Abstract
Mitochondria are double membrane organelles of endosymbiotic origin, best known for constituting the centre of energetics of a eukaryotic cell. They contain their own mitochondrial genome, which as a consequence of gradual reduction during evolution typically contains less than two dozens of genes. In this review, we highlight the extremely diverse architecture of mitochondrial genomes and mechanisms of gene expression between the three sister groups constituting the phylum Euglenozoa - Euglenida, Diplonemea and Kinetoplastea. The earliest diverging euglenids possess a simplified mitochondrial genome and a conventional gene expression, whereas both are highly complex in the two other groups. The expression of their mitochondrial-encoded proteins requires extensive post-transcriptional modifications guided by complex protein machineries and multiple small RNA molecules. Moreover, the least studied diplonemids, which have been recently discovered as a highly abundant component of the world ocean plankton, possess one of the most complicated mitochondrial genome organisations known to date.
Corresponding author: Julius Lukeš (jula@paru.cas.cz)

How to cite this article: Faktorová D, Dobáková E, Peña-Díaz P and Lukeš J. From simple to supercomplex: mitochondrial genomes of euglenozoan protists [version 2; referees: 2 approved] F1000Research 2016, 5(F1000 Faculty Rev):392 (doi: 10.12688/f1000research.8040.2)

Copyright: © 2016 Faktorová D et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: This work was funded by Gordon and Betty Moore Foundation grant GBMF4983 and Czech Grant Agency 15-21974S to Julius Lukeš.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare that they have no competing interests.

Introduction
The phylum Euglenozoa, which is part of the supergroup Excavata that significantly diverged from other eukaryotic lineages, is composed of three geographically ubiquitous groups of flagellated protists: Euglenida, Diplonemea, and Kinetoplastea (the fourth group, Symbiontida, has no molecular data available and thus will not be discussed here). The well-known representatives of these groups are *Euglena* for euglenids and *Trypanosoma* for kinetoplastids, whereas diplonemids are very poorly known. Although the Euglenozoa is a stable and highly supported group, mutual phylogenetic relationships among these three groups are not yet fully resolved, and euglenids likely constitute the earliest offshoot (Figure 1).

Euglenozoans have several common morphological features, such as subpellicular microtubules and a single flagellum or two heterodynamic flagella, protruding from an anterior pocket. Their lifestyles vary greatly, ranging from the free-living photosynthetic euglenids to intra- and extracellular parasites of plants, insects, and mammals, including humans. We do not yet know the predominant lifestyle of diplonemids, a group that recently came into the spotlight thanks to the Tara Oceans expedition, which revealed their global presence and extreme abundance in the world ocean. Indeed, diplonemids may comprise the sixth most abundant and third most species-rich group of marine eukaryotes. The euglenids and diplonemids display different modes of nutrition; the former are characterized by photoautotrophy, whereas the latter are likely phagotrophs, osmotrophs, or parasites or a combination of these. The predominantly parasitic kinetoplastids make use of the carbon sources provided by their hosts.

Figure 1. Schematic phylogenetic tree of representative genera of Euglenozoa depicting the organization of their mitochondrial genomes. The scheme is based on Adl et al. (2012). Different organization of their mitochondrial genomes (in blue) is shown for the three major lineages: Kinetoplastea (in red), Diplonemea (in purple), and Euglenida (in green). Whereas Euglenida possess an array of linear mitochondrial DNA molecules of variable length, Diplonemea and Kinetoplastida contain in their organelle circular DNA molecules in different arrangements. In Diplonemea, circular molecules of two sizes are non-catenated and supercoiled. The kinetoplast DNA of Eubodonida, Parabodonida, and Neobodonida is composed of numerous free, non-catenated relaxed or supercoiled DNA circles, whereas in Trypanosomatida it is constituted of thousands of relaxed circles, mutually interlocked into a single giant network composed of interlocked maxicircles and minicircles that together with proteins are packaged into a single compact disk.
A hallmark feature of euglenozoans is a single large mitochondrion, frequently reticulated and displaying cristae with a discoid structure. Like all mitochondria of aerobic protists, this organelle contains mitochondrial DNA (mtDNA). Although as vestigial as mtDNAs of other eukaryotes, this organelar genome evolved in euglenozoan protists into a stunning variety of structures and organizations, as described in more detail below. With the advent of affordable high-throughput sequencing, thousands of mt genomes are being assembled and annotated. However, their selection remains strongly skewed toward the metazoans, which mostly harbor standard, highly reduced, and streamlined mt genomes. Yet the majority of extant eukaryotic diversity is constituted by protists, of which only a very small fraction has their mtDNA characterized. Still, the available mt genomes of protists show a range of bizarre gene arrangements, modes of organization, and complex post-transcriptional maturations.

