environmental microbiology

Environmental Microbiology (2020) **22**(9), 3660–3670 doi:10.1111/1462-2920.15130

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

Drahomíra Faktorová,[1,2*†] **Binnypreet Kaur**,[1,2†] **Matus Valach**,[3†] **Lena Graf**,[2,4] **Corinna Benz**,[1] **Gertraud Burger**[3] **and Julius Lukeš**[1,2*]

[1]*Czech Academy of Sciences, Institute of Parasitology, Biology Centre, Czech Republic.*

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

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

[4]*Present address: Johannes Kepler University, Linz, Austria.*

Summary

Diplonemids are a group of highly diverse and abundant marine microeukaryotes that belong to the phylum Euglenozoa and form a sister clade to the well-studied, mostly parasitic kinetoplastids. Very little is known about the biology of diplonemids, as few species have been formally described and just one, *Diplonema papillatum*, has been studied to a decent extent at the molecular level. Following up on our previous results showing stable but random integration of delivered extraneous DNA, we demonstrate here homologous recombination in *D. papillatum*. Targeting various constructs to the intended position in the nuclear genome was successful when 5′ and 3′ homologous regions longer than 1 kbp were used, achieving N-terminal tagging with mCherry and gene replacement of α- and β-tubulins. For more convenient genetic manipulation, we designed a modular plasmid, pDP002, which bears a protein-A tag and used it to generate and express a C-terminally tagged mitoribosomal protein. Lastly, we developed an improved transformation protocol for broader applicability across laboratories. Our robust methodology allows the replacement, integration as well as endogenous tagging of *D. papillatum* genes, thus opening the door to functional studies in this species and establishing a basic toolkit for reverse genetics of diplonemids in general.

Received 9 May, 2020; revised 7 June, 2020; accepted 13 June, 2020. *For correspondence. E-mail dranov@paru.cas.cz. E-mail jula@paru.cas.cz, Tel: +420 387 775 433 Fax: +420 385 310 388.
†Equal contribution

Introduction

Diplonemids are heterotrophic protists belonging to Euglenozoa. They constitute the sister group to kinetoplastids, which include the well-studied pathogenic *Trypanosoma* and *Leishmania* species. The third major group of Euglenozoa that branches off basally to diplonemids and kinetoplastids are the free-living euglenids, which are important players in freshwater ecosystems (Adl *et al*., 2019). Rarely identified in the environment, diplonemids were considered a marginal and thus ecologically insignificant group of flagellates. Until recently, only three diplonemid genera were recognized (*Diplonema*, *Rhynchopus* and *Hemistasia*) with just a handful of formally described species (Simpson, 1997; Vickerman, 2000; von der Heyden *et al*., 2004; Roy *et al*., 2007; Massana, 2011). Initially, diplonemids were only known from brackish and marine habitats, frequently associated with sediments (López-García *et al*., 2007), and were largely overlooked by barcoding studies due to an atypical V4 region of the 18S rRNA gene that does not properly assemble. However, the more recent use of the V9 region of the 18 rRNA gene, which in diplonemids is more suitable for metagenomic approaches, allowed several comprehensive surveys of marine microbial diversity to recognize diplonemids as the most diverse and the fifth to sixth most abundant group of heterotrophic planktonic eukaryotes (de Vargas *et al*., 2015; Lukeš *et al*., 2015; Flegontova *et al*., 2016; Gawryluk *et al*., 2016). Although diplonemids had been encountered in large lakes (Yi *et al*., 2017; Mukherjee *et al*., 2019), they seem to be rare in freshwater habitats.

In phylogenetic analyses based on (parts of) the 18S rRNA gene, diplonemids split into four distinct lineages: (i) the so-called classic diplonemids hereafter referred to as Diplonemidae, consisting of the genera *Diplonema* and *Rhynchopus*, recently expanded by the genera

Lacrimia and *Sulcionema* (Tashyreva *et al*., 2018a, 2018b), (ii) a small planktonic clade containing the genus *Hemistasia* and two newly described genera *Artemidia* and *Namystynia* (Prokopchuk *et al*., 2019), (iii) a deep-sea pelagic diplonemid (DSPD) clade I, named Eupelagonemidae (Okamoto *et al*., 2019), and (iv) a DSPD clade II (Flegontova *et al*., 2016). In a revised taxonomy based on morphology and single-cell genomics, Eupelagonemidae emerged as by far the most abundant phylotype (Okamoto *et al*., 2019). Thus, although we know nothing about their ecological functions and biology, Eupelagonemidae must be a main planktonic component with an important, albeit undetermined role in the ecosystem of the world ocean (Flegontova *et al*., 2016). A more detailed study of this group is hampered by the fact that none of the species is available in culture.

The development of (high-throughput) genetic tools in as many planktonic protists as possible will be critical for tackling the functions of at least a small fraction of the over 100 million unique genes from across marine unicellulars (Carradec *et al*., 2018). In more than a dozen such lineages across the eukaryotic tree, many impervious to functional studies thus far, the expression of introduced genes was recently demonstrated (Faktorová *et al*., 2020). *Diplonema papillatum* – an easily cultured and comparably fast dividing diplonemid representative – has joined this suite of genetically tractable marine organisms. Being the only diplonemid for which both the nuclear (our unpubl. data) and mitochondrial genome sequences are available (Marande *et al*., 2005; Marande and Burger, 2007; Kiethega *et al*., 2013; Moreira *et al*., 2016; Faktorová *et al*., 2018), *D. papillatum* is the most suitable candidate to be established as a model organism. Indeed, diplonemids are attractive for molecular and cell biology studies due to the large number of exceptional features, such as extensive mitochondrial editing and *trans*-splicing (Valach *et al*., 2016; Kaur *et al*., 2020), mitochondrial DNA amount exceeding that of any other known organellar genome (Lukeš *et al*., 2018), uniquely remodelled and expanded respiratory chain complex I (Valach *et al*., 2018), as well as the presence of endosymbiotic bacteria with extremely reduced genomes (George *et al*., 2020).

Previously, several key steps necessary for the genetic manipulation of this flagellate were accomplished (Kaur *et al*., 2018). First, *D. papillatum* was shown to be sensitive to multiple selectable markers. Second, following electroporation, the extraneous DNA became not only stably integrated into the nuclear genome; but, the heterologous gene was indeed transcribed and translated. However, the major shortcoming of the method was the random integration of the extraneous DNA. That the gene was expressed a fortuitous consequence of the polycistronic nuclear transcription in diplonemids, a trait that these flagellates share with their sister group kinetoplastids (Clayton, 2016). In sum, until recently, our experiments fell short of targeted integration required for functional studies (Kaur *et al*., 2018; Faktorová *et al*., 2020).

At the outset of the study presented here, the *D. papillatum* nuclear genome was not yet completely assembled and annotated. Therefore, we selected tubulin genes, which have become traditional candidates for gene tagging and knock-outs in emerging model systems (Eichinger *et al*., 1999). Tubulins are among the major constituents of the eukaryotic cytoskeleton, which provides structural support and plays an important role in cell division, intracellular transport and DNA segregation (Jackson *et al*., 2006). In eukaryotes, the tubulin superfamily expanded into numerous groups, with α-, β- and γ-tubulins being omnipresent along with their specific regulatory arrangement. The α- and β-tubulin genes are usually organized in tandem arrays (McKean *et al*., 2001; Zhao *et al*., 2014).

