Protistology 15 (4), 304–311 (2021)

Protistology

The complete mitochondrial genome of an unusual strain of tiny vannellid amoeba (Amoebozoa, Discosea, Vannellida) isolated from the Niagara River (Canada)

Natalya I. Bondarenko[1], Anna A. Glotova[1], Elena S. Nassonova[2], Alexey E. Masharsky[3], and Alexey V. Smirnov[1]

[1] Department of Invertebrate Zoology, Faculty of Biology, St. Petersburg State University, 199034 St. Petersburg, Russia
[2] Laboratory of Cytology of Unicellular Organisms, Institute of Cytology RAS, 194064 St. Petersburg, Russia
[3]Core Facility Centre Biobank, St. Petersburg State University, Stary Peterhof 198504 St. Petersburg, Russia

| Submitted November 13, 2020 | Accepted December 9, 2020 |

Summary

We present a complete sequence and describe the mitochondrial genome organization of the strain of small vannellid amoeba isolated from the Niagara River (Canada) in the year 2007. The circular mitochondrial DNA of this strain has 52,924 bp in length and contains 30 protein-coding genes, two ribosomal RNAs, 25 transfer RNAs, and 13 open reading frames. It is the second in length mt genome among amoebae of the order Vannellida (Amoebozoa, Discosea). In contrast with the shorter mitochondrial genomes of crown vannellids, it shows no evidence of RNA editing. This finding supports the hypothesis on the independent origin of editing in the phylogenetic lineage corresponding to the order Vannellida (Amoebozoa, Discosea).

Key words: Amoebozoa, Vannellidae, *Ripella*, mitochondrion, mitochondrial genome

Abbreviations: MT – mitochondrial; *cox1-3*, – cytochrome oxidase subunit I, II, and III genes; *cob* – cytochrome b gene; *atp9* – ATP synthase subunit 9 gene; *nad1-7, 9, 11* – NADH dehydrogenase subunit 1-7, 9, 11 and 4L genes; *tRNA* – transfer RNA genes; *rrnL, rrnS* – ribosomal RNA genes; ORF – open reading frames; PCGs – protein-coding genes; *rps* - small ribosomal subunit protein genes; *rpl* – large ribosomal subunit protein genes; CDS – coding DNA sequence

doi:10.21685/1680-0826-2021-15-4-8

Introduction

Mitochondrial genomes (MT genomes) of Amoebozoa remain poorly studied despite significant efforts invested in this field in recent time (reviewed by Bondarenko et al., 2019b). Among amoebozoan lineages, the best taxonomic sampling in MT genome studies is achieved among the order Vannellida (Bondarenko et al., 2018a, 2018b, 2018c, 2019a). This group of amoebae demonstrated very different sizes of the mitochondrial genomes. The genus *Vannella*, the crown phylogenetic lineage of this order (Smirnov et al., 2007), has the length of MT genomes 29-34 kbp and shows extensive post-translational editing (Bondarenko et al., 2018a, 2018b). The MT genome of *Clydonella sawyeri*, which belongs to a more basal lineage of Vannellida (Kudryavtsev and Volkova, 2018), is 31 kbp in length but shoes low level of editing in five genes only. Simultaneously, the only studied basal lineage, represented with *Paravannella minima*, demonstrates a much longer genome, reaching up to 53 kbp, with no editing (Bondarenko et al., 2019a). This finding suggests an independent origin of editing in this branch of amoebozoan tree; however, a limited taxonomic sample does not allow one to locate the point of the origin of editing.

The present paper reports data on the mitochondrial genome of a strain of a small vannellid amoeba, belonging to a new, independent basal lineage among the order Vannellida and possessing a long mitochondrial genome (52 kbp), with no editing. This finding supports the idea of the independent origin of RNA editing in this amoebozoan lineage.

Material and methods

The culture of an amoeba, designated further as "Niagara strain," was isolated from the sample of the top layer of bottom sediment of theNiagara River in the area between the Niagara Falls and Ontario Lake, along the Niagara Park-way (Canada, grid reference 43.191408N, – 79.054652E). In a forthcoming paper, this strain will be described as "Simripella niagara" (should not be considered here as a taxonomic mentioning, and provided here in order to link the present data with a forthcoming publication by another author). According to the personal communication by Dr. Alexander Kudryavtsev, in the SSU phylogenetic tree, it forms an independent lineage, basal to most other vannellids. The strain is illustrated in Fig. 1.

