Chapter 6
Mitochondrial RNA Editing and Processing in Diplonemid Protists

Drahomíra Faktorová, Matus Valach, Binnypreet Kaur, Gertraud Burger, and Julius Lukeš

Abstract RNA editing and processing in the mitochondrion of *Diplonema papillatum* and other diplonemids are arguably the most complex processes of their kind described in any organelle so far. Prior to translation, each transcript has to be accurately trans-spliced from gene fragments encoded on different circular chromosomes. About half of the transcripts are massively edited by several types of substitution editing and addition of blocks of uridines. Comparative analysis of mitochondrial RNA processing among the three euglenozoan groups, diplonemids, kinetoplastids, and euglenids, highlights major differences between these lineages. Diplonemids remain poorly studied, yet they were recently shown to be extremely diverse and abundant in the ocean and hence are rapidly attracting increasing attention. It is therefore important to turn them into genetically tractable organisms, and we report here that they indeed have the potential to become such.

6.1 Introduction

6.1.1 General Overview

It is beyond reasonable doubt that the genome of all extant mitochondria is of bacterial origin and with high confidence derives from a single acquisition of an alpha-proteobacterium by an archaeal cell (Zimorski et al. 2014). The mitochondrial genome was then subject to progressive reduction by downsizing of the endosymbiont genome and via the transfer of genes into the nucleus and subsequent retargeting of their products into the organelle. This led to a stepwise conversion

D. Faktorová · B. Kaur · J. Lukeš (✉)
Institute of Parasitology, Biology Centre and Faculty of Sciences, University of South Bohemia, České Budějovice, Czech Republic
e-mail: jula@paru.cas.cz

M. Valach · G. Burger
Department of Biochemistry and Robert-Cedergren, Centre for Bioinformatics and Genomics, Université de Montréal, Montreal, Canada

J. Cruz-Reyes, M. Gray (eds.), *RNA Metabolism in Mitochondria*, Nucleic Acids and Molecular Biology 34, https://doi.org/10.1007/978-3-319-78190-7_6

of the endosymbiont into a mitochondrial organelle that is controlled largely from the nucleus (Lithgow and Schneider 2010; Gray 2012). Closest to the original proto-mitochondrial version seems to be the gene-rich mitochondrial genomes of jakobid flagellates, which belong to the supergroup Discoba (Burger et al. 2013). In several lineages, the gradual loss of genes resulted in a minimized genome containing just two protein-coding genes (Flegontov et al. 2015) or in a complete elimination of the mitochondrial genome (Maguire and Richards 2014). In other lineages that include both uni- and multicellular eukaryotes, organization of the mitochondrial genome acquired an almost limitless spectrum of forms and structures, which led some authors to postulate that "anything goes" in these organellar genomes (Burger et al. 2003). Recent research shows that this statement also applies to the expression of mitochondrial genes, as their transcripts are more often than not subject to diverse and complex forms of RNA editing, splicing, and processing.

Moreover, the structural and organizational diversity is not confined to the genome and transcriptome but also applies to the proteome of these organelles. Interestingly, only a minor fraction of proteins constituting the mitochondrion (= mitoproteome) is a remnant of the original alpha-proteobacterium, while most of them are of diverse prokaryotic (but other than alpha-proteobacterial) or eukaryotic origin (Szklarczyk and Huynen 2010). The evolution of the mitochondrial ribosome represents an illustrative example of numerous lineage-specific losses accompanied by gains of a substantial amount of novel proteins (Desmond et al. 2011). Since most of extant eukaryotic diversity is hidden in poorly studied protist lineages (Pawlowski et al. 2012), it is likely that their mitoproteomes will significantly differ from that of the prototypic ones in yeast and human. The mitoproteomes of these latter opisthokonts are by far the best studied and are at present the largest in terms of protein repertoire, as summarized in MitoCarta2.0 (Calvo et al. 2016). However, it seems that some protist mitoproteomes may be as complex as those of their multicellular relatives, as exemplified by the studies of the mitochondrion of *Acanthamoeba castellanii* (Gawryluk et al. 2014) and *Trypanosoma brucei* (Zíková et al. 2017).

T. brucei and related trypanosomatid flagellates contain a single canonical mitochondrion that generates ATP via oxidative phosphorylation, with oxygen being the terminal electron acceptor (Tielens and van Hellemond 2009; Škodová-Sveráková et al. 2015). It is likely that in terms of main metabolic setup, *Diplonema papillatum* (Fig. 6.1) and other diplonemids have a rather similar organelle (our unpublished data). This presumption and the relatedness with kinetoplastid flagellates indicate that the mitoproteome of diplonemids will be highly complex rather than reduced as is the case of disparate anaerobic or microaerophilic eukaryotes (Maguire and Richards 2014). The well-studied mitochondrion of *T. brucei* with over 1100 proteins (Dejung et al. 2016; Urbaniak et al. 2013; our unpublished data) is as complex as the mitochondrion of multicellular organisms. Moreover, its metabolism is highly adaptable to the drastically different environments of the insect vector and the bloodstream of the mammalian host (Verner et al. 2015). It is reasonable to assume that the mitochondrion of diplonemids (Fig. 6.1) will be more akin to the

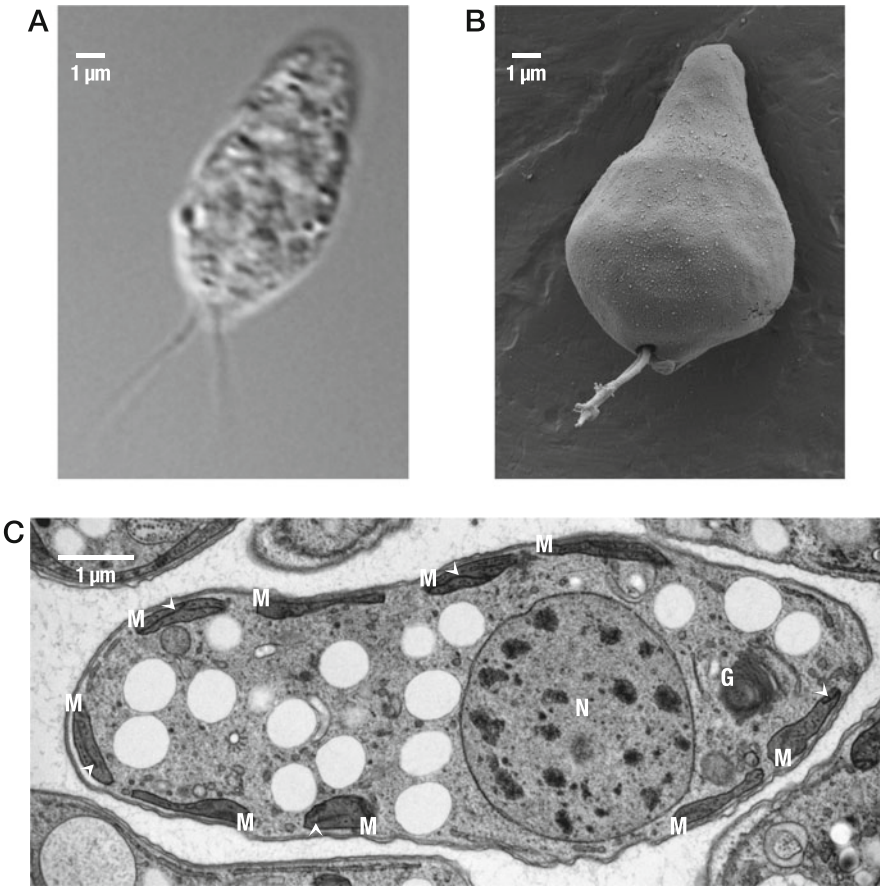

Fig. 6.1 Morphology of the model diplonemid, *Diplonema papillatum*. Light microscopy (**a**) and scanning electron microscopy (**b**) revealing the sac-like shape of the cell in culture and two heterodynamic flagella. (**c**) Transmission electron microscopy of a longitudinally sectioned cell with a prominent nucleus (N), single reticulated and peripherally located mitochondrion (M) with large discoidal cristae (arrowhead), and readily visible Golgi apparatus (G)

morphologically developed and metabolically highly active organelle of the insect-dwelling trypanosomes, especially since its nuclear genome is much larger (estimated at around 180 Mbp; our unpublished data) compared to that of the well-studied parasitic kinetoplastids (El-Sayed et al. 2005).

So far, diplonemids have been considered a marginal, rare, and rather insignificant group that received attention only thanks to its bizarre mitochondrial genome (see below). However, as they are emerging as major players in the world oceanic ecosystem, we predict that the era of diplonemids is just beginning.

6.1.2 Diplonemid Ecology, Taxonomy, and Phylogeny

With a single known exception (Triemer and Ott 1990), diplonemids seem to be confined to the marine environment including benthic waters. Yet in this largest planetary ecosystem, they are virtually omnipresent. In the frame of a global survey of marine microbial eukaryotes performed by the *Tara* Oceans expedition, based on the V9 region of the 18S ribosomal (r)RNA gene, over 85% of total eukaryotic plankton diversity is represented by unicellular eukaryotes (de Vargas et al. 2015). Diplonemids appeared among the most abundant groups, as they constitute the sixth most abundant (by reads of rRNA) and the third most diverse (by the number of operational taxonomic units, OTUs) eukaryotic group of the photic zone (de Vargas et al. 2015; Lukeš et al. 2015).

This came as a surprise since all the other prominently present eukaryotic groups were already well known, whereas diplonemids were until then considered rare and ecologically insignificant protists. In some stations of the *Tara* Oceans expedition, diplonemids reach up to 58% of all eukaryotes in the deeper mesopelagic zone (Flegontova et al. 2016) and were detected down to 6000 m in the poorly studied abyssopelagic zone (Eloe et al. 2011). Extensive sampling in the deeper pelagic layer, which is apparently the main habitat of diplonemids, further confirmed their prominent position among marine planktonic eukaryotes in terms of abundance and diversity (Flegontova et al. 2016).

The vast majority of marine diplonemids falls into a single clade dubbed the "deep-sea pelagic diplonemids" (DSPD) from deep oceanic environments (López-García et al. 2001, 2007; Lara et al. 2009) and was recently encountered at various depths ranging from surface to mesopelagic waters (Lukeš et al. 2015). The DSPD clade is also widespread in different geographical locations, ranging from tropical to polar regions, as well as from coastal to open ocean environments (Flegontova et al. 2016). Despite their diversity, ubiquity, and apparent abundance, we know close to nothing about the lifestyle, morphology, physiology, and biochemistry of the DSPD clade. Diplonemid species subjected to studies so far have been associated with parasitic or predatory lifestyles in plants, diatoms, and other marine protists (Schnepf 1994; Yabuki and Tame 2015). However, neither of the investigated species falls into the DSPD clade, which represents over 90% of diplonemid diversity.

The elusive DSPD diplonemids were, however, frequently encountered in a single-cell genomic survey of heterotrophic flagellates, conducted in the North Pacific Ocean (Gawryluk et al. 2016). Data generated from 10 individual cells, some of which belonged to OTUs most frequently represented in the *Tara* Oceans dataset, contain over 4000 protein-coding genes that fall into an ensemble of categories expected for heterotrophic protists. One striking feature is the high density of nonconventional introns that are absent from their kinetoplastid sister group (Gawryluk et al. 2016). Although we still have limited morphological and genetic information about the DSPD clade, it has now been formally described as a new class within Diplonemidea (Okamoto et al. 2018).

