

School of Doctoral Studies in Biological Sciences
University of South Bohemia in České Budějovice
Faculty of Science

**GENETIC MODIFICATIONS AND FUNCTIONAL
ANALYSES OF DIPLONEMIDS**

Ph.D. Thesis

Binnypreet Kaur

Supervisor: Prof. RNDr. Julius Lukeš, CSc.
Co-Supervisor: RNDr. Drahomíra Faktorová, Ph.D.

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

České Budějovice

2020

This thesis should be cited as: Kaur B., 2020: Genetic modifications and functional analyses of Diplonemids. Ph.D. Thesis Series, No. 3. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, 243 pp.

Annotation

This study is dedicated for the first time for the systematic and multifaceted research in the diplonemids which were overlooked for long time due to technical limitations but recently they were discovered as the most species rich among marine plankton protists. This study describes the phylogeny and morphology of diplonemids based on the 18S rRNA gene sequences. Since, *D. papillatum*, the only species studied in detail so far and the easily cultivable representative, thus we established it as a model organism to study the function of diplonemids genes which will further allow us to shed light on their amazing evolutionary success in contemporary oceans. Further investigation revealed that diplonemids are the record holders in mitochondrial genome content, gene structure and RNA editing.

Declaration

I hereby declare that I did all the work presented in this thesis by myself or in collaboration with co-authors of the presented papers. I properly cite all references and other sources that I used to work up the thesis. Those references and other sources are given in the list of references.

Prohlášení

Prohlašuji, že svoji disertační práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své disertační práce, a to v úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

České Budějovice

Binnypreet Kaur

This thesis originated from a partnership between the Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, supporting doctoral studies in the Molecular and Cell biology and Genetics study program.



Přírodovědecká
fakulta
Faculty
of Science



BIOLOGY
CENTRE
CAS



INSTITUTE OF PARASITOLOGY
Biology Centre CAS

Financial support

This work was supported by Grant agency of the University of South Bohemia in České Budějovice (grant 050/2016/P and 094/2018/P to Binnypreet Kaur), the Czech Grant Agency (15–21974S and 16–18699S to Julius Lukeš), the ERC CZ grant (LL1601 to J.L.), the Czech Ministry of Education (ERD Funds OPVVV16_019/ 0000759 to J.L.). This work was also supported by the Gordon and Betty Moore Foundation (grant GBMF-4983.01 to J.L. and G.B.), the Canadian Institute of Health Research (CIHR, MOP70309 to GB), the Natural Sciences and Engineering Research Council of Canada (RGPIN-2014-05286 and RGPIN-2019-04024 to G.B.).

Acknowledgments

I would like to sincerely thank my supervisor, Prof. Julius Lukeš, Ph.D. for his immense faith, constant encouragement and his extensive support throughout my studies. Without his expertise in the field of protists in general, my research and writing of this thesis would not be possible. It was my great pleasure to stay in the scientifically ecstatic atmosphere set by him.

I would also like to thank my co-supervisor, Dr. Drahomíra Faktorová, Ph.D. for her valuable time, advice and easy communication so that we could generate and analyze the diplomema data and to present my research at conferences.

Besides them, I also thank Prof. Gertraud Burger and Dr. Matus Valach Ph.D., for sharing their knowledge, for giving me the opportunity to improve my skills in their excellent lab. and for their helpful collaboration.

Here, I would also like to acknowledge my husband Samarjeet Singh, who motivated me to start my Ph.D., who always stands next to me with all my ups and downs and support me extensively throughout my studies and helped me a lot with organizing both work and home.

I would also like to thank all the people from Jula's lab, both former and present lab members, for their help in my work, for creating a friendly atmosphere in the Institute. I am also thankful for finding friends in my colleagues Daria Tashyreva, Ph.D. and Michael John Hammond, Ph.D. for enlightening discussions and valuable advice during the writing of my thesis.

I am also grateful to my friends who have supported me and made my life happy, easy and unforgettable in Ceske Budejovice.

Finally, I would like to thank my family, my parents and my lovely brother for their love, trust and support throughout my life. They are my true inspiration and this thesis would not be possible without them.

List of papers and author's contribution

The thesis is based on the following papers (listed chronologically):

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

<https://doi.org/10.1016/j.protis.2018.02.001>

Binnypreet Kaur participated in cell cultivation and light and electron microscopy observations

2. **Kaur B.**, Valach M., Pena Diaz C., Moreira S., Keeling P., Burger G., Lukeš J., Faktorová D. (2018) Transformation of *Diplonema papillatum*, the type species of the highly diverse and abundant marine micro-eukaryotes Diplonemida (Euglenozoa), *Environmental Microbiology* 20: 1030-1040. doi: 10.1111/1462-2920.14041.

Binnypreet Kaur participated in cell cultivation, transfection, microscopy, produced transgenic cell lines, growth rate measurements, Southern blotting, Western blotting and data analysis.

3. Faktorová D., Valach M., **Kaur B.**, Burger G., Lukeš J. (2018) Mitochondrial RNA editing and processing in diplomemid protists. *RNA Metabolism in Mitochondria. Nucleic Acids and Molecular Biology*; 34: 145–176, Cham Springer. https://doi.org/10.1007/978-3-319-78190-7_6.

Binnypreet Kaur prepared figures and contributed to writing of the chapter.

4. **Kaur B.**, Záhonová K., Valach M., Faktorová D., Prokopchuk G., Burger G., Lukeš J. (2020) Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplomemids. *Nucleic Acids Res.* 10. <https://doi.org/10.1093/nar/gkz1215>.

Binnypreet Kaur participated in cell cultivation, nucleic acid extraction, data curation, investigation and contributed to writing of the manuscript.

5. Faktorová D., Nisbet R.E.R., Fernández Robledo J.A., Casacuberta E., Sudek L., Allen A.E., Ares M. Jr., Aresté C., Balestreri C., Barbrook A.C., Beardslee P., Bender S., Booth D.S., Bouget F.-Y., Bowler C., Breglia S.A., Brownlee C., Burger G., Cerutti H., Cesaroni R., Chiurillo M.A., Clemente T., Coles D.B., Collier J.L., Cooney E.C., Coyne K., Docampo R., Dupont C.L., Edgcomb V., Einarsson E., Elustondo P.A., Federici F., Freire-Beneitez V., Freuria N.J., Fukuda K., García P.A., Girguis P.R., Gomaa F., Gornik S.G., Guo J., Hampl V., Hanawa Y., Haro-Contreras E.R., Hehenberger E., Highfield A., Hidakawa Y., Hopes A., Howe C.J., Hu I., Ibañez J., Irwin N.A.T., Ishii Y., Janowicz N.E., Jones A.C., Kachale A., Fujimura-Kamada K., **Kaur B.**, Kaye J.Z., Kazana E., Keeling P.J., King N., Klobutcher L.A., Lander N., Lassadi I., Li Z., Lin S., Lozano J.C., Luan F., Maruyama S., Matute T., Miceli C., Minagawa J., Moosburner M., Najle S.R., Nanjappa M., Nimmo I.C., Noble L., Novák Vanclová A.M.G., Nowacki M., Nuñez I., Pain A., Piersanti

A., Pucciarelli S., Pyrih J., Rest J.S., Rius M., Robertson D., Ruaud A., Ruiz-Trillo I., Sigg M.A., Silver P.A., Slamovits C.H., Smith G.J., Sprecher B.N., Stern R., Swart E., Tsaousis A., Tsy-pin L., Turkewitz A., Turnšek J., Valach M., Vergé V., von Dassow P., von der Haar T., Waller R.F., Wang L., Wen X., Wheeler G., Woods A., Zhang H., Mock T., Worden A.Z. & Lukeš J. (2020) Genetic tool development in marine protists: Emerging model organisms for experimental cell biology. (Nature Methods- resubmission).

bioRxiv preprint first posted online Aug. 1, 2019; doi: <http://dx.doi.org/10.1101/718239>.

Binnypreet Kaur performed cell cultivation of D. papillatum, designed tagging construct, produced transgenic cell lines, growth rate measurements, Western blotting, data analysis and contributed to writing of the D. papillatum part of the manuscript.

6. Faktorová D., **Kaur B.**, Graf L., Benz C., Valach M., Burger G., Lukeš J., Targeted integration by homologous recombination enables *in-situ* tagging and replacement of genes in marine micro eukaryote *Diplonema papillatum* (MS in preparation).

Binnypreet Kaur performed cell cultivation, designed tagging construct, produced transgenic cell lines, growth rate measurements, Western blotting, data analysis and contributed to writing.

Co-author agreement

Prof. RNDr. Julius Lukeš, the supervisor of this Ph.D. thesis and co-author of all presented papers, fully acknowledges the contribution of Binnypreet Kaur

Prof. RNDr. Julius Lukeš

Table of Contents

Abbreviations.....	1
Summary.....	2
Aims of the study.....	6
1. Introduction	
1.1 General Overview of diplomemids.....	7
1.2 Diplomemids: From insignificance to prominence.....	8
2. Genetically tractable system	
2.1 Selection of model organism.....	13
2.2 Transformation methods.....	16
2.3 CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats).....	19
2.4 RNA interference (RNAi).....	23
3. Nuclear genome, polycistronic transcription and mRNA splicing in diplomemids.....	24
4. Mitochondrial genome architecture and gene expression	
4.1 RNA editing pattern across diplomemids.....	33
4.2 Prediction of RNA editing machinery in diplomemids.....	34
5. References	36
6. Chapter 1: Phylogeny and morphology of new diplomemids from Japan.....	58
7. Chapter 2: Transformation of <i>Diplonema papillatum</i> , the type species of the highly diverse and abundant marine micro-eukaryotes Diplonemida (Euglenozoa).....	81
8. Chapter 3: Mitochondrial RNA editing and processing in diplomemid protists.....	93
9. Chapter 4: Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplomemids.....	126
10. Chapter 5: Genetic tool development in marine protists: Emerging model organisms for experimental cell biology.....	142
11. Chapter 6: Targeted integration by homologous recombination enables <i>in-situ</i> tagging and replacement of genes in the marine microeukaryote <i>Diplonema papillatum</i>	194
12. Conclusions.....	234
13. Curriculum vitae.....	236

ABBREVIATIONS

ATCC	American type culture collection
DSPD	deep-sea pelagic diplomonids
HR	homologous recombination
NHEJ	non-homologous end joining
dsDNA	double-stranded DNA
OTUs	operational taxonomy Units
CRISPR	clustered regularly interspaced short palindromic repeats
SL	spliced leader
DSB	double strand break
sgRNA	single stranded guide RNA
RNAi	RNA interference
RISC	RNA-Induced Silencing Complex
mt	mitochondrial
MURF	mitochondrial unidentified reading frame
ORF	open reading frame
PPR	pentatricopeptide repeat
RECC	RNA editing core complex
SSU	small subunit of ribosome
LSU	large subunit of ribosome
nt	nucleotide
A	adenine
C	cytosine
T	thymidine
U	uridine
UTR	untranslated region
ADAR	adenosine deaminase
TUTase 2	terminal uridylyltransferase 2

SUMMARY

Diplonemids are a sparsely studied group of heterotrophic marine protists. They belong to the phylum Euglenozoa, and represent a sister group of the well-studied free-living euglenids (Euglenida), as well as highly pathogenic trypanosomatids and free-living bodonids (Kinetoplastea). For a long time, diplonemids were considered a small and rare group of flagellates with only two known genera and less than a dozen species (Massana 2011; Simpson 1997; Vickerman 2000; von der Heyden *et al.* 2004). However, this group has attracted significant interest after the recent holistic assessment of the eukaryotic planktonic diversity in samples collected by the *Tara* Oceans expedition, with the dataset composed of ~800 million V9 18S rRNA barcodes. Unexpectedly, its analysis revealed that diplonemids are the most diverse marine eukaryotes, surpassing metazoans and dinoflagellates, and the 6th most abundant eukaryotic taxon populating the world oceans (de Vargas *et al.* 2015; Lukeš *et al.* 2015; Flegontova *et al.* 2016). Additional studies suggest that they are even more abundant in the deeper meso- and bathypelagic zones (our unpubl. data). Despite this huge diversity, only a handful of diplonemid species have been formally described (Simpson 1997; Maslov *et al.* 1999; Porter 1973; Busse and Preisfeld 2002; Montegut-Felkner and Triemer 1996; Triemer and Ott 1990; von der Heyden *et al.* 2004; Roy *et al.* 2007; David and Archibald 2016). However, all the described species belong to the clades rarely found in the available marine samples, while only a limited number of sequence data and no morphology was available for the representatives of the diplonemid group Eupelagonemidae (Gawryluk *et al.* 2016). Indeed this major clade remains virtually unknown yet it represents up to 97% of

all members of this group in the ocean (Flegontova *et al.* 2016; Okamoto *et al.* 2019).

The only diplonemid feature studied in some detail is the unique and complex mitochondrial genome. In their single reticulated organelle, these flagellates harbor a huge amount of mitochondrial DNA (Lukes *et al.* 2018). The mitochondrial genome is composed of thousands of circular non-catenated molecules, which carry fragmented genes, the transcripts of which have to be *trans*-spliced (Marande *et al.* 2007; Valach *et al.* 2016; Faktorová *et al.* 2018). Almost all the morphological, genomic and physiological information available about diplonemids has been obtained from the studies of a single species, *Diplonema papillatum* (e.g. Marande *et al.* 2005; Kiethega *et al.* 2011; Maslov *et al.* 1999; Valach *et al.* 2016; Valach *et al.* 2017). This is due to its capacity to tolerate cryopreservation and grow in an affordable medium to high densities, and also to the fact that no other species are available in culture collection. Moreover, together with available limited genomic and transcriptomic data, *D. papillatum* represents a suitable model for genetic manipulations and hence functional studies.

In the first chapter of this thesis, we report the discovery of new diplonemid species from surface marine water samples. We have established axenic cultures of these strains, describe their life cycles under the cultivation conditions and have performed their phylogenetic analysis based on the 18S rRNA gene sequences. We have also tried to comprehensively describe their morphology using light and electron microscopy. In total, we described five new diplonemid species and

established three new genera (**Chapter 1, Tashyreva *et al.* 2018a, *Protist* 169: 158–179**).

The second chapter of this thesis demonstrates a successful genetic transformation of *D. papillatum*. This is the first documented case of transformation in a euglenozoan protist outside the well-studied kinetoplastids. We showed that *D. papillatum* is sensitive to multiple drugs that can be used as selectable markers, we have established conditions for efficient electroporation, and demonstrated that electroporated DNA constructs can be stably integrated in its nuclear genome. Finally, we have confirmed that the heterologous genes are transcribed into mRNA with the spliced leader RNA attached and eventually translated into proteins. Although we achieved for the first time expression of a heterologous gene in diplomonids, our constructs were not integrated into the target position. (**Chapter 2, Kaur *et al.* 2018, *Environ. Microbiol.*, 20, 1030–1040**).

In the third chapter of this thesis, we summarize what is known about the mitochondrial RNA editing and RNA processing in *D. papillatum* and other diplomonids. (**Chapter 3, Faktorová *et al.* 2018, In: M.W. Gray and J. Cruz-Reyes (Eds.), *RNA Metabolism in Mitochondria*, Springer 145–176**).

In the fourth chapter, in a comparative manner we reveal a broader picture of the mitochondrial genome architecture, gene content, and post-transcriptional processing of six recently isolated, phylogenetically diverse diplomonid species. We document the most complex and extensive pattern of RNA editing in the hemistasiid clade,

which became the record holder in this respect for organellar transcriptomes (**Chapter 4, Kaur *et al.* 2020, *Nucleic Acids Res.* 10, <https://doi.org/10.1093/nar/gkz1215>**).

In the fifth and sixth chapters, we show the establishment of successful endogenous gene tagging and replacement techniques using homologous recombination in *D. papillatum* and provide evidence of stable transformants. Genetic modification of *D. papillatum* can now be used as a standard reverse genetics tool for the discovery of gene function. The aim of this work was a correct integration of the cassette, which was achieved by the extension of the 5' and 3' homologous regions of the constructs over 1.5 kb. Thus, we established this diplonemid as a promising model organism (**Chapter 5, Faktorova *et al.*, *Nature Methods*, in resubmission; Chapter 6, Manuscript in preparation**).

AIMS OF THE STUDY

Although diplomonads quite unexpectedly emerged as extremely diverse marine protists, so far only marginal information about them is available. Their lifestyle (free-living vs commensal vs parasitic) and preferences for abiotic factors (temperature, light, pressure, and oxygen concentration) are yet to be studied. Along with their abundance and diversity, diplomonads are of interest because of their bizarre mitochondrial genome and expression as well as unorthodox ultrastructural features and presence of a range of endosymbionts.

The main aims of this study were to:

- i) characterize the newly isolated strains and develop stable cultures; describe their life cycle and morphology, and replicate their development in the laboratory
- ii) establish at least one representative species as a genetically tractable system and initiate functional analyses
- iii) characterize the mitochondrial genomes of newly isolated and phylogenetically diverse species

1. INTRODUCTION

1.1. General Overview of diplomemids

Diplonemids are biflagellated heterotrophic protists belonging to the phylum Euglenozoa of the supergroup Excavata, which also contains kinetoplastids, euglenids, and symbiontids (Cavalier-Smith 1993, Adl *et al.* 2019, Burki *et al.* 2019) (Fig. 1). The kinetoplastids are so far the best-studied group composed of small colorless flagellates characterized by a mass of mitochondrial DNA also called the kinetoplast (k) DNA. They are subdivided into highly pathogenic trypanosomatids (such as *Trypanosoma* and *Leishmania*) and free-living bodonids. Mostly aerobic, free-living euglenids contain heterotrophic, mixotrophic and photosynthetic representatives. The best-studied euglenid is *Euglena gracilis*, for which genome and several transcriptomes are available (O'Neill *et al.* 2015; Yoshida *et al.* 2016; Ebenezer *et al.* 2019). We know very little about anaerobic free-living symbiontids represented by only three species and a handful of environmental DNA sequences (Yubuki and Leander *et al.* 2018).

Diplonemids were initially considered a small and rare sister group of medically important kinetoplastids and ecologically relevant euglenids (Moreira *et al.* 2001) and were thus poorly studied. This was reflected by their decades-old subdivision into only two genera - *Diplonema* and *Rhynchopus* (Lara *et al.* 2009) with a formal description of just a handful of species (Lukeš *et al.* 2015). Diplonemids are mostly marine free-living phagotrophs, although some *Rhynchopus* species were proposed to be parasitic (Roy *et al.* 2007). All of the known

diplonemids carry two heterodynamic flagella, a subpellicular corset of microtubules and a single mitochondrion with extremely large cristae (Schnepf 1994; Marande *et al.* 2005; Roy *et al.* 2007).

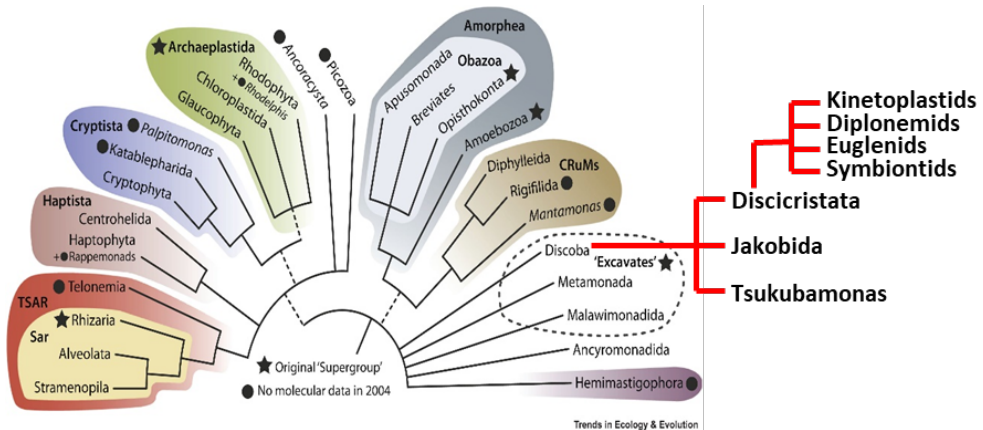


Figure 1. The eukaryotic tree of life. The tree is based on the eukaryotic phylogenomic studies. The supergroup (denoted by star symbol) Excavata indicated by black with broken lines contain phylum Euglenozoa. The taxonomic sub-division of Euglenozoa and has been corroborated and is shown with red arrows (Adapted from Burki *et al.* 2019).

The description of most of their representatives was based exclusively on morphological and ultrastructural observations, with sequence data available for only *D. papillatum*, *D. ambulator*, and *R. euleeides* (Roy *et al.* 2007).

1.2 Diplonemids: From insignificance to prominence

Recently, diplomemids emerged as one of the most abundant eukaryotic protist groups in the world oceans. They are present from the surface to the deep mesopelagic zone (200 to 1,000 meters) and in the deep aphotic

pelagic waters (1,000 to 4,000 meters) (Lara *et al.* 2009; de Vargas *et al.* 2015; Gawryluk *et al.* 2016; Lukeš *et al.* 2015; Flegontova *et al.* 2016). However, they were also detected down to 6,000 m in the poorly studied abyssopelagic zone (Scheckenbach *et al.* 2010; Eloë *et al.* 2011) and are also found in shallow sediments (Larsen and Patterson 1990), as well as in hydrothermal vents (López-García *et al.* 2007). This emergence was triggered by a global metabarcoding survey of marine eukaryotic diversity performed by the *Tara* Oceans expedition across the tropical, temperate and polar regions, which collected samples at 210 sites with the depths from the surface up to 2,000 m. The analysis was based on ~850 million sequences of the V9 motif of the 18S rRNA gene, which revealed that diplomonads constitute the 6th most abundant eukaryotic group in the marine plankton (Flegontova *et al.* 2016). They were present in the photic layer of all initially analyzed 45 worldwide-distributed sampling stations, showing their cosmopolitan distribution, with their abundance increasing with oceanic depth (de Vargas *et al.* 2015; Gawryluk *et al.* 2016; Lukeš *et al.* 2015; Flegontova *et al.* 2016).

Even more surprising is their enormous diversity, substantially exceeding that of the well-known dinoflagellates, diatoms, and metazoans (de Vargas *et al.* 2015; Lukeš *et al.* 2015). With over 45,000 operational taxonomic units (OTUs) identified, diplomonads ranked as the most diverse planktonic eukaryotes, and are thus inevitably important heterotrophic players in the largest ecosystem of our biosphere (Lukeš *et al.* 2015, Flegontova *et al.* 2016; Flegontova *et al.* 2018) (Fig. 2). Despite their significance, their biology and ecological roles remain virtually unknown. Currently, based on the 18S rRNA sequences, diplomonads have been subdivided into four major lineages: (i) 'classical' diplomonads

(Diplonemidae), which include both benthic and planktonic species; (ii) hemistasiids (Hemistasiidae), a small planktonic clade; (iii) an extremely diverse clade of deep-sea pelagic diplonemids - DSPD I, recently named Eupelagonemidae; and (iv) DSPD II- a relatively small clade of deep-sea pelagic diplonemids (Flegontova *et al.*, 2016; Okamoto *et al.*, 2019).

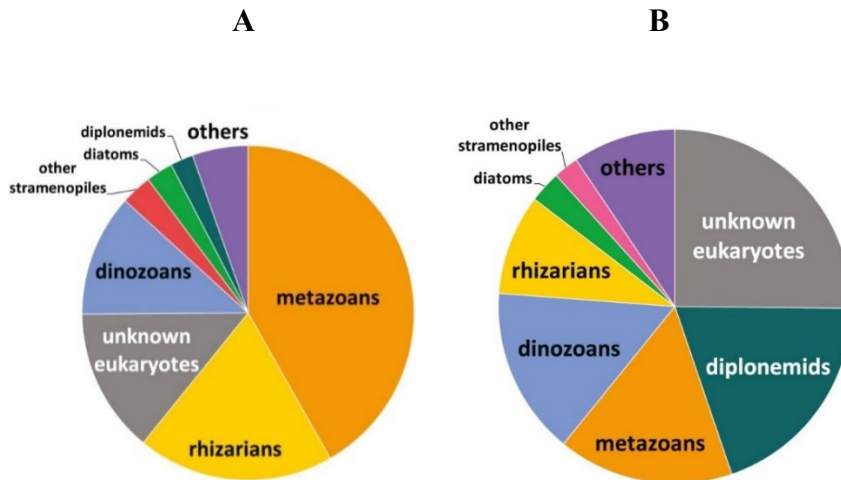


Figure 2. Pie chart showing diplomemids as the (A) 6th most abundant marine eukaryotes according to V9 18S rRNA region of 18S rRNA gene (B) the 1st most diverse group (45,197 swarm based OTUs, or 26% of all eukaryotic OTUs belonging to known clades) (Adapted from Flegontova *et al.* 2016).

Previously, only two diplomemids genera were known -*Diplonema* and *Rhynchopus* (Adl *et al.* 2012). Their cells are oval to sack-shaped and generally variable in size, equipped with a distinct apical papilla and a tubular cytopharynx contiguous with a deep flagellar pocket, both having separate opening (Triemer and Ott 1990; Tashyreva *et al.*, 2018a). A naked cell membrane is supported by a tightly packed corset of microtubules. The cells are highly metabolically active with twisting, gliding or swimming patterns (Tashyreva *et al.*, 2018a). *Diplonema* cells

are motile, bearing two equal size flagella supported with a paraflagellar rod that emerge from the subapical depression, while *Rhynchopus* cells possess two unequal heterodynamic flagella which are concealed in their trophic stages and fully developed in the swimming stages (Von der Heyden *et al.* 2004; Tashyreva *et al.* 2018). Species from both genera carry a single highly reticulated peripheral mitochondrion with lamellar cristae containing a large amount of mitochondrial DNA. Species ranked into the clade Hemistasiidae bear a prominent digestive vacuole, peripheral lacunae and extrusomes. They are predators of diatoms, dinoflagellates, and copepods (Elbrächter *et al.* 1996). Unexpectedly, some diplonemid species exhibit symbioses with prokaryotes. Endosymbiotic relationships have evolved multiple times independently in a number of unrelated groups of protists. In diplonemids, the endosymbionts known so far come from two from different orders, *Rickettsiaceae* and *Holosporaceae* (Tashyreva *et al.* 2018 b; our unpublished data).

The newly described clade Eupelagonemidae represents over 97% of known diplonemid diversity (Okamoto *et al.* 2019) but so far no culturable representative is available, although some morphological information and single-cell amplified genome fragments exist (Gawryluk *et al.* 2016; Okamoto *et al.* 2019). Recently, diplonemids were also found in freshwater lakes Baikal in Russia and Biwa and Ikeda in Japan (Yi *et al.* 2017; Mukherjee *et al.* 2019). Despite this huge diversity and abundance, close to nothing is known about this clearly significant group of marine protists. To change this, our laboratory is trying to turn at least one diplonemid species into a genetically tractable

system and establish it as a model organism to study the function of its genes.

2. Genetically tractable system

In recent decades, the extent of known diversity of marine protists has expanded exponentially with the advances in high throughput sequencing. Systematic analysis of the available wealth of information, represented by millions of genes with unknown function (Carradec *et al.* 2019), has the potential to improve our understanding of the biology of these pivotal members of the marine ecosystem. However, it is the availability of genetically tractable representatives, which will allow us to systematically analyze gene regulation and expression, and protein-protein interactions, shedding light onto their biology. Indeed, there is an urgent need to conduct functional studies by reverse genetic approaches, mainly using gene replacement *via* homologous recombination, RNA interference and CRISPR/Cas approaches. So far, stable and efficient transfections have been achieved in several marine protists with prominent examples including *Phaeodactylum tricorutum*, *Chaetoceros gracilis*, *Parabodo caudatus* and *Thalassiosira pseudonana* (Miyahara *et al.* 2013; Ifuku *et al.* 2015; Gomaa *et al.* 2017; Nymark *et al.* 2016; Stukenberg *et al.* 2018; Hopes *et al.* 2016). In any case, the rather short list of manipulatable marine unicellular eukaryotes has recently been significantly extended (Faktorová *et al.*, in communication). The first step in developing a transgenic organism is the selection of reliable strain that grows fast to high densities and establish reproducible transfection

protocols for it. Model organisms for each taxonomic unit are the key to understand the basic biological processes, metabolism, genes functions and behavior.

2.1 Selection of model organism

A typical model organism is one which is widely studied, its genome and transcriptome are available, has a short generation time and is easy to culture in the laboratory. A few diplomonid species are available from the American Type Culture Collection (ATCC) and are classified based on their 18S rRNA gene sequences, yet they still lack morphological description and sequenced genomes. So far, the best studied is *D. papillatum*, which has been morphologically described and is being sequenced (Burger *et al.* unpublished). Additionally, it has the potential to become a model species, as it grows quickly and relatively easy in artificial seawater, reaches high cell density (6×10^6 cells/ml) with a doubling time of 12 hours and can be easily cryo-preserved (Kaur *et al.* 2018).

D. papillatum was first described in 1973, designated as *Isonema papillatum*, and misplaced within the euglenid group before being isolated from seawater on the surface of eelgrass from New Hampshire, morphologically characterized and made available as ATCC50162 (Porter 1973). Like other diplomonids, *D. papillatum* is colorless, sack-shaped and 5-15 μm in length. The non-rigid cell body capable of pronounced shape changes is covered by a naked plasma membrane, with a prominent anterior papilla and possesses two short flagella inserted sup-

apically into a pronounced pocket (Griessmann 1913; Porter 1973). The two flagella are of different length (5.5 to 7.0 μm), arise at the base of the papilla and generally project anteriorly from the cell. Swimming usually involves spiral, oscillating, or erratic movements, and no encystment formation has been observed so far. The nucleolus persists throughout nuclear division.

Light and electron microscopy revealed a single mitochondrion with an ultrastructure atypical for Euglenozoa (Marande *et al.* 2005). In addition, DNA is evenly distributed throughout the organelle rather than being compacted as in bodonids. The mitochondrial lumen is homogeneously electron transparent and contains few but exceptionally large and irregularly arranged cristae, which are flat, uniformly thick, unbranched, and only rarely attached to the mitochondrial inner membrane. (Marande *et al.* 2005). The morphology of *D. papillatum* is shown in Fig. 3. *D. papillatum*, is known for the unprecedented complexity of its mitochondrial genome which is huge and multipartite with systematically fragmented genes (Marande *et al.* 2005; Marande and Burger 2007; Vlcek *et al.* 2011; Valach *et al.* 2017; Lukeš *et al.* 2018), while its well described transcriptome undergoes massive *trans*-splicing of transcripts and diverse and extensive RNA editing, the molecular mechanisms behind these clearly complex mechanisms remain unknown.

First metabolic studies were also performed on *D. papillatum*, revealing central position of gluconeogenesis in the carbon metabolism. Still, the cells are incapable of performing glycolysis, do not consume

glucose and preferentially use amino acids as their main carbon source, even though the glycolytic enzymes are present (Morales *et al.* 2016).

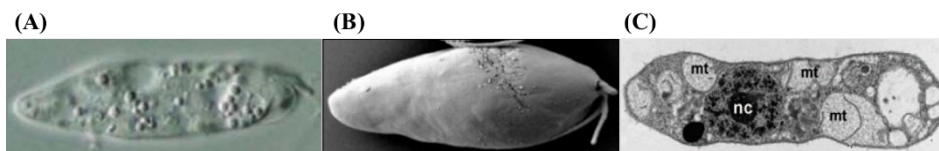


Figure 3. Morphology of *D. papillatum*. (A) light microscopy (B) scanning electron microscopy (C) transmission electron microscopy showing longitudinally section. Prominent flat cristae are visible in the organellar lumen. nc, nucleus; mt, mitochondrion (Adapted from Marande *et al.* 2005).

These enzymes seem to be compartmentalized into peroxisomes, which are modified glycosomes (Makiuchi *et al.* 2011; Morales *et al.* 2016). Another diplomonid species that captured some interest is *Hemistasia phaeocysticola*, which until recently has been the only cultured representative of hemistasiids. It has been formally described based on light microscopy, ultrastructure and molecular data (Elbrächter *et al.* 1996; Yabuki and Tame 2015; Prokopchuk *et al.* 2019). *H. phaeocysticola* belongs to the clade Hemistasiidae, a sister group to the recently established Eupelagonemidae and DSPDII clades (Adl *et al.* 2019; Cavalier-Smith 2016; Yabuki and Tame 2015). *H. phaeocysticola* was originally described as a predator of diatoms, haptophytes and other planktonic organisms (Schnepf, 1994), but can be cultured without prey in media similar to that of *D. papillatum* (Yabuki and Tame 2015). Cells are also biflagellated, pyriform to oblong in shape, 13 to 33 μm in length, with a flexible apical papilla and a spiral groove. Flagella are of unequal length and are approximately 1.5 times the cell length with the posterior flagellum being slightly longer than the anterior one. They are located subapically in the flagellar pocket, which wraps around the body

during feeding. Cells are fast-moving and can rapidly stop and change direction. The nucleus is small and spherical in shape, located at the lower end of the flagellar pocket (Elbrächter *et al.* 1996; Yabuki and Tame 2015; Prokopchuk *et al.* 2019).

Electron microscopy showed that *H. phaeocysticola* also possesses giant flat mitochondrial cristae similar to *D. papillatum* as well as an apical papillum and feeding apparatus vanes with fuzzy coats (Yabuki and Tame 2015). *H. phaeocysticola* is of interest due to the presence of extrusomes, ejectable rod-shaped structures, which have evolved independently several times during eukaryotic evolution and are employed for a variety of functions such as predation, cell invasion, and protection. Indeed, molecular phylogenies showed that *Hemistasia* is the closest known relative of the most diverse eupelagonemids (Gawryluk *et al.* 2016) and its laboratory culture has already been established. Like other diplomonads, *H. phaeocysticola* also possesses a large and highly unusual mitochondrial genome, with twice the degree of gene fragmentation than *D. papillatum* (Yabuki *et al.* 2016).

2.2 Transformation methods

Various transformation methods include viruses (*i.e.* adenoviruses, retroviruses and lentiviruses) and conventional non-viral modalities including electroporation, micro-injections, liposomal reagents and particle bombardment, which have been available for a long time. Out of these, electroporation has been extensively used for the transformation of the sister group of diplomonads, the trypanosomatids (Bellofatto *et al.*

1993; LeBowitz *et al.* 1995; Li *et al.* 2007; Dyer *et al.* 2016). Electroporation techniques can be used to introduce *in-situ* tags, such as enhanced green fluorescent protein and mCherry, to the endogenous loci which can be further monitored by either fluorescent microscopy, antibodies against the tag through western blot analysis or indirect immunofluorescence using anti-tag antibodies. During electroporation, the high voltage electrical pulses are used to permeabilize the cell membranes to transfect foreign DNA into the cells. The foreign DNA becomes integrated into the nuclear genome in a targeted manner using homologous recombination (HR) or randomly through non-homologous end joining (NHEJ). HR is a homology directed, universal and highly conserved double-strand DNA (dsDNA) repair process (Heyer *et al.* 2007, 2008; Prakash *et al.* 2015). It is a sequence induced recombination, in which 5' to 3' homologous strands act as a repair template to fix the double-strand break with the help of recombinases such as Rad51 (Rodgers and McVey 2016).

In *D. papillatum*, genomic data showed the presence of Rad51 (our unpublished data), a key component of HR which coats single-stranded DNA to form nucleoprotein filaments essential for the homology search and strand exchange. Thus, *in-situ* tagging can be achieved using machinery responsible for HR, which in diplonemids is very likely operational. The efficiency of HR recombination also varies between species and is reported to be upregulated by alterations in the Rad51 levels and with the use of enhancers like RS-1 (3-(benzylamino) sulfonyl)-4-bromo-N-(4-bromophenyl) benzamide), which stimulate RAD51 recombinase in organisms with low HR efficiencies (Choo *et al.* 2014, Jayathilaka *et al.* 2008, Song *et al.* 2015).

NHEJ is another double strand repair pathway conserved in all eukaryotes, although the levels of conservation of its components differ (Weller *et al.* 2002; Nenarokova *et al.* 2019). The NHEJ pathway exists into two main types - classical (C-NHEJ) and alternative (A-NHEJ), also named as microhomology-mediated end joining. It was reported recently that C-NHEJ pathway has been independently lost in several parasitic protist lineages (Nenarokova *et al.* 2019). Since it requires no template, it causes the re-joining of incorrect DSB ends or re-joining the resected DNA and thus leads to random integration, which causes various mutations. Despite this, in many eukaryotes NHEJ usually triumphs over HR due to their huge and highly repetitive genomes (Lieber. 2010; Malkova & Haber 2012; Rodgers & McVey 2016).

To solve the problem of random integration due to NHEJ, one solution is to enhance HR by disrupting the essential and conserved components of NHEJ, namely the following five core components: Ku70/Ku80 heterodimer (Ku), DNA-dependent protein kinase catalytic subunit (DNA-PKcs), DNA ligase IV (Lig4), and the XRCC4 and XLF protein (Waters *et al.* 2014; William *et al.* 2014; Her *et al.* 2018; Neal and Meek 2011; Neal *et al.* 2011; Emerson *et al.* 2018). The Ku heterodimer first recognizes and binds to the double-stranded break and serves as a scaffold on which other components of the NHEJ machinery can subsequently assemble and form a stable complex with further sealing performed by Lig4 (Dyanan & Yoo, 1998). Deletion of the Ku genes in *N. crassa* (Ninomiya *et al.* 2004), *K. lactis* (Kooistra *et al.* 2004), *A. nidulans* (Nayak *et al.* 2006) and *C. neoformans* (Goins *et al.* 2006) all resulted in increased targeted integration rates following different transformation approaches. Various inhibitors are available for

different components of NHEJ, such as the inhibition of the DNA-PKcs through the depletion of their cofactor, inositol hexakisphosphate (InsP₆) via the calmodulin antagonists W7 (N-(6-aminohexyl)-5-chloro-1-naphthalenesulfonamide) and chlorpromazine (2-chloro-10-(3-dimethylaminopropyl) phenothiazine hydrochloride) (Byrum *et al.* 2004). Although the benzochromenone 2-(morpholin-4-yl)-benzo[h]chomen-4-one and chromenone 8-dibenzothiophen-4-yl-2-morpholin-4-yl-chromen-4-one are also potent inhibitors of the DNA-PKcs (Durant *et al.* 2003; Leahy *et al.* 2004; Willmore *et al.* 2004), these molecules are toxic and determination of their optimal concentration for a given species is challenging.

2.3 CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats)

CRISPR is an array of alternating short palindromic repeats and spacer sequences that reflects a chronological history of the viruses and plasmids that have invaded bacteria and thus provides an adaptive immunity to prokaryotes (Barrangou *et al.* 2007). It was first observed in the 1980s in *Escherichia coli* (Ishino *et al.* 1987). In bacteria, CRISPR-based immunity involves three critical steps protecting them from invading viruses: adaptation, expression, and interference (van der Oost *et al.* 2009; Makarova *et al.* 2011). In the adaptation step, fragments of foreign DNA are incorporated into the CRISPR locus as new spacers. During expression, the transcription and subsequent processing of the CRISPR locus provide an RNA template for the recognition of

complementary protospacer sequences in the invading DNA. In the final interference step, the invading DNA is cleaved and inactivated by an RNA-guided Cas effector protein (van der Oost *et al.* 2009; Makarova *et al.* 2011). To allow the immune system to distinguish its own DNA from foreign DNA (already incorporated in the CRISPR array), stable binding by the effector protein and subsequent DNA cleavage requires a protospacer motif present only in the targeted foreign DNA. This mechanism has been repurposed for sequence-specific genome editing by constituting a programmable system to create double-strand breaks, thereby opening exciting and far reaching possibility of targeted and exact genome editing (Doudna and Charpentier 2014). Within the last couple of years, this method has been rapidly adopted and widely utilized by researchers due to its universal applicability and robustness when compared to the previously known genome editing methods, thus making CRISPR one of the most important technological breakthroughs of this century.

There are three main classes of the CRISPR/Cas system. Class I and III systems utilize a multi-protein effector complex to cleave invading DNA. By contrast, class II systems require only a single effector protein and are thus more advantageous for genome editing. The Cas9 protein from *Streptococcus pyogenes* is the effector protein most widely used for genome editing. More recently, the single effector Cpf1 has been demonstrated to be another highly effective tool for genome engineering (Zetsche *et al.* 2015a). The Cas9 effector protein from *S. pyogenes* revealed the existence of a two-component system; Cas9 endonuclease for generating targeted DSBs coupled with a single chimeric guide RNA (gRNA), which are short programmable sequences

that base pairs with complementary DNA in the targeted genome and provide a genomic address for the associated generic Cas9 protein to locate the target locus (Jinek *et al.* 2012) (Fig. 4). There is some evidence showing the toxic effects of Cas9 when expressed at high levels using binary expression systems (Port and Bullock 2016).

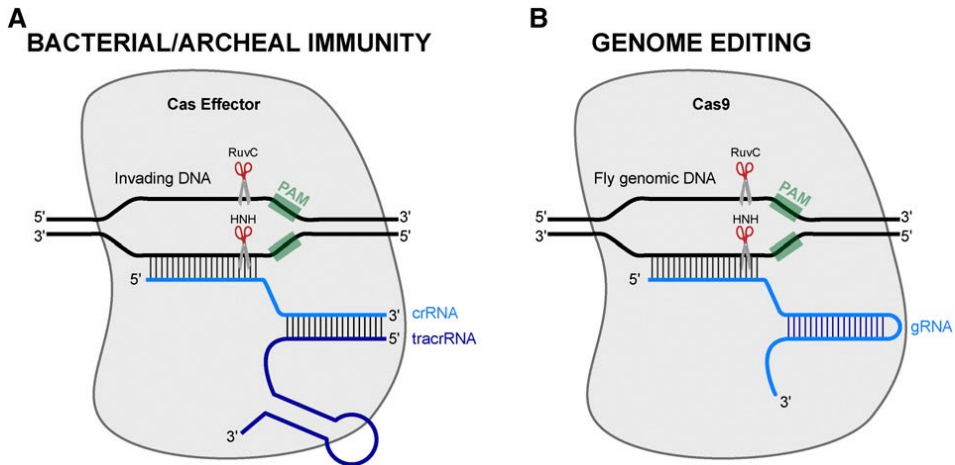


Figure 4. CRISPR-Cas system for two-component genome editing. (A) A Cas effector protein is targeted to a PAM-containing DNA target by a crRNA and tracrRNA. The invading DNA is subsequently cleaved by the RuvC and HNH nuclease domains, generating a DSB. (B) A two-component system composed of Cas9 and a single chimeric gRNA can cleave genomic DNA containing a PAM sequence. Target specificity is determined by 20 nucleotides at the 5' end of the gRNA, allowing the researcher to program the Cas9 cleavage (Adapted from Bier *et al.* 2017).

