Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

Kristin Leonore Lillebo Bentzen

Functional studies of two genes encoding closely related group II silicanins in the diatom *Thalassiosira pseudonana*

Master's thesis in Chemical Engineering and Biotechnology
Supervisor: Olav Vadstein, Tore Brembu
June 2021

NTNU
Norwegian University of
Science and Technology

Kristin Leonore Lillebo Bentzen

Functional studies of two genes encoding closely related group II silicanins in the diatom *Thalassiosira pseudonana*

Master's thesis in Chemical Engineering and Biotechnology
Supervisor: Olav Vadstein, Tore Brembu
June 2021

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

PREFACE

This master thesis was a continuation of the specialization course TBT4500. Hence, a lot of the content in the first three chapters, in addition to some of the content in result, originates from the report produced in the course TBT4500. The master thesis was conducted at the Department of Biotechnology and Food Science (IBT) at the Norwegian University of Science and Technology (NTNU), during the period from January to June 2021.

I would like to express my deepest gratitude towards my supervisors professor Olav Vadstein and research scientist Tore Brembu for making my participation in the project possible. Many thanks to Tore Brembu for the help and for the guidance during the master thesis, and for always being available for questions. I am impressed by the amount of interest and knowledge you have, as well as the whole research group. In particular, I want to thank PhD Candidate Annika Messemer for all the support during the thesis, for always being available for questions, the guidance and training in the laboratory, as well as the good laboratory environment you bring with you.

Finally, I wish to thank friends and family for all their encouragement and support, especially my sister Kaisa Leonore Lillebo Bentzen, as well as my fellow master students for creating a good social environment throughout my time as a master student at NTNU. This would not have been the same without you. I would also like to thank Christine Våge Sjevelås for her unconditional support and guidance, you made the countless hours we spent together in the laboratory a joy.

ABBREVIATIONS

bp	base pair
Cas	CRISPR associated protein
cDNA	Complementary DNA
CFP	Cyan Fluorescent Protein
CPEC	Circular Polymerase Extension Cloning
CRISPR	clustered regularly interspaced short palindromic repeats
DNA	deoxyribonucleic acid
dsDNA	Double stranded DNA
ER	Endoplasmic reticulum
EYFP	Enhanced Yellow Fluorescence Protein
FRET	Förster Resonance Energy Transfer
GFP	Green Fluorescent Protein
HDR	Homology-directed repair
Kb	Kilobases
KO	Knockout
LB	Luria Bertani
LCPA	long-chain polyamine
TpLHCF9	Major fucoxanthin Chla a/c protein (Light Harvesting Complex Protein)
mRNA	Messenger Ribonucleic acid
NHEJ	non-homologous end joining
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PTM	post-translational modification
SAP	Silica-associated protein
SDV	Silica deposition vesicle
SNARE	Soluble NSF Attachment Protein Receptor
Sin	Silicanin
SW	Seawater
UV	Ultraviolet
WT	Wild type
YFP	Yellow Fluorescence Protein

ABSTRACT

Diatoms are a type of unicellular eukaryotic algae that has a large and diversified group, with an estimated 100 000 species. They are known for their ability to create highly complex structures made up of species-specific porous silica patterns (SiO_2) that make up the cell wall, or frustule, which is formed inside the silica deposition vesicle (SDV). The process of frustule biosynthesis is complex, including a huge number of genes and compounds, and our understanding of it is still incomplete. However, many important components of the biosynthesis process have been identified.

In this work, single and double knockout lines for two genes encoding Tp24711 and Tp24708, which are closely related group II silicanins, in the diatom *Thalassiosira pseudonana* were used to try to create a mutant strain. The cells were transformed using a vector containing the genes for the sgDNA containing the specified target sequence and the appropriate gene for the Cas9 protein, and the CRISPR/Cas9 complex was delivered through bacterial conjugation. The goal is to understand more about how these silicanins affect frustule formation as well as the functions and properties of the individual genes. In order to achieve this the shape of frustules and cell development will be examined. In addition, two genes, Tp24711 from the silicanin group II and Tpbd856-1852 from the silicanin group III, were examined using fluorescence protein fusion with the fluorescent marker mTurquoise to identify the positions of the gene products in vivo. The plasmid containing the amplified fragments (inserts) would be conjugated into *T. pseudonana* and analysed under a microscope to determine the location of the gene product in vivo.

Sanger sequencing revealed no alterations in the genes encoding Tp24711 and Tp24708 in contrast to certain anomalies in the colony PCR. Despite the lack of evidence of Cas9-induced mutation, the findings show that the episome is present in a small number of cells, but not in an overwhelming amount. As a result, episome-containing cells require a longer incubation period, as the cells are most likely to undergo multiple cell divisions before modifications can be observed. Another option is that the bacterial conjugation cloning method is ineffective for *T. pseudonana*, and that different cloning methods are necessary for this diatom species. Furthermore, the findings imply that the selection marker *nat* (nourseothricin resistant gene) is insufficient to encourage the episome to be retained. Another selection marker is needed because the antibiotic NOU is suspected of interfering with protein synthesis. *T. pseudonana* cell lines expressing mTurquoise-tagged Tp24711 and Tpbd856-1852fusion proteins to detect location in vivo were not accomplished, implying that other cloning processes should be tested out.

SAMMENDRAG

Kiselalger er en type encellede eukaryote alger med en stor og diversifisert gruppe, estimert til rundt 100 000 arter. De er kjent for sin evne til å lage svært komplekse strukturer som består av artsspesifikke porøse silika mønstre (SiO_2) som utgjør celleveggen, eller frustulen som dannes i spesifikke organeller kalt silika-deponeringsvesikler (SDV). Cellevegg biosyntesen er en kompleks prosess, hvor et stort antall gener og komponenter som er involvert, og vår forståelse av den er fortsatt ufullstendig. Hittil har mange viktige komponenter i Cellevegg biosyntesen blitt identifisert.

I denne masteroppgaven ble det gjort et forsøk på å skape mutanter med enkle og doble knockout-linjer for to gener som koder for Tp24711 og Tp24708, som er nært beslektede gruppe II silikaniner i kiselalgen *Thalassiosira pseudonana*. Cellene ble transformert ved bruk av en vektor som inneholdt genene for sgDNA med den spesifiserte målsekvensen og det passende genet for Cas9-proteinet, hvor CRISPR / Cas9-komplekset ble levert gjennom bakterie konjugering. Målet er å bedre forstå hvordan disse silikaninene påvirker dannelse av celleveggen, samt funksjonene og egenskapene til de enkelte genene. For å oppnå dette vil formen på celleveggen og celleutvikling bli undersøkt. I tillegg ble to gener, Tp24711 fra silicanin-gruppe II og Tpbd856-1852 fra silicanin-gruppe III, undersøkt ved bruk av fluorescens proteinfusjon med den fluorescerende markøren mTurquoise for å identifisere posisjonene til genproduktene *in vivo*. Plasmidet som inneholder de amplifiserte fragmentene (innsatsene) vil bli konjugert i *T. pseudonana* og analysert under et mikroskop for å bestemme plasseringen av genproduktet in vivo.

Sanger-sekvensering avdekket ingen mutasjoner i genene som koder for Tp24711 og Tp24708 i motsetning til noen forskjeller i koloniene på gel bilder. Til tross for mangel på bevis for Cas9-indusert mutasjon, viser resultatene at episomet er til stede i et lite antall celler, men liten andel av cellene. Som et resultat krever cellene som har episomet en lengre inkubasjonsperiode, da cellene mest sannsynlig vil gjennomgå flere celledelinger før mutasjoner kan observeres. Et annet alternativ er at kloningsmetoden for bakteriekonjugering er ineffektiv for *T. pseudonana*, og at forskjellige klonings metoder er nødvendige for dette spesifikke kiselalgen. Videre antyder funnene at seleksjonsmarkøren nat (nourseothricin resistent gen) er utilstrekkelig for å få cellene til å beholdet episoden. En annen seleksjonsmarkør er nødvendig på det grunnlag at antibiotika NOU er mistenkt for å forstyrre proteinsyntese. *T. pseudonana*-linjer som uttrykker mTurquoise-merkede Tp24711 og Tpbd856-1852 proteiner for å oppdage plassering *in vivo* ble ikke oppnådd, noe som antyder at andre kloningsprosesser bør testes ut.

TABLE OF CONTENT

1. INTRODUCTION

1.1 Background

Phytoplankton have existed for approximately 100 million years, and today they generate most of the organic matter which serves as food for aquatic zooplankton. Phytoplankton are found in oceans and freshwater worldwide, and flourishes wherever there is sufficient nutrients and light (1). The phytoplankton consist of photosynthetic bacteria such as cyanobacteria, and microalgae such as stramenopiles (heterokonts), rhodophytes (red algae), and chlorophytes (green algae) (2). Diatoms are considered to be the most important group of the phytoplankton. They are a diverse group of microscopic eukaryotic phytoplankton responsible for around 20 % of the photosynthesis on earth (1) and are ecologically important microalgae with high biotechnological potential (3). The exact number of diatoms species are not known, around 100,000 different species are estimated to exist (4). Diatoms vary notably in shape and size, with their size ranging from microns to millimetres (3), and can exist as single cells, chains of cells or colonies (5). One characteristic feature is their ability to generate highly ornamented silica structures which constitutes the cell wall, or frustule. The frustule consists of porous, species-specific nano- and micro-scaled patterns of silica (2, 6). The synthesis of their cell wall is not entirely understood, even though there have been identified many components involved, nevertheless it is a great example of biomineralization (6, 7).
Diatoms are used in several different studies such as bio-, nano- and environmental technologies due to their ability to construct three-dimensional silica structures and their ecological importance (8). Examples are as a source of lipids for biofuel production, as an ingredient in fish feed, as a nitrogen-fixing biofertilizer, in industrial waste detoxification, and in the synthesis of biomaterial and computer chips (9-11). The silica structures produced by diatoms exhibit highly controlled nanopatterns that can be produced by biological self-assembly in large quantities and at low cost (5), which makes them an attractive model for nanotechnology studies (1). Another outcome from the studies of diatom genomes is to further help uncover how they will respond to today's rapidly changing conditions in the oceans, which is a critical factor for the health of the environment (1).

1.2 Diatom evolution

According to the endosymbiotic theory, red and green algae higher plants have evolved through a primary endosymbiotic event. Diatoms are secondary endosymbionts and a part of the heterokont lineage (5). Diatoms differ from red and green algae as their plastids have four membranes, whereas the plastids of red and green algae normally have two membranes (2). It is therefore believed that the ancestor of diatoms evolved following a different pattern, where it underwent two separate endosymbiosis events over millions of years ago (12). In the primary endosymbiotic scenario, a unicellular eukaryotic heterotroph cell successfully engulfed or was invaded by a cyanobacterium. Resulting in a photosynthetic plastid with a

highly reduced cyanobacterial derived genome (13). During the secondary endosymbiosis event, the heterotrophic eukaryote then engulfed an autotrophic green algae, followed by an engulfment of a autotrophic red algae. These events gave rise to a host cell containing red algae chloroplast with green algae genes, creating the stramenopiles (14), after which the cell was heavily silicified, and then further diversified into the diatom known today. The cell that arose as a result of the uptake of two endosymbionts is probably the origin of all stramenopiles, and that the diatoms arose at a later stage, as illustrated in Figure 1. The reason the plasmids in diatoms have four membranes instead of two is believed to be due to the secondary endosymbiosis. One membrane derives from the endomembrane of the host cell, the plasma membrane derives from the engulfed red algae, and the other two from the primary plastid (13).

It has been hypothesized that the endosymbiotic event with the green algae preceded that of the red algae. The genes that derived from the green algae are very abundant within the diatoms, compared to the genes from the red algae which are less so. This might be the reason for the high abundance of diatoms in the oceanic environment, as the combination of red algal chloroplast with green algal genes encoded on the nucleus gave them selective advantages in their environment (3).

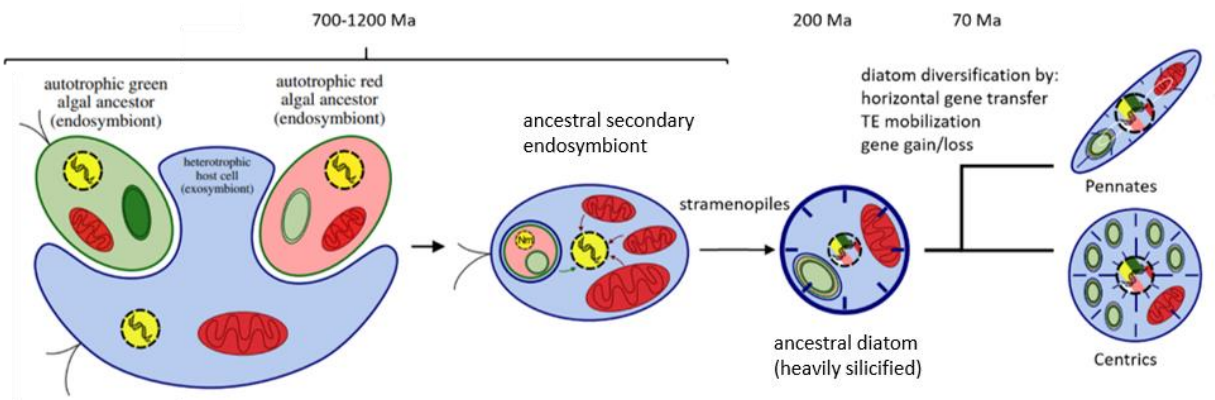

Figure 1: Schematic representation of the evelution, according to the endosymbiotic theory, from a heterotropic host cell to the evolved diatom, pennates and centrics. Figure modified from Benoiston et al. (3)

1.3 The diatom frustule

The most characteristic feature for diatoms is that they are surrounded by a cell wall, known as frustules (2), which is composed of amorphous silica (SiO_2) and organic-inorganic hybrid material (15). Due to the micro- and nano-scaled patterns of pores in the frustule each diatom species has their own characteristic pattern (5). To prevent the silica from dissolving in sea water, the cell wall is produced in an acidic silica deposition vesicle (SDV) and encapsulated in an organic matrix that is rich in proteins and sugars (1). The siliceous material of the frustule displays ornamented structures with highly regular patterns. Diatoms can be divided into two main groups, depending on the symmetry of their cell shape: pennate diatoms, which are elongated and bilaterally symmetrical, and centric diatoms, which are radially symmetrical (2).

Frustules are composed of two halves (theca), which are called epitheca and hypotheca. They are almost identical, the only difference being that the epitheca is slightly larger and overlaps the hypotheca, like a Petri dish, and together they completely enclose the protoplast (5). Each thecae are composed of a valve and girdle bands, as illustrated in Figure 2. The valve is located on the top of the thecae, displaying the larger external surface. The girdle bands consisting of a series of circular overlapping structures of silica called cingula, which is located at the edge of the thecae. The girdle band located at the point of attachment to the valve is called valvocopula, while the girdle bands located in the region where the epitheca and the hypotheca overlap is called pleural band (15).

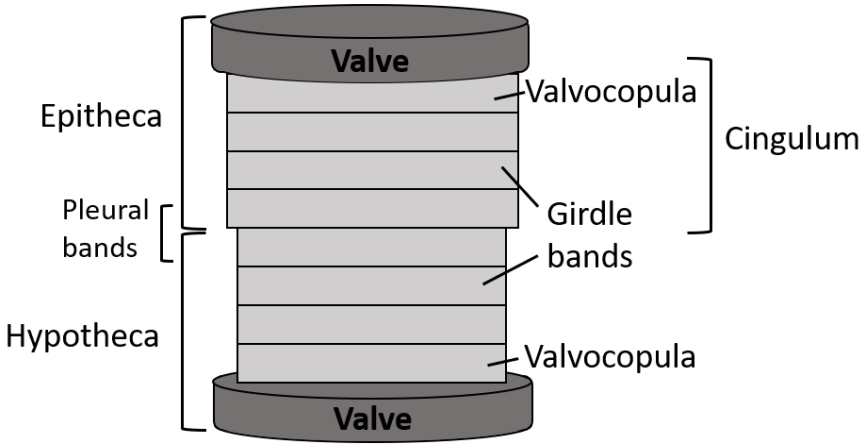

Figure 2: Schematic drawing of the frustule in diatom. Showing the two overlapping thecas, hypotheca and epitheca, as well as the valves, girdle bands, valvocopula and cingulum. Figure modified from Hildebrand et al. (6).

It has been suspected that the silica cell wall is one of the reasons for the evolutionary success of diatoms. One suggestion is that the silica cell wall may serve as a protective shield against phytoplankton predators (16). Another reason for their success is that biosynthesis of the silica cell wall is most likely associated with lower energy consumption, compared to cell walls made out of organic material (5).

During vegetative growth, the diatom cells divide mitotically. The stiffness and the layout of the silica frustule imposes restrictions when it comes to cell division and growth. Due to the rigidity, new valves can only be developed during cell division. The protoplast of the newly formed daughter cells is still retained inside the mother cell`s frustule after cytokinesis. Each daughter cell must synthesize a new theca *de novo* before separation, and therefore only inherits one theca from the parental cell (15). When generating the new valve inside the silica deposition vesicle, the SDV lays down a precise silica lattice work, followed by a coating consisting of organic matter (5). This is done to prevent dissolution when exocytosed and the two daughter cells are separated. Through this process, the inherited parental theca becomes the epitheca, whereas the newly synthesized one becomes the hypotheca. The two cells

separate from each other and each of them synthesize new girdle bands (15). Once separated, to prevent gaps when increasing the valve-to-valve distance as the cells grows, girdle bands are synthesised stepwise during interphase (Figure 3) (5). Diatoms reproduce primarily by mitotic division, which will result in a stepwise reduction of the average cell size for most of the diatoms (2). However, some diatom species (*Thalassiosira pseudonana*, *Phaeodactylum tricornutum* and *Cylindrotheca fusiformis*), for unknown reasons show no size reduction when grown in laboratory conditions (15). Another method for reproduction is by sexual reproduction, which will result in restoration of maximal cell size. With sexual reproduction the frustule is discarded and as a result the auxospore produce a new frustule with the restored maximal cell size (5).

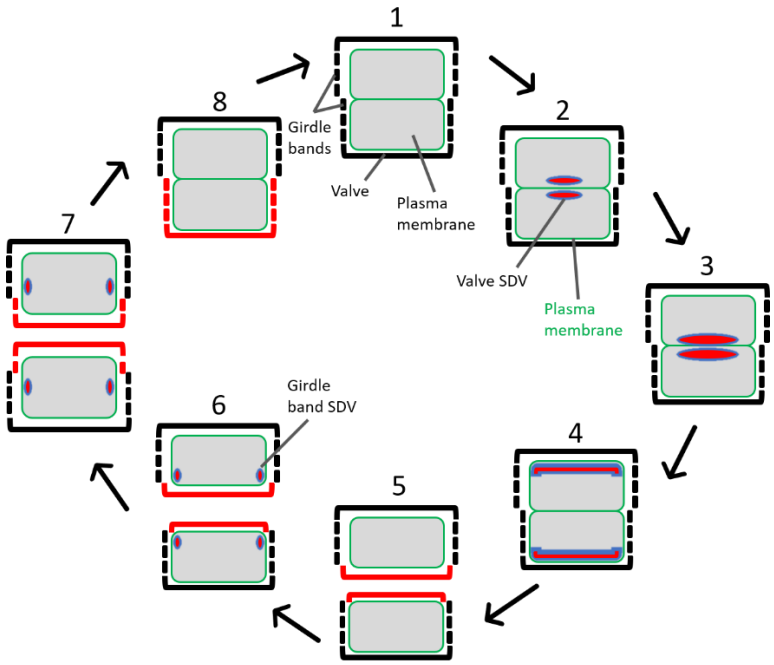

Figure 3: Schematic representation of the different stages of the ditaom cell cycle. The shaded gray represent the protoplast, the plasma membrane in green outline and the newly formed silica (red), and SDVs (blue). Valves are synthesized inside valve SDVs, which start to develop directly after cytokinesis (step 2-4). Fully developed valves are exocytosed to the cell surface, and the two daughter cells separate (step 5). Girdle bands are synthesized in separate girdle band SDVs and added one by one, in interphase (6-7). The final step, when the last girdle band is added, the cell expansion stops, and the cell cycle is completed (8). Figure modified from Kröger and Poulsen et al. (5).

1.4 Proteins associated with the silicalemma

As mentioned, the synthesis of the new components of diatoms frustule takes place inside the organelle called the SDV. The SDV is located in the cytoplasmic membrane and is surrounded by its own membrane called the silicalemma. The valve and the girdle bands, which are necessary building blocks for the cell plasma membrane, are first generated inside the SDV and

then exocytosed outside the membrane. Depending on what they synthesise, SDVs are categorized into two types, the valve-SDVs and girdle band-SDVs.

Cingulins, frustulins, pleuralins, silaffins, silacidins and long-chain polyamines (LCPAs), as well as scaffold structures like the cytoskeleton or chitin fibres are the organic molecules involved in frustule synthesis, that have been identified (15). Depending on a variation of different factors the morphology of the silica varies. Through examination of structure formation in diatom it indicates that three different scales are involved, normally referred to as micro-, nano- and meso-scale features in valves (17). The two families of proteins that are associated with the silicalemma that plays a role in the frustule morphogenesis is Silicalemma Associated Proteins (SAP) and silicanins (Sin). The molecular characterization and in vivo localization revealed that they both are membrane proteins with a longer, luminal, N-terminal domain separated from a short, cytosolic, C-terminal domain by a single trans-membrane domain.
The SAPs are characterized by their motif organization, which consist of a trans membrane domain, serine rich region and a conserved cytoplasmic domain. There have been identified two different SAPs in *T. pseudonana*, TpSAP1, TpSAP2 and TpSAP3. Additionally, the Silicanin-1 (Sin1) has been identified in *T. pseudonana*. Silicanins are characterized by a long intraluminal N-terminal region, a single transmembrane segment, a short cytoplasmic sequence and a glutamine and asparagine rich domain with eight conserved cysteines (15).

Knockdown experiments with SAP1 and SAP3, as well as knockout experiments with Sin1, have given a further understanding of their role in the silica morphogenesis. They both have a prominent role in silica morphogenesis. Fluorescent tagging of the three SAPs (TpSAP1, TpSAP2 and TpSAP3) in *T. pseudonana* identified the expression of TpSAP1 and TpSAP3 located at the site of the girdle bands and valve. Knockdown of TpSAP1 and TpSAP3 resulted in alteration in the frustule morphogenesis. Additionally, the knockout of TpSAP1 displayed malformed valves, where the position of the pattern forming centre were altered and showed abnormal distal pattering. Since the pattern forming centre are located by the microtubules it applies that the knockout of the TpSAP1 may interfere with the microtubule network and the interaction of SDV-associated proteins (18). Sin1 is associated with both valves and girdle bands, although it was later discovered that the Sin1 was also associated with the plasma membrane after their location in the SDV was determined (19). *T. pseudonana* knockout lines of Sin1 showed drastically compromised stiffness and strength of the cell wall, which was a result of the morphological aberrations and reduction of biosilica content (20).

There are a lot of different proteins involved in the synthesis of the frustule, but in this thesis the focus is on genes in the protein family silicanins. A phylogenetic analysis of the Sin1-like protein families revealed that there is four, possibly five subfamilies present in *T. pseudonana* and *P. tricornutum*, whereas 15 of 25 where identified members in *T. pseudonana* (Figure 4 a). All the genes in these families have the same overall structure, which is shown in Figure 4 b,

containing an N-terminal ER transit peptide, a canonical protease cleavage site (RXL) and a C-terminal transmembrane domain. There is one of the genes that were identical to Sin1, the Tp24701 gene in subgroup one. From the subfamilies the two closely related genes in subfamily two, Tp24711 and Tp24708, are the one focused on in this thesis, as well as one from subgroup four, Tpbd856-bd1852. Tpbd856 and Tpbd1852 constitute the 5' and 3' parts of the same gene (7).