Moreover, significantly more information is available on the morphology, ultrastructure, and behavior of marine diplonemids not falling into the DSPD clade, but constituting several sister clades. These sac-like cells, highly variable in size and shape, have invariably two heterodynamic flagella inserted into a pronounced flagellar pocket and a DNA-rich mitochondrion with prominent lamellar cristae (Fig. 6.1). As is expected for a newly emerging speciose group of protists, the taxonomy and phylogeny of diplonemids is likely to evolve in the upcoming years.

6.1.3 Relationship of Diplonemids to Other Members of Euglenozoa

Diplonemids are part of the supergroup Euglenozoa, which includes two other morphologically and biochemically distinct main groups, kinetoplastids and euglenids (Adl et al. 2012; Cavalier-Smith 2016). This triumvirate was extended by the addition of anaerobic symbiontids (also called postgaardids) that were until recently placed among euglenids (Cavalier-Smith 2016). Symbiontids, which have a uniquely modified feeding apparatus and owe their name to their dependence on surface bacteria, are a poorly studied small group with only three genera described so far—*Postgaardia*, *Calkinsia*, and *Bihospites*. They were isolated from anoxic or low-oxygen environment, mainly from marine sediments (Yubuki et al. 2009, 2013; Breglia et al. 2010). Dependence on surface-bound episymbiotic bacteria along with hydrogenosome-like mitochondria with reduced cristae indicate a tight mutualistic relationship. Recently, symbiontids were shown to be present worldwide, similarly to the other euglenozoan groups, and they also seem to be more diverse than appreciated so far (Breglia et al. 2010; Edgcomb et al. 2011; Yubuki et al. 2013).

6.1.4 Mitochondrial Genome and Gene Structure

Despite the fact that all mitochondria are most likely derived from a single endosymbiotic event, mitochondrial genomes have evolved into myriad forms (Burger et al. 2003). The most diverse mitochondrial genomes are to be found among protists belonging to the supergroup Discoba (Smith and Keeling 2015). Jakobida harbor the most gene-rich mitochondrial genomes known (Burger et al. 2013), while anaerobic Metamonada exhibit mitochondrial reduction and even complete organelle loss (Karnkowska et al. 2016).

Arguably one of the most complex forms of mitochondrial DNA (mtDNA) evolved in diplonemid flagellates. *D. papillatum* carries in its organelle the largest amount of mtDNA known so far. The presence of an extraordinarily high amount of nucleic acids in its single mitochondrion was indicated by centrifugations of total

DNA in cesium chloride density gradients (Maslov et al. 1999). Later on, this observation was corroborated by staining mtDNA in situ, which revealed a strong continuous signal throughout the lumen of the reticulated organelle (Marande et al. 2005).

Flow cytometry experiments indicate that the *D. papillatum* nuclear genome has a size of about 180 Mbp (our unpublished data). In a more recent study, the cultured cells were stained simultaneously with an A + T-selective and nonselective dye, and the nuclear and mitochondrial signals were distinguished by color deconvolution, followed by quantification (Wheeler et al. 2012). This approach revealed massive inflation of the *D. papillatum* mtDNA, which with its estimated size of 270 Mbp not only exceeds that of the corresponding nuclear DNA but also represents the largest amount of DNA documented in any bacterium-derived organelle (Lukeš et al. unpublished). However, this enormous inflation does not reflect the gene content, which is rather ordinary, specifying subunits of respiratory complexes (six identified ORFs have unknown function) and the large and small subunit mitoribosomal rRNAs (Vlcek et al. 2011; Valach et al. 2014; Moreira et al. 2016).

Members of the genera *Diplonema* and *Rhynchopus*, as well as *Hemistasia phaeocysticola* have a multipartite mitochondrial genome (Vlcek et al. 2011; Yabuki et al. 2016). In *D. papillatum*, mtDNA is composed of thousands of non-interlocked circular chromosomes of at least 81 sequence classes that fall into two size categories—6 kb and 7 kb long, also labelled classes A and B, respectively (Marande et al. 2005) (Fig. 6.2). Within each class, chromosomes are essentially identical in sequence except for a short region called "cassette." Representing only about 5% of the chromosome, each cassette is composed of short unique 5′ and 3′ regions that flank a coding sequence, which is invariably a single gene fragment. With the sole exception of the small mitoribosomal rRNA, all genes are broken into up to 11 fragments, each of which resides on an individual chromosome (Valach et al. 2016); contiguous gene versions were not detected in mtDNA or nuclear DNA of *D. papillatum* (Figs. 6.2, 6.3 and 6.4). As a consequence of systematic fragmentation, not a single gene could be recognized at the outset of investigating the mitochondrial genome (Burger et al. 2016).

6.2 From Fragmented Genes to Contiguous Transcripts Via RNA Splicing

6.2.1 Splicing Types Found in Nature

As detailed above, genes in diplonemid mitochondria are systematically fragmented. However, mRNAs and rRNAs are, as usually, in one piece. Therefore, some kind of posttranscriptional mending must take place, which we have investigated mostly in *D. papillatum* and to some degree in *D. ambulator*, *Diplonema* sp. 2 [recently renamed to *Flectonema neradi* (Tashyreva et al. 2018)], and *Rhynchopus euleeides*.

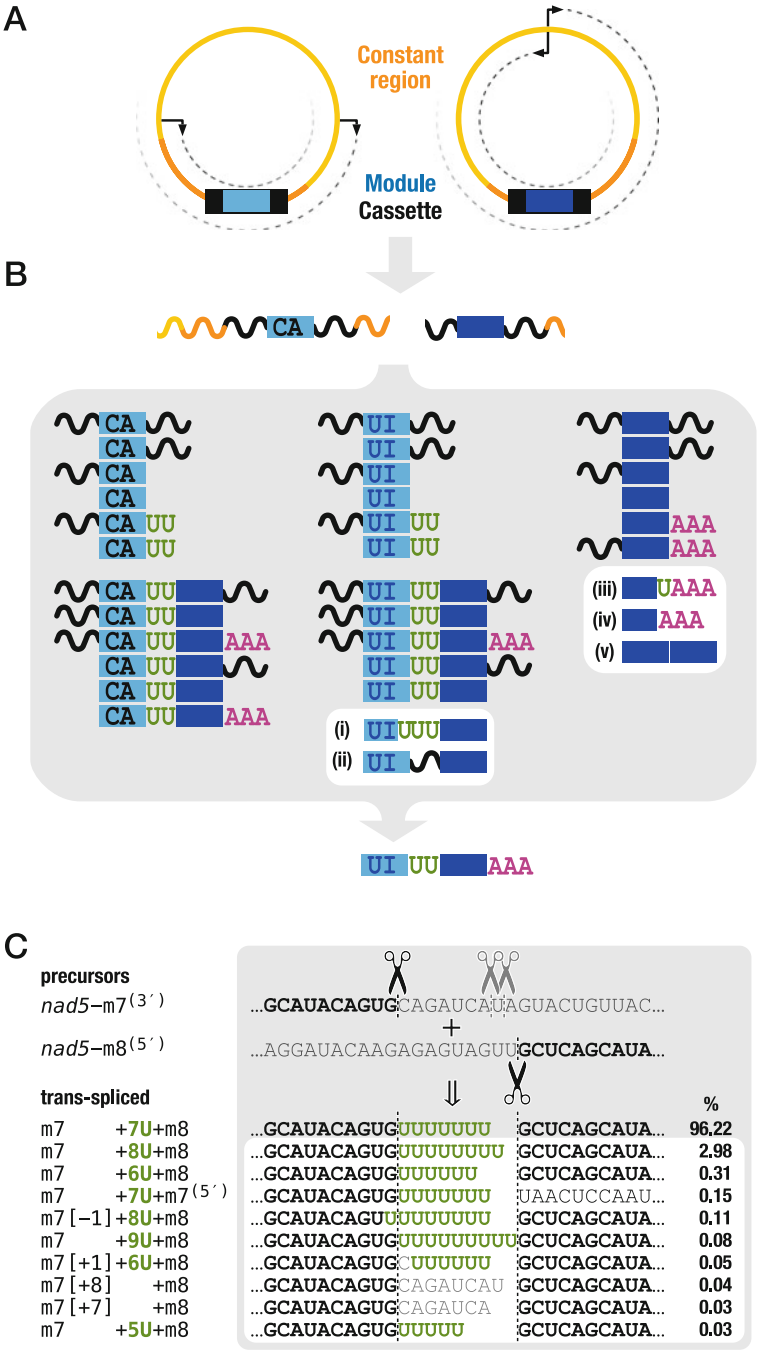

Fig. 6.2 Gene expression in diplonemid mitochondria. (**a**) Canonical circular mitochondrial chromosomes comprise a constant region of identical sequence across all members of a class

Our results reveal that the formation of contiguous mitochondrial mRNAs and rRNAs is diametrically different from conventional RNA splicing.

To summarize briefly, four major RNA splicing mechanisms exist across the various life forms and are classified according to the type of intervening sequence that is being eliminated: spliceosomal, tRNA (or archaeal), Group I, and Group II intron splicing (reviewed in Moreira et al. 2012). An additional less abundant type acting in fungi and vertebrates is IRE-mediated splicing that removes HAC1/XBP1 introns from pre-mRNA (Gonzalez et al. 1999). Each intron type is spliced by a distinct molecular machinery, be it a ribonucleoprotein complex (spliceosomal introns), catalytic RNA assisted by proteins (Groups I and II introns), or proteinaceous enzymes (tRNA and HAC1/XBP1 introns) (Hudson et al. 2015; Stahley and Strobel 2006; Zhao and Pyle 2017; Tanaka et al. 2011).

Initially, RNA splicing was viewed as an intramolecular (*cis*) reaction, removing an internal stretch of a pre-RNA and resealing adjacent exons. However, each of the abovementioned splicing types can also proceed in *trans*, i.e., the exons can reside on separate molecules, essentially representing halves of a pre-RNA broken apart within the intron.

6.2.2 RNA Processing Steps Prior to Trans-splicing in Diplonemid Mitochondria

Expression of fragmented genes in diplonemid mitochondria involves a unique mode of trans-splicing not seen before in any other system. The substrate for this particular trans-splicing is generated in a series of steps. First, gene pieces are

Fig. 6.2 (continued) [e.g., (**a**) and (**b**) in *D. papillatum*] and a unique cassette, which encloses a module (gene fragment). A cassette may be oriented in either sense relative to the constant region (illustrated at left and right). Long primary transcripts are initiated from the constant region by either two convergent promoters (left), or a bi-directional promoter (right), and extended into the other side of the constant region. (**b**) Separately transcribed single module precursors are processed in a highly parallelized process, which includes removal of 5′ and 3′ flanking noncoding regions, C-to-U, A-to-I, and U-appendage RNA editing of specific modules, 3′ polyadenylation of terminal modules, and trans-splicing of modules at processed ends (gray background). During the processing and trans-splicing, errors and their repair can take place: (i) exonucleolytic over-trimming of a module can be compensated for by a longer U-tract; (ii) 3′ flanking region of the upstream module can be retained instead of a U-tract; (iii) 3′ end over-trimming of can be compensated for by U-addition, even if the terminus is not normally a U-appendage site; (iv) polyadenylation of the terminal module may occur at over-trimmed sites; (v) two non-cognate modules can be joined together. Note that only the coding-strand transcripts are shown. (**c**) Examples of erroneous and error-compensating intermediates at the junction between the modules m7 and m8 of *nad5* detected in the total RNA from *Flectonema neradi* (*Diplonema* sp. 2). Coding and flanking noncoding regions are shown in black and gray, respectively. Note that the correctly processed, U-appendage-edited, and trans-spliced product represents the vast majority of detected RNAs

