University of South Bohemia in České Budějovice

Faculty of Science

Transformation of marine protist *Diplonema papillatum* - characterization of selected cell lines

Bachelor Thesis

Lena Graf

Supervisor: RNDr. Drahomíra Faktorová, Ph. D.

České Budějovice, 2020

This thesis should be cited as: Graf L., 2020: Transformation of marine protist *Diplonema papillatum* - characterization of selected cell lines. Bc. Thesis in English, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic, 53 p.

Annotation

The aim of this thesis was to analyze and characterize selected transformant cell lines of marine protist *Diplonema papillatum.*

Declaration

I hereby declare that I have worked on my Bachelor thesis independently and used only the sources listed in the bibliography.
I hereby declare that, in accordance with Article 47b of Act No. 111/1998 in the valid wording, I agree with the publication of my Bachelor thesis, in full to be kept in the Faculty of Science archive, in electronic form in publicly accessible part of the STAG database operated by the University of South Bohemia in České Budějovice accessible through its web pages.
Further, I agree to the electronic publication of the comments of my supervisor and thesis opponents and the record of the proceedings and results of the thesis defence in accordance with aforementioned Act No. 111/1998.
I also agree to the comparison of the text of my thesis with the Theses.cz thesis database operated by the National Registry of University Theses and a plagiarism detection system.

České Budějovice, 21.05.2020 Lena Graf

Acknowledgements

Foremost, I would like to express my profound gratitude to my supervisor, RNDr. Drahomíra Faktorová, Ph. D., for introducing me to this topic and allowing me to take part in this project. She guided me throughout the whole process of this thesis with patience and provided me with knowledge to tackle all tasks. Furthermore, she created and confided me with the cell lines I was working with during this project.

I additionally would like to thank Martina Tesařová (Laboratory of Electron Microscopy, Institute of Parasitology) and Galina Prokopchuk (Laboratory of Molecular Biology of Protists, Institute of Parasitology) for the help with preparation and observation of electron microscopy pictures, respectively.

Moreover, I would like to thank Priscila Peña-Diaz (currently in Evolutionary protistology laboratory, BIOCEV, Prague) for assisting in the designing-process of the primers for the long homology regions.

Lastly, I would like to thank Jorge Morales (Heinrich-Heine University, Düsseldorf, Germany) for providing our laboratory with the anti-enolase antibodies.

Abbreviations

bp	base pairs
cDNA	complementary DNA
D. papillatum	*Diplonema papillatum*
DSPD	Deep sea pelagic diplonemids
EtBr	ethidium bromide
Fw primer	forward primer
HR	homologous recombination
MMix	master mix
mRNA	messenger RNA
mtDNA	mitochondrial DNA
N/A	not applicable
NHEJ	non-homologous end joining
Neo^R	neomycin resistance marker
OTU	operational taxonomic unit
PCR	polymerase chain reaction
$Puro^R$	puromycin resistance marker
rRNA	ribosomal RNA
RT	reverse transcription
Rv primer	reverse primer
SL	spliced leader
TEM	transmission electron microscopy
UTR	untranslated region
WT	wild-type

Table of content

1 Introduction

1.1 Euglenozoa

The phylum Euglenozoa belongs together with Heterolobosea to the supergroup Discoba (Hampl *et al.* 2009; Burki *et al.*, 2020; Fig. 1).

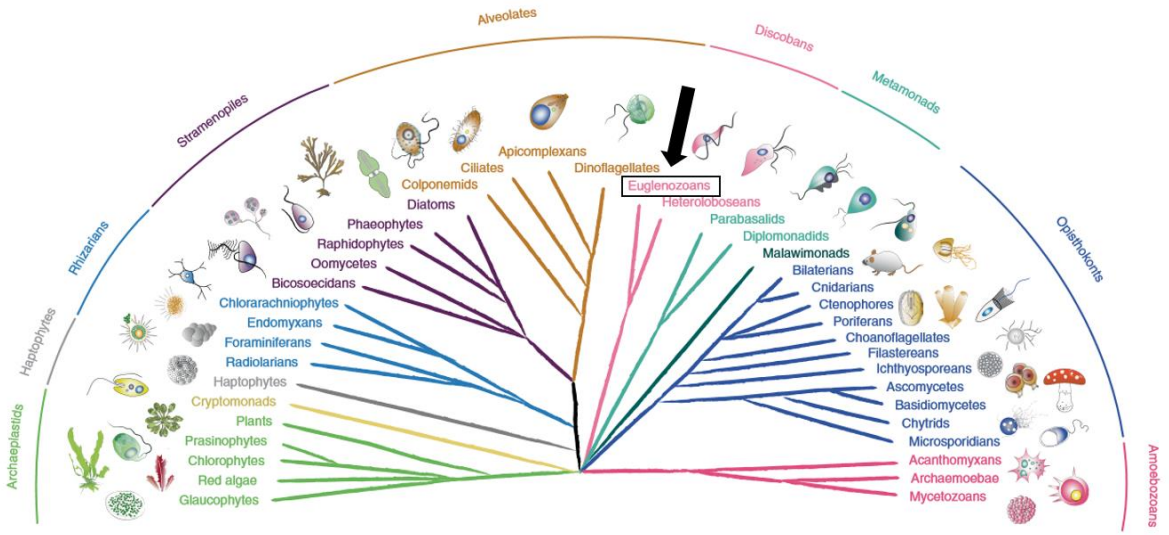


Figure 1: The phylum of Euglenozoa (see black arrow) on the eukaryotic tree of life (adapted from Faktorová et al. 2020). As the knowledge about eukaryotes is augmented the more studies with different approaches are arising, the harder it is to keep a definite grouping within this tree. Because there is still a significant amount of data missing, some branchings are not resolved yet till date (Burki et al., 2020)

In recent years, Euglenozoa attracted widespread interest as this group includes a significant number of organisms, belonging to either of the four subphyla - *Kinetoplastea Diplonemea, Euglenida* and *Symbiontida.* The members of this large group of unicellular flagellates display diverseness in morphology (Fig. 2), mode of nutrition as well as lifestyle - from free living to parasites of cattle and humans. In contrast to *Kinetoplastea*, represented by omnipresent and well-studied classes of trypanosomatids or bodonids and free living *Euglenida,* the other two groups were studied either poorly until recently – *Diplonemea,* or barely any data are available till now – *Symbiontida* (Adl *et al.*, 2019).

The major characteristics of Euglenozoa are comprised of having mostly two flagella with an apical or subapical pocket and mainly heteromorphic paraxonemal rods (Adl *et al.*,

2019). Initially, synamorphies were used to compile the phylum of Euglenozoa, which in detail were the presence of a distinct flagellar root pattern, paraxonemal rod structure and extrusomes. As paraxonemal rods are observed in *Euglenida* (euglenids), *Kinetoplastea* (kinetoplastids) and *Diplonemea* (diplonemids), these three lineages have been grouped together early on. The paraxonemal rods are divided into the anterior flagellum and the recurrent flagellum (Simpson *et al*., 1997).

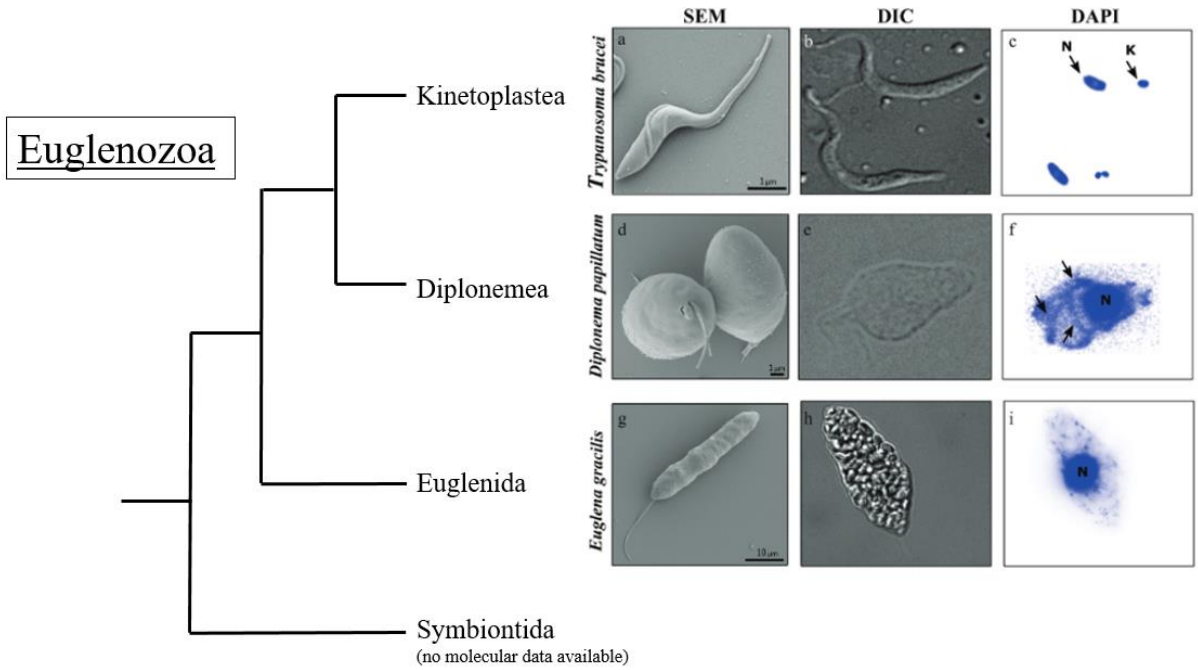


Figure 2: The scheme of Euglenozoa phylum and its groups - Kinetoplastea, Diplonemea, Euglenida and Symbiontida. The representatives of Kinetoplastea (T. brucei), Diplonemea (D. papillatum) and Euglenida (E. gracilis) are shown by scanning electron microscopy (SEM), light microscopy (DIC) and the distribution of their DNA throughout the cell is visualised by DAPI staining (Faktorová et al., 2016; Adl et al., 2019).

Within Euglenozoa, a basic motif of a ciliary apparatus composed of two kinetosomes, together with an asymmetrical pattern of three microtubular roots can be observed. Furthermore, a significant feature of Euglenozoa is the possession of a single mitochondrion, which predominently bear discoidal cristae (Adl *et al.*, 2019). The mitochondrion of diplonemids and kinetoplastids possesses an astonishingly massive DNA content; one theory is that the extensive DNA content in diplonemids and kinetoplastids arose after diverging of a common ancestor from *Euglena*. However, it is still unclear what and if any benefit such a large amount of mitochondrial DNA might provide (Lukeš *et al.*, 2018).

Diplonemea, *Kinetoplastea* and *Euglenida* are further connected by possessing kinetochores and a pre-replication complex, which hints towards a steadily loss of genes

connected to metabolically relevant enzymes in kinetoplastids. On the other hand, the free-living *Diplonemea* and *Euglenida* have remained metabolically more versatile even though being predisposed to undergo similar developments (Butenko *et al.*, 2020).

1.2 *Diplonemea*

The class of *Diplonemea* is comprised of a group of heterotrophic protists with exceptional metabolic capabilities, in addition to bearing two short flagella. As mentioned above, in contrast to its sister groups, diplonemids remained significantly understudied until recently, only a few genera and characterized species were known, namely *Diplonema*, *Rhynchopus* and *Hemistasia* (Simpson, 1997; Vickerman, 2000; von der Heyden *et al.*, 2004; Roy *et al.*, 2007; Massana, 2011).

Despite the fact that Eukaryotes consist mainly of unicellular organisms, the diversity of protists has been overlooked for a long time. Protists are major players in global primary production, which is to a half been contributed by the marine biosphere. This highlights the importance of the need to clarify unresolved ecological networks and address the lack of knowledge regarding morphological diversity, together with basic functional understanding (Baldauf, 2008; Massana, 2011; Worden *et al.*, 2015; Carradec *et al.*, 2018).

In order to reveal the diversity within marine eukaryotic plankton, samples were compiled during the global *Tara Oceans* expedition. The V9 region of the gene coding for 18S rDNA was chosen to enable differentiation by metabarcoding. On the basis of these metabarcodes, operational taxonomic units (OTUs) were created, presenting taxonomic relationships and lifestyles. As a result, over 85 % of the sequences acquired were associated with protistan eukaryotes and to a big surprise over 12,300 OTUs were attributed to diplonemids (de Vargas *et al.*, 2015). Moreover, further research augmented the numbers more than three times and more then 45,000 OTUs related to diplonemids were identified. This qualifies diplonemids as the most diverse eukaryotic plankton in the ocean, expressing significant global abundance (Flegentova *et al.*, 2016).

Comparing this pan-oceanic survey with past studies using universal eukaryotic primers leads to the observation that the number of diverse species associated with diplonemids is considerably higher than previously thought. Samples retrieved in prior studies were examined using primers for classical diplonemids and previously discovered

diplonemids from fluid-seawater interfaces of deep-sea hydrothermal vents (López-García *et al*., 2007; Lara *et al*., 2009).

Taken together, the phylogenetic analysis revealed that *Diplonemea* is indeed composed of four main groups: 1/ *Diplonemidae* – so called the 'classical' diplonemids, represented by genera *Diplonema*, *Rhynchopus*, *Lacrimia, Flectonema* and *Sulcionema* (Tashyreva *et al*., 2018a,b), 2/ *Hemistasiidae* - a small planktonic clade containing *Hemistasia*, *Artemidia* and *Namystynia* (Prokopchuk *et al.*, 2019), 3/ deep-sea pelagic diplonemids (DSPD) I and 4/ deep-sea pelagic diplonemids (DSPD) II. The DSPD I group (later renamed to *Eupelagonemidae* (Okamoto *et al.*, 2019) represents the most abundant class of marine diplonemids (97%) present primarily in the mesoplagic zone (200-1,000 m) (Lara *et al.*, 2009; Okamoto *et al*., 2019; Adl *et al.,* 2019; Fig. 3).

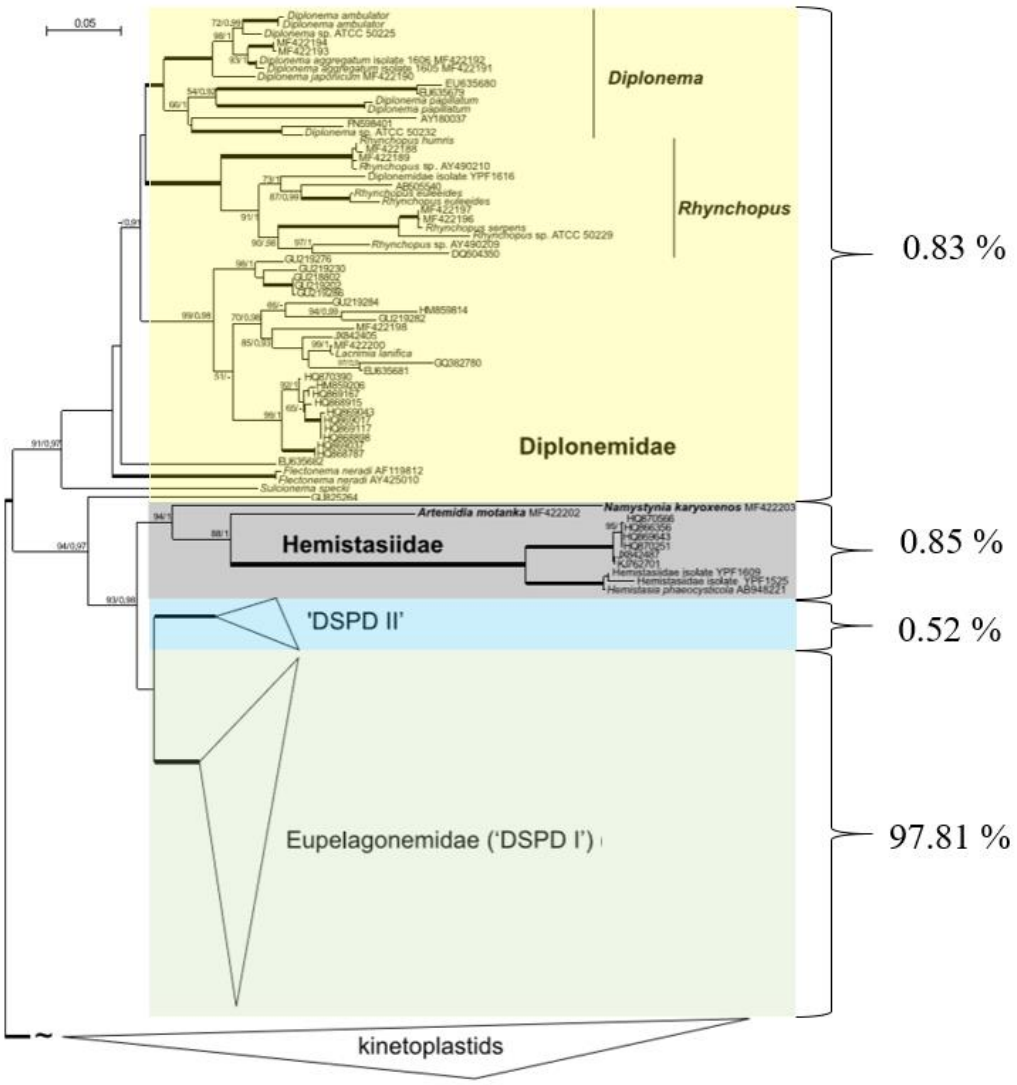


Figure 3: The phylogenetic tree of the four major groups of diplonemids.
The percentage next to each clade represent the percentage of the reads, which highlight the overwhelming abundance of the Eupelagonemidae (adapted from Prokopchuk et al., 2019 and Flegentova et al., 2016).

In contrast to the abundancy of *Eupelagonemidae* species and therefore expected significant importance for the marine ecosystem, no data regarding their ecological role currently exist as well as no species being available in culture yet (Flegontova *et al.*, 2016).

The only morphological data of *Eupelagonemidae* representatives were retrieved in the study of Gawryluk *et al.* (2016). In according study, ten diplonemids were visualized, displaying a range of shapes and sizes, which were obtained from the eastern North Pacific Ocean. As the result, these ten organisms were joined according to their appearance and the data of single-cell genomics (Fig. 4).

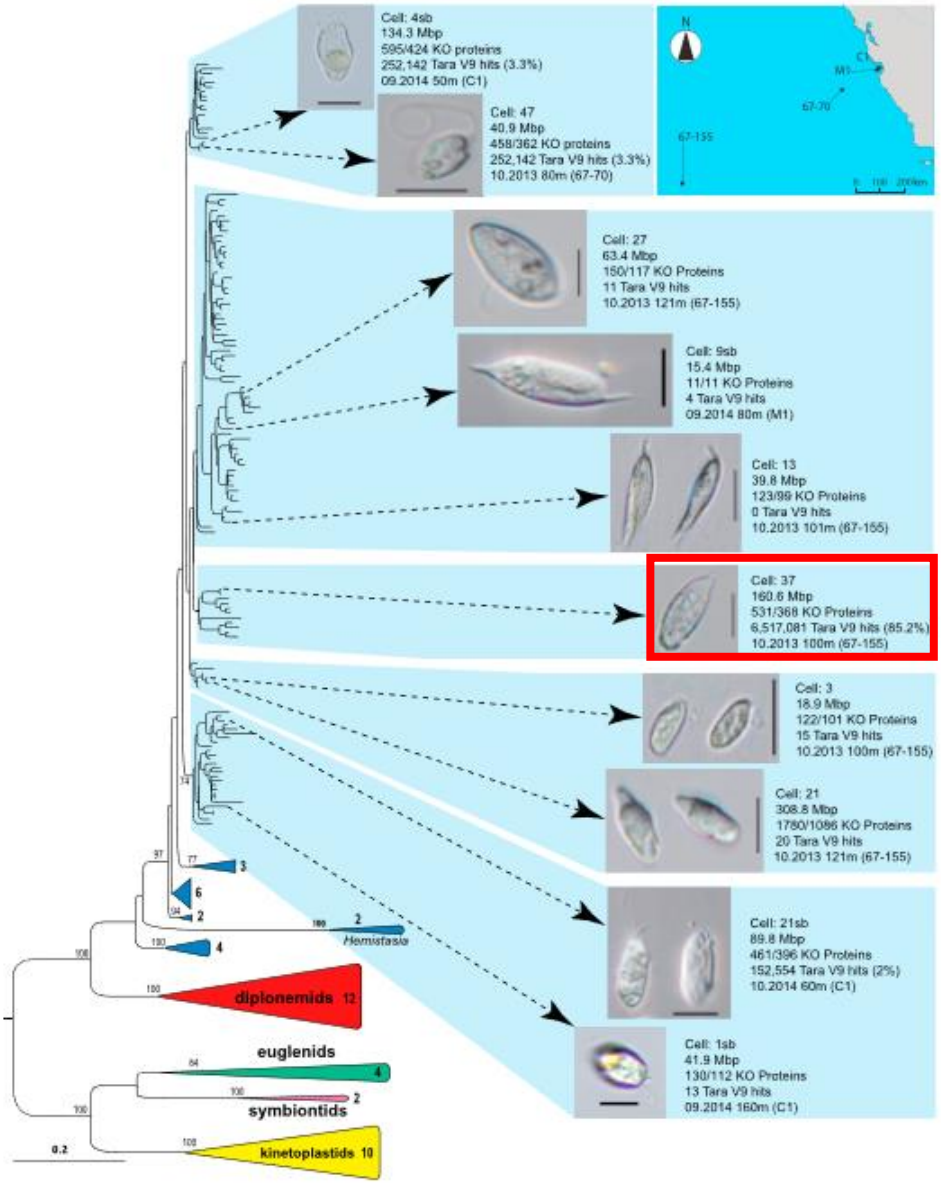


Figure 4:Scheme of the phylum Euglenozoa with emphasis on morphology and genomic characteristics of ten newly described Eupelagonemidae cells (adapted from Gawryluk et al., 2016).
Cell number 37 (shown in red frame) represents one of the most diverse heterotrophic eukaryotes in the ocean (more than 6.5 milions of hits to Tara Oceans V9 data and 85.2 % of total Discoba hits). The scale bars display 10 μm.

In contrast to this high number of marine diplonemids, they do not seem to be abundant in fresh waters, however, four OTUs belonging to *Diplonemea* were found in the freshwater Lake Baikal again using V9 region of 18S rDNA for sequencing (Yi *et al.*, 2017). Additionally, one OTU was identified in two out of six examined deep freshwater lakes in Japan using kinetoplastid-specific primers, while not being observable using universal eukaryote primers. This introduces further questions regarding possible misjudgement of the real content of diplonemids in environmental samples (Mukherjee *et al.*, 2019).

Even though diplonemids display a significant range in diversity and amount of species like numerous of other heterotrophic marine protist, they remain significantly understudied. This is due to the fact that these protists, and in particular diplonemids, propose a difficulty regarding their cultivation under laboratory conditions (Caron *et al.*, 2017, Carradec *et al.*, 2018).

1.3 *Diplonema papillatum*

The species of *Diplonema papillatum* (*D. papillatum*) was discovered on the surface of sea grass leaves in the sea around Friday Harbour in Washington and formerly was named *Isonema papillatum* (Porter, 1973). Its colourless, about 10-20 μm oblong shaped cell possesses two equal, subapical flagella.

The advantage of *D. papillatum* is that it is one of the few diplonemids that is possible to be cultured axenically in the laboratory and is available at the American Type Culture Collection under the code of ATCC 50162.

