CHARACTERIZATION OF *NANNOCHLOROPSIS OCEANICA* CCMP1779 GROWN IN LIGHT:DARK CYCLES INFORMS GENETIC ENGINEERING TOOL DEVELOPMENT

By

Eric Poliner

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Cell and Molecular Biology – Doctor of Philosophy

2017

ABSTRACT

CHARACTERIZATION OF *NANNOCHLOROPSIS OCEANICA* CCMP1779 GROWN IN
LIGHT:DARK CYCLES INFORMS GENETIC ENGINEERING TOOL DEVELOPMENT

By

Eric Poliner

Nannochloropsis is a genus of fast-growing microalgae that have a high lipid content.
Nannochloropsis species have a high triacylglycerol (TAG) content and contain a large amount
of the omega-3 long-chain polyunsaturated fatty acid, eicosapentaenoic acid (EPA). There is a
growing interest in *Nannochloropsis* species as models for the study of microalga lipid
metabolism and as a platform for synthetic biology. Genome sequences are available for several
species, and genetic engineering techniques are being introduced. In this study, I developed a
new generation of transgenic vectors for gene stacking and marker-free gene disruption in
Nannochloropsis oceanica CCMP1779. These tools enable gene specific studies and were
applied to investigate a lipid biosynthetic pathway that is co-expressed under different light
conditions.

As for all photosynthetic organisms, light plays an important role in driving metabolism
and regulation by photosensing in *Nannochloropsis* species. Each day photosynthetic organisms
must maximize their energy capture during the day and be able to sustain themselves during the
night. *Nannochloropsis* cultures synchronize cell division during a light:dark cycle, with cell
division occurring at night, along with the usage of stored metabolites that are accumulated
during the day. RNA-sequencing measures global transcript abundance, that ultimately might
lead to changes in enzymatic activity, metabolism and physiology. I investigated the role of
transcriptional regulation on metabolite levels and cell physiology using RNA-sequencing. In
the study I found coordination between cell growth, triacylglycerol and hexose content, and

transcript abundance of the genes in relevant pathways. Briefly anabolic processes were phased to the light period and catabolic processes phased to the dark period. Furthermore, promoters for transgenic expression were chosen based on transcriptomic measurements gathered in this study.

Eicosapentaenoic acid is a high-value fatty acid that is a necessary nutrient for humans, with a biosynthetic pathway consisting of 5 fatty acids desaturases (FADs) and a fatty acid elongase (FAE). Interestingly, the genes of this biosynthetic pathway were strongly co-expressed during light:dark cycles, and I set out to characterize the pathway. Expression of isolated cDNAs in *S. cerevisiae* resulted in the production of the expected long-chain polyunsaturated fatty acids (LC-PUFAs), and ultimately EPA when all 4 LC-PUFA FADs and an FAE were co-expressed. Selected FADs were overexpressed in *N. oceanica* and resulted in increased LC-PUFA and EPA content.

CRISPR/Cas9 is a potent tool for gene editing. The RNA-guided nuclease, Cas9, was tested as a fusion with green fluorescent protein (GFP) and NanoLuciferase (Nlux) reporters, and the Cas9-Nlux fusion was readily detectable for efficient screening of transformants for recombinant protein production. Single-guide RNAs (sgRNAs) when fused to 5' and 3' self-cleaving ribozymes efficiently targeted genes. The two components of the system were expressed from a bidirectional promoter. *N. oceanica* is capable of expressing transgenes from circular episomal DNA, and an episomal CRISPR construct was generated. The nitrate reductase gene was targeted and the mutants generated with frame-shifts in the coding sequence were unable to grow on nitrate. When antibiotic selection was removed, the episome was lost, and a mutant line that was "cured" of the episome was isolated. These tools are being utilized for gene specific studies in *N. oceanica*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

TAG - triacylglycerol

LC-PUFA - long-chain polyunsaturated fatty acids

EPA - eicosapentaenoic acid

RNA-seq - RNA-sequencing

CHIP-Seq - chromatin immunoprecipitation DNA sequencing

GC - guanine cytosine nucleotides

N - nitrogen

bZIP - basic leucine zipper

VCP - violaxanthin chlorophyll binding protein

Ribi - ribosomal subunit bidirectional promoter

UEP - ubiquitin extension protein

β-tub - β-tubulin

EF - elongation factor

lux - luciferase

bHLH - basic helix loop helix

WRI1 - WRINKLED1 transcription factor

UTR - untranslated region

DGDG - digalactosyldiacylglycerol

DGTS - diacylglyceryl-*NN*-trimethylhomoserine

DAG - diacylglycerol phosphate

MGDG - monogalactosyldiacylglycerol

SQDG - sulfoquinovsyldiacylglycerol

PC - phosphatidylcholine

PE - phosphatylethanolamine

ACP - acyl-carrier protein

ENR - enoyl-ACP reductase

KAS - ketoacyl-ACP synthase

KAR - ketoacyl-ACP reductase

TE - acyl-ACP thioesterase

LC-FACs - long-chain fatty acyl-CoA synthetase

PAP - phosphatidic acid phosphatase

LPAT - lysophosphatidyl acyltransferase

PDC - pyruvate dehydrogenase complex

PDCK - pyruvate dehydrogenase complex kinase

PDAT - phospholipidiacylglycerol acyltransferases

DGAT - diacylglycerol acyltransferase type I

DGTT - diacylglycerol acyltransferase type II

FAS - fatty acid synthase

DGDS - digalactosyl-1,2-sn-diacyl-glycerol synthase

ELO - elongase

ENO - enolase

GPAT - glycerol-3-phosphate acyltransferase

LPAAT - lysophosphatidic acid acyltransferase

GT - glycosyl transferases

GH - glycosyl hydrolases

M2HD - mannitol 2-dehydrogenase

MDH - malate dehydrogenase

ME - malic enzyme

MGDGS - monogalactosyl-1,2-sn-diacylglycerol synthase

MPDH - mannitol-1-phosphate dehydrogenase

MPP - manitol-1-phosphatase

PC - phosphoenolpyruvate carboxylase

PCK - pyruvate carboxykinase

PDH - pyruvate dehydrogenase

PFK - phosphofructokinase

PGI - phosphoglucoisomerase

PGK - phosphoglycerate kinase

PGLM - phosphoglycerase mutase

PGM - phosphoglucomutase

PK - pyruvate kinase

PP - phosphatidate phosphatase

PPDK - pyruvate phosphate dikinase

PYC - pyruvate carboxylase

SLS - sulfolipid synthase

TPI - triose-phosphase isomerase

TPT - triosephosphate transporter

UGPase - UDP-glucose pyrophosphylase

TFA - total fatty acids

NR - nitrate reductase

HR - homologous recombination

FP - Fluorescent Protein

Nlux - NanoLuciferase

NLS - nuclear localization signal

sgRNA - single-guide RNA

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

KO - knockout

HDV - hepatitis delta virus ribozyme

HH - hammerhead ribozyme

CEN/ARS – *Saccharomyces cerevisiae* centromere and autonomous replication sequence fusion

FP - fluorescent protein

CP - chromoprotein

BRET - bioluminescence resonance energy transfer

X-Gluc - 5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid

kb - kilobase

TF - transcription factor

TR - transcription regulator

TFBS - transcription factor binding site

PWM - position weighted matrix

TPR - third party-repository

PS - photosystem

LHC - light harvesting complex

ACC - acetyl-coA carboxylase

ACL - ATP citrate lyase

ACS - acetyl-coA synthetase

ADH - aldehyde dehydrogenase

ALD - aldolase

CA - carbonic anhydrase

CS - cellulose synthase

HS - heat shock protein

DES - desaturase

FbP - fructose 1,6-bisphosphase

FK - fructokinase -

GAPDH - glyceraldehyde 3-phosphase dehydrogenase

GBS - β-1,3-glucan synthase

16:0 - 16:0 fatty acid

2PGA - 2-phosphoglycerate

3PGA - 3-phosphoglycerate

6PG - 6-phosphogluconate

Ac-CoA - acetyl-CoA

Ace - acetate

Citr - citrate

DHAP - dihydroxyacetone phosphate

F1-6P - fructose 1,6-bisphosphate

F6P - fructose 6-phosphate

FA - fatty acid

Fru - fructose

G1-3-P - glycerate 1,3-bisphosphate

G1P - glucose 1-phosphate

G3P - glycerate 3-phosphate

G6P - glucose 6-phosphate

Glu - glucose

M1P - mannitol 1-phosphate

Mal - malate

Mnl - mannitol

OAA - oxaloacetate

PEP - phosphoenolpyruvate

Pyr - pyruvate

UDPG - UDP-glucose

Mito - mitochondria

FAD - fatty acid desaturase

FAE - fatty acid elongase

aa - amino acid

DOX - desaturase overexpression line

NR-KO - nitrate redctase knockout line

iEV - integrate empty-vector CRISPR line

WT - wild-type

construct - constructed DNA

PBS - phosphate buffered saline solution

Δ - delta

ω - omega

PCR - polymerase chain reaction

DAPI - 4'-6-Diamidino-2-Phenylindole

CLSM - confocal laser scanning microscopy

LDSP - lipid droplet surface protein

Chapter 1. Introduction

Algae are highly efficient at turning solar energy into biomass, and are sources of unique

bioproducts, such as omega-3 fatty acids (*1, 2*), carotenoids (*3*), and interesting polysaccharides,

such as agarose, alginate, and β-1,3-glucans (*3-5*). Several groups have screened algae for

productivity and production of valuable compounds (*4, 6, 7*). *Nannochloropsis* was identified as

a genus with rapid growth, and high lipid content, including triacylglycerol (TAG) (*6*) and the

omega-3 (*ω3*) long-chain polyunsaturated fatty acid, eicosapentaenoic acid (EPA) (*1*). Under

nutrient-replete conditions *Nannochloropsis* species have a lipid content of ~25-30% of dry-

weight (*6, 8-10*). Stresses, such as high-light or nutrient deprivation, in particular nitrogen (N)

deprivation, causes microalgae to pause growth and accumulate triacylglycerol or other storage

compounds, and drives *Nannochloropsis* species to accumulate high quantities of TAG, up to

60% of biomass (*6, 8-11*). In recent years genomes for several *Nannochloropsis* species have

become available (Table 1.1) (*11-13*) and molecular tools have been developed (Table 1.2) (*14-

17*), making this genus excellent microalgal models for comparative genomics (*18, 19*).

Synthetic biology is an emerging field based on rationally designing biological systems

(*20*). To develop systems that behave as desired, an approach described as design, build, test, is

used to iteratively test refinements and determine how elements of the system influence the

outcome (*21*). The design phase is often based on information drawn from genome-wide data

and databases of related systems. Organisms with high-quality genome-wide data and advanced

genetic engineering tools that can be redesigned are known as chassis organisms (Figure 1.1).

Metabolic maps are built using genome assemblies, functional annotation, and databases of

known enzymatic pathways. Regulatory networks are coming into focus through integrating

RNA-seq, chromatin immunoprecipitation DNA sequencing (CHIP-Seq), and databases of

transcription factors and their target DNA motifs. In order to build or refine biosynthetic pathways or develop chassis organisms, several molecular tools are needed to modify the genome of an organism. While molecular tools, such as mutant libraries, transgenic overexpression, and reporter protein fusions, are instrumental in gaining a molecular understanding of biological processes, they are by themselves insufficient to create optimized biological systems. Redesigned algae will require a new generation of tools that enable precise and marker-free gene disruption mutants, and high-capacity gene stacking systems that can robustly and predictably express multiple genes. Finally, in order to test the synthesized system, highly facile methods to select or screen for the desired modifications as quickly as possible post transformation are needed.

The *Nannochloropsis* genus is also an emerging algal model for genetic engineering of lipid accumulation (*17, 19, 22*). Several *Nannochloropsis* species seem particularly amenable to transgenic expression, with a moderate GC (guanine cytosine) content and simple gene structure facilitating genetic manipulation (*11, 18, 23*). Several endogenous promoters and terminators are in use, including bidirectional promoters that are helpful in stacking transgenes, i.e. expressing multiple genes (*14, 15, 23-25*). Methods exist for targeted DNA insertion by homologous recombination into the genome (*14, 25, 26*). We and others developed CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) based methods, making targeted gene disruption and editing possible (*17, 27*). Genetic engineering toolkits are becoming publicly available and should accelerate development of *Nannochloropsis* species as chassis organisms (Figure 1.1).

In order to develop the knowledge and tools to make *Nannochloropsis oceanica* CCMP1779 into a chassis organism I investigated the coordination between cell physiology,

metabolites, and the transcriptome during growth under light:dark cycles in Chapter 2. Based on this study I identified the EPA biosynthetic pathway as being co-expressed during light:dark cycles. In Chapter 3, I describe the isolation and characterization of the constituent genes of this pathway, the development of molecular tools for gene overexpression and gene stacking, and subsequent manipulation of the EPA pathway in *N. oceanica*. My development of a gene disruption technique based on an episomal CRISPR/Cas9 system, that can make non-transgenic gene disruption (knockout) mutants, is detailed in Chapter 4. The transcriptomic data gathered and genetic tools developed make *N. oceanica* gene-specific studies feasible and form the basis to develop *N. oceanica* as a chassis strain.

Taxomony and Evolution of the *Nannochloropsis* genus

Nannochloropsis is a genus in the heterokont phylum. Photosynthetic heterokonts are secondary endosymbionts originating from an unicellular heterotrophic eukaryotic cell that engulfed a red alga, which, over time, became a plastid. The evolutionary relationships of organisms in this clade are complex and still not well defined. The red-algal type plastid shows several signatures of its origin distinct from green lineage plastids (*28-30*). Brown algae and diatoms are also algae of the heterokont lineage that share 57% and 51% of genes with *N. gaditana* CCMP526 genome, respectively (*13*).