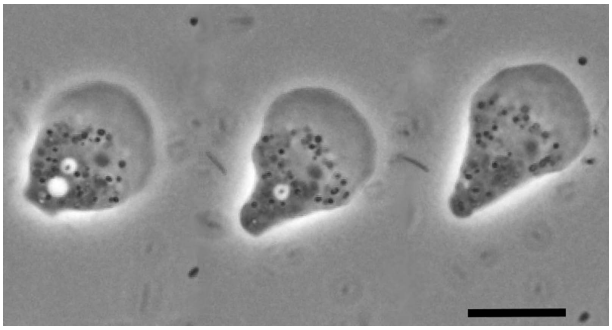

Fig. 1. Light-microscopic images of "Niagara strain." Phase contrast. Scale bar: 10 μm.

Amoebae were cultured in 90 mm Petri dishes filled with Millipore-sterilized (0.2 μm pore) artificial seawater (25‰) and one wheat grain per dish. Cells were concentrated and washed to remove bacteria as described earlier (Bondarenko et al., 2018a). Total DNA isolation was performed using NucleoSpin Tissue Kit (Macherey-Nagel, Germany) according to the manufacturer's instructions. Approximately 40,5 million reads with a length of 150 bp were obtained using HiSeq 2500 sequencing system (Illumina). Quality control check of raw sequence data was performed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), SPAdes assembler was used for *de novo* mitochondrial genome assembly (Bankevich et al., 2012). An annotation of mitochondrial genome sequence was performed using the MITOS web server (Bernt et al., 2013a). Artemis was used to visualize annotation files, manual correction of gene boundaries, and open reading frames (ORFs) search (version 16.0; Rutherford et al., 2000). All protein-coding genes (PCGs) boundaries were verified by manual comparison with the orthologs in other amoebozoans. Genes coding tRNAs were positioned with tRNAcan-SE Search Server v.1.21 (Lowe and Eddy, 1997). Strand asymmetry was calculated using the formulae: AT skew= [A-T]/ [A+T] and GC skew= [G-C]/ [G+C], for the H-strand (Perna and Kocher, 1995). The physical map was generated by our original script written in Python.

Results and discussion

The mitochondrial genome of "Niagara strain" is a double-stranded circular DNA molecule with a length of 52,924 bp (Fig. 2). It is another sizeable MT genome (over 40 kbp) among amoebae belonging to the family Vannellidae (see Bondarenko et al.,