Cas9 can be used either as DNA, mRNA, or in protein form, or can also be transgenically expressed (Bassett and Liu 2014; Housden *et al.* 2014, 2016; Gratz *et al.* 2015a,b; Housden and Perrimon 2016; Port and Bullock 2016b). Species-specific features such as recognition of different Protospacer Adjacent Motif (PAM) sequence and size of the nuclease within a given type of CRISPR/Cas system have also been used to diversify genome-engineering capabilities, as it expands the availability for targeting and the size of the nuclease, with smaller

variants being more easily delivered into the cells (Gasiunas *et al.* 2012; Cong *et al.* 2013; Esvelt *et al.* 2013; Hou *et al.* 2013; Kleinstiver *et al.* 2015; Ran *et al.* 2015; Zetsche *et al.* 2015a). PAM is a short stretch (3-8 bp) of DNA that serves as a binding signal for Cas9 and is present upstream of the target region. PAM sequences 5'-NGG-3' are relevant in the case of SpCas9, while saCas9 from *Staphylococcus aureus* and StCas9 from *Streptococcus thermophilus* specifically recognizes the PAM motifs 5'-NNG-3', 5'-NNGRR(N)-3' and 5'-NNAGAAW-3', respectively. In all cases, "N" represents any base followed by two guanines ("G"), where "R" is G or A and "W" represents T or A.

Various studies reported the application of CRISPR/Cas9 for gene editing in protists (Ghorbal *et al.* 2014; Grzybek *et al.* 2018). Indeed, successful knock-out and gene silencing have been described in trypanosomatids *Trypanosoma* spp., and *Leishmania* spp. (Lander *et al.* 2015; Peng *et al.* 2015), as well as multi-gene editing in *Dictyostelium* (Sekine *et al.* 2018) and gene editing in several species of diatoms (Hopes *et al.* 2016). CRISPR/Cas9 has been successfully used by different Cas9/sgRNA delivery methods into the cells through: (i) co-transfection of Cas9 protein/gRNA ribonucleoprotein complex *via* electroporation or liposome-mediated transfection (Soares Medeiros *et al.* 2017), (ii) electroporation of all-in-one vector containing Cas9 gene and sgRNA (Stukenberg *et al.* 2018), and (iii) by creating cell lines stably expressing Cas9, in which only sgRNAs are required to target specific genes (Peng *et al.* 2015; Aygul *et al.* 2018).

2.4 RNA interference (RNAi)

RNAi mechanism regulates gene expression through small (~20–30 nt), noncoding RNAs (ncRNA) that target transcripts for silencing in a sequence-specific manner. It was first described in *Caenorhabditis elegans* (Fire *et al.* 1998), after which it was identified in a wide array of organisms. In the classical RNAi pathway, a double-stranded RNA (dsRNA) molecule identical to the target sequence is recognized and cleaved by Dicer (RNaseIII enzyme) into short (20-30 nt) double-strand fragments that get loaded into an RNA-Induced Silencing Complex (RISC). The passive strand gets further degraded, activating RISC to use the other strand *i.e.* the guide strand, which has a perfect or near-perfect complementarity to target transcripts for Argonaute-mediated transcriptional or post-transcriptional silencing (Portnoy *et al.* 2011; Hannus *et al.* 2013). RNAi is a very powerful tool to study gene functions in organisms, especially in those, which lack HR and are thereby limited to traditional gene knock-out techniques. In protists, RNAi pathways are highly diverse, functionally specialized and composed of unique small ncRNA classes and different Argonaute proteins (Braun *et al.* 2010; Shi *et al.* 2000; Prucca *et al.* 2008; Zhang *et al.* 2008). Preliminary data showed the presence of Dicer and Argonaute in the *D. papillatum* genome, which suggests that it is meaningful to try this technique in diplomonads in the future (our unpublished data).

3. Nuclear genome, polycistronic transcription and mRNA splicing in diplomemids

The kinetoplastids are the only group within Euglenozoa for which high-quality nuclear genomes are available mainly due to their medical significance (Berriman *et al.* 2005; El-Sayed *et al.* 2005; Ivens *et al.* 2005). Moreover, a draft nuclear genome is being generated for *D. papillatum*. The approximate size of its nuclear genome is around 240 Mbp. It is of highly repetitive nature, especially in intergenic regions (unpublished data; G. Burger, personal communication). The unusual feature of *D. papillatum*, known only from a handful of eukaryotes including trypanosomes, is its chromosomal gene organization, in which genes are arranged in long polycistronic transcription units of huge number of open reading frames (unpublished data). The complex processing of nuclear primary transcripts involving mandatory *trans*-splicing is a common feature among Euglenozoa yet has been thoroughly studied only in trypanosomes (Huang and van der Ploeg 1991; Ullu *et al.* 1993) and to some extent in euglenoids (Tessier *et al.* 1991). A similar mRNA processing process exists in nematodes (Krause and Hirsh 1987) and in trematodes (Rajkovic *et al.* 1990), and more recently was also found in some chordates (Vandenberghe *et al.* 2001).

Primary nuclear transcripts are processed by *trans*-splicing of a short sequence (for example, 26-39 nt long in most euglenozoans) called spliced leader (SL RNA) (Walder *et al.* 1986; Tessier *et al.* 1991; Frantz *et al.* 2000; Mayor and Floeter-Winter 2005). Although they differ in numerous components, the players behind *cis* general *trans* as well as SL *trans* splicing are similar. In case of *cis*-splicing, the splice donor and

acceptor sites lie on the same RNA strand, separated by the intron sequence, which contains the branch point adenosine (Stover *et al.* 2005; Lasda & Blumenthal 2011). The spliceosomal small nuclear ribonucleoprotein particle recognizes the splice site boundaries and catalyzes the *trans*-esterification reaction, thus pairing a 5' (donor) and a 3' (acceptor) splice site, followed by the removal of the intronic RNA sequence between them (Stover *et al.* 2005). Both general *trans*-splicing and SL *trans*-splicing also rely on the presence of a 5' splice site sequence on one RNA molecule, a branchpoint polypyrimidine tract, and a 3' splice site sequence. The exact consensus sequence for each of these elements varies slightly among different organisms (Lasda & Blumenthal 2011).

The SL *trans*-splicing requires a special class of non-coding small nuclear RNAs, called SL RNAs. These molecules contain two functionally distinct moieties: the 5' part consists of the leader sequence that is spliced onto a pre-mRNA, along with the SL RNA methylguanosine cap; the 3' part contains a binding site for the Sm protein complex, which also binds the RNAs involved in intron splicing (Stover *et al.* 2005). These two moieties are separated by a splice donor site (GU dinucleotide). In 5' UTR of pre-mRNA the AG splice site (acceptor site) and a conserved adenosine, used as a branch point, are located. Between them a polypyrimidine (Py) tract is present (Mayor and Floeter-Winter 2005; Lasda & Blumenthal 2011; Preußner *et al.* 2012). Thus, a nuclear machinery *trans*-splices the leader sequence to the splice acceptor sites (AG dinucleotides) in the 5' region of target pre mRNAs. As a result, many or even all mRNAs in a given species that employes

SL *trans*-splicing have a common sequence at their 5' ends (Stover *et al.* 2005; Preußner *et al.* 2012).

Similarly to the much more widespread *cis*-splicing, the *trans*-splicing process is mediated by spliceosomes and proceeds through two

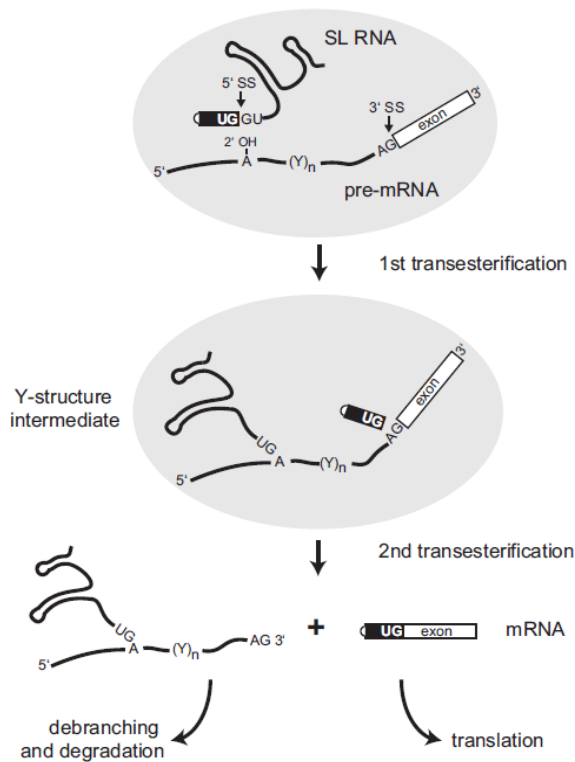


Fig. 5. Mechanism of *trans*-splicing. Depicted is the SL RNA and a pre-mRNA, including the conserved branch point adenosine, the polypyrimidine tract (Y)_n, and the protein-coding exon (as part of a long polycistronic transcript). Part of the consensus sequences at the 5' and 3' splice sites (5_{SS}; 3_{SS}) are shown as well. The first *trans*-esterification step involves a nucleophilic attack of the 2' OH of the branchpoint adenosine at the 5' splice site. Intermediates are the free SL mini-exon and the SL intron/protein-coding exon in form of a so-called Y-structure with the characteristic 2'–5' phosphodiester linkage. In the second step, the free 3' OH of the 5' SL mini-exon attacks the 3' splice site of the protein-coding exon. Joining of the SL RNA and the protein-coding exon into a single mRNA, and the intron is released in the form of a Y-structure, which undergoes debranching and degradation (Adapted from Preußner *et al.* 2012).

transesterification steps, but a so-called Y structure is formed instead of a lariat (Fig. 5). The first transesterification step involves a nucleophilic attack of the 2' OH of the branch point adenosine at the 5' splice site. Intermediates are the free SL mini-exon and the SL intron/protein-coding exon in the form of a Y-structure with the characteristic 2'-5' phosphodiester linkage. In the second step, the free 3' OH of the 5' SL attacks the 3' splice site of the protein-coding exon. As a result, the first 39 nucleotides of the SL RNA and the protein-coding exon are joined with each other, yielding a single mRNA with the released intron in the Y-structure, which subsequently undergoes debranching and degradation (Liang *et al.* 2003; Preußner *et al.* 2012).

The SL-RNAs sequence is weakly conserved among Euglenozoa but includes a consensus binding site (Frantz *et al.* 2000; Tessier *et al.* 1991). Its length ranges from 26 nt in euglenids to 39 nt in both kinetoplastids and diplomemids (Sturm *et al.* 2001). Recent single-cell genomic data from 10 deep-sea pelagic diplomemids confirmed the presence of the SL RNA genes in all of them (Gawryluk *et al.* 2016). In addition, in *D. papillatum* nuclear DNA, numerous SL like sequences are scattered throughout the genome, but only one sequence is the true SL RNA gene being present in about 500 copies) (our unpublished data). With SL RNA the m⁷G cap is added to each protein-coding mRNA, thus ensuring its stability and eventual translation (Stover *et al.* 2005).

Finally, polyadenylation is coupled with *trans*-splicing (Lasda & Blumenthal 2011; Clayton 2016). The pre-mRNA is first cleaved endonucleolytically at a specific site that is followed by sequential addition of AMPs to the hydroxyl group at the 3' end of the mRNA. The

cleavage site is usually a conserved hexanucleotide AAUAAA, located 10 to 30 nt upstream of this 3' end (Clayton 2016).

4. Mitochondrial genome architecture and gene expression

Euglenozoa in protists carry a single large mitochondrion per cell, while the least known symbiontids exhibit a reduced hydrogenosome-like mitochondrion with either reduced or absent cristae (Breglia *et al.* 2010; Edgcomb *et al.* 2011; Yubuki *et al.* 2013). In euglenids, mitochondrial (mt)DNA is linear and likely uniformly distributed in the mitochondrial lumen. Kinetoplastids contain circular covalently linked DNA, which is either supercoiled and free, or relaxed and interlocked into a single network (Faktorová *et al.* 2016). Their mtDNA is comprised of dozen of maxicircles and thousands of mutually concatenated minicircles (Lukeš *et al.* 2002; Povelones *et al.* 2014; Read *et al.* 2016). In all euglenozoans, mtDNA is transcribed into polycistronic tracts that are endonucleolytically cleaved and further *trans*-spliced and/or edited (Flegontov *et al.* 2011; Faktorová *et al.* 2016).

The mitochondrial genome architecture, gene structure and expression in diplomonids are extremely unusual. All extant mitochondria evolved from a single endosymbiotic event, but their genome was subject to dramatically different lineage-specific development and thus exhibits considerable diversity (Burger *et al.* 2016). In all diplomonid studied so far, mitochondrial chromosomes are circular (Roy *et al.* 2007; Kiethega *et al.* 2011; Yabuki *et al.* 2016) carrying highly fragmented genes, the transcripts of which are post-

transcriptionally assembled into contiguous mRNA by *trans*-splicing (Vlcek *et al.* 2011; 2013; Moreira *et al.* 2016).

The first molecular study of a diplomemid mitochondrial gene was carried out on the model species *D. papillatum*. Although a complete *cox1* transcript was recovered, no contiguous gene could be identified in the mtDNA (Maslov *et al.* 1999). Only later studies showed that gene identification is complicated by their highly fragmented nature. The mitochondrial genome of *D. papillatum* consists of a complex array of small covalently closed chromosomes of at least 81 different sequences, which fall into two size categories (6 and 7 kbp) (Fig. 6) (Marande *et al.* 2005).

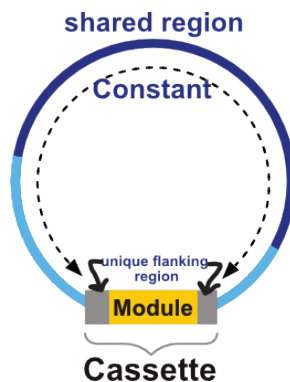


Figure 6: Transcription and transcript processing in *D. papillatum* mitochondrial DNA. Structure of a mitochondrial chromosome with the unique region (“cassette”), which encodes gene module and the non-coding constant region (Adapted from Kiethega *et al.* 2013).

About 95% of a chromosome’s length shares its sequence with that of the other members of its class, with only ~ 5% of the sequence being unique. This unique region referred to as a ‘cassette’, includes a coding sequence, notably a gene fragment (‘module’) 40 to 540 nt long

(Fig. 6) (Marande and Burger 2007; Kiethega *et al.* 2013). Each module is transcribed independently together with long neighboring regions. Next, each long precursor molecule undergoes RNA processing involving removal of the non-coding regions, 3' module polyadenylation, RNA editing by substitution (C-to-U and A-to-I) and by the addition of uridines (U-appendage) (Fig. 7).

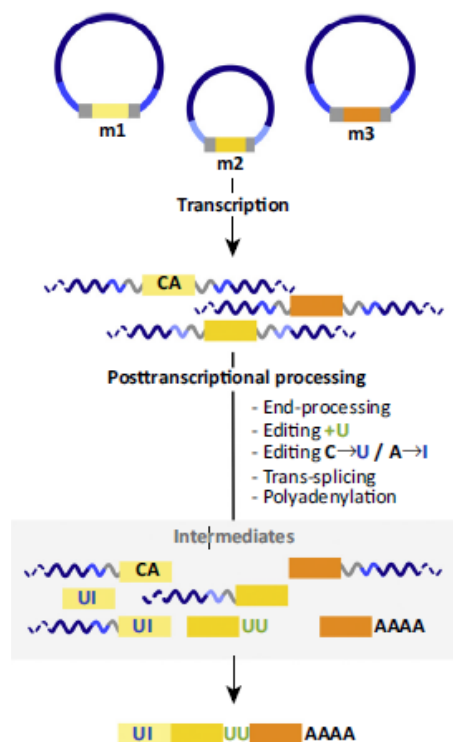


Figure 7. Mitochondrial gene structure and gene processing. The depicted gene consists of 3 modules, which are all transcribed separately. The primary transcript includes long stretches of the constant region which is common to both classes of mitochondrial chromosomes. The class-specific regions (bright and pale blue) are adjacent to the cassette (light gray). Each cassette contains a gene fragment (module m1 to m3, yellow shades). Transcription starts and ends within the constant region and the transcripts thus require extensive end processing. Substitution editing at clustered sites is illustrated for module 1, symbolized by the dinucleotide CA that is converted to UI.

U-appendage editing is shown for module 2, to which two Us are added at its 3' end prior to *trans*-splicing. Module 3 undergoes polyadenylation (Adapted from Burger *et al.* 2016).

All modules of a given gene are further *trans*-spliced together in a defined order, into a contiguous, single-gene mRNA (Marande *et al.* 2005; Marande and Burger 2007; Vlcek *et al.* 2011; Kiethega *et al.* 2013; Valach *et al.* 2014). The aforementioned RNA processing steps occur in parallel mode with *trans*-splicing. The only exception is uridine (U)-appendage editing, which must be completed before the corresponding terminal part of the module is joined to its downstream module or, in the case of terminal modules, before it is polyadenylated (Kiethega *et al.* 2013; Valach *et al.* 2014; Faktorová *et al.* 2018). The intermediates of the mitochondrial *trans*-splicing in *D. papillatum* occur at a detectable steady-state level. The transcripts assembly does not seem to proceed in a defined order (e.g. from 3' to 5' or *vice versa*). Instead, module transcript assembly appears to start simultaneously with any cognate pair and to proceed in parallel until completion (Kiethega *et al.* 2013 and Moreira *et al.* 2016).

Various molecular and *in-silico* studies proved that *trans*-splicing in diplomids is different from the intron-based (*trans*)-splicing, as neither of the conserved group I or group II intron sequence motifs, nor highly conserved spliceosomal intron motifs were detected in the regions adjacent to the coding modules (Kiethega *et al.* 2011). Additionally, module junction sites differ from all known insertion points of organellar introns which further diminishes the possibility of the presence of discontinuous group I and II introns. Hence, the nature of these *trans*-

acting factors and/or machinery involved in this process remains to be established (Kiethega *et al.* 2013; Valach *et al.* 2014; Moreira. *et al.* 2016).

The number of mitochondrial chromosomes and module/cassette arrangements and module content vary considerably among diplonemids. In *D. papillatum* 76 out of its 81 mapped chromosomes carry a single cassette with a single module (mono-module/mono-cassette organization), while three cassettes enclose two modules (two-module/mono-cassette) and two cassettes are without an identified module (Moreira *et al.* 2016, Valach *et al.* 2016). In case of *D. ambulator*, 21 out of 40 distinct chromosomes contain multi-modules either in the single cassette or multi-cassette arrangement. In case of *R. euleeides*, 10 chromosomes out of 66 carry multi-modules (Valach *et al.* 2016). Recent studies showed that the hemistasiids have even much more fragmented genomes, with many gene modules existing in multiple non-identical copies, leading to highly heterozygous gene sequences. Moreover, the length of modules is half of their average size in classical diplonemids (Yabuki *et al.* 2016).

The mitochondrial gene content in diplonemids studied so far is quite conventional, specifying proteins of the respiratory chain, and the large (LSU) and small subunit (SSU) rRNAs (Vlcek *et al.* 2011; Moreira *et al.* 2016; Valach *et al.* 2017). However, the amount of DNA in their mitochondrion is by all means unconventional. Indeed, the estimated size of the *D. papillatum* mitochondrial genome is 250 Mbp and represents the largest amount of DNA documented in any bacterium-derived organelle and is unusually GC rich (46.5% A+T), even more than

the corresponding nuclear genome (Lukes *et al.* 2018). Interestingly, *D. papillatum* also employs a non-standard genetic code in its mitochondrion with the UGA stop codon reassigned as tryptophan, and additional GUG initiation codon used in a few cases (Vlček *et al.* 2011).

4.1 RNA editing patterns in diplomemids

In addition to gene fragmentation and *trans*-splicing, RNA editing in the mitochondrial genome of diplomemids is unique and complex, well described at the level of nucleic acids, while the responsible machinery remains unknown. Diplomemids possess three distinct types of RNA editing, namely C-to-U and A-to-I substitution editing and the post-transcriptional uridine addition (U-appendage) at 3' ends of certain modules. All these events take place post-transcriptionally with some differences across species, as well as across transcripts (Valach *et al.* 2016). About half of the U-appendage sites are shared among four previously studied four diplomemids (*D. papillatum*, *D. ambulator*, *R. euleeides* and *F. neradi*).

RNA editing was first observed in the *cox1* gene of *D. papillatum* which is fragmented into 9 modules, with six Us inserted between its fourth and fifth modules (Marande 2007). Initially, this U-insertion was considered a rare event, but afterwards U-appendages were found in many transcripts, with most sites shared among the studied species. Earlier, *Physarum polycephalum*, a myxomycete was the only known eukaryote to use more than one type of editing in its mitochondria: co-transcriptional nucleotide insertions and occasional

deletions (Gott *et al.* 2005), as well as post-transcriptional C-to-U deaminations (Bundschuh *et al.* 2011). In the mitochondrial LSU rRNA of *D. papillatum*, 26 Us are added between modules 1 and 2 and it was demonstrated that Us were added upstream of module junctions before *trans*-splicing took place (Valach *et al.* 2014; 2016). RNA editing is also present in all other examined *Diplonema*, *Rhynchopus* and in *Hemistasia* species. Many U-appendage sites are shared among them while the pattern of substitution RNA editing varies more significantly (Moreira *et al.* 2016).

4.2 Prediction of RNA editing machinery in diplomemids

The three distinct types of RNA editing (C-to-U and A-to-I substitutions, and U-appendage) in the *D. papillatum* mitochondrion affect both pre-mRNAs and pre-rRNAs. Out of 18 *D. papillatum* mitochondrial genes, the transcripts of 15 of them are edited with a total of ~350 nt changed or added. So far, the corresponding RNA editing machinery remain completely unknown, yet it is likely that it is not directed by *cis*-motif to the editing sites (Kiethega *et al.* 2013; Moreira *et al.* 2016; Valach *et al.* 2016).

For the U-insertions and deletions, a “cut-and-reseal” strategy has been described in the mitochondrion of kinetoplastid flagellates. The information for the exact U-insertions and/or deletions is provided by small (~50-nt-long oligo-uridylated) and abundant RNA molecules called guide RNAs. The process is highly complex and requires the activity of up to 100 specialized proteins constituting several complexes

(Aphasizhev and Aphasizheva 2011; Read *et al.* 2016). Although similar small RNA molecules could guide *trans*-splicing and RNA editing also in diplomonids, such molecules seem to be absent from the *D. papillatum* mitochondrion (Faktorová *et al.* 2018).

C-to-U editing is particularly common in plants, while A-to-I substitutions are mostly restricted to the nucleus of mammals with both events catalyzed by cytidine and adenosine deaminases, respectively (Sun *et al.* 2016). Nucleus-encoded RNA-binding proteins, such as pentatricopeptide repeats (PPR) proteins in plants and likely also in other organisms such as Heterolobosea (Fu *et al.* 2014; Knoop *et al.* 2010) are known to contribute to the organellar RNA editing (Yagi *et al.* 2013). Both the land plants and the distantly related heterolobosean protists encode DYW-type PPR proteins (Rudinger *et al.* 2011; 2013; Fu *et al.* 2014), the presence of which in the *D. papillatum* genome would indicate C-to-U editing. Preliminary bioinformatics searches indeed showed that it contains multiple members of several families of RNA-binding proteins (PPR, KH and Zinc-finger domains), thus representing promising candidates for future functional studies in diplomonids (Burger and Valach 2018). The A-to-I and C-to-U substitutions in animal mRNAs and tRNAs are performed by RNA- or tRNA-specific adenosine deaminases (ADAR and ADAT) or the apolipoprotein B mRNA editing complex ("APOBEC") comprising a catalytic subunit. In all these enzymes, the deaminase domains, which share several catalytic features, are present (Gerber *et al.* 2001). Hence, the substitution RNA editing could be performed by their homologs in the *D. papillatum* mitochondrion (Valach *et al.* 2016).

The U-appendage editing has so far not been reported from outside of diplomids, but it is worth noting that homologs of uridylyltransferase 2 (TUTase 2), which is critical for the U editing in kinetoplastids, are present and thus constitute potential candidates that may play a role in the addition of terminal Us to the RNA modules in diplomids (Valach *et al.* 2016). In any case, the identification of proteins performing RNA editing and SL *trans*-splicing will require extensive functional and biochemical studies (Valach *et al.* 2016). Moreover, several RNA ligases that reseal RNA indel editing sites in the kinetoplastid mitochondrion or, alternatively, RtcB-type ligases that join the ends of tRNA halves after intron removal, were found in the mitochondrion of *D. papillatum* (G. Burger *et al.*, unpublished data) and represent other attractive candidates for knock-outs or knock-downs.

5. References

1. Adl S.M., Bass D., Lane C.E., Lukeš J., Schoch C.L., Smirnov A., Agatha S., Berney C., Brown M.W., Burki F. *et al.* (2019) Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.* 66: 4–119.
2. Aphasizhev R, Aphasizheva I (2011) Uridine insertion/deletion editing in trypanosomes: a playground for RNA-guided information transfer. *WIREs* 2: 669–685
3. Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., *et al.*, (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315: 1709–1712.

4. Bassett A., Liu J.L., (2014) CRISPR/Cas9 mediated genome engineering in *Drosophila*. *Methods* 69: 128–136.
5. Bellofatto V., Hartree D.E., Torres-Munoz J., 1993. *Leptomonas seymouri* as a model system for the analysis of gene expression in trypanosomatids. *J. Parasitol.* 79: 637-44.
6. Berriman M., Ghedin E., Hertz-Fowler C., Blandin G., Renaud H. *et al.*, (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309: 416–422.
7. Bier E, Harrison MM, O'Connor-Giles KM, Wildonger J (2018) Advances in engineering the fly genome with the CRISPR-Cas system. *Genetics* 208: 1-18.
8. Braun L, Cannella D, Ortet P, Barakat M, Sautel CF, Kieffer S, Garin J, Bastien O, Voinnet O, Hakimi MA. (2010) A complex small RNA repertoire is generated by a plant/fungal-like machinery and effected by a metazoan-like Argonaute in the single-cell human parasite *Toxoplasma gondii*. *PLoS Pathog.* 6: e1000920.
9. 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.
10. Bundschuh R., Altmüller J., Becker C., Nürnberg P., and Gott J.M., *et al.*, (2011) Complete characterization of the edited transcriptome of the mitochondrion of *Physarum polycephalum* using deep sequencing of RNA. *Nucleic Acids Res.* 39: 6044–6055.

11. Burger G., Moreira S., Valach M., (2016) Genes in hiding. *Trends Genet.* 32: 553–565.
12. Burger G., Valach M., (2018) Perfection of eccentricity: Mitochondrial genomes of diplomemids. *IUBMB Life* 70: 1197–1206.
13. Burki F., Roger A.J., Brown M.W. and Simpson A.G.B. (2019) The new tree of eukaryotes. *Trends Ecol. Evol.* 35: 43–55.
14. Busse I., Preisfeld A., (2002) Phylogenetic position of *Rhynchopus* sp. and *Diplonema ambulator* as indicated by analyses of euglenozoan small subunit ribosomal DNA. *Gene* 284: 83–91.
15. Byrum J., Jordan S., Safrany S.T., Rodgers W., (2004) Visualization of inositol phosphate-dependent mobility of Ku: depletion of the DNA-PK cofactor InsP6 inhibits Ku mobility. *Nucleic Acids Res.* 32: 2776–2784.
16. Carradec Q., Pelletier E., Da Silva C., Alberti A., Seeleuthner Y., *et al.*, (2018) A global ocean atlas of eukaryotic genes. *Nat Commun.* 25, 9: 373.
17. Cavalier-Smith T., (1993) Kingdom Protozoa and its 18 phyla. *Microbiol Rev.* 57: 953-994.
18. Cavalier-Smith T., (2016) Higher classification and phylogeny of Euglenozoa. *Eur. J. Protistol.* 56: 250–276.
19. Choo J.H., Han C., Kim J.Y., Kang H.A., (2014) Deletion of a KU80 homolog enhances homologous recombination in the thermotolerant yeast *Kluyveromyces marxianus*. *Biotechnol Lett.* 36: 2059–2067.
20. Clayton C. E., (2016) Gene expression in Kinetoplastids. *Curr Opin. Microbiol.* 32: 46-51.

21. Cong L., Ran F. A., Cox D., Lin S., Barretto R., *et al.*, (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819–823.
22. David V., Archibald J.M., (2016) Evolution: Plumbing the depths of diplonemid diversity. *Curr. Biol.* 26: R1290–R1292.
23. de Vargas C., Audic S., Henry N., Decelle J., Mahe F., *et al.* (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* 348: 1261605.
24. Doudna J.A., Charpentier E., (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science* 346: 1258096–1258096.
25. Durant S., Karran P., (2003) Vanillins a novel family of DNA-PK inhibitors. *Nucleic Acids Res.* 31: 5501–12.
26. Dyer P., Dean S., Sunter J., (2016) High-throughput gene tagging in *Trypanosoma brucei*. *J. Vis. Exp.* 12: 114.
27. Dynan W. S., Yoo S., (1998) Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids *Nucleic Acids Res.* 26: 1551-1559.
28. Ebenezer T.E., Zoltner M., Burrell A., Nenarokova A., Novák Vanclová A.M.G., *et al.* (2019) Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biol.* 17: 11.
29. Edgcomb V. P., Breglia S. A., Yubuki, N., Beaudoin, D., Patterson, D. J., *et al.* (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.
30. Elbrächter M., Schnepf E., Balzer I., (1996) *Hemistasia phaeocysticola* (Scherffel) comb. nov., redescription of a free-

- living, marine, phagotrophic kinetoplastid flagellate. *Arch. Protistenkd.* 147:125–136.
31. Eloë E.A., Shulse C.N., Fadrosch D.W., Williamson S.J., Allen E.E., Bartlett D.H., (2011) Compositional differences in particle-associated and free-living microbial assemblages from an extreme deep-ocean environment. *Environ. Microbiol. Rep.* 3: 449–458.
 32. Emerson C.H., Bertuch A.A., (2016) Consider the workhorse: nonhomologous end-joining in budding yeast. *Biochem. Cell Biol.* 94: 396–406.
 33. Emerson C.H., Lopez C.R., Ribes-Zamora A, Polleys E.J., Williams C.L., Yeo L, Zaneveld J.E., Chen R, Bertuch A.A. 2018 Ku DNA end-binding activity promotes repair fidelity and influences end-processing during non-homologous end-joining in *Saccharomyces cerevisiae*. *Genetics* 209: 115-128.
 34. Esvelt K. M., Mali P., Braff J. L., Moosburner M., Yaung S. J. *et al.*, (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* 10: 1116–1121.
 35. Faktorová D., Nisbet R.E.R., Fernández Robledo J.A., Casacuberta E., Sudek L., Allen A. E., *et al.*, (2020). Genetic tool development in marine protists: Emerging model organisms for experimental cell biology, *Nat. Methods*, doi.org/10.1101/718239.
 36. Faktorová D., Valach M., Kaur B., Burger G., Lukeš, J., (2018) Mitochondrial RNA editing and processing in diplomonid protists. In: Cruz-Reyes, J and Gray, M (eds.), RNA Metabolism in Mitochondria. *Nucleic Acids and Molecular Biology*. Springer, Cham 34: 145–176.

37. Faktorová D., Valach M., Kaur B., Burger G., Lukeš J., (2018) Mitochondrial RNA editing and processing in diplomemid. *Protists* In.pp. 145–176.
38. Fire A., Xu S., Montgomery M.K., Kostas S.A., Driver S.E., Mello C.C., (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 39: 806–811
39. Flegontov P., Gray M. W., Burger G., Lukeš J., (2011) Gene fragmentation: a key to mitochondrial genome evolution in Euglenozoa? *Curr. Genet.* 57: 225-232.
40. Flegontova O., Flegontov P., Malviya S., Audic S., Wincker P., *et al.*, (2016) Extreme Diversity of Diplonemid Eukaryotes in the Ocean. *Curr. Biol.* 26: 3060-3065.
41. Flegontova O., Flegontov P., Malviya S., Poulain J., de Vargas C., *et al.*, (2018) Neobodonids are dominant kinetoplastids in the global ocean. *Environ. Microbiol.* 20: 878–889.
42. Frantz, C., Ebel, C., Paulus, F., & Imbault, P. (2000). Characterization of *trans*-splicing in Euglenoids. *Current Genetics* 37: 349–355.
43. Fu CJ, Sheikh S, Miao W, Andersson SG, Baldauf SL. 2014 Missing genes, multiple ORFs, and C-to-U type RNA editing in *Acrasis kona* (Heterolobosea, Excavata) mitochondrial DNA. *Genome Biol. Evol.* 6: 2240-2257.
44. Gasiunas G., Barrangou R., Horvath P., Siksnys V., (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* 109: E2579–E2586.

45. Gawryluk R.M.R., del Campo J., Okamoto N., Strassert J.F.H., Lukeš J., Richards, T.A., Worden, A.Z., Santoro, A.E. and Keeling, P.J. (2016) Morphological Identification and Single-Cell Genomics of Marine Diplonemids. *Curr. Biol.* 26: 3053–3059.
46. Gerber AP, Keller W. 2001 RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem. Sci.* 26: 376-384.
47. Ghorbal M, Gorman M, Macpherson CR, Martins RM, Scherf A, Lopez-Rubio JJ. 2014 Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. *Nat. Biotechnol.* 32: 819–821.
48. Goins C.L., Gerik K.J., Lodge J.K., (2006). Improvements to gene deletion in the fungal pathogen *Cryptococcus neoformans*: absence of Ku proteins increases homologous recombination, and co-transformation of independent DNA molecules allows rapid complementation of deletion phenotypes. *Fung. Genet. Biol.* 43: 531–544.
49. Goma F., Garcia P.A., Delaney J., Girguis P.R., Buie C.R., Edgcomb V.P., (2017) Toward establishing model organisms for marine protists: successful transfection protocols for *Parabodo caudatus* (Kinetoplastida: Excavata). *Environ. Microbiol.* 19: 3487–3499.
50. Gott J.M., 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.

51. Gratz S. J., Rubinstein C. D., Harrison M. M., Wildonger J., O'Connor-Giles K. M., (2015b) CRISPR-Cas9 genome editing in *Drosophila*. *Curr. Protoc. Mol. Biol.* 111: 2.1–2.20.
52. Gratz S. J., Harrison M. M., Wildonger J., O'Connor-Giles K. M., (2015a) Precise genome editing of *Drosophila* with CRISPR RNA-guided Cas9. *Methods Mol. Biol.*, 1311: 335–348.
53. Griessman K., (1914) Ueber marine Flagellaten. *Arch Protistenkd* 32: 1–78.
54. Porter, D. (1973) *Isonema papillatum* sp. n., a new colorless marine flagellate: a light- and electron microscopic study. *J. Protozool.* 20: 351–356.
55. Grzybek M, Golonko A, Górska A, Szczepaniak K, Strachecka A, Lass A, Lisowski P. (2018) The CRISPR/Cas9 system sheds new lights on the biology of protozoan parasites. *Appl. Microbiol. Biotechnol.* 102: 4629-4640.
56. Her J., Bunting S.F., (2018). How cells ensure correct repair of DNA double strand breaks. *J. Biol. Chem.* 293: 10502–10511.
57. Heyer W.D., 2007. Biochemistry of eukaryotic homologous recombination. In *Molecular genetics of recombination* (ed. Aguilera A, Rothstein R), pp. 95–133. Springer, *Top Curr. Genet.*, 17: 95– 133
58. Hopes A., Nekrasov V., Kamoun S., Mock T., (2016). Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant Methods* 12: 49.
59. Hou Z., Zhang Y., Propson N. E., Howden S. E., Chu L. F. *et al.*, (2013) Efficient genome engineering in human pluripotent stem

- cells using Cas9 from *Neisseria meningitidis*. *Proc. Natl. Acad. Sci. USA* 110: 15644–15649.
60. Housden B. E., Perrimon N., (2016a) Cas9-mediated genome engineering in *Drosophila melanogaster*. *Cold Spring Harb. Protoc.* 9, doi: 10.1101/pdb.top086843
61. Housden B. E., Perrimon N., (2016b) Comparing CRISPR and RNAi-based screening technologies. *Nat. Biotechnol.* 34: 621–623.
62. Huang J., van der Ploeg L.H., (1991) Maturation of polycistronic pre-mRNA in *Trypanosoma brucei*: analysis of *trans* splicing and poly(A) addition at nascent RNA transcripts from the hsp70 locus. *Mol. Cell. Biol.* 11: 3180–3190.
63. Ifuku K, Yan D, Miyahara M, Inoue-Kashino N, Yamamoto YY, Kashino Y (2015) A stable and efficient nuclear transformation system for the diatom *Chaetoceros gracilis*. *Photosynth. Res.* 123: 203-211.
64. Ishemgulova A., Hlaváčová J., Majerová K., Butenko A., Lukeš J., *et al.*, (2018) CRISPR/Cas9 in *Leishmania mexicana*: A case study of LmxBTN1. *PLoS One* 13: e0192723.
65. Ishino Y., Shinagawa H., Makino K., Amemura M., Nakata A., (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* 169: 5429–5433
66. Ivens A.C., Peacock C.S., Worthey E.A., Murphy L., Aggarwal G., *et al.*, (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 309: 436-42.

67. Jayathilaka K., Sheridan S.D., Bold T.D., Bochenska K., Logan H.L., *et al.*, (2008) A chemical compound that stimulates the human homologous recombination protein RAD51. *Proc. Natl. Acad. Sci. USA* 105: 15848–15853.
68. Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., *et al.*, (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816–821.
69. Kaur B., Valach M., Peña-Díaz P., Moreira S., Keeling P.J., *et al.*, (2018) Transformation of *Diplonema papillatum*, the type species of the highly diverse and abundant marine microeukaryotes Diplonemida (Euglenozoa). *Environ. Microbiol.* 20: 1030–1040.
70. Kiethega G.N., Yan Y., Turcotte M., Burger G., (2013) RNA-level unscrambling of fragmented genes in *Diplonema* mitochondria. *RNA Biol.* 10: 301-313.
71. Kiethega G.N., Turcotte M., Burger G., (2011) Evolutionarily conserved *cox1* *trans*-splicing without *cis*-motifs. *Mol. Biol. Evol.* 28: 2425–2428.
72. Kleinstiver B.P., Prew M.S., Tsai S. Q., Topkar V. V., Nguyen N. T., *et al.*, (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523: 481–485.
73. Knoop, V. and Rüdinger, M. (2010) DYW type PPR proteins in a heterolobosean protist: plant RNA editing factors involved in an ancient horizontal gene transfer? *FEBS Lett.* 584: 4287– 4291.
74. Kooistra R., Hooykaas P.J., Steensma H.Y., (2004) Efficient gene targeting in *Kluyveromyces lactis*. *Yeast* 15: 781-92.
75. Krause M., Hirsh D., (1987) A *trans*-spliced leader sequence on actin mRNA in *Caenorhabditis elegans*. *Cell* 49:753–761.

76. Lander N., Chiurillo M., Vercesi A., Docampo R., (2017) Endogenous C terminal tagging by CRISPR/Cas9 in *Trypanosoma cruzi*. *Bio-Protocol* 7: e2299.
77. Lander N., Chiurillo M.A., Docampo R., (2016a) Genome editing by CRISPR/ Cas9: a game change in the genetic manipulation of protists. *J. Eukaryot. Microbiol.* 63:679–690.
78. Lara E., Moreira D., Vereshchaka A., López-García P., (2009) Pan-oceanic distribution of new highly diverse clades of deep-sea diplomonids. *Environ. Microbiol.* 11: 47–55.
79. Larsen J., Patterson D.J., (1990) Some flagellates (Protista) from tropical marine sediments. *J. Nat. Hist.*, 24: 801–937.
80. Lasda EL, Blumenthal T (2011) *Trans-splicing. Wiley Interdiscip. Rev. RNA* 2: 417–434.
81. Leahy J.J., Golding B.T., Griffin R.J., Hardcastle I.R., Richardson C., *et al.*, (2004) Identification of a highly potent and selective DNA-dependent protein kinase (DNA-PK) inhibitor (NU7441) by screening of chromenone libraries. *Bioorg. Med. Chem. Lett.* 14: 6083–7.
82. LeBowitz J. H., (1995). Transfection experiments with *Leishmania*. *Methods Cell Biol.* 45: 65-78.
83. Li X., Heyer W.D., (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res.* 18: 99-113.
84. Li Y., Sun Y., Hines J.C., Ray D.S., (2007). Identification of new kinetoplast DNA replication proteins in trypanosomatids based on predicted S-phase expression and mitochondrial targeting. *Eukaryot. Cell.* 6: 2303-10.

85. Liang X.H., Haritan A., Uliel S., Michaeli S., (2003) *trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot. Cell* 2: 830-40.
86. Lieber M.R., (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 79: 181-211.
87. 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.
88. Lukeš J., Flegontova O., Horák A., (2015) Diplonemids. *Curr. Biol.* 25: R702-704.
89. Lukeš J., Guilbride D. L., Votýpka J., Zíková A., Benne R., Englund P. T., (2002) Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot. Cell.* 1: 495-502.
90. Lukeš J., Wheeler R., Jirsová D., David V., Archibald J.M., (2018) Massive mitochondrial DNA content in diplomemid and kinetoplastid protists. *IUBMB Life* 70: 1267–1274.
91. Makarova K.S., Haft D.H., Barrangou R., Brouns S.J.J., Charpentier E., *et al.*, (2011) Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.* 9: 467–477.
92. Makiuchi T., Annoura T., Hashimoto M., Hashimoto T., Aoki T., Nara, T., (2011) Compartmentalization of a glycolytic enzyme in *Diplonema*, a non-kinetoplastid euglenozoan. *Protist* 162: 482–489.
93. Malkova A., Haber J.E., (2012). Mutations arising during repair of chromosome breaks. *Annu. Rev. Genet.* 46: 455– 473.

