NTNU
Norwegian University of
Science and Technology

Functional studies of two silicon transporter-associated kinases in the diatom *Thalassiosira pseudonana*

Maren Moen Skisland

M.Sc in Cell- and Molecular Biology
Submission date: June 8th 2021 (Trondheim)
Supervisor: Olav Vadstein
Co-supervisor: Tore Brembu

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biology

Acknowledgements

This master's thesis was conducted at the "Analysis and Control of Microbial Systems" (ACMS) group at the Norwegian University of Science and Technology (NTNU) as a part of the requirements for receiving the degree of Master of Science (M.Sc) in Cell- and Molecular Biology.

First and foremost, I would like to thank my supervisors, Professor Olav Vadstein and Research Scientist Tore Brembu for their help and guidance throughout my thesis. I especially want to thank Ph D. student Annika Messemer for all her help in the laboratory, both by being there when I had questions and by teaching me new methods. Additionally, I would like to thank other members of the ACMS group for their help and advice during the day-to-day lab work.

Lastly, I would like to thank my family and friends for all their love and support.

Abbreviations

ABC transporter	ATP binding cassette transporter
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Basepair
CAH	CEN6-ARSH4-HIS3
CaMK	Ca^{2+}/calmodulin-dependent protein kinase
Cas	CRISPR associated protein
cDNA	Complementary DNA
CRISPR	Clustered regularly interspaced short palindromic repeats
DAPI	4′,6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
DSB	Double-stranded break
EDTA	Ethylenediaminetetraacetic acid
f/2 medium	Nutrient-enriched seawater medium
FCPB	Fucoxanthin Chlorophyll a/c-binding Protein B
gDNA	Genomic DNA
HRM	High resolution melting
ICE	Interference of CRISPR edits
kb	Kilobase
KO	Knockout
LB medium	Luria-Bertani medium
LHCF	Major fucoxanthin Chla a/c protein
mNG	mNeonGreen
MQ water	Milli-Q water
mRNA	Messenger Ribonucleic acid
m.y.a	million years ago
Nat	N-acetyl transferase
NHEJ	Non-homologous end joining
NLS	Nuclear localization signal
Nou	Nourseothricin
NTC	No template control
OD	Optical density
PAM	Protospacer adjacent motif
PAR	Photosynthetically active radiation

PCR	Polymerase chain reaction
qRT-PCR	Quantitative real-time PCR
RNA	Ribonucleic acid
-RT	Reverse transcriptase
SD	Standard deviation
SDV	Silica deposition vesicle
sgRNA	Single guide RNA
SIT	Silicon transporter
SOC	Super optimal broth
TAE buffer	Tris-acetate-EDTA buffer
T_m	Melting temperature
TMD	Transmembrane domain
UV	Ultraviolet
WT	Wild type
YFP	Yellow fluorescent protein

Abstract

Diatoms consist of a large and diverse group of unicellular eukaryotes. These organisms are responsible for a substantial amount of the primary production in the marine environment, as well as having a great impact on several global systems. This includes the climate system by affecting the concentration of carbon dioxide in the atmosphere, as well as biological systems such as the flux of carbon and nutrients in the oceans. Their highly ornate and silicious cell wall, the frustule, makes them very attractive for studies involving processes such as biomineralization and biosynthesis.

The two closely related kinase-encoding genes that were studied in this thesis, *Tp264671* and *Tp14322*, were found through a transcriptomics analysis. These two genes have been shown to be very closely co-expressed with silicon transporters SIT1 and SIT2, respectively. Protein alignments of the genes found that CaMK was the catalytic domain, in addition to several regions with conserved amino acids that appeared in multiple alignments of the kinase domains.

The goal of this thesis was to create mutant cell lines of the diatom *Thalassiosira pseudonana* by knocking out *Tp264671* and *Tp14322* (as well as *Tp14242*, another closely related kinase) with the CRISPR/Cas9 technology. The similarity in the sequences of these three genes makes it possible to knock out multiple genes at once by using target sites found in all of them. Target sites were inserted into the vector pTpPUC3-Cas9-M-G1 and transferred into *T. pseudonana* by bacterial conjugation. Mutant screening of the algae cultures indicated that mutations had occured for clones containing certain target sites, although Sanger sequencing confirmed no mutations had happened. Problems with establishing mutants could be due to the genes being essential for the cell to work properly, or perhaps due to issues with plasmid efficiency (which could be controlled by multiple things such as plasmid components and sequence composition). Gene expression analysis (along with microscopy results), showing low expression of *Cas9* and the sgRNA, indicated that there might not have been enough mRNA from the gene-editing components of the plasmid to create knockout mutants. Low expression of the nourseothricin resistance gene implied that there might not have been enough antibiotics in the medium to create a sufficient selection pressure for the *T. pseudonana* cultures to keep the episome.

Another goal was to fuse *Tp264671* and *Tp14322* with the fluorescent protein mNeonGreen, followed by inserting the genes (with the tag) into the pTpPUC3 vector. None of the attempts to achieve this were successful and a microscopy study of expression and localization of the genes in *T. pseudonana* could therefore not be done.

Abstrakt

Kiselalger består av en stor og mangfoldig gruppe av encellede eukaryoter. Disse organismene står for en betydelig mengde av den primære produksjonen i det marine systemet, i tillegg til at de har stor innvirkning på flere globale systemer. Kiselalger påvirker blant annet klimasystemet ved at de er med på å regulere konsentrasjonen av karbondioksid i atmosfæren, samt biologiske systemer som omrøring av karbon og næringsstoffer i havene. Celleveggen (frustulen), som består av intrikate mønstre av silisium, gjør dem veldig attraktive for studier som involverer prosesser som biomineralisering og biosyntese.

To nært beslektede kinase-kodende gener som ble studert i denne oppgaven, *Tp264671* og *Tp14322*, ble funnet gjennom en transkriptomisk analyse. Disse to genene har vist seg å være ko-uttrykt med henholdsvis silisiumtransporter SIT1 og SIT2. Ved sammenligning av protein sekvensen til genene ble det funnet at CaMK var det katalytiske domenet, i tillegg til flere områder med konserverte aminosyrer som dukket opp i flere av sekvensanalysene av kinasedomenene.

Målet med denne oppgaven var å lage mutante cellelinjer av kiselalgen *Thalassiosira pseudonana* ved å slå ut genene *Tp264671* og *Tp14322* (samt *Tp14242*, en annen nær slektning) med CRISPR/Cas9-teknologien. Likheten i sekvensene til disse tre genene gjør det mulig å slå ut flere gener samtidig ved å bruke «target sites» som finnes i alle de tre genene. Ulike target sites ble satt inn i vektoren pTpPUC3-Cas9-M-G1 og overført til *T. pseudonana* ved konjugering via bakterien *Escherichia coli*. Mutant screening av algekulturene indikerte at mutasjoner hadde skjedd hos kloner som inneholdt spesifikke target sites, men Sanger-sekvensering avkreftet dette. Problemer med å etablere mutanter kan skyldes ved at genene er essensielle for at cellen skal kunne leve, eller kanskje på grunn av problemer med plasmideffektivitet (som kan kontrolleres av flere ting som plasmidkomponenter og sekvensstruktur). Genekspresjonsanalyse (sammen med mikroskopiresultater) som viste lavt uttrykk av Cas9 og sgRNA indikerte at det kanskje ikke var nok RNA til stedet for å produsere mutanter, og lavt uttrykk av nourseothricin-resistensgenet antydet at det kanskje ikke var nok antibiotika i mediet for å skape et tilstrekkelig seleksjonstrykk for *T. pseudonana*-kulturene.

Et annet mål var å tagge *Tp264671* og *Tp14322* med det fluorescerende proteinet mNeonGreen for så å sette inn genene i pTpPUC3-vektoren. Ingen av forsøkene på å oppnå dette var vellykkede, og mikroskopering av uttrykk og lokalisering av genene i *T. pseudonana* kunne derfor ikke gjøres.

Table of contents

Acknowledgements ... 2
Abbreviations ... 3
Abstract ... 5
Abstrakt ... 6
List of figures ... 9
List of tables ... 11
1 Introduction ... 12
1.1 Diatoms and their evolution ... 12
1.2 The frustule ... 14
1.3 Thalassiosira pseudonana ... 16
1.4 Silicic acid transporters ... 17
1.5 Kinases ... 19
1.6 Characterization of kinases co-expressed with TpSITs ... 21
1.7 CRISPR/Cas9 ... 22
1.8 Thesis objective ... 24
2 Materials and methods ... 25
2.1 Kinase domain analysis ... 25
2.2 Creating knockout mutants of *Tp14322* and *Tp264671* (and *Tp14242*) ... 25
2.2.1 Selection of PAM sites and primers ... 25
2.2.2 Modifying vector to create knockout cell lines ... 26
2.2.3 Transformation of DH5α and DH10β *Escherichia coli* (*E. coli*) with pTpPUC3 ... 28
2.2.4 Transfer of plasmid DNA from *E. coli* into *T. pseudonana* by conjugation ... 29
2.2.5 Amplification of 800-1000 bp amplicons before mutant screening ... 31
2.2.6 Screening for mutants with high resolution melting (HRM) analysis ... 32
2.2.7 Plasmid check ... 34
2.2.8 Screening for mutations in colonies from conjugation plate ... 34
2.2.9 Analysis of gene expression in *T. pseudonana* ... 35
2.2.10 Fluorescence microscopy ... 38
2.3 Fusion of mNeonGreen with *Tp14322* and *Tp264671* ... 39
2.3.1 Vector and linearization ... 39
2.3.2 Amplification of gene fragments and mNeonGreen ... 40
2.3.3 Cloning methods for inserting fragments into pTpPUC3 ... 41
3 Results ... 48

3.1 Kinase domain analysis ... 48
3.2 Creating knockout cell lines of *Tp14322* and *Tp264671* ... 50
3.2.1 Vector construction ... 50
3.2.2 Screening for mutations in *T. pseudonana* transformants ... 51
3.2.3 pTpPUC3-Cas9-M-G1 plasmid confirmation ... 59
3.2.4 Gene expression analysis with qRT-PCR ... 60
3.2.5 Microscopy of triple knockout PAM clones ... 62
3.3 Fusion of fluorescent protein mNeonGreen with *Tp14322* and *Tp264671* ... 64
3.3.1 Fragment amplification ... 64
3.3.2 pTpPUC3 plasmid confirmation ... 65
3.3.3 Cloning fragments into pTpPUC3 ... 66
4 Discussion ... 67
4.1 CaMK is the catalytic domain in *Tp264671* and *Tp14322* ... 67
4.2 Plasmid and mutation efficiency ... 68
4.3 Low expression of Cas9 and the nourseothricin resistance gene ... 70
4.4 Fluorescent protein cloning ... 72
5 Future aspects ... 73
6 Conclusion ... 74
References ... 75
Appendices ... 78
Appendix 1: Media and solutions ... 78
Appendix 2: Primers used for vector modification ... 80

List of figures

Introduction

Figure 1.1. The evolutionary history of diatoms.

Figure 1.2. Overview of the different structures of the diatom frustule

Figure 1.3. Schematic representation of the cell cycle and frustule formation of *T. pseudonana.*

Figure 1.4. Images of dried frustules of *T. pseudonana* showing the shape of the species.

Figure 1.5. Model of SIT-mediated silicon transport and transmembrane domains in SITs.

Figure 1.6. Overview of protein phosphorylation by a catalytic kinase and dephosphorylation by a phosphatase.

Figure 1.7. Phylogenetic tree based on protein alignments of *Tp14322* and *Tp264671*.

Figure 1.8. Representation of how the CRISPR/Cas9 complex binds to the target DNA to produce DSB, and DSB reparation methods.

Materials and methods

Figure 2.1. Schematic maps of the two kinase-encoding genes *Tp14322* and *Tp264671*.

Figure 2.2. Map of the pTpPUC3-Cas9-M-G1 vector used for mutant generation.

Figure 2.3. Map of the pTpPUC3 vector used for fluorescent protein fusion.

Figure 2.4. General overview of amplified fragments for fluorescent protein fusion and their structure in the vector.

Figure 2.5. Schematic representation of the Gibson assembly cloning method.

Results

Figure 3.1. Map of the conserved amino acids in the catalytic CaMK domain in *Tp264671* and *Tp14322*.

Figure 3.2. Colony screening of DH5α E. coli transformed with pTpPUC3 plasmid inserted with the different target sites.

Figure 3.3. Gel electrophoresis analysis of amplified regions from clones containing target sites for *Tp264671* r42 and *Tp14322* r77.

Figure 3.4. Gel electrophoresis analysis of amplified regions from clones containing target sites for *Tp264671* r291 and *Tp14322* r392.
Figure 3.5. High resolution melting analysis of single KO clones with target sequences *Tp264671* r42, *Tp264671* r291, *Tp14322* r77 and *Tp14322* r392.
Figure 3.6. Gel electrophoresis analysis of amplified regions from clones containing target site for f436 in both *Tp264671* and *Tp14322*.
Figure 3.7. High resolution melting analysis of double KO clones with the target sequence f436 for both *Tp264671* and *Tp14322*.
Figure 3.8. Gel electrophoresis analysis of amplified regions from clones (from the first conjugation) containing the target site f233 for *Tp264671*, *Tp14322* and *Tp14242*.
Figure 3.9. Gel electrophoresis analysis of amplified regions from clones (from the second conjugation) containing the target site f233 for *Tp264671*, *Tp14322* and *Tp14242*.
Figure 3.10. Gel electrophoresis analysis of plasmids from different parts of the mutant generation process cut with BamHI-HF.
Figure 3.11. Relative gene expression analysis of different genes in the pTpPUC3-Cas9-M-G1 plasmid.
Figure 3.12. Microscopy photos of clone 4 straight after the second conjugation.
Figure 3.13. Microscopy photos of clones from the second conjugation stained with DAPI.
Figure 3.14. Gel electrophoresis analysis of amplified fragment for fluorescent protein fusion.
Figure 3.15. Gel electrophoresis analysis of different pTpPUC3 plasmids used in the fluorescent protein fusion experiments.
Figure 3.16. Gel electrophoresis analysis of Gibson products both before and after purification.

List of tables

Materials and methods

Table 2.1. Red Taq PCR reaction.

Table 2.2. PCR program for colony screening with Red Taq.

Table 2.3. Phusion PCR reaction setup (20 µl)

Table 2.4. PCR thermoprofile for amplifying target sites with Phusion DNA polymerase.

Table 2.5. HRM master mix reaction setup.

Table 2.6. Thermoprofile for DNA amplification and HRM analysis.

Table 2.7. Q5 Master Mix reaction setup.

Table 2.8. PCR thermoprofile for amplifying target sites with Phusion DNA polymerase.

Table 2.9. QuantiTect reverse-transcription reaction.

Table 2.10. SYBR Green qRT-PCR reaction mix.

Table 2.11. Thermoprofile for qRT-PCR.

Table 2.12. Phusion PCR reaction (50 µl).

Table 2.13. Gibson assembly reaction mix.

Table 2.14. Gibson assembly reaction mix with purified fragments.

Table 2.15. DreamTaq PCR reaction setup.

Table 2.16. Thermoprofile for nested Gibson assembly.

Table 2.17. CPEC PCR reaction.

Table 2.18. CPEC PCR program.

Results

Table 3.1. Deviations in amino acids in different subdomains of the catalytic CaMK domain.

Table 3.2. Overview of amino acids found in the different PAM sites.

1 Introduction

1.1 Diatoms and their evolution

Diatoms are unicellular photoautotrophic eukaryotes that are considered the most successful group of eukaryotic phytoplankton in the ocean after they rose to their prominence about 100 million years ago (Falciatore & Bowler, 2002; Bowler et al., 2010; Armbrust, 2009). There are estimated to be around 100,000 different species of diatoms, where the size of the species can vary from a few micrometers to a few millimeters. They are found existing as single cells, colonies, or several cells connected to each other in a chain-like structure (Mann and Vanormelingen, 2013; Nelson et al., 1995). Diatoms are responsible for approximately 40% of all organic matter generated by photosynthetic production in the marine environment, and 20% of the global primary production. They therefore have great influence on global climate, function of the marine ecosystem, and concentrations of carbon dioxide in the atmosphere (Nelson et al., 1995; Armbrust, 2009). Due to one of their main characteristics, their heavy siliceous external wall (the frustule), diatoms are for example thought to have a dominant role in the biological carbon pump. Their importance in the carbon pump system is based on how the gravitational settling of dead and dying diatoms generates a net downward carbon flux in the ocean, which is a process that transfers carbon, electrons, and energy to the deeper parts of the ocean. For the organisms living beneath the photic zone of the water column, carbon flux is essential to survive (Bowler et al., 2009). In addition to their ecological importance, diatoms are of great interest in the field of biotechnology. By understanding the intricate process of frustule synthesis, we might discover new nanotechnological applications as well. Diatoms also produce highly unsaturated fatty acids that can be used as an alternative source of biofuels (Kroth, 2007; Dunahay et al., 1996).

The rise of diatoms and plants came from different endosymbiotic events (Figure 1.1), where cyanobacteria were found to be the origin of the oxygenic photosynthesis leading to the development of these organisms. Around 1.5 billion years ago the first initial primary endosymbiosis occurred, where a cyanobacterium was engulfed by (or invaded) a eukaryotic heterotroph. This merge formed the photosynthetic plastids of Plantae, a group consisting of red and green algae, as well as land plants (Yoon et al., 2004). The plastid acquired a photosynthetic apparatus when it became housed within an organelle containing two membranes (Prihoda et al., 2012). Approximately 10% of the nuclear genes of Plantae is derived from the endosymbiotic cyanobacteria after its genome was transferred to the host nucleus during this event (Reyes-Prieto et al., 2006). Under the second endosymbiosis, which happened around 500 million years later, a red alga was captured by a different heterotroph. As gene transfer went on from the nuclear and plastid genomes of the red alga to the host nucleus, the red algal endosymbiont was over time transformed into the plastids of the Stramenopiles. This group includes brown macroalgae, diatoms and plant parasites (Armbrust et al., 2004). The plastids in many of the organisms

that evolved from the second endosymbiotic event are surrounded by four membranes, rather than only the two membranes which was characteristic for the primary endosymbiotic event. These four membranes correspond to the exosymbiont endomembrane (membrane furthest out), the plasma membrane of the engulfed red alga, along with the two membranes of the primary plastid from the first event (inner membranes) (Prihoda et al., 2012). Although the diatom plastid is derived from red algae, green algae has also contributed with a significant amount of nuclear genes in diatoms (Becker et al., 2008). An analysis by Moustafa et al. (2009) estimates that 16% of the diatom nuclear coding potential is derived from green algae after identifying over 1700 green gene transfers.

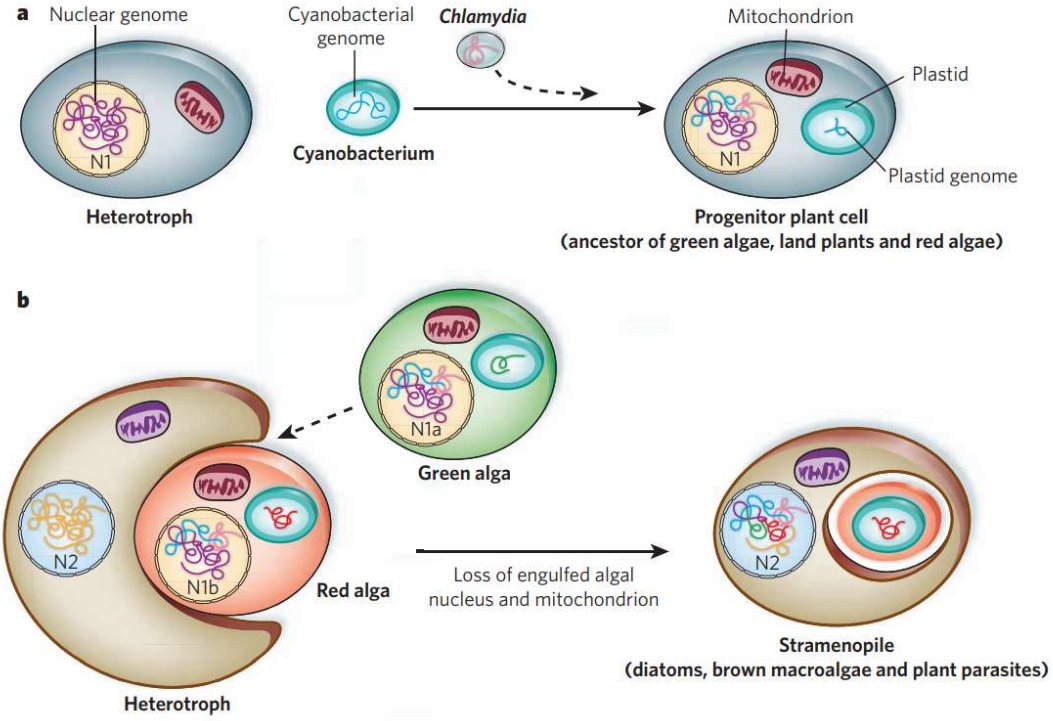

Figure 1.1. *The evolutionary history of diatoms. a) shows the primary endosymbiosis where the host nucleus (N1) receives the genome from the cyanobacteria. The plastid genome in the progenitor plant cell has retained few of the original genes and has further diverged into green and red alga as well as land plants. b) shows the secondary endosymbiosis where the eukaryotic red alga is engulfed by a different heterotroph. Green alga can potentially also be engulfed. The heterotrophic host nucleus (N2) retains the crucial algal nuclear and plastid genes, but algal mitochondrion and nucleus, and additional bacterial genes are lost. Image from Armbrust (2009).*

1.2 The frustule

The highly ornate frustule, which diatoms synthesize from amorphous silica [(SiO2)n(H_2O)], is developed in a process called biomineralization (Kroth, 2007). In this process, organisms will form mineralized structures by converting ions into solid minerals. This has a very important function in numerous organisms that have mineralized skeletons, as it can be involved in extracellular as well as intracellular events (Simkiss and Wilbur, 2012). Biomineral deposits can also act as storage systems for times where organisms might experience more physiological stress from the environment. Deposited ions may then be withdrawn and used for things like skeletal growth and repair (Simkiss and Wilbur, 2012). Diatoms have an important membrane-bound organelle within the protoplast, known as the silica deposition vesicle (SDV), where the frustule synthesis occurs. Silicic acid is transported to the SDV where it precipitates and can later be used for intracellular silica biogenesis. The membrane surrounding the SDV is the silicalemma, a lipid-based bilayer structure (Kröger and Poulsen, 2008).

The diatom frustules consist of two almost identical halves (theca) where one is slightly larger than the other, similarly to a Petri dish. The bigger half of the two is called the epitheca, while the smaller half is called the hypotheca (Figure 1.2). Each of these are further composed of circular siliceous girdle band that are attached to a larger outer surface, the valve (Falciatore and Bowler, 2002). The girdle bands are what keeps the two halves together and encloses the protoplast. Each theca has a terminal girdle band called the pleural band, which is at the region where the two theca overlap (Kröger and Poulsen, 2008). Based on the frustule structure, diatoms are divided into two groups: centric and pennate. Centric diatoms have a radial symmetry, while pennate diatoms have bilateral symmetry (Falciatore and Bowler, 2002).

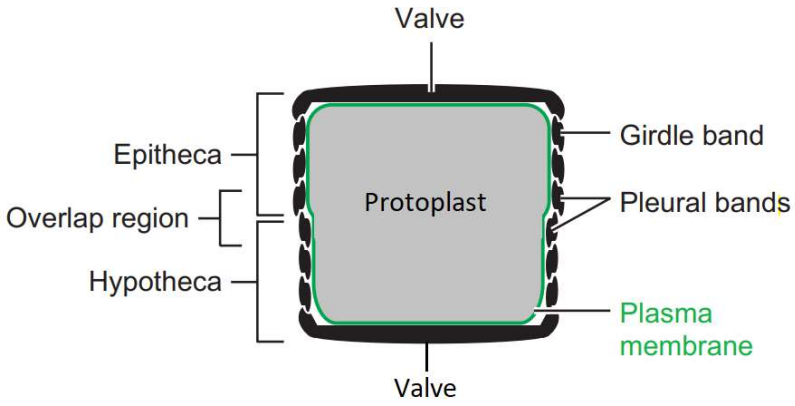

Figure 1.2. *An overview of the different structures of the diatom frustule. Adapted from Kröger and Poulsen (2008).*

Diatoms normally reproduce asexually, through mitosis (Figure 1.3). To prepare for division, the cell will slightly elongate, pushing the two thecas away from each other so the nucleus can divide. Girdle bands are synthesized stepwise to avoid gaps in the frustule as the cell increases in volume during interphase (G1, S and G2 of the cell cycle) (Falciatore and Bowler, 2002; Kröger and Poulsen, 2008). The plasma membrane will undergo invagination to divide the protoplast through cytokinesis. The thecas of the parent cell then becomes the epitheca of the daughter cell, meaning the new daughter cells must generate the hypotheca themselves. The formation of the hypotheca is very intricate and works by depositing silica in a precise lattice work before coating it with organic matrix that prevents dissolution. This entire structure is exocytosed from the SDV after generation, and the two daughter cells are then fully formed and separated (Falciatore and Bowler, 2002).

By reproducing mitotically, the average diatom cell size in a population will decrease over time (Kröger and Poulsen, 2008). Once they reach around 30-40% of their maximum diameter, they will resort to sexual reproduction by meiosis to obtain their original size. During gametogenesis, gametes from male and female individuals will combine to create a diploid specialized zygote, an auxospore. This auxospore will increase a lot in volume over a short period of time before it then continues the path of asexual division until the cell size later becomes unsustainable again (Falciatore and Bowler, 2002; Kröger and Poulsen, 2008).

