NTNU – Trondheim
Norwegian University of
Science and Technology

Studies of light responses and the development of a transformation system for the benthic diatom *Seminavis robusta*

Darshan Kirupamurthy

Biotechnology
Submission date: June 2014
Supervisor: Atle M. Bones, IBI
Co-supervisor: Per Winge, IBI

Norwegian University of Science and Technology
Department of Biology

Acknowledgments

This master's thesis was carried out at the Norwegian University of Science and Technology (NTNU) at the Department of Biology.

I would first of all wish to thank my supervisors Atle Magnar Bones and Per Winge. Atle for letting me be a part of his research group, the Cell and molecular biology group (aka the Bones group), and at the same time introducing me to the magnificent world of cell and molecular biology. I am truly honored to have had Per as my academic supervisor. I am grateful for the encouragements, guidance and patience he has shown me throughout my master's thesis. I'm also thankful for have gotten the chance to learn from Pers' enormous chamber of knowledge which must be deeper than Mimirs' well. I would also like to thank researcher Ralph Kissen for taking his time to help me in the lab every time I asked him for any assistance. He truly is a lab-guru, with a solution to all questions. I am grateful for the help that the principal engineer Torfinn Sparstad has done for me, helping to prepare experiments and for making the time spent in the lab fun. He has also been my fellow "*groruddøling*" in the Bones group. I am grateful to the rest of the Cell and Molecular biology group as well, for encouraging imputes and motivation, especially Tore Brembu, Leila Alipanah and Marianne Nymark for helping me give my diatoms proper care.

There are also a lot of people outside of the academic community I would like to thank. I am really thankful towards my uncle Koneswaran and auntie Saratha for inviting me for dinner every Friday and treating me like family. I will never forget your generosity. I consider myself lucky for all the support and kindness my friends and fellow master students have shown throughout these two years, the good, bad and the ugly times.

Last but not least, I would like to thank my family. I am forever grateful for the support and encouragements my mom and dad have given me during my student days throughout the ups and downs. I would also like to thank my sister and brother-in-law for all the moral support, kindness and understanding they have shown me while I was working with this master's thesis. I owe my sister a special gratitude for motivating me and convincing me to carry on during the hard periods when everything seemed like a never ending story. I would at the end like to thank my little niece, Tansi. Even though she may not be aware of it, her laughs and happy presence lightens up the days in a way no one else can.

Trondheim, 13th of June, 2014

Darshan Kirupamurthy

"What you have learned is a mere handful; What you haven't learned is the size of the world"
- Avvaiyar

Abstract

Diatoms are a divergent group of organisms with a complex evolutionary background, and belong to the Heterokonta group phylogenetically. Diatoms have gone through several endosymbiotic processes with both green algae and red algae, and have acquired a complex genome. Some of the genes and protein found in diatoms are similar to those found in plants and green algae, while others show similarities to animals (metazoa), yeast, bacteria and other unicellular microorganisms. This genomic diversity could be one of the reasons why diatoms dominate the primary production in many marine and fresh water ecosystems.

Seminavis robusta is a benthic diatom that lives on the bottom sediments and has a bilateral symmetry. One of the advantages with *S. robusta* is the size; it is bigger than both *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* which have been fully genome sequenced. *S. robusta* also have two mating types, which means that it is possible to do breeding between individuals and keep a pedigree. *S. robusta* has therefore been considered a good model organism for studies on benthic diatoms.

In this study two experiments were done on *S. robusta.* The first experiment was to make a vector construct that could be used to transform *S. robusta.* The endogenous AtpBE/D promoter in *S. robusta* expressing the ATP synthase beta subunit and ATP synthase epsilon or delta (in this thesis referred to as epsilon) subunit was chosen as promoter for the construct. The AtpBE promoter was chosen because of its small size and because it is bidirectional and can express two genes simultaneously. The AtpBE promoter and the belonging AtpB and AtpE terminators were isolated from genomic DNA from *S. robusta* and cloned into the pBluescript (KS+) vector. The *nourseothricin acetyltransferase* (*nat1*) gene was chosen as selectable maker. *Nat1* gives resistance to nourseothricin, the gene was modified and synthetically synthesized to correspond to the tRNA abundance of *S. robusta*. The *nat1* gene was cloned into the AtpE side of the promoter while the gene coding for Yellow Fluorescence Protein (*YFP*) was cloned into the AtpB side of the promoter. *YFP* gene was included to check the expression capability of the AtpB promoter direction. Results from transformation attempts showed that only *nat1* was expressed and translated in the cells. No YFP or YFP mRNA was detected, indicating that the gene was not transcribed.

The second experiment performed on *S. robusta* was a light experiment, where cells in exponential growth phase were kept in dark for 12 hours before eight flasks were exposed to blue light (BL, 80 µmol photons $m^{-2}s^{-1}$), four flasks were set for 0,5 hour and the other four for 6 hours. After the exposure, cells were harvested and RNA was isolated. The same procedure was done in white light (WL, 100 µmol photons $m^{-2}s^{-1}$) for 0,5 and 6 hours. The gene expressions from the treatments were studied with RT- qPCR and microarray technology. The study showed that some of the genes coding for Light Harvesting Complex proteins (LHCs), Aureochromes, Aureochrome-like proteins and Heat Shock Transcriptions factors (HSFs) had strong responses to the light treatments.

Sammendrag

Kiselalger (Diatomer) er en divergent gruppe av organismer med en kompleks evolusjon og er fylogenetisk plassert blant Heterokonta. Kiselalgene har gjennomgått en rekke endosymbiotiske prosesser, med både grønnalger og rødalger. Dette har ført til at de har utviklet et komplekst genom. Noen av genene og proteinene som finnes hos kiselalger ligner de man finner hos planter og grønnalger, mens andre har mer likhetstrekk med de hos dyr (metazoa), gjær, bakterier og andre encellede mikroorganismer. Denne genomiske diversiteten kan være en av årsakene til at kiselalger dominerer primærproduksjonen i mange marine- og ferskvanns- økosystemer.

Seminavis robusta er en bentisk diatom som lever i bunnsjiktet og har en bilateral symmetri. En av fordelene med å jobbe med *S. robusta* er størrelsen, den er større enn både *Phaeodactylum tricornutum* og *Thalassiosira pseudonana* som begge har blitt genomkarakterisert. *S. robusta* har to "mating" typer, noe som betyr at det er mulig å utføre kryssing mellom individer og ha en oversikt over kryssingene med en slektstavle. *S. robusta* har på grunn av disse egenskapene blitt ansett som er god modellorganisme for studier av bentiske kiselalger.

I denne oppgaven ble to forsøk utført på *S. robusta*. Det første forsøket gikk ut på å lage et vektorkonstrukt som kunne brukes til å transformere *S. robusta.* Den endogene AtpBE/D promotoren fra *S. robusta*, uttrykker ATP beta og epsilon eller delta (i denne oppgaven har man valgt å kalle det epsilon) subenhetene. Promotoren ble valgt fordi den var kort og kompakt, i tillegg er den bidireksjonell og kan derved utrykke begge genene. Både promotoren og de tilhørende terminatorsekvensene AtpEt og AtpBt ble amplifisert fra genomisk DNA isolert fra *S. robusta.* De amplifiserte sekvensene ble deretter klonet inn i en pBluescript (KS+) vektor. Genet *nourseothricin acetyltransferase* (*nat1*) ble brukt som seleksjonsmarkør. *Nat1* koder for resistens mot antibiotikumet nourseothricin. *Nat1* genet ble modifisert og produsert syntetisk slik at kodon sammensetningen i genet var tilpasset tRNA mengden i *S. robusta. Nat1* ble klonet inn vektoren slik at AtpE-retningen av bidireksjonelle promotoren styrte utrykket av genet. Genet som koder for "Yellow Fluorescence Protein" (*YFP*) ble satt inn på AtpB-siden av promotoren for å sjekke AtpB promotoren sin evne til å utrykke gener. Resultater fra forsøket viste at *nat1* ble uttrykt i cellene men transkripsjon eller translasjon av YFP ble ikke detektert.

Forsøk nummer to var et lysforsøk som ble utført på *S. robusta.* Celler i eksponentiell vekstfase ble først satt i mørke for 12 timer. Deretter ble åtte flasker med celler eksponert for blått lys (80 µmol fotoner $m^{-2}s^{-1}$), fire av flaskene ble eksponert for BL i en halv time mens de resterende fire ble eksponert i seks timer. Etter eksponeringen ble cellene høstet og RNA ble isolert. Celler ble også eksponert for hvitt lys (100 µmol fotoner $m^{-2}s^{-1}$) med samme fremgangsmåte. Genuttrykket til de behandlede cellene ble undersøkt ved å bruke kvantitativ PCR og DNA mikromatrise-chip teknologi. Forsøket viste at gener for Lyshøstningskompleksproteiner (LHCs), Aureochromer, Aureochrome-lignende proteiner og

varmesjokks-transkripsjonsfaktorer (HSFs) var blant de genene som hadde sterkest respons etter de ulike lysbehandlingene.

TABLE OF CONTENTS

Abbreviations

ACP	Acyl-acyl carrier protein
Amp	Ampicillin
AtpBt	AtpB terminator
AtpEt	AtpE terminator
AUREO1a	AUREOCHROME1a
BL	Blue Light
Bp	Base pairs
bZIP domain	Basic region leucine zipper domain
C	Carbon
^{0}C	Celsius
cRNA	Complementary RNA
Ct	Threshold cycle
CYP	Cytochromes p450
D9	Δ9 desaturase
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dsCYC2	diatom-specific cyclin 2
dsDNA	Double stranded DNA
ER	Endoplasmic Reticulum
EtBr	Ethidium Bromide
FA	Formaldehyde
FCP3	Fucoxanthin chlorophyll protein 3
FSW	Fresh Sea Water
gDNA	Genomic DNA
GFP	Green fluorescent protein
HSE	Heat Shock Element
HSF	Heat Shock transcription Factors
HSP	Heat Shock Proteins
InDel	Insertion and Deletion
LHC	Light Harvesting Complexe
LOV domain	light-, oxygen or voltage domain
Ma	Million years
mL	Milliliter
mRNA	Messenger RNA
MS	Malate synthase
µg	Mikrogram
µL	Microliter
µmol	Micromole
ng	Nanogram
NTC	Nourseothricin
PAE laboratory	Protistology and Aquatic Ecology laboratory

PAR	Photosynthetically available radiation
PCR	Polymerase Chain Reaction
PEP	Phosphoenolpyruvate
PK2	Pyruvate Kinase 2
qPCR	Real-time quantitative PCR
RIN	RNA Integrity Number
RNA	Ribonucleic acid
RNase	Ribonuclease
rRNA	Ribosomal RNA
RT	Reverse-transcription
-RT sample	No Reverse Transcription sample
RT-PCR	reverse transcriptase PCR
Si	Silicon
SNP	Single Nucleotide Polymorphism
TAE	Tris-acetate EDTA
VSP35	Vacuolar Sorting Protein 35
WL	White Light
YFP	Yellow Florescence Protein

1 Introduction

1.1 Diatoms

The name diatom comes from the Greek word *diatomas*, which means cut in half, as a reference to their distinctive cell wall (Armbrust, 2009). The diatoms, belonging to the phylum Bacillariophyceae are eukaryotic unicellular aqueous organisms that are a part of the division Heterokontophyta (Stramenopiles) (Chepurnov et al., 2008). There is a big diversity in the diatoms. They can be found in freshwater, marine and even in some terrestrial environments (Round et al., 1990; Apt et al., 1996; Chepurnov et al., 2008). In the oceans they can be found as deep as the photosynthetically available radiation (PAR) can penetrate (Falciatore and Bowler, 2002). What makes diatoms characteristic is their cell wall structure, which is made up by hydrated silica (SiO_2nH_2O) (Du Buf and Bayer, 2002). This silica wall is called a frustule (Falciatore and Bowler, 2002), and every diatom specie has its own characteristic frustule with pores, ridges spikes and channels that make them separable from one another (Hasle and Fryxell, 1970; Losic et al., 2006). It is believed that the frustules work as a protection against zooplankton predation (Nymark, 2013). The frustules are also used to classify the diatoms into two major groups: centric and pennate diatoms. The centric diatoms have frustules that are radially symmetric, while the pennate diatoms have more elongated, bilaterally asymmetric frustules (Falciatore and Bowler, 2002). Normally the centric diatoms tend to be planktonic whereas many of the pennate diatoms are benthic, and can live on surfaces like sediments. Diatoms are important in the oceans because of their ability to recycle elements like carbon (C) and Silicon (Si). When looking at the oceanic primary production, the diatoms can stand for as much as 40 – 45% of the carbon fixation (Mann, 1999; Falciatore and Bowler, 2002; Nymark et al., 2009).

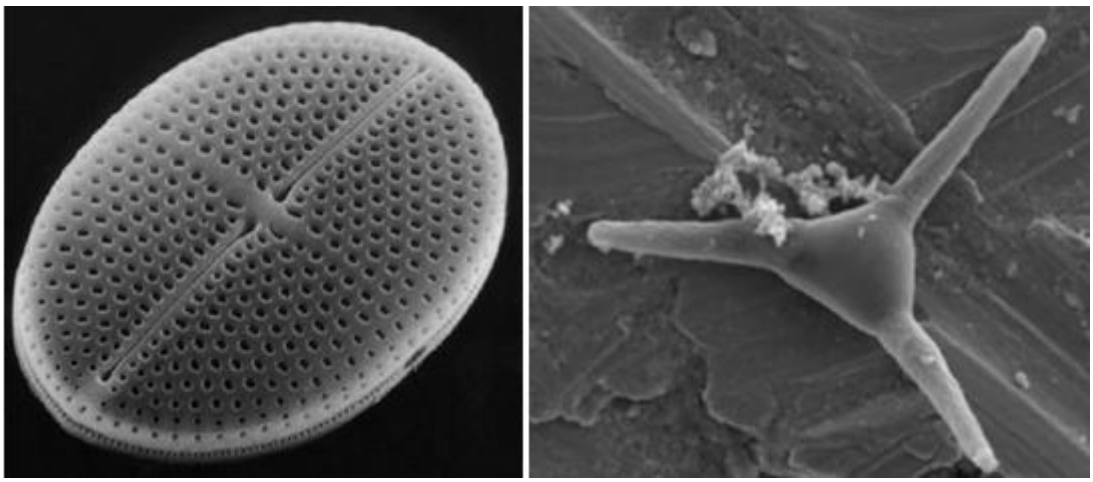

Figure 1: Electron micrograph of the diatom *Mastogloia binotata* (Left) (Falciatore and Bowler, 2002) *Phaeodactylum tricornutum* (Right) (Francius et al., 2008)

1.1.1 Origin

Compared to the photosynthetic organisms that are found on land, the diatoms have a complex evolutionary history (Armbrust, 2009). By looking at fossil records of phytoplankton one can try to predict the evolutionary history of the specie (Falkowski et al., 2004). The origin of the diatoms is little known. This is because the frustules (silica wall) that surrounds the diatoms easily dissolves in marine sediments (Falkowski et al., 2004). Estimates that have been done, which included phylogenetic analyses of ribosomal genes (Falkowski et al., 2004), showed that the diatoms had an average age at about 165 million years (Ma) (Medlin and Kaczmarska, 2004). During this time they have diverged into a many different species (Chepurnov et al., 2008) and could be found throughout the oceans and even in some non-marine environments (Falkowski et al., 2004). Today the vast majority of marine phytoplankton is cyanobacteria, followed by three eukaryotic phytoplankton, diatoms, dinoflagellates and coccolithophores. All the eukaryotic phytoplankton are believed to have originated through secondary or tertiary symbiotic processes (Falkowski et al., 2004).

The diatoms, which are photosynthetic heterokonts are evolutionary, fundamentally different from higher photosynthetic plants that dominate land (Armbrust et al., 2004). Green, red, glaucophyte algae and higher land plants are all believed to have evolved though primary endosymbiotic events where a nonphotosynthetic eukaryote engulfed or got invaded by a prokaryotic cyanobacterium and hence acquired a chloroplast (Armbrust et al., 2004). This event gave rise to two major plastid lineages, the chloroplasts and rhodoplasts (Falkowski et al., 2004). Eukaryotic phytoplankton like diatoms and haptophytes are thought to have arisen from secondary endosymbiosis, where a nonphotosynthetic eukaryote engulfed a red algae (rhodophytes), a photosynthetic eukaryote (Armbrust et al., 2004; Falkowski et al., 2004). Through this endosymbiotic event the cell acquired chloroplasts, and in the process many genes were transferred from the endosymbiont to the nuclear genome of the host (Armbrust et al., 2004). This theory can be strengthened when looking at the numbers of membranes surrounding the plastids in the cells. Diatoms have four membranes that enclose their plastids while land plants only have two membranes (Falciatore and Bowler, 2002). The process where a red algae gets engulfed and incorporated into a nonphotosynthetic eukaryote is shown in Figure 2 (Armbrust et al., 2004).

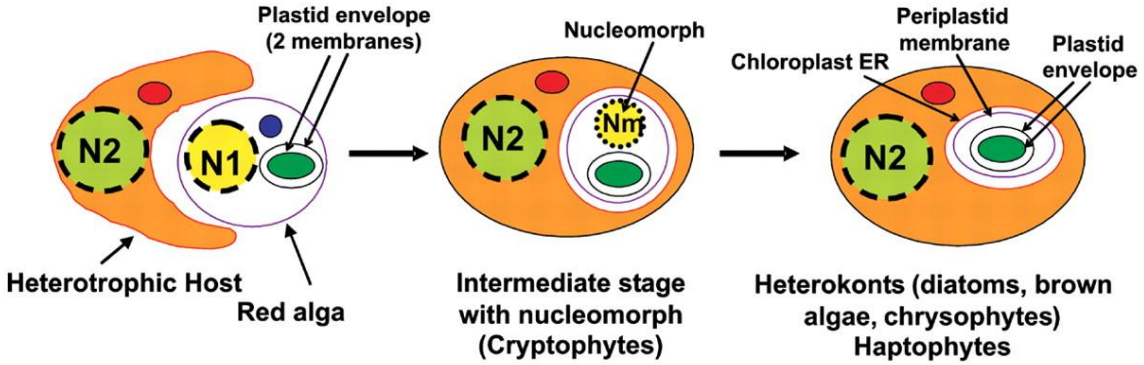

Figure 2: The origin of chloroplasts through secondary endosymbiosis with red algae. The red algae nucleus (N1) disappears after endosymbiosis. Some of the genes from N1 nucleus are transferred to the N2 nucleus of the host cell. After endosymbiosis four membranes surrounds the plastids in the cells. (Armbrust et al., 2004).

1.1.2 Life cycle

The life cycle traits of diatoms have only been studied in a few diatom species, but the ones that has been studied represent almost all of the principal diatom lineages (Chepurnov et al., 2008). These studies can therefore give an overview of the life cycle of diatoms. Diatoms have to stages in their life cycle. One vegetative stage, that lasts for months where the cells are diploid. The second stage is a sexual phase which lasts from hours to days. In the sexual phase the development of zygote to vegetative cell is also included (Chepurnov et al., 2008). In the vegetative phase the cells divide mitotically. During this phase the cells also reduce in size over time (Chepurnov et al., 2004). The cells enter the sexual phase when a certain size reduction is reached, approximately 30-40 % of the maximum diameter (Falciatore and Bowler, 2002). The reason for the size reduction is due to the restrictions the frustule imposes on diatoms during cell division. The silica wall of the diatoms consists of two overlapping halves called thecae. The two halves of the thecae form a "Petri dish-like" complex. When the cells divide, each of the daughter cells inherits one of the thecae. The thecae that the daughter cells inherit from the parent cell, epitheca, form the top of the new cell frustule. The smaller of the thecae, hypotheca, is produced during cell division and its shape is directed by the size of the epitheca. Because the frustule has a Petri dish form and the daughter cells always inherit the epitheca from the mother cell, repeated cell division leads to a reduced size in the population due to the smaller size of the hypotheca. The maximum cell size is reached during sexual reproduction when the cells go through gametogenesis and form a specialized cell type called an auxospore. The auxospore sloughs off the old thecae and expands to it reaches its maximum size. The new cell is formed inside the auxospore envelope, and starts to divide in a new round of vegetative multiplication. Sexual reproduction and size restitution are two features that are unique for the diatoms and a special feature of their life cycle (Falciatore and Bowler, 2002).

1.1.3 Light acclimation

Light availability is an essential factor that regulates the growth of photosynthetic organisms in the ocean (Falciatore and Bowler, 2002). Light is the primary energy source for algae but also provide them with positional information. Light responses in marine phytoplanktonic organisms are very sensitive to changes in light intensity and spectral quality. When comparing to photoperception and signal transduction in land plants, the marine environment have other restrictions (Falciatore and Bowler, 2002). The spectral distribution of solar irradiation is the same on land and the water surface. Under the water surface the light scattering properties and quality varies depending on the water and its constituents (Nymark, 2013). In clear ocean water, blue light (400–500 nm) belonging to the blue region of PAR is most predominant and the blue part of the spectrum increases with ocean depth (Falciatore and Bowler, 2002; Nymark, 2013). In addition, compared to land plants phytoplanktonic organisms that are drifting in the ocean constantly have to adapt to the different light conditions (Falciatore and Bowler, 2002). Some diatoms are also known to utilize movement as a response to light response. This is done to achieve optimal light conditions, and the process is called phototaxis. Light induced plastid reorientation is also observed in diatoms (Falciatore and Bowler, 2002).

1.1.4 Diatom research

Thalassiosira pseudonana and *Phaeodactylum tricornutum* were the first diatoms that got their whole genome sequenced (Bowler et al., 2008; Chepurnov et al., 2008) and they are currently the most important model organisms for diatoms. *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* are distantly related and belongs to each of the two major architectural types within the diatoms, centric and pennate diatoms respectively (Bowler et al., 2008). The diatoms satisfy most of the criteria used to describe a good model organism. They have short generation time, easy to maintain and experimentally tractable. This makes them suitable for laboratory work and there is also a great deal of cytological, physiological and biochemical data that is available. *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* were chosen as candidates for sequencing because of their small genome size and their easy handling in the laboratory where they grow easily and rapidly, with more than one division a day. They also lack the reduction–restitution cycle which is unique for diatoms (Chepurnov et al., 2008).

There has been developed some transformation system that has shown to be successful to introduce exogenous DNA into some diatoms. Transformation has been done with helium-accelerated particle bombardment (Apt et al., 1996; Chepurnov et al., 2004) and multi-pulse electroporation (Niu et al., 2012; Miyahara et al., 2013). One important factor to the transformation success was the use of endogenous promoter and regulatory sequences that are necessary for expression of foreign genes (Falciatore and Bowler, 2002).

Still there are some obstacles when using diatoms as an experimental organism, which is inherently related to their life cycle. There have only been developed successful breeding methods for a few diatoms with sexual reproduction (Falciatore and Bowler, 2002). With the absence of methods to control sex, desired traits and mutations are hard to maintain over generations in diploid organisms as diatoms. Other difficulties that one encounter is that laboratory studies of inheritance cannot be performed, so that a full analysis of the functional genome is not possible (Falciatore and Bowler, 2002).

1.2 Seminavis robusta

Seminavis robusta is a pennate benthic diatom and belongs to the group of *Naviculaceae* (Gillard et al., 2008). The mating system of *S. robusta* is heterothallic mating with two mating types (MT^+ and MT^-) (Vanstechelman et al., 2013). The diatom has a life cycle common with most diatoms, where there is a size reduction–restitution life cycle and a sexual phase with auxosporulation. The cells can be up to 80 µm long, which is large compared to other diatoms (Chepurnov et al., 2008). Because of their rather large size *S. robusta* and their ability to grow and move on surfaces like Petri dishes makes it easy to monitor the diatoms under low magnifications of an inverted microscope. A breeding program has also been initiated on *S. robusta* with a diatom pedigree by Chepurnov and coworkers (Chepurnov et al., 2008). They have among other things shown that *S. robusta* is highly tolerant to inbreeding (Chepurnov et al., 2008). *S. robusta* has a larger genome than *P. tricornutum and T. pseudonana* but it is probably smaller than the genome of the diatoms like *Pseudo-nitzschia multiseries.* Seminavis satisfy most of the requirements to be a good model organism, and with its genome now in draft format (unpublished) it may be the first benthic diatom model system established and therefore a good candidate for molecular genetic studies (Chepurnov et al., 2008).

Figure 3: *Seminavis robusta,* a benthic pennate diatom belonging to the *Naviculaceae* group.

1.2.1 Cytochrome P450

The cytochromes P450 (CYP) is a large superfamily of cysteinato-heme enzymes, and have got their name because of their carbonmonoxide-bound form have an absorption band at 450 nm (Werck-Reichhart and Feyereisen, 2000; Meunier et al., 2004). They are monooxygenases and catalyze reactions by incorporating single atoms from oxygen molecules into a substrate, and concomitantly reducing the other oxygen atom to water (Bernhardt, 2006). The reaction is shown in equation 1.

$$S + O_2 + 2e^- + 2H^+ \xrightarrow{\text{cytochrome P450}} SO + H_2O \quad (1)$$

The CYPs are present on most, but not all life forms (Werck-Reichhart and Feyereisen, 2000). They can be found in a variety of organisms like bacteria, plants and mammals, and it is thought that CYP superfamily in eukaryotes evolved from their prokaryotes ancestors. They have an crucial role in the oxidative transformation of exogenous and endogenous molecules in most organisms (Meunier et al., 2004). In mammals the CYPs have an essential role in drug metabolism, detoxifying xenobiotics, biosynthesis of steroid hormones and metabolism of vitamin D_3 (Werck-Reichhart and Feyereisen, 2000). In plant cells CYPs are thought to be associated with secondary metabolism of various componds such as pigments, plant steroids, lipid derived components, glucosinolates and others (Strøm Midthun, 2012). Studies done by Elise Strøm Midthun (Strøm Midthun, 2012) has shown that *S. robusta* has a particular high number of CYP coding genes, compared to other diatoms. While *P. tricornutum* has 8 CYP coding genes *S. robusta* has 68 genes, that is more than eight times as many genes (Strøm Midthun, 2012). This may be related to the ecological niche these diatoms occupy and could be an adaptation to a benthic habitat. In this master's thesis a gene expression study of selected CYP genes in *S. robusta* has been performed.

1.2.2 Light harvesting complexes (LHCs)

The light harvesting complexes (LHCs) are a part of the photosynthetic reaction center of the chloroplast, and consist of pigments and pigment-binding proteins (Nymark, 2013). The pigments fuel the photosynthetic reactions by harvesting energy from light. The genes coding for LHCs in diatoms can be divided into four groups, LHCFs, LHCRs, LI818/LHCSR-like LHCXs and LHCY (Nymark, 2013). Studies indicate that LHCF, LHCRI and LHCY proteins are present in light harvesting processes, while LHCX and LHCRII seems to be involved in photoprotection (Nymark et al., 2013). Two LCH genes (LCHF8 and LHCR6) from *S. robusta* were chosen for gene expression study.

1.2.3 Fatty acid desaturases

The desaturation of fatty acid chains is important for cells to keep the fluidity of biological membranes that consists of a bilayer of phospholipids and are important for the storage of excess energy in the form of triglycerides (Nakamura and Nara, 2004). Fatty acid desaturases have therefore been conserved in most organisms. There are classes of unsaturated fatty acids like arachidonic acid and docosahexaenoic acid that are essential in many physiological functions in animals. In humans three desaturases are known, they are stearoyl CoA desaturases (Δ9 desaturases), Δ6 desaturase and Δ5 desaturase (Nakamura and Nara, 2004).
Since unsaturated fatty acids are essential for cells, organisms have evolved various types of desaturases and elongases for production of specific unsaturated fatty acids. This is achieved by the introduction of a double bond into fatty acids where fatty acid desaturase catalyze the reaction aerobically.
Desaturases can be separated into two groups; membrane-bound desaturases and soluble desaturases (Nakamura and Nara, 2004). Soluble desaturases that are found in plants are Acyl-acyl carrier protein (ACP) desaturases. They are localized in plant plastids, and it is believed that they are connected to the electron transport system together with ferredoxin-NADPH reductase and ferredoxin (Shanklin and Cahoon, 1998).
The membranes-bound desaturases can be further divided into two subgroups. The two subgroups are acyl-lipid desaturases and acyl-coenzyme A (CoA) desaturases. The former group is found in membranes of cyanobacterial thylakoid, plant endoplasmic reticulum (ER) and plastids. The latter subgroup is present in the ER membrane of animals and fungi (Tocher et al., 1998), studies have also shown that they are present in organisms like insects and nematodes (Nakamura and Nara, 2004). Several desaturases has also been characterized in diatoms, including *P. tricornutum* where their gene expression are known to be influenced by light (Mühlroth et al., 2013).The Δ9 desaturase (D9) found in *S. robusta* was chosen for the gene expression study performed in this master's thesis.

1.2.4 Heat shock transcription factors HSF

Heat shock transcription factors (HSFs) are the transcription factors that control the transcription of heat shock proteins (HSPs) (Nakai, 1999). The HSFs are important transcription factors for the induction of many HSPs in most eukaryotes (Prändl et al., 1998). When cells are exposed to high temperatures heat shock proteins are induced as a protection mechanism for the cells. The HSFs control the transcription level of HSPs by detecting cellular metabolic changes like differentiation and stresses like heat shock, oxidative stress and exposure to heavy metals that cause denaturation of protein. The HSFs induce most of the HSP gene expressions by binding upstream of the heat shock element (HSE) (Nakai, 1999). Studies done on transgenic *Arabidopsis thaliana* by Prändl et al (Prändl et al., 1998) and Lee et al (Lee et al., 1995) show that the derepression of HSFs can increase basal thermotolerance (Lee et al., 1995; Prändl et al., 1998). In plants the HSFs are believed to be coded by up to five small gene families (Nover et al., 1996). Comparative analysis done on the fully genome sequenced *P. tricornutum* and *Thalassiosira pseudonana* (*T. pseudonana*) reveal that diatoms

have a high abundance of HSFs compared to other single-celled algae (Rayko et al., 2010). The HSFs have also shown to be important in non-stress situations in yeast cells (Nover et al., 1996; Morimoto, 1998). One HSF was also chosen for the gene expression study.

