

RESEARCH

Open Access



Typical structure of rRNA coding genes in diplomemids points to two independent origins of the bizarre rDNA structures of euglenozoans

Paweł Hałakuc*, Anna Karnkowska and Rafał Milanowski*

Abstract

Background: Members of Euglenozoa (Discoba) are known for unorthodox rDNA organization. In Euglenida rDNA is located on extrachromosomal circular DNA. In Kinetoplastea and Euglenida the core of the large ribosomal subunit, typically formed by the 28S rRNA, consists of several smaller rRNAs. They are the result of the presence of additional internal transcribed spacers (ITSs) in the rDNA. Diplonemea is the third of the main groups of Euglenozoa and its members are known to be among the most abundant and diverse protists in the oceans. Despite that, the rRNA of only one diplomemid species, *Diplonema papillatum*, has been examined so far and found to exhibit continuous 28S rRNA. Currently, the rDNA organization has not been researched for any diplomemid. Herein we investigate the structure of rRNA genes in classical (Diplonemidae) and deep-sea diplomemids (Eupelagonemidae), representing the majority of known diplomemid diversity. The results fill the gap in knowledge about diplomemid rDNA and allow better understanding of the evolution of the fragmented structure of the rDNA in Euglenozoa.

Results: We used available genomic (culture and single-cell) sequencing data to assemble complete or almost complete rRNA operons for three classical and six deep-sea diplomemids. The rDNA sequences acquired for several euglenids and kinetoplastids were used to provide the background for the analysis. In all nine diplomemids, 28S rRNA seems to be contiguous, with no additional ITSs detected. Similarly, no additional ITSs were detected in basal prokinetoplastids. However, we identified five additional ITSs in the 28S rRNA of all analysed metakinetoplastids, and up to twelve in euglenids. Only three of these share positions, and they cannot be traced back to their common ancestor.

Conclusions: Presented results indicate that independent origin of additional ITSs in euglenids and kinetoplastids seems to be the most likely. The reason for such unmatched fragmentation remains unknown, but for some reason euglenozoan ribosomes appear to be prone to 28S rRNA fragmentation.

Keywords: rRNA, rDNA, rRNA operon, Euglenozoa, Diplonemids, Euglenids, Kinetoplastids, Internal transcribed spacer

Background

Until a few years ago, Diplonemea (Euglenozoa, Discoba) was a rather neglected group. Only two diplomemid genera, *Diplonema* and *Rhynchopus*, have been described and cultured. This remained in sharp contrast to other well-studied euglenozoans: Kinetoplastea, sister to Diplonemea, and Euglenida [1]. Lately, metabarcoding

*Correspondence: pawelhalakuc@uw.edu.pl; r.milanowski@uw.edu.pl

Institute of Evolutionary Biology, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

surveys from the deep pelagic zone and deep-sea sediments have shown their unrivaled diversity [2–4]. Based on these metabarcoding data, two clades of deep-sea pelagic diplonemids (DSPD I and II) have been described, with the former grouping 97% of all known diplonemid diversity [1]. It also encompasses ten diplonemid single-cells for which genomes were acquired [5]. The majority of the metabarcoding sequences corresponded with a single cell known as Cell 37, and were later described as a new species *Eupelagonema oceanica* [6].

The genomes originating from single cells were incomplete and fragmented, primarily due to high repetitiveness caused by an unexpectedly high density of ‘noncanonical’ introns, similar to euglenid nonconventional introns [5, 7]. The second reason is the size of the genomes – acquired assemblies were up to 300 Mbp large, consistent with the previously reported expected genome size of diplonemids [8]. However, even for such incomplete assemblies, regions present in many copies—such as mitochondrial DNA or nuclear ribosomal RNA (rRNA) operon—can be extracted [9, 10].

Typically, eukaryotic ribosomes contain four rRNAs. Three of them: 18S (also known as SSU, small subunit), 5.8S and 28S rRNA (together also known as LSU, large subunit) are encoded in a single operon (rRNA or rDNA) and co-transcribed. Genes are separated by internal transcribed spacers (ITSs), which are removed during post-transcriptional processing to form mature rRNAs [11]. Such a structure of four continuous rRNAs has been confirmed in the single investigated diplonemid *Diplonema papillatum* [12]. However, this result is in opposition to two other euglenozoan groups: kinetoplastids and euglenids. In both of these groups, 28S rRNA is fragmented into several smaller molecules: 6 in kinetoplastids [13–15], and 13 in euglenids [16–18]. These smaller rRNAs together perform structural and catalytic functions of typical 28S rRNA. The fragmentation is caused by additional ITSs in the rRNA operon of both euglenids and kinetoplastids. While 28S rRNA fragmentation occasionally occurs in various eukaryotes [19–21], the extent of the fragmentation in Euglenozoa is unparalleled. The lack of studied rRNA operons in diplonemids puts the parsimonious (i.e., involving single ancestral acquisition) evolutionary path of euglenozoan rRNA operon into question, which we try to answer herein.

