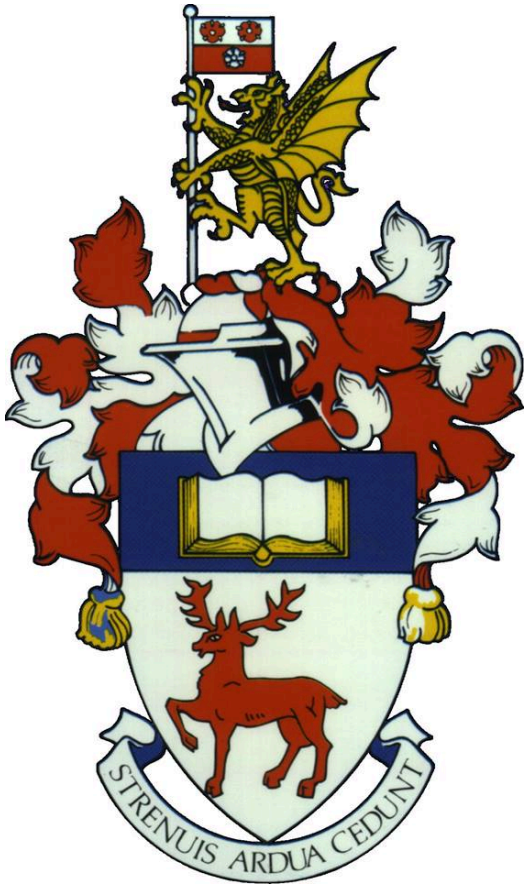
UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

School of Ocean and Earth Sciences



Biotechnological development of the marine microalgae *Nannochloropsis gaditana*

by

Harry Oliver Jackson

Thesis for the degree of Doctor of Philosophy

May 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

School of Ocean and Earth Sciences

Doctor of Philosophy

BIOTECHNOLOGICAL DEVELOPMENT OF THE MARINE MICROALGAE *NANNOCHLOROPSIS GADITANA*

Harry Oliver Jackson

Marine microalgae are promising systems for biotechnology due to both their diversity and ability to produce a vast array of metabolites, many of which are precursors in the production of high-value products including fine chemicals and pharmaceuticals. Algae are also a potential source of sustainable and carbon neutral biofuels although large-scale production systems are currently in their infancy. To fully realize the potential of marine microalgae, non-model biotechnologically relevant strains must be further improved using genetic engineering in order to increase efficiency of production for commercial exploitation. However, the necessary genetic tools for such approaches are in need of development on a species-to-species basis before these needs can be met. This thesis describes the establishment and development of principle molecular technologies in the oleaginous biofuel candidate alga *Nannochloropsis gaditana.* An *in situ* fluorescent reporter system was established and used to characterize novel genetic elements; including the main component of an inducible expression system capable of precise temporal control of expression of transgenes through alteration of the growth media with no negative effect on growth rates. Additionally, a promoter-trapping pipeline was designed for the discovery of promoters that are highly expressed under specific environmental conditions. Also in *N. gaditana*, high-throughput screening of mutant strains was used to isolate cell lines with increased resistance to oxidative stress that exhibited improved growth in a variety of high light treatments, as well as strains with increased lipid accumulation. These contributions to the molecular toolkit for *N. gaditana* will further develop this phenotypically favorable strain towards its amenability to highly sophisticated genetic engineering approaches. Several of the technologies and strategies described may also have wider implications for improvement of other species of marine microalgae for biotechnology.

Breathe, breathe in the air.
Don't be afraid to care.
Leave but don't leave me.
Look around and choose your own ground.

Long you live and high you fly
And smiles you'll give and tears you'll cry
And all you touch and all you see
Is all your life will ever be.

Run, rabbit run.
Dig that hole, forget the sun,
And when at last the work is done
Don't sit down it's time to dig another one.

For long you live and high you fly
But only if you ride the tide
And balanced on the biggest wave
You race towards an early grave.

Rick Wright, Roger Waters & David Jon Gilmour. 1973.

This humble piece of work is dedicated to Martin Brian Jackson, Diane Joy Jackson, Alexander William Jackson and Lucy Ellen Jackson.

Table of Contents

List of Tables

List of Figures

Declaration of Authorship

I, Harry Oliver Jackson, declare that this thesis titled, 'A genetic toolkit for biotechnological advancement of marine microalgae' and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. [Delete as appropriate] None of this work has been published before submission [or] Parts of this work have been published as: [please list references below]:

Acknowledgements

I would firstly like to thank Thomas Bibby for his patient and supportive mentoring. Over the course of my PhD it became increasingly apparent how fortunate I was to have Tom as my supervisor. I am immensely grateful for the opportunities made available to me, the consideration given to my development as a scientist and the inspiring and diverse group I was able to work in. Mark Moore and Matthew Terry were excellent co-supervisors of my project, their ability to assimilate information and provided pertinent, thought-provoking and useful feedback, often instantaneously, is greatly appreciated. Thank you.

A huge thank you to Adokiye Berepiki who was a great teacher in the lab and an equally great friend, large parts of the thesis would not have been possible without his input. Thank you also to Andreas Johansson, Despo Polyviou, John Gittins, Nicola Pratt, and Alison Baylay for their guidance and good company. Thank you to everyone at Algenuity for their invaluable support, in the form of expertise, time and for allowing me to use their state of the art technologies.

I would like to thank the Natural Environment Research Council for funding the project, the PHYCONET research council network for funding the collaboration with Algenuity and the organizers of the SPITFIRE DTP at Southampton.

Thank you to my good friends in Southampton, especially Elis Newham, Finn Illsley-Kemp, Joanna Harley, Jack Cowley, Luke Edwards, and Luke Holman for all the laughs, inspiration and trips to The Bookshop Alehouse. Thank you to The Bookshop Alehouse for allowing us admission to their establishment.

Finally, to my family, thank you for equipping me for the challenges of life, inspiring me, encouraging a love of knowledge (and funding it in the case of my parents), and for never missing a fishing trip, philosophical discussion or row over board game rules.

Abbreviations

ATP	**A**denosine **T**ri**P**hosphate
BP	**B**ase **P**air
EMS	**E**thane **M**ethanesulfonate
FAME	**F**atty **A**cid **M**ethyl **E**sters
FTIR	**F**ourier-**T**ransform **I**nfrared Spectroscopy
GFP	**G**reen **F**luorescent **P**rotein
Mb	**M**ega **B**ase
NADPH	**N**icotinamide **A**denine **D**inucleotide **P**hosphate
NPQ	**N**on-**P**hotochemical **Q**uenching
OES	**O**xygen-**E**volving **C**omplex
PMF	**P**roton **M**otive **F**orce
PSI/PSII	**P**hotosystem I / **P**hotosystem II
RACE	**R**apid **A**mplification of **c**DNA **E**nds
RC	**R**eaction **C**entre
ROS	**R**eactive **O**xygen **S**pecies
rDNA	**R**ibosomal **DNA**
RuBisCO	**R**ibulose- 1,5-**B**isphosphate **C**arboxylase/oxygenase
TAG	**T**ri**a**cyl **G**lycerol received

Chapter 1: Introduction

Microalgae are subjects of great attention as promising sources of sustainably produced fatty acids, carotenoids, vitamins, and other valuable compounds (De Jesus Raposo *et al*., 2013). However, natural products in microalgae to a large extent remain unexplored compared to those in land plants despite the fact that microalgae offer many advantages over terrestrial plants, including, rapid growth rates and the lack of a required competition for fresh water or arable land (Hlavova *et al*., 2015). Although microalgae have been used in many biotechnological applications, each new application requires a different species or strain with the required properties. The challenge is therefore to isolate, develop, characterize and optimize species that can express the desired properties. The work in this thesis was carried out in two strains of biotechnologically relevant species of microalgae, *Nannochloropsis gaditana* and *Dunaliella tertiolecta*.

Dunaliella belongs to the Chlorophyceae class of green algae, which are common in marine waters. *Dunaliella* is responsible for the majority of primary production in many hypersaline environments and was first described by Teodoresco in 1905 (Teodoresco, 1905). In the time since its discovery, *Dunaliella* has been a model organism for the study of salt adaptation in algae. Additionally, the large accumulation of *β*-carotene by some species, including *D. tertiolecta,* has lead to commercial applications. The lipid content of *D. tertiolecta* can reach 67% of total biomass, in combination with its high growth rates, the species is also a potential biofuel feedstock (Takagi *et al*., 2006). However, the nuclear genomes of *D. tertiolecta* and other *Dunaliella* sp. have not yet been sequenced, presumably due to the large effort required to sequence these medium size (~100 Mb) eukaryotic genomes (Rismani-Yazdi *et al*., 2011). Furthermore, the nuclear genome of *D. tertiolecta* has not yet been transformed, currently making this a non-model organism for the study of biofuel production. Although, the chloroplast of *D. tertiolecta* shows promise as a host for recombinant protein production (Georgianna *et al*., 2013).

Nannochloropsis belongs to the Eustigmatophyceae class of green algae. The genus of *Nannochloropsis* contains 6 species and was first termed by Hibberd in 1981 (Hibberd, 1981).

Nannochloropsis sp. are mainly found in marine environments but also occur in fresh and brackish water (Fawley and Fawley, 2007). Unlike related microalgae, *Nannochloropsis* has *chlorophyll a* but completely lacks *chlorophyll b* and *c*. Additionally, they are able produce high concentrations of a range of pigments including astaxanthin, zeaxanthin and canthaxanthin (Lubián *et al*., 2000). The main use for *Nannochloropsis* currently is as an energy-rich food source for fish larvae and rotifers. *Nannochloropsis* has been identified as a promising alga for industrial application because of its ability to accumulate high levels of polyunsaturated fatty acids (Sukenik *et al*., 1989; Boussiba *et al*., 1987). The nuclear genome of *N. gaditana* (27.5 Mb; 10,646 genes) has been sequenced and detailed transcriptomic data is also available; in addition, the alga is easily transformable (Radakovits *et al*., 2012; Corteggiani Carpinelli *et al*., 2014; Kilian *et al*., 2011). Sophisticated genetic engineering approaches have been applied in *N. gaditana*, including RNAi (RNA interference) and CRISPR/Cas9 (Ajjawi *et al*., 2017). Several "improved" mutant phenotypes of *N. gaditana* for commercial applications have been generated (Perin *et al*., 2015; Ajjawi *et al*., 2017). These advances have seen *N. gaditana* emerge as a model organism for research in biofuel production from photosynthetic organisms.

In this chapter the fundamentals of photosynthesis are outlined with a description of the molecular apparatus of photosynthesis. The potential of using marine microalgae for biotechnology is then outlined with a discussion of main targets for improvement. Lastly the current methods to achieve these targets and their limitations are addressed.

1.1 Photosynthesis

Photosynthesis is the most important biochemical reaction on the planet. As well as providing the oxygen that allowed the evolution of complex life, all of our food and energy past and present were derived from the process of photosynthesis. Photosynthesis, which literally means building with light, describes the process by which plants and photosynthetic microorganisms convert thermodynamically stable inorganic raw materials into biologically usable organic compounds in the presence of sunlight. The net primary productivity of photosynthesis is estimated at 85 - 120 Gt carbon/year (Imramovsky *et al*., 2011). Thus, understanding the fundamental and applied aspects of photosynthesis is essential to a wide range of sciences and technologies. The simplified equation for photosynthesis can be represented by the equation:

$$CO_2 + H_2O \xrightarrow{\text{sunlight}} \underset{\text{carbohydrate}}{[CH_2O]} + O_2 \qquad (1.1)$$

Through input of the suns energy, the energy-poor starting materials CO_2 and H_2O are converted into energy rich carbohydrates and O_2. The light energy used in this process is derived from

solar radiation between the waveband of 400 – 700 nm, termed 'photosynthetically active radiation' or PAR (Alados *et al*., 1996). The subsidiary reactions that constitute the overall equation shown above can be expressed on an oxidation-reduction scale ('redox potential'), where H_2O is the initial electron donor in oxygenic photosynthesis. Certain species of bacteria use alternative electron donors such as H_2S and AsO_3^{3-} and were thought to be the precursors of oxygenic photosynthesis, indeed the reason for the success of oxygenic photosynthesis (over other types) is that there is essentially a limitless supply of electrons in water (Kulp *et al*., 2008). Electrons are then passed through a series of protein complexes before finally reducing the electron carrier NADP+ to NADPH (nicotinamide adenine dinucleotide phosphate). The catalysts of this reaction are membrane proteins held within thylakoid membranes that also act to spatially separate the protons generated by water splitting from consumption of protons in the formation of NADPH such that a proton gradient is generated. The energy stored in this proton motive force (PMF) is used to convert ADP (adenosine diphosphate) into ATP (adenosine triphosphate) through additions of a molecule of inorganic phosphate (Pi) (Bolton, 1977). The capture of energy from sunlight and storage in these universal energy-storing molecules is commonly referred to as the light reactions, as light is required for the reactions to proceed; the energy stored in the form of ATP and NADPH is subsequently allocated to various processes as needed including the light independent (or dark reactions) that convert CO_2 into carbohydrate or biomass on which all life is dependent. The dark reactions are catalysed by soluble proteins, notably RuBisCO (Ribulose- 1,5-Bisphosphate Carboxylase/oxygenase) which is the most abundant protein on the planet (Griffiths, 2006).

Oxygenic photosynthesis can occur in both prokaryotic cells (cyanobacteia and prochlorophytes) and a wide variety of eukaryotic cells that originally acquired photosynthesis via the primary endosymbiosis event with a prokaryotic cell that subsequently became the chloroplast (Wallin, 1993).

The above text describes the photosynthetic electron transfer chain (PET). The main complexes involved in the electron transport chain of oxygenic photosynthesis include photosystem II (PSII), cytochrome b_6f and photosystem I (PSI). Membrane soluble carrier molecules mediate transfer of electrons between these complexes. The proton motive force generated in tandem with electron transfer is used to drive synthesis of ATP by the ATP synthase, an F-type ATPase also located in the thylakoid membrane (Falkowski and Raven, 1997) (see Figure 1.1).

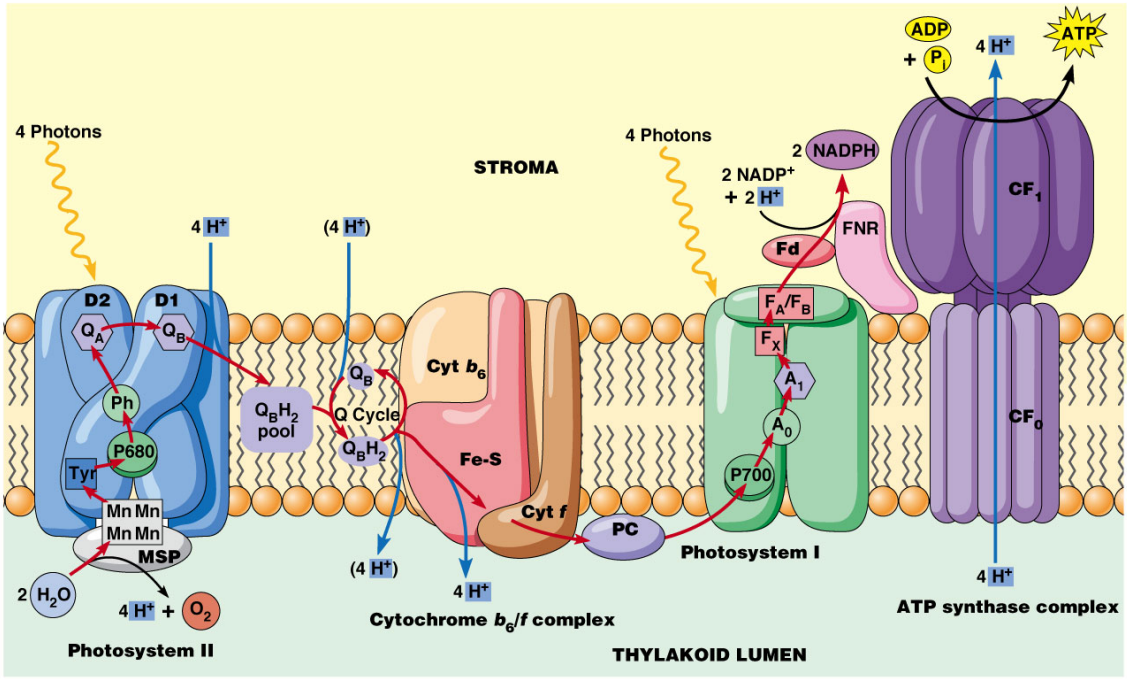


Figure 1.1: Highly simplified representation of oxygenic photosynthesis. PSII includes 25 subunits, subunits D1 and D2 bind chlorophyll P680 and components of the electron transport chain, four manganese atoms facilitate oxidation of water. The Cyt b_6/f contains seven subunits and includes two cytochromes (b_6 and f) and an iron-sulphur protein (Fe-S). PSI contains over seven subunits and binds chlorophyll P700 in addition to several other electron carriers. Electrons are transferred to NADPH via soluble ferredoxin (Fd) using ferredoxin-$NADP^+$ reductase (FNR). The ATP synthase complex uses the proton gradient generated by electron transport to synthesize ATP. Figure from Chernet (2010).

1.1.1 Energy transfer in photosynthesis

The process of oxygenic photosynthesis begins with the absorption of a quanta of light by pigment molecules surrounding PSII. This light energy is then transferred through pigment molecules in the form of excitation until it reaches a 'special pair' of chlorophyll molecules (P680) located at the centre of PSII, also known as the reaction centre (Vredenberg and Duysens, 1963). At the special pair, excitation energy causes an electron to be donated to a molecule of pheophytin – a modified chlorophyll molecule. This first redox reaction results in the oxidation of the P680 pigment to form $P680^+$ and reduction of pheophytin to $pheophytin^-$. The electron is then shuttled to the next electron carrier in the form of a quinone (Q_B) bound to the D2 subunit of PSII. This transfer takes place in 10^{-10} s, minimizing the risk of the electron being passed back to the special pair and being wasted as heat before subsequent transfer to a loosely associated quinone molecule (Q_B). Two high-energy electrons are required to fully reduce Q_B to QH_2, taking up two protons from the cytoplasm in the process. The reduced quinone (plasoquinol) diffuses through the membrane to cytochrome b_6/f. The splitting of H_2O at the oxygen-evolving complex (OEC) of PSII restores the electrons lost in this process (Hill,

1937). Electrons are passed from the OEC to the oxidized $P680^+$ via a tyrosine side chain situated on the D1 subunit of PSII (Nelson and Yocum, 2006). The reaction at the OEC can be summarized as the following:

$$2\,H_2O \longrightarrow O_2 + 4H^4 + 4\,e^- \qquad (1.2)$$

Four electrons are required to balance this equation. However, the reaction centre only has enough reducing power to reduce one electron at a time. At the centre of the OEC resides a manganese-calcium oxide cluster in a cubane-like structure (Umena *et al*., 2011), which serves to temporarily store reducing power until four manganese atoms have been reduced and the splitting of H_2O can proceed. Each photosystem II monomer contains more than 1,300 water molecules some forming extensive hydrogen-bonding networks that may serve as channels for protons, water or oxygen molecules (Ferreira *et al*., 2004). The OEC is located on the lumenal side of the membrane and thus helps to maintain the proton gradient across the thylakoid membrane as well as supplying electrons to PSII.

Two electrons are transferred from PSII to the cytochrome b_6/f complex via reduced membrane-soluble plastoquinones. The cytochrome b_6/f complex provides approximately two thirds of the protons that drive the synthesis of ATP (Hasan *et al*., 2013). Through the action of cyclic electron flow from PSI via ferrodoxin, the complex also plays an important role in balancing the ratio of NADPH and ATP. Electrons flow through an iron-sulphur centre located in the Rieske Fe-S protein to cytochrome f before being passed to a second electron carrier in the form of plastocyanin (Green and Parson, 2003). Firstly, two protons are extracted from the stromal side of the thylakoid membrane and transferred via plastoquinol to the lumenal side. Secondly, another plastoquinol binds to the complex, transferring an electron to plastocyanin via the iron-sulphur centre. Finally, uptake of two more protons from the stromal side regenerates the spent plastoquinone. Overall, four electrons are moved from the stroma to the lumen. Electrons can also be passed from PSI via ferrodoxin in a process known as cyclic electron transport (Barber *et al*., 1997).

PSI accepts electrons from plastocyanin and transfers them either to NADPH (noncyclic electron transport) or back to cytochrome b_6/f (cyclic electron transport). Similar to PSII, PSI is a transmembrane protein complex containing antenna chlorophylls, a reaction centre (P700), phylloquinine, and a number of iron-sulfur proteins that serve as intermediate redox carriers. Upon excitation of the special pair of reaction centre pigments (P700), an electron is donated to a chlorophyll molecule (A_0) and then onto phylloquinone (A_1) before transfer through three iron-sulphur centres (4Fe-4S), F_x, F_A and F_B. The electron carrier in the chain (F_x) reduces ferredoxin. An electron from plastocyanin oxidizes $P700^+$ to regenerate P700. The flow of

electrons through the transfer chain of oxygenic photosynthesis and increase in redox potential of electrons upon excitation at PSII and PSI is commonly referred to as the z-scheme (Figure 1.2) (Allen and Martin, 2007; Senge *et al*., 2014).

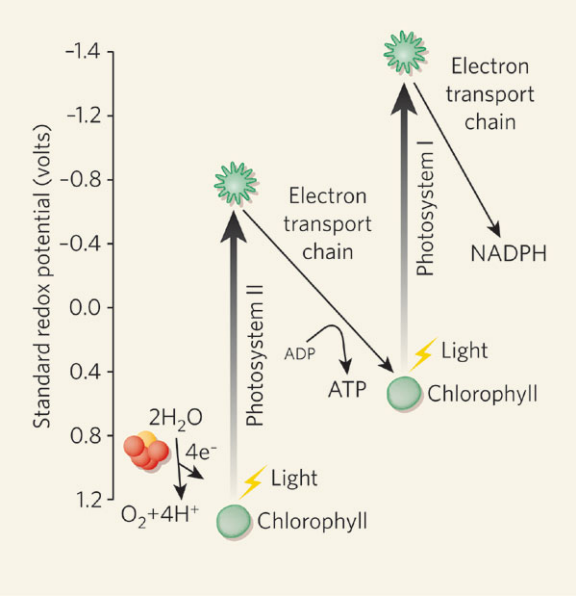


Figure 1.2: The z-scheme of photosynthesis. Photosystems I and II absorb light energy, which is subsequently converted into an electrochemical potential. PSII and PSI are connected in series electrically. The two 'light reactions' of photosynthesis form part of a chain of electron transfers that is coupled, through proton pumping, to synthesis of the energy-storage molecule adenosine triphosphate (ATP). The electron transport chain of photosynthesis ends with photosystem I delivering electrons to $NADP^+$ to form NADPH. ATP and NADPH drive the 'dark reactions' that transfer the electrons to CO_2 to provide the energy to make sugars and the other molecules of life. Figure from Allen and Martin (2007).

1.1.2 The photosynthetic machinery

Photosystem II (PSII) is a water-plastoquinone oxidoreductase comprised of at least 25 subunits. The core of PSII consists of a pseudo-symmetric heterodimer of two homologous proteins D1 and D2 (Rutherford and Faller, 2003). The oxygen-evolving complex is formed of three subunits (PsbO, PsbP and PsbQ). Two antenna proteins CP43 and CP47 form an intrinsic antenna with CP43 also binding the manganese centre. Each photosystem II contains at least 99 cofactors: 35x chlorophyll a, 12x beta-carotene, 2x pheophytin, 2x plastoquinone, 2x heme, 2x bicarbonate, 20x lipid, the Mn_4CaO_5 cluster (including two chloride ions), and one non heme Fe^{2+} and two putative Ca^{2+} ion per monomer (Guskov *et al*., 2009). The outer antenna of PSII is provided by LHCII in eukaryotic species or by phycobilisomes in cyanobacteria and red algae.

The cytochrome b_6f complex is a platoquinol-plastocyanin reductase with a dimer structure of two monomers each comprising eight subunits. Each monomer contains four large subunits: a cytochrome *f* with a c-type cytochrome, a cytochrome b_6f with a heme group, a Rieske iron-sulphur protein containing a [2Fe-2S] cluster and a subunit IV; and four small subunits: PetG, PetL, PetM and PetN (Whitelegge *et al*., 2002). The inter-membrane space within the cytochrome b_6f complex is contains lipids, which provides directionality to heme-heme electron transfer (Hasan and Cramer, 2014).

Photosystem I (PSI) is a plastocyanin-ferredoxin oxidoreductase composed of 14 subunits in eukaryotes. PSI has more than 110 co-factors, significantly more than PSII (Busch and Hippler, 2011). Combined, PSI and LHCI contain 200 chlorophyll molecules, three Fe-S clusters and two phylloquinones (Ben-Shem *et al*., 2003). The electron transfer components of the reaction centre of PSI are a primary electron donor (P700) and five electron acceptors: A0 (chlorophyll), A1 (a phylloquinone) and three 4Fe-4S centres: Fx, Fa, and Fb. Like PSII, PSI has two core homologous subunits, PsaA and PsaB, binding P700, A0, A1, and Fx. The Fx 4Fe-4S centre is bound by four cysteines, two provided by PsaA and two by PsaB. PsaC, PsaO, PsaH and PsaE form the docking site for ferredoxin on the stromal side of the membrane, while PsaN forms the plastocyanin docking site on the lumenal side (Fromme *et al*., 2001).

In eukaryotes, the associated antenna complex of PSII (LHCII) and PSI (LHCI) are composed of a family of apoproteins, Lhcb1 to Lhcb6 and Lhca1 to Lhca4, respectively. Four subcomplexes of LHCII (LHCIIa to LHCIIc) are formed by the apoproteins and associated pigments; two subcomplexes of LHCI are formed, LHCIa and LHCIb. Lhcb1 and Lhcb2 are the most numerous apoproteins and form trimmers, Lhcb4-6 form monomers (Grossman *et al*., 1995).

The ATP synthase consists of two regions, the F_O portion, embedded in the membrane, and the F_1 portion, located outside the membrane (Soares *et al*., 2008). The F_1 portion consists of the subunits, alpha (x3), beta (x3), gamma, delta, and epsilon. The F_O region of ATP synthase consists of subunits a, b (x2) and c (x10-15). Protons travel across the membrane from the intermembrane space to the matrix, down the concentration gradient generated by the electron transfer chain through the c-ring of the F_O region, which rotates during this process. The c-ring is attached to the asymmetric central stalk, composed mainly of the gamma subunit, causing it to rotate within the $alpha_3beta_3$ of the F_1 causing the three catalytic nucleotide binding sites to go through a series of conformational changes that leads to ATP synthesis (Boyer, 1997; Stock *et al*., 1999; Junge and Nelson, 2015). Figure 1.3 shows a more detailed representation of the main proteins and protein complexes involved in photosynthesis carried out in the thylakoid membrane of the chloroplast in eukaryotic species.

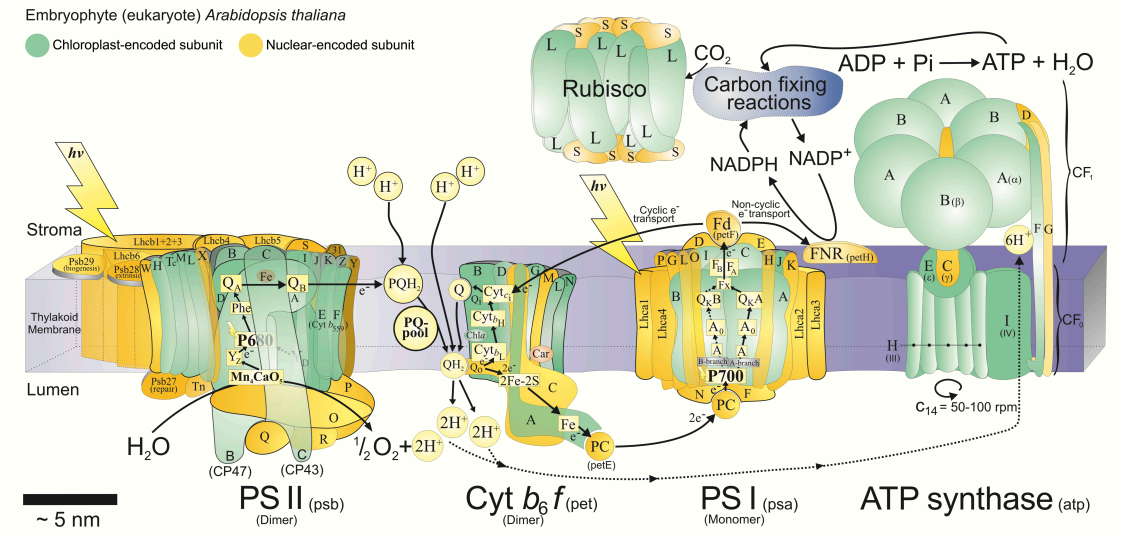


Figure 1.3: Major proteins and protein complexes of the chloroplast photosynthetic apparatus of higher plants exemplified by *Arabidopsis thaliana*. Photosystem II (PSII), cytochrome b_6f (Cyt b_6f), photosystem I (PSI) and ATP synthase. Polypeptide subunits encoded in the chloroplast are coloured green; polypeptide subunits encoded in the nucleus are coloured yellow. Figure from Allen *et al*. (2011).

1.1.3 Light harvesting

The pigments responsible for light harvesting in photosynthetic organisms are organised in light harvesting complexes, consisting of pigment complexes arranged on a protein backbone that tunes the pigments through its interactions. Light harvesting complexes increase the amount of energy available to the cell by enabling the absorption of light of a wider range of wavelengths (Croce and Van Amerongen, 2014). Figure 1.4 shows the absorption spectra of three common pigments. Table 1.1 lists the main pigments involved in photosynthesis.

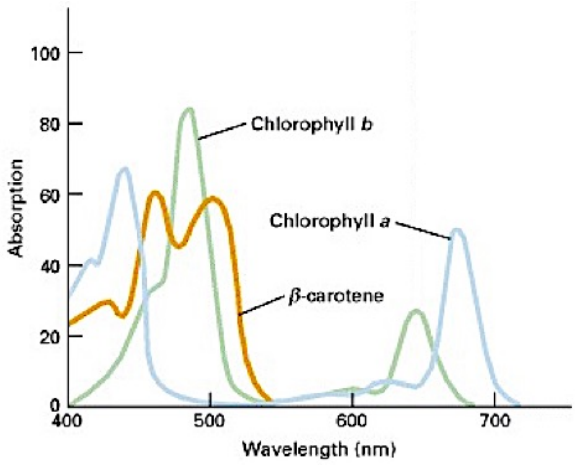


Figure 1.4: The absorption spectra for chlorophyll *a,* chlorophyll *b,* and β-carotene. Each spectrum shows how well light of different wavelengths is absorbed by one of the pigments. Figure adapted from Lodish *et al*. (2000).

Table 1.1: Pigments involved in photosynthesis. Adapted from (Grossman *et al*., 1995).

Pigment family	Pigment	Absorption maxima (nm)	Phylogenetic group
Chlorophyll	Chlorophyll *a*	662, 430	All groups
	Chlorophyll *b*	644, 430	Land plants, Chlorophyta, Euglenophyto, Prochlorophyta
	Chlorophyll *c*	628, 578, 444	Chromophytic algae, Cryptophyta, Pyrrophyta, Prochlorophyta
	β-Carotene	449, 477	All groups
	Lutein	443, 472	Land plants, Rhodophyta
Carotenoids	Zeaxanthin	451, 478	Land plants
	Fucoxanthin	438, 451, 469	Chromophytic algae
	Peridinin	475	Pyrrophyta
Phycobilins	Phycoerythrobilin	535-567	Cyanophyta, Rhodophyta, Crytophyta
	Phycocyanobilin	590-670	Cyanophyta, Rhodophyta, Crytophyta
	Phycourobilin	498	Cyanophyta, Rhodophyta

The chlorophyll pigment family is found in all groups of photosynthetic organisms (Grossman *et al*., 1995). Chlorophylls are cyclic tetrapyrroles, attached non-covalently and easily detachably to the apoproteins LHCI and LHCII (Grossman *et al*., 1995). Chlorophylls are complemented by phycobilins in cyanobacteria and red algae, which allow them to absorb a wider range of light. Phycobilins are linear tertapyrroles that are attached to the apoproteins covalently (Grossman *et al*., 1995). Phycobilins are maintained in phycobilisomes, LHCs located peripherally, attached to the photosystems on the stromal side of the thylakoid membrane. The LHCs also bind carotenoids that help dissipate excess light energy (Horton *et al*., 1996). With the exception of the 'green gap' from approx. 500 to 600 nm, the combination of these pigments cover the whole light spectra (see Figure 1.4) (Croce and Van Amerongen, 2014).

1.1.4 Carbon fixation

The light-independent reactions of photosynthesis (also known as the Calvin–Benson–Bassham (CBB) cycle) are chemical reactions that use the energy generated in the light-dependant reactions, in the form of ATP and NADPH, to fix inorganic atmospheric carbon dioxide (CO_2) to form complex organic carbohydrates (Figure 1.5). Despite the name, this process only occurs when light is available. This process takes place in the stroma of the chloroplast and begins with the carboxylation of Ribulose 1,5-bisphosphate (RuBP), a five carbon compound, into six

carbon intermediary compound which is rapidly degraded into two molecules of 3-Phosphoglycerate (PGA). The carboxylation of RuBP through combination with CO_2 is catalysed by the enzyme RuBisCo, probably the most abundant enzyme on Earth (Feller *et al*., 2008).

ATP and NADPH formed in the light-dependent reactions are used in the conversion of PGA to Glyceraldehyde 3-phosphate (G3P), regenerating $NADP^+$, ADP and inorganic phosphate (P_i) that are shuttled back into the light-dependant reactions. The majority of the G3P that is formed is used to regenerate RuBP for the continuation of the CBB cycle. The remaining ATP, NADPH and G3P are allocated to all the homeostatic processes of the cell. Under low light, more G3P will be allocated to the regeneration of RuBP, when light is not limiting, G3P can be utilised in starch production or exported from the chloroplast (Blankenship, 2008).

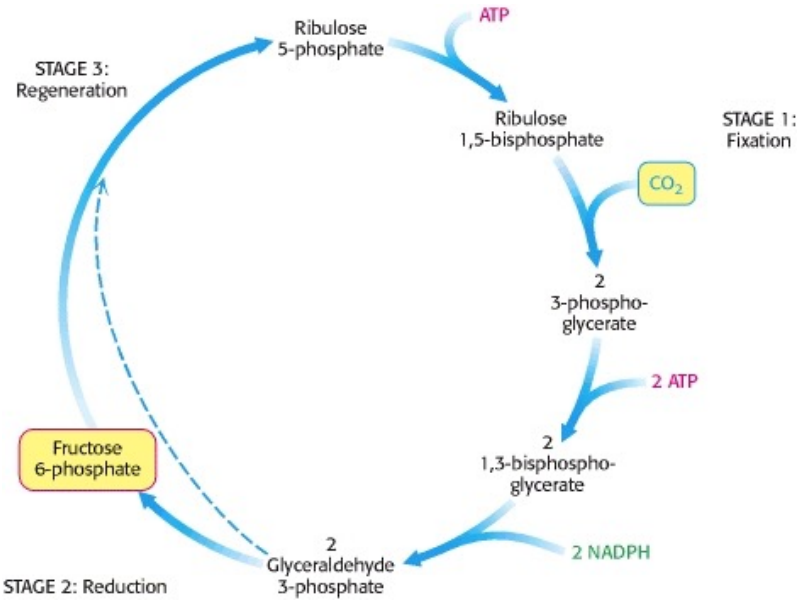


Figure 1.5: The Calvin cycle. Consisting of three stages: Stage 1 is the fixation of carbon by the carboxylation of ribulose 1,5-bisphosphate. Stage 2 is the reduction of the fixed carbon to begin the synthesis of hexose. Stage 3 is the regeneration of the starting compound, ribulose 1,5-bisphosphate. Figure from Sofiyanti *et al*. (2015).

1.1.5 The evolution of photosynthesis

It is agreed amongst biologists that oxygenic photosynthesis first arose in a prokaryote similar to extant cyanobacteria, however the evolutionary mechanism and time of this event are not known for certain (Blankenship, 2008). Photosynthesis is the only known significant source of oxygen, which enables the prediction of a minimum age for cyanobacteria of 2.3 billion years, based on an increase in atmospheric oxygen, in the geological record, to more than 10^{-5} of its present concentration (Bekker *et al*., 2004; Olson, 2006).

Anoxygenic photosynthesis is thought to be the precursor of oxygenic photosynthesis and can be dated back to at least 3.4 billion years ago (Björn and Govindjee, 2015). Bacteria carrying out anoxygenic photosynthesis use protein complexes similar to PSII or PSI but not both complexes in tandem (Bryant and Frigaard, 2006). *Rhodopseudomonas palustris* is a purple photosynthetic bacterium, which possesses only a PSII-like reaction centre that drives cyclic electron transport. *Chlorobium tepidum*, an anaerobic green bacterium, uses only a PSI-like reaction centre to exploit light energy, drawing electrons from hydrogen sulphide for linear electron transport. Phylogenetic analysis cannot account for the distribution of gene sets for both photosystems among different species, leading to the suggestion that genes encoding the photosystems have been transferred through lateral gene transfer (Raymond *et al*., 2003). An alternative explanation is the arrival of both photosystems in a 'protocyanobacterium' - a precursor of cyanobacteria with the ability to switch between a type I-like and type II-like system (Figure 1.6). Indeed, *Oscillatoria limnetica*, a species of cyanobacteria has the ability to switch off PSII in the presence of H_2S and revert from oxygenic to anoxygeneic photosynthesis (Oren and Padan, 1978). The ability of a species to turn off PSI has not yet been observed. *Chloroflexus aurantiacus* has a type-II photosystem in addition to an unusual kind of light-collecting antenna, called a chlorosome, specific to *Chlorobium*, a type-I-containing bacterium. This has led to the suggestion that *Chloroflexus* is a descendant from a protocyanobacterium that lost its type-I photosystem (Allen and Martin, 2007). Comparison of the genomes of several cyanobacteria and anoxygenic photosynthetic bacteria has revealed a core set of genes involved in photosynthesis, suggested to have arisen from a now extinct group of anoxygenic bacteria and since to have spread to other lineages (Mulkidjanian *et al*., 2006). PSI was suggested as the evolutionary precursor of PSII; the cores of the reaction-centres of the photosystems are similar in structure and it is likely that their divergence began with a duplication of this gene cluster (Schubert *et al*., 1998).

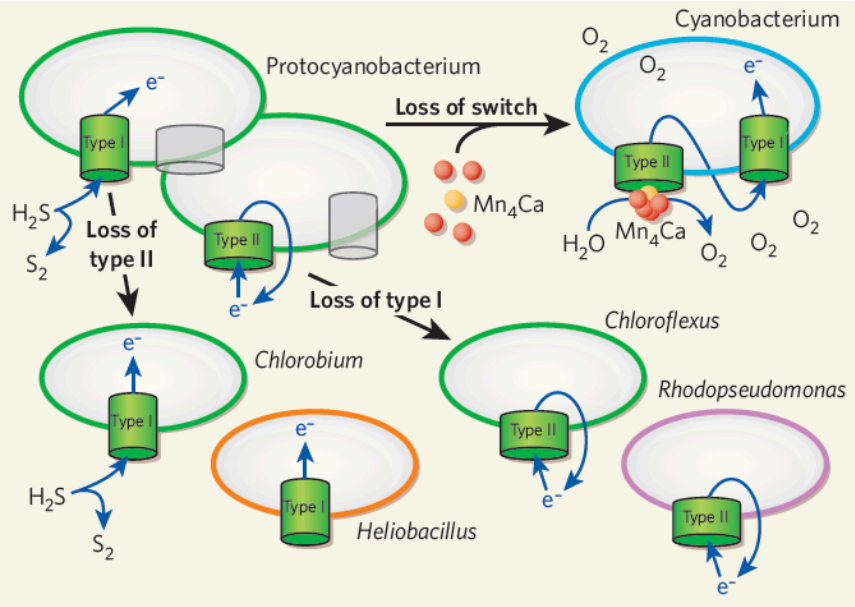


Figure 1.6: Discovery of a modern day 'protocyanobacterium' would be a missing link in the evolution of oxygenic photosynthesis. This species would contain genes for both PSII and PSI with the ability to switch between the two systems. If one of the photosystems were permanently inactive it would be its genes would be lost, giving rise to familiar type I or type II species. The arrival of a catalyst (Mn4Ca) to split O2 and allow for linear electron transport would result in the loss of the switch function and the birth of the first true cyanobacteria. Figure from Allen and Martin (2007).

The structure of the Mn_4Ca cluster of PSII shows a similarity to manganese oxide minerals, which could be an evolutionary indication of the environment, in which water splitting first occurred (Pushkar *et al*., 2008). Over absorption of ultraviolet light by Mn atoms in the water-splitting complex can cause them to dissociate from PSII, a phenomenon known as photoinhibition (Hakala *et al*., 2006). Before the arrival of water splitting, there was no ozone to filter out the sun's ultraviolet light. If an organism possessing both photosystems, but expressing only PSII (i.e. a protocyanobacterium), were introduced into an aquatic, Mn-containing environment, photooxidation of environmental Mn would result in a build up of electrons at PSII, disrupting the electron transport chain. It therefore stands to reason that the protocyanobacterium could have responded by expressing photosystem I, dissipating the build up of electrons from photosystem II, and creating the flow of electrons seen in modern cyanobacteria (Figure 1.6) (Allen and Martin, 2007).

In addition to bacteria, photosynthesis is also present in eukaryotes but not archeans. It is widely believed that the chloroplast of eukaryotes was inherited through an endosymbiosis event with a bacterium, further reinforcing the original evolution of photosynthesis in a bacterial host. The first eukaryotic phototrophs resulted from the incorporation of a cyanobacterial cell via endosymbiosis into a eukaryotic host, already containing a mitochondrion from a previous endosymbiotic event (van der Giezen, 2011). Two primary lineages originated from this early

ancestor. The ‘green’ plastid lineage, dominated by green algae and their descendants, land plants, characterized by the use of chlorophyll b as an accessory pigment. The second, ‘red’ lineage includes the red algae (rhodophytes), which retained pigmentation characteristic of cyanobacteria, and a diverse range of phytoplankton and seaweed whose plastids where derived from rhodophytes. With the exception of red algae, members of this group use chlorophyll c as an accessory pigment. Both groups are more closely related by their respective plastid physiology and photophysiology than the evolution of the host cell (Falkowski *et al*., 2004).

The diverse properties of microalgae, which make them of commercial interest, are represented in their linage as compared to that of modern land plants that descended exclusively from green algae (Figure 1.7).

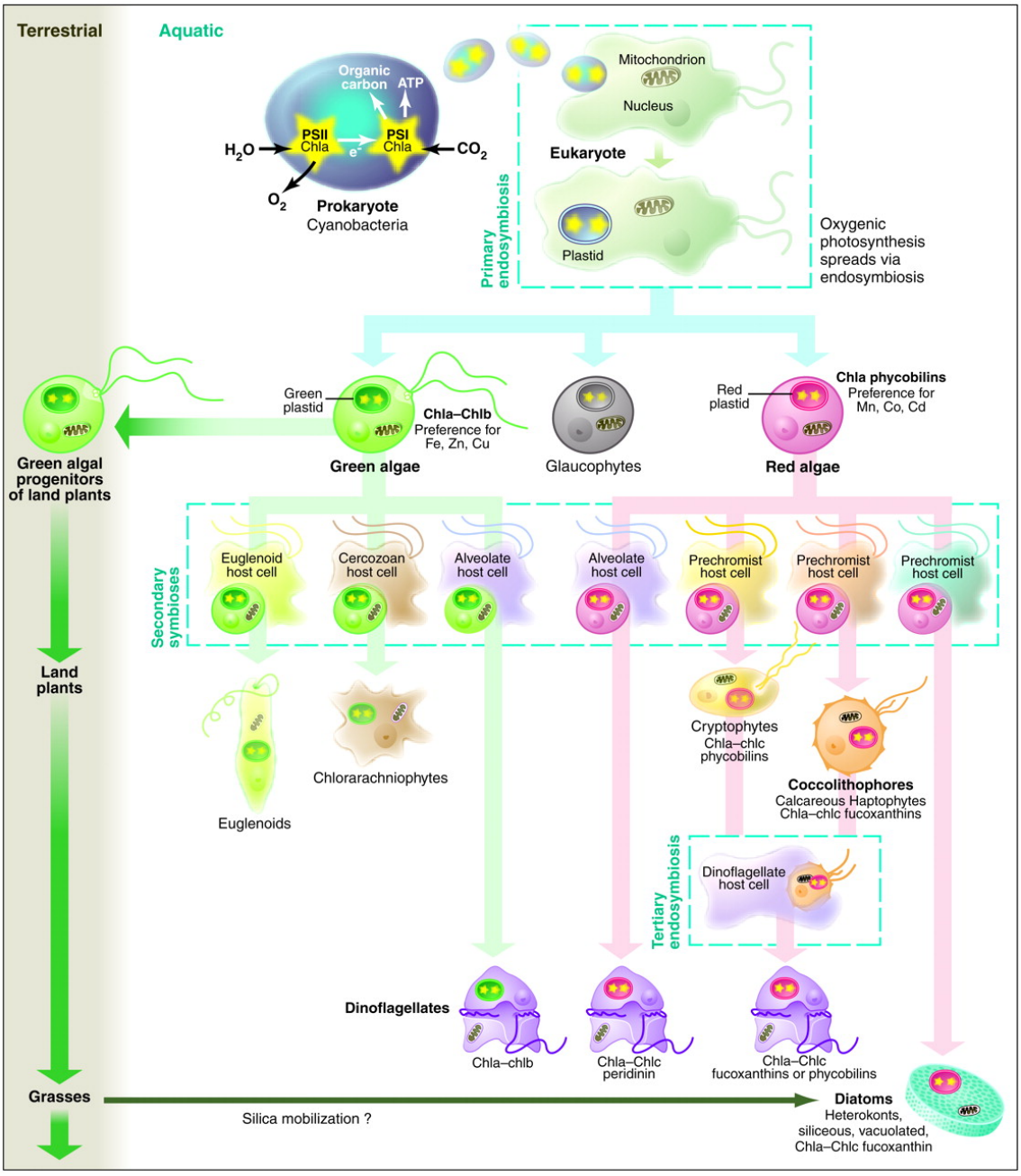


Figure 1.7: Pattern of plastid inheritance in eukaryotic phytoplankton. The original plastid was inherited by endosymbiosis of a cyanobacterium by a eukaryotic host cell to produce the first oxygenic eukaryote and three subsequent lineages, green algae, glaucophytes, and red algae. Further diversification, in addition to secondary and tertiary endosymbiotic events within these groups gave rise to microalgal groups that currently populate the worlds oceans, such as dinoflagellates, diatoms and coccolithophorids. Figure from Falkowski *et al*. (2004).

1.1.6 Aquatic photosynthesis

Despite the fact that the aggregate biomass of aquatic plants amounts to less than 1% of total global photosynthetic biomass, approximately 45% of global photosynthesis occurs in aquatic environments (Imramovsky *et al*., 2011). In almost all aquatic ecosystems, photosynthesis supplies the primary source of organic matter for the growth and metabolic demands of the other organisms in the ecosystem and hence, sets an upper limit on the overall biomass of the ecosystem. Marine microalgae play a crucial role in fixing 45 gigatons of carbon per year, which feeds the ocean's biogeochemical cycles and food-chains (Fenchel, 2008; Imramovsky *et al*., 2011) (see Figure 1.8).

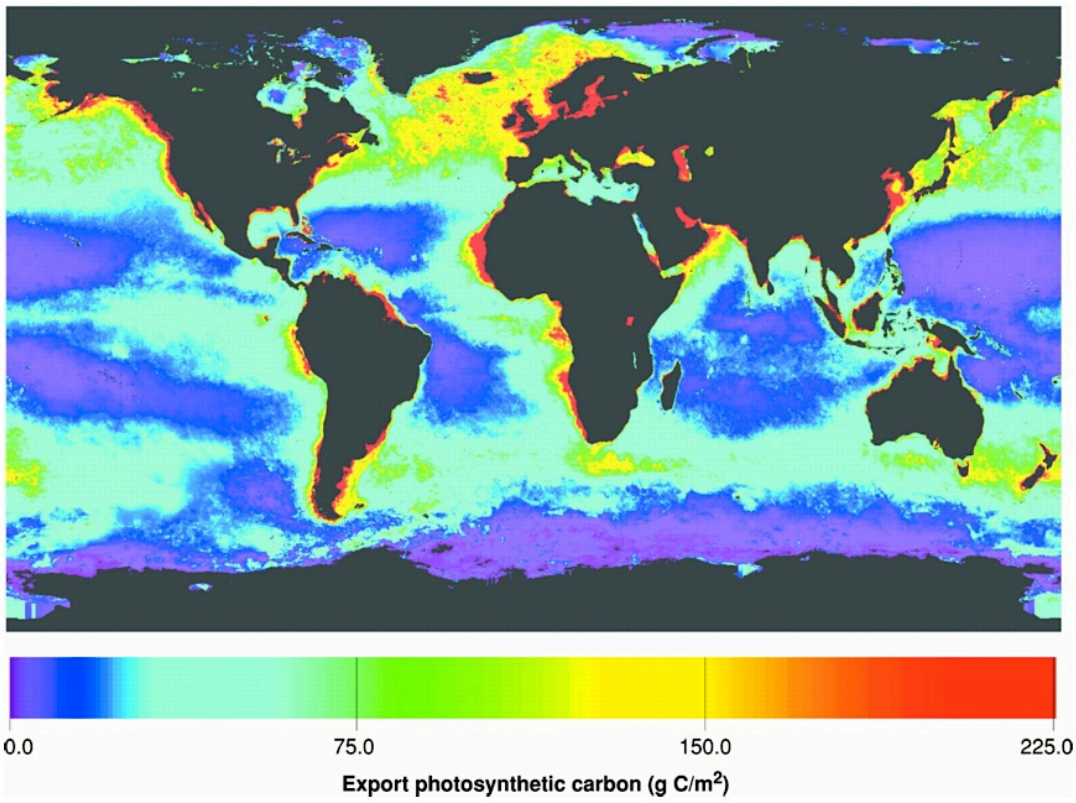


Figure 1.8: Global map of annual mean export of photosynthetic carbon in gC per meter squared. Figure from Falkowski *et al*. (1998).

The Earth's atmosphere was permanently altered over 2 billion years ago by the addition of highly reactive oxygen by aquatic photosynthetic organisms, which subsequently allowed multicellular organisms to evolve (Farquhar *et al*., 2001; Bekker *et al*., 2004). A small portion of the deposited fossilized organic remains of aquatic photosynthetic microorganisms has become the petroleum and natural gas that supports modern civilisation.

Most of the detailed biochemical, biophysical and molecular biological knowledge of photosynthetic process was learned through studies of model algae, such as *Chlamydomonas*,

Chlorella and *Phaeodactylum* (Harris, 2001; Merchant *et al*., 2010; Apt *et al*., 1996; Kaplan *et al*., 1995). The diversity of marine environments is reflected in the sheer diversity of microalgal species, conservatively estimated at 72,500 depending on characterization (Guiry, 2012). Additionally, there is sufficient knowledge available of the basic photosynthetic processes to understand how these species have adapted to their respective environments, these interpretations form the foundation for understanding community structure and global biogeochemical cycles in aquatic environments. A bottleneck in the full realization of the potential of microalgae is the availability of sequenced genomes, which are vital for the characterization of key metabolic pathways, generation of improved cell lines and discovery of useful genetic elements (Finazzi *et al*., 2010). As more genomes have been sequenced it is becoming apparent that the diversity and number of species of microalgae has likely been underestimated.

1.2 Commercial uses of algae

Microalgae have been used by numerous cultures as a source of nutrients for thousands of years. Microalgae first attracted biotechnological interest for the extraction of natural products such as high-value pigments and nutritional supplements several decades ago with the initial commercial culturing of *Spirulina, Chlorella* and *Dunaliella* sp (Borowitzka, 1999). More recently, advanced technologies have enabled the use of microalgae as a manufacturing platform for expression of foreign and native genes due to their quick photoautotrophic growth rates and low maintenance costs compared to land plants, mammalian cells, yeast and bacteria. Many algal species have developed pathways for the expression of commercially valuable products (Priyadarshani and Rath, 2012; Rasala *et al*., 2014). Adaptation of certain species to extreme conditions, for example the adaptation of *Dunaliella* for growth at high salinity, provides a useful means of reducing the risk of contamination in large-scale growth systems (Ben-Amotz and Avron, 1972).

High-value chemicals that can be extracted from cyanobacteria and eukaryotic algae include compounds used in the food industry, medicine and cosmetics (see Table 1.2). The advancement of algae for production of biofuel is currently under examination. Some of the main successful exploitations of algal biomass are food additives such as β-carotene, astaxanthine, and long-chain polyunsaturated fatty acids (PUFAs) (Varela *et al*., 2015; Borowitzka, 2013; Sharon-Gojman *et al*., 2015). Cell-wall sulphated polysaccharides are another line of exploitation, mainly for cosmetics (Arad and Levy-Ontman, 2010). Algae can also be used as components in novel biosensors for environmental application, such as sensitive monitoring of pollutants (Viji *et al*., 2014; Diaz *et al*., 2014). Additionally, algae have been used extensively for ecological

reef rehabilitation and bioremediation (Hagedorn *et al*., 2015; Vidyashankar and Ravishankar, 2016).

Table 1.2: Oil content and storage compounds of industrially relevant species. Adapted from (Vinayak *et al*., 2015).

Phylum	**Species**	**Oil content (% d.w.)**	**High value compounds**
Chlorophyta	*Tetraselmis suecica*	15-32	Carotenoids, chlorophyll, tocopherol, lipids
Chlorophyta	*Ankistrodesmus sp.*	28-40	Mycosporine-like amino acids, polysaccharides
Chlorophyta	*Dunaliella salina*	10	Carotenoid, β carotene, mycosporine-like amino acids, sporopollenin
Chlorophyta	*Dunaleilla teriolecta*	36-42	Carotenoid, β carotene, mycosporine-like amino acids
Chlorophyta	*Neochloris oleoabundans*	35-65	Fatty acids, starch
Chlorophyta	*Botryococcus braunii*	29-75	Isobotryococcene, botryococcene, triterpenes
Chlorophyta	*Chlorella vulgaris*	58	Neutral lipids
Chlorophyta	*Chlorella emersonii*	34	Neutral lipids
Chlorophyta	*Chlorella protothecoides*	15-55	Eicosapentaenoic acid (EPA), ascorbic acid
Chlorophyta	*Chlorella minutissima*	57	C16- and C18-lipids
Heterokontophyta	*Nitzschia laevi*	28-69	EPA
Heterokontophyta	Thalassiosira	21-31	Glycosylglycerides, neutral lipids, TAG
Heterokontophyta	*Schizochytrium limacinum*	50-77	Docosahexaenoic acid (DDHA)
Myzozoa	*Crypthecodinium cohni*	20	DDHA, Starch
Ochrophyta	*Cyclotella sp.*	42	Neutral lipids
Ochrophyta	*Nannochlropsis sp.*	46-68	EPA, TAG, omega-3, LC-PUFA

Microalgae are arguably the most diverse group of organisms on the planet, providing a large platform for selection of candidate strains for specific applications. The diversity of microalgae is also present in the storage compounds used by different groups. Green algae, dinoflagellates

and diatoms are high oil content algae, typically in the form of treacly glycerols (TAGs) (Hu *et al*., 2008), which can be converted into biofuel through the process of esterification (Schenk *et al*., 2008).

Commercial cultivation systems for microalgae consist of open systems such as raceways or shallow ponds, or closed systems in the form of photobioreactors (PBRs) (see Figure 1.9). The former are used for biofuel feedstock cultivation and the latter typically for cultivation of algae for production of high-value products or as a platform to generate the biomass required to inoculate a large-scale open system. Closed systems consist of a one of a variety of PBRs coupled with an associated light-generating component to optimize the availability of solar radiation. Raceways operate continuously with the addition of fresh culture and nutrients in close proximity to a paddle wheel, which circulates the system, and removal of biomass for harvest before the paddlewheel. While both systems are similar in that they incorporate an inlet and outlet, there are advantages and disadvantages associated with each system. Closed systems are more expensive to maintain but provide a more controllable internal environment, whereas open systems, essential for biofuel production, are considerably cheaper to run but are vulnerable to contamination by foreign microorganisms (Demirbas, 2011).

Microalgae have several inherent advantageous traits that make them suitable for commercial exploitation; namely, fast growth rates, the ability to grow on sunlight, the ability to be grown on non-arable land, the ability to sequester CO_2 from the environment during growth, and in eukaryotic systems, the ability to perform the necessary post-translational modifications of heterologously expressed eukaryotic proteins (Hlavova *et al*., 2015; Stephenson *et al*., 2011; Doron *et al*., 2016).

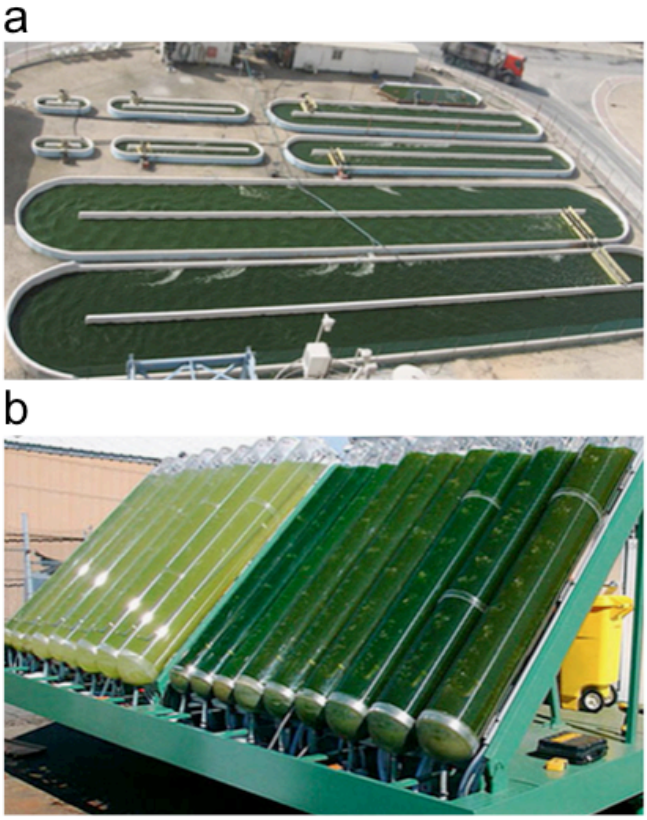


Figure 1.9: Examples of both open and closed algal cultivation systems. **(a)** Algal raceway ponds of increasing size (open system). **(b)** Inclined tubular bioreactors (closed system). Figure adapted from Bitog *et al*. (2011).

1.2.1 Algal biofuels

The recognition of diminishing global oil supplies, combined with the threat of global warming due to anthropogenic CO_2 emissions has generated massive interest in the development of renewable energy. The use of miroalgal biomass for biofuel has long been a focus of efforts, however the relatively high cost and low yield are of concern; biotechnological improvements and global market changes may, however, have major implications for the viability of microalgae as a biofuel source. Microalgae are the third generation of material used for biofuel production after edible feedstock (first generation) and biofuels derived from organic waste products and land grown energy crops (second generation), which compete with food crops for arable land (Moore, 2008). The conversion rate of sunlight into high-energy compounds is often an order of magnitude higher in microalgae relative to first generation biofuels (Melis, 2009). Another major advantage of algal derived biofuels is the ability to accumulate biomass rapidly, with a biomass turnover rate of approximately one week on average for the worlds oceans (Imramovsky *et al*., 2011).

The major challenge associated with efforts to develop a viable algae-to-biofuels process is the ability to produce such algal biofuels at the scale required and to do so economically (Starckx and Senne, 2012; Stephenson *et al*., 2010). It has been calculated that in order to produce 100,000 barrels of algae oil per year (equivalent to approximately 10% of Australia's daily requirement), ~650 ha of pond and almost 4 GL of water per year, assuming an aerial annual average biomass productivity of 20 g dry weight m^{-2} d^{-1}, a 30% total lipid content and a 100% conversion efficiency of the lipid to biodiesel (which is very unlikely to be achieved) (Borowitzka and Moheimani, 2013; Fon Sing *et al*., 2013). High lipid productivity is the first step in producing biofuels from algae (Griffiths and Harrison, 2009) but in order to achieve an economic process, all steps in the production process from algal culture to fuel production must be efficient at scale. Thus, the establishment of a viable algae-to-biofuel production system is not solely a biological or engineering challenge but a combination of both.

1.2.1.1 Bacterial communities in industrial algal production systems

Obtaining robust growth of algae in industrial production systems is also made difficult by the presence of complex bacterial communities that can either promoter or reduce growth rates (Fulbright *et al*., 2018). Even in closed bioreactors that are designed to reduced the risk of bacterial contamination, bacteria are known to enter and proliferate within the system (Fulbright *et al*., 2018). Little is known about these bacterial communities and characterization of these communities is recently emerging area of interest (Wang *et al*., 2016; Fuentes *et al*., 2016; Fulbright *et al*., 2018). A study of the bacterial communities within 275 *Nannochloropsis salina* biofuel feedstock culture samples during the scale up process has shown that larger cultures harbour richer bacterial communities, and further, that blooms of a specific bacteria correlated with poor growth rates of *N. salina* (Fulbright *et al*., 2018). 16 bacterial OTUs (operational taxonomic units) were found in 90 % of the *N. salina* cultures including one OTU that was present in all of the cultures (Fulbright *et al*., 2018). Gaining a better understanding of these complex ecosystems is of critical importance to the optimization of algal growth rates within large-scale industrial systems.

1.2.2 Improving algae for biotechnology

The quantum yield of photosynthesis sets a limit on the amount of biomass that can be accumulated by microalgae and ultimately converted to useful products, such as biofuel, and has been identified as the limiting factor for large-scale production systems, both for biofuel and specific compounds. The maximum efficiency of photosynthesis is 8-10% (Hambourger *et al*.,

2009). The factors that contribute to this low efficiency have spawned corresponding areas of research seeking to increase efficiencies towards the theoretical maximum (see Figure 1.10).

1.2.2.1 Engineering to improve light collection

Of the light that is absorbed by the antenna pigments, only the equivalent of the low-energy (red) wavelengths are used by PSI (680 nm) and PSII (700 nm). The remaining energy within the higher-energy energy (blue) wavelengths is reemitted as heat and fluorescence, a process known as non-photochemical quenching (NPQ) (Lambrev *et al*., 2012). Microalgae grown in mass culture can have as much as a 95% reduction in energy conversion efficiency due to a phenomena known as mass culture self-shading whereby cells near the surface of the culture receive too much light and are susceptible to photoinhibition caused by non-photochemical quenching, inhibiting the activities of PSII, and to a lesser extent PSI, while cells lower down in the culture are light limited due to poor light penetration (Polle *et al*., 2003; Mitra and Melis, 2008; Zavafer *et al*., 2015; Melis, 2009). Truncation of the light-harvesting antenna has been achieved through genetic modification of *Chlamydomonas reinhardtii* using both RNAi technology and genetic engineering, resulting in improvements to photosynthetic efficiency in excess of 50% (Masuda *et al*., 2003; Mussgnug *et al*., 2007; Polle *et al*., 2003) (see Figure 1.10b). However, these improvements have only been demonstrated under a narrow range of growth conditions (high light, high cell density, low CO_2 concentration), with lower productivities reported under other conditions (Page *et al*., 2012; Kirst *et al*., 2014; Lea-Smith *et al*., 2014). Therefore, the improvement gained through antennae truncation should be considered a slight move towards the theoretical maximum, which is likely cancelled by the requirement for specific growth conditions (Cotton *et al*., 2015).

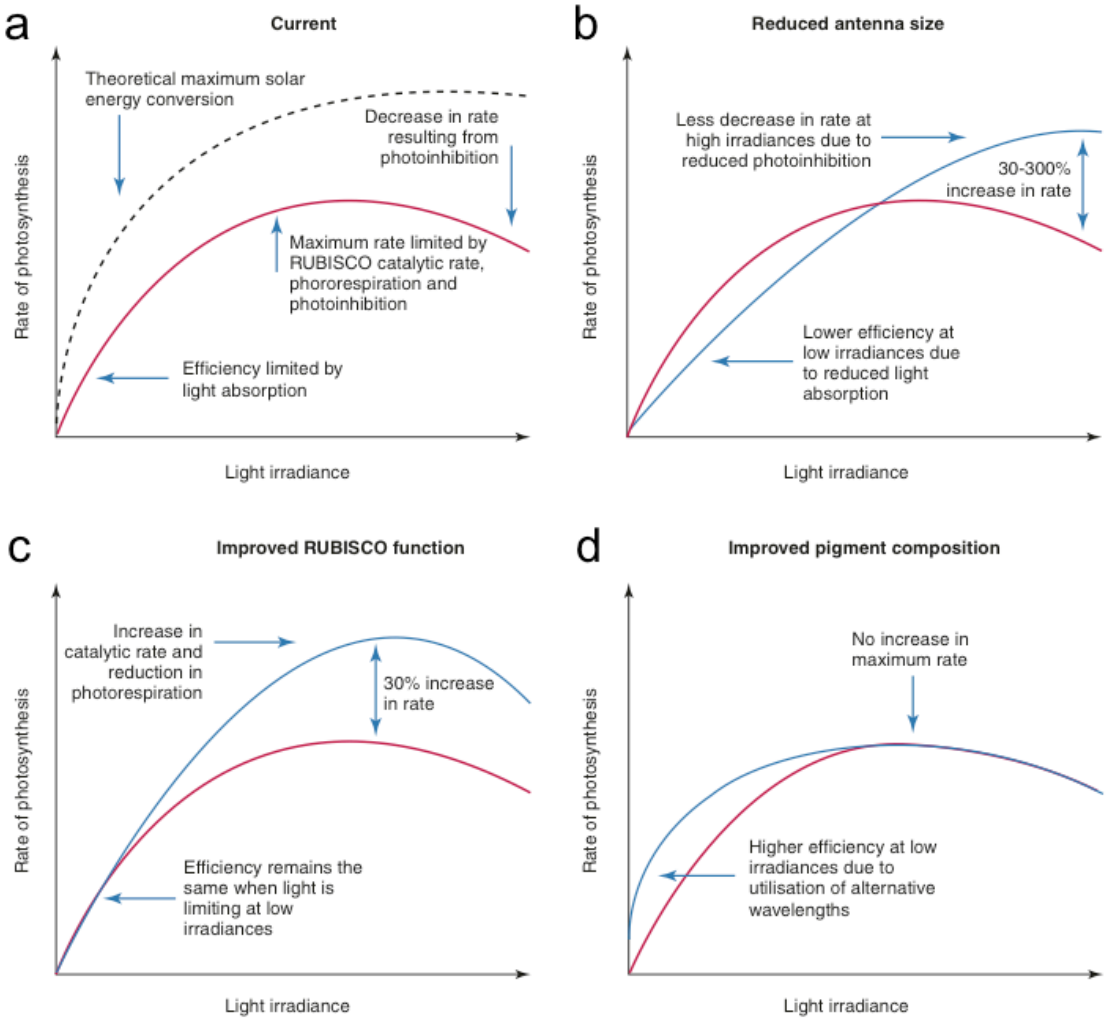


Figure 1.10: Photosynthesis irradiance curves (P/I), illustrating targets for improved efficiency. P/I curves describe the relationship between solar energy input and the amount of photosynthetic productivity measured by oxygen evolution or carbon fixation. **(a)** There is an initial linear relationship between energy input and photosynthetic output until photosynthetic saturation is reached (Pmax). At higher energy inputs, excess energy results in damage to the photosynthetic machinery and photoinhibition, reducing Pmax. Red line represents current photosynthetic organisms; dashed line represents the theoretical maximum. **(b)** Reducing the size of the light-harvesting antennae associated with the photosystems alleviates mass culture self-shading and photoinhibition. **(c)** Effect of increasing the specificity of RuBisCo for CO_2. **(d)** The effect of increasing the spectral range and efficiency of absorption of incident light. Figure from Stephenson *et al*. (2011).

1.2.2.2 Engineering to improve carbon fixation

The enzyme RuBisCo, which catalyses the carboxylation of ribulose-1,5-bisphosphate (RuBP), the rate limiting step of the Calvin cycle, is slow and inefficient (Savir *et al*., 2010). RuBisCo has a low affinity for CO_2 and can also catalyse the oxygenation of its substrate (Whitney *et al*., 2011). This results in another inefficiency owing to the consumption of oxygen, ATP and NADPH and release of fixed carbon in a competing reaction - photorespiration. Another branch of research focuses on increasing the efficiency of RuBisCo. Overexpression of RuBisCo as well as increasing the local concentration of CO_2 at the RuBisCo catalytic site are advantageous under carbon-limiting conditions and can increase the maximum rate of solar energy conversion (see Figure 1.10c). Tobacco cell lines with functional RuBisCo from the cyanobacterium *Synechococcus elongates* have been generated (Lin *et al*., 2014). Transformed lines were photosynthetically competent and their respective rates of CO_2 fixation per unit of enzyme were higher than the tobacco control (Lin *et al*., 2014). Directed evolution of a cyanobacterial RuBisCo produced a mutant with 2.9-fold improvement in activity with only a 9% loss in CO_2/O_2 specificity. However, *Synechocystis* sp. PCC 6803 expressing the improved Rubisco did not show growth improvements, instead producing 25% less Rubisco (Durão *et al*., 2015), consistent with the finding that altering RuBisCo activity has little effect on the rate of photosynthesis in the same species (Marcus *et al*., 2011). This may be due to carbon-concentrating mechanisms in cyanobacteria, which protect against high levels of O_2 around RuBisCo and reduce photorespiration (Dou *et al*., 2008; Mangan and Brenner, 2014). For decades it has not been possible to express RuBisCo in an easily manipulatable host; however, Aigner *et al*. recently succeeded in expressing a functional plant RuBisCo in *Escherichia coli,* which may enable mutational analysis of RuBisCo and production of improved variants (Aigner *et al*., 2017). Bypassing RuBisCo and using alternative aerobic carbon fixation cycles, such as the 3-hydroxypropionate bi-cycle (Zarzycki *et al*., 2009), and even newly designed pathways (Bar-Even *et al*., 2010; Erb, 2011), is being attempted to improve carbon fixation.

1.2.2.3 Engineering to improve photochemistry

Only 48.7% of the solar energy that reaches the photosystems is between the usable range of 400 – 700 nm, termed photosythetically active radiation (PAR) (Kruse *et al*., 2005). Increasing the quantity and variety of pigments will broaden this range and increase the amount of available solar energy for algal strains (Blankenship *et al*., 2011) (see Figure 1.10d). Additional pigments absorbing outside the previously defined PAR range have been discovered in extant biological systems, e.g. chlorophyll *f* (706 nm) and chlorophyll *d* (710 nm) (Chen *et al*., 2010; Gan *et al*., 2014). Chlorophyll *d* is known to be involved in primary charge separation in both

photosystem I and II in *Acaryochloris marina* (Renger and Schlodder, 2008), it is not yet known if chlorophyll *f* plays a role in photochemistry at the reaction centre. The energy of the 710 nm photon absorbed by chlorophyll *d* is 70 meV, less than that in the 680nm photon (1.82 eV) absorbed by chlorophyll *a*. Due to the strict energy requirements for water oxidation, this may result in an increased chance of the occurrence of back-reactions leading to damage caused by singlet oxygen production, particularly under variable light conditions (Rutherford *et al*., 2012). Broadening of the photosynthetic spectrum may be possible but would likely require costly tightly controlled growth conditions.

1.2.2.4 Source/sink optimization

In contrast to the optimization of individual components of photosynthesis, a promising alternative could be the optimization of source (light energy) versus sink (metabolic capacity); taking advantage of the inherent link between kinetics of the light and dark reactions (Paul and Foyer, 2001; Cotton *et al*., 2015). This has been illustrated by the heterologous expression of the proton/sucrose symporter CscB in *Synechococcus elongates*, resulting in the export of up to 36.1 mg l^{-1} h^{-1} of sucrose. The export of the photosynthetic sink (sucrose) helps to maintain an oxidized electron transport chain and minimize photodamage and also yielded increased PSII activity, carbon fixation, and chlorophyll content under optimized conditions in these strains (Ducat *et al*., 2012).

In summary, most of the suggested modifications to photosynthetic components would only be beneficial under a narrow range of culture conditions, which could be provided by photobioreactors but with a significant associated cost. “Sink-maximized” strains should be used for engineering of photosynthesis in order to avoid metabolic congestion and mismatches between dark and light reactions (Cotton *et al*., 2015). It may also be more beneficial to use this technology to produce complex, high-value chemicals rather than low-value products such as biomass and ethanol.

1.3 Genetic engineering in algal biotechnology

Despite the vast number of microalgal species, methods for genetic modification have only been developed in approximately 25 species, most of these initially chosen for their ease of culture as opposed to suitability for commercial application (Radakovits *et al*., 2010; Doron *et al*., 2016). Significant efforts have been invested in the establishment of tools that will allow realization of the potential of microalgae for biofuel and high-value product production, most having been originally developed in the model species *Chlamydomonas reinhartii* (Kindle *et al*., 1989). The

expression of transgenes in *Chlamydomonas* was based on identification and accumulation of regulatory elements such as promoters and untranslated regions (UTRs) (Stern *et al*., 2008). The techniques for genetic engineering can be broadly divided into forward and reverse genetic approaches (Table 1.3). Reverse genetic approaches use variety of techniques to target changes to a known endogenous gene and rely on transformation systems for the integration of transgenes into the host alga. Forward genetic approaches typically begin with the generation of a library of mutated strains, either by transformation or through the use of mutagens, followed by phenotypic screening to identify a strain of interest.

Table 1.3: Comparison of forward and reverse genetic approaches. Table adapted from Hlavova *et al*. (2015)**.**

Forward genetics	**Reverse genetics**
Targeted on desired feature, phenotype is selected for, causative gene primarily unknown	Targeted on modification of a known gene, phenotype primarily unknown, desired feature predicted not guaranteed
Results in mutants, not GMO	Results in (mostly) GMO
No prior information required	Requires prior information
Available in any organism	Available in selected organisms only

1.3.1 Transformation systems

Stable genetic transformation has been reported for the green (*Chlorophyta*), red (*Rhodophyta*), and brown (*Phaeophyta*) algae; diatoms; euglenids; and dinoflagellates. Over 30 strains of microalgae have been transformed to date (Doron *et al*., 2016). Successful transformation of microalgae was first developed in the chloroplast of *C. reinhardti*, however there are still fewer transformation systems established which target the chloroplast as compared to the nucleus (Shimogawara *et al*., 1998). One major advantage of plastid transformation is the ability to target integration of transgenes using homologous recombination, whereas nuclear transformation of microalgae usually results in random integration events (Doron *et al*., 2016). Chloroplast transformation methods for microalgae have since been optimized further (Purton, 2007; Purton *et al*., 2012).

Nuclear transformation in microalgae has several advantages, including the targeting of heterologous proteins for expression in organelles, such as the chloroplast, the ability to perform glycosylations and post-translational modifications of proteins, as well as the ability to target proteins for secretion (León-Bañares *et al*., 2004). These advantages are of significant relevance to the exploitation of microalgae for the industrial production of recombinant proteins, especially considering the difficulty associated with obtaining efficient expression of foreign

genes in microalgae (Eichler-Stahlberg *et al*., 2009). Several methods for gene delivery to the algal nucleus are available.

1.3.1.1 Methods of gene delivery

1.3.1.1.1 Electroporation

The application of an electrical field can be used to increase the permeability of cell membranes allowing the introduction of DNA to the cell. This technique was first used to transform cell wall-deficient mutants (protoplasts) and other algae with thin cell walls. It has been used to transform *C. reinhardtii* (Brown *et al*., 1991; Shimogawara *et al*., 1998; Yamano *et al*., 2013), *D. salina* (Sun *et al*., 2005)*, Chlorella vulgaris*, *Scenedesmus obliquus,* and *Nannochloropsis* sp (Radakovits *et al*., 2012; Kilian *et al*., 2011). The use of protoplasts has been shown to improve the efficiency of transformation in *Lobosphaera* (Zorin *et al*., 2014). Additionally, growth in silica deficient media can reduce cell wall thickness and has been utilized for transformation of *P. tricornutum* via electroporation (Niu *et al*., 2012). A recent study involving a high-throughput genetic screen, mapping 11,478 insert sites in electroporation-transformed *C. reinhardtii,* indicated that the site of integration of plasmid DNA within the genome occurs in an almost totally random manor, enabling the generation of mutant strains of nearly all genes (Zhang *et al*., 2014). The model of integration purposed proceeded via partial digestion of both genomic and plasmid DNA by sequence-specific endonucleases during transformation, facilitating the integration of plasmid DNA into double-stranded breaks; however, this can result in ligation of multiple DNA fragments into one insert site in rare cases (Zhang *et al*., 2014).

1.3.1.1.2 Biolistics

Biolistics, also known as micro-particle bombardment, is a robust and frequently used method of transformation. Biolistics uses gold or tungsten particles, coated in DNA, which are propelled into algal cells through the cell wall barrier at high-velocity via a biolistics particle delivery system or gene gun. *C. reinhardtii* (Kindle *et al*., 1989)*, Dunaliella salina* (Shabelnik *et al*., 2011)*, Haematoccucs pluvialis* (Steinbrenner and Sandmann, 2006)*,* and several diatoms (Dunahay *et al*., 1995; Apt *et al*., 1996; Falciatore *et al*., 1999; Zaslavskaia *et al*., 2000) have been successfully transformed via biolistics.

1.3.1.1.3 Glass beads method

An inexpensive and cheap method of transformation used for marine microalgae is agitation in the presence of glass beads, polyethylene glycol (PEG) and DNA. This method is often used with protoplasts or thin cell walled species. Transformation using the glass bead method has been established in *C. reinhardtii* (Kindle, 1998), *D. salina* (Feng *et al*., 2009), and *Chlorella ellipsodea* (Jarvis and Brown, 1991). A comparison of transformation efficiency in *D. salina* using glass beads, electroporation and biolistics found that transformation using the glass bead methods was most efficient (Feng *et al*., 2009). Transformation efficiencies can be increased through the use of linearized DNA as opposed to supercoiled DNA (Kindle, 1998).

1.3.1.1.4 *Agrobacterium*-mediated transformation

Agrobacterium tumefaciens is widely used for nuclear transformation of land plants and has been modified for use with microalga, using the pCAMBIA vector, where it has again proven to be widely applicable; several species of microalgae have been transformed via *Agrobacterium* including *Chlamydomonas* (Kumar *et al*., 2004; Pratheesh *et al*., 2014) in addition to other algae of biotechnological value, such as *Schizochytrium* (Cheng *et al*., 2012), *Haematococcus pluvialis* (Kathiresan *et al*. 2015), *Isochrysis* sp. (Prasad *et al*., 2014), and *Dunaliella bardawil* (Anila *et al*., 2011). Furthermore, these transformations most often resulted in the generation of single or low copy number transformants, which is important from a regulatory perspective and in order to minimize gene silencing.

Agrobacterium-based vectors utilize a natural bacterial system to introduce DNA into the nuclear genome of eukaryotic microalgae. The pathogenic process involves the formation of a tumour, the genes for which are encoded by the bacterium, at or close to the infection site which provides the bacterium with carbon and nitrogen in the form of opines (Zupan *et al*., 2000). *Agrobacterium* transfers a portion of its DNA (T-DNA) into the nuclear genome of the host via the tumour-inducing (Ti) plasmid, containing the T-DNA delimited by 25 bp repeats (known as the right and left borders) and ~35 virulence (*vir*) genes (Hellens *et al*., 2000). *Agrobacterium*-mediated transformations use 'dis-armed' *Agrobacterium* strains in which the tumour inducing genes within the T-DNA of the Ti plasmid have been removed (Hellens *et al*., 2000).

1.3.1.2 Nuclear control elements and promoters

Transcription of a particular gene is initiated by a promoter sequence which resides upstream of the transcriptional start site of the gene. Strong promoters are used to regulate efficient

expression of foreign genes; typically these are derived from viruses such as the 35S promoter of the Cauliflower Mosaic Virus (Odell *et al*., 1985). However, the use of the heterologous promoters is not always effective, for example, the promoter of the Cauliflower Mosaic Virus (CaMV35S) is not suitable for transformation of *C. ellipsoidea* and *C. reinhardtii* (Kim *et al*., 2002; Day *et al*., 1990; Blankenship and Kindle, 1992; Kumar *et al*., 2004). While heterologous promoters can be applicable, endogenous promoters of abundantly expressed genes are recommended to robustly express foreign genes in algae (Díaz-Santos *et al*., 2013). For example, the endogenous *RBCS2* promoter of a small subunit of RuBisCO has been used to efficiently express transgenes in *C. reinhardtii* (Stevens *et al*., 1996). Fusion of the *RBCS2* promoter to the *HSP70A* (heat shock protein 70A) promoter, which acts as a transcriptional enhancer, further increases the efficiency of expression, thus the *HSP70A/RBCS2* regulatory system is highly recommended for *C. reinhartii* (Schroda *et al*., 2000).

Other endogenous promoters of highly expressed genes that have been used to transform other species of algae include, the fucoxanthin-chlorophyll binding protein promoter (Falciatore *et al*., 1999), the constitutive promoter of the Glyceraldehude-3-Phosphate Dehydrogenase (GAPDH) gene (Jia *et al*., 2012), and the β-tubulin promoter (Radakovits *et al*., 2012). A "promoter trapping approach" has recently been applied in *C. reinhartii* to identify efficient endogenous promoters, capable of driving expression of a promoter-less selectable marker and generating antibiotic-resistant colonies (von der Heyde *et al*., 2015). This approach should facilitate the discovery of efficient promoters in the growing number of algae with sequenced genomes.

Transfer RNAs (tRNAs) are adaptor molecules composed of RNA that serve as the physical link between mRNA and the amino acid sequence of proteins. When tRNAs that are compatible with the codon usage of a given transgene are absent, a dramatic negative effect is observed on the efficiency of its expression (Heitzer *et al*., 2007). Differences in codon usage exist between species of algae (Leon-Banares *et al* 2004), hence, codon optimization of non-native genes can significantly increase protein expression (Zaslavskaia *et al*., 2000; Franklin *et al*., 2002).

In certain cases, the use of constitutive promoters is not always the most optimal approach for transgene expression in algae, for example, for the purpose of expression of a toxic compound for industrial or pharmaceutical use or for a silencing expression system such as RNAi (see section 1.3.2.3). In these cases, the use of an inducible system offers major advantages. Additionally, inducible expression systems may prove more efficient for large-scale production of heterologous proteins; allowing a culture to conserve energy and reach an optimal cell concentration before the onset of expression. Some algae can obtain nitrogen through conversion of nitrate to ammonium, a reaction catalysed by nitrate reductase (Fernández *et al*., 1989; Berges, 1997). The promoter of the nitrate reductase gene, which is switched on and off in the presence of nitrate or ammonium ions, has been used to regulate inducible expression of a

reporter gene in several species of algae including, *C. reinhardtii* (Schmollinger *et al*., 2010), *C. ellipsoidea* (Wang *et al*., 2004), *D. salina* (Li *et al*., 2008, 2007) and *P. triornutum* (Niu *et al*., 2012). The use of light-induced promoters (Blankenship and Kindle, 1992; Park *et al*., 2013; Baek *et al*., 2016), metal-responsive promoters (Quinn, 1995; Ferrante *et al*., 2011) and vitamin B_{12}-responsive (Helliwell *et al*., 2014) promoters are also being explored in algae. Another system for inducible expression is riboswitch biology, through a mechanism of alternative splicing of mRNA mediated by a regulatory element of the *THIC* gene, involved in the biosynthesis of thiamine pyrophosphate (TPP) (Ramundo *et al*., 2013).

1.3.1.3 Selection marker systems for nuclear transformation

Stable transformation of microalgae is dependent on the use of a proper selection marker. These include genes conferring resistance to antibiotics or herbicides, as well as metabolic enzymes that control growth under specific nutritional conditions. Table 1.4 lists selection systems for nuclear transformation, normally carried out on nutrient-rich agar post-transformation.

Table 1.4: Selection markers and selection modes for nuclear transformation. Adapted from (Doron *et al*., 2016).

Type of selection marker	**Selection gene**	**Gene product**	**Selection mode**	**Species/Genetic background**
Antibiotic resistance	*Ble*	Phleomycin-binding protein	Reistance to Zeocin/Phlomycin	*Chlamydomonas reinhardtii* *Dunaliella salina* *Volvox carteri* *Chlorella ellipsoidea* *Phaeodactylum tricornutum* *Cylindrotheca fusiformis* *Nannochloropsis* sp. *Nannochloropsis granulate/gaditana/oculata/ oceanica/salina*
	aphVIII	Aminoglycoside 3′-phosphotransferase	Resistance to Paromomycin	*Chlamydomonas reinhardtii* *Gonium pectoral* *Eudorina elegans*
	aadA	Aminoglycoside 3′-adenylyltransferase	Resistance to Spectinomycin/ Streptomycin	*Chlamydomonas reinhardtii*
	aph7	Aminoglycoside phosphotransferase	Resistance to Hygromycin	*Chlamydomonas reinhardtii* *Haematococcus pluvialis* *Volvox carteri* *Chlorella vulgaris* *Laminaria japonica*
	nptII	Neomycin phosphotransferase	Resistance to Neomycin	*Chlamydomonas reinhardtii* *Chlorella sorokiniana,* *Chlorella vulgaris* *Amphidinium* sp.*, and Symbiodinium microadriaticum* *Cyclotella cryptica,*

Table 1.4 continued

Type of selection marker	Selection gene	Gene product	Selection mode	Species/Genetic background
				Navicula saprophila *Phaeodactylum tricornutum*
	cat	Chloramphenicol acetyltransferase	Resistance to Chloramphenicol	*Dunaliella salina* *Chlorella vulgaris*
	CRY1-1	Cytosolic ribosomal protein S14	Resistance to Emetine	*Chlamydomonas reinhardtii*
Herbicide resistance	*GAT*	Glyphosate aminotransferase	Resistance to Glyphosate	*Chlamydomonas reinhardtii*
	ALS	Acetolacetate synthase	Resistance to Sulfometuron methyl	*Chlamydomonas reinhardtii* *Porphyridium* sp. *Parietochloris incisa*
	PDS1	Phytoene desaturase	Resistance to Norflurazo	*Haematoccucs pluvialis* *Chlorella zofingiensi*
Metabolic markers	*NIT1*	Nitrate reductase	Growth in the presence of nitrate salt	*Chlamydomonas reinhardtii (nit1-)* *Volvox carteri (nit1-)* *Dunaliella viridis (nit1-)* *Chlorella sorokiniana (nit1-)* *Chlorella ellipsoidea (nit1-)*
	ARG7	Argininosuccinate lyase	Growth in Arginine free media	*Chlamydomonas reinhardtii (arg7-)*
	NIC7	Quinolinate synthetase	Growth in Nicotinamide free media	*Chlamydomonas reinhardtii (nic7-)*
	OEE1	Oxygen-evolving enhancer protein1	Photoautotrophic growth	*Chlamydomonas reinhardtii (oee1-)*

1.3.2 Reverse genetic approaches

Reverse, or targeted, approaches direct specific genetic modifications at pre-determined sites in the genome; this could involve the deletion or addition of a gene, the up or down regulation of a gene, or introduction of a gene or a point mutation. Targeting constructs typically contain part of the gene to be targeted via homologous recombination, a reporter gene, a selectable marker, and transcriptional promoters and terminators. Targeted approaches rely on the availability of a method of transformation as well as genomic information and prior knowledge of gene function in order to identify targets. Development of targeted approaches has been hindered by the lack of sequenced genomes, low rates of homologous versus random integration of inserts during transformation and diploid life stages in certain algal groups (Cerutti *et al*., 2011; Parker *et al*., 2008).

1.3.2.1 Homologous recombination

Unlike transformation of the chloroplast, a low frequency of homologous recombination was initially observed in nuclear transformation microalgae (Sodeinde and Kindle, 1993; Nelson and Lefebvre, 1995). Homologous recombination in the nuclei of *Chlamydomonas* has been reported, however, not at the frequency required to allow its development as a technology for gene targeting (Dawson *et al*., 1997; Hallmann *et al*., 1997; Minoda *et al*., 2004; Zorin *et al*., 2005; Sodeinde and Kindle, 1993; Nelson and Lefebvre, 1995). Homologous recombination efficiency suitable for biotechnology has been reported in just two eukaryotic algae, *Cyanidioschyzon merolae* (Minoda *et al*., 2004) and *Nannochloropsis oceanica* (Kilian *et al*., 2011). However, in the case of *N. oceanica* homologous recombination was used to generate knock-out mutants of genes involved with nitrogen assimilation, providing a strong negative selection with which to enable easy identification of mutants demonstrating successful homologous recombination; given the biotechnological relevance of this species, this may be the reason that no further studies altering metabolically benign genes have since begun to emerge (Kilian *et al*., 2011).

