Protistology 12 (1), 47–54 (2018)

Protistology

The complete mitochondrial genome of *Clydonella sawyeri* (Amoebozoa, Discosea, Vannellida)

Natalya Bondarenko[1], Anna Glotova[1], Oksana Kamyshatskaya[2], Yelisei Mesentsev[1], Kirill Lotonin[1], Alexey Masharsky[3], Dmitry Polev[4], Elena Nassonova[1,5] and Alexey Smirnov[1]

[1] *Department of Invertebrate Zoology, Faculty of Biology, St. Petersburg State University, 199034 St. Petersburg, Russia*
[2] *Core Facilities Center "Culture Collection of Microorganisms", St. Petersburg State University, Stary Peterhof, 198504 St. Petersburg, Russia*
[3] *Core Facility Centre for Molecular and Cell Technologies, St. Petersburg State University, Stary Peterhof, 198504 St. Petersburg, Russia*
[4] *Core Facility Centre Biobank, St. Petersburg State University, Stary Peterhof, 198504 St. Petersburg, Russia*
[5] *Laboratory of Cytology of Unicellular Organisms, Institute of Cytology RAS, 194064 St. Petersburg, Russia*

| Submitted February 21, 2018 | Accepted April 10, 2018 |

Summary

We describe the complete sequence and organization of the mitochondrial genome from a brackish-water amoeba *Clydonella sawyeri* (Amoebozoa, Discosea, Vannellida). The circular mitochondrial DNA of this species has 31,131 bp in length and contains 17 protein-coding genes, 2 ribosomal RNAs, 21 transfer RNAs and 13 open reading frames. Length and gene content distinguish mitochondrial genome of *Clydonella sawyeri* from the mitochondrial genomes of other Amoebozoa species.

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

Introduction

Mitochondrial genomes (mt genomes) are an important instrument for the phylogenetic studies because of their accessibility and higher evolutionary rate compared to the nuclear DNA (Castro et al., 1998; Lang et al., 1999; Gray et al. 1999). There is much data on the mt genomes in many groups of organisms (Tan et al., 2017), while among Amoebozoa mitochondrial genomes remain relatively poorly studied. By now sequences of mt genomes for 13 amoebozoan species are available (see Bondarenko et al., 2018a, Table S1; 2018b). This dataset is dominated with the species that can be grown in axenic culture or in pure mass culture, like *Acanthamoeba, Balamuthia*, and *Dictyostelium,*

doi: 10.21685/1680-0826-2018-12-1-4
© 2018 The Author(s)
Protistology © 2018 Protozoological Society Affiliated with RAS

and it does not yet cover even all the major branches of Amoebozoa.

The evolutionary change of the gene order in the mt genomes of Amoebozoa appears to be rather rapid and the level of synteny between genomes may be rather low in different amoebozoan lineages (Heidel and Glöckner, 2008) and even among the closely related species belonging to the same phylogenetic lineage (Bondarenko et al., 2018b). The reasons for this are not yet clear. To understand this problem better we subsequently sequence mitochondrial genomes of the amoebae of the order Vannellida (Amoebozoa, Discosea), aiming to cover all major phylogenetic branches in this large and diverse group of naked lobose amoebae. The present paper reports data on the mitochondrial genome of the recently described species *Clydonella sawyeri*, the first *Clydonella* species, fully characterized both by the morphological and the molecular methods (Kudryavtsev and Volkova, 2018).

Material and methods

The type culture of *Clydonella sawyeri* isolated from the upper part of the littoral zone of Seldyanaya Bay, Chupa Inlet (Kandalaksha Bay, The White Sea, North-Western Russia) was used for this study (Kudryavtsev and Volkova, 2018). 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 manufacturer's instructions. Approximately 1.7 million reads with length 25-595 bp were obtained using Ion Torrent PGM system (Life Technologies). 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 MITOS web server (Bernt et al., 2013a). Artemis was used for visualization of 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. The *C. sawyeri* mitochondrial genome has been deposited in GenBank under the accession number MH094141.

Results and discussion

Mitochondrial genome of *C. sawyeri* is a double-stranded circular DNA molecule with the length of 31,131 bp (Fig. 1). Thus, it further completes the list of the relatively small (less than 40 kbp) mt genomes among Amoebozoa. Two other small mt genomes also belong to amoebae of the family Vannellidae (Bondarenko et al., 2018a, 2018b). Mitochondrial genome of *C. sawyeri* has GC content 24,5% (Table 1), which is a relatively 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). The nucleotide composition of *C. sawyeri* mitochondrial genome is significantly biased toward A and T bases which leads to the predominance of certain codons and amino acids in proteins (Table 2).