1.2.5 Aureochromes

Aureochromes are blue light photoreceptors that are only found in photosynthetic stramenopiles (Toyooka et al., 2011; Suetsugu and Wada, 2013). Photoreceptors are important for plants and algae because they give the cells information about the light conditions and help them fine tune the light responses (Toyooka et al., 2011). In marine environments blue light is predominant under the water surface because light with longer wavelength gets absorbed by the water masses (Toyooka et al., 2011; Suetsugu and Wada, 2013). Organisms like diatoms therefore use photoreceptors like aureochromes to detect blue light (Lockhart, 2013). The aureochromes have a light-, oxygen sensitive (LOV) domain, which is a subgroup of the of the Per-ARNT-Sim (PAS) super family (Toyooka et al., 2011). The LOV domain works as a sensory (Herman et al., 2013) and protein-protein interaction module (Toyooka et al., 2011). Another domain that also is found in aureochromes is a basic region leucine zipper (bZIP) domain which works as an effector (Herman et al., 2013). Aligned sequences on the LOV domains of the aureochromes show some similarities with domains in phototropins found in plants (Toyooka et al., 2011). The phototropins act as blue-light receptor in plants, and are well studied. The bZIP domain is an α-helix forming domain that binds to DNA; they work as transcription regulators and are present in a lot of eukaryotes. Because of these domains the aureochromes act as photosensors with DNA binding sites (Toyooka et al., 2011). A aureochrome gene was also chosen for the gene expression study.

1.3 Transforming algae

Because of their many different traits plus the fact that they are easy to growth and maintain in culture, algae have the potential for a variety commercial uses (Hallmann, 2007). If the algae is susceptible for genetic manipulation the use of algae is even more attractive. The sequencing of the genome of eukaryotic algae like *P. tricornutum* and *T.pseudonana* the recent years (Armbrust et al., 2004; Bowler et al., 2008), has given a lot of information about the individual organisms as well as their phylogeny (Armbrust et al., 2004; Bowler et al., 2008). This new genome information plus available transcriptome data is a good starting point for constructing vectors for transformation of algae (Hallmann, 2007). Successful transformation of algae are reported on a growing number of species over the last 30 years, and most of them have been nuclear transformation (Hallmann, 2007). Some of the species that have been transformed are *P. tricornutum* (Apt et al., 1996; Falciatore et al., 1999), *Thalassiosira weissflogii* (Falciatore et al., 1999), *Laminaria japonica* (Qin et al., 1999) and *Haematococcus pluvialis* (Steinbrenner and Sandmann, 2006).

Some basic questions that are need to be addressed when an algae specie is chosen to be transformed are:
Are there any closely related species of the algae that has been transformed?
What kind of vector construct should be used for the transformation?
Are there any promoters that can give a high expression level?
How should the DNA be transferred into the cells and later monitored?
Genetic transformation attempts on algae have shown that the total number of transformants produced and transformation efficiency vary between the different species. When comparing genetic transformation studies done on algae there seems to be one reoccurring factor effecting transformation efficiency. A general trend seems to be that the number of successful transformations decreases as the size and complexity of the algae increase (Hallmann, 2007).

To be able to transform organisms, appropriate promoters must be chosen (Hallmann, 2007). This is important because without the right promoters the genes transformed into the algae cannot be expressed. Some properties which has to be considered is whether the promoter should be inducible or continuously expressed (Hallmann, 2007). There are some promoters that have shown to work in several algae species like the CaMV35S and SV40 promoters. Preferably the promoters should be endogenous, and chosen from a well characterized gene (Falciatore and Bowler, 2002). This knowledge is important since promoters of most commercial selectable marker genes are not functional in algae (Hallmann, 2007).

The use of selectable marker genes is important in the transformation technology, because it makes it possible to monitor, identify and select individuals that have been transformed (Miki and McHugh, 2004). In algal transformation the selectable marker is often a gene that codes for antibiotic resistance (Hallmann, 2007). This new gene separates the transformed cells from the non-transformed cells because the transformants acquire a new antibiotic resistance that previously did not exist in the organism. There are many types of selectable marker genes that have successfully been used in algal transformation. In *P. tricornutum* the *ble* gene from *Strepto-alloteichus hindustanus* has shown to give resistance for zeomycin (Apt et al., 1996; Falciatore et al., 1999). The *sat-1* and *nat* genes can also be expressed in *P. tricornutum* and gives the diatom resistance to the antibiotic nourseothricin (Zaslavskaia et al., 2000).

In many cases of transformation the introduced genes are not expressed as desired in the algae, even though the right promoters have been chosen (Hallmann, 2007). The lack of expression can be a result of methylation caused by the position of the integrated DNA or other epigenetic mechanisms. The response can be a part of a defense mechanism against viruses, transposable elements or foreign DNA. Another issue that can affect the expression of a foreign protein is related to GC-content and codon use of the exogenous DNA. This is because codon usage can be different between species. It is therefore important to have a gene sequence that corresponds to the tRNA abundance in the target organism. If the genes don't correspond well to the tRNA abundance the translation of the gene will be low and hence effect the expression. The presence of introns in exogenous DNA transformed into the organisms can also lead to no expression of the protein. This is because the target organisms will often not be able to splice the DNA correctly (if it is derived from a distant related

organism). Genes containing only exons should therefore be used and in most cases cDNAs are used. However, sometimes there are cryptic introns in cDNAs which results in splicing of "intron-less" genes (ex. when GFP was first expressed in *Arabidopsis thaliana* it was mis-spliced). In some cases it has been shown that genes without introns also are expressed poorly. This issue can be solved by introducing homologous introns into the exogenous gene. There are also other difficulties when transforming algae that are common in plants as well as other eukaryotes, and they are related to failure of DNA delivery, failure of DNA integration into the genome or the lack of DNA transported through the plasma membrane and degradation of exogenous DNA (Hallmann, 2007).

When a vector construct has been made for the algae there are a few transformation methods that are available. The methods are based on incorporating the DNA into the cells by making the cell membrane temporarily permeable (Hallmann, 2007). After the DNA enters the cell it gets incorporated randomly in the genome by recombination events. The challenge with introducing DNA into cells is to make the cell membrane permeable only for a short moment and not kill the cells during the process. One transformation method that has given successful transformation in algae is the micro-particle bombardment also called gene gun transformation (Sanford, 1990; Hallmann, 2007). The technique is based on transforming cells by shooting metal micro-projectiles coated with DNA on the cells. The particles that are shot have a high velocity (Sanford, 1990), which makes it possible to transform cells with thick and rigid cell walls even diatoms with their silica walls (Hallmann, 2007). It is possible to do transformation of organelles as well, but that require specific vectors which contain promoters that are functional in the organelle, ex. chloroplast. Several diatoms have been transformed with gene gun transformation, some of them are *P. tricornutum* (Zaslavskaia et al., 2000), *Cyclotella cryptic* and *Navicula saprophila* (Dunahay et al., 1995).
A second transformation method is electroporation. During electroporation a high voltage is sent though the cells. The high electric voltage temporarily destabilizes the cell membrane by disrupting the phospholipids which allows DNA to pass though (Hallmann, 2007). It was recently shown that this method too can be used to transform the diatom *P. tricornutum* (Niu et al., 2012). A third transformation method that has been done on two algae species is *Agrobacterium tumefaciens* mediated transformation (Hallmann, 2007). *Agrobacterium tumefaciens*-mediated transformation has mainly been used to do transformation of plants, but has now shown ability to transform the red algae *Porphyra yezoensis* (Cheney et al., 2001) and the green algae *Chlamydomonas reinhardtii* (Kumar et al., 2004).

1.4 Aim of study

The primary aim with this study was to gain more knowledge and experience in the use of *S. robusta* in laboratory work. This aim was approached by:
First making at vector construct for *S. robusta* by using endogenous bidirectional promoter and terminator sequences and try to establish a transformation method for this diatom.
Secondly, an expression study was also done on *S. robusta*, by looking at the expression differences after exposure to blue light and white light with different time periods (0,5 and 6 hours).

Since *S. robusta* is a promising candidate for becoming a model organism for benthic diatoms, the availability of a vector construct can help provide a greater understanding of the molecular genetics in the algae. A light experiment will also reveal mechanisms on how *S. robusta* adapts to various light types. These responses were measured and compared with responses seen in other diatoms by the use of bioinformatical tools and transcriptional analyses.

2 Materials and Methods

2.1 Cultivation of algae

The cultures of *S. robusta*, matingtype 111-1 and 112-1 were prepared by thawing cryopreserved cells. The cells were acquired from the Protistology and Aquatic Ecology (PAE) laboratory at the department of biology at Ghent University in Belgium. Since the diatoms were marine, f/2 medium (Appendix I) was used as cultivation medium, which is normal for marine algae. Growth conditions were 18°C with a day and night cycle consisting of 16 hours constant light at 100 µmol photons $m^{-2}s^{-1}$ and 8 hour with absence of light. The cultures were grown in 50mL sterile cell tissue flasks (BD Biosciences, cat.no 353136). To maintain the cultures 5mL of cells were transferred from flasks with cell scraper (VWR, cat.no 734-2603) and pipettes with dense cultures to new flasks with 45mL f/2 medium, this was done approximately once a week.

2.1.1 Axenization of cells

Axenization of cells was performed as described as in Algal culturing techniques (Andersen, 2005) with modifications. The cultures were examined for bacterial contamination by axenization testing and also by microscopy. Axenization was done when bacterial contamination was detected.

Procedure:

1. Axenization was done by resuspending cells and transferring 0,5mL of cells from a culture flask where the cells had grown dense and exponentially to a 15 mL falcon tube (VWR, cat.no 525-0150).

2. A volume of 10mL, f/2 medium and 1,5mL antibiotic mixture (see appendix I) was added to the tube. A mixture of 0,5mL f/2 and 1mL peptone (g/L) was also added to stimulate bacterial growth.

3. The cultures where incubated horizontally for three days, before they were checked for bacterial growth. Then 0,5mL of cells were resuspended and transferred to a $25cm^2$ sterile growth flask (VWR, cat.no 734-2311) with 10mL f/2 medium.

4. After a week an axenization test was performed on the culture by resuspending the cells and transferring 0,5 mL of the cells to a 15mL falcon tube with 5mL f/2 medium and 1mL peptone (g/L). The tubes were wrapped in aluminum foil and incubated in room temperature for two weeks. This was done to check for bacterial growth, which could be detected if the solution became cloudy or opaque.

2.1.2 Harvesting cells

When harvesting cells for RNA isolation it's important that the procedure is performed fast, and thereby avoiding unnecessary exposures such as light, temperature and mechanical stress that can affect the RNA-pool. This can give false results when doing gene expression studies later based on the RNA.

Procedure:
The harvesting of cells was done by scraping the cells off the bottom from culture flasks with a cell scraper and then pouring the cells with the medium into a 50mL falcon tube. The tubes were then centrifuged at 3000 rpm for 5 minutes. The supernatant was decanted after the centrifugation, leaving the diatom pellet and ca 1,5mL supernatant in the falcon tube. The pellet was then resuspended with the supernatant by pipetting and transferred to a 2mL tube and centrifuged for 1min at 13000rpm on 4^{o}C. The rest of the supernatant was removed and the cells were flash frozen in liquid nitrogen, and later stored in a -80^{o}C freezer.

2.1.3 Counting cells

The counting and estimation of cells was done by counting cells in a Nageotte counting chamber with a Nikon Eclipse E800 microscope at total magnification at x100.

Culture flasks with *S. robusta* were first resuspended with a cell scraper and 1-2 mL of the suspension was transferred to a dram glass. The suspension in the dram glass was then added with lugol's solution. Lugol's solution was used to stain and immobilize the cells for the counting. To be able to estimate the number of cells in the flask a small volume of the solution in the drams glass was transferred to the Nageotte counting chamber. Eight stripes were counted in the chamber. The average of these stripes was then calculated and an estimate of the cell concentration was found.

2.1.4 Isolation of genomic DNA from *S.robusta*

Genomic DNA from *S. robusta* was provided by Dr. Tore Brembu and principal engineer Torfinn Sparstad. The isolation of genomic DNA was done as described in Brembu et al. (Brembu et al., 2013), which is a modification of the protocol from Bowler et al. (Bowler et al., 2008). For procedure see appendix I

Procedure:
For procedure see appendix I

2.2 Producing a transformation vector for *S. robusta*

A transformation vector for *S. robusta* was designed and the cloning strategy was developed by Dr. Tore Brembu and Per Winge (see figure 8). The promoter and associated terminators were amplified from genomic DNA, cloned and verified by sequencing.

2.2.1 Polymerase Chain Reaction

The polymerase chain reaction (PCR) is a technique where DNA sequences can be amplified *in vitro* (Gibbs, 1990). PCR is a technique that is frequently used in many areas of molecular biology. The technique utilizes DNA polymerase ability to synthesize DNA, and can make millions of copies from one single DNA strand. The reaction is divided into 3 phases. The first phase is a denaturing phase where double stranded DNA is denatured at a high temperature. The second phase is called the annealing phase. In the annealing phase the temperature is lowered. This allows the forward and reverse primers that are complementary to the 3' end of the sense and antisense DNA strands, to anneal to the target regions on the DNA strands. The third and final step of the PCR cycle is the extension phase. At this phase the temperature is set to the optimum temperature for the DNA polymerase, so that new DNA strands can be synthesized. During PCR amplification the number of DNA fragments determined by the flanking PCR primers is amplified exponentially. The increase in DNA product can be represented as 2^n, where n is number of cycles. The polymerase that is used in PCR is heat-resistant; this prevents the enzyme from degrading under the high temperatures. This makes it possible to use the same enzyme throughout the whole reaction without having to add new enzymes for every cycle (Gibbs, 1990).

Procedure:
A bidirectional promoter was chosen for the vector construct because of the small size and its ability to express two genes. The bidirectional promoter for AtpE – AtpB genes and the terminator sequences AtpEt and AtpBt were PCR amplified from *S. robusta* genomic DNA. The AtpE/D gene codes for the epsilon (or delta) subunit of the mitochondrial ATP synthase, while the AtpB gene codes for the beta subunit of the same enzyme. Primer pairs were designed for each of the sequences and ordered from Sigma. PCR primers with enzyme restriction sites were used to facilitate cloning of the PCR products. The PCR reactions were performed in 0,2ml PCR tubes. The reaction mixture for the PCR reaction is shown in Table 1, PCR cycle details are shown in **appendix III**. The volumes and concentrations for the PCR reaction mixture were used as recommended from the instructions from Takara BIO Inc for Ex Taq polymerase (Takara BIO Inc, cat.no. RR001A).

Table 1: PCR reaction mix.

Reactants	Volume (µL)
Autoclaved MQ water	37
10 x Ex-Taq Buffer	5
dNTP Mixture (2,5mM)	4
Template (gDNA)	2
Forward primer	1
Reverse primer	1
Total volume:	50

After the PCR reaction was complete a small aliquot of the PCR product (2µL) was visualized in a 1,2% agarose gel. This was done to see if the amplified PCR product had the expected product size. GeneRuler ladders at 1kb (Fermentas Life Sciences, cat.no SM1332) and 100 bp (Fermentas Life Sciences, cat.no SM0321) were used to decide the length of the bands on the gel.
The rest of the PCR products were stored at – 20 ^{o} C until they were used for further applications.

2.2.2 Gel electrophoresis

For analyzing DNA structure and integrity, gel electrophoresis is a commonly used technique (Johnson and Grossman, 1977; Bjornsti and Megonigal, 1999). When performing agarose gel electrophoresis the DNA fragments are separated according to the size and shape of the products (Bjornsti and Megonigal, 1999). The negative charge of the DNA molecules is utilized during electrophoresis (Bjornsti and Megonigal, 1999), where the strands migrate through the gel matrix away from the negative pole and towards the positive electrode. The migration of the DNA fragments is determined by several factors: the molecular size, the shape of the DNA (circular or linear DNA), conformation, the concentration of agarose (pore size) and the net charge through the gel and the ionic strength of the buffer (Johnson and Grossman, 1977; Bjornsti and Megonigal, 1999).

Procedure:

1. 100mL of 1 x TAE buffer was mixed with 1,2 mg agarose (Sigma, cat. no. A9539-500G) in a conical flask.

2. The mixture was then heated in a micro wave at maximum heat (900W), until all the agarose had dissolved and the solution had become clear. A lid was placed loosely on the top of the conical flask.

3. The solution was then let to cool down for a few minutes, until 5µL Gel Red™ (Biotium, cat. no. 41003-1-10ml) was added to the solution and mixed into the solution by carefully shaking the flask without making bobbles.

4. The mixture was further cooled for about 5min before it was poured into a gel tray with a comb, all which was provided by Bio-Rad. The gel was set to cool down and become stiff for 30min.

5. When the gel had solidified it was put in a gel tank filled with appropriate volume of 1 x TAE buffer and 20µL sample in each well, including 5µL DNA ladder (1 kb Plus DNA ladder) with fragments of known sizes in the well furthest to the right and left. The gel was put on a constant voltage at 70V for one hour.

6. After an hour the gel was analyzed to bands on the Bio-Rad Gel Doc 2000.

If the DNA bands on the gel were intended for future use, they were cut out and purified with a SV Gel and PCR Clean-Up system kit (Promega, cat.no A9282).

2.2.3 DNA purification

In molecular biology the possibility to manipulate high quality DNA is important, there are various ways to isolate and purify both single and double stranded DNA (Hawkins et al., 1994) . One method that is used is purification with the use of silica particles, as silica has a high affinity for DNA (Yang et al., 1998). When using purification methods involving silica based spin columns, one can concentrate and purify DNA at the same time (Yang et al., 1998). Silica based spin columns produced from Promega Wizard SV Gel and PCR Clean-Up kit was used for extract DNA from the gel separation.

Procedure:
The purification steps were done as described in the kit given from Promega.
After the DNA was purified the quantifications of the DNA was done with NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, USA).

2.2.4 DNA quantification – NanoDrop

It is important to know the amount and concentration of DNA in samples if they are to be used used in further molecular biology applications (O'Neill et al., 2011). The quantification of DNA can be decided with the use of NanoDrop 1000 spectrophotometer (Desjardins and Conklin, 2010).
Through UV spectrophotometry light is sent through a sample, and the light intensity passing through the sample is measured. This can give an estimate on the DNA concentration, and if there are any contaminants, e.g. RNA, proteins or other molecules(O'Neill et al., 2011). The

Nanodrop exploits DNA molecules ability to absorb light at 260nm to give an estimate of the purity of the DNA. Nucleic acids and proteins have absorbance maxima at 260 and 280 nm, respectively and the 260/280 ratio can be used to evaluate if the sample contain proteins, similarly the 260/230 ratio can be used to evaluate other contaminations, ex. carbohydrates) (O'Neill et al., 2011).

Procedure:

The DNA quantification was performed as described in the NanoDrop™1000 spectrophotometer manual. MilliQ water was used as blank sample (1µL).

2.2.5 Restriction endonucleases

Restriction endonucleases are enzymes that are useful and often used in recombinant DNA technology (Loenen et al., 2014). Based on their different characteristics restriction endonucleases can be divided into four main groups, respectively Type I, II, III and IV. In molecular biology Type II is the most commonly used restriction endonucleases. This is because of their property to recognize and cut specific DNA sequence motifs called recognition sites that often occur as palindromes (Loenen et al., 2014). Restrictions enzymes ability to cut at specific places can be used when constructing a vector, where genes / DNA fragments can be cut and inserted into a vector and studied.

Procedure:

A mixture was made with all components as shown in Table 2. All the solutions were transferred to a 1,5 mL eppendorf tube, and this was carried out on ice. The mixture was then incubated at the recommended temperatures provided from www.neb.com (New England BioLabs inc.).

Table 2: Reaction mix for restriction enzyme facilitated DNA cutting

Reactants	**Volume (µL)**
Autoclaved MQ water	15,5
10 x NEB Buffer	2
Template (DNA)	2
Restriction enzyme*	0,5
Total volume:	20

*Double restriction enzyme cutting was also done. When this was done 0,5 µL enzyme number 2 was added as the same volume of water was reduced. A list of all the restriction enzymes used for the vector construction is listed in appendix II.

After 1,5 hours a agarose gel electrophoresis was done on the samples to separate the DNA fragments that where cut, from the uncut samples. The samples were then purified with Promegas' SV Gel and PCR Clean-Up system kit, and later stored in a -20^{o}C freezer.

2.2.6 DNA ligation

DNA ligases have been isolated from various of organisms (Nilsson and Magnusson, 1982) and it is known that they contribute during DNA replication, repair and recombination by restoring nickes on single strands of the DNA double helix (Sgaramella and Ehrlich, 1978). Among these enzymes, DNA ligase from phage T4 (T4 DNA ligase) has shown the ability to ligate blunt ends (fully base-pared DNA chains) together (Sgaramella and Ehrlich, 1978; Nilsson and Magnusson, 1982). This makes T4 ligase useful in DNA cloning, since some type II restriction endonucleases produce blunt ends.

Procedure:

1. A calculated volume of MQ water, based on the NanoDrop values of the linear vector and DNA fragment of interest (insert) was added to an eppendof tube. The list of all the reagents is shown in Table 3.

2. The tube was kept on ice as the other reagents were added to the mixture. The calculated volumes of vector and insert were added, before the ligation buffer was added. The tube with the ligation buffer was properly vortexed before the buffer was added to the mixture. The T4 DNA ligase (Fermentas Life Sciences, cat.no EL0014) was added at the end.

3. The mixture was spun down quickly and incubated at room temperature for 1,5-2 hours.

4. After incubation the mixture was set on 65^{o}C for 10 min to denature the ligase and stored at -20^{o}C and was ready for transformation.

Table 3: Reaction mixture for ligation reaction.

Reagents	**Volume (µL)**
Autoclaved MQ water	Calculated so total volume is 10
Linear vector	Calculated after insert concentration
Linear insert	Calculated so that insert –vector ratio is 6:1
10 x T4 ligation buffer	1
T4 DNA ligase	0,5
Total volume	**10**

2.2.7 Vector - pBluescript KS+

pBluescript KS+ (Stratagene, cat.no 212207) is a multipurpose vector, and part of a series of pBluescript vectors with different f1 origin and polylinkers (Alting-Mees and Short, 1989). The vectors were originally designed to assist in the mapping of DNA insert (Alting-Mees and Short, 1989), based on theory proposed by Wahl et al (Wahl et al., 1987). A special feature with the pBluescript vector is that it has restriction enzyme sites in the β-gal gene, which is part of the polylinker, and thereby allow blue/white color selection on the recombinant clones after induction of the β-gal gene with IPTG and supplying the substrate Xgal (Alting-Mees and Short, 1989). Previous studies have shown that the pBluescript vector can be used on transformation of diatoms (Dunahay et al., 1995; Falciatore et al., 1999; Poulsen et al., 2006). The pBluescript KS+ was used in the construction of the *S. robusta* transformation vector.

2.2.8 Heat Shock transformation

The transformation of exogenous DNA into *Escherichia coli* (*E.coli*) cells is a method widely used and can be used to study gene functions (Sha et al., 2011). The method to chemically treat *E.coli* cells coupled with heat shock was first proposed by Mandel and Higa (Mandel and Higa, 1970) in their protocol in 1970 (Hanahan, 1983; Sha et al., 2011). Since Mandel and Higas' protocol there have been a wide range of modifications on the technique (Sha et al., 2011). The knowledge about what happens during heat shock transformation is not fully understood yet, but there are two models that try to explain the mechanism behind the transformation of bacterial cells. The first model explains that exogenous plasmid DNA binds to receptor complexes in the cell membrane and is then transported into the cells though transmembrane channels (Hanahan, 1983; Sha et al., 2011). The other model suggests that plasmid DNA passes into the cell during the heat shock step, because the cell membrane becomes destabilized during this short period (Sha et al., 2011).

Procedure:

1. A volume of 2 µL pBluescript KS+ was added to an eppendorf tube with chemically competent DH5α *E.coli* cells. The tube was gently mixed without pipetting or vortexing, and then set for incubation on ice for 30min.

2. The cells where put on heat shock treatment for 30 seconds at 42^{o}C in a water bath, and immediately put on ice.

4. The cells were then added 500 µL S.O.C. medium at room temperature. (See appendix II for recipe).

5. The tube was properly closed and put on one hour incubation at 37^{o}C with constant shaking.

6. Different volumes of the transformation mix, 10, 50 and 100 µL were then added to LB medium plates with ampicillin (Amp) (Sigma-Aldrich, cat.no A9518-25G) and incubated overnight at 37^{o}C.

7. Transformed colonies where picked, incubated in LB/Amp medium over night, and the plasmids were isolated. The plates with the transformed cells were later stored in 4^{o}C cold storage.

2.2.9 Plasmid-isolation based on column

There are many ways to isolate DNA from organisms. One way of doing plasmid isolation is based on first releasing DNA from the cells by adding a lysis buffer (Singh and Anthony Weil, 2002). The QIAGEN method uses the immobilization of DNA on a silica substrate with a chaotropic agent in a spin filter (Little, 1991; Tanaka and Ikeda, 2002). The DNA is then washed with buffer containing ethanol (Tanaka and Ikeda, 2002) and then eluted from the column by adding water or low salt buffer onto the column. The system is based on lysis of *E.coli* cells with a alkaline solution followed by the absorption of DNA to a silica membrane while there is a high salt buffer present. The whole process can be separated into three steps. The first step is the preparation and clearing of a bacterial lysis. The next step is the attachment of DNA to the QIAprep membrane by utilizing the buffers in the lysis solution. The final step consists of several washing procedures where salts are removed and the plasmid is purified. Plasmid DNA is finally eluted by adding water with pH between 7-8,5 causing the DNA to reattach itself from the membrane (QIAGEN, 2004; Zhang and Cahalan, 2007).

Procedure:

Plasmid isolation was done using the QIAGEN (Hilden, Germany) QIAprep® Spin miniprep kit (QIAGEN, cat.no 27106). The method was done as instructed by manufacturer.

DNA concentration was measured with NanoDrop 1000 after the isolation and stored at -20^{o}C

2.2.10 DNA sequencing for verification

DNA sequencing is the process in which the order and the nucleotides of a given DNA strand are determined. The ability to gain information about the nucleic acid composition in a DNA strand is important in many biological sciences (Pettersson et al., 2009). One of the most innovative sequencing techniques in biological research was the sanger sequencing when it was introduced in 1977. Since that time a variety more efficient sequencing techniques have become available, but still many of the new sequencing techniques use fluorescence-based sequencing based on the Sanger sequencing method (Pettersson et al., 2009).

Procedure:

All PCR products and intermediate vector products were sequenced for verification of successful amplification and insertion respectively. The sequencing was done with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Life technologies, cat.no 4337455) and sequencing primers were selected depending on the DNA product or plasmid vector used in the analysis. The PCR reaction was done at the NTNU lab, while sequencing was performed with Applied Biosystems 3130xl Genetic Analyzers at the DNA sequencing core facility to the University Hospital of North Norway. The sequencing results were e-mailed one to two business days after sequencing.

2.2.11 Biolistic transformation of Seminavis Robusta

The principle behind biolistic transformation is to use high velocity and micro projectiles coated with DNA, to transfer DNA into cells through cell walls and membranes. The technique is called biolistic (biological ballistics) transformation, because DNA is shot into the cells (Sanford, 1990; Taylor and Fauquet, 2002). The technique was initially used on plant cells and tissue, but later modified for transformation of other species (Taylor and Fauquet, 2002), including algae (Mayfield and Kindle, 1990). Studies on algae has shown that transformation with microparticle bombardment is possible on diatoms like *Phaeodactylum tricornutum* (Apt et al., 1996), *Cyclotella cryptic* and *Navicula saprophila* (Dunahay et al., 1995).

The Biolistic PDS-1000/He instrument (BioRad, cat.no 165-2257) from BioRad utilizes the principle of biolistic transformation to transforme cells. The system uses pressurized helium gas to accelerate DNA coated micro-projectiles for transformation (BioRad, 1995). The system consists of several units. The main unit is the bombardment chamber followed by a connective tubing where the vacuum source is connected and other components that are necessary for the delivery of high pressure helium. Figure 4 shows an overview of all the components in the Biolistic ® PDS-1000/He Particle Delivery System.

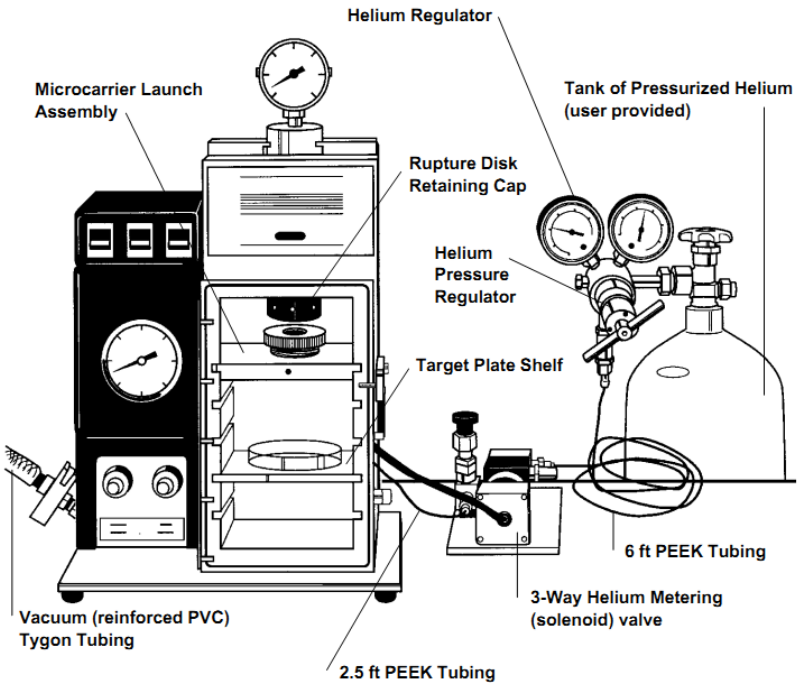

Figure 4: An overview of all the components in the Biolistic ® PDS-1000/He Particle Delivery System (BioRad, 1995).

During the biolistic process a high pressure of helium is released by a rupture disk (BioRad, cat.no 165-2332). This causes the macrocarrier (BioRad, cat.no 165-2335) which is coated with microcarriers (BioRad,cat.no 165-2267) to move towards the target cells with a high speed. The macrocarrier is however stopped before it can reach the target cells by a stopping screen. The stopping screen (BioRad, cat.no 165-2336) prevents the microcarrier sheet for reaching the cells, but lets the DNA coated microcarriers pass though and transform the target cells. The process of how the rupture disk releases the helium gas and the transformation of cells is illustrated in Figure 5.