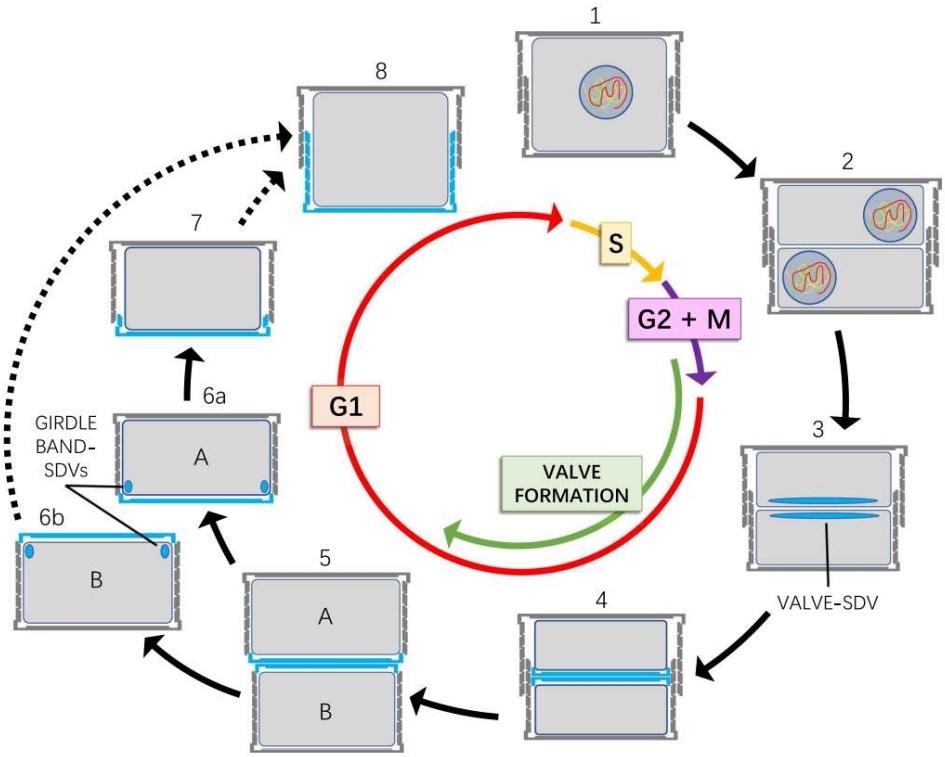

Figure 1.3. *Schematic representation of the cell cycle and frustule formation of T. pseudonana. The cell cycle phases in the center of the figure shows the relative lengths of each phase. In step (1), the cell starts a new cell cycle. (2) After DNA replication in the S phase, the cell enters the G2+M phase. Here the two daughter cells' protoplasts are formed by cytokinesis (end of M phase), although both remains inside the parental frustule. (3) An SDV in each protoplast is formed and synthesis of the building blocks for the hypotheca generation is started. In (4), the contents of the SDVs are exocytosed. (5) The daughter cells are now fully separated and can start the formation of their hypotheca. Cell A will be slightly larger than cell B since it inherited the parental epitheca. (6) Building blocks for formation of the girdle bands are synthesized in the SDVs, and in (7) they are exocytosed. (8) Synthesis of the girdle bands leads to increased girdle length and cell volume. A new round of cell cycle happens when the cells have reached the maximum number of girdle bands. Image from Fattorini and Maier (2021).*

1.3 Thalassiosira pseudonana

The centric (and cylindrical) diatom *Thalassiosira pseudonana* has become an increasingly favored model species of diatoms in research, especially for understanding frustule formation and the silica biomineralization processes behind it (Armbrust, 2004; Mock et al., 2008). *T. pseudonana* was the first diatom species to be chosen for whole genome sequencing. Its DNA sequence contains approximately 34 megabases including its mitochondrial, nuclear and plastid genomes, with a total of 11 242 protein-encoding genes overall (Armbrust et al. 2004).

The valves of *T. pseudonana* are circular and measured to be around 3.8 µm (±0.4 µm) in diameter on average. Valve formation is initiated in an area on the valve called the pattern center, and ribs are radiating out from that area and towards the rim of the valve. The surfaces of the valves on *T. pseudonana* consist of interspersed circular pores between the ribs, which are approximately 18 nm (±3.1) in diameter (Hildebrand et al., 2006). Figure 1.4 shows the shape and different structures of the frustule of *T. pseudonana.*

The strain used in this project is *T. pseudonana* CCMP1335. This strain is an estuarine isolate that was originally collected in 1958 from Moriches Bay (Long Island, New York, USA). It has been cultured continually since then (Nuester et al., 2012).

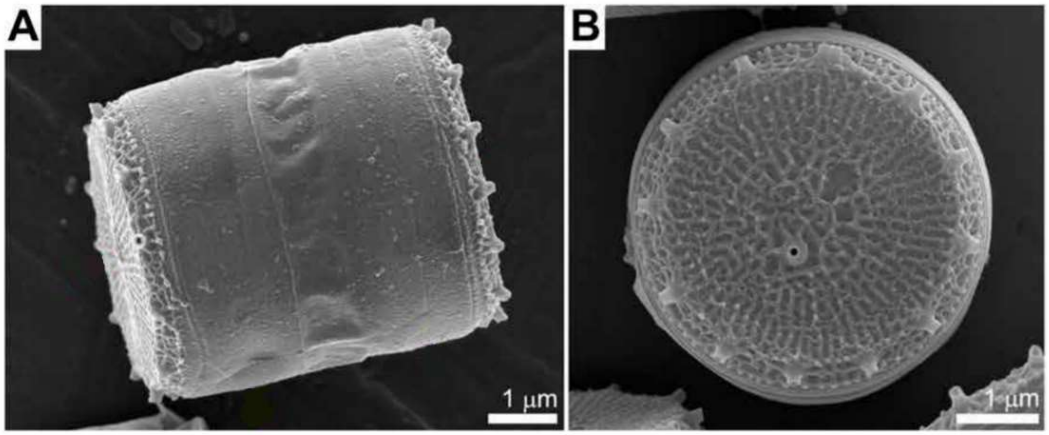

Figure 1.4. *Images of dried frustules of T. pseudonana showing the shape of the species. The pictures are taken with a scanning electron microscope (SEM). A) shows a side-view of the cell with the girdle bands, while B) gives a better view of the structure of its valve. Adapted from Kumar et al. (2020).*

1.4 Silicic acid transporters

To obtain silica, diatoms rely on silicic acid transporters (SITs) to actively pump silicic acid $Si(OH)_4$ into the cell from the environment (Figure 1.5 A) (De Tommasi et al., 2017). Transport through SITs occurs mainly when the concentrations of silicic acid in the environment is low (<30µM). However, at higher concentrations, diffusion is usually the way the cell acquires silicic acid, and the transport role of the SITs lower (Shrestha and Hildebrand, 2015). In marine species, like diatoms, SITs act like a co-transporter with sodium ions (Na^+) (Hildebrand et al., 1997). Their placement is in the plasma membrane, and there is predicted to be 10 transmembrane domains (TMDs) in each SIT, as well as a coiled-coil motif at the C-terminus (Figure 1.5 B) (De Tommasi et al., 2017). The exact role of this motif is not yet known. Centric diatoms, however, do not have the coiled-coil motif in their SITs, something that could mean that ancestral SITs lacked this. Furthermore, this could propose that either its function is unnecessary for centric diatoms, or that they accomplish their function through other mechanisms (Thamatrakoln et al., 2006). In a comparative analysis of amino acids done by Thamatrakoln et al. (2006), conserved regions in relation to silicon transport were identified. This included repeats of the sequence motif GXQ (X = Q, G, R, or M) at the transmembrane region TMD2+3, as well as GRQ at TMD7+8. These are believed to be responsible for coordinating the silicic acid during transportation.

SITs were some of the first silicon-responsive genes to be identified in any organism, and it appears that SIT genes have no known homologues in other organisms (Hildebrand et al., 1993; Kröger and Poulsen, 2008). Since SITs occur in taxonomically isolated lineages (in distantly related eukaryotic groups), it is

suggested that they have evolved independently, or by horizontal gene transfer (HGT) (Marron et al., 2016). Evidence also suggests that the TMDs has evolved independently multiple times though duplication and fusion of SIT-Ls, a family of related transporters (Marron et al., 2016). A phylogenetic analysis by Marron et al. (2016) proposed that SITs may originally have risen to counter the high concentrations of silicon (Si) in the Precambrian (4600-542 m.y.a.) oceans and to prevent Si toxicity in the cell. In the Phanerozoic era (542 m.y.a. to present day), the Si concentration in the oceans has reduced, leading to a widespread loss of SITs in different lineages. They concluded that the Si levels in the ancient oceans had a huge effect on the formation and development of different silicon transporter families. An overaccumulation of intracellular soluble silicon could be detrimental to the cell, as it could auto-polymerize. The process of silicon efflux and being able to control the intake of silicon has therefore been a very important trait in marine silicious organisms (Thamatrakoln et al., 2006). Shrestha and Hildebrand (2015), however, raise a question about SIT evolution, arguing that reduced levels of Si in the ocean should have increased the need for transporter-mediated uptake instead of leading to the loss of SITs. They suggest that SITs might not have been involved in the silicification process development unless transport across the intracellular membranes was required.

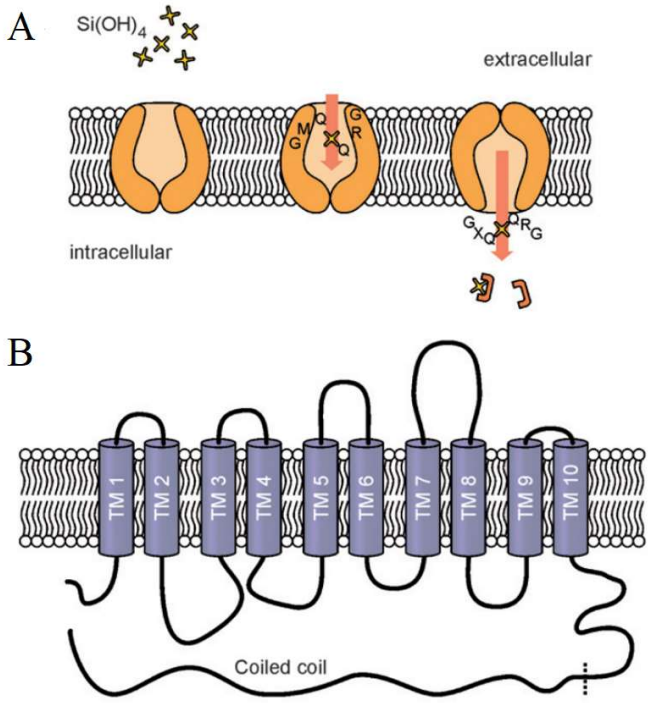

Figure 1.5. *A) Model of how the SIT-mediated silicon transport in the plasma membrane occurs. B) Overview of the predicted 10 transmembrane domains with the coiled-coil motif that is found in SITs. Adapted from Schröder et al. (2008).*

It is thought that the protein expression of SITs is closely related to the cell cycle. During cell division there will be a down regulation as the formation of the cell wall happens (De Tommasi et al., 2017). In the absence of silicon (Si), diatoms will eventually die because the cell division will halt (Kröger and Poulsen, 2008). In a study by Shrestha and Hildebrand (2015), knockdown mutants of two important SITs in *T. pseudonana*, SIT1 and SIT2, were analyzed and it was shown that the down regulation is inversely related to the silicic acid concentration. The results suggested that these transporters also have a role in sensing the levels of silicic acid and can therefore evaluate and control the cell division process and frustule synthesis independently of the transport.

1.5 Kinases

Many intracellular reactions and pathways are dependent on an array of different proteins and molecules in order to work properly. Kinases are enzymes that plays a crucial role in these cellular signaling pathways and comprise one of the largest superfamilies of homologous genes in eukaryotes. They work by catalyzing the transfer of phosphate groups from ATP (adenosine triphosphate) to specific substrates. The process of phosphorylation is highly conserved, and dependent on certain structural features of the active site found in all kinases (Jura et al., 2011). The kinase domain (catalytic domain) contains around 250-300 amino acid residues and is the structure that determines which kinases are related. The kinase superfamily is mainly divided into two subsections: protein-serine/threonine kinases and protein-tyrosine kinases (Hanks and Hunter, 1995).

The catalytic domain, in a complex with a divalent Mg^{2+} cation, is responsible for the binding of the ATP (or GTP) phosphate donor to a specific protein or peptide substrate. The donor phosphate will bind to the acceptor hydroxyl residue (Ser, Thr, or Tyr) of the substrate and activate it (Hanks and Hunter, 1995). Inactivation of the target protein can be done by dephosphorylation where a protein phosphatase enzyme removes a phosphate group from the amino acid residue on the substrate. The process of substrate phosphorylation by a kinase is showed in Figure 1.6.

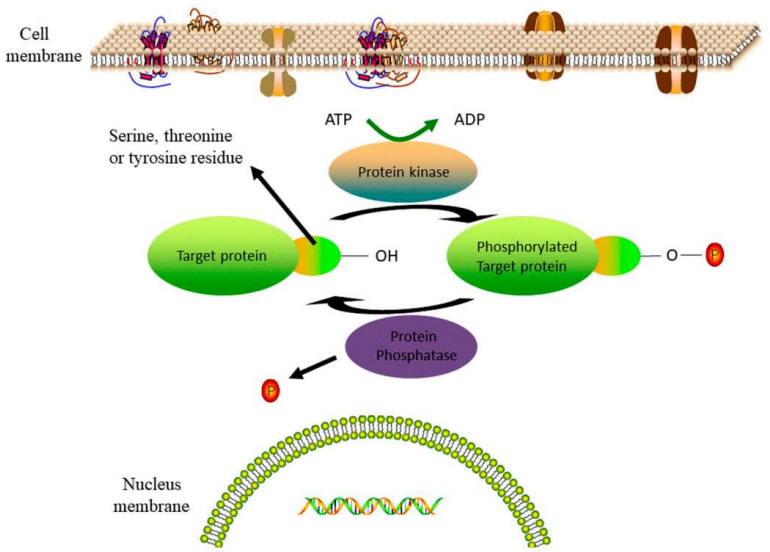

Figure 1.6. *Overview of the protein phosphorylation by a catalytic kinase and the protein dephosphorylation by a phosphatase. Image from Li et al. (2019).*

The Ca^{2+}/calmodulin-dependent protein kinase (CaMK) molecule is a cellular mediator of calcium (Ca^{2+}) signals (Bayer and Schulman, 2019). The calcium ion is a common second messenger in the cell that relays and amplifies signals received at the receptors on the cell surface. There are multiple versions of the CaMK proteins, and it is found in different forms, e.g., the monomeric CaMKI kinases and the multimeric holoenzyme CaMKII that forms a wheel-like structure from 6-12 kinase subunits. CaMKII is a versatile serine/threonine protein kinase that can, in response to increasing Ca^{2+} concentrations in the cell, phosphorylate multiple different proteins (Maier and Bers, 2007).

Each of the monomeric kinase subunits constructing the shape of the CaMKII contains a catalytic domain with partially overlapping CaM-binding and autoinhibitory regions at the N-terminal, and a regulatory domain at the C-terminal responsible for oligomerization. The autoinhibitory region will sterically block substrate access to the active site by being so close to the catalytic region. When the concentration of Ca^{2+} increases in the cell, calmodulin (CaM) will bind four Ca^{2+} ions to create the Ca^{2+}/CaM complex. This complex will bind to the CaMKII regulatory domain and thereby activating the enzyme by displacing the autoinhibitory domain. CaMKII can stay in this activated state, even after the Ca^{2+}/CaM complex has dissociated and the Ca^{2+} concentration has declined, by the autophosphorylation of the threonine on the autoinhibitory segment. Phosphatases can then dephosphorylate the autophosphorylated CaMKII to inactivate the enzyme (Maier and Bers, 2007).

1.6 Characterization of kinases co-expressed with TpSITs

In a study by Brembu et al. (2017), cultures of *T. pseudonana* were subjected to Si starvation and Si replenishment, and a following transcriptomics analysis discovered the two closely related kinase-encoding genes *Tp14322* and *Tp264671*. They have been shown to be closely co-expressed with silicon transporters SIT1 and SIT2, respectively. This co-expression was visible in the Si experiments where *Tp14322* showed similar up-regulation of gene expression to SIT1 during Si shift-down, and a similar down-regulation during Si shift-up. *Tp264671* showed the same trend with SIT2 in these Si shift-up/down experiments. As no further studies have been done, we do not know much about these genes and their function. However, they seem to have important roles in the cell as they are expressed in very high levels, and it is hypothesized that the two genes might be involved in silicon sensing. *Tp14322* and *Tp264671* are likely cytosolic as no putative ER signal peptides were identified for them. This may suggest that SITs could be a potential candidate for the substrates of the kinases, although no phosphorylation of them has been reported yet. Through a phylogenetic analysis, it was revealed that the kinase family *Tp14322* is a part of, is found exclusively in the Thalassiosirales order (and not in other diatoms), and that it is related to calcium-dependent protein kinases found in *T. pseudonana*. Furthermore, in Figure 1.7, it is shown that *Tp14242* is another kinase-encoding gene that is closely related to *Tp14322* and *Tp264671*. All of them are also very similar on a nucleotide level, which is why we can include *Tp14242* when trying to establish triple knockout lines of *T. pseudonana.*

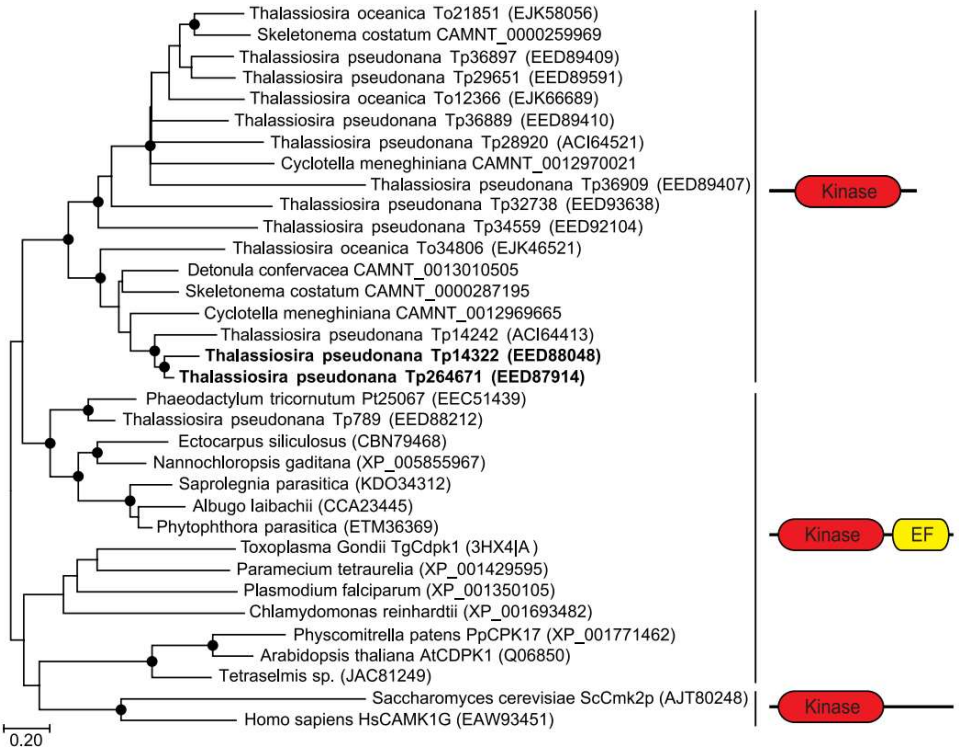

Figure 1.7. *Phylogenetic tree based on the protein alignment of Tp14322 and Tp264671 (shown in bold) and some related kinases. Tp14242 is within the same clade as Tp14322 and Tp264671, indicating that they are very closely related. To the right, domain composition of the different clades is shown. Image from Brembu et al. (2017).*

1.7 CRISPR/Cas9

The technology of clustered regularly interspaced palindromic repeats (CRISPR) is a relatively new genome-editing approach that can generate RNA-guided nucleases with customizable specificities, such as the CRISPR associated Cas protein (Sander and Joung, 2014). The CRISPR/Cas system was first discovered in prokaryotes where it works as an RNA-guided defense mechanism against plasmid and viral invasions. The prokaryote will integrate fragments of the viral DNA into its genome to serve as an adaptive immune system, and later infections will then lead to the cleavage of the foreign double-stranded DNA (dsDNA). More recently, a very powerful and efficient tool in gene-editing technology has been developed based on the bacterial CRISPR-associated Cas9 endonuclease found in *Streptococcus pyogenes* (Sander and Joung, 2014).

CRISPR/Cas9 is a complex that utilizes a single guide RNA (sgRNA) to guide the Cas9. The sgRNA is designed to have two important features: the 20-nucleotide sequence at the 5' end that directs the complex to a complementary DNA target sequence, and a double-stranded structure at the 3' end that binds to Cas9 (Doudna and Charpentier, 2014). The DNA target sites for the sgRNA must lie immediately 5' of a protospacer adjacent motif (PAM) sequence that matches the 5'-NGG trinucleotide sequence. A PAM is a short and conserved sequence motif consisting of 2-5 bp. Alternate PAM sequences (like 5'-NAG) have been recorded, although not as effective as 5'-NGG (Sander and Joung, 2014; Jiang and Doudna, 2017). When the sgRNA/Cas9 complex binds to the target site, the DNA is cleaved by the two nuclease domains of Cas9, HNH and RuvC, and a double-stranded break (DSB) is created. The HNH domain cleaves the DNA strand that is complementary to the sgRNA, while the RuvC domain cleaves the DNA strand that is opposite of the complementary strand. Any PAM site in the DNA can be targeted by Cas9 by altering the first 20 nucleotides of the sgRNA (Doudna and Charpentier, 2014). A schematic overview of the CRISPR/Cas9 system is shown in Figure 1.8 A.

The process of genome editing by CRISPR/Cas9 is dependent on the creation and repair of DSBs in the target dsDNA (Figure 1.8 B). To repair the nuclease-induced DSBs, either non-homologous end joining (NHEJ) or homology-directed repair (HDR) can be used. Both processes are intracellular pathways responsible for DNA repair in nearly every type of cells and organisms. Repair by NHEJ can lead to indel mutations of different lengths by insertion, deletion, or substitution, something that can disrupt the

translational reading frame of a coding sequence. HDR can, through homologous recombination of DNA "donor" templates, introduce specific point mutations or insert specific sequences into the target DNA (Sander and Joung, 2014).

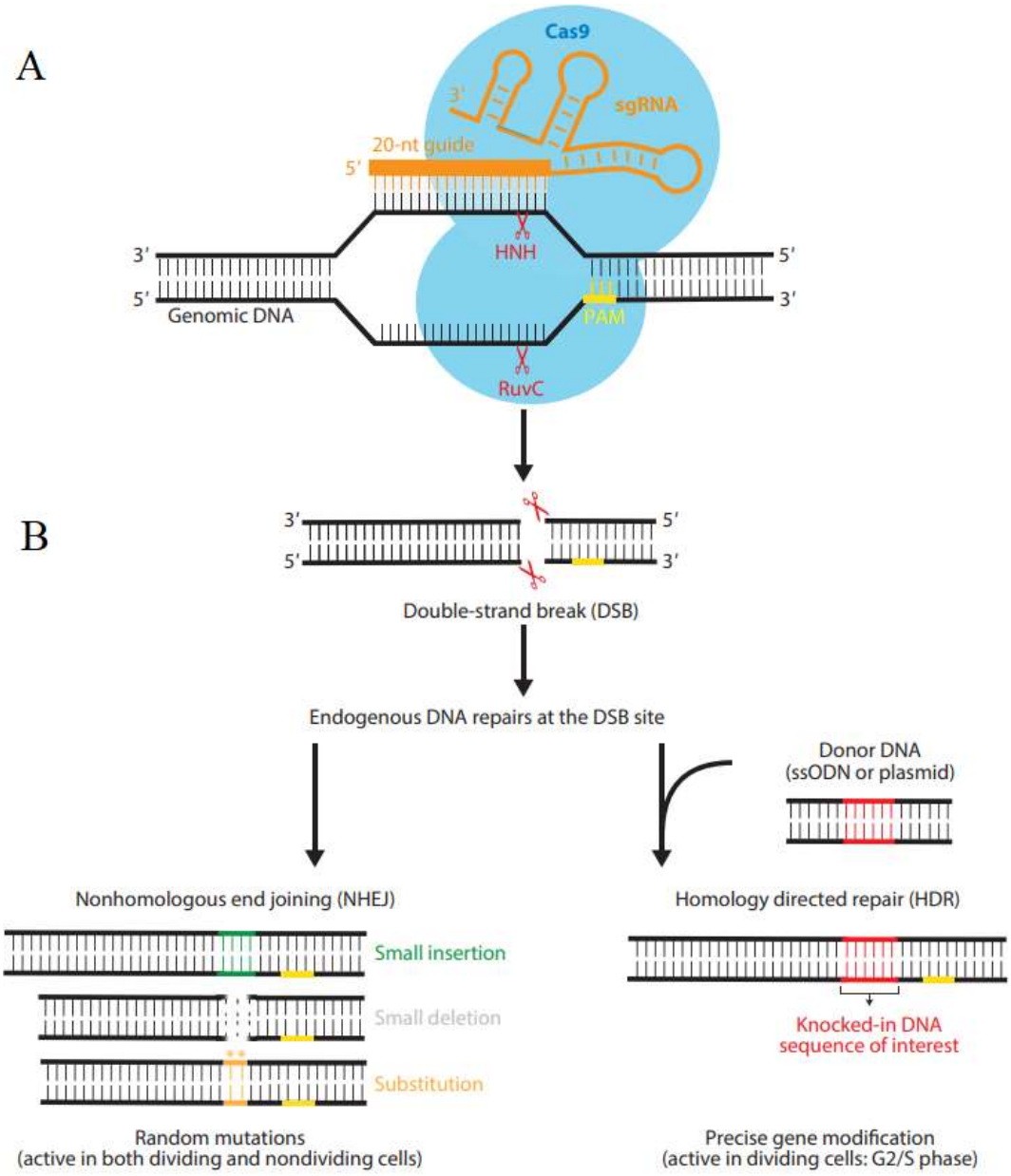

Figure 1.8. *Representation of how the CRISPR/Cas9 complex binds to target DNA to produce DSB, and how the DSB can be repaired by different methods. A) The sgRNA and Cas9 forms a complex that is directed to a complementary DNA sequence to the 5' of the sgRNA. The DNA site is then cleaved by the HNH and RuvC nuclease domains of Cas9. B) DSB reparation at the DNA target site can be done by NHEJ (left) which generates random indel mutations, or by HDR (right) which can be used to insert a DNA sequence into the target site by homologous recombination. Image from Jiang and Doudna (2017).*

1.8 Thesis objective

The main objective of this thesis is to create mutant strains of *T. pseudonana.* Single, double, and possibly triple (with *Tp14242*) knockout lines will be generated by targeting the two kinase-encoding genes *Tp14322* and *Tp264671* with the CRISPR/Cas9-based genome editing technology. HRM and Sanger sequencing will be used to screen and characterize edited loci and identify mutants. Confirmed mutant lines with an out-of-frame mutation will be subjected to Si limitation and Si replenishment to observe if there are any changes in phenotype. Cell growth and size will then be analyzed by flow cytometry. Frustules from the wild type and mutants will be isolated, cleansed and imaged using a scanning electron microscope (SEM). In addition, a phosphoproteomics analysis will be done on the wild type and mutants to identify substrates for the kinases.