The *Nannochloropsis* genus seems to possess a number of genes derived from endosymbiotic gene transfer, particularly in lipid biosynthetic and carbohydrate degradation pathways, e.g. glycosyl hydrolases (*18*). The diverse genetic background of the *Nannochloropsis* genus may have contributed to its oleaginousness, with a particularly large set of lipid biosynthetic genes. For example, *Nannochloropsis oceanica* possess 11 type-2 diacylglycerol

3

acyltransferases (DGATs) referred to also as DGTTs (*11, 18, 31, 32*), and a high copy number of

other predicted lipid biosynthetic genes such as: enoyl-ACP reductase (ENR), ketoacyl-ACP

synthase (KAS), ketoacyl-ACP reductase (KAR), acyl-ACP thioesterase (TE), long-chain fatty

acyl-CoA synthetase (LC-FACS), phosphatidic acid phosphatase (PAP), and lysophosphatidyl

acyltransferase (LPAT) (*18*). The evolutionary origins of these genes are still being investigated

and it is hypothesized that the ancestral heterotroph, red alga, and horizontal gene transfer

contributed to the present genome (*18*).

      The secondary endosymbiosis event also led to interesting cellular structure

characteristics, such as four membranes surrounding the plastid, complicating intracellular

trafficking (*30, 33, 34*). Trafficking of nuclear encoded proteins into the plastid has been

extensively studied in diatoms (*35*), and has led to the protein location prediction program

HECTAR (*36*). While in several *Nannochloropsis* species there are some examples of protein

localization by fluorescent protein (FP) fusions, including to the plastid (*24*), endoplasmic

reticulum (ER) (*15*), mitochondria (*37*) and lipid droplets (*25, 31*), there have not been

investigations into the signals and mechanisms of protein localization. The transport of

metabolites across subcellular compartments also has not been studied in the *Nannochloropsis*

genus, although it represents a plausible target for optimizing metabolite production (*38*).

      Genetic diversity has also arisen frequently by horizontal gene transfer during algae

evolution (*39, 40*), and enables adaption to unique environments or metabolic niches (*41*).

Several lipid biosynthetic genes appear to be related to bacterial orthologs and likely were

acquired by horizontal gene transfer, including KAR, PAP, ENR, KAS, TE, and LC-FACS genes

(*18*). An operon encoding proteins specialized in hydrogen generation possibly derived from

bacteria has also been identified in *N. oceanica* (*11*). Transkingdom gene transfer by bacterial

conjugation to diatoms and the capacity for heterokonts to maintain episomal DNA indicates a possible route of gene acquisition (*40, 42*). This evolutionary diversity of the *Nannochloropsis* genus and the genetic plasticity of the heterokonts may lead to an interesting model for symbiotic evolution and a chassis organism that can be robustly adapted for genetic engineering.

Genomes and transcriptomes across the *Nannochloropsis* genus

A complete genome sequence forms the foundation for gene-specific studies, and is a prerequisite for the drafting of metabolic and regulatory maps. Several genome assemblies have been generated for different species of *Nannochloropsis,* including multiple strains within a species, such as, *N. oceanica* CCMP1779 (*11*) and IMET1 (*18*), *N. gaditana* CCMP526 (*13, 27*) and B-31 (*12*), *N. salina* CCMP537, *N. oculata* CCMP525 (*18*), and *N. granulata* CCMP529 (Table 1.1) (*18*). The genomes of the examined *Nannochloropsis* species are approximately 30 megabases, and contain 7-11,000 genes each (*11, 13, 18*). The *N. gaditana* B-31 genome is estimated to be distributed over 30 chromosomes (*12*), and the presumed *N. oceanica* IMET1 chromosomes separated by pulse-field gel electrophoresis as 22 individual genome fragments (*18*).

The plastid and mitochondrial genomes of representatives from five *Nannochloropsis* species were used to produce a pangenome of each organelle (*28*). The *N. oceanica* IMET1 plastid genome is 117,548 bp and contains 160 genes consisting of 126 protein-coding genes, and 34 RNA genes, and the mitochondrial genome is 38,057 bp and contains 63 genes consisting of 35 protein-coding genes and 28 RNA genes (*28*). The *Nannochloropsis* genus plastid pangenome contains signatures of a red algal origin, including red algal-type Rubisco and Rubisco activase genes (*43*).

Extensive transcriptome data based on RNA-sequencing of cultures grown under different conditions reveal characteristic transcriptional changes, providing a whole-genome view of possible adjustments to maintain homeostasis. Different examined conditions included phosphorus (*44*) and N deprivation (*11, 12, 22*), alternating light:dark cycles (*45*)(Chapter 2), varying light intensities (*46*), and different growth phases of batch cultures (*13*), in various *Nannochloropsis* species (Table 1.1). These datasets suggest that different aspects of metabolism and other cellular processes, such as the cell cycle are coordinated on a transcriptional level in response to environmental conditions. For example, N deprivation, which generally leads to a transition from the normal cell division cycle to quiescence, also causes transcriptional downregulation of photosynthesis and protein production, while lipid biosynthesis is upregulated as observed in *Chlamydomonas* (*47, 48*), and *Nannochloropsis* species (*12, 13, 22*) . In Chapter 2, I show that when *N. oceanica* is grown in a light:dark cycle, there is phased expression of genes at different times of day, including those involved in cell division at night, and anabolic processes during the day (*45*). A majority (64%) of the DNA-binding transcription factors and 56% of other transcriptional regulators, have phased expression during light:dark cycles. This genome-wide dataset was an asset for further studies such as the identification of condition-specific promoters. Links to currently published genome-wide datasets are listed in Table 1.1.

Light regulation and photosynthesis

In heterokont algae, light, either by capture and conversion of solar energy during photosynthesis or by perception through photosensory regulatory proteins, affects metabolite levels (pigments, lipids, and carbohydrates) (*45, 49*), coordinates the cell cycle (*45, 49-52*), and may entrain a circadian clock (*53*). In *Nannochloropsis* species, high-intensity light results in accumulation of

TAG and a decrease in plastid size, thus maximizing energy conversion while avoiding photodamage (*10, 46, 54*). The day:night cycle influences most organisms to coordinate behavior and/or metabolism with either phase (*45, 49, 52*). The transitory storage compounds used by *N. oceanica* during a light:dark cycle are unknown, therefore I measured TAG and carbohydrates (in the form of hexoses) content throughout a light:dark cycle (Chapter 2). *Nannochloropsis* species accumulate TAG (*55*) and carbohydrates during the day, which are both metabolized during the night, in accordance with transcriptional changes in genes encoding enzymes of the respective biosynthesis and utilization pathways (*45*).

Nannochloropsis species are studied as a model for photosynthesis in secondary endosymbionts. The *Nannochloropsis* genus is notable for only possessing chlorophyll *a*, the unusual carotenoids violaxanthin and vaucheriaxanthin ester, and an alternative xanthophyll cycle utilizing violaxanthin, antheraxanthin, and zeaxanthin (*56-58*). Red algae and their derived endosymbionts contain LHCr type antenna proteins that link core complex pigment and protein components, and participate in energy transfer and photoprotection (*56, 58, 59*). Characterization of the photosystem II (PSII) of *N. gaditana* identified the light harvesting complex proteins of the classes LHCx, LHCf, Red-CLH-like LHC, and LHCr, that are characteristic of the red algae-type plastid (*59*). Characterization of the *N. gaditana* photosystem I (PSI) discovered the absence of several subunits (PsaH, PsaK, PsaG) that are typically present in plants, and identified the light harvesting complex proteins of the classes LHCr, LHCf, and LHCx associated with PSI (*58*).

Lipid and carbon metabolism

Lipid biosynthesis is the best characterized metabolic pathway in *Nannochloropsis* species, in particular the production of TAG and EPA. The *Nannochloropsis* genus is hypothesized to possess a cytosolic type-I fatty acid synthase (FAS) in addition to the plastid type-II FAS complex, but further studies to corroborate this hypothesis are needed (*11, 45, 46*). The TAG biosynthetic pathway involves the transfer of acyl chains to a glycerol backbone by the sequential action of glycerolphosphate acyltransferase (GPAT), LPAT, and DGAT, in addition phospholipidiacylglycerol acyltransferases (PDATs) have also been identified in *Nannochloropsis* species (*11*). The four LPATs of *N. oceanica* have been investigated for their roles in membrane lipids and TAG biosynthesis with LPAT1 and LPAT4 having primary roles in each process respectively, and LPAT2 and LPAT3 possibly playing roles in both processes (*25*). Of the 13 DGATs encoding genes, 6 are upregulated during N deprivation, a condition that also favors TAG accumulation (*31*). We identified a *N. oceanica* DGAT (DGTT5) that is able to increase TAG production in many different hosts (*31*).

The *Nannochloropsis* genus contains the omega-3 fatty acid EPA in its membrane lipids (15-30% total fatty acids - TFA) (*1, 11, 60, 61*). I reconstructed the *N. oceanica* EPA biosynthetic pathway in *S. cerevisiae* by introducing four LC-PUFA fatty acid desaturases (FADs) and a fatty acid elongase (FAE), resulting in the production of EPA (0.1% TFA) (*15*) (Chapter 3). FADs are named for the double bond introduced, a specific number of carbons from either the carboxyl (Δ, delta-) or methyl (*ω,* omega-) end of a fatty acid chain. Thus, omega-3 and delta-6 FADS act on the third carbon from the methyl end and the sixth carbon from the carboxyl end, respectively. The FADs of *N. oceanica* resemble those of other heterokont algae with their histidine box motifs for coordinating a diiron center, and in two cases (delta-5 and

delta-6) contain a cytochrome b domain (*15*). Eleven fatty acid elongases have been identified in

N. oceanica (*11*) and delta-6 and palmitic fatty acid specific elongases from *N. oceanica* and *N.*

gaditana, respectvely, have been characterized in some detail. The palmitic acid elongase

controls flux into the EPA pathway by conversion of 16:0 to 18:0 (*26*), while the delta-6

elongase converts 18:3 to 20:3, two intermediates with low *in vivo* abundance (*15*). EPA is likely

produced in the ER but ultimately accumulates on diacylglyceryl-trimethylhomoserine (DGTS)

and monogactosyl diacylglycerol (MGDG), the latter of which is assembled in the plastid (Figure

1.1). Thus it has been proposed that EPA is imported into the plastid by a DGTS mediated

transport mechanism (*61*).

Carbohydrates play structural, storage, and osmoprotectant roles in *Nannochloropsis*

species. In *N. oceanica* glucose is the predominant hexose in the total complex carbohydrate

fraction, which contains smaller amounts of mannose, and trace amounts of rhamnose, fucose,

arabinose, xylose, and galactose (*11*). Marine *Nannochloropsis* species reduce their level of the

sugar alcohol mannitol and the disaccharide trehalose content in response to low-salt stress

consistent with a role of these carbohydrates in osmoprotection (*62*). Heterokont algae lack

starch but produce β-1,3 linked polysaccharides (chrysolamarinin), by the activity of a β-1,3-

glucan synthase which ispredicted to be encoded in the genome of *Nannochloropsis* species (*11,*

12). Approximately 20% of alcohol insoluble polysaccharides are in this form in *N. oceanica*

(*11*). Chrysolamarinin is also a storage compound in diatoms (*49, 63*), and has been suggested to

have a similar role in *Nannochloropsis* species (*10, 22, 64*), but further studies are needed to

confirm this hypothesis. Cellulose is a major polysaccharide in *Nannochloropsis* species, with

approximately 80% of the of alcohol insoluble polysaccharides in this form in *N. oceanica* (*11*).

Cellulose serves as a major component of the cell wall (*11, 65, 66*), but the cell wall is quite

complex in *Nannochloropsis* species (*65*). Four putative cellulose synthase-encoding genes have been identified (*65, 66*). A large number of carbohydrate-degrading enzymes, 48-49 glycosyl hydrolases, with very diverse taxonomic relations, are found encoded across the pangenome (*18*). The complete repertoire of carbohydrate metabolism in the *Nannochloropsis* genus has yet to be fully established.

In order to understand the metabolic networks of *Nannochloropsis* species, a summary of possible chemical reactions in the form of a metabolic map has been generated. A mass-balanced metabolic map for *N. salina* CCMP537 has been produced by Loira and colleagues taking into account 9 organelles (as well as the plastid lumen) of the cell (*38*). The model was validated by modeling different growth conditions and comparing it to *in vivo* data. The conditions of N and phosphate deprivation, were used to maximize lipid production and determine essential nutrients, respectively. The iNS934 map-based model predicted several genes whose disruption may result in increased TAG.

Transcriptional Regulation

The dynamic control of metabolism in response to environmental changes and intracellular cues is multilayered but inevitably involves transcription factors (TF, which possess DNA binding activity) and transcriptional regulators (TR, regulators of TF activity) that modulate gene expression. Databases of known TFs and corresponding position-weighted matrices (PWM) (*67*) can be used to systematically determine potential interacting TF-DNA sequences. Having a large number of closely related organisms is a valuable resource for *in silico* predictions of conserved regulatory DNA motifs (*19*). Genome sequencing and cataloging of TFs as a first step has been undertaken for the *Nannochloropsis* genus (*11, 19*). Comparative genomic studies of the

heterokont lineage have identified TF signatures based upon organismal lifestyle (autotrophic, parasitic), multicellularity, or lifecycle stages (*68-70*). TF prediction of *N. oceanica* implied the presence of 115 putative TFs and 109 putative TRs, which combined represent about 2% of the genome (*11*). The *Nannochloropsis* genus has a reduced number of TF and TR families (20-26) compared to land plants and green algae, possibly due to a simple lifecycle and its unicellularity (*11, 19*). Putative TFs of the Myb family (29-35 members), a TF family known to regulate growth and metabolism in other organisms, are enriched in the *Nannochloropsis* genus (*11, 19*). Several TFs in *Nannochloropsis* species have been investigated but studies into their targets are only beginning. Despite the large amounts of predicted TFs and TRs their roles need to be experimentally corroborated.