Results

We successfully assembled rRNA operons of all three classical and six out of ten deep-sea diplonemids, including the most abundant *Eup. oceanica* (Additional file 2: Table S1). Lengths of all acquired operons and their subunits are typical for eukaryotes. Furthermore, we acquired and annotated sequences of rRNA operons for three

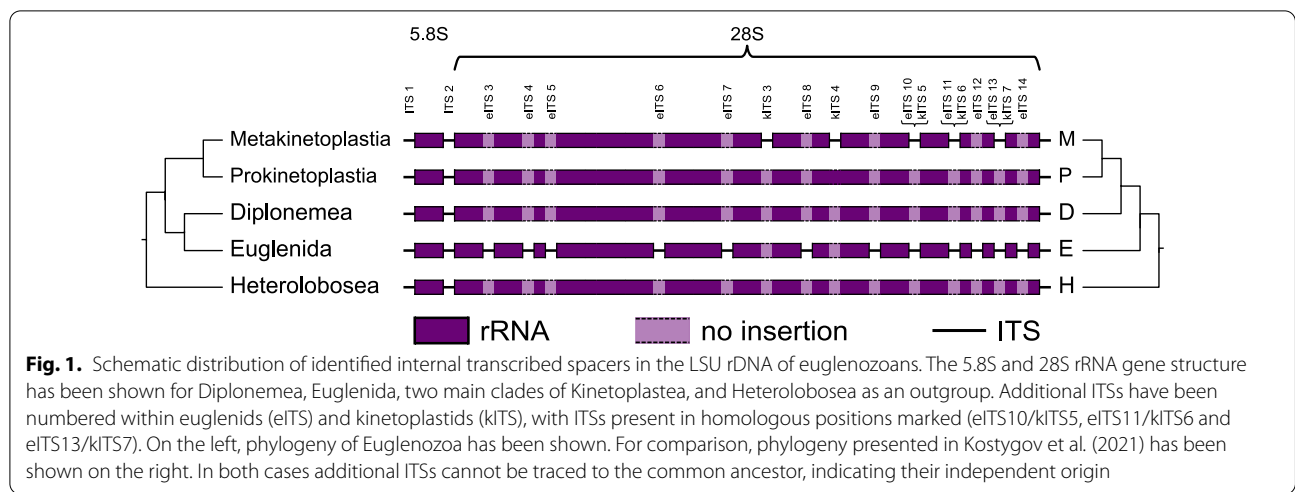
euglenids—one heterotrophic species and two phototrophs—and for nine kinetoplastids—six metakinetoplastids and three prokinetoplastids. In several cases a complete intergenic region (IGR) has not been recovered, hence only the 18S-5.8S-28S rRNA coding region has been analysed further.

Since it is not possible to automatically predict the very complex rRNA secondary structure, another approach has been utilised. We used previously described rRNA structures of *Euglena gracilis* [17], *Trypanosoma cruzi* [22] and *Leishmania major* [23] to identify structural elements, i.e., helices and loops composing the bulk structure of the ribosome. Subsequently, we modelled these structural elements for all other species (see “Methods” section) and marked them upon the alignment. The expected structure of the mature rRNA, which is the most conserved known biological feature, has been used to identify expansion elements. The fragments which would disrupt the ribosome structure are most likely removed during the maturation of rRNA. All alignments and annotations are available in the RepOD repository accompanying this paper (<https://doi.org/10.18150/J4Q2ES>).

We identified all conserved features of the 28S rRNA in all analysed diplonemids and no significant insertions or deletions were found (Additional file 2: Table S1, Fig. 1). For that reason, we conclude that no additional ITS is present in diplonemid rRNA operons, resulting in the typical eukaryotic continuous 28S rRNA.

On the other hand, the rRNA operons of all three euglenids are significantly elongated (10–13 kbp), mainly due to the large expansions within 18S and especially 28S rRNA genes (Additional file 2: Table S1). As previously suggested, almost all of the expansions occur in divergent regions of the rRNA, also known as expansion segments (ES) [24, 25]. We identified the ITSs described for *E. gracilis* [26] in both analysed euglenids. Moreover, two potentially novel additional ITS sites were found in *Rhabdomonas costata*. All expansions within 18S rRNA are shared between euglenids but it has been shown in *E. gracilis* that they are not removed from the mature 18S rRNA [16]; hence we do not indicate these as additional ITSs.

In kinetoplastids, two types of the rRNA operons can be distinguished: an elongated one in metakinetoplastids (trypanosomatids and bodonids), and a standard eukaryotic one in all prokinetoplastids (*Perkinsella* sp. and similar). Long rRNA operons in metakinetoplastids originated from the elongations of the 28S rRNA gene, which are present in the same positions and possess the same features and spatial distribution pattern as previously described additional ITSs [13–15]. All analysed metakinetoplastid sequences have exactly the same pattern of



additional ITSs as trypanosomatids (Fig. 1). In *Bodo saltans*, a basal metakinetoplastid, expansion in the kinetoplastid ITS3 (kITS3) site is short but still much longer than in prokinetoplastids and other analysed species. It suggests that *B. saltans* rRNA does contain the kinetoplastid ITS3. Furthermore, only three kinetoplastid ITSs share positions with euglenid ITSs: kITS5 and eITS10, kITS6 and eITS11, kITS7 and eITS13 (Fig. 1, Additional file 1: Figure S1).

No expansions were recognised as group I introns by RNAmmer [27]. No homology has been observed within and between additional ITSs of euglenids and kinetoplastids. The sequences of ITSs do not have distinct or conserved secondary structures and we did not recover any open reading frames (ORFs) longer than 20 amino acids. No significant blast hits (e-values < 0.001) to NCBI-nr and NCBI-nt have been recovered.

