First multigene analysis of Archamoebae (Amoebozoa: Conosa) robustly reveals its phylogeny and shows that Entamoebidae represents a deep lineage of the group

Tomáš Pánek^{a,1,*}, Eliška Zadrobílková^{a,b}, Giselle Walker^a, Mathew W. Brown^{c,d}, Eleni Gentekaki^e, Miluše Hroudová^f, Seungho Kang^{c,d}, Andrew Roger^{g,h}, Alexander K. Tice^{c,d}, Čestmír Vlček^f & Ivan Čepička^a

^a*Department of Zoology, Faculty of Science, Charles University in Prague, Vinicna 7, 128 44 Prague, Czech Republic*

^b*Centre for Epidemiology and Microbiology, National Institute of Public Health, Srobarova 48, 100 42 Prague, Czech Republic*

^c*Department of Biological Sciences, Mississippi State University, Mississippi State, MS, 39762, USA*

^d*Institute for Genomics, Biocomputing & Biotechnology, Mississippi State University, Mississippi State, MS, 39762, USA.*

^e*School of Science, Mae Fah Luang University, Chiang Rai, 57100, Thailand*

^f*Department of Genomics and Bioinformatics, Institute of Molecular Genetics, Czech Academy of Sciences, Videnska 1083, 142 20 Prague 4, Czech Republic*

^g*Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, NS B3H 4R2, Canada*

^h*Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS B3H 4R2, Canada*

* Corresponding author: Tomáš Pánek, e-mail address: mistrpanek@seznam.cz.

¹ *Present address: Department of Biology and Ecology, Faculty of Science, University of Ostrava, Chittussiho 10, 710 00 Ostrava, Czech Republic*

1 ABSTRACT

2 Archamoebae is an understudied group of anaerobic free-living or endobiotic protists that
3 constitutes the major anaerobic lineage of the supergroup Amoebozoa. Hitherto, the phylogeny
4 of Archamoebae was based solely on SSU rRNA and actin genes, which did not resolve
5 relationships among the main lineages of the group. Because of this uncertainty, several different
6 scenarios had been proposed for the phylogeny of the Archamoebae. In this study, we present the
7 first multigene phylogenetic analysis that includes members of Pelomyxidae, and
8 Rhizomastixidae. The analysis clearly shows that Mastigamoebidae, Pelomyxidae and
9 Rhizomastixidae form a clade of mostly free-living, amoeboid flagellates, here called
10 Pelobiontida. The predominantly endobiotic and aflagellated Entamoebidae represents a
11 separate, deep-branching lineage, Entamoebida. Therefore, two unique evolutionary events,
12 horizontal transfer of the nitrogen fixation system from bacteria and transfer of the sulfate
13 activation pathway to mitochondrial derivatives, predate the radiation of recent lineages of
14 Archamoebae. The endobiotic lifestyle has arisen at least three times independently during the
15 evolution of the group. We also present new ultrastructural data that clarifies the primary
16 divergence among the family Mastigamoebidae which had previously been inferred from
17 phylogenetic analyses based on SSU rDNA.

19 KEYWORDS

20 Conosa, Pelobiontida, evolution of parasitism, nitrogen fixation system, flagellar apparatus,
21 classification

23 1. INTRODUCTION

24 Archamoebae is a group of anaerobic amoeboid flagellates and amoebae. It was originally
25 created to unite two groups of previously-known, presumptively primarily-amitochondriate,
26 amoeboid protists – pelobionts and Entamoebae (Cavalier-Smith, 1983). Subsequently, remnants
27 of mitochondria were reported from several of its species (e.g. Clark and Roger, 1995; Tovar *et*
28 *al.*, 1999; Walker *et al.*, 2001; Gill *et al.*, 2007; Ptáčková *et al.*, 2013, Zadbóřilková *et al.* 2015a).
29 Using ultrastructural and molecular data, a close relationship between Archamoebae and
30 mycetozoan slime moulds (grouping together with some other taxa as Conosa) has been
31 hypothesised and discussed (Cavalier-Smith, 1998, 2013; Walker *et al.*, 2001; Cavalier-Smith *et*

1 *al.*, 2004); and the relationship has been strongly supported in recent multigene analyses
2 (Cavalier-Smith *et al.*, 2015). Very few other possibly anaerobic species (*Vannella peregriina*,
3 *Flamella citrensis*) have been described within the supergroup Amoebozoa so far (Bovee, 1956;
4 Smirnov and Fenchel, 1996), so Archamoebae constitutes the major anaerobic lineage of
5 Amoebozoa.

6 Currently, Archamoebae comprises approximately 450 nominal species, distributed among
7 five families – Entamoebidae, Pelomyxidae, Mastigamoebidae, Tricholimacidae, and
8 Rhizomastixidae. Most described species are free-living, but the group also contains numerous
9 endobionts (more than 100 nominal species) including the prevalent and significant human
10 parasite *Entamoeba histolytica*, and other protists infecting humans (Clark *et al.*, 2006; Stensvold
11 *et al.*, 2012).

12 In analyses of SSU rDNA, each of the four families containing more than a single species
13 appears robustly monophyletic; Tricholimacidae is monotypic and no molecular data exists for
14 *Tricholimax hylae* (Ptáčková *et al.*, 2013; Zadrobílková *et al.*, 2015; Zadrobílková *et al.*, in
15 press). Monophyly of Pelomyxidae (*Mastigella* + *Pelomyxa*) is further supported by actin gene
16 phylogeny (Zadrobílková *et al.*, 2015). In SSU rDNA trees, family Mastigamoebidae splits into
17 two diverse, statistically well-supported clades, provisionally called Mastigamoebidae A and B
18 (Ptáčková *et al.*, 2013). The latter clade also contains the endobiotic and aflagellate genera
19 *Iodamoeba* and *Endolimax* (Stensvold *et al.*, 2012; Ptáčková *et al.*, 2013).

20 Nevertheless, relationships within the Archamoebae are currently unclear, because
21 neither actin nor SSU rDNA trees are able to resolve relationships between families (Cavalier-
22 Smith *et al.*, 2004; Ptáčková *et al.*, 2013; Stensvold *et al.*, 2012; Zadrobílková *et al.*, 2015;
23 Zadrobílková *et al.*, in press). Both *Pelomyxa* and *Entamoeba* form very long branches in SSU
24 rDNA trees, and their phylogenetic positions are probably affected by long-branch attraction
25 (Ptáčková *et al.*, 2013). Sequence data for use in multigene phylogenetic analyses was hitherto
26 available just for two lineages – *Mastigamoeba balamuthi* (Mastigamoebidae A) and *Entamoeba*
27 *spp.* (Entamoebidae). Morphology is also ambiguous as to relationships between families and
28 even genera, in the absence of heuristic arguments as to which characters are genuinely
29 taxonomically informative (discussed in Walker *et al.* 2001; *c.f.* the revised interpretation of the
30 placement of *Mastigella* in Ptáčková *et al.*, 2015).

1 Anaerobic mitochondrial derivatives (MROs) of two species, *Entamoeba histolytica* and
2 *Mastigamoeba balamuthi*, have been biochemically characterized. It was shown that MROs of
3 each species have a sulfate activation pathway, which is not present in any other known
4 mitochondria, and whose key enzyme (ATP sulfurylase) has been acquired laterally from
5 bacteria (Mi-Ichi *et al.*, 2011; Nývltová *et al.*, 2015). Moreover, both species possess an ϵ -
6 proteobacterial nitrogen fixation system (NIF system), the only eukaryotes to do so. This system
7 has replaced the ancestral mitochondrial iron-sulfur cluster machinery (ISC machinery). It is
8 thought that the ISC machinery exports a sulfur-containing moiety from the mitochondrial matrix
9 to the cytoplasm, for use in cytoplasmic FeS protein biogenesis (CIA pathway). It has also been
10 shown that FeS cluster biogenesis is the only known function of yeast mitochondria that is
11 indispensable to cellular viability (see Lill, 2009). In both, *M. balamuthi* and *E. histolytica*, this
12 ancestral mitochondrial pathway has been lost and replaced by the NIF system, which is active
13 both in MROs and the cytosol of *Mastigamoeba balamuthi*, and in the cytosol of *E. histolytica*
14 (Nývltová *et al.*, 2013). As compared to MROs of *E. histolytica*, MROs of *M. balamuthi*
15 additionally contain a set of proteins typically involved in hydrogenosomal metabolism that
16 allows anaerobic acetyl CoA-dependent synthesis of ATP (Nývltová *et al.*, 2015).