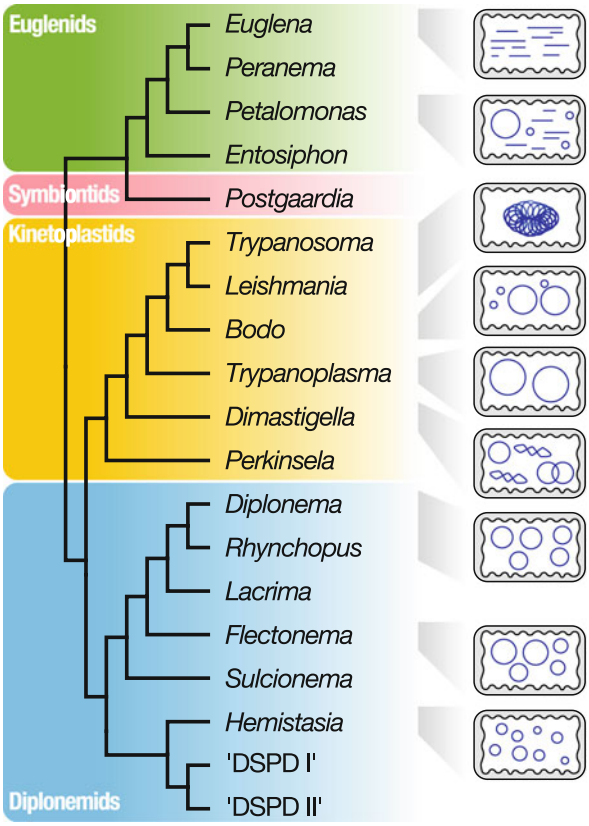

Fig. 6.3 Mitochondrial genome architectures and gene expression pathways in euglenozoans. Phylogenetic relationships among representative euglenozoan genera with their mitochondrial genome organization schematized. Euglenid mitochondria generally contain an assortment of linear molecules of variable length, though some species also harbor circular DNAs. Mitochondrial DNA of trypanosomatids, termed kinetoplast DNA (kDNA), is arranged into a single disc-shaped structure of catenated molecules. Bodonid species (*Bodo*, *Trypanoplasma*, and *Dimastigella*) contain non-catenated and relaxed or supercoiled circular molecules. Diplonemid mitochondrial circular chromosomes differ in size, with *Hemistasia* having particularly small chromosomes, as well as gene fragments

transcribed as long precursor molecules from a promoter located in the shared region of a mitochondrial chromosome. Although precise mapping of the transcription start site by in vitro capping experiments failed, the site was inferred to be located within the constant regions of chromosomes from precursor length determined by RNA circularization followed by RT-PCR across the ligation site (circRT-PCR) and amplicon sequencing (Kiethega et al. 2013) (Fig. 6.2a).

The promoter is most likely bi-directional (or two mirroring promoters exist in the constant region of circular chromosomes), since gene fragments are found encoded on either strand of the chromosome (plus and minus orientation of A-class and B-class

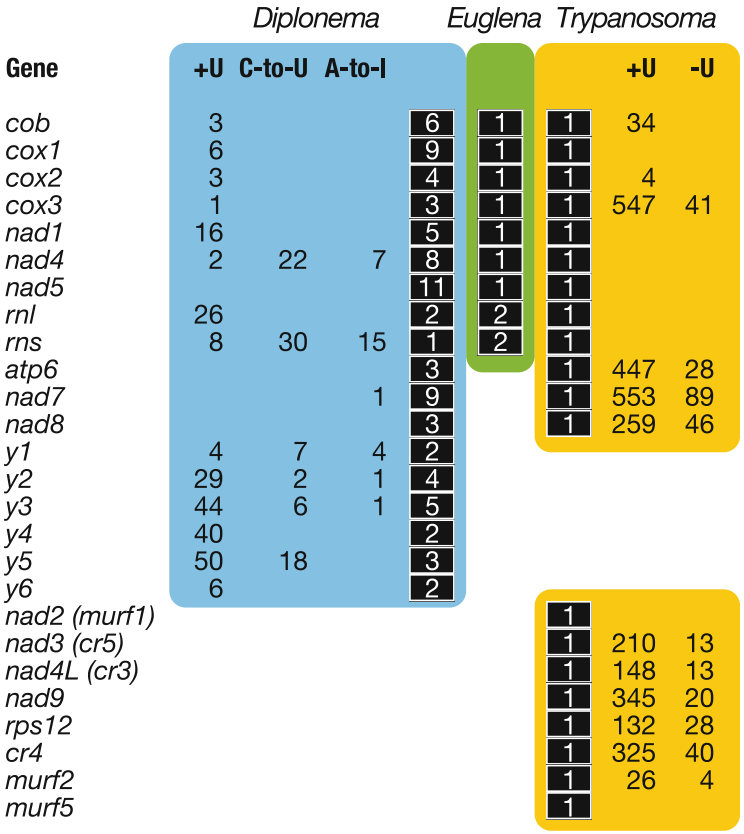

Fig. 6.4 Gene complement and editing site count across representative euglenozoans. Black rectangles indicate the presence of a gene (left column), with the number specifying the tally of precursor transcripts. Also shown is the total number of edits (+U, −U, C-to-U, A-to-I) in the corresponding mature transcript

chromosomes). In addition, antisense transcripts of individual gene fragments are detectable at low steady-state concentrations (Valach et al. 2014). Whether the amount of sense and antisense transcripts is regulated at the level of transcription initiation, transcription progressivity, or transcript degradation is currently not known.

The subsequent step in the expression of fragmented mitochondrial genes consists in end-processing of module transcripts. Processing intermediates, which are readily discernable by cDNA sequencing and circRT-PCR experiments, indicate that a combination of both endonucleolytic cuts and trimming are at work to generate transcripts that consist exclusively of coding regions (Fig. 6.2b). Only the 5′ ("first") module of protein-coding genes retains noncoding sequence, notably a 26- to 27-nt-long 5′ UTR (Kiethega et al. 2013).

Prior to trans-splicing, modules that will constitute the end of the mature transcripts undergo further maturation, notably addition of a homopolymer tail at the 3′

end. Transcripts of the "last" module from protein-coding genes are polyadenylated, forming the A-tail of mRNAs. Remarkably, A-tailed 3′-module transcripts belong to the most abundant precursors in total RNA, being present in certain cases (e.g., *cox1*) in a steady-state concentration comparable with that of the mature transcript (Marande and Burger 2007).

The last modules of both mito-rRNAs also receive a homopolymer tail. The large ribosomal subunit (mt-LSU) rRNA is polyadenylated. We reported previously that the transcript, once incorporated into the mitoribosome, has no A-tail (Valach et al. 2014). However, we realized recently that the result that led to this conclusion was due to an experimental artifact (see below "Limitations Encountered in Using the RNA-Seq Approach"). Reinvestigation of this issue by circRT-PCR demonstrates unambiguously that the A-tail length of mt-LSU rRNA (19–20 nt) remains unchanged after integration into the mitoribosome (Valach and Burger, unpublished data). The small ribosomal subunit (mt-SSU) rRNA from diplonemids studied so far is special in that its 3′ end carries a tail made from 8 Us. Curiously, in the kinetoplastid *Trypanosoma brucei*, both mt-rRNAs are modified by the addition of multiple terminal uridines (Adler et al. 1991).

Throughout eukaryotes, terminal adenylation or uridylation of rRNAs is generally a signal for degradation (Slomovic et al. 2010; Kuai et al. 2004). While exceptions to that rule have been reported for several taxa (Chaput et al. 2002; Mohanty and Kushner 2011), rigorous studies of either transcript stability or the state of rRNA actually incorporated into the ribosome are rare. Finally, prior to trans-splicing, certain modules will undergo RNA editing, which will be detailed in a later section.

6.2.3 *Succession of Posttranscriptional Processing Steps and Trans-splicing*

Contiguous mRNAs and mt-LSU rRNA of diplonemid mitochondria are formed through the joining of gene module transcripts that have been processed as described above. (Note that we use the term "module-transcript joining" synonymously with "trans-splicing") (Figs. 6.2, 6.3, 6.4 and 6.5). Intermediates of module-transcript end-processing, as well as trans-splicing, are readily detectable, not only in circRT-PCR experiments (Kiethega et al. 2013) and deep transcriptome sequencing (Moreira et al. 2016) but even in much less sensitive Northern hybridization (Marande and Burger 2007). This situation made *D. papillatum* an ideal system in which to investigate the temporal order of events.

Specifically, we observed a mixture of end-processing and trans-splicing intermediates, demonstrating that the succession of the individual posttranscriptional processing steps is not as strict as presented above (Fig. 6.2b). For example, module transcripts were detected that still carry adjacent, noncoding sequence at one terminus, while their other terminus is already trans-spliced to the neighbor module. This

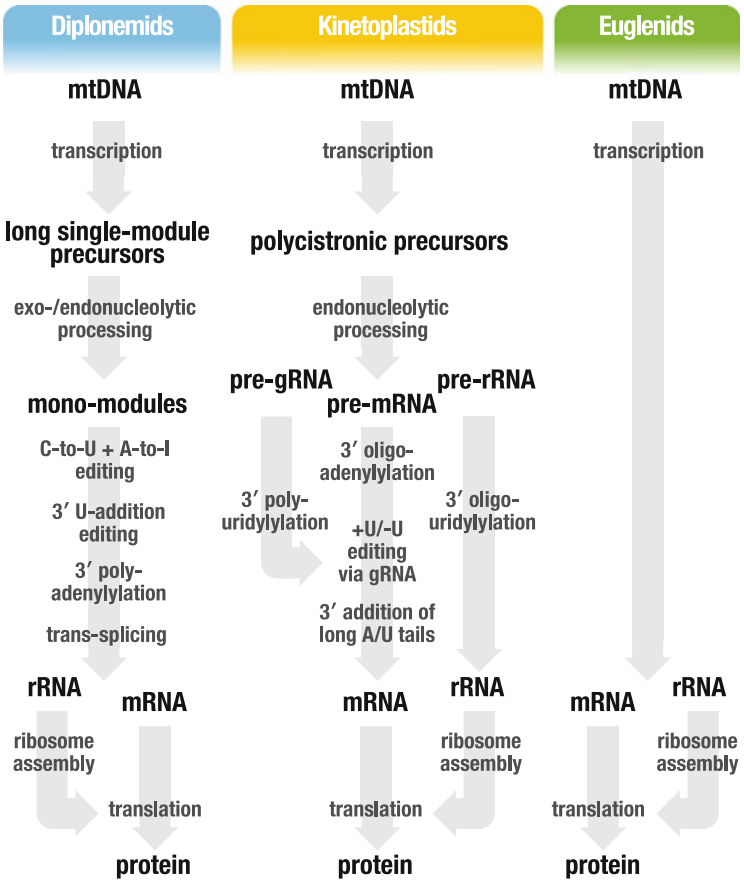

Fig. 6.5 Comparison of gene expression pathways among euglenozoans. Note that while the processes are sequential in kinetoplastids, diplonemids perform most processing and editing steps in parallel (see also Fig. 6.2)

shows that end-processing is not required to be completed for both module termini before trans-splicing can proceed. Further, polyadenylation of 3′-module transcripts is not a prerequisite for trans-splicing of their 5′ end to the upstream neighbor. Similarly, RNA editing of a module transcript via substitutions is not required to have taken place before trans-splicing. The only exception is U-appendage RNA editing. U-addition at module 3′ ends is completed before the corresponding terminus is joined to its downstream module or, in the case of terminal modules, before it is polyadenylated. Still, trans-splicing products with incompletely processed ends are the minority, as are those that are still pre-edited or not yet polyadenylated.

In summary, trans-splicing results in the correct sequential order of modules, yet proceeds without a particular directionality (e.g., 3′ to 5′). Thus transcript biogenesis in diplonemid mitochondria is a highly parallelized process (Kiethega et al. 2013).