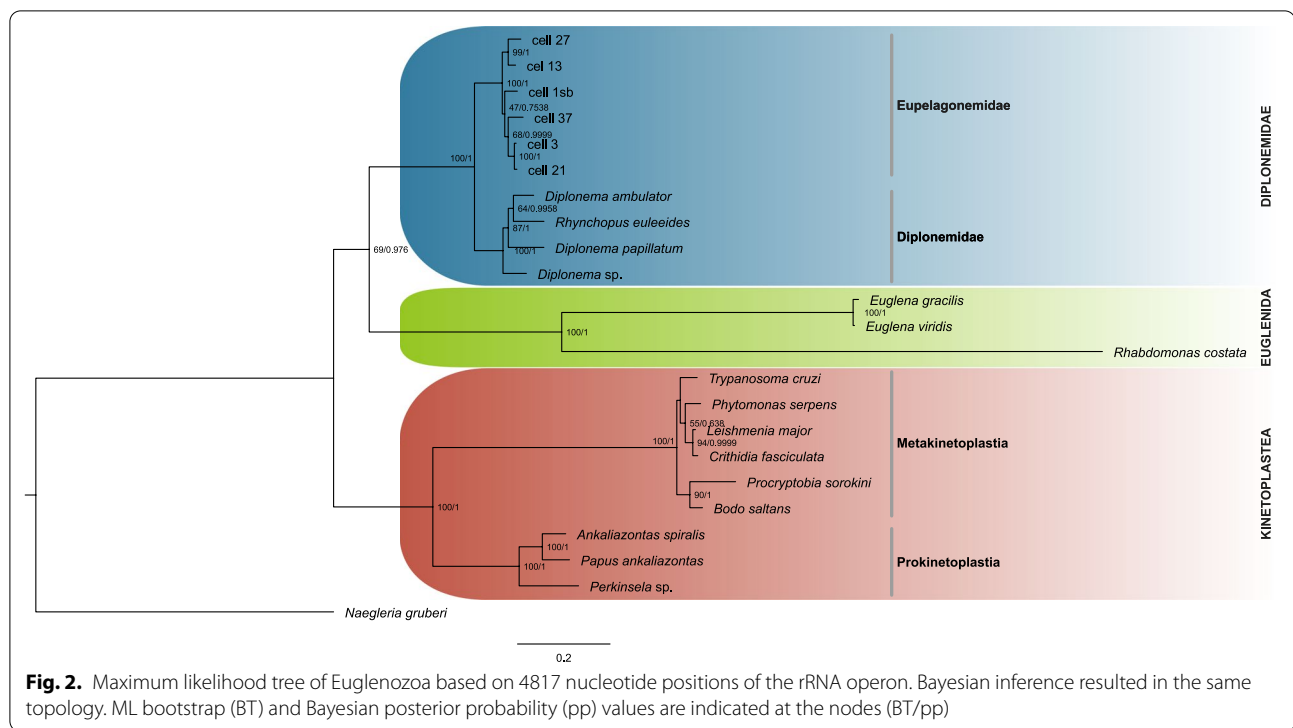
To provide background for structural analyses we reconstructed the phylogeny of Euglenozoa based on the 18S-5.8S-28S rRNA coding region (Fig. 2). All three major groups of euglenozoans form maximally supported clades (100 bootstrap for ML and 1.00 posterior probability for BI). In spite of that, relations between groups are not resolved, though this is typical for rRNA phylogenies of Euglenozoa [9, 28, 29]. The internal topology for euglenids, kinetoplastids and diplomemids is as expected [1], moreover, within diplomemids, the division between Diplonemidae and Eupelagonemidae is maximally supported, and the internal topologies of these families are in agreement with the previous analysis [6].

Discussion

The typical eukaryotic 18S-28S rDNA unit comprises co-transcribed 18S, 5.8S and 28S rRNA separated by ITS1 and ITS2, which are removed in post-transcriptional

processing. The ITS2 is a eukaryotic invention—the 23S rRNA present in prokaryotes comprises both 5.8S and 28S rRNA structure, and is separated from 16S rRNA (the prokaryotic equivalent of 18S) by a single ITS. The length and secondary structure of the ITSs are not conserved, with the shortest ITSs observed in the protist parasite *Giardia intestinalis*, and the longest—in multicellular eukaryotes [11, 30]. The elongation is usually a result of insertion of short tandem repeats, but the functional consequence of such elongations is unknown.

Fragments of the rRNA (both 18S and 28S) forming external (more distant from the site of peptidyl transfer) parts of the ribosome are much less conserved than the internal fragments. For this reason, externally located variable regions (or expansion segments, ES) show much greater variability in sequence, structure and length [25]. Expansion of these segments causes the size of mature 28S rRNA to vary from ~2500 bp in microsporidia to over 5000 bp in multicellular species, such as humans. Interestingly, the LSU rRNA of microsporidia is a fusion of 5.8S and 28S rRNA, with a structure more similar to prokaryotes than other eukaryotes [31]. In several distinct eukaryotic lineages an opposite process occurred, resulting in the formation of a fragmented mature 28S rRNA. The best-known example is the presence of the so-called “hidden break” in insects and other proto-stome animals, causing the RNA isolates to seem to be degraded [19, 32, 33]. An analogous situation is observed in several mammals, mainly rodents [20, 34, 35]. It is worth mentioning that insect and mammalian “hidden breaks”, or rather additional ITSs, are present in different expansions’ segments (ES19 and ES15, respectively). Furthermore, in the case of the rodent *Ctenomys*, the additional ITS is present within an intron. Said intron is excised or retained in a tissue-specific fashion, resulting



in the absence or presence of the “hidden break”, leading to continuous or fragmented mature 28S rRNA [20]. Different ribosome structures in different tissues may suggest the functional importance of the additional ITS in *Ctenomys*. Another notable example exists in malaria-causing apicomplexan *Plasmodium falciparum*, in which two types of 28S rRNA units are present: continuous A-type and fragmented S-type [36]. The expression of one or the other type is strictly regulated (e.g., by temperature and glucose concentration), with only the continuous A-type expressed in the vertebrate host [21, 37, 38]. Other non-homologous additional ITSs can be found in Amoebozoa [39], dinoflagellates [40] and in mitochondria or plastid 23–28S rRNA [25]. However, the number of additional ITSs present in kinetoplastids and euglenids is unmatched in any other taxa.