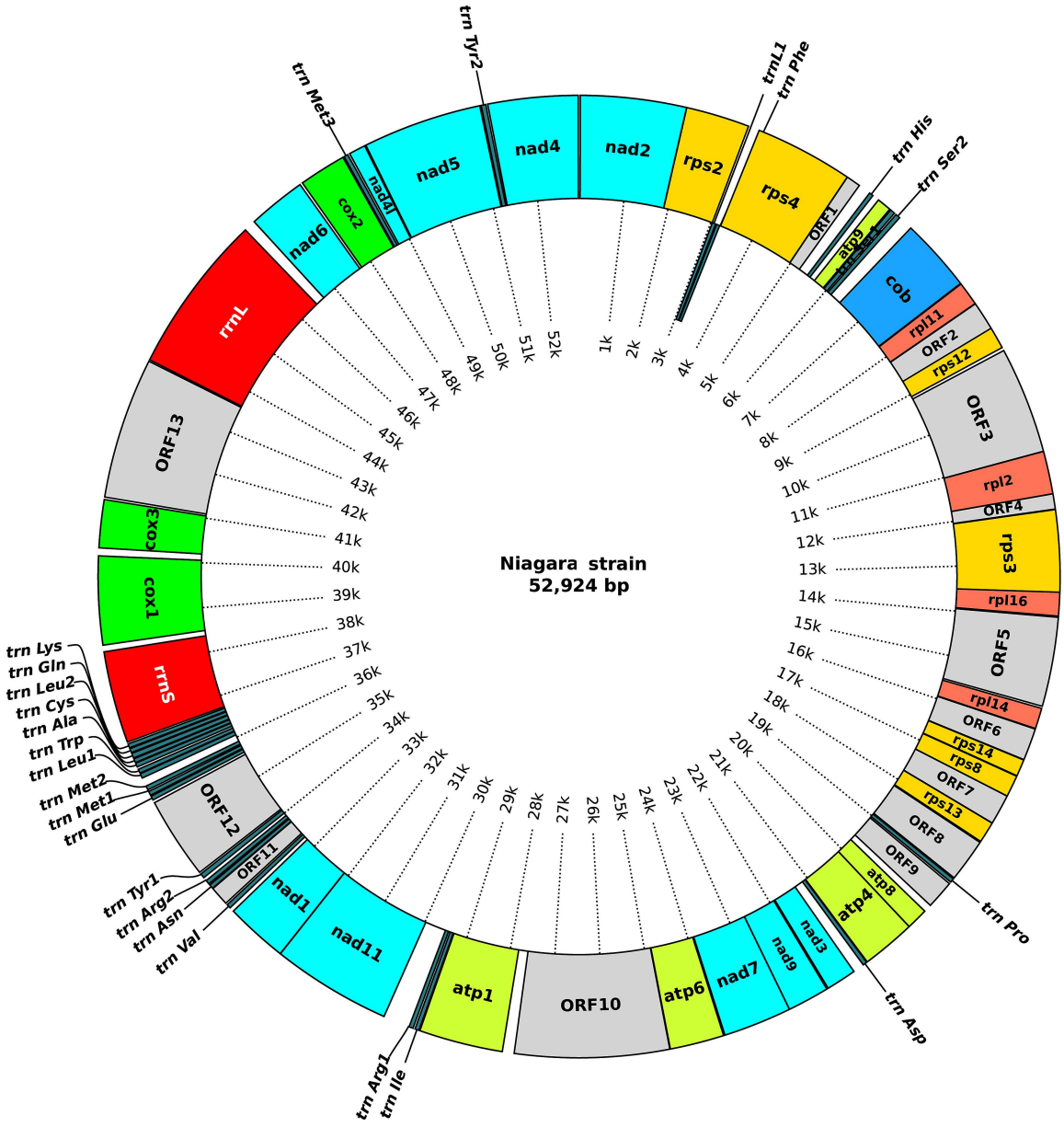

Fig. 2. Mitochondrial genome map of the "Niagara strain." The tRNA genes are labeled based on the IUPACIUB single-letter amino acid codes.

2019b). It shows GC content 27,1% (Table 1), which is a rather low level. The prevalence of thymine over adenine and guanine over cytosine in the majority strand provides negative AT-skew and positive GC-skew. This picture of AT-skew is similar to that observed in most other organisms (Bernt et al., 2013b; Bondarenko et al., 2018b). Therefore, the nucleotide composition of "Niagara strain" MT genome is significantly biased toward A and T bases which unavoidably leads to the predominance of certain codons and amino acids in proteins.

"Niagara strain" MT genome contains set of 30 PCGs (*atp1, 4, 6, 8, 9, cob, cox1-3, nad1-7, 9, 11, nad4L, rpl* and *rps* genes), 25 tRNA, two rRNA genes (*rrnL* and *rrnS*) and 13 open reading frames (ORFs) (Table 2). This MT genome in gene content is more complete than other sequenced MT genomes of vannellids (Bondarenko et al. ,2018a, 2018b, 2018c). The set of PCGs genes differs by the presence of *nad7* and *nad9* genes and includes all *atp* genes. The set of *rpl* and *rps* genes in the MT genome of this strain also differs from that in other known Amoebozoa

Table 1. LNucleotide composition of the mitochondrial genome of Niagara strain.

AT%	GC%	A%	T%	G%	C%	AT-skew	GC-skew
72,9	27,1	33,7	39,2	15,4	11,7	-0,076	0,135

MT genomes (Bondarenko et al. 2018a, 2018b, 2018c; Burger et al., 1995; Greninger et al., 2015; Ogawa et al., 2000; Tanifuji et al., 2017). Similar to the *Clydonella sawyeri* MT genome, fifty-three genes and thirteen ORFs are located on H-strand except for two tRNA genes on L-strand (Bondarenko et al., 2018c). The total length of all PCGs in the Niagara strain MT genome, excluding termination codons, is 28.135 bp which amounts to 53,16% of the total genome length. All genes in the MT genome of this strain contain no introns. All ORFs are unique to this MT genome and have no homologs among Vannellidae species as well as among other Amoebozoa. This situation is quite common, and in many other MT genomes of Amoebozoa we also met unique ORFs (Bondarenko et al., 2019b). As a remarkable character, it contains two rather long ORFs with lengths 2508 and 2739 bp (ORF 13 and 10, respectively). MT genome of Niagara strain has two small gene overlaps and eleven non-coding regions longer than 100 bp (Table 2). The largest overlap is 14 bp and is located between ORF2 and *rps12*. The non-coding regions constitute 3583 bp in total and 6,7% of the total MT genome size (Table 2) which is similar to the *Paravannella minima* MT genome (Bondarenko et al., 2019a). The largest non-coding region is 474 bp long and located between $tRNA^{Arg1}$ and *nad11*.