17 Here, we present the first multigene phylogenetic analysis that is based on seven protein-
18 coding genes and includes members of four families in the Archamoebae. Our results clearly
19 show that the predominantly endobiotic and parasitic family Entamoebidae represents a deep
20 lineage of Archamoebae (Entamoebida Cavalier-Smith, 1993), and the other three families form
21 the second clade of the group, the order Pelobiontida Page, 1976. Based on the results,
22 plesiomorphic features and convergent evolution within the Archamoebae is discussed. Using
23 transmission electron microscopy, we define the morphological characteristics of the two clades
24 that currently fall within Mastigamoebidae, “Mastigamoebidae A” and “B”.

26 2. MATERIAL AND METHODS

27 2.1. Organisms and RNA extraction

28 Four examined strains (*Mastigella eilhardi* ATCC 50342, *Rhizomastix libera* IND8, *R. elongata*
29 VAVRH, and *Mastigamoeba abducta* CHOM1) were grown in monoeukaryotic cultures with
30 various unidentified bacteria, following previously published protocols (Ptáčeková *et al.*, 2013;
31 Zadrobílková *et al.*, 2015; Zadrobílková *et al.*, in press). *Pelomyxa* sp. was isolated directly from

1 anaerobic sediment collected from a small freshwater farm pond outside of Fayetteville, AR,
2 USA in September, 2014.

3 Total RNA samples were extracted from 28 – 50 ml of the culture (2 ml of cell
4 suspension lying at the bottom of 10 ml of culture medium, in 15 ml Falcon tubes). Cells of
5 *Mastigella eilhardi* strain ATCC 50342 were filtered through a membrane filter with 5µm pores
6 (Whatman, GE Healthcare Bio-Sciences, USA) to remove bacteria and centrifuged at 1,200 g for
7 10 minutes. Cells of other strains were centrifuged at 1,200 g for 10 minutes without the
8 filtration step. Total RNA was extracted from harvested cells using TriReagent Solution
9 (Ambion, USA) according the manufacturer's instructions.

10 A single *Pelomyxa* sp. cell was transferred from a ~0.5ml drop of the anoxic suspension to a
11 fresh ~0.5ml drop of filter-sterilized natural spring water. Immediately the cell was processed as
12 follows. To remove any other contaminating eukaryotic cells, the *Pelomyxa* cell was
13 successively washed in five fresh aliquots of sterile spring water. Once free of any other potential
14 contaminating eukaryotes, the cell was picked up with a loop made from 32-gauge platinum
15 (<https://youtu.be/nSZuTOZ0QyY>) and transferred to a 200µL PCR tube containing cell lysis
16 buffer. Total RNA was extracted from *Pelomyxa* sp. using a modified version of Smart-seq2
17 (Picelli *et al.*, 2014) that includes an additional six rounds of a freeze thaw cycle to aid in cell
18 lysis in -80°C isopropanol and ~25°C H₂O respectively.

20 2.2.cDNA libraries construction, sequencing, cluster assembly

21 mRNAs from all four examined strains were isolated by selection with Dynabeads Oligo(dT)₂₅
22 (Invitrogen). Illumina sequencing libraries were prepared from mRNA (double polyA selection
23 in a case of IND8 and triple polyA selection in a case of 50342, CHOM1 and VAVRH strains)
24 using BIOO Scientific developed protocol NEXTflex RNA-Seq Kit and sequenced on
25 appropriate platform (MiSeq 150bp paired end for IND8 or HiSeq2000 100bp paired end for
26 50342, CHOM1 and VAVRH strains). Illumina sequence read data were filtered based on
27 quality scores with the fastq_quality_filter program of FASTXTOOLS
28 (http://hannonlab.cshl.edu/fastx_toolkit/), using a cut-off filter (a minimum 70% of bases must
29 have quality of 20 or greater). Filtered sequences were then assembled into clusters using the
30 INCHWORM assembler of the TRINITY package (Grabherr *et al.*, 2011).

The resultant *Pelomyxa* sp. single cell cDNA generated from the SmartSeq-2 method was fragmented using sonification in a Covaris S220 (Woburn, MA, USA) (set at Duty %: 10, Intensity: 5, Burst Cycle: 200, Time: 30s, Mode: Frequency Sweeping). The fragmented cDNA was used as the starting material for a DNA Illumina library generated using the NEBnext Ultra DNA library construction kit (New England Biolabs, Boston, USA) following the manufacturer's protocol. The resultant Illumina library was pooled with four other libraries from non-related protists from a different experiment (5-plex pool). The library pool was sequenced using the MiSeq platform (300bp paired end reads). Reads were demultiplexed by the MiSeq. Reads were trimmed to remove low-quality sequences using Trimmomatic v. 0.30 with a sliding window of 10 nucleotides and a PHRED33 quality threshold of 25 (Bolger et al. 2014). The quality trimmed reads were assembled using Trinity 2.0.4 (Grabherr et al. 2011). Open-reading frames (ORFs) were predicted using TransDecoder (from the Trinity package) and translated to protein sequences.

2.3. Protein Data sets Construction and Gene Searching in ESTs:

Single-gene protein datasets were based on datasets published by Pánek *et al.* (2015) and Cavalier-Smith *et al.* (2014, 2015). In addition, we included also sequences of *Phalansterium solitarium* (GenBank) and screened available ESTs of 11 amoebozoans deposited in MMETSP (Keeling *et al.*, 2014) classified as *Mayorella* sp. ATCC 50980, *Sapocribrum chincoteaguense* ATCC 50979, *Filamoeba nolandii* ATCC 50430, *Neoparamoeba aestuarina* SoJaBio B1-5/56/2, *Paramoeba atlantica* CCAP 1560/9, *Pessonella* sp. PRA-29, *Stereomyxa ramosa* ATCC 50982, *Trichosphaerium* sp. ATCC 40318, *Vannella robusta* DIVA3 518/3/11/1/6, *Vannella* sp. DIVA3 517/6/12, and *Vexillifera* sp. DIVA3 564/2. Those ESTs were screened for previously published gene orthologues of actin, α -tubulin, β -tubulin, cytosolic HSP70, cytosolic HSP90, EF-2, and EF-1 α genes using local BLAST (tBLASTN) in BioEdit 7.0.4.1. The tBLASTN hits were then translated to amino acid residues. Those sequences as well as sequences from five archamoebae obtained during this study were then added to the single-gene datasets and aligned with the help of the MAFFT 7.221 server (<http://mafft.cbrc.jp/alignment/software/>) at default settings, checked and trimmed manually. No stop codons were observed in the coding regions. To test for undetected paralogs or contaminants, we performed phylogenetic analyses of the single-gene alignments (Supplementary Material S1). Newly obtained sequences are deposited as Supplementary Material S2.

Final single-gene alignments were concatenated into the multi-protein data set, which contained 3585 aligned characters (amino acid residues) of seven genes, and is provided as Supplementary Material S3. We did not include alpha-tubulin and beta-tubulin gene sequences of *Entamoeba invadens* and *E. histolytica* in the final multi-protein dataset, since *Entamoeba* tubulins formed extremely long branches in the single gene trees. This trimmed dataset contained more than 3560 amino acid residues for all species of pelobionts excepting *Pelomyxa*, which was represented by 1360 residues. *Entamoeba* spp. was represented by 2615 and 2716 residues, respectively.