Clydonella sawyeri mt genome contains set of 17 PCGs (*atp1, 9, cob, cox1-3, nad1-6, nad4L, rpl* and *rps* genes), 21 tRNA, two rRNA genes (*rrnL* and *rrnS*) and 13 open reading frames (ORFs) (Table 3). The set of *rpl* and *rps* genes in *C. sawyeri* mt genome differs from that in *Vannella simplex* and *Vannella croatica* mt genomes (Bondarenko et al., 2018a, 2018b). The set of PCGs genes differs by absence of *nad11* and presence of *atp1* genes. Thirty-eight genes and thirteen ORFs are located on H-strand except for two tRNA genes on L-strand. The total length of all PCGs in *C. sawyeri* mt genome, excluding termination codons, is 15.065 bp. All genes in *C. sawyeri* mt genome contain no introns.

Table 1. Nucleotide composition characteristics of the mitochondrial genome of *Clydonella sawyeri*.

AT%	GC%	A%	T%	G%	C%	AT-skew	GC-skew
75,5	24,5	31,9	43,6	14,6	9,9	-0,155	0,194

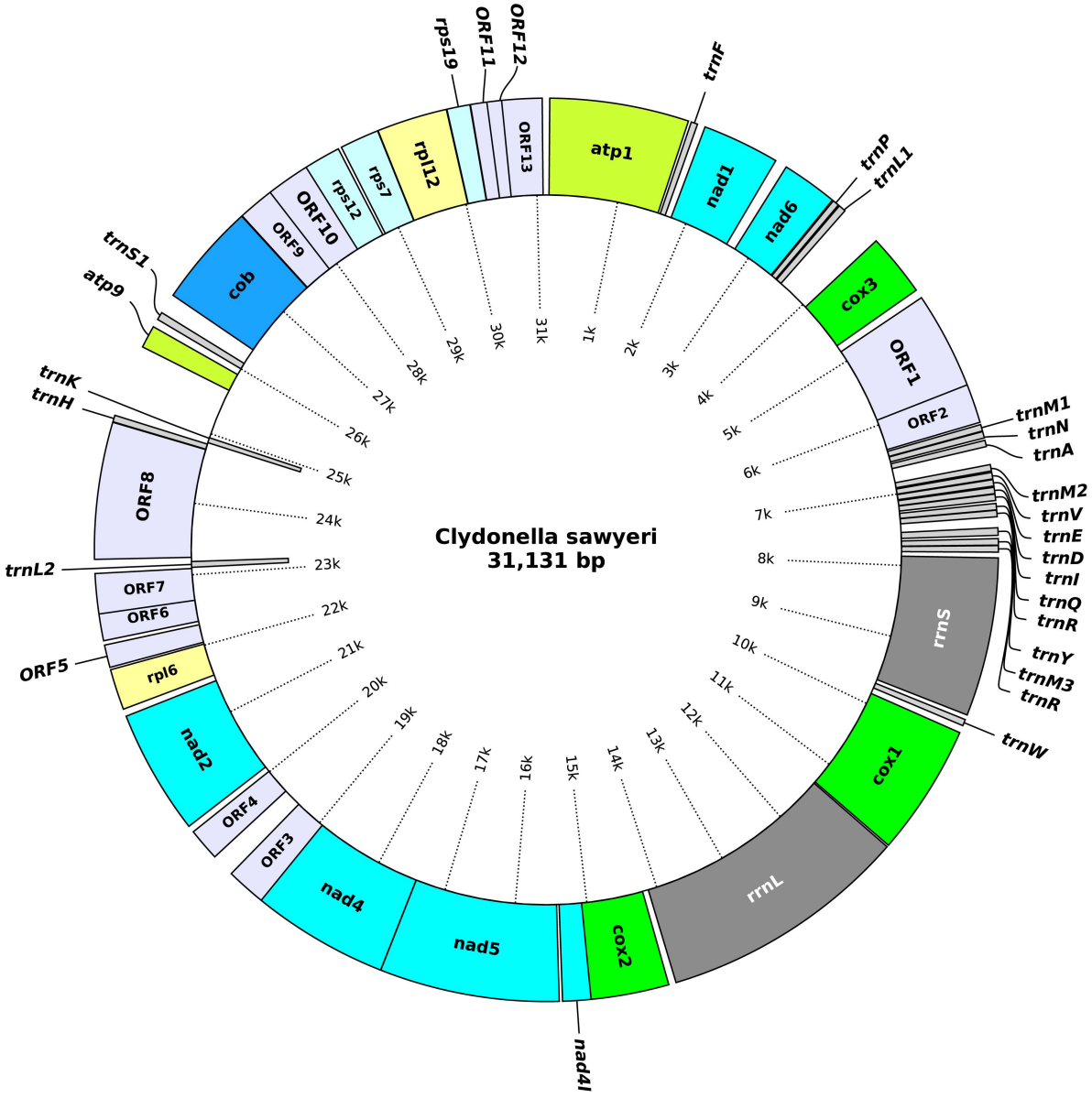

Fig. 1. *Clydonella sawyeri* mitochondrial genome map. The tRNA genes are labeled based on the IUPAC-IUB single letter amino acid codes.

Mitochondrial genome of *C. sawyeri* has five small gene overlaps and five non-coding regions longer than 100 bp. The largest overlap is 608 bp and located between *ORF11* and *ORF12*. The non-coding regions constitute 3225 bp in total and 10,35% of the total mt genome size (Table 3). The largest non-coding region is 715 bp long and located between $tRNA^{Lys}$ and *atp9*.

The large ribosomal RNA (*rrnL*) gene in *C. sawyeri* is located between cox1 and cox2 genes and small ribosomal RNA (*rrnS*) between $tRNA^{Arg}$ and $tRNA^{Trp}$ genes (Fig. 1). The length of *rrnL* and *rrnS* is 2789 bp and 1767 bp, respectively. tRNA genes have a total length of 1570 bp and most of them are located between *ORF2* and *rrnS* genes. All tRNAs have the typical cloverleaf secondary structure. tRNA genes are better represented in mt genome of *C. sawyeri* as compared to other mt genomes of vannellids (Bondarenko et al., 2018a, 2018b). Its genome contains additional arginine, leucine, serine and three methionine tRNA genes (Fig. 2).