The cell of *D. papillatum* bears only a single mitochondrion with its DNA spread throughout, which is unusual for the phylum of Euglenozoa. Furthermore, mitochondrial cristae appear to be flat and of considerable size. Additionally, mitochondrial chromosomes accommodate split up genes, which are allocated on more than one chromosome (Marande *et al.*, 2005). Due to these discontinuous genes, but identified continuous transcripts, a sophisticated machinery is needed to make expression possible. This process was shown to include unusual trans-splicing and RNA-editing (Marande and Burger, 2007; Kiethega *et al.*, 2013; Fig. 5).

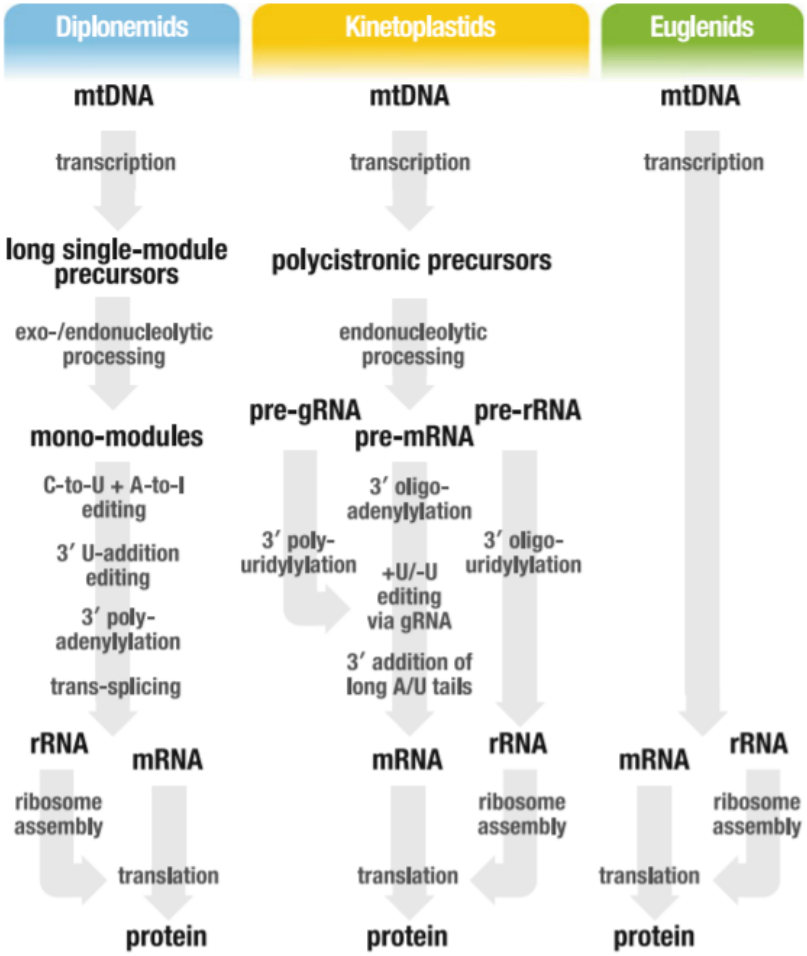


Figure 5: Different modes of gene expression within the Euglenozoa.
In contrast to euglenids, kinetoplastids and diplonemids show complicated posttranscriptional processing of their RNAs (Faktorová et al., 2018).

A remarkable feature of *D. papillatum* is the size of its mitochondrial genome, which accounts for approximately 260 x 10^6 base pairs (260 Mbp), therefore being larger than conventional content of genes in organelles and exceeding the amount of DNA (~180 Mbp) in its nucleus (Lukeš *et al*., 2018). While the mitochondrial genome of *D. papillatum* is exceptionally large, discovering the function and purpose of its discontinuous genes has been difficult. Recently, five *y* genes were discovered to code for ten subunits of the respiratory chain complex I (Valach *et al*., 2018).

Another interesting feature is the storage of energy by synthesis of a β-1,3-glucan polymer when *D. papillatum* is subjected to nutritionally poor media. Furthermore, it was observed that under these conditions, *D. papillatum* becomes highly motile, capable of active swimming as a way of stress reduction. It is possible that this transition is utilized to relocate themselves from regions with poor nutrient content to a more beneficial location as a form of survival mechanism (Škodová-Sveráková *et al.*, 2020).

1.4 Transformation of *Diplonema papillatum*

As *D. papillatum* can be cultured axenically in laboratory conditions, has been the most extensively studied to date and its genome having been sequenced, assembled and annotated, it was the species of choice for developing the transformation protocol that could allow understanding the role of these heterotrophic eukaryotes in the ocean.

The creation of a new model organism includes several steps (Fig. 6), starting from the selection of an organism, which requires the obtaining of basic data about the organism (Fig. 6A). Further three essential steps have to be fulfilled at the same time: 1/ determination of sensitivity of the organism to various antibiotics that can be further used as resistance markers (Fig. 6B), 2/ the suitable constructs have to be designed - either linear for incorporation into the genome, or circular, therefore to be kept in the cytoplasm (Fig. 6C) and 3/ the transformation technique for the introduction of this foreign DNA has to be chosen and tested (Fig. 6D). Finally, if the transformant cell lines are obtained, they have to be tested using numerous procedures in order to verify the validity of the established methodology (Fig. 6E).

Figure 6: The protocol for the establishment of a model organism, therefore serving as a ´transformation roadmap´ for researchers in order to establish a new genetically tractable organism (adapted from Faktorová et al., 2020).

Two studies attempted to establish a model organism from *D. papillatum* and used two different strategies: 1) N-terminal gene tagging of an α-tubulin gene with a fluorescent protein under the selection of resistance marker (Kaur *et al.*, 2018) and 2) inserting of the V5-tagged resistance marker into the genome of *D. papillatum* (Faktorová *et al*., 2020).

The study of Kaur *et al.* (2018) presented methodology for the transformation of the genome of *D. papillatum* and showed the possibility of genetical modification of this diplonemid species. In order to select appropriate genomic regions for integration of foreign

DNA, numerous parameters had to be fulfilled as the genome of *D. papillatum* bears excessive repetition (Gertraud Burger, unpublished data). Therefore, the region for targeting was chosen on the basis of being located in the annotated contig, having no introns but a significant expression level as well as being multicopy and bearing unique flanking regions.

As the nuclear genome of *D. papillatum* was not assigned yet at the start of the project, the region of choice was the α-tubulin gene, being a typical candidate for developing of model systems (Eichinger *et al.*, 1999). If the integration of the designed construct would be successful, its expression is visualisable by fluorescence microscopy as well as verifiable by Western blot analysis.

Puromycin was determined to be the antibiotic to which *D. papillatum* shows the highest sensitivity in comparison to the other tested antibiotics (Fig. 7). Therefore, the puromycin resistance gene (*Puro*^{R}) was selected as the first marker of choice.

Antibiotic	Concentration (μg/ml)
Puromycin	20
Blasticidin	50
Geneticin	75
Hygromycin	125
Nourseothricin	400
Phleomycin	500
Tetracycline	not sensitive

Figure 7: The sensitivity of D. papillatum to various antibiotics.
D. papillatum has been shown to be sensitive to six antibiotics, the highest sensitivity was found to puromycin. On the other hand, no response was observed in the case of tetracycline (Kaur et al. 2018).

The mCherry (red fluorescent tag) protein was selected for α-tubulin tagging. Additionally, a short Ty-tag bearing a start codon was added at the beginning of mCherry as another control, subsequently mCherry fluorescence gene was incorporated without a stop codon with the goal of enabling expression of the fused Ty-mCherry-α-tubulin protein.

The mCherry fluorescence tag and *Puro*^{R} gene are both framed with *D. papillatum* specific 5' and 3' untranslated regions (UTRs) and having 500 bp long homologous regions that should enable the integration in the correct position of *D. papillatum* genome. The whole, about 3 kbp long, construct (Fig. 8) was successfully electroporated, several transformant cells lines were obtained and the integration of the construct in *D. papillatum* genome was verified

by polymerase chain reaction (PCR) and Southern blot, however unfortunately not in the expected position.

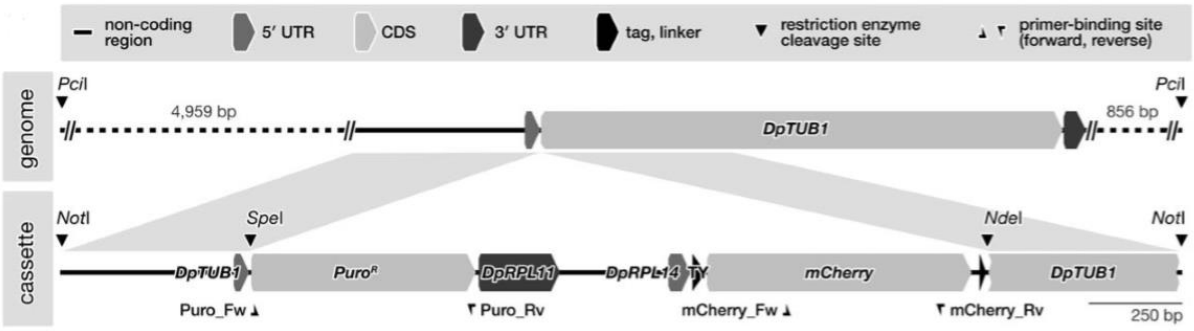


Figure 8: The scheme of N-terminal tagging of α-tubulin with Ty and mCherry tags and Puro^R resistance marker. The figure further shows the targeted region for the N-teminal tagging of the α-tubulin gene, here denoted as DpTUB1, which unfortunately was not achieved (adapted from Kaur et al. 2018).

This was found because the restriction digestion of transformant genomic DNA resulted in fragments smaller than expected, as well as PCR amplification using genomic DNA of the transformants as a template and primers outside the planned integrated position did not result in expected product size.

However, because of the polycistronic transcription fortunately being present in *Diplonema*, the construct is successfully transcribed. This was verified by isolation of RNA, its reverse transcription (RT) that resulted in the synthesis of complementary DNA (cDNA) and further PCR amplification on cDNA (so called RT-PCR) using specific primers for mCherry and *Puro^R*.

The construct is also properly post-transcriptionaly processed by addition of a spliced leader (SL) RNA sequence, which has a length of 39 nucleotides in diplonemids, and is added by trans-splicing on the 5' end of nuclear mRNAs (Sturm *et al.*, 2001; Kaur *et al.*, 2018). In this case cDNA was used as a template for so called SL RT-PCR (spliced leader RT-PCR), where SL-specific primers together with mCherry or Puro^R specific primers were used for a two stepped nested PCR reaction.

Even though transcription and post-transcriptional processing of both incorporated genes was confirmed, the translation of mCherry was not observed neither by Western blots (using anti-Ds red antibodies that should recognise mCherry), nor red fluorescence was observed by fluorescence microscopy.

In contrast, *Puro*^{R} gene is correctly translated as confirmed by Western blot analysis using anti-*Puro*^{R} antibodies (Kaur *et al.*, 2018; Fig. 9).

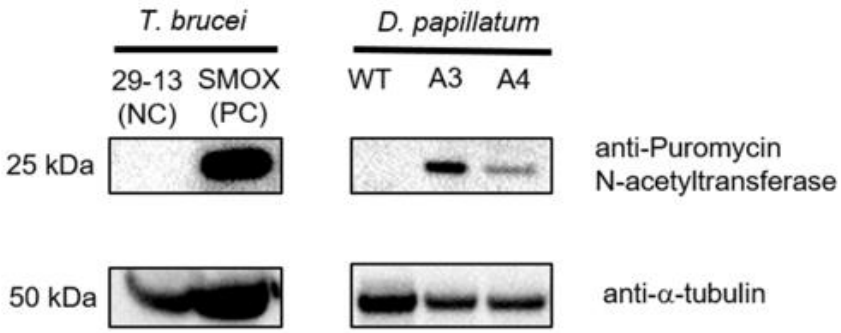


Figure 9: Confirmation of the Puro^{R} *gene expression in transformant cell lines A3 and A4 using Western blot analysis. In contrast to D. papillatum WT, the transformant cell lines A3 and A4 expressed the Puro*^{R}*. T. brucei cells were used as positive (PC) and negative (NC) controls (Kaur et al., 2018).*

In the second study, the wild-type (WT) of *D. papillatum* was electroporated with a construct consisting of the V5-tagged neomycin resistance gene (*Neo*^{R}), framed by partial regulatory sequences from the related kinetoplastid *Blastocrithidia* (strain p57). Seven cell lines showing resistance to neomycin were obtained and further examination focussed on approving transcription and expression of the V5-*Neo*^{R} protein, therefore RT-PCRs were performed, followed by Western blot analysis. These tests lead to confirmation of the expression of the tagged *Neo*^{R} protein in two cell lines (Fig. 10).

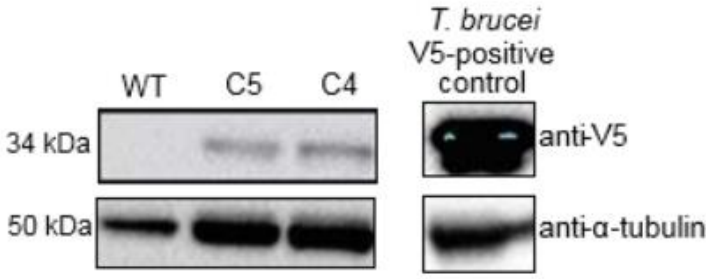


Figure 10: Verification of V5-Neo^R gene expression in transformant cell lines C4 and C5. The Western blot analysis using antibody against the V5 tag showed a signal in cell line C4 and C5 in contrast to D. papillatum WT. T. brucei cell line was used as a positive control (Faktorová et al., 2020).

When combining the results of Kaur *et al.* (2018) and Faktorová *et al.* (2020), it is clearly evident that the transformation of *D. papillatum* is possible and the results were promising, with the only complication seeming to be the random integration of the constructs into *D. papillatum* genome.

1.5 Targeted integration of designed constructs into the genome of *D. papillatum*

Because of the above-mentioned results, the remaining question was how to achieve the correct integration of the designed construct into the planned position in the genome as it is of considerable importance to enable tagging and knock-out of any *D. papillatum* genes.

There were two hypotheses for the failed integration:

1/ the machinery for homologous recombination (HR) in *D. papillatum* could be less efficient, therefore non-homologous end-joining (NHEJ) would serve as the main DNA repair/recombination pathway

2/ it might result from the fact that the genome of *D. papillatum* is very repetitive, so that the regions of homology needed to be extended

1/ Homologous recombination (HR) is utilized in order to correct double strand breaks with its strand exchange proteins (e.g. Rad50-Rad52) and its replication proteins (e.g. Rpa1-Rpa3) (Krejci *et al.*, 2012; Son and Hasty, 2019). On the other hand, non-homologous end-joining (NHEJ) with some of the core factors being ligase (LIG4) and the Ku70/80 heterodimer (XRCC5, XRCC6) is another possibility to fix such breaks (Waters *et al*., 2014).

Research on the genome showed that both these two pathways are present in *D. papillatum* (Gertraud Burger, unpublished data) and both are aiming to repair double strand breaks (Fig. 11). Failed integration of a designed construct by electroporation could be therefore explained by the fact that the NHEJ pathway is the main DNA repair/recombination pathway in *D. papillatum.* This hypothesis was tested and was not proven to be right (details stated in- Faktorová *et al.*, submitted).

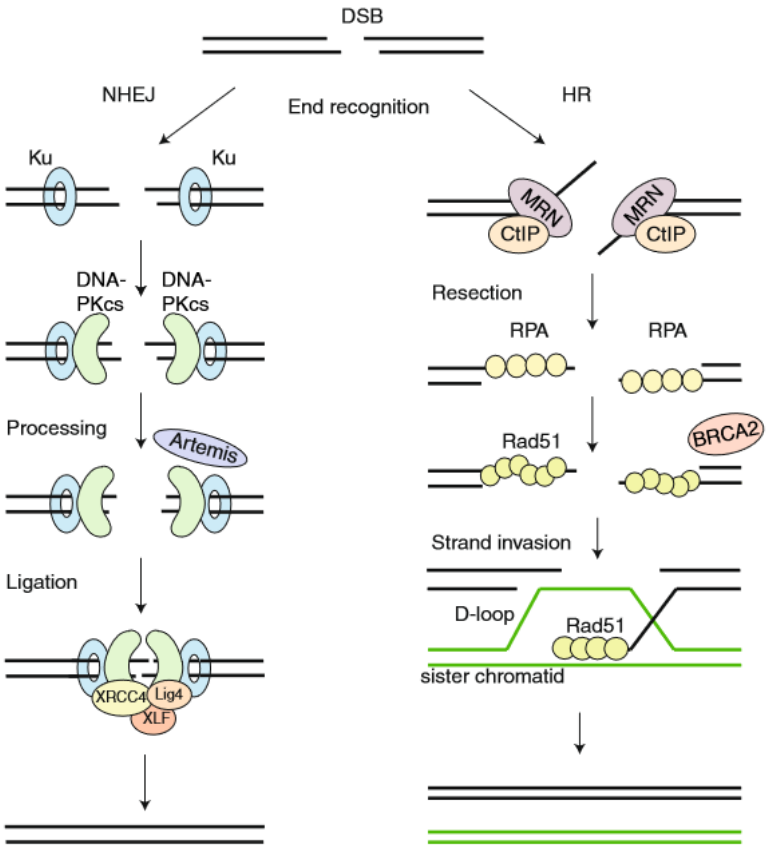


Figure 11: Repairment of a double strand break using HR or NHEJ pathway.
As HR and NHEJ are both mechanisms of coping with a double strand break, they are assumed to be competing and their functionality to be dependent on the conditions (Brandsma and Gent, 2012).

2/ Another possibility to explain random incorporation that a significant portion of the genome consists of repetitive sequences, accounting for approximately 60% (Gertraud Burger, unpublished data). Consequently, the plan was to increase the probability of the integration into the right position by extension of the 5′ and 3′ homologous arms of the construct from previous 500 bp to 1000-2000 bp.

The results that were obtained by testing of both hypotheses in our laboratory were summarised in a manuscript only recently submitted to *Environmental Microbiology* journal and being attached to this thesis as supplementary data in the appendix (Faktorová *et al.*, submitted).

I participated during the initial stages of the study when the main aim was to verify whether increasing of homologous arms of the constructs will lead to the integration into the correct position in the genome of *D. papillatum*. This is important as it is a crucial step to successful transformation enables to tag/knock-out any of *D. papillatum* gene and allow their functional analysis.

Initially, two constructs were designed:

1/ α-tubulin replacement with a cassette containing *Puro*^{R} (Fig. 12, construct #1)

2/ N-terminal tagging of α-tubulin using the *Puro*^{R}-mCherry cassette, formerly being called *DF_Dp_01* (Kaur *et al.*, 2018), with enlarged homologous arms (Fig. 12, construct #3) (Faktorová *et al.*, submitted).

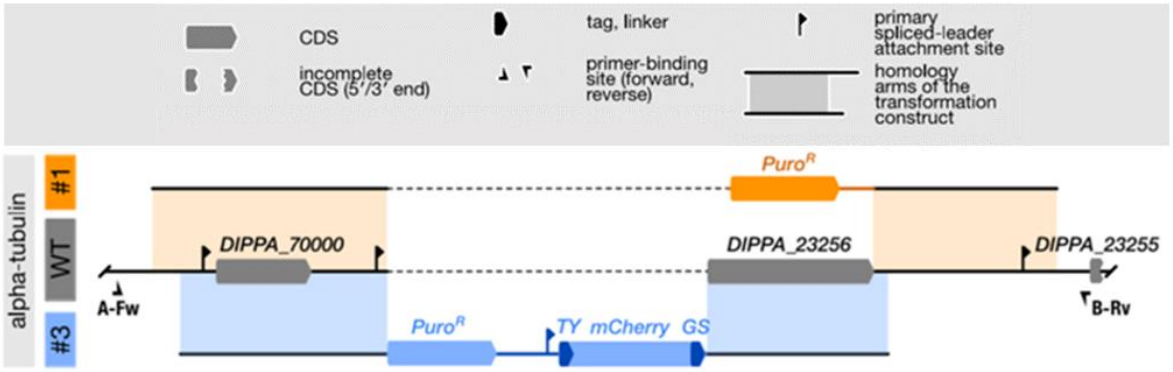


Figure 12: Replacement of the α-tubulin gene and N-terminal tagging under puromycin selection. Construct #1 for α-tubulin replacement is shown in orange highlight, while construct #3 for N-terminal tagging of α-tubulin is highlighted in blue (adapted from Faktorová et al., submitted).

The cell lines examined in this study were generated in the lab earlier and obtained from Drahomíra Faktorová. The following eight transformant cell lines were tested:

1/ replacing of α-tubulin with puromycin resistance gene (*Puro*^{R})

– three cell lines (number 8, 9, 10)

2/ N-terminal tagging of α-tubulin with mCherry fluorescent protein under puromycin selection

– five cell lines (number 13, 14, 15, 16, 24)

2 Aim of the thesis

The aim of this thesis was to analyse and characterize selected cell lines of the marine protist *Diplonema papillatum* that were obtained after transformation of wild-type (WT) using two different constructs with the plan of 1/ replacing (knock-out) or 2/ endogenous N-terminal tagging of one α-tubulin gene (*DIPPA_23256*).

The specific aims of the thesis:

1/ To isolate the genomic DNAs from the obtained cell lines and verify by PCR whether they contain the electroporated constructs

2/ To select and further characterize the cell lines containing constructs integrated in the expected position in the genomic DNA

3/ To proof whether the integrated genes are transcribed, properly post-transcriptionally processed (by SL RT-PCR) and translated (by Western blot analysis and fluorescence microscopy)

4/ To test the viability of the selected cell lines by measuring the growth curves

5/ To examine the selected cell lines by electron microscopy in order to identify if any alterations due to the replacement of α-tubulin gene are observable

3 Materials and methods

3.1 Used cell lines and cultivation

The cultures of *Diplonema papillatum* were cultivated in tissue culture flasks (TPP) in a sterile artificial sea salt liquid medium at 27°C.

	Diplonema cultivation medium
Sea salts (Sigma, S9883)	36 g
Tryptone	1 g
Fetal bovine serum	10 ml
Chloramphenicol (40 mg/ml)	2.5 ml (final concentration 100 µg/ml)
Distilled water	x
∑	1 l

Puromycin antibiotic (20 µl/ml) was added to the media for cultivation of the transformant cell lines 8, 9, 10 (construct #1), cell lines 13, 14, 15, 16, 24 (construct #3) as well as the cell line A3 obtained in previous study from Kaur *et al*. (2018) (to keep the selection as they express the puromycin resistance gene - $Puro^R$).

3.2 Cryoprotection of the cells

D. papillatum cells were frozen using cryoprotection. In details, 800 µl of the cells in the exponential phase of growth were mixed with 200 µl of 50 % glycerol (sterilized using 0.22µm filter) in the labelled cryotubes. Subsequently, they were placed in the freezing container (Mr. Frosty, Nalgene) and transferred to a –80° C freezer. The cryo-container was filled with isopropanol, which cools the sample slowly at a rate of 1° C/min. If the cell lines were not needed during the next days, the frozen tubes were transferred to liquid nitrogen to be preserved over longer periods of time.

To defreeze frozen cell cultures, the desired tubes were removed from the liquid nitrogen container and thawed at room temperature. The sample was transferred to a cultivation flask containing 5 ml of media with or without puromycin based on the defrost cell culture. The cells were then grown in an incubator at 27° C.

3.3 Isolation of genomic DNA

The genomic DNA (gDNA) from 10 ml (about $2x10^7$ cells) of WT and transformant cell lines was isolated using a DNA isolation kit (Qiagen, 69504).

3.4 Polymerase chain reaction (PCR)

PCR was used to verify the correct integration of the constructs in the genomic DNA of *D. papillatum* transformants as well as for amplification of *Puro*^{R} gene. WT and cell line A3 (from previous study) were used as controls.