2.4. Phylogenetic Analyses

The single-gene Maximum Likelihood trees (ML trees) were constructed in RAxML 8.0.19 (Stamatakis and Ott, 2008) using PROTGAMMAILG or PROTGAMMAILGF model of the sequence evolution with 10 ML tree searches and 500 non-parametric bootstrap replicates. The best partitioning scheme for the multi-protein data set was estimated in PartitionFinder Protein 1.1.0 (Lanfear *et al.*, 2012) under the Bayesian Information Criterion with greedy searching. The Best-scoring ML tree was found using 50 independent heuristic searches. Branch support was estimated from 1000 non-parametric bootstrap replicates. Because RAxML is not equipped with C-series models, we also ran IQ-TREE v. 1.3.3 (Nguyen *et al.* 2015) on the multigene dataset. The best-fitting model available under ML analyses as determined under the Akaike Information Criterion was LG+G4+C60+F with class weights optimized from the dataset (using the exchangeabilities from the LG Q-Matrix (LG+G4+FMIX{empirical,C60pi1...C60pi60})) (Wang, *et al.* 2014). Branch support was estimated from 1000 ultrafast bootstrap replicates in IQ-Tree. Besides two ML approaches, we also used Bayesian Monte Carlo Markov Chain software, PhyloBayes v. 3.3 (Lartillot and Philippe, 2004), which is able to use non-parametric methods for modeling heterogeneous site specific features of sequence evolution. PhyloBayes was run on the multi-protein data set using the CAT-Poisson model. Two independent chains were run until they converged (*i.e.* their maximum observed discrepancy was lower than 0.3 and the effective sample size of all model characteristics was at least 80). Topologies were congruent between chains. Consensus trees and posterior probabilities were calculated using the bpcomp program with the first 25 % of the generations as burn-in, sampling every 50 trees.

Light Microscopy and Transmission Electron Microscopy

As a representative of Mastigamoebidae A, *Mastigamoeba balamuthi* was obtained from the CCAP (Strain 1557/1) in 1999. This strain has subsequently been lost from the CCAP, but is the same strain as ATCC 30984. *M. balamuthi* was cultured in Jones' Horse Serum medium (<http://www.ccap.ac.uk/media/documents/HSM.pdf>), and was observed under the light microscope and prepared for Transmission Electron Microscopy as described previously for *Mastigella commutans* (Walker et al. 2001).

As a representative of Mastigamoebidae B, *Mastigamoeba guttula* was cultured and observed under the light microscope as described previously for the two strains HRADANAN and LU2HNS4 (Ptáčková et al., 2013). A cell suspension of strain HRADANAN was prepared by centrifugation of a 15 ml tube containing 12 ml culture, for ten minutes at $1,000 \times g$. The sample was high-pressure frozen using a Leica EM PACT2 (Leica Camera, Wetzlar, Germany), and cryosubstituted in a Leica EM AFS2 using 100% acetone with 2% OsO₄ as follows: -90 °C for 96 hours, 5 °C for 14 hours; -20 °C for 24 hours; 3 °C for 8 hours; 4 °C for 18 hours. Samples were then washed three times in 100 % acetone. Embedding was done at room temperature, using Epon resin (Poly/Bed 812/Araldite, Polysciences, Warrington, USA), having been infiltrated in an ascending series of concentrations changed after an hour (1:2, 1:1, 2:1). Samples were sectioned at 60 nm thickness using a diamond knife on an Ultracut E ultramicrotome (Reichert) and collected on copper mesh grids coated with formvar film. Ultrathin sections were stained with lead citrate and uranyl acetate (2–3 %) and examined using a TEM JEOL 1011 (Jeol, Tokyo, Japan) transmission electron microscope.

3. RESULTS AND DISCUSSION

3.1. Phylogenetic analyses

In the present study, we obtained EST data from five species, meaning that the broad diversity of Archamoebae could be represented in multigene analyses for the first time: *Rhizomastix libera* and *R. elongata* (Rhizomastixidae), *Mastigella eilhardi* and *Pelomyxa* sp. (Pelomyxidae), and *Mastigamoeba abducta* (Mastigamoebidae B). The latter species represents a lineage that also includes endobiotic genera *Endolimax* and *Iodamoeba* (Ptáčková et al., 2013).

Monophyly of the family Rhizomastixidae and Entamoebidae was highly supported across phylogenies of all seven used molecular markers (genes for actin, alpha-tubulin, beta-tubulin, elongation factor 1-alpha, elongation factor 2, cytosolic HSP70, and HSP90). On the other hand, the resolution of individual gene trees was not sufficient either to clearly show relationships

1 between other lineages of Archamoebae, or to show monophyly of Archamoebae itself (see Fig.
2 1 for RAxML bootstrap supports or Supplementary Material S1 for RaxML gene trees). No
3 robustly or moderately-supported conflicting nodes (bootstrap support > 70) were observed
4 among the gene trees inferred from individual molecular markers.

5 To resolve the internal phylogeny of Archamoebae more robustly, we concatenated
6 sequences from these markers in the final multi-protein phylogenetic analysis. The Bayesian and
7 Maximum Likelihood (ML) analyses yielded a highly congruent topology of Archamoebae.
8 Maximum likelihood analyses (RAxML and IQ-Tree) as well as Bayesian approach
9 (PhyloBayes) showed that Archamoebae form a clade with absolute statistical support. Similarly,
10 relationships within Archamoebae have been recovered with absolute statistical support using
11 Bayesian and IQ-tree ML analyses and at least 93 % using RAxML. These results (see Fig. 2)
12 allowed us to reconstruct the phylogeny of Archamoebae, make inferences about its evolution
13 and revise its taxonomy.

14 Three competing hypotheses about the phylogeny of Archamoebae have been proposed
15 so far: **(1)** Pelomyxidae and Entamoebidae form a clade, sister to the rest of Archamoebae
16 (Cavalier-Smith *et al.*, 2004; Cavalier-Smith, 2013); **(2)** Rhizomastixidae and Entamoebidae
17 constitute a clade (Ptáčková *et al.*, 2013); **(3)** Entamoebidae forms a separate lineage from the
18 rest of Archamoebae (Cavalier-Smith, 1991). Our analysis based on concatenated dataset clearly
19 showed Entamoebidae as the sister clade to the rest of the Archamoebae. This result conclusively
20 supports the third proposed hypothesis about the deep phylogeny of Archamoebae. The
21 phylogenetic position of Rhizomastixidae is surprising since previous actin and SSU rDNA
22 analyses have suggested possible close relationship between Rhizomastixidae and Entamoebidae.

23 Our Bayesian analysis of multiprotein dataset further recovered Amoebozoa as a clade,
24 although its statistical support is very low. As currently shown by Cavalier-Smith *et al.* (2015),
25 Amoebozoa splits into two clades, Conosa and Lobosa. Archamoebae, besides Macromycetozoa
26 and Variosea *sensu* Berney *et al.* (2015), is one of the three conosean lineages. Our analysis
27 demonstrated monophyly of each of them (see Fig. 2). On the other hand, it was unable to
28 resolve relationships between these lineages and to recover Conosa as monophyletic, since the
29 putative loboseans *Sapocribrum* and *Pessonella* (Amoebozoa: Lobosa) branched sister to
30 Variosea with high statistical support in Bayesian and IQ-Tree ML analyses (1 and 96,
31 respectively).

3.2. Transmission Electron Microscopy

Molecular markers (SSU rDNA trees) strongly support division of Mastigamoebidae into two clades, Mastigamoebidae A and B (see Ptáčková *et al.*, 2013). On the other hand, no morphological differences between the clades have been reported so far. Therefore, we decided to thoroughly examine the cell morphology of representatives of both clades (see Table 1 and Figs 3 and 4). Representatives of Mastigamoebidae A (e.g. *Mastigamoeba balamuthi*, *M. aspera*, or *M. punctachora*) generally show more morphological variation and are larger than representatives of Mastigamoebidae B (e.g. *Mastigamoeba guttula*, *M. simplex*).