Here, we show targeted integration of heterologous genes into the *D. papillatum* genome, facilitated by extended 5′ and 3′ homologous regions. We designed and successfully tested constructs for the replacement of the α- and β- tubulin genes using two different selection markers. Increasing the length of homologous regions was sufficient to achieve targeted integration of the previously published construct (Kaur *et al*., 2018) for the N-terminal tagging of α-tubulin with the fluorescent mCherry protein. As a means for systematic gene deletion, insertion and tagging, we next built a modular construct, which allows precise 5′ and 3′-tagging of the selected gene. Lastly, we elaborated an improved the transformation protocol, thus validating the devised methodologies for stable gene replacement and tagging in *D. papillatum* by homologous recombination as sufficiently robust for broad deployment.

Results

Homologous recombination and non-homologous end-joining pathways

In the previous study (Kaur *et al*., 2018), we failed to target the electroporated construct into the correct position in the *D. papillatum* nuclear genome, which is crucial for gene tagging and knock-outs. Therefore, we first assessed the type of DNA repair that may act in this protist and identified genes involved in the two principal repair mechanisms for example, Rad50–Rad52 and Rpa1–Rpa3 acting in the homologous recombination (HR) pathway (Krejci *et al*., 2012; Son and Hasty, 2019) and Lig4, Ku70 and Ku80 participating in the non-homologous end-joining (NHEJ) pathway (Waters *et al*., 2014). Hence, we wondered whether off-site

integration was due to a lower efficiency of HR relative to NHEJ, or due to the highly repetitive nature of the genome (repetitive sequences represent ~60% of the nuclear genome; our unpubl. data). Therefore, we experimentally tested two strategies to achieve targeted integration: (1) inhibiting the NHEJ pathway and (2) extending the length of regions homologous to the targeted site.

Inhibition of NHEJ pathway

It was technically not feasible to block the NHEJ pathway by knocking out the genes encoding the Ku70/80 proteins. Therefore, we attempted to inhibit the pathway using W7 (N-(6-aminohexyl)-5-chloro-1-naphthalenesulfonamide). W7 inhibits the production of Ku protein's cofactor inositol-hexakisphophate (InsP6) (Byrum *et al*., 2004) and was shown to significantly increase the rate of gene deletion in the pathogenic yeast *Cryptococcus neoformans* (Arras and Fraser, 2016).

The minimum inhibitory concentration of W7 for *D. papillatum* was 40 $\mu g\ ml^{-1}$, as inferred from the Alamar blue assay (Supporting Information Fig. S1A). To test transformation efficiency (for details, see Experimental procedures), we used the DF_Dp_01 construct, which contains *Puro*^{R} + *mCherry* cassette bearing 500 bp-long homologous arms and *Diplonema* UTRs (Kaur *et al*., 2018). We obtained five cell lines within 3–4 weeks under puromycin selection, yet none of the examined cell lines showed correct integration in the target site (Supporting Information Fig. S1B).

Increased length of homologous regions

In our second, alternative approach, we increased the length of the homologous regions of the construct by a fusion PCR method (Supporting Information Fig. S2), which previously proved successful for *Trypanosoma brucei* (Barnes and McCulloch, 2007). In total, five different constructs with 1–2 kb-long homology arms were electroporated. Integration into the genome was analysed by PCR and expression verified by Western blot analysis. In the case of N-terminal tagging of α-tubulin with mCherry, we also examined the transformants by fluorescent microscopy.

α- and β-tubulin replacement with puromycin resistance marker

Diplonema papillatum contains at least 30 α-, 27 β- and 3 γ-tubulin genes in the nuclear genome (our unpubl. data). Hence, we designed constructs for the replacement of α- and β-tubulin genes (constructs #1 and #2). We targeted well-expressed and intron-less alleles of α-tubulin [*DIPPA_23256*, same as in the previous study (Kaur *et al*., 2018)] and β-tubulin (*DIPPA_12526*), each bearing unique flanking regions. The goal was to replace the coding sequences with the puromycin resistance marker, that is, the puromycin-N-acetyltransferase gene (hereafter referred to as *Puro*^{R}), using native *D. papillatum* 5′ and 3′ UTRs (Fig. 1A and Supporting Information Fig. S3). We chose to target tubulins because these are multicopy genes and homologous integration of the *Puro*^{R} construct would therefore not be lethal. The *Puro*^{R} construct could be integrated elsewhere in the genome and would still likely be expressed due to the polycistronic transcription in diplonemids (Kaur *et al*., 2018; Faktorová *et al*., 2020). Therefore, the integration site was verified by PCR, which was possible because sequences flanking the chosen loci were unique.

Puro^R cell lines were recovered after cultivation in the presence of the selection drug for 10–14 days following electroporation, which was the time span required to ensure death of wild type (WT) cells. All the cell lines that survived puromycin selection were growing more slowly than WT cells and exhibited an unusual spherical morphology. To verify the integration of the construct into the target site, each cell line was propagated and PCRs were performed on genomic DNA together with the WT and the cell line A3 from our previous study (Kaur *et al*., 2018), used here as a positive control (Supporting Information Fig. S4A and B). In the case of the β-tubulin experiment, none of the six cell lines showed homologous integration (Supporting Information Fig. S4B); but α-tubulin replacement was successful in one (cell line 9) of three cell lines (Supporting Information Fig. S4A), which was verified by PCR with two primer pairs (Fig. 1B and Supporting Information Fig. S5A). Gene expression was further validated by spliced leader (SL) RT-PCR, showing that the corresponding mRNA is properly processed post-transcriptionally by the addition of the SL RNA to its 5′ end (Supporting Information Fig. S5B), and by Western blot analysis (Fig. 1C).

N-Terminal tagging of α-tubulin with mCherry under puromycin selection

These encouraging results led us to modify the DF_Dp_01 construct (Kaur *et al*., 2018) to allow N-terminal tagging of α-tubulin with mCherry under puromycin selection (construct #3 in Fig. 1A and Supporting Information Fig. S3). The DF_Dp_01 construct was modified so that the length of the 5′ and 3′ homologous arms was extended from 500 bp to 1784 and 1416 bp respectively. One of five *Puro*^{R} cell lines (cell line 15) yielded the correct PCR product (Supporting Information Fig. S4C). Hence, the extension of the homologous regions had a decisive impact on the proper integration of the construct into the target genomic locus. Cell line 15 (together with the controls) was also tested by two other primer pairs (Fig. 1B and Supporting