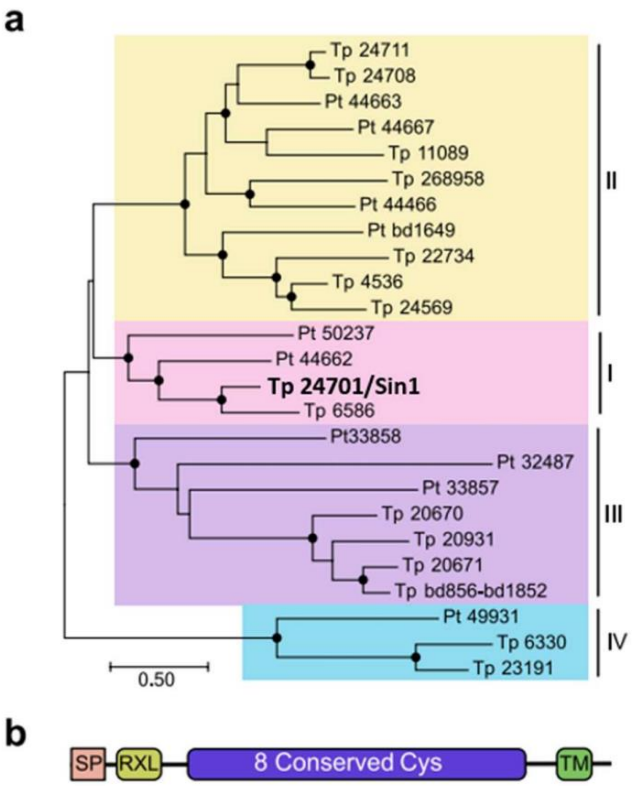

Figure 4: The Silicanins protein family (a) Representation of the phylogenetic tree of silicanins assumed to be involved in the frustule biosynthesis. The four groups dividing the proteins families are indicated in different colours and numbers. The Sin 1 protein is shown in bold. The closely related silicanins used to establish knockout-lines, Tp24711 and 24708, located in group II, Tpbd856-1852 from the used for localization belongs to group III. TP stand for *Thalassiosira pseudonana* and Pt stands for Phaeodactylum tricornutum. (b) a schematic representation of the predicted domain structure of Sin1 like proteins, consisting of four components. Signal peptide (SP) used to co-translate import into the endoplasmatic reticulum, the protease cleavage site (RXL), eight conserved cysteine (Cys) and a transmembrane domain (TM). Image from (7)

1.5 Transport of proteins

In a study from 2010 it was identified that an ER-resident kinase was shown to phosphorylate recombinant silaffins *in vitro*. Through this study a proposed model was made, in which proteins of the cell wall, after they are synthesized at the ER, are directed to the Golgi (21). Frustulins and pleuralins are two cell wall-associated proteins not involved in the silica biogenesis which are theorised to be transported by the secretory vesicle from the trans-Golgi network to the cytoplasmic membrane. In contrast, all the cell wall-associated proteins that

are involved in frustule morphogenesis (such as silaffins, silacidins, cingulins and silicanins) might be transported to the SDV by a different type of vesicle

The SDV route has a receptor-like protein within the ER and the Golgi, which should be responsible for specific recognition of the transport substrates that also are involved in the vesicle generation. Further specific factors, such as regulatory GTPase and SNARE protein, are necessary for fusion of the vesicles with the cytoplasmic membrane or the SDV (15, 22). One transport pathway is when the proteins from the trans-Golgi directed to the vacuole are transported via vesicles. Another known transport pathway is where the vesicles of ER origin bypass the Golgi apparatus, by directly targeting the vacuole. Since there are similarities between the vacuole and the SDV, it is safe to assume that a transport pathway where the ER-vacuole-SDV bypasses the Golgi exist as well (23).

Four possible mechanism and pathways have been hypothesised for the delivery of cell-wall associated proteins to their final cellular localization, after the synthesis of the new cell wall components within the SDV is complete. However, there are only one of the four theories which correspond with what recent studies has shown, where the membrane of the SDV is fusing with the plasma membrane. In this hypothesis, the plasmalemma and the distal silicalemma fuse at the centre of the valve. The pulling of the membranes allows the SDV content to be exocytosed, and the distal silicalemma is then taken up into the cytosol and recycled, where the proximal silicalemma becomes the new plasmalemma (15).

Through different studies for silaffins it appears that their amino acid composition and related post-translational modification (PTM) are more important for their function than the protein sequence itself. All the silaffins that are known today in *T. pseudonana* carries similar modification, consisting of phosphorylation of hydroxy-amino acids (polyamine-type), and lysine residues (complex glycosylation) and sulfation. Silacidins also undergoes extensive post-translational processing during maturation of the pro-peptides. These proteins have many phosphorylated serine residues and contains several copies of the spacer sequences RRL, which is the target site of proteolytic cleavage. All silaffins as well as the silicalemma membrane protein (silicanins and SAPs) have at least one R-X-L (the Protease cleavage site, as mentioned in Figure 4) motif and several PTM sites. There is no data today informing of cingulins post-translational modification, but all except one carry similar motif.

1.6 *Thalassiosira pseudonana*

The diatom *Thalassiosira pseudonana* was the first diatom to have its whole genome sequenced. *T. pseudonana* is today the model organism for studying and understanding silica biomineralization and frustule synthesis in diatoms. The structure is cylindrical with circular valves with average diameter of 3.8 µm (± 0.4 µm) (12). The length of the cell depends on the salinity of the growth medium, but at average have the length of 6.5 µm (±1.6 µm) (24). Its nuclear genome has a size of 34.5 Mbp genome, which encodes about 11 400 genes (12).

Recent studies on *T. pseudonana* have revealed more than 100 proteins which is potentially involved in the diatom silica formation (21).

1.7 Genome editing in marine algae

Sharma et al. (8) performed a transgene-free genome editing experiment in the marine diatom *P. tricornutum* by using two different methods to deliver the CRISPR/Cas9 system to the diatom and induce mutation in a common target gene. The two CRISPR/Cas9 delivery systems, bacterial conjugation and biolistic CRISPR/Cas9 transformation, were compared regarding mutation efficiency, and the potential problems connected to constitutive expression of Cas9. Biolistic DNA delivery for transformation of diatoms is a very common technique, but ever since it was demonstrated that *Cas9* gene editing could be achieved by bacterial conjugation, it has been less used (Nymark et al., 2015) (25). The difference between these two methods is that conjugative transformation allows the vector to be maintained as an episome in the recipient cell, whereas the biolistic transformation results in transgene integration of vector DNA into the algae genome. Both methods have similar percentage of CRISPR-induced targeted biallelic mutations, but to induce biallelic mutations when the CRISPR/Cas9 system is episomal an extended growth period might be needed. Constitutive expression of Cas9, independent of the CRISPR/Cas9 vector system, can cause re-editing of mutant lines with small indels. By using the episomal CRISPR/Cas9 system, complication associated with the biolistic transformation system, such as unstable mutant lines caused by constitutive expression of Cas9 and permanent and random integration of foreign DNA into the host genome, can be avoided.

Compared to the biolistic transformation method, the bacterial conjugation based CRISPR/Cas9 method strongly reduce the risk of integration of vector fragments into the diatom genome and ensures a higher number of transformed colonies. Additionally, no expensive equipment is required with the bacterial conjugation method. By using biolistic method, it has been observed higher mutation efficiency, compared to when these genes are located at an episomal vector. However, the somewhat higher efficiency does not outweigh the disadvantages connected to the continuous presence and expression of the Cas9 nuclease. Having the CRISPR/Cas9 in an episome makes it possible to eliminate the plasmid, by removing the selection marker (nourseothricin in the medium), the cells will not need the episome, and it will be lost within a few weeks. Which is done to decrease the chances for re-editing and off-target mutations accumulating over time, increasing the chance to create stable knock out lines. Through biolistic transformation makes it impossible to isolate and remove vectors integrated into the genome.

1.8 Thesis objective

In this project, two genes (Tp24711 and Tp24708) encoding closely related group II silicanins in the diatom *Thalassiosira pseudonana* will be studied. Mutated strains will be established by

implementing knockout-lines for the two silicanins using the gene editing technique CRISPR/Cas9. The aim is to study how these silicanins affect frustule synthesis and gain a better understanding of the functions and characteristics of the different genes. This will be accomplished by looking at frustule morphology and cell growth. Additionally, the aim is to verify the localization of two genes *in vivo* by tagging Tp24711 and Tpbd856-1852 with the fluorescent marker mTurquoise.

2. FUNDAMENTAL INSTRUMENT USED / THEORETICAL BACKGROUND

2.1 CRISPR/CAS9

The CRISPR technology is a relative new technique and stands for Clustered regularly interspaced short palindromic repeat (CRISPR). This technology is a powerful tool and is used to edit specific sites in the genome, preferably small insertions / deletions, by using customizable specificities, such as CRISPR associated protein (Cas protein) (26). This method is used a lot after it was discovered that the adaptive immune system in prokaryotes which work against plasmid or viral infection, can be used to recognize and cleave/inactivate foreign double stranded DNA (dsDNA). There exist several different Cas proteins that all are endonucleases and possess various properties, where some induce cuts in dsDNA while other cut RNA. Cas9 was the first Cas protein to be detected, and cuts dsDNA. The CRISPR/Cas9 complex consists of the Cas9 endonuclease and a single guide RNA (sgRNA). This technique works by having a sequence of approximately 20 nucleotides at the 5´ end of the sgRNA which recognizes and direct the complex to the complementary DNA target site. When the sgRNA and Cas9 complex binds it forms base pairs with the target sequence complementary to the sgRNA, which enables the nuclease domain of Cas9 to cleave the DNA, inducing a site-specific double stranded break in the DNA. Cas9 can be directed to any target 5´ of a protospacer adjacent motif (PAM) site by localizing the target site adjacent to a PAM and then altering the first 20 nucleotides of sgRNA (Figure 5 a) (27).

CRISPR/Cas9 based editing depends on the repair of double-stranded breaks in target dsDNA; it can either be repaired by homology-directed repair (HDR) repair or non-homologous end joining (NHEJ). When repaired by HDR, it is done by assisted recombination, which can either be done by using exogenous DNA or using the other allele as a template. NHEJ is a faster repair mechanism, where the disrupted genes are repaired by creating single stranded overhangs that are ligated together. When ligated together it can result in insertion or deletion of base pairs, which can cause different mutations, such as frameshift mutation causing knockout of the gene (Figure 5 b) (27).

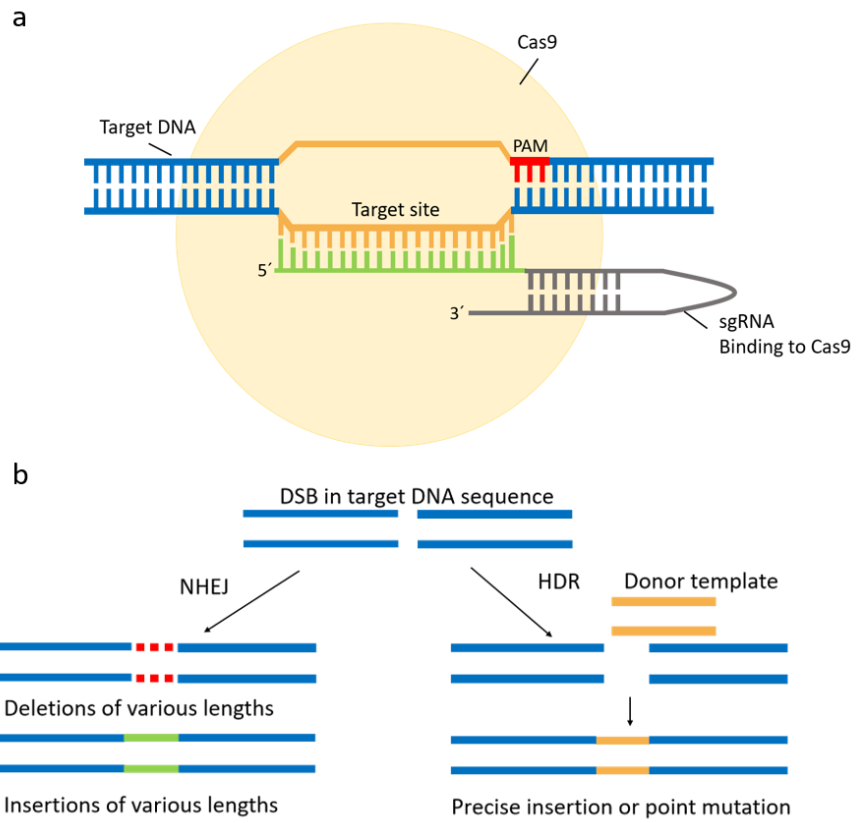

Figure 5: **(a)** Schematic representation of the CRISPR-Cas9 technology. The complex consist of a single gude RNA (sgRNA) directing the endonuclease (Cas9) to the DNA target site by a sequence of 20 nucleotides at the 5´ end of the sgRNA. The complex induces site specific double-stranded breaks in the target DNA, by forming base pairs with the matching sequence. Figure modified from Doudna and Charpentier et al. (27). **(b)** Repair mechanism of the endonuclease induced double-stranded break in the target DNA sequence. Repair by non-homologous end-joining (NHEJ) results in insertions of various lengths, while repair by homology-directed repair (HDR) results in precise insertion or specific point mutation of a template DNA sequence. Figure modified from Sander and Joung et al. (26).

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

Quantitative real-time PCR (qRT-PCR) is a technique where reversed transcription (RT) is followed by the polymerase chain reaction (PCR), which is used for detection and quantification of gene expression. This method is used to amplify cDNA which is reversed transcribed from the mRNA of a sample of interest. The cDNA is amplified in the presence of a fluorescence dye, that greatly enhance its fluorescence signal upon binding to the dsDNA. With increasing amount of PCR product, the signal from the fluorescent dye increases as well, which approximately doubles for each PCR cycle. When calculating the qRT-PCR result the quantification can either be absolute or relative (28). In absolute quantification to calculate an absolute value of gene expression, attempting to state the number of copies of the specific RNA per cell, either an internal or an external calibration curve can be used. Relative quantification is based on the comparison between the relative expression of a target gene and a reference gene (usually a stably expressed “housekeeping gene”), which is presented as

the relative quantity. The relative expression of the gene is determined by calculations based on the cycle threshold (Ct) value. The Ct value is the point where the fluorescence of the sample surpasses the fluorescence in the background, which are dependent on the mRNA concentration in the initial sample. Samples with more mRNA will have a lower Ct compared to the samples with less mRNA, which make the comparison of the relative mRNA levels between samples possible (29).

3. EXPERIMENTAL PROCEDURES

The materials, compounds and instrument used in these experiments are presented in Appendix D and Appendix E.

3.1 Genome editing in *T. pseudonana*

The generated CRISPR-vector pTpPUC3 with guide RNA was used to establish knockout-lines for Tp24711 and Tp24708, which is shown in Figure 6. To customize the plasmid for this experiment, the target sequence was changed of the single guide RNA to be specific for the PAM sites used by performing a restriction enzyme reaction and a ligation. sgRNA for four different target sites in each gene, Tp24708 and Tp24711, was used. The respective target sites were PAM1 and PAM2 for each gene, and since some sequences in the two genes are identical, it is possible to perform a double knockout of both PAMs in both genes (DoublePAM1 and DoublePAM2) Figure 7. The vector contains two BsaI restrictions sites placed at the 5´- to the target sequence in the sgRNA, which is used for restriction digestion and ligation of oligos to create the customized target sites. The target region that was ligated into the vector, targeted the selected genes which encode the two closely related silicanins in *T. pseudonana*. The sgRNAs were selected based on the low probability for off-target effects, creating the unique target sites in the genome of *T. pseudonana*. The pFCPB promoter and the terminator TpLHCF9 controls the *Cas9* gene, and it is tagged with the yellow fluorescent protein (EYFP). The pTpPUC3 vector also contains the genes for nourseothricin and kanamycin resistance. The oligonucleotides that were used are listed in Table 1.

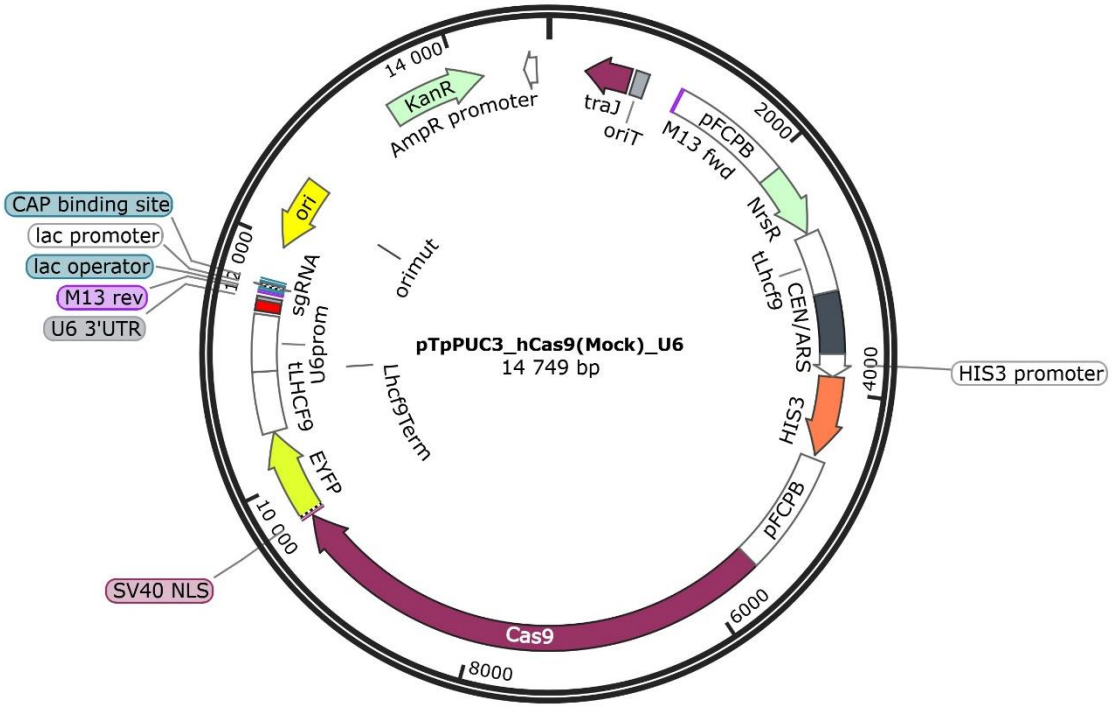

Figure 6: Plasmid map of the pTpPUC3-hCas9-U6 vector. This modified pTpPUC3 plasmid was used to establish knockout-lines by transforming the Cas9 genes into *T. pseudonana*. The vector contains a nourseothricin resistance cassette (NrsR) for resistance in *Thalassiosira pseudonana*, and a kanamycin resistance cassette (KamR) for resistance in E. coli, represented in the image as soft green. As well as the yeast derived sequence *CEN6-ARSH4-HIS3* (black), facilitating replication and segregation in diatoms, and an origin of transfer (oriT) (grey). Additionally, the plasmid contains the *Cas9* gene (purple) which is tagged with the fluorescent protein EYFP (yellow) controlled by the pFCPB promoter and the terminator tLHCF9. The sgRNA shown in red with the target region, is under the control of the U6 promoter. Plasmid map was created with SnapGene®.

Table 1: The list of the primers that were used for amplification of the genes by PCR. It shows the primer names and the corresponding sequence for each gene and their PAM. The location of these specific primers is represented in Figure 7.

Primer name	Orientation	Sequence (5`-3`)
Tp24708 F1	Forward	GCCUUCAAAACGACAAUGUGGA
Tp24708 R1	Reverse	GATTGATAGAGTTGTTCGCAAG
Tp24708 F2	Forward	GACTGGAATGTCCGTGAGTATG
Tp24708 R2	Reverse	GGCAGCCCTTTAGCTCTAGTGA
Tp24708 F3	Forward	CGATTTCCTCAACCGTCCTCG
Tp24711 F1	Forward	GAGACGACGAATCCTTTGCCAC
Tp24711 R1	Reverse	TCCTCCGTTCTCGGAACAGTAG
Tp24711 R2	Reverse	ATAGTCATGCCCTACAAACGTC
Tp24711 Pam1F	Forward	CCAGCACCAACAATACACCATG
Tp24711 Pam1R	Reverse	GCCTCGTCATTCCACACCTG

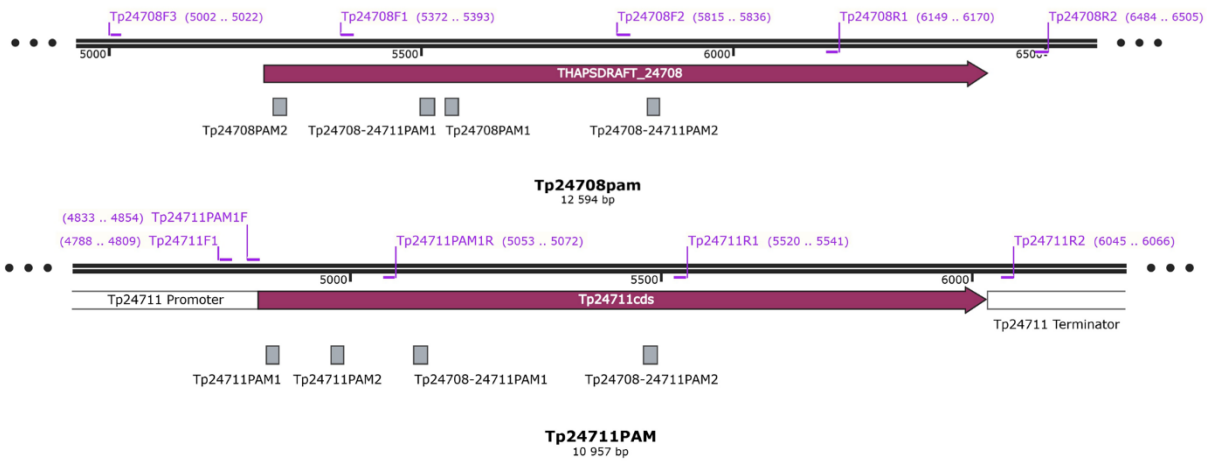

Figure 7: The map of the two genes in *Thalassiosira pseudonana*, *Tp24711* and *Tp24708*, and the location of the single and double PAM sites selected for CRISPR/Cas9-based gene editing. The gene *Tp24708* is indicated in purple and the same goes for the gene *Tp24711* with its respective promoter and terminator indicated in white. Figure was created with SnapGene®.

3.1.1 Constructing knockout-lines for Tp24708 and Tp24711

The pTpPUC3-Cas9 plasmid was cloned by changing the target sequence in the sgRNA. The target sequence was flanked by two BsaI cutting sites could therefore be customized to the specific for the PAM sites used, with oligo fragments ligated together.

The oligo fragments prepared by an annealing reaction set up which contained 1.4 µL each of the forward and reverse primers of each complementary oligo pair (Table 2), 1.4 µL T4 Ligase buffer and nuclease free water to a total volume of 50 µL. The reaction was incubated at 85°C for 10 minutes and then the temperature dropped 1°C per 90 seconds until the temperature reached 12°C.

Table 2: The list of the primers that were used for oligoannealing to create the target site of sgRNA.

Oligo name	Forward primer Sequence (5`-3`)	Reversed primer Sequence (5`-3`)
Tp24708 PAM1	ATTGTTGCTCCATGTCCAGCCCTGCA	AACTGCAGGGCTGGACATGGAGCAA
Tp24708 PAM2	ATTGTACCATCCTCTCAGCCGCCCTA	AAACTAGGGCGGCTGAGAGGATGGTA
Tp24711 PAM1	ATTGTGCCCTCCTCACCGCCGCCCTA	AAACTAGGGCGGCGGTGAGGAGGGCA
Tp24711 PAM2	ATTGTATTGTTCTCCTCGACGCGACA	AAACTGTCGCGTCGAGGAGAACAATA
Tp24708-24711PAM 1	ATTGTACGGAACCTCACAAGACGCTA	AAACTAGCGTCTTGTGAGGTTCCGTA
Tp24708-24711PAM 2	ATTGTCAACAACAACCGTAATCTCGA	AAACTCGAGATTACGGTTGTTGTTGA

The vector was linearized using the restriction enzyme BsaI-HFv2 by mixing 10 µl plasmid DNA (163,3 ng/µL), 5 µL CutSmart buffer and nuclease free water to a total volume of 50 µL. The reaction was incubated at 37°C for 60 minutes.