1.3.2.2 CRISPR/Cas9

The CRISPR/Cas9 system for gene editing, based on the use of a guide RNA sequence which directs precise DNA restrictions within the host genome, performed by the Cas9 nuclease, is being explored in microalgae (Jinek *et al*., 2012). Its first use in microalgae was in the model alga *C. reinhardti,* and while the components of the Cas9 system did function to cause targeted

gene modifications, it was not possible to obtain transformants exhibiting stable expression of Cas9 (or a gene encoding a Cas9 lacking nuclease activity) due to the inherent toxicity of the protein when constitutively expressed, calling for future modification of the system to stem Cas9 production and/or activity following gene editing (Jiang *et al*., 2014). More recently, CRISPR/Cas9 has been successfully applied in *N. gaditana* to enable insertional mutagenesis of 18 transcription factors identified as being responsible for putative negative regulation of lipid production under nitrogen deprivation (Ajjawi *et al*., 2017). One knockout mutant of a homolog of fungal $Zn(II)_2Cys_6$-encoding genes exhibited improved partitioning of total carbon to lipid from 20% in the wildtype to 40-50% in the mutant; this knockout strain grew poorly, however, attenuation of the gene using RNAi yielded strains producing twice as much lipid (~5.0 g m^{-2} d^{-1}) as the with little effect on growth (Ajjawi *et al*., 2017).

1.3.2.3 RNAi

The increasing number of sequenced algal genomes has enabled identification of components of the RNA-mediated silencing machinery, which can convert double-stranded RNA (dsRNA) into small interfering RNAs (siRNAs) which inhibit transcription of translation by neutralizing target mRNAs (Sontheimer and Carthew, 2005; Ghildiyal and Zamore, 2009; Voinnet, 2009; Fabian *et al*., 2010). RNAi has been used in *C. reinhardtii* to downregulate, or knockdown, a variety of genes (Sineshchekov *et al*., 2002; Rohr *et al*., 2004; Soupene *et al*., 2004; Schroda, 2006). Stable silencing of genes using RNAi has also been demonstrated in *V. carteri* (Ebnet, 1999; Cheng *et al*., 2006), *D. salina* (Jia *et al*., 2009), *P. tricornutum* (De Riso *et al*., 2009), and *E. gracilis* (Iseki *et al*., 2002). RNAi is a valuable technology where controllable attenuation of a gene is preferred over complete knockout of the gene, as in a recent study in *N. gaditana* where the expression of a *ZnCys* gene was lowered in a stepwise manor over several strains in order to generate a mutant with a 2-fold increase in lipid content without negatively effecting growth (see figure 1.11) (Ajjawi *et al*., 2017).

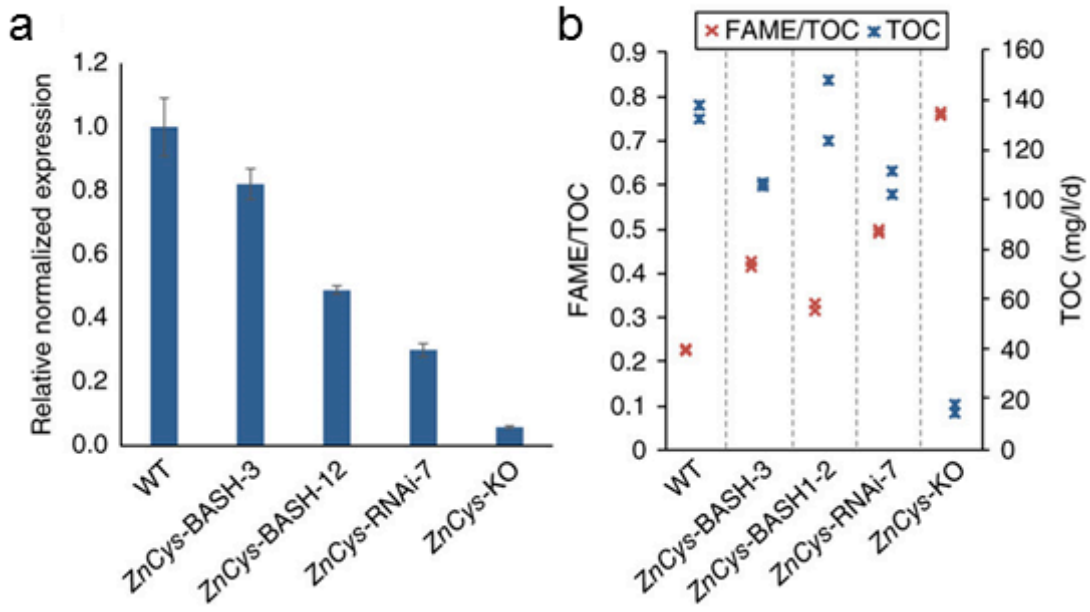


Figure 1.11: Generation of attenuated *ZnCys* alleles in *N. gaditana* using RNAi. (**a**) Steady-state mRNA levels of *ZnCys*-attenuated lines (*ZnCys*-BASH-3, *ZnCys*-BASH-12, *ZnCys*-RNAi- 7 and *ZnCys*-KO) relative to WT, determined by qRT–PCR. (**b**) TOC productivity (blue) and FAME/TOC (red) values of *ZnCys*-mutant lines. FAME: fatty acid methyl esters; lipid, TOC: total organic carbon. Figure adapted from Ajjawi *et al*. (2017).

1.3.3 Forward genetic approaches

A principle component of forward genetic approaches is mutagenesis. Mutagenesis is a process by which rates of mutation in a given organism are increased using a physical or chemical mutagen; the resulting population of mutants may then be screened for a desired phenotype (Kodym and Afza, 2003). Many different physical and chemical mutagens have been used to generate mutants with improved performance in the screening conditions applied.

1.3.3.1 UV light

Exposure to ultraviolet (UV) light induces point mutations in the DNA preferentially at methyl-CpG sites (Ikehata and Ono, 2011). UV is a useful method of mutagenesis when genomic information and a method of transformation are not available for the microalgal species it is being employed in. UV mutagenesis has been used to increase the eicosapentaenoic acid content of *Phaeodactylum tricornutum* (Depauw *et al*., 2012). Improvements to the biomass and lipid content of *Chlorella* strains have also been achieved (Liu *et al*., 2015). UV mutagenesis has also been combined with fluorescence-activated cell sorting (FACS) to analyse lipid bodies and increase lipid yield in *Chlamydomonas* (Terashima *et al*., 2015; Sharma *et al*., 2015). Strains of

Nannochloropsis salina with 3 fold higher lipid accumulation than have recently been generated using ethyl methanesulfonate (EMS), a chemical-based mutagen, followed by UV mutagenesis on the EMS-generated strain (Beacham *et al*., 2015). However, reduced growth rates affected the overall lipid productivity, which is a common problem with mutant strains with increased lipid content.

1.3.3.2 Chemical mutagens

Similar to UV mutagenesis, chemical mutagens are widely used to induce mutagenesis and are particularly useful when genomic information and a transformation system are unavailable. Preferential mutation of genomic regions with high GC-content make chemical mutagens bias in certain cases, thus limiting the pool of mutant strains (Harper and Lee, 2012). A commonly used chemical mutagen in *Escherichia coli* and other bacterial systems is N´-nitro-N-nitrosoguanidine (NTG) (Harper and Lee, 2012). NTG has also been used to generate mutants with enhanced carotenoid accumulation and increased astaxanthin content in green algae (Sandesh Kamath *et al*., 2008). Ethyl methanesulfonate (EMS) is also a popular choice of chemical mutagenesis (Kodym and Afza, 2003). Successful mutagenesis using EMS enabled generation of algal mutants with increased lipid accumulation; for example, *C. reinhardtii* (Lee *et al*., 2014). EMS has also been used to generate mutant strains of *Nannochloropsis* sp. with enhanced lipid accumulation and improved light use efficiency (Doan and Obbard, 2012; Beacham *et al*., 2015; Perin *et al*., 2015).

1.3.3.3 Insertional mutagenesis

Insertional mutagenesis involves the integration of foreign DNA into the host genome and concurrent disruption of endogenous genes and as such requires a method of transformation (Hlavova *et al*., 2015). However, insertional mutagenesis has inherent advantages over chemical and UV based methods. The integration site of chimeric DNA into the genome is reportedly close to random enabling generation of large mutant pools (Zhang *et al*., 2014). Additionally, the insertion of a large piece of DNA provides an anchor for PCR based approaches which can identify the flanking region of the insert and thus the mutation site (Hlavova *et al*., 2015). Insertional mutagenesis has been used in the model plant organism *Arabidopsis thaliana* to prepare a library of tagged knockouts containing 225,000 mutants (covering the entire genome with a probability of 96.6%), providing a powerful tool for basic research (Alonso, 2003). Mutants generated using insertional mutagenesis can therefore be useful in two ways: 1) by providing a phenotype that may be of biotechnological interest, and 2) providing knowledge of gene function. Indeed, insertional mutagenesis has been applied extensively in *Chlamydomonas*

reinhartii to study gene function (Pollock *et al*., 2002; Dent, 2005; Fang *et al*., 2006; Galván *et al*., 2007; Gonzalez-Ballester *et al*., 2011; Jungnick *et al*., 2014). Mutants of *Chlamydomonas* with increased lipid content have also been generated using a library of insertional mutants (Terashima *et al*., 2015). In addition to basic research, if applied in biotechnologically relevant species and with appropriate screening procedures, insertional mutagenesis could also be transformative to the field of algal biotechnology.

Table 1.5 summarises the available tools for the main model organisms for algal biotechnology.

Table 1.5: Model organisms for algal biotechnology and available tools. Adapted from (Hlavova *et al*., 2015).

Species	Taxonomic group	Available tools
Chlamydomonas reinhardtii	Green algae, Chlorophyceae	Genome sequence, genetic transformation, the most developed molecular toolkit, chloroplast gene targeting available, lacks nuclear gene targeting
Cyanidioschyzon merolae	Red algae, Cyanidiophyceae	Genome sequence, genetic transformation, developed molecular toolkit, gene targeting available
Nannochloropsis sp.	Heterokonts, Eustigmatophyceae	Genome sequence, genetic transformation, partially developed molecular toolkit, gene targeting available
Ostreococcus tauri	Green algae, Prasinophyceae	Genome sequence, genetic transformation, developed molecular toolkit, gene targeting available
Phaeodactylum tricornutum	Heterokonts, Bacillariophyceae	Genome sequence, genetic transformation, developed molecular toolkit, gene targeting available
Thalassiosira pseudonana	Heterokonts, Coscinodiscophyceae	Genome sequence, genetic transformation, partially developed molecular toolkit

1.4 Aims and objectives of this research

The aim of this thesis was to use transgenic approaches to further realize the biotechnological potential of two species of biotechnologically relevant green microalgae, *D. tertiolecta* and *N. gaditana*. The objective with *D. tertiolecta* was to establish a nuclear transformation system to unlock the possibility of sophisticated genome editing techniques in this species. The objectives with *N. gaditana,* which is already readily transformable, were more broad: to establish an *in*

situ eGFP (enhanced green fluorescent protein) reporter system for characterization of protein expression, to design a novel system for inducible expression of heterologous proteins in *N. gaditana* that will be a valuable tool for basic research and potentially commercial applications, and finally to use insertional mutagenesis to identify endogenous promoters for use in genetic engineering and to generate mutant phenotypes with potentially improved characteristics for large-scale biofuel production systems.

Chapter 1 introduces photosynthesis in a cellular and marine microalgal context and gives a synthesis of commercial uses and genetic engineering of marine microalgae.

Chapter 2 gives a detailed description of the materials and methods used in this thesis.

Chapter 3 presents an *in situ* eGFP (enhanced green fluorescent protein) reporter system for *N. gaditana* and demonstrates the effectiveness of this system through the characterization of a novel endogenous promoter of the nitrate reductase gene capable of promoting inducible expression of heterologous proteins.

Chapter 4 describes the use of a random insertional mutagenesis approach to identify endogenous promoters in *N. gaditana*, screening procedures are described in addition to characterization of the expression levels of the obtained promoters and PCR methods used to identify the corresponding genes. The efficiency of the promoters was then validated using the eGFP reporter system.

Chapter 5 describes the use of random insertional mutagenesis and high-throughput phenotypic screening to generate two mutant strains of *N. gaditana* with improved growth characteristics for biofuel production systems. Two mutants strains, both with increased resistance to ROS (reactive oxygen species) stress and increased tolerance to high light are isolated and characterised. The effected gene of one of these mutants is identified by PCR. Additionally, the components for a high-throughput FTIR (Fourier-transform infrared spectroscopy) based screen for increased lipid content are presented.

Chapter 6 outlines transformation parameters for *D. tertiolecta* and details attempts to establish a nuclear transform system using *Agrobacterium,* biolistics, electroporation, and glass beads.

Chapter 7 gives a synthesis of the biotechnological advancements made in each species and their suitability for genetic engineering. The significance of the genetic elements and technologies developed in *N. gaditana* is discussed in addition to that of the mutant phenotypes generated. The justification for further studies is presented based on the observations made in this thesis.

Chapter 2: Materials and Methods

2.1 Description of microalgal strains

2.1.1 Description of *N. gaditana*

N. gaditana CCMP526 was obtained from the National Centre for Marine Algae and Microbiota (NCMA) culture collection. This strain was originally isolated from Lagune de Oualidia, Morocco in 1985. The cells of *Nannochloropsis* sp. are indistinguishable on the basis of morphology and are characterized primarily by 18S rDNA sequence analysis (Andersen *et al*., 1998). All species are small, non-motile and spherical in shape with no distinct morphological features. They have a diameter of roughly 1-2 µm and a simple ultrastructure compared to closely related taxa (see Figure 2.1) (Kandilian *et al*., 2013).

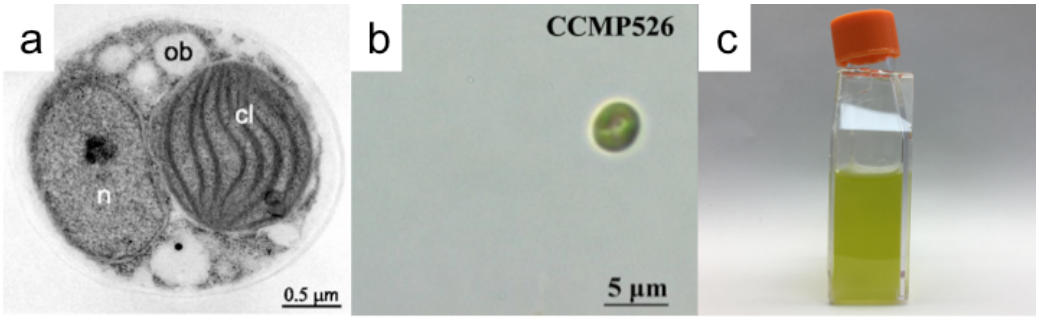


Figure 2.1: Morphology of *N. gaditana*. **(a)** Electron micrograph image of *N. gaditana*. Showing: n, nucleus; cl, chloroplast; ob, putative oil bodies. Scale bar: 0.5 µm. Figure adapted from Simionato *et al*. (2013). **(b)** Light microscope image of *N. gaditana.* Showing ~3 µm spherical cell with simple ultrastructure. Sale bar: 5 µm. **(c)** 40 mL *N. gaditana* culture in laboratory-scale culture flask.

2.1.2 Description of *D. tertiolecta*

D. tertiolecta CCMP364 was obtained from NCMA culture collection. This strain was originally isolated from Oslo fjord in Norway in 1928. *Dunaliella* sp. are motile, unicellular,

oval shaped cells ranging in diameter from 7-11 µm (see Figure 2.2). The cells lack an outer cell wall and have a stigma near the location where two flagella are inserted. The presence of a single cup shaped chloroplast enveloped in the inner cell membrane gives the cells a light green colour. The chloroplasts have a basal pyrenoid containing the majority of the RuBisCo – the enzyme responsible for fixation of inorganic carbon. The pyrenoid is visible due to the proximal accumulation of starch granules forming the amylosphere (Borowitzka and Siva, 2007). Also visible are small and numerous granules or droplets containing carotenoid pigments (Borowitzka and Siva, 2007). The other main features are the cell nucleus, and the Golgi apparatus, which is involved in the distribution of synthesized compounds including oil droplets (Andersson, 2001).

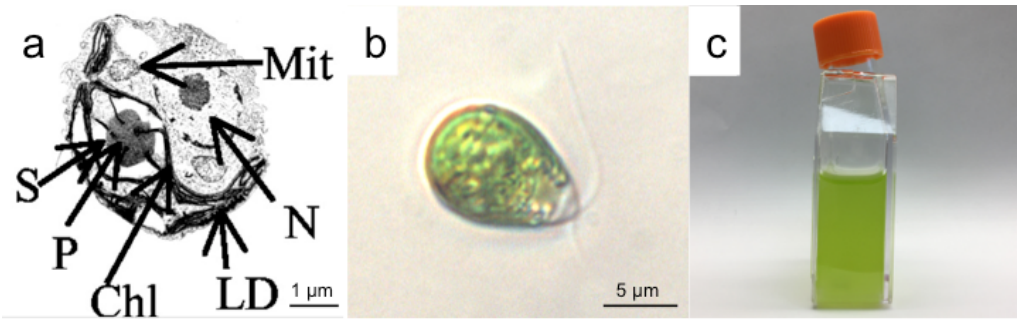


Figure 2.2: Morphology of *D. tertiolecta*. **(a)** Electron micrograph image of *D. tertiolecta.* Scale bar: 1 µm. Showing: Chl, Chloroplast; LD, lipid drops; Mit, mitochondria; N, nucleus; NS, nuclear space; P, pyrenoid; S, starch. Figure adapted from Segovia (2003). **(b)** Light microscope image of ~7 µm *D. tertiolecta* cell showing organelles and flagella. Scale-bar: 5 µm. **(c)** 40 mL *D. tertiolecta* culture in laboratory-scale culture flask.

2.2 Microbiological techniques

2.2.1 Culturing and maintenance of *N. gaditana*

N. gaditana was maintained in sterile conditions in a modified f/2 media (F2N) which is identical to f/2 media (Fallis, 1975; Guillard and Ryther, 1962) except that trace metals, vitamins, and phosphate solutions were added in fivefold higher concentrations, 10 mM Tris·HCl (pH 7.6) was added to maintain the pH, and 5 mM NH_4Cl was included as a nitrogen source (see Table 2.1) (Kilian *et al*., 2011). Cultures were maintained in liquid media and on F2N agar plates (92 x 21 mm). The media was prepared by adding 30 g of Tropic Marin artificial sea salts (Tropical Marine Centre LTD, Rickmansworth, UK) to 1 L of ddH_2O, autoclaving (121 °C for 40 min) and cooling before addition of the media. Agar plates were prepared by adding 1.5% (w/v) of agar to artificial seawater before autoclaving, cooling and

addition of media. For agar plates containing an antibiotic selective agent, artificial sea salts and NH_4Cl was lowered to 15 g/L and 2 nM, respectively, to prevent inactivation of the antibiotic. Liquid stock cultures were sub-cultured every second week, taking a 200 µL inoculum and transferring to 40 mL of fresh media in 25 cm^2 vented Corning flasks (Thermo Scientific UK Ltd, Loughborough, UK) and incubated under ~70 µmol photons m^{-2} s^{-1} fluorescent light (12 hr light, 12 hr dark cycle) at 24 – 30 °C. Agar plates were restreaked onto fresh media every 2 months and incubated under ~100 µmol photons m^{-2} s^{-1} continuous fluorescent light at 20 °C. All culture maintenance was performed in a flow hood to ensure sterility. All nutrients were obtained from Thermo Scientific UK Ltd.

Table 2.1: Chemical (mg/L) and molar (M) compositions of F2N media used for cultivation of *N. gaditana.*

Compound	Amount (mg/L)	Molar concentration (M)
NH_4Cl	75	4.41×10^{-3}
$NaH_2PO_4 \cdot H_2O$	25	1.81×10^{-4}
$Na_2SiO_3 \cdot 9H2O$	30	1.06×10^{-4}
$FeCl_3 \cdot 6H_2O$	16	5.90×10^{-5}
$Na_2EDTA \cdot 2H_2O$	22	5.90×10^{-5}
$CuSO_4 \cdot 5H_2O$	0.049	1.97×10^{-7}
$Na_2MoO_4 \cdot 2H_2O$	0.0315	1.30×10^{-7}
$ZnSO_4 \cdot 7H_2O$	0.11	3.83×10^{-7}
$CoCl_2 \cdot 6H_2O$	0.05	2.10×10^{-7}
$MnCl_2 \cdot 4H_2O$	0.9	4.55×10^{-6}
Cobalamine (Vitamin B_{12})	0.0025	1.85×10^{-9}
Biotin (Vitamin H)	0.00025	1.03×10^{-8}
Thiamine HCl (Vitamin B_1)	0.5	1.48×10^{-6}

2.2.1.1 *N. gaditana* growth curve

Growth experiments were typically set up by inoculating triplicate cultures in F2N media, normalized to an $A_{540\,nm}$ value of 0.02 (see section 2.2.4). $A_{540\,nm}$ measurements where then taken every one or two days depending on the experiment. Figure 2.3 shows a standard *N. gaditana* growth curve in F2N media. The onset of the logarithmic growth phase can be seen on day 2.

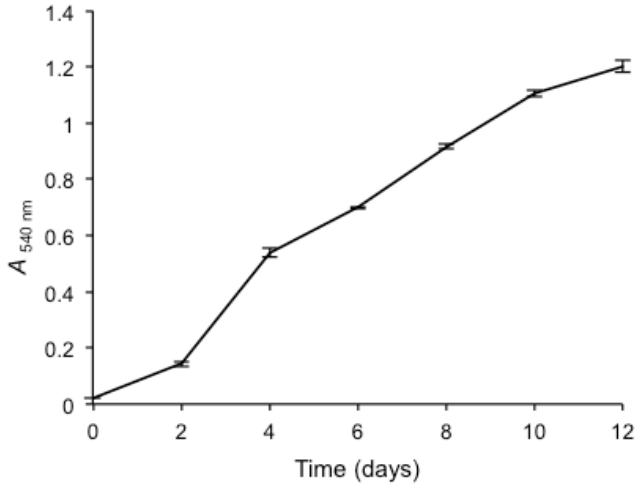


Figure 2.3: Standard growth curve of *N. gaditana* in F2N media measured by absorbance at wavelength of 540 nm ($A_{540\ nm}$). Values are means of triplicate measurements.

2.2.2 Culturing and maintenance of *D. tertiolecta*

D. tertiolecta was maintained in sterile conditions in f/2-Si media (see Table 2.2) (Fallis, 1975; Guillard and Ryther, 1962). Cultures were maintained in liquid and on f/2-Si agar plates (92 x 21 mm). The media was prepared by filtering seawater through Whatman Grade GF/A Glass microfiber (Thermo Scientific UK Ltd) filter disks followed by autoclaving (121°C for 40 min) and cooling before addition of the media. Agar plates were prepared by adding 1.5% (w/v) of agar to sterilized seawater before re-autoclaving, cooling and addition of media. Liquid stocks were sub-cultured every second week, taking a 50 µL inoculum and transferring to 40 mL of fresh media in 25 cm^2 vented Corning flasks (Thermo Scientific UK Ltd) and incubated under ~100 µmol photons m^{-2} s^{-1} fluorescent light (12 hr light, 12 hr dark cycle) at 20 °C. Agar plates were restreaked onto fresh media every 2 months and incubated under ~100 µmol photons m^{-2} s^{-1} continuous fluorescent light at 20 °C. All culture maintenance was performed in a flow hood to ensure sterility. All nutrients were obtained from Thermo Scientific UK Ltd.

Table 2.2: Chemical (mg/L) and molar (M) compositions of f/2-Si media used for cultivation of *D. tertiolecta.*

Compound	Amount (mg/L)	Molar concentration (M)
$NaNO_3$	75	8.82×10^{-4}
$NaH_2PO_4 \cdot H_2O$	5	3.62×10^{-5}
$FeCl_3 \cdot 6H_2O$	3.2	1.18×10^{-5}
$Na_2EDTA \cdot 2H_2O$	4.4	1.18×10^{-5}
$CuSO_4 \cdot 5H_2O$	0.0098	3.93×10^{-8}
$Na_2MoO_4 \cdot 2H_2O$	0.0063	2.60×10^{-8}
$ZnSO_4 \cdot 7H_2O$	0.022	7.65×10^{-8}
$CoCl_2 \cdot 6H_2O$	0.01	4.20×10^{-8}
$MnCl_2 \cdot 4H_2O$	0.18	9.10×10^{-7}
Cobalamine (Vitamin B_{12})	0.0005	3.69×10^{-10}
Biotin (Vitamin H)	0.00005	2.05×10^{-9}
Thiamine HCl (Vitamin B_1)	0.1	2.96×10^{-7}

2.2.3 Cell numbers and growth rates

Cell numbers were determined using a Beckman Multisizer 3 Coulter Counter (Beckman Coulter Life Sciences, Indiana, US). 100 µL of sample was added to 9.9 mL of diluent. Three technical repeats using 500 µL volumes of these samples were counted using a 70 µL aperture tube. The sizing threshold was 1.8 µm with a noise level of 1.6 µm. The current was set to 800 µA with a gain of 4. Filtered 3% NaCl solution prepared using deionized water was used as a diluent and electrolyte. The aperture was flushed before and between measurements.

Specific growth rates (µ) were calculated with the following equation:

$$\mu = \frac{\ln \frac{\chi_2}{\chi_1}}{t_2 - t_1} \tag{2.1}$$

where χ is the biomass of the culture at time t

2.2.4 Spectrophotometric measurements

Spectrophotometric measurements of cell density were taken using a Jenway 7315 spectrophotometer (Keison Products, Essex, UK). Absorbance values at a wavelength of 750 nm ($A_{750\ nm}$) and 540 nm ($A_{750\ nm}$) were used to measure the cell density of *N. gaditana* cultures. For each measurement, 1 mL of sample was loaded into 12.5 mm width polystyrene cuvette (Thermo Scientific UK Ltd). 1 mL of media was used to blank the spectrophotometer at the appropriate wavelengths.

To convert absorbance measurements to cell concentrations (cells per mL), a series of dilutions of *N. gaditana* wild type cells was prepared in F2N media. The cell concentration of each sample was measured using a Coulter counter, as described previously, in addition to its spectrophotometrically measured $A_{540\ nm}$ and $A_{750\ nm}$ value. These values were plotted together to generate calibration curves (cells per mL versus $A_{540\ nm}$ and $A_{750\ nm}$) (see Figure 2.4). Standard curves drawn to these plots were used to convert absorbance measurements to cells/mL for all transformations and experiments. $A_{540\ nm}$ values were used for cell density measurements thereafter. Where necessary, samples were diluted with media before measuring $A_{540\ nm}$.

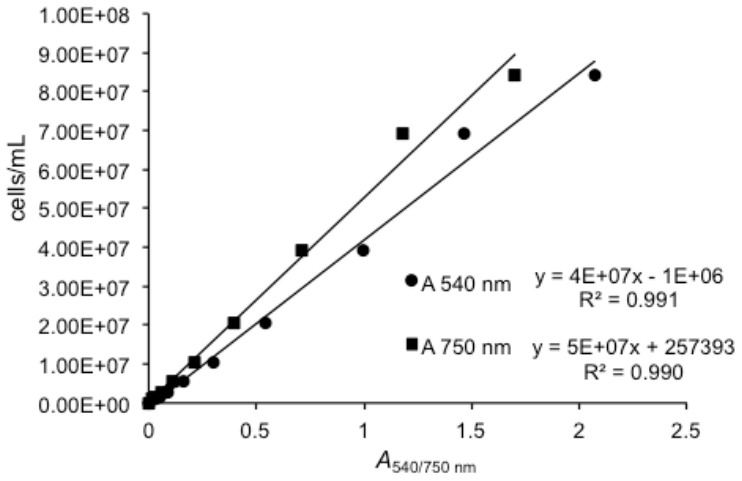


Figure 2.4: Standard curves of absorbance at 750 nm ($A_{750\ nm}$) and 540 nm ($A_{540\ nm}$) vs cells/mL obtained using spectrophotometry and a Coulter counter. The correlation between $A_{750\ nm/540\ nm}$ and cells/mL was determined by means of standard curve. y is cells/mL; x is $A_{750\ nm/540\ nm}$ value.

2.2.5 Microplate reader assays

2.2.5.1 Green fluorescent protein (GFP) measurements

The fluorescence of live cells was analysed using a FLUOstar OPTIMA microplate reader (BMG Labtech Ltd, Aylesbury, UK) in black-bottomed 96-well plates (Thermo Scientific UK Ltd) in triplicate. 100 µL or 200 µL samples were used for each measurement. eGFP fluorescence was measured using filter settings appropriate to the maximal emission and excitation wavelengths for eGFP, 488 nm and 509 nm, respectively. F2N media was used as a blank for all *N. gaditana* samples; average blank measurements (3*n*) were subtracted from all fluorescence measurements. Fluorescence measurements for each replicate were then normalised to its absorbance at a wavelength of 540 nm ($A_{540\ nm}$). Finally, average normalised wild type fluorescence values (3*n*) were then subtracted from all normalised fluorescence values to determine eGFP fluorescence or divided into normalised fluorescence values to determine fold changes.

For quantification of eGFP fluorescence as ng eGFP per 10^7 cells, normalised wild type fluorescence-subtracted fluorescence measurements were converted to eGFP concentrations in ng/µL using eGFP calibration curves (see Figure 2.5). For the eGFP calibration curves, 100 µL eGFP standards were prepared by diluting a recombinant eGFP (MyBioSource Inc, San Diego, US) in dH2O at the following concentrations (ng/µL): 0, 0.01, 0.05, 0.1, 0.25, 0.5, 1, 2.5, 5, and 10. The $A_{540\ nm}$ to cells/mL calibration curve was used to calculate the cell concentration in

cells/µL at an $A_{540\,nm}$ of 1, eGFP concentrations in ng/µL for each replicate were then divided by this value to determine the ng of eGFP per 10^7 cells.

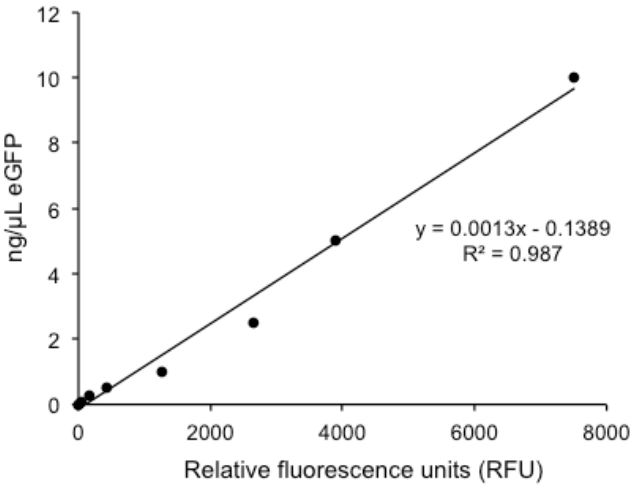


Figure 2.5: eGFP (enhanced green fluorescent protein) standard curve of relative fluorescence units (RFU) versus ng/µL eGFP. The correlation between RFU and ng/µL eGFP was determined by means of standard curve. y is ng/µL eGFP; x is RFU.

2.2.5.2 Reactive oxygen species (ROS) assay

5-(and-6)-carboxy-2',7'-dihydrofluorescein diacetate (H_2DFFDA; Invitrogen, California, US) was used as an indicator for reactive oxygen species (ROS), using methods described by Szivák *et al* with slight modifications (Szivák *et al*., 2009). Upon oxidation, H_2DFFDA is converted to fluorescent DFFDA, which can be detected at 488 nm excitation with a 520 nm emission filter. The following sterile stock solutions were prepared for the assay: 0.01 M 3-(N-morpholino)propanesulfonic acid (MOPS; pH 7.0; Thermo Scientific UK Ltd), 0.05 M ethylenediaminetetraacetic acid (EDTA; Thermo Scientific UK Ltd) and 10 mM H_2DFFDA in DMSO (Thermo Scientific UK Ltd). The H_2DFFDA stocks were split into 10 µL aliquots and stored at -20 °C.

15 mL cultures of each *N. gaditana* strain tested was harvested at 3 x 000 *g* (times gravity) for 10 min, washed in 0.01 M MOPS buffer solution twice and re-suspended in 5 mL of 0.01 M MOPS. 100 µL of each sample was then used to run cell counts (see section 2.2.3). 1 mL of each sample was normalized to 10^6 cells/mL. To each of these solutions was added, 1 µL of 10 mM H_2DFFDA in DMSO (10 µM final concentration) and 10 µL of 0.05 M EDTA (0.5 mM final concentration) as a permeabilization agent. Negative controls were included in each experiment containing DMSO, heat-killed cells (95 °C for 10 min), and EDTA. The solutions were mixed, incubated for 60 min in the dark, and then analysed on a FLUOstar OPTIMA microplate reader (BMG Labtech Ltd) in black-bottomed 96-well plates (Thermo Scientific UK

Ltd) in triplicate. 200 µL samples were used for each measurement. DFFDA fluorescence in each sample was measured using the appropriate filter settings and expressed as relative fluorescence units (RFU).

2.2.6 Fast repetition rate fluorometry (FRRf)

A MkII FastTracka Fast Repetition Rate fluorometer (FRRf) (Chelsea Technologies Group, West Moseley, Surrey, U.K.) was used to measure maximum photosynthetic efficiency (F_v/F_m) and absorption cross-section of photosystem II (σ_{PSII}).

Light that is absorbed by photosynthetic organisms follows three competing pathways. It may be used to produce ATP and NADPH through photosynthesis (photochemistry), re-emitted as fluorescence, or dissipated in the form of heat (Baker, 2008). F_v/F_m measurements are usually performed following a dark adaptation period in order to allow relaxation of any upregulation of the heat dissipation pathway (through a series of mechanisms collectively termed non-photochemical quenching) and hence the maximum amount of light to follow the photochemical and fluorescence pathways. The measurement is a representation of the difference in fluorescence between a state when all the reaction centres are open (F_o) and the maximum fluorescence (F_m) when they are all closed; this difference (F_v) is normalized to (F_m) to obtain a ratio of F_v/F_m (Falkowski and Raven, 1997). In maximum fluorescence, the maximum number of reaction centres have been reduced or closed by a saturating light source. In general, the greater the stress to the photosynthetic organism, the fewer functional reaction centres available, lowering the F_v/F_m ratio (see Figure 2.6). σ_{PSII} is a measure of the function cross-section of photosystem II in the dark-adapted state.

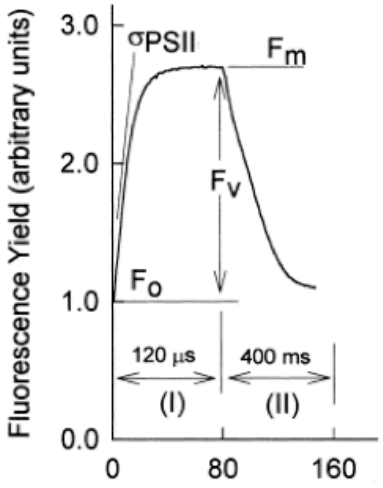


Figure 2.6: Principles of fast repetition rate fluorometry (FRRf) measurements of F_v/F_m and σPSII. The fluorescence yield in the energised state with all centres open (F_o) is briefly determined in the dark-adapted state. The maximum florescence (F_m) is then determined after a saturating light pulse. The maximum photosynthetic efficiency (F_v/F_m) is determined by dividing the difference between F_o and F_m (F_v) over F_m. ML, weak modulated measuring light; SP, saturating light pulse. Figure adapted from Kolber *et al*. (1998).

1 mL of each sample was dark adapted for 20 min before being loaded into the dark chamber of the instrument. The settings listed in Table 2.3 were used for all single turnover (ST) measurements. The photomultiplier (PMT) was adjusted during measurements to keep the maximum fluorescence (F_m) centered at 10000 ± 2000 (dimensionless).

Table 2.3: Measurement settings for MkII FastTracka Fast Repetition Rate fluorometer.

Parameter	**Setting**
Sequences per acquisition	24
Sequence interval (ms)	100
Acquisition pitch (s)	3
Flashlets in sat. phase	150
Flashlet duration in sat. phase (µs)	1
Flashlets in rel. phase	25
Flashlet duration in rel. phase (µs)	84
Saturation phase duration (ms)	0.3
Relaxation phase duration (ms)	2.1
LED set	50

Additionally, the F_v/F_m at different light levels was measured by running rapid light curves (RLCs) under light intensities ranging from 0 to 2000 µmol photons m^{-2} s^{-1}. A light challenge RLC was designed which consisted of a 60 second pre-RLC dark period, 12 light/dark steps and

a 1200 second post-RLC dark period. Each light step was for 240 seconds at 2000 µmol photons m^{-2} s^{-1} each dark step was for 30 seconds.

2.2.7 Fourier-transform infrared spectroscopy (FTIR)

Fourier-transform infrared spectroscopy (FTIR) can be used to measure the infrared absorption attributed to specific molecular vibrational modes (Murdock and Wetzel, 2009). Macromolecular groups including, proteins, lipids, carbohydrates and nucleic acids, can be quantified by measuring their characteristic absorption bands (see Table 2.4).

Table 2.4: Band assignments for infrared spectroscopy. Table adapted from Stehfest *et al*. (2005).

Wavenumber (cm^{-1})	Assignment
3000-2800	νC-H of saturated CH
~2925	νC-H of lipids
~1740	νC=O of ester groups, primarily from lipids and fatty acids
~1650	νC=O of amides from proteins (amide I band)
~1540	δN–H of amides from proteins (amide II band)
~1455	δ_{as} CH_3 and δ_{as} CH_2 of proteins
~1398	δ_s CH_3 and δ_s CH_2 of proteins, and ν_s C-O of COO^- groups
~1250-1230	ν_{as} P=O of phosphodiester groups of nucleic acids and phospholipids
~1200-900	νC–O–C of saccharides
~1075	νSi–O of silicate frustules

ν_{as}, asymmetric stretch; ν_s symmetric stretch; δ_{as}, asymmetric deformation; δ_s, symmetric deformation.

A Nicolet iS10 FTIR spectrometer (Thermo Scientific UK Ltd) with an XY Autosampler (PIKE Technologies, Fitchburg, US) was used to determine the relative lipid content of *N. gaditana* cell lines using previously described methods (Dean *et al*., 2010).

Samples were collected either from agar plates or pelleted from 1 mL of liquid culture and re-suspended in 30 µL of dH_2O and loaded into each well of the microplate of the XY Autosampler. The microplate was then incubated at 40 °C for 30 min to remove all water from the samples that would otherwise interfere with the absorption spectra. The OMNIC (Thermo Scientific UK Ltd) and AutoPRO (PIKE Technologies) software associated with the Nicolet iS10 and XY Autosampler was programed to take a baseline measurement in an empty well (360 scans) and then measure all other wells containing samples (60 scans).

For analysis of spectra, previously published peak attributions were identified in *N. gaditana* (see Figure 2.7) (Giordano *et al*., 2001; Pistorius *et al*., 2009; Stehfest *et al*., 2005). All spectra were baseline corrected using the automatic baseline function in the OMNIC software to minimize differences between spectra due to baseline shifts. Spectra were then scaled to the peak height of the amide II band (1545 cm^{-1}). Lipid accumulation was expressed as a lipid:amide I ratio using the peak height measured at either 2921 cm^{-1} or 1741 cm^{-1}, attributed to lipids, and at 1655 cm^{-1} attributed to amide I (protein). A macro was designed to automate this analysis using the OMNIC Macros Basic software (Thermo Scientific Uk Ltd).

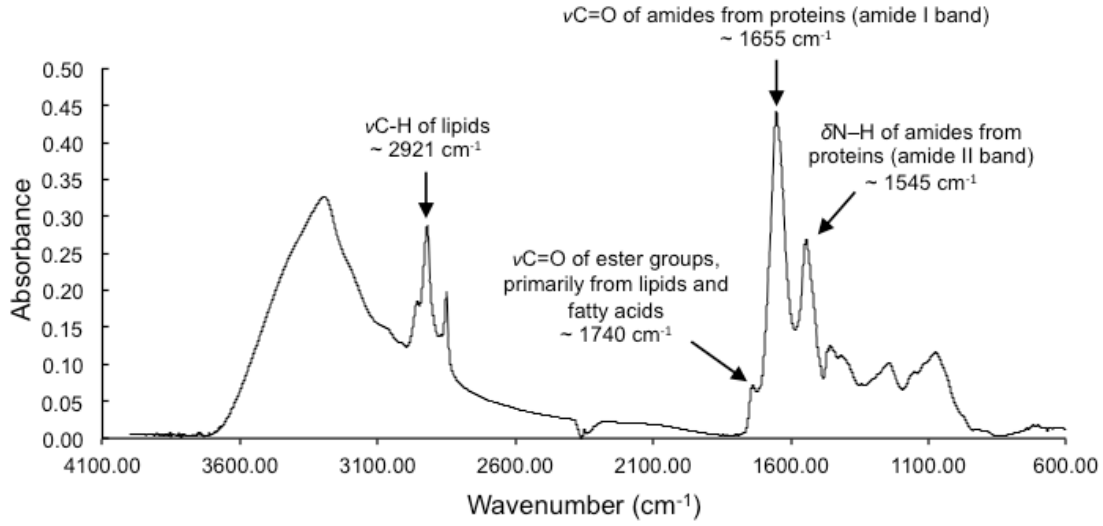


Figure 2.7: Baseline corrected *N. gaditana* FTIR spectra showing absorption bands used for lipid analysis: lipid peaks at ~2920 cm^{-1} and ~ 1740 cm^{-1}; amide I peak at ~1655 cm^{-1}; amide II peak at ~1545 cm^{-1}.

2.3 Molecular techniques

2.3.1 PCR amplification of DNA

Polymerase chain reactions (PCRs) were carried out routinely in order to amplify nucleotide sequences for vector assembly and to genotype cell lines using a SureCycler 8800 (Agilent, California, US). PCR reaction mixes contained 12.5 µL of either Phire Green Hot Start II PCR Master Mix (2x) (New England Biolabs Ltd, Hitchin, UK) or Q5 Hot Start PCR Master Mix (2x) (New England Biolabs Ltd), 9.5 µL of ddH_2O, 1 µL of forward primer (10 µM), 1 µL of reverse primer (10 µM) and 1 µL of template DNA for a final volume of 25 µL. PCR cycles began with hot-start activation at 95 °C for 1 min, followed by 30 amplification cycles consisting of denaturation at 95 °C for 10 seconds, annealing at ~60 °C (depending on T_m of primer set) for 10 seconds, elongation at 7 °C for 10 seconds, extension at 72 °C for 2 min.

Reactions were held at 4 °C before analysis by gel electrophoresis or long-term storage at -20 °C. All primers were obtained from Thermo Scientific UK Ltd.

2.3.2 Genotyping

To confirm the integration of transgenic DNA into the genome of *N. gaditana* cell lines, PCRs were performed using primers specific to the transgenic DNA. The primers used to genotype transformants generated with each plasmid used are listed in Table 2.5.

Table 2.5: Nucleotide sequences of primers used for genotyping of transgenic *N. gaditana* cell lines. Fw, forward primer; rv, reverse primer; bp, basepairs.

Primer	Sequence (5′ to 3′)	Plasmid used for transformation	Amplicon
Hygro F1 (fw)	ggaacttcaagcttaccatggatag	PAlg.Hyg.SD	Amplicon within *hph* gene (435 bp)
Hygro R1 (rv)	ctcgctgaattcgccgatgtccagc		
hph fw	atgggtaaaaagcctgaactcaccg	pLCP, pNR, pLDH, pHYP1, pCC7, pHYP2	*hph* gene (1029 bp)
hph rw	ttattcctttgccctcggacgagtgc		
B/C.SD F1 (fw)	atggccaagctgaccagcgccgttc	pCambia.Ble2AChr.SD, pAlg.Ble2AChr.SD.SD	Amplicon within *ble* gene (160 bp)
B/C.SD R1 (rv)	ggaccgcgctgatgaacagggtcac		

2.3.3 RNA extractions

RNA extractions were performed using the TRIzol method (Gauthier *et al*., 1997). Cells were isolated by centrifugation of approximately 10 mL of medium density culture at 2,500 x *g* for 10 min and re-suspended in 1 mL of TRIzol (Thermo Scientific UK Ltd). Samples were allowed to stand at room temperature for 5 min then 0.2 mL of chloroform (Thermo Scientific UK Ltd) was added. Samples were covered tightly, shaken vigorously for 15 seconds, and allowed to stand for a further 10-15 minutes at room temperature. The resulting mixture was centrifuged at 12,000 x *g* for 15 minutes at 2-8 °C. Centrifugation separates the mixture into 3 phases: a red organic phase (containing protein), an interphase (containing DNA), and a colourless upper aqueous phase (containing RNA). The aqueous phase was transferred to a fresh tube containing 0.5 mL of 2-propanol (Thermo Scientific UK Ltd) and mixed. Samples were allowed to stand at room temperature for 5-10 minutes and then centrifuged at 12,000 x *g* for 10 minutes at 2-8 °C. The RNA precipitate forms a pellet on the side and bottom of the tube. In order to wash the RNA, the supernatant was removed and 1 mL of 75% ethanol was added before vortexing the

samples and centrifuging again at 12,000 x *g* for 5 minutes at 2-8 °C. Supernatant were removed and pellets were air dried for 5 minutes in a Laminar Flow Cabinet. 5-10 µL of RNase-free ddH_2O was added to the pellets and mixed by repeated pipetting. RNA was quantified using a Nanodrop ND-1000 Spectrophotometer (Thermo Scientific UK Ltd), concentrations typically ranged from 500-1200 ng/µL; A_{260}/A_{280} ratios were ≥ 1.7.

2.3.4 3′ RACE

3′ RACE (rapid amplification of cDNA ends) was used to amplify nucleic acid sequences from a messenger RNA template between a defined internal site, in the form of an integrated vector sequence, and the 3′ end of the mRNA (Frohman *et al*., 1988).

Reverse transcription reactions were prepared on wet ice in PCR tubes; all reagents were vortexed briefly before use. Sample mixes were prepared by mixing 1.25 µL of 10x M-MulV RT buffer (New England Biolabs Ltd), 1.25 µL of the oligo-dT adaptor primer (Ap_oligo(dT)18), 0.5 µg of RNA, and ddH_2O to a final volume of 12.5 µL. Enzyme mixes were prepared by mixing 1.25 µL of 10x M-MuLV-RT buffer, 0.5 µL of 10 mM dNTPs (New England Biolabs Ltd), 0.5 µL of the RNase inhibitor RNasin (New England Biolabs Ltd, Hitchin, UK), 0.5 µL of M-MuLV-RT (New England Biolabs Ltd), and 9.75 µL of ddH_2O. Sample mixes were heated at 70 °C for 2 minutes then returned to wet ice before addition of enzyme mixes, bringing the total reaction volume to 25 µL. Samples were placed in a thermal cycler set at 42 °C for 60 minutes before the temperature was increased to 95 °C for 5 minutes. cDNA samples were then placed on ice or stored at -20 °C until processed further.

The primary and secondary nested PCR reactions (50 µL reaction volumes) were performed using either 2x Phire Green Hot Start PCR MasterMix (New England Biolabs Ltd) or the Advantage2 PCR Kit (Clontech Laboratories, Inc, Saint-Germain-en-Laye, France). The primary PCRs were performed using 2 µL of the cDNA reactions, a gene specific forward primer internal to the vector sequence, and a reverse primer specific to the adaptor sequence at the 3′ end of the RNA/cDNA template (AP-OUT), both primers at a final concentration of 0.5 µM. The cycling parameters for the first round of PCRs were a 2-step PCR with an annealing temperature of 68 °C and an extension time of 180 seconds, run for 40 cycles. The secondary nested PCRs were performed using 2 µL of the first round PCR products diluted 1:10 in ddH_2O, a gene specific forward primer nested in the 3′ direction of the gene specific primer used in the first round of PCRs, and a reverse primer specific to the adaptor sequence at the 3′ end of the RNA/cDNA template (AP-IN), both primers at a final concentration of 0.5 µM. The cycling parameters for the secondary nested PCRs were a 2-step PCR with an annealing temperature of 68 °C and an extension time of 180 seconds, run for 30 cycles. Positive control PCRs using

forward and reverse primers specific to the vector sequence were performed in the same run as the first round of PCRs in a separate reaction volume using the same cycling parameters. The products of secondary PCRs were separated by gel electrophoresis and gel extracted using the Zymoclean Gel DNA Recovery Kit (Cambridge Bioscience, Cambridge, England). The recovered fragments were sequenced with the Mix2Seq sequencing service (Eurofins Genomics, Anzinger, Germany) using the gene specific forward primer used in the secondary RACE PCRs. See Table 2.6 for a list of all of the primers used for the reverse transcription reaction and 3′ RACE PCRs.

Table 2.6: Nucleotide sequences of primers used for reverse transcription and 3´ RACE reactions. Fw, forward primer; rv, reverse primer.

Primer	Sequence (5′ to 3′)	Notes
Ap_oligo(dT)18	gccctaggcgagaacgagatctagctctagaattcggacgttttttttttttttttttvn	Reverse transcription reaction
AP-OUT	gccctaggcgagaacgagatctag	First RACE PCR adaptor primer
AP-IN	gagatctagctctagaattcggacg	Second RACE PCR adaptor primer
Hygro F1	ggaacttcaagcttaccatggatag	Fw internal control primer
Hygro R1	ctcgctgaattcgccgatgtccagc	Rw internal control primer
Hygro RT F2	cgaggtgctggccgattccggcaac	First RACE PCR gene specific primer
Hygro RT F3	attccggcaacaggcgcccttccac	Second RACE PCR gene specific primer
B/C.SD F1	atggccaagctgaccagcgccgttc	Fw internal control primer
B/C.SD R1	ggaccgcgctgatgaacagggtcac	Rw internal control primer
B/C.SD RT F2	cgatcgtggagcagtacgagcgcgc	First RACE PCR gene specific primer
B/C.SD RT F3	cacgggcggcatggacgagctgtac	Second RACE PCR gene specific primer
B/C.SD RT F2 b	cgatgacagagcgttgctgcctgtg	First RACE PCR gene specific primer
B/C.SD RT F3 b	ccgcggtttcaaaatcggctccgtc	Second RACE PCR gene specific primer

2.3.5 Arbitrary primed PCR

Arbitrary primed polymerase chain reactions (AP-PCRs) were performed to amplify genomic regions flanking transgenic plasmid DNA sequences following methods described previously (Tran *et al*., 2015; Das *et al*., 2005). Briefly, two arbitrary primed PCR reactions were performed, initially using genomic DNA as a temple, and an internal gene specific forward primer, central in location within the transgenic insert, in combination with an arbitrary reverse primer. The thermocycling parameters favour amplification of the amplicons that contain DNA from the flanking genomic region directed by the arbitrary primers. The products obtained in the first PCR were then used as temple DNA for a secondary nested AP-PCR reaction using a an external gene specific primer, close to the downstream end of the transgenic DNA, and a nested arbitrary reverse primer. The products of the second AP-PCRs were then sequenced to obtain the genomic flanking regions.

10 ng of genomic DNA was prepared using a DNeasy Blood and Tissue kit (Qiagen Ltd, Manchester, UK) and used as template DNA for all primary AP-PCRs. All AP-PCRs were performed using 2.5 U Taq polymerase (Qiagen Ltd) and 0.2 mM dNTPs (Qiagen Ltd), primers and primer concentrations were are described subsequently for each reaction.

When using methods described by Tran *et al*, the reaction mixture for the first AP-PCRs used 0.5 μM of the arbitrary primer Arb1 (Tran *et al*., 2015) and 0.2 μM of a gene specific forward primer (either B/C int or B/Cd int). The cycling parameters for the first AP-PCR were: 94 °C for 5 min; 6 cycles at 94 °C for 30 secs, 30 °C for 30 secs, 72 °C for 90 secs; 30 cycles at 94 °C for 30 secs, 58 °C for 30 secs, 90 °C for 90 secs; 72 °C for 5 min. The products of the first PCR were purified using a Zymo Clean and Concentrator kit (Cambridge Bioscience) and diluted 1:10 before being used as template for the secondary AP-PCR reactions. The secondary nested AP-PCRs used 0.2 μM of the nested arbitrary primer Arb2 (Tran *et al*., 2015) and 0.2 μM of the external gene specific primer (either B/C ext or B/Cd ext). The cycling parameters for the secondary PCRs were: 95 °C for 1 min; 30 cycles at 95 °C for 30 secs, 65 °C for 30 secs, 72 °C for 90 secs; 72 °C for 4 min. The purified products were then sent for sequencing using the external gene specific primers (Eurofins Genomics).

When following the methods of Das *et al*, the reaction mixture for the first AP-PCRs used 0.2 μM of the gene specific forward primer (either B/C int or B/Cd int) and 0.5 μM of the arbitrary primer Arb1-5 (Das *et al*., 2005). The cycling parameters for the first AP-PCR were: 94 °C for 5 min; 6 cycles at 94 °C for 30 secs, 30 °C for 30 secs, 72 °C for 90 secs; 30 cycles at 94 °C for 30 secs, 45 °C for 30 secs, 72 °C for 2 min; 72 °C for 4 min. The products of the first PCR were purified using a Zymo Clean and Concentrator kit (Cambridge Bioscience) and diluted 1:10 before being used as template for the secondary AP-PCR reactions. The secondary nested AP-

PCRs used 0.2 µM of the nested arbitrary primer Arb2b (Das *et al*., 2005) and 0.2 µM of the external gene specific primer (either B/C ext or B/Cd ext). The cycling parameters for the secondary PCRs were: 95 °C for 1 min; 30 cycles at 95 °C for 30 secs, 52 °C for 30 secs, 72 °C for 2 min; 72 °C for 4 min. The purified products were then sent for sequencing using the external gene specific primers (Eurofins Genomics).

See Table 2.7 for a list of all of the primers used for AP-PCRs.

Table 2.7: Nucleotide sequences of primers used for arbitrary primed PCR reactions.

Primer	Sequence (5′ to 3′)
Arb1	ggccaggcctgcagatgatgnnnnnnnnnngtat
Arb2	ggccaggcctgcagatgatg
Arb1-5	ggccacgcgtcgactagtcannnnnnnnnnctgct
Arb2b	ggccacgcgtcgactagtca
B/C int	agttcatctacaaggtgaagctgcg
B/C ext	atgaccgagatcggcgagcagcc
B/Cd int	cgtcagcgggagagccgttg
B/Cd ext	ggagggtagcatgttgattgtaa

2.3.6 Immunoblots

40 mL cultures of *N. gaditana* were used to prepare whole cell protein extracts. Cells were harvested by centrifugation at 3,000 x *g* for 10 min at room temperature. For homogenization, approximately 100 mg of 0.1 mm zirconia beads (Biospec Products, Oklahoma, US) were added to the pelleted cells in addition to an equal volume (~150 µL) of SDS lysis buffer (200 mM NaCl, 25 mM EDTA, 0.5% (w/v) SDS, 200 mM Tris-Cl, pH 8.5). Cells were then lyzed in a TissueLyser II tissue lyser (Qiagen Ltd) for 2 x 30 second cycles at a frequency of 30 oscillations per second. Samples were heated to 95°C in a heat block for 10 min, cooled briefly on ice, and then pelleted by centrifugation at 17,000 x *g* for 10 min at 4°C. The total protein content of the lysate was quantified using a Pierce bicinchoninic acid (BCA) protein assay (Thermo Scientific UK Ltd) with a bovine serum albumin (BSA) standard. Recombinant green fluorescence protein (rGFP; Sigma-Aldrich Company Ltd, Dorset, UK) or recombinant mCherry fluorescent protein (rmCherry; abcam, Cambridge, UK) was used to prepare protein standards for quantification of immunoblots. 10-20 µg of each protein sample was added to LDS loading buffer (Invitrogen) containing 50 mM DTT before heating to 70 °C for 10 min. Electrophoresis was carried out on 4-12% (w/v) gradient Bis-tris NuPAGE gels in MES buffer

(Invitrogen) in a Novex XCell SureLock Mini Cell (Invitrogen) for 35 min at 200 V. An rGFP standard was loaded at 20, 40, 80,160 and 320 ng per well.

Blotting was performed as described previously (Berepiki *et al*., 2016). Membranes were incubated first with an anti-GFP antibody (N-terminal antibody produced in rabbit; Sigma-Aldrich Company Ltd) or anti-mCherry (N-terminal antibody produced in rabbit; abcam) diluted to 1:1,000 in blocking solution and then with goat anti-rabbit IgG (Agrisera) diluted to 1:1,000 in blocking solution. Blots were then incubated with SuperSignal West Dura reagents (Thermo Scientific UK Ltd) and imaged using a C-DiGit Imaging system (LI-COR Biosciences, Nebraska, US). For quantification of immunoblots, images were analysed using ImageJ software (National Institute of Health, Maryland, US).

2.3.7 Gel electrophoresis

Nucleotide sequences were routinely visualized using gel electrophoresis. 1% (w/v) agarose gels were made with 0.3 g of molecular grade agarose, 30 mL of 1x tris-acetate-EDTA (TAE) buffer and 3 µL of SYBR Safe gel stain (Thermo Scientific UK Ltd). 1x TAE was used for the running buffer. 5 µL of HyperLadder I (Bioline, London, UK) was used as the molecular weight marker (MWM). 5 µL samples were added to each well, prepared by mixing 5 µL of PCR product solution with 1 µL of loading dye. Electrophoresis was carried out at 100V for 40 min before gels were visualized by UV-transillumination.

2.3.8 Molecular cloning and vector assembly

2.3.8.1 *In vivo* vector assembly

Unless otherwise stated, vectors were assembled *in vivo* in yeast. Frozen competent yeast cells (strain FY834; a gift from Adokiye Berepiki, Manchester Institute of Biotechnology, University of Manchester, Manchester, UK) were prepared and subsequently co-transformed with vector elements via heat shock, following the methods of Knop (Knop *et al*., 1999). Primers for amplification of genetic elements contained 30 bp extensions to adjacent elements or to the vector backbone sequence. The regions of homology, conferred by the extensions, undergo homologous recombination to form the final vector upon introduction into yeast (see Figure 2.8 for a basic example).

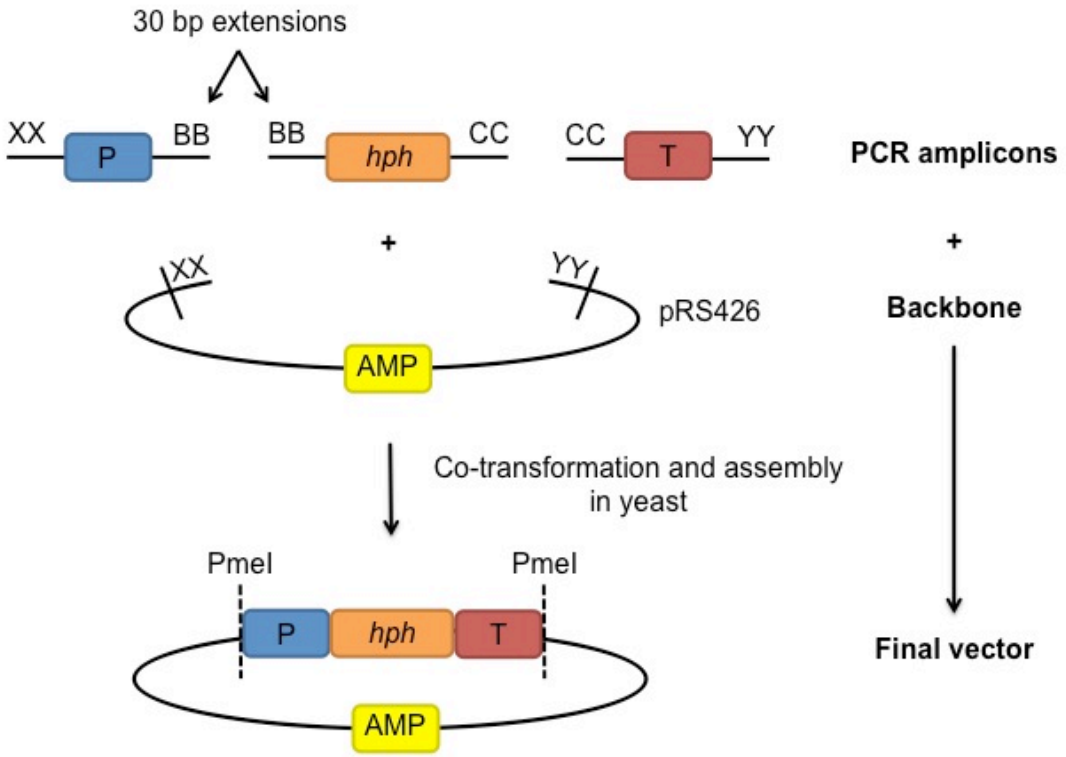


Figure 2.8: Cloning and assembly of basic vector by homologous recombination in yeast. Vector contains a basic resistance cassette containing the *hph* resistance gene conferring resistance to hygromycin, transcription is driven by a promoter (P) and terminated by and a terminator (T). Primers for vector elements contain 30 bp extensions to adjacent PCR amplicons or backbone sequence after digestion with PmeI. Following co-transformation, assembly and amplification in *E. coli*, the resistance cassette can be released from the vector, by digestion with PmeI, for transformation of eukaryotic algae and selection on hygromycin.

2.3.8.1.1 Preparation of frozen competent yeast

Yeast (strain FY834) from a fresh plate was inoculated into 10 ml of YPD (Sigma-Aldrich Company Ltd) and cultured at 30 °C overnight. Two flasks each containing 300 ml of YPD were then inoculated using 5 mL of the overnight culture and cultured again at 30 °C until a density of OD_{600nm} 0.6 – 08 (approximately 2 x 10^7 cells/mL). The cultures were then dispensed into 50 mL Falcon tubes (Thermo Scientific UK Ltd) and harvested by centrifugation at 2,000 x *g* for 5 minutes at room temperature. Harvested yeast cells were pooled into two 50 mL Falcons and washed once with 50 mL of sterile water (room temperature) and then centrifuged again before removal of supernatant and another wash step using 40 mL of room temperature SORB (100 mM LiAc dehydrate, 10 mM Tris–HCl pH 8, 1 mM EDTA/NaOH pH 8, 1 M sorbitol; Sigma-Aldrich Company Ltd). Cells were then re-suspended in a total volume of 360 µL SORB and 40 µL of carrier SS (salmon sperm) DNA (denatured at 95 °C for 10 min and cooled on ice;

Thermo Scientific Uk Ltd) at a concentration of 10 mg/mL, per 50 mL of starting culture (2.16 mL of SORB and 240 µL SS DNA was added to each 50 mL Falcon). 50 µL aliquots of cells were dispensed into 1.5 mL Eppendorf tubes (Thermo Scientific UK Ltd) at room temperature and then transferred to a -80 °C freezer (no shock freezing).

2.3.8.1.2 Transformation of frozen competent yeast

Frozen cells were allowed to thaw before addition and mixing of 10 µL of DNA containing, 1 µL of the vector backbone, typically pRS426 (Fungal Genetics Stock Centre, School of Biological Sciences, Missouri, US), linearized with the restriction PmeI, and 9 µL of PCR amplified elements (divided volumetrically by the number of elements). The pRS426 backbone confers resistance to ampicillin for selection of the assembled vector in yeast and *E. coli.*. To determine the efficiency of DNA assembly, no DNA negative controls were prepared using 10 µL of sterile ddH_2O. 360 µL of PEG/LiAC solution (100 mM LiAc dehydrate, 10 mM Tris-HCl pH 8, 1mM EDTA, 20 g PEG3350; Sigma-Aldrich Company Ltd) was added and reaction mixes were vortexed briefly before incubation at room temperature for approximately 30 minutes. 47 µL of DMSO (Thermo Scientific UK Ltd) was added (to give a final DMSO concentration of approximately 10%), mixes were vortexed again and placed in a water bath at 42 °C for 15 minutes. 1 mL of sterile ddH_2O was added to cells before centrifugation at 12,000 x *g* for 15 seconds. 1.3 mL of the supernatant was removed before re-suspending cells in the remaining ~150 µL. 20% of the total volume was plated onto SC-ura (Sigma-Aldrich Company Ltd) agar plates containing the respective antibiotic (100 µg/µL of ampicillin for pRS426 derived vectors) to confirm correct vector assembly; the remaining 80% was used to inoculate 20 mL liquid SC-ura cultures containing respective antibiotic (100 µg/µL of ampicillin for pRS426-derived vectors) for plasmid rescue. Plates and flasks were incubated at 30 °C for 2-3 days (180 rpm shaking for flasks). The appearance of 5-10 times more colonies on the transformation plate relative to the no DNA control verified correct assembly of vectors.

2.3.8.1.3 Plasmid rescue from yeast

Plasmid DNA (pDNA) was rescued from yeast following the methods of Robzyk and Kassir (Robzyk and Kassir, 1992). 20 mL of yeast culture was transferred to a 50 mL Falcon tube and centrifuged at 2,000 x *g* for 5 minutes at room temperature. The pellet was re-suspended in 1 mL of sterile ddH_2O and transferred to a 2 mL centrifuge tube. Cells were centrifuged at 12,000 x *g* for 15 seconds, the supernatant was then removed. The pellet was re-suspended in 100 µL of STET (8% sucrose, 50 mM Tris pH 8, 50 mM EDTA, 5% Triton X-100; Sigma-Aldrich Company Ltd) and approximately 0.3 g of 0.45 mm glass beads (Sigma-Aldrich Company Ltd)

were added before lysing cells in a tissue lyser twice for 30 seconds at full speed with a 30 second gap between cycles. Another 100 µL of STET was added before vortexing briefly and incubation in a heat block at 95 °C for 3 minutes. Cells were then cooled briefly on ice and centrifuged at 13,000 x *g* for 10 minutes. 100 µL of the supernatant was transferred to a fresh tube containing 50 µL of 7.5 M ammonium acetate (Thermo Scientific UK Ltd) and incubated at -20 °C for 1 hour before centrifugation at 13,000 x *g* for 10 minutes. 100 µL of the supernatant was added to 200 µL of ethanol and 10 µL 3 M sodium acetate. The DNA was then pelleted by centrifugation at 16,000 x *g* for 20 minutes. The supernatant was removed and 200 µL 70% of ethanol was added to the pellet before centrifugation at 16,000 x *g* for 5 minutes. The supernatant was removed and 20 µL of water or EB (10 mM Tris-Cl, pH 8.5) was added to resuspend the DNA once dry.

2.3.8.2 Transformation of *E. coli* and confirmation of correct vector assembly

Electroporation (BIO-RAD MicroPulser electroporator; BIO-RAD Laboratories Company, California, US) of *E. coli* (ATCC 77107) was carried out at 1.8 kV (Sando *et al*., 2007; Lessard, 2013) using 2 µL of pDNA rescued from yeast. *E. coli* transformants were plated on LB agar plates containing the respective antibiotic (100 µg/µL of ampicillin for pRS426 derived vectors). 4-6 of the resultant colonies were picked and grown in 2 mL of liquid LB media with the respective antibiotic (100 µg/µL of ampicillin for pRS426 derived vectors) overnight (37 °C; shaken at 200 rpm). The pDNA of these cultures was then extracted using a plasmid miniprep kit (Zyppy Reasearch Corp, California, US), leaving ~200 µL of culture for subsequent inoculations of cultures harbouring the correctly assembled vector, and digested with a pair of restriction enzymes, giving a distinct pattern of fragment sizes, before visualisation by gel electrophoresis and UV-transillumination to confirm correct assembly see Figure 2.9.

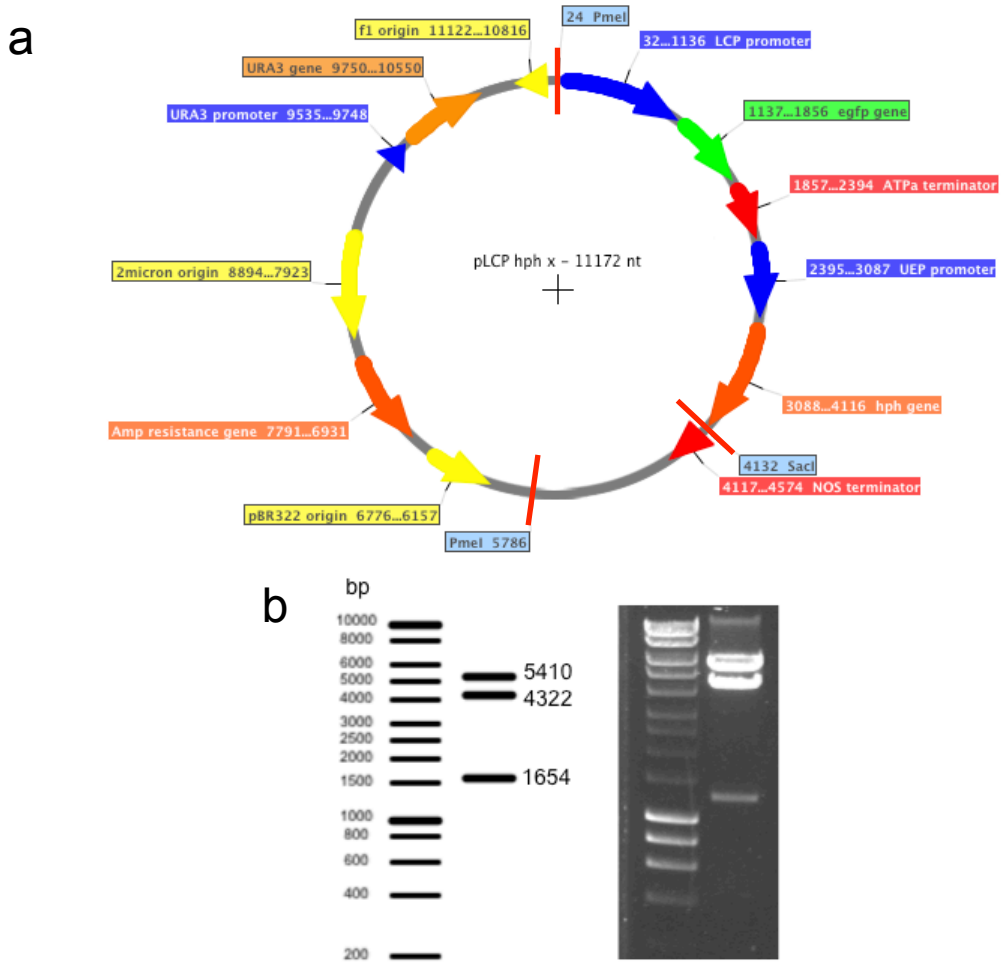


Figure 2.9: Restriction digest check of vectors assembled by homologous recombination**. (a)** Example vector map of pLCP showing PmeI and SacI restriction sites used for digestion confirmation of assembly, generating 3 fragments, 5410 bp, 4322 bp, and 1654 bp in length and confirmation by gel electrophoresis. **(b)** Confirmation of correct pLCP vector assembly by presence of expected fragmentation pattern visualized by gel electrophoresis.

2.3.8.3 Amplification and long-term storage of vectors in *E. coli*

20 mL cultures of an *E. coli* transformants containing the correctly assembled vectors were inoculated in 20 mL of liquid LB and respective antibiotic to maintain selection (100 µg/µL of ampicillin for pRS426 derived vectors) and incubated overnight (37 °C; shaken at 200 rpm). The next day, 600 µL of this culture was taken and mixed by pipetting with 400 µL of glycerol in a 1.5 mL microcentrifuge tube and cryopreserved at - 80 °C for long-term storage. The rest of the culture was used for purification of pDNA using a miniprep kit (Zyppy Reasearch Corp) before restriction enzyme digestion (using PmeI with pRS426-derived vectors) to linearize DNA and release resistance cassette and flanking DNA. Digested pDNA was then concentrated

to around ~500 µg/µL in ddH_2O using a gel extract kit (Qiagen), for transformations of eukaryotic algae by electroporation.

2.3.8.4 Assembly of p35BT vector by homologous recombination

A 35S promoter was amplified from pGWB402 using the primers 35S fw XX ext and 35S rv *ble* ext. A *ble* selection marker conferring resistance to zeocin was amplified from pP66 using primers *ble* fw and *ble* rv. A NOS terminator was amplified from pGWB402 using primers T_{NOS} fw *ble* ext and T_{NOS} rv YY ext. The amplicons generated have 30 bp extensions at the 5′ and 3′ ends that direct recombination with adjacent amplicons upon co-transformation into yeast. Amplicons were co-transformed into yeast along with the acceptor vector pRS426 (Fungal Genetics Stock Centre, School of Biological Sciences, Missouri, US) following linearization by PCR with primer pair YY fw PmeI ext and XX rv PmeI ext, for assembly via the endogenous recombination system of yeast. The assembled plasmids were transferred from yeast to *E. coli* and, following restriction digest screening the cassette was released from the backbone by digestion with PmeI before transformation into *D. tertiolecta*. See Table 2.8 for primer sequences, Figure 6.7 for vector map and appendix section A1 for complete Genbank format vector sequence for p35BT. The pP66 plasmid was provided by Adokiye Berepiki (Manchester Institute of Biotechnology, University of Manchester, Manchester, UK). The pGWB402 plasmid was provided by Matthew Terry (Life Sciences Building 85, University of Southampton, Southampton, UK).

2.3.8.5 Assembly of pNR and pLCP eGFP expression vectors by homologous recombination

To obtain the native *N. gaditana* promoter sequence of the nitrate reductase gene, a region of ~1.3 kb upstream of the gene (Naga_100699g1) was amplified primers P_{NR} fw XX ext and P_{NR} rv *egfp* ext. A region of ~1.1 kb upstream of the native lipocalin protein-coding gene (Naga_100131g14) was amplified using primers P_{LCP} fw XX ext and P_{LCP} rv eGFP ext. The eGFP sequence was amplified from pGAP-α-GFP (unpublished) using the *egfp* primer pair. The native ATP α subunit terminator was generated using primers $T_{ATPα}$ fw *egfp* ext and $T_{ATPα}$ rv CC ext. The UEP promoter (Radakovits *et al*. 2012) was amplified using primers P_{UEP} fw CC ext and P_{UEP} rv *hph* ext. The *hph* (hygromycin B) selection marker was amplified from pZC1 using the *hph* primer pair. The NOS terminator was amplified from the pGWB402 vector (Addgene, Massachusetts, US) using primers T_{NOS} fw hph ext and T_{NOS} rv YY ext. The amplicons generated have 30 bp extensions at the 5′ and 3′ ends that direct recombination with adjacent amplicons upon co-transformation into yeast. Amplicons were co-transformed into yeast along

with the acceptor vector pRS426 (Fungal Genetics Stock Centre, School of Biological Sciences, Missouri, US) following linearization by PCR with primer pair YY fw PmeI ext and XX rv PmeI ext, for assembly via the endogenous recombination system of yeast. The assembled plasmids were transferred from yeast to *E. coli* and, following restriction digest screening the cassette was released from the backbone by digestion with PmeI before transformation into *N. gaditana*. See Table 2.8 for primer sequences, Figure 3.3 for vector map of pNR, and appendix sections A2 and A3 for complete Genbank format vector sequences for pLCP and pNR, respectively.

2.3.8.6 Assembly of pLDH, pHYP1, pCC7 and pHYP2 eGFP expression vectors by homologous recombination

To construct the pLDH, pHYP1, pCC7 and pHYP2 eGFP expression vectors, the pLCP vector (see previous section) was used as an acceptor vector, keeping all of the pre-existing elements but switching the P_{LCP} promoter via homologous recombination. A region of ~1 kb upstream of the following genes was amplified to obtain the native *N. gaditana* promoter sequences: NGA_0640800, P_{LDH}; NGA_2110200, P_{HYP1}; NGA_2007210, P_{CC7}; NGA_2078400, P_{HYP2}. P_{LDH} was amplified with primers P_{LDH} fw XX ext and P_{LDH} *egfp* ext. P_{HYP1} was amplified with primers P_{HYP1} fw XX ext and P_{HYP1} *egfp* ext. P_{CC7} was amplified with primers P_{CC7} fw XX ext and P_{CC7} *egfp* ext. P_{HYP2} was amplified with primers P_{HYP2} fw XX ext and P_{HYP2} *egfp* ext. The amplicons generated have 30 bp extensions at the 5′ and 3′ ends that directed recombination with the acceptor vector. pLCP was used as the acceptor vector after linearization with primers pLCP XX rv and pLCP *egfp* fw. The promoter amplicons were co-transformed into yeast along with the acceptor vector for assembly via the endogenous recombination system of yeast. The assembled plasmids were transferred from yeast to *E. coli* and, following restriction digest screening the cassette was released from the backbone by digestion with PmeI before transformation into *N. gaditana*. See Table 2.8 for primer sequences, Figure 4.10a for vector map of pLDH and appendix sections A4, A5, A6, and A7 for complete Genbank format vector sequences for pLDH, pHYP1, pCC7 and pHYP2, respectively.

Table 2.8: Primer and template sequences for PCR amplification of vector elements. Fw, forward primer; rv, reverse primer; extensions capitalized.

Primer	5′ Extension	Sequence	Template
35S fw	XX	GGAAGGGCGATCGGTGCGGGCCGTTTAAACcgttgtaaaacgacggccag	pGWB402
35S rv	*ble*	CACCGGAACGGCACTGGTCAACTTGGCCATtctagagtcccccgtgttctc	pGWB402
ble fw	No extension	atggccaagttgaccagtgc	pP66
ble rv	No extension	tcagtcctgctcctcggc	pP66
T_{NOS} fw	*ble*	GTGCACTTCGTGGCCGAGGAGCAGGACTGAtgataacagcgcttagagctc	pGWB402
T_{NOS} rv	YY	TGGAATTGTGAGCGGATAACAAGTTTAAACcagaaataggcgcgcctaattc	pGWB402
P_{NR} fw	XX	GGAAGGGCGATCGGTGCGGGCCGTTTAAACggcttctcacgtcgttctcatc	*N. gaditana* gDNA
P_{NR} rv	*egfp*	GGTGAACAGCTCCTCGCCCTTGCTCACCATcgaccaccctggagtga	
P_{LCP} fw	XX	GGAAGGGCGATCGGTGCGGGCCGTTTAAACcgcattcaggtctagtaag	*N. gaditana* gDNA
P_{LCP} rv	*egfp*	GGTGAACAGCTCCTCGCCCTTGCTCACCATtattgcttctgttcaaatgt	
eGFP fw	No extension	atggtgagcaagggcgagga	peGFP
eGFP rv	No extension	tcacttgtacagctcgtcca	
$T_{ATP\alpha}$ fw	*gfp*	ACTCTCGGCATGGACGAGCTGTACAAGTGAtagcggcgcttttgacc	*N. gaditana* gDNA
$T_{ATP\alpha}$ rv	CC	TTGTGTCATGAATTAACAGTTAACGAATACggacagggatacctgtgttt	
P_{UEP} fw	CC	GTATTCGTTAACTGTTAATTCATGACACAAcatcctgctgtatgattttggcacaacg	*N. gaditana* gDNA
P_{UEP} rv	*hph*	CGTCGCGGTGAGTTCAGGCTTTTTACCCATagctgctgccccgaccgtatc	
hph fw	No extension	atgggtaaaaagcctgaactcaccg	pZC1
hph rv	No extension	ttattcctttgccctcggacgagtgc	
T_{NOS} fw	*hph*	CCCAGCACTCGTCCGAGGGCAAAGGAATAAtgataacagcgcttagagctc	pGWB402
T_{NOS} rv	YY	TGGAATTGTGAGCGGATAACAAGTTTAAACcagaaataggcgcgcctaattc	
YY fw	PmeI	gtttaaacttgttatccgctcacaattccacac	pRS426
XX rv	PmeI	gtttaaacggcccgcaccgatc	
P_{LDH} fw	XX	GGAAGGGCGATCGGTGCGGGCCGTTTAAACaatagttgattgcaacgactctg	*N. gaditana* gDNA
P_{LDH} rv	*egfp*	GGTGAACAGCTCCTCGCCCTTGCTCACCATgcttttacgctgtcataatcatt	
P_{HYP1} fw	XX	GGAAGGGCGATCGGTGCGGGCCGTTTAAACtcttgctcaagaaagtccaaatt	*N. gaditana* gDNA
P_{HYP1} rv	*egfp*	GGTGAACAGCTCCTCGCCCTTGCTCACCATgacctctatgtagcgagatagg	

Table 2.8 continued

P_{CC7} fw	XX	GGAAGGGCGATCGGTGCGGGCCGTTTAAACaacttttcccaagaaggatggta	*N. gaditana* gDNA
P_{CC7} rv	*egfp*	GGTGAACAGCTCCTCGCCCTTGCTCACCATaagtcgacgctcaaggc	
P_{HYP2} fw	XX	GGAAGGGCGATCGGTGCGGGCCGTTTAAACctttacaatagaatcgatgagaa	*N. gaditana* gDNA
P_{HYP2} rv	*egfp*	GGTGAACAGCTCCTCGCCCTTGCTCACCATgaacagacaaatagtccaac	
pLCP rv	XX	gtttaaacggcccgcaccga	pLCP
pLCP fw	*egfp*	atggtgagcaagggcgagga	

2.3.9 Colony PCRs

Colony PCRs were used to genotype colonies picked from selection plates. Colonies were picked and re-suspended in 50 µL of Triton X-100 lysis buffer (Wan *et al*., 2011). The solution was then incubated at 95 °C for 10 minutes, then cooled to 4 °C in a thermocycler. 1 µL of the supernatant was taken as template for PCRs.

2.4 Transformation procedures

2.4.1 *N. gaditana* electroporation procedure

Methods for transformation largely followed Radakovits (Radakovits *et al*., 2012). 200 mL cultures of *N. gaditana* were grown for transformation in F2N media (Kilian *et al*., 2011) in 75 cm^2 vented Corning flasks (Thermo Scientific UK Ltd) and incubated under 70 µmol s^{-1} m^{-2} fluorescent light on an orbital shaker (150 rpm) at 30 °C to mid-log phase. On the day of transformation, cell counts were taken and 5 x 10^8 cells were harvested per reaction by centrifugation at 3,000 x *g* for 10 minutes (4 °C). Depending on the required volume, cells were either spun down in 500 mL Corning centrifuge tubes (Thermo Scientific UK Ltd), combined, and then spun down again in 50 mL Falcon tubes, or harvested directly in 50 mL Falcon tubes. Cells were then washed three times in 375 nM sorbitol (cooled on ice; Thermo Scientific UK Ltd) and re-suspended to 100 µL per reaction in 375 nM sorbitol. 100 µL aliquots of this suspension were dispensed into 1.5 mL Eppendorf tubes. 5 µg of linearized pDNA was added to each reaction and mixed by gentle pipetting before 5 minutes of incubation on ice. The pAlg.Ble2AChr.SD and pAlg.Hyg.SD vectors were linearized with BamHI, all other vectors where linearized with PmeI. The equivalent volume of water was added to no DNA negative control reactions. Reaction mixtures were then transferred to pre-chilled 2 mm cuvettes and electroporated at 2.4 kV using a BIO-RAD MicroPulser electroporator. Cells were then

immediately recovered in 10 mL of F2N media and incubated overnight in 70 µmol s^{-1} m^{-2} fluorescent light at 24 °C. The following day cells were plated at 5 x 10^7 cells/plate on F2N media selection plates (92 x 21 mm; 1.5% agar, 50% artificial sea water, 2 mM NH_4Cl, 300 µg/mL of hygromycin B or 3 µg/mL of zeocin). Selection plates were incubated under 70 µmoL s^{-1} m^{-2} fluorescent light at 24 °C until transformants appeared (~4 weeks) and were ready for further processing (~5-6 weeks).

2.4.2 *D. tertiolecta Agrobacterium* procedure

All methods and materials used for the *Agrobacterium* transformations were provided by Algenuity (Eden Laboratory, Bedfordshire, United Kingdom).

D. tertiolecta was first re-streaked on DM02 agar plates (1.5% agar; see Table 2.9) for three generations to acclimate to low salt growth (0.2M NaCl); this is necessary for the co-incubation stage with the *Agrobacterium*. 2 x 50 mL cultures of *D. tertiolecta* were then inoculated in DM02 media and incubated underneath red/blue LEDs at 24 °C, shaken at 150 rpm, CO_2 diffused, for 24 hours. 2 x 5 mL cultures of *Agrobactrium* containing either the pCambia.ble.SD, pCambia.Ble2AChr.SD or pCambia.hyg.SD vector were inoculated in YEP (yeast extract peptone) media with 50 µg/mL of streptomycin and 50 µg/mL of kanamycin and incubated at 28 °C, shaken at 100 rpm, for 48 hours. The next day, *D. tertioletca* was harvested and plated at concentrations ranging from 5 x 10^6 and 1 x 10^8 cells/plate on 5 x DM02 plates containing 100 µM acetosyringone and incubated under red/blue LEDs for 24 hours.

For the co-incubation stage, the *Agrobacterium* cultures were harvested and re-suspended in DMO2 media containing 100 µM acetosyringone to an OD_{600} of 0.5. 200 µL of this solution was added to each DM02 *D. tertiolecta* plate. The co-incubation plates where the incubated under red/blue LEDs at 28 °C for 24-48 hours.

In order to harvest the co-incubation plates, a solution containing DM02 media and 200 µg/mL of cefotaxime was prepared; cefotaxime kills the *Agrobacterium* before further processing. 1 mL of this solution was added to the co-incubation plates. Cells were then scraped into the solution irrigating the top of the plate using a sterile L-shaped spreader and pipetted into an Eppendorf tube.

D. tertiolecta was plated in soft-agar following the co-incubation. Cell counts were performed on the solutions harvested from the co-incubation plates before 5 x 150 µL volumes were pipetted onto 5 x 1.5% agar DM02 bottom plates containing 50 µg/mL of zeocin, per co-incubation plate. DM02 soft agar (0.3%) containing 50 µg/mL of zeocin was then added to each plate and mixed with the cell suspension by gentle tilting. The preliminary antibiotic sensitivity

and plating efficiency trials (see section 6.2.1.2 and 6.2.1.3) in soft-agar were prepared in the same way, omitting zeocin in the case of the plating efficiency trials. Post-transformation plates were then incubated in either 200 μE fluorescent light or red/blue LEDs and at 24 or 28 °C.

Table 2.9: Composition of DMxx media. DM02 media contains 0.2 M NaCl; DM04 media contains 0.4 M NaCl. Combine the following ingredients before adding ddH_2O to 1 L before adding 1 g of $NaHCO_3$ L^{-1} to prevent precipitate.

Compound	Molarity	Quantity
NaCl	N/A	11.7 g (0.2 M final) or 23.4 g (0.4 M final)
KCl	2.0 M	5 mL
$MgCl_2 \cdot 6H_2O$	2.0 M	10 mL
$CaCl_2 \cdot 4H_2O$	1.0 M	10 mL
$MgSO_4 \cdot 7H_2O$	2.4 M	10 mL
$NaNO_3$	2.0 M	2.5 mL
Na_2SO_4	1.0 M	24 mL
$NaH_2PO_4 \cdot 2H_2O$	100 mM	1 mL
Fe_2EDTA	1.5 mM	1 mL
H_3BO_3	N/A	11.44 mg
$MnCl_2 \cdot 4H_2O$	N/A	1.385 mg
$ZnCl_2$	N/A	0.109 mg
$CoCl_2 \cdot 6H_2O$	0.002 mM	0.01 mL
$CuCl_2 \cdot 2H_2O$	0.0002 mM	0.001 mL
Tris-HCl pH 7.6	910 mM	22 mL

2.4.3 *D. tertiolecta* biolistics procedure

All methods and materials used for the biolistics transformations were provided by Algenuity (Eden Laboratory, Bedfordshire, United Kingdom).

D. tertiolecta was first re-streaked on DM02 agar plates (1.5% agar; see Table 2.9) for three generations to acclimate to low salt growth (0.2 M NaCl). A 50 mL starter culture of *D. tertiolecta* was then inoculated in DM02 media and incubated underneath red/blue LEDs at 24 °C, shaken at 150 rpm, CO_2 diffused, for 24 hours. This starter culture was harvested by centrifugation (3,000 x *g* for 10 min) at a cell concentration of 1 x 10^7 cells/mL and used to inoculate two 1 L cultures at a cell concentration of 2 x 10^5 cells/mL. After ~48 hours the 1 L

cultures were harvested by centrifugation (3,000 x *g* for 10 min) once a cell concentration of 2 x10^6 cells/mL (early-logarithmic phase) was reached and re-suspended in DM02 media to a cell concentration of 5 x 10^8 cells/mL. 200 µL volumes of this solution were plated on 10x DM02 agar plates with 100 µg/mL of Carbenicillin (92 x 21 mm; 1.5% agar; 1 x10^8 cells/plate) to be used for the biolistics transfromation. 2 of these plates were to be used as no DNA negative controls; 4 were to be used with the pAlg.Ble.SD vector and 4 with the pAlg.Ble2AChr.SD vector.