PCR amplification was performed using Phusion polymerase (NEB Biolabs, M0486S) and following PCR mixture and program:

PCR mix:

Reaction component	**25 µl reaction**
5x Phusion HF buffer	5 µl
10 mM dNTPs	0.5 µl
20 µM Fw primer	0.6 µl
20 µM Rv primer	0.6 µl
DMSO	0.75 µl
Phusion DNA polymerase	0.25 µl
gDNA (20ng/µl)	1 µl
milliQ water	16.3 µl

PCR program:

1. Initial denaturation	98° C	30”
2. Denaturation	98° C	10”
3. Annealing	67° C	30”
4. Extension	72° C	30 seconds/kb*
5. Final extension	72° C	5’
6. Hold	12° C	∞

35 cycles (repetition of steps 2-4) of PCR amplification were used.

* the length of extension was used based on the primer combinations

The products of the PCRs were separated using gel electrophoresis. Obtained amplicons were subsequently extracted from the gels and verified by sequencing (Eurofins Genomics).

List of primers used for gDNA PCR:

Primer name	Primer sequence (5' to 3')
S-Fw	TACAAGAAATTGAAGAACGATTCACTGGTAG
S-Rv	TGTAGACTGTTTCTGTTTGTTTGTTTCTTTC
W-Fw	GAACGTTTCTCGGTTTGATTCGCACAAAACT
W-Rv	ACATTCCTACCGTTCAGAAAGAGGGAGGAT
A-Fw	TCAGGTTGCCGGCATTTGGGAGCACAATCAG
B-Rv	ACCGGCTACCACCTACTCCCGCTGCTTTATGTG
Puro-Fw	ATGACCGAGTACAAGCCCAC
Puro-Rv	TCAGGCACCGGGGCTTGC
3LHR Fw1*	TAGGAATGTCTCTCGTTTTCTTTTG
3LHR Fw3*	GTGTAAAGGCAGCAATACATGTTC
3LHR Fw4*	AGGTTAGTCACTGTTCCTCGTGTAG

*These additional primers were designed for sequencing verification for the purpose of covering the whole region of the PCR products

3.5 Reverse transcription

In order to the verify whether the constructs integrated in cell lines 9 and 15 are correctly transcribed and posttranscriptionally processed, RNA was isolated and used as a template for reverse transcription resulting in the synthesis of complementary DNA (cDNA).

3.5.1 RNA isolation

Chemicals used:

	RNaseZAP (Ambion; AM9780)
	TriReagent (MRC; TR118)
	Chloroform
	Isopropanol
	EtOH (75 %)

Procedure:

1. The work bench, pipettes and gloves were cleaned with RNaseZap to remove any RNAses.
2. Approximately 1 x 10^8 cells were spinned down, the pellet was taken and resuspended in 1 ml of TriReagent. This mixture was left at room temperature for five minutes to ensure proper dissociation of nucleoprotein complexes.
3. 0.2 ml chloroform was added, the tube was shaken vigorously for 15 seconds and afterwards left 2 - 15 minutes at room temperature.
4. The solution was then centrifuged at 12000xg for 15 minutes at 4° C.
5. The aqueous phase was taken and separated from the organic phase into a new tube. 0.5 ml of isopropanol was added, the solution was incubated for 10 minutes at room temperature and spinned down at 12000xg for 10 minutes at 4° C. Afterwards, the supernatant was discarded.
6. 1 ml of 75 % EtOH was added, the mixture was shaken for 5 seconds and spinned down at 7500xg for 5 minutes at 4 ° C. The supernatant was again removed.
7. The pellet was then left to dry at room temperature for 5-15 minutes, where the tube should be lying.
8. The sample was then dissolved in 50 µl of 60°C warm RNase-free water. Next, the concentration can be measured and subsequently stored at – 80° C.

3.5.2 Reverse transcription and complementary DNA (cDNA) synthesis

The QuantiTect Reverse Transcription Kit (Quiagen, 205311) was used for cDNA amplification with random primers.

cDNA preparation:

Reaction component	Volume per reaction
gDNA Wipeout Buffer, 7x	2 µl
Template RNA (up to 1 µg including all types of RNA)	y
RNase-free water	x
∑	14 µl

Reaction component	Volume per reaction
Quantiscript Reverse Transcriptase (contains RNase inhibitor)	1 µl
Quantiscript RT buffer, 5 x (contains Mg^{2+} and dNTPs)	4 µl
RT Primer Mix	1 µl
Entire genomic DNA elimination reaction	14 µl
∑	20 µl

Procedure:

1. The RNA sample, gDNA Wipeout Buffer, Quantiscript Reverse Transcriptase, Quantiscript RT Buffer, RT Primer Mix and RNAse-free water were thawed, mixed, centrifuged and kept on ice.
2. For each sample, namely WT, A3, 9 and 15, two reactions were prepared based on the QuantiTect Reverse Transcription protocol – first set with reverse transcriptase (RT+), the second without (RT-). RT- reactions served as negative controls.

3.5.3 Spliced leader reverse transcription PCR (SL RT-PCR)

To check that the spliced-leader (SL) sequence was added (trans-spliced) at the 5´ end of the transcripts, SL RT-PCR using OneTaq polymerase (NEB Biolabs, M0486L) was performed.

From cDNA, the particular regions (5´parts of $Puro^R$, mCherry and mCherry + α-tubulin) were amplified by two sets of PCRs (nested PCR) using primers derived from the SL RNA gene (DpSL_Fw1 and DpSL_Fw2 primers) in combination with $Puro^R$ (SL_Puro_Rv1; SL_Puro_Rv2), mCherry (SL_mCherry_Rv1; SL_mCherry_Rv2) or α-

tubulin (SL_Atubulin_Rv1 and SL_Atubulin_Rv2) specific primers, respectively. Forward primers are denoted as “Fw primer” and the reverse primers as “Rv primer”.

The products of the PCR reactions were separated using gel electrophoresis. Obtained amplicons, containing the SL and N-terminal parts of *Puro*^{R}, mCherry or mCherry + α-tubulin were the extracted from the gel and verified by sequencing (Eurofins Genomics).

First PCR:

Reaction component	**1 x MMix**
cDNA	0.5 µl
20 µM Fw1 primer	0.5 µl
20 µM Rv1 primer	0.5 µl
2X Mastermix One *Taq* Quick-Load	12.5 µl
Distilled water	11 µl

PCR program:

1. Initial denaturation	94° C	30”
2. Denaturation	94° C	30”
3. Annealing	58° C	30”
4. Extension	68° C	1’ 30”
5. Final Extension	68° C	5’
6. Hold	14° C	∞

The steps of denaturation until extension are repeated for 30 cycles.

Second PCR:

For the second PCR, the products of the first PCR were used for specific amplification of the selected regions using the same program.

Reaction component	**1 x MMix**
First PCR reaction	0.5 µl
20 µM Fw2 primer	0.5 µl
20 µM Rv2 primer	0.5 µl
2X Mastermix	12.5 µl
Distilled water	11 µl

List of primers used for SL RT-PCR:

Primer name	Primer sequence (5’ to 3’)
DpSL_Fw1	CCAACGATTTAAAAGCTACAGTTTCT
DpSL_Fw2	AAAAGCTACAGTTTCTGTACTTTATTG
SL_Puro_Rv1	GCTCGTAGAAGGGGAGGTTG
SL_Puro_Rv2	CGTGAGGAAGAGTTCTTGCAG
SL_mCherry_Rv1	CTTCAGCTTCAGCCTCTGCT
SL_mCherry_Rv2	AAGCGCATGAACTCCTTGAT
SL_Atubulin_Rv1	AACAGAGCTCCAAGACCAGAAC
SL_Atubulin_Rv2	GTCTCAGAGAAGAAGGTGTTGTAGG

3.6 Gel-electrophoresis

Chemicals used:

Components	EtBr-gel
Agarose	800 mg
1 x TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA)	100 ml
Ethidium Bromide (EtBr)	1 µl

Procedure:

1. For the gel, agarose and 1 x TAE buffer were mixed according to the table above in an Erlenmeyer flask and heated in a microwave until the agarose is melted.
2. After it was cooled down to approximately 50 °C, EtBr was added and the solution is mixed carefully.
3. The electrophoretic chamber was assembled and the EtBr-gel was poured. Bubbles should be removed using a pipette tip.
4. The combs were chosen according to the use of the gel and put into the liquid gel.
5. After the gel is solid, 1 x TAE buffer is added to the chamber until it reaches the mark. The combs were removed, the wells can be loaded and run approximately at 90 V.

3.7 Gel extraction

The Expin Gel SV Kit (GeneAll, 102-150) was used for gel extraction.

Procedure:

1. The gel was placed on the transilluminator and DNA bands of interest were cut out from the gel with an EtOH-cleaned scalpel.
2. The gel pieces of each band were placed in different tubes and weighed.
3. For each volume of gel, three volumes of GB buffer were added.
4. The tubes are incubated at 50° C until the gel is melted, it is helpful to vortex the tubes each 2-3 minutes.
5. If the solution stayed yellow, one gel volume of isopropanol is added. If not, 10 µl of 3M sodium acetate (pH 5.0) should be added before.
6. These mixtures were transferred to SV columns and centrifuged for 1 minute at full speed. The flow-through was discarded. Subsequently, the SV tube was placed in the collection tube.
7. 700 µl of Buffer NW were added and the tubes were centrifuged for 30 seconds at full speed. The flow-through was discarded, the SV was reinserted in the collection tube.
8. The tubes were again centrifuged for 1 minute at full speed to get rid of any wash buffer left. The SV tubes were then transferred to new 1.5 ml tubes.
9. 25 µl of 70° C distilled water was added to the middle of the membranes, left to stand for 1 minute and centrifuged for 1 minute at full speed.

3.8 Western blot

Western blots were performed to analyse the expression of construct targeted into cell line 9 and cell line 15 on the protein level.

3.8.1 Running of the SDS-PAGE gel

Chemicals used:

	1x Running Buffer
Tris	25 mM
Glycine	192 mM
SDS	0.1 %

	Blotting Buffer
Methanol	20% (Vol.)
Glycine	38.6 mM
Tris	48 mM
SDS	1.3 mM

The cell lysates were prepared by resuspending of $5x10^5$ *D. papillatum* cells in 25 µl of 2x Laemli SDS sample buffer. NuPAGE 4-12 % Bis-Tris gels (Invitrogen, NP0322BOX) were used for separation of the samples.

Procedure:

1. The gels were taken out of the package and placed into the electrophoretic chamber, which was subsequently filled with 1x Running Buffer.
2. The wells were washed with a syringe before loading of the individual samples and 5 µl protein ladder (Precision Plus Protein Standards; Bio-Rad, 161-0373).
3. The gel was run at 100 V until approximately half of the gel were exceeded. Subsequently, the voltage was increased to 150 V.

3.8.2 Transfer of the proteins on a membrane

For the transfer of the proteins on the membrane, the wet blotting technique was chosen.

Procedure:

1. For each gel, one PVDF membrane (GE Healthcare; 10600023) and two thick filter paper sheets were cut out. The membrane was soaked in methanol to be activated. Additionally the filter papers should be soaked in blotting buffer to ensure that the "blotting sandwich" does not dry out.
2. The "blotting sandwich" was assembled as following: white panel, sponge, filter paper, membrane, SDS-PAGE gel, filter paper, sponge, black panel. Before closing the panels of the "blotting sandwich", the bubbles are removed.
3. The "blotting sandwich" were transferred into the blotting apparatus, a magnetic stirrer and a cooling pack is added together with the blotting buffer.
4. The chamber was closed, placed onto a stirrer apparatus, placed in a fridge and run at 100 V for 1 1/2 hours.

3.8.3 Treatment of the membrane

Chemicals used:

	Milk solution
dry nonfat milk	10 g
1 x PBS (0.05 % Tween-20)	200 ml

List of antibodies used for Western blot analysis:

Name	Primary or secondary	Animal host	Company	Ordering Number	Concen-tration
anti-puromycin N-acetyltransferase	primary	rabbit	Thermo-Fisher Scientific	702389	1:500
Anti-Ds Red	primary	rabbit	Clontech	632496	1:1,000
Anti-Ty	primary	mouse	Sigma-Aldrich	SAB4800032	1:1,000
Anti-α-tubulin	primary	mouse	Sigma-Aldrich	T9026	1:5,000
Anti-enolase	primary	rabbit	gift from Jorge Morales		1:2,000
Anti-rabbit	secondary	/	Sigma-Aldrich	A0545	1:1,000
Anti-mouse	secondary	/	Sigma-Aldrich	A9044	1:1,000

The Clarity Western ECL Substrate (Bio-Rad), which contains luminol and peroxide was used for the treatment of the membrane.

Procedure:

1. The membrane is carefully taken with tweezers on one edge and put in a 50 ml Falcon tube with the side with the proteins facing the inside of the tube.
2. 3 ml of milk together with the primary antibody (the details about the concentration used is stated in the table above) was added to the tube and incubated on a roller shaker at room temperature for 2 hours or in the fridge overnight.
3. The milk solution was discarded and the membrane was washed five times for five minutes with 1x PBS-Tween.
4. Subsequently, the 3 ml of milk solution together with the secondary antibody was added to the tube, incubated and washed as before.
5. A transparent film was cut so that the membrane fits in there. The membrane is carefully placed in this film with tweezers and further touching of the membrane should be avoided at all time.
6. 500 µl of Luminol and 500 µl Peroxide from the Clarity Western ECL Substrate Kit were mixed in a tube and pipetted onto the membrane. The film is closed and the solution is spread evenly so that the whole membrane is soaked.
7. After an incubation, the liquid was removed and the membrane was developed using Chemidoc MP imager (Bio-Rad) at different exposure times.

3.9 Fluorescence microscopy

With fluorescence microscopy, the occurrence of fluorescence of WT of *D. papillatum* and cell line 15 was analysed and compared. In brief, 5 µl of living cells of cell line 15 and WT were placed on a slide, respectively, covered with a coverslip and left for 2-5 minutes until the organisms stopped moving while still being alive. The slides were examined with an AxioPlan 2 imaging fluorescence microscope (Zeiss) and a video was recorded. The desired images were obtained using Media Player Classic program and processed with the Gimp 2.8.8 software.

3.10 Growth curves

To investigate the effect of the genome transformation on the growth pattern of cell line A3, 9 and 15, growth curves were performed. The density of the transformed cell lines and a wild type culture of *D. papillatum* were examined for 10 subsequent days. The cultures were kept in the logarithmic phase, thus they were diluted with the media either with or without puromycin, if the concentration of $1x10^6$ was considerably exceeded. To provide the means for accurate counting of the cells, the counting-sample of the three different cultures of *D. papillatum* is mixed with a cell fixation solution. This ensures that the cells are immobile by killing them without destroying their cell structure.

Chemicals used:

	Cell fixation solution
3,7 % Formaldehyde	5 ml
20 x SSC	2.5 ml
Distilled water	x
∑	50 ml

Procedure:

1. From the cultures of WT and cell line A3, 9 and 15 samples were taken and diluted to $1x10^6$ with the accurate type of media.
2. After 24 hours each in the span of the experiment, the amount of the cells in the cultures was ascertained by counting three samples of each type of the organisms to be tested with the Countess II FL Automated Cell Counter (ThermoFisher Scientific, AMQAF1000).

3. If the concentration of the cells exceeded $1x10^6$ considerably, the amount of culture and media needed was calculated and the density was reduced to $1x10^6$ again.

3.11 Electron microscopy

The transmission electron microscopy techniques was used to compare the ultrastructure of *D. papillatum* WT and cell line 9. The samples were prepared by high pressure freezing and observed using JEOL 1010 TEM microscope and the images were captured with an Olympus Mega View III camera.

4 Results

4.1 Isolation of the genomic DNA from the obtained cell lines and verification whether they contain the electroporated constructs

As mentioned in the Introduction, eight cell lines were obtained after electroporation of the following constructs in *D. papillatum* WT cells:

1/ replacing of α-tubulin with puromycin resistance gene (*Puro*^{R}, construct #1)

– **three cell lines (named as 8, 9, 10)**

2/ N-terminal tagging of α-tubulin with mCherry fluorescent protein under puromycin selection (construct #3)

– **five cell lines (named as 13, 14, 15, 16, 24)**

For details of the constructs including the position of the primers - see Fig. 13 and for the expected sizes of PCR products, see the table below.

Expected sizes of individual amplicons for WT and tested cell lines (in bp):

Primers combinations	**WT**	**A3**	**9**	**15**
S-Fw and S-Rv	4,750	4,750	4,223	not tested
W-Fw and W-Rv	3,244	3,244	not tested	5,166
Puro-Fw + Puro-Rv	no product	600	600	600
A-Fw and B-Rv	5,213	5,213	4,720	7,178
Puro-Fw + B-Rv	no product	no product	2664	5,162

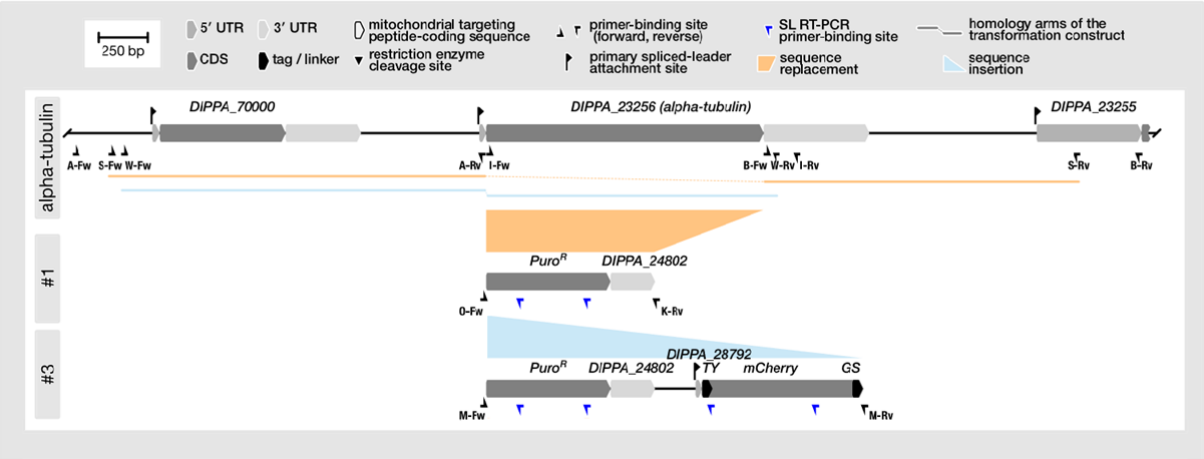


Figure 13: Designed constructs for α-tubulin replacement (construct #1) and N-terminal tagging (construct #3). The α-tubulin gene to be replaced or N-terminally tagged, here called DIPPA_23256, is shown together with neighboring situated genes up- and downstream (adapted from Faktorová et al., submitted).

4.1.1 Isolation of genomic DNA

The genomic DNA from *D. papillatum* WT, cell line A3 (obtained in previous study Kaur *et al.*, 2018, here used as a control) was isolated together with genomic DNA of eight tested transformant cell lines.

4.1.2 Verification of the integration of the construct in *D. papillatum* genome

To check whether the cell lines contain the electroporated constructs, PCR with primers used for the construct preparation were performed. The primer combination S-Fw + S-Rv, which was designed for the demonstration of the presence of construct #1 (Fig. 4A) and the primer combination W-Fw + W-Rv, designed for visualizing the presence of construct #3 (Fig. 4B), lead to positive results in few of the cell lines. The isolated gDNA served as a template for this validation.

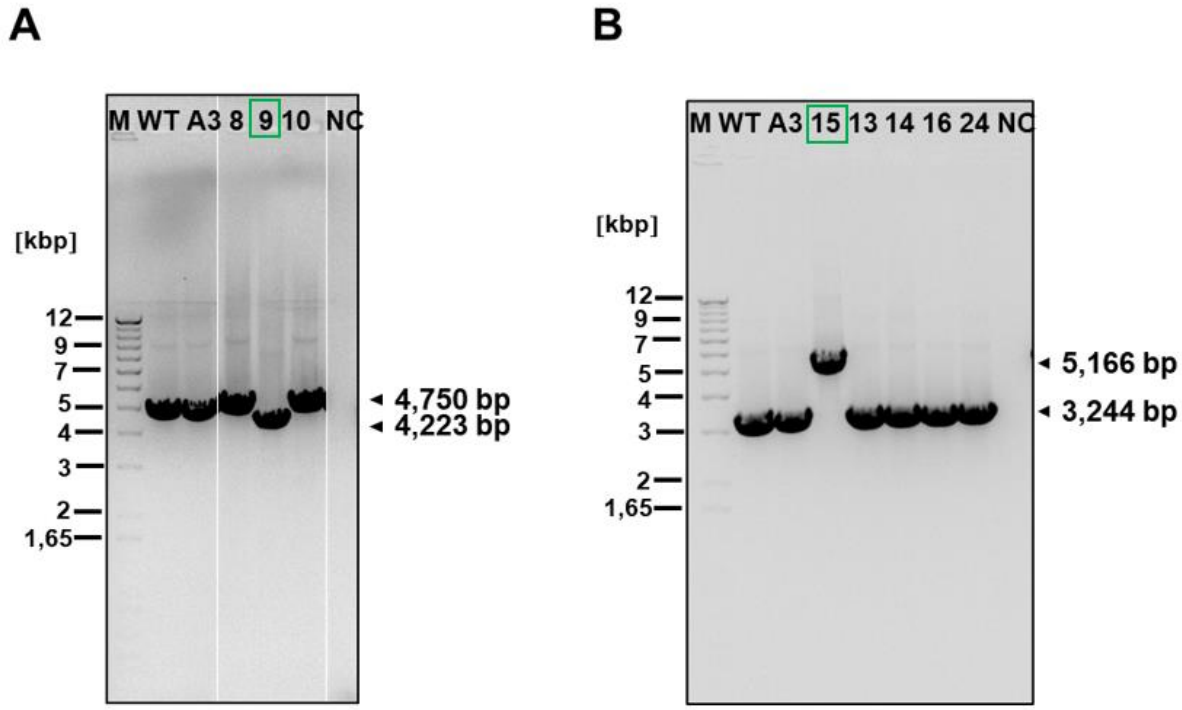


Figure 14: PCR verification of α-tubulin replacement (A) and N-terminal tagging (B) constructs in the D. papillatum genome, respectively.
(A) Cell line 9 showed a clearly visible shift (4,223 bp) compared to the controls WT and A3, therefore verifying the replacement of α-tubulin. (B) Cell line 15 displayed a significant increase in size (5,166 bp) and thus validating the insertion of construct #3.

The insertion of the respective construct into the target site (manifested by a shift of the band on DNA gel) was verified in two out of eight tested cell lines, namely one out of three obtained from α-tubulin replacement - **clone 9** and one out of five obtained from α-tubulin tagging cell lines – **clone 15**. These cell lines were selected and tested further.

4.2 Selection and characterization of the cell lines containing constructs integrated in the expected position in the genomic DNA

The presence of the respective constructs in cell lines 9 and 15 was further verified by PCR using various combination of primers.

4.2.1 PCR using Puro-Fw + Puro-Rv primers

First, the presence of *Puro*^{R} gene was verified in cell lines 9, 15 and A3 (used in previous study from Kaur *et al.*, 2018). The genomic DNA of WT was used as a control (Fig. 15).