Several approaches have been used to identify regulators of lipid biosynthesis in *Nannochloropsis* species. Hu and colleagues (*19*) took advantage of the extensive genome sequences available to predict conserved transcription factor binding sites (TFBS). They determined the enrichment of gene ontology (GO) terms associated with each motif, and enrichment of motifs associated with lipid biosynthetic genes. Using a TF catalogue and RNA-seq during N deprivation Hu et al. identified TFs that showed positive or negative co-expression with lipid biosynthetic genes. Finally, they predicted putative connections between TFs and motifs in lipid biosynthetic gene promoters based on a TF-PWM database (*71*). One of these predicted lipid biosynthesis regulating TFs (bZIP1) was recently investigated (*72*). In addition, Ajjawi and colleagues (*17*) identified 20 TFs possibly involved in TAG accumulation based on changes in expression under N deprivation, and used CRISPR/Cas9 disruption to assess the predictions.

As a photosynthetic organism, light sensing is likely important for tuning metabolism in *Nannochloropsis* species. Aureochromes are heterokont specific photosensitive transcription activators, with a bZIP DNA binding domain and a photosensing dimerization LOV domain (*11*). In diatoms the aureochromes are implicated in regulating several processes, including cell cycle and light acclimation (*50, 73, 74*). Cryptochromes are blue light photoreceptors derived from DNA repair enzymes, which have been characterized in diatoms (*75*) and found to oscillate during light:dark cycles in diatoms (*52*) and *N. oceanica* (*45*). Recently, the *Chlamydomonas reinhardtii* animal-like cryptochrome was reported to be involved in regulating the cell cycle and the circadian clock (*76, 77*). Finally, although phytochromes were thought to be absent in most heterokonts, they have recently been identified in diatoms (*78*). *N. oceanica* possesses three aureochromes, an animal-like cryptochrome, and lack phytochromes, but these proteins have not been functionally characterized (*11*).

Transformation and gene expression platforms

The most widely adapted method for transformation of *Nannochloropsis* species is by electroporation (*11, 13, 14*), but other protocols based on biolistics (*79, 80*) or agrobacterium have been developed (*81, 82*). For insertion of a transgene into the genome by electroporation a linear piece of DNA is required; a constructed DNA (construct) is therefore digested with restriction enzymes or PCR-amplified (*11, 14, 15, 83*). Each transformant is likely to have a distinct insertion site, and therefore may display different phenotypes due to genome context-specific regulation of the transgene's expression or disruption of endogenous genes by transgene insertion (*82*).

Introduction of circular DNA has so far had mixed success in producing transformants (*11, 14, 83*). Synthetic Genomics Inc. has described plasmid/episome maintenance in algae by use of autonomous replication sequences from a *Nannochloropsis* (*84*), and the utilization of an *S. cerevisiae* centromere – autonomous replication sequence (CEN/ARS) in pennate and centric diatoms (*40, 42, 85*). I developed an episomal expression system for *N. oceanica* based on the CEN/ARS region (Chapter 4). In diatoms, episomes are maintained under antibiotic selection but are gradually lost without selection pressure (*40, 42, 85*). The transgene expression levels from episomes are more uniform compared to genome integrated constructs between independent transformants, likely due to the absence of insertion site-specific effects using this approach (*42, 85*). The capacity of episomes for maintenance of foreign DNA have been reported to be up to 94 kilobases (kb) in diatoms (*42*).

Antibiotic resistance marker genes

Several antibiotics are effective depending on the *Nannochloropsis* strain, including hygromycin, zeocin, and blastidicin, and genes conferring resistance are used to isolate *Nannochloropsis* transformants (*11, 13, 14, 17, 25*). Zeocin in combination with its respective marker gene is the most widely used selection in *Nannochloropsis* species due to its stringency at low concentrations (Table 1.2). However, it is mutagenic and can lead to secondary mutations (*86*). In diatoms, the selection agents nourseothricin and G418 are frequently utilized with their respective antibiotic resistance genes (*42, 63, 87*). Mutated endogenous proteins in conjunction with competitive inhibitors, such as phytoene desaturase and the inhibitor norflurazon can also be used as a selection marker in some algae (*88*). I have made a number of vectors containing *N. oceanica*-adapted antibiotic selection marker genesthat will be available on Addgene

(www.addgene.com)(Table 1.3). When several selection agents and resistance genes are available,multiple transgenic tools can be used in conjunction in one transgenicline (Figure 1.2a). However, techniques for generation of transgenic algae without antibiotic resistance markers are necessary for deployment into open ponds. The removal of an antibiotic resistance marker gene could be achieved by the use of a recombinase or endogenous homologous recombination (*89*). The cotransformation of an episome carrying an antibiotic resistance gene with an insertion construct without a selection marker may also enable generation of marker-free mutants after episome loss in the absence of selection pressure.

Transgenic expression in *Nannochloropsis* species

A fundamental technique for genetic engineering and synthetic biology is the overexpression of target genes by increasing transcriptional and/or translational efficiency. Most often, strong promoters that mediate high transcription rates are utilized to express heterologous or endogenous genes at elevated levels. Several endogenous promoters from a variety of *Nannochloropsis* species have been isolated and applied to transgenic expression (Table 1.2), including those driving the genes encoding ubiquitin extension protein (UEP)(*13, 26, 79, 90*), β-tubulin (β-tub)(*13, 27, 37, 79, 83*), lipid droplet surface protein (LDSP)(*11, 15, 25, 91*), and elongation factor (EF)(*15, 31*). Several bidirectional promoters have also been utilized for transgenic expression, including those driving the expression of the genes encoding violaxanthin chlorophyll binding proteins (VCP)(*14, 24, 37*) or ribosomal subunits (Ribi) (Figure 1.2b) (*15*). To enhance translational efficiency of transgenes, a 5' UTR can include a consensus Kozak sequence (*92*) or leader enhancing sequence (*93-95*). Chapter 3 details the identification of the EF and Ribi promoters (*15*).

Reporter genes are useful for evaluating transgenic strategies and understanding gene/promoter function. In several *Nannochloropsis* species, members of the classes of the fluorescent- (FP), luminescent- (lux), and chromoproteins (CP) have been used (Table 1.2). Green fluorescent protein (GFP) is the most widely reported fluorescent protein and has been utilized for subcellular localization of fusion proteins throughout *Nannochloropsis* cells (*24, 25, 37*). Other fluorescent proteins such as a red fluorescent protein (RFP, sfCherry) (*79*), yellow FP (YFP, Venus variant) (*25, 31, 96*), and cyan FP (CFP, Cerulean variant)(*15*) have been employed in different *Nannochloropsis* species (Chapter 3 and Chapter 4). Luciferases have the advantage of a high signal to noise ratio, and specific substrates allow their use in combination. I developed codon-optimized firefly luciferase (Flux) and the ultra-bright NanoLuciferase (Nlux) for *in-vivo* assays in *N. oceanica* (Chapter 3 and Chapter 4)(*15*). The ultra-bright Nlux allows detection of very low protein quantities and is an effective photon donor for bioluminescence resonance energy transfer (BRET) (*97, 98*). Chromoproteins are colored and do not need a substrate, while the β-Glucuronidase (GUS) reporter is an enzyme that produces a blue stain after conversion of 5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid (X-Gluc). A purple chromoprotein (shPCP) from the sea anemone *Stichodacyla haddoni* was successfully produced in *N. oculata* and used for screening of transformants (*99*), while GUS has been utilized in *N. salina* (*83*).

Virus-derived 2A peptides allow production of two discrete proteins from a single transcript by preventing the formation of a peptide bond (peptide bond "skipping") (Figure 1.2c). The nascent protein interacts with the ribosome causing it to stall, then during the release of the ribosome a peptide bond is "skipped" between the final two amino acids (*100*). 2A peptides have been widely applied to different hosts and the efficiency of skipping depends on the peptide variant and host. After screening several variants and lengths in *N. oceanica*, I found the P2A

peptide of 60 amino acids to be most efficient (Chapter 3)(*15*). The F2A peptide has also been used in *N. salina* (*7*) and *Chlamydomonas* (*101*). In order to facilitate multigene expression, bidirectional promoters with P2A peptide sequences were assembled in the pNOC-stacked vector series and will be made available through Addgene (Chapter 3)(*15*). Toolkits for assembly of multiple transgenic expression cassettes into a single vector have been developed for some algae (*102, 103*), and would facilitate gene stacking in *Nannochloropsis* species (Figure 1.2d).

Generation of targeted gene disruption and transcriptional repression

RNA interference (RNAi) is a powerful technique that can suppress gene expression to varying degrees. RNAi by antisense or double-stranded RNA has been developed in a number of algae including *Chlamydomonas* (*104, 105*), diatoms (*63, 73, 106*), and different *Nannochloropsis* species (*16, 17, 37, 90*). RNAi using an inverted repeat of regions of a target gene has been found to be effective throughout the *Nannochloropsis* genus (Figure 1.3a) (Table 1.2) (*16, 17, 37, 90*). In order to generate a strong and stable repression effect several strategies are available including, fusing the interfering RNA to an antibiotic resistance gene (*16, 17, 106*), co-silencing a gene that can be counter selected (*105*), or expression of the interfering RNA from a bidirectional promoter that also drives expression of an antibiotic resistance gene. In case of essential genes, disruption of the target gene can result in slow growth or lethality while transcriptional repression targeting the same gene may result in a moderate phenotype (*17*).

Homologous recombination has been demonstrated in several *Nannochloropsis* species for insertion of an antibiotic resistance marker with flanking regions of 1 kb identical to the insertion site (Figure 1.3b). Targeting efficiency has been reported to be high when transforming low-density cultures (*14*). Several groups have used this technique for gene disruption (Table

16

1.2)(*14, 25, 26, 96*). Homologous recombination is adaptable to several purposes, such as, insertion of protein tags, insertion into neutral sites, or replacement of genes with altered functionality.

CRISPR/Cas9 is an RNA-guided nuclease based approach that is dramatically expanding the capabilities of biologists to modify genomes, particularly through the ability to disrupt specific genes or perform precise gene editing (*107, 108*). In this system two components have to be targeted to the nucleus , a single-guide RNA (sgRNA) and the Cas9 nuclease, which together form a ribonucleoprotein complex (Figure 1.3c). The sgRNA requires production without extraneous sequences or modifications on the termini; strategies for sgRNA production include the use of RNA polymerase III-driven promoters (most often the U6 promoter) (*109*), direct introduction into the cell (*17*), expression of modified tRNAs containing the sgRNA in a spliced region (*110*), co-expression of a ribonuclease and sgRNA with cleavage sites, and use of self-cleaving ribozymes (*110*). The U6 promoter from diatoms appear to be active and suitable for sgRNA production (*109*). However, there have been no reports of successful U6 promoter use in the *Nannochloropsis* genus, and two publications utilized alternative strategies (Table 1.2). Wang and colleagues (*27*) expressed the sgRNA from an V-ATPase promoter and had a low mutational efficiency. In a strategy developed by Ajjawi and coworkers (*17*), the sgRNA is synthesized and introduced by transformation into a *Nannochloropsis* strain expressing Cas9. In Chapter 4, I detail the development of a CRISPR/Cas9 system for *N. oceanica* based on a Cas9-Nlux fusion protein and a self-cleaving ribozyme flanked sgRNA, co-expressed by a bidirectional promoter. The generation of off-target mutations is an unresolved issue for CRISPR based gene editing, which has not been examined in the *Nannochloropsis* genus. Several strategies exist to reduce the number of potential off-targets, including transient expression of

one or both components (*17, 111, 112*), paired nickase Cas9 (*108*), or high-fidelity Cas9

enzymes (*113, 114*). The CRISPR/Cas9 system developed in Chapter 4 is based on an episome,

which can be removed after a mutation is produced. This method thus generates a marker-free

mutant without the continued presence of Cas9 and reduces the chances of potential off-target

mutations.

Insertional mutatagenesis, whereby an antibiotic resistance gene is randomly integrated

into a genome, results in gene disruptions and gene deletions (Figure 1.3d) (*47, 115*). Insertional

mutagenesis conducted in *N. gaditana* produced mutants with a variety of growth and

photosynthetic phenotypes, screening of which identified lines with enhanced light-use

efficiency(*116*). While insertional mutagenesis is an efficient method for forward-genetic

screens, ready-to-use lines with a disruption of a desired gene requires a mutant library that takes

a significant investment to establish (*115*).

Altering metabolism in *Nannochloropsis* species by protein engineering

The usefulness of the aforementioned genetic engineering tools has been demonstrated by

modifying different aspects of metabolism. The majority of studies have targeted lipid

biosynthesis, either to enhance TAG or EPA production (Table 1.2). In several cases endogenous

or heterologous (from *S. cerevisiae* or *C. reinhardtii*) DGATs have been overproduced in

different *Nannochloropsis* species (*31, 32, 81, 90, 95, 117*). Overexpression of the endogenous

DGAT1a-encoding gene in *N. oceanica* resulted in a 39% increase in TAG content per cell and

RNAi repression resulted in a 20% decrease in TAG content per cell following N deprivation

(*90*). We found that the overexpression of the endogenous DGGT5-encoding cDNA in *N.*

oceanica resulted in a 3.5 fold increase in TAG (as %TFA)(*31*). Furthermore, the DGTT7-

encoding cDNA has also been overexpressed in *N. oceanica* IMET1 resulting in 69% and 129%
increase in neutral lipid (% dry weight) content under N-replete and N-deprivation conditions
(*95*). The malonyl-CoA transacylase of *N. oceanica* IMET1, which loads the malonyl group onto
the acyl-carrier protein for fatty acid synthesis, has been characterized (*118*) and its
overproduction results in a 36% (% dry weight) increase in lipids without compromised growth
(*119*). In *N. salina* RNAi conducted against the pyruvate dehydrogenase complex (PDC) kinase
(PDCK), in order to increase acetyl-CoA levels for fatty acid production *(21)*, resulted in
enhanced TAG content at the expense of protein (*37*). I increased expression of the desaturase
genes in the EPA pathway using gene stacking techniques achieving up to a 25% increase in
EPA (Chapter 3)(*15*). Engineering efforts are only beginning and the enhanced productivity of
engineered *Nannochloropsis* strains will require further identification of targets to affect
metabolite partitioning into the desired pathways, and/or the reduction of final product turnover.

