Typical structure of rRNA coding genes in diplonemids 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 diplonemid 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 diplonemid. 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 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 diplonemids. The rDNA sequences acquired for several euglenids and kinetoplastids were used to provide the background for the analysis. In all nine diplonemids, 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
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 Leishmenia 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 (Perkinsela 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 diplonemids is as expected [1], moreover, within diplonemids, 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 protostome 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 diplonemids show that the lack of additional ITSs in D. papillatum is, in fact, typical for the Diplonemidae. This finding is significant for elusive taxa like diplonemids, 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 diplonemids 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 diplonemids, 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 ribo-
some [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 riboso-
mal 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 frag-
mentation. Similarly, a number of unique RNA modifica-
tions 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 eugle-
nids and diplonemids.

Conclusions
We acquired novel complete rRNA operons for six kine-
toplastids, two euglenids, three classical and six marine
diplonemids. All analysed diplonemids lack additional
ITSs known from other euglenozoans. Interestingly,
while all investigated metakinetoplastids have the exact
same pattern of ITSs as trypanosomes, the early branch-
ing 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 diplone-
mids 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 mitochon-
drial or contaminant rRNA operons and potential misas-
sembles only high scoring hits (e-value < 10–5) with high
coverage (> 5 x higher than genome average) were kept.
We have found that the newly performed metaSPAdes
assemblies contained rRNA operons of better qual-
ity, and therefore they were used for further analyses.
Assembly graphs were manually inspected in Bandage
[55] to identify potential misassembles. In such a case,
contigs containing rRNA operons were manually cor-
rected 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. graci-
ilis (M12677.1, X53361.2) and Crithidia fasciiculata
(Y00055.1) were used as queries for euglenids and kine-
toplastids, respectively. Genome assemblies of Perkin-
sela sp. (LFNC0100001.1) and Phytomonas serpens
(AIHY00000001.1) were accessed from GenBank [56,
57]. Raw reads were accessed (last time on 05/05/21)
for Euglena viridis (SRR14099996) [58], Rhodomonas
costata [59], B. saltans (ERR036178) [60], Papus anka-
lazontas (SRR13394431), Ankialazontas spiralis and
Procrpytobia sorokini (cocultured, SRR13394430) [61].
These were processed in the same way as diplonemid
assemblies. Lastly, the rRNA operon sequence of Naegle-
ria gruberi (AB298288.1) was accessed as a non-eugle-
nozoan outgroup.

Sequences of rRNA operons were aligned using
MAFFT einsi [62]. The obtained alignment was further
manually edited in Geneious v10.2.2 [63], based upon
annotated secondary structures. The secondary struc-
ture of E. gracilis rRNA has been predicted [16, 17], while
for several trypanosomatid ribosomes, cryo-electron
microscopy structures have been obtained [23, 46, 64].
RNAmpdb 2.0 web-server [65] was used to extract sec-
ondary structures from available cryo-electron micros-
copy models. Secondary structures of E. gracilis, T. cru-
zi and L. major have been modelled, based on previously
published structures. Determined helices were marked
upon the alignment and numbered following a previ-
ously 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]. Sev-
eral 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 num-
bered 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 × 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 100 million generations, trees sampled every 100 generations. 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.