Mitochondrial genome architecture and gene content

Standard mt genomes are usually represented by a circular or linear DNA molecule encoding an average of fewer than two dozen genes ranging from 2 to 66 proteins in Chromera velia and Andalucia godoyi, respectively. Although euglenozoans harbor a low number of genes in their mt genomes, they developed an extremely variable genome architecture. In euglenids and diplomids, mtDNA seems to be evenly distributed throughout the lumen of the organelle, whereas in kinetoplastids, the picture is more complex. In the obligatory parasitic trypanosomatids mtDNA is invariably compacted into a single disk-shaped structure of concatenated DNA termed the kinetoplast DNA (kDNA), the free-living or commensalic bodonids have their kDNA distributed either evenly or in foci in the mt lumen.

The best-studied kDNA is that of the human pathogen and model organism Trypanosoma brucei (for current reviews, see 17–19). It is composed of thousands of DNA circles mutually interlocked into a single network that is densely packed into a disk-shaped structure located close to the basal body of the flagellum. The kDNA network of T. brucei and related flagellates is formed of dozens of maxicircles, each about 20 kb long, and of approximately 5,000 uniformly sized (~1.0 kb) minicircles. The maxicircle is composed of a single conserved region, which contains all protein-coding and rRNA genes, and a shorter variable region of species-specific size and sequence, which probably plays a role in replication. The conserved region carries 18 protein-coding genes, mostly subunits of respiratory complexes (complex I: nad1, nad2, nad3, nad4, nad5, nad7, nad8, and nad9; complex III: cob; complex IV: cox1, cox2, and cox3; complex V: atp6), one ribosomal protein (rps12), small and large mito-rRNA genes (12S and 9S), and four open reading frames of unknown function (MURF2, MURF5, cr3, and cr4) (Figure 3). Each minicircle codes for

Figure 2. Morphology of representatives of Kinetoplastea (Trypanosoma brucei), Diplonemea (Diplonema papillatum), and Euglenida (Euglena gracilis). Scanning electron microscopy (SEM) (a, d, g) and differential interference contrast (DIC) (b, e, h) reveal cell morphology, whereas 4′,6-diamidino-2-phenylindole (DAPI) staining provides information about the amount and distribution of mitochondrial DNA (c, f, i). (c) Trypanosoma with distinct nucleus (N) and kinetoplast (K). (f) In Diplonema, arrows point to large amounts of mitochondrial DNA meandering through the cell. Scale bars = 1 μm (a, d) and 10 μm (g).
The mtDNA of *Euglena*, the model euglenid, is surprisingly simple when compared with the highly complex genomes of its sister groups. Indeed, the genome is streamlined in terms of both its architecture and gene content. The mtDNA of *Euglena* is represented by a pool of recombination-prone short linear molecules containing repeats, pieces of non-functional gene fragments, and full-length gene copies, which ensure the production of functional proteins. The set of only three protein-coding genes (coxl-3)\(^{36,37}\) was recently complemented by four intact genes encoding additional subunits of the respiratory chain (nad1, 4, and 5 and cob)\(^{18}\). The genes encoding subunits of complex V and ribosomal proteins are missing. The SSU and LSU of mito-rRNA are likely split into two fragments each (SSU-R, SSU-L and LSU-R, LSU-L). However, only the SSU-R fragment has been identified to date and this is most likely due to the same reasons as in the case of *Diplonema* LSU, namely the extreme divergence of the corresponding mito-rRNAs\(^{9,38}\).