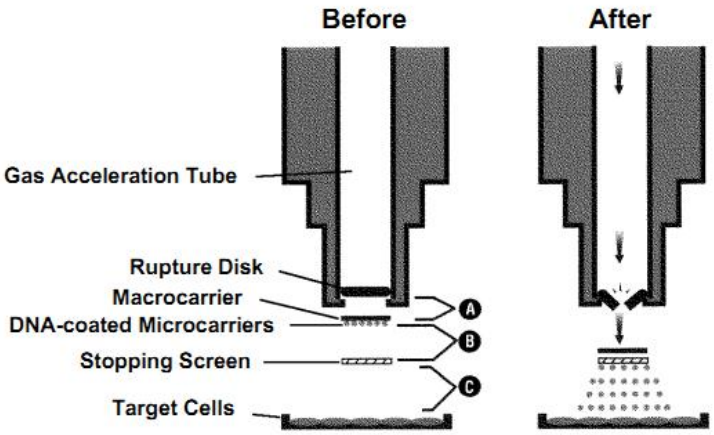

Figure 5: The high pressure helium start to build up in the gas acceleration tube, and when the right pressure is reached the rupture disk releases the gas. This causes the macrocarriers to move towards the stopping screen with high velocity. The stopping screens stops the macrocarrier, while the microcarriers pass though and transforms the target cells (BioRad, 1995).

The velocity of the microcarriers during the bombardments is dependent upon five factor:

1. The first is the helium pressure used, which is decided by the use of rupture disk.
2. The amount of vacuum in the bombardment chamber.
3. The distance between the the rupture disk and the macrocarrier.

4. The distance from the macrocarrier to the stopping screen and (5) the distance between the stopping screen and the target cells (BioRad, 1995).

Procedure:

Plating of diatoms for biolistic transformation

1. Agar plates with 50 % f/2 medium and 50 % autoclaved MQ water were prepared.

2. The cells where harvested from culture flasks and centrifuged in falcon tubes for 10 min at 4500 g. The concentration of cells was calculated by counting a 2mL sample of the cells in a Nageotte counting chamber. An appropriate concentration would be 1 x 10^6 cells/mL.

3. The supernatant was decanted and the algae pellet was resuspended in 1mL f/2 medium.

4. The resuspended cells were transferred to a prepared f/2 – MQ water agar plate. The Plates were set to dry for 10 min without lid, before they were sealed and put in a culture room for 24 hours before transformation.

Preparation of micro projectiles

1. 60 mg of tungsten particles (Microcarriers) were transferred to a 1,5 mL eppendorf tube, and mixed with 1 mL absolute ethanol. The mixture was then vortexed for 2 min with full speed ethanol removed and 1 mL new ethanol added. This was repeated 3 times.

2. The eppendof tube was centrifuged at 10 000 rpm for 1 min, and the supernatant was removed. The tungsten was resuspended in autoclaved MQ water and the previous vortexing and centrifuge steps were repeated twice.

3. While constantly vortexing the eppendof tube with tungsten particles, 50 µL aliquots from the tungsten mixture was transferred to 20 new eppendorf tubes. The tubes where stored at -20^{o}C until further use.

Tungsten particle DNA coating

The coating of the tungsten particles was done was described by Peter Kroth (Kroth, 2007), with some modifications.

1. A required amount of tungsten particle aliquots (50 µg) was thawed and vortexed at full speed for 3-5 min.

2. While the solution was thoroughly mixed, 5 µL (1 µg/ µL) linear plasmid DNA was added, quickly followed by 50 µL $CaCl_2$ (2,5 M) and 20 µL spermidine (0,1 M).

3. The effendorf tube was set on continuous vortex for 2 min and put on ice for 1 min to let the particles sediment.

4. The solution was centrifuged at 10 000 rpm for 2 seconds and the supernatant was removed.

5. The pellet was washed in 140 µL 70% ethanol, centrifuged, and the supernatant removed. The same procedure was repeated with absolute ethanol.

6. The tungsten particle pellet was resuspended in 60 µL absolute ethanol at low vortexing speed.

Preparation of micro bombardment

1. The hood and the biolistic device (Bio Rad, cat.no 165-2257) and all equipments for the experiment were washed with 70 % ethanol before transformation to avoid unnecessary contaminations with fungi and bacteria.

2. The rupture disk was dipped in 70 % isopropanol before fitted within the device as instructed in the manual from manufacturer (BioRad).

3. A test run with no cells and particles was done to clean the helium lines. While doing this test run the helium tank pressure regulator was set to 1700 psi for the transformation. The rupture disks chosen for the transformation where made so they would rupture at 1700 psi and release the high pressure gas. The interior of the device was cleaned again after the test run.

3. Macrocarriers were washed with ethanol and dried before placing them in macrocarrier holders, by the use of a seating tool.

4. The eppendorf tube with the DNA coated tungsten particles was vortexed as 10 µL was taken and placed on the middle a macrocarrier and set to dry.

5. The macrocarrier with the DNA-tungsten mix, the stopping screen and the rupture disk were all installed into the device as instructed in the Bio-Rad Biolistic® PDS-1000/He system manual.

6. A prepared agar plate with diatoms was placed in the chamber, on shelf 2 without the lid.

7. The chamber-door was closed and air was deflated by turning on the vacuum button.

8. When the pressure showed 26,5 psi the vacuum was turned off and the "fire" button was pushed and held in until the rupture disc ruptured. After the bombardment the vacuum was released and a new bombardment was prepared.

9. Each agar plate was shot 5 times to cover larger areas of the plate with tungsten particles.

10. All the agar plates were sealed with parafilm and placed in a culture room for 2 days for recovery before replating them on selective medium.

Replating and selection of transformants

1. 1 mL of fresh sea water (FSW) was added to the agar plates from the micro bombardment. The solution is spread across the plate by carefully stoking the surface with a cell scraper (VWR).

2. The plate was tilted carefully so that all the solution got gathered in one place, so that the resuspended cells could be transferred to a f/2 agar plate containing nourseothricin (100 µg/mL), (Jena Bioscience, cat.no. AB-102L), see appendix I

3. The cells were carefully spread on the selective medium using a cell scraper, and set to dry, before they were sealed with parafilm and incubated for 2-3 weeks in a culture room.

4. After 2-3 weeks, colonies were visible as uneven brownish spots. They were picked and transferred to a 6 well plate with f/2 media and a concentration gradient of nourseothricin. Well number 1-5 had the concentrations 100, 150, 175, 200 and 250 µg/mL, while well 6 only contained f/2. This test was done to check the sensitivity of the cells to nourseothricin.

2.3 Light exposure experiments

Based on results from previous light experiments performed in *P. tricornutum* (Nymark et al. 2013) and *S. robusta* (Strøm Midthun 2012) a set of CYP genes, light harvesting complex genes (LHCs), a heat shock factor (HSF4), a bZIP-PAS transcription factor (PAS9) and a fatty acid desaturase (D9) were chosen for further studies when *S. robusta* were grown under different light conditions. The genes which were studied for *S. robusta* are homologous genes of the ones that were identified as light responsive in *P. tricornutum.*

Procedure:

Axenic *S. robusta* cells where grown in 125mL volume flasks (VWR, cat.no 734-2315) under normal growth conditions as described in "Cultivation of algae", but the day before the light exposure treatment the cells had a 12 hour period with absence of light. After this period the cells were treated with constant blue light (BL) at 80 µmol photons $m^{-2}s^{-1}$ for 0,5 and 6 hours and harvested immediately after. Each treatment had four biological replicates. Additionally four biological replicates where exposed to white light (WL) at approximately 100 µmol photons $m^{-2}s^{-1}$ for the same exposure time, and harvested. Four biological replicates kept for 12 hours in darkness were also harvested, as a control (Dark) to determine which genes are turned on and off during the dark/light transition.

2.3.1 RNA isolation

When performing Real-time quantitative PCR (qPCR) gene expression analysis it is kown that the quality and quantity of the initial starting RNA can influence the accuracy of gene expression (Fleige and Pfaffl, 2006) It is therefore important that the starting RNA is of good quality so that it will not affect downstream analyses that often are time consuming and expensive (Fleige and Pfaffl, 2006).

Procedure:

Stainless steel beads (QIAGEN) 5 mm cooled to -80^{o}C where added to 2 mL tubes (Sarstedt, cat.no 72.695.500) containing cell pellets, stored at -80^{o}C. The pellet with the steal bead were homogenized (mechanically disrupted) in a tissueLyser (QIAGEN, cat. no. 85200) for 2x1 min at 25 Hz. During the first homogenization step the tubes were in a homogenization block that had a temperature around -80^{o}C to prevent any changes in the RNA degradation.
The Spectrum^{TM} Plant Total RNA kit (Sigma Aldrich, cat. no. STRN250) was used for RNA isolation. 500 µL of Lysis buffer was added to the sample and the tube was put in a homogenization block at room temperature. After 4 min incubation at 65^{o}C the samples were sentrifuged and the supernatant was transferred to the filtration column and centrifuged again. After the centrifugation, binding solution was added to the flow-through. The flow-through mixed with the binding solution was transferred to a RNA binding column that was centrifuged. The binding column was later treated with 500 µL of wash solution I, and centrifuged; this was done to remove proteins and other soluble components. DNA was removed with on-column digestion with RNase-Free DNase provided by QIAGEN (QIAGEN, cat. no. 79254). After the DNase treatment the washing step with wash solution I was repeated, before 500 µL of wash solution II was added and the columns were centrifuged. After the washing steps RNA was eluted by adding elution buffer.
After RNA was eluted from the columns, the concentration was determined with NanoDrop™1000 spectrophotometer by measuring the 260 nm absorbance. The purity of RNA was determined by the 260/280 nm ratio (Fleige and Pfaffl, 2006). 1,5 µL of the RNase inhibitor RNasin (Progmega, cat. no. N2611) was added to all the RNA samples, before they were stored in a -80^{o}C freezer.

2.3.2 RNA integrity control

To achieve good gene expression data, the RNA integrity has to be of a certain standard (Fleige and Pfaffl, 2006). To use molecular biology techniques like qPCR and DNA microarrays intact RNA is important. RNA integrity can be tested in various ways, some methods are spectrometric methods, gel electrophoresis and lab-on-chip technology like micro-fluidic capillary electrophoresis (Fleige and Pfaffl, 2006). When inspecting the RNA integrity with Formaldehyde (FA) gel electrophoresis or on-chip micro-fluidic capillary

electrophoresis the ratio between the 28S and 18S ribosomal RNA bands are compared (Bustin and Nolan, 2004; Fleige and Pfaffl, 2006). A high 28S:18S RNA ratio indicates that there is little RNA degradation in the sample. RNA samples are considered to be of high quality if they have sharp visible rRNA bands on the gel (QIAGEN, 2010) and a ratio at ca. 2:1 of the 28S:18S rRNA bands (Schroeder et al., 2006).

Procedure:

RNA integrity was examined with two different integrity tests. One was FA gel electrophoresis with ethidium bromide staining as described in the RNeasy Mini handbook provided by QIAGEN (QIAGEN, 2010). The gel was prepared as described in appendix II. The second method incorporated the use of the Agilent 2100 Bioanalyzer (Agilent Technologies, p/n G2938-90007), which is an on-chip-electrophoresis system based on micro-fluidic capillary electrophoresis. The analysis was performed as described in the Agilent RNA 6000 Nano kit (Agilent Technologies, p/n 5067-1511).

2.3.3 cDNA synthesis

When performing qPCR, the polymerase can only use DNA as template, so to be able to carry out qPCR, mRNA has to be converted to DNA. Reverse transcriptase can together with RNA, suitable primers, correct buffers and deoxyribonucleoside triphosphates (dNTPs) synthesize complementary DNA (cDNA). The first DNA product will be a complementary strand to the RNA template that will hybridize with the RNA (Krug, 1987). The hybridized DNA-RNA molecule can be treated with RNases so that unwanted RNA molecules can be removed (QIAGEN, 2009).

Procedure:

The cDNA synthesis was performed as described in QIAGENs' QuantiTect® Reverse Transcription Kit (QIAGEN, cat.no 205313). Before performing the reverse transcription, removal of genomic DNA was performed as recommended in the QuantiTect® Reverse Transcription Handbook.

1. All the reactants listed in Table 4 where added to 0,2 mL wells in a 96 PCR well plate.

Table 4: Components in the genomic DNA removal procedure

Components	**Volume**
RNA template	1 µg (Calculated)
Nuclease free water	(Calculated)
gDNA wipeout buffer	3 µL
Total volume	21 µL

2. The PCR tubes with the reactants where incubated for 2 min at 42^{o}C, and then put on ice.

3. An aliquot of 14 µL RNA (presumably without genomic DNA) was transferred to a new well on the 96 well plate.

4. A Reverse-transcription (RT) master mix was prepared as shown in Table 5

Table 5: Components in reverse-transcription (RT) master mix.

Components	**Volume per reaction (µL)**
Quantiscript Reverse Transcriptase	0,5
Quantiscript Reverse Transcription buffer	2
Quantiscript RT-primer mix	0,5
Totalt volum	**3**

5. 6 µL RT master mix was added to the "DNA-free" RNA samples, giving a total volume of 20 µL in the wells.

6. The samples were incubated for 15 min at 42^{o}C and later in 95^{o}C for 3 min for deactivation of the Quantiscript Reverse Transcriptase.

7. 1:10 aliquots of the cDNA samples were prepared before they were stored at -20^{o}C.

8. No Reverse Transcription (-RT) samples were also made by following the procedure described above, but without adding Quantiscript Reverse Transcriptase.

2.3.4 Quantitative real-time PCR

Quantitative real-time PCR (RT-qPCR) is a sensitive and powerful method that can be used to quantify mRNA expression (Giulietti et al., 2001). The method makes it possible for direct detection of PCR product at the exponential phase of the reaction, and makes it possible to combine amplification and detection to one single step. The method is a "high throughput" technique to a certain level and requires only small amounts of samples. After the development of the PCR method a number of PCR techniques have been developed to quantify mRNA expression. Using the converted cDNA (Krug, 1987; Giulietti et al., 2001) combined with the use of fluorescent techniques has led to quantitative or real-time PCR (Ramakers et al., 2003). In RT-qPCR gene expression can be analyzed, by using cDNA as template. The technique is fluorescence-based and can detect and quantify mRNA targets of low copy number in vivo (Ramakers et al., 2003; Huggett et al., 2005; Bustin et al., 2009). This quantitative PCR technique is often preferred over other quantitative techniques because it does not rely on end-point analysis that often can be affected by enzyme instability, a decrease of reaction components in time or product inhibition.

One dye that can be used during quantification of cDNA in real-time PCR is the non-sequence-specific, double strand DNA binding specific dye SYBR® Green I (Ramakers et al.,

2003; Guénin et al., 2009). The dye can be used to monitor the increase in fluorescence intensity at the end of each PCR cycle (Ramakers et al., 2003). With the use of the fluorescence, the threshold cycle (C_t) (Ramakers et al., 2003), the number of PCR cycles required to detect the amplicons, can be determined (Korbonits et al., 2001). The cycles where the PCR product is detectable with fluorescence above background levels marks the start of the exponential detection phase, and the amplicon concentration is proportional to the initial number of template copies (Korbonits et al., 2001). The C_t values varies depending on how much target DNA there was at the starting point of the PCR reaction, so samples with a high number of target sequence will have a low C_t-value (Guénin et al., 2009). To be able to compare mRNA transcription between samples, reference genes are needed (Radonić et al., 2004). Reference genes are used to normalize against the relative expression, which is calculated based on the differences in concentrations of target gene between the samples (Pfaffl et al., 2002; Bustin et al., 2009).

Procedure:

The cDNA obtained from the cDNA synthesis was used in qPCR. qPCR was performed with the LightCycler® 480 instrument from Roche Applied Science, with LightCycler® 480 SYBR Green Master kit (Roche Applied Science, cat.no 04707516001) made from the same producer. The procedure was done according to instructions provided from manufacturer (Roche Applied Science, 2011). The reactions were done in a LightCycler® 480 Multiwell Plate 96. All the samples and the plates were on ice, until the plates were added onto the LightCycler® 480 instrument.

1. Master mixes for the qPCR reactions were prepared in 1,5 mL eppendorf tubes with the components given in Table 6

Table 6: Reaction mix for qPCR

Components	**Volume per reaction (µL)**
Autoclaved distilled water	3
PCR primers 10µM (forward and Reverse)	2
LightCycler 480 Probes Master	10
Total volume	**15**

2. A volume of 15 µL master mix and 5 µL sample was added to each of the wells in the 96 well plate. The same volume of qPCR reaction mix was used for the –RT samples too. Each primer set also had a No Template Control, which was a well that contained 15 µL master mix and 5 µL with autoclaved distilled water.

3. The 96 well plate was sealed with a LightCycler® 480 Multiwell Sealing Foil and centrifuged at 1500 x g for 2 minutes, after all the samples had been applied to the well.

4. The samples were loaded into the LightCycler® 480 instrument (Roche Applied Science), with the cycle settings shown in Table 7

Table 7: qPCR cycle settings.

Step	Temperature (°C)	time	Cycle
Pre-Incubation	95	5 min	1
Amplification	95	10 sec	45
	55	10 sec	
	72	10 sec	
Melting Curve	95	5 sec	1
	65	1 min	
	97	-	
Cooling	40	10 sec	1

11 genes were chosen in total for qPCR analysis. 3 of these genes, SrExp, SrSec7 and qSrVPS35 were chosen as reference genes. SrVPS35 was chosen because it had been used in a previous study done on *S.robusta* by Midthun (Strøm Midthun, 2012). SrExp and SrSec7 were chosen after studying unpublished microarray data from *Phaeodactylum tricornutum* done by Marianne Nymark (Nymark, 2013). 5 of the target genes, SrLHCR6, SrLHCF8, SrD9, SrPAS9 and SrHSF4 were chosen based on studies done on *P. tricornutum* by Nymark. The remaining 3 target genes were chosen from studies done by Midthun (2012) that might have a function related to light adaptation (Strøm Midthun, 2012). Primers were made for the selected *S. robusta* genes (othologues of the light induced genes identified in *P. tricornutum*). The primers that were used are listed in appendix III.

2.3.5 qPCR data analysis

The raw data from the qPCR run were analyzed by using the software LinRegPCR (Ruijter and Ramakers, 2003), Rest2009 (Pfaffl, 2009) and qBase^{Plus} (Hellemans et al., 2007). By using the Tm Calling Analysis Module in the the LightCycler® 480 software, melting curves were calculated. The melting curves were examined for non-specific products, and primer dimers. The –RT samples were examined for the presence of genomic DNA.

1. LinReg was first used to determine a baseline fluorescence and uses a linear regression method to fit data from the PCR data set. The mean PCR efficiency was decided for each primerset as efficiency per amplicon group. The C_t values and baseline were determined in LinRegPCR.

2. The C_t values from LinRegPCR were then imported in to Rest2009 and qBase^{Plus}. The samples with dark treatment were set as control when comparing the treatments. The PCR efficiency for each group was also imported to both Rest2009 and qBase^{Plus}.

3. The relative quantification data were normalized to the reference genes. In Rest2009 a pair-wise fixed reallocation randomization test was used to see if there was any significance in the expression ratios. Expression ratios with p-values less than 0.05 were considered as significant up or down regulated compared to the control. In qBase^{Plus} it was possible to

analyze multiple samples and subgroups simultaneously, with the same test, something that is not possible in Rest2009.

2.3.6 Microarray

Microarray is a method that is used to study gene expression, DNA sequence variation, protein levels, tissues, cells in a big parallel format (Stears et al., 2003). When doing expression (transcript) profiling, the microarray is based on parallel quantification of large numbers of mRNA transcripts (Schulze and Downward, 2001). The principle of microarray is that mRNA isolated from cell samples or tissue is synthesized to cDNA (in some cases to complementary RNA, cRNA (Agilent Technologies, 2010)), labeled and later hybridized to DNA fragments that are attached to a solid surface in an ordered array (Schulze and Downward, 2001). With microarray many thousands of genes can be detected and quantified at the same time. The microarray system often used today can be divided into two groups. They are cDNA- and oligonucleotide based microarrays. Although array techniques can be used to study patterns of gene expression there are some differences between the methods. On cDNA microarrays probes are made though PCR and usually printed (spotted) on the glass slide or on nylon membranes as spots at defined locations. With this array technique a microscope slide can be fitted with 30 000 cDNAs (Schulze and Downward, 2001). The oligonucleotide arrays contain single stranded oligonucleotide probes and are not dependent on clones, PCR product or cDNA (Lipshutz et al., 1999). The oligonucleotide arrays can consist of 20-100 mers that have a unique part of a given transcript where they can hybridize (Schulze and Downward, 2001). The oligonucleotides on the arrays are synthesized *in situ* thought either ink-jet technology or photolithography onto silicon wafers, and if the oligonucleotides are pre-synthesized they can be printed onto glass slides. In this experiment an oligonucleotide array made by Agilent Technologies was used (Agilent Technologies, 2010) and a one-color microarray-based gene expression analysis were performed. With the one-color microarray approach fluorophore labeled (e.g cyanine-3 (cy-3)) cRNA from test samples and controls are hybridized to separate microarrays (non-competitive hybridization) (Oberthuer et al., 2010). In comparison, two-color microarray two samples are labeled with two different fluorophores (e.g Cy-3 and Cy-5) and hybridized on the same array. The results that are extracted from these two procedures are also different. In on-color microarray the abundance of the various mRNA in the sample is presumed to be visualized through absolute fluorescence intensity emitted from probes on the array. In two-color microarray however the fluorescence intensities show the different ratios of mRNA between two the samples for each probe (Oberthuer et al., 2010).

Procedure:

The procedure of the microarray experiment was performed by principal engineer Torfinn Sparstad, and the procedures were done as instructed in the One-Color Microarray-Based Gene Expression Analysis, Low Input Quick Amp Labeling kit (Agilent p/n 5190-2305) provided by Agilent Technologies (Agilent Technologies, 2010).

Samples of total RNA isolated form *S. robusta* treated with white light (WL 0,5h I-IV and WL 6h I-IV), blue light (BL 0,5h I-IV and BL 6h I-IV) and darkness (Dark I-IV) were used in the microarray experiment. Two biological samples for each treatment were used for the microarray experiment, for the qPCR analysis four biological replicates were used.

Total RNA was first reverse transcribed. A dilution of 10µL with 200 ng total RNA was made for all samples. The diluted samples were reverse transcribed to make cDNA. The cDNA was transcribed with the presence of cy-3 producing labeled complementary RNA (cRNA). The reverse transcription, labeling and production of cRNA were done using the Spike-In Kit, One Color (Agilent p/n 5188-5282). The labeled cRNA was purified by using RNeasy® Plant Mini Spin (50) (Qiagen, cat.no 74904), and quantified with NanoDrop 1000 spectrophotometer using microarray measurement settings. The hybridization of cRNA was done by using Gene Expression Hybridization Kit (Agilent p/n 5188-5242). The cRNA was diluted to an appropriate concentration (1650 ng) as recommended in the Gene Expression Hybridization Kit. Samples was later fragmented and 100 µL was hybridized to a 4x44k 60mer oligonucleotide microarray (Agilent Technologies, cat.no 0305096251) designed for *S. robusta* by Associate Professor Per Winge. Slides were set for hybridization for 22 hour at 65^{o}C in a rotation oven. After hybridization slides were washed twice with Gene Expression Wash Buffer 1 (ambient temperature) for 1 min, followed with 1 min wash in Gene Expression Wash Buffer 2 pre-warmed to 37°C. Both wash buffers were included in the Gene Expression Wash Buffer Kit (Agilent p/n 5188-5327). Following the washing the slides were put in a slide holder, with the Agilent barcode faces up. The slides were scanned with Agilent DNA microarray scanner (Agilent Technologies) with 5 µm resolution. Scanned images were analyzed with Agilent Feature Extraction software version 9.5.

2.3.7 Microarray data analysis

The data obtained from the Feature Extraction files were analyzed with the Limma package (version 3.20.1; Smyth et al 2005), R version 3.0.3 and were used for statistical analysis and identification of significant differentially expressed genes. No background subtraction was performed and spots identified as feature outliers were excluded from the analysis. Genes with adjusted P-values less than 0.01 were regarded as statistical significantly differentially expressed.

WORK FLOW FOR THE LIGHT EXPOSURE EXPERIMENT

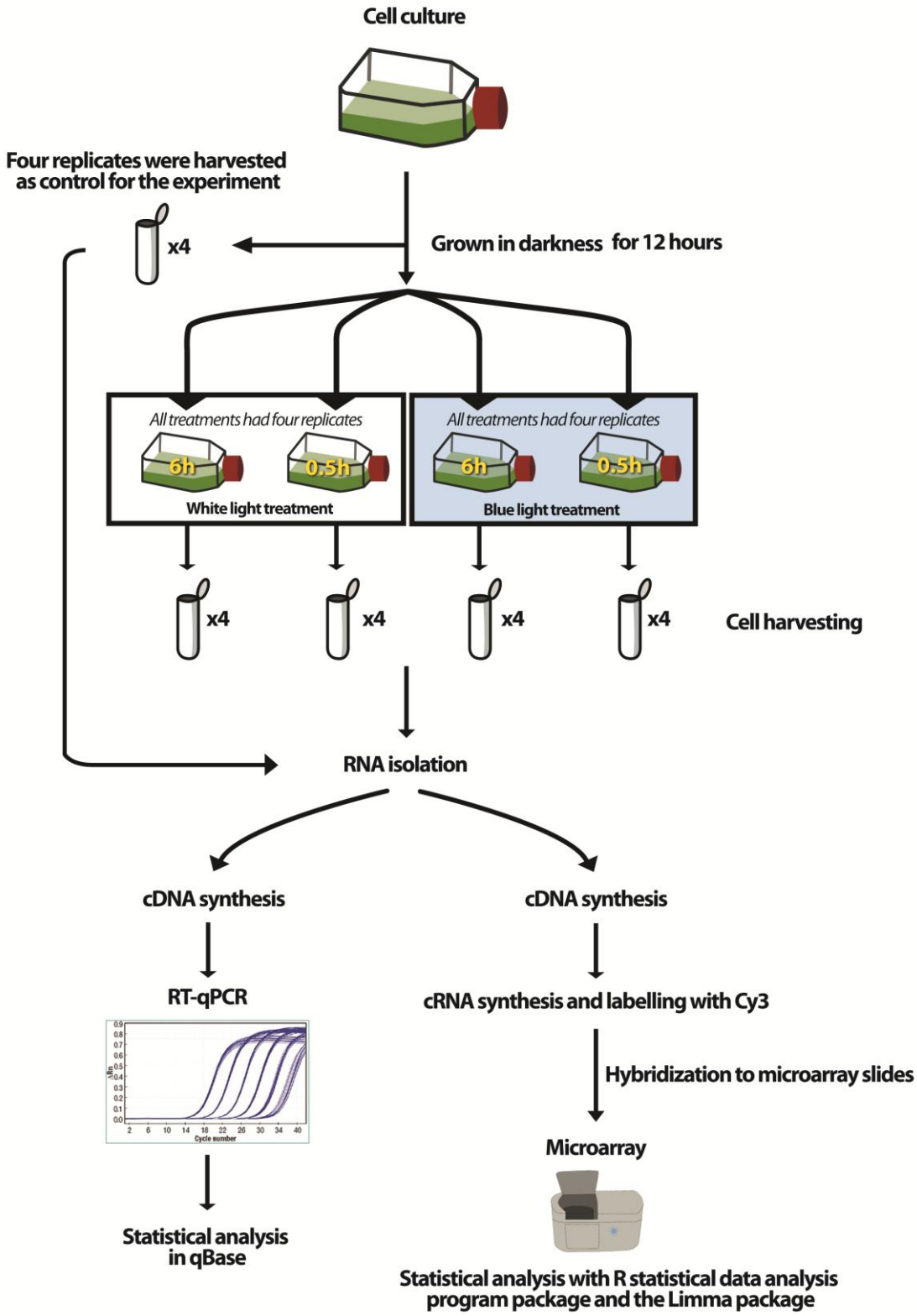

Figure 6: Schematic illustration of the light exposure experiment. Cells where first grown in 12 hours in darkness and immediately transferred to their light treatment. After the treatments cells where harvested and RNA isolated. Sample preparations of RNA were later done for analysis with RT-qPCR and Microarray.

3 Results

The aim of this study was to construct a bidirectional vector model that could be used to transform the diatom *Seminavis robusta*. A suitable, strong, constitutive and endogenous bidirectional promoter (AtpE/B) coding for the Atpase beta and epsilon/delta subunits with its two associated terminator sequences (AtpBt, AtpEt) were chosen (Tore Brembu unpublished results). The sequences where integrated stepwise into the pBluescript KS+ (pBKS) vector. To test the bidirectional transcription of the promoter AtpE/B, a selection marker coding for *nourseothricin* resistance (*nat1*) and *yellow florescence protein* (YFP) which worked as a reporter gene was placed on either side of the promoter. The "atpE/D promoter" driving *nat1* gene expression. The vector was later transformed into *S. robusta* through biolistic transformation. A gene expression analysis was also done on *S .robusta* with 8 candidate genes. The genes where chosen based on previous expression studies of cytochrome P450 genes in *S. robusta* by Elise Midthun (Strøm Midthun, 2012) and results from a light study done on *Phaeodactylum tricornutum* by Marianne Nymark (Nymark, 2013). The genes were chosen as candidate genes and may in the future be cloned into the expression vector for further functional studies. The expression of the candidate genes were investigated by exposing the cells for different light conditions as well as 12 hour darkness. After 12 hour darkness the cells were treated with blue light for 0,5 hours and 6 hours and white light for the same duration. RNA was later isolated and used for qPCR and DNA microarray analyses.

3.1 Designing the pBKS-NAT1-AtpBE-YFP vector

The cloning strategy of the making of the pBKS-NAT1-AtpBE-YFP was done step wise. The steps consisted of stepwise cloning of DNA fragments into the vector contruct. Figure 7 illustrates in four steps how the cloning strategy was executed.

.

