Ketting R, Plasterk R, Tabak HF. *Caenorhabditis elegans* has a single pathway to target matrix proteins to peroxisomes. *EMBO Rep*. 2000;1:40–6.
- Saveria T, Halbach A, Erdmann R, Volkmer-Engert R, Landgraf C, Rottensteiner H, et al. Conservation of PEX19-binding motifs required for protein targeting to mammalian peroxisomal and trypanosome glycosomal membranes. *Eukaryot Cell*. 2007;6:1439–49.
- Gorsky G, Palazzoli I, Fenaux R. Influence of temperature changes on oxygen uptake and ammonia and phosphate excretion, in relation to body size and weight, in *Oikopleura dioica* (Appendicularia). *Mar Biol*. 1987;94:191–201.
- Hopcroft RR, Roff JC. Zooplankton growth rates: extraordinary production by the larvacean *Oikopleura dioica* in tropical waters. *J Plankton Res*. 1995;17:205–20.
- Seo H-C, Kube M, Edvardsen RB, Jensen MF, Beck A, Spriet E, et al. Miniature genome in the marine chordate *Oikopleura dioica*. *Science*. 2001;294:2506.

36. Nishida H. Development of the appendicularian *Oikopleura dioica*: Culture, genome, and cell lineages. *Dev Growth Differ.* 2008;50:5239–56.
37. McCollum D, Monosov E, Subramani S. The pas8 mutant of *Pichia pastoris* exhibits the peroxisomal protein import deficiencies of Zellweger syndrome cells—the PAS8 protein binds to the COOH-terminal tripeptide peroxisomal targeting signal, and is a member of the TPR protein family. *J Cell Biol.* 1993;121:761–74.
38. Nakao MC, Nakai K. Improvement of PSORT II protein sorting prediction for mammalian proteins. *Genome Inform.* 2002;13:441–2.
39. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc.* 2007;2:953–71.
40. Bloisi W, Colombo I, Garavaglia B, Giardini R, Finocchiaro G, Didonato S. Purification and properties of carnitine acetyltransferase from human liver. *Eur J Biochem.* 1990;189:539–46.
41. Shimada H, Oginuma M, Hara A, Imamura Y. 9,10-phenanthrenequinone, a component of diesel exhaust particles, inhibits the reduction of 4-benzoylpyridine and all-trans-retinal and mediates superoxide formation through its redox cycling in pig heart. *Chem Res Toxicol.* 2004;17:1145–50.
42. Koonin EV. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* 2010;11:209.
43. Tovar J, León-Avila G, Sánchez LB, Sutak R, Tachezy J, van der Giezen M, et al. Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature.* 2003;426:172–6.
44. Mazzucco A, Benchimol M, De Souza W. Endoplasmic reticulum and Golgi-like elements in *Entamoeba*. *Micron.* 1997;28:241–7.
45. Dacks JB, Davis LAM, Sjögren ÅM, Andersson JO, Roger AJ, Doolittle WF. Evidence for Golgi bodies in proposed “Golgi-lacking” lineages. *Proc R Soc Lond B Biol Sci.* 2003;270 Suppl 2:S168–71.
46. Klute MJ, Melançon P, Dacks JB. Evolution and diversity of the Golgi. *Cold Spring Harb Perspect Biol.* 2011;3:a007849.
47. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402.
48. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
49. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59:307–21.
50. McIntosh MT, Elliott DA, Joiner KA. *Plasmodium falciparum*: Discovery of peroxidase active organelles. *Exp Parasitol.* 2005;111:133–6.
51. De Duve C. The peroxisome: A new cytoplasmic organelle. *Proc R Soc Lond B Biol Sci.* 1969;173:71–83.
52. de Duve C. The origin of eukaryotes: a reappraisal. *Nat Rev Genet.* 2007;8:395–403.
53. Kim PK, Mullen RT, Schumann U, Lippincott-Schwartz J. The origin and maintenance of mammalian peroxisomes involves a de novo PEX16-dependent pathway from the ER. *J Cell Biol.* 2006;173:521–32.
54. Blobel G. Intracellular protein topogenesis. *Proc Natl Acad Sci U S A.* 1980;77:1496–500.
55. Hart DT, Misset O, Edwards SW, Opperdoes FR. A comparison of the glycosomes (microbodies) isolated from *Trypanosoma brucei* bloodstream form and cultured procyclic trypomastigotes. *Mol Biochem Parasitol.* 1984;12:25–35.
56. Denoeud F, Henriot S, Mungpakdee S, Aury J-M, Silva CD, Brinkmann H, et al. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science.* 2010;330:1381–5.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



Mastigamoeba balamuthi* genome and the nature of the free living ancestor of *Entamoeba

Vojtěch Žárský¹, Vladimír Klimeš², Jan Pačes³, Čestmír Vlček³, Eva Nývltová¹, Ivan Hrdý¹,
Lael Barlow⁴, Marek Eliáš², Joel Dacks⁴, Neil Hall⁶, Andrew Roger⁵, Jan Tachezy^{1*}

¹Department of Parasitology, Faculty of Science, Charles University, BIOCEV, Průmyslová 595, 25242 Vestec, Czech Republic

²Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

³Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague 14220, Czech Republic

⁴Department of Cell Biology, University of Alberta, Edmonton, AB T6G 2H7, Canada

⁵Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS B3H 4R2, Canada

⁶School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ United Kingdom

***Corresponding author:** Jan Tachezy; e-mail tachezy@natur.cuni.cz, Department of Parasitology, Faculty of Science, Charles University, BIOCEV, Průmyslová 595, 25242 Vestec, Czech Republic, Phone +420 325 874 144, Fax: +420 224 919 704

Key words: evolution of parasitism, lateral gene transfer, pathways complexity, Archamoeba, Entamoeba, Mastigamoeba, chitinous cysts

Summary

Conosa is a group of protist that includes aerobic free living Eumycetozoa, and anaerobic free-living and parasitic members of Archamoebae. Phylogenetic analyses have robustly supported that ancestor of Conosa was an aerobic flagellate, from which the anaerobic lineage evolved. The anaerobes include *Mastigamoeba balamuthi* that inhabits low-oxygen organic-rich water sediments and *Entamoeba histolytica*, an important intestinal pathogen of humans. Here, we sequence genome and transcriptome of *M. balamuthi* and performed phylogenomic analysis of eight Conosa genomes to trace gene histories during adaptation to anaerobic environments, and during transition from free-living to parasitic lifestyle. These events were both associated with massive gene losses that in parasitic lineage resulted in reduction of structural features, in complete loss of some metabolic pathways and generally in reduction of metabolic complexity. The reconstruction of the character of the common ancestor of Archamoebae allows to estimate ancestral preconditions for the parasite evolution that include an ability to form chitinous resistant cysts, proteolytic enzyme equipment, compartmentalization of sulfate activation pathway, and anaerobic energy metabolism. After split of Entamoebidae, the parasite specific features were associated mainly with gain of genes for surface membrane proteins that are involved in host-parasite interactions. In contrast, gene gains identified in *M. balamuthi* were associated dominantly with polysaccharide catabolic processes, specifically with cell wall degradation. The phylogenetic analysis of acquired genes suggested an essential role of lateral gene transfer in the parasite evolution (*Entamoeba*) as well as in the adaptation to anaerobic water sediments (*Mastigamoeba*).

Introduction

Parasitic organisms evolved multiple times in all major eukaryotic lineages (Blaxter & Koutsovoulos, 2015), however evolutionary steps leading to their transition from presumably free-living ancestors remains elusive. The parasites are organisms that employed various strategies with the common aim to get nutritional benefits for growth and reproduction at the expense of the host. These strategies are reflected by specific features of parasites that resulted mainly from the character of free-living ancestor and from character of the host niche, to which evolving parasite needed to adapt. The best known parasitic features are those involved in the parasite pathogenicity and virulence. Because these features are usually absent in the host cells, they are seemingly unique to the parasite. However, some of these features could be lineage-specific and functional before origin of parasitic members and just absent in the host lineage (Janouskovec & Keeling, 2016)(Ginger & Field, 2016). For example, compartmentalization of glycolysis to glycosomes discovered in *Trypanosoma brucei* was initially thought to reflect adaptation to parasitism in the glucose-rich host bloodstream (Opperdoes & Borst, 1977), however soon after, glycosomes appeared to be common to kinetoplastids including free-living species (Jackson et al., 2016)(Gualdrón-López et al., 2012). In other words, some properties of parasites represent rather precondition to parasitism than a result of transition from free-living-to-parasitic life strategy. Once the organism becomes an obligate parasite that lives in relatively narrow and nutrient-rich environment, various functions and consequently genes became dispensable. This tendency is particularly apparent in intracellular parasites such as microsporidia and apicomplexans, in which reduction of metabolic pathways evolved together with dramatic reduction of the genome (Heinz et al., 2012)(Mathur et al., 2019)(Janouškovec et al., 2015). However, other parasites such as parasitic kinetoplastids and ciliates seem to keep most of physiological functions and underwent more a reduction of diversity within the genes families and consequently functional redundancy than complete

loss of particular physiological function (Jackson et al., 2016)(Coyne et al., 2011). In addition to gene losses, parasite genomes are shaped by parasite innovations to gain truly new parasite-specific functions. These innovations include expansions of certain gene families and paralog speciation, appearance of novel genes of unclear origin and acquisition of genes by lateral gene transfer (LGT) (Jackson et al., 2016)(Loftus et al., 2005)(Carlton et al., 2007)(Nývltová et al., 2015).

Investigations of gene histories to distinguish precondition and unique parasitic features are essential for understanding of parasite evolution. Such studies are dependent on availability of genomes from parasitic species together with close free-living relatives. However, although genomes of most important parasites have been sequenced, currently there are only few cases of free-living neighbors which could be used for comparative genomics. These include for example free living kinetoplastids related to *Trypanosomatids* (Jackson et al., 2016), chrompodellids that are neighbors of obligate parasite of Apicomplexa group (Mathur et al., 2019), and free-living ciliates related to pathogenic *Ichthyophthirius multifiliis* (Coyne et al., 2011). Of note, all these studies were focused on organisms that occupied aerobic niches. However, several important human parasites such as *Entamoeba histolytica* and *Trichomonas vaginalis* are anaerobes living under oxygen limited conditions of human gut and urogenital tract, respectively. The mode of evolutionary processes that led to appearance of anaerobic parasite is puzzling and there are not known primarily free-living relatives of *T. vaginalis* for comparative studies. In contrast, *Entamoebae* belong to *Archamoeba* group of anaerobic protists that includes free-living species such as *Mastigamoeba balamuthi* (Baptiste et al., 2002). Moreover, phylogenetic analyses revealed that Archamoebae clustered with aerobic Eumycetozoa in Conosa group and that the last common Conosa ancestor (LCCA) was probably aerobic free-living protists (Baptiste et al., 2002)(Pánek et al., 2016). Therefore, Conosa group provides unique opportunity to trace history of parasitism from aerobic LCCA, via common anaerobic ancestor of Archamoebae

(LCAA) to parasitic *Entamoeba*. In this sense, *M. balamuthi* represents a key species for predictions of LCAA features.

M. balamuthi was isolated from public well in Gambia (Chavez et al., 1986). It forms multinucleated amoeboid stages, uninuclear flagellates with a single flagellum, and cysts. Adaptation to anaerobiosis is reflected by transformation of mitochondria to hydrogenosomes that contain typical enzymes of anaerobic energy metabolism (Nývltová et al., 2015), and anaerobic type of peroxisomes (Le et al., 2020). There are several unusual features that *M. balamuthi* shares with *E. histolytica* and that were most likely gained by LCAA. These include loss of mitochondrial iron-sulfur cluster (ISC) assembly which was replaced by ϵ -proteobacterial nitrogen fixation system (NIF)(Nylvtova et al., 2013) and acquisition of components of a sulfate activation pathway that operate in *M. balamuthi* hydrogenosomes and *E. histolytica* mitosomes, the most reduced form of mitochondria (Nývltová et al., 2015)(F Mi-ichi et al., 2009).

Here we sequenced and analyzed a genome of *M. balamuthi* with aim to characterize the changes associated with adaptations from aerobic to anaerobic niches, to estimate common features of *M. balamuthi* and *Entamoeba* ancestor that may facilitate evolution of parasitism in *Entamoeba* lineage as well as a unique innovation that appeared separately in free-living and parasitic Archamoebae to adapt to fundamentally two different anaerobic environments.

Results and Discussion

Genome of *Mastigamoeba balamuthi*

Genome of axenically grown *M. balamuthi* was sequenced using 454 GS FLX Titanium system (Roche) and NexSeq (Illumina) and assembled into a 57.27 megabase pair genome over 1926 scaffolds

(N50=442.5 kbp)(Table S1). RNA-seq was used for sequencing of *M. balamuthi* transcriptome, which aided prediction of 16287 protein coding genes. Completeness of the genome was estimated using the Benchmarking Universal Single-Copy Orthologs (BUSCO) assessment that identified 82.8% conserved eukaryotic single-copy orthologs in *M. balamuthi* genome. Furthermore 98.6% of the transcriptome sequences were mapped to the genome using GMAP (Wu & Watanabe, 2005).

Characteristics of *M. balamuthi* genome are remarkably different in comparison to *Entamoeba* species (Fig. 1, Table S1). *M. balamuthi* genome is larger and codes for 30-96% more genes than *Entamoeba* species (Fig. 1, Table S2). The protein coding genes, that constitute 41.2% of *M. balamuthi* genome, are 6-41% longer than *Entamoeba* species. Interestingly intergenic regions that span 35.2% of *M. balamuthi* genome are of 8% and 66% longer in comparison to *E. histolytica* and *E. moshkovskii*, respectively, but 62% shorter when compared with *E. invadens*. *M. balamuthi* proteins coding genes are more rich for introns (3.38 introns/gene) with average intron size 137 bp (Table S1). The *M. balamuthi* genome is rich in GC content (60.7%) that is on the opposite end of the scale compared to low 24.3% GC content of *E. histolytica* genome (Fig. 1). Because there are no genomic sequences of *Archamoebae* available outside of *Entamoeba* lineage, we selected actin genes to compare GC content within this group. We can observe that obligate free-living organisms incline to higher GC content while the groups with both endobiotic representatives (*Rhizomastix*) and predominantly parasitic organisms (*Entamoeba*) tend to have progressively lower GC content. Outside of *Archamoebae*, low GC content comparable to *Entamoeba* species displayed only the slime mold *Dictyostelium discoideum* (Fig 1).

All expected structural RNA genes required for splicing (spliceosomal RNAs), tRNA maturation (Ribonuclease P), translation (ribosomal RNAs, transfer RNAs) and ER translocation (signal recognition particle (SRP) RNA) were detected in support of the genome completeness (Table S3 RNA). Unusual property of *Entamoeba* genome is very high number of tRNA genes (4500) that

form unique linear arrays (C. Graham Clark et al., 2006). Therefore we were interested in distribution of tRNA in *M. balamuthi*. We found considerably lower number of tRNA genes (219) that did not form such arrays and were randomly scattered within the genome (SI TableS4-tRNA). Interestingly we found a missing homolog of SRP RNA in *Entamoeba* that is coded as a part of threonine-tRNA coding array. The proximity of tRNA genes and SRP RNA may have functional role as they are transcribed by the same RNA polymerase III.

Analysis of trasposable elements (TEs) revealed that their overall contribution to the *M. balamuthi* genome size is about 1.7% (Fig. 2, Fig. S1, Table S5). The TEs localize predominantly to intergenic regions and scaffold ends (56.5% and 28.0% respectively), about 12.5% within coding sequences, and 3.0% within introns. TEs include 421 complete or degenerated retrotransposon (RT) and similar number (464) of DNA transposons. This is interesting in comparison with *Entamoeba* species that contain predominantly either RTs (*E. histolytica* and *E. dispar*) or DNA transposons (*E. invadens* and *E. moshkovskii*)(Fig. 2, (Pritham et al., 2005). Most of *M. balamuthi* RTs (268) belong to the order of Long terminal repeat (LTR) RTs that is absent in *Entamoeba* species (Pritham et al., 2005)(Table S5). Non-LTR RT in *M. balamuthi* includes three superfamilies (L1, R2 and I) of the order of long interspersed nuclear elements (LINE). DNA transposons in *M. balamuthi* include Tc1-*Mariner* that is dominant in *E. invadens* and *E. moshkovskii*, *Pif-Harbinger* that is absent in Entamoebidae and a number of unclassified TEs. *M. balamuthi* lacks DNA transposons of family *Mutator* that is predominant in *E. invades/E. moshkovskii*. LINE R2 appeared as the only common TEs to all compared species, which suggests that related TEs might be present in LCAA. However phylogenetic analysis of LINEs members did not support monophyletic origin of LINE R2 TEs in *M. balamuthi* and Entamoebidae (Fig. S2_phylogeny). The observed different assortment of TEs in *M. balamuthi* and *Entamoeba* species seems to be more consistent with a lateral transmission of TEs within specific

ecosystems (Gilbert et al., 2010) than a vertical transmission from a common ancestor (Pritham et al., 2005).

Gene histories and character of LCAA

To trace gene losses and gains that led to diversification of free living *M. balamuthi* and parasitic species of *Entamoeba* group we reconstructed gene histories within Archamoebae and their relatives. First we collected conserved orthologous groups (OGs) in predicted proteomes of 8 amoebozoans (*M. balamuthi*, *Entamoeba histolytica*, *E. moshkovskii*, *E. invadens*, *Dictyostelium discoideum*, *Acytostelium subglobosum*, *Planoprotostelium fungivorum* and *Acanthamoeba castellanii*) and 4 opisthokonts (*Homo sapiens*, *Amphimedon queenslandica*, *Saccharomyces cerevisiae* and *Batrachochytrium dendrobatidis*) by searching the EggNOG database of OGs using a profile hidden Markov mode (HMM) based search. Based on the evolutionary relationships between these species and with assumption of a vertical inheritance of OGs, we then reconstructed evolutionary history of each of the OGs: (i) We predicted that given OG was gained in the last common ancestor of species in which this OG was detected. (ii) We predicted losses of the OG in descending lineages in which given OG was not detected (Fig. 3, Table S6).

In Archamoebae lineage, the substantial loss of orthologous groups (1532) was associated with the first major evolutionary transition: the adoption the anaerobic/microaerophilic life style. This is supported by statistical evaluation of OG terms using Ontologizer tool (Bauer et al., 2008) that revealed losses of OG terms for “Mitochondrion” with the highest support $5.1e-34$, following by term “Oxidoreductase activity” ($5.1e-17$). We can expect that mitochondrion of LCAA was similar to hydrogenosome of *M. balamuthi* which lacks most of oxygen dependent pathways and mitochondrial components namely the oxidative phosphorylation (OXPHOS), and the respiratory chain complexes except the complex II (succinate dehydrogenase, SDH). Pyruvate dehydrogenase complex (PDC), and

TCA cycle were substituted by a simple anaerobic energy metabolism based on substrate level phosphorylation (Nývltová et al., 2015)(Gill et al., 2007).

Furthermore, we can predict that evolving anaerobic mitochondria lost cristae, presence of which correlates with OXPHOS (Davies et al., 2011). This is supported by lack of all components of mitochondrial contact site and cristae organizing system (MICOS) in both *M. balamuthi* and entamoebids. A concomitant event to OXPHOS loss is a loss of the redox potential at the inner mitochondrial membrane that is crucial for import of proteins into aerobic mitochondria. This is reflected by reduction and modifications of protein import machinery (Pyrihová et al., 2018)(Makki et al., 2019). In *E. histolytica*, only Tom40 component of outer mitochondrial translocase (TOM) is known together with unique subunit Tom60 (Makiuchi et al., 2013), and Sam50 (Dolezal et al., 2010), whereas, no component of inner membrane translocase (TIM) has been found so far. We expected that *M. balamuthi* hydrogenosome may retain more standard translocation machinery as hydrogenosome of *T. vaginalis*, however, only Tom40, and Sam50 were identified. We did not detect any components of TIM that are likely too divergent to be recognized by current bioinformatic tools. However, we identified a standard mitochondrial protein processing peptidase (MPP) that is required for maturation of proteins imported into mitochondrial matrix, and that is absent in *E. histolytica*.

Changes in physiological functions upon adaptation to anaerobiosis are further documented by correlation between a number of lost OGs and corresponding COG categories (Fig. 3). In addition to reduction in pathways of energy metabolism, this analysis also pointed out on strong reduction in COG category "lipid transport and metabolism" (Fig. 3) that include both biosynthetic and catabolic pathways. The data suggest a complete loss of fatty acid (FA) β -oxidation including acyl-CoA dehydrogenase, which in mitochondria is coupled with respiratory chain to maintain a redox balance. Alternatively, FA β -oxidation can be localized in peroxisomes where the FA degradation is initiated by oxygen-dependent acyl-CoA oxidase. However, all oxygen dependent enzymes in peroxisomes adapted

to anaerobiosis including catalase and acyl-CoA oxidase were possibly lost as observed in anaerobic peroxisomes of *M. balamuthi* (Le et al., 2020). Presence of FA biosynthesis in LCAA is not supported by available data. Neither *E. histolytica*, nor *M. balamuthi* possess complete FA synthesis pathways. Interestingly, *E. histolytica* encodes an unusual acetyl-CoA carboxylase (ACC) that includes 2 bacterial-type carboxylase domains, an acetyl-CoA carboxylase and a pyruvate carboxylase (C. G. Clark et al., 2007). This protein has been implicated in formation of malonyl-CoA, a starting substrate for fatty acid synthesis. We identified an ortholog of ACC in *M. balamuthi*, however its function needs to be tested. In addition, *M. balamuthi* possesses a single component of type II FAS: 3-oxoacyl-[acyl-carrier protein] reductase. Neither *E. histolytica*, nor *M. balamuthi* possess fatty acid synthase of type I FAS that is present in *D. discoideum*.

Adaptation to anaerobiosis further resulted in a loss of haem synthesis that in mitochondria requires oxygen in two biosynthetic steps, and loss of mitochondrial FeS cluster assembly (ISC), which was replaced by two copies of NIF system that operate in both mitochondria and the cytosol (Nyvltova et al., 2013). Interestingly, both *M. balamuthi* and entamoebids possess a standard pathway mediating cytosolic FeS cluster assembly (CIA)(Table S7). In organisms with ISC machinery, CIA pathway is dependent on an export of FeS intermediate that is formed in mitochondria (Pandey et al., 2019). Because ISC was most likely absent in LCAA with dual localization of NIF, we can expect that NIF machinery interacted directly with CIA scaffold proteins such as Nbp35. In support of this hypothesis, we found that in *M. balamuthi*, Nbp35 have dual localization as NIF. We identified three Nbp35 paralogs, of which one possess hydrogenosomal targeting sequence and its presence in hydrogenosomes was supported by mass spectrometry (Le et al., 2020).

The most prominent change in signaling pathways is loss of eukaryotic histidine kinases. No member of four eukaryotic families of histidine kinases were found in either *M. balamuthi* or *E.*

histolytica although they were identified in other amoebae such as *Discosea* and *Eumycetozoa* and thus likely presented in their common ancestor (Kabbara et al., 2019).

Adaptation of *M. balamuthi* to free-living lifestyle

M. balamuthi is living in low-oxygen organic-rich water sediments. Plant debris that is freely present in such a niche is known to be decomposed mainly by saprophytic fungi and bacteria (Artzi, Bayer, & Moraïs, 2017)(Cheng et al., 2018). Analysis of gene histories in *M. balamuthi* inferred a presence of a large set of genes that were gained to optimize *M. balamuthi* metabolism within this specific niche (Table S8). OG term analysis revealed the highest statistical support for gains of genes involved in “Carbohydrate metabolic processes” ($1e-12$)(Fig 3). Specifically, *M. balamuthi* gained proteins for plant cell wall degradation including enzyme for degradation of cellulose, pectins as well as pentose-based hemicellulose (xylans and arabinans) (Fig. S3). Cellulolytic enzymes include endoglucanase, exoglucanase and cellobiase that convert cellulose to glucose. Pentoses are catabolised based on bacterial isomerisation reaction catalyzed by xylose or arabinose isomerases. A unique enzyme is bifunctional xylulose 5-phosphate/fructose 6-phosphate phosphoketolase (Xfp). This enzyme converts D-xylulose 5-phosphate to D-glyceraldehyde and acetyl phosphate (acetyl-P) and thus links xylose metabolism with glycolysis. Alternatively, it metabolises fructose 6-phosphate to D-erythrose 4-phosphate and acetyl-P. Acetyl-P is a substrate for acetate kinase (Ack) that produced acetate and pyrophosphate. This so called pentose phosphoketolase pathway was known to be present only in some bacteria and fungi (Glenn et al., 2014). Interestingly, Ack was characterized in *E. histolytica* (Reeves & Guthrie, 1975), however enzymes synthesizing acetyl-P such as Xfp are absent and physiological role of Ack in *E. histolytica* is unknown (Pineda et al., 2016). In addition, *M. balamuthi* gained genes to defense against microbial flora in its vicinity. It has been shown that *M. balamuthi* acquired enzymes of

Clostridiales origin that allows production of bactericidal *p*-cresol to suppress growth of putative competitors (Nývltová et al., 2017).

