Development of a molecular toolkit to genetically engineer the microalga *Nannochloropsis gaditana* CCMP526 for biotechnology applications

A thesis submitted to the University of Technology Sydney for the degree of DOCTOR OF PHILOSOPHY (PhD)

Submitted October 2020

Margaret Ramarajan, BSc, MSc.

Supervisors: Professor Peter J. Ralph, Dr Michele Fabris,

Climate Change Cluster, School of Life Sciences, University of Technology Sydney

Certificate of original authorship

I, Margaret Ramarajan declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Life Science at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. Any help that I have received in my research work and the preparation of this thesis itself has been acknowledged. This research is supported by the Australian Government Research Training Program.

Signature:

Production Note:
Signature removed prior to publication.

Date: 02/10/2020

Acknowledgements

I thank God for his abundant grace and wisdom to fulfil my PhD research. I would like to thank Professor Peter Ralph for giving me this wonderful opportunity to work in Climate Change Cluster (C3) at the University of Technology Sydney, and my co-supervisor Dr Michele Fabris for their time, strong support and help throughout the entire duration of my PhD project. This research was supported by the UTS International Research Scholarship (UTS IRS) and UTS President's Research Scholarship.

I also would like to thank Dr Mathieu Pernice and Dr. Raffaela M. Abbriano for their help and support with data analysis. I recognise Dr Unnikrishnan Kuzhiumparambil for providing support in GC-MS analysis and Kun Xiao for the much-needed teaching and guidance in flow cytometry. I acknowledge Dr Tim Kahlke for his guidance in using bioinformatics tools. I am grateful for the custom script developed by Dr Chris Evenhuis for microscopic cell-count. I thank Dr Stephen Woodcock for his valuable guidance on statistical analysis. I acknowledge Dr Audrey Commault, Dr Justin Ashworth, Dr Christopher Hall, and Dr Leen Labeeuw for providing useful counsel and discussions.

I thank the technical staff: Paul Brooks, Gemma Armstrong, Scott Allchin, Taya Lapshina, Sue Fenech, and Rachel Keeys, for helping me with the various training and inductions and other lab-based needs. I am grateful for my fellow research scholars for making this PhD journey interesting and fun: Ana Cristina Jaramillo Madrid, Dylan Russell, Jack Adriaans, Jestin George, Lorenzo Barolo, Marco Giardina, Matthias Windhagauer, Navpreet Kaur Walia, Nathan Williams, Nine Le Reun and Shawn Price. Special thanks to Rachelle Tran and Abeeha Khalil for being supportive and all the ice-cream breaks.

Lastly, I extent my heartfelt gratitude for the love and support I received from my family during these three years of my PhD. My parents, my husband and my lovely six-year-old daughter. I thank them for their prayers and encouragement to undertake a PhD degree overseas and being a pillar of strength during hard and stressful times.

Preface

This thesis has been prepared for submission in a thesis by compilation format, whereby the thesis contains a combination of published and publishable work. Considering that this thesis is presented as a series of ready to submit manuscripts, there is a degree of repetition across chapters, particularly within the introductions and materials and methods sections of Chapter 2, 3 and 4. Published work (Chapter 2) has been incorporated into this thesis and appears as it was presented to the journal immediately prior to publication with the following modifications: i) the font and format was changed to maintain consistency across the thesis, ii) figures and tables were re-numbered to reflect the chapter numbering and iii) supplementary information for each chapter appear in the appendix and have been re-numbered accordingly. The referencing format used throughout this thesis conforms to the requirements of the journal Nature.

List of publications:

Chapter 2:

Novel endogenous promoters for genetic engineering of the marine microalga *Nannochloropsis gaditana* CCMP526

Margaret Ramarajan[a], Michele Fabris[a,b], Raffaela M. Abbriano[a] , Mathieu Pernice[a], Peter J. Ralph[a]

[a] University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo NSW 2007, Australia

[b] CSIRO Synthetic Biology Future Science Platform, GPO box 2583, Brisbane QLD 4001, Australia

Published in December 2019 within the journal of 'Algal Research- Biomass, Biofuels and Bioproducts' volume 44, as a research article.

Author Contribution: M.R. performed the experiments; M.F. and P.J.R. conceived the research project; M.F, R.A. and M. P. contributed in data analysis; M.R. wrote the

manuscript; all authors edited and reviewed the manuscript.

Chapter 4:

Genetic engineering of the microalga *Nannochloropsis gaditana* for the production of sesquiterpene β-caryophyllene

Margaret Ramarajan[a], Michele Fabris[a,b], Unnikrishnan Kuzhiumparambil[a] , Peter J. Ralph[a]

[a] University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo NSW 2007, Australia

[b] CSIRO Synthetic Biology Future Science Platform, GPO box 2583, Brisbane QLD 4001, Australia

Manuscript: In-prep for submission to Metabolic Engineering

Author Contribution: M.R. performed the experiments; M.R, M.F. and P.J.R. conceived the research project; U.K. performed GC-MS; M.F, and U.K. contributed in data analysis; M.R. wrote the manuscript; all authors edited and reviewed the manuscript.

Conferences:

- Poster presentation at the Plant Biology World Summit 2020. Title: Genetic engineering of the marine microalga *Nannochloropsis gaditana* CCMP526 for the production of sesquiterpene β-caryophyllene
- Poster presentation at Australia New Zealand Marine Biotechnology Symposium (ANZMBS 2019). Title: Novel endogenous promoters for genetic engineering of the marine microalga *Nannochloropsis gaditana* CCMP526
- Attended the Synthetic Biology Australasia Conference, Sydney 2017

Table of Contents

List of Figures

Supplementary figures:

List of Tables

Supplementary tables:

General Abbreviations A to Z

Acronym	Definition
AACT	Acetoacetyl-CoA thiolase
ABA	Abscisic acid
ABRE	ACGT-containing abscisic acid response element
AMT	Ammonium transporter
ANOVA	Analysis of variance
ANZMBS	Australia New Zealand Marine Biotechnology Symposium
AOC	Auto Injector/Auto Sampler
AP2	transcription factor APETALA2
ATP	Adenosine triphosphate
BCA	Bovine serum albumin
bHLH	Basic helix-loop-helix transcription factor
BKT	β-carotene ketolase
bZIP	basic leucine-zipper transcription factor
CCAP	The Culture Collection of Algae and Protozoa
CCM	Carbon concentration mechanisms
CDP-ME	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol
CDPMEK	4-diphosphocytidyl-2c-methyl-d-erythritol kinase
CMS	4-Diphosphocytidyl-2C-methyl-D-erythritol synthase
CMV	cytomegalovirus
CRIBI	Centro di Ricerca Interdipartimentale per le Biotecnologie Innovative
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CSM	2-c-methyl-d-erythritol 4-phosphate cytidylyltransferase
CYP	Cytochrome P450 monooxygenase
DCM	Dichloromethane
DHA	Docosahexaenoic acid
DMAPP	Dimethyl allyl pyrophosphate
DMSO	Dimethyl sulfoxide
DNA	Deoxy-ribonucleic acid
DOXP	Deoxyxylulose 5-Phosphate
DXP	1-deoxy-D-xylulose-5-phosphate
DXR	1-deoxy-D-xylulose-5-phosphate reductoisomerase
DXS	1-deoxy-d-xylulose 5-phosphate synthase
ECL	Enhanced chemiluminescence
EDP	Eukaryotic Promoter Database
EEP	Endonuclease/exonuclease/phosphatase
EIC	Extracted-ion chromatogram
EPA	Eicosapentaenoic acid
EPPSII	Extrinsic protein in photosystem II
ESAW	Enriched Seawater, Artificial Water

FACS	Fluorescence assisted cell sorting
FCCP	Fucoxanthin-chlorophyll a-c binding protein
FDA	Food and Drug Administration
FITC	Fluorescein Isothiocyanate
FMDV	Foot-and-mouth disease virus
FPKM	Fragments Per Kilo base of exon per Million fragments mapped
FPP	Farnesyl pyrophosphate
FPPS	Farnesyl pyrophosphate synthase
GAL	Promoter with Gal4p-binding sites
GDPD	Glycerophosphoryl diester phosphodiesterase
GEO	Gene Expression Omnibus
GFP	Green Fluorescent Protein
GGPP	Geranylgeranyl pyrophosphate
GPP	Geranyl pyrophosphate
GUS	β-glucuronidase
HDR	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
HDS	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase
HMBDP	4-hydroxy-3- methylbut-2-enyl diphosphate
HMBPP	hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
HMG-CoA	β-Hydroxy β-methylglutaryl-CoA
HMGR	β-Hydroxy β-methylglutaryl-CoA reductase
HMGS	β-Hydroxy β-methylglutaryl-CoA synthase
HSF	Heat shock transcription factor
HSP	Heat shock protein
HTPG	Heat or thermotolerance protein G
HYP	Hypothetical protein
IDI	Isopentenyl diphosphate:dimethylallyl diphosphate isomerase
IDT	Integrated DNA technologies
INR	Initiator element
IPM	Isopropyl myristate
IPP	Isopentenyl pyrophosphate
ISPD	Isoprenoid synthase domain
ISPF	2-c-methyl-d-erythritol–cyclodiphosphate
LDSP	Lipid droplet surface protein
LED	Light Emitting Diode
LIP	Light-inducible protein
MECDP	2-C-methyl-D-erythritol 2,4-cyclodiphosphate
MEP	Methyl-D-erythritol 4-phosphate
MVA	Mevalonate
MYB	Myeloblastosis proto-oncogene, transcription factor
MYC	Myeloblastosis rroto-oncogene, basic helix-loop-helix transcription factor
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NCBI	National Center for Biotechnology Information
NCLDV	Nucleocytoplasmic large DNA viruses

NF YB	Nuclear transcription factor-Y subunit beta
NF YC	Nuclear transcription factor-Y subunit gamma
NIES	Nitrate inducible expression system
NSW	New South Wales
NUDIX	Nucleoside diphosphates linked to some moiety X
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAM	Pulse-amplitude modulated fluorometry
PBS	Phosphate buffer saline
PCC	The Pasteur culture collection of cyanobacteria
PCR	Polymerase chain reaction
PMD	Mevalonate 5-pyrophosphate decarboxylase
PMK	Mevalonate 5-phosphate kinase
PSII	Photosystem II
PSU	Practical Salinity Unit
PTM	Post-translational modifications
PUFA	Polyunsaturated fatty acids
PVDF	Polyvinylidene difluoride
QLD	Queensland
RIGS	Repeat-induced gene silencing
RNA	Ribonucleic acid
RPKM	Reads Per Kilobase of transcript, per Million mapped reads
SDS	Sodium dodecyl sulfate
SIM	Selective Ion Monitoring
SIT	Silicon transporters
SNP	Single nucleotide polymorphisms
SPS	Sodium Phosphate Symporter
SQDG	Sulfoquinovosyldiacylglycerol
SQS	Squalene synthase
STRE	Stress response element
TAG	Triacylglycerols
TAIR	The Arabidopsis Information Resource
TCA	Tricarboxylic acid
TMHMM	Transmembrane hidden markov model
TSS	Transcription start sites
TSSP	Transcription start sites for plant
TUB	β-tubulin
UEP	Ubiquitin extension protein
USA	United States of America
UTR	Untranslated Region
UTS	University of Technology Sydney
VCP	Violaxanthin/Chlorophyll α-binding Protein
YAC	Yeast Artificial Chromosome
YFP	Yellow Fluorescent Protein

Thesis summary

Model microalgae such as *Chlamydomonas reinhardtii* and *Phaeodactylum tricornutum* are well suited for genetic engineering as they possess an array of well characterised genetic tools that can be used for gene and genome manipulations. *Nannochloropsis gaditana* is a photosynthetic oleaginous microalga that has been studied extensively due to a broad range of industrial applications such as oil, polyunsaturated fatty acids (PUFAs) and pigments. Even though genetic engineering and synthetic biology resources are rapidly advancing and widening this alga's industrial potential, there are still some limitations such as a narrow repertoire of functional promoters with limited diversity in terms of expression range and inducibility; and a lack of knowledge of specific regulatory elements that control transcription in *Nannochloropsis* spp. that would allow for predictable expression and rational design of promoter regions. To date, several studies involving gene over-expression have been conducted in *Nannochloropsis* spp. using endogenous promoters, but in-depth promoter analysis done in other model microalgae like *P. tricornutum* and *C. reinhardtii* is still missing with *Nannochloropsis* spp. By identifying and profiling a suite of promoters for use in *N. gaditana* CCMP526, we have expanded the genetic toolbox available for this industrially relevant microalgal species both for biotechnological applications such as metabolic engineering or recombinant protein production, and to understand the biology of *Nannochloropsis.* In this research we also explored *N. gaditana* as a novel platform for sesquiterpenoid biosynthesis and it represents the first report of terpenoid engineering in this microalga, which shed some light into metabolic pathway engineering for terpenoid production in *N. gaditana*. Whilst this PhD thesis offers new knowledge into terpenoid biosynthesis, it also provides candidate gene promoters, which can be used in a wide variety of biotechnology application in *N. gaditana*.

CHAPTER 1

General Introduction

Margaret Ramarajan[a], Michele Fabris[a,b], Peter J. Ralph[a]

[a] University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo NSW 2007, Australia

[b] CSIRO Synthetic Biology Future Science Platform, GPO box 2583, Brisbane QLD 4001, Australia

Author contribution: M.R. wrote the chapter; M.F. and P.J.R edited and reviewed the chapter.

Abstract

The oleaginous autotrophic tiny protist, *Nannochloropsis* spp. are found in a variety of habitats and can withstand a range of salinity and temperature fluctuations. *Nannochloropsis* strains are algal bio-factories that produce biomass including lipids by converting sunlight and CO_2 which can be utilised as feedstock for multiple industrial uses such as biofuel, aquaculture feed and other high value bio-products such as sterols, pigments and terpenoids. The oleaginous microalga *Nannochloropsis gaditana* has been studied extensively for biofuel production because of its high lipid content, but recent research has focused on exploiting this alga for other biotechnology applications. The availability of fully sequenced genomes, omics datasets, transformation protocols and gene editing tools are essential to propel this microalga become an industrially relevant species.

This chapter focuses on the molecular background of this organism highlighting the merits and specific areas that require further research. Research that would enable the development of molecular tools needed for an algal bio-factory or genetic engineering host that decades of research and improvement has provided to other biotechnological host systems (bacterial, mammalian and plant). We examine the role of promoters in gene expression, the different types of promoters currently available for transgene expression in *N. gaditana*, how these could be exploited for biotechnological and industrial applications, such as terpenoid production. We identify and report the gaps in research, which would play a crucial role in the advancement of this alga for biotechnology applications. In doing so, we would gain a better understanding of this microalga as a production host system that can be used to effectively supplement the ever-increasing demand for nutraceutical and pharmaceutical products such as β-glucans, terpenoids, pigments and sterols.

1.1 Introduction

Microalgae belonging to the genus *Nannochloropsis* were first placed in the class Eustigmatophyceae by Hibberd[1] after analysing several strains from culture collections; Cambridge, Texas, Gottingen, Prague and Innsbruck[1]. He explained," The stem of Nannochloris is Nannochlorid - and so the correct form of the new name is *Nannochloridopsis*. *Nannochloropsis* represents a shortened form of this name"[1]. Hibberd described *Nannochloropsis* to be very small, greenish coloured alga which apparently do not produce motile reproductive stages[1]. These tiny protists are about 2-3 µm in diameter, have a very simple ultra-structure, consisting of a nucleus, a single mitochondrion, a single Golgi body, a single chloroplast, a simple cell wall and usually a single carbohydrate storage "grain"; there is no evidence of basal bodies or any flagellar features[1,2]. *Nannochloropsis* spp. have several noticeable biochemical features, for example the absence of chlorophyll b and c[1] and the presence of β carotene, chlorophyll a and xanthophyll pigments like astaxanthin, zeaxanthin, violaxanthin, vaucheriaxanthin and canthaxanthin[3,4] and the ability to store lipids in its cytosol in the form of triacylglycerides (TAG)[5–7] and omega 3 fatty acids such as eicosapentaenoic acid (EPA)[8–10].

Nannochloropsis spp. are found in a variety of habitats such as freshwater, brackish and marine water[2]. *Nannochloropsis* species belong to the family Eustigmatophyceae and belongs to the stramenopile lineage[11,12] (Figure 1) after undergoing both primary and secondary endosymbiotic events[13] (Figure 2). In the primary endosymbiotic event, a cyanobacterium was engulfed into a heterotroph and was retained in its cytosol without digestion, ultimately transferring most of its genome to the nucleus of the host[14,15]. This is the progenitor plant cell from which three clades subsequently evolved – plants/ green algae, red algae and glaucophytes[13]. Independent secondary endosymbiosis events resulted in the uptake of this primary endosymbiotic cell by eukaryotic phagotrophs (heterotroph), which in the case of stramenopiles was the red algae[16,17]. Therefore like diatoms[18], the plastids of *Nannochloropsis* contain four enveloping membranes; the inner two are equivalent to the envelope membranes of plants, and the outer two which are similar to and continuous with the endoplasmic

reticulum[19] which functions as a barrier between the ER lumen and the primary chloroplast, together forming the nucleus-plastid continuum[12].

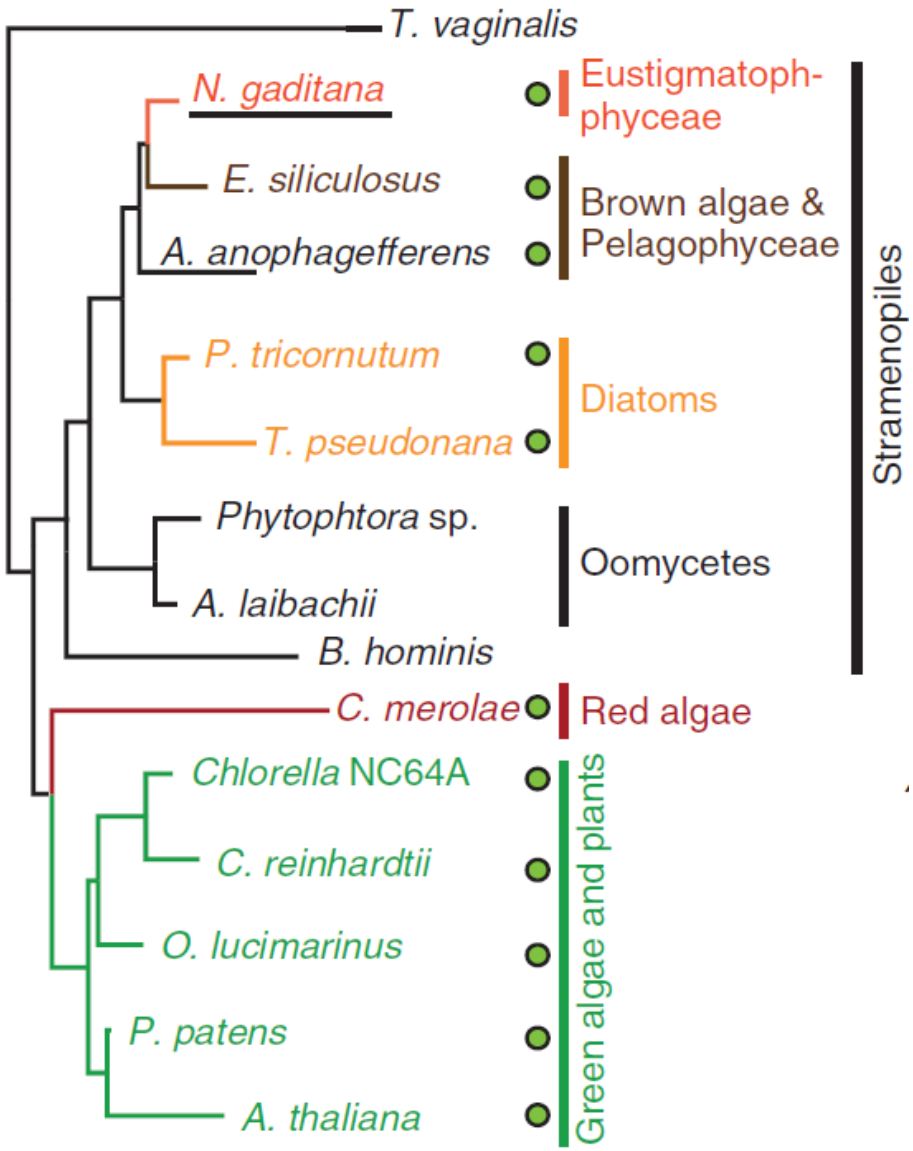


Figure 1. Schematic phylogenetic tree of stramenopiles and photosynthetic algae.

The tree was adapted from Baldauf et al[20].,Tyler et al[21]., by Radakovits et al., 2012[11] and is based on a concatenation of six highly conserved proteins (α-tubulin, β-tubulin, actin, and elongation factor 1-alpha)[20]. Filled green circles on the right indicate photosynthetic species. (Radakovits et al., 2012[11]).

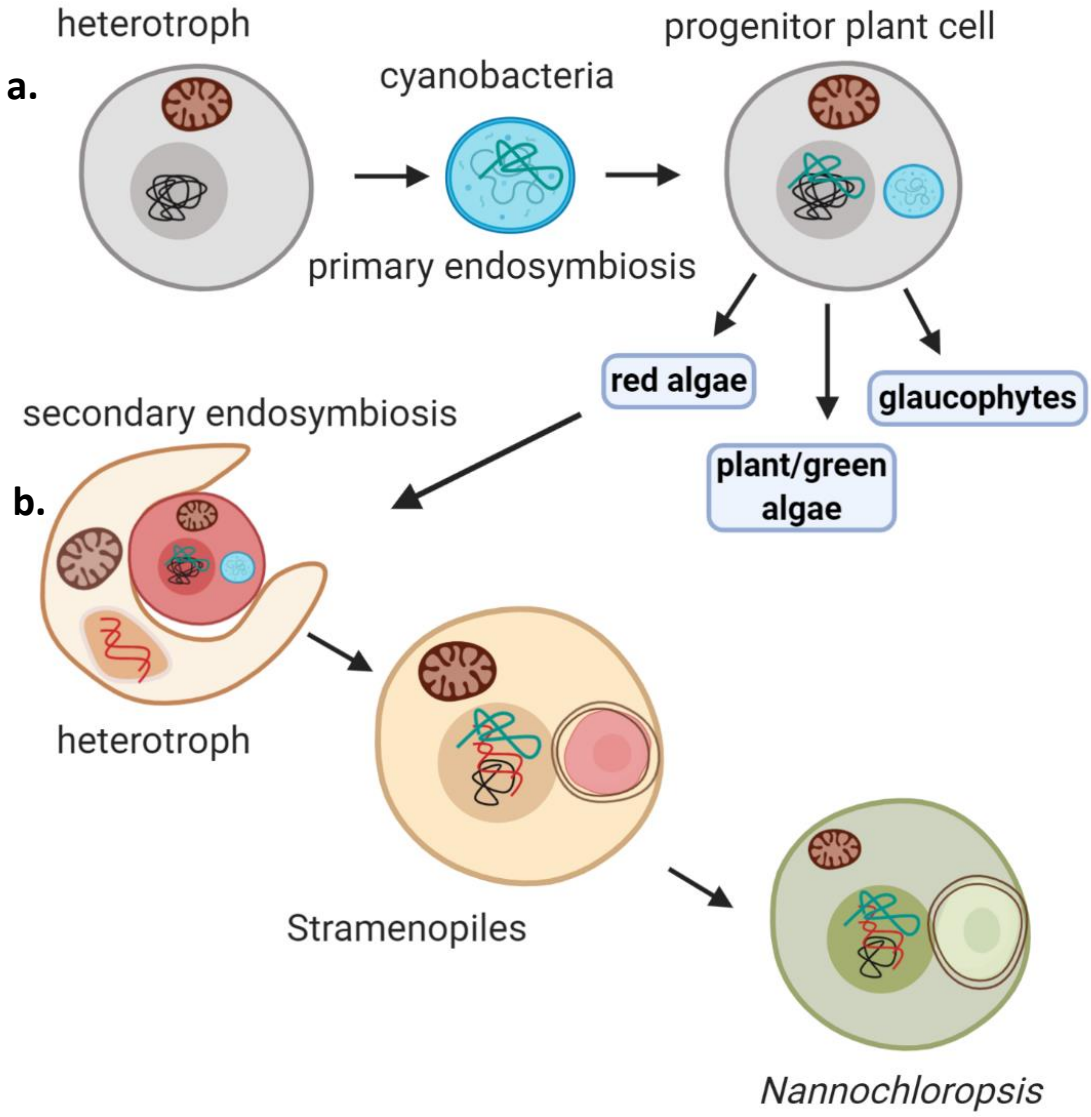


Figure 2. Diagram showing primary and secondary endosymbiosis events resulting in the photosynthetic forms with plastids.

a. Primary endosymbiosis where a cyanobacterial cell was engulfed by a heterotroph without cellular digestion leading to transfer of most of genomic content to the host nucleus forming the progenitor plant cell. **b**. One of the resulting algal cells (red algae) is engulfed by a eukaryotic phagotroph in an independent secondary endosymbiosis leading to origin of stramenopiles bearing chloroplast with four enveloping membranes[13,15,17].

These microalgae are also very robust due to the presence of a thick cell wall which comprises of an outer algaenan layer protecting an inner cellulose layer[22] and can hence tolerate a wide range of environmental conditions with regards to pH, temperature and salinity[11]. The currently recognised species are *N. gaditana*[11], *N. limnetica*[23], *N. oceanic*[12]*a, N. oculata*[1], *N. salina*[1], *N. granulata*[5,24] and *N. australis*[25].

Table 1 presents the taxonomic classification of the genus *Nannochloropsis*. *Nannochloropsis* spp. are non-motile spheres, morphologically indistinguishable under a light microscope[26]. The genome is haploid (n) and compact[27]. It is about 30 Mbp (similar to *Phaeodactylum tricornutum* genome- 27 Mbp[28] but smaller than *Chlamydomonas reinhardtii*- 120 Mbp[29]) with a rather small content of repeated regions[30] and about 10,486 predicted protein-coding genes with a density of one gene every 2.6 kbp[30].

Table 1. Taxonomic classification of microalgae *Nannochloropsis*[31]

(source: http://www.algaebase.org)

Empire	**Eukaryota**
Kingdom	Chromista
Phylum	Ochrophyta
Class	Eustigmatophyceae
Order	Eustigmatales
Family	Monodopsidaceae
Genus	*Nannochloropsis*

Nannochloropsis spp. follow an autotrophic lifestyle[32–34]. They grow optimally at temperatures between 20-24 °C[35] and have an average doubling time of approximately 14 hours[36] with a biomass yield of 650± 140 $mg.L^{-1}$ d^{-1} [11].

Nannochloropsis is capable of synthesising large amounts of lipids in the form of triacylglycerols (TAG) and long-chain polyunsaturated fatty acids (LC-PUFA)[37,38] such as arachidonic acid (AA, 20:4 n−6), eicosapentaenoic acid (EPA, 20:5 n-3) [39] and docosahexaenoic acid (DHA, 22:6 n−3)[40]. Different *Nannochloropsis* spp. are grown in outdoor ponds and photo-bioreactors for aquaculture[41] due to high proportion of LC-PUFA content which is an essential constituent of human nutrition. They are also considered a promising species for biofuel production, due to their ability to accumulate

storage triacylglycerols (TAGs) under conditions of nitrogen (N) starvation and saturated light[42] making up to half of their dry weight[38]. *Nannochloropsis* is also rich in pigments such as chlorophyll a and carotenoids[4,43]. These microalgae are an important source of beta glucans (βG) which accounts for around 14-21% dry weight[44]. In general, βG have the ability to stimulate the immune system of cultured organisms (aquaculture)[45] and are classified as biological response modifiers i.e. activation of cytotoxic macrophages, helper T cells, and NK cells, promotion of T cell differentiation etc. in mouse and other animal models[44,46,47].

1.2 Genetic engineering of *N. gaditana* for functional genetics and biotechnology

Genetic engineering in this microalga is still in its infancy. The availability of genomic data for *N. gaditana* B-31, CCMP526[11,48], transcriptome data sets and transformation protocols[11,12] are crucial for this organism to develop genetic tools such as mutant libraries, transgenic over-expression, reporter protein fusions, marker-free knockouts, and high-capacity gene stacking[49]. The genetic tools ae required to build metabolic maps, construct regulatory networks for metabolic engineering, understand the biological processes using mutant libraries, gene knockouts or transgenic over-expression. Also, tools such as high-capacity gene stacking systems for multi-gene expression are essential to produce novel compounds such as terpenoids (a diverse group of plant secondary metabolites[50,51]) in the microalgal host.

Microalgae are new and promising candidates for production of high-value terpenoids due to the following advantages: i) they are eukaryotic, ii) they have the capacity to carry out complex biological functions, iii) have inexpensive growth requirements compared to bacterial and mammalian systems. iv) the advances in genomic, transcriptomic and metabolomics of model microalgal species have led to discovery of new genetic tools for gene manipulation and metabolic engineering[49,52,53]. v) development of genome-scale metabolic models[54–58] for a few model species has enabled microalgae to emerge as a new contender to be a useful synthetic biology platform for high-value product synthesis[59].

The genetic tools specified above are dependent on a suite of promoters that could be used about a variety of growth conditions. Although few promoters have been

identified and used for transgene and reporter gene expression[49], it is still limited to a handful of growth conditions which calls for the need for new and improved promoters for future use. In general, these promoters along with the gene of interest are transformed into *N. gaditana* using plasmid-based transformation vectors, built in the laboratory in a suitable *Escherichia coli* host using recombinant DNA technology[60]. These vectors can be purified and used for transforming any desired host, which in this case is microalga *N. gaditana*.

1.2.1 *Nannochloropsis* transformation

Although transformation of bacterial and yeast systems is rather simple[61–64], the same does not apply for algae and plants. Bacterial and yeast cloning require competent cell preparation followed by heat shock or electroporation with the required plasmid. The entry of a desired plasmid into a plant cell is not simple, due to the presence of a cellulosic cell wall and other tough materials like lignin. Scientists have found multiple ways to overcome this barrier. For example, protoplasts can be prepared by digesting the cell wall using a mixture of cellulases and then treat the naked cells with recombinant DNA[65,66]. Agrobacterium-mediated transformation can be performed on plant callus, leaf discs, root tips or floral dip[67–69]. Also, micro-projectile bombardment (biolistics gene gun) using recombinant DNA coated gold or tungsten particles[70,71]. These same methods have been modified and adapted to the microalgae *Nannochloropsis* (details given below).

Nannochloropsis spp. transformation methods include electroporation using high voltage field strength of 12,000 V cm^{-1}. In *N. gaditana*[11] using the β-tubulin, heat shock protein 70 and the ubiquitin extension protein promotors with a transformation efficiency of 12.5 × 10^{-6} colonies. electroporated $cell^{-1}$, 27.8 × 10^{-6} colonies. electroporated $cell^{-1}$ and 166.7 × 10^{-6} colonies. electroporated $cell^{-1}$ respectively, was achieved. In *N. oceanica*[12] using LDSP promoter a transformation efficiency of 1.25x 10^{-6} per µg plasmid DNA was achieved. Biolistic transformation using 0.6 µm gold particles at 1550 psi[72] generated 25 *Nannochloropsis* colonies, *Agrobacterium*-mediated transformation of *Nannochloropsis* strain UMT-M3[73] using alternative phenolic compounds like cinnamic acid, coumarin and vanillin was also successful. Also, the episome-based, non-integrative expression of transgenes has been demonstrated to be

feasible with *Nannochloropsis spp*.,[74,75] similar to diatoms[76] and could provide a more stable platform for future studies and synthetic biology applications. Homologous recombination in *Nannochloropsis* strain W2J3B was reported by Kilian et al[36] and in *N. oceanica* by Gee et al[77], which is advantageous for site directed transgene integration. Other genetic modification tools have been tested as well such as transformation via electroporation using PCR product[78], over-expression of endogenous genes[38], heterologous expression of important gene product[79,80], and an inducible expression strategy for the Cre recombinase in *Nannochloropsis gaditana*[75].

1.2.2 Role of a promoter region

A gene promoter is a sequence of DNA that is located upstream of the transcription start site (TSS) of the gene. It also has multiple cis-acting elements (regions of non-coding DNA which regulate the transcription of neighbouring genes[81]) where specific proteins involved in transcription initiation and regulation can bind[82] leading to activation or repression of the gene[83]. In general, the eukaryotic promoter would comprise of two regions- proximal and distal as shown in Figure 3. The proximal or core region is closer to the TSS which can be identified by the presence of conserved elements such as TATA box at ~35 bp, CCAAT box at ~80 bp and GC box at ~200 bp above the TSS. These are predominant features in eukaryotic promoters of mammals, but not all plant promoters[83]. The regions present far away from the coding region of the gene, yet able to control its expression forms the distal part of the promoter which comprises of regulatory domains.

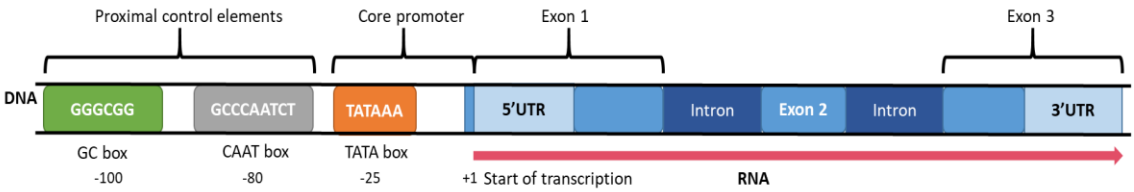


Figure 3. Model of a eukaryotic gene along with its promoter.

The gene has the 5'and 3' UTR, exons, and introns. The promoter contains the core and proximal control elements. These elements are recognition sites for transcription factors to bind which in turn aids the binding of RNA polymerase to start transcription of the downstream gene.

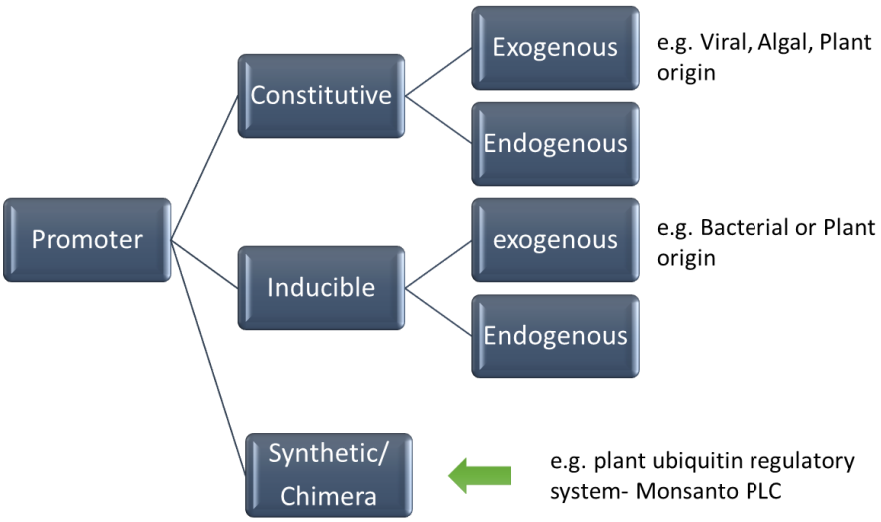


Figure 4. Schematic representation of the different types of promoters.

The gene promoter can be of different types based on function, origin, and activity.

The promoter sequences and their contributing elements are critical for fine regulation of introduced transgenes[82,83]. Promoters that continuously drive the expression under a wide range of conditions are called "constitutive" (Figure 4). If endogenous, they are mostly promoters of highly expressed genes that maintain the structural and functional integrity of the host cell. In plants, examples of constitutive promoters include- the *Arabidopsis* ubiquitin promoter[84], the soybean polyubiquitin promoter[85] and the strong MtHP promoter from the legume *Medicago truncatula*[86].

Table 2. Available promoters for *Nannochloropsis* spp.

Citation	Strain(s)	Promoter(s)	Type
Vieler et al. (2012)	*N. oceanica* CCMP1779	LDSP	Constitutive
Radakovits et al. (2012)	*N. gaditana* CCMP526	β-tub, UEP, HSP	Constitutive
Kilian et al. (2011) Moog et al. (2015)	*N. oceanica* W2J3B, *N. oceanica* CCMP1779	VCP	Bidirectional
Poliner et al. (2017)	*N. oceanica* CCMP1779	EF, Ribi	Constitutive, bidirectional
Zienkiewicz et al. (2017)	*N. oceanica* CCMP1779	EF	Constitutive

Kaye et al. (2015)	*N. oceanica* CCMP1779	LDSP	Constitutive
Li et al. (2016a)	*N. oceanica* CCMP1779	Hsp20	Constitutive
Wei et al. (2017b)	*N. oceanica* IMET1	β-tub, HSP70	Constitutive
Kang et al. (2017)	*N. salina* CCMP1776	β-tub, UEP	Constitutive
Kwon et al. (2017)	*N. salina* CCMP1776	β-tub, UEP	Constitutive
Jackson et al. (2019)	*N. gaditana* CCMP526	NR	inducible

In *Nannochloropsis* spp. few constitutive promoters have been identified[49], and the most frequently used is the β-tubulin promoter[11] (Table 2). Heterologous constitutive promoters on the other hand are mostly of viral origin[87] especially in plants and mammals. For example, the cauliflower mosaic virus (CaMV) 35S promoter is a plant nuclear promoter system which is constitutively active in several plant species[88,89] and for expression in mammalian cells viral promoters derived from cytomegalovirus (CMV)[90,91] or simian virus 40 (SV40)[92,93] have been used regularly. Strong expression was not observed with these viral promoters in microalgal hosts[94,95].

Alternatively, inducible promoters are not active always, but respond to an external trigger which can be environmental, stress-related, or chemical which activate the expression of the gene. In plants, for example the nopaline synthase (*nos*) promoter is auxin and wound inducible[96], steroid inducible expression system uses glucocorticoid hormone[97] and several patented promoter sequences such as water-deficit inducible promoter[98], stress inducible promoters[99], disease-inducible promoters[100], nitrate-inducible promoter[101], pathogen-inducible promoters[102], drought-inducible promoters[103], cold-inducible promoters[104]. In *Nannochloropsis,* for example, promoter of *sulfoquinovosyldiacylglycerol synthase 2* (*SQD2*), which encodes the sulfoquinovosyl transferase that catalyses the second step of sulfolipid biosynthesis was used as an inducible promoter which responds to phosphorous starvation in *Nannochloropsis* strain NIES-2145[105] and a nitrate-inducible expression system in *N. gaditana*[106].

On the other hand, synthetic promoters are artificial sequences constructed by engineering cis elements, which include enhancers, activators, or repressors directly upstream of the core promoter sequence. They are designed to be compact and efficient carrying only the conserved functional regions. In plants, several synthetic and chimeric promoters have been developed[107]. In microalgae, for example, the synthetic algal promoters (SAPs) of *C. reinhardtii* based on the chimeric *hsp70/rbs2* promoter, which were used to drive the expression of fluorescent reporter in *C. reinhardtii*[108]. To date, no such promoters have been constructed for use in *Nannochloropsis* spp.

The selection of promoters for genetic engineering depends on variables like source (i.e. endogenous, synthetic or viral) and gene expression levels (i.e. constitutive or induced)[109]. Native promoters have the advantage of being recognised by the microalgal transcriptional machinery. For example, in *C. reinhardtii*, the 5´untranslated region (5'-UTR) of the *RuBisCO small subunit* gene (*rbcS*), *heat shock protein 70A* gene (*hsp70A*) and a phosphorous starvation-inducible promoter are some of the promoter regions studied so far[110–112]. Heterologous promoters have also been used in different types of microalgae such as viral promoters CaMV35S, SV40, and CMV[109] which are commonly used for expressing genes in plant and mammalian cell lines. One drawback of using such viral promoter systems is low transgene expression in microalgal systems such as *C. reinhardtii*[95] and *Nannochloropsis*[73].

Endogenous promoters have been used to successfully transform *N. gaditana* CCMP526 and *N. oceanica* CCMP1779. Radakovits et al., selected three promoters from genes encoding β-*tubulin (TUB, Nga00092)*, *heat shock protein 70 (HSP, Nga07210)* and the *ubiquitin extension protein (UEP, Nga02115.1)*, which drove the expression of *Sh ble* gene that confers resistance to zeocin and bleomycin[11]. Later, Kilian et al., patented a bidirectional promoter from the *violaxanthin/chlorophyll α-binding protein (VCP)* genes, *VCP1* (GenBank accession no. *JF957601)* and *VCP2* (GenBank accession no*. JF946490)*[113] which drove the expression of *Sh ble* gene that conferred resistance to bleomycin/zeocin[36]. Here *VCP1* is unidirectional and *VCP2* is bidirectional and are light-inducible in nature. Another nuclear transformation method developed by Vieler et al., used a promoter sequence of the *stress-inducible endogenous lipid droplet surface protein* (*LDSP*, *NannoCCMP1779_4188*)[12] which drove the strong induction of the

hygromycin B resistance gene under nitrogen starvation conditions. Then in 2017, Poliner et al., in their over-expression study used a unidirectional *elongation factor* (*EF*, *NannoCCMP_10181*) promoter and bidirectional *Ribi* promoter, which was found in the intergenic region between two ribosomal subunit genes (*NannoCCMP_9668* and *NannoCCMP_9669*) to express multiple reporters and target transgenes in *N. oceanica*[114]. Recently, Ajjawi et al., used CRISPR/Cas9 and RNAi gene editing tools for disruption of genes in *N. gaditana* CCMP1894 and used novel endogenous promoters (*TCT, initiation factor 4AIII, 60S ribosomal protein L24, EIF3*) for reporter gene expression[115] while Jackson et al., demonstrated the use of an inducible expression system using nitrate with *N. gaditana*[106].

Promoters of varying strength whether inducible or constitutive are required to modulate the expression of transgenes and entire metabolic pathways, to ultimately engineer the production of valuable bio-products from these microalgae. For example, in plants, multiple use of a single promoter can induce gene silencing leading to transgene suppression[116]. Despite the availability of multiple promoters that could drive transgene expression in *N. gaditana,* a repertoire of functional promoters is limited in terms of diversity in expression range and inducibility. Further, a lack of knowledge of specific regulatory elements that control transcription in *Nannochloropsis spp.* that would allow for predictable expression and rational design of promoter regions. To date, several studies involving gene overexpression have been conducted in *Nannochloropsis* using endogenous promoters[117] but in-depth promoter analysis done in other model microalgae like *Phaeodactylum*[118–122] and *Chlamydomonas*[111] is still missing with *Nannochloropsis spp.*

1.2.3 Promoter activity assay

Experimental analysis of promoters and their putative elements can be carried out using transformation and transgene expression using reporter gene expression assay. Initially, the promoter activity was linked to expression of antibiotic resistance genes which was subject to time consuming protein assays, for example plant promoters expressing chloramphenicol acetyltransferase (CAT) used protein assay to measure CAT activity using [^{14}C] chloramphenicol[123]. Later, promoter analysis was done using fluorescent reporter genes such as *green fluorescent protein* (*GFP*)[124] or mVenus, which

is a variant of yellow fluorescent protein[125] or bioluminescent reporter like luciferase or histochemical reporter such as β-glucuronidase (GUS)[126]. Fluorescent reporters have the advantage of being used in high throughput assays whether it is using single cell imaging or flow cytometry or both[127,128]. Barnard et al., developed a split-GFP-based bimolecular fluorescence complementation (BiFC) assays in yeast and can be performed in living cells to provide information on protein interaction and sub-cellular location and can be performed in living cells to provide information on protein interaction and sub-cellular location[129]. Charvin et al., used fluorescence reporters for long-term fluorescent imaging in unperturbed dividing yeast cells with a microfluidic device[130]. Cormack et al., used an enhanced GFP as reporter for gene expression in *Candida albicans*[131].

On the other hand, *GUS* gene fusion systems have been used extensively in plants for spacio-temporal expression studies[132–135] as the expression of this gene is not deleterious to the plant host. The *GUS* gene codes for a β-glucuronidase (E.C.3.2.1.31) enzyme which has the catalytic activity to hydrolyse a wide variety of glucuronides including histochemical substrates such as 5-bromo-4-chloro-3-indolyl β-D- glucuronide (X-Gluc)[126] and can be used in laboratory model systems lacking endogenous GUS activity such as *Saccharomyces cerevisiae*, *Caenorhabididtis elegans*[126,136].

Luciferases catalyse the oxidation of substrate luciferins to yield non-reactive oxyluciferins with a release of light[137–139]. For example, the firefly (*Photinus* or *Luciola*) luciferase gene codes an enzyme that catalyses the light producing, adenosine triphosphate (ATP) dependent oxidation of luciferin[137,140,141]. Hence it can be used as an ATP assay as shown in Figure 5. This makes it extremely sensitive and measurable parameter using instruments like a microplate reader. In plants for example, a dual luciferase assay system was used to analyse promoter sequences[142], for analysis of viral promoter activity[143]. In microalgae, for example, in *C. reinhardtii* a synthetic luciferase gene derived from *Renilla reniformis* (RLuc) and adapted to the nuclear codon usage of *C. reinhardtii* was used to efficiently monitor the expression profile of selected nuclear genes using the standard bioluminescent assay with a micro-plate luminometer using automated substrate injection[144], and in *Nannochloropsis*, a codon optimised *NanoLuc* from deep sea shrimp[114] was used. This gene was expressed alone or fused to the zeocin resistance gene Sh *ble* under control of different promoter variants[114,144].

Another group lead by Bock has tested the luciferase gene from the marine copepod *Gaussia princeps* (*G-Luc*) for its suitability as a sensitive bioluminescent reporter of gene expression in *Chlamydomonas*[145]. They showed that this codon optimised version of the marker gene served as a highly sensitive reporter of gene expression from both constitutive and inducible algal promoters[145].

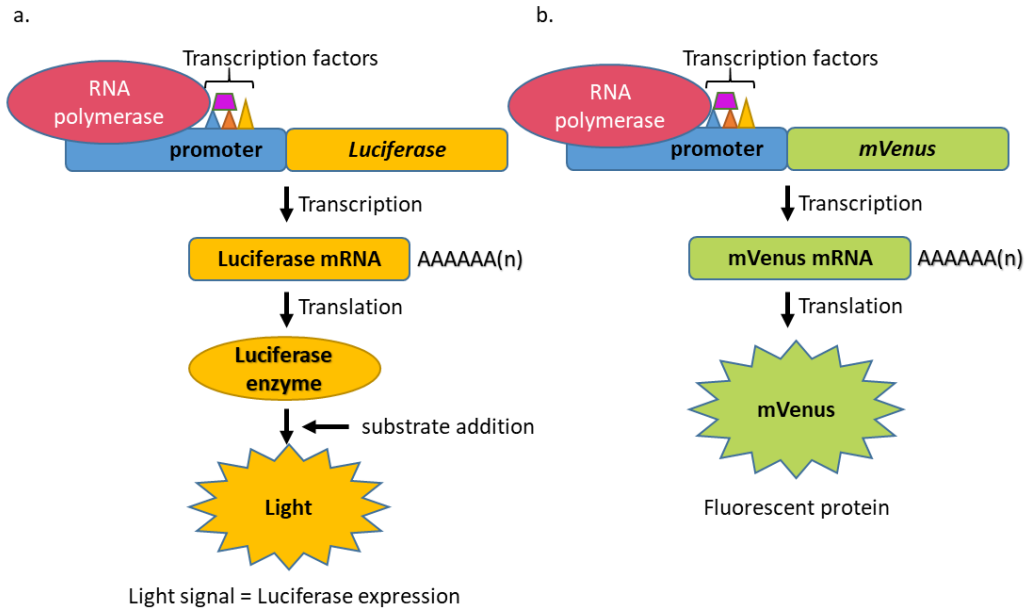


Figure 5. Schematic representation of the luciferase and reporter assay and mVenus expression.

a. In this assay the promoter strength is proportional to the amount of luciferase expression which can be measured as the amount of light emitted after substrate addition. **b**. the promoter activity is measured using reporter (mVenus) fluorescence.

Lauersen et al, have developed a modular vector toolkit in the model microalga *C. reinhardtii* where they have used the *Gaussia princeps* luciferase, as well as bright cyan, green, yellow and red fluorescent protein variants[146]. A study conducted by Heitzer group, proved *Gaussia*-luciferase to be a superior quantifiable reporter gene for the analysis of constitutive promoter sequences in *C. reinhardtii*[147].

Even though the green fluorescent protein (GFP) from *Aequorea victoria* has become a standard reporter in many biological systems, its use in higher plants was limited due to aberrant splicing and protein instibility[148]. Therefore, Davis et al., developed a soluble, highly fluorescent variants of green fluorescent protein[148]. Quaedvlieg et al., highlighted the use of bifunctional reporters (fusions between green

fluorescent protein and β-glucuronidasein) in higher plants[149]. Birnbaum et al., used GFP fluorescence to sort protoplasts from fluorescent reporter lines using fluorescence-activated cell sorter (FACS), where flow cytometry (a technique that simultaneously measures and analyses multiple physical characteristics of cells such as size, relative granularity, and relative fluorescence intensity as they flow in a single stream through a light beam[150]) was used to separate GFP expressing cells[151].

The *Nannochloropsis* promoters studied so far have driven the expression of antibiotic resistance genes like *sh-ble* (zeocin) and hygromycin[11,12,36] and reporter genes (e.g. firefly *luciferase, NanoLuc* from deep sea shrimp [114], *Cerulean*, *Venus* [125] but in-depth promoter analysis done in other model microalgae like *Phaeodactylum* [118–122] and *Chlamydomonas*[111] is missing with *Nannochloropsis spp*. Although, all the reporter assay systems mentioned above could be used for promoter analysis in this microalga, we were specifically looking for an inexpensive, high throughput analysis system using flow cytometry, hence we selected the fluorescent reporter system (mVenus) for measuring promoter activity. We intend to identify and characterize a suite of new promoters in this PhD research that would benefit the broader research community to explore the use of this microalga for recombinant protein production or high value bio-products such as terpenoid biosynthesis.

1.3 Plant terpenoids

Terpenoids (also known as terpenes or isoprenoids) are the largest group of natural products[152,153]. In plants, they are of chemically diverse set of secondary metabolites which are associated with developmental physiology, mutualistic and antagonistic plant–herbivore and plant–environment interactions[152,154]. Over 55,000 different terpenoids have been isolated[155] and this number is likely to increase over the years. In nature, terpenes occur predominantly as hydrocarbons, alcohols and their glycosides, ethers, aldehydes, ketones, carboxylic acids and esters[155]. Terpenoids are classified based on the number of isoprene units (C_5H_8) and the common ones are divided into mono-, sesqui-, and di-terpenes (C10, C15 and C20, respectively)[154]. Some of the essential roles played by plant terpenoids are i) photosynthetic metabolism, including light harvesting (phytol tail of chlorophyll), electron transfer (plastoquinone), and photoprotection (carotenoids), ii) respiration (ubiquinone), iii) regulation of

membrane structure and fluidity (sterols), and developmental regulation (phytohormones)[156]. While these terpenoids act mostly within the organism, several other terpenoids act between organisms as repellents, chemo-attractants, and toxins and play a significant role in the organism's defence mechanism and survival[157–159].

The universal building blocks of all terpenoids are isopentenyl pyrophosphate (IPP), and dimethyl allyl pyrophosphate (DMAPP); see Figure 6[156]. Two different enzymatic pathways have evolved to generate these two precursors, they are the mevalonate (MVA) pathway and the methyl-D-erythritol 4-phosphate (MEP) pathway[160,161]. Plants and some groups of algae (for example, red algae- *Galdieria sulphuraria* and *Cyanidium caldarium*; diatoms- *Phaeodactylum tricornutum*, *Nitzchia ovalis*; Chrysophyceae- *Ochromonas danica*[162]) use the cytosolic MVA pathway, but also have the bacterial MEP pathway for terpenoid production in the plastid[162]. Yet many unicellular algae have lost the MVA pathway after acquiring the MEP pathway and are solely dependent on the MEP pathway to produce isoprenoids such as the green alga *C. reinhardtii*, and the stramenopile *Nannochloropsis* sp[156,160].

The IPP units are added to a DMAPP precursor in a condensation reaction catalysed by prenyltransferases to generate pyrophosphate molecules that vary in chain length: 1) C10 geranyl pyrophosphate (GPP), 2) C15 farnesyl pyrophosphate (FPP), and 3) C20 geranylgeranyl pyrophosphate (GGPP)[156,160]. These prenyl pyrophosphates are the precursors of terpene synthases (TPSs), where GPP is the precursor to monoterpenes, FPP to sesquiterpenes, and GGPP to diterpenes, carotenoids, etc[156,160].

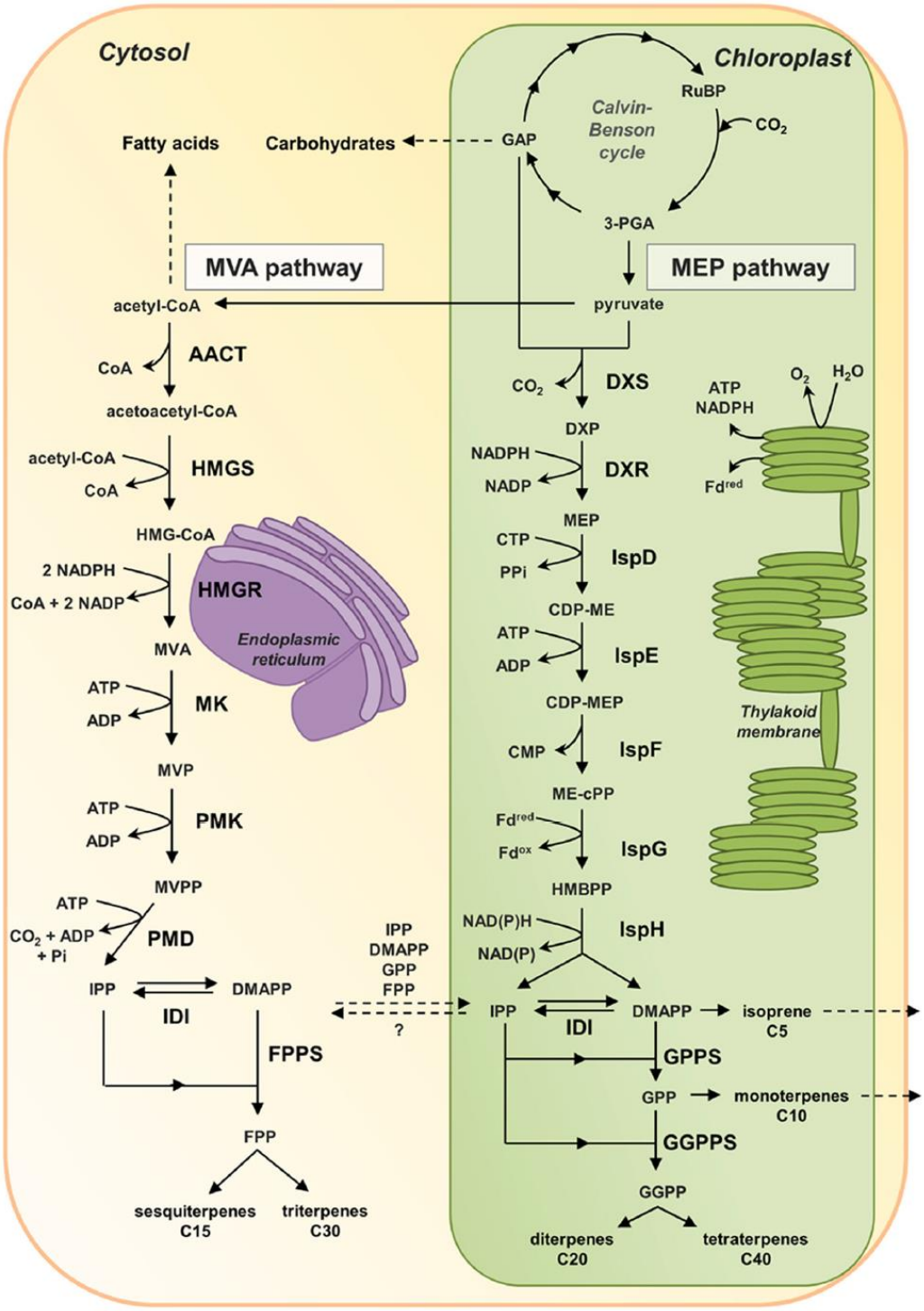


Figure 6. Subcellular compartmentalization of the MVA and MEP pathways in the plant cell[156].

The cytosolic MVA pathway generates IPP/DMAPP which is used for the synthesis of sesquiterpenes and triterpenes, whereas the plastidial MEP pathway produces IPP/DMAPP for the synthesis of isoprene, monoterpenes, diterpenes, and tetraterpenes. The photosynthetic machinery of the chloroplast generates energy and reducing power such as ATP, NADPH, and

reduced ferredoxin (Fdred) cofactors that are utilised by the MEP pathway. MVA pathway enzymes are: AACT acetoacetyl-CoA thiolase, HMGS HMGCoA synthase, HMGR HMG-CoA reductase, MK mevalonate kinase, PMK mevalonate 5-phosphate kinase, PMD mevalonate 5-pyrophosphate decarboxylase, and IDI IPP isomerase. MEP pathway enzymes are: DXS 1-deoxy-D-xylulose 5-phosphate synthase, and DXR 1-deoxy-D-xylulose 5-phosphate reductase. The prenyltransferases (GPPS, GGPPS and FPPS) (Image taken from Davies et al.,[156])

1.3.1 Industrial applications of plant terpenoids

The terpenoid family is very diverse which has led to many commercial applications. They are used as disinfectants, insect repellents, agrochemicals, flavouring agents, skin penetration enhancers, biofuels, perfumes and fragrances apart from high-value pharmaceuticals for their anti-microbial, anti-fungal, anti-parasitic, anti-viral, anti-allergenic, anti-spasmodic, anti-hyperglycemic, anti-inflammatory, chemotherapeutic and immunomodulatory properties[154,155,163]. Plants are the primary producers of these structural diverse, ubiquitous secondary metabolites[50,51,164]. Generally large natural resources are required to obtain sufficient quantities of desired terpenoid, and the associated ecological and environmental impacts are staggering[59] as these high-value plant terpenoids are present in very low abundance. Also, the plant terpenoids are often produced as mixtures which is extremely difficult and expensive to extract[59]. For example, Taxol, an anti-cancer drug, which gained first marketing approval from the U.S. Food and Drug Administration for the treatment of refractory ovarian cancer in 1992, has a limited supply from the original source, the bark of the Pacific yew (*Taxus brevifolia*) whose yield constitutes only between 0.01 and 0.02 % of bark dry weight[165]. Taxol is now largely produced by *Taxus* cell culture methods or by semi-synthetic means using advanced precursors[165]. On the other hand, chemical synthesis of terpenoids requires multiple steps, is very complex and expensive[59].

Alternately, efficient and scalable production of terpenoids is possible due to synthetic biology in microbial hosts. Today, the industrially work-horse organisms- *E. coli* and *S. cerevisiae* are used for production of commercial terpenoids. For example, synthetic biology approaches have been used to create heavily engineered hosts which can produce terpenes such as artemisinic acid[166] (a sesquiterpene endoperoxide with potent antimalarial properties, naturally produced by the plant *Artemisia annua* but industrially produced by engineered *Saccharomyces cerevisiae* strains with fermentation titres of 25 grams per litre of artemisinic acid[166]) and taxadiene[167] (the

potent anticancer drug Taxol was first isolated from the *Taxus brevifoliathe* and taxadiene is the first committed taxol intermediate biosynthesized in an engineered *Escherichia coli* strain with yield of approximately 1 gram per litre[167]). However, these organisms have not evolved to synthesise complex terpenoids. Hence the focus is slowly shifting to photosynthetic organisms that are capable of producing a much larger diversity of isoprenoids[59]. In general, photosynthetic organisms use CO_2 as feedstock and are capable of linking photosynthesis with the production of heterologous metabolites[59,168]. They can generate a surplus of redox power through photosynthesis[169], necessary for biosynthesis of terpenoids and other cellular functions. Also, the production of substantial amounts of isoprene by land plants and microalgae indicates a higher native flux capacity[59,170] which could be beneficial for heterologous terpene biosynthesis. Additionally, microalgae share evolutionary ancestry with higher plants and may contain more conducive cellular environments for plant terpene synthases than bacteria or yeast[171]. Also, algal cells are cultivated at temperatures similar to those used for plant growth[171]

1.3.2 Terpenoid biosynthesis in photosynthetic microorganisms

Cyanobacterial hosts have simple cellular organization, photosynthetic growth and have been involved in the production of terpenes for example, isoprene in *Synechocystis* sp. PCC 6803[172], *Synechococcus elongatus* PCC 7942[173], limonene/bisabolene in *Synechococcus* sp. PCC 7002[174], alpha-bisabolene production in *Synechocystis* sp. PCC 6803[175] and (E)-α-bisabolene under high-density conditions in *Synechocystis* sp. PCC 6803[176]. Eukaryotic, unicellular microalgae have a huge biodiversity and are largely unexplored[177].

Marine microalgae are well suited for large scale production due to a variety of reasons: i) growth in salt water ii) does not compete with land plants for fertile land space iii) fast turn over compared to plants iv) photosynthetic production using CO_2 as carbon source v) no exogenous administration of carbohydrate feedstock needed, as required for heterotrophic microorganisms (bacteria, yeast). Few microalgal species have been used for industrially relevant terpenoid production including the green alga *C. reinhardtii* and the diatom *P. tricornutum*[59]. The terpenes (E)-alpha-bisabolene, patchoulol and manoyl oxide have been produced in *C. reinhardtii*[178–180]. Lauersen et al.,

reported the expression of a patchoulol synthase from *Pogostemon cablin* in green microalga *C. reinhardtii* yielding 1.03 mg L^{-1} sesquiterpenoid patchoulol under photoautotrophic batch cultivation[180] and suggested that enzyme titre, rather than substrate (FPP) availability was the limiting factor for increasing sesquiterpene product yield[180]. Wichmann et al., demonstrated successful expression of *Abies grandis* bisabolene synthase and produced the sesquiterpene biodiesel precursor (E)-α-bisabolene to 10.3±0.7 mg g^{-1} cell dry weight under mixotrophic batch cultivation[178]. They identified pathways competing for carbon using artificial microRNA-based knock-down, which revealed two carbon flux gatekeeping targets[178]. Lauersen et al., also reported the nuclear transgene expression of codon optimised diterpene synthases 2 and 3 (CfTPS2 and CfTPS3) from *Coleus forskohlii* with synthetic interon and showed 80 mg 13R (+) manoyl oxide g^{-1} cell dry mass which was hydroxylated in the chloroplast using a truncated plant microsomal cytochrome P450 monooxygenase (CYP)[179]. They demonstrated that the *C. reinhardtii* cell had a natural tolerance to increased flux through the MEP pathway upon overexpression of non-native DXS and DXR from *Salvia pomifera* but overexpression of the protein ERG20, a GGPP synthase resulted in increased production of freely available GGPP substrate and noted that CfTPS2 does not outcompete native mechanisms for GGPP channelling[179]. D'Adamo et al., reported the production of the triterpenes betulin and its precursor lupeol in *P. tricornutum* by introducing three plant enzymes (a oxidosqualene cyclase from Lot*us japonicus* and a cytochrome P450 along with its native reductase from *Medicago truncatula*) and obtained betulinic acid titres of >0.1 mg g^{-1} dry biomass[181]. Recently, Papaefthimiou et al., expressed plant labdane-type diterpene cyclase, copal-8-ol diphosphate synthase from *Cistus creticus* in the chloroplast genome of *C. reinhardtii* and obtained labdane diterpenes at concentration 1.172 ± 0.05 µg mg^{-1} cell dry weight[182]. The terpenoid biosynthesis potential of *Nannochloropsis spp.* are yet to be explored.

Microalgal systems that have been explored for terpene synthesis are mentioned above and they are especially interesting hosts for engineering heterologous isoprenoid product biosynthesis by using native isoprenoid metabolism. However, achieving heterologous terpene biosynthesis in a microalgal host depends on several factors such as availability of characterised specific TPSs and CYPs needed for isoprenoid

biosynthesis[171]. Along with understanding the endogenous flux regulation of the isoprenoid pathways in the microalgal host[59]. Also understanding host endogenous gene silencing mechanisms of nuclear transgenes[59], boosting transgene expression rates and identifying competing pathways and gene-knockout targets[171,179] are vital. As well as a need for high-throughput screening to identify the best performing strains[59], identifying available substrate pool and means for providing sufficient pools for TPSs[59,171]. Future metabolic advancement in these photosynthetic microalgal platforms can provide more production options for expression of more complex plant pathways[59].

1.3.3 Is *Nannochloropsis* capable of heterologous terpenoid biosynthesis?

Nannochloropsis gaditana is an autotrophic microalgal species with several advantages for transgene expression, and novel terpenoid production. Firstly, microalgae generally have faster growth rates when compared to plant systems[11,36,70,71]. Secondly, they are not prone to contamination from pathogenic human viruses[183]. Thirdly, there is less risk of contamination by airborne contaminants because algae can be grown in sealed bioreactors, thus protecting the growth environment from transgene flow.

Heterologous production of terpenoids is not straight forward in photosynthetic microorganisms[59]. Even though considerable progress has been made in genetic and metabolic engineering of various photosynthetic hosts[59,171]. The knowledge about regulatory mechanisms that govern various metabolic pathway is limited, biomass and production of bio-products under a variety of growth conditions (auto/hetero/mixotrophic), temperature (cold/warm), light regimes (light: dark/continuous), salinity (hyper/hypo), nutrient stress (nitrogen/ phosphorous) is growing rapidly but not complete. The more dedicated genetic tools become available, the better we can understand the biology which would in turn aid in faster and easier engineering of these organisms. The aim being the development of algal bio-factories that could produce industrially relatable titres, and yields.

However, for *Nannochloropsis* the enzymes in the MEP pathway and the various diphosphate synthases have not been characterised yet, hence their substrate specificity, function, enzyme structure, kinetics, sub-cellular localisation and transport

across the chloroplast membrane are yet to be experimentally verified. One study on regulation of sterol biosynthesis in *N. oceanica* by Lu et al., 2014 identified the presence of genes for a full MEP pathway for the production of IPP (precursor) for isoprenoids, whereas a gene encoding hydroxy-methyl-glutaryl-CoA synthase (HMGS) was the only enzyme from the MVA pathway that was identified[184]. The study proposed a feedback system in place for the regulation of sterol and FA homeostasis[184]. Future research in this field would aid in designing effective metabolic strategies for terpenoid production in this microalga.

1.4 Thesis aims and objectives

The overall aim of this thesis is to develop the oleaginous microalga *Nannochloropsis gaditana* to produce industrially relevant bio-products by identifying and studying novel endogenous (both constitutive and inducible) and exogenous (viral) promoters and use them for heterologous gene expression in *N. gaditana*. By understanding the promoter activity at the transcript and protein level the best candidate promoters were used to engineer *N. gaditana* for the production of β-caryophyllene, a commercially relevant sesquiterpenoid with applications in perfume industry, beer brewing and high-density fuels.

Objectives

Identification and characterisation of novel endogenous promoters from *N.gaditana* CCMP526. To identify and assess the promoter activity of selected endogenous constitutive promoters in *N. gaditana* across the different growth stages and compare the promoter activity of these endogenous promoters with endogenous β-tubulin control promoter

Identification and characterisation of heterologous viral promoters and endogenous phosphate-inducible promoters for genetic engineering *N.gaditana* CCMP526. 1) To identify and examine the promoter activity of selected putative viral promoters in *N. gaditana* across the different growth stages and compare the promoter activity of these viral promoters with endogenous control promoter*;* then assess its use as a universal promoter in stramenopiles. **2)** To identify and analyse the promoter

activity of selected putative phosphate-inducible promoters in *N. gaditana* under phosphate-replete and deplete conditions.

Explore *N. gaditana* as a novel platform for terpenoid production. To computationally reconstruct the hypothetical terpenoid metabolism in *N. gaditana* and use the best candidate promoter from this study, for terpenoid production in *N. gaditana* using a caryophyllene synthase gene from *Artemisia anna* in *N. gaditana* and estimate the amount of β-caryophyllene produced in transgenic *N. gaditana* culture.

1.5 References

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

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

3. Whittle, S. J. & Casselton, P. J. The chloroplast pigments of the algal classes Eustigmatophyceae and Xanthophyceae. II. Xanthophyceae. *Br. Phycol. J.* **10**, 192–204 (1975).

4. Lubián, L. M. *et al.* Nannochloropsis (Eustigmatophyceae) as source of commercially valuable pigments. *J. Appl. Phycol.* **12**, 249–255 (2000).

5. Tibbetts, S. M., Bjornsson, W. J. & McGinn, P. J. Biochemical composition and amino acid profiles of *Nannochloropsis granulata* algal biomass before and after supercritical fluid CO_2 extraction at two processing temperatures. *Anim. Feed Sci. Technol.* **204**, 62–71 (2015).

6. Ma, X. N., Chen, T. P., Yang, B., Liu, J. & Chen, F. Lipid production from *Nannochloropsis*. *Mar. Drugs* **14**, (2016).

7. Alboresi, A. *et al.* Light remodels lipid biosynthesis in *Nannochloropsis gaditana* by modulating carbon partitioning between organelles. *Plant Physiol.* **171**, 2468–2482 (2016).

8. Hoffmann, M., Marxen, K., Schulz, R. & Vanselow, K. H. TFA and EPA productivities of *Nannochloropsis salina* influenced by temperature and nitrate stimuli in turbidostatic controlled experiments. *Mar. Drugs* **8**, 2526–2545 (2010).

9. Meng, Y. *et al.* The characteristics of TAG and EPA accumulation in *Nannochloropsis oceanica* IMET1 under different nitrogen supply regimes. *Bioresour. Technol.* **179**, 483–489 (2015).

10. Chen, C. Y., Chen, Y. C., Huang, H. C., Ho, S. H. & Chang, J. S. Enhancing the production of eicosapentaenoic acid (EPA) from *Nannochloropsis oceanica* CY2 using innovative photobioreactors with optimal light source arrangements. *Bioresour. Technol.* **191**, 407–413 (2015).

11. Radakovits, R. *et al.* Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropis gaditana*. *Nat. Commun.* **3**, 686 (2012).

12. Vieler, A. *et al.* Genome, Functional Gene Annotation, and Nuclear Transformation of the Heterokont Oleaginous Alga *Nannochloropsis oceanica* CCMP1779. *PLoS Genet.* **8**, (2012).

13. McFadden, G. I. Primary and secondary endosymbiosis and the origin of plastids. *J. Phycol.* **37**, 951–959 (2001).

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

15. Reyes-Prieto, A., Weber, A. P. M. & Bhattacharya, D. The origin and establishment of the plastid in algae and plants. *Annu. Rev. Genet.* **41**, 147–168 (2007).

16. Archibald, J. M. & Keeling, P. J. Recycled plastids: A 'green movement' in eukaryotic evolution. *Trends Genet.* **18**, 577–584 (2002).

17. Janouskovec, J., Horák, A., Oborník, M., Lukes, J. & Keeling, P. J. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10949–54 (2010).

18. Kroth, P. & Strotmann, H. Diatom plastids: Secondary endocytobiosis, plastid genome and protein import. *Physiol. Plant.* **107**, 136–141 (1999).

19. Agrawal, S. & Striepen, B. More Membranes, more Proteins: Complex Protein Import Mechanisms into Secondary Plastids. *Protist* **161**, 672–687 (2010).

20. Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science.* **290**, 972–977 (2000).

21. Tyler, B. M. *et al. Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science.* **313**, 1261–1266 (2006).

22. Scholz, M. J. *et al.* Ultrastructure and Composition of the *Nannochloropsis gaditana* Cell Wall. **13**, 1450–1464 (2014).

23. Freire, I. *et al. Nannochloropsis limnetica*: A freshwater microalga for marine aquaculture. *Aquaculture* **459**, 124–130 (2016).

24. Zhang, X. *et al.* First record of a large-scale bloom-causing species *Nannochloropsis granulata* (Monodopsidaceae, Eustigmatophyceae) in China Sea waters. *Ecotoxicology* **24**, 1430–1441 (2015).

25. Fawley, M. W., Jameson, I. & Fawley, K. P. The phylogeny of the genus *Nannochloropsis* (Monodopsidaceae, Eustigmatophyceae), with descriptions of *N. australis* sp. Nov. and *Microchloropsis* gen. Nov. *Phycologia* **54**, 545–552 (2015).