The ligated oligo fragments were ligated into the linearised plasmid by adding 2 µL 10 x T4 ligase buffer, 1 µL T4 DNA ligase and nuclease free water to a total volume of 20 µL. The

solution was incubated for 60 minutes at room temperature and afterwards heat-inactivated at 65°C for 10 minutes.

3.1.2 Heat-shock transformation of competent *Escherichia coli*

Heat shock transformation and the ligation products was performed, where 50 µL competent DH5α *E. coli* were thawed on ice, followed by an incubation on ice for 30 minutes with 5 µL ligation mix or 0.5 µl plasmid. After the incubation, the cells were heat shocked at 42°C for 45 seconds, followed by an incubation on ice for 2 minutes. 1 mL of LuriaBertani medium (LB medium) was added and incubated while shaking (220 rpm) at 37°C for 1 hour. The cells were centrifuged, and the supernatant was thrown. The pellet was resuspended in leftover medium. Everything was plated out on LB agar plates containing kanamycin (50 µg/mL) over night at 37°C. The clones that had grown on the plates were picked and again incubated in LB-medium at 37°C. To extract the plasmid a Miniprep was performed according to the GeneJet plasmid Miniprep protocol (Thermo scientific).

For 50 µL competent DH10β *E. coli* were thawed on ice for 30 minutes, followed by an incubation on ice with 1 µL plasmid for 30 minutes. After the incubation, the cells were heat-shocked at 42°C for 45 seconds, followed by an incubation on ice for 2 minutes. 1 mL of LB-medium were added and incubated with shaking (220 rpm) at 37°C for 1 hour. The 100 µL of the heat-shock transformed competent DH10β strain *E. coli* were then incubated on LB agar plates containing kanamycin (50 µg/mL) and gentamicin (20 µg/mL) over night at 37°C. a selection of the cells that had grown were picked and incubate them in 5 mL LB-medium with Kanamycin (50 µg/mL) and Gentamicin (20 µg/mL) overnight at 37°C.

PCR colony screening was used to look for the presence of the inserted oligos in the pTpPUC3 plasmid. This was done by picking single colonies from the selection plate for the different PAMs, streaking them out on a colony screen plate and transfer part of the colony to a PCR tube containing 0,5 µL insert-specific primers (the forward primer to each of the PAMs), 0,5 µL backbone-specific primers (Reversed M13), 12,5 µL VWR Red Taq DNA Polymerase Master Mix and nuclease free water to a total volume of 25 µL. The PCR program shown in Table 3 was used to obtain the PCR products.

Table 3: Thermocycling conditions for PCR amplifications using the RedTaq 2x Master mix.

Program	Temperature (°C)	Time	Cycle
Initiation	95	5 min	
Denaturation	95	10 sec	
Annealing	60	15 sec	X 34
Extension	72	20 sec	
Final extension	72	5 min	

The PCR products were separated on an agarose gel (1 % agarose in Tris-acetate-EDTA buffer) containing GelRed (5µL/100mL). To verify the inserted sequence and the orientation, plasmids that gave the correct band at 200 bp were submitted for Sanger sequencing. The plasmid DNA was purified and concentrated from *E. coli* overnight cultures according to the Thermo Scientific GeneJet plasmid Miniprep protocol. The manufacturer's instructions were followed with these exceptions; nuclease free water was used instead of elution buffer, and it was incubated for 5 minutes instead of 2 minutes. 7,5 µL of the flow thought was sent in, with 2,5 µL of either U6 Mock F or M13-rev for sequencing.

3.1.3 Conjugation

3.1.3.1 Preparation of E.coli cells and *T. pseudonana* culture

150 ml LB with kanamycin (50 µg/mL) and gentamicin (20 µg/mL) was inoculated with 3 ml transformed DH10β-*E.coli* overnight culture and incubated at 37°C with shaking (220 rpm). Culture (150 ml) grown to an OD600 value of > 0.3 was spun down for 10 minutes at 3 000 g, the supernatant was removed, and the cells were resuspended in 800 µl of SOC media.

Liquid culture of *T. pseudonana* were grown in f/2 medium made from sterile filtered (0.2 µm) and autoclaved local seawater (SW) and supplemented with nutrients and vitamins. The cells were grown under controlled conditions at 18°C with light (150 mmol m^{-2} s^{-1}), which is used throughout the whole experiment. The cells were harvested after approximately two weeks by centrifugation for 5 minutes at 4 000 g, 10°C, and the supernatant was removed. The cells were counted using BD ACCURI C6 flow cytometer and the concentration was adjusted to approximately 2 x 108 cells ml^{-1}.

3.1.3.2 Growing *T. pseudonana* culture

T. pseudonana cells were distributed to 6 different 1,5 ml eppendorf tubes and 200 µl of resuspended DH10β-*E.coli* bacteria in SOC medium were added and mixed by pipetting. The cells were then plated on f/2 + 5% LB plates and incubated in the dark at 30°C. After 90 minutes the plates were moved to 18°C in the light for 4 h. After incubation 1 ml of *T. pseudonana* medium was added to the plate and the cells were scraped. 200 µl of the 12 scraped cells was than plated on f/2 + Nou (100 µg/mL) plate, the rest (800 µL) was plated out on another f/2 + Nou (100 µg/mL) plate. The plates were incubated at 18°C under constant light. After approximately 7–14 days colonies appeared on the plates, and then 24 clones of each PAM were transferred to 24-well plates containing 1µL NOU and 1 mL f/2 + Si, and incubated further at 18 °C in constant light (30).

3.2 Characterization of mutants

3.2.1 Screening for mutants

DNA extraction was done by taking 100 µL of liquid grown colonies in the 24-well plates and spin for 1 minutes at 17 000 g. The supernatant was removed and the cells were resuspended in 20 µL lysis buffer by vortexing, followed by a 15 minute incubation on ice and 1 minute incubation at 95°C. 1 µl of extracted DNA was transferred to PCR tubes containing 1 µL of insert-specific primers (forward), 1 µL of insert-specific primers (reversed), 4 µl 5x Phusion HF buffer, 0.4 µl 10 mM dNTPs, 0.2 µL VWR Phusion hot start DNA Polymerase Master Mix and nuclease free water to a total volume of 20 µL. The PCR program shown in Table 4 was used to obtain the PCR product of interest, as well as the primers used for sequencing in Table E.1 Appendix E.

Table 4: Thermocycling conditions for PCR amplifications using VWR Phusion hot start DNA Polymerase Master Mix.

Program	Temperature (°C)	Time	Cycle
Initiation	98	∞	
Denaturation	98	10 sec	
Annealing	60	15 sec	X 34
Extension	72	15-30 sec/kb	
Final extension	72	5 min	

3.2.2 Flow cytometry (FCM) analysis

FCM analysis of *T. pseudonana* cells from the WT culture and from the three transformed *T. pseudonana* clones were performed using NovoCyte™ flow cytometer (ACEA Biosciences), to calculate absolute cell count and the chlorophyll content. Samples were excited by a 488 nm laser (Blue laser) and chlorophyll fluorescence emission collected on a detector with a 675/30nm bandpass filter (640 nm Red Laser). The YFP-fluorescence emission was collected on a detector with a 530/30nm bandpass filter.

3.2.3 Gene expression analysis

3.2.3.1 Cell harvesting

40 mL of the culture was centrifuged at 3 000 g for 10 minutes, and most of the supernatant was removed. The cells were resuspended in the remaining supernatant and transferred to 1.5 ml tubes and centrifuged again it at 3 000 g for 10 minutes, and the supernatant was removed, to make it as dry as possible. The pellet was snap-freezed in liquid nitrogen, and then immediately placed in a freezer (-80°C).

3.2.3.2 RNA isolation

One precooled (-80°C) 5 mm stainless steel bead was added to each of the tubes containing the frozen algae pellets and placed in a precooled (-80°C) adapter set. The frozen pellets were mechanically disrupted and homogenized by placing the adapter set, in a TissueLyser system (QIAGEN) and shaking for 2 minutes at 25 Hz. Afterward 450 µl lysis buffer RLT containing β-Mercaptoethanol (β-ME) (10µL β-ME/1mL buffer) was added and mixed by vortexing followed by a new mechanical disruption in a TissueLyser. The tubes were transferred to a room temperature QIAshredder spin column and then centrifuged for 2 minutes at full speed. To further lyse the samples, an incubation at 56°C for 4 minutes was performed. Total RNA was isolated from the lysates by using the protocol for Purification of total RNA from plant cells and tissues and filamentous fungi of the QIAGEN™ RNeasy Plant Mini Handbook, including the Appendix D: Optional On-Column DNase Digestion with the RNAse-Free DNase Set (QIAGEN) to remove residual contaminating genomic DNA (gDNA). A mixture of 10 µl DNase and 70 µl RDD buffer was added to each sample, followed by a 15 minutes benchtop incubation step. The purified RNA was eluted in (30 µl) elution solution and stored at -80°C.

3.2.3.3 RNA quality

In order to determine the concentration and purity of the isolated RNA NanoDropTM 1000 Spectrophotometer was used (Appendix C, Table C.1). The data was obtained by measuring the absorbance at 260 nm of two technical replicas per biological replica. To determine the quantity of RNA used in subsequent cDNA synthesis, the mean concentration of the samples was used. The purity of the RNA is indicated by the ratio of abundance at 260 and 280 nm, as well at 260 and 230 nm (the $A_{260/280}$ and $A_{260/230}$ ratios). RNA has an absorbance maximum of approximately 260 nm, in water. Absorbance of protein, phenols or other contaminants will at 280 nm influence the $A_{260/280}$ ratio of RNA by giving it a decrease. The RNA samples are generally considered pure when the ratio of $A_{260/280}$ is over 2,0. A decrease in the $A_{260/230}$ ratio is detected if carbohydrates or other contaminants are present in the sample, since it absorb light at 230nm. The RNA samples are generally considered pure when the ratio of $A_{260/230}$ is approximately in the range of 2,0-2,2 (31).

3.2.3.4 cDNA synthesis

Since it is difficult to maintain mRNA in a stable vector and difficult to manipulate, it is easier to convert it to complementary DNA (cDNA). cDNA is the conversion of RNA to DNA by reverse transcriptase, where an enzyme originating from retroviruses is responsible for the DNA is synthesized from mRNA (32). This was done to estimate the relative gene expression of specific genes in the sample originated from mRNA by qRT-PCR. For the synthesis of cDNA the QuantiTect Reverse Transcription kit (QIAGEN) was used from all the RNA samples. The kit was also used to prepare - reverse transcriptase (-RT) controls to check for gDNA contamination. A mix of 2 µl gDNA wipeout buffer, 7x (QIAGEN) and RNase free water (to a total volume of 14 µl) was added to 1 µg of RNA. To eliminate remaining gDNA, the reaction was incubated at

42°C for 10 minutes, Afterwards the gDNA elimination reaction was placed on ice. The 14 µl of the reaction was transferred to another tube for cDNA synthesis. For the -RT control sample, the total volume of the entire genomic DNA elimination reaction mix was 21 µL of which 14 µL was transformed to another tube for cDNA synthesis, whereas the remaining 7 µl was used for the -RT control. cDNA reaction mix with a total volume of 6 µl and 3 µl of the –RT reaction mix (Table 5) were added to the cDNA synthesis and the –RT control, respectively. Afterwards the mixture was incubated at 42°C for 30 minutes, followed by deactivation of the RT at 95°C for 3 minutes. The samples were diluted 1:10 and stored at -20°C.

Table 5: cDNA and -RT reaction mix components.

Component	cDNA mix	-RT mix
Quantiscript reverse transcriptase	1 µl	-
dH2O	-	0,5 µl
Reverse transcriptase buffer	4 µl	2 µl
Reverse transcriptase primer mix	1 µl	0,5 µl
Total rx. mix	6 µL	3 µL
Entire genomic DNA	14 µL	7 µL
Total volume	20 µL	10 µL

3.2.3.5 Quantitative realtime PCR (qRT-PCR) analysis

5 µL of the 1:10 diluted cDNA template and 15 µL master mix from LightCycler 480 SYBR Green I Master Kit was prepared for the qRT-PCR reaction in a LightCycler 480 Multiwell plate 96 (Table 6). Additionally, another dilution condition was tested for the YFP primer and the sgRNA primer, with 1:100 diluted plasmid. To check for primer dimerization and to confirm that there was no DNA contamination of the reagents no-template controls (NTCs) were included. NTCs were used as negative control and were prepared by exchanging the cDNA with autoclaved MQ water for each primer-pair. When the plate was prepared, it was sealed off with LightCycler 480 Sealing Foil and centrifuged for 2 minutes at 1500 g. The program used for the qRT-PCR reaction in a LightCycler 96 is presented in Table 7. Additionally, all primers used for genes of interest and reference genes are found in Table 8.

The data was analysed to obtain the melting temperature (Tm) and Ct values of the samples (using the LightCycler 96 software 1.1). The Tm is the temperature where half of the DNA product is double stranded, and half is single stranded. The melting curves were used for displaying the PCR product, and by analysing the melting peaks one can determine if any non-specific by-products or primer dimers are present together with the desired amplicon.

Table 6: qRT-PCR reaction mix from LightCucler 480 SYBR Green I Master kit.

Component	Volume (µl)
dH2O	3
PCR primers (10µM, 5µM of each)	2
2x LightCycler 480 SYBR Green I Master	10
Total	15 µL

Table 7: Thermocycling conditions setup for qRT-PCR amplification program.

Program	Temperature (°C)	Ramp (°C/s)	Time (s)	Cycles
Preincubation	95		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	2.2	30	1

Table 8: The list of the primers used for qRT-PCR analysis.

Primer name	Orientation	Sequence (5`-3`)
qNatF	Forward	GCCATCGAGGCACTGGATGGGT
gNatR	Reverse	CGTCGGGGAACACCTTGGTCAG
MS141_qGFP_for	Forward	TGAGCAAAGACCCCAACGAG
MS142_qGFP_rev	Reverse	TTGTACAGCTCGTCCATGCC
qsgRNA	Reverse	TCAAGTTGATAACGGACTAGCC
qTpLhcf9F	Forward	CCATGATGGGAATTCTTGGACT
qTpLhcf9R	Reverse	AGCCGAATGTAACCATTGTGCT
qhCas9F	Forward	GGCATAAGCCCGAGAATATC
qhCas9R	Reverse	TCCTCTTCATCCTTTCCCTAC

3.2.3.6 Plasmid extraction

The samples (10-20µL) containing the plasmid pTpPUC3 taken directly from grown *T. pseudonana* culture were first spun down at 1 700 g, for 5 minutes and then the pellet was treated with Tp Lysis buffer (1µL/1g), followed by an incubation for 15 minutes on ice, and on 95°C for 1 minutes. The rest of the preparation for the gel electrophoresis of the samples was done according to the Miniprep was performed according to the GeneJet plasmid Miniprep protocol (Thermo scientific), before loaded onto the 1% agarose gel, and run at 90 V for approximately 1 hour before the bands were visualized in the G:BOX gel documentation system, To determine the quantity of the plasmid pTpPUC3 in the *T. pseudonana* culture. If there quantity of the plasmid which was taken directly from the liquid grown culture, was to low to be detected by gel electrophoresis, two different procedures was performed. One of them was to amplify the plasmid by PCR using the primers in Table 9, before the gel electrophoresis was repeated. Another, were attempting to increase the amount of the plasmid by transforming it into DHα5 *E.coli*, and continue with a colony screening. The whole plasmid extraction method was repeated without performing the rest of the preparation for the gel electrophoresis of the samples, according to the Miniprep protocol GeneJet plasmid Miniprep protocol (Thermo scientific).

Table 9: The list of primers used for amplifying the whole plasmid.

Primer name	Orientation	Sequence (5`-3`)
OriMut F	Forward	GCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCA
TpPUC3R	Reverse	AGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
TpPUC3F	Forward	CTGCAATGATACCGCGTGACCCACGCTCACCGGCTCC
BsaIMutR	Reverse	TGGGTCACGCGGTATCATTGCAGCACTGGGGCCAGATGG

3.2.4 Analysing the morphology by microscopy

By using the fluorescence microscopy (Zeiss Axio Imager Z2 fluorescence microscope), the size and shape of the cells in the WT, mutant Tp24708 and Tp24711 single and double KO lines were assessed. 10 µl of each culture sample were transferred to a coverslip and a 24x24 mm cover glass was placed on top of the imaging spacer to seal the sample. The coverslip was placed in the microscope for imaging, using a 100X oil objective and a HXP-12 UV lamp. Three different images were taken with the fluorescence microscope, where one image was taken with bright field, another (Cy-5) had emission peak at 673 nm and the last one (Fluo-4) had a emission peak at 516 nm. The ZEN 2.3 software was used for image processing and to measure the length of the cells.

3.3 Tagging of fusion protein

To investigate subcellular localization of Tp24711 and Tpbd856-1852, fluorescent labelled proteins was created and expressed in *T. pseudonana*. To do a localization study, the genes were tagged with mNeonGreen and mTurquoise. Fluorescent labelled proteins are a useful

tool to investigate subcellular localization, mobility, transport routes and interactions of proteins in living cells (33). The Cyan Fluorescent Protein (CFP) mTurquoise is derived from *Aequorea victoria* GFP (34). Although CFPs are generally dimmer compared to Green Fluorescent Protein (GFP), CFP is a popular choice when performing Förster Resonance Energy Transfer (FRET) experiments, due to their blue-shifted fluorescence (35).

3.3.1 Vector modification

Figure 8 and 9 shows a representative map of the pTpPUC-mNeonGreen/mTurquoise vector which was used for transfer of the fusion proteins to *T. pseudonana*. The two gene of interest, *Tp24711* and *Tpbd856-1852* (from group III silicanins), was cloned and tagged with the fluorescence marker mTurquoise. Similar to the pTpPUC3 vector mention earlier, the pTpPUC-mNeonGreen/mTurquoise vector also contain the gene for kanamycin resistance. As indicated in Turquoise in Figure 8 and 9, the mTurquoise is inserted in the N-terminal part of the gene of interest, directly after the RXL cleavage Site. This is done to prevent the removal of the fluorescence marker by proteolysis when transported through a membrane. Since mNeonGreen and mTurquoise share the same sequence in the start and end, the same primers can be used. The pTpPUC3 was amplified by heat-shock transformation of competent DH5α *E. coli* cells with the plasmid.

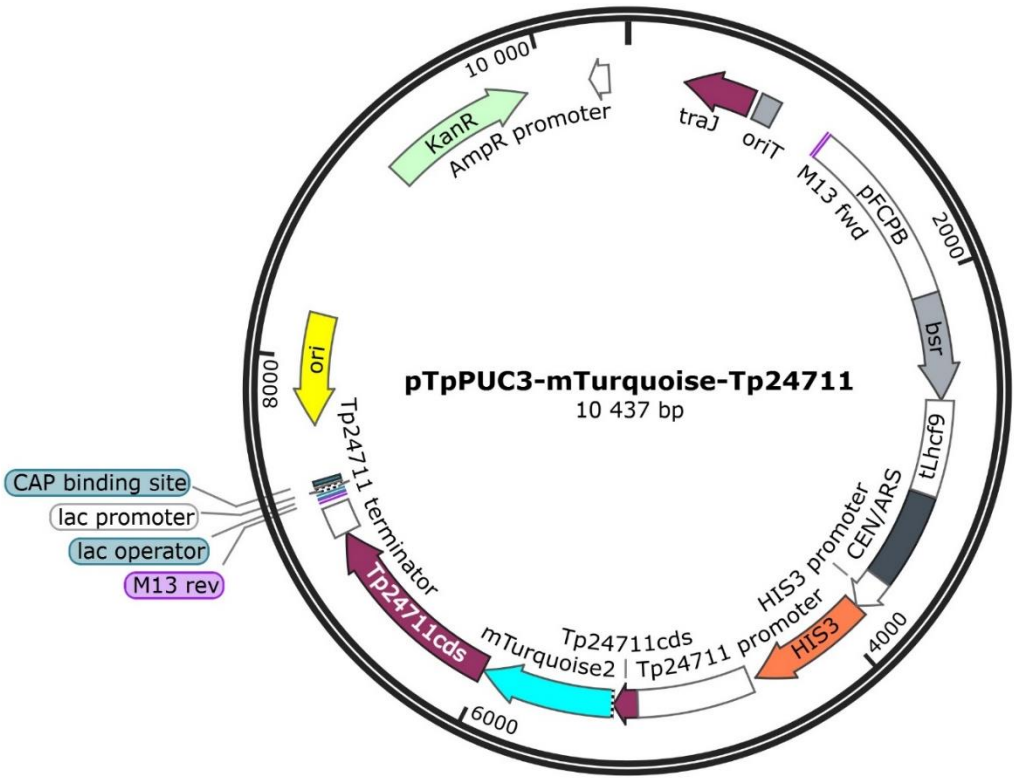

Figure 8: Plasmid map of the pTpPUC3-mTurquoise-Tp24711 vector. This modified pTpPUC3 plasmid was used to create mTurquoise constructs and contains a blasticidin resistance cassette (bsr) which was exchanged with nourseothricin resistance (NrsR), for resistance in *Thalassiosira pseudonana*, and a kanamycin resistance cassette (KamR) for resistance in E. coli, represented in the image as soft green. Additionally, the yeast derived sequence CEN6-ARSH4-HIS3 (black), facilitating replication and segregation in diatoms, and an origin of transfer (oriT)

(grey). mTurqouise2 (vibrant blue) is placed N-terminal to Tp24711, which is under control of the native promoter (Tp24711). Plasmid map was created with SnapGene®.

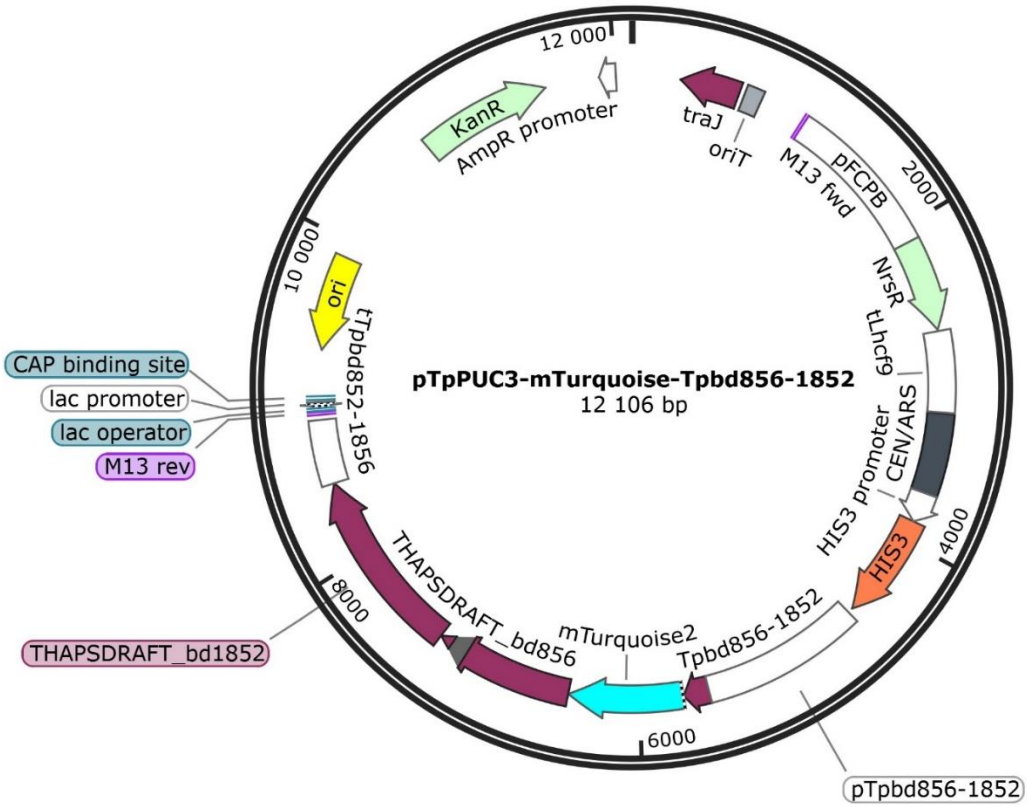

Figure 9: Plasmid map of the pTpPUC3-mTurquoise-Tpbd856-1852 vector. This modified pTpPUC3 plasmid was used to create mTurquoise constructs and contains a nourseothricin resistance cassette (NrsR) for resistance in *Thalassiosira pseudonana*, and a kanamycin resistance cassette (KamR) for resistance in E. coli, represented in the image as soft green. Additionally, the yeast derived sequence CEN6-ARSH4-HIS3 (black), facilitating replication and segregation in diatoms, and an origin of transfer (oriT) (grey). mTurqouise2 (vibrant blue) is placed N-terminal to Tpbd856-1852, which is under control of the native promoter (Tpbd856-1852). Plasmid map was created with SnapGene®.