94. Malkova A., Ira G., (2013) Break-induced replication: functions and molecular mechanism. *Curr. Opin. Genet. Dev.* 23: 271–279.
95. Marande W., Luke J., Burger G., (2005) Unique mitochondrial genome structure in diplomonids, the sister group of kinetoplastids. *Eukaryot. Cell* 4: 1137-1146.
96. Marande W., Burger, G., (2007) Mitochondrial DNA as a genomic jigsaw puzzle. *Science* 318- 415.
97. Maslov D.A., 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.
98. Massana R., (2011) Eukaryotic picoplankton in surface oceans. *Annu. Rev. Microbiol.* 65: 91–110.
99. Mayer M.G., Floeter-Winter L.M., (2005) Pre-mRNA trans-splicing: from kinetoplastids to mammals, an easy language for life diversity. *Mem. Inst. Oswaldo Cruz* 100: 501-13.
100. Miyahara M., Aoi M., Inoue-Kashino N., Kashino Y., Ifuku K., (2013) Highly efficient transformation of the diatom *Phaeodactylum tricorutum* by multi-pulse electroporation. *Biosci. Biotechnol. Biochem.* 77: 874-876.
101. Montegut-Felkner A.E., Triemer R.E., (1996) Phylogeny of *Diplonema ambulator* (Larsen and Patterson): 2. Homologies of the feeding apparatus. *Europ. J. Protistol.* 32:64–76.
102. Morales J., Hashimoto M., Williams T.A., Hirawake-Mogi H., Makiuchi T., *et al.*, (2016) Differential remodelling of peroxisome function underpins the environmental and metabolic

- adaptability of diplomemids and kinetoplastids. *Proc Biol Sci.* 283: 20160520.
103. Moreira D., Lopez-Garcia P., Rodriguez-Valera F., (2001) New insights into the phylogenetic position of diplomemids: G+C content bias, differences of evolutionary rate and a new environmental sequence. *Int. J. Syst. Evol. Microbiol.* 51: 2211-2219.
104. 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.
105. Mukherjee I., Hodoki Y., Okazaki Y., Fujinaga S., Ohbayashi K., and Nakano S.I., (2019) Widespread Dominance of kinetoplastids and unexpected presence of diplomemids in deep freshwater lakes. *Front. Microbiol.* 10: 2375.
106. Nayak T., Szewczyk E., Oakley C.E., Osmani A., Ukil L., Murray S.L., Hynes M.J., Osmani S.A., Oakley B.R., (2006) A versatile and efficient gene-targeting system for *Aspergillus nidulans*. *Genetics* 172: 1557–1566.
107. Neal J.A., Dang V., Douglas P., Wold M.S., Lees-Miller S.P., Meek K., (2011) Inhibition of homologous recombination by DNA-dependent protein kinase requires kinase activity, is titratable, and is modulated by autophosphorylation. *Mol. Cell Biol.* 31:1719–33.
108. Neal J.A., Meek K., (2011) Choosing the right path: does DNA-PK help make the decision? *Mutat. Res.* 711:73–86.
109. Nenarokova A., Záhonová K., Krasilnikova M., Gahura O., McCulloch R., (2019) Causes and Effects of Loss of Classical

- Nonhomologous End Joining Pathway in Parasitic Eukaryotes.
*mBio*10: e01541-19.
110. Ninomiya Y., Suzuki K., Ishii C., Inoue H., (2004) Highly efficient gene replacements in *Neurospora* strains deficient for nonhomologous end-joining. *Proc. Natl. Acad. Sci. USA* 101: 12248–12253.
 111. Nymark M., Sharma A. K., Sparstad T., Bones A. M., Winge, P., (2016) A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6:24951.
 112. Okamoto N., Gawryluk R.M.R., Campo J., Strassert J.F.H., Lukeš J., (2019) A revised taxonomy of diplomonads including the Eupelagonemidae n. fam. and a type species, *Eupelagonema oceanica* n. gen. & sp. *J. Eukaryot. Microbiol.* 66: 519–524.
 113. O'Neill E. C., Trick M., Hill L., Rejzek M., Dusi R. G., et al., (2015) The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Mol. Biosyst.* 11: 2808-2820.
 114. Peng D., Kurup S.P., Yao P.Y., Minning T.A., Tarleton R.L., (2014) CRISPR-Cas9- mediated single-gene and gene family disruption in *Trypanosoma cruzi*. *mBio* 6:e02097-14.
 115. Port F., Bullock S. L., (2016a) Augmenting CRISPR applications in *Drosophila* with tRNA-flanked sgRNAs. *Nat. Methods* 13: 852–854.
 116. Porter D., (1973) *Isonema papillatum* sp. n., a new colorless marine flagellate: a light- and electronmicroscopic study. *J. Protozool.* 20: 351–356.

117. Portnoy V., Huang V., Place R.F., Li L.C., (2011) Small RNA and transcriptional upregulation. *Wiley Interdiscip. Rev. RNA*. 2: 748-60.
118. Povelones M.L., (2014) Beyond replication: division and segregation of mitochondrial DNA in kinetoplastids. *Mol Biochem. Parasitol.* 196: 53–60.
119. Prakash R., Zhang Y., Feng W., Jasin M., (2015) Homologous recombination and human health: The roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harb. Perspect. Biol.* 7: a016600.
120. Preußner C., Palfi Z., Bindereif A., (2009) Special Sm core complex functions in assembly of the U2 small nuclear ribonucleoprotein of *Trypanosoma brucei*. *Eukaryot. Cell* 8: 1228–1234.
121. Prokopchuk G., Tashyreva D., Yabuki A., Horák A., Masařová P., Lukeš J., (2019) Morphological, Ultrastructural, Motility and Evolutionary Characterization of Two New Hemistasiidae Species. *Protist* 170: 259–282.
122. Prucca CG, Slavin I, Quiroga R, Elías EV, Rivero FD, Saura A, Carranza PG, Luján HD. 2008 Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature* 456: 750-4.
123. Rajkovic A., Davis R. E., Simonsen J. N., Rottman F. M., (1990) A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proc. Natl. Acad. Sci. USA* 87:8879–8883.

124. Ran F. A., Cong L., Yan W. X., Scott D. A., Gootenberg J. S. et al., (2015) *In-vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* 520: 186–191
125. Read L.K., Lukeš J., Hashimi H., (2016) Trypanosome RNA editing: the complexity of getting U in and taking U out. *Wiley Interdiscip. Rev. RNA* 7: 33-51.
126. Rodgers K., McVey M., (2016) Error-prone repair of DNA double-strand breaks. *J. Cell Physiol.* 231:15-24.
127. Roy J., Faktorová D., Benada O., Lukeš J., Burger G., (2007) Description of *Rhynchopus euleeides* n. sp. (Diplonemea), a free-living marine euglenozoan. *J. Eukaryot. Microbio.* 154: 137–145.
128. Rüdinger M, Szövényi P, Rensing SA, Knoop V 2011 Assigning DYW-type PPR proteins to RNA editing sites in the funariid mosses *Physcomitrella patens* and *Funaria hygrometrica*. *Plant J.* 67: 370-80.
129. Rüdinger M, Kindgren P, Zehrmann A, Small I, Knoop V. 2013. A DYW-protein knockout in *Physcomitrella* affects two closely spaced mitochondrial editing sites and causes a severe developmental phenotype. *Plant J.* 76: 420-32.
130. Scheckenbach F., Hausmann K., Wylezich C., Weitere M., Arndt H., (2010) Large-scale patterns in biodiversity of microbial eukaryotes from the abyssal sea floor. *Proc. Natl. Acad. Sci. USA* 107:115–120.
131. 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.

132. Sekine R., Kawata T., Muramoto T., (2018) CRISPR/Cas9 mediated targeting of multiple genes in *Dictyostelium*. *Sci. Rep.* 31: 8471.
133. Simpson A. G. B., (1997) The identity and composition of the Euglenozoa. *Archiv für Protistenkunde*, 148: 318-328.
134. Soares Medeiros L.C., South L., Peng D., Bustamante J.M., Wang W., (2017) Rapid, selection-free, high-efficiency genome editing in protozoan parasites using CRISPR-Cas9 ribonucleoproteins. *mBio* 8: e01788–e01717.
135. Song W., Dominska M., Greenwell P.W., Petes T.D., (2014) Genome-wide high-resolution mapping of chromosome fragile sites in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* 111: 2210–2218.
136. Stover, N. A., Kaye, M. S., and Cavalcanti, A. R. (2006). Spliced leader *trans*-splicing. *Curr. Biol.* 16: R8–R9.
137. Stukenberg D., Zauner S., Dell'Aquila G., Maier U.G., (2018) Optimizing CRISPR/Cas9 for the diatom *Phaeodactylum tricornutum*. *Front. Plant Sci.* 6; 9:740.
138. Sturm, N. R., M. C. Yu, and D. A. Campbell. 1999. Transcription termination and 3' end processing of the spliced leader RNA in kinetoplastids. *Mol. Cell. Biol.* 19:1595–1604.
139. Sun T., Bentolila S., Hanson M. R. (2016). The unexpected diversity of plant organelle RNA editosomes. *Trends Plant Sci.* 21: 962–973
140. Tashyreva D., Prokopchuk G., Yabuki A., Kaur B., Faktorová D., *et al.*, (2018) Phylogeny and morphology of new diplomonads from Japan. *Protist* 169: 158–179.

141. Tessier L. H., Keller M., Chan R. L., Fournier R., Weil J. H., Imbault P., (1991) Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by *trans*-splicing in *Euglena*. *EMBO J.* 10: 2621–2625.
142. Triemer R.E., Ott D.W., (1990) Ultrastructure of *Diplonema ambulator* Larsen & Patterson (Euglenozoa) and its relationship to *Isonema*. *Europ. J. Protistol.* 25:316–320.
143. Ullu E., Matthews K.R., Tschudi C., (1993) Temporal order of RNA-processing reactions in trypanosomes: rapid *trans* splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Mol. Cell. Biol.* 13: 720–725.
144. 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 diplomemids. *RNA Biol.* 13: 1204–1211.
145. Valach M., Moreira S., Hoffmann S., Stadler P.F., Burger G., (2017) Keeping it complicated: Mitochondrial genome plasticity across diplomemids. *Sci. Rep.* 7: 14166.
146. van der Oost J., Jore M.M., Westra E.R., Lundgren M., Brouns S.J., (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* 34: 401-7.
147. Vandenberghe A. E., Meedel T. H., Hastings K. E., (2001) mRNA 5'-leader *trans*-splicing in the chordates. *Genes Dev.* 15: 294–303.
148. Vickerman K., (2000) Diplonemids (Class: Diplonemea Cavalier Smith, 1993). In Lee JJ, Leedale GF, Bradbury P (eds) An

- Illustrated Guide to the Protozoa Vol. 2, 2nd edn, *Society of Protozoologists, Lawrence, Kansas, USA*, pp 1157–1159.
149. Vlcek C., Marande W., Teijeiro S., Lukeš J., Burger G., (2011) Systematically fragmented genes in a multipartite mitochondrial genome. *Nucleic Acids Res.* 39: 979–988.
150. von der Heyden S., Chao E.E., Vickerman K., Cavalier-Smith T., (2004) Ribosomal RNA phylogeny of bodonid and diplomemid flagellates and the evolution of euglenozoa. *J. Eukaryot. Microbiol.* 51: 402–416.
151. Walder J.A., Eder P.S., Engman D.M., Brentano S.T., Walder R.Y., *et al.*, (1986) The 35-nucleotide spliced leader sequence is common to all trypanosome messenger RNA's. *Science* 1; 233: 569-71.
152. Waters C.A., Strande N.T., Wyatt D.W., Pryor J.M., Ramsden D.A., (2014) Nonhomologous end joining: a good solution for bad ends. *DNA Repair (Amst)* 17: 39–51.
153. Weller G.R., Kysela B., Roy R., Tonkin L.M., Scanlan E., *et al.*, (2002) Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* 297: 1686 –1689.
154. Williams G.J., Hammel M., Radhakrishnan S.K., Ramsden D., Lees-Miller S.P., Tainer J.A., (2014) Structural insights into NHEJ: building up an integrated picture of the dynamic DSB repair super complex, one component and interaction at a time. *DNA Repair (Amst)* 17:110 –120.
155. Willmore E., de Caux S., Sunter N.J., Tilby M.J., Jackson G.H., *et al.*, (2004) A novel DNA-dependent protein kinase inhibitor,

- NU7026, potentiates the cytotoxicity of topoisomerase II poisons used in the treatment of leukemia. *Blood* 103: 4659–65.
156. Yabuki A., Tame A., (2015) Phylogeny and reclassification of *Hemistasia phaeocysticola* (Scherffel) Elbrachter & Schnepf, 1996. *J. Eukaryot. Microbiol.* 62: 426–429.
157. 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.
158. Yagi Y., Tachikawa M., Noguchi H., Satoh S., Obokata J., Nakamura T., (2013) Pentatricopeptide repeat proteins involved in plant organellar RNA editing. *RNA biology* 10: 1419–1425.
159. Yi Z., Berney C., Hartikainen H., Mahamdallie S., Gardner M., et al., (2017) High-throughput sequencing of microbial eukaryotes in Lake Baikal reveals ecologically differentiated communities and novel evolutionary radiations. *FEMS Microbiol. Ecol.* 93: 8.
160. Yoshida Y., Tomiyama T., Maruta T., Tomita M., Ishikawa T., Arakawa K., (2016) De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics* 17: 182.
161. Yubuki N., Simpson A. G., Leander B. S., (2013) Reconstruction of the feeding apparatus in *Postgaardia mariagerensis* provides evidence for character evolution within the Symbiontida (Euglenozoa). *Eur. J. Protistol.* 49: 32-39.
162. Zetsche B., Gootenberg J. S., Abudayyeh O. O., Slaymaker I. M., Makarova K.S., et al., (2015a) Cpf1 is a single RNA-guided

endonuclease of a class 2 CRISPR-Cas system. *Cell* 163: 759–771.

163. Zhang D., Xiong H., Shan J., Xia X., Trudeau V.L., (2008) Functional insight into Maelstrom in the germline piRNA pathway: a unique domain homologous to the DnaQ-H 3'-5' exonuclease, its lineage-specific expansion/loss and evolutionarily active site switch. *Biol. Direct* 3: 48.

Chapter 1

Phylogeny and morphology of new diplomids from Japan

ORIGINAL PAPER

Phylogeny and Morphology of New Diplonemids from Japan



Daria Tashyreva^{a,2}, Galina Prokopchuk^{a,2}, Akinori Yabuki^{b,2}, Binnypreet Kaur^{a,c,2},
Drahomíra Faktorová^{a,c}, Jan Votýpka^{a,d}, Chiho Kusaka^b, Katsunori Fujikura^b,
Takashi Shiratori^e, Ken-Ichiro Ishida^f, Aleš Horák^{a,c}, and Julius Lukeš^{a,c,1}

^aInstitute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice (Budweis), Czech Republic

^bDepartment of Marine Diversity, Japan Agency for Marine-Earth Science and Technology, Yokosuka, Japan

^cFaculty of Science, University of South Bohemia, České Budějovice (Budweis), Czech Republic

^dFaculty of Sciences, Charles University, Prague, Czech Republic

^eGraduate School of Life and Environmental Sciences

^fFaculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

Submitted June 28, 2017; Accepted February 5, 2018
Monitoring Editor: Alastair Simpson

Diplonemids were recently found to be the most species-rich group of marine planktonic protists. Based on phylogenetic analysis of 18S rRNA gene sequences and morphological observations, we report the description of new members of the genus *Rhynchopus* – *R. humris* sp. n. and *R. serpens* sp. n., and the establishment of two new genera – *Lacrimia* gen. n. and *Sulcionema* gen. n., represented by *L. lanifica* sp. n. and *S. specki* sp. n., respectively. In addition, we describe the organism formerly designated as *Diplonema* sp. 2 (ATCC 50224) as *Flectonema neradi* gen. n., sp. n. The newly described diplonemids share a common set of traits. Cells are sac-like but variable in shape and size, highly metabolic, and surrounded by a naked cell membrane, which is supported by a tightly packed corset of microtubules. They carry a single highly reticulated peripheral mitochondrion containing a large amount of mitochondrial DNA, with lamellar cristae. The cytopharyngeal complex and flagellar pocket are contiguous and have separate openings. Two parallel flagella are inserted sub-apically into a pronounced flagellar pocket. *Rhynchopus* species have their flagella concealed in trophic stages and fully developed in swimming stages, while they permanently protrude in all other known diplonemid species.

© 2018 Elsevier GmbH. All rights reserved.

Key words: *Diplonema*; ultrastructure; phylogeny; Euglenozoa; description; flagellates.

Introduction

Diplonemids are colorless heterotrophic, predominantly marine protists, equipped with two flagella. They belong to Euglenozoa and are thus related to ecologically important euglenids and econom-

¹Corresponding author; fax +420 387775416

²These authors contributed equally to this work.
e-mail jula@paru.cas.cz (J. Lukeš).

ically and medically relevant kinetoplastids (Adl et al. 2012; Maslov et al. 1999; Moreira et al. 2001). For a long time, diplomemids were considered a small and rare group of flagellates with only three genera and less than a dozen species formally described (Massana 2011; Simpson 1997; Vickerman 2000; von der Heyden et al. 2004). However, they emerged from obscurity thanks to molecular analysis of hundreds of planktonic samples collected across the globe by several research expeditions (de Vargas et al. 2015; Lara et al. 2009). The recent Tara Oceans 18S rRNA-based metabarcoding survey revealed remarkable diversity and abundance of marine diplomemids: with over 45,000 operational taxonomic units (OTUs), they qualify as the most species-rich marine planktonic eukaryotes (David and Archibald 2016; Flegontova et al. 2016). Furthermore, diplomemids show cosmopolitan distribution with different lineages being abundant and often dominant in most ocean niches, from shallow littoral sediments (Larsen and Patterson 1990) to deep aphotic pelagic waters (de Vargas et al. 2015; Flegontova et al. 2016; Gawryluk et al. 2016; Lara et al. 2009), hydrothermal vents (López-García et al. 2007), and down to poorly studied abyssopelagic zones (Eloe et al. 2011; Scheckenbach et al. 2010).

According to 18S rRNA-based phylogeny, diplomemids represent a monophyletic group which can be subdivided into four robustly supported lineages: (i) the so-called classic diplomemids, hereafter referred to as Diplonemidae, consisting of the genera *Diplonema* and *Rhynchopus*; (ii) a small planktonic clade containing the genus *Hemistasia*; (iii) a deep-sea pelagic diplomemids (DSPD) clade I and (iv) DSPD clade II (Flegontova et al. 2016). DSPD I diplomemids were recently formally described as Eupelagonemidae (Okamoto et al., submitted; this work). While this phylogeny shows a clear support for diplomemid monophyly, the relationships between these major lineages are unclear. Most of the species richness is apparently confined within the hyperdiverse Eupelagonemidae clade which, together with DSPD II, was until recently known exclusively from environmental sequences, and we still lack any cultured representative (Flegontova et al. 2016; Lara et al. 2009; Lukeš et al. 2015). However, the single-cell approach provided a first glance at morphological diversity and genomic characteristics of members of the Eupelagonemidae, although it was stressed that poor quality of genomic assemblies and possible contamination necessitates establishment of stable cultures (Gawryluk et al. 2016).

Despite their abundance, huge diversity and consequently importance in marine food webs, very little is known about the behavior as well as molecular, morphological and biochemical traits of diplomemids (David and Archibald 2016). Indeed, there are only four species altogether for which both morphological observations and sequence data, albeit limited, is available: two members of the genus *Diplonema* – *D. papillatum* (Maslov et al. 1999; Porter 1973) and *D. ambulator* (Busse and Preisfeld 2002; Montegut-Felkner and Triemer 1996; Triemer and Ott 1990; Triemer 1992), and a single species for each of the genera *Rhynchopus* and *Hemistasia*, namely *Rhynchopus euleeides* (von der Heyden et al. 2004; Roy et al. 2007) and *Hemistasia phaeocysticola* (Elbrächter et al. 1996; Yabuki and Tame 2015), respectively. However, virtually all molecular studies, such as the analysis of mitochondrial RNA editing and trans-splicing, were performed solely on *D. papillatum* (Kiethega et al. 2013; Marande et al. 2005; Marande and Burger 2007; Moreira et al. 2016; Vlcek et al. 2011;). Otherwise, a few other species known from early morphological studies, namely *Diplonema breviciliata* (Griessmann 1913), *D. nigricans* (Schuster et al. 1968), *D. metabolicum* (Larsen and Patterson 1990), *Rhynchopus amitus* (Skuja 1948), and *R. coscinodiscivorus* (Schnept 1994), lack any molecular data and are unavailable in culture. Moreover, several putative species have been classified according to their 18S RNA gene sequences and are available at the American Type Culture Collection (ATCC) but lack proper morphological description (von der Heyden et al. 2004).

Diplonemids belonging to the genera *Diplonema* and *Rhynchopus*, or Diplonemidae, were only marginally present in the global metabarcoding dataset (Flegontova et al. 2016). Still, the number of putative diplomemid OTUs revealed by environmental molecular barcodes available on public databases greatly exceeds the currently recognized number of species. Despite their newly discovered abundance and diversity, to the best of our knowledge, no new classic diplomemid species have recently been described. Therefore, we attempted to establish axenic cultures by manual picking of diplomemid-like cells from samples collected in surface waters around Japan. Based on phylogenetic analysis of nearly full-size 18S rRNA gene sequences and morphological observations using electron and light microscopy, we report the description of four new diplomemid species and two new genera. In addition to this, we create a novel

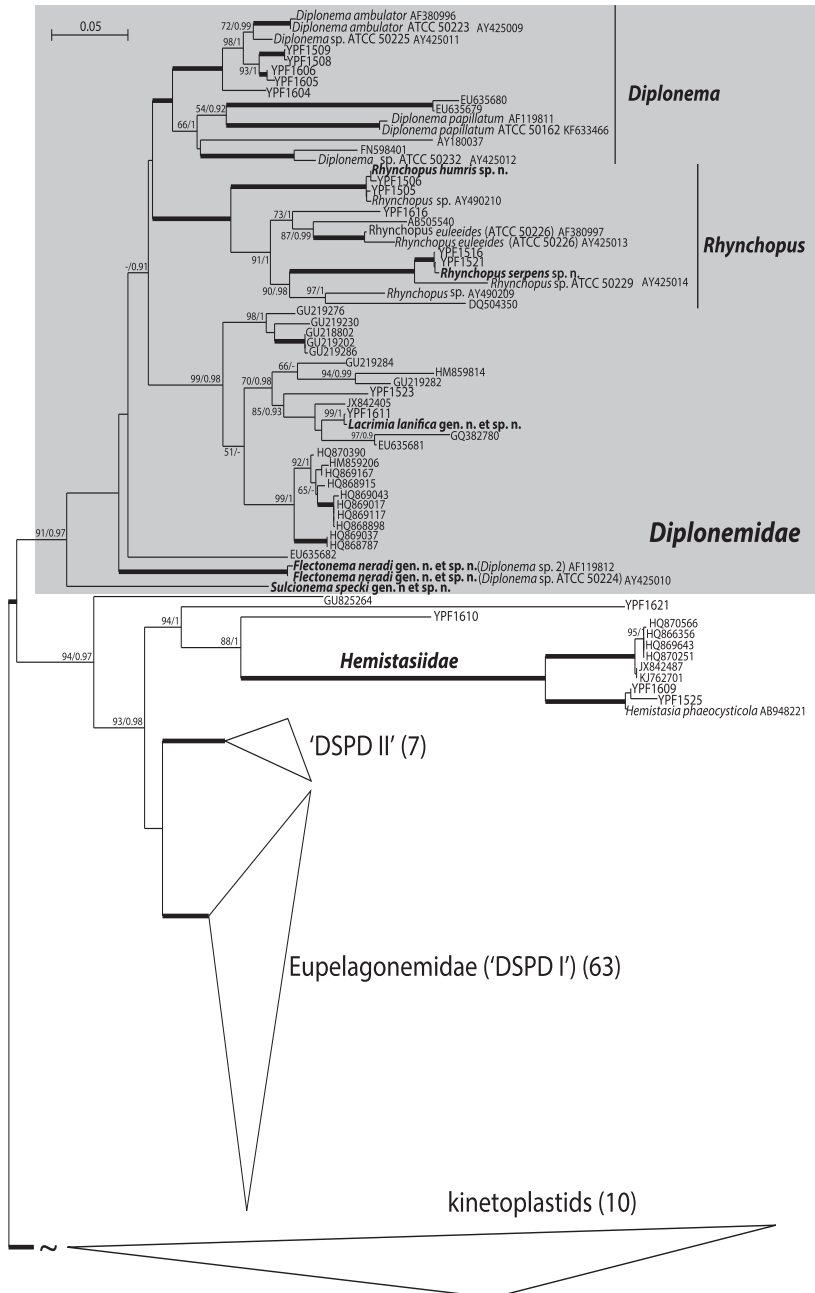


Figure 1. Maximum likelihood (ML) phylogeny of diplomemids based on 18S rRNA dataset K (kinetoplastid)

genus to accommodate a protist previously referred to as *Diplonema* sp. 2 (ATCC50224).

Results

Molecular Phylogeny

We have created three datasets differing by the composition of the outgroup and analyzed them using maximum likelihood and Bayesian inference (see Methods for details). Both methods applied on the dataset K (diplonemids rooted with kinetoplastids only) yielded a highly congruent topology, differing mainly by branching support of some clades (Fig. 1). The monophyly of Diplonemidae, including new members described in this study (see Taxonomic summary), is highly supported. While the internal relationships in Diplonemidae cannot be clearly resolved, there are three main robustly supported clades: (i) a clade containing *D. papillatum*, *D. ambulator* and *Diplonema* sp. ATCC50232, as well as several environmental sequences; (ii) a clade comprising of *Rhynchopus euleeides* (ATCC50226), *Rhynchopus* sp. (ATCC50229) and a few sequences from uncultured isolates, as well as two *Rhynchopus* spp. introduced into culture during of this study and formally described as new species *R. humris* sp. n. and *R. serpens* sp. n. (see below). The last main group identified within Diplonemidae is an abundant assemblage of environmental sequences (iii) described here for the first time, which is represented by an isolate introduced into culture and formally described as a new genus and species *Lacrimia lanifica* gen. n., sp. n. (see below). *Diplonema* sp. 2 (ATCC50224) is formally described here as *Flectonema neradi* gen. n., sp. n. (see below) branches outside of these three main clades along with one environmental sequence (i.e., EU635682), although this position is not highly supported. Finally, YPF1618, formally described in this study as *Sulcionema specki* gen. n., sp. n. (see below), is shown to be the most basal lineage of all classic diplomemids, including the new members.

We have tested the alternative positions of *F. neradi* and *S. specki* using the AU-test. The con-

strained ML tree in which *F. neradi* and *Diplonema* species form a clade is not rejected (p -value: 0.121) and the other constrained ML tree showing the monophyly of *F. neradi*, *S. specki* and *Diplonema* species is not rejected either (p -value: 0.077).

Analyses of auxiliary datasets with outgroup expanded by the addition of euglenids (E) or euglenids and heteroloboseans (H) yielded very similar topologies to that of dataset K. Composition of individual 'genus-level' clades remained the same, however, the branching order mainly within the *Diplonema* and *Rhynchopus* clade slightly differed in nodes with low support in dataset E. In dataset H, *Rhynchopus* branched with the new clade containing *Lacrimia lanifica* to the exclusion of *Diplonema* and the overall support for both datasets decreased, most likely due to the more diverged outgroup sequences and lower number of characters available for analysis (Supplementary Material Fig. SA1 A, B).

Light Microscopy

In order to characterize five strains that in phylogenetic analyses formed novel species, we proceeded to their morphological characterization. Differences on the light microscopy level, while not substantial, were sufficient for their distinction (Table A1A).

Rhynchopus humris

Cells cultivated axenically in a fresh, nutrient-rich medium (=trophic cells) are uniform in size and shape. The cells are of elongated to elliptical shape, dorsoventrally flattened, and narrowed at both ends (Fig. 2 A). The length ranges from 12.5 to 16.4 μm ($14.5 \pm 1.2 \mu\text{m}$; $n=25$), and the width is between 3.3 and 5.1 μm ($4.2 \pm 0.59 \mu\text{m}$; $n=25$). The cells carry two very short flagella buried in the flagellar pocket and thus are invisible by light microscopy. Prominent granulation is observed throughout the cytoplasm. The cells are typically attached to the flask bottom with a small fraction floating freely in the medium but gradually detach as cultures become older. The surface-attached cells move slowly by gliding and frequently change their direction (Fig. 3A). The cells are highly metabolic which

outgroup, 133 taxa, 2000 nucleotides) inferred using IQ-Tree 1.5b under the GTR model with six relaxed rate categories. Branching support (numbers at respective nodes) is represented by non-parametric bootstrapping (BS) estimated from 1000 replicates using thorough algorithm in IQ-Tree as well as by Bayesian posterior probabilities (PP) estimated in Phylobayes 4.1 (C40 + GTR model). For Bayesian inference (BI), 0.95 probability was used as a support criterion, d.t. means different topology of respective node in BI compared to ML. Bold lines represent absolute support (100 BS/1.0 PP).

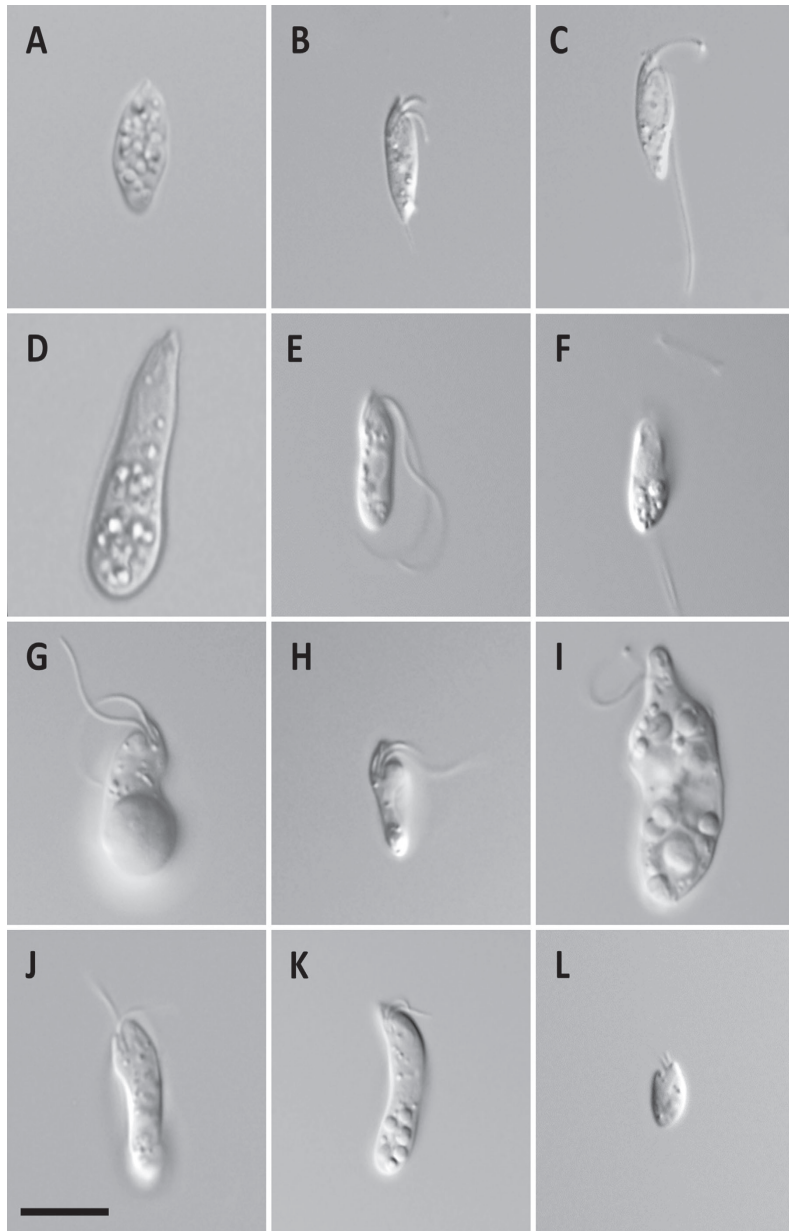


Figure 2. Differential interference contrast images of living cells. Cells of *Rhynchopus humris* (A), *Rhynchopus serpens* (D), *Lacrimia lanifica* (G), *Sulcionema specki* (I), and *Flectonema neradi* (K) from nutrient-rich medium. Note the absence of flagella in both *Rhynchopus* species. Cells of *R. humris* with partially (B) and fully developed

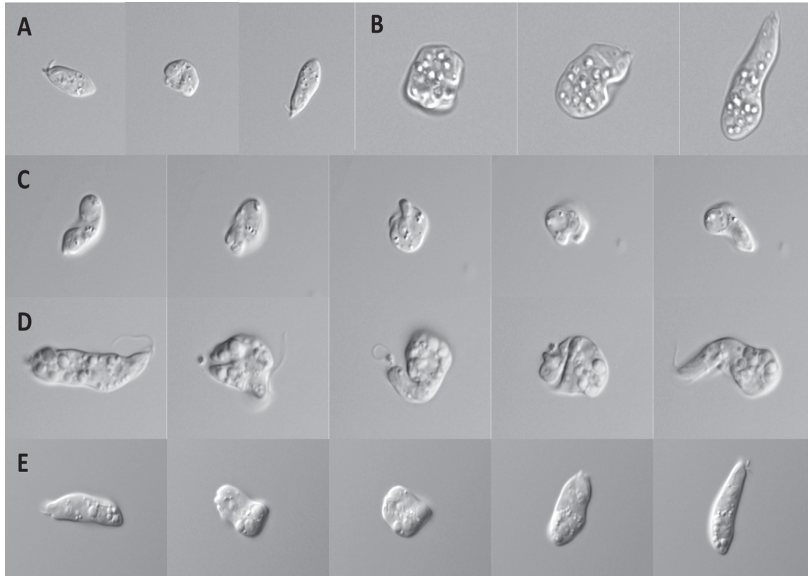


Figure 3. Differential interference contrast images of live cells. Contracting-extending and twisting metabolic movements of starved *R. humris* (A), *R. serpens* (B), starved *L. lanifica* (C), *S. specki* (D), and *F. neradi* (E). Scale bar is 10 μ m.

is displayed as frequent twisting and contracting-extending movements (Table A1A).

When starved, cells reduce their size up to $7.2 \times 3.2 \mu\text{m}$, and the prominent cytoplasmic granulation disappears. *Rhynchopus humris* gradually develops large number of swimming cells with unequally long heterodynamic flagella (Fig. 2B, C), which are about twice of the body length and used for fast cell propulsion. The swimming cells slightly oscillate and move in a straight line but can abruptly stop and change the direction. One flagellum is twisted around the anterior part of the cell wobbling like a lasso, and the other flagellum waving and stretched along the body (Fig. 2C). The flagella are parallel-orientated and arise from subapical flagellar pocket. No cysts were observed in either starved or old batch cultures (Table A1B). Cell division occurs by binary fission from the anterior to posterior end, producing two equal daughter cells.

DAPI staining of the DNA reveals that in *R. humris*, the nucleus is often located proximally to the

cell's periphery (Fig. 4A, B). In most cells, a network of mitochondrial DNA is located peripherally under the plasma membrane (Fig. 4A), clearly without any kinetoplast-like structure. The patchy character of mitochondrial DNA staining was often observed in *R. humris*, in both starved and non-starved cells (Fig. 4B).

Rhynchopus serpens

Cells in nutrient-rich medium (trophic cells) are consistent in morphology and have little variability in size. The cells are elongated, dorsoventrally flattened, tapered at the anterior, and rounded as well as widened at the posterior ends (Fig. 2D). They measure 21.3 to 28.7 μm ($26.0 \pm 1.7 \mu\text{m}$; $n=25$) in length, and 7.2 to 9 μm ($8.0 \pm 0.6 \mu\text{m}$; $n=25$) in width. Numerous cytoplasmic vesicles are concentrated at the posterior part of *R. serpens* cells. The short flagella stubs are concealed in the flagellar pocket and invisible under light microscopy. Cells

(C) flagella, *R. serpens* with partially (E) and fully developed (F) flagella following starvation treatment. *L. lanifica* (H), *S. specki* (J), and *F. neradi* (L) after 7-day starvation. Note the decrease of cytoplasmic granulation in all species. Scale bar is 10 μ m.

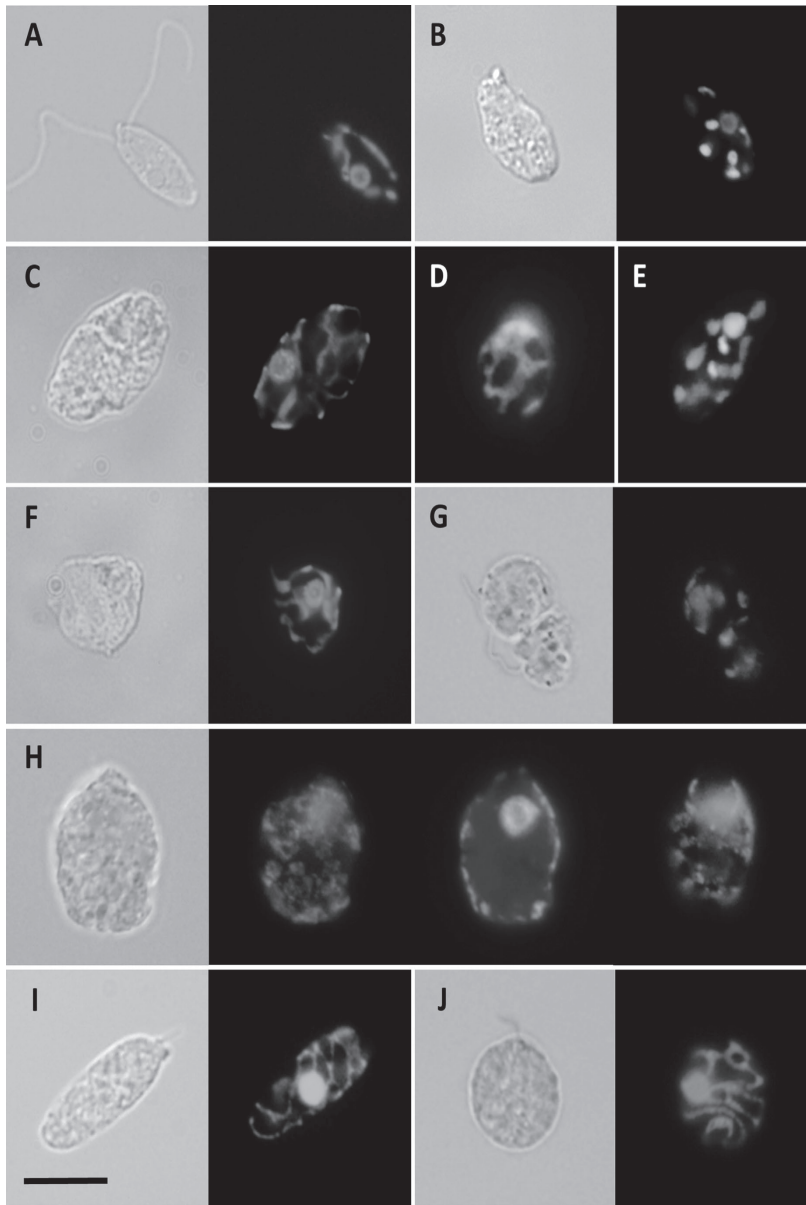


Figure 4. Light and fluorescence micrographs of fixed DAPI-stained cells (**A-C, F-J**). Swimming (**A**) and trophic (**B**) *R. humris* cells with peripheral nuclei; trophic *R. serpens* cell (**C**); non-starved *L. lanifica* cells (**F, J**); a series of images focused through non-starved *S. specki* cell with the nucleus in the anterior part (**H**); non-starved *F. neradi* cells (**I, J**). Note that mitochondrial DNA forms a network in all species but is split into isolated aggregates in *R. humris* (**B**) and *L. lanifica* (**G**). Fluorescence microscopy of starved live *R. serpens* cells stained with DiOC₆(3) reveals extensive mitochondrial network (**D**) or possibly fragmented mitochondria (**E**). Scale bar is 10 μm .

typically attach to the flask bottom, move slowly by gliding and show frequent metabolic movements (Table A1A; Fig. 3B).

As cultures become denser, the cells progressively display greater variability in size, and often detach from the surface. Numerous cytoplasmic refractive bodies prominently present in cells of well-growing cultures diminish or completely disappear in aging cultures and in the starvation medium, indicating that they possibly serve as food reserves. The cells attain a more symmetrical shape but preserve the acute apex and eventually halve in size ($13.5 \times 3.7 \mu\text{m}$). The swimming stages (Fig. 2E, F) occur less frequently than in *R. humris* cultures. The cells are symmetrical and elongated, notably dorsoventrally compressed, tapering at the cell apex. Two unequally long flagella (2 to 2.5 times of the body length) are parallel and inserted into a pronounced flagellar pocket. The swimming is slow when both *R. serpens* flagella are stretched along the body and perform waving movements (Fig. 2E; Table A1B), whereas fast propulsion occurs when the anterior flagellum forms a loop beating rapidly before the anterior end (Fig. 2F; Table A1B). Cysts are not produced. Cells divide by binary fission.

DAPI staining shows that the nucleus has a variable position within *R. serpens* cells, and reveals a network of mitochondrial DNA beneath the cell surface forming aggregates but lacking kinetoplast-like structures (Fig. 4C). Multiple attempts to stain the mitochondrion with specific dyes and immunolabelling worked only in starved *R. serpens*, in which a reticulated mitochondrion was stained with $6 \mu\text{M}$ DiOC₆(3), regardless of the presence of DMSO. In non-starved cells, the signal from the mitochondrion was hindered by very bright fluorescence of cytoplasmic inclusion bodies that non-specifically took up the DiOC₆(3) dye. The staining revealed that a mitochondrion, most likely single, forms a network, which occupies a substantial part of the cell's periphery (Fig. 4D). In some cases, the network seems to split into several smaller oval-shaped organelles, which remain located under the cell surface (Fig. 4E).

Lacrimia lanifica

Cells in well growing cultures are nearly isodiametric but often teardrop-shaped due to the presence of large digestive vacuoles at the posterior end, while the anterior end is narrowed (Fig. 2G). They are 10.4 to $15.3 \mu\text{m}$ long ($13.1 \pm 1.3 \mu\text{m}$; $n=25$) and 7.3 to $10.5 \mu\text{m}$ wide ($8.7 \pm 0.74 \mu\text{m}$; $n=25$). *Lacrimia lanifica* cells tend to form homogeneously suspended cultures, and invariably possess two

unequally long flagella (about the body length) inserted sub-apically into a conspicuous flagellar pocket. When suspended in the medium, *L. lanifica* constantly maintains rotational movement with its two flagella (Table A1A), and is also capable of gliding along the surface by attaching to it with its longer flagellum. In culture, *L. lanifica* displays metabolic movements mostly when attached temporarily to the surface with its body or flagellum, however, these movements become frequent when the cells are trapped under a coverslip or embedded in 0.5% ultralow gelling agarose (Fig. 3C). Cell division occurs by binary fission producing two equal daughter cells, both with digestive vacuoles, or only one of the daughter cells receives the vacuole.

In aged and starved cultures, the cells become notably smaller ($7.5 \times 3.6 \mu\text{m}$) and attain more elongated shape, however, the length of flagella mostly remains unchanged. The food vacuoles become small or disappear (Fig. 2H). Some of the starved *L. lanifica* cells are capable of fast swimming in a straight line through spiral oscillating movement (Table A1B). No cysts produced.

DAPI staining reveals the mitochondrial network beneath the cell's surface (Fig. 4F), and a nucleus in the anterior part of the cell. Regardless of the nutrition, the mitochondrial network sometimes splits into separate small aggregates (Fig. 4G).

Sulcionema specki

Cells are noticeably flattened with tapered, slightly crooked anterior along with rounded posterior ends, and contain conspicuous cytoplasmic granulation (Figs 2I, 3D). The cells range between 23.7 and $33.3 \mu\text{m}$ ($27.6 \pm 2.1 \mu\text{m}$; $n=25$) in length and 5.8 to $9.6 \mu\text{m}$ ($7.4 \pm 0.9 \mu\text{m}$; $n=25$) in width although smaller cells also occur, possibly as a result of recent binary fission. *Sulcionema specki* tends to form homogeneously suspended cultures, and possesses two equal to subequal flagella, which are about a third of the body length and inserted sub-apically in a flagellar pocket. The flagella support only erratic 'floundering'-like movement without cell propulsion. Cells are highly metabolic, which is displayed as frequent radical contortions, twisting, and contraction-extension reminiscent of amoeboid movement (Table A1A; Fig. 3D).