Altering metabolism in *Nannochloropsis* species by regulatory engineering

Manipulation of entire metabolic pathways on a greater scale could be accomplished by TF
engineering. Recently, TF overproduction, or inactivation (CRISPR) and repression (RNAi) of a
TF encoding gene have been used in various *Nannochloropsis* species with the goal of increasing
biomass and/or lipid production (Table 1.2)(*17, 72, 80, 120*). Overproduction of the bHLH2 TF,
in *N. salina* resulted in an increased growth rate and a greater biomass productivity, although the
transcriptional reprogramming was not described (*80*). Overproduction of the bZIP1 TF, a
predicted lipid biosynthetic TF (*19*), in *N. salina* resulted in enhanced growth, and under stress
conditions enhanced lipid content (*72*). In the bZIP1 overexpressor lines, the expression of
putative target genes involved in lipid biosynthesis (*19*) was increased under normal conditions

and more dramatically under stress conditions (*72*). Remarkably, the introduction of the *Arabidopsis thaliana* WRINKLED1 TF (WRI1), a regulator of seed oil production, into *N. salina* enhanced lipid accumulation, possibly by upregulating lipid biosynthetic genes containing the WRI1 motifs in the promoter (*120*). CRISPR inactivation of the ZnCys TF, likely involved in N assimilation, resulted in a large increase in TAG content in *N. gaditana* but a strong reduction of growth. An optimization of the balance between growth versus lipid accumulation was achieved by decreased expression of ZnCys by RNAi or by the CRISPR/Cas9 mediated insertion of an antibiotic resistance cassette into the 3' UTR of this gene (*17*).

Additional challenges for the development of improved *Nannochloropsis* strains

To develop improved strains optimized for biosynthetic yield, either by increasing flux into the target pathway or by disruption of competing pathways multiple specific modifications are likely to be required. For example, stacking gene modifications will be necessary to introduce new pathways or for the optimization of existing pathways. Furthermore, the development of marker-free strategies or the utilization of auxotrophic markers for selection of genetically modified strains are necessary for deployment of engineered strains into open ponds exposed to the environment.

Disrupting competing metabolic pathways may yield enhanced productivity of certain bioproducts. Carbohydrates are a competing sink for lipid production and polysaccharide biosynthetic genes have been targeted for repression in green algae and diatoms to increase lipid accumulation (*63, 121, 122*). The tough cell wall of the *Nannochloropsis* genus takes considerable cellular resources to construct and is an impediment to efficient processing of cells to obtain bioproducts. Therefore, *Nannochloropsis* strains with weakened walls may be superior

for bioproduction. Reducing the turnover of desired products may enhance their accumulation, and identification of the genes involved in product degradation will be an avenue towards enhancing the productivity of algae (*123, 124*).

The successful adoption of gene disruption technology will facilitate disruption of biosynthetic genes for essential metabolites, generating strains that require supplementation or gene complementation (auxotrophy). Nitrate reductase (NR)-deficient strains of *Chlamydomonas* strains can be complemented with a wild-type NR gene and selected (*125*). The NR gene has been targeted in several studies in various *Nannochloropsis* species (*14, 17, 27*) and diatoms (*126, 127*). In Chapter 4, I describe the generation of a NR knockout line that is non-transgenic using the episome CRISPR/Cas9 system. This line is an ideal chassis organism for development of an auxotrophic marker. Auxotrophic selection may be particularly useful when paired with episomal artificial chromosomes, enabling a nutrient selection pressure on episome maintenance.

Call for an open alga

Establishing a model organism (or bioproduct chassis) requires dissemination of the skills and tools developed to a wide network of scientists. A niche is developing for third-party repositories (TPR) that facilitate the maintenance of accumulated biological materials and are making these materials more accessible than ever before. However, for a TPR to be sustainable innovators must be willing to transfer their materials and utilize the TPR as part of their own workflow. A notable nonprofit TPR for plasmid and strain dissemination is Addgene that we are collaborating with by depositing the collection of *Nannochloropsis* engineering vectors developed in my studies. Some of the model *Nannochoropsis* species such as CCMP1779 (*N. oceanica*), CCMP526 (*N. gaditana*), and CCMP537 (*N. salina*) are publicly available from algae culture

collections (NCMA, https://ncma.bigelow.org/). We are making our engineered NR knockout strain publicly available through NCMA, as well (Chapter 4). As transgenic tools and knockout and chassis strains are produced, deposition with TPRs will accelerate innovation, allow researchers to do more with less, build genome-wide overexpression and gene disruption strain libraries, and establish quality controls and standardization in the field.

APPENDIX
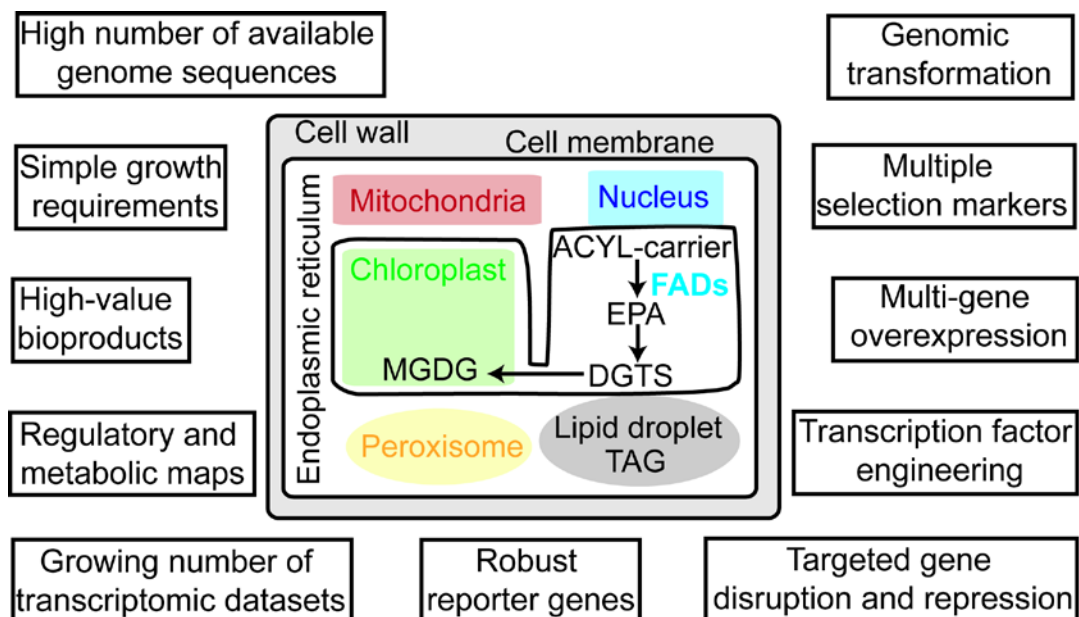
Figure 1.1 The *Nannochloropsis* genus as a chassis organism.
Advantageous growth characteristics and genome-wide datasets of the *Nannochloropsis* genus are listed on the left. Genetic engineering capabilities in the *Nannochloropsis* genus are listed on the right. A schematic of a *Nanochloropsis* cell is shown in the center, highlighting the EPA pathway and cell organization. Organelles and cell structures are identified by color and text, metabolites are in capital letters, and enzymes bolded.
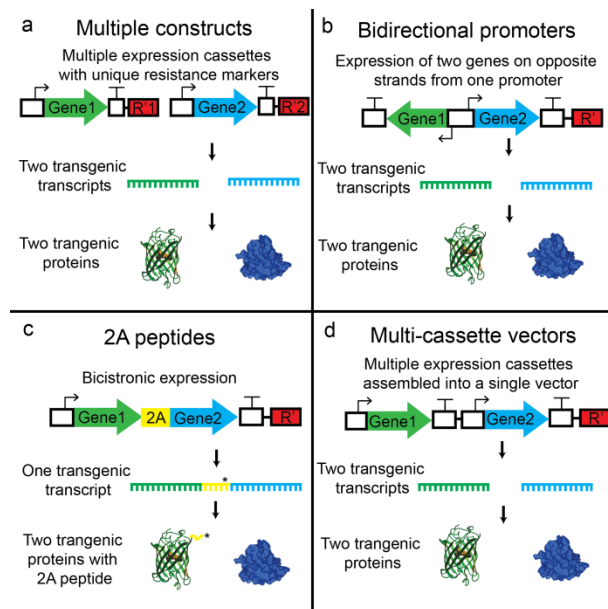
Figure 1.2 Multi-gene overexpression (gene stacking) strategies in *N. oceanica.*
Diagrams represent configuration for two protein coding transgenes. Gene1 represented by GFP is shown in green. Gene2 represented by NanoLuciferase is shown in blue. Promoters are indicated by arrows and terminators by a T (Synthetic Biology Open Language standard) (*128*). Resistance markers are indicated as R' (red). *Nannochloropsis* expression cassettes without a plasmid backbone are shown. a. Constructs with unique selection markers (indicated as R'1 and R'2, red) can be introduced into one line. b. Bidirectional promoters regulate transcription on both DNA strands, and express two transcripts. c. Sequences encoding 2A peptides are placed between two protein-encoding genes. A peptide bond is not formed between the two final amino acids (*) during translation. Two discrete proteins are produced, the N' terminal protein (green) contains the majority of the 2A peptide (yellow) and the C terminal protein (blue) contains the last amino acid of the 2A peptide (not shown). d. Assembly of multiple expression cassettes into a single construct to produce multiple transcripts from different promoters. The image of the GFP protein was obtained from the NIH Image Gallery (https://www.flickr.com/photos/nihgov/) and is adapted under the terms of CC BY-NC 2.0. The NanoLuciferase image is adapted from Hall et al. with permission (*97*).
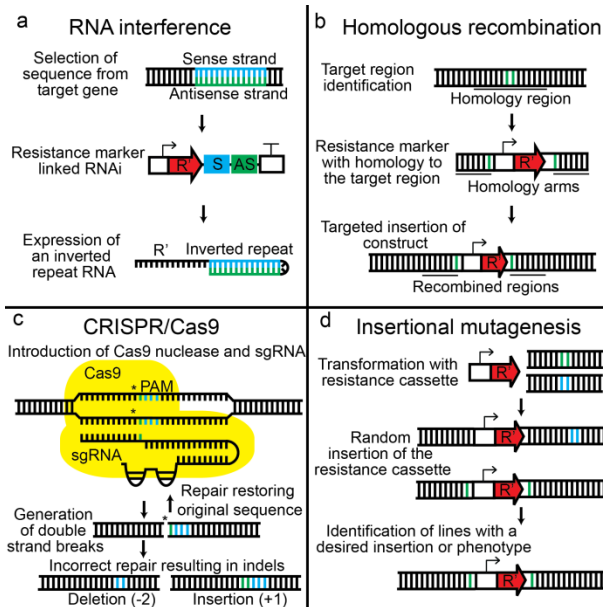
Figure 1.3 Gene repression and inactivation techniques in the *Nannochloropsis* genus. Promoters are indicated by arrows and terminators by a T (Synthetic Biology Open Language standard) (*128*). Resistance markers are indicated as R' (red). Extended DNA is indicated by an overhanging backbone, and free ends of DNA are indicated by blunt ends. a. RNA interference by expression of sense (blue) and antisense (green) sequences of a target gene at the 3' end of a resistance gene. The sense and antisense sequences form an inverted repeat on the resistance marker transcript. b. Homology arms are the sequences flanking a desired insertion site (green). A resistance cassette with flanking homology arms to a target region is introduced by transformation. Homologous recombination between the homology arms of the resistance cassette and the target region results in a disrupted gene. c. CRISPR/Cas9 techniques utilize the Cas9 nuclease and an sgRNA, forming ribonucleoprotein complex (yellow). The PAM site (blue) and the final 3' nucleotide of the guide sequence (green) are indicated. Double stranded cuts are produced in the 3' end of the target region. Incorrect repair of the strand break prevents further Cas9 action. Insertion or deletion mutations are most likely to occur. d. Insertional mutagenesis by transformation with a resistance cassette. The resistance cassette is randomly inserted throughout the genome, resulting in mutant lines with unique insertion sites. Desired mutants (green) are identified by a phenotype screen, target gene screen, or from a mutant library.

Table 1.1. Publicly available whole-genome datasets produced in *Nannochloropsis* species.
The first column contains the lead author and citation. Strains refers to the *Nannochloropsis* subspecies. The data and presentation are listed in the third column. A link to the complete data is listed in the fourth column (public access).