3.3.2 Creating the fragments

In order to create fusions of Tp24711 and Tpbd856-1852 with mTurquoise, the fragments to be assembled into pTpPUC3 were amplified by PCR. A 5 x 200 µL Tp-culture was spun down for 1 minutes at 17000 g, after the supernatant was removed they were resuspended in 20 µL lysis buffer and mixed by vortexing, followed by 15 minutes on ice and then incubated at 95°C for 10 minutes. 1 µL of the Tp lysate was then transferred to PCR tubes containing 1 µL of each of the insert-specific primers (forward and reversed), shown in Table 10, 10 µl VWR Phusion hot start High-fidelity PCR Master Mix and nuclease free water were added to a total volume of 20 µL (Table 11). The PCR thermoprofile shown in Table 12 was used to obtain the PCR product. In order to confirm the correct length of amplified PCR fragments, the products were analysed by gel electrophoresis.

Table 10: The list of primers and the corresponding annealing temperature for the amplification of the fragments. The last column shows which fragment created and size each fragment has. The DNA template used for promoter and gene + terminator was T. pseudonana and for mTurqouise it was pNCS-mNeonGreen plasmid.

Forward primer	Reverse primer	Annealing temp.	Amplified fragment
Tp24711			
P24711_F	P24711_mNeonGreenR	64,5°C	Promotor (781 bp)
P24711- mNeonGreenF	mNeonGreen-c24711R2	60°C	mTurquoise (749 bp)
mNeonGreen-C24711_F2	t24711R	72°C	Gene + terminator (1245 bp)
Tpbd856-1852			
PUC-pTpbd856_F	pTpbd856- mNeonGreenR	60°C	Promotor (1196 bp)
pTpbd856- mNeonGreenF	mNeonGreen-Tpbd856R	69,6°C	mTurquoise (733 bp)
mNeonGreen-Tpbd856_F	tTpbd1852-PUCR	64,5°C	Gene + terminator (2467 bp)

Table 11: Reaction setup for PCR amplification with Phusion DNA polymerase.

Component	Volume (µL)	Final concentration
5x Phusion HF Buffer	10.0	1x
dNTPs (10 mM each)	1.0	0.2 mM
Sense Primer (10 µM)	2.5	0.5 µM
Antisense Primer (10 µM)	2.5	0.5 µM
Phusion DNA Polymerase (2.0 U/µL)	0.5	0.02 U/µL
Template DNA	4.7	2 ng/µL
PCR Grade Water	28.8	-
Total reaction volume	50.0	-

Table 12: The PCR thermoprofile used to amplify the fragments. *see Table 12 for the corresponding temperature.

Program	Temperature (°C)	Time	Cycle
Hot start	98	∞	
Initiation	98	5 min?? (30 sec)	
Denaturation	98	10 sec	
Annealing	*	15 sec	X 34
Extension	72	15-30 sec/kb	
Final extension	72	5 min	

3.3.3 Cloning the vector containing the fragments

The obtained fragments were inserted into the linearized pTpPUC3 using three different methods, Gibson assembly, Nested Gibson and CPEC, which was followed by heat-shock transformation with the NEB stable *E.coli* cells and sent in for sequencing. The Gibson assembly

reaction was performed according to protocol of Gibson assembly® by New England Biolabs. Different concentration and incubation time was tested out on all the methods.

3.3.3.1 Purifying DNA from agarose gel

The obtained DNA fragment after an agarose gel electrophoresis (1% agarose, 90V, ~1h) was isolated and purified based on size by performing a gel purification. The band from all the three fragments ligated together, promoter + terminator + gene, were cut out from the gel and purified. After obtaining the fragments on the gel, it was transferred to a UV box where the desired DNA fragment was cut out with a sterile razor blade. The gel with the gel fragment of interest was weighted and placed in a tube. The DNA was isolated from the gel according to the gel purification kit, QIAquick Gel Extraction Kit. The isolated and purified DNA fragment was then inserted into the linearized pTpPUC3 using three different methods, Gibson assembly, Nested Gibson and CPEC, which was followed by heat-shock transformation with the NEB stable *E.coli* cells and sent in for sequencing.

3.3.3.2 Nested Gibson

If the transformation of a Gibson assembly fails to yield any colonies or it is revealed through testing that the assembled plasmid is incomplete, a method called Nested Gibson can be used. This method is used to increase the efficiency of Gibson assembly that contains many pieces and/or DNA fragments that are difficult to assemble, where the failed Gibson assembly product will function as a template for PCR. By using the same primers used to amplify the fragments, new primer pairs can amplify multiple fragments that has already been ligated together through the failed Gibson assembly. The Nested Gibson was performed according to the Nested Gibson Assembly protocol found at protocols.io (36), apart from the master mix used, which is presented in Table 13. And the Thermocycling conditions for PCR amplification when performing Nested Gibson is presented in Table 14.

Table 13: Reaction setup for PCR amplification for Nested Gibson.

Component/Reaction mixture	Volume (µL)
5 x Phusion HF buffer	4
Phusion DNA polymerase	0.2
dNTPs 10 mM	0.4
Failed Gibson Reaction	0.5
Forward primer [10 µM]	1
Reverse primer [10 µM]	1
MQ water	12.9
Total reaction volume	20

Table 14: Thermocycling conditions for PCR amplification when performing Nested Gibson.

Program	Temperature (°C)	Time	Cycle
Initiation	98	10	
Denaturation	98	10	
Annealing	Tm-5°C	15 sec	X 30
Extension	72	10 sec/kb	
Final extension	72	5 min	

3.3.3.3 Circular Polymerase Extension Cloning (CPEC)

The method circular polymerase extension cloning (CPEC) is an efficient, simple and economical circular DNA assembly. This method is free of restriction digestion, ligation and single-stranded homologous recombination, since it is based on polymerase overlap extension (37). 100 ng of the linearized vector backbone and equimolar amounts of the assembly fragments were mixed together to a total volume of 20 ml. The assembly reaction mixture was performed according to Table 15. The PCR program shown in Table 16 was used to obtain the assembly product. The PCR assembly products were checked for successful assembly by heat-shock transformation of competent *E. coli* (NEB stable *E. coli*) and separated on an agarose gel.

Table 15: Reaction setup for PCR amplification for CPEC.

Component/Reaction mixture	Volume (µL)
5 x Phusion HF buffer	4
Phusion DNA polymerase	0.2
dNTPs 10 mM	0.4
DNA fragments, promoter	0.5
DNA fragments, mTurquoise	0.5
DNA fragments, terminator	0.5
Linearized Backbone pTpPUC3	3.5
MQ water	10.4
Total reaction volume	20

Table 16: Thermocycling conditions for PCR amplification when performing CPEC.

Program	Temperature (°C)	Time	Cycle
Initiation	95	5	-
Denaturation	95	30 sec	
Annealing	Tm-5°C	30 sec	X 34
Extension	72	15 sec/kb	
Final extension	72	10 min	-

4. RESULTS

4.1 Knockout lines of Tp24711 and Tp24708

Maps of the two members of the silicanin-like proteins in *T. pseudonana*, Tp24711 and Tp24708, and the location for the single and double PAMs including the primers used in this experiment are shown in Figure 7. As mentioned earlier, there are identical sequences in the two genes, *Tp24708* and *Tp24711*, which makes it possible to knock both genes out by targeting the same region. The results below show the attempt to establish a stable knockout line for both single and double PAMs.

4.2 Screening for mutations

This project is based on the continuation of previous work where genome edit of *T. pseudonana* cells was attempted, which is summarized below. Modified pTpPUC3 vectors targeting the two genes, *Tp24708* and *Tp24711*, were used for CRISPR/Cas9-based genome editing of *T. pseudonana*. sgRNA for six different target sequences were used in this project, two of them for each gene (PAM1 and PAM2) used in the single knockout experiment. For the double knockout experiment, the two identical target sequences in both genes were used (doublePAM1 and DoublePAM2). The sgRNA for the six target sequences were made by exploiting the BsaI cutting sites in the pTpPUC3 vector and was transformed into *T. pseudonana* by bacterial conjugation (DH10β-*E.coli*). Clones from the initial conjugation were examined for indications that mutations had occurred. Screening, microscopy and high resolution melting (HRM) analyses of the clones showed indications of mutations in the cells, by displaying some interesting traits. However, upon closer examination of the previous conjugation it was concluded that there were no mutations, which is contrary to what the previous results indicated.

Consequently, a new conjugation was performed and screened for mutations. The transformants were investigated by Sanger sequencing to screen for mutations in the target DNA sites. 16 samples of Tp24708 PAM1 and Tp24708 PAM1 were sent in for sequencing, but the result showed no occurrence of mutations. The same was done for the *Tp24711* gene, where 3 samples of Tp24708-24711 PAM1 and 24 samples of Tp24708-24711 PAM2 were investigated for mutations. Similar to the previous results, the genes showed no mutations, despite indication of mutation when analysed by gel electrophoresis (Figure A.3 and A.4, Appendix A).

4.2.1 Cell morphology

In order to look for variations in the cell morphology, the size and shape of the different cell lines (Tp24708 and Tp24711) were investigated under a fluorescence microscope. The selected clones (Tp2708 PAM 1 clone F and C, Tp24708 PAM2 clone B and Tp24711 PAM 1 clone D, I and J) were collected from liquid *T. pseudonana* colonies inoculated in f/2 medium containing NOU (100 µg/mL). The images were taken using 100X oil objective with three different filter

sets: bright field, Chlorophyll (673 nm) and YFP (516 nm). The YFP-fluorescence is used to validate if Cas9 is present in the cells, since Cas9 does not fluorescence on its one, it is tagged with EYFP. One clone from each PAM investigated under the microscope is presented in Figure 10-13, whereas the images for other clones of each PAM is represented in Figure B.1-B.13 in 3Appendix B. When looking at them through the microscope the majority of the cells did not resemble the *T. pseudonana* WT cell at all. Additionally, it was hard to find a single cell, as most of the cells were clumped together in a way in which the shape was almost not recognizable. Almost every cell had apparent chloroplast autofluorescence in the microscope, which indicates that there were living cells in the culture. Furthermore, all YFP fluorescence in the cells overlapped and corresponded with the chloroplast autofluorescence, except a few, as indicated by the white arrow in Figure 12. The fact that all the YFP fluorescence overlap with the chloroplast autofluorescence, indicates that there are no Cas9 present in the cells. Some of the cells showed YFP fluorescence outside the nucleus, which may suggest that the YFP has been transported from the nucleus to the lysosomes to be degraded. There were some clones that did not produce fluorescence and YFP in the nucleus, which can be explained by low RNA expression. Additionally, the cells in Figure 13 look like they are in the process of dividing, as indicated by the blue arrows.

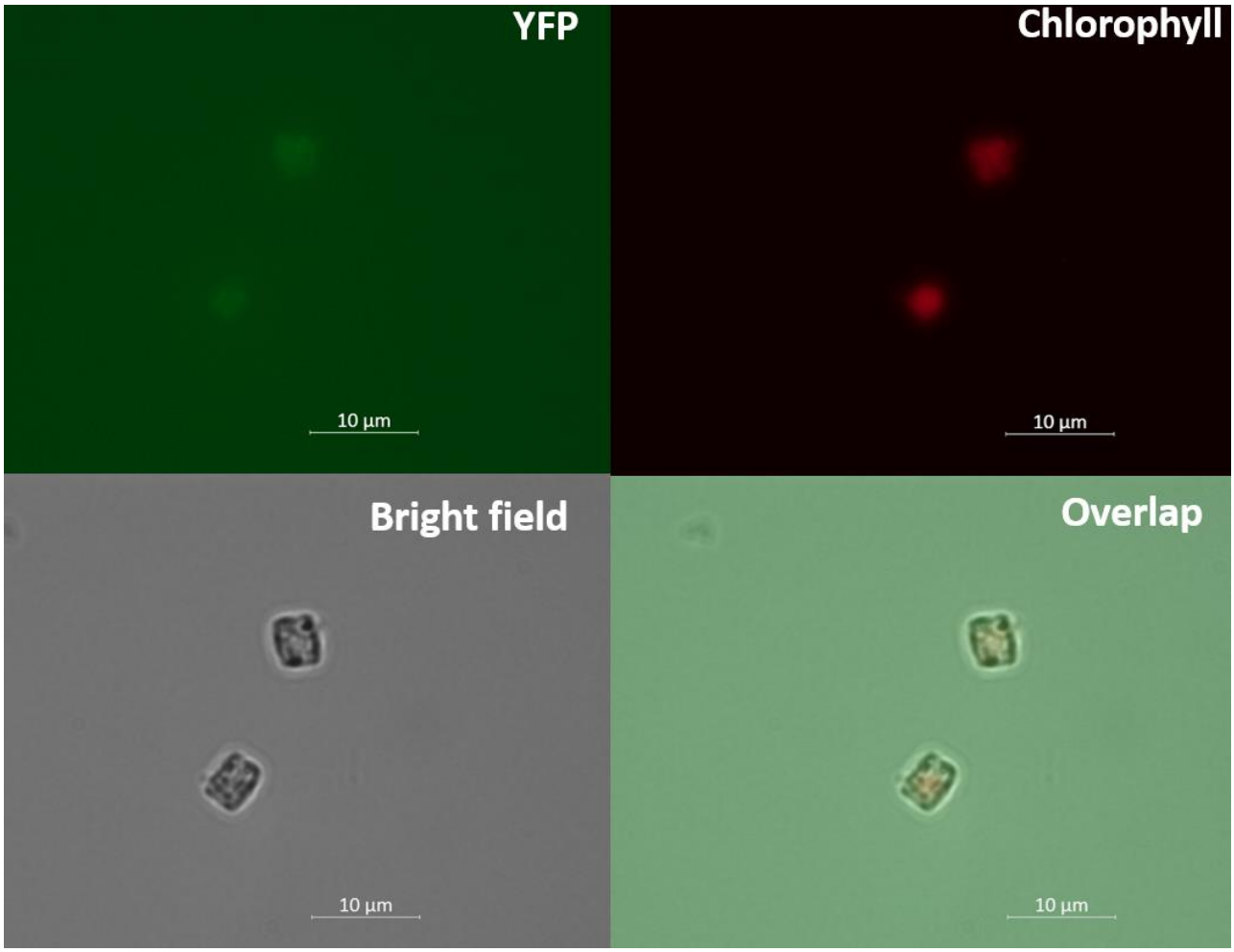

Figure 10: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Wild type cells. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

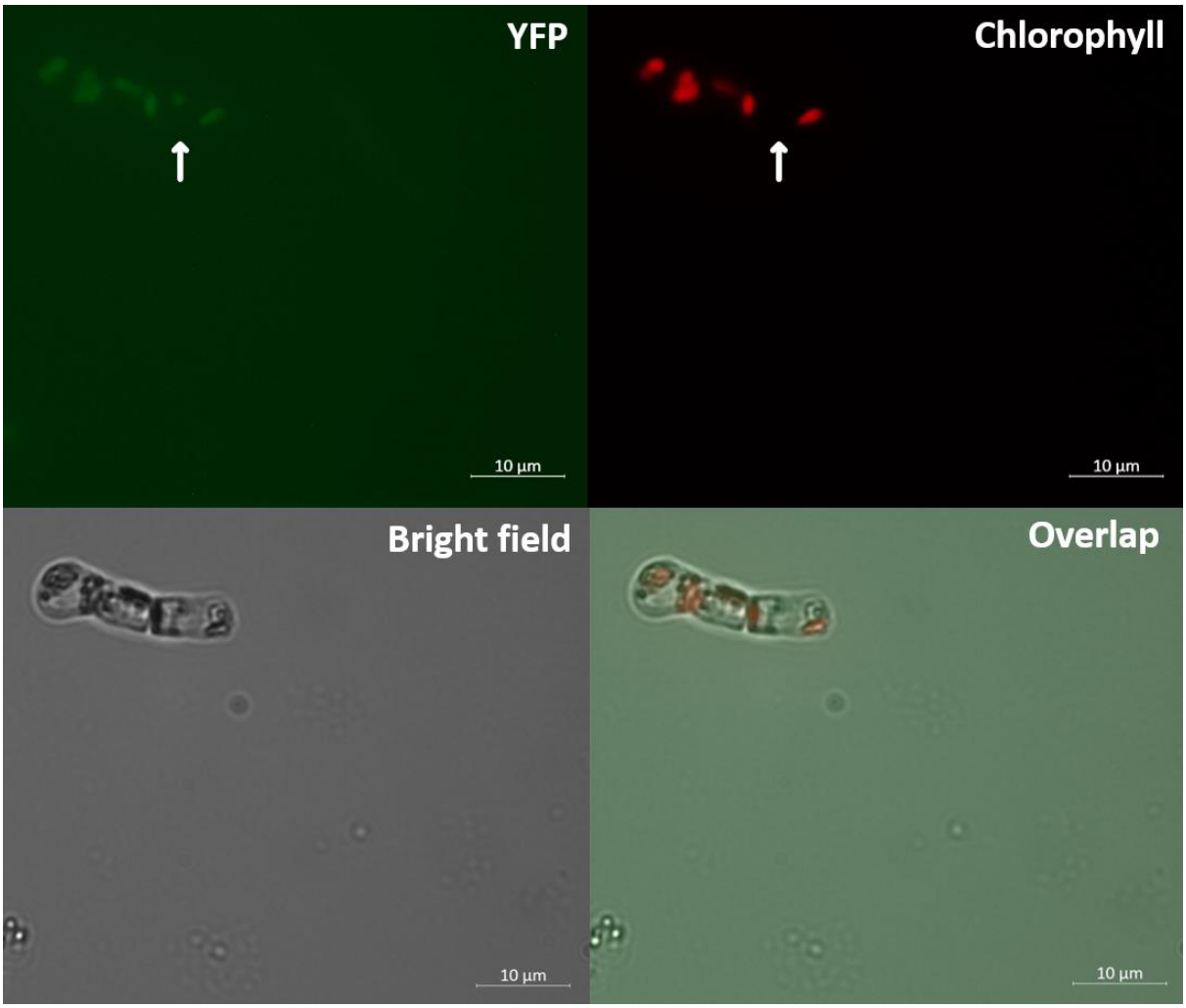

Figure 11: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24711 PAM 1 I. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

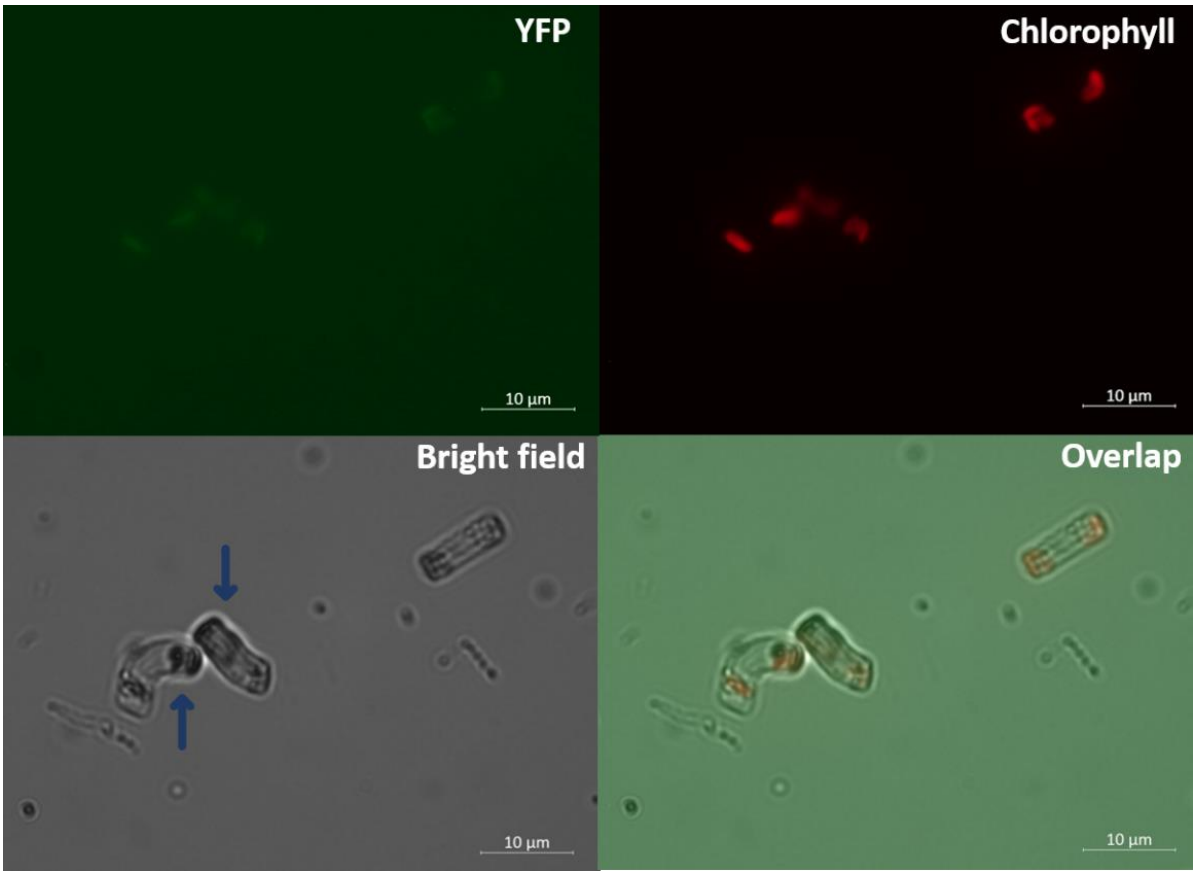

Figure 12: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24708 PAM 2 B. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

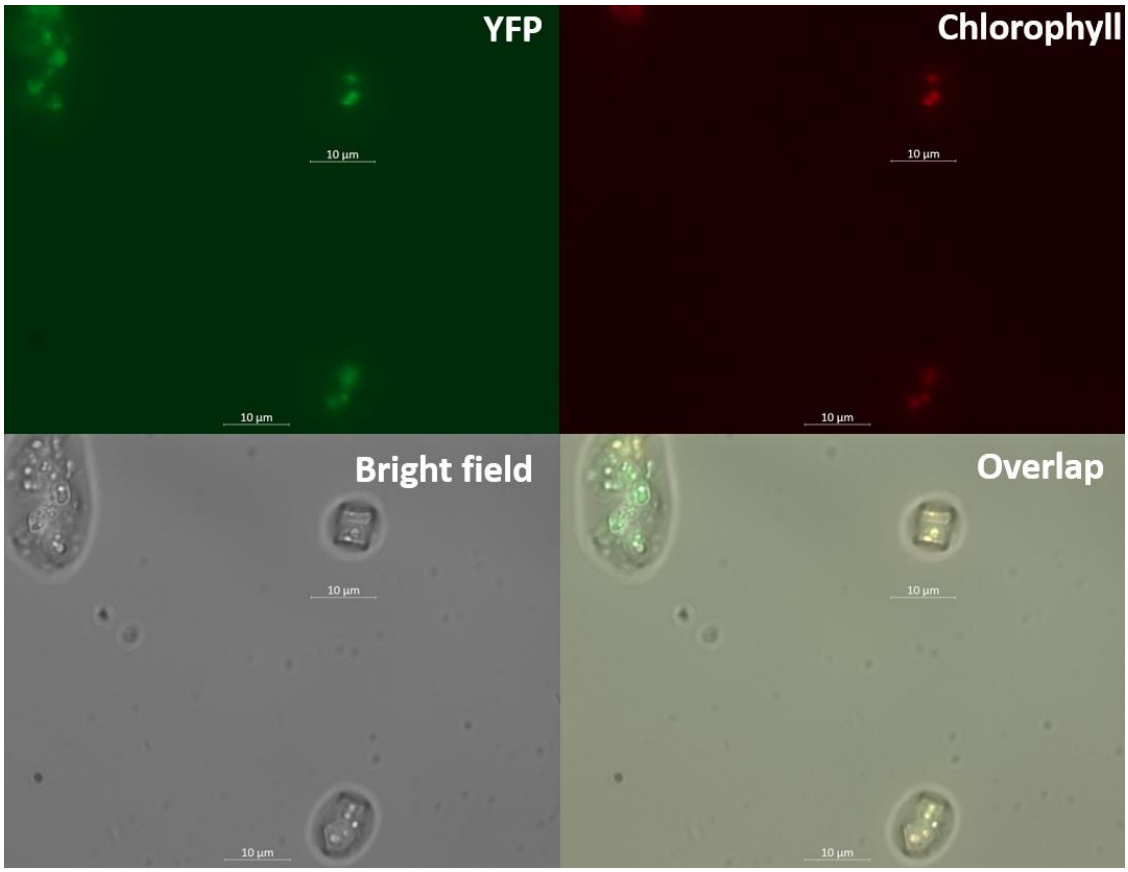

Figure 13: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24708 PAM 1 F. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

4.2.2 Expression of chlorophyll fluorescence

The analysis of selected transformants by flow cytometry was conducted to measure the total cell count and the indirect measurement of chlorophyll, through measurements of the chlorophyll fluorescence. For testing eight clones were selected from the liquid grown colonies after conjugation and harvested in the exponential growth phase. The eight clones are represented in Figure 14-16. Based on the data for each clone, two clones from Tp24708 PAM1 and Tp24708 PAM2 were selected for further work, as well as four clones from Tp24711 PAM1. The absolute cell count for the different clones varied from 673 000 cells/mL to 1 155 000 cells/mL, as shown in Figure 14.