To prepare the DNA coated beads for transformation, a 50 mg/mL stock of gold particles was sonicated before 60 µL was added to 40 µL of DNA binding buffer (30 mg/ml gold particle final concentration) in 1.5 mL Eppendorf tubes. 6 µg of plasmid DNA (in ~7-10 µL of elution buffer) was added to each solution; pAlg.Ble.SD vector or pAlg.Ble2AChr.SD vector. ~7-10 µL of elution buffer was added to the no DNA negative control solution. The solutions were vortexed, an equal volume of precipitation buffer was added, vortexed again and then left to stand for 3 min. The solutions were centrifuged at full speed for 10 seconds and the supernatant was removed. 500 µL of cold 100% ethanol was added and the solutions were vortexed and centrifuged at full speed for 10 seconds again. The supernatant was removed and the beads were re-suspended in 175 µL of 100% ethanol to give a final gold particle concentration of ~0.5 mg gold per 35 µL.

A Biolistic PDS-1000/He Particle Delivery System (BIO-RAD) was used to deliver the DNA-coated gold particles to the *D. tertiolecta* transformation plates. For the four transformation plates with each vector (pAlg.Ble.SD or pAlg.Ble2AChr.SD), 35 µL of the DNA-coated gold bead solutions were added to the microcarrier of the instrument and delivered to the transformation plates. Two negative control plates were set up using 35 µL of the no DNA gold bead solution. The transformation plates were incubated overnight in low light for recovery.

The following day the transformation plates were irrigated with 1 mL of DM02 media and harvested. For each vector used for transformation, the total harvested volume from two of the initial transformation plates was transferred to two separate bioassay dishes (245 x 245 x 25 mm) DM02 agar plates (1.5% agar) with 50 µg/mL zeocin for selection. The total harvested volume from the other two initial plates was split (50:50) over two selection plates. For the no DNA control initial transformation plates, all of the harvested volume from one plate was transferred to a single selection plate and the other was split 50:50 over two selection plates. Positive controls were also set up in the same conditions but with no antibiotic added to the selection plates using cells harvested from transformation plates (DNA included).

The selection plates were incubated under red/blue LEDs at 24 °C and monitored for growth.

2.4.4 *D. tertiolecta* electroporation procedure

The electroporation procedure was adapted from method previously tested that yielded inconclusive results (Walker *et al*., 2005b). 25 mL starter cultures of *D. tertiolecta* were grown in standard f/2-Si media (Fallis, 1975; Guillard and Ryther, 1962) in 75 cm^2 vented Corning flasks and incubated under 200 µmol photons m^{-2} s^{-1} fluorescent light (12 hr light, 12 hr dark cycle) at 24 °C. After 3 days of growth the starter cultures were inoculated into 2x 200mL f/2-Si in 75 cm^2 vented Corning flasks and grown until a cell density of ~1 x 10^6 cells mL^{-1} was reached (~3 days). The total volume of cells was harvested by centrifugation at 2,500 x *g* for 10 min, washed in 375 mM sorbitol (x3) and re-suspended in a total volume of 10 mL 375 nM sorbitol. 400 µL of this solution was transferred to a 1.5 mL Eppendorf tube and incubated with 8-10 µg of the plasmid DNA (p35BT, linearized with PmeI) on ice for 10 min before transferring to a sterile pre-chilled 2 mm gap cuvette and performing electroporation at 2 kV/cm. No DNA controls were included using reaction mixtures prepared in the same way as tests reactions except that the DNA solution was replaced with the equivalent volume of ddH_2O. Cells were immediately recovered in 25 mL of f/2-Si and grown for a further 3 days. Cells were harvested by centrifugation at 2,500 x *g* and 5 x 10^7 cells/plate were plated on f/2-Si agar plates (1.5% agar) with 50 µg/mL of zeocin. Positive control plates were inoculated using reaction mixtures containing the p35BT cassette on no antibiotic f/2-Si agar plates. Plates were incubated in continuous 200 µmol photons m^{-2} s^{-1} fluorescent light at 24 °C and monitored for appearance of transformants.

2.4.5 *D. tertiolecta* glass beads procedure

The glass beads procedure was adapted from previously described methods established in *D. salina* (Feng *et al*., 2009). 40 mL cultures of *D. tertiolecta* were inoculated in f/2-Si media (see Table 2.2) and grown to the logarithmic phase. Cultures were harvested by centrifugation (3,000 x *g*; 7 min), washed three times in f/2-Si media and then re-suspended to concentration of 10^5 cells/mL (solution A hereafter). For each transformation reaction a separate 200 µL solution (solution B hereafter) was prepared containing a 100 µL solution containing 60 µg of the pAlg.Ble.SD plasmid linearized with BamHI, 100 µL of 35% w/v PEG (Sigma-Aldrich Company Ltd) and 300 mg of 0.45 – 0.52 mm glass beads (Sigma-Aldrich Company Ltd). To sterile 1.5 mL Eppendorf tubes was added 800 µL of solution A and 200 µL of solution B. Tubes were inverted 4-5 times to mix the solutions and then agitated in a vortex mixer for 12 seconds at 2,400 rpm. The beads were allowed to settle before the cells were transferred to sterile test tubes and incubated overnight in dim light at 24 °C. The following day the each test tube was harvested by centrifugation (3,000 x *g*; 7 min), the supernatants were removed, leaving

~100 µL to re-suspend the pelleted cells. Each ~100 µL cell solution was plated on an f/2-Si agar plate (1.5% agar) with 50 µg/mL of zeocin. Negative controls were inoculated using no DNA reaction mixtures on plates containing zeocin; positive controls were inoculated using reaction mixtures containing the pAlg.Ble.SD cassette on agar plates containing zeocin. Plates were incubated in continuous 200 µmol photons m^{-2} s^{-1} fluorescent light at 24 °C and monitored for appearance of transformants. The pAlg.Ble.SD plasmid was a gift from Algenuity (Eden Laboratory, Bedfordshire, United Kingdom).

2.5 Screening procedures

2.5.1 Promoter trap screening procedure

For the gene-trap promoter screen (Chapter 4), a transformation cassette was amplified from the pCambia.Ble2AChr.SD vector (Algenuity) using the forward and reverse primers listed in Table 2.10. This transformation cassette was transformed into *N. gaditana* following the standard transformation procedure (see section 2.4.1).

Table 2.10: Primers used to amplify transformation cassette from pCambia.Ble2ACHr.SD. Fw, forward primer; rv, reverse primer.

Primer	Sequence (5′ to 3′)
Ble2AChr fw	gtcactgtcacaacgtcg
Ble2AChr rv	ccaattcactgttccttgca

After the transformation, 200 µL of F2N media (1.5% agar; 50% artificial sea water; 2 mM NH_4Cl; 300 µg/mL of hygromycin B) was loaded into each well of a sterile Costar 96-well clear microplate (Sigma-Aldrich Company Ltd). 96 colonies were then picked from the transformation plates using sterile p200 pipette tips (Thermo Scientific UK Ltd) and loaded into each well of the microplate.

Five 75 mL F2N agar plates were pre-prepared (1.5% agar; 50% artificial sea water; 2 mM NH_4Cl) in square bioassay dishes (245 x 245 x 25 mm; Thermo Scientific UK Ltd) with the following concentrations of zeocin (µg/mL): 0, 20, 40, 60, 80. A Scienceware 96-well replicator (Sigma-Aldrich Company Ltd) was then sequentially dipped in the microplate and then placed on the surface of each agar plate depositing ~5 µL from each well of the microplate on the agar in 96-well format. The square bioassay dishes were then incubated under continuous 200 µmol photons m^{-2} s^{-1} fluorescent light at 24 °C and monitored for growth.

2.5.2 Gene-trap FTIR screen procedure

For the FTIR screen (Chapter 5), F2N agar plates were pre-prepared (1.5% agar; 50% artificial sea water; 2 mM NH_4Cl; 300 µg/mL) in square bioassay dishes (245 x 245 x 25 mm; Thermo Scientific UK Ltd). 94 boxes (10 x 10 mm) were drawn on the bottom of each dish.

N. gaditana was transformed with the pAlg.Hyg.SD plasmid linearized with BamHI, following the standard transformation procedure (see section 2.4.1). After colonies had appeared they were picked from the initial transformation plate and streaked onto each box of the bioassay dishes, each dish contained 94 mutant cell lines. These dishes were then incubated under 200 µmol photons m^{-2} s^{-1} fluorescent light at 24 °C for twelve days.

After twelve days, the 94 cell lines on each bioassay dish were scraped and added to the FTIR microplate. 30 uL of ddH_2O was added to each well and mixed by pipetting with a multi-channel pipette. The microplate was then incubated at 40 °C for 30 mins before the lipid to amide ratio for each cell line was measured on the FTIR (see section 2.2.7).

2.6 Cryopreservation of *N. gaditana*

For cryopreservation of *N. gaditana* strains, 200 mL cultures were grown in F2N media. Cultures were grown on a shaker at 150 rpm at 24 °C under 100 µmol photons m^{-2} s^{-1} fluorescent light to a cell concentration of 1 x 10^7 cells/mL. Cultures were place in the dark for 20 hours before cryopreservation. A Biocision CoolCell freezing container (Sigma-Aldrich Company Ltd) was equilibrated to 4 °C. A filter sterilized 20% DMSO solution of DMSO in F2N media was prepared. The dark treated cells were harvested by centrifugation (2,500 x *g* for 15 min) in sterile 50 mL Falcon tubes. The cells were re-suspended in F2N media to an end concentration of 1.2 x 10^8 cells/mL. 1 mL of the 20% DMSO solution was dispensed into the ice-cold, pre-labelled 2 mL cryogenic vials (Sigma-Aldrich Company Ltd). In subdued light, 1 mL of the harvested cells were transferred to each cryogenic vials to reach a final DMSO concentration of 10% and a final cell concentration of 6 x 10^7 cells/mL. The vials were gently closed and inverted several times away from bright light. The cryogenic vials were equilibrated for 30 minutes on ice (4 °C) in darkness. The cryogenic vials were transferred to the pre-chilled CoolCell freezing container. All wells of the freezing container should contain a filled vial; if freezing a batch with fewer than 12 vials, each empty well was filled with a BioCision CoolCell Filler Vial (Sigma-Aldrich Company Ltd). After ensuring the vials could slide in and out of the wells freely, the lid of the freezing container was fully seated. The container was immediately transferred to a -80 °C freezer ensuring at least 1 inch of clearance around the module. The

containers were left at -80 °C for 3 hours to reach a temperature less than -50 °C. The freezing container ensures a standardized controlled rate of -1°C/minute cell freezing in a -80 °C freezer. The vials were then transferred to a storage box.

2.6.1 Thawing process

A water bath of ddH_2O was set to 27 °C. A floating vial-holder (Sigma-Aldrich Company Ltd) was placed in the water bath and used to hold the cryogenic vials for thawing. The cryogenic vials were transferred to the water bath. As soon as all of the ice in the cryogenic vials had thawed, each vial was cleaned with ethanol. Within 3 min the vials were transferred to 50 mL of F2N media. The DMSO concentration must be reduced to less than 0.2%. The recovered cultures were allowed to remain in darkness or in subdued light overnight. Cultures were then incubated in normal conditions.

Chapter 3: An inducible expression system for *N. gaditana*

3.1 Introduction

3.1.1 Advances in *Nannochloropsis* species

Recent advances have seen the emergence of *Nannochloropsis* sp. as a platform genus of marine microalgae for biotechnological applications. Reverse genetic approaches using random mutagenesis combined with high-throughput phenotypic screening have isolated *Nannochloropsis* strains with improved light-use efficiency and increased lipid accumulation (Doan and Obbard 2012; Beacham *et al*. 2015; Perin *et al*. 2015). Further, *Nannochloropsis* sp. are haploid and transformation of the nuclear genome with integration constructs has been established and shown to occur by high-efficiency homologous recombination in *Nannochloropsis oceanica*, allowing the generation of knockout mutants (Kilian *et al*. 2011; Radakovits *et al*. 2012). Wang *et al* reported the first use of a CRISPR/Cas9-based approach (see section 1.3.2.2) in a *Nannochloropsis* sp., generating mutant strains of *N. oceanica* with precise five-base deletions in the nitrate reductase gene, demonstrating the potential for sophisticated gene editing techniques (Wang *et al*. 2016).

3.1.2 *N. gaditana* for biotechnology

Within the genus, *N. gaditana* is a particularly promising organism for biofuel production systems due to its ability to accumulate high levels of lipids such as triacylglycerols (TAGs), which can exceed 60% of total biomass on a dry weight basis under nitrogen starvation (Rodolfi *et al*. 2009). A detailed understanding of key metabolic pathways in *N. gaditana* has been assisted by the draft assembly of the nuclear (~27.5 Mbp; 10,646 protein-coding genes) and organellar genomes (Radakovits *et al*. 2012; Jinkerson *et al*. 2013). Codon utilization analysis

of *N. gaditana* has shown no unused codons, an advantageous property for host systems for heterologous protein expression (Jinkerson *et al*. 2013). CRISPR/Cas9 has also been successfully implemented in *N. gaditana* to knock out 18 transcription factors involved in negative lipid regulation under nitrogen starvation (Ajjawi *et al*. 2017). Attenuation of one of the identified genes yielded strains with a 2-fold increase in lipid content and little effect on growth, which has previously not been possible in mutant strains with increased lipid accumulation (Ajjawi *et al*. 2017). These advances have resulted in the emergence of *N. gaditana* as a candidate platform species to enable further biotechnological development.

3.1.3 Inducible expression systems

The availability of a wide selection of endogenous promoters with well-characterized and novel expression traits is a key requirement for the development of a species from both a fundamental research and an industrial perspective (Doron *et al*. 2016). Endogenous promoters have been shown to be more effective for generating stable transformants and driving expression of heterologous genes (Walker *et al*. 2005). To our knowledge, only seven endogenous promoters have been used to successfully drive expression of heterologous genes in *N. gaditana* (see Table 3.1) (Radakovits *et al*. 2012; Ajjawi *et al*. 2017). Furthermore, these promoters were often uncharacterized or characterized based solely on their ability to generate transformants with no detailed analysis carried out.

Table 3.1: Endogenous promoters characterized in *N. gaditana.*

Gene of promoter	Method of characterization	Reference
β-tubulin	Transformation efficiency	(Radakovits *et al*., 2012)
Ubiquitin extension protein	Transformation efficiency	(Radakovits *et al*., 2012)
Heat shock protein 70	Transformation efficiency	(Radakovits *et al*., 2012)
TCT (full name not given)	Uncharacterized	(Ajjawi *et al*., 2017)
RPL24 (full name not given)	Uncharacterized	(Ajjawi *et al*., 2017)
4AIII (full name not given)	Uncharacterized	(Ajjawi *et al*., 2017)
ETF3 (full name not given)	Uncharacterized	(Ajjawi *et al*., 2017)

In particular, there is a requirement for inducible promoters that can be switched on and off, without detrimental effects on growth, to enable temporally controlled expression of toxic molecules such as Cas9 or for use in silencing systems such as RNAi (Jiang *et al*. 2014; Doron *et al*. 2016). This chapter describes the development of an inducible expression system for *N. gaditana*.

Various inducible expression systems have been developed in algae; however, many rely on induction methods that may have limitations in a photoautotrophic species. For example, the hsp70A promoter has been used in green algae to regulate heat-induced expression; however, the use of heat as a method of induction may deem this promoter unsuitable as a molecular switch for *N. gaditana* due to the possibility of detrimental effects on growth caused by changes in temperature (Schroda *et al*. 2000; Jakobiak *et al*. 2004). Light-based induction systems have also been explored in algae, initially by use of the CABII-I promoter in *Chlamydomonas*, but pose similar concerns regarding the effects of the method of induction on growth (Blankenship and Kindle 1992; Park *et al*. 2013; Baek *et al*. 2016).

Expression systems controlled by nutrients or chemicals that can be regulated in the growth media were therefore considered. Algae (including *Nannochloropsis* sp.) are able to utilise nitrate and subsequently convert it into ammonium through use of the enzyme nitrate reductase (NR) (Fernández *et al*. 1989; Berges 1997). Micromolar concentrations of nitrate elicit a rapid induction of expression of the nitrate reductase gene (Raven *et al*. 1992; Forde 2000; Galván and Fernández 2001; Llamas *et al*. 2002; Fernández and Galván 2008). However, the energy cost of direct ammonium assimilation is lower than that of nitrate (Bloom *et al*. 1992). Consequently, the presence of ammonium provides a negative feedback effect on nitrate assimilation and supresses the nitrate reductase gene (see Figure 3.1) (Ohresser *et al*. 1997; Loppes *et al*. 1999; Llamas *et al*. 2002; Fernández and Galván 2008). Hence, expression of the *NR* gene is induced in the presence of nitrate ions and repressed in the presence of ammonium ions in a range of algae, including *Cylindrotheca fusiformis*, *C. reinhardtii*, *C. ellipsoidea*, *D. salina*, *V. carteri*, and *Phaeodactilum triornutum* (Wang *et al*. 2004; Poulsen and Kröger 2005; Li *et al*. 2007; Schmollinger *et al*. 2010; Niu *et al*. 2012; von der Heyde *et al*. 2015).

The promoter of the nitrate reductase gene (*NR*) has been used to regulate inducible expression in a range of applications in several species of algae including, *C. reinhardtii* (Schmollinger *et al*. 2010), *C. ellipsoidea* (Wang *et al*. 2004), *D. salina* (Li *et al*. 2008; Li *et al*. 2007) and *P. tricornutum* (Niu *et al*. 2012). Schmollinger *et al*. used the *NR* promoter to regulate an RNA silencing system (see section 1.3.2.3) and saw reductions in the level of the targeted transcript approx. 2 hours after transfer from ammonium to nitrate-containing media (Schmollinger *et al*., 2010). Li *et al*. and Niu *et al*. used the *NR* promoter to drive expression of a resistance marker and found that transformed cell lines lost resistance when transferred from nitrate to ammonium (Li *et al*., 2007; Niu *et al*., 2012). Wang *et al*. reported the expression of the *NR* promoter using a GUS stain reporter gene and found that the expression pattern of the GUS gene was consistent with the transcriptional pattern of the *NR* gene, demonstrating that regulation of *NR* is mainly under transcriptional control (Wang *et al*., 2004).

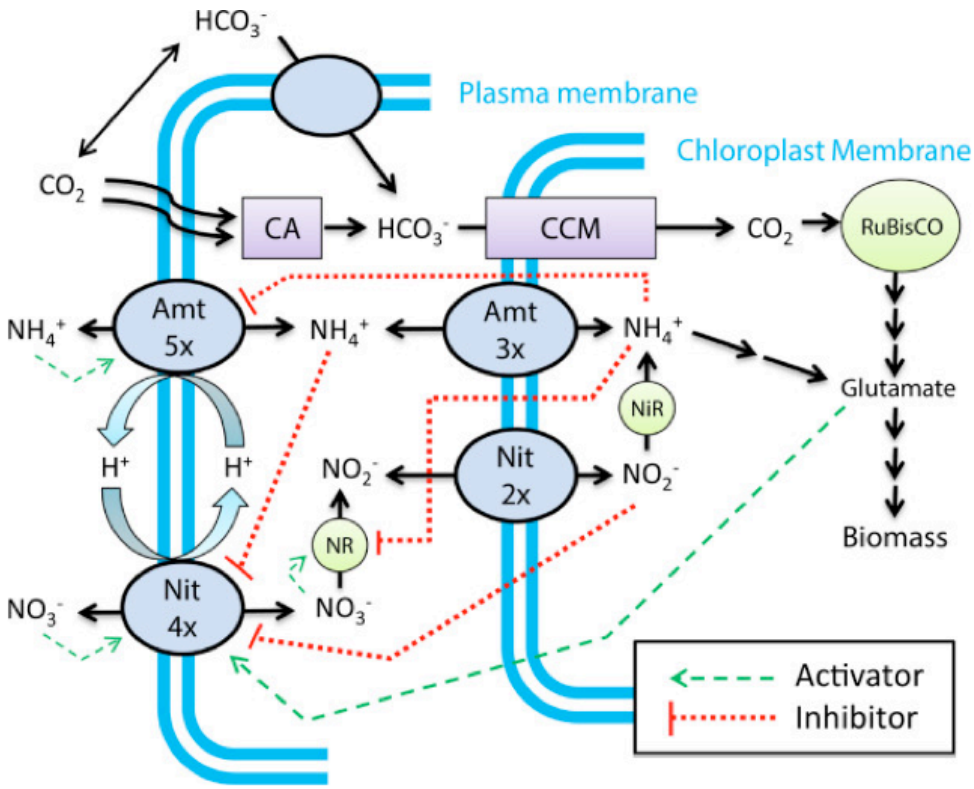


Figure 3.1: Schematic of cellular regulation of nitrogen assimilation in the model eukaryotic alga *C. reinhardtii*. Ammonium (NH_4^+) is assimilated into glutamate and subsequently biomass without prior reductions. Conversely, nitrate (NO_3^-) must be reduced firstly by nitrate reductase (NR), to form nitrite (NO_2^-), and then again by nitrite reductase (NiR) to form ammonium. Thus the cost of nitrate assimilation is greater than direct assimilation of ammonium; leading to suppression of expression of NR in the presence of NH_4^+ via the regulatory elements represented as dashed lines. Nit, nitrate transporter; CCM, carbon concentrating mechanisms; CA, carbonic anhydrase. Figure from Scherholz and Curtis (2013).

Using enhanced green fluorescent protein (eGFP) as a fluorescence reporter and immunoblot analysis, the ability of the endogenous *N. gaditana* nitrate reductase promoter to drive inducible expression of heterologous proteins in transgeneic cell lines was tested. In addition, the temporal regulation characteristics of the nitrate reductase promoter over a range of nitrogen conditions were explored to help define the biotechnological applicability of this promoter.

3.2 Results

3.2.1 Construction of the nitrate reductase expression vector

The efficiency of gene expression driven by the nitrate reductase promoter sequence (P_{NR}) of *N. gaditana* was analysed using the eGFP (enhanced green fluorescent protein) expression vector pNR. eGFP is a variant of GFP which contains chromophore mutations that make the protein up

to 35 times brighter than wild type GFP (Zhang *et al*., 1996). The promoter was defined as the region ~1.3 kb upstream of the nitrate reductase gene (*NR*; Naga_100699g1) from the *N. gaditana* B31 genome assembly (assembly accession GCA_000569095.1; Corteggiani Carpinelli et al., 2014). A length of ~1.3 kb was chosen in order to ensure all of the regulatory elements within the promoter sequence were included. Promoter analysis with TSSPlant (Shahmuradov et al., 2017) suggested a putative transcription start site at -633 bp relative to the *NR* start codon (although this has not been confirmed experimentally) with a TATA box at -681 bp. Analysis with PLACE and PlantCARE (Higo et al., 1999; Lescot, 2002) identified several putative CAAT box motifs upstream of the TATA box (see Figure 3.2 for an annotated sequence of the *NR* promoter). Further motif analysis of this promoter region of the *N. gaditana* nitrate reductase gene was performed against *NR* genes from *Chlamydomonas reinhardtii* (XP_001696697.1)*, Thalassiosira pseudonana* (XP_002294410.1)*, Micromonas pusilla* (XP_003058321.1)*, Chondrus crispus* (XP_005714489.1)*, Chlorella variabilis* (XP_005844793.1) using the MEME search tool (Multiple EM for Motif Elicitation; Bailey et al., 2009). No significantly conserved motifs between the promoter sequences were detected. This is not surprising given the low identity between the *NR* genes of these respective species (40 - 50%). Algal *NR* genes are diverse in sequence and have a larger numbers of introns than higher plants (Song and Ward, 2004). The promoter of the *NR* gene in *N. oceanica* is located between a divergent gene pair and may therefore by bidirectional; however, this is not the case in *N. gaditana* (Poliner et al., 2018). This analysis highlights the diversity in the promoter regions of *NR* genes, which warrants further investigation.

> P_{NR}, 1 – 1319 bp
GGCTTCTCACGTCGTTCTCATCGTCTGCTTTCTCCCCCTTTCCTTTTCTTCCGTCGTTTCTCCCACAGAA
CAATTACGGGAGATGCTGGATTTCTATGTGGAGGCTGACGTTTTCCGGGTCAGCGCCGATCGGACGATGA
TTGAACTCTTCAACACGGATTGAGAGAGCACTGTACGCAGTGCAGAGCGTGGTAGGACATAGAATCGCGC
AGGGAAATAGGTTGGAAGGGGAG**CAATT**CGCAGAAGGAAAGCTCAGCGATCGAGAAAAACATGGAGGGTG
AAGATAAGCTCGCTCTGAAAGTAAATCTGA**CAAAT**CAAGCACAGATGAGGTCACTGCGAAGTTGACCTGA
AAGCGTGTTTTTT**CAAAT**CGTCGCCCAGGCGAAGTGTTATTATTTCTTGCACATCTTGAAGATAGTCTATA
CGCAGGATGAAAAGGAGGATCAACTTTGCCATTTCTGCATACGTCATAATTCCACGATGACTAG**CAAT**GT
GGAATAGATCTTCCTTTTGAATTTCTCTGAATGGTTAGCACTTTCCTTTGTTTTTTTGTCCTATGAAATT
TAAATGGTAATTGATACCTCCACTAAGGAGT**CAAATCAAT**AGCTTTGATCACAAGCAGAAGAAATGATAA
GAGCGCAAAGCTTGGAATAATCTCTT**TATA**TGTCCCATTGATTAACATCTATCCTACAAGGGCGTGAACA
ACCACGATCAGAGTTTAGGAGAGCGATCAACTCCAAGCACTAAAAGGGCGTGAACAACCACTATAAGAAT
TGAAGAGGCGATCCTCCAAGAACGCTGCCCGACCAGGCATTAATTGCCGCATAATCTGACCGTCGTCTGG
ATAGCCTGGCAGACGGGAAAACTTCATGGAATAATGACGCCATTACTTGAAGTTTCGCGGGGCTTTTTTT
CCTGGCTCGGAAT**TATA**CTTTTTACGCCATAACTGCGTTTCCCCGTTCTCCAAGGGCTCCGCCGTCGGCA
ACTTTGTATGGGGAAGCGCGGTAGATTTTGTTTTCAAGGAAGGACCAAGGGTGTGAGGTGGCTTTTGAAA
CACTGGTGCCATGTATGGGACTATTCTGACGGATCTCCGCCGCGCAACTTAACTCCCGCCTCGGATTCCG
TGTTAGGCGTCGTAACGAACATTAGGAAGACATCGGATCTTAGATGAGAAATTATGCATCATGCGACTTG
TGTGAATGTGCATCAATCTTTTGCCTGTTGCATTACGCATGCTCACGCGACTTCATGAATATGTCTTCCC
TTACATACTCCTCCTGATCACGACAACACAACATTTCTCACATCACTCCAGGGTGGTCG

Figure 3.2: Annotated sequence of *N. gaditana* NR gene. Putative CAAT boxes (bold and underlined text), TATA box (bold, underlined and highlighted blue), and possible alternative TATA box (highlighted yellow) were identified with PLACE and PlantCARE. A putative transcription start site (highlighted green) was identified with TSSPlant, although this was not confirmed experimentally.

This *N. gaditana NR* sequence was incorporated into the pNR vector (Figure 3.3) at the 5′ end of an eGFP coding sequence (*egfp*) that was terminated by a ~0.5 kb terminator region ($T_{ATPα}$) of the endogenous ATP α subunit (NGA_2121710). The hygromycin phosphotransferase resistance gene (*hph*) from *E. coli*, conferring resistance to hygromycin B was used as a selectable marker for transformation. From 5′ to 3′, the resistance cassette of the pNR vector consisted of an endogenous ubiquitin extension protein promoter (P_{UEP}) (Radakovits *et al*. 2012), the hph gene, and an *Agrobacterium tumefaciens* nopaline synthase terminator (T_{NOS}). Two PmeI restriction sites flanking the expression cassette were used to release the cassette from the vector backbone and linearize the DNA for transformation of *N. gaditana*. In addition, a negative control expression cassette (pLCP) was constructed that was identical to pNR except that the nitrate reductase promoter was replaced with a constitutive promoter of an endogenous gene coding for a lipocalin protein (PLCP; Nga01014). See section 2.3.8 and 2.3.8.5 for details of vector construction and Appendix A for Genbank format vector maps for pNR and pLCP.

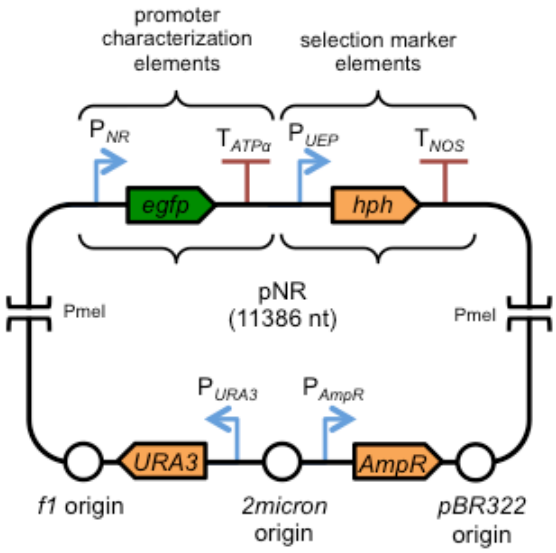


Figure 3.3: Vector used for characterization of the nitrate reductase promoter. The vector contained promoter-characterizing elements upstream of a *hph* resistance cassette, conferring resistance to hygromycin B. P_{NR}, nitrate reductase promoter: *egfp*, enhanced green fluorescent protein gene: $T_{ATPα}$, ATP α subunit terminator: *hph*, hygromycin phosphotransferase gene; T_{NOS}, Agrobacterium tumefaciens nopaline synthase terminator. Vector was linearized with PmeI before transformation of *N. gaditana* to improve transformation efficiency and remove accessory elements for vector preparation in *S. cerevisiae* and *E. coli*. Figure was drawn to SBOL standards (Galdzicki *et al*. 2014).

3.2.2 Generation of transgenic cell lines

N. gaditana was grown to the logarithmic phase in liquid media (see section 2.2.1.1) before being harvested for transformation by electroporation (see section 2.4.1). After selection on

hygromycin B, twenty transformants were recovered (NgNR1-20). These cell lines were sub-cultured once in liquid media containing hygromycin B to confirm antibiotic resistance and to remove any wild type cells (Figure 3.4a). They were subsequently maintained in standard media and on agar plates without antibiotic selection. Stable integration of the expression cassette into the *N. gaditana* genome in the NgNR cell lines was shown by PCR confirmation of the presence of the *hph* gene (Figure 3.4b).

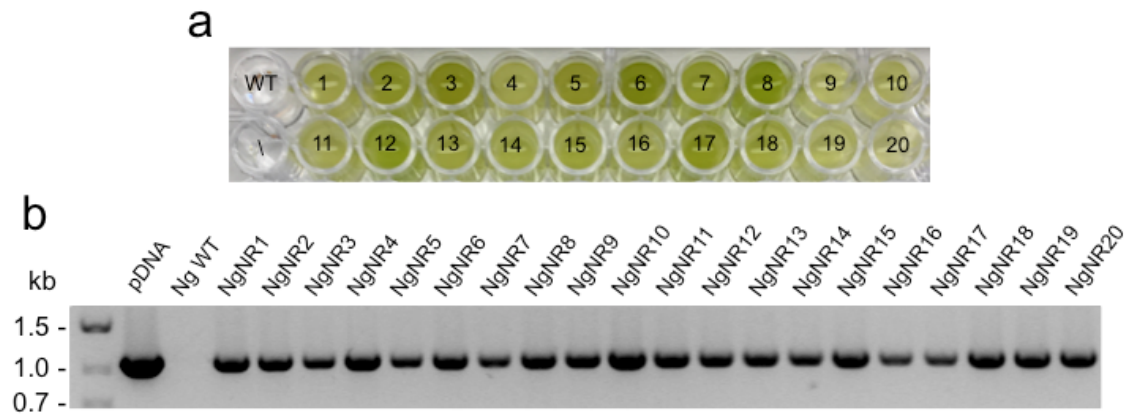


Figure 3.4: Confirmation of vector integration into 20 transgenic *N. gaditana* cell lines. (**a**) The recovered transformants (NgNR1-20) were sub-cultured in liquid media containing hygromycin B once to confirm resistance. WT, *N. gaditana* wild type; NgNR strains are numbered from 1 to 20. (**b**) Genotyping by PCR using primers binding to the 5′ and 3′ end of the *hph* gene, producing a product of 1029 bp in length. Template DNA: DNA ladder, lane 1; pNR vector (positive control), lane 2; *N. gaditana* wild type (negative control), lane 3; NgNR strains 1 to 20, lanes 3 to 23.

3.2.3 eGFP fluorescence analysis of transgenic cell lines

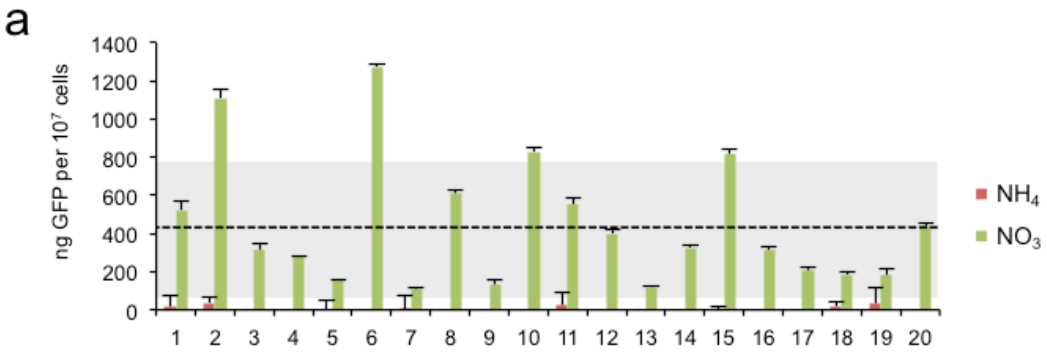
The integration sites of chimeric DNA following electroporation-based transformation are largely randomly distributed within the genome and positional effects can alter transgene expression levels (Zhang *et al*., 2014; Chen and Zhang, 2016; Thompson and Gasson, 2001). Accordingly, the eGFP expression levels of all twenty of the pNR transformants were analysed to determine an average expression level for the nitrate reductase promoter. NgNR1-20 and *N. gaditana* wild type cultures were grown to the logarithmic phase (see section 2.2.1.1) in media containing either 5 mM sodium nitrate or 5 mM ammonium chloride as the sole nitrogen source to induce or repress expression, respectively.

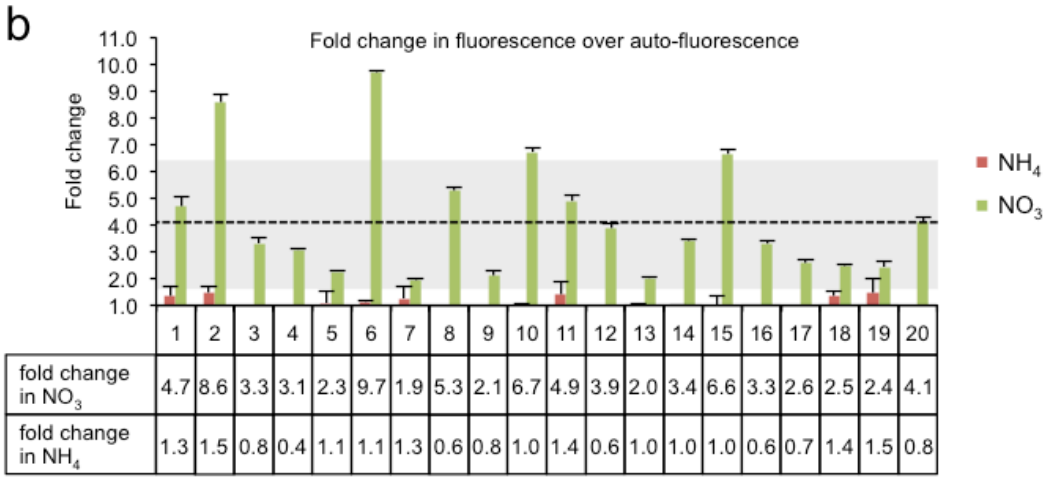
eGFP fluorescence was analysed as described previously; using a series of recombinant standards of the fluorescent protein (Rasala *et al*., 2013). Fluorescence from eGFP was measured in a series of recombinant eGFP standards, the *N. gaditana* wild type and NgNR1-20 enabling estimation of eGFP content volumetrically and per cell. Qualitative eGFP fluorescence

measurements were made quantitative using calibration curves as in Richards *et al*. 2003. A calibration curve for the eGFP standards (fluorescence vs ng/µL) was generated in order to convert the fluorescence measurements of the samples from relative fluorescence to ng eGFP per µL (see Figure 2.5). Additionally, a calibration curve of $A_{540\ nm}$ to cell concentration in cells per mL (see Figure 2.4) was used to calculate the cell concentration and subsequently an estimate of the concentration of eGFP per 10^7 cells using the following equation.

$$\frac{\text{g GFP L}^{-1}}{\text{cells L}^{-1}} = \text{g GFP L}^{-1} \qquad (3.1)$$

The eGFP expression levels varied across the twenty NgNR strains (Figure 3.5a). Average eGFP expression across all the NgNR strains in the presence of nitrate was 443.5 ± 330.1 ng of eGFP per 10^7 cells. In the ammonium-grown cultures the average eGFP expression level was undetectable above wild type auto-fluorescence, indicating strong suppression. The fold change in fluorescence over wild type auto-fluorescence was also calculated for each NgNR strain in both nitrate and ammonium (Figure 3.5b). The average fold change in fluorescence across all NgNR strains was 4.2 ± 2.2 in the presence of nitrate and 1.0 ± 0.4 (i.e. not statistically different from wild type fluorescence level; $p = 0.47$, Student's *t*-test) in the presence of ammonium. Having demonstrated that the regulation of expression is consistent in 20 independent strains, NgNR3, which displayed an average expression level and had low background fluorescence levels in the supressed state, was selected for further analysis.





	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
fold change in NO_3	4.7	8.6	3.3	3.1	2.3	9.7	1.9	5.3	2.1	6.7	4.9	3.9	2.0	3.4	6.6	3.3	2.6	2.5	2.4	4.1
fold change in NH_4	1.3	1.5	0.8	0.4	1.1	1.1	1.3	0.6	0.8	1.0	1.4	0.6	1.0	1.0	1.0	0.6	0.7	1.4	1.5	0.8

Figure 3.5: Analysis of eGFP expression levels in NgNR1-20. (**a**) ng GFP per 10^7 cells was calculated using calibration curves. The average eGFP expression level across the twenty NgNR strains (dotted line) was 443.5 ± 330.1 ng eGFP per 10^7 cells in the nitrate-grown cultures (green bars); in the ammonium-grown cultures (red bars) the average eGFP was undetectable above wild type auto-fluorescence. Values are means ± SD of triplicate measurements. (**b**) The fold change in fluorescence over wild type auto-fluorescence across all strains. The average fold increase across the NgNR strains (dotted line) was 4.2 ± 2.2 in the nitrate-grown cultures (green bars) and 1.0 (i.e. no change) in the ammonium-grown cultures (red bars). Values are means ± SD of triplicate measurements.

3.2.3.1 Control constitutive promoter expression cassette analysis

To confirm that the changes in eGFP fluorescence in the NgNR strains were due to regulation of the nitrate reductase promoter and not to other physiological responses due to the alteration of the nitrogen source from ammonium to nitrate, a control experiment was performed. An expression vector was constructed (pLCP) that was identical to pNR except that the nitrate reductase promoter was switched for an endogenous constitutive promoter coding for a lipocalin protein (LCP; Naga_100131g14). Transcriptomic data obtained in *N. gaditana* B-31 shows that native transcript levels of the lipocalin protein were unchanged under nitrogen deprivation: 464.18 RPKM (reads per kilobase of transcript per million mapped reads) on day 3 of growth in sodium nitrate and 436.61 RPKM on day 3 under nitrogen depletion; 555.90 RPKM on day 6 of

growth in sodium nitrate and 566.70 RPKM on day 6 of growth in nitrogen deplete conditions (see Figure 3.6) (Corteggiani Carpinelli *et al*., 2014).

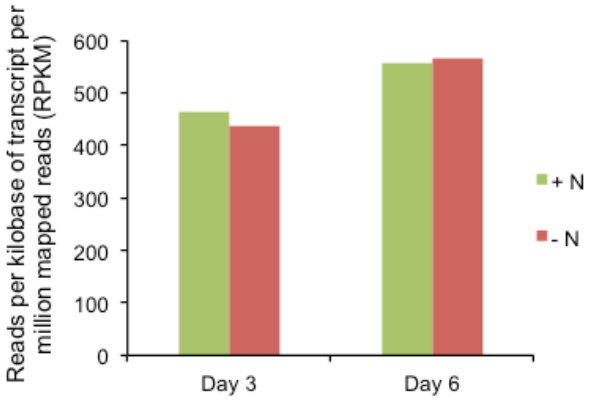


Figure 3.6: Native transcript levels of control lipocalin protein coding gene in *N. gaditana* B-31. Transcript levels are relatively unaffected by this alteration of nitrogen conditions on days 3 and day 6 of growth. + N, sodium nitrate as sole nitrogen source; - N, nitrogen deplete conditions.

Three transformants generated with the pLCP vector (NgLCP1-3) were grown to the logarithmic growth phase in 5 mM sodium nitrate and 5 mM ammonium chloride and analysed for eGFP fluorescence as described for the NgNR cultures (Figure 3.7). The wild type subtracted normalised fluorescence measurements in NgLCP1-3 were variable between replicates as was observed for the NgNR strains. However, there was no significant difference between eGFP fluorescence in each replicate when the nitrogen source was altered ($p = 0.44$, Student's *t*-test). Hence, the control analysis with the NgLCP strains showed no significant effect on eGFP fluorescence levels when the nitrogen source was switched from ammonium to nitrate, confirming that the changes observed in the NgNR strains were due to regulation of the nitrate reductase promoter.

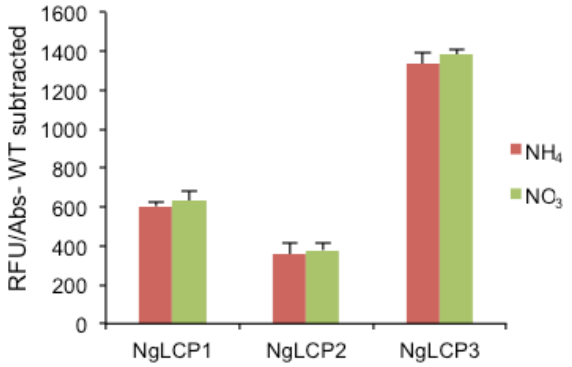


Figure 3.7: Control fluorescence level analysis of the constitutive lipocalin protein promoter. Expression of eGFP under the control of the lipocalin protein promoter is not significantly different ($p = 0.44$, Student's *t*-test) between cultures grown in ammonium chloride and sodium nitrate at the logarithmic growth phase. Values are means ± SD of triplicate measurements.

3.2.4 Immunoblot analysis

To confirm that eGFP was expressed, NgNR3 was further examined by immunoblot analysis (see section 2.6.6 for methods; Figure 3.8). Protein samples were extracted from wild type cells and NgNR3 grown to the logarithmic phase in media containing either 5 mM ammonium or 5 mM nitrate as the sole nitrogen source. eGFP was detected as a single band at 27 kDa in the nitrate-grown NgNR3 culture but was undetectable in both the *N. gaditana* wild type and NgNR3 grown in ammonium. Quantification of the protein band in the nitrate-grown NgNR3 culture using the recombinant eGFP standards calibration curve determined a concentration of 115 ng of eGFP per 10 µg of total protein; using the following equation:

$$\frac{\text{g eGFP}}{\text{g of total protein}} \text{ x } 100 = \% \text{ eGFP of total protein} \qquad (3.2)$$

this equates to approx. 1.15 % of total protein.

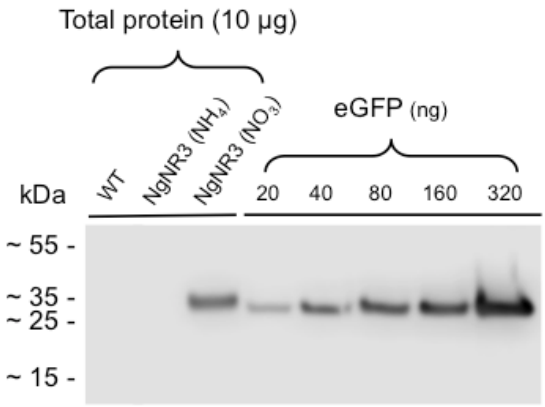


Figure 3.8: Analysis of eGFP in ammonium and nitrate-grown NgNR3 cultures via immunoblotting. NgNR3 Cultures were grown to the logarithmic phase in media containing either ammonium or nitrate as the sole nitrogen source before protein samples were extracted and probed with an antibody to eGFP. In the nitrate-gown culture samples and the eGFP standards, a band migrating to a distance equivalent to the expected mass of eGFP (27 kDa) was detected. In the ammonium-grown culture and the wild type negative control, this band was not detected. Lane 1; *N. gaditana* wild type (negative control), lane 2; ammonium-grown NgNR3 culture, lane 3; nitrate-grown NgNR3 culture, lane 4-8; eGFP standards (20, 40, 60, 80, 160, 320 ng).

3.2.5 Regulation of P_{NR} in different nitrogen conditions

The ability to induce expression of the nitrate reductase promoter in cells activity growing on other nitrogen sources is a potentially useful feature for heterologous protein expression in *N. gaditana*. eGFP expression and cell density in NgNR3 cultures pre-grown in different nitrogen conditions is shown in Figure 3.9. NgNR3 and wild type cells were grown to logarithmic phase in standard media, washed in nitrogen-deplete media three times and then used to inoculate cultures containing ammonium, urea (each at 5 mM) or no nitrogen. Cells were treated with 5 mM nitrate on day 5 to induce expression. Cell growth and eGFP expression in the NgNR3 cultures was measured every two days (see methods).

eGFP fluorescence was undetectable in NgNR3 for the first four days of growth on ammonium. After the addition of nitrate to the ammonium-grown cultures on day 5, eGFP fluorescence was detected the following day, showing that induction of P_{NR} is possible in the presence of both ammonium and nitrate. Fluorescence continued to increase and peaked on day 10 (Figure 3.9a). However, the GFP fluorescence level in NgNR3 on day 10 was ~13.9% of that achieved when NgNR3 was grown in media containing nitrate as the only source of nitrogen (Figure 3.9a). The growth of the ammonium-grown wild type and NgNR3 cultures was not statistically significantly different over the 12 days (p = 0.38, Student's t-test; Figure 3.9b).

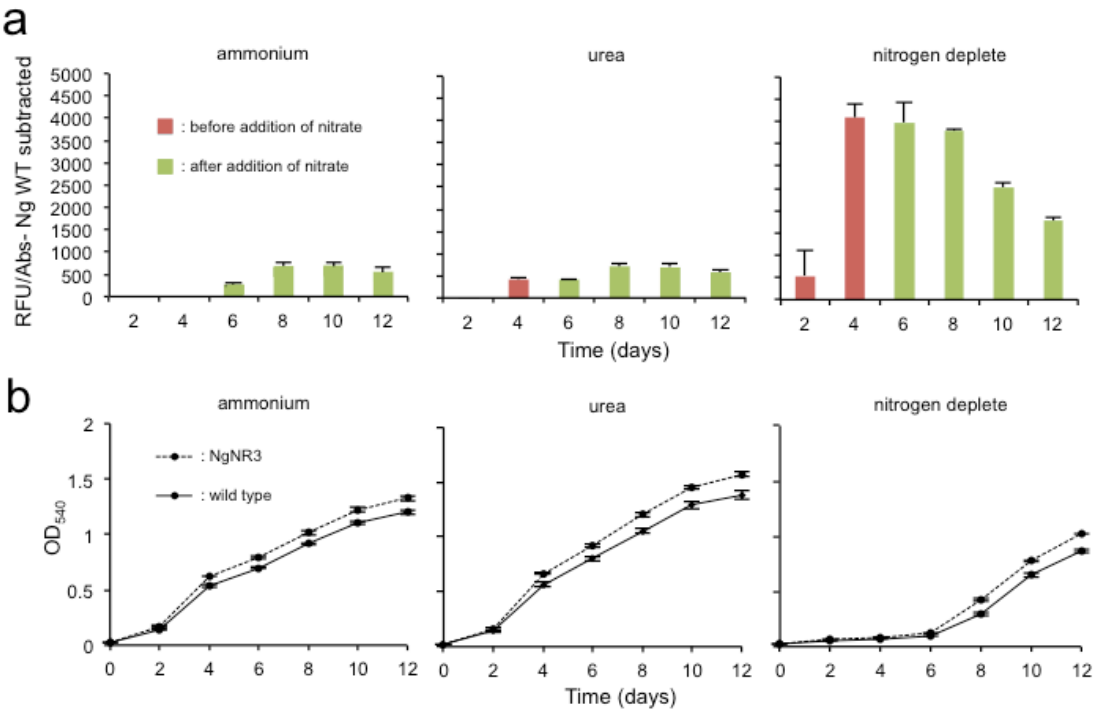


Figure 3.9: Regulation of P_{NR} with different nitrogen sources. *N. gaditana* wild type and NgNR3 cultures were grown on the following nitrogen sources before addition of 5 mM sodium nitrate to cultures on day 5: 5 mM ammonium chloride, 5 mM urea, and nitrogen-deplete media. (**a**) P_{NR} driven eGFP expression in NgNR3. eGFP fluorescence is expressed as normalized fluorescence with wild type auto-fluorescence subtracted. Grey bars indicate expression before addition of nitrate on day 5; black bars indicate expression after addition of nitrate. (**b**) Growth curves for NgNR3 grown on different nitrogen sources. Solid lines indicate *N. gaditana* wild type and dashed lines NgNR3. Data shown is mean ± SD of triplicate measurements.

When NgNR3 was grown on urea, eGFP fluorescence was observed before the addition of nitrate on day 5, indicating that the presence of ammonium is required for full suppression of P_{NR} in these conditions. The level of eGFP expression after the addition of nitrate to the urea-grown cells was similar to the ammonium-grown cells (equivalent to ~14.1% of the expression achieved in the presence of nitrate alone; Figure 3.9a). As in the ammonium-grown cultures, there was no statistically significant difference between growth of the wild type and NgNR3 over the 12 days ($p = 0.37$, Student's *t*-test; Figure 3.9b).

eGFP fluorescence in the nitrogen-deplete cells peaked on day 4, before the addition of nitrate on day 5, and was equivalent to 81.1% of the expression achieved in the presence of nitrate alone (see Figure 3.9a). This is the highest fluorescence level observed in this series of experiments, further indicating that the presence of ammonium is required for full suppression of expression from P_{NR} under these conditions. The absence of any nitrogen source until the addition of nitrate on day 5 resulted in low levels of growth up until this point. However, in these conditions, growth of NgNR3 was not statistically significantly different from the wild type ($p = 0.37$, Student's *t*-test; Figure 3.9b).

3.2.6 Long-term stability of NgNR3

The ability of transgenic biotechnological strains to withstand the effects of silencing of transgene expression is of particular importance. To confirm the long-term stability of the pNR cassette, NgNR3 was sub-cultured for 20 months in liquid culture without selective pressure before being re-tested for resistance to hygromycin B. Figure 3.10a shows that this line still retained resistance under these conditions. The presence of the *hph* gene was confirmed by PCR (Figure 3.10b) and eGFP expression was confirmed by fluorescence, as described previously (Figure 3.10c).

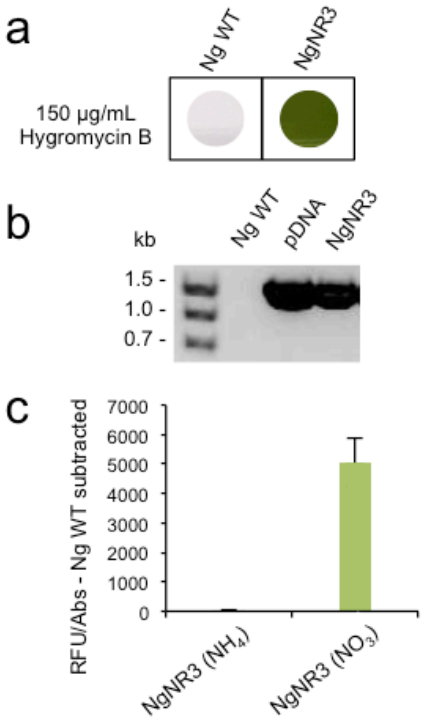


Figure 3.10: Long-term stability of the NgNR3 strain. After 20 months of sub-culturing in liquid media the following analyses were repeated to assess the long term-stability of the NgNR3 strain. (**a**) Resistance of NgNR to 150 µg/mL of hygromycin B in liquid culture. (**b**) The presence of the *hph* resistance gene was detected via PCR using the same primers as previously described. Template DNA: DNA ladder; lane 1, *N. gaditana* wild type (negative control); lane 2, pNR vector (positive control); lane 3, NgNR3; lane 4 (**c**) eGFP fluorescence expressed as A540 nm normalized fluorescence with wile type auto-fluorescence subtracted in ammonium and nitrate-grown NgNR3 cultures. Values are means ± SD of triplicate measurements.

3.3 Discussion

The use of constitutive promoters for the expression of toxic compounds for industrial or pharmaceutical uses or for establishing silencing systems can be problematic. Therefore,

inducible expression systems offer considerable advantages in these cases (Doron *et al*., 2016). The use of inducible systems is also advantageous for large-scale production of recombinant proteins in microalgae, enabling the cells to first reach an optimal concentration before expression of the transgene is initiated (Doron *et al*., 2016). The potential of the promoter of the endogenous nitrate reductase gene of *N. gaditana* to meet these demands was explored in this study.

The eGFP fluorescence-based analysis of expression levels over twenty strains transformed with the pNR constructed (NgNR1-20) allowed calculation of an average eGFP expression level of 443.5 ± 330.1 ng of eGFP per 10^7 cells in the presence of nitrate and suppression of expression to undetectable levels in ammonium (Figure 3.5a). Assuming a cellular protein content of approximately 3 pg of protein per cell (Fábregas *et al*., 2002), the average level of eGFP expression in the presences of nitrate in pg per cell (4.435 x 10^{-2}; Figure 3.5a) is equivalent to approximately 1.48 ± 1.10 % of total protein. This is comparable to expression systems established in the model organism *C. reinhardtii* in which heterologous protein production in the chloroplast typically ranges from 1-5% of total protein (Almaraz-Delgado *et al*., 2014; Rasala *et al*., 2011; Manuell *et al*., 2007). Using the nitrate reductase promoter to drive expression of eGFP, an average fold change in fluorescence of 4.2 was achieved (Figure 3.5b). Many factors influence the ability to detect fluorescent proteins in a cellular context, such as background auto-fluorescence, the strength of the promoter used to drive expression, and rates of mis-folding and successful maturation of the fluorescent protein. In a previous study, expression of a widely used *Chlamydomonas* codon-optimized GFP (CrGFP) in *C. reinhardtii* under the control of a robust promoter (hsp70/rbcs2) yielded a comparable fold change in fluorescence of 2.8 (Rasala *et al*., 2013). Although the pattern of suppression and induction was conserved across the twenty NgNR strains, there was a large degree of variability in expression levels in the P_{NR}-induced conditions (nitrate-grown). This variability is likely to be the result of random integration of the expression cassette and positional effects within the genome on expression, which is commonly observed in microbes transformed with random integration-based transformation procedures such as electroporation (Zhang *et al*., 2014; Chen and Zhang, 2016; Thompson and Gasson, 2001). The fluorescence-based approach of quantifying expression using calibration curves presented here (Figure 3.5) provides a quantitative method for rapidly analysing multiple strains with relative ease, enabling cell lines with the desired level of expression to be identified.

The series of experiments exploring nitrogen regulation (Figure 3.9) shed light on the ability of the nitrate reductase promoter to regulate expression under a variety of nitrogen conditions. Transcript levels of the *NR* gene of the eukaryotic alga *D. tertiolecta* have been shown to be supressed in the presence of ammonium and induced in the presence of nitrate; however, in the

presence of both ammonium and nitrate the transcript was fully supressed (Song and Ward, 2004). This observation at the transcript level in *D. tertiolecta* was not observed at the protein level in NgNR3 as eGFP expression was seen in the combined presence of ammonium and nitrate (Figure 3.9a). Therefore, removal of residual ammonium before induction of P_{NR}, which could be costly and impractical on a large-scale, may not be necessary for exploitation of this system in a biotechnological context. However, expression was lower than when nitrate was the sole source of nitrogen. When grown on urea, expression of eGFP was observed before the addition of nitrate, suggesting the strong effects of ammonium on expression may be required if total suppression is required (Figure 3.9a) (Forde, 2000; Galván and Fernández, 2001; Llamas *et al*., 2002; Fernández and Galván, 2008; Raven *et al*., 1992). The strong expression of eGFP in the nitrogen-deplete cultures before addition of nitrate suggests that P_{NR} is also induced by nitrogen deprivation. Transcript levels of genes involved in nitrogen uptake, scavenging mechanisms and assimilation, including nitrate reductase, have also been reported to increase under nitrogen deprivation in other algae as a scavenging mechanism (Alipanah *et al*., 2015). Expression from P_{NR} under nitrogen-depletion could offer another means of induction for industrial application whereby the nitrogen concentration of the growth media could be set such that cells grow to a desired density before depletion of nitrogen and concurrent onset of expression from P_{NR}.

As has been shown in other algae, the endogenous promoter of the nitrate reductase gene was able to efficiently regulate expression of heterologous protein in *N. gaditana* (Poulsen and Kröger, 2005; Schmollinger *et al*., 2010; Wang *et al*., 2004; Li *et al*., 2007; von der Heyde *et al*., 2015; Niu *et al*., 2012). Strong expression and suppression was possible in the sole presence of nitrate and ammonium, respectively. Furthermore, the ability to fully supress expression from the *NR* promoter through continuous growth on ammonium chloride and then induce protein production through addition of sodium nitrate, without the removal of ammonium, is a useful feature that will enable future applications requiring precise temporal induction of protein expression at a specific point in the growth phase, while also maintaining growth rates. Induction and suppression of P_{NR}-driven eGFP expression was greatly affected by the nitrogen source of the growth media, presenting a variety of ways to control expression; accordingly, nitrogen sources should be user-defined on the basis of the intended use of the promoter. The establishment and validation of the *in situ* eGFP reporter system and ability to temporally control transgene expression with the endogenous nitrate reductase promoter are valuable additions to the advancement of sophisticated genetic engineering technologies that will enable further development of the oleaginous alga *N. gaditana* as a model organism for biotechnology. Future studies on the *NR* promoter should be directed towards identifying key regulatory motifs within the promoter region and to elicitation of the transcription factors involved in its regulation. One strategy that could aid this would be to repeat the analysis of the *NR* promoter

with successively more truncated versions of the promoter region in order to identify the minimum possible length of the nucleotide sequence before the function of the promoter is lost (Ding *et al*., 2013). Additionally, further investigation of regulation strategies using different media compositions and onset timings should be performed in both lab-scale and large-scale commercial systems.

Chapter 4: Promoter trapping in *N. gaditana*

4.1 Introduction

Microalgae have been a focus of the biotechnology industry for several decades as a potential source of naturally derived high-value products, including nutraceuticals, pigments, cosmetics and increasingly so in the last few years, as a platform for biofuel production (Borowitzka, 2013). The fast growth rates of microalgae under phototrophic conditions, low cost of maintenance compared to plants, yeast and bacteria, and the ability to grow on non-arable land make them an attractive prospect for use as expression systems for synthesis of high-value compounds (Doron *et al*., 2016; Rasala *et al*., 2011; Almaraz-Delgado *et al*., 2014; Feng *et al*., 2014). In order to realize the full potential of microalgae, more advanced technologies for transgene expression are needed for commercially promising species. Such technologies have been developed in the model alga *Chlamydomas reinhardtii*, where efficient transgene expression was made possible by the accumulation of promoters and translated regions (Kindle *et al*., 1989; Stern *et al*., 2008). The number of microalgal species that have been successfully used for nuclear transformation has slowly risen and currently stands at ~25 species (Doron *et al*., 2016). More sophisticated techniques for chloroplast transgenic approaches have been developed in several species of microalgae, most notably *Chlamydomonas*, such as the extensive use of reverse genetic studies to gain insight into chloroplast gene function, techniques for marker recycling and the use of plastome deletion mutants as a host strain (Purton, 2007; Franklin *et al*., 2002; Rasala *et al*., 2011, 2014). However, the same level of development has not been observed for nuclear expression systems (Doron *et al*., 2016).

4.2 The importance of endogenous promoters

In addition to delivery of DNA to the nucleus, highly active endogenous promoters are one of the key requirements for achieving efficient transformation and/or expression of transgenes in microalgae. Heterologous promoters have been used to drive expression of selection markers in

green microalgae however the expression is often transient or low efficiency (Kim *et al*., 2002; Feng *et al*., 2009). Currently, endogenous promoters are best for generating high efficiency transformation and stable transgenic cell lines (Walker *et al*., 2005a). A large collection of constitutively and highly active promoters is available for *C. reinhardtii*, for example the fusion of the RuBisCO small subunit *RBCS2* promoter and the heat shock protein (*HSP70A*) promoter (Neupert *et al*., 2012). However, the use of endogenous promoters for transformation has only been explored in a limited number of green algae (Vila *et al*., 2012). Additionally, the isolation of strong endogenous promoters and use for transgene expression does not always result in successful transformation, as was made clear by the attempts made to transform *D. tertiolecta* using the endogenous *RBCS2* promoter, which resulted in low efficiency transient transformation (Walker *et al*., 2005b). However, the *D. tertiolecta RBCS2* flanking regions were effectively used as heterologous promoters in *C. reinhardtii* to drive expression of a bleomycin selection marker (*ble*) (Walker *et al*., 2005a). This illustrates that more work is required to develop technologies that can allow identification of endogenous promoters that can robustly drive transgene expression in a variety of microalgae.

4.3 Promoter trapping

New promoters can be identified using bioinformatic approaches, however this relies on available genome sequences. Alternatively, if genomic sequences are not available, DNA walk-based methods can be used to retrieve sequences flanking a protein-encoding sequence, which is the most commonly adopted strategy in these cases (Tan *et al*., 2005b). Another approach that has been successful in identifying a large number of novel promoters in higher plants is the promoter trapping method (Blanvillain and Gallois, 2008; Springer, 2000). Briefly, this method consists of generating a library of transformed strains each containing a randomly inserted reporter gene and the subsequent use of this known insert sequence to sequence the genomic regions surrounding it, through a variety of methods. However, the only algal species in which promoter trapping has been used is the model species *C. reinhardtii* (Haring and Beck, 1997; Auchincloss *et al*., 1999; Vila *et al*., 2012).

In order to expand the use of this promising technology in a more commercially relevant species, a promoter trapping system for the eukaryotic alga *N. gaditana* has been developed. *N. gaditana* is a fast growing oil-rich strain which has recently attracted attention due to its ease of transformation via electroporation and amenability to more sophisticated genome-editing techniques (see sections 2.1.1 and 3.1.2) (Radakovits *et al*., 2012; Perin *et al*., 2015; Ajjawi *et al*., 2017). In this work a gene-trap vector has been used containing a *ble* selection marker. The *ble* resistance product inactivates zeocin by binding to the antibiotic; this mode of action confers

a linear relationship between the expression of *ble* and the level of antibiotic resistance and is an effective means of isolating highly active trapped promoters. The vector also contained a mCherry fluorescent reporter coding sequence linked to the *ble* gene by a 2A sequence, which is self-cleaving to ensure un-fused expression of both proteins at a 1:1 ratio. Following random insertion of the vector, mutant strains were screened on an increasing scale of zeocin to reveal highly active promoters. Sequencing of the genomic regions flanking the vector subsequently allowed 1000 bp regions upstream of the corresponding genes to be amplified and cloned into an enhanced green fluorescent protein (eGFP) expression cassette for validation of the promoters in a non-native context in *N. gaditana*. This study shows that the promoter trap approach is an effective means of identifying robust endogenous promoters for transgene expression.

4.4 Results

4.4.1 The Ble2AmChr.SD gene trap cassette and promoter trapping strategy

The Ble2AmChr.SD cassette used for promoter trapping (Figure. 4.1) contained a bleomycin resistance gene (*ble)* conferring resistance to zeocin for selection of transgenic *N. gaditana* cell lines and a mCherry fluorescent protein-coding gene (*mCherry*). A flippase recognition target site (FRT) was added to the 3′ end of the *ble* gene to prevent translational linkage of the *ble* resistance product to the endogenous protein associated with the trapped endogenous gene. The *ble* and *mCherry* genes were expressed using the 2A peptide system to yield two separate products in equal quantity (Ble and mCherry). The cassette was flanked by a 35 bp splice acceptor site (SA) at the 3′ end and a 7 bp spice donor (SD) site at the 5′ end. The splice acceptor and donor sequences were consensus sequences for microalgae and direct integration of the cassette into an endogenous gene. The expression level of the cassette is set by the level of activity of the promoter associated to the trapped endogenous gene. The Ble2AmChr.SD vector was developed and supplied by Algenuity (Eden Laboratory, Bedfordshire, United Kingdom).

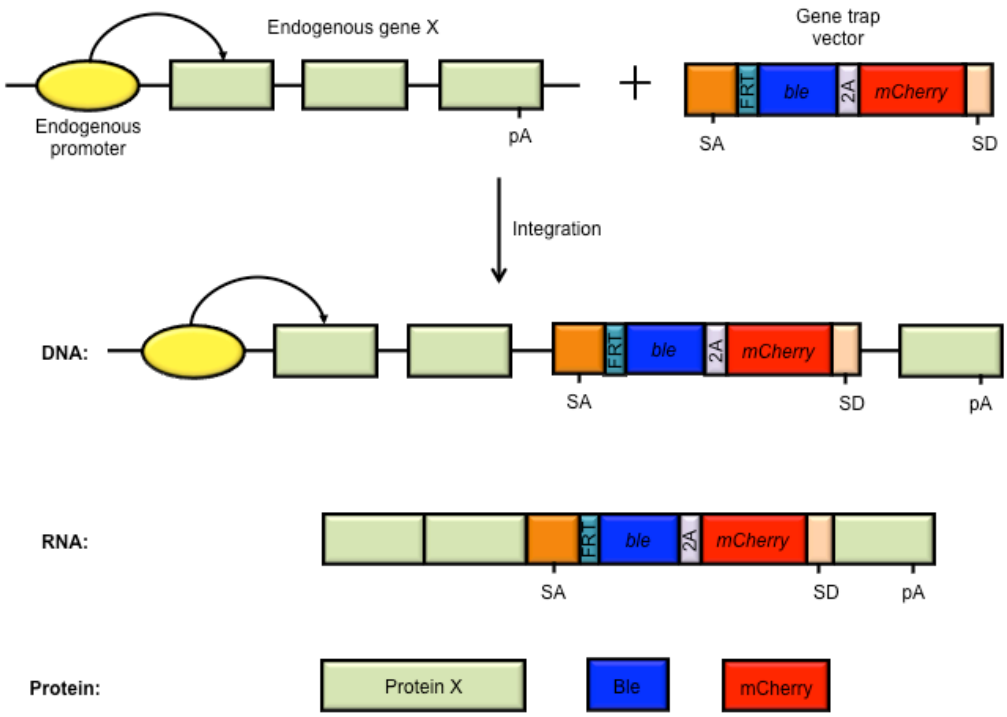


Figure 4.1: Integration and expression of the Ble2AmChr.SD cassette. Consensus splice acceptor and splice donor sequences direct integration of the vector into an endogenous gene X, generating a fusion transcript proportional in expression level to the strength of the endogenous promoter at the location of insertion. Translation of this fusion transcript generates three discrete products: a truncated endogenous protein X, the Ble resistance protein and the mCherry fluorescent protein. SA, splice acceptor site; SD, splice donor site; *ble,* bleomycin resistance gene; Ble, bleomycin resistance protein, conferring resistance to zeocin; *mCherry*, mCherry fluorescent reporter gene; mCherry, mCherry fluorescent reporter protein; FRT, flippase recognition target site; 2A, self-cleaving 2A peptide.

4.4.2 Transformation of *N. gaditana* with the Ble2AmChr.SD cassette and promoter strength screening

N. gaditana was grown to the logarithmic growth phase (see section 2.2.1.1) and transformed with 3 µg of the Ble2AmChr.SD cassette. A no DNA negative control was also included for the transformation, which did not yield any colonies. After transformed colonies had appeared on the transformation plates and were large enough to be picked, 96 colonies were screened for promoter strength. 96 colonies were selected and loaded into a 96 well microplate containing 200 µL of media in sterile conditions. A 96 well replicator was then used to dispense 5 µL of each cell line on agar plates containing 0, 20, 40, 60 and 80 µg/mL of zeocin. After 5 weeks of incubation under continuous fluorescent white light (200 µmol m^{-2} s^{-1}) at 24 °C, the number of viable cell lines at each concentration was recorded (Figure 4.2).

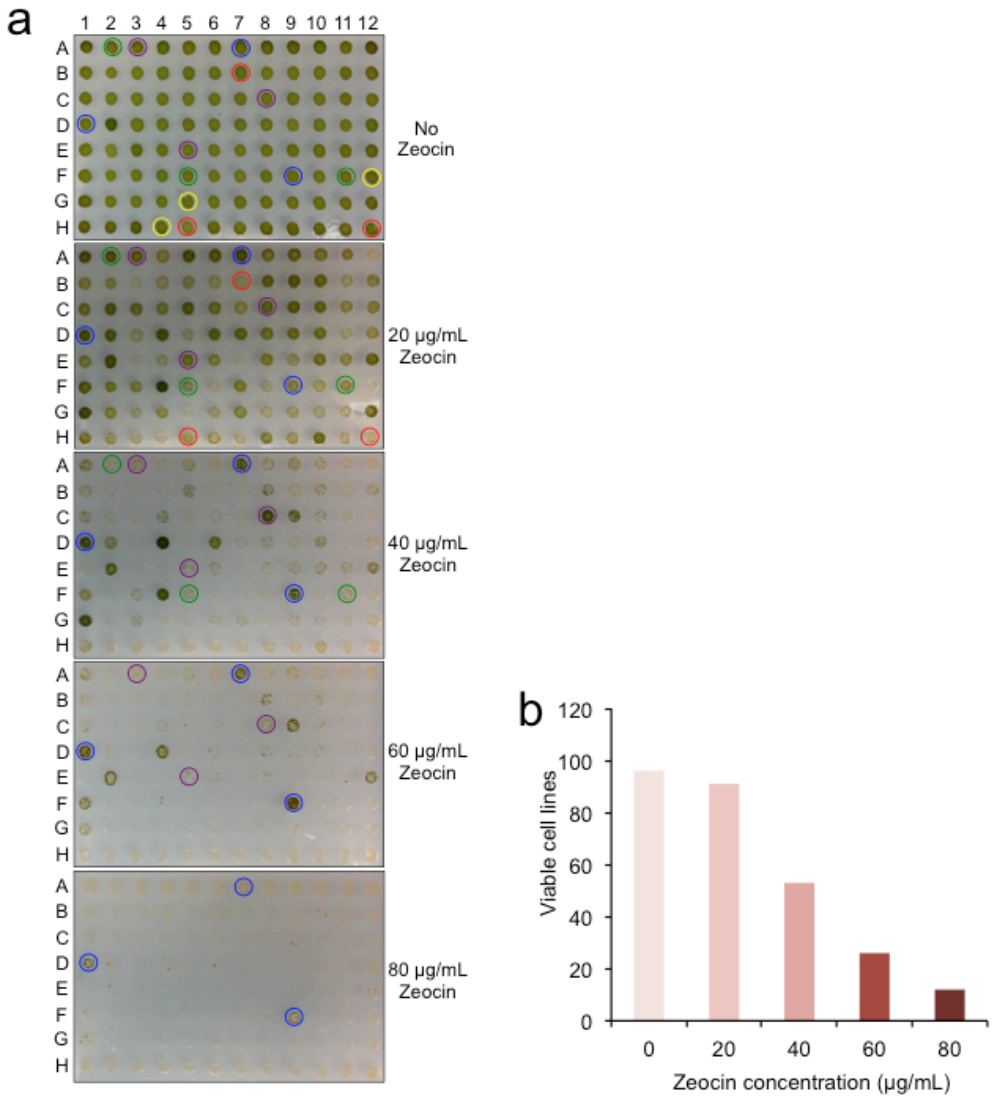


Figure 4.2: Selection of robust endogenous promoters. (**a**) Screening of *N. gaditana* cell lines transformed with the Ble2AmChr.SD cassette, conferring resistance to zeocin, on increasing concentrations of the antibiotic. The level of resistance is determined by the strength of the endogenous promoter trapped by the cassette. Yellow, red, green, purple and blue circle denote cell lines, chosen for further analysis, which were viable at 0, 20, 40, 60 and 80 µg/mL of zeocin, respectively. (**b**) Quantification of number of viable cell lines at each concentration of zeocin in the series shown in panel **a**.

All cell lines showing growth within the 5 µL spot at each concentration were classed as a viable cell line, although growth was limited to a few colonies in some cases. Three Ble2AmChr cell lines were selected from each plate/concentration (fifteen in total) for further analysis, ensuring that the cell lines selected at each concentration lost viability at the next concentration in the series (Figure 4a). These strains were sub-cultured once in F2N liquid media with 1.5 µg/mL of zeocin and sub-cultured in standard F2N liquid media and on standard F2N agar plates thereafter.

4.4.3 Confirmation of Ble2AChr.SD cassette integration

The selected Ble2AmChr.SD cell lines were named according to their level of zeocin resistance: z80.1 – z80.3, etc. In order to confirm integration of the Ble2AChr.SD cassette, all cell lines were initially genotyped by PCR with two gene specific primers binding within the *ble* gene (B/C.SD F1 and B/C.SD R1), giving a product of 160 bp in length (Figure 4.3). This amplicon was generated in the positive plasmid DNA control and all of the transformed *N. gaditana* cell lines and was absent in the wild type.

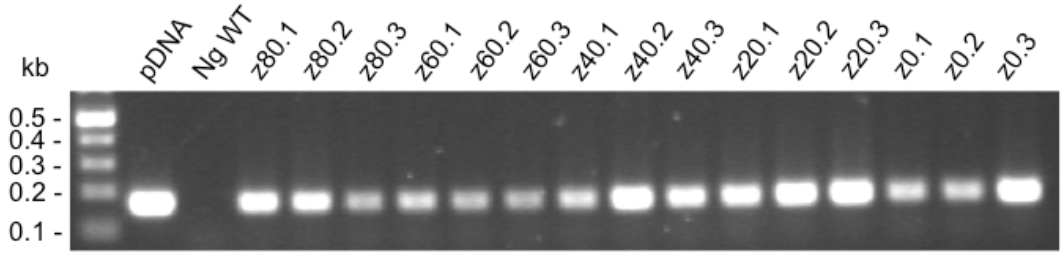


Figure 4.3: PCR genotyping of Ble2AmChr.SD cell lines to confirm integration of the Ble2AChr.SD vector. All reactions used the forward primer B/C.SD F1 and reverse primer B/C.SD R1, generating a 160 bp amplicon. Template DNA: Lane 1, DNA ladder; lane 2, pDNA, plasmid positive control DNA (Ble2AChr.SD); lane 3, *N. gaditana* wild type; lanes 4-18, *N. gaditana* Ble2AmChr.SD transgenic cell lines.

4.4.4 Identification of endogenous genes trapped by Ble2AChr.SD cassette

4.4.4.1 3′ RACE procedure

To identify the putative promoter regions responsible for expression of the Ble2AChr.SD cassette a 3′ RACE (Rapid Amplification of cDNA Ends) protocol was initially tested. RACE is a procedure for amplification of nucleic acid sequences from a messenger RNA template between a defined internal site and the 3′ or 5′ end of the mRNA (see Figure 4.4) (Frohman *et al*., 1988). mRNA was extracted from all of the selected Ble2AmChr.SD cell lines and the *N. gaditana* wild type and reverse transcription reactions were performed using the Ap_oligo(dT)18 adaptor primer, generating cDNA templates of all transcripts with the Ap_oligo(dT)18 adaptor sequence at the 3′ end. These cDNA templates were then used for two PCR reactions. The first set of PCRs (Figure 4.5a) used the gene specific B/C.SD RT F2 primer, binding within the *mCherry* gene, and the AP-OUT primer binding to the adaptor sequence at the 3′ end of the cDNA template. To increase specificity and amplification of the cDNA templates containing the Ble2AChr.SD cassette, a 1:10 dilution of these PCR reactions was then used as template DNA for the second set of nested RACE PCRs using the second gene specific

primer B/C.SD RT F3, nested downstream of the B/C.SD RT F2 primer and the AP-IN primer, nested upstream of the AP-OUT primer (Figure 4.5b). Due to the sequence specificity in the initial RACE PCRs being derived solely from the gene specific primer, it is not expected to see products in the initial RACE PCRs (Figure 4.5a). However, the efficiency of the nested RACE PCRs (Figure 4.5b) was low, with only one amplicon being generated for the z20.1 strain, indicating that the 3′ RACE procedure was unsuccessful.

In order to determine if the failure of the first 3´ RACE procedure (Figures 4.5a and 4.5b) was due to the distance between the gene specific primers and the adaptor primers at the 3′ end of the cDNA template being too large, a second set of 3´ RACE PCRs was performed (Figures 4.5c and 4.5d). All conditions remained the same, however, the gene specific primers were specific to sequences further downstream in the gene-trap cassette. The new B/C.SD RT F2 (b) primer was used for the initial RACE PCRs (Figure 4.5c) and the new B/C.SD RT F3 (b) primer was used for the nested PCRs (Figure 4.5d). However, the nested PCRs again showed poor efficiency (Figure 4.5d). Some bands were visible, however they appeared at the same migration level in multiple strains and were therefore believed to be non-specific. The 3′ RACE procedure was therefore unsuccessful and an arbitrary primer PCR approach was adopted to identify genes containing the Ble2AChr.SD cassette.

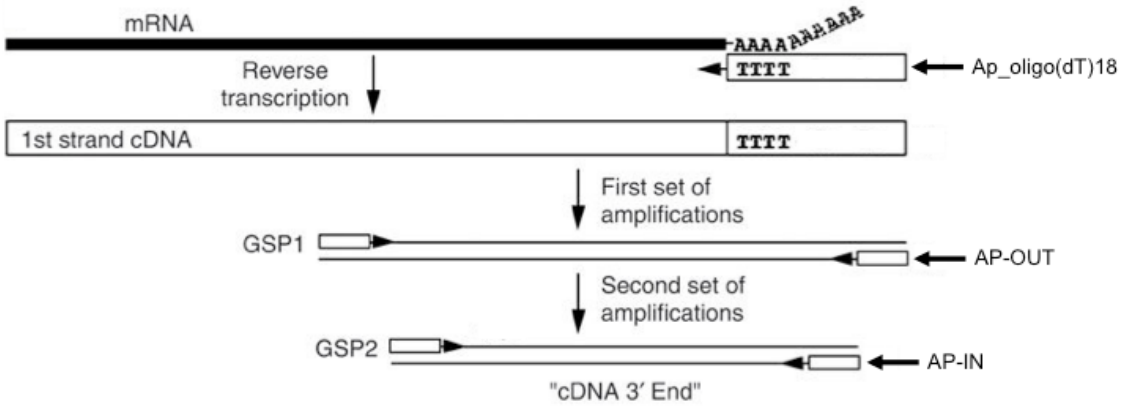


Figure 4.4: Schematic of 3´ RACE (rapid amplification of cDNA ends) procedure. A reverse transcription is first performed with the adaptor primer Ap_oligo(dT)18. Through the conversion of the transcript from mRNA into cDNA, this adaptor primer provides an anchor for the reverse primers used in the subsequent RACE PCRs (AP-OUT and AP-IN). The first set of RACE amplifications uses a gene specific primer (GSP1) that binds within the known vector sequence in combination with AP-OUT. The second set of nested amplifications increases specificity to the target transcript by nesting the second gene specific primer (GSP2) in the 3´ direction of GSP1 and nesting the AP-IN primer in the 5´ direction of AP-OUT. Figure adapted from Scotto-Lavino *et al*. (2007).

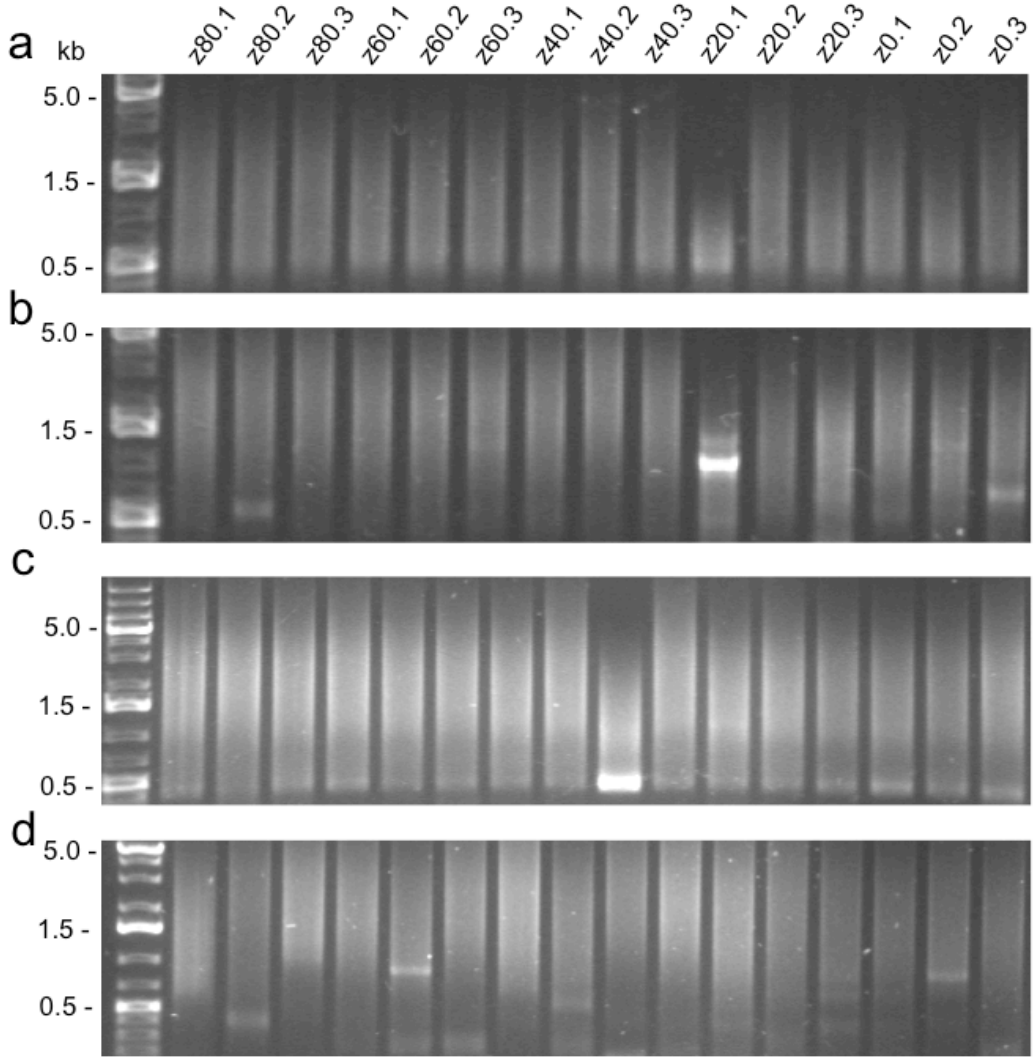


Figure 4.5: Experiments to identify endogenous genes driving expression of the Ble2AmChr.SD cassette using 3′ RACE (Rapid Amplification of cDNA Ends). (**a**) RACE PCRs using the B/C.SD RT F2 gene specific primer. Template DNA: lanes 1-15, mRNA/cDNA templates from Ble2AChr.SD strains. (**b**) Nested RACE PCRs using the B/C.SD RT F3 gene specific primer. Template DNA: lanes 1-15, 1:10 dilutions of initial PCRs performed on Ble2AChr.SD strains. (**c**) RACE PCRs using B/C.SD RT F2 (b) gene specific primer. Template DNA: lanes 1-15, mRNA/cDNA templates from Ble2AChr.SD strains. (**d**) Nested PCRs using B/C.SD RT F3 (b) gene specific primer. Template DNA: lanes 1-15, 1:10 dilutions of initial PCRs performed on Ble2AChr.SD strains. No genuine gene specific products were obtained in either experiment.

4.4.4.2 Arbitrary primed PCR

An arbitrary primed PCR (AP-PCR) procedure was adapted for the *N. gaditana* Ble2AmChr.SD cell lines following methods described previously (Tran *et al*., 2015; Das *et al*., 2005). Briefly, AP-PCR uses a primer containing an arbitrary sequence paired with a gene specific primer of known sequence binding within the Ble2AChr.SD cassette. It is expected that the arbitrary primer will bind at multiple locations in the genome, including adjacent to the gene specific sequence. Ideally, the arbitrary primer will produce an amplicon in conjunction with the gene specific primer. As illustrated in Figure 4.6, the procedure is performed in two rounds. In the first, an arbitrary primer (Arb1 or Arb1-5) is paired with a gene specific primer. The products of these reactions then serve as templates for second round reactions employing a nested gene specific primer and a second arbitrary primer (Arb2 or Arb2b).

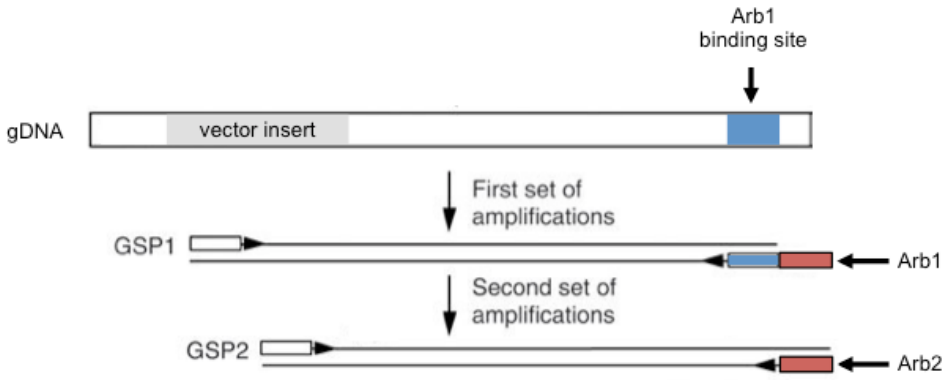


Figure 4.6: Schematic of arbitrary primed PCR (AP-PCR). A PCR is initially carried out using a gene specific primer (GSP1) binding within the vector sequence (grey shading) and an arbitrary primer (Arb1) that is expected to bind somewhere within the neighbouring genomic DNA. A second PCR is then performed with a nested gene specific primer (GSP2) and the second arbitrary primer (Arb2).

Genomic DNA, normalized to 10 ng/µL, was first extracted from *N. gaditana* wild type and the fifteen transgenic lines and used as template DNA for the first round of PCRs. The products of the first round of PCRs were purified and diluted 1:10 for the second round of PCRs.

Initially, an AP-PCR procedure was carried out using methods described by Tran *et al*. (2015). For the first round PCR, the B/C int gene specific primer (binding centrally to the *mCherry* gene) was used with the arbitrary primer Arb1; for the second round of PCRs, the B/C ext primer (binding towards the 3´ end of the *mCherry*) was used with the Arb2 primer (Figure 4.7a). No genuine products were obtained in the second round of PCRs, aside from some non-specific products appearing at the same migration in multiple PCRs.

Secondly, the same gene specific primers (B/C int and B/C ext; primer set a) were used with primers (Arb1-5 and Arb2b) and cycling parameters described by Das *et al*. (2005). Three

second round PCR products were obtained across the fifteen strains using this procedure (Figure 4.7b), which were gel extracted (Figure 4.7d) and sent for sequencing (Table 4.1; z60.3 (a), z20.3 (a) and z0.3 (a)). The results showed that the arbitrary primers had bound within the vector sequence and did not extend into the neighbouring *N. gaditana* genomic region.

A second gene specific primer set, binding further towards the 3´ end of the cassette was then designed (B/Cd int and B/Cd ext; primer set b) and used with the methods described by Das *et al*. (2005). Seven second round PCR products were obtained across the fifteen strains using this procedure (Figure 4.7c), which were gel extracted (Figure 4.7e) and sent for sequencing (Table 4.1; z80.1 (b), z60.1 (b), z60.3 (a), z20.2 (b), z20.3 (b), z0.2 (b) and z0.3 (b)). See Appendix B for all nucleotide sequences of AP-PCR sequencing results.

