Typical Structure of Rrna Coding Genes In Diplonemids Points To Two Independent Origins of The Bizarre Rdna Structures of Euglenozoans

Paweł Hałakuc (✉ pawelhalakuc@uw.edu.pl)

University of Warsaw

Anna Karnkowska

University of Warsaw

Rafał Milanowski

University of Warsaw

Research Article

Keywords: rRNA, rDNA, rRNA operon, Euglenozoa, diplonemids, euglenids, kinetoplastids, internal transcribed spacer

Posted Date: February 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1089468/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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 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 diplonemid species, *Diplonema papillatum*, has been examined so far and found to exhibit continuous 28S rRNA. The rDNA organization has not been researched for any diplonemid so far. Herein we investigate the structure of rRNA genes in classical (Diplonemidae) and deep-sea diplonemids (Eupelagonemidae), representing the majority of known diplonemid diversity. The results fill the gap in knowledge about diplonemid rDNA and allow for better understanding of the evolution of fragmented structure of the rDNA in Euglenozoa,

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

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

Background

Until a few years ago, Diplonemea (Euglenozoa, Discoba) was a rather neglected group. Only two diplonemid genera, *Diplonema* and *Rhynchopus*, have been described and cultured. This remained in sharp contrast to other well-researched euglenozoans: Kinetoplastea, sister to Diplonemea, and Euglenida [1]. Lately, metabarcoding surveys from the deep pelagic zone and deep-sea sediments have shown their unrivaled diversity [2–4]. Based on those 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]. 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, partially due to high repetitiveness, caused by an unexpectedly high density of ‘noncanonical’ introns, similar to euglenid nonconventional introns [5, 7].

However, even for such incomplete assemblies, regions present in many copies, such as mtDNA or nuclear rDNA operon, can be extracted [8, 9]

Typically, eukaryotic ribosomes contain four ribosomal RNAs. Three of them (18S, 5.8S and 28S rRNA) are encoded in a single operon (rDNA) and co-transcribed. Genes are separated by internal transcribed spacers (ITSs), which are removed during post-transcriptional processing to form mature rRNAs [10]. Such a structure of four continuous rRNAs has been confirmed in the single investigated diplomonid *Diplonema papillatum* [11]. 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 [12–14], and 13 in euglenids [15–17]. Those smaller rRNAs together perform structural and catalytic functions of typical 28S rDNA. The fragmentation is caused by additional ITSs in the rDNA operon of both euglenids and kinetoplastids. While 28S fragmentation occasionally occurs in various eukaryotes [18–20], the extent of the fragmentation in Euglenozoa is unparalleled. The lack of studied rDNA operons in diplomonids puts the parsimonious (i.e., involving single ancestral acquisition) evolutionary path of euglenozoan rDNA operon into question, which we try to answer herein.

Results

We successfully assembled complete rDNA operons of all three classical and six of ten deep-sea diplomonids, including the most abundant *Eupelagonema oceanica* (Supplementary Table S1). Lengths of all acquired operons and their subunits are typical for eukaryotes. We identified all conserved features of the 28S rRNA and no significant insertions or deletions were found (Supplementary Table S1, Figure 1). For that reason, we conclude that no additional ITS is present in the rDNA of analysed diplomonids, resulting in typical eukaryotic continuous 28S rRNA.

The rDNAs of all three euglenids – one heterotrophic species and two phototrophs – are significantly elongated (10,000-13,000 bp), mainly due to the large expansions within SSU and LSU genes. As previously suggested, almost all of the expansions occur in divergent regions of the rRNA, also known as expansion segments (ES) [21, 22]. We identified the ITSs described for *E. gracilis* [23] in both analysed euglenids. Moreover, two potential novel additional ITS sites were found in *Rhabdomonas costata*. None of the expansions were recognised as group I introns.

In kinetoplastids, two types of the rDNA operons can be distinguished: elongated one in metakinetoplastids (trypanosomatids and bodonids), and typical eukaryotic in all protokinetoplastids (*Perkinsella* sp. and similar). Long rDNA operons in metakinetoplastids originated from the elongations of the LSU, which are present in the same positions and possess the same features and spatial distribution pattern as previously described additional ITSs [12–14]. All analysed metakinetoplastid sequences have the exact same pattern of additional ITSs as trypanosomatids (Figure 1). In *Bodo saltans*, basal metakinetoplastid, expansion in the kinetoplastid ITS-3 site is short but still much longer than in protokinetoplastids and non-kinetoplastid species. It suggests that *Bodo saltans* rDNA does contain

kinetoplastid ITS-3. Furthermore, only three kinetoplastid ITSs share positions with euglenid ITSs (Figure 1).