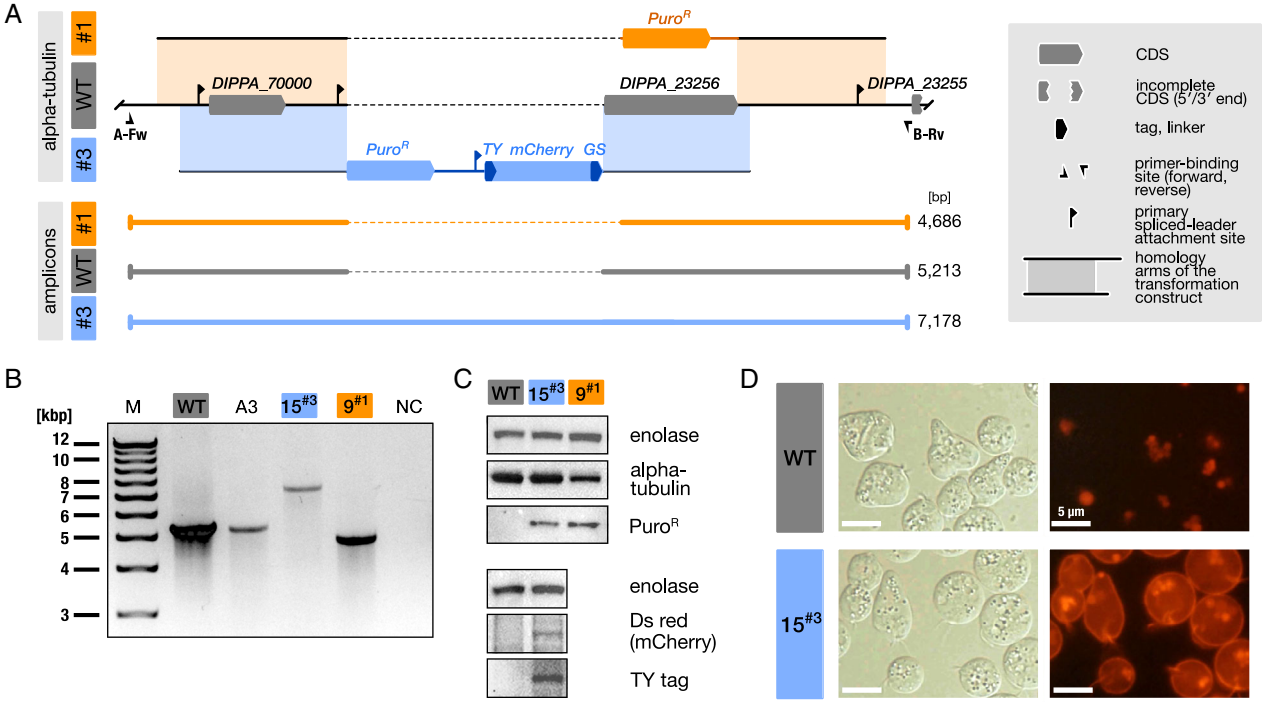

Fig 1. Replacement of the alpha-tubulin gene and N-terminal tagging under puromycin selection. A. Schema of the genomic neighbourhood of the wild-type (WT) alpha-tubulin locus (*DIPPA_23256*) and its comparison to the replacement (#1) and tagging (#3) constructs. The schema includes positions of the homology arms, primers used for PCR validation of the on-target integration, and expected sizes of the amplicons. (For an in-scale schema, see Supporting Information Fig. S3.) B. PCR of total DNA of *D. papillatum* WT and transformant cell lines 9 (construct #1) and 15 (construct #3) using primers (A-Fw, B-Rv) that bind outside the target region. Cell line A3 contains a type-#3 construct integrated into a heterologous location (see also (Kaur *et al*., 2018). Negative control PCR (NC) was performed without template DNA. (Uncropped gel is shown in Supporting Information Fig. S8.) C. Western blot analysis of *D. papillatum* wild-type and transformant cell lines $9^{\#1}$ and $15^{\#3}$. Enolase was used as a loading control. (Uncropped blots and detailed information on the used antibodies are shown in Supporting Information Fig. S9.) D. Representative epifluorescence micrographs of *D. papillatum* wild-type and transformant cell line $15^{\#3}$, which expresses the mCherry-alpha-tubulin fusion protein.

Information Fig. S5A). Next, the correct integration of the N-terminally tagged α-tubulin with mCherry was validated by SL RT-PCR (Supporting Information Fig. S5B–D), and the expression of Puro^R, Ty tag and mCherry was verified by Western blot analysis (Fig. 1C). Finally, using fluorescent microscopy, we confirmed that live, transformed cells, including their flagella, lit up as expected for fluorescence-tag labelled α-tubulin (Fig. 1D). Cell lines with successfully integrated constructs were stable even after several months in culture.

α- and β-tubulin replacement with V5-tagged neomycin resistance marker

Based on the results described above, we decided to test the extended homology arms in combination with the neomycin resistance marker (*Neo^R*; encoding the aminoglycoside 3′-phosphotransferase, APT). To facilitate Western blot-based detection of the translated protein, a triple V5 tag was fused to the 5′ terminus of the gene (constructs #4 and #5) (Fig. 2A and B and Supporting Information Fig. S3).

The V5-*Neo^R* fusion flanked by the partial 5′ and 3′ hexokinase UTRs of the trypanosomatid *Blastocrithidia* sp. p57 (GenBank: MN047315) was inserted into the *D. papillatum* genome and, despite its random integration, was efficiently expressed (Faktorová *et al*., 2020). Importantly, the V5-*Neo^R* fusion protein was catalytically active. To the best of our knowledge, this is the first report describing the deployment of a tagged resistance marker in eukaryotes. Our observation that the N-terminally tagged aminoglycoside 3′-phosphotransferase, which confers resistance to the neomycin/geneticin family of antibiotics, remains functional, is useful for verifying the expression of the resistance marker.

Next, we modified the V5-*Neo^R* fusion construct and flanked it with the same *D. papillatum* long homologous regions of the α- and β-tubulin ORFs (*DIPPA_23256* and *DIPPA_12526* respectively) (Fig. 2A and B). A total of 9 and 10 neomycin-resistant cell lines respectively were recovered after 8–10 days following electroporation with the constructs. Six α-tubulin and seven β-tubulin replacement cell lines contained the extraneous DNA integrated into the intended location (Fig. 2C and D). Two cell lines

