Mitochondrial genomes of Amoebozoa

Natalya Bondarenko¹, Alexey Smirnov¹, Elena Nassonova¹,², Anna Glotova¹,² and Anna Maria Fiore-Donno³

¹ Department of Invertebrate Zoology, Faculty of Biology, Saint Petersburg State University, 199034 Saint Petersburg, Russia
² Laboratory of Cytology of Unicellular Organisms, Institute of Cytology RAS, 194064 Saint Petersburg, Russia
³ University of Cologne, Institute of Zoology, Terrestrial Ecology, 50674 Cologne, Germany

Summary

In this mini-review, we summarize the current knowledge on mitochondrial genomes of Amoebozoa. Amoebozoa is a major, early-diverging lineage of eukaryotes, containing at least 2,400 species. At present, 32 mitochondrial genomes belonging to 18 amoebozoan species are publicly available. A dearth of information is particularly obvious for two major amoebozoan clades, Variosea and Tubulinea, with just one mitochondrial genome sequenced for each. The main focus of this review is to summarize features such as mitochondrial gene content, mitochondrial genome size variation, and presence or absence of RNA editing, showing if they are unique or shared among amoebozoan lineages. In addition, we underline the potential of mitochondrial genomes for multigene phylogenetic reconstruction in Amoebozoa, where the relationships among lineages are not fully resolved yet. With the increasing application of next-generation sequencing techniques and reliable protocols, we advocate mitochondrial genomes as a promising tool for understanding evolutionary patterns in Amoebozoa.

Key words: Amoebozoa, protists, amoeba, mitochondria, mitochondrial genome

Abbreviations: mt – mitochondrial; coxI-3 - cytochrome oxidase subunit I, II, and III genes; cob – cytochrome b gene; atp1,6,8,9 – ATP synthase subunit 1,6,8,9 genes; nad1-7,9,11 – NADH dehydrogenase subunit 1-7, 9,11 and 4L genes; trRNA – transfer RNA genes; rrlL, rrlS – ribosomal RNA genes; ORF – open reading frames; PCGs – protein-coding genes; rps - small ribosomal subunit protein genes; rpl – large ribosomal subunit protein genes; SSU rRNA – small-subunit ribosomal RNA; LSU rRNA – large-subunit ribosomal RNA

doi:10.21685/1680-0826-2019-13-4-1

© 2019 The Author(s)

Protistology © 2019 Protozoological Society Affiliated with RAS
Introduction

One of the most captivating and still unresolved questions in the evolutionary biology of eukaryotes is the origin and evolution of the mitochondrial genome and the evolutionary processes by which a basic alpha-proteobacterial gene core has been modified during eukaryogenesis (Gray et al., 2001; Burger et al., 2003a; Cavalier-Smith, 2009; Gray, 2015). Our knowledge considerably advanced with the advent of comparative genomics of protistan mitochondria, showing that there was no “typical mitochondrial genome” (Gray et al., 2004). Since it is assumed that >90% of the initial bacterial gene complement must have been lost during the transition from endosymbiont to organelle, not surprisingly the mitochondrial genomes display a wide range of size, structure, codon-usage, and coding capacity (Adams et al., 2003; Burger et al., 2003b). Analysis of evolutionary trends over time demonstrated that gene transfers to the nucleus were non-linear, that they occurred in waves of exponential decrease, and that much of them took place comparatively early, independently, and at lineage-specific rates (Janouškovec et al., 2017). This process led to differential gene retention so that each lineage should possess a specific core of shared mitochondrial genes (Kannan et al., 2014). Examining the mitochondrial genomes in early diverging eukaryotes could help to solve the gene composition of the “last mitochondrial common ancestor” (LMCA) and how it evolved.

To date, the most gene-rich mitochondrial genomes (ca 66 genes) were found in the jakobids (excavates). Based on this, it was postulated that all other eukaryotes would possess only a subset of the jakobid genes (Kannan et al., 2014). However, two gene-rich mitochondrial genomes (of Ancoracysta twista and Diphylella rotans), have been found in unrelated taxa, suggesting a perhaps less linear evolutionary pattern (Kamikawa et al., 2016; Janouškovec et al. 2017). The large mitochondrial genome of Diphylella is especially intriguing in this respect since the species belongs to Opimoda - the other main domain of eukaryotes, rather than excavates (the latter belong to Diphoda domain) (Derelle et al., 2015). Opimoda includes three main lineages, Amoebozoa, animals and fungi, and among them, the basal Amoebozoa is the lineage where the shortage of data is especially sensitive. The molecular study of Amoebozoa is hampered by large discrepancies between evolutionary rates (Pawlowski and Burk, 2009), difficulties in culturing and identifying taxa, and therefore suffers from drastic undersampling (Kang et al., 2017).

The last general review about protistan mitochondrial genomes was that of Gray et al. (2004). The review by Miller (2014) is focused on Amoebozoa, but it is dedicated to the comparison of three genomes only (Acanthamoeba, Dictyostelium, and Physarum). We provide here an updated summary of the current knowledge about amoebozoan mitochondrial genomes, discussing their organization and gene diversity.