The cultures were inoculated with the same amount of cell culture under det same condition, meaning the cells should in theory have the same amount of chlorophyll. From the graph displaying the chlorophyll data from flow cytometry for the selected clones (Figure 15), there is a clear difference in the distribution of chlorophyll between the clones and wild type cells. Both the wild type samples exhibited approximately 0.4 times more chlorophyll than the clones. Figure 16 shows the measured mean YFP- fluorescence, which represents the amount of YFP fluorescence each clone exhibits.

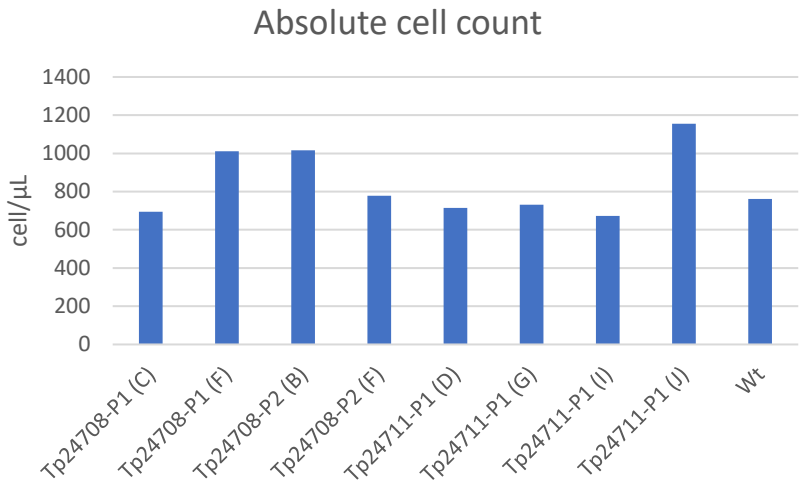

Figure 14: Flow cytometry data of *Thalassiosira pseudonana* cells, Indicate the absolute cell count, including both living and dead cells in the samples. Two clones from Tp24708 PAM1 and Tp24708 PAM2, as well as four clones from Tp24711 PAM1. WT is a wild type *T. pseudonana* cells.

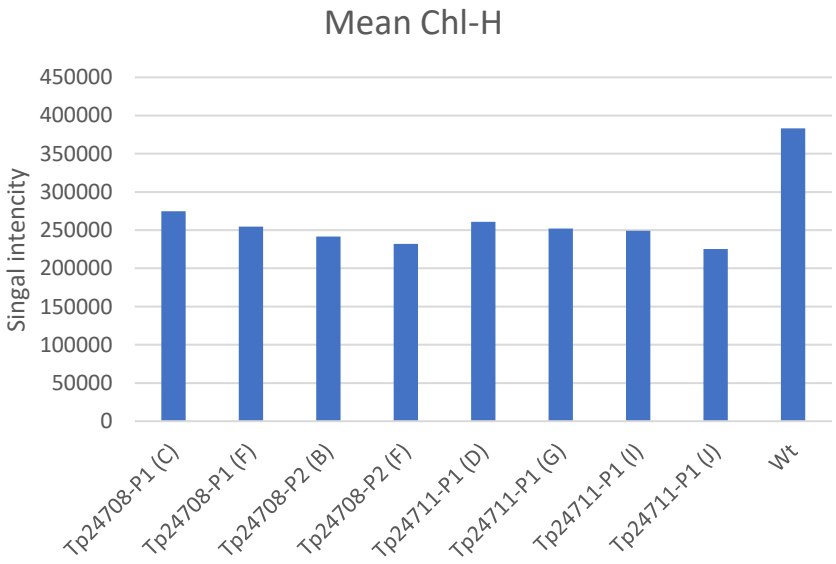

Figure 15: Flow cytometry data of *Thalassiosira pseudonana* cells, Indicate the measurements of the chlorophyll fluorescence. Two clones from Tp24708 PAM1 and Tp24708 PAM2, as well as four clones from Tp24711 PAM1. WT is a wild type *T. pseudonana* cells.

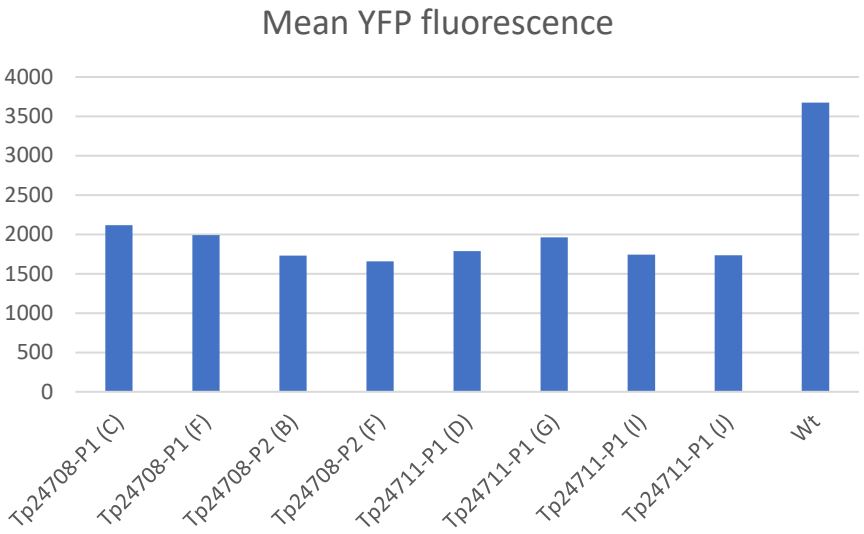

Figure 16: Flow cytometry data of *Thalassiosira pseudonana* cells, Indicate the measurements of the YFP fluorescence. Two clones from Tp24708 PAM1 and Tp24708 PAM2, as well as four clones from Tp24711 PAM1. WT is a wild type *T. pseudonana* cells.

4.2.3 Gene expression analysis of *T. pseudonana* cells

qRT-PCR was performed to investigate the expression level of different genes in the plasmid, the *Cas9*, *YFP*, *TpLHCF9*, the nourseothricin resistance gene (*NOU*) and sgRNA, and compare it to the expression level in WT. Both Cas9 and YFP primer was used to investigate the expression of Cas9, as they have different positions in the *Cas9* gene where the YFP is located at the end. TpLHCF9 is the shared terminator for both the *NOU* resistance gene and *Cas9*. By comparing the gene expression levels between the different genes, it is clear how the gene-editing component of the plasmid was expressed. Seven clones (24708p1 C and F, 24708p2 B and F, 24711p1 D, I and J) were selected for qRT-PCR analyses based on their different levels of YFP fluorescence from the flow cytometry result, were the highest and lowest level for each gene were selected (Figure 16). The graph (Figure 17) represents the mean Ct-value calculated from two samples of each mutant, apart from the result for Cas9 where only one sample of each mutant was used.

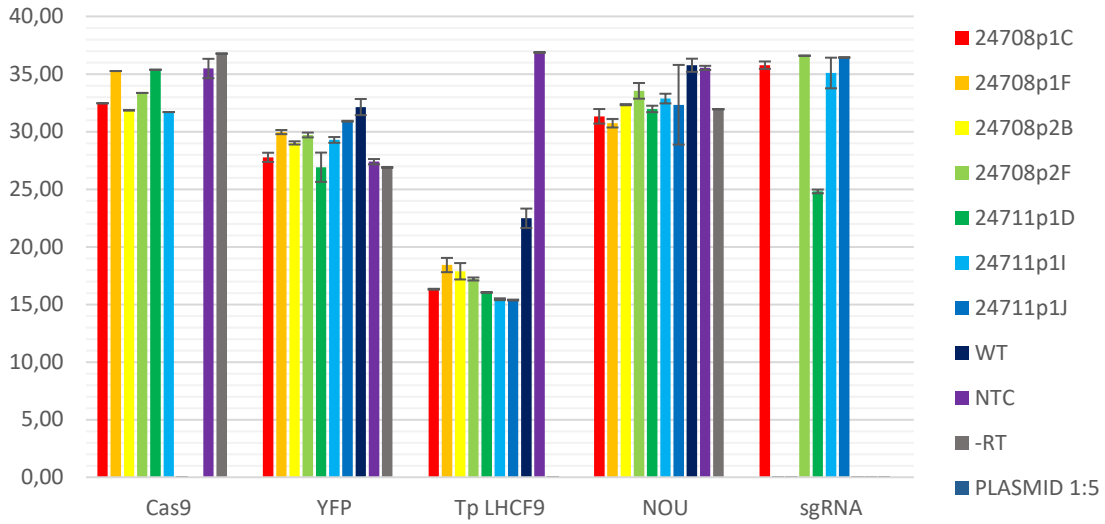

Figure 17: qRT-PCR analysis. Relative gene expression showing the Ct values of different genes from different clones. The Ct values were based on the mean of two copies of the same sample, and presented with a ±SD.

Ct-values of Cas9 ranged from 31.7 to 35.37, indicating little to no expression of *Cas9*. Both the NTC and -RT showed high Ct-values, 35.48 and 36.77, respectively. The Ct-values for the expression of YFP in the mutants, as well, as the NTC and the -RT ranged from 26.89-30.90, whereas the WT had the highest recorded Ct-value for the expressed YFP of 32.13. The low YFP expression supports the theory that there is a low expression level of *Cas9* genes in the plasmid. The Light harvesting complex (LHC) protein is highly expressed in *T. pseudonana* under light, and therefore should be highly expressed. As expected, the qRT-PCR result showed high expression of the TpLHCF9 in all mutants (15.38-18.42). The WT had a Ct value of 22.48 which is higher than the result of the mutants. The expression result of the *NOU* resistance gene was low, which is the opposite of what was expected. All the mutants, including the -RT, showed high Ct-values, ranging from 30.73 to 33.54. These result indicates that the mutants express very little of the gene for NOU resistance, if any at all. Seeing as the *T. pseudonana* culture is grown in f/2 medium containing NOU (100 µg/mL), the qRT-PCR result indicated that the cells have another mechanism for resisting the antibiotic. Additionally, the WT expressed for NOU also had a very high Ct-value, greater than 35, as expected since WT is a negative control for NOU (and Cas9, YFP and sgRNA), as the cells do not have plasmid. The qRT-PCR result of the expression of the sgRNA showed sufficient values in one of the mutants, Tp24711 PAM1 D, with a Ct value of 24.81, whereas the other mutants had a Ct value ranging from 35.09 to 36.59.

As expected, the WT showed tendencies of having a lower expression of all the genes compared to the mutants, with the exception of LHCF9. Additionally, the graph did not show a significant difference in the Ct value between the different mutants for each of the genes in the plasmid, apart from one mutant in the expression level for the sgRNA.
Positive controls containing diluted plasmid (1:100) was also analysed by the qRT-PCR along with the other samples (not included in the figure as the signal was negative). NTC (No

template control) indicated contamination or primer-dimer formation which could result in false positive results and -RT (no reverse transformation) gDNA contamination in the samples. The -RT and NTC values indicated that there was no contamination. The only one worth mentioning is the -RT and NTC in YFP. It still possesses a high Ct value, but not high enough to ignore it. The high CT value is most likely the result of primer dimers or unspecific bindings. The qRT-PCR results can vary depending on the primers from the kit, where some primers can be more efficient than others. The primers used in the qRT-PCR were tested on a gel to see if there could be any deficiencies. The gel images showed no evidence of deficiency for the primers used in the qRT-PCR (Figure A.1, Appendix A).

4.3 Plasmid extraction

Karas et al. (30) reported that from the 18 *T. pseudonana* colonies with the pTpPUC3/pTA-MOB plasmid used in conjugation, 16 of the *T. pseudonana* colonies had their plasmid successfully rescued in *E.coli* in which all had the correct plasmid. Several methods was tested in an effort to extract the plasmid, one of them being transformation with *E.coli*, but unlike other experiments carried out, the plasmid was not successfully extracted.

4.4 Fusion of fluorescent protein with Tp24711 and Tpbd856-bd1852

The aim for the fluorescence protein fusion experiment was to tag the two genes, *Tp24711* from the silicanin group II and *Tpbd856-1852* from the silicanin group III, with fluorescence markers to study the cells under a microscope. In order to identify the location of the gene product *in vivo*, the plasmid containing the amplified fragments (inserts) would be introduced into *T. pseudonana* by conjugation.

4.4.1 Amplifying fluorescence fragments

To identify the locations of the gene products *in vivo*, an attempt was made to tag the two genes from the silicanin protein family, *Tp24711* and *Tpbd856-1852*, with two fluorescence markers, mNeonGreen and mTurquoise. The amplified fragments containing the promoter, mNeonGreen/mTurquoise and the gene + terminator were ligated and assembled with the pTpPUC3-vector, followed by a heat-shock transformation into DH5α-*E. coli*. Clones containing the plasmid with the inserted fragments were investigated by colony screening. The size of amplified the individual fragments were compared to the expected fragment lengths, which are presented in Table 17.

Table 17: The expected size for the various constructs, when performing gel electrophoresis.

Gene	Amplified fragment	Size (Bp)
Tp24711	Promotor	781
Tp24711	mTurquoise	749
Tp24711	Gene + terminator	1 245
Tp24711	All fragments	2 775
Tp24711	Plasmid + fragments	10 437
Tpbd856-1852	Promotor	1 196
Tpbd856-1852	mTurquoise	733
Tpbd856-1852	Gene + terminator	2 467
Tpbd856-1852	All fragments	4 396
Tpbd856-1852	Plasmid + fragments	12 106

The mNeonGreen fluorescent marker was changed to mTurquoise as it was suspected that the genes used, interacted with an ankyrin repeated-containing protein in the SDV. A student had already cloned this gene with mNeonGreen (38), so to confirm this theory, the fluorescence marker was changed to mTurquoise with the goal of looking for fret in cells that were transformed with both plasmids.

The product of the digested fragments containing the promoter, mNeonGreen/mTurquoise, and the gene + terminator for the two different genes, *Tp24711* and *Tpbd856-1852*, are presented in Figure 18 and 19. To identify the optimal annealing temperature, different temperatures were tested for each fragment, using a temperature gradient for the annealing step. The temperature gradient is indicated in parentheses in the description. The successfully amplified fragments were used to assemble the plasmid containing the fragments.

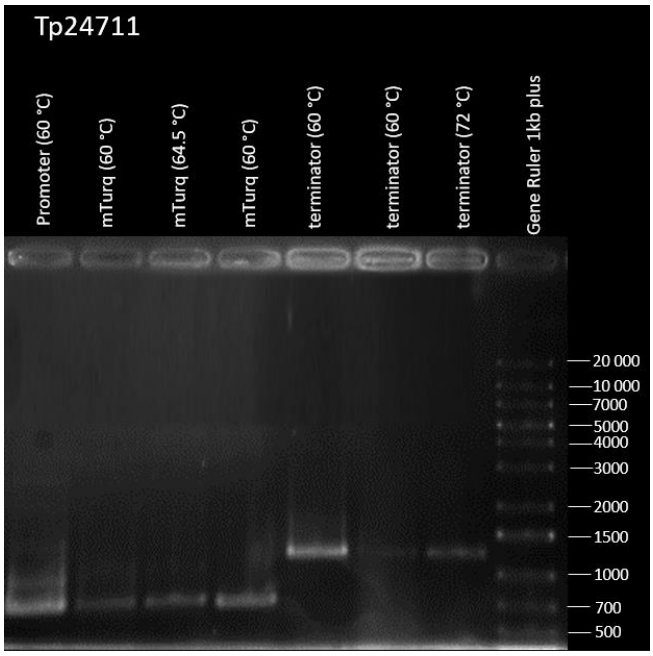

Figure 18: Gel electrophoresis analysis of PCR product with the amplified fragments used for assembly of vector containing the genes tagged with the fluorescence marker mNeonGreen and mTurquoise. The different fragments amplified with the numbers corresponding length of the different forward and reversed primers that was used, which is as follows; amplified promoter (781 bp), amplified mNeonGreen gene (749 bp), amplified *Tp24711* gene + terminator (1245). The numbers on the right indicate fragment sizes in the cursor (Gene Ruler 1kb plus).

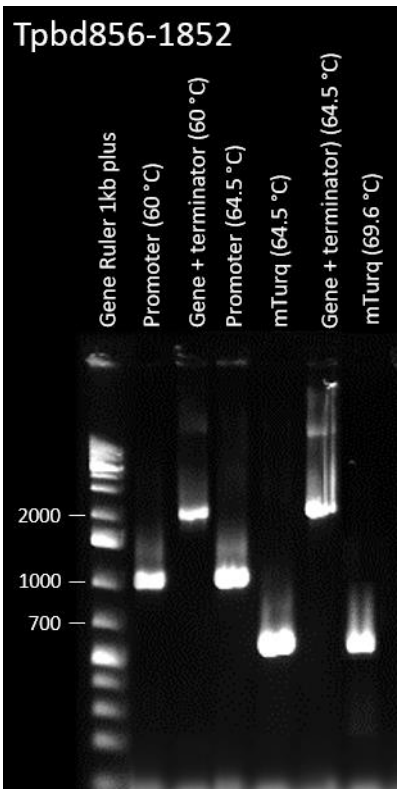

Figure 19: Gel electrophoresis analysis of the PCR product with the amplified fragments used for assembly of vector containing the genes tagged with the fluorescence marker mNeonGreen and mTurquoise. The different fragments amplified with the numbers corresponding length of the different forward and reversed primers that

was used, which is as follows; amplified promoter (1196 bp), amplified mNeonGreen gene (733 bp), amplified *Tpbd856-1852* gene and terminator (2467 bp). The numbers on the left indicate fragment sizes in the cursor (Gene Ruler 1kb plus).

4.4.2 Cloning fragments into pTpPUC3 vector

To clone the different fragments into the pTpPUC3 vector for both Tp24711 and Tpbd856-bd1852, several different cloning methods were used; Gibson assembly, nested Gibson, CPEC and purifying gel fragment. Despite multiple attempts and methods using different concentrations and incubation times for the assembly of the mTurquoise vector, the clones did not indicate that there had been a successful assembly. However, the result indicated that larger fragments had been produced and tested, but none with the correct size to have both the plasmid and all three fragments. Figure 20 (a) represents one of the attempts using CPEC. Additionally, the size of the vector was tested to look for unspecific bands and to ensure it had the correct size. The gel indicated several bands, the largest one having a size of approximately 3 000 bp for Tp24711. This result indicates that there had been a successful ligation for all three fragments, with the size of 2 775 bp. After running the Gibson assembly product on a gel, fragments containing the size of approximately 6 000 bp appeared for both Tp24711 and Tpbd856-bd1852, Figure 20 (b). If there had been a successful cloning, Tp24711 would have had a size of 10 437 and, Tpbd856-bd1852 would have had a size of 12 106, whereas the plasmid itself would have had a size of 7 796 bp. As the Gibson assembly failed, nested Gibson was performed (Figure 20 c). Similar to the other methods, nested Gibson did not show a band with the right size. However, it did give a band for Tpbd856-bd1852 at approximately 4 000 bp, indicating ligated product containing the promoter, mTurquoise, the gene and the terminator (4 395 bp). This was extracted from the gel and purified to use for cloning but did not show any different results than the previous attempts. Additionally, the vector was digested with PstI-HF and SacI and run on a gel to confirm that it had the correct size, Figure A.1 in Appendix A is showing no indication of such.

As none of the cloning methods to assemble the amplified fragments gave any satisfying result, no further work was done regarding the fluorescence protein cloning part of the project.

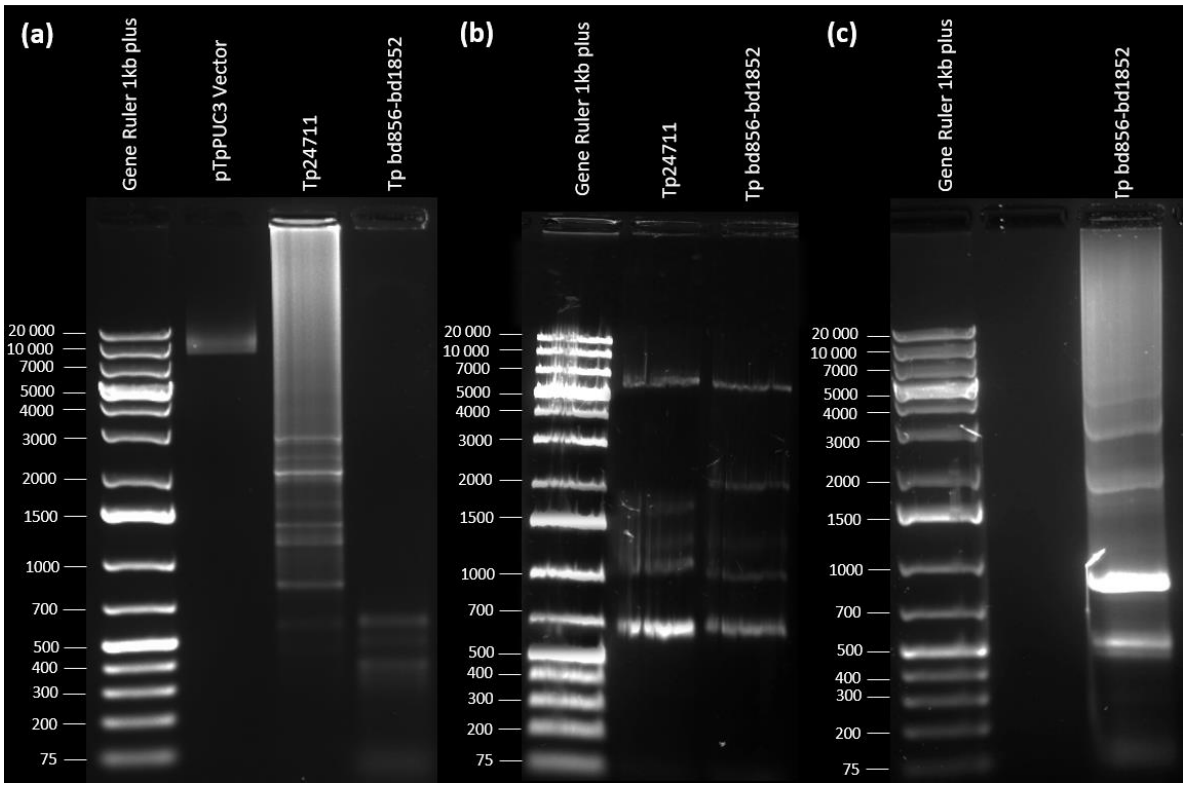

Figure 20: Gel electrophoresis (1,0% agarose) image from cloning the different fragments into the pTpPUC3 vector for both TP24711 and Tpbd856-bd1852, using three different cloning methods. (a) CPEC (b) Gibson Assembly (c) Nested Gibson. The different sizes of the amplified fragments for Tp24711, is as follows; promoter = 781 bp, mTurquoise = 749 bp and for terminator + Gene = 1245 bp. The different sizes of the fragments for Tpbd856-bd1852, is as follows; promoter = 1 195 bp, mTurquoise = 733 bp and for terminator + Gene = 2 467 bp. The numbers on the right for each individual image indicate fragment sizes in the cursor (Gene Ruler 1kb plus).