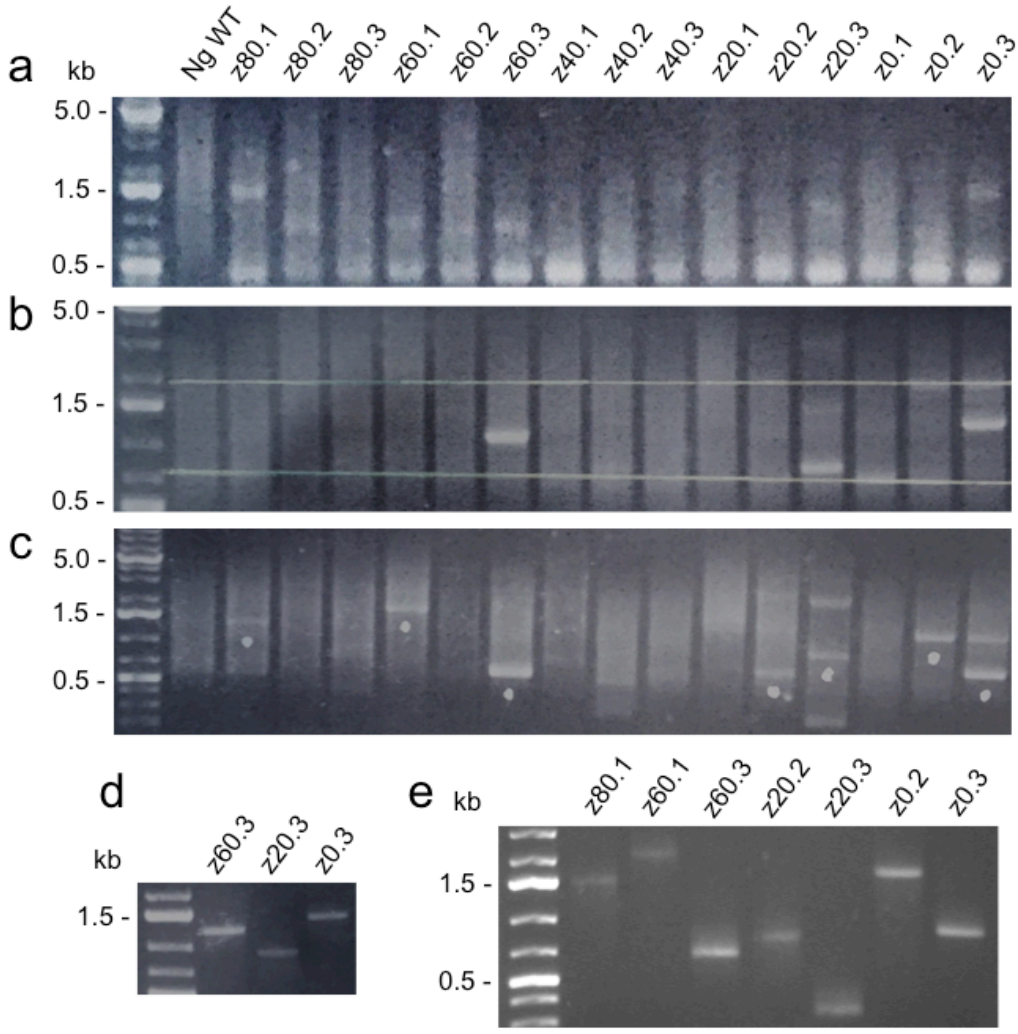


Figure 4.7: Arbitray primed PCR to identify endogenous genes driving expression of Ble2AmChr.SD cassette. (**a**) AP-PCR using primer set a (B/C int and B/C ext) and using the method of Tran *et al*. (2015). (**b**) AP-PCR using primer set a (B/C int and B/C ext) and methods described by Das *et al*. (2005). (**c**) Second AP-PCR using primer set b (B/Cd int and B/Cd ext) and methods described by Das *et al*. (2005). Dots denote products that were gel extracted. (**d**) The three gel extracted products from panel **b**. (**e**) The seven gel extracted products from panel **c**.

4.4.4.3 Analysis of amplicons generated using AP-PCR

The amplicons generated using the AP-PCR procedure were sequenced using the gene specific primer used in the second round of PCRs (B/C ext or B/Cd ext). The three gel extracted amplicons generated with gene specfic primer set a (B/C int and B/C ext; Figure 4.7d) were sequenced with the B/C ext primer. The sequences aligned to the Ble2AmChr.SD cassette; however, the sequences did not extend into the *N. gaditana* genome (Table 4.1; z60.3 (a), z20.3 (a) and z0.3 (a)).

The seven gel extracted amplicons generated with gene specific primer set b (B/Cd int and B/Cd ext) were sequenced with the B/Cd ext primer. Two of the sequences obtained (z80.1 (b), z60.1 (b); Table 4.1) aligned with the Ble2AChr.SD cassette, but did not include sequences from the host genome. Another sequence (z20.2 (b); Table 4.1) contained a 123 bp alignment to the Ble2AChr.SD cassette immediately followed in the 3´ direction by a 130 bp sequence that aligned to the *N. gaditana* genome. However, there were no corresponding genes annotated in the genomic region of this alignment. The sequences z60.3 (b), z20.2 (b), z20.3 (b) and z0.3 (b) all contained alignments to the Ble2AChr.SD cassette (130, 111, 123 and 123 bp, respectively) immediately followed by alignments (99.3-100% identity) to the *N. gaditana* genome (321, 146, 20, 285 bp, respectively) that corresponded to four different open reading frames (ORFs) (Table 4.1). These ORFs were taken to be the putative genes trapped by the Ble2AChr.SD vector.

Table 4.1: BLAST analysis of amplicons obtained from Ble2AChr.SD strains using AP-PCR. a, primer set a (B/C int and B/C ext); b, primer set b (B/Cd int and B/C ext).

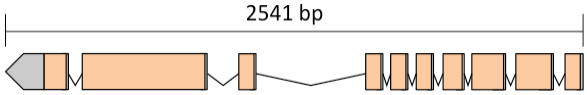
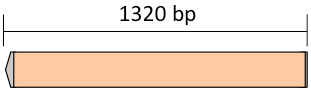
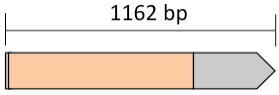
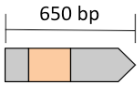
Gene identified	Strain	Query length (bp)	Align. length	%ID	Gene ID	Gene length (bp)	Gene function
No	z60.3 (a)	520	N/A. Sequences only aligned with vector DNA and did not reach *N. gaditana* genomic DNA				
	z20.3 (a)	637					
	z0.3 (a)	647					
	z80.1 (b)	10					
	z60.1 (b)	74					
	z20.2 (b)	130	130	100	N/A. No corresponding gene		
Yes	z60.3 (b)	321	321	99.7	NGA_0640800	2541	D-lactate dehydrogenase
	z20.2 (b)	246	146	99.3	NGA_2110200	1320	Hypothetical protein 1
	z20.3 (b)	20	20	100	NGA_2007210	1162	Chloride channel 7
	z0.3 (b)	285	285	100	NGA_2078400	650	Hypothetical protein 2

4.4.5 Molecular characterization of genes trapped by Ble2Chr.SD cassette

The insertion site of the Bles2AChr.SD cassette in the z60.3 strain corresponded to a 2541 bp protein-coding gene on the reverse strand coding a 746 amino acid (aa) protein annotated as D-

lactate dehydrogenase (NGA_0640800). The insertion site of the cassette in the z20.2 strain corresponded to a 1320 bp protein-coding gene on the reverse strand coding a 304 aa protein annotated as a hypothetical protein (referred to as hypothetical protein 1 hereafter). The insertion site of the cassette in the z20.3 strain corresponded to a 1162 bp protein-coding gene on the forward strand coding a 575 aa protein annotated as chloride channel 7. The insertion site of the cassette in the z0.3 strain corresponded to a 650 bp protein-coding gene on the forward strand coding a 75 aa protein annotated as a hypothetical protein (referred to hypothetical protein 2 hereafter). This information is summarised in Table 4.2.

Table 4.2: Molecular characterization of genes trapped by the Ble2AChr.SD vector in *N. gaditana* strains z60.3, z20.2, z20.3 and z0.3. Right to left gene directionality denotes reverse strand gene; left to right gene directionality denotes forward strand gene; orange, protein-coding region; grey, non-protein-coding region; aa, amino acids).

Strain	Gene of insertion	Protein function and length	Gene structure
z60.3	NGA_0640800	D-lactate dehydrogenase (746 aa)	
z20.2	NGA_2110200	Hypothetical protein 1 (304 aa)	
z20.3	NGA_2007210	Chloride channel 7 (575 aa)	
z0.3	NGA_2078400	Hypothetical protein 2 (75 aa)	

4.4.6 Analysis of mCherry expression level in Ble2AChr.SD strains

The mCherry protein expression levels in the Ble2AChr.SD strains were initially analysed on the basis of mCherry fluorescence in a microplate reader using cultures sampled in the logarithmic growth phase. However, the fluorescence levels measured with filter settings specific to the excitation/emission wavelengths of mCherry (587 and 610 nm, respectively) were undetectable above the endogenous autofluorescence level (data not shown).

To overcome the fluorescence detection issue, mCherry protein levels in the four Ble2AChr.SD strains for which the endogenous genes were identified (z60.3, z20.2, z20.3, and z0.3) were assessed by immunoblot. 20 μg of protein was extracted from these strains at the logarithmic growth phase and probed with the anti-mCherry antibody (Figure 4.8a). Protein bands were detected at the expected size for mCherry (28.8 kDa) in the standards. Protein bands were also detected in z60.3, z20.2 and z20.3, however, the bands migrated more slowly to a position that corresponded to the combined size of the mCherry and Ble protein (14.1 kDa). There were no bands at the expected size migration point for mCherry in the *N. gaditana* samples. This indicates therefore, that the 2A self-cleavage peptide was ineffective resulting in a fusion product of mCherry and Ble. Figure 4.8b shows quantification of the bands in z60.3, z20.2 and z20.3, which corresponded to 3.6×10^{-2}, 5.0×10^{-2} and 6.8×10^{-3}% of total protein, respectively.

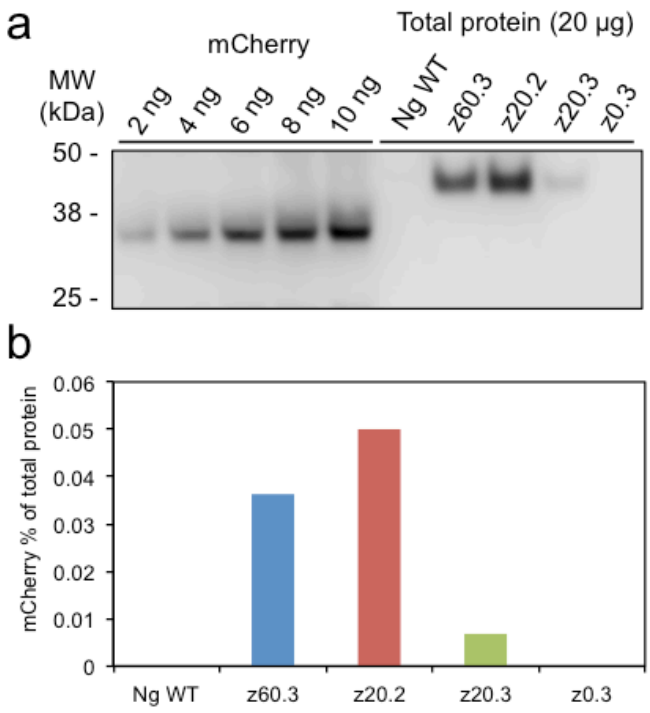


Figure 4.8: Immunoblot analysis of Ble2AChr.SD strains with anti-mCherry antibody. (**a**) Image of immunoblot. Bands were detected in z60.3, z20.2 and z20.3. The larger protein size in the samples as compared to the mCherry standards is due to fusion of mCherry to the Ble resistance protein. (**b**) Quantification of the immunoblot in panel **a** as percentage of total protein using recombinant mCherry standards.

4.4.7 Transcript analysis of genes identified in Ble2AmChr.SD strains in *N. gaditana* B-31

Chromosome scale transcript analysis has been carried out in the *N. gaditana* B-31 isolate in nitrogen replete and deplete conditions (Corteggiani Carpinelli *et al*., 2014). In order to compare the protein expression levels observed in the Ble2AChr.SD strains to the transcript level of the corresponding genes in *N. gaditana* B-31, the protein coding sequences of the genes trapped by the Ble2AChr.SD cassette (Table 4.2) were used to perform a BLASTP (protein BLAST) search to identify the homologs of these proteins in *N. gaditana* B-31. Homologs of the D-lactate dehydrogenase gene, hypothetical protein 1, and chloride channel 7 gene were identified in *N. gaditana* B-31, corresponding to the gene codes Naga_100082g16, Naga_100040g24, and Naga_100020g49, respectively. No homolog could be identified for the hypothetical protein 2 gene. The *N. gaditana* B-31 transcript levels of these genes after 3 and 6 days of growth in both nitrogen replete and deplete conditions are shown in Figure 4.9. The native transcript levels correlate with the relative pattern of mCherry expression levels observed in the Ble2AChr.SD strains (Figure 4.8) in all conditions and at all growth points with the highest expression level

observed for hypothetical protein 1 gene (z20.2), followed by the D-lactate dehydrogenase gene (z60.3) and finally the chloride channel 7 gene (z20.3).

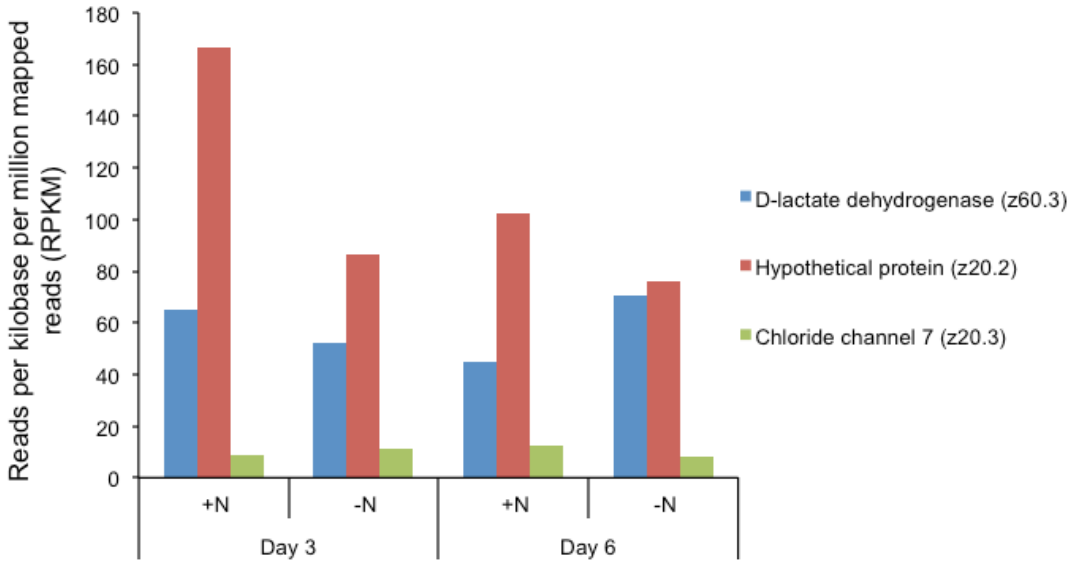


Figure 4.9: Transcript levels of genes identified in Ble2ACHr.SD strains in *N. gaditana* B-31. Transcript abundance (measured in reads per kilobase per million mapped reads (RPKM)) is shown on day 3 and day 6 of growth in liquid culture in both nitrogen replete (+N) and nitrogen deplete (-N) conditions.

4.4.8 Validation of promoter sequences using an eGFP expression cassette

4.4.8.1 Construction of the eGFP expression cassette

The genomic regions preceding the Ble2AChr.SD cassette in the transformants capable of high levels of antibiotic resistance should be capable of driving efficient expression of a transgene. To confirm this, the genomic regions upstream of the ATG codons of the identified genes were amplified from the *N. gaidtana* genome and cloned into enhanced green fluorescent protein (eGFP) expression cassettes (see Table 4.3 and Figure 4.10a). The pLDH, pHYP1, pCC7 and pHYP2 cassettes were transformed into *N. gaditana* and five transformants were isolated for each construct: NgLDH 1-5, NgHYP1 1-5, NgCC7 1-5 and NgHYP2 1-5. The transformants were genotyped for the presence of the *hph* selection marker and integration of the cassette (Figure 4.10b).

Table 4.3: Plasmids constructed for validation of promoter regions identified in Ble2AChr.SD strains.

Plasmid	Plasmid length (bp)	Promoter region and corresponding gene code	Promoter length (bp)
pLDH	11067	D-lactate dehydrogenase (NGA_0640800)	1000
pHYP1	11067	Hypothetical protein 1 (NGA_2110200)	1000
pCC7	11044	Chloride channel 7 (NGA_2007210)	977
pHYP2	11067	Hypothetical protein 2 (NGA_2078400)	1000

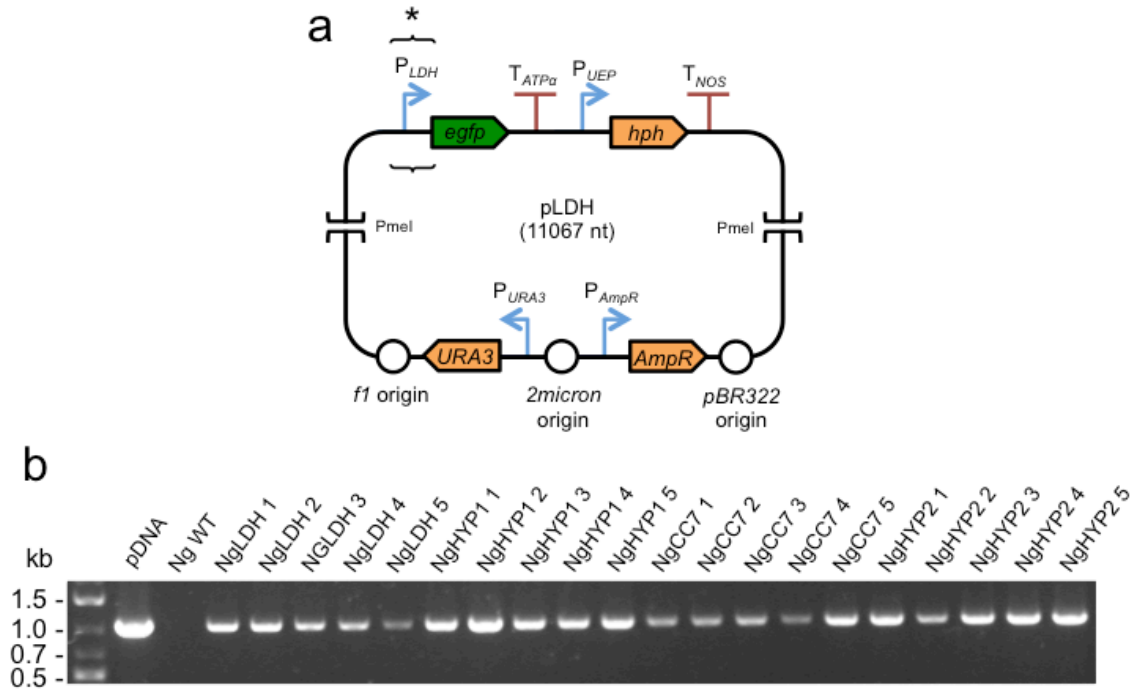


Figure 4.10: eGFP expression vector and genotyping of transgenic lines. (**a**) Example of eGFP expression vector used for validation of endogenous promoters. The pLDH cassette included a 1000 bp region upstream of the gene predicted as D-lactate dehydrogenase (P_{LDH}). The asterisk denotes the location of the promoter region to be validated in each expression cassette. *egfp*, enhanced green fluorescent protein gene; $T_{ATP\alpha}$, ATP synthase alpha subunit; $P_{UEP,}$ ubiquitin extension protein promoter; *hph*, hygromycin B resistance marker; T_{NOS}, NOS terminator. Figure drawn to SBOL standards (Galdzicki *et al*., 2014). (**b**) Genotyping of eGFP expression cassette transformants with *hph* forward and reverse primers to confirm integration of cassette. Lane 1, DNA ladder; lane 2, pDNA, pLDH cassette (positive control); lane 3, *N. gaditana* wild type; lanes 4-23, eGFP expression cassette strains.

4.4.8.2 Fluorescence analysis of eGFP expression cassette strains

To assess the effectiveness of the isolated promoters, eGFP fluorescence was measured in the transformants. The strains were grown to the logarithmic growth phase in standard F2N liquid

culture and analysed on a microplate reader with filter sets specific for eGFP. Figure 4.11a shows the eGFP expression level for all of the transformants. Figure 4.11b shows the mean and standard deviation for the five strains representing each promoter. All of the promoters were capable of driving expression of eGFP (Figure 4.11a). The pattern of expression levels was similar to that previously observed (Figures 4.8 and 4.9) with the highest expression level achieved with the hypothetical protein 1 promoter, followed by the D-lactate dehydrogenase promoter, the chloride channel 7 promoter and, finally, the hypothetical protein 2 promoter.

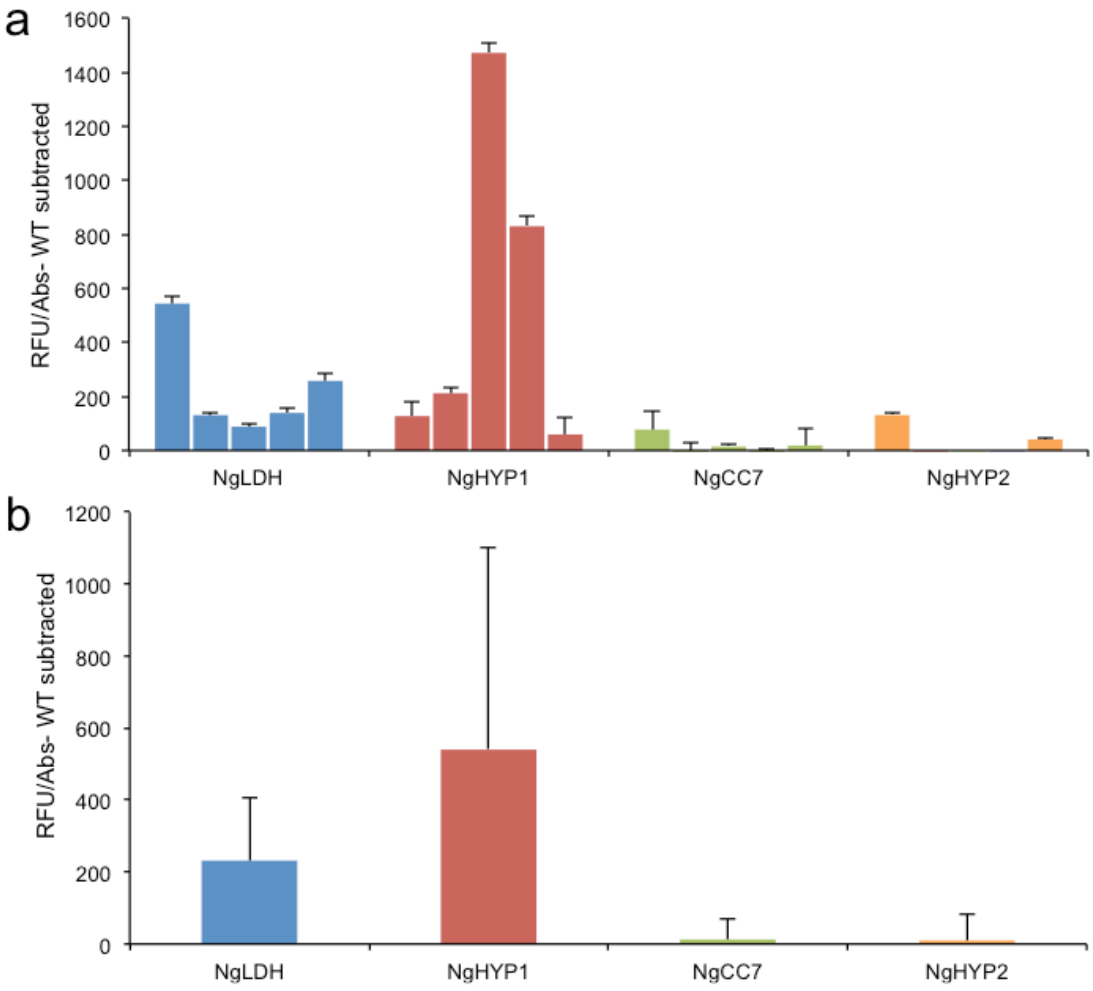


Figure 4.11: eGFP expression level analysis of *N. gaditana* eGFP expression vector cell lines. (**a**) eGFP fluorescence level analysis of five cell lines for each tested promoter: NgLDH, D-lactate dehydrogenase promoter strains; NgHYP1, hypothetical protein 1 promoter strains; NgCC7, chloride channel 7 protein promoter strains; NgHYP2, hypothetical protein 2 promoter strains. Each bar represents one of the five clones of each expression cassette/promoter. Values are means of triplicate measurements, error bars denote SD. (**b**) Mean eGFP fluorescence level for each set of cell lines shown in panel **a**. Values are means of measurements for all five strains, error bars denote SD.

4.5 Discussion

Promoter trapping is a powerful tool for identifying promoters that are highly expressed under specific environmental or nutritional conditions (Blanvillain and Gallois, 2008; Vila *et al*., 2012; Buzas *et al*., 2005; Kiechle *et al*., 2002). Few studies in marine microalgae have been carried out to this end, and these have been exclusively in the model species *C. reinhardtii* with none in biotechnologically relevant species such as *N. gaditana* (Haring and Beck, 1997; Auchincloss *et al*., 1999; Vila *et al*., 2012).

We have shown that transformation of *N. gaditana* with the Ble2AChr.SD promoter trapping cassette and selection on the antibiotic zeocin can be used to identify highly expressed endogenous promoters in *N. gaditana*. Figure 4.2 shows that stepwise increases in the concentration of zeocin result in a stepwise reduction in the number of viable transformed clones from 96/96 viable clones at a zeocin concentration of 0 μg/mL to 12/96 viable clones at a zeocin concentration of 80 μg/mL. Integration of the Ble2AChr.SD cassette was subsequently confirmed by PCR in the fifteen strains that were carried forward for further analysis (Figure 4.3).

3´ RACE was initially performed to identify the mRNA transcript containing the Ble2AChr.SD cassette and hence the trapped promoter. The 3´ RACE technique was ineffective in the conditions tested. Several factors may have determined the success of the 3´ RACE procedure, including the distance within the cDNA transcript template between the gene specific primer and the adaptor primer used in the RACE PCRs, the GC content of the transcript of integration, the abundance of the transcript of integration, degradation of RNA and incomplete cDNA synthesis (Bower and Johnston, 2010). The AP-PCR strategy was successful in amplifying eight amplicons from the fifteen strains. Sequencing of these amplicons allowed the identification of genomic DNA sequences flanking the Ble2AChr.SD cassette in four of the amplicons and corresponding transformants (Table 4.2). Similar to 3´RACE, AP-PCR reactions are not always successful (Bahrani-Mougeot *et al*., 2002; Knobloch *et al*., 2003; Burall *et al*., 2004). The arbitrary primers used in the AP-PCRs carried out in this study were initially tested as part of a panel of arbitrary primers in the bacterium *Staphylococcus epidermidis* (Das *et al*., 2005). The goal of Das *et al*. in this study was to optimize the parameters of the AP-PCRs. They concluded that template and primer concentrations were the most critical factors for efficiency (Das *et al*., 2005). All of the template and primer concentrations shown to be most optimal in their study were used in this study. The efficiency of the AP-PCRs was also reported to be affected by a pentamer at the 3´ end of the arbitrary primers and the GC content of the host genome, which was the basis for the selection of their panel of primers (Das *et al*., 2005). Other researchers have also advised designing arbitrary primers with 3´ pentamers that are frequently present in

the genome being analysed (Knobloch *et al*., 2003; O'Toole *et al*., 1999). Until more comprehensively optimized systems for identifying vector insertion sites have been established in marine microalgal systems, large panels of transformants should be screened in insertional mutagenesis-based approaches in order to account for low efficiencies.

An mCherry fluorescent reporter protein was expressed as part of the Ble2AChr.SD cassette; however, mCherry fluorescence was not detectable above the level of wild-type autofluorescence, even in clones capable of overcoming high concentrations of zeocin. This may suggest that mCherry is not a suitable fluorescent reporter for use in *N. gaditana*. However, it should be noted that the immunoblot undertaken on the transformants using an antibody binding to the mCherry protein (Figure 4.8) indicated that that the 2A self-cleavage peptide was not functional in this system. The resulting fusion of the mCherry fluorescent reporter to the Ble resistance protein may have affected the brightness of the mCherry fluorescence and hence detection of mCherry.

The immunoblot performed on the four gene identified strains (Figure 4.8) revealed that the highest level of mCherry protein expression was in z20.2 (5.0×10^{-2}% of total protein), which was resistant to a concentration of 20 μg/mL of zeocin. This was higher than the mCherry protein expression in z60.3 (3.6×10^{-2} % of total protein), which was resistant to a zeocin concentration of 80 μg/mL of Zeocin. This inconsistency may be due to the respective functions of the genes that were trapped, and concurrently knocked out, in each strain; i.e. the growth rate of z20.2 may have been negatively affected by the disruption of the predicted hypothetical protein coding gene to a lower degree than the effect on growth of the disruption of the D-lactate dehydrogenase gene in z60.3. Conversely, the disruption of the D-lactate dehydrogenase gene may have conferred a growth advantage, enabling z60.3 to overcome a higher level of zeocin with a lower expression level compared to z20.2. The presence of larger protein bands in the Ble2aChr.SD strains compared to the standards suggested that the 2A self-cleavage peptide is not cleaved in *N. gaditana*, generating a fusion product of Ble and mCherry. Despite the benefits of the 2A system, namely the production of separate gene products in a 1:1 ratio, it has not be widely adopted due to the lack of publicly available cloning vectors harbouring 2A sequences and the lack of comparative analysis of efficiencies of different 2A sequences, which have been shown to vary in different contexts (Kim *et al*., 2011). Consequently, systematic comparison of 2A peptide cleavage efficiencies is an ongoing area of research (Liu *et al*., 2017b). Analysis of the native transcript levels in *N. gaditana* B-31 of genes homologous to those trapped in the Ble2AChr.SD strains confirmed the same pattern on expression as observed in the Ble2AChr.SD strains (Figure 4.9). This reinforces that both the predictions of the trapped genes and analysis of the expression levels in the Ble2AChr.SD strains was accurate.

eGFP expression cassettes were constructed containing approx. 1000 bp regions upstream of the identified genes in order to validate the efficiency of the identified promoters. In this case, due to the availability of the *N. gaditana* genome, forward and reverse gene specific primers specific to the readily available genomic sequences could easily be identified. In instances where the genome of the host organism is not available, alternative methods for amplifying the genomic promoter region must be adopted, such as 5´ RACE (Walker *et al*., 2005a). The eGFP fluorescence analysis was consistent with the immunoblot (Figure 4.8) and transcript (Figure 4.9) analysis and showed that all of the promoters were capable of driving the expression of eGFP to levels detectable above wild type autofluorescence; however, the D-lactate dehydrogenase and hypothetical protein 1 promoters were the most highly expressed (Figure 4.11).

In conclusion, we have established the first promoter trapping strategy for *N. gaditana*, which is a valuable addition to the genetic toolkit for this host. Additionally, this work provides a foundation for the development of promoter trapping strategies in other non-model, biotechnologically relevant species of marine microalgae. The development of similar strategies in other non-model systems will be aided by improvement of PCR technologies capable of more reliable amplification of genomic sequences flanking vector insert sites to enable more efficient identification of trapped genes. The future direction of this research should also be to use the antibiotic concentration screening procedure (see Figure 4.2) in a variety of different, commercially relevant nutritional and environmental conditions in order to identify novel promoters that are highly expressed under these conditions. This could include reduced nitrogen content in the media or variations in temperature or irradiance.

Chapter 5: High-throughput phenotypic screening of *N. gaditana* mutants for biofuel production

5.1 Introduction

Microalgae are the largest group of photosynthetic organisms in the world and are a potentially viable feedstock for biofuels due to their fast growth rates, ability to convert sunlight to biomass, and ability to grow in saline water on non-arable land (Wijffels and Barbosa, 2010; Barbosa and Wijffels, 2013; Hu *et al*., 2008). The ability of microalgae to adapt to a broad range of environmental conditions such as, temperature, salinity, irradiance and nutrient availability, is reflected in their diverse lipid composition and ability to synthesis numerous other storage compounds (Thompson, 1996). It was initially believed that lipids played two roles within algal cells: as energy reserves and structurally in the lipid bilayers that compartmentalize subcellular organelles. However, it has since been realized that different lipid classes participate in various independent physiological functions such as signal transduction (Guschina and Harwood, 2006).

A panel of candidate strains for biofuel production have emerged, with lipid contents as high as 75 % of dry mass in the form of triacylglycerols (TAGs); TAGs can be converted to biodiesel through esterification (Franz *et al*., 2013). Microalgae derived biofuels constitute the third generation of biofuels and have an inherent advantage over terrestrial first and second generation biofuels in that there is a lesser demand on resources for the synthesis of structural components (Barbosa and Wijffels, 2013). However, to fully realize this potential and make microalgae an economically viable biofuel feedstock, more detailed understanding of key metabolic pathways is required to improve growth rates and cellular lipid content.

Nitrogen deprivation induces an upregulation in lipid content in many strains of microalgae, which has been explored as a means of enhancing production rates for biofuel growth systems (Figure 5.1); however, this upregulation is concurrent with an overall reduction in growth and biomass accumulation (Boussiba *et al*., 1987).

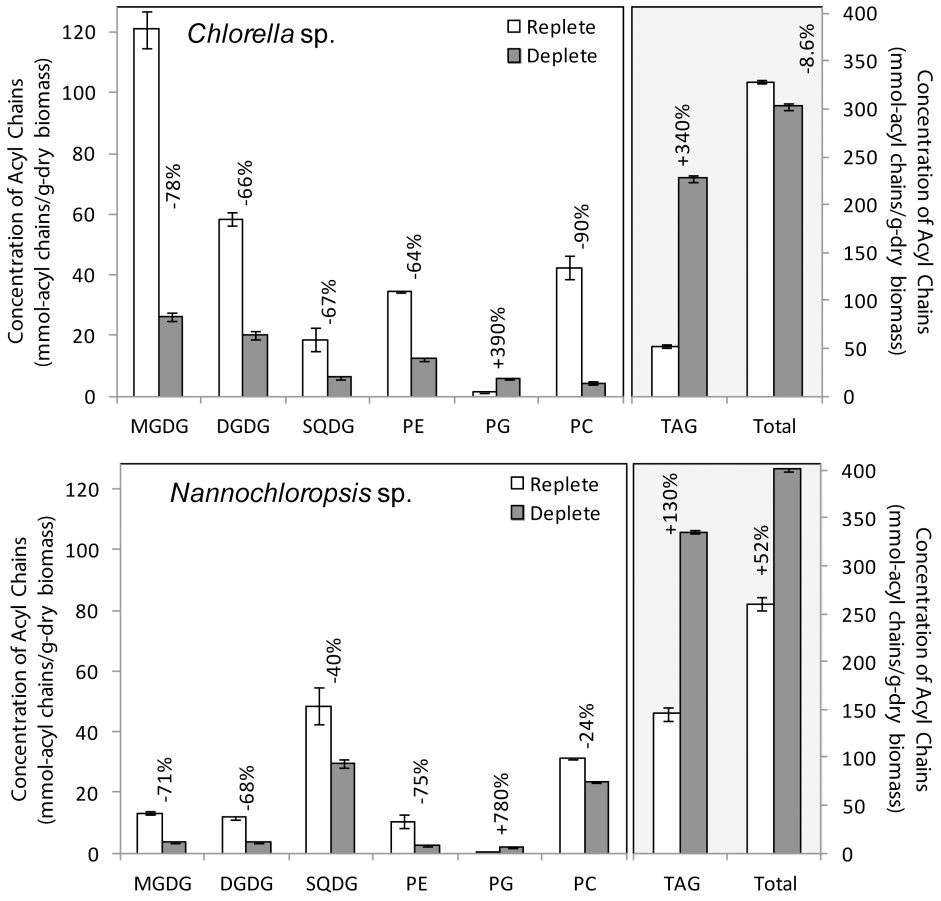


Figure 5.1: Total acyl chain concentrations within different lipid classes recovered from nitrogen replete (5 mM $NaNO_3$ media) and deplete (0.5 mM $NaNO_3$ media) *Chlorella* sp. and *Nannochloropsis* sp. MGDG, monogalactosyldiacylglycerol; DGDG, digalactosyldiacylglycerol; SQDG, sulfoquinovosyldiacylglycerol; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PC, phosphatidylcholine; TAG, triacylglycerides; total, total lipid. Error bars are standard deviations of triplicate measurements. Figure from Martin *et al*. (2014).

Genetic engineering offers another means by which to increase oil content. The alteration of specific genes and subsequent analysis of the resulting phenotypes, so called 'reverse genetics', has been used successfully to generate plants with increased/altered oil content (Yu *et al*., 2011). Over-expression of a homomeric acetyl-coenzyme A carboxylase (ACCase) from *Arabidopsis* has been shown to increase the fatty acid composition of *Brassica napus* seeds by 5 % (Roesler *et al*., 1997). Overexpression of cytosolic glycerol-3-phosphate dehydrogenase (*gpd1*) in *B. napus* has been reported to increase seed oil content by 40 % (Vigeolas and Geigenberger, 2004; Vigeolas *et al*., 2007). However, few attempts to use reverse genetic approaches in microalgae have lead to the generation of improved/altered phenotypes for oil production (Kilian *et al*., 2011; Boussiba *et al*., 1987; Yu *et al*., 2011). Deletion of genes in the starch biosynthesis pathway has been shown to lead to a reallocation of resources to lipid accumulation in the model organism *C. reinhardtii* yielding an increase in lipid production (Wang *et al*., 2009; Siaut *et al*., 2011; Li *et al*., 2010). However, similar reverse genetic

strategies have not been explored in non-model biofuel candidate strains until recently (Ajjawi *et al*., 2017).

An alternative approach is to generate libraries of mutant strains with altered phenotypes and screen these libraries for the desired phenotype before probing the genome for the associated genetic modification, so called 'forward genetics' (Hlavova *et al*., 2015). Such approaches have the advantage of being non-biased and can be carried out without prior knowledge of gene function, which is not always available (Hlavova *et al*., 2015). Perin *et al*. generated a library of *N. gaditana* mutants using both chemical and insertional mutagenesis screened this library for alterations to the photosynthetic apparatus (Perin *et al*., 2015). One of the isolated strains exhibited both a reduced cellular chlorophyll content and improved photosynthetic activity, conferring a 21 % increase biomass productivity in the artificial conditions tested (Perin *et al*., 2015). Additionally, many high-lipid strains of *C. reinhartdii* have been isolated using forward genetic screens (Terashima *et al*., 2015; Lee *et al*., 2014). Aside form biofuels; random mutagenesis has been used to obtain algal strains with enhancements in other high-value compounds. *D. tertiolecta* strains with 10-15% increases in zeaxanthin, a carotenoid used in the prevention of degenerative diseases, have recently been isolated (Kim *et al*., 2017).

This chapter describes the use of insertional mutagenesis and phenotypic screening to generate mutant strains of *N. gaditana* with (i) increased resistance to reactive oxygen species and in turn tolerance to increased irradiances and potentially an overall improvement in growth, and (ii) increased lipid content. The rational for the first screen is to increase lipid productivity by improving the growth characteristics of the alga while the former aims to increase lipid productivity the increasing the lipid content of the cell. The use of insertional mutagenesis via the pAlg.Hyg.SD gene-trap vector to disrupt genes through the insertion of a known nucleotide sequence provides an anchor for PCR, enabling the possibility of identifying the effected genes in the mutant strains using PCR-based techniques (see Figure 5.2).

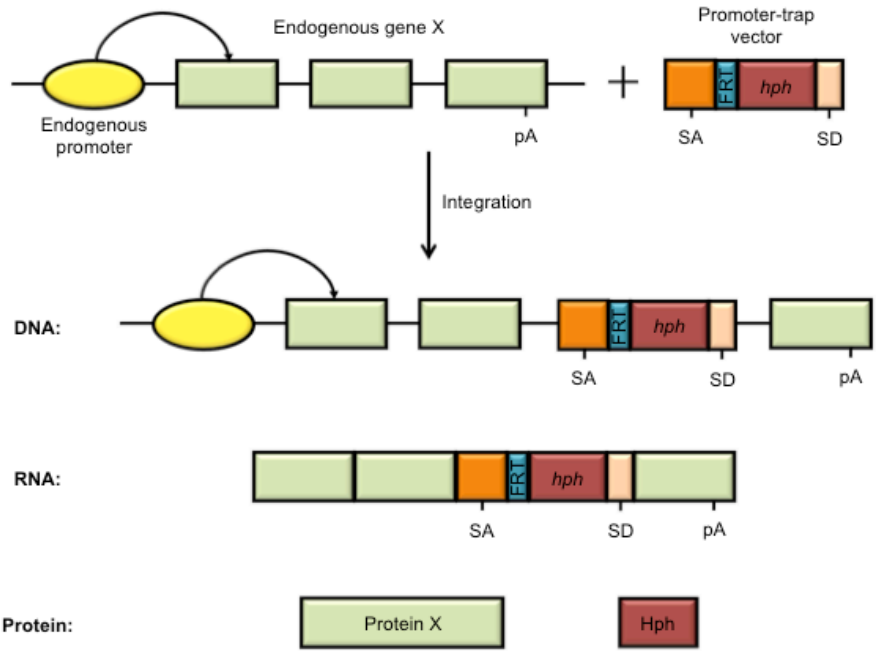


Figure 5.2: Integration of the pAlg.Hyg.SD 'gene-trap' vector, used for insertional mutagenesis and phenotypic screening. The vector sequence contains: a splice acceptor site (SA) and splice donor site (SD), a *hph* gene encoding the hygromycin-B-phosphotransferase resistance protein (Hph), conferring resistance to hygromycin B, and a Flp-*FRT* sequence to prevent translational fusion of Hph to the truncated endogenous protein X. The SA and SD sequences direct integration into endogenous gene X generating a fusion transcript coding for Hph in addition to knocking out endogenous gene X.

5.2 Results

5.2.1 Screening for increased resistance to reactive oxygen species

Absorption of light is essential for photosynthetic growth. However, drastic and sudden changes in irradiance, from high irradiances in full sunlight (approx. 2000 µmol photons m^{-2} s^{-1}) to shaded conditions with little light exposure, can cause damage to the photosynthetic machinery (Erickson *et al*., 2015). Initially, as irradiance intensity increases, the photosynthetic efficiency increases towards its maximal rate (P_{max}) (Björkman and Demmig, 1987). As irradiance intensity increases, photosynthesis becomes saturated, but excess light is still absorbed, beyond the capacity of photosynthesis (Figure 5.3). If this over-excitation cannot be dissipated, chemical intermediates and byproducts are produced that can cause photo-oxidative damage to photosystem II (PSII) and photosystem I (PSI) in addition to other proteins, nucleic acids or lipids within the cell (Niyogi, 1999). This damage results in photoinhibition; a decrease in the rate of photosynthesis and/or P_{max} (Kok, 1956). In addition to irradiance, ROS (reactive oxygen species) production is also regulated by temperature, pH and iron availability; depletion of iron

and increases in temperature have been shown to enhance ROS (Twiner and Trick, 2000; Liu *et al*., 2007). Microalgae have been shown to release large quantities ROS into their environment; ROS has also been implicated in the mortalities of fish, protists and bacteria (Oda *et al*., 1997; Marshall *et al*., 2002; Ishimatsu *et al*., 1996; Evans *et al*., 2006; Flores *et al*., 2012).

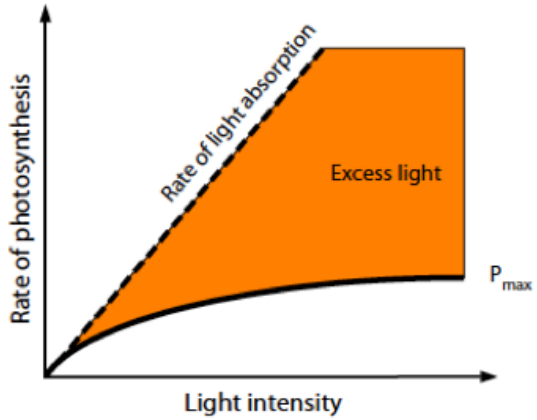


Figure 5.3: The effect of increasing light intensity on rates of photosynthesis. Initially, light absorption is limiting and photosynthetic efficiency is maximal. As light increases and photosynthesis saturates at the maximum rate of photosynthesis (P_{max}), excess light absorption leads to reduced rates of photosynthesis and reductions in P_{max} in extreme cases. Figure adapted from Erickson *et al*. (2015).

ROS, formed in aerobic conditions, are the most prevalent reactive byproducts of photosynthesis and include radical molecules, superoxide ($O_2^{\bullet-}$) and hydroxyl radicals ($OH^{\bullet}$), and non-radical species, singlet oxygen ($^1O_2$) and hydrogen peroxide (H_2O_2) (Suzuki *et al*., 2012). Singlet oxygen is generated mainly at the PSII reaction centre and is the main source of ROS-induced photo-oxidative damage (Krieger-Liszkay, 2005). H_2O_2, $O_2^{\bullet-}$, and $OH^{\bullet}$ are produced mainly at the acceptor side of PSI, through electron transfer to O_2 by NADP and ferredoxin or the PSI reaction centre (Asada, 2006; Tjus *et al*., 2001).

In high light intensities the surface layer of cells in a commercial mass culture of microalgae would absorb light far in excess of the rate that photosynthesis could dissipate them, subjecting these cells to photoinhibition caused by excessive ROS production (Niyogi, 1999; Holt *et al*., 2004; Powles, 1984; Mellis, 1999). Thus, it is conceivable that microalgal strains with increased resistance to ROS may also exhibit improved growth and enhanced productivities in mass culture.

5.2.1.1 Screening procedure

In order to screen for mutant strains of *N. gaditana* with increased resistance to ROS, menadione was used to induce oxidative stress. Redox cycling of menadione promotes the

conversion of triplet oxygen to superoxide anions (Figure 5.4) (Fahlbusch, 2003; Loor *et al*., 2010; Criddle *et al*., 2006; Castro *et al*., 2007).

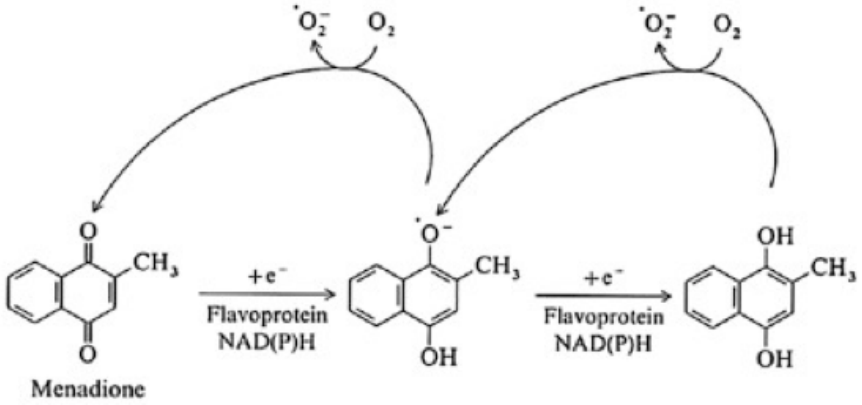


Figure 5.4: Redox cycling of menadione and induction of oxidative stress by reduction of triplet oxygen (O_2) to superoxide radical anions ($O_2^{\cdot-}$). NAD(P)H, reduced nicotinamide adenine dinucleotide phosphate. Figure adapted from Fahlbusch (2003).

Preliminary trials were carried out to determine the minimum concentration of menadione that would cause full die back of wild type cells in F2N agar plates (1.5 % agar; 50 % artificial sea water; 2 mM NaCl; 5 x 10^7 cells/plate) (Table 5.1). Concentrations of 2, 2.5, 3, 3.5, and 4 µg/mL of menadione were initially tested. Full killing was achieved at a concentration of 3 µg/mL, while growth was observed at a concentration of 2.5 µg/mL. A narrower range of concentrations was subsequently tested: 2.6, 2.7, 2.8 and 2.9 µg/mL. A concentration of 2.9 µg/mL of menadione was determined to be the minimum concentration to result in full killing of wild type cells. Transformations of *N. gaditana* were then carried out in the standard conditions (see methods), except that 2.9 µg/mL of menadione was included in the post-transformation plates in order to select for mutant strains with resistance to menadione above the level of the wild type. Ten transformation plates were inoculated, which yielded two colonies. These cell lines with putative increases in resistance to menadione, referred to as NgROS1 and NgROS2 hereafter, were carried forward for further analysis.

Table 5.1: Sensitivity trials to identify minimum concentration of menadione required to cause full die back of *N. gaditana* cells at a concentration of 5 x 10^7 cells/plate. A concentration of 2.9 µg/mL was the minimum concentration tested to give full killing of cells. Three plates were inoculated at each concentration and monitored for growth over thirty days.

Menadione concentration (µg/mL)	Growth
2	+
2.5	+
2.6	+
2.7	+
2.8	+
2.9	-
3	-
3.5	-
4	-

5.2.1.2 Genotyping of NgROS1 and NgROS2

NgROS1 and NgROS2 were genotyped by PCR with the *hph* forward reverse genotyping primers to confirm integration of the pAlg.Hyg.SD cassette (Figure 5.5). The 435 bp amplicon was generated in both NgROS1 and NgROS2 and was absent in the wild type.

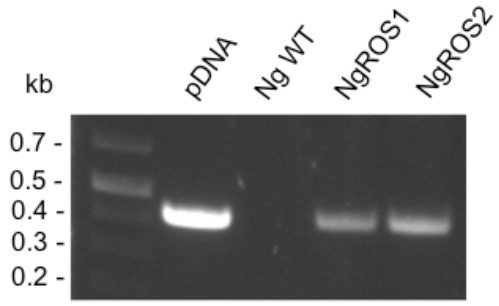


Figure 5.5: Genotyping of NgROS1 and NgROS2 with the *hph* genotyping primer set to confirm integration of the pAlg.Hyg.SD cassette. The 435 bp amplicon was present in NgROS1 and NgROS2 and the positive control plasmid DNA (pAlg.Hyg.SD) reaction but was absent in the *N. gaditana* wild type. Template DNA: lane 1, DNA ladder; lane 2, pDNA; lane 3, wild type; lane 4, NgROS1; lane 5, NgROS2.

5.2.1.3 NgROS1 and NgROS2 – reactive oxygen species resistance tests

Spot tests were performed to confirm that NgROS1 and NgROS2 had increased resistance to menadione. The *N. gaditana* wild type, NgROS1 and NgROS2 were grown to the logarithmic

phase, normalized to the same $A_{750\,nm}$ value and diluted 1:10 in F2N media. 5 µL spots were added to standard F2N agar plates containing menadione at concentrations ranging from 0 to 0.5 µM (Figure 5.6a). At a concentration of 0.2 µM menadione more growth was observed for NgROS1 and NgROS2 as compared to the wild type. No wild type growth was present at a concentration of 0.5 µM menadione; growth was observed at this concentration for NgROS1 and NgROS2 confirming that both strains had elevated levels of resistance in these conditions.

To determine if NgROS1 and NgROS2 had increased levels of resistance to other reactive oxygen species in addition to the superoxide anions generated by menadione. The spot tests were performed in the same way except that menadione was switched for hydrogen peroxide (H_2O_2) at concentrations ranging from 0 to 60 µg/mL (Figure 5.6b). The absence of growth of the wild type at a concentration of 60 µg/mL hydrogen peroxide and the presence of growth for NgROS1 and NgROS2 at the same concentration demonstrated that both strains also had elevated levels of resistance to hydrogen peroxide in these conditions. It appeared from the spot tests that growth was slightly higher for NgROS2 as compared to NgROS1.

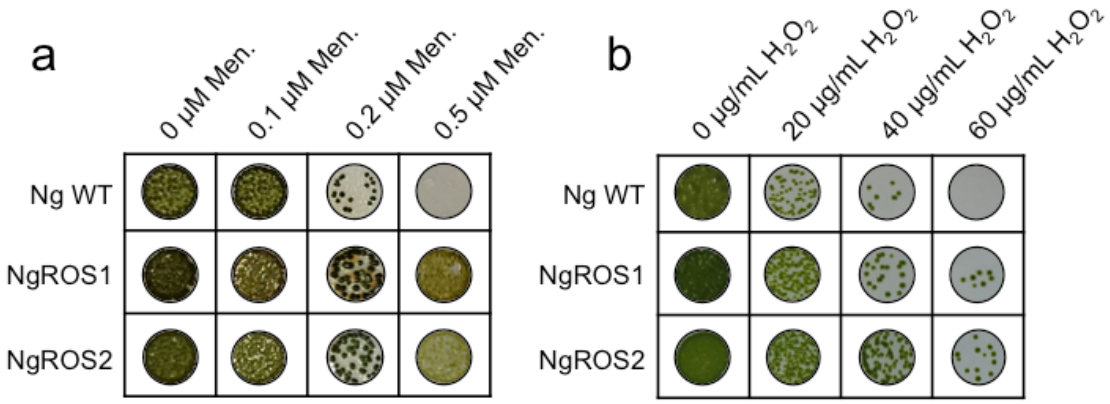


Figure 5.6: NgROS1 and NgROS2 resistance tests with menadione and hydrogen peroxide (H_2O_2). *N. gaditana* spot were 1:10 dilutions of normalized logarithmic phase cultures. NgROS1 and NgROS2 both exhibited elevated levels of resistance to both menadione and hydrogen peroxide. Spots are representatives of biological duplicate tests.

5.2.1.4 NgROS1 and NgROS2 – high light tolerance tests

To test if the increased resistance to reactive oxygen species in NgROS1 and NgROS2 also conferred an increased tolerance to high light, a series of experiments were undertaken in range of light treatments.

5.2.1.4.1 High light transition growth experiment

Initially, a growth experiment was inoculated with six replicate cultures of the wild type, NgROS1 and NgROS2 in 80 µmol photons m^{-2} s^{-1} light. After four days, three of replicate cultures of each strain were transferred to high light (1000 µmol photons m^{-2} s^{-1}). Over the course of the experiment, culture density ($A_{540\ nm}$), photosynthetic efficiency (F_v/F_m) and absorption cross-section of photosystem II (σ_{PSII}; nm^2) was measured every day (Figure 5.7).

After the light transition on day four, all strains exhibited a reduction in growth and σ_{PSII} and initially a reduction in F_v/F_m (Figure 5.7). In both the control and high light cultures, the highest cell density reached after the light transition on day eight was by NgROS2 followed by NgROS1; the wild type exhibited the lowest cell density by day eight (Figure 5.7a). In the high light cultures, the final cell density on day eight was significantly higher in NgROS1 ($p < 0.05$, Student's *t*-test) and NgROS2 ($p < 0.05$, Student's *t*-test) as compared to the wild type.

Within the control light group, the F_v/F_m (Figure 5.7b) of NgROS1 remained fairly constant (0.533 on day two, 0.528 on day eight). The wild type showed a slow decline in F_v/F_m (0.527 on day two, 0.448 on day eight). The F_v/F_m of NgROS2 also declined over the course of the experiment (0.541 on day two, 0.482 on day eight). In the high light group, the wild type had the lowest F_v/F_m on day five after the light transition (0.474); the F_v/F_m of the NgROS1 and NgROS2 cultures on day five was 0.508 in both cases. However, by the end of the experiment the F_v/F_m values for the high light cultures had converged at a similar level: 0.462, 0.452 and 0.482 for the wild type, NgROS1 and NgROS2 respectively.

On day four the σ_{PSII} values (Figure 5.7c) in all cultures were virtually the same (ranging from 1.12 – 1.15 nm^2). After the light transition on day five, the σ_{PSII} values of the high light cultures collectively decreased as compared to the control light group (ranging from 0.93 to 0.10 nm^2 in the control light group and 0.78 to 0.83 nm^2 in the high light group). There was little difference in the σ_{PSII} values for between strains within each light condition by day eight.

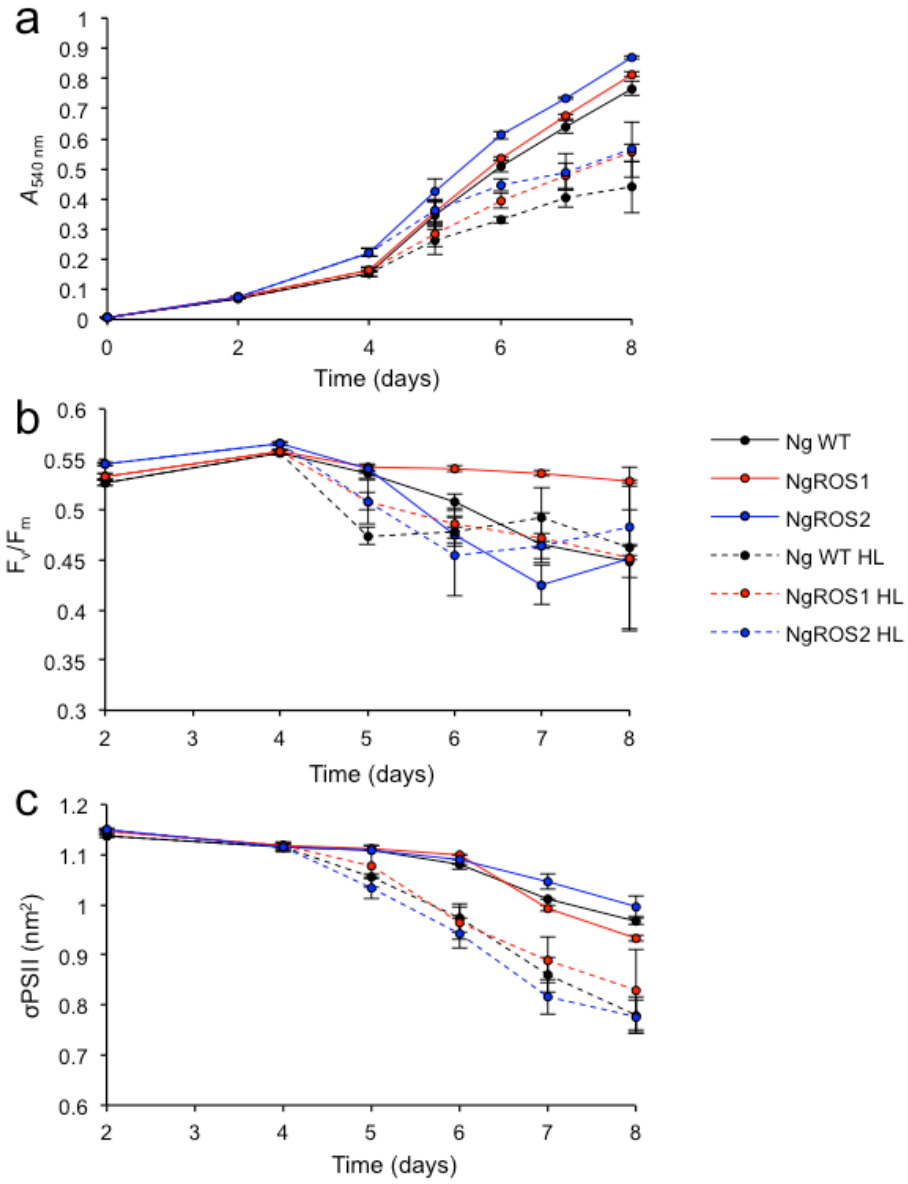


Figure 5.7: Growth and photosynthetic characteristics of *N. gaditana* wild type, NgROS1 and NgROS2 over eight days, before and a after a transition from a medium light intensity (80 µmol photons m^{-2} s^{-1}) to a high light intensity (1000 µmol photons m^{-2} s^{-1}) on day four. (**a**) Cell density measured as absorbance at a wavelength of 540 nm ($A_{540\,nm}$). (**b**) Photosynthetic efficiency (F_v/F_m,). (**c**) Absorption cross-section of photosystem II (σ_{PSII}; nm^2). Error bars denote SD of triplicate measurements.

5.2.1.4.2 One hour high light treatment

In order to gain a better understanding of the immediate effects of severe changes in irradiance on photosynthesis, on a more immediate time scale, a second experiment was undertaken with the wild type, NgROS1 and NgROS2. Six replicate cultures of each strain were grown to the logarithmic phase in 80 µmol photons m^{-2} s^{-1} light, before three replicate cultures of each strain were transferred to very high light (1800 µmol photons m^{-2} s^{-1}) for one hour. After the one hour treatment, the F_v/F_m and σ_{PSII} (nm^2) of each strain was measured (Figure 5.8).

In all strains the, the high light treatment caused a reduction in F_v/F_m. Within the control light group, the F_v/F_m values (Figure 5.8a) were not statistically significantly different between the wild type and NgROS1 ($p = 0.15$, Student's *t*-test) but were statistically significantly different between the wild type and NgROS2 ($p = 0.020$, Student's *t*-test). Within the high light treatment group, F_vF_m was 8.03 % higher ($p = 0.000087$, Student's *t*-test) for NgROS1 (0.523) as compared to the wild type (0.481). The F_v/F_m of NgROS2 (0.529) was also significantly higher (9.07 %; $p = 0.00005$, Student's *t*-test) than the wild type. Within the control light group, the σ_{PSII} values (Figure 5.8b) were not statistically significantly different between the wild type and NgROS1 ($p > 0.05$, Student's *t*-test) or between the wild type and NgROS2 ($p > 0.05$, Student's *t*-test). The same was true in for the strains in the high light treatment group. However, the σ_{PSII} value of the wild type was 1.23 % lower in the high light group (1.07 nm^2), as compared to the control group (1.09 nm^2), the values were not statistically significantly different. Conversely, the σ_{PSII} value of NgROS1 was 5.67 % higher ($p = 0.0012$, Student's *t*-test) in the high light group (1.07 nm^2) as compared to the control group (1.01 nm^2). Similar to NgROS1, the σ_{PSII} value of NgROS2 was 7.34 % higher ($p = 0.0036$, Student's *t*-test) in the high light group (1.07 nm^2) as compared to the control group (1.02 nm^2).

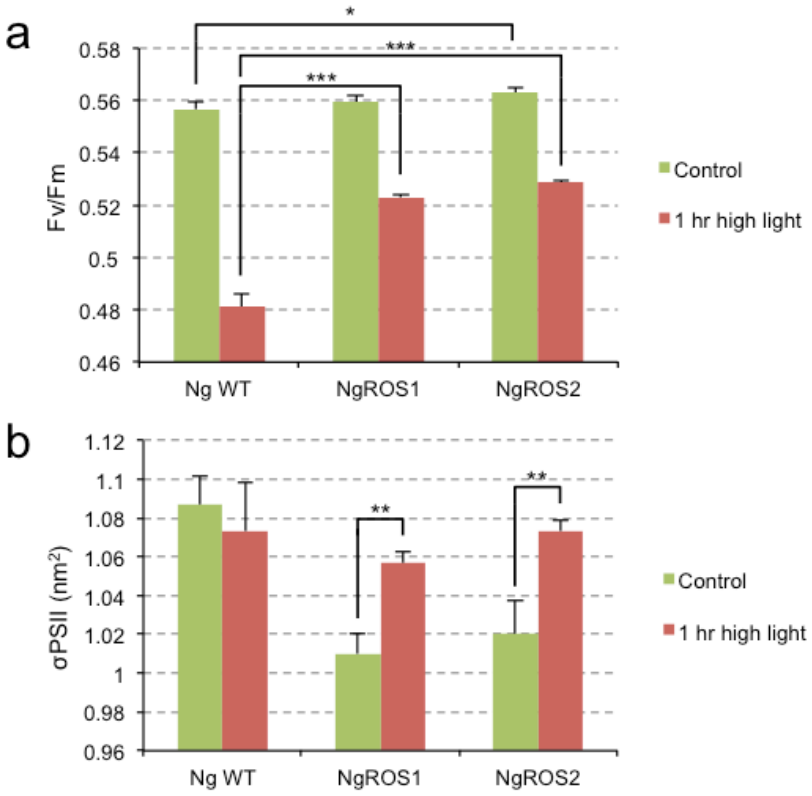


Figure 5.8: Response of *N. gaditana* wild type, NgROS1 and NgROS2 to one hour high light treatment**.** (**a**) Photosynthetic efficiency (F_v/F_m) and (**b**) absorption cross-section of photosystem II (σ_{PSII}; nm^2). The light levels in the control and high light groups were 80 and 1800 µmol photons m^{-2} s^{-1} respectively. Error bars denote SD of triplicate measurements; *, $p < 0.05$, **, $p < 0.05$ ***, $p < 0.001$.

5.2.1.4.3 Sequential light challenge

A rapid light curve (RLC) sequential light challenge was designed to further test the tolerance of NgROS1 and NgROS2 to increased and rapidly changing irradiances (see section 2.2.6 for information on FRRF procedures). The RLC consisted of a 60 second pre-RLC dark period to assess the baseline F_v/F_m of the strains followed by 12 light/dark steps (each light interval lasted 240 seconds at an irradiance of 2000 µmol photons m^{-2} s^{-1}; each dark interval lasted 30 seconds), and finally a 1200 second post-RLC recovery dark period. F_v/F_m was continually measured throughout the RLC with the expectation that the light/dark steps would cause photodamage to the cells, which would be observed by a reduction in F_v/F_m in each subsequent dark step, and that over the 1200 post-RLC recovery interval, F_v/F_m would increase towards the level of the baseline measurements provided full repair of any photodamaged PSII occurred. An indication of the long-term photodamage can then been assessed by comparing the baseline F_v/F_m values measured at the pre-RLC interval to the F_v/F_m values measured during and at the end of the recovery period.

The sequential light challenge was performed on triplicate cultures of the *N. gaditana* wild type, NgROS1 and NgROS2, sampled at the logarithmic phase (Figure 5.9). The first F_v/F_m measurements for each strain were similar; 0.587, 0.593 and 0.583 in the wild type, NgROS1 and NgROS2 respectively. These F_v/F_m values at this time point were not significantly different to the wild type for both NgROS1 ($p = 0.10$, Student's *t*-test) and NgROS2 ($p = 0.18$, Student's *t*-test). At the twelfth dark step, F_v/F_m values had reduced to approx. 0.300 in all strains, indicating a high level of non-photochemical quenching and/or photoinhibition reducing the photochemical efficiency following short-term dark exposure. Over the post-RLC recovery period, NgROS2 showed the fastest recovery, followed by NgROS1 and then the wild type. The final F_v/F_m measurement at the end of the recovery period was 0.504 for the wild type, 0.523 for NgROS1 and 0.528 for NgROS2. These results were statistically significantly higher than the wild type in the case of NgROS1 ($p = 0.025$, Student's *t*-test) and NgROS2 ($p = 0.010$, Student's *t*-test).

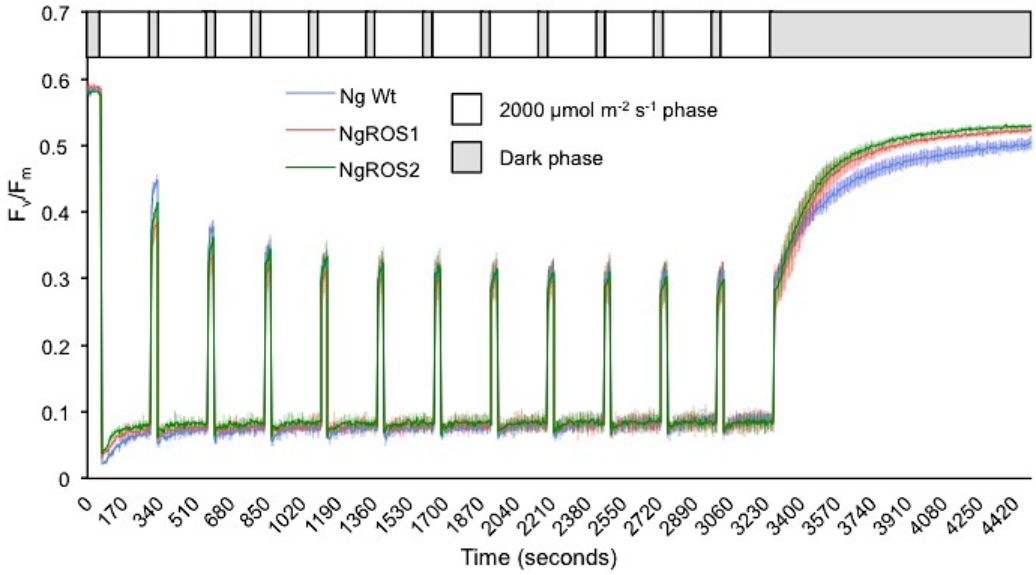


Figure 5.9: Sequential light challenge (rapid light curve; RLC) analysis of *N. gaditana* wild type (blue trace), NgROS1 (red trace) and NgROS2 (green trace). Photosynthetic efficiency (F_v/F_m) was measured every 5 seconds over the course of the RLC, which consisted of a 60 second pre-RLC dark period, 12 light/dark steps (each light interval lasted 240 seconds at an irradiance of 2000 µmol photons m^{-2} s^{-1}; each dark interval lasted 30 seconds), and a 1200 second post-RLC recovery dark period. Values are means (solid lines) and SD (lighter shading) of triplicate measurements. Light (clear areas) and dark (shaded areas) are denoted at the top of the plot.

5.2.1.5 Reactive oxygen species assay

The increased growth of NgROS1 and NgROS2 in the presence of superoxide anions and hydrogen peroxide (Figure 5.6) and increased tolerance to high light (Figure 5.7, 5.8 and 5.9)

could be explained by mutation conferring resistance to ROS. However, it is possible that these phenotypes could be a result of reduced rates of ROS production in NgROS1 and NgROS2. To test this, a ROS assay was carried out with triplicate cultures of wild type, NgROS1 and NgROS2 cultures, logarithmically grown in 80 µmol photons m^{-2} s^{-1} light (Figure 5.10). 5-(and-6)-carboxy-2',7'-dihydrofluorescein diacetate (H_2DFFDA) was used as an indicator for ROS (Szivák *et al*., 2009). Upon oxidation, H_2DFFDA is converted to fluorescent DFFDA, which can be detected at a 488 nm excitation wavelength with a 520 nm emission filter using a microplate reader. The relative fluoresces units (RFU) measured in the wild type was 151. The RFU was not significantly different in NgROS1 (151; $p = 0.49$, Student's *t*-test) or NgROS2 (146; $p = 0.44$, Student's *t*-test) (Figure 5.10).

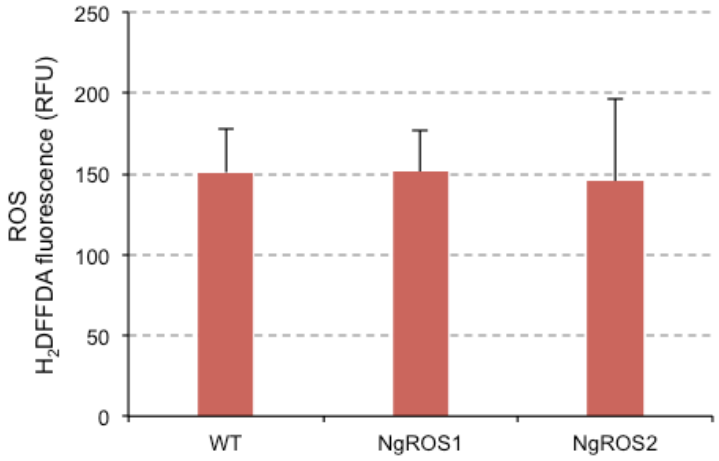


Figure 5.10: Reactive oxygen species (ROS) assay. The *N. gaditana* wild type (WT), NgROS1 and NgROS2 were grown to the logarithmic phase in 80 µmol photons m^{-2} s^{-1} light and stained with H_2DFFDA (5-(and-6)-carboxy-2',7'-dihydrofluorescein diacetate), which fluoresces when reduced to DFFDA. Values are means ± SD of triplicate measurements.

5.2.1.6 Molecular characterization of NgROS1 and NgROS2

5.2.1.6.1 3′RACE

In order to identify the mutated genes in NgROS1 and NgROS2, 3′ RACE (Rapid Amplifiction of cDNA Ends; see section 2.3.4 and Figure 4.4) was performed using the Hygro RT F2 for the first RACE PCR and the Hygro RT F3 primer for the second RACE PCR. After the first round of RACE PCRs, as expected, there were no highly visible products in the wild type, NgROS1 and NgROS2 PCRs (Figure 5.11). After the second round of nested RAC PCRs, two products were visible for NgROS1, a larger product of approx. 2.5 kb in length (ROS1 (a)) and a smaller product of approx. 0.2 kb in length (ROS1 (b)) (Figure 5.11). Two products were also visible for NgROS2, a larger product of approx. 1.0 kb in length (ROS2 (a)) and a smaller product of

approx. 0.25 kb in length (ROS2 (b)) (Figure 5.11). No bands were visible in the second round of RACE PCR for the wild type (Figure 5.11).

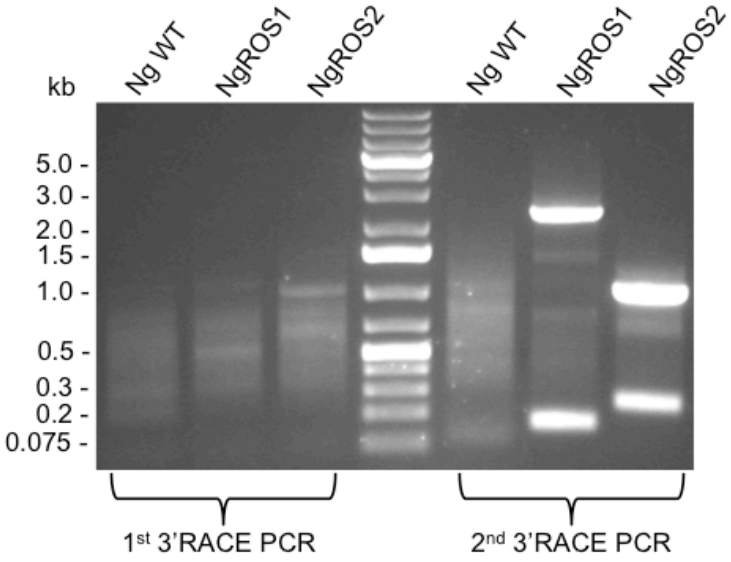
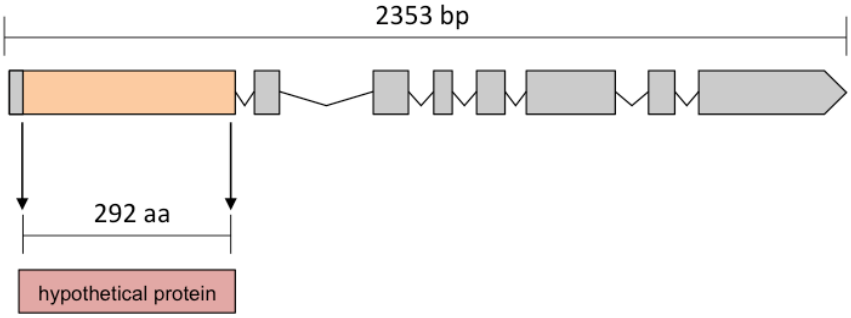


Figure 5.11: Results of 3′ RACE performed on NgROS1 and NgROS2. The products from the first round of RACE PCRs with the Hygro RT F2 and second round of PCRs with the Hygro RT F3 are shown on the left and right sides of the DNA ladder in the fourth lane, respectively. Ng WT, *N. gaditana* wild type.

5.2.1.6.2 Sequencing results for 3′RACE

The products from the second round of RACE PCRs (Figure 5.11) were sent for sequencing with the Hygro RT F3 primer. Sequences were obtained for all of the amplicons generated: ROS1 (a), ROS1 (b), ROS2 (a) and ROS2 (b) See Appendix B.2 for sequencing results. The sequences for the two products obtained for NgROS2 were 766 bp and 166 bp in length, respectively, both sequences aligned to the pAlg.Hyg.SD vector but neither sequence had alignments to the *N. gaditana* genome. This is believed to be due to the integration of an undigested pAlg.Hyg.SD vector sequence into the NgROS1 genome that contained a polyA tail that was preventing the sequence reaching the endogenous transcript DNA by providing a binding site for the adaptor primers upstream of the endogenous polyA tail of the gene of insertion. The sequences for the two products obtained for NgROS1 were 628 bp and 89 bp in length, respectively, both sequences aligned (ROS1 (a), 623 bp with 100% identitiy; ROS1 (b) 68 bp with 100% identity) to the same 2353 bp endogenous *N. gaditana* mRNA, consisting of 8 exons (NCBI ID: XM_005855667.1), the first exon predicted to code for a 292 amino acid hypothetical protein (NCBI ID: XP_005855729.1). The NCBI ID of the corresponding gene was NGA_0661400. Hence, this gene is the putative knock out in NgROS1 (Table 5.2).

Table 5.2: Gene, mRNA and protein associated with putative knock out in NgROS1. aa, amino acids; bp, nucleotide base pairs; boxes, exons; orange shading, protein coding region; grey shading, non-protein coding; ID, NCBI ID.

Gene; mRNA ID; CDS ID	**mRNA and protein structure**
NGA_0661400; XM_0058565.67.1; XP_005855729.1	

5.2.1.6.3 Structural and functional analysis of NgROS1 hypothetical protein

To search for homologs of the hypothetical protein (XP_005855729.1) in *N. gaditana* and all other available organisms. A BLASTP (Protein Basic Local Alignment Search Tool) search was carried out on the XP_005855729.1 protein sequence, shown below.

```
  1 MLCPKMRTRP FLLFLVGMLA LSTASAAPAF MYGIDDNNEI IQYDPVNKMT RLVQDTGLTK

 61 FQGSNAFAFD EVRNQMFWLY QGDATNPAGL YYWDQVTGTI DRIASQAQTW DNQRFPANAV

121 YYRDPSTGIS YFVWITEGGS TVNFLPITYD ASGNPTGVGA DIQRTISGPS FSPSFMRFGD

181 IAVQTSTKQL YLATSNGRFS KIDLTNAFGQ ALLPYTEIKT GNPSLQLAFD CEENILYGQR

241 YVATDTGSDN WYTINLATGV TTTIPNYSTE GARLSARDLG GSSCTDSPLT AR
```

Ten alignments were generated (see Table 5.3). Among the alignments was the query sequence (XP_005855729.1), in addition to two possible homologs (EWM23750.1 and EWM20154.1) and three additional *N. gaditana* sequences with lower scoring alignments. The only alignments with predicted protein function descriptions were for a VPS10 domain-containing receptor SorCS3 (otherwise referred to as sortilin related VPS10 domain containing receptor 3) protein native to bacteria *Paenibacillus jamilae* (KZE72571.1) and *Paenibacillus polymyxa* (WP_06479705.1). However, the % identity and alignment scores were low for these sequences.

Table 5.3: Table of ten significant BLASTP alignments to protein sequence coded by putative NgROS1 knock out gene.

Description	Score	E value	% ident.	Accession
hypothetical protein Naga_101493g1 [*Nannochloropsis gaditana*]	602	0.0	100	EWM23750.1
hypothetical protein NGA_0661400 [*Nannochloropsis gaditana* CCMP526]	599	0.0	100	XP_005855729.1
hypothetical protein Naga_101486g1 [*Nannochloropsis gaditana*]	584	0.0	97	EWM20154.1
hypothetical protein Naga_100764g3 [*Nannochloropsis gaditana*]	426	6×10^{-145}	72	EWM22905.1
hypothetical protein NGA_0695300 [*Nannochloropsis gaditana* CCMP526]	380	1×10^{-129}	70	XP_005852464.1
hypothetical protein NGA_0118200 [*Nannochloropsis gaditana* CCMP526]	117	2×10^{-29}	70	XP_005855371.1
hypothetical protein CBC61_08455 [*Alteromonadaceae bacterium* TMED101]	48.9	0.008	35	OUV32905.1
VPS10 domain-containing receptor SorCS3 [*Paenibacillus jamilae*]	42.0	1.0	30	KZE72571.1
VPS10 domain-containing receptor SorCS3 [*Paenibacillus polymyxa*]	41.2	2.2	30	WP_06479705.1
hypothetical protein CVT66_09615 [*Actinobacteria bacterium HGW-Actinobacteria-6*]	40.0	5.9	25	PKQ19567.1

The same protein sequence (XP_005855729.1) was scanned for matches against the InterPro protein signature databases, using the InterProScan tool (Jones *et al*., 2014; Finn *et al*., 2017). Two homologous superfamilies were predicted; a six-bladed beta-propeller, TolB-like from residues 29 – 168 and a WD40/YVTN repeat-like-containing domain from residues (169 – 282). Both of these superfamilies contain bladed-beta-propeller motifs; see Figure 5.12 for an example crystal structure for both superfamilies. Additionally, a search for signal peptide sequences was carried out using the SignalP 4.1 prediction server (Petersen *et al*., 2011). An N-terminal signal peptide was predicted from residues 1 – 26 (Figure 5.13) with a D-score (discrimination score used to discriminate signal peptides from non-signal peptides) of 0.908; this was above the minimum cut-off score (0.450), indicating that the sequence is likely a signal peptide. See Table 5.4 for a list of the predicted structures.

Table 5.4: List of possible structures and corresponding, amino acid residues, predicted for XP_005855729.1 protein using SignalP 4.1 and InterPro prediction servers. (Finn *et al*., 2017; Jones *et al*., 2014; Petersen *et al*., 2011).

Residues	Possible structure	Prediction tool
1 – 26	Signal peptide	SignalP 4.1 Server
29 – 168	Six-bladed beta-propeller, TolB-like	InterPro
169 – 282	WD40/YVTN repeat-like-containing domain	Interpro

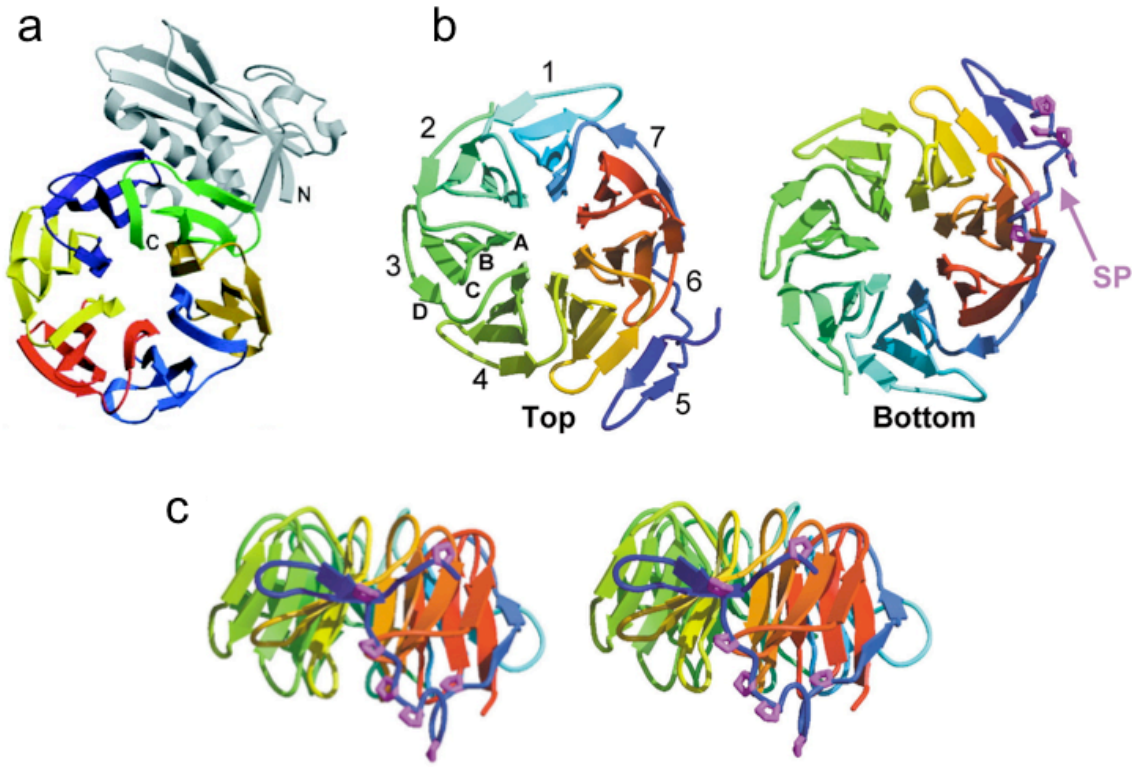


Figure 5.12: Secondary structures of homologous superfamily domains identified in NgROS1 hypothetical protein (XP_005855729.1). (**a**) A view down the central barrel of the C-terminal six-bladed β-propeller domain of the *E. coli* TolB protein. Each of the six repeated sequence motifs of the amino acid sequence that fold to give the β-propeller structure is shown in a different colour. This view demonstrates the mechanism of propeller closure in TolB; the C-terminal β strand from the final sequence motif forms the inner β strand of the four-stranded sheet that forms blade 1 (Carr *et al*., 2000). (**b**) Top ad bottom view of structure of human TLE1 C-terminal (WD40 repeat domain), showing the seven-bladed β-propeller structure. Each blade consists of a four-stranded β sheet (A–D), except for blade 5, which has an extra two strands provided by the C-terminal of the signal peptide domain (SP). (**c**) Stereo pair of the hTLE1 C-terminal viewed from the side (Pickles *et al*., 2002).

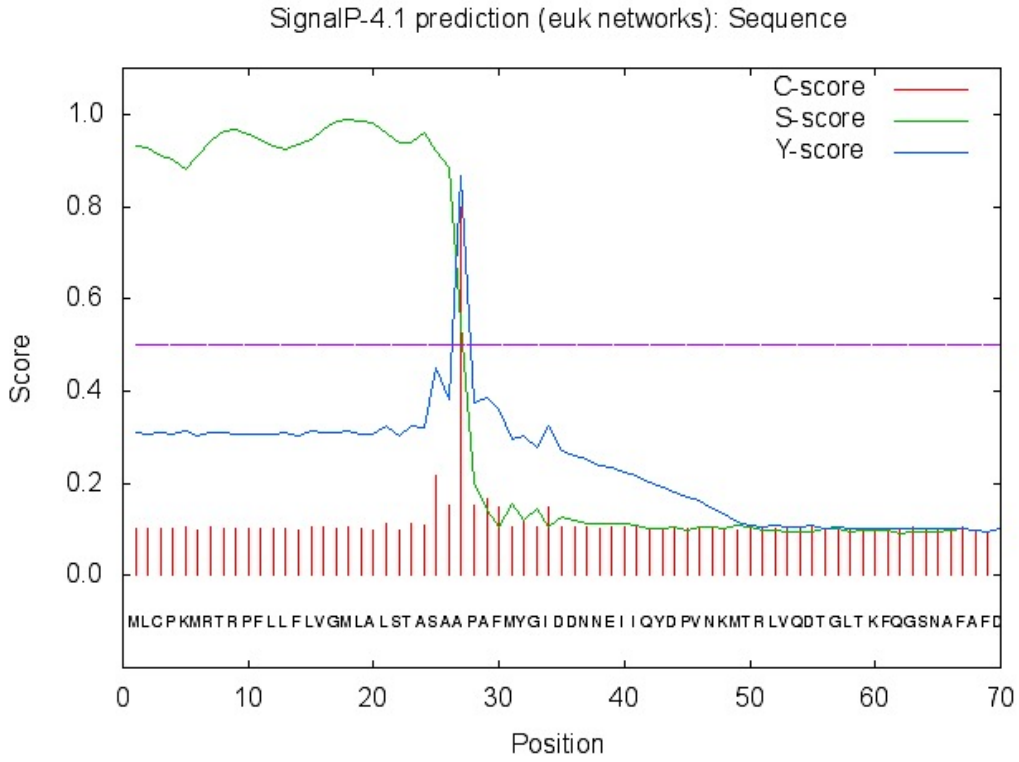


Figure 5.13: Signal peptide prediction results for NgROS1 hypothetical protein (XP_005855729.1). The SignalP 4.0 server predicted an N-terminal signal peptide from residues 1-26. The C-score (raw cleavage site score) is trained to be high at the position immediately after the cleavage site (the first residue in the mature protein). The S-score (signal peptide score) is trained to distinguish positions within signal peptides from positions in the mature part of the proteins and from proteins without signal peptides. The Y-score (combined cleavage site score) is a combination of the C-score and the slope of the S-score, giving a better cleavage site prediction than the raw C-score alone. The D-score (not shown) for the signal peptide prediction was 0.908 (above the minimum cut-off score of 0.450).

5.2.2 Screening for increased lipid content

Microalgal derived biofuel have been a source of growing interest over the past few decades (see sections 1.2.1 and 5.1). Two main biological barriers to such biofuels becoming economically feasible are insufficient rates of conversion of CO_2 into biomass (photosynthetic efficiency) and partitioning of carbon to lipids without affecting biomass productivity (Ajjawi *et al*., 2017). Although improved lipid productivities have been reported in model organisms such as *C. reinhardtii* (see section 5.1), the lipid productivities of *Nannochlopsis* strains are far higher. For example, the lipid-production capacity of *N. gaditana* is six fold higher than that of *C. reinhardtii* (Radakovits *et al*., 2012).