Available mitochondrial genomes in Amoebozoa

To date, complete mt genomes have been reported for 18 amoebozoan species (Fig. 1, Table 1). The first complete amoebozoan mt genome obtained using Sanger sequencing was that of Acanthamoeba castellanii strain Neff (Discosea: Centramoebia) (Burger et al., 1995). Successively, the mt genomes of the model organisms Dictyostelium discoideum (Ogawa et al., 2000) and Physarum polycephalum (Takano et al., 2001) were also obtained with Sanger sequencing. Focus on Dictyostelia as a model for intercellular communication provided, along with several nuclear genomes, the mt genomes of Dictyostelium fasciculatum (now Cavenderia fasciculata), D. citrinum and Polyphondylium pallidum (now Heterostelium pallidum) (Heidel and Göckner, 2008).

Later, genomes were obtained with next-generation sequencing of the total DNA. This expanded the number of mt genomes available in other major taxonomic divisions of Amoebozoa, i.e. three members of Variosea, Phalanisterium sp. (Pombert et al., 2013) and two Protostelium species (GenBank data). Among Tubulinea, the only species with sequenced mitochondrial genome remains Vermamoeba vermiformis (Echinamoebida) (Fukucikova and Lahr, 2016). Among Discosea, several strains of the pathogenic amoeba Balamuthia mandrillaris were added to the closely related Acanthamoeba (Greninger et al., 2015), as well as another species of Acanthamoeba, A. polyphaga (Karlyshev, 2019). Among other Discosea, the mt genomes of two marine species belonging to Dactylopodida were recently obtained, that of Neoparamoeba pemaqui-densis (Tanifuji et al., 2017) and that of Paramoeba aparasomata (Bondarenko et al., 2019b). Still in
Fig. 1. Schematic phylogenetic tree of Amoebozoa, drawn mainly from Kang et al. (2017), illustrating the number of genera per order (according to Adl et al. 2019 and http://eumycetozoa.com/data/index.php; last accessed Oct. 2019) and the coverage of mitochondrial genomes. Level of RNA editing: “high” = massive editing in many genes; “low” = few editing sites in several genes. Branch length not to scale.
Table 1. Mitochondrial genomes of amoebozoan species, GenBank accession number, and references. Genome size, number of protein-coding genes (PCGs) and number of tRNAs are provided. The level of RNA editing is indicated (when known) as “low” (few editing sites in several genes) or “extensive” (numerous editing sites in many genes). More detailed data on editing are provided only when available from the relevant publications. Translation tables are indicated according to the usual definitions (Table 1 – “standard code”; Table 4 – “mold, protozoan, and coelenterate mitochondrial code” and the “Mycoplasma/Spiroplasma code”). Sequences labeled as “unverified” by the NCBI staff are highlighted in gray and probable reasons are explained in the footnotes. Strain sources: ATCC – American Type Culture Collection (USA); CCAP – Culture Collection of Algae and Protozoa (UK); CCM of SPbSU – Culture collection of the Core facility center “Culturing of Microorganisms” of Saint Petersburg State University. Other strain designations are the author’s ones (according to GenBank data).