Another goal of the thesis is to fuse *Tp14322* and *Tp264671* with the fluorescent protein mNeonGreen and express them in *T. pseudonana* by conjugation. Transgenic lines that express mNeonGreen will be analyzed to confirm that the transgenic expression mirrors the expression of the kinases in the wild type. Fluorescence microscopy and laser scanning microscopy will be used to analyze the expression and localization of the kinases under Si deplete and Si replete conditions.

2 Materials and methods

Information about solutions and primers can be found in the appendices.

2.1 Kinase domain analysis

Protein alignments of the two kinase-encoding genes *Tp14322* and *Tp264671* against the wild type sequence were done in NCBI to investigate the catalytic domain and identify which amino acids were conserved in what domains of the kinases. The alignments in NCBI were compared to the consensus sequence for a number of kinases (including *Tp264671*, *Tp14322* and *Tp14242*) in a SnapGene file provided by supervisor Tore Brembu. This was done to get an extra confirmation that the same conserved amino acids were found in the same places in the alignments on both platforms.

The amino acid sequence of the catalytic domain of the kinases, which appeared to be CaMK, was compared to the CaMK2α amino acid sequence found in Hanks and Hunter (1995) to look for deviations in conserved amino acids both between the kinase alignments and the CaMK2α sequence, as well as in the sequences between the three kinases. This could give information about protein efficiency in addition to substrate specificity. Additionally, another small investigation was done to see where the different PAMs were located in the amino acid sequences of *Tp264671* and *Tp14322*.

2.2 Creating knockout mutants of *Tp14322* and *Tp264671* (and *Tp14242*)

2.2.1 Selection of PAM sites and primers

The PAMs used in this project (Fig. 2.1) were chosen by supervisor Tore Brembu, and primers were designed to amplify regions containing these CRISPR/Cas9 5'-NGG target sites. The nucleotide sequences of the sgRNA and the target DNA determines the target efficiency of the CRISPR/Cas9 system, and it is revealed that off-target effects can occur when there are more than three mismatches between the sgRNA and the target sequence (Naeem et al., 2020). PAMs were therefore chosen based on a low number of mismatches (not more than 2) and a low probability of off-target effects by Cas9. A script made by Associate professor Per Winge at NTNU was used to search for mismatches in the PAM target region. Primers (usually 20-23 bp long) that were used to amplify regions in the three closely related kinase-encoding genes were run through a primer BLAST (Basic Local Alignment Search Tool), another script made by Per Winge, to make sure they were specific to the target region.

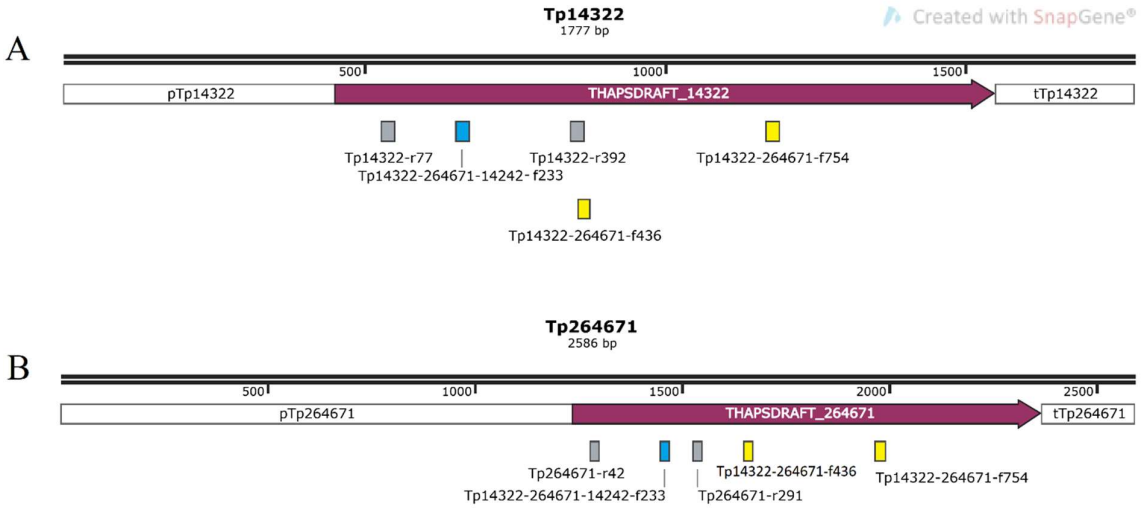

Figure 2.1. *Schematic maps of the two genes Tp14322* ***(A)*** *and Tp264671* ***(B)****, and the location of the PAMs. The name/number of the PAMs specifies the location in the coding region (shown in purple), and the letter in front of them indicates whether the PAM sequence is on the forward (f) or reverse (r) DNA strand. The different colors of the PAM regions are based on whether they were used for generating single (grey), double (yellow), or triple (blue) knockouts.*

2.2.2 Modifying vector to create knockout cell lines

The vector pTpPUC3-Cas9-M-G1 (Fig. 2.2) was used when trying to establish knockout mutants of *T. pseudonana.* To construct this vector for gene editing, a Cas9 gene and an sgRNA target region was inserted into the empty pTpPUC3 backbone (Addgene ID: Plasmid #62864), and the BsaI restriction site in the FCPB promoter upstream of the Nourseothricin (Nou) resistance gene was also mutated. As seen in Fig. 2.2, the expression of the Cas9 and the Nou resistance gene is controlled by the FCPB promoter, and they also both have a LHCF9 terminator. The Cas9 gene, however, is also tagged with the yellow fluorescent protein YFP at the C terminus. sgRNA expression is controlled by a U6 promoter and has a U6 terminator. Two BsaI restriction sites placed immediately 5'- to the sgRNA were used for restriction digestion and ligating in oligos for the two kinase-encoding genes *Tp14322* and *Tp264671* to create customized target sites for gene editing. In addition to containing a gene for resistance against Nou, a kanamycin resistance gene is also found in the pTpPUC3-Cas9-M-G1 vector. Antibiotics for the corresponding antibiotic resistance genes were added to the growth medium for selection pressure when culturing cells, Nou when culturing *T. pseudonana* and kanamycin when culturing *E. coli*.

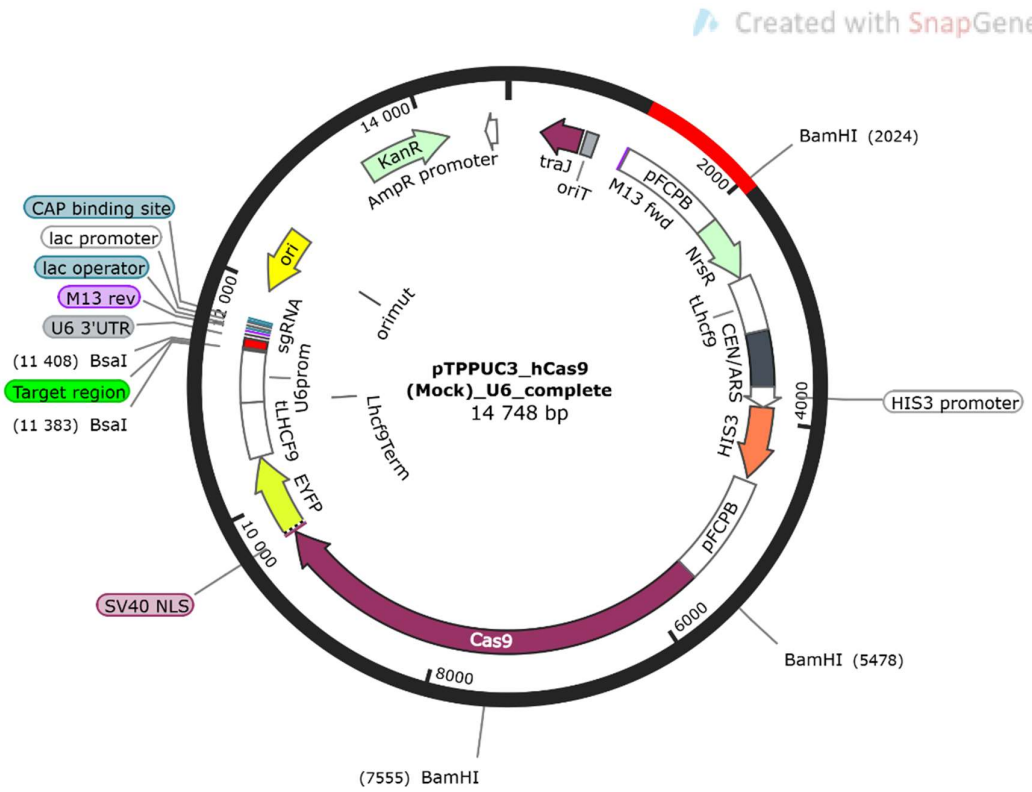

Figure 2.2. *Map of the pTpPUC3-Cas9-M-G1 vector that shows the organization of the different genes involved in the process of gene editing in this project.*

Oligo annealing reactions with target sequences for CRISPR/Cas9 were performed using 1.4 µl (100 µM stock concentration) of each oligo, forward and reverse, 5 µl T4 DNA ligase buffer (NEB), and 42.2 µl autoclaved MQ water to reach a total volume of 50 µl. The reaction was incubated on a T100 Thermal Cycler (Bio-Rad) for 10 minutes at 85 °C, and then cooled down 1°C every 90 sec until it reached 25°C. The complementary sequences for the oligos can be found in Table A2-1 (Appendix 2).

The plasmid was linearized by cutting 8 µl pTpPUC3-Cas9-M-G1 plasmid DNA (114.7 ng/µl) with 1 µl restriction enzyme BsaI-HFv2 (20,000 units/ml, NEB), and adding 5 µl 10x CutSmart Buffer (NEB) and 36 µl MQ water to get a total reaction volume of 50 µl. This reaction was incubated for 30 minutes at 37°C. The annealed oligos were ligated into the plasmid by adding 1 µl T4 ligase, 2 µl T4 ligase buffer, and 7 µl MQ water to a 20 µl reaction containing 5 µl cut plasmid and 5 µl annealed oligos. This reaction was incubated for 60 min at 25°C before being heat inactivated for 10 minutes at 65°C.

2.2.3 Transformation of DH5α and DH10β *Escherichia coli* (*E. coli*) with pTpPUC3

Competent DH5α *E. coli* cells (50 µl) were thawed on ice before adding 5 µl of the ligation product. This transformation mix was incubated on ice for 30 minutes before going through a heat-shock for 30 seconds at 42°C before being put back on ice. The heat-shock induces the formation of pores in the cell surface, allowing the double-stranded plasmid DNA to pass through and into the cell (Hasegawa et al., 2018). Further, 1 ml of room temperature liquid Luria-Bertani (LB; Table A1-1, Appendix 1) medium (Bertani, 1951) was added to the cells before incubating them on a Thermo-Shaker (TS-100C, BioSan) for 1 hour at 37°C and 360 rpm. After incubation, the cells were spun down at 3000 rpm for 5 minutes and most of the supernatant was decanted. The pellet was resuspended in the leftover medium and the transformation mix (~100 µl) was plated out on LB agar plates containing kanamycin (50 µg/ml). The plates were incubated overnight at 37°C. Clones that appeared on the plates after overnight incubation were picked for PCR colony screening to see if the *E. coli* colonies contained the plasmid with the inserted oligo. The colonies were restreaked onto new LB agar plates (with 50 µg/ml kanamycin) before being transferred into PCR tubes with a polymerase reaction (Table 2.1) containing the insert-specific forward primer for the PAMs (Table A2-1, Appendix 2) and the backbone-specific reverse primer M13 which is downstream of the sgRNA (Table A2-7, Appendix 2). A 200 bp region was amplified using the PCR program presented in Table 2.2.

Table 2.1. *Red Taq PCR reaction.*

Component	Volume (µl)	Final concentration
VWR Red Taq 2x DNA Polymerase Master Mix	12.5	1x
Forward primer (10 µM)	0.5	0.2 µM
Reverse primer (10 µM)	0.5	0.2 µM
MQ water	11.5 (to 25 µl)	
Template DNA	0 (picked)	<1000 ng
Total volume	25	

Table 2.2. *PCR program for colony screening with Red Taq.*

Program	Temperature (°C)	Time	Cycles
Initial denaturation	95	5 min	1
Denaturation	95	30 sec	
Annealing	60	30 sec	34
Extention (1 min/kb)	72	15 sec	
Final extention	72	5 min	1

The PCR products were loaded on a 1% agarose gel in 1x TAE (Tris-acetate-EDTA) buffer containing GelRed® Nucleic Acid Gel Stain (10,000x in water, Biotium), a staining solution that will fluoresce under UV-light when intercalated to DNA (Crisafuli et al., 2015). Colonies whose PCR product gave a ~200 bp band were inoculated in 5 ml liquid LB medium containing kanamycin (50 µg/ml) and incubated overnight at 37°C and 170 rpm. Plasmids from the liquid bacteria cultures were then isolated and purified by using the GeneJET Plasmid Miniprep Kit (Thermo Scientific) and following the manufacturer's protocol, with the exception that 60°C MQ water was used to elute the DNA instead of elution buffer. The concentrations of the plasmids were measured on a NanoDrop™ One Spectrophotometer (Thermo Scientific) before submitting 7.5 µl of each sample with 2.5 µl M13r primer for Sanger sequencing (Eurofins Genomics) to verify the presence of the inserts.

Plasmids with oligos that showed a clean and matching sequence to the WT after sequencing were transformed into 50 µl competent DH10β *E. coli* cells prior to conjugation. This strain of *E. coli* contains a sex pili, a hair-like structure on the bacterial cell surface that is used to transfer plasmid from a donor cell to a recipient (Virolle et al., 2020). The transformation procedure was the same as when transforming DH5α cells, with the exception that the transformation mix was plated out on LB agar plates containing kanamycin (50 µg/ml) and gentamicin (20 µg/ml). After overnight incubation at 37°C, clones were picked and inoculated in 5 ml liquid LB medium containing kanamycin (50 µg/ml) and gentamicin (20 µg/ml) to incubate at 37°C overnight.

2.2.4 Transfer of plasmid DNA from *E. coli* into *T. pseudonana* by conjugation

Bacterial conjugation is the process of gene transfer between two cells which allows the recipient cell to maintain the vector as an episome. The transfer of episomes by conjugation is most common among bacteria but can also occur between bacteria and eukaryotic cells (Llosa et al., 2002; Sharma et al., 2018; Karas et al., 2015). The procedure of DNA transfer to *T. pseudonana* via conjugation from *E. coli* in this project was done accordingly to the protocol described in Karas et al. (2015).

Erlermeyer flasks (250 ml) were filled with 150 ml LB containing kanamycin (50 µg/ml) and gentamicin (20 µg/ml), and 3 ml of the overnight DH10β cultures were added to the flasks before being incubated at 37°C and 170 rpm. The cultures were grown and measured continually on a U-5100 UV-Visible Spectrophotometer (Hitachi) until they reached an absorbance value > 3.0 at OD_{600}. They were then spun down for 10 min at 25°C and 3000 g before the supernatant was decanted and the pellet was resuspended in 800 µl SOC (super optimal broth) medium (Table A1-2, Appendix 1).

Cells from a culture of *T. pseudonana* (CCMP1335) were grown in liquid f/2 medium (Table A1-3, Appendix 1; Guillard, 1975) at 18°C and under constant white light with a photosynthetically active radiation (PAR) of 150 µmol m^{-2} s^{-1}. The cells were spun down for 5 min at 10°C and 4000 g, and the supernatant was decanted with a pipette as the algae pellet is more unstable and loose as opposed to a bacteria pellet. The pellet was resuspended in leftover medium before 400 µl of the mix was pipetted out to be cell counted on a BD Accuri™ C6 Flow Cytometer (BD Biosciences). The concentration of the cells was supposed to be approximately 2 x 10^8 cells ml^{-1}, however, the culture did not reach the desired density. The concentration of the algae cell culture was measured to be 1.28 x 10^8 cells ml^{-1}. All the cells were evenly distributed to 6 Eppendorf tubes, each containing 550 µl cell culture.

For conjugation, 200 µl *E. coli* cells from the SOC medium were added to the Eppendorf tube containing 550 µl *T. pseudonana* and mixed by pipetting up and down a few times. This mix (750 µl) was plated out on ½ f/2, 5% LB, 1% agar plates and incubated in the dark for 90 min at 30°C. After this incubation period, the plates were moved to 18°C conditions with light that had a PAR of 150 µmol m^{-2} s^{-1} and incubated for 4 hours. Further, 1 ml f/2 medium were added to the plates and the cells were scraped. The scraped cells were mixed a little by pipetting before 200 µl were plated out on ½ f/2, Nou (50 µg/ml), 1% agar plates and incubated at 18°C with constant light (150 µmol m^{-2} s^{-1}).

Colonies started appearing on the plates at around 2 weeks, and approximately 3-4 weeks after the conjugation, 24 clones from each plate (each transformation) were transferred to sterile 24-well Tissue Culture Plates (VWR) containing 1 ml f/2 with 1 µl Nou (50 µg/ml). These liquid cell culture plates were incubated at 18°C with constant light (150 µmol m^{-2} s^{-1}). The cultures were regularly (every 9-10 days) diluted with f/2 medium to ensure that they stayed alive.

DNA from the *T. pseudonana* liquid cell cultures with the different PAMs were extracted by transferring 100 µl of each clone to an Eppendorf tube and centrifuging at 17 000 g for 1 min. The supernatant was decanted with a pipette and the pellet was resuspended in 50 µl lysis buffer (10% TritonX-100, 20 mM Tris-HCl pH 8, 10 mM EDTA; Table A1-5, Appendix 1). This mix was incubated on ice for 15 min, followed by an incubation at 95°C for 10 min. Lysates were stored at -20°C and could later be used as templates for PCR.

A second conjugation was done using the same protocol as described previously in this section, except this time 800 µl of the scraped cells were also plated out on a ½ f/2, Nou (50 µg/ml), 1% agar plate in parallel with the 200 µl that was done the first time. The plates were incubated at 18°C with constant light (150 µmol m^{-2} s^{-1}) for approximately three weeks. The triple knockout (KO) was the one that was focused on this time, and clones that appeared on the triple KO conjugation plates were inoculated in a 24-well Tissue Culture Plate (VWR) containing 1 ml f/2 media with Nou (50 µg/ml) and incubated at

20°C with constant light (150-200 µmol m^{-2} s^{-1}). Half of the clones on the liquid culture plate were from the 200 µl plate, and the other half was from the 800 µl plate. A continuation of the work that was done on the clones from the second conjugation can be found in section 2.2.8.

2.2.5 Amplification of 800-1000 bp amplicons before mutant screening

The DNA sequences surrounding the target sites of the different PAM clones were amplified in a PCR reaction by using their lysates as DNA template. Phusion® High-Fidelity DNA Polymerase (Thermo Scientific) reactions with insert-specific primers (Table A2-2, Appendix 2) were made as described in Table 2.3. The amplicon lengths expected in a WT control sample would be 998 bp and 868 bp for *Tp14322* and *Tp264671*, respectively. However, as the samples could contain large indels (insertion/deletion of nucleotides) at the editing sites, a longer extension time (1 min) was necessary when running the PCR program (Table 2.4). The size of the PCR products and potential indels were investigated by running a 1% agarose gel in 1x TAE buffer (Table A1-6, Appendix 1) with GelRed (Biotium) to stain the DNA.

Table 2.3. *Phusion PCR reaction setup (20 µl).*

Component	Volume (µl)	Concentration
5x Phusion HF Buffer	4	1x
10 mM dNTPs	0.4	200 µM each
Forward primer (10 µM)	1.25	0.625 µM
Reverse primer (10 µM)	1.25	0.625 µM
Template DNA (lysates)	1	
DMSO	0.6	3%
Phusion Hot Start II DNA polymerase (2 U/µl)	0.2	0.02 U/µl
MQ water	11.3 (to 20 µl)	
Total volume	20	

Table 2.4. *PCR thermoprofile for amplifying target sites with Phusion DNA polymerase.*

Program	Temperature (°C)	Time	Cycles
Initial denaturation	98	30 sec	1
Denaturation	98	10 sec	
Annealing	60	30 sec	34
Extension (15-30 s/kb)	72	1 min	
Final extention	72	10 min	1

2.2.6 Screening for mutants with high resolution melting (HRM) analysis

Clones that showed signs of gene editing on the agarose gel were further screened by HRM to investigate putative mutations. HRM analysis is a method that detects sequence variations such as indels and single nucleotide polymorphisms in a PCR product (Twist et al., 2013). An amplification step of the target sequence is performed prior to the HRM analysis where a fluorescent saturating dye is bound to the dsDNA. The fragments that are amplified are usually 150-250 bp long as this technology is more sensitive to smaller fragments (Vossen et al., 2009). Following the amplification step, melting curves are produced from the fluorescence of PCR products that are heated up until they reach their melting temperature (T_m). Depending on chemical properties such as sequence composition, length, and GC content, different PCR products will have different T_ms (Herrmann et al., 2006).

A dilution series of the wild type and the successfully amplified PCR products from the clones was done in preparation for the HRM, and 5 µl of the 1:4 000 000 diluted PCR products were then used as templates when amplifying short regions (~100 bp) surrounding the target sites in the HRM program. The LightCycler® 480 High Resolution Melting Master kit (Roche) was used to prepare the master mix (Table 2.5), and specific HRM primers for single-, double-, and triple-knockout mutant screening were added (Table A2-3, Appendix 2). The HRM program to amplify and analyze the samples was performed on a LightCycler® 96 instrument (Roche) according to the manufacturer's protocol (Table 2.6). Data from the run was analyzed in the LightCycler® 96 Software Version 1.1.

Table 2.5. *HRM master mix reaction setup.*

Component	Volume (µl)	Concentration
HRM master mix (2x conc.)	10	1x
$MgCl_2$ (25 mM)	2.4	3 mM
Forward primer (4 µM)	1	0.2 µM
Reverse primer (4 µM)	1	0.2 µM
PCR-grade water	0.6	
Total volume	20	

Table 2.6. *Thermoprofile for DNA amplification and HRM analysis.*

Program	Temperature (°C)	Ramp (°C/s)	Time (s)	Cycles
Preincubation	95	4.4	600	1
3-step amplification	95	4.4	10	
	63	2.2	TD 63°C, 1 Cyc → 55°C (-1°C)	45
	72	4.4	20	
High resolution melting	95	4.4	60	
	40	2.2	60	1
	65	2.2	1	
	95	2.2	1	
Cooling	37	1.0	30	1

The clones that deviated from the wild type and were detected to have mutations on the HRM were diluted 1:100 000, and 200 µl of the final dilutions were plated out on f/2, 1% agar plates that were incubated at 18°C with constant light (150 µmol m^{-2} s^{-1}). The plates did not contain Nou as the intention was to remove the plasmid from clones that appeared positive for mutations. By losing the plasmid, the clones would keep their mutations as Cas9 could not further cut their genomes (Zhang et al., 2017).

PCR products from the clones that were plated out were purified with the ExS-Pure™ Enzymatic PCR clean-up kit (NimaGen). This kit will, according to the manufacturer, degrade excess nucleotides and primers in the sample. The clean-up mixtures, containing 2.5 µl PCR product and 1 µl ExS-Pure, were incubated on a T100 Thermal Cycler (Bio-Rad) for 4 min on 37°C for enzymatic purification, followed by heat inactivation at 90°C for 1 min. After this program, 4 µl MQ water and 2.5 µl of the gene-specific primer were added (Table A2-2, Appendix 2), and the purified products were submitted for Sanger sequencing to confirm mutations. ICE (Interference of CRISPR Edits) analysis (Synthego) was used to run a batch analysis of the Sanger sequencing data that came back. This analysis tool detects CRISPR edits by comparing the sample sequence to the wild type sequence (Hsiau et al., 2018).

Colonies that grew on the f/2 plates were not further analyzed or screened as results from Sanger sequencing did not show any mutations in the clones.

2.2.7 Plasmid check

Due to strange and inconclusive results from the HRM and sequencing results, several samples containing plasmids from different parts of the mutant generation process were checked to see if the plasmid had the correct length. This included testing the pTpPUC3-Cas9-M-G1 vector (49.9 ng/µl) used for cloning, plasmids ligated with the different PAMs, as well as plasmids from conjugation cultures that were purified using the GeneJET Plasmid Miniprep Kit (Thermo Scientific). Variable volumes of plasmids (to a total of ~100 ng) were digested using 0.1 µl (1 unit) of the restriction enzyme BamHI-HF (20,000 units/ml, NEB), 1 µl 10x CutSmart Buffer (NEB), and MQ water to a total volume of 10 µl. The samples were incubated at 37°C for 30 min and loaded onto a 1% agarose gel in 1x TAE buffer that was run on 70 V for 2 h. A lower voltage combined with a longer runtime result in better separation of large DNA fragments, making it easier to determine the correct length (Issaq et al., 1997). The gel was post-stained in 50 ml MQ water containing 15 µl GelRed for 20-30 minutes before being viewed and imaged in the Syngene™ G:BOX Gel Documentation and Analysis system.

2.2.8 Screening for mutations in colonies from conjugation plate

In a study by Hopes et al. (2016) where gene editing of the urease gene in *T. pseudonana* was done, clones were screened straight off the conjugation plate instead of being cultured in liquid media first. Using the same protocol as described in the paper by Hopes et al. (2016), screening of plates was also attempted with some of the clones from the second conjugation that contained the triple KO PAM. Colonies from both the 200 µl and 800 µl conjugation plate were picked and resuspended in Eppendorf tubes containing 20 µl f/2 medium. Half of each mix was plated out on ½ f/2, Nou (50 µg/ml), 1% agar plates which were incubated at 20°C with light (150-200 µmol m^{-2} s^{-1}). The remaining 10 µl were spun down for 1 min at 17 000 g and the supernatant was decanted with a pipette before the pellet was resuspended in 20 µl lysis buffer. The mix was incubated on ice for 15 min, followed by an incubation at 95°C for 10 min. The Q5® High-Fidelity 2X Master Mix (NEB) was used to set up PCR reactions according to Table 2.7, containing 1 µl lysate as template and gene-specific primers (Table A2-2, Appendix 2). Target sites were amplified using the PCR program presented in Table 2.8. A 1% agarose gel was run for 1 h at 100 V, and the PCR product from the clones that gave a target band was purified with the ExS-Pure™ Enzymatic PCR clean-up kit (NimaGen) before being submitted for Sanger sequencing with the forward primer that was used in the amplification.