Citation	Strain(s)	Datasets	Public access
Radakovits (*13*)	*N. gaditana* CCMP526	Genome assembly	http://nannochloropsis.genomeprojectsolutions-databases.com/
Corteggiani - Carpinelli (*12*)	*N. gaditana* B-31	Genome browser and BLAST	www.nannochloropsis.org
Wang (*18*)	*N. oceanica* IMET1 and CCMP531, *N. granulata* CCMP529, *N. oculata* CCMP525, *N. salina* CCMP537, *N. gaditana* CCMP526	Genome browser and BLAST	http://www.bioenergychina.org:8989/
Vieler (*11*)	*N. oceanica* CCMP1779	Genome browser and BLAST	https://genome.jgi.doe.gov/Nanoce1779/Nanoce1779.home.html
Poliner (*45*)	*N. oceanica* CCMP1779	RNA-Seq during light:dark cycles	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE69460
Mühlroth (*44*)	*N. oceanica* CCMP1779	RNA-seq under phosphate deprivation	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE95774
Hu (*19*)	*N. oceanica* IMET1 and CCMP531, *N. granulata* CCMP529, *N. oculata* CCMP525, *N. salina* CCMP537, *N. gaditana* CCMP526	Predicted regulatory connections of lipid biosynthetic genes	http://www.singlecellcenter.org /en/NannoRegulationDatabase/Download.htm

Table 1.2. Genetic tools developed for the *Nannochloropsis* genus.
The first column contains the lead author and citation. Strains refers to the *Nannochloropsis* subspecies. Methods describe the genetic techniques used in the study. The target gene refers to the function of the primary gene manipulated. Reporter or epitope indicates if a protein tag was included, reporter abbreviations are FP - fluorescent protein, CP - chromoprotein, and GUS - β-glucuronidase. Promoters used to drive transcription of transgenes *in vivo* are listed; the species of origin for non-*Nannochloropsis* promoters are indicated in italics. Endogenous promoter abbreviations are LDSP - lipid droplet surface protein, β-tub - β-tubulin, UEP - ubiquitin extension protein, HSP - heat shock protein, EF- elongation factor, VCP - violaxanthin–chlorophyll a binding protein, and Ribi - ribosomal subunit bidirectional promoter. Selection agent refers to antibiotic selection used to isolate *Nannochloropsis* transformants.

Citation	Method(s)	Strain(s)	Target gene(s)	Reporter(s) or epitope(s)	Promoter(s)	Selection agent(s)
Vieler (*11*)	Transformation protocol	*N. oceanica* CCMP1779			LDSP promoter	Hygromycin
Radakovits (*13*)	Transformation protocol	*N. gaditana* CCMP526			β-tub, UEP, HSP	Zeocin
Li (*83*)	Transformation protocol	*N. oceanica* PP983 and MBIC10090, *N. granulata* MBIC10054, *N. salina* MBIC10063, *N. gaditana* CCAP849/5, *N.* oculata CCAP 849/1, *N. limnetica* KR1998/3		GUS	β-tub	Zeocin
Cha (*82*)	Transformation protocol	*N. sp.* UMT-M3			*CMV* 35S	Hygromycin
Ma (*129*)	Transformation protocol	*N. oculata* CS-179			*C. reinhardtii* HSP70A::RBCS2	Zeocin
Moog (*24*)	Subcellular localization	*N. oceanica* CCMP1779		Green FP	VCP	Zeocin
Shih (*99*)	Chromoprotein screening	*N. oculata* NIES-2146		shCP	*C. reinhardtii* HSP70A::RBCS2	
Kang (*79*)	Reporter protein	*N. salina* CCMP1776		sfCherry	β-tub and UEP	Zeocin
Chen (*130*)	Transformation	*N. oculata*	Fish growth		*C. reinhardtii*	

Table 1.2 (cont'd)

	protocol			hormone		HSP70A::RBCS2	
	Overexpression						
Poliner (*15*)	Multi-gene overexpression	*N. oceanica* CCMP1779	delta-9, delta-12, delta-5 fatty acid desaturases	Firefly luciferase, NanoLuciferase, Cyan FP (Cerulean variant)	EF, LDSP, Ribi	Hygromycin, Zeocin
Zienkiewicz (*31*)		*N. oceanica* CCMP1779	diacylglycerol acyltransferase type 2-5	Yellow FP (Venus variant)	EF	Hygromycin
Kaye (*91*)		*N. oceanica* CCMP1779	delta-12 fatty acid desaturase		LDSP	Hygromycin
Li (*95*)		*N. oceanica* CCMP1779	diacylglycerol acyltransferase type 2-7	Flag tag	Hsp20	Zeocin
Kang (*80*)		*N. salina* CCMP1776	basic helix loop helix 2	Flag tag	β-tub, UEP	Zeocin
Beacham (*81*)		*N. salina* CCAP 849/3	*S. cerevisiae* DGA1		*CMV* Tef, 35S	Hygromycin
Chen (*119*)		*N. oceanica* IMET1	malonyl CoA-acyl carrier protein transacylase	Flag tag	Hsp20	Zeocin
Wei (*131*)		*N. oceanica* IMET1	RuBisCO activase		β-tub, HSP70	Zeocin

Table 1.2 (cont'd)

Iwai (*117*)		*N. oceanica* NIES-2145	*C. reinhardtii* diacylglycerol acyltransferase type 2-4		VCP, *C. reinhardtii* SQD	Zeocin
Kang (*120*)		*N. salina* CCMP1776	*A. thaliana* WRINKLED1	Flag tag	β-tub, UEP	Zeocin
Kwon (*72*)		*N. salina* CCMP1776	basic leucine zipper domain 1	Flag tag	β-tub, UEP	Zeocin
	Gene disruption/ Transcriptional repression					
Kilian (*14*)	Gene disruption by HR	*N. oceanica* W2J3B	nitrate reductase		VCP	Zeocin, Hygromycin, Blasticidin
Wang (*27*)	Gene disruption by CRISPR/Cas9	*N. oceanica* IMET1	nitrate reductase		VCP, β-tub, V-ATPase	Hygromycin
Ajjawi (*17*)	Gene disruption by CRISPR/Cas9, repression by CRISPR/Cas9, repression by RNAi	*N. gaditana* CCMP1894	ZnCys	Green FP, Flag tag	initiation factor 4AIII, 60S ribosomal protein L24, initiation factor 3, TCT	Hygromycin, Blasticidin
Wei (*16*)	Repression by RNAi	*N. oceanica* IMET1 and CCMP1779	carbonic anhydrase		β-tub	Zeocin
Dolch (*26*)	Gene disruption by HR	*N. gaditana* CCMP526	palmitate elongase		UEP	Zeocin

Table 1.2 (cont'd)

	Gene disruption or transcriptional repression, and overexpression					
Wei (*90*)	Overexpression and repression by RNAi	*N. oceanica* IMET1	diacylglycerol acyltransferase type 1-1A	Green FP	VCP, UEP, β-tub	Zeocin
Gee (*96*)	Gene disruption by HR, complementation	*N. oceanica* CCMP1779	carbonic anhydrase	Flag tag, Yellow FP (Venus variant)	UEP	Zeocin
Nobusawa (*25*)	Overexpression and gene disruption by HR	*N. oceanica* NIES-2145	lysophosphatidic acid acyltransferase 1-4	Green FP, Yellow FP (Venus variant)	LDSP	Zeocin, Hygromycin
Ma (*37*)	Overexpression and repression by RNAI	*N. salina* CCMP537	pyruvate dehydrogenase kinase	Green FP	β-tub, VCP	Zeocin

31

Table 1.3. Effective antibiotic selection agents and resistance genes for *N. oceanica*
The first, second, and third columns contain the antibiotic common name, its molecular class, and the lethal concentration for *N. oceanica* CCMP1779, respectively. The fourth, fifth, and six column contain the resistance gene name, organism from which it was isolated, and the number of amino acids of the resulting protein, respectively.

Antibiotic	Molecular class	Effective concentration (µg/ml)	Resistance gene	Originating organism	Amino acids
Hygromycin	aminoglycoside	100-500	Hygromycin-B phosphotransferase	*Escherichia coli*	441
Zeocin	glycopeptide	5	Bleomycin resistance protein	*Streptoalloteichus hindustanus*	124
Blasticidin	peptidyl nucleoside	50	Blasticidin-S deaminase	*Aspergillus terreus*	129
G418	aminoglycoside	250	Aminoglycoside 3'-phosphotransferase	*Klebsiella pneumoniae*	266
Nourseothricin	aminoglycoside	100	Nourseothricin acetyltransferase	*Streptomyces noursei*	189

Chapter 2. Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1779 under light:dark cycles

ABSTRACT

Nannochloropsis oceanica CCMP1779 is a marine unicellular heterokont and an emerging reference species for basic research on oleogenic microalgae with biotechnological relevance. We investigated its physiology and transcriptome under light:dark cycles. We observed oscillations in lipid content and a predominance of cell division in the first half of the dark phase. Globally, more than 60% of the genes cycled in *N. oceanica* CCMP1779, with gene expression peaking at different times of the day. Interestingly, the phase of expression of genes involved in certain biological processes was conserved across photosynthetic lineages. Furthermore, in agreement with our physiological studies we found the processes of lipid metabolism and cell division enriched in cycling genes. For example, there was tight coordination of genes involved in the lower part of glycolysis, fatty acid synthesis and lipid production at dawn preceding lipid accumulation during the day. Our results suggest that diel lipid storage plays a key role for *N. oceanica* CCMP1779 growth under natural conditions making this alga a promising model to gain a basic mechanistic understanding of triacylglycerol production in photosynthetic cells. Our data will help the formulation of new hypotheses on the transcriptional control of cell growth and metabolism in *Nannochloropsis*.

INTRODUCTION

The daily light:dark cycle is a major environmental change affecting photosynthetic organisms. During the light period photoautotrophs rely on solar energy to drive anabolic processes including the transient storage of a fraction of the fixed carbon for later use in cellular processes

during the dark period. Other processes controlled by light:dark cycles in photosynthetic organisms include cell division, stress sensitivity, chemotaxis, nutrient uptake, and phototaxis (*132-135*). Studies of circadian clock mutants that display disruptions in the timing of daily events indicate that this coordination is necessary for optimal growth and survival (*136, 137*). Moreover, these rhythms not only affect the physiology of single organisms but could also influence the structure of phytoplankton communities (*138*).

Nannochloropsis species are small unicellular algae belonging to the heterokont lineage. They live in marine, and fresh and brackish water environments. Marine *Nannochloropsis* species are used as a source of fish food and omega-3 fatty acids (*139*) and, due to their high lipid content, have been considered as a potential source of biofuels (*6, 140-142*). The genomes of several *Nannochloropsis* species collected at different locations around the globe have been recently sequenced (*18*). Their small genomes range from ~25 to ~32 Mb and contain ~9,000-12,000 genes. *Nannochloropsis* species all display an expansion of gene families involved in lipid biosynthesis (*11, 18*) and accumulate triacylglycerols, particularly under nutrient limiting conditions (*11-13, 22, 143*). In accordance, transcriptomic analysis of different *Nannochloropsis* species have been focused on identifying changes occurring under nitrogen deprivation (*11-13, 22, 143*).

In different *Nannochloropsis* species cell division is synchronized and carbohydrate, lipid and protein contents per cell oscillate under daily light:dark cycles. However, little is known about changes in gene expression occurring under diel conditions. In contrast to plants and green algae (*144-146*), information on global gene expression changes under diel cycles in heterkonts is limited. Only one study on the diatom *Phaeodactylum tricornutum* analyzed global gene expression changes over the course of a light:dark cycle (*49*). In this report, we characterize gene

34

expression, cell growth and metabolite content of *Nannochloropsis oceanica* CCMP1779 every three hours under day-night cycles of 12 h light and 12 h dark. We also extend the annotation of cell cycle and central metabolism genes and use changes in expression under diel cycle together with predictions of subcellular localization of central carbon metabolism genes to propose a model of primary metabolism under light:dark cycles in *N. oceanica* CCMP1779.

RESULTS

N. oceanica CCMP1779 growth and metabolite content under diel conditions

Many algae grown under light:dark cycles display synchronized cell division (*147*). We investigated cell division rates in *N. oceanica* CCMP1779 under diel conditions of 12 h light and 12 dark (LD) and observed that cell division was confined to the first 6 h of the dark period. Cell growth occurred only during the light period with cell mass increasing from a minimum at ZT0 (Zeitgeber time, h after last lights on) to a maximum at ZT12 (Figure 2.1). This pattern of growth is similar to what has been observed in other phytoplankton species under both laboratory conditions (*148-150*) as well as *in situ* (*151*). In contrast, in some diatom species cell division occurs continuously during exponential growth (*50-52*).

Nannochloropsis species are able to produce large amounts of lipids and when grown under light:dark cycles accumulate lipids only during the light phase (*55, 152*). We therefore quantified the content of neutral, storage and polar membrane lipids over a 24 h diel cycle in *N. oceanica* CCMP1779. The content of the different lipids followed the same pattern during the LD cycle increasing during the day and decreasing during the night including triacylglycerol (TAG), digalactosyldiacylglycerol (DGDG), monogalactosyldiacylglycerol (MGDG), sulfoquinovosuldiacylglycerol (SQDG), phosphatidylcholine, phosphatylethanolamine,

phosphatidylglycerol and phosphatylinositol (Figure 2.2). By the end of the light period TAG was the most abundant lipid followed by the membrane lipids MGDG and phosphatidylcholine (Figure 2.2). The total fatty acid content was similar to previous reports cellular fatty acid content during a light:dark cycle in other *Nannochloropsis* species (*55, 152*) and glycerolipids content as percent dry-weight was similar to previously reported values under similar light intensities (*22*). The oscillations of lipid content per cell might be just a reflection of cell growth during the day and division during the night. However, we observed that TAG was the only lipid that showed oscillations as a percentage of dry-weight indicating that in this case the change in content was independent of cell size changes (Figure 2.2). This result suggests a role of TAG in transitory carbon storage under LD cycles in *N. oceanica* CCMP1779. Glucose is the most abundant hexose in *N. oceanica* strains (*9, 11*). Total glucose content per cell and its level on a percent dry weight oscillated under diel conditions (Figure 2.2). The decrease in glucose content during the night could be due to shedding of the cell wall during cell division, which is composed mainly of cellulose (*11, 65*), and the use of glucose-containing storage compounds, such as intracellular laminarin (*9*).