5. DISCUSSION

This paper is a preliminary attempt to investigate the possibility of genome edited *T. pseudonana* cells by the bacterial conjugation method to deliver CRISPR/Cas9 system and induce single and double knockout-lines. The main purpose of this thesis was to investigate if two genes encoding proteins of the silicanin protein family, *Tp24708* and *Tp24711*, in *T. pseudonana* are involved in frustule synthesis. The aim for this research was to unravel the mechanisms of frustule synthesis and the functions of their proteins. Additionally, a localization study was attempted to identify the locations of two members of the silicanin-like proteins *in vivo*. The subcellular localization was going to be studied, by creating fusions with the fluorescent protein mTurquoise and expressing it in *T. pseudonana*. In the knockout experiments, the focus was on the two closely related group II silicanins, Tp24711 and Tp24708 in *T. pseudonana*. In the localization studies, two genes, *Tp24711* from the silicanin group II and *Tpbd856-1852* from the silicanin group III, were studied. Two main approaches were used during these experiments: (1) create and characterize knockout mutants by CRISPR/Cas9-based gene editing; and (2) studying subcellular localization. Step 1 and 2 was done by performing Sanger sequencing, flow cytometry, qRT-PCR and looking at the clones with fluorescence microscopy.

The creation of Tp24708 and Tp24711 knockout lines by CRISPR/Cas9-based gene editing was performed a second time as it was not achieved the first time around, during the project thesis. The vector containing the Cas9 endonuclease and the sgRNA for the target genes were assembled and successfully transferred to *T. pseudonana* by bacterial conjugation. A colony PCR screening was performed on the grown colonies where it revealed that the genes amplified from clones had some differences compared to the WT genes. However, since the gel electrophoresis is not a reliable screening method, the transformant was further investigated by Sanger sequencing. The screening result for the target DNA site revealed that there were no mutants for any of the six sgRNAs targeting these genes. This could be explained by a low conjugation efficiency where the plasmid was not efficiently transferred to the *T. pseudonana* cells, but there was no indication of such. It is likely that the plasmid was efficiently transferred, but the sgRNAs did not work or the plasmid was ineffective. Several students have attempted to induce single and double knockout-lines by bacterial conjugation with the same plasmid containing the Cas9, but all struggled to successfully achieve mutations. It was speculated that it may be due to an ineffective plasmid, and therefore a new plasmid was created where sgRNA was modified. The modification was done as the expression of the sgRNA should stay constant. Unfortunately, but there is a point where it stops, resulting in premature and degraded sgRNA, resulting in premature sgRNA and degradation. The new plasmid containing the mutated sgRNA does not degrade that easily, which results in more full length sgRNAs (39). A new conjugation was not carried out due to the time restriction, so this hypothesis was not validated.

The fact that no mutations was detected could be the result of the knock-out mutation in the two genes causing a lethal phenotype. If the targeted genes codes an essential protein in *T. pseudonana*, the Cas9-inflicted mutation could be lethal and therefore explain why only WT cells was detected. However, this is a highly unlikely explanation as the result did not indicate frameshift mutations. Additionally, Sharma et al. (8) reported that when delivering the CRISPR/Cas9 vector by bacterial conjugation, there might be a delay in detection of mutations compared to the traditional delivery by biolistic transformation in *P. tricornutum*. Seeing as the colonies in this study were screened shortly after they appeared, the presence of only wild type cells might be caused by an occurrence of multiple cell divisions before mutations are induced by the Cas9. This result suggests that an extended incubation time before screening would be preferred. Additionally, under the course of the experiments, it was observed that even in the absence of antibiotic *T. pseudonana* cells did not grow well on agar plates.

Another, more likely explanation for the lack of mutation may be that there is not enough pressure from the antibiotic NOU for the *T. pseudonana* cells to preserve the plasmid containing Cas9. Interestingly, the qRT-PCR result revealed an extremely low expression of the nourseothricin resistant gene (*nat*) (Figure 17). The little expression of *nat* may explain the lack of mutations during this experiment. The pTpPUC3 plasmids containing *nat* as a selectable marker, could potentially be one reason for the cell to preserve the plasmid containing the Cas9. Nourseothricin acetyl transferase (nat) is an enzyme which inactivates NOU by acetylating the beta-amino group of the beta-lysine residue (40). As previously mentioned, the *T. pseudonana* culture is grown in f/2 medium containing 100 µg mL^{-1} NOU, which suggests that the cells must have another mechanism for resisting the antibiotic. A plausible mechanism might be that the *T. pseudonana* cells have a higher expression of ABC transporters. These transporters have the ability to create resistance to antibiotics by using the transporters to pump out NOU from the cytoplasm of the cell (41). It was previously suspected that the low expression shown in qRT-PCR of the NOU resistance gene was caused by the primers not working properly due to variations in efficiency of the primers from the kit. However, as previously mentioned, the primers were tested and showed no evidence of deficiency.

In order to investigate whether NOU is an acceptable selection marker for *T. pseudonana*, one could test other selection markers. For *P. tricornutum* transformants, zeocin is frequently used for selection. Zeocin is a type of antibiotics causing double stranded breaks. The zeocin resistance gene *sh ble* binds the antibiotic zeocin in a 1:1 ratio, preventing the formation of double stranded breaks in the DNA (42). The 1:1 ratio of zeocin and resistance gene provides the cell enough pressure to keep the episome and therefor give rise to Cas9 expression and then causing mutations. It has been reported that the use of the antibiotic zeocin in *P. tricornutum* effectively kills diatom cells at low antibiotic concentrations, whereas NOU were relatively ineffective (43). These findings suggest that Zeocin is more effective as a selection marker. Zeocin as a selection marker in *T. pseudonana* was previously tested in Poulsen et al.

(44) using biolistic shooting. It resulted in very little efficiency. It was assumed that expression of She ble is too low to overcome toxicity of zeocin.
Although zeocin is proven to be the more efficient selection marker, NOU is most commonly used in *T. pseudonana* on the basis of NOU being approximately one hundred times less toxic to *T. pseudonana* than zeocin (44). The result indicates that NOU does not have the same 1:1 ratio as zeocin. One reason might be that the nat is extremely efficient, and the *T. pseudonana* cells do not need constant expression of *nat* to flourish in f/2 containing NOU. A few of the cells may have kept the episome in order to express nat and maintain their resistance of NOU for the whole community, causing the rest of the cells in the colony to lose the plasmid. One thing to consider is the efficiency of the promotor and terminator used in this study and also in previous studies (Poulsen 2006, et al (44)). Until now pFCPB and pNR (nitrate reductase) are the only promotors published for use in *T. pseudonana*. pNR is an inducible promotor that leads to aberrant morphology (Gorlich et al. (19)). pNR is therefore not the promotor of choice when investigating changes to the morphology of the frustule. Using Zeocin as a selection maker was unsuccessful and therefore discussed by the authors that the too little expression of *She ble* was the reason for that. For future research investigation new promotors could help overcome these issues. To summarize a different cloning strategy could resolve the issues connected to conjugation, such as biolistic transformation, as well as another selection marker, such as the bleaching herbicide PDS-norfurazon resistance-based selection system (45).

The YFP fluorescence results from the flow cytometry analysis was used to validate any differences in YFP signal. The results indicate that there is no functional YFP, which further indicates that there is no Cas9. The reason being that Cas9 is fused to YFP and cannot send out a fluorescence signal by itself. One important point to consider is that although the DNA is present and extractable, the episome might be silenced due to methylation. The results imply that some cells keep the episome and express the YFP and sgRNA, but only to a small degree. Considering the mean signal, one will not see a shift if less than 1% of the cell in the culture possesses a higher signal. A higher signal in less than 1% of the cell is not enough to shift the mean. Other analyses show that only a few cells have a higher signal for YFP (personal communication).

The measured autofluorescence of chlorophyll attained by flow cytometry shows that all samples contain low chlorophyll contents. Furthermore, there is a clear difference between the distribution of chlorophyll in the wild types and the clones. Wild type and transformant cultures were inoculated with the same amount of cell culture under det same conditions, meaning the cells should in theory have the same amount of chlorophyll fluorescence. One explanation could be that the cells are smaller, resulting in lower chlorophyll fluorescence content. Another reason for the low chlorophyll fluorescence could be that NOU blocks or inhibits protein synthesis in *T. pseudonana*, resulting in cells having problems producing the necessary proteins. Validating this theory using WT control inoculated in F/2 medium

containing NOU would not be possible, as the cells would not survive the antibiotic. However, not seeing the protein does not necessarily mean there is no transcript, which is why the clones were further investigated by qRT-PCR to confirm if there is a transcript to produce Cas9. Additionally, we do not know how stable Cas9 is, or how stable YFP is which is bound to the Cas9.

When examining the result from the qRT-PCR analysis it is interesting to look at the gene expression of *LHCF9* in the clones. This is the only gene analysed which is located in the genome, whereas the other genes are only located on pTpPUC3-Cas9, making it the only gene present in wild-type cells. Every clone had, as expected, a high expression of the *LHCF9* gene, whereas the WT showed lower *LHCF9* mRNA levels. In some of the clones (Tp24708p1F and Tp24708p2B) *LHCF9* expression seemed to have slightly higher Ct-values (not significantly), but then again so did the error bars. The LHCF9 promoter and terminator have been used to regulate *Cas9* and *NOU* expression in the episome (plasmid map in Figure 6). The only difference being that *Cas9* and *NOU* is episomal. Therefore, one would expect the expression of *Cas9* and *NOU* to be at the same level as *LHCF9*.

The gene expression analysis revealed low *Cas9* expression in the clones, almost equal to the negative control, which correlates with the absence of mutations in the cells. Although there was low cas9 expression, some episome could potentially be present, although not expressed. Then again, if there is 1 out of 1000 cells expressing Cas9, it would not appear in the mixture of thousands of cells. Additionally, the low YFP fluorescence expression from flow cytometry, supports the theory that there are low expression levels of *Cas9* genes in the plasmid. Regarding the pTpPUC3 plasmid, there was no result in the qRT-PCR. One reason may be that the concentration was too low.

Only one clone, Tp24711 PAM2 F, expressed a sufficient amount of sgRNA from the qRT-PCR analysis. The result confirmed that the population of the one clone expressing the sgRNA contains the episome, at least that part of it. This can be caused by two reasons, the first being that only the U6 and sgRNA was left, which is highly unlikely. The second reason being that the clones contain the episome. Seeing as the cells react somewhat to the antibiotics, the episome is most likely present. It implies that at least part of the episome is present, whereas the rest of the episome is probably not accessible. This theory can be explained by either the cells are heavily methylation, or histone rep, or not all the cells carrying the episome in the mixture as stated above.

Although a number of the cells did not resemble the wild type when examining the cells under a fluorescence microscopy, none of the clones showed any indication of possessing Cas9. This again correlates with the lack of mutations. Nevertheless, one clone (Tp24711 PAM1 Figure 13) exhibited YFP fluorescence not overlapped by chloroplast autofluorescence, which could be an indication of Cas9. However, all the fluorescence images correspond to the results

indicating that there was only a very little part of the cells that had Cas9 expression. This correlates with previous result, where the results suggest that there is low expression and therefore few cells possessing the cas9, resulting in a small portion of cells in the colony which is difficult to locate. This presents a problem when it comes to finding the genome edited clones.

Several other master students (38) have attempted to genome edit *T. pseudonana* by bacterial conjugation, but few achieved a successful result. A disadvantage of using biolistic transformation to genome edit diatom is unstable mutant lines caused by continuous presence and expression of the Cas9 endonuclease. Another disadvantage is the permanent and random integration of foreign DNA into the host genome, which is avoided by using conjugation. As biolistic transformation results in transgene integration of vector DNA into the diatom genome, it is not possible to remove vector integrated into the genome. By using bacterial conjugation the vector can be maintained as an episome in the recipient cell, and later shut off the episome by relieving the cell of antibiotic stress. As a consequence, the bacterial conjugation method is preferred over the biolistic transformation method in *P. tricornutum (8)*. However, the lack of successful genome editing of *T. pseudonana* using conjugation may suggest that this method is not efficient for *T. pseudonana*. It can therefore be argued that a different method, such as biolistic transformation, should be used instead.

The fragment containing the gene for mTurquoise, the promoter, and the gene and terminator of the two genes, *Tp24711* and *Tpbd856-1852*, were successfully amplified and ready to be assembled with the vector. Attempting to assemble the fragments with the plasmid and creating fusions with the fluorescent protein mTurquoise and expressing it in *T. pseudonana*, in order to identify the locations *in vivo*, proved to be more difficult than expected. The challenge was connected to the assembly of the pTpPUC3 vector backbone containing the mTurquoise tagged genes, resulting in not creating the fusion protein. Several others have reported difficulties (Tore Brembu, personal communications), especially with PCR amplification of the vector backbone. Had there been more time these problems could have been resolved by testing more clones and more screening, or using another cloning approach. An alternative cloning approach is based on having two restriction sites adjacent to each other, one of which is cut, and then the Gibson assembly is performed. The Gibson assembly is constructed in such a way that one of the restriction sites is removed while the other remains. Following Gibson construction, a restriction is applied to the Gibson in order to remove all of the backbone without insert, leaving only the clones with the insert.

6. CONCLUSION

The goal of this thesis was to investigate two silicanin-like proteins found in *T. pseudonana* (Tp24708 and Tp24711) in order to understand their function in the diatom and how they might be involved in frustule synthesis/formation. The findings are inconclusive, but they do suggest that there is a small number of cells containing the episome, resulting in a high number of false positives. The results revealed that the episome is present in some cells, and in one case confirmed that the clone had the episome (Tp24711 PAM2 F). Although the episome was present and may be accessible to expression, there was no evidence of Cas9-induced mutation. This implies that episome-containing cells require a longer incubation time, as it is likely that the cells go through numerous cell divisions before alterations can be detected. When attempting to locate the gen edited clones and isolate the cells of interest, the limited number of cells that might express Cas9 is an impediment.

There is a possibility that the bacterial conjugation cloning approach is not working for *T. pseudonana*, and that other cloning procedures, such as biolistic transformation, are required for this species of diatom. Furthermore, the data suggests that the selection marker nat (nourseothricin resistant gene) does not provide enough stress to the cell, and hence does not provide a sufficient incentive for the episome to be preserved. The antibiotic NOU is also thought/suspected to interfere with protein production, raising the question of whether another selection marker is required to achieve Cas9-induced genome editing in *T. pseudonana* cells.

The sub-goals of creating Tp24711 and Tpbd856-1852 *T. pseudonana* cell lines expressing mTurquoise-tagged fusion proteins to determine location *in vivo* were not achieved. The most challenging part was attempting to assemble the pTpPUC3 vector backbone with the mTurquoise tagged genes. It was thus not possible to analyse their position and confirm the notion that mNeonGreen interacts with an ankyrin repeated-containing proteins in the SDV, by looking for FRET in cells transformed with both fluorescence markers.

7. FUTURE ASPECTS

To increase our understanding of the functions and features of the distinct genes, repetition of the experiment and further investigate of the *T. pseudonana* strains with single and double knockout lines in the two genes, *Tp24708* and *Tp24711*, using the CRISPR / Cas9 gene editing technique is required. As the results indicate that certain cells possessed the episome, a cell sorter may be used to quickly identify a cell containing the desired features, such as sufficient *Cas9* expression identified by YFP fluorescence. Flow cytometry can sort out cells in specified wells using a fluorescence-activated cell sorter (FACs), allowing researchers to proceed with the cells of interest. A different cloning strategy, such as biolistic transformation, as well as another selection marker, such as the bleaching herbicide PDS-norfurazon resistance-based selection system (45), should be tried to improve genome editing of *T. pseudonana* cells. Additionally, examining frustule morphology in TEM/SEM will allow a better understanding of how mutations in different genes affect the frustule.

Furthermore, continuing the ongoing investigation involving the mTurquoise tagging of the Tp24711 and Tpbd856-1852 would help to better understand the function and enable in vivo verification of the localisation. This work could be enriched by employing different cloning procedures, such as those that employ two restriction sites which are adjacent to each other, with one of them disappearing in the final product. Future studies are encouraged to continue this investigation, as it may reveal useful information on their position *in vivo* and subsequently their role in the diatom. It would also be beneficial to use FRET to investigate putative relationships in order to learn more about the exact functions of the proteins.

REFERENCE

1. Armbrust EV. The life of diatoms in the world's oceans. Nature. 2009;459:185-192.
2. Falciatore A, Bowler C. Revealing the molecular secrets of marine diatoms. Annu Rev Plant Biol. 2002;53:109-130.
3. Benoiston AS, Ibarbalz FM, Bittner L, Guidi L, Jahn O, Dutkiewicz S, et al. The evolution of diatoms and their biogeochemical functions. Philos Trans R Soc B-Biol Sci. 2017;372:10.
4. Mann DG, Vanormelingen P. An Inordinate Fondness? The Number, Distributions, and Origins of Diatom Species. J Eukaryot Microbiol. 2013;60(4):414-420.
5. Kroger N, Poulsen N. Diatoms-From Cell Wall Biogenesis to Nanotechnology. Annual Review of Genetics. Annual Review of Genetics. 42. P 2008; 42: p. 83-107.
6. Hildebrand M, Lerch SJL, Shrestha RP. Understanding Diatom Cell Wall Silicification - Moving Forward. Front Mar Sci. 2018;125.
7. Brembu T, Chauton MS, Winge P, Bones AM, Vadstein O. Dynamic responses to silicon in Thalasiossira pseudonana - Identification, characterisation and classification of signature genes and their corresponding protein motifs. Sci Rep. 2017;7: 4865
8. Sharma AK, Nymark M, Sparstad T, Bones AM, Winge P. Transgene-free genome editing in marine algae by bacterial conjugation - comparison with biolistic CRISPR/Cas9 transformation. Sci Rep. 2018;8: 14401.
9. Jamali AA, Fariba Akbari MMG, Guardia Mdl, Khosroushahi AY. Applications of Diatoms as Potential Microalgae in Nanobiotechnology. BioImpacts. 2012;2:83-89.
10. Lebeau T, Robert JM. Diatom cultivation and biotechnologically relevant products. Part II: Current and putative products. Applied Microbiology and Biotechnology. 2003;60(6):624-32.
11. Bozarth A, Maier U-G, Zauner S. Diatoms in biotechnology: modern tools and applications. Applied Microbiology and Biotechnology. 2009;82(2):195-201.
12. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, et al. The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. Science. 2004;306(5693):79-86.
13. Prihoda J, Tanaka A, de Paula WBM, Allen JF, Tirichine L, Bowler C. Chloroplast-mitochondria cross-talk in diatoms. J Exp Bot. 2012;63(4):1543-57.
14. Parker MS, Mock T, Armbrust EV. Genomic Insights into Marine Microalgae. Annual Review of Genetics. Annual Review of Genetics. 42. Palo Alto: Annual Reviews; 2008. p. 619-45.
15. Fattorini N.UWE G.Maier, U. Targeting of proteins to the cell wall of the diatom *Thalassiosira pseudonana*. Springer. 2021;1:5.
16. Smetacek V. Diatoms and the ocean carbon cycle. Protist. 1999;150(1):25-32.
17. Hildebrand M, Lerch SJL. Diatom silica biomineralization: Parallel development of approaches and understanding. Semin Cell Dev Biol. 2015;46:27-35.

18. Tesson B, Lerch SJL, Hildebrand M. Characterization of a New Protein Family Associated With the Silica Deposition Vesicle Membrane Enables Genetic Manipulation of Diatom Silica. Sci Rep. 2017;7: 13457.
19. Görlich S, Pawolski D, Zlotnikov I, Kröger N. Control of biosilica morphology and mechanical performance by the conserved diatom gene Silicanin-1. Commun Biol. 2019;2: 245.
20. Kotzsch A, Groger P, Pawolski D, Bomans PHH, Sommerdijk N, Schlierf M, et al. Silicanin-1 is a conserved diatom membrane protein involved in silica biomineralization. BMC Biol. 2017;15:2016.
21. Sheppard V, Poulsen N, Kroger N. Characterization of an Endoplasmic Reticulum-associated Silaffin Kinase from the Diatom *Thalassiosira pseudonana*. J Biol Chem. 2010;285(2):1166-76.
22. Zhang CH, Hicks GR, Raikhel NV. Plant vacuole morphology and vacuolar trafficking. Front Plant Sci. 2014;5: 476.
23. Viotti C. ER and vacuoles: never been closer. Front Plant Sci. 2014;5:20.
24. Hildebrand M, York E, Kelz JI, Davis AK, Frigeri LG, Allison DP, et al. Nanoscale control of silica morphology and three-dimensional structure during diatom cell wall formation. J Mater Res. 2006;21(10):2689-98.
25. Nymark M, Sharma AK, Sparstad T, Bones AM, Winge P. A CRISPR/Cas9 system adapted for gene editing in marine algae. Sci Rep. 2016;6(1):24951.
26. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 2014;32(4):347-55.
27. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science. 2014;346(6213):1077-1087.
28. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001;29(9): e45.
29. Freeman WM, Walker SJ, Vrana KE. Quantitative RT-PCR: Pitfalls and potential. Biotechniques. 1999;26(1):112-125.
30. Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips APR, Noddings CM, et al. Designer diatom episomes delivered by bacterial conjugation. Nat Commun. 2015;6:10: 6925.
31. Wilfinger WW, Mackey K, Chomczynski P. Effect of pH and ionic strength on the spectro-photometric assessment of nucleic acid purity. Biotechniques. 1997;22(3):474-481.
32. Reece RJ. Analysis of genes and genomes. John Wiley & Sons L, editor. University of Manchester, UK: John Wiley & Sons, Ltd; 2004.
33. Lippincott-Schwartz J, Snapp E, Kenworthy A. Studying protein dynamics in living cells. Nature Reviews Molecular Cell Biology. 2001;2(6):444-56.
34. Goedhart J, Van Weeren L, Hink MA, Vischer NOE, Jalink K, Gadella TWJ. Bright cyan fluorescent protein variants identified by fluorescence lifetime screening. Nature Methods. 2010;7(2):137-9.