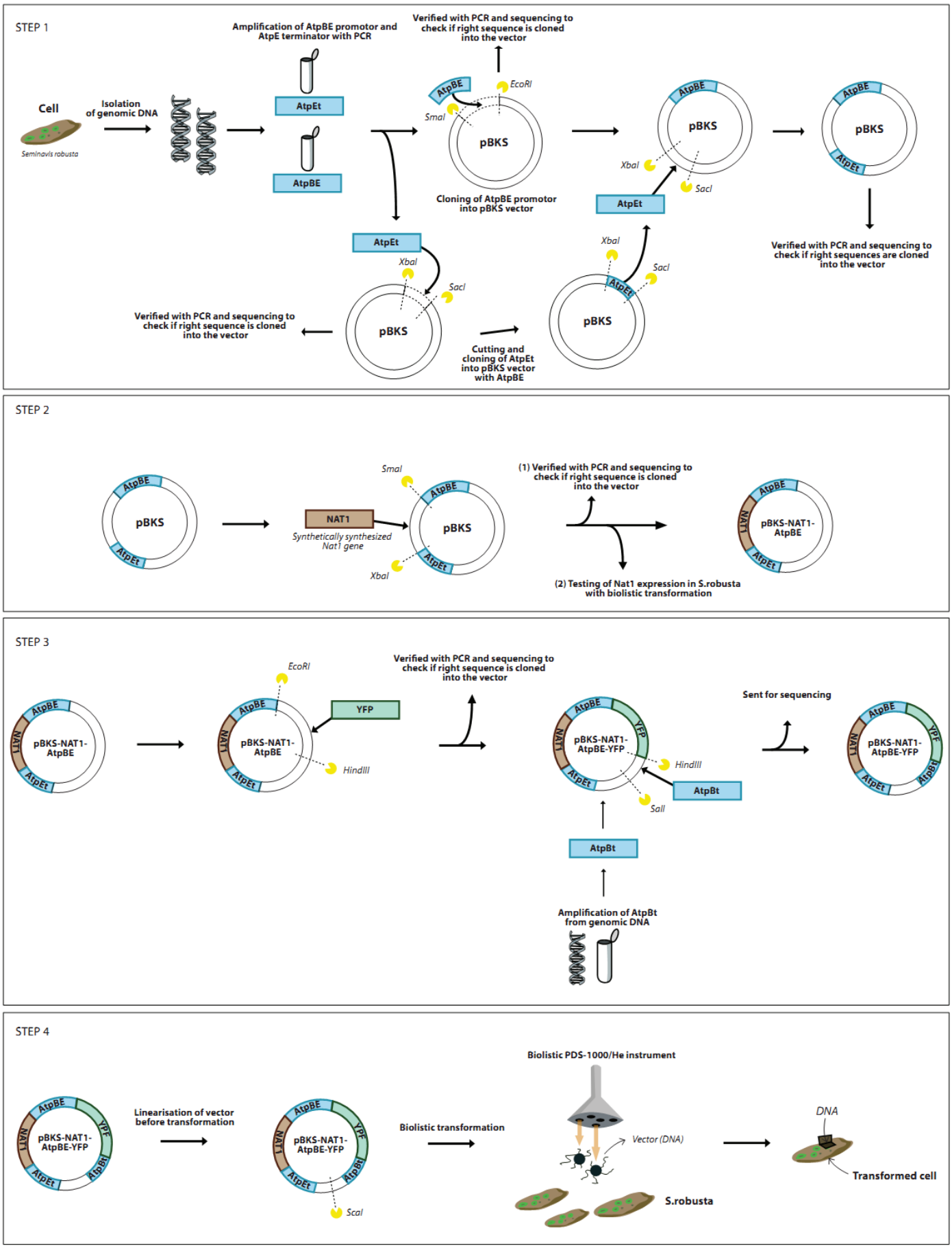

Figure 7: Schematic illustration of how the pBKS-Nat1-AtpBE-YFP construct for *S. robusta* was made. All the steps are simplified; restriction enzymes are visualized as small yellow figures. The vector construct was made stepwise. AtpBE ans AtpEt was first inserted into the same vector before step two where Nat1 was cloned into the vector and tested for expression in *S. robusta* by transformation. In step 3, YFP and AtpBt was cloned into the vector. In step 4 the vector was cut with ScaI to make the vector linear. It was later transformed into *S.robusta* with biolistic transformation.

First the promoter sequence was amplified with PCR from genomic DNA isolated from *S. robusta*. The PCR product was verified by sequencing and the presence of polymorphisms was examined. The same procedures were done with the AtpE terminator, amplification from genomic DNA, verification by sequencing and examination of polymorphisms. The next step consisted of cloning the promoter and terminator into separate pBKS KS(+) vectors. The vectors were later propagated in DH5α *E.coli* cells by transformation and later isolated. Since the both the AtpBE promoter and AtpE terminator had two or more polymorphisms, one of the polymorph sequences were chosen for further downstream applications. The terminator sequence was cloned into the same vector as the atpBE promoter sequence. The cloning of the two fragments into the pBKS vector were verified by sequencing. The next step was to clone the selective marker gene into the vector. Before the *nat1* gene could be cloned into the pBKS vector it was codon optimized to match the tRNA abundance in *S. robusta* (the gene was synthesized by Eurofins/MWG and delivered in a pEX-A2 vector). This was done by examining the general codon usage in *S. robusta* which was available from transcriptome and genome data (unpublished). The modified *nat1* gene was synthesized with appropriate restriction sites on both sides of the gene, ready for cloning. After the *nat1* gene had been cloned into the pBKS vector and verified the pBKS vector construct was transformed in *S.* robusta to check if the AtpE direction of the promoter was functional and capable of expressing *nat1*. After the gene expression of *nat1* was confirmed in the transformed *S. robusta* cells the functionality of "AtpB direction of the promoter" was analyzed. First the AtpB terminator (AtpBt) was amplified with PCR from genomic DNA from *S. robusta* and sent for sequencing to verify the PCR product. After the verification AtpBt was cloned into the pBKS-NAT1-AtpBE vector. To check the functionality of the AtpB promoter Yellow fluorescence protein (YFP) was cloned into the vector construct between the AtpB promoter and AtpB terminator. The YFP was obtained from the pEarleyGate 104 (pEG104) plasmid (Earley et al., 2006). When amplifying YFP from pEG104 the primers were restriction site-modified so that the EcoRI and HindIII restriction sites where included in the amplified YFP genes. After YFP had been cloned into the vector construct, the YFP and atpBt sequences were verified by sequencing. The vector was transformed into *S. robusta* cells and a strong fluorescence from the YFP protein would provide a check if the AtpB direction of the promoter was functional. The results from each of the steps during the vector construction are described in detail further down.

3.1.1 PCR amplification, verification and cloning of AtpE/AtpB, AtpEt and *Nat1* into pBKS

PCR products of the bidirectional AtpBE promoter and the associated gene terminators were amplified from genomic DNA using restriction site-modified PCR primers (primers are listed in appendix III).

Cloning of AtpEB and AtpEt into pBKS vector

The restriction sites used during the vector construction were included into the sequences through restriction-site modified PCR primers so that the sequences could be cut with specific restriction enzymes and cloned into the polylinker of the bluescript vector. The PCR amplified AtpBE promoter and AtpE terminator sequences were run on a agarose gel as shown in Figure 8A to confirm the presence of PCR product and to remove contaminants.
A gel extraction was done on the AtpBE promoter and AtpE terminator shown in Figure 8A by using the Wizard® SV Gel and PCR Clean-Up System (Promega). Followed by this the AtpBE promoter and atpB terminator were cloned into separate pBKS vectors. AtpBE promoter and a pBKS vector were cut with restriction enzymes EcoRI and SmaI and ligated together using T4 DNA ligase. The same procedure was done with the AtpE terminator and a pBKS vector but with the restriction enzymes SacI and XbaI. After the cloning of AtpBE and AtpB into their separate vectors rubidium chloride competent DH5α *E.coli* cells were transformed with the vector constructs to maintain and amplify the vectors.

Colony PCR was performed on the transformed DH5α cells to confirm successful cloning of the PCR fragments. The forward and reverse primers for AtpBE and AtpEt were used to verify the inserts; the expected fragment lengths for AtpBE and AtpEt were 448 bp and 295 bp respectively. The PCR product showed bands on the gel that were approximately 500 bp for AtpBE and 250 bp for AtpEt, as shown in Figure 8B and C respectively. The results of an agarose gel electrophoreses only gives an indication of the fragment size. To verify the successful cloning of AtpBE and AtpEt in the pBKS vector a small fragment of the pBKS vector construct was sequenced. The fragment from pBKS was amplified for sequencing by using the pBKS primers, T7 and KST3.

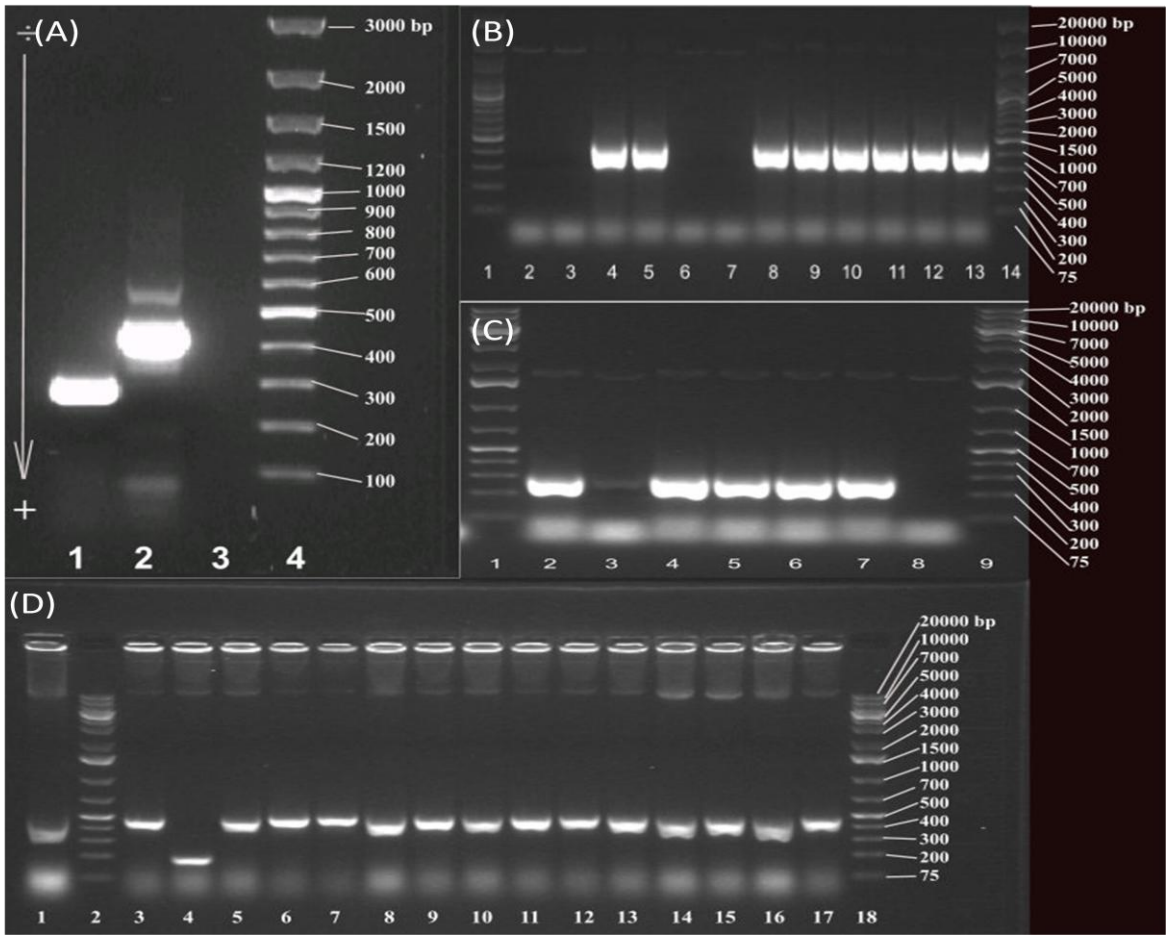

Figure 8: A) The isolation of the promoter AtpBE and the terminator atpEt after from genomic DNA with PCR. B) Colony PCR was done to confirm the insertion of the AtpBE promoter into the pBKS vector. The promoter was verified in well 4, 5, 8, 9, 10, 11, 12 and 13. C) The Verification of the atpEt terminator was done with colony PCR. It was confirmed in 5 out of 7 colonies. Well 2, 4, 5, 6 and 7 had the terminator sequence. D) The *nat1* gene was also verified with colony PCR, and the gene was present in 15 out of 16 colonies. The wells with the Nat1 gene were 1, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18.

3.1.2 Sequencing data from AtpBE

The sequencing verified that the promoter region had successfully been cloned into the pBKS vector. Results from the sequencing are shown in Figure 9 confirmed that the length of the promoter was 423bp, and that there were two types of polymorphisms in the AtpBE promoter. One of the promoter alleles had a deletion of five base pairs. There were also other differences like transitions and transversions, indicating the genetic diversity within *S. robusta* is quite high. In total four samples were sent for sequencing, two of the samples had a promoter with a deletion and two were without the deletion. The polymorphisms were also verified and detected in the sequencing reads from the genome and transcriptome data sets. For downstream applications with the vector construct the promoter sequence without the deletion was chosen.

Figure 9: Results from the sequencing of the AtpE/B promoter. The sequencing showed that there are several polymorphisms in the promoter and one of the sequences have a deletion of five base pairs. The consensus sequence of the promoter based on genomic sequence data (not published) is shown in the bottom line.

3.1.3 Sequencing data from AtpEt

The sequencing data of AtpEt showed that there were polymorphisms in the terminator. One region at the end of the sequence showed presence of a transversion between adenine and thymine as shown in Figure 10.

The vector containing the AtpEt polymorphism with thymine was chosen for the downstream applications in the vector construction and cloned into the pBKS vector containing the AtpBE promoter sequence without deletions.

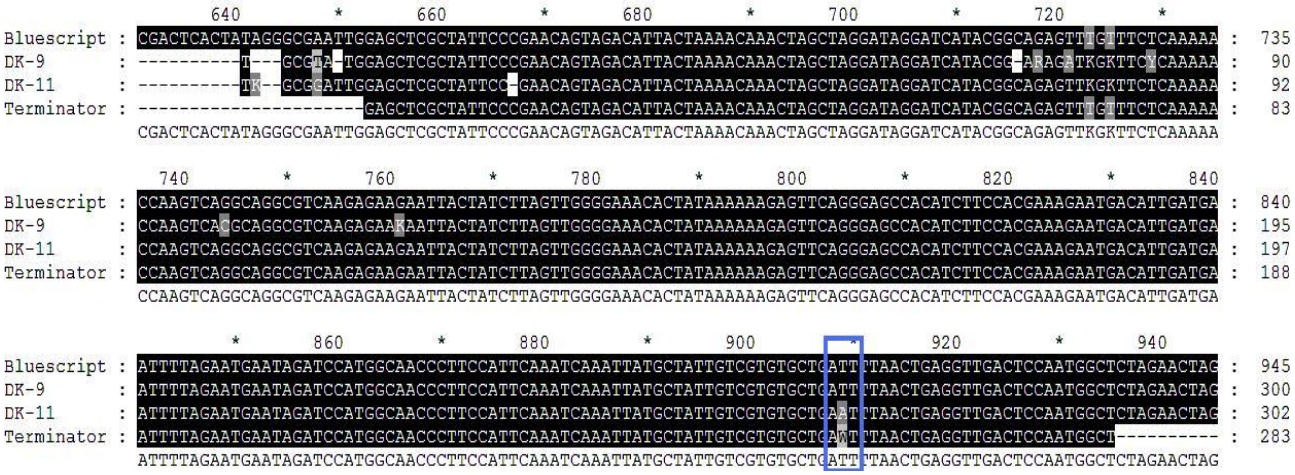

Figure 10: Results from the sequencing of the AtpEt terminator. The sequencing showed that there is one polymorphic site in the terminator. At one spot there is a transversion between adenosine and thymine.

3.1.4 Cloning of *Nat1* into the pBKS vector

The codon sequence of the selectable marker gene *nat1* was modified to correspond to the tRNA abundance in *S. robusta* (codon optimized), using compiled cDNA data from the transcriptome. The *nat1* gene was cloned into the pBKS vector containing the AtpBE promoter and AtpEt sequences. After the *nat1* gene and pBKS vector construct was cut with the restriction enzymes SmaI and XbaI and later ligated, DH5α cells were transformed with the vector. The *nat1* gene was inserted in the vector such that it was under regulation of the AtpE promoter, (between the bidirectional AtpBE promoter and the AtpE terminator). The Insertion of the *nat1* gene in the pBKS vector with AtpBE and AtpEt was pointed out with colony PCR and gel electrophoresis, shown in Figure 8D. A region of the vector containing the *nat1* gene and the AtpBE promoter was amplified and sent for sequencing, which verified the insertion of *nat1*.

To check if the AtpE promoter and AtpEt terminator were functional and capable of giving a proper expression of the *nat1* gene, biolistic transformation of *S. robusta* with the vector construct was performed. A schematic description of the construct is illustrated in Figure 11.

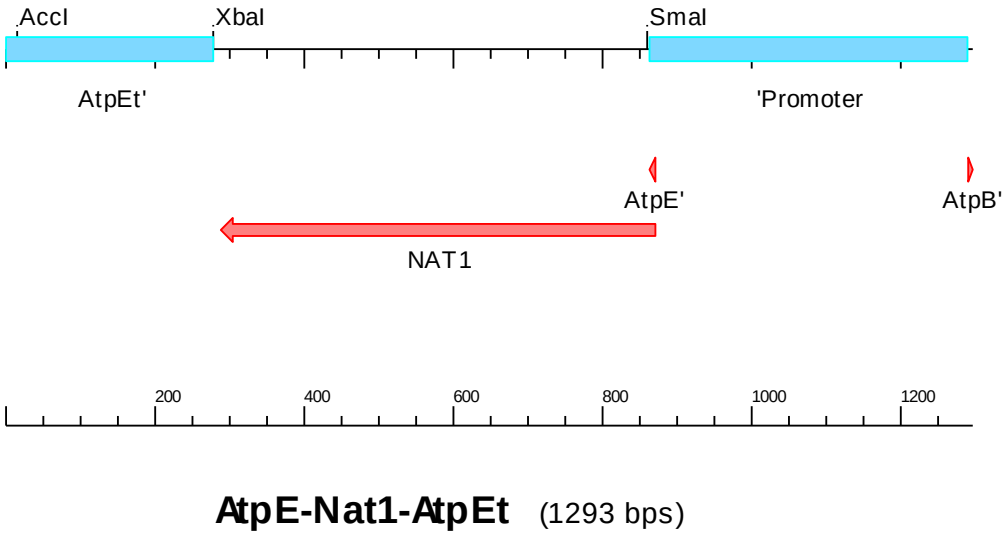

Figure 11: Part of the pBKS vector that was transformed and expressed in *S. robusta*. AtpEt: terminator sequence. *nat1* selection marker under control of the AtpE promoter,

Nine Petri dishes with *S. robusta* were used for biolistic transformation. Six of the dishes "shot" with circular vector, and three dishes were "shot" with linearized vector. Since previous studies had indicated that the transformation efficiency was higher with linear vectors this was also tested. The linearization was done with the use of the restriction enzyme ScaI. ScaI was chosen because it did not cut within the area that was intended to be expressed. Out of the 9 Petri dishes with *S. robusta* only one dish contained transformants. None of the cells transformed with circular vector showed any signs of transgenetic properties. After 3 weeks on selectable medium containing nourseothricin (NTC) resistant cells were transferred to a liquid medium containing 100, 150, 200 and 250 µg/mL of NTC. An additional control to verify presence of the *nat1* gene was done with colony PCR on the cells. The primers for the *nat1* gene were used to detect the presence *nat1* transcription (571bp), the atpEt primers were also used, as positive control. The results are shown in Figure 12 in wells 4-9, confirming that the cells have the *nat1* gene inserted. In well 2 and three the AtpEt terminator is shown.

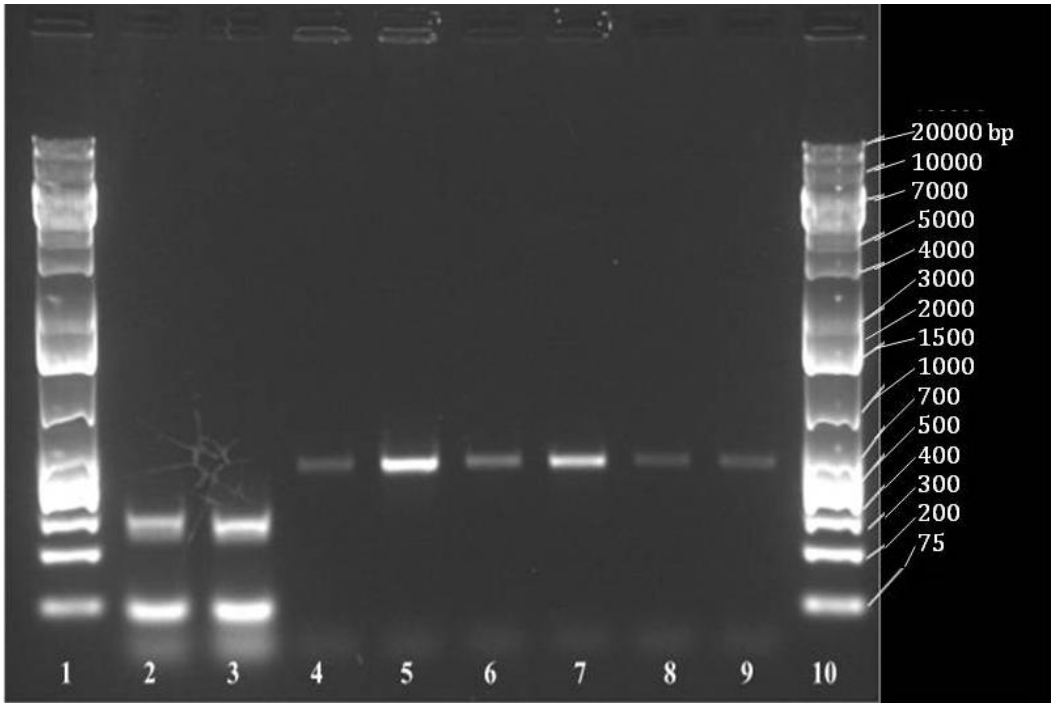

Figure 12: PCR products after performing colony PCR on transformed *S. robusta*. cells. Well 1 and 10 contain 1kb ladder. Well 2 and 3 show the length of the terminator AtpEt, while wells 4-9 show the *nat1* gene with parts of the terminator sequence. No negative control was included in the colony PCR.

3.1.5 Isolation of AtpB terminator and Yellow Fluorescent Protein

After the verification of the vector construct containing the AtpE-NAT1-AtpEt module, the AtpB terminator and YFP gene were isolated. The AtpBt was PCR amplified from *S. robusta* genomic DNA, while YFP was amplified from a pEG104 vector. Both AtpBt and YFP were amplified through PCR with primers that introduced restriction sites to the amplified PCR products. Gel photo in Figure 13 shows the isolated PCR products. The PCR products were later isolated from the gel, purified and cloned into the AtpEt-Nat1-AtpE/B vector.

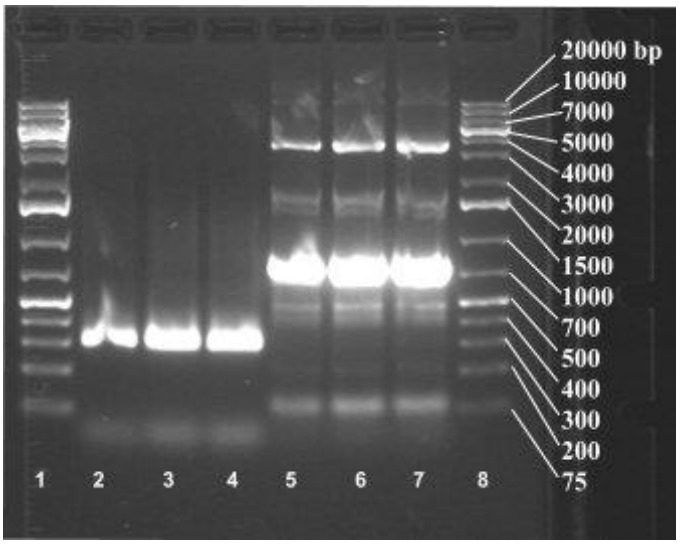

Figure 13: Isolation of atpBt (281 bp) and YFP (728 bp). The atpB terminator is located in well 1-3, while YFP is located in wells 5-7

The YFP gene was first cloned into the AtpEt-Nat1-AtpBE vector before the terminator sequence, AtpBt, was inserted. To control that both fragments had been successfully cloned a restriction fragment analysis was performed. Wells 4-6 contained samples of the AtpEt-Nat1-AtpBE-YPF vector, and was used to compare against the fragments in wells 2,3 and 8-13, which were tested for the presence AtpBt. Well 8 was a control sample, where a pBKS vector only containing the AtpEt-Nat1-AtpBE was cut to visualize the size different between a vector with and without the YPF insert. The smallest fragments in wells 2,3 and 8-13 indicated that both YFP and AtpBt could be present in the AtpEt-Nat1-AtpBE vector. The expected length of the region XbaI and SalI cut was 2015 bp if AtpBt was present, in the absence of AtpBt the length would be 1734 bp. Because small differences in the wandering of the bands the comparison of the fragments was difficult, however four samples (Wells 1,2, 4 and 7) thought to have both YFP and AtpBt were sent for sequencing.

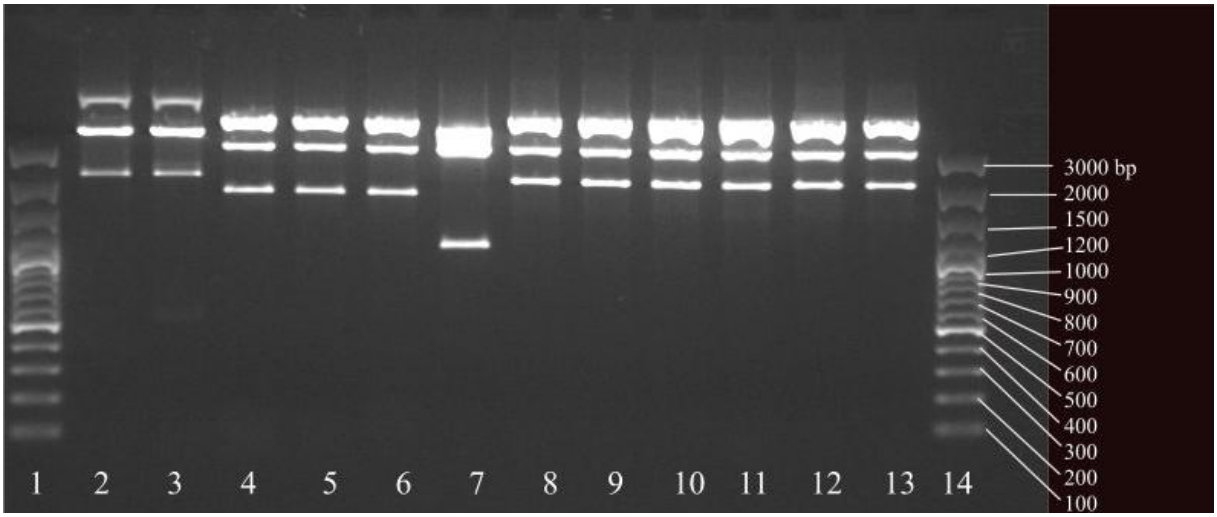

Figure 14: Cutting test with the restriction enzymes XbaI and SalI to check if YFP and AtpEt is inserted into the vector. The expected length of the lowest band was 2015 bp if AtpBt was present. In the absence of the AtpBt the length was expected to the 1734 bp. Wells 2,3 and 8-13 where tested for the presence of AtpBt and YFP. Well 7 was a negative control where the lowest band visualized the region between XbaI and SalIs cutting site without YFP and AtpBt. The lowest bands in wells 4-6 show the length between the XbaI and SalI cutting sites without the AtpBt. Wells 4-6 where used to compare against the other wells (except well 8) to detect the presence of AtpBt. Because of small differences in the wandering of the bands the comparison of the fragments was difficult, however four samples (Wells 1,2, 4 and 7) thought to have both YFP and AtpBt were sent for sequencing for verification.

The sequencing confirmed that the YFP and AtpBt sequences were present in the vector. Figure 15 shows the results from the sequencing.

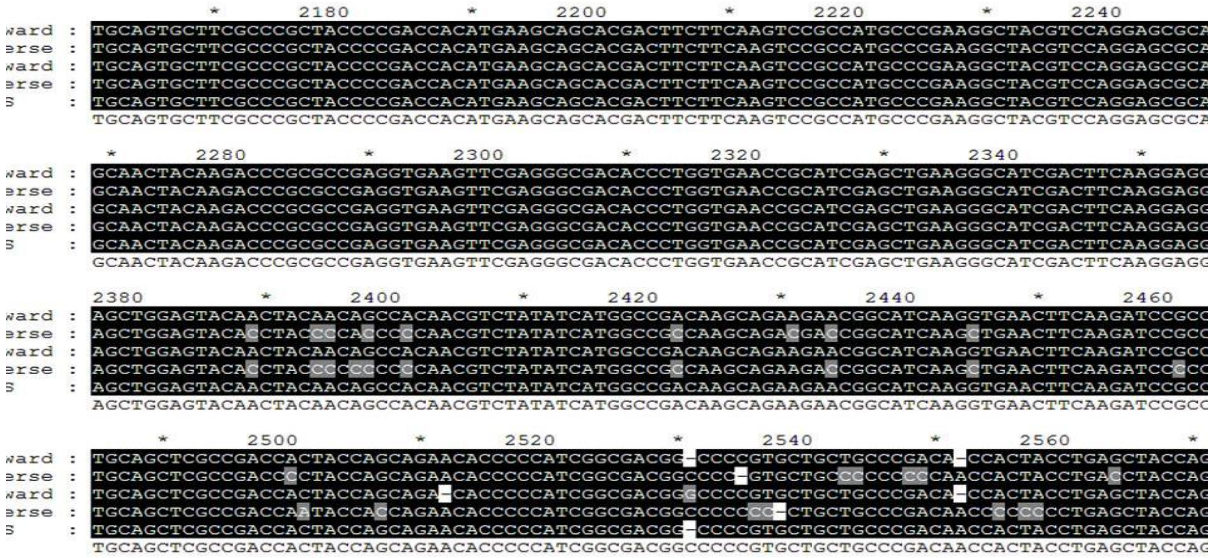

Figure 15: The sequencing of the pBKS-Nat1-atpBE-YFP verified that the atpb terminator and YFP were inserted into the vector. The alignment only show the part of the vector sequence where the YPF gene ends and the AtpB terminator starts. The sequence obtained from the genome sequencing project is shown in the top row.