| Species and strain | Size of the genome (bp) | Number of protein coding genes | Number of tRNAs | Level and notes on RNA editing | GenBank accession number and translation table | Reference (according to GenBank database) |
|--------------------|-------------------------|--------------------------------|-----------------|-------------------------------|-----------------------------------------------|------------------------------------------|
| Acanthamoeba castellanii strain Neff; ATCC 30010 | 41,591 | 32 | 16 | Low (In several clusters of tRNA genes) | U12386 Table = 4 | Burger et al., 1995 |
| Acanthamoeba castellanii strain TN | 41,588 | 32 | NA | No editing | KX580904 Table = 4 | Greninger et al., 2016, unpublished |
| Acanthamoeba castellanii isolate BCP-EM3VF21-1 | 39,205 | 36 | 13 | Low (5 sites in tRNA genes) | KT185628 Table = 4 | Fučíková and Lahr, 2016 |
| Acanthamoeba polyphaga strain Lnc Ap-1 | 39,215 | 32 | 15 | No editing | KP054475.2 Table = 4 | Karlyshev, 2019 |
| Dictyostelium discoideum strain AX3, partially X22* | 55,564 | 36 | 18 | Low (in tRNA genes) | AB000109 Table = 1 | Ogawa et al., 2000 |
| Dictyostelium fasciculatum strain SH3 | 54,563 | 40 | 17 | No editing | EU275727 Table = 1 | Heidel and Glöckner, 2008 |
| Dictyostelium citrinum | 57,820 | 43 | 19 | No editing | DQ36395.4 Table = 1 | Heidel and Glöckner, 2008 |
| Physarum polycephalum | 62,862 | 20 | No data | Extensive (in PCG, ORFs and tRNAs) | AB027295 Table = 1 | Takano et al., 2001 |
| Polysphondylium pallidum strain CK8 | 47,653 | 35 | 19 | No data | AY700145 Table = 1 | Burger et al., 2004, unpublished |
| Polysphondylium pallidum strain PNS00 | 48,042 | 40 | 20 | No editing | EU275726 Table = 1 | Heidel and Glöckner, 2008 |
| Phalansterium sp. strain PJK-2012 | 53,614 | 19 | 24 | Extensive (in tRNA genes) | KC121006 Table = 1 | Pombert et al., 2013 |
| Balamuthia mandrillaris strain GAM-19 | 41,570 | 33 | 13 | No editing | KT030673 Table = 4 | Greninger et al., 2015 |
| Balamuthia mandrillaris strain ok1 | 42,823 | 33 | 13 | No editing | KT030672 Table = 4 | Greninger et al., 2015 |
| Balamuthia mandrillaris strain V188 | 42,177 | 33 | 13 | No editing | KT030670 Table = 4 | Greninger et al., 2015 |
| Balamuthia mandrillaris strain v039 | 39,996 | 32 | 13 | No editing | KT175741 Table = 4 | Greninger et al., 2015 |
| Balamuthia mandrillaris strain 2046 | 41,656 | 33 | 13 | No editing | KT175740 Table = 4 | Greninger et al., 2015 |
| Balamuthia mandrillaris strain CDC-V039 | 39,894 | No data | No data | No data | CM003363 no translation provided | Detering and Kiderlen ,2015, unpublished |
| Balamuthia mandrillaris strain BeSh | 42,177 | 33 | 13 | No data | NC_027736 Table = 4 | Greninger et al., 2015, unpublished |
| Vermamoeba vermiformis isolate BCP-EM3VF21-2 | 52,068 | 37 | 25 | No editing | KT185627 Table = 4 | Fučíková and Lahr, 2016 |
| Vermamoeba vermiformis (as Hartmannella) | 51,645 | 37 | 25 | No editing | GU828005 Table = 4 | Bullerwell et al., 2010 |
| Neoparamoeba pemaquidensis CCAP 1560/4 (as Paramoeba) | 48,522 | 34 | 19 | No editing | KX611830 Table = 4 | Tanifuji et al., 2017 |
Discosea, mt genomes of four species of Vannellida were recently sequenced: *Vannella croatica* (Bondarenko et al., 2018a), *V. simplex* (Bondarenko et al., 2018b), *Clydonella sawyeri* (Bondarenko et al., 2018c) and *Paravannella minima* (Bondarenko et al., 2019a). Among them the mt genome of *Vannella croatica* was obtained into two steps: the purified mitochondrial DNA that was isolated by pulsed-field gel electrophoresis was sequenced using NGS. Further, the obtained contig was checked by Sanger sequencing of the genomic DNA using a set of custom-made primers. The results of the two sequencing methods were almost identical (Bondarenko et al., 2018a).

### Differences in size and gene content

All mt genomes of Amoebozoa sequenced to date are circular and display little size variation, from 28.9 to 62.8 kb, within the range found in animals, i.e. 11 to 77 kb (Lavrov and Pett, 2016) and in fungi, 16–110 kb (database of curated fungal mt genomes: http://mitofun.biol.uoa.gr/, last accessed Oct. 2019). In contrast, mitochondrial genomes of plants are significantly larger and range from 180 to 600 kb (Lynch et al., 2006).

A typical amoebozoan mitochondrial genome contains on average 30 protein-coding genes (Fig. 2). They are involved in the electron transport chain (three cytochrome oxidase subunits, ten NADH dehydrogenase subunits and apocytochrome b), ATP synthesis (seven ATP synthase subunits) and several mitochondrial ribosomal protein genes (rpL and rpS) (Burger et al., 1995; Ogawa et al., 2000; Takano et al., 2001; Pombert et al., 2013; Greninger et al., 2015; Fucíková and Lahr, 2016; Tanifuji et al., 2017 Bondarenko et al., 2018a, 2018b, 2018c, 2019a, 2019b). Similar to metazoans, most of the amoebozoan mt genomes have duplicated tRNA genes for serine and leucine and in some cases for methionine (*Vannella simplex, V. croatica, Clydonella sawyeri, Paravannella minima, Phalansterium sp., Vermamoeba vermiformis, Physarum polycephalum, Dictyostelium citrinum*), isoleucine (*Vermamoeba vermiformis, Dictyostelium discoideum, D. citrinum, Acanthamoeba castellanii strain TN, A. polyphaga*), phenylalanine (*Polysphondylium pallidum*), lysine (*Vannella simplex, V. croatica*) and arginine (*Clydonella sawyeri, Paravannella minima, Vermamoeba vermiformis*) (Heidel and Glöckner, 2008; Bondarenko et al., 2018a, 2018b, 2018c, 2019a, 2019b). Along with the genes of known function, there are open reading frames (ORF) without recognizable counterpart in other mt genomes (Fig. 2). These ORFs might code for additional proteins whose sequences have diverged too much to be identified as such, and/or because of the incompleteness of the reference database, especially lacking representatives of free-living protists (del Campo et al., 2014). They also may represent a captured DNA, such as mitochondrial plasmids (Nakagawa et al., 1998). Their number depends on the species: *Acanthamoeba castellanii* isolate BCP-EM3VF21-1 is devoid of them, while *Protostelium mycophagum* has the maximum number of 19. The length of