Newly acquired rRNA structures of nine diplomonids show that the lack of additional ITSs in *D. papillatum* is, in fact, typical for the Diplonemea. This finding is significant for elusive taxa like diplomonids, known mostly from metabarcoding data. Continuous 28S rRNA allows the employment of third-generation sequencing (PacBio, MinION) in both DNA and RNA surveys [41–43].

However, such a result is a surprise from the evolutionary point of view. The presence of additional ITSs in both euglenids and kinetoplastids suggests that it may be another ancestral feature of Euglenozoa, especially since three of them share positions between groups [44]. In

such a scenario, continuous 28S rRNA in *D. papillatum* could be coincidental – species-specific secondary losses of additional ITSs (the aforementioned “hidden breaks”) are known in insects [19]. Lack of additional ITSs in all diplomonids rules out this possibility. Similarly, the lack of additional ITSs in Prokinetoplastida indicates that the last common ancestor of kinetoplastids had a continuous 28S rRNA. In such a case, additional ITSs found in kinetoplastids could be common only for Trypanosomatidae and Bodonidae, but exact pinpointing of their origin requires additional surveys across kinetoplastids [1]. If additional ITSs are neither an ancestral feature of kinetoplastids nor present in diplomonids, they cannot be common in Euglenozoa.

Based on the obtained results, it seems that the kinetoplastids’ and euglenids’ additional ITSs emerged independently. However, it is highly unlikely that the occurrence of additional ITSs in such unparalleled numbers in two closely related groups is a coincidence. It seems most probable that some factor in euglenozoan biology makes fragmentation of the 28S rRNA more feasible than in other eukaryotes. One possible explanation is the ribosomal protein repertoire unique to this group. It has been shown that post-transcriptional removal of additional ITSs in *T. brucei* is guided by ribosomal proteins [45]. In general, kinetoplastid ribosomal proteins exhibit a number of unusual features interacting with unusual rRNA [22, 46, 47]. Even more oddities have

been found in cryo-EM structures of the *E. gracilis* ribosome [48]. Small rRNAs termini colocalise, mostly in two focal points. Several ribosomal proteins exhibit unusual elongations interacting with the expansion segments of *E. gracilis* rRNA, and four novel *Euglena*-specific ribosomal proteins have been found, three of them interacting with unique LSU rRNA motifs/deletions. Furthermore, *E. gracilis* rRNA was found to bear the highest number of ribosomal post-transcriptional modifications reported to date [49]. The frequency of modifications is much higher in the LSU, correlating with a high level of rRNA fragmentation. Similarly, a number of unique RNA modifications have been found in the proximity of additional ITSs in *T. brucei* [50]. A group of such modifications appears late in the maturation of the ribosome, at the same stage as ITSs removal. In any case, the co-presence of RNA modifications, unusual ribosomal proteins and additional ITSs suggests close correlation. Answering the “chicken or egg” question about their origin will require additional data with a better phylogenetic representation of euglenids and diplomonads.

Conclusions

We acquired novel complete rRNA operons for six kinetoplastids, two euglenids, three classical and six marine diplomonads. All analysed diplomonads lack additional ITSs known from other euglenozoans. Interestingly, while all investigated metakinetoplastids have the exact same pattern of ITSs as trypanosomes, the early branching prokinetoplastids do not possess any additional ITSs. These results suggest that additional ITSs in euglenids and kinetoplastids are of independent origins.

Methods

Genome assemblies of classical and deep-sea diplomonads were accessed [5]. During the initial analysis of the original assemblies, we have found highly fragmented rRNA operons only. For that reason, raw reads for each species were obtained and reassembled (Additional file 2: Table S2). The quality of raw reads was evaluated using FastQC v0.11.5 [51] and trimmed in Trimmomatic v0.36 [52]. Processed reads were assembled using metaSPAdes v3.10.1 [53, 54]. Acquired assemblies were searched by BLASTn with rRNA sequences of *Diplonema papillatum* (KF633466-8) as queries. To exclude potential mitochondrial or contaminant rRNA operons and potential misassemblies only high scoring hits ($e\text{-value} < 10^{-5}$) with high coverage ($> 5 \times$ higher than genome average) were kept. We have found that the newly performed metaSPAdes assemblies contained rRNA operons of better quality, and therefore they were used for further analyses. Assembly graphs were manually inspected in Bandage [55] to identify potential misassemblies. In such a case,

contigs containing rRNA operons were manually corrected and replaced in the assemblies. Furthermore, the acquired operons were manually checked for mismatches since metaSPAdes does not support mismatch correction [54].

To provide phylogenetic background, we searched genomes of kinetoplastids and euglenids for rRNA operons by BLASTn. The operons of *E. gracilis* (M12677.1, X53361.2) and *Crithidia fasciculata* (Y00055.1) were used as queries for euglenids and kinetoplastids, respectively. Genome assemblies of *Perkinsella* sp. (LFNC01000001.1) and *Phytomonas serpens* (AIHY00000000.1) were accessed from GenBank [56, 57]. Raw reads were accessed (last time on 05/05/21) for *Euglena viridis* (SRR14099996) [58], *Rhabdomonas costata* [59], *B. saltans* (ERR036178) [60], *Papus ankaliizontas* (SRR13394431), *Ankaliizontas spiralis* and *Procrystobia sorokini* (cocultured, SRR13394430) [61]. These were processed in the same way as diplomonad assemblies. Lastly, the rRNA operon sequence of *Naegleria gruberi* (AB298288.1) was accessed as a non-euglenozoan outgroup.