6.2.4 Partner Selection in Trans-splicing

In diplonemids of the D/R clade, about 80 mitochondrial module transcripts have to be trans-spliced to their correct partner, raising the question how cognate module recognition is achieved. Cis sequence elements such as those adjacent to trans-splicing sites of conventional introns are not discernable, nor are conserved primary or secondary structure elements that are shared by all splice sites (Kiethega et al. 2011). Therefore, we posit trans-acting factors that recognize module transcripts to be joined and align them tail to head for trans-splicing.

Kinetoplastids possess guide RNAs, an abundant species of ~50-nt-long transcripts with a 5′-triphosphate and a U-tail, which are involved in mitochondrial uridine insertion and deletion RNA editing (Aphasizhev and Aphasizheva 2011; Read et al. 2016). We speculated initially that such molecules might guide trans-splicing in diplonemids, yet *Diplonema* mitochondria do not contain such an RNA species (Kiethega et al. 2013). Another conceivable kind of trans-acting splice guides would be full-length antisense mRNAs and antisense rRNA, serving as a single template for all splice junctions of a given gene. Since full-length genes are not present in *Diplonema* nuclear or mitochondrial DNA, these antisense transcripts would have to be produced by an RNA-dependent RNA polymerase (Valach et al. 2014), using sense transcripts as a template. Alternatively, there might be multiple (i.e., a total of 69) short antisense RNAs, each complementary to a single module junction.

We tested the "antisense RNA hypothesis" for *cox1* and *rnl*. For *cox1*, we performed exhaustive in silico analyses in an attempt to detect potential splice guides. Indeed, for each of the junctions, sequences were identified in the mitochondrial and nuclear genome that have the potential to be transcribed into splice guides (Kiethega et al. 2011); in turn, RT-PCR experiments indicated the existence of splice guides for five of the eight junctions (Kiethega et al. 2013). For *rnl*, which is ~100× more highly expressed, RT-PCR returned a readily discernable antisense product, Northern experiments showed a weak and smeary signal, and deep sequencing of a stranded cDNA library made from total RNA using an approach, which produces di-tagged first-strand cDNAs (ScriptSeq kit), yielded ~2.5% read coverage of the complementary strand bridging the *rnl*-m1/*rnl*-m2 junction. This rate is more than two times above the 1% of spurious antisense reads considered typical for the methodology (Valach et al. 2014).

Yet, these results must be considered with caution. The RT-PCR technique may produce artifactual antisense products, e.g., by polymerase template switching. Moreover, in more recent RNA-Seq experiments (Valach et al., unpublished data), we noted a considerable variation in the depth of junction-crossing antisense reads between libraries made from different RNA preparations (1–6%). Furthermore, total RNA-Seq libraries, made using the first-strand dUTP-cDNA approach (as implemented in the TruSeq kit) (Parkhomchuk et al. 2009) of *D. ambulator*, *F. neradi* (*Diplonema* sp. 2), and *R. euleeides*, showed only a coverage of 1% (i.e., background level). In sum, at the current time, it is uncertain if *Diplonema* cells

indeed produce a significant steady-state level of genuine junction-crossing antisense RNAs for *cox1* or mt-LSU rRNA.

Instead of RNA guides, trans-splicing could also be directed by guide proteins. Such proteins must be capable of binding selectively to specific RNA sequence motifs. Sequence-specific RNA-binding proteins are generally composed of several conserved RNA-binding domains that engage in base-dependent interactions with RNA and form a three-dimensional shape that is complementary to that of the recognized RNA motif (Ban et al. 2015). The most common and best-studied RNA-binding proteins are characterized by either tristetraprolin (TTP)-type tandem zinc finger domains, pentatricopeptide repeat protein (PPR) domains, Pumilio-FBF (Puf) domains, or RNA recognition motif (RRM) domains. We detected genes from the three latter families in the preliminary version of the *D. papillatum* nuclear genome sequence (our unpublished data). It remains to be confirmed, in silico and experimentally, which of these predicted proteins are located in the mitochondrion.

6.2.5 *Accuracy of Module Trans-splicing*

With several dozen distinct gene module transcripts in the diplonemid mitochondrion, what is the trans-splicing accuracy? Deep transcriptome sequencing of total *D. papillatum* RNA shows on average ~0.1% mis-spliced transcripts, with certain modules being considerably more "promiscuous" than others (Fig. 6.2c). For example, in a total RNA library, as much as ~16% of trans-spliced *cox1*-m7 3′ termini have been joined incorrectly, i.e., predominantly to *cox1*-m6 instead of *cox1*-m8. In contrast, poly-A libraries contain only about 0.4% of mis-joined *cox1*-m7 products. Thus, incorrectly joined modules appear to be eliminated by some quality control mechanism in mature polyadenylated mRNAs. Mis-splicing might be caused by short identical sequence motifs. A preliminary search (≥6-nt-long motifs within 20-nt from the junction) did not reveal recurrent patterns. The analysis has to be extended to more distant regions and also consider secondary structure motifs.

Interestingly, a recent investigation of the distantly related diplonemid *H. phaeocysticola* (Yabuki et al. 2016) recovered rare cases of mitochondrial transcripts in which the first *cox1* module was joined to downstream modules other than the expected module 2. These findings were interpreted as indicative of an mRNA assembly pathway containing a step of module-transcript insertion in contrast to a "concatenation" model described for *D. papillatum*. Although this suggestion is an intriguing possibility that merits further study, in the light of the existence of module mis-joining in all D/R diplonemids studied thus far (Valach et al. 2016), it seems more plausible that the rare transcripts with unexpected module order in *Hemistasia* also represent dead-end intermediates.

6.2.6 Speculations on the Trans-splicing Reaction and Machinery

While the process of trans-splicing in diplonemid mitochondria is quite well characterized, open questions remain about the reaction itself. Given the absence of conserved nucleotides at the splice junctions, a ribozyme reaction mechanism is unlikely, thus favoring the hypothesis of an enzyme-based ligation of module transcripts. For example, splicing of conventional tRNA introns and of HAC/XBP1 involves an end-joining reaction catalyzed by RNA ligases of the T4 Rnl or the RtcB family (Popow et al. 2012). Preliminary analyses of the nuclear genome draft from *D. papillatum* show that it encodes proteins of the RtcB family. Some family members will be involved in the splicing of nuclear tRNA introns, while others might join mitochondrial module transcripts. We postulate that module ligation and matchmaking are performed by an integrated molecular machinery—the hypothetical joinosome (Valach et al. 2016)—whose identification is our priority.

Interestingly, a second case of unorthodox trans-splicing has been reported in mitochondria of certain dinoflagellates. One of the mitochondrion-encoded gene, *cox3*, is broken up into two separate pieces, while its transcript is contiguous (Jackson and Waller 2013). Whether the machinery involved shares communalities with the system in diplonemids remains to be investigated.

6.2.7 Limitations Encountered in Using the RNA-Seq Approach

By investigating the mitochondrial transcriptome of diplonemids, we became aware of several limitations of the RNA-Seq approach (see also Ozsolak and Milos 2011; Levin et al. 2010). One problem is that read coverage only partially represents the actual steady-state level of a transcript, especially when the library construction protocol, as in our case, uses hexamer primers for initiating first-strand cDNA synthesis. Not only does coverage drop strongly toward the template's extremities, but also internally drastic fluctuations occur, probably due to differences in efficiency of primer annealing to particular sequence contexts. An important challenge in our analyses was the low read coverage in homopolymer tracts, likely caused by inefficient progressivity at the stage of reverse transcription and sequencing.

Strand specificity is another issue, especially when analyzing the level of genuine antisense transcripts. We noted that the degree of spurious antisense reads depends on the sequence of the gene in question and may be above or below the overall vendor-stated rate of a given library construction protocol. To determine exactly the level of spurious antisense products, controls with an in vitro-synthesized RNA should be performed so that the portion of genuine antisense transcripts in the sample can be reliably assessed. Our approach was to synthesize in vitro an ~200-nt-long RNA that covers the *rnl*-m1/*rnl*-m2 junction and—for cost reasons—mix it into an

RNA preparation of another organism to be sequenced, construct a stranded RNA-Seq library of this mix, and sequence it.

Further, we encountered the problem that capture probes used for eliminating over-abundant transcripts, such as rRNAs, are not always removed completely from the sample prior to library construction. Remnants of the capture probe prime reverse transcriptase during first-strand synthesis, generating artificially profuse amounts of reads all starting at the same position. Capture probes are biotinylated for easy removal with streptavidin-coated magnetic beads after annealing with their target rRNA. We assume that sample contamination occurred because of a too low ratio of beads to capture probe and/or because of incomplete biotinylation of the oligonucleotide.

A final unexpected issue was that different RNA-Seq library construction kits are not equally effective in reproducing A-tails. Control experiments with circRT-PCR confirmed that mt-LSU rRNA has indeed 19–20 As at its 3′ end, just as determined via the reads from the ScriptSeq library (Valach et al. 2014), while the TruSeq library returned only 0–2-nt-long A-tails. We assume that the particular mix of random primers used by the TruSeq protocol is biased against annealing with A-tracts.

6.3 From Defective to Functional Products Via RNA Editing

As alluded to in the previous section, the convoluted mitochondrial gene expression in diplonemids does not stop at ribonucleolytic processing and covalent joining to form the functional mRNA or rRNA. Certain module transcripts undergo additional maturation steps, which result in nucleotide-level changes of the transcript sequence.

6.3.1 Types of RNA Editing Systems in Mitochondria

In general, RNA modifications corresponding to nucleotide insertions, deletions, or substitutions are referred to as RNA editing (reviewed in Knoop 2010). They may take place directly during transcription or at later maturation stages, may involve a variety of enzymatic activities (e.g., base deaminases, nucleases, ligases, 3′ or 5′ polymerases), and may affect mRNAs, rRNAs, or tRNAs, as well as other types of transcripts like miRNAs, ncRNAs, or retrotransposons (reviewed in Knoop 2010; Nishikura 2016). We first briefly overview the diversity of RNA editing mechanisms in mitochondria (Table 6.1), with emphasis on five instances where the enzymatic players have been characterized, before addressing the peculiarities of the diplonemid RNA editing.

C-to-U substitution is commonly encountered in land plant organelles, with hundreds to thousands of editing events per genome (reviewed in Takenaka et al.

Table 6.1 Diversity and distribution of RNA editing types in mitochondria

Type of change		Distribution	Transcript category	Selected references
Substitution	C-to-U	Land plants	mRNA	Reviewed in Takenaka et al. (2013)
		Slime molds	mRNA	Bundschuh et al. (2011)
		Heteroloboseans	mRNA	Rüdinger et al. (2011) and Fu et al. (2014)
		Diplonemids	mRNA, rRNA	Moreira et al. (2016)
		Malawimonads	mRNA	Authors' unpublished data
	U-to-C	Land plants	mRNA	Reviewed in Takenaka et al. (2013)
	A-to-I	Diplonemids	mRNA, rRNA	Moreira et al. (2016)
	Various	Dinoflagellates	mRNA, rRNA	Lin et al. (2002) and Jackson et al. (2007)
Insertion	Predominantly C (also U, A, G)	Slime molds	mRNA, rRNA, tRNA	Bundschuh et al. (2011), Mahendran et al. (1991) and Chen et al. (2012)
	Predominantly G (also A, C, U)	Heteroloboseans	mRNA, rRNA, tRNA	Yang et al. (2017)
	A	Dinoflagellates	mRNA, rRNA	Jackson et al. (2007)and Jackson and Waller (2013)
		Metazoans	tRNA	Yokobori and Pääbo (1995)
	U	Metazoans	mRNA	Vanfleteren and Vierstraete (1999) and Lavrov et al. (2016)
		Kinetoplastids	mRNA	Reviewed in Read et al. (2016)
		Diplonemids	mRNA, rRNA	Moreira et al. (2016)
	Various (5′ end)	Amoebozoans	tRNA	Jackman et al. (2012)
		Fungi	tRNA	Laforest et al. (1997)
		Heteroloboseans	tRNA	Authors' unpublished data
	Various (3′ end)	Jakobids	tRNA	Leigh and Lang (2004)
		Metazoans	tRNA	Segovia et al. (2011)
Deletion	A	Slime molds	mRNA	Gott et al. (2005)
	U	Kinetoplastids	mRNA	Reviewed in Read et al. (2016)

2013; see also Chap. 9). Although the enzyme responsible for the deamination reaction has not yet been unambiguously identified, a plethora of ancillary cofactors has been catalogued (reviewed in Sun et al. 2016) and the indirect experimental evidence has been converging on the DYW family of PPR proteins as the catalytic component (Salone et al. 2007; Shikanai 2015).