### Table 1. (Continuation).

| Species and strain | Size of the genome (bp) | Number of protein coding genes | Number of tRNAs | Level and notes on RNA editing | GenBank accession number and translation table | Reference (according to GenBank database) |
|--------------------|------------------------|-------------------------------|-----------------|-------------------------------|-----------------------------------------------|--------------------------------------------|
| *Paramoeba aparasomata* | 46,254                 | 31                            | 19              | No editing                    | MK518072 Table = 4                             | Bondarenko et al., 2019                  |
| *Paravannella minima* | 53,464                 | 30                            | 23              | No editing                    | MH910097 Table = 1                             | Bondarenko et al., 2019                  |
| *Vannella croatica* | 28,933                 | 12                            | 16              | Extensive                     | MF508648 Table = 4                             | Bondarenko et al., 2018a                  |
| *Vannella simplex* | 34,145                 | 27                            | 17              | Extensive                     | MF496657 Table = 4                             | Bondarenko et al., 2018b                  |
| *Clydonella sawyeri* | 31,131                 | 17                            | 21              | Low (in 5 PCG)                | MH094141 Table = 4                             | Bondarenko et al., 2018c                  |
| *Protostelium mycophagum* | 48,607                | 16                            | 24              | No data                       | KY775056 Table = 4                             | Glöckner, 2017, unpublished               |
| *Protostelium sp.* | 44,490                 | 16                            | 22              | No data                       | KY775057 Table = 4                             | Glöckner, 2017, unpublished               |

* The mt genome of *Dictyostelium discoideum* is a composite, originating from two strains
** This *B. mandrillaris* strain mt genome is identical to that of the strain V451 KT030670, and has not yet been subject to final NCBI review.
*** Labeled in GenBank as “unverified” because the type of editing is not known.
**** Labeled in GenBank as “unverified”, presumably because of incomplete annotation.
Fig. 2. Linearized maps of 15 amoebozoan mitochondrial genomes. The same functional groups of genes share the same colour code. The map illustrates gene order and functional groups, but not the mt genome size and relative length of genes, so it is drawn not to scale.
ORF may vary from hundreds to several thousand nucleotides. The longest known ORF belongs to *Paramoeba aparasomata* and consists of 5,871 bp.

The protein-coding genes in amoebozoan mt genomes are predominantly initiated with the start codons AUG, AUU or AUA and terminated with UAA or UAG (with UGA coding for tryptophan) that are also used in the unrelated protistan lineages Excavata: Euglenozoa and Alveolata: Ciliata. Some amoebozoans use only AUG as starting codon (which is the standard genetic code for nuclear genes): *Physarum polycephalum, Polysphondylium pallidum, Phalanterium sp.*, and *Paravannella minima* (Table 1).

There is a huge variability in non-coding regions in the amoebozoan mt genomes, usually ranging in length from 50 to 2,500 bp. The longest one reaches 11,701 bp and was found in *Physarum polycephalum* (Takano et al. 2001). In some mt genomes, e.g. in *Paravannella minima* and *Vannella croatica*, the non-coding regions are scarce and short: they do not exceed 250 bp in length (Bondarenko et al., 2018a, 2019a).

### The diversity of gene composition and gene order

Despite the conserved size and similarity in the overall gene content of the amoebozoan mt genome, gene composition can significantly vary even between closely related species (Fig. 2). For example, genes encoding the ATP synthase subunits alpha and C are missing in all known mt genomes, except in that of *Balamuthia mandrillaris* (Fig. 2), which makes it stand out from its relative *Acanthamoeba* spp. Similarly, the genes *en1, en2* were identified only in two dictyostelids (*Polysphondylium pallidum* strain PN500 and *Dictyostelium fasciculatum* strain SH3) (Heidel and Glöckner, 2008). The *tufA* gene was found in *Vermamoeba vermiformis* (Fig. 2). In addition to variations in protein-coding genes, the number of tRNA genes can range from 13 to 25 among different amoebozoan species (Fig. 2).

In contrast to animals, where mt genes are predominantly encoded on both strands (Burger et al., 2003a), in many species of Amoebozoa the number of genes encoded in the minus strand is always low and concerns mainly tRNA genes. Genes encoded on the minus strand were found in Discosea (*Vannella croatica, Paravannella minima, Clydonella sawyeri, Acanthamoeba castellanii* isolate BCP-EM3VF21-1, and *Neoparamoeba pemaquidensis*) and in Evosea: *Physarum polycephalum* and *Dictyostelium mycophagum* (see Table 1 for relevant references).