There are two alternative start codons in the "Niagara strain" MT genome. Most of PCGs and ORFs use ATG as a start codon, and only *nad11* use ATT. There are only two stop codons in the "Niagara strain" MT genome (TAA and TAG); TGA stop codon was not found in this mt genome. Similar to the *Paravannella minima* MT genome, the "Niagara strain" MT genome does not have TAA stop codons within CDS, which leads to reading frameshifts in the MT genomes of otherVannellidae mt genomes (Bondarenko et al., 2018a, 2018b, 2018c). Similar to most sequenced MT genomes of Amoebozoa, the Niagara strain uses the genetic code 4.

The large ribosomal RNA (*rrnL*) gene in the "Niagara strain" MT genome is located between *ORF13* and *nad6* gene, and the small ribosomal RNA (*rrnS*) is situated between $tRNA^{Lys}$ and *cox1* genes (Fig. 2). The length of *rrnL* and *rrnS* is 2838 bp and 1634 bp, respectively. tRNA genes have a total length of 1895 bp, and most of them are located between *ORF12* and *rrnS* genes. All tRNAs have the typical cloverleaf secondary structure. This MT genome contains additional arginine, serine, tyrosine, three leucine, and methionine tRNA genes. Among known Amoebozoa MT genomes, only *Phalansterium* sp. and *Clydonella sawyeri* have three methionine tRNA genes (Bondarenko et al., 2018c; Pombert et al., 2013). As for leucine, no one known amoebozoan MT genome has three tRNA genes. We observed the difference in the nucleotide composition between $tRNA^{Leu1}$ located in the L-strand and the other two duplications of this gene located on the H-strand. These differences in nucleotide composition and location of these duplications suggest about ancient nature of $tRNA^{Leu1}$ duplication. $tRNA^{Arg}$, $tRNA^{Ser}$, and $tRNA^{Met}$ duplications also have ancient nature. In contrast, $tRNA^{Tyr}$ gene duplication occurs for the first time. The copies of these duplicated genes show a small difference in the nucleotide composition indicating the "young" nature of this duplication.

The MT genome of "Niagara strain" shows no evidence of post-translational editing. The same is true for the recently sequenced species *Paravannella minima* (Bondarenko et al., 2019a). However, little editing was found in another lineage – *Clydonella sawyeri* (Bondarenko et al. 2018c) and extensive editing – in the species belonging to the genus *Vannella* – the crown lineage of Vannellida, namely – *Vannella croatica* and *V. simplex* (Bondarenko et al., 2018a, 2018b). Among other amoebozoan lineages, RNA editing is known among plasmodial slime molds - Myxogastria (see Houtz et al., 2018) – the group belonging to Evosea, which is the crown group of the entire Amoebozoa tree. Discosea is located more basally (Kang et al., 2017). At the same time, among Discosea lineage, no editing was found in the MT genomes of the genera *Paramoeba* and *Neoparamoeba* (Tanifuji et al., 2017; Bondarenko et al., 2020). Both belong to Dactylopodida clade, which in the phylogenetic tree of Discosea is more basal than the Vannellida clade. So, the present finding further confirms the suggestion on the possibility of the independent origin of editing in individual amoebozoan lineages (Bondarenko et al., 2019b).

Table 2. Organization of Niagara strain mitochondrial genome.