A different process takes place in the mitochondria of slime molds such as *Physarum polycephalum* (Bundschuh et al. 2011; Mahendran et al. 1991), where the plentiful mono- and dinucleotide insertions at internal sites in mitochondrial transcripts occur co-transcriptionally, probably relying on the interplay between the RNA polymerase complex and its substrate DNA (Visomirski-Robic and Gott 1997; see also Chap. 8). However, the exact mechanism of this system remains to be elucidated. Much better understood is the editing of mt tRNA at their 5′ and 3′ ends, a posttranscriptional nucleotide insertion process observed in various eukaryotic clades (Table 6.1). Several amoebozoans replace 5′ terminal nucleotides of their mt tRNA employing an unconventional 3′ to 5′ polymerase of the Thg1 family (Abad et al. 2011; see also Chap. 7). Editing of tRNA at 3′ end can proceed via polyadenylation by a 3′ terminal adenylyltransferase (poly-A polymerase), generating a missing secondary structure element (reviewed in Rammelt and Rossmanith 2016).

Finally, one of the best understood RNA editing processes takes place in the mitochondrion of kinetoplastids (Benne et al. 1986), where most mRNAs undergo extensive insertion and/or deletion of U residues by a complex ribonucleoprotein machinery (reviewed in Read et al. 2016; see also Chap. 5). The multicomponent editosome includes endonuclease, U-specific exoribonuclease, terminal uridylyltransferase (TUTase), and ligase activities, which for each edited site complete a cycle consisting of cleaving the transcript, inserting/deleting a number of Us specified by a partially complementary guide RNA, and religating the broken strand.

6.3.2 Idiosyncratic RNA Editing in Diplonemid Mitochondria

6.3.2.1 Appendage of Uridines

RNA editing in *D. papillatum* mitochondria was noted early on in the *cox1* cDNA, which contained six nonencoded Ts inserted between its modules 4 and 5 (Marande and Burger 2007). Once high-throughput cDNA sequencing technologies made possible a comprehensive investigation of RNA editing sites, a more complex picture emerged: in this diplonemid, 240 Us are inserted at 18 sites distributed across 14 out of its 18 genes (Moreira et al. 2016) (Figs. 6.2b, c and 6.4). Insertions of no other nucleotide besides U, nor nucleotide deletions, have been detected in any diplonemid analyzed to date.

While many U-tracts are shorter than the one in *cox1*, a stretch of as many as 26 Us is added in the middle of the mt-LSU rRNA (*rnl*) (Valach et al. 2014). Recently, we have confirmed the presence of even more impressive 50 Us in a row in the mature transcript of the (unassigned) gene *y5* (Valach et al. 2017). Such long U-tracts blur the line between the conventional definition of RNA editing (a single or a couple of affected nucleotides at a single site) and posttranscriptional modifications traditionally not considered to represent RNA editing, such as terminal polyuridylylation.

In *Diplonema*, the U residues are not inserted in a cut-add-reseal strategy as in kinetoplastid RNA editing. Instead, they are appended to 3′ termini of processed

modules prior to trans-splicing. First, all identified U insertions are confined to module junctions or to 3′ ends of last modules, just upstream of poly-A tails in case of mRNAs. Second, circular RT-PCR, 5′ and 3′ RACE, and primer extension assays showed that in *D. papillatum*, no *cox1* mRNA trans-splicing intermediate lacks the six Us after the modules 4 and 5 have been joined together, nor does it contain the six Us attached to the downstream module 5. It is exclusively the 3′ end of the upstream module 4 to which the U-tract is appended (Kiethega et al. 2013). Comprehensive investigation of the entire transcriptome further confirmed that only the 3′ end-processed module transcripts are uridylylated, irrespective of the maturation state of the 5′ end of that same module (Moreira et al. 2016). In this respect, the U-appendage pathway is similar to trans-splicing, which can also proceed even if the opposite end of a module that does not participate in module joining is incompletely processed (see the previous section; Fig. 6.2b).

The close relationship between module transcript joining and uridylylation has been further corroborated by our deep-coverage transcriptome data from *D. papillatum* and three additional diplonemids, revealing transient errors or "background noise." At a frequency around 0.1%, Us (mostly 1 to 3) are being added even at module 3′ ends that normally do not undergo U-appendage RNA editing. Interestingly, the vast majority of these abnormal U-addition events occur in trans-spliced transcripts whose upstream partner's 3′ end is several nucleotides shorter, with the Us compensating for the missing sequence (Valach et al. 2017). We thus hypothesize that the same process ensuring the usual U-appendage RNA editing can also repair a deletion at a module junction, which could have arisen from erroneous over-trimming during module transcript end-processing (Fig. 6.2b, c). Curiously, at certain, but not all, junctions usually separated by a U-tract (e.g., *nad5*-m7/m8 in *F. neradi* [= *Diplonema* sp. 2]), we also observe rare (<1%) occurrences of two cognate modules being joined together without the U-tract (Valach et al. 2017). However, in these cases, the missing sequence is compensated by a sequence stretch originating from the upstream module's 3′ flanking region, which is present instead of the expected U-tract (Fig. 6.2b, c). It remains to be seen whether these defective trans-spliced products are translated or rather are discarded by some downstream control mechanism, as is apparently the case for mis-joined, non-cognate modules (see Sect. 2). In any case, these two observations—gap filling by U-addition or partial retention of a 3′ flanking region—imply that some kind of a molecular ruler measures the length of the module transcripts or the distance between the two RNA ends to be joined. Likely candidates are the factors involved in junction recognition (see below).

6.3.2.2 Clustered Substitutions of Adenosines and Cytidines

The screening for cDNA vs. genome differences further unveiled 85 cytidine-to-uridine (C-to-U) substitutions, well known from organelles of many species (Table 6.1). In addition, we discovered 29 adenosine-to-guanosine (A-to-G) substitutions in half of the *D. papillatum* genes (Moreira et al. 2016) (Fig. 6.4). These substitutions indicate C-to-U and A-to-I base deamination (inosine is read as

guanosine during reverse transcription). Indeed, A-to-I deamination could readily be demonstrated experimentally (Moreira et al. 2016). While this type of deamination is common for tRNAs, ours was the first report of its kind for mitochondrial mRNAs and rRNAs. Diplonemid mitochondria also show an exceptionally high rate (>95%) of RNA editing at a given site, and further, in most instances, diplonemid editing sites congregated in clusters denser even than those of the so-called hyper-edited segments in metazoan nuclear transcripts (Wahlstedt and Ohman 2011) (Fig. 6.4).

The latter two features are particularly intriguing. As a general rule, we considered as a cluster a group of adjacent sites where more than half of the potentially editable residues (As + Cs) in a row were indeed edited. In *D. papillatum*, in all but one cluster (*y5*-m1), every single C in a cluster is edited, as are most As (Moreira et al. 2016). For example, in an 85 nt-long region of mt-SSU rRNA, all 15 As and all 30 Cs are substituted. Although most sites are edited to high levels, there are few partially edited (5–40% rate) sites, with editing rates generally slightly higher for C-to-U than for A-to-I substitutions. Still, all of these occur within a cluster or at its boundaries and thus may indicate "misfiring" of the editing enzyme(s).

Our comprehensive analyses of trans-splicing and editing intermediates in *D. papillatum* also revealed that substitution RNA editing in a cluster progresses stochastically and not directionally. As mentioned in the previous section on trans-splicing, substitution editing is essentially completed before trans-splicing begins; pre-edited or partially edited module transcripts that are already trans-spliced are found only at below 5% (Moreira et al. 2016) (Fig. 6.2).

6.3.3 *Functional Consequences of RNA Editing*

Both types of RNA editing in diplonemid mitochondria appear to be critical for the function of the affected transcripts. For example, in the case of *rnl*, the long U-tract is predicted to form segments of two helices of the mt-LSU's central domain 0 (Valach et al. 2014). The six Us of the *cox1* mRNA add codons for amino acids that restore the three-dimensional structure of the protein (Kiethega et al. 2011), whereas the two Us of the *nad4* transcript rectify the reading frame of the coding sequence (Moreira et al. 2016). In several mature transcripts (e.g., *cox3*, *y3*), U-appendage together with polyadenylation creates the termination codon, and in *nad1* mRNA, the 16 nt-long U- tract at its 3′ end adds codons for five additional phenylalanyl residues to the polypeptide, thus completing the C-terminal membrane-spanning helix. Similarly, substitution RNA editing of *nad4* mRNA leads to a protein that contains all its hydrophobic transmembrane helices instead of lacking the second helix (Moreira et al. 2016). Comparative analysis of the gene across four diplonemid species demonstrated that the proteins encoded by edited mRNAs became more similar to one another, as well as to homologs from other organisms.

Interestingly, dense C-to-U and A-to-I RNA editing results in codons rich in U and G (I) residues, which mostly specify apolar amino acids. In addition,

uridylylation creates UUU codons, which code for the hydrophobic phenylalanine residue. Apolar and hydrophobic amino acids being favored in membrane-embedded or membrane-anchored proteins, one can easily imagine how these two types of RNA editing in particular could become evolutionarily fixed for mending the deterioration of diplonemid genes, which all encode proteins of this class.

6.3.4 Predicted Components of the Editing Machineries

Based on our insights into diplonemid mitochondrial RNA editing described above, we have attractive working hypotheses about the nature of the enzymes involved in the two types of RNA editing. Akin to kinetoplastids, diplonemids add Us at the 3′ end of mitochondrial transcripts, suggesting that U-appendage RNA editing is performed by a TUTase enzyme similar to RET2 of the trypanosome editosome. For substitution RNA editing, a nucleotide/base excision-replacement system is conceivable, but our current data rather indicate that the mechanism relies on base deamination. Since the C-to-U and A-to-I edits are closely spaced and display no ordering of pre-edited and edited positions in transcript intermediates, we speculate that an enzyme able to deaminate both Cs and As is involved. Interestingly, a precedent for such an enzyme was discovered in the kinetoplastid *T. brucei* (Rubio et al. 2007). According to our preliminary analyses, several genes potentially encoding proteins with a nucleotidyltransferase or deaminase domain are present in the draft nuclear genome of *D. papillatum* (our unpublished data).

As in the case of trans-splicing (Kiethega et al. 2011), no cis-elements have been identified in the genome sequence that have the potential to direct the enzymatic machinery to the RNA editing sites (Moreira et al. 2016). This led us to postulate that all three processes—the trans-splicing, U-appendage, and substitution RNA editing—are guided by trans-acting factors (Valach et al. 2016). Among the numerous RNA-binding protein (RBP) families that were mentioned in the previous section and that could be implicated in mitochondrial RNA processing in *D. papillatum*, PPR proteins have emerged as primary candidates. They are not only the most notable cofactors of C-to-U editing in land plant organelles (Sun et al. 2016) but also serve as cofactors of numerous other organellar RNA transactions in a wide variety of organisms (Manna 2015).