While closely related amoebozoan species show almost complete synten in their mt genomes (Bondarenko et al., 2019b; Karlyshev, 2019), the order of genes is altered and the same pattern is no more recognizable with increasing phylogenetic distance (Bondarenko et al., 2018a, 2018b, 2018c). The genome alignment made for available mt genomes of Vannellida, Dactylopodida and *Acanthamoeba* (Fig. 3) clearly indicated that within the same genus the level of synten between genomes is relatively high (e.g. within the genus *Vannella* or the genus *Acanthamoeba*); the same occurs between phylogenetically closely related genera (*Neoparamoeba, Paramoeba*). However, it is much lower between more distant lineages. For example, the genera *Clydonella* and *Paravannella* are separated by significant evolutionary distance from *Vannella* (Kudryavtsev, 2014), and the level of synten between these genera and *Vannella* is much lower. The homologies between orders are always low, and even gene blocks recognized as homologous may differ in their gene content and their position in the genome (Fig. 3). The evolution of the mt genome structure is potentially a powerful marker to test the phylogenetic hypotheses, however, at present, its power is limited by the low number of available genomes.

Besides changes in the gene order, amoebozoan mt genomes show size differences due to introns and duplications of protein-coding genes. Introns were found in the mt genomes of *Acanthamoeba castellani* strain Neff and strain TN (Burger et al., 1995), *Dictyostelium discoideum, D. fasciculatum, D. citrinum* (Ogawa et al., 2000; Heidel and Glöckner, 2008) and all strains of *Balamuthia mandrillaris* (Greninger et al., 2015). In *Dictyostelium discoideum, Physarum polycephalum, Balamuthia mandrillaris* and *Acanthamoeba castellani* strain Neff the introns have been characterized as Group I introns. In *Dictyostelium discoideum* mt genome Group I introns encode homing endonucleases, active in self-splicing (Ogawa et al., 2000; Lang et al., 2007; Greninger et al., 2015). An extra source of polymorphism may be a recombination of mt DNA in presence of plasmids, like mF plasmid, shown to enhance this process in *Physarum polycephalum* (Nomura et al., 2005).

Duplication of protein-coding genes is not a common event in the mt genomes of Amoebozoa. It is known in three amoebozoan species only. In
Fig. 3. Alignment of several amoebozoan mitochondrial genomes made by Mauve genome aligner (http://darlinglab.org/mauve/mauve.html). Potentially homologous gene blocks share the same colour. The connecting lines show how their respective positions vary within Discosea. Note that recognition of homologous fragments by Mauve is done ex novo, so that it may not correspond to available annotations of mt genomes. Only protein-coding genes are listed. Please refer to Fig. 2 and GenBank annotations for exact gene order and length, including ORFs and tRNAs.
Neoparamoeba pemaquidensis and Phalansterium sp. the cob gene (part of oxidative phosphorylation chain) is duplicated (Pombert et al., 2013; Tanifuji et al., 2017), and in Polysphondylium pallidum the duplicated gene is rps3 (participates in the ribosomal complex and DNA repair) (Burger et al., 2004 unpublished, source - GenBank data AY700145).

RNA editing in amoebozoan mitochondrial genomes

While RNA editing is observed across many taxa and all domains of life, it involves fundamentally different modes and underlying mechanisms, suggesting that it is a derived trait within the lineages in which it is found, rather than a primitive feature inherited from a common evolutionary ancestor (Horton and Landweber, 2000; Gray, 2012). RNA editing converts primary RNA transcripts into mature and functional transcripts and occurs in mitochondria, plastids or nuclei of a wide range of eukaryotes (Burger, 2016; Moreira et al., 2016; Valach et al., 2017). It can affect both protein-coding RNAs and ribosomal and transfer RNAs (Yang et al., 2017). Among protists, massive editing of RNAs is observed in the kinetoplast of trypanosomes (Gölinger, 2012), in Heterolobosea (Yang et al., 2017) and in diplonemids (Valach et al., 2017). The latter one has attracted much attention in the last years since it involves the assembly of fragmented genes (Valach et al., 2017).

There are little data on the distribution and occurrence of editing among basal clades of Amoebozoa (Fig. 1). However, Amoebozoa contains a group of organisms that perhaps provides a greater challenge to RNA editing than any other living organisms. These are the plasmodial slime molds (Myxogastria), mostly studied through the model organism Physarum polycephalum. Of them, Physarum carries out one of the most complex sets of RNA editing events yet described (Houtz et al., 2018 and citations therein) — with the site-specific insertion of over 1,300 nucleotides (sometimes dinucleotides) and base conversions.

Our recent studies revealed the existence of mitochondrial RNA editing in amoebae of the order Vannellida, although with an interesting distribution along its phylogeny. We observed extensive RNA editing in two crown species — Vannella croatica and Vannella simplex, while members of two other, more basal, genera — Clydonella and Paravannella — had little or no editing (Bondarenko et al., 2018a, 2018b, 2018c, 2019a).

In some Amoebozoa lineages, there are no sequenced mt genomes, but sequences of the CoxI gene are available. We looked for evidence of editing by translating the gene using Expasy (https://web.expasy.org/translate/, last accessed Nov. 2019). There was no evidence of editing in the genus Korotnevella (Zlatogurski et al., 2016), in Parvamoeba rugata (JN202434 sequence), Cochliopodium pentatrilobatum (KC489470), C. megatetrastylus (KC747719), C. actinophorum (CQ354207) and Squamamoeba japonica (JN6380333). Certainly, results based on translation of a single gene are no proof of the absence of editing in the entire lineage. In contrast, numerous stop-codons in all possible variants of translation were found in the sequences of the Cox I gene of testate amoebae (Kosakyan et al., 2011), suggesting a scattered distribution of RNA editing in the order Arcellinida. These data further advocate for the independent origin of RNA editing in different amoebozoan lineages.