26. Fawley, K. P. & Fawley, M. W. Observations on the Diversity and Ecology of Freshwater *Nannochloropsis* (Eustigmatophyceae), with Descriptions of New Taxa. *Protist* **158**, 325–336 (2007).

27. Wang, D. *et al. Nannochloropsis* Genomes Reveal Evolution of Microalgal Oleaginous Traits. **10**, (2014).

28. Bowler, C. *et al.* The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–244 (2008).

29. Merchant, S. S. *et al.* The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions. *Science.* **318**, 245–250 (2007).

30. Corteggiani Carpinelli, E. *et al.* Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion. *Mol. Plant* **7**, 323–335 (2014).

31. Lubian, L. M. (*Nannochloropsis gaditana* sp. nov., a new marine Eustigmatophyceae, Cadiz Bay). *Lazaroa* **4**, 287–293 (1982).

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

33. Rocha, J. M. S., Garcia, J. E. C. & Henriques, M. H. F. Growth aspects of the marine microalga *Nannochloropsis gaditana*. *Biomol. Eng.* **20**, 237–242 (2003).

34. Poliner, E. *et al.* Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1779 under light / dark cycles. **83**, 1097–1113 (2015).

35. V, E. S. P. B., James, C. M. & Salman, A. E. Growth and Omega 3 Fatty Acid and Amino Acid Composition of Microalgae Under Different Temperature Regimes1. **77**, 337–351 (1989).

36. Kilian, O., Benemann, C. S. E., Niyogi, K. K. & Vick, B. High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proc. Natl. Acad. Sci.* **108**, 21265–21269 (2011).

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

38. Kaye, Y. *et al.* Metabolic engineering toward enhanced LC-PUFA biosynthesis in *Nannochloropsis oceanica*: Overexpression of endogenous δ12 desaturase driven by stress-inducible promoter leads to enhanced deposition of polyunsaturated fatty acids in TAG. *Algal Res.* **11**, 387–398 (2015).

39. Zittelli, G. C. *et al.* Production of Eicosapentaenoic acid by *Nannochloropsis* sp . cultures in outdoor tubular photobioreactors. **70**, 299–312 (1999).

40. Tonon, T., Harvey, D., Larson, T. R. & Graham, I. A. Long chain polyunsaturated fatty acid production and partitioning to triacylglycerols in four microalgae. *Phytochemistry* **61**, 15–24 (2002).

41. Ferreira, M., Coutinho, P., Seixas, P., Fábregas, J. & Otero, A. Enriching Rotifers with " Premium " Microalgae . *Nannochloropsis gaditana*. **11**, 585–595 (2009).

42. Sukenik, A., Carmeli, Y. & Berner, T. Regulation of fatty acid composition by irradiance level in the eustigmatophyte *Nannochloropsis* sp. *J. Phycol.* **25**, 686–692 (1989).

43. Bai, M. Der, Cheng, C. H., Wan, H. M. & Lin, Y. H. Microalgal pigments potential as byproducts in lipid production. *J. Taiwan Inst. Chem. Eng.* **42**, 783–786 (2011).

44. Jan, M. & Kazik, P. *Nannochloropsis: Biology, Biotechnological Potential and*

Challenges. (Nova Science Publishers, Incorporated, 2017).

45. Kumar, P., Jain, K. K., Munilkumar, S. & Chalal, R. S. Beta Glucan : Avaluable Nutraceuticalfor Promoting Health in Aquaculture. *African J. Basic Appl. Sci.* **5**, 220–227 (2013).

46. Bohn, J. A. & BeMiller, J. N. (1→3)-β-d-Glucans as biological response modifiers: a review of structure-functional activity relationships. *Carbohydr. Polym.* **28**, 3–14 (1995).

47. Veverka, M. *et al.* Beta-glucan complexes with selected nutraceuticals: Synthesis, characterization, and stability. *J. Funct. Foods* **8**, 309–318 (2014).

48. Jinkerson, R. E., Radakovits, R. & Posewitz, M. C. Genomic insights from the oleaginous model alga *Nannochloropsis gaditana*. *Bioengineered* **4**, (2013).

49. Poliner, E., Farré, E. M. & Benning, C. Advanced genetic tools enable synthetic biology in the oleaginous microalgae *Nannochloropsis* sp. *Plant Cell Rep.* **37**, 1383–1399 (2018).

50. Zwenger Chhandak Basu, S., Zwenger, S. & Basu, C. Plant Terpenoids: Applications and Future Potentials. *Biotechnol. Mol. Biol. Rev.* **3**, 1–007 (2008).

51. Singh, B. & Sharma, R. A. Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. *3 Biotech* **5**, 129–151 (2015).

52. Slattery, S. S. *et al.* An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in *Phaeodactylum tricornutum*. *ACS Synth. Biol.* **7**, 328–338 (2018).

53. Crozet, P. *et al.* Birth of a Photosynthetic Chassis: A MoClo Toolkit Enabling Synthetic Biology in the Microalga *Chlamydomonas reinhardtii*. *ACS Synth. Biol.* **7**, 2074–2086 (2018).

54. Kim, J. *et al.* Flux balance analysis of primary metabolism in the diatom *Phaeodactylum tricornutum*. *Plant J.* **85**, 161–176 (2016).

55. Fabris, M. *et al.* The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *Plant J.* **70**, 1004–1014 (2012).

56. Boyle, N. R. & Morgan, J. A. Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. *BMC Syst. Biol.* **3**, 1–14 (2009).

57. Shah, A. R., Ahmad, A., Srivastava, S. & Jaffar Ali, B. M. Reconstruction and analysis of a genome-scale metabolic model of *Nannochloropsis gaditana*. *Algal Res.* **26**, 354–364 (2017).

58. Chang, R. L. *et al.* Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol. Syst. Biol.* **7**, (2011).

59. Vavitsas, K., Fabris, M. & Vickers, C. E. Terpenoid metabolic engineering in photosynthetic microorganisms. *Genes (Basel).* **9**, (2018).

60. Lewin, B. *Lewin ' s GENES XI*. (Jones & Bartlett Learning, 2014).

61. Chung, C. T., Niemela, S. L. & Miller, R. H. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2172–2175 (1989).

62. Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).

63. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28 (1990).

64. Orr-Weaver, T. L., Szostak, J. W. & Rothstein, R. J. Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 6354–8 (1981).

65. McElroy, D., Zhang, W., Cao, J. & Wu, R. Isolation of an efficient actin promoter for use in Rice transformation. *Plant Cell* **2**, 163–171 (1990).

66. Rhodes, C. a, Pierce, D. a, Mettler, I. J., Mascarenhas, D. & Detmer, J. J. Genetically transformed maize plants from protoplasts. *Science* **240**, 204–207 (1988).

67. Toki, S. *et al. Agrobacterium*-mediated transformation of cereals: a promising approach crossing barriers. *Plant Cell Rep.* **22**, 612–617 (2004).

68. Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W. & Chua, N.-H. *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat. Protoc.* **1**, 641–646 (2006).

69. Sridevi, G., Parameswari, C., Sabapathi, N., Raghupathy, V. & Veluthambi, K. Combined expression of chitinase and β-1,3-glucanase genes in *indica* rice (Oryza sativa L.) enhances resistance against *Rhizoctonia solani*. *Plant Sci.* **175**, 283–290 (2008).

70. Christou, P. Genetic transformation of crop plants using microprojectle bombardment. *Plant J.* **2**, 275–281 (1992).

71. Christou, P., McCabe, D., Swain, W. F. & Barton, A. K. Particle-mediated transformation of Soybean plants and lines. *United States Patent* **19**, (1991).

72. Bertrand Vick, E. & Oliver Kilian, A. Transformation of Algal cells. (2014).

73. Cha, T. S., Chen, C. F., Yee, W., Aziz, A. & Loh, S. H. Cinnamic acid, coumarin and vanillin: Alternative phenolic compounds for efficient Agrobacterium-mediated transformation of the unicellular green alga, *Nannochloropsis* sp. *J. Microbiol. Methods* **84**, 430–434 (2011).

74. Poliner, E., Takeuchi, T., Du, Z.-Y., Benning, C. & Farre, E. Non-transgenic marker-free gene disruption by an episomal CRISPR system in the oleaginous microalga, *Nannochloropsis oceanica* CCMP1779. *ACS Synth. Biol.* **7**, 962–968 (2018).

75. Verruto, J. *et al.* Unrestrained markerless trait stacking in *Nannochloropsis gaditana* through combined genome editing and marker recycling technologies . *Proc. Natl. Acad. Sci.* **115**, E7015–E7022 (2018).

76. Karas, B. J. *et al.* Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* **6**, 6925 (2015).

77. Gee, C., Plant, K. N., Biology, M. & Berkeley, U. C. The carbonic anhydrase CAH1 is an essential component of the carbon-concentrating mechanism (CCM) of the marine alga, *Nannochloropsis oceanica*. *Proc Natl Acad Sci U S A.* **114**, 4537–4542 (2017).

78. Li, F., Gao, D. & Hu, H. High-efficiency nuclear transformation of the oleaginous marine *Nannochloropsis* species using PCR product. *Biosci. Biotechnol. Biochem.* **78**, 812–817 (2014).

79. Chen, H. L., Li, S. S., Huang, R. & Tsai, H. J. Conditional production of a functional fish growth hormone in the transgenic line of *Nannochloropsis oculata* (Eustigmatophyceae). *J. Phycol.* **44**, 768–776 (2008).

80. Kang, N. K. *et al.* Heterologous overexpression of sfCherry fluorescent protein in *Nannochloropsis salina*. *Biotechnol. Reports* **8**, 10–15 (2015).

81. Griffiths AJF, JH, M. & Suzuki DT, et al. Transcription: an overview of gene regulation in eukaryotes. in *An Introduction to Genetic Analysis.* (New York: W. H. Freeman, 2000).

82. Hernandez-Garcia, C. M. & Finer, J. J. Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci.* **217**–**218**, 109–119 (2014).

83. Porto, M. S. *et al.* Plant promoters: An approach of structure and function. *Mol. Biotechnol.* **56**, 38–49 (2014).

84. Callis, J., Raasch, J. A. & Vierstra, R. D. Ubiquitin Extension Proteins of *Arabidopsis thaliana*. *J. Biol. Chem.* **265**, 12486–12493 (1990).

85. Hernandez-Garcia, C. M., Martinelli, A. P., Bouchard, R. A. & Finer, J. J. A soybean (*Glycine max*) polyubiquitin promoter gives strong constitutive expression in transgenic soybean. *Plant Cell Rep.* **28**, 837–849 (2009).

86. Xiao, K., Zhang, C., Harrison, M. & Wang, Z. Y. Isolation and characterization of a novel plant promoter that directs strong constitutive expression of transgenes in plants. *Mol. Breed.* **15**, 221–231 (2005).

87. Potenza, C., Aleman, L. & Sengupta-Gopalan, C. Invited review: Targeting transgene expression in research, agricultural, and environmental applications: Promoters used in plant transformation. *Vitr. Cell. Dev. Biol. Plant* **40**, 1–22 (2004).

88. Odell, J. T., Nagy, F. & Chua, N.-H. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* **313**, 810–812 (1985).

89. Kay, R., Chan, A., Daly, M. & Mcpherson, J. Duplication of CaMV 35S Promoter Sequences Creates a Strong Enhancer for Plant Genes. *Science.* **236**, 1299–1302 (1987).

90. Boshart, M. *et al.* A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* **41**, 521–530 (1985).

91. Thomsen, D. R., Stenberg, R. M., Goins, W. F. & Stinski, M. F. Promoter-regulatory region of the major immediate early gene of human cytomegalovirus. *Proc. Natl. Acad. Sci.* **81**, 659–663 (1984).

92. Takahashi, K. *et al.* Requirement of stereospecific alignments for initiation from the simian virus 40 early promoter. *Nature* **319**, 121–126 (1986).

93. Gruss, P., Dhar, R. & Khoury, G. Simian virus 40 tandem repeated sequences as an element of the early promoter. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 943–947 (1981).

94. Walker, T. L., Purton, S., Becker, D. K. & Collet, C. Microalgae as bioreactors. *Plant Cell Rep.* **24**, 629–641 (2005).

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

96. An, G., Costa, M. A. & Ha, S. B. *Nopaline synthase* promoter is wound inducible and auxin inducible. *Plant Cell* **2**, 225–233 (1990).

97. Schena, M., Lloyd, A. M. & Davis, R. W. A steroid-inducible gene expression system for plant cells. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 10421–10425 (1991).

98. Hinchey, B. S., Bensen, R. J. & Nelson, D. E. Water-Deficit inducible plant promoters. **1**, 12–17 (2004).

99. Hinchey, B. S. Stress-inducible plant promoters. **1**, (2005).

100. Luc Adam, Hayward, C., Lynne Reuber, San Mateo, C. & Karen S. Century, Albany, C. Disease-inducible promoters. **2**, (2011).

101. Hidenori Ozawa. Nitrate inducible promoter. **2**, (2007).

102. Karimi;, M., Barthels, N. & Godelieve Gheysen. Pathogen-inducible plant promoters. **1**, (2001).

103. Medrano, L., Feldmann, K. A., Alexandrov, N., Fang, Y. & Nester Apuya. Drought inducible promoters and uses thereof. **2**, (2008).

104. Mine;, T., Ohyama;, A., Hiyoshi;, T. & Keisuke Kasaoka. Cold-inducible promoter sequences. **12**, 2–4 (1943).

105. Iwai, M., Hori, K., Sasaki-Sekimoto, Y., Shimojima, M. & Ohta, H. Manipulation of oil synthesis in *Nannochloropsis* strain NIES-2145 with a phosphorus starvation-inducible promoter from Chlamydomonas reinhardtii. *Front. Microbiol.* **6**, 1–15 (2015).

106. Jackson, H. O. *et al.* An inducible expression system in the alga *Nannochloropsis gaditana* controlled by the nitrate reductase promoter. *J. Appl. Phycol.* **31**, 269–279 (2018).

107. Liu, W. & Stewart, C. N. Plant synthetic promoters and transcription factors. *Curr. Opin. Biotechnol.* **37**, 36–44 (2016).

108. Scranton, M. A. *et al.* Synthetic promoters capable of driving robust nuclear gene expression in the green alga *Chlamydomonas reinhardtii*. *Algal Res.* **15**, 135–142 (2016).

109. Urtubia, H. O., Betanzo, L. B. & Vásquez, M. Microalgae and Cyanobacteria as Green Molecular Factories: Tools and Perspectives. in *Algae - Organisms for Imminent Biotechnology* (eds. Dhanasekaran, D. & Thajuddin, N.) 1–28 (IntechOpen, 2016).

110. Stevens, D. R., Rochaix, J. D. & Purton, S. The bacterial phleomycin resistance gene ble as a dominant selectable marker in *Chlamydomonas*. *Mol. Gen. Genet.* **251**, 23–30 (1996).

111. Schroda, M., Blöcker, D. & Beck, C. F. The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. *Plant J.* **21**, 121–131 (2000).

112. Iwai, M., Ikeda, K., Shimojima, M. & Ohta, H. Enhancement of extraplastidic oil synthesis in *Chlamydomonas reinhardtii* using a type-2 diacylglycerol acyltransferase with a phosphorus starvation-inducible promoter. *Plant Biotechnol. J.* **12**, 808–819 (2014).

113. Oliver Kilian, A. & Bertrand Vick, E. Bidirectional promoters in *Nannochloropsis*. **1**, (2013).

114. Poliner, E. *et al.* A toolkit for *Nannochloropsis oceanica* CCMP1779 enables gene stacking and genetic engineering of the eicosapentaenoic acid pathway for enhanced long-chain polyunsaturated fatty acid production. *Plant Biotechnol. J.* **16**, 298–309 (2017).

115. Ajjawi, I. *et al.* Lipid production in *Nannochloropsis gaditana* is doubled by decreasing expression of a single transcriptional regulator. *Nat. Biotechnol.* **35**, 647–652 (2017).

116. Wilde, C. De *et al.* Plants as bioreactors for protein production : avoiding the problem of transgene silencing. *Plant Mol. Biol.* **43**, 347–359 (2000).

117. Mark Hildebrand , Aubrey Davis , Raffaela Abbriano , Haley R. Pugsley , Jesse C. Traller , Sarah R. Smith , Roshan P. Shrestha , Orna Cook , Eva L. Sánchez-Alvarez , Kalpana Manandhar- Shrestha, and B. A. Applications of Imaging Flow Cytometry for Microalgae. in *Imaging Flow Cytometry: Methods and Protocols* **1389**, 47–67 (Springer Science+Business Media New York, 2016).

118. Watanabe, Y., Kadono, T., Kira, N., Suzuki, K. & Iwata, O. Marine Genomics Development of endogenous promoters that drive high-level expression of introduced genes in the model diatom *Phaeodactylum tricornutum*. *Mar. Genomics* **42**, 41–48 (2018).

119. Lin, H.-Y. *et al.* Alkaline phosphatase promoter as an efficient driving element

for exogenic recombinant in the marine diatom *Phaeodactylum tricornutum*. *Algal Res.* **23**, 58–65 (2017).

120. Erdene-Ochir, E. *et al.* Cloning of a novel endogenous promoter for foreign gene expression in *Phaeodactylum tricornutum*. *Appl. Biol. Chem.* **59**, 1–7 (2016).

121. Adler-Agnon (Shemesh), Z., Leu, S., Zarka, A., Boussiba, S. & Khozin-Goldberg, I. Novel promoters for constitutive and inducible expression of transgenes in the diatom *Phaeodactylum tricornutum* under varied nitrate availability. *J. Appl. Phycol.* **30**, 2763–2772 (2018).

122. Seo, S., Jeon, H., Hwang, S., Jin, E. & Chang, K. S. Development of a new constitutive expression system for the transformation of the diatom *Phaeodactylum tricornutum*. *Algal Res.* **11**, 50–54 (2015).

123. An, G. Development of Plant Promoter Expression Vectors and Their Use for Analysis of Differential Activity of Nopaline Synthase Promoter in Transformed Tobacco Cells. *Plant Physiol.* **81**, 86–91 (1986).

124. Tsien, R. Y. the Green Fluorescent Protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).

125. Nagai, T. *et al.* A variant of YFP with fast and efficient maturation for cell-biological applications. *Nature* **20**, 1585–1588 (2002).

126. Jefferson RA. The GUS reporter gene system. *Nature* **342**, 837–838 (1989).

127. Dupuy, D. *et al.* Genome-scale analysis of in vivo spatiotemporal promoter activity in *Caenorhabditis elegans*. *Nat. Biotechnol.* **25**, 663–668 (2007).

128. Mattioli, K. *et al.* High-throughput functional analysis of lncRNA core promoters elucidates rules governing tissue specificity. *Genome Res.* **29**, 344–355 (2019).

129. Barnard, E., McFerran, N. V., Trudgett, A., Nelson, J. & Timson, D. J. Development and implementation of split-GFP-based bimolecular fluorescence complementation (BiFC) assays in yeast. *Biochem. Soc. Trans.* **36**, 479–482 (2008).

130. Charvin, G., Cross, F. R. & Siggia, E. D. A microfluidic device for temporally controlled gene expression and long-term fluorescent imaging in unperturbed dividing yeast cells. *PLoS One* **3**, (2008).

131. Cormack, B. P. *et al.* Yeast-enhanced green fluorescent protein (yEGFP): A reporter of gene expression in *Candida albicans*. *Microbiology* **143**, 303–311 (1997).

132. de Freitas, F. A., Yunes, J. A., da Silva, M. J., Arruda, P. & Leite, A. Structural characterization and promoter activity analysis of the γ-kafirin gene from sorghum. *MGG Mol. Gen. Genet.* **245**, 177–186 (1994).

133. Xie, Y., Liu, Y., Meng, M., Chen, L. & Zhu, Z. Isolation and identification of a super strong plant promoter from cotton leaf curl Multan virus. *Plant Mol. Biol.* **53**, 1–14 (2003).

134. Zhang, W., McElroy, D. & Wu, R. Analysis of rice Act1 5′ region activity in transgenic rice plants. *Plant Cell* **3**, 1155–1165 (1991).

135. Kishi-Kaboshi, M., Aida, R. & Sasaki, K. Parsley ubiquitin promoter displays higher activity than the CaMV 35S promoter and the chrysanthemum actin 2 promoter for productive, constitutive, and durable expression of a transgene in *Chrysanthemum morifolium*. *Breed. Sci.* **69**, 536–544 (2019).

136. Jefferson, R. A., Klass, M., Wolf, N. & Hirsh, D. Expression of chimeric genes in *Caenorhabditis elegans*. *J. Mol. Biol.* **193**, 41–46 (1987).

137. Greer, L.F.; Szalay, A. . Imaging of light emission fron the expression of liciferases in living cells and organisms: a review. *Luminescence* 43–74 (2002).

138. Gould, S. J. & Subramani, S. Firefly luciferase as a tool in molecular and cell biology. *Anal Biochem* **175**, 5–13 (1988).

139. Hastings, J. W. Chemistries and colors of bioluminescent reactions: A review. *Gene* **173**, 5–11 (1996).

140. Ow, D. W. *et al.* Transient and Stable Expression of the Firefly Luciferase Gene in Plant Cells and Transgenic Plants. *Science.* **234**, 856–859 (1986).

141. Nakajima, Y. & Ohmiya, Y. Bioluminescence assays: multicolor luciferase assay, secreted luciferase assay and imaging luciferase assay. *Expert Opin. Drug Discov.* **5**, 835–849 (2010).

142. Hellens, R. P. *et al.* Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods* **1**, 1–14 (2005).

143. Bai, H. *et al.* Major capsid protein of *Autographa californica* multiple nucleopolyhedrovirus contributes to the promoter activity of the very late viral genes. *Virus Res.* **273**, 197758 (2019).

144. Fuhrmann, M., Hausherr, A., Ferbitz, L., Schödl, T. & Hegemann, P. Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. *Plant Mol. Biol.* **55**, 869–881 (2004).

145. Shao, N. & Bock, R. A codon-optimized luciferase from *Gaussia princeps* facilitates the in vivo monitoring of gene expression in the model alga *Chlamydomonas reinhardtii*. *Curr. Genet.* **53**, 381–388 (2008).

146. Lauersen, K. J., Kruse, O. & Mussgnug, J. H. Targeted expression of nuclear transgenes in *Chlamydomonas reinhardtii* with a versatile, modular vector toolkit. *Appl. Microbiol. Biotechnol.* **99**, (2015).

147. Ruecker, O., Zillner, K., Groebner-Ferreira, R. & Heitzer, M. Gaussia-luciferase as a sensitive reporter gene for monitoring promoter activity in the nucleus of the green alga *Chlamydomonas reinhardtii*. *Mol. Genet. Genomics* **280**, 153–162 (2008).

148. Davis, S. J. & Vierstra, R. D. Soluble, highly fluorescent variants of green

fluorescent protein (GFP) for use in higher plants. *Plant Mol. Biol.* **36**, 521–528 (1998).

149. Quaedvlieg, N. E. M. *et al.* Erratum: Fusions between green fluorescent protein and β-glucuronidase as sensitive and vital bifunctional reporters in plants. *Plant Mol. Biol.* **38**, 917 (1998).

150. Weaver, J. L. *Introduction to flow cytometry. Methods (San Diego, Calif.)* **21**, (2000).

151. Birnbaum, K. *et al.* Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nat. Methods* **2**, 615–619 (2005).

152. Gershenzon, J. & Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **3**, 408–414 (2007).

153. Buckingham, J. *Dictionary of natural products.* (Chapman & Hall, 1994).

154. Kainer, D. *et al.* Plant-Derived Terpenes: A Feedstock for Specialty Biofuels. *Trends Biotechnol.* **35**, 227–240 (2016).

155. Breitmaier, E. *Terpenes: Flavors, Fragrances, Pharmaca, Pheromones*. *Wiley-VCH* (2006).

156. Davies, F. K., Jinkerson, R. E. & Posewitz, M. C. Toward a photosynthetic microbial platform for terpenoid engineering. *Photosynth. Res.* **123**, 265–284 (2015).

157. Pichersky, E. & Gershenzon, J. The formation and function of plant volatiles: Perfumes for pollinator attraction and defense. *Curr. Opin. Plant Biol.* **5**, 237–243 (2002).

158. Kessler, A. & Baldwin, I. . Defensive Function of Herbivore-Induced Plant Volatile Emissions in Nature. *Science.* **291**, 2141–2144 (2001).

159. Baldwin, I. T., Kessler, A. & Halitschke, R. Volatile signaling in plant–plant–herbivore interactions: what is real? *Curr. Opin. Plant Biol.* **5**, 681–685 (2002).

160. Lichtenthaler, H. K., Rohmer, M. & Schwender, J. Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants. *Physiol. Plant.* **101**, 643–652 (1997).

161. HK., L. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 47–65 (1999).

162. Lohr, M., Schwender, J. & Polle, J. E. W. Isoprenoid biosynthesis in eukaryotic phototrophs: A spotlight on algae. *Plant Sci.* **185–186**, 9–22 (2012).

163. Ajikumar, P. K. *et al.* Terpenoids: Opportunities for Biosynthesis of Natural Product Drugs Using Engineered Microorganisms. *Mol. Pharm.* **5**, 167–190 (2008).

164. Schrader, J. *Biotechnology of Isoprenoids*. **148**, (2015).

165. Croteau, R., Ketchum, R. E. B., Long, R. M., Kaspera, R. & Wildung, M. R. Taxol

biosynthesis and molecular genetics. *Phytochem. Rev.* **5**, 75–97 (2006).

166. Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–532 (2013).

167. Parayil Kumaran Ajikumar, Xiao, W.-H., Tyo, K. E. J., Wang, Y. & Stephanopoulos, F. G. Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in *Escherichia coli*. *Science.* **330**, 70–74 (2010).

168. Nielsen, A. Z. *et al.* Extending the biosynthetic repertoires of cyanobacteria and chloroplasts. *Plant J.* **87**, 87–102 (2016).

169. Mellor, S. B., Vavitsas, K., Nielsen, A. Z. & Jensen, P. E. Photosynthetic fuel for heterologous enzymes: the role of electron carrier proteins. *Photosynth. Res.* **134**, 329–342 (2017).

170. Vickers, C. E., Gershenzon, J., Lerdau, M. T. & Loreto, F. A unified mechanism of action for volatile isoprenoids in plant abiotic stress. *Nat. Chem. Biol.* **5**, 283–291 (2009).

171. Lauersen, K. J. Eukaryotic microalgae as hosts for light-driven heterologous isoprenoid production. *Planta* **249**, 155–180 (2019).

172. Chaves, J. E. & Melis, A. Biotechnology of cyanobacterial isoprene production. *Appl. Microbiol. Biotechnol.* **102**, 6451–6458 (2018).

173. Gao, X. *et al.* Engineering the methylerythritol phosphate pathway in cyanobacteria for photosynthetic isoprene production from CO2. *Energy Environ. Sci.* **9**, 1400–1411 (2016).

174. Davies, F. K., Work, V. H., Beliaev, A. S. & Posewitz, M. C. Engineering limonene and bisabolene production in wild type and a glycogen-deficient mutant of *Synechococcus* sp. PCC 7002. *Front. Bioeng. Biotechnol.* **2**, 1–11 (2014).

175. Sebesta, J. & Peebles, C. A. Improving heterologous protein expression in *Synechocystis* sp. PCC 6803 for alpha-bisabolene production. *Metab. Eng. Commun.* **10**, e00117 (2020).

176. Dienst, D., Wichmann, J., Mantovani, O., Rodrigues, J. & Lindberg, P. High density cultivation for efficient sesquiterpenoid biosynthesis in *Synechocystis* sp. PCC 6803. *Sci. Rep.* **10**, 5932 (2020).

177. Sasso, S., Pohnert, G., Lohr, M., Mittag, M. & Hertweck, C. Microalgae in the postgenomic era: A blooming reservoir for new natural products. *FEMS Microbiol. Rev.* **36**, 761–785 (2012).

178. Wichmann, J., Baier, T., Wentnagel, E., Lauersen, K. J. & Kruse, O. Tailored carbon partitioning for phototrophic production of (E)-α-bisabolene from the green microalga *Chlamydomonas reinhardtii*. *Metab. Eng.* **45**, 211–222 (2018).

179. Lauersen, K. J. *et al.* Phototrophic production of heterologous diterpenoids and a hydroxy-functionalized derivative from *Chlamydomonas reinhardtii*. *Metab. Eng.* **49**, 116–127 (2018).

180. Lauersen, K. J. *et al.* Efficient phototrophic production of a high-value sesquiterpenoid from the eukaryotic microalga *Chlamydomonas reinhardtii*. *Metab. Eng.* **38**, 331–343 (2016).

181. D'Adamo, S. *et al.* Engineering the unicellular alga *Phaeodactylum tricornutum* for high-value plant triterpenoid production. *Plant Biotechnol. J.* **17**, 75–87 (2019).

182. Papaefthimiou, D. *et al.* Heterologous production of labdane-type diterpenes in the green alga *Chlamydomonas reinhardtii*. *Phytochemistry* **167**, 112082 (2019).

183. Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E. & Villaverde, A. Microbial factories for recombinant pharmaceuticals. *Microb. Cell Fact.* **8**, 17 (2009).

184. Lu, Y. *et al.* Regulation of the cholesterol biosynthetic pathway and its integration with fatty acid biosynthesis in the oleaginous microalga *Nannochloropsis oceanica*. *Biotechnol. Biofuels* **7**, 1–15 (2014).

CHAPTER 2

Novel endogenous promoters for genetic engineering of the marine microalga *Nannochloropsis gaditana* CCMP526

Margaret Ramarajan[a], Michele Fabris[a,b], Raffaela M. Abbriano[a] , Mathieu Pernice[a], Peter J. Ralph[a]

[a] University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo NSW 2007, Australia

[b] CSIRO Synthetic Biology Future Science Platform, GPO box 2583, Brisbane QLD 4001, Australia

Author contributions: M.R. performed the experiments; M.F. and P.J.R. conceived the research project; M.F, R.A. and M. P. contributed in data analysis; M.R. wrote the manuscript; all authors edited and reviewed the manuscript.

Published in December 2019 within the journal of 'Algal Research- Biomass, Biofuels and Bioproducts' as a research article (See appendix 4)

Novel endogenous promoters for genetic engineering of the marine microalga *Nannochloropsis gaditana* CCMP526

Margaret Ramarajan[a], Michele Fabris[a,b], Raffaela M. Abbriano[a] , Peter J. Ralph[a]

[a] University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo NSW 2007, Australia

[b] CSIRO Synthetic Biology Future Science Platform, GPO box 2583, Brisbane QLD 4001, Australia

Abstract

Nannochloropsis spp. are marine microalgae from the *Eustigmatophyceae* stramenopile lineage that have been studied extensively due to a broad range of industrial applications, mostly related to their oil and pigment production. However, tools to genetically engineer members of this group, and therefore further understand and maximise their industrial potential are still limited. In order to expand the potential industrial uses of this organism, several molecular tools, including gene promoters of different strength, are needed. A comprehensive and diverse set of well-characterised promoters is key to a number of genetic engineering and synthetic biology applications, such as the assembly of complex biological functions or entire metabolic pathways.

In this study, we measured the promoter activity of three endogenous constitutive promoters from *N. gaditana* genes *EPPSII (Nga02101); HSP90 (Nga00934); ATPase (Nga06354.1)* in driving the expression of a *Sh ble- mVenus* fluorescent reporter fusion protein. Through a combined approach that includes flow cytometry, RT-qPCR and immunoblotting, we profiled the activity of these promoters at both the transcript and protein level. Two promoters *HSP90 (Nga00934)* and *EPPSII (Nga02101)* outperformed the widely used *β-tubulin* promoter, exhibiting 4.5 and 3.1-fold higher mVenus fluorescence, respectively. A third promoter *ATPase (Nga06354.1)* was also able to drive the expression of transgenes, albeit at lower levels. We show that the new promoters identified in this study are valuable tools, which can be used for genetic engineering and functional genetics studies in *N. gaditana.*

CHAPTER 3

Novel endogenous inducible promoter and viral promoters for genetic engineering the microalga *Nannochloropsis gaditana*

Margaret Ramarajan[a], Michele Fabris[a,b], Raffaela M. Abbriano[a] , Peter J. Ralph[a]

[a] University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo NSW 2007, Australia

[b] CSIRO Synthetic Biology Future Science Platform, GPO box 2583, Brisbane QLD 4001, Australia

Author contribution: M.R. performed the experiments; M.F. and P.J.R. conceived the research project; M.F. and R.A. contributed in data analysis; M.R. wrote the manuscript; all authors edited and reviewed the manuscript.

Abstract

Availability of a wide range of promoters that differ in their capacity to regulate the temporal and spatial expression of the transgene can radically increase the effective application of transgenic technology. Heterologous constitutive promoters with elevated activity, are available for higher plants such as CamV35S promoter and animals such as CMV and SV40 promoter. Strong, viral promoters with orthogonal expression capability could be useful in engineering a variety of microalgae. However, such promoters have not being identified yet for microalgae of the group *Nannochloropsis*.

Nannochloropsis spp. are oleaginous microalgae with capacity to cope with phosphorous limitation because of the presence of phosphorus (P) reserves and efficient P recycling which allow photosynthesis to continue, leading to cell growth. Phosphate limitation is a stress condition known to trigger increased accumulation of lipids and pigments in this photosynthetic organism. Finding a highly inducible promoter that is capable of actively expressing exogenous genes in *Nannochloropsis* under phosphorous stress will further improve the utility of this organism as a low-cost industrial platform for multiple biotechnology applications.

In this study, a putative promoter candidate (*EsV-1-89*; NCBI gene ID: 920795) from the *Ectocarpus siliculosus* virus 1 was identified which was able to drive transgene expression in *N. gaditana,* but the promoter activity was around 26% of the *Hsp90* endogenous control promoter and furthermore no reporter gene expression was observed in the diatom *P. tricornutum.* This suggests that the viral promoter has moderate activity and cannot be used as a universal promoter across the stramenopile taxa.

In this study, we identified two promising phosphate-inducible gene candidates (*SPS*, EWM20309.1; *GDPD*, EWM28518.1) and used their putative promoter regions to drive the expression of a reporter gene (*mVenus*) under phosphate-replete and deplete conditions. Even though phosphate limitation induced promoter activity was evident in the initial screen, it could not be confirmed due to the instability of the generated transgenic lines.

3.1 Introduction

Molecular genetics approaches have been used to identify and characterise cis-acting DNA sequences essential for eukaryotic gene regulation[1]. These sequences are modular recognition elements that interact with specific transcription factors for precise and efficient initiation of transcription and for controlling expression of a gene[2,3]. A typical promoter region comprises of a core and regulatory regions[4,5]. Some known promoter elements include TATA box, GC-box, CCAAT-box or a CpG, INR[6–16]. The proximal and distal regions of the promoter contain diverse regulatory sequences such as enhancers, repressors, insulators (cis-elements) that contribute to the regulation of gene expression[17,18]. Hence transcription regulation depends on the presence and activity of transcription factors and the number, type and position of regulatory elements[18,19]. Gene regulation may occur at different stages of gene expression and can be extremely important during transcription.[20] The promoters that power transgene expression warrant this control[20]. These promoters can be endogenous or from a variety of origins such as viral[21] or synthetic[22] and can be constitutive[23] or inducible[24] in nature (Figure 4).

Over time, several promoters have been isolated from a wide range of organisms and applied to genetic engineering systems such as bacteria, yeasts, algae, higher plants and mammalian[25–32]. Whether the goal is basic research or strain/crop improvement or biopharming (production and use of genetically engineered plants and animals to produce pharmaceutical substances[33]), selection of promoters based on host system and transgene type is critical for successful gene expression[20]. Examples of promoters that can drive constitutive gene expression in higher plants include the *Arabidopsis thaliana ubiquitin* (NCBI gene ID: 824425) promoter[34], the soybean *polyubiquitin* (U310508) promoter[23] and the strong MtHP promoter from the legume *Medicago truncatula*[25]. Microbial promoter include, the *alcohol dehydrogenase isozyme I* (*ADH1*) promoter in yeast[35] and *elongation factor 2* (*EF2*) promoter in *Phaeodactylum tricornutum*[36]. In *Nannochloropsis* sp. few constitutive promoters (Table 2) have been identified[37] including the ones characterised in Chapter 2[38]. Heterologous constitutive promoters on the other hand are mostly of viral origin[17,20,39–41]. For example, the cauliflower mosaic virus (CaMV) 35S promoter is a plant nuclear promoter system which

is constitutively active in several plant species[21,29] and green algae such as *Chlamydomonas reinhardtii*[42,43]. For expression in mammalian cells, viral promoters derived from cytomegalovirus (CMV)[44,45] or simian virus 40 (SV40)[46,47] are routinely used.

In case of CaMV, the genes in its double-strand DNA genome are transcribed by host nuclear RNA polymerase II[21] and the 35S promoter does not depend on any trans-acting viral gene products[21,48] whereas, the CMV promoter is derived from an immediate early gene (*IE1/Ul123*, NCBI gene ID: 3077513)[45,49] responsible for the most abundant viral RNAs production in infected host cells at immediate-early (IE) times after infection and is widely used because of its extremely high transcription levels in a variety of primary and stable cell lines and in cell-free transcription systems[41]. Such a strong expression was not observed with viral promoters such as CaMV promoter in microalgal hosts such as *C. reinhardtii*, *Nannochloropsis* strain UMT-M3 and *N. salina*[43,50–52] which may be due the dissimilarity in available transcription factors governing RNAs in higher plants/animals and microalgal systems[53]. Recently, an expression system (Algevir[54]) for microalgae based on viral vectors was introduced, where expression of transgene using is an inducible geminiviral vector was explored in the marine microalga *Schizochytrium* sp.[54].

Phycodnaviridae are an ancient family of large icosahedral viruses that infect eukaryotic algae[55,56]. Phycodnaviridae are evolutionarily connected to large DNA viruses, referred to as the nucleocytoplasmic large DNA viruses (NCLDV)[55]. Phycodnaviridae are grouped into six genera: *Chlorovirus, Coccolithovirus, Prasinovirus, Prymnesiovirus, Phaeovirus* and *Raphidovirus*[55]. As there are no reports on viruses infecting *Nannochloropsis* spp., we therefore considered the *Phaeovirus- Ectocarpus siliculosus* virus (EsV-1)[57] (Figure 9), which is a lysogenic virus that infects the marine stramenopile brown alga, *Ectocarpus siliculosus* that is phylogenetically closely related to *N. gaditana*[58,59] (Figure 9). Essentially, EsV-1 is completely dependent on the host transcriptional machinery (similar to CamV)[55] and does not code for any viral-specific tRNAs[57] which highlights the fact that such a virus should poses promoters recognisable by the transcriptional machinery of the microalgal host.

Here, we set out to identify and explore new viral promoter candidates that could serve as microalgal counterparts for efficient and high gene expression in *N. gaditana* and possibly as universal promoter for stramenopiles. It could also help us compare the promoter strength of viral promoters with endogenous ones.

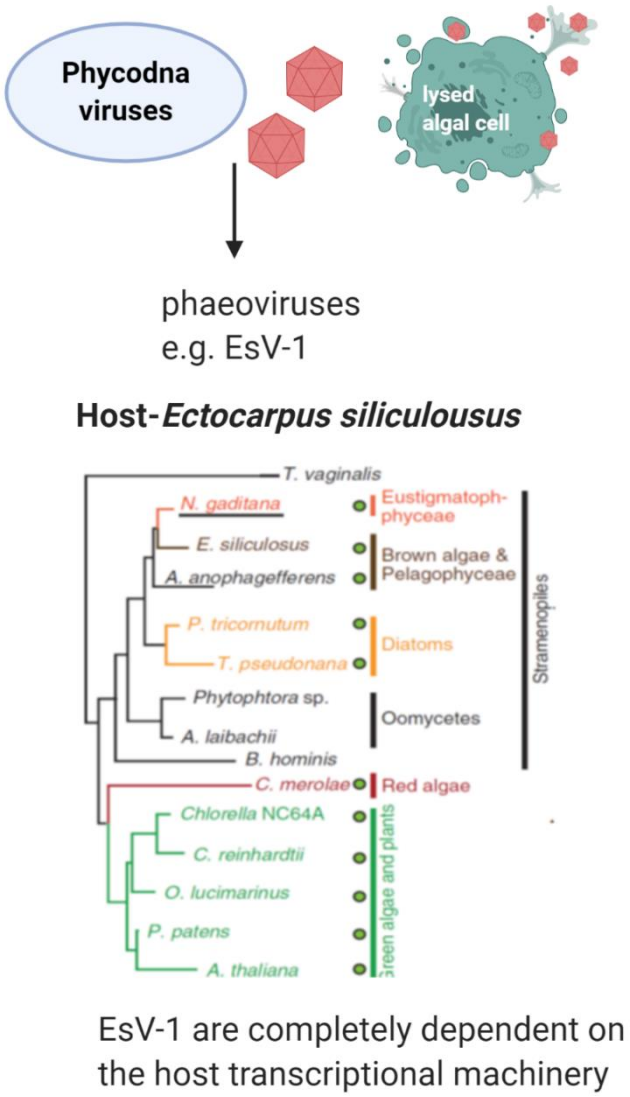


Figure 9. Selection of *Ectocarpus siliculosis* virus 1 (EsV-1) gene promoter

The EsV-1 belongs to the phaeoviruses. The data on EsV-1 was obtained from virusite.org and the phylogenetic tree was adapted from Radakovits et al., 2012[58].

On the other hand, constitutive promoters may not be suitable for expressing highly deleterious or lethal/toxic genes such as *TphoA*, a chimeric protein consisting of the transmembrane domain from Vibrio cholera *ToxR* (GenBank® accession no. M21249) joined to bacterial alkaline phosphatase *PhoA* (GenBank® accession no. U00096.2) in a

E. coli host[60] and genes that could affect or alter metabolic or developmental pathways that would inhibit the growth and multiplication of the host[61]. For example, a tetracycline-regulated promoter system was used to express proteins that are either cytotoxic or interfere with replication of the adenovirus host[62]. In such cases, an inducible gene expression system would prove to be extremely useful. The promoters of inducible genes are not active always, but respond to an external trigger which can be environmental, stress-related or chemical and are highly regulated for rapid and specific activation in response to a stimuli[27]. In plants, for example the nopaline synthase (*Nos*) promoter is both auxin and wound-inducible[24] (Table 4), while a steroid-inducible expression system uses the glucocorticoid hormone[63] and several patented promoter sequences such as water-deficit inducible promoter[64], stress-inducible promoters[65], disease-inducible promoters[66], nitrate-inducible promoter[67], pathogen-inducible promoters[68], drought-inducible promoters[69] and cold-inducible promoters[70] (Table 4).

Table 4. Types of inducible promoters

The different types of plant inducible promoter types with example

Promoter type	**Example**
Hormone inducible	Nopaline synthase promoter[24] (Auxin inducible)
Disease inducible	Disease inducible promoters[66], pathogen-inducible promoter[68]
Steroid inducible	GRE-CaMV promoter[63]
Stress inducible	RAB-17, CA4H promoters, drought-inducible promoter[64,65,69]
Nutrient inducible	Nitrate-inducible promoter[67]
Temperature inducible	Cold-inducible promoter[70]

The criteria used to select our induction system included i) being inexpensive and ii) uses an inducer that could be fully controlled leading to dose-dependent expression of the gene of interest with minimal basal activity[27,61]. A variety of inducible expression systems have been developed in microalgae i) temperature-inducible

promoters such as *heat shock protein 70A (HSP70A)* gene promoter from the green alga *C. reinhardtii*[71–73]; ii) light-inducible promoters such as *fucoxanthin chlorophyll a/c binding protein (FcpA)* gene promoter from diatoms[74–76] *and light-inducible protein (*LIP*)* gene promoter from *Dunaliella* sp.[77]; iii) gene expression systems regulated by nutrients or chemicals such as nitrate, for example, the *nitrate reductase* (*NR, Naga_100699g1*) gene promoter from *N. gaditana*, is active in the presence of nitrate and inactive in the presence of ammonia[78]; similarly, the *nitrate reductase* (*NR, AY579336.1*) gene promoter in *P. tricornutum*[79–81] and the *lipid droplet surface protein (LDSP)* promoter in *N.oceanica,* which has robust expression under nitrogen starvation conditions[82]. Silica is another nutrient whose presence in the media represses the expression of *silicon transporters* (*SITs*), which only become active under silicon limitation. SIT promoter has been successfully used as an inducible promoter in the diatoms *Thalassiosira pseudonana* and *Cyclotella cryptica*[83]. *Carbonic anhydrase (CA)* gene promoter is another example of an inducible promoter that can be turned on at low (less than 5%) CO_2 concentration in *P. tricornutum*[84] and induced under high salinity conditions in *Dunaliella salina* [85]. Finally, phosphate-inducible promoters which drive transgene expression under phosphate limitation conditions have been shown to work in multiple microalgal species, for example, *the sulphoquinovosyldiacylglycerol 2 (SQD2)* gene promoter from *C. reinhardtii*[86,87]*, alkaline phosphatase* gene (*APase*) promoter from *P. tricornutum*[88]*.*

N. gaditana has been found to increase intracellular accumulation of specific lipids, or carotenoids under nutrient deficiency, in particular nitrogen and phosphorous. While nitrogen limitation leads to i) general down regulation of protein synthesis excluding glycolysis and fatty acid synthesis; ii) decreased photosynthetic efficiency; iii) lipid accumulation and iv) autophagy of plastids[89,90]. Phosphate limitation may not be so harsh on the cells as *Nannochloropsis* cells are shown to preferentially store phosphorous as polyphosphate granules while P is abundant[91] and to cope with P limitation, an interplay of P recycling and lipid composition changes are observed, which allows photosynthesis to continue leading to cell growth[87,92]. This was the motivation for selecting stress inducible (phosphorous) promoter targets.

In this study, we set out to identify and characterise i) putative viral promoters from EsV-1 and ii) putative endogenous gene promoters which might be inducible under phosphate limitation. In case of viral promoter targets, one putative promoter region of ORF89 (*EsV-1-89;* NCBI gene ID: *920795*) was synthesised and transformed into the *N. gaditana* genome along with *mVenus* reporter. The promoter activity was measured using flow cytometry and RT-qPCR. The viral promoter was further tested in *N. oceanica* and *P. tricornutum* to assess for promoter activity in other stramenopile model species.

For our inducible promoter target, the putative phosphate-inducible promoter candidates were identified by mining the *N. oceanica* CCMP1779[92] transcriptomic data for genes that were upregulated under phosphate starvation condition and searching for their othologs in *N. gaditana* followed by extraction and cloning of their upstream putative promoter regions. The promoter activity was monitored *in vivo* using mVenus fluorescence by flow cytometry.

3.2 Materials and methods

3.2.1 Strains and culture conditions

N. gaditana CCMP526, *N. oceanica* CCMP1779 and *P. tricornutum* CCAP 1055/1 strains were purchased from the Provasoli-Guillard National Centre for Culture of Marine Phytoplankton, USA. The cells were grown in liquid modified F/2 medium[93], where the seawater component was replaced with Enriched Seawater, Artificial Water (ESAW)[94], and the concentration of the vitamin was doubled (primary vitamin stock solution: Thiamine · HCl- 400 mg, Biotin- 2 g/L). Salinity of the media was maintained at 31 PSU.

Axenic cultures of *N. gaditana* were maintained under cool white LED light panel set to a light intensity of 100 µmol photons m^{-2} s^{-1}, under a continuous light regime, at 21 °C in a shaker incubator (Kühner, Switzerland). *P. tricornutum* was maintained under the same set of conditions except the light intensity was 200 µmol photons m^{-2} s^{-1}. Batch cultures were maintained in 250 mL Erlenmeyer flasks closed with foam plugs to allow gas exchange. Cultures were checked for bacterial contamination every two months by streaking on marine broth agar plates.

All promoter analysis experiments were carried out in the conditions described above, except for incubation of the electroporated culture, which was carried out at a light intensity of 80 µmol photons m^{-2} s^{-1}. For the phosphate regulation experiment, liquid cultures were grown to mid-log phase, pelleted, and washed in phosphate-deplete medium (x3), and re-suspended in phosphate-deplete growth medium.

3.2.2 Identification of viral promoter from *Ectocarpus siliculosus* virus

Nannochloropsis gaditana is phylogenetically closely related to *Ectocarpus siliculosus* (a filamentous brown alga)[58,59] and *Ectocarpus siliculosus* virus 1[57] is a DNA virus that infects *Ectocarpus* sp. which has a fully sequenced genome. The genome sequence was accessed using the genome browser http://www.virusite.org/[95] and the putative promoter region of *ORF89* (*EsV-1-89*; NCBI gene ID: 920795) was extracted and analysed with PlantCARE[96] and Softberry TSSP[97] software. Also, the putative promoter sequence was manually screened for conserved promoter motifs found in nucleo-cytoplasmic large DNA viruses (NCLDV)[98] such as a highly conserved 10 nucleotide sequence AAAAATAnTT observed in Chlorellaviruses[99], three AT rich conserved regions (ArnTTAAnA, AATGACA and GTnGATAyr)[100] and sequences (ATGACAA and TATAAAT) similar to eukaryotic "TATA box" observed in chloroviruses[101]. A region of 833 bp was selected and synthesised (Integrated DNA Technologies, USA).

3.2.3 Identification of phosphate inducible genes in *N. gaditana*

A published transcriptomic data set of *N oceanica* CCMP1779[92] was used to identify genes that showed high inducibility under phosphate-limiting condition; however, as such a dataset is not currently available for *N. gaditana*. Three candidate genes with negligible basal expression in phosphate replete state and 2 to 4-fold higher gene expression in phosphate deplete condition were selected (Table 5). The orthologs of these genes were obtained using the web-based *N. gaditana* BLASTp search tool using default parameters, in the CRIBI Genomics *Nannochloropsis* genome portal[102]. The putative upstream promoter regions (600 bp–1000 bp) of these genes were analysed using PlantCARE[96] and Softberry TSSP[97] software.

3.2.4 Construction of transformation vectors

The plasmid pNaga4.mVenus (Chapter 2-Materials and methods section 2.4)[38] was used as the vector backbone which was amplified using Q5® High Fidelity DNA

polymerase (New England Biolabs, USA) along with the terminator region of a *N. gaditana* gene putatively annotated as *brix domain containing 1* (*Nga00944*) and a *N. gaditana* putative *hsp90* gene (*Nga00934)* promoter region. The primers used for amplification are specified in Table S10. These amplified fragments were ligated using the Gibson Assembly® Master Mix (New England Biolabs, USA) to form the pNaga4.mV.DC.N2 vector (Figure S9).

The transformation vector pNaga4.mV.DC.N2 (control vector) was used as the backbone and restriction digested using *Hpa*I and *Sbf*I restriction enzyme to remove the control (HSP90) promoter sequence. The putative promoter sequences of genes (*EEP, EWM28686.1; SPS, EWM20309.1; GDPD, EWM28518.1* and *ORF89, EsV-1-89*; NCBI gene ID: 920795) were amplified using Q5® High Fidelity DNA polymerase (New England Biolabs, USA) and sequence-specific primers with *Hpa*I and *Sbf*I restriction sites (Table S10). The amplified putative promoter regions were purified using a Monarch® DNA Gel Extraction Kit (New England Biolabs, USA). These gel-purified promoter sequences were independently ligated with the double digested vector backbone using the Quick Ligation Kit (New England Biolabs, USA).

The pPTPBR11_p49202-mVENUS episomal vector with a yeast centromeric sequence (Fabris et al., submitted), was used as the vector backbone and restriction digested using *NdeI* and *SpeI* restriction enzyme to remove the endogenous 49202 promoter sequence from *P. tricornutum.* The putative promoter sequence of *ORF89* (*EsV-1-89*; NCBI gene ID: 920795) was amplified using Q5® High Fidelity DNA polymerase (New England Biolabs, USA) and sequence-specific primers (Table S10) and purified using a Monarch® DNA Gel Extraction Kit (New England Biolabs, USA). The gel-purified putative viral promoter fragment was ligated with the double digested vector backbone using Gibson Assembly Master Mix (New England Biolabs, USA) to create the pPTPBR11_pEsv1-mVENUS.

DH5-alpha *Escherichia coli* electrocompetent cells (New England Biolabs, USA) were used for DNA cloning and were grown in Luria broth (Miller's LB base) medium (Invitrogen, Life Technologies, USA) with 100 µg mL^{-1} ampicillin. Plasmid DNA was isolated using the Zippy Plasmid Miniprep Kit (Zymo Research, USA). The insert

sequences in the resultant vector variants of pNaga4.mV.DC.N2 and pPTPBR11_pEsv1-mVENUS were confirmed by DNA sequencing (Macrogen, South Korea).

3.2.5 *N. gaditana*, *N. oceanica* and *P. tricornutum* transformation and screening

The *N. gaditana* strain CCMP526 and *N. oceanica* strain CCMP1779 were transformed with the pNaga4.mV.DC.N2 and the promoter variants. The linearised plasmids were introduced into the cell via electroporation (Chapter 2 Materials and methods section 2.5)[38]. The *P. tricornutum* Bohlin strain CCAP1055/1 was transformed with the circular pPTPBR11_pEsV1-mVenus using bacterial conjugation[103,104]. Zeocin-resistant colonies were screened by flow cytometry using a flow cytometer (CytoFLEX LX, Beckman Coulter Life sciences, USA) to check for mVenus reporter fluorescence using a 488 nM laser and 525/40 optical filter.

3.2.6 *N. gaditana* chlorophyll α fluorescence measurement

Pulse-amplitude modulated fluorometry (PAM) was used to measure the maximum quantum yield of photosystem II (Fv/Fm calculated as Fv = Fm − Fo, where Fo is the minimum fluorescence in the dark and Fm the maximum fluorescence) as a proxy for photosynthetic health. The measurement is a rapid assessment of transgenic *N. gaditana* photosynthetic activity to determine photosystem stress due to transgene introduction via electroporation and photosystem damage due to random gene integration when compared to *N. gaditana* wild type and empty vector control. A Pocket-PAM (Gademann Instruments GmbH, Germany) was used to take the measurements after the cultures were acclimated for 20 mins in darkness at room temperature. The settings used are blue light, saturation pulse intensity – 12 (>5000 µmol photons m^{-2} s^{-1}), saturation pulse width of 0.6 s with gain.

3.2.7 Flow cytometry analysis of *N. gaditana* transformants

Four selected *N. gaditana* clones expressing *mVenus* under the control of (*SPS*, EWM20309.1; *GDPD*, EWM28518.1) gene promoters were inoculated into 50 mL culture media in 250 mL Erlenmeyer flasks in triplicate. Flow cytometric analysis of cells was performed once each day to estimate the cell count, measure mVenus fluorescence using a 488 nM laser and 525/40 optical filter and chlorophyll measurement using the 690/50 bandpass filter. Empty vector control was used to estimate the chlorophyll

background fluorescence. The gating used to select the mVenus positive cells is shown in Figure S10. The mean and median mVenus fluorescence intensity and the percentage of cells expressing mVenus protein were calculated using *N. gaditana* wild-type strain and *N. gaditana* transformant without mVenus protein as negative controls. The net mVenus fluorescence was calculated as a product of mVenus intensity/cell and number of cells expressing mVenus fusion protein.

3.2.8 Gene expression analysis by Real-Time quantitative PCR (RT-qPCR)

Fifty mL of mid-log phase cultures of *N. gaditana* CCMP526 ($1x10^5$ cells mL^{-1}) were pelleted and RNA extracted using the protocol described in Chapter 2 (section 2.2.3). A total of 320 ng of RNA was used for each sample for cDNA synthesis using iScript™ gDNA clear cDNA Synthesis Kit (Biorad, USA). Data analysis was done in single threshold mode. RT-qPCR primers used to amplify *mVenus* and β *actin* genes are given in Table S11 including their primer efficiency and R^2 values. The starting quantity (SQ) of *mVenus* was normalised using the average of β *actin* SQ (which has been used as a housekeeping gene in previous studies[105]) and *G3P* (glyceraldehyde 3-phosphate) SQ to give the relative abundance of *mVenus* mRNA in all the samples analysed.

3.2.9 Statistical analysis

Statistical analyses were done using GraphPad Prism 8.0. 2-way ANOVA parametric test was applied, and significant effects were analysed using Dunnett's multiple comparison test for the cell count and Fv/Fm data and Turkey's multiple comparison test for mVenus fluorescence intensity measurements for all *N. gaditana* transgenic lines including controls. The analyses tested the null hypothesis that there was no difference in growth, chlorophyll α fluorescence in all the lines tested, and all transgenic lines under study showed no difference in mVenus fluorescence in phosphate replete and depleted medium. Significant differences between the control (*Hsp90*) promoter and the EcV-1 viral (*Ecto*) promoter was calculated using a non-parametric t-test (Wilcoxon matched-pairs signed rank test).

3.3 Results and discussion

3.3.1a Identification of putative viral promoter candidate from *Ectocarpus siliculosus virus 1 genome*

E. siliculosus virus (EsV-1) is pandemic in populations of the marine brown macroalga *E. siliculosus*[57]. The genome of EsV-1 is made of double-stranded DNA[57,106]. The genome browser http://www.virusite.org/[95] was used to extract the putative promoter region (Table S13) of multiple ORFs which code for putative proteins (Table 5), but most sequences were not suitable for synthesis due to the presence of tandem repeats[55] hence the putative promoter of *ORF89* (*EsV-1-89*; NCBI gene ID: 920795) whose sequence was compatible for synthesis was extracted. This was followed by manual screening (no online promoter analysis tools available to analyse putative promoters from algal viruses) to identify conserved promoter motifs that are found to be conserved in members of phycodynaviruses[98]. These motifs include a highly conserved 10 nucleotide sequence AAAAATAnTT observed in chlorellaviruses[99], three AT rich conserved regions (ArnTTAAnA, AATGACA and GTnGATAyr)[100] and sequences (ATGACAA and TATAAAT) similar to eukaryotic "TATA box" observed in chloroviruses[101] (y- pyrimidine, r- purine, n- any nucleotide) along with PlantCARE[96] and TSSP analysis[107] (Table S15). Based on these promoter motifs, a 700–1300 bp of intergenic sequence was selected as a putative promoter region (Table 5). Out of the four regions selected, only one had sequence compatibility for synthesis, hence the 833 bp before ORF89 was sent for fragment synthesis.

Table 5. The selected ORFs of *Ectocarpus siliculosus virus-1*[57,106] and their NCBI gene and protein IDs along with putative annotations.

Organism	Open Reading Frame	NCBI gene ID	NCBI protein ID	Putative promoter sequence size	Putative annotation
EcV1	ORF89	920795	NP_077574	833 bp	NA
EcV1	ORF66	920718	NP_077551	750 bp	helicase

EcV1	ORF115	920597	NP_077600.1	1243 bp	Protein with iron-sulphur binding region
EcV1	ORF171	920659	NP_077656.1	1459 bp	NA

3.3.1b Identification of phosphate-inducible promoter candidates in *N. gaditana*

Under phosphate limited conditions, the gene expression dataset of *N. oceanica*[92] was obtained using NCBI Gene Expression Omnibus[108] (GEO) database and then the differentially expressed genes under phosphate-replete and deplete conditions were screened using GEO2R[109] as such a dataset is not available for *N. gaditana.* Three genes that showed a logFC value (log 2-fold change of significant expressed genes between phosphate-replete and deplete conditions) greater than 4 were selected (Table 6). The putative annotation of these genes such as phospholipid recycling (*GDPD,* EWM28518.1), inorganic orthophosphate transporter (*SPS*, EWM20309.1), purple acid phosphatases (*EEP*, EWM28686.1)[92], suggested that they could be involved in functions related to phosphate acquisition, transport and metabolism. Promoters of similar genes such as alkaline phosphatase (XP_002184454.1[110]) from *P. tricornutum*[88] and sulphoquinovosyldiacylglycerol 2 (*SQD2*, XM_001689610.1) from *C. reinhardtii*[86,87] were found to be phosphate-inducible. In plants for example, the soybean vegetative storage protein gene *VspB* encoding vacuolar glycoprotein acid phosphatase was found to be inducible in the presence of sugars and under phosphate limitation[111], a gene encoding β-glucosidase was found to be induced under phosphate-limiting conditions in the root in *Arabidopsis thaliana*[112,113].

The potential orthologs of the three *N. oceanica* putative phosphate-inducible genes in the *N. gaditana* genome were identified, and their upstream putative promoter sequences (Table S12) were extracted using the CRIBI Genomics *Nannochloropsis* genome portal[102]. Since computational tools to analyse the structure of algal promoter are not available, plant-specific tools PlantCARE[96] and Softberry TSSP[97,107] (Table S14 & S16) were used to predict transcription factor binding sites and cis-regulatory elements

such as TATA and CAAT boxes and their position with respect to the start of transcription (TSS).

The computational data showed the presence of putatively conserved promoter domains such as CAAT-boxes, TATA-boxes (core promoter elements that are usually 80-150 bp and 30 bp upstream of the start codon) and cis-acting regulatory elements such as G box which are involved in light responsiveness in plants[114–117].

Other putative transcription factor (TF) binding sites such as TGACG-motif which is a cis-acting regulatory element involved in the MeJA-responsiveness[118], MYB (TF family involved in biotic and abiotic stresses, development, differentiation, metabolism, defense in plants)[119,120] were present (Table S12). Also, the ABRE element (cis-acting element involved in the abscisic acid responsiveness)[121,122] were also found in all the tested promoters, whereas the stress response element (STRE)[123] was only found in (*SPS*, EWM20309.1) gene promoter. The transcription factors that bind to these elements such as MYB (myeloblastosis proto-oncogene, transcription factor), NF YB (nuclear transcription factor-Y subunit beta), NF YC (nuclear transcription factor-Y subunit gamma), bZIP (basic leucine-zipper transcription factor). bHLH (basic helix-loop-helix transcription factor), E2F (cellular transcription factor involved in transcription control during the cell growth cycle), HSF (heat shock transcription factor) and AP2 (transcription factor APETALA2- integrase-type DNA-binding superfamily) found in higher plants are also present in *N. gaditana*[53,124]. Based on these promoter motif predictions, a 600 –1000 bp of intergenic sequence was selected as a putative promoter for each gene (Table S12).

Out of the three promoters, only two were amplifiable (*SPS* and *GDPD*). The *HSP90* promoter region in the pNaga4.mV.DC.N2 vector was replaced with these two putative promoter regions, independently.

Table 6. Genes selected from *N. oceanica* CCMP1779 transcriptomic data[92] and their corresponding orthologs identified in the *N. gaditana* genome using NCBI BLASTp and their annotation.

The raw P-values, moderated t-statistic (t), B-statistic i.e log-odds that the gene is differentially expressed, Log2-fold change between two experimental conditions (phosphate replete and deplete conditions) were obtained using GEO2R[109] web tool.

Abbreviation	ID	Genmodel ID *N. oceanica* 1779	P.Value	t	B	logFC	*N. gaditana* GeneBank ID	Annotation	E-value	*N. gaditana* promoter sequence location	promoter size (bp)
EEP	6829	NannoCCMP1779_9015	9.7E-10	12.96	11.817	4.434	EWM28686.1	Endonuclease/exonuclease/phosphatase	3e-139	NG_chr04:459440..460045	605
SPS	12076	NannoCCMP1779_1557	1.01E-06	7.75	5.847	4.054	EWM20309.1	Sodium phosphate symporter	9e-047	NG_scf33:7650..8644	994
GDPD	1519	NannoCCMP1779_4935	2.46E-06	7.21	5.024	4.043	EWM28518.1	Glycerophosphoryl diester phosphodiesterase	0	NG_chr04:1048440..1049438	998

3.3.2 Characterisation of transgenic *N. gaditana* cells carrying the putative viral promoters

The transformation of *N. gaditana* via electroporation generated only 12 zeocin-resistant colonies with the viral promoter construct. The level of mVenus fluorescence varied significantly among clones that were analysed (Figure S11, S12 & S13). Of the 12 transgenic clones that were obtained with the viral promoter (EsV-1-89; NCBI gene ID: 920795) construct, only 8 survived subculturing and were subsequently screened for mVenus expression. Finally, only two maintained the construct after three months. Also, not all transgenic lines that were screened resulted positive for mVenus fluorescence, as the selection and reporter cassettes are independent of each other. It is plausible that only a portion of the vector was integrated into the genome and not the *mVenus* reporter gene, giving rise to antibiotic resistant colonies that did not exhibit any mVenus fluorescence when analysed using flow cytometry.

Similar to what was observed in Chapter 2, in some cases of random chromosomal integration of transgenes, the level of mean mVenus fluorescence varied amongst transformants, presumably due to multiple factors[38], including i) position effects due to random chromosomal transgene integration and influence of neighboring regulatory elements on transgene expression[125–127]; ii) number of intact transgene integrations[128]; iii) integrative fragmentation or DNA re-arrangement, including deletion or translocation, which can also influence the growth and fitness of the host[129–131]; and iv) various forms of epigenetic and homology-dependent gene silencing[132,133]. Also, amongst the transgenic lines that were positive for mVenus expression, the population was heterogeneous (Figure S10), similar to what has been observed in other microalgae[83,134,135]. A similar phenomenon has been reported in mammalian systems as well[125], where the heterogeneity was found to be due to silencing position effects, which could lead to multiple cell populations with varying levels of expression within the same transgenic line[136]. Variations within transgenic cell lines can also be non-genetic and dependent on other factors, such as cell age, unequal cell division and cell cycle stage in non-synchronised cultures[137,138]. Hence, measuring the transgene expression from individual cells provides an accurate account compared to bulk measurements[139].

Three clones expressing *mVenus* under the control of *EsV-1* promoter was selected and compared to three clones expressing the *mVenus* under the control of the *HSP90* promoter. Since the random chromosomal integration of transgenic constructs may result in rearrangements and deletions of genes relevant for fitness and physiology of the alga, growth using cell count and PAM to measure the maximum quantum yield of photosystem II were measured daily.

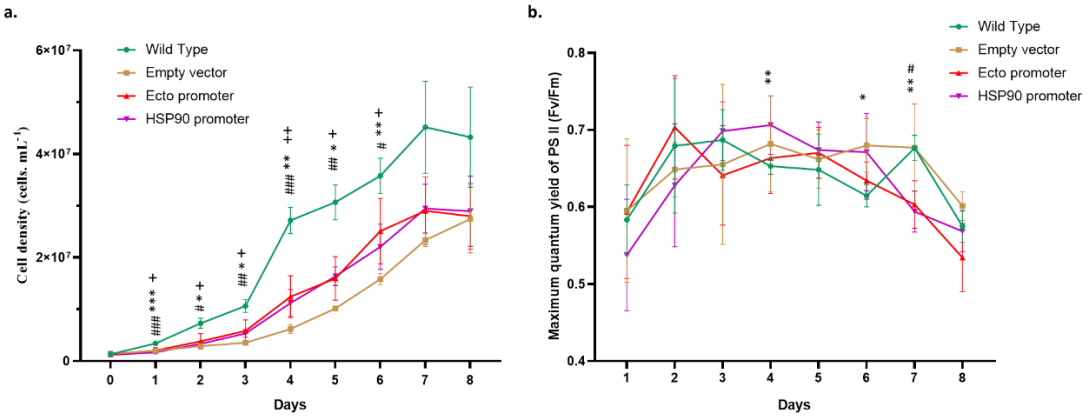


Figure 10. Growth and photosynthetic activity of wild type and transgenic *N. gaditana* cultures measured over a period of 9 days.

a. The growth curve of all the transgenic lines under viral promoter study including wild type and negative control. **b.** Photosynthetic activity of wild type and transgenic *N. gaditana* cell lines. Error bars represent the standard deviation for wild type and empty vector cell lines (n=3) and for the putative viral promoter cell lines and positive (HSP90) control (n=9). Significant differences in growth between wild type and the transgenic lines (empty vector- +, (Hsp90) control promoter- *, (Ecto) viral promoter- #) were calculated using 2-way ANOVA Dunnett's multiple comparison test * P<0.05, ** P<0.005, *** P<0.0005).

Wild type and empty vector control performed better than the rest of the transgenic cell lines when considering growth rate and photosynthetic activity. Differences in growth (Figure 10a) and photosynthetic activity (Figure 10b) between wild-type and transgenic lines suggest that random genomic integration may have had detrimental effects on algal photosynthetic fitness due to random insertional mutagenesis, such as genetic rearrangements and disruption of relevant genetic loci[129–131,140]. Photosynthetic activity is linked to growth of cells and if photosynthetic ability to produce energy is affected it would invariably affect algal growth rate.

3.3.3 Characterisation of transgenic *N. gaditana* cells carrying the putative phosphate-inducible promoters

The transformation of *N. gaditana* via electroporation generated 20-50 zeocin-resistant colonies per promoter construct. As observed in transgenic *N. gaditana* lines with the viral promoter (Figure S13), the level of mean mVenus fluorescence varied significantly among the transgenic *N. gaditana* lines with the putative phosphate-inducible promoters (Figure S11 and S12) that were analysed; this is possibly due to the same factors discussed earlier[38].

In order to test the effect of random transgene integration into the transgenic *N. gaditana* lines, four transgenic cell lines were selected per promoter construct, including positive control (*HSP90*) along with empty vector control and wild type *N. gaditana*. These lines were up-scaled to 50 mL in 250 mL culture flasks in triplicate. The growth curve and maximum quantum yield of photosystem II of all the transgenic and wild-type *N. gaditana* lines are shown in Figure 11. The cultures were introduced to a phosphate-limiting growth medium on Day 6 (mid-log) to confirm promoter activity of the putative phosphate-limitation inducible promoters in the transgenic *N. gaditana* lines.

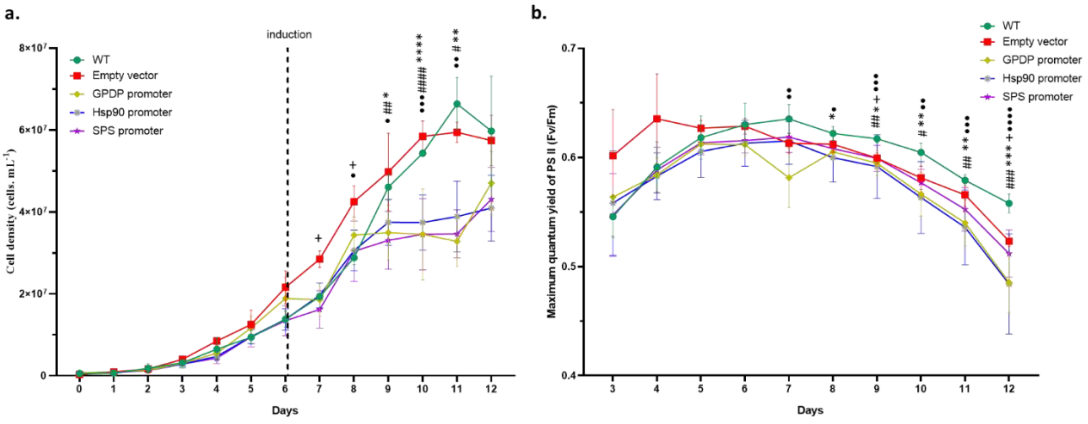


Figure 11. Growth of photosynthetic activity of wild type and transgenic *N. gaditana* cultures measured over a period of 12 days.

a. The growth curve of all the transgenic lines under P-inducible promoter study including wild type and negative control. **b.** Photosynthetic activity of wild type and transgenic *N. gaditana* cell lines measured under phosphate replete and phosphate deplete conditions. Error bars represent the standard deviation for wild type and empty vector cell lines (n=3) and for the putative P-inducible (GDPD and SPS) promoter cell lines and positive (HSP90) control (n=9). Significant differences in growth between wild type and the transgenic lines (empty vector- +,

Hsp90 promoter- *, SPS promoter- #, GDPD promoter- •) were calculated using 2-way ANOVA Dunnett's multiple comparison test * P<0.05, ** P<0.005, *** P<0.0005, **** P<0.0001.

