Molecular Life Sciences
DOI 10.1007/978-1-4614-6436-5_118-2


# Mitochondrial Genomes of Excavata

Julius Lukeš*
Biology Centre, Institute of Parasitology, Czech Academy of Sciences, and Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

## Synopsis

Supergroup Excavata comprises a variety of biflagellated aerobic and anaerobic protists whose defining characteristic is a central grooved cytostome (cell mouth). Of all the eukaryotic supergroups, excavates exhibit the widest variety of mitochondrial genome forms and gene content. One group, the jakobids, possesses the most ancestral (least derived) mitochondrial genome yet characterized: a circular-mapping DNA containing the largest known mitochondrial gene set and displaying bacterial operon-like gene arrangements and expression signals. At the other extreme, certain excavates have lost mtDNA entirely. One excavate phylum, Euglenozoa, contains three well-delineated lineages: kinetoplastids, diplonemids, and euglenids. Some kinetoplastids, including many parasitic genera such as *Trypanosoma* (the causative agent of African sleeping sickness), have a complex kinetoplast DNA that consists of interlocked mass of maxicircles and minicircles. Maxicircles correspond to the mitochondrial genome in other organisms, but their protein-coding genes are "cryptic" in that their transcripts must undergo an intricate process of posttranscriptional uridine insertion/deletion editing in order to become translatable. Minicircles encode small guide RNAs that provide the information in *trans* for this RNA editing. In *Diplonema papillatum* (a diplonemid), the mtDNA consists of numerous small circular molecules encoding only portions of genes, whose transcripts must be correctly *trans*-spliced in order to be translated. In contrast, in *Euglena gracilis* (a euglenid), the mtDNA takes the form of small linear fragments, and the few identified genes are embedded in a sea of noncoding repeated sequence and interspersed with small fragments of authentic genes.

## Introduction

Excavata is a taxonomic unit usually considered to be at the rank of a superkingdom, comprising aerobic and anaerobic protists having a central cytostome in the form of a groove. Major phyla with representative genera include Euglenozoa (*Euglena*, *Trypanosoma*), Heterolobosea (*Naegleria*), Jakobida (*Jakoba*, *Reclinomonas*), Preaxostyla (*Trimastix*), Fornicata (*Giardia*), and Parabasalia (*Trichomonas*). In most cases, one flagellum is positioned to bring food into the groove, while the second, forward-directed flagellum propels the organism. However, numerous members of this group have lost some or all of these features and are placed in this superkingdom on the basis of molecular phylogenetic evidence only (Hampl et al. 2009).

Excavata is of particular interest because of its possibly basal position in the eukaryotic phylogenetic tree. Because light and electron microscopic studies have failed to find typical cristae-containing, double-membrane-bound vesicles in a number of excavates, these were at one time considered to be amitochondriate protists that had branched off the eukaryotic tree before the

*Email: jula@paru.cas.cz

DOI 10.1007/978-1-4614-6436-5_118-2


endosymbiotic event that led to mitochondria. However, with the advent of molecular methods, which have been instrumental in the identification of mitochondrial or mitochondrion-derived prokaryotic genes in all carefully examined excavates, this view has been abandoned.

At present, organelles containing the products of "mitochondrial" genes are classified into three morphologically and functionally distinct categories: mitochondrion, hydrogenosome, and mitosome. It is worth noting, however, that the distinctions among these three categories are not as clearly defined as was earlier thought. Recent studies indicate the existence of a continuum of organelles, in which a conventional mitochondrion represents one extreme and a highly reduced mitosome the other. Therefore, given the current paucity of information, the conservative approach is to refer to the cristae-deficient vesicles of most excavates as "mitochondrion-related organelles" (MROs).

## Mitochondrial Genome Diversity in Excavates

In Excavata, only the ultrastructurally typical mitochondria, usually containing respiratory complexes, are known to contain mtDNA. Interestingly, in terms of amount of DNA, number of (protein-coding) genes, and overall organization, the mitochondrial genome is extremely variable in these protists (Gray et al. 2004). In fact one can argue that excavates exhibit higher mitochondrial genome diversity than the rest of eukaryotes combined: one group, the jakobids, harbors extremely gene-rich mitochondrial genomes, whereas the "petite" mutants of trypanosomes lack any organellar DNA whatsoever. Apparently, the progressive transfer of genes from the endosymbiont-derived mitochondrion to the cell nucleus occurred at dramatically different rates in the various lineages of Excavata.