Gene	Strand	Location	Size (bp)	Anticodon	Start	Stop	Intergenic nucleotides
nad2	+	22-1899	1878		ATG	TAA	36
rps2	+	1903-3033	1131		ATG	TAA	3
tRNA^{Leu1}	-	3036-3114	79	TAA			2
tRNA^{Phe}	-	3139-3211	73	GAA			24
rps4	+	3233-4981	1749		ATG	TAA	21
ORF1	+	4985-5239	255		ATG	TAA	3
tRNA^{His}	+	5463-5535	73	GTG			223
atp9	+	5660-5914	255		ATG	TAA	124
tRNA^{Ser1}	+	5937-6023	87	GCT			22
tRNA^{Ser2}	+	6059-6143	85	TGA			35
cob	+	6409-7713	1305		ATG	TAG	265
rpl11	+	7727-8149	423		ATG	TAA	13
ORF2	+	8156-8677	522		ATG	TAA	6
rps12	+	8664-9053	390		ATG	TAA	-14
ORF3	+	9113-11011	1899		ATG	TAA	59
rpl2	+	11015-11733	759		ATG	TAA	3
ORF4	+	11775-12044	270		ATG	TAA	1
rps3	+	12054-13505	1452		ATG	TAA	9
rpl16	+	13505-13924	418		ATG	TAA	-1
ORF5	+	13946-15529	1584		ATG	TAA	41
rpl14	+	15569-15937	369		ATG	TAA	39
ORF6	+	15948-16517	570		ATG	TAA	10
rps14	+	16526-16825	300		ATG	TAA	8
rps8	+	16836-17219	384		ATG	TAG	10
ORF7	+	17222-17758	537		ATG	TAA	2
rps13	+	17763-18140	378				4
ORF8	+	18162-18929	768		ATG	TAA	21
tRNA^{Pro}	+	18952-19024	73	TGG			22
ORF9	+	19060-19566	507		ATG	TAA	35
atp8	+	19707-20090	384		ATG	TAA	140
atp4	+	20092-21051	960		ATG	TAA	1
tRNA^{Asp}	+	21065-21138	74	GTC			13
nad3	+	21338-21877	540		ATG	TAA	199
nad9	+	21905-22648	744		ATG	TAA	27
nad7	+	22650-23867	1218		ATG	TAA	1

Table 2. Continuation.

atp6	+	23891-24853	963		ATG	TAA	23
ORF10	+	24864-27602	2739		ATG	TAA	10
atp1	+	27813-29303	1491		ATG	TAA	210
tRNA^{Ile}	+	29322-29394	73	CAT			18
tRNA^{Arg1}	+	29432-29504	73	ACG			37
nad11	+	29979-32105	2127		ATT	TAA	474
nad1	+	32109-33218	1110		ATG	TAA	3
tRNA^{Val}	+	33299-33372	74	TAC			80
ORF11	+	33409-33750	342		ATG	TAA	36
tRNA^{Asn}	+	33761-33832	72	GTT			10
tRNA^{Arg2}	+	33860-33933	74	TCT			27
tRNA^{Tyr1}	+	34000-34082	83	GTA			66
ORF12	+	34118-35572	1455		ATG	TAA	35
tRNA^{Glu}	+	35628-35699	72	TTC			55
tRNA^{Met1}	+	35722-35793	72	CAT			22
tRNA^{Met2}	+	35826-35898	73	CAT			32
tRNA^{Leu2}	+	36055-36136	82	TAG			156
tRNA^{Trp}	+	36150-36222	73	CCA			13
tRNA^{Ala}	+	36267-36338	72	TGC			44
tRNA^{Cys}	+	36351-36421	71	GCA			12
tRNA^{Leu3}	+	36438-36522	85	CAA			16
tRNA^{Gln}	+	36540-36612	73	TTG			17
tRNA^{Lys}	+	36630-36701	72	TTT			17
rrnS	+	36736-38369	1634		ATG	TAA	34
cox1	+	38488-40056	1569		ATG	TAA	118
cox3	+	40197-41048	852		ATG	TAG	140
ORF13	+	41093-43600	2508		ATG	TAA	44
rrnL	+	43629-46466	2838		ATG	TAA	28
nad6	+	46677-47711	1035		ATG	TAG	210
cox2	+	47762-48601	840		ATG	TAA	50
tRNA^{Met3}	+	48625-48698	74	CAT			23
nad4l	+	48736-49032	297		ATG	TAA	37
nad5	+	49055-51172	2118		ATG	TAA	22
tRNA^{Tyr2}	+	51198-51280	83	GTA			25
nad4	+	51313-52908	1596		ATG	TAA	32

Acknowledgments

Supported by the RSF project 20-14-00195. This study utilized equipment of the Core Facilities Centers "Culture Collection of Microorganisms," "Centre for Molecular and Cell Technologies," "Biobank," and "Computing Centre SPbU" of the Research Park of Saint-Petersburg State University.