35. Markwardt ML, Kremers G-J, Kraft CA, Ray K, Cranfill PJC, Wilson KA, et al. An Improved Cerulean Fluorescent Protein with Enhanced Brightness and Reduced Reversible Photoswitching. Plos One. 2011;6(3):e17896.
36. Garza E, Bielinski V. Nested Gibson Assembly v1 protocolsio. 2020. dx.doi.org/10.17504/protocols.io.bbikikcw
37. Quan JY, Tian JD. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc. 2011;6(2):242-51.
38. Harris IE. Functional studies of a family of ankyrin repeat-containing proteins in the diatom *Thalassiosira pseudonana* hypothesized to be involved in frustule morphogenesis [Master]. Master's thesis in Chemical Engineering and Biotechnology: NTNU; 2021.
39. Gao ZL, Herrera-Carrillo E, Berkhout B. Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III. Mol Ther-Nucl Acids. 2018;10:36-44.
40. Kochupurakkal BS, Iglehart JD. Nourseothricin N-Acetyl Transferase: A Positive Selection Marker for Mammalian Cells. Plos One. 2013;8(7):e68509.
41. Mendes P, Girardi E, Superti-Furga G, Kell DB. Why most transporter mutations that cause antibiotic resistance are to efflux pumps rather than to import transporters. bioRxiv. 2020. doi: https://doi.org/10.1101/2020.01.16.909507.
42. Gatignol A, Durand H, Tiraby G. Bleomycin resistance conferred by a drug-binding protein. FEBS Letters. 1988;230(1-2):171-5.
43. Zaslavskaia LA, Lippmeier JC, Kroth PG, Grossman AR, Apt KE. Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a variety of selectable marker and reporter genes. J Phycol. 2001;36(2):379-86.
44. Poulsen N, Chesley PM, Kröger N. molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (bacillariophyceae). J Phycol. 2006;42(5):1059-65.
45. Taparia Y, Zarka A, Leu S, Zarivach R, Boussiba S, Khozin-Goldberg I. A novel endogenous selection marker for the diatom Phaeodactylum tricornutum based on a unique mutation in phytoene desaturase 1. Sci Rep. 2019;9

Appendix A: Additional gel images

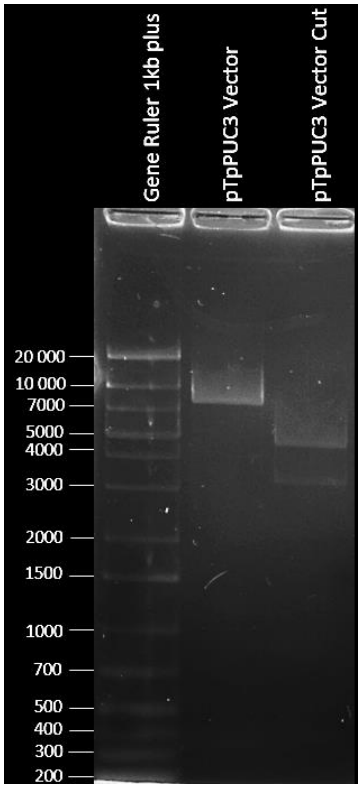

Figure A. 1: Gel image of pTpPUC3 cut vector for 2h with both PstI-HF and SAC1, the whole plasmid = 7 796, the fragment from the plasmid cut with two restriction enzyme; PstI-HF = 4 596 and SAC1 = 3 200.

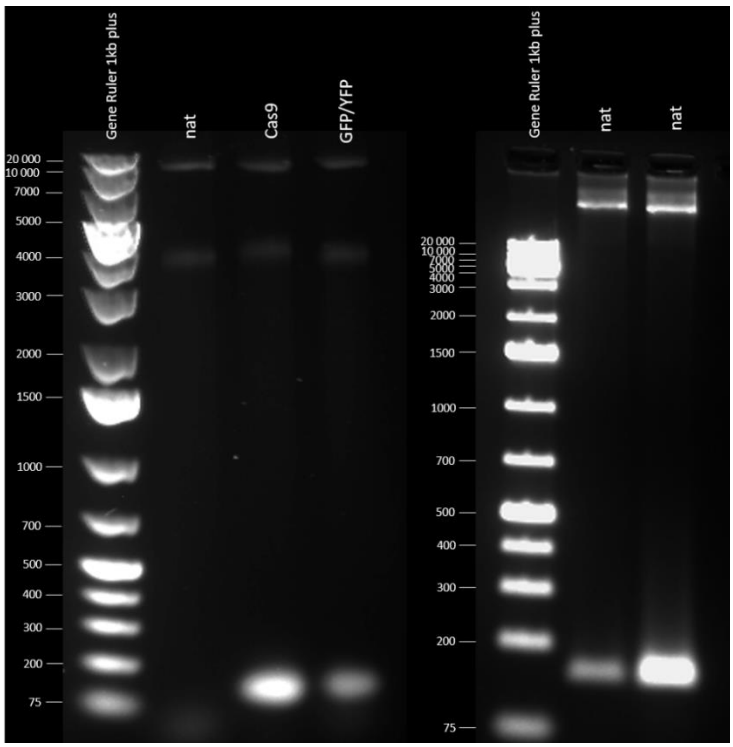

Figure A. 2: Gel images of primers used in the qRT-PCR. Nat (124 bp), Cas9 (93 bp) and GFP/YFP (94 bp).

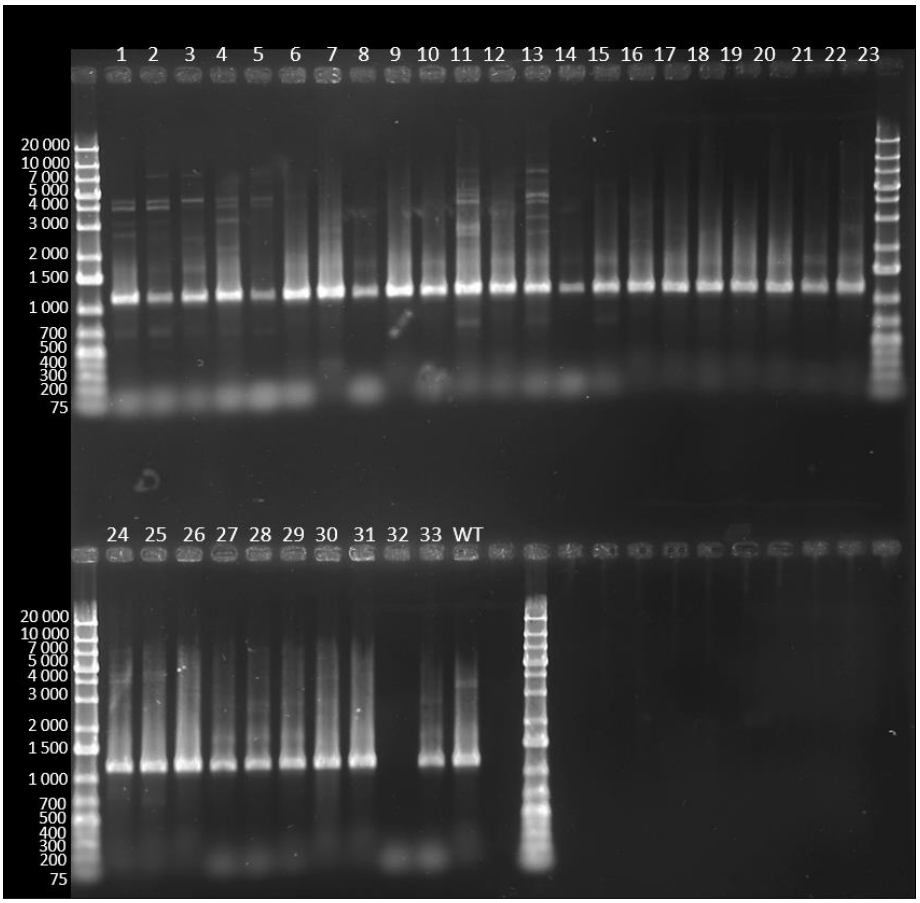

Figure A. 3: Tp24708 colony screening plate 2 new conjugation (Gene Ruler 1kb plus)

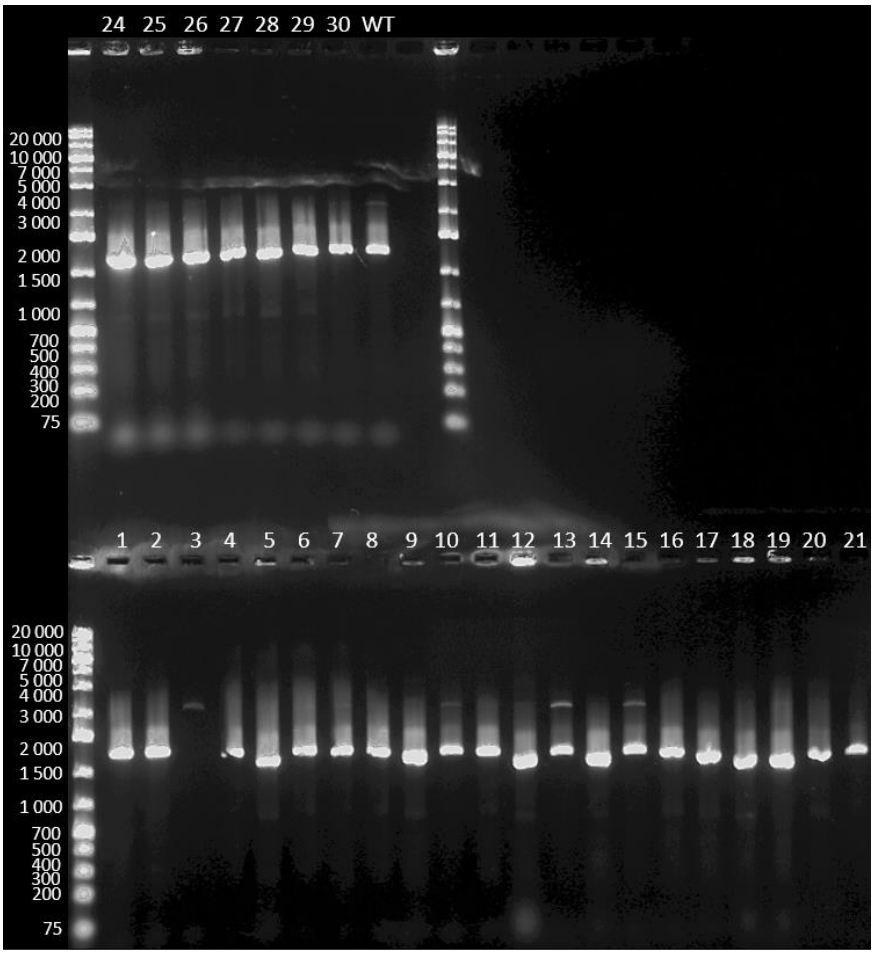

Figure A. 4: Tp24711 colony screening plate 2 new conjugation (Gene Ruler 1kb plus)

Appendix B: Additional fluorescence microscopy images

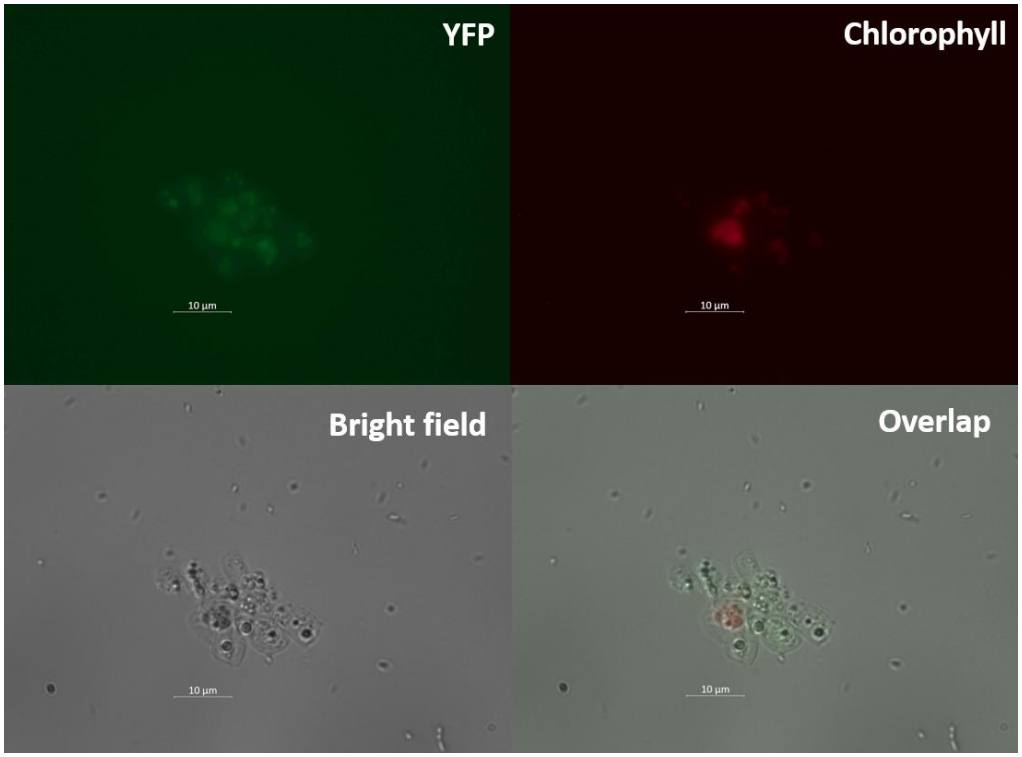

Figure B. 1: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24708 PAM 1 C. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

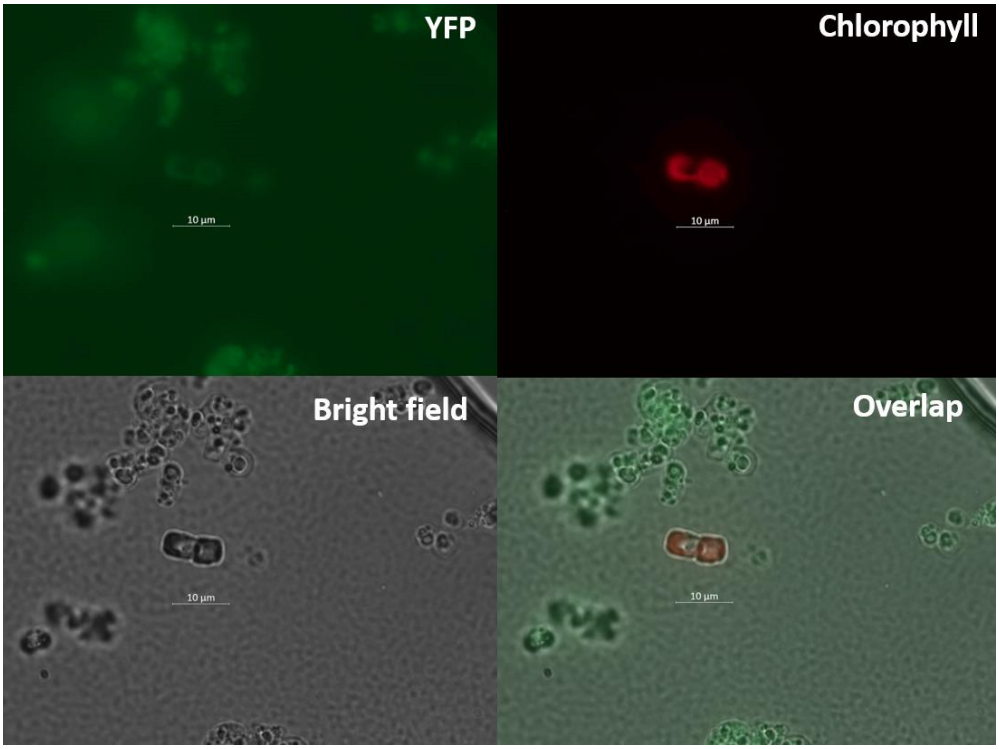

Figure B. 2: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24711 PAM 1 D. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

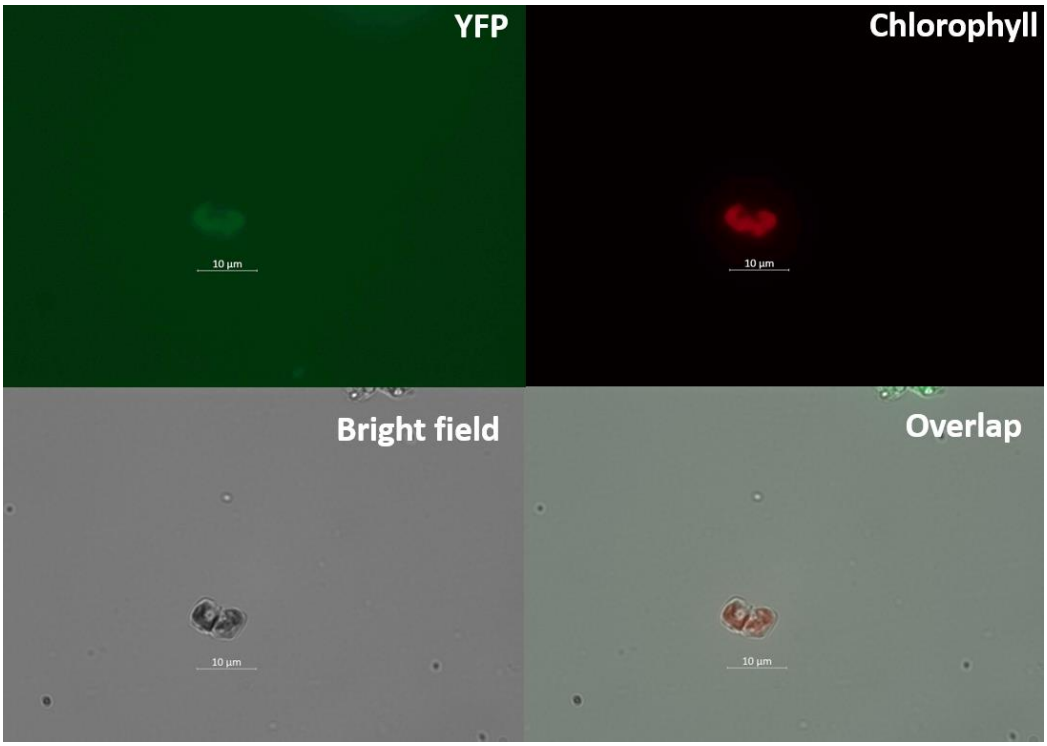

Figure B. 3: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24711 PAM 1 D. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

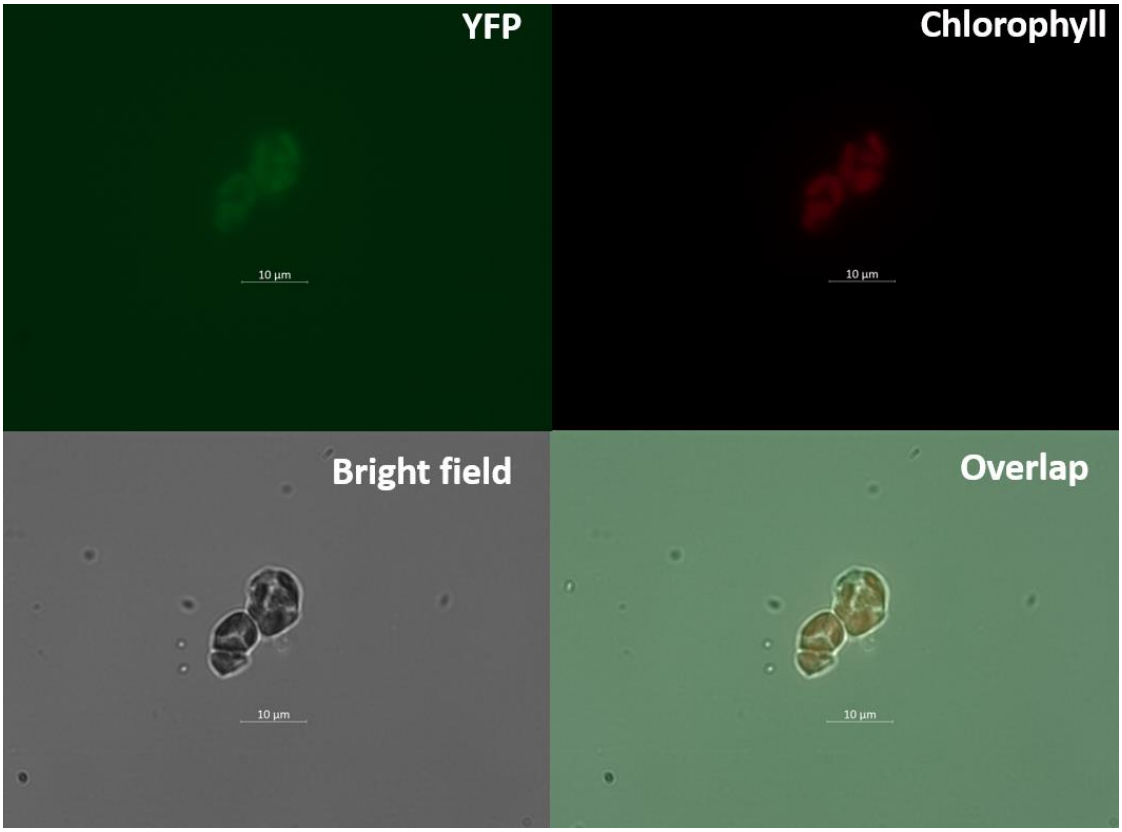

Figure B. 4: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24711 PAM 1 D. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

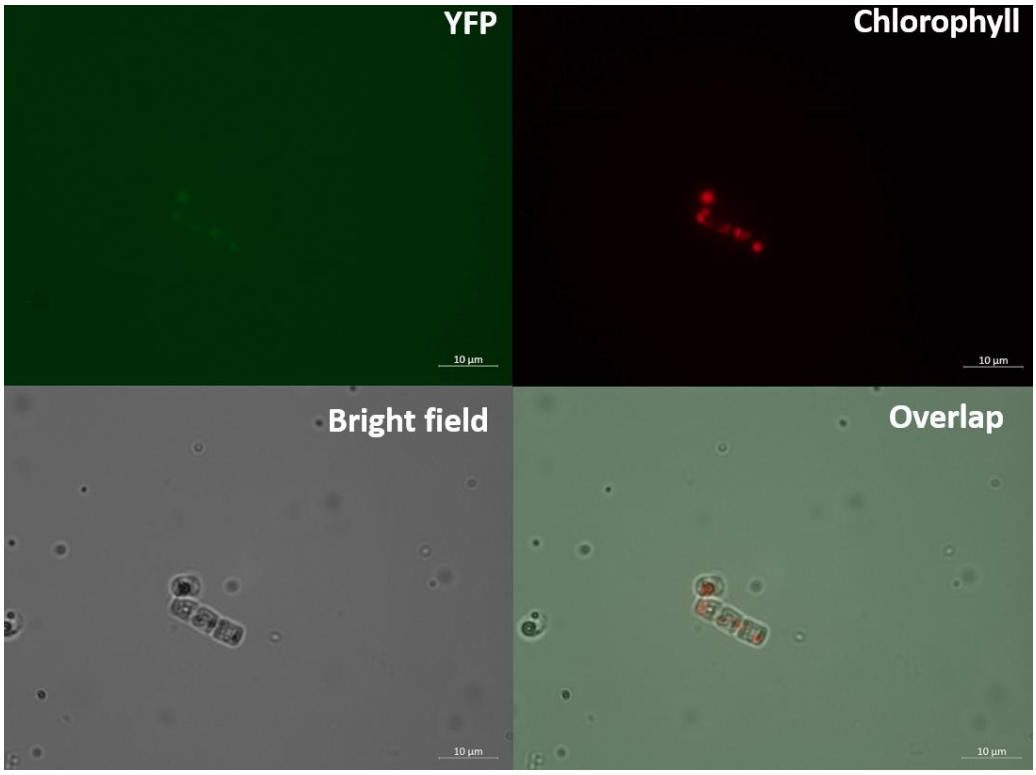

Figure B. 5: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24711 PAM 1 I. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

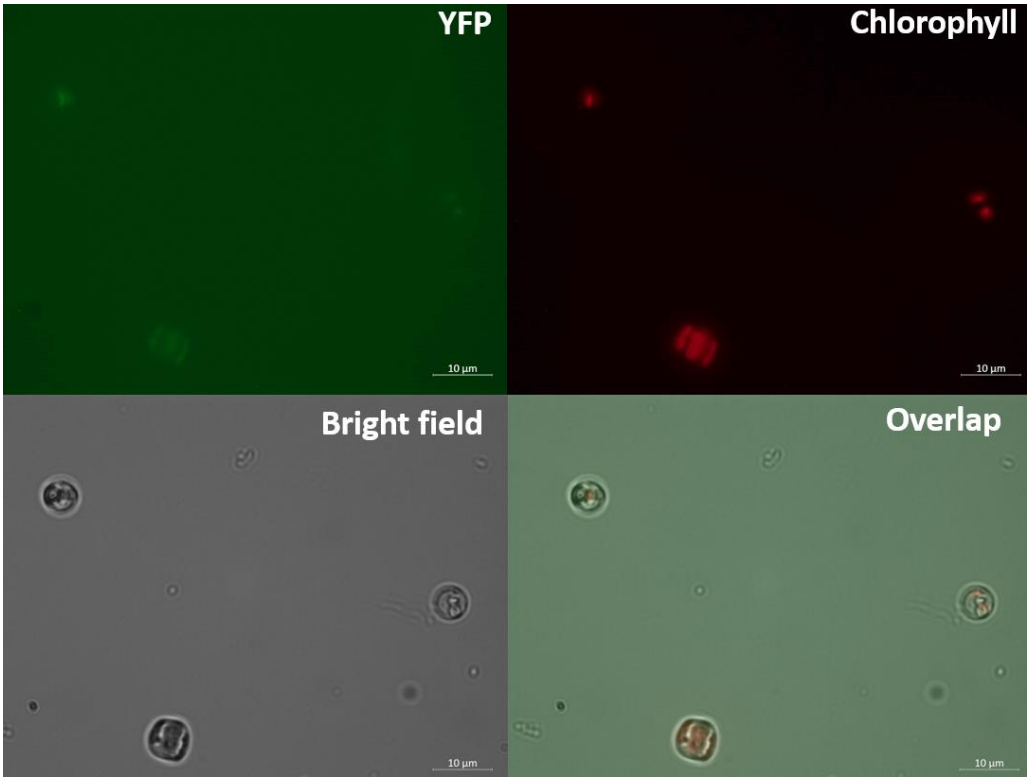

Figure B. 6: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24711 PAM 1 I. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