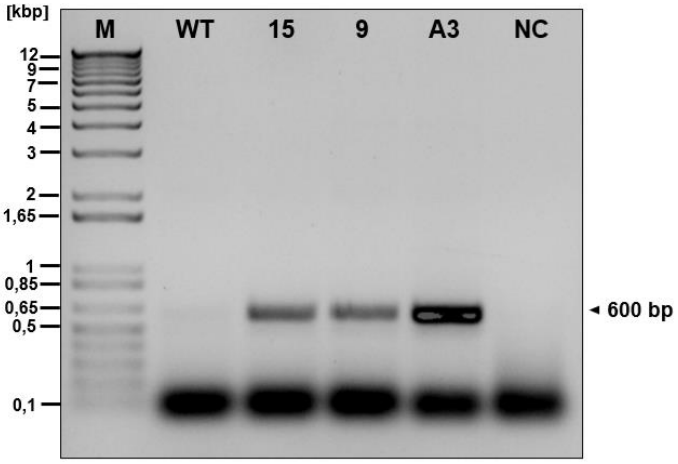


Figure 15: PCR verification of presence of Puro^{R} *gene in studied cell lines.*
While there are clear bands of 600 bp visible in cell line 15, 9 and A3, WT on the other hand showed no signal in that size region. Because the band size is equal in all three cell lines and A3 was the positive control, it can be assumed that the Puro^{R} *gene is present in cell line 15 and 9.*

4.2.2 PCR using A-Fw + B-Rv and Puro-Fw + B-Rv primers

To double-check whether the constructs were really correctly targeted into the intended positions of *D. papillatum* genome, PCR with both primers outside the electroporated construct (A-Fw + B-Rv; Fig. 16A) and combination of one primer inside the construct and one outside (Puro-Fw + B-Rv; Fig. 16B) were used:

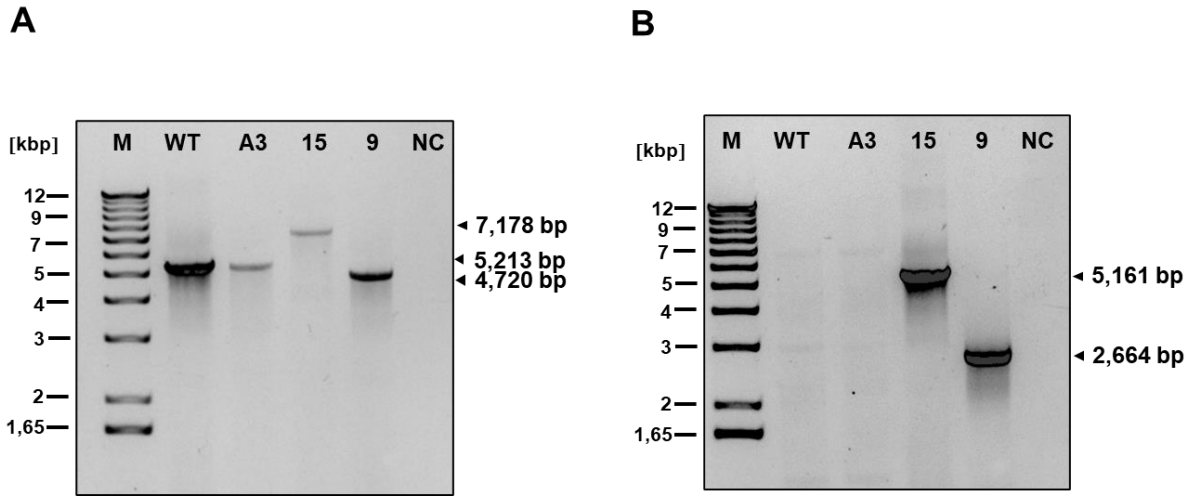


Figure 16: PCR verification of the correct integration of the constructs in D. papillatum genome in studied cell lines. (A) Examination with primers outside of the construct indicates the presence of respective constructs in cell lines 15 and 9 as the size of the product for cell line 15 is significantly larger (7,178 bp) and the size of the product for cell line 9 significantly smaller (4,720 bp) than the size of the band for the unaltered regions of α-tubulin in A3 and WT (5,213 bp). (B) Using one primer inside and one primer outside the intended region of the construct resulted in no signal in WT and A3, while displaying the presence of respective constructs of expected sizes in cell line 15 (5,161 bp) and 9 (2,664 bp), respectively.

All results of PCR reactions show the expected sizes of the PCR products and therefore verifies the correct integration of both constructs in cell lines 9 and 15.

4.3 Verification of expression of the constructs

4.3.1 Transcription and post-transcriptional processing

The correct transcription and post-transcriptional processing of the integrated constructs were verified using spliced leader reverse transcription PCR (SL RT-PCR) that should confirm that the SL RNA sequence was added at the 5' end of particular mRNAs – *Puro*^{R} (Fig. 17A), mCherry (Fig. 17B), and mCherry-tubulin (Fig. 17C).

To perform this experiment, RNA was first isolated from the cultures, then reverse transcribed to complementary DNA (cDNA) that was further used as a template for a 2 step PCR reaction (see Materials and Methods). For expected sizes of the PCR products see the table below, where primers which are not applicable for the respective cell line are denoted as "N/A".

Expected sizes of individual amplicons for WT and tested cell lines (in bp):

Primers combination:	**WT**	**A3**	**9**	**15**
DpSL_Fw2 and SL_Puro_Rv2	N/A	232	232	232
DpSL_Fw2 and SL_mCherry_Rv2	N/A	180	N/A	180
DpSL_Fw2 and SL_Atubulin_Rv2	N/A	N/A	no product	1041

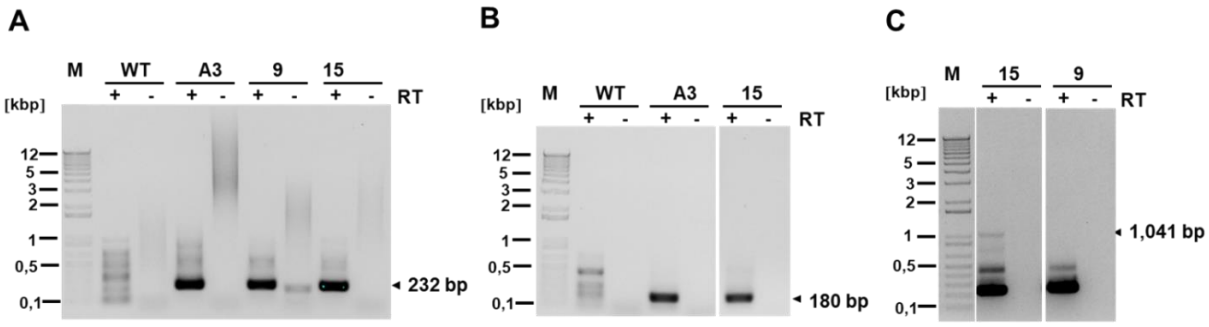


Figure 17: Agarose gel with PCR-amplified genes of 15, 9, A3 and WT.
(A) As the primers used are positioned around the SL of Puro^R, the obtained bands of equal size (232 bp) in cell line A3, 9 and 15 verify the correct addition of the SL to this region. (B) The presence of the SL on mCherry mRNAs is confirmed by strong bands of the same size (180 bp) in cell line A3 and 15. (C) Correct addition of SL is approved by a band of 1,041 bp in cell line 15.

All results of SL RT-PCR reactions show the expected sizes of the PCR products and therefore verify the correct transcription and post-transcriptional processing of both constructs in cell lines 9 and 15.

4.3.2 Verification of protein expression by Western blot analysis

In order to examine the translation of introduced heterologous genes, the expression of *Puro^R*, mCherry and Ty tag was tested by Western blot together with α-tubulin. Enolase was used as a loading control.

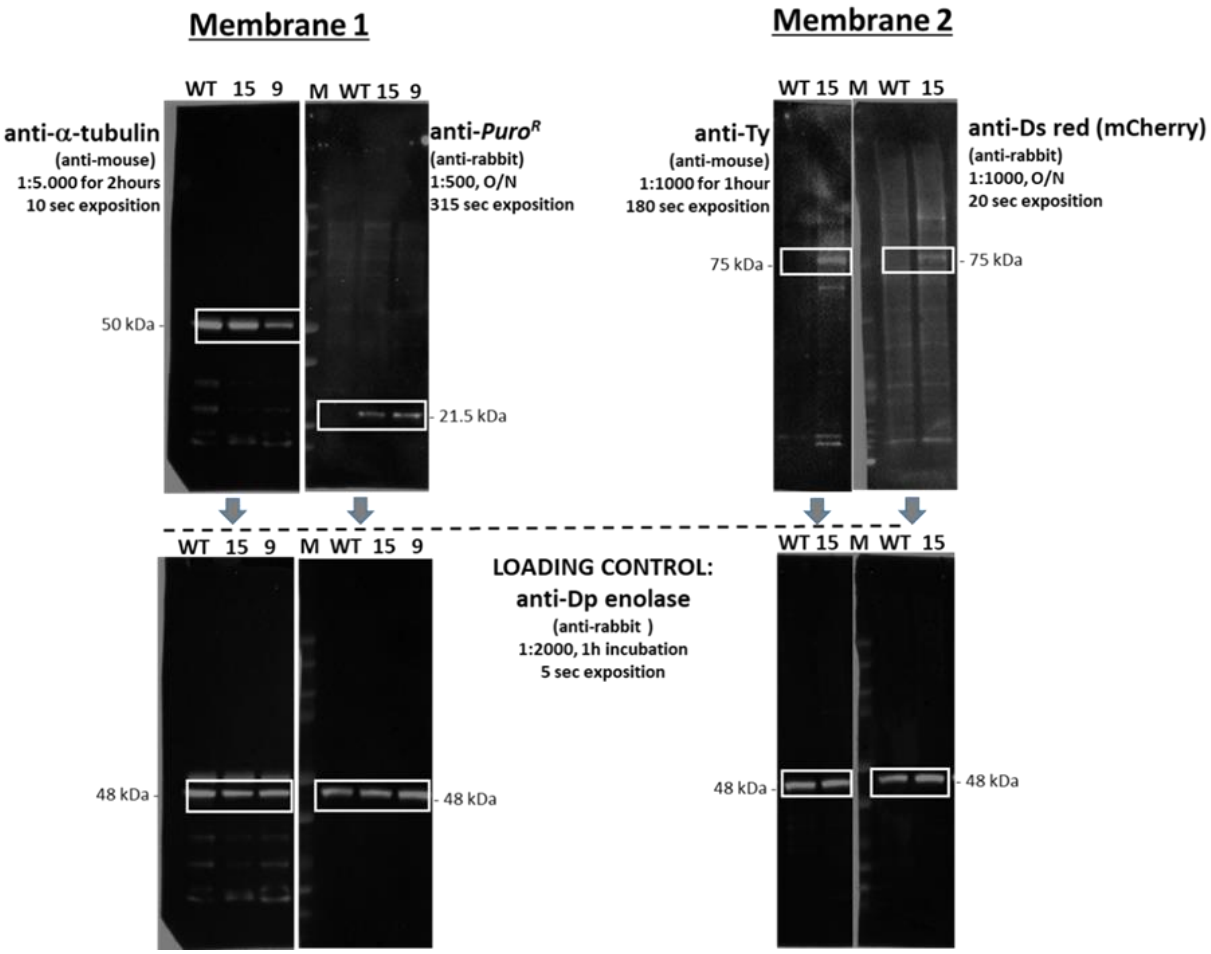


Figure 18: Western blot analysis of D. papillatum WT and tested cell lines A3, 9 and 15.
Membrane 1: The antibodies against α-tubulin resulted in bands of equal sizes (50 kDa) for WT, cell line 15 and 9. In detail, strong signals are achieved for WT and cell line 15 while the signal for cell line 9 is slightly less. On the other hand, anti-Puro^R antibodies reveal the expression of Puro^R for cell line 15 and 9 by means of a clear band being visible on the membrane (21.5 kDa).
Membrane 2: Antibodies against the Ty tag of mCherry and against mCherry directly detected respective protein in cell line 15 with clearly visible bands of the size of 75 kDa. The loading control, which was anti-Dp enolase prove the validity of the tests with unambiguous bands of the same size (48 kDa) for all of the cell lines tested.

Western blot analysis of *D. papillatum* cell lines 9 and 15 confirm the expression of *Puro^R* protein in both cell lines as well as cell line A3 in contrast to WT. Moreover, expression of mCherry in cell line 15 was verified using anti-Ty and anti-Ds red antibodies (Fig. 18).

Furthermore, anti-α-tubulin antibodies were used to test whether its level is not influenced. In addition, anti-Dp enolase served as a loading control.

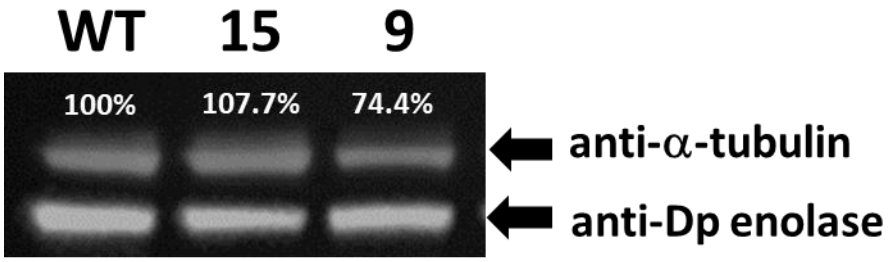


Figure 19: Western blot analysis of D. papillatum WT and tested cell lines 9 and 15 using anti-α-tubulin antibody. When comparing the intensity of the signals with WT being the reference therefore attributed with 100% intensity, a slightly stronger signal is visible in cell line 15 (107.7%), while the signal for cell line 9 was found to be decreased to 74.4%. The loading control with anti-Dp enolase antibodies verifies the validity of the experiment.

The level of α-tubulin expression seems to be decreased in cell line 9 in comparison to WT and cell line 15. The signal was quantified and enolase was used as a loading control.

4.3.3 Verification of protein expression by fluorescence microscopy

To verify the results of Western blot analysis, the expression of the mCherry gene using fluorescence microscopy was examined in cell line 15. Fluorescence microscopy of living cells of clone 15 show the red fluorescence of the whole cells including flagella that correspond to the expected α-tubulin localisation (Fig. 20). The fluorescence was observed in the whole population of the culture, in contrast to wild-type that exhibited only autofluorescence of the vesicles inside the cells.

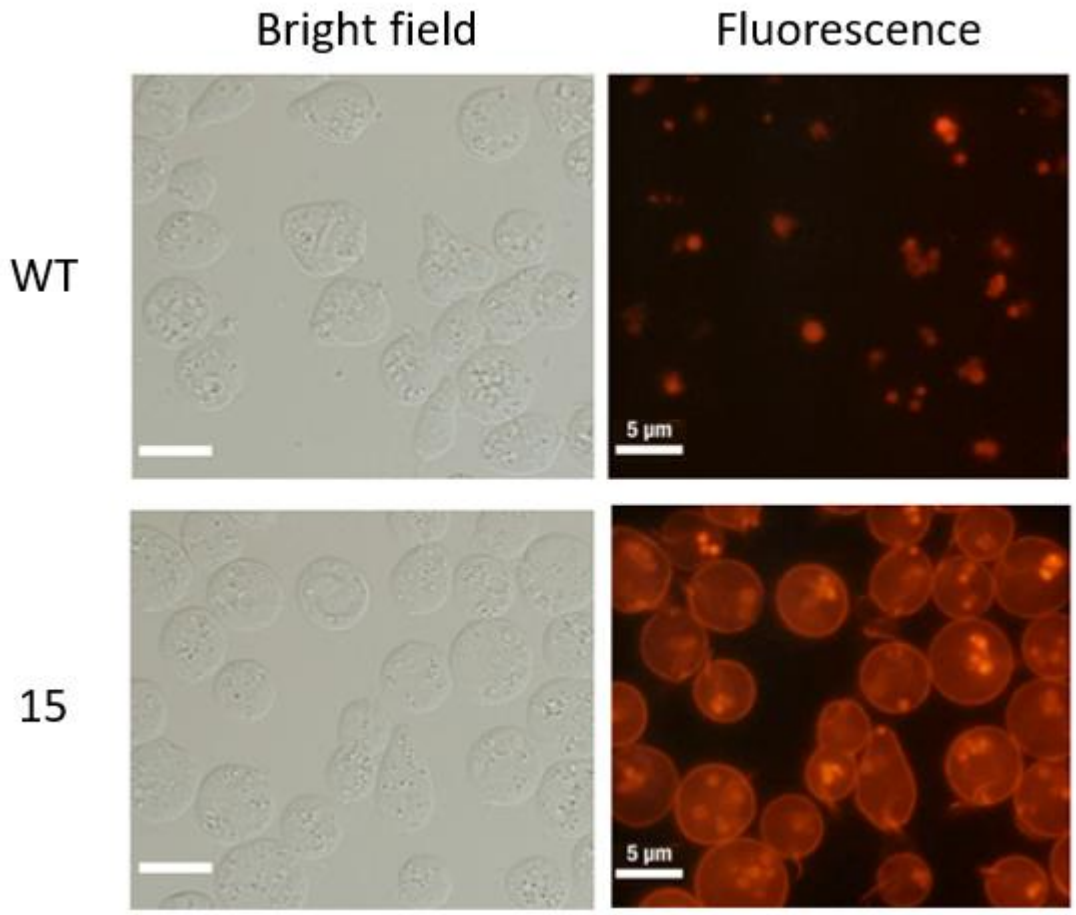


Figure 20: Fluorescence microscopy of WT and cell line 15 – left: bright field, right: fluorescence. As it is clearly visible from this figure, when comparing the pictures taken of WT and cell line 15, a distinct fluorescent signal was obtained for cell line 15 using fluorescent microscopy. The scale bars correspond to the size of 5 µm.

4.4 Measuring of the growth curves

In order to determine whether the insertion of any of the constructs in the genome and their expression influences the viability of *D. papillatum* cells, growth curves were measured.

The growth of cell lines 9 and 15 was compared to WT and cell line A3 (from previous study, here used as a control). The cell density was measured and the cells were diluted to the starting concentration of $1x10^6$ every 24 hours over the period of 10 days. A representative figure of three independent experiments is shown in Fig. 21, with the respective doubling-time being visualized in Fig. 22.

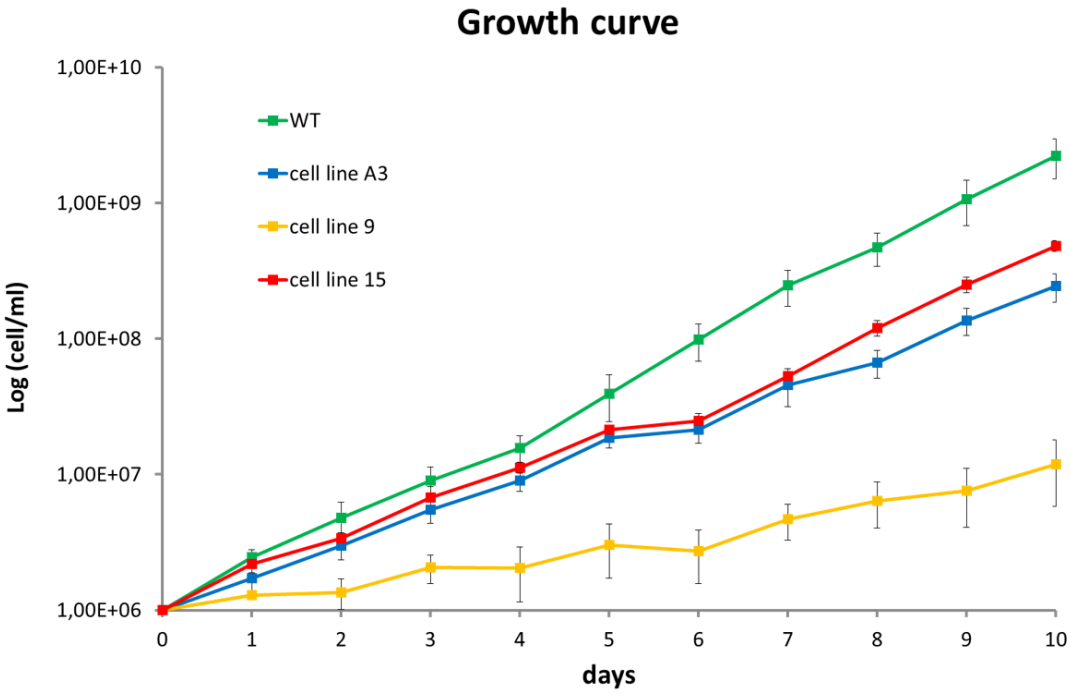


Figure 21: Growth curves of WT and cell lines A3, 9 and 15.
As these graphs clearly present, cell line 9 (yellow) grew significantly slower than the other cell lines in the time span of this experiment (ten days). The other two transformed cell lines, namely cell line A3 (blue) and cell line 15 (red) show a similar growth pattern.

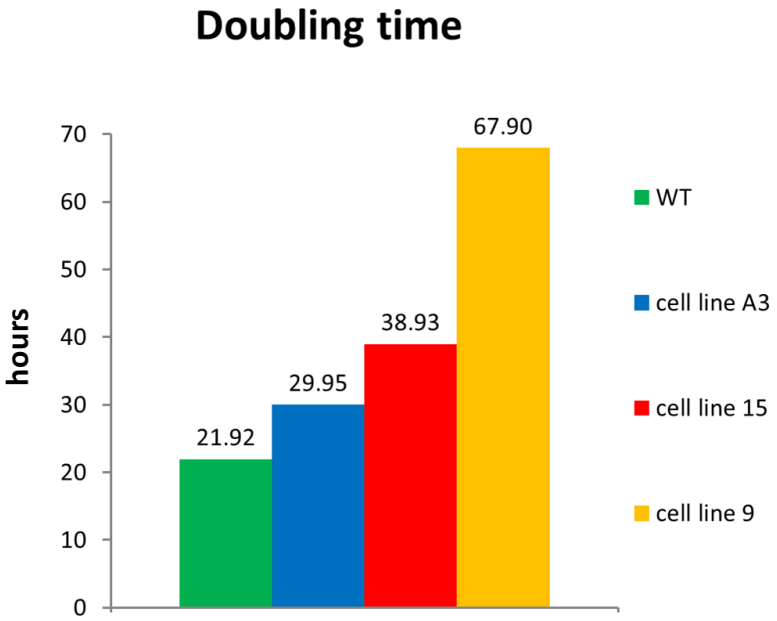


Figure 22: Doubling time of WT and transformed cell lines.
While the doubling time was found to be the shortest for WT (green), the doubling time of cell line A3 (blue), 15 (red) and 9 (yellow) was found to be longer, in the later significantly longer.

All three transformed cell lines exhibited slower growth compared to the WT. While in case of cell line A3 and 15, it was most likely caused by addition of puromycin (an antibiotic

used for the cell line selection) to the growth media, the growth of cell line 9 was significantly slower. The doubling time (the time which takes the cells to double their amount in the culture) was about two times slower (68 hours) in contrast to the cell lines A3 or 15 (30 and 39 hours, respectively) and more then tree times compared to the WT. This finding was surprising taking into the consideration that other 29 copies of α-tubulin are present in the genome.

Eventhough no significant changes were observed under the light microscope, it was decided to examine the ultrastructure of the cells using the transmittion electron miscoscopy.

4.5 Transmission electron microscopy

In order to look for any alterations due to the replacement of α-tubulin gene in cell line 9, as indicated by the results of growth curves and Western blot analysis, transmission electron microscopy (TEM) was performed, with special attention to details of the structure of the tubules in the flagella. WT cells were used as a control (Fig. 23).