The vector construct with all the inserts included is illustrated in Figure 16. Restriction sites that do not interfere with the coding regions are included. Figure 16 illustrates the final composition of the vector construct before it was transformed into *S. robusta.*

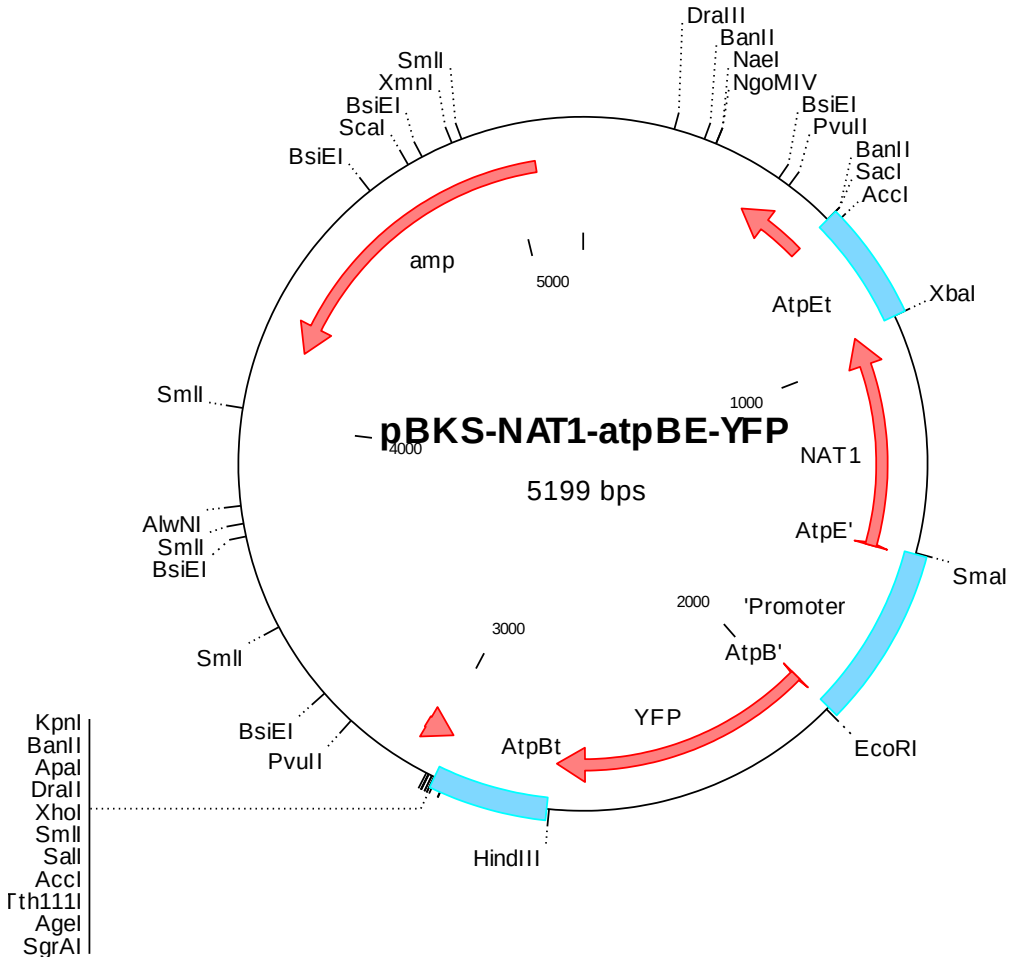

Figure 16: Vector construct with all inserts included.

3.1.6 Transformation of *S. robusta* with the pBKS-NAT1-AtpBE-YFP vector

The pBKS-NAT1-AtpBE-YFP vector was linearized with the ScaI restriction enzyme and the fragment purified (as described in 2.2.3 DNA purification) and tungsten particles were coated. The Petri dishes had a density of $1x10^5$cells/mL. In total there were four Petri dishes and each dish was shot five times with DNA coated tungsten particles. All the plates were marked with a number, the first plate was number 1 and the last one 4. The cells were set in a culture room for two days after transformation and transferred to a selective medium the third day after transformation and put back in a culture room for 2-3 weeks.

One visible colony formed in Petri dish number 4 as a light brown cloud. The colony was picked and transferred to a six wells plate containing liquid f/2 medium. The wells contained different concentrations of nourseothricin, respectively 0, 100, 150, 200, 250 and 300 µg/mL. There was sign of viable cells in all of the wells, and cells where later transferred to culture flasks.

3.1.6.1 Expression of the YFP gene and detection of YFP protein by confocal microscopy

When the cells had grown densely in the flasks 1mL f/2 of the cell culture was extracted from a flask with transformed cells to inspect the cells by confocal microscopy. Under the appropriate settings on the microscope the cells were studied for the presence of YFP.

During the microscopy no cells with YFP expression were detected, visualized in Figure 17. In Figure 17A fluorescence would have been detected if YFP protein was correctly processed and expressed in sufficient quantities. In Figure 17B the cell is visualized through a normal lens while Figure 17C shows the autofluorescence that is emitted from the cell.

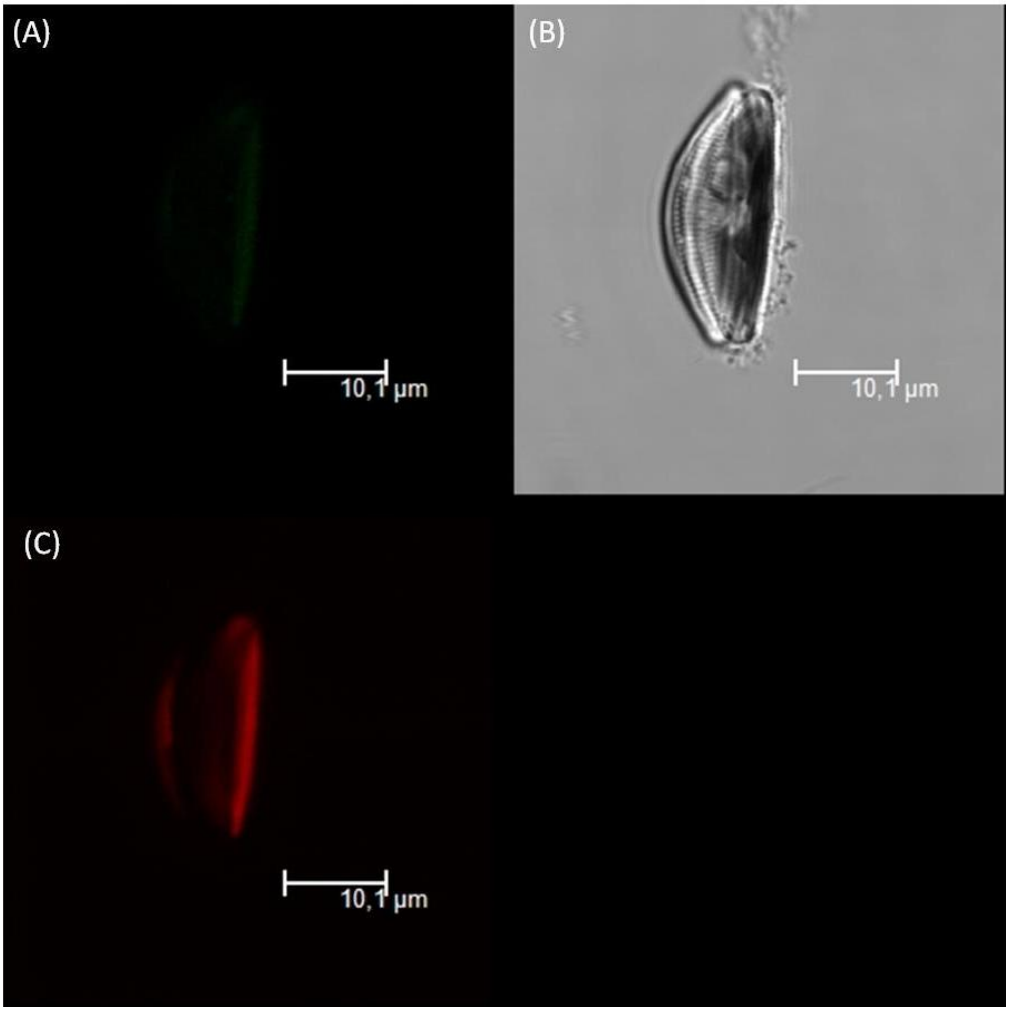

Figure 17: A) A transformed *S. robusta* cell visualized through confocal microscopy. If the YFP expression was sufficient and protein processing was correct fluorescence from the protein would have been detected. B) *S. robusta* seen through a normal lens. C) Autofluorescence emitted from *S. robusta*.

None of the cells that were viewed through the microscope showed YFP fluorescence.

3.1.7 Reverse Transcriptase PCR of transformed *S. robusta* RNA

Since it was not detected any YFP expression in the *S. robusta* cells, reverse transcriptase PCR (RT-PCR) was performed to verify that the AtpB promoter was able to drive the expression of the YFP gene. RNA was isolated from the transformed cells and reverse transcription was performed on the isolated RNA. PCR was later done on the cDNA from the transformed cells with primers that were complementary to the YFP gene. The PCR did not produce any bands when the PCR product was applied on agarose gel electrophoresis as shown in Figure 18

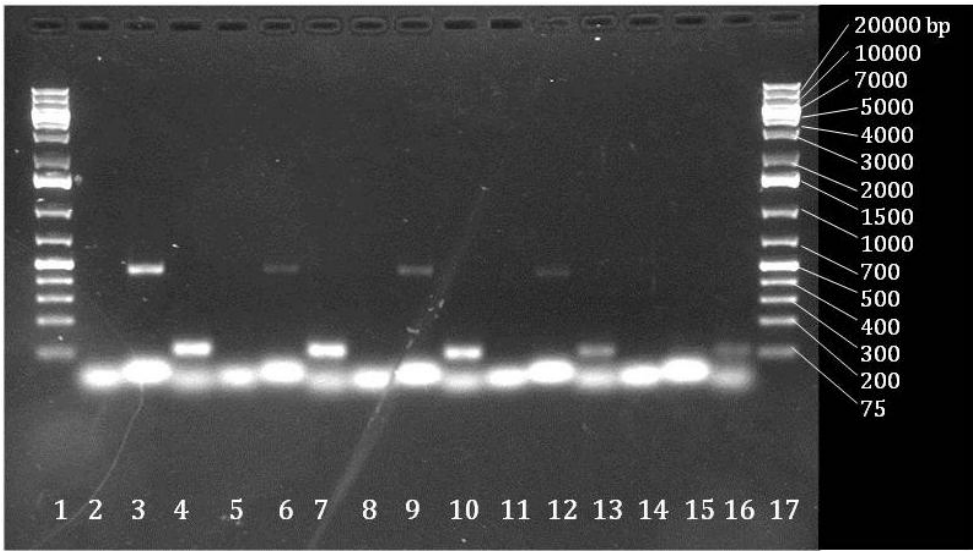

Figure 18: RT-PCR of transformed *S. robusta* cells. No YFP transcripts (728 bp) were detected, but the *nat1* gene (581 bp) was present in all the transformed cells. Cells transformed with the YFP gene would have shown bands in wells 1, 4 and 7 if the YFP gene was expressed in the cells. Wells 2, 5, 8 and 11 show the presence of the *Nat1* gene while wells 3, 6, 9, 12 and 15 were positive controls and show the presence of the endogenous gene exportin-1 (88 bp). The ladder used to evaluate the length of the fragments was a 1kb GeneRuler ladder (Fermentas Life Sciences).

3.2 Light exposure experiment

The expression of heat shock transcription factors (HSF), some light harvesting complexes (LHCs), a fatty acid desaturase and an aurechrome-like gene were studied after exposure to blue and white light. The expression measurements from the blue light (BL) treated samples were then compared with cells harvested from white light (WL) exposure. The cells were exposed for two different time periods in the two light treatments, 0,5 and 6 hours. As control for the experiment, cells harvested after 12 hour darkness (Dark) were used as control against the blue light and white light samples. All the treatments had four biological replicates given a name after the light treatment, exposure time and replica number, BL0,5h I-IV, BL6h I-IV, WL0,5h I-IV and WL6h I-IV. The control samples were given the name Dark I-IV.

3.2.1 RNA integrity and quality test

After the isolation of RNA from the light exposed cells, the RNA quantity was measured with NanoDrop 1000 spectrophotometer. All the samples had applicable quality values of RNA on the NanoDrop. To further control the integrity and quality of RNA the samples were run on a FA gel and on the Agilent 2100 Bioanalyzer. Both techniques control the integrity of the RNA, but the Agilent 2100 Bioanalyzer is a newer, more exact and less time consuming technique. The two techniques were compared to see if they gave the same results on the RNA integrity.

Results from the FA gel showed that all the samples had intact RNA that was not degraded, seeing that the 28S rRNA bands were sharper than the 18S rRNA bands and there were no clear smears were detected. The results from the FA gel electrophoresis visualizing the integrity of the RNA samples are shown in Figure 19.

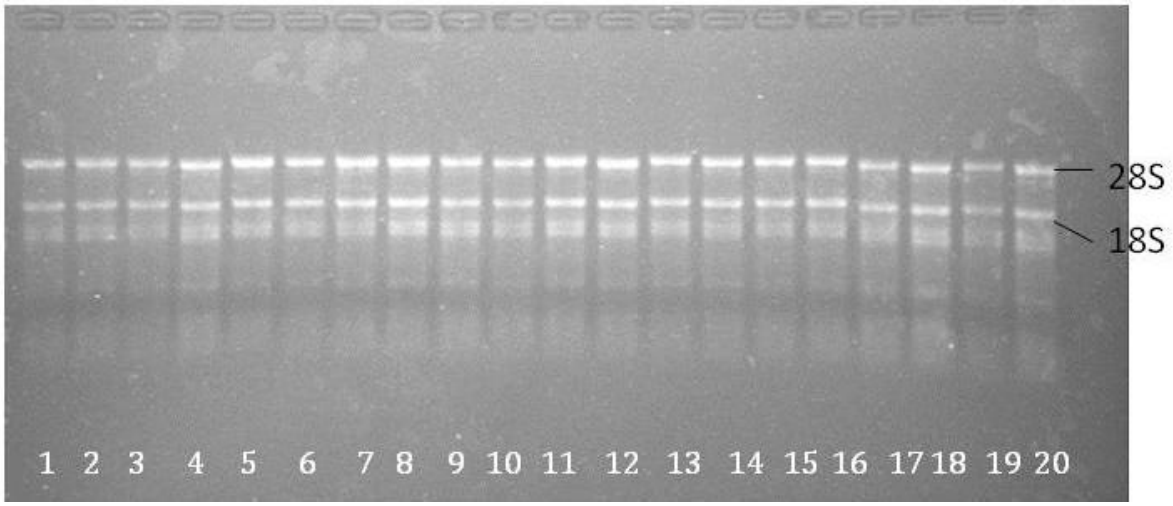

Figure 19: FA gel showing the integrity of total RNA of all the extracted samples. Wells 1-4 are biological replicates of for cells exposed in blue light (BL) for 0,5 hours, BL I, BL II BL III and BL IV respectively. The following next four wells, 5-8 were total RNA from cells exposed in blue light for 6 hours (BL 6h). The biological replicates were in the order BL 6h I, BL 6h II, BL 6 III and BL 6 IV. Wells 9-12 show the integrity of RNA isolated from the four replicates exposed to 0,5 hours in white light (WL 0,5h), WL 0,5h I, WL 0,5 II, WL 0,5 III and WL 0,5 IV. The RNA integrity of the white light, 6 hours treated cells (WL 6h) WL 6h I, WL 6h II, WL 6 III and WL 6 IV are visualized in wells 13-16. RNA from the cultures treated with 12 hours darkness (Dark) which was used as control samples were placed in wells 17-20 in the order Dark I, Dark II, Dark III and Dark IV. After evaluating and comparing the 28S rRNA bands against the 18S rRNA bands none of the samples were considered to be degraded.

The same test done on the Agilent 2100 Bioanalyzer also confirmed that the RNA quality was applicable for further applications. Agilent 2100 Bioanalyzer uses software algorithm that calculates RNA Integrity Numbers (RIN) values, which is used to give a quality estimate of the RNA. The RIN values can range from 1 - which means the RNA is completely degraded to 10 – which indicate RNA with no degradation. Samples with RIN values above 7 are in general thought to be decent RNA. Samples BL 0,5h II, BL 0,5h III, BL 6h I and BL 6h II had a RIN value above 8. The rest of the samples had RIN values between 7 and 8, except for sample Dark IV which had a RIN value of 6,9.

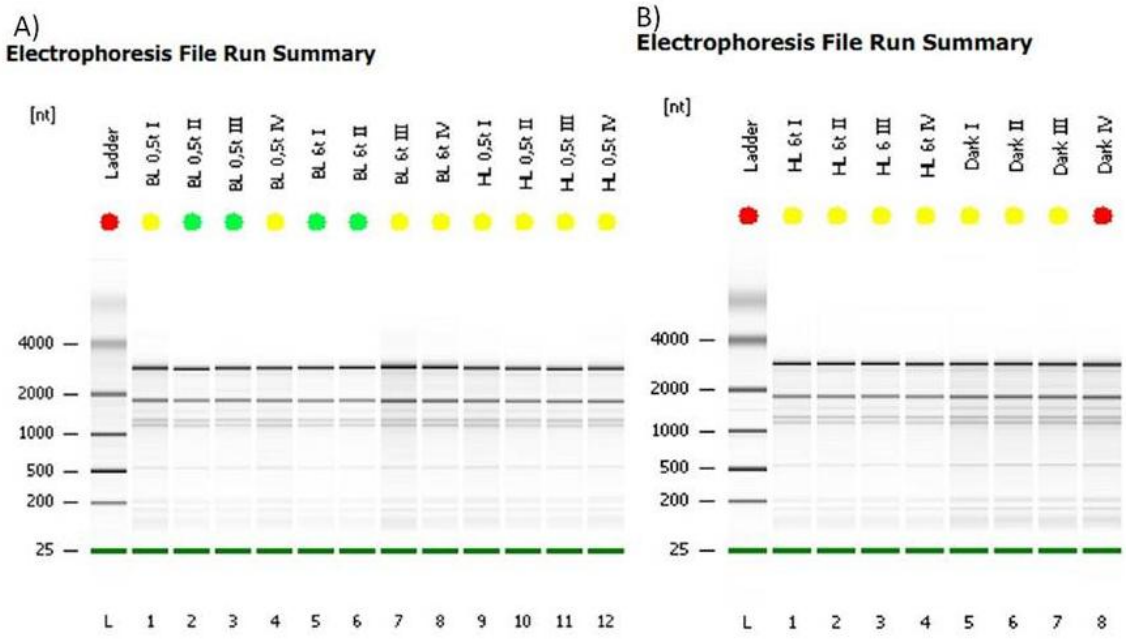

Figure 20: Gel image generated by the Agilent 2100 Bioanalyzer. All the samples had a RNA Integrity Number (RIN) value that was above 6,9. None of the RNA samples showed any clear sign of RNA degradation and all samples were usable for downstream applications. Well marked with L is the nucleotide ladder provided by the Agilent for Agilent 2100 Bioanalyzer, while the other samples are replicas from the light treatments. The color spots under each of the sample names show if the samples had RNA with RIN values within the primary or secondary RIN value thresholds that were set. The primary threshold was set for a RIN value of 8 or higher, and is indicated with a green spot. The secondary threshold was set on RIN value equal to 7 or higher, and is indicated with a yellow spot. Samples with RIN values less than 7 are indicated with a red spot. A) In the first four wells (1-4) the rRNA from replicates of cultures exposed for 0,5 hour blue light (BL0,5h I-IV) are shown, followed by rRNA from BL 6 hour treated cells (BL6h I-IV) in wells 5-8. Wells 9-12 show the white light (WL) treated cells for 0,5 hour (WL0,5 I-IV). B) Wells 1-4, WL6h I-IV replica and in wells 5-8 the control samples (Dark) are located from Dark I-IV.

3.2.2 qPCR analysis of light responses in *S. robusta*

The gene expression of the target genes HSF4, PAS9, D9, LHCR6, LHCF8, CYP15, CYP21, and CYP23 were examined after cDNA synthesis was performed on the total RNA (On all four replicates from each light treatment). Expression was investigated by doing qPCR on the cDNA produced from total RNA. The gene expression of all target genes is illustrated in Figure 21, where the expression is compared against the expression of the dark treated samples. The genes analyzed were normalized using two reference genes, SrVPS35 and SrEXP. Initially three reference genes were chosen, but one of them, SrSec, was influenced by the light treatment and was therefore excluded before doing REST2009 and qBase analyses. The statistical analysis done in qBase was a One way Anova test.

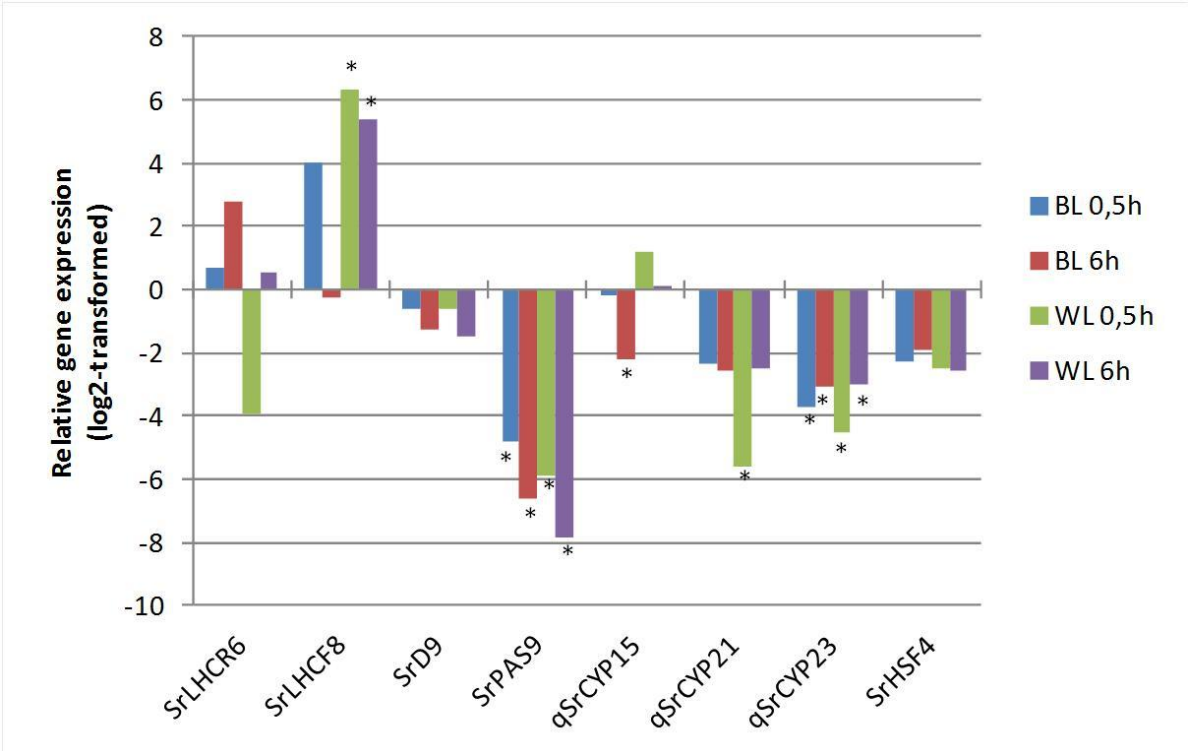

Figure 21: The gene expressions of the genes studied during the qPCR. The difference in gene expression between the treatments was compared by scaling the expression levels from the treated cells against the dark sample expressions. The dark samples were cell cultures harvested after 12 hours in darkness, and functioned as control samples. The relative expression is visualized in a $\log_2$ scale. Cells treated with blue light for 0,5 hour (BL 0,5h) are visualized with blue bars, cells treated with blue light for 6 hours (BL 6h) are indicated with red bars. Cells treated with white light for 0,5 and 6 hours are shown with green and purple bars, respectively. All the samples (BL 0,5h, BL 6h, WL0,5h and WL 6h) were normalized by using SrExp and VPS35 as reference genes. Genes that were significantly regulated are indicated with an asterisk.

From Figure 21 it can be seen that the SrLHCR6 expression is not significantly changed. It appear to be up regulated after 0,5 hours and 6 hour in blue light treatment when compared to the dark samples. After 0.5 hours white light treatment the SrLHCR6 expression was down regulated, while samples exposed for 6 hour were slightly up regulated.

The SrLHCF8 gene expression was up regulated after 0,5 hour of blue light treatment, but unchanged after 6 hours blue light exposure. The expression levels for samples treated with white light were significantly up regulated at both the 0,5 and 6 hour time points.

The overall expression of SrD9 was unaffected by the light treatments, possibly slightly down regulated, but none of the changes were significant.

SrPAS9 gene had significant changes in expression after all treatments, as shown in Figure 21. All expressions were down regulated compared to the control samples. Comparisons between the exposure times shows that samples treated for 6 hours were more down regulated (more than 60 times reduced expression) then samples treated for 0,5 hour.

Expression of the SrCYP15 gene varies with the different treatments. After 0,5 hour blue light exposure the gene expression is not affected (slightly down regulated), but after 6 hours the

expression is significantly down regulated. Cells treated with white light showed no significant regulation of the CYP15 gene, but it may be slightly up regulated after 0,5 hour.

Expression of the SrCYP15 gene varies with the different treatments. After 0,5 hour blue light exposure the gene expression is not affected (slightly down regulated), but after 6 hours the expression is significantly down regulated. Cells treated with white light had a slight up regulation of CYP15 gene expression after 0,5 hour, but they were unaffected after 6 hours treatment (had a small up regulation).

None of the SrCYP21 gene were significantly down regulated by any of the treatments, except for the samples treated with white light for 0,5 hour.

Expression of the SrCYP23 gene shows a significant difference from the dark treated control group. The expression is down regulated in all treatments. Samples treated for 0,5 hour are more affected than the samples treated for 6 hours.

Expression of the HSF4 gene was down regulated for all the samples compared to the control, but none were significant.

Since the some of the genes chosen to be studied in the qPCR experiment where chosen based on gene expressions in *P. tricornutum* a table was made comparing the expression of the genes between the diatoms.

Table 8: A comparison of gene expression in *P. tricornutum and S. robusta*.(DT48D=48 hour dark treatment).

Gene	Treatment							
	Log2-transformed ratio							
	BL0,5h /DT48 D	BL 0,5h/ Dark	BL6h/ DT48D	BL6h/ Dark	WL0,5h/ DT48D	WL 0,5h/ Dark	WL6h/DT48 D	WL 6h/ Dark
	P.tricornutum	*S.robusta*	*P.tricornutum*	*S.robusta*	*P.tricornutum*	*S.robusta*	*P.tricornutum*	*S.robusta*
LHCR6	0,57	0,70	5,78	2,76	-0,27	-3,94	0,06	0,51
LHCF8	6,74	4,02	1,70	-0,24	6,95	6,33	-2,25	5,42
Desaturase delta 9 desaturase (PTD9)	-2,63	-0,61	2,07	-1,29	-2,61	-0,62	-0,13	-1,45
bZIP5_PAS (PAS9)	-7,36	-4,80	-2,59	-6,64	-4,19	-5,91	1,07	-7,85
Heat shock transcription factor (HSF4)	-3,40	-2,26	-3,21	-1,92	-3,21	-2,53	1,13	-2,59

The genes chosen for the qPCR experiment were chosen from a previous microarray experiment results done on *P. tricornutum.* The similarities and differences in gene expression of the genes can be seen in Table 8. Most of the genes show the same regulation tendencies in

both organisms, but there are some genes that act differently. The genes that are expressed in different expression directions are LHCF8 treated in WL 6h. In *P. tricornutum* the expression is down-regulated while in *S. robusta* the expression is up-regulated. The PAS9 and HSF4 genes also show different expressions in the two diatoms. In *P. tricornutum* the expression of the genes are slightly up-regulated, but in *S.robusta* PAS9 and HSF4 are both down-regulated.

3.3 Microarray analyses of light responses in *S. robusta*

RNA samples from *S. robusta* were used to do a microarray experiment, where the whole transcriptome of light treated cells was investigated. Two biological replicates from each treatment (BL 0,5h, BL 6h, WL 0,5 and WL 6h) and four control samples (Dark) were hybridized on three slides (4*44K spots), in total 12 hybridizations.

3.3.1 Significantly up and down regulated genes

Many genes showed signs of significantly up and down-regulation by the light treatment the cells were exposed for. 6 genes that were registered to be either significantly up-regulated or down-regulated in some of the treatments were chosen and compared between the different treatments. First the samples treated for 0,5 hour with different light treatments were compared against each other, results are shown in Table 9. Expression levels are given with a comparison against the dark samples (control). A comparison between BL and WL is also included in Table 9 to check for significant changes in gene expression between the two treatments.

Table 9: Six genes that were significantly up- or down regulated from the microarray experiment were chosen for further comparison between the different treatments. BL 0,5 and WL 0,5 expressions are compared against the control (Dark). Expression of the various genes were also compared with BL against WL.

cDNA ContigID	Gene	BL 0,5h/Dark		WL 0,5/Dark		BL 0,5h/WL 0,5h	
		log2	adj.P.Val	log2	adj.P.Val	log2	adj.P.Val
g31394	Pyruvate Kinase 2 (PK2)	8,159	4,01E-06	4,487	5,66E-05	3,672	0,00091
05727	fucoxanthin chlorophyll a/c-binding protein precursor	5,845	1,39E-05	6,942	6,45E-06	Not significant	Not significant
g34780	DNA-binding heat shock factor	-2,155	0,00238	-5,355	1,51E-05	3,201	0,00101
g18889	Malate synthase (MS)	-4,972	0,00015	-6,712	2,30E-05	Not significant	Not significant
g32753	bZIP7_PAS, bZIP transcription factor family protein	5,701	1,54E-05	Not significant	Not significant	6,882	0,00011
g40889	Fucoxanthin chlorophyll protein 3 (FCP3) (LHCF5)	4,997	9,59E-06	6,505	4,37E-06	Not significant	Not significant

A comparison with the same genes was also done on the BL 6 hours and WL 6 hours treated cells, seen in Table 10. Gene expression in both treatments where compared with the control (Dark). The BL and WL treatments were also compared against each other (BL 6h/WL 6h) to see if there were any significant expression changes between the treatments.