When considering the largest OGs of *M. balamuthi* and *E. histolytica* we can see an expansion of proteins involved in environmental sensing, interactions and signalling (Fig. 4_Ten largest EggNOG A B). In both organisms have expanded tyrosine kinases, small GTPases, their exchange factors and cathepsins. *M. balamuthi* has a unique expansion of guanylate cyclases and cGMP-dependent phosphodiesterases pointing to an intricate cGMP-dependent signaling. Interestingly, although *M. balamuthi* lacks eukaryotic histidine kinases, we identified three histidine kinases with a generic structure Per-ARNT-Sim (PAS) domains, and a histidine kinase domain. Phylogenetic analysis revealed that these histidine kinases are of bacterial origin (Fig. S4).

Genome streamlining of *Entamoebae*

The transition to endobiosis/parasitism in the *Entamoeba* lineage was the second major evolutionary step that was accompanied with massive loss of OGs (2552). These losses correspond either to complete absence of certain physiological or structural features, or to reduced complexity of mainly metabolic pathways. Structurally, the most prominent change is the loss of flagellum and associated microtubular cone that subtends the nucleus and defines Conosa group (Cavalier-Smith, 1998). This is supported by significantly reduced OG term for “Cillium” ($p=1e-13$).

Loss of flagellum: Presence of flagellated cells is common among Amoebozoa including *M. balamuthi*, however currently there is not any survey of amoebozoan flagellum and associated components due to the bias of sequenced genomes towards aflagellated dictyostelids and *Entamoebae*. Our search for flagellar components in *M. balamuthi* genome revealed a complex set of proteins associated with centriole, basal body including BBSome of Bardet-Biedl Syndrome proteins (Nachury et al., 2007), and axonemes (Fig. S5. Flagellum, Table S9). Interestingly, *M. balamuthi* possesses

almost complete set of BBSome proteins including an ortholog of BBS6 that seemed to be present until now only in metazoans and it is absent in flagellated fungi. We did not detect any outer dynein arm components, which supports an ultrastructural observations (Walker et al., 2001). In *Entamoeba* lineage, all components associated with biogenesis and function of the flagellum were completely lost (Fig. S5). The loss of flagella is most likely associated with the character of environments. *M. balamuthi* is living without flagella in form of trophozoite however, it can form flagellated non-dividing organisms most likely upon environmental changes as *Naegleria* species to move to a new habitat. However, intestinal Archamoebae are living in narrow defined type of environment from which they can escape only in the form of cysts. Thus, their flagellated forms become dispensable.

The nuclear pore complex: M. balamuthi trophozoites in axenic culture possess one to tens pyriform nuclei with typical pores arranged in rows between cone microtubules (Chavez et al., 1986)(Walker et al., 2001). The nuclear pore complex (NPC) consists of approximately 30 nucleoporins (Nups) that form cytoplasmic fibrils, inner pore ring including scaffold subunits and phenylalanin-glycin repeat-nucleoporins (FG-NUPs) of the central channel, nuclear rings and the peripheral basket (Beck & Hurt, 2017). Searches for Nups in *M. balamuthi* detected almost the complete set of Nups (23 proteins) (Fig. 5)(Table S10). They include transmembrane protein Ndc1, scaffold subunits Nup205 and Nup188, and the basket subunit Nup153 that were previously suggested to be lost in Amoebozoa (Neumann et al., 2010). Finding of Ndc1 in Amoebozoa supports the previous prediction that Ndc1 together with Gp210 and Nup35 represent an ancestral system anchoring central core within the nuclear membrane (Neumann et al., 2010). In contrast, most Nups seems to be lost in *E. histolytica*, in which only 5 Nups were found so far. Particularly, no scaffold protein was identified and only one candidate of three FG-NUPs of the central channel was predicted (Neumann et al., 2010). This is surprising as the nuclear pore is one of the most conserved and defining features of eukaryotic cell, although the primary protein structure could be rather variable. Therefore we used profile-HMMs build on *M. balamuthi* Nups to

search *E. histolytica* sequences and discovered six more Nups including three scaffold proteins components of the inner pore ring and a complete set of FG-NUPs in *E. histolytica* (Fig. 5.).

Nevertheless, most components of cytoplasmic ring, and nuclear ring and all basket components seem to be absent. Reduction of Nups is often observed in parasitic species e.g. microsporidians *Encephalotozoon cuniculi*, excavates *G. intestinalis* and *T. vaginalis*, and apicomplexan *Theileria parva* (Neumann et al., 2010)(Mans et al., 2004). However, what is a causal relationship, if any, between NPC and adaptation to parasitism remains to be elucidated. We also cannot rule out that Nups in these parasitic species are too divergent to be detected.

Membrane trafficking complexes: Comparison of membrane trafficking complexes revealed that *M. balamuthi* retains most components found in free-living *D. discoideum* or *A. castellanii*, whereas *E. histolytica* lost all components of the HOPS and CORVET complexes that are involved in tethering of endosomes and lysosomes, and substantially reduced components of TRANsport Protein Particle I complex (TRAPPI) and oligomeric Golgi (COG) complex. We were able to identify only a single case where *M. balamuthi* share a partial loss of a complex with *Entamoeba*. The TSET complex that was recently described to be involved in endocytosis from the cell surface is complete in *D. discoideum*, partial in *M. balamuthi* and absent entirely from *E. histolytica*. In contrast, there are more gene copies for COPI and COPII components in *E. histolytica* (Fig S6 Joel Dacks). The examples of loss in *Entamoeba* seem to reflect a remodeling of the endocytic machinery after the split with *Mastigamoeba* and at the same time, *Entamoeba* has an expanded retromer complement (retrograde transport from endosomes to TGN), has all 5 adaptins, including the ones that act in the endosomal system in mammals, and uses late endosomal Rabs extensively.

To compare metabolic pathways complexity, we mapped enzyme coding genes of *D. discoideum*, *M. balamuthi* and *E. histolytica* to Kyoto Encyclopedia of Genes and Genome (KEEG) pathways. We were able to map 802 unique enzyme coded in *D. discoideum* genome, 677 enzymes of *M. balamuthi*

and only 301 enzymes of *E. histolytica* (Fig. 6). The reduction of unique enzyme spectrum in *E. histolytica* is distributed across majority of tested pathways. The most striking losses are apparent in nucleotides metabolism. Whereas, and *M. balamuthi* possesses *de-novo* purine and pyrimidine synthesis (Fig. S7B and C, KEGG pathways) these synthetic pathways are absent in *E. histolytica* and the parasite entirely relies on the nucleotide salvage pathways (C. G. Clark et al., 2007). The reduction of purine and pyrimidine metabolism most likely reflects adaptation of *E. histolytica* and other parasite such as *G. intestinalis* and *T. vaginalis* to nutrient rich environment of the host. *M. balamuthi* also possesses ribonucleotide reductase that was lost in *E. histolytica* but it was retained in *E. invadens* and *E. moshkovskii* (Loftus et al., 2005). Furthermore, *E. histolytica* substantially reduced or lost enzymes of amino acids synthesis (Fig. S7D and E, KEEG pathways)(C. G. Clark et al., 2007).

Preconditions to parasitism

Comparison of free living *M. balamuthi* and its parasitic relatives provides unique opportunity to estimate features that were present in LCAA and may represent preconditions to parasitism and those that were specifically gained in the pathogenic lineage as well as specific properties associated with free-living lifestyle of *M. balamuthi* (Fig. 7).

Key preconditions for Archamoebae parasitism include an ability of LCAA to live under anaerobic conditions, an ability to form resistant cysts, proteolytic enzyme equipment, and compartmentalization of sulfate activation pathway. Adaptation to anaerobiosis is certainly advantageous for colonizing of low-oxygen niches such as human gut. Nevertheless, efficient antioxidant system is still needed during periods of aerobic stress caused by the host immune response and during the invasion of oxygenated tissues. Both *M. balamuthi* and *E. histolytica* share antioxidant system characteristic for anaerobic and microaerophilic eukaryotic protists (Table S11 Virulence). Their antioxidant machinery rely on the thioredoxin-based system (thioredoxin, NADPH:flavin oxidoreductase, peroxiredoxin), Fe-superoxid

dismutase, rubrerythrin and flavodiiron proteins, the latter two being good candidates for LGT from anaerobic bacteria. However both *M. balamuthi* and *E. histolytica* and possibly LCAA lack glutathione-based pathways, catalases and peroxidases.

Another crucial precondition for parasitism is an ability to form resistant chitinous cysts. *M. balamuthi* forms cysts that allows to survive adverse conditions e.g. oxygen exposure or desiccation stress. The cysts are Periodic Acid-Schiff (PAS) positive (Chavez et al., 1986) and labeled by calcofluor (Fig. 8), which is used to label polysaccharides on chitin. This is reminiscent of the *Entamoebae* chitinous cysts that have been studied in *E. histolytica* and *E. invadens* (Chatterjee et al., 2009). The *Entamoebae* cyst wall is known to composed, in addition to chitin, of a unique set of lectins including chitinase, Jacob and Jessie with cysteine-rich chitin-binding domains (CBDs). In *M. balamuthi*, we found a complete pathway for chitin synthesis and degradation, as well as Jessie and Jacob lectins, which were previously thought to be unique to *Entamoebae* (Van Dellen et al., 2002) (Fig. S8). We identified a single *M. balamuthi* chitin synthase (chs) with a C-terminal transmembrane domain (TMD) that is monophyletic with membrane-anchored chs-1 of *E. invadens* (Fig. S9 Chitin synthase). There are five *M. balamuthi* chitinases of which three possesses N-terminal CBD, whereas two more chitinases have no CBD but they possess a TMD domain at the C-terminus (Fig. S10 chitinase tree). A common feature of Jessie lectins is presence of N-terminal CBD with 8 conserved cysteines. *Entamoeba* Jessie-1 and Jessie-2 are short Jessie variants, whereas Jessie-3 contains except CBD an additional C-terminal specific domain. We found three orthologs of Jessie-3 in *M. balamuthi* based on phylogeny of CBD domains (Fig. S11A): a short variant *MbJassie-A*, a long variant with C-terminal domain of unknown function (*MbJassie-B*), and another long variant (*MbJassie-C*) with the Jessie-3 specific C-terminal domain (Fig. S11A). Homology searches revealed that this domain corresponds to a catalytic domain of a prokaryotic chitinase, whose structure was recently determined (Fig. S11B)(5XSV_A) (Nishitani et al., 2018). The four proposed catalytic residues (Glu532, Asp566, Glu572, His693) are

conserved in all Jessie-3 proteins and *MbJassie-C* (Fig. S11C). *Entamoeba* Jacob lectins have a variable number (2-6) of CBDs (Frisardi et al., 2000). *M. balamuthi* possesses a single Jacob ortholog with two CBDs and C-terminal TMD (Fig S12). Although *Entamoeba* and *M. balamuthi* possess all required lectins, the detail cyst wall structure might be different. It has been proposed that during initial stage of encystation, Jacob binds to the surface of encysting *Entamoebae* by Gal/GalNAc lectin, then chitin is synthesized and finally the wall is solidified by the addition of Jessie (Chatterjee et al., 2009). However, Gal/GalNAc lectin is absent in *M. balamuthi*. Noteworthy, unlike *Entamoebae* Jessie and Jacob proteins, *MbJassie-C* and Jacob possess C-terminal TMDs that might function as a direct anchor to the plasma membrane instead of Gal/GalNAc lectin and the chitin cyst wall is possibly degraded during excystation by the chitinase activity (Fig. 8) (Chavez et al., 1986). The use of indirect attachment of *Entamoeba* chitin-binding moieties through Gal/GalNAc lectin may be a useful innovation that could allow to hatch fast through a perforation of the cyst wall (Makioka et al., 2005). Altogether, these findings suggest that the free living LCAA was most likely able to form the chitin-based cyst and that this ability may represents a well prepared aid for evolving parasitic *Entamoebids* to survive the passage through the host stomach and in the outer environment.

LCAA most likely possessed sulfate activation pathway in anaerobic type of mitochondria (Nývltová et al., 2015)(F Mi-ichi et al., 2009). This pathway produced 3'-phosphoadenosine 5'-phosphosulfate (PAPS) that is exported to the cytosol via mitochondrial carrier family protein in *E. histolytica* (EhMCP)(Mi-Ichi et al., 2015) and most likely in *M. balamuthi* that possesses EhMCP ortholog (Fig. S13). In *E. histolytica*, PAPS is then utilized by sulfotransferases (SULTs) to form sulfolipids (SLs) including cholesteryl sulphate that is a substrate for SULT6 and is involved in regulation of cyst formation (Fumika Mi-ichi et al., 2015). Catabolic pathways of sulfated molecules then includes sulfatases (SFs)(Fumika Mi-ichi et al., 2017). *M. balamuthi* was shown to have no capacity to synthesize sulfolipids and absence of SULTs and SFs has been suggested (Fuika Mi-ichi et

al., 2015). We identified four genes for SULTs and two genes for SFs in *M. balamuthi*. Phylogenetic analysis suggested that SULTs of *E. histolytica* and *M. balamuthi* are related and they were possibly acquired via LGT by LCAA (Fig. S14). However, origin of SFs is different. *M. balamuthi* SFs are related to eukaryotic enzymes (Fig. S15), whereas *E. histolytica* counterparts are related to bacterial Zn-dependent alkylSFs from *Bacteroides* (Fumika Mi-ichi et al., 2017). It is plausible that after split of *Entamoeba*, SULTs underwent gene duplication, paralogous SULTs gained various substrate specifications associated with adaptation to parasitism (Fumika Mi-ichi et al., 2017) and consequently acquired different catabolic SFs to metabolize *Entamoebae* specific sulfated molecules, whereas *M. balamuthi* retained eukaryotic SFs.

Cysteine proteases (CPs) are important virulent factors of *E. histolytica* (Lidell et al., 2006) (Hellberg et al., 2001). The most important are CPs of C1 papain superfamily that consists of 36 members categorized into three clades (EhCP-A, EhCP-B and EhCP-C) (C. G. Clark et al., 2007) (Bruchhaus et al., 2003). We found 54 C1 papain CPs in *E. histolytica* and considerably larger repertoire in *M. balamuthi* (156 CPs) (Table S11 Virulence). In phylogenetic analyses *M. balamuthi* CPs clustered to ten groups that include EhCPs (Fig. S16 tree). There two *M. balamuthi* CPs closely related with EhCP-B and a single CP with EhCP-C. Importantly, although overall CP repertoire is reduced in *E. histolytica*, a number of paralogs in each *E. histolytica* clade is expanded. For example, there is 13 *E. histolytica* EhCP-B paralogs and 18 EhCP-C paralogs (Fig. S16 tree). In the contrary, some of *M. balamuthi* clades that have no orthologs in *E. histolytica* are highly expanded such as the 84 paralogs of CP m51a1_g146. Therefore, we can predict that LCAA was equipped with broad spectrum of CPs, from which *E. histolytica* CPs associated with virulence evolved.

E. histolytica amoebapore is a pore-forming protein of saposin-like protein (SAPLIP) family that is involved in the degradation of bacteria (Andrä et al., 2003) and possibly host cells (Bujanover et al., 2003)(Leippe et al., 1991). We found a homolog of amoebapore in *M. balamuthi* genome, which

resembles the animal prosaposin gene, composed of three saposin domains (Table S11 Virulence). Thus, members of SAPLIP family were likely present in LCAA.

Features that could be considered as specific adaptation to parasitism and that are absent in free-living *M. balamuthi* include unique membrane proteins that are involved in parasite-host cell interactions. These proteins includes Gal/GalNAc lectin (Petri et al., 2002), highly expanded family of BspA adhesine (111 paralogs in *E. histolytica*, 1320 in *E. moshkovskii*), Ariel1 surface antigen family protein (Willhoeft et al., 1999), a lysine- and glutamic acid-rich protein (KERP1)(Seigneur et al., 2005), and *E. histolytica* serine-, threonine and isoleucine-rich proteins (EhSTIRP) involved in adhesion and cytotoxicity (MacFarlane & Singh, 2007). *M. balamuthi* also lack small GTPases AIG1 (Nakada-Tsukui et al., 2018) that is involved in *E. histolytica* virulence via regulation of the host cell adhesion. These features were most likely absent in LCAA.

The role of LGT

The comparison of OGs within Conosa revealed that *Archamoebae* and *M. balamuthi* in particular gained a number of genes that provided specific functions to these organisms. Many of these genes were likely transferred from prokaryotic lineages via LGT (Gill et al., 2007)(Nyvltova et al., 2013) (Nývltová et al., 2015). To asses overall contribution of LGT on *M. balamuthi* gene repertoire we conducted a large scale analysis of gene phylogenies. Based on homology searches against the Uniprot database we selected conserved domains of *M. balamuthi* proteins and build phylogenies for each of them. For the LGT evaluation we introduced an LGT score as a tool to estimate cases when following two conditions are met: (i) the gene transfer is directed from prokaryotes to eukaryotes, and (ii) a gene tree topology is more plausible to be interpreted as a result of LGT rather than independent losses of an ancestral eukaryotic gene. Altogether we calculated LGT score for 5254 genes, eliminating redundant

paralogs. Out of these, 528 genes have a strong support for an LGT (LGT score > 0.75) (Fig. 9A LGT) (Table S12).

To evaluate for which functions genes more often are transferred from prokaryota, we correlated LGT score with functional annotation of *M. balamuthi* proteins (Fig. 9D). The most significant positive correlation between LGT score and COG enrichment was observed for enzymes and small molecule transporters, namely for genes functioning in amino acid metabolisms, carbohydrate transport and metabolism, and energy conversion (pearson correlation coefficient 0.31, 0.25, 0.23 respectively). Conversely, the genes involved in genetic information processing, signal transduction and intracellular trafficking correlated negatively with the LGT score except COG “cell wall/membrane/envelope biogenesis” that include genes for cyst wall components (Fig. 9D).

To estimate possible prokaryotic donor for the best 528 LGT candidates, we calculated domain boundary frequency (DBF) value for each taxon within all 528 phylogenetic trees. DBF aims to evaluate prokaryotic taxons that are in the closest proximity to the boundary between the eukaryotes and prokaryotes (see Material and Methods for detail, Table S12, Fig. S17). This analysis revealed only 32 out of 528 LGT candidates with a strong support for rooting within the specific prokaryotic taxon (Table S12). These genes were associated mostly with Actinobacteria (8 genes) that are ubiquitous aquatic and terrestrial bacteria, and mostly anaerobic bacteria of Bacteroidetes group (6 genes)(Fig. 9B_LGT). As an example of well-supported recent LGT, we can name two enzymes of proline synthesis pathway: glutamate 5-kinase (G5K) and glutamate-5-semialdehyde dehydrogenase (G5SD) that were acquired most likely from *Verrucomicrobiae* bacterium (Fig. S19 Proline synthesis). Because these enzymes are functionally interdependent, we can expect that corresponding genes were genetically linked in the donor and transferred together. Indeed G5K and G5SD are part of *Verrucomicrobiae* bacterium operon and corresponding two coding loci were located next to each other in *M. balamuthi* genome. The closest *Verrucomicrobiae* bacterium sequences were obtained from

samples of fresh water ponds of about 10-20 meters depth (Cabello-Yeves et al., 2017), which could be cohabited by *Mastigamoeba*. Neither of these genes is present in entamoebids, which supports their possibly recent acquisition after *M. bamamuthi/Entamoeba* split. We attribute the small number of well-supported prokaryotic groups as LGT sources to several factors: insufficient resolution of single gene phylogenies, lack of taxon sampling or poor taxonomic placement of metagenomic sequences of prokaryota and possible complicated non-vertical history of genes in prokaryotes.

Finally, we were interested to which extend *M. balamuthi* share LGT candidates with other eukaryotes, particularly with entamoebids. To do so, we search for eukaryotic lineages that form a monophyletic group with the *M. balamuthi* LGT candidates. This analysis revealed that 358 out of 528 LGT candidates were unique to *M. balamuthi* and 170 LGT candidates were shared with various eukaryotes (Fig. 9C_LGT). As expected, the most LGT candidates (34 OGs) were shared with Entamoebidae and 24 out of 34 OGs were unique to *M. balamuthi* and Entamoebidae (Table S12). NIF system components that were previously suggested to be acquired by LCAA were recovered using this approach (Nyvltova et al., 2013). Furthermore, the unique 24 OGs include several enzymes crucial to the anaerobic metabolism such as [Fe,Fe] hydrogenase, pyruvate-phosphate dikinase, pyruvate kinase, acetate kinase, and iron-sulfur flavoprotein as well as enzymes of aminoacid metabolisms including serine O-acetyltransferase of *de novo* cysteine biosynthetic pathway, methionine-gamma-lyase, and arginine decarboxylase. Interestingly, the later enzyme mediates survival of the passage of enteric bacteria through acidic parts of digestive tract as needs *E. histolytica* (Iyer et al., 2003). Type A flavoprotein is shared in addition to *M. balamuthi* and entamoebids by a free-living anaerobic excavate *Stygiella incarcerata*. Dikaryotic fungi were the second group with the highest number of the common LGT OGs (26). These OGs includes important enzyme for polysaccharide degradation such as xylan 1,4-beta-xylosidase, α - amylase, and cellulase (Table S12, Fig. 9D). Base on the three topologies, it seems more plausible that the gene transfer was directed from dikaryotic fungi to *M. balamuthi*,

although we cannot exclude a possibility that both eukaryotes acquired these enzymes from the same prokaryotic source of organic matter decomposer living in the common niche.

Conclusions

Analysis of the *M. balamuthi* genome allowed us to reason about the transition of the common ancestor of Dictyostelida and Archamoebae to common anaerobic ancestor of *M. balamuthi* and *Entamoeba* species (LCAA) and to separate the evolution of parasitism of *Entamoeba* lineage from adaptation of free-living amoeba to low-oxygen organic-rich water sediments. Two events, adaptation to anaerobiosis and adaptation to parasitism were accompanied with massive gene losses, although these were quite different. During the transition to anaerobic lifestyle, the losses of gene families were related mainly to oxygen-dependent pathways and function of organelles, particularly mitochondria and peroxisomes. In the second step, gene losses in parasitic *Entamoeba* are significantly associated with amino acid and nucleotide metabolism, however, majority of losses are distributed across most OGs resulting in overall decrease in metabolic complexity. In contrast, there is not such a dramatic gene loss in *M. balamuthi*.

Gene losses were compensated by numerous LGTs from prokaryotes to acquire new metabolic capacity. There is a growing list of reports suggesting that LGT from various prokaryotes to different eukaryotic lineages played a significant role in eukaryotic adaptation to specific niches. For example, *E. histolytica* and *T. vaginalis* interact with the rich and dense mucosal microbiota within the host and it seems plausible that these eukaryotes acquired genes from the mucosal bacterial sources. Comparison of putative gene donors between anaerobic *E. histolytica* and aerobic free-living *D. discoideum* showed that the former was enriched for Bacteroidetes and Firmicutes donors, while then latter for Proteobacteria. This analysis could be interpreted as a result of LGTs in the different habitats for the two Amoebozoa, and further support acquisition of genes from the gut microbiota by *E. histolytica*. However, this picture did not consider adaptation to anaerobiosis that most likely preceded parasitism

and consequently most LGTs related with adaptation to oxygen-poor niche happen in free habitat. Our analysis of *M. balamuthi* LGT donors revealed that Bacterioidetes as putative gene donors are also enriched in this free-living organism. As these bacteria are present in the gut but also widely distributed in water, sediments and soil, we can predict that LCAA gained Bacterioidetes genes related to anaerobiosis from environmental bacteria not from the gut microbiota. After split of *M. balamuthi*/Entamoebidae, *M. balamuthi* seems to gain genes to increase the metabolic capacity as decomposer, while entamoebids acquired gene related to pathogenicity and virulence. Genes providing these specific properties are indeed of different origin. For example *Entamoeba* BspA surface adhesine were suggested to be acquired from Bacterioidetes, while most OGs of *M. balamuthi* are related to Actinobacteria. However, we cannot rule out the possibility that some gains that we suggested for *M. balamuthi* lineage were gained already by free-living LCAA and subsequently lost in entamoebids. Our study of gene histories of Conosa group underlined an importance of comparative genomics for understanding how aerobic protists become anaerobes and how free living anaerobes become parasites.

Materials and methods

Cell culture

Axenic culture of *M. balamuthi* strain ATCC 30984 (kindly provided by M. Müller, Rockefeller University, USA) was cultured in PYGC medium (Chavez et al., 1986) at 24°C in 50 ml tissue culture flasks. The cells were transferred to fresh medium weekly.

DNA isolation and sequencing

DNA from axenically grown *M. balamuthi* culture was isolated using phenol-chloroform extraction.

Shotgun (500-800 bp) and paired-end (8-20 kb) fragment genomic libraries were prepared and sequenced using 454 GS FLX platform with Titanium chemistry (Roche Diagnostic, Rotkreuz, Switzerland) that generate 4.5 mil reads, and Illumina MiSeq (San Diego, USA) at EMBL (Heidelberg, Germany)

RNA isolation and sequencing

Total RNA was isolated using TRIzol (Total RNA Isolation Reagent, Gibco BRL) and purified from DNA with RNeasy Minelute Cleanup Kit (Qiagen). The cDNA was prepared using Transcriptor High Fidelity cDNA Synthesis Kit (Roche). The transcriptome was sequenced using Illumina RNA-seq. Transcripts were assembled using Trinity (Grabherr et al., 2011).