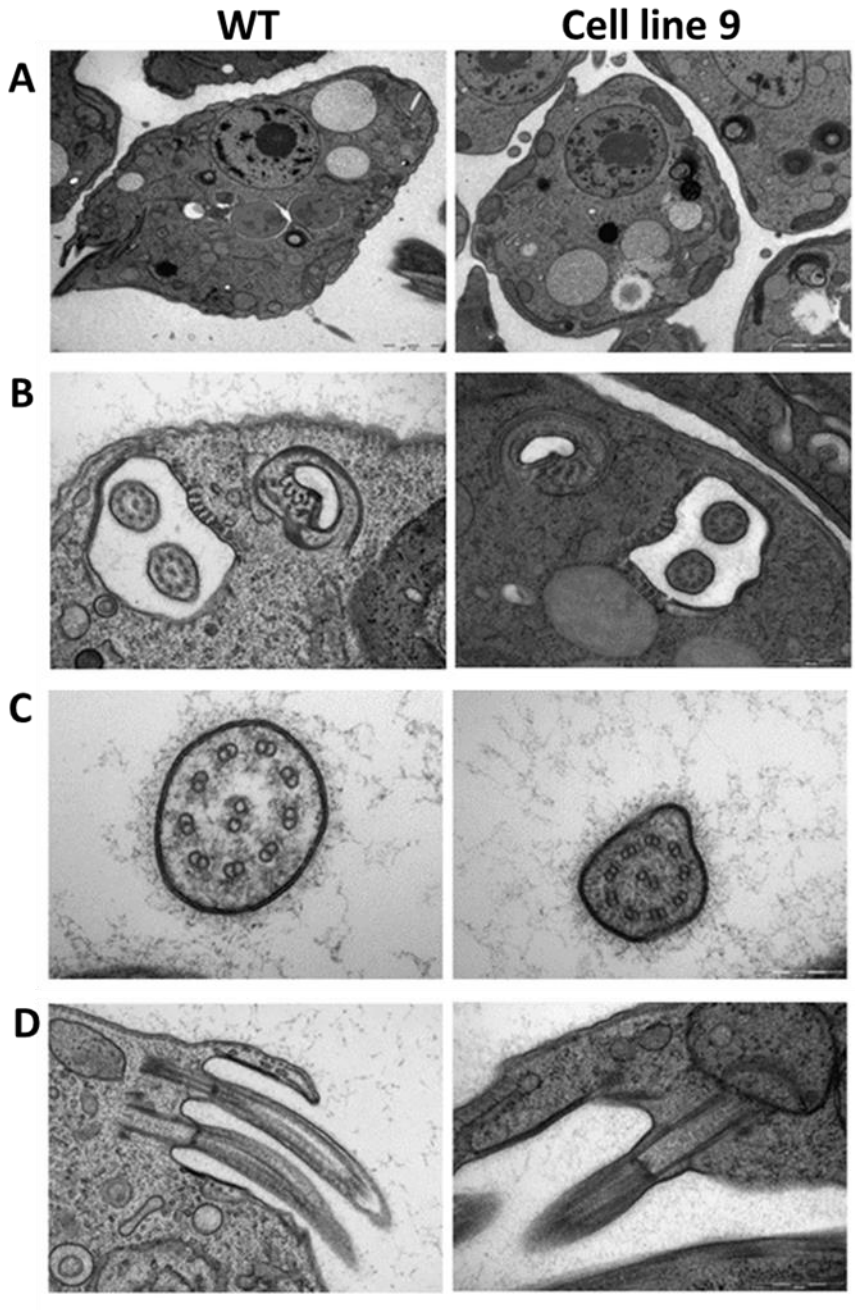


Figure 23: Comparison of whole cells and flagella of WT (left) with cell line 9 (right).
(A) The overall structure and composition of cell line 9 and WT does not appear to show any significant differences. (B) The TEM pictures provide evidence that both of the flagella are present in cell line 9 and WT and they seem to be equal in size. (C) The horizontal cross sections of the flagella visualize that the doublet of centric tubules and nine combinations of α- and β-tubules in an outer circle are present in both cells and do not show any significant changes. (D) The vertical cuts of the flagella do not give any hint towards any alterations of the basal body or the peripheral structure of cell line 9 when compared to WT.

Comparison of the flagella of WT and cell line 9 do not show any significant difference in the arrangement or amount of tubules.

5 Discussion and Conclusions

Protists are miniature, in the vast majority unicellular, eukaryotic organisms. Their classification is undergoing constant alterations as more and more new species are being discovered. When considering their positions on the eukaryotic tree of life, it is appearant that it practically completely covered by them (Fig. 1) eventhought it seems that we are still at the beginning of discovering their diversity. Thus, most of the eukaryotic species diversity (both terrestrial and marine) is hidden in the protozoan groups (O'Malley *et al.*, 2012; Pawlowski *et al.*, 2012; Burki, 2014; del Campo *et al.*, 2014; Adl *et al.*, 2019; Burki *et al.*, 2020). However till date, very little about most of them is known, eventhough plankton, as one of the many other groups and is mainly composed of marine protists, produces more than 50% oxygen on Earth. Together with the production of oxygen, they also absorb carbon dioxide and play an important role in the marine food chain (Falkowski *et al.*, 1998; Sekerci and Petrovskii, 2015).

The last decade has been crucial and has brought a significant shift in knowledge about the composition of protozoa in the seas and oceans. This was mainly due to the scientific expedition of *Tara Oceans* (2009-2012), which recovered hundreds of seawater samples from different depths of the world's oceans. Sequencing and analysis of the more than one billion sequences obtained using the V9 regions of ribosomal RNAs (rRNAs) uncovered the composition of microbial populations, revealed new species, and lead to numerous unexpected results (de Vargas *et al.*, 2015).

Among the biggest astonishment was the worldwide distribution, species abundance and diversity of diplonemids - a group of protist belonging to phylum Euglenozoa. It turned out that instead of several species described so far, there exist at least 40 thousand species, mainly at depths from 200 to 1,000 m (Flegontova *et al.*, 2016).

These surprising findings have deepened the interest in them. Apart from their sequences, however, the knowledge about these organisms is significantly limited. How do they look like? Do they live freely, parasitically or as commensals? In any case, they seem to be key players in the oceans. Their study can very likely provide information about relationships in microbial communities, symbiosis with other organisms and the overall evolution of life on Earth (Ptacnik *et al.*, 2008; Betts *et al.*, 2018).

To understand their role, it is needed to study their metabolism, the function of their genes and therefore to introduce protocols for the alteration of their genetic information in addition to the need to create a new experimental system for this group. However, this type of

research is generally very complicated, time consuming, risky and does not always leads to a successful outcome (Faktorová *et al.* 2020).

In our laboratory of Molecular biology of protists (Institute of Parasitology Biology Center, AS CR and Faculty of Science, University of South Bohemia), *Diplonema papillatum* was selected to be established as model organism for diplonemids because it grows axenically, fast and to high concentrations in a relatively simple liquid medium. Furthermore, *D. papillatum* is one of the most studied of diplonemids. To implement the transformation protocol, knowledge from its sister species *Trypanosoma brucei*, with which our laboratory has been working for many years, was used. Fortunatelly, many procedures proved to be applicable to *D. papillatum* as well and the first steps to creation of model organism and initial genetic tools has been already accomplished (Kaur *et al.,* 2018; Faktorová *et al.*, 2020). However the crucial step of targeted integration in the genome failed since the previously designed constructs were integrated stably, but randomly.

Therefore, two strategies were decided to be tested to overcome the random integration as described in the Introduction.

My work focused on testing the second hypothesis, namely whether the increased length of homologous regions will lead to correct integration.

Before my project in the lab started, two new constructs for modification of one α-tubulin gene (*DIPPA_23256*) were designed, amplified, transformed into *D. papillatum* WT cells and subsequently the transformants were selected by D. Faktorová.

Together with *D. papillatum* WT and cell line A3 (from the previous study, Kaur *et al.*, 2018), eight transformant cell lines for further examination were obtained: 1/ three originated from the replacement of α-tubulin and 2/ five from N-terminal tagging of α-tubulin with the fluorescent mCherry protein (the same contruct as used in Kaur *et al.*, 2018, but with increased arms of homology).

The aim of my thesis was mainly to select and further characterize the cell lines containing constructs integrated in the expected position in the genomic DNA and verify the correct expression of the integrated heterologous genes (Fig. 6E).

The PCR analysis revealed that one out of three knock-out cell lines (cell line 9) and one out of five N-terminal tagging cell lines (cell line 15) contained the respective correctly integrated construct in the genome. This denoted a breakthrough in *Diplonema* transformation as it means that homologous recombination is working in *D. papillatum* and that the tagging/knocking-out of any of its genes and their functional analysis will be possible.

These two cell lines were selected to be examined further and the corresponding results verify that the heterologous genes are correctly transcribed and post-transcriptionally processed. Their translation was confirmed by Western blots and in case of cell line 15 (N-terminal tagging of α-tubulin with mCherry) also by fluorescent microscopy.

The Western blot analysis of cell line 9 (α-tubulin replacement with *Puro*^{R}) resulted in a decrease of about 25% of the signal using α-tubulin antibodies, which would assume a decreased level of the α-tubulin proteins in the cell compared to WT. This was unexpected since, based on the preliminary research on the genome, there should be about 30 copies of α-tubulin genes present in *D. papillatum* genome. Therefore, it was decided to measure the growth curves and visualise the ultrastructure of cell line 9 compared to WT. While the growth of cell line 9 was affected in comparison to the assessment of growth pattern of the other two transformant cell lines grown in the same media, TEM analysis does not show any significant changes in the cell ultrastructure. The lower level of α-tubulin proteins in cell line 9 could be explained by preferential expression of the *DIPPA_23256* α-tubulin gene, however does not have effect on the cell ultrastructure.

Taken together, the findings obtained in this study are crucial for further steps of establishing methodology for gene tagging, gene knock-outs or knock-ins. Additional successes were achieved in the laboratory until the submission of this thesis, like the replacement of α- and β-tubulin with another resistant marker (*V5-Neo*^{R}), together with the design of the modular construct pDP002, which serves as a template for the tagging of any *D. papillatum* gene.

All these results are summarised in a manuscript only recently submitted to *Environmental Microbiology* journal which is attached to this thesis as supplementary data in the appendix (Faktorová et al., submitted).

6 References

Adl, S. M., Bass, D., Lane, C. E., Lukeš, J., Schoch, C. L., Smirnov, A., Agatha, S., Berney, C., Brown, M. W., Burki, F., Cárdenas, P., Čepička, I., Chistyakova, L., Del Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., Heiss, A. A., … Zhang, Q. (2019). Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *The Journal of Eukaryotic Microbiology, 66*(1), 4–119. https://doi.org/10.1111/jeu.12691

Baldauf, S. L. (2008). An overview of the phylogeny and diversity of eukaryotes. *Journal of Systematics and Evolution, 46*(3), 263–273. DOI: 10.3724/SP.J.1002.2008.08060.

Betts, H. C., Puttick, M. N., Clark, J. W., Williams, T. A., Donoghue, P., & Pisani, D. (2018). Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nature Ecology & Evolution*, *2*(10), 1556–1562. https://doi.org/10.1038/s41559-018-0644-x

Brandsma, I., & Gent, D. C. (2012). Pathway choice in DNA double strand break repair: observations of a balancing act. *Genome Integrity*, *3*(1), 9. https://doi.org/10.1186/2041-9414-3-9

Burki F. (2014). The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harbor perspectives in biology*, *6*(5), a016147. https://doi.org/10.1101/cshperspect.a016147

Burki, F., Roger, A. J., Brown, M. W., & Simpson, A. (2020). The New Tree of Eukaryotes. *Trends in Ecology & Evolution*, *35*(1), 43–55. https://doi.org/10.1016/j.tree.2019.08.008

Butenko, A., Opperdoes, F. R., Flegontova, O., Horák, A., Hampl, V., Keeling, P., Gawryluk, R., Tikhonenkov, D., Flegontov, P., & Lukeš, J. (2020). Evolution of metabolic capabilities and molecular features of diplonemids, kinetoplastids, and euglenids. *BMC Biology*, *18*(1), 23. https://doi.org/10.1186/s12915-020-0754-1

Caron, D. A., Alexander, H., Allen, A. E., Archibald, J. M., Armbrust, E. V., Bachy, C., Bell, C. J., Bharti, A., Dyhrman, S. T., Guida, S. M., Heidelberg, K. B., Kaye, J. Z., Metzner, J., Smith, S. R., & Worden, A. Z. (2017). Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nature Reviews. Microbiology*, *15*(1), 6–20. https://doi.org/10.1038/nrmicro.2016.160

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., Lima-Mendez, G., Rocha, F., Tirichine, L., Labadie, K., Kirilovsky, A., Bertrand, A., Engelen, S., Madoui, M. A., Méheust, R., Poulain, J., Romac, S., Richter, D. J., Yoshikawa, G., Dimier, C., … Wincker, P. (2018). A global ocean atlas of eukaryotic genes. *Nature Communications*, *9*(1), 373. https://doi.org/10.1038/s41467-017-02342-1

Cavalier-Smith T. (2016). Higher classification and phylogeny of Euglenozoa. *European Journal of Protistology*, *56*, 250–276. https://doi.org/10.1016/j.ejop.2016.09.003

Cavalier-Smith, T., Chao, E. E., & Vickerman, K. (2016). New phagotrophic euglenoid species (new genus Decastava; *Scytomonas saepesedens*; *Entosiphon oblongum*), Hsp90 introns, and putative euglenoid Hsp90 pre-mRNA insertional editing. *European Journal of Protistology*, *56*, 147–170. https://doi.org/10.1016/j.ejop.2016.08.002

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., … Karsenti, E. (2015). Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science (New York, N.Y.)*, *348*(6237), 1261605. https://doi.org/10.1126/science.1261605

del Campo, J., Sieracki, M. E., Molestina, R., Keeling, P., Massana, R., & Ruiz-Trillo, I. (2014). The others: our biased perspective of eukaryotic genomes. *Trends in Ecology & Evolution*, *29*(5), 252–259. https://doi.org/10.1016/j.tree.2014.03.006

Eichinger, L., Lee, S. S., & Schleicher, M. (1999). Dictyostelium as model system for studies of the actin cytoskeleton by molecular genetics. *Microscopy Research and Technique*, *47*(2),

124–134. https://doi.org/10.1002/(SICI)1097-0029(19991015)47:2<124::AID-JEMT5>3.0.CO;2-8

Faktorová, D., Dobáková, E., Peña-Diaz, P., & Lukeš, J. (2016). From simple to supercomplex: mitochondrial genomes of euglenozoan protists. *F1000Research*, *5*, F1000 Faculty Rev-392. https://doi.org/10.12688/f1000research.8040.2

Faktorová, D., Nisbet, 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., … Lukeš, J. (2020). Genetic tool development in marine protists: emerging model organisms for experimental cell biology. *Nature Methods*, *17*(5), 481–494. https://doi.org/10.1038/s41592-020-0796-x

Faktorová, D., Valach, M., Kaur, B., Burger, G., & Lukeš, J. (2018). Mitochondrial RNA Editing and Processing in Diplonemid Protists. In J. Cruz-Reyes & M. W. Gray (Hrsg.), *RNA Metabolism in Mitochondria* (p. 145–176). Springer International Publishing. https://doi.org/10.1007/978-3-319-78190-7_6

Falkowski, P. G., Barber, R. T., & Smetacek, V., V (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science (New York, N.Y.)*, *281*(5374), 200–207. https://doi.org/10.1126/science.281.5374.200

Flegontova, O., Flegontov, P., Malviya, S., Audic, S., Wincker, P., de Vargas, C., Bowler, C., Lukeš, J., & Horák, A. (2016). Extreme Diversity of Diplonemid Eukaryotes in the Ocean. *Current Biology*, *26*(22), 3060–3065. https://doi.org/10.1016/j.cub.2016.09.031

Gawryluk, R., Del Campo, J., Okamoto, N., Strassert, J., Lukeš, J., Richards, T. A., Worden, A. Z., Santoro, A. E., & Keeling, P. J. (2016). Morphological Identification and Single-Cell Genomics of Marine Diplonemids. *Current Biology*, *26*(22), 3053–3059. https://doi.org/10.1016/j.cub.2016.09.013

Hampl, V., Hug, L., Leigh, J. W., Dacks, J. B., Lang, B. F., Simpson, A. G., & Roger, A. J. (2009). Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proceedings of the National Academy of Sciences of the United States of America*, *106*(10), 3859–3864. https://doi.org/10.1073/pnas.0807880106

Jackson, A. P., Vaughan, S., & Gull, K. (2006). Evolution of tubulin gene arrays in Trypanosomatid parasites: genomic restructuring in Leishmania. *BMC Genomics, 7*, 261. https://doi.org/10.1186/1471-2164-7-261

Kaur, B., Valach, M., Peña-Diaz, P., Moreira, S., Keeling, P. J., Burger, G., Lukeš, J., & Faktorová, D. (2018). Transformation of Diplonema papillatum, the type species of the highly diverse and abundant marine microeukaryotes Diplonemida (Euglenozoa). *Environmental Microbiology*, *20*(3), 1030–1040. https://doi.org/10.1111/1462-2920.14041

Kiethega, G. N., Yan, Y., Turcotte, M., & Burger, G. (2013). RNA-level unscrambling of fragmented genes in *Diplonema* mitochondria. *RNA Biology, 10*(2), 301–313. https://doi.org/10.4161/rna.23340

Krejci, L., Altmannova, V., Spirek, M., & Zhao, X. (2012). Homologous recombination and its regulation. *Nucleic Acids Research, 40*(13), 5795–5818. https://doi.org/10.1093/nar/gks270

Lara, E., Moreira, D., Vereshchaka, A., & López-García, P. (2009). Pan-oceanic distribution of new highly diverse clades of deep-sea diplonemids. *Environmental Microbiology, 11*(1), 47–55. https://doi.org/10.1111/j.1462-2920.2008.01737.x

López-García, P., Vereshchaka, A., & Moreira, D. (2007). Eukaryotic diversity associated with carbonates and fluid-seawater interface in Lost City hydrothermal field. *Environmental Microbiology*, *9*(2), 546–554. https://doi.org/10.1111/j.1462-2920.2006.01158.x

Lukeš, J., Flegontova, O., & Horák, A. (2015). Diplonemids. *Current Biology*, *25*(16), R702–R704. https://doi.org/10.1016/j.cub.2015.04.052

Lukeš, J., Wheeler, R., Jirsová, D., David, V., & Archibald, J. M. (2018). Massive mitochondrial DNA content in diplonemid and kinetoplastid protists. *IUBMB Life*, *70*(12), 1267–1274. https://doi.org/10.1002/iub.1894

Marande, W., & Burger, G. (2007). Mitochondrial DNA as a genomic jigsaw puzzle. *Science (New York, N.Y.)*, *318*(5849), 415. https://doi.org/10.1126/science.1148033

Marande, W., Lukeš, J., & Burger, G. (2005). Unique mitochondrial genome structure in diplonemids, the sister group of kinetoplastids. *Eukaryotic Cell*, *4*(6), 1137–1146. https://doi.org/10.1128/EC.4.6.1137-1146.2005

Massana R. (2011). Eukaryotic picoplankton in surface oceans. *Annual Review of Microbiology*, *65*, 91–110. https://doi.org/10.1146/annurev-micro-090110-102903

Morales, J., Hashimoto, M., Williams, T. A., Hirawake-Mogi, H., Makiuchi, T., Tsubouchi, A., Kaga, N., Taka, H., Fujimura, T., Koike, M., Mita, T., Bringaud, F., Concepción, J. L., Hashimoto, T., Embley, T. M., & Nara, T. (2016). Differential remodelling of peroxisome function underpins the environmental and metabolic adaptability of diplonemids and kinetoplastids. *Proceedings. Biological sciences*, *283*(1830), 20160520. https://doi.org/10.1098/rspb.2016.0520

Mukherjee, I., Hodoki, Y., Okazaki, Y., Fujinaga, S., Ohbayashi, K., & Nakano, S. I. (2019). Widespread Dominance of Kinetoplastids and Unexpected Presence of Diplonemids in Deep Freshwater Lakes. *Frontiers in Microbiology*, *10*, 2375. https://doi.org/10.3389/fmicb.2019.02375

Okamoto, N., Gawryluk, R., Del Campo, J., Strassert, J., Lukeš, J., Richards, T. A., Worden, A. Z., Santoro, A. E., & Keeling, P. J. (2019). A Revised Taxonomy of Diplonemids Including the Eupelagonemidae n. fam. and a Type Species, Eupelagonema oceanica n. gen. & sp. *The Journal of Eukaryotic Microbiology*, *66*(3), 519–524. https://doi.org/10.1111/jeu.12679

O'Malley, M. A., Simpson, A. G. B., & Roger, A. J. (2013). The other eukaryotes in light of evolutionary protistology. *Biology & Philosophy*, *28*(2), 299–330. https://doi.org/10.1007/s10539-012-9354-y

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S. S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A. M., Gile, G. H., Holzmann, M., Jahn, R., Jirků, M., Keeling, P. J., Kostka, M., Kudryavtsev, A., Lara, E., Lukeš, J., … de Vargas, C. (2012). CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biology*, *10*(11), e1001419. https://doi.org/10.1371/journal.pbio.1001419

Porter, D. (1973). *Isonema papillatum* sp. n., a new colorless marine flagellate: a light- and electronmicroscopic study. *The Journal of Protozoology*, *20*(3), 351–356. https://doi.org/10.1111/j.1550-7408.1973.tb00895.x

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*(3), 259–282. https://doi.org/10.1016/j.protis.2019.04.001

Ptacnik, R., Solimini, A. G., Andersen, T., Tamminen, T., Brettum, P., Lepistö, L., Willén, E., & Rekolainen, S. (2008). Diversity predicts stability and resource use efficiency in natural phytoplankton communities. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(13), 5134–5138. https://doi.org/10.1073/pnas.0708328105

Roy, J., Faktorová, D., Benada, O., Lukes, J., & Burger, G. (2007). Description of Rhynchopus euleeides n. sp. (Diplonemea), a free-living marine euglenozoan. *The Journal of Eukaryotic Microbiology*, *54*(2), 137–145. https://doi.org/10.1111/j.1550-7408.2007.00244.x

Sekerci, Y., & Petrovskii, S. (2015). Mathematical Modelling of Plankton-Oxygen Dynamics Under the Climate Change. *Bulletin of Mathematical Biology*, *77*(12), 2325–2353. https://doi.org/10.1007/s11538-015-0126-0

Simpson, A. G. B. (1997). The identity and composition of the Euglenozoa. *Archiv Für Protistenkunde*, *148*(3), 318–328. https://doi.org/10.1016/S0003-9365(97)80012-7

Škodová-Sveráková, I., Prokopchuk, G., Peña-Diaz, P., Záhonová, K., Moos, M., Horváth, A., Šimek, P., & Lukeš, J. (2020). Unique Dynamics of Paramylon Storage in the Marine Euglenozoan *Diplonema papillatum*. *Protist*, *171*(2), 125717. https://doi.org/10.1016/j.protis.2020.125717

Son, M. Y., & Hasty, P. (2019). Homologous recombination defects and how they affect replication fork maintenance. *AIMS Genetics*, *5*(4), 192–211. https://doi.org/10.3934/genet.2018.4.192

Sturm, N. R., Maslov, D. A., Grisard, E. C., & Campbell, D. A. (2001). Diplonema spp. possess spliced leader RNA genes similar to the Kinetoplastida. *The Journal of Eukaryotic Microbiology*, *48*(3), 325–331. https://doi.org/10.1111/j.1550-7408.2001.tb00321.x