### Mechanisms of mitochondrial gene expression

In all euglenozoans, mtDNA is transcribed into polycistrons, which undergo endonucleolytic cleavage and further editing or processing (or both) into translatable mRNAs (Figure 4). In kinetoplastids, the majority of mt-encoded genes exist in a cryptic form, as their corresponding transcripts have to undergo extensive post-transcriptional RNA editing of the uridine insertion/deletion type (for recent reviews, see 19,27,28). In *Diplonema*, only a few insertions of blocks of uridines have been documented initially, but the recent comprehensive count amounts to ~200\(^{11,12,33}\). RNA editing in *Trypanosoma* and related flagellates is extremely complex, as the insertions and deletions of hundreds and dozens of uridines, respectively, are performed by a multitude of gRNAs and several protein complexes that interact in a highly dynamic manner\(^{28,39}\). Upon the addition of complex poly-U/A tails\(^{37,28}\) (Figure 4), the fully edited transcripts are translated on protein-rich and RNA-poor ribosomes\(^{40}\), but the role of the additional 45S ribosomal subunit of unique protein composition is still unclear\(^{41}\). Likely owing to their extreme hydrophobicity, only a few of the *de novo* synthesized mt proteins have been observed\(^{42,43}\).

The post-transcriptional processing is very different in *Diplonema*, where fragments of genes transcribed from individual DNA circles are trans-spliced\(^{31,33}\). The genome and mitoproteome of *Diplonema* may eventually shed light on this unique processing, but so far the proteins (and potentially also small RNAs) involved in it remain completely unknown. In any case, it is highly likely that the fully trans-spliced mRNAs are *in organello* translated (our unpublished data). Indeed, the recently described simplicity of mt mRNA processing in *Euglena*\(^7\) and likely other euglenids is in stunning contrast with the baroque complexity of RNA editing or trans-splicing (or both) in kinetoplastids and diplomonds (Figure 4 and Table 1).

### Why are mitochondrial genomes in Euglenozoa so diverse?

Soon after its discovery, RNA editing in kinetoplastids was explained as a remnant of the RNA world\(^{44}\). A more plausible explanation postulates that gene fragmentation in the euglenozoan last common ancestor was “the seed of future chaos”, leading
Figure 4. The comparison of RNA processing of the cox1 transcript in representative euglenozoans. In contrast to *Trypanoplasma borreli* and *Diplonema papillatum*, the cytochrome *c* oxidase subunit 1 (cox1) transcript of *Euglena gracilis* (in green) does not undergo RNA editing, splicing, or polyadenylation prior to its translation. In *Trypanoplasma*, cox1 undergoes RNA editing in the form of numerous uridine insertions (small blue “u”) and deletions (blue star), followed by polyadenylation (in red). The *Diplonema* cox1 transcript is formed by trans-splicing of nine small fragments called modules M1 thru M9 (in green), which is accompanied by the insertion of six uridines between the modules M4 and M5. Finally, the transcript is polyadenylated and translated on mitochondrial ribosomes.