Gene assembly and annotation

The genome sequence was assembled using MaSuRCA genome assembler (Zimin et al., 2013). Gene prediction was performed using Augustus (Stanke & Morgenstern, 2005) with 298 manually curated training gene model dataset that yielded 16287 protein coding genes. The gene annotation was performed using Interproscan (Jones et al., 2014) and Blast2GO (Conesa et al., 2005). Structural RNA genes were predicted and annotated using Rfam database (Griffiths-Jones et al., 2005).

Detection of TEs

Initially, ORFs coding for transposable elements were detected using HMMER search against translated genome of *M. balamuthi* with low threshold (e-value 10). As a query, reverse transcriptase related (“RNA dependent RNA polymerase” Pfam clan) and DDE endonuclease related (“Ribonuclease H-like superfamily” Pfam clan) Pfam domains were used to predict possible retortransposons and DNA transposons, respectively. This analysis detected 4935 putative ORFs that were further analyzed by

HHsearch to detect conserved protein domains (Pfam, <https://pfam.xfam.org/>) and searched against the TrepDB database (Wicker et al., 2002) using protein BLAST. Repeats and palindroms in the vicinity of ORFs were detected using local nucleotide BLAST. Based on the combined data, 885 ORFs were classified according to Wicker (Wicker et al., 2007). Those ORFs that showed homology to a group of TEs but lack any specific features (protein domains, flanking repeats) were manually annotated as incomplete.

Orthologous groups prediction

Prediction of orthologous groups was based on the eggNOG database ver. 4.5 (Powell et al., 2014), 2014) using HMMER (Finn et al., 2011). Proteomes of selected members of Amoebozoa and Opisthokonta (Table S6) were searched against the EggNOG database using HMMER. Only Eggnog Orthologous Groups (OGs) with best hit e-value $\leq 1e-10$ were considered. For each protein sequence non-overlapping hits with e-value $\leq 1e-5$ were selected.

Phylogenetic analysis

Phylogenies were builded for each predicted *M. balamuthi* protein. *M. balamuthi* protein sequences were used as queries to search the Uniprot protein database using DIAMOND protein aligner (version diamond/0.9.10) in a sensitive mode. DIAMOND hits with e-value higher than $1e-5$ were discarded. 10000 best hits were used for further analysis. NCBI taxonomy was assigned to the homologous sequences (hits) and hit regions homologous to the query sequence were clustered at 80% identity level using CD-HIT (Fu et al., 2012). Presence of hits from both eukaryota and prokaryota in the 80% identity group was suspected as contamination. The source of contamination was estimated based on the number of unique taxons in the 80% identity group and the contamination was removed. Only one sequence per taxonomic group of the 80% identity group was chosen for further analysis to limit redundancy. The multiple sequence alignment was constructed using MAFFT with default settings

(Katoh & Standley, 2013) and trimmed with BMGE (Crisuolo & Gribaldo, 2010) using BLOSUM30 matrix and block size 1. Phylogenetic trees were computed with IQ-TREE (Nguyen et al., 2015) with 1000 ultra-fast bootstraps and automatic model selection. Nodes with statistical support below 90% were removed. Further analysis of the trees was implemented using the ETE package (Huerta-Cepas et al., 2016).

LGT evaluation

LGT evaluation was based on calculation of two coefficients: (i) LGT Prokaryotic Coefficient (LGTPC) to estimate whether the gene transfer was directed from a prokaryotic lineage to eukaryotes. LGTPC compares the number of all unique taxons of the prokaryotes in the phylogenetic tree and the number of eukaryotic sequences that are monophyletic with the *M. balamuthi* query:

$$LGTPC = \frac{N_{\text{proktaxons}}}{N_{\text{proktaxons}} + N_{\text{euktaxons}}}$$

$N_{\text{proktaxons}}$, the number of unique prokaryotic taxons; $N_{\text{euktaxons}}$, the number of unique eukaryotic taxons that are monophyletic with *M. balamuthi*.

(ii) LGT Eukaryotic Coefficient (LGTEC) to estimate based on the principle of maximal parsimony whether the gene was present in the common ancestor of eukaryotes, or whether it was possibly transferred by LGT to more eukaryotic lineages. We modeled both scenarios, estimated a minimal number of transfers and losses for each scenario and compared the results:

$$LGTEC = \frac{N_{\text{ancestralLosses}}}{N_{\text{ancestralLosses}} + N_{\text{newTransfers}} + N_{\text{newLosses}}}$$

$N_{\text{ancestralLosses}}$, minimal number of losses if the gene was present in the common ancestor of eukaryotes;
 $N_{\text{newTransfers}}$, minimal number of transfers between eukaryotic lineages if the gene was gained after the radiation of eukaryotes; $N_{\text{newLosses}}$, minimal number of losses if the gene was gained after the radiation of eukaryotes. As an LGT value we selected the lower number calculated for the two coefficients (Fig S17). The strong support for LGT was considered for gene if both coefficients were ≥ 0.75 (Table S12, Fig. S17).

Domain boundary frequency

To assess possible source of prokaryota-to- eukaryota LGTs based on the phylogenetic tree, we introduce a measure that we call domain boundary frequency (DBF) that is calculated for each taxon at each node as follows:

T, taxon (defined in Table 12)

P_T , proportion of taxon T at given node.

N_{children} , number of children at the note

i, identifier of a child (1, 2, 3..., N_{children})

P_{T_i} ... taxon frequency for the child i

$$P_T = \frac{\sum_{i=1}^{N_{\text{children}}} f_{T_i}}{N_{\text{children}}}$$

The Domain boundary frequencies (DBFs) are the prokaryotic taxonomy frequencies at the root of eukaryotic subtree i.e. at the boundary between eukaryotic and prokaryotic sequences. The most significant is prokaryotic taxon with the highest DBF value. We classify the DBFs into three classes: (i)

$0 < \text{DBF} \leq 0.3$: unclear origin, (ii) $0.25 < \text{DBF} \leq 0.75$: possible sister relationship with given taxon, (iii) $0.75 < \text{DBF} \leq 1$: given taxon is likely source of the LGT.

Data Availability

The genomic sequences have been deposited in the online resource for community annotation of eukaryotes (ORCAE, <https://bioinformatics.psb.ugent.be/orcae/overview/Masba>).

Acknowledgments

This work was supported by Czech Science Foundation (P305/11/1061, 16-06123S), NPU II (LQ1604) provided by the Ministry of Education, Youth and Sport (MEYS) of the Czech Republic, CePaViP (CZ.02.1.01/0.0/0.0/16_019/0000759) provided by ERD Funds, and MICROBION funded from EU H2020 (No 810224).

Declaration of Interests

The authors declare no competing interests.

References

- Andrä, J., Herbst, R., & Leippe, M. (2003). Amoebapores, archaic effector peptides of protozoan origin, are discharged into phagosomes and kill bacteria by permeabilizing their membranes. *Developmental and Comparative Immunology*, 27(4), 291–304. [https://doi.org/10.1016/S0145-305X\(02\)00106-4](https://doi.org/10.1016/S0145-305X(02)00106-4)
- Artzi, L., Bayer, E. A., & Morais, S. (2017). Cellulosomes: Bacterial nanomachines for dismantling plant polysaccharides. In *Nature Reviews Microbiology* (Vol. 15, Issue 2, pp. 83–95). Nature Publishing Group. <https://doi.org/10.1038/nrmicro.2016.164>
- Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W., Gordon, P., Duruflé, L., Gaasterland, T., Lopez, P., Müller, M., & Philippe, H. (2002). The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3), 1414–1419. <https://doi.org/10.1073/pnas.032662799>
- Bauer, S., Grossmann, S., Vingron, M., & Robinson, P. N. (2008). Ontologizer 2.0--a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics (Oxford, England)*, 24(14), 1650–1651. <https://doi.org/10.1093/bioinformatics/btn250>
- Beck, M., & Hurt, E. (2017). The nuclear pore complex: Understanding its function through structural insight. In *Nature Reviews Molecular Cell Biology* (Vol. 18, Issue 2, pp. 73–89). Nature Publishing Group. <https://doi.org/10.1038/nrm.2016.147>
- Blaxter, M., & Koutsovoulos, G. (2015). The evolution of parasitism in Nematoda. *Parasitology*, 142, S26–S39. <https://doi.org/10.1017/S0031182014000791>
- Bruchhaus, I., Loftus, B. J., Hall, N., & Tannich, E. (2003). The intestinal protozoan parasite Entamoeba histolytica contains 20 cysteine protease genes, of which only a small subset is expressed during in vitro cultivation. *Eukaryotic Cell*, 2(3), 501–509. <https://doi.org/10.1128/ec.2.3.501-509.2003>
- Bujanover, S., Katz, U., Bracha, R., & Mirelman, D. (2003). A virulence attenuated amoebapore-less mutant of Entamoeba histolytica and its interaction with host cells. *International Journal for Parasitology*, 33(14), 1655–1663. [https://doi.org/10.1016/s0020-7519\(03\)00268-6](https://doi.org/10.1016/s0020-7519(03)00268-6)
- Cabello-Yeves, P. J., Ghai, R., Mehrshad, M., Picazo, A., Camacho, A., & Rodriguez-Valera, F. (2017). Reconstruction of Diverse Verrucomicrobial Genomes from Metagenome Datasets of Freshwater

- Reservoirs. *Frontiers in Microbiology*, 8(NOV), 2131. <https://doi.org/10.3389/fmicb.2017.02131>
- 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., Sicheritz-ponten, T., Noel, C. J., Dacks, J. B., Foster, P. G., Simillion, C., Peer, Y. Van De, Miranda-saavedra, D., Barton, G. J., Westrop, G. D., ... Johnson, P. J. (2007). Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science*, 315(January), 207–212. <https://doi.org/10.1126/science.1132894>
- Cavalier-Smith, T. (1998). A revised six-kingdom system of life. *Biol.Rev.Camb.Philos.Soc.*, 73(1464-7931 (Print)), 203–266.
- Chatterjee, A., Ghosh, S. K., Jang, K., Bullitt, E., Moore, L., Robbins, P. W., & Samuelson, J. (2009). Evidence for a “wattle and daub” model of the cyst wall of entamoeba. *PLoS Pathogens*, 5(7), e1000498. <https://doi.org/10.1371/journal.ppat.1000498>
- Chavez, L. A., Balamuth, W., & Gong, T. (1986). A light and electron microscopical study of a new, polymorphic free-living amoeba, *Phreatamoeba balamuthi* n. g., n. sp. *Journal of Protozoology*, 33(0022-3921 (Print)), 397–404.
- Cheng, Y., Shi, Q., Sun, R., Liang, D., Li, Y., Li, Y., Jin, W., & Zhu, W. (2018). The biotechnological potential of anaerobic fungi on fiber degradation and methane production. *World Journal of Microbiology & Biotechnology*, 34(10), 155. <https://doi.org/10.1007/s11274-018-2539-z>
- Clark, C. G., Alsmark, U. C. M., Tazreiter, M., Saito-Nakano, Y., Ali, V., Marion, S., Weber, C., Mukherjee, C., Bruchhaus, I., Tannich, E., Leippe, M., Sicheritz-Ponten, T., Foster, P. G., Samuelson, J., Noël, C. J., Hirt, R. P., Embley, T. M., Gilchrist, C. A., Mann, B. J., ... Hall, N. (2007). Structure and Content of the Entamoeba histolytica Genome. In *Advances in Parasitology* (Vol. 65, pp. 51–190). [https://doi.org/10.1016/S0065-308X\(07\)65002-7](https://doi.org/10.1016/S0065-308X(07)65002-7)
- Clark, C. Graham, Ali, I. K. M., Zaki, M., Loftus, B. J., & Hall, N. (2006). Unique organisation of tRNA genes in Entamoeba histolytica. *Molecular and Biochemical Parasitology*, 146(1), 24–29. <https://doi.org/10.1016/j.molbiopara.2005.10.013>
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal annotation and visualization tool in functional genomics research. Application note. *Bioinformatics*, 21, 3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
- Coyne, R. S., Hannick, L., Shanmugam, D., Hostetler, J. B., Brami, D., Joardar, V. S., Johnson, J., Radune, D., Singh, I., Badger, J. H., Kumar, U., Saier, M., Wang, Y., Cai, H., Gu, J., Mather, M. W., Vaidya, A. B., Wilkes, D. E., Rajagopalan, V., ... Clark, T. G. (2011). Comparative genomics

- of the pathogenic ciliate *Ichthyophthirius multifiliis*, its free-living relatives and a host species provide insights into adoption of a parasitic lifestyle and prospects for disease control. *Genome Biology*, 12(10), R100. <https://doi.org/10.1186/gb-2011-12-10-r100>
- Criscuolo, A., & Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, 10, 210. <https://doi.org/10.1186/1471-2148-10-210>
- Davies, K. M., Strauss, M., Daum, B., Kief, J. H., Osiewacz, H. D., Rycovska, A., Zickermann, V., & Kühlbrandt, W. (2011). Macromolecular organization of ATP synthase and complex I in whole mitochondria. *Proceedings of the National Academy of Sciences of the United States of America*, 108(34), 14121–14126. <https://doi.org/10.1073/pnas.1103621108>
- Dolezal, P., Dagley, M. J., Kono, M., Wolyneć, P., Likić, V. A., Foo, J. H., Sedinová, M., Tachezy, J., Bachmann, A., Bruchhaus, I., & Lithgow, T. (2010). The essentials of protein import in the degenerate mitochondrion of *Entamoeba histolytica*. *PLoS Pathogens*, 6(3). <https://doi.org/10.1371/journal.ppat.1000812>
- Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, 39(Web Server issue), W29-37. <https://doi.org/10.1093/nar/gkr367>
- Frisardi, M., Ghosh, S. K., Field, J., Van Dellen, K., Rogers, R., Robbins, P., & Samuelson, J. (2000). The most abundant glycoprotein of amebic cyst walls (Jacob) is a lectin with five Cys-rich, chitin-binding domains. *Infection and Immunity*, 68(7), 4217–4224. <https://doi.org/10.1128/IAI.68.7.4217-4224.2000>
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Gilbert, C., Schaack, S., Pace, J. K., Brindley, P. J., & Feschotte, C. (2010). A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature*, 464(7293), 1347–1350. <https://doi.org/10.1038/nature08939>
- Gill, E. E., Diaz-Trivino, S., Barbera, M. J., Silberman, J. D., Stechmann, A., Gaston, D., Tamas, I., & Roger, A. J. (2007). Novel mitochondrion-related organelles in the anaerobic amoeba *Mastigamoeba balamuthi*. *Molecular Microbiology*, 66(0950-382X (Print)), 1306–1320.
- Ginger, M., & Field, M. C. (2016). Making the pathogen: Evolution and adaptation in parasitic protists.

Molecular and Biochemical Parasitology, 209(1–2), 1–2.

<https://doi.org/10.1016/j.molbiopara.2016.11.002>

- Glenn, K., Ingram-Smith, C., & Smith, K. S. (2014). Biochemical and kinetic characterization of xylulose 5-phosphate/ fructose 6-phosphate phosphoketolase 2 (Xfp2) from *Cryptococcus neoformans*. *Eukaryotic Cell*, 13(5), 657–663. <https://doi.org/10.1128/EC.00055-14>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652. <https://doi.org/10.1038/nbt.1883>
- Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., & Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Research*, 33(Database issue), D121-4. <https://doi.org/10.1093/nar/gki081>
- Gualdrón-López, M., Brennand, A., Hannaert, V., Quiñones, W., Cáceres, A. J., Bringaud, F., Concepción, J. L., & Michels, P. A. M. (2012). When, how and why glycolysis became compartmentalised in the Kinetoplastea. A new look at an ancient organelle. *International Journal for Parasitology*, 42(1), 1–20. <https://doi.org/10.1016/j.ijpara.2011.10.007>
- Heinz, E., Williams, T. A., Nakjang, S., Noël, C. J., Swan, D. C., Goldberg, A. V., Harris, S. R., Weinmaier, T., Markert, S., Becher, D., Bernhardt, J., Dagan, T., Hacker, C., Lucocq, J. M., Schweder, T., Rattei, T., Hall, N., Hirt, R. P., & Embley, T. M. (2012). The Genome of the Obligate Intracellular Parasite *Trachipleistophora hominis*: New Insights into Microsporidian Genome Dynamics and Reductive Evolution. *PLoS Pathogens*, 8(10). <https://doi.org/10.1371/journal.ppat.1002979>
- Hellberg, A., Nickel, R., Lotter, H., Tannich, E., & Bruchhaus, I. (2001). Overexpression of cysteine proteinase 2 in *Entamoeba histolytica* or *Entamoeba dispar* increases amoeba-induced monolayer destruction in vitro but does not augment amoebic liver abscess formation in gerbils. *Cellular Microbiology*, 3(1), 13–20. <https://doi.org/10.1046/j.1462-5822.2001.00086.x>
- Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*, 33(6), 1635–1638. <https://doi.org/10.1093/molbev/msw046>
- Iyer, R., Williams, C., & Miller, C. (2003). Arginine-Agmatine Antiporter in Extreme Acid Resistance

in *Escherichia coli*. *Journal of Bacteriology*, 185(22), 6556–6561.

<https://doi.org/10.1128/JB.185.22.6556-6561.2003>

Jackson, A. P., Otto, T. D., Aslett, M., Armstrong, S. D., Bringaud, F., Schlacht, A., Hartley, C., Sanders, M., Wastling, J. M., Dacks, J. B., Acosta-Serrano, A., Field, M. C., Ginger, M. L., & Berriman, M. (2016). Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism. *Current Biology : CB*, 26(2), 161–172.

<https://doi.org/10.1016/j.cub.2015.11.055>

Janouskovec, J., & Keeling, P. J. (2016). Evolution: Causality and the origin of parasitism. In *Current Biology* (Vol. 26, Issue 4, pp. R174–R177). Cell Press. <https://doi.org/10.1016/j.cub.2015.12.057>

Janouškovec, J., Tikhonenkov, D. V., Burki, F., Howe, A. T., Kolísko, M., Mylnikov, A. P., & Keeling, P. J. (2015). Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proceedings of the National Academy of Sciences of the United States of America*, 112(33), 10200–10207. <https://doi.org/10.1073/pnas.1423790112>

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>

Kabbara, S., Hérivieux, A., Dugé de Bernonville, T., Courdavault, V., Clastre, M., Gastebois, A., Osman, M., Hamze, M., Cock, J. M., Schaap, P., & Papon, N. (2019). Diversity and Evolution of Sensor Histidine Kinases in Eukaryotes. *Genome Biology and Evolution*, 11(1), 86–108.

<https://doi.org/10.1093/gbe/evy213>

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780.

<https://doi.org/10.1093/molbev/mst010>

Le, T., Žárský, V., Nývltová, E., Rada, P., Harant, K., Vancová, M., Verner, Z., Hrdý, I., & Tachezy, J. (2020). Anaerobic peroxisomes in *Mastigamoeba balamuthi*. *Proceedings of the National Academy of Sciences*, 201909755. <https://doi.org/10.1073/pnas.1909755117>

Leippe, M., Ebel, S., Schoenberger, O. L., Horstmann, R. D., & Müller-Eberhard, H. J. (1991). Pore-forming peptide of pathogenic *Entamoeba histolytica*. *Proceedings of the National Academy of Sciences of the United States of America*, 88(17), 7659–7663.

<https://doi.org/10.1073/pnas.88.17.7659>

- Lidell, M. E., Moncada, D. M., Chadee, K., & Hansson, G. C. (2006). Entamoeba histolytica cysteine proteases cleave the MUC2 mucin in its C-terminal domain and dissolve the protective colonic mucus gel. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(24), 9298–9303. <https://doi.org/10.1073/pnas.0600623103>
- Loftus, B., Anderson, I., Davies, R., Alsmark, U. C., Samuelson, J., Amedeo, P., Roncaglia, P., Berriman, M., Hirt, R. P., Mann, B. J., Nozaki, T., Suh, B., Pop, M., Duchene, M., Ackers, J., Tannich, E., Leippe, M., Hofer, M., Bruchhaus, I., ... Hall, N. (2005). The genome of the protist parasite *Entamoeba histolytica*. *Nature*, *433*(1476-4687 (Electronic)), 865–868.
- MacFarlane, R. C., & Singh, U. (2007). Identification of an *Entamoeba histolytica* serine-, threonine-, and isoleucine-rich protein with roles in adhesion and cytotoxicity. *Eukaryotic Cell*, *6*(11), 2139–2146. <https://doi.org/10.1128/EC.00174-07>
- Makioka, A., Kumagai, M., Kobayashi, S., & Takeuchi, T. (2005). Entamoeba invadens: cysteine protease inhibitors block excystation and metacystic development. *Experimental Parasitology*, *109*(1), 27–32. <https://doi.org/10.1016/j.exppara.2004.10.003>
- Makiuchi, T., Mi-ichi, F., Nakada-Tsukui, K., & Nozaki, T. (2013). Novel TPR-containing subunit of TOM complex functions as cytosolic receptor for Entamoeba mitochondrial transport. *Scientific Reports*, *3*, 1129. <https://doi.org/10.1038/srep01129>
- Makki, A., Rada, P., Žárský, V., Kerešiče, S., Kováčik, L., Novotný, M., Jores, T., Rapaport, D., & Tachezy, J. (2019). Triplet-pore structure of a highly divergent TOM complex of hydrogenosomes in *Trichomonas vaginalis*. *PLoS Biology*, *17*(1), e3000098. <https://doi.org/10.1371/journal.pbio.3000098>
- Mans, B. J., Anantharaman, V., Aravind, L., & Koonin, E. V. (2004). Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle*, *3*(12), 1625–1650. <https://doi.org/10.4161/cc.3.12.1316>
- Mathur, V., Kolísko, M., Hehenberger, E., Irwin, N. A. T., Leander, B. S., Kristmundsson, Á., Freeman, M. A., & Keeling, P. J. (2019). Multiple Independent Origins of Apicomplexan-Like Parasites. *Current Biology*, *29*(17), 2936-2941.e5. <https://doi.org/10.1016/j.cub.2019.07.019>
- Mi-Ichi, F., Nozawa, A., Yoshida, H., Tozawa, Y., & Nozaki, T. (2015). Evidence that the *Entamoeba histolytica* mitochondrial carrier family links mitochondrial and cytosolic pathways through exchange of 3'-phosphoadenosine 5'-phosphosulfate and ATP. *Eukaryotic Cell*, *14*(11), 1144–1150. <https://doi.org/10.1128/EC.00130-15>

- Mi-ichi, F, Abu, Y. M., Nakada-Tsukui, K., & Nozaki, T. (2009). Mitosomes in *Entamoeba histolytica* contain a sulfate activation pathway. *Proc.Natl.Acad.Sci.U.S.A*, *106*(1091-6490 (Electronic)), 21731–21736.
- Mi-ichi, Fumika, Miyamoto, T., Takao, S., Jeelani, G., Hashimoto, T., Hara, H., Nozaki, T., & Yoshida, H. (2015). Entamoeba mitosomes play an important role in encystation by association with cholesteryl sulfate synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(22), E2884–E2890. <https://doi.org/10.1073/pnas.1423718112>
- Mi-ichi, Fumika, Miyamoto, T., & Yoshida, H. (2017). Uniqueness of Entamoeba sulfur metabolism: sulfolipid metabolism that plays pleiotropic roles in the parasitic life cycle. *Molecular Microbiology*, *106*(3), 479–491. <https://doi.org/10.1111/mmi.13827>
- Nachury, M. V, Loktev, A. V, Zhang, Q., Westlake, C. J., Peränen, J., Merdes, A., Slusarski, D. C., Scheller, R. H., Bazan, J. F., Sheffield, V. C., & Jackson, P. K. (2007). A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. *Cell*, *129*(6), 1201–1213. <https://doi.org/10.1016/j.cell.2007.03.053>
- Nakada-Tsukui, K., Sekizuka, T., Sato-Ebine, E., Escueta-de Cadiz, A., Ji, D., Tomii, K., Kuroda, M., & Nozaki, T. (2018). AIG1 affects in vitro and in vivo virulence in clinical isolates of *Entamoeba histolytica*. *PLoS Pathogens*, *14*(3), e1006882. <https://doi.org/10.1371/journal.ppat.1006882>
- Neumann, N., Lundin, D., & Poole, A. M. (2010). Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor. *PloS One*, *5*(10), e13241. <https://doi.org/10.1371/journal.pone.0013241>
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Nishitani, Y., Horiuchi, A., Aslam, M., Kanai, T., Atomi, H., & Miki, K. (2018). Crystal structures of an archaeal chitinase ChiD and its ligand complexes. *Glycobiology*, *28*(6), 418–426. <https://doi.org/10.1093/glycob/cwy024>
- Nývtová, E., Stairs, C. W., Hrdý, I., Rídl, J., Mach, J., Pařes, J., Roger, A. J., & Tachezy, J. (2015). Lateral gene transfer and gene duplication played a key role in the evolution of mastigamoeba balamuthi hydrogenosomes. *Molecular Biology and Evolution*, *32*(4). <https://doi.org/10.1093/molbev/msu408>
- Nývtová, E., Šut'ák, R., Žárský, V., Harant, K., Hrdý, I., & Tachezy, J. (2017). Lateral gene transfer of