References

Bankevich A., Nurk S., Antipov D., Gurevich A. et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

Bernt M., Donath A., Jühling F., Externbrink F. et al. 2013a. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69, 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Bernt M., Bleidorn C., Braband A., Dambach J. et al. 2013b. A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. Mol. Phylogenet. Evol. 69, 352–364. https://doi.org/10.1016/j.ympev.2013.05.002

Bondarenko N.I., Nassonova E.S., Mijanovic O., Glotova A.A. et al. 2018a. Mitochondrial genome of *Vannella croatica* (Amoebozoa, Discosea, Vannellida). J. Eukaryot. Microbiol. 65, 820-827. https://doi.org/doi:10.1111/jeu.12523

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. https://doi.org/10.1016/j.ejop.2018.01.006

Bondarenko N., Glotova A., Kamyshatskaya O., Mesentsev Y. et al. 2018c. The complete mitochondrial genome of *Clydonella sawyeri* (Amoebozoa, Discosea, Vannellida). Protistology. 12, 47–54. https://doi.org/10.21685/1680-0826-2018-12-1-4

Bondarenko N., Glotova A., Nassonova E., Masharsky A., Polev D. and Smirnov A. 2019a. The complete mitochondrial genome of *Paravannella minima* (Amoebozoa, Discosea, Vannellida). Europ. J. Protistol. 68, 80–87. https://doi.org/10.1016/j.ejop.2019.01.005

Bondarenko N., Smirnov A, Nassonova E., Glotova A. and Fiore-Donno Anna-Maria. 2019b. Mitochondrial genomes of Amoebozoa. Protistology. 13, 179–191. https://doi.org/10.21685/1680-0826-2019-13-4-1

Bondarenko N., Volkova E., Masharsky A., Kudryavtsev A. and Smirnov A. 2020. A comparative characterization of the mitochondrial genomes of *Paramoeba aparasomata* and *Neoparamoeba pemaquidensis* (Amoebozoa, Paramoebidae). J. Eukaryot. Microbiol. 67, 167–175. https://doi.org/10.1111/jeu.12767

Burger G., Plante I., Lonergan K.M. and Gray M.W. 1995. The mitochondrial DNA of the amoeboid protozoon, *Acanthamoeba castellanii*: complete sequence, gene content and genome organization. J. Mol. Biol. 245, 522–537. https://doi.org/10.1006/jmbi.1994.0043

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. https://doi.org/10.1007/978-3-319-78190-7_8

Greninger A.L., Messacar K., Dunnebacke T., Naccache S.N. et al. 2015. Clinical metagenomic identification of *Balamuthia mandrillaris* encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing. Genome Med. 7, 113. https://doi.org/10.1186/s13073-015-0235-2

Kang S., Tice A.K., Spiegel F.W., Silberman J.D. et al. 2017. Between a pod and a hard test: the deep evolution of amoebae. Mol. Biol. Evol. 34, 2258–2270. https://doi.org/10.1093/molbev/msx162

Kudryavtsev A. and Volkova E. 2018. *Clydonella sawyeri* n. sp. (Amoebozoa, Vannellida): Morphological and molecular study and a re-definition of the genus *Clydonella* Sawyer, 1975. Europ. J. Protistol. 63, 62–71. https://doi.org/10.1016/j.ejop.2018.01.008

Lowe T.M. and Eddy S.R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964. https://doi.org/10.1093/nar/25.5.955

Ogawa S., Yoshino R., Angata K., Iwamoto M. et al. 2000. The mitochondrial DNA of *Dictyostelium discoideum*: complete sequence, gene content, and genome organization. Mol. Gen. Genet. 263, 514–519. https://doi.org/10.1007/PL00008685

Perna N.T. and Kocher T.D. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41,

353–358. https://doi.org/10.1007/BF00186547

Pombert J.-F., Smirnov A., James E.R., Janouskovec J., Gray M.W. and Keeling P.J. 2013. The complete mitochondrial genome from an unidentified *Phalansterium* species. Protist Genomics. 1, 25–32. https://doi.org/10.2478/prge-2013-0002

Rutherford K., Parkhill J., Crook J., Horsnell T. et al. 2000. Artemis: sequence visualization and annotation. Bioinformatics. 16, 944–945. https://doi.org/10.1093/bioinformatics/16.10.944

Smirnov A.V., Nassonova E.S., Chao E. and Cavalier-Smith T. 2007. Phylogeny, evolution, and taxonomy of vannellid amoebae. Protist. 158, 295–324. https://doi.org/10.1016/j.protis.2007.04.004

Tanifuji G., Cenci U., Moog D., Dean S. et al. 2017. Genome sequencing reveals metabolic and cellular interdependence in an amoeba-kinetoplastid symbiosis. Sci. Rep. 7, 11688. https://doi.org/10.1038/s41598-017-11866-x

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