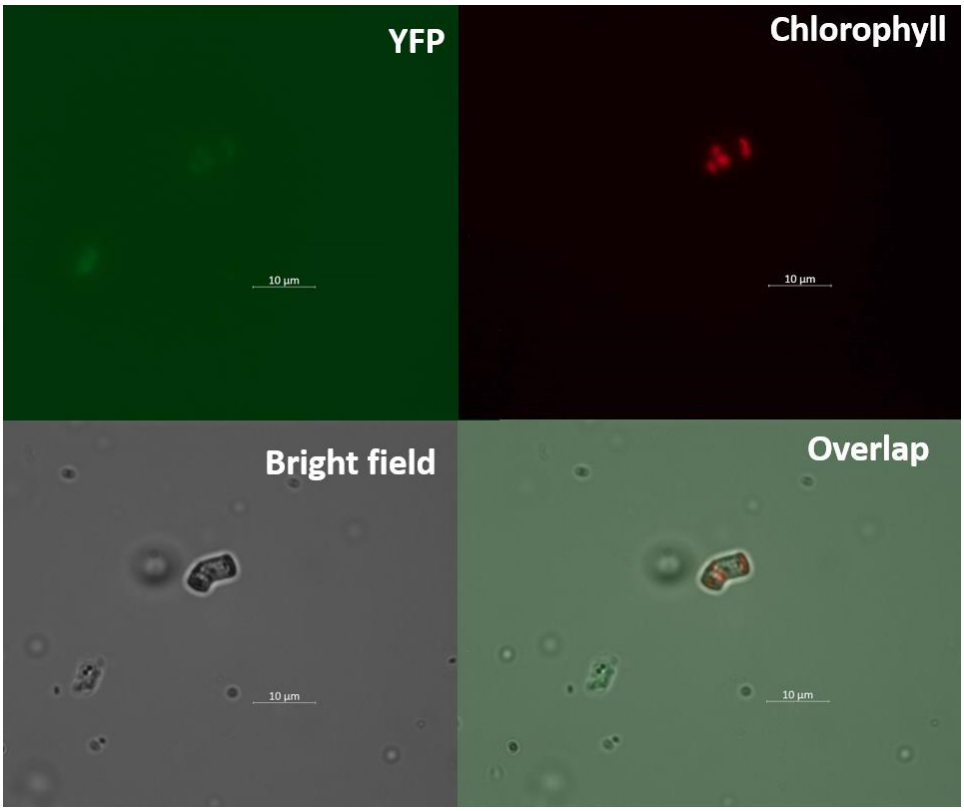

Figure B. 7: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24711 PAM 1 J. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

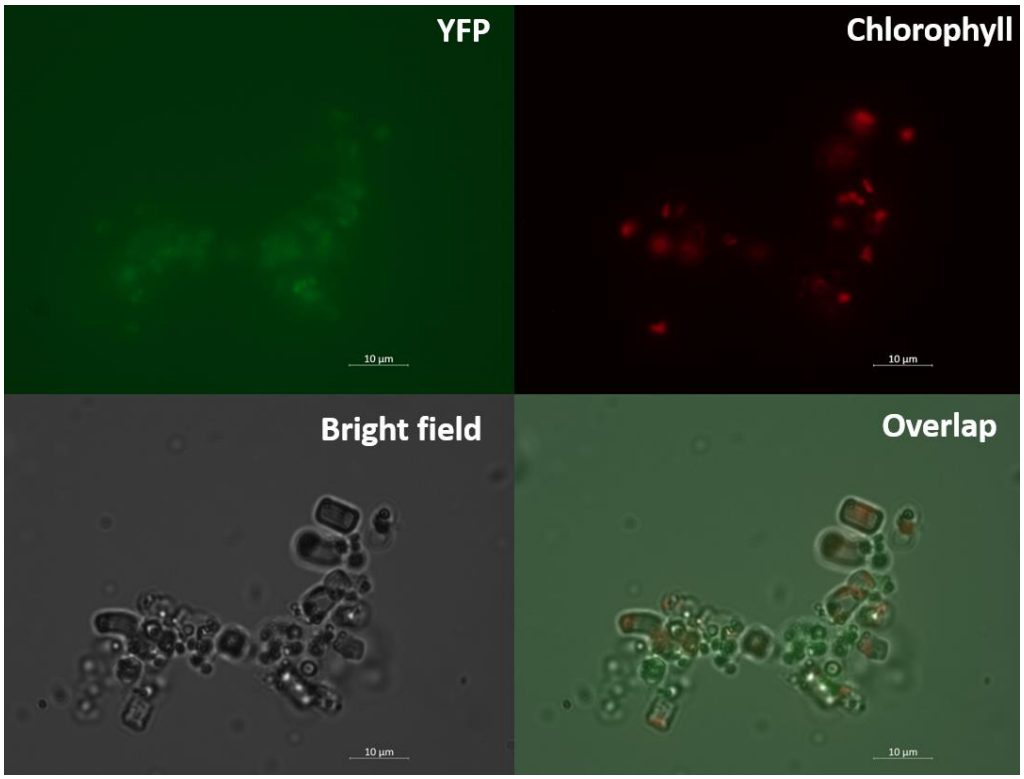

Figure B. 8: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24708 PAM 2 B. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

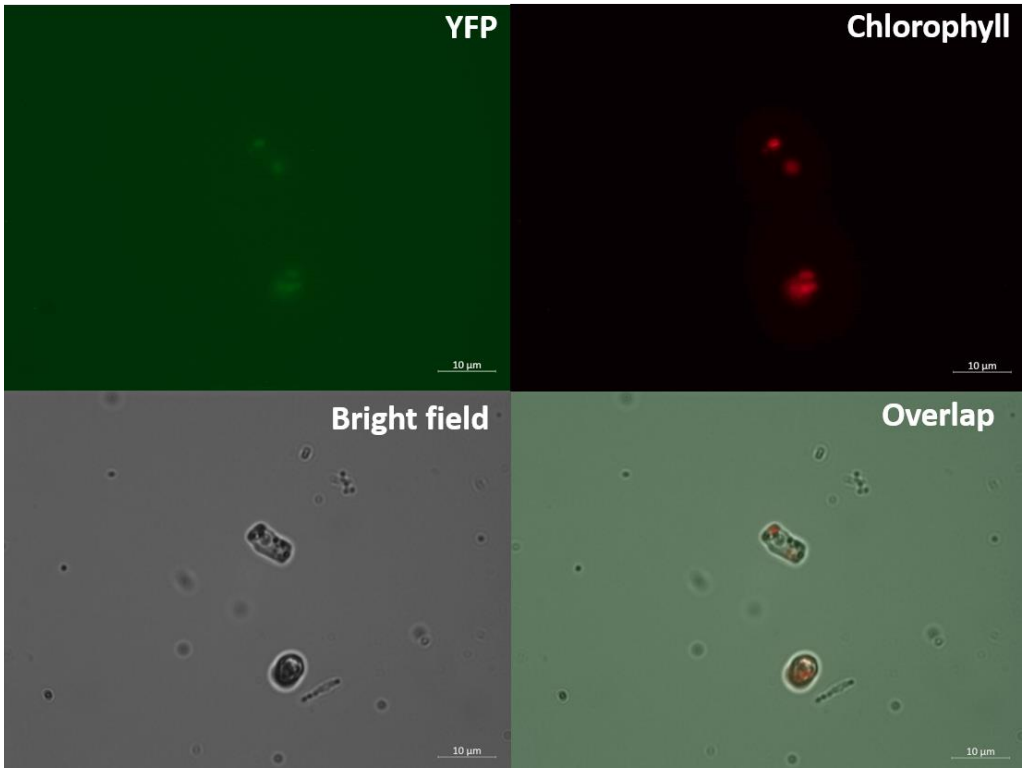

Figure B. 9: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24711 PAM 1 J. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

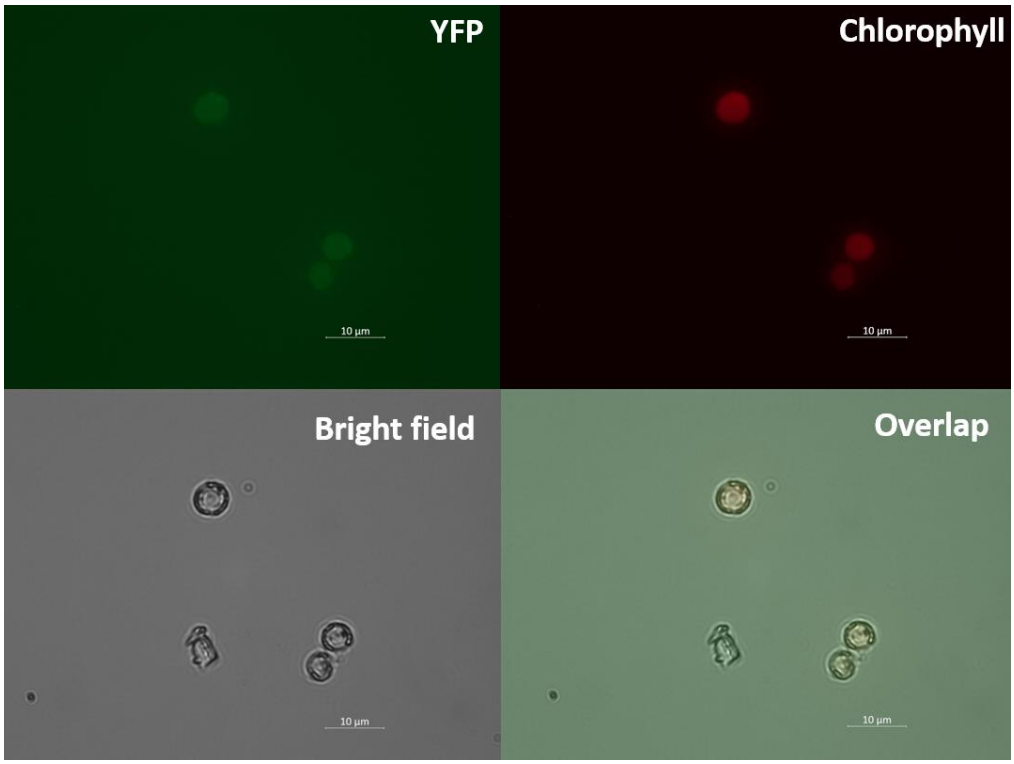

Figure B. 10: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing WT. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

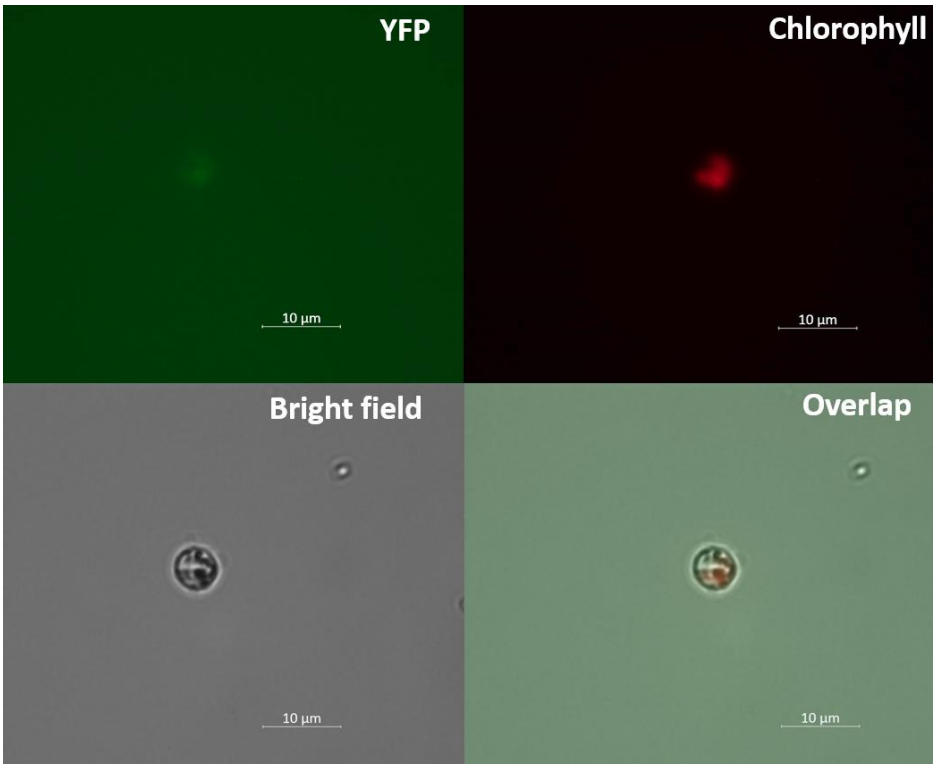

Figure B. 11: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing WT. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

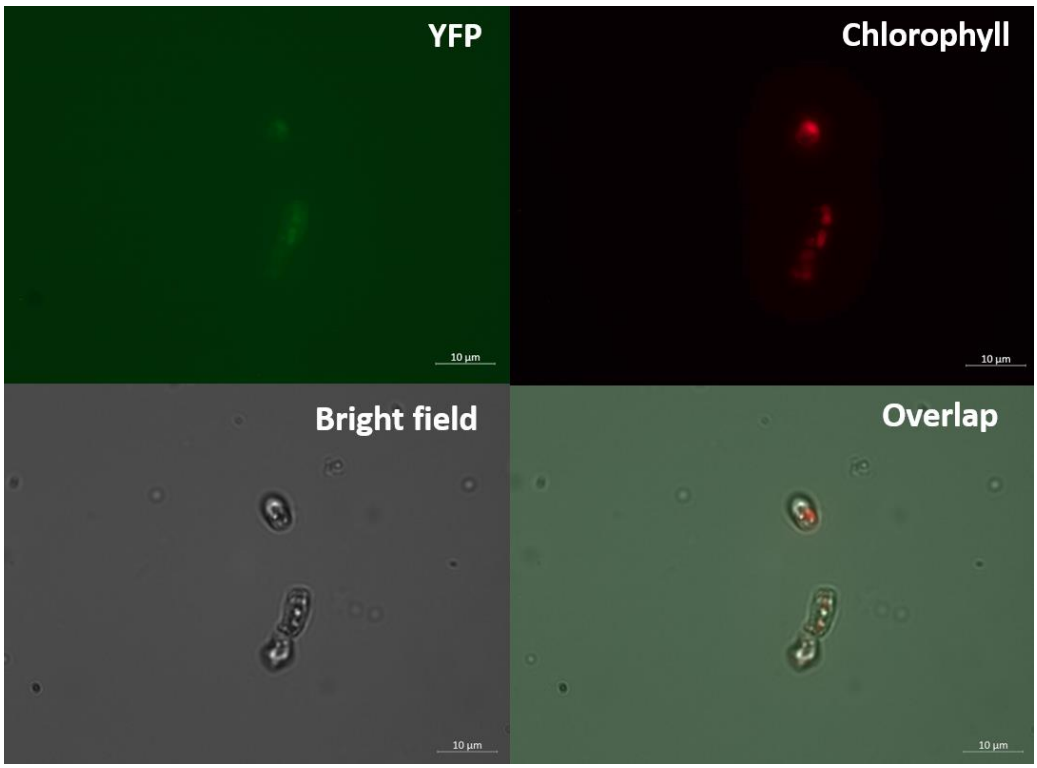

Figure B. 12: Fluorescence microscopy image of *Thalassiosira pseudonana* cell expressing Tp24711 PAM 1 J. Upper left: YFP fluorescence (green); upper right: chlorophyll autofluorescence (red) from chloroplasts; lower left: bright field image; lower right: overlap all channels. Scale bars are 10 µm.

Appendix C: Additional data

Table C. 1: NanoDrop - The amount (µL) used to prepare for qRT-PCR was calculated from the concentration obtained with Nanodrop, and the cell selected for the -RT, were the one which exhibited the highest concentration (Tp24708 P1 F).

Sample	ng/µL	µL (=312ng)	A260/A280	A260/A230
Tp24708 PAM 1 C	96,9	9,6	2,14	0,58
Tp24708 PAM 1 F	144,4	3,27	2,14	1,99
Tp24708 PAM 2 B	84,2	12	2,13	1,74
Tp24708 PAM 2 F	100,2	8,67	2,08	0,10
Tp24711 PAM 1 D	32,5	3,22	2,4	1,33
Tp24711 PAM 1 G	95,3	3,24	2,13	0,42
Tp24711 PAM 1 I	26,0	3,71	2,25	0,04
Tp24711 PAM 1 J	36,0	3,11	2,05	1,02
WT	105,6	2,96	2,04	0,10

D: Culture media and solutions

Table D. 1: Liquid Luria-Bertani medium (LB-medium). Components below were added to 1L distilled water, and autoclaved for 20 minutes at 120 °C. For agar plates, antibiotics were added after the solution was cooled to below 55 °C.

Components	Amount (g/L dH2O)	Suppliers
Trypton	10	VWR life science
yeast extract	5	oxoid, LP0021
NaCl	5	Sigma life science
agar (only for agar plates)	15	oxiod

Table D. 2: F/2 medium recipe. seawater was autoclaved (20 minutes, 121 °C), Sterile filtered (0.2 µm) nutrients were added. Quantities given below are for 1 L F/2 medium.

Stocks		Per liter
(1)	Trace elements	
	Na2EDTA	4,36 g
	FeCl2 * 6H2O	3.15 g
	CuSO4 * 5H2O	0,01 g
	ZnSO4 *7H20	0,022 g
	CoCl2 * 6H20	0,01 g
	MnCl2 * 4H2O	0,18 g
	Na2MO4 * 2H2O	0,006 g
(2)	Vitamin mix	
	Cyanocobalamin (Vitamin B12)	0,0005 g
	Thiamine HCL (Vitamin B1)	0,1 g
	Biotin	0,0005 g
(3)	Sodium metasilicate	
	Na2SIO2 * 9H2O	30,0 g
Medium		Per liter
	NaNO3	0,075 g
	NaH2PO4 * 2H2O	0,00565 g
	Trace element stock solution (1)	1,0 mL
	Vitamin mix solution (2)	1,0 mL
	Sodium metasilicate stock solution (3)	1,0 mL
	Agar (only for agar plates)	15 g/L

Table D. 3: SOC medium. Components below were added to distilled water.

	Amount	Suppliers
Trypton	20 g/L	vwr life science
Glucose	3.6 g/L	Life Technologies AS (Invitrogen Dynal AS)
MnCl2 × 2H2O	5.08 g/L	
KCl	2.5 mM	

Table D. 4: Lysis buffer for *T. pseudonana* cells. Sterile filtered (0.2 µm) compounds were added distilled water.

Lysis buffer	Concentration
tripton x-100	10 %
Tris-HCl ph 8	20 mM
EDTA	10 mM

Table D. 5: Components used to make 1 X TEA buffer.

	Amount per L
Tris-base	4.84 g
glacial acetic acid	1.142 mL
EDTA	0.0012 M

Appendix E: Various compounds and instruments

Table E. 1: Primers used for sanger sequencing and for colony screening.

Primer name	Orientation	Sequence (5`-3`)
M13-rev	Reverse	CAGGAAACAGCTATGAC
M13-F	Forward	TGTAAAACGACGGCCAGT
CEN-ARS-R 100	Reverse	TGTGGTCTTCTACACAGACA
U6_Mock F	Reverse	CTGCTCCAGTTCTCCCTTCATCAAGAGAGCAACCAACA
Cas9M_F1	Forward	GGCTCGATATCGGCACAAAC
Cas9M_R5	Reverse	CCGTCCAGCTCGACCAG
His3SeqF	Forward	TGTTCCCTCCACCAAAG

Table E. 2: A list of various compounds used during the experiment as well as the supplier.

Compound	Suppliers
CUT smart buffer	New England Biolabs, B7204s
T4 ligase buffer,10X buffer for t4 DNA ligase with 10 mM ATP	New England Biolabs, M0202A
GelRed nucleic acid stain, 10 000 x in water, 5 µl / 100 mL	Biotium
Loading dye, 6x DNA loading dye	Thermo Scientific, R0611
Gene ruler 1Kb DNA ladder and 1kbplus	Thermo scientific
dNTPs	VWR Life Science
SeaKem LE agarose	Lanza
2-Mercaptoethanol	Aldrich

Table E. 3: A list of various enzymes used during the experiment as well as the supplier.

Enzyme	Supplier
T4 ligase	New England Biolabs, M0202S
1 x BsaI-HF v2	New England Biolabs, R3733S
PstI-HF	New England Biolabs, R3140S
SacI	New England Biolabs, R3138S

Table E. 4: A list of various Antibiotics stock used during the experiment as well as the supplier.

Antibiotic	Suppliers
kanamycin sulfate biochemica	PanReac Applichem
Gentamicin	Gibco, life technology limited (ref: 15750-037)
Nourseothricin	Jena Biosciences

Table E. 5: A list of various Strains used during the experiment as well as the supplier.

Microorganism	Strain:
Ecoli	DH5α
	DH10β
	Stable
Thalassiosira pseudonana	CCMP1335

Table E. 6: A list of various Vectors used during the experiment as well as the supplier.

Vectors	Suppliers
pTpPUC3	AddGene (Plasmid 62864) 31
pNCS-mNeonGreen	Allele biotechnology
pBKS-mTurquoise AddGene	(plasmid 98886)

Table E. 7: A list of various kits used during the experiment as well as the supplier.

Kit	Manufacturer
Phusion hot start II High-fidelity PCR master mix	Thermo Scientific
Phusion hot start II High-fidelity PCR DNA polymerase	Thermo scientific
High Resolution Melting Master, Light Cycler 480	Rocher, version 07
Master mix (rød) redtaq 2X Master mix 1,5mM MgCl2	VWR life science
Exs-pure enzymatic pcr purification kit	NimaGen
GeneJET plasmid Miniprep kit	Thermo scientific
LightCycler 480 SYBR Green I Master Kit	Roche

Table E. 8: A list of various Instruments used during the experiment as well as the supplier.

Instrument	Suppliers
Microcentrifuge	VWR, himac, CT15E
Centrifuge	Thermo scientific Heraeus multifuge, x1R centrifuge
Flow-cytometer	NovoCyte™ flow cytometer ACEA Biosciences
Thermo-Shaker	Grant-bio, version: V.4GW (PSC24N)
Gel electrophoresis	Consort EV1450 & Consort EV261
Gel tray	thermo scientific OWL easycast B2/B1/B1A
G:Box gel documentation system	Syngene
PCR	Biorad t100 Thermo cycler
Incubator with shaking	Multitron, INFORS AG CH-4103 BOTTMINGEN
	INFORS HT Ecotron INFORS AG CH-4103 BOTTMINGEN
Fluorescence microscopy	Zeiss Axio Imager Z2 fluorescence microscope
Cover glass	VWR
TissueLyser	QIAGEN
Nano drop	NanoDrop™ 1000 Spectrophotometer, thermo scientific
LightCycler 96	Roche
Multiwell plate 96	LightCycler 480, Roche
Sealing Foil	LightCycler 480, Roche

NTNU

Norwegian University of
Science and Technology