Like what observed in 3.3.2, wild type *N. gaditana* performed better than the rest of the transgenic cell lines concerning growth rate and photosynthetic activity. Differences in growth (Figure 11a) and photosynthetic activity (Figure 11b) between wild-type and transgenic lines suggest that random genomic integration may have had detrimental effects on algal photosynthetic fitness due to random insertional mutagenesis, such as genetic rearrangements and disruption of relevant genetic loci[129–131,140]. Also, the growth of wild type and transgenic *N. gaditana* was not affected in phosphate-limiting growth media, as evident from Figure 11a, they maintained a steady exponential growth till day 11 (wild type) i.e. till 120 hrs after resuspension in phosphate-limiting media, Day 10 (empty vector control) i.e. till 96 hrs after resuspension in phosphate-limiting media and Day 9 (phosphate-inducible promoter transgenic lines) i.e. till 72 hrs after resuspension in phosphate-limiting media. Iwai et al., monitored *Nannochloropsis* strain NIES-2145 cell growth under phosphate deprivation for 13 days and highlighted that growth was not affected under this condition[87]. Muhlroth et al., report exponential growth of *N. oceanica* under phosphate-starvation for up to 48 hrs and slowly decline after 96 hrs[92]. As stated earlier, *Nannochloropsis* spp. have internal reserves or P storage and has a mechanism to cope with phosphate limitation via P recycling and changes in lipid classes hence photosynthesis occurs and cells do not stop growing[91,92]. Therefore, promoters that are inducible under phosphate limitation are suitable candidates that require further exploration to understand their regulation, as they can serve as good components for construction of industrially relevant chimeric promoters.

3.3.4 No phosphate limitation induced promoter activity was evident in the transgenic *N. gaditana* lines

The mVenus fluorescence of all transgenic *N. gaditana* cell lines under the putative phosphate-inducible promoters was tracked for 6 days under phosphate replete and 6 days under phosphate deplete condition (Figure 12). The transgenic lines expressing mVenus under the control of *SPS*, EWM20309.1 gene promoter showed low basal expression which was steadily maintained under both conditions and no

phosphate limitation induced promoter activity was observed. On the other hand, the transgenic lines expressing mVenus under the control of *GDPD,* EWM28518.1 gene promoter did not show any mVenus expression under both conditions (mVenus fluorescence levels similar to empty vector control). The *N. gaditana* cell lines expressing mVenus under the control of *Hsp90* (Nga00934) positive control showed mVenus fluorescence with a mean mVenus intensity which was 5.2-fold higher than the wild type background and showed steady decline under phosphate limitation, suggesting that gene expression via the *Hsp90* (Nga00934) promoter under phosphate-limitation was not ideal.

The percentage of cells expressing mVenus (Figure 12b) in the transgenic lines expressing mVenus under the control (*SPS*, EWM20309.1 and *GDPD,* EWM28518.1) was very low (close to 2%) which could be a possible explanation as to the lack of induction observed here when a clear phosphate limitation induced mVenus expression was observed while screening (Figure S11 and S12). The reason being the stability of the transgenic lines i.e. the transgenic lines expressing mVenus under the control of these putative phosphate-inducible promoters might not have stably integrated into the nuclear genome of *N. gaditana,* unlike the positive (*Hsp90*) transgenic lines which displayed stable mVenus fluorescence over a period of six months (data not shown).

Based on these results, we conclude that even though no phosphate limitation induced promoter activity was evident in the transgenic *N. gaditana* lines carrying (*SPS*, *EWM20309.1* and *GDPD, EWM28518.1*) gene promoters, which could be due to stability of the transgenic lines; it is possible that new stable transgenic *N. gaditana* lines could be generated in the future to evaluate the promoter activity of these putative promoters under phosphate-replete and deplete conditions. These stable transgenic lines with at least 50% of the sub-population expressing the mVenus could be used to confirm promoter activity under phosphate-limitation, which was observed in the initial screening (Figure S11 and S12). If these putative promoter candidates were to be confirmed, it would enrich the suite of inducible promoters available for researchers to use in a variety of biotechnology applications.

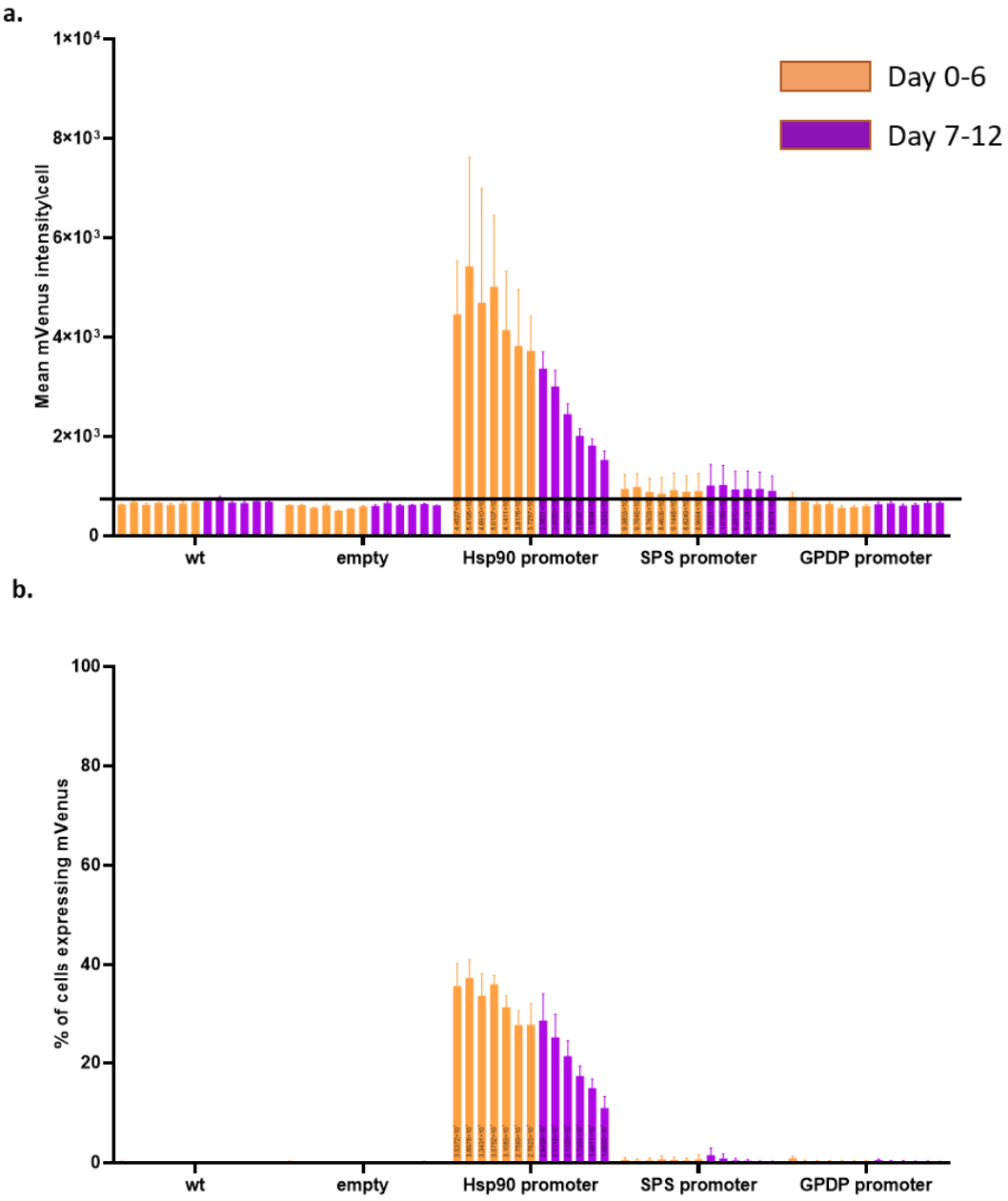


***Figure 12. mVenus* fluorescence measured over twelve days in transgenic *N. gaditana* expressing putative phosphate-inducible promoters.**

a. Mean mVenus fluorescence per cell; **b**. Percentage of cells expressing mVenus. Error bars represent the standard deviation for wild type and empty vector cell lines (n=3) and for the putative phosphate-inducible promoter cell lines and positive (HSP90) control (n=9). The bars in orange represent phosphate-replete condition and purple bars denote phosphate-deplete condition.

3.3.5 The putative viral *EsV-1-89* promoter drove mVenus expression in transgenic *N. gaditana*

The transgenic lines carrying the viral promoter (*EsV-1-89*; NCBI gene ID: *920795*)

were analysed over a period of 9 days to detect the promoter activity in these lines by measuring mVenus fluorescence *in vivo* and determine if the expression is steady across the various growth stages (Figure 13). The viral promoter was able to drive the reporter gene expression with consistent expression across the different growth stages (Figure 13a). Transgenic lines expressing mVenus under the control of *Hsp90* (*Nga00934*) promoter[38] were used as a positive control to test the performance of the viral promoter. For all the days tested, the net fluorescence intensity (Figure 13) of the transgenic *N. gaditana* lines with the viral promoter was significantly lower than the control promoter, where the promoter activity based on mVenus fluorescence was 26% of the control (*Hsp90*) promoter i.e. the viral promoter was less efficient at driving transgene expression compared to the endogenous control promoter. In order to study this viral promoter further, mVenus expression in these transgenic lines carrying the EcV-1 viral promoter were evaluated based on mRNA.

3.3.6 Viral promoter activity estimated using transcript abundance and mVenus fluorescence

Based on the growth data (Figure 10), we know that under the laboratory condition used, the logarithmic phase occurred from Day 1–7 and we chose the mid-log phase (Day 3-4) as the point of harvest for two reasons i) to obtain sufficient biomass, and ii) have optimal mVenus fluorescence (Figure 10). To measure transcript abundance of mVenus, three independent transgenic cell lines expressing the viral and control constructs were scaled-up to 250 mL in 500 mL culture flask in triplicate. Transgenic *N. gaditana* lines were cultivated in batch mode over a period of 5 days, while keeping the culture parameters consistent with the previous pilot experiment. No significant differences in growth were observed between wild type and transgenic lines expressing the viral and positive control promoter (Figure 13). Only significant differences in growth were observed between wild type and empty vector control (Figure 13). mVenus fluorescence was tracked daily to follow the expression level *in vivo* of the fusion protein, until Day 4 (Figure 13). A temporal expression pattern similar to the initial pilot experiment (Figure 13) was observed, with increasing net mVenus fluorescence as the cultures approached exponential phase (Figure 13). The activity of the viral promoter (*EsV-1-89*; NCBI gene ID: *920795*) based on the net fluorescence intensity was

significantly lower than the constitutive promoter control, which was 4.5% of the control (*Hsp90*) promoter which was also much lower than what was observed in the pilot experiment (26%). The culture parameters were kept consistent, but a significant difference was observed due to the fluctuation in the sub-population of cells expressing mVenus in the transgenic *N. gaditana* lines (which changed from 25-40% to 10-15%) (Figure 13 and 14) and not the mean mVenus fluorescence intensity which remained the same in both the experiments (pilot- Figure 13 and large scale- Figure 14).

Once the cells were harvested on Day 4, the relative transcriptional promoter activity for each promoter construct was determined by quantification of transgene transcripts using RT-qPCR. At the transcriptional level, the average relative promoter activity measured was 20.6% of the control (*Hsp90)* promoter. The mVenus transcript abundance observed corelated with the mean mVenus fluorescence detected (Figure 14). Based on these results, we concluded that the putative viral promoter (*EsV-1-89*; NCBI gene ID: *920795*) could drive transgene expression in *N. gaditana,* but its activity is not stronger than the endogenous *Hsp90 (Nga00934)* promoter[38].

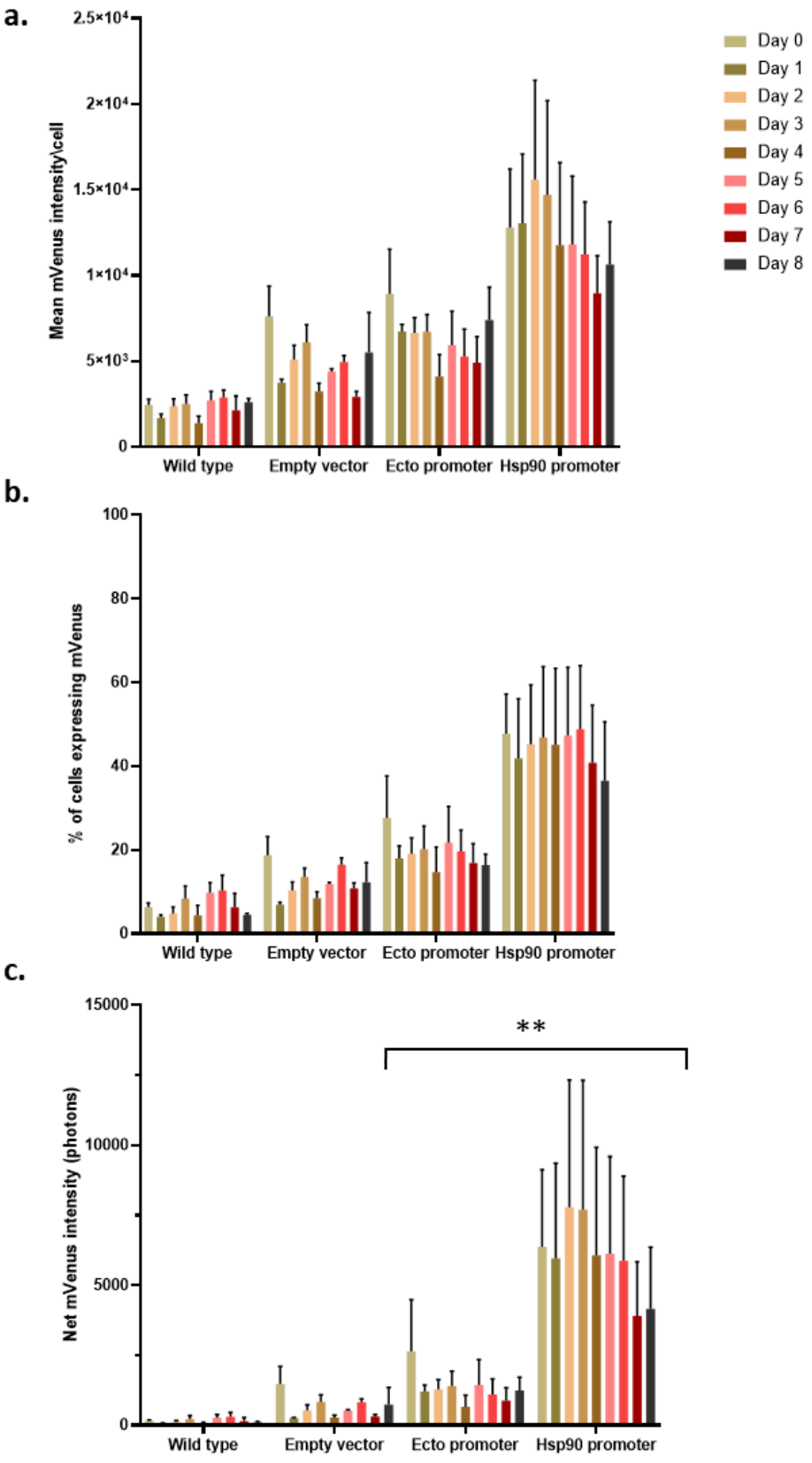


***Figure 13. mVenus* fluorescence measured over nine days in transgenic *N. gaditana* expressing putative viral promoter.**

a. Mean *mVenus* fluorescence per cell; **b.** percentage of cells expressing *mVenus*; **c.** Net *mVenus* fluorescence. Error bars represent the standard deviation for wild type and empty vector cell lines (n=3) and for the putative viral promoter cell lines and positive (HSP90) control (n=9). Significant differences between the control (Hsp90) promoter and the EcV-1 viral (Ecto) promoter was calculated using a non-parametric t-test: Wilcoxon matched-pairs signed rank test. ** P<0.05

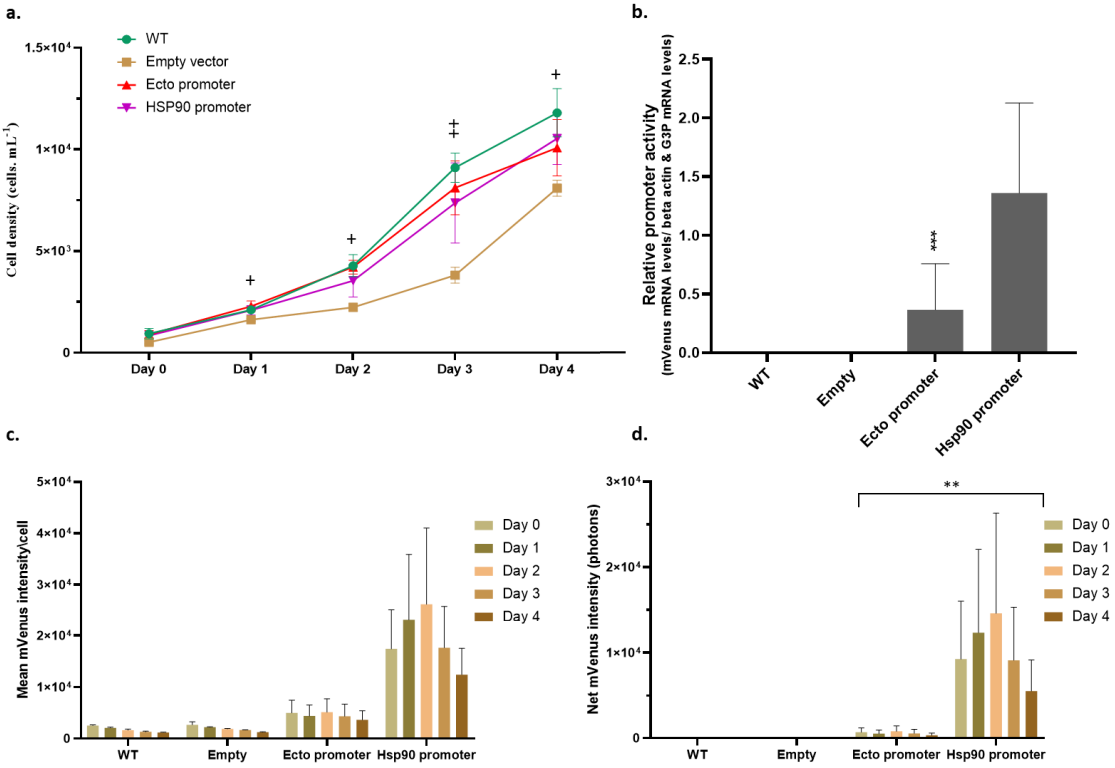


Figure 14. Relative mRNA and protein expression of *mVenus* in transgenic *N. gaditana* carrying viral promoter.

a. The growth curve of all the transgenic lines including wild type and empty vector control; **b**. Relative quantification of *mVenus* transcript abundance normalised to β-*actin and G3P* mRNA in all *N. gaditana* transformant lines, driven by the viral promoter, Hsp90 control promoter and in the empty vector control; **c.** Mean mVenus fluorescence per cell; **d.** Net mVenus fluorescence calculated as a product of mVenus intensity/cell and number of cells expressing mVenus. Error bars represent the standard deviation for wild type and empty vector cell lines (n=3) and for the putative viral promoter cell lines and positive (HSP90) control (n=9). Significant differences in growth only observed between wild type and the empty vector control were calculated using 2-way ANOVA Dunnett's multiple comparison test * P<0.05, ** P<0.005). Significant differences between the relative abundance of mVenus under the control (Hsp90) promoter and the EcV-1 viral (Ecto) promoter was calculated using paired t-test (2 tailed) *** P= 0.0008. Significant differences between Net fluorescence intensity of the control (Hsp90) promoter and the EcV-1 viral (Ecto) promoter was calculated using a non-parametric t-test: Wilcoxon matched-pairs signed rank test. ** P<0.05

3.3.7 Characterisation of transgenic *P. tricornutum* and *N. oceanica* cells carrying the putative viral promoter

The activity of the viral promoter (*EsV-1-89*; NCBI gene ID: *920795*) was confirmed. It was found to have moderate promoter activity unlike the viral promoters used in higher plants[28,29,48] and animal systems[141–143]. In some cases, these viral promoters are universal as most of them depend on host transcription machinery for gene expression such as the CamV 35S viral promoter[21,28,29]. We hypothesized that the viral promoter (*EsV-1-89*; NCBI gene ID: *920795*) would serve as a universal promoter and function in other taxa of the stramenopile group. Therefore, we selected the model diatom *P. tricornutum* and another species of *Nannochloropsis* (*N. oceanica*) to test this hypothesis. *N. oceanica* was transformed using pNaga4.mV.DC.N2.Ecto vector via electroporation, but no transgenic colonies expressing mVenus were obtained. *P. tricornutum* was transformed using the pPTPBR11_pEsv1-mVENUS episomal vector via bacterial conjugation and >200 zeocin-resistant colonies were obtained. 48 individual colonies were randomly selected and screened for mVenus expression using flow cytometry Figure S14. Bacterial conjugation enables transfer of intact episomal plasmid into the *P. tricornutum* cell[104] and the yeast centromeric sequence facilitates autonomous replication of the introduced plasmid[103,104] without nuclear genome integration which indicates that both zeocin-resistance gene cassette and mVenus reporter gene cassette are present in all the 48 *P. tricornutum* transgenic lines. But these zeocin resistant transgenic lines did not display mVenus fluorescence and had a mean mVenus fluorescence intensity similar to wild type *P. tricornutum* when analysed using a flow cytometer (Figure S14). This implies that the putative viral promoter region from EsV-1 does not function in *P. tricornutum.* The non-functional nature of this putative viral promoter in an episomal setting raises the question of its authentic gene expression capacity in *N. gaditana*. Maybe the two transgenic *N. gaditana* lines that showed stable mean mVenus fluorescence (Figure 7c) under the control of the putative viral promoter (Ecto) might be due to the influence of neighboring endogenous promoter-enhancer elements[144,145]. Such elements might influence reporter gene expression and can only be confirmed through whole genome sequencing and analysis of the genomic DNA from the above-mentioned transgenic *N. gaditana* lines, but it is beyond the scope of this

study.

3.4 Conclusion

The establishment of algal synthetic biology tools and resources has led to broader applications of industrially relevant microalgae. *Nannochloropsis spp,* are oleaginous model microalgae, a relatively new genetic engineering candidate. To date, several whole-genome datasets produced in *Nannochloropsis* species have been made available, and so are the genomic transformation protocols and multiple over-expression or repressor studies using constitutive and viral promoters[37]. But viral promoters from algal viruses have not been examined for transgene expression in *N. gaditana*. Such investigations have been conducted with another stramenopile *P. tricornutum*[39,146]*.* In this study, we identified and analysed a viral promoter of *ORF89* (*EsV-1-89*; NCBI gene ID: *920795*) from *Ectocarpus siliculosus* virus 1 and demonstrated that it could drive reporter gene expression to a moderate level compared to the *HSP90 (Nga00934)* promoter control in *N. gaditana,* but not in *P. tricornutum.*

Then again, over-expression studies using inducible promoters are still in its infancy in *Nannochloropsis*[78] with a single nitrate inducible promoter characterised by Jackson et al., and similar promoters are available for other model microalgae such as *P. tricornutum*[88,147]*, Cyclotella* cryptica[83] and *C. reinhardtii*[77,87]*.* No literature is available on stress-inducible promoter study such as phosphate limitation in *N. gaditana*. The putative phosphate-inducible promoter candidates selected still holds promise for use as a stress-inducible expression system in spite of no detectable induction which was due to lack of stability of the transgenic *N. gaditana* lines. Further research should be undertaken to generate new and stable transgenic lines and evaluating transgene expression under phosphate replete and deplete conditions.

In conclusion, the new promoters tested expand the suite of promoters available for genetic engineering this microalga for a broad range of biotechnology applications.

3.5 Acknowledgements

This research was supported by UTS: C3 funding. MF is supported by a CSIRO Synthetic

Biology Future Science Fellowship, co-funded by CSIRO and the University of Technology Sydney. The authors would like to thank Kun Xiao for technical assistance with flow cytometry, Stephen Woodcock for statistical analysis.

Statement of informed consent, human/animal rights

No conflicts, informed consent, human or animal rights applicable.

3.6 References

1. Maniatis, T. O. M., Goodbourn, S. & Fischer, J. A. Regulation of inducible and tissue-specific gene expression. *Science.* **236**, 1237–1245 (1987).

2. Ayoubi, T. A. Y. & Van De Ven, W. J. M. Regulation of gene expression by alternative promoters. *FASEB J.* **10**, 453–460 (1996).

3. Smale, S. T. Core promoters: Active contributors to combinatorial gene regulation. *Genes Dev.* **15**, 2503–2508 (2001).

4. Smale, S. T. & Kadonaga, J. T. The RNA Polymerase II Core Promoter. *Annu. Rev. Biochem.* **72**, 449–479 (2003).

5. Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* **11**, 394–403 (2004).

6. Porto, M. S. *et al.* Plant promoters: An approach of structure and function. *Mol. Biotechnol.* **56**, 38–49 (2014).

7. Gagniuc, P. & Ionescu-Tirgoviste, C. Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters. *BMC Genomics* **13**, 1 (2012).

8. Bucher, P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**, 563–578 (1990).

9. Kiran, K. *et al.* The TATA-box sequence in the basal promoter contributes to determining light-dependent gene expression in plants. *Plant Physiol.* **142**, 364–376 (2006).

10. Yamamoto, Y. Y. *et al.* Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics* **8**, 1–23 (2007).

11. Fujimori, S., Washio, T. & Tomita, M. GC-compositional strand bias around transcription start sites in plants and fungi. *BMC Genomics* **6**, 1–11 (2005).

12. Molina, C. & Grotewold, E. Genome wide analysis of *Arabidopsis* core promoters. *BMC Genomics* **6**, 25 (2005).

13. Ioshikhes, I. P. & Zhang, M. Q. Large-scale human promoter mapping using CpG islands. *Nat. Genet.* **26**, 61–63 (2000).

14. Juo, Z. S. *et al.* How proteins recognize the TATA box. *J. Mol. Biol.* **261**, 239–254 (1996).

15. Tatarinova, T., Brover, V., Troukhan, M. & Alexandrov, N. Skew in CG content near the transcription start site in Arabidopsis thaliana. *Bioinformatics* **19**, i313–i314 (2003).

16. Mantovani, R. The molecular biology of the CCAAT-binding factor NF-Y. *Gene* **239**, 15–27 (1999).

17. Lee, T. I. & Young, R. A. Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**, 77–137 (2000).

18. Hernandez-Garcia, C. M. & Finer, J. J. Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci.* **217**–**218**, 109–119 (2014).

19. Zou, C. *et al.* Cis-regulatory code of stress-responsive transcription in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14992–14997 (2011).

20. Potenza, C., Aleman, L. & Sengupta-Gopalan, C. Invited review: Targeting transgene expression in research, agricultural, and environmental applications: Promoters used in plant transformation. *Vitr. Cell. Dev. Biol. - Plant* **40**, 1–22 (2004).

21. Odell, J. T., Nagy, F. & Chua, N.-H. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* **313**, 810–812 (1985).

22. Liu, W. & Stewart, C. N. Plant synthetic promoters and transcription factors. *Curr. Opin. Biotechnol.* **37**, 36–44 (2016).

23. Hernandez-Garcia, C. M., Martinelli, A. P., Bouchard, R. A. & Finer, J. J. A soybean (*Glycine max*) polyubiquitin promoter gives strong constitutive expression in transgenic soybean. *Plant Cell Rep.* **28**, 837–849 (2009).

24. An, G., Costa, M. A. & Ha, S. B. Nopaline synthase promoter is wound inducible and auxin inducible. *Plant Cell* **2**, 225–233 (1990).

25. Xiao, K., Zhang, C., Harrison, M. & Wang, Z. Y. Isolation and characterization of a novel plant promoter that directs strong constitutive expression of transgenes in plants. *Mol. Breed.* **15**, 221–231 (2005).

26. Yamamoto, A., Hen, R. & Dauer, W. T. The ons and offs of inducible transgenic technology: A review. *Neurobiol. Dis.* **8**, 923–932 (2001).

27. Weake, V. M. & Workman, J. L. Inducible gene expression: Diverse regulatory mechanisms. *Nat. Rev. Genet.* **11**, 426–437 (2010).

28. Comai, L., Moran, P. & Maslyar, D. Novel and useful properties of a chimeric plant promoter combining CaMV 35S and MAS elements. *Plant Mol. Biol.* **15**, 373–381 (1990).

29. Kay, R., Chan, A., Daly, M. & Mcpherson, J. Duplication of CaMV 35S Promoter Sequences Creates a Strong Enhancer for Plant Genes. *Science.* **236**, 1299–1302 (1987).

30. Romanos, M. A., Scorer, C. A. & Clare, J. J. Foreign gene expression in yeast: a review. *Yeast* **8**, 423–488 (1992).

31. Struhl, K., Chen, W., Hill, D. E., Hope, I. A. & Oettinger, M. A. Constitutive and coordinately regulated transcription of yeast genes: promoter elements, positive and negative regulatory sites, and DNA binding proteins. *Cold Spring Harb. Symp. Quant. Biol.* **50**, 489–503 (1985).

32. Qin, J. Y. *et al.* Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS One* **5**, 3–6 (2010).

33. Goven, J. Biopharming. in *Encyclopedia of Food and Agricultural Ethics* (eds. Thompson, P. B. & Kaplan, D. M.) 1–10 (Springer Netherlands, 2013).

34. Callis, J., Raasch, J. A. & Vierstra, R. D. Ubiquitin Extension Proteins of *Arabidopsis thaliana*. *J. Biol. Chem.* **265**, 12486–12493 (1990).

35. Ruohonen, L., Aalto, M. K. & Keränen, S. Modifications to the ADH1 promoter of *Saccharomyces cerevisiae* for efficient production of heterologous proteins. *J. Biotechnol.* **39**, 193–203 (1995).

36. Seo, S., Jeon, H., Hwang, S., Jin, E. & Chang, K. S. Development of a new constitutive expression system for the transformation of the diatom *Phaeodactylum tricornutum*. *Algal Res.* **11**, 50–54 (2015).

37. Poliner, E., Farré, E. M. & Benning, C. Advanced genetic tools enable synthetic biology in the oleaginous microalgae *Nannochloropsis* sp. *Plant Cell Rep.* **37**, 1383–1399 (2018).

38. Ramarajan, M., Fabris, M., Abbriano, R. M., Pernice, M. & Ralph, P. J. Novel endogenous promoters for genetic engineering of the marine microalga *Nannochloropsis gaditana* CCMP526. *Algal Res.* **44**, 101708 (2019).

39. Sakaue, K., Harada, H. & Matsuda, Y. Development of gene expression system in a marine diatom using viral promoters of a wide variety of origin. *Physiol. Plant.* **133**, 59–67 (2008).

40. McKnight, S. & Tjian, R. Transcriptional selectivity of viral genes in mammalian cells. *Cell* **46**, 795–805 (1986).

41. Macias, M. P., Huang, L., Lashmit, P. E. & Stinski, M. F. Cellular or viral protein binding to a cytomegalovirus promoter transcription initiation site: effects on transcription. *J. Virol.* **70**, 3628–35 (1996).

42. Ruecker, O., Zillner, K., Groebner-Ferreira, R. & Heitzer, M. Gaussia-luciferase as a sensitive reporter gene for monitoring promoter activity in the nucleus of the green alga *Chlamydomonas reinhardtii*. *Mol. Genet. Genomics* **280**, 153–162 (2008).

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

44. Boshart, M. *et al.* A very strong enhancer is located upstream of an immediate

early gene of human cytomegalovirus. *Cell* **41**, 521–530 (1985).

45. Thomsen, D. R., Stenberg, R. M., Goins, W. F. & Stinski, M. F. Promoter-regulatory region of the major immediate early gene of human cytomegalovirus. *Proc. Natl. Acad. Sci.* **81**, 659–663 (1984).

46. Takahashi, K. *et al.* Requirement of stereospecific alignments for initiation from the simian virus 40 early promoter. *Nature* **319**, 121–126 (1986).

47. Gruss, P., Dhar, R. & Khoury, G. Simian virus 40 tandem repeated sequences as an element of the early promoter. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 943–947 (1981).

48. Guilley, H., Dudley, R. K., Jonard, G., Balàzs, E. & Richards, K. E. Transcription of cauliflower mosaic virus DNA: detection of promoter sequences, and characterization of transcripts. *Cell* **30**, 763–773 (1982).

49. Isomura, H. & Stinski, M. F. The Human Cytomegalovirus Major Immediate-Early Enhancer Determines the Efficiency of Immediate-Early Gene Transcription and Viral Replication in Permissive Cells at Low Multiplicity of Infection. *J. Virol.* **77**, 3602–3614 (2003).

50. Walker, T. L., Purton, S., Becker, D. K. & Collet, C. Microalgae as bioreactors. *Plant Cell Rep.* **24**, 629–641 (2005).

51. Cha, T. S., Chen, C. F., Yee, W., Aziz, A. & Loh, S. H. Cinnamic acid, coumarin and vanillin: Alternative phenolic compounds for efficient *Agrobacterium*-mediated transformation of the unicellular green alga, *Nannochloropsis* sp. *J. Microbiol. Methods* **84**, 430–434 (2011).

52. Beacham, T. A. & Ali, S. T. Growth dependent silencing and resetting of DGA1 transgene in *Nannochloropsis salina*. *Algal Res.* **14**, 65–71 (2016).

53. Thiriet-Rupert, S. *et al.* Transcription factors in microalgae: Genome-wide prediction and comparative analysis. *BMC Genomics* **17**, 1–16 (2016).

54. Bañuelos-Hernández, B., Monreal-Escalante, E., González-Ortega, O., Angulo, C. & Rosales-Mendoza, S. Algevir: An expression system for microalgae based on viral vectors. *Front. Microbiol.* **8**, 1–6 (2017).

55. Wilson, W. H., Etten, J. L. Van & Allen, M. J. The Phycodnaviridae : The Story of How Tiny Giants Rule the World. *Curr. Top. Microbiol. Immunol.* **328**, (2009).

56. Van Etten, J. L., Graves, M. V., Müller, D. G., Boland, W. & Delaroque, N. Phycodnaviridae - Large DNA algal viruses. *Arch. Virol.* **147**, 1479–1516 (2002).

57. Delaroque, N., Wolf, S., Muller, D. G. & Knippers, R. The brown algal virus EsV-1 particle contains a putative hybrid histidine kinase. *Virology* **273**, 383–390 (2000).

58. Radakovits, R. *et al.* Draft genome sequence and genetic transformation of the

oleaginous alga *Nannochloropis gaditana*. *Nat. Commun.* **3**, 686 (2012).

59. Jinkerson, R. E., Radakovits, R. & Posewitz, M. C. Genomic insights from the oleaginous model alga *Nannochloropsis gaditana*. *Bioengineered* **4**, (2013).

60. Giacalone, M. J. *et al.* Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system. *Biotechniques* **40**, 355–364 (2006).

61. Corrado, G. & Karali, M. Inducible gene expression systems and plant biotechnology. *Biotechnol. Adv.* **27**, 733–743 (2009).

62. Massie, B. *et al.* Inducible overexpression of a toxic protein by an adenovirus vector with a tetracycline-regulatable expression cassette. *J. Virol.* **72**, 2289–2296 (1998).

63. Schena, M., Lloyd, A. M. & Davis, R. W. A steroid-inducible gene expression system for plant cells. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 10421–10425 (1991).

64. Hinchey, B. S., Bensen, R. J. & Nelson, D. E. Water-Deficit inducible plant promoters. **1**, 12–17 (2004).

65. Hinchey, B. S. Stress-inducible plant promoters. **1**, (2005).

66. Luc Adam, Hayward, C., Lynne Reuber, San Mateo, C. & Karen S. Century, Albany, C. Disease-inducible promoters. **2**, (2011).

67. Hidenori Ozawa. Nitrate inducible promoter. **2**, (2007).

68. Karimi;, M., Barthels, N. & Godelieve Gheysen. Pathogen-inducible plant promoters. **1**, (2001).

69. Medrano, L., Feldmann, K. A., Alexandrov, N., Fang, Y. & Nester Apuya. Drought inducible promoters and uses thereof. **2**, (2008).

70. Mine;, T., Ohyama;, A., Hiyoshi;, T. & Keisuke Kasaoka. Cold-inducible promoter sequences. **12**, 2–4 (1943).

71. Kropat, J. & Beck, C. F. Characterization of Photoreceptor and Signaling Pathway for Light Induction of the *Chlamydomonas* Heat-Shock Gene HSP70A. *Photochem. Photobiol.* **68**, 414–419 (1998).

72. Schroda, M., Blöcker, D. & Beck, C. F. The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. *Plant J.* **21**, 121–131 (2000).

73. Lodha, M., Schulz-Raffelt, M. & Schroda, M. A New Assay for Promoter Analysis in *Chlamydomonas* Reveals Roles for Heat Shock Elements and the TATA Box in HSP70A Promoter-Mediated Activation of Transgene Expression . *Eukaryot. Cell* **7**, 172–176 (2008).

74. Poulsen, N. & Kröger, N. A new molecular tool for transgenic diatoms: Control of

mRNA and protein biosynthesis by an inducible promoter-terminator cassette. *FEBS J.* **272**, 3413–3423 (2005).

75. Miyagawa, A. *et al.* Research note: High efficiency transformation of the diatom Phaeodactylum tricornutum with a promoter from the diatom *Cylindrotheca fusiformis*. *Phycol. Res.* **57**, 142–146 (2009).

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

77. Park, S., Lee, Y., Lee, J. H. & Jin, E. S. Expression of the high light-inducible Dunaliella LIP promoter in *Chlamydomonas reinhardtii*. *Planta* **238**, 1147–1156 (2013).

78. Jackson, H. O. *et al.* An inducible expression system in the alga *Nannochloropsis gaditana* controlled by the nitrate reductase promoter. *J. Appl. Phycol.* **31**, 269–279 (2018).

79. Hempel, F., Bullmann, L., Lau, J., Zauner, S. & Maier, U. G. ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. *Mol. Biol. Evol.* **26**, 1781–1790 (2009).

80. Zaslavskaia, L. a, Lippmeier, J. C., Kroth, P. G., Grossman, A. R. & Apt, K. E. Transformation Of the Diatom *Phaeodactylum tricornutum* With a Variety of Selectable Marker and Reporter Genes. *J. Phycol.* **386**, 379–386 (2000).

81. Allen, A. E., Ward, B. B. & Song, B. Characterization of diatom (Bacillariophyceae) nitrate reductase genes and their detection in marine phytoplankton communities. *J. Phycol.* **41**, 95–104 (2005).

82. Kaye, Y. *et al.* Metabolic engineering toward enhanced LC-PUFA biosynthesis in Nannochloropsis oceanica: Overexpression of endogenous δ12 desaturase driven by stress-inducible promoter leads to enhanced deposition of polyunsaturated fatty acids in TAG. *Algal Res.* **11**, 387–398 (2015).

83. Shrestha, R. P. & Hildebrand, M. Development of a silicon limitation inducible expression system for recombinant protein production in the centric diatoms *Thalassiosira pseudonana* and *Cyclotella cryptica*. *Microb. Cell Fact.* **16**, 1–14 (2017).

84. Harada, H. *et al.* Regulation of the Expression of Intracellular β-Carbonic Anhydrase in Response to CO_2 and Light in the Marine Diatom *Phaeodactylum tricornutum* Published by : American Society of Plant Biologists (ASPB) Linked references are available on JSTOR for this a. *Plant Physiol.* **139**, 1041–1050 (2005).

85. Chai, Y., Lv, Y., Wang, T., Hou, W. & Xue, L. Heterologous gene expression driven by carbonic anhydrase gene promoter in *Dunaliella salina*. *ASAE Annu. Int. Meet. 2004* **24**, 7563–7573 (2004).

86. Iwai, M., Ikeda, K., Shimojima, M. & Ohta, H. Enhancement of extraplastidic oil

synthesis in *Chlamydomonas reinhardtii* using a type-2 diacylglycerol acyltransferase with a phosphorus starvation-inducible promoter. *Plant Biotechnol. J.* **12**, 808–819 (2014).

87. Iwai, M., Hori, K., Sasaki-Sekimoto, Y., Shimojima, M. & Ohta, H. Manipulation of oil synthesis in *Nannochloropsis* strain NIES-2145 with a phosphorus starvation-inducible promoter from *Chlamydomonas reinhardtii*. *Front. Microbiol.* **6**, 1–15 (2015).

88. Lin, H.-Y. *et al.* Alkaline phosphatase promoter as an efficient driving element for exogenic recombinant in the marine diatom *Phaeodactylum tricornutum*. *Algal Res.* **23**, 58–65 (2017).

89. Dong, H.-P. *et al.* Responses of *Nannochloropsis oceanica* IMET1 to Long-Term Nitrogen Starvation and Recovery. *Plant Physiol.* **162**, 1110–26 (2013).

90. Simionato, D. *et al.* 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 (2013).

91. Sforza, E., Calvaruso, C., La Rocca, N. & Bertucco, A. Luxury uptake of phosphorus in *Nannochloropsis salina*: Effect of P concentration and light on P uptake in batch and continuous cultures. *Biochem. Eng. J.* **134**, 69–79 (2018).

92. Mühlroth, A. *et al.* Mechanisms of phosphorus acquisition and lipid class remodelling under P limitation in a marine microalga. *Plant Physiol.* **175**, pp.00621.2017 (2017).

93. Guillard, R. R. L. Culture of Phytoplankton for Feeding Marine Invertebrates. in *Culture of Marine Invertebrate Animals: Proceedings -1st Conference on Culture of Marine Invertebrate Animals Greenport* (eds. Smith, W. L. & Chanley, M. H.) 29–60 (Springer US, 1975).

94. Berges, J. A., Franklin, D. J. & Harrison, P. J. Evolution of an artificial seawater medium: Improvements in enriched seawater, artificial water over the last two decades. *J. Phycol.* **37**, 1138–1145 (2001).

95. Stano, M., Beke, G. & Klucar, L. viruSITE—integrated database for viral genomics. *Database* **2016**, (2016).

96. Lescot, M. *et al.* PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **30**, 325–327 (2002).

97. Solovyev, V. V, Shahmuradov, I. A. & Salamov, A. A. Identification of Promoter Regions and Regulatory Sites. in *Computational Biology of Transcription Factor Binding* (ed. Ladunga, I.) 57–83 (Humana Press, 2010).

98. Oliveira, G. *et al.* Promoter Motifs in NCLDVs: An Evolutionary Perspective. *Viruses* **9**, 16 (2017).

99. Kang, M. *et al.* Genetic diversity in chlorella viruses flanking kcv, a gene that encodes a potassium ion channel protein. *Virology* **326**, 150–159 (2004).

100. Fitzgerald, L. A. *et al.* Putative gene promoter sequences in the *Chlorella* viruses. *Virology* **380**, 388–393 (2008).

101. Kawasaki, T., Tanaka, M., Fujie, M., Usami, S. & Yamada, T. Immediate early genes expressed in chlorovirus infections. *Virology* **318**, 214–223 (2004).

102. Giorgio Valle group. *Nannochloropsis gaditana* Genomes portal - CRIBI, University of Padua. (2013). Available at: http://www.nannochloropsis.org/index.php.

103. Karas, B. J. *et al.* Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* **6**, (2015).

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

105. Cao, S. *et al.* Evaluation of putative internal reference genes for gene expression normalization in *Nannochloropsis* sp. by quantitative real-time RT-PCR. *Biochem. Biophys. Res. Commun.* **424**, 118–123 (2012).

106. Delaroque, N., Maier, L., Knippers, R. & Müller, D. G. Persistent virus integration into the genome of its algal host, *Ectocarpus siliculosus* (Phaeophyceae). *J. Gen. Virol.* **80**, 1367–1370 (1999).

107. Shahmuradov, I. A., Umarov, R. K. & Solovyev, V. V. TSSPlant: A new tool for prediction of plant Pol II promoters. *Nucleic Acids Res.* **45**, (2017).

108. Edgar, R. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).

109. Barrett, T. *et al.* NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Res.* **41**, 991–995 (2013).

110. Lin, H. Y. *et al.* Identification and Characterization of an Extracellular Alkaline Phosphatase in the Marine Diatom *Phaeodactylum tricornutum*. *Mar. Biotechnol.* **15**, 425–436 (2013).

111. Sadka, A., DeWald, D. B., May, G. D., Park, W. D. & Mullet, J. E. Phosphate modulates transcription of soybean VspB and other sugar-inducible genes. *Plant Cell* **6**, 737–749 (1994).

112. Malboobi, M. A. & Lefebvre, D. D. A phosphate-starvation inducible β-glucosidase gene (psr3.2) isolated from *Arabidopsis thaliana* is a member of a distinct subfamily of the BGA family. *Plant Mol. Biol.* **34**, 57–68 (1997).

113. Lefebvre, D. D. & Malboobi, M. A. Phosphate deficiency inducible promoter. **2**, (2001).

114. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. The ovalbumin gene - sequence of putative control regions. *Nucleic Acids Res.* **8**, 127–142 (1980).

115. McKnight, S. L. & Kingsbury, R. Transcriptional control signals of a eukaryotic protein-coding gene. *Science (80-.).* **217**, 316–324 (1982).

116. Menkens, A. E., Schindler, U. & Cashmore, A. R. The G-box: a ubiquitous regulatory DNA element in plants bound by the GBF family of bZIP proteins. *Trends Biochem. Sci.* **20**, 506–510 (1995).

117. Yamaguchi-shinozaki, K. & Shinozaki, K. Organization of cis -acting regulatory elements in osmotic- and cold-stress- responsive promoters. *Trends Plant Sci.* **10**, 88–94 (2005).

118. Rouster, J., Leah, R., Mundy, J. & Cameron-Mills, V. Identification of a methyl jasmonate-responsive region in the promoter of a lipoxygenase 1 gene expressed in barley grain. *Plant J.* **11**, 513–523 (1997).

119. Abe, H. *et al.* Role of *Arabidopsis* MYC and MYB Homologs in Drought- and Abscisic Acid-Regulated Gene Expression. *Plant Cell* **9**, 1859 (2007).

120. Ambawat, S., Sharma, P., Yadav, N. R. & Yadav, R. C. MYB transcription factor genes as regulators for plant responses: An overview. *Physiol. Mol. Biol. Plants* **19**, 307–321 (2013).

121. Shen Qingxi & Ho Tuan-Hua, D. Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA-responsive complexes each containing a G-box and a novel cis-acting element. *Plant Cell* **7**, 295–307 (1995).

122. Hobo, T., Asada, M., Kowyama, Y. & Hattori, T. ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. *Plant J.* **19**, 679–689 (1999).

123. Moskvina, E., Imre, E. M. & Ruis, H. Stress factors acting at the level of the plasma membrane induce transcription via the stress response element (STRE) of the yeast *Saccharomyces cerevisiae*. *Mol. Microbiol.* **32**, 1263–1272 (1999).

124. Hu, J. *et al.* Genome-wide identification of transcription factors and transcription-factor binding sites in oleaginous microalgae *Nannochloropsis*. *Sci. Rep.* **4**, 5454 (2014).

125. Feng, Y.-Q., Lorincz, M. C., Fiering, S., Greally, J. M. & Bouhassira, E. E. Position Effects Are Influenced by the Orientation of a Transgene with Respect to Flanking Chromatin. *Mol. Cell. Biol.* **21**, 298–309 (2001).

126. Clark, A. J. *et al.* Chromosomal position effects and the modulation of transgene expression. *Reprod. Fertil. Dev.* **6**, 589–594 (1994).

127. Peach, C. & Velten, J. Transgene expression variability (position effect) of CAT and GUS reporter genes driven by linked divergent T-DNA promoters. *Plant Mol. Biol.* **17**, 49–60 (1991).

128. Hobbs, S. L. A., Warkentin, T. D. & DeLong, C. M. O. Transgene copy number can be positively or negatively associated with transgene expression. *Plant Mol. Biol.* **21**, 17–26 (1993).

129. Maqbool, S. B. & Christou, P. Multiple traits of agronomic importance in transgenic indica rice plants: analysis of transgene integration pattern, expression levels and stability. *Mol. Breed.* **5**, 471–480 (1999).

130. Al-Shawi, R., Kinnaird, J., Burke, J. & Bishop, J. O. Expression of a foreign gene in a line of transgenic mice is modulated by a chromosomal position effect. *Mol. Cell. Biol.* **10**, 1192–1198 (2015).

131. Takano, M., Egawa, H., Ikeda, J. E. & Wakasa, K. The structures of integration sites in transgenic rice. *Plant J.* **11**, 353–361 (1997).

132. Assaad, F. F., Tucker, K. L. & Signer, E. R. Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis*. *Plant Mol. Biol.* **22**, 1067–1085 (1993).

133. Cerutti H, Johnson, A., Gillham, N. & Boynton, J. Epigenetic Silencing of a Foreign Gene in Nuclear Transformants of *Chlamydomonas*. *Plant Cell* **9**, 925–945 (2002).

134. Davis, A., Crum, L. T., Corbeil, L. B. & Hildebrand, M. Expression of Histophilus somni IbpA DR2 protective antigen in the diatom *Thalassiosira pseudonana*. *Appl. Microbiol. Biotechnol.* **101**, 5313–5324 (2017).

135. George J, Kahlke T, Abbriano RM, Kuzhiumparambil U, Ralph PJ, Fabris M. Metabolic Engineering Strategies in Diatoms Reveal Unique Phenotypes and Genetic Configurations with Implications for Algal Genetics and Synthetic Biology. Front Bioeng Biotechnol. **8**:513. (2020).

136. Martin, D. I. K. & Whitelaw, E. The vagaries of variegating transgenes. *BioEssays* **18**, 919–923 (1996).

137. Laney, S. R., Olson, R. J. & Sosik, H. M. Diatoms favor their younger daughters. *Limnol. Oceanogr.* **57**, 1572–1578 (2012).

138. Aguilaniu, H., Gustafsson, L., Rigoulet, M. & Nyström, T. Asymmetric inheritance of oxidatively damaged proteins during cytokinesis. *Science.* **299**, 1751–1753 (2003).

139. Traller, J. C. & Hildebrand, M. High throughput imaging to the diatom *Cyclotella cryptica* demonstrates substantial cell-to-cell variability in the rate and extent of triacylglycerol accumulation. *Algal Res.* **2**, 244–252 (2013).

140. Dmytruk, K. V., Voronovsky, A. Y. & Sibirny, A. A. Insertion mutagenesis of the yeast *Candida famata* (*Debaryomyces hansenii*) by random integration of linear DNA fragments. *Curr. Genet.* **50**, 183–191 (2006).

141. Thomsen, D. R., Stenberg, R. M., Goins, W. F. & Stinski, M. F. Promoter-regulatory region of the major immediate early gene of human cytomegalovirus.

Proc. Natl. Acad. Sci. **81**, 659–663 (1984).

142. Schmidt, E. V, Christoph, G., Zeller, R. & Leder, P. The cytomegalovirus enhancer: a pan-active control element in transgenic mice. *Mol. Cell. Biol.* **10**, 4406–4411 (1990).

143. Chapman, B. S., Thayer, R. M., Vincent, K. A. & Haigwood, N. L. Effect of intron A from human cytomegalovirus (Towne) immediate-early gene on heterologous expression in mammalian cells. *Nucleic Acids Res.* **19**, 3979–3986 (1991).

144. Van Leeuwen, W., Ruttink, T., Borst-Vrenssen, A. W. M., Van Der Plas, L. H. W. & Van Der Krol, A. R. Characterization of position-induced spatial and temporal regulation of transgene promoter activity in plants. *J. Exp. Bot.* **52**, 949–959 (2001).

145. Meyer, P. Understanding and controlling transgene expression. *Trends Biotechnol.* **13**, 332–337 (1995).

146. Kadono, T. *et al.* Characterization of marine diatom-infecting virus promoters in the model diatom *Phaeodactylum tricornutum. Sci. Rep.* **5**, 1–13 (2015).

147. Adler-Agnon (Shemesh), Z., Leu, S., Zarka, A., Boussiba, S. & Khozin-Goldberg, I. Novel promoters for constitutive and inducible expression of transgenes in the diatom *Phaeodactylum tricornutum* under varied nitrate availability. *J. Appl. Phycol.* **30**, 2763–2772 (2018).

CHAPTER 4

Genetic engineering of the microalga *Nannochloropsis gaditana* for the production of sesquiterpene β-caryophyllene

Margaret Ramarajan[a], Michele Fabris[a,b], Unnikrishnan Kuzhiumparambil[a] , Peter J. Ralph[a]

[a] University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo NSW 2007, Australia

[b] CSIRO Synthetic Biology Future Science Platform, GPO box 2583, Brisbane QLD 4001, Australia

Contributors: M.R. performed the experiments; M.R, M.F. and P.J.R. conceived the research project; U.K. performed GC-MS; M.F, and U.K. contributed in data analysis; M.R. wrote the manuscript; all authors edited and reviewed the manuscript.

Manuscript: In-prep

Genetic engineering the microalga *Nannochloropsis gaditana* for the production of the sesquiterpene β-caryophyllene

Margaret Ramarajan[a], Michele Fabris[a,b], Unnikrishnan Kuzhiumparambil[a] , Peter J. Ralph[a]

[a] University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo NSW 2007, Australia

[b] CSIRO Synthetic Biology Future Science Platform, GPO box 2583, Brisbane QLD 4001, Australia

Abstract

The plant secondary metabolite β-caryophyllene is a bicyclic sesquiterpene traditionally used in cosmetics and perfumes. It is a phytocannabinoid and can function as a cannabinoid receptor type 2 (CB2) agonist and might offer a potential therapeutic strategy for the treatment of inflammation and pain. β-caryophyllene is also an important sesquiterpene found in essential oils of hops (beer brewing) and recently been studied as a precursor for high-density biofuels. Currently, β-caryophyllene is purified from plant-based oleoresins; however, increasing demand has led to significant interest in genetic engineering for industrial-scale production as plant-based extraction requires large biomass which is not environmentally sustainable. To date, *Escherichia coli* and the cyanobacterium *Synechocystis* have been engineered to produce this metabolite but these work horses have not evolved to synthesize complex terpenoids like photosynthetic organisms.

As photosynthetic lifestyles of eukaryotic microalgae require effective pathways for generating isoprenoid pigments and electron carriers, both of which play important roles in photosynthesis. Therefore, the metabolic flux for isoprenoid biosynthesis might be higher in photoautotrophic microalgae than the heterologous fermentative microbial hosts such as bacteria and yeasts. The interest in photosynthetic organisms is fresh, as they have a number of benefits over heterotrophic hosts such as a chloroplast, photosynthesis derived supply of reducing equivalents, relatively inexpensive growing conditions and use CO_2 as the sole carbon source aiding mitigation of greenhouse gas emission while yielding valuable biomass and potentially a suite of co-products such as lipids, carbohydrates, proteins, and pigments. In this study, we explored the capacity of the marine microalga *N. gaditana* as a novel terpenoid production platform. The β-caryophyllene synthase gene (*QHS1*) from *Artemisia annua* was introduced as a fusion gene along with a reporter (*mVenus*) under the control of endogenous *Hsp90* promoter. Under the laboratory growing conditions we were able to extract 20 ng. g of dry cell weight^{-1} of β-caryophyllene from cytoplasmic expression of the QHS1-mVenus fusion in *N. gaditana*. This research has expanded our knowledge on the unmapped *Nannochloropsis* terpenoid metabolism. This work represents the first expression of a heterologous gene for terpenoid production in *N. gaditana* with the prospect for more

intricate pathway engineering directed at the heterologous production of sesquiterpenes.

4.1 Introduction

Plants produce a chemically diverse array of secondary metabolites which are associated with developmental physiology, mutualistic and antagonistic plant–herbivore as well as plant–environment interactions[1–3]. Terpenoids form a major group of these plant secondary metabolites and are divided into mono-, sesqui- and di-terpenes (C10, C15 and C20, respectively), based on the number of isoprene units (C_5H_8) used in its synthesis[2]. The universal building blocks of all terpenoids are isopentenyl pyrophosphate (IPP), and dimethyl allyl pyrophosphate (DMAPP)[4]. In nature, sesquiterpenes occur as hydrocarbons and as oxygenated forms (acids, alcohols, aldehydes, lactones and ketones) which are made up of three isoprene units[5]. Several individual cyclases catalyse the conversion of farnesyl pyrophosphate (FPP) to a variety of sesquiterpenes[6]. The sesquiterpenoid pathway and types of sesquiterpenes are shown in Figure S15[7].

β-caryophyllene is a bicyclic sesquiterpene compound found in the essential oils of many higher plants[8]. It has a woody and spicy odour[9] and has been used as fragrance and flavouring agent since 1930s[8,10,11]. Studies have found that β-caryophyllene has local anaesthetic activity[12], neuroprotective activity[13], anti-inflammatory[14–16], anti-microbial[17], was found to have insulinotropic effect on streptozotocin-induced hyperglycemic rats[18]. It also has a potential for treating or preventing hepatic injury associated with oxidative stress, inflammation and steatosis[19] and can function as a cannabinoid receptor type 2 (CB2) agonist[20]. Compounds targeting CB2 might offer a potential therapeutic strategy for the treatment of inflammation, pain, atherosclerosis, and osteoporosis[20]. β-caryophyllene is an important sesquiterpene which along with its major oxidation product caryophyllene oxide is found in essential oils of hops which is used in beer brewing[21,22]. Recently, β-caryophyllene has also been studied as a precursor of high density biofuel[23,24].

Plants are the major producers of β-caryophyllene[11,25–29]. This sesquiterpene is induced as a defense molecule in response to herbivory[30] (for example, in maize roots

β-caryophyllene, was emitted when attacked by western corn rootworm- *Diabrotica virgifera*[30]), bacterial pathogen[27](for example, in *Arabidopsis* this sesquiterpene is shown to have defensive activity against infection by *Pseudomonas syringae*[27]*)* or insect attack[25,26]. β-caryophyllene is one of the biologically active compounds in clove essential oil from clove- *Eugenia caryophyllata* (13%)[29,31,32]. Chemical synthesis of this compound is very complex[32,33] hence biological synthesis is preferred. This volatile sesquiterpene has been over-expressed in higher plants using homologous and heterologous gene expression[34], in *Escherichia coli*[24,35] and cyanobacterium *Synechocystis* sp. strain PCC6803[36]. Plants are not an ideal source for large-scale terpenoid synthesis and extraction due to low amounts of terpenoids produced and the need for large biomass which is not sustainable[37]. Often the plant terpenoids are produced as mixtures which is extremely difficult and expensive to extract and seperate[37]. No reports are available on heterologous gene expression in yeast (*Saccharomyces cerevisiae*).

Synthetic biology has paved the way for efficient and scalable production of terpenoids in microbial hosts[38–40]. Heavily engineered strains of *E. coli* have been used for the production of β-caryophyllene under aerobic fed-batch fermentation yielding 0.34 mg h^{-1} g^{-1} dry cells[35] and 1.15 mg h^{-1} g^{-1} dry cells using acetic acid as substrate[41] while Wu et al., 2018 demonstrated that their engineered strains could produce 100 mg/L β-caryophyllene using marine microalgal hydrolysate as the sole carbon source[24]. But these industrial workhorses have not evolved to synthesise complex terpenoids like photosynthetic organisms[37]. For example, numerous cytochrome P450 enzymes are needed for the biosynthesis of many plant-derived terpenoids and these enzymes the need for redox power in the form of NADPH, cofactor availability, are poorly expressed in heterologous hosts like bacteria[42]. Hence in some cases dual-organism production systems are used, For example, the taxadiene produced using an engineered *E. coli* strain was oxygenated by a yeast expressing taxadiene 5α-hydroxylase, which is a cytochrome P450 and the P450 reductase[43]. P450 driven metabolic engineering has some challenges such as protein stability and activity, expensive co-factors, toxicity of substrate/intermediates to host, lack of specificity/efficiency and metabolic flux control that have been addressed in a variety of yeast model systems[42].

As photosynthetic lifestyles of eukaryotic microalgae require effective pathways for generating isoprenoid pigments and electron carriers, both of which play important roles in photosynthesis[44]. Therefore, the metabolic flux for isoprenoid biosynthesis might be higher in photoautotrophic microalgae than the heterologous fermentative microbial hosts such as bacteria and yeasts[45]. The recent exploratory works of a few research groups have highlighted the significance of microalgae for heterologous terpenoid production[46–49]. Photosynthetic organisms have a number of advantages over heterotrophic hosts such as a chloroplast, photosynthesis powered supply of reducing equivalents[37] i.e. carbon capture and conversion using carbon concentrating mechanisms (CCMs) to generate reducing equivalent energy by photosynthesis and reactions such as the Calvin-Benson cycle, microalgae are able to perform and grow utilising CO_2 from the environment[45]. Microalgae have relatively inexpensive growing conditions and use CO_2 as the sole carbon source aiding mitigation of CO_2 emission while producing valuable biomass and a suite of co-products such as lipids, carbohydrates, proteins, vitamins and pigments[50].

As stated above, the precursors (IPP and DMAPP) are essential for any terpenoid biosynthesis. The inheritance of genes involved in IPP biosynthesis in eukaryotes via the mevalonate (MVA) pathway is of archaeal/eukaryotic origin, whilst the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway is of eubacteria[51,52]. Specifically, the prokaryotic MEP pathway in the eukaryotes with plastids has been acquired via cyanobacterial endosymbiosis and the genes subsequently transferred to the nuclear genome[4,51,52]. Several stramenopiles such as diatoms *P. tricornutum* and *Nitzschia ovalis*[53,54]and Chrysophyceae *Ochromonas danica*[44] contain both the MEP pathway and the MVA pathway 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway for terpenoid biosynthesis, but the Eustigmatophyceae *Nannochlorpsis* spp. have acquired the MEP pathway (chloroplastic) but have lost the MVA pathway[4,55,56] similar to Chlorophytes such as *Scenedesmus obliquus*, *Chlamydomonas reinhardtii* and *Chlorella fusca*[4,57,58] where the MEP pathway is solely responsible for providing isoprenoids for the cytosol and chloroplast[4,53,59]. Also, similar to diatoms such as *P. tricornutum*[60], the outer membrane of the chloroplast is continuous with the endoplasmic reticulum and nuclear envelope[61–63].

Transcriptomic and proteomic responses to very low CO_2 suggest multiple carbon concentrating mechnisms in *Nannochloropsis* distinct from those of *C. reinhardtii* and *P. tricornutum* with an upregulation of a number of key C4-like pathway enzymes suggesting a C4-based CCM[55,64,65] and the fixed inorganic carbon is channelled through glycolysis and the Calvin cycle to form pyruvate and glyceraldehyde-3-phsophate (MEP pathway substrates)[64–67] . Further, *Nannochloropsis* has the natural capacity to produce precursor molecules for terpene based pigments such as chlorophyll a, carotenoids[68] and sterols such as cholesterol and phytosterols[56].

However, these microalgae remain unexplored for heterologous terpenoid biosynthesis. In this study, the available *N. gaditana* genomic data was used to computationally map the enzymes of the MEP pathway and their sub-cellular localisations *in-silico* to inform a suitable strategy for production of β-caryophyllene, a cannabinoid sesquiterpene[8,20,69–71]. With a broader vision to expand to more complex terpene engineering we succeeded in the production of β-caryophyllene as the first proof-of-concept terpenoid engineering in this species using a β-caryophyllene synthase (*QHS1*)[72] from *Artemisia annua L* (sweet wormwood) driven by the endogenous Hsp90 promoter characterised in Chapter 2. This study demonstrates the first expression of a heterologous gene for terpenoid biosynthesis in this microalga.

4.2 Materials and methods

4.2.1 Strain and growth conditions

N. gaditana strain CCMP526 was purchased from the Provasoli-Guillard National Centre for Culture of Marine Phytoplankton, USA. The strain was maintained on modified F/2 medium [73] where the seawater component was replaced with Enriched Seawater, Artificial Water (ESAW) [74] with 1% agar. Standard liquid cultures were grown in modified F/2 media as well. Both the liquid cultures and plates were grown under continuous light of 100 µmol photons m^{-2} s^{-1} from a LED light panel at 21°C. Liquid cultures were aerated in a shaker incubator set at 90 rpm maintained in 250 mL Erlenmeyer flasks closed with foam plugs to allow gas exchange and were maintained in batch mode. DH5-alpha *Escherichia coli* electro-competent cells (New England Biolabs, USA) were used for DNA cloning, and were grown in Luria broth (Millers LB base) medium (Invitrogen, Life Technologies, USA) with 100 µg mL^{-1} ampicillin at 37 °C.

4.2.2 Bioinformatics analyses

The *Arabidopsis thaliana* genes coding for the MEP pathway enzymes were obtained from TAIR[75,76]. Their respective protein sequences were extracted as fasta files from NCBI and *N. gaditana* orthologs were identified using blastp (protein blast) in *N. gaditana* genome browser (Table S17) and NCBI (Table S18). Enzyme putative localisation was inferred using DeepLoc[77], TargetP[78], SignalP v3.1[79], and ASAFind[79,80] (Table 7).

Also the TMHMM Server v. 2.0[81] was used to predict the transmembrane helices in protein query sequences. Based on this bioinformatics analyses, a hypothetical model of isoprenoid biosynthesis via MEP pathway in *N. gaditana* was constructed including the predicted sub-cellular localisation of all the pathway enzymes (Figure 15).

4.2.3 Construction of pNaga4.mV.DC.N2.QHS1 vector

The pNaga4.mV.DC.N2 vector (see Chapter 3 Materials and methods section 3.2.4) was linearised using *BamHI* and the entire coding region of the β-*caryophyllene synthase* gene (synthesised as a gBlocks® Gene Fragment, IDT, USA) from *Artemisia annua* (QHS1, GenBank No. AF472361.1) [72] was introduced in-frame before *mVenus* reporter gene using the Gibson Assembly® Master Mix (New England Biolabs, USA) and gene-specific primers (Table S19). The pNaga4.mV.DC.N2.QHS1 (Figure S16) vector sequence was confirmed by DNA sequencing (Macrogen, South Korea).

4.2.4 *E. coli* transformation

The DH5-alpha *Escherichia coli* (50 µL) electro-competent cells (New England Biolabs, USA) were mixed with 2 µL of Gibson ligated pNaga4.mV.DC.N2 and pNaga4.mV.DC.N2.QHS1 plasmids, individually and electroporated at 3000V, with a resistance of 200 Ω, a capacitance of 25 µF in a 0.2 cm cuvette using a Gene Pulser Xcell™ Electroporation System (Biorad, USA). The ampicillin-resistant *E. coli* colonies were screened by colony PCR using GoTaq Flexi polymerase (Promega, USA) and gene or region-specific primers (Table S19). Plasmid extraction was done using the Zippy Plasmid Miniprep Kit (Zymo Research, USA). The vector sequences were confirmed by DNA sequencing (Macrogen, South Korea).

4.2.5 *N. gaditana* transformation and transformant screening

The *N. gaditana* strain CCMP526 was transformed with the pNaga4.mV.DC.N2 and pNaga4.mV.DC.N2.QHS1 plasmids using electroporation (see Chapter 2 Materials and methods section 2.5). Zeocin-resistant colonies were screened by flow cytometry using a flow cytometer (CytoFLEX S, Beckman Coulter Life sciences, USA) to check for mVenus reporter fluorescence using a 488 nM laser and 525/40 optical filter.

4.2.6 β-caryophyllene extraction and GC-MS analysis

β-caryophyllene is a volatile sesquiterpene, hence cultivation of *N. gaditana* in the presence of 3.2% isopropyl myristate (Sigma-Aldrich, USA) was tested. 50 mL of *N. gaditana* wild type cultures were grown under above growth condition in a modified F/2 media containing 32 µL/mL of molecular grade isopropyl myristate (Sigma-Aldrich, USA) in triplicate. This ester was used to trap the volatile terpenes from the culture.

For *E. coli*, 500 mL of bacterial cultures carrying the pNaga4.mV.DC.N2 and pNaga4.mV.DC.N2.QHS1 plasmids were pelleted independently at 6000 g for 5 mins. For *N. gaditana,* the wild type and transgenic lines expressing QHS1 as well as the mVenus control cultures were cultivated in bi-phasic mode. The cultures were harvested at stationary phase and the culture pellet and isopropyl myristate layer collected for subsequent analysis.

The bacterial and microalgal pellets were lyophilized using a freeze drier (Alpha 2-4LDplus, Martin Christ Gefriertrocknungsanlagen GmbH, Germany) under a vacuum pressure of 0.10 mbar at -84 °C. Dried biomass pellets were extracted using dichloromethane (DCM). 500 µL of solvent was added to biomass and homogenised using a bead beater, centrifuged and supernatant was collected. Extraction was repeated twice with 250 µL DCM and supernatant were pooled. Pooled supernatants were analysed using GC-MS as per the method detailed below. Limonene was used as internal standard. Simultaneously, the isopropyl myristate layer was diluted with ethyl acetate at 1:3 ratio, limonene added as internal standard and analysed using GC-MS.

Samples were run on a GC-MS (QP2020, Shimadzu Corporation, Kyoto Japan) equipped with an AOC-20i/s auto sampler (Shimadzu Corporation). The column used was an SH-Rxi-5Sil MS fused silica capillary column (30.0 m x 0.25 mm x 0.25 µm)

operating in electron impact mode at 70 eV. Helium was used as the carrier gas at a constant flow of 1.0 mL min^{-1} and an injection volume of 1 µL, with an injector temperature of 280 °C and an ion source temperature of 230 °C. The temperature gradient of the oven was 70 °C for 1 min, then 7 °C per minute to 220 °C, hold for 5 mins and temperature ramped to 325 °C at 20 °C per minute.

Quantitative analysis was run in Selective Ion Monitoring (SIM) mode, where selected ions were monitored for 20 ms each. The peak areas were converted into concentrations in comparison with calibration curves plotted against a range known concentrations of standards.

4.2.7 *N. gaditana* chlorophyll α fluorescence measurement

Pulse-amplitude modulated fluorometry (PAM) was used to measure the maximum quantum yield of photosystem II (Fv/Fm calculated as Fv = Fm − Fo, where Fo is the minimum fluorescence in the dark and Fm the maximum fluorescence) as a proxy for photosynthetic health. The measurement is a quick assessment of transgenic *N. gaditana* photosynthetic activity to determine photosystem stress due to transgene introduction via electroporation and photosystem damage due to random gene integration when compared to *N. gaditana* wild type and empty vector control. A Pocket-PAM (Gademann Instruments GmbH, Germany) was used to take the measurements after the cultures were acclimated for 20 mins in darkness at room temperature. The settings used are blue light, saturation pulse intensity – 12 (>5000 µmol photons m−2 s−1), saturation pulse width of 0.6 s with gain.

4.2.8 Subcellular localisation of mVenus-QHS1 fusion protein in transgenic *N. gaditana*

The wild type and transgenic *N. gaditana* (3 mL) expressing the pNaga4.mV.DC.N2.QHS1 or pNaga4.mV.DC.N2 constructs were cultivated for 3 days in modified F/2 medium. Microscopic slides were prepared with 10 µL of the culture and imaged using a confocal laser scanning microscope (Nikon A1 Plus, Nikon Instruments Inc, Japan) under a 100X objective with oil immersion. mVenus and chlorophyll autofluorescence were measured using 488 nm and 637 nm lasers.

4.2.9 Statistical analysis

Statistical analyses were done using GraphPad Prism 8.0. 2-way ANOVA parametric test was applied, and significant effects were analysed using Dunnett's multiple comparison test for the cell count, Fv/Fm data and for mVenus fluorescence intensity measurements for all *N. gaditana* transgenic lines including controls. The analyses tested the null hypothesis that there was no difference in growth, chlorophyll α fluorescence in all the lines tested, and all transgenic lines showed no difference in mVenus fluorescence in wild type and transgenic *N. gaditana* lines expressing QHS1. Significant differences between control and QHS1 expressing *E. coli* was calculated using paired t-test. Significant differences between *N. gaditana* wild type and the transgenic lines carrying QHS1 were calculated using 2-way ANOVA Dunnett's multiple comparison test.

4.3 Results and discussion

4.3.1 Computational analyses indicate to no natural sesquiterpene production in *N. gaditana*

N. gaditana uses the MEP pathway for terpenoid synthesis and has the innate capacity for the production of pigments[68], sterols[56], phytol[63] and fatty acids[63,82,83]. Synthesis of other C5, C10 and C15 isoprenoids have not been reported in this microalga. To evaluate the endogenous terpenoid pathway that would supply precursors for heterologous terpenoid production, computational tools were used to reconstruct the metabolic network for terpenoid biosynthesis in *N. gaditana*. Using *A. thaliana* MEP pathway enzymes as reference, orthologs were searched in the *N. gaditana* genome using NCBI blastp[84] and identified the complete set of *in silico* genes encoding MEP pathway enzymes including prenyl phosphate biosynthesis (Table 7). The putative MEP pathway enzymes in *N. gaditana* were analysed using the subcellular localisation prediction tools DeppLoc[77] and ASAFind[79,80] (Table 7). The putative sub-cellular localisation of these enzymes was essential to putatively identify the location of diphosphates (GGPP, FPP and GPP) pool, specifically farnesyl diphosphate. ASAFind was used because it is specific for algae with secondary plastids of the red lineage to identify nuclear-encoded plastid proteins and looks for conserved "ASAFAP"-motifs and transit peptides based on a SignalP output[79,80]. DeepLoc-1.0 predicts the subcellular localisation

of eukaryotic proteins (10 different localisations: nucleus, cytoplasm, extracellular, mitochondrion, cell membrane, endoplasmic reticulum, chloroplast, Golgi apparatus, lysosome/vacuole and peroxisome) with a accuracy rate of 78% for 10 categories[77]. Based on DeepLoc and ASAFind predictions (Table 7) a hypothetical model for isoprenoid biosynthesis via MEP pathway and their enzyme sub-cellular localisation was constructed (Figure 15).