The majority of mitochondrial genomes in-

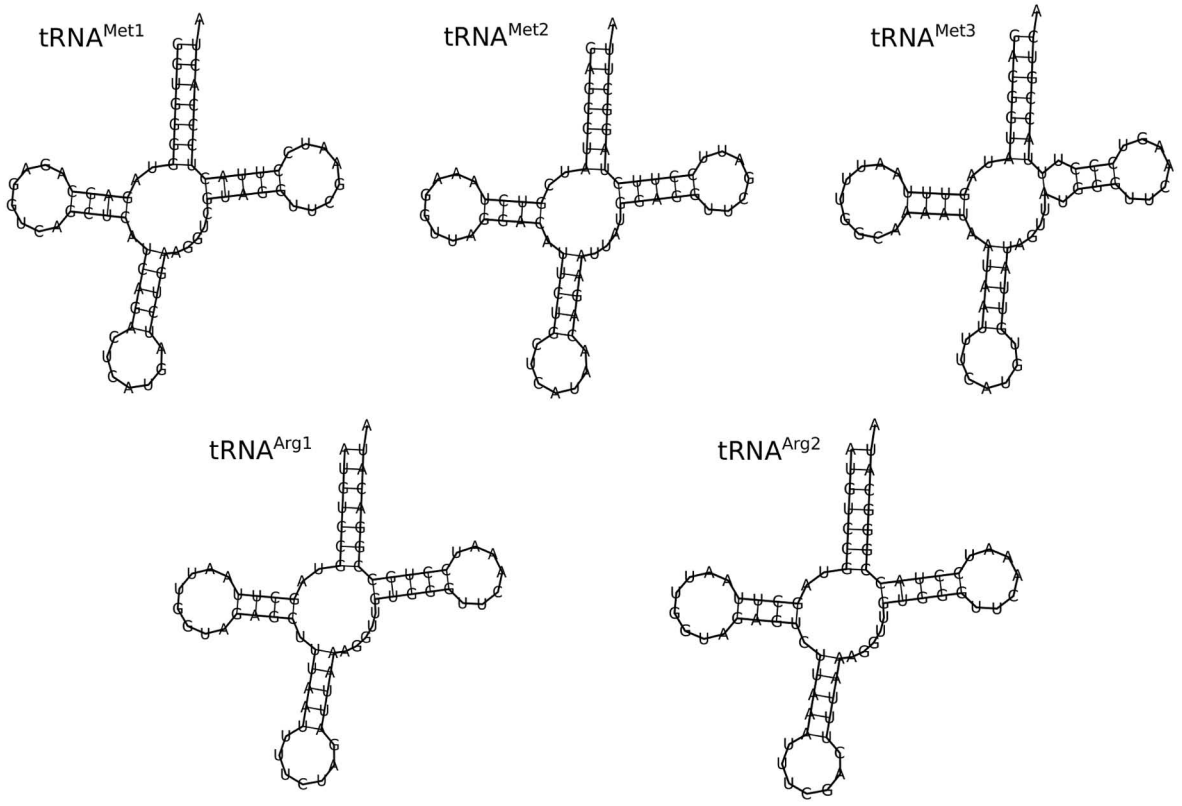

Fig. 2. Predicted secondary structure of the *tRNA*^{Met} and *tRNA*^{Arg} genes of *Clydonella sawyeri*.

Table 2. Amino acid composition of genes and ORFs located in the *Clydonella sawyeri* mitochondrial genome.

	nad1	nad2	nad4l	nad4	nad5	nad6	cob	atp1	atp9	cox1	cox2	cox3	rps12	rps7	rps19	rpl12	rpl6
Ala	20	18	8	19	30	8	21	38	11	33	11	19	8	5	1	9	1
Arg	5	6	3	9	11	4	10	24	1	8	6	7	12	6	2	19	6
Asn	11	17	7	27	23	16	15	30	2	16	22	5	7	17	9	15	1
Asp	7	13	3	9	19	4	9	28	1	16	16	4	2	5	0	4	5
Cys	3	5	0	9	13	0	3	2	1	2	4	4	2	3	2	5	1
Gln	4	4	1	8	7	3	7	21	1	6	5	5	6	3	2	2	1
Glu	11	11	2	9	12	4	7	25	3	9	16	6	1	6	3	6	4
Gly	19	19	5	27	43	8	27	44	10	49	11	18	9	5	4	24	4
His	2	6	0	6	13	3	11	4	0	15	7	10	1	1	0	5	1
Ile	31	48	23	55	64	35	38	39	7	50	33	12	7	15	6	21	8
Leu	44	68	18	81	97	31	52	58	13	55	32	28	7	15	8	27	12
Lys	8	23	2	13	27	8	7	36	3	8	14	3	28	27	19	32	43
Met	10	5	4	17	20	6	8	6	3	22	8	4	1	3	2	3	1
Phe	35	89	11	82	117	30	40	24	10	51	27	29	5	7	12	20	34
Pro	13	13	0	15	14	2	24	16	1	21	12	8	6	5	2	8	0
Ser	25	48	6	39	56	19	21	44	7	30	18	18	7	6	7	30	6
Thr	8	13	1	24	23	4	16	15	4	27	11	12	6	2	2	8	3
Trp	5	3	1	9	7	1	11	1	0	12	6	8	0	2	0	2	0
Tyr	10	31	2	28	28	4	22	16	1	26	15	13	1	2	2	6	6
Val	18	25	7	25	39	17	34	44	4	29	21	15	13	5	4	15	3

Table 3. Organization of *Clydonella sawyeri* mitochondrial genome.