Sequences of rRNA operons were aligned using MAFFT eini [62]. The obtained alignment was further manually edited in Geneious v10.2.2 [63], based upon annotated secondary structures. The secondary structure of *E. gracilis* rRNA has been predicted [16, 17], while for several trypanosomatid ribosomes, cryo-electron microscopy structures have been obtained [23, 46, 64]. RNApdbe 2.0 web-server [65] was used to extract secondary structures from available cryo-electron microscopy models. Secondary structures of *E. gracilis*, *T. cruzi* and *L. major* have been modelled, based on previously published structures. Determined helices were marked upon the alignment and numbered following a previously published structure of *Saccharomyces cerevisiae* rRNA structure [66]. Using this profile, structures of particular domains and regions have been predicted for all species using the RNAfold WebServer [67, 68]. Several intervals have been used in each case to best identify structural elements, i.e., helices and loops composing the bulk structure of the ribosome. Helices have been numbered in the same manner and marked upon sequences of all newly analysed species. Based on this annotation, homologous helices were manually aligned to prepare structure-based alignment which was used to identify irregularities in the lengths of the analysed structures.

All identified expansions have been investigated for possible homologies on sequence or structure level and checked for presence of open reading frames or other potentially coding fragments. Their sequences were searched by RNAmmer [27], BLASTn against NCBI-nt and BLASTx against NCBI-nr. The expansions occurring

in sites of known additional ITSs in *E. gracilis* and *T. cruzi* were described as corresponding ITS. Unusually large ($>4 \times$ longer than in other species) expansions found in other divergent regions were marked as a potential novel additional ITSs.

An alignment produced by MAFFT was used for phylogenetic analyses. Fragments with very high variance and no conserved secondary structure were manually removed (e.g., ITSs), with retained alignment trimmed using trimAL v1.2rev59 with `-automated1` option [69]. The remaining 4817 positions were used for phylogenetic analyses. A maximum-likelihood tree (ML) was calculated using raxml-ng [70], with GTR + I + G4 model of substitution chosen by modeltest-ng [71]. The best tree was estimated using 20 different starting trees and 1000 bootstraps. Bayesian inference was performed in MrBayes v3.2.6 [72]. Two runs of a Markov Chain Monte Carlo were carried out with four chains (one cold and three heated), with GTR + I + G model of substitution, 10 million generations, trees sampled every 100 generations and the burn-in set to the first 25% of the sample.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-022-02014-9>.

Additional file 1. Figure S1. Location of additional ITSs within secondary structure of LSU.

Additional file 2. Table S1-2. Identified segments of the rDNA and their lengths. List of analysed species and source of data.

Acknowledgements

We would like to thank Vladimir Hampl for providing access to genomic data of *Rhabdomonas costata* prior to its publication. We thank Kacper Maciszewski and Alicja Fells for review and providing comments that improved the quality and clarity of the manuscript.

Author contributions

PH and RM designed the study. PH carried out the analyses. PH, AK and RM interpreted the data. PH prepared figures, tables and wrote the manuscript. PH, AK and RM inspected the manuscript critically and took part in the revision of manuscript. All authors read and approved the final manuscript.

Funding

This work was funded by Grant 2015/19/B/NZ8/00166 to R.M. from the National Science Centre, Poland. A.K. has been supported by the EMBO Installation Grant.

Availability of data and materials

All the datasets generated and/or analysed during the current study are available in the RepOD repository, <https://doi.org/10.18150/J4Q2ES>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 17 November 2021 Accepted: 22 April 2022