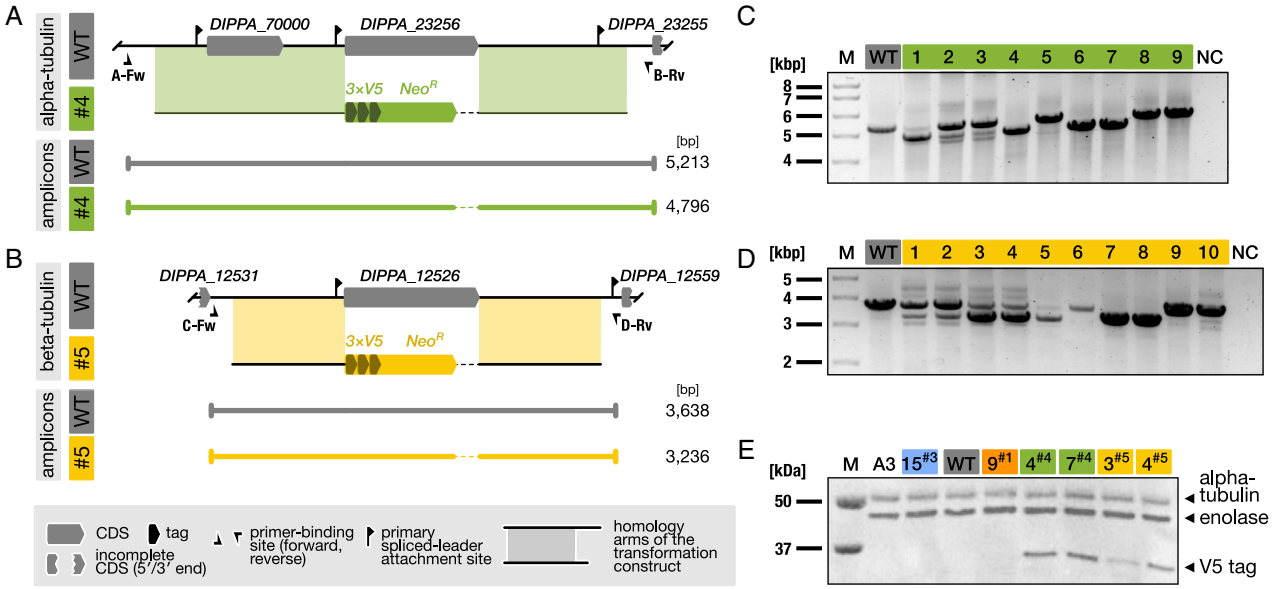

Fig 2. Replacement of alpha- and beta-tubulin genes with a V5-*Neo*^{R} fusion. A. Schema of the genomic neighbourhood of the wild-type (WT) alpha-tubulin locus (*DIPPA_23256*) and its comparison to the V5-*Neo*^{R} replacement (#4) construct. B. Schema of the genomic neighbourhood of the wild-type beta-tubulin locus (*DIPPA_12526*) and its comparison to the V5-*Neo*^{R} replacement (#5) construct. The schemas in A and B include positions of the homology arms, primers used for PCR validation of the on-target integration, and expected sizes of the amplicons. (For in-scale schemas, see Supporting Information Fig. S3.) C and D. PCR of total DNA of *D. papillatum* wild-type and transformant cell lines containing the construct #4 (C) or #5 (D) using primers that bind outside the target region. (Uncropped gels are shown in Supporting Information Fig. S8.) E. Western blot analysis of *D. papillatum* wild-type and selected transformant cell lines. Enolase was used as a loading control. (Uncropped blots and detailed information on the used antibodies are shown in Supporting Information Fig. S9.)

of each tubulin replacement experiment (4$^{\#4}$, 7$^{\#4}$, 7$^{\#5}$ and 8$^{\#5}$), in which the WT allele was not present, were selected for further validation of gene expression (Supporting Information Fig. S6A) and translation (Fig. 2E and Supporting Information Fig. S6B). Two other cell lines (3$^{\#5}$ and 4$^{\#5}$), containing a band of the WT allele together with the replaced one, were also tested and V5-*Neo*^{R} expression was confirmed as well (Fig. 2E).

pDP002 plasmid – a modular construct for N- and C-terminal tagging

Based on the aforementioned observations, we designed a construct named pDP002 to be used for high-throughput tagging of any chosen gene in the *D. papillatum* nuclear genome (Fig. 3A; GenBank: MT232523). The design of this construct is based on the modular pPOT (PCR only tagging) and pLENT (*Leishmania* endogenous tagging) series of plasmids, which were recently developed for use in the trypanosomatids *Trypanosoma brucei* and *Leishmania mexicana,* respectively (Dean *et al*., 2015; Dean *et al*., 2017) and have been deployed successfully worldwide (Goos *et al*., 2017; Benz and Urbaniak, 2019; Sunter *et al*., 2019), including our laboratory (Peña-Diaz *et al*., 2018; Faktorová *et al*., 2018). The pDP002 plasmid is primarily intended to serve as a template for C- and/or N-terminal tagging and carries diplonemid codon-optimized versions of the Protein A (*PrA*) tag and the two resistance marker genes, *Hyg*^{R} (hygromycin B phosphotransferase, *hph*; for N-terminal protein tagging) and *Neo*^{R} (for C-terminal protein tagging). The Protein A tag (25 kDa) can increase the solubility and/or expression of heterologous proteins (Sambrook *et al*., 1989), is easily detected by commercially available antibodies and is commonly used for protein immunoprecipitations (IP) with the aim to investigate the composition of protein complexes (Trahan *et al*., 2016). As indicated above, transformation efficiency may depend on the selection marker. We observed in *D. papillatum* that although puromycin was the most efficient antibiotic (Kaur *et al*., 2018), *Puro*^{R}-based selection took longer and the transformation efficiency was somewhat lower compared to *Neo*^{R} (Faktorová *et al*., 2020) or *Hyg*^{R} (our unpubl. data).

C-Terminal tagging of DpMRPS49 with PrA *under* Neo^R *selection*

The first gene that we tagged using the pDP002 template encodes a protein of the small mitoribosomal subunit, *DpMRPS49* (*DIPPA_31280*; construct #6). As the gene contains an N-terminal targeting signal for import into the mitochondrion, we added the *PrA* tag to its C-terminus (Fig. 3B and Supporting Information Fig. S3). Testing of the cell lines by PCR showed the correct location of integration in both tested cell lines (Fig. 3D). The expression