Oscillations in global gene expression in *N. oceanica* under day:night cycles

To understand the mechanisms driving the oscillations in growth and metabolism we analyzed global changes in gene expression in LD in *N. oceanica* CCMP1779 using transcriptome sequencing (RNA-Seq). We analyzed gene expression in two independent biological replicates every three hours over the course of one day. We confirmed similar expression patterns by reverse transcription quantitative PCR for genes expressed at different times of day in independent experiments (Figure 2.3). For data analysis and graphic representation using heat

maps we used the two biological replicates as representing two consecutive days as has been done for analyses of diel datasets in other species (*153*). COSPOT (*154*) and an application of the discrete Fourier transform (DFT) were used for the identification of cycling genes (see Experimental Procedures). We found that 7433 genes (61.8% of the genome) in *N. oceanica* CCMP1779 were predicted to be cyclic. Of these predictions, the majority (59%) were common to both COSPOT and the DFT analysis methods. The overlap between the two methods was high since the intersection between COPSOT and DFT covered ~70% of DFT predictions and ~80% of COSPOT predictions (Table 2.1). We also observed many genes displaying oscillatory expression patterns with narrow peaks of RNA levels that were not identified as cycling in our initial analysis. We therefore used EdgeR to identify an additional 199 cycling genes. In summary, a total of 7632 genes or 63.7% of the genome cycled under LD conditions in *N. oceanica* CCMP1779 with phases of expression at different times of the day (Figure 2.4a, Data 2.1). This percentage is higher than what has been observed for land plants, where 30-40% of genes cycle under day:night conditions (*145, 153, 155*). It is also higher than a recent estimate of ~45% cycling genes for the diatom *Phaedodactylum tricornutum* (*49*) but lower than those for some cyanobacteria (*156*). The distribution of expression phase was approximately bimodal with most genes peaking close to dawn (ZT21-ZT0) and in anticipation of dusk (ZT9) (Figure 2.4a). This pattern is similar to the one observed in many photosynthetic species (*145, 146, 153, 155-158*).

Using an existing genome annotation (*11*), the enrichment GO terms were assessed within each phase group and 195 terms contained a significantly over-represented number of cycling genes in at least one phase group (multiple testing adjusted *p*<0.05; Figure 2.4b). A summary of the top enriched GO terms for each phase can be found in Figure 2.4c. The phase

groups at ZT9 and ZT21 hours account for 60% of enriched GO terms with 67 and 50 terms respectively. These results indicate that changes in gene expression can be associated with different physiological processes occurring at specific times of day. For example, we found enriched GO terms relevant to terpenoid and chlorophyll biosynthesis in the early morning, DNA replication in the middle of the light period and terms associated with translation at the end of the night period (Figure 2.4c). Several of these phase-specific enrichments are similar to those observed in other photosynthetic organisms grown in both the laboratory as well as under natural conditions (*145, 146, 153, 155-159*).

Genes involved in cell division display strong diurnal oscillations in *N. oceanica* CCMP1779

Synchronization of the cell cycle in unicellular alga could affect the synchronization of global gene expression under diel conditions and play a role in the regulation of metabolism (*160-162*). We identified several cell cycle genes in *N. oceanica* CCMP1779 including 10 cyclin dependent kinases (CDK) and 10 cyclin related genes (CYC) (Figure 2.5, 2.6; Data 2.2). All these genes displayed oscillations in mRNA levels during the LD cycle (Figure 2.7). However, some of these genes were not identified as cycling using our initial analysis pipeline, probably due to their narrow peaks of expression. We had observed an enrichment of GO-terms associated with DNA synthesis at ZT6 (Figure 2.4c). The interaction between a retinoblastoma protein and E2F/DP transcription factors regulate the G1-S transition in most eukaryotes (*163*). The expression of a putative retinoblastoma-like gene (NannoCCMP1779|10679, for brevity 10679) preceded the expression of one putative *DP-1* transcription regulator (3316) one *E2F* gene (4416), which peaked at ZT6 (Figure 2.7). In the diatom *Phaedodactylum tricornutum* the G1-S transition is light regulated and the expression of several cell cycle genes oscillates under light:dark cycles

38

(*51*). Interestingly, in *N. oceanica* CCMP1779 putative CDKA1s (4688), CDKD1 (2601) and CDKA2 (1586) genes display a similar phase of expression as their homologs in *P. tricornutum* indicating that cell cycle progression might occur in similar fashion as in diatoms under LD conditions (Figure 2.7). In order to further characterize the cell cycle in *N. oceanica* CCMP1779 we performed DNA content analyses by flow cytometry under diel conditions (Figure 2.7b, c). These results confirmed that under our growth conditions not all cells divided each day. Most of the cells were in G1 in the morning and DNA synthesis (S-phase) started about ZT6. At ZT15, three hours into the dark period some cells had two and four times the G1 DNA content (Figure 2.7b). Accordingly, cells with two and four nuclei were present during the dark period (Figure 2.7d) and some cells divided into four daughter cells at that time (Figure 2.8). These results correspond well with RNA levels of cell cycle genes.

We also observed that the RNA levels of genes potentially involved in organelle division, such as genes encoding FtsZ-like proteins (10028, 11167, 10560), an ARC5/DRP5B-like protein (7886) and one MinD-like protein (5101) cycled. Their expression peaked between the middle and end of the day (Figure 2.7). These results suggest a coordination of organelle and cell division during the diel cycle as has been observed in diatoms (*161*). As is the case for other heterokonts, *Nannochloropsis* plastids are surrounded by four membranes. The outer most membrane continues directly to the outer nuclear envelope and tight coordination of the nuclear and plastid division has been observed in *Nannochloropsis oculata* (*33*).

The RNA content of genes involved in carbon assimilation peaks at dawn

The *N. oceanica* CCMP1779 genome encodes all the enzymes necessary for the Calvin-Benson cycle (*11*). Several reactions of the Calvin-Benson cycle are shared with other metabolic

pathways, and we therefore focused our analysis on reactions specific to carbon assimilation. We observed that the strongest expressed genes encoding phosphoribulokinase (11364) and both putative sedoheptulose bisphosphatases (2947, 4791) were coregulated peaking within the first 3 h of the light period (Data 2.2, Figure 2.9). Interestingly, the highest expressed genes potentially involved in carbon concentrating mechanisms were on at dawn (Data 2.2). For example, the highest expressed putative α-carbonic anhydrase gene (6698) displayed strong oscillations with changes in amplitude of almost 60-fold with a maximum in RNA content at 3 h into the light period. We have also identified two genes that peak at dawn encoding putative carbonate transporters. Our phylogenetic analysis showed that these genes belong to the SLC4 family of proteins present in other marine heterokonts and marine green algae but are likely to be different than the recently characterized *Phaeodactylum* bicarbonate pumps (*164*) (Figure 2.10, Data 2.2). These observations are consistent with the hypothesis that these genes are involved in processes occurring during the light period (Figure 2.9). Finally, most genes involved in the light reactions of photosynthesis also peaked close to dawn or during the first half of the light period (Data 2.2), suggesting a tight transcriptional control of the photosynthetic process in *N. oceanica* CCMP1779.

The presence of genes encoding a phosphoenolpyruvate (PEP) carboxylase (3970, PEPC), malic enzymes (9004, 4675, ME), a PEP carboxykinase (6821, PCK), pyruvate phosphate dikinase genes (2768) in *Nannochloropsis* species has led to the speculation of the presence of a C4-type carbon fixation mechanism (*11-13*). The expression of these genes peaked at the end of the night period (Data 2.2). The highest expressed of these genes, *PEPC* (3970), encodes an enzyme with a putative mitochondrial localization signal and might be similar to the *T. pseudonana* mitochondria-localized PEPC (*165*). One of the two ME contains a signal

peptide, is similar to other NADP-ME, and could localize in the plastid (Figure 2.11, Data 2.2). In addition to a potential role in a CO_2 concentrating mechanism it has been suggested that these proteins could function in the generation of reducing equivalents for lipid biosynthesis (*166, 167*), the dissipation of excess light energy, and/or pH regulation (*168*).

Carbohydrate metabolism under light:dark cycles

Glucose and mannitol in monomeric and polymeric forms are the major intracellular carbohydrates in *N. oceanica* IMET1 (*9*) and cellulose is the main carbohydrate component of the cell walls of different *Nannochloropsis* species (*11, 65*). *Nannochloropsis* genomes encode for large families of glycosyl transferases (GT) and glycosyl hydrolases (GH)(*11, 18, 65*). However, the details of polysaccharide biosynthesis in *Nannochloropsis* species are still unknown. Clustering analysis revealed several large sets of highly coregulated *GT* and *GH* genes (Figure 2.12). For example, we found three large groups of GT and GH encoding genes peaking at distinct times between ZT9 and ZT15 that could be involved in specific steps necessary for cell wall restructuring and biosynthesis during cell division.

Laminarin digestibility assays indicate that 20% of *N. oceanica* CCMP1779 insoluble polysaccharides are made up of β-1,3-glucans (*11*). The gene encoding the only β-1,3-glucan synthase (9088)(GBS) in *N. oceanica* CCMP1779, which is closely related to GT48 family proteins in other heterokonts (Figure 2.13), peaked at ZT15 (Data 2.2). These results suggest that it might be involved in the synthesis of polymers similar to the ones found in other heterokonts. It is expressed earlier than genes involved in UDP-glucose synthesis such as phospholucoisomerase (11835, 2243), phosphoglucomutase (4161) and UDP-glucose pyrophosphorylase encoding genes (6938, 11841; Data 2.2), which peak at dawn indicating a

potential role in polysaccharide biosynthesis during the light period. In contrast, we observed that genes involved in the metabolization of hexoses, such as a hexokinase (11432), two of the three putative phosphofructokinases (6373, 10476) and the two glucose-6-phosphate dehydrogenases (640, 2822) peaked at ZT9 (Data 2.2) indicating that they might be involved in storage carbohydrate metabolization during the night period (Figure 2.2, Figure 2.9).

Genes potentially involved in mannitol biosynthesis also display diel oscillations in *N. oceanica* CCMP1779. For example genes encoding a mannitol 1-phosphate dehydrogenase (10023) and a partial mannitol 1-phosphate phosphatase gene (10755) with similarity to *E. siliculosus* EsM1Pase1 and EsM1Pase2, peaked in the middle of the light period (Data 2.2, Figure 2.14). In contrast, the expression of a gene encoding a putative mannitol 2-dehydrogenase (5782, *M2DH*) with 43% identity to *E. siliculosus* M2DH (NCBI: CBJ29121.1), which catalyzes the conversion of mannitol to fructose, peaked at ZT21 (Figure 2.14). Moreover, *M2DH* was tightly co-expressed with a putative fructokinase-encoding gene (6825) that could close the mannitol cycle generating fructose-6-phosphate (Figure 2.14, Figure 2.9). In brown algae mannitol content as well as the expression of mannitol metabolic genes also oscillate under light:dark cycles (*169-171*) and our results suggest this might also occur in *N. oceanica* CCMP1779.

Acetyl-CoA metabolism is temporally and spatially segregated in *N. oceanica* CCMP1779

N. oceanica CCMP1779 plastids appear to contain all the enzymes necessary for the conversion of hexose-phosphates to acetyl-CoA (Data 2.2). Our current predictions indicate that *N. oceanica* CCMP1779 mitochondria are also able to metabolize triose-phosphates to acetyl-CoA, however this species seems to lack a cytosolic enolase (Figure 2.9, Data 2.2). This compartmentalization

is similar to what has been proposed in diatoms, where a complete cytosolic glycolytic pathway is lacking in several species (*172*). Based on the observed increase in lipid content during the light period in *N. oceanica* CCMP1779 (Figure 2.2) we expected a higher expression of genes involved in acetyl-CoA biosynthesis in the first part of the day. In accordance, the genes encoding enzymes of both the plastid and mitochondrial bottom parts of glycolysis displayed high degree of co-expression with maxima close to dawn (Figure 2.9, Data 2.2). In particular, all four pyruvate kinase genes (PK) (497, 5759, 10510, 8741) displayed strong oscillations with peaks within 3 h of dawn (Figure 2.9, Data 2.2). This observation contrasts with the case of *P. tricornutum* in which the genes encoding the highest expressed PK isoenzymes are strongly expressed at the end of the day (*49*). Moreover genes encoding putative plastid pyruvate dehydrogenase complex (PDC) subunits displayed very similar expression profiles during the LD cycle peaking close to dawn (Figure 2.15). In contrast, several putative mitochondrial PDC subunits displayed only weak RNA oscillations with higher levels during night period (Figure 2.15). Even though gene expression does not directly indicate metabolic flux, based on our results one might hypothesize that there is not a large flux from carbohydrates into acetyl-CoA at night in *N. oceanica* CCMP1779. This hypothesis implies that at night most acetyl-CoA is generated by β-oxidation of lipids.

We also identified other enzymes that could be involved in acetyl-CoA production in the cytosol in *N. oceanica* CCMP1779, including a putative cytosolic ATP-citrate lyase (10051) and two acetyl-CoA synthetases (956, 1243). The RNA levels of the ATP-citrate lysase gene peaked at dusk and the highest expressed putative acetyl-CoA synthetases encoding gene (956) peaked at ZT15 (Data 2.2, Figure 2.9). Based on their lack of co-expression with other fatty acid biosynthesis genes, these genes are not likely to be involved in lipid biosynthesis and might

43

generate acetyl-CoA for other processes such as histone acetylation in the nucleus (*173*). However, radiolabeling studies in *Nannochloropsis* sp. suggest that long chain polyunsaturated fatty acids (LC-PUFA) can be produced by an extraplastidic pathway, which could use a cytosolic acetyl-CoA source (*61*).