The first step of the MEP pathway is a condensation reaction of pyruvate and glyceraldehyde 3-phosphate (G3P) to form 1-deoxy-D-xylulose-5-phosphate (DXP) by DXP synthase (EWM22144.1, DXS) which has a predicted high likelihood (0.9994) to be localised to the chloroplast. DXP is later converted into 2-C-methyl-D-erythritol 4-phosphate (MEP) by DXP reductoisomerase (EWM24175.1, DXR), which has a high likelihood score (0.9781) and is predicted to be targeted to the mitochondria (DeepLoc prediction), but its *A. thaliana* counterpart has been experimentally confirmed to be located in the chloroplast[59,85,86]. As the MEP pathway is in the chloroplast, it is highly unlikely that DXR would be in the mitochondria, hence experimental confirmation is required. Next, MEP is converted to 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CDP-ME) using 2-c-methyl-d-erythritol 4-phosphate cytidylyltransferase (EWM22399.1, CSM) which has a predicted high likelihood (0.9934) to be localised to the chloroplast. CDP-ME is converted to CDP-ME-2-phosphate (CDP-MEP) using 4-diphosphocytidyl-2c-methyl-d-erythritol kinase (EWM29944.1, CDPMEK) which has a predicted moderate likelihood (0.5842), but high confidence level (ASAFind) to be localised to the chloroplast. Next, CDP-MEP is converted to 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MECDP) by 2-c-methyl-d-erythritol–cyclodiphosphate (EWM23179.1, ISPF) which has a predicted high likelihood (0.862) and high confidence level (ASAFind) to be localised to the chloroplast. MECDP is converted to 4-hydroxy-3-methylbut-2-enyl diphosphate (HMBDP) by 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (EWM25176.1, HDS) which has a predicted low likelihood score of 0.395 and is targeted to the cytosol, while its *A. thaliana* counterpart has been experimentally confirmed to be located in the chloroplast[59,87]. The final step to produce IPP and DMAPP (terpenoid precursors) from HMBDP is catalysed by 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (EWM26997.1, HDR) which has a predicted

high likelihood (0.7576) and high confidence level (ASAFind) to be localised to the chloroplast. Although HDR catalyses the above branching reaction, IPP and DMAPP are produced in a ratio, for example, in plants its around 85:15[88] and the plastidic IDI catalyses IPP isomerisation for substrate optimisation[88].

In *N. gaditana*, Radakovits et al., (2010) highlighted the presence of two types of IDI, one is a predicted bi-functional IDI/squalene synthase (IDI/SQS) without a chloroplast transit peptide sequence which suggests a cytosolic localisation, whereas the other independent IDI enzyme has a predicted chloroplast transit peptide[4,64,89,90]. In general, in heterokonts, the SQS enzyme is hypothesized to be associated with the membrane of the endoplasmic reticulum and SQS is proposed to physically interacts with farnesyl pyrophosphate synthase (FPPS)[4,91]. In photosynthetic heterokonts, the IDI/SQS fusion protein appears to be conserved[4]. EWM26314.1 has a putative squalene synthase domain and could be located in the chloroplast-ER continuum similar to *Haslea ostrearia* HoIDISQS[91] but has a Deeploc predicted moderate likelihood (0.558) to be mitochondrial which may not be plausible and the other putative IDI enzymes (EWM27595.1, EWM25927.1) are predicted to be plastidic (DeepLoc and ASAFind). In microalgae with only the MEP pathway such as *Nannochloropsis* spp. and *Chlamydomonas* spp., the terpenoid precursor molecules IPP and/or DMAPP might be transported across the chloroplast envelope to the cytosol for squalene/sterol biosynthesis[4,92], but the mechanism of transport whether active or passive is presently unknown.

Four putative prenyltransferases (EWM28775.1, GGPP; EWM28148.1, GPP; EWM21458.1, FPP; EWM26314.1, FPP) were predicted to be localised in the mitochondrion for the synthesis of isoprenoids such as polyprenols and ubiquinones[93] while two putative prenyltransferases (EWM28774.1, GGPP; EWM26176.1, GPP) were predicted to be localised in the chloroplast for the synthesis of precursors for chlorophyll, carotenoids, tocopherol and plastoquinones[93]. The various prenyl diphosphate substrates produced by the action of the above synthases require IPP and DMAPP which would necessitate the transport of IPP/DMAPP across the plastic membrane into the cytosol and then into the other organelles through a yet unknown mechanism.

The function of the putative prenyltransferases in *N. gaditana* has not been experimentally validated yet. Characterisation of these enzymes to understand their substrate specificity, enzyme kinetics and identify their sub-cellular localisation within *Nannochloropsis* would be a significant advantage in designing metabolic engineering strategies and targeting heterologous genes for terpenoid production. To date, no enzyme characterisation data is available for *N. gaditana* prenyltransferases. Studies involving the availability of the prenyl diphosphate substrate pool in the cytoplasm or the different organelles is lacking in this microalga and no targeting signal peptides have been characterised so far that would allow plastidic or ER or mitochondrial protein expression. The *N. gaditana* genome was mined for sesquiterpene synthases, but no homologues were found in *N. gaditana* genome for *A. thaliana* sesquiterpene synthases encoded by genes At5g23960 and At5g44630 which are responsible for the formation of all sesquiterpenes found in *A. thaliana* (Table S20). This suggests that enzymes necessary for sesquiterpene synthesis are not present in *N. gaditana*. The computational analyses were very useful in designing a suitable experimental strategy for terpene synthase expression in this host. Keeping in mind the above mentioned challenges, we finalised a sesquiterpene target biosynthesis based on a cytosolic expression of a terpene synthase as we hypothesized a FPP pool in the cytoplasm even though the amount of FPP substrate might be dynamic and ever-changing[48,49].

Table 7. MEP pathway genes in *A. thaliana* and their orthologs in *N. gaditana.*

The *A. thaliana* MEP pathway enzymes locus, function and enzyme commission number were taken from TAIR[75] website. The protein sequences from *N. gaditana* were analysed using DeppLoc[77] and ASAFind[79,80]. *N. gaditana* enzyme accession numbers and annotation were obtained from NCBI. The table also provides the Query coverage, percentage identity and E value of the *N. gaditana* sequence when using *A. thaliana* protein sequence query.

N.gaditana NCBI accession number	***N. gaditana Annotation***	***Tair Locus of A. thaliana query***	***EC number of query***	***Function of the A. thaliana query***	***A. thaliana NCBI accession number***	***Query cover***	***E value***	***% identify***	***N. gaditana ASAFind prediction (confidence)***	***N. gaditana DeepLoc prediction (Likelihood)***
EWM22144.1	1-deoxy-d-xylulose 5-phosphate synthase	AT4G15560 (DXS/CLA1)	2.2.1.7	A protein with 1-deoxyxylulose 5-phosphate synthase activity involved in the MEP pathway	NP_193291.1	0.79	0	0.5772	No	Chloroplast (0.9994)
EWM24175.1	1-deoxy-d-xylulose 5-phosphate reductoisomerase	AT5G62790 (DXR)	1.1.1.267	1-Deoxy-d-xylulose 5-phosphate reductoisomerase (DXR) catalyses the first committed step of the 2-C-methyl-d-erythritol 4-phosphate pathway for isoprenoid biosynthesis.	NP_201085.1	0.27	0.000004	0.6857	No	Mitochondrion (0.9781)

EWM22399.1	2-c-methyl-d-erythritol 4-phosphate cytidylyltransferase	AT2G02500 (ISPD/CMS)	2.7.7.60	A protein with 4-Diphosphocytidyl-2C-methyl-D-erythritol synthase activity	NP_001325418.1	0.74	3E-75	0.5279	Chloroplast (High)	Chloroplast (0.9934)
EWM29944.1	4-diphosphocytidyl-2c-methyl-d-erythritol kinase	AT2G26930 (CDPMEK)	2.7.1.148	A 4-(cytidine 5'-phospho)-2-C-methyl-D-erithritol kinase	NP_180261.1	0.67	9E-70	0.4572	Chloroplast (High)	Chloroplast (0.5842)
EWM23179.1	2-c-methyl-d-erythritol -cyclodiphosphate	AT1G63970 (ISPF)	4.6.1.12	A protein with 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase activity	NP_850971.1	0.65	4E-69	0.6433	Chloroplast (High)	Chloroplast (0.862)
EWM25176.1	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, partial	AT5G60600 (HDS)	1.17.7.1	A chloroplast-localised hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate (HMBPP) synthase (HDS), catalyses the formation of HMBPP from 2-C-methyl-	NP_001119467.1	0.34	3E-11	0.4576	No	Cytoplasm (0.395)

				D-erythrytol 2,4-cyclodiphosphate (MEcPP).						
EWM26997.1	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	AT4G34350 (HDR)	1.17.7.4	A protein with 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase activity	NP_567965.1	0.88	8E-171	0.5524	Chloroplast (High)	Chloroplast (0.7576)
EWM28775.1	geranylgeranyl pyrophosphate synthase	AT4G36810 (GGPS1)	2.5.1.29, 2.5.1.10, 2.5.1.1	A protein with geranylgeranyl pyrophosphate synthase activity involved in isoprenoid biosynthesis.	NP_195399.1	0.95	4E-114	0.5665	No	Mitochondrion (0.6829)
EWM28774.1	geranylgeranyl pyrophosphate synthase	AT4G36810 (GGPS1)	2.5.1.29, 2.5.1.10, 2.5.1.1	A protein with geranylgeranyl pyrophosphate synthase activity involved in isoprenoid biosynthesis.	NP_195399.1	0.93	9E-114	0.5096	Chloroplast (High)	Chloroplast (0.9988)
EWM28148.1	decaprenyl-diphosphate synthase subunit 1	AT2G34630 (GPS1)	2.5.1.1	geranyl diphosphate synthase	NP_001031483.1	0.65	7E-75	0.4236	No	Mitochondrion (0.7604)

EWM26176.1	Polyprenyl synthetase	AT2G34630 (GPS1)	2.5.1.1	geranyl diphosphate synthase	NP_001031483.1	0.56	4E-81	0.4236	No	Chloroplast (0.6623)
EWM21458.1	farnesyl pyrophosphate synthase	AT4G17190 (FPS2)	2.5.1.10, 2.5.1.1	A protein with farnesyl diphosphate synthase activity, which catalyses the rate limiting step in isoprenoid biosynthesis.	NP_193452.1	0.87	5E-117	0.4798	No	Mitochondrion (0.5982)
EWM26314.1	farnesyl-diphosphate farnesyltransferase	AT4G34640 (SQS1)	2.5.1.21	Squalene synthase, which converts two molecules of farnesyl diphosphate (FPP) into squalene.	NP_195190.1	0.48	3E-84	0.3949	No	Mitochondrion (0.558)
EWM27595.1	isopentenyl-diphosphate delta-isomerase	AT3G02780 (IPP2)	5.3.3.2	A protein with isopentenyl diphosphate:dimethylallyl diphosphate isomerase activity.	NP_186927.1	0.83	2E-47	0.3656	Chloroplast (low)	Chloroplast (0.9975)
EWM25927.1	NUDIX hydrolase domain-like protein	AT1G79690 (NUDT3)	5.3.3.2	A nudix hydrolase homolog 3	NP_001320544.1	0.56	3E-24	0.4148	Chloroplast (High)	Cytoplasm (0.6947)

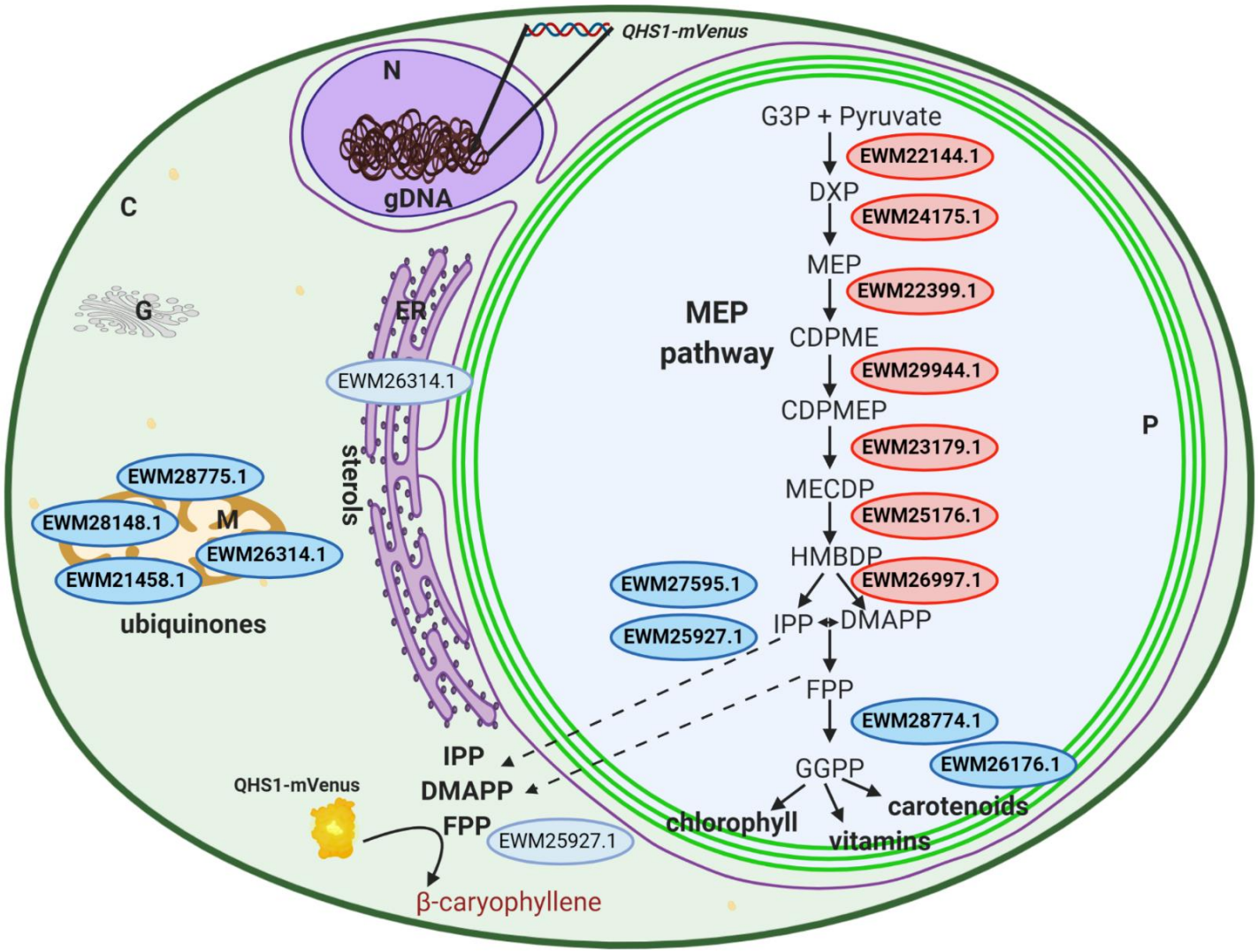

QHS1-mVenus
N
gDNA
C
G
ER
EWM26314.1
sterols
EWM28775.1
EWM28148.1
M
EWM26314.1
EWM21458.1
ubiquinones
QHS1-mVenus
IPP
DMAPP
FPP
EWM25927.1
β-caryophyllene
G3P + Pyruvate
EWM22144.1
DXP
EWM24175.1
MEP
EWM22399.1
MEP pathway
CDPME
EWM29944.1
CDPMEP
EWM23179.1
MECDP
EWM25176.1
HMBDP
EWM26997.1
EWM27595.1
EWM25927.1
IPP
DMAPP
FPP
EWM28774.1
GGPP
EWM26176.1
chlorophyll
carotenoids
vitamins
P

Figure 15. A hypothetical model for isoprenoid biosynthesis via MEP pathway and their enzyme localisation in *N. gaditana.*

The microalga *N. gaditana* produces the precursors IPP and DMAPP solely through the MEP pathway as mevalonate pathway is absent. The precursor pool present in the chloroplast and presumably transported to the cytosol and other organelles via unknown means (dashed arrow) is used for the synthesis of prenyl diphosphates by specific prenyltransferases in multiple sub-cellular compartments within the cell. N- nucleus, P- plastid, C- cytosol, ER- endoplasmic recticulum, G- golgi, M- mitochondrion, gDNA- genomic DNA of N. gaditana, QHS1-mVenus is the fusion protein β-caryophyllene synthase from A. annua and a yellow fluorescent protein (YFP) variant-mVenus, DXP- 1-deoxy-D- xylulose 5-phosphate; MEP 2-C-methyl-D-erythritol 4-phosphate, CDP-ME- 4-(cytidine 5'-diphospho)- 2-C-methyl-D-erythritol, CDP-MEP CDP-ME 2-phosphate, MECDP- 2-C-methyl-D-erythritol 2,4-cyclodiphosphate, HMBDP- 4-hydroxy-3-methylbut-2-enyl diphosphate. The putative sub-cellular localisations were inferred from the bioinformatics analysis reported in Table 7.

4.3.2 AaQHS1 is successfully expressed under the control of Hsp90 promoter in the cytosol of transgenic *N. gaditana*

Under the assumption that a cytosolic pool of FPP would be available for terpenoid production; a sesquiterpene cyclase cDNA from *Artemisia annua* (AF472361) that encodes a β-caryophyllene synthase (QHS1) with an open reading frame of 1,644 bp and yields a 60 kD protein[72] was synthesised, cloned in-frame with *mVenus* reporter gene and expressed under the control of endogenous *Hsp90* promoter (Chapter 2) in *N. gaditana*[94]. This promoter was selected because it had the highest promoter activity of all the promoters studied (Chapter 2 and 3) with maximum performance in early to mid-log phase of culture.

The QHS1 enzyme catalytically converts farnesyl diphosphate (FPP) derived from the MEP pathway (Figure 15) into β caryophyllene, but was not active with geranylgeranyl diphosphate (C20)[72]. No *N. gaditana* signal peptides were incorporated with the synthetised QHS1 enzyme and a cytosolic expression was expected. A few colonies were screened using mVenus specific primers (Table S19). The transgenic *E. coli* control was analysed for leaky expression of mVenus reporter using flow cytometry (Figure S17), hence scaled up and analysed for β-caryophyllene production using GC-MS analysis. The GC-MS extracted-ion chromatogram (EIC) of transgenic QHS1 *E. coli* cells showed a β-caryophyllene peak with a GC retention time identical to that of a trans-β-caryophyllene analytical standard (Figure 16, S17) which was absent in control *E. coli* cells. This confirms that the fusion protein produced in transgenic *E. coli* cells is active and resulted in detectable amounts of β-caryophyllene (Figure 16, S17) which was approximately 20 ng of β-caryophyllene. mg^{-1} dry weight of *E. coli* (Figure 17).

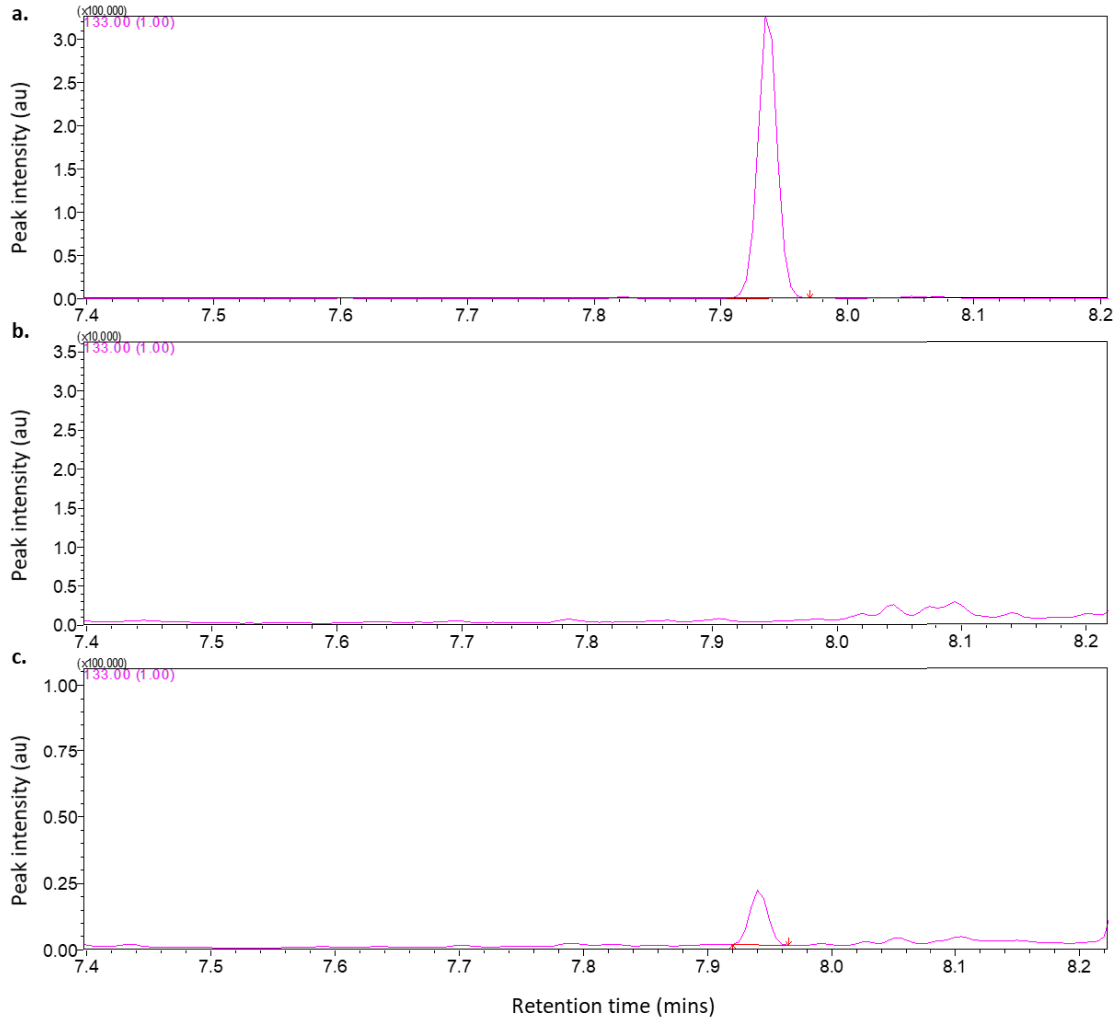


Figure 16. GC-MS extracted-ion chromatogram (EIC) of caryophyllene (m/z 133) in control (b) and transgenic (c) *E. coli.*

a. The β-caryophyllene analytical standard peak at a retention time of 7.94 mins; **b.** β-caryophyllene peak is absent in the control *E. coli* sample **c.** β-caryophyllene peak is present in the transgenic *E. coli* sample expressing *QHS1-mVenus* gene fusion.

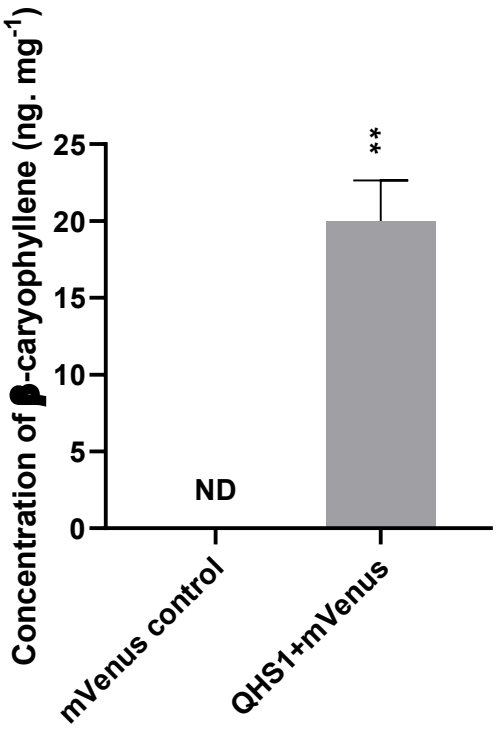


Figure 17. The concentration of β-caryophyllene produced in transgenic *E. coli*.

The transgenic *E. coli* expressing pNaga4.mV.DC.N2 (control vector) or pNaga4.mV.DC.N2.QHS1 were analysed using GC-MS for β-caryophyllene production. Error bars represent the standard deviation for control and QHS1 expressing *E. coli* (n=3). Significant differences between control and QHS1 expressing *E. coli* was calculated using Paired t-test. ** P value- 0.0058.

N. gaditana was transformed using the above-mentioned constructs and the zeocin-resistant colonies were screened using flow cytometry (Figure S19) to check for stable nuclear expression of *QHS1-mVenus* fusion reporter and identify the clones with the highest mVenus expression. Before cultivating *N. gaditana* with isopropyl myristate for volatile terpenoid capture, the effects of this ester on the growth and photosynthetic health of *N. gaditana* was evaluated using cell count and PAM to measure the maximum quantum yield of photosystem II. The cultivation with 3.2% isopropyl myristate did not show a significant effect on growth or photosynthetic health of wild type *N. gaditana* (Figure 18).

Out of the 18 transgenic lines (Figure S19), 10 mVenus positive transgenic lines were selected and scaled up (50 mL) for GC-MS analysis. None of the lines showed detectable amounts of β-caryophyllene (data not shown). The transgenic lines were not

stable and lost *QHS1-mVenus* fusion gene within a couple of months of sub-culture (Figure S20), hence GC-MS analysis using a bigger culture volume was not possible. To obtain stable transgenic *N. gaditana* lines expressing *QHS1-mVenus* fusion *Nannochloropsis* transformation was repeated and 12 additional colonies were obtained. These colonies were screened for mVenus expression using flow cytometry (Figure S21) to check for stable *QHS1-mVenus* expression over a period of 3 months. After three months of sub-culture only four transgenic lines stably maintained the expression of *QHS1-mVenus* construct. Three transgenic lines were selected to verify the sub-cellular localisation of the QHS1-mVenus fusion protein in *N. gaditana* transgenic lines; these transgenic lines including controls were examined by confocal microscopy[95]. The mVenus (YFP) signal was observed to be cytosolic and was independent of the chlorophyll auto-fluorescence from the chloroplast (Figure 19). *In vivo* localisation studies using fluorescent proteins have been conducted in *N. oceanica* CCMP1779 by Moog et al., where they show that despite the tiny size, *Nannochloropsis* sp. has huge potential as a model system for in vivo protein localisation studies[95]. They were able to successfully target GFP fusion protein to the cytosol, endoplasmic reticulum (along with compartments connected via vesicle transfer such as golgi), mitochondria, chloroplast stroma and the periplastidal compartment using signal peptides from marker proteins for different cellular compartments[95].

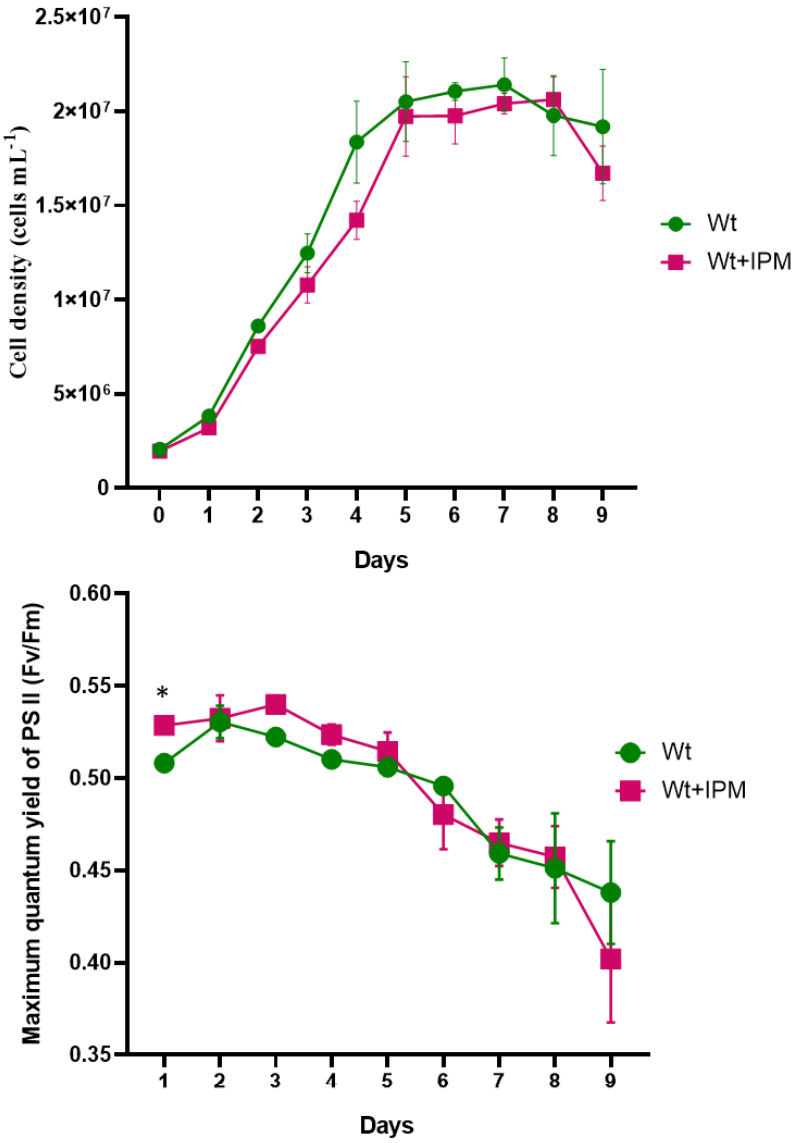


Figure 18. The growth and photosynthetic activity of wild type *N. gaditana.*

a. The growth curve of wild type *N. gaditana* in presence and absence of isopropyl myristate grown as a bi-phasic culture. Error bars represent the standard deviation for wild type (n=3). Significant differences in growth between wild type in presence and absence of isopropyl myristate was calculated using 2-way ANOVA Sidak's multiple comparisons test. **b.** Photosynthetic activity of wild type *N. gaditana* in presence and absence of isopropyl myristate. Error bars represent the standard deviation for wild type (n=3). Significant differences in growth between wild type in presence and absence of isopropyl myristate was calculated using 2-way ANOVA Sidak's multiple comparisons test. * P value- 0.0407.

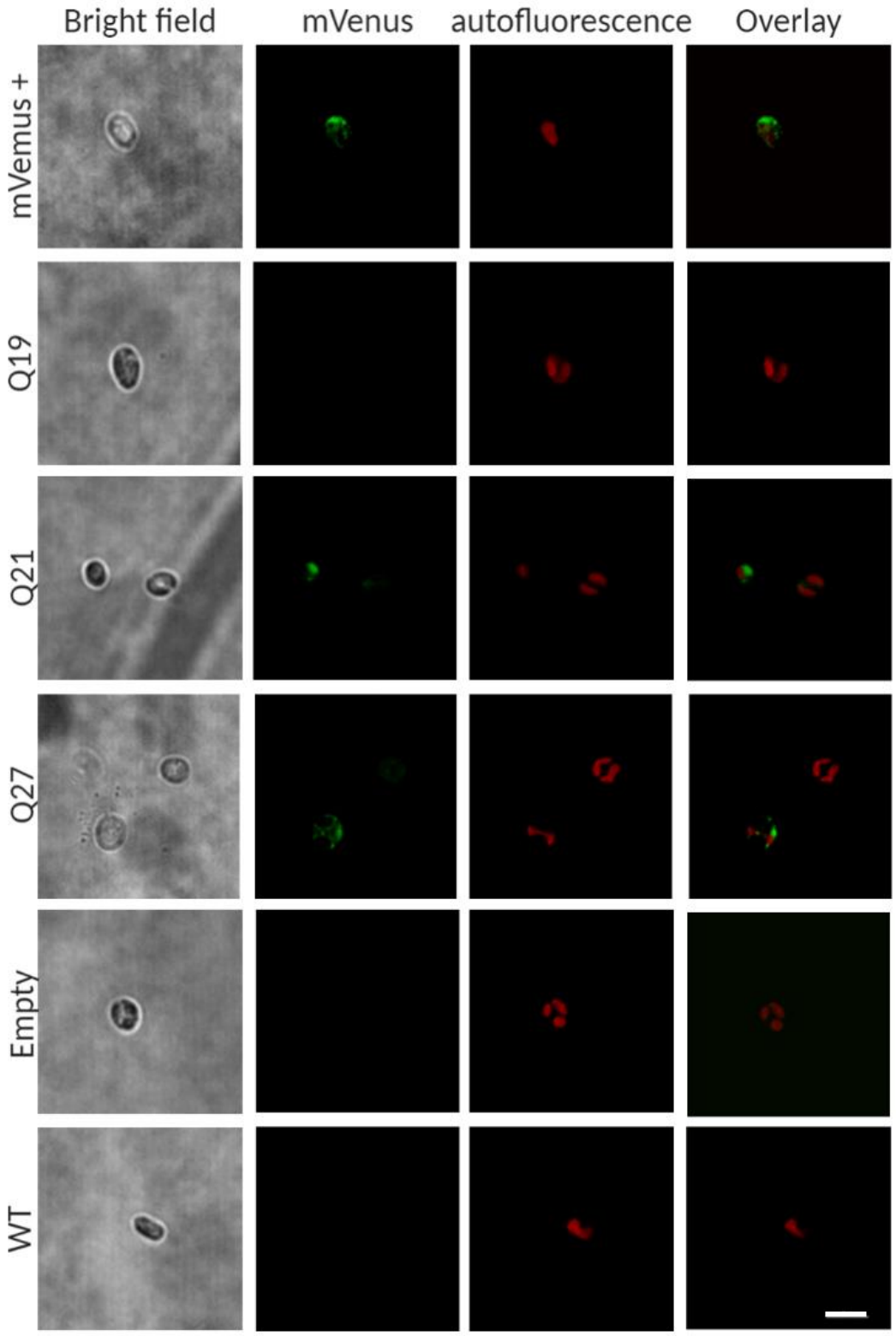


Figure 19. Confocal microscopy images of transgenic representative *N. gaditana* cell lines expressing *AaQHS1-mVenus* in the cytosol and controls.

N. gaditana expressing mVenus positive control and QHS1-mVenus fusion protein (Q19, Q21, Q27) along with *N. gaditana* carrying empty vector control and wild type cells where no mVenus

fluorescence was observed. Microscope magnification: 1000×, the red and green fluorescence was emitted by chloroplasts and mVenus (YFP) fluorescence, respectively. Scale bar is 5µm.

4.3.3 Random chromosomal integration of transgene has impact on growth and photosynthetic health of the transgenic *N. gaditana* lines

In order to test the effect of random transgene integration in the transgenic *N. gaditana* lines, three lines were scaled up. The growth curve and maximum quantum yield of photosystem II of all the transgenic and wild-type *N. gaditana* lines are shown in Figure 20 and Figure S23. Initially, the growth rate of the transgenic lines was significantly different than that of the wild type but stabilised when the cells entered mid-log phase (Figure S22). The significant differences in quantum yield of PS II between wild type and all the transgenic lines analysed suggest that random genomic integration may have had detrimental effects on algal photosynthetic fitness. This fitness trend was observed in the transgenic lines studied in previous chapters (2.3.4, 3.3.2 and 3.3.3) as well. On the other hand, there was no significant difference in growth between transgenic *N. gaditana* mVenus control line and the transgenic lines expressing QHS1-mVenus fusion except Q27 on day 4 and 6, suggesting that QHS1 expression did not influence the growth of this microalga in 50 mL cultures grown under laboratory conditions. It might be due to the fact that β-caryophyllene accumulation was negligible to impact growth, hence the same cannot be said for higher volume cultures. Similarly, the overall maximum quantum yield of photosystem II (photosynthetic health) of transgenic *N. gaditana* mVenus control line and the transgenic lines expressing QHS1-mVenus fusion was not significantly different except on Day 5 when the mVenus control had a lower maximum quantum yield compared to Q21 and Q19 (Figure 20 and S23). The QHS1 transgenic line Q27 had lower maximum quantum yield compared to mVenus control on Day 4 and Day 2 respectively (Figure 20 and S23). In general, it was observed in transgenic maize plants that over-expressed caryophyllene synthase constitutively that the β-caryophyllene production affected seed germination, plant growth and yield[96].

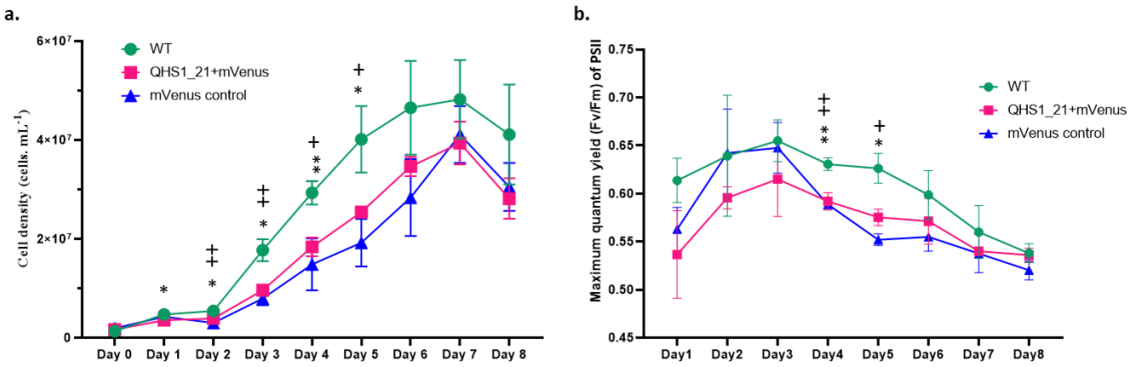


Figure 20. The growth, photosynthetic health of transgenic *N. gaditana*.

a. Growth curve of wild type and transgenic *N. gaditana* cultures (Q21) expressing QHS1-mVenus fusion protein including positive control measured over a period of 8 days. Rest of the transgenic lines are included in supplementary data. Error bars represent the standard deviation for wild type and transgenic lines (n=3). **b.** Photosynthetic activity of *N. gaditana* wild type and transgenic lines under study. Error bars represent the standard deviation for wild type and transgenic lines (n=3). Significant differences between wild type and the transgenic lines were calculated using 2-way ANOVA Dunnett's multiple comparison test. Significant differences between wild type and (positive control- +, Q21- *). * P<0.05, ** P<0.005.

4.3.4 Hsp90 promoter drove *AaQHS1-mVenus* fusion expression which lead to β-caryophyllene biosynthesis in *N. gaditana*

The median mVenus fluorescence intensity was also measured for all the transgenic *N. gaditana* lines under study, including wild type and mVenus control which carried the *mVenus* reporter gene alone under the control of *Hsp90* promoter (Figure 21). The median mVenus fluorescence of positive control was 12-fold higher than the background fluorescence observed in wild type cells which is similar to the median mVenus fluorescence observed in transgenic lines with Hsp90 promoter (Chapter 2, data not shown) which was 8 to 9-fold higher than the background fluorescence observed in wild type cells.

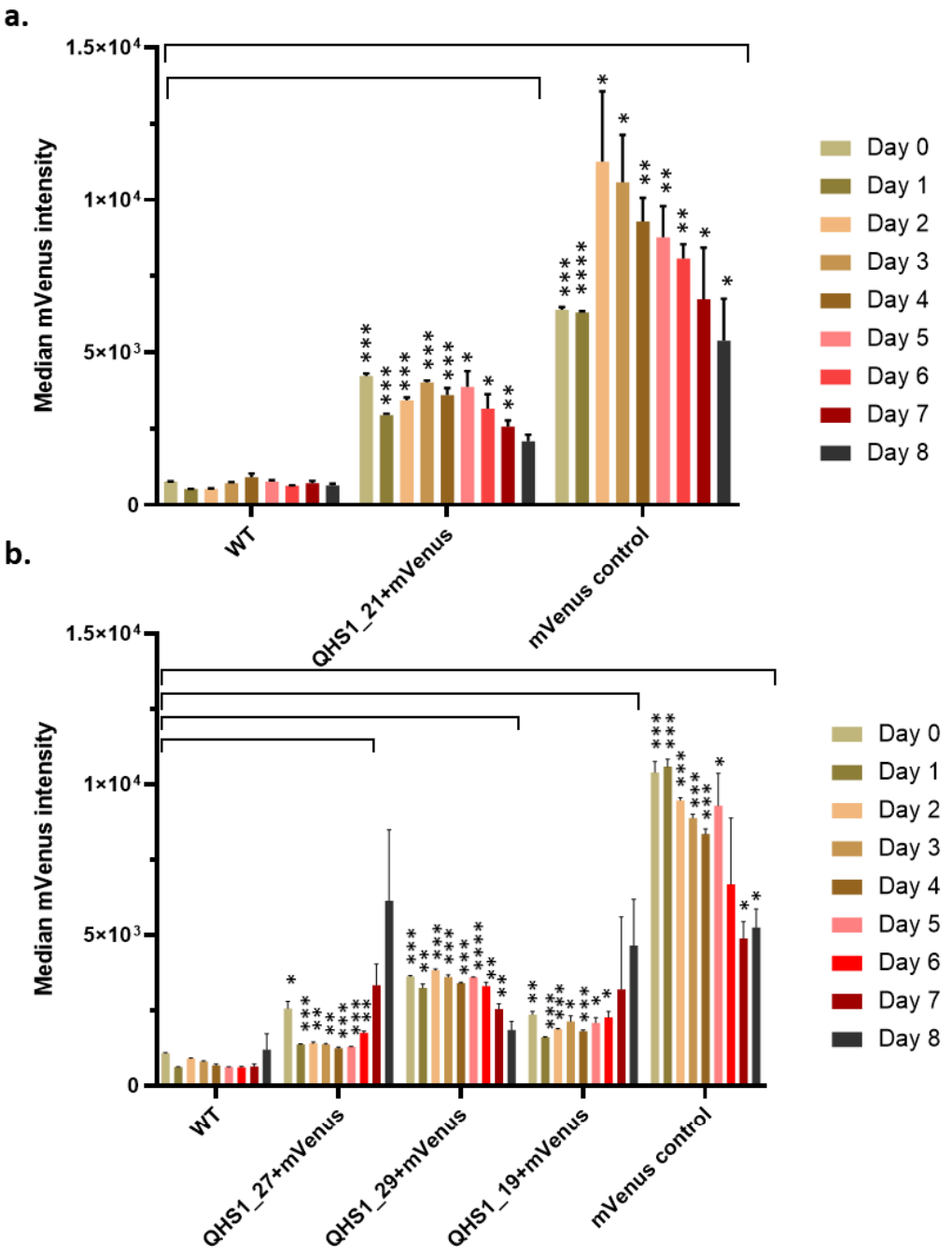


Figure 21. The median mVenus fluorescence of transgenic *N. gaditana*.

a. The median fluorescence intensity of mVenus in wild type and transgenic *N. gaditana* line Q21**. b.** The median fluorescence intensity of mVenus in wild type and transgenic *N. gaditana* lines Q19, Q27 and Q27 measured over a period of eight days. Error bars represent the standard deviation for wild type and transgenic lines (n=3). Significant differences between wild type and the transgenic lines were calculated using 2-way ANOVA Dunnett's multiple comparison test. * P<0.05, ** P<0.005, *** P<0.0005, **** P<0.0001.

The same cannot be said for the transgenic *N. gaditana* cell lines carrying QHS1-mVenus fusion protein, the median mVenus fluorescence of these lines was in the range of 2.8 – 4.9-fold higher than the background fluorescence observed in the wild type. We hypothesize that only the clonal lines with moderate *QHS1-mVenus* expression survive and the high expressing clones do not survive due to cytotoxicity of β-caryophyllene accumulation[96]. For e.g. Q18 clonal line which did not grow properly in liquid culture and was lost within a few sub-cultures had the highest mean mVenus fluorescence (Figure S19). At this point, the cytotoxicity of β-caryophyllene is just speculation and needs to be experimentally verified.

The *Hsp90* promoter activity was highest in the early to mid-log phase of culture (based on flow cytometry) similar to the observation made in Chapter 2 (section 3.4). The three transgenic lines (Q19, Q21, Q27) along with negative control (expressing mVenus alone) were scaled up to 1 L and the culture pelleted on Day 5 (mid-exponential phase) to get sufficient biomass from an exponential stage culture for β-caryophyllene detection using GC-MS. The dry biomass obtained from the mVenus control culture was approximately four fold higher than the *QHS1* expressing transgenic lines (dry weight: control-200.3 mg, Q19-35.1 mg, Q21-98.5 mg, Q27-44.2 mg). Even though previously, no significant difference in growth rate between mVenus control line and the QHS1-mVenus expressing transgenic lines were observed on Day 5 (Figure S22). It might be that in a bigger culture volume (1 L and above) the concentration of β-caryophyllene negatively affects the overall rate of cell division and health of the transgenic *N. gaditana* expressing QHS1-mVenus. Hence production using an inducible promoter was suggested, but the putative phosphate-inducible promoters tested in Chapter 3 did not provide any positive outcome and could not be used for p-inducible terpenoid production in *N. gaditana*.

Out of the three transgenic *N. gaditana* lines expressing QHS1-mVenus, the transgenic line with the highest observed mVenus expression (Q21) using flow cytometry was the only one where trace amount of β-caryophyllene was detected in the biomass fraction and not in the IPM layer (Figure 7). The SIM spectra of this transgenic line have a peak with a retention time 7.94 mins, similar to the β-caryophyllene standard used (Figure 22). In the other transgenic lines, the β-caryophyllene levels were below

the detection level. It was assumed that the cell wall[97] of *N. gaditana* was thick enough to contain β-caryophyllene within the cell. Later, Sadre et al., highlighted the function of cytosolic lipid droplets as traps for the terpenoids in the cells of *Nicotiana benthamiana* leaves[98]. As *N. gaditana* produces lipid droplets in the cytosol[99], assuming that the lipid in the cell prevented the release of β-caryophyllene outside the *N. gaditana* cell. Under the described laboratory growing conditions 20 ng. g of dry cell weight^{-1} of β-caryophyllene was extracted from cytoplasmic expression of the QHS1-mVenus fusion in *N. gaditana*. As stated above, this terpenoid synthase uses FPP as substrate which is a precursor for sterols and ubiquinone (UQ) and not photosynthetic terpenoids (precursor-GGPP/GPP). Furthermore, *Nannochloropsis spp.* lacks the mevalonate pathway[55,56,64], the FPP pool must be produced from IPP/DMAPP generated within the plastid. It is still unclear whether the transport of these precursors occurs through active transport with specialised transporters or passive transport. This factor may be a rate-limiting step in the production of β-caryophyllene in the cytosol. It is plausible that the cytosol has limited amounts of FPP and a larger pool is available in the mitochondria or ER or chloroplast[93,100,101]. The estimation of an available FPP pool in the various cellular organelles has not been reported, yet. Availability of such data would be valuable for targeting the specific terpene synthases to the respective cellular compartments.

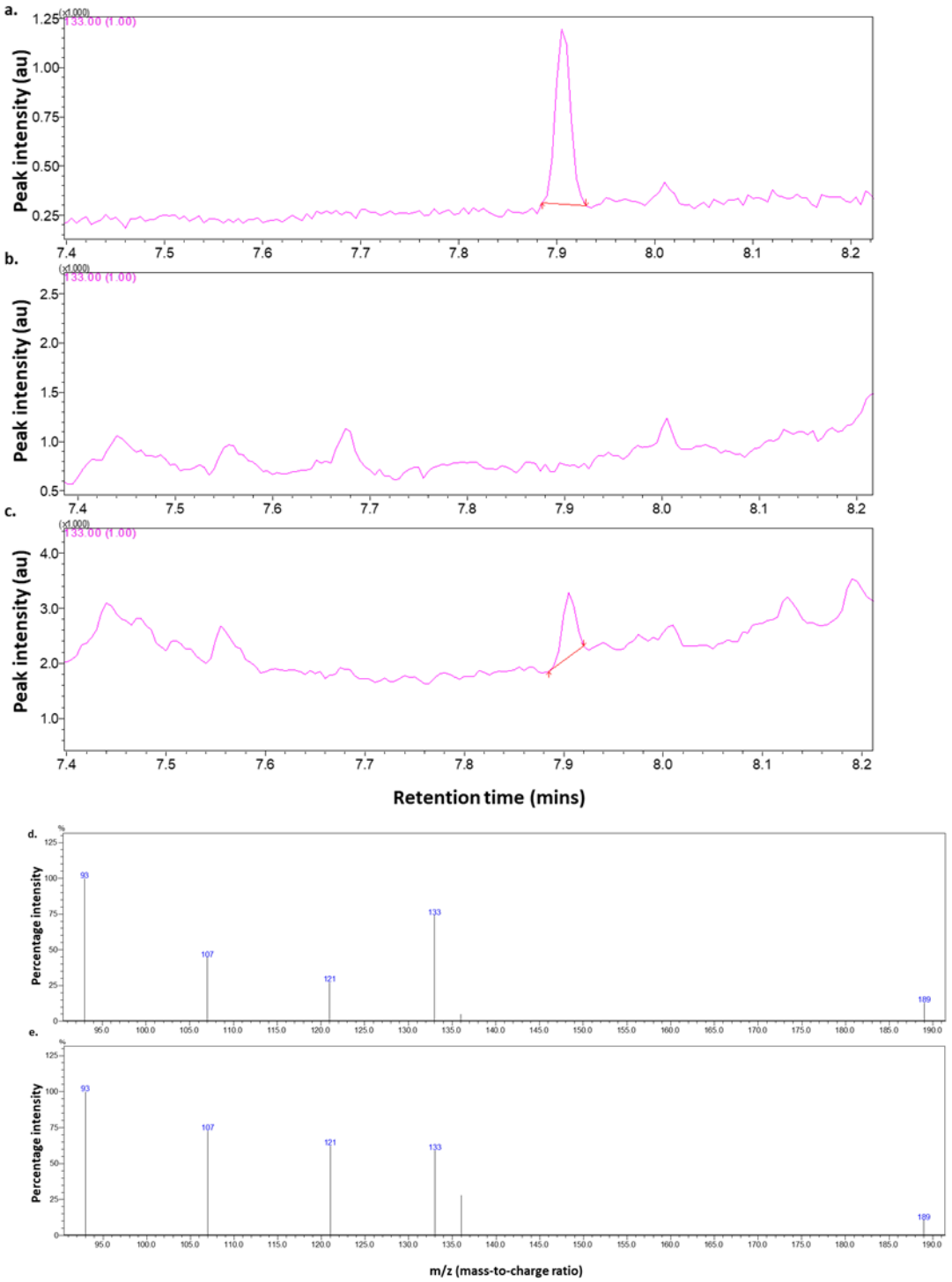


Figure 22. GC-MS extracted-ion chromatogram (EIC) of caryophyllene (m/z 133) in control (b) and transgenic (c) *N. gaditana.*

a. The β-caryophyllene analytical standard peak at a retention time of 7.94 mins; **b.** β-caryophyllene peak is absent in the control *N. gaditana* **c.** β-caryophyllene peak is present in the transgenic *N. gaditana* (Q21) expressing *QHS1-mVenus* gene fusion. **d.** The SIM spectra of β-caryophyllene standard (along with internal standard limonene). **e.** The SIM spectra of β-caryophyllene produced by the Q21 transgenic *N. gaditana* cell line (internal standard limonene).

Various photosynthetic hosts have been used for sesquiterpene production in the past such as the cyanobacteria *Synechococcus*[102], *Synechocystis*[36], *Anabaena* sp.[103] and the moss *Physcomitrella patens*[104]. As specified earlier, the sesquiterpenoid β-caryophyllene has been produced in engineered *E.coli* under aerobic fed-batch fermentation yielding 0.34 mg h^{-1} g^{-1} dry cells[35] and 1.15 mg h^{-1} g^{-1} dry cells using acetic acid as substrate[41] while Wu et al., 2018 demonstrated that their engineered *E. coli* strains could produce 100 mg/L β-caryophyllene using marine microalgal hydrolysate as the sole carbon source[24]. Reinsvold et al., were able to produce β-caryophyllene in a transgenic strain of cyanobacterium *Synechocystis* to approximately 3.7 µg. g of dry cell $weight^{-1}$. Even-though the amount detected in *N. gaditana* is many magnitudes lower than the other models, it was just the first attempt at terpenoid synthesis in this microalga using heterologous gene expression. Understanding the drivers of carbon flux in the MEP metabolic pathway, various prenyltransferases sub-cellular localisations and action could aid in development of metabolic engineering strategies for optimal production of novel terpenoids in this microalga. In this case, the production of β-caryophyllene is dependent on the presence of a pool of available FPP and its sub-cellular location, sufficient enzyme titer (which were not estimated in this study) and examining these are promising avenues for future research into strain improvement.

Specific to terpenoids, the identification and characterisation of prenyltransferases in *N. gaditana* would shed light into their substrate specificity, kinetics and localisation of these enzymes in the cell, which along with the information of substrate (GPP, FPP and GGPP) pool availability and location can greatly benefit researches develop strategies to target the heterologous terpene synthase enzymes (whether single or several) to the ideal sub-cellular locations. Understanding the metabolic network and carbon flux regulatory mechanisms[37] would assist in identifying gene targets that could be under or over-expressed to boost the flux through the desired pathway for better accumulation of enzyme substrates and down regulation of competing pathways without compromising the performance health of the microalgal host. *Nannochloropsis* has the capacity to become a GM platform that could support an array of commercially viable products provided these microalgae are given the much needed research attention by developing dedicated genetic tools and techniques[105].

4.4 Conclusion

This work represents the first report of terpenoid engineering in *N. gaditana* and is a proof of concept for using this microalga as a sesquiterpenoid production platform. We expect that significant improvement could be attained through metabolic, genetic and bioprocess engineering approaches. There is emerging interest in photosynthetic microbial platforms for the production of metabolites as they use CO_2 as carbon source, have a high NADPH content and offer higher native flux capacity[106]. In light of the limitations stated previously such as lack of knowledge of availability of prenyl diphosphate substrate pool in the cytoplasm or the different organelles and no targeting signal peptides have been characterised, the cytosolic expression of the terpene synthase was a first obliged choice, resulting in detectable amounts of β-caryophyllene. It is currently unknown what the upper limit of sesquiterpene production could be from the native pool of FPP which might be present in the various sub-cellular organelles and could be used for non-native products. There is enormous room for strain improvement provided metabolic research data on pathway regulation, flux, location and activity of the various pathway enzymes, transport mechanism of IPP/DMAPP precursors becomes available. Having characterised signal peptide sequences for sub-cellular targeting which is already available for other model organisms will be advantageous to design and execute future metabolic engineering experiments in this promising microalga.

4.5 Acknowledgement

This research was supported by the UTS President's Research Scholarship and UTS International Research Scholarship. I would like to thank Kun Xiao for technical assistance with flow cytometry and FACS and Taya Lapshina for technical assistance with GC-MS

Statement of informed consent, human/animal rights

No conflicts, informed consent, human or animal rights applicable.

4.6 References

1. Gershenzon, J. & Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **3,** 408–414 (2007).

2. Kainer, D. *et al.* Plant-Derived Terpenes: A Feedstock for Specialty Biofuels. *Trends Biotechnol.* **35,** 227–240 (2016).

3. Moses, T., Pollier, J., Thevelein, J. M. & Goossens, A. Bioengineering of plant (tri)terpenoids: From metabolic engineering of plants to synthetic biology in vivo and in vitro. *New Phytol.* **200,** 27–43 (2013).

4. Davies, F. K., Jinkerson, R. E. & Posewitz, M. C. Toward a photosynthetic microbial platform for terpenoid engineering. *Photosynth. Res.* **123,** 265–284 (2015).

5. Awouafack, M. D., Tane, P., Kuete, V. & Eloff, J. N. *Sesquiterpenes from the Medicinal Plants of Africa*. *Medicinal Plant Research in Africa: Pharmacology and Chemistry* (Elsevier Inc., 2013).

6. Cane, D. E. Enzymatic Formation of Sesquiterpenes. *Chem. Rev.* **90,** 1089–1103 (1990).

7. Kanehisa, M. & Susumu, G. KEGG:Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).

8. Liu, H. *et al.* Physicochemical characterization and pharmacokinetics evaluation of β-caryophyllene/β-cyclodextrin inclusion complex. *Int. J. Pharm.* **450,** 304–310 (2013).

9. 'caryophyllene'2013,The Merck Index Online. (2013). Available at: https://www.rsc.org/Merck-Index/monograph/m3145/caryophyllene?q=authorize.

10. Sköld, M., Karlberg, A. T., Matura, M. & Börje, A. The fragrance chemical β-caryophyllene - Air oxidation and skin sensitization. *Food Chem. Toxicol.* **44,** 538–545 (2006).

11. Sabulal, B. *et al.* Caryophyllene-rich rhizome oil of *Zingiber nimmonii* from South India: Chemical characterization and antimicrobial activity. *Phytochemistry* **67,** 2469–2473 (2006).

12. Ghelardini, C., Galeotti, N., Di Cesare Mannelli, L., Mazzanti, G. & Bartolini, A. Local anaesthetic activity of β-caryophyllene. *Farmaco* **56,** 387–389 (2001).

13. Chang, H.-J., Kim, J.-M., Lee, J.-C., Kim, W.-K. & Chun, H. S. Protective Effect of β-Caryophyllene, a Natural Bicyclic Sesquiterpene, Against Cerebral Ischemic Injury. *J. Med. Food* **16,** 471–480 (2013).

14. Lourens, A. C. U., Reddy, D., Başer, K. H. C., Viljoen, A. M. & Van Vuuren, S. F. In vitro biological activity and essential oil composition of four indigenous South African *Helichrysum* species. *J. Ethnopharmacol.* **95,** 253–258 (2004).

15. Bento, A. F. *et al.* β-caryophyllene inhibits dextran sulfate sodium-induced colitis in mice through CB2 receptor activation and PPARγ pathway. *Am. J. Pathol.* **178,** 1153–1166 (2011).

16. Oliveira-Tintino, C. D. de M. *et al.* Anti-inflammatory and anti-edematogenic action of the *Croton campestris* A. St.-Hil (Euphorbiaceae) essential oil and the compound β-caryophyllene in in vivo models. *Phytomedicine* **41,** 82–95 (2018).

17. Santos, A. O. *et al.* Effect of Brazilian copaiba oils on *Leishmania amazonensis. J. Ethnopharmacol.* **120,** 204–208 (2008).

18. Basha, R. H. & Sankaranarayanan, C. Protective role of β-caryophyllene, a sesquiterpene lactone on plasma and tissue glycoprotein components in streptozotocin-induced hyperglycemic rats. *J. Acute Med.* **5,** 9–14 (2015).

19. Varga, Z. V. *et al.* β-Caryophyllene protects against alcoholic steatohepatitis by attenuating inflammation and metabolic dysregulation in mice. *Br. J. Pharmacol.* **175,** 320–334 (2018).

20. Gertsch, J. *et al.* Beta-caryophyllene is a dietary cannabinoid. *Proc. Natl. Acad. Sci.* **105,** 9099 (2008).

21. Yang, X., Lederer, C., McDaniel, M. & Deinzer, M. Hydrolysis Products of Caryophyllene Oxide in Hops and Beer. *J. Agric. Food Chem.* **41,** 2082–2085 (1993).

22. Nance, M. R. & Setzer, W. N. Volatile components of aroma hops (*Humulus lupulus* L.) commonly used in beer brewing. *J. Brew. Distill.* **2,** 16–22 (2011).

23. Harvey, B. G., Merriman, W. W. & Koontz, T. A. High-density renewable diesel and jet fuels prepared from multicyclic sesquiterpanes and a 1-hexene-derived synthetic paraffinic kerosene. *Energy and Fuels* **29,** 2431–2436 (2015).

24. Wu, W., Liu, F. & Davis, R. W. Engineering *Escherichia coli* for the production of terpene mixture enriched in caryophyllene and caryophyllene alcohol as potential aviation fuel compounds. *Metab. Eng. Commun.* **6,** 13–21 (2018).

25. Cheng, A. X. *et al.* The rice (E)-β-caryophyllene synthase (OsTPS3) accounts for the major inducible volatile sesquiterpenes. *Phytochemistry* **68,** 1632–1641 (2007).

26. Jayaramaiah, R. H. *et al.* Functional characterization and transient expression manipulation of a new sesquiterpene synthase involved in β-caryophyllene accumulation in Ocimum. *Biochem. Biophys. Res. Commun.* **473,** 265–271 (2016).

27. Huang, M. *et al.* The major volatile organic compound emitted from *Arabidopsis thaliana* flowers, the sesquiterpene (E)-β-caryophyllene, is a defense against a bacterial pathogen. *New Phytol.* **193,** 997–1008 (2012).

28. Walter, R. H. β-Caryophyllene in native clove bud oil. *Phytochemistry* **11,** 405–406 (1972).

29. Chaieb, K. *et al.* The Chemical Composition and Biological Activity of Clove Essential Oil, Eugenia caryophyllata (*Syzigium aromaticum* L. Myrtaceae): A Short Review. *Phyther. Res.* **21,** 501–506 (2007).

30. Rasmann, S. *et al.* Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature* **434,** 732–737 (2005).

31. Zheng, G. Q., Kenney, P. M. & Lam, L. K. T. Sesquiterpenes from clove (eugenia caryophyllata) as potential anticarcinogenic agents. *J. Nat. Prod.* **55,** 999–1003 (1992).

32. Collado, I. G., Hanson, J. R. & Macías-Sánchez, A. J. Recent advances in the chemistry of caryophyllene. *Nat. Prod. Rep.* **15,** 187–204 (1998).

33. Dowling, M. S. & Vanderwal, C. D. Ring-closing metathesis of allylsilanes as a flexible strategy toward cyclic terpenes. Short syntheses of teucladiol, isoteucladiol, poitediol, and dactylol and an attempted synthesis of caryophyllene. *J. Org. Chem.* **75,** 6908–6922 (2010).

34. Salvagnin, U. *et al.* Homologous and heterologous expression of grapevine E-(β)-caryophyllene synthase (VvGwECar2). *Phytochemistry* **131,** 76–83 (2016).

35. Yang, J., Li, Z., Guo, L., Du, J. & Bae, H. J. Biosynthesis of β-caryophyllene, a novel terpene-based high-density biofuel precursor, using engineered *Escherichia coli*. *Renew. Energy* **99,** 216–223 (2016).

36. Reinsvold, R. E., Jinkerson, R. E., Radakovits, R., Posewitz, M. C. & Basu, C. The production of the sesquiterpene β-caryophyllene in a transgenic strain of the cyanobacterium *Synechocystis*. *J. Plant Physiol.* **168,** 848–852 (2011).

37. Vavitsas, K., Fabris, M. & Vickers, C. E. Terpenoid metabolic engineering in photosynthetic microorganisms. *Genes (Basel).* **9,** (2018).

38. Ajikumar, P. K. *et al.* Terpenoids: Opportunities for Biosynthesis of Natural Product Drugs Using Engineered Microorganisms. *Mol. Pharm.* **5,** 167–190 (2008).

39. Markus Lange, B. & Ahkami, A. Metabolic engineering of plant monoterpenes, sesquiterpenes and diterpenes-current status and future opportunities. *Plant Biotechnol. J.* **11,** 169–196 (2013).

40. Moser, S. & Pichler, H. Identifying and engineering the ideal microbial terpenoid production host. *Appl. Microbiol. Biotechnol.* **103,** 5501–5516 (2019).

41. Yang, J. & Nie, Q. Engineering *Escherichia coli* to convert acetic acid to β-caryophyllene. *Microb. Cell Fact.* **15,** 1–9 (2016).

42. Renault, H., Bassard, J. E., Hamberger, B. & Werck-Reichhart, D. Cytochrome P450-mediated metabolic engineering: Current progress and future challenges. *Curr. Opin. Plant Biol.* **19,** 27–34 (2014).

43. Zhou, K., Qiao, K., Edgar, S. & Stephanopoulos, G. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nat. Biotechnol.* **33,** 377–383 (2015).

44. Lohr, M., Schwender, J. & Polle, J. E. W. Isoprenoid biosynthesis in eukaryotic phototrophs: A spotlight on algae. *Plant Sci.* **185–186,** 9–22 (2012).

45. Lauersen, K. J. Eukaryotic microalgae as hosts for light-driven heterologous isoprenoid production. *Planta* **249,** 155–180 (2019).

46. D'Adamo, S. *et al.* Engineering the unicellular alga *Phaeodactylum tricornutum* for high-value plant triterpenoid production. *Plant Biotechnol. J.* **17,** 75–87 (2019).

47. Lauersen, K. J. *et al.* Phototrophic production of heterologous diterpenoids and a hydroxy-functionalized derivative from *Chlamydomonas reinhardtii*. *Metab. Eng.* **49,** 116–127 (2018).

48. Wichmann, J., Baier, T., Wentnagel, E., Lauersen, K. J. & Kruse, O. Tailored carbon partitioning for phototrophic production of (E)-α-bisabolene from the green microalga *Chlamydomonas reinhardtii*. *Metab. Eng.* **45,** 211–222 (2018).

49. Lauersen, K. J. *et al.* Efficient phototrophic production of a high-value sesquiterpenoid from the eukaryotic microalga *Chlamydomonas reinhardtii*. *Metab. Eng.* **38,** 331–343 (2016).

50. Jan, M. & Kazik, P. *Nannochloropsis: Biology, Biotechnological Potential and Challenges*. (Nova Science Publishers, Incorporated, 2017).

51. HK., L. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50,** 47–65 (1999).

52. Lange, B. M., Rujan, T., Martin, W. & Croteau, R. Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 13172–13177 (2000).

53. Lichtenthaler, H. K. Chapter 7 The Non-mevalonate DOXP/MEP (Deoxyxylulose 5-Phosphate/Methylerythritol 4-Phosphate) Pathway of Chloroplast Isoprenoid and Pigment Biosynthesis. in *The Chloroplast: Basics and Applications* (eds. Rebeiz, C. A. et al.) 95–118 (Springer Netherlands, 2010).

54. Rohmer, M. & Cvejic, J. H. CO 2 as main carbon source for isoprenoid biosynthesis via the mevalonate-independent methylerythritol 4-phosphate route in the marine diatoms *Phaeodactylum tricornutum* and *Nitzschia ovalis*. *Phytochemistry* **53,** 21–28 (2000).

55. Vieler, A. *et al.* Genome, Functional Gene Annotation, and Nuclear Transformation of the Heterokont Oleaginous Alga *Nannochloropsis oceanica* CCMP1779. *PLoS Genet.* **8,** (2012).

56. Lu, Y. *et al.* Regulation of the cholesterol biosynthetic pathway and its integration with fatty acid biosynthesis in the oleaginous microalga *Nannochloropsis oceanica*. *Biotechnol. Biofuels* **7,** 1–15 (2014).

57. Schwender, J. *et al.* Incorporation of 1-deoxy-D-xylulose into isoprene and phytol by higher plants and algae. *FEBS Lett.* **414,** 129–134 (1997).

58. Schwender, J., Gemünden, C. & Lichtenthaler, H. K. Chlorophyta exclusively use the 1-deoxyxylulose 5-phosphate/2-C-methylerythritol 4-phosphate pathway for the biosynthesis of isoprenoids. *Planta* **212,** 416–423 (2001).

59. Hsieh, M. H., Chang, C. Y., Hsu, S. J. & Chen, J. J. Chloroplast localization of methylerythritol 4-phosphate pathway enzymes and regulation of mitochondrial genes in ispD and ispE albino mutants in *Arabidopsis*. *Plant Mol. Biol.* **66,** 663–673 (2008).

60. Apt, K. E. In vivo characterization of diatom multipartite plastid targeting signals. *J. Cell Sci.* **115,** 4061–4069 (2002).

61. Sukenik, A., Carmeli, Y. & Berner, T. Regulation of fatty acid composition by irradiance level in the Eustigmatophyte *Nannochloropsis sp*. *J. Phycol.* **25,** 686–692 (1989).

62. Suda, S., Atsumi, M. & Miyashita, H. Taxonomic characterization of a marine *Nannochloropsis* species, N. oceanica sp. nov. (Eustigmatophyceae). *Phycologia* **41,** 273–279 (2002).

63. López-Rosales, A. R. *et al.* Fatty acids, hydrocarbons and terpenes of *Nannochloropsis* and Nannochloris isolates with potential for biofuel production. *Energies* **12,** (2019).

64. Radakovits, R. *et al.* Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropis gaditana*. *Nat. Commun.* **3,** 686 (2012).

65. Wei, L. *et al.* Transcriptomic and proteomic responses to very low CO2 suggest multiple carbon concentrating mechanisms in *Nannochloropsis oceanica*. *Biotechnol. Biofuels* **12,** 1–21 (2019).

66. Li, Y. *et al.* Effects of inorganic carbon and light on acetate assimilation by *Nannochloropsis oceanica* (Eustigmatophyceae) in mixotrophic cultivation. *Eur. J. Phycol.* **00,** 1–12 (2019).

67. Banerjee, A., Maiti, S. K., Guria, C. & Banerjee, C. Metabolic pathways for lipid synthesis under nitrogen stress in *Chlamydomonas* and *Nannochloropsis*. *Biotechnol. Lett.* **39,** 1–11 (2017).

68. Lubi, L. M. *et al. Nannochloropsis* (Eustigmatophyceae) as source of commercially valuable pigments. 249–255 (2000).

69. Fidyt, K., Fiedorowicz, A., Strządała, L. & Szumny, A. B-Caryophyllene and B-Caryophyllene Oxide—Natural Compounds of Anticancer and Analgesic Properties.

Cancer Med. **5,** 3007–3017 (2016).

70. Cho, K. S. *et al.* Terpenes from forests and human health. *Toxicol. Res.* **33,** 97–106 (2017).

71. Rufino, A. T. *et al.* Evaluation of the anti-inflammatory, anti-catabolic and pro-anabolic effects of E-caryophyllene, myrcene and limonene in a cell model of osteoarthritis. *Eur. J. Pharmacol.* **750,** 141–150 (2015).

72. Cai, Y. *et al.* A cDNA clone for beta-caryophyllene synthase from *Artemisia annua*. *Phytochemistry* **61,** 523–529 (2002).

73. Guillard, R. R. L. Culture of Phytoplankton for Feeding Marine Invertebrates. in *Culture of Marine Invertebrate Animals: Proceedings - 1st Conference on Culture of Marine Invertebrate Animals Greenport* (eds. Smith, W. L. & Chanley, M. H.) 29–60 (Springer US, 1975).

74. Berges, J. A., Franklin, D. J. & Harrison, P. J. Evolution of an artificial seawater medium: Improvements in enriched seawater, artificial water over the last two decades. *J. Phycol.* **37,** 1138–1145 (2001).

75. Lamesch, P. *et al.* The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* **40,** 1202–1210 (2012).

76. Berardini, T. Z. *et al.* The *Arabidopsis* information resource: Making and mining the 'gold standard' annotated reference plant genome. *Genesis* **53,** 474–485 (2015).

77. Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H. & Winther, O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33,** 3387–3395 (2017).

78. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2,** 953–971 (2007).

79. Bendtsen, J. D., Nielsen, H., Von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340,** 783–795 (2004).

80. Gruber, A., Rocap, G., Kroth, P. G., Armbrust, E. V. & Mock, T. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J.* **81,** 519–528 (2015).

81. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305,** 567–580 (2001).

82. Solovchenko, A. *et al.* Interactive effects of salinity, high light, and nitrogen starvation on fatty acid and carotenoid profiles in *Nannochloropsis oceanica* CCALA 804. *Eur. J. Lipid Sci. Technol.* **116,** 635–644 (2014).

83. Chen, J. W. *et al.* Identification of a malonyl CoA-acyl carrier protein transacylase and

its regulatory role in fatty acid biosynthesis in oleaginous microalga *Nannochloropsis oceanica*. *Biotechnol. Appl. Biochem.* **64,** 620–626 (2017).

84. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).

85. Carretero-paulet, L. *et al.* Expression and Molecular Analysis of the *Arabidopsis* DXR Gene Encoding 1-Deoxy-D-Xylulose 5-Phosphate Reductoisomerase, the First Committed Enzyme of the 2-C-Methyl-D-Erythritol 4-Phosphate Pathway. *Plant Physiol.* **129,** 1581–1591 (2002).