Gene	Strand	Location	Size (bp)	Anticodon	Start	Stop	Intergenic nucleotides
atp1	+	37-1584	1548		ATG	TAA	81
tRNA^{Phe}	+	1617-1689	73	gaa			32
nad1	+	1777-2637	861		ATG	TAG	87
nad6	+	2771-3394	624		ATG	TAA	133
tRNA^{Pro}	+	3408-3479	72	tgg			13
tRNA^{Leu1}	+	3494-3575	82	tag			14
cox3	+	4050-4748	699		ATG	TAA	474
ORF1	+	4848-5936	1089		ATG	TAG	99
ORF2	+	5930-6364	435		ATG	TAA	-5
tRNA^{Met1}	+	6390-6463	74	cat			25
tRNA^{Asn}	+	6470-6542	73	gtt			6
tRNA^{Ala}	+	6569-6639	71	tgc			26
tRNA^{Met2}	+	6841-6912	72	cat			1
tRNA^{Val}	+	6929-7001	73	tac			16
tRNA^{Glu}	+	7009-7080	72	ttc			7
tRNA^{Asp}	+	7094-7164	71	gtc			13
tRNA^{Ile}	+	7169-7240	72	gat			4
tRNA^{Gln}	+	7272-7343	72	ttg			31
tRNA^{Arg1}	+	7347-7419	73	tct			3
tRNA^{Tyr}	+	7540-7623	84	gta			20
tRNA^{Met3}	+	7659-7731	73	cat			35
tRNA^{Arg2}	+	7736-7808	73	tcg			4
rrnS	+	7871-9637	1767				62
tRNA^{Trp}	+	9692-9763	72	cca			54
cox1	+	9847-11314	1468		ATG	TAA	83
rrnL	+	11336-14124	2789				21
cox2	+	14207-15084	878		ATG	TAA	82
nad4l	+	15074-15388	315		ATG	TAA	-9
nad5	+	15424-17424	2001		ATG	TAG	35
nad4	+	17426-18961	1536		ATG	TAA	1
ORF3	+	18948-19388	441		ATG	TAA	-12
ORF4	+	19646-20008	363		ATG	TAA	257
nad2	+	20085-21479	1395		ATG	TAA	76
rpl6	+	21562-22017	456		ATG	TAA	82
ORF5	+	22044-22295	252		ATG	TAA	26
ORF6	+	22344-22715	372		ATA	TAA	48
ORF7	+	22642-23088	447		ATA	TAA	-72
tRNA^{Leu2}	-	23104-23185	82	taa			15
ORF8	+	23244-24767	1524		ATG	TAA	58
tRNA^{His}	+	24773-24847	75	gtg			5

Table 3. Continuation.

Gene	Strand	Location	Size (bp)	Anticodon	Start	Stop	Intergenic nucleotides
tRNA^{Lys}	-	24869-24941	73	ttt			21
atp9	+	25657-25908	252		ATG	TAA	715
tRNA^{Ser}	+	26001-26088	88	gct			92
cob	+	26315-27466	1152		ATG	TAA	226
ORF9	+	27473-27889	417		ATG	TAA	6
ORF10	+	27873-28379	507		ATG	TAA	-15
rps12	+	28354-28761	408		ATG	TAA	-24
rps7	+	28799-29235	437		ATG	TAA	37
rpl12	+	29240-30025	786		ATG	TAA	4
rps19	+	30031-30330	300		ATG	TAA	5
ORF11	+	30294-31085	792		ATG	TAA	-5
ORF12	+	30476-30730	255		ATG	TAA	-608
ORF13	+	30639-31085	447		ATA	TAA	90

clude two tRNA genes for serine and leucine and only one tRNA gene for each of the other 18 amino acids (Attardi, 1985; Cantatore and Saccone, 1987; Ogawa et al., 2000). Among known Amoebozoa mt genomes *Acanthamoeba castellanii, Balamuthia mandrillaris, Hartmannella vermiformis, Neoparamoeba pemaquidensis* (deposited as *Paramoeba*) and *Vannella croatica* have only two methionine tRNA genes (Burger et al., 1995; Bondarenko et al., 2018a; Greninger et al., 2015) and only *Phalansterium* sp. has three methionine tRNA genes (Pombert et al., 2013). The ancient nature of *tRNA^{Met}* duplication in *C. sawyeri* mt genome is evidenced by the large difference in the nucleotide composition of these genes. In contrast, *tRNA^{Arg}* gene duplication occurs for the first time and has the small difference in the nucleotide composition indicating the "young" nature of this duplication. The functional significance of these two duplications in *C. sawyeri* is not clear yet.

All PCGs and ten ORFs in *C. sawyeri* use ATG as a start codon, three ORFs use ATA as an alternative start codon. There are two stop codons in *C. sawyeri* mt genome (TAA and TAG). TGA stop codon wasn't found in this mt genome. Several genes have numerous TAA stop codons within CDS. The same situation was observed in *Vannella croatica* and *Vannella simplex* mt genomes (Bondarenko et al., 2018a, 2018b). In contrast to the above mentioned mt genomes, where most of the genes and ORFs have stop codons, thus presuming that those are editing sites (or reading frameshifts) (Bondarenko et al., 2018a, 2018b), in *C. sawyeri* mt genome the stop codons were found within five genes only, namely *cox1, cox2, nad2, nad5* and *rps7*. The mechanism of resolving these stop codons is not clear. Like in the other vannellids, it may be either RNA editing or reading frame shifts. We translated these five genes using insertional editing. Cytosine insertion was used, as it takes place in Cox I gene of some other vannellid amoebae (Nassonova et al., 2010) and using frameshifts. Obtained amino acid sequences were aligned manually for every gene with those from all other available amoebozoan mitochondrial genomes. This analysis showed that translation with cytosine insertions leads to non-synonymous replacements in conserved regions of proteins and amino acid sequences, resulting in considerable divergence with the sequences of other amoebozoans. Translation with using reading frame shifts showed better results; the number of non-synonymous replacements is lower, and sequences could be better aligned. Certainly, the insertional editing is not limited to cytosine insertions used here, several different kinds of editing are possible within the same mitogenome (Byrne and Gott, 2004; Gott et al., 2010), hence this question requires further study.