Full details of the flagellar apparatus characteristics known from each group are summarised in Table 1 and Figure 4, but the main differences are presented briefly here. The cone of microtubules arising from the basal body originates differently in each group: in members of Mastigamoebidae A, microtubules of the cone arise laterally, from along the sides of the basal body, and in some cases arise from the base of the basal body as well; whereas those in Mastigamoebidae B arise longitudinally, close to the base of the basal body, in a single layer. There may be an MTOC present immediately below the basal body, in some taxa of Mastigamoebidae A, but not B. In Mastigamoebidae A, the flagellar transition zone is long and, in some taxa, contains either a dense column or a spiral (potential homologies of the dense column are discussed in Walker *et al.* 2001); whereas the transition zone is short and no extra elements have been seen hitherto in members of Mastigamoebidae B. These results provide synapomorphies for the yet-unclassified groups Mastigamoebidae A and B, which were originally identified by molecular phylogenetics of the SSU rRNA gene, and have subsequently been confirmed by the detailed analyses presented here. Formal classification of these two groups is not given below, as it would require a complete revision of all nominal species of *Mastigamoeba*, *Iodameba* and *Endolimax*, which is beyond the scope of the current paper. However, our comparative analysis (see Table 1) indicates that *Mastigamoeba aspera*, the type species of the genus *Mastigamoeba*, belongs to the Mastigamoebidae “A” clade. Some of the present authors are currently working on a detailed taxonomic revision of Archamoebae which will reflect these findings and will deal with classification of species and genera within Mastigamoebidae.

3.3. Classification of Archamoebae

Currently, the class Archamoebae Cavalier-Smith, 1983 contains five families: Pelomyxidae Schulze, 1877; Mastigamoebidae Goldschmidt, 1907; Rhizomastixidae Ptáčeková *et al.*, 2013; Tricholimacidae Cavalier-Smith, 2013; Entamoebidae Chatton, 1925. Based on our results, we distinguish two orders, Pelobiontida Page, 1976 and Entamoebida Cavalier-Smith 1993. Furthermore, we have divided Pelobiontida Page, 1976 into two suborders - Pelomyxina Starobogatov, 1980 (stat. nov.) and Mastigamoebina Frenzel, 1997 (stat. nov.). We have defined taxa using node-based and branch-based phylogenetic definitions. We have classified the genera *Entamoeba* and *Mastigina*, and the family Tricholimacidae as Archamoebae *incertae sedis*.

Order Pelobiontida Page, 1976

(Eukaryota: Amoebozoa: Conosa: Archamoebae)

Definition: The clade consisting of *Mastigella eilhardi* Bürger, 1905 and all organisms or species that share a more recent common ancestor with *Mastigella eilhardi* Bürger, 1905 than with *Entamoeba histolytica* Schaudinn, 1903. This is a branch-based definition; qualifying clause – the name does not apply if *Protothorax articulatus* Olive & Stojanovich, 1972, *Dictyostelium discoideum* Raper, 1935, or *Filamoeba nolandii* Page, 1967 fall within the specified clade.

Remarks: The term Pelobiontida originally included only the genus *Pelomyxa* Greef, 1874. The term has since been emended by Griffin (1988) specifically to include *Pelomyxa* and mastigamoebids on the grounds both have flagella; it has subsequently been used to include different lineages of Archamoebae (see Ptáčeková *et al.*, 2013). Using the definition presented here, Pelobiontida is composed of three families (Pelomyxidae, Rhizomastixidae, Mastigamoebidae), of which Mastigamoebidae includes aflagellated mastigamoebids—genera *Endolimax* and *Iodamoeba*—that have in the past been treated as entamoebids. We choose to use this name for the whole order because it indicates the typical life style of most species (they live in freshwater sediments; greek word “pelos” means mud), and because Pelobiontida or pelobionts has been used to refer to the whole group in numerous publications from the past two decades.

Suborder Mastigamoebina Frenzel, 1997 (stat. nov.)

(Eukaryota: Amoebozoa: Conosa: Archamoebae: Pelobiontida)

Definition: The least inclusive clade containing *Rhizomastix libera* Ptáčeková *et al.*, 2013, *Mastigamoeba balamuthi* (Chávez *et al.*, 1986), and *Mastigamoeba abducta* Ptáčeková *et al.*

2013. This is a node-based definition: it is intended to apply to a crown clade; qualifying clause – the name does not apply if *Entamoeba histolytica* Schaudinn, 1903, *Mastigella eilhardi* Bürger, 1905, *Pelomyxa palustris* Greeff, 1874, or *Dictyostelium discoideum* Raper, 1935 fall within the specified clade.

Remarks: Mastigamoebina encompasses two families, Mastigamoebidae Goldschmidt, 1907 and Rhizomastixidae Ptáčeková *et al.* 2013. We do not list the family Endolimacidae here since it has been shown that both its genera, *Endolimax* and *Iodamoeba*, branch within Mastigamoebidae (Stensvold *et al.* 2012) and the genus *Endolimax* was transferred to the family Mastigamoebidae by Ptáčeková *et al.* 2013. Mastigamoebidae encompasses two clades, currently named Mastigamoebidae “A” (e.g. *Mastigamoeba balamuthi*, *M. punctachora*, *M. schizophrenia*) and Mastigamoebidae “B” (e.g. *Mastigamoeba simplex*, *M. guttula*, *Endolimax* spp., *Iodamoeba butschlii*) (Ptáčeková *et al.* 2013), which should be given the rank of subfamily upon formal revision of the nominal species and genera contained within them (synapomorphies are defined in section 3.2).

Suborder Pelomyxina Starobogatov, 1980 (stat. nov.)

(Eukaryota: Amoebozoa: Conosa: Archamoebae: Pelobiontida)

Definition: The clade consisting of *Pelomyxa palustris* Greeff, 1874 and all organisms or species that share a more recent common ancestor with *Pelomyxa palustris* Greeff, 1874 than with *Mastigamoeba balamuthi* (Chávez *et al.*, 1986). This is a branch-based definition; qualifying clause – the name does not apply if *Rhizomastix libera* Ptáčeková *et al.*, 2013, *Entamoeba histolytica* Schaudinn, 1903, *Dictyostelium discoideum* Raper, 1935, or *Mastigamoeba guttula* Ptáčeková *et al.*, 2013 fall within the specified clade.

Remarks: Pelomyxina encompasses a single family, Pelomyxidae. We formally transfer *Mastigamoeba bovis* Liebetanz, 1910 to the genus *Mastigella* as ***Mastigella bovis* comb. nov.**, because it shows no connection between the nucleus and the flagellum.

Order Entamoebida Cavalier-Smith, 1993

(Eukaryota: Amoebozoa: Conosa: Archamoebae)

Definition: The clade consisting of *Entamoeba histolytica* Schaudinn, 1903 and all organisms or species that share a more recent common ancestor with *Entamoeba histolytica* Schaudinn, 1903 than with *Mastigella eilhardi* Bürger, 1905. This is a branch-based definition; qualifying clause –

the name does not apply if *Dictyostelium discoideum* Raper, 1935, *Pelomyxa palustris* Greeff, 1874, or *Mastigamoeba balamuthi* (Chávez *et al.*, 1986) fall within the specified clade.

Remarks: Entamoebida encompasses a single family, Entamoebidae. Cavalier-Smith (1993) did not specify genera included in the family Entamoebidae. Because genera *Endolimax* and *Iodamoeba* have been transferred to the different family several years later (Cavalier-Smith *et al.*, 2004), it is clear that originally, Entamoebida was composed of genera *Entamoeba*, *Endamoeba*, *Endolimax* and *Iodamoeba*. We consider genera *Endolimax* and *Iodamoeba* as members of Mastigamoebidae.

Archamoebae *incertae sedis*

The phylogenetic position of some taxa within Archamoebae remains unclear and needs to be resolved using molecular methods. Currently, no molecular data from these organisms are available and the morphology of the taxa does not suggest obvious synonymy with any of the taxa defined above. 1. Family Tricholimacidae Cavalier-Smith, 2013 with sole genus and species *Tricholimax hylae*. 2. Genus *Endamoeba*. 3. Genus *Mastigina*.