Mitochondrial genomes as a potential tool to reconstruct amoebozoan phylogeny

Orthologous mitochondrial protein-coding genes are valuable tools for multigene phylogenies (Kannan et al., 2014; Janouškovec et al., 2017). In Amoebozoa especially, mt genomes would represent a valuable complement to full genomes, since the latter are notoriously difficult to assemble due to their large size, high repeat content and numerous and large introns (Clarke et al., 2013; Detering et al., 2015; Schaap et al., 2015). Mitochondrial genomes provide up to 37 genes widely distributed in eukaryotes that can be used for inferring evolutionary trends (Janouškovec et al., 2017); at least 25 of them are present in Amoebozoa (Bondarenko et al., 2019c). Among them, there are 13 genes of the respiratory chain (e.g. nad1–6, nad4L, cox1–3, cytb, atp6, and atp8), two subunits of the ribosomal RNA and transfer RNAs. In particular, the cox1 gene — a recognized DNA barcode for many groups of animals (e.g. butterflies, birds) has been used in the phylogeny and systematics of some taxa of Amoebozoa (Nassonova et al., 2010; Kosakyan et al., 2012; Zlatogursky et al., 2016) and thus shows a potential as a phylogenetic marker for the whole group.
As an example, we performed a phylogenetic analysis of an alignment consisting of 19 mt genes of 17 amoebozoans. The number of analysed amino-acid positions varied from 3,685 in *Clydonella sawyeri* to 5,195 in *Dictyostelium citrinum* mt genomes. Differences in the number of retained positions depended on genome length, annotation quality and level of homology between genes. The resulting tree showed highly supported branches for all groups recovering well-established monophyletic taxa (Fig. 4), i.e. Vannellida, Centramoebida, Dactylopodida, Dictyostelia. The Variosea were recovered as a paraphyletic group, with both sequences of *Protostelium* robustly grouping together. However, the basal branching of our tree and the relative positions of the three classes are weakly supported or are contradictory between Maximum Likelihood and Bayesian analyses. Only the class Discosea is monophyletic, as in most trees (Cavalier-Smith et al., 2015; Kang et al., 2017), while Evosea is paraphyletic to it (compare with Fig. 1). This configuration is probably an artifact due to the current unbalanced representation of taxa. Any robust conclusions will be possible only after obtaining more amoebozoan mitochondrial genomes.

**Acknowledgements**

Supported with RSF 17-14-01391 grant. The present study utilized equipment of the Core facility centres “Development of molecular and cell technologies”, “Biobank”, “Computing Centre SPSU” and “Culture Collection of Microorganisms” of the research park of Saint Petersburg State University.
References

Adams K.L. and Palmer J.D. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol. Phyl. Evol. 29, 380–395.

Bondarenko N., Nassonova E., Mijanovich O., Glotova A., Kamyshtskaya O., Kudryavtsev A., Masharsky A., Polev D. and Smirnov A. 2018a. Mitochondrial genome of Vannella croatica (Amoebozoa, Discosea, Vannellida). J. Eukar. Microbiol. 65, 820–827.

Bondarenko N., Glotova A., Nassonova E., Masharsky A., Kudryavtsev A. and Smirnov A. 2018b. The complete mitochondrial genome of Vannella simplex (Amoebozoa, Discosea, Vannellida). Europ. J. Protistol. 63, 83–95.

Bondarenko N., Glotova A., Nassonova E., Masharsky A., Kudryavtsev A. and Smirnov A. 2018c. The complete mitochondrial genome of Clydonella sawyeri (Amoebozoa, Discosea, Vannellida). Protistology 12, 47–54.

Bondarenko N., Volkova E., Masharsky A., Kudryavtsev A. and Smirnov A. 2019a. The complete mitochondrial genome of Paravannella minima (Amoebozoa, Discosea, Vannellida). Europ. J. Protistol. 68, 80–87.

Bondarenko N., Volkova E., Masharsky A., Kudryavtsev A. and Smirnov A. 2019b. A comparative characterization of the mitochondrial genomes of Paramoeba aparasomata and Neoparamoeba pemaquidensis (Amoebozoa, Paramoebidae). J. Eukar. Microbiol. https://doi.org/10.1111/jeu.12767.

Gray M. W. 2012. Evolutionary origin of RNA editing. Biochemistry. 51, 5235–5242.

Gray M. W. 2015. Mosaic nature of the mitochondrial proteome: implications for the origin and evolution of mitochondria. Proc. Natl. Acad. Sci. USA 112, 10133–10138.

Gray M. W., Burger G. and Lang B.F. 2001. The origin and early evolution of mitochondria. Genome Biol. 2. doi: 10.1186/gb-2001-2-6-reviews1018.

architecture in unicellular relatives of animals. Proc. Natl. Acad. Sci. USA, 100, 892–897.

Burger G. 2016. Non-functional genes repaired at the RNA level. C. R. Biol. 339, 289–295.