The cell shape, dimensions, and cytoplasmic granulation vary considerably depending on the age of cultures and content of nutrients in the medium. With the age of cultures and after transfer to the starvation medium, the *S. specki* cells begin to display great morphological variability, ranging from short oval to considerably elongated cells, with

both ends either rounded or constricted (Fig. 2J). The smallest cell measured only $16.3 \times 5 \mu\text{m}$, with flagella the same length as in the nutrient-rich medium. The conspicuous cytoplasmic inclusions greatly reduce in size or completely disappear. Although the cells become noticeably smaller, no fast swimming cells were observed even after 10-day incubation in the starvation medium, and the character of their movement remained unchanged (Table A1B). No cysts were observed.

Mitochondrial DNA, stained with DAPI, appears as numerous agglomerates occupying nearly the entire cell's subsurface (Fig. 4H). The nucleus is mostly located at the posterior part of cells (Fig. 4H).

Flectonema neradi

The cells in nutrient-rich medium show little variability in size and shape. They are dorsoventrally flattened, thin and crooked as well as constricted at both ends, which gives a crescent-like shape, with granulation in the posterior region (Fig. 2K). The length ranges from 15.9 to $20.9 \mu\text{m}$ ($18.3 \pm 1.45 \mu\text{m}$; $n=25$), and the width is between 3.5 and $5.8 \mu\text{m}$ ($4.0 \pm 0.6 \mu\text{m}$; $n=25$). *Flectonema neradi* cells are mostly attached to the flask surface, display pronounced metaboly (Fig. 3E), move slowly by gliding, and frequently change their direction. The species invariably possesses two equal flagella emerging sub-apically, which are about a fifth of the body length. Cells often temporarily attach to the flask surface with one of their short flagella and rotate around their anterior end (Table A1A). Cell division occurs by binary fission from the anterior to posterior end, producing two equal daughter cells.

No fast swimming stages or cysts can be observed in old batch cultures and after the starvation treatment. Starved cells are short and mostly rounded, or have constricted posterior ends (Fig. 2L). The dimensions of cells reduces (down to $6.3 \times 3.1 \mu\text{m}$) along with the length of their flagella. The cytoplasmic inclusions are small or absent (Table A1B).

Mitochondrial DNA is organized as a fine network anastomosing peripherally under the plasma membrane (Fig. 4I, J). The nucleus can occupy various positions within cells.

Electron Microscopy

Scanning EM revealed relatively slight differences in general morphology between the examined isolates (Fig. 5). *Rhynchopus humris* (Fig. 5A, F), *R. serpens* (Fig. 5B, G), *S. specki* (Fig. 5D, I) exhibit

long cylindrical or rounded cells, as is the case of *L. lanifica* (Fig. 5C, H), which has the cells tapered anteriorly with subapically inserted flagella. A cylindrical cell is also a characteristic of *F. neradi* (Fig. 5E), whose distinctive feature is that the opening of the flagellar pocket is extended into a curved groove, giving the anterior end a swirl-like appearance (Fig. 5J). The cell surface of all described species is generally smooth, though some pimples can occasionally be found in the anterior part of *R. serpens* (Fig. 5B). Invariably, the opening to the feeding apparatus contiguous with the flagellar pocket is located at the anterior end. The opening is characterized by a prominent collar-like cytostome with an apical papillum (Fig. 5F-J).

Comparison of ultrastructural features observed by TEM revealed similarity in the common organelles and structures in all species examined (Fig. 6). Immediately beneath the cell membrane lies a single-layer corset of peripheral, evenly spaced microtubules interlinked by fine lateral bridges (Fig. 7A-E). These microtubules seem to enfold the cell in a helical pattern. The corset is absent along the zone of flagellar attachment (Fig. 8A-E), and forms a perpendicular junction in the flagellar pocket region (Fig. 7A-E). Tubules of endoplasmic reticulum (ER) are often seen under the microtubular corset (Figs 6, 7). Beneath the microtubules is a dense mitochondrial network anastomosing around the periphery of the cell (Fig. 7F-J). The mitochondria contain variable but overall exceptionally large lamellar cristae, with the attachment to the mitochondrial inner membrane seen only rarely. Unlike other species, *F. neradi* contains numerous short cristae that are arranged transversely rather than longitudinally (Fig. 7O). Another prominent feature of the organelle is numerous patches of electron-dense DNA distributed throughout the mitochondrial matrix (Fig. 7K-O). Longitudinally and cross-sectioned cells show large mitochondrial profiles extending to the areas of the flagellar pocket and feeding apparatus (Figs 8B, D, 9F-I), although they do not reach the zone of flagellar attachment (Fig. 8A-E).

Analysis of longitudinally sectioned cells showed that the flagellar pocket is deep and morphologically similar in all analyzed species (Fig. 8A-E). Flagella arise from parallel basal bodies located in the proximal region of the flagellar pocket. Basal bodies are supported by the asymmetrically distributed ventral (Fig. 8B, E), intermediate (Fig. 8C), and dorsal (not shown) microtubular roots. Flagella of *L. lanifica*, *F. neradi*, *S. specki* invariably have conventional $9+2$ axonemal arrangement of micro-

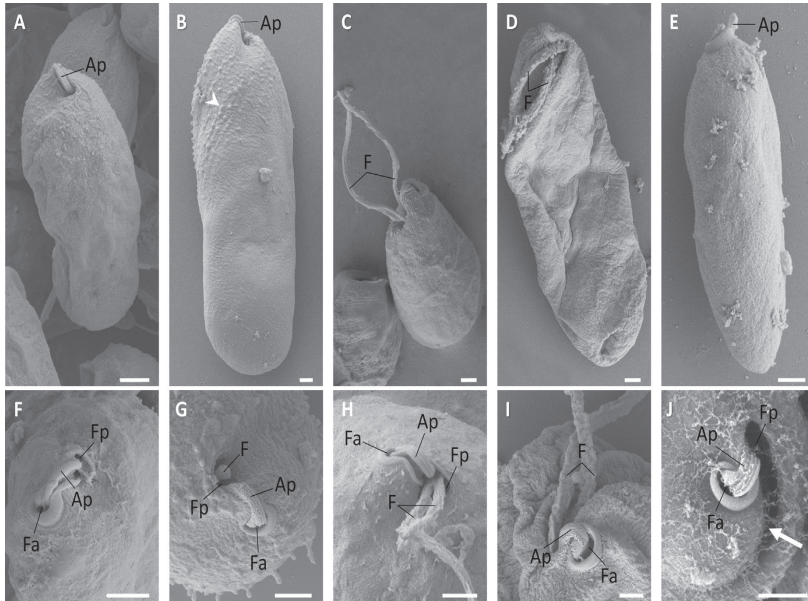


Figure 5. Scanning electron micrographs of *R. humris* (A, F), *R. serpens* (B, G), *L. lanifica* (C, H), *S. specki* (D, I), and *F. neradi* (E, J). (A-E) General appearance with apical papillum (Ap) and flagellum (F). Note a smooth cellular surface and pimples in the anterior part of *R. serpens* (arrowhead). (F-J) Detailed views of the anterior part showing subapical emergence of flagella from the pocket (Fp), and adjacent feeding apparatus (Fa) decorated with an apical papillum. Note a curved groove extended from the flagellar pocket in *F. neradi* (arrow). Scale bar = 1 µm.

tubules, and contain a prominent paraflagellar rod, which has a lattice-like structure (Fig. 8C-E, H-J). The paraflagellar rod joins the axoneme immediately after the distal transitional plate within the flagellar pocket, and runs along its length almost to the tip of the flagellum (Fig. 8A-E). The same structural organization is also preserved in fully developed flagella of starved *R. humris* and *R. serpens* (Fig. 8L, M), whereas, one or both flagella of trophic *Rhynchopus* cells are often rudimentary, with the paraflagellar rod and some axonemal tubules missing (Fig. 8F, G, K). Outer surfaces of the feeding apparatus, flagellar pocket and flagella are normally covered with a prominent glycocalyx (Fig. 8). In *R. humris*, they also coated with dense hairs (Fig. 8F, K, L), a feature that is in a less prominent form occasionally found also in the apex of *S. specki* and *L. lanifica* (data not shown).

Part of the flagellar pocket close to the basal bodies is marked with a complex of reinforcing microtubules (MTR) surrounded by a zone of exclusion along its entire length (Fig. 8A-E). The MTR extends into the pharyngeal complex of the feeding

apparatus, which is located adjacent to the flagellar pocket, and supports both the flagellar pocket and pharyngeal complex along their length (data not shown). The pharyngeal complex opens with a cytostome and extends as a cytopharynx down into the cell reaching up to half of its length (Fig. 9F-J). Figure 6F shows a deep pharyngeal lumen at the level of the nucleus. In longitudinal view, pharynx appears as a horn-like structure (Fig. 9F-J). In all studied species, the feeding apparatus consists of fibrils arranged in a series of vanes (ribs) with supporting elements (rods) composed of fibrils and densely packed microtubules, and a row of longitudinally and transversely oriented microtubules on the opposite side (Fig. 9A-E). In its vicinity, the cytoplasm is surrounded by ER (Fig. 9A-E) and contains numerous vesicles that eventually seem to spread uniformly throughout the cell (Fig. 9F-J). Their content varies in terms of electron density and granulation and they tend to increase in size during spreading away from the feeding apparatus.

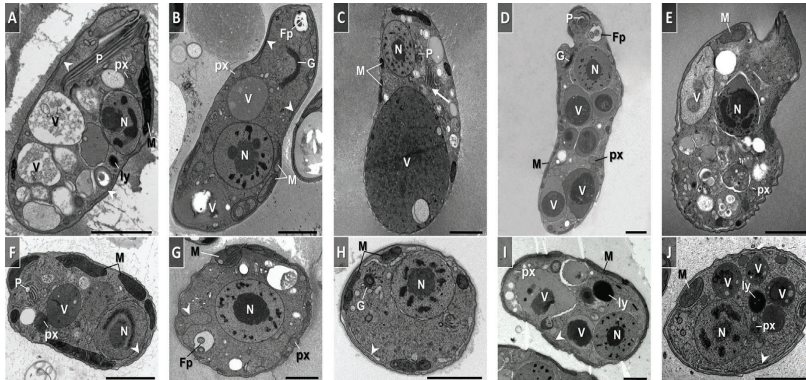


Figure 6. Transmission electron micrographs of *R. humris* (A, F), *R. serpens* (B, G), *L. lanifica* (C, H), *S. specki* (D, I), and *F. neradi* (E, J). (A–E) Longitudinal sections through the cell. (F–J) Transverse sections through the nucleus. (B) Nucleus (N) located in posterior part of the cell. (A, C–E, G) Nucleus in anterior region next to the bottom of the flagellar pocket (Fp), at the end of the pharyngeal lumen (P). Note large convoluted tubular vessels (arrow) near the ceasing pharyngeal lumen (P) and one large amorphous digestion vacuole (V) (C), numerous smaller digestion vacuoles of different contents (A, B, D, E), and little diverse vacuolar inclusions in the cytoplasm of all cells: peroxisomes (px), lysosomes (ly). Mitochondrial network (M) is located at the periphery; Golgi apparatus (G); ribosomes are electron-dense dots within the cytoplasm; arrowheads point to endoplasmic reticulum. Scale bar = 2 μ m.

A single very large digestive vacuole is invariably present in the non-starved *L. lanifica* cells (Fig. 6C), while all other species possess numerous smaller vacuoles of different content (Fig. 6A–B, D–E). A conspicuous Golgi apparatus, with a variable number of both linear and circularized foci, is usually found in the perinuclear region (Fig. 10F–J). The cytoplasm is rich in free ribosomes, lysosomes, peroxisomes, small refractive granules, likely of reserve nature, and sparse endoplasmic reticula that are scattered among the food vacuoles (Figs 6, 7, 9, 10).

The nucleus is invariably large, spherical or oval, confined by a prominent nuclear membrane with readily visible pores (Fig. 10A–E). A persistent densely granular nucleolus is located either centrally or eccentrically within the nucleus, and may occupy as much as its third. The heterochromatin is highly condensed and distributed at the periphery of the homogeneous nucleoplasm (Fig. 10A–E). The nucleus commonly appears next to the bottom of the flagellar pocket, except for *R. serpens*, where it can be located both in the anterior and posterior region of the cell (Fig. 6).

A prominent feature of *L. lanifica* is the presence of large convoluted tubular vessels which lie near the proximal end of the feeding apparatus (Figs 6C, 9H). They look like a single, mostly O-shaped organelle, which somewhat resembles Golgi cis-

ternae, but is larger and less electron-dense. Occasionally, numerous small membrane-bound droplets are seen within the vessels (data not shown). A direct connection between the pharynx and this organelle was not observed.

Discussion

Ribosomal RNA-based phylogenies confirm the existence of two main diplomonid clades, (provisionally) named Diplonemea and Eupelagonemea (Okamoto et al., submitted; this work). Not surprisingly, recent interest was directed to the hyperdiverse eupelagonemid diplomonids, as these may constitute one of the key players of the oceanic ecosystem (David and Archibald 2016; Flegontova et al. 2016; Gawryluk et al. 2016; Lukeš et al. 2015). Despite their enormous diversity, they are, at least on the level of 18S rRNA, quite uniform, which suggest a recent rapid speciation (Flegontova et al. 2016). However, exhaustive sampling of 18S rRNA sequences among Diplonemidae presented here shows that both *Hemistasia* and especially classic diplomonids contain previously unknown phylogenetic structuring and diversity. Classic diplomonids were traditionally composed of just two genera: *Rhynchopus* and *Diplonema* (Adl et al. 2012). Our 18S rRNA phylogeny supported by extensive sam-

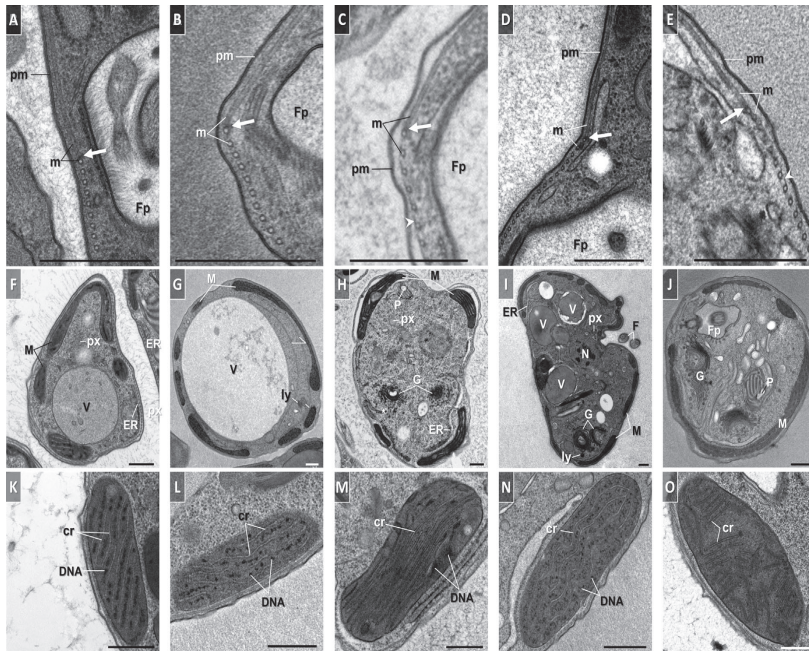


Figure 7. Transmission electron micrographs showing the surface ultrastructure and mitochondrial arrangement of *R. humris* (A, F, K), *R. serpens* (B, G, L), *L. lanifica* (C, H, M), *S. specki* (D, I, N), and *F. neradi* (E, J, O). (A-E) Cross section demonstrating plasma membrane (pm) and single row of microtubular corset (m) underneath it. Arrows point to the junction of microtubules in the region of the flagellar pocket (Fp). (F-J) Transverse section showing peripheral location of the mitochondrion (M). (K-O) Mitochondria displaying characteristic morphology: large lamellar cristae (cr) and abundant dense patches of DNA; nucleus (N); digestion vacuole (V); pharynx (P); Golgi apparatus (G); endoplasmic reticulum at ER; peroxisomes at px; lysosomes at Ly. Ribosomes are electron-dense dots within the cytoplasm. Arrowheads point to fine lateral bridge. Scale bar = 0.5 μ m.

pling of ‘environmental’ sequences, as well as newly described species, reveals the existence of at least five well-supported lineages. While *S. specki* and *F. neradi* fell into rare and species-poor clades, the third novel clade, represented here by the newly described *L. lanifica*, is the most diverse of the classic diplonemids and hence deserves further attention. Although most of the major diplonemid clades are robustly monophyletic, one should keep in mind that the 18S rRNA gene lacks the phylogenetic resolution to uncover their interrelationships and allow better understanding of the evolution of these remarkable protists. This situation shall eventually be improved by more extensive phylogenomic analyses employing the existence of several novel diplonemid species in culture and/or the power of single-cell transcriptomics.

All diplonemid species described to date share a common set of morphological and ultrastructural traits. The cells are sac-like but variable in shape, surrounded by a naked plasma membrane, which is supported by a tightly packed corset of parallel interconnected microtubules that follows a spiral course from the apex toward the posterior end. Such arrangement of the microtubular corset likely allows for extreme plasticity in shape and movement, as was observed in diplonemids and many euglenids (Arroyo et al. 2012; Jeuck and Arndt 2013; Simpson 1997; Swale 1973; Vickerman 1977; this work). The ER cisternae underlying the microtubular corset are likely involved in this active movement (Arroyo et al. 2012). As in the two other euglenozoan groups (Euglenida and Kinetoplastea), diplonemids carry a single peripheral

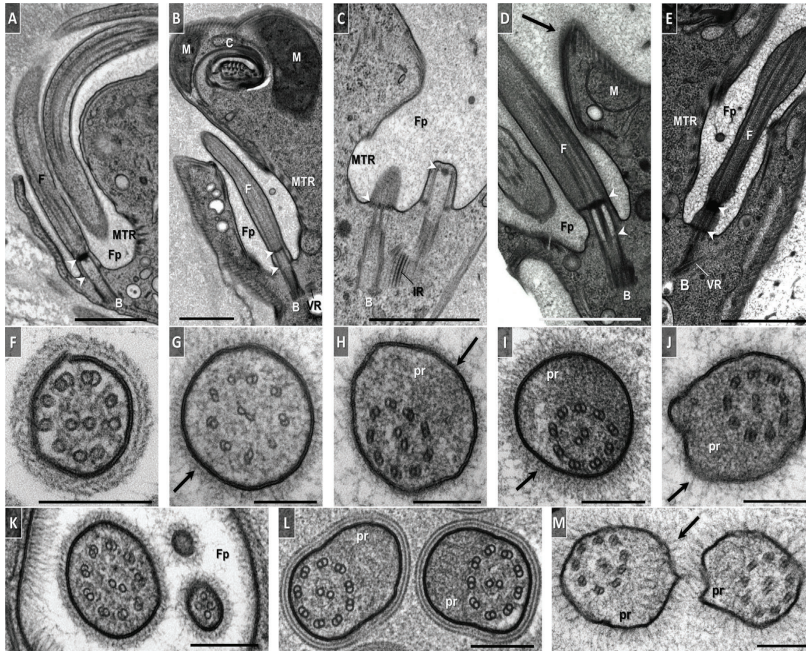


Figure 8. Ultrastructural architecture of the flagellar apparatus in *R. humris* (A, F, K, L), *R. serpens* (B, G, M), *L. lanifica* (C, H), *S. specki* (D, I), and *F. neradi* (E, J). (A-E) Longitudinal sections through flagella (F) and the area of flagellar attachment. In all strains, flagellar pocket (Fp) is deep and contains morphologically similar parallel basal bodies (B) located in the proximal region; ventral (VR) and intermediate (IR) roots. Flagellar pocket (Fp) is bordered by microtubular elements, reinforcing microtubules (MTR), which extend along and support the flagellar pocket and the pharyngeal wall (C). Note disrupted microtubular arrangement in the flagellar attachment zone and portions of large mitochondrion (M) in the areas of flagellar pocket (Fp) and feeding apparatus (C). Arrowheads indicate distal and proximal transitional plates. (F-K) Cross sections showing flagellar arrangement in trophic cells. Note characteristic 9 + 2 axonemal structure and lattice-like paraflagellar rod (pr) in *L. lanifica* (H), *F. neradi* (I), *S. specki* (J), and lack of B-tubules of the outer doublets and paraflagellar rod in *R. humris* (F) and *R. serpens* (G). (K) Likely, transitional stage of *R. humris*, where one of the flagella is fully developed yet missing paraflagellar rod and another appears as a rudiment. (L-M) Cross sections showing presence of conventional 9 + 2 arrangement plus paraflagellar rod in flagella of starved *R. humris* (L) and *R. serpens* (M). Note flagellar hairs in *R. humris* (F, K, L). Arrows point to glycoalyx. Scale bars = 1 μm (A-E) and 0.2 μm (F-M).

highly reticulated mitochondrion containing a large amount of mitochondrial DNA (Elbrächter et al. 1996; Faktorová et al. 2016; Marande et al. 2005; Maslov et al. 1999; Roy et al. 2007). However, as shown here, it is invariably arranged in numerous agglomerates lacking the kinetoplast-like structure characteristic for related trypanosomatids (Jensen and Englund 2012). Similarly dispersed distribution of the mitochondrial DNA was also described in euglenids and some bodonids (Hayashi and Ueda 1989; Lukeš et al. 2002).

Although we are the first to report a successful staining of the diplomonid mitochondria, allowing

the observation of its complex structure, earlier 3D reconstruction and staining of the mitochondrial DNA provided evidence that this organelle is reticulated in other *Diplonema* and *Rhynchopus* species as well (Marande et al. 2005; Roy et al. 2007). In all ultrastructurally examined diplomonids, the single branched mitochondrion contains a few long lamellar cristae arranged in parallel (Elbrächter et al. 1996; Maslov et al. 1999; Porter 1973; Roy et al. 2007; Schnepf 1994; Vickerman 1977). The only known exception is *F. neradi* (this work), which not only contains a much higher number of cristae than other diplomonids, but also has them arranged

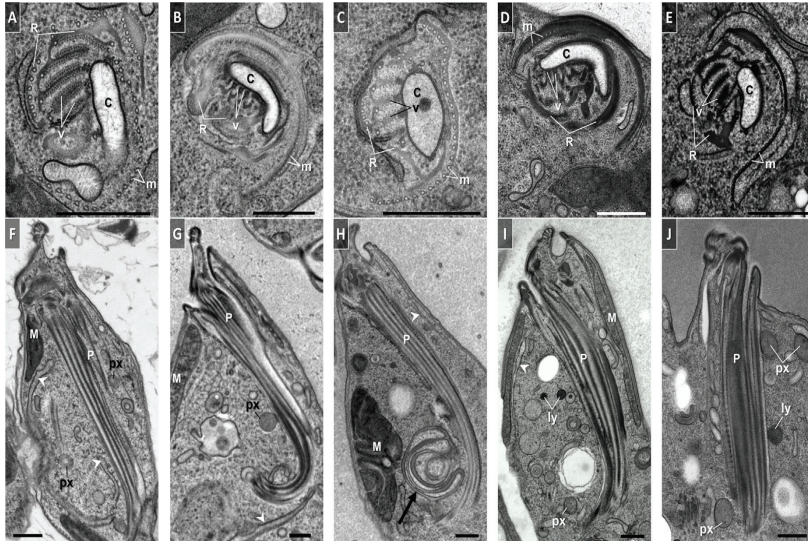


Figure 9. Ultrastructural architecture of the feeding apparatus in *R. humris* (A, F), *R. serpens* (B, G), *L. lanifica* (C, H), *S. specki* (D, I), and *F. neradi* (E, J). (A-E) Cross section showing the very top (B, D) and the intermediate part (A, C, E) of cytostome (C), which is the opening into the pharynx. Pharynx consists of a series of fibrils, the vanes (v), arranged in the form of a partial rosette, and supported by rods (R) and a row of microtubules (m). (F-J) Longitudinal section through the pharynx, which has a horn-like structure. The ribs forming the rosettes appear as a series of longitudinally-oriented dense lines. Note large convoluted tubular vessels (arrow) at the proximal end of pharynx (P) in H. Arrowheads indicate endoplasmic reticulum. The microbodies, numerous membrane-bounded vesicles, in early stages of formation surrounding the pharynx in the cytoplasm: peroxisomes at px; lysosomes at ly. Note portions of reticulated mitochondrion (M) surrounding the feeding apparatus and the density of ribosomes in the cytoplasm. Scale bar = 0.5 μ m.

transversely rather than longitudinally. Other conspicuous cellular features of diplomonads are a prominent Golgi apparatus, which is much larger than in the sister kinetoplastids (Han et al. 2013) but similar to euglenids (Becker and Melkonian 1996; Leedale 1982), and the exceptionally large nucleoli within the vesicular nuclei.

The architecture of the cell apex equipped with characteristic flagellar and feeding apparatuses is nearly identical in all studied diplomonads. The flagellar complex has a basic configuration of two nearly parallel basal bodies and asymmetrically distributed microtubular roots (Elbrächter et al. 1996; Montegut-Felkner and Triemer 1994, 1996; Schnepf 1994). Such arrangement was also observed in bodonids (Brugerolle et al. 1979; Brugerolle 1985) and euglenids (Kivic and Walne 1984). In all diplomonads, the characteristic feeding apparatus is arranged parallel to the longitudinal cell axis and consists of fibrillar ribs arranged in the form of a partial rosette, supported

by rod bundles and accompanying microtubules, which extend through the length of the pharynx (Montegut-Felkner and Triemer 1996; Porter 1973; Schuster et al. 1968). Another common feature is a conspicuous collar-like cytostome with an apical papillum, contiguous with the flagellar pocket (Larsen and Patterson 1990; Montegut-Felkner and Triemer 1996). The available SEM and TEM images clearly document two separate openings for the feeding apparatus and flagellar pocket in all diplomonads (Elbrächter et al. 1996; Montegut-Felkner and Triemer 1994, 1996; Porter 1973; Schuster et al. 1968; Triemer and Farmer 1991; this work), including *R. euleeides*, in which these were incorrectly interpreted as merged into a single opening (Roy et al. 2007). A feature distinguishing *R. humris* from the other diplomonads described herein is the presence of dense hair coat on both flagella and inside the cytostome and flagellar pocket. Flagellar hairs, although of varying fine structure, have been described in the diplomon-

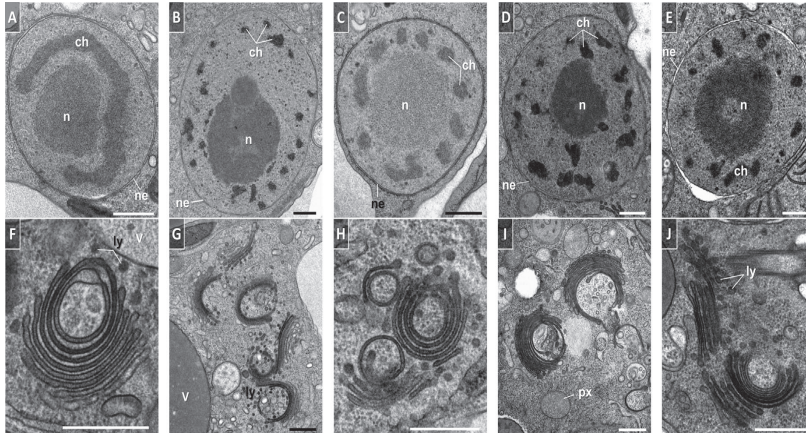


Figure 10. Transmission electron micrographs of nucleus and Golgi apparatus in *R. humris* (A, F), *R. serpens* (B, G), *L. lanifica* (C, H), *S. specki* (D, I), and *F. neradi* (E, J). (A-E) Section through large nuclei of vesicular pattern, with nuclear envelope (ne). Prominent centrally or eccentrically located nucleolus is surrounded by dense aggregations of chromatin (ch). (F-J) Cross sections of Golgi apparatus, which appears in either linear and/or circularized profiles, and is usually located adjacent to the nucleus/vacuoles (V). Peroxisomes at px; lysosomes at ly. Scale bar = 0.5 μ m.

mid *H. phaeocysticola* (Elbrächter et al. 1996), and the kinetoplasts of the genera *Bodo* (Eyden 1977), *Cryptobia* (Vickerman 1977), *Rhynchobodo* (Brugerolle 1985) and *Phyllomitus* (Mylnikov 1986).

Members of the genus *Rhynchopus* studied so far produce two distinct stages during their life cycle: (i) big non-flagellated trophic cells with prominent cytoplasmic granulation and/or digestive vacuoles, which move by gliding, and (ii) smaller, fully motile stages equipped with two long flagella, which usually lack digestive vacuoles and cytoplasmic inclusions. Both stages were observed in *R. euleeides* (Roy et al. 2007), *R. serpens* and *R. humris* (this work), *Rhynchopus* sp. ATCC 50230 (Simpson 1997) and in several other *Rhynchopus* species (von der Heyden et al. 2004), whereas for *R. coscinodiscivorus* (Schnepf 1994) and an *Isonema*-like (a junior synonym of *Diplonema*; Triemer and Ott, 1990) flagellate (Kent et al. 1987), only a trophic stage has been described. This can be explained by the fact that no observations were made under starvation conditions, which are known to trigger the emergence of the swimming stages (von der Heyden et al. 2004; this study). Since all other known diplomemid species have two permanently protruding flagella (Elbrächter et al. 1996; Larsen and Patterson 1990; Porter 1973; Schuster et al. 1968; Triemer and Ott 1990; this work), we

suggest that the existence of non-flagellated trophic stages is a discriminatory morphological feature for the genus *Rhynchopus*.

Flagellar stubs in trophic *Rhynchopus* cells often lack ordered axonemes with 9+2 microtubular arrangement (Schnepf 1994; Simpson 1997; this work). However, long flagella in their respective swimming stages develop regular axonemes accompanied by paraflagellar rods, which likely helps hydrodynamic propulsion (Hughes et al. 2012). This was shown for *Rhynchopus* sp. ATCC 50230 (Simpson 1997), *R. humris* and *R. serpens* (this work), while ultrastructural data are missing for swimming stages of *R. euleeides* (Roy et al. 2007), *R. coscinodiscivorus* (Schnepf 1994) and the *Isonema*-like flagellate (Kent et al. 1987). Paraflagellar rods are absent in trophic stages of *D. nigricans* (Schuster et al. 1968), *D. papillatum* (Porter 1973), and *D. ambulator* (Montegut-Felkner and Triemer 1994) in agreement with the apomorphy suggested for diplomemids by Adl et al. (2012). Here, we show that the flagella in trophic *F. neradi*, *S. specki* and *L. lanifica* cells are paraflagellar rod-bearing. While the latter diplomemid is capable of fast swimming under starvation conditions, the functions of flagella in *F. neradi* and *S. specki* should be further investigated, since these are used neither for substrate attachment nor for cell propulsion. However, we might have failed to find conditions

under which cells with long swimming flagella are produced.

Diplonemids studied herein can be distinguished by their characteristic morphology under nutrient-rich conditions. However, in every species the size, shape, length of flagella and cytoplasmic granulation are subject to high variability when nutrients are limited. Hence, species determination in natural samples or under different conditions may be difficult or even impossible. Distantly related species may exhibit similar morphology as in case of *L. lanifica* and the 4sb cell belonging to the Eupelagonemidae clade (Gawryluk et al. 2016; Okamoto et al., submitted; this work). The same is true for the swimming stages of different *Rhynchopus* species (Roy et al. 2007; this work). At the same time, the subcellular organization is strikingly similar among the representatives of different diplonemid genera as described in this work and elsewhere (Roy et al. 2007; Schnepf 1994; Triemer and Ott 1990) and hence, cannot aid species identification. Consequently, the determination and systematics of diplonemids, at least at the species level, will rather have to rely on the sequence data.

Due to little interspecies difference among *Rhynchopus* and their great morphological variability depending on nutrient availability, we argue that identification of species should not be based solely on morphological traits, and that observations under both nutrient-rich and starvation conditions are needed. The freshwater lifestyle of *R. amitus* suggests that this species is different from *R. humris* and *R. serpens*. The size and shape of *R. coscinodiscivorus* (pear-shaped; 20-25 × 10-12 μm) are similar to those of *R. serpens* (elongated, pear-shaped; 21.3-28.7 × 7.2-9 μm) but the 18S rRNA sequence is unavailable for the former species, precluding their more detailed comparison, whereas parasitizing or predation on diatoms is not known for *R. serpens*. Nevertheless, based on the cell morphology, we conclude that *L. lanifica*, *F. neradi* and *S. specki* are different species from previously described non-sequenced diplonemids. All three species are noticeably smaller, even under nutrient-rich conditions, than *D. metabolicum* (30-48 μm long) and *D. nigricans* (40-50 μm long), and never produce conical cells with broad anterior and tapered posterior ends (Schuster et al., 1968; Larsen and Patterson 1990). The length of their flagella when in the trophic phase readily distinguishes them from all described *Rhynchopus* species.

All diplonemids described in this study were isolated as free-swimming cells from water samples. However, a *Rhynchopus* isolate with 99% of

18S rRNA sequence identity to *R. humris* (Fig. 1), was reported to cause massive infections of the *Nephrops norvegicus* lobster (von der Heyden et al. 2004). Since our culture was established from a free-swimming motile cell, we propose that it may represent a dispersive and/or invasive stage, while the trophic stage develops under nutrient-rich conditions, such as are inside the host. It seems that parasitism, at least transient, as well as predation, are common life strategies for members of the genus *Rhynchopus*. Indeed, several species were found to parasitize crabs, lobsters and clams (Bodammer and Sawyer 1981; Kent et al. 1987; von der Heyden et al. 2004), while *R. coscinodiscivorus* predares/parasitizes on diatoms by invading the frustule (Schnepf 1994). So far, the literature refers to diplonemids mostly as to parasites or epibionts of plants and invertebrates or as predators of algae rather than bacterivores or detritus feeders (Vickerman 2000). Specifically, *D. ambulator* is involved in *Cryptocoryne* plant disease (Triemer and Ott 1990), *D. papillatum* is associated with drifting eelgrass (Porter 1973), *D. metabolicum* feeds on leaves of the *Halophila* seagrass (Larsen and Patterson 1990), and *H. phaeocysticola* feeds on diatoms, dinoflagellates and copepods (Elbrächter et al. 1996). Alternatively, *D. ambulator* and *R. euleeides* may also graze bacteria and ingest detritus particles (Larsen and Patterson 1990; Roy et al. 2007). However, new extensive sequence data from the pelagic samples (Flegontova et al. 2016; Gawryluk et al. 2016) point to the free-living heterotrophic life style of most species.

Taxonomic Summary

Phylum Euglenozoa Cavalier-Smith 1981, emend. Simpson 1997

Class Diplonemea Cavalier-Smith 1993, emend. Simpson 1997

Genus *Rhynchopus* Skuja (1948)

Rhynchopus humris sp. n. Tashyreva, Prokopchuk, Horák and Lukeš (2018)

Description: trophic cells elongated, dorsoventrally flattened, narrowed at both ends; 12.5-16.4 μm (14.5 ± 1.2 μm; n=25) long and 3.3-5.1 μm (4.2 ± 0.59 μm; n=25) wide; prominent granulation throughout cytoplasm. Short flagellar stubs with disordered axonemes, lacking paraflagellar rods, concealed inside flagellar pocket. Highly metabolic, move by gliding on surfaces. Fast

swimming stage in starved cultures; dimensions $7.2 \times 3.2 \mu\text{m}$; dorsoventrally compressed; cytoplasmic inclusions absent. Flagella up to 2.5 times of body length, inserted subapically, both with regular axonemes and paraflagellar rods. Anterior flagellum loops around cell apex; posterior flagellum stretched. Nucleus in central part proximally to cell periphery; mitochondrial cristae arranged longitudinally; flagella pocket hairs. No cysts and extrusomes.

Etymology: The species name is a Czech word for lobster, which is the host of a species with 99% 18S rRNA sequence identity to YPF1608.

Type strain: YPF1608

Type material: hapantotype (OsO₄-fixed slide) and genomic DNA sample deposited at the protistological collection of the Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, no. IPCAS Prot 37

Gene sequence: MF422204, 18S rRNA gene, partial sequence.

Type locality: sand filter of 'D-1' tank in Enoshima Aquarium, Fujisawa, Kanagawa, Japan.

Rhynchopus serpens sp. n. Tashyreva, Prokopchuk, Horák and Lukeš (2018)

Description: trophic cell cells elongated, dorsoventrally flattened, tapered at the anterior, rounded and widened posteriorly; numerous cytoplasmic inclusions in posterior half; $21.3\text{--}28.7 \mu\text{m}$ ($26.0 \pm 1.7 \mu\text{m}$; $n=25$) long and $7.2\text{--}9 \mu\text{m}$ ($8.0 \pm 0.6 \mu\text{m}$; $n=25$) wide. Short flagellar stubs with disordered axonemes, lacking paraflagellar rods, concealed inside flagellar pocket. Highly metabolic, move by gliding on surfaces. Swimming stage in starved cultures small ($13.5 \times 3.7 \mu\text{m}$), cytoplasmic granulation reduced or absent, dorsoventrally compressed; slow swimming with both flagella stretched along body or fast swimming cells with anterior flagellum forming loop and posterior flagellum stretched. Flagella up to 2.5 times of body length, inserted subapically, both with regular axonemes and paraflagellar rods. Nucleus position variable; mitochondrial cristae arranged longitudinally. No cysts and extrusomes.

Etymology: The species name describes the crawling movement of the cells.

Type strain: YPF1515

Type material: hapantotype (OsO₄-fixed slide) and genomic DNA sample deposited at the protistological collection of the Institute of Parasitology,

Biology Centre, Czech Academy of Sciences, České Budějovice, no. IPCAS Prot 38

Gene sequence: MF422195, 18S rRNA gene, partial sequence.

Type locality: Kaiike-lagoon, Koshikijima Island, Satsuma-Sendai, Kagoshima, Japan ($31^{\circ}51'36''\text{N}$, $129^{\circ}52'29''\text{E}$), surface water.

Phylum Euglenozoa Cavalier-Smith 1981, emend. Simpson 1997

Class Diplonemea Cavalier-Smith 1993, emend. Simpson 1997

Genus ***Lacrimia*** gen. n. Tashyreva, Prokopchuk, Horák and Lukeš (2018)

Marine diplomemids distinguished from similar genera by molecular phylogenetic analyses.

Type species: *Lacrimia lanifica*.

Lacrimia lanifica sp. n. Tashyreva, Prokopchuk, Horák and Lukeš (2018)

Description: cells in well growing cultures nearly isodiametric but often teardrop-shaped, large digestive vacuoles at posterior end, narrowed anterior end; $10.4\text{--}15.3 \mu\text{m}$ ($13.1 \pm 1.3 \mu\text{m}$; $n=25$) long and $7.3\text{--}10.5 \mu\text{m}$ ($8.7 \pm 0.74 \mu\text{m}$; $n=25$) wide. Two subapical subequal body-length flagella, with regularly arranged axonemes and paraflagellar rods. Cells homogeneously suspended, metabolic, maintaining constant rotation movements; gliding along surface by attaching with longer flagellum. Cells in aged and starved cultures elongated and small ($7.5 \times 3.6 \mu\text{m}$), lacking food vacuoles. Occasional fast swimming cells move in straight line through spiral oscillating movement. Nucleus in anterior half; mitochondrial cristae arranged longitudinally. No cysts and extrusomes produced.

Etymology: The name (feminine) reflects the tear-drop shape (lacrima) and spinning movement (lanifica)

Type strain: YPF1601

Type material: hapantotype (OsO₄-fixed slide) and genomic DNA sample deposited at the protistological collection of the Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, no. IPCAS Prot 39

Gene sequence: MF422199, 18S rRNA gene, partial sequence

Type locality: Tokyo Bay ($35^{\circ}19'10''\text{N}$, $139^{\circ}39'04''\text{E}$), surface water.

Phylum Euglenozoa Cavalier-Smith 1981, emend.

Simpson 1997

Class Diplonemea Cavalier-Smith 1993, emend. Simpson 1997

Genus *Sulcionema* gen. n. Tashyreva, Prokopchuk, Horák and Lukeš (2018)

Marine diplomemids distinguished from similar genera by molecular phylogenetic analyses.

Type species: *Sulcionema specki*.

Sulcionema specki sp. n. Tashyreva, Prokopchuk, Horák and Lukeš (2018)

Description: cells in nutrient-rich medium noticeably flattened with tapered, slightly crooked anterior along with rounded posterior ends, conspicuous cytoplasmic granulation; 23.7-33.3 μm ($27.6 \pm 2.1 \mu\text{m}$; $n=25$) long and 5.8-9.6 μm ($7.4 \pm 0.9 \mu\text{m}$; $n=25$) wide. Two equal to subequal flagella third of body length, with paraflagellar rods and regular axonemes. Cells suspended in medium, highly metabolic, flagella support only erratic movement. Starved cells reduce size to $16.3 \times 5 \mu\text{m}$, morphologically variable, cytoplasmic granulation reduced or absent. Swimming stage, cysts and extrusomes not observed. Nucleus mostly in posterior half; mitochondrial cristae arranged longitudinally.

Etymology: The generic name (neuter) reflects in Latin the sausage-like shape of the cells, and the species name in German ("bacon") refers to lipid-like cell inclusions.

Type strain: YPF1618

Type material: hapantotype (OsO₄-fixed slide) and genomic DNA sample deposited at the protistological collection of the Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, no. IPCAS Prot 40

Gene sequence: MF422201, 18S rRNA gene, partial sequence.

Type locality: Oura Beach, Shikinejima Island, Tokyo, Japan (34°19'48"N, 139°12'29"E), surface water.

Phylum Euglenozoa Cavalier-Smith 1981, emend. Simpson 1997

Class Diplonemea Cavalier-Smith 1993, emend. Simpson 1997

Genus *Flectonema* gen. n. Tashyreva, Prokopchuk, Horák and Lukeš (2018)

Marine diplomemids distinguished from similar genera by molecular phylogenetic analyses. Type

species: *Flectonema neradi*.

Flectonema neradi gen. n., sp. n. Tashyreva, Prokopchuk, Horák and Lukeš (2018)

Description: cells in nutrient-rich medium dorsoventrally flattened, constricted at both ends, thin and crooked, with granulation in the posterior region; 15.9-20.9 μm ($18.3 \pm 1.45 \mu\text{m}$; $n=25$) long and 3.5-5.8 μm ($4.0 \pm 0.6 \mu\text{m}$; $n=25$) wide; slowly glide on surfaces, display pronounced metaboly, attach to surfaces with one flagellum and rotate. Two equal subapical flagella fifth of body length, both with regular axonemes and paraflagellar rods. No fast swimming stages or cysts in old batch cultures and starvation medium. Starved cells short and rounded, some with constricted posterior ends, $6.3 \times 3.1 \mu\text{m}$, lack cytoplasmic inclusions. Nucleus position variable; mitochondrial cristae arranged transversely; opening of flagellar pocket extended into curved groove. Extrusomes not observed.