The remainder of this chapter describes the use of a Fourier-transform infrared (FTIR) spectroscopy-based screening procedure to identify mutant *N. gaditana* strains with increased or

altered lipid content. FTIR spectroscopy (see section 2.2.7) has been indicated as an efficient and reliable method for high-throughput lipid determination (Dean *et al*., 2010). However, this technology has not yet been used for high-throughout screening of a biofuel candidate microalga for mutant strain with altered lipid content. Using FTIR spectroscopy, lipid content can be expressed qualitatively as a ratio between the height of characteristic absorption peaks associated with lipid molecules and protein molecules, eliminating the need to normalise samples based on biomass which would compromise the efficiency of the screen (see section 2.2.7) (Dean *et al*., 2010; Stehfest *et al*., 2005). A lipid:amide I ratio, between the height of the lipid peak at ~ 2920 cm^{-1} and the amide I peak (ubiquitous to protein molecules) at ~ 1655 cm^{-1} was used to as the basis for the screen (see section 2.2.7).

The pAlga.Hyg.SD gene-trap vector was used to generate the library of *N. gaditana* mutants used for the screen, enabling the possibility of identifying the genes associated with the phenotypes of mutant strains exhibiting increased lipid content. Hence, in addition to the generation of improved phenotypes, key negative regulators of lipid production may also be identified.

5.2.2.1 Preliminary experiments for lipid analysis of *N. gaditana* using FTIR

To validate the lipid:amide I analysis, a nitrogen deprivation experiment was carried out in order to elicit an easily detectable and clear upregulation of lipid production in *N. gaditana*. Triplicate cultures (40 mL) of *N. gaditana* were grown to the logarithmic phase, washed in nitrogen deplete F2N media three times. Each replicate was then split into either nitrogen replete (5 mM ammonium chloride) or nitrogen deplete (no nitrogen) media (40 mL) and grown for a further 48 hours before analysis on the FTIR (see section 2.2.7).

Two different peaks were tested for the lipid component of the lipid:amide I analysis: the peak at 2921 cm^{-1} (stretching of lipid C-H bonds) and 1741 cm^{-1} (stretching of C=O bonds, primarily from lipids and fatty acids). Figure 5.14a shows the spectra of each *N. gaditana* replicate in both conditions normalized to the amide II peak (1545 cm^{-1}); a large increase in the absorbance of the 1741 cm^{-1} lipid peak can clearly be seen in the nitrogen deplete cultures. Figure 5.14a shows the lipid:amide I ratio both nitrogen conditions using the 1741 cm^{-1} peak; Figure 5.14c shows the lipid:amide I ratio in both nitrogen conditions using the 2921 cm^{-1} peak. The former analysis showed a 201 % increase in the lipid:amide I ratio from 0.323 ± 0.08 to 0.971 ± 0.20 in the nitrogen deplete cultures (p = 0.003, Student's t-test; Figure 5.14b). The latter showed a similar increase of 215 % in the lipid:amide I ratio from 0.401 ± 0.08 to 1.264 ± 0.16 in the nitrogen deplete cultures (p = 0.0004, Student's t-test; Figure 5.14c). The 2921 cm^{-1} peak was used in all subsequent calculations of lipid:amide I ratio.

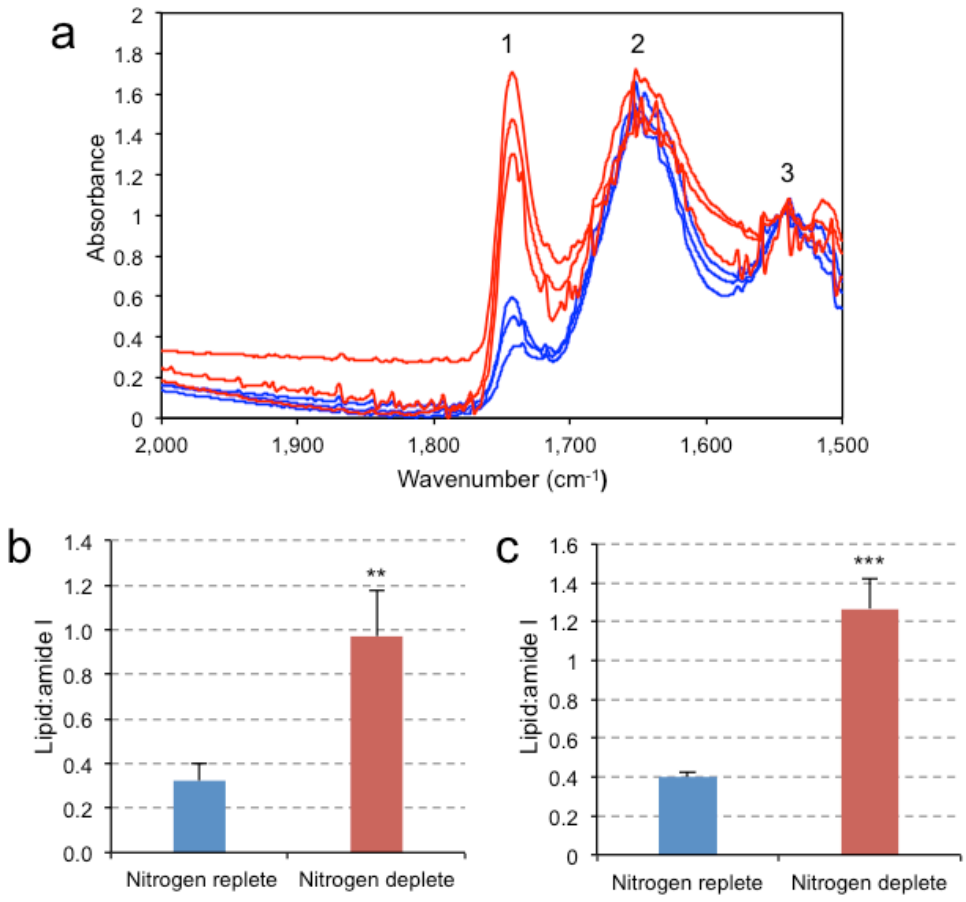


Figure 5.14: Lipid:amide I analysis of *N. gaditana* in nitrogen replete and deplete conditions. (**a**) 2,000 to 1,000 cm^{-1} region of the FTIR spectra showing normalization to the amide II protein peak (labelled 3), the amide I protein (labelled 2) and an increase in the 1741 cm^{-1} lipid peak (labelled 1). Blue spectra, replicates in nitrogen replete conditions (5 mM ammonium chloride); red spectra, replicates in nitrogen deplete conditions. (**b**) Lipid:amide I analysis using the 1741 cm^{-1} peak shown in panel **a**. (**c**) Lipid:amide I analysis using the 2921 cm^{-1} peak. Blue bars, nitrogen replete conditions; red bars, nitrogen deplete conditions; **, p <0.01; ***, p <0.001.

5.2.2.2 Screening procedure

N. gaditana was transformed with the pAlg.hyg,SD vector following the standard electroporation procedure (2.4.1). After transformants had appeared and were ready to be further processed (~ 6 weeks), arrays of 94 mutants strains were inoculated on bioassay dishes using the same media as used for the initial transformation (both containing a reduced nitrogen concentration of 2 mM ammonium nitrate). Colonies were picked and re-streaked into each chamber of the grid. These plates were then incubated under the same conditions as the initial transformation plates for a further 12 days before being analysed on the FTIR. Figure 5.15 shows eight gridded array plates making up the mutant library. In order to minimize natural lipid variance, all gridded array plates were measured on the FTIR exactly 12 says after

inoculation and were positionally rotated daily to account for slight changes in irradiance or temperature within the incubator. See section 2.5.2 for a full description of methods. Each strain was named NgLX.Y, where X is the number of the bioassay array plate the mutant strain was selected from and Y is the number of the mutant within the bioassay array plate.

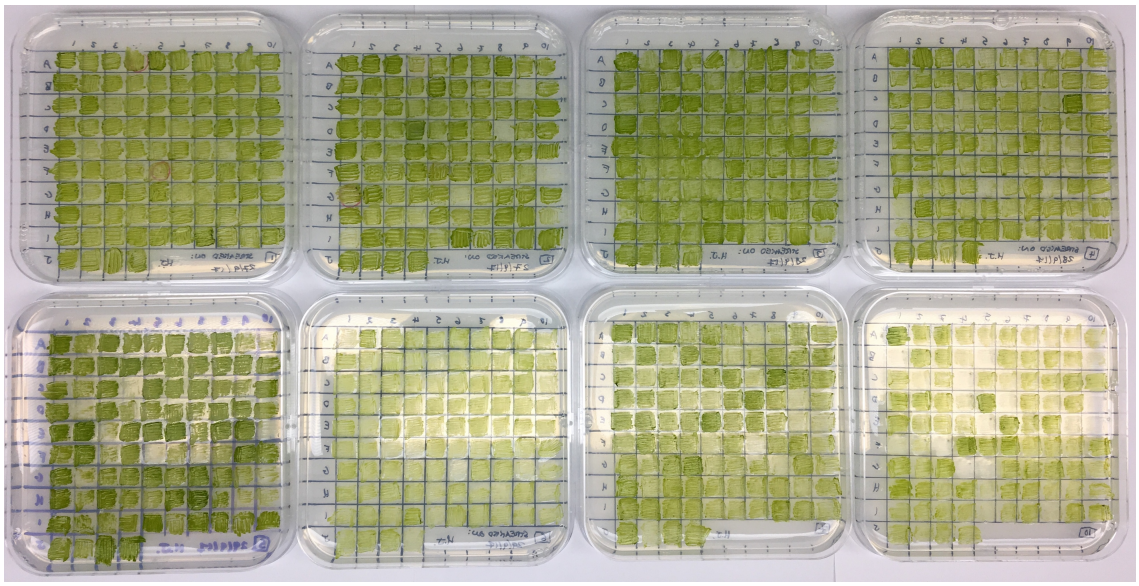


Figure 5.15: Formant of mutant library for FTIR screen for increased lipid:amide I ratio. Each gridded bioassay dish contains 94 mutant strains. Note: image was taken several weeks after analysis.

5.2.2.3 Lipid content analysis of mutant library

After 12 days, the cell lines from each plate were loaded into the FTIR for analysis. The lipid:amide I value for each cell line in the mutant library was then compiled into one data set for analysis (Figure 5.16). All spectra were visually inspected before analysis. Wells containing insufficient sample biomass yielded spectra that were not fit for analysis; these spectra were removed from the lipid:amide I analysis. The average lipid:amide I ratio across all of the mutant cell lines was 0.587 ± 0.159. This average was higher than the lipid:amide I ratio measured in the *N. gaditana* wild type in standard F2N media (0.401). This was presumably due to additional stress factors in the mutant library strains due to the use of solid agar media with a reduced salt concentration (50% artificial see water) and nitrogen content (2 mM NH_4Cl). The three strains with the largest lipid:amide I ratio were NgL1.4, NgL1.55 and NgL2.61, the respective lipid:amide I ratios were 1.04, 1.21 and 1.06, corresponding to ratios >2xSD greater than the mean. These strains were recovered in liquid culture for further analysis.

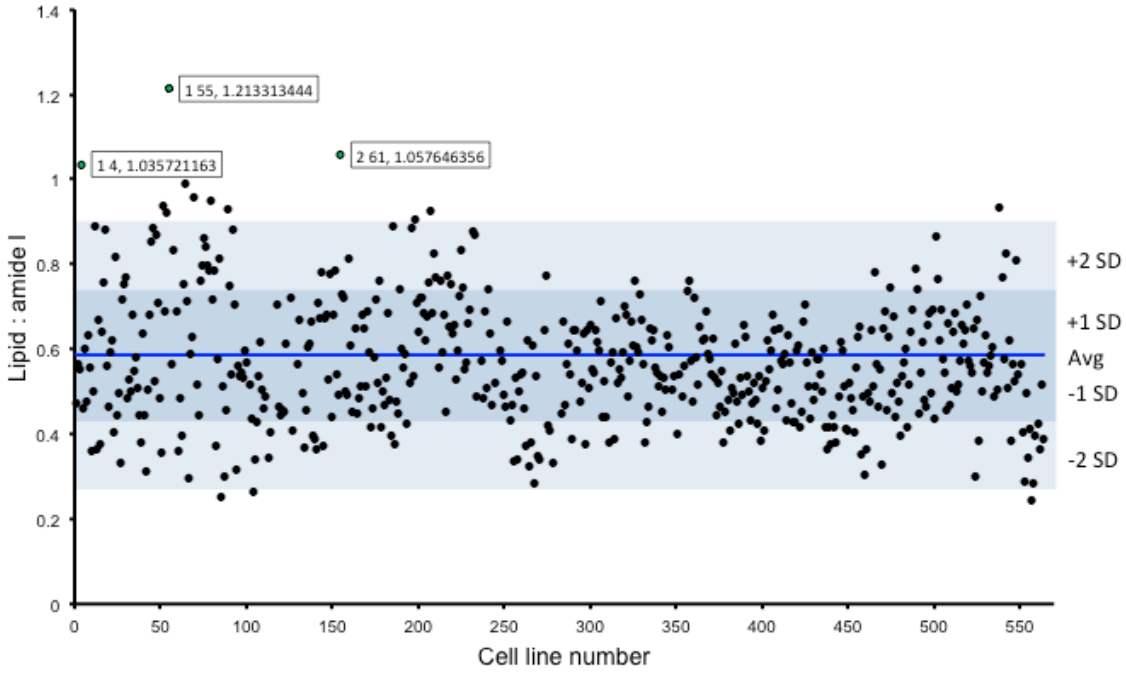


Figure 5.16: Lipid:amide I ratio values for mutant library measured by FTIR spectroscopy. Each data point represents a different cell line. The average lipid:amide I value across all of the cell lines is denoted by the blue line; the blue shading denotes and one (blue shading) and two (light blue shading) standard deviations from the mean. The three strains with the highest lipid:amide I values, NgL1.4, NgL1.55 and NgL2.61, are shown in green.

5.2.2.4 Genotyping of NgL1.4, NgL1.55 and NgL2.61

NgL1.4, NgL1.55 and NgL2.61 were genotyped by PCR with the *hph* forward reverse genotyping primers to confirm integration of the pAlg.Hyg.SD cassette (Figure 5.17). The 435 bp amplicon was generated in NgL1.4, NgL1.55 and NgL2.61 and was absent in the wild type.

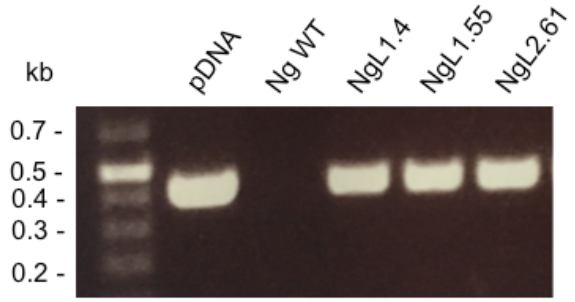


Figure 5.17: Genotyping of NgL1.4, NgL1.55 and NgL2.61 with *hph* genotyping primer set to confirm integration of the pAlg.Hyg.SD cassette. The 435 bp amplicon was present in NgL1.4, NgL1.55, NgL2.61 and the positive control plasmid DNA (pAlg.Hyg.SD) reaction but was absent in the *N. gaditana* wild type. Template DNA: lane 1, DNA ladder; lane 2, pDNA; lane 3, wild type; lane 4, NgL1.4; lane 5, NgL1.55; lane 6, NgL2.61.

5.2.2.5 Lipid analysis of NgL1.4, NgL1.55 and NgL2.61 in liquid culture

To assess the lipid:amide I ratio of the NgL1.4, NgL1.55 and NgL2.61 strains, triplicate 40 mL liquid cultures were inoculated in standard F2N media and sampled at the logarithmic phase for analysis on the FTIR (Figure 5.18). As expected from previous data (Figures 5.14 and 5.16), the lipid:amide I values were lower as compared to in the screening format. In these conditions, the wild type had the lowest lipid:amide I value (0.345 ± 0.03); the lipid:amide I value of NgL1.4 was (0.451 ± 0.04), which was significantly higher than the wild type ($p = 0.0012$, Student's *t*-test); the lipid:amide I value of NgL1.55 was (0.382 ± 0.03) and was not significantly different to wild type at $p < 0.05$ (Student's *t*-test); the lipid:amide I value of NgL2.61 was (0.414 ± 0.02), which was significantly higher than the wild type ($p = 0.015$, Student's *t*-test).

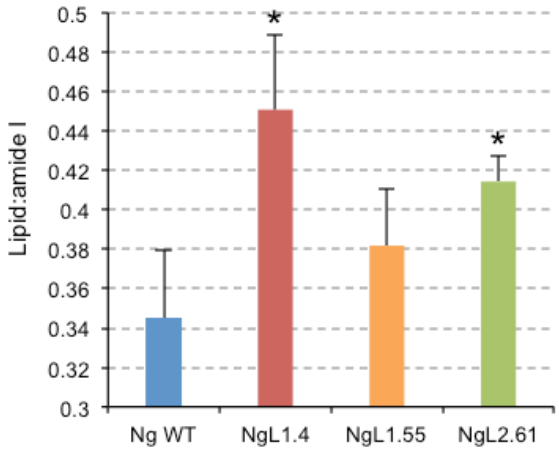


Figure 5.18: Lipid:amide I values for the *N. gaditana* wild type (blue bar), NgL1.4 (red bar), NgL1.55 (yellow bar) and NgL2.61 (green bar) in liquid culture. All cultured were grown to the logarithmic phase in standard F2N media in triplicate 40 mL cultures. *; significantly different from wild type at $p < 0.05$, Student's *t*-test.

5.3 Discussion

The prospect of algal derived biofuels is attracting great interest and is a rapidly growing area of research; however, improvement of naturally high oil producing strains using genetic engineering is required to make such production systems economically viable (see section 5.1) (Radakovits *et al*., 2010; Barbosa and Wijffels, 2013; Fon Sing *et al*., 2013; Doron *et al*., 2016). An increase in total lipid yield could be achieved by engineering strains with increased growth rates and overall biomass accumulation or increased cellular lipid content. This chapter described the use of random insertional mutagenesis and phenotypic screen to isolate mutant strains with potentially relevant phenotypes for biotechnology, specifically biofuel production. The goal of the first screen was to isolate strains with increased resistance to oxidative damage, induced by high light, which could potentially result in an overall increase in biomass accumulation (Erickson *et al*., 2015). The second screen was designed to isolate mutant strains with increased lipid content.

5.3.1 Screening for increased resistance to reactive oxygen species

The use of an increasing series of concentrations of menadione in the media as a screening tool provided an effective means of isolating strains with increased resistance because only the strains that were able to grow above the level of resistance of the wild type formed colonies on the transformation plates, while strains not able to grow at this level were killed. If a membrane transporter were required for the entrance of menadione into the cell, its use as a selective agent

for ROS resistance could conceivably be impeded by selection of mutations resulting in disruption of its transporter and reduced uptake into the cell, as apposed to ROS resistance. However, menadione is hydrophobic and passes through the cell membrane via diffusion (Burg *et al*., 2002).

Two strains, NgROS1 and NgROS2, were isolated which exhibited increased resistance to menadione, in addition to hydrogen peroxide (H_2O_2), which was not used as a selection parameter in the screen (Figure 5.6). This could suggest that the mechanism of resistance is not through the dissipation of a specific reactive oxygen species.

The initial light transition growth experiment showed an increase in biomass accumulation by the end of the growth experiment in NgROS1 and NgROS2 in both the control light group and in the high light group (Figure 5.7a). The high light transition had less of a negative effect on photosynthetic efficiency the following day (five) for both NgROS1 and NgROS2; however, the photosynthetic efficiency data became hard to interpret over the remainder of the experiment (Figure 5.7b). The high light transition elicited a decrease in the absorption cross-section of photosystem II in all strains of the high light group as compared to the control group; however, there was little difference between the cell lines within each group (Figure 5.7c) and again, the data was hard to interpret.

The second light experiment (Figure 5.8) was carried out over a shorter time interval with a larger increase in irradiance (80 to 2000 μmol photons m^{-2} s^{-1}). After the one hour high light treatment, both NgROS1 and NgROS2 had photosynthetic efficiencies significantly higher than that of the wild type (Figure 5.8a). The high light treatment caused a 1.23 % reduction in the absorption cross-section of photosystem II in the wild type (Figure 5.8b), which is in line with prevalent model that σPSII, is reduced in high light and restored in the dark (Lambrev *et al*., 2012). Interestingly, the absorption cross-section of photosystem II of NgROS1 and NgROS2 showed 5.67 % and 7.34 % increases in the high light treatment, respectively. Figure 5.7c showed a continual reduction in σPSII over the days after the high light transition in NgROS1 and NgROS2, which would suggest that the upregulation in σPSII seen after an hour of high light (Figure 5.8b) was a short lived effect. PSII complexes are arranged in clusters of one or more with an energetically shared network of LHCIIs (light harvesting complex II) (Hankamer *et al*., 2001; Dubinsky, 2013). It is conceivable that partial photoinhibition of PSII complexes within such networks could cause the light energy absorbed from the LHCII complexes associated to a neighbouring photoinhibited PSII complex to be redirected to another PSII reaction centre within the network, thus increasing σPSII (see Figure 5.19a). Presumably, this would be happening in tandem with structural rearrangement of the PSII-LHCII macro-organization, driving σPSII down (Figure 5.19b), which would eventually predominate (Dubinsky, 2013; Lambrev *et al*., 2012).

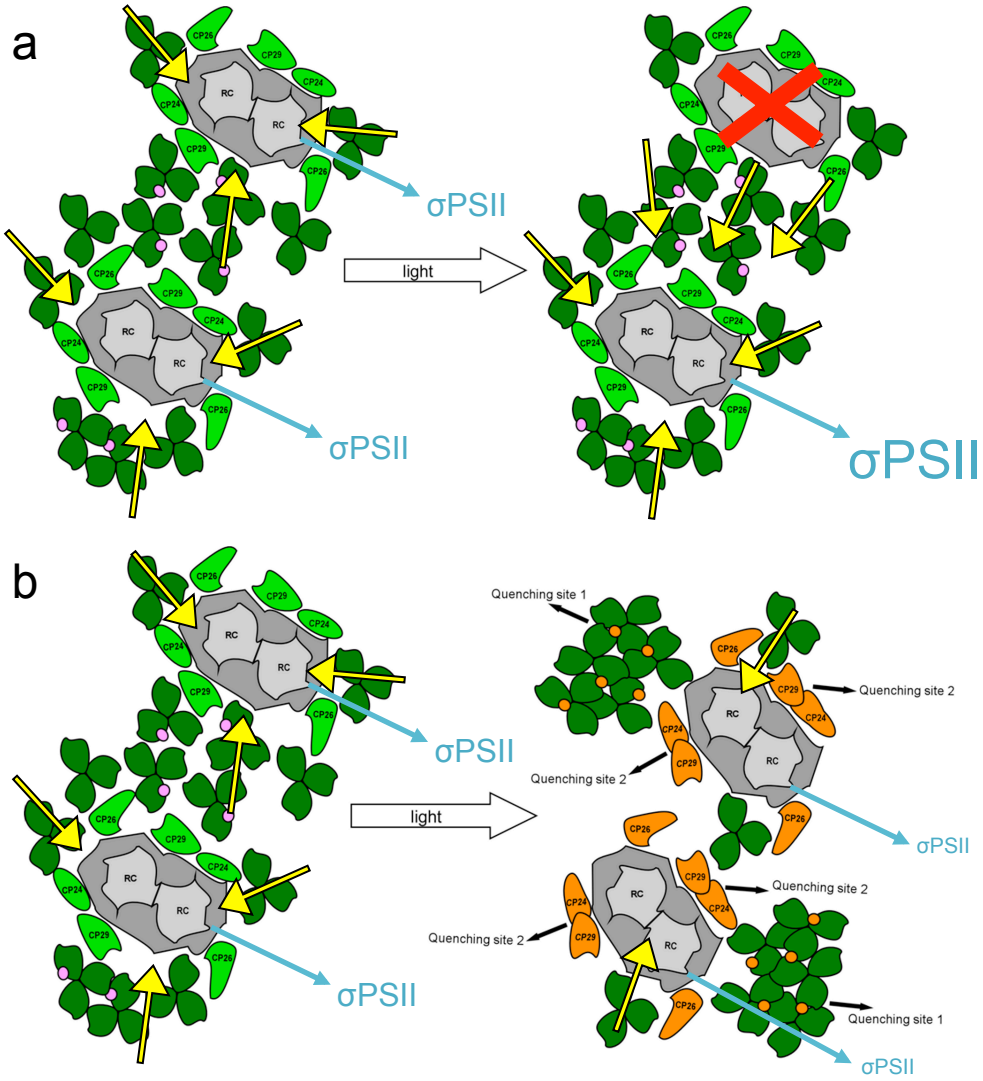


Figure 5.19: (**a**) Possible model for photoinhibition leading to an increase in measurements of the functional antenna cross-section of photosystem II (σPSII). Exposure to light leads to photoinhibition (red cross) of a PSII complex and redirection of light energy absorbed from its associated light harvesting complexes to a neighbouring PSII complex, in turn, increasing the σPSII of the remaining PSII complex. (**b**) Dissociation of LHCII complexes from PSII under high light leading to a reduction in σPSII. The pink circles in the medium light conditions state (left panel) denote violaxanthin, which is converted to zeaxanthin (orange circles) in the high light adapted state (right panel) causing aggregation and disassociation of LHCII trimers from the PSII-LHCII supercomplex. The conversion of the minor antenna from light green in the left panel to orange in the right panel also denotes conversion of violaxanthin to zeaxanthin. PSII complexes (dark grey structures) contain two reaction centres (RC), the associated LHCII trimmers are represented in dark green. Figure adapted from Dubinsky (2013).

The sequential light challenge experiment (Figure 5.9) further demonstrated the high light tolerance of NgROS1 and NgROS2. After the twelve rapid transitions from dark to 2000 µmol photons m^{-2} s^{-1}, the *N. gaditana* wild type restored photosynthetic efficiency to 85.9 % of its previous level. For NgROS1 and NgROSII the percentage restoration of photosynthetic efficiencies after the rapid light curve were significantly higher, 88.2 % and 90.6 %

respectively. The D1 reaction centre protein of PSII (see Figure 1.1) is a major target site for photodamage and is normally maintained by the PSII repair cycle (Kato *et al*., 2015). In theory, the improved recovery of NgROS1 and NgROS2 seen in the rapid light curve could thus either be due to a lower level of photodamage accumulation or an increase in the rates of repair of D1.

NgROS2 had the highest level of resistance to menadione and hydrogen peroxide and also the highest tolerance to high light (Figure 5.6, 5.8 and 5.9), while NgROS1 was at an intermediate level respective to the wild type, which could suggest a correlation between resistance to ROS and increased resistance to high light in the data shown. The ROS stain assay (Figure 5.10) showed that in standard light conditions (80 μmol photons m^{-2} s^{-1}) there was no significant difference in ROS production in NgROS1 or NgROS2 respective to the wild type.

It was not possible to identify the insertion site of the vector in NgROS2 due to the presence a polyA tail from the vector DNA that was interfering with the 3′ RACE reaction (see section 5.2.1.6.1). To overcome this issue, alternative methods must be used that do not rely on primer binding to an adaptor sequence added to the polyA of the endogenous transcript, as is the case with 3′RACE; this could be wither 5′ RACE or an arbitrary primed PCR procedure (Scotto-Lavino *et al*., 2007; Das *et al*., 2005; Tran *et al*., 2015). However, the vector insertion site of the NgROS1 strain was obtained. Analysis of the predicted hypothetical protein (see section 5.2.1.6.3) revealed a high likelihood that the protein contains a signal peptide at the N-terminal end, from residues 1-26, and two domains from 29 – 168 and 169 – 282 that contain beta-propeller motifs (see Table 5.4, Figure 5.12 and 5.13). The presence of a signal peptide dictate that the hypothetical protein is targeted to the secretory pathway, residing in either the endoplasmic reticulum, golgi, endosomes, cellular membrane or the protein may be secreted from the cell (Hemminger *et al*., 1998). Given that the beta-propeller motif domains are immediately adjacent to each other, it is conceivable that they could form one continues beta-propeller domain. Residues 169 – 282 were predicted as to constitute a WD40/YVTN repeat-like-containing domain. WD40 domain-containing proteins are a large family of proteins found in all eukaryotes and are involved in a range of functions, including: signal transduction, transcriptional regulation, cell cycle control, autophagy and apoptosis (Stirnimann *et al*., 2010). Despite the fact that WD40 domains have been implicated in many critical roles, they have been studied to a lesser extent than other common protein domains, making interpretation of its possible role in the NgROS1 hypothetical protein difficult (Stirnimann *et al*., 2010). Additionally, WD40 domains are highly promiscuous in their interactions due to their ability to interact with a wide range of proteins, peptides and nucleic acids, which further adds to the difficulty in interoperating the role of the possible WD40 domain in the NgROS1 hypothetical protein (Stirnimann *et al*., 2010). YVTN domains are similar in structure to WD40 domains and

may play a similar role as a rigid scaffold for a range of protein interactions (Jing *et al*., 2002; Datta *et al*., 2003; Chen *et al*., 1998).

To summarise, the menadione screen for increased ROS resistance yielded two strains, NgROS1 and NgROS2, with increased resistance to both menadione and hydrogen peroxide. Additionally, both strains exhibited elevated tolerance to the extreme light conditions tested. However, further work is required to resolve the structural and functional role of the genes disrupted by the integration of the transformation vector in both NgROS1 and NgROS2 and the wider implications this may have to the understanding of ROS dissipation mechanisms and the generation of mutant strains of other organism with increased resistance to ROS.

5.3.2 Screening for increased lipid content

FTIR spectroscopy can be used to rapidly determine the macromolecular composition of algal cells (Dean *et al*., 2010; Stehfest *et al*., 2005; Wagner *et al*., 2010). Lipid content can be assessed qualitatively with relative ease and efficiency by expressing the peak height ratio of lipid to amide I (Dean *et al*., 2010). However, this technology has not yet been employed as a tool for high throughput screening. Section 5.2.2 describes the establishment of a high throughput 96-well format screening platform and the use of this system to assess the relative lipid content of over 500 *N. gaditana* mutant cell lines.

The preliminary experiments assess changes in lipid accumulation in response to nitrogen deprivation (Figure 5.14) demonstrated that the FTIR is capable of measuring lipid content effectively. A 215 % increase in lipid content was observed in the nitrogen deprivation treatment tested (Figure 5.14c). This increase in total lipid is similar to those reported previously in *N. gaditana* (Simionato *et al*., 2013). Simionato *et al* reported total lipid increases of approx. 270 % using quantitative extraction-based methods; however, the nitrogen deprivation period before analysis in their study was significantly longer (7 days) (Simionato *et al*., 2013). The use of the amide I peak as an internal measure of biomass eliminated the need for normalization on the basis of cell count, which would have severely reduced the efficiency of the screen. The macro writing software provided by the FTIR supplier (see methods section 2.2.7) allowed for full automation of the lipid:amide I analysis. The major bottlenecks in the efficiency of the screen were: (a) the process of picking colonies from the initial transformation plates and arraying them in a gridded format and (b) transfer of cell lines from the gridded format to the FTIR microplate for analysis, which was all done manually (see methods section 2.2.7). However, colony picking robots and 96-well format liquid handling robots are currently revolutionising the speed and automation of high throughput microbial screening (Hartley *et al*., 2009; Liu *et al*., 2017a; Tillich *et al*., 2014). Such technologies could be applied to the screening

procedure presented here with no further development, which would greatly improve the potential of FTIR spectroscopy as a platform for phenotypic screening. Another highly advantageous trait of FTIR is that it can be used to assess a wide range of macromolecules, which may include those of high commercial value. For example, the carotenoid lutein, a high-value pharmaceutical and nutraceutical, expressed in several algae, including *Nannochloropsis* sp., can be measured using FTIR spectroscopy (Liu *et al*., 2012; Borowitzka, 2013; Markou and Nerantzis, 2013).

The lipid:amide I analysis performed on the mutant library revealed a large degree of variability among the strains (Figure 5.16). Although the respective mutations in each strain are likely to have contributed to this, it is also likely that a large proportion of this variability was due to subtle differences in the way colonies were streaked and subsequently grew in addition to environmental variation due to slight differences in temperature or irradiance within the incubator. This variability represents another aspect of the screening platform that would be improved through the use of liquid handling robots capable of higher levels of consistency. Cytometry-based lipid screens using methods such as fluorescence activated cell sorting (FACS) are less vulnerable to these effects of cell line history as they can be performed on a pooled batch culture of mutated strains that are all subject to the same growth regime (and stain procedures if used) before being analysed on a cell-by-cell basis (Terashima *et al*., 2015).

The lipid:amide I ratios of all strains were significantly lower in the nitrogen replete F2N media used for the liquid culture tests (Figure 5.16) as compared to the values observed in the screen due to the reduction in stress to the cell; presumably most significantly, due to nitrogen limitation in the screening format and the use of solid media. However, liquid culture analysis of the three strains exhibiting the highest lipid:amide I values, NgL1.4, NgL1.56 and NgL2.61, showed that both NgL1.4 and NgL2.61 had significantly higher relative lipid contents ($p < 0.05$, Student's *t*-test) compared to the wild type (Figure 5.18). The central dilemma of generating algal strains with enhanced lipid production is that this often comes with an associated reduction in growth rates (Tan and Lee, 2016). A recent study reporting an algal strain with a 2-fold increase in lipid productivity and close to wild type growth rates has demonstrated that it may be possible to overcome this dilemma by attenuating the expression of negative lipid regulators, as opposed to knocking them out completely (Ajjawi *et al*., 2017). Thus, it should be noted that analysis of the growth characteristics of these strains would be required before any statements on overall lipid productivity could be made, which were not possible in the timeframe of the PhD. Additionally, validation of the respective lipid contents of each strain should be validated with other methods of lipid quantification.

In conclusion, FTIR spectroscopy shows promise as a tool for microbial screening for strains with altered macromolecular compositions. A workflow has been designed and presented here

which allowed the rapid determination of the relative lipid contents of over 500 mutant strains in relatively short order and the subsequent isolation of two strains with significantly higher lipid contents. Natural variability in lipid accumulation poses a challenge for screens such as this that are performed over several temporally distinct instalments, each with slight differences in incubation conditions and growth of mutant strains. Further characterization of NgL1.4 and NgL2.61 is required, in order to determine if improved lipid productivity has been achieved, and if so, to characterize the genetic molecular basis for these respective phenotypes.

Chapter 6: *D. tertiolecta* transformation procedure trials

6.1 Introduction

113 years have past since the first description of the *Dunaliella* genus. *Dunaliella* are single-celled halotolerant green alga, responsible for most of the primary production in hypersaline environments (Oren, 2005). The large accumulation of *β*-carotene in certain strains has already lead to commercial application; *D. salina*, is grown in mass culture for *β*-carotene production in several countries including, Australia, USA and China (Hosseini Tafreshi and Shariati, 2009). *Dunaliella* has also become a model for study of salt adaptation in algae due to its extremely halotolerant phenotype (Oren, 2005).

D. tertiolecta endogenously stores several high-value products such as, carotenoids, *β*-carotene and mycosporine-like amino acids (Vinayak *et al*., 2015). In addition, *D. tertiolecta* has been utilized for bioremediation due to its ability to grow rapidly in a wide variety of environments (Hosseini Tafreshi and Shariati, 2009). The oil content in *D. tertiolecta* ranges from 36-42% of dry weight, and can increase further under nitrogen deprivation, which has lead to the consideration of this species for biofuel production (Barbosa and Wijffels, 2013; Minowa *et al*., 1995). Furthermore, the fast growth rates of *D. tertiolecta* in hyper saline environments aid in the maintenance of pure cultures and the lack of a rigid cell wall (Figure 2.2a and 2.2b) enables easier product extraction (Goyal, 2007; Goyal and Gimmler, 1989; Barzegari *et al*., 2010; Hosseini Tafreshi and Shariati, 2009).

D. tertiolecta is regarded as a "non-model" organism (Georgianna *et al*., 2013). Knowledge of functional genomics and molecular genetics are essential for metabolic engineering efforts towards enhancing the quality and quantity of biofuel feedstock derived form algae; however, neither the nuclear or organellar genomes of *D. tertiolecta* have yet to be sequenced, limiting description of key pathways to transcriptome sequencing (Rismani-Yazdi *et al*., 2011). Another prerequisite technology for full exploitation of a species for genetic engineering is the ability to

transform the nuclear genome with foreign DNA in order to direct multigene traits (Scott *et al*., 2010; Rasala *et al*., 2014). The nuclear genomes of several *Dunaliella* species have been transformed stably via electroporation, glass beads, *Agrobacterium tumifican* and micro-particle bombardment (Sun *et al*., 2005; Feng *et al*., 2009; Anila *et al*., 2011; Shabelnik *et al*., 2011). Nuclear transformation of *D. tertiolecta* has not yet been established. An electroporation procedure has been explored in one study, however, analysis using Southern hybridization revealed that plasmid DNA was entering the cells but was being rapidly degraded, resulting in transient expression of the transgene (Walker *et al*., 2005b). Recombinant nuclear genes are often vulnerable to "silencing" which poses a significant challenge to stable transformation (Cerutti, 1997). The plastid of eukaryotic algae is of prokaryotic origin and void of the silencing machinery; hence, stable integration of transgenes for extended periods is made possible in this organelle (Newman *et al*., 1990; Blowers, 1989). Indeed, the chloroplast of *D. tertiolecta* has been successfully transformed and has showed promise as a versatile host for production of measureable quantities of recombinant enzymes of five different classes: xylanase, α-galactosidase, phytase, phosphate anhydrolase, and β-mannanase (Georgianna *et al*., 2013). Table 6.1 lists the main biotechnologically relevant species of *Dunaliella* and the current level of development of each species.

Table 6.1: Main biotechnologically relevant species of *Dunaliella* and available tools.

Species	**Available tools**	**References**
D. salina	Nuclear transformation, chloroplast transformation, organellar genome sequences, transciptomic information, most developed molecular tools of the genus	(Sun *et al*., 2005; Feng *et al*., 2009; Li *et al*., 2008, 2011; Tan *et al*., 2005a; Smith *et al*., 2010; Zhao *et al*., 2011)
D. bardawil	Nuclear transformation, little development of molecular tools	(Anila *et al*., 2011)
D. tertiolecta	Chloroplast transformation, recombinant enzyme expression in chloroplast, little development of molecular tools	(Georgianna *et al*., 2013; Walker *et al*., 2005b, 2005a)

This chapter describes efforts to further develop *D. tertiolecta* as a candidate strain for biotechnology by establishing a robust nuclear transformation protocol. An *Agrobacterium*-mediated protocol was initially chosen due to it's wide application in marine microalgae (Cheney *et al*., 2008; Pratheesh *et al*., 2014; Anila *et al*., 2011). In addition to the *Agrobacterium*-mediated transformation trials, several other transformation methods established in microalgal species were tested with adaptions for *D. tertiolecta*; namely, biolistics, agitation in the presence of glass beads, and an electroporation procedure developed in *D. tertiolecta* in a

previous study yielding inconclusive results (Walker *et al*., 2005b). The *Agrobacterium* and biolistics transformation protocols were provided by industry collaborators (Algenuity) and adapted for *D. tertiolecta*. See section 2.4 for full descriptions of the methods for each transformation procedure.

6.2 Results

6.2.1 *Agrobacterium*-mediated procedure

Agrobacterium-mediated transformation was chosen for exploration in *D. tertilecta* due to its advantageous characteristics and applicability in eukaryotic algae (see section 1.3.1.1.4). The *Agrobacterium* strains used in these trials harboured a pCAMBIA.ble2AmCher vector, which uses 'gene-trap' technology to drive transcription of the vector from an endogenous promoter, in turn disrupting the coding sequence of the endogenous gene. The T-DNA contained, from 5′ to 3′: a splice acceptor site, a Flp-*FRT* recombination site, a *ble* gene encoding the bleomycin resistance protein (Ble) conferring resistance to zeocin, a 2A self-cleavage site, an *mCherry* gene encoding an mCherry fluorescent reporter protein and a splice donor site. The splice acceptor and splice donor sequences are consensus sequences for eukaryotic microalgae, which direct integration of the vector into a gene within the host organism genome. Transcription of the 'trapped' gene results in the production of an RNA fusion transcript containing the vector, flanked by the endogenous coding sequences. Translation of the spliced mRNA generates a truncated endogenous protein, which would likely be rapidly degraded, the Ble protein and the mCherry protein. The Flp-*FRT* and 2A sequences ensure that the Ble protein and mCherry reporter are not fussed together or to the endogenous truncated protein (see Figure 6.1).

The use of this vector has the advantage of using endogenous promoters, which are less vulnerable to silencing, in a nonbiased manner, to drive expression of Ble and mCherry. Interruption of the endogenous coding sequence generates knockout strains, which can be phenotypically screened for improved phenotypes. Additionally, the level of the Ble and mCherry proteins produced is proportional to the strength of the endogenous promoter, which enables selection of promoters of increasing strength by plating the transformants on and increasing scale of the zeocin selective agent and validation of these expression levels via spectrofluorometeric analysis of the mCherry reporter protein.

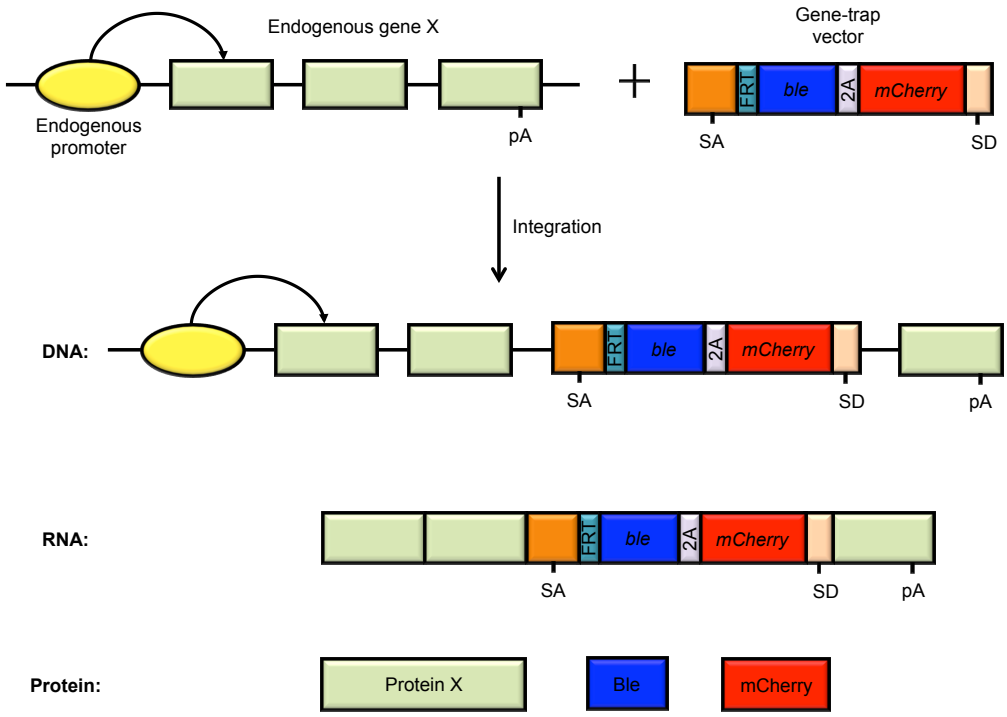


Figure 6.1: Integration of the 'gene-trap' vector into the host genome and subsequent production of separate resistance protein and mCherry fluorescent reporter protein. The vector sequence contains: a splice acceptor site (SA) and splice donor site (SD), a *ble* gene encoding the bleomycin resistance protein (Ble), conferring resistance to zeocin, an *mCherry* gene encoding an mCherry fluorescent reporter protein, and Flp-*FRT* and 2A sequences that ensure the production of separate Ble and mCherry proteins. The SA and SD sequences direct integration into endogenous gene X generating a fusion transcript, proportional in expression to the strength of the endogenous promoter. Translation of this fusion transcript generates three translationally discrete products: a truncated endogenous protein X, the Ble protein and the mCherry protein.

The methods used for the *Agrobacterium* transformation procedure are detailed in section 2.4.2; see Figure 6.2 for an overview of the procedure.

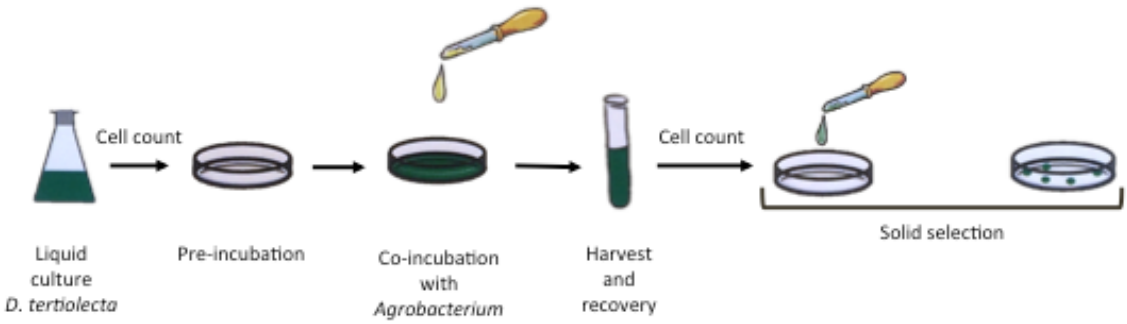


Figure 6.2: Simplified workflow for *Agrobacterium* transformation procedure. Firstly, *D. tertiolecta* is grown in liquid culture before being harvested by centrifugation. A cell count is performed on the harvested cells to determine the required volume to be plated on agar. After the agar plates have been incubated for 24 hours, a separate culture of *Agrobacterium tumefacien* harbouring the transformation vectors is added to the *D. tertiolecta* agar plates. This co-incubation is left for 24-48 hours before *D. tertiolecta* is harvested from the plates by irrigation with media. A cell count is again performed on the harvested cells before *D. tertiolecta* is plated on antibiotic selection plates at the required volume. See section 2.4.2 for a full description of the methods used for the *Agrobacterium* transformation procedure.

6.2.1.1 Preliminary spot-tests to ascertain appropriate selective agent

In order to identify an appropriate antibiotic selective agent and approximate effective concentration for *D. tertiolecta* to be used in the *Agrobacterium*-mediated transformtion, initial spot tests were set up using a panel of antibiotics at a range of concentrations. The antibiotics tested were: hygromycin B (at concentrations of 10 – 500 µg/mL), geneticin G418 (at concentrations of 5 – 200 µg/mL) and zeocin (at concentrations of 5 – 200 µg/mL). *D. tertiolecta* cultures were grown to mid-log phase in 0.4 M NaCl DMxx media and used to prepare solutions of the following concentrations in cells/mL: 1 x 10^7, 5 x 10^6, 2.5 x 10^6, 1 x 10^6 and 5 x 10^5. These solutions were spotted onto agar plates (1.5% agar) containing each concentration of antibiotic, along with positive controls containing no antibiotic, and assessed for growth after four, thirteen and twenty days (see Table 6.2, Table 6.3 and Table 6.4). The goal of the tests is to obtain the maximum possible cell suspension concentration where killing is achieved with the minimum amount of antibiotic. This is to facilitate the appearance of transformants in the final transformation while guaranteeing dieback of non-transformed cells.

Table 6.2: Initial spot tests to ascertain an appropriate selective agent and concentration for *D. tertiolecta* four days after inoculation. The antibiotics tested were: hygromycin B, geneticin G418 and zeocin. Each spot was assessed for the presence greening: '+' donates greening; '-' donates no greening. $n = 2$ for biological replicates.

Selective agent	**Cell suspension concentration (cells/mL)**				
	1 x 10^7	**5 x 10^6**	**2.5 x 10^6**	**1 x 10^6**	**5 x 10^5**
None	+	+	+	+	+
10 µg/mL Hyg B	+	+	+	+	+
25 µg/mL Hyg B	+	+	+	+	+
50 µg/mL Hyg B	+	+	+	+	+
100 µg/mL Hyg B	+	+	+	+	+
200 µg/mL Hyg B	+	+	+	+	+
400 µg/mL Hyg B	+	+	+	+	+
500 µg/mL Hyg B	+	+	+	+	+
500 µg/mL Hyg B	+	+	+	+	+
5 µg/mL G418	+	+	+	+	+
10 µg/mL G418	+	+	+	+	+
25 µg/mL G418	+	+	+	+	+
50 µg/mL G418	+	+	+	+	+
100 µg/mL G418	+	+	+	+	+
200 µg/mL G418	+	+	+	+	+
5 µg/mL Zeocin	+	+	+	+	+
10 µg/mL Zeocin	+	+	+	-	-
25 µg/mL Zeocin	+	+	-	-	-
50 µg/mL Zeocin	+	-	-	-	-
100 µg/mL Zeocin	-	-	-	-	-
200 µg/mL Zeocin	-	-	-	-	-

Table 6.3: Initial spot tests to ascertain an appropriate selective agent and concentration for *D. tertiolecta* sixteen days after inoculation. The antibiotics tested were: hygromycin B, geneticin G418 and zeocin. Each spot was assessed for the presence greening: '+' donates greening; '-' donates no greening. $n = 2$ for biological replicates.

Selective agent	**Cell suspension concentration (cells/mL)**				
	1 x 10^7	**5 x 10^6**	**2.5 x 10^6**	**1 x 10^6**	**5 x 10^5**
None	+	+	+	+	+
10 µg/mL Hyg B	+	+	+	+	+
25 µg/mL Hyg B	+	+	+	+	+
50 µg/mL Hyg B	+	+	+	+	+
100 µg/mL Hyg B	+	+	+	+	+
200 µg/mL Hyg B	+	+	+	+	+
400 µg/mL Hyg B	+	+	+	+	+
500 µg/mL Hyg B	+	+	+	+	+
500 µg/mL Hyg B	+	+	+	+	+
5 µg/mL G418	+	+	+	+	+
10 µg/mL G418	+	+	+	+	+
25 µg/mL G418	+	+	+	+	+
50 µg/mL G418	+	+	+	+	+
100 µg/mL G418	+	+	+	+	+
200 µg/mL G418	+	+	+	+	+
5 µg/mL Zeocin	+	+	+	+	+
10 µg/mL Zeocin	+	+	+	-	-
25 µg/mL Zeocin	-	-	-	-	-
50 µg/mL Zeocin	-	-	-	-	-
100 µg/mL Zeocin	-	-	-	-	-
200 µg/mL Zeocin	-	-	-	-	-

Table 6.4: Initial spot tests to ascertain an appropriate selective agent and concentration for *D. tertiolecta* twenty days after inoculation. The antibiotics tested were: hygromycin B, geneticin G418 and zeocin. Each spot was assessed for the presence greening: '+' donates greening; '-' donates no greening. $n = 2$ for biological replicates.

Selective agent	**Cell suspension concentration (cells/mL)**				
	1 x 10^7	**5 x 10^6**	**2.5 x 10^6**	**1 x 10^6**	**5 x 10^5**
None	+	+	+	+	+
10 µg/mL Hyg B	+	+	+	+	+
25 µg/mL Hyg B	+	+	+	+	+
50 µg/mL Hyg B	+	+	+	+	+
100 µg/mL Hyg B	+	+	+	+	+
200 µg/mL Hyg B	+	+	+	+	+
400 µg/mL Hyg B	+	+	+	+	+
500 µg/mL Hyg B	+	+	+	+	+
500 µg/mL Hyg B	+	+	+	+	+
5 µg/mL G418	+	+	+	+	+
10 µg/mL G418	+	+	+	+	+
25 µg/mL G418	+	+	+	+	+
50 µg/mL G418	+	+	+	+	+
100 µg/mL G418	+	+	+	+	+
200 µg/mL G418	+	+	+	+	+
5 µg/mL Zeocin	+	+	+	+	+
10 µg/mL Zeocin	+	+	+	-	-
25 µg/mL Zeocin	-	-	-	-	-
50 µg/mL Zeocin	-	-	-	-	-
100 µg/mL Zeocin	-	-	-	-	-
200 µg/mL Zeocin	-	-	-	-	-

After twenty days from inoculation there were no changes in the state of each spot. The results show that both hygromycin B and geneticin G418 were unable to kill *D. tertiolecta* in the tested conditions. zeocin was shown to be effective for complete killing at a cell suspension concentration of 1 x 10^6 cells/mL and a zeocin concentration of 10 µg/mL (see Figure 6.3). Thus, zeocin was selected for further whole-plate tests to establish the selection parameters for the post-transformation soft agar plates.

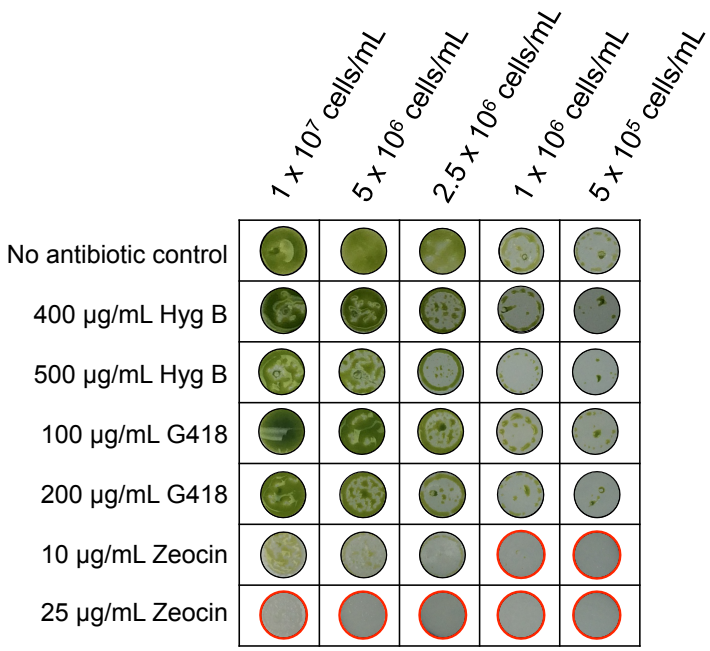


Figure 6.3: Representatives spots from initial spot tests to ascertain appropriate selective agent and concentration for *D. tertiolecta* twenty days after inoculation. The antibiotics tested were: hygromycin B, geneticin G418 and zeocin. Each spot was assessed for the presence greening. Zeocin was the only effective antibiotic in the conditions tested. Red outline donates full killing of cells. $n = 2$ for biological replicates.

6.2.1.2 Preliminary whole-plate antibiotic sensitivity tests

In order to determine the appropriate concentration of antibiotic and number of cells per plate for the *Agrobacterium* transformation, a series of whole-plate tests were set up in soft agar (see section 2.4.2 for details of soft agar preparation). Based on the results of the initial spot tests (see 6.2.1.1), whole-plate trials only included zeocin, which was the only effective antibiotic. The goal of the trials was to determine the minimum concentration of antibiotic that would give full killing at the highest possible number of cells per plate in order to facilitate the growth of transformants while killing all untransformed cells. The concentrations of zeocin tested were 25, 50 and 100 µg/mL; this was increased from the results of the initial spot tests due to the increased quantity of cells and the exhaustion and encapsulation effects on the antibiotic in soft agar. Quantities of cells/plate tested were: 10^8, 10^7 and 10^6. Plates were assessed for growth after 24 days; a concentration of 50 µg/mL was sufficient for full killing at 10^8 cells/plate. At 25 µg/mL, growth was observed on all plates after 24 days. The results are displayed in Table 6.5 and Figure 6.4.

Table 6.5: Whole-plate zeocin antibiotic trials in soft agar to ascertain appropriate concentration for *D. tertiolecta* at a range of cell concentrations (cells/plate). Each plate was assessed for the presence growth after twenty-four days: '+' donates greening; '-' donates no greening. $n = 2$ for biological replicates.

Zeocin Concentration (µg/mL)	Cells/plate		
	10^8	10^7	10^6
25	+	+	+
50	-	-	-
100	-	-	-

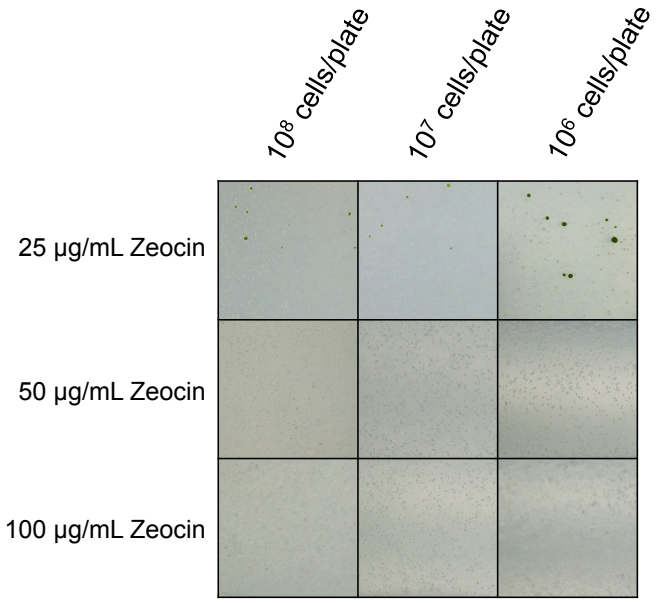


Figure 6.4: Whole-plate zeocin antibiotic trials in soft agar to ascertain appropriate concentration for *D. tertiolecta* at a range of cell concentrations (cells/plate). Each plate was assessed for the presence growth after twenty-four days. Colonies appeared on all plates containing 25 µg/mL; no growth was detected in any of the conditions tested at higher concentrations. $n = 2$ for biological replicates; 1 representative for each condition tested shown.

6.2.1.3 Preliminary whole-plate plating efficiency tests

Plating efficiency is a measure of the number of colonies originating from single cells, and is expressed as the percentage of cells that form colonies when a given number of cells are plated. The plating efficiency of *D. tertiolecta* in soft agar containing no antibiotic selective agent was calculated in order to provide an estimate for the number of cells that should be added to each plate following transformation. Wild type *D. tertiolecta* cells at mid-log phase were plated

quantities of 10, 100 and 1000 cells/plate in triplicate; the number of colonies formed on each plate were then counted. A plating efficiency of 75.0 ± 1.9% was calculated, the results were linear in the range of cells/plate tested ($R^2 = 0.99$). Results are shown in Table 6.6 and Figure 6.5.

Table 6.6: Plating efficiency of *D. tertiolecta* in soft agar. 10, 100 and 1000 cells were added per plate and the final average percentage of single cells plated which formed colonies was determined ($n = 3$ for biological replicates).

Cells/plate	Colonies formed	Average	St Dev
10 (1*n*)	8		
10 (2*n*)	4	7.33	3.06
10 (3*n*)	10		
10^2 (1*n*)	78		
10^2 (2*n*)	77	77	1.00
10^2 (3*n*)	76		
10^3 (1*n*)	748		
10^3 (2*n*)	744	747.33	3.06
10^3 (3*n*)	750		
Plating efficiency (%)		75.01	1.87

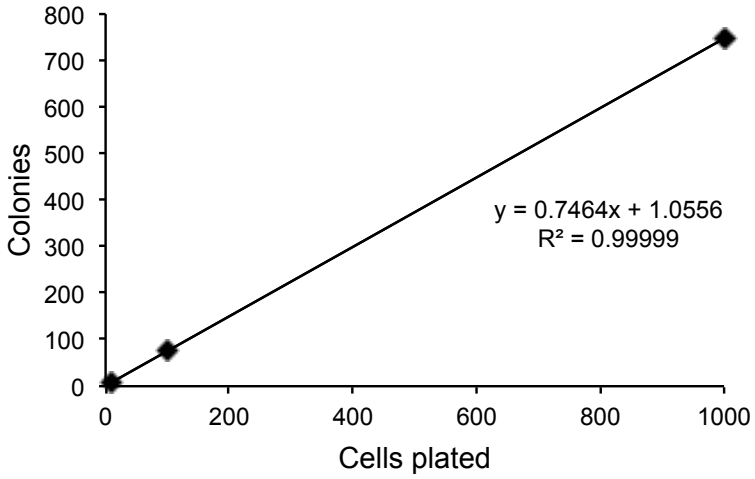


Figure 6.5: Plating efficiency of *D. tertiolecta* in soft agar. Plot of colonies formed versus cells plated. 10, 100 and 1000 cells were added per plate (n = 3 for biological replicates). A plating efficiency of 75.01% was calculated and was linear over the range tested.

6.2.1.4 *Agrobacterium*-mediated procedure results

Four trials were undertaken for the *Agrobacterium* transformation, however none of these trials yielded transformants of *D. tertiolecta.* A number of key parameters were adjusted. Two different vectors were used: pCambia.ble2aChr.SD (see Figure 6.1), and a simpler vector, pCambia.ble.SD, which is identical to the pCambia.ble.2aChr.SD vector except that no mCherry fluorescent reporter protein and 2A sequence is included. The number of *D. tertiolecta* cells co-incubated with *Agrobacterium* before transferring to the antibiotic selection plates was adjusted, the cell concentrations tested were: 5 x 10^6, 1 x 10^6, and 1 x 10^8 cells/plate. Additionally, the co-incubation phase was tested at 24 and 48 hours. The post-transformation selection plates were incubated in two different light conditions: blue/red LEDs and 200 μE fluorescent light. Finally, the post-transformation selection plates were incubated at 24 and 28 °C. For each transformation attempt, positive and negative controls were included. *D. tertiolecta* cells co-incubated with *Agrobacterium* strains harbouring the pCambia.ble2aChr.SD and pCambia.ble.SD vectors conferring resistance to zeocin were plated on no-antibiotic post-transformation plates as a positive control (n = 2 for biological replicates). Growth was observed on all of the positive control plates in the conditions tested, confirming that the *D. tertiolecta* cells had not lost viability over the course of the transformations. For the negative controls, *D. tertiolecta* was co-incubated with *Agrobacterium* harbouring a pCambia.hyg,SD vector conferring resistance to hygromycin B and plated on the standard zeocin-containing plates used for the test groups (n = 2 for biological replicates). No growth was observed on any of the negative control plates confirming that the zeocin antibiotic was ensuring cell death of

non-transformed cells. See section 2.4.2 for a full description of the methods used for the *Agrobacterium* transformation procedure. A summary of the parameters tested is given in Table 6.7.

Table 6.7: Summary of the parameters tested for *Agrobacterium*-mediated transformation of *D. tertiolecta.*

Trial No.	Vector	*D. tertiolecta* cell concentration at co-incubation step (cells/plate)	Co-incubation length (hours)	Post-transformation Light conditions	Post-transformation Temperature (°C)
1	pCambia.ble.SD and pCambia.ble2aChr.SD	5×10^6	24	red/blue LEDs	24
2	pCambia.ble.SD and pCambia.ble2aChr.SD	1×10^6	48	red/blue LEDs and 200 µE fluorescent light	24
3	pCambia.ble.SD and pCambia.ble2aChr.SD	1×10^8	48	red/blue LEDs and 200 µE fluorescent light	24 and 28
4	pCambia.ble2aChr.SD	1×10^6	48	200 µE fluorescent light	24

6.2.2 Biolistics procedure

Another method of DNA delivery was subsequently tested; a biolistics procedure, which have proven to be robust in other species of microalgae (see section 1.3.1.1.2). *D. tertiolecta* cells are relatively large (aprox. 7 µM average diameter; see Figure 2.2a and 2.2b), providing a large target for the DNA-coated gold beads. See section 2.4.3 for a full description of the methods used.

D. tertiolecta was acclimated to low salt media (0.2 M NaCl) and grown to early logarithmic phase in liquid media for transformation with two different vectors: pAlg.Ble2AChr.SD, containing the same vector elements as those shown in Figure 6.1, and the simpler pAlg.Ble.SD vector, containing only the *ble* selection marker. The selection plates where inoculated at two different concentrations; either 100 % or 50 % of the initial transformation plates.

After 6 weeks of incubation, no transformants appeared on any of the selection plates. No growth was observed on the no DNA negative control plates, further indicating that the

antibiotic had been fully effective; growth was observed as a lawn of cells on the no antibiotic positive control plates proving that the cells were viable after the transformation procedure.

6.2.3 Electroporation procedure

6.2.3.1 Preliminary antibiotic sensitivity tests

The electroporation procedure was performed in standard f/2-Si media. Preliminary tests were carried out in order to ascertain the appropriate concentration of zeocin for the selection plates (Figure 6.6). Cells were grown to the logarithmic phase and plated on 0, 10, 25, 50, 75 μg/mL at the cell concentration to be used for the transformation: 5×10^7 cells/plate. The plates were then monitored for growth for thirty days; after this period effective killing was achieved with a concentration of 50 μg/mL; at all lower concentrations growth was present. A zeocin concentration of 50 μg/mL was therefore used for the electroporation transformation procedure.

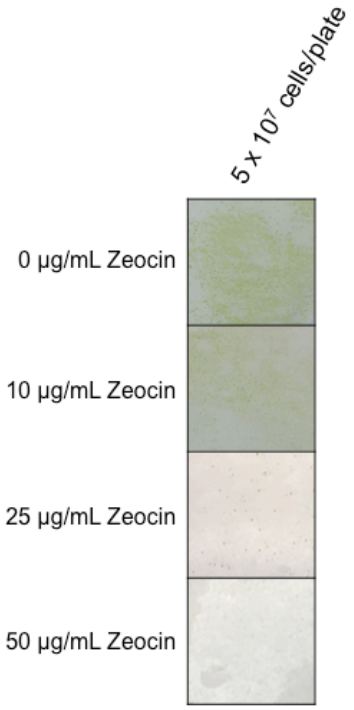


Figure 6.6: Antibiotic trials on f/2-Si agar plates to ascertain appropriate concentration for *D. tertiolecta* at a cell concentration of 5 x 10^7 cells/plate. Each plate was assessed for the presence growth after thirty days. Growth was present on all plates containing 25 µg/mL of zeocin or below; no growth was detected at a concentration of 5 x 10^7 cells/mL. n = 2 for biological replicates; 1 representative for each condition tested shown.

6.2.3.2 Electroporation procedure results

Electroporation was performed based on methods previously developed in *D. tertiolecta* but which did not yield stably transformed cell lines (Walker *et al*., 2005b). See section 2.4.4 for a full description of the methods used.

The p35BT vector used for the electroporation procedure (Figure 6.7) was assembled via homologous recombination in yeast and contained a cauliflower mosaic virus promoter (CaMV35S) amplified from the pGWB402 vector, a *ble* selection marker conferring resistance to zeocin amplified from the pP66 vector and a NOS terminator amplified from the pGWB402 vector.

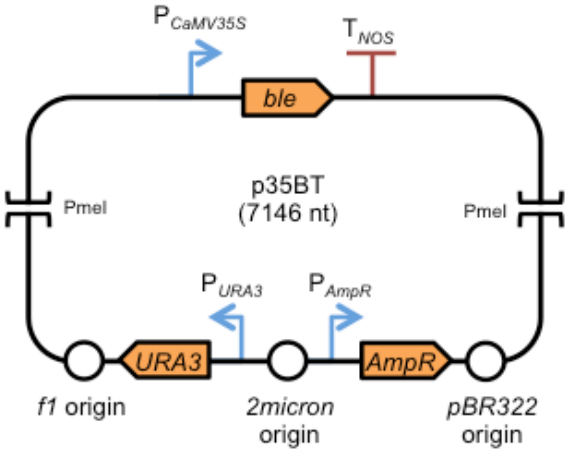


Figure 6.7: p35BT vector used for *D. tertiolecta* electroporation procedure. The p35BT cassette included a cauliflower mosaic virus 35S promoter (P_{CMV35S}), a *ble* selection marker conferring resistance to zeocin and a NOS terminator (T_{NOS},). The cassette was released from the vector backbone at PmeI restriction sites. Figure drawn to SBOL standards (Galdzicki *et al*., 2014).

D. tertiolecta was grown in standard f/2-Si media before being harvested for electroporation with the p35BT vector. After a three day recovery period the transformation cultures were plated along with control plates and incubated in continuous 200 µmol photons m^{-2} s^{-1} fluorescent light at 24 °C and monitored for growth. No transformants were appeared on any of the transformation plates and no growth was present on the negative control plates. The positive control plates did show growth, proving that the electroporation procedure had not caused a loss of cell viability. The procedure was repeated twice more with the same results.

6.2.4 *D. tertiolecta* glass beads procedure

Lastly, a glass beads procedure, adapted from previously described methods established in *D. salina* (Feng *et al*., 2009), was carried out. See section 2.4.5 for a full description of the glass beads transformation procedure.

D. tertiolecta was grown to the logarithmic phase in f/2-Si media before being harvested and agitated in the presence of glass beads with the pAlg.Ble.SD cassette conferring resistance to zeocin. The cells were recovered in f/2-Si media and incubated overnight in dim light. The following day the transformation cultures were plated on 50 µg/mL zeocin f/2-Si selection plates. The plates were incubated in continuous 200 µmol photons m^{-2} s^{-1} fluorescent light at 24 °C and monitored for appearance of transformants. Negative controls were inoculated using no DNA reaction mixtures on plates containing zeocin; positive controls were inoculated using reaction mixtures containing the pAlg.Ble.SD cassette on agar plates containing zeocin. No transformants appeared on any of the transformation plates and no growth was present on the

negative control plates. The positive control plates did show growth, proving that the electroporation procedure had not caused a loss of cell viability. The procedure was repeated once more with the same results.

6.3 Discussion

Dunaliella sp. have been of commercial interest for several decades due to the innate halotolerance among the genus and the array of native fine chemicals produced: lipids, carotenoids, glycerol, vitamins, minerals and proteins (Hosseini Tafreshi and Shariati, 2009). Furthermore, the salinity tolerance and ease of culture of *Dunaliella* sp. make them good candidates for molecular level study of environmental stress resistance mechanisms. *Dunaliella salina* is the best known source of *β*-cerotene and is grown commercially in mass culture worldwide (Hosseini Tafreshi and Shariati, 2009).

D. tertiolecta is one of the more well characterized species within the genus and has been proposed for application as a source of single cell protein and minerals and as an ecological indicator (Hosseini Tafreshi and Shariati, 2009). Additionally, the chloroplast of *D. tertiolecta* shows promise for production of recombinant enzymes (Georgianna *et al*., 2013). Despite this, a method for transformation of the nuclear genome has not been established. To address this, a variety of transformation techniques were tested to enable the possibility of carrying out transgenic research based on integration to the nuclear genome. Four transformation techniques were tested: *Agrobacterium*-mediated transfection, biolistics, electroporation and agitation in the presence of glass beads. However, none of these methods were successful in generating transformed cell lines. For all of the methods tested, the presence of growth on the positive control plates demonstrated that the transformation processes did not prevent cell growth.

The tumour inducing properties of *Agrobacterium* and the potential exploitation of its mode of action for transgenic research has been a focus of research for over 100 years (Nester, 2011). Agrobacterium has an extremely large host range; in addition to plant genetic engineering where *Agrobacterium* is used widely and routinely, transformation has been achieved in fungi (Bundock *et al*., 1995), human cells (Kunik *et al*., 2001) and algae (Kumar *et al*., 2004). This would suggest that Agrobacterium would be a suitable choice for transformation of the eukaryote *D. tertiolecta.* Indeed, successful transformation has been achieved in the closely related alga *D. bardawil* (Anila *et al*., 2011). *Dunaliella* lack a rigid cell wall (see Figure 2.2a and 2.2b) and are therefore potentially amenable to transformation methods that relay on increasing the permeability of the cell membrane such as electroporation and agitation glass beads. *D. salina* has been transformed by both of these methods (Geng *et al*., 2003; Feng *et al*., 2009). Electroporation was used to establish stable transformation with the hepatitis B surface

antigen under the control of a ubiquitin-Ω promoter (Geng *et al*., 2003). Glass beads was subsequently used to generate stable transformants with improved transformation efficiency (Feng *et al*., 2009).

Transient and stable transformation of *D. tertiolecta* with the *ble* selection marker under the control of the endogenous *RbcS1* promoter and terminator by electroporation has been reported (Walker *et al*., 2005b). Low transformation efficiency was observed and was likely due to rapid degradation of foreign DNA, which was shown by Southern blot analysis (Walker *et al*., 2005b) (Figure 6.8). At present this analysis provides the most likely explanation for the absence of a robust transformation procedure for *D. tertiolecta*. The establishment of a chloroplast transformation system for *D. tertiolecta* may be explained by the absence of this degradation effect (Georgianna *et al*., 2013).

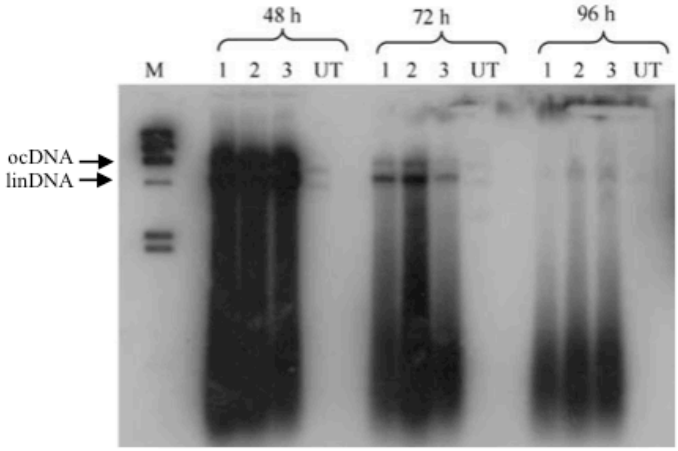


Figure 6.8: Southern blot showing degradation of foreign DNA in transiently transformed *D. tertiolecta* cells. A DIG-labelled *ble* gene probe was used for hybridization and revealed the presence of plasmid DNA as a smear with a lower molecular weight than the transforming plasmid. The first three lanes for each time interval are cells electroporated in the presence of plasmid DbleFLAG1.2; the fourth lane is a no DNA control (UT). M, DIG-labelled DNA marker; ocDNA, open circular plasmid DNA; linDNA, linear plasmid DNA. Figure adapted from Walker *et al*. (2005b).

A very highly expressed non-native CaMV35S promoter was used in our electroporation tests, which were unsuccessful (see section 6.2.3). Walker *et al*. also used a highly expressed constitutive promoter but that of an endogenous *RbcS1* gene; this promoter was also used heterologously to successfully generate transformants of *C. reinhardtii* (Walker *et al*., 2005b, 2005a). Although generally preferable, constitutive promoters with high levels of expression can lead to gene silencing and prevention of cell growth (Geng *et al*., 2003; Jiang *et al*., 2005). Li *et al*. addressed this issue by designing a vector containing an inducible promoter of the nitrate reductase gene capable of promoting temporally controllable expression in *D. salina* that could be induced by nitrate and supressed by ammonium (Li *et al*., 2008). The development of a similar system in *D. tertiolecta* may facilitate the generation of stable transgenic cell lines.

Genetic engineering in higher eukaryotes such as *Dunaliella* is also impeded by the low expression levels of randomly inserted transgenes due to the effects of chromatin structure and regulatory elements, referred to as 'position effects' (Peach and Velten, 1991; Chen and Zhang, 2016). One potential approach to alleviate this problem is the introduction of defined chromatin elements with no negative regulatory or chromatin structure effects on transgene expression (Kim *et al*., 2004). Wang *et al*. deployed this strategy; isolating a DNA fragment (DSM 2) from *D. salina* and showing that this sequence could improve the expression of transgenes by more than 5 fold.

In summary, several possible factors may explain the difficulties associated with nuclear transformation of *Dunaliella* sp., in particular *D. tertiolecta*: (i) the ability to introduce foreign DNA to the host cell, (ii) the ability of this foreign DNA to withstand host degradation systems, (iii) the ability of the transgenes to withstand gene silencing and position effects, (iv) the efficiency of the promoter used to drive expression of the selection marker, and (v) the ability of host cells to maintain viability and grow efficiently in the post-transformation incubation conditions. The degradation of foreign DNA by *D. tertiolecta* shown by Walker *et al*. has revealed a potential barrier to the establishment of a nuclear transformation system (Walker *et al*., 2005b). However, it is evident that further investigations are required to develop more efficient and stable transformation procedures and to identify regulatory elements capable of promoting robust expression of transgenes in *Dunaliella*. The inherent difficulty in transformation of this organism and the lack of a nuclear genome emphasizes the importance of strain selection considering the large role played by these technologies in algal biotechnology.

Chapter 7: Synthesis

7.1 Development of *N. gaditana* and *D. tertiolecta* for biotechnology

Microalgae are promising sources of biofuels and high value bio-based chemicals and are advantageous over terrestrial plants as feedstock for several reasons (see section 1.2). However, the cost of cultivation and rates of productivity of microalgae must be improved by selection of robust strains and subsequent genetic modification to further improve these strains is of urgent importance to develop microalgae-based biorefineries (Ng *et al*., 2017).

The applicability of technologies for genetic engineering in marine microalgae, in particular transgenic approaches, are highly species specific and must be established and developed on a species-to-species basis (Doron *et al*., 2016; Scott *et al*., 2010; Fon Sing *et al*., 2013). The aim of this thesis was to improve the biotechnological potential of the biofuel candidate strains *N. gaditana* and *Dunaleilla tertiolecta* using transgenic approaches.

Despite the commercial potential of *D. tertiolecta* and the fact that it is still widely studied within the field of biotechnology, no efforts to establish a nuclear transformation system have been successful, including those presented in this thesis (Chapter 6) (Tan *et al*., 2016; Georgianna *et al*., 2013; Kim *et al*., 2017). This is a major barrier to the employment of more sophisticated genome editing techniques and limits methods for strain modification to the use of UV or chemical mutagens to induce random mutations (Lee *et al*., 2014; Liu *et al*., 2015; Doan and Obbard, 2012). In order to address this issue, more emphasis should be placed on recognizing transformation systems as a key component when selecting species for biotechnological development. Additionally, reports of unsuccessful approaches, which are currently sparse, should be represented in the literature in order to better assess the applicability of each method for a given species.

It should also be emphasized that the dynamics of key metabolic pathways also vary significantly among different species of microalgae. For example, a recent study on the response of *D. tertiolecta* to nitrogen deprivation showed a reduction in neutral lipids to 1% of dry cell

weight (DCW), while starch was rapidly accumulated up to 46% DCW (Tan *et al*., 2016). This response is not seen in other oleaginous species, such as *Nannochloropsis* sp., *C. reinhardtii* or *Chlorella vulgaris*, where nitrogen deprivation is classically used to upregulate lipid accumulation (Tan *et al*., 2016). These differences illustrate that metabolic pathways should also be considered on a species-to-species basis, and further emphasize the importance of the ability to transform the host organism, given that caution should be taken when designing directed transgenic approaches based on metabolic information from a different species.

In contrast, *N. gaditana* is becoming regarded as a potential biofuel feedstock strain amenable to more advanced genetic engineering techniques which could conceivably lead to increases in large-scale productivities (Ajjawi *et al*., 2017). This was initiated by the sequencing of the nuclear and organellar genomes and establishment of an electroporation transformation procedure (Radakovits *et al*., 2012; Jinkerson *et al*., 2013). To build upon this progress, a library of standardized parts for synthetic biology should be accumulated to facilitate design-build-test cycles for genetic modification (Ramey *et al*., 2015).

See Table 7.1 for a summary of the tools and technologies currently available for genetic engineering of *N. gaditana* and *D. tertiolecta* and Figure 7.1 for a flow diagram for selection and development of algae for biotechnology.

Table 7.1: Biotechnological evaluation of *N. gaditana* and *D. tertiolecta.* Italics denote tools established or developed in this thesis.

Species	**Applications**	**Tools and technologies for genetic engineering**	**References**
D. tertiolecta	Biofuels, recombinant enzyme expression	Chemical and UV mutagenesis, Transcriptomic data, Chloroplast transformation	(Kim *et al*., 2017) (Rismani-Yazdi *et al*., 2011) (Georgianna *et al*., 2013)
N. gaditana	Biofuels	Nuclear transformation, Genomic and organellar genome sequences, Transcriptomic data, Chemical and UV mutagenesis, *Insertional mutagenesis,* CRISPR/Cas9, RNA interference, *Fluorescent reporter proteins,* *Promoter trapping systems,* *Inducible expression systems*	(Radakovits *et al*., 2012) (Radakovits *et al*., 2012; Finazzi *et al*., 2010) (Corteggiani Carpinelli *et al*., 2014) (Perin *et al*., 2015) (Ajjawi *et al*., 2017) (Ajjawi *et al*., 2017)

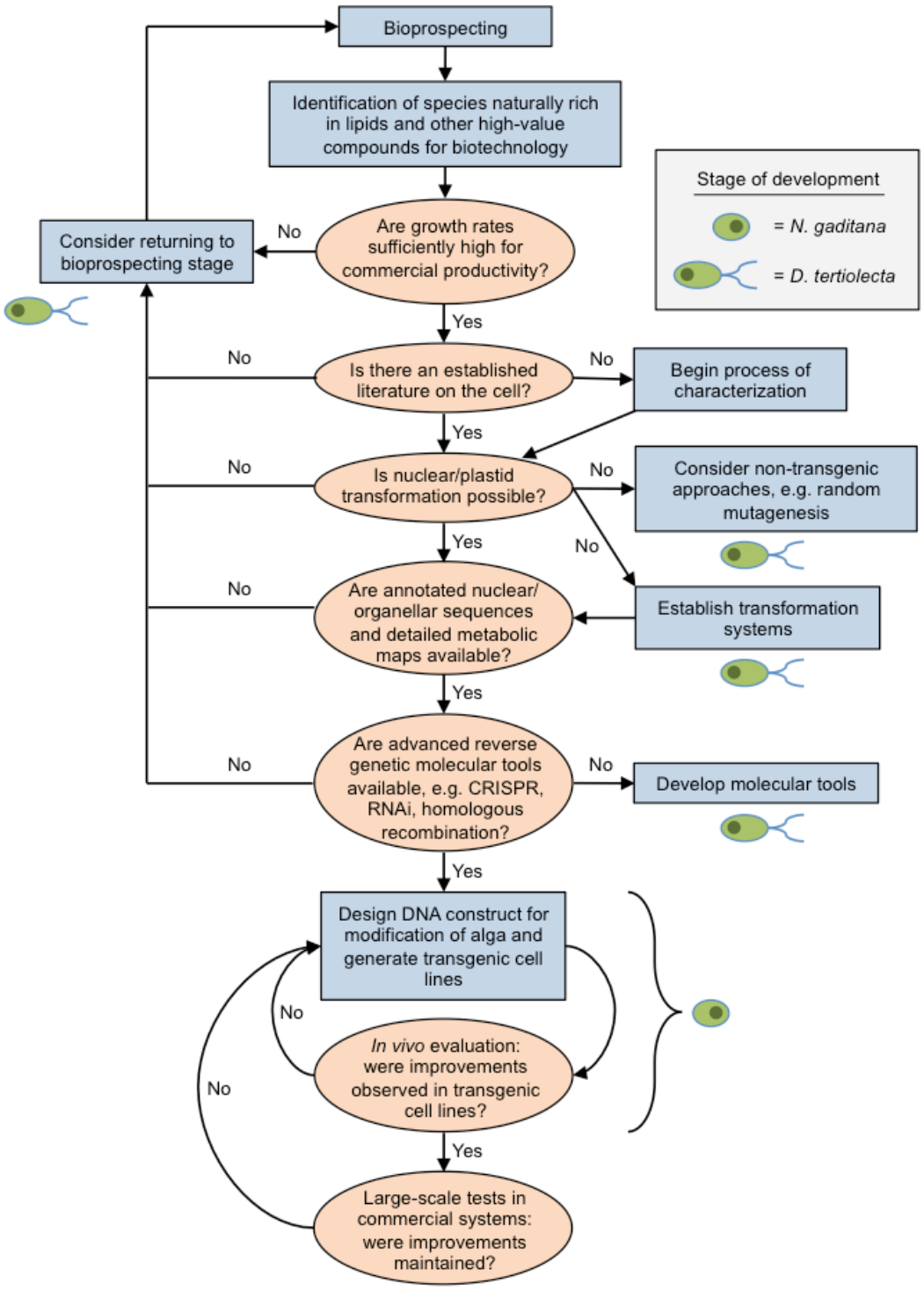


Figure 7.1: Flow diagram for selection and development of algae for biotechnology. The stage of development for *N. gaditana* and *D. tertiolecta* are indicated.

7.2 Expanding the molecular toolkit for *N. gaditana*

Although *N. gaditana* is currently one of the more suitable oleaginous strains for molecular biology and genetic engineering, there are many key technologies that must be established in

order for the host to obtain a similar level of development as model organisms such as *C. reinhardtii* (Purton, 2007; Rasala *et al*., 2014). This section outlines the tools and technologies that have been developed for use in *N. gaditana* in this thesis.

7.2.1 Inducible expression

To build upon the advances in principle technologies in *N. gaditana*, a targeted genetic approach was employed to design an inducible system capable of temporally controllable expression of heterologous genes (see Chapter 3). The inducible expression vector (Figure 3.3) used a promoter sequence for the nitrate reductase gene, isolated using genomic information, to regulate expression (Figure 3.5, 3.8 and 3.9) (Radakovits *et al*., 2012). The inducible system could be controlled with no concurrent reductions on growth and a variety of regulation strategies were developed (Figure 3.9). Inducible expression systems are of great importance to basic research, allowing controllable expression of toxic proteins and facilitating other transgenic techniques such as RNA interference (RNAi), which can be used to attenuate expression of a specific gene (Doron *et al*., 2016; Wei *et al*., 2017; Rohr *et al*., 2004; Mussgnug *et al*., 2007; Jia *et al*., 2009).

7.2.2 Green fluorescent protein reporters

The expression characteristics of the nitrate reductase promoter were reported using an enhanced green fluorescent protein (eGFP; Figure 3.5). This was the first use of GFP in *N. gaditana*, and it proved to be an effective fluorescent gene reporter. Fluorescence proteins are used widely in molecular biology both to report the expression of genes and to tag and visualize proteins of interest using live-cell fluorescence microscopy (Yuste, 2005; Berepiki *et al*., 2010, 2011). It has been demonstrated here that eGFP may be a suitable fluorescent protein for such non-invasive *in situ* applications.

7.2.3 Promoter trapping

While targeted approaches for promoter characterization are most favourable in instances where a gene of interest is already known, as in the characterization of the nitrate reductase promoter (Chapter 3), random approaches are useful for discovery of novel promoters whose functions are not yet known; this was demonstrated by the promoter trapping system presented in Chapter 4. Promoter trapping has been used extensively to identify several novel genetic elements in plants but has until now has only been explored in marine microalgae in the model organism *C.*

reinhartii (Blanvillain and Gallois, 2008; Buzas *et al*., 2005; Vila *et al*., 2012). The ability to select for highly expressing promoters in *N. gaditana* using selection of insertional mutants on zeocin (Figure 4.2) and subsequently to identify the corresponding gene (Figure 4.5; Table 4.2) could be carried out in different nutritional and environmental conditions to identify novel promoters for biotechnology.

Due to the inherent advantages of both targeted and random techniques, both strategies should be employed in efforts to ascertain useful promoter sequences in *N. gaditana* and other marine microalgae in future studies. Using random approaches like the promoter trap to discover new promoters of unknown function and subsequently detailed characterization using the eGFP reporter.

7.2.4 The challenge of position effects

Any transgenic approach in which integration of plasmid DNA is random within the genome is vulnerable to position effects, whereby the expression level of elements within the plasmid are affected by the position of the plasmid within the genome (Chen and Zhang, 2016; Peach and Velten, 1991). This can make ascertaining the expression level of promoter sequences challenging. Figure 3.5 shows one strategy that can be used to overcome this; multiple cell lines transformed with the expression cassette were analysed in order to determine an average expression level. It should be noted that when analysing the expression level of an endogenous promoter using an expression cassette, the promoter has been taken out of its native context, which will also have been subject to position effects. Therefore the true expression level cannot be definitively determined using such techniques. In addition, the use of fluorescent reporters analyses expression at the protein level and does not discriminate transcriptional and translational effects. Analysis of the native transcript levels of the corresponding genes can be used to compliment expression cassette analysis of promoters as is shown in section 4.4.7 (Corteggiani Carpinelli *et al*., 2014).

An alternative way to circumvent the issue of position effects is to target integration of the transgenic DNA to a specific location in the genome using homologous recombination, thereby eliminating inconsistency in the analysis due to random integration (see Figure 7.1) (Sodeinde and Kindle, 1993; Weeks, 2011). Homologous recombination of transgenic DNA has been established in *Nannochloropsis oceanica*; however, the targeted gene (nitrate reductase) was one with an easily confirmable phenotype, allowing rapid isolation of strains in which homologous recombination had been successful (Kilian *et al*., 2011). It has yet to be seen whether efforts to target genes with less critical and easily assayable roles will be met with the same success.

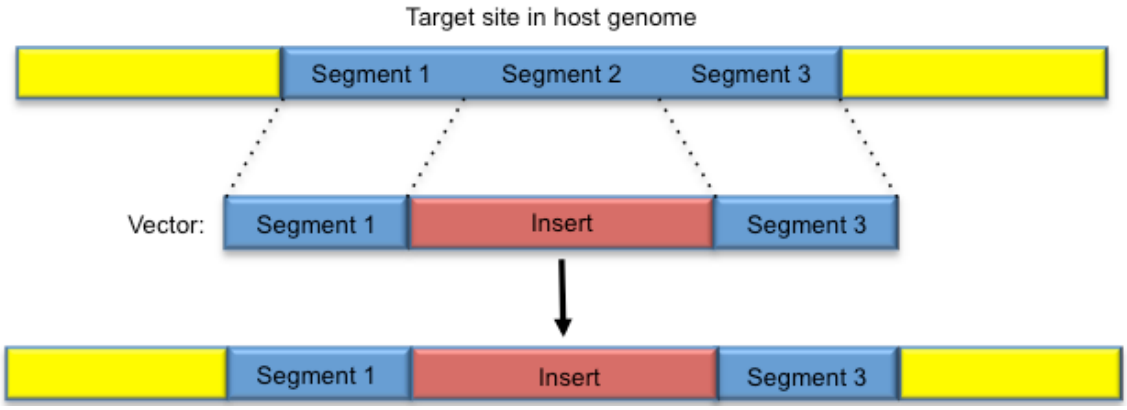


Figure 7.2: Targeted integration of vector DNA via homologous recombination. Transformation of the host with a vector containing homologous sequences (segments 1 and 3) to those outside the targeted region of integration (segment 2) results in replacement of the target site with the vector insert in the transformed cell lines. Homologous recombination allows targeted gene knock out and replacement.