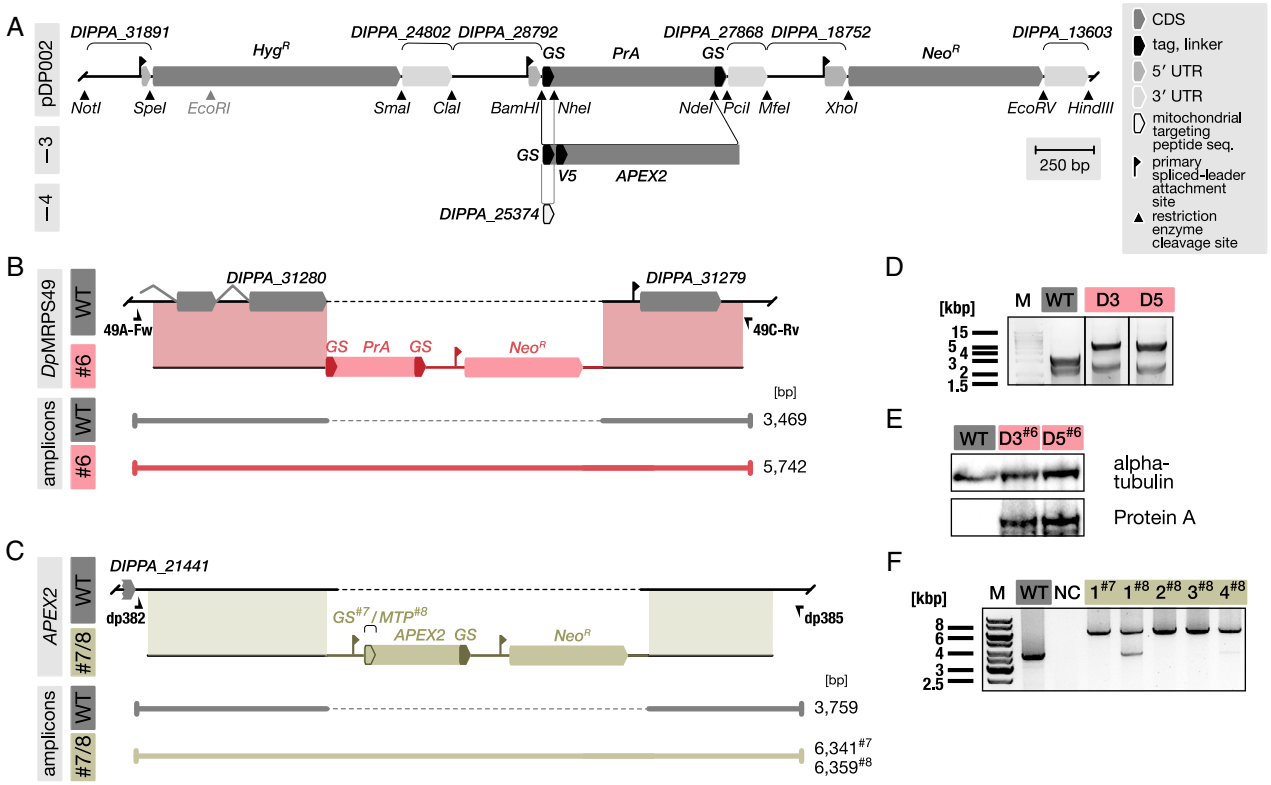

Fig 3. Precise gene tagging and replacement using the modular pDP002 plasmid and its variants. A. In-scale schema of the pDP002 plasmid, which allows N- and C-terminal Protein A (*PrA*) tagging, and plasmid variants pDP003 and pDP004 for the expression of APEX2 targeted to the cytosol and mitochondrion, respectively. The schema includes positions of the restriction enzyme sites and regulatory elements that drive the expression of the selection markers and tagging coding sequence. Systematic names of genes, from which the elements were derived, are shown on the uppermost track. B. Schema of the genomic neighbourhood of the wild-type (WT) MRPS49 locus (*DIPPA_31280*) and its comparison to the C-terminal *PrA*-tagging construct (#6). C. Schema of the genomic neighbourhood of the wild-type locus between genes *DIPPA_21441* and *DIPPA_21439* and its comparison to the APEX2-insertion constructs (#7, cytosolic APEX2; #8, mitochondrial APEX2).The schemas in B and C include positions of the homology arms, primers used for PCR validation of the on-target integration and expected sizes of the amplicons. (For in-scale schemas, see Supporting Information Fig. S3.) D. PCR of total DNA of *D. papillatum* wild-type and six selected transformant cell lines using primers that bind outside the target region. (Uncropped gel is shown in Supporting Information Fig. S8.) E. Western blot analysis of *D. papillatum* WT and selected transformant cell lines. Alpha-tubulin was used as a loading control. (Uncropped blot is shown in Supporting Information Fig. S9.) F. PCR of total DNA of *D. papillatum* wild-type and transformant cell lines containing the construct #7 or #8 using primers that bind outside the target region. Negative control PCR (NC) was performed without template DNA. (Uncropped gel is shown in Supporting Information Fig. S8).

of the tagged protein was confirmed by Western blot analysis using anti-PrA antibody (Fig. 3E).

pDP003 and pDP004 – modified constructs used for APEX2 expression

Two modified constructs, pDP003 and pDP004 (Fig. 3A; GenBank: MT232524 and MT232525), were designed for the expression of V5-tagged cytosolic and mitochondrial activity-enhanced ascorbate peroxidase (APEX2) (Lam *et al*., 2015), respectively. The goal was to eventually conduct with the corresponding transformants proximity-labelling experiments of proteins (and RNAs) with biotin-phenol (see constructs #7 and #8 in Fig. 3C and Supporting Information Fig. S3). For this experiment, we employed a slightly modified transformation protocol (see Experimental procedures) to test the robustness of the earlier established protocol. As the integration site, we chose here a moderately repetitive intergenic region (between the genes *DIPPA_21441* and *DIPPA_21439*).

The APEX2 construct was correctly integrated in all five cell lines obtained; in two cases, we also observed a fainter, WT size PCR product (Fig. 3F). All cell lines produced mRNA with a SL attached to the 5′ end of the transcript (Supporting Information Fig. S7A). However, our attempt to confirm APEX2 translation by Western blot analysis employing the anti-V5 antibody was unsuccessful (Supporting Information Fig. S7B).

Discussion

Diplonemids have remained a largely enigmatic group despite their abundance and ubiquity in the marine ecosystem. For functional studies of their cellular components,

a high-quality nuclear genome sequence, as well as methods for genetic manipulation, are needed. Although the release of the genome assembly and annotation of *D. papillatum* is underway (Burger *et al*., unpublished), we recently implemented the genetic tools in this species (Kaur *et al*., 2018; Faktorová *et al*., 2020).

An extraneous DNA stably introduced into a cell by electroporation usually has one of the following fates: (i) correct integration in the target genomic locus via HR, (ii) random integration via the NHEJ pathway, or (iii) retention in the form of an extrachromosomal plasmid. As in most multicellular eukaryotes NHEJ works more efficiently compared to HR, the chances of random integration of the introduced DNA are higher (Malkova and Haber, 2012; Rodgers and McVey, 2016). Only relatively few eukaryotes, for example, the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, display high HR efficiency (Hegemann *et al*., 2014). Nevertheless, in organisms with inefficient HR, this pathway can be enhanced by increasing the length of the homologous sequences. Alternatively, the relative efficiency of the HR pathway can be increased by blocking NHEJ, for example via mutating the Ku70/80 proteins, which are the central players of this pathway (Ninomiya *et al*., 2004; Goins *et al*., 2006; Nayak *et al*., 2006; Nenarokova *et al*., 2019).

Here, we report the knock-out of α- and a β-tubulin genes of *D. papillatum* using constructs with extended homology arms. Moreover, we provide evidence for the expression of fusion proteins, the N-terminally tagged α-tubulin and the C-terminally tagged mitoribosomal protein. Transformations with all constructs were successful, yet viability and transformation efficiency was higher with the neomycin than with the puromycin resistance marker. In all but one construct, at least one cell line with correct integration was obtained, which leads us to conclude that HR is functional in *D. papillatum*. We also demonstrated that targeted integration was largely dependent on the length of the homology arms. We built a total of five constructs with >1 kbp-long homology arms, four for the replacement of the α- and β-tubulin genes, and one for the N-terminal tagging of α-tubulin, yielding in total 33 clonal cell lines. Furthermore, we designed the modular plasmids pDP002, pDP003 and pDP004 for high-throughput N- and C-terminal tagging. Lastly, a variant transformation protocol was tested independently, confirming the robustness of the devised strategy.

Taken together, out of 40 obtained cell lines, the integration into the correct position was not achieved in 18. Conversely, in at least 13 cell lines, we found no sign of the WT locus together with a correctly integrated construct. In additional nine cell lines, we observed WT-sized PCR amplicons in addition to the construct-sized band, which we attribute to untransformed cells remaining in the population. As a complete replacement of the targeted region is unlikely in a diploid (or higher ploidy), we infer that *D. papillatum* is most likely haploid, though further evidence is required.

In summary, we provide tools that allow gene tagging, knock-out and knock-in strategies in *D. papillatum* and likely other diplonemids. These tools are suited for single-copy genes. Targeting multi-copy genes will require the implementation of the CRISPR/Cas9 technology, which we are currently pursuing. Another possibility is to use RNA interference because the required machinery is present in *D. papillatum* (our unpubl. data). Because this diplonemid is resistant to tetracycline, the RNA interference toolkit that was developed for the kinetoplastid *T. brucei* (Matthews *et al*., 2015) is a promising candidate. We believe that the availability of the straightforward and efficient transformation strategies described here will pave the way for a systematic inquiry about diplonemid cell biology by reverse genetics.

Experimental procedures

Strain, cultivation and determination of resistance to W7 inhibitor

Diplonema papillatum (ATCC 50162) was cultivated axenically as described previously (Kaur *et al*., 2018; Valach *et al*., 2018). The experiment was performed using Alamar Blue assay, which measures the viability by fluorescence, as described previously (Kaur *et al*., 2018), to determine the optimal concentration and to address possible toxic effects of W7 (N-(6-aminohexyl)-5-chloro1-naphthalene sulfonamide) inhibitor. A total of 5×10^7 cells (2×10^6 cells/ml) were pre-incubated in 5 μg ml^{-1} of W7 inhibitor for 4 h before electroporation. Cells were harvested and electroporated as described below. The transfectants were subjected to selection with increasing concentrations (12–40 μg ml^{-1}) of puromycin, and the genomic DNA of transformants were examined by genomic DNA PCR.