Sequencing of the entire circular-mapping mitochondrial genome of one excavate, the jakobid *Reclinomonas americana*, was a true breakthrough. The total number of 98 mitochondrial genes, of which 67 are protein coding, greatly surpassed the highest number of mitochondrion-encoded genes known up to that time (Lang et al. 1997) and still represents the largest known gene repertoire for any mitochondrial genome. While most of the genes in *R. americana* mtDNA code for subunits of respiratory complexes, 18 of them including 4 subunits of a eubacterial-type RNA polymerase were found to be associated with a mitochondrial genome for the first time (Lang et al. 1997). Since all other eukaryotes encode a single-subunit, phage T3/T7-like mitochondrial RNA polymerase in their nucleus, the most plausible explanation of these findings is that the nucleus-encoded, phage-type RNA polymerase replaced the original mitochondrion-encoded, eubacterial enzyme, retained only in *R. americana*, in all other eukaryotes.

Subsequently, mitochondrial genomes of a number of other jakobids have been sequenced, revealing that indeed all of them are extremely gene- as well as A+T-rich, with only relatively small differences among the encoded gene sets (Gray et al. 2004). The mitochondrial genome of *Jakoba libera* ($\sim$100 kb in size) is notable for its linearity and the telomere-like repeats that cap both ends. Finally, RNA editing has been noted in two tRNAs encoded by the mitochondrial genome of another jakobid, *Seculamonas ecuadoriensis*.

While the MRO (hydrogenosome) of the anaerobic ciliate *Nyctotherus* possesses a genome (Akhmatova et al. 1998), no DNA is present in the homologous organelle of the excavate *Trichomonas* and related protists. The well-studied MROs of the latter organism were originally characterized as molecular-hydrogen-producing vesicles but were recently shown also to be site of iron-sulfur cluster synthesis and perhaps other processes (Tachezy 2008). MROs (mitosomes) are found in another excavate, the anaerobic intestinal parasite *Giardia* spp., but similar organelles evolved

DOI 10.1007/978-1-4614-6436-5_118-2


independently in other eukaryotic lineages as well. Although early studies claimed that the mitosomes of the non-excavate *Entamoeba histolytica* contain DNA, the present view is that no DNA has been retained in any of the mitosomes studied so far. MROs have been observed in ultrastructural sections of a number of free-living excavates (e.g., *Trimastix*, *Carpediemonas*, *Retortamonas*, *Dysnectes*, and *Chilomastix*), as well as their parasitic or commensalic relatives (e.g., diplomonads, retortamonads, oxymonads, and parabasalids). Biochemical properties of these organelles remain largely unknown, but in all cases where extensive expressed sequence tag projects were launched, homologues of genes coding for mitochondrial proteins have been found, testifying to an origin of these genome-lacking organelles from an ancestral, genome-containing mitochondrion (Tachezy 2008).

## Mitochondrial DNA in Kinetoplastids

Kinetoplastid flagellates (phylum Euglenozoa) harbor a large and complex mass of DNA, termed kinetoplast (k) DNA, composed of so-called maxicircles and minicircles. Maxicircles are the counterpart of a typical mitochondrial genome (Jensen and Englund 2012), as they encode mitoribosomal RNAs as well as protein subunits of respiratory complexes and mitochondrial ribosomes; however, the function of minicircles was established only after the discovery of RNA editing (Stuart et al. 2005) (see below).

Kinetoplast DNA exists principally in two different arrangements – as free DNA circles in the suborder Bodonidae and as a single, three-dimensional kDNA network in the suborder Trypanosomatidae (Lukeš et al. 2002). Bodonidae is a paraphyletic group ancestral to Trypanosomatidae and includes parasitic, commensal, and free-living protists with two flagella. The kDNA in these groups comes in several forms: (1) pro-kDNA, found in *Bodo* spp., is characterized by non-interlocked relaxed DNA circles, concentrated in a single bundle located near the basal body of the flagella; (2) pan-kDNA, described in members of the genus *Cryptobia*, is composed of tens of thousands of minicircles that are either free or form small catenanes, are invariably supercoiled, and are evenly distributed throughout the mitochondrial lumen; (3) poly-kDNA arrangement is characteristic of *Dimastigella* spp. and *Cruzella marina*, the minicircles of which are non-interlocked and relaxed and distributed throughout the organellar lumen at multiple foci; (4) finally, in *Trypanoplasma borreli*, kDNA exists in a unique arrangement, termed mega-kDNA, in which minicircles are tandemly linked into large circular molecules, more or less evenly spread throughout the mitochondrial lumen. In most bodonids, the kDNA is truly gigantic, in size approaching that of the respective nuclear genomes (Lukeš et al. 2002). It is likely that yet another type of kDNA organization will be described in *Perkinsella* (or *Ichthyobodo*-like organism), an early-branching aflagellar kinetoplastid that became an endosymbiont of another protist, *Neoparamoeba*.