Table 2.7. *Q5 Master Mix reaction setup.*

Component	Volume (µl)	Concentration
Q5 High-Fidelity 2X Master Mix	12.5	1x
Forward primer (10 µM)	1.25	0.5 µM
Reverse primer (10 µM)	1.25	0.5 µM
Template DNA (lysates)	1	
PCR-grade water	9	
Total volume	25	

Table 2.8. *PCR thermoprofile for amplifying target sites with Phusion DNA polymerase.*

Program	Temperature (°C)	Time	Cycles
Initial denaturation	98	30 sec	1
Denaturation	98	10 sec	
Annealing	72	30 sec	40
Extension (15-30 s/kb)	72	30 sec	
Final extention	72	2 min	1

2.2.9 Analysis of gene expression in *T. pseudonana*

2.2.9.1 Cell harvesting

Liquid cell cultures from the triple knockout (after the second conjugation) were measured for YFP (yellow fluorescent protein) on a NovoCyte® Flow Cytometer (ACEA Biosciences), and clones with different levels of fluorescence were added to 25 cm^2 Tissue Culture Flasks (VWR) containing f/2 medium and Nou (50 µg/ml), and incubated at 20°C with constant light (150-200 µmol m^{-2} s^{-1}). A flask containing the wild type was also grown in these same conditions. A week before harvest, cells from the different flasks were counted on the flow cytometer and different volumes of the clones were diluted in new flasks containing 40 ml f/2 with 40 µl Nou (50 µg/ml) to have a total cell number of 4000 cells/µl. The purpose of this was for the clones to reach roughly the same amount of cells before the harvest so they have the same baseline for further experiments.

Cell cultures were harvested when they reached an absolute cell count of ~900-1000 cells/µl. The liquid cell cultures were spun down at 3000 g for 10 min before removing most of the supernatant and transferring the rest (with the pellet) to Eppendorf tubes which were centrifuged at 3000 g for 10 min. Most of the supernatant was again removed from the tubes to make the pellets as dry as possible before they were snapfreezed in liquid nitrogen and immediately placed in the -80°C freezer.

2.2.9.2 RNA extraction

The tubes with the frozen pellets were evenly placed in two precooled (-80°C) TissueLyser Adapter Sets (QIAGEN) and one precooled (-80°C) 5 mm stainless steel bead (QIAGEN) was added to each tube. Disruption and mechanical lysis were done by placing the adapter sets in a TissueLyser instrument (QIAGEN) and shaking the samples at 25 Hz for 2 min. RLT lysis buffer (450 µl) containing 2-Mercaptoethanol (10 µl 2-Mercaptoethanol per 1 ml RLT buffer) was prepared using the RNeasy® Plant Mini Kit (QIAGEN) protocol and added to each tube before they were placed back on the TissueLyser for another 2 min of shaking at 25 Hz. The samples were then further lysed by incubating at 56°C for 4 min, and total RNA from the lysates was isolated according to the Rneasy® Plant Mini Kit (QIAGEN) protocol. In between adding the ethanol and the washing steps, an on-column Dnase digestion using the Rnase-Free Dnase Set (QIAGEN) was performed to remove residual contaminating genomic DNA (gDNA). Each sample was added a mixture of 10 µl Dnase stock solution with 70 µl RDD buffer before being incubated at room temperature for 15 min. The protocol that was used for the Dnase digestion was found in the Rneasy® Mini Handbook (Third Edition). The purified RNA was eluted in 40 µl Rnase-free water and the concentration was measured on the NanoDrop. The measured RNA concentrations were very low, so the RNA was eluted in another 40 µl before storing the tubes of total purified RNA (80 µl) at -80°C.

2.2.9.3 Synthesis of cDNA by reverse transcription of RNA

Synthesis of complementary DNA (cDNA) is done by reverse transcription of mRNA that has been extracted from an experimental sample. The cDNA that is produced can then be used as template in a fluorescence-based qRT-PCR reaction for gene expression analysis (Kuang et al., 2018).

The QuantiTect® Reverse Transcription Kit (QIAGEN) was used for cDNA synthesis. The RNA samples from the -80°C freezer were thawed on ice, and gDNA elimination reactions were prepared by mixing 2 µl gDNA Wipeout Buffer 7x, variable volumes of Template RNA (with a total of 271.2 ng), and Rnase-free water to a volume of 14 µl. A -reverse transcriptase (-RT) control, used to check gDNA contamination, was also prepared from the RNA sample that had the highest concentration. Extra volumes of the reaction components were added to this sample, where 3 µl gDNA Wipeout Buffer 7x and Rnase-free water (to a total volume of 21 µl) were added to 271.2 ng RNA. The reactions were incubated at 42°C for 10 min for gDNA elimination, followed by immediately being put back on ice where 7 µl of the 21 µl reaction were transferred to another tube as -RT control. Reverse-transcription master mix for the cDNA reactions and -RT control was then prepared (Table 2.9) and added to the tubes with template RNA. The reactions were incubated at 42°C for 30 min, followed by an incubation at 95°C for 3 min for inactivation of the reverse transcriptase before they were stored at -20°C.

Table 2.9. *QuantiTect reverse-transcription reaction.*

Component	cDNA mix (µl)	-RT mix (µl)
Reverse-transcription master mix		
Quantiscript Reverse Transcriptase	1	-
Quantiscript RT Buffer, 5x	4	2
RT Primer Mix	1	0.5
Rnase-free water	-	0.5
Template RNA		
gDNA elimination reaction	14	7
Total volume	20	10

2.2.9.4 Quantitative real-time PCR (qRT-PCR) analysis

qRT-PCR is a method that enables quantitative analyses of gene expression by comparing the relative levels of mRNA between biological samples. A fluorescent reporter dye that binds to the cDNA prepared from mRNA samples is added to the qRT-PCR reaction to measure the increase in the fluorescent signal as the DNA is amplified. The intensity of the signal corresponds to the amount of DNA that is synthesized with each PCR cycle, meaning that the signal will get stronger as more DNA is produced. The fluorescence of the sample will at some point surpass the background fluorescence, which results in a Ct (cycle threshold) value that represent the relative gene expression in that specific sample. Since the results are relative in comparative analyses such as this, housekeeping genes that are stably expressed are needed as a consistent reference. The initial concentration of mRNA plays a role when comparing samples to each other as higher concentrations of mRNA will give a lower Ct than a sample with less mRNA. A lower Ct is therefore equal to a higher gene expression in qPCR analyses (Jozefczuk and Adjaye, 2011).

The cDNA samples (15 µl) were diluted 1:3, and the plasmid ligated with the triple KO PAM was diluted 1:100, 1:1000 and 1:10000 before reactions were prepared according to Table 2.10 using the LightCycler® 480 SYBR Green I Master (Roche). All the primers that were added for the different genes can be found in Table A2-4 (Appendix 2). Reaction mix (15 µl) was pipetted into the wells of a LightCycler® 480 Multiwell Plate 96 (Roche), and 5 µl cDNA template/plasmid dilutions/-RT control was added to the wells. The -RT control was used to check for contaminating gDNA which might not have been completely removed during the gDNA elimination reaction when synthesizing cDNA by reverse transcriptase. No template controls (NTCs), which also served to detect gDNA contamination, were prepared for the different primer pairs where 5 µl PCR-grade water was added to the wells instead of DNA template. The 96-well plate was sealed with MicroAmp™ Optical Adhesive Film (Thermo Scientific) and spun down at 1500 rpm for 2 min. The program presented in Table 2.11 was used when running the qRT-PCR on the LightCycler® 96 instrument (Roche) and the LightCycler 96 software 1.1 (Roche) was used for analyzing the results, including the T_m and the Ct values.

Table 2.10. *SYBR Green qRT-PCR reaction mix.*

Component	(µl)
2x LightCycler 480 SYBR Green I Master	10
PCR primers (5 µM of each)	2
PCR-grade water	3
Total volume	15

Table 2.11. *Thermoprofile for qRT-PCR.*

Program	Temperature (°C)	Ramp (°C/s)	Time (s)	Cycles
Preincubation	95	4.4	600	1
3-step amplification	95	4.4	10	
	55	2.2	10	40
	72	4.4	10	
High resolution melting	95	4.4	60	
	65	4.4	60	1
	65-90	2.2	-	
Cooling	37	1.0	30	1

A confirmation that the different primers for the qRT-PCR analysis worked can be seen in Figure A2-1 (Appendix 2).

2.2.10 Fluorescence microscopy

Cell culture samples (10 µl) of WT and triple KO clones from the second conjugation were pipetted onto VWR® Microscope Slides which were covered with 20x20 mm Cover Glass before being placed in the Zeiss Axio Imager.Z2 microscope to study cell shape and size. Imaging of the cells was done though a 40x objective with the Zeiss Axiocam MRm microscope camera.

Nucleic acid stain with DAPI (4′,6-diamidino-2-phenylindole) was also done on the WT and a few clones from the triple KO in order to visualize the nucleus and its location in the cells. This was done by adding 1 µl DAPI (1 µg/ml, Merck) in MQ water to 200 µl cell culture and incubating the mix in the dark for 15 min before studying the cells in the microscope. These clones were imaged with a 40x objective. The fluorescence channels used for imaging were a green fluorescence channel to detect YFP, a red fluorescence channel to detect autofluorescence from the chloroplasts, and a blue fluorescence channel to detect DAPI.

2.3 Fusion of mNeonGreen with *Tp14322* and *Tp264671*

2.3.1 Vector and linearization

The vector pTpPUC3 (Figure 2.3), which was also used in a study by Karas et al. (2015), was used for cloning of the gene fragments from the two genes as well as the fluorescent protein mNeonGreen. This vector, along with an inserted Cas9 gene (with EYFP tag) and sgRNA region, was the one that was used for mutant creation in section 2.2. It therefore contains the same regions and genes such as for the antibiotic resistance. The vector also contains several restriction sites that were used to linearize it in experiments, such as SalI, PstI, XbaI and EcoRI.

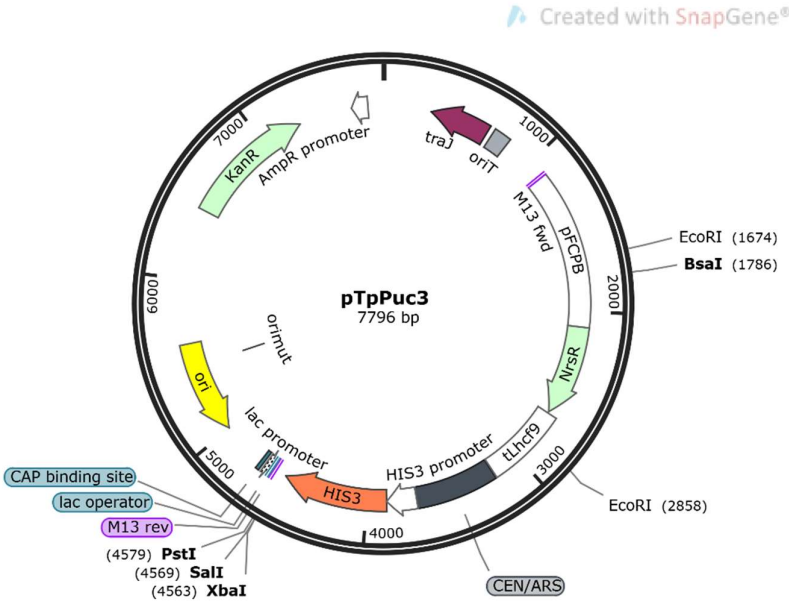

Figure 2.3. *Map of the pTpPUC3 vector that was used to insert gene fragments along with the fluorescent protein mNeonGreen.*

A restriction digestion was done to linearize the plasmid before insertion of PCR-products. The plasmid was cut by adding 1 µl SalI-HF restriction enzyme to 15 µl pTpPUC3 vector DNA (36.8 ng/µl), as well as 5 µl 10x CutSmart Buffer (NEB) and MQ water to a total volume of 50 µl. This mixture was incubated at 37°C for 15 min before being stored at -20°C for later use.

A new pTpPUC3 plasmid with a fixed (mutated) ori was also later used in a couple experiments for gene fusion. This plasmid has a single base pair mutation in the bacterial ori which makes the plasmid more stable.

2.3.2 Amplification of gene fragments and mNeonGreen

Primers (Table A2-5, Appendix 2) were designed for amplification of three fragments for the two genes (Fig. 2.4): the promoter of the gene and coding region, mNeonGreen, and the terminator of the gene. The amplification of the fragments was done by using the Phusion® High-Fidelity DNA Polymerase (Thermo Scientific). The DNA template that was added to the master mix was extracted from 50 µl *T. pseudonana* culture that was lysed the same way as described in section 2.2.4. For amplification of the mNeonGreen fragment, a specific mNeonGreen template was used. The rest of the reaction components, as well as the primers, were mixed as shown in Table 2.12, and the same Phusion PCR program as in Table 2.4 was run, except for changing the annealing temperature to 55°C and the extension time to 30 sec.

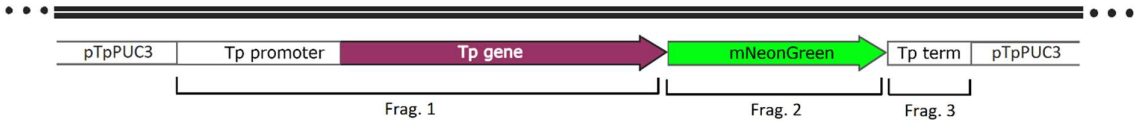

Figure 2.4. *General overview of the different amplified fragments and how they should be structured when inserted into the plasmid. Both Tp14322 and Tp264671 are structured like this.*

Table 2.12. *Phusion PCR reaction (50 µl).*

Component	Volume (µl)	Concentration
5x Phusion HF Buffer	10	1x
10 mM dNTPs	1	200 µM each
Forward primer (10 µM)	1.25	0.25 µM
Reverse primer (10 µM)	1.25	0.25 µM
Template DNA / mNeonGreen template	1 / 0.5	
Phusion Hot Start II DNA polymerase (2 U/µl)	0.5	0.02 U/µl
MQ water	35 (to 50)	
Total volume	50	

A portion of the PCR products (5 µl) was loaded on a 1% agarose gel in 1x TAE buffer and run on 100 V for 1 h before gel documentation was done in the G:BOX (Syngene). Certain fragments (*Tp264671* promoter+gene and terminator for both genes) appeared to be challenging to amplify, thus new primers (Table A2-5, Appendix 2) were designed and ordered. Four samples of each missing fragment were prepared as shown in the Phusion reaction in Table 2.3, except that 0.6 µl $MgCl_2$ was also added to the mix. The ThermoCycler was programmed to contain a temperature gradient (57.0°C-67.0°C) in the annealing step, and each of the four samples were placed in the following rows with temperature: A-

67°C, C-65°C, E-60.8°C and H-57°C. Extension times were also further adjusted in correspondence to the fragment lengths, with 70 sec for *Tp264671* promoter+gene, and 7 sec for the terminators. The PCR program was otherwise followed as in Table 2.4. Successful amplification of the remaining fragments was confirmed by running a 1% agarose gel in 1x TAE buffer at 110 V for 1 h. All amplified fragments were then sent in for Sanger sequencing to confirm that they were of good quality and ready to be cloned into pTpPUC3.

2.3.3 Cloning methods for inserting fragments into pTpPUC3

With successful cloning, the plasmid with all the fragments is supposed to be structured as shown in Figure 2.4, with mNeonGreen in between the coding region of the gene and the terminator. However, cloning all the fragments into the plasmid appeared to be a challenging task, and different cloning strategies were therefore tested.

2.3.3.1 Gibson assembly

Gibson assembly (Figure 2.5) is a method of enzymatic assembly of DNA fragments concerted by the actions of a mix containing a DNA polymerase, a DNA ligase and a 5' exonuclease (Gibson et al., 2009). Inserts that contain sequences on both ends that overlap with the sequence close to the cleavage site in the vector are amplified by PCR, and T5 exonuclease will chew back on the ends of the vector and insert to create 3' overhangs that expose the homologous sequences. The vector and the insert anneal before 3'-5' exonuclease activity by the DNA polymerase removes the protruding 3' overhangs. Gaps in the sequence are filled in by the DNA polymerase, and nicks in the filled-in gaps are repaired by the ligase to form a closed double-stranded plasmid (Wang et al., 2015).

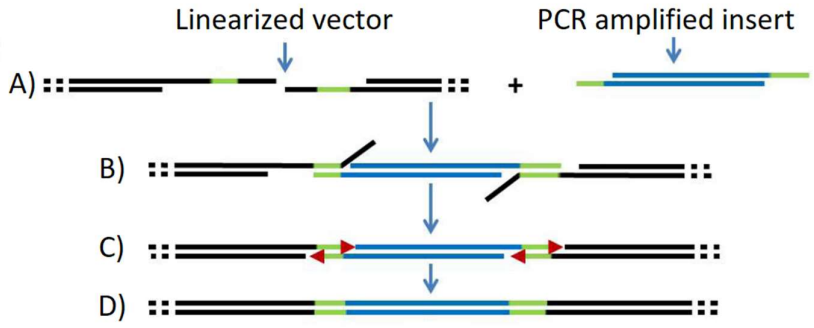

Figure 2.5. *Schematic representation of the Gibson assembly cloning method. A) Long overhangs of the insert and the linearized vector are generated by the 5' exonuclease which expose the homologous sequences (showed in green) between the two fragments. B) Protruding overhangs are eliminated by 3'-5' exonuclease activity by the DNA polymerase after the fragments have annealed. C) The polymerase and ligase fill in and repair in the gaps (indicated by red arrows) to D) form a double-stranded plasmid. Image adapted from Wang et al. (2015).*

Different versions to assemble a product using the Gibson assembly cloning method were attempted as explained in the following subsections:

2.3.3.1.1 Assembly based on fragment length

In two tubes, one for each gene, variable amounts of the different amplified fragments were mixed with components from the Gibson Assembly® Master Mix (NEB) as presented in Table 2.13. The length of the fragments, which was used to decide the volume of the insert in the reaction mix, can be found in Table A2-5 (Appendix 2) as amplicon length for the different primer pairs.

Table 2.13. *Gibson assembly reaction mix.*

Component	**Volume (µl)**	
pTpPUC3 plasmid (36.8 ng/µl)	3	
PCR fragments	*Tp14322*	*Tp264671*
Promoter+gene	2	3
mNeonGreen	1.5	0.5
Terminator	0.5	0.5
Gibson Assembly Master Mix (2x)	10	
MQ water	3	
Total volume	20	

The tubes were incubated at 50°C for 60 min, followed by 2 µl of the Gibson assembly reactions being transformed into 50 µl DH5α *E. coli* cells as described in section 2.2.3. After overnight incubation 10-15 colonies on each plate were picked for colony screening. The master mix was prepared as in Table 2.1, containing a gene-specific forward primer (Table A2-2, Appendix 2) as well as M13 as the reverse primer (Table A2-7, Appendix 2). The PCR program was run as showed in Table 2.2, with the exception that the annealing temperature was set to 55°C, and the extension time was changed to 3 min, 24 sec. The PCR products were run a 1% agarose gel (60 ml mixed with 3 µl GelRed) in 1x TAE buffer at 120 V for 1 h.

Gibson products were also transformed into 50 µl Stable Competent *E. coli* (NEB). According to the manufacturer NEB, these cells give a higher efficiency with less mistakes for unstable inserts. The transformation process (and colony screen) was the same as in section 2.2.3, with the exception that incubation was set to 30°C on all steps instead of 37°C. A gel was run at 90 V for 120 min, and clones that looked like they had a target band were inoculated and incubated at 30°C overnight. A plasmid isolation using the GeneJET Plasmid Miniprep Kit (Thermo Scientific) was done before measuring the concentrations on the NanoDrop and sending the samples in for Sanger sequencing (7.5 µl sample with 2.5 µl M13r primer).

A restriction digestion was done on the clones that were sent in for Sanger sequencing to check the length of the plasmid. Restriction enzymes XbaI and EcoRI (0.25 µl each) were added to variable volumes of sample (to a total of ~100 ng), as well as 1 µl 10x CutSmart Buffer (NEB) and MQ water to a total reaction volume of 10 µl. The reactions were incubated at 37°C for 30 min before the full volume was loaded on a 1% agarose gel. The gel was run at 70 V for 2 h and post-stained by placing the gel in 50 ml MQ water containing 15 µl GelRed for 20-30 minutes.

NEBuilder, an online assembly tool, was used to check the cloning strategy, and as there appeared to be no faults in it, new assemblies were attempted with the same plasmid and fragments.

2.3.3.1.2 Assembly with purified PCR products

A purification with the ExS-Pure™ Enzymatic PCR clean-up kit (NimaGen) was done as described in section 2.2.6. The concentration (ng/µl) of the purified products were then measured on a Qubit™ 3.0 Fluorometer (Invitrogen). The Qubit working solution was prepared accordingly to the manufacturer's manual, and 1 µl PCR product was mixed with 199 µl working solution. The mixtures were vortexed for 2-3 sec and incubated at room temperature for 2 min before inserting the assay tube into the fluorometer for reading. As there seemed to be sufficient amounts of product after measuring, purified PCR products were set up in a new Gibson assembly reaction (same as in Table 2.13). The products were transformed into DH5α cells and colonies were screened the same way as described previously in this section. Gibson products from both before and after enzymatic purification were run on a 1% agarose gel in 1x TAE buffer at 100 V for 1 h to check if the plasmid contained the inserts.

2.3.3.1.3 Assembly of fragments with known concentrations

New reactions for all the fragments were set up as in Table 2.12. DMSO (1 µl) was also added to the mix this time, as well as new primer working solutions being made. The PCR program was the same as in Table 2.4, except for changing the annealing temperature to 55°C, and the extension times to 60 sec for the promoter+gene and 30 sec for mNeonGreen and terminator for both gene fragments. The PCR products were run on a 1% agarose gel at 100 V for 1 h. Two of the products gave strong and clear bands, but as other unspecific bands were also present, an isolation of the target band was preferable. All the remaining sample of the two products were loaded onto a very thick 1% agarose gel which was run at 100 V for 1 h. A scalpel was used to cut out the target band, making sure to get as much of the band as possible without excessive surrounding gel, and it was put in an Eppendorf tube. The DNA extraction from the gel was done by using the Monarch® DNA Gel Extraction Kit (NEB) and following the recommended protocol. The DNA was eluted in 15 µl room temperature MQ water. These remaining fragments, as well as former amplified ones, were purified by using the Monarch® PCR & DNA Cleanup

Kit (5 µg) (NEB). By following the provided protocol, sample volumes of 20 µl were purified and eluted in 15 µl MQ water. The concentration of all the purified samples were measured on the NanoDrop before being entered into the NEBioCalculator, a tool that calculates the insert mass required at different insert:vector ratios for ligation reactions. A new Gibson Assembly reaction was set up (Table 2.14) using the purified PCR products as well as a longer cut version of the same pTpPUC3 plasmid that was used earlier. The insert:vector ratio in the reactions varied from 2:1 to 4:1 depending on the concentration and sample volume.

Table 2.14. *Gibson Assembly reaction mix with purified fragments.*

Component	**Volume (µl)**	
pTpPUC3 plasmid (36.8 ng/µl)	1	
PCR fragments	*Tp14322*	*Tp264671*
Promoter+gene	6	7.5
mNeonGreen	1	1
Terminator	2	0.5
Gibson Assembly Master Mix (2x)	10	
Total volume	20	

The reactions were incubated at 50°C for 60 min on the ThermoCycler before 5 µl were transformed into 50 µl Stable Competent *E. coli* (NEB). Colony screening reactions were set up using the DreamTaq™ DNA Polymerase (Thermo Scientific) (Table 2.15). The PCR program was run as showed in Table 2.2, with the exception that the annealing temperature was set to 55°C, and the extension time was changed to 3 min, 24 sec. The PCR products were run a 1% agarose gel (60 ml mixed with 3 µl GelRed) in 1x TAE buffer at 120 V for 1 h.

Table 2.15. *DreamTaq PCR reaction setup.*

Component	**Volume (µl)**	**Concentration**
10x DreamTaq Buffer	2	1x
dNTP mix (2 mM each)	0.4	200 µM each
Forward primer (10 µM)	1	0.5 µM
Reverse primer (10 µM)	1	0.5 µM
DreamTaq DNA Polymerase (5 U/µl)	0.2	0.05 U/µl
DNA template	0 (pick)	<1000 ng
MQ water	15.4 (to 20)	
Total volume	20	

2.3.3.1.4 Nested Gibson assembly

Nested Gibson assembly is a method that increases efficiency in reactions containing many fragments as the number of DNA pieces going into the Gibson reaction decreases. Larger regions containing multiple fragments are amplified by PCR before being ligated into the vector. Gibson products from failed assemblies can be used as template in the amplification as some DNA pieces might still be ligated together in the failed reaction (Garza and Bielinski, 2020).

Two fragments per gene, promotor+gene and mNeonGreen+terminator, were amplified instead of three fragments as done previously. The Q5® High-Fidelity 2X Master Mix (NEB) was used to set up PCR reactions containing 12.5 µl Q5 High-Fidelity 2X Master Mix, 1.25 µl (10 µM) of each forward and reverse primer (Table A2-5, Appendix 2), 1 µl of the Gibson assembly product from section 2.3.3.1.1 as template, and 9.5 µl MQ water to a total reaction volume of 25 µl. The reactions were run on the cycling conditions shown in Table 2.16, with varied extension times depending on the fragment length. A 1% agarose gel was run for 1 h at 100 V.

Table 2.16. *Thermoprofile for nested Gibson assembly.*

Program		Temperature (°C)	Time (s)	Cycles
Denaturation		98	10	30
Annealing		65	15	
Extension	***Tp264671* promoter+gene**	72	24	
	***Tp14322* promoter+gene**		16	
	mNeonGreen+terminator		10	

A pTpPUC3 plasmid check was done in order to see whether the plasmid was properly cut or not. Approximately 15 µl (552 ng/µl) was loaded onto a gel that was run at 110 V for 50 min and post-stained.