6.4 Comparison of Mitochondrial Gene Expression Across Euglenozoa

In molecular biology textbooks, expression of genetic information is simple and straightforward. However, in some organisms it is surprisingly derived, incomprehensible, and gratuitously inefficient. This applies not only to diplonemids but also

to those protists from other euglenozoan groups. Since no molecular data are currently available about symbiontids, we will compare the expression of mitochondrial genes among the three other euglenozoan groups—diplonemids (besides *Diplonema papillatum* also represented by *Diplonema ambulator, Flectonema neradi* (*D.* sp.2), and *Rhynchopus euleides*), trypanosomatids (represented by *Trypanosoma brucei*), and euglenids (represented *by Euglena gracilis*, *Peranema trichophorum*, and *Petalomonas cantuscygni*). What is currently known about the organization of their mtDNA and about the mitochondrial gene expression of *D. papillatum*, *T. brucei*, and *E. gracilis* is summarized in Figs. 6.3, 6.4 and 6.5.

6.4.1 *Mitochondrial A + T Content and Gene Complement Throughout Euglenozoa*

All euglenozoans carry a single mitochondrion with discoidal cristae, with possibly the only exception being the euglenid *P. trichophorum*, which possesses several small elongated mitochondria (Roy et al. 2007). While packaging of mtDNA into a dense single kinetoplast remains a character exclusive to kinetoplastids, mtDNA in diplonemids and euglenids is homogenously distributed throughout the organellar lumen and is only exceptionally organized into tiny bodies or foci. The A + T content of mtDNA varies across euglenozoans—it has a typically higher A + T content in *T. brucei*, *R. euleides, E. gracilis*, and *P. cantuscygni*, but in *P. trichophorum* and *D. papillatum*, the A + T content is unusually low (Roy et al. 2007; Dobáková et al. 2015).

Regardless of its structure (Fig. 6.3), the mitochondrial genome of euglenozoans has a very similar gene composition. It is typically composed of subunits of four respiratory complexes, complex I (NADH dehydrogenase; *nad* genes), complex III (ubiquinone-cytochrome *c* oxidoreductase; gene *cob*), complex IV (cytochrome *c* oxidase; *cox* genes), and complex V (ATP synthase; gene *atp6*), and two mitoribosomal RNAs (*rnl* and *rns*) (Faktorová et al. 2016) (Fig. 6.4). Moreover, mtDNA in *T. brucei* also encodes ribosomal protein Rps12 (Alfonzo et al. 1997). No tRNA genes have been identified in any euglenozoan mitochondrial genome and therefore have to be imported from the cytoplasm (Alfonzo and Söll 2009).

6.4.2 *Comparison of* D. papillatum *Genome Structure with Other Diplonemids*

Diplonemid species studied to date possess several classes of circular chromosomes. Compared to the 6.0 kb and 7.0 kb classes in the case of *D. papillatum*, the sizes in the other studied species vary from 4.5 kbp to >6.7 kbp (with a majority at ~5 kbp) in *D. ambulator*, from ~5 kbp to ~10 kbp in *Diplonema* sp. 2 (= *F. neradi*;

Tashyreva et al. 2018), and from 5 kbp to 12 kbp (with a majority at ~7 kb and ~8 kb) in *R. euleeides* (Kiethega et al. 2011; Valach et al. 2017). In *D. papillatum* almost every gene split into fragments (up to 11) and each piece is encoded on a separate chromosome (Moreira et al. 2016) (Fig. 6.4). While the fragmentation pattern is essentially identical in the other investigated D/R clade species, up to eight gene pieces were found to be encoded on the same chromosome (Valach et al. 2017).

In *H. phaeocysticola*, the size of mitochondrial chromosomes sequenced so far is significantly smaller (2.7–3.2 kb), with twice as many half-sized gene fragments (Yabuki et al. 2016), and a similar situation seems to be the case in the newly isolated species belonging to the same clade (our unpublished data). Moreover, currently we are trying to shed more light on this group by studying the mitochondrial genome structure of newly described diplonemid species that belong to the genus *Rhynchopus* (*Rhynchopus humris* and *Rhynchopus serpens*) or to the newly described environmental clade (*Lacrimia lanifica*), and even a novel early-branching clade, represented by *Sulcionema specki* (Tashyreva et al. 2018).

6.4.3 Kinetoplastids: Uridines In and Out

Kinetoplastids are either free-living (e.g., *Bodo saltans*) or parasitic protists, which include human parasites of major medical importance, such as members of the genera *Trypanosoma* and *Leishmania*. They are characterized by a kinetoplast, a compact mass of mtDNA composed of dozens of maxicircles and thousands of minicircles (Shapiro and Englund 1995; Stuart and Feagin 1992). Maxicircles (~20 kbp) represent functional equivalents of mtDNA in other organisms. Most of the mitochondrial genes (12 out of 18) are literally encrypted (Fig. 6.4). This means that their transcripts have to undergo a process of RNA editing, which restores meaningful open reading frames that are translatable (Fig. 6.5). Since its first description in *T. brucei* (Benne et al. 1986), many distinct and unrelated types of RNA editing have been described in organisms across the entire tree of life (Read et al. 2016). In *T. brucei* and other kinetoplastid protists, RNA editing is guided by small minicircle-encoded molecules called guide RNAs (gRNAs) that serve as template for the insertions and/or deletions of uridines into the pre-edited sequence at specific positions (Aphasizhev and Aphasizheva 2011).

Interestingly, about a thousand distinct minicircle-encoded gRNAs, together with more than 70 different nucleus-encoded proteins, are necessary for proper expression of the small complement of 18 mitochondrion-encoded genes (Alfonzo et al. 1997; Verner et al. 2015; Read et al. 2016). More specifically, in addition to well-described RNA editing core complex (RECC) or the 20S editosome (Göringer 2012), several other ribonucleoprotein complexes, e.g., the MRB1 complex, were recently shown to be involved in the RNA editing and processing machinery (Ammerman et al. 2012; Read et al. 2016; Dixit et al. 2017).

The uridine insertion/deletion type of RNA editing in kinetoplastids, and even more the obscure and still unrecognized machinery for trans-splicing associated with uridine insertions and cytidine-to-uridine and adenine-to-inosine substitution RNA editing in diplonemids, appears extremely costly in comparison to their benefits. So far, no advantages of these strategies have been proposed, leading to the speculation that they most likely originated as a result of constructive neutral evolution (Flegontov et al. 2011; Lukeš et al. 2011).

6.4.4 Euglenids: Surprises in Their Own Right

It was hoped that elucidation of the structure and expression of mitochondrial genome in euglenids, the sister group to kinetoplastids, would shed light on the origin of the latter groups's bizarre mtDNA structure and RNA processing. Therefore, it was quite surprising when the mitochondrial genome of *E. gracilis* was recently shown to be extremely streamlined, without any evidence of RNA editing (Dobáková et al. 2015) (Fig. 6.5). This mitochondrial genome consists of a heterogeneous population of 1 to 9 kbp-long linear fragments. Up to now, only seven protein-coding genes have been discovered, as well as two mito-rRNAs (mtSSU and mtLSU), which are each split into two fragments (Spencer and Gray 2011; Dobáková et al. 2015) (Fig. 6.4).

Nonetheless, transmission electron microscopy of the early-branching euglenid *P. cantuscygni* revealed a structure in its mitochondrion resembling the kinetoplast of the kinetoplastid flagellates (Leander et al. 2001; Lee and Simpson 2014). Observations of the mtDNA fraction by electron microscopy confirmed that linear DNA molecules are most frequent, but also small (1 to 2.5 kbp) and large (~40 kbp) circular molecules have been infrequently noted. This observation together with the absence in the sequenced mtDNA segments of some highly conserved mitochondrion-encoded subunits of respiratory complexes III and IV suggest that some kind of RNA editing and gene encryption may exist in this species (Roy et al. 2007).

6.5 Genetic Manipulation of *D. papillatum*

The recently recognized diversity and abundance of diplonemids (Flegontova et al. 2016; Gawryluk et al. 2016) makes it mandatory to turn at least one species into a genetically tractable organism. Indeed, in order to understand their biology, interactions, ecology, and more specifically functions of individual proteins, a crucial step is to establish protocols that would allow genetic manipulations of diplonemids. We have started to develop a transformation system of the type species *D. papillatum*, the genome of which is being sequenced (our unpublished data). Even more

importantly, it can be easily cultivated axenically in the laboratory, reaches high cell density, grows in large volumes, and can be cryopreserved.

Nuclear gene expression of *D. papillatum* is similar to that in other euglenozoans. Its genes are transcribed polycistronically, and individual mRNAs are then trans-spliced, with the short spliced-leader (SL) RNA gene being added to the 5′ end of each transcript. On one hand, the 39-nt-long SL RNA of *D. papillatum* is quite conserved at the sequence level even in the planktonic diplonemids from the DSPD clade. On the other hand, the situation seems much more complex when it comes to nuclear spliceosomal introns, as the nuclear DNA of the DSPD species displays a high density of noncanonical introns that await further characterization (Gawryluk et al. 2016). The genome and transcriptome of *D. papillatum* have been sequenced, and their assembly and annotation are under way (our unpublished data). Knowing the full set of genes will be essential not only for turning this diplonemid into a model species but also for our understanding of its metabolism and other features.

The first obvious task is to get foreign DNA into the *D. papillatum* cells. To ensure stable integration, several crucial steps have to be fulfilled. One is to find resistance markers that can be used for selection of transformants. In the next step, optimal transformation conditions and strategy have to be designed. Last but not least, constructs have to be obtained that will not only stably integrate into the genome, but even more importantly, allow expression, including transcription, posttranscriptional processing and modifications, so that the ensuing transcripts can be finally translated on cytosolic ribosomes. We have accomplished all these steps (Kaur et al. 2018), although efficiency is still moderate and requires optimization.

More specifically, so far we have found seven selection markers to which *D. papillatum* is sensitive. Using available genomic data, we have selected genes that are suited for replacement, namely, those that are nonessential are highly expressed and contain 5′ and 3′ untranslated regions (UTRs) longer than 100 nucleotides. Moreover, we have established a protocol for DNA uptake in a reproducible fashion and have created linear constructs bearing fluorescent protein and selection marker flanked by diplonemid 5′ and 3′ UTRs. We have also confirmed stable incorporation of foreign DNA into the *D. papillatum* genome and have evidence that both the fluorescence gene and the resistance marker on the electroporated constructs are transcribed. Sequencing results showed that the SL RNA sequence is trans-spliced to the 5′ end of the corresponding transcripts. The antibiotic resistance of selected clones provides indirect evidence that the integrated genes are translated (Kaur et al. 2018).

In principle, homologous recombination should be possible, since the genes involved in the corresponding machinery are present in the *D. papillatum* genome, but so far, the inserted DNA has failed to integrate into the target locus. We believe that this can be remedied by further extension of the 5′ and 3′ homologous regions of the constructs. We also plan to use the CRISPR/Cas9-based approach to achieve proper integration of the introduced genes. Attempts to maintain circular plasmids as non-integrated episomes, or to transform the cells with a virus vector carrying green fluorescent protein, were not successful (our unpublished data).