Table 10: Seven genes that were significantly up- or down regulated from the microarray experiment were chosen for further comparison between the different treatments. BL 0,5 and WL 0,5 expressions are compared against the control (Dark). Expression of the various genes was also compared with BL against WL.

cDNA ContigID	Gene	BL 6h/Dark		WL 6h/Dark		BL 6h/WL 6h	
		log2	adj.P.Val	log2	adj.P.Val	log2	adj.P.Val
31394	Pyruvate Kinase 2 (PK2)	5,405	1,42E-05	2,474	0,00103	2,931	0,00103
05727	fucoxanthin chlorophyll a/c-binding protein precursor	6,594	2,76E-06	8,031	1,81E-06	Not significant	Not significant
34780	DNA-binding heat shock factor	-1,768	0,00506	-2,936	0,00019	Not significant	Not significant
18889	Malate synthase (MS)	Not significant	Not significant	Not significant	Not siginificant	Not significant	Not significant
32753	bZIP7_PAS, bZIP transcription factor family protein	7,118	1,95E-06	5,642	6,45E-06	Not significant	Not significant
40889	Fucoxanthin chlorophyll protein 3 (FCP3) (LHCF5)	3,319	4,58E-05	6,794	1,48E-06	-3,474	9,14E-05

3.3.2 Comparison of microarray data against qPCR data

A comparison was done between the results acquired from the qPCR- and the microarray experiment, to control how many of the studied genes gave the same significant changes. The comparison was also done to evaluate similarities between the qPCR and microarray data. The comparison is shown in Table 11.

Table 11: Comparison of significantly up or down-regulated genes investigated with qPCR and microarray.

Treatment comparison	Gene	qPCR ratio	p-value	significant with qPCR	Microarray ratio	adjusted p-value
BL 0,5h/Dark	PAS9	-4,80	6,03E-02	Yes	-3.077	0.0222
	SrCYP23	-3,76	3,67E-03	Yes	not significant	not significant
	LHCF8	4,02	2,61E-01	No	4,017	0,00013
BL 6h/Dark	PAS9	-6,64	4,07E-03	Yes	-2.84	0.0244
	SrCYP15	-2,23	4,97E-02	Yes	-3,394	1,30E-05
	SrCYP23	-3,05	4,08E-02	Yes	not significant	not significant
WL 0,5h/Dark	PAS9	-5,91	2,70E-03	Yes	-2.601	0.0289
	SrCYP15	1,20	2,50E-01	No	1,597	0,00099
	SrCYP23	-4,49	2,70E-03	Yes	not significant	not significant
	SrCYP21	-5,61	-5,61E+00	Yes	not significant	not significant
	LHCF8	6,33	1,28E-03	Yes	5,459	1,99E-05
WL 6h/Dark	PAS9	-7,85	4,58E-03	Yes	-5,049	0,00088
	SrCYP23	-2,99	8,14E-03	Yes	not significant	not significant
	LHCF8	5,42	2,59E-03	Yes	4,475	2,68E-05
	LHCR6	0,51	6,70E-01	No	2,318	0,00262

In Table 11 it is shown that 4 genes showed significant changes of expression in both of the gene expression tests. The rest of the results that showed significant changes in genes expression were either only in the qPCR data or from the data acquired from the microarray experiment. The reference genes used in the qPCR experiment did not show any significant expression values in the microarray data (data not included).

4 Discussion

To be able to use *Seminavis robusta* as a model there are several requirements that needs to be fulfilled, as mentioned by Chepurnov (Chepurnov et al., 2008). One of the criteria was that the organisms needed to be susceptible for genetic manipulation, like transformation (Chepurnov et al., 2008). In this thesis a vector construct was made for *S. robusta*, and it's performance and usefulness in transformation was tested. Some studies have stated that bidirectional promoters are only found in mammalian eukaryotes (Koyanagi et al., 2005), but this however contradicts to recent studies that suggested bidirectional promoters also exist in other eukaryotes (Dhadi et al., 2009; Singh et al., 2009). The nuclear promoter chosen for the vector construct was small and bidirectional, so that two genes could be transcribed simultaneously and from the same promoter. An expression study of the two genes, AtpE and AtpB, which are normally regulated by the promoter, was previously performed by Tore Brembu (unpublished results) and indicated a moderate to high expression of the genes. The gene products are predicted to be localized in the mitochondrion.
In the qPCR gene expression study the blue and white light responses of *S. robusta* was examined and compared to those observed in *Phaeodactylum tricornutum*. The genes included in the expression study were evaluated as potential candidate genes for further studies and expression in the newly produced vector system. The microarray experiment performed gave also further information of gene responses connected to light exposure in *S robusta*.

4.1 Construction of an expression vector for *Seminavis robusta*

The atpBE promoter was chosen because of its relative high expression of atpB and atpE genes (Tore Brembu unpublished results), and the atpB and atpE promoters have been used to express exogenous DNA in previous studies in other organisms (Schauder et al., 1987; Suarez et al., 1997; Bateman and Purton, 2000; Xie and Allison, 2002). Studies of the chloroplastic AtpB and AtpE genes in tobacco plants (Kapoor et al., 1994) and Odontella sinensis (Kroth-Pancic et al., 1995) shows that they are co-transcribed. This may also be the case for the AtpB and AtpE genes coding for mitochondrial ATP synthase subunits beta and epsilon. Analysis of the S. robusta AtpE subunit suggest it belongs to the epsilon/delta class (AtpE/D), the atpB, atpE genes and the promoter atpBE promoter are found in the nuclear genome, and the proteins have to been imported into the mitochondrion. The presence of the mitochondrial genes in the nuclear genome is most likely a result of endosymbiotic gene transfer, something that has been observed in many eukaryotes (Keeling and Palmer, 2008; Jiroutová et al., 2010). The promoters' presence in the nuclear genome was the reason why it was chosen for the expression vector, since the vector was designed for nuclear transformation. The promoter was noticed by Tore Brembu (Dr. Scient) and Per Winge (associate professor) after studying sequence data from *Thalassiosira pseudonana*, *P. tricornutum* and *S. robusta*. The sequence comparison between the diatom species showed that the bidirectional atpBE promoter exist in the genome in all diatoms where genome data is available. The atpB and atpE terminators were also compared, and related sequences were found in the terminators between the

different species. The terminators are needed because they mark the end of a gene during transcripton. The advantage with a bidirectional promoter is that the vector construct can be very compact and express the selective marker gene simultaneously with the gene one wants to express or silence. Often the transformation of algae requires two vectors, one coding the selective marker and the other coding the cDNA of interest (Walker et al., 2005; Hallmann, 2007). When transforming cells with a vector that expresses two genes simultaneously one does not have the same difficulties with having to screen and check if both transgenes are present in the transformed algae (Amendola et al., 2005; Chaturvedi et al., 2006).

4.1.1 Cloning of the AtpBE promoter and associated terminators

Since there were several polymorphisms and an insertion and deletion (indel) in the sequenced promoter sequence as shown in Figure 9, we decided to use the promoter allele without the "deletion" in the vector construct. There is no published work that has documented polymorphisms in the AtpBE promoter for the atpBE genes coding for mitochondrial ATP synthase beta and epsilon/delta subunits. The promoter sequence was modified with the insertion of the restriction sites EcoRI and SmaI as illustrated in Figure 16. When cutting the vector and the promoter sequence with restriction enzymes this had to be done in two steps. First cutting with EcoRI, cleaning the DNA with Promegas' Wizard SV Gel and PCR Clean-Up kit and then proceed with the next cutting (SmaI). The cutting was performed in this order because the restriction enzymes did not have a buffer where both enzymes worked optimally. The EcoRI enzyme worked only optimally in the NEB buffer EcoRI from NEB, while SmaI had only full functionality in NEB buffer 4. The enzymes also had different optimum temperatures 37^{o}C for EcoRI and 25^{o}C for SmaI. This resulted in a low DNA yield of both promoter and linearized vector when the ligation procedure was initiated. Because of this issue a successful insertion of the promoter sequence into the pBKS vector was more difficult and time consuming then first anticipated.

Following the insertion of the promoter, the atpE terminator was inserted into the vector. The cloning of the atpE terminator was done with the restriction enzymes SacI and XbaI. Both of the enzymes had an activity of 100% with NEBuffer 4, so double digestion of the atpE terminator sequence and the vector was done as recommended by New England bioLabs. Since double enzyme digestion was possible with the atpE terminator, the cloning of the fragment into the pBKS vector was less cumbersome. The vector with the atpEt sequence was transformed into competent DH5α cells and verified by sequencing. The atpE terminator had only one SNP and one of the alleles were randomly selected (the thymine containing SNP allele), seen in Figure 10. The sequence was inserted into the vector containing the atpBE promoter. When all the expression initiation and termination components of the atpE expression side of the vector were verified the selective marker was chosen.

4.1.2 Production of a nourseothricin acetyltransferase (*Nat1*) selection marker for S. robusta

Initially Kanamycin (kan) was chosen for selecting transformed cells from untransformed cells. A Kan resistance gene, *neomycin phosphotransferase* (*nptII*) (Klein et al., 1988; Falciatore et al., 1999) was supposed to be placed between the atpE promoter and the atpE terminator. The decision to use *nptII* as selectable marker was revised after testing the antibiotic on wild type S.robusta of mating types 111-1 and 112-1. The S. robusta cells were able to grow on all concentrations of kan tested and showed signs of survival and growth in all concentrations from 50, 100, 150, 300, 200, 300 and 500µg/mL (not included in the results). There is no published work stating that S. robusta is kan resistant or tolerant but, similar tolerance has been observed in P. tricornutum (Apt et al., 1996). As a substitute for the *ntpII* gene, the *nourseothricin acetyltransferase* (*nat1*) gene was chosen. *Nat1* gives resistance against the aminoglycoside antibiotic nourseothricin (McDade and Cox, 2001). The *nat1* gene was chosen based on previous antibiotic exposure experiments performed by Dr. Tore Brembu, NTNU, and because the *nat1* gene had been successfully expressed in a diatom *P. tricornutum* (Zaslavskaia et al., 2000). To achieve the most favorable expression of the *nat1* gene, the nucleotide sequence of the gene was modified to match the codon usage of *S. robusta*, as mentioned in the thesis (Hallmann, 2007; Heitzer et al., 2007). Codon usage data for *S. robusta* was acquired from unpublished sequencing data (Brembu, Winge and Bones). The modified *nat1* gene (Appendix IV) was synthesized by Eurofins and delivered in a pEX-A2 vector. The restriction sites for *SmaI* and *XbaI* were included on either sides of the codon optimized version of the *nat1* gene as illustrated in Figure 11. After the verification and successful insertion of the *nat1* gene into the atpE-AtpEt vector, shown in Figure 8D, *S. robusta* was transformed to confirm that the atpE promoter and terminator were functional and capable of expressing the *nat1* gene.

Transformed nourseothricin resistant *S. robusta* cells were visible after 14 days incubation and were able to grow on agar plates containing 100µg/mL nourseothricin. To further test the cells antibiotic resistance the cells were subjected to higher concentrations of NTC, as mentioned in the results. This was done as a secondary selection, similar to what is mentioned in Apt et al (Apt et al., 1996). The colony PCR performed on the cells, Figure 12 also confirmed that the exogenous gene was present in the cells. Because cells grew fine in the presence of high concentrations of NTC it was concluded that *S. robusta* could be transformed with *nat1* as selectable marker with atpE and atpEt as promoter and terminator respectively. It was not possible to give a good estimate of the transformation efficiency of *S. robusta* since the cells are motile and do not produce distinct colonies. This made it very difficult to separate colonies and migrating cells, because migrating cells could form secondary cell clusters. The transformation efficiency was therefore not calculated.

4.1.3 Checking the bidirectionality of the promoter by expressing the marker gene YFP

To check the functionality of the "atpB direction" of the promoter a gene coding *yellow fluorescent protein* (YFP) was inserted between the atpB promoter and atpB terminator. YFP was chosen since it was readily available and have been used as a reporter gene in a transformations of *P. tricornutum*, (Brembu, unpublished results and master thesis by Martin Vejle Andersen (Andersen, 2012)). The use of YFP as a reporter gene has also been documented in other studies done on diatoms (Siaut et al., 2007; Vardi et al., 2008; Jiroutová et al., 2010). The YFP gene was PCR amplified from the pEarleyGate Vector pEG104 and a stop codon was introduced (the pEG104 vector is used to produce N-terminal YFP-fusion proteins so the stop codon was removed), no codon optimization was done. The restriction sites for *EcoRI* and *HindIII* was introduced into the YFP PCR product to facilitate cloning, seen in Figure 16. After YFP and the atpB terminator had been incorporated into the pBKS vector, this was verified through agarose gel electrophoresis (Figure 14).

The transformed cells were placed on selective medium and observed for 2-3 weeks for NTC resistant cells. The number of transformed cells was not as many as initially expected, only one cell cluster was observed in one out of four Petri dishes. Compared to the first transformation where only the *nat1* gene was present in the vector construct, the cell division was slower and cell density was lower in the second transformation. The reasons for poor growth of the cells expressing the pBKS-NAT1-atpBE-YFP vector construct is unclear, but may be related to the integration site in the genome. The length of the vector construct is small (5.2 kb) and should not pose any problems, it has been reported that cells can be biolistic transformed with vectors up to 20-30kb in size (Sanford et al., 1993).

Because there was no previous literature describing the optimal density of *S. robusta* cells for biolistic transformation, an estimate was therefore done based on the cell count that gave the first successful transformed cells. Studies have indicated that algae have different requirements for how dense the cells should be to achieve a high transformation rate, and this seems to be species-specific (Walker et al., 2005; Hallmann, 2007). The cell density is important to achieve a good transformation (Sanford et al., 1993). The cell density could therefore have been better optimized to achieve a higher transformation rate, but because of time limitations the optimization of cell density was not done. Because of the big size of the *S. robusta* cells (Chepurnov et al., 2008) the cell density was much lower compared to what is normally used during biolistic transformation of *P. tricornutum*. It could be that a higher initial number of cells on the agar plates during the biolistic transformation could have resulted in a greater yield of transformed cells.

The presence of transformed cells was a clear indication that the *nat1* gene was expressed in the cells, making it possible for them to grow on medium containing 100 µg/ml NTC. This was again a confirmation that the atpBE promoter was able to drive the expression of the *nat1* gene. To further test the cells resistance to NTC, cells were transferred from the

solid medium to six well plates containing liquid medium with different concentrations of NTC. The cells were tested for NTC concentrations from 100, 150, 200 and 250 µg/ml and all transformed cells showed viable signs.

When inspecting the cells under a confocal microscope no fluorescence was detected in the cells. The absence of YFP could be caused by several factors, among others inefficient translation of the protein, mis-folding of the protein, and regulatory sequences in the first intron of the atpB gene could be missing leading to low or no expression. Some studies have demonstrated that the absence of endogenous introns in cDNA transcription units can prevent or decrease expression of exogenous genes (Buchman and Berg, 1988; Snowden et al., 1996; Hallmann, 2007). As indicated in Figure 22 in appendix V the endogenous atpB gene has a short first exon, 31 bp coding sequence, followed by a 245 bp intron. This intron is evolutionary conserved and is present in both *P. tricornutum* and *T. pseudonana*. The absence of this intron in the YFP gene could be a possible explanation for why YFP was not successfully expressed in *S .robusta.* One other factor that could cause YFP to not be expressed properly can be because the codons used in YFP did not correspond to the tRNA abundance in *S. robusta.* This may result in inefficient translation of the protein. Unlike the *nat1* gene, YFP was not codon optimized. As previously mentioned, if exogenous genes have codon usage similar to the endogenous genes of the species they are placed in, the success of gene expression increases (Walker et al., 2005; Hallmann, 2007). The functionality of the AtpB promoter was tested by performing a RT-PCR on extracted RNA from transformed *S. robusta* cells. The PCR results came up negative with no amplification of YFP, indicating that the gene may not have been transcribed in the cells. The *nat1* gene, which was used as a positive control was successfully detected, indicating that the cells were transformed. However there could be other reasons for the absence of the YPF after the PCR reaction. The degradation of mRNA is important and prevalent in all organisms from all kingdoms (Beelman and Parker, 1995; Houseley and Tollervey, 2009). The *YFP* mRNA may have been detected as foreign nonsense mRNA because of false or poor polyadenylation and been degraded (Doran, 2006). One further possibility is that the YFP gene is silenced through a RNA interference coupled process (Gutiérrez et al., 1999).

4.2 Light exposure experiment

The light exposure treatment was done by exposing cells for white or blue light over a period of 0,5 and 6 hours before the cultures were harvested and flash frozen. To observe the transcriptional regulation in the transition between darkness and light, cell cultures were harvested after 12 hours in darkness, and importantly, without any light exposure before harvesting. Total RNA was isolated from all the samples and the quality and integrity was tested. When the RNA quality and integrity was confirmed cDNA synthesis was performed followed by quantitative PCR. Gene expression results from the PCR were normalized with linRegPCR (Ruijter and Ramakers, 2003) while statistical analyzes were done with qBase (Hellemans et al., 2007).

4.2.1 RNA integrity and quality test

The RNA integrity and quality was tested with FA gel electrophoresis (Qiagen) and Agilent 2100 Bioanalyzer (Agilent Technologies, USA) as shown in the results. The workload and the usability of these two techniques were compared. In both of the techniques the RNA integrity is estimated by studying the ratio of 28S:18S ribosomal RNA. In FA gel electrophoresis the 28S:18S ribosomal RNA is compared visually. While the Agilent 2100 Bioanalyzer uses the curves from an electrophoresis program to calculate RNA Integrity Number (RIN) values. The RIN value gives a quality estimate of the total RNA by using specific algorithms that are tailor made for plant and animal samples (provided by Agilent Technologies). The RIN values are presented in a scale ranging from 1 to 10, where 1 is the most degraded RNA and 10 is the most intact RNA (Fleige and Pfaffl, 2006). The Agilent 2100 Bioanalyzer also presents the data as a virtual gel image seen in Figure 18. The separation technology used by the Agilent 2100 Bioanalyzer gives a higher resolution and therefore more exact information than the regular FA gel electrophoresis. The Agilent 2100 Bioanalyzer is chip based and is therefore less time consuming, compared to Qiagens FA gel, where the gel has to be prepared for every gel electrophoresis. Ethidium bromide (EtBr**)** is required when preparing FA gel electrophoresis to stain the RNA. This is a drawback since EtBr is considered as mutagenic, carcinogenic and teratogenic (Saeidnia and Abdollahi, 2013). FA gel electrophoresis is however less expensive than the Agilent 2100 Bioanalyzer and the use of the different techniques should be determined depending on the downstream applications. If the downstream application is qPCR, either of the techniques is applicable, as illustrated in the results.

4.2.2 Evaluation of reference genes used in qPCR analysis

Initially three reference genes where chosen based on gene expression studies of *P. tricornutum* by Ph.D Marianne Nymark (Nymark, 2013) and *S. robusta* by Elise Strøm Midthun (Strøm Midthun, 2012). The reference genes were chosen based on how stable the gene expression was under various exposures and the abundance of the mRNA. The genes that were used as reference genes were Exp1 and Vsp35. The *Exp1* gene codes for Exportin-1-like protein while the *Vsp35* (vacuolar sorting protein 35) gene codes for a peripheral membrane protein. The function of the VPS35 protein is not fully understood, though it is thought to be associated with the late Golgi transmembrane protein, VPS10 (Seaman et al., 1997). Orthologous genes to *P. tricornutum Exp1* and *Vsp35* were identified in *S. robusta* and primers were made. A third reference gene was also identified, the *Sec7* gene. Sec7 was however later dismissed as a reference gene because qPCR data indicated that the gene was induced by the light treatments. The Sec7 gene products are thought to have a function concerning vesicle transport on the cytoplasmic surface of the Golgi apparatus (Achstetter et al., 1988; Deitz et al., 1996). Why Sec7 was induced by the light treatments is uncertain. It could be that the processing of glycoproteins and the synthesis of complex polysaccharides in

the Golgi apparatus (Zhang and Staehelin, 1992) increased because of the light treatment and hence also increasing the expression of Sec7. This is however just a speculation and has not been documented.

4.2.3 qPCR analysis of light responsive genes

The genes that were investigated with quantitative PCR were *SrLHCR6, SrLHCF8, SrD9, SrPAS9, SrCYP15, SrCYP21, SrCYP23 and HSF4*. All the gene expressions were compared against cell cultures harvested from 12 hours in darkness. The expression is presented in Figure 21, where the expression is given in Log_2 scale against the dark samples. The results showed that there were changes in the gene expression, and that some of the genes had significant changes in their expression. Because there were no published papers describing a similar light exposure experiment on *S. robusta*, responses were thought to be similar to the ones observed in the diatoms *P. tricornutum.* Gene expression data from light exposure experiments done by Nymark et al (not published) was therefore used to compare with data acquired from the present study.

SrLHCR6 gene expression:

The *SrLHCR6* gene is an orthologue of the fucoxanthin chlorophyll a/c light harvesting protein coding gene *LHCR6* in *P. tricornutum*. The expression of *SrLHCR6* was moderate up-regulated after 0,5 and more up-regulated after 6 hours in blue light as shown in Figure 21. Microarray data of *P. tricornutum* exposed to blue light with lower intensity for 30 min and 6 hours showed similar responses (Table 8), even though the cells had been in darkness for 48h before light exposure. When the cells are exposed to white light with the same time periods (0,5 and 6 hours), the response in *S. robusta* resembles the responses observed in *P. tricornutum.* There is a down-regulation of the gene in both cells, but after the 6 hour treatment the expression is slightly up-regulated again almost at the same level as in the control. None of the *SrLHCR6* expression were detected as significant (p-value < 0,05).

SrLHCF8 gene expression:

The *SrLHCF8* gene did not have any significant changes in the BL treatment, but significant changes were however detected in the WL treatments. The *SrLHCF8* gene, which codes for a LHCF8-like protein was up regulated after 0,5 hour in BL, while after 6 hours in BL treatment the expression was unregulated. This was also the case when the expression was compared against the data from *P. tricornutum* which showed that *LHCF8* was higher expressed at 0,5 hour than 6 hours. Compared with the results provided by Nymark, the initial response to BL after 0,5 hour were almost the same between *S. robusta* and *P. tricornutum* as seen in Table 8. There are several genes closely related to *LHCF8* in Seminavis and this could be one explanation for the different expression results. *SrLHCF8* was up regulated after exposure with white light for 0,5 and 6 hours. As shown in Figure 21 the expression is highest in the cells after 0,5 hour, and then there was a small decrease in the expression. The expression is nevertheless present and seems to be more induced in white light then blue light.

This pattern was also observed by Nymark et al (Nymark et al., 2013), who reported that genes encoding LHCF proteins were highly expressed in *P. tricornutum* exposed for white light.

SrD9 gene expression:

Among the other genes that were studied was the *Δ9 desaturases-like* gene (*SrD9*). Unlike the other genes that were examined, *SrD9* showed few signs of being influenced by the light treatments, and did not have any significant changes. As indicated in Figure 21 the expression of *SrD9* had barely a 2-fold down regulation from 0,5 to 6 hours treatment in both light treatments. Previous literature has suggested that the D9 enzymes might be important components in lipogenesis in cells (Nakamura and Nara, 2004). In human cells the *D9* gene is reported to be functioning in storing of excess energy as triglyceride. In the yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) a Δ9 desaturase has also been detected. It is a part of the maintenance system of the physical property of the cell membrane. It is also shown that the mRNA of *OLE1* a D9, increases in cold temperatures and causes an increase of monounsaturated fatty acids in the membrane phosolipids (Nakamura and Nara, 2004). The fact that the expression of SrD9 decreases after light exposure could mean that the lipogenesis also decreases in the cells. This could indicate that the storage of energy as triglycerides decreased when the *S.robusta* cells were exposed in blue and white light. Studies done on *Arabidopsis thaliana* (*A. thaliana)* mutants with deficient chloroplast desaturase suggest that the unsaturation of fatty acids can decide the temperature range in which chloroplasts can function optimally in plants. The mutant *A. thaliana* plants showed however more tolerance to higher temperatures seemed to increase with the reduced activity of chloroplast ω 3 desaturase (Nakamura and Nara, 2004). It could be that the slight decrease in *SrD9* in *S. robusta* was an adaptation to higher temperature, caused by the light treatments. This is only a theory and further studies must be done on this to be sure.

SrPAS9 gene expression:

One of the genes that showed significant changes to all the treatments was the *SrPas9* gene. As observed in Figure 21, *SrPAS9* gene was down regulated in both the white and blue light treatments when compared to the dark samples. In a previous study by Marianne et al (Nymark et al., 2013) it is suggested that a related PAS gene observed in *P. tricornutum* has a PER-ARNT-SIM domain and probably is a aureochrome. Other studies also suggests that the bZIP-PAS genes found in diatoms might be aureochrome-like proteins (Huysman et al., 2013b). Huysman et al mentions that the bZIP-PAS proteins might represent a class of photoreceptors that have putative light-sensitive and DNA-binding domains. The aureochromes have a bZIP domain and a LOV (light, oxygen, or voltage) domain belonging to the PAS superfamily of domains (Ogura et al., 2008) which is also found in blue-light photoreceptors of the phototropin family (Huysman et al., 2013b). The aureochromes are thought to be blue light induced (Herman et al., 2013; Huysman et al., 2013a; Kianianmomeni

and Hallmann, 2014) and they are expected to play a role in cell growth (Huysman et al., 2013b). In *S. robusta* the *SrPAS9* expression is down regulated in blue and white light. Almost similar behavior was also observed in the study done by Marianne et al (Nymark et al., 2013), where the expression of the PAS gene (Phatr2_45142) was highest expressed in darkness and down regulated after exposure to 6 hours in white light. One theory for why the expression is down regulated could be that there is a negative feedback loop that regulates the expression of *SrPAS9* when it initiates downstream signaling for cell growth and division. It has been reported that a related bZIP-PAS protein AUREOCHROME1a (AUREO1a) induce the transcription of a diatom-specific cyclin 2 (dsCYC2) in *P. tricornutum* (Huysman et al., 2013a). The dsCYC2 is thought to have a part in the cell division upon illumination (Lockhart, 2013) and controlling a G1-to-S light-dependent cell cycle checkpoint in *P. tricornutum* (Huysman et al., 2013a). There is however not much data about the *SrPAS9* gene that can shed more light about what specific functions the gene might have and if it might have a role similar to AUREO1a.

The gene expression of three CYP genes in *S. robusta*, *SrCYP15*, *SrCYP21* and *SrCYP23*, were analyzed by qPCR. The same genes have previously been studied and characterized by Elise Midthun (Strøm Midthun, 2012). In the study performed by Strøm Midthun the cells were exposed to different temperatures and day/night cycles with white light, while in the present study the gene expression of the CYP genes were investigated under blue and white light in ambient temperature.

SrCYP15 gene expression:

SrCYP15 is thought to belong to the CYP97 family (Strøm Midthun, 2012), and codes for a carotenoid hydroxylase related protein (Kim et al., 2009). The caroteniods have been shown to be present in most photosynthetic organisms, ranging from cyanobacteria to higher plants (Kim et al., 2009). It is believed that the SrCYP15 has a light-harvesting or light protection function in *S. robusta* (Strøm Midthun, 2012). The gene expression of SrCYP15 is not highly expressed in the control condition (dark sample) or the treated cells, an interesting observation however was that the gene expression was down regulated in blue light and up regulated in white light. In the 6 hour blue light treatment the gene showed a significant down regulation. The results might indicate that SrCYP15 is connected to light harvesting rather than light protection, since the responses are opposite between blue and white light. This could also explain why *SrCYP15* was up regulated in the white light exposure because white light contains the appropriate photons for photosynthesis, hence light harvesting could be needed. After 6 hours exposure to white light, the expression went down to almost the same as the control, which could indicate acclimatization to the white light. Further study has to be done before to the function of SrCYP15 can be predicted. As the results illustrate the expression levels of *SrCYP15* were low and shows that *SrCYP15* gene is responsive to light treatment.

SrCYP21 gene expression:

Studies done by Midthun (2012) suggested that eight CYP enzymes are diatom specific CYPs (Strøm Midthun, 2012). SrCYP21 was one of the diatom specific enzymes and was chosen in this study. In the dark control treatment the expression of *SrCYP21* was low, but with a higher compared to the expression from the light treatments. Only the WL 0,5h treatments showed significant expression changes. The expression was down regulated in all treatments when compared to the dark reference. The expression in the light treatments was so low that it might seem like the gene had been turned off during the light treatments. The down regulation in white light 0,5 hour was significant, however there were only two replicas included in the results from this treatment. Two of the biological replicates did not fulfill the sample quality control criteria, and as a consequence this result might be caused by insufficient number of replicates. One reason for the change in the expression could be that the gene has a connection to cold-responses. It could be that the absence of light lowered the temperature and caused the up regulation of *SrCYP21*. This could be one explanation for the down regulation in the light treatments, since it has previously been stated that *SrCYP21* could be cold-response coupled (Strøm Midthun, 2012).

SrCYP23 gene expression:

A phylogenetic study of the *S. robusta* CYP genes has identified three CYP genes that show low similarity to CYP enzymes in other organisms (Strøm Midthun, 2012). *SrCYP23* was one of these genes and was investigated in this study. There have been found genes similar to SrCYP23 in *P. tricornutum* and *T. pseudonana*, and it is therefore thought that the *SrCYP23* gene might be a part of a diatom-specific-CYP-family (Strøm Midthun, 2012). *SrCYP23* had a low expression in the dark control and in the light treatments the expression was almost completely turned off. The expression changes were however significant in all the treatments, with a down-regulation in all treatments. Because there was a very low expression of the gene in the light treatments, it could be as indicated by Strøm Midthun (2012) that the enzyme has a function related to temperature changes (Strøm Midthun, 2012). To be certain of this more temperature related experiments must be done to check the response of *SrCYP23*.

SrHSF4 gene expression:

The *SrHSF4*, a heat shock transcription factor gene present in *S.robusta* was also studied during the qPCR experiment. The HSFs are thought to register environmental signals concerning high temperature stress and convey the information onwards to the transcriptional machinery (Czarnecka-Verner et al., 1997). The HSFs are thought to be important in many eukaryotic organisms and work as transcription factors for many important heat shock proteins (HSPs) (Prändl et al., 1998). *SrHSF4* had a low expression in the control treatment and in the treated cells the gene expression was further down regulated. The analysis of the gene also showed that none of the treatments gave any expression changes that were

significant. As shown in Figure 21 the there is a small down regulation of the gene expression in all the treatments, and the expression was almost turned off. This is similar to the responses Nymark registered in *P. tricornutum* (unpublished data) were the HSF4 homolog HSF4.3a (Phatr2_49594) is down regulated by both blue and white light. Further research needs to be done before a prediction can be made on what kind of a function the gene could have.

4.2.4 Microarray analysis of light response in *S. robusta*

During a microarray experiment expression of thousands of genes can be monitored and studied at the same time (Stears et al., 2003). This is possible because a parallel quantification is done on the large number of mRNA transcripts isolated from cells (Schulze and Downward, 2001). When performing microarray experiment huge amounts of data are acquired. To give an impression of which types of genes were affected a few genes with significant expression changes were chosen. Seven genes that were either significantly up or down regulated where chosen and compared and they are listed in Table 9 and Table 10.