7.3 Fourier-transform infrared (FTIR) spectroscopy in algal biotechnology

FTIR spectroscopy can be used to rapidly analyse the macromolecular profile of cellular biomass; this has been shown in microalgal species (Wagner *et al*., 2010; Dean *et al*., 2010; Stehfest *et al*., 2005). Section 5.2.2 presents the first use of FTIR spectroscopy in *N. gaditana.* Using previously published peak attributions it was possible to identify several distinctive absorption peaks corresponding to different compounds (see section 2.2.7). The ability to rapidly and accurately assess the relative lipid content of *N. gaditana* using the lipid:amide I peak height ratio is demonstrated in Figure 5.13. A procedure for high-throughput analysis of cell lines in 96-well format is presented in section 2.5.2 and 5.2.2 that can be used as a basis for the design of similar screening platforms for *N. gaditana* and other species of microalgae. While this analysis was qualitative, methods for quantitative lipid analysis using reference compound standards are in the early stages of development and may soon become applicable to high-throughput screening platforms such as these (Wagner *et al*., 2010). Table 7.2 provides a comparison of the different methods currently available for lipid analysis in microalgae.

Table 7.2: Summary of lipid analysis techniques in microalgae. N, quantitative; L, qualitative; S, quantitative with standards; H, semi quantitative. Table adapted from Hounslow *et al*. (2017).

Technique type	Quantitative / qualitative	Advantages	Disadvantages
Gravimetric	N	Direct measurements	Extracts other compounds than lipids Requires large amount of culture Solvent type favours certain lipid types
Fluorescent lipid dyes	L / S (depending on dye used)	*In situ* Relatively fast procedure Process many samples at once	Dye efficiency varies between species Requires optimization
Colorimetric	S	Can be done on small volumes	Relies on extraction Signal decreases with chain length No structural information
Raman microscopy	H	Small samples needed *In situ* and non destructive Information of lipid properties	Photosynthetic pigments can interfere with signal Components of low abundance can not be detected
Density equilibrium			
Direct transesterification/ *in situ* esterification	N	Measures true potential of algal biomass to be converted to FAME Detailed structural information	
Nuclear magnetic resonance spectroscopy (NMR)	S	Detailed structural information	Interference from other compounds, which is inconsistent between samples Water can interfere so samples must be freese-dried
Gas chromatography mass spectroscopy (GSMS)	N	Lipid classes obtained as well as quantification	Assumes derivatized fatty acids can correctly identify all lipid portions in correct proportions Assumes that all lipid types will ionize at the same rate

Table 7.2 continued

Technique type	Quantitative / qualitative	Advantages	Disadvantages
Fourier-transform infrared spectroscopy (FTIR)	L / S	No extraction required Rapid analysis Little sample preparation Analyse many samples at once No need for normalization of samples for qualitative analysis Small culture volume required	Not specific to different lipid types

7.4 Commercial exploitation of *N. gaditana* for biofuel production

Microalgae are currently the best suited feedstock for biofuel production; however further optimization of fast growing and oil rich strains is required for such systems to become economically viable (Doron *et al*., 2016; Markou and Nerantzis, 2013; Mata *et al*., 2010; Borowitzka and Moheimani, 2013).

7.4.1 Forward and reverse strategies for species improvement

Reverse genetic strategies, whereby a known gene is altered to generate an improved phenotype have been explored comparatively less in marine microalgae than forward genetic strategies whereby a phenotype is isolated from a pool of mutant cell lines before the genetic basis is probed (see Table 1.3 and section 5.1) (Doron *et al*., 2016). This is due to the dependency of reverse approaches on the availability of transformation systems, sequence information and detailed metabolic pathways (Doron *et al*., 2016). A forward genetic approach, consisting of random insertional mutagenesis and phenotypic screening was used in Chapter 5 to select for strains with mutations that may increase overall lipid productivity.

7.4.1.1 Forward genetic screening for increased ROS resistance

Reactive oxygen species (ROS) are produced in algal cells as a byproduct of the saturation of photosynthesis under high light; ROS may then cause photoinhibition – a reduction in the rate of photosynthesis and/or the maximum photosynthetic efficiency (Erickson *et al*., 2015). A

phenotypic screen for increased resistance to menadione (a ROS inducing agent) yielded two strains of *N. gaditana,* NgROS1 and NgROS2*,* with increased resistance to both menadione and the ROS hydrogen peroxide (Figure 5.5). Both strains also exhibited improved growth characteristics in a series of high light treatments (Figure 5.6, 5.7 and 5.8).

Although the mutated gene in NgROS1 was identified (Figure 5.10 and Table 5.2), further molecular characterization of both strains is required in order to shed light on the genetic basis of these potentially improved phenotypes. To confirm that the mutations conferred by the integration of the transformation vector are responsible for the NgROS1 and NgROS2 phenotypes, it would be necessary to compliment these cell lines with the respective genes that were disrupted and observe if the wild type phenotype is restored. Further physiological experiments should also be performed to robustly show that NgROS1 and NgROS2 exhibit improved growth and productivity in conditions akin to commercial growth systems. Should these molecular and physiological requirements be met, targeting these genes in different species of algae and observing the effects may reveal the broader implications of the discovery of these genes for algal biotechnology.

7.4.1.2 Forward genetic screening for increased lipid content

The FTIR screen for mutant strains of *N. gaditana* with an increased ratio of lipid to protein (see section 5.2.2) showed a high degree of variance between the cell lines (Figure 5.16). Of the three highest lipid:amideI exhibiting strains that were characterized again in liquid culture, two had a significantly higher lipid:amideI ratio than the wild type (NgL1.4 and NgL2.61; Figure 5.18), which further suggests that the variance can be attributed in part to natural metabolic responses. Screening for increased lipid content in agar plate format (Figure 5.15) may have catalysed the variance due to slight differences in the way each cell line was re-streaked from the initial transformation plate and subsequently incubated, despite the fact that measures were taken to ensure an even exposure to nutrients, temperature and light among the mutants. FACS-based approaches that are carried out on a pooled culture of mutants may therefore be more suitable for isolating increased lipid mutants (Terashima *et al*., 2015). However, the establishment of a 96-well liquid culture format screening platform aided by liquid handling robots may allow for normalization of cell concentrations, and in turn growth phase effects, which would likely facilitate unlocking the full potential of the FTIR as a screening tool for lipid content.

The NgL1.4 and NgL2.61 mutants were isolated close to the end of the PhD and as such it was not possible to characterize the mutations and physiology of the strains in depth. However, given the challenges of the agar plate-based screen and the modest increases in lipid content in

the mutants compared to other studies, it may be more beneficial to focus on optimization of the FTIR screening platform itself, which shows the potential to become a powerful screening tool for other macromolecular composition changes in addition to lipid content (Ajjawi *et al*., 2017; Wagner *et al*., 2010; Dean *et al*., 2010; Stehfest *et al*., 2005).

7.4.1.3 Bridging the gap between forward and reverse strategies

The fundamental advantage of random forward genetic approaches is that they are non-biased; however one of the main caveats of these approaches is that there can be off-target effects and the genetic basis of isolated phenotypes can often be hard to discern (Doron *et al*., 2016). Although, reverse genetic approaches have the advantage of targeting specific genes, finding target genes and altering them is made difficult by a lack of more advanced tools and technologies, which must be established and optimized on a species-to-species basis (Doron *et al*., 2016). In addition, reverse genetic approaches do not allow for a non-bias assessment of gene function, which limits the possibility of discovery of new novel genetic elements.

In a recent study by Ajjawi and colleagues, a suite of techniques were combined, incorporating elements of both forward and reverse approaches, leading to the production of an *N. gaditana* mutant with a 2-fold increase in lipid content and little reduction in growth compared to the wild type. Firstly, RNA-seq was used to identify 20 possible negative regulators of lipid production; the use of RNA-seq is a non-biased approach akin to forward genetic strategies. Subsequently, a reverse genetic pipeline was deployed to knockout 18 of these transcription factors, revealing one transcription factor (a homolog of fungal $Zn(II)_2Cys_6$-encoding genes) that conferred an increase in portioning of total carbon to lipid from 20% to 40-55% when knocked out. Finally, RNAi was used to attenuate expression of this gene to fine-tune the balance between growth and lipid production in the final mutant strain (Ajjawi *et al*., 2017) . This work illustrates the importance of the availability of a suite of techniques in a given species and shows that it is possible to overcome paradigmatic challenges in algal biotechnology by deploying them in combination (see Figure 7.2).

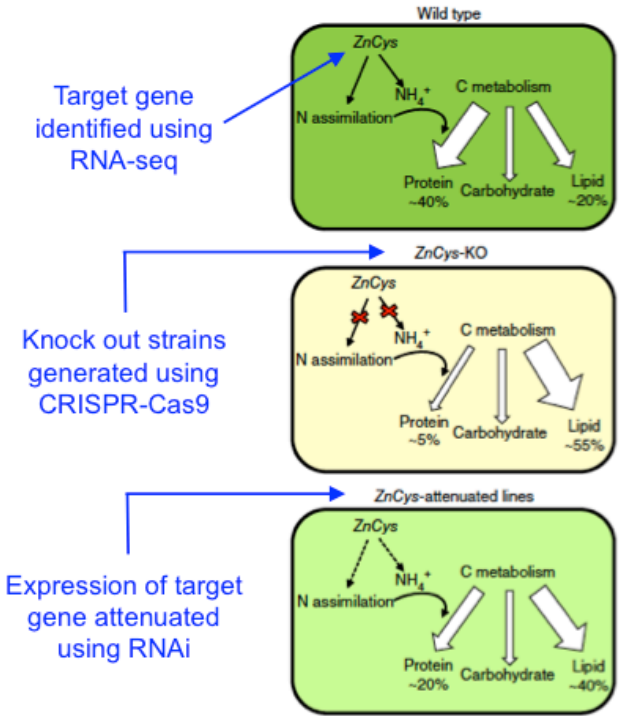


Figure 7.3: Schematic of carbon partitioning to protein, lipid and carbohydrate in wild type, *ZnCys*-KO and *ZnCys*-attenuated cell lines of *N. gaditana* and techniques used to identify and fine-tune *ZnCys* in order to generate cell line with doubled lipid productivity. Figure adapted from Ajjawi *et al*. (2017).

7.4.2 Challenges and bottlenecks in algal strain improvement

Forward and reverse genetic approaches have many strengths and the potential to generate strains that may one day be suitable for large-scale commercial growth systems. However, the limitations associated with the respective techniques used gives rise to bottlenecks in exploitation of microalgae for biotechnology.

The most critical step of random mutagenesis approaches is the screening method. In order to isolate the desired phenotype, tens of thousands of cell lines may have to be screened. Thus, there is a demand for the screening process to be simple and fast. The most effective types of screen are those that are based on survival of mutant and wild type cells under a certain condition compared to the control population. This was demonstrated by the selection of *N. gaditana* mutants on menadione at levels fatal to wild type cells in section 5.2.1.1.

For any mutant screen it is essential to ensure that all the mutants in the screen are subject to the same environmental conditions, which can often be challenging, especially considering the large populations of mutants that must be screened. In addition, mutants that have the desired phenotype but in conditions that are different from the screening procedures will be undetected

as false negatives. On the other hand, mutants with false positive phenotypes can be identified relatively easy by re-testing in more controlled conditions, for example in section 5.2.2.5.

Following the isolation of a verified mutant, another challenge is to maintain the cell line such that there is little chance of reversion to a wild type phenotype. The chance of reversion to wild type increases the more times the cell line is subcultured in addition to the possibility of the reverted cells outcompeting the mutant cells. Maintaining the isolated mutant in the presence of antibiotic either continuously or periodically can lower the likelihood of reversion. Alternatively, preparing cryopreserved cells and taking fresh aliquots periodically can minimize subculturing of cell lines and hence the chance of reversion. Highly expressed foreign genes are more vulnerable to gene silencing; this could be avoided partially through the use of inducible promoters (see Figure 3.10).

A major issue with any newly generated strain is its suitability for scale up to a larger industrial growth system, where conditions are more diverse and less controlled than at the level of a culture maintained in a laboratory. As a result of this, productivity in such systems never reaches that of an optimized laboratory culture (Hlavova *et al*., 2015). Unless large-scale experiments are performed it is uncertain how well a given strain will cope with the scale up process.

The field of algal biotechnology is still in its infancy and the basic tools required for strain improvement are still in the process of development. Proof-of-principle experiments are possible and will serve to bridge the gap between laboratory cultures, closed system bioreactors and finally open outdoor systems (Terashima et al., 2015; Hlavova *et al*., 2015; Ajjawi et al., 2017). In order to accelerate this progress, there is a need to increase the number of microalgal strains amenable to transgenic approaches and to develop more sophisticated genome-editing techniques in industrially relevant strains.

7.5 Conclusion

Microalgae have been cultivated on a large scale for several decades. This has so far been limited to wild type strains, identified through bioprospecting due to their optimal properties. However, for the field of algal biotechnology to be further advanced, improvements must be made to wild type strains at the commercial scale using synthetic biology, in order for biofuel and high-value product production systems to become economically feasible. So far, more advanced genetic tools that may make such improvements possible have only been established expansively in model organisms such as *C. reinhardtii*. This thesis describes the establishment of several principle technologies in the oleaginous alga *N. gaditana*; namely, an inducible

expression system for precise temporal control of transgene expression at no cost to growth, the use of an enhanced green fluorescent protein to report the expression of genes and a promoter trapping workflow for discovery of novel endogenous promoter regions that will also serve as a template for the establishment of similar strategies in other microalgal strains. These new genetic tools will contribute to the development of this non-model organism towards its commercial viability. In addition, high-throughput forward genetic screening was used to isolate two mutant strains of *N. gaditana* with increased resistance to photoinhibition, caused by reactive oxygen species, which also exhibited improved growth under a variety of high light treatments. Gaining a better understanding of the genetic basis of these mutant strains may have broad implications for the field of biotechnology if the phenotypes can maintained in larger growth systems. The elements for an FTIR-based high-throughput screen for mutant strains of *N. gaditana* with increased lipid content were assembled, leading to the isolation of two mutants strains with a verified increase in lipid content. Comprehensive genetic and physiological characterization of these mutants was not possible in the time frame of the PhD. However, this was the first demonstrated use of FTIR spectroscopy as a high-throughput screening tool and indicating how the technology can be used to rapidly analyse changes in biomass composition on a relevant timescale and is worthy of further exploration for this application. The future directions of this research are to further characterize the phenotypes of the mutant strains generated and to better understand their genetic basis. Should this be successful it may be possible to combine the mutations of distinct cell lines, if proven to be beneficial, into one cell line, in view of integrating all of the phenotypic improvements made.

Appendices

Appendix A

A.1 p35BT – Genbank formatted vector sequence

```
LOCUS       p35BT.xdna    7146 bp      DNA     circular   15/11/2017
DEFINITION
ACCESSION
VERSION
SOURCE
  ORGANISM
COMMENT     Serial Cloner Genbank Format
COMMENT     SerialCloner_Type=DNA
COMMENT     SerialCloner_Comments=Homologous recombination reaction between [p35s] as a
target
             sequence  and [PCR(pGWB402.xdna)] as an inserted sequence.
COMMENT     SerialCloner_Ends=0,0,,0,
FEATURES             Location/Qualifiers
     source          1..1863
                     /label=source [5'] [5']
                     /SerialCloner_Color=&h2F4F4F
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=organism="pRS426" - mol_type="other DNA" -
1..5726 \
     source          3608..7146
                     /label=source [3'] [3']
                     /SerialCloner_Color=&h2F4F4F
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=organism="pRS426" - mol_type="other DNA" -
1..5726 \
     misc_feature    complement(29..51)
                     /label=pGEX_3_primer
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pGEX_3_primer" - complement(29..51) \
     promoter        202..415
                     /label=URA3_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3_promoter" - 202..415 \
     gene            417..1217
                     /label=URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - 417..1217 \
```

```
     CDS             417..1220
                     /label=ORF frame 3_URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - note="ORF frame 3"
- translation="MSKATYKERAATHPSPVAAKLFNIMHEKQTNLCASLDVRTTKEL -
LELVEALGPKICLLKTHVDILTDFSMEGTVKPLKALSAKYNFLLFEDRKFADIGNTVK -
LQYSAGVYRIAEWADITNAHGVVGPGIVSGLKQAAEEVTKEPRGLLMLAELSCKGSLS - TGEYTKGTVDIAKSDKDFVIGFI \
     rep_origin      complement(1483..1789)
                     /label=f1_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="f1_origin" - complement(1483..1789) \
     misc_feature    complement(1822..1863)
                     /label=lacZ_a [5']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="lacZ_a" - complement(1822..1965) \
     misc_feature    complement(3605..1863)
                     /label=M13_pUC_rev_primer [3'] [5'] [3']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="M13_pUC_rev_primer" -
complement(2176..2198) \
     rep_origin      complement(3970..4589)
                     /label=pBR322_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pBR322_origin" - complement(2550..3169) \
     CDS             complement(4744..5604)
                     /label=ORF frame 3_Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
note="ORF frame 3" - translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY -
IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE -
YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL - DRWEPELNEAI \
     gene            complement(4744..5604)
                     /label=Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
complement(3324..4184) \
     promoter        complement(5646..5674)
                     /label=AmpR_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="AmpR_promoter" - complement(4226..4254) \
     rep_origin      complement(5736..6707)
                     /label=2micron2_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="2micron2_origin" - complement(4316..5287)
\
     misc_feature    complement(6699..6746)
                     /label=FRT
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="FRT" - complement(5279..5326) \
     misc_feature    3599..3628
```

```
                     /label=YY
                     /SerialCloner_Color=&h0000FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1841..1870
                     /label=XX
                     /SerialCloner_Color=&hFF0000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2766..3140
                     /label=Sh ble
                     /SerialCloner_Color=&h008080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1919..2753
                     /label=35S promoter_regulatory
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=regulatory_class="promoter" - note="35S
promoter" - 325..1159 \
     misc_feature    3156..3565
                     /label=NOS terminator_regulatory
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=regulatory_class="terminator" - note="NOS
terminator" - 2911..3320 \
     misc_feature    3172..3431
                     /label=NOS terminator_regulatory
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=regulatory_class="terminator" - note="NOS
terminator" - complement(3354..3613) \
ORIGIN
        1 tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca
       61 cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg
      121 ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc
      181 accataccac agcttttcaa ttcaattcat catttttttt ttattctttt ttttgatttc
      241 ggtttctttg aaatttttttt gattcggtaa tctccgaaca gaaggaagaa cgaaggaagg
      301 agcacagact tagattggta tatatacgca tatgtagtgt tgaagaaaca tgaaattgcc
      361 cagtattctt aacccaactg cacagaacaa aaacctgcag gaaacgaaga taaatcatgt
      421 cgaaagctac atataaggaa cgtgctgcta ctcatcctag tcctgttgct gccaagctat
      481 ttaatatcat gcacgaaaag caaacaaact tgtgtgcttc attggatgtt cgtaccacca
      541 aggaattact ggagttagtt gaagcattag gtcccaaaat ttgtttacta aaaacacatg
      601 tggatatctt gactgatttt tccatggagg gcacagttaa gccgctaaag gcattatccg
      661 ccaagtacaa ttttttactc ttcgaagaca gaaaatttgc tgacattggt aatacagtca
      721 aattgcagta ctctgcgggt gtatacagaa tagcagaatg ggcagacatt acgaatgcac
      781 acggtgtggt gggcccaggt attgttagcg gtttgaagca ggcggcagaa gaagtaacaa
      841 aggaacctag aggccttttg atgttagcag aattgtcatg caagggctcc ctatctactg
      901 gagaatatac taagggtact gttgacattg cgaagagcga caaagatttt gttatcggct
      961 ttattgctca aagagacatg ggtggaagag atgaaggtta cgattggttg attatgacac
     1021 ccggtgtggg tttagatgac aagggagacg cattgggtca acagtataga accgtggatg
     1081 atgtggtctc tacaggatct gacattatta ttgttggaag aggactattt gcaaagggaa
     1141 gggatgctaa ggtagagggt gaacgttaca gaaaagcagg ctgggaagca tatttgagaa
     1201 gatgcggcca gcaaaactaa aaaactgtat tataagtaaa tgcatgtata ctaaactcac
     1261 aaattagagc ttcaatttaa ttatatcagt tattacccta tgcggtgtga aataccgcac
     1321 agatgcgtaa ggagaaaata ccgcatcagg aaattgtaaa cgttaatatt ttgttaaaat
     1381 tcgcgttaaa tttttgttaa atcagctcat tttttaacca ataggccgaa atcggcaaaa
     1441 tcccttataa atcaaaagaa tagaccgaga tagggttgag tgttgttcca gtttggaaca
     1501 agagtccact attaaagaac gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg
     1561 gcgatggccc actacgtgaa ccatcaccct aatcaagttt tttggggtcg aggtgccgta
     1621 aagcactaaa tcggaaccct aaagggagcc cccgatttag agcttgacgg ggaaagccgg
     1681 cgaacgtggc gagaaaggaa gggaagaaag cgaaaggagc gggcgctagg gcgctggcaa
     1741 gtgtagcggt cacgctgcgc gtaaccacca cacccgccgc gcttaatgcg ccgctacagg
     1801 gcgcgtcgcg ccattcgcca ttcaggctgc gcaactgttg GGAAGGGCGA TCGGTGCGGG
     1861 CCGTTTAAAC cgttgtaaaa cgacggccag tgccaagctt gcatgcctgc aggtccccag
     1921 attagccttt tcaatttcag aaagaatgct aacccacaga tggttagaga ggcttacgca
     1981 gcaggtctca tcaagacgat ctacccgagc aataatctcc aggaaatcaa ataccttccc
     2041 aagaaggtta aagatgcagt caaaagattc aggactaact gcatcaagaa cacagagaaa
     2101 gatatatttc tcaagatcag aagtactatt ccagtatgga cgattcaagg cttgcttcac
     2161 aaaccaaggc aagtaataga gattggagtc tctaaaaagg tagttcccac tgaatcaaag
     2221 gccatggagt caaagattca aatagaggac ctaacagaac tcgccgtaaa gactggcgaa
```

```
2281 cagttcatac agagtctctt acgactcaat gacaagaaga aaatcttcgt caacatggtg
2341 gagcacgaca cacttgtcta ctccaaaaat atcaaagata cagtctcaga agaccaaagg
2401 gcaattgaga cttttcaaca aagggtaata tccggaaacc tcctcggatt ccattgccca
2461 gctatctgtc actttattgt gaagatagtg gaaaaggaag gtggctccta caaatgccat
2521 cattgcgata aaggaaaggc catcgttgaa gatgcctctg ccgacagtgg tcccaaagat
2581 ggaccccccac ccacgaggag catcgtggaa aaagaagacg ttccaaccac gtcttcaaag
2641 caagtggatt gatgtgatat ctccactgac gtaagggatg acgcacaatc ccactatcct
2701 tcgcaagacc cttcctctat ataaggaagt tcatttcatt tggagagaac acgggggact
2761 ctagaATGGC CAAGTTGACC AGTGCCGTTC CGGTGCTCAC CGCGCGCGAC GTCGCCGGAG
2821 CGGTCGAGTT CTGGACCGAC CGGCTCGGGT TCTCCCGGGA CTTCGTGGAG GACGACTTCG
2881 CCGGTGTGGT CCGGGACGAC GTGACCCTGT TCATCAGCGC GGTCCAGGAC CAGGTGGTGC
2941 CGGACAACAC CCTGGCCTGG GTGTGGGTGC GCGGCCTGGA CGAGCTGTAC GCCGAGTGGT
3001 CGGAGGTCGT GTCCACGAAC TTCCGGGACG CCTCCGGGCC GGCCATGACC GAGATCGGCG
3061 AGCAGCCGTG GGGGCGGGAG TTCGCCCTGC GCGACCCGGC CGGCAACTGC GTGCACTTCG
3121 TGGCCGAGGA GCAGGACTGA tgataacagc gcttagagct cgaatttccc cgatcgttca
3181 aacatttggc aataaagttt cttaagattg aatcctgttg ccggtcttgc gatgattatc
3241 atataatttc tgttgaatta cgttaagcat gtaataatta acatgtaatg catgacgtta
3301 tttatgagat gggtttttat gattagagtc ccgcaattat acatttaata cgcgatagaa
3361 aacaaaatat agcgcgcaaa ctaggataaa ttatcgcgcg cggtgtcatc tatgttacta
3421 gatcgggaat tggttccgga accaattcgt aatcatggtc atagctgttt cctgtgtgaa
3481 attgttatcc gctcacaatt ccacacaaca tacgagccgg aagcataaag tgtaaagcct
3541 ggggtgccta atgagtgagc taactcacat taggctgaat taggcgcgcc tatttctgGT
3601 TTAAACTTGT TATCCGCTCA CAATTCCAca caacatagga gccggaagca taaagtgtaa
3661 agcctggggt gcctaatgag tgaggtaact cacattaatt gcgttgcgct cactgcccgc
3721 tttccagtcg ggaaacctgt cgtgccagct gcattaatga atcggccaac gcgcggggag
3781 aggcggtttg cgtattgggc gctcttccgc ttcctcgctc actgactcgc tgcgctcggt
3841 cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggt tatccacaga
3901 atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg
3961 taaaaaggcc gcgttgctgg cgtttttcca taggctccgc cccctgacg agcatcacaa
4021 aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat accaggcgtt
4081 tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgccgctta ccggatacct
4141 gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct gtaggtatct
4201 cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc
4261 cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa gacacgactt
4321 atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtgc
4381 tacagagttc ttgaagtggt ggcctaacta cggctacact agaaggacag tatttggtat
4441 ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa
4501 acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa
4561 aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaacga
4621 aaactcacgt taagggattt tggtcatgag attatcaaaa aggatcttca cctagatcct
4681 tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa cttggtctga
4741 cagttaccaa tgcttaatca gtgaggcacc tatctcagcg atctgtctat ttcgttcatc
4801 catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct taccatctgg
4861 ccccagtgct gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat
4921 aaaccagcca gccggaaggg ccgagcgcag aagtggtcct gcaactttat ccgcctccat
4981 ccagtctatt aattgttgcc gggaagctag agtaagtagt tcgccagtta atagtttgcg
5041 caacgttgtt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc
5101 attcagctcc ggttcccaac gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa
5161 agcggttagc tccttcggtc ctccgatcgt tgtcagaagt aagttggccg cagtgttatc
5221 actcatggtt atggcagcac tgcataattc tcttactgtc atgccatccg taagatgctt
5281 ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc ggcgaccgag
5341 ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca catagcagaa ctttaaaagt
5401 gctcatcatt ggaaaacgtt cttcggggcg aaaactctca aggatcttac cgctgttgag
5461 atccagttcg atgtaaccca ctcgtgcacc caactgatct tcagcatctt ttactttcac
5521 cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc
5581 gacacggaaa tgttgaatac tcatactctt cctttttcaa tattattgaa gcatttatca
5641 gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg
5701 ggttccgcgc acatttcccc gaaaagtgcc acctgaacga agcatctgtg cttcattttg
5761 tagaacaaaa atgcaacgcg agagcgctaa tttttcaaac aaagaatctg agctgcattt
5821 ttacagaaca gaaatgcaac gcgaaagcgc tattttacca acgaagaatc tgtgcttcat
5881 ttttgtaaaa caaaaatgca acgcgagagc gctaattttt caaacaaaga atctgagctg
5941 cattttaca gaacagaaat gcaacgcgag agcgctattt taccaacaaa gaatctatac
6001 ttctttttg ttctacaaaa atgcatcccg agagcgctat ttttctaaca aagcatctta
6061 gattactttt tttctccttt gtgcgctcta taatgcagtc tcttgataac tttttgcact
6121 gtaggtccgt taaggttaga agaaggctac tttggtgtct attttctctt ccataaaaaa
6181 agcctgactc cacttcccgc gtttactgat tactagcgaa gctgcgggtg catttttca
6241 agataaaggc atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca
6301 gaaagtgata gcgttgatga ttcttcattg gtcagaaaat tatgaacggt ttcttctatt
6361 ttgtctctat atactacgta taggaaatgt ttacattttc gtattgtttt cgattcactc
6421 tatgaatagt tcttactaca attttttttgt ctaaagagta atactagaga taaacataaa
6481 aaatgtagag gtcgagttta gatgcaagtt caaggagcga aaggtggatg ggtaggttat
6541 atagggatat agcacagaga tatatagcaa agagatactt ttgagcaatg tttgtggaag
6601 cggtattcgc aatattttag tagctcgtta cagtccggtg cgtttttggt tttttgaaag
6661 tgcgtcttca gagcgctttt ggttttcaaa agcgctctga agttcctata ctttctagag
6721 aataggaact tcggaatagg aacttcaaag cgtttccgaa aacgagcgct tccgaaaatg
6781 caacgcgagc tgcgcacata cagctcactg ttcacgtcgc acctatatct gcgtgttgcc
6841 tgtatatata tatacatgag aagaacggca tagtgcgtgt ttatgcttaa atgcgtactt
6901 atatgcgtct atttatgtag gatgaaaggt agtctagtac ctcctgtgat attatcccat
6961 tccatgcggg gtatcgtatg cttccttcag cactaccctt tagctgttct atatgctgcc
7021 actcctcaat tggattagtc tcatccttca atgctatcat ttcctttgat attggatcat
```

```
     7081 actaagaaac cattattatc atgacattaa cctataaaaa taggcgtatc acgaggccct
     7141 ttcgtc

//
```

A.2 pLCP – Genbank formatted vector sequence

```
LOCUS       pLCP    11172 bp      DNA     circular   15/11/2017
DEFINITION
ACCESSION
VERSION
SOURCE
  ORGANISM
COMMENT     Serial Cloner Genbank Format
COMMENT     SerialCloner_Type=DNA
COMMENT     SerialCloner_Comments=Homologous recombination reaction between [pPGM2 KO
HPH.xdna
            ] as a target sequence  and [PCR (PLCP-eGFP-Tatpa)] as an in
            serted sequence.
COMMENT     SerialCloner_Ends=0,0,,0,
FEATURES             Location/Qualifiers
     misc_feature    complement(9362..9384)
                     /label=pGEX_3_primer
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pGEX_3_primer" - complement(29..51) \
     promoter        9535..9748
                     /label=URA3_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3_promoter" - 202..415 \
     gene            9750..10550
                     /label=URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - 417..1217 \
     CDS             9750..10553
                     /label=ORF frame 3_URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - note="ORF frame 3"
- translation="MSKATYKERAATHPSPVAAKLFNIMHEKQTNLCASLDVRTTKEL -
LELVEALGPKICLLKTHVDILTDFSMEGTVKPLKALSAKYNFLLFEDRKFADIGNTVK -
LQYSAGVYRIAEWADITNAHGVVGPGIVSGLKQAAEEVTKEPRGLLMLAELSCKGSLS - TGEYTKGTVDIAKSDKDFVIGFI \
     rep_origin      complement(10816..11122)
                     /label=f1_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="f1_origin" - complement(1483..1789) \
     misc_feature    complement(11155..24)
                     /label=lacZ_a [5']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="lacZ_a" - complement(1822..1965) \
     misc_feature    complement(5792..24)
                     /label=M13_pUC_rev_primer [3'] [5'] [3']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="M13_pUC_rev_primer" -
complement(2176..2198) \
     rep_origin      complement(6157..6776)
                     /label=pBR322_origin
                     /SerialCloner_Color=&hF08080
```

```
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pBR322_origin" - complement(2550..3169) \
     CDS             complement(6931..7791)
                     /label=ORF frame 3_Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
note="ORF frame 3" - translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY -
IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE -
YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL - DRWEPELNEAI \
     gene            complement(6931..7791)
                     /label=Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
complement(3324..4184) \
     promoter        complement(7833..7861)
                     /label=AmpR_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="AmpR_promoter" - complement(4226..4254) \
     rep_origin      complement(7923..8894)
                     /label=2micron2_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="2micron2_origin" - complement(4316..5287)
\
     misc_feature    complement(8886..8933)
                     /label=FRT
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="FRT" - complement(5279..5326) \
     misc_feature    2..31
                     /label=XX
                     /SerialCloner_Color=&h008000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    5786..5815
                     /label=YY
                     /SerialCloner_Color=&h8080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2365..2394
                     /label=CC
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4575..4604
                     /label=DD
                     /SerialCloner_Color=&h800040
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4117..4574
                     /label=NOS terminator_regulatory
                     /SerialCloner_Color=&hFF0000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=regulatory_class="terminator" - note="NOS
terminator" - 2911..3320 \
     misc_feature    2395..3087
                     /label=Puep
                     /SerialCloner_Color=&hFF80C0
                     /SerialCloner_Show=True
```

```
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4605..4626
                     /label=P3
                     /SerialCloner_Color=&hFFFF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    5766..5785
                     /label=P4
                     /SerialCloner_Color=&hFFFF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     gene            3088..4116
                     /label=hph_gene
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=gene="hph" - 485..1513 \
     CDS             3088..4116
                     /label=hph_CDS
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=gene="hph" - codon_start=1 - transl_table=11 -
product="hygromycin B phosphotransferase" - protein_id="CAR31381.1" -
db_xref="GI:194473371" - translation="MGKKPELTATSVEKFLIEKFDSVSDLMQLSEGEESRAFSFDVGG -
RGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQ \
     misc_feature    3088..3112
                     /label=hph fw
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4091..4116
                     /label=hph rv
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2..31
                     /label=XX ext
                     /SerialCloner_Color=&h00FFFF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1137..1166
                     /label=GFP ext
                     /SerialCloner_Color=&h00FFFF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1827..1856
                     /label=gfp ext
                     /SerialCloner_Color=&h996633
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1857..1873
                     /label=Tatpa P1 fw
                     /SerialCloner_Color=&hFF00FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2365..2394
                     /label=CC ext
                     /SerialCloner_Color=&h996633
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2345..2364
                     /label=Tatpa P2 rv
                     /SerialCloner_Color=&hFF00FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1117..1136
```

```
                     /label=PLCP P2 rv
                     /SerialCloner_Color=&h0000FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    32..50
                     /label=PLCP P1 fw
                     /SerialCloner_Color=&h0000FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1137..1856
                     /label=gfp
                     /SerialCloner_Color=&h00FF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
ORIGIN
        1 gGGAAGGGCG ATCGGTGCGG GCCGTTTAAA Ccgcattcag gtctagtaag ccgaagctag
       61 aggcactcgg ctcagtggcc acttgccagg cgcagttagg ccatttgacg tcagttttgc
      121 ctcgatctga tggcattgga gccaaaaagg ggtatttttt atcgaagtgc tgatactgct
      181 ggctgggatg aggtactaaa acggtagcac cagtcatgct gggctcgcct acaggaggta
      241 tagcaaagtt tgtggcctcg agtacggcgt cgctggctga gcttaaaaaa aaaggtggaa
      301 tatggagaat gaaggccaag taggatagta tcgatttccc catagttctg gccttgtgta
      361 cagtacaaag aatgatgaaa gttagtttgg tgagctgtcg cttgtcgtgg tcaatgggcc
      421 gtgggaacaa cctcccctgc ccttaaggga gacaaaatcg tggtacgtgc ttgctcgata
      481 aagcccagat tcgtcctacg catttataca aaaatgccac cgctgttgtg caacgcaact
      541 acgagttaaa tccactatcg tgcgctggcg gaacaacgag cacggccgca acaacatcgt
      601 gtgcagggcc ctgctacctg gtattgtttt gttttctgta gggaagacgt cgcttctttc
      661 cgtgttgctg actcggcgac gagcatcagg aaagcccaag agacccaacc cttcttcgtg
      721 caggggaatg tcgtctgtag tatttcgtta ttttcgtgaa aacgatattt gtgcgggggg
      781 tcgagctggt gggttgagga ctacatgctt ggcaccggag gctggttttc tttcttcccc
      841 cttatgaatt cgcctcgcga accagaagcg gcgaattgct catcgcggag aacatgacat
      901 ttaaatcaga atgtgccttg catccttcac cacgtatcgc tatatcttgc atctgaagtg
      961 cttctacgca gcggtcatga tgaaattctt gagacatttt ggtgtgtttt gagtcctgtc
     1021 ttgaagcccg cacggagtgg ctcataaggc gaggaacaaa tgtgatgcca ctcccacctc
     1081 gaaccatttt ttcctatgtc caccctacgc agccacacat ttgaacagaa gcaataATGG
     1141 TGAGCAAGGG CGAGGAGCTG TTCACCgggg tggtgcccat cctggtcgag ctggacggcg
     1201 acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc acctacggca
     1261 agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg cccaccctcg
     1321 tgaccaccct gacctacggc gtgcagtgct tcagccgcta ccccgaccac atgaagcagc
     1381 acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc atcttcttca
     1441 aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac accctggtga
     1501 accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg gggcacaagc
     1561 tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag aagaacggca
     1621 tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag ctcgccgacc
     1681 actaccagca gaacaccccc atcggcgacg gccccgtgct gctgcccgac aaccactacc
     1741 tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac atggtcctgc
     1801 tggagttcgt gaccgccgcc gggatcACTC TCGGCATGGA CGAGCTGTAC AAGTGAtagc
     1861 ggcgcttttg accgggaact cacgcattta atgtattcga aacacgcttt ttaagatctc
     1921 aaatgaaaca tgtatccgta tgctctccac atccgtcctc acgtacggtg tgatgttttt
     1981 ttaatttttt atatgaaagt gtaaggcttc tccattttac tttttgtgtt tctgtcttca
     2041 cagtctttct ctgcgagctg tgagagggtc atgcccttta tgggcacgac ggcaggattt
     2101 tggtttttga aagaatttct gtcggctgta agtcttcaaa aaagaagaaa gcggggacgg
     2161 aaacttcttg cttgtgtgtt atacattaca aagtcctcta caattctaca gtcctagttg
     2221 ttcacgcgac ttcacgcagc agaccagatt tatagctttg catgatcgtc tccgagcatt
     2281 tcagaatgct ataatatgga atgatcttcc gctggcacct cattatgcga tacgcactcc
     2341 atataaacac aggtatccct gtccGTATTC GTTAACTGTT AATTCATGAC ACAAcatcct
     2401 gctgtatgat tttggcacaa cgacgcggtt ggagggtgag ggagatggtg gtgagacatc
     2461 agaaatggag agaataacgg tagagaaaaa aaggaacaca cgcgcccagg atagaaccag
     2521 gaccacattc caagtccctc accagcaaca agagggacac cacgtcggtg aagaggtgcg
     2581 tgagacagag gacaccagac aacgtgcaca aagcgtcgga aagggaggcg accgcttcgc
     2641 tggaggggcg ccgtccacag cctgcgccca accacacaca ttcctcccgc gaggctcaag
     2701 tgccatgccg ccgagtgcca tgctcctgca tccggggcag gtcgacgagc cagcatggtt
     2761 tggtcgccga caacgcagat ttgatggagc ttgggcattt ctttaccgga cacgtcgctt
     2821 cttacatttt tacattccaa ttcagaggtc acacccgcga tcgagaaagc agagagtcgg
     2881 acccccttca agcaggaaag gcaagcaaca tacccatcgt taagtctgat gatggctggt
     2941 gcaaagccaa ggcagccagt cgatatgtta atgatcgggt acgtttaaag gcgtgggcaa
     3001 ggttatttaa cgctggtagc atcgtagagt ttttaacgca actgaagatt tcatgtctga
     3061 cttggagata cggtcggggc agcagctatg ggtaaaaagc ctgaactcac cgcgacgtct
     3121 gtcgagaagt ttctgatcga aaagttcgac agcgtctccg acctgatgca gctctcggag
     3181 ggcgaagaat ctcgtgcttt cagcttcgat gtaggagggc gtggatatgt cctgcgggta
     3241 aatagctgcg ccgatggttt ctacaaagat cgttatgttt atcggcactt tgcatcggcc
     3301 gcgctcccga ttccggaagt gcttgacatt ggggaattca gcgagagcct gacctattgc
     3361 atctcccgcc gtgcacaggg tgtcacgttg caagacctgc ctgaaaccga actgcccgct
     3421 gttctgcagc cggtcgcgga ggccatggat gcgatcgctg cggccgatct tagccagacg
     3481 agcgggttcg gcccattcgg accgcaagga atcggtcaat acactacatg gcgtgatttc
     3541 atatgcgcga ttgctgatcc ccatgtgtat cactggcaaa ctgtgatgga cgacaccgtc
     3601 agtgcgtccg tcgcgcaggc tctcgatgag ctgatgcttt gggccgagga ctgccccgaa
     3661 gtccggcacc tcgtgcacgc ggatttcggc tccaacaatg tcctgacgga caatggccgc
```

```
3721 ataacagcgg tcattgactg gagcgaggcg atgttcgggg attcccaata cgaggtcgcc
3781 aacatcttct tctggaggcc gtggttggct tgtatggagc agcagacgcg ctacttcgag
3841 cggaggcatc cggagcttgc aggatcgccg cggctccggg cgtatatgct ccgcattggt
3901 cttgaccaac tctatcagag cttggttgac ggcaatttcg atgatgcagc ttgggcgcag
3961 ggtcgatgcg acgcaatcgt ccgatccgga gccgggactg tcgggcgtac acaaatcgcc
4021 cgcagaagcg cggccgtctg gaccgatggc tgtgtagaag tactcgccga tagtggaaac
4081 cgacgcccca gcactcgtcc gagggcaaag gaataatgat aacagcgctt agagctcgaa
4141 tttccccgat cgttcaaaca tttggcaata aagtttctta agattgaatc ctgttgccgg
4201 tcttgcgatg attatcatat aatttctgtt gaattacgtt aagcatgtaa taattaacat
4261 gtaatgcatg acgttattta tgagatgggt ttttatgatt agagtcccgc aattatacat
4321 ttaatacgcg atagaaaaca aaatatagcg cgcaaactag gataaattat cgcgcgcggt
4381 gtcatctatg ttactagatc gggaattggt tccggaacca attcgtaatc atggtcatag
4441 ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatacg agccggaagc
4501 ataaagtgta aagcctgggg tgcctaatga gtgagctaac tcacattagg ctgaattagg
4561 cgcgcctatt tctgTTAGTG TCAAACAGTC AAACCAGTTC TACGtgtaga acaggatgca
4621 tggctgctgc aaggctgttt gtgccaacgc cgcaaatgga aaagaaagta agaaaaaaaa
4681 tcctacacag acaacctttc tagcacgaaa atgcccttgc gtacatcttt ttgcgtctca
4741 gtaatattgt tttcttaatc gaaaccatta ttgttcaagt catcgctaag acacttccat
4801 cgtgtgctct tagccgcatc tcggtcctcc ctcgccatgg aacatgtcat ttagggcgac
4861 aaaccacaaa taacatagaa gctttcatcg ataatgcccc atcagtccta tttaaatagt
4921 ccaatacaaa tatataaggg tttgtcattt ctcgtttcct ccggctgtcc catgctgtag
4981 acatgtgtac gtagctagaa ggagccccca tatgttcttt ccgacgatac acatggcttt
5041 aatactatcc ccatagtgac aaagcgtgga cacagaaaca cacacacatg ccactcacat
5101 agaaacggcc aatccacgat cgttacatct ccgctttcgt ttccggcaac gacttgaata
5161 attggggcat ctgctcacgc tcgttttacc cgaccgcaaa aaaaaatgga gagtggccac
5221 catgcttcaa ggagaagctc ggtatatgtc catcccgcgt gcctgcaaag acaaggtgaa
5281 gatgcacccg gcgaccatag tgtgtcatgt acctgtttga cgtttactac agggtgccat
5341 gctttattgg ctggaattgt ggacgcttgt tttctggaat acgaatattt ccccgccctg
5401 tattaattcc cttcaagcag caagctatct taacaaacgt tgtgcggcca aaggaaaata
5461 gccgccgagt ccatatgggg tcttctgact tcaaaatctc gcgcattcca acgtgtaaat
5521 gtgcgatcga aagagaaagc attgacacag ggaataaaaa ctccatcaat tggaaagatc
5581 tctggaatcg gtggcaccct gttccgtgaa gaagcgaagc tttgcttgcg gccgtatgct
5641 cattactcac ccactgactc gttcacataa taaaagtccc gctaaaaccc tacaatttcg
5701 cttccgacgc tccatcattc catcatagaa attgacaggt ctactgaaat catagttcca
5761 cccgttggtg cgtgtccttc ctgcaGTTTA AACTTGTTAT CCGCTCACAA TTCCAcacaa
5821 cataggagcc ggaagcataa agtgtaaagc ctggggtgcc taatgagtga ggtaactcac
5881 attaattgcg ttgcgctcac tgcccgcttt ccagtcggga aacctgtcgt gccagctgca
5941 ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt attgggcgct cttccgcttc
6001 ctcgctcact gactcgctgc gctcggtcgt tcggctgcgg cgagcggtat cagctcactc
6061 aaaggcggta atacggttat ccacagaatc aggggataac gcaggaaaga acatgtgagc
6121 aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag
6181 gctccgcccc cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc
6241 gacaggacta taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt
6301 tccgaccctg ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct
6361 ttctcatagc tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg
6421 ctgtgtgcac gaacccccgg ttcagcccga ccgctgcgcc ttatccggta actatcgtct
6481 tgagtccaac ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat
6541 tagcagagcg aggtatgtag gcggtgctac agagttcttg aagtggtggc ctaactacgg
6601 ctacactaga aggacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa
6661 aagagttggt agctcttgat ccggcaaaca aaccaccgct ggtagcggtg gttttttgt
6721 ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc
6781 tacggggtct gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt
6841 atcaaaaagg atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta
6901 aagtatatat gagtaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat
6961 ctcagcgatc tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac
7021 tacgatacgg gagggcttac catctggccc cagtgctgca atgataccgc gagacccacg
7081 ctcaccggct ccagatttat cagcaataaa ccagccagcc ggaagggccg agcgcagaag
7141 tggtcctgca actttatccg cctccatcca gtctattaat tgttgccggg aagctagagt
7201 aagtagttcg ccagttaata gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt
7261 gtcacgctcg tcgtttggta tggcttcatt cagctccggt tcccaacgat caaggcgagt
7321 tacatgatcc cccatgttgt gcaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt
7381 cagaagtaag ttggccgcag tgttatcact catggttatg gcagcactgc ataattctct
7441 tactgtcatg ccatccgtaa gatgcttttc tgtgactggt gagtactcaa ccaagtcatt
7501 ctgagaatag tgtatgcggc gaccgagttg ctcttgcccg gcgtcaatac gggataatac
7561 cgcgccacat agcagaactt taaaagtgct catcattgga aaacgttctt cggggcgaaa
7621 actctcaagg atcttaccgc tgttgagatc cagttcgatg taacccactc gtgcacccaa
7681 ctgatcttca gcatctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca
7741 aaatgccgca aaaaagggaa taagggcgac acggaaatgt tgaatactca tactcttcct
7801 ttttcaatat tattgaagca tttatcaggg ttattgtctc atgagcggat acatatttga
7861 atgtatttag aaaaataaac aaatagggggt tccgcgcaca tttccccgaa aagtgccacc
7921 tgaacgaagc atctgtgctt cattttgtag aacaaaaatg caacgcgaga gcgctaattt
7981 ttcaaacaaa gaatctgagc tgcatttta cagaacagaa atgcaacgcg aaagcgctat
8041 tttaccaacg aagaatctgt gcttcatttt tgtaaaacaa aaatgcaacg cgagagcgct
8101 aatttttcaa acaaagaatc tgagctgcat ttttacagaa cagaaatgca acgcgagagc
8161 gctattttac caacaaagaa tctatacttc ttttttgttc tacaaaaatg catcccgaga
8221 gcgctatttt tctaacaaag catcttagat tacttttttt ctcctttgtg cgctctataa
8281 tgcagtctct tgataacttt ttgcactgta ggtccgttaa ggttagaaga aggctacttt
8341 ggtgtctatt ttctcttcca taaaaaaagc ctgactccac ttcccgcgtt tactgattac
8401 tagcgaagct gcgggtgcat tttttcaaga taaaggcatc cccgattata ttctataccg
8461 atgtggattg cgcatacttt gtgaacagaa agtgatagcg ttgatgattc ttcattggtc
```

```
     8521 agaaaattat gaacggtttc ttctattttg tctctatata ctacgtatag gaaatgttta
     8581 cattttcgta ttgttttcga ttcactctat gaatagttct tactacaatt tttttgtcta
     8641 aagagtaata ctagagataa acataaaaaa tgtagaggtc gagtttagat gcaagttcaa
     8701 ggagcgaaag gtggatgggt aggttatata gggatatagc acagagatat atagcaaaga
     8761 gatacttttg agcaatgttt gtggaagcgg tattcgcaat attttagtag ctcgttacag
     8821 tccggtgcgt ttttggtttt ttgaaagtgc gtcttcagag cgcttttggt tttcaaaagc
     8881 gctctgaagt tcctatactt tctagagaat aggaacttcg gaataggaac ttcaaagcgt
     8941 ttccgaaaac gagcgcttcc gaaaatgcaa cgcgagctgc gcacatacag ctcactgttc
     9001 acgtcgcacc tatatctgcg tgttgcctgt atatatatat acatgagaag aacggcatag
     9061 tgcgtgttta tgcttaaatg cgtacttata tgcgtctatt tatgtaggat gaaaggtagt
     9121 ctagtacctc ctgtgatatt atcccattcc atgcggggta tcgtatgctt ccttcagcac
     9181 taccctttag ctgttctata tgctgccact cctcaattgg attagtctca tccttcaatg
     9241 ctatcatttc ctttgatatt ggatcatact aagaaaccat tattatcatg acattaacct
     9301 ataaaaatag gcgtatcacg aggccctttc gtctcgcgcg tttcggtgat gacggtgaaa
     9361 acctctgaca catgcagctc ccggagacgg tcacagcttg tctgtaagcg gatgccggga
     9421 gcagacaagc ccgtcagggc gcgtcagcgg gtgttggcgg gtgtcggggc tggcttaact
     9481 atgcggcatc agagcagatt gtactgagag tgcaccatac cacagctttt caattcaatt
     9541 catcattttt tttttattct tttttttgat ttcggtttct ttgaaatttt tttgattcgg
     9601 taatctccga acagaaggaa gaacgaagga aggagcacag acttagattg gtatatatac
     9661 gcatatgtag tgttgaagaa acatgaaatt gcccagtatt cttaacccaa ctgcacagaa
     9721 caaaaacctg caggaaacga agataaatca tgtcgaaagc tacatataag gaacgtgctg
     9781 ctactcatcc tagtcctgtt gctgccaagc tatttaatat catgcacgaa aagcaaacaa
     9841 acttgtgtgc ttcattggat gttcgtacca ccaaggaatt actggagtta gttgaagcat
     9901 taggtcccaa aatttgttta ctaaaaacac atgtggatat cttgactgat ttttccatgg
     9961 agggcacagt taagccgcta aaggcattat ccgccaagta caatttttta ctcttcgaag
    10021 acagaaaatt tgctgacatt ggtaatacag tcaaattgca gtactctgcg ggtgtataca
    10081 gaatagcaga atgggcagac attacgaatg cacacggtgt ggtgggccca ggtattgtta
    10141 gcggtttgaa gcaggcggca gaagaagtaa caaaggaacc tagaggcctt ttgatgttag
    10201 cagaattgtc atgcaagggc tccctatcta ctggagaata tactaagggt actgttgaca
    10261 ttgcgaagag cgacaaagat tttgttatcg gctttattgc tcaaagagac atgggtggaa
    10321 gagatgaagg ttacgattgg ttgattatga cacccggtgt gggtttagat gacaagggag
    10381 acgcattggg tcaacagtat agaaccgtgg atgatgtggt ctctacagga tctgacatta
    10441 ttattgttgg aagaggacta tttgcaaagg gaagggatgc taaggtagag ggtgaacgtt
    10501 acagaaaagc aggctgggaa gcatatttga gaagatgcgg ccagcaaaac taaaaaactg
    10561 tattataagt aaatgcatgt atactaaact cacaaattag agcttcaatt taattatatc
    10621 agttattacc ctatgcggtg tgaaataccg cacagatgcg taaggagaaa ataccgcatc
    10681 aggaaattgt aaacgttaat attttgttaa aattcgcgtt aaatttttgt taaatcagct
    10741 catttttaa ccaataggcc gaaatcggca aaatccctta taaatcaaaa gaatagaccg
    10801 agatagggtt gagtgttgtt ccagtttgga acaagagtcc actattaaag aacgtggact
    10861 ccaacgtcaa agggcgaaaa accgtctatc agggcgatgg cccactacgt gaaccatcac
    10921 cctaatcaag ttttttgggg tcgaggtgcc gtaaagcact aaatcggaac cctaaaggga
    10981 gcccccgatt tagagcttga cggggaaagc cggcgaacgt ggcgagaaag gaagggaaga
    11041 aagcgaaagg agcgggcgct agggcgctgg caagtgtagc ggtcacgctg cgcgtaacca
    11101 ccacacccgc cgcgcttaat gcgccgctac agggcgcgtc gcgccattcg ccattcaggc
    11161 tgcgcaactg tt

//
```

A.3 pNR – Genbank formatted vector sequence

```
LOCUS       pNR_hph    11386 bp      DNA     circular   15/11/2017
DEFINITION
ACCESSION
VERSION
SOURCE
  ORGANISM
COMMENT     Serial Cloner Genbank Format
COMMENT     SerialCloner_Type=DNA
COMMENT     SerialCloner_Comments=Homologous recombination reaction between [pPGM2 KO
HPH.xdna
            ] as a target sequence  and [PCR (PNR-eGFP-Tatpa)] as an ins
            erted sequence.
COMMENT     SerialCloner_Ends=0,0,,0,
FEATURES             Location/Qualifiers
     misc_feature    complement(9576..9598)
                     /label=pGEX_3_primer
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pGEX_3_primer" - complement(29..51) \
     promoter        9749..9962
                     /label=URA3_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3_promoter" - 202..415 \
     gene            9964..10764
                     /label=URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - 417..1217 \
     CDS             9964..10767
                     /label=ORF frame 3_URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - note="ORF frame 3"
- translation="MSKATYKERAATHPSPVAAKLFNIMHEKQTNLCASLDVRTTKEL -
LELVEALGPKICLLKTHVDILTDFSMEGTVKPLKALSAKYNFLLFEDRKFADIGNTVK -
LQYSAGVYRIAEWADITNAHGVVGPGIVSGLKQAAEEVTKEPRGLLMLAELSCKGSLS - TGEYTKGTVDIAKSDKDFVIGFI \
     rep_origin      complement(11030..11336)
                     /label=f1_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="f1_origin" - complement(1483..1789) \
     misc_feature    complement(11369..24)
                     /label=lacZ_a [5']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="lacZ_a" - complement(1822..1965) \
     misc_feature    complement(6006..24)
                     /label=M13_pUC_rev_primer [3'] [5'] [3']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="M13_pUC_rev_primer" -
complement(2176..2198) \
     rep_origin      complement(6371..6990)
                     /label=pBR322_origin
                     /SerialCloner_Color=&hF08080
```

```
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pBR322_origin" - complement(2550..3169) \
     CDS             complement(7145..8005)
                     /label=ORF frame 3_Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
note="ORF frame 3" - translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY -
IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE -
YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL - DRWEPELNEAI \
     gene            complement(7145..8005)
                     /label=Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
complement(3324..4184) \
     promoter        complement(8047..8075)
                     /label=AmpR_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="AmpR_promoter" - complement(4226..4254) \
     rep_origin      complement(8137..9108)
                     /label=2micron2_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="2micron2_origin" - complement(4316..5287)
\
     misc_feature    complement(9100..9147)
                     /label=FRT
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="FRT" - complement(5279..5326) \
     misc_feature    2..31
                     /label=XX
                     /SerialCloner_Color=&h008000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    6000..6029
                     /label=YY
                     /SerialCloner_Color=&h8080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2579..2608
                     /label=CC
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4789..4818
                     /label=DD
                     /SerialCloner_Color=&h800040
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4331..4788
                     /label=NOS terminator_regulatory
                     /SerialCloner_Color=&hFF0000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=regulatory_class="terminator" - note="NOS
terminator" - 2911..3320 \
     misc_feature    2609..3301
                     /label=Puep
                     /SerialCloner_Color=&hFF80C0
                     /SerialCloner_Show=True
```

```
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4819..4840
                     /label=P3
                     /SerialCloner_Color=&hFFFF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    5980..5999
                     /label=P4
                     /SerialCloner_Color=&hFFFF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     gene            3302..4330
                     /label=hph_gene
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=gene="hph" - 485..1513 \
     CDS             3302..4330
                     /label=hph_CDS
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=gene="hph" - codon_start=1 - transl_table=11 -
product="hygromycin B phosphotransferase" - protein_id="CAR31381.1" -
db_xref="GI:194473371" - translation="MGKKPELTATSVEKFLIEKFDSVSDLMQLSEGEESRAFSFDVGG -
RGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQ \
     misc_feature    3302..3326
                     /label=hph fw
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4305..4330
                     /label=hph rv
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2..31
                     /label=XX ext
                     /SerialCloner_Color=&h00FFFF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1351..1380
                     /label=GFP ext
                     /SerialCloner_Color=&h00FFFF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2071..2087
                     /label=Tatpa P1 fw
                     /SerialCloner_Color=&hFF00FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2579..2608
                     /label=CC ext
                     /SerialCloner_Color=&h996633
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2559..2578
                     /label=Tatpa P2 rv
                     /SerialCloner_Color=&hFF00FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1334..1350
                     /label=PNR P2 rv
                     /SerialCloner_Color=&h0000FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    32..53
```

```
                     /label=PNR P1 fw
                     /SerialCloner_Color=&h0000FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2041..2070
                     /label=gfp ext
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1351..2070
                     /label=gfp
                     /SerialCloner_Color=&h00FF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
ORIGIN
        1 gGGAAGGGCG ATCGGTGCGG GCCGTTTAAA Cggcttctca cgtcgttctc atcgtctgct
       61 ttctccccct ttccttttct tccgtcgttt ctcccacaga acaattacgg gagatgctgg
      121 atttctatgt ggaggctgac gttttccggg tcagcgccga tcggacgatg attgaactct
      181 tcaacacgga ttgagagagc actgtacgca gtgcagagcg tggtaggaca tagaatcgcg
      241 cagggaaata ggttggaagg ggagcaattc gcagaaggaa agctcagcga tcgagaaaaa
      301 catggagggt gaagataagc tcgctctgaa agtaaatctg acaaatcaag cacagatgag
      361 gtcactgcga agttgacctg aaagcgtgtt tttcaaatcg tcgcccaggc gaagtgttat
      421 tatttcttgc acatcttgaa gatagtctat acgcaggatg aaaaggagga tcaactttgc
      481 catttctgca tacgtcataa ttccacgatg actagcaatg tggaatagat cttccttttg
      541 aatttctctg aatggttagc actttccttt gtttttttgt cctatgaaat ttaaatggta
      601 attgatacct ccactaagga gtcaaatcaa tagctttgat cacaagcaga agaaatgata
      661 agagcgcaaa gcttggaata atctctttat atgtcccatt gattaacatc tatcctacaa
      721 gggcgtgaac aaccacgatc agagtttagg agagcgatca actccaagca ctaaaagggc
      781 gtgaacaacc actataagaa ttgaagaggc gatcctccaa gaacgctgcc cgaccaggca
      841 ttaattgccg cataatctga ccgtcgtctg gatagcctgg cagacgggaa aacttcatgg
      901 aataatgacg ccattacttg aagtttcgcg gggctttttt tcctggctcg gaattatact
      961 ttttacgcca taactgcgtt tccccgttct ccaagggctc cgccgtcggc aactttgtat
     1021 ggggaagcgc ggtagatttt gttttcaagg aaggaccaag ggtgtgaggt ggcttttgaa
     1081 acactggtgc catgtatggg actattctga cggatctccg ccgcgcaact taactcccgc
     1141 ctcggattcc gtgttaggcg tcgtaacgaa cattaggaag acatcggatc ttagatgaga
     1201 aattatgcat catgcgactt gtgtgaatgt gcatcaatct tttgcctgtt gcattacgca
     1261 tgctcacgcg acttcatgaa tatgtcttcc cttacatact cctcctgatc acgacaacac
     1321 aacatttctc acatcactcc agggtggtcg ATGGTGAGCA AGGGCGAGGA GCTGTTCACC
     1381 ggggtggtgc ccatcctggt cgagctggac ggcgacgtaa acggccacaa gttcagcgtg
     1441 tccggcgagg gcgagggcga tgccacctac ggcaagctga ccctgaagtt catctgcacc
     1501 accggcaagc tgcccgtgcc ctggcccacc ctcgtgacca ccctgaccta cggcgtgcag
     1561 tgcttcagcc gctaccccga ccacatgaag cagcacgact tcttcaagtc cgccatgccc
     1621 gaaggctacg tccaggagcg caccatcttc ttcaaggacg acggcaacta caagacccgc
     1681 gccgaggtga agttcgaggg cgacaccctg gtgaaccgca tcgagctgaa gggcatcgac
     1741 ttcaaggagg acggcaacat cctggggcac aagctggagt acaactacaa cagccacaac
     1801 gtctatatca tggccgacaa gcagaagaac ggcatcaagg tgaacttcaa gatccgccac
     1861 aacatcgagg acggcagcgt gcagctcgcc gaccactacc agcagaacac ccccatcggc
     1921 gacggccccg tgctgctgcc cgacaaccac tacctgagca cccagtccgc cctgagcaaa
     1981 gaccccaacg agaagcgcga tcacatggtc ctgctggagt tcgtgaccgc cgccgggatc
     2041 ACTCTCGGCA TGGACGAGCT GTACAAGTGA tagcggcgct tttgaccggg aactcacgca
     2101 tttaatgtat tcgaaacacg ctttttaaga tctcaaatga aacatgtatc cgtatgctct
     2161 ccacatccgt cctcacgtac ggtgtgatgt ttttttaatt ttttatatga aagtgtaagg
     2221 cttctccatt ttacttttttg tgtttctgtc ttcacagtct ttctctgcga gctgtgagag
     2281 ggtcatgccc tttatgggca cgacggcagg attttggttt ttgaaagaat ttctgtcggc
     2341 tgtaagtctt caaaaaagaa gaaagcgggg acggaaactt cttgcttgtg tgttatacat
     2401 tacaaagtcc tctacaattc tacagtccta gttgttcacg cgacttcacg cagcagacca
     2461 gatttatagc tttgcatgat cgtctccgag catttcagaa tgctataata tggaatgatc
     2521 ttccgctggc acctcattat gcgatacgca ctccatataa acacaggtat ccctgtccGT
     2581 ATTCGTTAAC TGTTAATTCA TGACACAAca tcctgctgta tgattttggc acaacgacgc
     2641 ggttggaggg tgagggagat ggtggtgaga catcagaaat ggagagaata acggtagaga
     2701 aaaaaaggaa cacacgcgcc caggatagaa ccaggaccac attccaagtc cctcaccagc
     2761 aacaagaggg acaccacgtc ggtgaagagg tgcgtgagac agaggacacc agacaacgtg
     2821 cacaaagcgt cggaaaggga ggcgaccgct tcgctggagg ggggccgtcc acagcctgcg
     2881 cccaaccaca cacattcctc ccgcgaggct caagtgccat gccgccgagt gccatgctcc
     2941 tgcatccggg gcaggtcgac gagccagcat ggtttggtcg ccgacaacgc agatttgatg
     3001 gagcttgggc atttctttac cggacacgtc gcttcttaca tttttacatt ccaattcaga
     3061 ggtcacaccc gcgatcgaga aagcagagag tcggacccca ttcaagcagg aaaggcaagc
     3121 aacataccca tcgttaagtc tgatgatggc tggtgcaaag ccaaggcagc cagtcgatat
     3181 gttaatgatc gggtacgttt aaaggcgtgg gcaaggttat ttaacgctgg tagcatcgta
     3241 gagtttttaa cgcaactgaa gatttcatgt ctgacttgga gatacggtcg gggcagcagc
     3301 tatgggtaaa aagcctgaac tcaccgcgac gtctgtcgag aagtttctga tcgaaaagtt
     3361 cgacagcgtc tccgacctga tgcagctctc ggagggcgaa gaatctcgtg ctttcagctt
     3421 cgatgtagga gggcgtggat atgtcctgcg ggtaaatagc tgcgccgatg gtttctacaa
     3481 agatcgttat gtttatcggc actttgcatc ggccgcgctc ccgattccgg aagtgcttga
     3541 cattggggaa ttcagcgaga gcctgaccta ttgcatctcc cgccgtgcac agggtgtcac
     3601 gttgcaagac ctgcctgaaa ccgaactgcc cgctgttctg cagccggtcg cggaggccat
     3661 ggatgcgatc gctgcggccg atcttagcca gacgagcggg ttcggcccat tcggaccgca
```

```
3721 aggaatcggt caatacacta catggcgtga tttcatatgc gcgattgctg atccccatgt
3781 gtatcactgg caaactgtga tggacgacac cgtcagtgcg tccgtcgcgc aggctctcga
3841 tgagctgatg ctttgggccg aggactgccc cgaagtccgg cacctcgtgc acgcggattt
3901 cggctccaac aatgtcctga cggacaatgg ccgcataaca gcggtcattg actggagcga
3961 ggcgatgttc ggggattccc aatacgaggt cgccaacatc ttcttctgga ggccgtggtt
4021 ggcttgtatg gagcagcaga cgcgctactt cgagcggagg catccggagc ttgcaggatc
4081 gccgcggctc cgggcgtata tgctccgcat tggtcttgac caactctatc agagcttggt
4141 tgacggcaat ttcgatgatg cagcttgggc gcagggtcga tgcgacgcaa tcgtccgatc
4201 cggagccggg actgtcgggc gtacacaaat cgcccgcaga agcgcggccg tctggaccga
4261 tggctgtgta gaagtactcg ccgatagtgg aaaccgacgc cccagcactc gtccgagggc
4321 aaaggaataa tgataacagc gcttagagct cgaatttccc cgatcgttca aacatttggc
4381 aataaagttt cttaagattg aatcctgttg ccggtcttgc gatgattatc atataatttc
4441 tgttgaatta cgttaagcat gtaataatta acatgtaatg catgacgtta tttatgagat
4501 gggtttttat gattagagtc ccgcaattat acatttaata cgcgatagaa aacaaaatat
4561 agcgcgcaaa ctaggataaa ttatcgcgcg cggtgtcatc tatgttacta gatcgggaat
4621 tggttccgga accaattcgt aatcatggtc atagctgttt cctgtgtgaa attgttatcc
4681 gctcacaatt ccacacaaca tacgagccgg aagcataaag tgtaaagcct ggggtgccta
4741 atgagtgagc taactcacat taggctgaat taggcgcgcc tatttctgTT AGTGTCAAAC
4801 AGTCAAACCA GTTCTACGtg tagaacagga tgcatggctg ctgcaaggct gtttgtgcca
4861 acgccgcaaa tggaaaagaa agtaagaaaa aaaatcctac acagacaacc tttctagcac
4921 gaaaatgccc ttgcgtacat ctttttgcgt ctcagtaata ttgttttctt aatcgaaacc
4981 attattgttc aagtcatcgc taagacactt ccatcgtgtg ctcttagccg catctcggtc
5041 ctccctcgcc atggaacatg tcatttaggg cgacaaacca caaataacat agaagctttc
5101 atcgataatg ccccatcagt cctatttaaa tagtccaata caaatatata agggtttgtc
5161 atttctcgtt tcctccggct gtcccatgct gtagacatgt gtacgtagct agaaggagcc
5221 cccatatgtt ctttccgacg atacacatgg ctttaatact atccccatag tgacaaagcg
5281 tggacacaga aacacacaca catgccactc acatagaaac ggccaatcca cgatcgttac
5341 atctccgctt tcgtttccgg caacgacttg aataattggg gcatctgctc acgctcgttt
5401 tacccgaccg caaaaaaaaa tggagagtgg ccaccatgct tcaaggagaa gctcggtata
5461 tgtccatccc gcgtgcctgc aaagacaagg tgaagatgca cccggcgacc atagtgtgtc
5521 atgtacctgt ttgacgttta ctacagggtg ccatgcttta ttggctggaa ttgtggacgc
5581 ttgttttctg gaatacgaat atttccccgc cctgtattaa ttcccttcaa gcagcaagct
5641 atcttaacaa acgttgtgcg gccaaaggaa aatagccgcc gagtccatat ggggtcttct
5701 gacttcaaaa tctcgcgcat tccaacgtgt aaatgtgcga tcgaaagaga aagcattgac
5761 acagggaata aaaactccat caattggaaa gatctctgga atcggtggca ccctgttccg
5821 tgaagaagcg aagctttgct tgcggccgta tgctcattac tcacccactg actcgttcac
5881 ataataaaag tcccgctaaa accctacaat ttcgcttccg acgctccatc attccatcat
5941 agaaattgac aggtctactg aaatcatagt tccacccgtt ggtgcgtgtc cttcctgcaG
6001 TTTAAACTTG TTATCCGCTC ACAATTCCAc acaacatagg agccggaagc ataaagtgta
6061 aagcctgggg tgcctaatga gtgaggtaac tcacattaat tgcgttgcgc tcactgcccg
6121 ctttccagtc gggaaacctg tcgtgccagc tgcattaatg aatcggccaa cgcgcgggga
6181 gaggcggttt gcgtattggg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg
6241 tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag
6301 aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc
6361 gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg ccccccctgac gagcatcaca
6421 aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt
6481 ttcccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc
6541 tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc tgtaggtatc
6601 tcagttcggt gtaggtcgtt cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc
6661 ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc caacccggta agacacgact
6721 tatcgccact ggcagcagcc actggtaaca ggattagcag agcgaggtat gtaggcggtg
6781 ctacagagtt cttgaagtgg tggcctaact acggctacac tagaaggaca gtatttggta
6841 tctgcgctct gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca
6901 aacaaaccac cgctggtagc ggtggttttt ttgtttgcaa gcagcagatt acgcgcagaa
6961 aaaaaggatc tcaagaagat cctttgatct tttctacggg gtctgacgct cagtggaacg
7021 aaaactcacg ttaagggatt ttggtcatga gattatcaaa aaggatcttc acctagatcc
7081 ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa acttggtctg
7141 acagttacca atgcttaatc agtgaggcac ctatctcagc gatctgtcta tttcgttcat
7201 ccatagttgc ctgactcccc gtcgtgtaga taactacgat acgggagggc ttaccatctg
7261 gccccagtgc tgcaatgata ccgcgagacc cacgctcacc ggctccagat ttatcagcaa
7321 taaaccagcc agccggaagg gccgagcgca gaagtggtcc tgcaacttta tccgcctcca
7381 tccagtctat taattgttgc cgggaagcta gagtaagtag ttcgccagtt aatagtttgc
7441 gcaacgttgt tgccattgct acaggcatcg tggtgtcacg ctcgtcgttt ggtatggctt
7501 cattcagctc cggttcccaa cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa
7561 aagcggttag ctccttcggt cctccgatcg ttgtcagaag taagttggcc gcagtgttat
7621 cactcatggt tatggcagca ctgcataatt ctcttactgt catgccatcc gtaagatgct
7681 tttctgtgac tggtgagtac tcaaccaagt cattctgaga atagtgtatg cggcgaccga
7741 gttgctcttg cccggcgtca atacgggata ataccgcgcc acatagcaga actttaaaag
7801 tgctcatcat tggaaaacgt tcttcggggc gaaaactctc aaggatctta ccgctgttga
7861 gatccagttc gatgtaaccc actcgtgcac ccaactgatc ttcagcatct tttactttca
7921 ccagcgtttc tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg
7981 cgacacggaa atgttgaata ctcatactct tccttttttca atattattga agcatttatc
8041 agggttattg tctcatgagc ggatacatat ttgaatgtat ttagaaaaat aaacaaatag
8101 gggttccgcg cacatttccc cgaaaagtgc cacctgaacg aagcatctgt gcttcatttt
8161 gtagaacaaa aatgcaacgc gagagcgcta atttttcaaa caaagaatct gagctgcatt
8221 tttacagaac agaaatgcaa cgcgaaagcg ctattttacc aacgaagaat ctgtgcttca
8281 tttttgtaaa acaaaaatgc aacgcgagag cgctaatttt tcaaacaaag aatctgagct
8341 gcatttttac agaacagaaa tgcaacgcga gagcgctatt ttaccaacaa agaatctata
8401 cttctttttt gttctacaaa aatgcatccc gagagcgcta tttttctaac aaagcatctt
8461 agattacttt ttttctcctt tgtgcgctct ataatgcagt ctcttgataa ctttttgcac
```

```
     8521 tgtaggtccg ttaaggttag aagaaggcta ctttggtgtc tattttctct tccataaaaa
     8581 aagcctgact ccacttcccg cgtttactga ttactagcga agctgcgggt gcatttttc
     8641 aagataaagg catccccgat tatattctat accgatgtgg attgcgcata ctttgtgaac
     8701 agaaagtgat agcgttgatg attcttcatt ggtcagaaaa ttatgaacgg tttcttctat
     8761 tttgtctcta tatactacgt ataggaaatg tttacatttt cgtattgttt tcgattcact
     8821 ctatgaatag ttcttactac aatttttttg tctaaagagt aatactagag ataaacataa
     8881 aaaatgtaga ggtcgagttt agatgcaagt tcaaggagcg aaaggtggat gggtaggtta
     8941 tatagggata tagcacagag atatatagca aagagatact tttgagcaat gtttgtggaa
     9001 gcggtattcg caatatttta gtagctcgtt acagtccggt gcgtttttgg ttttttgaaa
     9061 gtgcgtcttc agagcgcttt tggttttcaa aagcgctctg aagttcctat actttctaga
     9121 gaataggaac ttcggaatag gaacttcaaa gcgtttccga aaacgagcgc ttccgaaaat
     9181 gcaacgcgag ctgcgcacat acagctcact gttcacgtcg cacctatatc tgcgtgttgc
     9241 ctgtatatat atatacatga gaagaacggc atagtgcgtg tttatgctta aatgcgtact
     9301 tatatgcgtc tatttatgta ggatgaaagg tagtctagta cctcctgtga tattatccca
     9361 ttccatgcgg ggtatcgtat gcttccttca gcactaccct ttagctgttc tatatgctgc
     9421 cactcctcaa ttggattagt ctcatccttc aatgctatca tttcctttga tattggatca
     9481 tactaagaaa ccattattat catgacatta acctataaaa ataggcgtat cacgaggccc
     9541 tttcgtctcg cgcgtttcgg tgatgacggt gaaaacctct gacacatgca gctcccggag
     9601 acggtcacag cttgtctgta agcggatgcc gggagcagac aagcccgtca gggcgcgtca
     9661 gcgggtgttg gcgggtgtcg gggctggctt aactatgcgg catcagagca gattgtactg
     9721 agagtgcacc ataccacagc ttttcaattc aattcatcat ttttttttta ttctttttt
     9781 tgatttcggt ttctttgaaa tttttttgat tcggtaatct ccgaacagaa ggaagaacga
     9841 aggaaggagc acagacttag attggtatat atacgcatat gtagtgttga agaaacatga
     9901 aattgcccag tattcttaac ccaactgcac agaacaaaaa cctgcaggaa acgaagataa
     9961 atcatgtcga aagctacata taaggaacgt gctgctactc atcctagtcc tgttgctgcc
    10021 aagctattta atatcatgca cgaaaagcaa acaaacttgt gtgcttcatt ggatgttcgt
    10081 accaccaagg aattactgga gttagttgaa gcattaggtc ccaaaatttg tttactaaaa
    10141 acacatgtgg atatcttgac tgatttttcc atggagggca cagttaagcc gctaaaggca
    10201 ttatccgcca agtacaattt tttactcttc gaagacagaa aatttgctga cattggtaat
    10261 acagtcaaat tgcagtactc tgcgggtgta tacagaatag cagaatgggc agacattacg
    10321 aatgcacacg gtgtggtggg cccaggtatt gttagcggtt tgaagcaggc ggcagaagaa
    10381 gtaacaaagg aacctagagg ccttttgatg ttagcagaat tgtcatgcaa gggctcccta
    10441 tctactggag aatatactaa gggtactgtt gacattgcga agagcgacaa agattttgtt
    10501 atcggcttta ttgctcaaag agacatgggt ggaagagatg aaggttacga ttggttgatt
    10561 atgacacccg gtgtgggttt agatgacaag ggagacgcat tgggtcaaca gtatagaacc
    10621 gtggatgatg tggtctctac aggatctgac attattattg ttggaagagg actatttgca
    10681 aagggaaggg atgctaaggt agagggtgaa cgttacagaa aagcaggctg ggaagcatat
    10741 ttgagaagat gcggccagca aaactaaaaa actgtattat aagtaaatgc atgtatacta
    10801 aactcacaaa ttagagcttc aatttaatta tatcagttat taccctatgc ggtgtgaaat
    10861 accgcacaga tgcgtaagga gaaaataccg catcaggaaa ttgtaaacgt taatattttg
    10921 ttaaaattcg cgttaaattt ttgttaaatc agctcatttt ttaaccaata ggccgaaatc
    10981 ggcaaaatcc cttataaatc aaaagaatag accgagatag ggttgagtgt tgttccagtt
    11041 tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaagggcg aaaaaccgtc
    11101 tatcagggcg atggcccact acgtgaacca tcaccctaat caagtttttt ggggtcgagg
    11161 tgccgtaaag cactaaatcg gaaccctaaa gggagccccc gatttagagc ttgacgggga
    11221 aagccggcga acgtggcgag aaaggaaggg aagaaagcga aaggagcggg cgctagggcg
    11281 ctggcaagtg tagcggtcac gctgcgcgta accaccacac ccgccgcgct taatgcgccg
    11341 ctacagggcg cgtcgcgcca ttcgccattc aggctgcgca actgtt
//
```