Design and preparation of transformation cassettes

All cassettes (except for APEX2 constructs; see below) were prepared by a fusion PCR approach (Supporting Information Fig. S2) using Phusion or Q5 polymerase (NEB Biolabs, M0530S and M0491S respectively). In brief, the first three individual PCRs were used to amplify 5′ long homology region (PCR A), the cassette designed to replace/tag a gene of interest (PCR B), and 3′ long homology region (PCR C). PCR B-Fw and PCR B-Rv primers were designed to overlap with PCR A-Rv and PCR C-Fw primers respectively. The length of homologous arms depended on the length of the non-repetitive sequence in the vicinity of the genes and varied from ∼1 to 2 kbp. Nested primers (PCR D-Fw and Rv) were used

for joining all three pieces in the final product. Used primers are listed in Supporting Information Table 1. PCR-amplified cassettes (Figs 1A, 2A, B, 3B and C and Supporting Information Fig. S3) were gel purified by the GeneAll Expin Combo GP purification kit (112–102), ethanol-precipitated and DNA was then electroporated into the cells. The details on the preparation of the individual transformation cassettes are specified hereafter.

Replacement of α and β-tubulin genes with Puro^R *or V5 +* Neo^R*.* For amplification of the replacement cassette (PCR B), puromycin resistant marker (*Puro^R*) flanked with UTRs of α-tubulin gene (*DIPPA_23256*) was PCR amplified from the previously described construct (Kaur *et al*., 2018). Similarly, V5 + *Neo^R* cassette was amplified from p57-V5 + Neo^R plasmid (GenBank MN047315). Regions surrounding α-tubulin (2016 bp and 1840 bp) and β-tubulin (952 bp and 1333 bp) genes were used for amplification of long homology arms (PCR A and C). The final transformation cassettes of 1841 bp + 1553 bp and 890 bp + 1203 bp homology regions (PCR D) were amplified using nested primers for α and β-tubulin respectively. For details, see Figs 1A, 2A and B and Supporting Information Fig. S3 and Table 1.

Endogenous N-terminal tagging of α-tubulin using extended homologous arms. A similar N-terminal tagging approach and the same construct as described elsewhere (Kaur *et al*., 2018) were used; but here, we extended the 5′ and 3′ homologous regions using the nested PCR approach to enhance the probability of integration of the construct into the targeted locus. A schematic representation of this construct is shown in Fig. 1A and Supporting Information Fig. S3.

Endogenous C-terminal tagging of DpMRPS49 with PrA *under* Neo^R *selection.* The modular plasmid pDP002 (Fig. 3A) with *D. papillatum* codon-optimized *Hyg^R*, *PrA* and *Neo^R* coding sequences was synthesized by Eurofins Genomics (Ebersberg, Germany). To tag *DpMRPS49* with a Protein A tag at its C-terminus, pDP002 was used as a template for amplification of the tag and the downstream *Neo^R* marker (Fig. 3B, Supporting Information Fig. S3). The 5′ and 3′ homology arms were 1575 bp and 1605 bp long respectively. The final nested PCR product was A-tailed, cloned into pTOPO 2.1 (Thermo Fisher) and validated by sequencing. For transformation of *D. papillatum*, the tagging cassette was amplified from this plasmid using nested gene-specific primers. About 5 μg of purified PCR product was used for electroporation.

V5–tagged heterologous ascorbate peroxidase – integration into intergenic region. Diplonema papillatum codon-optimized V5-APEX2 coding sequence was synthesized as a gBlocks gene fragment by Integrated DNA Technologies (Coralville, IA, USA). The plasmids pDP003 and pDP004 (Fig. 3A) were created by replacing the coding sequence of Protein A in pDP002 by cloning via the restriction sites *NheI* and *NdeI* (cytosolic APEX2 in pDP003) or *BamHI* and *NdeI* (mitochondrion-targeted APEX2 in pDP004). Constructs for transformation were prepared by cloning homology arms amplified from *D. papillatum* genomic DNA with dp382 + dp383 (upstream arm) and dp384 + dp385 (downstream arm) into pDP003 (or pDP004) via the restriction sites *EcoRI* and *ClaI* (upstream arm) and *HindIII* (downstream arm). As the resulting plasmid had two *EcoRI* sites, one in each homology arm, the insertion cassette for *Diplonema* transformation was prepared by digesting the plasmid by *EcoRI-HF* (New England Biolabs). The reaction was heat-inactivated, precipitated with 1.5 M NaCl and 1.4 V isopropanol overnight at 4 °C, and finally solubilized in 10 mM Tris–HCl, pH8.0 to the concentration ~400 ng μL^{-1}. For a single electroporation, 2 μg (5 μl) of the digested plasmid were used.

Electroporation and transformant selection

All constructs (except APEX2; see below) were transformed as described previously (Kaur *et al*., 2018), with a DNA-free negative control. The applied protocol is similar to the one frequently used for *Trypanosoma brucei*. Briefly, 5×10^7 cells were pelleted, resuspended in the electroporation buffer, supplemented with 3–5 μg of the DNA construct, placed in cuvettes and electroporated using an Amaxa Nucleofector II apparatus, program X-001. The concentrations of all tested antibiotics had to be higher as compared with those used for *T. brucei*, with the duration of selection depending on the antibiotic used. The transformation efficiency is slightly higher in *T. brucei*.

In the case of the W7 inhibitor experiment, the cells were pre-incubated in W7 prior to electroporation (see above). Eight to sixteen hours after the electroporation, transfectants were subjected to selection with increasing concentrations of puromycin (12–40 μg ml^{-1}) for *Puro^R* or G418 (25–80 μg ml^{-1}) for *Neo^R* containing constructs.

APEX2 constructs were transformed according to a modified protocol, whose details are available at https://doi.org/10.17504/protocols.io.bedxja7n. Briefly, 10^7 *D. papilatum* cells from exponential growth phase were electroporated in 0.2 mm cuvettes in a transformation buffer (25 mM HEPES, pH 7.5, 25 mM KCl, 0.15 mM $CaCl_2$, 10 mM NaH_2PO_4, pH 7.5, 2.5 mM $MgCl_2$, 1 mM EDTA, 30 mM [0.5%] glucose, 145 mM [4.35%] sucrose, 0.1 mg mL^{-1} BSA, 1 mM ITP) using Gene Pulser Xcell apparatus (Bio-Rad) at 1500 V for 0.3 ms. After a 6 h-

long recuperation period, transformed cells were selected in the presence of G418 (100 μg mL^{-1}).

PCR on genomic DNA

For verification of the correct integration of the constructs in *D. papillatum*, the genomic DNA was isolated by DNA isolation kit (Qiagen, 69504) or by phenol-chloroform extraction and used as a template. Primers used are shown in Figs 1A, 2A, B, 3B and C and Supporting Information Fig. S3 and their sequences are listed in Supporting Information Table 1. PCR amplification was performed with Phusion or Q5 DNA polymerase using the manufacturer-recommended PCR program. All cassette integrations were confirmed by sequencing of the PCR products.

RNA isolation and cDNA synthesis, SL RT-PCR

Total RNA was isolated using TriReagent (MRC, TR118) or by a home-made Trizol substitute (Rodríguez-Ezpeleta *et al*., 2009). cDNA was prepared using the QuantiTect Reverse Transcription Kit (Qiagen, 205311) or SuperScript IV reverse transcriptase (Thermo) with random primers. PCR was performed on cDNA with primers shown in Supporting Information Fig. S3 (for primer sequences, see Supporting Information Table 1), and Q5 or OneTaq polymerase, as described previously (Kaur *et al*., 2018). DpSL_Fw1 and DpSL_Fw2 primers derived from the SL-RNA gene were used in combination with primers targeting CDSs of mCherry (SL_mCherry_Rv1; SL_mCherry_Rv2), $Puro^R$ (SL_Puro_Rv1; SL_Puro_Rv2), α-tubulin (SL_Atubulin_Rv1; SL_Atubulin_Rv2), V5 + Neo^R (SL_NeoR_Rv1; SL_NeoR_Rv2), Protein A (SL_protA_Rv1; SL_protA_Rv2) or APEX2 (dp375). The position of primers and the expected size of PCR products are shown in Supporting Information Figs S3, S5B–D, S6A and S7A. The obtained amplicons were verified by sequencing.