- p-cresol- and indole-producing enzymes from environmental bacteria to *Mastigamoeba balamuthi*. *Environmental Microbiology*, 19(3). <https://doi.org/10.1111/1462-2920.13636>
- Nyvtova, E., Sutak, R., Harant, K., Sedinova, M., Hrdy, I., Paces, J., Vlcek, C., & Tachezy, J. (2013). NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of *Mastigamoeba balamuthi*. *Proc.Natl.Acad.Sci.U.S.A*, 110(1091-6490 (Electronic)), 7371–7376.
- Opperdoes, F. R., & Borst, P. (1977). Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei*: the glycosome. *FEBS Letters*, 80(2), 360–364. [https://doi.org/10.1016/0014-5793\(77\)80476-6](https://doi.org/10.1016/0014-5793(77)80476-6)
- Pandey, A. K., Pain, J., Dancis, A., & Pain, D. (2019). Mitochondria export iron-sulfur and sulfur intermediates to the cytoplasm for iron-sulfur cluster assembly and tRNA thiolation in yeast. *The Journal of Biological Chemistry*, 294(24), 9489–9502. <https://doi.org/10.1074/jbc.RA119.008600>
- Pánek, T., Zadrobílková, E., Walker, G., Brown, M. W., Gentekaki, E., Hroudová, M., Kang, S., Roger, A. J., Tice, A. K., Vlček, Č., & Čepička, I. (2016). First multigene analysis of Archamoebae (Amoebozoa: Conosa) robustly reveals its phylogeny and shows that Entamoebidae represents a deep lineage of the group. *Molecular Phylogenetics and Evolution*. <https://doi.org/10.1016/j.ympev.2016.01.011>
- Petri, W. A., Haque, R., & Mann, B. J. (2002). The bittersweet interface of parasite and host: lectin-carbohydrate interactions during human invasion by the parasite *Entamoeba histolytica*. *Annual Review of Microbiology*, 56, 39–64. <https://doi.org/10.1146/annurev.micro.56.012302.160959>
- Pineda, E., Vázquez, C., Encalada, R., Nozaki, T., Sato, E., Hanadate, Y., Néquiz, M., Olivos-García, A., Moreno-Sánchez, R., & Saavedra, E. (2016). Roles of acetyl-CoA synthetase (ADP-forming) and acetate kinase (PPi-forming) in ATP and PPi supply in *Entamoeba histolytica*. *Biochimica et Biophysica Acta*, 1860(6), 1163–1172. <https://doi.org/10.1016/j.bbagen.2016.02.010>
- Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., Gabaldón, T., Rattei, T., Creevey, C., Kuhn, M., Jensen, L. J., von Mering, C., & Bork, P. (2014). eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Research*, 42(Database issue), D231-9. <https://doi.org/10.1093/nar/gkt1253>
- Pritham, E. J., Feschotte, C., & Wessler, S. R. (2005). Unexpected diversity and differential success of DNA transposons in four species of entamoeba protozoans. *Molecular Biology and Evolution*, 22(9), 1751–1763. <https://doi.org/10.1093/molbev/msi169>

- Pyrihová, E., Motycková, A., Voleman, L., Wandyszewska, N., Fišer, R., Seydlová, G., Roger, A., Kolísko, M., & Doležal, P. (2018). A Single Tim Translocase in the Mitosomes of *Giardia intestinalis* Illustrates Convergence of Protein Import Machines in Anaerobic Eukaryotes. *Genome Biology and Evolution*, *10*(10), 2813–2822. <https://doi.org/10.1093/gbe/evy215>
- Reeves, R. E., & Guthrie, J. D. (1975). Acetate kinase (pyrophosphate). A fourth pyrophosphate-dependent kinase from *Entamoeba histolytica*. *Biochemical and Biophysical Research Communications*, *66*(4), 1389–1395. [https://doi.org/10.1016/0006-291x\(75\)90513-6](https://doi.org/10.1016/0006-291x(75)90513-6)
- Seigneur, M., Mounier, J., Prevost, M.-C., & Guillén, N. (2005). A lysine- and glutamic acid-rich protein, KERP1, from *Entamoeba histolytica* binds to human enterocytes. *Cellular Microbiology*, *7*(4), 569–579. <https://doi.org/10.1111/j.1462-5822.2005.00487.x>
- Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, *33*(SUPPL. 2). <https://doi.org/10.1093/nar/gki458>
- Van Dellen, K., Ghosh, S. K., Robbins, P. W., Loftus, B., & Samuelson, J. (2002). *Entamoeba histolytica* lectins contain unique 6-Cys or 8-Cys chitin-binding domains. *Infection and Immunity*, *70*(6), 3259–3263. <https://doi.org/10.1128/IAI.70.6.3259-3263.2002>
- Walker, G., Simpson, A. G., Edgcomb, V. P., Sogin, M. L., & Patterson, D. J. (2001). Ultrastructural identities of *Mastigamoeba punctachora*, *Mastigamoeba simplex* and *Mastigella commutans* and assessment of hypotheses of relatedness of the pelobionts (Protista). *European Journal of Protistology*, *37*, 25–49.
- Wicker, T., Matthews, D. E., & Keller, B. (2002). TREP: A database for Triticeae repetitive elements. In *Trends in Plant Science* (Vol. 7, Issue 12, pp. 561–562). Elsevier Ltd. [https://doi.org/10.1016/S1360-1385\(02\)02372-5](https://doi.org/10.1016/S1360-1385(02)02372-5)
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. In *Nature Reviews Genetics* (Vol. 8, Issue 12, pp. 973–982). Nature Publishing Group. <https://doi.org/10.1038/nrg2165>
- Willhoeft, U., Buss, H., & Tannich, E. (1999). DNA sequences corresponding to the ariel gene family of *Entamoeba histolytica* are not present in *E. dispar*. *Parasitology Research*, *85*(8–9), 787–789. <https://doi.org/10.1007/s004360050633>
- Wu, T. D., & Watanabe, C. K. (2005). GMAP: A genomic mapping and alignment program for mRNA

and EST sequences. *Bioinformatics*, 21(9), 1859–1875.

<https://doi.org/10.1093/bioinformatics/bti310>

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, 29(21), 2669–2677.

<https://doi.org/10.1093/bioinformatics/btt476>

Figure Legends

Figure 1. Comparison of genome sizes, number of genes and GC contents in selected members of *Amoebozoa*. Parasitic lifestyle is highlighted.

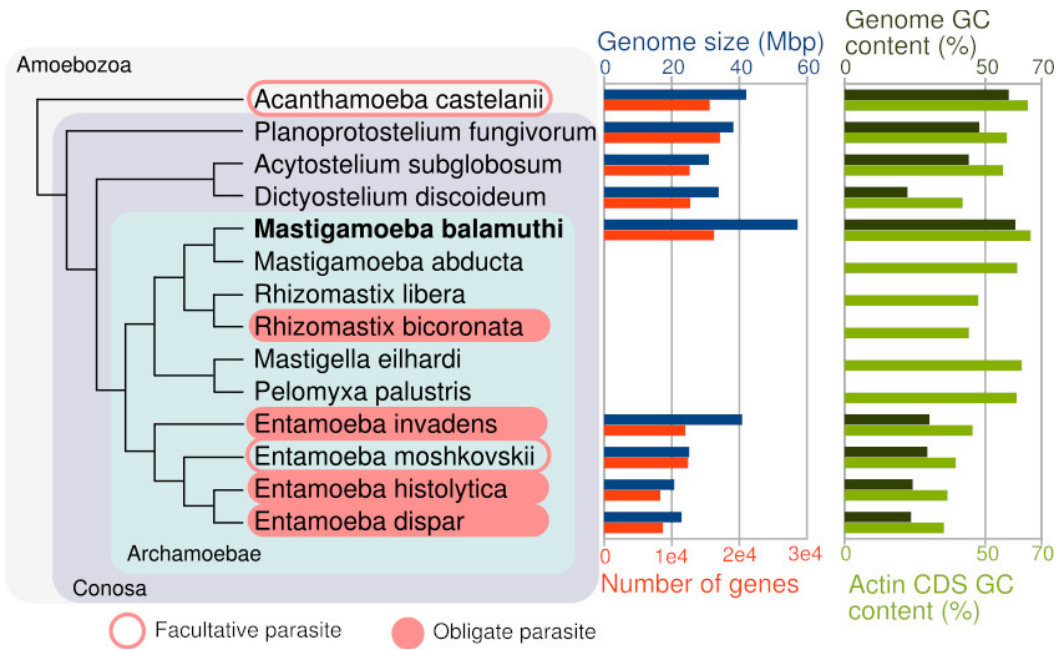


Figure 2. Comparison of DNA transposons and retrotransposons identified in *M. balamuthi* and two *Entamoeba* species. Reverse transcriptase related (“RNA dependent RNA polymerase” Pfam clan) and DDE endonuclease related (“Ribonuclease H-like superfamily” Pfam clan) Pfam domains were used for TE searches in *M. balamuthi* genome. Identified TEs were classified according to the TREP database. TEs in *E. histolytica* and *E. invadens* are according to Pritham et al. (2005).

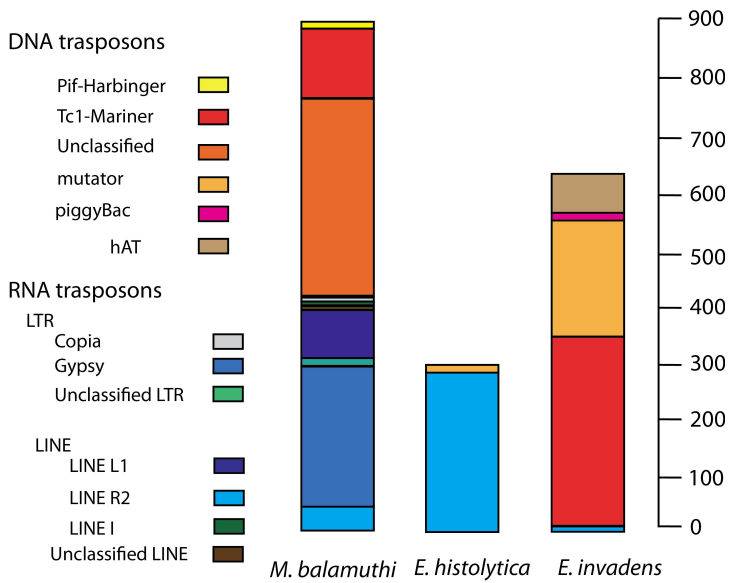


Figure 3. Losses and gains of orthologous groups in dictyostelids and Archamoebae. Predicted proteins of amoebozoans and selected representatives of opisthokonts were assigned to the orthologous groups (OGs) of the EggNOG database (Powell et al., 2014) using profile-HMM homology search (HMMER). Gains and losses of OGs were reconstructed for each node of the organismal phylogeny with the assumption of vertical inheritance of these OGs. For each branch these parameters are shown: (i) Number of gains (in blue) and losses (in red) of OGs that were predicted to occur at that branch and resulting predicted number of OGs (in black). (ii) Specific terms of the Gene Ontology enriched in the group of gained (blue) and lost (red) with high statistical significance using Ontologizer tool (Bauer, 2008). Statistical significance is shown as p-value next to the term. (iii) Net graphs show pearson correlation coefficients between the COG database functional categories (Tatusov, 2003), and gains (blue line) and losses (red line) respectively. The gray areas demarcate correlation coefficient values with low statistical significance ($p\text{-value} \geq 0.01$). At the terminal branches color-coded presence/absence matrices of OGs with an assigned COG functional category are shown. The description of the COG functional categories is shown in the bottom left corner.

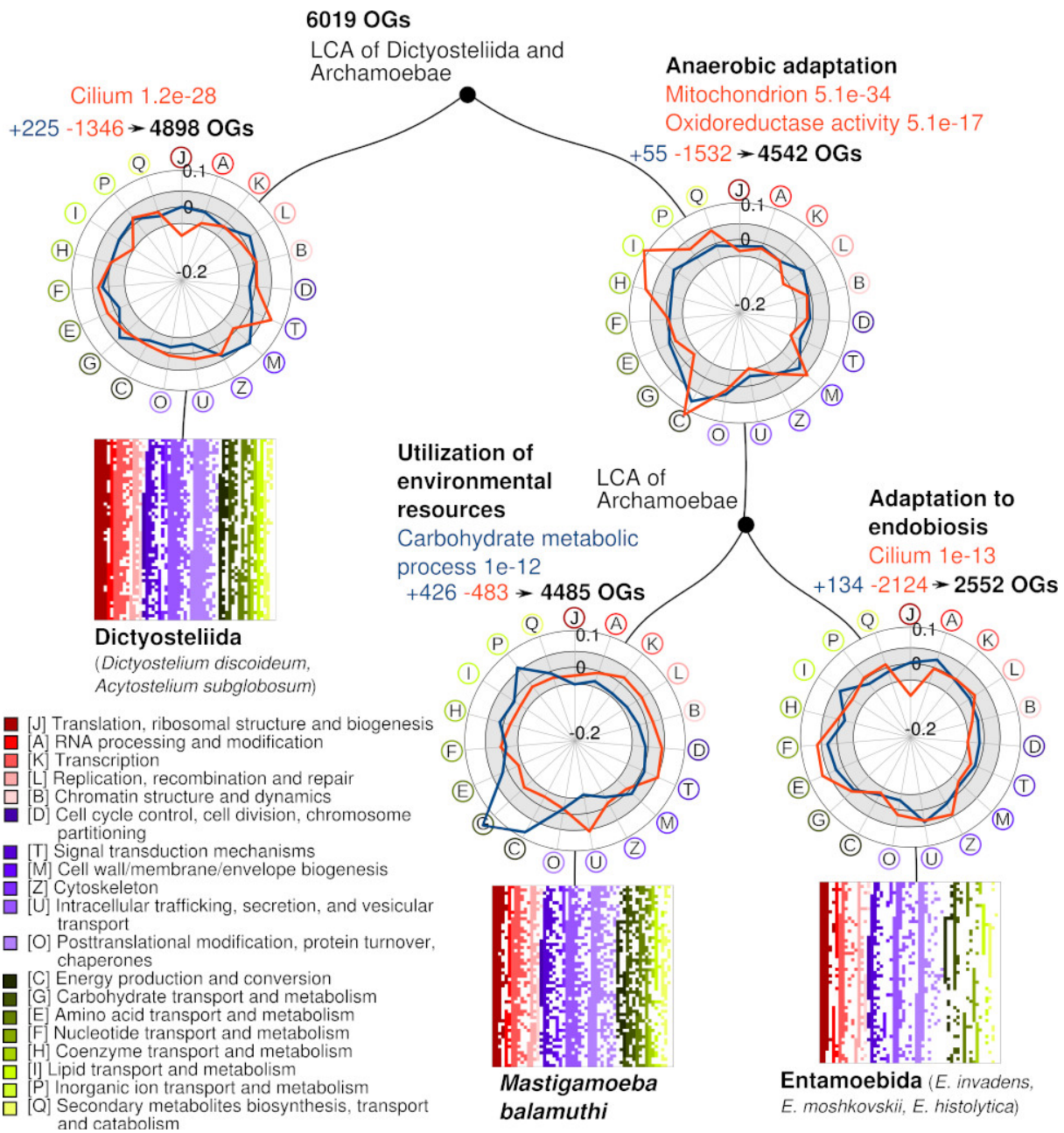


Figure 4. Ten largest EggNOG orthologous groups of *M. balamuthi* (A) and *E. histolytica* (B)

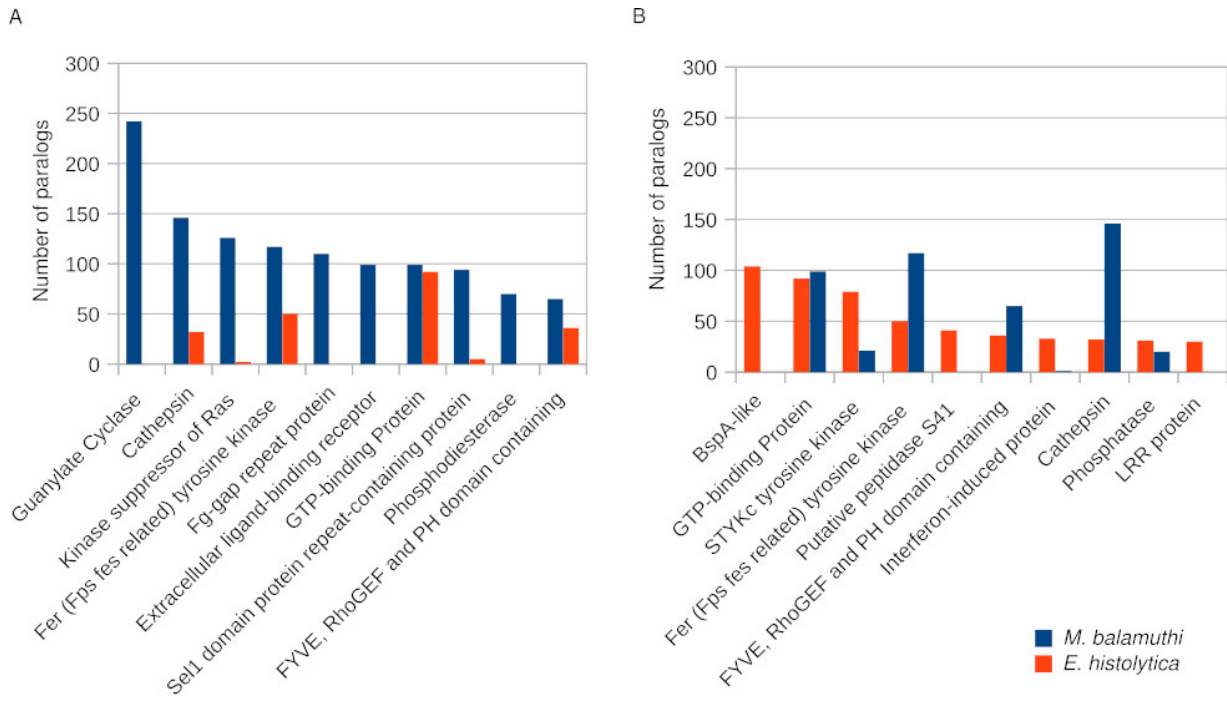


Figure 5. Comparison of nuclear pore complexes in *M. balamuthi* and *E. histolytica*. Components of cytoplasmic filaments are in green box, Nups of cytoplasmic, nuclear and inner ring complexes are in pink boxes, FG-Nups of central channel are in orange, the transmembrane Nups are in yellow box, nuclear basket components are in violet. NPC subunits of *E. histolytica* identified in this work are in red. TPR, translocated promoter region; Ndc1, nuclear division 1; Gp201, glycoprotein 210; Rae1, mRNA export factor 1; Gle1, glycine-leucin-phenylalanine-glycine lethal. Topology of the components is according to (Beck & Hurt, 2017).

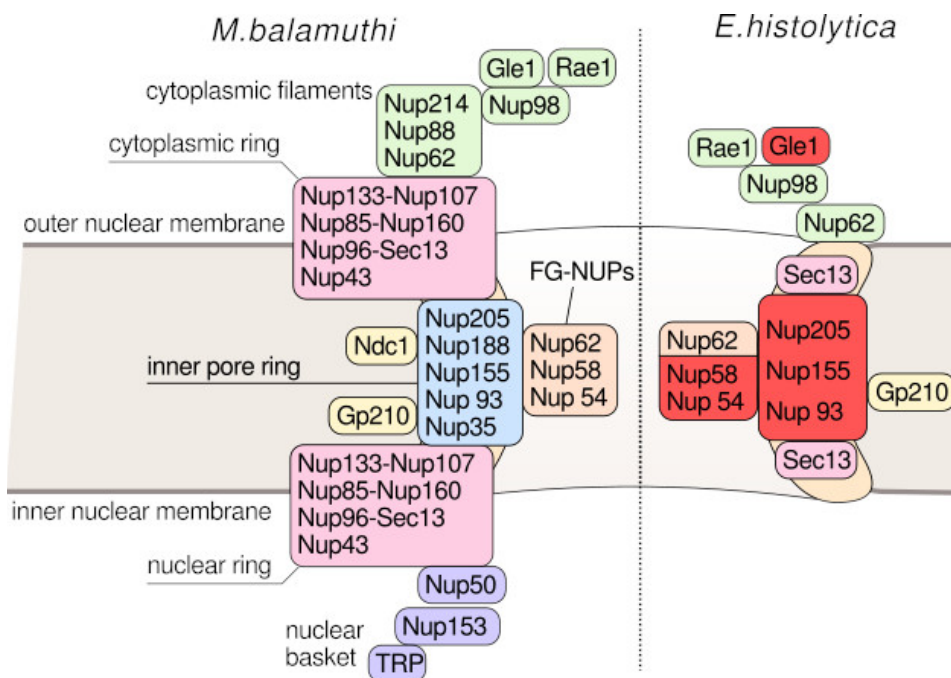


Figure 6. Comparison of metabolic pathways complexity in *D. discoideum*, *M. balamuthi* and *E. histolytica* using KEGG pathways mapping. KEGG enzymes were detected in the genomes using KofamKOALA (Aramaki et al., 2019). Numbers of unique enzymes (not counting paralogs) of each KEGG pathway are shown.

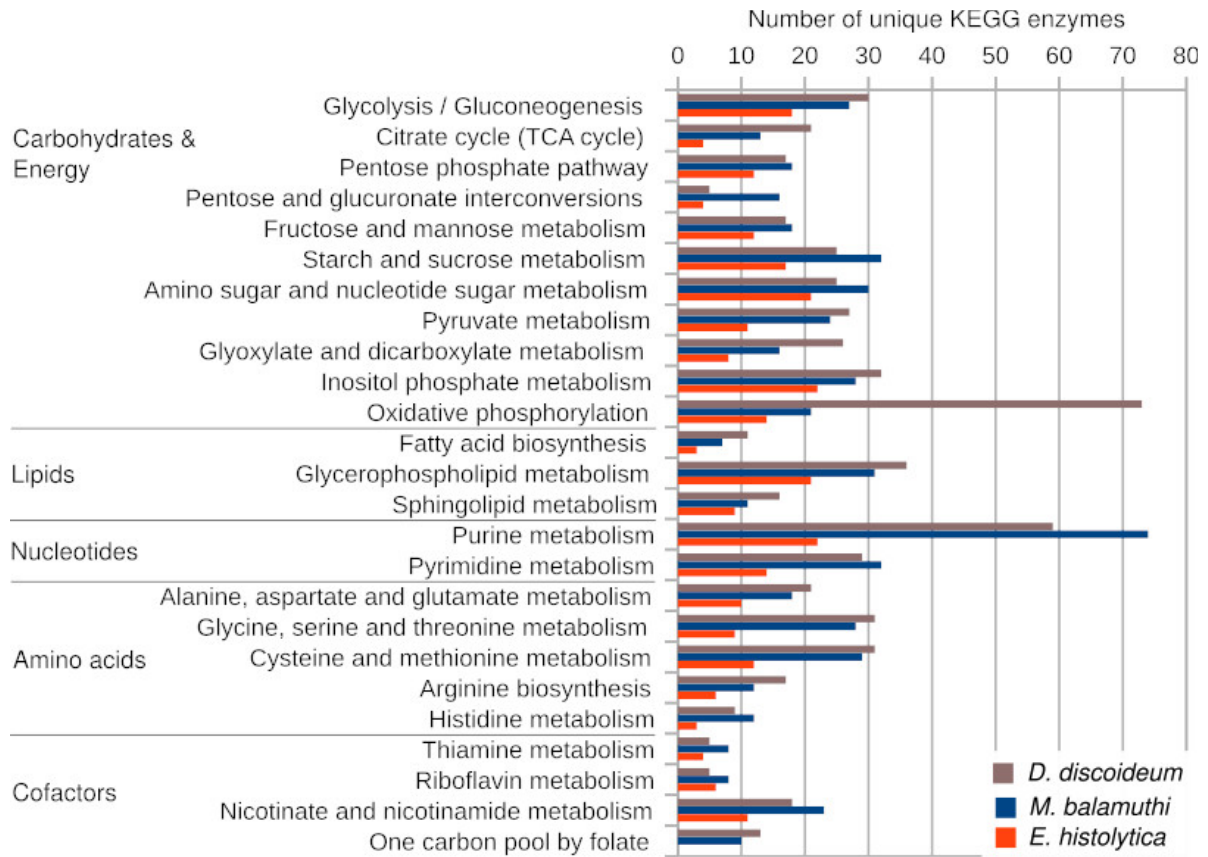


Figure 7. Summary of features representing pre-conditions and adaptations for parasitic and free-living lifestyle in anaerobic niches. Predicted preconditions are in brown, *E. histolytica* specific features are in red, *M. balamuthi* specific features are in green. A-pore, amoebapore/saplip; Chs, chitin synthase; Chase, chitinase; AS, ATP sulphurylase; APSK, adenosine phosphosulfate kinase; IPP, inorganic pyrophosphatase; MCF, mitochondrial carrier; SPLP, saposin-like protein; CL, cysteine proteases cathepsin L; CF, cysteine protease cathepsin F; CP C1, cysteine protease C1 family; Rom, rhomboid protease; Fe-SOD, iron-dependent superoxide dismutase; NADH ox., NADH oxidase; NADPH ox., NADPH oxidase; NFO, NADPH: flavin oxidoreductase; Prxs, peroxiredoxines; Rbr, rubrerythrin; Trx, thioredoxin; TrxR, thioredoxin reductase; FDP, flavodiiron protein; ISF, FeS flavoprotein.