With regard to kDNA structure, the situation is simpler in the other group within Kinetoplastida, Trypanosomatidae, which encompasses solely parasitic forms. Some, such as *Trypanosoma* spp., *Leishmania* spp., and *Phytomonas* spp., are serious pathogens of humans, other vertebrates, and economically important plants, as well as monoxenous parasites of insects. In *Trypanosoma brucei*, the most intensively studied kinetoplastid and causative agent of African sleeping sickness, and in the model flagellate, *Crithidia fasciculata*, replication and maintenance of the kDNA network has been investigated in great detail. The maxicircles and minicircles are mutually interlocked into a single huge network within the mitochondrion, densely packed into a disk located close to the basal body of the flagellum (Jensen and Englund 2012). The diameter and the thickness of the disk are

DOI 10.1007/978-1-4614-6436-5_118-2


correlated with the number and size of minicircles, respectively. A unique feature of the kDNA circles is that they are not supercoiled but relaxed. Their replication occurs via highly organized enzymatic machinery (in conjunction with which individual circles are released from the network), replicated in a specialized zone and reattached to the network at two antipodal sites. So far, six DNA polymerases and six DNA helicases have been identified as part of the replication machinery, which is estimated to comprise more than 100 different proteins, including topoisomerases, primases, ligases, and single- and double-strand-binding proteins. Somewhat surprisingly, only a single RNA polymerase is responsible for the transcription of both maxicircles and minicircles (Jensen and Englund 2012).

It has been known for decades that some *Trypanosoma* species contain less kDNA than others. Some have lost parts of or the entire maxicircle component but have retained minicircles (=dyskinetoplastic trypanosomes), whereas others have lost all kDNA (=akinetoplastic trypanosomes). Only recently was it shown that these forms are to various extents "petite" mutants of *T. brucei* and that the gradual loss of kDNA can also be induced under laboratory conditions (Lai et al. 2008).

As mentioned above, most maxicircle-encoded genes exist in an encrypted form, which means that their protein-coding transcripts have to be extensively edited via a complex process of uridine (U) insertion and deletion in order to become translatable. RNA editing occurs in all studied kinetoplastids; however, its extent varies depending on species and gene. Although RNA editing has been extensively studied only in *T. brucei* and *Leishmania tarentolae*, available data indicate that the underlying mechanism is highly conserved within Kinetoplastida. Information for the posttranscriptional, site-specific U insertions into and U deletions from pre-edited mRNAs is provided by small (50–70-bp-long) RNA molecules termed guide RNAs, which are encoded predominantly in the kDNA minicircles (Stuart et al. 2005). Moreover, several large protein complexes assist in this intricate task. In *T. brucei*, it is estimated that up to 1,000 different guide RNAs and over 70 proteins are involved in the editing and processing of 18 maxicircle-encoded transcripts. The best-studied complex is the editosome, which encompasses the main enzymatic activities for editing events, in particular (1) cleavage of pre-edited mRNA at the editing site, specified by guide RNA; (2) U insertion or deletion by a terminal uridylyl transferase or exonuclease, respectively; and (3) sealing by RNA ligase (Aphasizhev and Aphasizheva 2011). The MRP1/2 complex is thought to facilitate the formation of duplexes between guide RNAs and pre-edited mRNAs (Stuart et al. 2005). A mitochondrial poly-A polymerase (kPAP) complex modulates, in collaboration with additional proteins, the synthesis of 3′ poly-A/U tails in pre-edited, edited, and never-edited transcripts (Aphasizhev and Aphasizheva 2011). Finally, a recently described mitochondrial RNA binding complex (MRP1) is composed of dozens of proteins that usually lack any domains and are conserved only within the kinetoplastid flagellates. The composition of the MRP1 complex is uncertain, as different purification protocols lead to different sets of protein subunits. Moreover, knockdowns for individual subunits result in dramatically different phenotypes, ranging from general destabilization of guide RNAs through impact on all mRNAs to a limited effect on partially or extensively edited transcripts (Aphasizhev and Aphasizheva 2011). Evidently additional proteins awaiting discovery are involved in the byzantine and extremely complex editing and processing of mitochondrial transcripts in trypanosomes. Although the evolutionary advantage of this U insertion/deletion type of editing is unclear, the edited transcripts are translated in the organelle and the resulting, functional proteins are incorporated into mitochondrial respiratory complexes and ribosomes, as in conventional mitochondria in which no such editing process exists.