Reconstructed phylogenies are shown in Figure 2. All three major groups of euglenozoans form maximally supported clades (100 bootstrap for ML and 1.00 posterior probability for BI). Relations between groups are not resolved, which is typical for rDNA phylogenies of Euglenozoa [8, 24, 25]. We recovered the expected internal topology for euglenids and kinetoplastids [1]. Within diplomonids, the division between Diplomonadidae and Eupelagomonadidae is maximally supported, and the internal topologies of those 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 ITS-1 and ITS-2, which are removed in post-transcriptional processing. The ITS-2 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 weakly conserved, with the shortest ITSs observed in the parasite *Giardia intestinalis*, and the longest – in multicellular eukaryotes [10, 26]. The elongation is usually the result of insertion of short tandem repeats, but biological consequence of such elongations is unknown.

Fragments of the rRNA (both 18S and 28S) forming external parts of the ribosome are much less conserved than internal fragments. For this reason, externally located variable regions (or expansion segments, ES) show much greater variability in sequence, structure and length [22]. Expansion of those segments causes the size of mature 28S rRNA to vary from ~2,500 nt in microsporidia to over 5,000 nt in multicellular species, such as human. 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 [27]. In several distinct eukaryotic lineages, an opposite process occurred, resulting in formation of a fragmented mature 28S rRNA. The best-known example is the presence of the so-called “hidden break” in insects and other protostome animals, causing the RNA isolates to seem to be degraded [18, 28, 29]. An analogous situation is observed in several mammals, mainly rodents [19, 30, 31]. It is worth mentioning that insect and mammalian “hidden breaks”, or rather additional ITSs, are present in different expansion 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 [19]. Different ribosome structures in different tissues may suggest the functional importance of 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 [32]. 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 host [20, 33, 34]. Other non-homologous additional ITSs can be found in Amoebozoa [35], dinoflagellates [36] and mitochondrial or plastid

LSUs [22]. However, the number of additional ITSs present in kinetoplastids and euglenids is unmatched in any other taxa.

Newly acquired rDNA structures of nine diplonemids show that the lack of additional ITSs in *Diplonema papillatum* is, in fact, typical for the Diplonemea. This finding is significant for elusive taxa like diplonemids, known mostly from metabarcoding data. Continuous LSU allows for the employment of third-generation sequencing (PacBio, MinION) in both DNA and RNA surveys [37–39].

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 an ancestral feature of Euglenozoa, especially since three of them share positions between groups. In such a scenario, continuous LSU in *Diplonema papillatum* could be coincidental – species-specific secondary losses of additional ITSs (the aforementioned “hidden breaks”) are known in insects [18]. Lack of additional ITSs in all diplonemids rules out this possibility. Similarly, the lack of additional ITSs in Protokinetoplastia indicates that the last common ancestor of kinetoplastids had a continuous LSU. In such a case, additional ITSs found in kinetoplastids could be common only to 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 diplonemids, they cannot be common to the Euglenozoa.

Based on the obtained results, the kinetoplastids’ and euglenids’ additional ITSs emerged independently. 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 LSU more feasible than in other eukaryotes. One possible explanation is the unique to this group ribosomal protein repertoire. It has been shown that post-transcriptional removal of additional ITSs in *T. brucei* is guided by ribosomal proteins [40]. In general, kinetoplastid ribosomal proteins exhibit a number of unusual features interacting with unusual rRNA [41–43]. Even more oddities have been found in cryo-EM structures of *E. gracilis* ribosome [44]. Small rRNAs termini colocalise, mostly in two focal points. Several ribosomal proteins exhibit unusual elongations interacting with 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. 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*. A group of such modifications appears late in the maturation of ribosome, at the same stage as ITSs removal [45]. In any case, the co-presence of RNA modifications, unusual riboproteins and additional ITSs suggests close correlation. Answering the “chicken and egg” question about their origin will require additional data with a better phylogenetic representation of euglenids and diplonemids.

Conclusions

We acquired novel complete rDNA 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 protokinetoplastids do not possess any additional ITSs. Those 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 only found highly fragmented rDNA operons. For that reason, raw reads for each species were also obtained and reassembled (Supplementary Table S2). The quality of raw reads was evaluated using FastQC v0.11.5 [46] and trimmed in Trimmomatic v0.36 [47]. Processed reads were assembled using metaSPAdes v3.10.1 [48, 49]. Acquired assemblies were searched by blastn with rRNA sequences of *Diplonema papillatum* (KF633466-8), and high scoring hits with high coverage (>5x higher than genome average) were kept. We have found that newly performed metaSPAdes assemblies contained rDNA operons of better quality, and therefore they were used for further analyses. Assembly graphs were manually inspected in Bandage [50] to identify potential misassemblies. In such a case, contigs containing rDNA were manually corrected and replaced in the assemblies. Furthermore, acquired operons were manually checked for mismatches since metaSPAdes does not support mismatch correction [49].