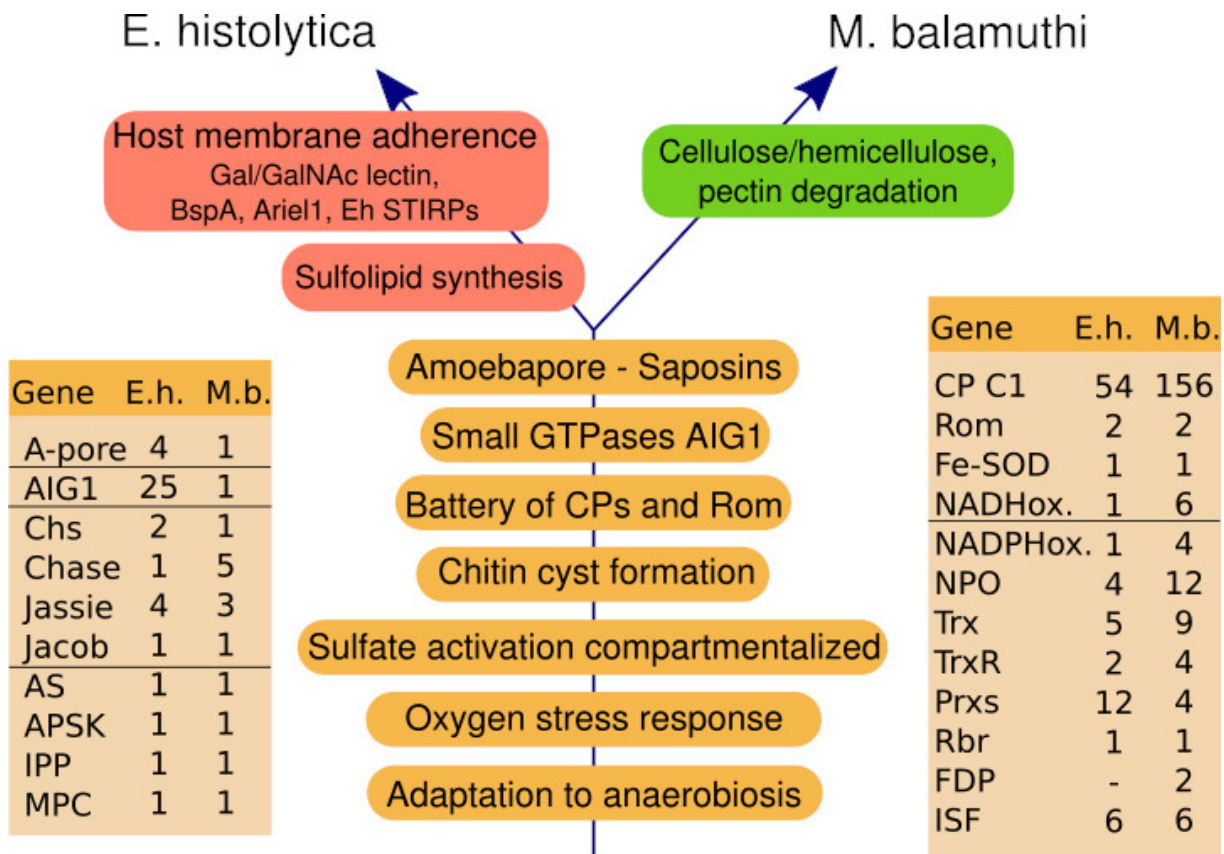


Figure 8. Cysts of *M. balamuthi*. (A) Labelling of *M. balamuthi* cyst by calcofluor white stain. Bar is X μm . (B) Scanning electron microscopy of *M. balamuthi* cyst. Bar is X μm . (C) Comparison of cyst wall components in *M. balamuthi* and *E. invadens* and their predicted structure. The model for *E. invadens* is based on (Chatterjee et al., 2009).

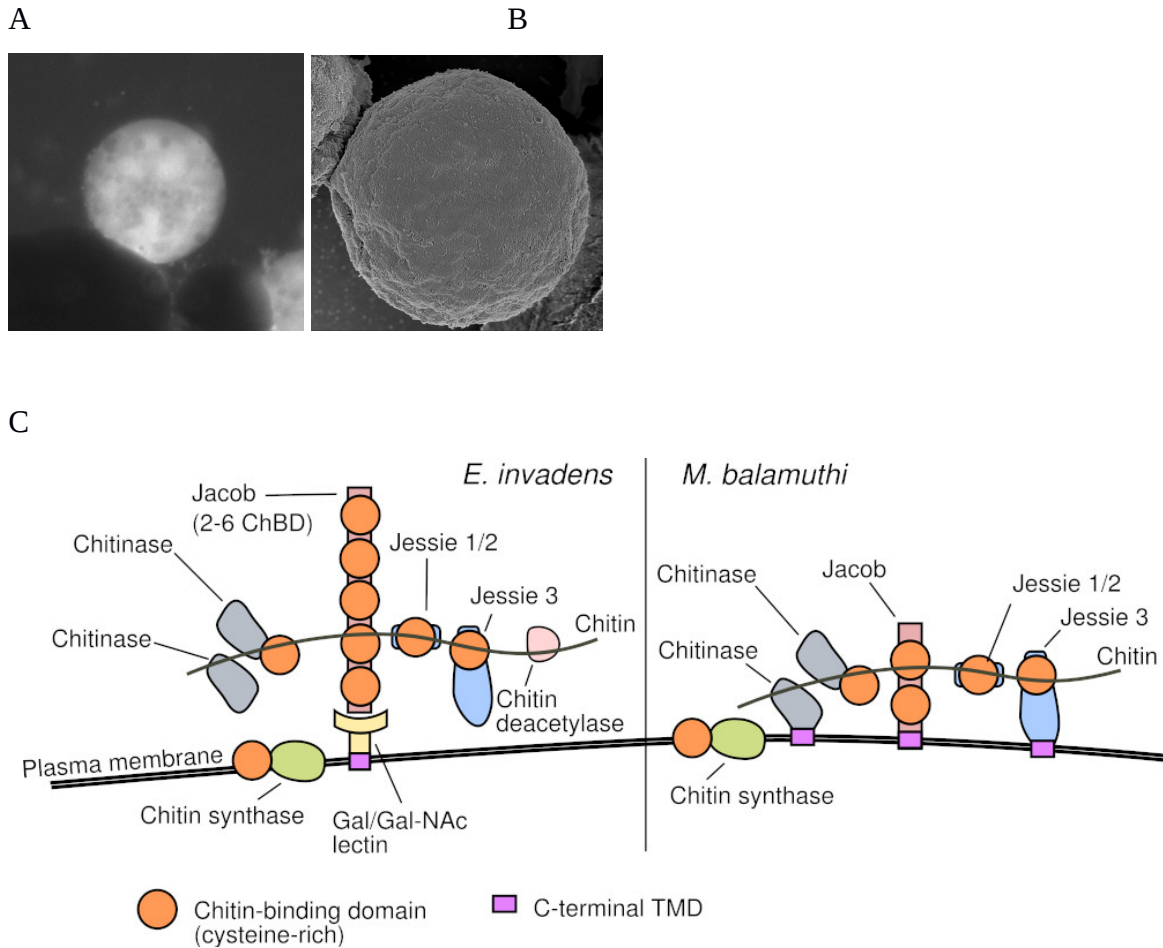
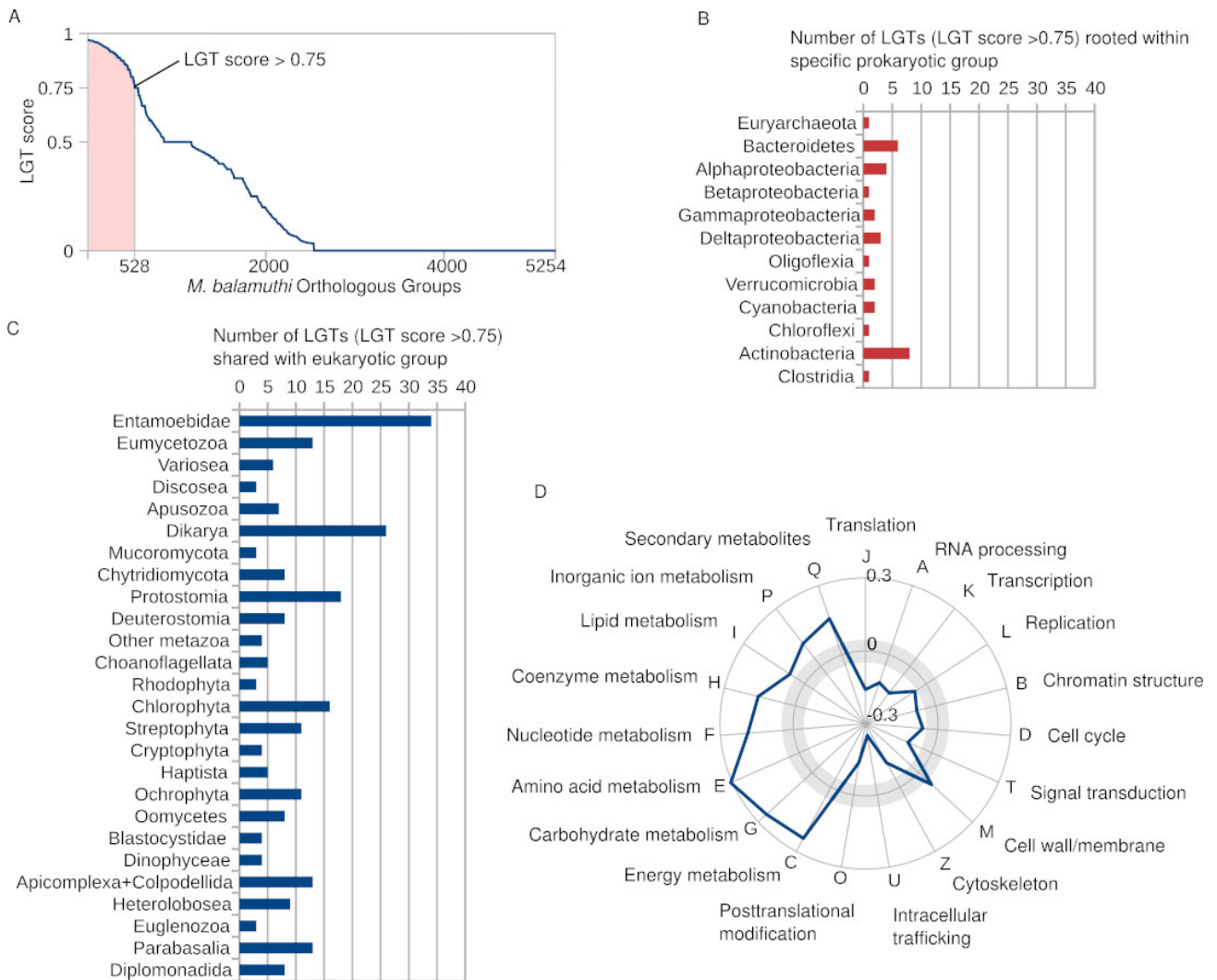


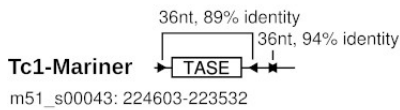
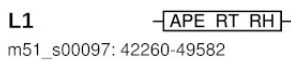
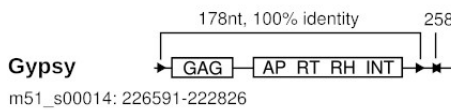
Figure 9. Evaluation of LGT contribution to *M. balamuthi* genome. (A) Distribution of LGT score among *M. balamuthi* orthologous groups. (B) Predicted prokaryotic sources of LGT (for LGT score > 0.75). (C) A number of genes acquired by LGT in eukaryotic groups that are shared with *M. balamuthi* (LGT score > 0.75). (D) Pearson correlation coefficients between the LGT score and COG functional categories. The gray area demarcates correlation coefficient values with low statistical significance (p-value ≥ 0.01).



Supplementary Figures

Fig S1. Analysis of transposable elements in *M. balamuthi*

		Complete	Incomplete	
Retrotransposons	Long terminal repeat (LTR)	Gypsy (RLG)	7	261
		Copia (RLC)	0	1
		unclassified LTR	0	13
	Long interspersed nuclear element (LINE)	L1 (RIL)	6	72
		R2 (RIR)	4	43
		I (RII)	0	7
		unclassified LINE	0	7
	DNA transposon	Tc1-Mariner (DTT)	45	72
		Pif-Harbinger (DTH)	5	14
unclassified DNA transposon		0	328	



TASE - Transposase
 RT - Reverse transcriptase
 APE - Apurinic endonuclease
 RH - RNase H
 GAG - Capsid protein
 AP - Aspartic proteinase
 INT - Integrase
 ▶ - Repeat
 ✱ - Palindrom

Fig S2. Tree R2

Reverse transcriptase domains of LINE and LTR elements from amoebozoans were aligned using MAFFT and a phylogeny was built using FastTree. The leaves are color coded according to the organism.

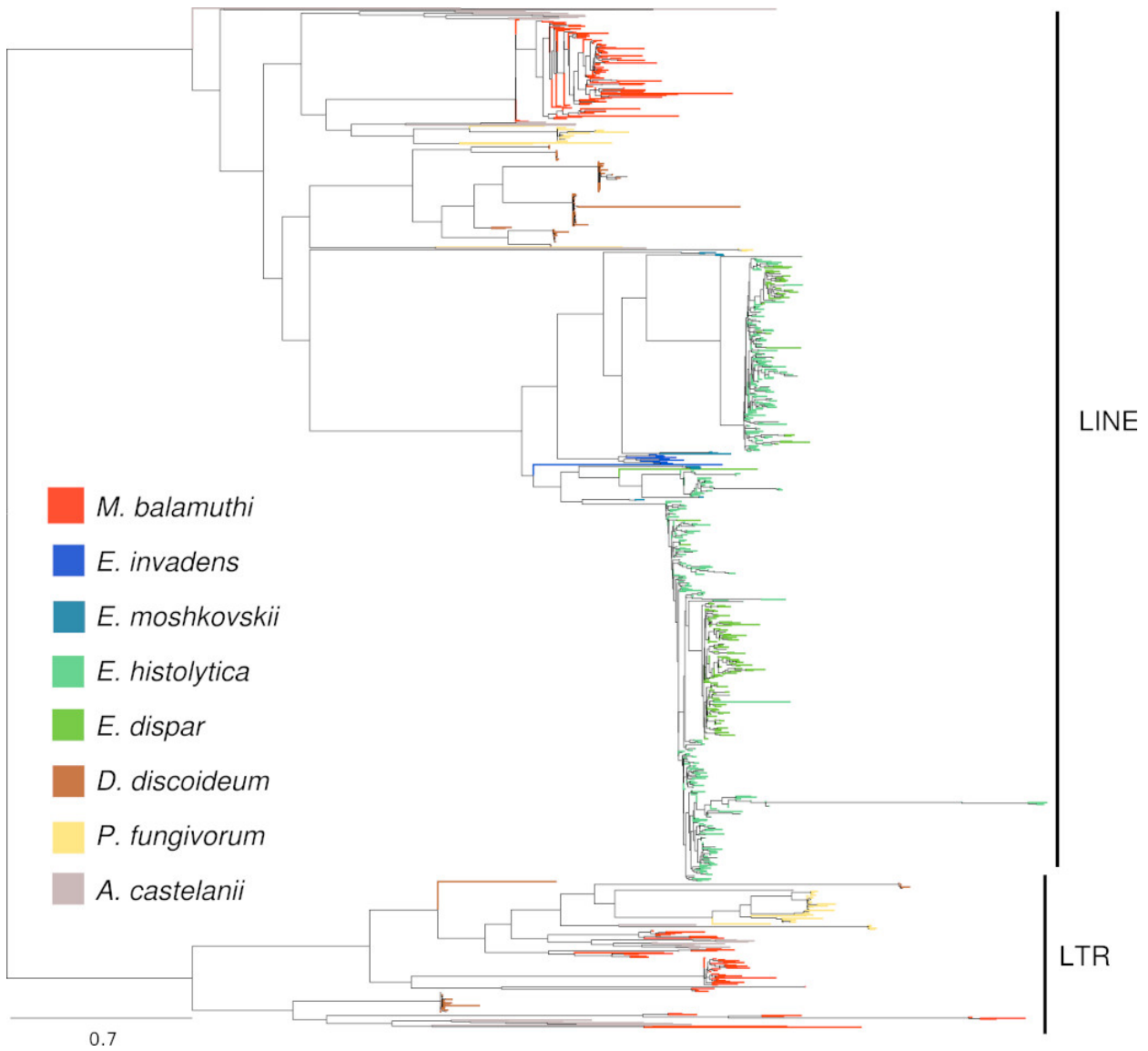
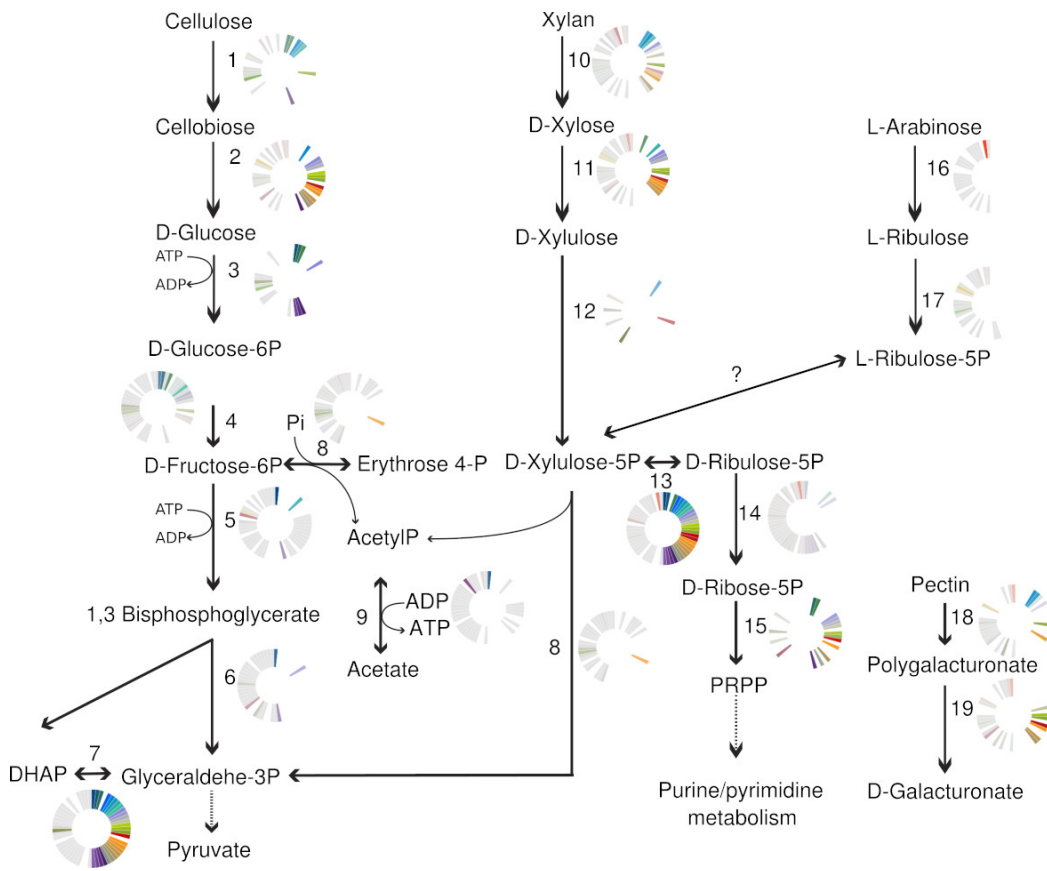


Fig S3. Cell wall degrading enzymes in *M. balamuthi*. (A) Identified enzyme with color-coded circles that represent an excerpt of the *M. balamuthi* gene phylogeny (see B for details). (1) 3.2.1.4 Endoglucanase (m51a1_g3901), 3.2.1.91 Exocellulase (m51a1_g1275); (2) 3.2.1.21 Celobiase (m51a1_g271); (3) Hexokinase (m51a1_g287); (4) Glucose-6-phosphate isomerase (m51a1_g1654); (5) PPI dependent Phosphofructokinase (m51a1_g10351); (6) Aldolase (m51a1_g12673); (7) Triosephosphate isomerase (m51a1_g5662); (8) Bifunctional xylulose 5-phosphate/fructose 6-phosphate phosphoketolase (m51a1_g11684); (9) Acetate kinase (m51a1_g5669); (10) 3.2.1.8. Endo-1,4-beta-xylanase (m51a1_g11098), 3.2.1.37 Exo-1,4-beta-xylosidase (m51a1_g3206); (11) 5.3.1.5 Xylose isomerase (m51a1_g7379); (12) 2.7.1.17 Xylulokinase (m51a1_g8016); (13) 5.1.3.1 D-ribulosephosphate-3-epimerase (m51a1_g1638); (14) 5.3.1.6 Ribose-5-phosphate isomerase (m51a1_g11023); (15) 2.7.6.1 Ribose-phosphate diphosphokinase (m51a1_g4408); (16) 5.3.1.4. L-Arabinose isomerase (m51a1_g3016); (17) 2.7.1.16 Ribulokinase (m51a1_g3020); (18) 3.1.1.11 Pectinesterase (m51a1_g7028); (19) 3.2.1.15 Polygalacturonase (m51a1_g8114); (?) 5.1.3.4 L-ribulose-5-phosphate 4-epimerase, a candidate not identified.

B. A color-coded excerpt of *M. balamuthi* gene phylogeny. The left half shows estimated prokaryotic sources of the gene. The right half shows estimated eukaryotic taxons that are possibly monophyletic with *M. balamuthi* gene. White field means that given prokaryotic or eukaryotic taxons were not included to the analysis. A gray field means there isn't support for given prokaryotic taxon being source of LGT or given eukaryotic taxon doesn't form a eukaryotic monophylum with *M. balamuthi* gene. A colored field means given prokaryotic taxon is supported to be a possible LGT donor or given eukaryotic taxon is probably monophyletic with *M. balamuthi* gene.

A



B

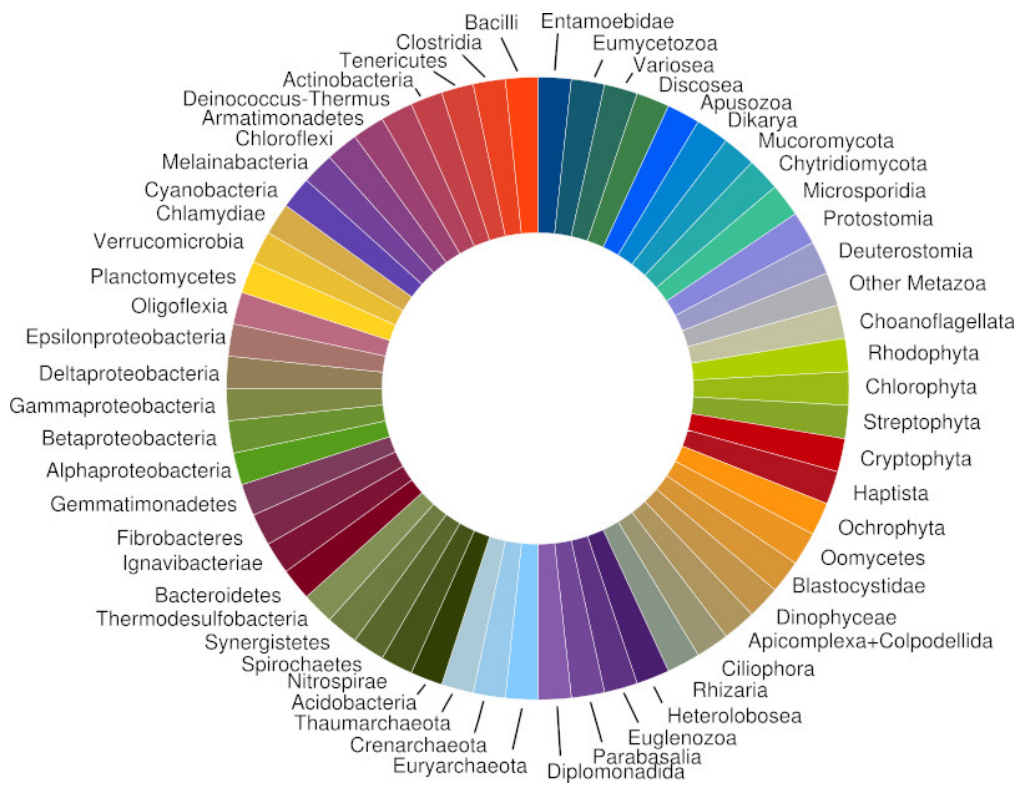


Fig S4. Phylogeny of *M. balamuthi* histidine kinases. A maximum-likelihood tree was constructed with IQ-TREE using 60 protein sequences and 122 sites. Bootstrap support values were calculated using 1000 ultrafast bootstrap replicates.