Published online: 09 May 2022

References

- Kostygov AY, Karnkowska A, Votýpka J, Tashyreva D, Maciszewski K, Yurchenko V, et al. Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open Biol.* 2021;11:200407. <https://doi.org/10.1098/rsob.200407>.
- Flegontova O, Flegontov P, Malviya S, Audic S, Wincker P, de Vargas C, et al. Extreme diversity of diplomonid eukaryotes in the ocean. *Curr Biol.* 2016;26:3060–5. <https://doi.org/10.1016/j.cub.2016.09.031>.
- Schoenle A, Hohlfeld M, Hermanns K, Mahé F, de Vargas C, Nitsche F, et al. High and specific diversity of protists in the deep-sea basins dominated by diplomonids, kinetoplastids, ciliates and foraminiferans. *Commun Biol.* 2021;4:1–10. <https://doi.org/10.1038/s42003-021-02012-5>.
- de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R, et al. Eukaryotic plankton diversity in the sunlit ocean. *Science.* 2015;348:1261605–1261605. <https://doi.org/10.1126/science.1261605>.
- Gawryluk RMR, del Campo J, Okamoto N, Strassert JFH, Lukeš J, Richards TA, et al. Morphological identification and single-cell genomics of marine diplomonids. *Curr Biol.* 2016;26:3053–9. <https://doi.org/10.1016/j.cub.2016.09.013>.
- Okamoto N, Gawryluk RMR, del Campo J, Strassert JFH, Lukeš J, Richards TA, et al. A revised taxonomy of Diplomonads including the Eupelagonemidae n. fam. and a type species, *Eupelagonema oceanica* n. gen. & sp. *J Eukaryot Microbiol.* 2019;66:519–24. <https://doi.org/10.1111/jeu.12679>.
- Milanowski R, Gumińska N, Karnkowska A, Ishikawa T, Zakryś B. Intermediate introns in nuclear genes of euglenids - Are they a distinct type? *BMC Evol Biol.* 2016;16:1–11. <https://doi.org/10.1186/s12862-016-0620-5>.
- Lukeš J, Wheeler R, Jirsová D, David V, Archibald JM. Massive mitochondrial DNA content in diplomonid and kinetoplastid protists. *IUBMB Life.* 2018;70:1267–74. <https://doi.org/10.1002/iub.1894>.
- Záhonová K, Lax G, Sinha SD, Leonard G, Richards TA, Lukeš J, et al. Single-cell genomics unveils a canonical origin of the diverse mitochondrial genomes of euglenozoans. *BMC Biol.* 2021;19:1–14. <https://doi.org/10.1186/s12915-021-01035-y>.
- Wideman JG, Lax G, Leonard G, Milner DS, Rodríguez-Martínez R, Simpson AGB, et al. A single-cell genome reveals diplomonidlike ancestry of kinetoplastid mitochondrial gene structure. *Philos Trans R Soc B Biol Sci.* 2019. <https://doi.org/10.1098/rstb.2019.0100>.
- Torres-Machorro AL, Hernández R, Cevallos AM, López-Villaseñor I. Ribosomal RNA genes in eukaryotic microorganisms: Witnesses of phylogeny? *FEMS Microbiol Rev.* 2010;34:59–86. <https://doi.org/10.1111/j.1574-6976.2009.00196.x>.
- Valach M, Moreira S, Kiethega GN, Burger G. Trans-splicing and RNA editing of LSU rRNA in *Diplonema* mitochondria. *Nucleic Acids Res.* 2014;42:2660–72. <https://doi.org/10.1093/nar/gkt1152>.
- Spencer DF, Collings JC, Schnare MN, Gray MW. Multiple spacer sequences in the nuclear large subunit ribosomal RNA gene of *Crithidia fasciculata*. *EMBO J.* 1987;6:1063–71. <https://doi.org/10.1002/j.1460-2075.1987.tb04859.x>.
- Hernández R, Díaz-de León F, Castañeda M. Molecular cloning and partial characterization of ribosomal RNA genes from *Trypanosoma cruzi*. *Mol Biochem Parasitol.* 1988;27:275–9. [https://doi.org/10.1016/0166-6851\(88\)90047-3](https://doi.org/10.1016/0166-6851(88)90047-3).
- Martínez-Calvillo S, Sunkin SM, Yan S, Fox M, Stuart K, Myler PJ. Genomic organization and functional characterization of the *Leishmania major* Friedlin ribosomal RNA gene locus. *Mol Biochem Parasitol.* 2001;116:147–57. doi:[https://doi.org/10.1016/S0166-6851\(01\)00310-3](https://doi.org/10.1016/S0166-6851(01)00310-3).
- Schnare MN, Gray MW. Sixteen discrete RNA components in the cytoplasmic ribosome of *Euglena gracilis*. *J Mol Biol.* 1990;215:73–83. [https://doi.org/10.1016/S0022-2836\(05\)80096-8](https://doi.org/10.1016/S0022-2836(05)80096-8).
- Smallman DS, Schnare MN, Gray MW. RNA:RNA interactions in the large subunit ribosomal RNA of *Euglena gracilis*. *Biochim Biophys Acta - Gene Struct Expr.* 1996;1305:1–6. [https://doi.org/10.1016/0167-4781\(95\)00204-9](https://doi.org/10.1016/0167-4781(95)00204-9).