Cavalier-Smith T. 2009. Predation and eukaryote cell origins: A coevolutionary perspective. Int. J. Biochem. Cell. Biol. 41, 307–322.

Cavalier-Smith T., Fiore-Donno A.M., Chao E., Kudryavtsev A., Berney C., Snell E.A. and Lewis R. 2015. Multigene phylogeny resolves deep branching of Amoebozoa. Mol. Phylogenet. Evol. 83, 293–304.

Clarke M., Lohan A. J., Liu B., Lagkouvardos I., Roy S., Zafar N., Bertelli C., Schild C., Kianianmomeni A., Bürglin T.R., Frech C., Turcotte B., Kopec K.O., Synnott J.M., Choo C., Paponov I., Finkler A., Heng Tang C.S., Hutchins A.P., Weinmeier T., Rattei T., Chu J.S., Gimenez I., Irimia M., Rigden D.J., Fitzpatrick D.A., Lorenzo-Morales J., Bateman A., Chiu C.H., Tang P., Hegemann P., Fromm H., Raoul D., Greub G., Miranda-Saavedra D., Chen N., Nash P., Ginger M.L., Horn M., Schaap P., Caler L. and Loftus B.J. 2013. Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. Genome Biol. 14:R11. doi:10.1186/gb-2013-14-2-r11.

Burger G. 2016. Non-functional genes repaired at the RNA level. C. R. Biol. 339, 289–295.

Bondarenko N., Volkova E., Masharsky A., Kudryavtsev A. and Smirnov A. 2019b. A comparative characterization of the mitochondrial genomes of Paramoeba aparasomata and Neoparamoeba pemaquidensis (Amoebozoa, Paramoebidae). J. Eukar. Microbiol. https://doi.org/10.1111/jeu.12767.

Del Campo J., Sieracki M.E., Molestina R.E., Keeling P.J., Massana R. and Ruiz-Trillo I. 2014. The others: our biased perspective of eukaryotic genomes. Trends Ecol. Evol. 29, 251–259.

Derelle R., Torruella G., Kliměš V., Brinkmann H., Kim E., Vlček Č., Lang F.B. and Eliaš M. 2015. Bacterial proteins pinpoint a single eukaryotic root. Proc. Natl. Acad. Sci. USA. 112, e693–e699.

Detering H., Aebischer T., Dabrowski P.W., Radonic A., Nitsche A. and Renard B. 2015. First draft genome sequence of Balamuthia mandrillaris, the causative agent of amoebic encephalitis. Genome Announc. 3. doi: 10.1128/genomeA.01013-15.

Fučíková K. and Lahr D. J. 2016. Uncovering cryptic diversity in two amoebozoan species using complete mitochondrial genome sequences. J. Eukar. Microbiol. 63, 112–122.

Gray M. W. 2012. Evolutionary origin of RNA editing. Biochemistry. 51, 5235–5242.

Gray M. W. 2015. Mosaic nature of the mitochondrial proteome: implications for the origin and evolution of mitochondria. Proc. Natl. Acad. Sci. USA 112, 10133–10138.

Gray M. W., Burger G. and Lang B.F. 2001. The origin and early evolution of mitochondria. Genome Biol. 2. doi: 10.1186/gb-2001-2-6-reviews1018.
Gray M.W., Lang F.B. and Burger G. 2004. Mitochondria of protists. Annu. Rev. Genet. 38, 477–524.

Göringer U.H. 2012. ‘Gestalt,’ composition and function of the Trypanosoma brucei editosome. Annu. Rev. Microbiol. 66, 65–82.

Greninger A.L., Messcar K., Dunnebacke T., Naccache S.N., Federman S., Bouquet J., Mirsky D., Nomura Y., Yagi S., Glaser C., Vollmer M., Press C.A., Kleinschmidt–DeMasters B.K., Dominguez S.R. and Chiu C.Y. 2015. Clinical metagenomic identification of Balamuthia mandrillaris encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing. Genome Med. 1; doi: 10.1186/s13073-015-0235-2.

Heidel A., J. and Glöckner G. 2008. Mitochondrial genome evolution in the social amoebae. Mol. Biol. Evol. 25, 1440–1450.

Horton T.L. and Landweber L.F. 2000. Evolution of four types of RNA editing in myxomycetes. RNA 6, 1339–1346.

Houtz J., Cremona N. and Gott J.M. 2018. Editing of mitochondrial RNAs in Physarum polycephalum. In: RNA metabolism in mitochondria (Eds: Cruz-Reyes J. and Gray M.W.). Springer, Cham. pp. 199–222.

Janouškovec J., Tikhonenkov D.V., Burki F., Howe A.T., Rohwer F. L., Mylnikov A.P. and Keeling P.J. 2017. A new lineage of eukaryotes illuminates early mitochondrial genome reduction. Curr. Biol. 27, 3717–3724.

Kamikawa R., Shiratori T., Ishida K., Miyashita H. and Roger A.J. 2016. Group II intron-mediated trans-splicing in the gene-rich mitochondrial genome of an enigmatic eukaryote, Diphylelia rotans. Genome Biol. Evol. 8, 458–466.