Tashyreva, D., Prokopchuk, G., Yabuki, A., Kaur, B., Faktorová, D., Votýpka, J., Kusaka, C., Fujikura, K., Shiratori, T., Ishida, K. I., Horák, A., & Lukeš, J. (2018). Phylogeny and Morphology of New Diplonemids from Japan. *Protist*, *169*(2), 158–179. https://doi.org/10.1016/j.protis.2018.02.001

Valach, M., Léveillé-Kunst, A., Gray, M. W., & Burger, G. (2018). Respiratory chain Complex I of unparalleled divergence in diplonemids. *The Journal of Biological Chemistry*, *293*(41), 16043–16056. https://doi.org/10.1074/jbc.RA118.005326

Valach, M., Moreira, S., Faktorová, D., Lukeš, J., & Burger, G. (2016). Post-transcriptional mending of gene sequences: Looking under the hood of mitochondrial gene expression in diplonemids. *RNA Biology*, *13*(12), 1204–1211. https://doi.org/10.1080/15476286.2016.1240143

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.

von der Heyden, S., Chao, E. E., Vickerman, K., & Cavalier-Smith, T. (2004). Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of euglenozoa. *The Journal of Eukaryotic Microbiology*, *51*(4), 402–416. https://doi.org/10.1111/j.1550-7408.2004.tb00387.x

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*, *17*, 39–51. https://doi.org/10.1016/j.dnarep.2014.02.008

Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., & Keeling, P. J. (2015). Environmental science. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science (New York, N.Y.)*, *347*(6223), 1257594. https://doi.org/10.1126/science.1257594

Yi, Z., Berney, C., Hartikainen, H., Mahamdallie, S., Gardner, M., Boenigk, J., Cavalier-Smith, T., & Bass, D. (2017). High-throughput sequencing of microbial eukaryotes in Lake Baikal reveals ecologically differentiated communities and novel evolutionary radiations. *FEMS Microbiology Ecology*, *93*(8), 10.1093/femsec/fix073. https://doi.org/10.1093/femsec/fix073

7 List of supplementary data

Appendix A Faktorová, D, Kaur, B, Valach, M., Graf, L., Benz, C., Burger, G., Lukeš, J. (submitted). Targeted integration by homologous recombination enables *in-situ* tagging and replacement of genes in the marine microeukaryote *Diplonema papillatum*. *Environmental Microbiology*

Appendix B Supporting information from Faktorová, D, Kaur, B, Valach, M., Graf, L., Benz, C., Burger, G., Lukeš, J. (submitted). Targeted integration by homologous recombination enables *in-situ* tagging and replacement of genes in the marine microeukaryote *Diplonema papillatum*. *Environmental Microbiology*

8 Supplementary data

environmental microbiology
environmental microbiology reports

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

Journal:	*Environmental Microbiology and Environmental Microbiology Reports*
Manuscript ID	EMI-2020-0707
Journal:	Environmental Microbiology
Manuscript Type:	EMI - Research article
Date Submitted by the Author:	09-May-2020
Complete List of Authors:	Faktorová, Drahomíra; Biology Center, AS CR, Institute of Parasitology Lukes, Julius; Institute of Parasitology, České Budejovice, Burger, Getraud; Université de Montréal, Department of Biochemistry and Robert-Cedergren Centre for Bioinformatics and Genomics Kaur, Binnypreet; Biology Center, AS CR, Institute of Parasitology; University of South Bohemia, Faculty of Sciences Valach, Matus; Université de Montréal, Department of Biochemistry and Robert-Cedergren Centre for Bioinformatics and Genomics
Keywords:	gene targeting; diplonemids; marine protist; resistance marker; model organism

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,@], Corinna Benz[1], Gertraud Burger[3], Julius Lukeš[1,2,*]

[1] Institute of Parasitology, Biology Centre, Czech Academy of Sciences and [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

[#] Equal contribution

[@] Present address: Johannes Kepler University, Linz, Austria

* To whom correspondence should be addressed: Drahomíra Faktorová (dranov@paru.cas.cz), Julius Lukeš (jula@paru.cas.cz)

Running title: Gene tagging and replacement in *Diplonema papillatum*

Key words (max 5-10): gene targeting; diplonemids; marine protist; resistance marker; model organism

Originality-Significance Statement

The recent development of a transformation protocol for *Diplonema papillatum* represents a crucial advancement towards studying the function of individual genes of this single-celled eukaryote. Here, we follow up on the previous study and demonstrate the feasibility of stable gene replacement, integration, and tagging by homologous recombination, turning the type species of the highly diverse and abundant diplonemid flagellates into a genetically tractable organism. Our work thus lays the foundations for unravelling the ecological role of these heterotrophic protists in the world oceans.

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.

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 (Ebenezer *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 technical issues. 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 5th to 6th 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,b), (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 the by far 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*., 2018a), *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 remodeled 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 was 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- (*DpTUB1*; *DIPPA_23256*) and β- (*DpTUB2*; *DIPPA_12526*) 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 (*DIPPA_23256*) with the fluorescent mCherry protein. As a means for systematic gene deletion, insertion, and tagging, we next built the modular construct pDP002, with which we achieved precise 3′-tagging of the mitoribosomal protein gene *DpMRPS49* (*DIPPA_31280*) and intergenic insertion of the heterologous ascorbate peroxidase gene *APEX2*. 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 (our unpubl. data), 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 μg, 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 to 2 kb-long homology arms were electroporated. Integration into the genome was analyzed 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

D. 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, i.e., 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**). Should a construct insert at a different genomic location than intended, *Puro*^{R} would likely be expressed due to the polycistronic transcription in diplonemids (Kaur *et al.*, 2018, Faktorová *et al.,* 2020). On the other hand, if homologous integration takes place, this would not be lethal, because the targeted genes exist in multiple copies in the genome. The location of integration can be determined accurately, because sequences flanking the chosen loci are unique.

Puro^R cell lines were recovered after cultivation in the presence of the selection drug for 10 to 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 Figs. S4A,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) out 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 verified 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 1,784 and 1,416 bp, respectively. One out 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 positive impact. Cell line

15 (together with the controls) was also tested by two other primer pairs (**Fig. 1B** and **Supporting Information Fig. S5A**). Next, the correct integration of the N-terminally tagged α-tubulin with mCherry was validated by SL RT-PCR (**Supporting Information Figs. 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 labeled 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) (**Figs. 2A,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 our knowledge, this is the first report of a successful deployment of a tagged resistance marker, which can be useful for any organism where no other antibodies are available to test the expression of the introduced construct.

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) (**Figs. 2A,B**). A total of 9 and 10 neomycin-resistant cell lines, respectively, were recovered after 8 to 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 (**Figs. 2C,D**). Two cell lines 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 since have been deployed successfully worldwide (Goos *et al.*, 2017; Sunter *et al.*, 2019; Benz and Urbaniak, 2019), including our laboratory (Faktorová *et*

al., 2018b; Peña-Diaz *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 even though 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 DpMRPL76 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 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. While 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. Since 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, e.g., 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; Nayak *et al.*, 2006; Goins *et al*., 2006 and 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. Since 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). Since 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

D. 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 $5x10^7$ cells ($2x10^6$ cells/ml) were pre-incubated in 5 μg/ml 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) 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, 2B, 3B, 3C** 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:

(i) 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 (2,016 bp and 1,840 bp) and β-tubulin (952 bp and 1,333 bp) genes were used for amplification of long homology arms (PCR A and C). The final transformation cassettes of 1,841 bp + 1,553 bp and 890 bp + 1,203 bp homology regions (PCR D)

were amplified using nested primers for α and β-tubulin, respectively. For details see **Figs. 1A, 2A, 2B,** and **Supporting Information Fig. S3** and **Table 1.**

(ii) 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**.

(iii) 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 1,575 bp and 1,605 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.

(iv) V5–tagged heterologous ascorbate peroxidase – integration into intergenic region

D. papillatum codon-optimized V5-APEX2 coding sequence was synthesized as a gBlocks gene fragment by Integrated DNA Technologies (Coralville, 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 1.5M NaCl and 1.4V isopropanol overnight at 4 °C, and finally solubilized in 10 mM Tris-HCl, pH 8.0 to 400 ng/µL. 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 into 5 x 10^7 cells by electroporation (Amaxa Nucleofector II, program X-001), as described previously (Kaur *et al*. 2018), with a DNA-free negative control. In the case of the W7 inhibitor experiment, the cells were pre-incubated in W7 prior to electroporation (see above). Eight to 16 hours after the electroporation, transfectants were subjected to selection with increasing concentrations of puromycin (12–40 µg/ml) for *Puro*^{R} or G418 (25-80 µg/ml) 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 BSA, 1 mM ITP) using Gene Pulser Xcell apparatus (Bio-Rad) at 1,500 V for 0.3 ms. After a six hour-long recuperation period, transformed cells were selected in the presence of G418 (100 µg/mL).

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, 2B, 3B, 3C** 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. S5B-D, S6A and S7A**. The obtained amplicons were verified by sequencing.

Western blot analysis

Cell lysates were prepared by resuspending 5x10^5 cells in 25 µl of 2x 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 1x 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 1:1,000; Sigma-Aldrich, SAB4800032) for the Ty-tag; anti-Ds Red antibody 1:1,000 (produced in-rabbit; Clontech, 632496) for mCherry; anti-V5 antibody (produced in rabbit [PA1-993] or mouse [R960-25], used at 1:1,000; Thermo-Fisher

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 1:5,000; Sigma-Aldrich, T9026). As a loading control, the anti-enolase antibody (produced in rabbit, used at 1:2,000; 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 1:2,000 dilution (Sigma) at room temperature for 1 hr and visualized using Clarity Western ECL substrate (Bio-Rad).

Fluorescence microscopy

Five µl 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 (16–18699S to J.L.), the ERC CZ (LL1601 to J.L.), the Czech Ministry of Education (ERD Funds OPVVV16_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

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

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

7. 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.

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

9. 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.

10. Ebenezer, T., Zoltner, M., Burrel, A., Nenarokova, A., Vanclová, A., Prasad, B., *et al.* (2019) Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biol.* **17**: 11.

11. 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.

12. Faktorová, D., Valach, M., Kaur, B., Burger, G., and Lukeš, J. (2018a) Mitochondrial RNA editing and processing in diplonemid protists. In: M.W. Gray and J. Cruz-Reyes (Eds.), RNA Metabolism in Mitochondria, *Springer* 145–176.

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

14. 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, *Nature Methods*, doi: 10.1038/s41592-020-0796-x.

15. 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.

16. 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.

17. 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.

18. 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.

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

20. 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.

21. 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.

22. Kaur, B., Valach, M., Peña-Diaz, 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 Diplonemida (Euglenozoa). *Environ Microbiol* **20:** 1030–1040.

23. Kaur, B., Záhonová, K., Valach, M., Faktorová, D., Prokopchuk, G., Burger, G.

600 and Lukeš, J. (2020) Gene fragmentation and RNA editing without borders: eccentric
601 mitochondrial genomes of diplonemids. *Nucleic Acids Res* **48**: 2694-2708.
602 24. Kiethega, G.N., Yan, Y., Turcotte, M., and Burger, G. (2013) RNA-level
603 unscrambling of fragmented genes in *Diplonema* mitochondria. *RNA Biol* **10:** 301–313.
604 25. Krejci, L., Altmannova, V., Spirek M., and Zhao, X. (2012) Homologous
605 recombination and its regulation. *Nucleic Acids Res* **40:** 5795–5818.
606 26. Lam, S.S., Martell, J.D., Kamer, K.J., Deerinck, T.J., Ellisman, M.H., Mootha,
607 V.K., and Ting, A.Y. (2015) Directed evolution of APEX2 for electron microscopy and
608 proximity labeling. *Nat Methods* **12:** 51–54.
609 27. López-García, P., Vereshchaka, A., and Moreira, D. (2007) Eukaryotic diversity
610 associated with carbonates and fluid sea water interface in lost city hydrothermal field.
611 *Environ Microbiol* **9:** 546–554.
612 28. Lukeš, J., Flegontova, O., and Horák, A. (2015) Diplonemids. *Curr Biol* **25:**
613 R702–704.
614 29. Lukeš, J., Wheeler, R., Jirsová, D., David, V., and Archibald, J.M. (2018)
615 Massive mitochondrial DNA content in diplonemid and kinetoplastid protists. *IUBMB*
616 *Life* **70:** 1267–1274.
617 30. Malkova, A, and Haber, JE. (2012) Mutations arising during repair of
618 chromosome breaks. *Annu Rev Genet* **46:** 455–473.
619 31. Marande, W., Lukeš, J., and Burger, G. (2005) Unique mitochondrial genome
620 structure in diplonemids, the sister group of kinetoplastids. *Eukaryot Cell* **4:**1137–1146.
621 32. Marande, W., and Burger, G. (2007) Mitochondrial DNA as a genomic jigsaw
622 puzzle. *Science* **318:** 415.

623 33. Massana, R. (2011) Eukaryotic picoplankton in surface oceans. *Annu Rev*
624 *Microbiol* **65:** 91–110.
625 34. Matthews, K.R., McCulloch, R., and Morrison, L.J. (2015) The within-host
626 dynamics of African trypanosome 726 infections. Philosophical transactions of the
627 Royal Society of London Series B, *Biological sciences* 370.
628 35. McKean, P.G., Vaughan, S., and Gull, K. (2001) The extended tubulin
629 superfamily. *J Cell Sci* **114:** 2723–2733.
630 36. Moreira, S., Valach, M., Aoulad-Aissa, M., Otto, C., and Burger, G. (2016)
631 Novel modes of RNA editing in mitochondria. *Nucleic Acids Res* **44:** 4907–4919.
632 37. Mukherjee, I., Hodoki, Y., Okazaki, Y., Fujinaga, S., Ohbayashi, K., and
633 Nakano, S.I. (2019) Widespread dominance of kinetoplastids and unexpected presence
634 of diplonemids in deep freshwater lakes. *Front Microbiol* **10:** 2375.
635 38. Nayak, T., Szewczyk, E., Oakley, C.E., Osmani, A., Ukil, L., Murray, S.L., *et al.*
636 (2006) A versatile and efficient gene-targeting system for *Aspergillus nidulans*. *Genetics*
637 **172:** 1557–1566.
638 39. Nenarokova, A., Záhonová, K., Krasilnikova, M., Gahura, O., McCulloch, R.,
639 Zíková, A., *et al.* (2019) Causes and effects of loss of classical nonhomologous end
640 joining pathway in parasitic eukaryotes. *mBio* **10:** e01541–19.
641 40. Ninomiya, Y., Suzuki, K., Ishii, C., and Inoue, H. (2004) Highly efficient gene
642 replacements in *Neurospora* strains deficient for nonhomologous end-joining. *Proc Natl*
643 *Acad Sci USA* **101:** 12248–12253.
644 41. Okamoto, N., Gawryluk, R.M.R., Campo, J., Strassert, J.F.H., Lukeš, J.,
645 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.

42. 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.

43. 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.

44. Rodgers, K., and McVey, M. (2016) Error-prone repair of DNA double-strand breaks. *J Cell Physiol* **231:** 15–24.

45. 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). Vol. 533. Methods in Molecular Biology. Totowa, NJ: Humana Press. pp. 33–47.

46. 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* **l54:** 137–145.

47. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

48. 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.

669 49. Simpson, A.G.B. (1997) The identity and composition of the Euglenozoa. *Archiv*
670 *für Protistenkunde* **148:** 318–328.
671 50. Son, M.Y., and Hasty, P. (2019) Homologous recombination defects and how
672 they affect replication fork maintenance. *AIMS Genet* **5:** 192–211.
673 51. Tashyreva, D., Prokopchuk, G., Yabuki, A., Kaur, B., Faktorová, D., Votýpka,
674 J., *et al.* (2018a) Phylogeny and morphology of new diplonemids from Japan. *Protist*
675 **169:** 158–179.
676 52. Tashyreva, D., Prokopchuk, G., Votýpka, J., Yabuki, A., Horák, A., and Lukeš,
677 J. (2018b) Life cycle, ultrastructure, and phylogeny of new diplonemids and their
678 endosymbiotic bacteria. *mBio* **9:** e02447–17.
679 53. Trahan, C., Aguilar, L.-C., and Oeffinger, M. (2016) Single-step affinity
680 purification (ssAP) and mass spectrometry of macromolecular complexes in the yeast *S.*
681 *cerevisiae*. *Methods Mol Biol* **1361:** 265–287.
682 54. Valach, M., Moreira, S., Faktorová, D., Lukeš, J., and Burger, G. (2016) Post-
683 transcriptional mending of gene sequences: Looking under the hood of mitochondrial
684 gene expression in diplonemids. *RNA Biol* **13:** 1204–1211.
685 55. Valach, M., Léveillé-Kunst, A., Gray, M.W., and Burger, G. (2018) Respiratory
686 chain complex I of unparalleled divergence in diplonemids. *J Biol Chem* **293:** 16043–
687 16056.
688 56. Vickerman, K. (2000) Diplonemids (Class: Diplonemea Cavalier Smith, 1993)
689 In Lee JJ, Leedale GF, Bradbury P (*eds*) An Illustrated Guide to the Protozoa Vol. 2,
690 2nd edn, *Society of Protozoologists*, *Lawrence, Kansas, USA*, pp 1157–1159.

691 57. von der Heyden, S., Chao, E.E., Vickerman, K., and Cavalier-Smith, T. (2004)
692 Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of
693 euglenozoa. *J Eukaryot Microbiol* **51:** 402–416.
694 58. Waters, C.A., Strande, N.T., Wyatt, D.W., Pryor, J.M., and Ramsden, D.A.
695 (2014) Nonhomologous end joining: a good solution for bad ends. *DNA Repair* **17:** 39–
696 51.
697 59. Yi, Z., Berney, C., Hartikainen, H., Mahamdallie, S., Gardner, M., Boenigk, J.,
698 Cavalier-Smith, T., and Bass, D. (2017) High-throughput sequencing of microbial
699 eukaryotes in Lake Baikal reveals ecologically differentiated communities and novel
700 evolutionary radiations. *FEMS Microbiol Eco.* **93:** 8.
701 60. Zhao, Z., Liu, H., Luo, Y., Zhou, S., An, L., Wang, C., Jin, Q., Zhou, M., and
702 Xu, J.R. (2014) Molecular evolution and functional divergence of tubulin superfamily in
703 the fungal tree of life. *Sci Rep* **23:** 6746.
704
705
706
707
708

709 **FIGURE LEGENDS**

710 **Fig. 1**: Replacement of the alpha-tubulin gene and N-terminal tagging under puromycin
711 selection.
712 **A.** Schema of the genomic neighborhood of the wild-type (WT) alpha-tubulin locus
713 (*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.

Fig. 2: Replacement of alpha- and beta-tubulin genes with a V5-*Neo*^{R} fusion.

A. Schema of the genomic neighborhood 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 neighborhood of the wild-type alpha-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**.)

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 neighborhood 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 neighborhood 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).

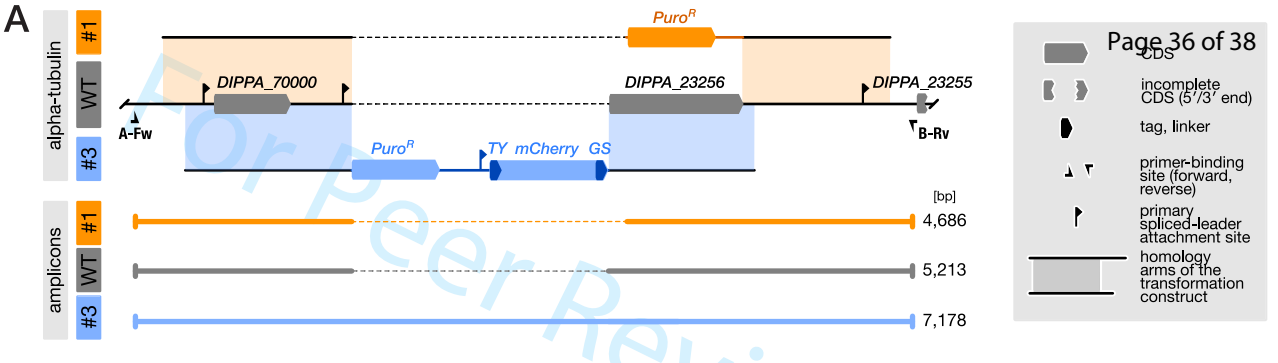
The schemas in B and C include positions of the homology arms, primers used for PCR validation of the on-target integration, and expected sizes of the amplicons. (For in-scale schemas, see **Supporting Information Fig. S3**.)

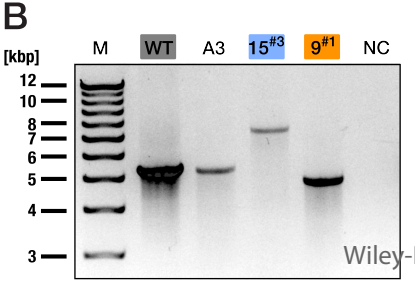
D. PCR of total DNA of *D. papillatum* wild-type and six selected transformant cell lines

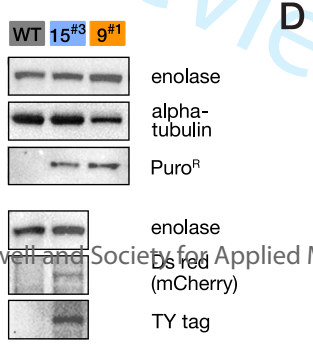
using primers that bind outside the target region. (Uncropped gel is shown in **Supporting Information Fig. S8**.)

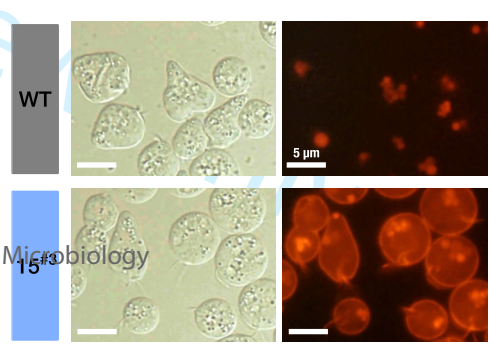
E. Western blot analysis of *D. papillatum* WT and selected transformant cell lines. Alpha-tubulin was used as a loading control. (Uncropped blot is shown in **Supporting Information Fig. S9**.)

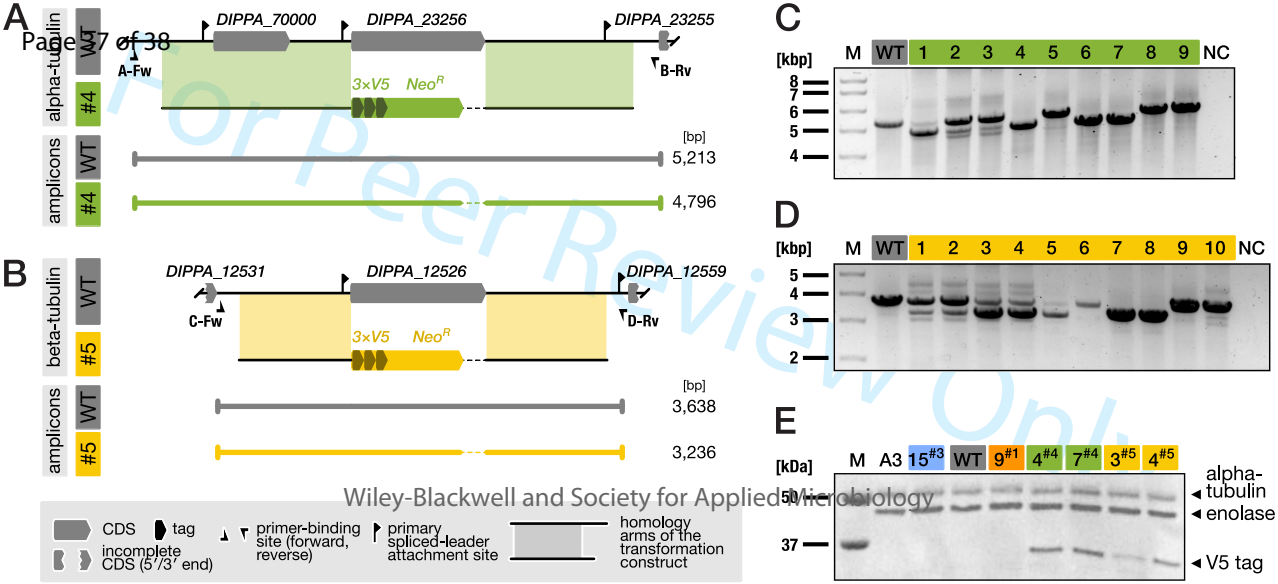
F. PCR of total DNA of *D. papillatum* wild-type and transformant cell lines containing the construct #7 or #8 using primers that bind outside the target region. Negative control PCR (NC) was performed without template DNA. (Uncropped gel is shown in **Supporting Information Fig. S8**).