3.4. Ancestral features and evolutionary trends in Archamoebae

The last common ancestor of all Archamoebae was an anaerobic amoeboid flagellate. Subsequently, several groups within Archamoebae partially lost flagellum-mediated movement, and some have even lost the entire flagellar apparatus, as seen in Entamoebidae and some mastigamoebids (*Endolimax*, *Iodamoeba*). *Mastigamoeba balamuthi* (Mastigamoebidae A) and *Entamoeba histolytica* (Entamoebidae) differ significantly in the complexity of their mitochondrial metabolism, with *E. histolytica* possessing an extremely reduced mitochondrial derivative, the mitosome, while *M. balamuthi* possessing a hydrogenosome (Tovar *et al.*, 1999; Mi-Ichi *et al.*, 2011; Nývltová *et al.*, 2015). Based on our phylogenetic analysis, we can conclude that both of the laterally transferred biochemical pathways that have been found in those species, *i.e.* the ϵ -proteobacterial NIF system for FeS cluster assembly and the mitochondrial-targeted sulphur activation pathway (see above), were present in the last common ancestor of Archamoebae.

Now, we can also conclude that the ancestors of Archamoebae, Pelobiontida, Pelomyxina, and Mastigamoebina were free-living protists. The endobiotic life style appeared at least three times during the evolution of Archamoebae: in the ancestor of Entamoebidae; in the

1 ancestor of the ‘*Endolimax* + *Iodamoeba*’ clade (in Mastigamoebidae B), and within the genus
 2 *Rhizomastix*. Although we cannot exclude the hypothesis that the last common ancestor of
 3 Mastigamoebina, or even of the whole Archamoebae, was endobiotic, we consider such
 4 scenarios much less plausible because other conoseans and most Archamoebae species are free-
 5 living. Almost all members of Pelomyxidae and Mastigamoebidae are free-living, with
 6 *Tricholimax hylae*, *Mastigella bovis* and *Mastigamoeba* sp. (‘Mastigamoebidae A’) found in
 7 vacuoles in *Pelomyxa belevskii* as potential exceptions (Ptáčková *et al.*, 2013, see below). In
 8 SSU rRNA gene trees the closely related genera of endobiotic mastigamoebids, *Endolimax* and
 9 *Iodamoeba*, form a internal branch of otherwise free-living Mastigamoebidae (Stensvold *et al.*,
 10 2012; Ptáčková *et al.*, 2013). The deepest split in the genus *Rhizomastix* is between free-living
 11 *Rhizomastix libera* and other *Rhizomastix* spp. Some of these species are endobiotic, while *R.*
 12 *elongata* was isolated from abandoned cesspit and was suspected to be free-living as well
 13 (Zadrobílková *et al.*, in press). Thus an “early-endobiotic” scenario would require multiple
 14 reversions of the endobiotic lifestyle within Pelobiontida. Endobiotic-to-free-living transitions
 15 are rare in the nature and thus less probable than *vice versa*. Nevertheless, Archamoebae is one
 16 of very few protistan lineages that are suspected to contain secondarily free-living organisms that
 17 evolved from endobiotic ancestors. The most studied example is *Entamoeba moshkovskii* (Clark
 18 *et al.*, 2006; Clark and Diamond, 1997), which has repeatedly been isolated from animal or
 19 human stool as well as water sediments (see Heredia *et al.*, 2014) and is probably amphizoic
 20 (*i.e.*, both free-living and endobiotic). Recently, *Entamoeba marina*, that is closely related to *E.*
 21 *moshkovskii*, has been isolated from tidal flat sediment (Shiratori and Ishida, in press). It
 22 indicates that Entamoebida is still an undersampled group of protists.

23 There remain three endobiotic lineages of Archamoebae whose phylogenetic position is
 24 uncertain, and for which no sequence data are currently available: *Tricholimax hylae*,
 25 *Entamoeba* spp., and *Mastigella bovis*. *Tricholimax hylae* shares morphological features both
 26 with Pelomyxidae and Mastigamoebidae (see Brugerolle 1982, 1991; Walker *et al.*, 2001) and
 27 was recently assigned as a sole genus and species in the family Tricholimacidae Cavalier-Smith,
 28 2013. *Mastigella bovis* was described from the rumen of cattle as a member of the genus
 29 *Mastigamoeba* (Liebentanz, 1910) and listed as a probable member of the genus *Mastigella* by
 30 Ptáčková *et al.* (2013); we transferred it to the genus *Mastigella*, see above. All these three
 31 lineages could possibly represent other independent transitions between free-living and

1 endobiotic lifestyle. However, their presumed phylogenetic positions have to be clarified by
2 analyses of molecular data.

3 **3.5.Flagellar apparatus of Conosa and Archamoebae**

4 The last common ancestor of Conosa (= Variosea, Macromycetozoa, and Archamoebae) was
5 very probably an aerobic, biflagellated protist equipped with both anterior and recurrent
6 flagellum (Cavalier-Smith, 1998, 2013; Cavalier-Smith *et al.*, 2015). Assuming that what is now
7 known to be the most phylogenetically-widespread flagellar morphology is ancestral, the basal
8 bodies of its flagella may have been associated with five different microtubular elements defined
9 as MTA1–MTA5 (Wright *et al.*, 1979); the basal body of the anterior flagellum would have been
10 associated with MTA1–3, while the posterior basal body would have been associated with
11 MTA4 and 5. Yubuki and Leander (2013) synonymized MTA3 with the eukaryotic root R3 and
12 MTA2 with superficial microtubules that originate on it. Further, they hypothesized that MTA4
13 is homologous to the eukaryotic root R2, and MTA5 corresponds to the root R1. These four
14 cytoskeletal elements must have arisen very early in the evolution of eukaryotes (Yubuki and
15 Leander, 2013). Some flagellated members of the Conosa also possess MTA1, which arises from
16 a microtubule organizing center (MTOC) located at the proximal part of the basal body of the
17 anterior flagellum. In those conoseans that have the flagellar apparatus associated with the
18 nucleus, MTA1 microtubules (if present) extend from the MTOC towards the apical part of the
19 nucleus and follow its surface (*e.g.* Wright *et al.*, 1979; Spiegel, 1981).

20 The clade Conosa was morphologically defined by a monolayer of microtubules partially
21 or completely surrounding the anterior basal body and diverging towards the nucleus and cell
22 posterior as a half or three-quarters open cone (Variosea and Mycetozoa) or a complete cone
23 (Archamoebae) (Cavalier-Smith, 1998, 2013, 2015). The cone as defined by Cavalier-Smith is
24 the same structure as MTA2 in Wright's terminology, and the superficial microtubules of Yubuki
25 and Leander's terminology (Yubuki and Leander, 2013). However, conoseans equipped with
26 both posterior and anterior flagella possess a more complex cone, formed not only by
27 microtubules of MTA2, but also MTA3–MTA5 (Wright *et al.* 1979). In addition, the MTA1 of
28 Macromycetozoa *sensu* Berney *et al.* (2015) and protosteloid Variosea usually form an inner
29 cone associated with the apical part of the nucleus (*e.g.* Wright *et al.*, 1979; Spiegel, 1981;
30 Walker *et al.*, 2001, 2003). During the evolution of Conosa, the ancestral flagellar apparatus has
31 been transformed into many different variants. In some lineages, the posterior flagellum and

1 associated cytoskeleton (Pelobiontida, *Planoprotostelium*, *Cavostelium*, *Phalansterium*), or even
2 the whole flagellar apparatus (Entamoebidae, ‘*Endolimax* and *Iodamoeba*’ clade in the
3 Pelobiontida, *Flamella*, *Acramoeba*, *Grellamoeba*, *Filamoeba*) have been lost (see Spiegel,
4 1981; Appendix 3 in Walker *et al.*, 2001; Cavalier-Smith *et al.*, 2015, Ptáčková *et al.*, 2013).