Etymology: The name (neuter) is derived from the bent shape (flecto-), the species name is after Thomas Nerad, the isolator of the strain.

Type strain: ATCC50224

Type material: hapantotype (OsO₄-fixed slide) and genomic DNA sample deposited at the protistological collection of the Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, no. IPCAS Prot 41.

Gene sequence: AF119812, 18S rRNA gene, partial sequence.

Type locality: Gaithersburg, Maryland, USA.

Methods

Isolation and cultivation: Samples from aquaria, lagoon and sandy beach seawater were collected around Japan. The details on the collected samples are summarized in the Supplementary Material Table SA2. The samples were inoculated into seawater-based Hemi medium (designed in this study), containing 3.6% sea salts (Sigma-Aldrich), enriched with 1% (v/v) heat-inactivated horse serum (Sigma-Aldrich) and 0.025 g/l LB broth powder (Amresco). The medium was supplemented with 10 $\mu\text{l/ml}$ antibiotics cocktail (P4083, Sigma-Aldrich) and sterilized by filtering through a 0.22 μm filter. The samples were then incubated at 20 °C. Possible diplomemid cells were isolated under the inverted microscope CKX31 (Olympus) using glass microcapillaries. In total, 18 strains from isolated single cells were established. A culture of *Diplonema* sp. 2 (ATCC 50224) was obtained from ATCC. All cultures were grown axenically at 15 or 20 °C in antibiotic-free Hemi medium. Additionally, for some experiments, the cultures were starved by incubation in a medium diluted 1:10 with seawater for up to 10 days.

DNA isolation, PCR amplification and sequencing: Total genomic DNA was isolated from cultures with DNeasy Blood & Tissue Kit (Qiagen) following the protocol A as described by the manufacturer. The nearly full-size 18S rRNA gene was amplified with universal eukaryotic primers SA (5'-AACCTGGTTGATCCTGCCAGT-3') and SB (5'-TGATCCTCCTGCAGGTCACCT-3'), the amplicons were purified and sequenced.

Phylogenetic analyses: 18S rRNA sequences of novel diplomid species described here were added to the exhaustive dataset of euglenozoan 18S rRNA extracted from public databases using the EukRef approach (eukref.org). The dataset was aligned together with newly sequenced cultures using the genepair algorithm as implemented in Mafft v7.305b (Katoch and Standley 2013). Ambiguous regions were removed by eye in Seaview 4 (Gouy et al. 2010). At this step, we created three datasets differing by outgroup composition. In dataset K, only kinetoplastids were used as an outgroup. This dataset composed of 133 taxa and 2000 nucleotide sites. In datasets E (149 taxa and 1835 sites) and H (152 taxa, 1835 sites) outgroups were expanded by the addition of euglenids and heteroloboseans and euglenids, respectively. The best fitting model of evolution for each dataset, as well as maximum likelihood phylogenies under these selected models, were estimated using IQ-Tree (Nguyen et al. 2015) with thorough non-parametric bootstrap analysis of 1,000 replicates as a measure of branching support. Robustness of observed topologies was also tested using Bayesian posterior probabilities as inferred by Phylobayes 4.1 (Lartillot et al. 2009) under the empirical mixture model C40 combined with exchange rates as defined by GTR matrix (C40 + GTR model). Two independent MCMC chains were ran until convergence was reached (i.e. maximum observed discrepancy was lower than 0.1 and effective sample size of observed statistics was at least 100). Alternative hypotheses on topologies of selected taxa (see Results) were tested using an approximately unbiased (AU) test in Consel (Shimodaira and Hasegawa 2001). For this, we first forced alternative topologies, re-optimized the ML tree and computed per-site log-likelihood scores in RAxML 8.28b (Stamatakis 2014).

Light microscopy: For imaging, live cells were trapped between a slide and coverslip with edges sealed either with nail polish or immobilized by suspending in 0.5 to 1% ultralow gelling agarose solution (Sigma). Microscopy slides for agarose-immobilized cells were prepared according to Reize and Melkonian (1989). Light microscopy observations were performed with either Olympus BX53 equipped with differential interference contrast (DIC) or Zeiss Primovert. Videos and images were taken with a DP72 microscope digital camera at 1600 × 1200-pixel resolution using CellSens software v. 1.11 (Olympus). The images were processed using GIMP v. 2.8.14, Irfan view v. 4.41 and Image J v. 1.51 software.

Fluorescence staining and microscopy: For DNA staining, cell pellets were fixed with seawater-based Parduć fixative consisting of 6:1 volumes of 2% OsO₄ and saturated HgCl₂ solutions (Parduć 1967) for 15 min, and washed thoroughly with distilled water. Cell suspensions were applied on glass slides, air-dried, and mounted in ProLong Gold antifade

reagent with 4',6-diamidino-2-phenylindole, or DAPI (Life Technologies). For mitochondrial staining, live cells in serum-free medium were treated for 30 min with (i) 100 or 200 nM MitoTracker Green FM, (ii) 100 or 500 nM MitoTracker Red CMXRos, (iii) 60 nM tetramethylrhodamine ethyl ester (TMRE), and (iv) 2, 6, 10 or 20 μM DiOC₆(3) either with or without addition of 5% v/v dimethyl sulfoxide (DMSO) (all dyes Life Technologies, USA). Slides were prepared as described above and observed with an AxioPlan 2 fluorescence microscope (Zeiss). In addition, we attempted to visualize mitochondria with immunofluorescence assay targeting mitochondrial heat shock protein (HSP) 70 with antibodies generated against *Trypanosoma brucei* HSP70, as described in Ziková et al. (2009).

Electron microscopy: For scanning electron microscopy (SEM), pellets were fixed with the above-described Parduć fixative (pH 7.4) either alone or followed by instantaneous freezing in liquid nitrogen and freeze-drying at -60 °C under high vacuum for 12 hours following a procedure described elsewhere (Small and Marszałek 1969). Alternatively, fixed cells were adhered to poly-L-lysine coated glass coverslips and dehydrated in an increasing gradient of acetone (30% to 100%), in sets of 15 min, and critical point dried using CO₂. Dehydrated samples were then transferred onto SEM specimen stubs, coated with gold/palladium in Sputter Coater Polaron chamber, and examined using a JEOL JSM-7401-F microscope at an accelerating voltage of 4 kV. Conventional fixation procedure with 4% paraformaldehyde or 2.5% glutaraldehyde followed by the dehydration in ethanol series could not be used, as it resulted in near-complete disintegration of cells. For transmission electron microscopy (TEM), samples were prepared by high pressure freezing technique (HPF) as described previously (Yurchenko et al. 2014). Ultrathin sections were observed in a JEOL 1010 TEM microscope at accelerating voltage of 80 kV. Images were captured with an Olympus Mega View III camera.

Acknowledgements

This work was supported by the European Research Council CZ LL1601 (to JL), the Czech Grant Agency projects Nos. P506-12-P9 (to AH) and 14-23986S (to JL), Czech-Biolmaging RI project (LM2015062 funded by MEYS CR) and grant from the Japanese Society for the Promotion of Science No. 26840133 (to AY). We thank Dr. S Tsuchida (JAMSTEC) and Messers M Kawato (JAMSTEC), H Kishi (Satsumasendai City Office), Y Suzuki, S Nemoto and M Sugimura (Enoshima Aquarium) for providing a part of the samples.

Appendix A.

Table A1A. Trophic cells in nutrient-rich medium.

Species	<i>R. humris</i>	<i>R. serpens</i>	<i>L. lanifica</i>	<i>S. specki</i>	<i>F. neradi</i>
Cell shape	elongated, terminally narrowed	elongated, acute anterior and round posterior ends	nearly round or teardrop-shaped	elongated, laterally flattened	elongated, posteriorly rounded, bent
Cell size, μm^a	12.5-16.4 \times 3.3-5.1	21.3-28.7 \times 7.2-9	10.4-15.3 \times 7.3-10.5	23.7-33.3 \times 5.8-9.6	15.9-20.9 \times 3.5-5.8
Flagella	stubs, buried in flagellar pocket	stubs, buried in flagellar pocket	subequal, as body length	subequal, a third of body length	equally long, a fifth of body length
Movement	gliding, metabolic	gliding, metabolic	rotation, gliding with flagellum, metabolic	erratic, metabolic	gliding, rotation, metabolic
Cytoplasmic inclusions	big vesicles	big posterior vesicles	big posterior vacuole	conspicuously big vesicles	posterior vesicles

^ameasured in extended state.**Table A1B.** Cells after starvation treatment.

Species	<i>R. humris</i>	<i>R. serpens</i>	<i>L. lanifica</i>	<i>S. specki</i>	<i>F. neradi</i>
Cell shape	elongated, terminally narrowed	elongated with both ends rounded	oval	elongated with both ends rounded	elongated or rounded
Cell size, μm^a	7.2 \times 3.2	13.5 \times 3.7	7.5 \times 3.6	16.3 \times 5	6.3 \times 3.1
Flagella	unequally long, 2 times of body length	unequally long, 2 to 2.5 times of body length	unequally long, 1 to 2.5 times of body length	subequal, a third of body length	equally long, a fifth of body length
Movement	fast swimming, metabolic	fast swimming, metabolic	rotation, gliding, metabolic, swimming	erratic, metabolic	gliding, rotation, metabolic
Cytoplasmic inclusions	absent	absent	absent	small or absent	small or absent

^asmallest measured cells.

Appendix B. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.protis.2018.02.001>.

References

- 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 EAD, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick L, Shadwick L, Schoch CL, Smirnov A, Spiegel FW (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* **59**:429–493
- Arroyo M, Heltai L, Millan D, DeSimone A (2012) Reverse engineering the euglenoid movement. *Proc Natl Acad Sci USA* **109**:17874–17879
- Becker B, Melkonian M (1996) The secretory pathway of protists: spatial and functional organization and evolution. *Microbiol Rev* **60**:697–721
- Bodammer JE, Sawyer TK (1981) Aufwuchs protozoa and bacteria on the gills of the rock crab, *Cancer irroratus* say: a survey by light and electron microscopy. *J Protozool* **28**:35–46
- Brugerolle G (1985) Des trichocystes chez les bodonides, un caractère phylogénétique supplémentaire entre Kinetoplastida et Euglenida. *Protistologica* **21**:339–348
- Brugerolle G, Lom J, Nohýnkova E, Joyon L (1979) Comparaison et évolution des structures cellulaires chez plusieurs espèces de Bodonidés et Cryptobiidés appartenant aux genres *Bodo*, *Cryptobia* et *Trypanoplasma* (Kinetoplastida, Mastigophora). *Protistologica* **15**:197–221
- Busse I, Preisfeld A (2002) Phylogenetic position of *Rhynchopus* sp. and *Diplonema ambulator* as indicated by analyses of euglenozoan small subunit ribosomal DNA. *Gene* **284**: 83–91

- David V, Archibald JM (2016) Evolution: plumbing the depths of diplomemid diversity. *Curr Biol* **26**:R1290–R1292
- 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, Tara Oceans Coordinators, Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E (2015) Ocean plankton. *Eukaryotic plankton diversity in the sunlit ocean*. *Science* **348**:1261605
- Elbrächter M, Schnepf E, Balzer I (1996) *Hemistasia phaeocysticola* (Scherffel) comb. nov., redescription of a free-living, marine, phagotrophic kinetoplastid flagellate. *Arch Protistenkd* **147**:125–136
- Eloe EA, Shulze CN, Fadrosch 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
- Eyden BP (1977) Morphology and ultrastructure of *Bodo designis* Skuja 1948. *Protistologica* **13**:169–179
- Faktorová D, Dobáková E, Peña-Díaz P, Lukeš J (2016) From simple to supercomplex: mitochondrial genomes of euglenozoan protists. *F1000Research* **5**:392
- Flegontova O, Flegontov P, Malviya S, Audic S, Wincker P, de Vargas C, Bowler C, Lukeš J, Horák A (2016) Extreme diversity of diplomemid eukaryotes in the ocean. *Curr Biol* **26**:3060–3065
- Gawryluk RMR, Del Campo J, Okamoto N, Strasser JFH, Lukeš J, Richards TA, Worden AZ, Santoro AE, Keeling PJ (2016) Morphological identification and single-cell genomics of marine diplomemids. *Curr Biol* **26**:3053–3059
- Gouy M, Guindon S, Gascuel O (2010) SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**:221–224
- Griessmann K (1913) Über marine Flagellaten. *Arch Protistenkd* **32**:1–78
- Han HM, Bouchet-Marquis C, Huebinger J, Grabenbauer M (2013) Golgi apparatus analyzed by cryo-electron microscopy. *Histochem Cell Biol* **140**:369–381
- Hayashi Y, Ueda K (1989) The shape of mitochondria and the number of mitochondrial nucleoids during the cell cycle of *Euglena gracilis*. *J Cell Sci* **93**:565–570
- Hughes LC, Ralston KS, Hill KL, Zhou ZH (2012) Three-dimensional structure of the trypanosome flagellum suggests that the paraflagellar rod functions as a biomechanical spring. *PLoS ONE* **7**:e25700, <http://dx.doi.org/10.1371/journal.pone.0025700>
- Jensen RE, Englund PT (2012) Network news: the replication of kinetoplast DNA. *Annu Rev Microbiol* **66**:473–491
- Jeuck A, Arndt H (2013) A short guide to common heterotrophic flagellates of freshwater habitats based on the morphology of living organisms. *Protist* **164**:842–860
- Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software version 7: improvements in performance and usability. *Mol Biol Evol* **30**:772–780
- Kent ML, Elston RA, Nerad TA, Sawyer TK (1987) An *Isonema*-like flagellate (Protozoa: Mastigophora) infection in larval geoduck clams, *Panope abrupta*. *J Invertebr Pathol* **50**:221–229
- Kiethega GN, Yan Y, Turcotte M, Burger G (2013) RNA-level unscrambling of fragmented genes in *Diplonema* mitochondria. *RNA Biol* **10**:301–313
- Kivic PA, Walne PL (1984) An evaluation of a possible phylogenetic relationship between the Euglenophyta and Kinetoplastida. *Orig Life* **13**:269–288
- Lara E, Moreira D, Vereshchaka A, López-García P (2009) Pan-oceanic distribution of new highly diverse clades of deep-sea diplomemids. *Environ Microbiol* **11**:47–55
- Larsen J, Patterson DJ (1990) Some flagellates (Protista) from tropical marine sediments. *J Nat Hist* **24**:801–937
- Lartillot N, Lepage T, Blanquart S (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**:2286–2288
- Leedale GF (1982) Ultrastructure. In Buetow D (ed) *The Biology of Euglena*. Academic Press, New York, pp 1–27
- 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, Flegontova O, Horák A (2015) Diplomemids. *Curr Biol* **25**:R702–R704
- Lukeš J, Guilbride DL, Votýpka J, Zíková A, Benne R, Englund PT (2002) 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 diplomemids, 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
- Massana R (2011) Eukaryotic picoplankton in surface oceans. *Annu Rev Microbiol* **65**:91–110
- Montegut-Felkner AE, Triemer RE (1994) Phylogeny of *Diplonema ambulator* (Larsen and Patterson): 1. Homologies of the flagellar apparatus. *Europ J Protistol* **30**:227–237
- Montegut-Felkner AE, Triemer RE (1996) Phylogeny of *Diplonema ambulator* (Larsen and Patterson): 2. Homologies of the feeding apparatus. *Europ J Protistol* **32**:64–76
- Moreira D, López-García P, Rodríguez-Valera F (2001) New insights into the phylogenetic position of diplomemids: G + C content bias, differences of evolutionary rate and a new environmental sequence. *Int J Syst Evol Microbiol* **51**:2211–2219

- 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
- Mylnikov AP** (1986) Ultrastructure of a colourless flagellate, *Phyllomitus apiculatus* Skuja 1948 (Kinetoplastida). *Arch Protistenkd* **132**:1–10
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ** (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**:268–274
- Okamoto N, Gawryluk RMR, del Campo J, Strasser JFH, Lukeš J, Richards TA, Worden AZ, Santoro AE, Keeling PJ** (2018) *Eupelagonema oceanica* n. gen. & sp. and a revised diplomemid taxonomy. *J Eukaryot Microbiol* (in press)
- Párducz B** (1967) Ciliary movement and coordination in ciliates. *Int Rev Cytol* **21**:91–128
- Porter D** (1973) *Isonema papillatum* sp. n., a new colorless marine flagellate: a light- and electronmicroscopic study. *J Protozool* **20**:351–356
- Reize IB, Melkonian M** (1989) A new way to investigate living flagellated/ciliated cells in the light microscope: immobilization of cells in agarose. *Bot Acta* **102**:145–151
- Roy J, Faktorová D, Benada O, Lukeš J, Burger G** (2007) Description of *Rhynchopus euleeides* n. sp. (Diplonemea), a free-living marine euglenozoan. *J Eukaryot Microbiol* **54**:137–145
- Shimodaira H, Hasegawa M** (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**:1246–1247
- Scheckenbach F, Hausmann K, Wylezich C, Weitere M, Arndt H** (2010) Large-scale patterns in biodiversity of microbial eukaryotes from the abyssal sea floor. *Proc Natl Acad Sci USA* **107**:115–120
- Schnepf E** (1994) Light and electron microscopical observations in *Rhynchopus coccinodiscivorus* spec. nov., a colorless, phagotrophic Euglenozoan with concealed flagella. *Arch Protistenkd* **144**:63–74
- Schuster F, Goldstein S, Hershenv B** (1968) Ultrastructure of a flagellate, *Isonema nigricans* nov. gen. nov. sp., from a polluted marine habitat. *Protistologica* **4**:141–154
- Simpson AGB** (1997) The identity and composition of the Euglenozoa. *Arch Protistenkd* **148**:318–328
- Skuja H** (1948) Taxonomie des Phytoplanktons einiger Seen in Uppland, Schweden. *Symb Bot Ups* **9**:5–399
- Small EB, Marszalek DS** (1969) Scanning electron microscopy of fixed, frozen, and dried protozoa. *Science* **163**:1064–1065
- Stamatakis A** (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313
- Swale EMF** (1973) A study of the colourless flagellate *Rhynchomonas nasuta* (Stokes) Klebs. *Biol J Linn Soc* **5**:255–264
- Triemer RE** (1992) Ultrastructure of mitosis in *Diplonema ambulator* Larsen and Patterson (Euglenozoa). *Europ J Protistol* **28**:398–404
- Triemer RE, Farmer MA** (1991) An ultrastructural comparison of the mitotic apparatus, feeding apparatus, flagellar apparatus and cytoskeleton in euglenoids and kinetoplastids. *Protoplasma* **164**:91–104
- Triemer RE, Ott DW** (1990) Ultrastructure of *Diplonema ambulator* Larsen & Patterson (Euglenozoa) and its relationship to *Isonema*. *Europ J Protistol* **25**:316–320
- Vickerman K** (1977) DNA throughout the single mitochondrion of a kinetoplastid flagellate: observations on the ultrastructure of *Cryptobia vaginalis* (Hesse, 1910). *J Protozool* **24**:221–233
- Vickerman K** (2000) Diplonemids (Class: Diplonemea Cavalier Smith, 1993). In Lee JJ, Leedale GF, Bradbury P (eds) *An Illustrated Guide to the Protozoa Vol. 2*, 2nd edn, Society of Protozoologists, Lawrence, Kansas, USA, pp 1157–1159
- Vícek C, Marande W, Teijeiro S, Lukeš J, Burger G** (2011) Systematically fragmented genes in a multipartite mitochondrial genome. *Nucleic Acids Res* **39**:979–988
- von der Heyden S, Chao EE, Vickerman K, Cavalier-Smith T** (2004) Ribosomal RNA phylogeny of bodonid and diplomemid flagellates and the evolution of euglenozoa. *J Eukaryot Microbiol* **51**:402–416
- Yabuki A, Tame A** (2015) Phylogeny and reclassification of *Hemistasia phaeocysticola* (Scherffel) Elbrächter & Schnepf, 1996. *J Eukaryot Microbiol* **62**:426–429
- Yurchenko V, Votýpka J, Tesařová M, Klepetková H, Kraeva N, Jirků M, Lukeš J** (2014) Ultrastructure and molecular phylogeny of four new species of monoxenous trypanosomatids from flies (Diptera: Brachycera) with redefinition of the genus *Wallaceina*. *Folia Parasitol* **61**:97–112
- Ziková A, Schnaufer A, Dalley RA, Panigrahi AK, Stuart KD** (2009) The FOF1-ATP synthase complex contains novel subunits and is essential for pro-cyclic *Trypanosoma brucei*. *PLoS Pathog* **5**:e1000436, <http://dx.doi.org/10.1371/journal.ppat.1000436>

Available online at www.sciencedirect.com

ScienceDirect

Chapter 2

Transformation of *Diplonema papillatum*, the type species of the highly diverse and abundant marine micro-eukaryotes Diplonemida (Euglenozoa), Environmental Microbiology 20: 1030-1040.

Transformation of *Diplonema papillatum*, the type species of the highly diverse and abundant marine microeukaryotes Diplonemida (Euglenozoa)

Binnypreet Kaur,^{1,2} Matus Valach,³
Priscila Peña-Díaz,¹ Sandrine Moreira,³
Patrick J. Keeling,⁴ Gertraud Burger,³
Julius Lukeš^{1,2} and Drahomíra Faktorová^{1,2*}

¹Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice (Budweis), Czech Republic.

²Faculty of Sciences, University of South Bohemia, České Budějovice (Budweis), Czech Republic.

³Department of Biochemistry and Robert-Cedergren Centre for Bioinformatics and Genomics, Université de Montréal, Montreal, Canada.

⁴Botany Department, University of British Columbia, Vancouver, Canada.

Summary

Diplonema papillatum is the type species of diplomids, which are among the most abundant and diverse heterotrophic microeukaryotes in the world's oceans. Diplonemids are also known for a unique form of post-transcriptional processing in mitochondria. However, the lack of reverse genetics methodologies in these protists has hampered elucidation of their cellular and molecular biology. Here we report a protocol for *D. papillatum* transformation. We have identified several antibiotics to which *D. papillatum* is sensitive and thus are suitable selectable markers, and focus in particular on puromycin. Constructs were designed encoding antibiotic resistance markers, fluorescent tags, and additional genomic sequences from *D. papillatum* to facilitate vector integration into chromosomes. We established conditions for effective electroporation, and demonstrate that electroporated constructs can be stably integrated in the *D. papillatum* nuclear genome. In *D. papillatum* transformants, the heterologous puromycin resistance gene is transcribed into mRNA and

translated into protein, as determined by Southern hybridization, reverse transcription, and Western blot analyses. This is the first documented case of transformation in a euglenozoan protist outside the well-studied kinetoplastids, making *D. papillatum* a genetically tractable organism and potentially a model system for marine microeukaryotes.

Introduction

Diplonemids are biflagellate protists inhabiting marine ecosystems. Diplonemids, together with symbiontids (a small poorly studied group of species living in anoxic or low-oxygen marine environment, for which no molecular data are available [Breglia *et al.*, 2010]), euglenids (important component of the freshwater ecosystems), and kinetoplastids (including free-living *Bodo*, as well as the highly pathogenic *Trypanosoma* and *Leishmania* spp.), form Euglenozoa. Both the kinetoplastids and euglenids have long been recognized as ecologically omnipresent and species-rich, while diplomemids were considered an insignificant group both in terms of diversity and ecology, and they remained poorly-studied (Lukeš *et al.*, 2015).

The first diplomemid, *Diplonema breviciliata*, was described by Griessman in 1914 (Griessman, 1914), and only less than half a dozen species in two genera have been described since then (Skuja, 1948; Schuster *et al.*, 1968; Porter, 1973; Larsen and Patterson, 1990; Schnepf, 1994; Simpson, 1997; Roy *et al.*, 2007). More recent ecological surveys tracked down two clades of previously unrecognized diplomemids in deep-sea pelagic waters, notably the deep sea pelagic diplomemids clade I and II (DSPD I, II) (Lara *et al.*, 2009). In a comprehensive marine survey, diplomemids emerged even as the 3rd most diverse and 6th most abundant group of oceanic eukaryotes (de Vargas *et al.*, 2015; Lukeš *et al.*, 2015) populating virtually all stations and all depths examined in the comprehensive Tara Oceans expedition (de Vargas *et al.*, 2015; Flegontova *et al.*, 2016). Analysis of the V9 18S rRNA sequence suggests that diplomemids may even be the most species-rich among all marine eukaryotes (Flegontova *et al.*, 2016). At present, only diplomemids from the *Diplonema*/

Received 3 November, 2017; revised 13 December, 2017; accepted 15 December, 2017. *For correspondence. E-mail dranov@paru.cas.cz; Tel. +420 387775433; Fax +420 385310388.

© 2018 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Rhynchopus (D/R) clade can be grown axenically and are available from the American Type Culture Collection (ATCC). Representatives of the hyper-diverse DSPD clades are not yet in culture. An initial examination of 10 DSPD isolates has been performed by single-cell genomics (Gawryluk *et al.*, 2016).

Among the common structural features of diplomonids is their sac-like shape with cell sizes ranging from 5 to 50 μm , their highly metabolic movement, and the presence of two sub-apical flagella. They possess a single mitochondrial network lining the inside of the plasma membrane and containing a large amount of mitochondrial DNA (mtDNA) (Marande *et al.*, 2005). Whether diplomonids have a parasitic, commensal, or free-living life-style is still an open question. Hence the ecological role of diplomonids in the oceanic ecosystem remains elusive (Lukeš *et al.*, 2015). This important gap in knowledge reflects our ignorance about the ecological functions of marine protists in general (Worden and Wilken, 2016), and the lack of a suitable model system for diplomonids in particular.

Establishing a model organism requires several crucial steps to be fulfilled, the most prominent being their availability in culture, and the ability to genetically modify the species. Ideally, a genetic system includes a vector, a transformation system, and a selectable marker, all of which facilitate expression of introduced genes, including transcription, post-transcriptional processing, and translation. All these steps require extensive experimentation and optimization for each target organism, but once a protocol established, it is generally straightforward to create any range of constructs to address many functional questions.

Only a single diplomonid species, *Diplonema papillatum*, has been examined at the molecular level in much detail. *D. papillatum* is a free-living protist with two short heterodynamic flagella that was first isolated from seawater at Friday Harbor, Washington (Porter, 1973). This species is available from the American Type Culture Collection (ATCC), can be easily cultivated axenically in the laboratory at high cell densities, and can be cryopreserved. The mitochondrial genome and transcriptome have been analyzed in detail, revealing novel modes of post-transcriptional gene expression (Marande and Burger, 2007; Kiethega *et al.*, 2013; Moreira *et al.*, 2016; Valach *et al.*, 2016; Faktorová *et al.*, 2018). Further, *D. papillatum* has been investigated regarding its compartmentalization of gluconeogenesis (Makiuchi *et al.*, 2011; Morales *et al.*, 2016). Finally, the nuclear genome has recently been sequenced and assembled, and functional annotation is under way (unpubl. data).

For all these reasons, establishing a tractable genetic system for *D. papillatum* and establishing it as a model system would be highly desirable. Here, we describe all key steps necessary for genetic manipulation of this marine flagellate.

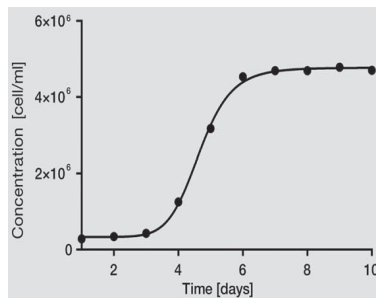


Fig. 1. The growth curve of *D. papillatum* cells during 10 days in *Diplonema* cultivation medium. Inoculation titer at time 0 was 2×10^5 cells/ml.

Results

Optimization of cultivation and electroporation conditions

In the standard seawater medium (see Experimental Procedures), *D. papillatum* cultures reach the exponential phase at a cell density of $2\text{--}4 \times 10^6$ cells/ml. The highest cell density obtained is 6×10^6 cells/ml with a doubling time of 12 h (Fig. 1).

To introduce foreign DNA into *D. papillatum*, we first tested electroporation as the technique of choice, as it is being successfully and extensively used for transforming *Trypanosoma brucei* and several other trypanosomatids (Beverley and Clayton, 1993). Since diplomonids and trypanosomatids are closely related and share a number of fine-structural features including a corset of subpellicular microtubules (Roy *et al.*, 2007; Tashyreva *et al.*, 2018), we reasoned that electroporation may also be suitable for *D. papillatum*.

As a first step, we confirmed that *D. papillatum* cells survive for at least 10 min in electroporation buffers (Cytomix - used for BTX electroporator or Amaxa - used for Amaxa Nucleofector II) and thus can be electroporated in standard electroporation buffers. Survival rate was $>90\%$. Next, two different electroporation apparatuses were tested: BTX electroporator and Amaxa Nucleofector II (see Experimental Procedures), for which we optimized several parameters, such as the composition of electroporation buffer, the electroporation program, and the amount of DNA used for transformation (Table 1). Specifically, cells were pelleted by a short centrifugation, re-suspended in electroporation buffer, subjected to the electric pulse, transferred back into seawater-containing medium, and then observed under the microscope. While electroporation with BTX killed the majority of cells, the Amaxa procedures yielded more than 20% survival (see Table 1), with cells retaining their original shape, and demonstrating full recovery after a few hours. The Amaxa Nucleofector II

Table 1. Used electroporation programs and survival rate of *D. papillatum* cells.

Electroporation machine	Used program	Survival rate
BTX	1600V, 25Ω, 50 μF	10%
Amaxa Nucleofector II	preset program X-001	50%-60%
Amaxa Nucleofector II	preset program X-014	20%-30%

programs proved to be best suited and therefore were chosen for further experiments (Table 1).

Identification of selectable markers

To identify selection markers, we tested the sensitivity of *D. papillatum* to seven antibiotics—hygromycin, geneticin, phleomycin, puromycin, blasticidin, nourseothricin and tetracycline. With the exception of tetracycline, these drugs have been extensively used for genetic manipulation of *T. brucei* and related trypanosomatids (<http://tryps.rockefeller.edu/>). The antibiotic concentration for selection of *D. papillatum* transformants was

assessed by determining cell viability by the Alamar Blue assay (Ráz *et al.*, 1997).

Diplonema papillatum was found to be sensitive to all the selection markers tested, except tetracycline. The effective concentrations for selection of electroporated *D. papillatum* are shown in Fig. 2 and Table 2.

The highest antibiotic sensitivity of *Diplonema* is to puromycin.

Rationale for construct design for *Diplonema* transformation

Our overall strategy for transformation is to incorporate foreign DNA via genomic integration, since nothing is known about DNA replication in this group and no native plasmids are known. Due to the high number of repetitive regions in the nuclear genome sequence of *D. papillatum* (unpubl. data), we restricted our selection of genomic regions to target for integration to genes that met several criteria. The genes must (i) be contained inside validated genomic contigs and (ii) be intron-less. Further the corresponding

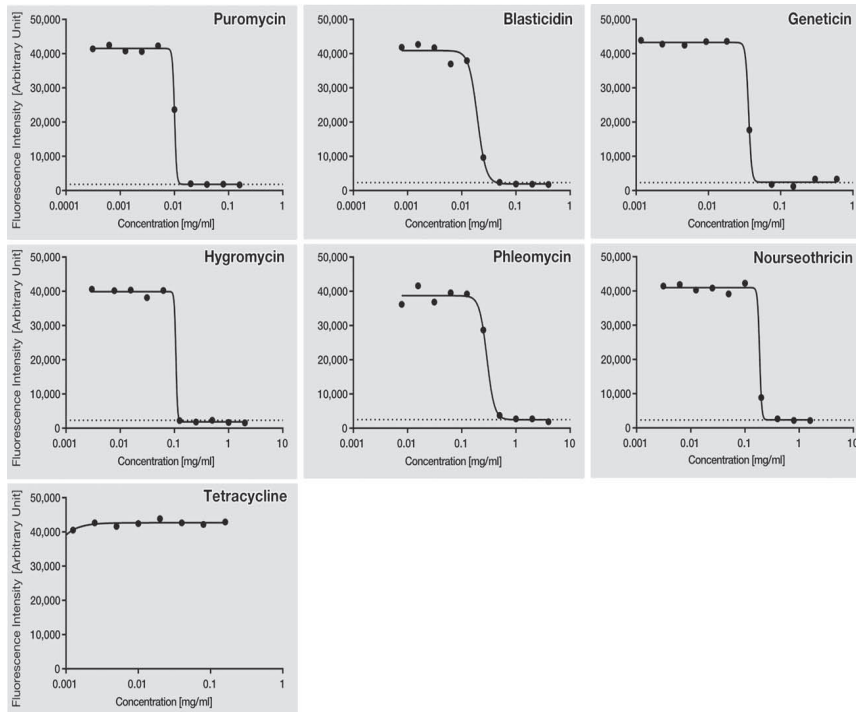


Fig. 2. Antibiotic sensitivity of *D. papillatum*. Effect of various concentrations of antibiotics (mg/ml; x axis) on survival of *Diplonema* cells, as determined by the Alamar blue assay, which measures viability by fluorescence (see Experimental Procedures). The fluorescence intensity corresponding to 1% cell survival is indicated by a dotted horizontal line.

Table 2. Tested resistance markers and their concentration used for selection.

Antibiotic	Concentration ($\mu\text{g/ml}$)
Puromycin	20
Blasticidin	50
Geneticin	75
Hygromycin	125
Nourseothricin	400
Phleomycin	500
Tetracycline	not sensitive

mRNAs must (iii) have a high steady state level (rank among the top hundred by their expression level) and (iv) carry a spliced leader (SL) at their 5' end (Sturm *et al.*, 2001).

In order to confirm proper integration, the target gene was tagged with a cassette that is composed of a fluorescent protein tag and a resistance gene, both flanked by 5' and 3' UTRs from *Diplonema*. To direct the integration of this cassette into the proper genomic position via homologous recombination, we appended to the cassette appropriate sequences of *D. papillatum* genome (targeting regions). This strategy, successfully used in numerous model systems (Stretton *et al.*, 1998; Janke *et al.*, 2004; Lai *et al.*, 2010; Wang *et al.* 2017), should lead to antibiotic-resistant transformants expressing a fluorescently labelled protein that can be detected with a fluorescent microscope or by commercially available antibodies.

Endogenous N-terminal tagging of α -tubulin

Based on the above described strategy, we have designed a construct for tagging the N-terminus of the *Diplonema* α -tubulin gene. The cassette contains the puromycin resistance gene (*pac*), here called puromycin^R (encoding the puromycin N-acetyltransferase of *Streptomyces alboniger*), which from all tested drugs was selectable at the lowest concentrations (Table 2). Moreover, it also contains the sequence encoding the fluorescent protein mCherry which lacks the stop codon, as we intended to create a fused mCherry- α -tubulin protein. We chose mCherry, which emits red light, to avoid overlap with the quite strong green autofluorescence of *D. papillatum* (our unpubl. data). The puromycin^R-mCherry cassette (~ 2 kbp) was flanked by *D. papillatum* sequences including 5' and 3' UTRs and homology regions about 500 bp long, which ought to enhance the integration of the construct into the targeted locus. A schematic representation of this construct is shown in Fig. 3A. A similar tagging approach was recently used for cytoskeleton studies in *T. brucei* (Sheriff *et al.*, 2014).

Transfection of *Diplonema* leads to stable chromosomal integration of foreign DNA

Both *NotI*-linearized and circular constructs were then electroporated into *Diplonema* cells in parallel and the transfectants were subjected to selection with increasing concentrations of puromycin to ensure stringent selection.

A total of 10 puromycin-resistant *Diplonema* clones were recovered after 8–10 days, at which time point a negative-control culture represented by wild type cells without the construct and in the presence of the drug did not display any viable cells.

Clones A3 and A4 (labelled according to their position in the 24-well plate) were investigated in detail by PCR and amplicon sequencing showing the presence of the puromycin^R-mCherry cassette in *Diplonema* genomic DNA (Fig. 3B; Supporting Information Figs. S1 and S4). To verify the cassette's presence in the genome, we performed Southern blot analysis using radioactively labelled probes against mCherry (Fig. 3C). The expected size of the *SpeI*-*NdeI* restriction fragment containing the mCherry CDS is 1969 bp, which corresponds to the detected band size (Fig. 3C, right). Digestion of genomic DNA with *PciI* should produce a fragment of 9699 bp in case the cassette has integrated into the expected site by homologous recombination; however, we detected bands of ~ 3.5 kbp (Fig. 3C, left). We, therefore, conclude that the complete cassette has been incorporated in the two *D. papillatum* transformants, but at heterologous sites in the genome. In line with this observation, amplification of the cassette with primers outside the cassette failed (data not shown).

Introduced heterologous gene can be transcribed, post-transcriptionally processed and translated in *Diplonema*

To examine whether mCherry and puromycin^R are transcribed in the *D. papillatum* transformants A3 and A4, we conducted reverse transcription followed by PCR amplification of the first-strand cDNA (RT-PCR). This experiment produced a single band of expected size confirming transcription of the two heterologous genes (Fig. 4B; Supporting Information Fig. S5A). We also confirmed, by nested RT-PCR, that the mCherry and puromycin^R mRNAs are properly processed post-transcriptionally by addition of the SL RNA to their 5' end (Fig. 4C; Supporting Information Figs. S2 and S5B).

To verify translation of the heterologous genes in A3 and A4 clones, we checked mCherry fluorescence and performed Western blots with an anti-mCherry antibody. However, in both cases no signal was detected, as expected due to the integration of the cassette into a different site. In contrast, translation of the puromycin^R gene could be demonstrated by two different immunoassays,

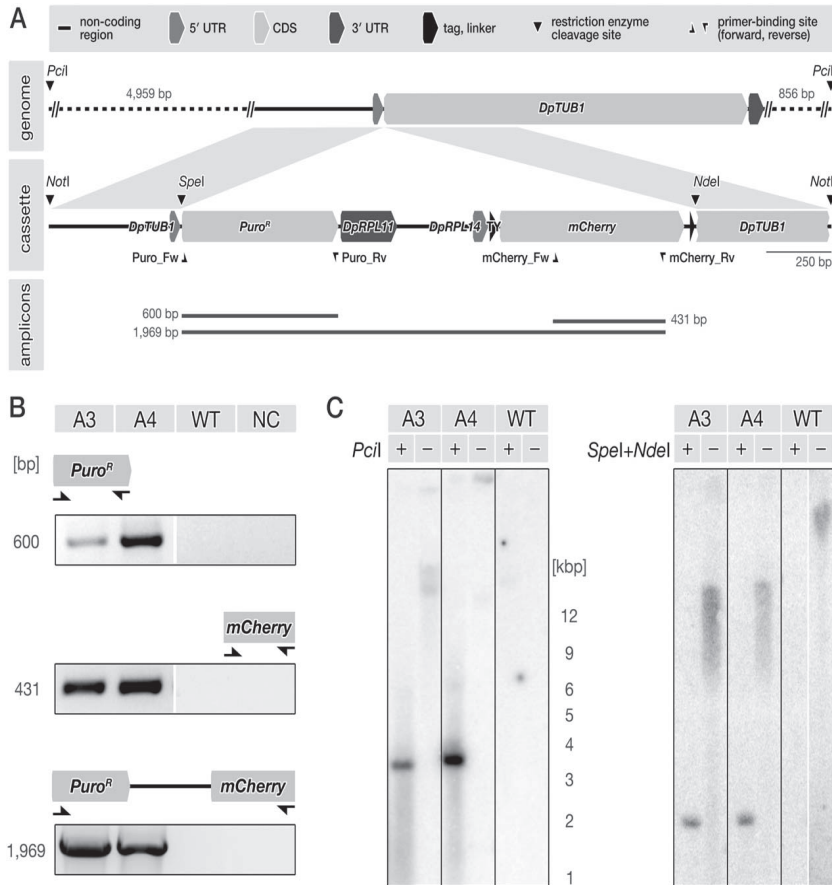


Fig. 3. Confirmation that the electroporated construct is integrated in the *D. papillatum* genome.

A. Scheme of the puromycin^R (puro^R) + mCherry cassette (2995 bp *NotI* fragment) including restriction sites, positions of the primers and expected sizes of the amplicons.

B. PCR of total DNA of *D. papillatum* wild type (WT) and selected transformants (A3 and A4) using specific primers for amplification of puromycin^R and mCherry CDSs. Negative control PCR (NC) was performed without template DNA. The explicit sequences are shown in Supporting Information Fig. S1 and the whole gels are shown in Supporting Information Fig. S4.

C. Southern hybridization of total DNA from *D. papillatum* wild type (WT) and transformants A3 and A4 using a DNA fragment of mCherry as a radiolabelled probe. Right panel, total DNA digested with *SpeI*+*NdeI*, which cut inside the cassette (+), and undigested (-). The *SpeI*+*NdeI* band is of expected size—1969 bp. Left panel, total DNA digested *PciI*, which cuts outside the cassette (+), and undigested (-). In case of homologous integration, the expected size of the *PciI* band is 9699 bp, however, we detected bands of ~3.5 kbp.

one using the anti-Puromycin antibody, and the other using the anti-Puromycin N-acetyltransferase antibody.

Puromycin is an aminonucleoside antibiotic that functions as an inhibitor of protein synthesis by disrupting peptide transfer on ribosomes, causing premature chain termination during translation. Since its structure resembles the 3' end of aminoacylated tRNA, puromycin enters the acceptor site of the ribosome and is added to the growing polypeptide chain, which leads to premature

termination of the newly synthesized proteins (Pestka, 1971). Anti-Puromycin antibody binds puromycin-containing newly synthesized proteins, visible in Western blot as a smear due to their different molecular weights. *Diplonema* wild type cells show indeed a strong signal of a broad molecular weight range, while the A3 and A4 clones do not (Supporting Information Fig. S3). This indicates that puromycin is not incorporated in the proteins of the clones A3 and A4, providing indirect evidence for the translation

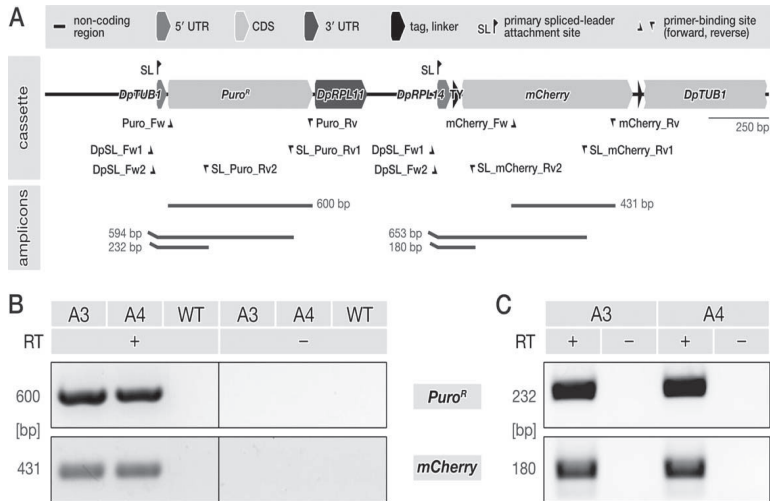


Fig. 4. Validation of proper transcription and post-transcriptional 5' end-processing of the transcripts produced from heterologous genes in *D. papillatum*.