Pyruvate Kinase 2 (PK2) gene expression:

One of the genes that were up-regulated in all the treatments was PK2 when controlled against the control (Dark). This indicates that light induces the expression of PK2. PK2 is involved in the glycolysis were it removes a phosphate group from phosphoenolpyruvate (PEP), phosphorylates ADP to ATP and produce pyruvate (Strominger, 1955). The expression of PK2 is significantly up-regulated in BL and WL compared to the control (Dark), but interestingly the comparison of BL and WL shows that the expression of PK2 is significantly more up-regulated in BL than WL. This could indicate that the chloroplast produce more carbohydrates-glucose when exposed to blue light and thereby increase the activity of glycolysis pathway. Studies done on chlorophyll-free, carotenoid-containing mutant *Chlorella vulgaris* has shown that PK was highly expressed in the cells exposed to blue light (Ruyters, 1980). Other studies also mention that blue light induces responses in some photosynthetic microalgae that provide substrates for PEP dephosphorylation (Sanchez and Voltolina, 1994), which might also indicate that the expression of PK is high in the cells. The expression levels of PK in *S. robusta* are shown in Table 9 and Table 10. A clear difference between BL and WL can be observed.

fucoxanthin chlorophyll a/c-binding protein (FCP) precursor expression:

Fucoxanthin chlorophyll a/c-binding proteins are LHC proteins that can be found in diatoms (Lang and Kroth, 2001). As indicated in Table 9and Table 10 the expression of the FCP precursor gene is significantly up-regulated in BL and WL both in the 0,5 hour and 6 hours treatments. This was expected since the diatoms are very sensitive to light changes, and the FCPs are thought to be associated with mechanisms regarding light harvesting (Lang and Kroth, 2001) and excess energy dissipation (Depauw et al., 2012). When comparing the expression levels of BL and WL against the control in Table 9 and Table 10 the expression of FCP precursor gene seems to be higher in WL treatments compared to the BL. This could

indicate that WL induces more expression of the FCP precursor gene, which could mean that the gene is linked to photosynthetic activity. Comparisons of the BL and WL treatments (BL/WL) in Table 9 and Table 10 however show that there is no significant difference in the expression between the two treatments. To understand the function of the gene further study need to be done on the gene.

DNA-binding heat shock factor expression:

DNA-binding HSFs are transcription factors that control the transcription of HSPs (Nakai, 1999). The HSFs often control the expression of HSPs by detecting cellular metabolic changes like differentiation and stresses like heat shock, oxidative stress and exposure to heavy metals that cause denaturation of protein. One HSF factor (cDNAcontig34780) was significantly down-regulated in all treatments when compared against the control (Dark), seen in Table 9 and Table 10. The down regulation is higher in cells treated with WL 0,5h that cells treated with BL for the same time period. A down regulation was also registered with HSF4 in the qPCR experiment with the same trends, where the cells treated with WL were more down regulated compared to BL treated. Cells treated for 0,5 hour show an significant expression difference between the BL and WL treatments. Cells treated with BL had a higher expression of the HSF (cDNAcontig34780). Cells treated for 6 hours however did not show any significant expression difference between the two light treatments as indicated in Table 10. The down-regulation of the HSF gene after 6 hours is moderate in BL and WL when compared to the control samples. There were also no significant expression difference between BL 6h and WL 6h. This might indicate that there has been some kind of an adaptation to the light treatment in the cells. Studies done by Ashworth et al also suggest that HSFs may have an important role in diatoms transition from nighttime conditions to light (Ashworth et al., 2013). This might also might be one reason for why the HSF gene expression is higher in the 0,5 hour treatment compared to the 6 hours treatment.

Malate synthase (MS) expression:

MS is one of the enzymes that are involved in the pyruvate pathway for isoleucine biosynthesis (Howell et al., 1999; Xu et al., 2004) and malate synthase is also one of the central enzymes in the glyoxylate cycle, together with isocitrate lyase. Expression of the MS gene is significantly down-regulated in the 0,5 hour light treatments. The down-regulation is strongest in the WL 0,5h treated cells, the expression difference between the two treatments (BL 05h/WL 0,5) is however not significant as shown in Table 9. The gene does not have any significant changes in the 6 hours light treatment, as indicated in Table 10. The expression of the MS gene is therefore only significantly down-regulated in the 0,5 light treatment, something that can indicate that the gene expression is stimulated by light exposure after darkness. In *P. tricornutum* MS has also been observed, and been registered as a fast light-

responding gene (Chauton et al., 2013). The fact that MS is highly expressed in *S. robusta* in darkness (and *P. tricornutum*) might indicate that the cells use the glyoxylate cycle to generate intermediates that can be used to synthesize glucose when no light is present (Chauton et al., 2013).

bZIP7_PAS expression:

bZIP7_PAS is a bZIP transcription factor family protein closely related to the bZIP_PAS7 protein in P. tricornutum. In plants basic region/leucine zipper motif (bZIP) transcription factors are thought to regulate processes like pathogen defence, seed maturation, flower development and light and stress signaling (Jakoby et al., 2002). In diatoms, classes of bZIP factors are thought to function as blue light photoreceptors (Rayko et al., 2010). The expression of the bZIP7_PAS-like gene show significant, strong up-regulation in the cells treated with BL 0,5h, while there was no significant expression in WL 0,5h treated cells. Table 9 also shows that there is a significant expression difference when the BL 05h treated cells are compared to the WL 0,5h treatment. In the 6 hours treatments there is a significant up-regulation of bZIP7_PAS in both BL and WL compared to the control, as shown in Table 10. The up-regulation is higher in the BL, which could be an indication for that bZIP7_PAS is partly blue light regulated. The comparison between the BL 6h and WL 6h treatments however show that there is no significant expression difference between the treatments (BL 6h/WL 6h). Other studies suggests that might regulate several other transcriptional regulators (Ashworth et al., 2013).

Fucoxanthin chlorophyll protein 3 (FCP3 and LHCF5) expression:

The expression of FCP3 is similar to the responses observed from the FCP precursor gene. There is a significant up-regulation of gene expression in the BL 05h and WL 0,5h treatments compared to the control. The expression is higher in WL than BL, but the difference in expression is not significant. When comparing the BL 6h and WL 6h in Table 10 the expression of the genes are still significantly up-regulated. The gene expression has however decreased slightly in the BL treatment, between 0,5h and 6 hours. The gene expression in WL is the same for both time points. Comparisons of the two light treatments (BL 6h/WL 6h) show that the difference in FCP3 gene expression is significant between the treatments after 6 hours. This might indicate that WL has a stronger induction of FCP3 than BL, and that the photosynthetic activity was higher in the cells exposed to WL compared to cells in BL. This is however only a speculation and further analysis must be done on the FCP3 gene and protein to confirm this. One study done on *Laminaria japonica* even suggest that the expression of LHCF5 varies between mating types (Zou et al., 2009).

4.2.5 Comparison of light responsive genes in qPCR and microarray analysis

The genes studied with qPCR were compared against the gene data that was acquired with microarray. qPCR is sometimes used as a validation tool to confirm gene expression results that are acquired through microarray analysis (Morey et al., 2006). Data from microarray and qPCR analysis can however sometimes show different results for same genes (Morey et al., 2006). A comparison between the expressions of the two datasets was therefore done to check for similarities.

The comparison between genes that gave significant expression levels from the qPCR experiment with the microarray experiment indicated that not all the results were the same between the gene expression studies. All the genes that showed significant gene expression in the qPCR analysis are listed in Table 11. The results from the same genes from the microarray experiment is also listed, and compared against the qPCR data. Genes that initially did not show significant expression changes in the qPCR but significant changes in the microarray data were also included in Table 11. From the BL 0,5h treatment none of the significant gene expression changes were detected in the microarray experiment. Expression of the LHCF8 gene was significantly up regulated in the microarray experiment but this was not detected in the qPCR experiment. In the WL 0,5h treatment there was only one gene that had significant gene expression change in both datasets. SrCYP15 showed a down-regulation in both datasets, even though microarray data indicate a slightly more down-regulation than the qPCR data, as shown in Table 11. In the WL 0,5h treatment LHCF8 was significantly up-regulated in both datasets. None of the other genes were significantly up- or down-regulated in both datasets. Two genes showed similar significant gene expressions in both datasets from the WL 6h treatment. The PAS9 gene was down-regulated in both datasets as well the expression of LHCF8, which was up-regulated. In Table 11 there are also other genes that show significant gene expression values, but they are only present in one of the two datasets. There could be various reasons for why the gene expressions did not correlate between the datasets. One major difference is that for the qPCR experiment there were 4 biological samples (replicas) for each treatment, while for the microarray analysis there were just two samples for the blue and white light treatments. This means that many genes with moderate regulation never will be identified as significantly expressed (adjusted p-value < 0.01) (Jørstad et al., 2007). With 3 or more biological replicates the list of significantly expressed genes will increase. Another issue is sensitivity. qPCR is more sensitive than DNA microarray and genes with low expression are not easily detected with microarrays and are often identified as unregulated. It is also important to point out that the two methods also use different normalization methods. While qPCR uses only a few genes for normalization (reference genes), in DNA microarrays there are thousands of genes that contribute to the normalization and thereby avoid bias introduced by the reference genes.

5 Conclusion

The aim of this study was to create transgenic *S. robusta* strains by transforming cells with a specifically designed expression vector containing a bidirectional promoter capable of expressing two genes simultaneously. A RT-qPCR and DNA microarray transcriptional analyses were also done on the diatoms after exposure to blue and white light after 12 hours in darkness. A further goal with this study was to promote the establishment of *S. robusta* as the first benthic diatom model system and a candidate for molecular genetic studies.

- The vector construct made for *S. robusta* with the endogenous AtpBE promoter showed only partial functionality when transformed into the diatom. Only the selectable marker (*nat1*) was transcribed in the cells from the AtpE direction of the promoter. The YFP gene that should have been expressed by the AtpB part of the promoter did not show any signs of expression, which might indicate that some essential parts of the promoter is missing and thus preventing gene expression.

- From the qPCR experiment 5 genes from the study had significant expression changes to the light treatments the cells were exposed for. Many of the genes that had significant expression changes have been documented to be connected to light responses, either light harvesting or photoprotection mechanisms in *S. robusta* and *P. tricornutum.*

- The microarray experiment performed on the light treatment samples revealed that several genes were up- and down-regulated. Six of the most regulated genes were included and discussed in this study. Some of the genes that were significantly differentially expressed in the microarray experiment showed similar regulation in the qPCR experiment. Genes connected to the pyruvate and glycolysis pathway were particularly affected and this is in good agreement with previous experiments performed on *P. tricornutum* (Chauton 2012). The data acquired from this work can be valuable in future studies if mapping genes connected to light responses and circadian rhythms is done.

- Comparison of the genes studied with RT-qPCR and microarray revealed that not all the genes had the same significant changes in gene expression in both datasets. Out of the 12 significantly regulated gene expressions in the RT-qPCR data only four of the genes had significant changes in both datasets. This may be due to sample size differences between the two experiments. Since the two different technologies use different approaches to process their data, the datasets need to be orderly filtered to yield comparable results.

Recommendations for further work

The development of an expression system for *S. robusta* that is able to express two proteins from a bidirectional promoter was not fully completed and further work remains.

Since the AtpE direction of the promoter is functional, an insertion of a new promoter into the already exiting vector could be an opportunity to make a vector that expresses two genes. A strong promoter can be chosen for the vector based on the microarray data acquired from this experiment and previous sequencing data from *S. robusta.* Another option is to include the missing intron from AtpB in the construct as it may contain regulatory sequence elements. However this will add a N-terminal protein tag to the expressed protein and can have unwanted consequences. An alternative will be to move the *nat1* gene to the "AtpB side" and see if it can tolerate the N-terminal protein tag.

An over expression of some of the genes discussed in this study can be done to further investigate their functions. HSF4 could be one of the candidates for this kind of study. Developing this vector system to include the expression of RNAs that fold into dsRNA structures may be used for RNA silencing. It could also be used to develop a vector for genome editing by exploiting the possibilities of the CRISPR/Cas system.

Since recent studies has shown that it is possible to transform *P. tricornutum* thought multi-pulse electroporation, an attempt to do this with *S. robusta* could be interesting. Since electroporation is less time consuming and produce a great number of transformants compared to biolistic transformation, this is a possibility that should be further investigated. If successful it might make *S. robusta* more usable in molecular genetic studies.

The blue light study done on *S. robusta* and the use of microarray technology has acquired a lot of data about the organisms' responses when they are transferred from darkness to blue and white light. The data from this experiment should be further investigated and compared to other diatoms and algae. Little is known about the mechanisms concerning carbon metabolism, storage, regulation and their connection to light exposure in *S. robusta* and further studies of the current DNA microarray data set will provide new insights into "the secret life of benthic diatoms".

6 References

ACHSTETTER, T., FRANZUSOFF, A., FIELD, C. & SCHEKMAN, R. 1988. SEC7 encodes an unusual, high molecular weight protein required for membrane traffic from the yeast Golgi apparatus. *J Biol Chem,* 263**,** 11711-11717.

AGILENT TECHNOLOGIES. 2010. *One-Color Microarray-Based Gene Expression Analysis* [Online]. Agilent Technologies. Available: http://www.genome.duke.edu/cores/microarray/services/agilent-microarrays/documents/LIQA_G4140-90040_GeneExpression_One-color_v6.5.pdf [Accessed 18.05.14 2014].

ALTING-MEES, M. A. & SHORT, J. M. 1989. pBluescript II: gene mapping vectors. *Nucleic Acids Res,* 17.

AMENDOLA, M., VENNERI, M. A., BIFFI, A., VIGNA, E. & NALDINI, L. 2005. Coordinate dual-gene transgenesis by lentiviral vectors carrying synthetic bidirectional promoters. *Nat Biotech,* 23**,** 108-116.

ANDERSEN, M. V. 2012. *Purification and characterization of an unusual DNA glycosylase in diatoms.*

ANDERSEN, R. A. 2005. *Algal culturing techniques,* Amsterdam, Elsevier.

APT, K. E., GROSSMAN, A. & KROTH-PANCIC, P. 1996. Stable nuclear transformation of the diatom Phaeodactylum tricornutum. *Molecular and General Genetics MGG,* 252**,** 572-579.

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

ARMBRUST, E. V., BERGES, J. A., BOWLER, C., GREEN, B. R., MARTINEZ, D., PUTNAM, N. H., ZHOU, S., ALLEN, A. E., APT, K. E., BECHNER, M., BRZEZINSKI, M. A., CHAAL, B. K., CHIOVITTI, A., DAVIS, A. K., DEMAREST, M. S., DETTER, J. C., GLAVINA, T., GOODSTEIN, D., HADI, M. Z., HELLSTEN, U., HILDEBRAND, M., JENKINS, B. D., JURKA, J., KAPITONOV, V. V., KRÖGER, N., LAU, W. W. Y., LANE, T. W., LARIMER, F. W., LIPPMEIER, J. C., LUCAS, S., MEDINA, M., MONTSANT, A., OBORNIK, M., PARKER, M. S., PALENIK, B., PAZOUR, G. J., RICHARDSON, P. M., RYNEARSON, T. A., SAITO, M. A., SCHWARTZ, D. C., THAMATRAKOLN, K., VALENTIN, K., VARDI, A., WILKERSON, F. P. & ROKHSAR, D. S. 2004. The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism. *Science,* 306**,** 79-86.

ASHWORTH, J., COESEL, S., LEE, A., ARMBRUST, E. V., ORELLANA, M. V. & BALIGA, N. S. 2013. Genome-wide diel growth state transitions in the diatom Thalassiosira pseudonana. *Proceedings of the National Academy of Sciences,* 110**,** 7518-7523.

BATEMAN, J. & PURTON, S. 2000. Tools for chloroplast transformation in Chlamydomonas: expression vectors and a new dominant selectable marker. *Molecular and General Genetics MGG,* 263**,** 404-410.

BEELMAN, C. A. & PARKER, R. 1995. Degradation of mRNA in eukaryotes. *Cell,* 81**,** 179-183.

BERNHARDT, R. 2006. Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology,* 124**,** 128-145.

BIORAD. 1995. *Instruction Manual, Biolistic PDS-1000/He Particle Delivery System, Rev C* [Online]. Available: http://www.bio-rad.com/webroot/web/pdf/lsr/literature/M1652249LEASE.pdf [Accessed 23.02.14].

BJORNSTI, M.-A. & MEGONIGAL, M. D. 1999. Resolution of DNA molecules by one-dimensional agarose-gel electrophoresis. *DNA topoisomerase protocols.* Springer, 9-17.

BOWLER, C., ALLEN, A. E., BADGER, J. H., GRIMWOOD, J., JABBARI, K., KUO, A., MAHESWARI, U., MARTENS, C., MAUMUS, F., OTILLAR, R. P., RAYKO, E., SALAMOV, A., VANDEPOELE, K., BESZTERI, B., GRUBER, A., HEIJDE, M., KATINKA, M., MOCK, T., VALENTIN, K., VERRET, F., BERGES, J. A., BROWNLEE, C., CADORET, J.-P., CHIOVITTI, A., CHOI, C. J., COESEL, S., DE MARTINO, A., DETTER, J. C., DURKIN, C., FALCIATORE, A., FOURNET, J., HARUTA, M., HUYSMAN, M. J. J., JENKINS, B. D., JIROUTOVA, K., JORGENSEN, R. E., JOUBERT, Y., KAPLAN, A., KROGER, N., KROTH, P. G., LA ROCHE, J., LINDQUIST, E., LOMMER, M., MARTIN-JEZEQUEL, V., LOPEZ, P. J., LUCAS, S., MANGOGNA, M., MCGINNIS, K., MEDLIN, L. K., MONTSANT, A., SECQ, M.-P. O.-L., NAPOLI, C., OBORNIK, M., PARKER, M. S., PETIT, J.-L., PORCEL, B. M., POULSEN, N., ROBISON, M., RYCHLEWSKI, L., RYNEARSON, T. A., SCHMUTZ, J., SHAPIRO, H., SIAUT, M., STANLEY, M., SUSSMAN, M. R., TAYLOR, A. R., VARDI, A., VON DASSOW, P., VYVERMAN, W., WILLIS, A., WYRWICZ, L. S., ROKHSAR, D. S., WEISSENBACH, J., ARMBRUST, E. V., GREEN, B. R., VAN DE PEER, Y. & GRIGORIEV, I. V. 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature,* 456**,** 239-244.

BREMBU, T., WINGE, P., TOOMING-KLUNDERUD, A., NEDERBRAGT, A. J., JAKOBSEN, K. S. & BONES, A. M. 2013. The chloroplast genome of the diatom Seminavis robusta: New features introduced through multiple mechanisms of horizontal gene transfer. *Mar Genomics,* 21**,** 00080-00089.

BUCHMAN, A. R. & BERG, P. 1988. Comparison of intron-dependent and intron-independent gene expression. *Mol Cell Biol,* 8**,** 4395-4405.

BUSTIN, S. A., BENES, V., GARSON, J. A., HELLEMANS, J., HUGGETT, J., KUBISTA, M., MUELLER, R., NOLAN, T., PFAFFL, M. W. & SHIPLEY, G. L. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry,* 55**,** 611-622.

BUSTIN, S. A. & NOLAN, T. 2004. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J Biomol Tech,* 15**,** 155-166.

CHATURVEDI, C. P., SAWANT, S. V., KIRAN, K., MEHROTRA, R., LODHI, N., ANSARI, S. A. & TULI, R. 2006. Analysis of polarity in the expression from a multifactorial bidirectional promoter designed for high-level expression of transgenes in plants. *Journal of Biotechnology,* 123**,** 1-12.

CHAUTON, M. S., WINGE, P., BREMBU, T., VADSTEIN, O. & BONES, A. M. 2013. Gene regulation of carbon fixation, storage, and utilization in the diatom Phaeodactylum tricornutum acclimated to light/dark cycles. *Plant physiology,* 161**,** 1034-1048.

CHENEY, D., METZ, B. & STILLER, J. 2001. Agrobacterium-mediated genetic transformation in the macroscopic marine red alga Porphyra yezoensis. *J Phycol,* 37.

CHEPURNOV, V. A., MANN, D. G., SABBE, K. & VYVERMAN, W. 2004. Experimental studies on sexual reproduction in diatoms. *International review of cytology,* 237**,** 91-154.

CHEPURNOV, V. A., MANN, D. G., VON DASSOW, P., VANORMELINGEN, P., GILLARD, J., INZE, D., SABBE, K. & VYVERMAN, W. 2008. In search of new tractable diatoms for experimental biology. *BioEssays,* 30**,** 692-702.

CZARNECKA-VERNER, E., YUAN, C.-X., NOVER, L., SCHARF, K.-D., ENGLICH, G. & GURLEY, W. 1997. Plant heat shock transcription factors: positive and negative aspects of regulation. *Acta Physiologiae Plantarum,* 19**,** 529-537.

DEITZ, S. B., WU, C., SILVE, S., HOWELL, K. E., MELANÇON, P., KAHN, R. A. & FRANZUSOFF, A. 1996. Human ARF4 expression rescues sec7 mutant yeast cells. *Molecular and cellular biology,* 16**,** 3275-3284.

DEPAUW, F. A., ROGATO, A., D'ALCALÁ, M. R. & FALCIATORE, A. 2012. Exploring the molecular basis of responses to light in marine diatoms. *Journal of Experimental Botany,* 63**,** 1575-1591.

DESJARDINS, P. & CONKLIN, D. 2010. NanoDrop microvolume quantitation of nucleic acids. *J Vis Exp,* 22.

DHADI, S. R., KROM, N. & RAMAKRISHNA, W. 2009. Genome-wide comparative analysis of putative bidirectional promoters from rice, Arabidopsis and Populus. *Gene,* 429**,** 65-73.

DORAN, P. M. 2006. Foreign protein degradation and instability in plants and plant tissue cultures. *Trends in biotechnology,* 24**,** 426-432.

DU BUF, H. & BAYER, M. M. 2002. *Automatic diatom identification,* River Edge, N.J., World Scientific.

DUNAHAY, T. G., JARVIS, E. E. & ROESSLER, P. G. 1995. Genetic transformation of the diatoms Cyclotella cryptica and Navicula saprophila. *Journal of Phycology,* 31**,** 1004-1012.

EARLEY, K. W., HAAG, J. R., PONTES, O., OPPER, K., JUEHNE, T., SONG, K. & PIKAARD, C. S. 2006. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J,* 45**,** 616-629.

FALCIATORE, A. & BOWLER, C. 2002. Revealing the molecular secrets of marine diatoms. *Annual review of plant biology,* 53**,** 109-130.

FALCIATORE, A., CASOTTI, R., LEBLANC, C., ABRESCIA, C. & BOWLER, C. 1999. Transformation of Nonselectable Reporter Genes in Marine Diatoms. *Marine Biotechnology,* 1**,** 239-251.

FALKOWSKI, P. G., KATZ, M. E., KNOLL, A. H., QUIGG, A., RAVEN, J. A., SCHOFIELD, O. & TAYLOR, F. 2004. The evolution of modern eukaryotic phytoplankton. *Science,* 305**,** 354-360.

FLEIGE, S. & PFAFFL, M. W. 2006. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine,* 27**,** 126-139.

FRANCIUS, G., TESSON, B., DAGUE, E., MARTIN-JÉZÉQUEL, V. & DUFRÊNE, Y. F. 2008. Nanostructure and nanomechanics of live Phaeodactylum tricornutum morphotypes. *Environmental microbiology,* 10**,** 1344-1356.

GIBBS, R. A. 1990. DNA amplification by the polymerase chain reaction. *Analytical Chemistry,* 62**,** 1202-1214.

GILLARD, J., DEVOS, V., HUYSMAN, M. J., DE VEYLDER, L., D'HONDT, S., MARTENS, C., VANORMELINGEN, P., VANNERUM, K., SABBE, K. & CHEPURNOV, V. A. 2008. Physiological and transcriptomic evidence for a close coupling between chloroplast ontogeny and cell cycle progression in the pennate diatom Seminavis robusta. *Plant physiology,* 148**,** 1394-1411.

GIULIETTI, A., OVERBERGH, L., VALCKX, D., DECALLONNE, B., BOUILLON, R. & MATHIEU, C. 2001. An overview of real-time quantitative PCR: applications to quantify cytokine gene expression. *Methods,* 25**,** 386-401.

GUÉNIN, S., MAURIAT, M., PELLOUX, J., VAN WUYTSWINKEL, O., BELLINI, C. & GUTIERREZ, L. 2009. Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *Journal of Experimental Botany,* 60**,** 487-493.

GUTIÉRREZ, R. A., MACINTOSH, G. C. & GREEN, P. J. 1999. Current perspectives on mRNA stability in plants: multiple levels and mechanisms of control. *Trends in plant science,* 4**,** 429-438.

HALLMANN, A. 2007. Algal transgenics and biotechnology. *Transgenic Plant J,* 1**,** 81-98.

HANAHAN, D. 1983. Studies on transformation of Escherichia coli with plasmids. *Journal of molecular biology,* 166**,** 557-580.

HASLE, G. R. & FRYXELL, G. A. 1970. Diatoms: Cleaning and Mounting for Light and Electron Microscopy. *Transactions of the American Microscopical Society,* 89**,** 469-474.

HAWKINS, T. L., O'CONNOR-MORIN, T., ROY, A. & SANTILLAN, C. 1994. DNA purification and isolation using a solid-phase. *Nucleic acids research,* 22**,** 4543.

HEITZER, M., ECKERT, A., FUHRMANN, M. & GRIESBECK, C. 2007. Influence of codon bias on the expression of foreign genes in microalgae. *Transgenic Microalgae as Green Cell Factories.* Springer, 46-53.

HELLEMANS, J., MORTIER, G., DE PAEPE, A., SPELEMAN, F. & VANDESOMPELE, J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome biology,* 8**,** R19.

HERMAN, E., SACHSE, M., KROTH, P. G. & KOTTKE, T. 2013. Blue-Light-Induced Unfolding of the Jα Helix Allows for the Dimerization of Aureochrome-LOV from the Diatom Phaeodactylum tricornutum. *Biochemistry,* 52**,** 3094-3101.

HOUSELEY, J. & TOLLERVEY, D. 2009. The many pathways of RNA degradation. *Cell,* 136**,** 763-776.

HOWELL, D. M., XU, H. & WHITE, R. H. 1999. (R)-citramalate synthase in methanogenic archaea. *J Bacteriol,* 181**,** 331-333.

HUGGETT, J., DHEDA, K., BUSTIN, S. & ZUMLA, A. 2005. Real-time RT-PCR normalisation; strategies and considerations. *Genes and immunity,* 6**,** 279-284.

HUYSMAN, M. J., FORTUNATO, A. E., MATTHIJS, M., COSTA, B. S., VANDERHAEGHEN, R., VAN DEN DAELE, H., SACHSE, M., INZÉ, D., BOWLER, C. & KROTH, P. G. 2013a. AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (Phaeodactylum tricornutum). *The Plant Cell Online,* 25**,** 215-228.

HUYSMAN, M. J., VYVERMAN, W. & DE VEYLDER, L. 2013b. Molecular regulation of the diatom cell cycle. *Journal of Experimental Botany***,** ert387.

JAKOBY, M., WEISSHAAR, B., DROGE-LASER, W., VICENTE-CARBAJOSA, J., TIEDEMANN, J., KROJ, T. & PARCY, F. 2002. bZIP transcription factors in Arabidopsis. *Trends Plant Sci,* 7**,** 106-111.

JIROUTOVÁ, K., KOŘENÝ, L., BOWLER, C. & OBORNÍK, M. 2010. A gene in the process of endosymbiotic transfer. *PLoS One,* 5**,** e13234.

JOHNSON, P. H. & GROSSMAN, L. I. 1977. Electrophoresis of DNA in agarose gels. Optimizing separations of conformational isomers of double-and single-stranded DNAs. *Biochemistry,* 16**,** 4217-4225.

JØRSTAD, T. S., LANGAAS, M. & BONES, A. M. 2007. Understanding sample size: what determines the required number of microarrays for an experiment? *Trends in plant science,* 12**,** 46-50.

KAPOOR, S., WAKASUGI, T., DENO, H. & SUGIURA, M. 1994. An atpE-specific promoter within the coding region of the atpB gene in tobacco chloroplast DNA. *Current Genetics,* 26**,** 263-268.

KEELING, P. J. & PALMER, J. D. 2008. Horizontal gene transfer in eukaryotic evolution. *Nature Reviews. Genetics,* 9**,** 605-618.

KIANIANMOMENI, A. & HALLMANN, A. 2014. Algal photoreceptors: in vivo functions and potential applications. *Planta,* 239**,** 1-26.

KIM, J., SMITH, J. J., TIAN, L. & DELLAPENNA, D. 2009. The evolution and function of carotenoid hydroxylases in Arabidopsis. *Plant and Cell Physiology,* 50**,** 463-479.

KLEIN, T. M., HARPER, E. C., SVAB, Z., SANFORD, J. C., FROMM, M. E. & MALIGA, P. 1988. Stable genetic transformation of intact Nicotiana cells by the particle bombardment process. *Proceedings of the National Academy of Sciences,* 85**,** 8502-8505.

KORBONITS, M., BUSTIN, S. A., KOJIMA, M., JORDAN, S., ADAMS, E. F., LOWE, D. G., KANGAWA, K. & GROSSMAN, A. B. 2001. The Expression of the Growth Hormone Secretagogue Receptor Ligand Ghrelin in Normal and Abnormal Human Pituitary and Other Neuroendocrine Tumors. *The Journal of Clinical Endocrinology & Metabolism,* 86**,** 881-887.

KOYANAGI, K. O., HAGIWARA, M., ITOH, T., GOJOBORI, T. & IMANISHI, T. 2005. Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. *Gene,* 353**,** 169-176.

KROTH-PANCIC, P. G., FREIER, U., STROTMANN, H. & KOWALLIK, K. V. 1995. MOLECULAR STRUCTURE AND EVOLUTION OF THE CHLOROPLAST atpB/E GENE CLUSTER IN THE DIATOM ODONTELLA SINENSIS. *Journal of Phycology,* 31**,** 962-969.

KROTH, P. G. 2007. Genetic Transformation. *Protein Targeting Protocols.* Springer, 257-267.

KRUG, M. S. 1987. First-strand cDNA synthesis primed with oligo (dT). *Methods Enzymol.;(United States),* 152.

KUMAR, S. V., MISQUITTA, R. W., REDDY, V. S., RAO, B. J. & RAJAM, M. V. 2004. Genetic transformation of the green alga-Chlamydomonas reinhardtii by Agrobacterium tumefaciens. *Plant science,* 166**,** 731-738.