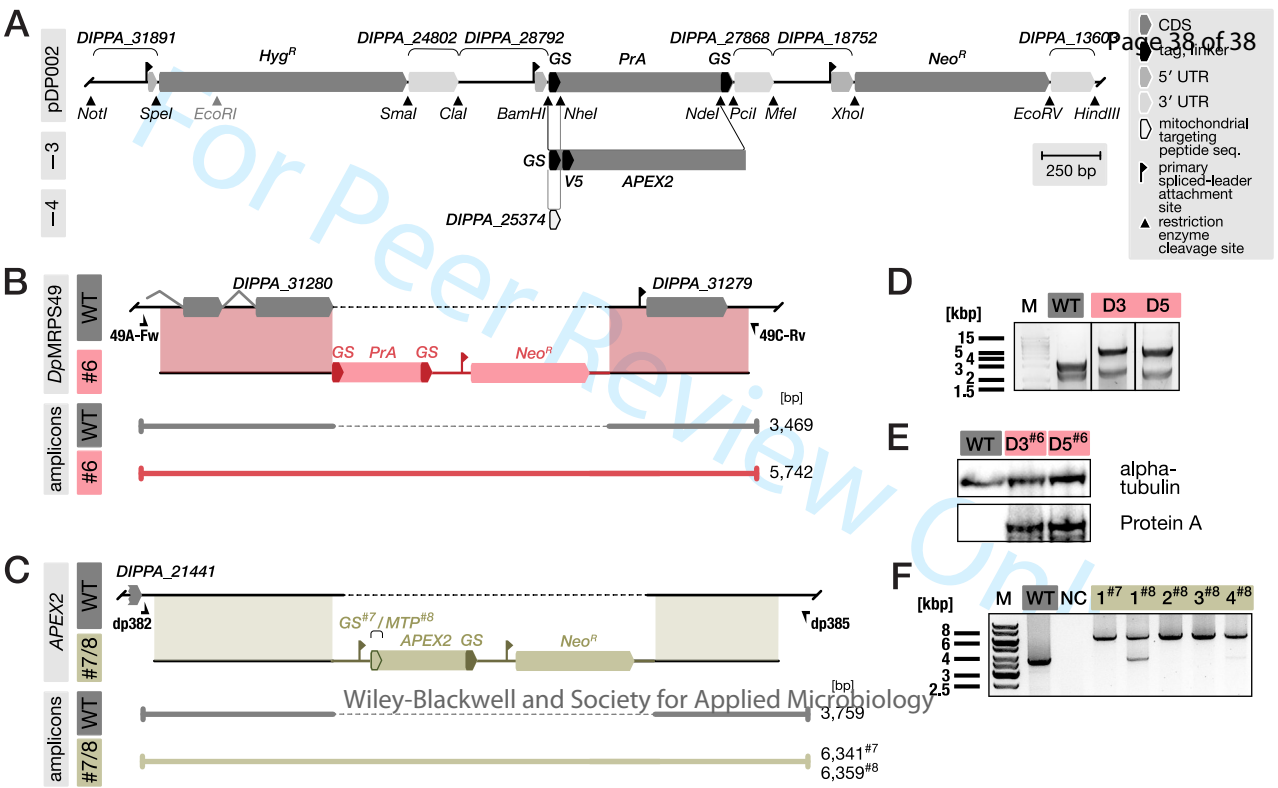












SUPPORTING INFORMATION

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,@], Corinna Benz[1], Gertraud Burger[3], Julius Lukeš[1,2,*]

[1] Institute of Parasitology, Biology Centre, Czech Academy of Sciences and [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

[#] Equal contribution
[@] Present address: Johannes Kepler University, Linz, Austria
* To whom correspondence should be addressed: Drahomíra Faktorová (dranov@paru.cas.cz), Julius Lukeš (jula@paru.cas.cz)

SUPPORTING INFORMATION

Table 1: List of primers used in this study.

PCR reaction	Primer name	Primer sequence (5′ to 3′)	PCR product size (bp)
Construct #1: Alpha-tubulin replacement with Puro^R			
PCR A	A-Fw	TCAGGTTGCCGGCATTTGGGAGCACAATCAG	2016
	A-Rv	TTTCGAAGACTTTTGGTTTTGATTTTTTGGTTTTCTGGAGAAATGGTTGT	
PCR B	O-Fw	AAATCAAAACCAAAAGTCTTCGAAAACTAGTATGACCGAGTACAAGCCCA	874
	K-Rv	GTGTGCGCGCTGACTGGTGAGCCGAAAAGTATCATTGCTGCCGGT	
PCR C	B-Fw	CACCAGTCAGCGCGCACACAGAAACACACATC	1840
	B-Rv	ACCGGCTACCACCTACTCCCGCTGCTTTATGTG	
PCR D (Nested)	S-Fw	TACAAGAAATTGAAGAACGATTCACTGGTAG	4223
	S-Rv	TGTAGACTGTTTCTGTTTGTTTGTTTCTTTC	
Construct #2: Beta-tubulin replacement with Puro^R			
PCR A	C-Fw	TGAGTACACGCGTTCGAGCAGCACCAATTCCAG	952
	C-Rv	AGAAAAGTGAGTAAATCCTGGTTGGTTGTTAA	
PCR B	Q-Fw	GATTTACTCACTTTTCTCAACTCGTCAAAACTAGTATGACCGAGTACAAG	889
	Q-Rv	GTGGGTCGCGGATAAGCTTCCAGGCGCATGCGCCGAAAAGTATCATTGCT	
PCR C	D-Fw	GCCTGGAAGCTTATCCGCGACCCACCTCAACTC	1333
	D-Rv	TTTGAGAACTTGTTAGGGGGCGTGATGGATTTAC	
PCR D (Nested)	T-Fw	CGTCCGTAAAGGGTTAAGCTATGCTCTTCAG	2928
	T-Rv	GGGAACTGAGAAACCGAATTGTTCATTTGTT	
Construct #3: Alpha-tubulin tagging with Ty-mCherry			
PCR A	A-Fw	TCAGGTTGCCGGCATTTGGGAGCACAATCAG	2016
	A-Rv	TTTCGAAGACTTTTGGTTTTGATTTTTTGGTTTTCTGGAGAAATGGTTGT	
PCR B	M-Fw	AAAACCAAAAGTCTTCGAAAATGACCGAGTACAAGCCCACGGTGC	2008
	M-Rv	TGGATCGAGATGCATTCCCTCATATGCACACGGGACCTGAGTCCT	
PCR C	I-Fw	ATGAGGGAATGCATCTCGATCCACATCGGC	1519
	I-Rv	GAAGGTAAGTTTTGAAGTCGTTCTGT	
PCR D (Nested)	W-Fw	GAACGTTTCTCGGTTTGATTCGCACAAAACT	5165
	W-Rv	ACATTCCTACCGTTCAGAAAGAGGGAGGAT	
Construct #4: Alpha-tubulin replacement with V5+Neo^R			
PCR A	A-Fw	TCAGGTTGCCGGCATTTGGGAGCACAATCAG	2016
	A-Rv	TTTCGAAGACTTTTGGTTTTGATTTTTTGGTTTTCTGGAGAAATGGTTGT	
PCR B	AK-Fw	AAATCAAAACCAAAAGTCTTCGAAAATGGGTAAGCCTATCCCTAACCCTC	989
	AK-Rv	TTTCTGTGTGCGCGCTGACTGGTGATTAGAAGAACTCGTCAAGAAGGCGA	
PCR C	B-Fw	CACCAGTCAGCGCGCACACAGAAACACACATC	1840
	B-Rv	ACCGGCTACCACCTACTCCCGCTGCTTTATGTG	
PCR D (Nested)	S-Fw	TACAAGAAATTGAAGAACGATTCACTGGTAG	4333
	S-Rv	TGTAGACTGTTTCTGTTTGTTTGTTTCTTTC	
Construct #5: Beta-tubulin replacement with V5+Neo^R			
PCR A	C-Fw	TGAGTACACGCGTTCGAGCAGCACCAATTCCAG	952
	C-Rv	AGAAAAGTGAGTAAATCCTGGTTGGTTGTTAA	
PCR B	BK-Fw	GGATTTACTCACTTTTCTCAACTCGTCAAAATGGGTAAGCCTATCCCTAACCCT	994
	BK-Rv	GTGGGTCGCGGATAAGCTTCCAGGCTTAGAAGAACTCGTCAAGAAGGCGA	
PCR C	D-Fw	GCCTGGAAGCTTATCCGCGACCCACCTCAACTC	1333
	D-Rv	TTTGAGAACTTGTTAGGGGGCGTGATGGATTTAC	
PCR D (Nested)	T-Fw	CGTCCGTAAAGGGTTAAGCTATGCTCTTCAG	3032
	T-Rv	GGGAACTGAGAAACCGAATTGTTCATTTGTT	

PCR reaction	Primer name	Primer sequence (5′ to 3′)	PCR product size (bp)
Construct #6: C-terminal tagging of DpMRPS49			
PCR A	49A-Fw	AGATGTTTGAAGCGCCTGGGTACCG	1741
	49A-Rv	CCCGACCCGGAGCCGCTGCCCATCGCGTCCCACTTCTGGAGATGC	
PCR B	protA-Fw	GGCAGCGGCTCCGGGTCGGG	2276
	protA-Rv	AAGCTTAAACTTGCAGCACGTAACAC	
PCR C	49C-Fw	GTGTTACGTGCTGCAAGTTTAAGCTTTAGTTTTCCGGGCCGGGAAAAGTGC	1771
	49C-Rv	TACTGGGCACCACTATCTCGGTGTC	
PCR D (Nested)	49D-Fw	CGGTTCCGCTGCACTACATCCATCG	5456
	49D-Rv	CACTACTGCGCTTGAGCAGGCTAGC	
Constructs #7 & 8: Integration of APEX2 into an intergenic region			
PCR A	dp382	TTCCGATCCATTGGGCGAAG	1639
	dp383	CTACTATCGATCACGAGCGGACTGACTAGAAA	
PCR C	dp384	AAGAGAAGCTTGCCAGTAGTTGTTGAGAGT	1958
	dp385	AGAGGTAACCTTTTGCTCAGTCCG	
Additional primers	mCherry_Fw	TTCATGTACGGCTCCAAGGC	
	mCherry_Rv	TAGTCCTCGTTGTGGGAGGT	
	Puro_Fw	ATGACCGAGTACAAGCCCAC	
	Puro_Rv	TCAGGCACCGGGGCTTGC	
	DpSL_Fw1	CCAACGATTTAAAAGCTACAGTTTCT	
	DpSL_Fw2	AAAAGCTACAGTTTCTGTACTTTATTG	
	DF_Fw1	AACTGTAATGCAAATTGCTAAGATGTTCACAC	
	DF_Rv1	AATCAGACGGTTCAAGTTGGTGTATGTC	
	SL_Puro_Rv1	GCTCGTAGAAGGGGAGGTTG	
	SL_Puro_Rv2	CGTGAGGAAGAGTTCTTGCAG	
	SL_mCherry_Rv1	CTTCAGCTTCAGCCTCTGCT	
	SL_mCherry_Rv2	AAGCGCATGAACTCCTTGAT	
	SL_Atubulin_Rv1	AACAGAGCTCCAAGACCAGAAC	
	SL_Atubulin_Rv2	GTCTCAGAGAAGAAGGTGTTGTAGG	
	SL_NeoR_Rv1	TTCAGTGACAACGTCGAGCA	
	SL_NeoR_Rv2	GAACCTGCGTGCAATCCATC	
	SL_protA_Rv1	TATCGGCCTTCGGTGCTTGGG	
	SL_protA_Rv2	GTTCCGTTGCGCCTCGTTGAG	
	dp375	GGTGGGAACAACGAGGCAC	

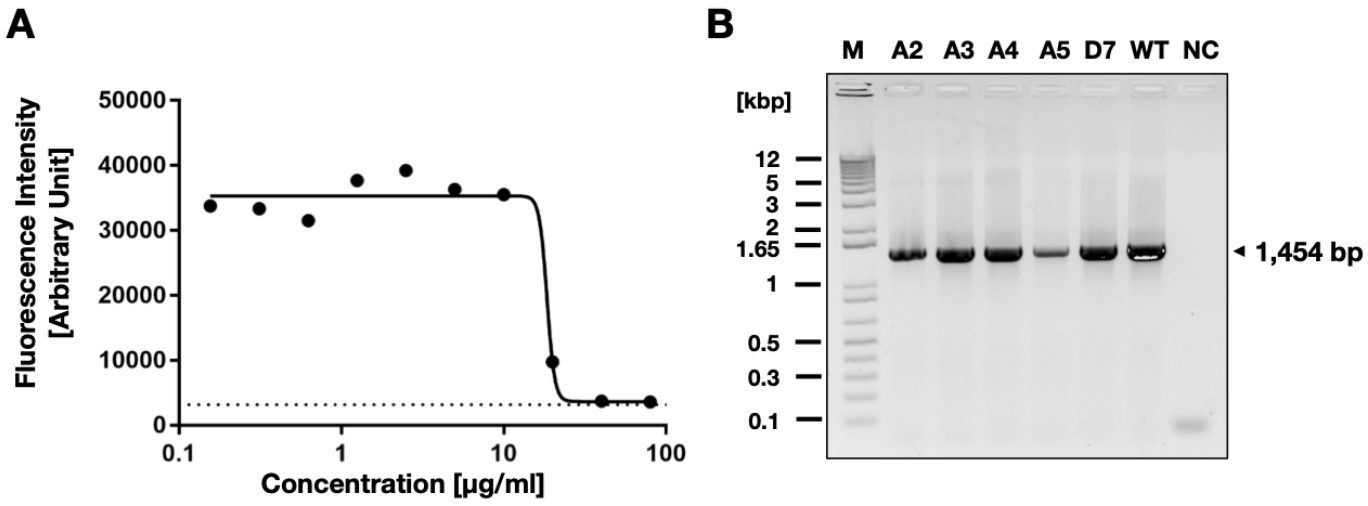


Figure S1: Sensitivity of *D. papillatum* to the NHEJ inhibitor W7 and its effect on the correct integration of the DF_Dp_01 construct.
A. Effect of increasing concentrations of W7 inhibitor (µg/ml shown; x-axis) on the survival of *D. papillatum* cells was tested using the Alamar blue assay, which measures the cell viability by fluorescence (for details, see **Experimental procedures**).
B. Integration of the DF_Dp_01 construct (Kaur *et al*., 2018) tested by PCR on genomic DNA using DF_Fw1 and DF_Rv1 primers. The expected amplicon size of the correctly integrated construct is 3,425 bp.

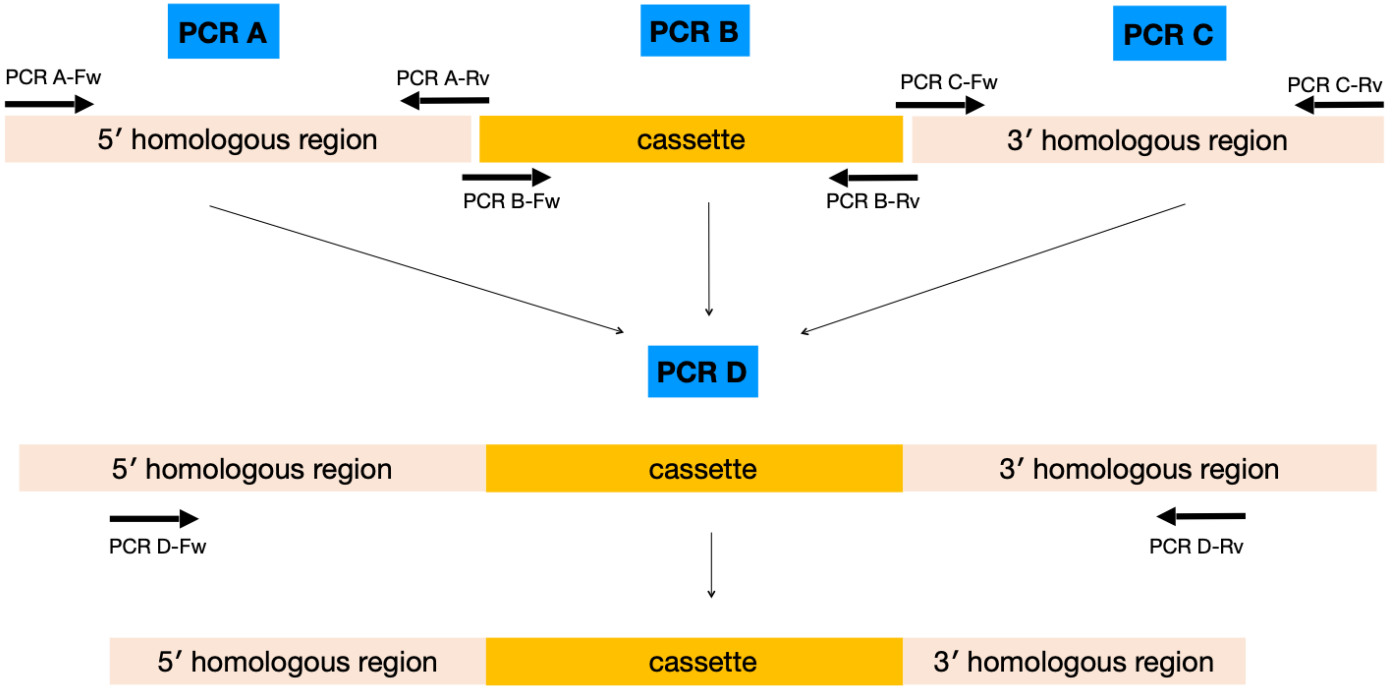


Figure S2: Schema of the fusion PCR approach used for construct preparation.
All transformation cassettes (except for the APEX2 constructs) were prepared by this PCR approach. First, three individual PCRs were used to amplify the 5′ long homology region (PCR A), cassette designed to replace/tag a gene of interest (PCR B), and 3′ long homology region (PCR C). Nested primers (PCR D-Fw and -Rv) were then used to join all three pieces in the final amplicon (PCR D). Used primers are listed in **Table S1**. For details, see **Experimental procedures**.

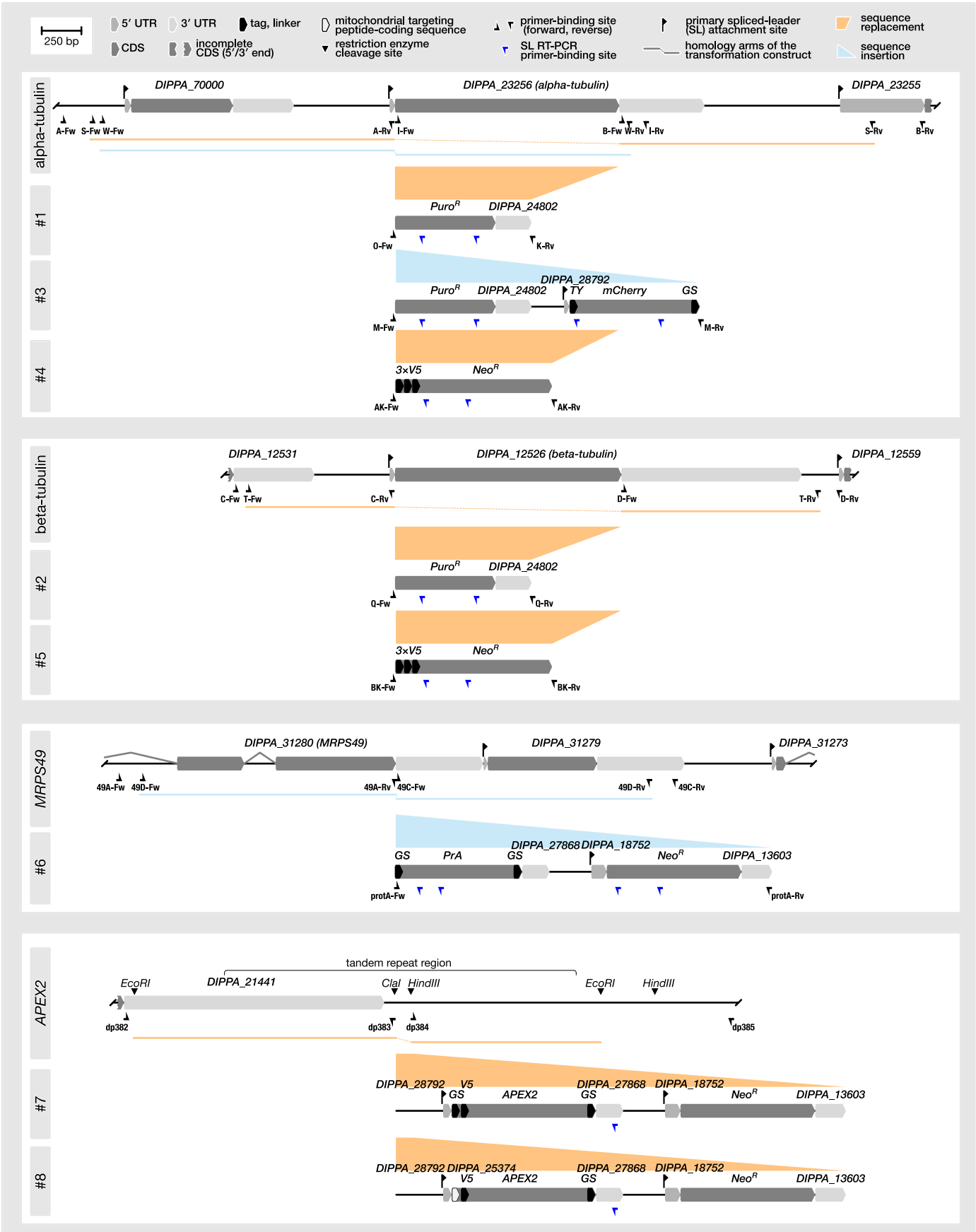


Figure S3: In-scale schemas of targeted genomic regions of *D. papillatum* and all the constructs (#1 to 8) used in this study. For details, see **Experimental procedures**.