A. Scheme of the puromycin^R (puro^R) + mCherry cassette including positions of the spliced leader (SL) primers and expected sizes of the amplicons of 5' region from the puromycin^R transcript and the mCherry transcript. RT-PCR primers used for amplifying mCherry and puromycin^R (puro^R) cDNAs are the same as shown in Fig. 3B.

B. Demonstration that heterologous genes in *D. papillatum* wild type (WT) and transformants A3 and A4 was used as a template for RT-PCR. Expected sizes of the RT-PCR products are indicated. Reactions with reverse transcriptase added and negative controls without reverse transcriptase are indicated by (+) and (-). The whole gels are shown in Supporting Information Fig. S5A.

C. The amplification of the 5' region from the puromycin^R (puro^R) transcript and the mCherry transcript respectively. Nested SL RT-PCR using total RNA from the transformants A3 and A4 as templates, and two sets of primers that anneal to the conserved SL sequence of *D. papillatum* (forward primers) and to the 5' end of the transcripts (reverse primers). The explicit sequences are shown in Supporting Information Fig. S2 and the whole gels are shown in Supporting Information Fig. S5B.

of the puromycin^R gene in the transformants (Supporting Information Fig. S3).

Similarly, Western immunoassays with the anti-Puromycin N-acetyltransferase antibody show clear expression of the heterologous puromycin^R gene in A3

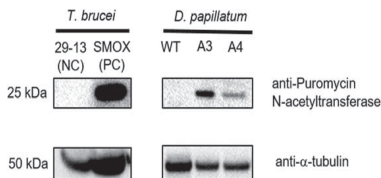


Fig. 5. Western blot analysis of *D. papillatum* wild type (WT) and of transformants A3 and A4 that express the puromycin^R gene (puromycin N-acetyltransferase).

Monoclonal rabbit anti-Puromycin N-acetyltransferase antibodies (1:500) and secondary anti-rabbit antibodies coupled to horseradish peroxidase (1:1000). 29-13, *T. brucei* procyclic-stage cell line 29-13, which does not express the puromycin^R gene, is used as a negative control (NC). SMOX, *T. brucei* cell line SMOX P9, which does express the puromycin^R gene, is used as a positive control (PC). Anti-α-tubulin antibodies were used as a loading control. The whole gels are shown in Supporting Information Fig. S6.

and A4, revealing distinct bands of the expected size (Fig. 5; Supporting Information Fig. S6). In both experiments, puromycin-resistant and sensitive cell lines from *T. brucei* served as negative and positive controls.

Discussion

Very few marine protists have been successfully transformed at present, most notably members of the green and red microalgae, diatoms, chlorarachniophytes (for review see Gong *et al.*, 2011), and further the alveolate *Perkinsus marinus* (Fernández-Robledo *et al.*, 2008), the prasinophyte *Ostreococcus tauri* (van Ooijen *et al.*, 2012), the haptophyte *Pleurochrysis carterae* (Endo *et al.*, 2016) and recently the kinetoplastid *Parabodo caudatus* (Gomaa *et al.*, 2017). Different transfection methods were used, notably polyethylene glycol-mediated approach for the haptophyte, otherwise microprojectile bombardment or electroporation.

The challenge of electroporating marine organisms

Electroporation induces a transient destabilization of the cell membrane, which then becomes highly permeable to

foreign DNA, proteins or small molecules. Electroporation is used particularly for suspension cultures as all cells are essentially transfected simultaneously (Heiser, 2000). For transforming *D. papillatum*, we have opted for electroporation, which is an established technique in kinetoplastids, the sister group of diplomemids (e.g., Adl *et al.*, 2012), specifically in the medically important *Trypanosoma* and *Leishmania* (Beverley and Clayton, 1993).

The technical challenge of electroporating marine organisms is that salts in the medium cause electrical discharge (arcing), which reduces the viability of the organism (Potter and Heller, 2003). In the case of *Parabodo caudatus* (Gomaa *et al.*, 2017), electroporation succeeded because this microeukaryote lives in both freshwater and marine environments and is, therefore, only moderately sensitive to low salt concentrations. Fortunately, *D. papillatum* is also tolerant to brackish water, and thus endures short-term exposure to the low-salt electroporation buffer.

Integration and expression of heterologous genes in the *Diplonema nucleus*

For two transformed *D. papillatum* clones, we demonstrated by PCR, DNA sequencing, and Southern blot analysis that the transfection constructs are readily integrated into the nuclear genome. However, the constructs integrated at apparently random genomic positions, a common issue in many genetic systems including mammalian and plant systems (Sargent *et al.*, 1997; Gorbunova and Levy, 1999). Ectopic integration may be due to the fact that the genome of *Diplonema* is highly repetitive, or, alternatively, because homologous recombination (HR) may be less efficient in this organism than microhomology-mediated end-joining (MMEJ) or the classical non-homologous end-joining (NHEJ) DNA repair/recombination pathways (for a review see Michael, 2010). Note that the genes known to be involved in the corresponding machineries are present in the *D. papillatum* genome (our unpubl. data).

Importantly, however, transfected heterologous genes are nevertheless expressed in *Diplonema*. The protein tag (mCherry) and resistance gene (puromycin^R) used here are transcribed, and the SL is correctly *trans*-spliced onto the 5' end of the transcripts. For the puromycin^R gene, we were able to demonstrate translation into protein and observe the expected resistant phenotype. In contrast, translation of the mCherry transcript was neither observed nor expected, because a stop codon and a 3' UTR were absent, as it was designed to be fused to the α -tubulin gene as an N-terminal tag. Due to the lack of a stop codon and/or poly-(A) tail, the mRNA may have been degraded by one of the quality control mechanisms of the eukaryotic cell (Klauer and van Hoof, 2012).

Since the purpose of this construct was to create a fused protein composed of N-terminally tagged α -tubulin

with mCherry, the stop codon at the end of the mCherry gene was intentionally removed. However, as this construct was not integrated into a proper position, no expression of mCherry could be observed.

An alternative approach to obtain resistant cell line emitting red fluorescence would be to design a cassette containing intact genes and corresponding UTRs to replace an endogenous gene such as tubulin, or insert the cassette into a particular position in the *Diplonema* genome, f.e. in a transcriptionally silent region, a routinely strategy in other organisms. While this approach would not produce a labelled endogenous protein, it would allow the expression of the mCherry protein. In our future work, we will apply different strategies with the aim to achieve correct integration of the cassette.

Future prospects for *D. papillatum* as a genetic system

Taken together, these data show that *D. papillatum* has all prerequisites for becoming a genetically tractable organism. Successful transformation provides first example of heterologous gene expression in an euglenozoan protist outside the kinetoplastids, and raises questions about how widely these methods can be applied within the group, and what other tools can now be used in model diplomemids. Currently, only half a dozen diplomemids are available in the ATCC collection and can be cultivated in the laboratory. For the time being, with a representative of the species-rich DSPD I clade yet to be brought into culture, the next candidate for transformation is *Hemistasia phaeocysticola* (Yabuki and Tame, 2015; Yabuki *et al.*, 2016), which is more closely related to the most abundant diplomemids. However, this species is also much more challenging to work with, as it prefers live diatoms as a food source, reaches only low cell densities and so far could not be cryopreserved (our unpubl. data).

The availability of a methodology to transfect *Diplonema* will also facilitate investigation of the machineries that drive the unique post-transcriptional processes in their mitochondria, such as U-appendage RNA editing and *trans*-splicing of fragmented genes (for reviews see Valach *et al.*, 2016; Faktorová *et al.*, 2018). For example, fluorescence tagging of terminal uridyl transferases or RNA ligases would identify the respective enzyme that acts in mitochondria, and open the venue for uncovering other components in the hypothetical editosome or *trans*-spliceosome.

Future work may achieve targeted integration by further extension of the 5' and 3' homologous regions of the constructs to more than 1500 bp, as shown in trypanosomes (Barnes and McCulloch, 2007). Furthermore, it would be worthwhile to explore the CRISPR/Cas9 strategy, which was recently successfully implemented in kinetoplastids (Lander *et al.*, 2016; Beneke *et al.*, 2017). In this strategy, a circular plasmid that is in the transformed organism

retained either as an episome or as a linearized plasmid, or a PCR product containing homologous targeting regions for genome integration can be used. For *Diplonema*, PCR amplicons are of higher priority, as attempts to maintain circular plasmids as non-integrated episomes have failed thus far. Last, but not least, we will also try to enhance the efficiency of the homologous recombination pathway using inhibitors of the non-homologous end joining pathway, as recently described in *Cryptococcus neoformans* (Arras et al., 2016). As a high quality genome and transcriptome of *D. papillatum* shall soon be available, it would be beneficial to have a transformation protocol in place. We envisage its application in functional analysis of proteins with potential ecological significance, such as those involved in pathways that respond to environmental stress.

Experimental procedures

Strains, cultivation and growth curves

D. papillatum (ATCC 50162) was cultivated axenically at 27°C in an artificial sea salt mixture 40 g/l (Sigma, S9883), 0.1% (w/v) tryptone, 1% (v/v) fetal bovine serum and 100 µg/ml of chloramphenicol. Cell density was measured manually by the Neubauer cell chamber. Before measurement, cells had to be fixed in 3.7% (v/v) formaldehyde in SSC to retain their shape and to prevent them from moving prior to counting.

Determination of resistance to antibiotics using Alamar Blue assay

The inhibition concentration where 99% of the cell population is dead (IC99) was determined using the fluorescence viability indicator Alamar blue (Resazurin sodium salt, Sigma, R7017) as described in Gould et al. (2013). It is based on resazurin, a non-toxic, permeable and weakly fluorescent dye used as a redox indicator. Its reduced form or resorufin is pink and highly fluorescent, with its fluorescence intensity being proportional to the number of respiring (e.g., metabolically active) cells. Cells were inoculated into the 96-well flat-bottomed microtiter plate (Costar) at a concentration 1×10^5 cells/ml. Seven drugs (hygromycin, puromycin, phleomycin, geneticin, blastidin, nourseothricin and tetracycline) were tested in triplicates using serial dilutions at 27°C for 48 h; wells without the drug at the end of each row were used as a control. Resazurin was prepared at a concentration of 0.125 mg/ml in PBS and added to the plate after 48 h of incubation with the drugs. The plate was subsequently incubated for another 24 h and fluorescence was read at 585 nm using a Tecan Infinite M200. Data were analyzed using Prism 5.0 software (GraphPad, San Diego, CA) and IC99 values were derived from sigmoidal dose-response curves with variable slopes.

Design and preparation of transformation cassette

Design of the puromycin^R + mCherry cassette (Fig. 3A) was performed using the UTRs from the *D. papillatum* genome to drive the expression. The entirely modular cassette (2996 bp-long) was designed for the N-terminal tagging and contained

the following fragments ordered in the 5'–3' direction, starting with a *NotI* restriction site: 499 nt of the sequence just before the beginning of α -tubulin gene *DpTUB1* including its 5'UTR, puromycin N-acetyltransferase gene (puromycin^R) with the stop codon, 3' UTR of *DpRPL11*, 5' UTR of *DpRPL14*, a TY tag and the mCherry CDS without the stop codon followed by a linker, 510 nt of the 5' of the α -tubulin gene *DpTUB1*, followed by a second *NotI* site. The cassette was synthesized by Biomatik (<http://www.biomatik.com>), cloned into pBluescript II SK(+) vector and the sequence was submitted to GenBank (Acc. No. MG490656). The vector was isolated using QIAprep Spin Miniprep Kit (Qiagen, 27106) and either directly used for electroporation or cut with *NotI* restriction enzyme, isolated from the gel using QIAquick Gel Extraction Kit (Qiagen, 28706), and re-suspended in 10 µl of deionized distilled water before electroporation.

Electroporation and obtaining of the transformants

A total of 5×10^7 cells (2×10^6 cells/ml) was harvested by centrifugation at 1300 g for 10 min at 25°C and re-suspended either in 100 µl cytomix buffer (van den Hoff et al., 1992) and electroporated with the BTX machine (1600V, 25Ω, 50 µF), or in 100 µl of AMAXA buffer (81.8 µl of Human T-cell nucleofector solution + 18.2 µl of Supplement) for the electroporation by Amaxa Nucleofector II.

Ten to 15 µg of DNA were mixed with electroporation buffer and *Diplonema* cells before transferring into the electroporation cuvette and electroporation. After the electroporation pulse was applied to the cuvette, the mixture was immediately transferred into 10 ml of *Diplonema* growth media. Subsequently, the cells were allowed to recover for about 8 h and the transfectants were subjected to selection with increasing concentrations (12–40 µg/ml) of puromycin. While this wide range makes the experiment more time consuming, it ensured stringent selection of transformants. Clones A3 and A4 examined in details were cultivated in 24 and 28 µg/ml of puromycin, respectively.

Following an expansion of each clone to a volume of 20 ml transformants had been cultured for up to 8 weeks prior to testing by PCR, which proved that all of them indeed contain integrated constructs. We were able to freeze-store the transformants in 10% glycerol-containing medium at –80°C or in liquid nitrogen for several months.

PCR using genomic DNA

Genomic DNA was isolated using Qiagen DNA isolation kit (Qiagen, 69504). Primer pairs used for verification of the integration are shown in Fig. 3A, for primer sequences see Supporting Information Table S1. PCR amplification was done using OneTaq polymerase (NEB Biolabs, M0486L) and the following program: initial denaturation 94°C for 3 min, denaturation at 94°C for 30 s, annealing at 58°C for 60 s, extension at 68°C for 3 min and a final extension at 68°C for 10 min, for 30 cycles. To amplify mCherry and puromycin^R, a lower extension of 68°C for 1 min 30 s and a final extension of 68°C for 5 min were used. Cassette integration in the genome of *D. papillatum* was confirmed by sequencing of the PCR products (Eurofins Genomics).

Southern blot hybridization

Southern blot analysis was performed as previously described (Lai *et al.*, 2008). Total DNA was isolated from wild type cells and transformed clones, digested with selected restriction enzymes and subjected to agarose gel electrophoresis. Samples were blotted overnight to a Zeta-Probe membrane (Bio-rad) by capillarity and subsequently cross-linked with UV light as described (Vondrušková *et al.*, 2005). The mCherry probe was PCR amplified using mCherry_Fw and mCherry_Rv primers and radiolabelled with [α - 32 P]dATP using Radioactive DNA Labelling kit (Thermo Scientific DecaLabel DNA Labelling Kit, #K0622) and hybridization with the probe was performed overnight at 60°C according to the Zeta-Probe membrane manual. The membrane was exposed to a Fuji Imaging phosphor screen (28956475, BAS-MS 2025) and scanned in a Phosphorimager (Amersham, GE Healthcare, Typhoon 9400) for signal detection.

RNA isolation and cDNA synthesis, RT-PCR and SL RNA-PCR

Total RNA was isolated using TriReagent (MRC, TR118) and pellet was re-suspended in 30 μ l of RNase-free water. cDNA was prepared using QuantiTect Reverse Transcription Kit (Qiagen, 205311) with random primers. PCR was performed on cDNA with primers shown in Fig. 3A (for primer sequences, see Supporting Information Table S1), and OneTaq polymerase (NEB Biolabs, M0486L), with the following protocol: 30 s at 94°C; 30 cycles of 30 s at 94°C, 30 s at 60°C and 1 min 30 s at 68°C, and a final extension for 5 min at 68°C. The same reactions without reverse transcriptase (RT-) were used as negative controls.

The cDNA was used as a template for nested PCR with the following program: 30 s at 94°C; 30 cycles of 30 s at 94°C, 30 s at 58°C and 1 min 30 s at 68°C, followed by a final extension (5 min at 68°C). DpSL_Fw1 and DpSL_Fw2 primers derived from the SL RNA gene were used in combination with mCherry (SL_mCherry_Rv1; SL_mCherry_Rv2) or puromycin^R specific primers (SL_Puro_Rv1; SL_Puro_Rv2; Supporting Information Table S1). The position of primers and expected size of PCR products is shown in Fig. 5. An amplicon containing the SL and N-terminus of puromycin^R or mCherry was obtained and verified by sequencing (Eurofins Genomics; Supporting Information Fig. S2).

Western blot

Cell lysates were prepared in 2 \times SDS sample buffer using 1 \times 10⁷ cells per lane and separated on a 12% (v/v) SDS-PAGE gel and proteins were subsequently transferred onto the PVDF membrane by electroblotting. Membranes were blocked overnight with 5% (w/v) non-fat milk prepared in PBS 0.5% (v/v) Tween 20 and probed with the primary monoclonal mouse anti-Puromycin antibody (1:500) (Merck, MABE343) overnight at 4°C. The membrane was subsequently incubated with the secondary anti-mouse polyclonal antibody conjugated with horseradish peroxidase (1:1000) (Sigma) and visualized using Clarity western ECL substrate (Bio-Rad).

For the detection of puromycin N-acetyltransferase (*pac* gene product), cell lysates were prepared in NuPAGE LDS sample buffer (Invitrogen) using 1 \times 10⁷ cells per lane

separated on Bolt 4%–12% Bis-Tris polyacrylamide gels (Invitrogen) and transferred to a Amersham Hybond P PVDF membrane (GE Healthcare), blocked overnight with 5% (w/v) non-fat milk prepared in PBS 0.5% (v/v) Tween-20 and subsequently hybridized with the primary monoclonal rabbit anti-Puromycin N-acetyltransferase antibody (1:500) (ThermoFisher Scientific) and subsequently with secondary anti-rabbit polyclonal antibody coupled to horseradish peroxidase (1:1000). Monoclonal anti- α -Tubulin antibody produced in mouse (1:1000) (Sigma, T9026) was used as a loading control.

Acknowledgements

We thank Jorge Morales (Heinrich Heine University Düsseldorf) for discussion in the initial stages of the project; Anzhelika Butenko and Pavel Flegontov (Institute of Parasitology) for help with the selection of genes and design of the constructs; Galina Prokopchuk and Daria Tashyreva (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 Institute of Health Research (CIHR, MOP 70309 to GB) and from the Gordon and Betty Moore Foundation (GBMF4983.01 to JL, GB and PK) are kindly acknowledged.

Conflict of Interest

Authors have no conflict of interest to declare.

References

- Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., *et al.* (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* **59**: 429–493.
- Arras, S.D.M., Fraser, J.A., and Lustig, A.J. (2016) Chemical inhibitors of non-homologous end joining increase targeted construct integration in *Cryptococcus neoformans*. *PLoS One* **11**: e0163049.
- Barnes, R.L., and McCulloch, R. (2007) *Trypanosoma brucei* homologous recombination is dependent on substrate length and homology, though displays a differential dependence on mismatch repair as substrate length decreases. *Nucleic Acids Res* **35**: 3478–3493.
- Beneke, T., Madden, R., Makin, L., Valli, J., Sunter, J., and Gluenz, E. (2017) A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplasts. *R Soc Open Sci* **4**: 170095.
- Beverly, S.M., and Clayton, C.E. (1993) Transfection of *Leishmania* and *Trypanosoma brucei* by electroporation. *Methods Mol Biol* **21**: 333–348.
- Breglia, S.A., Yubuki, N., Hoppenrath, M., and Leander, B.S. (2010) Ultrastructure and molecular phylogenetic position of a novel euglenozoan with extrusive epibiotic bacteria: *Bihospites bacati* n. gen. et sp. (Symbiontida). *BMC Microbiol* **10**: 145.
- de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., *et al.* (2015) Eukaryotic plankton diversity in the sunlit global ocean. *Science* **348**: 1261605.

- Endo, H., Yoshida, M., Uji, T., Saga, N., Inoue, K., and Nagasawa, H. (2016) Stable nuclear transformation system for the coccolithophorid alga *Pleurochrysis carterae*. *Sci Rep* **6**: 22252.
- Faktorová, D., Valach, M., Kaur, B., Burger, G., and Lukeš, J. (2018) Mitochondrial RNA editing and processing in diplomemid protists. "RNA Metabolism in Mitochondria", Springer series "Nucleic Acids and Molecular Biology", in press.
- Fernández-Robledo, J.A., Lin, Z., and Vasta, G.R. (2008) Transfection of the protozoan parasite *Perkinsus marinus*. *Mol Biochem Parasitol* **157**: 44–53.
- Flegontova, O., Flegontov, P., Malviya, S., Audic, S., Wincker, P., de Vargas, C., et al. (2016) Extreme diversity of diplomemid eukaryotes in the ocean. *Curr Biol* **26**: 3060–3065.
- Gawryluk, R.M.R., del Campo, J., Okamoto, N., Strassert, J.F.H., Lukeš, J., Richards, T.A., et al. (2016) Morphological identification and single-cell genomics of marine diplomemids. *Curr Biol* **26**: 3053–3059.
- Gong, Y., Hu, H., Gao, Y., Xu, X., and Gao, H. (2011) Microalgae as platforms for production of recombinant proteins and valuable compounds: progress and prospects. *J Ind Microbiol Biotechnol* **38**: 1879–1890.
- Gomaa, F., Garcia, P.A., Delaney, J., Girguis, P.R., Buie, C.R., and Edgcomb, V.P. (2017) Toward establishing model organisms for marine protists: successful transfection protocols for *Parabodo caudatus* (Kinetoplastida: Excavata). *Environ Microbiol* **19**: 3487–3499.
- Gorbunova, V.V., and Levy, A.A. (1999) How plants make ends meet: DNA double-strand break repair. *Trends Plant Sci* **4**: 263–269.
- Gould, M.K., Bachmaier, S., Ali, J.A., Alsford, S., Tagoe, D.N., Munday, J.C., et al. (2013) Cyclic AMP effectors in African trypanosomes revealed by genome-scale RNA interference library screening for resistance to the phosphodiesterase inhibitor CpdA. *Antimicrob Agents Chemother* **57**: 4882–4893.
- Griessman, K. (1914) Ueber marine Flagellaten. *Arch Protistenkd* **32**: 1–78.
- Heiser, W.C. (2000) Optimizing electroporation conditions for the transformation of mammalian cells. *Methods Mol Biol* **130**: 117–134.
- Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., et al. (2004) A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**: 947–962.
- Klauer, A.A., and van Hoof, A. (2012) Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. *Wiley Interdiscip Rev RNA* **3**: 649–660.
- Kiethega, G.N., Yan, Y., Turcotte, M., and Burger, G. (2013) RNA-level unscrambling of fragmented genes in *Diplonema* mitochondria. *RNA Biol* **10**: 301–313.
- Lai, D.H., Hashimi, H., Lun, Z.R., Ayala, F.J., and Lukeš, J. (2008) Adaptations of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *Trypanosoma equiperdum* and *Trypanosoma evansi* are petite mutants of *T. brucei*. *Proc Natl Acad Sci U S A* **105**: 1999–2004.
- Lai, J., Ng, S.K., Liu, F.F., Patkar, R.N., Lu, Y., Chan, J.R., et al. (2010) Marker fusion tagging, a new method for production of chromosomally encoded fusion proteins. *Eukaryot Cell* **9**: 827–830.
- Lander, N., Chiurillo, M.A., and Docampo, R. (2016) Genome editing by CRISPR/Cas9: A game change in the genetic manipulation of protists. *J Eukaryot Microbiol* **63**: 679–690.
- Lara, E., Moreira, D., Vereshchaka, A., and López-García, P. (2009) Panoceanic distribution of new highly diverse clades of deep-sea diplomemids. *Environ Microbiol* **11**: 47–55.
- Larsen, J., and Patterson, D.J. (1990) Some flagellates (Protista) from tropical marine sediments. *J. Nat. Hist* **24**: 801–937.
- Lukeš, J., Flegontova, O., and Horák, A. (2015) Diplomemids. *Curr Biol* **25**: R702–R704.
- Makiuchi, T., Annoura, T., Hashimoto, M., Hashimoto, T., Aoki, T., and Nara, T. (2011) Compartmentalization of a glycolytic enzyme in *Diplonema*, a non-kinetoplastid euglenozoan. *Protist* **162**: 482–489.
- Marande, W., and Burger, G. (2007) Mitochondrial DNA as a genomic jigsaw puzzle. *Science* **318**: 415.
- Marande, W., Lukeš, J., and Burger, G. (2005) Unique mitochondrial genome structure in diplomemids, the sister group of kinetoplastids. *Eukaryot Cell* **4**: 1137–1146.
- Michael, R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end joining pathway. *Annu Rev Biochem* **79**: 181–211.
- Morales, J., Hashimoto, M., Williams, T.A., Hirawake-Mogi, H., Makiuchi, T., Tsubouchi, A., et al. (2016) Differential remodeling of peroxisome function underpins the environmental and metabolic adaptability of diplomemids and kinetoplastids. *Proc Biol Sci* **283**: 20160520.
- Moreira, S., Valach, M., Aoulad-Aissa, M., Otto, C., and Burger, G. (2016) Novel modes of RNA editing in mitochondria. *Nucleic Acids Res* **44**: 4907–4919.
- Pestka, S. (1971) Inhibitors of ribosome functions. *Annu Rev Microbiol* **25**: 487–562.
- Porter, D. (1973) *Isonema papillatum* sp. n., a new colorless marine flagellate: a light- and electron microscopic study. *J Protozool* **20**: 351–356.
- Potter, H., and Heller, R. (2003) Transfection by electroporation. *Curr Protoc Mol Biol* Chapter 9, Unit-9.3.
- Ráz, B., Iten, M., Grether-Bühler, Y., Kaminsky, R., and Brun, R. (1997) The Alamar Blue assay to determine drug sensitivity of African trypanosomes (*T.b. rhodesiense* and *T.b. gambiense*) in vitro. *Acta Trop* **68**: 139–147.
- Roy, J., Faktorová, D., Benada, O., Lukeš, J., and Burger, G. (2007) Description of *Rhynchopus euleioides* n. sp. (Diplonemea), a free-living marine euglenozoan. *J Eukaryot Microbiol* **54**: 137–145.
- Sargent, R.G., Breneman, M.A., and Wilson, J.H. (1997) Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol Cell Biol* **17**: 267–277.
- Schnepf, E. (1994) Light and electron microscopical observations in *Rhynchopus coccinodiscivorus* sp. nov., a colorless, phagotrophic Euglenozoan with concealed flagella. *Arch Protistenkd* **144**: 63–74.
- Schuster, F.L., Goldstein, S., and Hershenov, B. (1968) Ultrastructure of a flagellate, *Isonema nigricans* nov. gen. nov. sp., from a polluted marine habitat. *Protistologica* **4**: 141–149.
- Sheriff, O., Lim, L.F., and He, C.Y. (2014) Tracking the biogenesis and inheritance of subpellicular microtubule in *Trypanosoma brucei* with inducible YFP- α -tubulin. *Biomed Res Int* **2014**: 893272.

- Simpson, A.G.B. (1997) The identity and composition of the Euglenozoa. *Arch Protistenkd* **148**: 318–328.
- Skuja, H. (1948) Taxonomie des phytoplanktons einiger seen in Uppland, Schweden. *Symb Bot Upsal* **9**: 5–399.
- Stretton, S., Techkarnjanaruk, S., McLennan, A.M., and Goodman, A.E. (1998) Use of green fluorescent protein to tag and investigate gene expression in marine bacteria. *Appl Environ Microbiol* **64**: 2554–2559.
- Sturm, N.R., Maslov, D.A., Grisard, E.C., and Campbell, D.A. (2001) *Diplonema* spp. possess spliced leader RNA genes similar to the Kinetoplastida. *J Eukaryot Microbiol* **48**: 325–331.
- Tashyreva, D., Prokopchuk, G., Yabuki, A., Kaur, B., Faktorová, D., Votyčka, J., et al. (2018) Phylogeny and morphology of diplomemids from the Sea of Japan. *Protist* (in press).
- Valach, M., Moreira, S., Faktorová, D., Lukeš, J., and Burger, G. (2016) Post-transcriptional mending of gene sequences: Looking under the hood of mitochondrial gene expression in diplomemids. *RNA Biol* **13**: 1204–1211.
- van den Hoff, M.J.B., Moorman, A.F.M., and Lamers, W.H. (1992) Electroporation in 'intracellular' buffer increases cell survival. *Nucleic Acids Res* **20**: 2902.
- van Ooijen, G., Knox, K., Kis, K., Bouget, F.Y., and Millar, A.J. (2012) Genomic transformation of the picoeukaryote *Ostreococcus tauri*. *J Vis Exp* **65**: 4074.
- Vondrušková, E., Van Den Burg, J., Ziková, A., Ernst, N., Stuart, K., Benne, R., and Lukeš, J. (2005) RNA interference analyses suggest a transcript-specific regulatory role for mitochondrial RNA-binding proteins MRP1 and MRP2 in RNA editing and other RNA processing in *Trypanosoma brucei*. *J Biol Chem* **280**: 2429–2438.
- Wang, Q., Xue, H., Li, S., Chen, Y., Tian, X., Xu, X., et al. (2017) A method for labeling proteins with tags at the native genomic loci in budding yeast. *PLoS One* **12**: e0176184.
- Worden, A.Z., and Wilken, S. (2016) A plankton bloom shifts as the ocean warms. *Science* **354**: 287–288.
- Yabuki, A., and 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., and Fujikura, K. (2016) Hyper-eccentric structural genes in the mitochondrial genome of the algal parasite *Hemistasia phaeocysticola*. *Genome Biol Evol* **8**: 2870–2878.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. List of primers used.

Fig. S1. PCR amplicons (from Fig. 3B) obtained with primers that amplify the puromycin^R region (A) or mCherry region (B) of clones A3 and A4 were verified by sequencing. Beginning parts of the amplicon sequences are aligned with the reference sequence.

Fig. S2. SL-PCR amplicons (from Fig. 4C) were verified by sequencing. Part of the amplicon sequences obtained with primers that amplify the puromycin^R region (A) or mCherry region (B) are aligned to the reference sequence.

Fig. S3. Western blot analysis of wild type (WT) *D. papillatum* and transformants A3 and A4 after 24 h incubation in a medium containing 20 µg/ml puromycin. *Trypanosoma brucei* SMOX P9 procyclic stage cells (SmOx) expressing the puromycin^R gene were used as a positive control. Monoclonal mouse anti-Puromycin antibodies (1:1000) and secondary anti-mouse antibodies coupled to horseradish peroxidase (1:1000) were used for visualization employing the ECL kit.

Fig. S4. Confirmation that the electroporated construct is integrated in the *D. papillatum* genome. (A) Scheme of the puromycin^R (puro^R) + mCherry cassette (2995 bp *NotI* fragment) including restriction sites, positions of the primers and expected sizes of the amplicons. The whole gels with PCR of total DNA of *D. papillatum* wild type (WT), selected transformants (A3 and A4; underlined) and some other obtained transformants (A2, A5, B4, D2, B2, C2) using specific primers for amplification of puromycin^R (B), mCherry (C) or puromycin^R + mCherry (D). Negative control PCR (NC) was performed without template DNA.

Fig. S5. Validation of proper transcription and post-transcriptional 5' end-processing of the transcripts produced from heterologous genes in *D. papillatum*. (A) The whole gels of RT-PCR shown in Fig. 4B. RNA from *D. papillatum* wild type (WT) and selected transformants (A3 and A4; underlined) and some other obtained transformants (C3, B3, B4, A5) was used as a template for RT-PCR. Expected sizes of the RT-PCR products are indicated. Reactions with reverse transcriptase added and negative controls without reverse transcriptase are indicated by (+) and (-). (B) The amplification of the 5' region from the puromycin^R (puro^R) transcript and the mCherry transcript respectively. Nested SL RT-PCR using total RNA from the transformants A3 and A4 as templates, and two sets of primers that anneal to the conserved SL sequence of *D. papillatum* (forward primers) and to the 5' end of the transcripts (reverse primers), see Fig. 4A. The whole gels of both SL RT-PCR reaction are shown here together with the sizes of expected product.

Fig. S6. Whole gels of western blot analysis (shown in Fig. 5) of *D. papillatum* wild type (WT) and transformants A3 and A4 that express the puromycin^R gene (puromycin N-acetyltransferase). Monoclonal rabbit anti-Puromycin N-acetyltransferase antibodies (1:500) and secondary anti-rabbit antibodies coupled to horseradish peroxidase (1:1,000). 29-13, *T. brucei* procyclic-stage cell line 29-13, which does not express the puromycin^R gene, is used as a negative control. SMOX, *T. brucei* cell line SMOX P9, which does express the puromycin^R gene, is used as a positive control. Anti-α-tubulin antibodies were used as a loading control.

Chapter 3

Mitochondrial RNA editing and processing in diplomemid protists.

In: M.W. Gray and J. Cruz-Reyes (Eds.), RNA Metabolism in Mitochondria, Springer 145–176.

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

© Springer International Publishing AG, part of Springer Nature 2018
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

145

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 (Szkłarczyk 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 diplomemids have a rather similar organelle (our unpublished data). This presumption and the relatedness with kinetoplastid flagellates indicate that the mitoproteome of diplomemids 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 diplomemids (Fig. 6.1) will be more akin to the

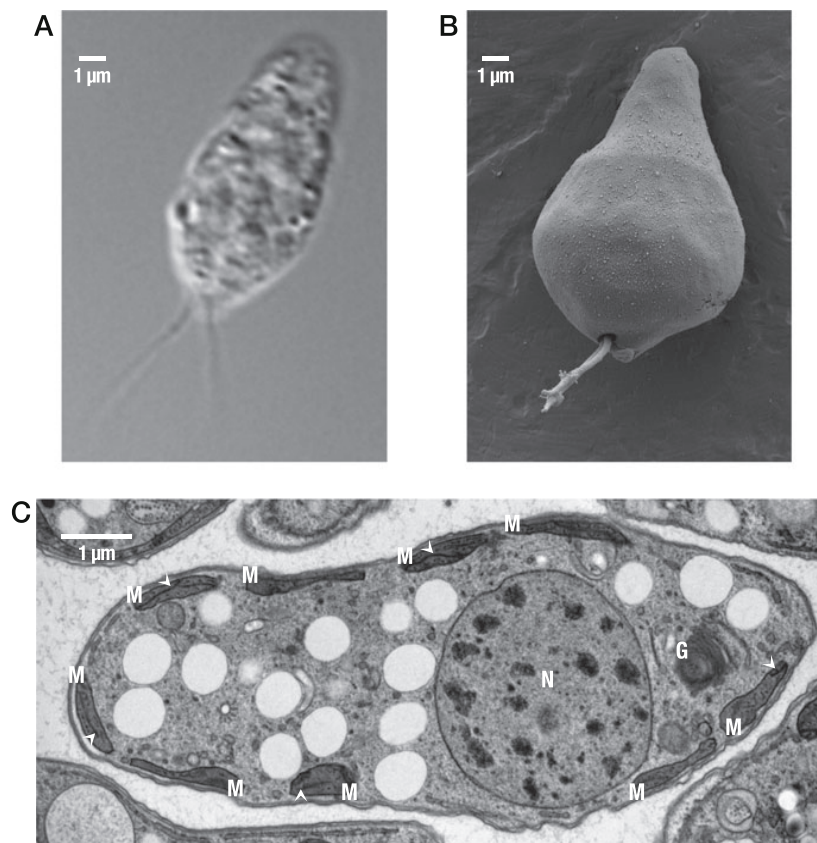


Fig. 6.1 Morphology of the model diplonemid, *Diplonema papillatum*. Light microscopy (a) and scanning electron microscopy (b) revealing the sac-like shape of the cell in culture and two heterodynamic flagella. (c) Transmission electron microscopy of a longitudinally sectioned cell with a prominent nucleus (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, ultra-structure, and behavior of marine diplomemids 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 diplomemids 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 epibiotic 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 diplomemid 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 non-selective 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 diplomid 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*.

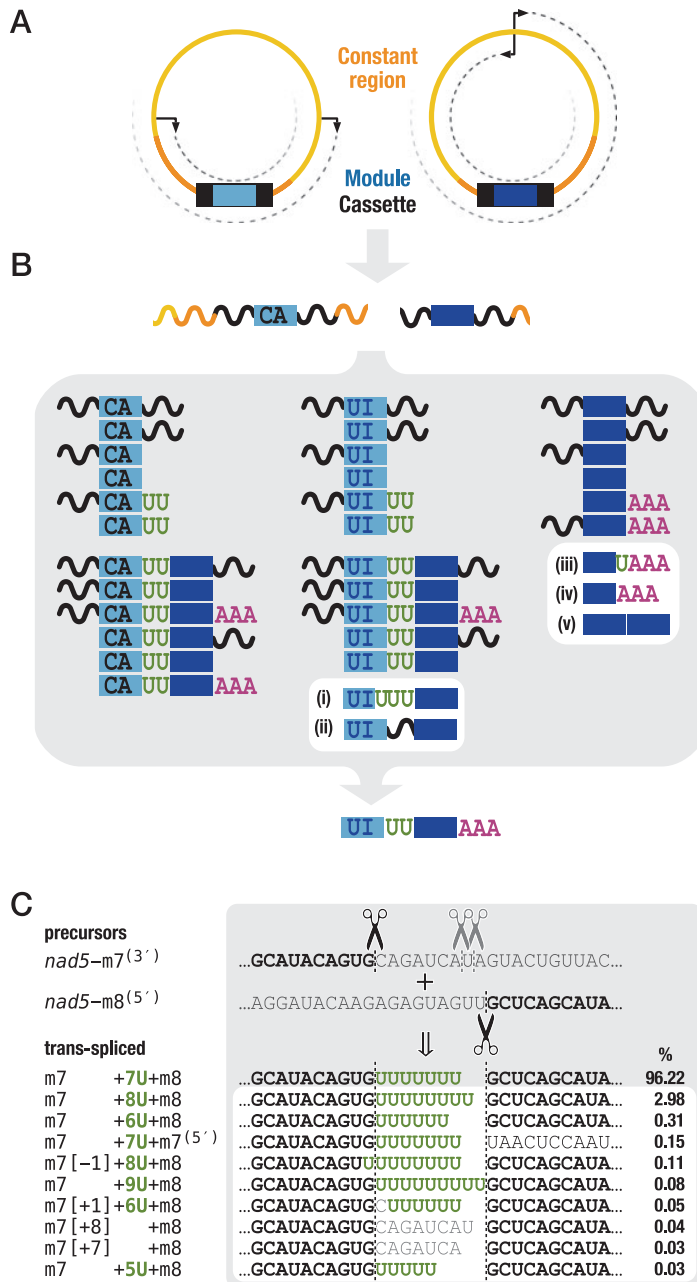


Fig. 6.2 Gene expression in diplonemid mitochondria. (a) Canonical circular mitochondrial chromosomes comprise a constant region of identical sequence across all members of a class

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

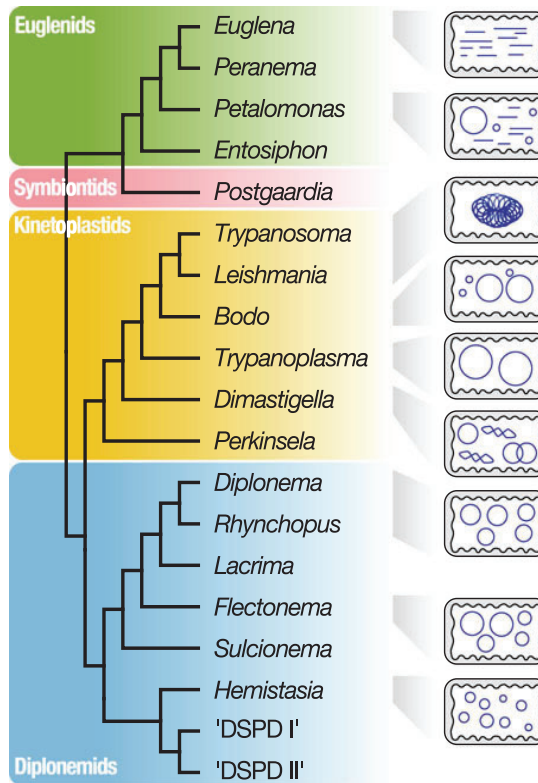


Fig. 6.3 Mitochondrial genome architectures and gene expression pathways in euglenozoans. Phylogenetic relationships among representative euglenozoan genera with their mitochondrial genome organization schematized. Euglenid mitochondria generally contain an assortment of linear molecules of variable 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

Gene	Diplonema			Euglena		Trypanosoma	
	+U	C-to-U	A-to-I		+U	-U	
<i>cob</i>	3		6	1	1	34	
<i>cox1</i>	6		9	1	1		
<i>cox2</i>	3		4	1	1	4	
<i>cox3</i>	1		3	1	1	547	41
<i>nad1</i>	16		5	1	1		
<i>nad4</i>	2	22	7	1	1		
<i>nad5</i>			11	1	1		
<i>rnl</i>	26		2	2	1		
<i>rns</i>	8	30	15	1	1		
<i>atp6</i>			3		1	447	28
<i>nad7</i>			9		1	553	89
<i>nad8</i>			3		1	259	46
<i>y1</i>	4	7	4	2			
<i>y2</i>	29	2	1	4			
<i>y3</i>	44	6	1	5			
<i>y4</i>	40			2			
<i>y5</i>	50	18		3			
<i>y6</i>	6		2				
<i>nad2 (murf1)</i>					1		
<i>nad3 (cr5)</i>					1	210	13
<i>nad4L (cr3)</i>					1	148	13
<i>nad9</i>					1	345	20
<i>rps12</i>					1	132	28
<i>cr4</i>					1	325	40
<i>murf2</i>					1	26	4
<i>murf5</i>					1		

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 diplomemids 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 diplomemid 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

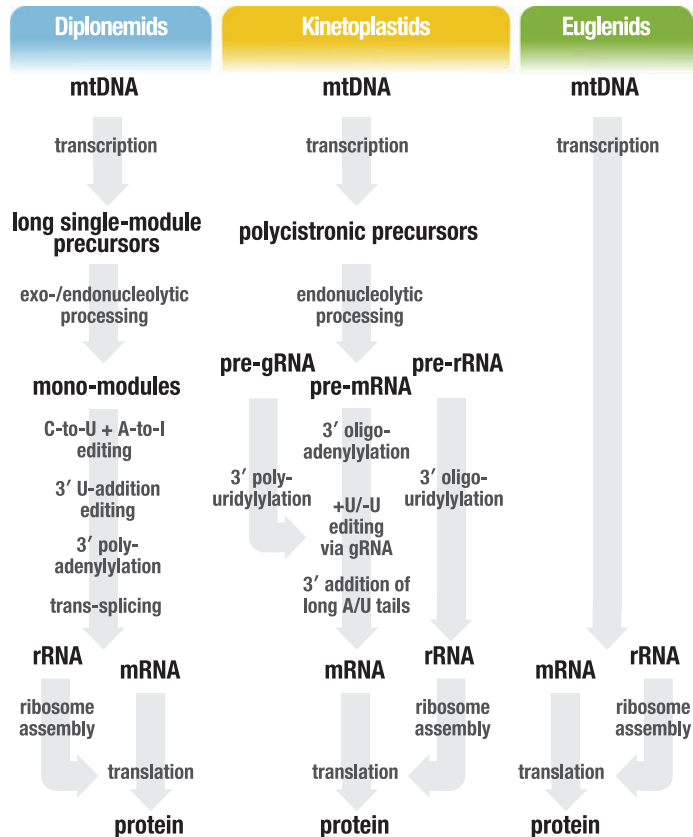


Fig. 6.5 Comparison of gene expression pathways among euglenozoans. Note that while the processes are sequential in kinetoplastids, diplomemids perform most processing and editing steps in parallel (see also Fig. 6.2)

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 diplomemid mitochondria is a highly parallelized process (Kiethega et al. 2013).