To provide phylogenetic background, we searched genomes of kinetoplastids and euglenids for rDNA operons by BLASTn. The rDNA 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 [51, 52]. Raw reads were accessed (last time on 05/05/21) for *Euglena viridis* (SRR14099996) [53], *Rhabdomonas costata* [54], *Bodo saltans* (ERR036178) [55], *Papus ankaliazontas* (SRR13394431), *Ankaliazontas spiralis* and *Procrystobia sorokini* (cocultured, SRR13394430) [56]. Those were processed in the same way as diplomonad assemblies. Lastly, the rDNA sequence of *Naegleria gruberi* (AB298288.1) was accessed as a non-euglenozoan outgroup.

Sequences of rRNA operons were aligned using MAFFT einsl [57]. Obtained alignment was further manually edited in Geneious v10.2.2 [58], based upon annotated secondary structures. The secondary structure of *E. gracilis* rRNA has been predicted [15, 16], while for several trypanosomatid ribosomes, cryo-electron microscopy structures have been obtained [42, 59, 60]. RNApdbee 2.0 web-server [61] was used to extract secondary structures from available cryo-electron microscopy models. Secondary structures of *E. gracilis*, *T. cruzi* and *L. major* have been annotated following previously published structures [62]. Determined helices were marked upon the alignment. Using this profile, it was possible to predict, describe and mark the secondary structure of all other species. This step was done manually and confirmed by the RNAfold WebServer [63, 64].

In parallel, homologous helices were manually aligned to prepare structure-based alignment, which was used to identify irregularities in the lengths of analysed structures. The most common elongations within rRNA genes result from group I intron insertions [10]. In order to check if a given expansion is an intron, its sequence was searched by RNAmmer [65] and by BLASTn against NCBI-nt. The expansions occurring in sites of known additional ITSs in *E. gracilis* and *T. cruzi* were described as corresponding ITS. Unusually large (>4x 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 [66]. The remaining 4817 positions were used for phylogenetic analyses. A maximum-likelihood tree (ML) was calculated using raxml-ng [67], with GTR+I+G4 model of substitution chosen by modeltest-ng [68]. The best tree was estimated using 20 different starting trees and 1000 bootstraps. Bayesian inference was performed in MrBayes v3.2.6 [69]. 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.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

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>.

Competing interests

The authors declare that they have no competing interests.

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.

Author Contributions

P.H. and R.M. designed the study. P.H. carried out the analyses. P.H., A.K. and R.M. interpreted the data. P.H. prepared figures, tables and wrote the manuscript. P.H., A.K. and R.M. inspected the manuscript critically and took part in the revision of manuscript. All authors read and approved the final manuscript.

Acknowledgments

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

References

1. 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.
2. Flegontov P, Butenko A, Firsov S, Kraeva N, Eliáš M, Field MC, et al. Supplementary Information: Genome of *Leptomonas pyrrhocoris*: a high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Sci Rep.* 2016;6:23704. doi:10.1038/srep23704.
3. 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 diplomonads, kinetoplastids, ciliates and foraminiferans. *Commun Biol.* 2021;4:1–10. doi:10.1038/s42003-021-02012-5.
4. de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, et al. Eukaryotic plankton diversity in the sunlit ocean. *Science* (80-). 2015;348. doi:10.1126/science.1261605.
5. Gawryluk RMR, del Campo J, Okamoto N, Strasser JFH, Lukeš J, Richards TA, et al. Morphological Identification and Single-Cell Genomics of Marine Diplomonads. *Curr Biol.* 2016;26:3053–9. doi:10.1016/j.cub.2016.09.013.
6. Okamoto N, Gawryluk RMR, del Campo J, Strasser 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.
7. 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.
8. 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.
9. Wideman JG, Lax G, Leonard G, Milner DS, Rodríguez-Martínez R, Simpson AGB, et al. A single-cell genome reveals diplomonadlike ancestry of kinetoplastid mitochondrial gene structure. *Philos Trans R Soc B Biol Sci.* 2019;374.