18. Greenwood SJ, Gray M. Processing of precursor rRNA in *Euglena gracilis*: identification of intermediates in the pathway to a highly fragmented large subunit rRNA. *Biochim Biophys Acta Gene Struct Expr*. 1998;1443:128–38. [https://doi.org/10.1016/S0167-4781\(98\)00201-2](https://doi.org/10.1016/S0167-4781(98)00201-2).
19. Macharia RW, Ombura FL, Aroko EO. Insects' RNA profiling reveals absence of "hidden break" in 28S ribosomal RNA molecule of onion thrips. *Thrips tabaci J Nucleic Acids*. 2015. <https://doi.org/10.1155/2015/965294>.
20. Melen GJ, Pesce CG, Rossi MS, Kornblihtt AR. Novel processing in a mammalian nuclear 28S pre-rRNA: Tissue-specific elimination of an "intron" bearing a hidden break site. *EMBO J*. 1999;18:3107–18. <https://doi.org/10.1093/emboj/18.11.3107>.
21. Fang J, Sullivan M, McCutchan TF. The effects of glucose concentration on the reciprocal regulation of rRNA promoters in *Plasmodium falciparum*. *J Biol Chem*. 2004;279:720–5. <https://doi.org/10.1074/jbc.M308284200>.
22. Liu Z, Gutierrez-Vargas C, Wei J, Grassucci RA, Ramesh M, Espina N, et al. Structure and assembly model for the *Trypanosoma cruzi* 60S ribosomal subunit. *Proc Natl Acad Sci U S A*. 2016;113:12174–9. <https://doi.org/10.1073/pnas.1614594113>.
23. Shalev-Benami M, Zhang Y, Matzov D, Halfon Y, Zackay A, Rozenberg H, et al. 2.8-Å Cryo-EM structure of the large ribosomal subunit from the eukaryotic parasite *Leishmania*. *Cell Rep*. 2016;16:288–94. <https://doi.org/10.1016/j.celrep.2016.06.014>.
24. Hassouna N, Mithot B, Bachelier J-P. The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. *Nucleic Acids Res*. 1984;12:3563–83. doi:<https://doi.org/10.1093/nar/12.8.3563>.
25. Gerbi SA. Expansion segments: regions of variable size that interrupt the universal core secondary structure of ribosomal RNA. In: *Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis*. 1996. p. 71–87.
26. Schnare MN, Cook JR, Gray MW. Fourteen internal transcribed spacers in the circular ribosomal DNA of *Euglena gracilis*. *J Mol Biol*. 1990;215:85–91. [https://doi.org/10.1016/S0022-2836\(05\)80097-X](https://doi.org/10.1016/S0022-2836(05)80097-X).
27. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res*. 2007;35:3100–8. <https://doi.org/10.1093/nar/gkm160>.
28. Cavalier-Smith T. Higher classification and phylogeny of Euglenozoa. *Eur J Protistol*. 2016;56:250–76. <https://doi.org/10.1016/j.ejop.2016.09.003>.
29. Kolisko M, Flegontova O, Karnkowska A, Lax G, Maritz JM, Pánek T, et al. EukRef-excavates: Seven curated SSU ribosomal RNA gene databases. *Database*. 2020;2020:1–11. <https://doi.org/10.1093/database/baaa080>.
30. Schultz J, Maisel S, Gerlach D, Müller T, Wolf M. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA*. 2005;11:361–4. <https://doi.org/10.1261/rna.7204505>.
31. Peyretailade E, Biderre C, Peyret P, Duffieux F, Méténier G, Gouy M, et al. Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. *Nucleic Acids Res*. 1998;26:3513–20. <https://doi.org/10.1093/nar/26.15.3513>.
32. Winnebeck EC, Millar CD, Warman GR. Why does insect RNA look degraded? *J Insect Sci*. 2010;10:159. <https://doi.org/10.1673/031.010.14119>.
33. Fujiwara H, Ishikawa H. Molecular mechanism of introduction of the hidden break into the 28S rRNA of insects: implication based on structural studies. *Nucleic Acids Res*. 1986;14:6393–401. <https://doi.org/10.1093/nar/14.16.6393>.
34. Henras AK, Plisson-Chastang C, O'Donohue MF, Chakraborty A, Gleizes PE. An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip Rev RNA*. 2015;6:225–42. <https://doi.org/10.1002/wrna.1269>.
35. Natsidis P, Schiffer PH, Salvador-Martinez I, Telford MJ. Computational discovery of hidden breaks in 28S ribosomal RNAs across eukaryotes and consequences for RNA Integrity Numbers. *Sci Rep*. 2019;9:1–10. <https://doi.org/10.1038/s41598-019-55573-1>.
36. Dame JB, McCutchan TF. Cloning and characterization of a ribosomal RNA gene from *Plasmodium berghei*. *Mol Biochem Parasitol*. 1983;8:263–79. [https://doi.org/10.1016/0166-6851\(83\)90048-8](https://doi.org/10.1016/0166-6851(83)90048-8).
37. Fang J, McCutchan TF. Malaria: Thermoregulation in a parasite's life cycle. *Nature*. 2002;418:742. <https://doi.org/10.1038/418742a>.
38. Waters AP, Syn C, McCutchan TF. Developmental regulation of stage-specific ribosome populations in *Plasmodium*. *Nature*. 1989;342:438–40. <https://doi.org/10.1038/342438a0>.
39. D'Alessio JM, Harris GH, Perna PJ, Paule MR. Ribosomal ribonucleic acid repeat unit of *Acanthamoeba castellanii*: cloning and restriction endonuclease map. *Biochemistry*. 1981;20:3822–7. <https://doi.org/10.1021/bi00516a024>.
40. Lenaers G, Maroteaux L, Michot B, Herzog M. Dinoflagellates in evolution A molecular phylogenetic analysis of large subunit ribosomal RNA. *J Mol Evol*. 1989;29:40–51. <https://doi.org/10.1007/BF02106180>.
41. Tedersoo L, Tooming-Klunderud A, Anslan S. PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytol*. 2018;217:1370–85. <https://doi.org/10.1111/nph.14776>.
42. Heeger F, Bourne EC, Baschien C, Yurkov A, Bunk B, Spröer C, et al. Long-read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments. *Mol Ecol Resour*. 2018;18:1500–14. <https://doi.org/10.1111/1755-0998.12937>.
43. Jamy M, Foster R, Barbera P, Czech L, Kozlov A, Stamatakis A, et al. Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. *Mol Ecol Resour*. 2020;20:429–43. <https://doi.org/10.1111/1755-0998.13117>.
44. Vesteg M, Hadariová L, Horváth A, Estraño CE, Schwartzbach SD, Krajčovič J. Comparative molecular cell biology of phototrophic euglenids and parasitic trypanosomatids sheds light on the ancestor of Euglenozoa. *Biol Rev*. 2019;94:1701–21. <https://doi.org/10.1111/brv.12523>.
45. Jaremko D, Ciganda M, Christen L, Williams N. *Trypanosoma brucei* L11 is essential to ribosome biogenesis and interacts with the kinetoplastid-specific proteins P34 and P37. *Sphere*. 2019;4:1–16. <https://doi.org/10.1128/msphere.00475-19>.
46. Hashem Y, Des Georges A, Fu J, Buss SN, Jossinet F, Jobe A, et al. High-resolution cryo-electron microscopy structure of the *Trypanosoma brucei* ribosome. *Nature*. 2013;494:385–9. <https://doi.org/10.1038/nature11872>.
47. Brito Querido J, Mancera-Martínez E, Vicens Q, Bochler A, Chicher J, Simonetti A, et al. The cryo-EM structure of a novel 40S kinetoplastid-specific ribosomal protein. *Structure*. 2017;25:1785–1794.e3. <https://doi.org/10.1016/j.str.2017.09.014>.
48. Matzov D, Taoka M, Nobe Y, Yamauchi Y, Halfon Y, Asis N, et al. Cryo-EM structure of the highly atypical cytoplasmic ribosome of *Euglena gracilis*. *Nucleic Acids Res*. 2020;48:11750–61. <https://doi.org/10.1093/nar/gkaa893>.
49. Schnare MN, Gray MW. Complete modification maps for the cytosolic small and large subunit rRNAs of *Euglena gracilis*: Functional and evolutionary implications of contrasting patterns between the two rRNA components. *J Mol Biol*. 2011;413:66–83. <https://doi.org/10.1016/j.jmb.2011.08.037>.
50. Chikne V, Rajan KS, Shalev-Benami M, Decker K, Cohen-Chalamish S, Madmoni H, et al. Small nucleolar RNAs controlling rRNA processing in *Trypanosoma brucei*. *Nucleic Acids Res*. 2019;47:2609–29. <https://doi.org/10.1093/nar/gky1287>.
51. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data. 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
52. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
53. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77. <https://doi.org/10.1089/cmb.2012.0021>.
54. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: A new versatile metagenomic assembler. *Genome Res*. 2017;27:824–34. <https://doi.org/10.1101/gr.213959.116>.
55. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics*. 2015;31:3350–2. <https://doi.org/10.1093/bioinformatics/btv383>.
56. Koreny L, Sobotka R, Kovarova J, Gnypova A, Flegontov P, Horvath A, et al. Aerobic kinetoplastid flagellate *Phytomonas* does not require heme for viability. *Proc Natl Acad Sci*. 2012;109:3808–13.
57. David V, Flegontov P, Gerasimov E, Tanifuji G, Hashimi H, Logacheva MD, et al. Gene loss and error-prone RNA editing in the mitochondrion of *perkinsella*, an endosymbiotic kinetoplastid. *MBio*. 2015;6:e01498-e1515. <https://doi.org/10.1128/mBio.01498-15>.