6.2.4 Partner Selection in Trans-splicing

In diplomemids 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 diplomemids, 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 anti-sense RNAs for *coxI* 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 *coxI*-m7 3' termini have been joined incorrectly, i.e., predominantly to *coxI*-m6 instead of *coxI*-m8. In contrast, poly-A libraries contain only about 0.4% of mis-joined *coxI*-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 *coxI* 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 diplomids 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 diplomid 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)
Insertion	Predominantly C (also U, A, G)	Dinoflagellates	mRNA, rRNA	Lin et al. (2002) and Jackson et al. (2007)
		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
	U	Metazoans	tRNA	Yokobori and Pääbo (1995)
		Kinetoplastids	mRNA	Vanfleteren and Vierstraete (1999) and Lavrov et al. (2016)
		Diplonemids	mRNA, rRNA	Reviewed in Read 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
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 Rossmannith 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 (*ml*) (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 ($\gamma 5$ -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 *rml*, 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*, $\gamma 3$), 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,

uridylation 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 homogeneously 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 diplomonads, 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 diplomonads (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 diplomonads. 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 diplomonads 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 diplomonad 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 diplomonad 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, Pawłowski 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 epibiotic 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 *Mycobacteres*. *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, Gailf F, Gorsky G, Grimsley N, Hingcamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki M, Speich S, Stemann 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, Fadrosch 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, Ivans 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-Díaz 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 diplomonids 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 diplomonids. *Curr Biol* 26:3053–3059
- Gonzalez TN, Sidrauskis 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
- Karkowska 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-Díaz 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 diplomemids. *Environ Microbiol* 11:47–55
- Lavrov DV, Adamski M, Chevalloné 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 diplomemids, 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/organelle 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 diplomemid 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, Krobisch 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, Rossmannith 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 coscinodiscivoris* 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, Puijn 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 kinetoplasts. *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 diplomemids 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 diplomemids. *RNA Biol* 13:1204–1211
- Valach M, Moreira S, Hoffmann S, Stadler PF, Burger G (2017) Keeping it complicated: mitochondrial genome plasticity in diplomemids. *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 *Postgaardia 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

Chapter 4

Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplomids.

Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplomonids

Binnypreet Kaur^{1,2,†}, Kristína Záhonová^{1,3,†}, Matus Valach^{4,*},
Drahomíra Faktorová^{1,2}, Galina Prokopchuk¹, Gertraud Burger⁴ and Julius Lukeš^{1,2,*}

¹Institute of Parasitology, Biology Centre, Czech Academy of Sciences, 37005 České Budějovice (Budweis), Czech Republic, ²Faculty of Sciences, University of South Bohemia, 37005 České Budějovice (Budweis), Czech Republic, ³Faculty of Science, Charles University, BIOCEV, 25250 Vestec, Czech Republic and ⁴Department of Biochemistry and Robert-Cedergren Centre for Bioinformatics and Genomics, Université de Montréal, H3T 1J4 Montreal, Canada

Received October 30, 2019; Revised December 14, 2019; Editorial Decision December 16, 2019; Accepted January 08, 2020

ABSTRACT

Diplomonids are highly abundant heterotrophic marine protists. Previous studies showed that their strikingly bloated mitochondrial genome is unique because of systematic gene fragmentation and manifold RNA editing. Here we report a comparative study of mitochondrial genome architecture, gene structure and RNA editing of six recently isolated, phylogenetically diverse diplomonid species. Mitochondrial gene fragmentation and modes of RNA editing, which include cytidine-to-uridine (C-to-U) and adenosine-to-inosine (A-to-I) substitutions and 3' uridine additions (U-appendage), are conserved across diplomonids. Yet as we show here, all these features have been pushed to their extremes in the Hemistasiidae lineage. For example, *Namystynia karyoxenos* has its genes fragmented into more than twice as many modules than other diplomonids, with modules as short as four nucleotides. Furthermore, we detected in this group multiple A-appendage and guanosine-to-adenosine (G-to-A) substitution editing events not observed before in diplomonids and found very rarely elsewhere. With >1,000 sites, C-to-U and A-to-I editing in *Namystynia* is nearly 10 times more frequent than in other diplomonids. The editing density of 12% in coding regions makes *Namystynia*'s the most extensively edited transcriptome described so far. Diplomonid mitochondrial genome architecture, gene structure and post-transcriptional processes display such high complexity that they challenge all other currently known systems.

INTRODUCTION

Diplomonids are heterotrophic marine flagellates belonging to the phylum Euglenozoa, which also includes the well-studied parasitic kinetoplastids and free-living euglenids (1,2) (Figure 1A). Diplomonids have been largely overlooked due to technical limitations, because the SSU rRNA V4 region, typically amplified in the metabarcoding approach, has expanded beyond typical lengths in diplomonids. In a recent survey, which targeted the more conserved V9 region, they have been detected in virtually every sample of seawater (3) and are currently ranked among the most diverse and abundant eukaryotic groups in the world's oceans (4–7). The ecological role of diplomonids in marine environments has only recently begun being appreciated (6,8,9).

We know close to nothing about the lifestyle of diplomonids (10). Their varied morphology and the recent finding of various bacterial endosymbionts in their cells (11–13) indicate that they have a versatile *modus vivendi*, likely enabling them to occupy widely different niches within the oceanic ecosystem. According to the 18S rRNA V9 region-based phylogenies, diplomonids fall in four major lineages: (i) 'classical' diplomonids (Diplomonidae) including both benthic and planktonic species of the genera *Diplonema*, *Rhynchopus*, *Lacrimia*, *Flectonema* and *Sulcionema*; (ii) hemistasiids (Hemistasiidae), a small planktonic clade composed of the genera *Hemistasia*, *Artemidia* and *Namystynia*; (iii) an extremely diverse clade of deep-sea pelagic diplomonids (or DSPD I, recently named Eupelagonemidae); and (iv) a second but relatively small clade of deep-sea pelagic diplomonids (DSPD II) (2,11,12,13,14).

The most conspicuous features of diplomonids are their unique and complex mitochondrial genome architecture

*To whom correspondence should be addressed. Tel: +420 38 777 5416; Fax: +420 38 531 03882; Email: jula@paru.cas.cz

Correspondence may also be addressed to Matus Valach. Tel: +1 514 343 6111; (Ext. 5172); Fax: +1 514 343 2210; Email: matus.a.valach@gmail.com

[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

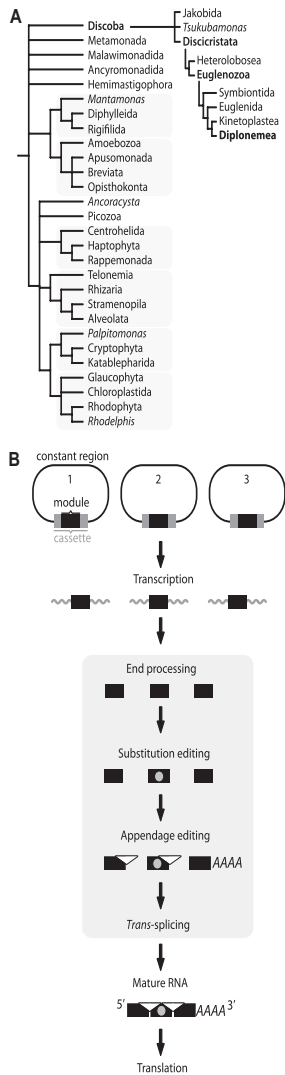


Figure 1. Phylogenetic position of diplomonids and their mitochondrial DNA structure and gene expression. (A) Recent classification of eukaryotes (based on (1)) highlighting the position of diplomonids. (B) The various steps of mitochondrial gene expression in diplomonids. The model gene consists of three pieces, also referred to as modules, each encoded in a unique region (cassette) on a different chromosome. Modules, together with surrounding regions, are transcribed separately from a promoter located in the constant region (25). Primary transcripts are end-processed, removing 5' and 3' non-coding regions from the primary transcripts. Certain module transcripts undergo substitution RNA editing and/or appendage RNA editing (nucleotide additions at the module's 3' end). The module transcript that will constitute the transcript's 3' end is poly-adenylated (mRNAs and mtLSU rRNA) or poly-uridylyated (mtSSU rRNA) (16,19). Finally, modules are joined together (*trans*-spliced) yielding mature RNA (mRNA or rRNA). Note that all post-transcriptional processes (gray background) occur in parallel in the diplomonid mitochondrion (19); thus, the arrows do not imply strict sequentiality.

and gene expression (Figure 1B), studied in depth in the type species *Diplonema papillatum* (reviewed in (15–17)). For example, the amount of mitochondrial DNA (mtDNA), estimated at 250 Mbp, is so far the highest recorded for an organelle (18). Further, its mtDNA is composed of >80 covalently closed non-catenated 6 and 7 kbp-long circular chromosomes. Except for a short unique region called the 'cassette', chromosomes consist mostly of repetitive sequence termed 'constant region', which is essentially identical across chromosomes of a given size class (Figure 1B). It is the unique cassette that typically encloses a single gene fragment, also called module, the size of which ranges from 40 to 540 bp. Chromosomes including cassettes are transcribed separately, then the non-coding portions are removed leaving behind module-only transcripts that are subsequently joined to their cognate neighboring modules derived from other circles (16,19). In this way, mature transcripts (mRNAs and rRNAs) are assembled via massive *trans*-splicing, the mechanism of which remains unknown.

Trans-splicing in diplomonid mitochondria differs from that observed in organelles and in the nucleus of nematodes and other eukaryotes including diplomonids (20–22), because the former process is apparently catalyzed by neither spliceosomes nor Group I or Group II splicing machineries (16,19,23–26). With the exception of the mitochondrial small subunit ribosomal RNA (mtSSU rRNA), all *D. papillatum* genes undergo this assembly process, making the extent of *trans*-splicing unprecedented.

Moreover, in addition to gene fragmentation compensated by *trans*-splicing, *D. papillatum* mitochondrial transcripts are subject to extensive RNA editing of two fundamentally different types: post-transcriptional uridine addition (U-appendage) at 3' ends of certain modules (unique to diplomonids), and deaminations of adenosines to inosine (A-to-I) and cytidines to uridines (C-to-U) at numerous positions within coding regions (16,27).

Studies of three other diplomonid species (*D. ambulador*, *Flectonema neradi* and *Rhynchopus euleicides*) showed little deviation from the features observed in the type species (27,28), except that across these taxa, the size of mitochondrial chromosomes ranges from 2 to 12 kbp (27,28). However, it was reported recently that gene fragmentation, as well as U-additions and A and C substitutions in transcripts are much more frequent in *Hemistasia phaeocysticola*, the single hemistasiid species examined until now (27,29). While *D. papillatum* has at most 11 modules per gene (16), the genes of *H. phaeocysticola* are fragmented twice as much (29).

Does the single examined hemistasiid species represent an exceptional case, or can fragmentation and RNA editing of mitochondrial transcripts reach even higher levels of complexity? To address this question, we examined species that have recently become available in culture, with their morphology, ultrastructure and life cycles described (11–13). Here, we present the most extensive comparative study of diplomonid mtDNA performed thus far. We show that the degree of mitochondrial RNA editing and gene fragmentation can reach unprecedented complexity, highlighting several questions about the role and evolution of these remarkable features.

MATERIALS AND METHODS

Strains, culture conditions and nucleic acids extraction

The six diplomonid species used in this study (Supplementary Table S1) were recently isolated from marine water collected in aquaria, lagoons and sandy beaches of Japan (11–13). The species were axenically cultivated in a medium containing 3.6% sea salts (Sigma-Aldrich, S9883), supplemented with 1% (v/v) heat-inactivated horse serum (Sigma-Aldrich, H0146) and 0.025 g/l LB broth powder (Sigma, L3522). The medium was filter-sterilized using a 0.22- μ m filter.

Total DNA from exponentially growing cultures was isolated using MasterPure Complete DNA and RNA Purification Kit (Lucigen, MC85200) specially designed for the isolation of DNA from marine organisms. RNA was extracted from whole cells using TriReagent (MRC, TR118) to prevent the loss of small RNAs corresponding to processing and *trans*-splicing intermediates. Residual DNA was removed by DNase treatment followed by extraction with a homemade Trizol substitute (30).

Reverse transcription, RT-PCR, 5' and 3' RACE, and poly-A tail site mapping

Reverse transcription was performed with First Strand cDNA Synthesis Kit for subsequent PCR (Roche) or with SuperScript IV Reverse Transcriptase (Thermo). Complementary DNA was amplified with Q5 High-Fidelity DNA Polymerase (New England Biolabs). PCR products were purified using the QIAquick Gel Extraction kit (Qiagen), Wizard SV Gel and PCR Clean-Up system (Promega) or Monarch DNA Gel Extraction Kit (New England Biolabs). Mapping of 3' polyadenylation sites was performed using an oligo-dT primer and a gene-specific primer. To determine the 5' and 3' ends of modules (5' and 3' RACE), the RNA adapter-oligonucleotide dp124 and the 5' RACE Adapter (from FirstChoice RLM-RACE Kit, Invitrogen) was ligated to the RNA using T4 RNA ligase I (New England Biolabs) and *Ec*RtcB RNA ligase (New England Biolabs), respectively. Detailed protocols are available at <https://www.protocols.io/researchers/matus-valach>. RT-PCR was performed using specific primers, and amplicons were sequenced at the IRIC Genomics Core Facility (Montreal, Canada) or at Eurofins Genomics (Ebersberg, Germany). Primer and adaptor sequences are listed in Supplementary Table S2.

Genome and transcriptome sequencing and assembly

Both library preparation and sequencing of genomes and transcriptomes were outsourced to the Genome Quebec Innovation Centre (Montreal, Canada). Single DNA-Seq and RNA-Seq library per species were produced due to limited material availability. Illumina genomic paired-end libraries were constructed from total DNA and sequenced in a single Illumina MiSeq lane. DNA reads were assembled using SPAdes v3.11.1 (31) and alternatively, the Tadpole assembler (part of the BBTools suite; <https://jgi.doe.gov/data-and-tools/bbtools/>).

To avoid the huge variety of *trans*-splicing and RNA editing intermediates present in diplomonid mitochondria (16,19), which complicate analysis and interpretation, we opted for the enrichment of mature mitochondrial transcripts, i.e. the polyadenylated (poly-A) RNA fraction, which was isolated from total RNA to construct strand-specific RNA-Seq libraries with an average insert size of ~200 nt. Libraries were sequenced on an Illumina HiSeq platform. For *de novo* assembly, Trinity v2.2.0 software was used with default parameters (32). Read counts and lengths for both DNA and RNA sequencing are listed in Supplementary Table S1. The raw sequencing data are available at NCBI (<https://www.ncbi.nlm.nih.gov/>) as BioProject PRJNA525750.

Two of the examined species (*D. japonicum* and *N. karyoxenos*) contain endosymbiotic bacteria (12,13). For the purpose of this study focusing on mitochondrial sequences, it was not necessary to estimate relative abundance of bacterial sequences in the datasets. However, we did perform RNA-Seq read mapping to DNA contigs and found only negligible differences between mapping rates to sequences of endosymbiont-bearing and -lacking species, which suggested that possessing an endosymbiont did not introduce any significant bias to our strategy.

Identification of transcripts and annotation of genomic modules

Candidate contigs originating from the mitochondrial transcriptome were identified by BLASTx searches (33) using protein sequences of previously identified mature mitochondrial mRNAs from *D. papillatum*, *D. ambulador*, *H. phaeocysticola*, *F. neradi* and *R. eleeides* as queries, and the transcriptome assemblies of the new diplomonids as queried databases. Genomic modules for each species were annotated based on BLASTn searches using predicted transcript sequences as queries. To validate module assignments, modules were aligned with the respective transcript using the built-in aligner of the Geneious 10.1.3 software (34) and visually inspected. To infer protein-coding ORFs, the nucleotide sequences were conceptually translated using NCBI's genetic code Table 4 (TGA = Trp). Identified modules are cataloged in Supplementary Table S3.

Completion of mitochondrial transcripts

RNA-Seq reads were mapped to reference sequences with Bowtie2 (35). Especially in the case of *nad* genes, terminal modules were missing. They were recovered by mapping RNA-Seq reads onto partial transcript contigs, and subsequently by extending the sequences via RT-PCR up to the polyA tail using oligo-dT and gene-specific primers (see above). To screen the contigs across the investigated species for the highly divergent *nad* genes (the previously designated *y* genes (36), we employed HMMER 3.1b2, a most sensitive method based on profile hidden Markov models (37).

Chromosome classification

The module-containing contigs were first extended by read mapping using the mapper software implemented in the

Geneious 10.1.3 package (34) and then compared with each other by BLASTn. Cassettes were identified by the same criteria as described previously (27). Briefly, cassettes are unique sequences surrounding modules, and are flanked by the constant regions of a chromosome, which we define as sequences with >90% identity over >100 bp adjacent to cassettes (Figure 1B). Contigs bearing the same cassette-flanking regions were assigned to the same chromosome class. For multiple sequence alignments, we used MAFFT v7.388 (38). Chromosome classes were ordered from the highest to the lowest member count and named A, B etc. To calculate mean read coverage of cassettes or modules, DNA-Seq reads were first mapped onto mitochondrial contigs by the Geneious 10.1.3 software. Aligning reads were merged with BBMerge (rem k = 62 extend2 = 50 ecct) and deduplicated with Dedupe (k = 31 ac = f) from BBTools. The resulting reads were mapped with Bowtie2 onto the reference sequence, and mean coverage was calculated using the Pileup tool from BBTools. For *N. karyoxenos* mitochondrial chromosome sequences, which are highly polymorphic, DNA-Seq reads were mapped with the Geneious 10.1.3 software without subsequent Bowtie2 mapping. The final list of identified chromosomes has been compiled in Supplementary Table S4.

In silico identification of RNA editing and DNA polymorphic sites

RNA editing clusters and longer insertions were identified by comparing the genomic contigs and mature transcript sequences by BLASTn. To distinguish between RNA editing sites and genomic polymorphisms, RNA-Seq and DNA-Seq reads were mapped onto mitochondrial transcripts and genomic contigs, respectively, with the built-in aligner of the Geneious 10.1.3 software. DNA polymorphic sites were identified as those exhibiting two (or sometimes more) nucleotides in the mapped reads, while in the case of RNA editing sites, the consensus nucleotide in genomic reads differed from that in RNA-Seq reads. A genomic position with >10% reads displaying difference from the reference was considered a polymorphic site. For RNA editing, a position was annotated as an editing site if at least 50% reads carried a base change. Note that a vast majority of sites was edited to >90%. RNA editing and DNA polymorphic sites of each transcript are detailed in Supplementary Table S5.

Phylogenomic analysis

We used all 15 assigned mitochondrially encoded protein sequences (Atp6, Cob, Cox1/2/3 and Nad1/2/3/4/4L/5/6/7/8/9) from 11 diplomemids and the corresponding homologs from other discobans, namely *Trypanosoma brucei*, *Bodo saltans* and *Perkinsela* sp. (Kinetoplastida); *Euglena gracilis* (Euglenida), *Acrasis kona*, *Naegleria gruberi* and *Stachyamoeba lipophora* (Heterolobosea); *Tsukubamonas globosa* (Tsukubamonadida); and *Andalucia godoyi*, *Reclinomonas americana* (ATCC 50394) and *Ophirina amphinema* (Jakobida). Sequences were downloaded from the NCBI GenBank Protein database except those of *B. saltans*. We retrieved the latter through tblastn searches (using *T. brucei* mitochondrial proteins

as queries) from transcripts that we assembled by Trinity v2.2.0 using RNA-Seq data deposited in the NCBI Bioproject PRJEB3146. Multiple sequence alignments (MSAs) of proteins were generated with MAFFT v7.38 (38) using the E-INS-i algorithm and default parameters. Protein alignments were stripped of hyper-variable sites (20% gap threshold) with trimAl v1.4.22 (39). Subsequently, protein sequences were concatenated for each species, with the final MSA containing 22 taxa and 5,063 positions. Phylogenetic inferences were performed by a Bayesian approach using posterior probabilities as support values (PhyloBayes v4.1 (40), MrBayes v3.2.6 (41)) and by maximum likelihood with bootstrapping (IQ-TREE v1.6.10 (42) and RAxML v8.2.11 (43)). Bayesian methods were executed in two independent chains and the first 25% cycles were discarded as burn-in. For PhyloBayes, we chose the substitution model CAT-GTR and site rate variation modeled as a Dirichlet process (ratecat option); the chains were stopped after they converged (i.e. maxdiff below 0.1 at ~750 cycles corresponding to ~25,000 generations). For MrBayes, we chose the GTR model with six discrete categories of gamma rate variation and 200,000 MCMC generations. For ML computations, we chose the substitution matrix LG for amino acid frequencies, which was determined as the best model by Model Finder (44). For IQ-TREE, we used default parameters with the option to calculate 1,000 ultrafast bootstrap replicates. For RAxML, additional parameters were: 50 categories for rate heterogeneity (CAT option), the algorithm 'rapid bootstrap analysis' and 100 distinct alternative runs on distinct starting trees for bootstrap support values. To evaluate the reliability of the inferred tree, we further analyzed the gene- and site-concordance factors (gCF and sCF, respectively) for each branch, as implemented in IQ-TREE v1.7 (45), with default parameters and the option to merge models across loci.

RESULTS

Gene repertoire

We examined mitochondrion-encoded genes from four Diplonemidae species (*Diplonema japonicum* strain YPF1604, *Rhynchopus humris* YPF1608, *Lacrimia lanifica* YPF1601 and *Sulcionema specki* YPF1618) and two Hemistasiidae species (*Artemidia motanka* YPF1610 and *Namystynia karyoxenos* YPF1621). In all six species, we identified the same set of genes described earlier in four Diplonemidae species (27), namely genes encoding ATP synthase subunit 6 (*atp6*), cytochrome *b* (*cob*), three cytochrome *c* oxidase subunits (*cox1*, *cox2* and *cox3*), 10 NADH dehydrogenase subunits (*nad1*, *nad2* [previously *y3*], *nad3* [*y1*], *nad4*, *nad4L* [*y6*], *nad5*, *nad6* [*y5*], *nad7*, *nad8* and *nad9* [*y2*] (36), as well as small and large subunit mitochondrial ribosomal RNAs (*rns* and *rnl*) (Figure 2; Supplementary Tables S3 and S4). In the two hemistasiids, we failed to detect *nad6* [*y5*], but this was presumably due to the gene's high divergence and not to its genuine absence. Moreover, *Lacrimia*, *Sulcionema* and *Namystynia* also encoded *y4*, a gene first discovered in *D. papillatum*; in *R. humris*, we found a candidate corresponding to module 2 from *D. papillatum* (*y4-m2*), but not *m1*. With homologs of the

latter gene at hand, we revisited data from a previous study (27), which enabled us to detect the two *y4* modules in *R. euleeides*, unrecognized previously because of extensive overlaps with *cox3-m1* and *nad5-m11* in this species. In *Lacrimia*, we found an additional gene, named *y7* (single module), which potentially codes for a protein of 67 amino acid residues. No tRNAs were found; as in other euglenozoan species, they are apparently not mitochondrially encoded, but reside on nuclear DNA and are imported to the mitochondria.

The inferred mitochondrion-encoded proteins of all species analyzed here and those studied previously (27) displayed an exceptionally low level of sequence conservation, which made detection of most genes challenging. *Cox1* was the most conserved protein across the diplomids (32.7% identity across 11 species), while *Nad3* with a mere 2.5% sequence identity was on the other end of the spectrum (Supplementary Figure S1).

Module numbers and sizes

The four Diplonemidae species investigated here build their 17–19 identified mitochondrial genes from essentially the same number of modules as does the type species *D. papillatum*. The sole exception is *Sulcionema rms*, which is encoded by two modules instead of one in all other ‘classical’ diplomids. In the two Hemistasiidae species, the total number of modules is doubled (Table 1 and Figure 2), while module sizes are halved (Figure 3). In fact, a given module observed in Diplonemidae is typically split into two to four modules in Hemistasiidae, since gene breakpoints are typically conserved across diplomids (for sequences that can be confidently aligned, occasional shifts are less than 6 bp).

About 4% of modules in hemistasiids are shorter than 20 bp—even as short as 3 bp—and referred to in the following as mini-modules (Figure 4A and Supplementary Figure S2). It should be noted that mini-modules cannot be unambiguously distinguished from appendage RNA editing (see also below) by inspection of DNA–RNA sequence differences alone. Still, two lines of evidence support mini-modules. First, we confirmed by 3' RACE and subsequent sequencing of *Artemidia cox3* the existence of an mRNA *trans*-splicing intermediate containing the putative *cox3-m9* mini-module. The 3' RACE RT-PCR sampled two amplicon populations: a minor one with 5 Us appended to the 3' terminus of the upstream module 8 and a major one with the triplet CAG, corresponding to the mini-module 9, joined to the aforementioned U-tract (Figure 4B). The triplet thus appeared to have been added as a whole, i.e. *trans*-spliced, rather than as a succession of unrecognized editing events. However, to completely rule out the latter alternative, a more extensive sampling of 3'-end RNA processing and editing intermediates by RNA-Seq would be necessary. Second, in the available RNA-Seq data, we detected RNA processing and *trans*-splicing intermediates, which contained in addition to the diminutive module its flanking sequence, thus indicating where in the genome it resided. This way, we could trace back the genomic source of four and eight such gene pieces in RNA-Seq reads of *Artemidia* and *Namystynia*, respectively. (Note, however, that for six additional short segments of 2–6 nt in five transcripts of *Namystynia* this was not pos-

sible and the corresponding regions have been marked as unresolved [Figure 2F].) For the putative *cox3-m9* mini-module, we could detect two RNA-Seq reads containing the CAG triplet joined upstream to the cognate *cox3-m8* (with the appended U-tract) and downstream to *cox1-m21* (Figure 4C). This indicated that *cox3-m9* and *cox1-m21* of *Artemidia* were actually juxtaposed on the same chromosome U01. The observation of the intermediates containing mini-modules with their flanking sequences led us to hypothesize a possible assembly scenario for mini-modules, where a larger precursor acts as a mini-module carrier (Figure 4D).

Classes of mitochondrial chromosomes

Knowing the module sequences from transcriptome data allowed us to identify the corresponding regions in genomic contigs, while the repetitive sequences adjacent to the unique module-flanking regions allowed us to delimit cassettes (see Figure 1). Further, cassette-flanking repetitive sequences were presumed to be part of the constant regions of chromosomes, according to the classification scheme of mitochondrial chromosomes in *D. papillatum* (16), and thus allowed categorization of chromosomes into multiple classes (Table 1; see ‘Materials and Methods’ section for details). In this way, 4 classes were established in *Sulcionema*, 5 in *D. japonicum* and *R. humris*, 8 in *Lacrimia* and 17 in *Artemidia*.

For *Namystynia* chromosomes, in contrast to the other species, we could not employ the criterium of recurring cassette-flanking sequences (representing the constant regions), because sequences around modules frequently consisted of unequally spaced tandem and dispersed repeated homooligomeric motifs that could not be unambiguously aligned. However, numerous chromosomes shared motif 1 (5'-GGGCCA AAAA-3') upstream and motif 2 (5'-TTTTGGGCC-3') downstream of the cassettes. Consequently, all chromosomes bearing these motifs were classified as class X, which is much more diverse than the classes from the other diplomids. Finally, in every species, a handful of chromosomes did not fit into a defined class and therefore were grouped into the category ‘unclassified’. Total counts of classified and unclassified chromosomes for each species are summarized in Table 1 and Supplementary Table S4.

Since the sequence repeats prevented the assembly of whole chromosomes from the available short reads (except for two cases in *Sulcionema*; see below), chromosome sizes remain unknown. Nevertheless, the assembled genomic contigs indicate that the overall chromosome architecture conformed to that previously observed in other diplomids (24,27,28), namely cassette sizes varied from ~0.2 to ~2 kbp with median length ~330 (±50) bp. Two types of deviations were observed. First, the median size of *D. japonicum* cassettes was ~1 kbp; as we detail below, this was due to the unusually high number of modules per cassette. Second, *Sulcionema* and *Artemidia* chromosomes contained cassettes with sizes well above the 2 kbp mark (from 3.2 to 12.3 kbp). Based on the complete assembly of three *Sulcionema* class D chromosomes (Supplementary Table S4), we calculated that these long cassettes covered

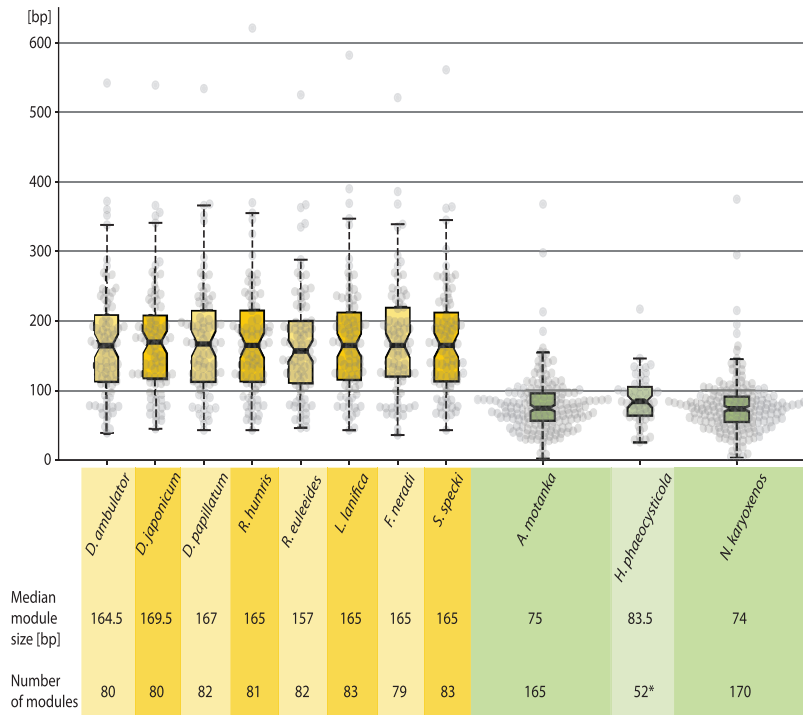


Figure 3. Average gene module size in diplo-nemids. The average module sizes of mitochondrial genes from all diplo-nemids for which data are available. Modules in Hemistasiidae species (right) are about half the size compared to those from the Diplonemidae clade (left). Four diplo-nemid species and *H. phaeocysticola* (light hues) were studied previously. Asterisk for *H. phaeocysticola*, the structure of only four genes is known (*cob*, *cox1*, *cox2* and *nad7*).

Table 1. Mitochondrial chromosomes in studied diplo-nemids

Group	Species	Chromosomes classes			Unclassified chromosomes				Total chromosome count (No. of modules)	
		No. of classes (types)	No. of mono-module chromosomes (types)	No. of multi-module chromosomes (types)	No. of empty chromosomes (types)	No. of mono-module chromosomes	No. of multi-module chromosomes	No. of chromosomes with one module		No. of chromosomes with multiple modules (No. of modules)
Diplonemidae	<i>D. japonicum</i>	5 (A–E)	4 (C, D)	20 (A, B, E)	–	3	3	7	23 (73 ^a)	30 (80 ^a)
	<i>R. humris</i>	5 (A–E)	59 (A–E)	5 (A, B, D)	2 (A)	6	1	65	6 (16)	73 (81)
	<i>L. lanifica</i>	8 (A–H)	71 (A–H)	4 (A, B, D)	7 (B, C, E)	4	–	75	4 (8)	86 (83)
	<i>S. specki</i>	4 (A–D)	38 (A)	9 (A–D)	16 (A)	1	1	39	10 (44 ^b)	65 (83 ^b)
	<i>A. motanka</i>	17 (A–Q)	107 (A–K, M–Q)	15 (A, B, E–G, I, L, N, P)	n.d.	22	2	128	18 (37 ^c)	146 (165 ^c)
Hemistasiidae	<i>N. karyoxenos</i>	1 ^d (X)	137 (X)	5 (X)	n.d.	20	1	157	6 (13)	163 (170)

n.d., not determined.

^a*atp6-m1* on two different chromosomes (counted once).

^b*nad2-m2* on two different chromosomes (counted once).

^c*nad3-m5* on three, and *nad4-m3* and *nad4-m5* on two different chromosomes (counted once).

^dChromosomes assigned to a class based on different criteria than in other species.

83–90% of the circles, the complete opposite of the situation in other analyzed diplo-nemid chromosomes, where a cassette represents only 5–10% of the chromosome length (24,27,28).

Module content and arrangement

Diplonema papillatum has 81 distinct mitochondrial chromosomes, 76 of which carry a single cassette that in turn

contains a single module (mono-module/mono-cassette organization). The remaining chromosomes contain one cassette each that encloses two modules (three instances; multi-module/mono-cassette organization) or cassettes without any identified module (two instances). In the other diplo-nemids examined previously and here, additional arrangements coexist, notably three or more (up to 11) modules per cassette and also two cassettes per chromosome (multi-module/multi-cassette organization) (27). In these

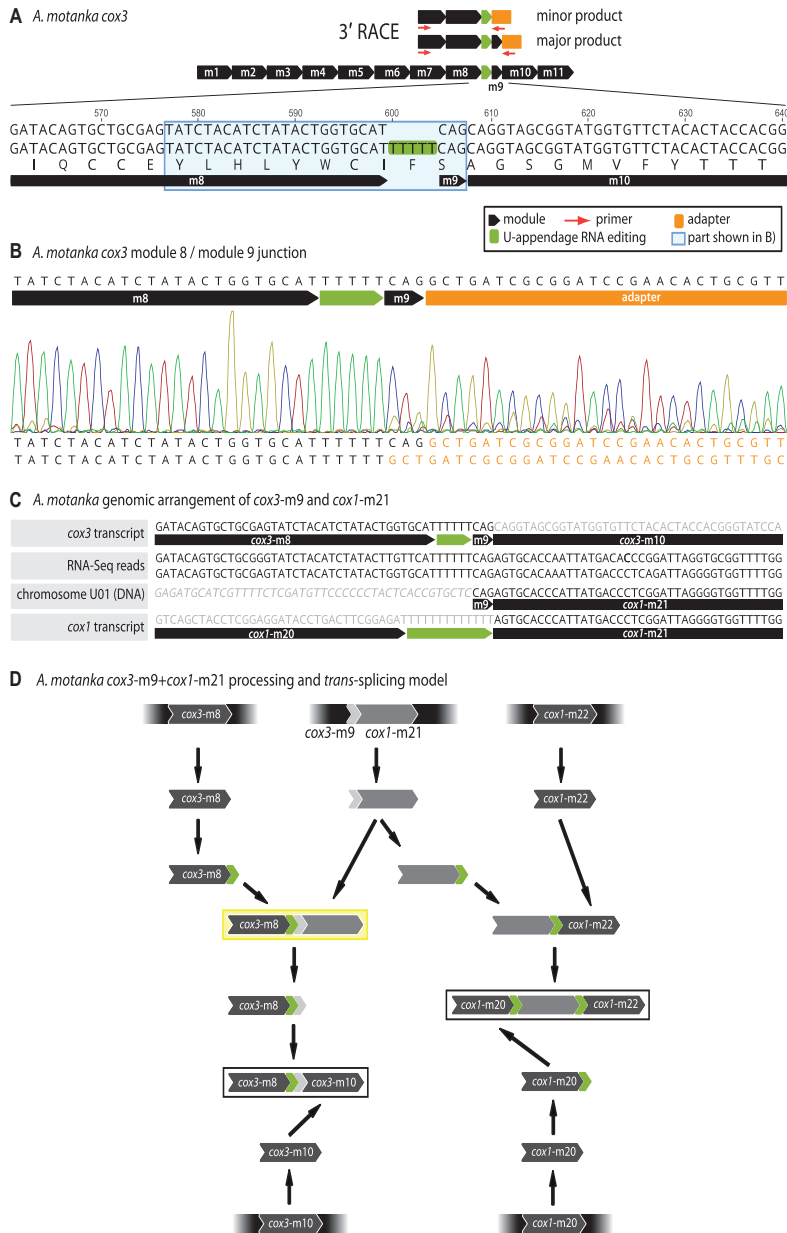


Figure 4. Mini-modules and new RNA editing types. (A) Example of the putative 3 bp-long mini-module *cox3*-m9 from *Artemidia*. The upper map shows the 3' RACE approach and the location of the mini-module in the transcript. (B) Sequence chromatogram of a 3' RACE amplicon including *cox3*-m9. Note the mixed chromatogram peaks downstream of the T-tract indicating a mixture of RT-PCR products with or without the CAG triplet, causing a 3-nt phase shift. (C) The *cox3*-m9 mini-module is encoded adjacent to *cox1*-m21 in the chromosome U01. The mini-module-encoding locus was inferred from the shown RNA-Seq reads that cover *cox3*-m8 with its appended U-tract, followed by the CAG triplet flanked by the *cox1*-m21 sequence. The non-coding region of the chromosome is set in italics. (D) Hypothetical scenario of the RNA processing pathway of the adjacent *cox3*-m9 and *cox1*-m21 modules and *trans*-splicing to their cognate partners. In this model, the larger precursor acts as a mini-module 'carrier'. The yellow box indicates the RNA intermediate identified in panel (C). The intermediates in black frames illustrate the expected, cognate modules up- and downstream of *cox3*-m9 and *cox1*-m21.

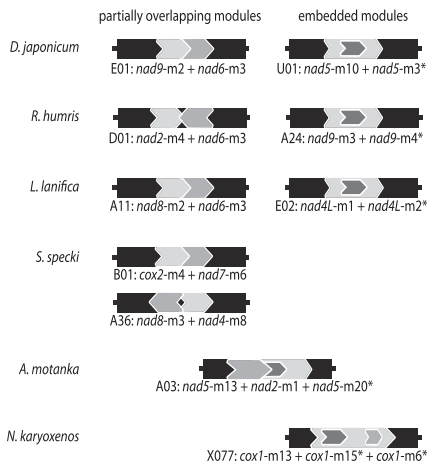


Figure 5. Overlapping gene module arrangements. Scheme of representative cassettes with overlapping modules detected in the six diplomonids studied here. Note that *S. specki* does not contain embedded modules and that *N. karyoxenos* lacks partially overlapping modules. Constant regions of chromosomes (indicated by chromosome IDs E01, U01, etc.; see also Supplementary Table S4) are depicted as black rectangles. Modules are represented by dark- and light-gray filled arrows. The arrow tip indicates the direction of module transcription. >>, modules encoded on the same strand; <>, 5'-ends of modules encoded on opposite strands overlap; ><, 3'-ends of modules encoded on opposite strands overlap.

latter instances, modules are either separated, overlapping or nested.

The six species analyzed here differed considerably in their total number of distinct chromosomes, ranging from 30 in *D. japonicum* to 163 in *Namystynia* (Table 1 and Supplementary Table S4). These differences were not only due to a different number of modules in a given species, but also to the fact that some chromosomes encoded multiple modules. For example, *D. japonicum* contained 23 multi-module chromosomes, the highest number in this category among all species analyzed (27), but only 7 mono-module chromosomes (Table 1). In contrast, among the 86 chromosomes of *Lacrimia*, only four contained multiple modules. The highest number of modules detected in a single chromosome was 11 in *Sulcionema*; this species also contained by far the largest number of apparently module-less cassettes (all 16 from its A class chromosomes; Supplementary Table S4).

About 60% of multi-module chromosomes of the diplomonids examined here contained partially overlapping or nested modules (Supplementary Table S6), an arrangement also noted before in other classical diplomonids (27). Overlaps were only conserved among closely related species (*nad5-m10 + nad5-m3*, *nad6-m2 + nad4L-m2* and *nad9-m2 + nad6-m3* in *D. ambulator* and *D. japonicum*; *nad9-m3 + nad9-m4* in *R. euleiides* and *R. humris*), and most overlapping modules were encoded on the same strand. No embedded modules were detected in *Sulcionema* (similarly to *D. papillatum*), whereas in *Namystynia*, all overlapping modules were completely nested (Figure 5 and Supplementary Table S6).

Conserved types of mitochondrial RNA editing and their distribution

In all species studied here, we identified mitochondrial transcripts that underwent multiple events of C-to-U and A-to-I substitution editing and U-appendage editing (Figures 2 and 6; Supplementary Table S5). These types of editing had also been described previously in four Diplonemidae species and *H. phaeocysticola* (16,27). In the type species, the presence of inosines in transcripts has been demonstrated experimentally indicating that post-transcriptional A-to-G DNA-RNA differences arose by deamination, a process that most certainly applies to C-to-U changes as well (16). Substitution editing sites occurred in clusters in similar gene regions across species, although individual sites did not necessarily coincide. Earlier reported substitution editing clusters (e.g. *nad2-m4*, *nad3-m2*, *nad4-m1*, *nad6-m1*, *nad7-m3* and *m5*, *nad9-m1* and *rns*) were present in the species studied here as well, although with some exceptions, such as the complete absence of substitution editing sites in *nad3*, *nad7* and *nad9* of *Lacrimia* (Figure 2). We also identified new sites of C-to-U and A-to-I substitutions and U appendage. These included one and two new C-to-U editing sites in *Lacrimia cob-m5* and *cox3-m1*, respectively. Further, *Sulcionema* possessed a novel A-to-I substitution site in *cox3-m3*, and in *R. humris*, we discovered two new editing sites at the junction of *m1* and *m2* of *cox3* (Supplementary Table S5).

All the above editing types were much more frequent in the hemistasiids and affected every single transcript (Figures 2 and 6; Supplementary Table S5). In addition to the editing clusters documented previously (16,27), we identified several novel instances, mostly located at the ends of modules (which complicated the recognition of the corresponding modules). Although more numerous, substitution editing sites in *Artemidia* (493 A-to-I and 620 C-to-U sites in >100 editing clusters) were amassed in half the number of clusters compared to *Namystynia* (458 A-to-I and 588 C-to-U sites in >210 editing clusters; Supplementary Table S5). Interestingly, certain editing sites in *Namystynia* coincided with one of the >300 genomic polymorphisms (mostly single-nucleotide polymorphisms, SNPs) dispersed across modules (Supplementary Table S5 and Supplementary Figure 3A).