A.4 pLDH – Genbank formatted vector sequence

```
LOCUS       pLDH    11067 bp      DNA     circular   15/11/2017
DEFINITION
ACCESSION
VERSION
SOURCE
  ORGANISM
COMMENT     Serial Cloner Genbank Format
COMMENT     SerialCloner_Type=DNA
COMMENT     SerialCloner_Comments=Homologous recombination reaction between [pLCP hph]
as a ta
            rget sequence  and [PLDH] as an inserted sequence.
COMMENT     SerialCloner_Ends=0,0,,0,
FEATURES             Location/Qualifiers
     misc_feature    complement(9257..9279)
                     /label=pGEX_3_primer
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pGEX_3_primer" - complement(29..51) \
     promoter        9430..9643
                     /label=URA3_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3_promoter" - 202..415 \
     gene            9645..10445
                     /label=URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - 417..1217 \
     CDS             9645..10448
                     /label=ORF frame 3_URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - note="ORF frame 3"
- translation="MSKATYKERAATHPSPVAAKLFNIMHEKQTNLCASLDVRTTKEL -
LELVEALGPKICLLKTHVDILTDFSMEGTVKPLKALSAKYNFLLFEDRKFADIGNTVK -
LQYSAGVYRIAEWADITNAHGVVGPGIVSGLKQAAEEVTKEPRGLLMLAELSCKGSLS - TGEYTKGTVDIAKSDKDFVIGFI \
     rep_origin      complement(10711..11017)
                     /label=f1_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="f1_origin" - complement(1483..1789) \
     misc_feature    complement(11050..24)
                     /label=lacZ_a [5']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="lacZ_a" - complement(1822..1965) \
     misc_feature    complement(5687..24)
                     /label=M13_pUC_rev_primer [3'] [5'] [3']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="M13_pUC_rev_primer" -
complement(2176..2198) \
     rep_origin      complement(6052..6671)
                     /label=pBR322_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
```

```
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pBR322_origin" - complement(2550..3169) \
     CDS             complement(6826..7686)
                     /label=ORF frame 3_Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
note="ORF frame 3" - translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY -
IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE -
YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL - DRWEPELNEAI \
     gene            complement(6826..7686)
                     /label=Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
complement(3324..4184) \
     promoter        complement(7728..7756)
                     /label=AmpR_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="AmpR_promoter" - complement(4226..4254) \
     rep_origin      complement(7818..8789)
                     /label=2micron2_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="2micron2_origin" - complement(4316..5287)
\
     misc_feature    complement(8781..8828)
                     /label=FRT
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="FRT" - complement(5279..5326) \
     misc_feature    5681..5710
                     /label=YY
                     /SerialCloner_Color=&h8080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4470..4499
                     /label=DD
                     /SerialCloner_Color=&h800040
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4012..4469
                     /label=NOS terminator_regulatory
                     /SerialCloner_Color=&hFF0000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=regulatory_class="terminator" - note="NOS
terminator" - 2911..3320 \
     misc_feature    2290..2982
                     /label=Puep
                     /SerialCloner_Color=&hFF80C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4500..4521
                     /label=P3
                     /SerialCloner_Color=&hFFFF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    5661..5680
                     /label=P4
                     /SerialCloner_Color=&hFFFF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
```

```
                     /SerialCloner_Arrow=True
     gene            2983..4011
                     /label=hph_gene
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=gene="hph" - 485..1513 \
     CDS             2983..4011
                     /label=hph_CDS
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=gene="hph" - codon_start=1 - transl_table=11 -
product="hygromycin B phosphotransferase" - protein_id="CAR31381.1" -
db_xref="GI:194473371" - translation="MGKKPELTATSVEKFLIEKFDSVSDLMQLSEGEESRAFSFDVGG -
RGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQ \
     misc_feature    2983..3007
                     /label=hph fw
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    3986..4011
                     /label=hph rv
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2..31
                     /label=XX ext
                     /SerialCloner_Color=&h00FFFF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1722..1751
                     /label=gfp ext
                     /SerialCloner_Color=&h996633
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1752..1768
                     /label=Tatpa P1 fw
                     /SerialCloner_Color=&hFF00FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2260..2289
                     /label=CC ext
                     /SerialCloner_Color=&h996633
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2240..2259
                     /label=Tatpa P2 rv
                     /SerialCloner_Color=&hFF00FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1032..1751
                     /label=gfp
                     /SerialCloner_Color=&h00FF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    32..1031
                     /label=PLDH
                     /SerialCloner_Color=&h0000FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1032..1061
                     /label=GFP ext
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
ORIGIN
        1 gGGAAGGGCG ATCGGTGCGG GCCGTTTAAA Caatagttga ttgcaacgac tctgTTTCTG
```

```
  61 AGTCAAACCC CCACAGGAAG TCAGCAGACG TGTCGATAGC GTCGGACGGC GCGGCCAGGC
 121 GACTTATCAA CGCTAAAGCG CAGGGACCAG AACCAATTAC CGCAATATCA AGACGATTCA
 181 TTGGCAAATT TTTCTAGTGG AATATAGGAA ACCCAGGCAC TTATGACGCC GACCAACAAC
 241 TGTTGCGTGA AAGCACGCGG CATGAGAAAC AAAGTCATCG ACAAATCTCC ACTTCACATG
 301 GTGGAGAGAA TAAACATCCC GCACGACTTC GCTTCACTGT TCACGTTATT ATAATTTCAT
 361 TCATATAAAT ATAGCGAGAA ATAATTAATA TTTCACGTAA AACCCTTCCA CATATTCTCG
 421 TTAATTTGCA CTCGTTCTTT TTGCGTCGTC GGATGAAAGA GCCGTCTGCG GATTATTCCA
 481 CGGGAGGTGA TGCCACCGTT GTTGTACGTA TCCACCACAA GGCAATTTCT ATGGGGCTAA
 541 ACACCCCCAA AAAACTCATT CATGCCAAAA AAATACCCAA CAATGTAACC TGCCCCCAGA
 601 ATTACCTCTT CAACCACGAA CCACCGGGAA AAAGAAACTT TGTCCACGAG GGGCAGCCCT
 661 TGATTTGAAA CTTCAACCGT CGAGCTCCTT GCAACGGCTC AAGCGTGCGT CACATTCTGA
 721 GAAACAATGG GGAGTTGAAA CCCGAGGCAC GCACACAATC TACGCGGTTC ATACCCGTGT
 781 ATGTTCCATT CTTGTTCACG GCTTGACCTT AACCGCCAAA CAAAGACCAC GGAGAGAACA
 841 TAATCTGCTA ATAGGTCTTT TTTATGTGTA CCATGAAGGC CGGCATTCGG CATTGTTCTC
 901 TCAACTGCGA GCACTTCGGT CGAAATAAAA GAACACTACC TCGTTCCCGG CCTGGCTTGG
 961 TTTGCCTGCA ATGAAGTCGA TTCACAATAT TTTTAAGTAC TCATCGTCaa tgattatgac
1021 agcgtaaaag cATGGTGAGC AAGGGCGAGG AGCTGTTCAC Cggggtggtg cccatcctgg
1081 tcgagctgga cggcgacgta aacggccaca agttcagcgt gtccggcgag ggcgagggcg
1141 atgccaccta cggcaagctg accctgaagt tcatctgcac caccggcaag ctgcccgtgc
1201 cctggcccac cctcgtgacc accctgacct acggcgtgca gtgcttcagc cgctaccccg
1261 accacatgaa gcagcacgac ttcttcaagt ccgccatgcc cgaaggctac gtccaggagc
1321 gcaccatctt cttcaaggac gacggcaact acaagacccg cgccgaggtg aagttcgagg
1381 gcgacaccct ggtgaaccgc atcgagctga agggcatcga cttcaaggag gacggcaaca
1441 tcctggggca caagctggag tacaactaca acagccacaa cgtctatatc atggccgaca
1501 agcagaagaa cggcatcaag gtgaacttca agatccgcca caacatcgag gacggcagcg
1561 tgcagctcgc cgaccactac cagcagaaca cccccatcgg cgacggcccc gtgctgctgc
1621 ccgacaacca ctacctgagc acccagtccg ccctgagcaa agaccccaac gagaagcgcg
1681 atcacatggt cctgctggag ttcgtgaccg ccgccgggat cACTCTCGGC ATGGACGAGC
1741 TGTACAAGTG Atagcggcgc ttttgaccgg gaactcacgc atttaatgta ttcgaaacac
1801 gctttttaag atctcaaatg aaacatgtat ccgtatgctc tccacatccg tcctcacgta
1861 cggtgtgatg tttttttaat tttttatatg aaagtgtaag gcttctccat tttacttttt
1921 gtgtttctgt cttcacagtc tttctctgcg agctgtgaga gggtcatgcc ctttatgggc
1981 acgacggcag gattttggtt tttgaaagaa tttctgtcgg ctgtaagtct tcaaaaaaga
2041 agaaagcggg gacggaaact tcttgcttgt gtgttataca ttacaaagtc ctctacaatt
2101 ctacagtcct agttgttcac gcgacttcac gcagcagacc agatttatag ctttgcatga
2161 tcgtctccga gcatttcaga atgctataat atggaatgat cttccgctgg cacctcatta
2221 tgcgatacgc actccatata aacacaggta tccctgtccG TATTCGTTAA CTGTTAATTC
2281 ATGACACAAc atcctgctgt atgattttgg cacaacgacg cggttggagg gtgagggaga
2341 tggtggtgag acatcagaaa tggagagaat aacggtagag aaaaaaagga acacacgcgc
2401 ccaggataga accaggacca cattccaagt ccctcaccag caacaagagg gacaccacgt
2461 cggtgaagag gtgcgtgaga cagaggacac cagacaacgt gcacaaagcg tcggaaaggg
2521 aggcgaccgc ttcgctggag gggggccgtc cacagcctgc gcccaaccac acacattcct
2581 cccgcgaggc tcaagtgcca tgccgccgag tgccatgctc ctgcatccgg ggcaggtcga
2641 cgagccagca tggtttggtc gccgacaacg cagatttgat ggagcttggg catttcttta
2701 ccggacacgt cgcttcttac atttttacat tccaattcag aggtcacacc cgcgatcgag
2761 aaagcagaga gtcggacccc cttcaagcag gaaaggcaag caacataccc atcgttaagt
2821 ctgatgatgg ctggtgcaaa gccaaggcag ccagtcgata tgttaatgat cgggtacgtt
2881 taaaggcgtg ggcaaggtta tttaacgctg gtagcatcgt agagttttta acgcaactga
2941 agatttcatg tctgacttgg agatacggtc ggggcagcag ctatgggtaa aaagcctgaa
3001 ctcaccgcga cgtctgtcga gaagtttctg atcgaaaagt tcgacagcgt ctccgacctg
3061 atgcagctct cggagggcga agaatctcgt gctttcagct tcgatgtagg agggcgtgga
3121 tatgtcctgc gggtaaatag ctgcgccgat ggtttctaca aagatcgtta tgtttatcgg
3181 cactttgcat cggccgcgct cccgattccg gaagtgcttg acattgggga attcagcgag
3241 agcctgacct attgcatctc ccgccgtgca cagggtgtca cgttgcaaga cctgcctgaa
3301 accgaactgc ccgctgttct gcagccggtc gcggaggcca tggatgcgat cgctgcggcc
3361 gatcttagcc agacgagcgg gttcggccca ttcggaccgc aaggaatcgg tcaatacact
3421 acatggcgtg atttcatatg cgcgattgct gatccccatg tgtatcactg gcaaactgtg
3481 atggacgaca ccgtcagtgc gtccgtcgcg caggctctcg atgagctgat gctttgggcc
3541 gaggactgcc ccgaagtccg gcacctcgtg cacgcggatt tcggctccaa caatgtcctg
3601 acggacaatg gccgcataac agcggtcatt gactggagcg aggcgatgtt cggggattcc
3661 caatacgagg tcgccaacat cttcttctgg aggccgtggt tggcttgtat ggagcagcag
3721 acgcgctact tcgagcggag gcatccggag cttgcaggat cgccgcggct ccgggcgtat
3781 atgctccgca ttggtcttga ccaactctat cagagcttgg ttgacggcaa tttcgatgat
3841 gcagcttggg cgcagggtcg atgcgacgca atcgtccgat ccggagccgg gactgtcggg
3901 cgtacacaaa tcgcccgcag aagcgcggcc gtctggaccg atggctgtgt agaagtactc
3961 gccgatagtg gaaaccgacg ccccagcact cgtccgaggg caaaggaata atgataacag
4021 cgcttagagc tcgaatttcc ccgatcgttc aaacatttgg caataaagtt tcttaagatt
4081 gaatcctgtt gccggtcttg cgatgattat catataattt ctgttgaatt acgttaagca
4141 tgtaataatt aacatgtaat gcatgacgtt atttatgaga tgggttttta tgattagagt
4201 cccgcaatta tacatttaat acgcgataga aaacaaaata tagcgcgcaa actaggataa
4261 attatcgcgc gcggtgtcat ctatgttact agatcgggaa ttggttccgg aaccaattcg
4321 taatcatggt catagctgtt tcctgtgtga aattgttatc cgctcacaat tccacacaac
4381 atacgagccg gaagcataaa gtgtaaagcc tggggtgcct aatgagtgag ctaactcaca
4441 ttaggctgaa ttaggcgcgc ctatttctgT TAGTGTCAAA CAGTCAAACC AGTTCTACGt
4501 gtagaacagg atgcatggct gctgcaaggc tgtttgtgcc aacgccgcaa atggaaaaga
4561 aagtaagaaa aaaaatccta cacagacaac ctttctagca cgaaaatgcc cttgcgtaca
4621 tctttttgcg tctcagtaat attgttttct taatcgaaac cattattgtt caagtcatcg
4681 ctaagacact tccatcgtgt gctcttagcc gcatctcggt cctccctcgc catggaacat
4741 gtcatttagg gcgacaaacc acaaataaca tagaagcttt catcgataat gccccatcag
4801 tcctatttaa atagtccaat acaaatatat aagggtttgt catttctcgt ttcctccggc
```

```
4861 tgtcccatgc tgtagacatg tgtacgtagc tagaaggagc ccccatatgt tctttccgac
4921 gatacacatg gctttaatac tatccccata gtgacaaagc gtggacacag aaacacacac
4981 acatgccact cacatagaaa cggccaatcc acgatcgtta catctccgct ttcgtttccg
5041 gcaacgactt gaataattgg ggcatctgct cacgctcgtt ttacccgacc gcaaaaaaaa
5101 atggagagtg gccaccatgc ttcaaggaga agctcggtat atgtccatcc cgcgtgcctg
5161 caaagacaag gtgaagatgc acccggcgac catagtgtgt catgtacctg tttgacgttt
5221 actacagggt gccatgcttt attggctgga attgtggacg cttgttttct ggaatacgaa
5281 tatttccccg ccctgtatta attcccttca agcagcaagc tatcttaaca aacgttgtgc
5341 ggccaaagga aaatagccgc cgagtccata tggggtcttc tgacttcaaa atctcgcgca
5401 ttccaacgtg taaatgtgcg atcgaaagag aaagcattga cacagggaat aaaaactcca
5461 tcaattggaa agatctctgg aatcggtggc accctgttcc gtgaagaagc gaagctttgc
5521 ttgcggccgt atgctcatta ctcacccact gactcgttca cataataaaa gtcccgctaa
5581 aaccctacaa tttcgcttcc gacgctccat cattccatca tagaaattga caggtctact
5641 gaaatcatag ttccacccgt tggtgcgtgt ccttcctgca GTTTAAACTT GTTATCCGCT
5701 CACAATTCCA cacaacatag gagccggaag cataaagtgt aaagcctggg gtgcctaatg
5761 agtgaggtaa ctcacattaa ttgcgttgcg ctcactgccc gctttccagt cgggaaacct
5821 gtcgtgccag ctgcattaat gaatcggcca acgcgcgggg agaggcggtt tgcgtattgg
5881 gcgctcttcc gcttcctcgc tcactgactc gctgcgctcg gtcgttcggc tgcggcgagc
5941 ggtatcagct cactcaaagg cggtaatacg gttatccaca gaatcagggg ataacgcagg
6001 aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct
6061 ggcgtttttc cataggctcc gccccccctga cgagcatcac aaaaatcgac gctcaagtca
6121 gaggtggcga aacccgacag gactataaag ataccaggcg tttcccccctg gaagctccct
6181 cgtgcgctct cctgttccga ccctgccgct taccggatac ctgtccgcct ttctcccttc
6241 gggaagcgtg gcgctttctc atagctcacg ctgtaggtat ctcagttcgg tgtaggtcgt
6301 tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct gcgccttatc
6361 cggtaactat cgtcttgagt ccaacccggt aagacacgac ttatcgccac tggcagcagc
6421 cactggtaac aggattagca gagcgaggta tgtaggcggt gctacagagt tcttgaagtg
6481 gtggcctaac tacggctaca ctagaaggac agtatttggt atctgcgctc tgctgaagcc
6541 agttaccttc ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag
6601 cggtggtttt tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga
6661 tcctttgatc ttttctacgg ggtctgacgc tcagtggaac gaaaactcac gttaagggat
6721 tttggtcatg agattatcaa aaaggatctt cacctagatc cttttaaatt aaaaatgaag
6781 ttttaaatca atctaaagta tatatgagta aacttggtct gacagttacc aatgcttaat
6841 cagtgaggca cctatctcag cgatctgtct atttcgttca tccatagttg cctgactccc
6901 cgtcgtgtag ataactacga tacgggaggg cttaccatct ggccccagtg ctgcaatgat
6961 accgcgagac ccacgctcac cggctccaga tttatcagca ataaaccagc cagccggaag
7021 ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc atccagtcta ttaattgttg
7081 ccgggaagct agagtaagta gttcgccagt taatagtttg cgcaacgttg ttgccattgc
7141 tacaggcatc gtggtgtcac gctcgtcgtt tggtatggct tcattcagct ccggttccca
7201 acgatcaagg cgagttacat gatcccccat gttgtgcaaa aaagcggtta gctccttcgg
7261 tcctccgatc gttgtcagaa gtaagttggc cgcagtgtta tcactcatgg ttatggcagc
7321 actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga ctggtgagta
7381 ctcaaccaag tcattctgag aatagtgtat gcggcgaccg agttgctctt gcccggcgtc
7441 aatacgggat aataccgcgc cacatagcag aactttaaaa gtgctcatca ttggaaaacg
7501 ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt cgatgtaacc
7561 cactcgtgca cccaactgat cttcagcatc ttttactttc accagcgttt ctgggtgagc
7621 aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg gcgacacgga aatgttgaat
7681 actcatactc ttccttttc aatattattg aagcatttat cagggttatt gtctcatgag
7741 cggatacata tttgaatgta tttagaaaaa taaacaaata ggggttccgc gcacatttcc
7801 ccgaaaagtg ccacctgaac gaagcatctg tgcttcattt tgtagaacaa aaatgcaacg
7861 cgagagcgct aatttttcaa acaaagaatc tgagctgcat ttttacagaa cagaaatgca
7921 acgcgaaagc gctattttac caacgaagaa tctgtgcttc atttttgtaa aacaaaaatg
7981 caacgcgaga gcgctaattt ttcaaacaaa gaatctgagc tgcatttta cagaacagaa
8041 atgcaacgcg agagcgctat tttaccaaca aagaatctat acttcttttt tgttctacaa
8101 aaatgcatcc cgagagcgct atttttctaa caaagcatct tagattactt tttttctcct
8161 ttgtgcgctc tataatgcag tctcttgata actttttgca ctgtaggtcc gttaaggtta
8221 gaagaaggct actttggtgt ctattttctc ttccataaaa aaagcctgac tccacttccc
8281 gcgtttactg attactagcg aagctgcggg tgcatttttt caagataaag gcatccccga
8341 ttatattcta taccgatgtg gattgcgcat actttgtgaa cagaaagtga tagcgttgat
8401 gattcttcat tggtcagaaa attatgaacg gtttcttcta ttttgtctct atatactacg
8461 tataggaaat gtttacattt tcgtattgtt ttcgattcac tctatgaata gttcttacta
8521 caatttttt gtctaaagag taatactaga gataaacata aaaaatgtag aggtcgagtt
8581 tagatgcaag ttcaaggagc gaaaggtgga tgggtaggtt atatagggat atagcacaga
8641 gatatatagc aaagagatac ttttgagcaa tgtttgtgga agcggtattc gcaatatttt
8701 agtagctcgt tacagtccgg tgcgtttttg gttttttgaa agtgcgtctt cagagcgctt
8761 ttggttttca aaagcgctct gaagttccta tactttctag agaataggaa cttcggaata
8821 ggaacttcaa agcgtttccg aaaacgagcg cttccgaaaa tgcaacgcga gctgcgcaca
8881 tacagctcac tgttcacgtc gcacctatat ctgcgtgttg cctgtatata tatatacatg
8941 agaagaacgg catagtgcgt gtttatgctt aaatgcgtac ttatatgcgt ctatttatgt
9001 aggatgaaag gtagtctagt acctcctgtg atattatccc attccatgcg gggtatcgta
9061 tgcttccttc agcactaccc tttagctgtt ctatatgctg ccactcctca attggattag
9121 tctcatcctt caatgctatc atttcctttg atattggatc atactaagaa accattatta
9181 tcatgacatt aacctataaa aataggcgta tcacgaggcc ctttcgtctc gcgcgtttcg
9241 gtgatgacgg tgaaaacctc tgacacatgc agctcccgga gacggtcaca gcttgtctgt
9301 aagcggatgc cgggagcaga caagcccgtc agggcgcgtc agcgggtgtt ggcgggtgtc
9361 ggggctggct taactatgcg gcatcagagc agattgtact gagagtgcac cataccacag
9421 cttttcaatt caattcatca ttttttttttt attcttttt ttgatttcgg tttctttgaa
9481 attttttga ttcggtaatc tccgaacaga aggaagaacg aaggaaggag cacagactta
9541 gattggtata tatacgcata tgtagtgttg aagaaacatg aaattgccca gtattcttaa
9601 cccaactgca cagaacaaaa acctgcagga aacgaagata aatcatgtcg aaagctacat
```

```
     9661 ataaggaacg tgctgctact catcctagtc ctgttgctgc caagctattt aatatcatgc
     9721 acgaaaagca aacaaacttg tgtgcttcat tggatgttcg taccaccaag gaattactgg
     9781 agttagttga agcattaggt cccaaaattt gtttactaaa aacacatgtg gatatcttga
     9841 ctgatttttc catggagggc acagttaagc cgctaaaggc attatccgcc aagtacaatt
     9901 ttttactctt cgaagacaga aaatttgctg acattggtaa tacagtcaaa ttgcagtact
     9961 ctgcgggtgt atacagaata gcagaatggg cagacattac gaatgcacac ggtgtggtgg
    10021 gcccaggtat tgttagcggt ttgaagcagg cggcagaaga agtaacaaag gaacctagag
    10081 gccttttgat gttagcagaa ttgtcatgca agggctccct atctactgga gaatatacta
    10141 agggtactgt tgacattgcg aagagcgaca aagattttgt tatcggcttt attgctcaaa
    10201 gagacatggg tggaagagat gaaggttacg attggttgat tatgacaccc ggtgtgggtt
    10261 tagatgacaa gggagacgca ttgggtcaac agtatagaac cgtggatgat gtggtctcta
    10321 caggatctga cattattatt gttggaagag gactatttgc aaagggaagg gatgctaagg
    10381 tagagggtga acgttacaga aaagcaggct gggaagcata tttgagaaga tgcggccagc
    10441 aaaactaaaa aactgtatta taagtaaatg catgtatact aaactcacaa attagagctt
    10501 caatttaatt atatcagtta ttaccctatg cggtgtgaaa taccgcacag atgcgtaagg
    10561 agaaaatacc gcatcaggaa attgtaaacg ttaatatttt gttaaaattc gcgttaaatt
    10621 tttgttaaat cagctcattt tttaaccaat aggccgaaat cggcaaaatc ccttataaat
    10681 caaaagaata gaccgagata gggttgagtg ttgttccagt ttggaacaag agtccactat
    10741 taaagaacgt ggactccaac gtcaaagggc gaaaaaccgt ctatcagggc gatggcccac
    10801 tacgtgaacc atcaccctaa tcaagtttt tggggtcgag gtgccgtaaa gcactaaatc
    10861 ggaaccctaa agggagcccc cgatttagag cttgacgggg aaagccggcg aacgtggcga
    10921 gaaaggaagg gaagaaagcg aaaggagcgg gcgctagggc gctggcaagt gtagcggtca
    10981 cgctgcgcgt aaccaccaca cccgccgcgc ttaatgcgcc gctacagggc gcgtcgcgcc
    11041 attcgccatt caggctgcgc aactgtt

//
```

A.5 pHYP1 – Genbank formatted vector sequence

```
LOCUS       pHYP1    11067 bp      DNA     circular   15/11/2017
DEFINITION
ACCESSION
VERSION
SOURCE
  ORGANISM
COMMENT     Serial Cloner Genbank Format
COMMENT     SerialCloner_Type=DNA
COMMENT     SerialCloner_Comments=Homologous recombination reaction between [pLCP hph]
as a ta
            rget sequence  and [PHYP1] as an inserted sequence.
COMMENT     SerialCloner_Ends=0,0,,0,
FEATURES             Location/Qualifiers
     misc_feature    complement(9257..9279)
                     /label=pGEX_3_primer
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pGEX_3_primer" - complement(29..51) \
     promoter        9430..9643
                     /label=URA3_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3_promoter" - 202..415 \
     gene            9645..10445
                     /label=URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - 417..1217 \
     CDS             9645..10448
                     /label=ORF frame 3_URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - note="ORF frame 3"
- translation="MSKATYKERAATHPSPVAAKLFNIMHEKQTNLCASLDVRTTKEL -
LELVEALGPKICLLKTHVDILTDFSMEGTVKPLKALSAKYNFLLFEDRKFADIGNTVK -
LQYSAGVYRIAEWADITNAHGVVGPGIVSGLKQAAEEVTKEPRGLLMLAELSCKGSLS - TGEYTKGTVDIAKSDKDFVIGFI \
     rep_origin      complement(10711..11017)
                     /label=f1_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="f1_origin" - complement(1483..1789) \
     misc_feature    complement(11050..24)
                     /label=lacZ_a [5']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="lacZ_a" - complement(1822..1965) \
     misc_feature    complement(5687..24)
                     /label=M13_pUC_rev_primer [3'] [5'] [3']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="M13_pUC_rev_primer" -
complement(2176..2198) \
     rep_origin      complement(6052..6671)
                     /label=pBR322_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
```

```
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pBR322_origin" - complement(2550..3169) \
     CDS             complement(6826..7686)
                     /label=ORF frame 3_Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
note="ORF frame 3" - translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY -
IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE -
YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL - DRWEPELNEAI \
     gene            complement(6826..7686)
                     /label=Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
complement(3324..4184) \
     promoter        complement(7728..7756)
                     /label=AmpR_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="AmpR_promoter" - complement(4226..4254) \
     rep_origin      complement(7818..8789)
                     /label=2micron2_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="2micron2_origin" - complement(4316..5287)
\
     misc_feature    complement(8781..8828)
                     /label=FRT
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="FRT" - complement(5279..5326) \
     misc_feature    2..31
                     /label=XX
                     /SerialCloner_Color=&h008000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    5681..5710
                     /label=YY
                     /SerialCloner_Color=&h8080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4470..4499
                     /label=DD
                     /SerialCloner_Color=&h800040
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4012..4469
                     /label=NOS terminator_regulatory
                     /SerialCloner_Color=&hFF0000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=regulatory_class="terminator" - note="NOS
terminator" - 2911..3320 \
     misc_feature    2290..2982
                     /label=Puep
                     /SerialCloner_Color=&hFF80C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4500..4521
                     /label=P3
                     /SerialCloner_Color=&hFFFF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
```

```
                     /SerialCloner_Arrow=True
     misc_feature    5661..5680
                     /label=P4
                     /SerialCloner_Color=&hFFFF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     gene            2983..4011
                     /label=hph_gene
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=gene="hph" - 485..1513 \
     CDS             2983..4011
                     /label=hph_CDS
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=gene="hph" - codon_start=1 - transl_table=11 -
product="hygromycin B phosphotransferase" - protein_id="CAR31381.1" -
db_xref="GI:194473371" - translation="MGKKPELTATSVEKFLIEKFDSVSDLMQLSEGEESRAFSFDVGG -
RGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQ \
     misc_feature    2983..3007
                     /label=hph fw
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    3986..4011
                     /label=hph rv
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1722..1751
                     /label=gfp ext
                     /SerialCloner_Color=&h996633
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1752..1768
                     /label=Tatpa P1 fw
                     /SerialCloner_Color=&hFF00FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2260..2289
                     /label=CC ext
                     /SerialCloner_Color=&h996633
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2240..2259
                     /label=Tatpa P2 rv
                     /SerialCloner_Color=&hFF00FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1032..1751
                     /label=gfp
                     /SerialCloner_Color=&h00FF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    32..1031
                     /label=PHYP1
                     /SerialCloner_Color=&h0000FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1032..1061
                     /label=gfp ext
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
ORIGIN
        1 gGGAAGGGCG ATCGGTGCGG GCCGTTTAAA Ctcttgctca agaaagtcca aattCTGGTA
```

```
  61 TAGCCGCACG TACTCGTCAT ACTGCCGTGC AAGCTCTTTT TCCAACATGT TGTGTTCTTC
 121 TACCAAAGAT GGCCTGAGAA CATATGATGT CGACATGTCT CAGTAAGAAA GCGTTCGCCG
 181 GGTTAGCCGT GTGTAATTAT ATTTTCGGCT TCATATTCGT GGGAAAGTAC CCCTGATTGC
 241 GGCTCATGCT GCCATCTCAT GCCAATTTCT TGATCCAATT GAAACACGTA CCGTACACTC
 301 TGTAAACTTT GAAGACGGGC CTCGCTGCGT TTGAGTTCGC CTTTCTTTTC CTCGACCTGT
 361 GTACGGCAAA GCTAAAAGTT AACATATAAT GCCATCATGA AAACCACGAG ATGCTATGTG
 421 TACTTACCTT CAGCTCCAGC ATTCGCTCCT CTGCATCCAG TGCAGCATGC TGCCGTTGCA
 481 ACGATTCAAG TGTGCGGCGG AGAGATGGCA CGTCTTCTGT ACTTGTCAAA GTGGTCATGA
 541 CACCAAGAGT CCTGCGTGGA GAGGGGGGAT GACGGCCTCG GATCAAAAAA GAAATAAGGC
 601 GAGCAACGAA GCCATCCCTC GGGCATACTC TTCCATGACT GTAACAAAAT AGTAGTTGGA
 661 ACACTGTTTT TCGTAGAACA TGCAATCTGG CTACGTCCCA AAGCCAAAAG GGGCAGCAAA
 721 GGCAAATTTT CGGCTTTGGT TCTTCCAATT ATTCTTGTTC TCCCAAGGCC ACCCCTACCA
 781 TTCCTTGCCC CAGCCCGAAC TTCAACTCGG AGCCTGACAC CAGCATTCCA CACCAGACTC
 841 GACCGCATTA CCGGCGTGGA AACGCGGGGA ATCCCCACTC GAAGATCTCT TGTTGGTCTT
 901 TGTCCCAGCT TTTCCTTGAG ACGATTTCTC ACGTATTTCA AATGCATTTG ATACCTCAAT
 961 TCTTTGACGG CACATAGAGT TCTAGTGTTC AACACCCTCA TGCCGACTCc ctatctcgct
1021 acatagaggt cATGGTGAGC AAGGGCGAGG AGCTGTTCAC Cggggtggtg cccatcctgg
1081 tcgagctgga cggcgacgta aacggccaca agttcagcgt gtccggcgag ggcgagggcg
1141 atgccaccta cggcaagctg accctgaagt tcatctgcac caccggcaag ctgcccgtgc
1201 cctggcccac cctcgtgacc accctgacct acggcgtgca gtgcttcagc cgctacccg
1261 accacatgaa gcagcacgac ttcttcaagt ccgccatgcc cgaaggctac gtccaggagc
1321 gcaccatctt cttcaaggac gacggcaact acaagacccg cgccgaggtg aagttcgagg
1381 gcgacaccct ggtgaaccgc atcgagctga agggcatcga cttcaaggag gacggcaaca
1441 tcctggggca caagctggag tacaactaca acagccacaa cgtctatatc atggccgaca
1501 agcagaagaa cggcatcaag gtgaacttca agatccgcca caacatcgag gacggcagcg
1561 tgcagctcgc cgaccactac cagcagaaca cccccatcgg cgacggcccc gtgctgctgc
1621 ccgacaacca ctacctgagc acccagtccg ccctgagcaa agaccccaac gagaagcgcg
1681 atcacatggt cctgctggag ttcgtgaccg ccgccgggat cACTCTCGGC ATGGACGAGC
1741 TGTACAAGTG Atagcggcgc ttttgaccgg gaactcacgc atttaatgta ttcgaaacac
1801 gctttttaag atctcaaatg aaacatgtat ccgtatgctc tccacatccg tcctcacgta
1861 cggtgtgatg tttttttaat tttttatatg aaagtgtaag gcttctccat tttacttttt
1921 gtgtttctgt cttcacagtc tttctctgcg agctgtgaga gggtcatgcc ctttatgggc
1981 acgacggcag gattttggtt tttgaaagaa tttctgtcgg ctgtaagtct tcaaaaaaga
2041 agaaagcggg gacggaaact tcttgcttgt gtgttataca ttacaaagtc ctctacaatt
2101 ctacagtcct agttgttcac gcgacttcac gcagcagacc agatttatag ctttgcatga
2161 tcgtctccga gcatttcaga atgctataat atggaatgat cttccgctgg cacctcatta
2221 tgcgatacgc actccatata aacacaggta tccctgtccG TATTCGTTAA CTGTTAATTC
2281 ATGACACAAc atcctgctgt atgattttgg cacaacgacg cggttggagg gtgagggaga
2341 tggtggtgag acatcagaaa tggagagaat aacggtagag aaaaaaagga acacacgcgc
2401 ccaggataga accaggacca cattccaagt ccctcaccag caacaagagg gacaccacgt
2461 cggtgaagag gtgcgtgaga cagaggacac cagacaacgt gcacaaagcg tcggaaaggg
2521 aggcgaccgc ttcgctggag gggggccgtc cacagcctgc gcccaaccac acacattcct
2581 cccgcgaggc tcaagtgcca tgccgccgag tgccatgctc ctgcatccgg ggcaggtcga
2641 cgagccagca tggtttggtc gccgacaacg cagatttgat ggagcttggg catttcttta
2701 ccggacacgt cgcttcttac atttttacat tccaattcag aggtcacacc cgcgatcgag
2761 aaagcagaga gtcggacccc cttcaagcag gaaaggcaag caacataccc atcgttaagt
2821 ctgatgatgg ctggtgcaaa gccaaggcag ccagtcgata tgttaatgat cgggtacgtt
2881 taaaggcgtg ggcaaggtta tttaacgctg gtagcatcgt agagttttta acgcaactga
2941 agatttcatg tctgacttgg agatacggtc ggggcagcag ctatgggtaa aaagcctgaa
3001 ctcaccgcga cgtctgtcga gaagtttctg atcgaaaagt tcgacagcgt ctccgacctg
3061 atgcagctct cggagggcga agaatctcgt gctttcagct tcgatgtagg agggcgtgga
3121 tatgtcctgc gggtaaatag ctgcgccgat ggtttctaca aagatcgtta tgtttatcgg
3181 cactttgcat cggccgcgct cccgattccg gaagtgcttg acattgggga attcagcgag
3241 agcctgacct attgcatctc ccgccgtgca cagggtgtca cgttgcaaga cctgcctgaa
3301 accgaactgc ccgctgttct gcagccggtc gcggaggcca tggatgcgat cgctgcggcc
3361 gatcttagcc agacgagcgg gttcggccca ttcggaccgc aaggaatcgg tcaatacact
3421 acatggcgtg atttcatatg cgcgattgct gatccccatg tgtatcactg gcaaactgtg
3481 atggacgaca ccgtcagtgc gtccgtcgcg caggctctcg atgagctgat gctttgggcc
3541 gaggactgcc ccgaagtccg gcacctcgtg cacgcggatt tcggctccaa caatgtcctg
3601 acggacaatg gccgcataac agcggtcatt gactggagcg aggcgatgtt cggggattcc
3661 caatacgagg tcgccaacat cttcttctgg aggccgtggt tggcttgtat ggagcagcag
3721 acgcgctact tcgagcggag gcatccggag cttgcaggat cgccgcggct ccgggcgtat
3781 atgctccgca ttggtcttga ccaactctat cagagcttgg ttgacggcaa tttcgatgat
3841 gcagcttggg cgcagggtcg atgcgacgca atcgtccgat ccggagccgg gactgtcggg
3901 cgtacacaaa tcgcccgcag aagcgcggcc gtctggaccg atggctgtgt agaagtactc
3961 gccgatagtg gaaaccgacg ccccagcact cgtccgaggg caaaggaata atgataacag
4021 cgcttagagc tcgaatttcc ccgatcgttc aaacatttgg caataaagtt tcttaagatt
4081 gaatcctgtt gccggtcttg cgatgattat catataattt ctgttgaatt acgttaagca
4141 tgtaataatt aacatgtaat gcatgacgtt atttatgaga tgggttttta tgattagagt
4201 cccgcaatta tacatttaat acgcgataga aaacaaaata tagcgcgcaa actaggataa
4261 attatcgcgc gcggtgtcat ctatgttact agatcgggaa ttggttccgg aaccaattcg
4321 taatcatggt catagctgtt tcctgtgtga aattgttatc cgctcacaat tccacacaac
4381 atacgagccg gaagcataaa gtgtaaagcc tggggtgcct aatgagtgag ctaactcaca
4441 ttaggctgaa ttaggcgcgc ctatttctgT TAGTGTCAAA CAGTCAAACC AGTTCTACGt
4501 gtagaacagg atgcatggct gctgcaaggc tgtttgtgcc aacgccgcaa atggaaaaga
4561 aagtaagaaa aaaaatccta cacagacaac ctttctagca cgaaaatgcc cttgcgtaca
4621 tctttttgcg tctcagtaat attgttttct taatcgaaac cattattgtt caagtcatcg
4681 ctaagacact tccatcgtgt gctcttagcc gcatctcggt cctccctcgc catggaacat
4741 gtcatttagg gcgacaaacc acaaataaca tagaagcttt catcgataat gccccatcag
4801 tcctatttaa atagtccaat acaaatatat aagggtttgt catttctcgt ttcctccggc
```

```
4861 tgtcccatgc tgtagacatg tgtacgtagc tagaaggagc ccccatatgt tctttccgac
4921 gatacacatg gctttaatac tatccccata gtgacaaagc gtggacacag aaacacacac
4981 acatgccact cacatagaaa cggccaatcc acgatcgtta catctccgct ttcgtttccg
5041 gcaacgactt gaataattgg ggcatctgct cacgctcgtt ttacccgacc gcaaaaaaaa
5101 atggagagtg gccaccatgc ttcaaggaga agctcggtat atgtccatcc cgcgtgcctg
5161 caaagacaag gtgaagatgc acccggcgac catagtgtgt catgtacctg tttgacgttt
5221 actacagggt gccatgcttt attggctgga attgtggacg cttgttttct ggaatacgaa
5281 tatttccccg ccctgtatta attcccttca agcagcaagc tatcttaaca aacgttgtgc
5341 ggccaaagga aaatagccgc cgagtccata tggggtcttc tgacttcaaa atctcgcgca
5401 ttccaacgtg taaatgtgcg atcgaaagag aaagcattga cacagggaat aaaaactcca
5461 tcaattggaa agatctctgg aatcggtggc accctgttcc gtgaagaagc gaagctttgc
5521 ttgcggccgt atgctcatta ctcacccact gactcgttca cataataaaa gtcccgctaa
5581 aaccctacaa tttcgcttcc gacgctccat cattccatca tagaaattga caggtctact
5641 gaaatcatag ttccacccgt tggtgcgtgt ccttcctgca GTTTAAACTT GTTATCCGCT
5701 CACAATTCCA cacaacatag gagccggaag cataaagtgt aaagcctggg gtgcctaatg
5761 agtgaggtaa ctcacattaa ttgcgttgcg ctcactgccc gctttccagt cgggaaacct
5821 gtcgtgccag ctgcattaat gaatcggcca acgcgcgggg agaggcggtt tgcgtattgg
5881 gcgctcttcc gcttcctcgc tcactgactc gctgcgctcg gtcgttcggc tgcggcgagc
5941 ggtatcagct cactcaaagg cggtaatacg gttatccaca gaatcagggg ataacgcagg
6001 aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct
6061 ggcgtttttc cataggctcc gccccccctga cgagcatcac aaaaatcgac gctcaagtca
6121 gaggtggcga aacccgacag gactataaag ataccaggcg tttcccccctg gaagctccct
6181 cgtgcgctct cctgttccga ccctgccgct taccggatac ctgtccgcct ttctcccttc
6241 gggaagcgtg gcgctttctc atagctcacg ctgtaggtat ctcagttcgg tgtaggtcgt
6301 tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct gcgccttatc
6361 cggtaactat cgtcttgagt ccaacccggt aagacacgac ttatcgccac tggcagcagc
6421 cactggtaac aggattagca gagcgaggta tgtaggcggt gctacagagt tcttgaagtg
6481 gtggcctaac tacggctaca ctagaaggac agtatttggt atctgcgctc tgctgaagcc
6541 agttaccttc ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag
6601 cggtggtttt tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga
6661 tcctttgatc ttttctacgg ggtctgacgc tcagtggaac gaaaactcac gttaagggat
6721 tttggtcatg agattatcaa aaaggatctt cacctagatc cttttaaatt aaaaatgaag
6781 ttttaaatca atctaaagta tatatgagta aacttggtct gacagttacc aatgcttaat
6841 cagtgaggca cctatctcag cgatctgtct atttcgttca tccatagttg cctgactccc
6901 cgtcgtgtag ataactacga tacgggaggg cttaccatct ggccccagtg ctgcaatgat
6961 accgcgagac ccacgctcac cggctccaga tttatcagca ataaaccagc cagccggaag
7021 ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc atccagtcta ttaattgttg
7081 ccgggaagct agagtaagta gttcgccagt taatagtttg cgcaacgttg ttgccattgc
7141 tacaggcatc gtggtgtcac gctcgtcgtt tggtatggct tcattcagct ccggttccca
7201 acgatcaagg cgagttacat gatcccccat gttgtgcaaa aaagcggtta gctccttcgg
7261 tcctccgatc gttgtcagaa gtaagttggc cgcagtgtta tcactcatgg ttatggcagc
7321 actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga ctggtgagta
7381 ctcaaccaag tcattctgag aatagtgtat gcggcgaccg agttgctctt gcccggcgtc
7441 aatacgggat aataccgcgc cacatagcag aactttaaaa gtgctcatca ttggaaaacg
7501 ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt cgatgtaacc
7561 cactcgtgca cccaactgat cttcagcatc ttttactttc accagcgttt ctgggtgagc
7621 aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg gcgacacgga aatgttgaat
7681 actcatactc ttccttttc aatattattg aagcatttat cagggttatt gtctcatgag
7741 cggatacata tttgaatgta tttagaaaaa taaacaaata ggggttccgc gcacatttcc
7801 ccgaaaagtg ccacctgaac gaagcatctg tgcttcattt tgtagaacaa aaatgcaacg
7861 cgagagcgct aatttttcaa acaaagaatc tgagctgcat ttttacagaa cagaaatgca
7921 acgcgaaagc gctattttac caacgaagaa tctgtgcttc atttttgtaa aacaaaaatg
7981 caacgcgaga gcgctaattt ttcaaacaaa gaatctgagc tgcatttta cagaacagaa
8041 atgcaacgcg agagcgctat tttaccaaca aagaatctat acttcttttt tgttctacaa
8101 aaatgcatcc cgagagcgct atttttctaa caaagcatct tagattactt tttttctcct
8161 ttgtgcgctc tataatgcag tctcttgata actttttgca ctgtaggtcc gttaaggtta
8221 gaagaaggct actttggtgt ctattttctc ttccataaaa aaagcctgac tccacttccc
8281 gcgtttactg attactagcg aagctgcggg tgcatttttt caagataaag gcatccccga
8341 ttatattcta taccgatgtg gattgcgcat actttgtgaa cagaaagtga tagcgttgat
8401 gattcttcat tggtcagaaa attatgaacg gtttcttcta ttttgtctct atatactacg
8461 tataggaaat gtttacattt tcgtattgtt ttcgattcac tctatgaata gttcttacta
8521 caatttttt gtctaaagag taatactaga gataaacata aaaaatgtag aggtcgagtt
8581 tagatgcaag ttcaaggagc gaaaggtgga tgggtaggtt atatagggat atagcacaga
8641 gatatatagc aaagagatac ttttgagcaa tgtttgtgga agcggtattc gcaatatttt
8701 agtagctcgt tacagtccgg tgcgtttttg gttttttgaa agtgcgtctt cagagcgctt
8761 ttggttttca aaagcgctct gaagttccta tactttctag agaataggaa cttcggaata
8821 ggaacttcaa agcgtttccg aaaacgagcg cttccgaaaa tgcaacgcga gctgcgcaca
8881 tacagctcac tgttcacgtc gcacctatat ctgcgtgttg cctgtatata tatatacatg
8941 agaagaacgg catagtgcgt gtttatgctt aaatgcgtac ttatatgcgt ctatttatgt
9001 aggatgaaag gtagtctagt acctcctgtg atattatccc attccatgcg gggtatcgta
9061 tgcttccttc agcactaccc tttagctgtt ctatatgctg ccactcctca attggattag
9121 tctcatcctt caatgctatc atttcctttg atattggatc atactaagaa accattatta
9181 tcatgacatt aacctataaa aataggcgta tcacgaggcc ctttcgtctc gcgcgtttcg
9241 gtgatgacgg tgaaaacctc tgacacatgc agctcccgga gacggtcaca gcttgtctgt
9301 aagcggatgc cgggagcaga caagcccgtc agggcgcgtc agcgggtgtt ggcgggtgtc
9361 ggggctggct taactatgcg gcatcagagc agattgtact gagagtgcac cataccacag
9421 cttttcaatt caattcatca ttttttttttt attcttttt ttgatttcgg tttctttgaa
9481 attttttga ttcggtaatc tccgaacaga aggaagaacg aaggaaggag cacagactta
9541 gattggtata tatacgcata tgtagtgttg aagaaacatg aaattgccca gtattcttaa
9601 cccaactgca cagaacaaaa acctgcagga aacgaagata aatcatgtcg aaagctacat
```

```
     9661 ataaggaacg tgctgctact catcctagtc ctgttgctgc caagctattt aatatcatgc
     9721 acgaaaagca aacaaacttg tgtgcttcat tggatgttcg taccaccaag gaattactgg
     9781 agttagttga agcattaggt cccaaaattt gtttactaaa aacacatgtg gatatcttga
     9841 ctgatttttc catggagggc acagttaagc cgctaaaggc attatccgcc aagtacaatt
     9901 ttttactctt cgaagacaga aaatttgctg acattggtaa tacagtcaaa ttgcagtact
     9961 ctgcgggtgt atacagaata gcagaatggg cagacattac gaatgcacac ggtgtggtgg
    10021 gcccaggtat tgttagcggt ttgaagcagg cggcagaaga agtaacaaag gaacctagag
    10081 gccttttgat gttagcagaa ttgtcatgca agggctccct atctactgga gaatatacta
    10141 agggtactgt tgacattgcg aagagcgaca aagattttgt tatcggcttt attgctcaaa
    10201 gagacatggg tggaagagat gaaggttacg attggttgat tatgacaccc ggtgtgggtt
    10261 tagatgacaa gggagacgca ttgggtcaac agtatagaac cgtggatgat gtggtctcta
    10321 caggatctga cattattatt gttggaagag gactatttgc aaagggaagg gatgctaagg
    10381 tagagggtga acgttacaga aaagcaggct gggaagcata tttgagaaga tgcggccagc
    10441 aaaactaaaa aactgtatta taagtaaatg catgtatact aaactcacaa attagagctt
    10501 caatttaatt atatcagtta ttaccctatg cggtgtgaaa taccgcacag atgcgtaagg
    10561 agaaaatacc gcatcaggaa attgtaaacg ttaatatttt gttaaaattc gcgttaaatt
    10621 tttgttaaat cagctcattt tttaaccaat aggccgaaat cggcaaaatc ccttataaat
    10681 caaaagaata gaccgagata gggttgagtg ttgttccagt ttggaacaag agtccactat
    10741 taaagaacgt ggactccaac gtcaaagggc gaaaaaccgt ctatcagggc gatggcccac
    10801 tacgtgaacc atcaccctaa tcaagtttt tggggtcgag gtgccgtaaa gcactaaatc
    10861 ggaaccctaa agggagcccc cgatttagag cttgacgggg aaagccggcg aacgtggcga
    10921 gaaaggaagg gaagaaagcg aaaggagcgg gcgctagggc gctggcaagt gtagcggtca
    10981 cgctgcgcgt aaccaccaca cccgccgcgc ttaatgcgcc gctacagggc gcgtcgcgcc
    11041 attcgccatt caggctgcgc aactgtt
//
```

A.6 pCC7 – Genbank formatted vector sequence

```
LOCUS       pCC7    11044 bp      DNA     circular   15/11/2017
DEFINITION
ACCESSION
VERSION
SOURCE
  ORGANISM
COMMENT     Serial Cloner Genbank Format
COMMENT     SerialCloner_Type=DNA
COMMENT     SerialCloner_Comments=Homologous recombination reaction between [pLCP hph]
as a ta
            rget sequence  and [PCC7] as an inserted sequence.
COMMENT     SerialCloner_Ends=0,0,,0,
FEATURES             Location/Qualifiers
     misc_feature    complement(9234..9256)
                     /label=pGEX_3_primer
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pGEX_3_primer" - complement(29..51) \
     promoter        9407..9620
                     /label=URA3_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3_promoter" - 202..415 \
     gene            9622..10422
                     /label=URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - 417..1217 \
     CDS             9622..10425
                     /label=ORF frame 3_URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - note="ORF frame 3"
- translation="MSKATYKERAATHPSPVAAKLFNIMHEKQTNLCASLDVRTTKEL -
LELVEALGPKICLLKTHVDILTDFSMEGTVKPLKALSAKYNFLLFEDRKFADIGNTVK -
LQYSAGVYRIAEWADITNAHGVVGPGIVSGLKQAAEEVTKEPRGLLMLAELSCKGSLS - TGEYTKGTVDIAKSDKDFVIGFI \
     rep_origin      complement(10688..10994)
                     /label=f1_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="f1_origin" - complement(1483..1789) \
     misc_feature    complement(11027..24)
                     /label=lacZ_a [5']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="lacZ_a" - complement(1822..1965) \
     misc_feature    complement(5664..24)
                     /label=M13_pUC_rev_primer [3'] [5'] [3']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="M13_pUC_rev_primer" -
complement(2176..2198) \
     rep_origin      complement(6029..6648)
                     /label=pBR322_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
```

```
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pBR322_origin" - complement(2550..3169) \
     CDS             complement(6803..7663)
                     /label=ORF frame 3_Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
note="ORF frame 3" - translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY -
IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE -
YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL - DRWEPELNEAI \
     gene            complement(6803..7663)
                     /label=Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
complement(3324..4184) \
     promoter        complement(7705..7733)
                     /label=AmpR_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="AmpR_promoter" - complement(4226..4254) \
     rep_origin      complement(7795..8766)
                     /label=2micron2_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="2micron2_origin" - complement(4316..5287)
\
     misc_feature    complement(8758..8805)
                     /label=FRT
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="FRT" - complement(5279..5326) \
     misc_feature    2..31
                     /label=XX
                     /SerialCloner_Color=&h008000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    5658..5687
                     /label=YY
                     /SerialCloner_Color=&h8080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2237..2266
                     /label=CC
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4447..4476
                     /label=DD
                     /SerialCloner_Color=&h800040
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    3989..4446
                     /label=NOS terminator_regulatory
                     /SerialCloner_Color=&hFF0000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=regulatory_class="terminator" - note="NOS
terminator" - 2911..3320 \
     misc_feature    2267..2959
                     /label=Puep
                     /SerialCloner_Color=&hFF80C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
```

```
                     /SerialCloner_Arrow=True
     misc_feature    4477..4498
                     /label=P3
                     /SerialCloner_Color=&hFFFF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    5638..5657
                     /label=P4
                     /SerialCloner_Color=&hFFFF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     gene            2960..3988
                     /label=hph_gene
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=gene="hph" - 485..1513 \
     CDS             2960..3988
                     /label=hph_CDS
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=gene="hph" - codon_start=1 - transl_table=11 -
product="hygromycin B phosphotransferase" - protein_id="CAR31381.1" -
db_xref="GI:194473371" - translation="MGKKPELTATSVEKFLIEKFDSVSDLMQLSEGEESRAFSFDVGG -
RGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQ \
     misc_feature    2960..2984
                     /label=hph fw
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    3963..3988
                     /label=hph rv
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1699..1728
                     /label=gfp ext
                     /SerialCloner_Color=&h996633
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1729..1745
                     /label=Tatpa P1 fw
                     /SerialCloner_Color=&hFF00FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2217..2236
                     /label=Tatpa P2 rv
                     /SerialCloner_Color=&hFF00FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1009..1728
                     /label=gfp
                     /SerialCloner_Color=&h00FF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    32..1008
                     /label=PCC7
                     /SerialCloner_Color=&h0000FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1009..1038
                     /label=gfp ext
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
ORIGIN
        1 gGGAAGGGCG ATCGGTGCGG GCCGTTTAAA Caacttttcc caagaaggat ggtaCGTTGC
```

```
  61 GTTCGTCCGG CCCATCTTCT CAGATAGGGC AAGGAAATGC GTGATATCCC AGAACGTGGA
 121 CGCTAACGGA CTGATTTCAG CAGCGTTATG AAGCACTTTG TTTTGAACTA TTTGGCCTTT
 181 GTCTTAAAAA CAGAAATCAT GAGAAGCCCT GCAACCCCTG ATCACCAAGA CGAGTCGCAC
 241 AATTTTCGGA CCGGAAAAAA GGAGGGATTG GTGAGAAACG TGACAAGCCC TACTGTACGT
 301 CTTCGCTGCC CGTCTCCTCG AAGGAAAAAA ATATCGAACA AGCTACTATT TAAGTGGAGT
 361 AATCGAGCAC GGTGCAGCGA GGCCGCCACA GCGTTTGAGA CTATGGCCAC TCCCATTAAA
 421 AGTCCAACAC AAAAAAATAT GAGCCCCGAT TCTGAAGTGG AGAGTCACAA CTTTGAACCA
 481 CATCACAGTG AGGTATGGAG AATACATGAG CGGAAACGGA GGCAGGTTGA ATTTATCGAC
 541 CCGCAGAGCG GCCGACAACG TGCCTTCATT CGATGGATGC TTACCGCCCT TTTGGCAATT
 601 TTAAGCGAGA TTTTGTGCGT CGTAATCATT GCATGCACCG TCGAGTTGCA GGCTCTAAAA
 661 TTTGACACAA CGCTCAAAAT CGCATCTCAT CGCAAGTGGT ATACCAGCGT GGCAGCTTTC
 721 TTTCTTTTTC TTACATTTTC TCTGCTATTT GCGGTGGTTG CTGGGGTCGT GACGCTTTTG
 781 CAGCCAGCTG TAGGAGGTTC AGGGATGGCA GAGGTAAAGA GTTACTTAAA TGGCGTGAAT
 841 CTTCCTCACG TAATGCACTT CAACACTTTA TATGGCAAGG CTGTTGGGGT AATATTTGCG
 901 CAAGCTTCAG GATTGCCTTT GGGATTCGAA GGTCCCATTG TCTCTTGTTG CGCGATTTTA
 961 GGTTCTTTGT TTTCGCAAGG CAAAATTTGG Ggccttgagc gtcgacttAT GGTGAGCAAG
1021 GGCGAGGAGC TGTTCACCgg ggtggtgccc atcctggtcg agctggacgg cgacgtaaac
1081 ggccacaagt tcagcgtgtc cggcgagggc gagggcgatg ccacctacgg caagctgacc
1141 ctgaagttca tctgcaccac cggcaagctg cccgtgccct ggcccaccct cgtgaccacc
1201 ctgacctacg gcgtgcagtg cttcagccgc taccccgacc acatgaagca gcacgacttc
1261 ttcaagtccg ccatgcccga aggctacgtc caggagcgca ccatcttctt caaggacgac
1321 ggcaactaca agacccgcgc cgaggtgaag ttcgagggcg acaccctggt gaaccgcatc
1381 gagctgaagg gcatcgactt caaggaggac ggcaacatcc tggggcacaa gctggagtac
1441 aactacaaca gccacaacgt ctatatcatg gccgacaagc agaagaacgg catcaaggtg
1501 aacttcaaga tccgccacaa catcgaggac ggcagcgtgc agctcgccga ccactaccag
1561 cagaacaccc ccatcggcga cggccccgtg ctgctgcccg acaaccacta cctgagcacc
1621 cagtccgccc tgagcaaaga ccccaacgag aagcgcgatc acatggtcct gctggagttc
1681 gtgaccgccg ccgggatcAC TCTCGGCATG GACGAGCTGT ACAAGTGAta gcggcgcttt
1741 tgaccgggaa ctcacgcatt taatgtattc gaaacacgct ttttaagatc tcaaatgaaa
1801 catgtatccg tatgctctcc acatccgtcc tcacgtacgg tgtgatgttt ttttaatttt
1861 ttatatgaaa gtgtaaggct tctccatttt acttttgtg tttctgtctt cacagtcttt
1921 ctctgcgagc tgtgagaggg tcatgccctt tatgggcacg acggcaggat tttggttttt
1981 gaaagaattt ctgtcggctg taagtcttca aaaaagaaga aagcggggac ggaaacttct
2041 tgcttgtgtg ttatacatta caaagtcctc tacaattcta cagtcctagt tgttcacgcg
2101 acttcacgca gcagaccaga tttatagctt tgcatgatcg tctccgagca tttcagaatg
2161 ctataatatg gaatgatctt ccgctggcac ctcattatgc gatacgcact ccatataaac
2221 acaggtatcc ctgtccGTAT TCGTTAACTG TTAATTCATG ACACAAcatc ctgctgtatg
2281 attttggcac aacgacgcgg ttggagggtg agggagatgg tggtgagaca tcagaaatgg
2341 agagaataac ggtagagaaa aaaaggaaca cacgcgccca ggatagaacc aggaccacat
2401 tccaagtccc tcaccagcaa caagagggac accacgtcgg tgaagaggtg cgtgagacag
2461 aggacaccag acaacgtgca caaagcgtcg gaaagggagg cgaccgcttc gctggagggg
2521 ggccgtccac agcctgcgcc caaccacaca cattcctccc gcgaggctca agtgccatgc
2581 cgccgagtgc catgctcctg catccggggc aggtcgacga gccagcatgg tttggtcgcc
2641 gacaacgcag atttgatgga gcttgggcat ttctttaccg gacacgtcgc ttcttacatt
2701 tttacattcc aattcagagg tcacacccgc gatcgagaaa gcagagagtc ggacccccct
2761 caagcaggaa aggcaagcaa catacccatc gttaagtctg atgatggctg gtgcaaagcc
2821 aaggcagcca gtcgatatgt taatgatcgg gtacgtttaa aggcgtgggc aaggttattt
2881 aacgctggta gcatcgtaga gtttttaacg caactgaaga tttcatgtct gacttggaga
2941 tacggtcggg gcagcagcta tgggtaaaaa gcctgaactc accgcgacgt ctgtcgagaa
3001 gtttctgatc gaaaagttcg acagcgtctc cgacctgatg cagctctcgg agggcgaaga
3061 atctcgtgct ttcagcttcg atgtaggagg gcgtggatat gtcctgcggg taaatagctg
3121 cgccgatggt ttctacaaag atcgttatgt ttatcggcac tttgcatcgg ccgcgctccc
3181 gattccggaa gtgcttgaca ttggggaatt cagcgagagc ctgacctatt gcatctcccg
3241 ccgtgcacag ggtgtcacgt tgcaagacct gcctgaaacc gaactgcccg ctgttctgca
3301 gccggtcgcg gaggccatgg atgcgatcgc tgcggccgat cttagccaga cgagcgggtt
3361 cggcccattc ggaccgcaag gaatcggtca atacactaca tggcgtgatt tcatatgcgc
3421 gattgctgat ccccatgtgt atcactggca aactgtgatg gacgacaccg tcagtgcgtc
3481 cgtcgcgcag gctctcgatg agctgatgct ttgggccgag gactgccccg aagtccggca
3541 cctcgtgcac gcggatttcg gctccaacaa tgtcctgacg gacaatggcc gcataacagc
3601 ggtcattgac tggagcgagg cgatgttcgg ggattcccaa tacgaggtcg ccaacatctt
3661 cttctggagg ccgtggttgg cttgtatgga gcagcagacg cgctacttcg agcggaggca
3721 tccggagctt gcaggatcgc cgcggctccg ggcgtatatg ctccgcattg gtcttgacca
3781 actctatcag agcttggttg acggcaattt cgatgatgca gcttgggcgc agggtcgatg
3841 cgacgcaatc gtccgatccg gagccgggac tgtcgggcgt acacaaatcg cccgcagaag
3901 cgcggccgtc tggaccgatg gctgtgtaga agtactcgcc gatagtggaa accgacgccc
3961 cagcactcgt ccgagggcaa aggaataatg ataacagcgc ttagagctcg aatttccccg
4021 atcgttcaaa catttggcaa taaagtttct taagattgaa tcctgttgcc ggtcttgcga
4081 tgattatcat ataatttctg ttgaattacg ttaagcatgt aataattaac atgtaatgca
4141 tgacgttatt tatgagatgg gtttttatga ttagagtccc gcaattatac atttaatacg
4201 cgatagaaaa caaaatatag cgcgcaaact aggataaatt atcgcgcgcg gtgtcatcta
4261 tgttactaga tcgggaattg gttccggaac caattcgtaa tcatggtcat agctgtttcc
4321 tgtgtgaaat tgttatccgc tcacaattcc acacaacata cgagccggaa gcataaagtg
4381 taaagcctgg ggtgcctaat gagtgagcta actcacatta ggctgaatta ggcgcgccta
4441 tttctgTTAG TGTCAAACAG TCAAACCAGT TCTACGtgta gaacaggatg catggctgct
4501 gcaaggctgt ttgtgccaac gccgcaaatg gaaaagaaag taagaaaaaa aatcctacac
4561 agacaacctt tctagcacga aaatgccctt gcgtacatct ttttgcgtct cagtaatatt
4621 gttttcttaa tcgaaaccat tattgttcaa gtcatcgcta agacacttcc atcgtgtgct
4681 cttagccgca tctcggtcct ccctcgccat ggaacatgtc atttagggcg acaaaccaca
4741 aataacatag aagctttcat cgataatgcc ccatcagtcc tatttaaata gtccaataca
4801 aatatataag ggtttgtcat ttctcgtttc ctccggctgt cccatgctgt agacatgtgt
```

```
4861 acgtagctag aaggagcccc catatgttct ttccgacgat acacatggct ttaatactat
4921 ccccatagtg acaaagcgtg gacacagaaa cacacacaca tgccactcac atagaaacgg
4981 ccaatccacg atcgttacat ctccgctttc gtttccggca acgacttgaa taattggggc
5041 atctgctcac gctcgtttta cccgaccgca aaaaaaaatg gagagtggcc accatgcttc
5101 aaggagaagc tcggtatatg tccatcccgc gtgcctgcaa agacaaggtg aagatgcacc
5161 cggcgaccat agtgtgtcat gtacctgttt gacgtttact acagggtgcc atgctttatt
5221 ggctggaatt gtggacgctt gttttctgga atacgaatat ttccccgccc tgtattaatt
5281 cccttcaagc agcaagctat cttaacaaac gttgtgcggc caaaggaaaa tagccgccga
5341 gtccatatgg ggtcttctga cttcaaaatc tcgcgcattc caacgtgtaa atgtgcgatc
5401 gaaagagaaa gcattgacac agggaataaa aactccatca attggaaaga tctctggaat
5461 cggtggcacc ctgttccgtg aagaagcgaa gctttgcttg cggccgtatg ctcattactc
5521 acccactgac tcgttcacat aataaaagtc ccgctaaaac cctacaattt cgcttccgac
5581 gctccatcat tccatcatag aaattgacag gtctactgaa atcatagttc cacccgttgg
5641 tgcgtgtcct tcctgcaGTT TAAACTTGTT ATCCGCTCAC AATTCCAcac aacataggag
5701 ccggaagcat aaagtgtaaa gcctggggtg cctaatgagt gaggtaactc acattaattg
5761 cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa
5821 tcggccaacg cgcggggaga ggcggtttgc gtattgggcg ctcttccgct tcctcgctca
5881 ctgactcgct gcgctcggtc gttcggctgc ggcgagcggt atcagctcac tcaaaggcgg
5941 taatacggtt atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc
6001 agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc
6061 cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac
6121 tataaagata ccaggcgttt ccccctggaa gctccctcgt gcgctctcct gttccgaccc
6181 tgccgcttac cggatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcata
6241 gctcacgctg taggtatctc agttcggtgt aggtcgttcg ctccaagctg ggctgtgtgc
6301 acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca
6361 acccggtaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag
6421 cgaggtatgt aggcggtgct acagagttct tgaagtggtg gcctaactac ggctacacta
6481 gaaggacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg
6541 gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggtttttt gtttgcaagc
6601 agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt
6661 ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaga ttatcaaaaa
6721 ggatcttcac ctagatcctt ttaaattaaa aatgaagttt taaatcaatc taaagtatat
6781 atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct atctcagcga
6841 tctgtctatt tcgttcatcc atagttgcct gactccccgt cgtgtagata actacgatac
6901 gggagggctt accatctggc cccagtgctg caatgatacc gcgagaccca cgctcaccgg
6961 ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga agtggtcctg
7021 caactttatc cgcctccatc cagtctatta attgttgccg ggaagctaga gtaagtagtt
7081 cgccagttaa tagtttgcgc aacgttgttg ccattgctac aggcatcgtg gtgtcacgct
7141 cgtcgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga gttacatgat
7201 cccccatgtt gtgcaaaaaa gcggttagct ccttcggtcc tccgatcgtt gtcagaagta
7261 agttggccgc agtgttatca ctcatggtta tggcagcact gcataattct cttactgtca
7321 tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca ttctgagaat
7381 agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat accgcgccac
7441 atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga aaactctcaa
7501 ggatcttacc gctgttgaga tccagttcga tgtaacccac tcgtgcaccc aactgatctt
7561 cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg caaaatgccg
7621 caaaaaaggg aataagggcg acacggaaat gttgaatact catactcttc cttttcaat
7681 attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt gaatgtattt
7741 agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcca cctgaacgaa
7801 gcatctgtgc ttcattttgt agaacaaaaa tgcaacgcga gagcgctaat ttttcaaaca
7861 aagaatctga gctgcatttt tacagaacag aaatgcaacg cgaaagcgct attttaccaa
7921 cgaagaatct gtgcttcatt tttgtaaaac aaaaatgcaa cgcgagagcg ctaatttttc
7981 aaacaaagaa tctgagctgc atttttacag aacagaaatg caacgcgaga gcgctatttt
8041 accaacaaag aatctatact tctttttgt tctacaaaaa tgcatcccga gagcgctatt
8101 tttctaacaa agcatcttag attacttttt ttctcctttg tgcgctctat aatgcagtct
8161 cttgataact ttttgcactg taggtccgtt aaggttagaa gaaggctact ttggtgtcta
8221 ttttctcttc cataaaaaaa gcctgactcc acttcccgcg tttactgatt actagcgaag
8281 ctgcgggtgc atttttcaa gataaaggca tccccgatta tattctatac cgatgtggat
8341 tgcgcatact ttgtgaacag aaagtgatag cgttgatgat tcttcattgg tcagaaaatt
8401 atgaacggtt tcttctattt tgtctctata tactacgtat aggaaatgtt tacattttcg
8461 tattgttttc gattcactct atgaatagtt cttactacaa tttttttgtc taaagagtaa
8521 tactagagat aaacataaaa aatgtagagg tcgagtttag atgcaagttc aaggagcgaa
8581 aggtggatgg gtaggttata tagggatata gcacagagat atatagcaaa gagatacttt
8641 tgagcaatgt ttgtggaagc ggtattcgca atattttagt agctcgttac agtccggtgc
8701 gtttttggtt ttttgaaagt gcgtcttcag agcgcttttg gttttcaaaa gcgctctgaa
8761 gttcctatac tttctagaga ataggaactt cggaatagga acttcaaagc gtttccgaaa
8821 acgagcgctt ccgaaaatgc aacgcgagct gcgcacatac agctcactgt tcacgtcgca
8881 cctatatctg cgtgttgcct gtatatatat atacatgaga agaacggcat agtgcgtgtt
8941 tatgcttaaa tgcgtactta tatgcgtcta tttatgtagg atgaaaggta gtctagtacc
9001 tcctgtgata ttatcccatt ccatgcgggg tatcgtatgc ttccttcagc actacccttt
9061 agctgttcta tatgctgcca ctcctcaatt ggattagtct catccttcaa tgctatcatt
9121 tcctttgata ttggatcata ctaagaaacc attattatca tgacattaac ctataaaaat
9181 aggcgtatca cgaggccctt tcgtctcgcg cgtttcggtg atgacggtga aaacctctga
9241 cacatgcagc tcccggagac ggtcacagct tgtctgtaag cggatgccgg gagcagacaa
9301 gcccgtcagg gcgcgtcagc gggtgttggc gggtgtcggg gctggcttaa ctatgcggca
9361 tcagagcaga ttgtactgag agtgcaccat accacagctt ttcaattcaa ttcatcattt
9421 ttttttttatt ctttttttg atttcggttt ctttgaaatt tttttgattc ggtaatctcc
9481 gaacagaagg aagaacgaag gaaggagcac agacttagat tggtatatat acgcatatgt
9541 agtgttgaag aaacatgaaa ttgcccagta ttcttaaccc aactgcacag aacaaaaacc
9601 tgcaggaaac gaagataaat catgtcgaaa gctacatata aggaacgtgc tgctactcat
```

```
     9661 cctagtcctg ttgctgccaa gctatttaat atcatgcacg aaaagcaaac aaacttgtgt
     9721 gcttcattgg atgttcgtac caccaaggaa ttactggagt tagttgaagc attaggtccc
     9781 aaaatttgtt tactaaaaac acatgtggat atcttgactg atttttccat ggagggcaca
     9841 gttaagccgc taaaggcatt atccgccaag tacaatttt tactcttcga agacagaaaa
     9901 tttgctgaca ttggtaatac agtcaaattg cagtactctg cgggtgtata cagaatagca
     9961 gaatgggcag acattacgaa tgcacacggt gtggtgggcc caggtattgt tagcggtttg
    10021 aagcaggcgg cagaagaagt aacaaaggaa cctagaggcc ttttgatgtt agcagaattg
    10081 tcatgcaagg gctccctatc tactggagaa tatactaagg gtactgttga cattgcgaag
    10141 agcgacaaag attttgttat cggctttatt gctcaaagag acatgggtgg aagagatgaa
    10201 ggttacgatt ggttgattat gacacccggt gtgggtttag atgacaaggg agacgcattg
    10261 ggtcaacagt atagaaccgt ggatgatgtg gtctctacag gatctgacat tattattgtt
    10321 ggaagaggac tatttgcaaa gggaagggat gctaaggtag agggtgaacg ttacagaaaa
    10381 gcaggctggg aagcatattt gagaagatgc ggccagcaaa actaaaaaac tgtattataa
    10441 gtaaatgcat gtatactaaa ctcacaaatt agagcttcaa tttaattata tcagttatta
    10501 ccctatgcgg tgtgaaatac cgcacagatg cgtaaggaga aaataccgca tcaggaaatt
    10561 gtaaacgtta atattttgtt aaaattcgcg ttaaattttt gttaaatcag ctcatttttt
    10621 aaccaatagg ccgaaatcgg caaaatccct tataaatcaa aagaatagac cgagataggg
    10681 ttgagtgttg ttccagtttg gaacaagagt ccactattaa agaacgtgga ctccaacgtc
    10741 aaagggcgaa aaaccgtcta tcagggcgat ggcccactac gtgaaccatc accctaatca
    10801 agtttttgg ggtcgaggtg ccgtaaagca ctaaatcgga accctaaagg gagcccccga
    10861 tttagagctt gacggggaaa gccggcgaac gtggcgagaa aggaagggaa gaaagcgaaa
    10921 ggagcgggcg ctagggcgct ggcaagtgta gcggtcacgc tgcgcgtaac caccacaccc
    10981 gccgcgctta atgcgccgct acagggcgcg tcgcgccatt cgccattcag gctgcgcaac
    11041 tgtt

//
```

A.7 pHYP2 – Genbank formatted vector sequence

```
LOCUS       pHYP2    11067 bp      DNA     circular   15/11/2017
DEFINITION
ACCESSION
VERSION
SOURCE
  ORGANISM
COMMENT     Serial Cloner Genbank Format
COMMENT     SerialCloner_Type=DNA
COMMENT     SerialCloner_Comments=Homologous recombination reaction between [pLCP hph]
as a ta
            rget sequence  and [PHYP1] as an inserted sequence.
COMMENT     SerialCloner_Ends=0,0,,0,
FEATURES             Location/Qualifiers
     misc_feature    complement(9257..9279)
                     /label=pGEX_3_primer
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pGEX_3_primer" - complement(29..51) \
     promoter        9430..9643
                     /label=URA3_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3_promoter" - 202..415 \
     gene            9645..10445
                     /label=URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - 417..1217 \
     CDS             9645..10448
                     /label=ORF frame 3_URA3_URA3
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="URA3" - gene="URA3" - note="ORF frame 3"
- translation="MSKATYKERAATHPSPVAAKLFNIMHEKQTNLCASLDVRTTKEL -
LELVEALGPKICLLKTHVDILTDFSMEGTVKPLKALSAKYNFLLFEDRKFADIGNTVK -
LQYSAGVYRIAEWADITNAHGVVGPGIVSGLKQAAEEVTKEPRGLLMLAELSCKGSLS - TGEYTKGTVDIAKSDKDFVIGFI \
     rep_origin      complement(10711..11017)
                     /label=f1_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="f1_origin" - complement(1483..1789) \
     misc_feature    complement(11050..24)
                     /label=lacZ_a [5']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="lacZ_a" - complement(1822..1965) \
     misc_feature    complement(5687..24)
                     /label=M13_pUC_rev_primer [3'] [5'] [3']
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="M13_pUC_rev_primer" -
complement(2176..2198) \
     rep_origin      complement(6052..6671)
                     /label=pBR322_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
```

```
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="pBR322_origin" - complement(2550..3169) \
     CDS             complement(6826..7686)
                     /label=ORF frame 3_Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
note="ORF frame 3" - translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY -
IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE -
YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL - DRWEPELNEAI \
     gene            complement(6826..7686)
                     /label=Ampicillin_Ampicillin
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="Ampicillin" - gene="Ampicillin" -
complement(3324..4184) \
     promoter        complement(7728..7756)
                     /label=AmpR_promoter
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="AmpR_promoter" - complement(4226..4254) \
     rep_origin      complement(7818..8789)
                     /label=2micron2_origin
                     /SerialCloner_Color=&hF08080
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="2micron2_origin" - complement(4316..5287)
\
     misc_feature    complement(8781..8828)
                     /label=FRT
                     /SerialCloner_Color=&hDC143C
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=label="FRT" - complement(5279..5326) \
     misc_feature    2..31
                     /label=XX
                     /SerialCloner_Color=&h008000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    5681..5710
                     /label=YY
                     /SerialCloner_Color=&h8080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2260..2289
                     /label=CC
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4470..4499
                     /label=DD
                     /SerialCloner_Color=&h800040
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    4012..4469
                     /label=NOS terminator_regulatory
                     /SerialCloner_Color=&hFF0000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=regulatory_class="terminator" - note="NOS
terminator" - 2911..3320 \
     misc_feature    2290..2982
                     /label=Puep
                     /SerialCloner_Color=&hFF80C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
```

```
                     /SerialCloner_Arrow=True
     misc_feature    4500..4521
                     /label=P3
                     /SerialCloner_Color=&hFFFF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    5661..5680
                     /label=P4
                     /SerialCloner_Color=&hFFFF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     gene            2983..4011
                     /label=hph_gene
                     /SerialCloner_Color=&hFF4500
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=gene="hph" - 485..1513 \
     CDS             2983..4011
                     /label=hph_CDS
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=False
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=gene="hph" - codon_start=1 - transl_table=11 -
product="hygromycin B phosphotransferase" - protein_id="CAR31381.1" -
db_xref="GI:194473371" - translation="MGKKPELTATSVEKFLIEKFDSVSDLMQLSEGEESRAFSFDVGG -
RGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQ \
     misc_feature    2983..3007
                     /label=hph fw
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    3986..4011
                     /label=hph rv
                     /SerialCloner_Color=&h0080FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1722..1751
                     /label=gfp ext
                     /SerialCloner_Color=&h996633
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1752..1768
                     /label=Tatpa P1 fw
                     /SerialCloner_Color=&hFF00FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    2240..2259
                     /label=Tatpa P2 rv
                     /SerialCloner_Color=&hFF00FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1032..1751
                     /label=gfp
                     /SerialCloner_Color=&h00FF00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    32..1031
                     /label=PHYP1
                     /SerialCloner_Color=&h0000FF
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    1032..1061
                     /label=gfp ext
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
ORIGIN
        1 gGGAAGGGCG ATCGGTGCGG GCCGTTTAAA Ctcttgctca agaaagtcca aattCTGGTA
```

```
  61 TAGCCGCACG TACTCGTCAT ACTGCCGTGC AAGCTCTTTT TCCAACATGT TGTGTTCTTC
 121 TACCAAAGAT GGCCTGAGAA CATATGATGT CGACATGTCT CAGTAAGAAA GCGTTCGCCG
 181 GGTTAGCCGT GTGTAATTAT ATTTTCGGCT TCATATTCGT GGGAAAGTAC CCCTGATTGC
 241 GGCTCATGCT GCCATCTCAT GCCAATTTCT TGATCCAATT GAAACACGTA CCGTACACTC
 301 TGTAAACTTT GAAGACGGGC CTCGCTGCGT TTGAGTTCGC CTTTCTTTTC CTCGACCTGT
 361 GTACGGCAAA GCTAAAAGTT AACATATAAT GCCATCATGA AAACCACGAG ATGCTATGTG
 421 TACTTACCTT CAGCTCCAGC ATTCGCTCCT CTGCATCCAG TGCAGCATGC TGCCGTTGCA
 481 ACGATTCAAG TGTGCGGCGG AGAGATGGCA CGTCTTCTGT ACTTGTCAAA GTGGTCATGA
 541 CACCAAGAGT CCTGCGTGGA GAGGGGGGAT GACGGCCTCG GATCAAAAAA GAAATAAGGC
 601 GAGCAACGAA GCCATCCCTC GGGCATACTC TTCCATGACT GTAACAAAAT AGTAGTTGGA
 661 ACACTGTTTT TCGTAGAACA TGCAATCTGG CTACGTCCCA AAGCCAAAAG GGGCAGCAAA
 721 GGCAAATTTT CGGCTTTGGT TCTTCCAATT ATTCTTGTTC TCCCAAGGCC ACCCCTACCA
 781 TTCCTTGCCC CAGCCCGAAC TTCAACTCGG AGCCTGACAC CAGCATTCCA CACCAGACTC
 841 GACCGCATTA CCGGCGTGGA AACGCGGGGA ATCCCCACTC GAAGATCTCT TGTTGGTCTT
 901 TGTCCCAGCT TTTCCTTGAG ACGATTTCTC ACGTATTTCA AATGCATTTG ATACCTCAAT
 961 TCTTTGACGG CACATAGAGT TCTAGTGTTC AACACCCTCA TGCCGACTCc ctatctcgct
1021 acatagaggt cATGGTGAGC AAGGGCGAGG AGCTGTTCAC Cggggtggtg cccatcctgg
1081 tcgagctgga cggcgacgta aacggccaca agttcagcgt gtccggcgag ggcgagggcg
1141 atgccaccta cggcaagctg accctgaagt tcatctgcac caccggcaag ctgcccgtgc
1201 cctggcccac cctcgtgacc accctgacct acggcgtgca gtgcttcagc cgctacccg
1261 accacatgaa gcagcacgac ttcttcaagt ccgccatgcc cgaaggctac gtccaggagc
1321 gcaccatctt cttcaaggac gacggcaact acaagacccg cgccgaggtg aagttcgagg
1381 gcgacaccct ggtgaaccgc atcgagctga agggcatcga cttcaaggag gacggcaaca
1441 tcctggggca caagctggag tacaactaca acagccacaa cgtctatatc atggccgaca
1501 agcagaagaa cggcatcaag gtgaacttca agatccgcca caacatcgag gacggcagcg
1561 tgcagctcgc cgaccactac cagcagaaca cccccatcgg cgacggcccc gtgctgctgc
1621 ccgacaacca ctacctgagc acccagtccg ccctgagcaa agaccccaac gagaagcgcg
1681 atcacatggt cctgctggag ttcgtgaccg ccgccgggat cACTCTCGGC ATGGACGAGC
1741 TGTACAAGTG Atagcggcgc ttttgaccgg gaactcacgc atttaatgta ttcgaaacac
1801 gctttttaag atctcaaatg aaacatgtat ccgtatgctc tccacatccg tcctcacgta
1861 cggtgtgatg tttttttaat tttttatatg aaagtgtaag gcttctccat tttacttttt
1921 gtgtttctgt cttcacagtc tttctctgcg agctgtgaga gggtcatgcc ctttatgggc
1981 acgacggcag gattttggtt tttgaaagaa tttctgtcgg ctgtaagtct tcaaaaaaga
2041 agaaagcggg gacggaaact tcttgcttgt gtgttataca ttacaaagtc ctctacaatt
2101 ctacagtcct agttgttcac gcgacttcac gcagcagacc agatttatag ctttgcatga
2161 tcgtctccga gcatttcaga atgctataat atggaatgat cttccgctgg cacctcatta
2221 tgcgatacgc actccatata aacacaggta tccctgtccG TATTCGTTAA CTGTTAATTC
2281 ATGACACAAc atcctgctgt atgattttgg cacaacgacg cggttggagg gtgagggaga
2341 tggtggtgag acatcagaaa tggagagaat aacggtagag aaaaaaagga acacacgcgc
2401 ccaggataga accaggacca cattccaagt ccctcaccag caacaagagg gacaccacgt
2461 cggtgaagag gtgcgtgaga cagaggacac cagacaacgt gcacaaagcg tcggaaaggg
2521 aggcgaccgc ttcgctggag gggggccgtc cacagcctgc gcccaaccac acacattcct
2581 cccgcgaggc tcaagtgcca tgccgccgag tgccatgctc ctgcatccgg ggcaggtcga
2641 cgagccagca tggtttggtc gccgacaacg cagatttgat ggagcttggg catttcttta
2701 ccggacacgt cgcttcttac atttttacat tccaattcag aggtcacacc cgcgatcgag
2761 aaagcagaga gtcggacccc cttcaagcag gaaaggcaag caacataccc atcgttaagt
2821 ctgatgatgg ctggtgcaaa gccaaggcag ccagtcgata tgttaatgat cgggtacgtt
2881 taaaggcgtg ggcaaggtta tttaacgctg gtagcatcgt agagttttta acgcaactga
2941 agatttcatg tctgacttgg agatacggtc ggggcagcag ctatgggtaa aaagcctgaa
3001 ctcaccgcga cgtctgtcga gaagtttctg atcgaaaagt tcgacagcgt ctccgacctg
3061 atgcagctct cggagggcga agaatctcgt gctttcagct tcgatgtagg agggcgtgga
3121 tatgtcctgc gggtaaatag ctgcgccgat ggtttctaca aagatcgtta tgtttatcgg
3181 cactttgcat cggccgcgct cccgattccg gaagtgcttg acattgggga attcagcgag
3241 agcctgacct attgcatctc ccgccgtgca cagggtgtca cgttgcaaga cctgcctgaa
3301 accgaactgc ccgctgttct gcagccggtc gcggaggcca tggatgcgat cgctgcggcc
3361 gatcttagcc agacgagcgg gttcggccca ttcggaccgc aaggaatcgg tcaatacact
3421 acatggcgtg atttcatatg cgcgattgct gatccccatg tgtatcactg gcaaactgtg
3481 atggacgaca ccgtcagtgc gtccgtcgcg caggctctcg atgagctgat gctttgggcc
3541 gaggactgcc ccgaagtccg gcacctcgtg cacgcggatt tcggctccaa caatgtcctg
3601 acggacaatg gccgcataac agcggtcatt gactggagcg aggcgatgtt cggggattcc
3661 caatacgagg tcgccaacat cttcttctgg aggccgtggt tggcttgtat ggagcagcag
3721 acgcgctact tcgagcggag gcatccggag cttgcaggat cgccgcggct ccgggcgtat
3781 atgctccgca ttggtcttga ccaactctat cagagcttgg ttgacggcaa tttcgatgat
3841 gcagcttggg cgcagggtcg atgcgacgca atcgtccgat ccggagccgg gactgtcggg
3901 cgtacacaaa tcgcccgcag aagcgcggcc gtctggaccg atggctgtgt agaagtactc
3961 gccgatagtg gaaaccgacg ccccagcact cgtccgaggg caaaggaata atgataacag
4021 cgcttagagc tcgaatttcc ccgatcgttc aaacatttgg caataaagtt tcttaagatt
4081 gaatcctgtt gccggtcttg cgatgattat catataattt ctgttgaatt acgttaagca
4141 tgtaataatt aacatgtaat gcatgacgtt atttatgaga tgggttttta tgattagagt
4201 cccgcaatta tacatttaat acgcgataga aaacaaaata tagcgcgcaa actaggataa
4261 attatcgcgc gcggtgtcat ctatgttact agatcgggaa ttggttccgg aaccaattcg
4321 taatcatggt catagctgtt tcctgtgtga aattgttatc cgctcacaat tccacacaac
4381 atacgagccg gaagcataaa gtgtaaagcc tggggtgcct aatgagtgag ctaactcaca
4441 ttaggctgaa ttaggcgcgc ctatttctgT TAGTGTCAAA CAGTCAAACC AGTTCTACGt
4501 gtagaacagg atgcatggct gctgcaaggc tgtttgtgcc aacgccgcaa atggaaaaga
4561 aagtaagaaa aaaaatccta cacagacaac ctttctagca cgaaaatgcc cttgcgtaca
4621 tctttttgcg tctcagtaat attgttttct taatcgaaac cattattgtt caagtcatcg
4681 ctaagacact tccatcgtgt gctcttagcc gcatctcggt cctccctcgc catggaacat
4741 gtcatttagg gcgacaaacc acaaataaca tagaagcttt catcgataat gccccatcag
4801 tcctatttaa atagtccaat acaaatatat aagggtttgt catttctcgt ttcctccggc
```

```
4861 tgtcccatgc tgtagacatg tgtacgtagc tagaaggagc ccccatatgt tctttccgac
4921 gatacacatg gctttaatac tatccccata gtgacaaagc gtggacacag aaacacacac
4981 acatgccact cacatagaaa cggccaatcc acgatcgtta catctccgct ttcgtttccg
5041 gcaacgactt gaataattgg ggcatctgct cacgctcgtt ttacccgacc gcaaaaaaaa
5101 atggagagtg gccaccatgc ttcaaggaga agctcggtat atgtccatcc cgcgtgcctg
5161 caaagacaag gtgaagatgc acccggcgac catagtgtgt catgtacctg tttgacgttt
5221 actacagggt gccatgcttt attggctgga attgtggacg cttgttttct ggaatacgaa
5281 tatttccccg ccctgtatta attcccttca agcagcaagc tatcttaaca aacgttgtgc
5341 ggccaaagga aaatagccgc cgagtccata tggggtcttc tgacttcaaa atctcgcgca
5401 ttccaacgtg taaatgtgcg atcgaaagag aaagcattga cacagggaat aaaaactcca
5461 tcaattggaa agatctctgg aatcggtggc accctgttcc gtgaagaagc gaagctttgc
5521 ttgcggccgt atgctcatta ctcacccact gactcgttca cataataaaa gtcccgctaa
5581 aaccctacaa tttcgcttcc gacgctccat cattccatca tagaaattga caggtctact
5641 gaaatcatag ttccacccgt tggtgcgtgt ccttcctgca GTTTAAACTT GTTATCCGCT
5701 CACAATTCCA cacaacatag gagccggaag cataaagtgt aaagcctggg gtgcctaatg
5761 agtgaggtaa ctcacattaa ttgcgttgcg ctcactgccc gctttccagt cgggaaacct
5821 gtcgtgccag ctgcattaat gaatcggcca acgcgcgggg agaggcggtt tgcgtattgg
5881 gcgctcttcc gcttcctcgc tcactgactc gctgcgctcg gtcgttcggc tgcggcgagc
5941 ggtatcagct cactcaaagg cggtaatacg gttatccaca gaatcagggg ataacgcagg
6001 aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct
6061 ggcgtttttc cataggctcc gccccccctga cgagcatcac aaaaatcgac gctcaagtca
6121 gaggtggcga aacccgacag gactataaag ataccaggcg tttcccccctg gaagctccct
6181 cgtgcgctct cctgttccga ccctgccgct taccggatac ctgtccgcct ttctcccttc
6241 gggaagcgtg gcgctttctc atagctcacg ctgtaggtat ctcagttcgg tgtaggtcgt
6301 tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct gcgccttatc
6361 cggtaactat cgtcttgagt ccaacccggt aagacacgac ttatcgccac tggcagcagc
6421 cactggtaac aggattagca gagcgaggta tgtaggcggt gctacagagt tcttgaagtg
6481 gtggcctaac tacggctaca ctagaaggac agtatttggt atctgcgctc tgctgaagcc
6541 agttaccttc ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag
6601 cggtggtttt tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga
6661 tcctttgatc ttttctacgg ggtctgacgc tcagtggaac gaaaactcac gttaagggat
6721 tttggtcatg agattatcaa aaaggatctt cacctagatc cttttaaatt aaaaatgaag
6781 ttttaaatca atctaaagta tatatgagta aacttggtct gacagttacc aatgcttaat
6841 cagtgaggca cctatctcag cgatctgtct atttcgttca tccatagttg cctgactccc
6901 cgtcgtgtag ataactacga tacgggaggg cttaccatct ggccccagtg ctgcaatgat
6961 accgcgagac ccacgctcac cggctccaga tttatcagca ataaaccagc cagccggaag
7021 ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc atccagtcta ttaattgttg
7081 ccgggaagct agagtaagta gttcgccagt taatagtttg cgcaacgttg ttgccattgc
7141 tacaggcatc gtggtgtcac gctcgtcgtt tggtatggct tcattcagct ccggttccca
7201 acgatcaagg cgagttacat gatcccccat gttgtgcaaa aaagcggtta gctccttcgg
7261 tcctccgatc gttgtcagaa gtaagttggc cgcagtgtta tcactcatgg ttatggcagc
7321 actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga ctggtgagta
7381 ctcaaccaag tcattctgag aatagtgtat gcggcgaccg agttgctctt gcccggcgtc
7441 aatacgggat aataccgcgc cacatagcag aactttaaaa gtgctcatca ttggaaaacg
7501 ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt cgatgtaacc
7561 cactcgtgca cccaactgat cttcagcatc ttttactttc accagcgttt ctgggtgagc
7621 aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg gcgacacgga aatgttgaat
7681 actcatactc ttccttttc aatattattg aagcatttat cagggttatt gtctcatgag
7741 cggatacata tttgaatgta tttagaaaaa taaacaaata ggggttccgc gcacatttcc
7801 ccgaaaagtg ccacctgaac gaagcatctg tgcttcattt tgtagaacaa aaatgcaacg
7861 cgagagcgct aatttttcaa acaaagaatc tgagctgcat ttttacagaa cagaaatgca
7921 acgcgaaagc gctattttac caacgaagaa tctgtgcttc atttttgtaa aacaaaaatg
7981 caacgcgaga gcgctaattt ttcaaacaaa gaatctgagc tgcatttta cagaacagaa
8041 atgcaacgcg agagcgctat tttaccaaca aagaatctat acttcttttt tgttctacaa
8101 aaatgcatcc cgagagcgct atttttctaa caaagcatct tagattactt tttttctcct
8161 ttgtgcgctc tataatgcag tctcttgata actttttgca ctgtaggtcc gttaaggtta
8221 gaagaaggct actttggtgt ctattttctc ttccataaaa aaagcctgac tccacttccc
8281 gcgtttactg attactagcg aagctgcggg tgcatttttt caagataaag gcatccccga
8341 ttatattcta taccgatgtg gattgcgcat actttgtgaa cagaaagtga tagcgttgat
8401 gattcttcat tggtcagaaa attatgaacg gtttcttcta ttttgtctct atatactacg
8461 tataggaaat gtttacattt tcgtattgtt ttcgattcac tctatgaata gttcttacta
8521 caatttttt gtctaaagag taatactaga gataaacata aaaaatgtag aggtcgagtt
8581 tagatgcaag ttcaaggagc gaaaggtgga tgggtaggtt atatagggat atagcacaga
8641 gatatatagc aaagagatac ttttgagcaa tgtttgtgga agcggtattc gcaatatttt
8701 agtagctcgt tacagtccgg tgcgtttttg gttttttgaa agtgcgtctt cagagcgctt
8761 ttggttttca aaagcgctct gaagttccta tactttctag agaataggaa cttcggaata
8821 ggaacttcaa agcgtttccg aaaacgagcg cttccgaaaa tgcaacgcga gctgcgcaca
8881 tacagctcac tgttcacgtc gcacctatat ctgcgtgttg cctgtatata tatatacatg
8941 agaagaacgg catagtgcgt gtttatgctt aaatgcgtac ttatatgcgt ctatttatgt
9001 aggatgaaag gtagtctagt acctcctgtg atattatccc attccatgcg gggtatcgta
9061 tgcttccttc agcactaccc tttagctgtt ctatatgctg ccactcctca attggattag
9121 tctcatcctt caatgctatc atttcctttg atattggatc atactaagaa accattatta
9181 tcatgacatt aacctataaa aataggcgta tcacgaggcc ctttcgtctc gcgcgtttcg
9241 gtgatgacgg tgaaaacctc tgacacatgc agctcccgga gacggtcaca gcttgtctgt
9301 aagcggatgc cgggagcaga caagcccgtc agggcgcgtc agcgggtgtt ggcgggtgtc
9361 ggggctggct taactatgcg gcatcagagc agattgtact gagagtgcac cataccacag
9421 cttttcaatt caattcatca ttttttttttt attcttttt ttgatttcgg tttctttgaa
9481 attttttga ttcggtaatc tccgaacaga aggaagaacg aaggaaggag cacagactta
9541 gattggtata tatacgcata tgtagtgttg aagaaacatg aaattgccca gtattcttaa
9601 cccaactgca cagaacaaaa acctgcagga aacgaagata aatcatgtcg aaagctacat
```

```
     9661 ataaggaacg tgctgctact catcctagtc ctgttgctgc caagctattt aatatcatgc
     9721 acgaaaagca aacaaacttg tgtgcttcat tggatgttcg taccaccaag gaattactgg
     9781 agttagttga agcattaggt cccaaaattt gtttactaaa aacacatgtg gatatcttga
     9841 ctgatttttc catggagggc acagttaagc cgctaaaggc attatccgcc aagtacaatt
     9901 ttttactctt cgaagacaga aaatttgctg acattggtaa tacagtcaaa ttgcagtact
     9961 ctgcgggtgt atacagaata gcagaatggg cagacattac gaatgcacac ggtgtggtgg
    10021 gcccaggtat tgttagcggt ttgaagcagg cggcagaaga agtaacaaag gaacctagag
    10081 gccttttgat gttagcagaa ttgtcatgca agggctccct atctactgga gaatatacta
    10141 agggtactgt tgacattgcg aagagcgaca aagattttgt tatcggcttt attgctcaaa
    10201 gagacatggg tggaagagat gaaggttacg attggttgat tatgacaccc ggtgtgggtt
    10261 tagatgacaa gggagacgca ttgggtcaac agtatagaac cgtggatgat gtggtctcta
    10321 caggatctga cattattatt gttggaagag gactatttgc aaagggaagg gatgctaagg
    10381 tagagggtga acgttacaga aaagcaggct gggaagcata tttgagaaga tgcggccagc
    10441 aaaactaaaa aactgtatta taagtaaatg catgtatact aaactcacaa attagagctt
    10501 caatttaatt atatcagtta ttaccctatg cggtgtgaaa taccgcacag atgcgtaagg
    10561 agaaaatacc gcatcaggaa attgtaaacg ttaatatttt gttaaaattc gcgttaaatt
    10621 tttgttaaat cagctcattt tttaaccaat aggccgaaat cggcaaaatc ccttataaat
    10681 caaaagaata gaccgagata gggttgagtg ttgttccagt ttggaacaag agtccactat
    10741 taaagaacgt ggactccaac gtcaaagggc gaaaaaccgt ctatcagggc gatggcccac
    10801 tacgtgaacc atcaccctaa tcaagtttt tggggtcgag gtgccgtaaa gcactaaatc
    10861 ggaaccctaa agggagcccc cgatttagag cttgacgggg aaagccggcg aacgtggcga
    10921 gaaaggaagg gaagaaagcg aaaggagcgg gcgctagggc gctggcaagt gtagcggtca
    10981 cgctgcgcgt aaccaccaca cccgccgcgc ttaatgcgcc gctacagggc gcgtcgcgcc
    11041 attcgccatt caggctgcgc aactgtt

//
```

Appendix B

B.1 Sequencing results for amplicons generated in Ble2AChr.SD strains using AP-PCR

z60.3 (a)>
tccacgggcggcatggacgagctgtacaagtgagtgagtgactgaccgcggaggtaagtgcgcgctgtct
gccggtgcccttctaccactgcttcggcatggtcatgggcaactcgcgcgcctgactgactgacgaagtt
cctattccgaagttcctattctctagaaagtataggaacttcgggacccacggatccattgggtcatcgg
atgccgggaccgacgagtgcagaggcgtgcaagcgagcttggcgcaatgtgttattaagttgtctaagcg
tcaatttgtttacaccacaatatatcctgccaccagccagccaacagctccccgaccggcagctcggcac
aaaatcaccactcgatacaggcagcccatcagtccgggacggcgtcagcgggagagccgttgtaaggcgg
cagactttgctcatgttaccgatgctattcggaagaacggcaactaagctgccgggtttgaaacacggat
gatctcgcggagggtagcatgttgattgta

z20.3 (a)>
actccacgggcggcatggacgagctgtacaagtgagtgagtgactgaccgcggaggtaagtgcgcgctgt
ctgccggtgcccttctaccactgcttcggcatggtcatgggcaactcgcgcgcctgactgactgacgaag
ttcctattccgaagttcctattctctagaaagtataggaacttcgggacccacggatccattgggtcatc
ggatgccgggaccgacgagtgcagaggcgtgcaagcgagcttggcgcaatgtgttattaagttgtctaag
cgtcaatttgtttacaccacaatatatcctgccaccagccagccaacagctccccgaccggcagctcggc
acaaaatcaccactcgatacaggcagcccatcagtccgggacggcgtcagcgggagagccgttgtaaggc
ggcagactttgctcatgttaccgatgctattcggaagaacggcaactaagctgccgggtttgaaacacgg
atgatctcgcggagggtagcatgttgattgtaacgatgacagagcgttgctgcctgtgatcaccgcggtt
tcaaaatcggctccgtcgatactatgttatacgccaactttgaaaacaactttgaaaaagctgttttctg
gtattta

z0.3 (a)>
ctcacgggcggcatggacgagctgtacaaggtgagtgagtgactgaccgcggaggtaagtgcgcgctgtc
tgccggtgcccttctaccactgcttcggcatggtcatgggcaactcgcgcgcctgactgactgacgaagt
tcctattccgaagttcctattctctagaaagtataggaacttcgggacccacggatccattgggtcatcg
gatgccgggaccgacgagtgcagaggcgtgcaagcgagcttggcgcaatgtgttattaagttgtctaagc
gtcaatttgtttacaccacaatatatcctgccaccagccagccaacagctccccgaccggcagctcggca
caaaatcaccactcgatacaggcagcccatcagtccgggacggcgtcagcgggagagccgttgtaaggcg
gcagactttgctcatgttaccgatgctattcggaagaacggcaactaagctgccgggtttgaaacacgga
tgatctcgcggagggtagcatgttgattgtaacgatgacagagcgttgctgcctgtgatcaccgcggttt
caaaatcggctccgtcgatactatgttatacgccaactttgaaaacaactttgaaaaagctgttttctgg
tatttaaggttttagaa

z80.1 (b)>
gtgatcaccg

z60.1 (b)>
ggtttcaaatcggctccgtcgatactatgttatacgccaactttgaaaacaactttgaaaaagctgtttt
cagg

z0.2 (b)>
tgtgatcaccgcggtttcaaatcggctccgtcgatactatgttatacgccaactttgaaaacaactttga
aaaagctgttttctggtatttaaggttttagaatgcaaggaacagtgaattggcctcgtctgcagccacg
tgcgggtacagtagattcaccagagagtggaacaggggtcctcgctgcgtggaatacgcggtaagatca
gcaatggcacagaaccccacgcgtctatattgacacccagcac

z60.3 (b)>
tgctgcctgtgatcaccgcggtttcaaatcggctccgtcgatactatgttatacgccaactttgaaaaca
actttgaaaaagctgttttctggtatttaaggttttagaatgcaaggaacagtgaattggtctttttttc
cgcccccaagggagaggtttgtgacgctacccttcgcaccggtgttgaagaaacccatgatttgatcgac
tgggctgctggtggtagacgaggtcgcagatatgctacttacggtacctgtacgcgtcggcccggtgcta
ttataatttatgtcgaaatggttcaaaagtgagcgtgcatacatgagcgtgcatgaatgattaccttaaa
cccacacatccccagaaacgcacaacaacaacatttaatctatataaataccaagtatactagaaaaatt
tcgagaaatgcactgagaatgttaacaagca

z20.2 (b)>
gtgatcaccgcggtttcaaaatcggctccgtcgatactatgttatacgccaactttgaaaacaactttga
aaaagctgttttctggtatttaaggttttagaatgcaaggatctcccaaggccacccctaccattccttg
ccccagcccgaacttcaactcggatcctgacaccagcattccacaccagactcgaccgcattaccggcgt
ggaaacgcggggaatccccactcgaagatctcttgttggtctttgtc

z20.3 (b)>
tgtgatcaccgcggtttcaaatcggctccgtcgatactatgttatacgccaactttgaaaacaactttga
aaaagctgttttctggtatttaaggttttagaatgcaaggaacagtgaattggacccatcttgacaaaaa
tag

z0.3 (b)>
gtgatcaccgcggtttcaaaatcggctccgtcgatactatgttatacgccaactttgaaaacaactttga
aaaagctgttttctggtatttaaggttttagaatgcaaggaacagtgaattggcgcaggcggtgtcttct
ctgtcctacaacttctggtggattgtgccgtcacaaatgactggggagggatcaacggcaatgccctcaa
gttttatcttggtttattttccattgctttcgacctgcttttcatgatacagcacttcgttctgtatccg
acgtctgcaactgcagtaacgccgtcgaaatggagtaaatgagttcgatcgaatctgactttccattctt
tgcggcggacaaaacggcaagtctgaggttttaaaggcgagaaatcacttttgttgtg

B.2 Sequencing results for amplicons generated in NgROS1 and NgROS2 using 3′RACE

ROS1 (a)>

gagggacggtatcgacgacaacaatgagattattcaatacgatccggttaacaagatgaccaggctcgtc
caagatacaggtttgacgaaatttcaaggaagtaatgcatttgcctttgacgaggtcaggaaccagatgt
tttggctctaccagggcgatgctaccaatcctgccggcttgtactattgggaccaggtcaccggaaccat
cgaccgtatcgccagccaagcgcaaacctgggacaatcaacgcttcccggcaaatgcggtgtattaccgc
gaccccagcactggcatatcatatttcgtttggatcactgagggtgggagcacggtcaatttcctcccca
tcacttatgatgcttccgggaatcctactggagtaggagcagatattcaacgaactatttcggggccttc
tttcagcccaagctttatgaggttcggcgatatagccgttcagacgagcaccaagcaattatatttggca
acaagcaacgggcgattctccaagattgacctcaccaacgcattcgggcaggccctgcttccttacacag
aaatcaagacgggcaatccctcccttcagcttgcattcgactgcgaagaaaatatcttgtacggtcag

ROS1 (b)>

agggacggtatcgacgacaacaatgagattattcaatacgatccggttaacaagatgaccaggctcgtcc
aaaaaaaaaaaaaacccc

ROS2 (a)>

gggatcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaag
gcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaa
aggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccctgacgagcatca
caaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccct
ggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctccctt
cgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaa
gctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgag
tccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggt
atgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttgg
tatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaacc
acccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaggatcct

ROS2 (b)>

agggatcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaa
ggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaa
aaggccaggaacgaaaaaaaaaaaaa

Bibliography

Aigner, H., Wilson, R.H., Bracher, A., Calisse, L., Bhat, J.Y., Hartl, F.U., and Hayer-Hartl, M. (2017). Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. Science (80-.). **358**: 1272–1278.