5 Walker *et al.* (2001), Cavalier-Smith (2013) and Yubuki and Leander (2013) have all
6 suggested that the microtubular cone is homologous to the MTA2. This interpretation is,
7 however, problematic in two main aspects: (1) the superficial microtubules of other Conosa and
8 eukaryotes form a sheet anchoring close to the dorsal side of the anterior basal body, so MTA-2
9 would have had to undergo intricate rearrangements during the evolution of Archamoebae. (2)
10 Radiating microtubules arise in multiple layers from the basal body of several Archamoebae (see
11 Brugerolle 1982, 1991; Simpson *et al.* 1997; Walker *et al.*, 2001; Frolov *et al.* 2011). Such
12 architecture is highly unusual for superficial microtubules since they form a monolayer in other
13 eukaryotes. The other possible homology would be with MTA1 as defined by Wright *et al.*
14 (1979). However, if this were correct, then the cone of Archamoebae would not be homologous
15 to the cone of other Conosa as defined by Cavalier-Smith (2013). In our opinion, it is currently
16 impossible to decide between these two interpretations of homology of the archamoeban cone
17 because the flagellar apparatus of Archamoebae is too simplified and derived. Individual
18 microtubular elements cannot be unequivocally homologized with microtubular ribbons of other
19 Conosa and Eukaryota.

20 Regardless of which scenario is correct, it is clear that flagella of most *Pelomyxa* spp. and
21 *Tricholimax hylae* have lost motility secondarily, and their non-‘9+2’ pattern of axoneme
22 microtubules (see Walker *et al.*, 2001) is aberrant. Besides, members of the genus *Pelomyxa*
23 have multiplied the flagellar apparatus and nucleus (*e.g.* Chistyakova *et al.*, 2014). Ancestors of
24 Entamoebida and ‘*Endolimax* + *Iodamoeba*’ clade have completely lost the flagellar apparatus.

25 We conclude that the last common ancestor of Pelobiontida, or possibly of all of the
26 Archamoebae, resembled members of the genera *Mastigamoeba* and *Mastigella*: possessing a
27 single motile, anterior flagellum with the classical ‘9+2’ pattern of axoneme microtubules,
28 lacking a posterior flagellum, and outer dynein arms in the anterior flagellar axoneme; and with
29 the anterior basal body having a cone of radiating microtubules and lateral microtubular ribbon.

30 4. CONCLUSIONS

Our work provides the first robust evidence for the primary divergence at the base of Archamoebae between Entamoebida and a major clade containing all flagellate Archamoebae (Pelobiontida). Based on these results, we revised the higher classification of Archamoebae and concluded that the bacterial nitrogen fixation system was present in the last common ancestor of Archamoebae (LCAA), mitochondrial derivatives of the LCAA contained a sulfate activation pathway, and that the endobiotic life-style has arisen at least three times during the evolution of the group. Our comparative ultrastructural analysis of Mastigamoebidae „A“ and „B“ showed synapomorphies of these two clades and indicates that *Mastigamoeba aspera*, the type species of the genus, belongs to the Mastigamoebidae „A“. Future studies on individual lineages included in the present study may help us to elucidate the evolution of anaerobic metabolism, via lateral gene transfer; as well as to understand the transition from free-living to endobiotic and parasitic lifestyles.

ACKNOWLEDGEMENTS

This work was supported by the Grant Agency of Charles University (project 521112), SVV (project 260 208/2015). This project was supported in part by the National Science Foundation Grant DEB 1456054 (<http://www.nsf.gov>), awarded to MWB. Mississippi State University's High Performance Computing Collaboratory provided some computational resources. The access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the program LM2010005, is also appreciated. We wish to thank Dr. Franck Gael Carbonero at the University of Arkansas for running the MiSeq lane of *Pelomyxa* sp.

REFERENCES

- Berney, C., Geisen, S., Van Wichelen, J., Nitsche, F., Vanormelingen, P., Bonkowski, M., Bass, D., 2015. Expansion of the 'reticulosphere': diversity of novel branching and network-forming amoebae helps to define Variosea (Amoebozoa). *Protist* 166, 271-295.
- Bovee, E.C., 1956. Some observations on the morphology and activities of a new ameba from citrus wastes, *Flamella citrensis* n. sp. *J Protozool* 3, 151–155.

- 1 Bolger, A. M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina
2 sequence data. *Bioinformatics* 30, 2114–2120.
- 3
- 4 Brugerolle, G. 1982. Caractères ultrastructuraux d'une mastigamibe: *Mastigina hylae*
5 (Frenzel). *Protistologica*, 18 227–235.
- 6
- 7 Brugerolle, G., 1991. Cell organization in free-living amitochondriate heterotrophic
8 flagellates. *The Biology of Free-living Heterotrophic Flagellates*. Clarendon Press, Oxford, pp.
9 133-148.
- 10
- 11 Cavalier-Smith, T., 1983. A six-kingdom classification and a unified phylogeny, in:
12 Schwemmler, W., Schenk, H.E.A. (Eds), *Endocytobiology II. Intracellular Space as Oligogenetic*
13 *Ecosystem*. De Gruyter, Berlin, pp. 1027–1034.
- 14
- 15 Cavalier-Smith, T., 1991. Archamoebae: the ancestral eukaryotes? *Biosystems* 25, 25–38.
- 16
- 17 Cavalier-Smith, T., 1998. A revised six-kingdom system of life. *Biol Rev* 73, 203–266.
- 18
- 19 Cavalier-Smith, T., 2013. Early evolution of eukaryote feeding modes, cell structural diversity,
20 and classification of the protozoan phyla Loukozoa, Sulcozoa, and Choanozoa. *Eur J Protistol*
21 49, 115-178.
- 22
- 23 Cavalier-Smith, T., Chao, E.E.Y., Oates, B., 2004. Molecular phylogeny of Amoebozoa and the
24 evolutionary significance of the unikont *Phalansterium*. *Eur J Protistol* 40, 21–48.
- 25
- 26 Cavalier-Smith, T., Chao, E.E., Snell, E.A., Berney, C., Fiore-Donno, A.M., Lewis, R., 2014.
27 Multigene eukaryote phylogeny reveals the likely protozoan ancestors of opisthokonts (animals,
28 fungi, choanozoans) and Amoebozoa. *Mol Phylogenet Evol*, 81, 71–85.
- 29

1 Cavalier-Smith, T., Fiore-Donno, A.M., Chao, E., Kudryavtsev, A., Berney, C., Snell, E.A.,
 2 Lewis, R., 2015. Multigene phylogeny resolves deep branching of Amoebozoa, *Mol Phylogenet*
 3 *Evol* 83, 293–304.
 4
 5 Chistyakova, L.V., Berdieva, M.A., Frolov, A.O., Goodkov, A.V., 2014. Reisolation and
 6 redescription of pelobiont *Pelomyxa paradoxa* Penard, 1902 (Archamoebae, Pelobiontida). *Cell*
 7 *Tissue Biol* 8, 504–512.
 8
 9 Clark, G.C., Diamond, L.S., 1997. Intraspecific variation and phylogenetic relationships in the
 10 genus *Entamoeba* as revealed by riboprinting. *J Euk Microbiol* 44, 142–154.
 11
 12 Clark, G.C., Roger, A.J., 1995. Direct evidence for secondary loss of mitochondria in *Entamoeba*
 13 *histolytica*. *Proc Natl Acad Sci USA* 92, 6518–6521.
 14
 15 Clark, C.G., Kaffashian, F., Tawari, B., Windsor, J.J., Twigg-Flesner, A., Davies-Morel, M.C.,
 16 Blessmann, J., Ebert, F., *et al.*, 2006. New insights into the phylogeny of *Entamoeba* species
 17 provided by analysis of four new small-subunit rRNA genes. *Int J Syst Evol Microbiol* 56, 2235–
 18 2239.
 19
 20 Derelle, R., Torruella, G., Klimeš, V., Brinkmann, H., Kim, E., Vlček, Č. *et al.*, 2015. Bacterial
 21 proteins pinpoint a single eukaryotic root. *Proc Natl Acad Sci USA* 112, E693-E699.
 22
 23 Frolov, A.O., Chistyakova, L.V., Malysheva, M.N., 2011. Light and electron microscopic study
 24 of *Pelomyxa flava* sp. n. (Archamoebae, Pelobiontida). *Cell Tissue Biol* 5, 81–89.
 25
 26 Gill, E.E., Diaz-Trivino, S., Barbera, M.J., Silberman, J.D., Stechmann, A., Gaston, D., Tamas,
 27 I., Roger, A.J., 2007. Novel mitochondrion-related organelles in the anaerobic amoeba
 28 *Mastigamoeba balamuthi*. *Mol Microbiol* 66, 1306–1320.
 29
 30 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X.,
 31 Fan, L., Raychowdhury, R., Zeng, Q., *et al.*, 2011. Full-length transcriptome assembly from
 32 RNA-Seq data without a reference genome. *Nat Biotechnol* 29, 644–652.