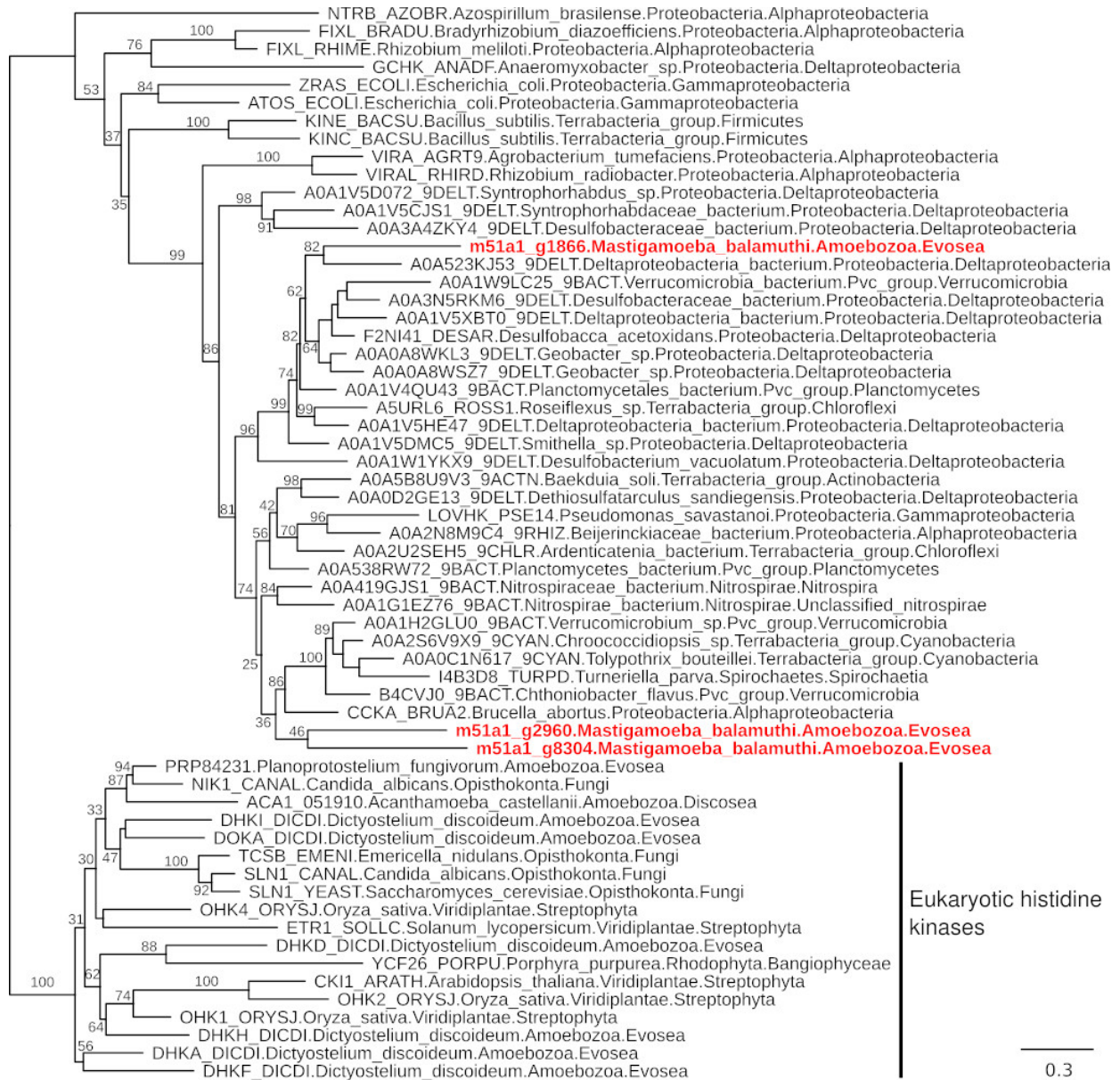


Figure S5. *M. balamuthi* flagellum reconstruction. A. Coulson plot of flagellar proteins found in representatives of the *Conosa* (uniflagelleted cells) and *Chlamydomonas reinhardtii* (biflagelleted cells).

B. Model of *M. balamuthi* flagellum that lack outer dynein arms.

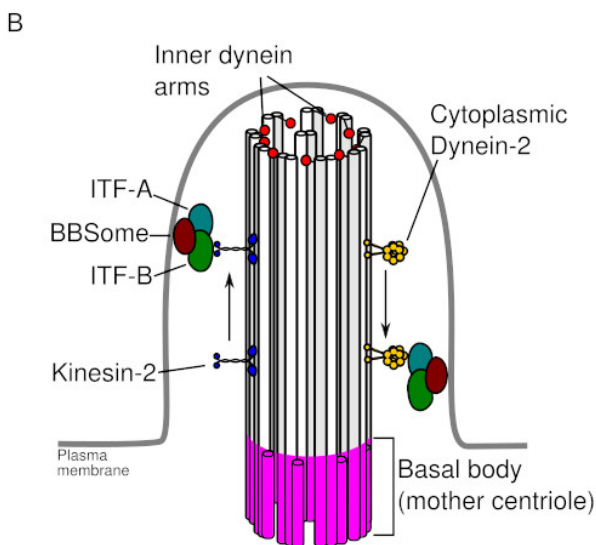
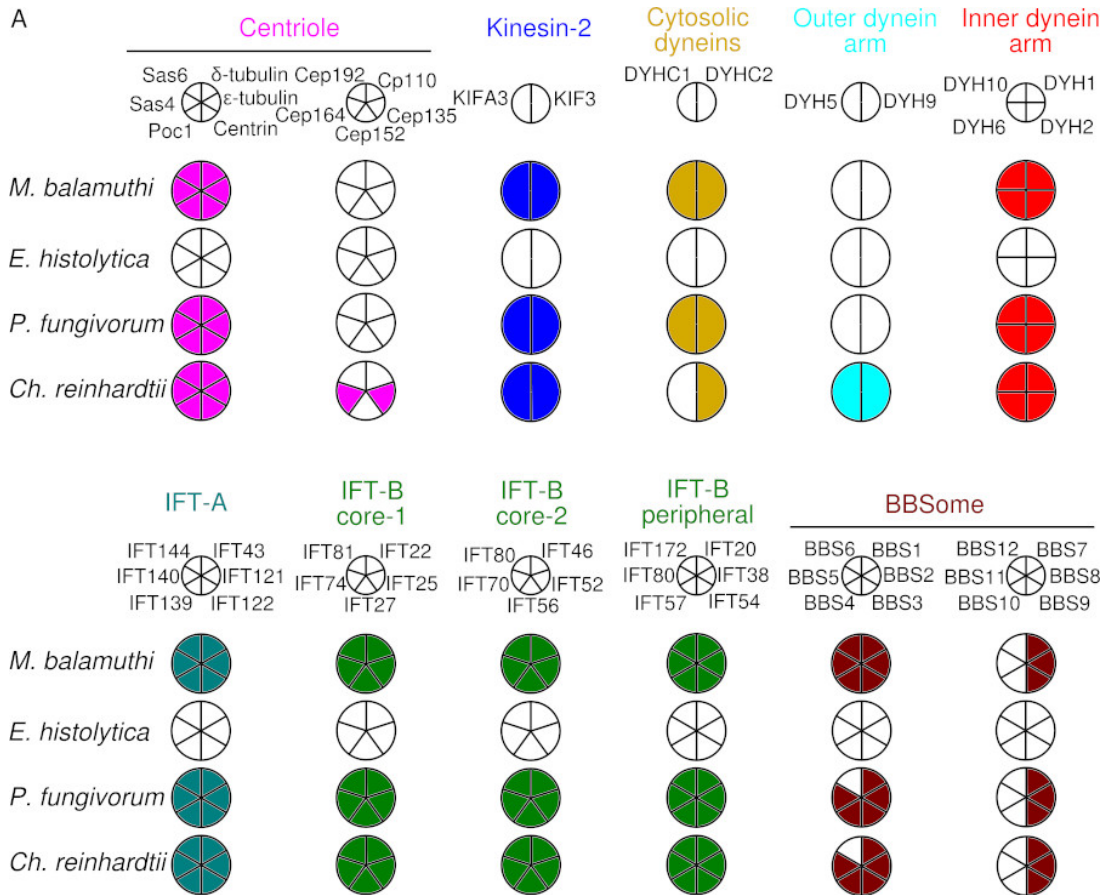


Fig S6. Detection of vesicular transport complexes in Amoebozoa.

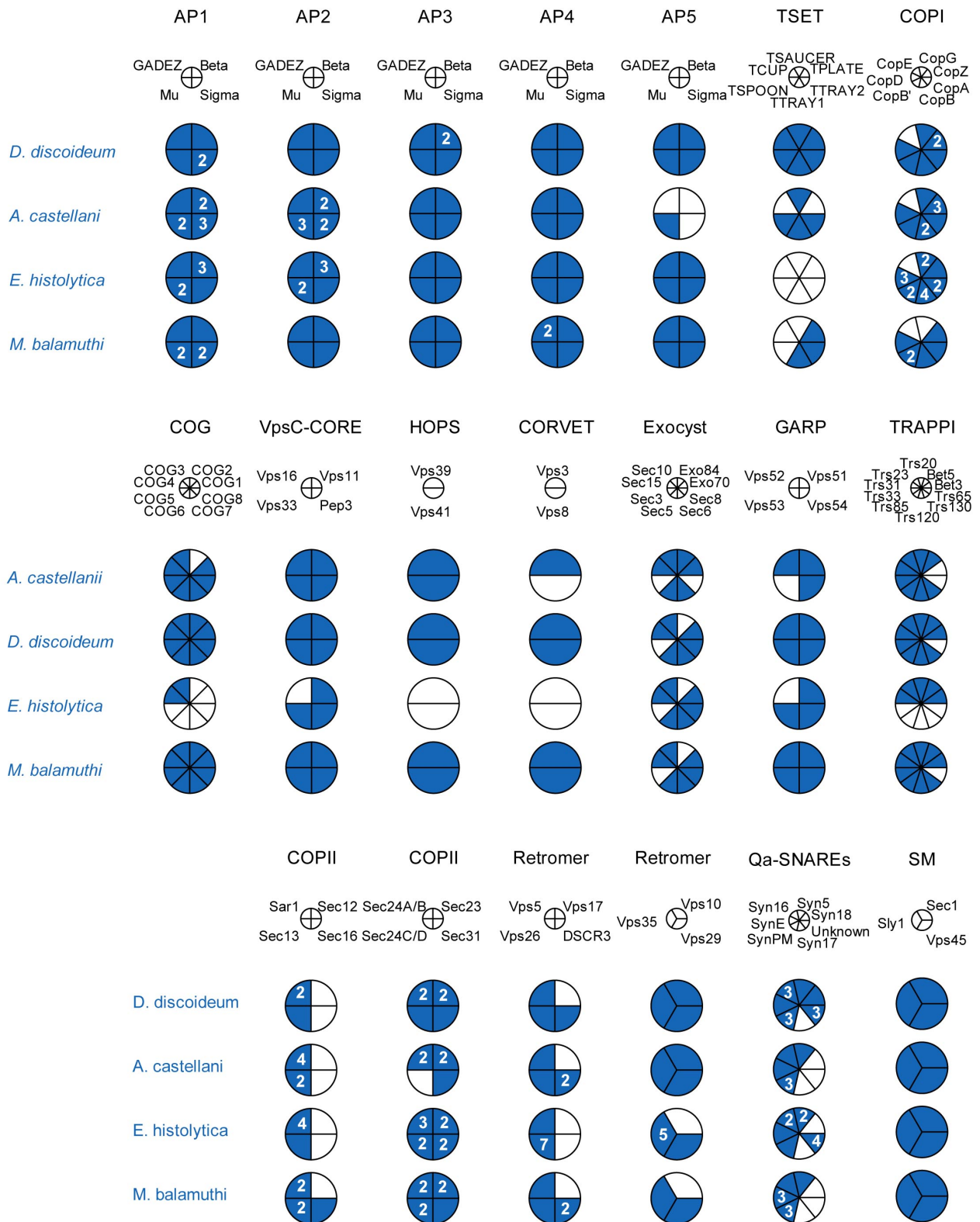







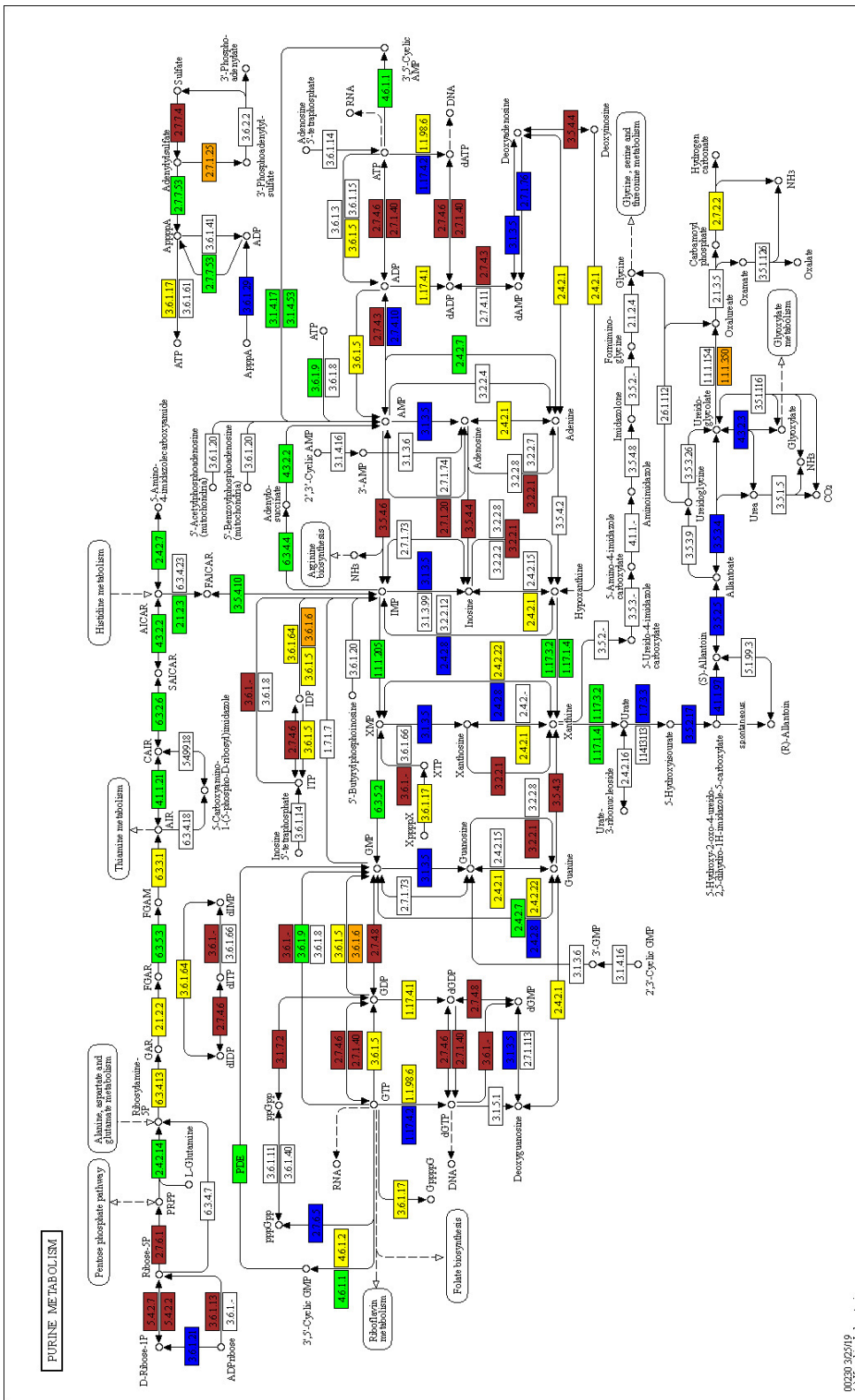


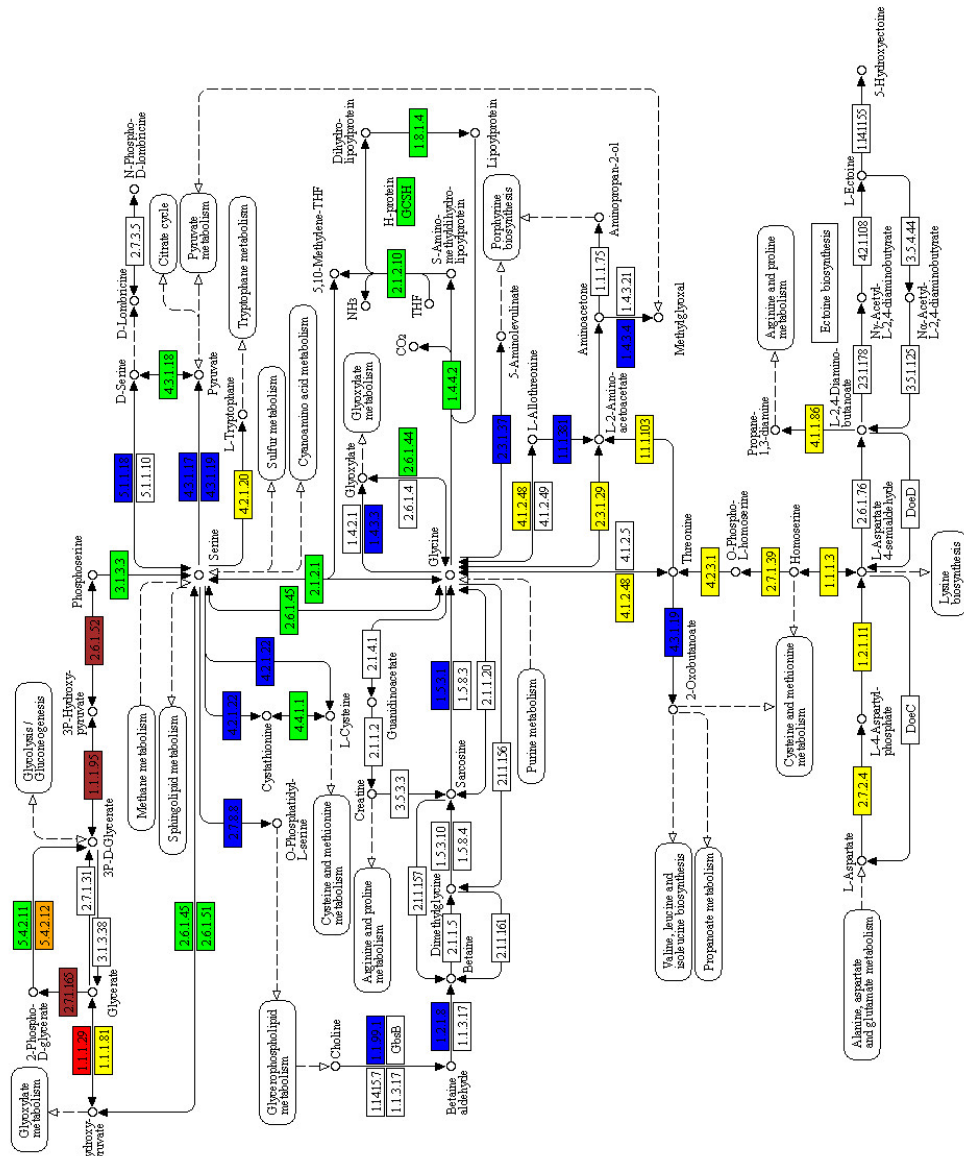
Fig S7. Selected KEEG pathways. KEEG pathways were predicted in predicted proteomes of *D. discoideum*, *M. balamuthi* and *E. histolytica* using KofamKOALA. Here we show selected pathways with color coded enzymes to indicate distribution among these three species. A. Legend. B. Purine metabolism. C. Pyrimidine metabolism. D. Glycine, serine and threonine metabolism. E. Histidine metabolism.

A

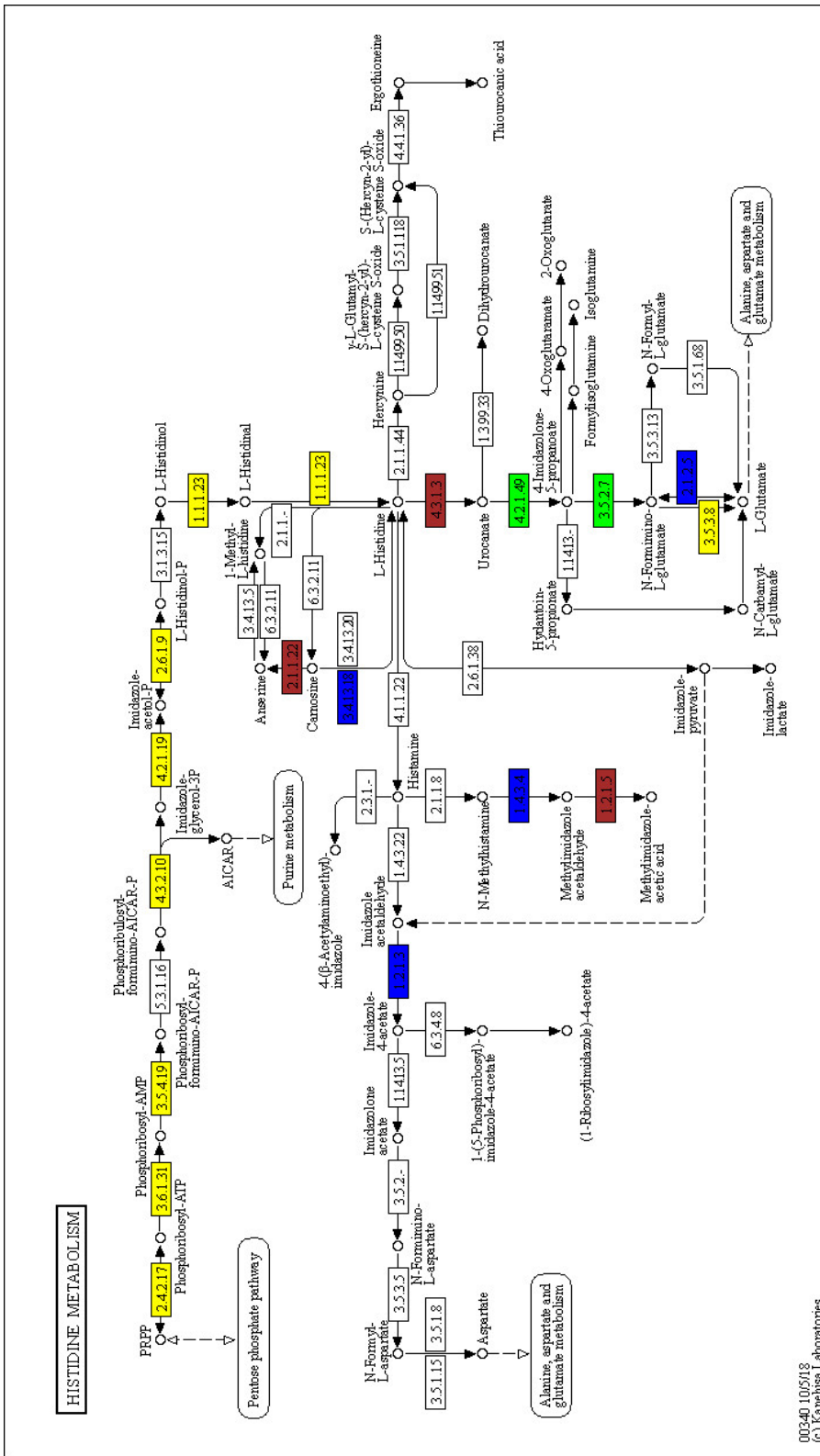
-  *D. discoideum*
-  *M. balamuthi*
-  *E. histolytica*
-  *D. discoideum* + *M. balamuthi*
-  *M. balamuthi* + *E. histolytica*
-  *D. discoideum* + *E. histolytica*
-  *D. discoideum* + *M. balamuthi* + *E. histolytica*



GLYCINE, SERINE AND THREONINE METABOLISM



00360_461119
© Karolinska Laboratories



00340.10/5/18
 (c) Kanelissa Laboratories

Fig S8. Chitin synthesis and degradation in *M. balamuthi*. Color-coded circles next to the enzymes represent an excerpt of the *M. balamuthi* gene phylogeny (see Figure S3 B). (1) Hexokinase (m51a1_g287); (2) Glucose-6-phosphate isomerase (m51a1_g1654); (3) 3.5.99.6 Glucosamine-6P isomerase (m51a1_g13796); (4) 2.3.1.4 Glucosamine 6P N-acetyltransferase (m51a1_g8246); (5) 5.4.2.10 Phosphoglucosamine mutase (m51a1_g1707); (6) 2.7.7.23 UDP-N-acetylglucosamine diphosphorylase (m51a1_g2078); (7) 2.4.1.16 Chitin synthase (m51a1_g11433); (8) 3.2.1.14 Chitinase (m51a1_g3348); (9) 3.2.1.52 Beta-N-acetylhexosaminidase (m51a1_g2284); (10) 2.7.1.59 N-acetylglucosamine kinase (m51a1_g11792); (11) 3.5.99.6 Glucosamine-6-phosphate deaminase (m51a1_g13796).