86. Phillips, M. A., León, P., Boronat, A. & Rodríguez-Concepción, M. The plastidial MEP pathway: unified nomenclature and resources. *Trends Plant Sci.* **13,** 619–623 (2008).

87. Hsieh, M. H. & Goodman, H. M. The *Arabidopsis* IspH homolog is involved in the plastid nonmevalonate pathway of isoprenoid biosynthesis. *Plant Physiol.* **138,** 641–653 (2005).

88. Tritsch, D., Hemmerlin, A., Bach, T. J. & Rohmer, M. Plant isoprenoid biosynthesis via the MEP pathway: In vivo IPP/DMAPP ratio produced by (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase in tobacco BY-2 cell cultures. *FEBS Lett.* **584,** 129–134 (2010).

89. Fabris, M. *et al.* Tracking the sterol biosynthesis pathway of the diatom *Phaeodactylum tricornutum*. *New Phytol.* **204,** 521–535 (2014).

90. Jinkerson, R. E., Radakovits, R. & Posewitz, M. C. Genomic insights from the oleaginous model alga *Nannochloropsis gaditana*. *Bioengineered* **4,** (2013).

91. Athanasakoglou, A. *et al.* Isoprenoid biosynthesis in the diatom *Haslea ostrearia*. *New Phytol.* **222,** 230–243 (2019).

92. Athanasakoglou, A. *et al. Isoprenoid biosynthesis in the diatom Haslea ostrearia*. *New Phytologist* (2018).

93. Schrader, J. *Biotechnology of Isoprenoids*. **148,** (2015).

94. Ramarajan, M., Fabris, M., Abbriano, R. M., Pernice, M. & Ralph, P. J. Novel endogenous promoters for genetic engineering of the marine microalga *Nannochloropsis gaditana* CCMP526. *Algal Res.* **44,** 101708 (2019).

95. Moog, D., Stork, S., Reislöhner, S., Grosche, C. & Maier, U. G. In vivo Localization Studies in the Stramenopile Alga *Nannochloropsis oceanica*. *Protist* **166,** 161–171 (2015).

96. Robert, C. A. M. *et al.* Genetically engineered maize plants reveal distinct costs and benefits of constitutive volatile emissions in the field. *Plant Biotechnol. J.* **11,** 628–639 (2013).

97. Scholz, M. J. *et al.* Ultrastructure and Composition of the *Nannochloropsis gaditana*

Cell Wall. **13,** 1450–1464 (2014).

98. Sadre, R. *et al.* Cytosolic lipid droplets as engineered organelles for production and accumulation of terpenoid biomaterials in leaves. *Nat. Commun.* **10,** 853 (2019).

99. Nobusawa, T., Hori, K., Mori, H., Kurokawa, K. & Ohta, H. Differently Localized Lysophosphatidic Acid Acyltransferases Crucial for Triacylglycerol Biosynthesis in the Oleaginous Alga *Nannochloropsis*. *Plant J.* 1–13 (2017).

100. Rohmer, M. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* **16,** 565–574 (1999).

101. Bach, T. J. & Rohmer, M. *Isoprenoid Synthesis in Plants and microorganisms*. (Springer, 2012).

102. Davies, F. K., Work, V. H., Beliaev, A. S. & Posewitz, M. C. Engineering limonene and bisabolene production in wild type and a glycogen-deficient mutant of *Synechococcus* sp. PCC 7002. *Front. Bioeng. Biotechnol.* **2,** 1–11 (2014).

103. Halfmann, C., Gu, L., Gibbons, W. & Zhou, R. Genetically engineering cyanobacteria to convert CO2, water, and light into the long-chain hydrocarbon farnesene. *Appl. Microbiol. Biotechnol.* **98,** 9869–9877 (2014).

104. Zhan, X., Zhang, Y. H., Chen, D. F. & Simonsen, H. T. Metabolic engineering of the moss *Physcomitrella patens* to produce the sesquiterpenoids patchoulol and α/β-santalene. *Front. Plant Sci.* **5,** 1–10 (2014).

105. Poliner, E., Farré, E. M. & Benning, C. Advanced genetic tools enable synthetic biology in the oleaginous microalgae *Nannochloropsis* sp. *Plant Cell Rep.* **37,** 1383–1399 (2018).

106. Nielsen, A. Z. *et al.* Extending the biosynthetic repertoires of cyanobacteria and chloroplasts. *Plant J.* **87,** 87–102 (2016).

CHAPTER 5

Synthesis and future prospects

Margaret Ramarajan[a], Michele Fabris[a,b], Peter J. Ralph[a]

[a] University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo NSW 2007, Australia

[b] CSIRO Synthetic Biology Future Science Platform, GPO box 2583, Brisbane QLD 4001, Australia

Author contribution: M.R. wrote the chapter; M.F. and P.J.R edited and reviewed the chapter.

5.1 Overview

The *E. coli* and yeast strains we see today are a product of genetic improvement and an array of genetic tools made available because of decades of work from multiple laboratories across the world. We stand today in the cusp of innovative biotechnology where development of genetic tools no longer takes decades but if such tools are made available for the microalga *Nannochloropsis gaditana*, strain improvement would be relatively easy. A fundamental need for transgene expression or over-expression studies in this alga required the identification of strong promoters. Several promoters have been identified and used for protein expression in *Nannochloropsis* spp.[1] but only three endogenous promoters were identified in *N. gaditana*[2]. This lead to the conception of this PhD project (2016) to characterise strong endogenous and exogenous promoters of various strength and inducible properties for transgene expression in this microalga This suite of promoters could lay the foundation to create an industrially useful chimeric promoter, similar to the promoters used in plant biotechnology (e.g. plant ubiquitin regulatory system-Monsanto UK Ltd[3]). The year following the start of my thesis, Ajjawi et al., 2017 reported the use of four new endogenous promoters for reporter/section marker expression in a gene repression study using CRISPR/Cas9 in *N. gaditana*[4]. Recently, Jackson et al., demonstrated the use of an inducible expression system using nitrate in N. *gaditana*[5] but no phosphate-inducible promoters have been characterised in spite the capacity of *N. gaditana* to cope with phosphorous (P) stress because of the presence of P reserves and an interplay of P recycling[6,7]. We have characterised three endogenous promoters (Chapter 2), one heterologous putative viral promoter (Chapter 3) from *Ectocarpus siliculosus* virus (EsV-1)[8] and two phosphate inducible promoters (Chapter 3). In Chapter 4, the best promoter candidate (*Hsp90*) was selected and used for a direct biotechnological application, to drive the expression of a terpene synthase gene for β-caryophyllene (plant terpenoid) production in *N. gaditana* using farnesyl diphosphate (FPP) as substrate. To the best of our knowledge, this represents a first, seminal proof-of-concept of the heterologous production of plant terpenoids in this microalgal species. The metabolic engineering strategy was based on a reconstructed isoprenoid metabolic pathway (Chapter 4, Figure 15,

and Table 7) in *N. gaditana* including their sub-cellular localisations *in-silico*, constructed using *N. gaditana* genomic data[2,9].

5.2 Synthesis and future prospects

Nannochloropsis spp. are oleaginous model microalgae. *Nannochloropsis* spp. are a relatively new platform that have the potential to be used as an industrially relevant production system through genetic engineering. In the past decade rapid advancement in the field of molecular biology have been made after i) the publication of the draft genome sequences of multiple species of *Nannochloropsis*[1–3]; ii) development of transformation protocols[1,2,4–10]; iii) availability of transcriptomic datasets under different conditions (continuous light, nitrogen depletion, phosphate depletion, varying light intensities and regimes)[2,11–14], which along with development of advanced genetic engineering tools can lead to developing *N. gaditana* into a chassis organism. In systems biology, chassis are organisms with simplified genome and a metabolic network for effective synthesis of products[15]; in synthetic biology, chassis are organisms that house and support genetic components and provide resources that allow them to function[15]. Well recognised chassis microorganisms such as *Escherichia coli* and *Saccharomyces cerevisiae* continue to dominate the field[16]. Synthetic biology has advanced from genetic mutations to designing complex genetic circuits in order to create artificial chromosomes and genomes[15,16]. In doing so, such simplified genomes and metabolic networks would be able to produce the desired product more efficiently[15,17,18]. Such systems allow a design, build and test approach with the help of advanced genetic tools and genome-wide data availability[19]. Photosynthetic chassis have recently gained significant interest such as the model organism *Chlamydomonas reinhardtii* and *Phaeodactylum tricornutum*[20] which offer sustainable production compared to heterologous hosts[21]. These photosynthetic production systems are productive sources of biomass[22,23]. A Modular Cloning toolkit for *C. reinhardtii* based on Golden Gate cloning system was used for building engineered cells[21]. A similar toolkit was built for cyanobacteria called CyanoGate[24] and a uLoop system for diatoms[25,26]. A minimal genome factory was launched in Japan to construct microorganisms for industrial use[15]. Further, development of metabolic maps and understanding the regulatory network via

RNA-seq and chromatin immunoprecipitation DNA sequencing will aid to develop such chassis organisms[19]. Toolkits to genetic engineer *Nannochloropsis* spp. are becoming publicly available that could lead to development of this oleaginous model species as a chassis organism[23] . Such a development could enable increased production of nutraceuticals such as omega 3 fatty acids (DHA and EPA) using this microalga which is naturally rich is LC-PUFAs[27]. The ever-increasing consumer demand is currently being meet by seafood industry which has undesirable effects such as overfishing and reduction of wild fish stocks[27]. Development of Nannochloropsis spp. into a chassis could greatly benefit the aquaculture, fisheries and nutraceutical industry[23,27]. Developments such as sequencing of the microalgal genome[2,28], availability of advanced genetic tools[19] and multi-fold gene stacking capacity[23,27] have facilitated pathway engineering[27,29]. The development and availability of modular promoters, untranslated regions, terminators, signal peptides for sub-cellular localisation, reporters would enable customise transgene expression in *Nannochloropsis* spp. in the near future.

As stated above, whether the goal is to build or refine metabolic pathway or understand the biology of the organism or develop chassis organisms, several molecular tools are needed. Tools to enable high-capacity gene stacking systems[23], marker-free knockout, mutant libraries, transgenic over-expression are vital[19]. All of which require a robust array of promoters that can be exploited. A few promoter candidates are available for this microalga[19]. In this PhD study, three endogenous constitutive promoters (*EPPSII*, *Nga02101; HSP90, Nga00934; ATPase, Nga06354.1)* were successfully characterised (Chapter 2). The promoter activity of two of the characterised promoters was significantly higher compared to the widely used β-tubulin promoter[1,19] while the third promoter had a similar promoter activity to β-tubulin promoter[1,19]. Having promoters of variable strengths is beneficial for fine-tuning the expression of transgenes and entire metabolic pathways and are also useful in avoiding gene silencing issues due to promoter repetition[30–32].

Constitutive promoters have shown to be very advantageous for expression of transgene, but may not be suitable for expressing highly deleterious or lethal/toxic

genes[33,34] and genes that could affect or alter the metabolic or developmental pathways that inhibit the growth of the host[35]. In such cases, inducible promoters are handy tools, hence focused on characterising phosphate-inducible promoters (Chapter 3), where a couple of putative phosphate-inducible genes were identified (*SPS, EWM20309.1; GDPD, EWM28518.1*) and the promoters tested for induction under phosphate limitations. Even though the pilot experiments produced promising results, the final transgenic lines were not stable, and confirmation of these promoter targets was not possible.

Strain instability is common in transgenic lines which has previously been observed in model algal species such as *Chlamydomonas*[36,37], *Phaeodactylum*[38,39], *Chlorella*[40] and mosses such as *Physcomitrella*[41]. Strain instability is also evident in transgenic plants and has been attributed to heterologous constitutive promoter expression[42], position effect[43–45] and epigenetic transgene suppression[46]. The random integration of the transgenic DNA in a variety of arrangements and in multiple copies is responsible for position effect and epigenetic suppression[47–49]. Initially transgene copy number was estimated using techniques such as Southern blotting by selecting probes that would bind to heterologous transgene DNA frangments[4,6,50] and subsequently quantitative real time PCR was used to identify the transgene copy number[51]. These techniques do not provide information whether the transgene copy is an intact cassette or truncated[52–54]. Recently, DNA sequencing technology has revealed that transgene integrations can be messy with multiple copies of the transgene re-arranged in different orientations forming integration islands[47]. Still the data cannot identify whether intact transgene copies could code for the desired protein as transcription depends on the presence of upstream regulatory elements including promoters[45,55–57]. Hence presence of multiple copies of the transgene in a transgenic line does not translate to high protein levels[55]. Reverse transcriptase-qPCR[52] and Nano-string[52] could be used to estimate the relative amount of transgenic mRNA but is not an accurate estimate of the intact transgene copies as mRNA levels fluctuate depending on several factors or if a transgene is suppressed or is transcriptionally inactive due to site of integration[43,44,58,59]. Hence the genotype of a transgenic cell line can tell us the intact copy number[60] of the heterologous transgene but cannot be corelated to protein levels

obtained[52]. Targeted transgene integration on the other hand would allow a controlled spatio-temporal expression of the desired transgene via homologous recombination or using site specific integration using integrases[9,61,62].

Later in the same chapter, a heterologous putative viral promoter (Chapter 3) from *Ectocarpus siliculosus* virus (EsV-1)[63] was tested for promoter activity in *N. gaditana* and *P. tricornutum* with the expectation of identifying a universal viral promoter for use in stramenopiles. The viral promoter successfully drove transgene expression in *N. gaditana,* but no expression was observed in *P. tricornutum.* While the data collection for Chapter 3 was in progress, an inducible expression system using nitrate reductase promoter was characterised by Jackson et al., (2018) for use in *N. gaditana*[64]. The suite of promoters characterised in this PhD study can be used for the development of a variety of molecular tools described below:

5.2.1 Evaluation of promoter activity under different growth and nutrient conditions to identify optimal working range and functional environments

As stated in Chapter 2, the constitutive promoters were profiled in batch mode under continuous light and nutrient conditions. In general, the heat shock proteins (HSPs) are molecular chaperones functionally involved in fundamental cellular processes and are induced by heat shock and other stimulants such as heavy metals, nitric oxide, hormones, microbial infections etc[65]. Similarly, *EPPSII* (*PsbQ*) is a part of the oxygen-evolving complex of photosystem II (*PSII*) and is required for the PSII reaction centre accumulation under low light[66,67]. Stephanou et al., demonstrated that the mammalian *Hsp-70* and *Hsp-90* promoters had enhanced activity when treated with interferon-γ[68]. Chang et al., found that okadaic acid synergistically acted on heat-induced human Hsp70 promoter activity[69].Park et al., highlighted the significance of *light-inducible protein (LIP)* gene promoter of the marine alga *Dunaliella* sp., which was introduced into *C. reinhardtii* and could be induced by high light (over 500 µmol photons· m^{-2} s^{-1}) in a light intensity-dependent manner[70]. Wang et al., demonstrated the β-*carotene ketolase (BKT)* gene promoter to be regulated by sodium acetate and light in *C. reinhardtii*[71]. Similarly, the promoter activity of these

promoters (*EPPSII, Nga02101; HSP90, Nga00934)* maybe variable (up or downregulated) under light, temperature, or heavy metal stress in *N. gaditana* (Figure 23). These different growth conditions can be tested in the future, to identify a more suitable set of conditions to use these promoters under for maximum transgene expression.

5.2.2 Construction of promoter libraries for characterising transcription factor binding domains and obtain an array of promoters of varying promoter strengths

The metabolic flux in a eukaryotic cell is delimited by a distinct series of regulatory controls at the transcriptional, translational, and protein levels[72]. The axiom that gene promoter can alter this flux at the transcription level has been widely exploited[73]. Synthetic biology, genetic and metabolic engineering strategies rely strongly on promoter detection and characterisation, as a wide range of gene expression capabilities are required[73]. In such application, strong yet tightly control promoters are wells suited to maximise protein production with low toxicity or metabolic stress to the cell[74]. Hence promoter engineering endeavours to modulate promoter transcriptional strength by mutating, enhancing, or otherwise altering the sequence of the promoter[73]. For example, in yeast, for pathway engineering for arteminisic acid[75] and other isoprenoids[76,77], the GAL promoters (a synthetic hybrid promoter system generated by the fusion of Gal4p-binding sites to the minimal core promoter sequence[78]) were used.

Strong over-expression is not always the answer for all genes, some may require low expression for optimal metabolite production, and therefore an array of promoters of variable strengths and capacities is essential. For this purpose, promoter libraries are created using i) random mutagenesis (error-prone PCR, for example, in *E. coli*, a promoter library of engineered promoters of varying strengths was created using mutagenesis)[79]; ii) hybrid promoter engineering, for example, in a yeast *Yarrowia lipolytica*, four hybrid promoter systems were created using copies of tandem repeats (upstream enhancer elements) and core promoter[80]. Promoter libraries can consist of artificial (synthesised fragments) promoters[81], mutated native promoters or enhancers[82,83] or intrinsic promoters[84,85].

In microalgae, specifically *Nannochloropsis spp*, such promoter libraries can be created using the suite of available promoters and the ones characterised in this study. These promoter libraries can be used to identify or build industrially relevant promoters of variable strengths and transcriptional regulation[86] (Figure 23). Generating promoter libraries would also assist in the detection and characterisation of transcription factor binding sites, to increase or decrease the promoter activity of known promoters[87–89]. In the future, high-throughput characterisation of transcription factor binding sites through single nucleotide polymorphisms (SNP)[90,91] and analysis of promoters at single-cell level on a microfluidic array[92]. Also, biochemical determination of promoters based on protein-DNA interaction using ChIP (chromatin immunoprecipitation)-chip[93,94] will further our understanding of eukaryotic gene promoters in microalgae and such data could be utilised to construct predictive models of promoter function[73] specific for microalgae. Single cell analysis techniques have been heralded as the next frontier of omics technology and such analysis potentially could identify new and unique information from cellular heterogeneity[95].For example, the single cell RNA-sequencing technologies could profile the genetic, epigenetic, proteomic and hereditary information from individual cells[96,97]

5.2.3 Gene stacking tools for pathway engineering in *N. gaditana*

Pathway engineering often requires the expression of multiple transgenes[23]. Modular characterised toolkits are needed and are available for model organisms such as *E. coli*[98], yeast[99], mammalian cells[100], higher plants[101], cyanobacteria[102], microalgae *C. reinhardtii*[21] and *N.oceanica*[23]. Several multi-gene expression vectors are available for metabolic engineering in plants and microalgae[27,103,104]. Pathway engineering for the biosynthesis of long- chain polyunsaturated fatty acids (LC-PUFAs)[104–108] and isoprenoids[51,109] have been demonstrated in microalgae including advanced engineering using CRISPR/Cas9 and Cre recombinase in *N. gaditana*[110]. Also, the episome-based, non-integrative expression of transgenes has been demonstrated to be feasible with *Nannochloropsis spp*[110,111]. Episomes are extrachromosomal DNA with centromere like region which have been shown to successfully replicate and segregate into *Phaeodactylum* daughter cells[112,113]. Episomes can be successfully introduced into microalgal hosts such as

P. tricornutum, *T. pseudonana*, *N. oceanica* using bacterial conjugation which is more efficient compared to electroporation and biolistic delivery[111,113]. These extrachromosomal vectors allow transgene expression efficiently and consistently by avoiding problems brought about by random chromosomal integration via biolistic or electroporation such as multiple fragment or partial expression cassette insertion, disruption of endogenous genes, variability in transgene expression due to the influence of position effect[47,112–114]. These episomes have shone to be faithfully maintained in the nucleus of the host without chromosomal integration[47,111,112]. In diatoms, episomal gene expression has been more uniform between independent transgenic lines when compared to random integration possibly by avoiding position effects[47,112,113]. George et al., compared in *P. tricornutum* the profound differences observed in the phenotype of cells transformed via episomal and random chromosomal integration[47]. The difference observed were related to transgene expression levels and sub-clonal population make-up[47] where transformants with episomes exhibited consistent phenotypes compared to transformants obtained via random integration which exhibited high degree of variability but overall higher transgene expression levels due to the fact that such lines carried highly concatenated transgenic DNA insertion as large gene islands within actively transcribed chromosomal regions[47]. Metabolic engineering in microalgal species could combine the reproducibility of episomal transformation and high transgene expression of random integration by utilizing site-directed chromosomal integration[9,23,47]. Such tools are coupled with effective promoters of varying strengths and capacities to enable repeated changes to biosynthetic pathways with several transgenes[23]. The promoters characterised in this study can be efficiently used in such gene stacking systems for pathway/ metabolic engineering of *N. gaditana.*

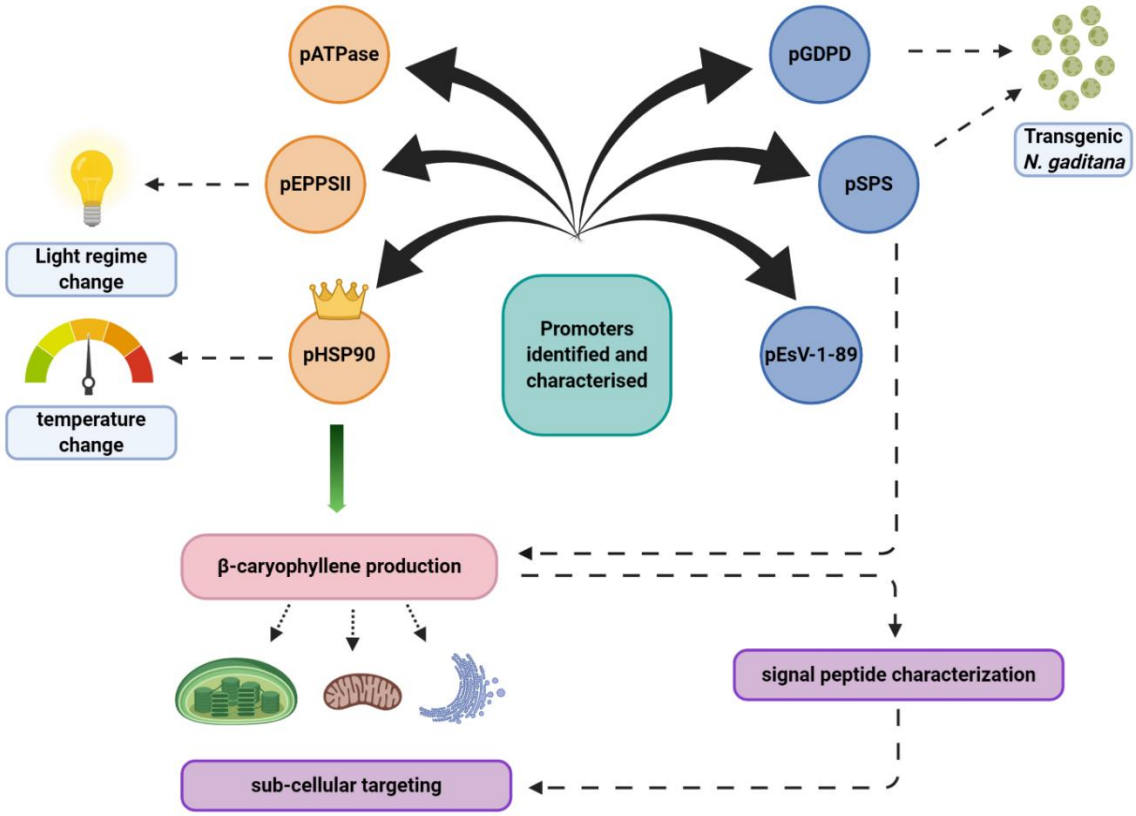


Figure 23. Schematic representation of the thesis overview.

Three endogenous constitutive promoters (orange circles; *EPPSII*, *Nga02101; HSP90, Nga00934; ATPase, Nga06354.1*) characterised in Chapter 2, two P-inducible (maybe) promoters (blue circles; *SPS, EWM20309.1; GDPD, EWM28518.1*) and one viral promoter (blue circle; EsV-1-89; NCBI gene ID: 920795) characterised in Chapter 3. The best promoter (Hsp90) used to drive the β-caryophyllene synthase (*QHS1*) gene from *Artemisia annua L* for β-caryophyllene biosynthesis (pink oblong, Chapter 4). Solid arrows- work completed in this PhD study. Dotted arrows- future prospects.

5.2.4 *N. gaditana* as a biofuel production platform

Nannochloropsis are a potential feedstock for fuels and high-value products because they can tolerate wide environmental and culture conditions, grow rapidly and produce substantial amounts of triacylglycerols (TAG) and eicosapentaenoic acid (EPA)[115]. *N. gaditana* can reach high cell densities (> 10 g L^{-1}) and tolerate fluctuations in pH, temperature and salinity[1]. *N. gaditana* cultures grown in F/2 medium yield approximately 0.65 g L^{-1} d^{-1} biomass and 0.31 g L^{-1} d^{-1} total lipids with a lipid content of 47.5 %[1] and EPA

around 4% of dry biomass[116]. Rodolfi et al., attained 60% lipid content after nitrogen starvation in *Nannochloropsis* sp. using 20 L flat alveolar panel photobioreactor[117] and Xiao et al., observed a significant increase in lipid content from 44.8% (Day 7-exponential phase) to 69.1% (Day 23- deceleration phase) in *N. oceanica* strain IMET1[118].

Nannochloropsis spp. genome has higher gene enrichment for cellular lipid metabolism than in *C. reinhardtii*[3] and play an important role in shifting carbon fluxes toward TAG metabolism. For example, there are 11 copies of diacylglycerol acyltransferase-2 (DGAT-2) genes in the genome of *Nannochloropsis* sp. compared to four DGAT-2 copies in *Thalassiosira pseudonana*, *P. tricornutum* and five DGAT-2 copies in *C. reinhardtii*[3,13]. Such high gene dose in *Nannochloropsis sp.* likely originated from ancient genomes of secondary endosymbiosis host and the engulfed green and red algae[3]. Multiple-genome pooling and horizontal genetic exchange from bacteria, collectively with the selective inheritance of lipid synthesis genes have formed a massive genetic organisation for oleaginousness witnessed among *Nannochloropsis* sp[3] and therefore an appropriate candidates for development into a model organism for algal biofuel production especially *N. gaditana*, which boasts a higher lipid production rate compared to other *Nannochloropsis* species[1].

Manipulation of metabolic pathways for lipid synthesis has been attempted in *Nannochloropsis* sp.[119] Iwai et al., reported triacylglycerols accumulation under phosphate starvation–dependent over-expression of *C. reinhardtii* type-2 *diacylglycerol acyl-CoA acyltransferase* (*XM_001693137.1*) using a *C. reinhardtii sulfoquinovosyldiacylglycerol synthase 2* (*SQD2, XM_001689610.1*) promoter in *Nannochloropsis* strain NIES-2145[120,121]. Similarly, the putative phosphate-inducible promoters, which were identified (*SPS, EWM20309.1; GDPD, EWM28518.1*) in Chapter 3 could serve well in expressing transgenes or over-expressing host lipid synthesis genes in *N. gaditana* (after confirmation of promoter activity under phosphate limitation in the near future). Furthermore, over-expression studies have been carried out in *N. oceanica* and *N. salina* to express transcription factors such as basic helix loop helix 2, basic leucine zip- per domain 1, *Arabidopsis thaliana* WRIN-KLED1 under the control of endogenous β-tubulin (*TUB*, *Nga00092*) and the ubiquitin

extension protein (*UEP, Nga02115.1*) for the regulation of lipid metabolic pathway to improve lipid production[122–124]. Two of the endogenous promoters characterised in Chapter 2 (*EPPSII*, *Nga02101; HSP90, Nga00934*) would perform better than the β-tubulin promoter in over-expressing the above specified transcription factors in *N. gaditana* to improve lipid biosynthesis.

5.2.5 The various avenues to explore to develop *Nannochloropsis* into a terpenoid production host

The 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway is responsible for isoprenoid biosynthesis in *N. gaditana*[1,125]. This pathway is responsible for the production of pigments[126], sterols[127], phytol[128] and fatty acids[128–130]. *N. gaditana* inherited the MEP genes via cyanobacterial endosymbiosis and the genes subsequently transferred to the nuclear genome[125,131,132]. This MEP pathway has never been described in details yet, thus in the course of this work was re-constructed *in-silico* using *A. thaliana* MEP pathway enzymes as reference, and their putative locations inferred using specific algorithms such as DeppLoc[133] and ASAFind[134,135]. The complete re-construction of the pathway model allowed the identification of all enzymatic steps and key substrate hubs for terpenoid engineering, such as prenylphosphates biosynthesis. This aided in developing a preliminary and exploratory metabolic engineering strategy for a cytosolic production of β-caryophyllene, a commercially relevant sesquiterpene, hypothesizing the presence of a FPP pool in the cytoplasm of *N. gaditana*.

Therefore, in Chapter 4, an endogenous constitutive (*HSP90, Nga00934)* promoter with highest promoter activity at early to mid-log phase (characterised in Chapter 2)was used to drive the β-caryophyllene synthase (*QHS1*)[136] gene from *Artemisia annua L* (sweet wormwood). No phosphate-inducible promoters were confirmed (Chapter 3) that could have been used to test β-caryophyllene production under the control of an inducible promoter. The amount of β-caryophyllene produced by the *N. gaditana* strain when targeted to the cytoplasmic was minimal, but there is room for strain enhancement. i) **Metabolic flux**: Basic understanding of the metabolic network and carbon flux regulatory

mechanisms[137] of *Nannochloropsis* would assist in identifying gene targets that could be under or over-expressed to boost the flux through the desired pathway for better accumulation of enzyme substrates and down-regulation of competing pathways without compromising the performance of the microalgal host; ii) **Terpenoid trap:** Sadre et al., highlighted the function of cytosolic lipid droplets as traps for the terpenoids in the cells of *Nicotiana benthamiana* leaves[138]. They showed that directing the biosynthetic steps onto the surface of lipid droplets lead to efficient production of functionalised terpenoids[138]. As *N. gaditana* produces large lipid droplets in the cytosol[139], these could serve as a great advantage and be used as natural traps to catch heterologous lipophilic plant isoprenoids within the cell without requiring solvent-culture overlays such as dodecane and isopropyl myristate; iii) **Chloroplast transformation**: to date, no successful protocols for chloroplast transformation exist for *Nannochloropsis.* Chloroplastic expression has advantages due to high-expression levels than nuclear expression, targeted insertion by homologous recombination, and protection from protein degradation due to the presence of chloroplast envelope[140]; iv) **Targeting signal peptide**: the MEP pathway is in the chloroplast and based on the hypothetical model (described in Chapter 4, section 4.3.1) for isoprenoid biosynthesis a dynamic substrate (FPP and GGPP) pool would be available in the chloroplast. Specific chloroplast targeting signal peptides are required to target the terpene synthases to the chloroplast. Availability of characterised signal peptide sequences for sub-cellular targeting which are already available for other model organisms[141–144] will be advantageous to design and execute future metabolic engineering experiments in *N. gaditana*; v) **Terpenoid bio-synthesis pathway**: the identification and characterisation of prenyltransferases in *N. gaditana* would shed light on their substrate-specificity, kinetics and localisation of these enzymes in the cell, which along with the size of the substrate (GPP, FPP and GGPP) pool availability and location can greatly benefit researchers in developing strategies to target the heterologous terpene synthase enzymes to the ideal sub-cellular locations.

The above steps would not just aid in boosting β-caryophyllene production, but an array of commercially viable terpenoids using this in this promising microalga as a photosynthetic GM host.

5.2.6 *Nannochloropsis* taught to use sugars like yeast systems

Media used for cultivation can have drastic impact on the microalgal biomass[145,146]. The three major microalgal cultivation modes are phototrophic, heterotrophic and mixotrophic cultures[147]. *Nannochloropsis* is a photoautotroph[148] but studies have shown that it can to a lesser extent maintain a mixotrophic culture in the presence of glucose, glycerol and lactic acid[149,150]. Also, the hexose symporter gene from a microalgae or yeast could be engineered into *N. gaditana* for uptake of glucose/galactose/sucrose, for example, The HUP1 from *Chlorella kessleri* was introduced into *C. reinhardtii* which resulted in a mutant that could use externally supplied glucose for hydrogen production[151–153]. A mixotrophic mode of cultivation is a good strategy that combines the advantages of photo and heterotrophy to achieve higher growth rate and biochemical accumulation[154]. For example, *Chlorella vulgaris* cultures had a higher growth rate in mixotrophic culture medium with glycerol[155]; In *Haematococcus pluvialis* cultures, glycerol addition to growth media accelerate astaxanthin production under low light due to increase in pyruvate[156]. Also, studies on *Chlorella protothecoides* showed that under mixotrophic cultivation, 57% enhancement of lipid content was observed[157]. It opens new perspectives on future strategies to exploit *N. gaditana*, an industrially relevant microalga.

5.3 Concluding remarks

Decades of molecular work has promoted this microalga as a prospective contender for future industrial use. The wealth of scientific data available in the genome, transcriptome and proteome level for this microalga has further aided the research community to develop valuable genetic engineering and genome editing tools. The novel promoter tools identified and characterised in this study has broadened the repertoire of promoters. These tools can be used but not limited to, understanding the biology of *Nannochloropsis*, for the metabolic engineering of this species and for overexpression of

heterologous protein in *N. gaditana*. Our research is the first to use *N. gaditana* as a host for terpenoid production and highlights that alternate expression systems like this marine microalga can effectively supplement the growing demand for plant based terpenoid products.

5.4 References

1. Radakovits, R. *et al.* Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropis gaditana*. *Nat. Commun.* **3**, 686 (2012).

2. Vieler, A. *et al.* Genome, Functional Gene Annotation, and Nuclear Transformation of the Heterokont Oleaginous Alga *Nannochloropsis oceanica* CCMP1779. *PLoS Genet.* **8**, (2012).

3. Wang, D. *et al. Nannochloropsis* Genomes Reveal Evolution of Microalgal Oleaginous Traits. **10**, (2014).

4. Li, F., Gao, D. & Hu, H. High-efficiency nuclear transformation of the oleaginous marine *Nannochloropsis* species using PCR product. *Biosci. Biotechnol. Biochem.* **78**, 812–817 (2014).

5. Cha, T. S., Chen, C. F., Yee, W., Aziz, A. & Loh, S. H. Cinnamic acid, coumarin and vanillin: Alternative phenolic compounds for efficient *Agrobacterium*-mediated transformation of the unicellular green alga, Nannochloropsis sp. *J. Microbiol. Methods* **84**, 430–434 (2011).

6. Ma, X. *et al.* Genetic transformation of *Nannochloropsis oculata* with a bacterial phleomycin resistance gene as dominant selective marker. *J. Ocean Univ. China* **15**, 351–356 (2016).

7. Moog, D., Stork, S., Reislöhner, S., Grosche, C. & Maier, U. G. In vivo localization studies in the stramenopile alga *Nannochloropsis oceanica*. *Protist* **166**, 161–171 (2015).

8. Kang, N. K. *et al.* Heterologous overexpression of sfCherry fluorescent protein in *Nannochloropsis salina*. *Biotechnol. Reports* **8**, 10–15 (2015).

9. Kilian, O., Benemann, C. S. E., Niyogi, K. K. & Vick, B. High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis sp. Proc. Natl. Acad. Sci.* **108**, 21265–21269 (2011).

10. Chen, H. L., Li, S. S., Huang, R. & Tsai, H. J. Conditional production of a functional fish growth hormone in the transgenic line of *Nannochloropsis oculata* (Eustigmatophyceae). *J. Phycol.* **44**, 768–776 (2008).

11. Corteggiani Carpinelli, E. *et al.* Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion. *Mol. Plant* **7**, 323–335 (2014).

12. Poliner, E. *et al.* Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1779 under light / dark cycles. **83**, 1097–1113 (2015).

13. Alboresi, A. *et al.* Light remodels lipid biosynthesis in *Nannochloropsis gaditana* by modulating carbon partitioning between organelles. *Plant Physiol.* **171**, 2468–2482 (2016).

14. Mühlroth, A. *et al.* Mechanisms of phosphorus acquisition and lipid class remodelling under P limitation in a marine microalga. *Plant Physiol.* **175**, 1543–1559 (2017).

15. Chi, H. *et al.* Engineering and modification of microbial chassis for systems and synthetic biology. *Synth. Syst. Biotechnol.* **4**, 25–33 (2019).

16. Adams, B. L. The Next Generation of Synthetic Biology Chassis: Moving Synthetic Biology from the Laboratory to the Field. *ACS Synth. Biol.* **5**, 1328–1330 (2016).

17. Zhang, L. yan, Chang, S. hua & Wang, J. How to make a minimal genome for synthetic minimal cell. *Protein Cell* **1**, 427–434 (2010).

18. Vickers, C. E., Blank, L. M. & Krömer, J. O. Grand Challenge Commentary: Chassis cells for industrial biochemical production. *Nat. Chem. Biol.* **6**, 875–877 (2010).

19. Poliner, E., Farré, E. M. & Benning, C. Advanced genetic tools enable synthetic biology in the oleaginous microalgae *Nannochloropsis* sp. *Plant Cell Rep.* **37**, 1383–1399 (2018).

20. Fabris, M. *et al.* Emerging Technologies in Algal Biotechnology: Toward the Establishment of a Sustainable, Algae-Based Bioeconomy. *Front. Plant Sci.* **11**, 1–22 (2020).

21. Crozet, P. *et al.* Birth of a Photosynthetic Chassis: A MoClo Toolkit Enabling Synthetic Biology in the Microalga *Chlamydomonas reinhardtii*. *ACS Synth. Biol.* **7**, 2074–2086 (2018).

22. Rodolfi, L. *et al.* Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnol. Bioeng.* **102**, 100–112 (2009).

23. Poliner, E., Clark, E., Cummings, C., Benning, C. & Farre, E. M. A high-capacity gene stacking toolkit for the oleaginous microalga, *Nannochloropsis oceanica* CCMP1779. *Algal Res.* **45**, 101664 (2020).

24. Vasudevan, R. *et al.* Cyanogate: A modular cloning suite for engineering cyanobacteria based on the plant moclo syntax. *Plant Physiol.* **180**, 39–55 (2019).

25. Pollak, B. *et al.* Loop Assembly : a simple and open system for re- cursive fabrication of DNA circuits. (2018).

26. Pollak, B. *et al.* Universal loop assembly: Open, efficient and cross-kingdom DNA fabrication. *Synth. Biol.* **5**, (2019).

27. Poliner, E. *et al.* A toolkit for *Nannochloropsis oceanica* CCMP1779 enables gene

stacking and genetic engineering of the eicosapentaenoic acid pathway for enhanced long-chain polyunsaturated fatty acid production. *Plant Biotechnol. J.* **16**, 298–309 (2018).

28. Jinkerson, R. E., Radakovits, R. & Posewitz, M. C. Genomic insights from the oleaginous model alga *Nannochloropsis gaditana*. *Bioengineered* **4**, (2013).

29. Poliner, E., Takeuchi, T., Du, Z.-Y., Benning, C. & Farré, E. M. Nontransgenic Marker-Free Gene Disruption by an Episomal CRISPR System in the Oleaginous Microalga, *Nannochloropsis oceanica* CCMP1779. *ACS Synth. Biol.* **7**, 962–968 (2018).

30. Wilde, C. De *et al.* Plants as bioreactors for protein production : avoiding the problem of transgene silencing. *Plant Mol. Biol.* **43**, 347–359 (2000).

31. Matzke, A. J. M., Neuhuber, F., Park, Y. D., Ambros, P. F. & Matzke, M. A. Homology-dependent gene silencing in transgenic plants: epistatic silencing loci contain multiple copies of methylated transgenes. *Mol. Gen. Genet.* **244**, 219–229 (1994).

32. Y.-D. Park *et al.* Gene silencing mediated by promoter homology occurs at the level of transcription and results in meiotically heritable alterations in methylation and gene activity. *Plant J.* **9**, 183–194 (1996).

33. Giacalone, M. J. *et al.* Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system. *Biotechniques* **40**, 355–364 (2006).

34. Massie, B. *et al.* Inducible overexpression of a toxic protein by an adenovirus vector with a tetracycline-regulatable expression cassette. *J. Virol.* **72**, 2289–2296 (1998).

35. Corrado, G. & Karali, M. Inducible gene expression systems and plant biotechnology. *Biotechnol. Adv.* **27**, 733–743 (2009).

36. Fuhrmann, M., Hausherr, A., Ferbitz, L., Schödl, T. & Hegemann, P. Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. *Plant Mol. Biol.* **55**, 869–881 (2004).

37. Neupert, J., Karcher, D. & Bock, R. Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *Plant J.* **57**, 1140–1150 (2009).

38. Niu, Y. F. *et al.* Transformation of diatom *Phaeodactylum tricornutum* by electroporation and establishment of inducible selection marker. *Biotechniques* **52**, 1–3 (2012).

39. Siaut, M. *et al.* Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* **406**, 23–35 (2007).

40. Lau, C. C., Loh, S. H., Aziz, A. & Cha, T. S. Effects of disrupted omega-3 desaturase gene construct on fatty acid composition and expression of four fatty acid biosynthetic genes in transgenic *Chlorella vulgaris*. *Algal Res.* **26**, 143–152 (2017).

41. Kaewsuwan, S., Bunyapraphatsara, N., Cove, D. J., Quatrano, R. S. & Chodok, P. High level production of adrenic acid in *Physcomitrella patens* using the algae *Pavlova* sp. Δ5-elongase gene. *Bioresour. Technol.* **101**, 4081–4088 (2010).

42. Ho, M. W., Ryan, A. & Cummins, J. Cauliflower mosaic viral promoter - A recipe for disaster? *Microb. Ecol. Health Dis.* **11**, 194–197 (1999).

43. Feng, Y.-Q., Lorincz, M. C., Fiering, S., Greally, J. M. & Bouhassira, E. E. Position Effects Are Influenced by the Orientation of a Transgene with Respect to Flanking Chromatin. *Mol. Cell. Biol.* **21**, 298–309 (2001).

44. Clark, A. J. *et al.* Chromosomal position effects and the modulation of transgene expression. *Reprod. Fertil. Dev.* **6**, 589–594 (1994).

45. Peach, C. & Velten, J. Transgene expression variability (position effect) of CAT and GUS reporter genes driven by linked divergent T-DNA promoters. *Plant Mol. Biol.* **17**, 49–60 (1991).

46. Cerutti H, Johnson, A., Gillham, N. & Boynton, J. Epigenetic Silencing of a Foreign Gene in Nuclear Transformants of *Chlamydomonas*. *Plant Cell* **9**, 925–945 (2002).

47. George, J. *et al.* Metabolic Engineering Strategies in Diatoms Reveal Unique Phenotypes and Genetic Configurations With Implications for Algal Genetics and Synthetic Biology. *Front. Bioeng. Biotechnol.* **8**, 1–19 (2020).

48. Huang, W. & Daboussi, F. Genetic and metabolic engineering in diatoms. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, (2017).

49. Gangl, D. *et al.* Biotechnological exploitation of microalgae. *J. Exp. Bot.* **66**, 6975–6990 (2015).

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

51. D'Adamo, S. *et al.* Engineering the unicellular alga P*haeodactylum tricornutum* for high-value plant triterpenoid production. *Plant Biotechnol. J.* **17**, 75–87 (2019).

52. Chen, W. & Wang, P. Molecular Analysis for Characterizing Transgenic Events. in *Transgenic Plants* 397–410 (2019). doi:10.1007/978-1-4939-8778-8_24

53. Ingham, D. J., Beer, S., Money, S. & Hansen, G. Quantitative real-time PCR assay for determining transgene copy number in transformed plants. *Biotechniques* **31**, 132–140 (2001).

54. Bubner, B., Gase, K. & Baldwin, I. T. Two-fold differences are the detection limit for determining transgene copy numbers in plants by real-time PCR. *BMC Biotechnol.* **4**, 1–11 (2004).

55. Meyer, P. Understanding and controlling transgene expression. *Trends Biotechnol.*

13, 332–337 (1995).

56. Martin, D. I. K. & Whitelaw, E. The vagaries of variegating transgenes. *BioEssays* **18**, 919–923 (1996).

57. Fiering, S., Whitelaw, E. & Martin, D. I. K. To be or not to be active: The stochastic nature of enhancer action. *BioEssays* **22**, 381–387 (2000).

58. al-Shawi, R., Kinnaird, J., Burke, J. & Bishop, J. O. Expression of a foreign gene in a line of transgenic mice is modulated by a chromosomal position effect. *Mol. Cell. Biol.* **10**, 1192–1198 (2015).

59. Robertson, G. *et al.* Position-dependent variegation of globin transgene expression in mice. *Proc. Natl. Acad. Sci.* **92**, 5371–5375 (2006).

60. Liu, J. *et al.* Genome-Scale Sequence Disruption Following Biolistic Transformation in Rice and Maize. *Plant Cell* **31**, 368–383 (2019).

61. Roberts, J. A. *et al.* Targeted transgene integration overcomes variability of position effects in zebrafish. *Development* **141**, 715–724 (2014).

62. Lauersen, K. J., Kruse, O. & Mussgnug, J. H. Targeted expression of nuclear transgenes in *Chlamydomonas reinhardtii* with a versatile, modular vector toolkit. *Appl. Microbiol. Biotechnol.* **99**, (2015).

63. Delaroque, N., Wolf, S., Muller, D. G. & Knippers, R. The brown algal virus EsV-1 particle contains a putative hybrid histidine kinase. *Virology* **273**, 383–390 (2000).

64. Jackson, H. O. *et al.* An inducible expression system in the alga *Nannochloropsis gaditana* controlled by the nitrate reductase promoter. *J. Appl. Phycol.* **31**, 269–279 (2018).

65. Kiang, J. G. & Tsokos, G. C. Heat shock protein 70 kDa: Molecular biology, biochemistry, and physiology. *Pharmacol. Ther.* **80**, 183–201 (1998).

66. Bricker, T. M. & Frankel, L. K. Auxiliary functions of the PsbO, PsbP and PsbQ proteins of higher plant Photosystem II: A critical analysis. *J. Photochem. Photobiol. B Biol.* **104**, 165–178 (2011).

67. Roose, J. L., Frankel, L. K., Mummadisetti, M. P. & Bricker, T. M. The extrinsic proteins of photosystem II: update. *Planta* **243**, 889–908 (2016).

68. Stephanou, A., Isenberg, D. A., Nakajima, K. & Latchman, D. S. Signal transducer and activator of transcription-1 and heat shock factor-1 interact and activate the transcription of the Hsp-70 and Hsp-90β gene promoters. *J. Biol. Chem.* **274**, 1723–1728 (1999).

69. Chang, N. T., Huang, L. E. & Liu, A. Y. C. Okadaic acid markedly potentiates the heat-induced hsp 70 promoter activity. *J. Biol. Chem.* **268**, 1436–1439 (1993).

70. Park, S., Lee, Y., Lee, J. H. & Jin, E. S. Expression of the high light-inducible *Dunaliella* LIP promoter in *Chlamydomonas reinhardtii*. *Planta* **238**, 1147–1156 (2013).

71. Wang, C. *et al.* A β-carotene ketolase gene (bkt1) promoter regulated by sodium acetate and light in a model green microalga *Chlamydomonas reinhardtii*. *Algal Res.* **20**, 61–69 (2016).

72. Blazeck, J. & Alper, H. Systems metabolic engineering: Genome-scale models and beyond. *Biotechnol. J.* **5**, 647–659 (2010).

73. Blazeck, J. & Alper, H. S. Promoter engineering: Recent advances in controlling transcription at the most fundamental level. *Biotechnol. J.* **8**, 46–58 (2013).

74. Terpe, K. Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **72**, 211–222 (2006).

75. Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).

76. Shiba, Y., Paradise, E. M., Kirby, J., Ro, D. K. & Keasling, J. D. Engineering of the pyruvate dehydrogenase bypass in *Saccharomyces cerevisiae* for high-level production of isoprenoids. *Metab. Eng.* **9**, 160–168 (2007).

77. Maury, J. *et al.* Reconstruction of a bacterial isoprenoid biosynthetic pathway in *Saccharomyces cerevisiae*. *FEBS Lett.* **582**, 4032–4038 (2008).

78. Blazeck, J., Garg, R., Reed, B. & Alper, H. S. Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnol. Bioeng.* **109**, 2884–2895 (2012).

79. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12678–12683 (2005).

80. Madzak, C., Tréton, B. & Blanchin-Roland, S. Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. *J. Mol. Microbiol. Biotechnol.* **2**, 207–216 (2000).

81. Gertz, J., Siggia, E. D. & Cohen, B. A. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* **457**, 215–218 (2009).

82. Patwardhan, R. P. *et al.* High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* **27**, 1173–1175 (2009).

83. Kwasnieski, J. C., Mogno, I., Myers, C. A., Corbo, J. C. & Cohen, B. A. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19498–19503 (2012).

84. Zeevi, D. *et al.* Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome Res.* **21**, 2114–2128 (2011).

85. Rajkumar, A. S., Dénervaud, N. & Maerkl, S. J. Mapping the fine structure of a eukaryotic promoter input-output function. *Nat. Genet.* **45**, 1207–1215 (2013).

86. de Boer, H. A., Comstock, L. J. & Vasser, M. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 21–25 (1983).

87. Haldimann, A., Daniels, L. L. & Wanner, B. L. Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the *Escherichia coli* phosphate regulon. *J. Bacteriol.* **180**, 1277–1286 (1998).

88. Sengstag, C. & Hinnen, A. A 28-bp segment of the *Saccharomyces cerevisiae* PHO5 upstream activator sequence confers phosphate control to the CYC1-lacZ gene fusion. *Gene* **67**, 223–228 (1988).

89. Guarente, L., Lalonde, B., Gifford, P. & Alani, E. Distinctly regulated tandem upstream activation sites mediate catabolite repression of the CYC1 gene of *S. cerevisiae*. *Cell* **36**, 503–511 (1984).

90. Patwardhan, R. P. *et al.* Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30**, 265–270 (2012).

91. Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012).

92. Dénervaud, N. A microfluidic live-cell imaging platform to study large collections of microbial genotypes. *Infoscience* **5593**, thesis (2013).

93. Kim, T. H. & Ren, B. Genome-Wide Analysis of Protein-DNA Interactions. *Annu. Rev. Genomics Hum. Genet.* **7**, 81–102 (2006).

94. Heintzman, N. D. & Ren, B. The gateway to transcription: Identifying, characterizing and understanding promoters in the eukaryotic genome. *Cell. Mol. Life Sci.* **64**, 386–400 (2007).

95. Wang, D. & Bodovitz, S. Single cell analysis: The new frontier in 'omics'. *Trends Biotechnol.* **28**, 281–290 (2010).

96. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257–272 (2019).

97. Mazutis, L. *et al.* Single-cell analysis and sorting using droplet-based microfluidics. *Nat. Protoc.* **8**, 870–891 (2013).

98. Iverson, S. V, Haddock, T. L., Beal, J. & Densmore, D. M. Cidar MoClo: Improved

MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synth. Biol.* **5**, 99–103 (2016).

99. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).

100. Martella, A., Matjusaitis, M., Auxillos, J., Pollard, S. M. & Cai, Y. Emma: An Extensible Mammalian Modular Assembly Toolkit for the Rapid Design and Production of Diverse Expression Vectors. *ACS Synth. Biol.* **6**, 1380–1392 (2017).

101. Collier, R., Thomson, J. G. & Thilmony, R. A versatile and robust *Agrobacterium*-based gene stacking system generates high-quality transgenic Arabidopsis plants. *Plant J.* **95**, 573–583 (2018).

102. Hou, L. *et al.* An Open-Source System for in Planta Gene Stacking by Bxb1 and Cre Recombinases. *Mol. Plant* **7**, 1756–1765 (2014).

103. Shih, P. M. *et al.* A robust gene-stacking method utilizing yeast assembly for plant synthetic biology. *Nat. Commun.* **7**, 1–8 (2016).

104. Hamilton, M. L. *et al.* Towards the industrial production of omega-3 long chain polyunsaturated fatty acids from a genetically modified diatom *Phaeodactylum tricornutum*. *PLoS One* **10**, 1–15 (2015).

105. Rasala, B. A., Chao, S. S., Pier, M., Barrera, D. J. & Mayfield, S. P. Enhanced genetic tools for engineering multigene traits into green algae. *PLoS One* **9**, (2014).

106. Poliner, E. *et al.* A toolkit for *Nannochloropsis oceanica* CCMP1779 enables gene stacking and genetic engineering of the eicosapentaenoic acid pathway for enhanced long-chain polyunsaturated fatty acid production. *Plant Biotechnol. J.* **16**, 298–309 (2017).

107. Kaye, Y. *et al.* Metabolic engineering toward enhanced LC-PUFA biosynthesis in *Nannochloropsis oceanica*: Overexpression of endogenous δ12 desaturase driven by stress-inducible promoter leads to enhanced deposition of polyunsaturated fatty acids in TAG. *Algal Res.* **11**, 387–398 (2015).

108. Wang, X. *et al.* Enhanced polyunsaturated fatty acid production using food wastes and biofuels byproducts by an evolved strain of *Phaeodactylum tricornutum*. *Bioresour. Technol.* **296**, 122351 (2019).

109. Lauersen, K. J. *et al.* Phototrophic production of heterologous diterpenoids and a hydroxy-functionalized derivative from *Chlamydomonas reinhardtii*. *Metab. Eng.* **49**, 116–127 (2018).

110. Verruto, J. *et al.* Unrestrained markerless trait stacking in *Nannochloropsis gaditana* through combined genome editing and marker recycling technologies . *Proc. Natl. Acad. Sci.* **115**, E7015–E7022 (2018).

111. Poliner, E., Takeuchi, T., Du, Z.-Y., Benning, C. & Farre, E. Non-transgenic marker-free gene disruption by an episomal CRISPR system in the oleaginous microalga, *Nannochloropsis oceanica* CCMP1779. *ACS Synth. Biol.* **7**, 962–968 (2018).

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

113. Karas, B. J. *et al.* Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* **6**, 6925 (2015).

114. Scaife, M. A. & Smith, A. G. Towards developing algal synthetic biology. *Biochem. Soc. Trans.* **44**, 716–722 (2016).

115. Wang, D., Lu, Y., Huang, H. & Xu, J. Establishing Oleaginous Microalgae Research Models for Consolidated Bioprocessing of Solar Energy. in *Biotechnology in China III: Biofuels and Bioenergy* (eds. Bai, F.-W., Liu, C.-G., Huang, H. & Tsao, G. T.) 69–84 (Springer Berlin Heidelberg, 2012).

116. Zittelli, G. C. *et al.* Production of eicosapentaenoic acid by *Nannochloropsis* sp . cultures in outdoor tubular photobioreactors. **70**, 299–312 (1999).

117. Rodolfi, L. *et al.* Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnol. Bioeng.* **102**, 100–112 (2009).

118. Xiao, Y., Zhang, J., Cui, J., Feng, Y. & Cui, Q. Metabolic profiles of *Nannochloropsis oceanica* IMET1 under nitrogen-deficiency stress. *Bioresour. Technol.* **130**, 731–738 (2013).

119. Banerjee, C., Dubey, K. K. & Shukla, P. Metabolic engineering of microalgal based biofuel production: Prospects and challenges. *Front. Microbiol.* **7**, (2016).

120. Iwai, M., Ikeda, K., Shimojima, M. & Ohta, H. Enhancement of extraplastidic oil synthesis in *Chlamydomonas reinhardtii* using a type-2 diacylglycerol acyltransferase with a phosphorus starvation-inducible promoter. *Plant Biotechnol. J.* **12**, 808–819 (2014).

121. Iwai, M., Hori, K., Sasaki-Sekimoto, Y., Shimojima, M. & Ohta, H. Manipulation of oil synthesis in *Nannochloropsis* strain NIES-2145 with a phosphorus starvation-inducible promoter from *Chlamydomonas reinhardtii*. *Front. Microbiol.* **6**, 1–15 (2015).

122. Kang, N. K. *et al.* Effects of overexpression of a bHLH transcription factor on biomass and lipid production in *Nannochloropsis salina*. *Biotechnol. Biofuels* **8**, 200 (2015).

123. Kang, N. K. *et al.* Increased lipid production by heterologous expression of AtWRI1 transcription factor in *Nannochloropsis salina*. *Biotechnol. Biofuels* **10**, 1–14 (2017).

124. Kwon, S. *et al.* Enhancement of biomass and lipid productivity by overexpression of a bZIP transcription factor in *Nannochloropsis salina*. *Biotechnol. Bioeng.* **115**, 331–340 (2018).

125. Davies, F. K., Jinkerson, R. E. & Posewitz, M. C. Toward a photosynthetic microbial platform for terpenoid engineering. *Photosynth. Res.* **123**, 265–284 (2015).

126. Lubián, L. M. *et al. Nannochloropsis* (Eustigmatophyceae) as source of commercially valuable pigments. *J. Appl. Phycol.* **12**, 249–255 (2000).

127. Lu, Y. *et al.* Regulation of the cholesterol biosynthetic pathway and its integration with fatty acid biosynthesis in the oleaginous microalga *Nannochloropsis oceanica*. *Biotechnol. Biofuels* **7**, 1–15 (2014).

128. López-Rosales, A. R. *et al.* Fatty acids, hydrocarbons and terpenes of *Nannochloropsis* and *Nannochloris* isolates with potential for biofuel production. *Energies* **12**, (2019).

129. Solovchenko, A. *et al.* Interactive effects of salinity, high light, and nitrogen starvation on fatty acid and carotenoid profiles in *Nannochloropsis oceanica* CCALA 804. *Eur. J. Lipid Sci. Technol.* **116**, 635–644 (2014).

130. Chen, J. W. *et al.* Identification of a malonyl CoA-acyl carrier protein transacylase and its regulatory role in fatty acid biosynthesis in oleaginous microalga *Nannochloropsis oceanica*. *Biotechnol. Appl. Biochem.* **64**, 620–626 (2017).

131. HK., L. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 47–65 (1999).

132. Lange, B. M., Rujan, T., Martin, W. & Croteau, R. Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 13172–13177 (2000).

133. Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H. & Winther, O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**, 3387–3395 (2017).

134. Bendtsen, J. D., Nielsen, H., Von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004).

135. Gruber, A., Rocap, G., Kroth, P. G., Armbrust, E. V. & Mock, T. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J.* **81**, 519–528 (2015).

136. Cai, Y. *et al.* A cDNA clone for beta-caryophyllene synthase from *Artemisia annua*. *Phytochemistry* **61**, 523–529 (2002).

137. Vavitsas, K., Fabris, M. & Vickers, C. E. Terpenoid metabolic engineering in photosynthetic microorganisms. *Genes (Basel).* **9**, (2018).

138. Sadre, R. *et al.* Cytosolic lipid droplets as engineered organelles for production and accumulation of terpenoid biomaterials in leaves. *Nat. Commun.* **10**, 853 (2019).

139. Nobusawa, T., Hori, K., Mori, H., Kurokawa, K. & Ohta, H. Differently Localized Lysophosphatidic Acid Acyltransferases Crucial for Triacylglycerol Biosynthesis in the Oleaginous Alga *Nannochloropsis*. *Plant J.* **90**, 547–559 (2017).

140. Potvin, G. & Zhang, Z. Strategies for high-level recombinant protein expression in transgenic microalgae: A review. *Biotechnol. Adv.* **28**, 910–918 (2010).

141. Apt, K. E. In vivo characterization of diatom multipartite plastid targeting signals. *J. Cell Sci.* **115**, 4061–4069 (2002).

142. Maeda, Y., Sunaga, Y., Yoshino, T. & Tanaka, T. Oleosome-associated protein of the oleaginous diatom *Fistulifera solaris* contains an endoplasmic reticulum-targeting signal sequence. *Mar. Drugs* **12**, 3892–3903 (2014).

143. Sunaga, Y. *et al.* Chloroplast-targeting protein expression in the oleaginous diatom *Fistulifera solaris* JPCC DA0580 toward metabolic engineering. *J. Biosci. Bioeng.* **119**, 28–34 (2015).

144. Liu, G. *et al.* The functionality of α-kafirin promoter and α-kafirin signal peptide. *Plant Cell, Tissue Organ Cult.* **128**, 133–143 (2017).

145. Zittelli, G. C., Biondi, N., Rodolfi, L. & Tredici, M. R. Photobioreactors for Mass Production of Microalgae. in *Handbook of Microalgal Culture* 225–266 (John Wiley & Sons, Ltd, 2013).

146. Richmond, A. Biological Principles of Mass Cultivation of Photoautotrophic Microalgae. in *Handbook of Microalgal Culture* 169–204 (John Wiley & Sons, Ltd, 2013).

147. Wang, J., Yang, H. & Wang, F. Mixotrophic cultivation of microalgae for biodiesel production: Status and prospects. *Appl. Biochem. Biotechnol.* **172**, 3307–3329 (2014).

148. Dianursanti, D., Sugiarto, C. W., Muharam, Y. & Susanto, B. H. The effect of nutrient arrangement on biomass growth and lipid content of microalgae *Nannochloropsis oculata* in internally illuminated bubble column photobioreactor. *AIP Conf. Proc.* **2024**, (2018).

149. Xu, F., Hu, H., Cong, W., Cai, Z. & Ouyang, F. Growth characteristics and eicosapentaenoic acid production by *Nannochloropsis* sp. in mixotrophic conditions. *Biotechnol. Lett.* **26**, 51–53 (2004).

150. Das, P., Lei, W., Aziz, S. S. & Obbard, J. P. Enhanced algae growth in both phototrophic and mixotrophic culture under blue light. *Bioresour. Technol.* **102**, 3883–3887 (2011).

151. Doebbe, A. *et al.* Functional integration of the HUP1 hexose symporter gene into the genome of *C. reinhardtii*: Impacts on biological H2 production. *J. Biotechnol.* **131**, 27–33 (2007).

152. Schenk, P. M. *et al.* Second Generation Biofuels: High-Efficiency Microalgae for Biodiesel Production. *BioEnergy Res.* **1**, 20–43 (2008).

153. Gimpel, J. A., Henr??quez, V. & Mayfield, S. P. In metabolic engineering of eukaryotic microalgae: Potential and challenges come with great diversity. *Front. Microbiol.* **6**, 1–14 (2015).

154. Kean, T. A. N., Mohd, M., Kassim, A. & Cheirsilp, B. *Mixotrophic Cultivation : Biomass and Biochemical Biosynthesis for Biofuel Production*. *Microalgae Cultivation for Biofuels Production* (Elsevier Inc., 2020).

155. Kong, W. B. *et al.* Effect of glycerol and glucose on the enhancement of biomass, lipid and soluble carbohydrate production by *Chlorella vulgaris* in mixotrophic culture. *Food Technol. Biotechnol.* **51**, 62–69 (2013).

156. Zhang, L., Zhang, C., Liu, J. & Yang, N. A strategy for stimulating astaxanthin and lipid production in *Haematococcus pluvialis* by exogenous glycerol application under low light. *Algal Res.* **46**, 101779 (2020).

157. Wang, Y., Rischer, H., Eriksen, N. T. & Wiebe, M. G. Mixotrophic continuous flow cultivation of *Chlorella protothecoides* for lipids. *Bioresour. Technol.* **144**, 608–614 (2013).

APPENDIX 1

Chapter 2 supplementary data

Novel endogenous promoters for genetic engineering of the marine microalga *Nannochloropsis gaditana* CCMP526

Margaret Ramarajan[a], Michele Fabris[a,b *], Raffaela M. Abbriano[a], Mathieu Pernice[a], Peter J. Ralph[a]

[a] Climate Change Cluster (C3), University of Technology Sydney, Broadway, Ultimo, NSW 2007, Australia

[b] CSIRO Synthetic Biology Future Science Platform, GPO box 2583, Brisbane, QLD, 4001, Australia

---

The following supporting information is available for this article:

Materials and methods:

Table S1 : PCR primers used in this study

Table S2 : qRT-PCR primers used in this study

Figure S1 : *Nannochloropsis* transformation vector- pNaga4.mVenus

Figure S2 : *Nannochloropsis* transformation vector control- pNaga4

Table S3 : Electroporation parameters and culture conditions tested for *N. gaditana* transformation

Figure S3 : PCR amplification of *mVenus* gene in selected *N.gaditana* transformants

Figure S4 : Example of gating strategy used to select chlorophyll positive and mVenus positive cells in flow cytometry

Results and discussion:

Figure S5 : Comparison of Ct values (Cq) of homologous genes in *N. gadiatana*, based on those selected from *N. oceanica* transcriptome [1] using quantitative real time RT-qPCR

Table S4 : *N. gaditana* selected gene promoter regions and their sequences

Table S5 : List of endogenous gene promoters used for transgene expression in *N. gaditana* and their transformation efficiencies

Figure S6 : Initial screening of *Nannochloropsis* transformants

Figure S7 : Exponential growth rate of wild type and transgenic *N. gaditana* cultures measured from day 1–4

Table S6 : One-way ANOVA Dunnett's multiple comparisons test

Figure S8 : mVenus fluorescence measured over 5 days

Materials and methods:

Table S 1. The PCR primers used in this study for qPCR analysis, promoter region amplification and to assemble pNaga4.mVenus and pNaga4.empty vectors.

Primer number	Sequence	F/ R	Binds to	Template used
MF283	tctagaGCCTGGGGTGCCTAATGAG	R	pUC19	pUC19
MF284	actcattaggcaccccaggctctagaTCGAGAGATTTTGTGT CTC	F	pTUBB	*N. gaditana* gDNA
MF285	gtggtggtggagctcGCTTCACAAAAAAGACAGC	R	pTUBB	*N. gaditana* gDNA
MF286	tttgtgaagcgagctccaccaccaccaccaccacATGGCCAAGT TGACCAGTG	F	sh-ble	pGOP3
MF287	gagaaaggatatcgaattcTCAGTCCTGCTCCTCGGC	R	sh-ble	pGOP3
MF288	gcaggactgagaattcgatatcCTTTCTCCGGTTTTTAATTT G	F	pTubb-3'UTR	*N. gaditana* gDNA
MF289	agaaaataccgcatcaggcgctgcagAATAATTCCCGCCTGT TC	R	pTubb-3'UTR	*N. gaditana* gDNA
MF290	ctgcagCGCCTGATGCGGTATTTTCTC	F	pUC19	pUC19
MF291	acttggccatgagctcGCTTCACAAAAAAGACAGC	R	pTubb	*N. gaditana* gDNA
MF292	tttgtgaagcgagctcATGGCCAAGTTGACCAGTG	F	*sh-ble*	pGOP3
MF293	tcaccggggcGTCCTGCTCCTCGGCCAC	R	*sh-ble*	pGOP3
MF294	ggagcaggacGCCCCGGTGAAGCAGACC	F	2A+Tag+ MCS	pChlamy4
MF295	tgctcaccatGGATCCTGGTACCCCCTGG	R	2A+Tag+ MCS	pChlamy4
MF296	accaggatccATGGTGAGCAAGGGCGAG	F	*mCerulean*	pmCerulean3-N1
MF297	ggatatcgcggccgcTTACTTGTACAGCTCGTCCATG	R	*mCerulean*	pmCerulean3-N1
MF298	gtacaagtaagcggccgcgatatcCTTTCTCCGGTTTTTAATT TG	F	pTubb-3'UTR	*N. gaditana* gDNA
MF299	agaaaataccgcatcaggcgctgcagAATAATTCCCGCCTGT TC	R	pTubb-3'UTR	*N. gaditana* gDNA
MF502	ccaaggatccatggtgagcaagggc	F	*Venus*	pTrans9xD3BCluc Norm
MF503	aaaagcggccgctcacttgtacagctcgtc	R	*Venus*	pTrans9xD3BCluc Norm
MF313	GCGAGACCTGACAGAGTATTTG	F	*Nga07090*	*N. gaditana* cDNA

MF314	TTCAGCTCCTGGTCAAAGTC	R		*N. gaditana* cDNA
MF221	CCCCATAAGGTGCATATGAAAG	F	*Nga01608*	*N. gaditana* cDNA
MF222	CTCGTTAAAGTGTGAGAGTCGG	R		*N. gaditana* cDNA
MF227	TCCCCTTTTGTGCAGGAC	F	*Nga02101*	*N. gaditana* cDNA
MF228	CAATCTTGTCCACACGCTTC	R		*N. gaditana* cDNA
MF231	TGGCATTAGGATTGGGCAG	F	*Nga21005*	*N. gaditana* cDNA
MF232	AGGACAAATACAAGGGCGG	R		*N. gaditana* cDNA
MF233	CCATTCTCCTTGCCCTCTG	F	*Nga01002*	*N. gaditana* cDNA
MF234	TTCGCCTTCACACCCTTG	R		*N. gaditana* cDNA
MF241	CTTATCGAACGACTCAGCCTG	F	*Nga00164*	*N. gaditana* cDNA
MF242	TCGACCTAAATTGGCCCATG	R		*N. gaditana* cDNA
MF251	TTATCAACACCTTCTACTCCAACA	F	*Nga00934*	*N. gaditana* cDNA
MF252	CAGGCGGATCTCTAATTCAGG	R		*N. gaditana* cDNA
MF253	ACGAGCGTGAGATTCAGC	F	*Nga06354.1*	*N. gaditana* cDNA
MF254	AGACTTTGCATCCGACTCATC	R		*N. gaditana* cDNA
MF255	CACACCTGCTGCATTTTCG	F	*Nga21085*	*N. gaditana* cDNA
MF256	ACCTCAAGGGCGTTAAGAAC	R		*N. gaditana* cDNA
MF262	CTCAAGGTCTCAAGCGAACG	F	*Nga06853*	*N. gaditana* cDNA
MF263	GGTCAAAGGTCAGGTCTCCC	R		*N. gaditana* cDNA
MF229	ACCAGCCAAACCAGAGAAG	F	*Nga20972*	*N. gaditana* cDNA
MF230	AGAACGTGGGTAAGATGCAAC	R		*N. gaditana* cDNA
MF613	AAAGACCCCAACGAGAAGC	F	*Venus*	pTrans9xD3BCluc Norm
MF614	GTCCATGCCGAGAGTGATC	R		pTrans9xD3BCluc Norm
MF446	ATATAcagctgTCGTTCATTTTTCCTTGT	F	Nga01608	*N. gaditana* gDNA
MF447	AGAGAgagctcGAAGAAAATAAAAGTTGTTA	R		*N. gaditana* gDNA
MF448	ATATAcagctgCCTTACTGCCTTTCGCTT	F	Nga02101	*N. gaditana* gDNA
MF449	AGAGAgagctcTATATCAAAGTCGTAAGGGG	R		*N. gaditana* gDNA
MF450	ATATAcagctgACACGTCATCATGGAGCC	F	Nga21005	*N. gaditana* gDNA
MF451	ATATAgagctcCAACGTGGCAGAGCAGGC	R		*N. gaditana* gDNA
MF452	AGAGAcagctgAGTTTAACATGCCGTGCC	F	Nga01002	*N. gaditana* gDNA
MF453	AGAGAgagctcGGTGTTTTGCGAGGTTTT	R		*N. gaditana* gDNA
MF454	ATATAcagctgGGGGTGAACGGCTCGGAA	F	Nga00934	*N. gaditana* gDNA
MF455	AGAGAgagctcTATACATAGAGGAAACAGAG	R		*N. gaditana* gDNA
MF456	ATATAcagctgCCCACTCAGATACCCCTCT	F	Nga06354.1	*N. gaditana* gDNA
MF457	AGAGAgagctcTTCCTGATTTTGTGATGGC	R		*N. gaditana* gDNA

Table S 2. qRT-PCR primers used in this study along with their $R^{2,}$ primer efficiency, amplicon length and melting temperature (T_m).