Ajjawi, I., Verruto, J., Aqui, M., Soriaga, L.B., Coppersmith, J., Kwok, K., Peach, L., Orchard, E., Kalb, R., Xu, W., Carlson, T.J., Francis, K., Konigsfeld, K., Bartalis, J., Schultz, A., Lambert, W., Schwartz, A.S., Brown, R., and Moellering, E.R. (2017). Lipid production in *Nannochloropsis gaditana* is doubled by decreasing expression of a single transcriptional regulator. Nat. Biotechnol. **35**: 647–652.

Alados, I., Foyo-Moreno, I., and Alados-Arboledas, L. (1996). Photosynthetically active radiation: measurements and modelling. Agric. For. Meteorol. **78**: 121–131.

Alipanah, L., Rohloff, J., Winge, P., Bones, A.M., and Brembu, T. (2015). Whole-cell response to nitrogen deprivation in the diatom *Phaeodactylum tricornutum*. J. Exp. Bot. **66**: 6281–6296.

Allen, J.F. and Martin, W. (2007). Out of thin air. Nature **445**: 610–612.

Allen, J.F., de Paula, W.B.M., Puthiyaveetil, S., and Nield, J. (2011). A structural phylogenetic map for chloroplast photosynthesis. Trends Plant Sci. **16**: 645–655.

Almaraz-Delgado, A.L., Flores-Uribe, J., Pérez-España, V.H., Salgado-Manjarrez, E., and Badillo-Corona, J.A. (2014). Production of therapeutic proteins in the chloroplast of *Chlamydomonas reinhardtii*. AMB Express **4**: 57.

Alonso, J.M. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science (80-.). **301**: 653–657.

Andersen, R.A., Brett, R.W., Potter, D., and Sexton, J.P. (1998). Phylogeny of the Eustigmatophyceae Based upon 18S rDNA, with Emphasis on *Nannochloropsis*. Protist **149**: 61–74.

Andersson, M.X. (2001). Chloroplast biogenesis. regulation of lipid transport to the thylakoid in chloroplasts isolated from expanding and fully expanded leaves of pea. Plant Physiol. **127**: 184–193.

Anila, N., Chandrashekar, A., Ravishankar, G.A., and Sarada, R. (2011). Establishment of *Agrobacterium tumefaciens*-mediated genetic transformation in *Dunaliella bardawil*. Eur. J. Phycol. **46**: 36–44.

Apt, K.E., Kroth-Pancic, P.G., and Grossman, A.R. (1996). Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. Mol. Gen. Genet. **252**: 572–579.

Arad, S. and Levy-Ontman, O. (2010). Red microalgal cell-wall polysaccharides: Biotechnological aspects. Curr. Opin. Biotechnol. **21**: 358–364.

Asada, K. (2006). Production and scavenging of reactive oxygen species in chloroplasts and their functions. Plant Physiol. **141**: 391–396.

Auchincloss, A.H., Loroch, A.I., and Rochaix, J.D. (1999). The argininosuccinate lyase gene of *Chlamydomonas reinhardtii*: Cloning of the cDNA and its characterization as a selectable shuttle marker. Mol. Gen. Genet. **261**: 21–30.

Baek, K., Lee, Y., Nam, O., Park, S., Sim, S.J., and Jin, E. (2016). Introducing *Dunaliella* LIP promoter containing light-inducible motifs improves transgenic expression in *Chlamydomonas reinhardtii*. Biotechnol. J. **11**: 384–392.

Bahrani-Mougeot, F.K., Buckles, E.L., Lockatell, C. V., Hebel, J.R., Johnson, D.E., Tang, C.M., and Donnenberg, M.S. (2002). Type 1 fimbriae and extracellular polysaccharides are preeminent uropathogenic *Escherichia coli* virulence determinants in the murine urinary tract. Mol. Microbiol. **45**: 1079–1093.

Baker, N.R. (2008). Chlorophyll fluorescence: a probe of photosynthesis *In Vivo*. Annu. Rev. Plant Biol. **59**: 89–113.

Bar-Even, A., Noor, E., Lewis, N.E., and Milo, R. (2010). Design and analysis of synthetic carbon fixation pathways . Proc. Natl. Acad. Sci. U. S. A. **107**: 8889–8894.

Barber, J., Nield, J., Morris, E., Zheleva, D., and Hankamer, B. (1997). The structure, function and dynamics of photosystem two. Physiol. Plant. **100**: 817–827.

Barbosa, M.J. and Wijffels, R.H. (2013). Biofuels from microalgae. Handb. Microalgal Cult. Appl. Phycol. Biotechnol. **24**: 566–577.

Barzegari, A., Hejazi, M.A., Hosseinzadeh, N., Eslami, S., Mehdizadeh Aghdam, E., and

Hejazi, M.S. (2010). *Dunaliella* as an attractive candidate for molecular farming. Mol. Biol. Rep. **37**: 3427–3430.

Beacham, T.A., Macia, V.M., Rooks, P., White, D.A., and Ali, S.T. (2015). Altered lipid accumulation in *Nannochloropsis salina* CCAP849/3 following EMS and UV induced mutagenesis. Biotechnol. Reports **7**: 87–94.

Bekker, A., Holland, H.D., Wang, P.-L., Rumble, D., Stein, H.J., Hannah, J.L., Coetzee, L.L., and Beukes, N.J. (2004). Dating the rise of atmospheric oxygen. Nature **427**: 117–120.

Ben-Amotz, A. and Avron, M. (1972). Photosynthetic activities of the halophilic alga *Dunaliella parva*. Plant Physiol. **49**: 240–243.

Ben-Shem, A., Frolow, F., and Nelson, N. (2003). Crystal structure of plant photosystem I. Nature **426**: 630–635.

Berepiki, A., Hitchcock, A., Moore, C.M., and Bibby, T.S. (2016). Tapping the unused potential of photosynthesis with a heterologous electron sink. ACS Synth. Biol.: acssynbio.6b00100.

Berepiki, A., Lichius, A., and Read, N.D. (2011). Actin organization and dynamics in filamentous fungi. Nat. Rev. Microbiol. **9**: 876–887.

Berepiki, A., Lichius, A., Shoji, J.Y., Tilsner, J., and Read, N.D. (2010). F-actin dynamics in *Neurospora crassa*. Eukaryot. Cell **9**: 547–557.

Berges, J. (1997). Miniview: algal nitrate reductases. Eur. J. Phycol. **32**: 3–8.

Bitog, J.P., Lee, I.B., Lee, C.G., Kim, K.S., Hwang, H.S., Hong, S.W., Seo, I.H., Kwon, K.S., and Mostafa, E. (2011). Application of computational fluid dynamics for modeling and designing photobioreactors for microalgae production: A review. Comput. Electron. Agric. **76**: 131–147.

Björkman, O. and Demmig, B. (1987). Photon yield of O2 evolution and chlorophyll fluorescence characteristics at 77 K among vascular plants of diverse origins. Planta **170**: 489–504.

Björn, L.O. and Govindjee (2015). The evolution of photosynthesis and its environmental impact. In Photobiology: The Science of Light and Life, Third Edition, pp. 207–230.

Blankenship, J.E. and Kindle, K.L. (1992). Expression of chimeric genes by the light-regulated cabII-1 promoter in *Chlamydomonas reinhardtii*: a *cabII-1/nit1* gene functions

as a dominant selectable marker in a *nit1- nit2-* strain. Mol. Cell. Biol. **12**: 5268–5279.

Blankenship, R.E., Tiede, D.M., Barber, J., Brudvig, G.W., Fleming, G., Ghirardi, M. Gunner, M.R., Junge, W., Kramer, D.M., Melis, A., Moore, T.A., Moser, C.C., Nocera, D.G., Nozik, A.J., Ort, D.R., Parson, W.W., Prince, R.C., and Sayre, R.T. (2011). Comparing photosynthetic and photovoltaic efficiencies and recognizing the potential for improvement. Science **332**: 805–9.

Blankenship, R.E. (2008). Molecular mechanisms of photosynthesis.

Blanvillain, R. and Gallois, P. (2008). Promoter trapping system to study embryogenesis. Methods Mol. Biol. **427**: 121–35.

Blowers, A.D. (1989). Studies on *Chlamydomonas* chloroplast transformation: foreign DNA can be stably maintained in the chromosome. Plant Cell Online **1**: 123–132.

Bolton, J.R. (1977). Photochemical conversion and storage of solar energy. J. Solid State Chem. **22**: 3–8.

Borowitzka, M.A. (1999). Commercial production of microalgae: ponds, tanks, tubes and fermenters. In Journal of Biotechnology, pp. 313–321.

Borowitzka, M.A. (2013). High-value products from microalgae-their development and commercialisation. J. Appl. Phycol. **25**: 743–756.

Borowitzka, M.A. and Moheimani, N.R. (2013). Sustainable biofuels from algae. Mitig. Adapt. Strateg. Glob. Chang. **18**: 13–25.

Borowitzka, M.A. and Siva, C.J. (2007). The taxonomy of the genus *Dunaliella* (Chlorophyta, Dunaliellales) with emphasis on the marine and halophilic species. J. Appl. Phycol. **19**: 567–590.

Boussiba, S., Vonshak, A., Cohen, Z., Avissar, Y., and Richmond, A. (1987). Lipid and biomass production by the halotolerant microalga *Nannochloropsis salina*. Biomass **12**: 37–47.

Bower, N.I. and Johnston, I.A. (2010). Targeted rapid amplification of cDNA ends (T-RACE) - An improved RACE reaction through degradation of non-target sequences. Nucleic Acids Res. **38**.

Boyer, P.D. (1997). The ATP synthase - a splendid molecular machine. Annu. Rev. Biochem **66**: 717–49.

Brown, L.E., Sprecher, S.L., and Keller, L.R. (1991). Introduction of exogenous DNA into

Chlamydomonas reinhardtii by electroporation. Mol. Cell. Biol. **11**: 2328–32.

Bryant, D.A. and Frigaard, N.U. (2006). Prokaryotic photosynthesis and phototrophy illuminated. Trends Microbiol. **14**: 488–496.

Bundock, P., den Dulk-Ras, A., Beijersbergen, A., and Hooykaas, P.J. (1995). Trans-kingdom T-DNA transfer from *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*. EMBO J **14**: 3206–3214.

Burall, L.S., Harro, J.M., Li, X., Lockatell, C.V., Himpsl, S.D., Hebel, J.R., Johnson, D.E., and Mobley, H.L.T. (2004). *Proteus mirabilis* genes that contribute to pathogenesis of urinary tract infection: identification of 25 signature-tagged mutants attenuated at least 100-Fold. Infect. Immun. **72**: 2922–2938.

Burg, D., Wielinga, P., Zelcer, N., Saeki, T., Mulder, G.J., and Borst, P. (2002). Inhibition of the multidrug resistance protein 1 (MRP1) by peptidomimetic glutathione-conjugate analogs. Mol. Pharmacol. **62**: 1160–6.

Busch, A. and Hippler, M. (2011). The structure and function of eukaryotic photosystem I. Biochim. Biophys. Acta - Bioenerg. **1807**: 864–877.

Buzas, D.M., Lohar, D., Sato, S., Nakamura, Y., Tabata, S., Vickers, C.E., Stiller, J., and Gresshoff, P.M. (2005). Promoter trapping in *Lotus japonicus* reveals novel root and nodule GUS expression domains. Plant Cell Physiol. **46**: 1202–1212.

Carr, S., Penfold, C.N., Bamford, V., James, R., and Hemmings, A.M. (2000). The structure of TolB, an essential component of the tol-dependent translocation system, and its protein-protein interaction with the translocation domain of colicin E9. Structure **8**: 57–66.

Castro, F.A. V, Herdeiro, R.S., Panek, A.D., Eleutherio, E.C.A., and Pereira, M.D. (2007). Menadione stress in *Saccharomyces cerevisiae* strains deficient in the glutathione transferases. Biochim. Biophys. Acta - Gen. Subj. **1770**: 213–220.

Cerutti, H. (1997). Epigenetic silencing of a foreign gene in nuclear transformants of *Chlamydomonas*. Plant Cell Online **9**: 925–945.

Cerutti, H., Ma, X., Msanne, J., and Repas, T. (2011). RNA-mediated silencing in algae: biological roles and tools for analysis of gene function. Eukaryot. Cell **10**: 1164–1172.

Chen, L., Doi, M., Durley, R.C.E., Chistoserdov, A.Y., Lidstrom, M.E., Davidson, V.L., and Mathews, F.S. (1998). Refined crystal structure of methylamine dehydrogenase from *Paracoccus denitrificans* at 1.75 Å resolution. J. Mol. Biol. **276**: 131–149.

Chen, M., Schliep, M., Willows, R.D., Cai, Z.-L., Neilan, B.A., and Scheer, H. (2010). A Red-Shifted Chlorophyll. Science (80-.). **329**: 1318–1319.

Chen, X. and Zhang, J. (2016). The genomic landscape of position effects on protein expression level and noise in yeast. Cell Syst. **2**: 347–354.

Cheney, D., Metz, B., and Stiller, J. (2008). *Agrobacterium*-mediated genetic transformation in the macroscopic marine red alga *Porphyra Yezoensis*. J. Phycol. **37**: 11–11.

Cheng, Q., Hallmann, A., Edwards, L., and Miller, S.M. (2006). Characterization of a heat-shock-inducible *hsp70* gene of the green alga *Volvox carteri*. Gene **371**: 112–120.

Cheng, R., Ma, R., Li, K., Rong, H., Lin, X., Wang, Z., Yang, S., and Ma, Y. (2012). *Agrobacterium tumefaciens* mediated transformation of marine microalgae *Schizochytrium*. Microbiol. Res. **167**: 179–186.

Chernet, B. (2010). The world of the cell 8th ed. (Pearson).

Corteggiani Carpinelli, E., Telatin, A., Vitulo, N., Forcato, C., D'Angelo, M., Schiavon, R., Vezzi, A., Giacometti, G.M., Morosinotto, T., and Valle, G. (2014). Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion. Mol. Plant **7**: 323–335.

Cotton, C.A.R., Douglass, J.S., De Causmaecker, S., Brinkert, K., Cardona, T., Fantuzzi, A., Rutherford, A.W., and Murray, J.W. (2015). Photosynthetic constraints on fuel from microbes. Front. Bioeng. Biotechnol. **3**: 36.

Criddle, D.N., Gillies, S., Baumgartner-Wilson, H.K., Jaffar, M., Chinje, E.C., Passmore, S., Chvanov, M., Barrow, S., Gerasimenko, O. V., Tepikin, A. V., Sutton, R., and Petersen, O.H. (2006). Menadione-induced reactive oxygen species generation via redox cycling promotes apoptosis of murine pancreatic acinar cells. J. Biol. Chem. **281**: 40485–40492.

Croce, R. and Van Amerongen, H. (2014). Natural strategies for photosynthetic light harvesting. Nat. Chem. Biol. **10**: 492–501.

Das, S., Noe, J.C., Paik, S., and Kitten, T. (2005). An improved arbitrary primed PCR method for rapid characterization of transposon insertion sites. J. Microbiol. Methods **63**: 89–94.

Datta, S., Ikeda, T., Kano, K., and Mathews, F.S. (2003). Structure of the phenylhydrazine adduct of the quinohemoprotein amine dehydrogenase from *Paracoccus denitrificans* at 1.7 Å resolution. Acta Crystallogr. - Sect. D Biol. Crystallogr. **59**: 1551–1556.

Dawson, H.N., Burlingame, R., and Cannons, A.C. (1997). Stable transformation of *Chlorella*: rescue of nitrate reductase-deficient mutants with the nitrate reductase gene. Curr. Microbiol. **35**: 356–362.

Day, A., Debuchy, R., van Dillewijn, J., Purton, S., and Rochaix, J.-D. (1990). Studies on the maintenance and expression of cloned DNA fragments in the nuclear genome of the green alga Chlamydomonas reinhardtii. Physiol. Plant. **78**: 254–260.

Dean, A.P., Sigee, D.C., Estrada, B., and Pittman, J.K. (2010). Using FTIR spectroscopy for rapid determination of lipid accumulation in response to nitrogen limitation in freshwater microalgae. Bioresour. Technol. **101**: 4499–4507.

Demirbas, M.F. (2011). Biofuels from algae for sustainable development. Appl. Energy **88**: 3473–3480.

Dent, R.M. (2005). Functional genomics of eukaryotic photosynthesis using insertional mutagenesis of *Chlamydomonas reinhardtii*. Plant Physiol. **137**: 545–556.

Depauw, F.A., Rogato, A., D'Alcalá, M.R., and Falciatore, A. (2012). Exploring the molecular basis of responses to light in marine diatoms. J. Exp. Bot. **63**: 1575–1591.

Díaz-Santos, E., de la Vega, M., Vila, M., Vigara, J., and León, R. (2013). Efficiency of different heterologous promoters in the unicellular microalga *Chlamydomonas reinhardtii*. Biotechnol. Prog. **29**: 319–328.

Diaz, M., Mora, V., Pedrozo, F., Nichela, D., and Baffico, G. (2014). Evaluation of native acidophilic algae species as potential indicators of polycyclic aromatic hydrocarbon (PAH) soil contamination. J. Appl. Phycol. **27**: 321–325.

Ding, W.T., Zhang, G.C., and Liu, J.J. (2013). 3′ truncation of the *GPD1* promoter in Saccharomyces cerevisiae for improved ethanol yield and productivity. Appl. Environ. Microbiol. **79**: 3273–3281.

Doan, T.T.Y. and Obbard, J.P. (2012). Enhanced intracellular lipid in *Nannochloropsis* sp. via random mutagenesis and flow cytometric cell sorting. Algal Res. **1**: 17–21.

Doron, L., Segal, N., and Shapira, M. (2016). Transgene expression in microalgae - from tools to applications. Front. Plant Sci. **7**: 505.

Dou, Z., Heinhorst, S., Williams, E.B., Murin, C.D., Shively, J.M., and Cannon, G.C. (2008). CO_2 fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO_2. J. Biol. Chem. **283**: 10377–10384.

Dubinsky, Z. (2013). Photosynthesis (InTech 2013).

Ducat, D.C., Avelar-Rivas, J.A., Way, J.C., and Silvera, P.A. (2012). Rerouting carbon flux to enhance photosynthetic productivity. Appl. Environ. Microbiol. **78**: 2660–2668.

Dunahay, T.G., Jarvis, E.E., and Roessler, P.G. (1995). Genetic Transformation of the Diatoms *Cyclotella cryptica* and *Navicula saprophila*. J. Phycol. **31**: 1004–1012.

Durão, P., Aigner, H., Nagy, P., Mueller-Cajar, O., Hartl, F.U., and Hayer-Hartl, M. (2015). Opposing effects of folding and assembly chaperones on evolvability of Rubisco. Nat. Chem. Biol. **11**: 148–155.

Ebnet, E. (1999). Volvoxrhodopsin, a light-regulated sensory photoreceptor of the spheroidal green alga *Volvox carteri*. Plant Cell Online **11**: 1473–1484.

Eichler-Stahlberg, A., Weisheit, W., Ruecker, O., and Heitzer, M. (2009). Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. Planta **229**: 873–883.

Erb, T.J. (2011). Carboxylases in natural and synthetic microbial pathways. Appl. Environ. Microbiol. **77**: 8466–8477.

Erickson, E., Wakao, S., and Niyogi, K.K. (2015). Light stress and photoprotection in *Chlamydomonas reinhardtii*. Plant J. **82**: 449–465.

Evans, C., Malin, G., Mills, G.P., and Wilson, W.H. (2006). Viral infection of *Emiliania huxleyi* (Prymnesiophyceae) leads to elevated production of reactive oxygen species. J. Phycol. **42**: 1040–1047.

Fabian, M.R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. Annu. Rev. Biochem. **79**: 351–379.

Fábregas, J., Maseda, A., Domínguez, A., Ferreira, M., and Otero, A. (2002). Changes in the cell composition of the marine microalga, *Nannochloropsis gaditana*, during a light : dark cycle. Biotechnol. Lett. **24**: 1699–1703.

Fahlbusch, K.-G. (2003). Ullmann's encyclopedia of industrial chemistry. In Ullmann's Encyclopedia of Industrial Chemistry, pp. 73–198.

Falciatore, A., Casotti, R., Leblanc, C., Abrescia, C., and Bowler, C. (1999). Transformation of nonselectable reporter genes in marine diatoms. Mar. Biotechnol. (NY). **1**: 239–251.

Falkowski, P.G., Barber, R.T., and Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. Science (80-.). **281**: 200–206.

Falkowski, P.G., Katz, M.E., Knoll, A.H., Quigg, A., Raven, J. a, Schofield, O., and Taylor, F.J.R. (2004). The Evolution of modern eukaryotic phytoplankton. Science (80-.). **305**: 354–360.

Falkowski, P.G. and Raven, J. a (1997). Aquatic Photosynthesis. Freshw. Biol. **69**: 375.

Fallis, A.. (1975). Culture of marine invertebrate animals. In Journal of Chemical Information and Modeling, pp. 1689–1699.

Fang, S.C., De Los Reyes, C., and Umen, J.G. (2006). Cell size checkpoint control by the retinoblastoma tumor suppressor pathway. PLoS Genet. **2**: 1565–1579.

Farquhar, G.D., von Caemmerer S, and Berry, J.A. (2001). Models of photosynthesis. Plant Physiol. **125**: 42–45.

Fawley, K.P. and Fawley, M.W. (2007). Observations on the diversity and ecology of freshwater *Nannochloropsis* (Eustigmatophyceae), with descriptions of new taxa. Protist **158**: 325–336.

Feller, U., Anders, I., and Mae, T. (2008). Rubiscolytics: fate of Rubisco after its enzymatic function in a cell is terminated. In Journal of Experimental Botany, pp. 1615–1624.

Fenchel, T. (2008). The microbial loop - 25 years later. J. Exp. Mar. Bio. Ecol. **366**: 99–103.

Feng, S., Li, X., Xu, Z., and Qi, J. (2014). *Dunaliella salina* as a novel host for the production of recombinant proteins. Appl. Microbiol. Biotechnol. **98**: 4293–4300.

Feng, S., Xue, L., Liu, H., and Lu, P. (2009). Improvement of efficiency of genetic transformation for *Dunaliella salina* by glass beads method. Mol. Biol. Rep. **36**: 1433–1439.

Fernández, E. and Galván, A. (2008). Nitrate assimilation in *Chlamydomonas*. Eukaryot. Cell **7**: 555–559.

Fernández, E., Schnell, R., Ranum, L.P., Hussey, S.C., Silflow, C.D., and Lefebvre, P. a (1989). Isolation and characterization of the nitrate reductase structural gene of *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. U. S. A. **86**: 6449–53.

Ferrante, P., Diener, D.R., Rosenbaum, J.L., and Giuliano, G. (2011). Nickel and low CO_2-controlled motility in *Chlamydomonas* through complementation of a paralyzed flagella mutant with chemically regulated promoters. BMC Plant Biol. **11**: 22.

Ferreira, K.N., Iverson, T.M., Maghlaoui, K., Barber, J., and Iwata, S. (2004). Architecture of the photosynthetic oxygen-evolving centre. Science (80-.). **303**: 1831–1839.

Finazzi, G., Moreau, H., and Bowler, C. (2010). Genomic insights into photosynthesis in eukaryotic phytoplankton. Trends Plant Sci. **15**: 565–572.

Finn, R.D. Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.Y., Dosztanyi, Z., El-Gebali, S., Fraser, M., Gough, J., Haft, D., Holliday, G.L., Huang, H., Huang, X., Letunic, I., Lopez, R., Lu, S., Marchler-Bauer, A., Mi, H., Mistry, J., Natale, D.A., Necci, M., Nuka, G., Orengo, C.A., Park, Y., Pesseat, S., Piovesan, D., Potter, S.C., Rawlings, N.D., Redaschi, N., Richardson, L., Rivoire, C., Sangrador-Vegas, A., Sigrist, C., Sillitoe, I., Smithers, B., Squizzato, S., Sutton, G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Xenarios, I., Yeh, L.S., Young, S.Y., and Mitchell, A.L. (2017). InterPro in 2017 - beyond protein family and domain annotations. Nucleic Acids Res. **45**: D190–D199.

Flores, H.S., Wikfors, G.H., and Dam, H.G. (2012). Reactive oxygen species are linked to the toxicity of the dinoflagellate *Alexandrium* spp. to protists. Aquat. Microb. Ecol. **66**: 199–209.

Fon Sing, S., Isdepsky, A., Borowitzka, M.A., and Moheimani, N.R. (2013). Production of biofuels from microalgae. Mitig. Adapt. Strateg. Glob. Chang. **18**: 47–72.

Forde, B.G. (2000). Nitrate transporters in plants: structure, function and regulation. Biochim. Biophys. Acta - Biomembr. **1465**: 219–235.

Franklin, S., Ngo, B., Efuet, E., and Mayfield, S.P. (2002). Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast. Plant J. **30**: 733–744.

Franz, A.K., Danielewicz, M.A., Wong, D.M., Anderson, L.A., and Boothe, J.R. (2013). Phenotypic screening with oleaginous microalgae reveals modulators of lipid productivity. ACS Chem. Biol. **8**: 1053–1062.

Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci U S A **85**: 8998–9002.

Fromme, P., Jordan, P., and Krauss, N. (2001). Structure of photosystem I. Biochim. Biophys. Acta - Bioenerg. **1507**: 5–31.

Fuentes, J.L., Garbayo, I., Cuaresma, M., Montero, Z., González-Del-Valle, M., and Vílchez, C. (2016). Impact of microalgae-bacteria interactions on the production of algal biomass and associated compounds. Mar. Drugs **14**.

Fulbright, S.P., Robbins-Pianka, A., Berg-Lyons, D., Knight, R., Reardon, K.F., and

Chisholm, S.T. (2018). Bacterial community changes in an industrial algae production system. Algal Res. **31**: 147–156.

Galdzicki, M., Clancy, K.P., Oberortner, E., Pocock, M., Quinn, J.Y., Rodriguez, C.A., Roehner, N., Wilson, M.L., Adam, L., Anderson, J.C., Bartley, B.A., Beal, J., Chandran, D., Chen, J., Densmore, D., Endy, D., Grünberg, R., Hallinan, J., Hillson, N.J. Johnson, J.D., Kuchinsky, A., Lux, M., Misirli, G., Peccoud, J., Plahar, H.A., Sirin, E., Stan, G.B., Villalobos, A., Wipat, A., Gennari, J.H., Myers, C.J., and Sauro, H.M. (2014). The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. Nat. Biotechnol. **32**: 545–550.

Galván, A. and Fernández, E. (2001). Eukaryotic nitrate and nitrite transporters. Cell. Mol. Life Sci. **58**: 225–233.

Galván, A., González-Ballester, D., and Fernández, E. (2007). Insertional mutagenesis as a tool to study genes/functions in *Chlamydomonas*. Adv. Exp. Med. Biol. **616**: 77–89.

Gan, F., Zhang, S., Rockwell, N.C., Martin, S.S., Lagarias, C., and Bryant, D.A. (2014). Extensive remodeling of a cyanobacterial photosynthetic apparatus in far-red light. Science (80-.). **345**: 1–11.

Gauthier, E.R., Madison, S.D., and Michel, R.N. (1997). Rapid RNA isolation without the use of commercial kits: application to small tissue samples. Pflugers Arch. Eur. J. Physiol. **433**: 664–668.

Geng, D., Wang, Y., Wang, P., Li, W., and Sun, Y. (2003). Stable expression of hepatitis B surface antigen gene in *Dunaliella salina* (Chlorophyta). J. Appl. Phycol. **15**: 451–456.

Georgianna, D.R., Hannon, M.J., Marcuschi, M., Wu, S., Botsch, K., Lewis, A.J., Hyun, J., Mendez, M., and Mayfield, S.P. (2013). Production of recombinant enzymes in the marine alga *Dunaliella tertiolecta*. Algal Res. **2**: 2–9.

Ghildiyal, M. and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat. Rev. Genet. **10**: 94–108.

van der Giezen, M. (2011). Mitochondria and the rise of eukaryotes. Bioscience **61**: 594–601.

Giordano, M., Kansiz, M., Heraud, P., Beardall, J., Wood, B., and McNaughton, D. (2001). Fourier Transform Infrared Spectroscopy as a novel tool to investigate changes in intracellular macromolecular pools in the marine microalga *Chaetocerus muellerii* (Bacilliariophyceae). J. Phycol. **37**: 271–279.

Gonzalez-Ballester, D., Pootakham, W., Mus, F., Yang, W., Catalanotti, C., Magneschi, L.,

de Montaigu, A., Higuera, J.J., Prior, M., Galvan, A., Fernandez, E., and Grossman, A.R. (2011). Reverse genetics in *Chlamydomonas*: a platform for isolating insertional mutants. Plant Methods **7**: 24.

Goyal, A. (2007). Osmoregulation in *Dunaliella*, Part II: Photosynthesis and starch contribute carbon for glycerol synthesis during a salt stress in *Dunaliella tertiolecta*. Plant Physiol. Biochem. **45**: 705–710.

Goyal, A. and Gimmler, H. (1989). Osmoregulation in *Dunaliella tertiolecta*: effects of salt stress, and the external pH on the internal pH. Arch. Microbiol. **152**: 138–142.

Green, B.R. and Parson, W.W. (2003). Light-harvesting antennas in photosynthesis.

Griffiths, H. (2006). Plant biology: designs on rubisco. Nature **441**: 940–941.

Griffiths, M.J. and Harrison, S.T.L. (2009). Lipid productivity as a key characteristic for choosing algal species for biodiesel production. J. Appl. Phycol. **21**: 493–507.

Grossman, A.R., Bhaya, D., Apt, K.E., and Kehoe, D.M. (1995). Light-harvesting complexes in oxygenic photosynthesis: diversity, control, and evolution. Annu. Rev. Genet. **29**: 231–88.

Guillard, R.R.L. and Ryther, J.H. (1962). Studies of marine planktonic diatoms: *I. Cyclotella nana* Hustedt, and *Detonula confervacea* (cleve) Gran. Can. J. Microbiol. **8**: 229–239.

Guiry, M.D. (2012). How many species of algae are there? J. Phycol. **48**: 1057–1063.

Guschina, I.A. and Harwood, J.L. (2006). Lipids and lipid metabolism in eukaryotic algae. Prog. Lipid Res. **45**: 160–186.

Guskov, A., Kern, J., Gabdulkhakov, A., Broser, M., Zouni, A., and Saenger, W. (2009). Cyanobacterial photosystem II at 2.9-A resolution and the role of quinones, lipids, channels and chloride. Nat. Struct. Mol. Biol. **16**: 334–342.

Hagedorn, M., Carter, V.L., and Voolstra, C.R. (2015). Seasonal preservation success of the marine dinoflagellate coral symbiont, *Symbiodinium* sp. PLoS One **10**.

Hakala, M., Rantamäki, S., Puputti, E.M., Tyystjärvi, T., and Tyystjärvi, E. (2006). Photoinhibition of manganese enzymes: insights into the mechanism of photosystem II photoinhibition. In Journal of Experimental Botany, pp. 1809–1816.

Hallmann, A., Rappel, A., and Sumper, M. (1997). Gene replacement by homologous recombination in the multicellular green alga *Volvox carteri*. Proc Natl Acad Sci U S A **94**: 7469–7474.

Hambourger, M., Moore, G.F., Kramer, D.M., Gust, D., Moore, A.L., and Moore, T.A. (2009). Biology and technology for photochemical fuel production. Chem. Soc. Rev. **38**: 25–35.

Hankamer, B., Morris, E., Nield, J., Gerle, C., and Barber, J. (2001). Three-dimensional structure of the photosystem II core dimer of higher plants determined by electron microscopy. J. Struct. Biol. **135**: 262–269.

Haring, M.A. and Beck, C.F. (1997). A promoter trap for *Chlamydomonas reinhardtii*: development of a gene cloning method using 5' RACE-based probes. Plant J. **11**: 1341–1348.

Harper, M. and Lee, C.J. (2012). Genome-wide analysis of mutagenesis bias and context sensitivity of N-methyl-N'-nitro-N-nitrosoguanidine (NTG). Mutat. Res. - Fundam. Mol. Mech. Mutagen. **731**: 64–67.

Harris, E.H. (2001). *Chlamydomonas* as a model organism. Annu. Rev. Plant Physiol. Plant Mol. Biol. **52**: 363–406.

Hartley, T., Stewart, C., Stewart, R., and Munroe, D.J. (2009). Cost-effective addition of high-throughput colony picking capability to a standard liquid-handling platform. J. Lab. Autom. **14**: 22–26.

Hasan, S.S. and Cramer, W.A. (2014). Internal lipid architecture of the hetero-oligomeric cytochrome b_6f complex. Structure **22**: 1008–1015.

Hasan, S.S., Yamashita, E., Baniulis, D., and Cramer, W.A. (2013). Quinone-dependent proton transfer pathways in the photosynthetic cytochrome b_6f complex. Proc. Natl. Acad. Sci. U. S. A. **110**: 4297–302.

Heitzer, M., Eckert, A., Fuhrmann, M., and Griesbeck, C. (2007). Influence of codon bias on the expression of foreign genes in microalgae. Adv. Exp. Med. Biol. **616**: 46–53.

Hellens, R., Mullineaux, P., and Klee, H. (2000). A guide to *Agrobacterium binary* Ti vectors. Trends Plant Sci. **5**: 446–451.

Helliwell, K.E., Scaife, M.A., Sasso, S., Araujo, A.P.U., Purton, S., and Smith, A.G. (2014). Unraveling vitamin B12-responsive gene regulation in algae. Plant Physiol. **165**: 388–397.

Hemminger, W., Sarge, S.M., and Bundesanstalt, P. (1998). Post-targeting functions of signal peptides. Protein Transp. into Endoplasmic Reticulum **1**: 1–16.

von der Heyde, E.L., Klein, B., Abram, L., and Hallmann, A. (2015). The inducible *nitA*

promoter provides a powerful molecular switch for transgene expression in *Volvox carteri*. BMC Biotechnol.: 1–13.

Hibberd, D.J. (1981). Notes on the taxonomy and nomenclature of the algal classes Eustigmatophyceae and Tribophyceae (synonym Xanthophyceae). Bot. J. Linn. Soc. **82**: 93–119.

Hill, R. (1937). Oxygen evolved by isolated chloroplasts. Nature **139**: 881–882.

Hlavova, M., Turoczy, Z., and Bisova, K. (2015). Improving microalgae for biotechnology - from genetics to synthetic biology. Biotechnol. Adv. **33**: 1194–1203.

Holt, N.E., Fleming, G.R., and Niyogi, K.K. (2004). Toward an understanding of the mechanism of nonphotochemical quenching in green plants. Biochemistry **43**: 8281–8289.

Horton, P., Ruban, A. V., and Walters, R.G. (1996). Regulation of light harvesting in green plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **47**: 655–684.

Hosseini Tafreshi, A. and Shariati, M. (2009). *Dunaliella* biotechnology: methods and applications. J. Appl. Microbiol. **107**: 14–35.

Hounslow, E., Noirel, J., Gilmour, D.J., and Wright, P.C. (2017). Lipid quantification techniques for screening oleaginous species of microalgae for biofuel production. Eur. J. Lipid Sci. Technol. **119**: 1–24.

Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M., and Darzins, A. (2008). Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. Plant J. **54**: 621–639.

Ikehata, H. and Ono, T. (2011). The mechanisms of UV mutagenesis. J. Radiat. Res. **52**: 115–125.

Imramovsky, A., Pesko, M., Ferriz, J.M., Kralova, K., Vinsova, J., and Jampilek, J. (2011). Photosynthesis - Inhibiting efficiency of 4-chloro-2-(chlorophenylcarbamoyl)phenyl alkylcarbamates. Bioorganic Med. Chem. Lett. **21**: 4564–4567.

Iseki, M., Matsunaga, S., Murakami, A., Ohno, K., Shiga, K., Yoshida, K., Sugai, M., Takahashi, T., Hori, T., and Watanabe, M. (2002). A blue-light-activated adenylyl cyclase mediates photoavoidance in *Euglena gracilis*. Nature **415**: 1047–1051.

Ishimatsu, A., Oda, T., Yoshida, M., and Ozaki, M. (1996). Oxygen radicals are probably involved in the mortality of yellowtail by *Chattonella marina*. Fish. Sci. **62(5)**: 836–837.

Jarvis, E.E. and Brown, L.M. (1991). Transient expression of firefly luciferase in protoplasts of the green alga *Chlorella ellipsoidea*. Curr. Genet. **19**: 317–321.

De Jesus Raposo, M.F., De Morais, R.M.S.C., and De Morais, A.M.M.B. (2013). Health applications of bioactive compounds from marine microalgae. Life Sci. **93**: 479–486.

Jia, Y., Li, S., Allen, G., Feng, S., and Xue, L. (2012). A novel glyceraldehyde-3-phosphate dehydrogenase (GAPDH) promoter for expressing transgenes in the halotolerant alga *Dunaliella salina*. Curr. Microbiol. **64**: 506–513.

Jia, Y., Xue, L., Liu, H., and Li, J. (2009). Characterization of the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene from the halotolerant alga *Dunaliella salina* and inhibition of its expression by RNAi. Curr. Microbiol. **58**: 426–431.

Jiang, G., Lü, Y., Niu, X., and Xue, L. (2005). The actin gene promoter-driven bar as a dominant selectable marker for nuclear transformation of *Dunaliella salina*. Yi Chuan Xue Bao **32**: 424–433.

Jiang, W., Brueggeman, A.J., Horken, K.M., Plucinak, T.M., and Weeks, D.P. (2014). Successful transient expression of Cas9 and single guide RNA genes in *Chlamydomonas reinhardtii*. Eukaryot. Cell **13**: 1465–1469.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (80-.). **337**: 816–821.

Jing, H., Takagi, J., Liu, J., and Springer, T.A. (2002). Archaeal surface layer proteins contain β propeller, PKD, and β helix domains and are related to metazoan cell surface proteins. Structure **10**: 1453–1464.

Jinkerson, R.E., Radakovits, R., and Posewitz, M.C. (2013). Genomic insights from the oleaginous model alga *Nannochloropsis gaditana*. Bioengineered **4**: 37–41.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y. Lopez, R., and Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics **30**: 1236–1240.

Junge, W. and Nelson, N. (2015). ATP synthase. Annu. Rev. Biochem. **84**: 631–657.

Jungnick, N., Ma, Y., Mukherjee, B., Cronan, J.C., Speed, D.J., Laborde, S.M., Longstreth, D.J., and Moroney, J. V. (2014). The carbon concentrating mechanism in *Chlamydomonas reinhardtii*: finding the missing pieces. In Photosynthesis Research, pp.

159–173.

Kandilian, R., Lee, E., and Pilon, L. (2013). Radiation and optical properties of *Nannochloropsis oculata* grown under different irradiances and spectra. Bioresour. Technol. **137**: 63–73.

Kaplan, D., Heimer, Y.M., Abeliovich, A., and Goldsbrough, P.B. (1995). Cadmium toxicity and resistance in *Chlorella* sp. Plant Sci. **109**: 129–137.

Kato, Y., Ozawa, S.I., Takahashi, Y., and Sakamoto, W. (2015). D1 fragmentation in photosystem II repair caused by photo-damage of a two-step model. Photosynth. Res. **126**: 409–416.

Kiechle, M., Manivasakam, P., Eckardt-Schupp, F., Schiestl, R.H., and Friedl, A. a (2002). Promoter-trapping in *Saccharomyces cerevisiae* by radiation-assisted fragment insertion. Nucleic Acids Res. **30**: e136.

Kilian, O., Benemann, C.S.E., Niyogi, K.K., and Vick, B. (2011). High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. Proc. Natl. Acad. Sci. **108**: 21265–21269.

Kim, D.-H., Kim, Y.T., Cho, J.J., Bae, J.-H., Hur, S.-B., Hwang, I., and Choi, T.-J. (2002). Stable integration and functional expression of flounder growth hormone gene in transformed microalga, *Chlorella ellipsoidea*. Mar. Biotechnol. (NY). **4**: 63–73.

Kim, J.-M., Kim, J.-S., Park, D.-H., Kang, H.S., Yoon, J., Baek, K., and Yoon, Y. (2004). Improved recombinant gene expression in CHO cells using matrix attachment regions. J. Biotechnol. **107**: 95–105.

Kim, J.H., Lee, S.R., Li, L.H., Park, H.J., Park, J.H., Lee, K.Y., Kim, M.K., Shin, B.A., and Choi, S.Y. (2011). High cleavage efficiency of a 2A peptide derived from *Porcine teschovirus*-1 in human cell lines, zebrafish and mice. PLoS One **6**.

Kim, M., Ahn, J., Jeon, H., Jin, E.S., Cutignano, A., and Romano, G. (2017). Development of a *Dunaliella tertiolecta* strain with increased zeaxanthin content using random mutagenesis. Mar. Drugs **15**.

Kindle, K.L. (1998). High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. Methods Enzymol. **297**: 27–38.

Kindle, K.L., Schnell, R.A., Fernandez, E., and Lefebvre, P.A. (1989). Stable nuclear transformation of *Chlamydomonas* using the *Chlamydomonas* gene for nitrate reductase. J. Cell Biol. **109**: 2589–2601.

Kirst, H., Formighieri, C., and Melis, A. (2014). Maximizing photosynthetic efficiency and culture productivity in cyanobacteria upon minimizing the phycobilisome light-harvesting antenna size. Biochim. Biophys. Acta - Bioenerg. **1837**: 1653–1664.

Knobloch, J.K.M., Nedelmann, M., Kiel, K., Bartscht, K., Horstkotte, M.A., Dobinsky, S., Rohde, H., and Mack, D. (2003). Establishment of an arbitrary PCR for rapid identification of *Tn917* insertion sites in *Staphylococcus epidermidis*: vharacterization of biofilm-negative and nonmucoid mutants. Appl. Environ. Microbiol. **69**: 5812–5818.

Knop, M., Siegers, K., Pereira, G., Zachariae, W., Winsor, B., Nasmyth, K., and Schiebel, E. (1999). Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. Yeast **15**: 963–972.

Kodym, A. and Afza, R. (2003). Physical and chemical mutagenesis BT - plant functional genomics. In Plant functional genomics, pp. 189–204.

Kok, B. (1956). On the inhibition of photosynthesis by intense light. Biochim. Biophys. Acta **21**: 234–244.

Kolber, Z.S., Prášil, O., and Falkowski, P.G. (1998). Measurements of variable chlorophyll fluorescence using fast repetition rate techniques: defining methodology and experimental protocols. Biochim. Biophys. Acta - Bioenerg. **1367**: 88–106.

Krieger-Liszkay, A. (2005). Singlet oxygen production in photosynthesis. In Journal of Experimental Botany, pp. 337–346.

Kruse, O., Rupprecht, J., Mussgnug, J.H., Dismukes, G.C., and Hankamer, B. (2005). Photosynthesis: a blueprint for solar energy capture and biohydrogen production technologies. Photochem. Photobiol. Sci. **4**: 957–970.

Kulp, T.R., Hoeft, S.E., Asao, M., Madigan, M.T., Hollibaugh, J.T., Fisher, J.C., Stolz, J.F., Culbertson, C.W., Miller, L.G., and Oremland, R.S. (2008). Arsenic(III) fuels anoxygenic photosynthesis in hot spring biofilms from Mono Lake, California. Science (80-.). **321**: 967–970.

Kumar, S.V., Misquitta, R.W., Reddy, V.S., Rao, B.J., and Rajam, M.V. (2004). Genetic transformation of the green alga *Chlamydomonas reinhardtii* by *Agrobacterium tumefaciens*. Plant Sci. **166**: 731–738.

Kunik, T., Tzfira, T., Kapulnik, Y., Gafni, Y., Dingwall, C., and Citovsky, V. (2001). Genetic transformation of HeLa cells by *Agrobacterium*. Proc. Natl. Acad. Sci. **98**: 1871–1876.

Lambrev, P.H., Miloslavina, Y., Jahns, P., and Holzwarth, A.R. (2012). On the relationship between non-photochemical quenching and photoprotection of Photosystem II. Biochim. Biophys. Acta - Bioenerg. **1817**: 760–769.

Lea-Smith, D.J., Bombelli, P., Dennis, J.S., Scott, S.A., Smith, A.G., and Howe, C.J. (2014). Phycobilisome-deficient strains of *Synechocystis* sp. PCC 6803 have reduced size and require carbon-limiting conditions to exhibit enhanced productivity. Plant Physiol. **165**: 705–714.

Lee, B., Choi, G.G., Choi, Y.E., Sung, M., Park, M.S., and Yang, J.W. (2014). Enhancement of lipid productivity by ethyl methane sulfonate-mediated random mutagenesis and proteomic analysis in *Chlamydomonas reinhardtii*. Korean J. Chem. Eng. **31**: 1036–1042.

León-Bañares, R., González-Ballester, D., Galván, A., and Fernández, E. (2004). Transgenic microalgae as green cell-factories. Trends Biotechnol. **22**: 45–52.

Lessard, J.C. (2013). Transformation of *E. coli* via electroporation. Methods Enzymol. **529**: 321–327.

Li, D., Han, X., Zuo, J., Xie, L., He, R., Gao, J., Chang, L., Yuan, L., and Cao, M. (2011). Construction of rice site-specific chloroplast transformation vector and transient expression of *EGFP* gene in *Dunaliella Salina*. J. Biomed. Nanotechnol. **7**: 801–806.

Li, J., Xue, L., Yan, H., Liu, H., and Liang, J. (2008). Inducible EGFP expression under the control of the nitrate reductase gene promoter in transgenic *Dunaliella salina*. J. Appl. Phycol. **20**: 137–145.

Li, J., Xue, L., Yan, H., Wang, L., Liu, L., Lu, Y., and Xie, H. (2007). The nitrate reductase gene-switch: a system for regulated expression in transformed cells of *Dunaliella salina*. Gene **403**: 132–142.

Li, Y., Han, D., Hu, G., Sommerfeld, M., and Hu, Q. (2010). Inhibition of starch synthesis results in overproduction of lipids in *Chlamydomonas reinhardtii*. Biotechnol. Bioeng. **107**: 258–268.

Lin, M.T., Occhialini, A., Andralojc, P.J., Parry, M.A.J., and Hanson, M.R. (2014). A faster Rubisco with potential to increase photosynthesis in crops. Nature **513**: 547–550.

Liu, G., Lanham, C., Buchan, J.R., and Kaplan, M.E. (2017a). High-throughput transformation of *Saccharomyces cerevisiae* using liquid handling robots. PLoS One **12**.

Liu, S., Zhao, Y., Liu, L., Ao, X., Ma, L., Wu, M., and Ma, F. (2015). Improving cell growth and lipid accumulation in green microalgae *Chlorella* sp. via UV irradiation. Appl.

Biochem. Biotechnol. **175**: 3507–3518.

Liu, W., Au, D.W.T., Anderson, D.M., Lam, P.K.S., and Wu, R.S.S. (2007). Effects of nutrients, salinity, pH and light:dark cycle on the production of reactive oxygen species in the alga *Chattonella marina*. J. Exp. Mar. Bio. Ecol. **346**: 76–86.

Liu, Y., Zhang, Y., Xing, H.H., and Liu, C.H. (2012). Evaluate the quality of *Tagetes ercta* petal by Fourier transform infrared spectroscopy. 2012 Symp. Photonics Optoelectron. SOPO 2012: 1–3.

Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L., and Liu, J. (2017b). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci. Rep. **7**.

Llamas, A., Igeño, M.I., Galván, A., and Fernández, E. (2002). Nitrate signalling on the nitrate reductase gene promoter depends directly on the activity of the nitrate transport systems in *Chlamydomonas*. Plant J. **30**: 261–271.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000). Molecular cell biology. 4th edition.

Loor, G., Kondapalli, J., Schriewer, J.M., Chandel, N.S., Vanden Hoek, T.L., and Schumacker, P.T. (2010). Menadione triggers cell death through ROS-dependent mechanisms involving PARP activation without requiring apoptosis. Free Radic. Biol. Med. **49**: 1925–1936.

Lubián, L., Montero, O., Moreno-Garrido, I., Huertas, I.E., Sobrino, C., González-del Valle, M., and Pares, G. (2000). *Nannochloropsis* (Eustigmatophyceae) as source of commercially valuable pigments. J. Appl. Phycol. **12**: 249–255.

Mangan, N. and Brenner, M. (2014). Systems analysis of the CO_2 concentrating mechanism in cyanobacteria. Elife **2014**.

Manuell, A.L., Beligni, M.V., Elder, J.H., Siefker, D.T., Tran, M., Weber, A., McDonald, T.L., and Mayfield, S.P. (2007). Robust expression of a bioactive mammalian protein in *Chlamydomonas* chloroplast. Plant Biotechnol. J. **5**: 402–412.

Marcus, Y., Altman-Gueta, H., Wolff, Y., and Gurevitz, M. (2011). Rubisco mutagenesis provides new insight into limitations on photosynthesis and growth in *Synechocystis* PCC6803. J. Exp. Bot. **62**: 4173–4182.

Markou, G. and Nerantzis, E. (2013). Microalgae for high-value compounds and biofuels production: a review with focus on cultivation under stress conditions. Biotechnol. Adv.

31: 1532–1542.

Marshall, J.-A., Hovenden, M., Oda, T., and Hallegraeff, G.M. (2002). Photosynthesis does influence superoxide production in the ichthyotoxic alga *Chattonella marina* (Raphidophyceae). J. Plankton Res. **24**: 1231–1236.

Martin, G.J.O., Hill, D.R.A., Olmstead, I.L.D., Bergamin, A., Shears, M.J., Dias, D.A., Kentish, S.E., Scales, P.J., Botté, C.Y., and Callahan, D.L. (2014). Lipid profile remodeling in response to nitrogen deprivation in the microalgae *Chlorella* sp. (Trebouxiophyceae) and *Nannochloropsis* sp. (Eustigmatophyceae). PLoS One **9**.

Masuda, T., Tanaka, A., and Melis, A. (2003). Chlorophyll antenna size adjustments by irradiance in *Dunaliella salina* involve coordinate regulation of chlorophyll a oxygenase (CAO) and *Lhcb* gene expression. Plant Mol. Biol. **51**: 757–771.

Mata, T.M., Martins, A.A., and Caetano, N.S. (2010). Microalgae for biodiesel production and other applications: a review. Renew. Sustain. Energy Rev. **14**: 217–232.

Melis, A. (2009). Solar energy conversion efficiencies in photosynthesis: minimizing the chlorophyll antennae to maximize efficiency. Plant Sci. **177**: 272–280.

Mellis, A. (1999). Photosystem II damage and repair cycle in chloroplasts: what modulates the rate of photodamage *in vivo*? Trends Plant Sci. **4**: 130–135.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, J., Witman, G.B., Terry, A., Salamov, A., Fritz-laylin, L.K., Maréchal-drouard, L., Marshall, W.F., Qu, L., Nelson, D.R., Sanderfoot, A., Spalding, M.H., Kapitonov, V.V., Ren, Q., Cardol, P., Cerutti, H., Chanfreau, G., Ferris, P., Fukuzawa, H., González-ballester, D., Mueller-roeber, B., Rajamani, S., Sayre, R.T., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y.W., Luo, Y., Martínez, D., Chi, W., Ngau, A., Otillar, B., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Igor, V., Rokhsar, D.S., and Grossman, A.R. (2010). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science (80-.). **318**: 245–250.

Minoda, A., Sakagami, R., Yagisawa, F., Kuroiwa, T., and Tanaka, K. (2004). Improvement of culture conditions and evidence for nuclear transformation by homologous recombination in a red alga, *Cyanidioschyzon merolae* 10D. Plant Cell Physiol. **45**: 667–671.

Minowa, T., Yokoyama, S. ya, Kishimoto, M., and Okakura, T. (1995). Oil production from algal cells of *Dunaliella tertiolecta* by direct thermochemical liquefaction. Fuel **74**: 1735–1738.

Mitra, M. and Melis, A. (2008). Optical properties of microalgae for enhanced biofuels production. Opt. Express **16**: 21807–21820.

Moore, A. (2008). Biofuels are dead: long live biofuels(?) - part two. N. Biotechnol. **25**: 96–100.

Mulkidjanian, A.Y., Koonin, E. V, Makarova, K.S., Mekhedov, S.L., Sorokin, A., Wolf, Y.I., Dufresne, A., Partensky, F., Burd, H., Kaznadzey, D., Haselkorn, R., and Galperin, M.Y. (2006). The cyanobacterial genome core and the origin of photosynthesis. Proc. Natl. Acad. Sci. **103**: 13126–13131.

Murdock, J.N. and Wetzel, D.L. (2009). FTIR microspectroscopy enhances biological and ecological analysis of algae. Appl. Spectrosc. Rev. **44**: 335–361.

Mussgnug, J.H., Thomas-Hall, S., Rupprecht, J., Foo, A., Klassen, V., McDowall, A., Schenk, P.M., Kruse, O., and Hankamer, B. (2007). Engineering photosynthetic light capture: impacts on improved solar energy to biomass conversion. Plant Biotechnol. J. **5**: 802–814.

Nelson, J. a and Lefebvre, P. a (1995). Targeted disruption of the *NIT8* gene in *Chlamydomonas reinhardtii*. Mol. Cell. Biol. **15**: 5762–5769.

Nelson, N. and Yocum, C.F. (2006). Structure and function of Photosystems I and II. Annu. Rev. Plant Biol. **57**: 521–565.

Nester, E. (2011). *Agrobacterium*: The natural genetic engineer (100 years later). APSnet Featur.

Neupert, J., Shao, N., Lu, Y., and Bock, R. (2012). Genetic transformation of the model green alga *Chlamydomonas reinhardtii*. Methods Mol. Biol. **847**: 35–47.

Newman, S.M., Boynton, J.E., Gillham, N.W., Randolph-Anderson, B.L., Johnson, A.M., and Harris, E.H. (1990). Transformation of chloroplast ribosomal RNA genes in *Chlamydomonas*: molecular and genetic characterization of integration events. Genetics **126**: 875–888.

Ng, I.S., Tan, S.I., Kao, P.H., Chang, Y.K., and Chang, J.S. (2017). Recent developments on genetic engineering of microalgae for biofuels and bio-based chemicals. Biotechnol. J. **12**.

Niu, Y.F., Yang, Z.K., Zhang, M.H., Zhu, C.C., Yang, W.D., Liu, J.S., and Li, H.Y. (2012). Transformation of diatom *Phaeodactylum tricornutum* by electroporation and establishment of inducible selection marker. Biotechniques **52**.

Niyogi, K.. (1999). Photoprotection revisited: genetic and Molecular Approaches. Annu. Rev. Plant. Physiol. Plant. Mol. BioL **50**: 333–359.

O'Toole, G.A., Pratt, L.A., Watnick, P.I., Newman, D.K., Weaver, V.B., and Kolter, R. (1999). Genetic approaches to study of biofilms. Methods Enzymol. **310**: 91–109.

Oda, T., Nakamura, A., Shikayama, M., Kawano, I., Ishimatsu, A., and Muramatsu, T. (1997). Generation of reactive oxygen species by raphidophycean phytoplankton. Biosci Biotechnol Biochem **61**: 1658–1662.

Odell, J.T., Nagy, F., and Chua, N.-H. (1985). Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature **313**: 810–812.

Olson, J.M. (2006). Photosynthesis in the Archean era. Photosynth. Res. **88**: 109–117.

Oren, A. (2005). A hundred years of *Dunaliella* research: 1905-2005. Saline Systems **1**: 2.

Oren, A. and Padan, E. (1978). Induction of anaerobic, photoautotrophic growth in the cyanobacterium *Oscillatoria limnetica*. J. Bacteriol. **133**: 558–563.

Page, L.E., Liberton, M., and Pakrasi, H.B. (2012). Reduction of photoautotrophic productivity in the cyanobacterium *Synechocystis* sp. strain PCC 6803 by phycobilisome antenna truncation. Appl. Environ. Microbiol. **78**: 6349–6351.

Park, S., Lee, Y., Lee, J.H., and Jin, E.S. (2013). Expression of the high light-inducible *Dunaliella LIP* promoter in *Chlamydomonas reinhardtii*. Planta **238**: 1147–1156.

Parker, M.S., Mock, T., and Armbrust, E.V. (2008). Genomic Insights into marine microalgae. Annu. Rev. Genet. **42**: 619–645.

Paul, M.J. and Foyer, C.H. (2001). Sink regulation of photosynthesis. J. Exp. Bot. **52**: 1383–1400.

Peach, C. and Velten, J. (1991). Transgene expression variability (position effect) of *CAT* and *GUS* reporter genes driven by linked divergent T-DNA promoters. Plant Mol. Biol. **17**: 49–60.

Perin, G., Bellan, A., Segalla, A., Meneghesso, A., Alboresi, A., and Morosinotto, T. (2015). Generation of random mutants to improve light-use efficiency of *Nannochloropsis gaditana* cultures for biofuel production. Biotechnol. Biofuels **8**: 161.

Petersen, T.N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8**: 785–786.

Pickles, L.M., Roe, S.M., Hemingway, E.J., Stifani, S., and Pearl, L.H. (2002). Crystal structure of the C-terminal WD40 repeat domain of the human Groucho/TLE1 transcriptional corepressor. Structure **10**: 751–761.

Pistorius, A.M.A., DeGrip, W.J., and Egorova-Zachernyuk, T.A. (2009). Monitoring of biomass composition from microbiological sources by means of FTIR spectroscopy. Biotechnol. Bioeng. **103**: 123–129.

Polle, J.E.W., Kanakagiri, S.-D., and Melis, A. (2003). *tla1*, a DNA insertional transformant of the green alga *Chlamydomonas reinhardtii* with a truncated light-harvesting chlorophyll antenna size. Planta **217**: 49–59.

Pollock, a., Eger, K. a., Godfrey, A.C., Adams, J.E., Mason, C.B., James, V., and Moroney, B. (2002). Use of the bleomycin resistance gene to generate tagged insertional mutants of *Chlamydomonas reinhardtii* that require elevated CO_2 for optimal growth. Funct. Plant Biol. **29**: 231–241.

Poulsen, N. and Kröger, N. (2005). A new molecular tool for transgenic diatoms: control of mRNA and protein biosynthesis by an inducible promoter-terminator cassette. FEBS J. **272**: 3413–3423.

Powles, S.B. (1984). Photoinhibition of photosynthesis induced by visible light. Annu. Rev. Plant Physiol. **35**: 15–44.

Prasad, B., Vadakedath, N., Jeong, H.-J., General, T., Cho, M.-G., and Lein, W. (2014). *Agrobacterium tumefaciens*-mediated genetic transformation of haptophytes (*Isochrysis* species). Appl. Microbiol. Biotechnol. **98**: 8629–8639.

Pratheesh, P.T., Vineetha, M., and Kurup, G.M. (2014). An efficient protocol for the *Agrobacterium*-mediated genetic transformation of microalga *Chlamydomonas reinhardtii*. Mol. Biotechnol. **56**: 507–515.

Priyadarshani, I. and Rath, B. (2012). Commercial and industrial applications of microalgae – a review. J. Algal Biomass Util. **3**: 89–100.

Purton, S. (2007). Tools and techniques for chloroplast transformation of *Chlamydomonas*. Adv. Exp. Med. Biol. **616**: 34–45.

Purton, S., Szaub, J.B., Wannathong, T., Young, R., and Economou, C.K. (2012). Genetic engineering of algal chloroplasts: progress and prospects. Russ. J. Plant Physiol. **31**: 479–494.

Pushkar, Y., Yano, J., Sauer, K., Boussac, A., and Yachandra, V.K. (2008). Structural

changes in the Mn_4Ca cluster and the mechanism of photosynthetic water splitting. Proc. Natl. Acad. Sci. **105**: 1879–1884.

Quinn, J.M. (1995). Two copper-responsive elements associated with the *Chlamydomonas Cyc6* gene function as targets for transcriptional activators. Plant Cell Online **7**: 623–638.

Radakovits, R., Jinkerson, R.E., Darzins, A., and Posewitz, M.C. (2010). Genetic engineering of algae for enhanced biofuel production. Eukaryot. Cell **9**: 486–501.

Radakovits, R., Jinkerson, R.E., Fuerstenberg, S.I., Tae, H., Settlage, R.E., Boore, J.L., and Posewitz, M.C. (2012). Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropis gaditana*. Nat. Commun. **3**: 686.

Ramey, C.J., Barón-Sola, Á., Aucoin, H.R., and Boyle, N.R. (2015). Genome engineering in Cyanobacteria: where we are and where we need to go. ACS Synth. Biol. **4**: 1186–1196.

Ramundo, S., Rahire, M., Schaad, O., and Rochaix, J.-D. (2013). Repression of essential chloroplast genes reveals new signaling pathways and regulatory feedback loops in *Chlamydomonas*. Plant Cell **25**: 167–186.

Rasala, B.A., Barrera, D.J., Ng, J., Plucinak, T.M., Rosenberg, J.N., Weeks, D.P., Oyler, G.A., Peterson, T.C., Haerizadeh, F., and Mayfield, S.P. (2013). Expanding the spectral palette of fluorescent proteins for the green microalga *Chlamydomonas reinhardtii*. Plant J. **74**: 545–556.

Rasala, B.A., Chao, S.S., Pier, M., Barrera, D.J., and Mayfield, S.P. (2014). Enhanced genetic tools for engineering multigene traits into green algae. PLoS One **9**.

Rasala, B.A., Muto, M., Lee, P.A., Jager, M., Cardoso, R.M.F., Behnke, A., Kirk, P., Hokanson, C.A., Crea, R., Mendez, M., and Mayfield, S.P. (2011). Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. Plant Biotechnol. J. **8**: 719–733.

Raven, J.A., Wollenweber, B., and Handley, L.L. (1992). A Comparison of ammonium and nitrate as nitrogen-sources for photolithotrophs. New Phytol. **121**: 19–32.

Raymond, J., Zhaxybayeva, O., Gogarten, J.P., and Blankenship, R.E. (2003). Evolution of photosynthetic prokaryotes: a maximum-likelihood mapping approach. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **358**: 223–30.

Renger, T. and Schlodder, E. (2008). The primary electron donor of photosystem II of the cyanobacterium *Acaryochloris marina* is a chlorophyll d and the water oxidation is driven by a chlorophyll a/chlorophyll d heterodimer. J. Phys. Chem. B **112**: 7351–7354.

Richards, H.A., Halfhill, M.D., Millwood, R.J., and Stewart, C.N. (2003). Quantitative GFP fluorescence as an indicator of recombinant protein synthesis in transgenic plants. Plant Cell Rep. **22**: 117–121.

Rismani-Yazdi, H., Haznedaroglu, B.Z., Bibby, K., and Peccia, J. (2011). Transcriptome sequencing and annotation of the microalgae *Dunaliella tertiolecta*: pathway description and gene discovery for production of next-generation biofuels. BMC Genomics **12**: 148.

De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C., and Falciatore, A. (2009). Gene silencing in the marine diatom *Phaeodactylum tricornutum*. Nucleic Acids Res. **37**.

Robzyk, K. and Kassir, Y. (1992). A simple and highly efficient procedure for rescuing autonomous plasmids from yeast. Nucleic Acids Res. **20**: 32000.

Roesler, K., Shintani, D., Savage, L., Boddupalli, S., and Ohlrogge, J. (1997). Targeting of the *Arabidopsis* homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. Plant Physiol. **113**: 75–81.

Rohr, J., Sarkar, N., Balenger, S., Jeong, B.R., and Cerutti, H. (2004). Tandem inverted repeat system for selection of effective transgenic RNAi strains in *Chlamydomonas*. Plant J. **40**: 611–621.

Rutherford, A.W. and Faller, P. (2003). Photosystem II: evolutionary perspectives. Philos. Trans. R. Soc. B Biol. Sci. **358**: 245–253.

Rutherford, A.W., Osyczka, A., and Rappaport, F. (2012). Back-reactions, short-circuits, leaks and other energy wasteful reactions in biological electron transfer: redox tuning to survive life in O_2. FEBS Lett. **586**: 603–616.

Sandesh Kamath, B., Vidhyavathi, R., Sarada, R., and Ravishankar, G.A. (2008). Enhancement of carotenoids by mutation and stress induced carotenogenic genes in *Haematococcus pluvialis* mutants. Bioresour. Technol. **99**: 8667–8673.

Sando, S., Narita, A., and Aoyama, Y. (2007). Light-up hoechst-DNA aptamer pair: generation of an aptamer-selective fluorophore from a conventional DNA-staining dye. ChemBioChem **8**: 1795–1803.

Savir, Y., Noor, E., Milo, R., and Tlusty, T. (2010). Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. Proc. Natl. Acad. Sci. U. S. A. **107**: 3475–80.

Schenk, P.M., Thomas-Hall, S.R., Stephens, E., Marx, U.C., Mussgnug, J.H., Posten, C., Kruse, O., and Hankamer, B. (2008). Second generation biofuels: high-efficiency

microalgae for biodiesel production. BioEnergy Res. **1**: 20–43.

Scherholz, M.L. and Curtis, W.R. (2013). Achieving pH control in microalgal cultures through fed-batch addition of stoichiometrically-balanced growth media. BMC Biotechnol. **13**.

Schmollinger, S., Strenkert, D., and Schroda, M. (2010). An inducible artificial microRNA system for *Chlamydomonas reinhardtii* confirms a key role for heat shock factor 1 in regulating thermotolerance. Curr. Genet. **56**: 383–389.

Schroda, M. (2006). RNA silencing in *Chlamydomonas*: mechanisms and tools. Curr. Genet. **49**: 69–84.

Schroda, M., Blöcker, D., and Beck, C.F. (2000). The *HSP70A* promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. Plant J. **21**: 121–131.

Schubert, W.D., Klukas, O., Saenger, W., Witt, H.T., Fromme, P., and Krauss, N. (1998). A common ancestor for oxygenic and anoxygenic photosynthetic systems: a comparison based on the structural model of photosystem I. J. Mol. Biol. **280**: 297–314.

Scott, S.A., Davey, M.P., Dennis, J.S., Horst, I., Howe, C.J., Lea-Smith, D.J., and Smith, A.G. (2010). Biodiesel from algae: challenges and prospects. Curr. Opin. Biotechnol. **21**: 277–286.

Scotto-Lavino, E., Du, G., and Frohman, M.A. (2007). 3′ end cDNA amplification using classic RACE. Nat. Protoc. **1**: 2555–2562.

Segovia, M. (2003). Cell death in the unicellular Chlorophyte *Dunaliella tertiolecta*. A hypothesis on the evolution of apoptosis in higher plants and metazoans. Plant Physiol. **132**: 99–105.

Senge, M.O., Ryan, A.A., Letchford, K.A., MacGowan, S.A., and Mielke, T. (2014). Chlorophylls, symmetry, chirality, and photosynthesis. Symmetry (Basel). **6**: 781–843.

Shabelnik, M.Y., Kovalevska, L.M., Yurchenko, M.Y., Shlapatska, L.M., Rzepetsky, Y., and Sidorenko, S.P. (2011). Differential expression of *PKD1* and *PKD2* in gastric cancer and analysis of *PKD1* and *PKD2* function in the model system. Exp. Oncol. **33**: 206–211.

Sharma, S.K., Nelson, D.R., Abdrabu, R., Khraiwesh, B., Jijakli, K., Arnoux, M., O'Connor, M.J., Bahmani, T., Cai, H., Khapli, S., Jagannathan, R., and Salehi-Ashtiani, K. (2015). An integrative Raman microscopy-based workflow for rapid *in situ* analysis of microalgal lipid bodies. Biotechnol. Biofuels **8**: 164.

Sharon-Gojman, R., Maimon, E., Leu, S., Zarka, A., and Boussiba, S. (2015). Advanced methods for genetic engineering of *Haematococcus pluvialis* (Chlorophyceae, Volvocales). Algal Res. **10**: 8–15.

Shimogawara, K., Fujiwara, S., Grossman, A., and Usuda, H. (1998). High-efficiency transformation of *Chlamydomonas reinhardtii* by electroporation. Genetics **148**: 1821–1828.

Siaut, M., Cuiné, S., Cagnon, C., Fessler, B., Nguyen, M., Carrier, P., Beyly, A., Beisson, F., Triantaphylidès, C., Li-Beisson, Y., and Peltier, G. (2011). Oil accumulation in the model green alga *Chlamydomonas reinhardtii*: characterization, variability between common laboratory strains and relationship with starch reserves. BMC Biotechnol. **11**.

Simionato, D., Block, M.A., La Rocca, N., Jouhet, J., Maréchal, E., Finazzi, G., and Morosinotto, T. (2013). The response of Nannochloropsis gaditana to nitrogen starvation includes *De novo* biosynthesis of triacylglycerols, a decrease of chloroplast galactolipids, and reorganization of the photosynthetic apparatus. Eukaryot. Cell **12**: 665–676.

Sineshchekov, O.A., Jung, K.-H., and Spudich, J.L. (2002). Two rhodopsins mediate phototaxis to low- and high-intensity light in *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. U. S. A. **99**: 8689–94.

Smith, D., Lee, R.W., Cushman, J.C., Magnuson, J.K., Tran, D., and Polle, J.E. (2010). The *Dunaliella salina* organelle genomes: Large sequences, inflated with intronic and intergenic DNA. BMC Plant Biol. **10**: 83.

Soares, M.C.S., Marinho, M.M., Huszar, V.L.M., Branco, C.W.C., and Azevedo, S.M.F.O. (2008). The effects of water retention time and watershed features on the limnology of two tropical reservoirs in Brazil. Lakes Reserv. Res. Manag. **13**: 257–269.

Sodeinde, O.A. and Kindle, K.L. (1993). Homologous recombination in the nuclear genome of *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. U. S. A. **90**: 9199–203.

Sofiyanti, N., Fitmawati, D.I., and Roza, A.A. (2015). <i>Stenochlaena riauensis<i> (Blechnaceae), a new fern species from riau, Indonesia. In Bangladesh Journal of Plant Taxonomy, pp. 137–141.

Song, B. and Ward, B.B. (2004). Molecular characterization of the assimilatory nitrate reductase gene and its expression in the marine green alga *Dunaliella tertiolecta* (Chlorophyceae). J. Phycol. **40**: 721–731.

Sontheimer, E.J. and Carthew, R.W. (2005). Silence from within: endogenous siRNAs and

miRNAs. Cell **122**: 9–12.

Soupene, E., Inwood, W., and Kustu, S. (2004). Lack of the Rhesus protein *Rh1* impairs growth of the green alga *Chlamydomonas reinhardtii* at high CO_2. Proc. Natl. Acad. Sci. U. S. A. **101**: 7787–7792.

Springer, P.S. (2000). Gene traps: tools for plant development and genomics. Plant Cell **12**: 1007–1020.

Starckx and Senne (2012). A place in the sun - algae is the crop of the future. Geel Flanders Today **7**: 703–705.

Stehfest, K., Toepel, J., and Wilhelm, C. (2005). The application of micro-FTIR spectroscopy to analyze nutrient stress-related changes in biomass composition of phytoplankton algae. Plant Physiol. Biochem. **43**: 717–726.

Steinbrenner, J. and Sandmann, G. (2006). Transformation of the green alga *Haematococcus pluvialis* with a phytoene desaturase for accelerated astaxanthin biosynthesis. Appl. Environ. Microbiol. **72**: 7477–7484.

Stephenson, A.L., Kazamia, E., Dennis, J.S., Howe, C.J., Scott, S.A., and Smith, A.G. (2010). Life-cycle assessment of potential algal biodiesel production in the united kingdom: a comparison of raceways and air-lift tubular bioreactors. Energy and Fuels **24**: 4062–4077.

Stephenson, P.G., Moore, C.M., Terry, M.J., Zubkov, M. V., and Bibby, T.S. (2011). Improving photosynthesis for algal biofuels: toward a green revolution. Trends Biotechnol. **29**: 615–623.

Stern, D.B., Witman, G., and Harris, E.H. (2008). The *Chlamydomonas* sourcebook. Elsevier Sci. B.V. **ISBN: 978**-: 1–2000.

Stevens, D.R., Rochaix, J.D., and Purton, S. (1996). The bacterial phleomycin resistance gene *ble* as a dominant selectable marker in *Chlamydomonas*. Mol. Gen. Genet. **251**: 23–30.

Stirnimann, C.U., Petsalaki, E., Russell, R.B., and Müller, C.W. (2010). WD40 proteins propel cellular networks. Trends Biochem. Sci. **35**: 568–574.

Stock, D., Leslie, A.G., and Walker, J.E. (1999). Molecular architecture of the rotary motor in ATP synthase. Science **286**: 1700–5.

Sukenik, A., Carmeli, Y., and Berner, T. (1989). Regulation of fatty acid composition by irradiance level in the Eustigmatophyte *Nannochloropsis* Sp. J. Phycol. **25**: 686–692.

Sun, Y., Yang, Z., Gao, X., Li, Q., Zhang, Q., and Xu, Z. (2005). Expression of foreign genes in *Dunaliella* by electroporation. Mol. Biotechnol. **30**: 185–192.

Suzuki, N., Koussevitzky, S., Mittler, R., and Miller, G. (2012). ROS and redox signalling in the response of plants to abiotic stress. Plant, Cell Environ. **35**: 259–270.

Szivák, I., Behra, R., and Sigg, L. (2009). Metal-induced reactive oxygen species production in *Chlamydomonas reinhardtii* (chlorophyceae). J. Phycol. **45**: 427–435.

Takagi, M., Karseno, and Yoshida, T. (2006). Effect of salt concentration on intracellular accumulation of lipids and triacylglyceride in marine microalgae *Dunaliella* cells. J. Biosci. Bioeng. **101**: 223–226.

Tan, C., Qin, S., Zhang, Q., Jiang, P., and Zhao, F. (2005a). Establishment of a micro-particle bombardment transformation system for *Dunaliella salina*. J. Microbiol. **43**: 361–5.

Tan, G., Gao, Y., Shi, M., Zhang, X., He, S., Chen, Z., and An, C. (2005b). SiteFinding-PCR: a simple and efficient PCR method for chromosome walking. Nucleic Acids Res. **33**: 1–7.

Tan, K.W.M. and Lee, Y.K. (2016). The dilemma for lipid productivity in green microalgae: importance of substrate provision in improving oil yield without sacrificing growth. Biotechnol. Biofuels **9**.

Tan, K.W.M., Lin, H., Shen, H., and Lee, Y.K. (2016). Nitrogen-induced metabolic changes and molecular determinants of carbon allocation in *Dunaliella tertiolecta*. Sci. Rep. **6**: 1–13.

Teodoresco, E.C. (1905). Organisation et developpement du *Dunaliella*, nouveau genre de Volvocacee-Polyblepharidee. Bot Zentralbl Beih **18**: 215–232.

Terashima, M., Freeman, E.S., Jinkerson, R.E., and Jonikas, M.C. (2015). A fluorescence-activated cell sorting-based strategy for rapid isolation of high-lipid *Chlamydomonas mutants*. Plant J. **81**: 147–159.

Thompson, A. and Gasson, M.J. (2001). Location effects of a reporter gene on expression levels and on native protein synthesis in *Lactococcus lactis* and *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. **67**: 3434–3439.

Thompson, G.A. (1996). Lipids and membrane function in green algae. Biochim. Biophys. Acta - Lipids Lipid Metab. **1302**: 17–45.

Tillich, U.M., Wolter, N., Schulze, K., Kramer, D., Brödel, O., and Frohme, M. (2014). High-throughput cultivation and screening platform for unicellular phototrophs. BMC Microbiol. **14**: 239.

Tjus, S.E., Scheller, H. V, Andersson, B., and Møller, B.L. (2001). Active oxygen produced during selective excitation of photosystem I is damaging not only to photosystem I, but also to photosystem II. Plant Physiol. **125**: 2007–2015.

Tran, K.T., Maeda, T., Sanchez-Torres, V., and Wood, T.K. (2015). Beneficial knockouts in *Escherichia coli* for producing hydrogen from glycerol. Appl. Microbiol. Biotechnol. **99**: 2573–2581.

Twiner, M.J. and Trick, C.G. (2000). Possible physiological mechanisms for production of hydrogen peroxide by the ichthyotoxic flagellate *Heterosigma akashiwo*. J. Plankton Res. **22**: 1961–1975.

Umena, Y., Kawakami, K., Shen, J.-R., and Kamiya, N. (2011). Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. Nature **473**: 55–60.

Varela, J.C., Pereira, H., Vila, M., and León, R. (2015). Production of carotenoids by microalgae: achievements and challenges. Photosynth. Res. **125**: 423–436.

Vidyashankar, S. and Ravishankar, G.A. (2016). Algae-based bioremediation.

Vigeolas, H. and Geigenberger, P. (2004). Increased levels of glycerol-3-phosphate lead to a stimulation of flux into triacylglycerol synthesis after supplying glycerol to developing seeds of *Brassica napus* L. in planta. Planta **219**: 827–835.

Vigeolas, H., Waldeck, P., Zank, T., and Geigenberger, P. (2007). Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. Plant Biotechnol. J. **5**: 431–441.

Viji, S., Anbazhagi, M., Ponpandian, N., Mangalaraj, D., Jeyanthi, S., Santhanam, P., Devi, A.S., and Viswanathan, C. (2014). Diatom-based label-free optical biosensor for biomolecules. Appl. Biochem. Biotechnol. **174**: 1166–1173.

Vila, M., Díaz-Santos, E., De La Vega, M., Rodriǵuez, H., Vargas, Á., and León, R. (2012). Promoter trapping in microalgae using the antibiotic paromomycin as selective agent. Mar. Drugs **10**: 2749–2765.

Vinayak, V., Manoylov, K.M., Gateau, H., Blanckaert, V., Hérault, J., Pencréac'H, G., Marchand, J., Gordon, R., and Schoefs, B. (2015). Diatom milking? A review and new

approaches. Mar. Drugs **13**: 2629–2665.

Voinnet, O. (2009). Origin, biogenesis, and activity of pant microRNAs. Cell **136**: 669–687.

Vredenberg, W.J. and Duysens, L.N.M. (1963). Transfer of energy from bacteriochlorophyll to a reaction centre during bacterial photosynthesis. Nature **197**: 355–357.

Wagner, H., Liu, Z., Langner, U., Stehfest, K., and Wilhelm, C. (2010). The use of FTIR spectroscopy to assess quantitative changes in the biochemical composition of microalgae. J. Biophotonics **3**: 557–566.

Walker, T.L., Becker, D.K., and Collet, C. (2005a). Characterisation of the *Dunaliella tertiolecta RbcS* genes and their promoter activity in *Chlamydomonas reinhardtii*. Plant Cell Rep. **23**: 727–735.

Walker, T.L., Becker, D.K., Dale, J.L., and Collet, C. (2005b). Towards the development of a nuclear transformation system for *Dunaliella tertiolecta*. J. Appl. Phycol. **17**: 363–368.

Wallin, I.E. (1993). Symbionticism in the light of recent cytological investigations - Ivan E. Wallin, 1969. BioSystems **31**: 181–183.

Wan, M., Rosenberg, J.N., Faruq, J., Betenbaugh, M.J., and Xia, J. (2011). An improved colony PCR procedure for genetic screening of *Chlorella* and related microalgae. Biotechnol. Lett. **33**: 1615–1619.

Wang, H., Hill, R.T., Zheng, T., Hu, X., and Wang, B. (2016). Effects of bacterial communities on biofuel-producing microalgae: stimulation, inhibition and harvesting. Crit. Rev. Biotechnol. **36**: 341–352.

Wang, P., Sun, Y., Li, X., Zhang, L., Li, W., and Wang, Y. (2004). Rapid isolation and functional analysis of promoter sequences of the nitrate reductase gene from *Chlorella ellipsoidea*. J. Appl. Phycol. **16**: 11–16.

Wang, Z.T., Ullrich, N., Joo, S., Waffenschmidt, S., and Goodenough, U. (2009). Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtit*. Eukaryot. Cell **8**: 1856–1868.

Weeks, D.P. (2011). Homologous recombination in *Nannochloropsis*: a powerful tool in an industrially relevant alga. Proc. Natl. Acad. Sci. **108**: 20859–20860.

Wei, L., Xin, Y., Wang, Q., Yang, J., Hu, H., and Xu, J. (2017). RNAi-based targeted gene knockdown in the model oleaginous microalgae *Nannochloropsis oceanica*. Plant J. **89**: 1236–1250.

Whitelegge, J., Zhang, H., and Aguilera, R. (2002). Full subunit coverage liquid chromatography electrospray ionization mass spectrometry (LCMS+) of an oligomeric membrane protein cytochrome b_6f Complex. Mol. Cell. … **1**: 816–827.

Whitney, S.M., Houtz, R.L., and Alonso, H. (2011). Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. Plant Physiol. **155**: 27–35.

Wijffels, R.H. and Barbosa, M.J. (2010). An outlook on microalgal biofuels. Science (80-.). **329**: 796–799.

Yamano, T., Iguchi, H., and Fukuzawa, H. (2013). Rapid transformation of *Chlamydomonas reinhardtii* without cell-wall removal. J. Biosci. Bioeng. **115**: 691–694.

Yu, W.L., Ansari, W., Schoepp, N.G., Hannon, M.J., Mayfield, S.P., and Burkart, M.D. (2011). Modifications of the metabolic pathways of lipid and triacylglycerol production in microalgae. Microb. Cell Fact. **10**.

Yuste, R. (2005). Fluorescence microscopy today. Nat. Methods **2**: 902–904.

Zarzycki, J., Brecht, V., Müller, M., and Fuchs, G. (2009). Identifying the missing steps of the autotrophic 3-hydroxypropionate CO_2 fixation cycle in *Chloroflexus aurantiacus*. Proc. Natl. Acad. Sci. U. S. A. **106**: 21317–22.

Zaslavskaia, L.A., Casey Lippmeier, J., Kroth, P.G., Grossman, A.R., and Apt, K.E. (2000). Transformation of the diatom *Phaeodactylum tricornutum* (Bascillariophyceae) with a variety of selectable marker and reporter genes. J. Phycol. **36**: 379–386.

Zavafer, A., Cheah, M.H., Hillier, W., Chow, W.S., and Takahashi, S. (2015). Photodamage to the oxygen evolving complex of photosystem II by visible light. Sci. Rep. **5**: 16363.

Zhang, G., Gurtu, V., and Kain, S.R. (1996). An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. Biochem. Biophys. Res. Commun. **227**: 707–711.

Zhang, R., Patena, W., Armbruster, U., Gang, S.S., Blum, S.R., and Jonikas, M.C. (2014). High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. Plant Cell **26**: 1398–1409.

Zhao, R., Cao, Y., Xu, H., Lv, L., Qiao, D., and Cao, Y. (2011). Analysis of expressed sequence tags from the green alga *Dunaliella Salina* (Chlorophyta). J. Phycol. **47**: 1454–1460.

Zorin, B., Grundman, O., Khozin-Goldberg, I., Leu, S., Shapira, M., Kaye, Y., Tourasse, N., Vallon, O., and Boussiba, S. (2014). Development of a nuclear transformation system for oleaginous green alga *Lobosphaera* (Parietochloris) *incisa* and genetic complementation of a mutant strain, deficient in arachidonic acid biosynthesis. PLoS One **9**.

Zorin, B., Hegemann, P., and Sizova, I. (2005). Nuclear-gene targeting by using single-stranded DNA avoids illegitimate DNA integration in *Chlamydomonas reinhardtii*. Eukaryot. Cell **4**: 1264–1272.

Zupan, J., Muth, T.R., Draper, O., and Zambryski, P. (2000). The transfer of DNA from *Agrobacterium tumefaciens* into plants: a feast of fundamental insights. Plant J. **23**: 11–28.