- 1 Griffin, J.L., (1988) Fine structure and taxonomic position of the giant amoeboid flagellate
2 *Pelomyxa palustris*. J Protozool 35, 300–315.
- 3
- 4 Heredia, R.D., Fonseca, J.A., López, M.C., 2012. *Entamoeba moshkovskii* perspectives of a new
5 agent to be considered in the diagnosis of amebiasis. Acta Trop 123, 139–145.
- 6
- 7 Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust,
8 E.V., Archibald, J.M., *et al.* (2014) The Marine microbial eukaryote transcriptome sequencing
9 project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through
10 transcriptome sequencing. PLoS Biol 12: e1001889.
- 11
- 12 Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. PartitionFinder: combined selection of
13 partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29,
14 1695–1701.
- 15
- 16 Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the
17 amino-acid replacement process. Mol Biol Evol 21: 1095–1109.
- 18
- 19 Liebetanz E., 1910. Die parasitischen protozoen des Wiederkäuermagens. Arch Protistenk 19,
20 19–80.
- 21
- 22 Lill R., 2009. Function and biogenesis of iron-sulphur proteins. Nature 460, 831–838.
- 23
- 24 Mi-ichi, F., Makiuchi, T., Furukawa, A., Sato, D., Nozaki, T., 2011. Sulfate activation in
25 mitochondria plays an important role in the proliferation of *Entamoeba histolytica*. PLoS Negl Trop
26 Dis 5, e1263.
- 27
- 28 Nguyen, L. T., Schmidt, H. A., von Haeseler, A., Minh, B. Q., 2015. IQ-tree: a fast and effective
29 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32, 268–
30 274.
- 31

1 Nývltová, E., Šuták, R., Harant, K., Šedinová, M., Hrdý, I., Pačes, J., Vlček, Č., Tachezy, J.,
 2 2013. NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the
 3 mitochondria and cytosol of *Mastigamoeba balamuthi*. Proc Natl Acad Sci USA 110, 7371–
 4 7376.
 5
 6 Nývltová, E., Stairs, C. W., Hrdý, I., Rídl, J., Mach, J., Pačes, J., Roger, A.J, Tachezy, J., 2015.
 7 Lateral gene transfer and gene duplication played a key role in the evolution of *Mastigamoeba*
 8 *balamuthi* hydrogenosomes. Mol Biol Evol 32, 1039–1055.
 9
 10 Page, F.C., 1976. A revised classification of the Gymnamoebia (Protozoa: Sarcodina). Zool J
 11 Linn Soc 58, 61–77.
 12
 13 Pánek, T., Táborský, P., Pechiadaki, M.G., Hroudová, M., Vlček, Č., Edgcomb, V.P., Čepička,
 14 I., 2015. Combined culture-based and culture-independent approaches provide insights into
 15 diversity of jakobids, an extremely plesiomorphic eukaryotic lineage. Front Microbiol 6, 1288.
 16
 17 Ptáčková, E., Kostygov, A.Y., Chistyakova, L.V., Falteisek, L., Frolov, A.O., Patterson, D.J.,
 18 Walker, G., Cepicka, I., 2013. Evolution of Archamoebae: morphological and molecular
 19 evidence for pelobionts including *Rhizomastix*, *Entamoeba*, *Iodamoeba*, and
 20 *Endolimax*. Protist 164, 380–410.
 21
 22 Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R., 2014. Full-length
 23 RNA-seq from single cells using Smart-seq2. Nat Protoc 9, 171–181.
 24
 25 Simpson, A.G.B., Bernard, C., Fenchel, T., Patterson, D.J., 1997. The organisation of
 26 *Mastigamoeba schizophrenia* n. sp.: more evidence of ultrastructural idiosyncrasy and simplicity
 27 in pelobiont protists. Eur J Protistol 33, 87–98.
 28
 29 Smirnov, A.V., and Fenchel, T., 1996. *Vahlkampfia anaerobica* n. sp. and *Vannella peregrinia* n.
 30 sp.(Rhizopoda) - Anaerobic amoebae from a marine sediment. Arch Protistenkd 147, 189–198.
 31

Shiratori, T., Ishida, K., in press. *Entamoeba marina* n. sp.; a new species of *Entamoeba* isolated from tidal flat sediment of Iriomote island, Okinawa, Japan. *J Eukaryot Microbiol*, in press.

Spiegel, F.W., 1981. Phylogenetic significance of the flagellar apparatus in protostelids (Eumycetozoa). *BioSystems*, 14, 491–499.

Stamatakis, A., Ott, M., 2008. Efficient computation of the phylogenetic likelihood function on multi-gene alignments and multi-core architectures. *Phil. Trans. R. Soc. B.* 363, 3977–3984.

Stensvold, C.R., Lebbad, M., Clark, C.G., 2012. Last of the human protists: The phylogeny and genetic diversity of *Iodamoeba*. *Mol Biol Evol*, 29, 39–42.

Tovar, J., Fischer, A., Clark, C.G., 1999. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol Microbiol*, 32, 1013–1021.

Walker, G., Simpson, A.G., Edgcomb, V., 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). *Eur J Protistol*, 37, 25–49.

Walker, G., Silberman, J.D., Karpov, S.A., Preisfeld, A., Foster, P., Frolov, A.O., et al., 2003. An ultrastructural and molecular study of *Hyperamoeba dachnaya*, n. sp., and its relationship to the mycetozoan slime moulds. *Eur J Protistol*, 39, 319–336.

Wang, H.C., Susko, E., Roger, A.J., 2014. An amino acid substitution-selection model adjusts residue fitness to improve phylogenetic estimation. *Mol Biol Evol.* 31, 779–792.

Wright, M., Moisand, A., Mir, L., 1979. The structure of the flagellar apparatus of the swarm cells of *Physarum polycephalum*. *Protoplasma*, 100, 231–250.

Yubuki, N., Leander, B.S., 2013. Evolution of microtubule organizing centers across the tree of eukaryotes. *Plant J* 75, 230–244.

- 1 Zadbílková, E., Walker, G., Čepička, I., 2015. Morphological and molecular evidence support
2 a close relationship between the free-living archamoebae *Mastigella* and *Pelomyxa*. Protist 166,
3 14–41.
4
5 Zadbílková, E., Smejkalová, P., Walker, G., Čepička, I., in press. Morphological and
6 molecular diversity of the neglected genus *Rhizomastix* Alexeieff, 1911 (Amoebozoa:
7 Archamoebae) with description of five new species. J Eukaryot Microbiol, in press.
8

Legends to tables and figures

Figure 1: Bootstrap support of Archamoebae and its internal nodes as seen in single-gene and seven-gene phylogenetic analyses. In the lower side of the table, all alternative groups that were recovered with bootstrap support >50 in at least one single gene tree are also presented.

Figure 2: Phylogenetic tree of eukaryotes based on concatenation of seven protein-coding genes: actin, α -tubulin, β -tubulin, EF1 α , EF2, HSP70, HSP90. The tree is based on alignment of 3585 positions and 78 taxa. The topology was constructed in PhyloBayes under CAT Poisson model. The values at nodes represent PhyloBayes posterior probabilities, RAXML non-parametric bootstraps, and IQ-tree bootstrap support. The values lower than 50% or 0.5 are marked by „*“; branches that were missing in the best ML tree topology are marked by „-“. Clades supported by statistical support higher than 0.98/90/90 are marked by thick branches. Taxa whose ESTs were newly sequenced are in bold. Photos: (A) – *Rhizomastix libera* strain IND8, (B) – *Mastigella eilhardi* strain ATCC 50342. Scale Bar = 10 μ m. With respect to results published by Derelle et al. (2015), we did not mark *Malawimonas* as a member of Excavata.