The expression of fatty acid synthesis genes precede lipid accumulation during the day

We found that most genes likely to be involved in fatty acid synthesis cycled under LD conditions in *N. oceanica* CCMP1779. Acetyl-coA carboxylase (ACC) catalyzes the carboxylation of acetyl-coA to form malonyl-coA, which is then used for fatty acid biosynthesis. Several heterokonts, including diatoms, contain two forms of monomeric ACC, one of which contains a signaling peptide and could localize to the plastid (*174*). However, as in other *Nannochloropsis* species (*12, 22*), we did not find genes encoding full length monomeric ACCs in *N. oceanica* CCMP1779. We identified eight genes containing ACC specific domains that might encode subunits of heteromeric ACCs, as found in seed plants, or form part of a yet unidentified monomeric ACC (*175*)(Data 2.2). Visual inspection indicated that the expression of five of these genes peaked at dawn in a similar manner as a plastid malonyltransferase (6562) and putative components of the plastidtype II fatty acid synthase (FAS) (Figure 2.15). The peak of expression of genes encoding putative components of the mitochondria type II FAS peaked slightly earlier during the mid-late night (Figure 2.15). Taken together our observations suggest the coordination between the glycolysis and fatty acid synthesis for lipid production in *N. oceanica* CCMP1779 (Figure 2.9).

Three putative type I FAS-like genes had been previously identified in *N. oceanica* CCMP1779 (3502, 6720, 1983) (*11*). Phylogenetic analysis showed that these proteins are more

similar to fungal polyketide synthases (Figure 2.16). Moreover, the expression of these three genes cycle at different phases of the LD cycle (Data 2.2) indicating that each of them might have slightly different functions and act independently of fatty acid biosynthesis in this alga.

Regulation of gene expression related to lipid biosynthesis over the diel cycle.

In order to identify the mechanism responsible for the observed diel oscillations in lipid content in *N. oceanica* CCMP1779 (Figure 2.2) we also analyzed the expression of genes putatively involved in lipid biosynthesis in our dataset. By visual inspection the expression of 88% (62/70) of genes involved in lipid biosynthesis cycled under diel conditions. Genome analyses indicate the presence of a plastid as well as an endoplasmatic reticulum (ER) localized glycerolipid biosynthetic pathways in *N. oceanica* CCMP1779 (*176*). We observed that all three genes involved in the plastid diacylglycerol (DAG) synthesis (4533, 2512, 4742) displayed coordinated expression with a peak at the dark to light transition (Figure 2.15). The putative plastid galactolipid synthases MGD1 (10634) and DGD1 (4384) (*11*) cycled with expression peaking also at dawn (Figure 2.15). The SQDG biosynthetic genes (6637, 4348) also showed cyclic expression with a maximum during the light period (Figure 2.15). In summary, the expression of the plastid glycerolipid genes in *N. oceanica* CCMP1779 displayed high coordination with fatty acid synthesis genes at the transcriptional level so their expression preceded the observed lipid accumulation during the light period (Figure 2.9). In contrast, lipid biosynthesis in the ER did not appear to be as tightly coordinated over the diel cycle (Figure 2.15). Although most of the genes involved in DAG production in the ER cycled with a peak close to dawn, the genes involved in phosphatidylcholine, phosphatylethanolamine and phosphatylinositol biosynthesis peaked at different times of day (Data 2.2).

Nannochloropsis species produce high amounts of LC-PUFA (*11, 22, 141, 177*). We observed that in *N. oceanica* CCMP1779 all genes encoding fatty acid desaturases (11542, 6416, 2179, 5130, 5794, 126) were strongly co-expressed throughout the LD cycle with maxima at ZT3 (Figure 2.15). In contrast, the expression of most elongase encoding genes peaked at different times between dusk and dawn (Figure 2.15). Most of the EPA in *N. oceanica* CCMP1779 is found in MGDG and the betain lipid DGTS (*11*). The expression of the bifunctional monomeric DGTS synthase (10012) rapidly increased during the first hours of the light period and was co-expressed with the fatty acid desaturase genes (Figure 2.15).

We observed oscillations in TAG content on a per dry weight basis in *N. oceanica* CCMP1779 (Figure 2.2). TAG can be generated from DAG via diacylglycerol acyltransferase (DGAT) and the *Nannochloropsis* genus displays a large expansion of the DGAT/MGAT gene family (*18*). The *N. oceanica* CCMP1779 genome encodes for one DGAT-1 and 10 DGAT-2 genes (*11*). The expression of most DGAT-2 genes cycled with peaks of expression at different times during the light period (Figure 2.15). In contrast, DGAT-1 (3520) RNA levels peaked at dusk (Figure 2.15). TAG can also be synthesized by the transfer of a glycerolipid-bound fatty acid to DAG. The two phospholipid:DAG tranferases identified in *N. oceanica* CCMP1779 (2212, 8602) were expressed during the light period (Figure 2.15). TAG is stored in lipid droplets formed by a membrane lipid monolayer and a neutral lipid core. Lipid droplet proteins have been characterized in a number of algal species, including the major lipid droplet (MLDP) in *Chlamydomonas* and lipid droplet surface protein (LDSP) in *N. oceanica* CCMP1779 (*104, 176*). MLDP and LDSP protein levels were found to correlate with TAG content during nitrogen stress (*104, 176*). We observed that *LDSP* RNA levels strongly increase during the first half of

the day and that LDSP protein levels correlate with TAG content over the diel cycle (Figure 2.17).

Lipid degradation and the TCA cycle.

Genome analyses indicate that β-oxidation occurs in both the peroxisome and mitochondria in *N. oceanica* CCMP1779 (*11*). The expression of most genes involved in the early steps of β-oxidation peaked between dusk and dawn during diel cycles and might be involved in the degradation of lipids in the dark (Figure 2.18). For example, the expression of the strongest expressed mitochondrial acyl-CoA dehydrogenase encoding gene (6983), the mitochondrial multifunctional enzyme α-subunit (4030) and three putative β-subunits/thiolases (2420, 11662, 3591) peaked at dusk (Figure 2.18). We also identified an acyl-CoA synthetase (LACS) with a putative mitochondrial localization signal expressed at dusk (11454), suggesting a role in mitochondrial β-oxidation (Figure 2.18). In contrast, the expression of genes involved in peroxisomal β-oxidation, such as a peroxisomal acyl-CoA oxidase (10648), the peroxisomal multifunctional enzyme (6411) and a putative peroxisomal LACS (4001) peaked later during the night period (Figure 2.18). We had previously identified several genes that could be involved in lipid degradation (*11*). Although most of these genes were expressed either close to dusk or during the night period in a similar manner as β-oxidation genes, a group was expressed preferentially during the early light phase indicating that they might be involved in acyl editing during lipid biosynthesis (*178*) (Figure 2.19).

We found that most genes involved in the mitochondrial TCA cycle were also co-expressed with peaks in RNA levels around dusk (Figure 2.18). Interestingly, all four genes encoding pyruvate carboxylases (PYC) in *N. oceanica* CCMP1779 (6175, 7693, 4504, 7838)

displayed similar patterns of expression with maximum RNA levels at the end of the day (Data 2.2). In contrast, in the diatom *Phaeodactylum triconutum* genes encoding a plastid PYC and a mitochondrial PYC are expressed at dawn and early during the night period respectively (*49*). Our results indicate that PYCs could play an anapleurotic role in the TCA cycle at night, further suggesting that in the dark most of the acetyl-CoA derives from fatty acid degradation in *N. oceanica* CCMP1779.

Cyclic expression of transcriptional regulators

We observed that the mRNA levels of about 60% of the *N. oceanica* CCMP1779 genes cycle under diel conditions (Figure 2.4a). These changes could be a result of diurnal oscillations of light and/or metabolite levels or directly caused by a circadian oscillator. In order to identify potential regulatory proteins we analyzed the expression of genes encoding transcription factors. About 64% of the DNA binding transcription factors and 56% of other transcriptional regulators identified in *N. oceanica* CCMP1779 (*11*) displayed cycling mRNA levels under diel conditions (Figure 2.20a). We did not observe a particular enrichment in either phase of expression or number of cycling members in the different gene families, but we found that AP2, CCAAT, TIG, BSD and HMG families contain less than 50% of cycling members based on computational analysis (Data 2.2). We had identified four genes encoding putative blue light receptors in *N. oceanica* CCMP1779, three Aureochrome-like proteins (AUREO2/3/4) and a cryptochrome/photolyase family protein (CPF1) (*11*). In other organisms these proteins are able to directly affect gene expression. For example, AUREO1a in *Phaeodactylum tricornutum* regulates the expression of the diatom specific cyclin 2 gene by binding to its promoter (*47*, and PtCPF1 is able to act as transcriptional repressor (*75*). The expression of *AUREO3* (10447) and 4

(5385) as well as *CPF1* (4809) displayed oscillations in gene expression with peaks close to dawn (Figure 2.20b) indicating that they might be involved in light dependent regulation of gene expression at dawn. Interestingly, Pt*CPF1* and Pt*AUREO1c* also have higher expression levels at dawn than at dusk (*52*).

Nucleosome modifications lead to changes in chromatin state and can influence gene expression (*179*). Using HMMER analysis (*180*) we identified genes encoding proteins potentially involved in the regulation of histone modifications, these included putative histone deacetylases, histone methylases, histone acetyltransferases, Chromosome Condensation 1-like proteins and SNF2 related proteins. We found that the expression of large number of these genes peaked at ZT21, with a second group peaking during the first half of the day (Figure 2.21). These results show that *N. oceanica* CCMP1779 tightly controls the diurnal expression of chromatin modification factors, which could be involved in the global cyclic gene expression as has been shown to occur in mammals (*181*).

Some plastid encoded genes displayed cycling RNA levels (Data 2.2). Although the *N. oceanica* CCMP1779 chloroplast genome has not been assembled, we identified a few genes likely to be chloroplast encoded based on the localization in the *N. oceanica* IMET1 chloroplast genome (*28*). It has been recently shown that in *Arabidopsis* the nuclear encoded sigma 70-like factors affect the diel and circadian oscillations of chloroplast encoded genes (*182*). All five sigma 70-like transcriptional regulator encoding genes in *N. oceanica* CCMP1779 displayed diel oscillations in gene expression at different times of the day (Figure 2.20c). Thus, they could be involved in regulating cyclic plastid expression in *N. oceanica* CCMP1779. Moreover, the expression of three of the four putative mitochondrial transcription termination factor-related

gene (mTERF) also displayed diel oscillations in *N. oceanica* (Figure 2.20d) indicating that mitochondrial gene expression might also be diurnally controlled (*183*).

DISCUSSION

The large number of genes displaying cyclic expression indicates the existence of significant transcriptional control of growth and metabolic rhythms in *N. oceanica* CCMP1779. We found that the timing of expression of genes related to some metabolic pathways was conserved among taxa of photosynthetic organisms. These include the expression close to dawn of carbon fixation and protein synthesis related genes, the expression of photosynthesis and pigment biosynthesis genes during the first half of the light period and the expression of genes related to the TCA cycle and β-oxidation close to dusk (*49, 145, 158, 160*). Our results indicate that lipids play a major role as photosynthetic carbon sinks in *N. oceanica*. We found strong co-expression of genes involved in the lower part of glycolysis in both plastid and mitochondria with fatty acid biosynthesis genes at dawn suggesting cooperation for lipid production. Our results also suggest that *N. oceanica* CCMP1779 use TAG partly as transitory carbon storage under diel conditions. In addition, we observed that the RNA level of genes involved in TAG synthesis and storage as well as the expression of a lipid droplet protein displayed oscillations under diel conditions that correlated with TAG content. We expect that the data presented here will provide not only a resource to the community interested in the evolution of cyclic gene expression, but will also aide the future exploration of TAG biosynthesis and degradation during natural day-night cycles. Understanding this process will be crucial for optimizing TAG yield in biotechnologically relevant algae.

EXPERIMENTAL PROCEDURES

Culture conditions

For all experiments *N. oceanica* CCMP1779 cells were grown in flasks containing f/2 media under 12 h light:12 h dark cycles at 22°C under agitation (100 rpm). Light intensity during light periods was 40 µmol s^{-1} m^{-2}. ZT0-9 and ZT12-21 sets of samples were grown in two identical incubators set to reversed light:dark cycles. The ZT0 samples were collected in the dark and ZT12 samples in the light. Cell numbers were determined using a Beckman Coulter Z/2.

Lipid analyses

For each biological replicate, we used one 100 ml culture grown in a 500 ml flask. Cell numbers were ~ 30 10^6 ml^{-1} at time of harvest. For dry weight determination 40 ml of culture were filtrated through pre-dried and pre-weighed Whatman GF/F or C filters and dried for 48 hours in a vacuum desiccator. For metabolite analysis, 40 ml of each culture were collected by centrifugation at 4000 g for 10 minutes. Cells were transferred to 1.5 ml tubes and frozen in liquid nitrogen after a final centrifugation step. For polar lipid separation shown in Figure 2.2, lipids were extracted from the frozen pellet using the Bligh Dyer protocol followed by a 0.9 % KCl wash (*184*). Lipids were separated by thin layer chromatography using either (80:32:16:3:3) chloroform: acetone: methanol: acetic acid: water (*55*), or (25:25:25:10:4) methyl acetate: isopropanol: chloroform: methanol: KCl (0.25%) (*185*) based on the separation resolution of known standards and the use of headgroup specific dyes. For data shown in Figure 2.4, frozen pellets were ground using a TissueLyser II (Qiagen) and resuspended in 400 µl methanol. For the lipid extraction, 200 µl of the methanol suspension were used for the Bligh Dyer protocol followed by a 0.9 % KCl wash (*184*). The organic phase was separated and dried to completion

under nitrogen and the pellet was resuspended in 400 μl chloroform. TAG was separated over 8 cm with 80:20:1 petroleum ether: diethyl ether: acetic acid, and total polar lipids were collected by scraping the origin. An internal standard of 5 μg of pentadecanoic acid was added to each tube containing scraped bands and fatty acid methyl esters (FAMEs) were prepared with 1M methanolic HCl at 80°C for 30 minutes. FAMES were phase separated with hexane and 0.9% NaCl, dried to completion under nitrogen and resuspended in 50-75 μl of hexane. Gas chromatography and flame ionization detection (Agilent) were used to separate and quantify the fatty acid profile (*11*).