Primer number	Tm (°C)	Amplicon length (bp)	R^2	Primer efficiency (%)
MF313	62.4	138	0.99776	103.7036334
MF314	61.8			
MF221	61	147	0.98996	104.2276717
MF222	62.2			
MF227	61.9	148	0.92162	95.34898413
MF228	61.3			
MF231	61.8	146	0.94311	108.9725378
MF232	62.6			
MF233	61.8	136	0.94083	92.38726755
MF234	62.1			
MF241	62.1	146	0.9638	92.39513252
MF242	62.3			
MF251	61.9	143	0.98947	108.7550762
MF252	61.8			
MF253	61.7	109	0.99407	97.71961563
MF254	62.1			
MF255	61.9	144	0.98733	115.330176
MF256	62.1			
MF262	58.94	121	0.98837	106.0933185
MF263	59.67			
MF229	61.8	75	0.93001	93.42285027
MF230	62.6			
MF613 MF614	62.1 61.7	78	0.99396	103.8405996

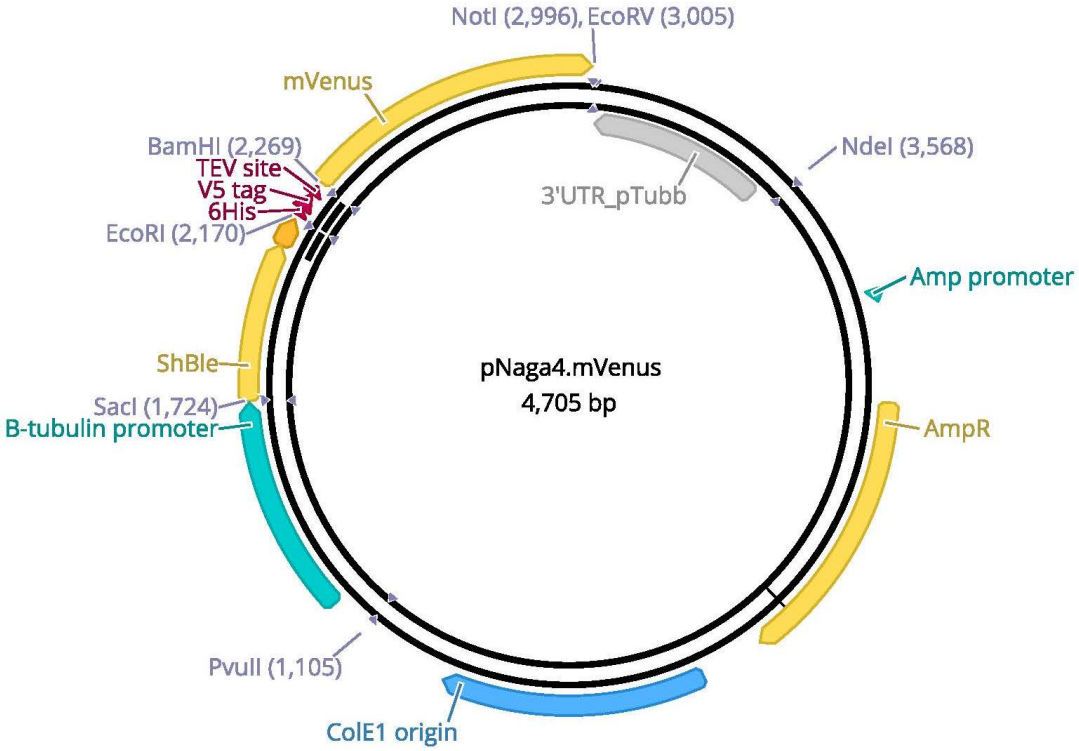


Figure S 1. *Nannochloropsis* transformation vector - pNaga4.mVenus constructed using a partial pUC19 backbone and carrying the antibiotic resistance marker *Sh ble* and reporter gene mVenus as a fusion under the control of *β-tubulin* (TUB) promoter and *β-tubulin* (TUB) terminator.

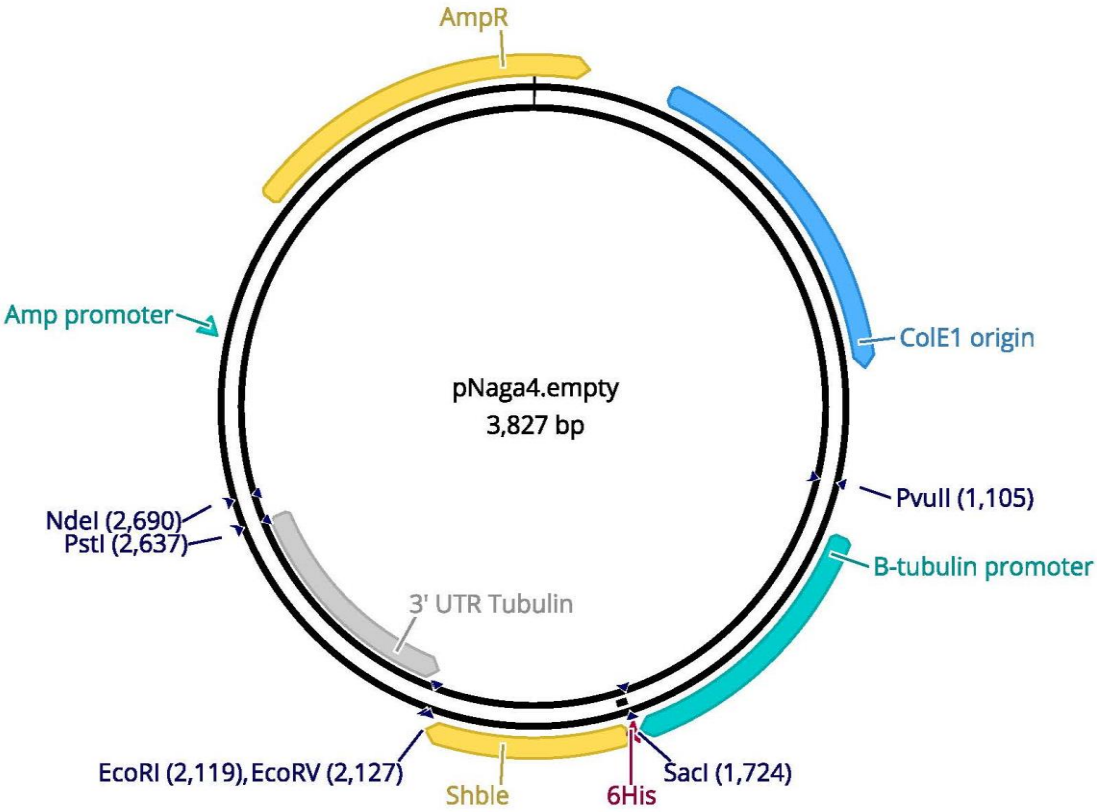


Figure S 2. *Nannochloropsis* transformation control vector - pNaga4.empty constructed using a partial pUC19 backbone and carrying the antibiotic resistance marker *Sh ble* under the control of *β-tubulin* (TUB) promoter and a *β-tubulin* (TUB) terminator.

Table S 3. Different electroporation parameters and culture conditions tested for *N. gaditana* transformation and the observed colony count.

Number of cells	Parameters used	Voltage, Resistance, Capacitance	Cuvette used	Time output	Sample resistance (Ohms)	Culture condition	Number of colonies
approx $5x10^7$ cells/mL	374 mM Sorbitol 3 washes, 1.5 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	12.6 ms	N/A	half salt F/2+ESAW medium, 20°C, 200uE light intensity (12h light/dark)	0
approx $5x10^7$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	14 ms	N/A	half salt F/2+ESAW medium, 20°C, 200uE light intensity (12h light/dark)	0
approx $5x10^7$ cells/mL	374 mM Sorbitol 3 washes, 1.5 µg linearised plasmid	2100 V, 500 Ohm, 50 µF	2 mM	12.4 ms	N/A	half salt F/2+ESAW medium, 20°C, 200uE	0

						light intensity (12h light/dark)	
approx $5x10^7$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2100 V, 500 Ohm, 50 µF	2 mM	14.4 ms	N/A	half salt F/2+ESAW medium, 20°C, 100uE light intensity (12h light/dark)	0
approx $5x10^7$ cells/mL	374 mM Sorbitol 3 washes, 2 µg double digest plasmid	2100 V, 500 Ohm, 50 µF	2 mM	13.5 ms	N/A	half salt F/2+ESAW medium, 20°C, 100uE light intensity (12h light/dark)	0
approx 1x 10^9 cells/mL	374 mM Sorbitol 3 washes, 4.6 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	10.5 ms	N/A	half salt F/2+ESAW medium, 20°C, 100uE light intensity (12h light/dark)	0
approx 1x 10^9 cells/mL	374 mM Sorbitol 3 washes, 4.6 µg linearised plasmid	2200 V, time constant- 23.4 ms, 50 µF	2 mM	N/A	468	half salt F/2+ESAW medium, 20°C, 100uE light intensity (12h light/dark)	0
approx 7.5x10^7 cells/mL	374 mM Sorbitol 5 washes, 2.5 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	14.3 ms	1031	half salt F/2+ESAW medium, 20°C, 100uE light intensity (12h light/dark)	0
approx 7.5x10^7 cells/mL	374 mM Sorbitol 5 washes, 2.5 µg linearised plasmid	2100 V, 500 Ohm, 50 µF	2 mM	13.5 ms	1000	half salt F/2+ESAW medium, 20°C, 100uE light intensity (12h light/dark)	0
approx 7.5x10^7 cells/mL	374 mM Sorbitol 5 washes, 2.5 µg linearised plasmid	2200 V, time constant- 20 ms, 25 µF	2 mM	N/A	894	half salt F/2+ESAW medium, 20°C, 100uE light intensity (12h light/dark)	0
approx 1x 10^9 cells/mL	374 mM Sorbitol 3 washes, 3 µg linearised plasmid	2400 V, 500 Ohm, 50 µF	2 mM	14.4 ms	>1100	half salt F/2+ESAW medium, 22°C, 80uE light intensity (continuous light)	0
approx 1x 10^9 cells/mL	Max effeciency b µFfer 3 washes, 3.5 µg linearised plasmid	2400 V, 500 Ohm, 50 µF	2 mM	8.1 ms	564	half salt F/2+ESAW medium, 22°C, 80uE light intensity (continuous light)	0
approx 2x10^6 cells/mL	374 mM Sorbitol 3 washes, 1.8 µg linearised plasmid	1200 V, 500 Ohm, 50 µF	1 mM	14.6 ms	>1100	half salt F/2+ESAW medium, 22°C, 80uE light intensity (continuous light)	0
approx 2x10^6 cells/mL	374 mM Sorbitol 3 washes, 1.8 µg linearised plasmid	1.2 K V, 500 Ohm, 50 µF	1 mM	15 ms	>1100	half salt F/2+ESAW medium, 22°C, 80uE light intensity (continuous light)	0
approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, 1.5 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	25.4 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	24
approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, 1.1 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	12.1 ms	1014	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	6
approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	22.7 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE	0

						light intensity (continuous light)	
approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, 2.3 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	24.2 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, 1.3 µg double digest plasmid	2200 V, 600 Ohm, 50 µF	2 mM	14.2 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, 1.5 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	13 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, no plasmid	2200 V, 600 Ohm, 50 µF	2 mM	17.8 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 2.4x 10^9 cells/mL	375 mM Sorbitol 3 washes, 1.2 µg linearised plasmid	2200 V, 500 Ohm, 50 µF	2 mM	12.2 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 2.4x 10^9 cells/mL	375 mM Sorbitol 3 washes, 1.2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	13.7 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 2.4x 10^9 cells/mL	375 mM Sorbitol 3 washes, 2.6 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	13.5 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 2.4x 10^9 cells/mL	375 mM Sorbitol 3 washes, no plasmid	2200 V, 600 Ohm, 50 µF	2 mM	15 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, 1 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	16.7 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, co-transformation pNaga4+5, total .6 µg	1100 V, 500 Ohm, 50 µF	1 mM	14.4 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, 1.1 µg linearised control (Cas9+) plasmid	1400 V, 500 Ohm, 40 µF	2 mM	16.4 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, 1.6 µg linearised control (Cas9+) plasmid	2200 V, 600 Ohm, 50 µF	2 mM	17.3 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, co-transformation pNaga5_Cas9+, total 2 µg	1100 V, 500 Ohm, 50 µF	1 mM	13.6 ms	968	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0

approx 2x 10^9 cells/mL	375 mM Sorbitol 3 washes, no plasmid	2200 V, 600 Ohm, 50 µF	2 mM	17.5 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 3x 10^9 cells/500µL	375 mM Sorbitol 3 washes, 1.1 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	9.4 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 3x 10^9 cells/500µL	375 mM Sorbitol 3 washes, 1.5 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	10.5 ms	951	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 3x 10^9 cells/500µL	375 mM Sorbitol 3 washes, 1.5 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	12.8 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 3x 10^9 cells/500µL	375 mM Sorbitol 3 washes, 1.5 µg linearised control (cas9+) plasmid	2200 V, 600 Ohm, 50 µF	2 mM	10.5 ms	986	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	49
approx 3x 10^9 cells/500µL	375 mM Sorbitol 3 washes, 1.5 µg linearised control (cas9+) plasmid	2200 V, 600 Ohm, 50 µF	2 mM	12.6 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	63
Approx 1.5x10^9 cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	18.6 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (day-night)	108
Approx 1.5x10^9 cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	18.1 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (day-night)	82
Approx 1.5x10^9 cells/mL	374 mM Sorbitol 3 washes, 1.5 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	19.6 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (day-night)	41
Approx 1.5x10^9 cells/mL	374 mM Sorbitol 3 washes, 1.5 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	18.9 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (day-night)	22
Approx 1.5x10^9 cells/mL	374 mM Sorbitol 3 washes, no plasmid control	2200 V, 600 Ohm, 50 µF	2 mM	20.7 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (day-night)	0
Approx 1.5x10^9 cells/mL	374 mM Sorbitol 3 washes, 2 µg circular plasmid	2200 V, 600 Ohm, 50 µF	2 mM	19.6 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (day-night)	0
Approx 3x10^9 cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	17.6 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 6x10^9 cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	16.9 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 6x10^9 cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	19.8 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx 6x10^9 cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	15.1 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE	63

						light intensity (continuous light)	
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	15.2 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	42
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	17.6 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	39
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	15.9 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	16 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	19 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	17.2 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	17.8 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	18 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	13.9 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	15.8 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	33
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	17 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	36
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	15.7 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	22
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	15.8 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	12
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	18.5 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE	61

						light intensity (continuous light)	
approx $6x10^9$ cells/mL	374 mM Sorbitol 3 washes, no plasmid control	2200 V, 600 Ohm, 50 µF	2 mM	22.5 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
Approx $2x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	19.7 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
Approx $2x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	18 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
Approx $2x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	19.2 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
Approx $2x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	18.6 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
Approx $2x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	18 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
Approx $2x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	18.8 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
Approx $2x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	14.9 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
Approx $2x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	15.3 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
Approx $2.5x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2.3 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	14.4 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	1
Approx $2.5x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2.3 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	14.9 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	2
Approx $2.5x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2.3 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	14.7 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
Approx $2.5x10^9$ cells/mL	374 mM Sorbitol 3 washes, 3.2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	15.3 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0
Approx $2.5x10^9$ cells/mL	374 mM Sorbitol 3 washes, no plasmid control	2200 V, 600 Ohm, 50 µF	2 mM	16.7 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE	0

						light intensity (continuous light)	
approx $1x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2.5 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	24.5 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	68
approx $1x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	26.6 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	43
approx $1x10^9$ cells/mL	374 mM Sorbitol 3 washes, 4 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	26.6 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	67
approx $1x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	26 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	48
approx $1x10^9$ cells/mL	374 mM Sorbitol 3 washes, 2.1 µg linearised plasmid	2200 V, 600 Ohm, 50 µF	2 mM	27.3 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	23
approx $1x10^9$ cells/mL	374 mM Sorbitol 3 washes, no plasmid control	2200 V, 600 Ohm, 50 µF	2 mM	27.3 ms	>1100	half salt F/2+ESAW medium, 22°C, 100uE light intensity (continuous light)	0

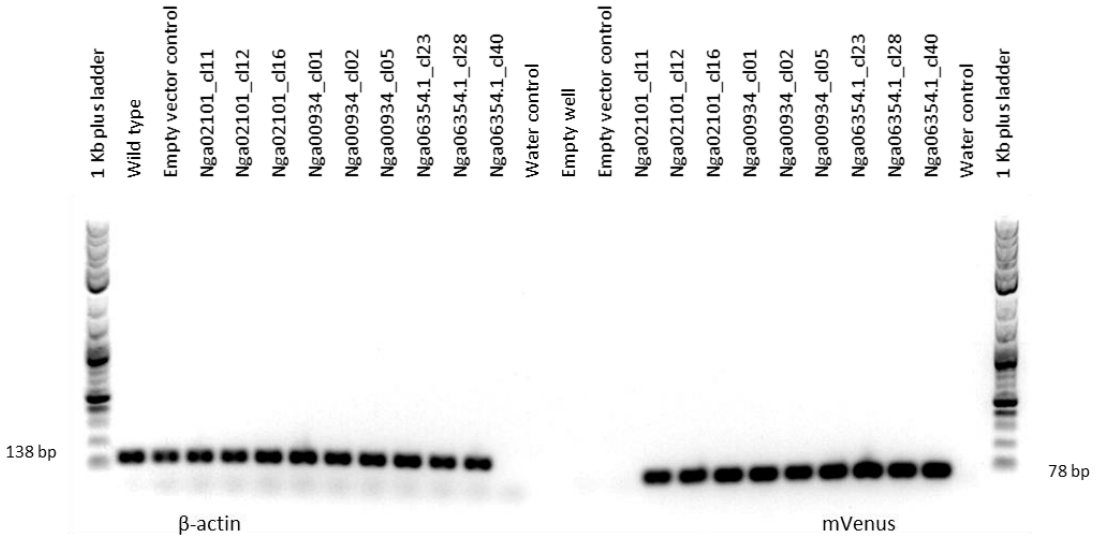
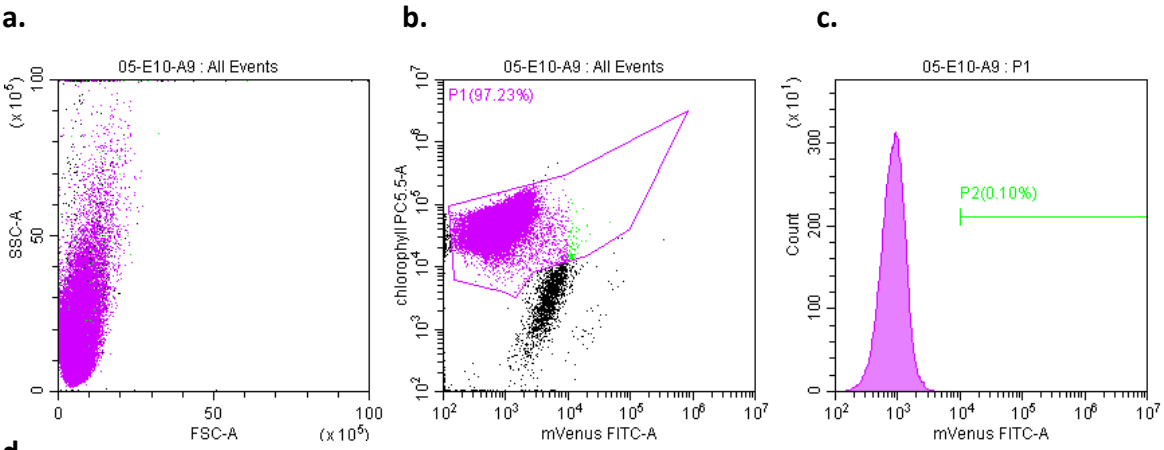


Figure S 3. PCR amplification of the mVenus reporter gene fragment and native *β-actin* gene fragment (positive control) in selected *N. gaditana* transformants.

Genomic DNA from wild-type, empty construct *pNaga4* carrying clone, three (*EPPSII, Nga02101*) promoter transgenic cell lines, three *(HSP90, Nga00934)* promoter transgenic cell lines, and three (*ATPase, Nga06354.1*) transgenic cell lines were isolated and analyzed using PCR. Specific primers for mVenus and *β-actin* were used (Table S1).

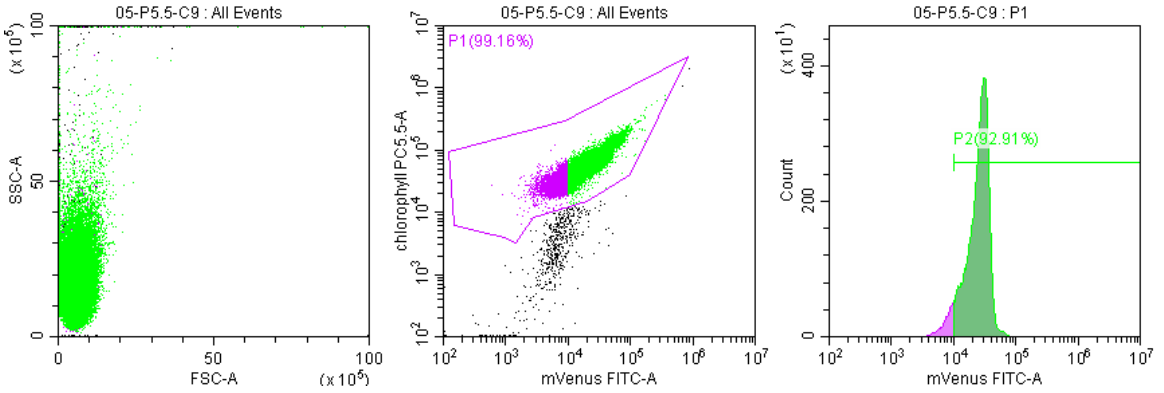


d.

e Name: 05-E10-A9

Sample ID:

Population	Events	% Total	% Parent
All Events	75552	100.00%	100.00%
P1	73456	97.23%	97.23%
P2	70	0.09%	0.10%



h.

Tube Name: 05-P5.5-C9

Sample ID:

Population	Events	% Total	% Parent
All Events	67520	100.00%	100.00%
P1	66956	99.16%	99.16%
P2	62206	92.13%	92.91%

Figure S 4. Example of gating strategy used to select chlorophyll positive and mVenus positive cells by flow cytometry.

a. All events recorded using forward and side scatter (empty control sample). **b.** P1 gate to select chlorophyll positive cells (empty control sample). **c.** P2 gate for cells expressing mVenus (empty

control sample). **d.** Events and percentage of cells in each gate (empty control sample). **e.** All events recorded using forward and side scatter (experimental sample). **f.** P1 gate to select chlorophyll positive cells (experimental sample). **g.** P2 gate for cells expressing mVenus (experimental sample). **h.** Events and percentage of cells in each gate (experimental sample).

Results and discussion:

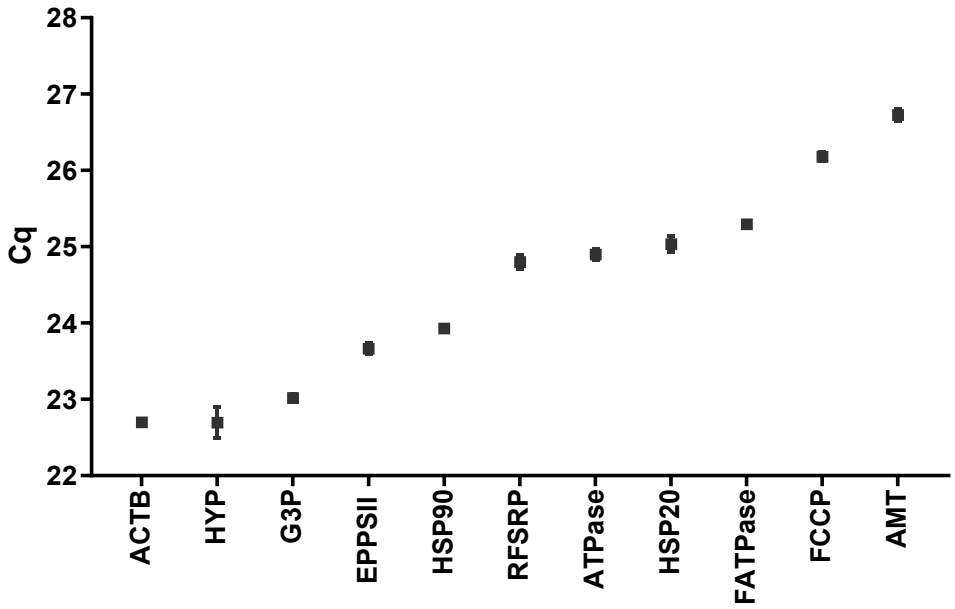


Figure S 5. Comparison of Ct values (Cq) of selected genes using quantitative real time RT-qPCR in *N. gaditana* grown under continuous light at 100 µmol photons m^{-2} s^{-1} light intensity in 50 mL liquid modified F/2-Si medium of culture density $1x10^7$ cells/mL. Lower Cq values corresponds to higher *mRNA/cDNA* levels. β-actin (*ACTB, Nga07090*) was used as a reference gene.

Table S 4. The promoter region of selected genes extracted from the *N. gaditana* genome browser along with their location co-ordinates

.

N. gaditana gene ID	Location co-ordinates	Putative promoter sequence
Nga01608	NG_chr09:558123..558943	TCGTTCATTTTTCCTTGTCATATACTCGCCAAGAAAGCAGCGGGAT CGGGTGCACGCTACAGGGCAGTACATCAGAAAAGCAAGGGGGGC ATAGCAAACTTGAGACGTCTTTTTCCATCGCCGGTCTGACTCACAC AAAAGTAGTCGCTCGTGGTTTTAAGTTATACATGGCGCCATTAC GCCTGCACGTCGGACTGCCGGAGCATTCACGTGCTTTAGCGGCGC GAGGATCCTCTCGCACGGGCGAGAGCATGCATCTGCTTCAAGGTG TCAAGCCGACGGCGCCGTCTGCGGCCTCGCGTGTCAATAATGAAA GAAAAGCTCCACATCCTGCATCATGAAGCGAGAGGAATGCATGTG

		TGTCATTGCTTCATCTTTTTTCGAAGGTCGTGTTGAACACTGTGGGC TAGGCTAAGCACTTTGACCTCTCAGAGGCAACCTCCCACCCGATCC CTTCCCCCCCCAACTATTCTGTAGTGAGATCCTGGGGTAGCCCTTTT CACATTGATCACGTGATTTTGGATAAATAACCACGCACTTTTGGGT AAATATAAATAGGTTTTGGTCGTTCATTCTTGATCTGCTGGCATCGT GTGCCAATCACGACTTCCTATAGGAGTCGGAAATGCACTTGAATTT CTTGAGCTTGAGAACCGAAGCAGAACATCAACTTGGCCTGCAAGA CTTTCACATGTGTTTTCCTTTTTAGATGAATTGAATTAAGACGTGGG AAAACCCACAGTTACGACAAAAGATTTGTTGAGACTGGAGGTGTG TCTTCCCTCGAGGATTTTCACACATTAACAACTTTTATTTTCTTC
Nga02101	NG_chr15:632052..632762	CCTTACTGCCTTTCGCTTTGGACGCTATGTGGTGTTTAATCTTTCTG TAACGTTTGAGTGCCATTCTCCGCGATCGGGTTGTTAATTAAGCTG TTAATTACAACTCTATATCCCGTCCTAGATGCCGTAATCTCAACGCT GCGGCCGCTCAGCCTCCCCTTCTCCGATTCTCCAAAGCCAAAACCC TCAAAATCAATGCCAGCAGGTGGCACTTGGCCGTcCTCGCCCTCCT TTCCGGCTTGAGGCCGTCCATCGCGGGTGCAGTGGTATCTTGAGC GTGGCCAGACCTCGATCTAAGTCCCCATCCCCCCTTTCATGGATGT CACTTGACTATGGCATATCTTTCTCTTCAACATGTCATTGCACGACT TGTCCGTAGAAAACCCTGCCCTCTTCCGCACATTGCAATTATTGAAC GAAAAGTCGCACGAGCAACCTTGTCGTCTGGAGGACTTGAAGGCC TGCAGCCATCAGGCAACGACATATTAAAGGAATTATATGACCAATT CACGACCGTTGATCTTAAATTTTGGCCACAGGAGCATGTTCTAAGA TCTATAGAGGCTAATGTAGATGCAAAGCTTTCATGTCACCCCTCAC CGGCTGGACACATGTGCAGAAGGCGCTGCCGTGAGGCGGGCGTT CCTGGGCCTTTTCTGCATCTCTATTTCTCACCCAACACATCCCTTcCC CCCTTACGACTTTGATATA
Nga21005	NG_scf09:187230..188006	ACACGTCATCATGGAGCCGCACACCCCGCAGACGAGCCACGGAGG TTTGGCAATTTGGCCCTGATGTTGCTTTGCGTCTGACCAGACGGAG ATCTTTGCTTTATCATATCTGAGCTGGTCTTGCATTTAATAGCAAAG TGGCGATGCGCATCACTTCTTCATTGTTTCTCTCCCACGACCTTCTC TGGAAGTTCATCTTGTTCTTGCACAAGATCCTCCGTCTCCAAAAAG CGTTGTTCGGCTTGCGTGTCGGAAGCTGCGCTTACAAGACCCAATC TTCTTGTTGTTAGAAGTATACTTTTCGAAATAATTCGCGGTGTATGG ACGCACTCCCTGCATCTGAAAAGCTTGTTTGAAGCCCGTCAACCTT CATGGTATCTTTTCACCCTTACGTTACCTGCCGAGCGCAACGTTCAC ACGACGGCTTTGTGTGGTGAGTGGCTGCTGCAACGCGACAGTAAT GTCCAGATCTTTTATTCTTATTTTACCTTCCCCTCTTGCGTAAGAACG GGCTTGGCCCCTGTGTGCACAATACAACCTCTTCTCTGAGAAGCAT TGTTTGTGTCAGATGAATTGGAAAAATGCATGCCGATCTTTGGCGG TCGATGCATTGGGCACAGATGCAGGAGGAAAGATTGCTCTTGATG AAAGGAGATTGGGATTTGACACGGGCCAGACATTCTGACCCGACA CGGTTACTCCTTGACTTGCACGCGCATTGCGATGCGTATTTTAACG GCACTGCTATTTCCTCTCCTTGCCTGCTCTGCCACGTTG
Nga01002	NG_chr15:254581..255308	AGTTTAACATGCCGTGCCCGACGGCCGAGTACACGCGCGCCGAAG CGGTTTGAAACGCAAGCGAAAAATATGCCTTTTCATAGCCAAGGTT TCGCGCCAGACCTTCGAAGTCGGTCTTCCCTCGGACAGTCCTCGAC TTGAAGTCGCCGTAATCCCTTCGAAAGTCCGAAAGGCGTCAGCCG

		TAATGCTAAGTCAGGCCGCAAGGAAGGCCCAAGCCCCCCCTCGCT CATACCACGACGGAGGCTCAAGATACTCACTTCCAGAGGCCCCAT GCTCGGAACGCGGACAGGGCCTTCCCCGCGCCCACGGACGACGA GGAAACTATCCTATTGCTCGTGGGAAGGAATGGTGTTTTGTTCAGT CGACCGATGGCTTTTGGCCCGGTGTGGACTGAGAGGAAAGGCTTT GCTTTCTGCGTGGGAAGGACACGAGCCAAACCCTGGTGTCGCCCT TTCATTGAGGACAGGGTCCAAGTGGAAATGCCCTCCTATTCCCCTT CCCGGCCATCTCACGGACGTATCCTATCCCAGCGAACTATTGATGG TTCTGCTATCTCACCCACGCATTGGTGGCGCGCGCATGTGAAGCAG GCATAATCGTGGCCACTTGTCTGACGGAGGGGAGTTGCCCATGGA TCCATCGCCCCTGAAGCTGGTGCTGGGTGCTCAGTGTGGCTTTGAA TGGTTGAATACCCGCGGGTCTCACCTCCAAAAACCTCGCAAAACAC C
Nga00164	NG_chr08:830300..831036	AAGTTGAGGAACTGAAGGAGATGAATATATAATATAGAAAATAAT CAGCTATCAAGCAACAGTTCTTTTTGGTATGATGCATGTACACATG TATTAAAAAAGAGGATAGGTGAGGGTAGGGGTTGGAAATATCGAA CCAAGAGCCTCTGTCTGTACTACACAAGTCAGACAACAATTATTTC CGGTGTGGCTCAGTGAGTTGATACCGCGCCCCTTATGGGCTGTCTG GGGAAAGCCATCAGTCTCCTCTGTCGCGAATGTAGCAAATACATG GACAGTGAATTCTATACAACAATGAGAGCACTAAAAAAATTAACG CGGAATATGTCATGTACGAAAGGGTACAGGCTAAAGGGGTCTTGT AGTGTCCCGACCTTGGGGGATTTCCTTCCGGCGACGCTCGGTCTTT TCCGGTTGTATTTCTTGTTCTAGAAGGGAAGCATGGGCTTTCACAT TCTAGACTGTCGTGTTGTCCGCTTCACATGAGATCATATTGTTTCCA AGGAGCTGTTCTCTTATGTTATGTTCTCAAATAATGTTTTCCCAGGC TTTCACGCCCCGTAACATTCCATGGAAAACAAACAAAAGAAGGCTT ATTCAACATCAACCTCCTGCCGTTGTACATGTCCTTTCCACAACTTC TGCAATATCTGATCTTGTTCAGATACATCTCCGCTCCACACCCATGG TGACAATATCACATCTGCCCTATTACAGGAAAGAAACCACAAGTCA ACA
Nga00934	NG_chr02:1136568..1137299	GGGGTGAACGGCTCGGAACGGTCGACGGCTCCTTCGACGAAGGC CATCGAATGTTCTGGATATCTGACATGTTCCGTGGGCACCACAATT GTTCTATTTTCGACGTCAGTCTCTGAGTGGTGGTGGCATTTTTTCAC TGTCTACGCTTGTACTGGTGCGTACATTTGGTATTTCATCTGAGAG CATAAATTACTCAATCTGTATGCATCGGGGAAGGATTTCAACTCTG CGTGCATCACTTTTTTTGCTCGAACAAAGCGCAGGAAGCAACGAA GGCTGCTTCTTTCTTTTTCCTCTTGCGATATTTCTTGCAAGGAGTCT GTAAGACAGCGACTTGCTTGCGAGAGCAACGTTGTAGGTAACATA TACTCTCTCGTGTTGTACTGTTTATTGACGTGCGGTCAGTTCTGTAG GGTTTTTGGTGCAGACAAGGATTTTCTCATGCCGTTGTGTGAGTGT GTGGGTGGGGCAGGGTGCACTGTTTTTGACACCAAATAAAAACAA CCCTTGACACGAGGAGGCCAGTAATATTTTCCCGTGTCGATGAAG ATATGTCTGCCCAGGCCTTCGTCGTAATGTTGGAGGTAGTCATTGG ATGTGAAATGATGGGCGGGTTAAATACGAAAACATCTTCGCCCCT ATGGAAGGGGAAGGGTGGGGGGCAAATGAACTCCCGTCCTATTC CCATGCCCGGCCTTCATCGAGATTCTTACTCTGTTTCCTCTATGTAT A

Nga06354.1	NG_chr10:109730..110324	CCCACTCAGATACCCCTCTTTTAATTCTCCGTCGCTCACAACGTCTG ACTTTGTCACGAAAGGTATATAATGGACGAAATAATTTTATACTTTT GCTTGCTTACCATAGATAAGTGAGCATCAAAATGCCACGGAGTAG ACCGATCAAAAAAGTACCTCTGATCGAAAACTTCCACAAGCCATCA TTTGATAAGTAATTGATGCCGTGCTAAAAGATAAGATGATTTCGTG ACTTACTGAATGAGATTCCACCTGATCTGTCTGTAATTCCAACGTGT AGACTTAACTTGGCGAAAAGAGGCACAGCTCTTGCCCTTCCACAG AGCGTTCTTAGCCCTATCCCTTTCTTGCGATAAGAAAGAAAAGCTG GTGCAACTGTGCCTAAAACACCATGTGACAGGCGAGACGTACCAT CTCTGACTCTCTTAACGTGCAGGCATTTTTTTAATTGGTCATTTCCG GCTCTGGGCGCGGCCACGACCCCGTCAAGTCAGCGTCCTGGCCCT CTTGCCGACCGAGGCCGGGGTTTTCCGAGCTTTGCTTGCCTTTACT TAACGGCCGAATATTCAGTCATCAGCCATCACAAAATCAGGAA
Nga21085	NG_chr12:337162..337973	CACCCCAGCATCATTGCGTTCGCTCCGTCTTTAGGAAGAAAGAGTC CCGATTCTGGAACCCATTGCAAGCCTCTCGTGGGGTCCTAAAGGCT CATCCTCCAGCCCGTCAGATATCTGAATCCTCTCGCTGCCTCTCGCC CCGTGGGCTGCAAGCGCACGCTGCCGGCAAGGTGGACGACACGC CTACCTGAGCGAGGGTTCCAAAAAGGCAAGCCGTCAATACATCAC TTGCGTGCGTCTTTGCGATCGTTTGACACTGGTGGTTCTGTTCGTA CATGACCATTTATTTGCTGCTGACCCACGTGCTGCTGGTTGGCTGG CGGGCGGTGTGCCCTTCTCCGTCTGGCAACGCCCGCCGGCCCCCCC TAGTCATCATGCATCTTCTCGACCCGCGTGCTTTACGCCTAGGATG GACGTCATTGTCCCATGTACCTGCCGCCTCTCAAAAAACAGTGTCT GTACTTTTCGATTTATTCGTTCTTGGGCGTGATGTATGGTCGTAAGT CACGGTACGAAGACGACAAGTGTAAGGGGACGCACGAGGACGAG CAAAGCTTTGGCTTGGCCCCCTCTCCTTGTCTCTTAATTAGTTGCGT CCTTGCACAACAACCAACACAAACTCGTGTAGTTGACACACTATTG AATAAGAGTTGGAAGAGGATGGTGCGATTTCAGGATTTTGGCATC TAGTAGACCTACAAGCGACATGCTCTTCAATGCATTTGGGAGTGGC AGAGTGCTCAAGACAAGCTGTCTTCCAGAAATCCATACCACACCCT CACTTACGTGCCCCCTACTTGAAAACACTTTTA
Nga06853	NG_scf04:79450..80203	TTATCACGTTGTCCAAGACCAAACCAGGAGTGGTGCATGTTGACG CGTGAACAAAACCTGACCACCCTGGAAATAAAAACGTGAAGCGGC GCAACTGGAGCATTTTTGCTATAATGTGTTTGAACTCACATCTTCTC GCCGTTGAGGCAAGCCTTCGAGCCCACGGTCGTCGCTTTCGGGCC CCGGCAGCAGTGGAAGACGAATCCTTGCCGCCCTCCTCGCCGGCG GACAGCGCTTGCTCCTCTTTGGAATGCTTGGATGATTTCGATACAA ATATCATCGGGAGTAGTGGCCGTTACTATGCGGGAGGAGAGGCCT CTGTGACTTCAGAGGGAGTGGTAGATGAGCGTGCATAAATGACCT CCTATCCCAACCTGTGCGACCTGCCCTCGACGGGTCGTTGAGGAGT CGAGAGCCGTCCTCTGCACCTATCTCACTTTTGTCCAGACTCCATCG CGAACGATTGTGACGCAGCGATACAATCATGTGTCTCCTACGAGCC TCTTGGACTTGTCTCATATTTCGTATCTCATGGTGCCGTACCGTCTC GTTCCCTCAAGTCACCCTAGTCTGTTCACGAGATCCATTAGGTCGC CCGTATGGCCGCAGCCATTTTTTCTCTTGATCGAGTTTACTTTGCGC AAATAGGGGCTTTATCGATGTCAATGCAACGTGCACGGCCTCACG TACCCATCTCCGCCCCCACCTTCCCACATGTCACACAGCCTACTTCC TCTCTTCATACAGAGCACCA

Nga20972	NG_chr06:961019..961669	GGGAGAGATAGGACAGGCCATCACAAGAAGGAGACCAAGTGAATACGTTATTACCCAGCCCAGAGGCTGCGATGTGGGAAAAGAGGAAGGGCTATTTTGTTGAAATGGGCTCAGGAGGGGGGTGGTCGGAGGGTGGGGAAGAAAACGGCCGCGGGGACGGGGCAGAGGCATCTGATCGAAAGAACCAAATACTTCATCCGCAGCGACACTCCACCGGGCCTGATCCTCTCCATCCCCCTTGATTTCTGACATTGCCATTGTGCCAATTACCCTTGCCGAGTGGATATGCAGGTCCTAGCGCCTTCCCTGGCGCTTCTCCTTATCATGCAAGAACATCAAAGAAATTGTTGAAGAGCATGAAGAACATGATCACGCGTGTCGCTTTTTACTTTTGCATCCCTCTCCGTCACGCTTTCAGCGGTAGGTAGTTTCCCGTCTCTCCCTGGTACGTTTCGACACAAGGCAGAACCAAGCATCAATTTGCCATCGACGACCCGGCAACCACCCATCGACACTGCAGGAAAGTTTCCCGACGCGTCGCCTCCTGCGGCTCGCATTCCGAGGAAGAGCGCTCCACCAGCCAACCACCAAAGAAAAAACACACAACGCATACCAGCAGTTTGACCATCGCCTCGATTTCCCCCCTTGGTT

Table S 5. List of endogenous gene promoters used for transgene expression in *N. gaditana* along with their gene IDs and average transformation efficiencies.

A total of four electroporations were done per promoter construct and spread on four modified f/2 media plates containing 5 µg/mL zeocin. The transformation efficiency shown is an average of the number of colonies formed on four plates per 5x 10^8 cells plated by the amount of plasmid used (µg) with standard deviation.

Promoter	Gene ID	Average transformation efficiency (colonies cells^{-1} µg DNA^{-1}, n=4)	Standard deviation
TUB	*Nga00092*	7X10^{-9}	6.26X10^{-9}
EPPSII	*Nga02101*	8x10^{-8}	5.23X10^{-8}
HSP90	*Nga00934*	3x10^{-9}	3.51X10^{-9}
ATPase	*Nga06354.1*	2x10^{-8}	2.79X10^{-8}
HYP	*Nga01608*	0	0
G3P	*Nga21005*	0	0

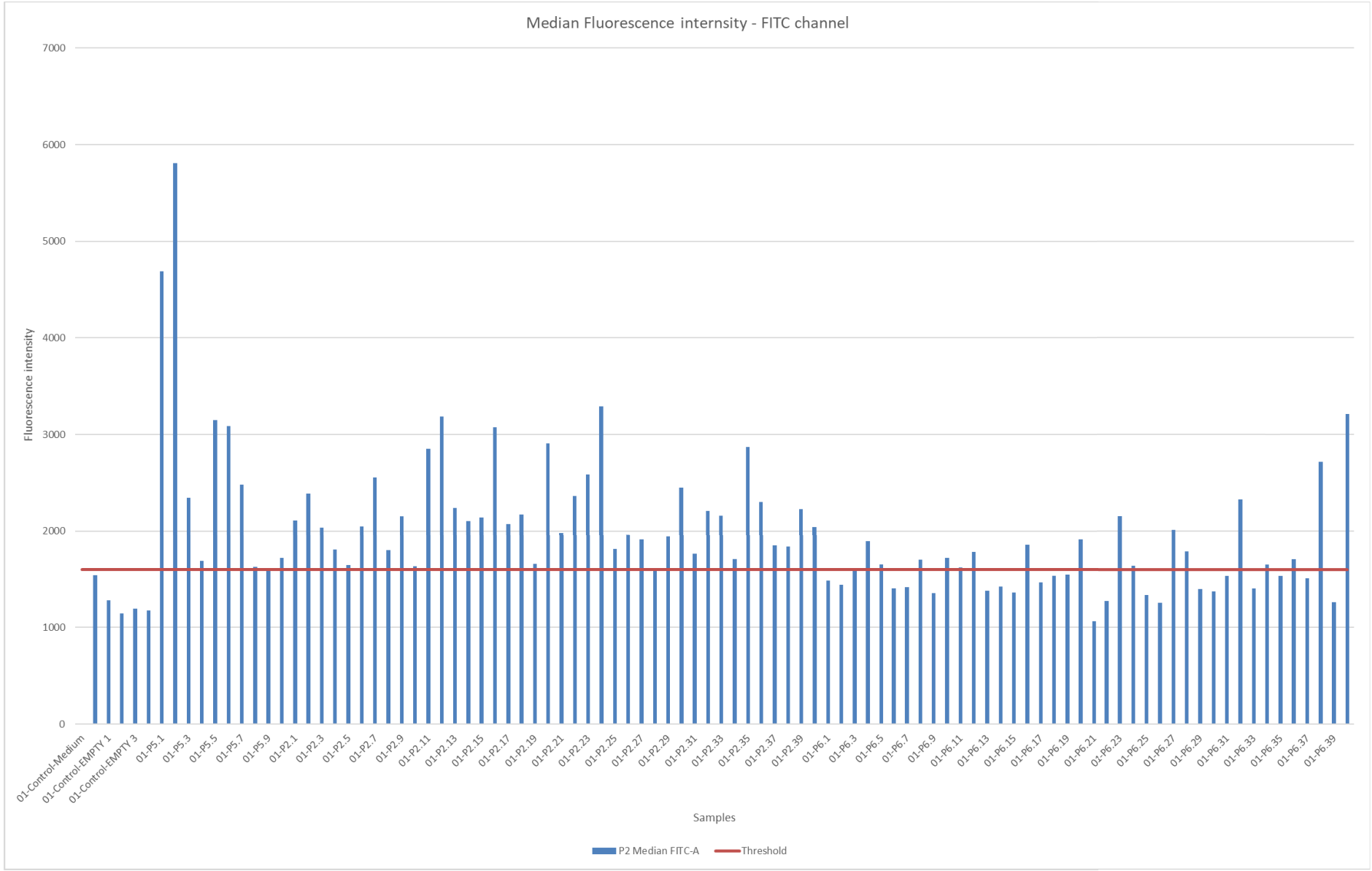

Median Fluorescence internsity - FITC channel
Fluorescence intensity
7000
6000
5000
4000
3000
2000
1000
0
01-Control-Medium
01-Control-EMPTY 1
01-Control-EMPTY 3
01-P5.1
01-P5.3
01-P5.5
01-P5.7
01-P5.9
01-P2.1
01-P2.3
01-P2.5
01-P2.7
01-P2.9
01-P2.11
01-P2.13
01-P2.15
01-P2.17
01-P2.19
01-P2.21
01-P2.23
01-P2.25
01-P2.27
01-P2.29
01-P2.31
01-P2.33
01-P2.35
01-P2.37
01-P2.39
01-P6.1
01-P6.3
01-P6.5
01-P6.7
01-P6.9
01-P6.11
01-P6.13
01-P6.15
01-P6.17
01-P6.19
01-P6.21
02-P6.23
01-P6.25
01-P6.27
01-P6.29
01-P6.31
01-P6.33
01-P6.35
01-P6.37
01-P6.39
Samples
P2 Median FITC-A
Threshold

Figure S 6. Initial screening of transgenic *N. gaditana* cell lines based on median mVenus fluorescence intensity per cell.

The controls used are culture media and transgenic cell lines carrying the empty vector construct. P2- *EPPSII, Nga02101* transgenic clones; *P5- HSP90* transgenic clones, *Nga00934* and P6- *ATPase, Nga06354.1* transgenic clones. P2-11, 12, 16; P5-1, 2, 5 and P6-23, 28, 40 were selected for further analysis

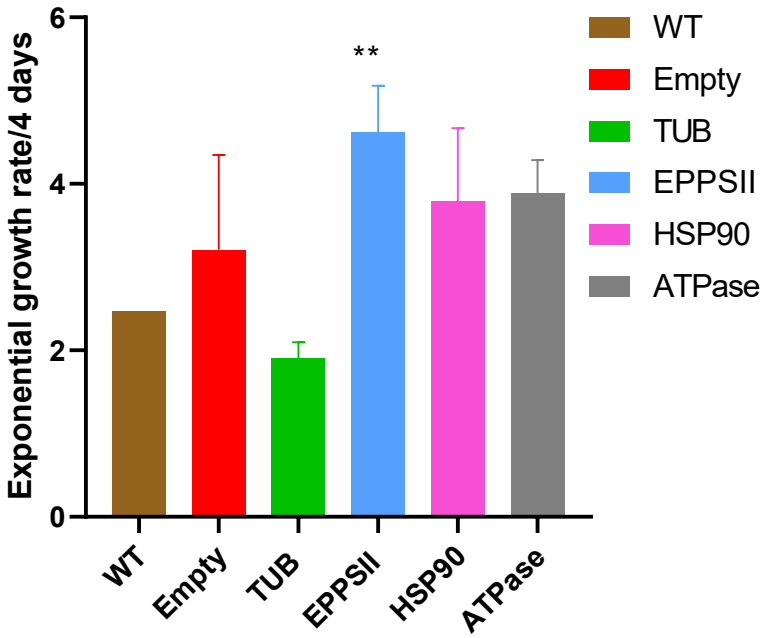


Figure S 7. Exponential growth rate of wild type and transgenic *N. gaditana* cultures measured from day 1 to 4.

Values represented are averages of three biological replicates (n=3). Error bars represent standard deviation for transgenic cell lines and for wild type. Significant differences between wild type and transgenic cell lines of *N. gaditana* calculated using one-way ANOVA- Dunnett's multiple comparisons test: ** P<0.005.

Table S 6. One-way ANOVA - Dunnett's multiple comparisons test comparing wild type *N. gaditana* culture exponential growth rate per 4 days with all transgenic *N. gaditana* lines under study.

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Su mMary	Adjusted P Value
WT vs. Empty	-0.7391	-2.296 to 0.8175	No	ns	0.5325
WT vs. TUB	0.5650	-0.9916 to 2.122	No	ns	0.7415
WT vs. EPPSII	-2.144	-3.700 to -0.5870	Yes	**	0.0072
WT vs. HSP90	-1.317	-2.874 to 0.2397	No	ns	0.1080
WT vs. ATPase	-1.418	-2.975 to 0.1384	No	ns	0.0783

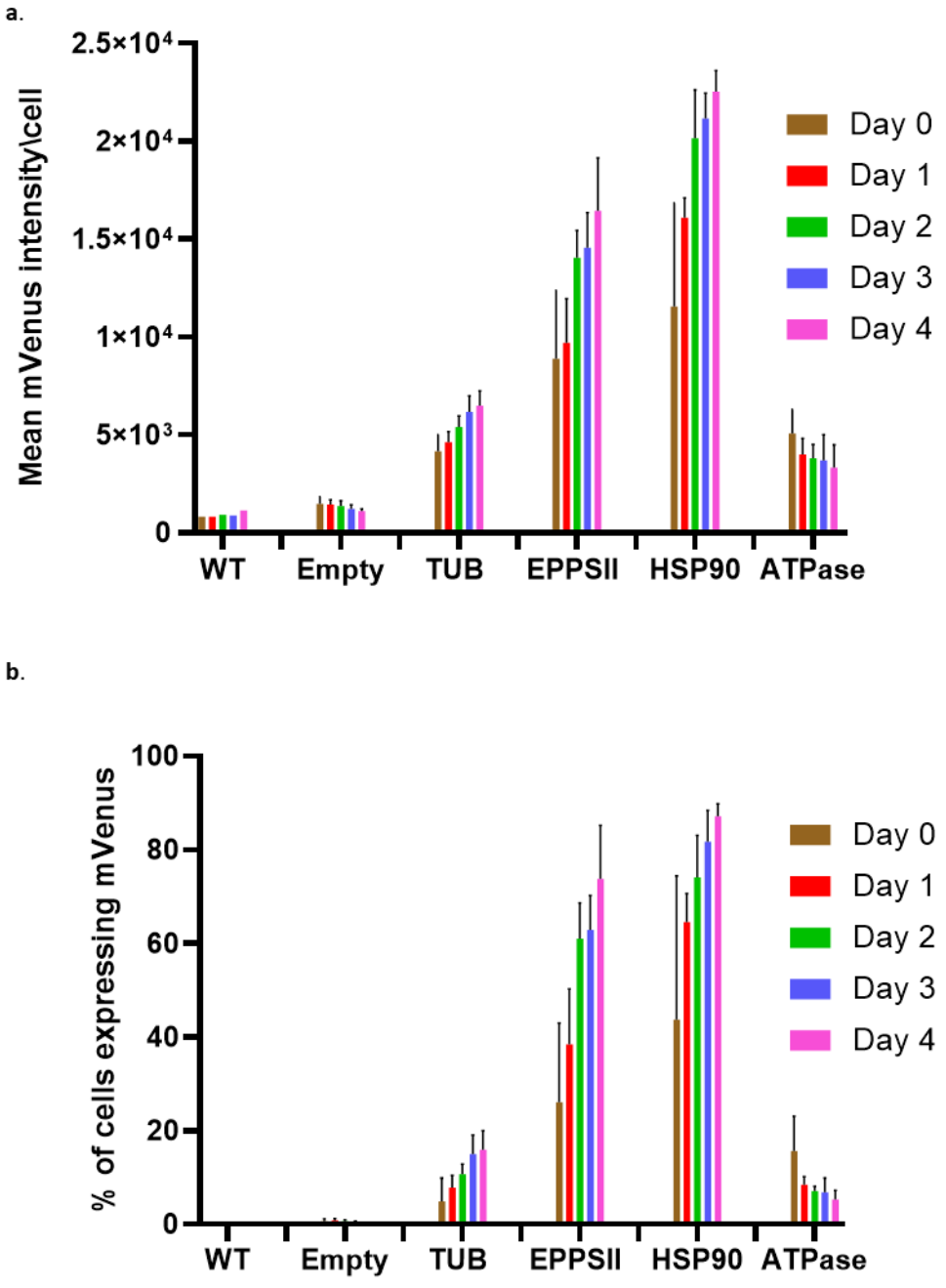


Figure S 8. mVenus fluorescence measured over 5 days.

a. Mean mVenus fluorescence per cell. **b.** Percentage of cells expressing mVenus**.** Error bars represent standard deviation for transgenic cell lines (n=9) and for wild type (n=3).

Table S 7. Putative endogenous gene promoter sequence of *N. gaditana* analysis using PlantCARE[2].

It includes the site or domain names, the conserved sequence, its annotated function along with source organism.

Promoter name	***Site Name***	***Organism***	***Position***	***Strand***	***Matrix score.***	***sequence***	***function***
extrinsic protein of PSII promoter (EPPSII, Nga02101)	A-box	Petroselinum crispum	112	+	6	CCGTCC	cis-acting regulatory element
	A-box	Petroselinum crispum	245	+	6	CCGTCC	cis-acting regulatory element
	A-box	Petroselinum crispum	216	+	6	CCGTCC	cis-acting regulatory element
	CAAT-box	Hordeum vulgare	193	+	4	CAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Hordeum vulgare	410	-	4	CAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Hordeum vulgare	401	-	4	CAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Glycine max	503	+	5	CAATT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Hordeum vulgare	359	-	4	CAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Nicotiana glutinosa	401	-	4	CAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Nicotiana glutinosa	410	-	4	CAAT	common cis-acting element in promoter and enhancer regions

	CAAT-box	Nicotiana glutinosa	405	+	4	CAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Nicotiana glutinosa	503	+	4	CAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Nicotiana glutinosa	193	+	4	CAAT	
	CAAT-box	Nicotiana glutinosa	359	-	4	CAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Arabidopsis thaliana	502	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Glycine max	405	+	5	CAATT	common cis-acting element in promoter and enhancer regions
	CCAAT-box	Hordeum vulgare	513	-	6	CAACGG	MYBHv1 binding site
	CCGTCC-box	Arabidopsis thaliana	112	+	6	CCGTCC	cis-acting regulatory element related to meristem specific activation
	CCGTCC-box	Arabidopsis thaliana	245	+	6	CCGTCC	cis-acting regulatory element related to meristem specific activation
	CCGTCC-box	Arabidopsis thaliana	216	+	6	CCGTCC	cis-acting regulatory element related to meristem specific activation
	CCGTCC-box	Petroselinum hortense	112	+	6	CCGTCC	
	CCGTCC-box	Petroselinum hortense	245	+	6	CCGTCC	
	CCGTCC-box	Petroselinum	216	+	6	CCGTCC	

		hortense					
	G-box	Zea mays	363	+	6	CACGAC	cis-acting regulatory element involved in light responsiveness
	G-box	Zea mays	508	+	6	CACGAC	cis-acting regulatory element involved in light responsiveness
	HD-Zip 1	Arabidopsis thaliana	405	-	8.5	CAAT(A/T)ATTG	element involved in differentiation of the palisade mesophyll cells
	HD-Zip 2	Arabidopsis thaliana	405	-	8	CAAT(G/C)ATTG	element involved in the control of leaf morphology development
	MYB recognition site	Arabidopsis thaliana	513	+	6	CCGTTG	
	MYC	Arabidopsis thaliana	608	-	6	CATGTG	
	MYC	Arabidopsis thaliana	610	+	6	CATGTG	
	STRE	Arabidopsis thaliana	691	-	5	AGGGG	
	STRE	Arabidopsis thaliana	155	-	5	AGGGG	
	STRE	Arabidopsis thaliana	307	-	5	AGGGG	
	STRE	Arabidopsis thaliana	592	-	5	AGGGG	
	O2-site	Zea mays	584	-	9	GATGACA	cis-acting regulatory element involved in

					TGA	zein metabolism regulation
Sp1	Zea mays	152	+	5	CC(G/A)CCC	light responsive element
Sp1	Zea mays	689	+	5	CC(G/A)CCC	light responsive element
Sp1	Zea mays	305	+	5	CC(G/A)CCC	light responsive element
TATA-box	Arabidopsis thaliana	106	+	4	TATA	core promoter element around -30 of transcription start
TATA-box	Arabidopsis thaliana	495	-	4	TATA	core promoter element around -30 of transcription start
TATA-box	Brassica napus	493	+	6	ATTATA	core promoter element around -30 of transcription start
TATA-box	Arabidopsis thaliana	707	-	4	TATA	core promoter element around -30 of transcription start
TATA-box	Glycine max	483	-	5	TAATA	core promoter element around -30 of transcription start
TATA-box	Arabidopsis thaliana	556	-	4	TATA	core promoter element around -30 of transcription start
TATA-box	Arabidopsis thaliana	494	-	5	TATAA	core promoter element around -30 of transcription start
TGA-element	Brassica oleracea	476	+	6	AACGAC	auxin-responsive element
Unnamed__1	Zea mays	276	+	5	CGTGG	

	Unnamed__4	Petroselinum hortense	65	+	4	CTCC	
	Unnamed__4	Petroselinum hortense	538	-	4	CTCC	
	Unnamed__4	Petroselinum hortense	161	+	4	CTCC	
	Unnamed__4	Petroselinum hortense	227	+	4	CTCC	
	Unnamed__4	Petroselinum hortense	153	+	4	CTCC	
	Unnamed__4	Petroselinum hortense	446	-	4	CTCC	
	Unnamed__4	Petroselinum hortense	169	+	4	CTCC	
	WRE3	Pisum sativum	203	-	6	CCACCT	
	chs-CMA2a	Petroselinum crispum	322	+	8	TCACTTGA	part of a light responsive element
heatshock protein 90 promoter (HSP90, Nga00934)	A-box	Petroselinum crispum	673	+	6	CCGTCC	cis-acting regulatory element
	AAGAA-motif	Avena sativa	280	-	7	GAAAGAA	
	ABRE	Arabidopsis thaliana	393	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	CAAT-box	Nicotiana glutinosa	87	-	4	CAAT	

	CAAT-box	Nicotiana glutinosa	389	-	4	CAAT	
	CAAT-box	Pisum sativum	162	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Pisum sativum	662	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Nicotiana glutinosa	85	+	4	CAAT	
	CAAT-box	Pisum sativum	491	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Nicotiana glutinosa	194	+	4	CAAT	
	CAAT-box	Arabidopsis thaliana	590	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
	CCAAT-box	Hordeum vulgare	444	-	6	CAACGG	MYBHv1 binding site
	CCGTCC motif	Nicotiana tabacum	673	+	6	CCGTCC	
	CCGTCC-box	Petroselinum hortense	673	+	6	CCGTCC	
	CGTCA-motif	Hordeum vulgare	103	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
	CGTCA-motif	Hordeum vulgare	391	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
	G-box	Zea mays	392	-	6	CACGTC	cis-acting regulatory element involved in

						light responsiveness
GT1-motif	Arabidopsis thaliana	612	+	6	GGTTAA	light responsive element
MYB recognition site	Arabidopsis thaliana	444	+	6	CCGTTG	
MYC	Arabidopsis thaliana	85	+	6	CAATTG	
MYC	Arabidopsis thaliana	662	-	6	CATTTG	
MYC	Arabidopsis thaliana	161	+	6	CATTTG	
STRE	Arabidopsis thaliana	635	-	5	AGGGG	
STRE	Arabidopsis thaliana	645	+	5	AGGGG	
Sp1	Oryza sativa	607	+	6	GGGCGG	light responsive element
TATA-box	Helianthus annuus	726	-	6	TATACA	core promoter element around -30 of transcription start
TATA-box	Arabidopsis thaliana	728	-	4	TATA	core promoter element around -30 of transcription start
TATA-box	Arabidopsis thaliana	364	+	4	TATA	core promoter element around -30 of transcription start
TCT-motif	Arabidopsis thaliana	707	+	6	TCTTAC	part of a light responsive element

	TCT-motif	Arabidopsis thaliana	321	-	6	TCTTAC	part of a light responsive element
	TGACG-motif	Hordeum vulgare	103	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
	TGACG-motif	Hordeum vulgare	391	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
	Unnamed__1	Zea mays	74	+	5	CGTGG	
	Unnamed__4	Petroselinum hortense	28	+	4	CTCC	
	Unnamed__4	Petroselinum hortense	670	+	4	CTCC	
	Unnamed__4	Petroselinum hortense	516	-	4	CTCC	
	Unnamed__4	Petroselinum hortense	314	-	4	CTCC	
	Unnamed__4	Petroselinum hortense	580	-	4	CTCC	
	WUN-motif	Nicotiana glutinosa	186	+	8	AAATTACT	
	as-1	Arabidopsis thaliana	103	-	5	TGACG	
	as-1	Arabidopsis thaliana	391	+	5	TGACG	
p type H+ ATPase promoter (ATPase,	AAGAA-motif	Avena sativa	356	+	7	GAAAGAA	

Nga06354.1)							
	ABRE	Arabidopsis thaliana	272	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	Arabidopsis thaliana	428	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ACE	Petroselinum crispum	402	+	9	GCGACGTACC	cis-acting element involved in light responsiveness
	AT~TATA-box	Arabidopsis thaliana	63	+	6	TATATA	
	CAAT-box	Pisum sativum	184	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Arabidopsis thaliana	447	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Nicotiana glutinosa	196	-	4	CAAT	
	CGTCA-motif	Hordeum vulgare	483	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
	DRE core	Arabidopsis thaliana	509	+	6	GCCGAC	
	G-Box	Pisum sativum	271	-	6	CACGTT	cis-acting regulatory element involved in light responsiveness
	G-Box	Pisum sativum	427	-	6	CACGTT	cis-acting regulatory element involved in light responsiveness
	G-box	Zea mays	475	+	6	CACGAC	cis-acting regulatory element involved in light responsiveness

	GA-motif	Arabidopsis thaliana	105	+	8	ATAGATAA	part of a light responsive element
	GATA-motif	Arabidopsis thaliana	212	+	10	AAGATAAGATT	part of a light responsive element
	GATA-motif	Pisum sativum	334	-	7	GATAGGG	part of a light responsive element
	MBS	Arabidopsis thaliana	372	+	6	CAACTG	MYB binding site involved in drought-inducibility
	MYC	Arabidopsis thaliana	183	+	6	CATTTG	
	MYC	Arabidopsis thaliana	390	+	6	CATGTG	
	Myb	Arabidopsis thaliana	372	+	6	CAACTG	
	Myc	Arabidopsis thaliana	421	+	7	TCTCTTA	
	STRE	Arabidopsis thaliana	12	-	5	AGGGG	
	TATA	Arabidopsis thaliana	81	-	8	TATAAAAT	
	TATA-box	Arabidopsis thaliana	63	+	6	TATATA	core promoter element around -30 of transcription start
	TATA-box	Helianthus annuus	83	-	6	TATAAA	core promoter element around -30 of transcription start
	TATA-box	Arabidopsis thaliana	65	+	4	TATA	core promoter element around -30 of transcription start

	TATA-box	Arabidopsis thaliana	85	+	4	TATA	core promoter element around -30 of transcription start
	TATA-box	Brassica oleracea	64	+	6	ATATAA	core promoter element around -30 of transcription start
	TATA-box	Arabidopsis thaliana	84	-	5	TATAA	core promoter element around -30 of transcription start
	TATA-box	Pisum sativum	82	-	7	TATAAAA	core promoter element around -30 of transcription start
	TGACG-motif	Hordeum vulgare	483	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
	Unnamed__1	Zea mays	128	-	5	CGTGG	
	Unnamed__1	Zea mays	474	-	5	CGTGG	
	Unnamed__2	Zea mays	520	-	6	CCCCGG	
	Unnamed__4	Petroselinum hortense	26	+	4	CTCC	
	Unnamed__4	Petroselinum hortense	132	-	4	CTCC	
	WRE3	Pisum sativum	248	+	6	CCACCT	
	as-1	Arabidopsis thaliana	483	-	5	TGACG	
	chs-CMA1a	Daucus carota	547	+	8	TTACTTAA	part of a light responsive element

glyceraldehyde-3-phosphate promoter (G3PD, Nga21005)	AAGAA-motif	Avena sativa	356	+	7	GAAAGAA	
	ABRE	Arabidopsis thaliana	272	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	Arabidopsis thaliana	428	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ACE	Petroselinum crispum	402	+	9	GCGACGT ACC	cis-acting element involved in light responsiveness
	AT~TATA-box	Arabidopsis thaliana	63	+	6	TATATA	
	CAAT-box	Pisum sativum	184	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Arabidopsis thaliana	447	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	Nicotiana glutinosa	196	-	4	CAAT	
	CGTCA-motif	Hordeum vulgare	483	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
	DRE core	Arabidopsis thaliana	509	+	6	GCCGAC	
	G-Box	Pisum sativum	271	-	6	CACGTT	cis-acting regulatory element involved in light responsiveness
	G-Box	Pisum sativum	427	-	6	CACGTT	cis-acting regulatory element involved in light responsiveness
	G-box	Zea mays	475	+	6	CACGAC	cis-acting regulatory element involved in

						light responsiveness
GA-motif	Arabidopsis thaliana	105	+	8	ATAGATAA	part of a light responsive element
GATA-motif	Arabidopsis thaliana	212	+	10	AAGATAAGATT	part of a light responsive element
GATA-motif	Pisum sativum	334	-	7	GATAGGG	part of a light responsive element
MBS	Arabidopsis thaliana	372	+	6	CAACTG	MYB binding site involved in drought-inducibility
MYC	Arabidopsis thaliana	183	+	6	CATTTG	
MYC	Arabidopsis thaliana	390	+	6	CATGTG	
Myb	Arabidopsis thaliana	372	+	6	CAACTG	
Myc	Arabidopsis thaliana	421	+	7	TCTCTTA	
STRE	Arabidopsis thaliana	12	-	5	AGGGG	
TATA	Arabidopsis thaliana	81	-	8	TATAAAAT	
TATA-box	Arabidopsis thaliana	63	+	6	TATATA	core promoter element around -30 of transcription start
TATA-box	Helianthus annuus	83	-	6	TATAAA	core promoter element around -30 of transcription start

	TATA-box	Arabidopsis thaliana	65	+	4	TATA	core promoter element around -30 of transcription start
	TATA-box	Arabidopsis thaliana	85	+	4	TATA	core promoter element around -30 of transcription start
	TATA-box	Brassica oleracea	64	+	6	ATATAA	core promoter element around -30 of transcription start
	TATA-box	Arabidopsis thaliana	84	-	5	TATAA	core promoter element around -30 of transcription start
	TATA-box	Pisum sativum	82	-	7	TATAAAA	core promoter element around -30 of transcription start
	TGACG-motif	Hordeum vulgare	483	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
	Unnamed__1	Zea mays	128	-	5	CGTGG	
	Unnamed__1	Zea mays	474	-	5	CGTGG	
	Unnamed__2	Zea mays	520	-	6	CCCCGG	
	Unnamed__4	Petroselinum hortense	26	+	4	CTCC	
	Unnamed__4	Petroselinum hortense	132	-	4	CTCC	
	WRE3	Pisum sativum	248	+	6	CCACCT	
	as-1	Arabidopsis thaliana	483	-	5	TGACG	

	chs-CMA1a	Daucus carota	547	+	8	TTACTTAA	part of a light responsive element
hypothetical protein promoter (HP, Nga01608)	AAGAA-motif	Avena sativa	311	+	7	GAAAGAA	
	ABRE	Arabidopsis thaliana	208	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	Arabidopsis thaliana	510	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	Arabidopsis thaliana	185	-	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	Arabidopsis thaliana	724	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	Arabidopsis thaliana	207	+	6	CACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	Arabidopsis thaliana	509	-	6	CACGTG	cis-acting element involved in the abscisic acid responsiveness
	ARE	Zea mays	151	-	6	AAACCA	cis-acting regulatory element essential for the anaerobic induction
	CAAT-box	Nicotiana glutinosa	304	+	4	CAAT	
	CAAT-box	Nicotiana glutinosa	597	+	4	CAAT	
	CAAT-box	Nicotiana glutinosa	503	-	4	CAAT	
	CAAT-box	Pisum sativum	754	-	5	CAAAT	common cis-acting element in promoter and enhancer regions

	CAAT-box	Nicotiana glutinosa	364	-	4	CAAT	
	CAAT-box	Nicotiana glutinosa	713	-	4	CAAT	
	CAAT-box	Arabidopsis thaliana	596	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
	DRE core	Arabidopsis thaliana	274	+	6	GCCGAC	
	G-Box	Pisum sativum	207	+	6	CACGTG	cis-acting regulatory element involved in light responsiveness
	G-Box	Pisum sativum	509	-	6	CACGTG	cis-acting regulatory element involved in light responsiveness
	G-box	Zea mays	185	+	6	CACGTC	cis-acting regulatory element involved in light responsiveness
	G-box	Zea mays	723	-	6	CACGTC	cis-acting regulatory element involved in light responsiveness
	G-box	Zea mays	386	-	6	CACGAC	cis-acting regulatory element involved in light responsiveness
	G-box	Zea mays	601	+	6	CACGAC	cis-acting regulatory element involved in light responsiveness
	G-box	Arabidopsis thaliana	207	+	6	CACGTG	cis-acting regulatory element involved in light responsiveness
	G-box	Arabidopsis thaliana	509	-	6	CACGTG	cis-acting regulatory element involved in light responsiveness
	GCN4_motif	Oryza sativa	126	-	7	TGAGTCA	cis-regulatory element involved in

							endosperm expression
	GTGGC-motif	Arabidopsis thaliana	587	+	10	CATCGTG TGGC	part of a light responsive element
	MYB	Arabidopsis thaliana	527	+	6	TAACCA	
	MYB-like sequence	Arabidopsis thaliana	527	+	6	TAACCA	
	MYC	Arabidopsis thaliana	354	+	6	CATGTG	
	MYC	Arabidopsis thaliana	690	+	6	CATGTG	
	MYC	Arabidopsis thaliana	688	-	6	CATGTG	
	Myb	Arabidopsis thaliana	739	-	6	TAACTG	
	STRE	Arabidopsis thaliana	82	+	5	AGGGG	
	TATA-box	Arabidopsis thaliana	20	+	4	TATA	core promoter element around -30 of transcription start
	TATA-box	Brassica oleracea	548	+	6	ATATAA	core promoter element around -30 of transcription start
	TATA-box	Arabidopsis thaliana	549	-	4	TATA	core promoter element around -30 of transcription start
	TATA-box	Arabidopsis thaliana	162	+	4	TATA	core promoter element around -30 of transcription start

	TATA-box	Arabidopsis thaliana	611	-	4	TATA	core promoter element around -30 of transcription start
	TATA-box	Arabidopsis thaliana	161	-	5	TATAA	core promoter element around -30 of transcription start
	TCA-element	Nicotiana tabacum	370	+	9	CCATCTTTTT	cis-acting element involved in salicylic acid responsiveness
	TGA-element	Brassica oleracea	564	-	6	AACGAC	auxin-responsive element
	Unnamed__1	Zea mays	149	+	5	CGTGG	
	Unnamed__1	Glycine max	711	+	11	GAATTTAATTAA	60K protein binding site
	Unnamed__1	Zea mays	530	-	5	CGTGG	
	Unnamed__1	Zea mays	725	+	5	CGTGG	
	Unnamed__4	Petroselinum hortense	199	-	4	CTCC	
	Unnamed__4	Petroselinum hortense	767	-	4	CTCC	
	Unnamed__4	Petroselinum hortense	439	+	4	CTCC	
	Unnamed__4	Petroselinum hortense	321	+	4	CTCC	
	Unnamed__	Petroselinum	615	-	4	CTCC	

	4	hortense					
	W box	Arabidopsis thaliana	420	+	6	TTGACC	

Table S 8. Putative endogenous gene promoter sequences of *N. gaditana* analysis using TSSP[3,4].

It includes the gene ID, sequence length, number of promoters predicted by the software, position of the promoter in the query sequence.