Kosakyan A., Heger T.J., Leander B.S., Todoro M., Mitchell E.A.D. and Lara E. 2012. COI barcoding of nebelid testate amoebae (Amoebozoa: Arcellinida): extensive cryptic diversity and redefinition of family Hyalospheniidae Schultze. Protist. 163, 415–434.

Kudryavtsev A. 2014. Paravannella minima n.g. n.sp. (Discosea, Vannellidae) and distinction of the genera in the vannellid amoebae. Europ. J. Protistol. 50, 258–269.

Lang B.F., Laforest M.J. and Burger G. 2007. Mitochondrial introns: a critical view. Trends Genet. 23, 119–125.

Le P., Fisher P.R. and Barth C. 2009. Transcription of the Dictyostelium discoideum mitochondrial genome occurs from a single initiation site. RNA 15, 2321–2330.

Lavrov D.V. and Pett W. 2016. Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. Genome. Biol. Evol. 8, 2896–2913.

Miller D. 2014. Mitochondrial genomes in Amoebozoa. In: Molecular life sciences. Springer, New York. doi.org/10.1007/978-1-4614-6436-5_114-2.

Moreira S., Valach M., Aoulad-Aissa M., Otto C. and Burger G. 2016. Novel modes of RNA editing in mitochondria. Nucleic Acids Res. 44, 4907–4919.

Nakagawa C.C., Jones E.P. and Miller D.L. 1998. Mitochondrial DNA rearrangements associated with mF plasmid integration and plasmoidal longevity in Physarum polycephalum. Curr. Genet. 33, 178–187.

Nassonova E., Smirnov A., Fahrni J. and Pawlowski J. 2010. Barcoding amoebae: comparison of SSU, ITS and COI genes as tools for molecular identification of naked lobose amoebae. Protist. 161, 102–115.

Nomura H., Moriyama Y. and Kawano S. 2005. Rearrangements in the Physarum polycephalum mitochondrial genome associated with a transition from linear mF-mtDNA recombinants to circular molecules. Curr. Genet. 47, 100–110.

Ogawa S., Yoshino R., Angata K., Iwamoto M., Mi M., Kuroe K., Matsuo K., Morio T., Urushihara H., Yanagisawa K. and Tanaka Y. 2000. The mitochondrial DNA of Dictyostelium discoideum: complete sequence, gene content and genome organization. Mol. Gen. Genet. 263, 514–519.
Pombert J.-F., Smirnov A., James E. R., Janouškovec J., Gray M. W. and Keeling P. J. 2013. The complete mitochondrial genome from an unidentified *Phalansterium* species. Protist Genomics. 1, 25–32.

Schaap P., Barrantes I., Minx P., Sasaki N., Anderson R. W., Bénard M., Biggar K. K., Buchler N. E., Bundschuh R., Chen X., Fronick C., Fulton L., Goldner G., Jahn N., Knoop V., Landweber L. F., Marie C., Miller D., Noegel A. A., Peace R., Pierron G., Sasaki T., Schallenberg-Rüdinger M., Schleicher M., Singh R., Spaller T., Storey K.B., Suzuki T., Tomlinson C., Tyson J.J., Warren W. C., Werner E. R., Werner-Felmayer G., Wilson R.K., Takano H., Abe T., Sakurai R., Moriyama Y., Miyazawa Y., Nozaki H., Kawano S., Sasaki N. and Kuroiwa T. 2001. The complete DNA sequence of the mitochondrial genome of *Physarum polycephalum*. Mol. Gen. Genet. 264, 539–545.

Tanifuji G., Cenci U., Moog D., Dean S., Nakayama T., David V., Fiala I., Curtis B.A., Sib-bald S. J., Onodera N. T., Colp M., Flegontov P., Johnson-MacKinnon J., McPhee M., Inagaki Y., Hashimoto T., Kelly S., Gull K., Lukeš J. and Archibald J. M. 2017. Genome sequencing reveals metabolic and cellular interdependence in an amoeba-kinetoplastid symbiosis. Sci. Rep. 7, 11688.

Valach M., Moreira S., Hoffmann F., Stadler P. F. and Burger G. 2017. Keeping it complicated: Mitochondrial genome plasticity across diplonemids. Sci. Rep. 7, 14166.

Yang J., Harding T., Kamikawa R., Simpson A.G.B. and Roger A.J. 2017. Mitochondrial genome evolution and a novel RNA editing system in deep-branching heteroloboseids. Genome Biol. Evol. 9:1161–1174.

Zlatogursky V.V., Kudryavtsev A., Udalov I. A., Bondarenko N., Pawlowski J. and Smirnov A. 2016. Genetic structure of a morphological species within the amoeba genus *Korotnevella* (Amoebozoa: Discosea), revealed by the analysis of two genes. Europ. J. Protistol. 56, 102–111.

Address for correspondence: Natalya Bondarenko. Department of Invertebrate Zoology, Faculty of Biology, Saint Petersburg State University, Universitetskaya nab. 7/9, 199034 Saint Petersburg, Russia; e-mail: n.bondarenko@spbu.ru.