### References

1. Kostygov AY, Karnowska A, Votyipka J, Tashyvra 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.
2. Flegontova O, Flegontov P, Malviya S, Audic S, Wincker P, de Vargas C, et al. Extreme diversity of diplomedin eukaryotes in the ocean. Curr Biol. 2016;26:3060–5. https://doi.org/10.1016/j.cub.2016.09.031.
3. Schoene A, Hoffeld M, Hermanns K, Mahé F, de Vargas C, Nitzche F, et al. High and specific diversity of protists in the deep-sea basins dominated by diplomedin eukaryotes, kinetoplastids, ciliates and foraminifersans. Commun Biol. 2021;4:1–10. https://doi.org/10.1038/s44284-2021-00120-5.
4. de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, et al. Eukaryotic plankton diversity in the sunlit ocean. Science. 2015;348:1261605–1261605. https://doi.org/10.1126/science.1261605.
5. Gayruluk RM, del Campo J, Okamoto N, Strasert JH, Lukel J, Richards TA, et al. Morphological identification and single-cell genomics of marine diplomedin eukaryotes. Curr Biol. 2016;26:3053–9. https://doi.org/10.1016/j.cub.2016.09.013.
6. Okamoto N, Gayruluk RM, del Campo J, Strasert JH, Lukel J, Richards TA, et al. A revised taxonomy of Diplomedin eukaryotes, the Eupelagomidae n. fam. and a type species, Eupelagomina eocena n. gen. & sp. J Eukaryot Microbiol. 2019;66:519–24. https://doi.org/10.1111/jeu.12679.
7. Milanowski R, Gumińska N, Karnowska A, Ishikawa T, Zakryl 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.
8. Lukel J, Wheeler R, Jirsova D, David V, Archibald JM. Massive mitochondrial DNA content in diplomedin and kinetoplastid protists. IUBMB Life. 2018;70:1267–74. https://doi.org/10.1002/ubb.12619.
9. Záhonová K, Lax G, Sinha SD, Leonard G, Richards TA, Lukel J, et al. Single-cell genomics unveils a canonical origin of the diverse mitochondrial genomes of euglenoza. BMC Biol. 2021;19:1–14. https://doi.org/10.1186/s12915-021-01035-y.
10. Wideman JG, Luk G, Leonard G, Milner DS, Rodriguez-Martinez R, Simp-son AGB, et al. A single-cell genome reveals diplomedinlike ancestry of kinetoplastid mitochondrial gene structure. Philos Trans R Soc B Biol Sci. 2019. https://doi.org/10.1098/rstb.2019.0100.
11. Torres-Machorro AL, Hernández R, Cevallos AM, López-Villaseñor I. Riboso- mal RNA genes in eukaryotic microorganisms: Witnesses of phylogeny? FEBS Microbiol Rev. 2010;34:59–88. https://doi.org/10.1002/1574-6976.2009.00196.x.
12. Valach M, Moreira S, Kiethga GN, Burger G. Trans-splicing and RNA editing of LSU rRNA in Diplomedia mionucleus. Nucleic Acids Res. 2014;42:2660–72. https://doi.org/10.1093/nar/gkt1152.
13. Spencer DF, Collings J, 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/10.1002/1460-2075.1987.tb04859.x.
14. 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.
15. 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. https://doi.org/10.1016/S0166-6851(01)00310-3.
16. 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/10.1016/0022-2836(90)90096-8.
17. 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.
58. Nelson DR, Hazzouri KM, Lauersen KJ, Jaiswal A, Chaiboonchoe A, Mys-
tikou 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, Hidá S, Karkowska A, Milanowski R, Szabolov J, Hrdaliová M,
et al. Heterotrophic Euglenid Rhodomonas 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 associ-
ated with the origins of parasitism. Curr Biol. 2016;26:161–72. https://doi.
org/10.1016/j.cub.2015.11.055.
61. Tiškonenkov DV, Gawryluk RM, 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/ms3010.
63. Kearsie 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, Jun Ayub M, Levin MJ, Frank J. The structure of the 80S ribosome
from Trypanosoma cruzi reveals unique RNA components. Proc Natl
Acad Sci U S A. 2005;102:10206–11. https://doi.org/10.1073/pnas.05009
26102.
65. Zok T, Antczak M, Zurkowsi M, Popenda M, Blazewicz J, Adamiak RW,
et al. RNApbee 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. Ribovi-
sion 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 website. Nucleic Acids Res. 2008;36:70–4. https://doi.org/10.1093/
bt28348.
68. 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. 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/
btz348.
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/bioin
formatics/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;37:291–4. https://doi.org/10.
1101/612903.
72. Ronquist F, Teslenko M, Van Der Mark P, Ayres DI, 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.