These preliminary data allow us to state that *D. papillatum* can be transformed and has a solid potential to become a genetically tractable organism. Once a robust, reproducible transfection protocol for gene replacement and tagging has been established in *D. papillatum,* we plan to apply the procedure to other diplonemid species—key to understanding the biology of the group as a whole. For the time being, with a representative of the species-rich DSPD clade yet to be brought into culture, the next candidate for transformation is *H. phaeocysticola*. However, this species is much more challenging to work with, as it prefers live diatoms as a food source and reaches only low cell densities. Moreover, in contrast to *D. papillatum*, *H. phaeocysticola* can apparently not be cryopreserved (our unpublished data).

6.6 Conclusions and Outlook

Within the last couple of years, diplonemids have emerged from obscurity as one of the most diverse groups of marine eukaryotes. They are also among the half dozen most abundant eukaryotes. Since their cell numbers seem to expand with depth, one can expect that the importance of diplonemids for the marine ecosystem is widely underappreciated.

Two steps are key for further exploration of these fascinating and ecologically highly relevant protists: (1) complete genome and transcriptome sequences from a broad range of diplonemid species have to become available, and (2) diplonemid species must become amenable to reverse genetic methods, allowing stable integration, transcription, and translation of introduced genes. Given the steadily growing interest in diplonemids, we are optimistic on both accounts.

Acknowledgments We thank Daria Tashyreva, Galina Prokopchuk, and Anzhelika Butenko (Institute of Parasitology) for sharing unpublished data. Support from the Grant Agency of University of South Bohemia (050/2016/P to BK), the Czech Grant Agency (15-21974S and 16-18699S to JL), the ERC CZ (LL1601 to JL), the Canadian Institutes of Health Research (CIHR, MOP 70309 to GB), and from the Gordon and Betty Moore Foundation (GBMF4983.01 to GB and JL) is kindly acknowledged.

References

Abad MG, Long Y, Willcox A, Gott JM, Gray MW, Jackman JE (2011) A role for tRNA(his) guanylyltransferase (Thg1)-like proteins from *Dictyostelium discoideum* in mitochondrial 5′-tRNA editing. RNA 17:613–623

Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, Le Gall L, Lynn DH, McManus H, Mitchell EA, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick L, Schoch CL, Smirnov A, Spiegel FW (2012) The revised classification of eukaryotes. J Eukaryot Microbiol 59:429–493

Adler BK, Harris ME, Bertrand KI, Hajduk SL (1991) Modification of *Trypanosoma brucei* mitochondrial rRNA by posttranscriptional 3′ polyuridine tail formation. Mol Cell Biol 11:5878–5884

Alfonzo JD, Söll D (2009) Mitochondrial tRNA import—the challenge to understand has just begun. Biol Chem 390:717–722

Alfonzo JD, Thiemann O, Simpson L (1997) The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria. Nucleic Acids Res 25:3571–3579

Ammerman ML, Downey KM, Hashimi H, Fisk JC, Tomasello DL, Faktorová D, Kafková L, King T, Lukeš J, Read LR (2012) Architecture of the trypanosome RNA editing accessory complex, MRB1. Nucleic Acids Res 40:5637–5650

Aphasizhev R, Aphasizheva I (2011) Uridine insertion/deletion editing in trypanosomes: a playground for RNA-guided information transfer. WIREs 2:669–685

Ban T, Zhu JK, Melcher K, Xu HE (2015) Structural mechanisms of RNA recognition: sequence-specific and non-specific RNA-binding proteins and the Cas9-RNA-DNA complex. Cell Mol Life Sci 72:1045–1058

Benne R, van den Burg J, Brakenhoff JP, Sloof P, van Boom JH, Tromp MC (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell 46:819–826

Breglia SA, Yubuki N, Hoppenrath M, Leander BS (2010) Ultrastructure and molecular phylogenetic position of a novel euglenozoan with extrusive episymbiotic bacteria: *Bihospites bacati* n. Gen. Et sp. (Symbiontida). BMC Microbiol 10:145

Bundschuh R, Altmüller J, Becker C, Nürnberg P, Gott JM (2011) Complete characterization of the edited transcriptome of the mitochondrion of *Physarum polycephalum* using deep sequencing of RNA. Nucleic Acids Res 39:6044–6055

Burger G, Gray MW, Lang BF (2003) Mitochondrial genomes: anything goes. Trends Genet 19:709–716

Burger G, Gray MW, Forget L, Lang BF (2013) Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol Evol 5:418–438

Burger G, Moreira S, Valach M (2016) Genes in hiding. Trends Genet 32:553–565

Calvo SE, Clauser KR, Mootha VK (2016) MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res 44:D1251–D1257

Cavalier-Smith T (2016) Higher classification and phylogeny of Euglenozoa. Eur J Protistol 56:250–276

Chaput H, Wang Y, Morse D (2002) Polyadenylated transcripts containing random gene fragments are expressed in dinoflagellate mitochondria. Protist 153:111–122

Chen C, Frankhouser D, Bundschuh R (2012) Comparison of insertional RNA editing in *Myxomycetes*. PLoS Comput Biol 8:e1002400

de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury J-M, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horák A, Jaillon O, Lima Mendez G, Lukeš J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Acinas SG, Bork P, Bowler C, Gaill F, Gorsky G, Grimsley N, Hingcamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki M, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E, Boss E, Follows M, Karp-Boss L, Krzic U, Reynaud EG, Sardet C, Sullivan MB, Velayoudon D (2015) Eukaryotic plankton diversity in the sunlit global ocean. Science 348:1261605

Dejung M, Subota I, Bucerius F, Dindar G, Freiwald A, Engstler M, Boshart M, Butter F, Janzen CJ (2016) Quantitative proteomics uncovers novel factors involved in developmental differentiation of *Trypanosoma brucei*. PLoS Pathog 12:e1005439

Desmond E, Brochier-Armanet C, Forterre P, Gribaldo S (2011) On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes. Res Microbiol 162:53–70

Dixit S, Müller-McNicoll M, David V, Zarnack K, Ule J, Hashimi H, Lukeš J (2017) Differential binding of mitochondrial transcripts by MRB8170 and MRB4160 regulates distinct editing fates of mitochondrial mRNA in trypanosomes. MBio 8:e02288-16

Dobáková E, Flegontov P, Skalický T, Lukeš J (2015) Unexpectedly streamlined mitochondrial genome of the euglenozoan *Euglena gracilis*. Genome Biol Evol 7:3358–3367

Edgcomb VP, Breglia SA, Yubuki N, Beaudoin D, Patterson DJ, Leander BS, Bernhard JM (2011) Identity of epibiotic bacteria on symbiontid euglenozoans in O2-depleted marine sediments: evidence for symbiont and host co-evolution. ISME J 5:231–243

Eloe EA, Shulse CN, Fadrosh DW, Williamson SJ, Allen EE, Bartlett DH (2011) Compositional differences in particle-associated and free-living microbial assemblages from an extreme deep-ocean environment. Environ Microbiol Rep 3:449–458

El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Hass BJ, Tran A-N, Wortman JR, Alsmark UCM, Angiuoli S, Anupama A, Badger J, Bringaud F, Cadag E, Carlton JM, Cerqueira GC, Creasy T, Delcher AL, Djikeng A, Embley TM, Hauser C, Ivens AC, Kummerfeld SK, Pereira-Leal JB, Nilsoon D, Peterson J, Salzberg SL, Shallom J, Silva JC, Sundaram J, Westenberger S, White O, Melville SE, Donelson JE, Andersson B, Stuart KD, Hall N (2005) Comparative genomics of trypanosomatid parasitic protozoa. Science 309:404–409

Faktorová D, Dobáková E, Peña-Diaz P, Lukeš J (2016) From simple to supercomplex: mitochondrial genomes of euglenozoan protists. F1000 Res 5:392

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

Flegontov P, Michálek J, Janouškovec J, Lai D-H, Jirků M, Hajdušková E, Tomčala A, Otto TD, Keeling PJ, Pain A, Oborník M, Lukeš J (2015) Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites. Mol Biol Evol 32:1115–1131

Flegontova O, Flegontov P, Malviya S, Audic S, Wincker P, de Vargas C, Bowler C, Lukeš J, Horák A (2016) Unexpected diversity and abundance of planktonic diplonemids in the world ocean. Curr Biol 26:3060–3065

Fu C-J, Sheikh S, Miao W, Andersson SGE, Baldauf SL (2014) Missing genes, multiple ORFs, and C-to-U type RNA editing in *Acrasis kona* (Heterolobosea, Discoba) mitochondrial DNA. Genome Biol Evol 6:2240–2257

Gawryluk RMR, Chisholm KA, Pinto DM, Gray MW (2014) Compositional complexity of the mitochondrial proteome of a unicellular eukaryote (*Acanthamoeba castellanii*, supergroup Amoebozoa) rivals that of animals, fungi, and plants. J Proteome 109:400–416

Gawryluk RMR, del Campo J, Okamoto N, Strassert JFH, Lukeš J, Richards TA, Worden AZ, Santoro AE, Keeling PJ (2016) Morphological identification and single-cell genomics of marine diplonemids. Curr Biol 26:3053–3059

Gonzalez TN, Sidrauski C, Dörfler S, Walter P (1999) Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway. EMBO J 18:3119–3132

Göringer HU (2012) 'Gestalt,' composition and function of the *Trypanosoma brucei* editosome. Annu Rev Microbiol 66:65–82

Gott JM, Parimi N, Bundschuh R (2005) Discovery of new genes and deletion editing in *Physarum* mitochondria enabled by a novel algorithm for finding edited mRNAs. Nucleic Acids Res 33:5063–5072

Gray MW (2012) Mitochondrial evolution. Cold Spring Harb Perspect Biol 4:a011403

Hudson AJ, Stark MR, Fast NM, Russell AG, Rader SD (2015) Splicing diversity revealed by reduced spliceosomes in *C. merolae* and other organisms. RNA Biol 12:1–8

Jackman JE, Gott JM, Gray MW (2012) Doing it in reverse: 3′-to-5′ polymerization by the Thg1 superfamily. RNA 18:886–899

Jackson CJ, Waller RF (2013) A widespread and unusual RNA trans-splicing type in dinoflagellate mitochondria. PLoS One 8:e56777

Jackson CJ, Norman JE, Schnare MN, Gray MW, Keeling PJ, Waller RF (2007) Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. BMC Biol 5:41

Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Žárský V, Barlow LD, Herman EK, Soukal P, Hroudová M, Doležal P, Stairs CW, Roger AJ, Eliáš M, Dacks JB, Vlček Č, Hampl V (2016) A eukaryote without a mitochondrial organelle. Curr Biol 26:1274–1284

Kaur B, Valach M, Peña-Diaz P, Moreira S, Keeling PJ, Burger G, Lukeš J, Faktorová D (2018) Transformation of *Diplonema papillatum*, the type species of the highly diverse and abundant marine microeukaryotes Diplonemida (Euglenozoa). Environ Microbiol 20:1030–1040

Kiethega GN, Turcotte M, Burger G (2011) Evolutionary conserved *cox1* trans-splicing without cis-motifs. Mol Biol Evol 28:2425–2458

Kiethega GN, Yan Y, Turcotte M, Burger G (2013) RNA-level unscrambling of fragmented genes in *Diplonema* mitochondria. RNA Biol 10:301–313

Knoop V (2010) When you can't trust the DNA: RNA editing changes transcript sequences. Cell Mol Life Sci 68:567–586

Kuai L, Fang F, Butler JS, Sherman F (2004) Polyadenylation of rRNA in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 101:8581–8586

Laforest MJ, Roewer I, Lang BF (1997) Mitochondrial tRNAs in the lower fungus *Spizellomyces punctatus*: tRNA editing and UAG 'stop' codons recognized as leucine. Nucleic Acids Res 25:626–632

Lara E, Moreira D, Vereshchaka A, López-García P (2009) Panoceanic distribution of new highly diverse clades of deep-sea diplonemids. Environ Microbiol 11:47–55