Novel types of RNA editing

By inspecting DNA-RNA differences, we detected two new types of editing not documented before in diplomonids. First, the two hemistasiids carried G-to-A substitutions (Figure 7A; Supplementary Figure S3A,B), notably 14 unambiguous sites in six different transcripts of *Namystynia* and one such site in *nad5* of *Artemidia* (Figures 2 and 6; Supplementary Table S5); a second site may exist in *Artemidia nad1*; however, the corresponding A in RNA could have also originated by A-appendage (see below).

The second new type of editing was detected in *Sulcionema* and *Namystynia*. The *nad4* transcript of the former contained between *m4* and *m5* not only a non-encoded U but also an additional A (Figure 2), which we confirmed by 3' RACE (Supplementary Figure S3C). Such A-appendage editing appeared far more frequent in *Namysty-*

nia, in which we spotted one to four such sites in eight different transcripts, summing up to a total of 17 editing positions and 41 post-transcriptionally added As (Figure 2 and Supplementary Table S5). The A-appendage site of *Sulcionema nad4* also occurred in the same transcript of *Namystynia* (m13-m14 junction). Curiously, at all sites except the one in *nad4L*, A-appendage in *Namystynia* coincided with U-appendage, occasionally in an interspersed fashion, as for example between *nad4*-m1 and m2 (5'-UAAUUUUUUUUU-3').

In *Namystynia*, we also observed six cases of apparent G-additions between modules, and again intermixed with other post-transcriptionally added nucleotides, for example 5'-GAUUU-3' between *nad5*-m12 and m13 (Supplementary Table S5). The Gs could have arisen in two ways, by (i) genuine G-appendage editing or (ii) A-appendage followed by A-to-I deamination. While editing by rare G-insertions had been observed in an amoebozoan and a heterolobosean (46,47), we considered the second alternative more likely because it would not imply an additional machinery required to specifically add G residues. Further, RNA-Seq read mapping to the junction of *nad5*-m12 and *nad5*-m13, where such a G-addition was noted, revealed a minor population of RNA-Seq reads that displayed 3' terminal A-tracts, which we interpret as not-yet edited intermediates (Figure 7B). It will be interesting to validate this hypothesis by biochemical assays once *Namystynia* becomes more amenable to experimental work.

Phylogenetic relationships among diplomonids

In molecular phylogenies based on 18S rRNA (11,12,14) (Figure 6), the genera *Diplonema*, *Rhynchopus*, *Lacrimia*, *Flectonema* and *Sulcionema* formed Diplomonidae, which are also referred to as 'classical' diplomonids, to the exclusion of Hemistasiidae, Eupelagonemidae (formerly DSPD I (48)), and the lineage currently described by its acronym 'DSPD II'. To build a more robust phylogeny, we used here the concatenated protein sequences inferred from 15 different mitochondrial transcripts, identified in this and previous studies. The resulting tree (Figure 6, right tree; Supplementary Figure S4A) resolved the relationships between diplomonids with high confidence but differs in topology from the nuclear 18S rRNA-based trees including the same species (Figure 6, left tree). First, in the mitochondrial phylogeny, *Lacrimia* grouped together with *Diplonema*, while it is placed at the base of the *Diplonema* + *Rhynchopus* clade in the 18S rRNA trees. Second, *Namystynia*, and not *Artemidia*, was the sister taxon of *Hemistasia*. Finally, the most significant deviation was the position of *Sulcionema*, as it branched together with Diplomonidae in the nuclear trees, but with Hemistasiidae in the mitochondrial tree.

To examine possible reasons for this incongruence, we calculated gene- and site-concordance factors for each branch in the tree (45) (Supplementary Figure S4B). Compared to the concatenated dataset, single-locus phylogenies showed comparably low support for the positions of *Lacrimia* and *D. papillatum*; however, the majority of informative sites in the concatenated dataset supported the tree topology. The conflicting positions within Hemistasiidae were mainly due to the limited data currently avail-

able for *H. phaeocysticola* (i.e. four proteins instead of 15). More importantly, most single-protein phylogenies placed *Sulcionema* prior to the divergence of hemistasiids, i.e. the topology shown here (Supplementary Figure S4A), but the site-wise support for this topology was almost identical with that of the two other mutually exclusive topologies, in which *Sulcionema* formed a sister group to either all diplomonids or the Diplomonidae clade. Further taxon sampling of basal diplomonids should allow to resolve the described inconsistencies.

DISCUSSION

In the past, most studies were performed on the type species *D. papillatum* which, incidentally, has recently become genetically tractable (49). Together with three other classical diplomonids that had formerly been examined at the molecular level (11–14), these species represent a tiny and ecologically restricted fraction of the highly diverse group (4,14). Here, we have considerably expanded the knowledge about the diplomonid mitochondrial genomes by examining six recently isolated species from both the Diplomonidae and Hemistasiidae clades, thus covering a substantial part of diplomonid diversity (11–13).

Diplomonids are record holders in mitochondrial genome content and organization

It is worth noting that these protists, which were neglected until recently, carry the largest amount of mtDNA documented in an organelle, which in *D. papillatum* even exceeds that of nuclear DNA (18). In comparison to human mtDNA, which typically constitutes only about 1% of total cellular DNA and consists of a single, circular-mapping 16.5 kbp molecule encoding complete protein-coding genes (13 in human versus 16 in diplomonids), mtDNA in diplomonids is unprecedented in its magnitude, while its gene expression mode adds a supplementary layer of complexity. Most of the diplomonid mtDNA is non-coding, with the extensive constant regions apparently carrying only the origin of replication and transcription initiation signals. The baroque organization of the mitochondrial genome and transcriptome in diplomonids is partially met by the well-studied case of sister trypanosomatids, which tells us that sustaining this extravagancy must require an enormously complex cellular machinery. Moreover, observing such unusual features in the free-living diplomonids challenges the common view that extreme oddities are synonymous with a parasitic lifestyle.

The enigmatic $\gamma 4$ gene

All diplomonids analyzed in this and previous studies (27) encode the same set of mitochondrial genes (Figure 6). The only gene with patchy distribution, encountered in half of the species, is $\gamma 4$ encoding a hypothetical protein, which is poorly conserved at the sequence level and for which no homolog was found outside diplomonids. The Y4 protein of *D. papillatum* was recently detected by mass spectrometry in a respirasome supercomplex and, therefore, might represent a novel diplomonid-specific subunit of one of the respiratory chain complexes (36). Alternatively, Y4 might specify a

highly derived mitochondrion-encoded mitoribosomal protein. For example, the kinetoplastid *RPS12* (encoding the mitoribosomal uS12m) and *MURF5* are renowned for their extreme sequence divergence, and the latter has been uncovered as *RPS3* (uS3m) only by structure determination of the *Trypanosoma brucei* mitoribosome (50). No homologs encoding uS12m and uS3m were detected in the *D. papillatum* nuclear genome (our unpublished observations), which suggests that Y4 may be an extremely divergent mitoribosomal protein. Structure determination of the *Diplonema* respirasome and mitoribosome will be the ultimate test of these hypotheses.

Mitochondrial gene fragmentation at new heights in hemistasiids

An earlier study of four genes indicated high fragmentation in *Hemistasia* mtDNA (29). Our more systematic investigation of complete mitochondrial transcriptomes from two other hemistasiids generalized this finding, uncovering putative mini-modules as short as 3 bp. The vast majority of mini-modules is embedded in another module, which indicates that increasing gene fragmentation facilitates double use of coding sequence for distinct genes. Importantly, reuse can be even multiple: in *Namystynia*, *cox1*-m15 and *cox1*-m6 are both embedded in *cox1*-m13, while in *Artemidia*, *nad2*-m1 and *nad5*-m20 extensively overlap with *nad5*-m13, with a 9-bp region contributing to all three gene pieces (Figure 5 and Supplementary Table S6).

Ultra-short (1–30 nt) coding sequences are found in nuclear and mitochondrial genomes of numerous organisms (51–53). These micro-exons are joined to their neighbors—by the spliceosome or the Group I or Group II splicing machineries—via *cis*-splicing, thus relying on a physical connection between exons for proper joining. The hemistasiid case suggests that mini-module joining might proceed through an intermediate where a larger precursor acts as a ‘carrier’ of the mini-module (Figure 4C and D), ensuring the correct *trans*-splicing of a sequence that alone is presumably too short to ensure specificity. The actual mechanism of module transcript match-making still remains an intriguing puzzle.

RNA editing at an unprecedented level

Among diplonemids, the largest number of mitochondrial editing sites was counted in *Namystynia*, notably over 1,000 A-to-I and C-to-U substitutions, 14 G-to-A changes, and 94 U+A-tracts that sum up to >600 nt added to modules (Supplementary Table S5). As in previously studied diplonemids (16), RNA editing had an overall restorative effect on coding sequences, allowing production of functional proteins from *a priori* defective gene pieces.

The myxomycete *Physarum polycephalum*, several dinoflagellates, and the lycophytes *Isoetes engelmannii* and *Selaginella uncinata*, are renowned for extensive organellar editing with 1,333 sites in *P. polycephalum* mitochondria, 1,782 in *I. engelmannii* mitochondria and 3,415 in *S. uncinata* plastids (47,54–56). When comparing the number of edits per number of residues in a given transcriptome, the editing is most pervasive in *Isoetes* (6.7%), several

dinoflagellates (5.4–6.5%), followed by *Selaginella* (4.3%) and *Physarum* (3.5%). Diplonemidae rank lower (1.9–2.5%; Supplementary Table S7), yet Hemistasiidae surpass all previous records. With an editing density of 12.2%, *Namystynia* has the most extensively edited transcriptome documented so far (Supplementary Table S7).

Physarum polycephalum is also one of the few species known to employ more than one mode of mitochondrial RNA editing: co-transcriptional nucleotide insertions and occasional deletions (57), and post-transcriptional C-to-U substitutions (47,58), while diplonemids feature substitution (C-to-U, A-to-I and G-to-A) as well as U- and A-appendage editing.

New types of RNA editing

In hemistasiids, we detected two types of editing novel for diplonemid mitochondria, which involve G-to-A substitutions and A-appendage to internal modules. G-to-A editing has been only rarely reported in mitochondria, e.g. in dinoflagellates such as *Hematodinium* (59), while such events are extremely uncommon in the nucleus (60,61). Attesting to the importance of this type of editing in hemistasiids, the G-to-A substitution site in *nad5*-m6 of *Artemidia* is also conserved in *Namystynia*. The editing event contributes to the replacement of a Ser by an Asp codon that corresponds to the function-critical residue at position 179 in mammalian Nad5, an amino acid involved in the proton relay of complex I (62).

The molecular mechanism of G-to-A editing remains a matter of speculation; while C and U interconversion can proceed by transamination (U-to-C) and deamination (C-to-U)—since the two bases differ only in the absence or presence of an amino group—G and A differ in two groups, and no single chemical reaction is known to interconvert these two bases.

More concrete notions exist about A-appendage editing, which is a crucial step in the maturation of the dinoflagellate *cox3* transcript (63,64). The reaction is presumably catalyzed by the poly(A) polymerase that otherwise adds poly-A tails to mitochondrial transcripts. The corresponding enzymes have been characterized in mammals and trypanosomes (65). In the latter, the 3' tails are actually a mix of A+U residues, generated in two steps. Prior to editing, which in trypanosomes involves U-insertions and U-deletions (66–68), a 20–25 residue-long 3' A-tail is added. Once editing is completed, this tail is elongated to a 200–300 nt-long A+U heteropolymer, earmarking the transcript for translation and allowing its association with the mitochondrial ribosome (69). Both short and long tails are synthesized by the kinetoplast poly(A) polymerase 1 (70), which forms a complex with two pentatricopeptide proteins called kinetoplast polyadenylation/uridylation factors (KPAFs) 1 and 2 (71). RNA-editing terminal uridylyl transferase 1, which forms a complex with 3' exonuclease (72), is involved in the formation of long mRNA 3' tails (73,74), and possibly also in uridylation of rRNAs and gRNAs.

We expect a similar protein complex to operate in diplonemid mitochondria. However, in diplonemids, the addition of As—frequently together with Us—takes place in two fundamentally different contexts: not only at the

3' end of terminal modules, thus generating mRNA tails, but also of internal modules, as we documented here for *Namystynia* (Figures 2 and 7A; Supplementary Table S5). It will be interesting to examine in diplonemids whether a single protein complex is responsible for generating both A-tails and A-appendages or whether distinct specialized complexes have evolved for the two purposes.

Mitochondrial genes of *Sulcionema*—primitively simple or reduced upon divergence?

The mitochondrial system of *Sulcionema* appears in several aspects less complex than that of the other diplonemids. Furthermore, this species has the shortest branch in both mitochondrial (Figure 6A and Supplementary Figure 4A) and nuclear phylogenies (11). Some of these features might have been reduced upon divergence, while others might be primitive.

Of particular interest is the absence of the post-transcriptionally added Us between modules m4 and m5 of *Sulcionema cox1* (Supplementary Figure S5). Besides being the first editing site identified in a diplonemid (23), this U-appendage had apparently an important evolutionary impact on the Cox1 protein structure. In all species except diplonemids, the protein region corresponding to junction m4/m5 (loop 1) is positively charged and, in the folded protein, interacts with a downstream loop 2 composed of small and hydrophobic residues. The inverse situation applies to all diplonemids that feature *cox1* U-appendage editing. Here, the U-tract added at the m4/m5 junction and its environs specify a hydrophobic patch, whereas the downstream loop contains a polar Arg residue (28). Interestingly, the *Sulcionema* Cox1 protein has exactly the same hydrophobicity pattern in loops 1 and 2 as the other diplonemids, only the Us at the m4/m5 junction are genome-encoded as part of m4. More extensive taxon sampling will be necessary to untangle the order of the two evolutionary events, loop 1/loop 2-polarity switching and U-appendage editing.

Other less complex features in *Sulcionema* include the lack of nested modules (Supplementary Table S6) and the low RNA editing frequency (Supplementary Table S7) compared to the other diplonemids. Furthermore, the range of copy numbers across its module-bearing chromosomes is only ~13 (~30 when including its 16 module-less chromosomes), but 50–150 in the other Diplonemidae, and even ~600 in the hemistasiid *Artemidia* (up to ~300 of its B class chromosomes alone) (Supplementary Table S4) (27). It was noted that unequal copy numbers of chromosomes in multipartite genomes may cause chromosome loss during random mtDNA segregation to daughter cells (75). One solution to this problem is to over-amplify mtDNA, which in *D. papillatum* represents >50% of total cellular DNA (18,75). It would thus be interesting to see whether the more even chromosome copy number distribution in *Sulcionema* correlates with a lower mtDNA to nuclear DNA ratio.

Gene fragmentation and complexity in the most diverse diplonemids

Given the very high estimate of diplonemid species in the ocean (5,7), it can be safely predicted that species with even

more complex mitochondrial genomes and transcriptomes will eventually be discovered, especially among hemistasiids and eupelagonemids. For example, when reanalyzing the published genome sequences from 10 single-cell eupelagonemids (4), we detected in the data from 'cell 13' three potential mitochondrial modules encoding highly conserved regions of *cox1* and *nad7*. One candidate module corresponds exactly to the hemistasiid *nad7*-m2, the second to the upstream half of the hemistasiid *nad7*-m3, and the third is a homolog of *cox1*-m11 from *Hemistasia* (Supplementary Figure S5). This indicates that the mitochondrial genomes of Eupelagonemidae species are similar to those of the Hemistasiidae clade with respect to module sizes and RNA editing (Figure 2 and Table 1; Supplementary Table S5).

CONCLUSIONS AND OUTLOOK

Since diplonemids are a highly successful group as to their geographic distribution and habitat diversity, their extravagantly complex mitochondrial system has apparently little if any impact on their fitness. We see in it an excellent example of constructive neutral evolution (76–79), which postulates that a stepwise increase in the complexity of a given cellular machinery can occur with no associated selective consequence. Regardless of whether the flexible gene module structure allows sequence reuse for unrelated genes and/or *de novo* generation of 'improved' gene pieces, the variety of splicing and editing events makes the mitochondrial genome of diplonemids a laboratory for new inventions. Indeed, in a diplonemid cell, it takes 'the whole village' to decode the handful of mitochondrial genes. Our next challenge is to identify the individual players involved in decoding and to establish how the complex gene expression is coordinated.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Jan Votýpka (Charles University, Prague) and Akinori Yabuki (JAMSTEC, Yokohama) for help with the isolation of strains.

Author contributions: M.V., G.B., J.L.: conceptualization; K.Z., M.V.: methodology; K.Z.: software; K.Z., M.V., B.K.: data curation; K.Z., M.V.: formal analysis; B.K., K.Z., M.V.: investigation; K.Z., M.V., B.K.: visualization; B.K., K.Z., D.F., M.V.: writing—original draft; M.V., G.B., J.L.: writing—review and editing; G.P., G.B.: resources; J.L., G.B., M.V., D.F.: supervision; J.L.: project administration; J.L., G.B., B.K.: funding acquisition.

FUNDING

ERC CZ grant [LL1601 to J.L.]; Czech Ministry of Education (ERD Funds) [OPVVV16_019/ 0000759 to J.L.]; Gordon and Betty Moore Foundation [GBMF-4983.01 to J.L., G.B.]; Natural Sciences and Engineering Research Council of Canada [RGPIN-2014-05286, RGPIN-2019-04024 to

G.B.]; Grant Agency of the University of South Bohemia [094/2018/P to B.K.]. Funding for open access charge: Gordon and Betty Moore Foundation.

Conflict of interest statement. None declared.

REFERENCES

- Burki,F., Roger,A.J., Brown,M.W. and Simpson,A.G.B. (2019) The new tree of eukaryotes. *Trends Ecol. Evol.*, **35**, 43–55.
- Adl,S.M., Bass,D., Lane,C.E., Lukeš,J., Schoch,C.L., Smirnov,A., Agatha,S., Berney,C., Brown,M.W., Burki,F. *et al.* (2019) Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.*, **66**, 4–119.
- de Vargas,C., Audic,S., Henry,N., Decelle,J., Mahe,F., Logares,R., Lara,E., Berney,C., Le Bescot,N., Probert,I. *et al.* (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science*, **348**, 1261605–1261605.
- Gawryluk,R.M.R., del Campo,J., Okamoto,N., Strassert,J.F.H., Lukeš,J., Richards,T.A., Worden,A.Z., Santoro,A.E. and Keeling,P.J. (2016) Morphological identification and single-cell genomics of marine diplomonads. *Curr. Biol.*, **26**, 3053–3059.
- Flegontova,O., Flegontov,P., Malviya,S., Audic,S., Wincker,P., de Vargas,C., Bowler,C., Lukeš,J. and Horák,A. (2016) Extreme diversity of diplomonid eukaryotes in the ocean. *Curr. Biol.*, **26**, 3060–3065.
- David,V. and Archibald,J.M. (2016) Evolution: Plumbing the depths of diplomonid diversity. *Curr. Biol.*, **26**, R1290–R1292.
- Flegontova,O., Flegontov,P., Malviya,S., Poullain,J., de Vargas,C., Bowler,C., Lukeš,J. and Horák,A. (2018) Neobodonids are dominant kinetoplastids in the global ocean. *Environ. Microbiol.*, **20**, 878–889.
- Sibbald,S.J. and Archibald,J.M. (2017) More protist genomes needed. *Nat. Ecol. Evol.*, **1**, 0145.
- Keeling,P.J. and Campo,J. del (2017) Marine protists are not just big bacteria. *Curr. Biol.*, **27**, R541–R549.
- Lukeš,J., Flegontova,O. and Horák,A. (2015) Diplomonads. *Curr. Biol.*, **25**, R702–R704.
- Tashyreva,D., Prokopchuk,G., Yabuki,A., Kaur,B., Faktorová,D., Votýpka,J., Kusaka,C., Fujikura,K., Shiratori,T., Ishida,K.-I. *et al.* (2018) Phylogeny and morphology of new diplomonids from Japan. *Protist*, **169**, 158–179.
- Tashyreva,D., Prokopchuk,G., Votýpka,J., Yabuki,A., Horák,A. and Lukeš,J. (2018) Life cycle, ultrastructure, and phylogeny of new diplomonids and their endosymbiotic bacteria. *mBio*, **9**, e02447-17.
- Prokopchuk,G., Tashyreva,D., Yabuki,A., Horák,A., Masařová,P. and Lukeš,J. (2019) Morphological, ultrastructural, motility and evolutionary characterization of two new hemistasiidae species. *Protist*, **170**, 259–282.
- Okamoto,N., Gawryluk,R.M.R., Campo,J., Strassert,J.F.H., Lukeš,J., Richards,T.A., Worden,A.Z., Santoro,A.E. and Keeling,P.J. (2019) A revised taxonomy of diplomonids including the Eupelagonemidae n. fam. and a type species, *Eupelagonema oceanica* n. gen. & sp. *J. Eukaryot. Microbiol.*, **66**, 519–524.
- Valach,M., Moreira,S., Faktorová,D., Lukeš,J. and Burger,G. (2016) Post-transcriptional mending of gene sequences: Looking under the hood of mitochondrial gene expression in diplomonids. *RNA Biol.*, **13**, 1204–1211.
- Moreira,S., Valach,M., Aoulad-Aissa,M., Otto,C. and Burger,G. (2016) Novel modes of RNA editing in mitochondria. *Nucleic Acids Res.*, **44**, 4907–4919.
- Faktorová,D., Valach,M., Kaur,B., Burger,G. and Lukeš,J. (2018) Mitochondrial RNA editing and processing in diplomonid protists. In: Cruz-Reyes,J. and Gray,M (eds.), *RNA Metabolism in Mitochondria. Nucleic Acids and Molecular Biology*. Springer, Cham, **34**, 145–176.
- Lukeš,J., Wheeler,R., Jirsová,D., David,V. and Archibald,J.M. (2018) Massive mitochondrial DNA content in diplomonid and kinetoplastid protists. *IUBMB Life*, **70**, 1267–1274.
- Kiethega,G.N., Yan,Y., Turcotte,M. and Burger,G. (2013) RNA-level unscrambling of fragmented genes in *Diplonema* mitochondria. *RNA Biol.*, **10**, 301–313.
- Sturm,N.R., Maslov,D.A., Grisard,E.C. and Campbell,D.A. (2001) *Diplonema* spp. possess spliced leader RNA genes similar to the kinetoplastida. *J. Eukaryot. Microbiol.*, **48**, 325–331.
- Lasda,E.L. and Blumenthal,T. (2011) *Trans-splicing*. Wiley *Interdiscip. Rev. RNA*, **2**, 417–434.
- Glanz,S. and Kück,U. (2009) Trans-splicing of organelle introns—a detour to continuous RNAs. *BioEssays*, **31**, 921–934.
- Marande,W., Lukeš,J. and Burger,G. (2005) Unique mitochondrial genome structure in diplomonids, the sister group of kinetoplastids. *Eukaryot. Cell*, **4**, 1137–1146.
- Marande,W. and Burger,G. (2007) Mitochondrial DNA as a genomic jigsaw puzzle. *Science*, **318**, 415.
- Vleck,C., Marande,W., Teijeiro,S., Lukeš,J. and Burger,G. (2011) Systematically fragmented genes in a multipartite mitochondrial genome. *Nucleic Acids Res.*, **39**, 979–988.
- Valach,M., Moreira,S., Kiethega,G.N. and Burger,G. (2014) Trans-splicing and RNA editing of LSU rRNA in *Diplonema* mitochondria. *Nucleic Acids Res.*, **42**, 2660–2672.
- Valach,M., Moreira,S., Hoffmann,S., Stadler,P.F. and Burger,G. (2017) Keeping it complicated: mitochondrial genome plasticity across diplomonids. *Sci. Rep.*, **7**, 14166.
- Kiethega,G.N., Turcotte,M. and Burger,G. (2011) Evolutionarily conserved *cox1* trans-splicing without *cis*-motifs. *Mol. Biol. Evol.*, **28**, 2425–2428.
- Yabuki,A., Tanifuji,G., Kusaka,C., Takishita,K. and Fujikura,K. (2016) Hyper-eccentric structural genes in the mitochondrial genome of the algal parasite *Hemistasia phaeocysticola*. *Genome Biol. Evol.*, **8**, 2870–2878.
- Rodríguez-Ezpeleta,N., Teijeiro,S., Forget,L., Burger,G. and Lang,B.F. (2009) Construction of cDNA libraries: Focus on protists and fungi. In: Parkinson,J (ed.), *Expressed Sequence Tags (ESTs). Methods in Molecular Biology (Methods and Protocols)*. Humana Press, Totowa, NJ, Vol. **533**, pp. 33–47.
- Bankevich,A., Nurk,S., Antipov,D., Gurevich,A.A., Dvorkin,M., Kulikov,A.S., Lesin,V.M., Nikolenko,S.I., Pham,S., Pribelski,A.D. *et al.* (2012) SPADes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.*, **19**, 455–477.
- Haas,B.J., Papanicolaou,A., Yassour,M., Grabherr,M., Blood,P.D., Bowden,J., Couger,M.B., Eccles,D., Li,B., Lieber,M. *et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.*, **8**, 1494–1512.
- Altschul,S. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Kearse,M., Moir,R., Wilson,A., Stones-Havas,S., Cheung,M., Sturrock,S., Buxton,S., Cooper,A., Markowitz,S., Duran,C. *et al.* (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.
- Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
- Valach,M., Léveillé-Kunst,A., Gray,M.W. and Burger,G. (2018) Respiratory chain complex I of unparalleled divergence in diplomonids. *J. Biol. Chem.*, **293**, 16043–16056.
- Eddy,S.R. (2009) A new generation of homology search tools based on probabilistic inference. *Genome Inform.*, **23**, 205–211.
- Katoh,K. and Standley,D.M. (2013) MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
- Capella-Gutiérrez,S., Silla-Martínez,J.M. and Gabaldón,T. (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
- Lartillot,N., Lepage,T. and Blanquart,S. (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*, **25**, 2286–2288.
- Ronquist,F., Teslenko,M., van der Mark,P., Ayres,D.L., Darling,A., Höhna,S., Larget,B., Liu,L., Suchard,M.A. and Huelsenbeck,J.P. (2012) MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.*, **61**, 539–542.
- Nguyen,L.-T., Schmidt,H.A., von Haeseler,A. and Minh,B.Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*, **32**, 268–274.
- Stamakis,A. (2014) RAXML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

44. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. and Jermin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods*, **14**, 587–589.
45. Minh, B.Q., Hahn, M. and Lanfear, R. (2018) New methods to calculate concordance factors for phylogenomic datasets. bioRxiv doi: <https://doi.org/10.1101/487801>, 05 December 2018, preprint: not peer reviewed.
46. Yang, J., Harding, T., Kamikawa, R., Simpson, A.G.B. and Roger, A.J. (2017) Mitochondrial genome evolution and a novel RNA editing system in deep-branching heteroloboseids. *Genome Biol. Evol.*, **9**, 1161–1174.
47. Bundschuh, R., Altmüller, J., Becker, C., Nürberg, P. and Gott, J.M. (2011) Complete characterization of the edited transcriptome of the mitochondrion of *Physarum polycephalum* using deep sequencing of RNA. *Nucleic Acids Res.*, **39**, 6044–6055.
48. Lara, E., Moreira, D., Vereshchaka, A. and López-García, P. (2009) Pan-oceanic distribution of new highly diverse clades of deep-sea diplomonads. *Environ. Microbiol.*, **11**, 47–55.
49. Kaur, B., Valach, M., Peña-Díaz, P., Moreira, S., Keeling, P.J., Burger, G., Lukeš, J. and Faktorová, D. (2018) Transformation of *Diplonema papillatum*, the type species of the highly diverse and abundant marine microeukaryotes Diplomonada (Euglenozoa). *Environ. Microbiol.*, **20**, 1030–1040.
50. Ramrath, D.J.F.F., Niemann, M., Leibundgut, M., Bieri, P., Prange, C., Horn, E.K., Leitner, A., Boehringer, D., Schneider, A. and Ban, N. (2018) Evolutionary shift toward protein-based architecture in trypanosomal mitochondrial ribosomes. *Science*, **362**, eaau7735.
51. Weyn-Vanhenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, X., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q. *et al.* (2014) HTS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep.*, **6**, 1139–1152.
52. Osigus, H.-J., Eitel, M. and Schierwater, B. (2017) Deep RNA sequencing reveals the smallest known mitochondrial micro exon in animals: The placozoan *cox1* single base pair exon. *PLoS One*, **12**, e0177959.
53. Ustianenko, D., Weyn-Vanhenryck, S.M. and Zhang, C. (2017) Microexons: discovery, regulation, and function. *Wiley Interdiscip. Rev. RNA*, **8**, e1418.
54. Grewe, F., Herres, S., Viehöver, P., Polsakiewicz, M., Weisshaar, B. and Knoop, V. (2011) A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.*, **39**, 2890–2902.
55. Oldenkott, B., Yamaguchi, K., Tsuji-Tsukinoki, S., Knie, N. and Knoop, V. (2014) Chloroplast RNA editing going extreme: more than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte *Selaginella uncinata*. *RNA*, **20**, 1499–1506.
56. Klinger, C.M., Paoli, L., Newby, R.J., Wang, M.Y.W., Carroll, H.D., Leblond, J.D., Howe, C.J., Dacks, J.B., Bowler, C., Cahoon, A.B. *et al.* (2018) Plastid transcript editing across dinoflagellate lineages shows lineage-specific application but conserved trends. *Genome Biol. Evol.*, **10**, 1019–1038.
57. Gott, J.M. (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.
58. Schallenberg-Rüdinger, M., Lenz, H., Polsakiewicz, M., Gott, J.M. and Knoop, V. (2013) A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. *RNA Biol.*, **10**, 1549–1556.
59. Jackson, C.J., Gornik, S.G. and Waller, R.F. (2012) The mitochondrial genome and transcriptome of the basal dinoflagellate *Hematodinium* sp.: Character evolution within the highly derived mitochondrial genomes of dinoflagellates. *Genome Biol. Evol.*, **4**, 59–72.
60. Wang, I.X., Grunseich, C., Chung, Y.G., Kwak, H., Ramratan, G., Zhu, Z. and Cheung, V.G. (2016) RNA-DNA sequence differences in *Saccharomyces cerevisiae*. *Genome Res.*, **26**, 1544–1554.
61. Daneck, P., Nellaker, C., McIntyre, R.E., Buendia-Buendia, J.E., Bumpstead, S., Ponting, C.P., Flint, J., Durbin, R., Keane, T.M. and Adams, D.J. (2012) High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biol.*, **13**, R26.
62. Agip, A.-N.A., Blaza, J.N., Bridges, H.R., Viscomi, C., Rawson, S., Muench, S.P. and Hirst, J. (2018) Cryo-EM structures of complex I from mouse heart mitochondria in two biochemically defined states. *Nat. Struct. Mol. Biol.*, **25**, 548–556.
63. Jackson, C.J., Norman, J.E., Schnare, M.N., Gray, M.W., Keeling, P.J. and Waller, R.F. (2007) Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. *BMC Biol.*, **5**, 41.
64. Burger, G., Jackson, C.J. and Waller, R.F. (2012) Unusual mitochondrial genomes and genes. In: Bulterwell, C. (ed.), *Organelle Genetics*. Springer, Berlin, Heidelberg, pp. 41–77.
65. Chang, J.H. and Tong, L. (2012) Mitochondrial poly(A) polymerase and polyadenylation. *Biochim. Biophys. Acta*, **1819**, 992–997.
66. Zimmer, S.L., Simpson, R.M. and Read, L.K. (2018) High throughput sequencing revolution reveals conserved fundamentals of U-indel editing. *Wiley Interdiscip. Rev. RNA*, **9**, e1487.
67. Cruz-Reyes, J., Moers, B.H.M., Doharey, P.K., Meehan, J. and Gulati, S. (2018) Dynamic RNA holo-editosomes with subcomplex variants: Insights into the control of trypanosome editing. *Wiley Interdiscip. Rev. RNA*, **9**, e1502.
68. Alfonzo, J., Thiemann, O. and Simpson, L. (1997) The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria. *Nucleic Acids Res.*, **25**, 3751–3759.
69. Aphasizhev, R. and Aphasizheva, I. (2011) Uridine insertion/deletion editing in trypanosomes: a playground for RNA-guided information transfer. *Wiley Interdiscip. Rev. RNA*, **2**, 669–685.
70. Etheridge, R.D., Aphasizheva, I., Gershon, P.D. and Aphasizhev, R. (2008) 3' adenylation determines mRNA abundance and monitors completion of RNA editing in *T. brucei* mitochondria. *EMBO J.*, **27**, 1596–1608.
71. Aphasizheva, I., Maslov, D., Wang, X., Huang, L. and Aphasizhev, R. (2011) Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes. *Mol. Cell*, **42**, 106–117.
72. Zhang, L., Sement, F.M., Suematsu, T., Yu, T., Monti, S., Huang, L., Aphasizhev, R. and Aphasizheva, I. (2017) PPR polyadenylation factor defines mitochondrial mRNA identity and stability in trypanosomes. *EMBO J.*, **36**, 2435–2454.
73. Ryan, C.M. and Read, L.K. (2005) UTP-dependent turnover of *Trypanosoma brucei* mitochondrial mRNA requires UTP polymerization and involves the RET1 TUTase. *RNA*, **11**, 763–773.
74. Aphasizheva, I. and Aphasizhev, R. (2010) RET1-catalyzed uridylylation shapes the mitochondrial transcriptome in *Trypanosoma brucei*. *Mol. Cell Biol.*, **30**, 1555–1567.
75. Burger, G. and Valach, M. (2018) Perfection of eccentricity: Mitochondrial genomes of diplomonads. *IUBMB Life*, **70**, 1197–1206.
76. Gray, M.W., Lukeš, J., Archibald, J.M., Keeling, P.J. and Doolittle, W.F. (2010) Irremediable complexity? *Science*, **330**, 920–921.
77. Flegontov, P., Gray, M.W., Burger, G. and Lukeš, J. (2011) Gene fragmentation: a key to mitochondrial genome evolution in Euglenozoa? *Curr. Genet.*, **57**, 225–232.
78. Lukeš, J., Archibald, J.M., Keeling, P.J., Doolittle, W.F. and Gray, M.W. (2011) How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life*, **63**, 528–537.
79. Stoltzfus, A. (2012) Constructive neutral evolution: Exploring evolutionary theory's curious disconnect. *Biol. Direct*, **7**, 35.

Chapter 5

Genetic tool development in marine protists: Emerging model organisms for experimental cell biology

The following passage of this scope "**Genetic tool development in marine protists: Emerging model organisms for experimental cell biology**" contains classified information which is available only in the archived original of the graduation thesis deposited at the relevant USB Faculty

Chapter 6

Targeted integration by homologous recombination enables *in-situ* tagging and replacement of genes in the marine microeukaryote *Diplonema papillatum*

The following passage of this scope "**Targeted integration by homologous recombination enables *in-situ* tagging and replacement of genes in the marine microeukaryote *Diplonema papillatum***" contains classified information which is available only in the archived original of the graduation thesis deposited at the relevant USB Faculty.

13. Curriculum vitae

BINNYPREET KAUR

Date of birth: 1st March 1988
Nationality: Indian
Home address: Branišovská 40, České Budějovice 370 05,
Czech Republic
Telephone number: +420 702845554
E-mail: binny@paru.cas.cz
Work address: Biology Centre of the Czech Academy of
Sciences, Institute of Parasitology & University of
South Bohemia, Faculty of Science, Branišovská
31, České Budějovice 370 05, Czech Republic

EDUCATION

2015 – present: **Ph.D. student** of Molecular and Cell Biology and
Genetics, Department of Molecular Biology and
Genetics, Faculty of Science, University of South
Bohemia, České Budějovice, Czech Republic
Institute of Parasitology, Czech Academy of
Sciences, Czech Republic
Thesis Title: Genetic modifications and
functional analyses of diplomemids

2012- 2014: **Senior Research Fellow** at National Dairy
Research Institute, Karnal, Haryana, India

Project Title: Elucidating the physiological and genomic regulation process of follicular development, oocytes maturation and embryogenesis in buffalo”

2008-2010: **M.Sc. Honors in Biotechnology** from Lovely Professional University, Punjab, India
Thesis: Extraction, purification and identification of immunomodulatory components of *Cannabis sativa*.

PUBLICATIONS

1. **Kaur B.**, Záhonová K., Valach M., Faktorová D., Prokopchuk G., Burger G., Lukeš J. (2020) Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplonemids. *Nucleic Acids Res.* 10, doi:10.1093/nar/gkz1215.
2. Faktorová D., Nisbet R.E.R., Fernández Robledo J.A., Casacuberta E., Sudek L., Allen A.E., Ares M. Jr., Aresté C., Balestreri C., Barbrook A.C., Beardslee P., Bender S., Booth D.S., Bouget F.-Y., Bowler C., Breglia S.A., Brownlee C., Burger G., Cerutti H., Cesaroni R., Chiurillo M.A., Clemente T., Coles D.B., Collier J.L., Cooney E.C., Coyne K., Docampo R., Dupont C.L., Edgcomb V., Einarsson E., Elustondo P.A., Federici F., Freire-Beneitez V., Freuria N.J., Fukuda K., García P.A., Girguis P.R., Goma F., Gornik S.G., Guo J., Hampl V., Hanawa Y.,

Haro-Contreras E.R., Hehenberger E., Highfield A., Hirakawa Y., Hopes A., Howe C.J., Hu I., Ibañez J., Irwin N.A.T., Ishii Y., Janowicz N.E., Jones A.C., Kachale A., Fujimura-Kamada K., **Kaur B.**, Kaye J.Z., Kazana E., Keeling P.J., King N., Klobutcher L.A., Lander N., Lassadi I., Li Z., Lin S., Lozano J.-C., Luan F., Maruyama S., Matute T., Miceli C., Minagawa J., Moosburner M., Najle S.R., Nanjappa M., Nimmo I.C., Noble L., Novák Vanclová A.M.G., Nowacki M., Nuñez I., Pain A., Piersanti A., Pucciarelli S., Pyrih J., Rest J.S., Rius M., Robertson D., Ruaud A., Ruiz-Trillo I., Sigg M.A., Silver P.A., Slamovits C.H., Smith G.J., Sprecher B.N., Stern R., Swart E., Tsaousis A., Tsypin L., Turkewitz A., Turnšek J., Valach M., Vergé V., von Dassow P., von der Haar T., Waller R.F., Wang L., Wen X., Wheeler G., Woods A., Zhang H., Mock T., Worden A.Z. & Lukeš J. (2020) Genetic tool development in marine protists: Emerging model organisms for experimental cell biology. (*Nature Methods*-resubmission).

bioRxiv preprint first posted online Aug. 1, 2019; doi: <http://dx.doi.org/10.1101/718239>.

3. **Kaur B.**, Valach M., Pena Diaz C., Moreira S., Keeling P., Burger G., Lukeš J., Faktorová D. (2018) Transformation of *Diplonema papillatum*, the type species of the highly diverse and abundant marine micro-eukaryotes Diplonemida (Euglenozoa), *Environmental Microbiology* 20: 1030-1040.
4. Faktorová D., Valach M., **Kaur B.**, Burger G., Lukeš J. (2018) Mitochondrial RNA editing and processing in diplomemid

protists. *RNA Metabolism in Mitochondria. Nucleic Acids and Molecular Biology*; 34: 145–176, Cham Springer

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

CONFERENCES/SYMPOSIUM/WORKSHOPS

Talk: **Kaur B.**, Faktorová D., Valach M., Burger G., Lukeš J. (2019). *Diplonema papillatum*, a representative of the highly diverse and abundant marine microeukaryotes, can be genetically manipulated. VIII European Congress of Protistology- ISOP Joint meeting. Rome, Italy, 28 July- 2 August 2019

Talk: **Kaur B.**, Faktorová D., Valach M., Záhonová K., Burger G., Lukeš J. (2019). DNA fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplomemids. Diplonema Workshop, Institute of Parasitology, CAS, Biology Centre. Czech Republic 17 July 2019

Talk: **Kaur B.**, Faktorová D., Valach M., Burger G., Lukeš J. (2018). The mitochondrial genome

structure and arrangement across diplomonads.
Ph.D. Retreat, The Department of Molecular
Biology and Genetics, Bělá-Nová Pec, Czech
Republic. 19th Oct to 21st Oct 2018

Talk: **Kaur B**, Faktorová D., Pena Diaz C., Burger G.,
Lukeš J. (2016). Genetic modifications and
functional analyses of Diplonemids, Genome
Engineering: The CRISPR-Cas Revolution
Conference, Cold Spring Harbor Laboratory, New
York, USA. 17 Aug-20 Aug 2016.

Talk: **Kaur B**, Faktorová D., Pena Diaz C., Burger G.,
Lukeš J. (2017). Genetic modifications and
functional analyses of Diplonemids. Ph.D. Retreat
2017. Department of Molecular Biology and
Genetics, Ceske Budejovice, Czech Republic.

Talk: **Kaur B**, Faktorová D., Pena Diaz C., Burger G.,
Lukeš J. (2016). Genetic modifications and
functional analyses of Diplonemids in 46th
Jirovec's Protozoological Days of Czech Society
for Parasitology, The Protozoological section of
the Czech Society for Parasitology, Rumburak,
Czech Republic 2016.

Poster **Kaur B**, Faktorová D., Pena Diaz C., Burger G.,
Lukeš J. (2016). Developing a genetic

modification system for marine protist *Diplonema papillatum* BSP meeting, České Budějovice, Czech Republic, September 4-7, 2016

PROFESSIONAL EXPERIENCE

Employment:

2015 – present: **Graduate student**, Institute of Parasitology,
Biology Centre of the Czech Academy of Sciences

2015 – present: **Researcher**, Institute of Parasitology,
Biology Centre of the Czech Academy of Sciences

2012- 2014: **Senior Research Fellow** at National Dairy
Research Institute, Karnal, Haryana, India

RESEARCH STAY

2017 Robert-Cedergren Center for Bioinformatics and Genomics, Department of Biochemistry and Molecular Medicine, Faculty of Medicine, **Université de Montréal**, Montréal, QC, Canada. (3 months stay); Advisor: **Prof. Gertraud Burger** and **Dr. Matus Valach**

Title: Investigation of the mitochondrial gene structure and RNA editing in several new diplomemid strains

GRANTS AND AWARDS

2015-2017	Grant Agency of USB of individual project Student grant
2017-2018	Grant Agency of USB of individual project Student grant
2017	Aesthetic image competition -Second Prize

TEACHING AND STUDENTS

2017-2018	Teaching assistant at Advances in Molecular Biology course (KMB-770/2018), Faculty of Science, University of South Bohemia, České Budějovice.
------------------	---

© for non-published parts Binnypreet Kaur
binny@paru.cas.cz

Genetic modifications and functional analyses of Diplonemids
Ph.D. Thesis Series, 2020, No. 3

All rights reserved
For non-commercial use only

Printed in the Czech Republic by Typodesign
Edition of 20 copies

University of South Bohemia in České Budějovice
Faculty of Science
Branišovská 1760
CZ-37005 České Budějovice, Czech Republic

Phone: +420 387 776 201
www.prf.jcu.cz, e-mail: sekret-fpr@prf.jcu.cz