Promoter name	Gene Id	Thresholds		Length of Query sequence	TSS position	TSS score	TATA-box position	TATA-box score
		TATA promoters	**TATA-less promoters**					
extrinsic protein of PSII promoter (EPPSII)	Nga02101	1.52	-0.04	711	658	1.8706		
					267	1.88		
heat shock protein 90 promoter (HSP90)	Nga00934	1.52	-0.04	732	643	1.8842	185	3.5731
					218	1.9176		
p type H+ ATPase promoter (ATPase)	Nga06354.1	1.52	-0.04	595	462	1.8746		
glyceraldehyde-3-phosphate promoter (G3PD)	Nga21005	1.52	-0.04	777	488	1.8577		
hypothetical protein promoter (HP)	Nga01608	1.52	-0.04	821	585	1.9735	551	8.1926

Table S 9. Expression of *N. oceanica* CCMP1779 genes under light/dark cycles and prediction of cyclic gene expression[1].

FPKM values (Fragments Per Kilobase of transcript per Million mapped reads) for each sample collected at the indicated times determined by RNA-seq (data-Poliner et al., 2015) and their *N. gaditana* orthologs identified using the *N. gaditana* genome browser[5] blastp tool.

	Biological replicate 1								Biological replicate 2								
	0	3	6	9	12	15	18	21	0	3	6	9	12	15	18	21	(hrs)
***N. oceanica* track_id**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	**FPKM**	***N.gaditana* ortholog**
NannoCCMP1779|4578	3749.8	15745.3	17841.3	10107.5	6356.47	2825.09	4218.13	5870.07	3458.38	15427.6	17531.7	11960	5871.6	3005.46	4443.63	6616.39	Nga02101
NannoCCMP1779|5366	957.535	919.371	884.714	400.631	369.952	179.135	140.935	97.2187	621.044	811.783	848.896	437.065	333.964	179.772	147.342	146.862	Nga00934
NannoCCMP1779|8848	837.838	939.936	314.761	87.0406	65.2478	272.127	715.514	922.004	833.525	1023.52	364.198	111.521	64.4175	221.435	753.855	1121.97	Nga06354.1
NannoCCMP1779|8392	4183.24	1373.75	2072.79	3097.45	1779.67	1361.72	1340.69	942.246	5091.76	1540.05	2139.52	2046.36	1630.76	1140.15	1386.66	741.122	Nga01608
NannoCCMP1779|3236	2434.09	3732.04	1359	412.571	262.568	156.79	486.153	2356.42	2726.75	3931.08	1458.79	581.58	281.948	147.137	507.09	2627.88	Nga21005
NannoCCMP1779|8259	1787.84	679.006	730.756	534.721	457.003	236.019	187.06	159.61	1318	590.629	667.768	516.036	450.526	216.019	204.488	205.834	Nga00164
NannoCCMP1779|9984	834.655	1010.27	813.293	768.942	976.881	789.42	1081.49	1237.91	870.946	1024.71	838.753	833.792	952.724	855.185	1126.86	1316.58	Nga21085

NannoCCMP1779|9414	2096.87	5682.56	7225.23	4121.7	2893.03	1342.26	1784.71	2872.65	2081.13	5834.88	6821.2	4970	2835.28	1533.93	1863.8	3103.9	Nga06853
NannoCCMP1779|5503	2635.38	1165.26	377.224	397.793	963.842	1785.77	3212.87	5135.14	3064.52	1163.8	370.872	647.297	1115.9	1975.7	3112.81	5534.65	Nga20972

References:

1. Poliner, E. *et al.* Transcriptional coordination of physiological responses in Nannochloropsis oceanica CCMP1779 under light / dark cycles. *The Plant J.* **83,** 1097–1113 (2015).

2. Lescot, M. *et al.* PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **30,** 325–327 (2002).

3. Solovyev, V. V, Shahmuradov, I. A. & Salamov, A. A. Identification of Promoter Regions and Regulatory Sites. in *Computational Biology of Transcription Factor Binding* (ed. Ladunga, I.) 57–83 (Humana Press, 2010).

4. Shahmuradov, I. A., Umarov, R. K. & Solovyev, V. V. TSSPlant: A new tool for prediction of plant Pol II promoters. *Nucleic Acids Res.* **45,** (2017).

5. Giorgio Valle group. Nannochloropsis gaditana Genomes portal - CRIBI, University of Padua. (2013). Available at: http://www.nannochloropsis.org/index.php.

APPENDIX 2

Chapter 3 supplementary data

Novel endogenous inducible promoter and viral promoters for genetic engineering the microalga *Nannochloropsis gaditana*

Contributors: Margaret Ramarajan[a], Michele Fabris[a,b], Raffaela M. Abbriano[a] , Peter J. Ralph[a]

[a] University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo NSW 2007, Australia

[b] CSIRO Synthetic Biology Future Science Platform, GPO box 2583, Brisbane QLD 4001, Australia

The following supporting information is available for this chapter:

Materials and methods:

Results and discussion:

Materials and methods:

Table S 10. The PCR primers used in this study for promoter region amplification and to assemble *pNaga4.mV.DC.N2* and its variants

Primer number	Sequence	F/R	Binds to	Template used
MF769	atgcctgcaggCGAATTCCTCGAGCACCAC	F	pNaga4.mV	pNaga4.mV
MF787	gagctctcaGTCCTGCTCCTCGGCCA	R		
MF788	ccgaggagcaggactgagagctcAATCTACGCGCATAATGTCCG	F	Brix-3'UTR	*N.gaditana* gDNA
MF766	accccactagtgttaacTCCTCCACCCAGACGGTG	R		
MF767	gtggaggagttaacactagtGGGGTGAACGGCTCGGAA	F	Hsp90-promoter	*N.gaditana* gDNA
MF768	gtgctcgaggaattcgcctgcaggcatTATACATAGAGGAAACAGAGTAAGAATCTCGATG	R		
MF704	atatatgttaacATAAAGGGCGTTAATTTCATGTTG	F	GPDP di esterase promoter	*N.gaditana* gDNA
MF705	atatatcctgcaggGGCGTCAACTCTTGTTGAAGTGGG	R		
MF706	atatatgttaacCTTTGTCAAAACCTGCGTTGAA	F	sodium phosphate symporter promoter	*N.gaditana* gDNA
MF707	atatatcctgcaggTAGAACCCAACAAAAATGGATG	R		
MF710	atatgttaacAAATTCTTATTTTTACATGTTTACTTTTACC	F	endonuclease gene promoter	*N.gaditana* gDNA
MF711	atatatcctgcaggCCCCTCGGCACCAGTGCAAACA	R		
MF797	atatgttaacCTTAGCCGCGAGATTGGACAGCA	F	Ectocarpus putative promoter	Synthetic DNA fragment
MF798	atatcctgcaggTAGTCATCCCCATCAAGCACGA	R		

***T*able S 11. The R^2, efficiency %, slope and Y-intercept of the reference and target gene primers used in qRT-PCR**

Target	Efficiency %	Slope	Y-Intercept	R^2	Type
β-Actin	101.1030708	-3.295775747	31.86699971	0.98834405	Reference gene
G3P	98.13628667	-3.367411317	32.56438749	0.995137272	Reference gene
Venus	108.7346873	-3.128963755	32.10888858	0.995245856	Target gene

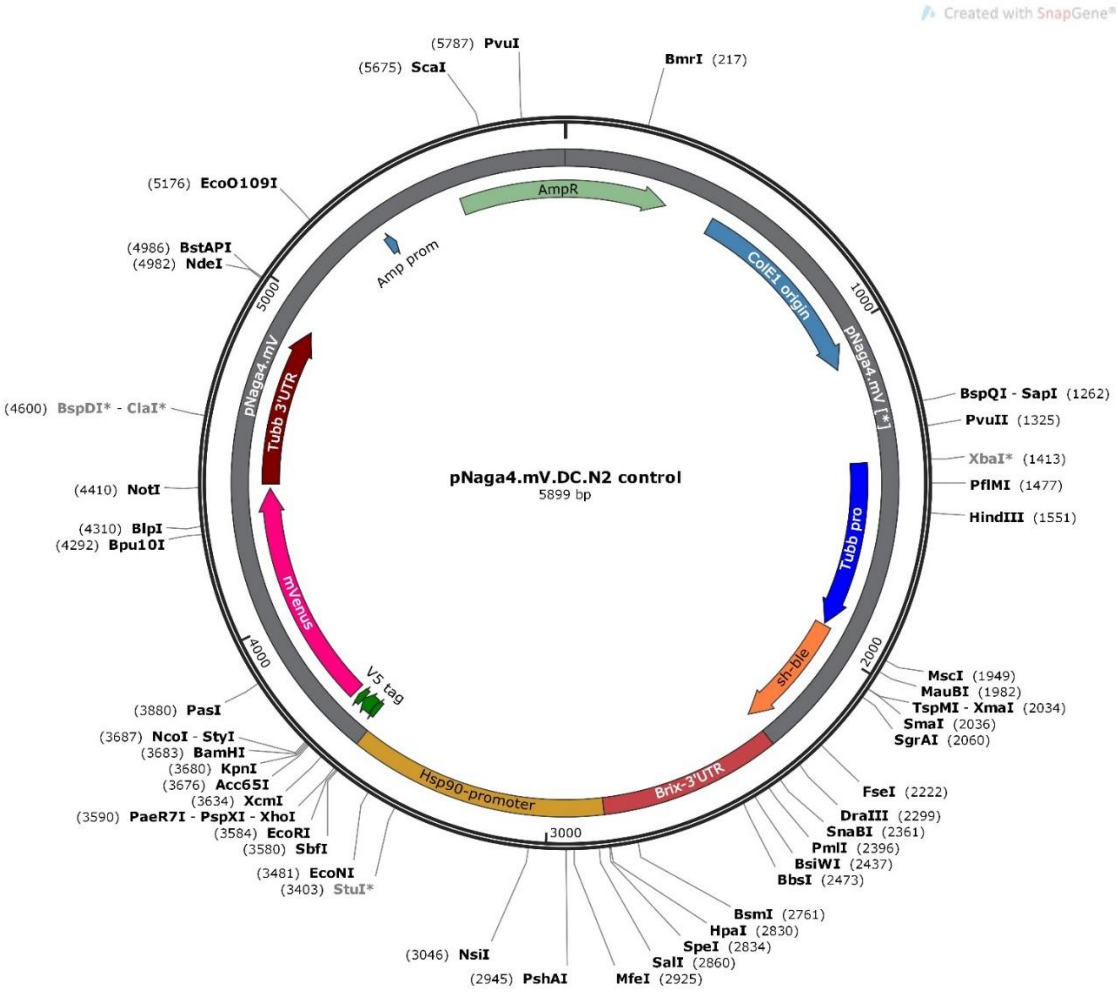
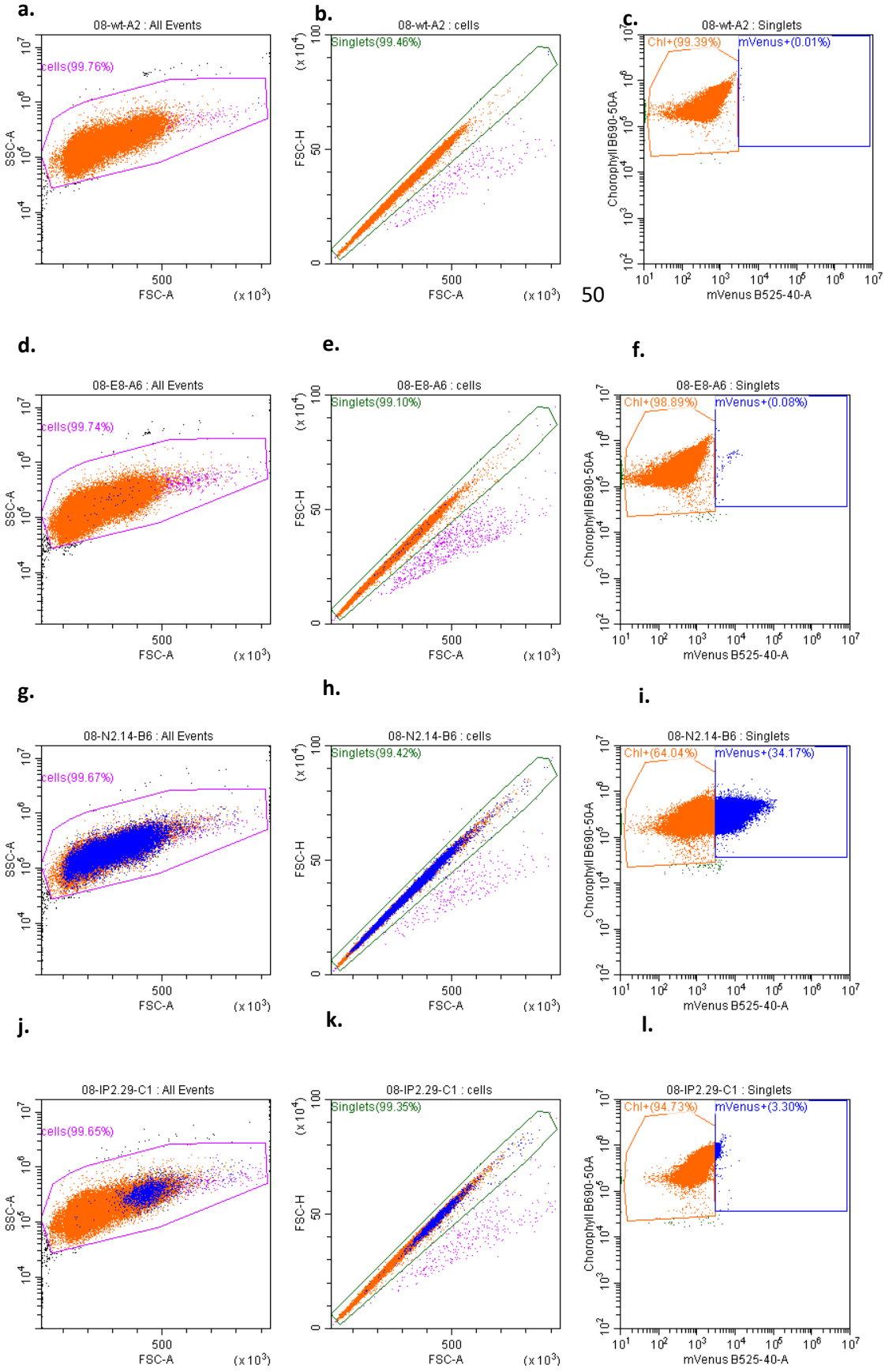


Figure S 9. Nannochloropsis transformation vector - pNaga4.mV.DC.N2 (control) constructed using the pNaga4.mVenus backbone carrying the antibiotic resistance marker Sh ble under the control of *β-tubulin* (TUB) promoter and (Brix domain) terminator and the mVenus fluorescent reporter gene under the control of a HSP90 promoter and *β-tubulin* (TUB) terminator.


a.
08-wt-A2 : All Events
cells(99.76%)
SSC-A
FSC-A
(x 10^3)
b.
08-wt-A2 : cells
Singlets(99.46%)
FSC-H
FSC-A
(x 10^3)
c.
08-wt-A2 : Singlets
Chl+(99.39%)
mVenus+(0.01%)
Chorophyll B690-50-A
mVenus B525-40-A
d.
08-E8-A6 : All Events
cells(99.74%)
e.
08-E8-A6 : cells
Singlets(99.10%)
f.
08-E8-A6 : Singlets
Chl+(98.89%)
mVenus+(0.08%)
g.
08-N2.14-B6 : All Events
cells(99.67%)
h.
08-N2.14-B6 : cells
Singlets(99.42%)
i.
08-N2.14-B6 : Singlets
Chl+(64.04%)
mVenus+(34.17%)
j.
08-IP2.29-C1 : All Events
cells(99.65%)
k.
08-IP2.29-C1 : cells
Singlets(99.35%)
l.
08-IP2.29-C1 : Singlets
Chl+(94.73%)
mVenus+(3.30%)

50

Figure S 10. Example of gating strategy used to select chlorophyll positive and mVenus positive cells by flow cytometry.

a. All events recorded using forward and side scatter (wild type sample). **b.** Singlets gate to select for single wild type cells. **c**. Chl+ gate to select for chlorophyll positive cells (wild type sample) and mVenus+ gate for cells expressing mVenus (wild type). **d.** All events recorded using forward and side scatter (empty vector control sample). **e.** Singlets gate to select for single cells in empty vector control. **f**. Chl+ gate to select for chlorophyll positive cells (empty vector control sample) and mVenus+ gate for cells expressing mVenus (empty vector control). **g.** All events recorded using forward and side scatter (positive control sample). **h.** Singlets gate to select for single cells in positive control. **i**. Chl+ gate to select for chlorophyll positive cells (positive control sample) and mVenus+ gate for cells expressing mVenus (positive control). **j.** All events recorded using forward and side scatter (experimental sample). **k.** Singlets gate to select for single cells in example experimental sampple. **l**. Chl+ gate to select for chlorophyll positive cells (empty vector control sample) and mVenus+ gate for cells expressing mVenus (example experimental sample).

Table S 12. Putative p-inducible genes from *N. gaditana*, their upstream promoter sequence along with location co-ordinates.

N.gaditana GeneBank ID	*Location co-ordinates*	*Promoter sequence*
EWM28686.1	NG_chr04:459440..460045	GGGGAGCCGTGGTCACGTTTGTGGGAATCTTATGGATACACAAGTGCATTGGGGCGCATG
		TTCTACACGTGCAGCGGACGAGCGAGGCAAGGGGAATGACAGATTCTCCCCTTTGAAATT
		CCTTAGAGGGAAGATCCCACCACCCAAAACCATCACCACGCGACACCATCGCCATCCCAT
		GTTGCGGCAAGTGTTGCATCCACAGGATCCTCGCGCCAAGACCCGTCGTGCGCAAGCCGT
		GCGGGGGGTGCGAGGGAATGCCGGGTGCAGGTGACGACAAGGGCGGAGCGAGAGGGAGTC
		CACGGCCCCGGTTCAAGGAGGCTTTGGGGCATACGCGGTCACGCCCACGCCGCTTTTTTT
		CCCCGTCGGGGCAATTTTCAGCCACGGTCGTCCCGAAATCCACCTTGAAAGGAGCTTTGT
		GTTTGCATCCTTGACATGCGTACAGCATGTTGGCGTTGGAATATTTTCACCGCACCACGT
		AGTACTTCTCTCGCGATTAGGCTTGTCCAGAAGACTGGACCCGTTTCTACCAGGCACCAA
		GTTTCTCCTATAAACATACAAGAGCGCAAGTTTCACCATTTTCATTTGTACATTTTTATT
		CTTAAA
EWM20309.1	NG_scf33:7650..8644	CTTTGTCAAAACCTGCGTTGAAGATGATAATGGTGTTGTAAAGCAAGATCATAGCGCTCA
		AGTATTAAGGAGAGACAGCATCGCGCTGTTGAGCAGAATAAGCAAGCTGCCAGCAAAAAA
		AGCCATGTAGACCACAGATTTCACCAGAACCAAAGCAATACAGCATCCATTGTCATACAT
		CAAGGACCCAGTTTGCTTGCGGCCATGCTTCTTCGCAGGAAATTCGCAGGACTGGATTAT
		TTCCGCTTTGGAAGTGTTCGCTCTTTAATGTTCGGCCAAGTGAATCTTGGCCCGACGGCT
		TCTTACGCCCCTCCAAACAATCGAATCTCGTCCAAGGAAGCTTTCATCTTGTCCATGCGG
		TCGGACAATGCCATGCATTAAATGAATTGAATCGTAATCATGTGGTGTAGACAGTTACAT
		TTTGGAGGAGGGAATCGAATCATGATTGATGCTTGAATGTGCTGAAAGGCTCCCTGTATG
		GCGTTGATTATCTCCGGACAAGCGTGGAAAGATGTTGCAGAAGTCATGGATTAAGATTTA
		AGCCCTCCTAAACGTTGTCGTCCGAAGGAAGTGATCATCGTCGACAAGGACTTTGCAAGG
		ATCGGGGAAATAAATTTACCCTTCCATGATATGCAATCCCGTCGCCTCGTATTGTCTTAC
		GGTCATGAGTTCAGGCACCAGTTGTTGAGCGGTTCGTTGCTTCTAGTGACCAGGGTGTGA
		CGAATTTGCGTCGTAATAGCATCGTCCGCCATGTTTGAGTGTGACAAACGTGACTGACTT

		GACAGCTTGCGTGGCCTGGGGAGGCTTGGGCGTCGAGCACAGGGAATAGAGAGACTCGTT
		GATGCTGGAGGGGGGgTTGAGGTTGACTTGTATTCACCACCACTCATACCCTATGCGTGG
		TTGCCTTAAAGTCTTGAAAGGTACATCGCATACACGCAAAATAGATTTTGATCGCCTCGC
		ACATGTCCTCTTTCATCCATTTTTGTTGGGTTCTA
EWM28518.1	NG_chr04:1048440..1049438	ATAAAGGGCGTTAATTTCATGTTGATACCTAGATGATTGACTAGCCATTTTCCACCCAAA
		AAGTATACGACACTCCGAAGTGGATTTTTTGCTAGGTTGCGACTCAGAATTTGTATTTAA
		CAAGGAAGAGATATCCTCCCCCCGGAGTGAGGTGAATGGGACAGGTAGCCAACAATAGAC
		ACCTCATGGATACAAAGTGCACAGCTGATTGCATTTTGCACTCGTATTTCTCCACAGGTA
		TGGCACCGAAAAGTCTTTGCGTCGGGATTCAAAATTCTCTAAAGAAATTTGCTTTCGTAT
		ATTTACCATTTCCATTCCTTTGCGTTTCTTTGTCATGGCGCTACCTCGGCCAATAGTAGA
		AAAAATAGTGAGATAACTCAACAACGAAAAACACTGTAATCAAATGACGAAACGAAGCAA
		TAAACGGTCGACGAATTGCTTCGATCGTCACATGCGGACATGCACGGAGTTAAATTGGAT
		CCGATGTACGGCAATATGTCCTCCCCAGGAAAAAACTTTAGCGAGACGAGCGAGTTGTTA
		AAAAAGCGACCCCCGGACTATAACAAAATTAGGGCGTACCACGCCAACGCTTATGCCGCC
		TGCGTTACCCCATACCAGGACACGTTACGGGAACCGCTCCTGCGTGGGACGCACTGTGAC
		GCATGTCGCCGTGCAGTTGTTATGTCAGGCACGGTGCGGGCGTGTCTGGTATTTTTGTGT
		GGTAACTACAGGCCAGGCGTCCCTCGGGGCAGGTAACGTCCTTGGCACGACACGGACACG
		ACGGGAGGATGGGAAATATCGTGCGAATCTTGCCTCATTCACAGGCCCTCCCCGCTCTGC
		CCGCGTCGAGAGAGCGCACGACGAGATCCTTCCCTTGTCATGACTGATATATGGACTCTA
		TCGGTTGCATTCGACGCATTTCCATGTATTTTGTATGAATTCGTGACTGTGACCATGGCA
		TTTTCCTCATTTCCTCCCACTTCAACAAGAGTTGACGCC

Table S 13. Putative viral ORFs from *Ectocarpus siliculosis virus-1*, their upstream promoter sequence along with location co-ordinates.

Ectocarpus siliculosis virus-1 gene ID	*Location co-ordinates*	*Promoter sequence*
920795	115176 – 116006	CTTAGCCGCGAGATTGGACAGCAACGAAATTGATTTTTTGTTGCCAACGACTTCACTGAAGCACAGTGGTCGGTATTTTT
		CAACCCAAGGCTGTGCGGTCATTTGACTACAACAACAACACCGCCACTGTTGGTGCGACTTCCATACCACAAGGATTTAA
		TAAAAATACTTACAGTACGTTTATCAACCCGCGTCCGATGCATTCAAAGTTGAAAATCATCACACAAGAGTGTCGTCGCG
		CTTGTTCTTGTTCTTTTGCTGCTCCGCGTACTTTGTGTGCCACGCCCTCCTGATCCACTCTTTTTTCTGTGAATCCCGCA
		TGGAAGACATGATGGTCCTGTACAGTACGTTTCGCTTTGCCGGTTTTTCGAGAACCATGCTTTCTGATATGGACGCGCAC
		AGTTCTCTTGTCTCGGTCGATGAGTCCGCCGTGATGATGACAAGTGGACACGCCTTGTTACACGTATCCCCTTCGCGCAA
		GTGGTAGGCGCACGTGACTCCAGACATGACGGGCATGATGATGTCCAGACATATGAGGTCATAGTCCTTCGTCTTGACCT
		TTTCAAGACATTCCAAAGCACCATTGCACAGATCAACCTCTATATCGACGTGCATCATCATTTTTTCAAACTGGCGTCTG
		GCGATGGCAGAATCGTCAACCACTAGGATGCAAGGTCTCGTTGCCGTATTGGAAATCTCGTTCTCTTGGAATATGGTAGA
		CTCCTCGGCAATGTCTTCCTCCACGCGATAAACACTCGATCCCCTTCTGAATGTCAAGGCCATGCCGTCATGTTGTAAAC
		GAAGCTCTTTCTCGTGCTTGATGGGGATGACTA
920718	93360 – 94109	TCTCTACGCCGGGAAAAGTGGAAACTATGGAGGAGCGTATTCCCCGGGCGATTACGCCGATGACGGCGGTTGTGTTACTC
		TTGTCGTATATTGGCTGAAGAATAACTAGGCCCGACTCTCCCGTGTCAATGTAAATGATGGGGTCGGACAAGTCCACTTT
		CTTGTTTGTGACCATGCTGTCAATTCCCGTGCGGATCTCTTCAGACGTGTACATATCATACCCGATCAAAGCAGGGTTCT
		CTGGGTACCTGTAGTAAATGGGCCACGACACACTTTGCGAATCGTTATCGTTTGTTATGGGTGCTAATTCGCCGTCGACT
		ATTTCCGTCATCACAACACCAAAGCCATGCACATTCGATGCTTCTTCTTCGAACTCCTGTTGCGTCAAACCGTTTGGTGA
		CACCAAAGGCGAGTATATCAACAAATCGATTTCGTTTCTGTTTGCATACGTCTCGGCTGCAAGCGTGAAGTCCTCGAGTG
		ACACCTGAAACACATTAAACATCACCGTCAACACAGAGGGTGTGACAGCATTTACCATTAAATCCGAAATGTCAGTAATG
		ACCACGGCTTTCGCCAGTTCATACTTATCAGATCGTGACTTAACAAGCGTTTGGTAAATACTTTCTGATGCAAAGTAGCA
		ACCCAGAAACAGACCCGCAAACGTTGCGAAGCATACCATTCTTGCAAAACCACGCATATTTTTCTTCGATCGTCGTTTCA
		ACCACGTATTAAAAAATATTTTTTTACTTA
920597	144026 – 145268	CTCCGGGTTGGCGTTCAAATGATCATTTTGGCGGCCCTGGGCTGCAATCTGGATATTGTTCTACACGCCTTGTGTGGCTA

	(complementary)	TTCTGGTATAGCCGCACAAGGTGTCCATGTGCTTCCCACGAATCAATTTTCAATCATCGTGAACACACATGGCCCTGCTC
		GAAACGAGAGTTACCTCGACTTGCAAGGGAATGGAACGAACATTAAGCGAACGTAGGTTTGTGGAGAAACCCTTAGTTCA
		CCAATTGGTGACCCGTTCTCTAGGGATTGACTTCACACCTTTCGATGTGGGCCCACCCACGGTCTTACCCTCTCCGTTCG
		CAAGGAACAGACACAGCTTTGACCGTTTTCACCAATTTCGCCACGAAGGTACGCCGGGTTTGGACATATACTTCTCACCC
		AGACTGTCGTCCAGGTTACGCAGGACCTCTTTTCCACGATGAGATAGATCGATCCATACGCTGACTAGGGGGTTGGAATA
		TCGCCGTTAGTTCGACGAGAGATTTCAGCCGTGGAGAGAAATCTCGGTAGCCTTGCGACCACCTATCCCCCTCTTAGGGG
		CTATCCAAAGGCAACCCCTCCTCCTCGATATAAGGTCGGTGTGTATTGACACCGACCCCCGTATGCTTCCCGAGACTCAA
		TTTTCAATCATCTCGATCACACACAGCTCACTCTCGTACTCGGAGTTACCCGAACTTGAGTGAGACTAGACTGTACATTA
		AGCCACATAGATTTGAAAAATCCTTAGCAACCGACTCTTGACCCAGTCGTTAGCGACGGAAGGCATCCCTTTCGGGATAA
		TTCCACCGACGGTCTCTGCACGAGTCTCGAGACGCACAAGACTCATACCTTTGACCGTTTTCAAGAATCTCGCCCCGACC
		CAATCACAACCATGTAGGTATGTGGAGTGAGGATCTCACCCGGACTGTCGTCCCGGTTGTATGTACCCAGGTTCTCGTGA
		GAGAGCCCGCGAGTAGCATACGGACTAGGTGGTAACAATGACATCGTTGGTTCGATGACGGGATGACTCAACGTATTCCC
		CAAAAGCCTTGCGACCTTTCATCCACCTGTTGACGGTTTGAGGTTAGCCGGCATGATTGATGTAGATGGTTGAGTAAACC
		CTGGCAACAAAATAATTCTCGAGATCGTGCAATCGTTCGTCCTTTTTATTACTCATCGTCATTATTTTCGGAAAAGATTG
		CCGTGTTTCGCGGGGGTACTTACCGTCACAAATCGTTCTGACC
920659	232370 – 232828	ATCATCTATTATTTAGCTATTCAGTCTACAGCAGCATCATCAAATTCGGTGGTTCCGCATGGCTCAAGTGCCTCGGTCAG
		CCAACCGTCTGGCACTATCTCTCACTCCCCCCACTCCCCCCACTCCCACCACTCCCACCACTCCCACCACTTCCCCCGGG
		CCCACACTCCATCGCTTCCGATCGCACCTTCTTCTACCAACACCAACCATGCGCTCCAACAACGACTTCATTACGCCGGA
		AAAGAGGAGCATCTTTACCCCCGGTGCCCCCACCAAGAAAAGGATGAAAAATACCCAAGAGACGCCAACGGAAAACGCCC
		TTCTGTTGATTTCGACCAAGGACGACGGGGACATTGCCACCTACTGTAGCAACCACTCCGTGTTGTCCGGGATTAGCGGC
		CACTTTGGCAGTATCGAGTTGTTCTTGTCCGATGAGCTTTTCGAGAGCGTGCTTGGCGTCGTAGGTGGCAGCGGCTACTC
		CTCGGTGAACCGCAGTCTTTACGTGCCGGAGGAAATCGCCGAACGTAAGAAGAGCGCTCTGCTGAAATTGTTTTCTCTGG
		AAATTCATGGTGGTGGGGAGGCACCGACGGGTGAAGTCTTTAGGGGGACCATCGTAGCGTAGGTGGTTTTTCAAACGTTT
		TTTAAAAACATACCCCGGGGTTTTGTAGAGATATTTCGTTAATGACGCACACACGGGTTGTTGCACTCGTCCAAGGGAAA
		AATTATCTCAATAACACAAAATCCATCGGTACGACCTGAAATCACCACCCCCAAAGAGCGCAACAAGGCAGTGCACACAC
		CAAAATCATGCCTATGGTCAGAAATAATAAGTACGCCGAGGTGGCAAGCGATCGCGCATCCGCGTGAAGCCTCGTGCATG
		ATATGAGATAATCCTCGTGAAGAGCAACCAGAGACGCACAGTCATCATACCCCGTCGTTTCCATCACATATCTCATCTCC

		CACCCCAACGATGGCGAATTCTTCGGATGAGAATCACAAATGGAGTCAGTATTAAAGTGGTGGTCATTTTGAATGATAGC
		CACCGCGACATTTTCAGGCACGTGGTTCCAACACTCGTTCCCTATCGGATGCCGAACGGCGAGGGTGACTCCGAACGAAA
		AAATCAGAGACATGTACGACACAATTTTAAGGTCCGCGTCCATGTTCACGTGGCACTCGCCGGAAGTCGTGTTCCAGATC
		TGAGCGTTGTCCGTCGAAGAACAAACTGCCAAGTCGCCCGATTCATACATAAAGTCCGGACAGATCGGACCACACGAAGC
		CCCGGAAGCAGACGCCATGCAGAAGAAGAGAAGACCCGGGCCAAACAAGGTTGATCTCATGGTTTTGTCGACTGCGGGTG
		ATCTAACCACACGATGAGTTGTTCAAATTTTCCACGTAATAATGAACGTTCGTTCGTTAACCATTGACATTCGCTAAATA
		TTTTGTTGCGAGTACACAA

Table S 14. Putative p-inducible gene promoter sequence analysis using PlantCARE[1].

It includes the site or domain names, the conserved sequence, its annotated function along with source organism.

Promoter name	*Site Name*	*Organism*	*Position*	*Strand*	*Matrix score.*	*sequence*	*function*
endonuclease promoter	ABRE	*Arabidopsis thaliana*	13	-	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	66	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	65	+	6	CACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	475	-	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE3a	*Zea mays*	475	-	6	TACGTG	
	ABRE4	*Zea mays*	475	+	6	CACGTA	
	AE-box	*Arabidopsis thaliana*	538	-	8	AGAAACTT	part of a module for light response
	ARE	*Zea mays*	146	+	6	AAACCA	cis-acting regulatory element essential for the anaerobic induction
	CAAT-box	*Arabidopsis thaliana*	47	-	5	CCAAT	common cis-acting element in promoter and enhancer regions

	CAAT-box	*Pisum sativum*	583	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	*Nicotiana glutinosa*	371	+	4	CAAT	
	CGTCA-motif	*Hordeum vulgare*	271	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
	G-Box	*Pisum sativum*	13	+	6	CACGTT	cis-acting regulatory element involved in light responsiveness
	G-Box	*Pisum sativum*	65	+	6	CACGTG	cis-acting regulatory element involved in light responsiveness
	G-box	*Arabidopsis thaliana*	475	-	6	TACGTG	cis-acting regulatory element involved in light responsiveness
	G-box	*Arabidopsis thaliana*	65	+	6	CACGTG	cis-acting regulatory element involved in light responsiveness
	G-box	*Zea mays*	224	-	6	CACGAC	cis-acting regulatory element involved in light responsiveness
	GC-motif	*Zea mays*	241	-	6	CCCCCG	enhancer-like element involved in anoxic specific inducibility
	Gap-box	*Arabidopsis thaliana*	578	-	9.5	CAAATGAA (A/G)A	part of a light responsive element
	LTR	*Hordeum vulgare*	392	+	6	CCGAAA	cis-acting element involved in low-temperature responsiveness
	MYC	*Arabidopsis thaliana*	582	+	6	CATTTG	
	STRE	*Arabidopsis thaliana*	89	+	5	AGGGG	
	STRE	*Arabidopsis thaliana*	107	-	5	AGGGG	
	Sp1	*Oryza sativa*	280	+	6	GGGCGG	light responsive element
	TATA-box	*Arabidopsis thaliana*	546	+	9	ccTATAAAa a	core promoter element around -30 of transcription start
	TATA-box	*Arabidopsis thaliana*	548	-	4	TATA	core promoter element around -30 of transcription start
	TGACG-motif	*Hordeum vulgare*	271	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
	Unnamed 1	*Zea mays*	7	+	5	CGTGG	

	Unnamed 1	*Zea mays*	381	-	5	CGTGG	
	Unnamed 1	*Zea mays*	299	-	5	CGTGG	
	Unnamed 1	*Zea mays*	474	-	5	CGTGG	
	Unnamed 1	*Zea mays*	155	-	5	CGTGG	
	Unnamed 1	*Zea mays*	344	-	5	CGTGG	
	Unnamed 2	*Zea mays*	305	+	6	CCCCGG	
	Unnamed 4	*Petroselinum hortense*	2	-	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	316	-	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	284	-	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	544	+	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	105	+	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	410	-	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	294	-	4	CTCC	
	WRE3	*Pisum sativum*	399	+	6	CCACCT	
	as-1	*Arabidopsis thaliana*	271	+	5	TGACG	
sodium phosphate symporter promoter	ABRE	*Arabidopsis thaliana*	657	+	7	TACGGTC	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Hordeum vulgare*	957	+	9	CGCACGTGTC	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	767	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ACE	*Petroselinum crispum*	919	-	9	GCGACGTACC	cis-acting element involved in light responsiveness
	CAAT-box	*Nicotiana glutinosa*	155	+	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	650	-	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	385	-	4	CAAT	
	CAAT-box	*Pisum sativum*	723	-	5	CAAAT	common cis-acting element in promoter and enhancer regions

	CAAT-box	*Nicotiana glutinosa*	317	+	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	633	+	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	444	-	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	168	-	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	365	+	4	CAAT	
	CGTCA-motif	*Hordeum vulgare*	717	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
	G-Box	*Pisum sativum*	766	-	6	CACGTT	cis-acting regulatory element involved in light responsiveness
	MBS	*Arabidopsis thaliana*	678	-	6	CAACTG	MYB binding site involved in drought-inducibility
	MYB	*Arabidopsis thaliana*	85	-	6	CAACAG	
	MYB	*Arabidopsis thaliana*	897	-	6	CAACCA	
	MYC	*Arabidopsis thaliana*	398	+	6	CATGTG	
	MYC	*Arabidopsis thaliana*	959	-	6	CATGTG	
	Myb	*Arabidopsis thaliana*	678	-	6	CAACTG	
	Myb	*Arabidopsis thaliana*	411	-	6	TAACTG	
	Myb-binding site	*Nicotiana tabacum*	85	-	6	CAACAG	
	STRE	*Arabidopsis thaliana*	848	+	5	AGGGG	
	STRE	*Arabidopsis thaliana*	307	-	5	AGGGG	
	TCA	*Pisum sativum*	17	-	9	TCATCTTCAT	
	TCT-motif	*Arabidopsis thaliana*	654	+	6	TCTTAC	part of a light responsive element
	TCT-motif	*Arabidopsis thaliana*	300	+	6	TCTTAC	part of a light responsive element
	TGACG-motif	*Hordeum vulgare*	717	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
	Unnamed__1	*Zea mays*	502	+	5	CGTGG	
	Unnamed__1	*Zea mays*	895	+	5	CGTGG	
	Unnamed__1	*Zea mays*	789	+	5	CGTGG	

	Unnamed 4	*Petroselinum hortense*	68	-	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	544	+	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	469	+	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	846	-	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	423	-	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	799	-	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	491	+	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	310	+	4	CTCC	
	Unnamed 4	*Petroselinum hortense*	426	-	4	CTCC	
	as-1	*Arabidopsis thaliana*	717	+	5	TGACG	
GPDP promoter	ABRE	*Hordeum vulgare*	618	-	9	GCAACGTGTC	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	620	-	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	CAAT-box	*Nicotiana glutinosa*	35	-	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	350	+	4	CAAT	
	CAAT-box	*Pisum sativum*	286	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	*Arabidopsis thaliana*	473	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	*Nicotiana glutinosa*	172	+	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	417	+	4	CAAT	
	CAAT-box	*Arabidopsis thaliana*	349	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	*Nicotiana glutinosa*	491	+	4	CAAT	
	CAAT-box	*Pisum sativum*	108	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	*Pisum sativum*	400	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	*Nicotiana glutinosa*	207	-	4	CAAT	

CAAT-box	*Nicotiana glutinosa*	434	-	4	CAAT	
CGTCA-motif	*Hordeum vulgare*	404	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
CGTCA-motif	*Hordeum vulgare*	656	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
CGTCA-motif	*Hordeum vulgare*	992	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
CGTCA-motif	*Hordeum vulgare*	445	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
G-Box	*Pisum sativum*	620	+	6	CACGTT	cis-acting regulatory element involved in light responsiveness
G-box	*Zea mays*	765	+	6	CACGAC	cis-acting regulatory element involved in light responsiveness
G-box	*Zea mays*	856	+	6	CACGAC	cis-acting regulatory element involved in light responsiveness
G-box	*Zea mays*	776	+	6	CACGAC	cis-acting regulatory element involved in light responsiveness
GC-motif	*Zea mays*	138	+	6	CCCCCG	enhancer-like element involved in anoxic specific inducibility
GC-motif	*Zea mays*	549	+	6	CCCCCG	enhancer-like element involved in anoxic specific inducibility
LTR	*Hordeum vulgare*	245	+	6	CCGAAA	cis-acting element involved in low-temperature responsiveness
MBS	*Arabidopsis thaliana*	673	-	6	CAACTG	MYB binding site involved in drought-inducibility
MYC	*Arabidopsis thaliana*	400	-	6	CATTTG	
MYC	*Arabidopsis thaliana*	448	-	6	CATGTG	
Myb	*Arabidopsis thaliana*	673	-	6	CAACTG	
TATA-box	*Arabidopsis thaliana*	63	+	4	TATA	core promoter element around -30 of transcription start
TATA-box	*Arabidopsis thaliana*	887	-	4	TATA	core promoter element around -30 of transcription start
TATA-box	*Brassica napus*	886	-	6	ATATAT	core promoter element around -30 of transcription start

TATA-box	*Oryza sativa*	928	-	7	TACAAAA	core promoter element around -30 of transcription start
TATA-box	*Arabidopsis thaliana*	297	+	4	TATA	core promoter element around -30 of transcription start
TATA-box	*Arabidopsis thaliana*	558	-	4	TATA	core promoter element around -30 of transcription start
TC-rich repeats	*Nicotiana tabacum*	273	+	9	ATTCTCTAAC	cis-acting element involved in defense and stress responsiveness
TGACG-motif	*Hordeum vulgare*	404	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
TGACG-motif	*Hordeum vulgare*	656	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
TGACG-motif	*Hordeum vulgare*	445	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
TGACG-motif	*Hordeum vulgare*	992	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
Unnamed 1	*Zea mays*	578	-	5	CGTGG	
Unnamed 1	*Zea mays*	642	+	5	CGTGG	
Unnamed 2	*Zea mays*	139	+	6	CCCCGG	
Unnamed 2	*Zea mays*	550	+	6	CCCCGG	
Unnamed 4	*Petroselinum hortense*	72	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	636	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	465	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	827	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	143	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	783	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	500	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	973	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	135	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	229	+	4	CTCC	

	WUN-motif	*Nicotiana glutinosa*	281	-	9	AAATTTCTT	
	as-1	*Arabidopsis thaliana*	404	+	5	TGACG	
	as-1	*Arabidopsis thaliana*	656	+	5	TGACG	
	as-1	*Arabidopsis thaliana*	445	-	5	TGACG	
	as-1	*Arabidopsis thaliana*	992	+	5	TGACG	

Table S 15. Putative viral gene promoter sequence analysis using PlantCARE[1].

It includes the site or domain names, the conserved sequence, its annotated function along with source organism.

Promoter name	***Site Name***	***Organism***	***Position***	***Strand***	***Matrix score.***	***sequence***	***function***
ORF 115	ABRE	*Arabidopsis thaliana*	2	-	7	AACCCGG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	373	-	7	AACCCGG	cis-acting element involved in the abscisic acid responsiveness
	ACA-motif	*Arabidopsis thaliana*	881	+	12	AATCACAACCATA	part of gapA in (gapA-CMA1) involved with light responsiveness
	CAAT-box	*Pisum sativum*	15	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	*Nicotiana glutinosa*	637	+	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	241	+	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	1095	-	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	130	+	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	880	+	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	352	+	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	1196	-	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	54	-	4	CAAT	

CAAT-box	*Pisum sativum*	730	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Arabidopsis thaliana*	351	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Nicotiana glutinosa*	1149	+	4	CAAT	
CAAT-box	*Arabidopsis thaliana*	240	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Nicotiana glutinosa*	995	+	4	CAAT	
CAAT-box	*Nicotiana glutinosa*	604	-	4	CAAT	
CAAT-box	*Pisum sativum*	1228	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Nicotiana glutinosa*	44	+	4	CAAT	
CAAT-box	*Nicotiana glutinosa*	644	+	4	CAAT	
CAAT-box	*Arabidopsis thaliana*	243	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Arabidopsis thaliana*	879	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Nicotiana glutinosa*	123	+	4	CAAT	
CAAT-box	*Nicotiana glutinosa*	265	-	4	CAAT	
CGTCA-motif	*Hordeum vulgare*	1015	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
CGTCA-motif	*Hordeum vulgare*	1176	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
CGTCA-motif	*Hordeum vulgare*	1070	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
CGTCA-motif	*Hordeum vulgare*	1223	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
DRE1	*Zea mays*	521	-	7	ACCGAGA	
GC-motif	*Zea mays*	615	+	6	CCCCCG	enhancer-like element involved in anoxic specific inducibility
GC-motif	*Zea mays*	1210	-	6	CCCCCG	enhancer-like element involved in anoxic specific inducibility

GCN4_motif	*Oryza sativa*	1023	-	7	TGAGTCA	cis-regulatory element involved in endosperm expression
LTR	*Hordeum vulgare*	789	-	6	CCGAAA	cis-acting element involved in low-temperature responsiveness
LTR	*Hordeum vulgare*	1185	-	6	CCGAAA	cis-acting element involved in low-temperature responsiveness
MYB	*Arabidopsis thaliana*	886	+	6	CAACCA	
MYB	*Arabidopsis thaliana*	1106	-	6	CAACCA	
MYB	*Arabidopsis thaliana*	1066	-	6	CAACAG	
MYC	*Arabidopsis thaliana*	15	-	6	CATTTG	
MYC	*Arabidopsis thaliana*	145	-	6	CATGTG	
MYC	*Arabidopsis thaliana*	105	+	6	CATGTG	
MYC	*Arabidopsis thaliana*	241	+	6	CAATTG	
Myb-binding site	*Nicotiana tabacum*	1066	-	6	CAACAG	
STRE	*Arabidopsis thaliana*	466	+	5	AGGGG	
STRE	*Arabidopsis thaliana*	555	+	5	AGGGG	
STRE	*Arabidopsis thaliana*	547	-	5	AGGGG	
STRE	*Arabidopsis thaliana*	574	-	5	AGGGG	
TATA-box	*Arabidopsis thaliana*	86	+	4	TATA	core promoter element around -30 of transcription start
TATA-box	*Brassica oleracea*	587	+	6	ATATAA	core promoter element around -30 of transcription start
TATA-box	*Arabidopsis thaliana*	386	+	4	TATA	core promoter element around -30 of transcription start
TATA-box	*Arabidopsis thaliana*	588	+	4	TATA	core promoter element around -30 of transcription start
TCA	*Pisum sativum*	1172	+	9	TCATCTTCAT	
TCT-motif	*Arabidopsis thaliana*	302	+	6	TCTTAC	part of a light responsive element
TGA-element	*Brassica oleracea*	765	-	6	AACGAC	auxin-responsive element

TGACG-motif	*Hordeum vulgare*	1015	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
TGACG-motif	*Hordeum vulgare*	1176	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
TGACG-motif	*Hordeum vulgare*	1070	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
TGACG-motif	*Hordeum vulgare*	1223	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
Unnamed__1	*Zea mays*	115	-	5	CGTGG	
Unnamed__1	*Zea mays*	509	+	5	CGTGG	
Unnamed__1	*Zea mays*	360	-	5	CGTGG	
Unnamed__1	*Zea mays*	296	-	5	CGTGG	
Unnamed__1	*Zea mays*	433	-	5	CGTGG	
Unnamed__4	*Petroselinum hortense*	222	-	4	CTCC	
Unnamed__4	*Petroselinum hortense*	580	+	4	CTCC	
Unnamed__4	*Petroselinum hortense*	512	-	4	CTCC	
Unnamed__4	*Petroselinum hortense*	903	-	4	CTCC	
Unnamed__4	*Petroselinum hortense*	311	+	4	CTCC	
Unnamed__4	*Petroselinum hortense*	681	-	4	CTCC	
Unnamed__4	*Petroselinum hortense*	577	+	4	CTCC	
W box	*Arabidopsis thaliana*	338	+	6	TTGACC	
W box	*Arabidopsis thaliana*	850	+	6	TTGACC	
W box	*Arabidopsis thaliana*	757	+	6	TTGACC	
WRE3	*Pisum sativum*	538	+	6	CCACCT	
WRE3	*Pisum sativum*	1062	+	6	CCACCT	

	WRE3	*Pisum sativum*	986	-	6	CCACCT	
	as-1	*Arabidopsis thaliana*	1015	+	5	TGACG	
	as-1	*Arabidopsis thaliana*	1176	-	5	TGACG	
	as-1	*Arabidopsis thaliana*	1070	+	5	TGACG	
	as-1	*Arabidopsis thaliana*	1223	-	5	TGACG	
ORF 89	ABRE	*Arabidopsis thaliana*	460	-	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	607	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	490	-	6	CACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Hordeum vulgare*	488	+	9	CGCACGTGTC	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	491	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE3a	*Zea mays*	460	-	6	TACGTG	
	ABRE4	*Zea mays*	460	+	6	CACGTA	
	CAAT-box	*Arabidopsis thaliana*	12	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	*Nicotiana glutinosa*	582	-	4	CAAT	
	CAAT-box	*Petunia hybrida*	41	+	7	TGCCAAC	common cis-acting element in promoter and enhancer regions
	CAAT-box	*Nicotiana glutinosa*	728	+	4	CAAT	
	CAAT-box	*Nicotiana glutinosa*	28	-	4	CAAT	
	CAAT-box	*Arabidopsis thaliana*	687	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
	CAAT-box	*Pisum sativum*	100	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
	CAT-box	*Arabidopsis thaliana*	122	+	6	GCCACT	cis-acting regulatory element related to meristem expression
	CGTCA-motif	*Hordeum vulgare*	506	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness

CGTCA-motif	*Hordeum vulgare*	785	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
CGTCA-motif	*Hordeum vulgare*	653	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
DRE1	*Zea mays*	410	-	7	ACCGAGA	
G-Box	*Pisum sativum*	490	-	6	CACGTG	cis-acting regulatory element involved in light responsiveness
G-box	*Arabidopsis thaliana*	460	-	6	TACGTG	cis-acting regulatory element involved in light responsiveness
G-box	*Zea mays*	606	-	6	CACGTC	cis-acting regulatory element involved in light responsiveness
G-box	*Arabidopsis thaliana*	490	-	6	CACGTG	cis-acting regulatory element involved in light responsiveness
MYB	*Arabidopsis thaliana*	126	-	6	CAACAG	
MYB	*Arabidopsis thaliana*	656	+	6	CAACCA	
MYC	*Arabidopsis thaliana*	99	+	6	CATTTG	
Myb-binding site	*Nicotiana tabacum*	126	-	6	CAACAG	
STRE	*Arabidopsis thaliana*	467	-	5	AGGGG	
STRE	*Arabidopsis thaliana*	760	-	5	AGGGG	
TATA-box	*Arabidopsis thaliana*	600	-	4	TATA	core promoter element around -30 of transcription start
TCA-element	*Brassica oleracea*	760	-	9	TCAGAAGAGG	cis-acting element involved in salicylic acid responsiveness
TGA-element	*Brassica oleracea*	45	+	6	AACGAC	auxin-responsive element
TGACG-motif	*Hordeum vulgare*	506	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
TGACG-motif	*Hordeum vulgare*	785	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
TGACG-motif	*Hordeum vulgare*	653	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
Unnamed__1	*Zea mays*	279	-	5	CGTGG	

	Unnamed__1	*Zea mays*	740	-	5	CGTGG	
	Unnamed__4	*Petroselinum hortense*	261	+	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	738	+	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	497	+	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	286	+	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	720	+	4	CTCC	
	W box	*Arabidopsis thaliana*	553	+	6	TTGACC	
	as-1	*Arabidopsis thaliana*	506	+	5	TGACG	
	as-1	*Arabidopsis thaliana*	785	-	5	TGACG	
	as-1	*Arabidopsis thaliana*	653	-	5	TGACG	
ORF 171	ABRE	*Arabidopsis thaliana*	500	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	1059	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	1167	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	1058	-	6	CACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	1392	-	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	1166	-	6	CACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE3a	*Zea mays*	499	+	6	TACGTG	
	ABRE3a	*Zea mays*	1392	-	6	TACGTG	
	ABRE4	*Zea mays*	499	-	6	CACGTA	
	ABRE4	*Zea mays*	1392	+	6	CACGTA	

ARE	*Zea mays*	623	-	6	AAACCA	cis-acting regulatory element essential for the anaerobic induction
ARE	*Zea mays*	1339	-	6	AAACCA	cis-acting regulatory element essential for the anaerobic induction
CAAT-box	*Nicotiana glutinosa*	546	-	4	CAAT	
CAAT-box	*Pisum sativum*	996	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Pisum sativum*	40	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Pisum sativum*	1383	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Nicotiana glutinosa*	352	-	4	CAAT	
CAAT-box	*Nicotiana glutinosa*	1141	+	4	CAAT	
CAAT-box	*Nicotiana glutinosa*	728	+	4	CAAT	
CAAT-box	*Nicotiana glutinosa*	1422	-	4	CAAT	
CAT-box	*Arabidopsis thaliana*	398	+	6	GCCACT	cis-acting regulatory element related to meristem expression
CCAAT-box	*Hordeum vulgare*	305	+	6	CAACGG	MYBHv1 binding site
CGTCA-motif	*Hordeum vulgare*	682	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
G-Box	*Pisum sativum*	1058	-	6	CACGTG	cis-acting regulatory element involved in light responsiveness
G-Box	*Pisum sativum*	1166	-	6	CACGTG	cis-acting regulatory element involved in light responsiveness
G-box	*Arabidopsis thaliana*	499	+	6	TACGTG	cis-acting regulatory element involved in light responsiveness
G-box	*Zea mays*	1185	-	6	CACGAC	cis-acting regulatory element involved in light responsiveness
G-box	*Arabidopsis thaliana*	1164	-	9	GCCACGTGGA	cis-acting regulatory element involved in light responsiveness
G-box	*Arabidopsis thaliana*	1392	-	6	TACGTG	cis-acting regulatory element involved in light responsiveness

G-box	*Arabidopsis thaliana*	1058	-	6	CACGTG	cis-acting regulatory element involved in light responsiveness
G-box	*Brassica napus*	1391	+	8	CCACGTAA	cis-acting regulatory element involved in light responsiveness
G-box	*Arabidopsis thaliana*	1166	-	6	CACGTG	cis-acting regulatory element involved in light responsiveness
GARE-motif	*Brassica oleracea*	321	+	7	TCTGTTG	gibberellin-responsive element
GATA-motif	*Pisum sativum*	1079	-	7	GATAGGG	part of a light responsive element
GC-motif	*Zea mays*	152	+	6	CCCCCG	enhancer-like element involved in anoxic specific inducibility
GC-motif	*Zea mays*	257	+	6	CCCCCG	enhancer-like element involved in anoxic specific inducibility
GT1-motif	*Arabidopsis thaliana*	1416	-	6	GGTTAA	light responsive element
GTGGC-motif	*Arabidopsis thaliana*	1365	-	10	CATCGTGTGG C	part of a light responsive element
MYB	*Arabidopsis thaliana*	1363	+	6	TAACCA	
MYB	*Arabidopsis thaliana*	1417	+	6	TAACCA	
MYB	*Arabidopsis thaliana*	203	+	6	CAACCA	
MYB	*Arabidopsis thaliana*	369	+	6	CAACCA	
MYB	*Arabidopsis thaliana*	322	-	6	CAACAG	
MYB	*Arabidopsis thaliana*	904	+	6	CAACCA	
MYB recognition site	*Arabidopsis thaliana*	305	-	6	CCGTTG	
MYB-like sequence	*Arabidopsis thaliana*	1363	+	6	TAACCA	
MYB-like sequence	*Arabidopsis thaliana*	1417	+	6	TAACCA	
MYC	*Arabidopsis thaliana*	996	-	6	CATTTG	
Myb-binding site	*Nicotiana tabacum*	322	-	6	CAACAG	

STRE	*Arabidopsis thaliana*	601	+	5	AGGGG	
TATA-box	*Oryza sativa*	660	-	7	TACAAAA	core promoter element around -30 of transcription start
TATA-box	*Oryza sativa*	1245	+	8	TACATAAA	core promoter element around -30 of transcription start
TCT-motif	*Arabidopsis thaliana*	524	-	6	TCTTAC	part of a light responsive element
TGA-element	*Brassica oleracea*	220	+	6	AACGAC	auxin-responsive element
TGA-element	*Brassica oleracea*	933	-	6	AACGAC	auxin-responsive element
TGACG-motif	*Hordeum vulgare*	682	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
Unnamed__1	*Zea mays*	1060	+	5	CGTGG	
Unnamed__1	*Zea mays*	1391	-	5	CGTGG	
Unnamed__1	*Zea mays*	1168	+	5	CGTGG	
Unnamed__2	*Zea mays*	153	+	6	CCCCGG	
Unnamed__2	*Zea mays*	1279	+	6	CCCCGG	
Unnamed__2	*Zea mays*	652	+	6	CCCCGG	
Unnamed__2	*Zea mays*	258	+	6	CCCCGG	
Unnamed__2	*Zea mays*	654	-	6	CCCCGG	
Unnamed__4	*Petroselinum hortense*	104	+	4	CTCC	
Unnamed__4	*Petroselinum hortense*	375	+	4	CTCC	
Unnamed__4	*Petroselinum hortense*	140	+	4	CTCC	
Unnamed__4	*Petroselinum hortense*	956	+	4	CTCC	
Unnamed__4	*Petroselinum hortense*	122	+	4	CTCC	
Unnamed__4	*Petroselinum hortense*	507	-	4	CTCC	

	Unnamed__4	*Petroselinum hortense*	213	+	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	1108	+	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	113	+	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	477	+	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	166	+	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	1001	-	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	131	+	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	576	-	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	245	-	4	CTCC	
	WRE3	*Pisum sativum*	356	+	6	CCACCT	
	WRE3	*Pisum sativum*	620	-	6	CCACCT	
	WRE3	*Pisum sativum*	838	-	6	CCACCT	
	WRE3	*Pisum sativum*	462	-	6	CCACCT	
	as-1	*Arabidopsis thaliana*	682	+	5	TGACG	
	box S	*Arabidopsis thaliana*	1037	+	7	AGCCACC	
ORF 66	ABRE	*Arabidopsis thaliana*	204	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE	*Arabidopsis thaliana*	722	-	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
	ABRE3a	*Zea mays*	722	-	6	TACGTG	
	ABRE4	*Zea mays*	722	+	6	CACGTA	
	ACE	*Petroselinum crispum*	444	-	9	GACACGTATG	cis-acting element involved in light responsiveness

ARE	*Zea mays*	686	+	6	AAACCA	cis-acting regulatory element essential for the anaerobic induction
CAAT-box	*Arabidopsis thaliana*	89	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Nicotiana glutinosa*	180	+	4	CAAT	
CAAT-box	*Nicotiana glutinosa*	126	+	4	CAAT	
CAAT-box	*Pisum sativum*	421	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CGTCA-motif	*Hordeum vulgare*	325	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
CGTCA-motif	*Hordeum vulgare*	382	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
CGTCA-motif	*Hordeum vulgare*	60	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
CGTCA-motif	*Hordeum vulgare*	505	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
G-box	*Zea mays*	263	+	6	CACGAC	cis-acting regulatory element involved in light responsiveness
G-box	*Arabidopsis thaliana*	722	-	6	TACGTG	cis-acting regulatory element involved in light responsiveness
G-box	*Zea mays*	203	-	6	CACGTC	cis-acting regulatory element involved in light responsiveness
LTR	*Hordeum vulgare*	543	+	6	CCGAAA	cis-acting element involved in low-temperature responsiveness
MYB	*Arabidopsis thaliana*	718	+	6	CAACCA	
MYB	*Arabidopsis thaliana*	376	-	6	CAACAG	
Myb-binding site	*Nicotiana tabacum*	376	-	6	CAACAG	
TATA-box	*Arabidopsis thaliana*	413	-	4	TATA	core promoter element around -30 of transcription start
TATA-box	*Arabidopsis thaliana*	86	+	4	TATA	core promoter element around -30 of transcription start
TGA-element	*Brassica oleracea*	711	-	6	AACGAC	auxin-responsive element

	TGACG-motif	*Hordeum vulgare*	325	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
	TGACG-motif	*Hordeum vulgare*	382	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
	TGACG-motif	*Hordeum vulgare*	60	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
	TGACG-motif	*Hordeum vulgare*	505	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
	Unnamed__1	*Zea mays*	262	-	5	CGTGG	
	Unnamed__1	*Zea mays*	689	-	5	CGTGG	
	Unnamed__1	*Zea mays*	561	-	5	CGTGG	
	Unnamed__1	*Zea mays*	721	-	5	CGTGG	
	Unnamed__2	*Zea mays*	41	+	6	CCCCGG	
	Unnamed__4	*Petroselinum hortense*	28	-	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	117	+	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	31	-	4	CTCC	
	Unnamed__4	*Petroselinum hortense*	373	+	4	CTCC	
	as-1	*Arabidopsis thaliana*	60	+	5	TGACG	
	as-1	*Arabidopsis thaliana*	382	-	5	TGACG	
	as-1	*Arabidopsis thaliana*	325	-	5	TGACG	
	as-1	*Arabidopsis thaliana*	505	-	5	TGACG	

Table S 16. Putative p-inducible gene promoter sequence analysis using TSSP[2,3].

It includes the gene ID, sequence length, number of promoters predicted by the software, position of the promoter in the query sequence, transcription factor binding sites along with position of the conserved binding sites.

Promoter name	*Gene Id*	*Length of Query sequence*	*No of promoters predicted*	*Position*	*Transcription factor binding sites*
endonuclease promoter	EWM28686.1	606 bp	1	558bp	472 (+) RSP00004 tagaCACGTaga
					481 (-) RSP00010 cACGTG
					484 (-) RSP00011 ctccACGTGgt
					271 (-) RSP00016 caTGCAC
					424 (+) RSP00026 gcttttgaTGACtTcaaacac
					480 (-) RSP00065 ACGTGgcgc
					480 (-) RSP00066 ACGTGccgc
					481 (-) RSP00069 tACGTG
					269 (+) RSP00102 aaaTGACGaaaatgc
					281 (-) RSP00114 CTTGTCGTCA
					368 (-) RSP00148 CGACG
					357 (-) RSP00161 WAAAG
					482 (-) RSP00186 cTACGTGgcca
					336 (+) RSP00241 CGGTCA
					355 (+) RSP00304 TTTTTTCC
					452 (-) RSP00308 CAACA
					273 (-) RSP00309 CACCTG
					481 (-) RSP00069 TACGTG

					476 (+) RSP00427 CACGTA
					378 (-) RSP00463 atttcatggCCGACctgcttttt
					378 (-) RSP00464 acttgatggCCGACctctttttt
					378 (-) RSP00465 aatatactaCCGACcatgagttct
					373 (-) RSP00466 actaCCGACatgagttccaaaaagc
					478 (-) RSP00469 GNGGTG
					473 (-) RSP00469 GNGGTG
					478 (-) RSP00470 GTGGNG
					349 (+) RSP00483 GCCGC
					351 (+) RSP00508 gcaTTTTTatca
					352 (+) RSP00508 gcaTTTTTatca
					353 (+) RSP00508 gcaTTTTTatca
SPS promoter	EWM20309.1	995 bp	0		
GPDP promoter	EWM28518.1	999 bp	1	569 bp	459 (+) RSP00016 caTGCAC
					561 (+) RSP00092 TAACAAA
					402 (+) RSP00102 aaaTGACGaaaatgc
					453 (-) RSP00102 aaaTGACGaaaatgc
					429 (+) RSP00148 CGACG
					280 (+) RSP00161 WAAAG
					542 (+) RSP00161 WAAAG
					520 (-) RSP00161 WAAAG
					449 (+) RSP00252 CACATG
					515 (+) RSP00265 ACTTTA
					390 (+) RSP00269 atcttatgtcattgaTGACGacctcc
					465 (-) RSP00269 atcttatgtcattgaTGACGacctcc
					515 (-) RSP00304 TTTTTTCC
					379 (+) RSP00308 CAACA

					459 (+) RSP00327 CATGCA
					273 (-) RSP00398 TTTGAA
					319 (+) RSP00404 TTTGCGT
					274 (-) RSP00463 atttcatggCCGACctgcttttt
					274 (-) RSP00464 acttgatggCCGACctcttttttt
					274 (-) RSP00465 aatatactaCCGACcatgagttct
					269 (-) RSP00466 actaCCGACatgagttccaaaaagc
					542 (-) RSP00477 TTTAA
					474 (-) RSP00477 TTTAA
					548 (-) RSP00508 gcaTTTTTatca
					547 (-) RSP00508 gcaTTTTTatca
					518 (-) RSP00508 gcaTTTTTatca
					517 (-) RSP00508 gcaTTTTTatca
					394 (-) RSP00508 gcaTTTTTatca
					368 (-) RSP00508 gcaTTTTTatca
					367 (-) RSP00508 gcaTTTTTatca

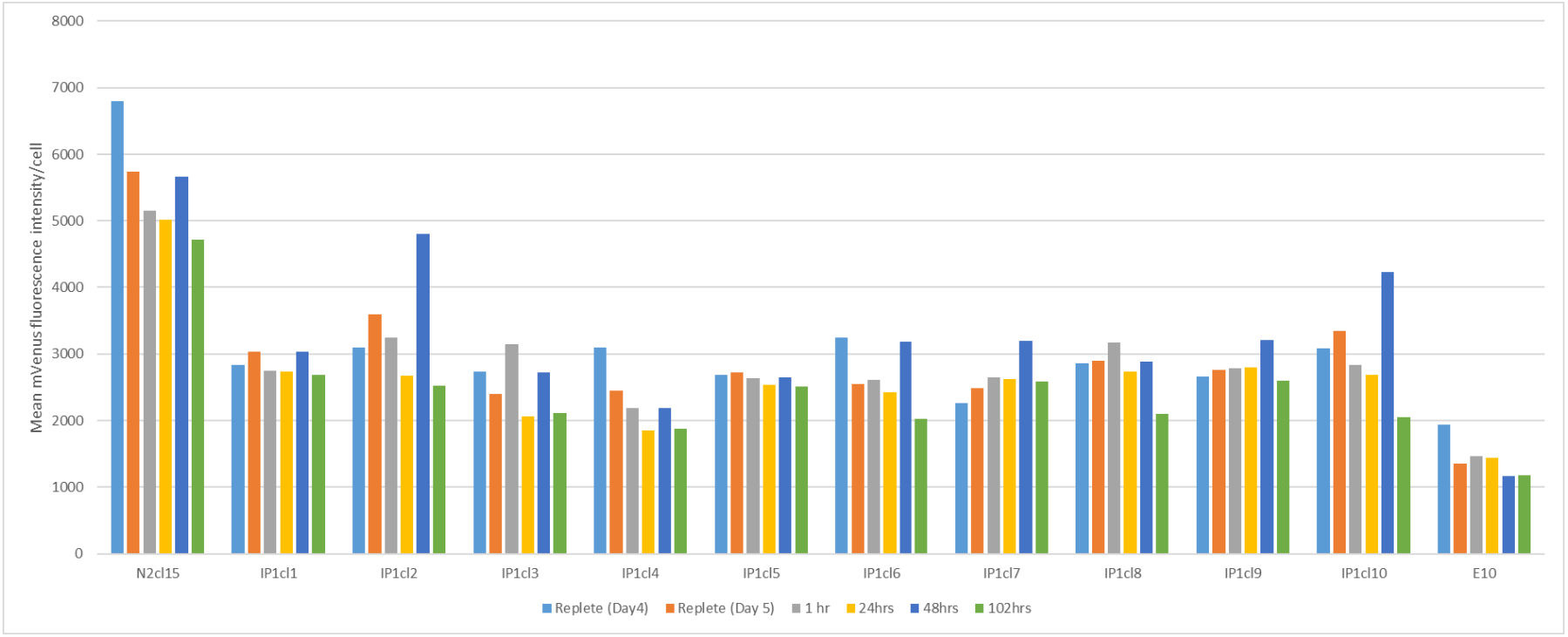


Figure S 11. Screening of GDPD, EWM28518.1 transgenic *N. gaditana* lines.

Mean mVenus fluorescence intensity per cell observed in all the transgenic lines (IP1cl1-10) tested including mVenus positive control (N2cl15) and empty vector control (E10) that does not contain the mVenus cassette, under phosphate replete (day 4- light blue and day 5- organge) and deplete (day 5- grey, day 6- yellow, day 7- dark blue and day 8- green) conditions.

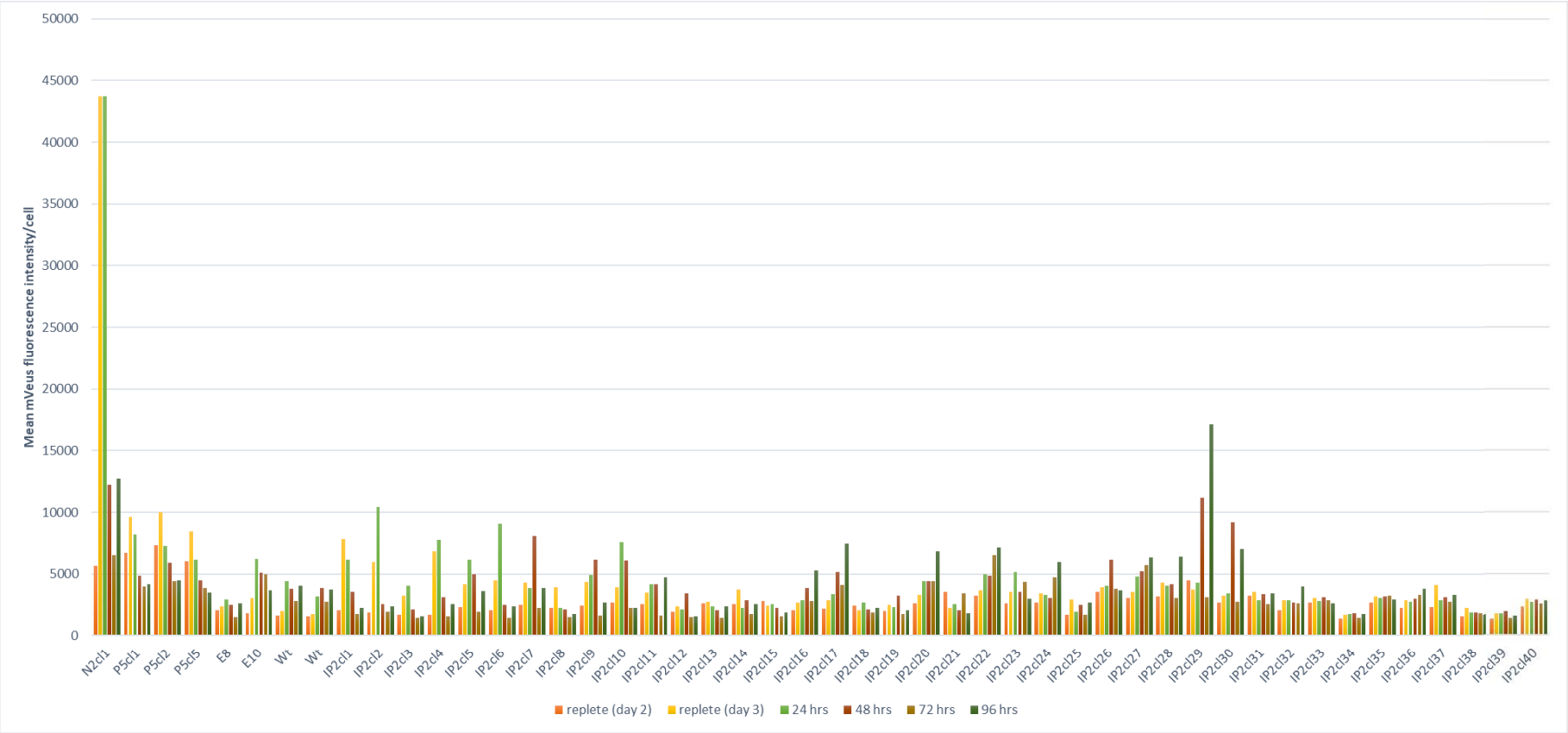


Figure S 12. Screening of SPS, EWM20309.1 transgenic *N. gaditana* lines.

Mean mVenus fluorescence intensity per cell observed in all the transgenic lines (IP2cl1-40) tested including mVenus positive control (N2cl15), empty vector control (E8 and 10) that does not contain the mVenus cassette and wild type, under phosphate replete (day 2- orange and day 3- yellow) and deplete (day 4- green, day 5- brown, day 6- light brown and day 7- dark green) conditions.