Western blot analysis

Cell lysates were prepared by resuspending 5×10^5 cells in 25 μl of 2× SDS sample buffer and separated on 4–12% (v/v) NuPAGE gels (Invitrogen, NP0322BOX) for anti-Ds Red and anti-puromycin N-acetyltransferase antibodies and 4–12% Tris-Glycine gels (Invitrogen, XP04122BOX) for the anti-V5 antibody. After the run, proteins were transferred onto a PVDF membrane by electroblotting. Membranes were blocked with 5% (w/v) skimmed milk prepared in 1× PBS + 0.5% (v/v) Tween 20 and probed with specific primary antibodies: anti-puromycin N-acetyltransferase antibody (produced in rabbit used at 1:500; Thermo Fisher, 702389) for $Puro^R$; anti-Ty antibody (produced in mouse used at 1:1000; Sigma-Aldrich, SAB4800032) for the Ty-tag; anti-Ds Red antibody 1:1000 (produced in-rabbit; Clontech, 632496) for mCherry; anti-V5 antibody [produced in rabbit (PA1-993) or mouse (R960-25), used at 1:1000; ThermoFisher Scientific] for the V5 tag; anti-Protein A antibody (produced in rabbit used at 1:20,000; Sigma-Aldrich, P3775) for protein A and anti-α-tubulin antibody (produced in mouse used at 1:5000; Sigma-Aldrich, T9026). As a loading control, the anti-enolase antibody (produced in rabbit, used at 1:2000; gift of J. Morales) was used to determine the level of enolase in *D. papillatum*. The membrane was subsequently incubated with a horseradish peroxidase-conjugated secondary anti-mouse or anti-rabbit polyclonal antibody at 1:2000 dilution (Sigma) at room temperature for 1 h and visualized using Clarity Western ECL substrate (Bio-Rad).

Fluorescence microscopy

Five microlitres of *D. papillatum* live cells were placed on a slide, covered with a coverslip, cells were allowed to immobilize for 2–5 min and were subsequently observed under an AxioPlan 2 imaging fluorescence microscope (Zeiss). A video was recorded and individual images were obtained using Media Player Classic program and processed using Gimp 2.8.8 software.

Sequence accession numbers

The DNA sequences reported here were deposited in GenBank under the accession numbers MT232523–MT232525.

ACKNOWLEDGEMENTS

We thank Priscila Peña-Diaz and Michaela Svobodová (Biology Centre) for help with designing of the primers for long homology regions and PCR amplification, respectively, and Jorge Morales (Heinrich-Heine University, Dusseldorf, Germany) for anti-enolase antibodies. Support from the Czech Grant Agency (20-07186S to J.L.), the ERC CZ (LL1601 to J.L.), the Czech Ministry of Education (project Centre for research of pathogenicity and virulence of parasites r.n.: CZ.02.1.01/0.0/0.0/16_019/0000759 to J.L.), the Natural Sciences and Engineering Research Council of Canada (NSERC; RGPIN-2019-04024 to G.B.), and from the Gordon and Betty Moore Foundation (GBMF4983.01 to G.B. and J.L.) is kindly acknowledged.

Authors have no conflict of interest to declare.

AUTHOR CONTRIBUTIONS

D.F., B.K., M.V., C.B.: construct design; D.F., B.K., M.V., L.G., C.B.: cell cultivation, PCR, SL RT-PCR and Western

blotting; D.F., L.G.: fluorescence microscopy; G.B.: analysis of genes involved in recombination; D.F., B.K.: writing of the initial manuscript draft; J.L., M.V., D.F., G.B.: review and editing; M.V.: visualization; J.L., G.B.: supervision.

REFERENCES

Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., *et al*. (2019) Revisions to the classification, nomenclature, and diversity of eukaryotes. *J Eukaryot Microbiol* **66**: 4–119.

Arras, S.D.M., and Fraser, J.A. (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.

Benz, C., and Urbaniak, M.D. (2019) Organising the cell cycle in the absence of transcriptional control: dynamic phosphorylation co-ordinates the *Trypanosoma brucei* cell cycle post-transcriptionally. *PLoS Pathog* **15**: e1008129.

Byrum, J., Jordan, S., Safrany, S.T., and 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.

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., *et al*. (2018) A global atlas of eukaryotic genes. *Nat Commun* **9**: 373.

Clayton, C.E. (2016) Gene expression in Kinetoplastids. *Curr Opin Microbiol* **32**: 46–51.

Dean, S., Sunter, J.D., and Wheeler, R.J. (2017) TrypTag.org: a trypanosome genome-wide protein localisation resource. *Trends Parasitol* **33**: 80–82.

Dean, S., Sunter, J., Wheeler, R.J., Hodkinson, I., Gluenz, E., and Gull, K. (2015) A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. *Open Biol* **5**: 140197.

Eichinger, L., Lee, S.S., and Schleicher, M. (1999) *Dictyostelium* as model system for studies of the actin cytoskeleton by molecular genetics. *Microsc Res Tech* **47**: 124–134.

Faktorová, D., Bär, A., Hashimi, H., McKenney, K., Horák, A., Schnaufer, A., et al. (2018) TbUTP10, a protein involved in early stages of pre-18S rRNA processing in *Trypanosoma brucei*. *Mol Biochem Parasitol* **225**: 84–93.

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* **17**: 481–494. https://doi.org/10.1038/s41592-020-0796-x.

Faktorová, D., Valach, M., Kaur, B., Burger, G., and Lukeš, J. (2018) Mitochondrial RNA editing and processing in diplonemid protists. In *RNA Metabolism in Mitochondria*, Gray, M.W., and Cruz-Reyes, J. Cham, Switzerland: (eds): Springer International Publishing AG, pp. 145–176.

Flegontova, O., Flegontov, P., Malviya, S., Audic, S., Wincker, P., de Vargas, C., *et al*. (2016) Extreme diversity of diplonemid 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 diplonemids. *Curr Biol* **26**: 3053–3059.

George, E., Husnik, P.F., Tashyreva, D., Prokopchuk, G., Horák, A., Kwong, W.K., *et al*. (2020) Highly reduced genomes of protist endosymbionts show evolutionary convergence. *Curr Biol* **30**: 925–933.

Goins, C.L., Gerik, K.J., and 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.

Goos, C., Dejung, M., Janzen, C.J., Butter, F., and Kramer, S. (2017) The nuclear proteome of *Trypanosoma brucei*. *PLoS One* **12**: e0181884.

Hegemann, J.H., Heick, S.B., Pöhlmann, J., Langen, M.M., and Fleig, U. (2014) Targeted gene deletion in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Methods Mol Biol* **1163**: 45–73.

von der Heyden, S., Chao, E.E., Vickerman, K., and Cavalier-Smith, T. (2004) Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of euglenozoa. *J Eukaryot Microbiol* **51**: 402–416.

Jackson, A.P., Vaughan, S., and Gull, K. (2006) Evolution of tubulin gene arrays in trypanosomatid parasites: genomic restructuring in *Leishmania*. *BMC Genomics* **18**: 261.