DOI 10.1007/978-1-4614-6436-5_118-2


## Mitochondrial DNA in Euglenids and Diplonemids

In addition to Kinetoplastida, phylum Euglenozoa contains two other well-defined monophyletic groups – Euglenida and Diplonemida. Euglenids are free-living, ecologically significant protists, carrying a secondarily acquired green plastid, which was lost in some lineages. Diplonemids are a small, poorly studied group of parasites and commensals that branches off between kinetoplastids and euglenids and has recently been considered to be more closely related to the former.

Despite the fact that the plastid genome of *Euglena gracilis* was the first completely sequenced plastid DNA, the mitochondrial genome of this protist, as well as of related species, has proven very refractory to study, and our knowledge about it remains quite limited. Since the description of the mtDNA-encoded gene specifying cytochrome *c* oxidase subunit 1 (*cox1*) in the mid-1990s (Refs. 294 and 320 in Gray et al. 2004), only one study has been published further characterizing *E. gracilis* mtDNA (Spencer and Gray 2011).

The mitochondrial genome of *E. gracilis* is represented by a collection of heterodisperse linear DNA fragments, most of which are about 4 kb long, although small fragments centered around 7.5 kb were also seen. This observation is in agreement with early studies of the physical structure of *E. gracilis* mtDNA, in which only small DNA pieces were detected. In *E. gracilis* mtDNA, genes encoding all three subunits of the cytochrome *c* oxidase complex (*cox1*, *cox2*, and *cox3*) have been found, in most cases flanked by repeat sequences of varying size. However, full-size open reading frames were encountered rather rarely, whereas DNA fragments containing small pieces of the *cox* genes (as well as rRNA genes), in various arrangements, were abundant. Transcripts of protein-coding genes appear not to require RNA editing or RNA splicing. The small and large mitoribosomal rRNAs are present in the form of two fragments each, yet their transcripts do not appear to be spliced together. Each half of the small subunit rRNA is encoded by a separately transcribed subgenomic module.

Although in all other aspects of cellular and molecular biology, diplonemids are the less well-known organisms when compared with euglenids, a substantial amount of information is available about diplonemid mtDNA. Described by (Vlček et al. 2011) as "[a]rguably the most bizarre mitochondrial DNA," the mitochondrial genome of *Diplonema papillatum* is located in a single large reticulated organelle, for which a low number of exceptionally large and flat mitochondrial cristae is characteristic (Marande et al. 2005). The organellar DNA is not concentrated in a single or a few locations, as in the related kinetoplastids, but is distributed in a seemingly random fashion throughout the lumen of the mitochondrion. Staining with DNA-specific dyes strongly indicates that the amount of mtDNA is extremely high and may even approach that of the nuclear genome. Electron microscopy and agarose gel electrophoresis of the mtDNA, separated from nuclear DNA, revealed that it comprises two categories of circular chromosomes, 6 and 7 kb in size (designated A and B, respectively), that exist as relaxed and supercoiled circles, each present in multiple copies in free form (Marande et al. 2005).

While the sequence data obtained in conjunction with this initial study already revealed that at least the *cox1* gene is split into several fragments, each residing on a different circular chromosome (Marande et al. 2005), it was not obvious at that point whether these modules were simply nonfunctional gene fragments present in addition to a full-size functional copy of this gene or whether no intact version of *cox1* existed. A follow-up study demonstrated that the *cox1* fragments are indeed transcribed individually from different circular chromosomes, with the contiguous, translatable *cox1* transcript being generated by splicing together, in *trans* and in a perfectly orderly fashion, the separate transcripts of nine *cox1* subgenomic modules located in the mtDNA (Marande and Burger 2007). Moreover, this work revealed a single case of insertion of a stretch of six uridines

DOI 10.1007/978-1-4614-6436-5_118-2


between two modules (Marande and Burger 2007), reminiscent of the uridine insertion/deletion type of RNA editing characteristic of kinetoplastids (see previous chapter). While nothing is yet known about the machinery responsible for this insertion, it has been shown that in terms of size and location, exactly the same insertion of six non-encoded uridines occurs in the related diplonemids *D. ambulator*, *Diplonema* sp. 2, and *Rhynchopus euleeides*.