**Table 1. Architecture and gene content of mitochondrial genomes of representative euglenozoans and humans.**

<table>
<thead>
<tr>
<th></th>
<th>Trypanosoma brucei</th>
<th>Diplonema papillatum</th>
<th>Euglena gracilis</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of mitochondrial cristae</td>
<td>Discoidal</td>
<td>Discoidal</td>
<td>Discoidal</td>
<td>Tubular</td>
</tr>
<tr>
<td>Mode of life</td>
<td>Parasitic</td>
<td>Phagotrophic osmotrophic</td>
<td>Mixotrophic heterotrophic</td>
<td>Heterotrophic</td>
</tr>
<tr>
<td>Habitat</td>
<td>Insect gut mammalian bloodstream</td>
<td>Marine</td>
<td>Freshwater</td>
<td>Predominantly terrestrial</td>
</tr>
<tr>
<td>Genome structure</td>
<td>Circular</td>
<td>Circular</td>
<td>Linear</td>
<td>Circular</td>
</tr>
<tr>
<td>Protein-coding genes</td>
<td>18</td>
<td>10</td>
<td>7*</td>
<td>13</td>
</tr>
<tr>
<td>rRNA (SSU/LSU)</td>
<td>+/+</td>
<td>+/+</td>
<td>+/?</td>
<td>+/+</td>
</tr>
<tr>
<td>tRNA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>22</td>
</tr>
<tr>
<td>Chromosome size</td>
<td>Maxicircles: ~20.0 kb Minicircles: ~1.0 kb</td>
<td>Class A: 6.0 kb Class B: 7.0 kb</td>
<td>~1.0 to ~9.0 kb</td>
<td>16.6 kb</td>
</tr>
<tr>
<td>Genome copy number</td>
<td>Maxicircles: ~dozens Minicircles: ~5,000</td>
<td>~100</td>
<td>?</td>
<td>~50 to 1×10^5</td>
</tr>
<tr>
<td>mRNA polyadenylation</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Trans-splicing</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Uridine insertions</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Uridine deletions</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Introns</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

*Seven complete genes, together with multiple gene fragments.*
to the emergence of extremely complex post-transcriptional mechanisms, differing in each sister lineage, yet eventually correcting the scrambling on the post-transcriptional level\(^6\). In fact, it was argued that the kinetoplastid RNA editing is a prime example of “irremediable complexity”, a rampant mechanism that does not provide any selective advantage yet fixes the problem\(^6\). The recent finding of a mt genome in *Euglena*\(^6\) implies that the irreversible scrambling originally implied for the mtDNA of the euglenozoan last common ancestor\(^6\) did happen at a later stage in evolution, probably in the predecessor of diplomonads and kinetoplastids. Although despite the available sequence data the mutual relationships among the three euglenozoan lineages remain unresolved, we can predict, on the basis of their mt genomes and transcriptomes, that the mostly free-living photosynthetic euglenids constitute the earliest offshoot of the long euglenozoan branch.

### Competing interests

The authors declare that they have no competing interests.

### Grant information

This work was funded by Gordon and Betty Moore Foundation grant GBMF4983 and Czech Grant Agency 15-21974S to JL.

### Acknowledgments

The scanning electron micrographs of *Diplonema* and *Trypanosoma* were kindly provided by Binnypreet Kaur and Eva Horákův (Biology Centre, České Budějovice, respectively.

### References


Open Peer Review

Current Referee Status: ✔️ ✔️

Editorial Note on the Review Process
F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty and are edited as a service to readers. In order to make these reviews as comprehensive and accessible as possible, the referees provide input before publication and only the final, revised version is published. The referees who approved the final version are listed with their names and affiliations but without their reports on earlier versions (any comments will already have been addressed in the published version).

The referees who approved this article are:

Version 2

1 Daniel Sloan, Colorado State University, Fort Collins, CO, USA
   Competing Interests: No competing interests were disclosed.

2 Thomas Becker, University of Freiburg, Freiburg, Germany
   Competing Interests: No competing interests were disclosed.

Version 1

1 Daniel Sloan, Colorado State University, Fort Collins, CO, USA
   Competing Interests: No competing interests were disclosed.

2 Thomas Becker, University of Freiburg, Freiburg, Germany
   Competing Interests: No competing interests were disclosed.