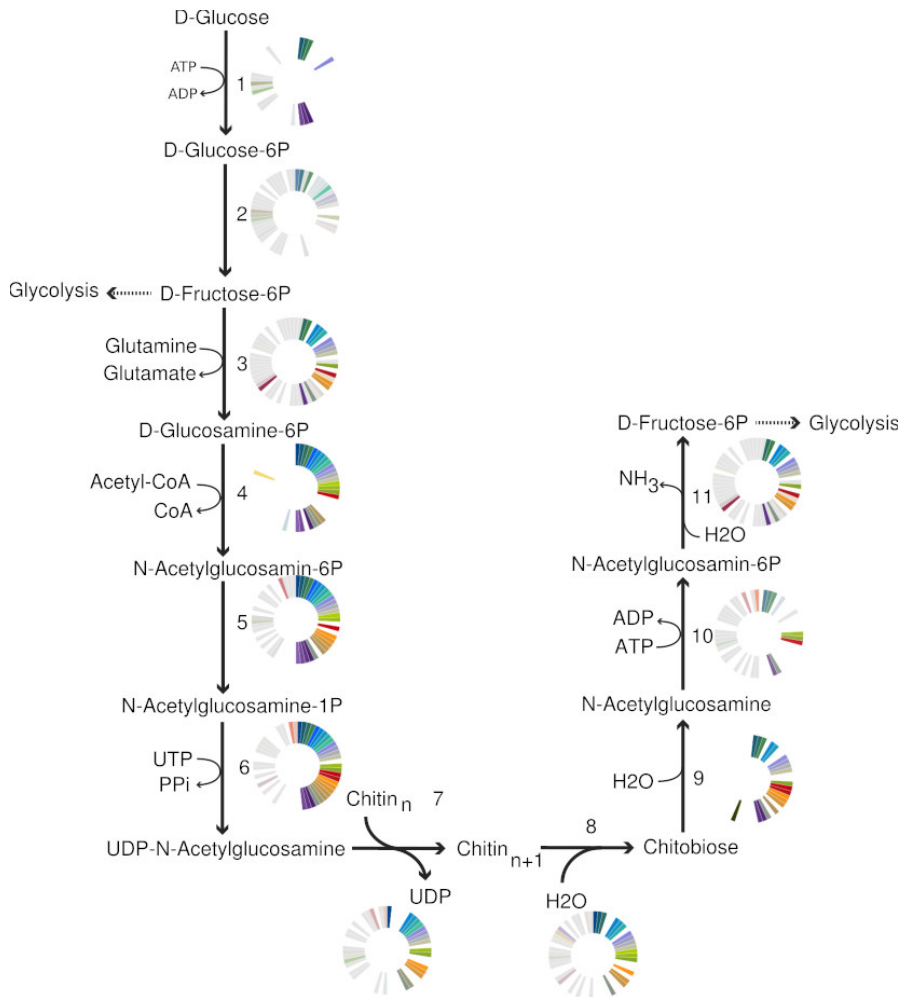


Fig S9. Monophyly of *M. balamuthi* and *Entamoeba* chitin synthase-1. A maximum-likelihood tree was constructed with IQ-TREE using 48 protein sequences and 209 sites. Bootstrap support values were calculated using 1000 ultrafast bootstrap replicates. Branches with high support (bootstrap support value = 100%) are depicted with black circles. Numbers in black triangles indicate a number of analyzed genes.

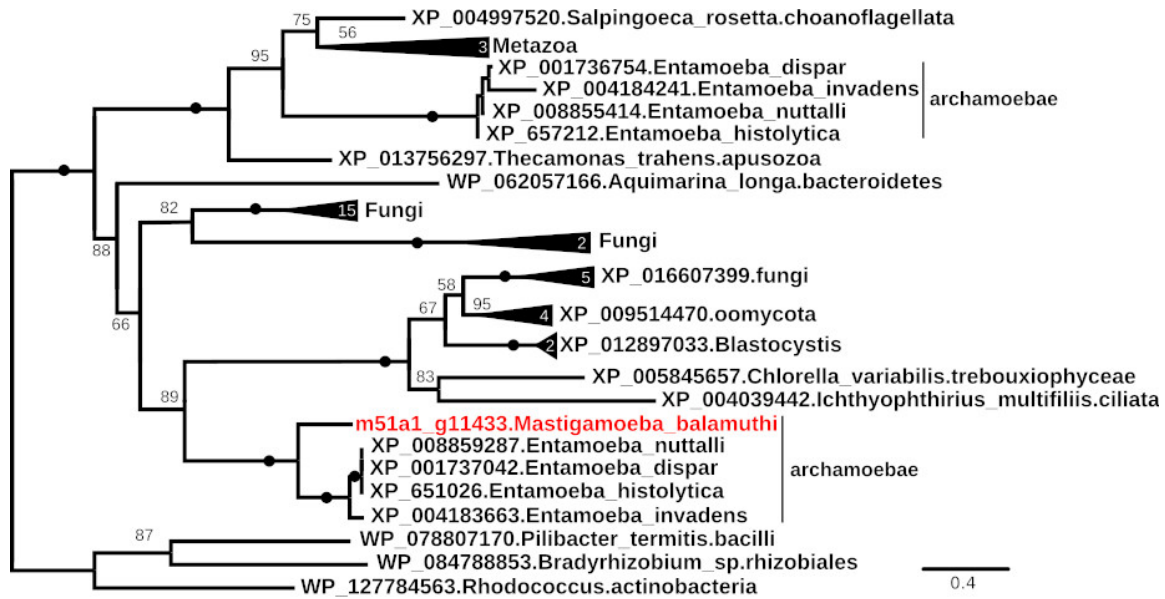


Fig S10. Phylogenetic analysis of catalytic domains of chitinases (glycosyl hydrolases family 18). A maximum-likelihood tree was constructed with IQ-TREE using 82 protein sequences and 119 sites. Bootstrap support values were calculated using 1000 ultrafast bootstrap replicates. Branches with high support (bootstrap support value = 100%) are depicted with black circles. Chitinase structure is indicated: brown circle indicates chitin-binding domain, violet box indicates transmembrane domain, gray structures indicate a catalytic domain.

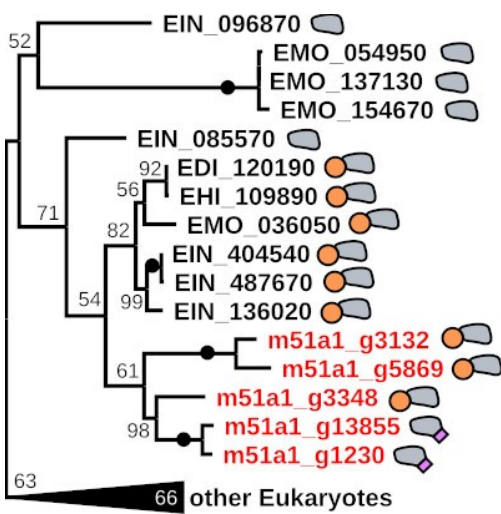
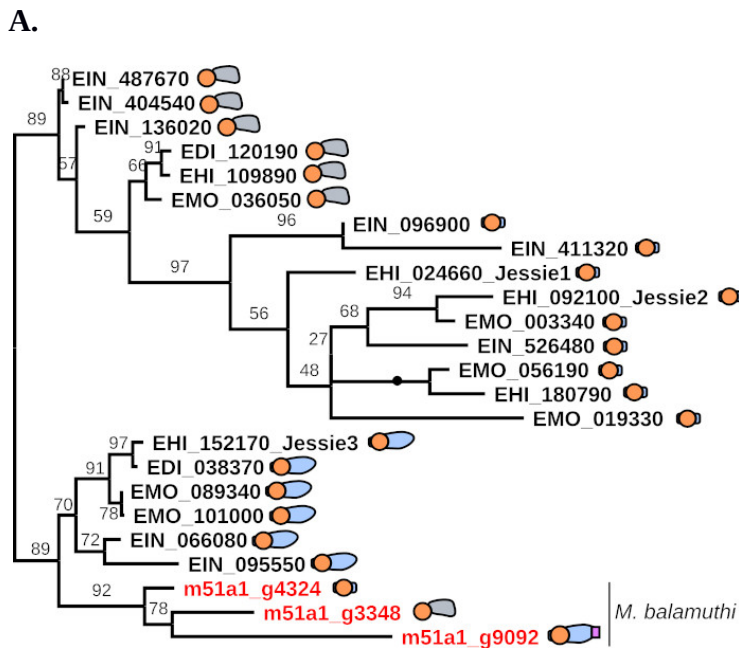
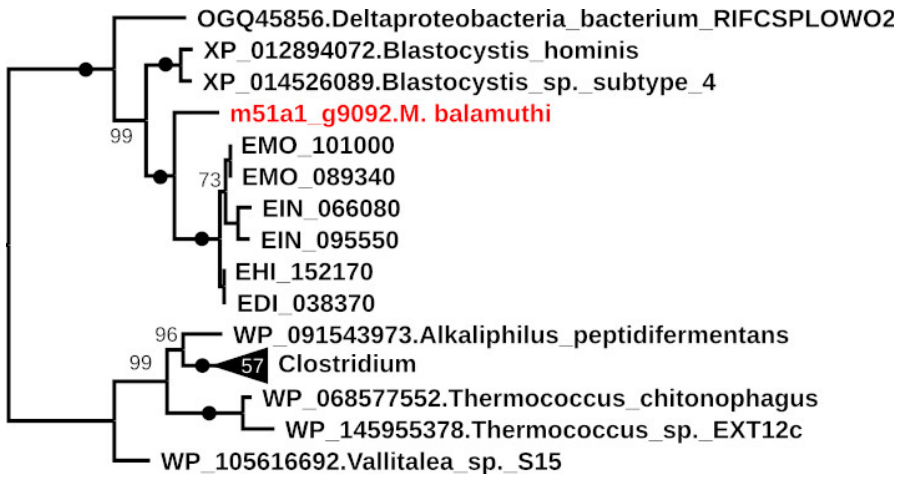


Fig. S11. Comparison of *M. balamuthi* and *Entamoebae* Jessie lectin. (A) Phylogenetic analysis of N-terminal chitin-binding domains. A maximum-likelihood tree was constructed with IQ-TREE using 24 protein sequences and 69 sites. Bootstrap support values were calculated using 1000 ultrafast bootstrap replicates. Jessie structure is indicated: brown circle indicates N-terminal chitin-binding domain, violet box indicates transmembrane domain, gray structure indicates Jessie specific domain of unknown function, blue structure indicates Jessie-3 specific domain with homology to a prokaryotic chitinases. (B) Phylogeny of C-terminal domains of Jessie-3. A maximum-likelihood tree was constructed with IQ-TREE using 71 protein sequences and 322 sites. Bootstrap support values were calculated using 1000 ultrafast bootstrap replicates. A number in black triangle indicates the number of sequences of *Clostridium* genera in the analysis. (C) Protein sequence alignment of partial *M. balamuthi* and *Thermococcus chitinophagus* chitinase. Conserved catalytic amino acid residues are marked by red circles (Nishitani et al., 2018).



B



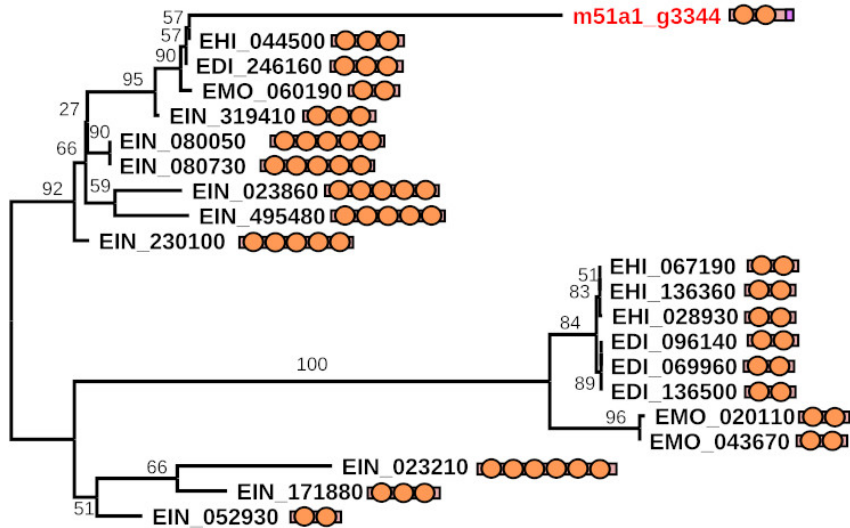
C

EhJessie3	230	NPTQIWIGRPSQSVTISYTPGVAIRLPDKYYGMFLGYAMDAYGLNPGMLIGLGAKESFSF	289
		+PTQI IG ++ +P+KY+ M L A ++ +NP L+ L AKE++	
5XSV	158	HPTQIIIGLADIDYPLNLASSARMWVPNKYFAMGLALAYEWFKVNPNFLMALAAKENWGT	217
EhJessie3	290	ARYQATDDGSYFIVANENEHYDCYSSSQRGLCRDGNLDGPFQVETGGMSTDVAILPNRFW	349
		A + Y ++ +E E+Y + + DG FQVE+G + A P+ F	
5XSV	218	AVTKDPAFKGYKVIIDEEYY-----WPVQIDHPDGIFQVESGNFNQIKAYYPDIF-	268
EhJessie3	350	IGESTTPKSERKIKYMFDFNEVLTYAGFRAYHDYFTLNYGRAFVLTSDFHFRHNLVMMMK	409
		P + YM + D + + ++ ++ L+	
5XSV	269	-----PDTADHDDYM-----KVSLDPNDTAWITSPIVAAVSLTMERELLY---	308
EhJessie3	410	KVGLRDALSKRTKREQRDSLEFATAMYTYNRGVFDTQLIQMLN-----GCNADMDPCTDC	464
		+ D ++ + + E + YNRGV + +++ + NA++ +	
5XSV	309	-AAVGDKYNEFLRLAKDPWAETEIIDFGYNRGVGAIEALKIFSDNWEKAINAEV-LWKEF	366
EhJessie3	465	KLDGYGGHTTDIRIVCKAVDSVPNKEVYDYNLSKEDVEYFISTIYSTY----PFNNVDWK	520
		++G+GGH + + +D + + +YD NL+ +D+EYF + ++ + ++ +W	
5XSV	367	NMEGFGGHVPTVINITATMD-METERIYDANLTWDDIEYFFTVVRQKFFRPGAISDEEWN	425
EhJessie3	521	TLRSVDVTAYDYLKSQRNSNAISFRYDWRLLAVIRMHLP	560
		+ DV AYD L + IS+RYD+ T+L V H P	
5XSV	426	AMMRDVKRAYDLLSQHWGGDHISYRYDFLTILRVAMKHWP	465

● proposed catalytic residues

Fig S12. Analysis of Jacob lectins. A. Phylogenetic analysis of C-terminal chitin-binding domains. A maximum-likelihood tree was constructed with IQ-TREE using 21 protein sequences and 57 sites. Bootstrap support values were calculated using 1000 ultrafast bootstrap replicates. Jacob structure is depicted: brown circle indicates chitin-binding domain (CBD), violet box indicates transmembrane domain in *M. balamuthi* sequence (m51a1_g3344). B. Protein sequence alignment of CBD sequences in *M. balamuthi* (m51a1_g3344), and *Entamoeba* species. EHI, *E. histolytica*; EDI, *E. dispar*; EIN, *E. invadens*. Conserved cysteine residues involved in chitin binding are indicated in red.

A



B

<i>m51a1_g3344</i>	AKC	APADD	GLFC	YPGVGH	DNY	YLEC	RPGM	QSGWRK	CAVG	TQC	-	-	TVAG	KHDQN	PCRR								
<i>EHI_044500</i>	FQC	-	-	TQD	GLFC	VIDGVHN	KQYYEC	ATTFK	-	GFRPC	ASG	TWC	QGEQ	EGPYQYN	PCVW								
<i>EDI_246160</i>	FQC	-	-	TQD	GLFC	VIDGVHN	KQYYEC	ATTFK	-	GFRPC	APG	TWC	QGEQ	EGPYQYN	PCVW								
<i>EMO_060190</i>	FNC	-	-	TKD	GLFC	VIDG	IHN	KQYYEC	AA	TFK	-	GFRPC	APG	TWC	QGEQ	EDPYQYN	PCVW						
<i>EIN_319410</i>	FVC	-	-	KED	GLFC	VIDG	IHN	NQYYEC	AN	TFK	-	GFRPC	APG	TWC	QGEQ	EKPYGYN	PCVW						
<i>EIN_080050</i>	LKC	-	-	KED	GLFC	VIDG	KHN	DQYFQC	SEFY	T	-	GFRPC	SKG	TWC	KGEQ	EKPYD	TD	PCVW					
<i>EIN_080730</i>	LKC	-	-	KED	GLFC	VIDG	KHN	DQYFQC	SEFY	T	-	GFRPC	SKG	TWC	KGEQ	EKPYD	TD	PCVW					
<i>EIN_023860</i>	VEC	-	-	KKD	GF	FL	IDG	KHN	DQYFQC	SS	EF	T	-	GFRQC	PKD	TWC	KKEQ	SEPF	EES	PCVW			
<i>EIN_495480</i>	FKC	-	-	RGL	GLFC	I	IDG	IHN	DQYYQC	SEQFN	-	GTRHC	PKN	TWC	KGEQ	KLPY	TES	PCVF					
<i>EIN_230100</i>	LQC	-	-	KQD	GLYC	V	IDG	KHN	EMY	YQC	SDF	FN	-	GFRAC	PKG	TWC	KGEQ	DKPY	ETN	PCVW			
<i>EHI_067190</i>	VNC	TEVKE	GYC	YEEES	KN	TEYYWC	VNSVG	-	YEMKC	PNG	TTC	-	-	HTKD	IGYNN	PCY	-						
<i>EHI_136360</i>	VNC	TEVKE	GYC	YEEES	KN	TEYYWC	VNSVG	-	YEMKC	PNG	TTC	-	-	HTKD	IGYNN	PCY	-						
<i>EHI_028930</i>	VNC	TEVKE	GYC	YEEES	KN	TEYYWC	VNNVG	-	YEMKC	PNG	TTC	-	-	HTKD	IGYNN	PCY	-						
<i>EDI_096140</i>	VNC	TEVKE	GYC	YEEES	KN	TEYYWC	VNSVG	-	YEMKC	PKG	TTC	-	-	HTKD	NGYNS	PCY	-						
<i>EDI_069960</i>	VNC	TEVKE	GYC	YEEES	KN	TEYYWC	VNSVG	-	YEMKC	PKG	TTC	-	-	HTKD	NGYNS	PCY	-						
<i>EDI_136500</i>	VNC	TEVKE	GYC	YEEES	KN	TEYYWC	VNSVG	-	YEMKC	PKG	TTC	-	-	HTKD	NGYNS	PCY	-						
<i>EMO_020110</i>	VDC	SKVEN	GYC	YEGEG	NEN	RY	YWC	SNGVG	-	YEMQC	ASG	TVC	-	-	HTREN	PIEN	PCGY						
<i>EMO_043670</i>	VDC	SKVEN	GYC	YEGEG	NEN	KY	YWC	SNGVG	-	YEMQC	SSG	TVC	-	-	HTREN	PIEN	PCGY						
<i>EIN_023210</i>	DIC	-	-	EKD	GYMC	IMDG	THDKY	YYVC	SNSYE	-	GFLPC	PKGER	C	NGK	KYMN	FSEN	PC	EV					
<i>EIN_171880</i>	DIC	-	-	IAD	GFMC	VNDGL	HNQ	FYMC	SSGYK	-	GF	MNC	P	TGAK	C	SGA	I	LMP	F	TQ	S	PC	VT
<i>EIN_052930</i>	GIC	-	-	QSD	GYC	VNDGE	HN	DQYFMC	SDS	FK	-	GFLKC	PRG	TKC	GNV	LNN	PCH	TN	PC	VN			

Fig S13. Phylogenetic analysis of mitochondrial carrier proteins (MPC). A maximum-likelihood tree was constructed with IQ-TREE using 195 protein sequences and 273 sites. Bootstrap support values were calculated using 1000 ultrafast bootstrap replicates.

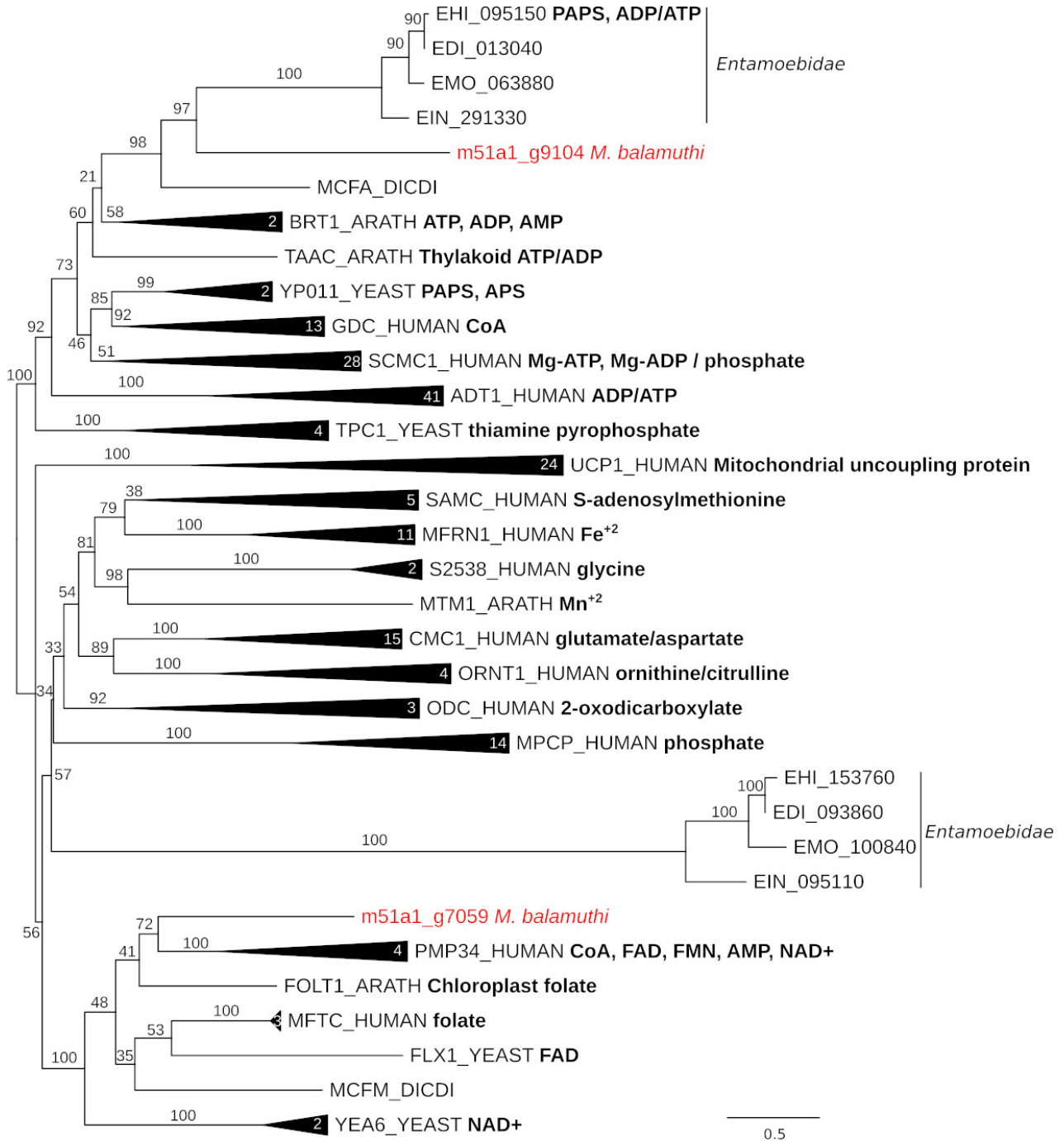


Fig S14. Phylogeny of sulfotransferases. A maximum-likelihood tree was constructed with IQ-TREE using 59 protein sequences and 342 sites. Bootstrap support values were calculated using 1000 ultrafast bootstrap replicates.