2.3.3.1.5 Gibson assembly with fixed ori plasmid

A new Gibson assembly using the pTpPUC3 FO4 (fixed ori) (~1 µg/3 µl) was done to ligate in a FCPB promotor, mNeonGreen/mTurquoise and a LHCF terminator. This way only one fragment, the gene of interest, could be ligated in instead of multiple fragments as attempted earlier, and thereby increasing the efficiency. Volumes of the inserts used varied based on their length (primers and length of inserts can be found in Table A2-6 in Appendix 2) as the concentration of the inserts were not measured. The pTpPUC3 FO4 was cut with 0.5 µl PstI restriction enzyme, and 3 µl of the plasmid were transferred to two new tubes, one reaction for mNeonGreen and one for mTurquoise. The 3 µl pTpPUC3 was mixed

with 2 µl pFCPB, 1 µl mNeonGreen/mTurquoise, 1 µl tLHCF9, 10 µl Gibson Assembly Master Mix, and 3 µl MQ water for a total volume of 20 µl. The two reactions were incubated at 50°C for 60 min before 5 µl Gibson product were transformed into 50 µl Stable Competent *E. coli* (NEB). Colonies that appeared on the plate after overnight incubation on 30 °C were screened as done in section 2.2.3.

2.3.3.2 Circular polymerase extension cloning (CPEC)

CPEC is a one-step method that can assemble and clone multiple inserts with any vector by using the mechanism of polymerase extension. Overlapping DNA fragments are joined together during the annealing step in the PCR program to form a double-stranded circular form, such as a plasmid. No restriction digestion or ligation is therefore required in this method (Quan and Tian, 2014).

CPEC reactions for the two genes were set up as shown in Table 2.17 by using the Phusion® High-Fidelity DNA Polymerase (Thermo Scientific) and the same volume of inserts was used as in section 2.2.3.1.1. A new restriction digestion of the pTpPUC3 plasmid (36.8 ng/µl) was done the same way as in section 2.3.1, except that it was incubated at 37°C for 1 h instead of 15 min. The fragments were assembled according to the PCR program in Table 2.18.

Table 2.17. *CPEC PCR reaction.*

Component	**Volume (µl)**	
pTpPUC3 plasmid (36,8 ng/µl)	3	
PCR fragments	*Tp14322*	*Tp264671*
Promoter+gene	2	3
mNeonGreen	1.5	0.5
Terminator	0.5	0.5
5x Phusion HF Buffer	5	
10 mM dNTPs	1	
DMSO	0.75	
Phusion Hot Start II DNA polymerase (2 U/µl)	0.5	
MQ water	10.75	
Total volume	25	

Table 2.18. *CPEC PCR program.*

Program	**Temperature (°C)**	**Time**	**Cycles**
Initial denaturation	98	30 sec	1
Denaturation	98	10 sec	15
Annealing	55	30 sec	
Extention (15 sec/kb) ***Tp14322*****/ Tp264761**	72	2 min 35 sec/ 2 min 50 sec	
Final extention	72	5 min	1

The CPEC products were run on a 1% agarose gel at 100 V for 1 h before 5 µl were transformed into 50 µl stable *E. coli* cells (NEB). No screening was done as there were no growth on the plates. Another CPEC was therefore set up as in Table 2.17, except that the pTpPUC3 FO4 plasmid (~1 ng/3 µl) was used, and the fragment volumes were changed to 2 µl of the FCPB promoter and 1 µl of both mNeonGreen/mTurquoise and the LHCF9 terminator. Both the pFCPB and tLHCF9 were DpnI (20,000 units/ml, NEB) digested before insertion. According to the manufacturer NEB, DpnI is a restriction enzyme that may enhance cloning by cutting backbone that might be in the sample after amplification, but not the PCR product itself.

3 Results

3.1 Kinase domain analysis

A protein alignment analysis of *Tp264671* and *Tp14322* against the WT sequence was done in NCBI to investigate which amino acids were conserved in what domains of the kinases. The analyzed sequences of the genes both got a "domain hit" and a "specific hit" (as seen in Figure 3.1) that showed that CaMK is the catalytic domain. The E-values, which were very low, 8.23e-132 for *Tp264671* and 1.75e-129 for *Tp14322*, indicates that there is a very high probability that the catalytic domain is a CaMK-type kinase. The E-value is, according to NCBI, a parameter that is described as the random background noise that is caused by the number of hits expected to see by chance when comparing a pasted sequence against a database of a specific size. A consensus sequence in SnapGene® Software showed the most frequently occurring amino acids at certain position in the kinase alignments, also including *Tp14242*. This confirmed the results of what was found in the NCBI alignment as the same conserved amino acids were found at the same places in the sequence in both analyses. The conservation number (describing how often an amino acid occur in a set of alignments) was generally very high for the conserved amino acids that was found in the three kinase alignments.

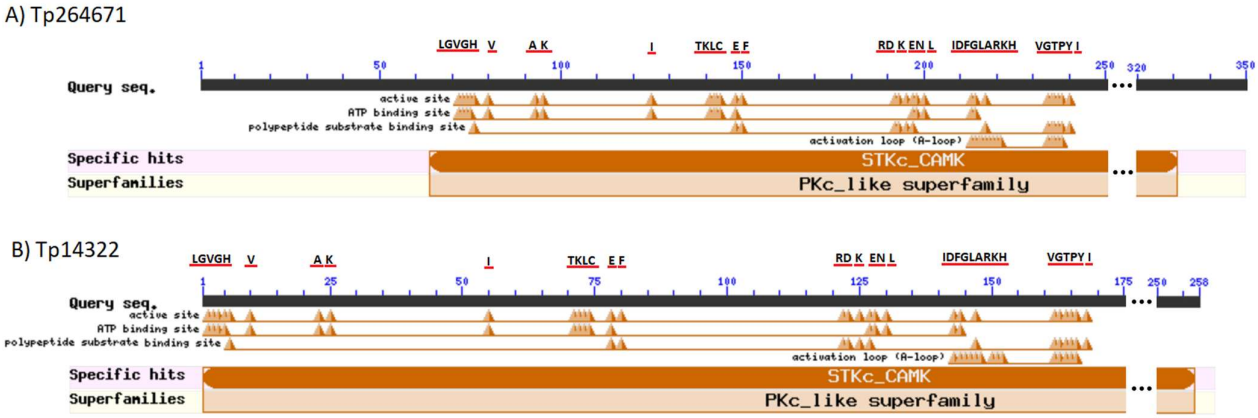

Figure 3.1. *Map (from NCBI) of the conserved amino acids in the catalytic CaMK domain in A) Tp264671 and B) Tp14322. Peaks in the map signifies conserved amino acids found in the sequences, which are depicted with a red underscore above the query sequence.*

The sequence alignments for the three kinase-encoding genes in SnapGene were also compared to the CaMK2α sequence found in Hanks and Hunter (1995) to see if there were any deviations in the conserved amino acids in the different subdomains of the catalytic domain (Table 3.1). Deviations between the three genes were also investigated. Differences in conserved amino acids between the sequences for the three kinase-encoding genes in the SnapGene file (in comparison to the Hanks and Hunter (1995) CaMK2α sequence) were as followed: in subdomain VI B both leucine, found for *Tp14242*, and isoleucine, found for *Tp264671* and *Tp14322*, are nonpolar amino acids; in subdomain VIII both glutamate for *Tp14242* and aspartate for *Tp264671* and *Tp14322* are acids; in subdomain IX there is a difference in the physical-chemical properties where isoleucine (nonpolar) is found for *Tp14242* and *Tp264671* while threonine (polar) is found in the same position for *Tp14322*.

Table 3.1. *Deviations in amino acids in a few different subdomains of the catalytic CaMK domain. HH: CaMK2α amino acid sequence from Hanks and Hunter (1995). SG: Kinase domain alignment (of catalytic domain in Tp264671, Tp14322 and Tp14242) from SnapGene file provided by supervisor Tore Brembu. Red letters indicate collective deviations in the conserved amino acids between the kinase-encoding genes and the CaMK2α amino acid sequence. The blue "X" indicate a difference in amino acids in between the three kinase-encoding genes in the SnapGene file. Not all subdomains are included.*

Alignment platform	**Subdomain**	**Conserved amino acids**	**Deviating amino acid between the three kinases (X)**
HH	VI B	VVHRDLKPENLLL	*Tp14242*: L
SG		VVHRDXKPENILF	*Tp264671*/14322: I
HH	VIII	TPGY SPEVL	*Tp14242*: E
SG		TPYY APXVL	*Tp264671*/14322: D
HH	IX	WA GVI YILL GYPP	*Tp14242*/264671: I
SG		WS GVI YXLL GYPP	*Tp14322*: T

3.2 Creating knockout cell lines of *Tp14322* and *Tp264671*

Oligos containing different target sites were inserted into the pTpPUC3-Cas9-M-G1 plasmid for CRISPR/Cas9-based genome editing in the kinase-encoding genes *Tp14322*, *Tp264671* and the related *Tp14242*. Plasmids with successfully inserted target sites were introduced into *T. pseudonana* by bacterial conjugation and cultured in medium containing Nourseothricin for selection of the clones containing the plasmid. The expressed Cas9 and sgRNA could thereby combine to cut in the algae cells genome to produce mutations in the process of generating of single-, double-, and triple-knockouts. In addition, a small investigation was done to find out where in the amino acid sequence the PAMs were located (Table 3.2).

Table 3.2. *Overview of amino acids found in the different PAM sites.*

Gene	PAM	Strand	General location of PAM in gene sequence	Position of PAM in amino acid sequence
Tp264671	r42	Reverse	Upstream of kinase domain	33 (Leu)
	r291	Reverse	In kinase domain	116 (Arg)
Tp14322	r77	Reverse	Upstream of kinase domain	26 (Pro)
	r392	Reverse	In kinase domain	131 (Ala)
***Tp264671*-14322-14242**	f233	Forward	In kinase domain	75 (Gly)
***Tp264671*-14322**	f436	Forward	In kinase domain	142 (Glu)
	f754	Forward	In kinase domain	246 (Lys)

3.2.1 Vector construction

pTpPUC3 Cas9 M G1 plasmids were transformed into DH5α *E. coli* cells, and three colonies from each plate with each target site were picked for colony screening with gel electrophoresis. As seen in Figure 3.2, at least two clones from each plate with a specific PAM gave a product with the intended length of 200 bp, except for clones inserted with the target site containing PAM f754, which gave no product. More colonies for the f754 PAM were screened and new attempts to insert the oligo were done, however none of them were successful. Further screening and experiments on this double KO PAM were therefore not done. Confirmation that the oligos were successfully inserted was done by Sanger sequencing.

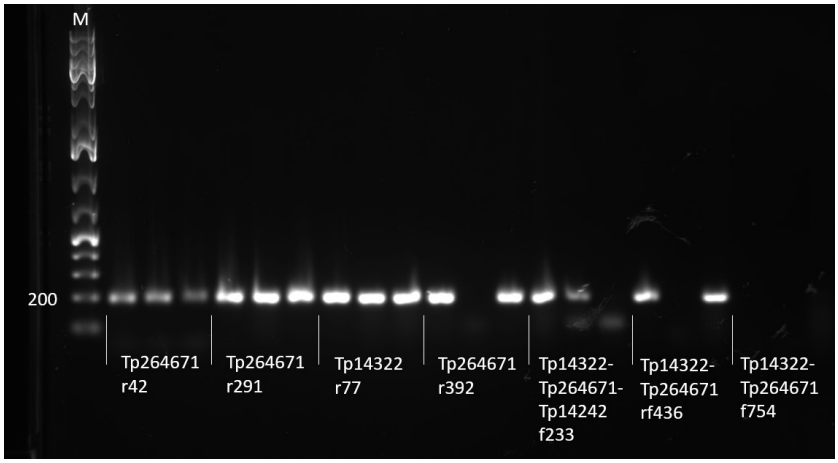

Figure 3.2. *Colony screening of DH5α E. coli transformed with pTpPUC3 plasmid inserted with the different target sites. Three colonies were picked from each transformation plate containing the different target sites and screened to identify which ones contained the insert. GeneRuler 1 kb Plus DNA ladder (Thermo Scientific) was used as a marker to identify the sizes of the products (M). The number to the left of the marker indicates the length (bp) of the fragment.*

3.2.2 Screening for mutations in *T. pseudonana* transformants

Lysates that were extracted from liquid algae cultures were used to amplify regions with the specific target sites. PCR products were analyzed by gel electrophoresis to check for deviations from the WT, something that could indicate mutations. Most of the clones, especially the ones that deviated significantly from the WT, were picked for further screening with HRM to potentially also discover smaller mutations in the sequences. Investigation of putative mutations was also done by Sanger sequencing.

3.2.2.1 Screening mutations in single knockout PAM clones

T. pseudonana clones with target site for *Tp264671* r42 had one clone, number 4, which deviated significantly from the WT (868 bp for *Tp264671*) as it gave bands with sizes ranging from ~700-3000 bp, indicating that this clone might have a large insertion (Figure 3.3 A). The rest of the clones with this target site did not show any signs of mutations on the gel; however, smaller mutations that would not result in a band on the gel could still be present. The reason *Tp264671* r42 clone 1, 2, 7, 9, 11 and 12 did not give a PCR product could be due to unsuccessful amplification in general, or perhaps by larger deletions removing the binding site for either one or both primers, leading to no products being amplified. Deletion of sites for primer binding was, however, confirmed to not be the case. On the other hand, several clones with a target site for *Tp14322* r77 deviated from the WT (998 bp for *Tp14322*) (Figure 3.3 B and C). Clones such as 5, 6, 11, 15, 20 and 24 for the r77 target site contained three bands with sizes from 500-1500 bp, while two clones (7 and 8) had two bands at approximately 700 bp and 1500 bp.

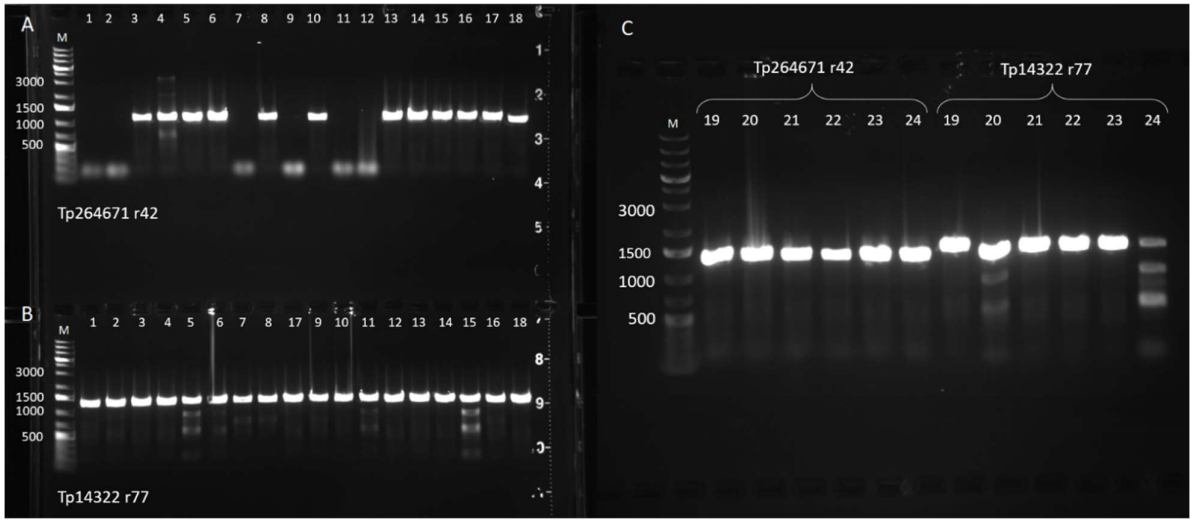

Figure 3.3. *Gel electrophoresis analysis with 1.0% agarose showing amplified regions containing the target sites for Tp264671 r42 and Tp14322 r77. A) shows the screening of Tp264671 r42 clone 1-18, B) Tp14322 r77 clone 1-18, and C) clone 19-24 for both Tp264671 r42 and Tp14322 r77. The numbers on the top of each subfigure signifies the picked clones for the specific target sites. GeneRuler 1 kb Plus DNA ladder (Thermo Scientific) was used as a marker to identify the sizes of the products (M). The numbers to the left of the markers indicate the length (bp) of the fragments.*

Another single target site for each of the genes, containing r291 for *Tp264671* and r392 for *Tp14322*, were colony screened and assessed by gel electrophoresis. Clones with target site for r291 (Figure 3.4 A and C) showed a lot of variety in band sizes and patterns as opposed to clones with target site for r392 (Figure 3.4 B and C), which did not indicate any mutations or indels on the gel. Band sizes for r291 clones ranged from approximately 500-3000 bp, and about half of the clones did not give a band where the WT would normally give one (868 bp for *Tp264671*). The reason for this could be, similarly to six of the clones for *Tp264671* r42, that the amplification was unsuccessful.

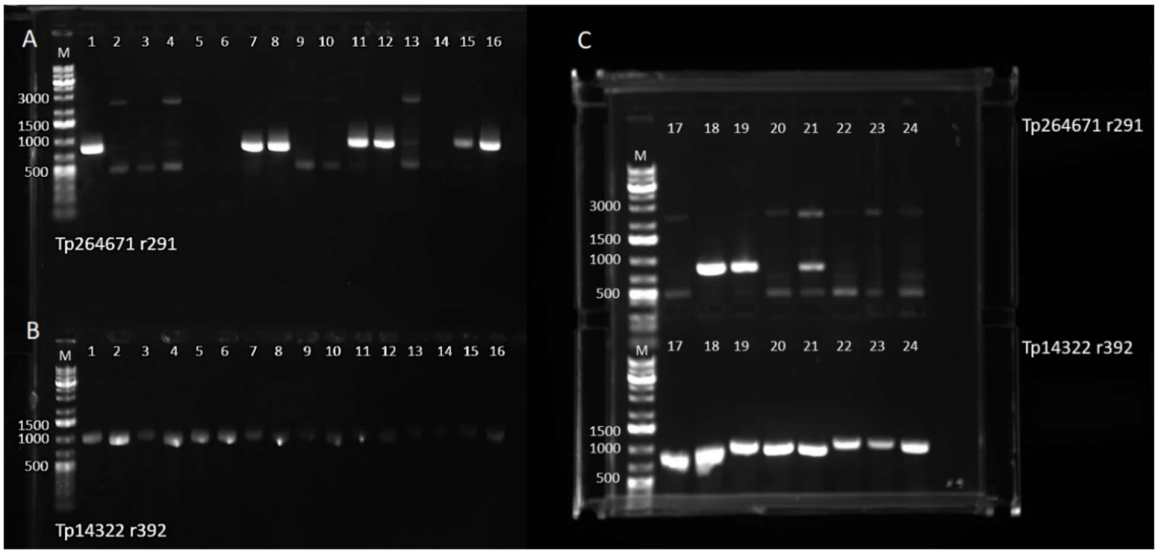

Figure 3.4. *Gel electrophoresis analysis with 1.0% agarose showing amplified regions containing the target sites for Tp264671 r291 and Tp14322 r392. A) shows the screening of Tp264671 r291 clone 1-16, B) Tp14322 r392 clone 1-16, and C) clone 17-24 for both Tp264671 r291 and Tp14322 r392. The numbers on the top of each subfigure signifies the picked clones for the specific target sites. GeneRuler 1 kb Plus DNA ladder (Thermo Scientific) was used as a marker to identify the sizes of the products (M). The numbers to the left of the markers indicate the length (bp) of the fragments.*

Clones were screened for mutations in the target sites by using the PCR products for the initial screening as template in an HRM analysis. By using HRM, smaller mutations that are not apparent using gel electrophoresis can be detected in the target DNA sequence as smaller regions are amplified and analyzed. As seen in Figure 3.5, clones with certain target sites seemed to be more efficient for gene editing than others as they differed more from the WT. Clones with target site for *Tp264671* r42, *Tp14322* r77 and *Tp14322* r392 seemed to have little deviation from the WT (Figure 3.5 A, C, D), although the samples showed a slightly higher T_m than that of the WT. Clones for r291, however, deviated significantly from the WT as the T_m for the clones was generally ~7-8°C lower than the T_m for the WT (Figure 3.5 B). This big gap in the melting temperatures could be a sign that mutations occurred in the r291 clones. Two of the *Tp264671* r291 clones (2 and 13) also seemed to have a slight double-peak, something that could indicate a mix of WT and putative mutant cells in the cultures (Figure 3.5 D). An almost double-peaked melting curve for the WT in Figure 3.5 D could be a sign of a polymorphism, although sequencing results confirmed that was not the case.

Clones that deviated the most from the WT were sent in for Sanger sequencing: 1, 5, 13-15 for r77; 4, 6, 17, 18, 23 for r42; 1-3, 9, 11, 12 for r391; and 2, 4, 7, 9, 13, 15, 21, 24 for r291. By batch analyzing the sequencing results in ICE (Synthego), it was confirmed that no mutations had occurred in any of the clones with the different target sites.

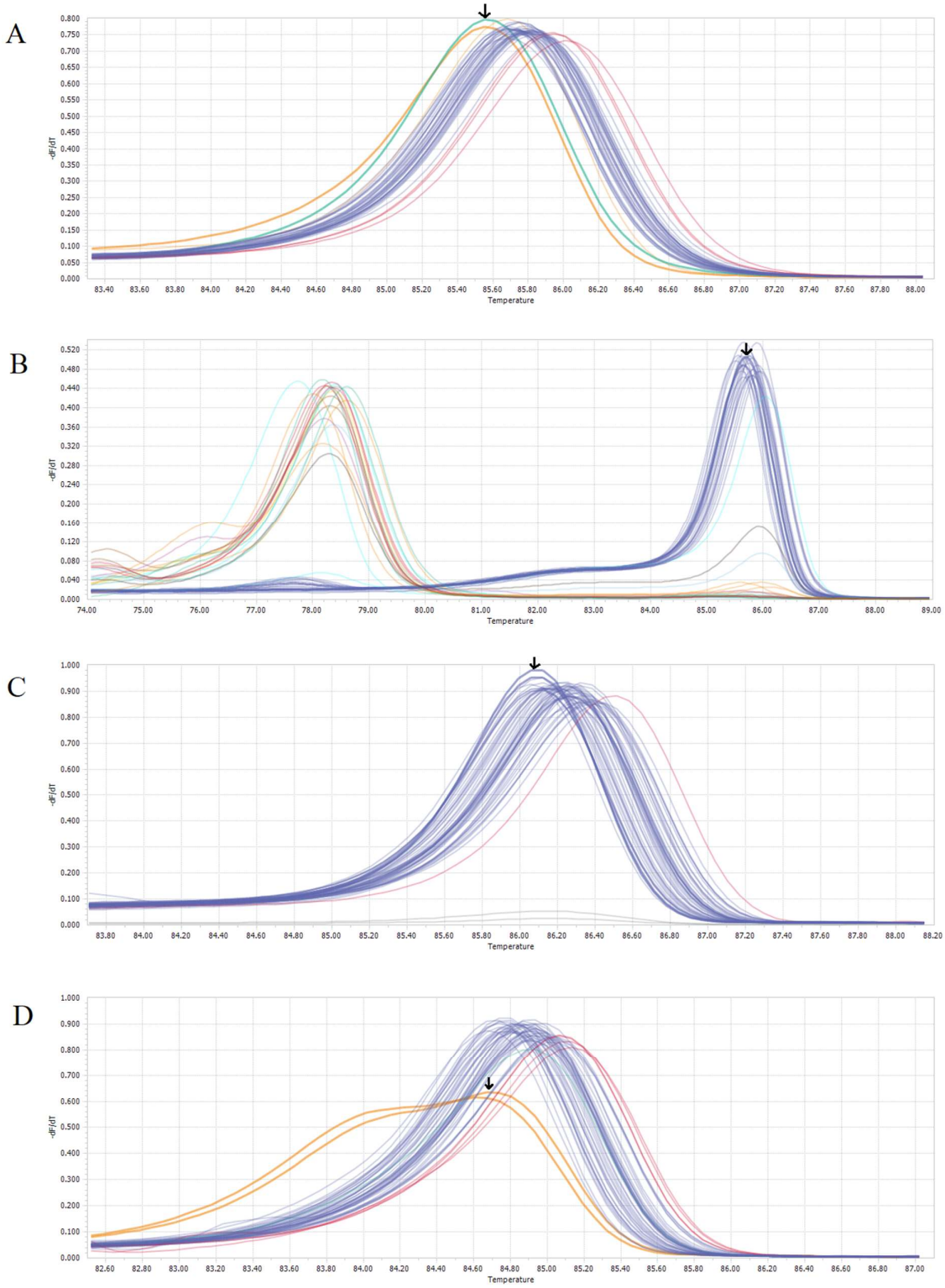

Figure 3.5. *High resolution melting curves from screening of single KO clones with target sequence A) Tp264671 r42, B) Tp264671 r291, C) Tp14322 r77, and D) Tp24311 r392. Melting peaks for WT are marked by the black arrows.*

3.2.2.2 Screening for mutations in double knockout PAM clones

The two genes, *Tp264671* and *Tp14322*, were also investigated for mutations in the target region for PAM f436, which has an identical sequence in both genes and can therefore be used to generate double KOs. Gel electrophoresis analysis showed that no mutations seemed to be detected for the *Tp264671* clones (Figure 3.6 A and C). Multiple weak bands could, however, be seen for some of the *Tp14322* clones (5, 7, 9, 10, 12, 15 and 16), with sizes ranging from about 500-1500 bp (expected WT size is 998 bp) (Figure 3.6 B). As there were no successful attempts to insert the oligo for PAM f754 (another double KO target region) into the pTpPUC3 vector, no analyses were done with this target site.