58. Nelson DR, Hazzouri KM, Lauersen KJ, Jaiswal A, Chaiboonchoe A, Mystikou A, et al. Large-scale genome sequencing reveals the driving forces of viruses in microalgal evolution. *Cell Host Microbe*. 2021;29:250-266.e8. <https://doi.org/10.1016/j.chom.2020.12.005>.
59. Soukal P, Hrdá Š, Karnkowska A, Milanowski R, Szabová J, Hradilová M, et al. Heterotrophic Euglenid *Rhabdomonas Costata* Resembles Its Phototrophic Relatives in Many Aspects of Molecular and Cell Biology. *Sci Rep*. 2021;11:1–17. <https://doi.org/10.1038/s41598-021-92174-3>.
60. Jackson AP, Otto TD, Aslett M, Armstrong SD, Bringaud F, Schlacht A, et al. Kinetoplastid phylogenomics reveals the evolutionary innovations associated with the origins of parasitism. *Curr Biol*. 2016;26:161–72. <https://doi.org/10.1016/j.cub.2015.11.055>.
61. Tikhonenkov DV, Gawryluk RMR, Mylnikov AP, Keeling PJ. First finding of free-living representatives of Prokinetoplastina and their nuclear and mitochondrial genomes. *Sci Rep*. 2021;11:1–21. <https://doi.org/10.1038/s41598-021-82369-z>.
62. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80. <https://doi.org/10.1093/molbev/mst010>.
63. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28:1647–9. <https://doi.org/10.1093/bioinformatics/bts199>.
64. Gao H, Juri Ayub M, Levin MJ, Frank J. The structure of the 80S ribosome from *Trypanosoma cruzi* reveals unique rRNA components. *Proc Natl Acad Sci U S A*. 2005;102:10206–11. <https://doi.org/10.1073/pnas.0500926102>.
65. Zok T, Antczak M, Zurkowski M, Popenda M, Blazewicz J, Adamiak RW, et al. RNApdbee 2.0: multifunctional tool for RNA structure annotation. *Nucleic Acids Res*. 2018;46:W30–5. <https://doi.org/10.1093/nar/gky314>.
66. Bernier CR, Petrov AS, Waterbury CC, Jett J, Li F, Freil LE, et al. RiboVision suite for visualization and analysis of ribosomes. *Faraday Discuss*. 2014;169:195–207. <https://doi.org/10.1039/C3FD00126A>.
67. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Res*. 2008;36:70–4. <https://doi.org/10.1093/nar/gkn188>.
68. Lorenz R, Bernhart SH, Höner zu Siederdisen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. *Algorithms Mol Biol*. 2011;6:26. <https://doi.org/10.1186/1748-7188-6-26>.
69. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3. <https://doi.org/10.1093/bioinformatics/btp348>.
70. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. 2019;35:4453–5. <https://doi.org/10.1093/bioinformatics/btz305>.
71. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. Model-Test-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol Biol Evol*. 2020;31:291–4. <https://doi.org/10.1101/612903>.
72. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:539–42. <https://doi.org/10.1093/sysbio/sys029>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