Figure 3: Mastigamoebidae A taxa show more morphological variation and are generally larger than Mastigamoebidae B taxa. Microtubules of the cone in Mastigamoebidae A arise laterally, from along the sides of the basal body; whereas those in Mastigamoebidae B arise longitudinally, close to the base of the basal body. (A–J) show a representative of Mastigamoebidae A, *Mastigamoeba balamuthi*. (A–F) Light microscopy, DIC optics (A, B) Small amoebae with 2–4 nuclei, the dominant life cycle stage of *M. balamuthi*; (C) "Giant" amoeba form with ca. 50 nuclei; (D) Binucleate gliding flagellate; (E) Uninucleate flagellate; (F) Swimming form with posterior pseudopodia. (G–J) Transmission electron-microscopy of the flagellar apparatus, serial sections, 90 nm apart: Flagellar apparatus, showing lateral emergence of cone microtubules from the basal body (Bb), transitional cylinder at the base of the transition zone (TC), and electron-dense column (DC) at the top of the transition zone, from which the central pair of axonemal microtubules emerges. A microtubular root (MR) emerges laterally from the basal body, with a bilaminar root sheet (RS) on its distal edge. **(K–P) show a representative of Mastigamoebidae B, *Mastigamoeba guttula*.** (K, L) Gliding cells, strains

LUH2NS4 and HRADANAN respectively; (M) Aflagellate cell, LUH2NS4; (N–P) Strain HRADANAN, transmission electron-microscopy of the flagellar apparatus; (N) Transverse section through the basal body, close to the base, showing longitudinal alignment of cone microtubules (MC) and the lateral emergence of the microtubular root (MR); (O–P) Longitudinal sections through the flagellar apparatus, showing the laterally-emerging microtubular root (MR), the transitional cylinder (TC), and the longitudinally-emerging microtubular cone. The transition zone is ca. 200 nm long, which is short, similar to that seen in *Mastigamoeba simplex* (Walker et al. 2001). Scale bar in K = 20 µm for a, b; 25 µm for c; 15 µm for d; 20 µm for e; 750 nm for f–j; 10 µm for k, l, m; 500 nm for n; 750 nm for o, p. Micrographs K, L, M are reproduced from Ptáčeková *et al.* 2013 with permission from Elsevier.

Figure 4: Representative flagellar apparatuses from Mastigamoebidae A and B. (A) Schematic diagram of the microtubular flagellar apparatus of *Mastigamoeba punctachora*, a representative of Mastigamoebidae A. Note that the cone of microtubules (MC) arises laterally from both the sides and the base of the basal body. The flagellar transition zone (TZ) is long and contains a dense column (DC). An MTOC below the basal body has not been confirmed in *M. punctachora* so this characteristic of some members of Mastigamoebidae A is not shown here. Fl, flagellar axoneme; TC, transition zone cylinder; Bb, basal body; RS, bi-laminar root sheet; MR, microtubular root; SMt, side microtubules (part of the microtubular cone). **(B) Schematic diagram of the microtubular flagellar apparatus of *Mastigamoeba simplex*, a representative of Mastigamoebidae B.** Note that the cone of microtubules (MC) arises longitudinally from near the base of the basal body. The flagellar transition zone (TZ) is short and contains no extra elements. Both figures are reproduced from Walker *et al.* 2001, with permission from Elsevier.

Table 1: Flagellar apparatus characteristics of representatives of Mastigamoebidae A and B clades.

Figure 1 (RGB)

CLADE	Bootstrap support in phylogenetic trees (RAxML)							
	ACT	ATUB*	BTUB	EF1A	EF2	HSP70	HSP90*	ALL
Archamoebae			57			83	66	100
Entamoebida	99	100	100	100	100	100	100	100
Pelobiontida					91		85	100
Pelomyxina	76			96	50	99		100
Mastigamoebina				53		97		96
Mastigamoebidae			85		93			93
Rhizomastixidae	100	100	100	100	100	100	100	100
Pelomyxina + Mastigamoebinae	52							
Pelomyxina + Rhizomastixidae					51			
Pelomyxina + Entamoebida						69		
<i>Mastigella</i> + Mastigamoebidae							52	
Mastigamoeb. B + Rhizomastixidae						59		

* Sequences of these genes are missing in *Pelomyxa* data

Figure 1 (grey scale)

CLADE	Bootstrap support in phylogenetic trees (RAxML)							
	ACT	ATUB*	BTUB	EF1A	EF2	HSP70	HSP90*	ALL
Archamoebae			57			83	66	100
Entamoebida	99	100	100	100	100	100	100	100
Pelobiontida					91		85	100
Pelomyxina	76			96	50	99		100
Mastigamoebina				53		97		96
Mastigamoebidae			85		93			93
Rhizomastixidae	100	100	100	100	100	100	100	100
Pelomyxina + Mastigamoebinae	52							
Pelomyxina + Rhizomastixidae					51			
Pelomyxina + Entamoebida						69		
<i>Mastigella</i> + Mastigamoebidae							52	
Mastigamoeb. B + Rhizomastixidae						59		

* Sequences of these genes are missing in *Pelomyxa* data

Figure 2 (RGB)

Figure 2 (grey scale)

Figure 3

Figure 4

Table 1

Larger group/ Species	Flagellum	9+2 axoneme	Dynein arms missing	Transition Zone length	Transitional Dense Column or Spiral	Transition Zone cylinder	Cartwheel in base of Bb	Bb length	Lateral root of MT	Bi-laminar root sheet	MT of cone emerging from sides of Bb	MT of cone emerging from Bb base	Electron-dense material around Bb	Electron-dense MTOC below Bb	References
Mastigamoebidae “A”															
<i>Mastigamoeba aspera</i>	Long (>2× cell)	9+2	?	"Short"	?	?	?	250nm	Yes	?	Many	Most of cone	No	Triangle	Chystyakova <i>et al.</i> 2012
<i>Mastigamoeba balamuthi</i>	Long (>2× cell)	9+2	Yes, outer	700nm	DC	Yes	?	250nm	Yes	Yes	Most of cone	Few	No	No	Chavez <i>et al.</i> 1986, Brugerolle 1991, this paper
<i>Mastigamoeba punctachora</i>	Long (>2× cell)	9+2	Yes, outer	1000nm	DC	Yes	No	250nm	Yes	Yes	Most of cone	Few	No	Possibly?	Bernard <i>et al.</i> 2000, Walker <i>et al.</i> 2001
<i>Mastigamoeba schizophrenia</i>	Long (>2× cell)	9+2	Yes, outer	700nm	DC	Yes	Yes	250nm	Yes	Yes	Single layer at base, lateral	No	No	No	Simpson <i>et al.</i> 1997
<i>Mastigamoeba</i> sp.	Long (>2× cell)	9+2	?	700nm	Spiral + central filament	Yes	?	250nm	Yes	Yes	Single layer at base, lateral	No	No	Ring	Brugerolle 1991
Summary	Long	9+2	Yes	Long	Transit- ional element present	Yes	?	250nm	Yes	Yes	MT arise laterally from sides of BB	Some taxa	No	Some taxa	
Mastigamoebidae “B”															
<i>Mastigamoeba simplex</i>	Long (>2× cell)	9+2	Yes, outer	200nm	No	Yes	Yes	250nm	Yes	Yes	Single layer at base, in longitudinal axis	No	No	No	Walker <i>et al.</i> 2001
<i>Mastigamoeba guttula</i>	Long (>2× cell)	9+2	?	200nm	No	Yes	?	250nm	Yes	Yes	Single layer at base, in longitudinal axis	No	No	No	This paper
Summary	Long	9+2	?	Short	No	Yes	?	250nm	Yes	Yes	MT arise longitudinally	No	No	No