LANG, M. & KROTH, P. G. 2001. Diatom fucoxanthin chlorophyll a/c-binding protein (FCP) and land plant light-harvesting proteins use a similar pathway for thylakoid membrane insertion. *Journal of Biological Chemistry,* 276**,** 7985-7991.

LEE, J. H., HÜBEL, A. & SCHÖFFL, F. 1995. Derepression of the activity of genetically engineered heat shock factor causes constitutive synthesis of heat shock proteins and increased thermotolerance in transgenic Arabidopsis. *The Plant Journal,* 8**,** 603-612.

LIPSHUTZ, R. J., FODOR, S. P., GINGERAS, T. R. & LOCKHART, D. J. 1999. High density synthetic oligonucleotide arrays. *Nature genetics,* 21**,** 20-24.

LITTLE, M. C. 1991. Process for the purification of DNA on diatomaceous earth. Google Patents.

LOCKHART, J. 2013. Blue Light Checkpoint: How Blue Light Controls the Onset of Cell Division in Diatoms. *The Plant Cell Online,* 25**,** 1.

LOENEN, W. A. M., DRYDEN, D. T. F., RALEIGH, E. A., WILSON, G. G. & MURRAY, N. E. 2014. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic acids research,* 42**,** 3-19.

LOSIC, D., ROSENGARTEN, G., MITCHELL, J. G. & VOELCKER, N. H. 2006. Pore architecture of diatom frustules: potential nanostructured membranes for molecular and particle separations. *Journal of Nanotechnological Sciences,* 6**,** 1-8.

MANDEL, M. & HIGA, A. 1970. Calcium-dependent bacteriophage DNA infection. *Journal of molecular biology,* 53**,** 159-162.
MANN, D. G. 1999. The species concept in diatoms. *Phycologia,* 38**,** 437-495.
MAYFIELD, S. P. & KINDLE, K. L. 1990. Stable nuclear transformation of Chlamydomonas reinhardtii by using a C. reinhardtii gene as the selectable marker. *Proceedings of the National Academy of Sciences,* 87**,** 2087-2091.
MCDADE, H. & COX, G. 2001. A new dominant selectable marker for use in Cryptococcus neoformans. *Medical Mycology,* 39**,** 151-154.
MEDLIN, L. K. & KACZMARSKA, I. 2004. Evolution of the diatoms: V. Morphological and cytological support for the major clades and a taxonomic revision. *Phycologia,* 43**,** 245-270.
MEUNIER, B., DE VISSER, S. P. & SHAIK, S. 2004. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chemical reviews,* 104**,** 3947-3980.
MIKI, B. & MCHUGH, S. 2004. Selectable marker genes in transgenic plants: applications, alternatives and biosafety. *Journal of Biotechnology,* 107**,** 193-232.
MIYAHARA, M., AOI, M., INOUE-KASHINO, N., KASHINO, Y. & IFUKU, K. 2013. Highly efficient transformation of the diatom Phaeodactylum tricornutum by multi-pulse electroporation. *Biosci. Biotechnol. Biochem,* 77**,** 874-876.
MOREY, J. S., RYAN, J. C. & VAN DOLAH, F. M. 2006. Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. *Biological Procedures Online,* 8**,** 175-193.
MORIMOTO, R. I. 1998. Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. *Genes & development,* 12**,** 3788-3796.
MÜHLROTH, A., LI, K., RØKKE, G., WINGE, P., OLSEN, Y., HOHMANN-MARRIOTT, M. F., VADSTEIN, O. & BONES, A. M. 2013. Pathways of Lipid Metabolism in Marine Algae, Co-Expression Network, Bottlenecks and Candidate Genes for Enhanced Production of EPA and DHA in Species of Chromista. *Marine drugs,* 11**,** 4662-4697.
NAKAI, A. 1999. New aspects in the vertebrate heat shock factor system: Hsf3 and Hsf4. *Cell stress & chaperones,* 4**,** 86.
NAKAMURA, M. T. & NARA, T. Y. 2004. Structure, function, and dietary regulation of Δ6, Δ5, and Δ9 desaturases. *Nutrition,* 24.
NILSSON, S. V. & MAGNUSSON, G. 1982. Sealing of gaps in duplex DNA by T4 DNA ligase. *Nucleic Acids Res,* 10**,** 1425-1437.
NIU, Y., YANG, Z., ZHANG, M., ZHU, C., YANG, W., LIU, J. & LI, H. 2012. Transformation of diatom Phaeodactylum tricornutum by electroporation and establishment of inducible selection marker. *Biotechniques***,** 1-3.
NOVER, L., SCHARF, K.-D., GAGLIARDI, D., VERGNE, P., CZARNECKA-VERNER, E. & GURLEY, W. B. 1996. The Hsf world: classification and properties of plant heat stress transcription factors. *Cell stress & chaperones,* 1**,** 215.
NYMARK, M. 2013. *Light responses in the marine diatom Phaeodactylum tricornutum.*
NYMARK, M., VALLE, K. C., BREMBU, T., HANCKE, K., WINGE, P., ANDRESEN, K., JOHNSEN, G. & BONES, A. M. 2009. An integrated analysis of molecular acclimation to high light in the marine diatom Phaeodactylum tricornutum. *PLoS One,* 4**,** e7743.
NYMARK, M., VALLE, K. C., HANCKE, K., WINGE, P., ANDRESEN, K., JOHNSEN, G., BONES, A. M. & BREMBU, T. 2013. Molecular and photosynthetic responses to

prolonged darkness and subsequent acclimation to re-illumination in the diatom Phaeodactylum tricornutum. *PLoS One,* 8**,** e58722.
O'NEILL, M., MCPARTLIN, J., ARTHURE, K., RIEDEL, S. & MCMILLAN, N. Year. Comparison of the TLDA with the Nanodrop and the reference Qubit system. *In:* Journal of Physics: Conference series, 2011. IOP Publishing, 012047.
OBERTHUER, A., JURAEVA, D., LI, L., KAHLERT, Y., WESTERMANN, F., EILS, R., BERTHOLD, F., SHI, L., WOLFINGER, R. D., FISCHER, M. & BRORS, B. 2010. Comparison of performance of one-color and two-color gene-expression analyses in predicting clinical endpoints of neuroblastoma patients. *Pharmacogenomics J,* 10**,** 258-266.
OGURA, Y., TOKUTOMI, S., WADA, M. & KIYOSUE, T. 2008. PAS/LOV proteins: A proposed new class of plant blue light receptor. *Plant signaling & behavior,* 3**,** 966.
PETTERSSON, E., LUNDEBERG, J. & AHMADIAN, A. 2009. Generations of sequencing technologies. *Genomics,* 93**,** 105-111.
PFAFFL, M. 2009. Rest 2009 Software user guide. *Qiagen, Hilden, Germany*.
PFAFFL, M. W., HORGAN, G. W. & DEMPFLE, L. 2002. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic acids research,* 30**,** e36.
POULSEN, N., CHESLEY, P. M. & KRÖGER, N. 2006. MOLECULAR GENETIC MANIPULATION OF THE DIATOM THALASSIOSIRA PSEUDONANA (BACILLARIOPHYCEAE) 1. *Journal of Phycology,* 42**,** 1059-1065.
PRÄNDL, R., HINDERHOFER, K., EGGERS-SCHUMACHER, G. & SCHÖFFL, F. 1998. HSF3, a new heat shock factor from Arabidopsis thaliana, derepresses the heat shock response and confers thermotolerance when overexpressed in transgenic plants. *Molecular and General Genetics MGG,* 258**,** 269-278.
QIAGEN. 2004. *QIAprep® Miniprep Handbook* [Online]. Available: http://public.wsu.edu/~kahn_sci/Flow/E2-QIAprep_Miniprep_Handbook.pdf [Accessed 19.05.2013].
QIAGEN. 2009. *QuantiTect® Reverse Transcription Handbook* [Online]. Available: http://www.qiagen.com/resources/Download.aspx?id={F0DE5533-3DD1-4835-8820-1F5C088DD800}&lang=en&ver=1 [Accessed 13.03.14].
QIAGEN. 2010. *RNeasy® Mini Handbook* [Online]. Available: http://www.qiagen.com/resources/Download.aspx?id={14E7CF6E-521A-4CF7-8CBC-BF9F6FA33E24}&lang=en&ver=1 [Accessed 16.03.14].
QIN, S., SUN, G.-Q., JIANG, P., ZOU, L.-H., WU, Y. & TSENG, C.-K. Year. Review of genetic engineering of Laminaria japonica (Laminariales, Phaeophyta) in China. *In:* Sixteenth International Seaweed Symposium, 1999. Springer, 469-472.
RADONIĆ, A., THULKE, S., MACKAY, I. M., LANDT, O., SIEGERT, W. & NITSCHE, A. 2004. Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications,* 313**,** 856-862.
RAMAKERS, C., RUIJTER, J. M., DEPREZ, R. H. L. & MOORMAN, A. F. M. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters,* 339**,** 62-66.
RAYKO, E., MAUMUS, F., MAHESWARI, U., JABBARI, K. & BOWLER, C. 2010. Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytologist,* 188**,** 52-66.
ROUND, F. E., CRAWFORD, R. M. & MANN, D. G. 1990. *The diatoms: the biology and morphology of the genera*, Cambridge University Press.
RUIJTER, J. M. & RAMAKERS, C. 2003. LinRegPCR: Analysis of real-time quantitative PCR data. *User Manual provided with software*.

RUYTERS, G. 1980. Blue light-effects on enzymes of the carbohydrate metabolism in Chlorella 1. Pyruvate kinase. *The blue light syndrome.* Springer, 361-367.
SAEIDNIA, S. & ABDOLLAHI, M. 2013. Are other fluorescent tags used instead of ethidium bromide safer? *Daru,* 21**,** 71.
SANCHEZ, S. M. & VOLTOLINA, D. 1994. chemical composition of Chaetoceros sp.(Bacillariophyceae) under different light conditions. *Comparative biochemistry and physiology. B: Comparative biochemistry*.
SANFORD, J., SMITH, F. & RUSSELL, J. 1993. Optimizing the biolistic process for different biological applications. *Methods in enzymology,* 217**,** 483.
SANFORD, J. C. 1990. Biolistic plant transformation. *Physiologia Plantarum,* 79**,** 206-209.
SCHAUDER, B., BLÖCKER, H., FRANK, R. & MCCARTHY, J. E. G. 1987. Inducible expression vectors incorporating the Escherichia coli atpE translational initiation region. *Gene,* 52**,** 279-283.
SCHROEDER, A., MUELLER, O., STOCKER, S., SALOWSKY, R., LEIBER, M., GASSMANN, M., LIGHTFOOT, S., MENZEL, W., GRANZOW, M. & RAGG, T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC molecular biology,* 7**,** 3.
SCHULZE, A. & DOWNWARD, J. 2001. Navigating gene expression using microarrays - a technology review. *Nat Cell Biol,* 3**,** E190-E195.
SCHWARTZ, A., KOMAROVA, T., SKULACHEV, M., ZVEREVA, A., DOROKHOV, Y. L. & ATABEKOV, J. 2006. Stability of plant mRNAs depends on the length of the 3′-untranslated region. *Biochemistry (Moscow),* 71**,** 1377-1384.
SEAMAN, M. N., MARCUSSON, E. G., CEREGHINO, J. L. & EMR, S. D. 1997. Endosome to Golgi retrieval of the vacuolar protein sorting receptor, Vps10p, requires the function of the VPS29, VPS30, and VPS35 gene products. *The Journal of cell biology,* 137**,** 79-92.
SGARAMELLA, V. & EHRLICH, S. D. 1978. Use of the T4 Polynucleotide Ligase in The Joining of Flush-Ended DNA Segments Generated by Restriction Endonucleases. *European Journal of Biochemistry,* 86**,** 531-537.
SHA, J., WANG, Y., WANG, J., REN, L., TU, Q., LIU, W., WANG, X., LIU, A., WANG, L. & WANG, J. 2011. Capillary-composited microfluidic device for heat shock transformation of Escherichia coli. *Journal of Bioscience and Bioengineering,* 112**,** 373-378.
SHANKLIN, J. & CAHOON, E. B. 1998. Desaturation and Related Modifications of Fatty Acids1. *Annu Rev Plant Physiol Plant Mol Biol,* 49**,** 611-641.
SIAUT, M., HEIJDE, M., MANGOGNA, M., MONTSANT, A., COESEL, S., ALLEN, A., MANFREDONIA, A., FALCIATORE, A. & BOWLER, C. 2007. Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum. *Gene,* 406**,** 23.
SINGH, A., SAHI, C. & GROVER, A. 2009. Chymotrypsin protease inhibitor gene family in rice: Genomic organization and evidence for the presence of a bidirectional promoter shared between two chymotrypsin protease inhibitor genes. *Gene,* 428**,** 9-19.
SINGH, M. V. & ANTHONY WEIL, P. 2002. A method for plasmid purification directly from yeast. *Analytical biochemistry,* 307**,** 13-17.
SNOWDEN, K. C., BUCHHHOLZ, W. G. & HALL, T. C. 1996. Intron position affects expression from thetpi promoter in rice. *Plant molecular biology,* 31**,** 689-692.
STEARS, R. L., MARTINSKY, T. & SCHENA, M. 2003. Trends in microarray analysis. *Nature medicine,* 9**,** 140-145.

STEINBRENNER, J. & SANDMANN, G. 2006. Transformation of the green alga Haematococcus pluvialis with a phytoene desaturase for accelerated astaxanthin biosynthesis. *Applied and environmental microbiology,* 72**,** 7477-7484.

STROMINGER, J. L. 1955. Enzymic synthesis of guanosine and cytidine triphosphates: A note on the nucleotide specificity of the pyruvate phosphokinase reaction. *Biochimica et Biophysica Acta,* 16**,** 616-618.

STRØM MIDTHUN, E. 2012. *Characterization of the Cytochrome p450 Family in the unique Diatom Seminavis robusta.*

SUAREZ, A., GÜTTLER, A., STRÄTZ, M., STAENDNER, L. H., TIMMIS, K. N. & GUZMÁN, C. A. 1997. Green fluorescent protein-based reporter systems for genetic analysis of bacteria including monocopy applications. *Gene,* 196**,** 69-74.

SUETSUGU, N. & WADA, M. 2013. Evolution of Three LOV Blue Light Receptor Families in Green Plants and Photosynthetic Stramenopiles: Phototropin, ZTL/FKF1/LKP2 and Aureochrome. *Plant and Cell Physiology,* 54**,** 8-23.

TANAKA, J. & IKEDA, S. 2002. Rapid and Efficient DNA Extraction Method from Various Plant Species Using Diatomaceous Earth and a Spin Filter. *Breeding Science,* 52**,** 151-155.

TAYLOR, N. J. & FAUQUET, C. M. 2002. Microparticle bombardment as a tool in plant science and agricultural biotechnology. *DNA and cell biology,* 21**,** 963-977.

TOCHER, D. R., LEAVER, M. J. & HODGSON, P. A. 1998. Recent advances in the biochemistry and molecular biology of fatty acyl desaturases. *Progress in Lipid Research,* 37**,** 73-117.

TOYOOKA, T., HISATOMI, O., TAKAHASHI, F., KATAOKA, H. & TERAZIMA, M. 2011. Photoreactions of Aureochrome-1. *Biophysical Journal,* 100**,** 2801-2809.

VANSTECHELMAN, I., SABBE, K., VYVERMAN, W., VANORMELINGEN, P. & VUYLSTEKE, M. 2013. Linkage mapping identifies the sex determining region as a single locus in the pennate diatom Seminavis robusta. *PLoS One,* 8**,** e60132.

VARDI, A., BIDLE, K. D., KWITYN, C., HIRSH, D. J., THOMPSON, S. M., CALLOW, J. A., FALKOWSKI, P. & BOWLER, C. 2008. A Diatom Gene Regulating Nitric-Oxide Signaling and Susceptibility to Diatom-Derived Aldehydes. *Current Biology,* 18**,** 895-899.

WAHL, G. M., LEWIS, K. A., RUIZ, J. C., ROTHENBERG, B., ZHAO, J. & EVANS, G. A. 1987. Cosmid vectors for rapid genomic walking, restriction mapping, and gene transfer. *Proceedings of the National Academy of Sciences,* 84**,** 2160-2164.

WALKER, T. L., COLLET, C. & PURTON, S. 2005. Algal Transgenics in the Genomic Era. *Journal of Phycology,* 41**,** 1077-1095.

WERCK-REICHHART, D. & FEYEREISEN, R. 2000. Cytochromes P450: a success story. *Genome Biol,* 1**,** 3003.3001-3003.3009.

XIE, G. & ALLISON, L. 2002. Sequences upstream of the YRTA core region are essential for transcription of the tobacco atpB NEP promoter in chloroplasts in vivo. *Current Genetics,* 41**,** 176-182.

XU, H., ZHANG, Y., GUO, X., REN, S., STAEMPFLI, A. A., CHIAO, J., JIANG, W. & ZHAO, G. 2004. Isoleucine biosynthesis in Leptospira interrogans serotype lai strain 56601 proceeds via a threonine-independent pathway. *J Bacteriol,* 186**,** 5400-5409.

YANG, D. Y., ENG, B., WAYE, J. S., DUDAR, J. C. & SAUNDERS, S. R. 1998. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *American journal of physical anthropology,* 105**,** 539-543.

ZASLAVSKAIA, L. A., LIPPMEIER, J. C., KROTH, P. G., GROSSMAN, A. R. & APT, K. E. 2000. Transformation of the diatom Phaeodactylum tricornutum

(Bacillariophyceae) with a variety of selectable marker and reporter genes. *Journal of Phycology*, 36**,** 379-386.

ZHANG, G. F. & STAEHELIN, L. A. 1992. Functional Compartmentation of the Golgi Apparatus of Plant Cells: Immunocytochemical Analysis of High-Pressure Frozen- and Freeze-Substituted Sycamore Maple Suspension Culture Cells. *Plant physiology*, 99**,** 1070-1083.

ZHANG, S. & CAHALAN, M. D. 2007. Purifying plasmid DNA from bacterial colonies using the QIAGEN Miniprep Kit. *Journal of visualized experiments: JoVE*.

ZOU, D., BI, Y. & ZHOU, Z. 2009. Effects of light on the relative transcription of lhcf5 and lhcf6 genes from the gametophytes of Laminaria japonica. *Journal of Fishery Sciences of China*, 16**,** 850-858.

APPENDIX I – Solutions for diatom culturing

F/2 medium	
Components (g/L dH2O)	**Volume (mL)**
$NaNO_3$ (75)	1
NaH_2PO_4 x H2O (5.65)	1
Na_2SiO_3 x 9H2O (30)	1
Trace metal solution (*)	1
Vitamin solution (**)	0,5
Sterile filtered, autoclaved sea water	1000

f/2 trace metal solution	
Components (g/ L dH2O)	**Volume (mL)**
$CoCl_2$ x 6H2O (10.0)	1
$CuSO_4$ x 5H2O (10.0)	1
$MnCl_2$ x 4H2O (180.0)	1
Na_2MoO4 x 2H2O (6.3)	1
$ZnSO_4$ x 7H2O (22.0)	1
Sterile filtered, autoclaved sea water	950

Vitamin solution	
Components	**Amount (g)**
Biotin	0.0005
Cyanocolbalamin (Vitamin B_{12})	0.0005
Thiamine HCl (Vitamin B_1)	0.1

The mixture was dissolved in 1000mL sterile filtered, autoclaved sea water and the solution was adjusted to pH 8 with 1M HCl or NaOH. The mixture was sterile filtered though a 0.20 µm filter and stored at 4^oC ready for use.

Antibiotic mix for axenization	
Components	**Volume (mL)**
Cefortaxim (Claforan) (250 mg/ml)	0.5
Gentamycin (20 mg/ml)	1
Penicillin G (100 000 U/ml)	0.5
Polymixin B (250 000 U/ml)	0,050
Milli Q-water	5

To prevent precipitation Cefortaxim must be added as last components in the mixture.

f/2 agar plates	
Components	Volume (mL)
Sterile filtered sea water	450
$NaNO_3$ (75)	1
NaH_2PO_4 x H2O (5.65)	1
Na_2SiO_3 x 9H2O (30)	1
Trace metal solution (*)	1
Vitamin solution (**)	0,5
MilliQ water	500
Bacto agar	10 g
Nourseothricin (100mg/mL)	1 mL

The bacto agar, sterile filtered sea water and milliQ water was first autoclaved. The solution was set too cool down before rest of the components were added. The mixed solution was then poured into Petri dishes and stored in 4^o after the solution had solidified.

ISOLATION OF GENOMIC DNA FROM S.ROBUSTA

The isolation of genomic DNA was done as described in Brembu et al. (Brembu et al., 2013), which is a modification of the protocol from Bowler et al. (Bowler et al., 2008).

Procedure:
The isolation of genomic DNA was performed by principal engineer Torfinn Sparstad and researcher Tore Brembu and was done as described in Brembu et al. (Brembu et al., 2013), which is a modification of the protocol from Bowler et al. (Bowler et al., 2008).

Procedure:
1. A culture of six liters of *S. robusta* in late exponential phase was centrifuged for ten minutes at 2000 g at four Celsius.

2. The pellet was then collected and frozen in liquid nitrogen and then resuspended in a lysis buffer (50 mM Tris – HCl (pH 8,0, 50 mM) EDTA (pH 8,0), 1% SDS, 10 mM DTT and10 mg/mL of proteinase K). A volume of 10 ml lysis buffer was added to every one liter culture.

3. The resuspention was incubated for 45 min at 50 ^{o}C.

4. To remove proteins, three phenol/chloroform extractions were done. After the first phenol/chloroform extraction RNase was added to the lysate, and the solution was incubated for 60 min at 37^o C.

5. To remove phenol residues from the solution an extraction was made with chloroform isoamyl alcohol. By using ethanol and NaCl, genomic DNA was precipitated and collected with the use of a glass rod.

6. The DNA was transferred to a 15mL tube and mixed with 70% ethanol and incubated overnight at 4^{o} C.
7. The next day the DNA pellet was again washed in 70% ethanol and air-dried, before it was resuspended in TE-buffer.

8. The concentration of DNA was determined with spectrophotomertry, fluorometry and agarose gel (0.8%).

Appendix II – Buffers and solutions

1% Agarose gel (50mL)	
Agarose	0,5 g
1 x TAE buffer (1mM EDTA)	50 mL
GelRed (Biotium, cat. no. 41003-1-10ml)	2,5 µL

Restriction enzymes used for vector construction

Restriction enzyme	**Buffer**	**Optimums temperature (°C)**	**New England cat.no**	**Used to insert**
EcoRI	NEBuffer EcoRI	37	R0101	AtpBE promoter and *nat1* gene
HindIII	NEBuffer 2	37	R0104	*nat1* gene and AtpB terminator
SacI	NEBuffer 4	37	R0156	AtpE terminator
SalI	NEBuffer 3	37	R0138	AtpE terminator
SmaI	NEBuffer 4	25	R0141	*nat1* gene and AtpBE promoter
XbaI	NEBuffer 4	37	R0145	AtpE terminator
ScaI	NEBuffer 3	37	R0122	Linearization of vector

Recipe for S.O.C medium

Components	Amount
Bacto-Tryptone (Becton, Dickinson and company, cat. no 21705)	2 g
Bacto™ Yeast extract (Becton, Dickinson and company, cat. no 212750)	0,5 g
NaCl	0,05 g
KCl	0,02 g
$MgCl_2$	0,09 g
MilliQ water	98 mL
Sterile filtered Glucose (1M)*	2 mL

*The sterile filtered glucose was added after the rest of the components had been autoclaved.

Appendix III – PCR cycle and primers

The PCR program that was used

Step	Temperature setting	Duration
1	94°C	5 min
2	94 °C	30 sec
3	55 °C*	30 sec
4	72 °C	2 min
5	Repeat steps 2-4	30 repeats
6	4 °C	∞

* The annealing temperature was adjusted to match the melting temperature for the primers used for the specific reaction.

Primers used for vector construct

Primer name	Used to isolate	Melting temperature (°C)	Sequence (5'-3')	Included restriction site
AtpEt_Sac1	AtpE terminator	76,6	CTGCGACTCGCTATTCCCGAACAGTAGAC	SacI
AtpEt_Xba1	AtpE terminator	73,5	CATGTCTAGAGCCATTGGAGTCAACCTCAGT	XbaI
AtpEpSma1	AtpB/E promoter	77,8	CGGTCCCGGGCAACATGGTCAATCTAATAGT	SmaI
AtpBpEcoRI	AtpB/E promoter	69,9	GTCGGAATTCTATCATTGTCTCGTATTGTTG	EcoRI
SrTBHf	AtpB terminator	79,1	CTGCAAGCTTGGCAATGTAAACTGACTGATGTGGA	HindIII
SrTBSr	AtpB terminator	75,6	GACTGTCGACATGGTCGTCTCCAATACACC	SalI
SrYFPf	YFP	74,1	CATTTAGAATTCACCATGGGCAAGGGGC	EcoRI
SrYFPr	YFP	72	TCGAAAGCTTAGTCCTTACTTGTACAGCTCGTC	HindIII

Primers used for RT-PCR

Primer name	Used to verify	Melting temperature (°C)	Sequence (5'-3')
NAT1F	Nat1 gene	67,6	TACCGCTTACCGATACCGAACC
NAT1R	Nat1 gene	71,3	CATTCGACGGTATGCGTGGATG
YFPf	YFP	67	GACCCTGAAGTTCATCTGCACC
YFPr	YFP	66,9	GCGGATCTTGAAGTTCACCTTG

Primers used for qPCR		
Primer name	Target gene	Sequence (5'-3')
SrLHCR6F	LHCR6	GAGACTCCACTCTTCTTCAAGC
SrLHCR6R		GAACCCTCAGCTCGCTCTTAGT
SrLHCF8f	LHCF8	CTTCTTGGTCTTATGGTTCACG
SrLHCf8r		GCAGTGAGAGAATCCCCGTTGT
SrD9F	D9	GAAGCCTTTTCCATTGGTCTCG
SrD9R		GAAGCCTCTGGTAGCGTTGTGA
SrPAS9f	PAS9	TGTGCATCTTCGACAAACATCC
SrPAS9r		GTAACTTCGTCGTTTGCATTGG
SrHSF4f	HSF4	GTAGGGTTGCATCAAGCATGTC
SrHSF4r		GAATGCCACCTATCATCTCAGT
SrEXPf	EXP	CCTTCGGGAGCATATATCGTC
SrEXPr		GAACATACCCTCTACGAATCTG
Sec7F	Sec7	CACTGGAGACCATGTCGATTAC
Sec7R		TCCAGATCTATTGTCCATTAGG
SrCyp15F	CYP15	GCCATTCCGCCTGAAGAAGTTG
SrCyp15R		GAACCATTGCCAAAGACGGATG
SrCyp21F	CYP21	TAGCTGTCATGCAAAGTTGAGC
SrCyp21R		TGTATCCTCTTGTATGGCGAGT
SrCyp23F	CYP23	CAGATGTTTGGTGATCAAACTG
SrCyp23R		TGCGAGATGAACAAGCTAGGTT

Appendix IV–Vector constructs parts

Modified *nat1* gene

```
ATGGGAACCACTCTGGACGATACCGCTTACCGATACCGAACCAGTGTCCCAGGAGACGCC
GAAGCCATTGAAGCACTGGATGGATCCTTCACTACCGAAACTGTCTTCCGAGTCACCGCC
ACTGGAGAAGGCTTCACCCTGCGAGAGGTGCCAGTGGACCCACCCCTGACTAAGGTGTTC
CCAGATGACGAATCCGATGACGAATCGGATGACGGAGAGGATGGAGACCCAGATTCCCGG
ACGTTCGTCGCCTATGGGGACGACGGCGATCTGGCAGGCTTCGTGGTCGTCTCGTATTCC
GGATGGAACCGACGACTGACCGTCGAGGATATCGAGGTCGCCCCAGAGCACCGCGGACAC
GGTGTCGGACGCGCTTTGATGGGCCTCGCGACAGAGTTTGCCCGAGAACGGGGTGCCGGA
CACCTCTGGCTGGAGGTCACTAACGTCAACGCACCTGCCATCCACGCATACCGTCGAATG
GGTTTTACTCTCTGCGGCCTGGATACCGCCCTGTACGATGGAACCGCCTCGGATGGCGAA
CAAGCGCTCTACATGAGCATGCCCTGCCCATAGCTCGAGGCATGCGGTACCCACGTGTTA
TGCTTCCGGCTCCTATGTTGTGTGGAATTGTG
```

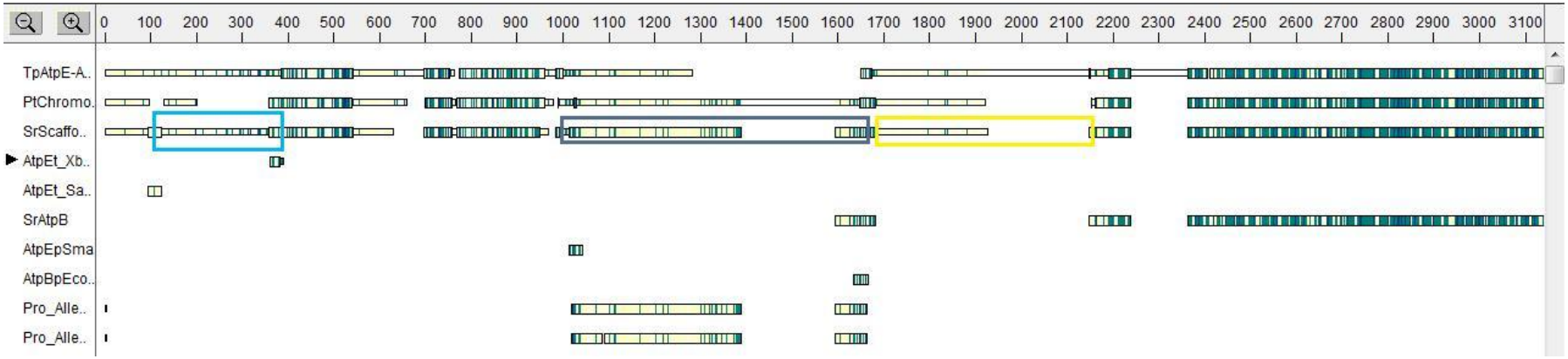

Figure 22: The AtpBE promoter is marked with a dark blue box, showing where in the scaffold the promoter is. The yellow box indicates the intron sequence that is present in the AtpB gene, which might contain transcriptional information required for gene expression.