Extensive sequencing of the mitochondrial genome of *D. papillatum* revealed that it contains a typical set of protein-coding genes, which are, however, all split into fragments in the manner originally described for *cox1* (Vlček et al. 2011). So far, a complete set of modules for cytochrome *b*, *cox1*, *cox2*, *cox3*, and *nad7* (encoding NADH dehydrogenase subunit 7 of respiratory complex I), as well as dozens of modules varying in length from 60 to 350 bp and encoding other subunits of the respiratory chain, have been mapped. Some genes, often present and conserved in mitochondrial genomes, such as *rps12* (ribosomal protein subunit 12) and *nad9* (NADH dehydrogenase subunit 9), have so far not been encountered in the *D. papillatum* sequences, while the absence of tRNA genes in diplonemids (Vlček et al. 2011) is a character shared with kinetoplastid mitochondrial genomes. Furthermore, as in the case of *cox1* transcripts, the gene modules encoding other mitochondrial proteins are transcribed and the products *trans*-spliced into mature transcripts. The position of the gene modules in the A and B circles is highly conserved, with only one module invariably present per circle. Using comparative sequence analysis, a constant region in each circle has been identified, in addition to the gene-encoding portion flanked by conserved motifs (Vlček et al. 2011).

The existence of small mtDNA fragments in diplonemids and euglenids indicates that the common ancestor of Euglenozoa likely already had a fragmented mitochondrial genome, which might have had important consequences for extant mitochondrial genomes within this lineage (Flegontov et al. 2011). As proposed recently by (Spencer and Gray 2011), antisense transcripts of gene fragments might exemplify the ancestral form of guide RNAs, genes for which are abundantly present in the kinetoplast DNA.

## References

Akhmatova A, Voncken F, Van Alen T, Van Hoek A, Boxma B, Vogels G, Veenhuis M, Hackstein JHP (1998) A hydrogenosome with a genome. Nature 396:527–528

Aphasizhev R, Aphasizheva I (2011) Mitochondrial RNA processing in trypanosomes. Res Microbiol 162:655–663

Flegontov P, Gray MW, Burger G, Lukeš J (2011) Gene fragmentation: a key to mitochondrial genome evolution in Euglenozoa? Curr Genet 57:225–232

Gray MW, Lang BF, Burger G (2004) Mitochondria in protists. Annu Rev Genet 38:477–524

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AG (2009) Phylogenetic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proc Natl Acad Sci U S A 106:3859–3864

Jensen RE, Englund PT (2012) Network News: The replication of kinetoplast DNA. Ann Rev Microbiol 66:473–491

Lai D-H, Hashimi H, Lun Z-R, Ayala FJ, Lukeš J (2008) Adaptation of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *T. equiperdum* and *T. evansi* are petite mutants of *T. brucei*. Proc Natl Acad Sci U S A 105:1999–2004

Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 387:493–497

DOI 10.1007/978-1-4614-6436-5_118-2


Lukeš J, Guilbride DL, Votýpka J, Zíková A, Benne R, Englund PT (2002) The kinetoplast DNA network: evolution of an improbable structure. Eukaryot Cell 1:495–502

Marande W, Burger G (2007) Mitochondrial DNA as a genomic jigsaw puzzle. Science 318:415

Marande W, Lukeš J, Burger G (2005) Unique mitochondrial genome structure in diplonemids, the sister group of kinetoplastids. Eukaryot Cell 4:1137–1146

Spencer DF, Gray MW (2011) Ribosomal RNA genes in *Euglena gracilis* mitochondrial DNA: fragmented genes in a seemingly fragmented genome. Mol Genet Genomics 285:19–31

Stuart K, Schnaufer A, Ernst NL, Panigrahi AK (2005) Complex management: RNA editing in trypanosomes. Trends Biochem Sci 30:97–105

Tachezy J (ed) (2008) Hydrogenosomes and mitosomes: mitochondria of anaerobic eukaryotes. Springer-Verlag Berlin Heidelberg

Vlček C, Marande W, Teijeiro S, Lukeš J, Burger G (2011) Systematically fragmented genes in a multipartite mitochondrial genome. Nucleic Acids Res 39:979–988