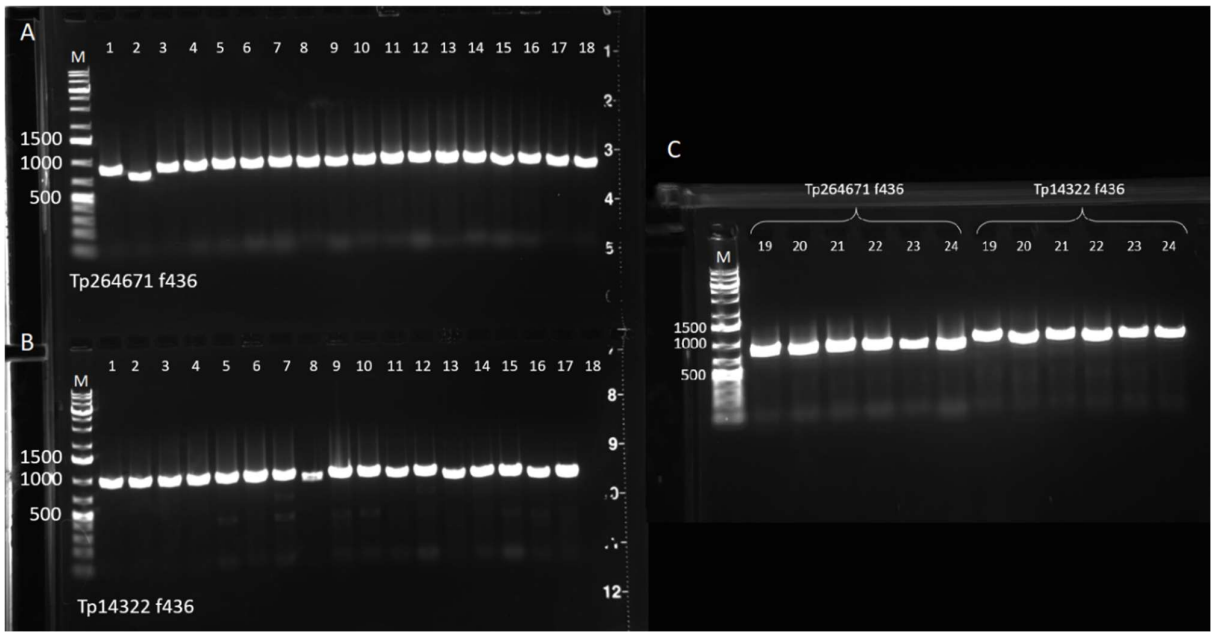

Figure 3.6. *Gel electrophoresis analysis with 1.0% agarose showing amplified regions containing PAM f436 for target site in both Tp264671 and Tp14322. A) shows the screening of Tp264671 f436 clone 1-18, B) Tp14322 f436 clone 1-18, and C) clone 19-24 for both Tp264671 f436 and Tp14322 f436. The numbers on the top of each subfigure signifies the picked clones for the specific target sites. GeneRuler 1 kb Plus DNA ladder (Thermo Scientific) was used as a marker to identify the sizes of the products (M). The numbers to the left of the markers indicate the length (bp) of the fragments.*

The amplified region containing the f436 target site seemed to be more efficient for editing in *Tp14322* than in *Tp264671* when investigating putative mutations with HRM (Figure 3.7). Clones from *Tp264671* showed minimal deviation from WT (Figure 3.7A), while clones from *Tp14322* deviated a lot more (Figure 3.7 B). One clone in particular, number 4 from *Tp14322*, stood out as it had a T_m that was ~3.5°C lower than the T_m of the WT. Melting curves with a double-peak were also observable for multiple *Tp14322* clones (3-7, 15) in the HRM analysis, indicating that there might be a mix of putative mutant cells and WT. Although no mutations seemed to have occurred in the *Tp264671* clones, several of the *Tp14322* clones appear to show signs of editing. Clone 3-6, 17, 18, 20 and 21, which seemed to deviate

the most from the WT in *Tp14322*, were sent in for Sanger sequencing. Analyzing the results from the sequencing in ICE (Synthego) confirmed that no mutations had occurred in any of the clones.

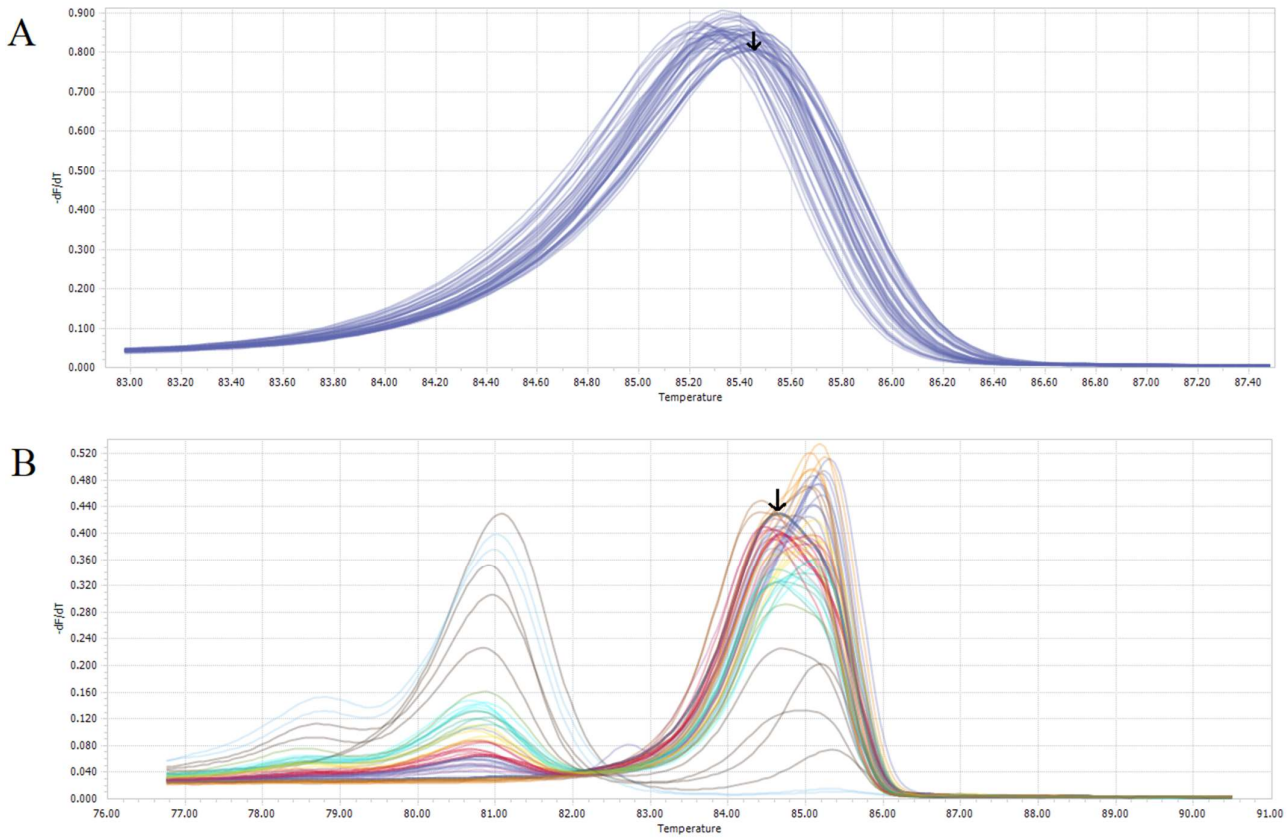

Figure 3.7. *High resolution melting curves from screening of double KO clones with target sequence A) Tp264671 f436 and B) Tp14322 f436. Melting peaks for WT are marked by the black arrows.*

3.2.2.3 Screening for mutations in triple knockout PAM clones

The target site for the triple KO PAM f233, which targets editing of *Tp14242* in addition to *Tp264671* and *Tp14322*, was also investigated for mutations by gel electrophoresis. As few colonies grew on the conjugation plate after the first conjugation, only 3 clones were picked, and the target region was amplified for each of the genes (Figure 3.8). Aside from a faint band for clone 3 amplified with primers specific for *Tp14322*, no indication that any mutations had occurred in the clones were found by the gel analysis. HRM analyses were not conducted on the clones containing the f233 target site. Clones pictured in Figure 3.8 were not sent in for Sanger sequencing to check.

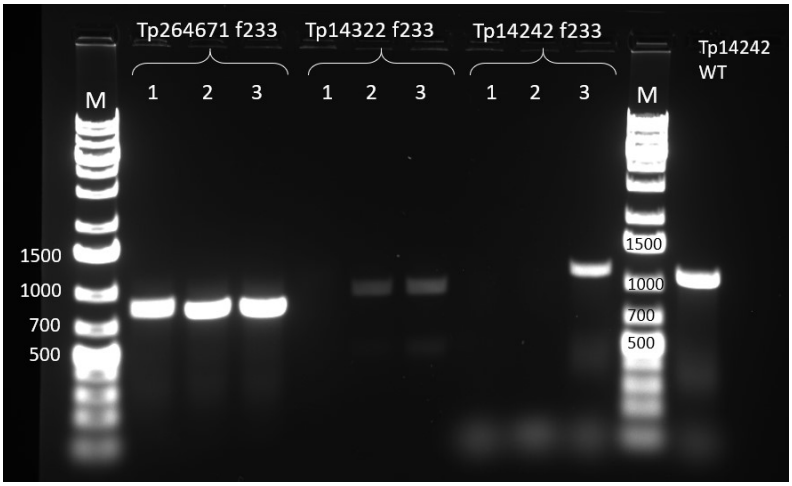

Figure 3.8. *Gel electrophoresis analysis with 1.0% agarose showing amplified regions containing PAM f233 for target site in Tp264671, Tp14322 and Tp14242 after the first conjugation. The samples (1, 2 and 3), which is amplified DNA from three different liquid cultures containing the f233 target site, were screened for edits in the three kinase-encoding genes listed at the top of the figure. GeneRuler 1 kb Plus DNA ladder (Thermo Scientific) was used as a marker to identify the sizes of the products (M). The numbers to the left (and in the middle) of the markers indicate the length (bp) of the fragments.*

Another method for screening was tested where colonies that grew on the f/2 plates (after the second conjugation) were picked straight off the plate without being cultured in liquid medium first. In this conjugation both 200 µl and 800 µl scraped cells (mix of *E. coli* and *T. pseudonana*) were plated out on each their plate and incubated. DNA from the colonies that grew on the 200 µl and 800 µl conjugation plates was amplified (only for *Tp264671*) and PCR products were analyzed by gel electrophoresis (Figure 3.9). The expected band size for the *Tp264671* F1 and R2 primers were 1406 bp; however, all the clones that grew on the 800 µl plate (Figure 3.9 B) seemed to have extra bands at ~700 bp and ~1000 bp, indicating that mutations might have happened on this plate and not on the 200 µl plate (where no bands apart from WT was present; Figure 3.9 A). This seems strange as the cells that were plated out on the two conjugation plates came from the same transformation mix. Deviations in the two plates could be explained by multiple things: deletions could indeed have occurred in clones from the 800 µl plate, there could be more DNA in the background as colonies on the 800 µl plate had been grown slightly longer (screened 7 days apart), or biproducts might have been amplified as a consequence of primers not being specific enough. Primers were run through a primer BLAST (NBCI) which showed that the primers were specific to only the target region. Sanger sequencing results were batch analyzed with ICE (Synthego), which confirmed that there were no mutations in the clones from either of the conjugation plates (200 µl and 800 µl).

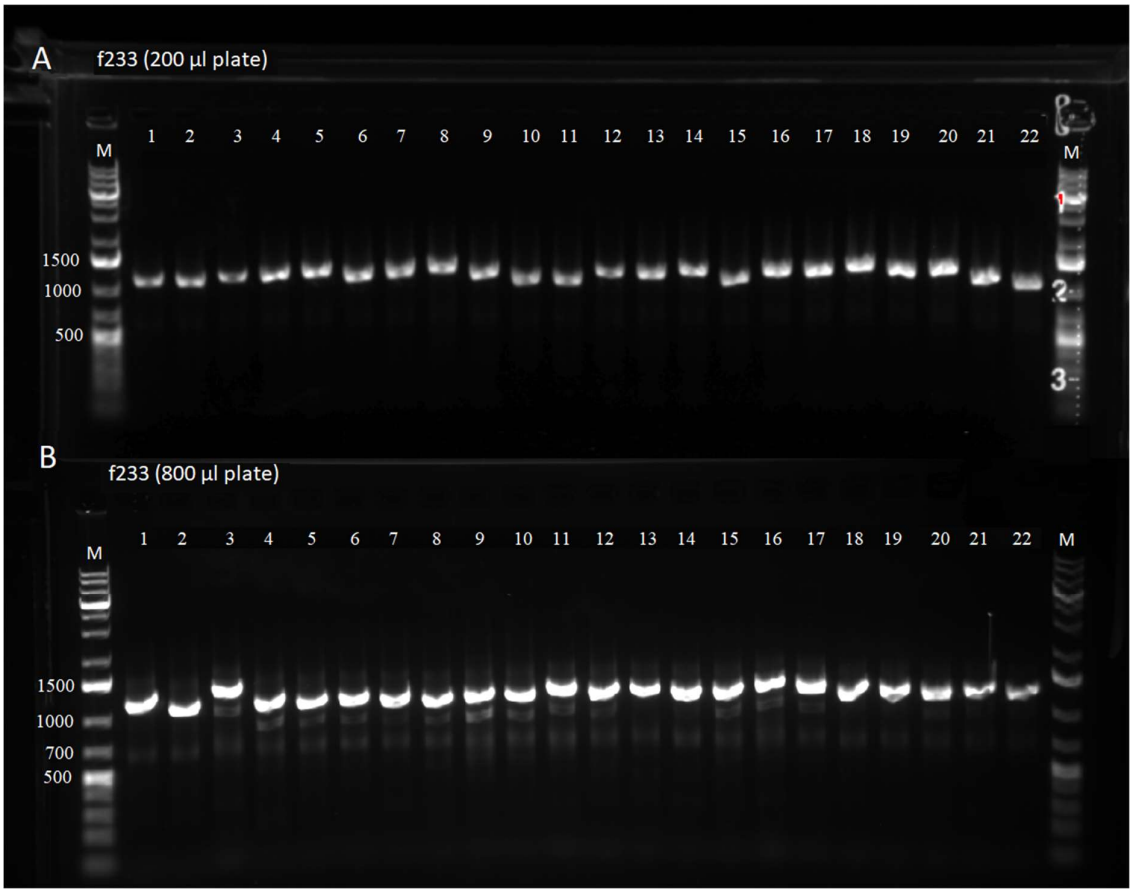

Figure 3.9. *Gel electrophoresis analysis with 1.0% agarose showing amplified regions by Tp264671 primers containing PAM f233 from colonies from the second conjugation. A) Clones picked from the 200 µl plate, and B) clones from the 800 µl plate. The numbers on the top of each subfigure signifies the picked clones for the specific target sites. GeneRuler 1 kb Plus DNA ladder (Thermo Scientific) was used as a marker to identify the sizes of the products (M). The numbers to the left of the markers indicate the length (bp) of the fragments.*

Further analyses and results of liquid culture clones from the second conjugation can be found in the following sections 3.2.4 and 3.2.5.

3.2.3 pTpPUC3-Cas9-M-G1 plasmid confirmation

Plasmids from different part of the mutant generation process were cut with the restriction enzyme BamHI-HF to ensure that they had the correct length and did not contain any deletions. This included checking the pTpPUC3-Cas9-M-G1 vector and plasmids ligated with the different oligos (Figure 3.10 A), as well as conjugation cultures containing plasmids with the target sites (Figure 3.10 B). As BamHI-HF has three restriction sites in the plasmids, three bands with sizes 2077 bp, 3454 bp and 9217 bp (empty vector would be about 26 bp shorter for the longest fragment) were expected to be visualized on the gel. This corresponded well with the bands from the gel electrophoresis as all the samples in the different lanes seemed to have the correct bands sizes. The results of the plasmid check on the gel indicated that the plasmid has been present in the *E. coli* cells during the conjugation procedure.

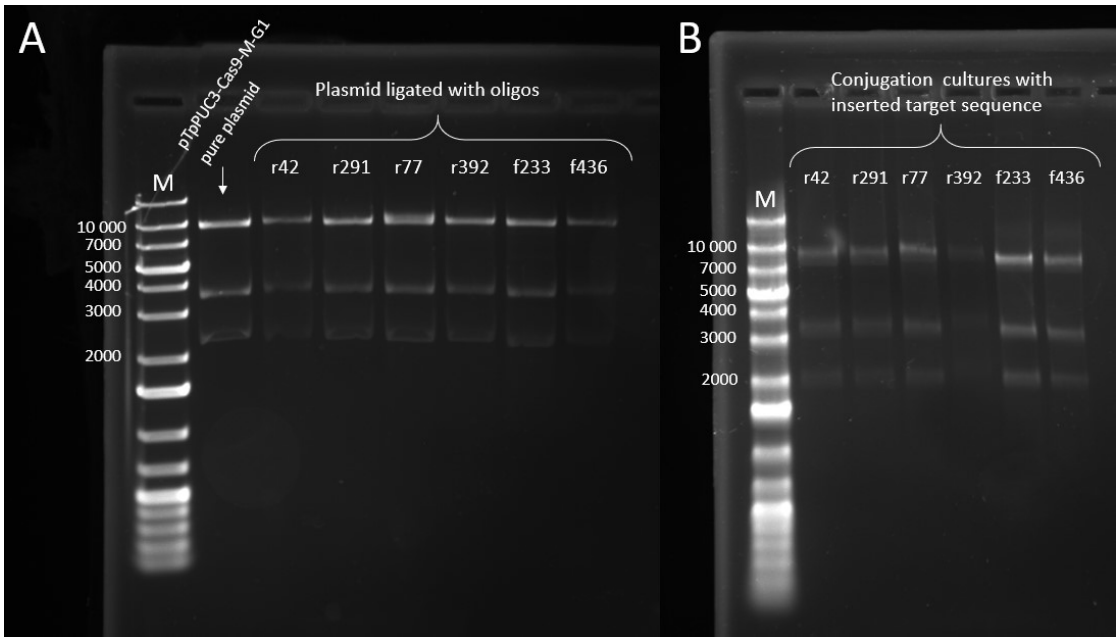

Figure 3.10. *Gel electrophoresis analysis (1.0% agarose) of plasmids from different parts of the mutant generation process cut with BamHI-HF. A) Empty pTpPUC3-Cas9-M-G1 plasmid and plasmids ligated with oligos for the different target sites, and B) conjugation cultures (from the second conjugation) confirmed to contain the specific target sites were checked. GeneRuler 1 kb Plus DNA ladder (Thermo Scientific) was used as a marker to identify the sizes of the products (M). The numbers to the left of the markers indicate the length (bp) of the fragments.*

3.2.4 Gene expression analysis with qRT-PCR

An analysis of gene expression was done on selected clones in order to measure the expression levels of different genes in the plasmid, including *Cas9*, *YFP*, *TpLHCF9* (for light harvesting complex), the nourseothricin resistance gene (*Nou*), and the sgRNA. As the nourseothricin resistance gene and Cas9 have the same LHCF9 terminator, comparing the expression between these two genes was done to see how high the gene-editing component of the plasmid was expressed.

Clone 2, 4, 15, 16, 18, 23 and 24 from the second conjugation, which were inserted with the target sequence for triple KO PAM f233, were analyzed. These specific clones were chosen based on their different levels of YFP when being measured on a flow cytometer. Some of them had low fluorescence (18 and 24), some high fluorescence (15, 16 and 23), and some with fluorescence values in between (2 and 4).

Results in Figure 3.11 shows the different Ct values of the different genes that were measured in the selected clones. The average Ct value that was calculated from the clones for each gene was calculated to be approximately 21.8 for *TpLHCF9*, 31.7 for *YFP*, 31.8 for the sgRNA, 33.5 for *Nou*, and 35.3 for *Cas9*. *TpLHCF9* had the lowest Ct value and thereby the highest expression, while *Cas9* was found to be the least expressed gene of the ones that were analyzed. Fewer clones were, however, measured for *Cas9*. Clone 4 was not included for analysis of *Cas9* expression as there were no space left on the qPCR plate, while the WT and clone 24 did not seem to be amplified during the analysis, and therefore did not give any signal or value. Certain clones seemed to have a higher transgene expression than others, such as clone 18 which had one of the lowest Ct values among the clones for the different analyzed genes. Clone 2 also stood out as it had higher expression of both *YFP* and the sgRNA than the other clones. It was concluded that a general low RNA expression of the genes meant that there was a low concentration of RNA to begin with.

Positive controls with 1:100, 1:1000 and 1:10 000 plasmid dilutions were made and analyzed along with the rest of the samples, although they did not give any signal and were therefore not included in Figure 3.11. No signal for the -RT control suggest that there were no gDNA contamination in the samples. The NTCs did not give any signal either, indicating that there were no factors such as contamination or primer-dimer formation that could give false positive results in the other analyzed samples. The low amount of RNA could also be a reason why the controls did not give any signal.

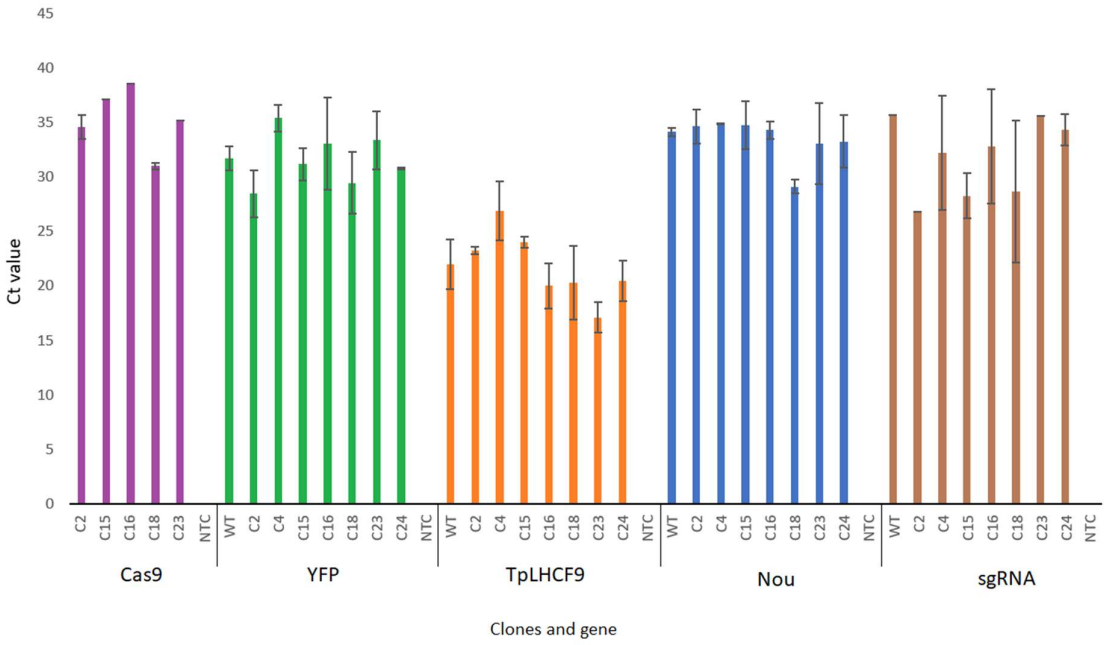

Figure 3.11. *Relative gene expression analysis showing the Ct value of different genes (Cas9, YFP, TpLHCF9, Nou and sgRNA) from different clones. The Ct values were based on the mean of two copies of the same samples, and presented with a ± SD.*

3.2.5 Microscopy of triple knockout PAM clones

Clones from the second conjugation were studied under the microscope in order to assess cell size and shape. Cells from clone 4, which was the clone that grew the most after inoculation in f/2 medium, were looked at ~4 weeks after the second conjugation. Figure 3.12 shows that the cells looked alive and good as chloroplast autofluorescence (indicating living cells) was apparent in all the cells.

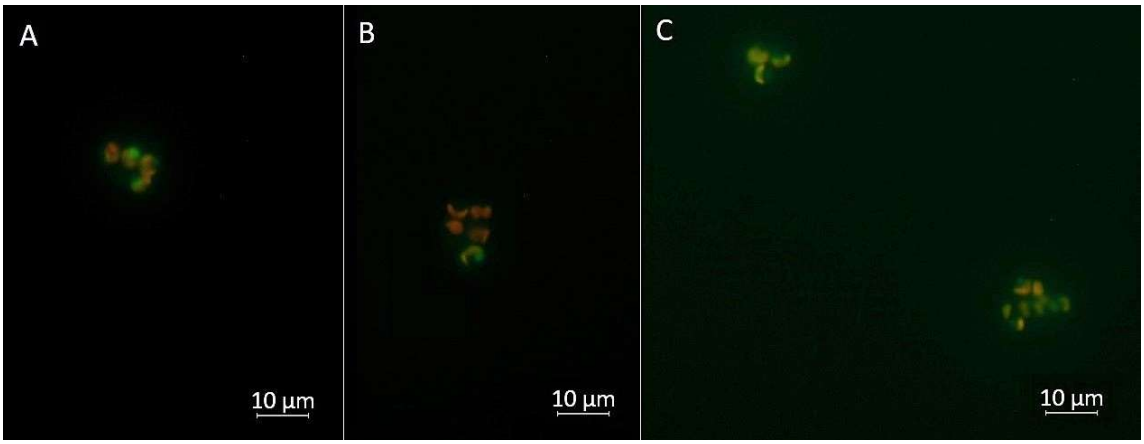

Figure 3.12. *Microscopy images of clone 4 (All A), B) and C)) straight after the second conjugation. Red fluorescence is caused by chlorophyll, the pigment found in the chloroplasts in T. pseudonana (indicating that the cell is alive), while the green fluorescence is caused by chloroplast as well as YFP (which is inserted into pTpPUC3-Cas9-M-G1). Images taken with a 40x objective.*

A selection of clones from the liquid cultures were assessed and imaged again at ~5 months after the second conjugation. Clone 2, 15 and 18, which expressed the highest amount of *YFP* from the qRT-PCR, were stained with DAPI in order to study colocalization of the fluorescent protein YFP with fluorescence from the nucleus (Figure 3.13).

The microscopy pictures in Figure 3.13 correspond nicely with the results from the qRT-PCR as there seem to be low amounts of RNA and RNA expression in the clones. Although *YFP* had the highest relative gene expression of the analyzed genes, colocalization was not visible in any of the cells that were found, except for in some of the dead ones (Figure 3.13 B, E, F, G). Some living cells (Figure 3.13 C, D) had YFP fluorescence outside the nucleus, something that could indicate that the YFP might have been transported from the nucleus to the lysosomes to be degraded. Cells had to be actively searched for under the microscope as the cultures did not seem to have a very high cell density. Another sign of low RNA expression could be explained by the fact that not all the clones that were found had fluorescence and signs of YFP in the nucleus.