Acknowledgments

Supported by the RSF project 17-14-01391 (culturing, treatment of cells and molecular work).

Bioinformatics supported by RFBR project 16-34-60111 to Natalya Bondarenko. This study utilized equipment of the Core Facilities Centers "Culture Collection of Microorganisms", "Centre for Molecular and Cell Technologies" and "Computing Centre SPbU" of Saint-Petersburg State University. Infrastructure supported with SPSU equipment grants 1.40.539.2017 and 1.40.509.2017. The authors thank A.A. Kudryavtsev for providing *C. sawyeri* culture prior to depositing it as a type strain to the collection of the Core Facilities Center "Culture Collection of Microorganisms".

References

Attardi G. 1985. Animal mitochondrial DNA, an extreme example of genetic economy. Int Rev Cytol. 93, 93–145.

Bankevich A., Nurk S., Antipov D., Gurevich A., Dvorkin M., Kulikov A. S., Lesin V., Nikolenko S., Pham S., Prjibelski A., Pyshkin A., Sirotkin A., Vyahhi N., Tesler G., Alekseyev M.A. and Pevzner P.A. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19, 455–477.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M. and Stadler P.F. 2013a. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69, 313–319.

Bernt M., Bleidorn C., Braband A., Dambach J., Donath A., Fritzsch G., Golombek A., Hadrys H., Jühling F., Meusemann K., Middendorf M., Misof B., Perseke M., Podsiadlowski L., von Reumont B., Schierwater B., Schlegel M., Schrödl M., Simon S., Stadler P.F, Stöger I. and Struck T.H. 2013b. A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. Mol. Phylogenet. Evol. 69, 352–364.

Bondarenko N.I., Nassonova E.S., Mijanovic O., Glotova A.A., Kamyshatskaya O.G., Kudryavtsev A.A., Masharsky A.E., Polev D.E. and Smirnov A.V. 2018a. Mitochondrial genome of *Vannella croatica* (Amoebozoa, Discosea, Vannellida). J. Eukaryot. Microbiol. accepted, doi will be added in proofs.

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.

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.

Byrne E. M. and Gott J. M. 2004. Unexpectedly complex editing patterns at dinucleotide insertion sites in *Physarum mitochondria*. Mol. Cell Biol. 24, 7821–7828.

Cantatore P. and Saccone C. 1987. Organization, structure, and evolution of mammalian mitochondrion genes. Int. Rev. Cytol. 108, 149–208.

Greninger A.L., Messacar 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. 7, 113.

Gott J.M., Somerlot B.H. and Gray M.W. 2010. Two forms of RNA editing are required for tRNA maturation in *Physarum mitochondria*. RNA 16, 482–488.

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

Kudryavtsev A. and Volkova E. 2018. *Clydonella sawyeri* n. sp. (Amoebozoa, Vannellida): Morphological and molecular study and a redefinition of the genus *Clydonella* Sawyer, 1975. Europ. J. Protistol. 63, 62–71.

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.

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.

Ogawa S., Yoshino R., Angata K., Iwamoto M., Pi 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.

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.

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.

Tan M.H., Gan H.M., Lee Y.P., Poore G., Austin C.M. 2017. Digging deeper: new gene order rearrangements and distinct patterns of codons usage in mitochondrial genomes among shrimps from the Axiidea, Gebiidea and Caridea (Crustacea: Decapoda). PeerJ 5:e2982 https://doi.org/10.7717/peerj.2982.

Address for correspondence: Natalya Bondarenko. Department of Invertebrate Zoology, Faculty of Biology, St. Petersburg State University, Universitetskaya nab. 7/9, 199034 St. Petersburg, Russia;
e-mail: *natalya.lannik@gmail.com*