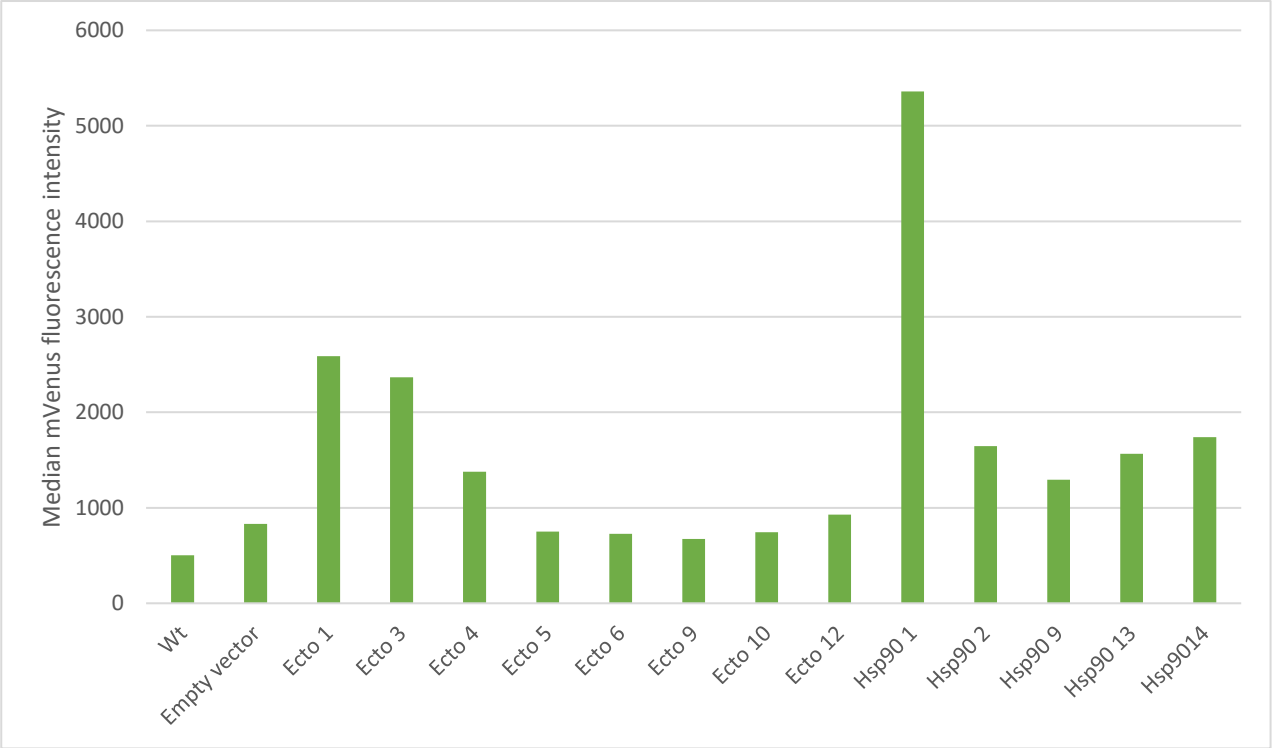


Figure S 13. Screening of transgenic *N. gaditana* lines carrying the viral promoter (Ecto).

Median mVenus fluorescence intensity observed in all the transgenic lines tested (Ecto 1, 3-6, 9, 10, 12) and the control transgenic N. gaditana lines expressing mVenus under the control of Hsp90 promoter (Hsp90 1, 2, 13 and 14) including empty vector control and wild type.

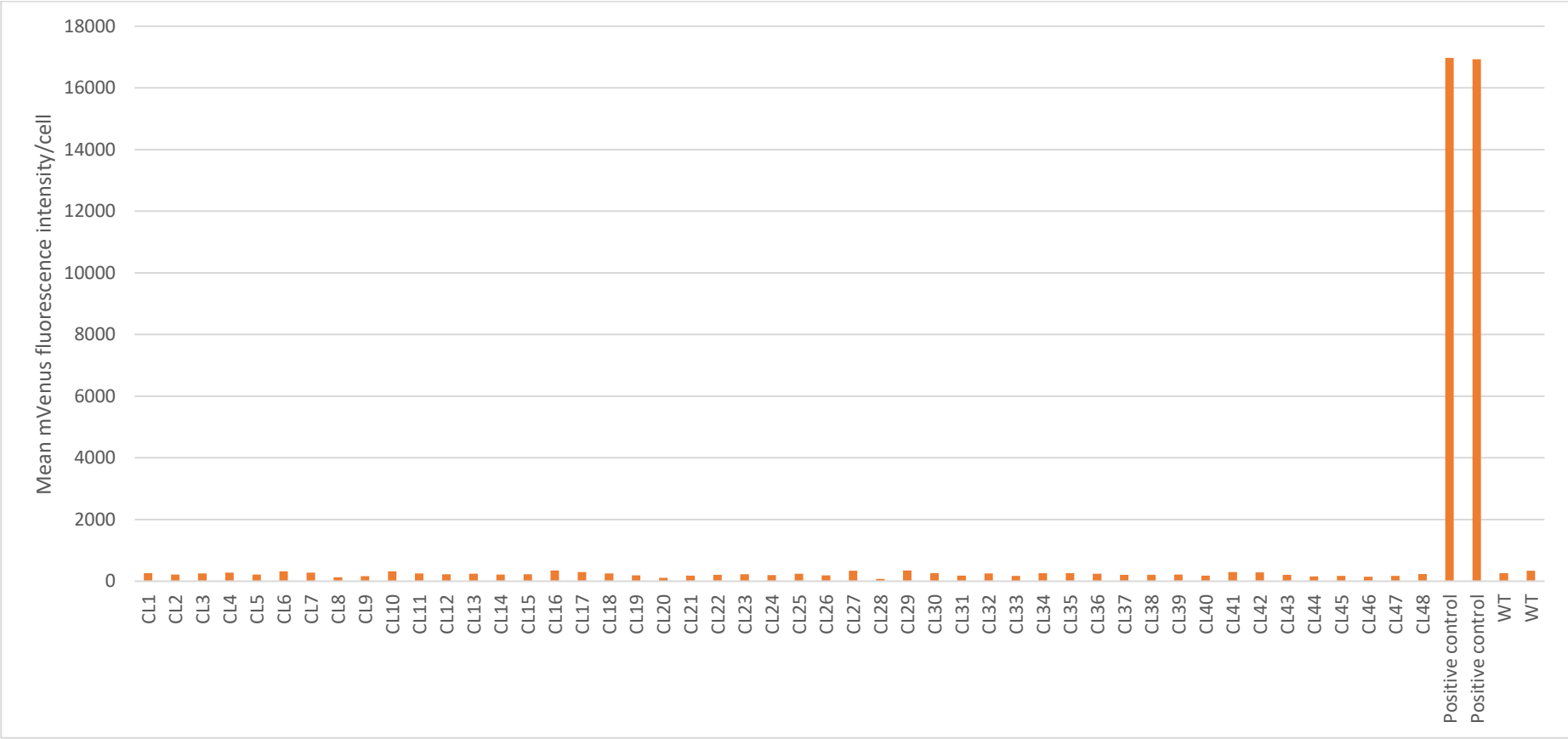


Figure S 14. Screening of transgenic *P. tricornutum* lines carrying the viral promoter (Ecto).

Mean mVenus fluorescence intensity observed in all the transgenic lines tested (CL 1- 48) and the control transgenic *P. tricornutum* line expressing mVenus under the control of 49202 Pt promoter including wild type.

References

1. Lescot, M. *et al.* PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **30,** 325–327 (2002).

2. Solovyev, V. V, Shahmuradov, I. A. & Salamov, A. A. Identification of Promoter Regions and Regulatory Sites. in *Computational Biology of Transcription Factor Binding* (ed. Ladunga, I.) 57–83 (Humana Press, 2010).

3. Shahmuradov, I. A., Umarov, R. K. & Solovyev, V. V. TSSPlant: A new tool for prediction of plant Pol II promoters. *Nucleic Acids Res.* **45,** (2017).

APPENDIX 3

Chapter 4 supplementary data

Genetic engineering of the microalga *Nannochloropsis gaditana* for the production of sesquiterpene β-caryophyllene

Contributors: Margaret Ramarajan[a], Michele Fabris[a,b], Raffaela M. Abbriano[a] , Unnikrishnan Kuzhiumparambil[a] , Peter J. Ralph[a]

[a] University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo NSW 2007, Australia

[b] CSIRO Synthetic Biology Future Science Platform, GPO box 2583, Brisbane QLD 4001, Australia

---

The following supporting information is available for this chapter:

Introduction:

Materials and methods:

Results and discussion:

Figure S23 : The growth, photosynthetic health of transgenic *N. gaditana*

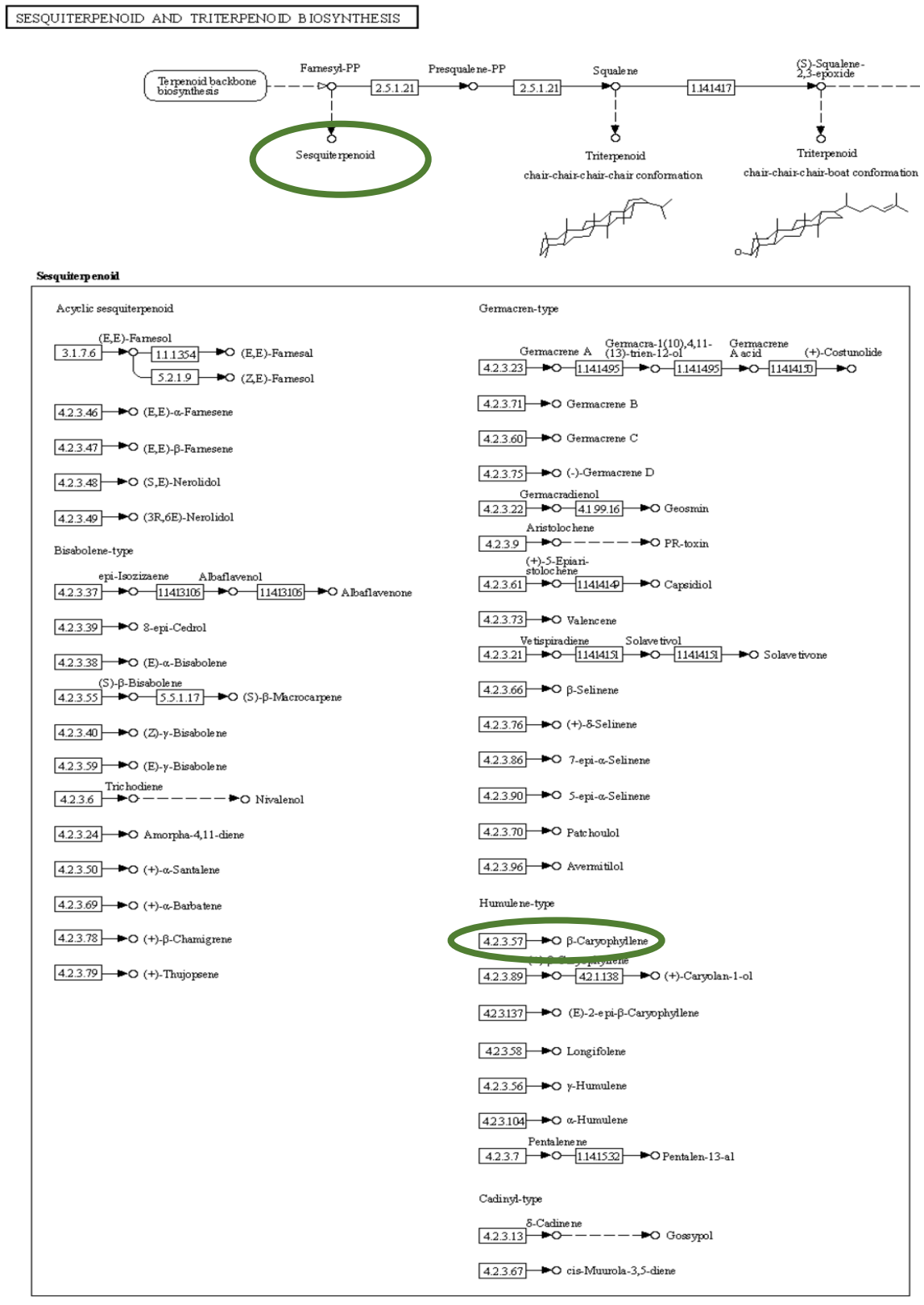


Figure S 15. KEGG pathway- Sesquiterpenoid and Triterpeneoid pathway.

The sesquiterpenoids are derived from Farnesyl-PP substrate using single or double or rarely multi-enzymatic reactions. The terpenoid synthase (EC 4.2.3.57) selected in this study and its product are highlighted.

Materials and methods:

Table S 17. **MEP pathway genes in *A. thaliana* and their orthologs in *N. gaditana* obtained using the *N. gaditana* genome browser[1].**

The protein sequences of MEP pathway enzymes from *A. thaliana* blasted against the *N. gaditana* genome using blastp tool and the corresponding orthologous gene, its % identity, length of the *N. gaditana* protein sequence, reference coordinates, e-value, alignment score, function, EC number and gene expression (RPKM) under normal conditions and up or down regulation under nitrogen starvation were collected from the output file.

NCBI		*N. gaditana* genome browser								Gene expression (RPKM)		Under nitrogen depletion
Gene	**NCBI protein accession number_A. thaliana**	**Subject gene**	**% id**	**Length**	**e-value**	**Score**	**Co-ordinates**	**Function**	**EC number**	**Day 3**	**Day 6**	**Up/Down**
CLA1/DXS	NP_193291.1	Naga_100064g23.1	57.27%	674	0	714	NG_scf03 (222 468-226305)	1-deoxy-d-xylulose 5-phosphate synthase	EC:2.2.1.7	86.435	97.01	up
DXR	NP_001190600.1	Naga_101672g1.1	55.17%	58	8.00E-11	64.7	NG_chr14 (693 759-694385)	hypothetical protein		30.805	22.265	down
DXR	NP_201085.1	Naga_101672g1.1	55.17%	58	8.00E-11	64.7	NG_chr14 (693 759-694385)	hypothetical protein		30.805	22.265	down
MTC/CMS	NP_565286.1	Naga_100067g6.1	53.39%	221	7.00E-62	233	NG_scf02 (285 926-287690)	2-c-methyl-d-erythritol 4-phosphate cytidylyltransferase	EC:2.7.7.60	49.922	20.778	down
MTC/CMS	NP_001325418.1	Naga_100067g6.1	53.39%	221	7.00E-62	233	NG_scf02 (285 926-287690)	2-c-methyl-d-erythritol 4-phosphate cytidylyltransferase	EC:2.7.7.60	49.922	20.778	down
CMK	NP_180261.1	Naga_100247g7.1	46.25%	253	8.00E-47	184	NG_chr02 (267 955-269257)	4-diphosphocytidyl-2c-methyl-d-erythritol kinase	EC:2.7.1.148	25.398	13.349	no change
MDS/MCS	NP_850971.1	Naga_100484g5.1	64.33%	157	5.00E-56	213	NG_chr18 (132 656-133625)	2-c-methyl-d-erythritol - cyclodiphosphate synthase	EC:4.6.1.12	17.493	12.336	up
MDS/MCS	NP_001319313.1	Naga_100484g5.1	68.97%	87	5.00E-32	132	NG_chr18 (132 656-133625)	2-c-methyl-d-erythritol - cyclodiphosphate synthase	EC:4.6.1.12	17.493	12.336	up

HDS	NP_001332136.1	Naga_102438g1.1	45.76%	59	8.00E-10	62	NG_chr12 (157601-158159)	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	EC:1.17.7.1	45.38	32.179	no change
HDS	NP_001119467.1	Naga_102438g1.1	45.76%	59	8.00E-10	62	NG_chr12 (157601-158159)	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	EC:1.17.7.1	45.38	32.179	no change
HDS	NP_851233.1	Naga_102438g1.1	45.76%	59	1.00E-09	62	NG_chr12 (157601-158159)	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	EC:1.17.7.1	45.38	32.179	no change
HDS	NP_001332137.1	Naga_102438g1.1	45.76%	59	8.00E-10	62	NG_chr12 (157601-158159)	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	EC:1.17.7.1	45.38	32.179	no change
HDS	NP_200868.2	Naga_102438g1.1	45.76%	59	9.00E-10	62	NG_chr12 (157601-158159)	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	EC:1.17.7.1	45.38	32.179	no change
HDR	NP_567965.1	Naga_100079g5.1	58.44%	397	4.00E-138	487	NG_chr07 (933490-936027)	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	EC:1.17.1.2	213.108	408.706	no change
GGPPS11 (At4g36810)	NP_195399.1	Naga_100002g113.1	58.00%	300	5.00E-85	310	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
		Naga_100072g15.1	32.95%	258	7.00E-21	97.8	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
		Naga_100004g170.1	29.72%	286	1.00E-20	97.1	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change
GGPPS3 (At2g18640)	NP_179454.1	Naga_100002g113.1	49.16%	299	3.00E-67	251	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
GGPPS7 (At3g14550)	NP_001319550.1	Naga_100002g113.1	54.52%	299	1.00E-74	276	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
		Naga_100072g15.1	32.42%	256	2.00E-16	82.8	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
		Naga_100004g170.1	30.38%	260	3.00E-16	82.4	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change
GGPPS7 (At3g14550)	NP_001326728.1	Naga_100002g113.1	54.52%	299	1.00E-74	276	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
		Naga_100072g15.1	32.42%	256	2.00E-16	82.8	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
		Naga_100004g170.1	30.38%	260	3.00E-16	82.4	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change

GGPPS7 (At3g14550)	NP_188073.2	Naga_100002g113.1	54.52%	299	1.00E-74	276	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
		Naga_100072g15.1	32.42%	256	2.00E-16	82.8	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
		Naga_100004g170.1	30.38%	260	3.00E-16	82.4	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change
GGPPS6 (At3g14530)	NP_188071.1	Naga_100002g113.1	55.18%	299	7.00E-76	280	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
		Naga_100004g170.1	30.74%	244	4.00E-16	81.6	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change
		Naga_100072g15.1	32.03%	256	2.00E-15	79.3	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
GGPPS2 (At2g18620)	NP_179452.1	Naga_100002g113.1	57.19%	299	7.00E-88	320	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
		Naga_100004g170.1	31.08%	251	2.00E-20	95.9	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change
		Naga_100072g15.1	29.15%	271	5.00E-18	88.2	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
GGPPS8 (At3g20160)	NP_001325827.1	Naga_100002g113.1	46.46%	297	1.00E-61	233	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
GGPPS8 (At3g20160)	NP_188651.1	Naga_100002g113.1	50.67%	300	1.00E-65	246	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
GGPPS9 (At3g29430)	NP_189589.1	Naga_100002g113.1	54.52%	299	1.00E-71	266	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
		Naga_100004g170.1	29.55%	264	2.00E-14	76.6	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change
		Naga_100072g15.1	31.52%	257	5.00E-13	71.6	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
GGPPS10 (At3g32040)	NP_189747.1	Naga_100002g113.1	53.18%	299	6.00E-65	244	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
GGPPS1 (At1g49530)	NP_175376.1	Naga_100002g113.2	47.35%	302	6.00E-64	240	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
		Naga_100004g170.1	31.01%	258	3.00E-16	82	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change

GGPPS4 (At2g23800)	NP_179960.1	Naga_100002g113.2	52.51%	299	2.00E-75	278	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
		Naga_100072g15.1	29.63%	297	1.00E-16	83.6	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
		Naga_100004g170.1	28.67%	279	2.00E-15	79.3	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change
GGPPS12 (At4g38460)	NP_195558.1	Naga_100002g113.2	44.93%	227	4.00E-49	191	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
FPS1 (AT5g47770)	NP_199588.1	Naga_100019g42.1	46.72%	351	3.00E-94	341	NG_scf06 (163780-165539)	farnesyl pyrophosphate synthase	EC:2.5.1.10	57.462	30.639	down
FPS2 (AT4g17190)	NP_193452.1	Naga_100019g42.1	47.98%	346	2.00E-95	345	NG_scf06 (163780-165539)	farnesyl pyrophosphate synthase	EC:2.5.1.10	57.462	30.639	down
GPS1 (At2g34630)	NP_001031483.1	Naga_100072g15.1	42.36%	347	2.00E-66	249	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
		Naga_100004g170.1	44.79%	288	3.00E-64	241	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change
		Naga_100002g113.1	34.78%	184	3.00E-19	92.4	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
GPS1 (At2g34630)	NP_850234.1	Naga_100004g170.1	46.95%	262	8.00E-61	230	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change
		Naga_100072g15.1	42.64%	326	1.00E-60	229	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
		Naga_100002g113.2	34.78%	184	4.00E-19	91.7	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
SPS1 (At1g78510)	NP_177972.2	Naga_100004g170.1	41.70%	271	3.00E-53	205	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change
		Naga_100072g15.1	37.11%	353	4.00E-53	204	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
		Naga_100002g113.1	34.25%	219	2.00E-24	110	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
SPS1 (At1g78510)	NP_001077840.1	Naga_100004g170.1	41.70%	271	2.00E-53	206	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change
		Naga_100072g15.1	39.20%	324	6.00E-53	204	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
		Naga_100002g113.1	34.25%	219	1.00E-24	110	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
SPS2 (At1g17050)	NP_173148.2	Naga_100004g170.1	39.19%	296	3.00E-55	212	NG_chr05 (586361-588558)	decaprenyl-diphosphate synthase subunit 1	EC:2.5.1.1	40.788	45.686	no change

		Naga_100072g15.1	39.93%	298	9.00E-54	207	NG_chr09 (745820-747907)	Polyprenyl synthetase	EC:2.5.1.1	51.663	50.139	up
		Naga_100002g113.1	33.18%	223	9.00E-25	110	NG_chr04 (447813-451347)	geranylgeranyl pyrophosphate synthase	EC:2.5.1.10, EC:2.5.1.29, EC:2.5.1.1	15.989	10.02	no change
IDI/IPP1 ((At5g16440)	NP_197148.3	Naga_100093g4.1	40.91%	264	3.00E-48	188	NG_chr09 (437368-440042)	farnesyl-diphosphate farnesyltransferase	EC:2.5.1.21, EC:5.3.3.2	24.515	13.734	up
		Naga_100015g76.1	34.98%	283	4.00E-40	160	NG_chr06 (681709-683050)	isopentenyl-diphosphate delta-isomerase	EC:5.3.3.2	100.366	156.391	up
IDI/IPP2 ((At3g02780)	NP_186927.1	Naga_100093g4.1	42.15%	242	7.00E-48	186	NG_chr09 (437368-440042)	farnesyl-diphosphate farnesyltransferase	EC:2.5.1.21, EC:5.3.3.2	24.515	13.734	up
		Naga_100015g76.1	36.56%	279	1.00E-40	162	NG_chr06 (681709-683050)	isopentenyl-diphosphate delta-isomerase	EC:5.3.3.2	100.366	156.391	up
IDI/IPP2 (At3g02780)	NP_001325698.1	Naga_100093g4.1	42.54%	228	3.00E-45	177	NG_chr09 (437368-440042)	farnesyl-diphosphate farnesyltransferase	EC:2.5.1.21, EC:5.3.3.2	24.515	13.734	up
		Naga_100015g76.1	38.99%	218	3.00E-38	154	NG_chr06 (681709-683050)	isopentenyl-diphosphate delta-isomerase	EC:5.3.3.2	100.366	156.391	up
IPP1	AAC49932.1	Naga_100093g4.1	42.15%	242	2.00E-47	185	NG_chr09 (437368-440042)	farnesyl-diphosphate farnesyltransferase	EC:2.5.1.21, EC:5.3.3.2	24.515	13.734	up
		Naga_100015g76.1	35.34%	266	4.00E-40	160	NG_chr06 (681709-683050)	isopentenyl-diphosphate delta-isomerase	EC:5.3.3.2	100.366	156.391	up
IPP2	AAC49920.1	Naga_100093g4.1	42.15%	242	7.00E-48	186	NG_chr09 (437368-440042)	farnesyl-diphosphate farnesyltransferase	EC:2.5.1.21, EC:5.3.3.2	24.515	13.734	up
		Naga_100015g76.1	36.56%	279	1.00E-40	162	NG_chr06 (681709-683050)	isopentenyl-diphosphate delta-isomerase	EC:5.3.3.2	100.366	156.391	up

Table S 18. MEP pathway genes in *A. thaliana* and their orthologs in *N. gaditana*.

The protein sequences of MEP pathway enzymes from A. thaliana and *N. gaditana* were analysed using DeppLoc[2], TargetP[3], ChloroP[4] and TMHMM[5] transmembrane and sub-cellular location prediction tools. ASAFind[6,7] was used only for *N. gaditana* protein sequences to identify nuclear-encoded plastid proteins as it is specific for algae with secondary plastids. The table also provides the Query coverage, percentage identity and E value of the *N. gaditana* sequence when using A. thaliana protein sequence query along with the confirmed sub cellular locations of *A. thaliana* enzymes and their citations.

A. thaliana Gene	NCBI mRNA accession number	NCBI protein accession number	DeepLoc.1 prediction	TargetP	Chloro P	TMHMM predicted domains	Confirmed Location	Query cover	% identity	E value	*N.gaditana* NCBI accession number	*ASAFind*	DeepLoc.1 prediction	TargetP	Chloro P	TMHMM predicted domains	Citation
CLA1/DXS	NM_117647.3	NP_193291.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	89%	57.72%	0	XP_005854509.1 EWM22144.1	No No	Chloroplast	Mitochondrion	Yes Yes	0	[8,9]
DXR	NM_001203671.1	NP_001190600.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	11% 7%	55.17% 68.57%	1e-13 2e-07	EWM24174.1 EWM24175.1	No No	Nuclear Mitochondrion	Nil Mitochondrion	No No	0 1	[9,10]
DXR	NM_125674.3	NP_201085.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	12% 7%	55.17% 68.57%	1e-13 2e-07	EWM24174.1 EWM24175.1	No No	Nuclear Mitochondrion	Nil Mitochondrion	No No	0 1	[9,10]
MTC/CMS	NM_126305.3	NP_565286.1	Chloroplast	Mitochondrion	Yes	0	Chloroplast	77%	52.79%	5e-79	EWM22399.1	Chloroplast	Chloroplast	secretory	Yes	1	[9,11]
MTC/CMS	NM_001335124.1	NP_001325418.1	Chloroplast	Mitochondrion	Yes	0	Chloroplast	77%	52.79%	5e-79	EWM22399.1	Chloroplast	Chloroplast	secretory	Yes	1	[9,11]
CMK	NM_128250.4	NP_180261.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	68% 66%	45.72% 44.83%	1e-74 6e-70	EWM29944.1 XP_005853675.1	Chloroplast No	Chloroplast Chloroplast	secretory nil	Yes Yes	1 0	[9,11]
MDS/MCS	NM_180640.3	NP_850971.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	67% 67%	64.33% 63.69%	5e-71 2e-70	EWM23179.1 XP_005854081.1	Chloroplast Chloroplast	Chloroplast Chloroplast	secretory secretory	No No	1 0	[9,12]
MDS/MCS	NM_001334142.1	NP_001319313.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	54% 54%	68.97% 67.82%	2e-40	EWM23179.1 XP_005854081.1	Chloroplast	Chloroplast Chloroplast	secretory secretory	No No	1 0	[9,12]

										9e-40		Chloroplast					
HDS	NM_001345427.1	NP_001332136.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	8%	45.76%	2e-11	EWM25176.1	No	Cytoplasm	nil	No	0	[9,13]
HDS	NM_001125995.2	NP_001119467.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	8%	45.76%	2e-11	EWM25176.1	No	Cytoplasm	nil	No	0	[9,13]
HDS	NM_180902.5	NP_851233.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	7%	45.76%	2e-11	EWM25176.1	No	Cytoplasm	nil	No	0	[9,13]
HDS	NM_001345428.1	NP_001332137.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	8%	45.76%	2e-11	EWM25176.1	No	Cytoplasm	nil	No	0	[9,13]
HDS	NM_125453.6	NP_200868.2	Chloroplast	Chloroplast	Yes	0	Chloroplast	7%	45.76%	2e-11	EWM25176.1	No	Cytoplasm	nil	No	0	[9,13]
HDR	NM_119600.4	NP_567965.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	84%	58.44%	1e-170	EWM26997.1	Chloroplast	Chloroplast	secretory	Yes	1	[9,14]
GGPPS11 (At4g36810)	NM_119845.3	NP_195399.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	80% 80% 61%	58.33% 58.33%32.43 %	8e-110 1e-109 2e-24	EWM28775.1 EWM28774.1 EWM28148.1	No Chloroplast No	Mitochondrion Chloroplast Mitochondrion	Mitochondrion Secretory Mitochondrion	Yes Yes No	0 0 0	[15,16]
GGPPS3 (At2g18640)	NM_127420.2	NP_179454.1	ER	Secretory	No	1	ER	79% 79% 56%	49.16% 49.16% 30.67%	6e-93 3e-92 8e-20	EWM28775.1 EWM28774.1 EWM28148.1	No Chloroplast No	Mitochondrion Chloroplast Mitochondrion	Mitochondrion Secretory Mitochondrion	Yes Yes No	0 0 0	[15,16]
GGPPS7 (At3g14550)	NM_001338122.1	NP_001319550.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	82% 82% 64% 64%	54.52% 54.52% 32.31% 34.38%	2e-102 5e-102 1e-24 7e-25	EWM28775.1 EWM28774.1 EWM28148.1 EWM26176.1	No Chloroplast No No	Mitochondrion Chloroplast Mitochondrion Chloroplast	Mitochondrion Secretory Mitochondrion Mitochondrion	Yes Yes No Yes	0 0 0 1	[15,16]
GGPPS7 (At3g14550)	NM_001338123.1	NP_001326728.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	82% 82% 64% 64%	54.52% 54.52% 32.31% 34.38%	2e-102 5e-102 1e-24 7e-25	EWM28775.1 EWM28774.1 EWM28148.1 EWM26176.1	No Chloroplast No No	Mitochondrion Chloroplast Mitochondrion Chloroplast	Mitochondrion Secretory Mitochondrion Mitochondrion	Yes Yes No Yes	0 0 0 1	[15,16]
GGPPS7 (At3g14550)	NM_112315.4	NP_188073.2	Chloroplast	Chloroplast	Yes	0	Chloroplast	82% 82% 64% 64%	54.52% 54.52% 32.31% 34.38%	2e-102 5e-102 1e-24 7e-25	EWM28775.1 EWM28774.1 EWM28148.1 EWM26176.1	No Chloroplast No No	Mitochondrion Chloroplast Mitochondrion Chloroplast	Mitochondrion Secretory Mitochondrion Mitochondrion	Yes Yes No Yes	0 0 0 1	[15,16]
GGPPS6 (At3g14530)	NM_112313.3	NP_188071.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	82% 82% 60% 64%	55.18% 55.18% 32.10% 33.98%	1e-104 5e-104 2e-24 1e-23	EWM28775.1 EWM28774.1 EWM28148.1 EWM26176.1	No Chloroplast No No	Mitochondrion Chloroplast Mitochondrion Chloroplast	Mitochondrion Secretory Mitochondrion Mitochondrion	Yes Yes No Yes	0 0 0 1	[16]
GGPPS2 (At2g18620)	NM_127418.2	NP_179452.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	85% 85% 64%	57.19% 57.19%31.08 %	8e-109 4e-108 3e-22	EWM28775.1 EWM28774.1 EWM28148.1	No Chloroplast No	Mitochondrion Chloroplast Mitochondrion	Mitochondrion Secretory Mitochondrion	Yes Yes No	0 0 0	[16]

GGPPS8 (At3g20160)	NM_001338457.1	NP_001325827.1	Chloroplast	Mitochondrion	No	0	Chloroplast	85% 85% 68%	46.46% 46.46% 28%	4e-82 3e-81 8e-15	EWM28775.1 EWM28774.1 EWM28148.1	No Chloroplast No	Mitochondrion Chloroplast Mitochondrion	Mitochondrion Secretory Mitochondrion	Yes Yes No	0 0 0	16
GGPPS8 (At3g20160)	NM_112907.2	NP_188651.1	Chloroplast	Mitochondrion	No	0	Chloroplast	86% 86% 71%	50.67% 50.67% 29.18%	3e-95 2e-94 5e-21	EWM28775.1 EWM28774.1 EWM28148.1	No Chloroplast No	Mitochondrion Chloroplast Mitochondrion	Mitochondrion Secretory Mitochondrion	Yes Yes No	0 0 0	16
GGPPS9 (At3g29430)	NM_113869.2	NP_189589.1	Chloroplast	Mitochondrion	Yes	0	Chloroplast	83% 83% 65% 64%	54.52% 54.52% 31.06% 33.46%	1e-100 6e-100 2e-23 8e-22	EWM28775.1 EWM28774.1 EWM28148.1 EWM26176.1	No Chloroplast No No	Mitochondrion Chloroplast Mitochondrion Chloroplast	Mitochondrion Secretory Mitochondrion Mitochondrion	Yes Yes No Yes	0 0 0 1	16
GGPPS10 (At3g32040)	NM_114027.3	NP_189747.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	82% 82% 65% 70%	53.18% 53.18% 31.82% 32.13%	7e-98 3e-97 9e-24 2e-22	EWM28775.1 EWM28774.1 EWM28148.1 EWM26176.1	No Chloroplast No No	Mitochondrion Chloroplast Mitochondrion Chloroplast	Mitochondrion Secretory Mitochondrion Mitochondrion	Yes Yes No Yes	0 0 0 1	16
GGPPS1 (At1g49530)	NM_103841.3	NP_175376.1	Mitochondrion	Mitochondrion	No	0	Mitochondria	89% 89% 69%	47.35% 47.35% 31.01%	2e-85 9e-85 2e-21	EWM28775.1 EWM28774.1 EWM28148.1	No Chloroplast No	Mitochondrion Chloroplast Mitochondrion	Mitochondrion Secretory Mitochondrion	Yes Yes No	0 0 0	15–17
GGPPS4 (At2g23800)	NM_127943.3	NP_179960.1	ER	Secretory	Yes	1	ER	78% 78% 65%	52.51% 52.51% 28.78%	9e-98 9e-97 1e-19	EWM28775.1 EWM28774.1 EWM28148.1	No Chloroplast No	Mitochondrion Chloroplast Mitochondrion	Mitochondrion Secretory Mitochondrion	Yes Yes No	0 0 0	15,16
GGPPS12 (At4g38460)	NM_120007.4	NP_195558.1	Chloroplast	Chloroplast	Yes	0	Chloroplast	67% 84% 47% 45%	44.93% 39.40% 31.90% 29.94%	4e-59 3e-58 6e-11 9e-09	EWM28775.1 EWM28774.1 EWM28148.1 EWM26176.1	No Chloroplast No No	Mitochondrion Chloroplast Mitochondrion Chloroplast	Mitochondrion Secretory Mitochondrion Mitochondrion	Yes Yes No Yes	0 0 0 1	16
FPS1 (AT5g47770)	NM_124151.3	NP_199588.1	Mitochondrion	Mitochondrion	No	0	Cytosol/Mitochondria	89%	46.72%	6e-116	EWM21458.1	No	Mitochondrion	Secretory	No	2	18,19
FPS2 (AT4g17190)	NM_117823.4	NP_193452.1	Cytoplasm	nil	No	0	Cytosol	99% 19% 19%	47.98% 29.85% 29.85%	6e-118 0.26 0.26	EWM21458.1 EWM21639.1 XP_005854209.1	No No No	Mitochondrion ER ER	Secretory Nil Secretory	No No No	2 8 8	18,19
GPS1 (At2g34630)	NM_001036406.3	NP_001031483.1	Mitochondrion	Mitochondrion	No	0	Mitochondria	80% 71% 40% 40%	42.36% 42.36% 34.78% 34.78%	4e-82 9e-76 4e-21 8e-21	EWM26176.1 EWM28148.1 EWM28775.1 EWM28774.1	No No No Chloroplast	Chloroplast Mitochondrion Mitochondrion Chloroplast	Mitochondrion Mitochondrion Mitochondrion Secretory	Yes No Yes Yes	1 0 0 0	20,21
GPS1 (At2g34630)	NM_179903.3	NP_850234.1	Cytoplasm	nil	No	0	Chloroplast	99% 85% 53% 53%	42.64% 44.10% 34.78% 34.78%	8e-76 1e-72 2e-21 7e-21	EWM26176.1 EWM28148.1 EWM28775.1 EWM28774.1	No No No Chloroplast	Chloroplast Mitochondrion Mitochondrion Chloroplast	Mitochondrion Mitochondrion Mitochondrion Secretory	Yes No Yes Yes	1 0 0 0	21,22
SPS1 (At1g78510)	NM_106498.4	NP_177972.2	ER	Chloroplast	Yes	0	ER	82% 66% 54% 54%	37.11% 39.60% 34.32% 34.32%	2e-64 1e-63	EWM26176.1 EWM28148.1 EWM28775.1 EWM28774.1	No No No Chloroplast	Chloroplast Mitochondrion Mitochondrion	Mitochondrion Mitochondrion	Yes No Yes Yes	1 0 0 0	23

										1e-28 2e-28			Chloroplast	Mitochondrion Secretory			
SPS1 (At1g78510)	NM_001084371.1	NP_001077840.1	Mitochondrion	Mitochondrion	No	0	-	82% 74% 60% 60%	39.20% 39.60% 34.32% 34.32%	1e-64 4e-64 5e-29 1e-28	EWM26176.1 EWM28148.1 EWM28775.1 EWM28774.1	No No No Chloroplast	Chloroplast Mitochondrion Mitochondrion Chloroplast	Mitochondrion Mitochondrion Mitochondrion Secretory	Yes No Yes Yes	1 0 0 0	23
SPS2 (At1g17050)	NM_101565.6	NP_173148.2	Chloroplast	Chloroplast	Yes	0	Chloroplast	76% 64% 50% 50%	38.05% 39.19% 33.18% 33.18%	2e-65 2e-64 1e-27 2e-27	EWM26176.1 EWM28148.1 EWM28775.1 EWM28774.1	No No No Chloroplast	Chloroplast Mitochondrion Mitochondrion Chloroplast	Mitochondrion Mitochondrion Mitochondrion Secretory	Yes No Yes Yes	1 0 0 0	23
IDI/IPP1 (At5g16440)	NM_121649.6	NP_197148.3	Chloroplast	Chloroplast	Yes	0	Chloroplast	96% 83% 15% 5%	38.87% 34.98% 44.44% 29.41%	2e-54 1e-47 4e-05 6.0	EWM26314.1 EWM27595.1 EWM25927.1 EWM28806.1	No Chloroplast Chloroplast No	Chloroplast Chloroplast Cytoplasm ER	Mitochondrion Secretory Mitochondrion Secretory	Yes Yes No No	3 0 1 1	24,25
IDI/IPP2 (At3g02780)	NM_111146.4	NP_186927.1	Chloroplast	Chloroplast	Yes	0	Mitochondria	79% 84% 15% 27% 9%	42.15% 36.56% 40.00% 27.16% 30.77%	3e-54 1e-48 6e-04 2.3 2.8	EWM26314.1 EWM27595.1 EWM25927.1 EWM26380.1 EWM28806.1	No Chloroplast Chloroplast No No	Chloroplast Chloroplast Cytoplasm Nucleus ER	Mitochondrion Secretory Mitochondrion Secretory nil	Yes Yes No Yes No	3 0 1 0 1	24,25
IDI/IPP2 (At3g02780)	NM_001337434.1	NP_001325698.1	Cytoplasm	nil	No	0	Cytosol	89% 76% 18% 33% 10%	42.54% 38.99% 40.00% 27.16% 30.77%	2e-51 8e-46 0.001 1.7 5.2	EWM26314.1 EWM27595.1 EWM25927.1 EWM26380.1 EWM28806.1	No Chloroplast Chloroplast No No	Chloroplast Chloroplast Cytoplasm Nucleus ER	Mitochondrion Secretory Mitochondrion Secretory nil	Yes Yes No Yes No	3 0 1 0 1	24,25
IPP1	U47324.1	AAC49932.1	Cytoplasm	nil	No	0	Chloroplast	96% 98% 19%	42.15% 35.34% 44.44%	3e-54 3e-48 8e-05	EWM26314.1 EWM27595.1 EWM25927.1	No Chloroplast Chloroplast	Chloroplast Chloroplast Cytoplasm	Mitochondrion Secretory Mitochondrion	Yes Yes No	3 0 1	24,25
IPP2	U49259.1	AAC49920.1	Chloroplast	Chloroplast	Yes	0	Mitochondria/Cytosol	79% 84% 15% 27% 9%	42.15% 36.56% 40.00% 27.16% 30.77%	3e-54 1e-48 6e-04 2.3 2.8	EWM26314.1 EWM27595.1 EWM25927.1 EWM26380.1 EWM28806.1	No Chloroplast Chloroplast No No	Chloroplast Chloroplast Cytoplasm Nucleus ER	Mitochondrion Secretory Mitochondrion Secretory nil	Yes Yes No Yes No	3 0 1 0 1	24,25

Table S 19. The PCR primers used in this study for amplification and to assemble the pNaga4.mV.DC.N2.QHS1 vector

Primer number	Sequence	F/ R	Binds to	Template used
MF758	tccaggggtaccaggaATGTCTGTTAAAGAAGAGA AAGTAATTC	F	QHS1	Synthetic QHS1 sequence from *Artemisia anna*
MF759	cccttgctcaccatgccTATAGGTATAGGATGAACG AGC	R		
MF803	TTTGGGCACTAGGTGTCTATTT	F	QHS1	pNaga4.mV.DC.N2. QHS1
MF804	CACAAGGGATTCTCGGGTTAT	R		
MF805	GGCAGACCAGTTTCTCATCTT	F	QHS1	pNaga4.mV.DC.N2. QHS1
MF806	CTACATCGTTGGTGAGGGATTC	R		
MF502	ccaaggatccatggtgagcaagggc	F	*Venus*	Screening
MF503	aaaagcggccgctcacttgtacagctcgtc	R	*Venus*	Screening
MF613	AAAGACCCCAACGAGAAGC	F	*Venus*	Screening
MF614	GTCCATGCCGAGAGTGATC	R	*Venus*	Screening

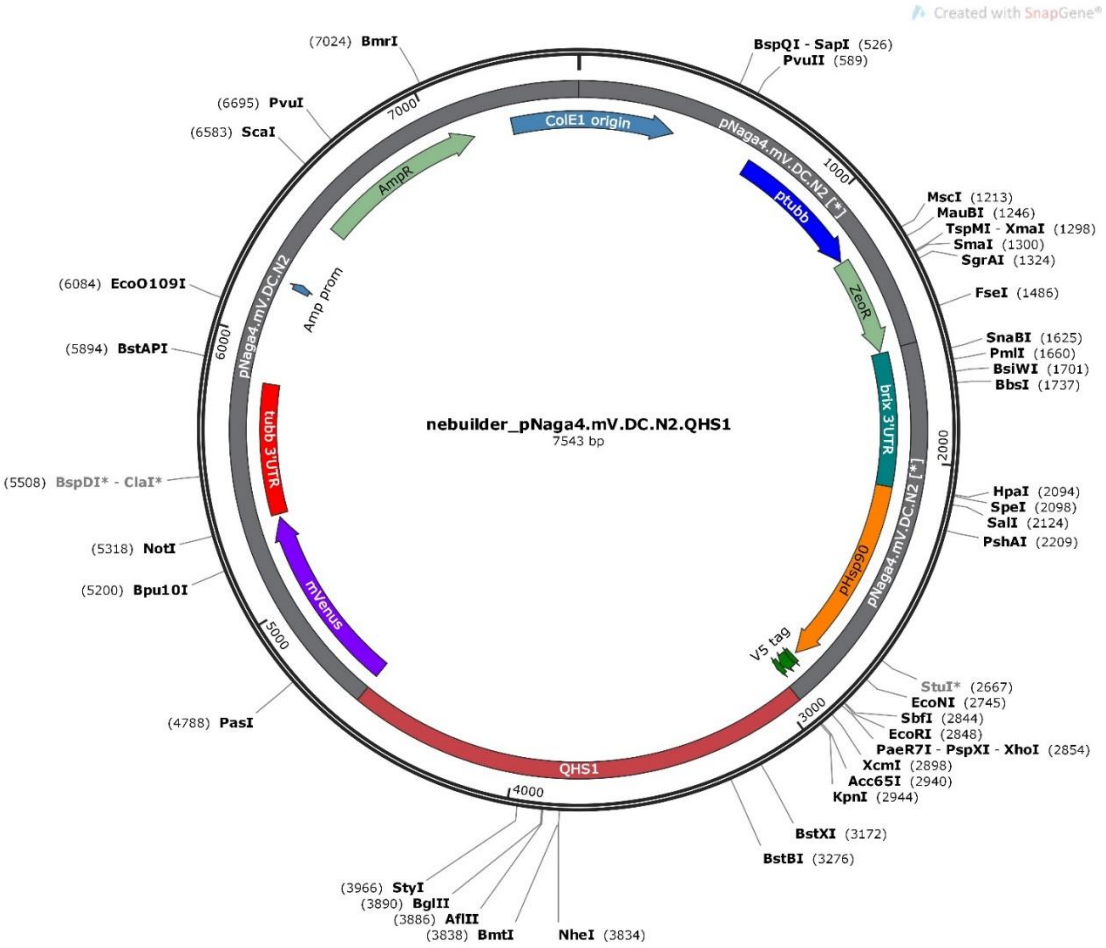


Figure S16. ***Nannochloropsis*** **transformation vector**

pNaga4.mV.DC.N2.QHS1 constructed using the pNaga4.mV.DC.N2 backbone carrying the antibiotic resistance marker *Sh ble* under the control of β-*tubulin* (TUB) promoter and (Brix domain) terminator and the *QHS1-mVenus* fluorescent reporter fusion under the control of a HSP90 promoter and β-*tubulin* (TUB) terminator.

Results and discussion:

Table S 20. Sesquiterpene genes in *A. thaliana* blasted against the *N. gaditana* genome to identify orthologs. No hits obtained.

Gene (TAIR)	NCBI Gene ID	NCBI protein accession number of *A. thaliana*	Blastp (*N. gaditana*)
Terpene synthase 21 (At5g23960)	832461	NP_001190374.1	no hits
Terpene synthase 21 (At5g23960)	832461	NP_197784.2	no hits
Sesquiterpene synthase (At5g44630)	834491	NP_199276.1	no hits

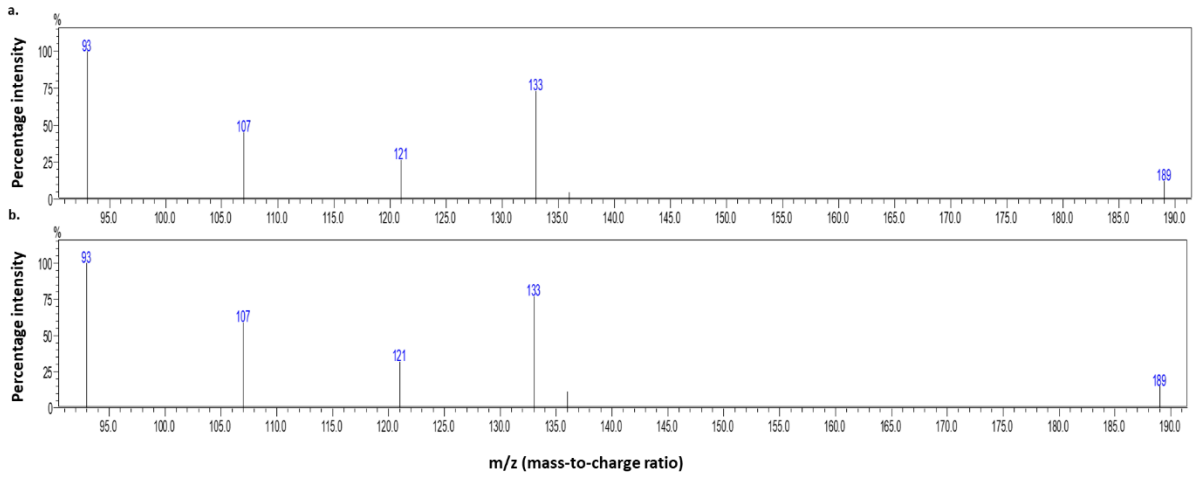


Figure S 17. The SIM spectra of β-caryophyllene.

a. The SIM spectra of *β-caryophyllene* standard (along with internal standard limonene). **b.** The SIM spectra of *β-caryophyllene* produced by the transgenic *E.coli* carrying pNaga4.mV.DC.N2.QHS1 (internal standard limonene)

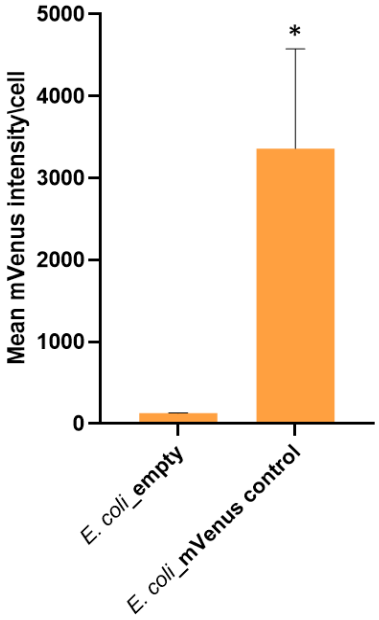


Figure S 18. Expression of mVenus reporter in *E. coli* carrying mVenus control plasmid pNaga4.mV.DC.N2.

The transgenic *E. coli* cells carrying the pNaga4.mV.DC.N2 control plasmid with mVenus reporter was found to express mVenus under the leaky control of Amp^r promoter.

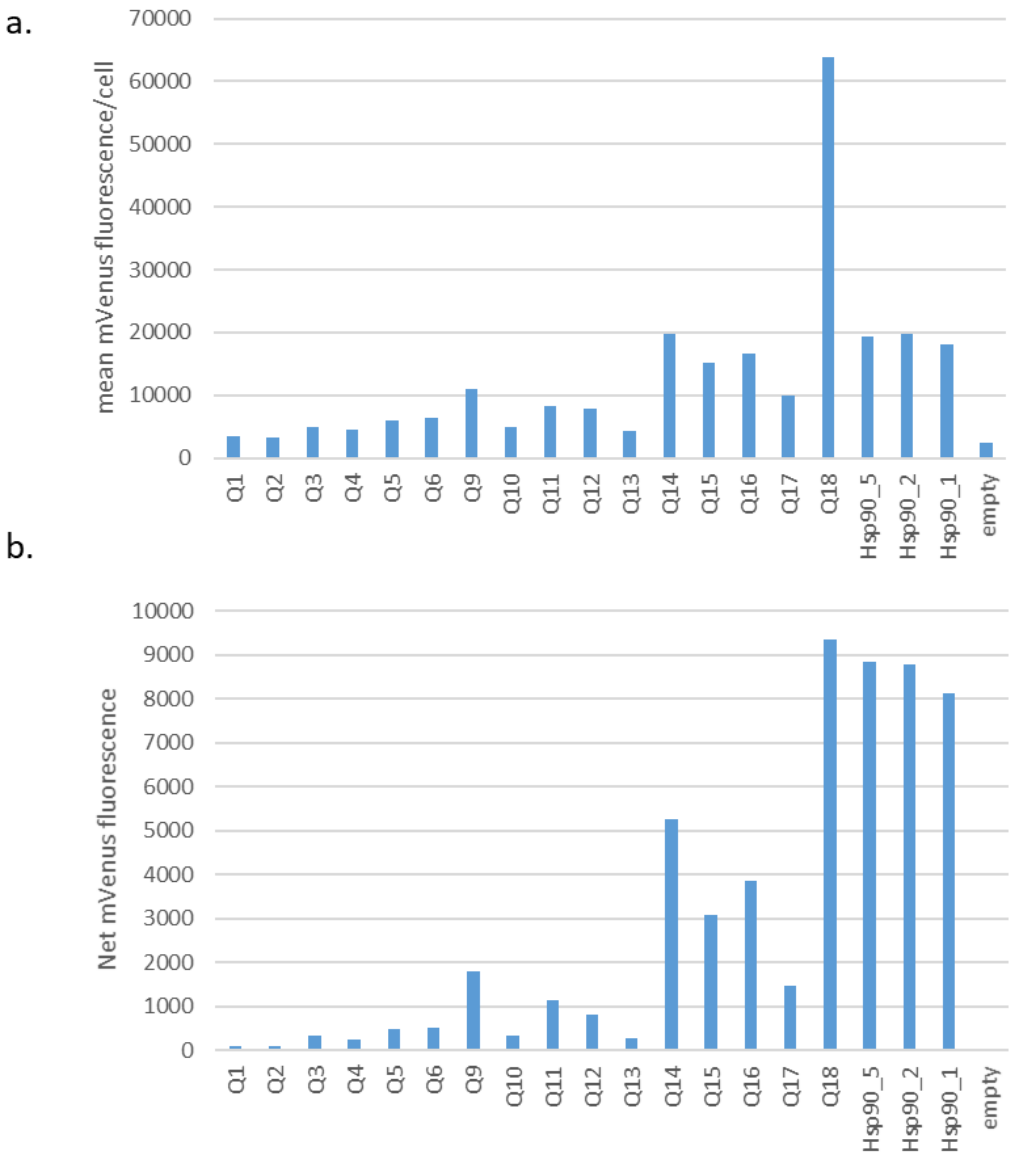


Figure S 19. Expression of QHS1-mVenus fusion protein in *N. gaditana* transformants.

a. Mean mVenus fluorescence per cell; **b.** Net mVenus fluorescence calculated as a product of mVenus intensity per cell, and number of cells expressing mVenus. Q1-Q18 are the *N. gaditana* transgenic lines carrying *QHS1-mVenus* fusion under the control of (*Hsp90*, Nga00934) promoter. Hsp90_1, Hsp90_2, Hsp90_5 are three control transgenic lines and empty is empty vector control carrying only the antibiotic (*Sh ble*) resistance cassette.

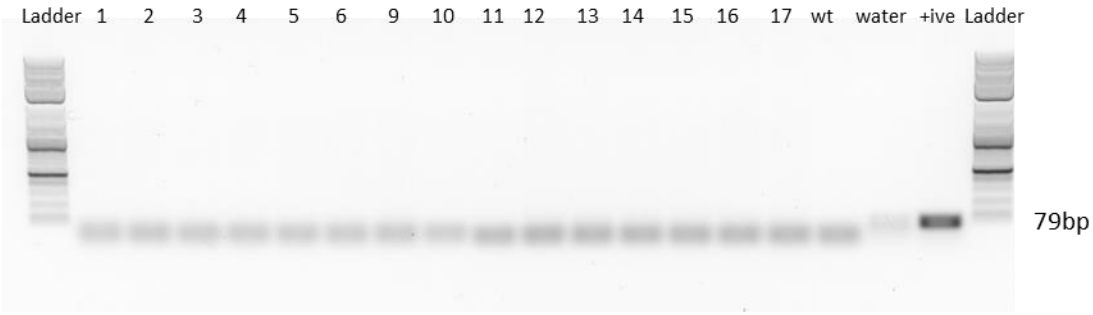


Figure S 20. PCR amplification of the mVenus reporter gene fragment in selected *N. gaditana* transformants.

Genomic DNA from wild-type, transgenic cell lines carrying mVenus (positive control) and transgenic cell lines carrying mVenus+ QHS1 (1-6,9-17) were isolated and analysed using PCR. Specific primers for mVenus were used (Table S15).

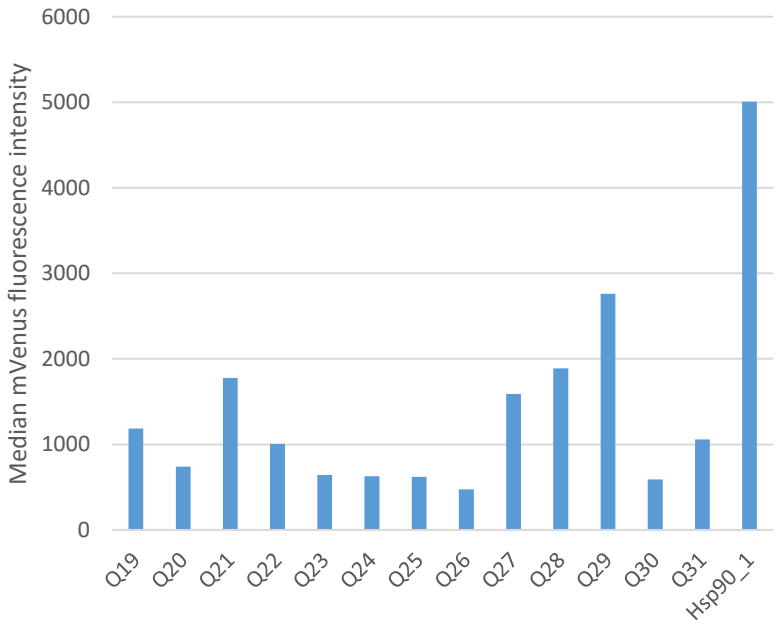


Figure S 21. Expression of QHS1-mVenus fusion protein in *N. gaditana* transformants

Median mVenus fluorescence intensity of Q19-Q31 in the *N. gaditana* transgenic lines carrying *QHS1-mVenus* fusion under the control of (*Hsp90*, Nga00934) promoter. Hsp90_1 is the positive control with mVenus reporter alone.

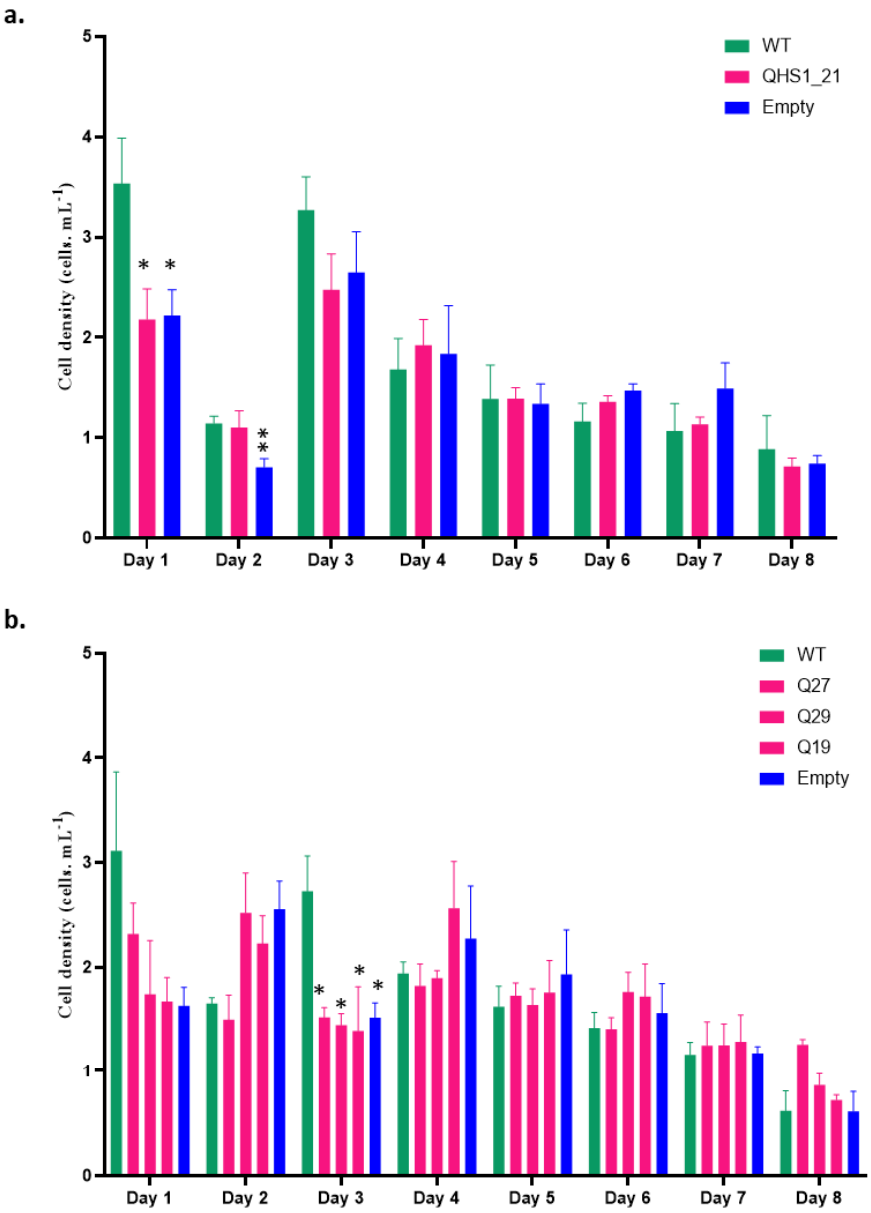


Figure S 22. Growth rate of transgenic *N. gaditana*.

a. The growth rate of wild type, transgenic *N. gaditana* cultures expressing *QHS1* (Q21) and empty vector measured from Day 1–8. **b.** The growth rate of wild type, transgenic *N. gaditana* cultures expressing *QHS1* (Q27, Q29, Q19) and empty vector measured from Day 1–8. Error bars represent standard deviation for three biological replicates (n=3). Significant differences between wild type and transgenic cell lines of *N. gaditana* calculated using one-way ANOVA- Dunnett's multiple comparisons test : * P<0.05, ** P<0.005.

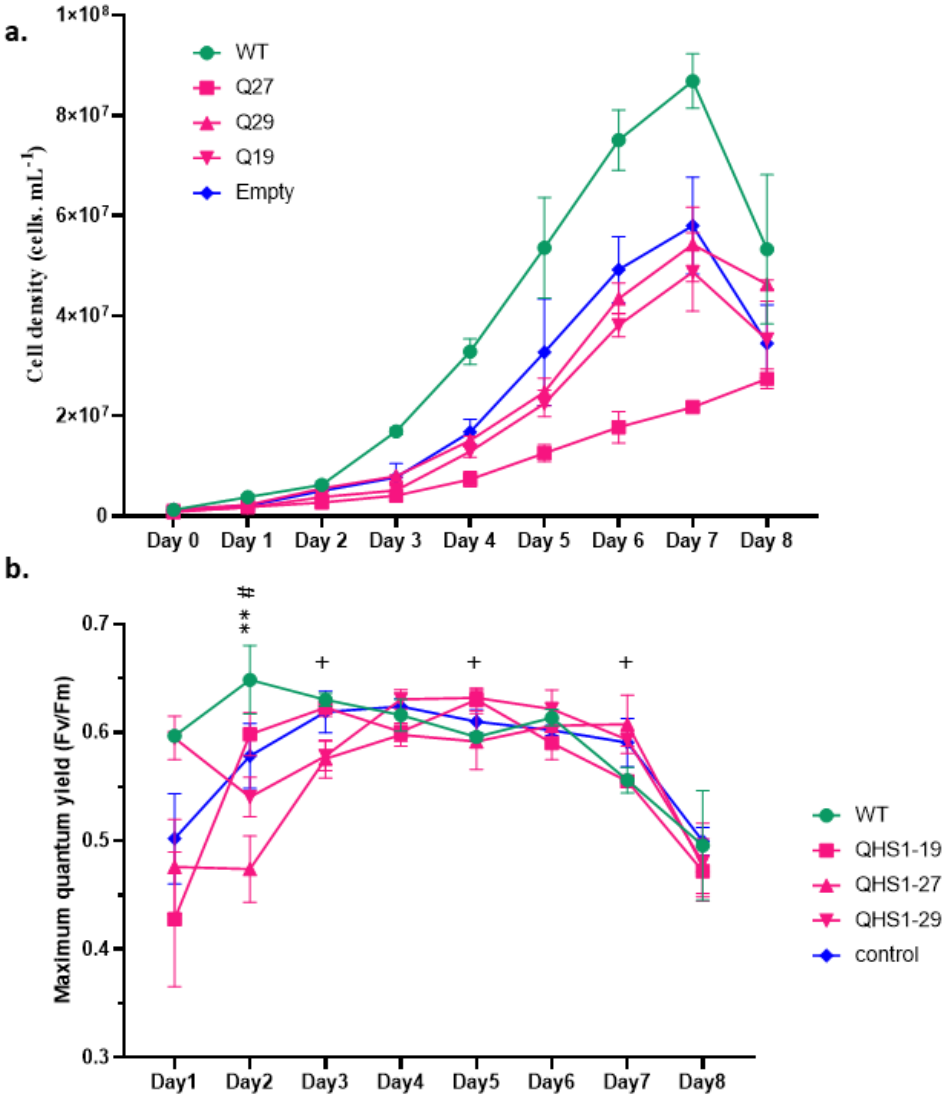


Figure S 23. The growth, photosynthetic health of transgenic *N. gaditana*.

a. Growth curve of wild type and transgenic *N. gaditana* cultures (Q19, Q27, Q29) expressing QHS1-mVenus fusion protein including positive control measured over a period of 8 days. Rest of the transgenic lines are included in supplementary data. Error bars represent the standard deviation for wild type and transgenic lines (n=3). **b.** Photosynthetic activity of *N. gaditana* wild type and transgenic lines under study. Error bars represent the standard deviation for wild type and transgenic lines (n=3). Significant differences between wild type and the transgenic lines were calculated using 2-way ANOVA Dunnett's multiple comparison test. Significant differences between wild type and (Q19- *, Q27- #, Q29- +). * P<0.05, ** P<0.005.

Reference:

1. Giorgio Valle group. Nannochloropsis gaditana Genomes portal - CRIBI, University of Padua. (2013). Available at: http://www.nannochloropsis.org/index.php.

2. Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H. & Winther, O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33,** 3387–3395 (2017).

3. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2,** 953–971 (2007).

4. Emanuelsson, O., Nielsen, H. & Heijne, G. Von. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* **8,** 978–984 (1999).

5. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305,** 567–580 (2001).

6. Bendtsen, J. D., Nielsen, H., Von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340,** 783–795 (2004).

7. Gruber, A., Rocap, G., Kroth, P. G., Armbrust, E. V. & Mock, T. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J.* **81,** 519–528 (2015).

8. Mandel, M. A., Feldmann, K. A., Herrera-estrella, L., Rocha-sosa, M. & Lebn, P. CLA 1, a novel gene required for chloroplast development, is highly conserved in evolution. *Plant J.* **9,** 649–658 (1996).

9. Phillips, M. A., León, P., Boronat, A. & Rodríguez-Concepción, M. The plastidial MEP pathway: unified nomenclature and resources. *Trends Plant Sci.* **13,** 619–623 (2008).

10. Budziszewski, G. J. *et al.* Arabidopsis genes essential for seedling viability: Isolation of insertional mutants and molecular cloning. *Genetics* **159,** 1765–1778 (2001).

11. Hsieh, M. H., Chang, C. Y., Hsu, S. J. & Chen, J. J. Chloroplast localization of methylerythritol 4-phosphate pathway enzymes and regulation of mitochondrial genes in ispD and ispE albino mutants in Arabidopsis. *Plant Mol. Biol.* **66,** 663–673 (2008).

12. Hsieh, M. H. & Goodman, H. M. Functional evidence for the involvement of Arabidopsis IspF homolog in the nonmevalonate pathway of plastid isoprenoid biosynthesis. *Planta* **223,** 779–784 (2006).

13. Gutiérrez-Nava, M. D. L. L., Gillmor, C. S., Jiménez, L. F., Guevara-García, A. & Léon, P. Chloroplast Biogenesis genes act cell and noncell autonomously in early chloroplast development. *Plant Physiol.* **135,** 471–482 (2004).

14. Hsieh, M. H. & Goodman, H. M. The Arabidopsis IspH homolog is involved in the plastid nonmevalonate pathway of isoprenoid biosynthesis. *Plant Physiol.* **138,** 641–653 (2005).

15. Okada, K., Saito, T., Nakagawa, T., Kawamukai, M. & Kamiya, Y. Five geranylgeranyl diphosphate synthases expressed in different organs are localized into three subcellular compartments in Arabidopsis. *Plant Physiol.* **122,** 1045–1056 (2000).

16. Beck, G. *et al.* Characterization of the GGPP synthase gene family in Arabidopsis thaliana. *Plant Mol. Biol.* **82,** 393–416 (2013).

17. Zhu, X. F. *et al.* Geranylgeranyl pyrophosphate synthase encoded by the newly isolated gene GGPS6 from Arabidopsis thaliana is localized in mitochondria. *Plant Mol. Biol.* **35,** 331–341 (1997).

18. Cunillera, N. *et al.* Arabidopsis thaliana Contains two differentially expressed farnesyl-diphosphate synthase genes. *J. Biol. Chem.* **271,** 7774–7780 (1996).

19. Cunillera, N., Boronat, A. & Ferrer, A. The Arabidopsis thaliana FPS1 gene generates a novel mRNA that encodes a mitochondrial farnesyl-diphosphate synthase isoform. *J. Biol. Chem.* **272,** 15381–15388 (1997).

20. Ducluzeau, A. L. *et al.* Gene network reconstruction identifies the authentic trans-prenyl diphosphate synthase that makes the solanesyl moiety of ubiquinone-9 in Arabidopsis. *Plant J.* **69,** 366–375 (2012).

21. Chen, Q., Fan, D. & Wang, G. Heteromeric geranyl(geranyl) diphosphate synthase is involved in monoterpene biosynthesis in arabidopsis flowers. *Mol. Plant* **8,** 1434–1437 (2015).

22. Bouvier, F., Suire, C., D'Harlingue, A., Backhaus, R. A. & Camara, B. Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. *Plant J.* **24,** 241–252 (2000).

23. Jun, L., Saiki, R., Tatsumi, K., Nakagawa, T. & Kawamukai, M. Identification and subcellular localization of two solanesyl diphosphate synthases from Arabidopsis thaliana. *Plant Cell Physiol.* **45,** 1882–1888 (2004).

24. Phillips, M. A., D'Auria, J. C., Gershenzon, J. & Pichersky, E. The Arabidopsis thaliana type I isopentenyl diphosphate isomerases are targeted to multiple subcellular compartments and have overlapping functions in isoprenoid biosynthesis. *Plant Cell* **20,** 677–696 (2008).

25. Campbell, M., Hahn, F. M., Poulter, C. D. & Leustek, T. Analysis of the isopentenyl disphosphate isomerase gene from Arabidopsis thaliana. *Plant Mol. Biol.* **36,** 323–328 (1998).

APPENDIX 4

Published work

Algal Research 44 (2019) 101708

Contents lists available at ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

Novel endogenous promoters for genetic engineering of the marine microalga *Nannochloropsis gaditana* CCMP526

Margaret Ramarajan[a], Michele Fabris[a,b,*], Raffaela M. Abbriano[a], Mathieu Pernice[a], Peter J. Ralph[a]

[a] University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo, NSW, 2007, Australia
[b] CSIRO Synthetic Biology Future Science Platform, GPO Box 2583, Brisbane, QLD, 4001, Australia

ARTICLE INFO

Keywords:
Promoter
Nuclear transformation
Genetic engineering
Nannochloropsis
Algae biotechnology

ABSTRACT

Nannochloropsis is a marine microalga from the *Eustigmatophyceae* stramenopile lineage that has been studied extensively due to a broad range of industrial applications, mostly related to their oil and pigment production. However, tools to genetically engineer members of this group, and therefore further understand and maximise their industrial potential are still limited. In order to expand the potential industrial uses of this organism, several molecular tools, including gene promoters of different strength, are needed. A comprehensive and diverse set of well-characterized promoters is key to a number of genetic engineering and synthetic biology applications, such as the assembly of complex biological functions or entire metabolic pathways.

In this study, we measured the promoter activity of three endogenous constitutive promoters from *N. gaditana* genes *EPPSII (Nga02101); HSP90 (Nga00934); ATPase (Nga06354.1)* in driving the expression of a *Sh ble-mVenus* fluorescent reporter fusion protein. Through a combined approach that includes flow cytometry, RT-qPCR and immunoblotting, we profiled the activity of these promoters at both the transcript and protein level. Two promoters *HSP90 (Nga00934)* and *EPPSII (Nga02101)* outperformed the widely used *β-tubulin* promoter, exhibiting 4.5 and 3.1-fold higher mVenus fluorescence, respectively. A third promoter *ATPase (Nga06354.1)* was also able to drive the expression of transgenes, albeit at lower levels. We show that the new promoters identified in this study are valuable tools, which can be used for genetic engineering and functional genetics studies in *N. gaditana.*

* Corresponding author at: University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo, NSW, 2007, Australia.
E-mail address: Michele.Fabris@uts.edu.au (M. Fabris).

https://doi.org/10.1016/j.algal.2019.101708
Received 12 June 2019; Received in revised form 24 October 2019; Accepted 24 October 2019
Available online 19 November 2019