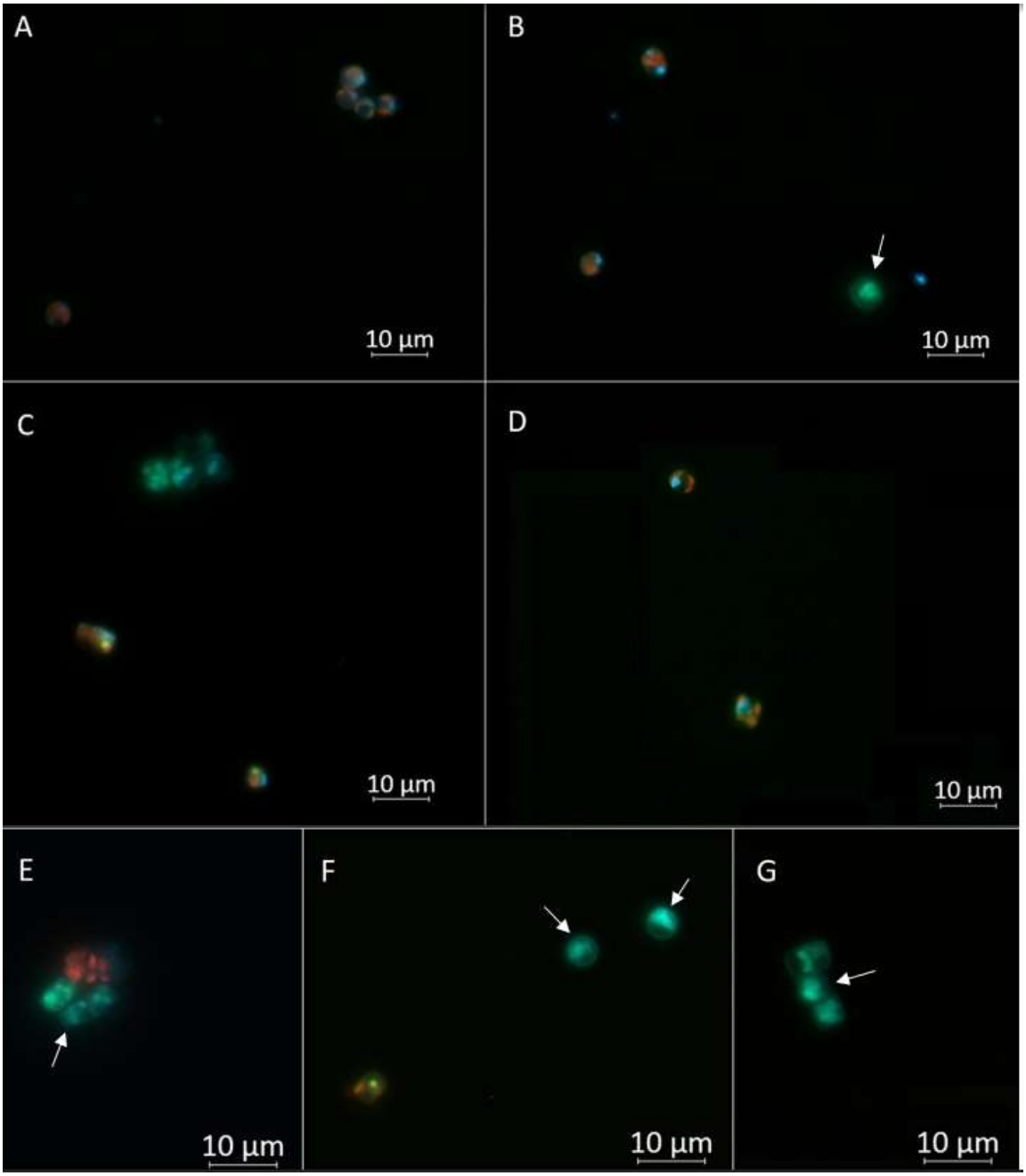

Figure 3.13. *Fluorescence microscopy images of clones from the second conjugation stained with DAPI. Pictured is A)-C) clone 2, E) clone 15, F) and G) clone 18. Red fluorescence is caused by chlorophyll, the pigment found in the chloroplasts in T. pseudonana (indicating that the cell is alive), while the green fluorescence is caused by chloroplast as well as YFP (which is inserted into pTpPUC3-Cas9-M-G1). The blue fluorescence indicates the location of the nucleus in the algae cells. Colocalization of YFP and DAPI is indicated by white arrows. Images taken with a 40x objective.*

3.3 Fusion of fluorescent protein mNeonGreen with *Tp14322* and *Tp264671*

The aim of the fluorescent protein fusion experiments was to tag the two genes of interest with mNeonGreen to act as a marker when studying the cells in the microscope. The genes along with mNeonGreen were inserted into a plasmid which were then introduced into *T. pseudonana* by bacterial conjugation.

3.3.1 Fragment amplification

Amplifying fragments that were going to be inserted into pTpPUC3 appeared to be challenging as not all the reactions for the specific fragments would give a product. Changes that were done in the reaction setup included trying different PCR kits for the amplification, as well as adding $MgCl_2$ and DMSO to the reaction mix. The PCR program was also modified, e.g. by lowering the annealing temperature, increasing the extension time, and setting up a temperature gradient for the annealing step. However, it appeared that new primers were necessary, and the remaining fragments were amplified by using a temperature gradient ranging from 57°C-67°C. All successfully amplified fragments that were used in the fluorescent protein fusion experiments are pictured in Figure 3.14. As the products that were amplified with lower temperatures tended to give multiple bands as seen in Figure 3.14 B, the fragments that were amplified at 65°C and 67°C were used for protein fusion first as they did not have biproducts.

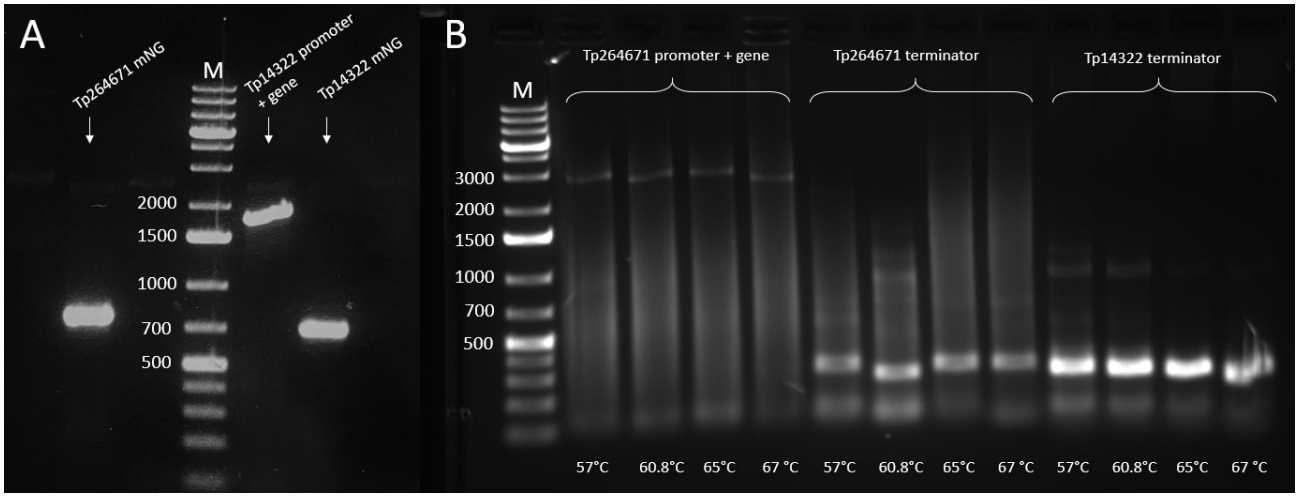

Figure 3.14. *Representative photo of fragments amplification from A) first attempt and B) with new primers and a temperature gradient from 57°C-67°C. GeneRuler 1 kb Plus DNA ladder (Thermo Scientific) was used as a marker to identify the sizes of the products (M). The numbers to the left of the markers indicate the length (bp) of the fragments.*

3.3.2 pTpPUC3 plasmid confirmation

The different plasmids (pTpPUC3 and pTpPUC3 FO4) that were used for cloning were run on a gel to make sure they had the correct length and to see if other unspecific bands would appear. Figure 3.15 shows the gel electrophoresis analysis of the plasmids, which were digested for different amounts of time with restriction enzymes SalI-HF (for pTpPUC3; Figure 3.15 A and B) and PstI (for pTpPUC3 FO4; Figure 3.15 C). All the bands on the gel seemed to correlate well to the expected band size of 7796 bp for the plasmids, something that indicates that the plasmids looked properly cut and were ready to be used for cloning. The pTpPUC plasmid that was cut for 15 min (Figure 3.15 A) appeared to have two very faint extra bands, one at ~6000 bp and one at ~18 000 bp. Another batch of the same plasmid was therefore cut again, this time for 1 h (Figure 3.15 B), to make sure it was completely cut and try to eliminate the possibility of amplifying biproducts.

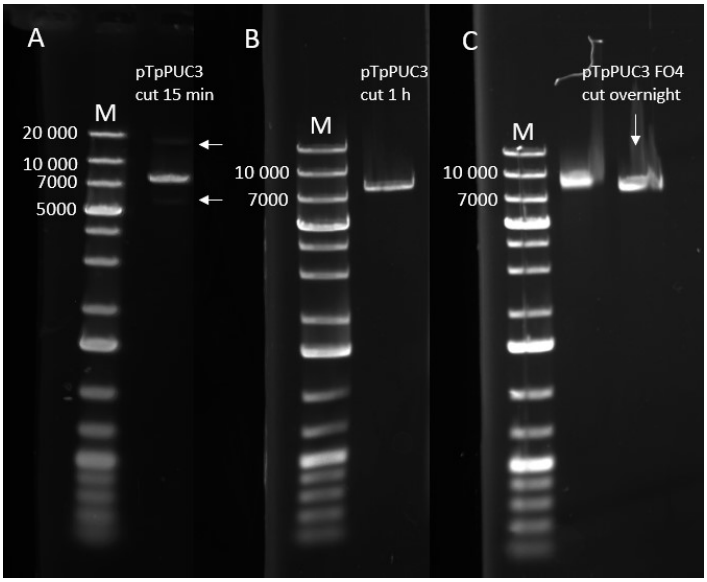

Figure 3.15. *Gel electrophoresis analysis (1.0% agarose) of different pTpPUC3 plasmids (cut for different periods of time) used for experiments: A) pTpPUC3 cut for 15 min (with two very faint biproducts indicated by white arrows), B) pTpPUC3 cut for 1 h, and C) pTpPUC3 FO4 cut overnight. GeneRuler 1 kb Plus DNA ladder (Thermo Scientific) was used as a marker to identify the sizes of the products (M). The numbers to the left of the markers indicate the length (bp) of the fragments.*

3.3.3 Cloning fragments into pTpPUC3

Several cloning methods were tested when attempting to clone the different fragments into the pTpPUC3 vector. This included Gibson assembly (ligation of fragments with overlapping ends), nested Gibson assembly (fewer fragments to ligate together), and CPEC (assembly by a single polymerase).
Multiple attempts to transform *E. coli* with Gibson products (plasmids with fragments for *Tp264671* and *Tp14322*) were unsuccessful, either due to no colonies growin or due to not getting a PCR product when screening the colonies. Changes such as purifying the products and transforming them into stable *E. coli* did not help either. However, after running the Gibson products (both before and after purification) on a gel (Figure 3.16), fragments with the right sizes seemed to be present (11 086 bp for *Tp264671* and 10 277 for *Tp14322*), indicating that the plasmids were likely to contain the inserts, although they might not have been successfully transformed into the *E. coli* cells. Insertion of the terminators could be hard to confirm on a gel as they are very small (~220 bp). Figure 3.16 also shows several biproducts for the different samples, although the concentration of these seems to be lower in the purified samples.

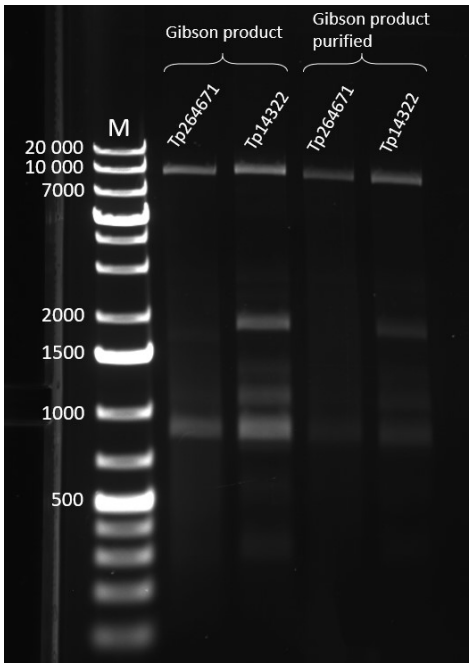

Figure 3.16. *Gel electrophoresis (1.0% agarose) analysis of Gibson products, both before and after purification. GeneRuler 1 kb Plus DNA ladder (Thermo Scientific) was used as a marker to identify the sizes of the products (M). The numbers to the left of the markers indicate the length (bp) of the fragments.*

As none of the cloning methods to assemble the amplified fragments ended up working, no further work was done with the fluorescent protein cloning part of the project.

4 Discussion

Two main objectives were set for this thesis: 1) creating mutants of *T. pseudonana* by knocking out the kinase-encoding genes *Tp264671*, *Tp14322* and *Tp14242*, and 2) fusing the fluorescent protein mNeonGreen with *Tp264671* and *Tp14322* to express it in *T. pseudonana* and analyze the expression and localization of the kinases by microscopy. An analysis to find the catalytic domain of the kinases and where the PAMs are located in the amino acid sequence was also performed. Results that came from the experiments are discussed below.

4.1 CaMK is the catalytic domain in *Tp264671* and *Tp14322*

Through a kinase domain analysis, it was found that the catalytic domain of both *Tp264671*and *Tp14322* was most similar to CaMK domains. The sequence of the catalytic domain for the two genes were aligned with CaMK domains of other genes in order to investigate conserved domains in the amino acid sequence. Results in NCBI showed that the kinases had the same conserved amino acids at the same sites in the sequence, while a more detailed analysis by comparing the alignments of the kinase-encoding genes in SnapGene with the CaMK2α sequence in Hanks and Hunter (1995) showed changes in certain conserved amino acids at specific places in the subdomains of the catalytic domain.

The differences in amino acids that were found between the three kinases are shown in table 3.1. Although two of the sites had different amino acids, the physical-chemical properties were still the same; the L and I in subdomain VI B where both are nonpolar, and E and D from subdomain VIII where both are acidic). One site in subdomain IX, however, had a change in the physical-chemical properties of the amino acids that differed, as one was I (nonpolar) and one was T (polar). Such differences can affect the efficiency of the kinase-binding to the substrate, in addition to altering the substrate specificity. The binding of a kinase to a substrate is highly specific, despite many kinase proteins exhibiting similar sequences and structures (Li et al., 2003). The reason for unique substrate recognition is due to specificity-determining residues in the kinase sequence.

4.2 Plasmid and mutation efficiency

Oligos containing target site for gene editing were inserted into the pTpPUC3-Cas9-M-G1 plasmid, and results showed that certain target sites initially seemed to work better and be more efficient than others, however, establishing mutants appeared to be a challenging task.

A general trend that was seen throughout the mutant screening process was that the clones that showed signs of possibly having mutations in the gel electrophoresis and HRM analysis, did not seem to contain mutations after all when analyzing their sequencing results. The fact that there were differences in the melting curves for several of the replicates made it harder to interpret the results. Errors in the preparations or pipetting could be the source of this issue. Even several clones that showed significant deviations from the WT in the HRM analysis, such as ones containing the target site for *Tp264671* r291 (Figure 3.5 B) and *Tp14322* f436 (Figure 3.7 B), did not show any differences when their sequence was aligned with the WT sequence. These same figures also show that a few of the clones display double-peaks, indicating that there might be a mix of the putative mutant and WT cells in the liquid cultures. A mix of the two could lead to the WT outcompeting the mutant cells, and no mutations would thereby be detected in that specific culture.

Another source of error as to why mutant creation was unsuccessful could be that the genes are essential for survival and knocking out the genes would be lethal for *T. pseudonana.* Du (2020) reports that a genome-wide knockout analysis in yeasts found that up to a quarter of the genes in a genome can be essential. Although an estimate of the number of essential genes in *T. pseudonana* has not yet been reported, it should not be disregarded that a similar situation could be the case in this diatom, where the kinase-encoding genes *Tp264671*, *Tp14322* and *Tp14242* might be essential. A phosphoproteomics analysis by mass spectrometry was planned to be conducted in order to identify substrates for the WT and mutants of *T. pseudonana*, but the analysis was not done as no mutants were created. Information about what substrates the kinases phosphorylate could have been valuable to determine what processes and pathways they might be a part of, and thereby perhaps get a better idea of whether they are essential or not.

The results also raise questions regarding the plasmid used. The pTpPUC3 vector (used in the fluorescent protein fusion part of the project) was modified in order to be suited for gene editing by the CRISPR/Cas9 technology. When analyzing pure plasmid and samples containing plasmids with the different target sites by gel electrophoresis (Figure 3.10), nothing appeared to be wrong with the plasmid as gel bands showed the correct length of the fragments. However, suspicion towards the plasmid efficiency was raised as further results and sequencing confirmed that no mutations had happened.

Other reasons as to why no mutants were obtained could also be that the structure of the plasmid or certain components might affect the efficiency of the plasmid. Local chromatin structure and chromosomal packaging of the DNA with the help of histones has been known to affect the ability for proteins to bind to the DNA, meaning the accessibility of the Cas9 to find the PAM site and bind to the DNA would decrease (Doench et al., 2014). Another reason could be that the sgRNA could be inefficient in making DSBs in the target sites of the DNA as a result of poor Cas9 or target site recognition. Cell fitness is in general negatively affected by Cas9 expression as it generates DSBs in the genome. Through structural analyses of sgRNA activity, it is discovered that the sequence of the sgRNA impacts the on-target activity of CRISPR/Cas9 gene editing. A GC content of 40-60% in the sgRNA sequence is known to stabilize DNA:RNA duplex and destabilize off-target binding, thereby increasing the on-target activity (Naeem et al., 2020). However, the GC content should ideally be between 50% and 75% (Kaulich and Dowdy, 2015). The site-specific sgRNA sequences that was inserted into the pTpPUC3-Cas9-M-G1 plasmid had a GC content ranging from 48-61%, something that suggests that the GC contents should have been sufficient and mostly within the ideal range for a stable target site.

By measuring sgRNA activity, a predictive model of 1841 sgRNAs was made by Doench et al. (2014) in order to improve sgRNA design for gene editing. It was observed that certain nucleotides were preferred in the 5'NGG PAM, where cytosine was favored and thymine was the least favored. Additionally, preferences were found in the nucleotide that is immediately adjacent to the PAM, where guanine was strongly preferred while cytosine was strongly unfavorable. This leads to the conclusion that GCGG formatted PAMs had the highest probability of successful gene editing, whereas the CTGG formatted PAMs had the least (Doench et al., 2014; Kaulich and Dowdy, 2015). These results could correlate with what was found for *Tp264671* r291 target site. Clones containing a target site with the r291 PAM, formatted as GCGG, seemed to have high activity of gene editing when analyzing them by gel electrophoresis (Figure 3.4 A and C) and HRM (Figure 3.5 B). These clones seemed to deviate significantly from the WT on HRM analysis as the T_m for some of the putative mutants were found to be ~7-8°C lower than the T_m for the WT. The results from screening the clones with the r291 target site might indicate that this is a very effective site for editing, and the reason for unsuccessful mutant establishment might lay elsewhere.

An alteration of the pTpPUC3-Cas9-M-G1 plasmid to increase the efficiency when cloning was attempted by mutating a region in the beginning of the sgRNA from TTTT to TTTA. About 20 bp downstream from the TTTT→TTTA region, a region on the complementary strand had to be mutated from TTTT to ATTT in order for the sgRNA loop to form. A homopolymer stretch of TTTT should be avoided in a sgRNA sequence as it is known as a transcription termination signal for RNA polymerase III (Sekine et al., 2018). Similarly, in the U6 promotor, a stretch of T bases signals the polymerase III to stop transcription. A stretch of TTTT act as a minimal terminator with 75% termination efficiency,

while a stretch ≥ 6 signals complete termination (Gao et al., 2018). Three stretches containing ≥ 4 T bases are found in the U6 promoter in the pTpPUC3 Cas9 M G1 plasmid, even including one stretch of 9 T bases. Multiple regions with these T stretches might decrease the expression of the sgRNA and thereby affect the plasmid efficiency. The U6 promoter along with the sgRNA sequence was the same as used in Hopes et al. (2016), a study where these sequences appeared to work. This modified plasmid was not used in the project as it was not confirmed to work.

Karas et al. (2015) was the first to demonstrated successful conjugation of the pTpPUC3 vector into *T. pseudonana* by bacterial conjugation. They experienced that the yeast CEN6-ARSH4-HIS3 (CAH) region in the pTpPUC3 plasmid resulted in an increase in conjugation efficiency when transferring the plasmid from *E. coli* to *T. pseudonana,* as compared to the pTpPUC4 control, which lack this region. The CAH region (~1.4 kb) is comprised of CEN6 and ARSH4, which encode plasmid maintenance functions in the yeast centromeric plasmids, as well as the HIS3 gene, which act as a histidine marker and complements yeast histidine auxotrophy (Karas et al., 2015). This region, conferring the low-copy episomal replication ability in *T. pseudonana*, is the reason the episome can be successfully rescued consistently. A similar study conducted by Diner et al. (2016), albeit on the pennate diatom *Phaeodactylum tricornutum*, found that contiguous fragments of the CAH greater than 500 bp and with low GC were required to support sturdy episome maintenance. In addition to the CAH region, the inclusion of an origin of transfer (OriT), which allows the plasmids to be mobilized to the recipient by conjugation, as well as nourseothricin as an antibiotic selectable marker, is necessary for an optimal conjugation of the pTpPUC3 plasmid (Karas et al., 2015). Instead of bacterial conjugation, biolistic CRISPR/Cas9 transformation could be done to possibly increase gene editing efficiency. In this method, the foreign DNA will integrate into the genome of the recipient cell (Sharma et al., 2018).

4.3 Low expression of Cas9 and the nourseothricin resistance gene

An expression analysis by qRT-PCR was done on some clones from the second conjugation that had various levels of YFP when measured on a flow cytometer. The analysis found that certain genes had a higher expression than others when comparing the average Ct values of the clones measured for a specific gene. Results showed that the *LHCF9* had the highest expression, followed by *YFP*, the sgRNA, *Nou*, and lastly *Cas9* (Figure 3.11). The *LHCF9* gene (which is the only one that belongs to the nuclear genome, and not the episome) appeared to be most expressed, indicating that the quality of the RNA extraction and cDNA production was sufficient to reflect *LHCF9* expression. The level of *LHCF9* expression, which is affected by light exposure, could also be explained by the fact that they were cultivated under constant light, meaning that the expression was kept relatively high and steady.

Comparing the expression of the other measured genes to the *LHCF9* might therefore not give the most precise results as a normalization of the genes was not done.

The reason behind the genes having a lower expression might, as mentioned earlier, be because of flaws in the plasmid, something that could lower its overall efficiency. This could affect whether the sgRNA is sufficiently expressed. Results showed that the sgRNA was not very highly expressed, meaning it might not have been that efficient in the guiding the Cas9 to the intended target sites. The low expression of the sgRNA and *Cas9* implies that there was probably never enough of these to create knockouts.

A low expression of *Nou* might indicate that there is a lower efficiency in the selection of the transformants with the plasmid. An enzyme kinetics analysis of Nat (N-acetyl transferase), the enzyme for Nou resistance, could be investigated to see how fast the enzyme uses up concentrations of Nou in the medium. This information could help in determining what amount of Nou should be added to the medium for the most efficient growth of *T. pseudonana*, and thereby make sure there is enough antibiotic pressure for the diatom to keep the episome. Another reason for low expression could be that *T. pseudonana* have transporters that pump Nou out of the cells. ABC (ATP binding cassette) transporters, which are membrane proteins that are responsible for ATP-powered translocation of molecules, are able to function as importers by bringing nutrients into the cell, or as exporters to pump out potentially harmful chemicals and toxins as well as other substances such as drugs and lipids (Rees et al., 2009).

Microscopy photos of clones (with different levels of YFP) from the second conjugation showed similar results to what was seen in the qRT-PCR analysis, where there was a low expression of RNA. DAPI was used to stain the nucleus and overlap of the blue fluorescence channel with the green fluorescence channel made it possible to observe potential colocalization of the nucleus with YFP. The pTpPUC3-Cas9-M-G1 plasmid contains a motif comprised of only seven amino acids (PKKKRKV), the SV40 NLS (nuclear localization signal), which leads it to the nucleus where it can cut the genomic DNA of *T. pseudonana* and produce mutations. Seeing colocalization of the nucleus with YFP in living cells is desirable as it could have indicated that Cas9 was present in the nucleus, which is a prerequisite for mutant generation. Microscopy images did, however, show that colocalization was only seen in dead cells (Figure 3.13 B, E, F and G), and not in any living ones. Fluorescence outside the nucleus could indicate that the protein had been transported from the nucleus to the lysosomes to be degraded, which could mean that the Cas9 might have been in the nucleus at one point. This might explain how colocalization was visible exclusively in the dead cells; they did not have the ability to transport Cas9 out of the nucleus as the cellular machinery was down. The cells had to be actively searched for under the microscope, and not all of the cells that were found had fluorescence or YFP in the nucleus.

4.4 Fluorescent protein cloning

In the fluorescent protein fusion part of the project, I attempted to clone gene fragments along with mNeonGreen into the pTpPUC3 vector. Several different methods and protocols described in section 2.3 were tried, although none of them were successful. One of the problems was due to the fact that we have not been able to eliminate clones without insert (false positives). As discussed in section 4.2, inefficiency in the cloning might come from the plasmid that was used. Other things that could have caused a low cloning efficiency includes doing a gel extraction of PCR product and thereby losing a lot of product, and not always measuring the concentration of the PCR product to calculate the best insert:vector ratio volumes, which might also have resulted in a higher concentration of the vector, something that is not beneficial.

The gel electrophoresis analysis of the Gibson products (Figure 3.16) suggested that all of the fragments had been successfully inserted into the pTpPUC3 plasmid. However, colony screening after transformation into both DH5α *E. coli* and Stable competent *E. coli* showed that the plasmid was not adopted by the few clones that did grow on the LB plates.

5 Future aspects

In future work, methods might be changed or modified in order for the cloning procedure to be more efficient. This includes plating out *T. pseudonana* cells on plates with Nou of different concentrations and observing how many survives. The sgRNA will be optimized (or perhaps using the TTTT → TTTA mutated plasmid), and clones that look good under the microscope and show YFP will be isolated. Incorporation of the plasmid with the Cas9 could be attempted by biolistic transformation instead of conjugation to possibly increase efficiency. Another idea to make the conjugation work could be by cell sorting, where fluorescence will be measured and cells that have a high level of YFP fluorescence will be collected and cultivated to be screened again later.