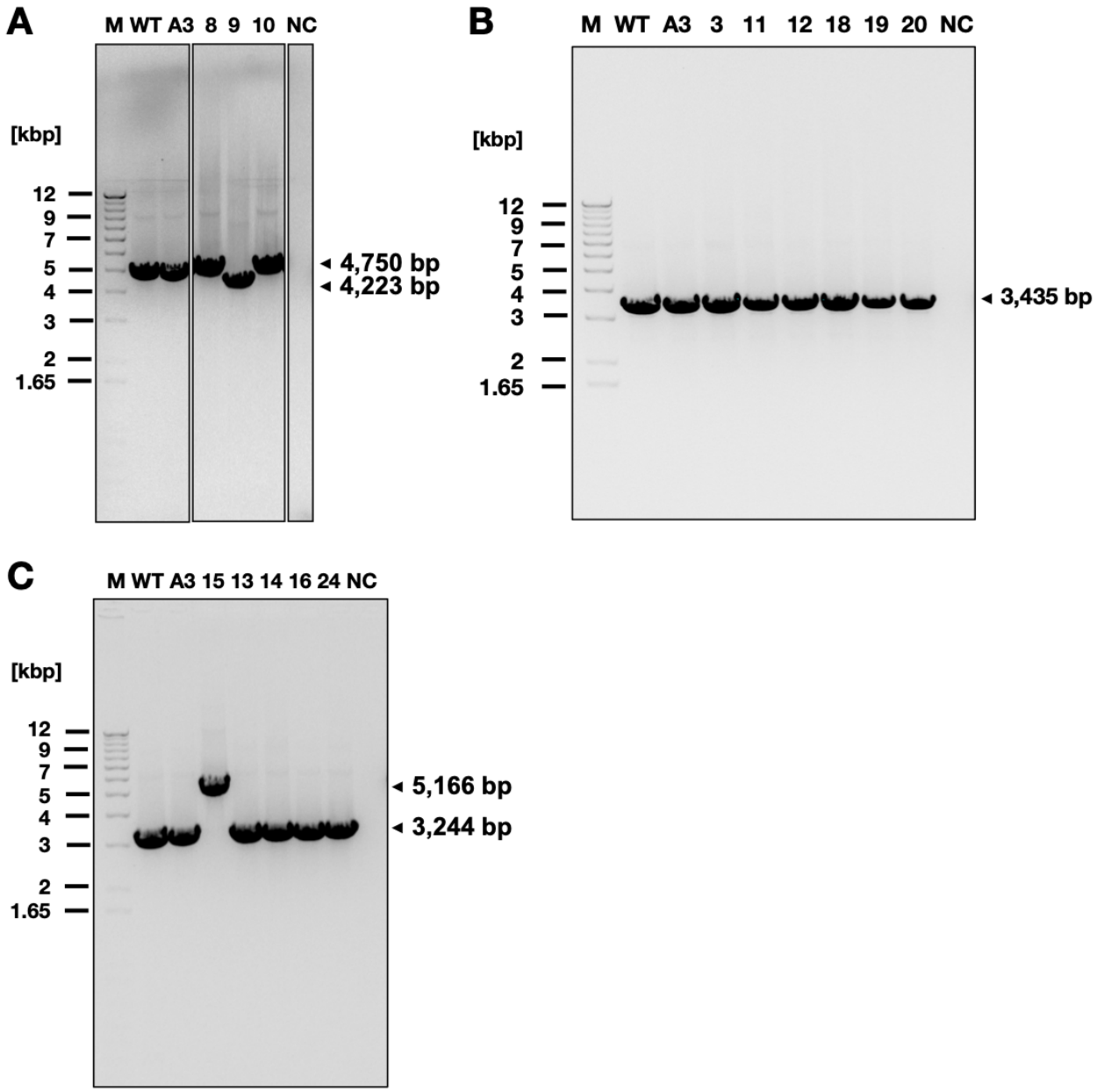


Figure S4: Verification of integration by genomic PCR.
PCR of total DNA of *D. papillatum* wild type (WT), cell line A3 (from a previous study by Kaur *et al*., 2018) and selected transformant cell lines obtained from the electroporation of the constructs #1 (alpha tubulin replacement with *Puro*^{R}; **A**), #2 (beta tubulin replacement with *Puro*^{R}; **B**), and #3 (N-terminal tagging of alpha tubulin with Ty-mCherry; **C**) using corresponding PCR D primers (see **Fig. S3** and **Table S1**). The expected amplicon size of the correctly integrated construct #2 (**B**) is 2,928 bp. Negative control (NC) was performed without template DNA.

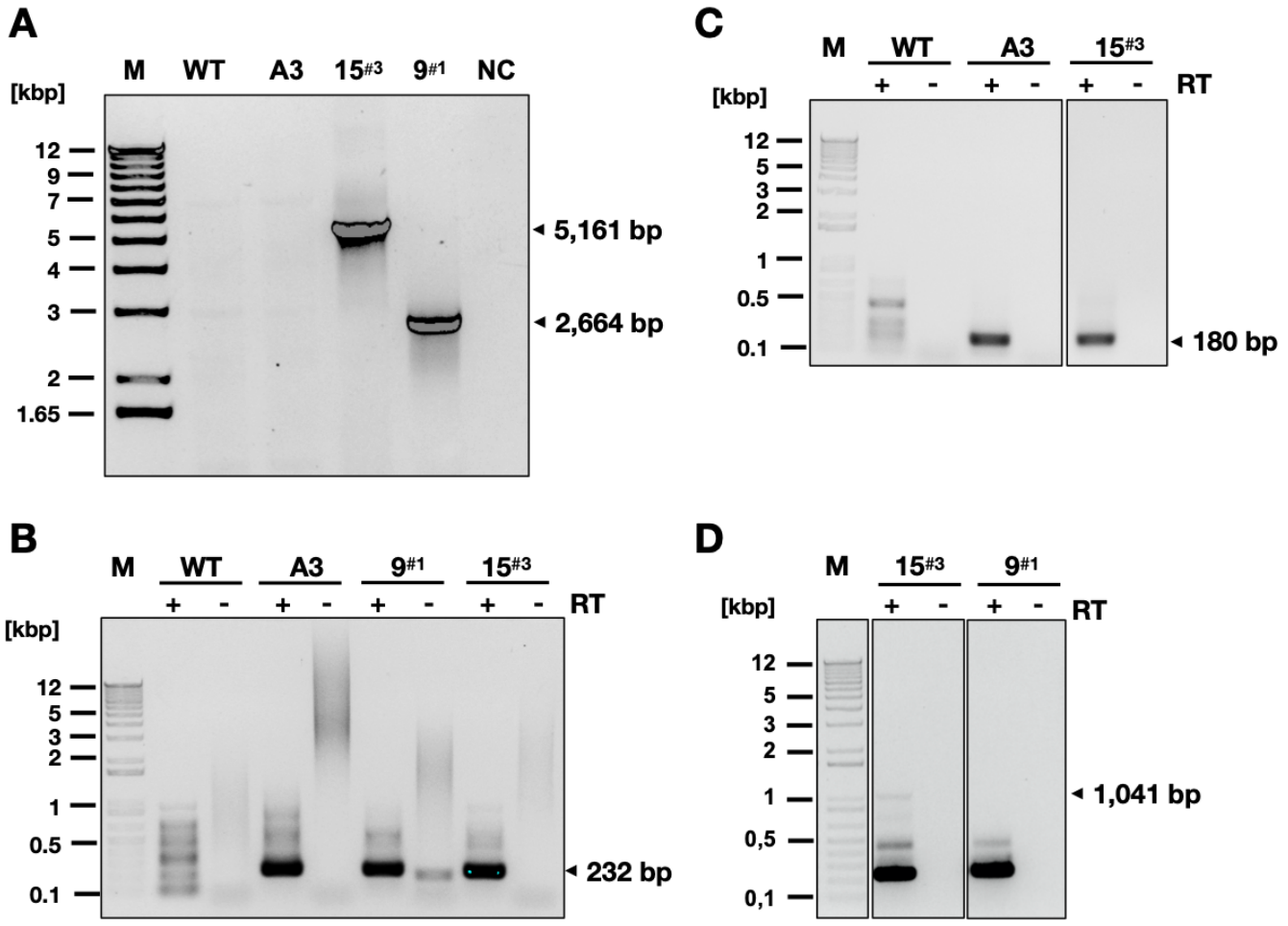


Figure S5: Genomic PCR and SL RT-PCR of transformant cell lines $9^{\#1}$ and $15^{\#3}$.
A. PCR of total DNA of *D. papillatum* wild type (WT), cell line A3 (from the previous study by Kaur *et al*., 2018) and transformants $9^{\#1}$ and $15^{\#3}$ using Puro_Fw and B_Rv primers (see **Fig. S3** and **Table S1**). The negative control (NC) was performed without template DNA.
B–D. Amplification of the 5′ region of *Puro*^{R} (**B**), *mCherry* (**C**), and *mCherry-tubulin* (**D**) transcripts. Nested SL RT-PCR was performed using total RNA from transformant cell lines $9^{\#1}$ and $15^{\#3}$ and two sets of primers that anneal to the conserved spliced-leader (SL) sequences of *D. papillatum* (forward primers) and to the 5′ proximal region of the transcripts (reverse primers). For details, see **Fig. S3** and **Table S1**.

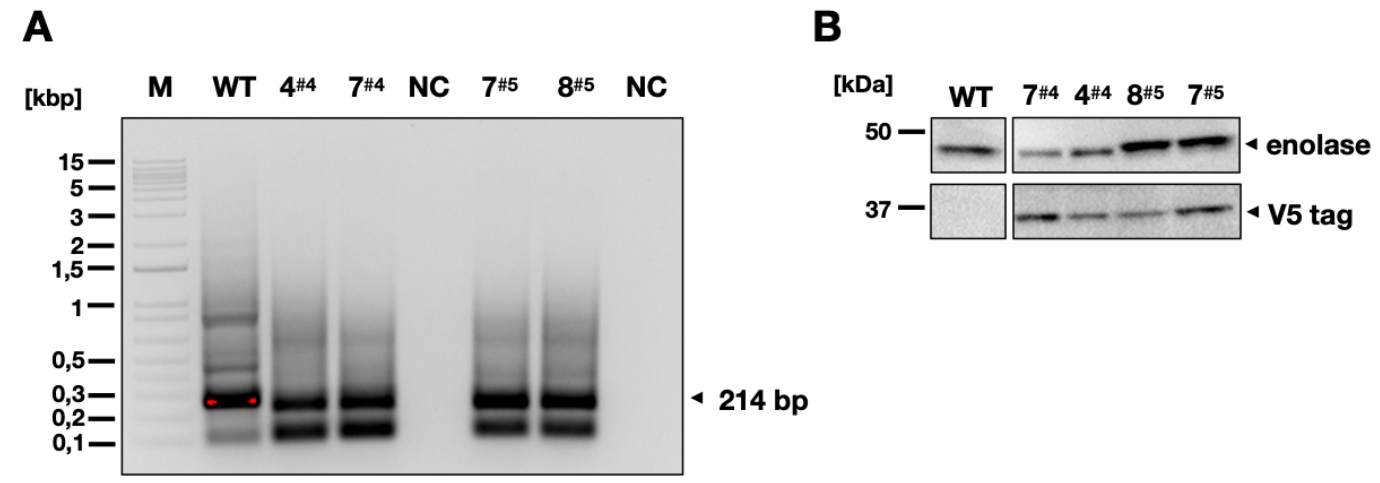


Figure S6: Expression of V5+Neo^{R} constructs.
A. SL RT-PCR of alpha (#4) and beta (#5) tubulin replacement with V5+Neo^{R}.
B. Western blot using the anti-V5 antibody. Enolase was used as a loading control.

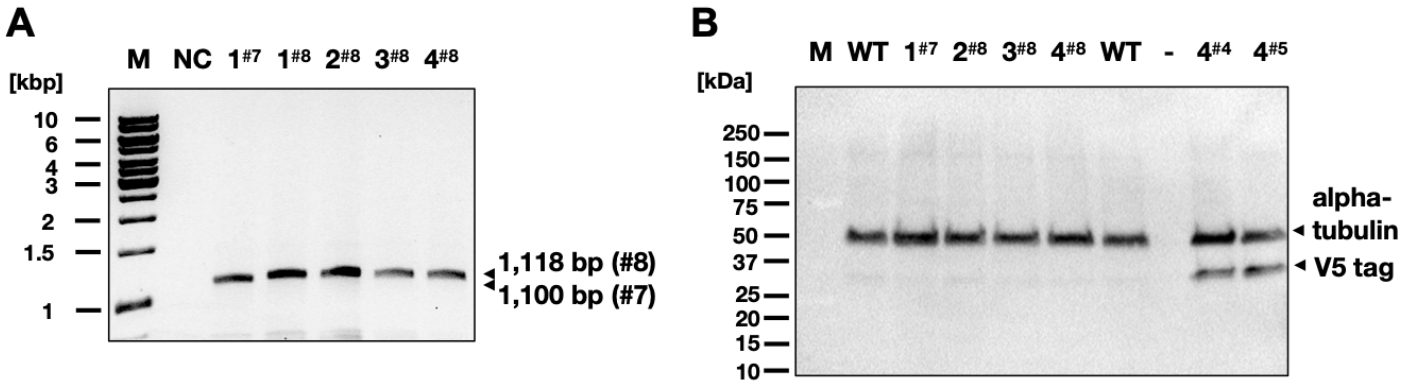


Figure S7: Expression of APEX2 in selected transformant cell lines.
A. SL RT-PCR using DpSL_Fw1 (dp400) and dp375 primers. For details, see **Fig. S3** and **Table S1**.
B. Western blot analysis using the anti-V5 antibody. Cell lines 4$^{#4}$ and 4$^{#5}$ were used as positive controls. Tubulin serves as a loading control.

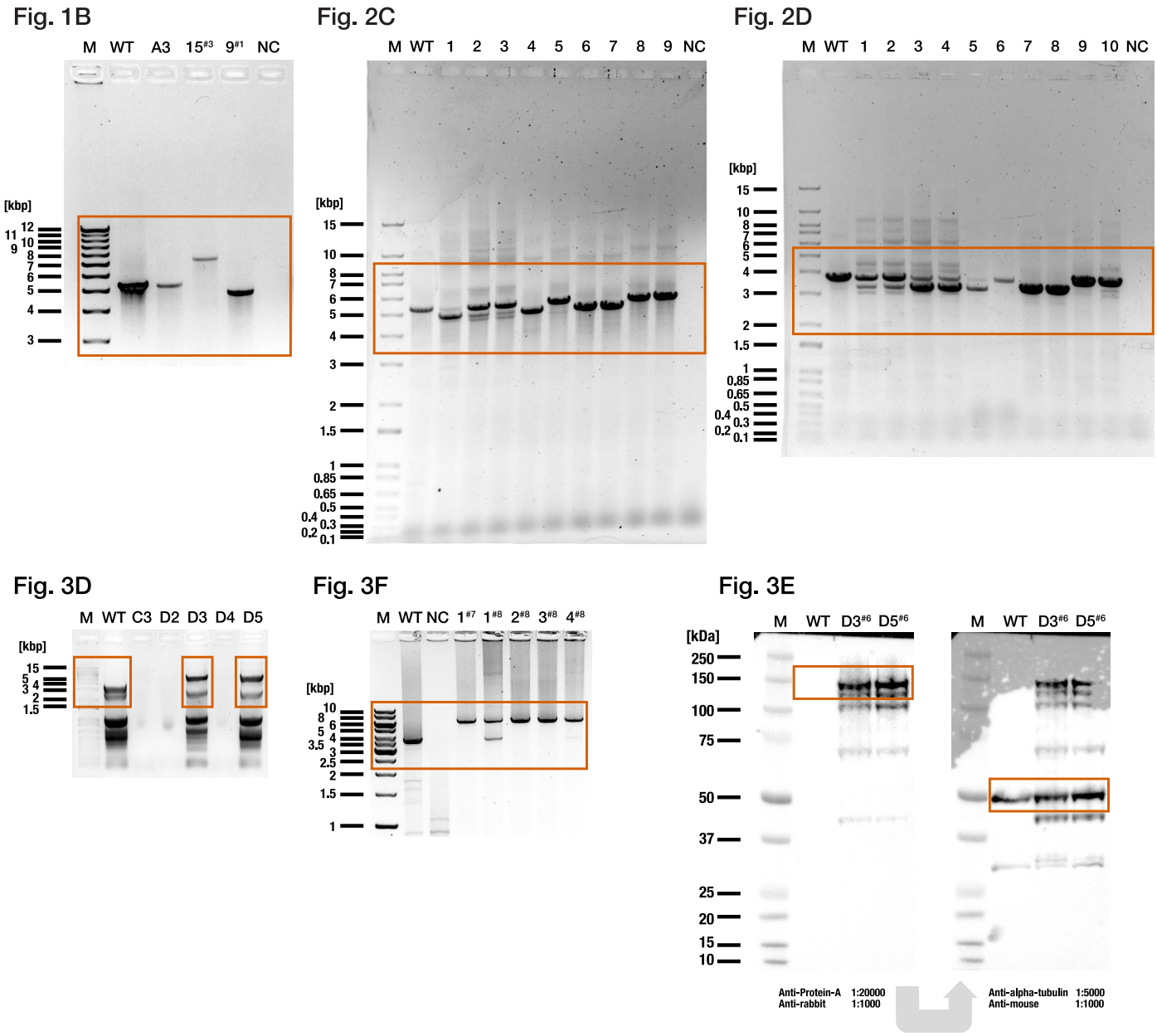


Figure S8: Uncropped gels (from **Fig. 1B**, **2C**, **2D**, **3D**, and **3F**) and Western blot (from **Fig. 3E**). Red frames indicate parts shown in **Figures 1–3**.

Fig. 1C

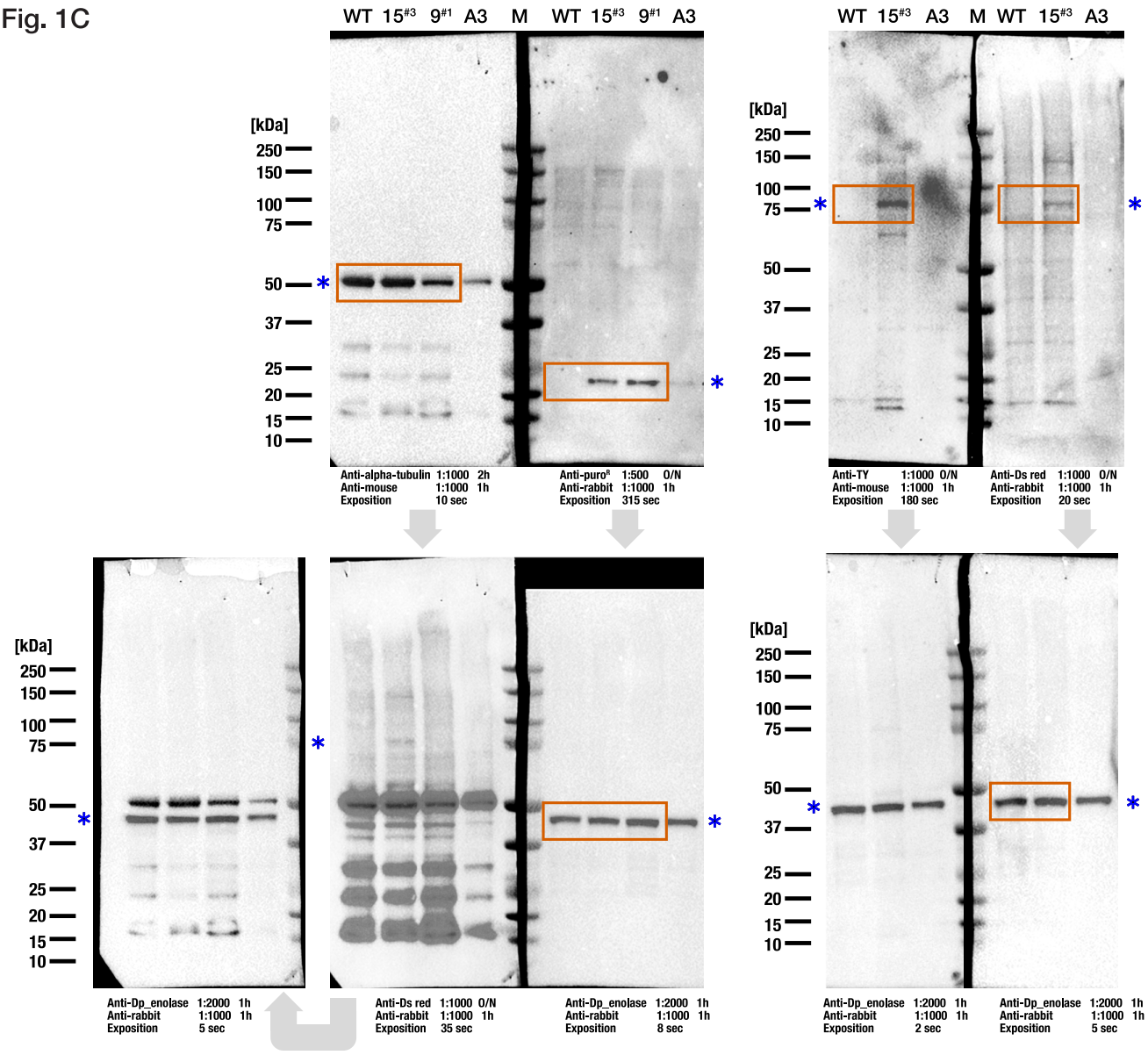


Fig. 2E

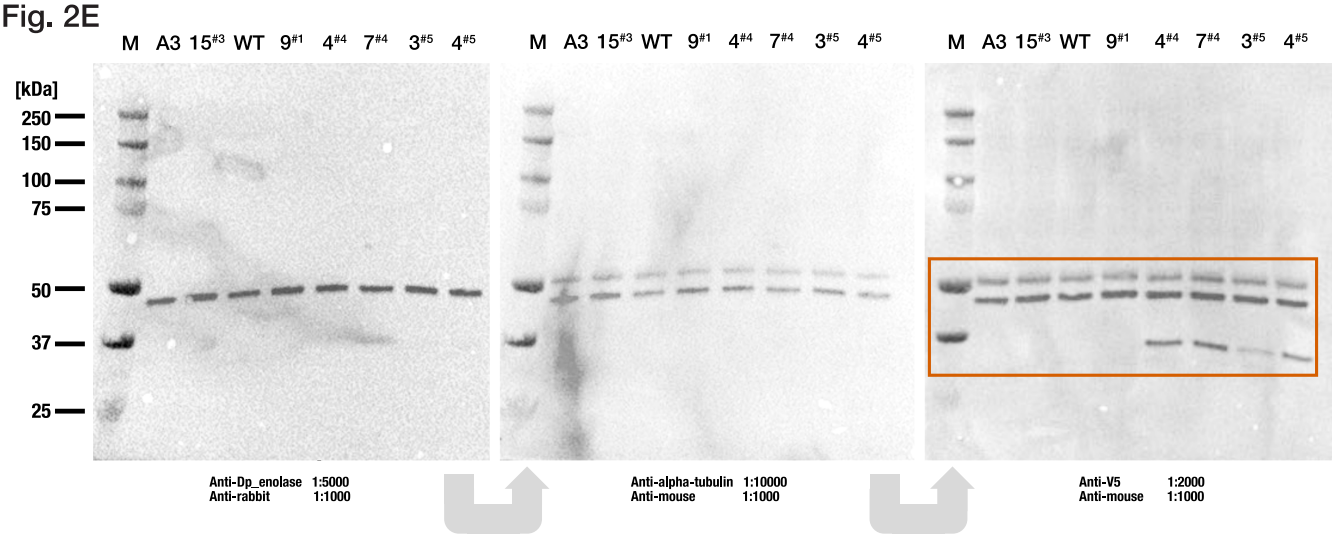


Figure S9: Uncropped Western blots showing the successive probing of the membranes (from **Fig. 1C** and **2E**). Red frames indicate parts shown in **Figures 1–2**. Blue asterisks mark the expected size of the target protein.