10. 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.
11. 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.
12. 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.
13. Hernández R, Díaz-de León F, Castañeda M. Molecular cloning and partial characterization of ribosomal RNA genes from *Trypanosoma cruzi*. *Mol Biochem Parasitol.* 1988;27:275–80. doi:[https://doi.org/10.1016/0166-6851\(88\)90047-3](https://doi.org/10.1016/0166-6851(88)90047-3).
14. 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).
15. Schnare MN, Gray MW. Sixteen discrete RNA components in the cytoplasmic ribosome of *Euglena gracilis*. *J Mol Biol.* 1990;215:73–83. doi:10.1016/S0022-2836(05)80096-8.
16. 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. doi:[https://doi.org/10.1016/0167-4781\(95\)00204-9](https://doi.org/10.1016/0167-4781(95)00204-9).
17. Greenwood SJ, W. 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. doi:[https://doi.org/10.1016/S0167-4781\(98\)00201-2](https://doi.org/10.1016/S0167-4781(98)00201-2).
18. 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;2015.
19. 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.
20. 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.
21. Hassouna N, Mithot B, Bachellerie 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:10.1093/nar/12.8.3563.
22. 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.

23. 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.
24. Cavalier-Smith T. Higher classification and phylogeny of Euglenozoa. *Eur J Protistol.* 2016;56:250–76. doi:10.1016/j.ejop.2016.09.003.
25. 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.
26. 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.
27. 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.
28. Winnebeck EC, Millar CD, Warman GR. Why does insect RNA look degraded? *J Insect Sci.* 2010;10:159.
29. 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.
30. 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.
31. Natsidis P, Schiffer PH, Salvador-Martínez 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.
32. Dame JB, McCutchan TF. Cloning and Characterization of a ribosomal RNA gene from *Plasmodium berghei*. *Mol Biochem Parasitol.* 1983;8:263–79.
33. Fang J, McCutchan TF. Malaria: Thermoregulation in a parasite's life cycle. *Nature.* 2002;418:742. doi:10.1038/418742a.
34. Waters AP, Syin C, McCutchan TF. Developmental regulation of stage-specific ribosome populations in *Plasmodium*. *Nature.* 1989;342:438–40.
35. 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.
36. 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.

37. Tedersoo L, Tooming-Klunderud A, Anslan S. PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytol.* 2018;217:1370–85. doi:10.1111/nph.14776.
38. 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.
39. 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.
40. 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. *mSphere.* 2019;4:e00475-19. doi:10.1128/msphere.00475-19.
41. 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.
42. 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. doi:10.1038/nature11872.
43. 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.
44. 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.
45. 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.
46. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data. 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
47. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20.
48. 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. doi:10.1089/cmb.2012.0021.
49. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: A new versatile metagenomic assembler. *Genome Res.* 2017;27:824–34. doi:10.1101/gr.213959.116.

50. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics*. 2015;31:3350–2.
51. Koreny L, Sobotka R, Kovarova J, Gnipova 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.
52. 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:e01498-15. doi:10.1128/mBio.01498-15.
53. 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. doi:10.1016/j.chom.2020.12.005.
54. 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. doi:10.1038/s41598-021-92174-3.
55. 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. doi:10.1016/j.cub.2015.11.055.
56. Tikhonenkov D V, 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. doi:10.1038/s41598-021-82369-z.
57. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
58. 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.
59. 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.
60. 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. doi:10.1016/j.celrep.2016.06.014.
61. 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. doi:10.1093/nar/gky314.

62. Petrov AS, Bernier CR, Gulen B, Waterbury CC, Hershkovits E, Hsiao C, et al. Secondary structures of rRNAs from all three domains of life. PLoS One. 2014;9:e88222.
63. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. Nucleic Acids Res. 2008;36 Web Server issue:70–4.
64. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6:26. doi:10.1186/1748-7188-6-26.
65. 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.
66. 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.
67. 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.
68. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol. 2020;31:291–4.
69. 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.

Figures

Figure 1

Schematic distribution of identified internal transcribed spacers in the LSU rDNA of euglenozoans.

The LSU structure has been shown for Diplonemea, Euglenida, two main clades of Kinetoplastea, and Heterolobosea as an outgroup. Additional ITSs have been numbered within euglenids (eITS) and kinetoplastids (kITS), with ITSs present in homologous positions marked (eITS-10/kITS-5 and eITS-13/kITS-7). On the left, phylogeny of Euglenozoa has been shown. For comparison, phylogeny presented

in Kostygov *et al.* (2021) has been shown on the right. In both cases additional ITSs cannot be traced to the common ancestor, indicating their independent origin.

Figure 2

Maximum likelihood tree of Euglenozoa based on 4817 nucleotide positions of the rDNA operon.

Bayesian inference resulted in the same topology. ML bootstrap (BT) and Bayesian posterior probability (pp) values are indicated at the nodes (BT/pp).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableS12010222.xlsx](#)