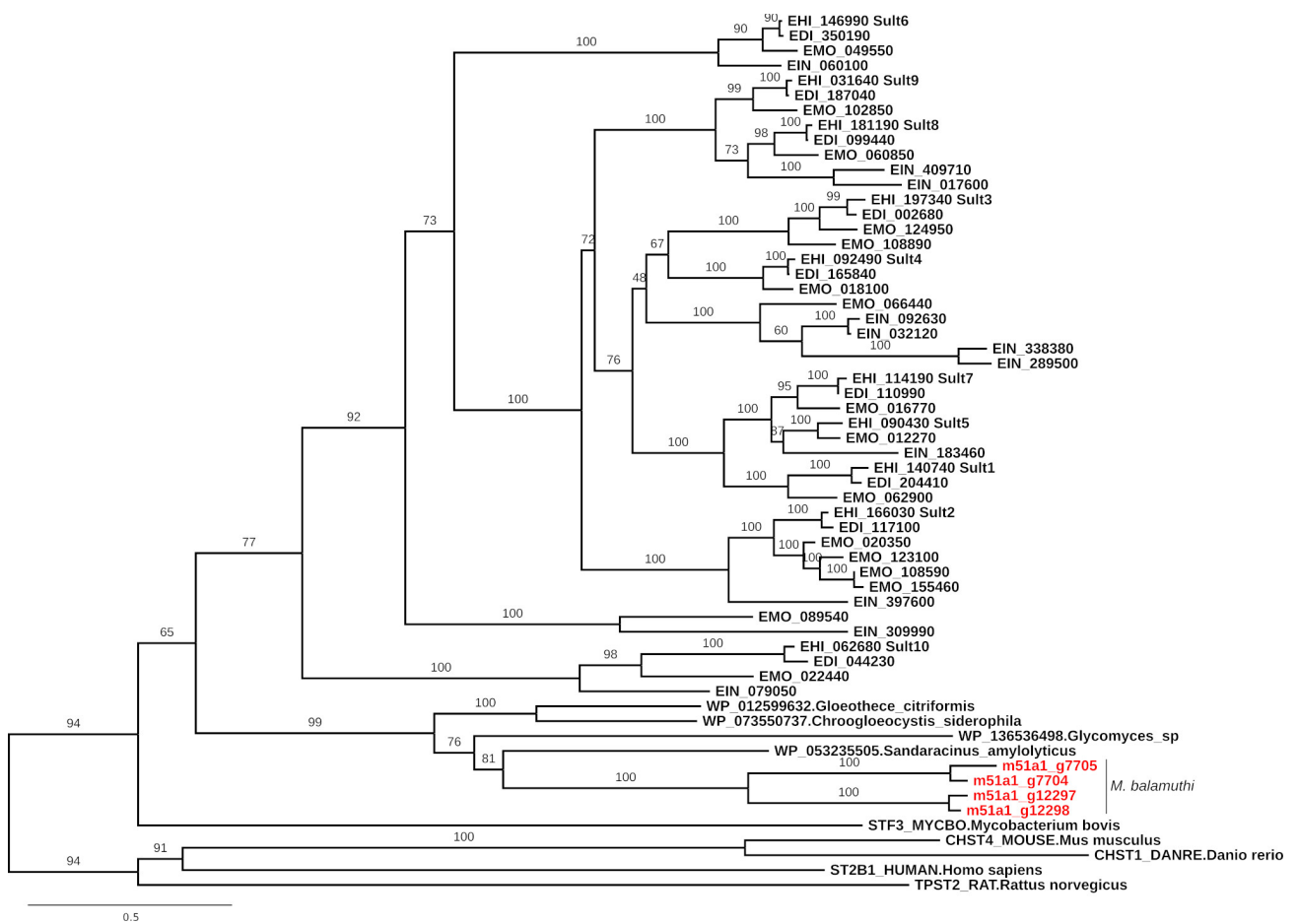


Fig S15. Phylogeny of sulfatases. A maximum-likelihood tree was constructed with IQ-TREE using 52 protein sequences and 361 sites. Bootstrap support values were calculated using 1000 ultrafast bootstrap replicates.

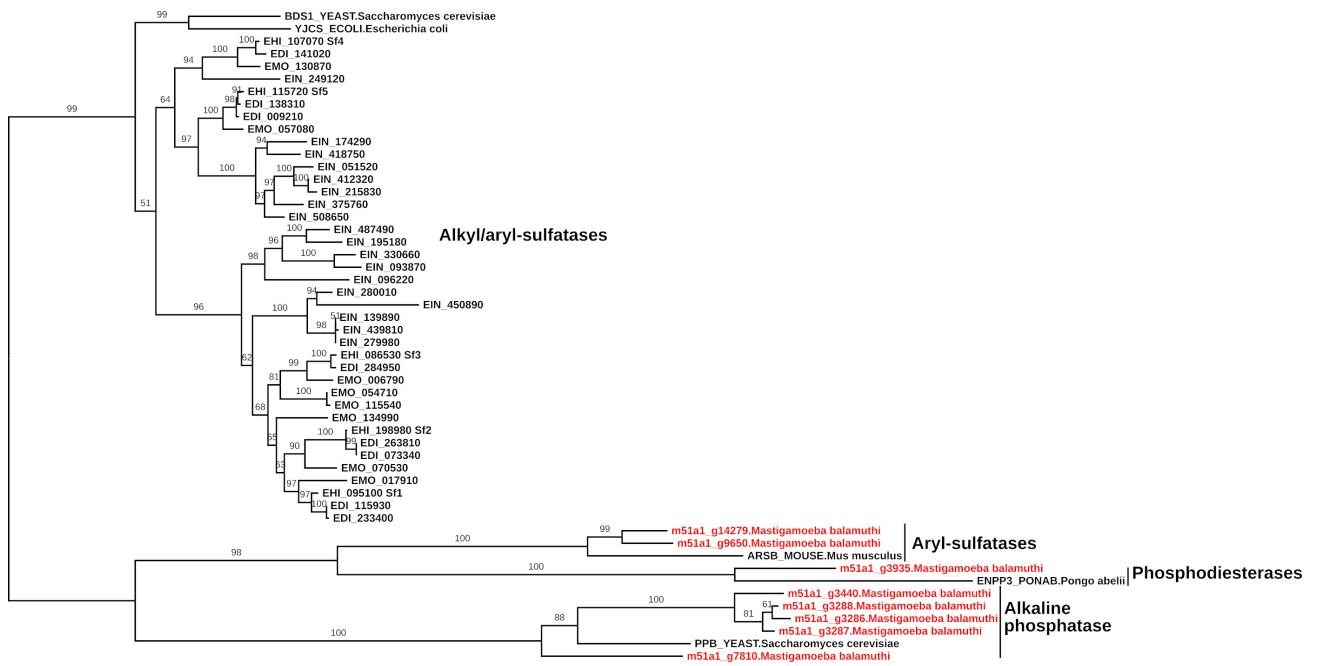
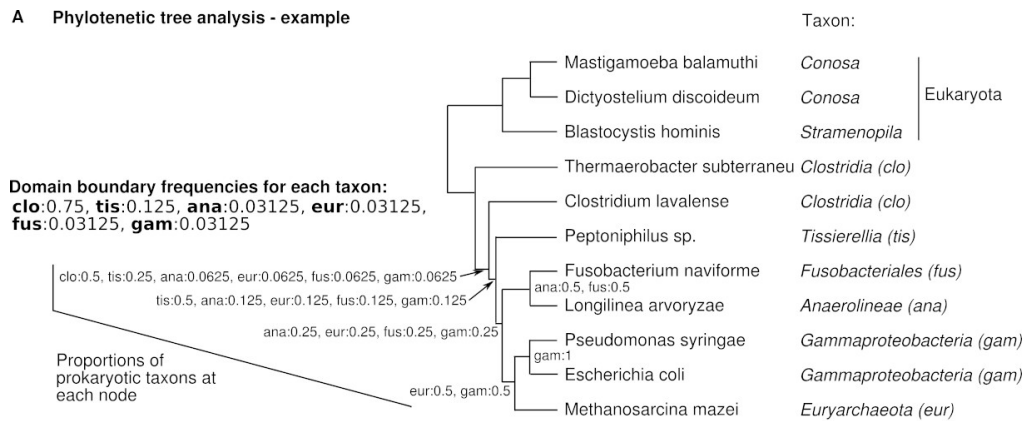


Fig S16. Phylogeny of cysteine proteases. A maximum-likelihood tree was constructed with IQ-TREE using 570 protein sequences and 121 sites. Bootstrap support values were calculated using 1000 ultrafast bootstrap replicates. Concatenated branches of *M. balamuthi* and *Entamoeba* are in yellow and pink respectively. Numbers of paralogs for each *Entamoeba* species are shown. EHI, *E. histolytica*; EDI, *E. dispar*; EMO, *E. moshkovskii*; EIN, *E. invadens*.

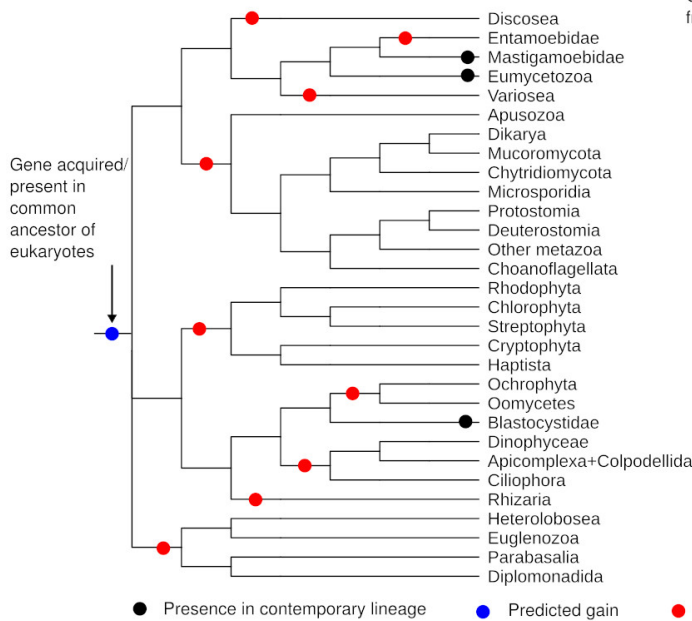


Figure S17. *Domain Boundary Frequency* (DBF) value. A. An example phylogenetic tree. B. Count of events in scenario when the gene is an ancestral eukaryotic gene. C. Count of events in scenario allowing for a late LGT from prokaryotes and further LGTs between eukaryotic lineages. D. Calculation of Domain boundary frequencies for prokaryotic taxons (possible sources of LGT) and LGT score.

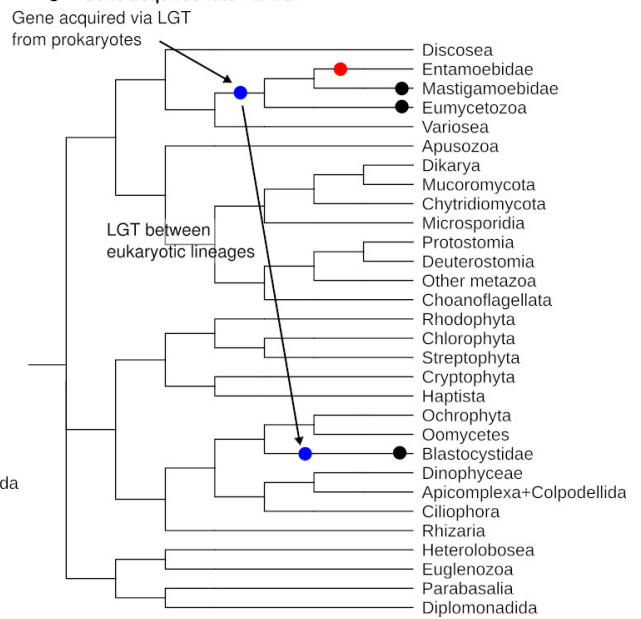
A Phylogenetic tree analysis - example



B Ancestral eukaryotic gene



C Gene acquired late via LGT



D LGT score calculation:

N Euk. taxons = 2

N Prok. taxons = 6

$$\text{LGT prokaryotic coeff} = \frac{N_{\text{proktaxons}}}{N_{\text{proktaxons}} + N_{\text{euktaxons}}} = 6/(6+2) = 0.75$$

Losses if the gene was present in common ancestor of Eukaryotes (N=9):

Discosea, Variosea, Entamoebidae, Obazoa, Archaeplastida+Cryptophyta+Haptista, Ochrophyta+Oomycetes, Alveolata, Rhizaria, Excavata

Minimum number of additional transfers in case of LGT (N=1):

Conosa -> Blastocystidae (or vice versa)

Minimum number of losses in case of LGT (N=1):

Entamoebidae

$$\text{LGT eukaryotic coeff} = \frac{N_{\text{ancestralLosses}}}{N_{\text{ancestralLosses}} + N_{\text{newTransfers}} + N_{\text{newLosses}}} = 29/(9+1+1) = 0.82$$

$$\text{LGT score} = \text{minimum}(\text{LGT prokaryotic coeff}, \text{LGT eukaryotic coeff}) = \text{minimum}(0.75, 0.82) = 0.75$$

Figure S18. Proportion (in percent) of prokaryotic taxa at the prokaryota/eukaryota boundary of *M. balamuthi* genes with high LGT score (> 0.75)

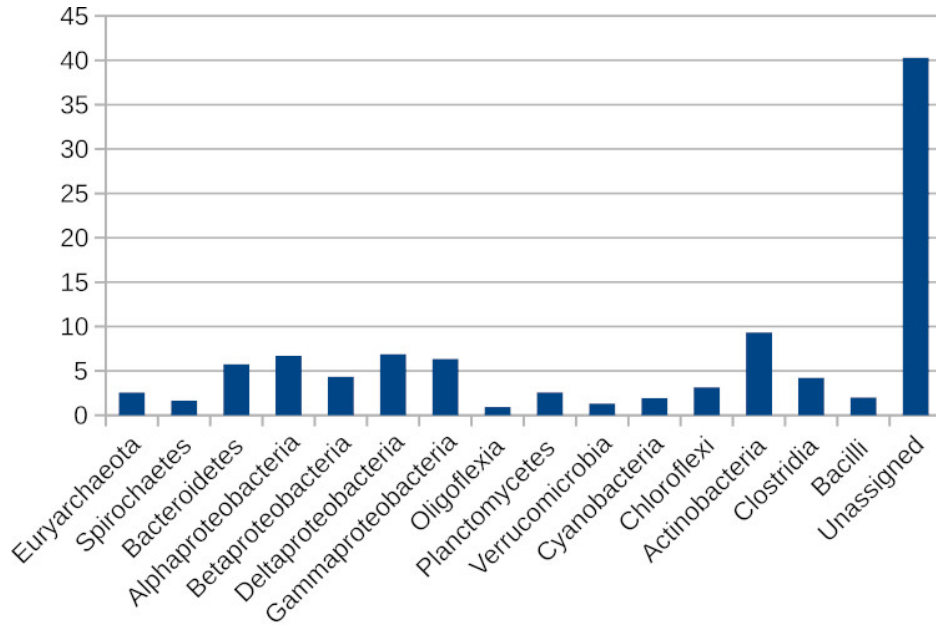
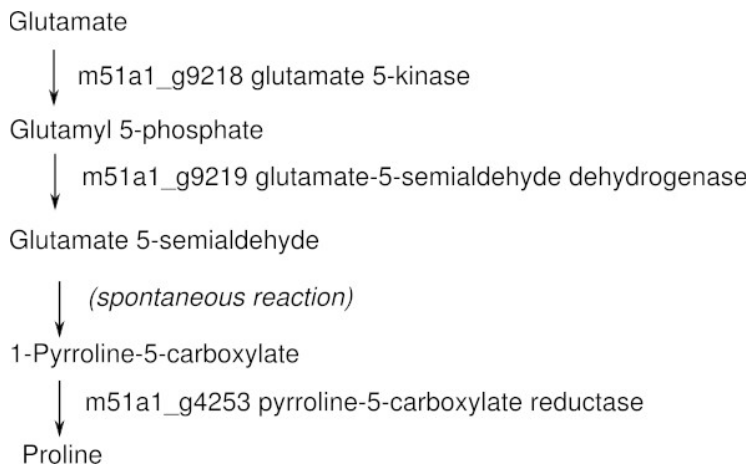
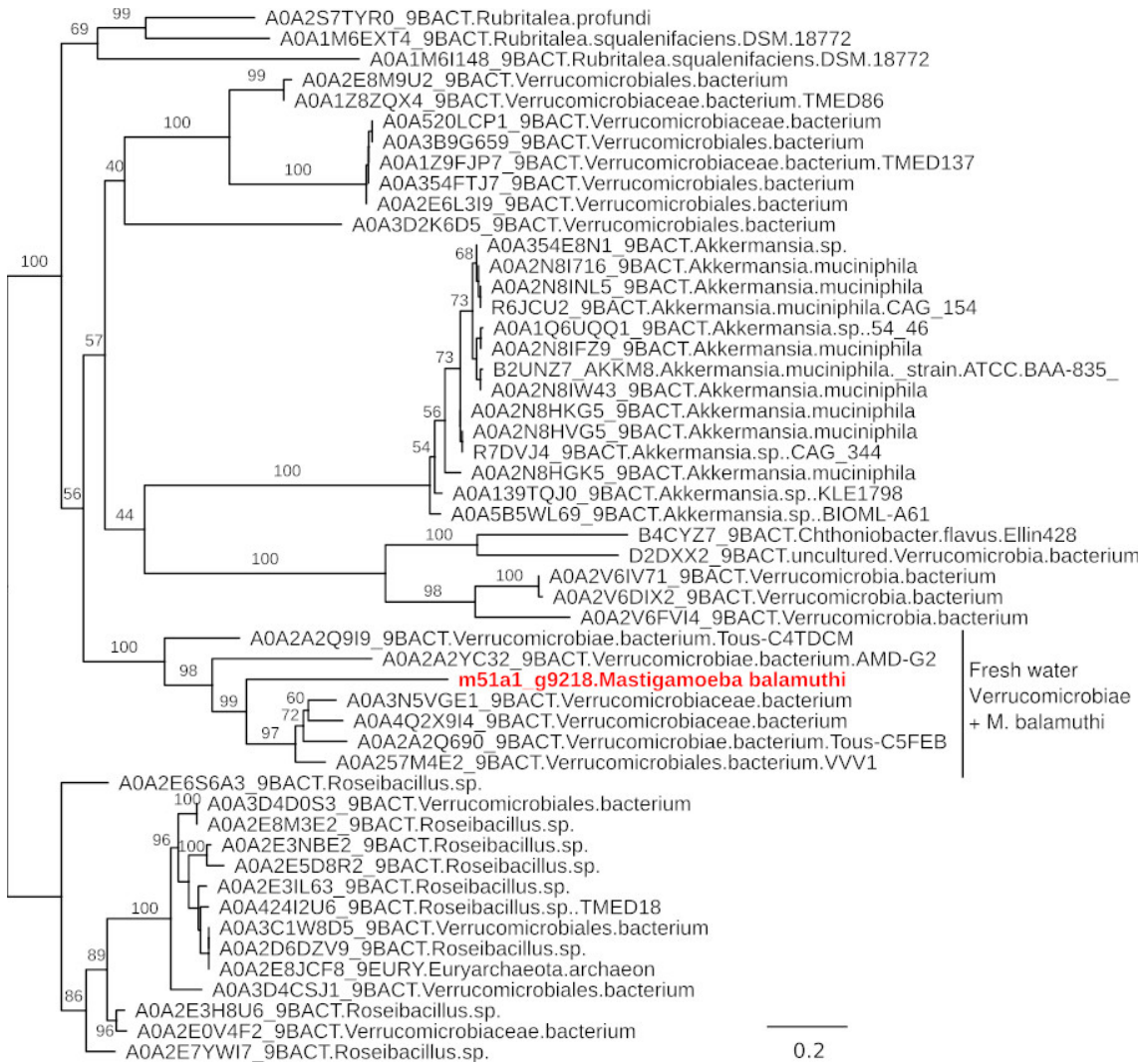


Figure S19. Example of metabolic pathway acquired by LGT. Phylogenies were constructed using IQ-TREE and 1000 ultrafast bootstraps. A. An overview of the proline synthesis pathway from glutamate. The first two enzymes seem to be a recent LGT from a *Verrucomicrobiae* bacterium with sequence identities > 60%. The closest *Verrucomicrobiae* genomes were isolated from fresh water ponds from a depth of about 10 meters. B. Phylogeny of the glutamate 5-kinase from *M. balamuthi* using 51 protein sequences and 244 sites. C. Phylogeny of the glutamate-5-semialdehyde dehydrogenase from *M. balamuthi* using 51 protein sequences and 406 sites.

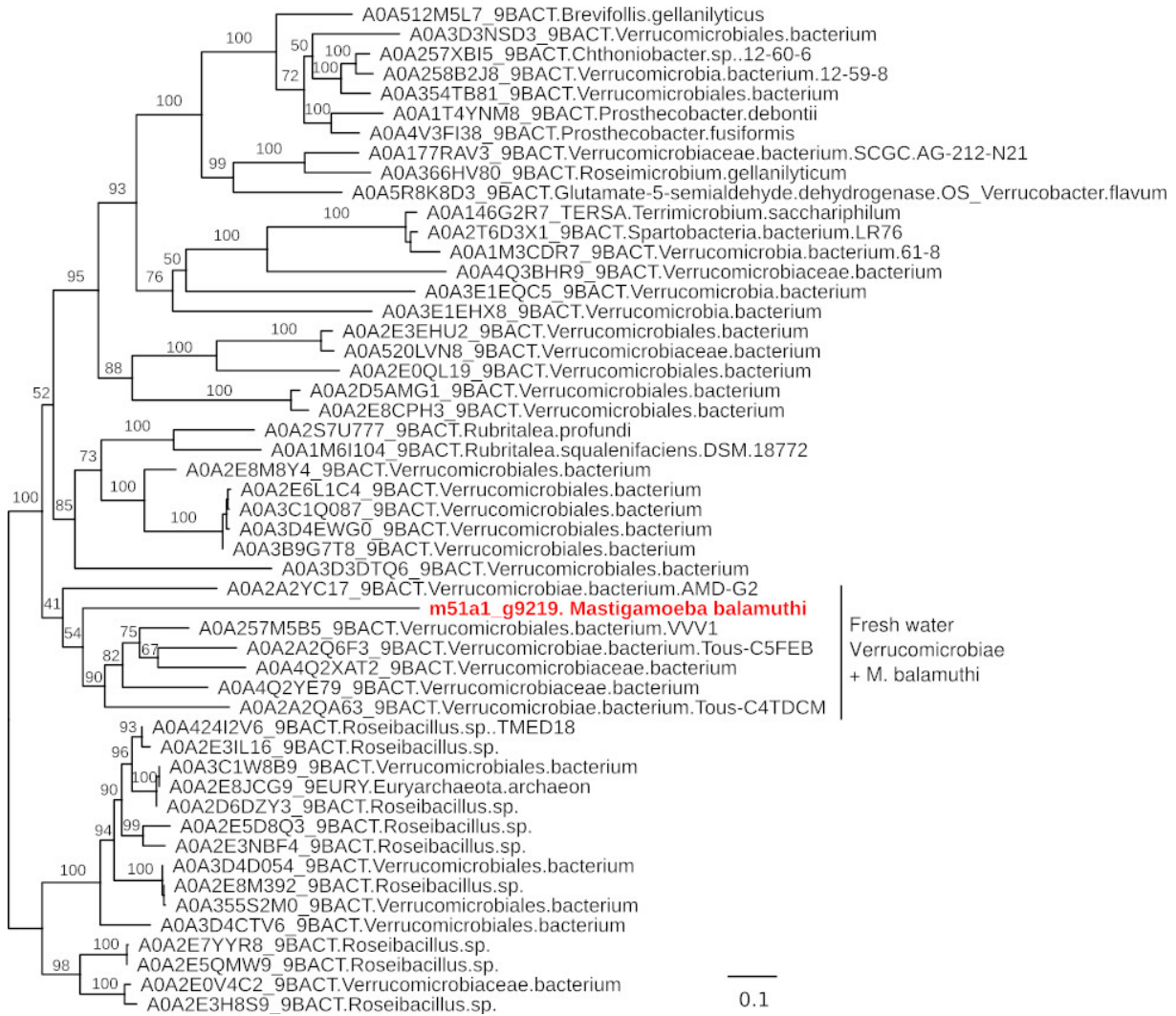
A



B



C



Supplementary Tables:

Table S1. Characteristics of *M. balamuthi* genome sequence

Table S2. Comparison of *M. balamuthi* and three *Entamoeba species* genome properties

Table S3. RNA coding genes

Table S4. Distances of *M. balamuthi* tRNA genes.

Table S5. Analysis of transposable elements in *M. balamuthi* genome.

Table S6. OGs analysis

Table S7. CIA pathway

Table S8. *M. balamuthi* gains – polysaccharide metabolism

Table S9. Flagellum

Table S10 Nuclear pore complex

Table S11. Virulence factors

Table S12. LGT candidates

Link for download:

https://drive.google.com/open?id=1a0YolPVVAFGqFDqosZwf-NYvE9S_S9fQ