Kaur, B., Valach, M., Peña-Diaz, P., Moreira, S., Keeling, P. J., Burger, G., *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.

Kaur, B., Záhonová, K., Valach, M., Faktorová, D., Prokopchuk, G., Burger, G., and Lukeš, J. (2020) Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplonemids. *Nucleic Acids Res* **48**: 2694–2708.

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.

Krejci, L., Altmannova, V., Spirek, M., and Zhao, X. (2012) Homologous recombination and its regulation. *Nucleic Acids Res* **40**: 5795–5818.

Lam, S.S., Martell, J.D., Kamer, K.J., Deerinck, T.J., Ellisma n, M.H., Mootha, V.K., and Ting, A.Y. (2015) Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* **12**: 51–54.

López-García, P., Vereshchaka, A., and Moreira, D. (2007) Eukaryotic diversity associated with carbonates and fluid sea water interface in lost city hydrothermal field. *Environ Microbiol* **9**: 546–554.

Lukeš, J., Flegontova, O., and Horák, A. (2015) Diplonemids. *Curr Biol* **25**: R702–704.

Lukeš, J., Wheeler, R., Jirsová, D., David, V., and Archibald, J.M. (2018) Massive mitochondrial DNA content

in diplonemid and kinetoplastid protists. *IUBMB Life* **70**: 1267–1274.
Malkova, A., and Haber, J.E. (2012) Mutations arising during repair of chromosome breaks. *Annu Rev Genet* **46**: 455–473.
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 diplonemids, the sister group of kinetoplastids. *Eukaryot Cell* **4**: 1137–1146.
Massana, R. (2011) Eukaryotic picoplankton in surface oceans. *Annu Rev Microbiol* **65**: 91–110.
Matthews, K.R., McCulloch, R., and Morrison, L.J. (2015) The within-host dynamics of African trypanosome 726 infections. *Philos Trans R Soc Lond B Biol Sci* **370**: 20140288.
McKean, P.G., Vaughan, S., and Gull, K. (2001) The extended tubulin superfamily. *J Cell Sci* **114**: 2723–2733.
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.
Mukherjee, I., Hodoki, Y., Okazaki, Y., Fujinaga, S., Ohbayashi, K., and Nakano, S.I. (2019) Widespread dominance of kinetoplastids and unexpected presence of diplonemids in deep freshwater lakes. *Front Microbiol* **10**: 2375.
Nayak, T., Szewczyk, E., Oakley, C.E., Osmani, A., Ukil, L., Murray, S.L., *et al*. (2006) A versatile and efficient gene-targeting system for *Aspergillus nidulans*. *Genetics* **172**: 1557–1566.
Nenarokova, A., Záhonová, K., Krasilnikova, M., Gahura, O., McCulloch, R., Zíková, A., *et al*. (2019) Causes and effects of loss of classical nonhomologous end joining pathway in parasitic eukaryotes. *mBio* **10**: e01541–19.
Ninomiya, Y., Suzuki, K., Ishii, C., and Inoue, H. (2004) Highly efficient gene replacements in *Neurospora* strains deficient for nonhomologous end-joining. *Proc Natl Acad Sci USA* **101**: 12248–12253.
Okamoto, N., Gawryluk, R.M.R., Campo, J., Strassert, J.F. H., Lukeš, J., Richards, T.A., *et al*. (2019) A revised taxonomy of diplonemids including the Eupelagonemidae n. fam. and a type species, *Eupelagonema oceanica* n. gen. & sp. *J Eukaryot Microbiol* **66**: 519–524.
Peña-Diaz, P., Mach, J., Kriegová, E., Poliak, P., Tachezy, J., and Lukeš, J. (2018) Trypanosomal mitochondrial intermediate peptidase does not behave as a classical mitochondrial processing peptidase. *PLoS One* **13**: e0196474.
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.
Rodgers, K., and McVey, M. (2016) Error-prone repair of DNA double-strand breaks. *J Cell Physiol* **231**: 15–24.
Rodríguez-Ezpeleta, N., Teijeiro, S., Forget, L., Burger, G., and Lang, B.F. (2009) Construction of cDNA libraries: focus on protists and fungi. In *Methods in Molecular Biology*, Vol. **533**, Parkinson, J. (ed) . Totowa, NJ: Humana Press, pp. 33–47.
Roy, J., Faktorová, D., Benada, O., Lukeš, J., and Burger, G. (2007) Description of *Rhynchopus euleeides* n. sp. (Diplonemea), a free-living marine euglenozoan. *J Eukaryot Microbiol* **I54**: 137–145.
Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory.
Simpson, A.G.B. (1997) The identity and composition of the Euglenozoa. *Archiv für Protistenkunde* **148**: 318–328.
Son, M.Y., and Hasty, P. (2019) Homologous recombination defects and how they affect replication fork maintenance. *AIMS Genet* **5**: 192–211.
Sunter, J.D., Yanase, R., Wang, Z., Catta-Preta, C.M.C., Moreira-Leite, F., Myskova, J., *et al*. (2019) *Leishmania* flagellum attachment zone is critical for flagellar pocket shape, development in the sand fly, and pathogenicity in the host. *Proc Natl Acad Sci U S A* **116**: 6351–6360.
Tashyreva, D., Prokopchuk, G., Votýpka, J., Yabuki, A., Horák, A., and Lukeš, J. (2018a) Life cycle, ultrastructure, and phylogeny of new diplonemids and their endosymbiotic bacteria. *mBio* **9**: e02447–17.
Tashyreva, D., Prokopchuk, G., Yabuki, A., Kaur, B., Faktorová, D., Votýpka, J., *et al*. (2018b) Phylogeny and morphology of new diplonemids from Japan. *Protist* **169**: 158–179.
Trahan, C., Aguilar, L.-C., and Oeffinger, M. (2016) Single-step affinity purification (ssAP) and mass spectrometry of macromolecular complexes in the yeast *S. cerevisiae*. *Methods Mol Biol* **1361**: 265–287.
Valach, M., Léveillé-Kunst, A., Gray, M.W., and Burger, G. (2018) Respiratory chain complex I of unparalleled divergence in diplonemids. *J Biol Chem* **293**: 16043–16056.
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 diplonemids. *RNA Biol* **13**: 1204–1211.
de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., *et al*. (2015) Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**: 1261605.
Vickerman, K. (2000) Diplonemids (class: Diplonemea Cavalier Smith, 1993). In *An Illustrated Guide to the Protozoa*, Vol. **2**, 2nd edn, Lee, J.J., Leedale, G.F., and Bradbury, P. (eds). Lawrence, Kansas, USA: Society of Protozoologists, pp. 1157–1159.
Waters, C.A., Strande, N.T., Wyatt, D.W., Pryor, J.M., and Ramsden, D.A. (2014) Nonhomologous end joining: a good solution for bad ends. *DNA Repair* **17**: 39–51.
Yi, Z., Berney, C., Hartikainen, H., Mahamdallie, S., Gardner, M., Boenigk, J., *et al*. (2017) High-throughput sequencing of microbial eukaryotes in Lake Baikal reveals ecologically differentiated communities and novel evolutionary radiations. *FEMS Microbiol Eco* **93**. https://doi.org/10.1093/femsec/fix073.
Zhao, Z., Liu, H., Luo, Y., Zhou, S., An, L., Wang, C., *et al*. (2014) Molecular evolution and functional divergence of tubulin superfamily in the fungal tree of life. *Sci Rep* **23**: 6746.

Supporting Information

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

Appendix S1: Supporting Information