Lavrov DV, Adamski M, Chevaldonné P, Adamska M (2016) Extensive mitochondrial mRNA editing and unusual mitochondrial genome organization in calcaronean sponges. Curr Biol 26:86–92

Leander BS, Triemer RE, Farmer MA (2001) Character evolution in heterotrophic euglenids. Eur J Protistol 37:337–356

Lee WJ, Simpson AGB (2014) Morphological and molecular characterisation of *Notosolenus urceolatus* Larsen and Patterson 1990, a member of an understudied deepbranching euglenid group (petalomonads). J Eukaryot Microbiol 61:463–479

Leigh J, Lang BF (2004) Mitochondrial 3′ tRNA editing in the jakobid *Seculamonas ecuadoriensis*: a novel mechanism and implications for tRNA processing. RNA 10:615–621

Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods 7:709–715

Lin S, Zhang H, Spencer DF, Norman JE, Gray MW (2002) Widespread and extensive editing of mitochondrial mRNAs in dinoflagellates. J Mol Biol 320:727–739

Lithgow T, Schneider A (2010) Evolution of macromolecular import pathways in mitochondria, hydrogenosomes and mitosomes. Philos Trans R Soc B Sci 365:799–817

López-García P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409:603–607

López-García P, Vereshchaka A, Moreira D (2007) Eukaryotic diversity associated with carbonates and fluid seawater interface in Lost City hydrothermal field. Environ Microbiol 9:546–554

Lukeš J, Archibald JM, Keeling PJ, Doolittle WF, Gray MW (2011) How a neutral evolutionary ratchet can build cellular complexity. IUBMB Life 63:528–537

Lukeš J, Flegontova O, Horák A (2015) Diplonemids. Curr Biol 25:R702–R704

Maguire F, Richards TA (2014) Organelle evolution: a mosaic of 'mitochondrial' functions. Curr Biol 24:R518–R520

Mahendran R, Spottswood MR, Miller DL (1991) RNA editing by cytidine insertion in mitochondria of *Physarum polycephalum*. Nature 349:434–438

Manna S (2015) An overview of pentatricopeptide repeat proteins and their applications. Biochimie 113:93–99

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
Maslov DA, Yasuhira S, Simpson L (1999) Phylogenetic affinities of *Diplonema* within the Euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences. Protist 150:33–42
Mohanty BK, Kushner SR (2011) Bacterial/archaeal/organellar polyadenylation. WIREs 2:256–276
Moreira S, Breton S, Burger G (2012) Unscrambling of genetic information at the RNA level. WIREs 3:213–228
Moreira S, Valach M, Aoulad-Aissa M, Otto C, Burger G (2016) Novel modes of RNA editing in mitochondria. Nucleic Acids Res 44:4907–4919
Nishikura K (2016) A-to-I editing of coding and non-coding RNAs by ADARs. Nat Rev Mol Cell Biol 17:83–96
Okamoto N, Gawryluk RMR, del Campo J, Strassert JFH, Lukeš J, Richards TA, Worden AZ, Santoro AE, Keeling PJ (2018) *Eupelagonema oceanica* n. gen. & sp. and a revised diplonemid taxonomy. J Eukaryot Microbiol (in press)
Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 12:87–98
Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res 37:e123
Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, Bowser SS, Čepička I, Decelle J, Dunthorn M, Fiore-Donno A-M, Gile HG, Holzmann M, Jahn R, Jirků M, Keeling PJ, Kostka M, Kudryavtsev A, Lara E, Lukeš J, Mann GD, Mitchell ADE, Nitsche F, Romeralo M, Saunders WG, Simpson AGB, Smirnov VA, Spouge J, Stern FR, Stoeck T, Zimmermann J, Schindel D, de Vargas C (2012) CBOL Protist working group: barcoding eukaryotic richness beyond the animal, plant and fungal kingdoms. PLoS Biol 10:e1001419
Popow J, Schleiffer A, Martinez J (2012) Diversity and roles of (t)RNA ligases. Cell Mol Life Sci 69:2657–2670
Rammelt C, Rossmanith W (2016) Repairing tRNA termini: news from the 3′ end. RNA Biol 13:1182–1188
Read LK, Lukeš J, Hashimi H (2016) Trypanosome RNA editing: the complexity of getting U in and taking U out. WIREs 7:33–51
Roy J, Faktorová D, Lukeš J, Burger G (2007) Unusual mitochondrial genome structures throughout the Euglenozoa. Protist 158:385–396
Rubio MAT, Pastar I, Gaston KW, Ragone FL, Janzen CJ, Cross GAM, Papavasiliou FN, Alfonzo JD (2007) An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. Proc Natl Acad Sci U S A 104:7821–7826
Rüdinger M, Fritz-Laylin L, Polsakiewicz M, Knoop V (2011) Plant-type mitochondrial RNA editing in the protist *Naegleria gruberi*. RNA 17:2058–2062
Salone V, Rüdinger M, Polsakiewicz M, Hoffmann B, Groth-Malonek M, Szurek B, Small I, Knoop V, Lurin C (2007) A hypothesis on the identification of the editing enzyme in plant organelles. FEBS Lett 581:4132–4138
Schnepf E (1994) Light and electron microscopical observations in *Rhynchopus coscinodiscivorus* spec. Nov., a colorless, phagotrophic Euglenozoon with concealed flagella. Arch Protistenkd 144:63–74
Segovia R, Pett W, Trewick S, Lavrov DV (2011) Extensive and evolutionarily persistent mitochondrial tRNA editing in velvet worms (phylum Onychophora). Mol Biol Evol 28:2873–2881
Shapiro TA, Englund PT (1995) The structure and replication of kinetoplast DNA. Annu Rev Microbiol 49:117–143
Shikanai T (2015) RNA editing in plants: machinery and flexibility of site recognition. Biochim Biophys Acta 1847:779–785

Škodová-Sveráková I, Verner Z, Skalický T, Votýpka J, Horváth A, Lukeš J (2015) Lineage-specific activities of a multipotent mitochondrion of trypanosomatid flagellates. Mol Microbiol 96:55–67

Slomovic S, Fremder E, Staals RH, Pruijn GJ, Schuster G (2010) Addition of poly(a) and poly(a)-rich tails during RNA degradation in the cytoplasm of human cells. Proc Natl Acad Sci U S A 107:7407–7412

Smith DR, Keeling PJ (2015) Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. Proc Natl Acad Sci U S A 112:10177–10184

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

Stahley MR, Strobel SA (2006) RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. Curr Opin Struct Biol 16:319–326

Stuart K, Feagin JE (1992) Mitochondrial DNA of kinetoplastids. Int Rev Cytol 141:65–88

Sun T, Bentolila S, Hanson MR (2016) The unexpected diversity of plant organelle RNA editosomes. Trends Plant Sci 21:962–973

Szklarczyk R, Huynen MA (2010) Mosaic origin of the mitochondrial proteome. Proteomics 10:4012–4024

Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A (2013) RNA editing in plants and its evolution. Annu Rev Genet 47:335–352

Tanaka N, Meineke B, Shuman S (2011) RtcB, a novel RNA ligase, can catalyze tRNA splicing and HAC1 mRNA splicing *in vivo*. J Biol Chem 286:30253–30257

Tashyreva D, Prokopchuk G, Yabuki A, Kaur B, Faktorová D, Votýpka J, Kusaka C, Fujikura K, Shiratori T, Ishida K, Horák A, Lukeš J (2018) Phylogeny and morphology of diplonemids from the Sea of Japan. Protist 169:158–179

Tielens AGM, van Hellemond JJ (2009) Surprising variety in energy metabolism within Trypanosomatidae. Trends Parasitol 25:482–490

Triemer RE, Ott DW (1990) Ultrastructure of *Diplonema ambulator* Larsen & Patterson (Euglenozoa) and its relationship to *Isonema*. Eur J Protistol 25:316–320

Urbaniak MD, Martin DM, Ferguson MA (2013) Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of *Trypanosoma brucei*. J Proteome Res 12:2233–2244

Valach M, Moreira S, Kiethega GN, Burger G (2014) Trans splicing and RNA editing of LSU rRNA in *Diplonema* mitochondria. Nucleic Acids Res 42:2660–2672

Valach M, Moreira S, Faktorová D, Lukeš J, Burger G (2016) Post-transcriptional mending of gene sequences: looking under the hood of mitochondrial gene expression in diplonemids. RNA Biol 13:1204–1211

Valach M, Moreira S, Hoffmann S, Stadler PF, Burger G (2017) Keeping it complicated: mitochondrial genome plasticity in diplonemids. Sci Rep 7:14166

Vanfleteren JR, Vierstraete AR (1999) Insertional RNA editing in metazoan mitochondria: the cytochrome b gene in the nematode *Teratocephalus lirellus*. RNA 5:622–624

Verner Z, Basu S, Benz C, Dixit S, Dobáková E, Faktorová D, Hashimi H, Horáková E, Huang Z, Paris Z, Pena-Diaz P, Ridlon L, Týč J, Wildridge D, Zíková A, Lukeš J (2015) Malleable mitochondrion of *Trypanosoma brucei*. Int Rev Cell Mol Biol 315:73–151

Visomirski-Robic LM, Gott JM (1997) Insertional editing of nascent mitochondrial RNAs in *Physarum*. Proc Natl Acad Sci U S A 94:4324–4329

Vlcek Č, Marande W, Teijeiro S, Lukeš J, Burger G (2011) Gene fragments scattered across a multipartite mitochondrial genome. Nucleic Acids Res 39:979–988

Wahlstedt H, Ohman M (2011) Site-selective versus promiscuous A-to-I editing. WIREs 2:761–771

Wheeler RJ, Gull K, Gluenz E (2012) Detailed interrogation of trypanosome cell biology via differential organelle staining and automated image analysis. BMC Biol 10:1

Yabuki A, Tame A (2015) Phylogeny and reclassification of *Hemistasia phaeocysticola* (Scherffel) Elbrachter & Schnepf, 1996. J Eukaryot Microbiol 62:426–429

Yabuki A, Tanifuji G, Kusaka C, Takishita K, Fujikura K (2016) Hyper-eccentric structural genes in the mitochondrial genome of the algal parasite *Hemistasia phaeocysticola*. Genome Biol Evol:8, 2870–2878

Yang J, Harding T, Kamikawa R, Simpson AGB, Roger AJ (2017) Mitochondrial genome evolution and a novel RNA editing system in deep-branching heteroloboseids. Genome Biol Evol 9:1161–1174

Yokobori S, Pääbo S (1995) Transfer RNA editing in land snail mitochondria. Proc Natl Acad Sci U S A 92:10432–10435

Yubuki N, Edgcomb VP, Bernhard JM, Leander BS (2009) Ultrastructure and molecular phylogeny of *Calkinsia aureus*: cellular identity of a novel clade of deep-sea euglenozoans with epibiotic bacteria. BMC Microbiol 27:16

Yubuki N, Simpson AG, Leander BS (2013) Reconstruction of the feeding apparatus in *Postgaardi mariagerensis* provides evidence for character evolution within the Symbiontida (Euglenozoa). Eur J Protistol 49:32–39

Zhao C, Pyle AM (2017) Structural insights into the mechanism of group II intron splicing. Trends Biochem Sci 42:470–482

Zíková A, Verner Z, Nenarokova A, Michels PAM, Lukeš J (2017) A paradigm shift: the mitoproteomes of procyclic and bloodstream *Trypanosoma brucei* are comparably complex. PLoS Pathog 13:e1006679

Zimorski V, Ku C, Martin W, Gould SB (2014) Endosymbiotic theory for organelle origins. Curr Opin Microbiol 22:38–48