6 Conclusion

The aim of this study was to create single-, double-, and triple knockout mutant cell lines of *T. pseudonana* by utilizing the CRISPR/Cas9 technology for gene editing in the closely related (and very similar) kinase-encoding genes *Tp264671*, *Tp14322* and *Tp14242*. Different screening analyses such as gel electrophoresis and HRM showed strong indications that mutations had occured in several clones for the different target sites, although analyses by Sanger sequencing confirmed that no mutations had happened. Results did, however, indicate that some target sites, especially the single KO PAM *Tp264671* r291, seemed more efficient than others as clones from this transformant deviated significantly from the WT in the mutant screening analyses (except for in sequencing). Mixed results from the screening process suggested that issues with establishing mutant strains might be due to the genes being essential for the cell to survive and knocking them out will simply cause the cell to die, or the problem might be due to efficiency issues with the pTpPUC3-Cas9-M-G1 plasmid that was used. Plasmid efficiency relies on many things, such as the genes and components it is comprised of as well as sequence characteristics and composition. No faults were, however, found in the plasmid when checking it in different stages of the mutant creation process. A gene expression analysis that was conducted on a set of clones found that there was a very low expression of certain genes (*Cas9*, *YFP*, *Nou* and the sgRNA) in the plasmid, while there was a slightly higher expression of *LHCF9*. Low expression of *Cas9* and the sgRNA suggested (in correlation to studies of the cells by microscopy) that the concentration of mRNA from the plasmid was probably too low to create knockouts, while low expression of *Nou* could indicate that there was not enough selection pressure when culturing the *T. pseudonana* clones.

In addition to creating mutants, I attempted to fuse the fluorescent protein mNeonGreen with *Tp264671* and *Tp14322*. Different methods were tried in order to fuse the fragments together, although none of them were successful. Expression and localization studies as a tool to uncover the function of these genes were therefore not done.

A protein alignment analysis of the three genes identified CaMK as the catalytic domain of the kinases. Comparing the kinase alignments with alignments for the genes found on NCBI as well as with a CaMK sequence from a previous study suggested that certain regions of amino acids in the different subdomains of the catalytic domain were well conserved as they were found in all the alignments.

References

Armbrust, E. V. (2009). The life of diatoms in the world's oceans. *Nature, 459*(7244), 185-192.

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., . . . Bechner, M. (2004). The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. *science, 306*(5693), 79-86.

Bayer, K. U., & Schulman, H. (2019). CaM kinase: still inspiring at 40. *Neuron, 103*(3), 380-394.

Becker, B., Hoef-Emden, K., & Melkonian, M. (2008). Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. *BMC evolutionary biology, 8*(1), 1-18.

Bertani, G. (1951). Studies on lysogenesis I.: the mode of phage liberation by lysogenic Escherichia coli1. *Journal of bacteriology, 62*(3), 293.

Bowler, C., Karl, D. M., & Colwell, R. R. (2009). Microbial oceanography in a sea of opportunity. *Nature, 459*(7244), 180-184.

Bowler, C., Vardi, A., & Allen, A. E. (2010). Oceanographic and biogeochemical insights from diatom genomes. *Annual review of marine science, 2*, 333-365.

Brembu, T., Chauton, M. S., Winge, P., Bones, A. M., & Vadstein, O. (2017). Dynamic responses to silicon in Thalasiossira pseudonana-Identification, characterisation and classification of signature genes and their corresponding protein motifs. *Scientific reports, 7*(1), 1-14.

Crisafuli, F., Ramos, E., & Rocha, M. (2015). Characterizing the interaction between DNA and GelRed fluorescent stain. *European Biophysics Journal, 44*(1-2), 1-7.

De Tommasi, E., Gielis, J., & Rogato, A. (2017). Diatom frustule morphogenesis and function: a multidisciplinary survey. *Marine genomics, 35*, 1-18.

Diner, R. E., Bielinski, V. A., Dupont, C. L., Allen, A. E., & Weyman, P. D. (2016). Refinement of the diatom episome maintenance sequence and improvement of conjugation-based DNA delivery methods. *Frontiers in bioengineering and biotechnology, 4*, 65.

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., . . . Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nature biotechnology, 32*(12), 1262-1267.

Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *science, 346*(6213).

Du, L.-L. (2020). Resurrection from lethal knockouts: Bypass of gene essentiality. *Biochemical and Biophysical Research Communications, 528*(3), 405-412.

Dunahay, T. G., Jarvis, E. E., Dais, S. S., & Roessler, P. G. (1996). Manipulation of microalgal lipid production using genetic engineering. *Applied biochemistry and biotechnology, 57*(1), 223-231.

Falciatore, A., & Bowler, C. (2002). Revealing the molecular secrets of marine diatoms. *Annual review of plant biology, 53*(1), 109-130.

Fattorini, N., & Maier, U. G. (2021). Targeting of proteins to the cell wall of the diatom Thalassiosira pseudonana. *Discover Materials, 1*(1), 1-10.

Garza, E., & Bielinski, V. A. (2020). *Nested Gibson Assembly.* Protocols.io

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods, 6*(5), 343-345.

Guillard, R. R. (1975). Culture of phytoplankton for feeding marine invertebrates. In *Culture of marine invertebrate animals* (pp. 29-60): Springer.

Hanks, S. K., & Hunter, T. (1995). The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification 1. *The FASEB journal, 9*(8), 576-596.

Hasegawa, H., Suzuki, E., & Maeda, S. (2018). Horizontal plasmid transfer by transformation in Escherichia coli: environmental factors and possible mechanisms. *Frontiers in microbiology, 9*, 2365.

Herrmann, M. G., Durtschi, J. D., Bromley, L. K., Wittwer, C. T., & Voelkerding, K. V. (2006). Amplicon DNA melting analysis for mutation scanning and genotyping: cross-platform comparison of instruments and dyes. *Clinical chemistry, 52*(3), 494-503.
Hildebrand, M., Higgins, D. R., Busser, K., & Volcani, B. E. (1993). Silicon-responsive cDNA clones isolated from the marine diatom Cylindrotheca fusiformis. *Gene, 132*(2), 213-218.
Hildebrand, M., Volcani, B. E., Gassmann, W., & Schroeder, J. I. (1997). A gene family of silicon transporters. *Nature, 385*(6618), 688-689.
Hildebrand, M., York, E., Kelz, J. I., Davis, A. K., Frigeri, L. G., Allison, D. P., & Doktycz, M. J. (2006). Nanoscale control of silica morphology and three-dimensional structure during diatom cell wall formation. *Journal of Materials Research, 21*(10), 2689-2698.
Hopes, A., Nekrasov, V., Kamoun, S., & Mock, T. (2016). Editing of the urease gene by CRISPR-Cas in the diatom Thalassiosira pseudonana. *Plant Methods, 12*(1), 1-12.
Hsiau, T., Maures, T., Waite, K., Yang, J., Kelso, R., Holden, K., & Stoner, R. (2018). Inference of CRISPR edits from Sanger trace data. *BioRxiv*, 251082.
Issaq, H. J., Chan, K. C., & Muschik, G. M. (1997). The effect of column length, applied voltage, gel type, and concentration on the capillary electrophoresis separation of DNA fragments and polymerase chain reaction products. *Electrophoresis, 18*(7), 1153-1158.
Jiang, F., & Doudna, J. A. (2017). CRISPR–Cas9 structures and mechanisms. *Annual review of biophysics, 46*, 505-529.
Jozefczuk, J., & Adjaye, J. (2011). Quantitative real-time PCR-based analysis of gene expression. *Methods in enzymology, 500*, 99-109.
Jura, N., Zhang, X., Endres, N. F., Seeliger, M. A., Schindler, T., & Kuriyan, J. (2011). Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. *Molecular cell, 42*(1), 9-22.
Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P., Noddings, C. M., . . . Jablanovic, J. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nature communications, 6*(1), 1-10.
Kaulich, M., & Dowdy, S. F. (2015). Combining CRISPR/Cas9 and rAAV templates for efficient gene editing. *nucleic acid therapeutics, 25*(6), 287-296.
Kroth, P. (2007). Molecular biology and the biotechnological potential of diatoms. *Transgenic microalgae as green cell factories*, 23-33.
Kröger, N., & Poulsen, N. (2008). Diatoms—from cell wall biogenesis to nanotechnology. *Annual review of genetics, 42*, 83-107.
Kuang, J., Yan, X., Genders, A. J., Granata, C., & Bishop, D. J. (2018). An overview of technical considerations when using quantitative real-time PCR analysis of gene expression in human exercise research. *PloS one, 13*(5), e0196438.
Kumar, S., Rechav, K., Kaplan-Ashiri, I., & Gal, A. (2020). Imaging and quantifying homeostatic levels of intracellular silicon in diatoms. *Science advances, 6*(42), eaaz7554.
Li, L., Shakhnovich, E. I., & Mirny, L. A. (2003). Amino acids determining enzyme-substrate specificity in prokaryotic and eukaryotic protein kinases. *Proceedings of the National Academy of Sciences, 100*(8), 4463-4468.
Li, T., Wang, N., Zhang, T., Zhang, B., Sajeevan, T. P., Joseph, V., . . . Naman, C. B. (2019). A systematic review of recently reported marine derived natural product kinase inhibitors. *Marine drugs, 17*(9), 493.
Llosa, M., Gomis-Rüth, F. X., Coll, M., & Cruz, F. d. l. (2002). Bacterial conjugation: a two-step mechanism for DNA transport. *Molecular microbiology, 45*(1), 1-8.
Maier, L. S., & Bers, D. M. (2007). Role of Ca2+/calmodulin-dependent protein kinase (CaMK) in excitation–contraction coupling in the heart. *Cardiovascular research, 73*(4), 631-640.
Mann, D. G., & Vanormelingen, P. (2013). An inordinate fondness? The number, distributions, and origins of diatom species. *Journal of eukaryotic microbiology, 60*(4), 414-420.
Marron, A. O., Ratcliffe, S., Wheeler, G. L., Goldstein, R. E., King, N., Not, F., . . . Richter, D. J. (2016). The evolution of silicon transport in eukaryotes. *Molecular biology and evolution, 33*(12), 3226-3248.
Mock, T., Samanta, M. P., Iverson, V., Berthiaume, C., Robison, M., Holtermann, K., . . . Rodesch, M. (2008). Whole-genome expression profiling of the marine diatom Thalassiosira pseudonana

identifies genes involved in silicon bioprocesses. *Proceedings of the National Academy of Sciences, 105*(5), 1579-1584.
Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K., & Bhattacharya, D. (2009). Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *science, 324*(5935), 1724-1726.
Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A., & Quéguiner, B. (1995). Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles, 9*(3), 359-372.
Nuester, J., Vogt, S., & Twining, B. S. (2012). LOCALIZATION OF IRON WITHIN CENTRIC DIATOMS OF THE GENUS THALASSIOSIRA 1. *Journal of phycology, 48*(3), 626-634.
Prihoda, J., Tanaka, A., de Paula, W. B., Allen, J. F., Tirichine, L., & Bowler, C. (2012). Chloroplast-mitochondria cross-talk in diatoms. *Journal of experimental botany, 63*(4), 1543-1557.
Quan, J., & Tian, J. (2014). Circular polymerase extension cloning. In *DNA cloning and assembly methods* (pp. 103-117): Springer.
Rees, D. C., Johnson, E., & Lewinson, O. (2009). ABC transporters: the power to change. *Nature reviews Molecular cell biology, 10*(3), 218-227.
Reyes-Prieto, A., Hackett, J. D., Soares, M. B., Bonaldo, M. F., & Bhattacharya, D. (2006). Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Current Biology, 16*(23), 2320-2325.
Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology, 32*(4), 347-355.
Schröder, H. C., Wang, X., Tremel, W., Ushijima, H., & Müller, W. E. (2008). Biofabrication of biosilica-glass by living organisms. *Natural product reports, 25*(3), 455-474.
Sekine, R., Kawata, T., & Muramoto, T. (2018). CRISPR/Cas9 mediated targeting of multiple genes in Dictyostelium. *Scientific reports, 8*(1), 1-11.
Sharma, A. K., Nymark, M., Sparstad, T., Bones, A. M., & Winge, P. (2018). Transgene-free genome editing in marine algae by bacterial conjugation–comparison with biolistic CRISPR/Cas9 transformation. *Scientific reports, 8*(1), 1-11.
Shrestha, R. P., & Hildebrand, M. (2015). Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses. *Eukaryotic cell, 14*(1), 29-40.
Simkiss, K., & Wilbur, K. M. (2012). *Biomineralization*: Elsevier.
Thamatrakoln, K., Alverson, A. J., & Hildebrand, M. (2006). COMPARATIVE SEQUENCE ANALYSIS OF DIATOM SILICON TRANSPORTERS: TOWARD A MECHANISTIC MODEL OF SILICON TRANSPORT 1. *Journal of phycology, 42*(4), 822-834.
Twist, G. P., Gaedigk, R., Leeder, J. S., & Gaedigk, A. (2013). High-resolution melt analysis to detect sequence variations in highly homologous gene regions: application to CYP2B6. *Pharmacogenomics, 14*(8), 913-922.
Virolle, C., Goldlust, K., Djermoun, S., Bigot, S., & Lesterlin, C. (2020). Plasmid Transfer by Conjugation in Gram-Negative Bacteria: From the Cellular to the Community Level. *Genes, 11*(11), 1239.
Vossen, R. H., Aten, E., Roos, A., & den Dunnen, J. T. (2009). High-Resolution Melting Analysis (HRMA)—More than just sequence variant screening. *Human mutation, 30*(6), 860-866.
Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G., & Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. *Molecular biology and evolution, 21*(5), 809-818.
Zhang, H., Cheng, Q.-X., Liu, A.-M., Zhao, G.-P., & Wang, J. (2017). A novel and efficient method for bacteria genome editing employing both CRISPR/Cas9 and an antibiotic resistance cassette. *Frontiers in microbiology, 8*, 812.
Wang, J.-W., Wang, A., Li, K., Wang, B., Jin, S., Reiser, M., & Lockey, R. F. (2015). CRISPR/Cas9 nuclease cleavage combined with Gibson assembly for seamless cloning. *Biotechniques, 58*(4), 161-170.

Appendices

Appendix 1: Media and solutions

Table A1-1. Recipe for Luria-Bertani (LB) medium/agar plates. Components were added to 1 L dH_2O and the solution was autoclaved for 20 min at 120°C. LB solution containing agar was cooled down to <55°C before adding antibiotics and pouring it onto agar plates. The plates were left to solidify under a sterile hood before being stored at 4°C.

Compound	Amount (g/L dH_2O)
Trypton	10
Yeast extract	5
NaCl	5
Bactoagar (only for agar plates)	15

Table A1-2. Recipe for SOC medium. Components were added to 1 L dH_2O and the solution was autoclaved for 20 min at 120°C. The solution was then stored at 4°C.

Compound	Amount
Trypton	20 g/L
Glucose	3.6 g/L
$MnCl_2$ x H_2O	5.08 g/L
KCl	2.5 mM

Table A1-3. Recipe for f/2 + Si medium/agar plates (* with stocks from Table A1-4).

Component	Per liter
$NaNO_3$	0.075 g
NaH_2PO_4 x H_2O	0.00565 g
Trace elements *	1.0 ml
Vitamin mix *	1.0 ml
Sodium metasilicate *	1.0 ml
Agar (only for agar plates)	15 g/L

Table A1-4. Recipe for trace element, vitamin mix and sodium metasilicate solutions. Components were dissolved in 1 L dH_2O before being sterile filtered and stored at 4°C.

Stocks	Per liter (g)
Trace elements	
Na2EDTA	4.36
$FeCl_2$ x $6H_2O$	3.15
$CuSO_4$ x $5H_2O$	0.01
$ZnSO_4$ x $7H_2O$	0.022
$CoCl_2$ x $6H_2O$	0.01
$MnCl_2$ x $4H_2O$	0.18
Na_2MO_4 x $2H_2O$	0.006
Vitamin mix	
Cyanocobalamin (Vitamin B12)	0.0005
Thiamine HCl (Vitamin B1)	0.1
Biotin	0.0005
Sodium metasilicate	
Na_2SiO_2 x $9H_2O$	30.0

Table A1-5. Recipe for *T. pseudonana* lysis buffer.

Compound	Concentration
Triton x-100	10 %
Tris-HCl, pH 8	20 mM
EDTA	10 mM

Table A1-6. Recipe for 1x TAE buffer used when running agarose gels.

Compound	Amount per liter
Tris (base)	4.84 g
Glacial acetic acid	1.442 ml
EDTA	0.0012 M

Appendix 2: Primers used for vector modification

Table A2-1. *Complementary oligos for constructing PAM target sites.*

Primer name	Orientation	Sequence (5'-3')
Tp14322-r77PF	Forward	ATTGTAGTACAGCACAGATCACACGA
Tp14322-r77PR	Reverse	AAACTCGTGTGATCTGTGCTGTACTA
Tp14322-r392PF	Forward	ATTGTAGTGACAAGGTGGACGTACTA
Tp14322-r392PR	Reverse	AAACTAGTACGTCCACCTTGTCACTA
Tp264671-42PF	Forward	ATTGTCATCGTAGGGAGCTGATACGA
Tp264671-42PR	Reverse	AAACTCGTATCAGCTCCCTACGATGA
Tp264671-291PF	Forward	ATTGTTGACCGATGGCTCAGACTTGA
Tp264671-291PR	Reverse	AAACTCAAGTCTGAGCCATCGGTCAA
Tp14322-264671-14322-f233PF	Forward	ATTGTGTATTGGGAGTTGGTCATCAA
Tp14322-264671-14322-f233PR	Reverse	AAACTTGATGACCAACTCCCAATACA
Tp14322-264671-f436PF	Forward	ATTGTCCTTGTCACTAAACTGTGCGA
Tp14322-264671-f436PR	Reverse	AAACTCGCACAGTTTAGTGACAAGGA
Tp14322-264671-f754PF	Forward	ATTGTTTCAGAGAAGCAGCCGTGCTA
Tp14322-264671-f754PR	Reverse	AAACTAGCACGGCTGCTTCTCTGAAA

Table A2-2. *Primers for gene amplification of 800-1000 amplicons.*

Primer name	Orientation	Sequence (5'-3')	Amplicon size (bp)
Tp14322F1	Forward	gattcctaaccaaagatccacg	998
Tp14322R1	Reverse	gatccgtttgcagtagtttacg	
Tp264671F1	Forward	cattatgcacgacagccaatca	868
Tp264671R1	Reverse	gctctttcgtagaacatcaggt	
Tp14322F2	Forward	tgaagcacgacagcatcattca	932
Tp14322R2	Reverse	acttgaagctcacatcaaccga	
Tp264671F2	Forward	gctcgtgagattgcattgcttg	955
Tp264671R2	Reverse	gtgtcattggcctcagcttcaa	
Tp14242F1	Forward	tcatcatcatccgacgcaatga	843
Tp14242R1	Reverse	atagtgctttcgtgccaatcca	

Table A2-3. *Primers for HRM analysis.*

Primer name	Orientation	Sequence (5'-3')	Amplicon size (bp)
Tp14322 triple f233/r77 hrm F	Forward	gacgacgataccattccatcacg	224
Tp14322 triple f233/r77 hrm R	Reverse	tgctctccagtgtatcggttgat	
Tp14322 r392 f436 hrm F	Forward	tcgtgaagtagctcttcttgttg	179
Tp14322 r392 f436 hrm R	Reverse	cttgatgttcagagaagcagccg	
Tp264671 r42 hrm F	Forward	cgcatcaccaattgaaatgacaaaagttg	120
Tp264671 r42 hrm R	Reverse	cttgcacagggtgatacaggtgtag	
Tp264671 r291 hrm F	Forward	catggatctgtacgacaatgtgttg	127
Tp264671 r291 hrm R	Reverse	tctcatcaagcaatgcaatctcacg	
Tp264671 f436 hrm F	Forward	gctcgtgagattgcattgcttgatg	199
Tp264671 f436 hrm R	Reverse	gatgcaatatcttcgcagcttcatg	

Table A2-4. *Primers for qRT-PCR analysis.*

Primer name	Orientation	Sequence (5'-3')	Amplicon size (bp)
qsgRNA	Forward	tcaagttgataacggactagcc	
qhCas9F	Forward	GGCATAAGCCCGAGAATATC	93
qhCas9R	Reverse	TCCTCTTCATCCTTTCCCTAC	
MS141 qGFP for	Forward	TGAGCAAAGACCCCAACGAG	94
MS142 qGFP rev	Reverse	TTGTACAGCTCGTCCATGCC	
qNatF	Forward	GCCATCGAGGCACTGGATGGGT	124
qNatR	Reverse	CGTCGGGGAACACCTTGGTCAG	
qTpLhcf9F	Forward	CCATGATGGGAATTCTTGGACT	141
qTpLhcf9R	Reverse	AGCCGAATGTAACCATTGTGCT	

Table A2-5. *Primers for amplification of fragments for fluorescent protein fusion.*

Primer name	Orientation	Sequence (5'-3')	Amplicon size (bp)
PUC-pTp14322-F	Forward	aggagtctggacttgacctctagagatgagttggggtcccacct	1585
Tp14322-mNeon-R	Reverse	CCTTGCTCACCATtggtacaatatcaatttggcaaccgc	
Tp14322-mNeon-F	Forward	gatattgtaccaATGGTGAGCAAGGGCGAG	736
mNeon-tTp14322-R	Reverse	acaactctgcaatTTACTTGTACAGCTCGTCCATGCC	
mNeon-tTp14322-F	Forward	CTGTACAAGTAAattgcagagttgtgaatctgctgt	256
tTp14322-PUC-R	Reverse	gcttgcatgcctgcaggttagaaaacggaaagctcttggacga	
mNeon-tTp14322-F new	Forward	GTACAAGTAAattgcagagttgtgaatctgc	257
tTp14322-PUC-R new	Reverse	caagcttgcatgcctgcaggttagaaaacggaaagctcttggac	
PUC-pTp264671-F	Forward	aggagtctggacttgacctctagaggtttggtggttatgaagtcaaggttgaaag	2402
Tp264671-mNeon-R	Reverse	CCTTGCTCACCATtacaaaggctatttgacgattcacaactC	
PUC-pTp264671-F new	Forward	ggacttgacctctagaggtttggtggttatgaagtcaagg	2391
Tp264671-mNeon-R new	Reverse	TGCTCACCATtacaaaggctatttgacgattcac	
Tp264671-mNeon-F	Forward	atagcctttgtaATGGTGAGCAAGGGCGAG	736
mNeon-tTp264671-R	Reverse	atgtagcatatatTTACTTGTACAGCTCGTCCATGC	
mNeon-tTp264671-F	Forward	CTGTACAAGTAAatatatgctacatgttgaactaaatgttataaataga atgct	248
tTp264671-PUC-R	Reverse	gcttgcatgcctgcaggtttggggagagagttggtgagga	
mNeon-tTp264671-F new	Forward	CTGTACAAGTAAatatatgctacatgttgaactaaatgttataaatag	246
tTp264671-PUC-R new	Reverse	ttgcatgcctgcaggtttggggagagagttggtg	

Table A2-6. *Primers for amplification of inserts for pTpPUC3 FO4.*

Primer name	Orientation	Sequence (5'-3')	Amplicon size (bp)
pFCPB_fwd	Forward	acctctagagtcgacctgcaAGCTTGCGCTTTTTCCGAG	1010
pFCPB_rev	Reverse	tcacagtactTTTGGTATCGGTTTGGTAAATC	
mNG_fwd	Forward	cgataccaaaAGTACTGTGAGCAAGGGC	734
mNG_rev	Reverse	gcggccgcgcTTACTTGTACAGCTCGTCC	
mTurq_fwd	Forward	cgataccaaaAGTACTGTGAGCAAGGGC	734
mTurq_rev	Reverse	gcggccgcgcTTACTTGTACAGCTCGTCC	
Lhcf9term_fwd	Forward	GTACAAGTAAGCGCGGCCGCATACTGGAttg	543
Lhcf9term_rev	Reverse	gccaagcttgcatgcctgcaGGGAGAACTGGAGCAGCTACTAC	

Table A2-7. *Primer for sequencing.*

Primer name	Orientation	Sequence (5'-3')
M13 Reverse	Reverse	CAGGAAACAGCTATGAC

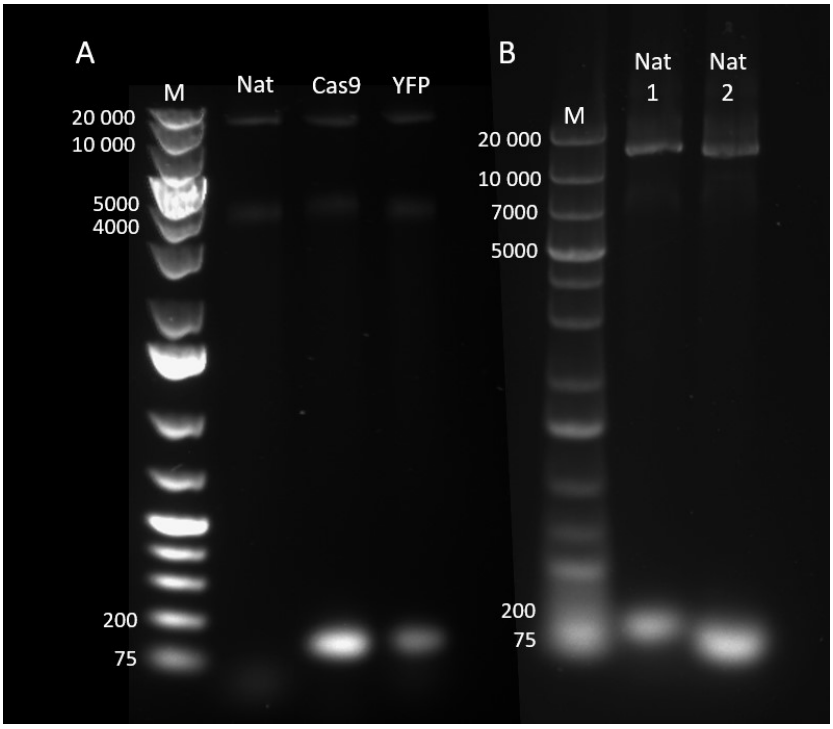

Figure A2-1. *Gel electrophoresis analysis with 1.0% agarose confirming that the primers used for amplification of the genes that were analyzed by qRT-PCR worked. Expected band sized (seen in Table A2-5, Appendix 2) ranged from ~90-150 bp. A) shows product Cas9 and YFP, but not for Nat (the enzyme for nourseothricin resistance). In another attempt B) regions for Nat were amplified with old Phusion polymerase (Nat 1) and new Phusion polymerase (Nat 2). GeneRuler 1 kb Plus DNA ladder (Thermo Scientific) was used as a marker to identify the sizes of the products (M). The numbers to the left of the markers indicate the length (bp) of the fragments.*