Carbohydrate analysis

Cells were grown and harvested as described for the lipid analysis. Frozen ground pellets were resuspended in 400 μl methanol. 20 μl of this methanol suspension, which was the same one used for lipid analysis, and 10 μg of *myo*-inositol internal standard were dried under vacuum before the addition of 150 μl 3.3N trifluoroacetic acid (TFA). TFA hydrolysis and the preparation of alditol acetates were performed as described previously (*186*). Sugar alditol acetates were quantified gas chromatography-mass spectrometry (Agilent) against the *myo*-inositol internal standard and a 0.625-20 μg standard curve using target ion counts (*186*).

Analysis of gene expression by RNA-Seq

For each biological replicate we used a 50 ml culture grown in a 250 ml flask (~20-30 10^6 cells ml^{-1}). Cells were collected every 3 h and two biological replicates were harvested for each time point. Cells were collected at 4,000 g for 10 min at 4°C, the cell pellet was resuspended in 1.5 ml f/2 media and transferred to 2 ml tubes. The cells were centrifuged again at 18,000 g for 10 min

at 4°C and flash frozen in liquid nitrogen. Samples that were in the dark at the time of collection were placed in black or amber tubes to prevent light exposure. RNA was extracted from ground frozen pellets using the Omega eZNA Plant RNA kit (Omega Biotek). RNA quality was checked with the Bioanalyzer (Agilent). All samples had a RIN of 6.9-7.9. Samples were sequenced at the MSU-Research Technology Support Facility using an Illumina HiSeq 2500 and a single-end 50 nucleotide run. Eight samples were sequenced in each lane using a custom bar-coding, but the two biological replicates from the same time point were run in separate lanes. The average number of RNA-seq reads per sample was 23,434,157 and they ranged between 16,783,501 and 30,197,823. The reads of each of the 16 samples (8 time points, 2 samples each time point) were mapped to the *N. oceanica* CCMP1779 V1 (*11*) genome using Tophat (*187*) with default parameters except for intron length (min 13, max 8712) and max-multi-hits (1). Expression level estimates in Fragments Per Kilobase of exon per Million fragments mapped (FPKM) was calculated using Cufflinks (*188*) with parameter –I 8712.

Accession Numbers

Nannochloropsis oceanica CCMP1779 genome data and gene IDs are from the Nannochloropsis oceanica CCMP1779_V1 (*11*) (https://bmb.natsci.msu.edu/about/directory/faculty/christoph-benning/nannochloropsis-oceanica-ccmp1779/). For clarity, in the main text and figures, gene IDs in the format NannoCCMP1779|xxxxx, have been abbreviated to the digits xxxxx. Raw read data have been deposited in NCBI's Gene Expression Omnibus (*189*) and are accessible through GEO Series accession number GSE69460 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE69460)

Analysis of cyclic gene expression

The determination of cycling genes was performed using COSPOT (*154*) and the discrete Fourier transform (DFT) as described previously (*144*). In this parallel study of *Chlamydomonas reinhardtii*, we had compared the performance of these methods, both separately and in conjunction, using a gold standard set of cyclic genes (*144*). Since no such set was available for *N. oceanica* CCMP1779 and the relationship between the COSPOT p-value and "cyclic score" derived from the DFT appears to be the same (Figure 2.22), we applied the same cutoff thresholds that we used in *C. reinhardtii*. A gene in *N. oceanica* CCMP1779 was called "cyclic" if its expression vector had a COPSOT p-value < 0.02 or a cyclic score > 0.800, which is equivalent to the p-value of 0.02 in a population of randomized expression vectors derived from the *N. oceanica* FPKM data set.

Neither COSPOT nor DFT have high coverage of oscillatory expression profiles that peak only at a single time point (i.e. narrow peak of expression). To detect these patterns, we used EdgeR (*190*) to identify cases where read counts were non-uniformly distributed across time samples. A gene was called sharply expressed if the read counts were significantly over-expressed (p-value < 0.05) at one time point in the light-dark cycle relative to every other time point.

Analysis of gene expression by RT-qPCR

RNA was collected as described above and used to produce cDNA using the iScript cDNA Synthesis Kit (Bio-Rad). Primers with annealing temperatures of 55ºC were designed for RT-qPCR and checked for efficiency using serial dilutions of cDNA template and for specificity by melting curve analysis (Table 2.2). We used an elongation factor encoding gene 10181 as

reference, since its expression was constitutive in our experiment (Data 2.1) and had been previously suggested to be a constitutively expressed gene in *Nannochloropsis* species (Cao *et al.,* 2012). Quantitative PCR was performed using Sybr Green Master Mix (Life Technologiess) and a Mastercyler ep *realplex* (*Eppendorf).* Relative gene expression was obtained by the ΔCT method.

Flow cytometry analysis

Cells grown under diel conditions were fixed over night in 70% ethanol at 4ºC, washed twice in phosphate-buffered saline pH 7.2 (PBS) and stained in PBS/0.1% Triton/1µg ml^{-1} 4', 6'-diamidino-2-phenylindole (DAPI). Flow cytometry analysis was conducted on an Influx sorter (BD Biosciences, USA) using a 355nm UV laser and a 460/50 band pass filter, and 10,000 cells were analyzed per sample. The percentage of cells in 1C (unreplicated haploid genome), S1 (first round of DNA synthesis), 2C (replicated haploid genome), S2 (second round of DNA synthesis) and 4C (cells with four haploid genomes) where estimated using WinList (Verity Software House).

Microscopic analysis

Cells grown under diel conditions were harvested at ZT 18 in the dark. Cells were fixed in a mixture of 4% (w/v) paraformaldehyde and 0.25% (v/v) glutaraldehyde prepared in PBS overnight at 4 °C. DNA was stained with DAPI. The samples were analyzed with a Olympus FluoView 1000 confocal laser scanning microscope with excitation at 405 nm and emission 440 nm. The three-dimensional optical sections were acquired at 0.45 µm step intervals. The final

images represent projection of an image stack. For image processing and analysis we used the Olympus FluoView 1000 software.

Gene functional annotation

Candidate genes were first identified by BLASTp searches using functionally annotated *Phaeodactylum tricornutum* (*39*), *Ectocarpus siliculosus* (*191*), *N. gaditana* (*12*) or *N. oceanica* IMET1 (*18*) sequences and by PFAM domain analysis using HMMER3 (*180*), and checked by reciprocal BLAST against NCBI. For the phylogenetic analyses sequences were aligned using MUSCLE (*192*) and trees constructed using the maximum likelihood method implemented in MEGA 5 (*193*), with bootstrap replicates and gap removal. For each phylogenetic analysis the best combination of evolutionary model and rates was found using MEGA 5. Subcellular localization was predicted using HECTAR (*36*). In addition, subcellular analysis for a few genes was also performed using PredPlantPTS1 (*194*) and MitoProtII (*195*) and ASAFind (*196*). For ASAFind, we used SignalP 4.1 with default settings.

Protein analysis

N. oceanica CCMP1779 liquid cultures were grown in f/2 medium to mid-log phase under light dark cycles. 5.0×10^7 cells were transferred to a 1.7 ml tube and centrifuged at 13,000 g for 1 minute. Ground cell pellets were resuspended in 1 ml of Laemmli buffer, mixed by vortexing and left at 22°C for 10 min. 20 µl (1 million cells) were loaded on a BioRad Mini-PROTEAN TGX 10 well pre-cast 4%-20% gel (#456-1094). The LDSP protein was detected using an anti-LDSP antibody (*176*) used in a 1:10,000 dilution. Detection occurred using a BioRad goat anti-Rabbit IgG HRP (170-6515, 1:20,000) and Thermo Fisher SuperSignal West Femto (#34095).

Blots were imaged on a BioRad ChemiDoc MP system. Direct-Blue 71 (*197*) staining was used as indication of total protein content on the blots.

APPENDICES
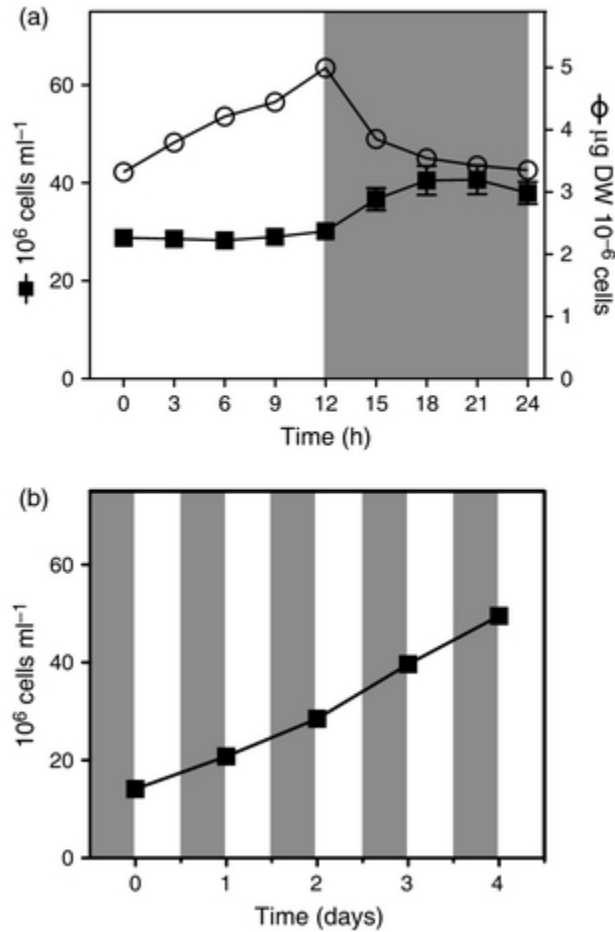
Figure 2.1. *N. oceanica* **CCMP1779 cell growth under diel cycles.**
Cells were grown in 12 h light:12 h dark diel cycle. Shaded areas indicate dark periods. (a) Cell number (filled squares) and cell dry weight (DW, open circles) were quantified at different times of the day. Values are the average of 3-6 independent cultures from three experiments. (b) Cell growth during four days under diel conditions. Values are the average of three independent cultures. Error bars represent SEM (some are smaller than the symbol).
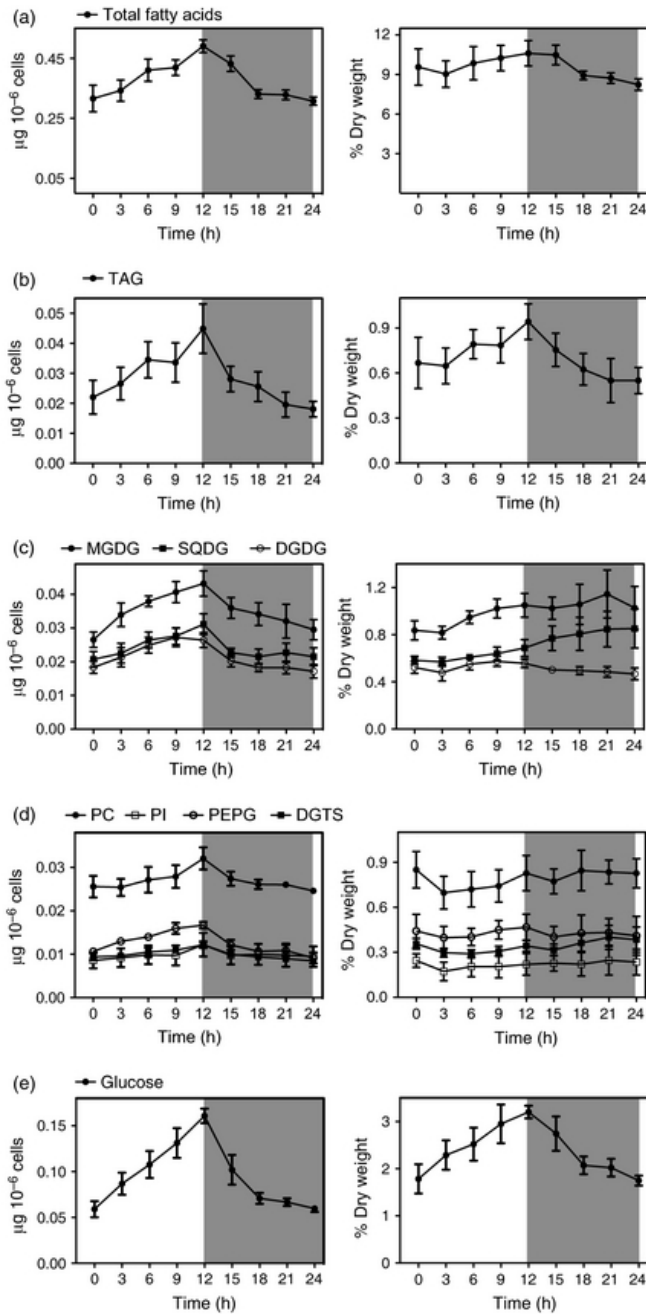
Figure 2.2. Metabolite content under diel cycles.
Cells were grown in 12 h light:12 h dark cycles. Lipids were quantified as esterified fatty acids associated with each lipid class. Shaded areas indicate dark periods. Values are the average of 3-6 different cultures from three independent experiments. Error bars represent SEM.
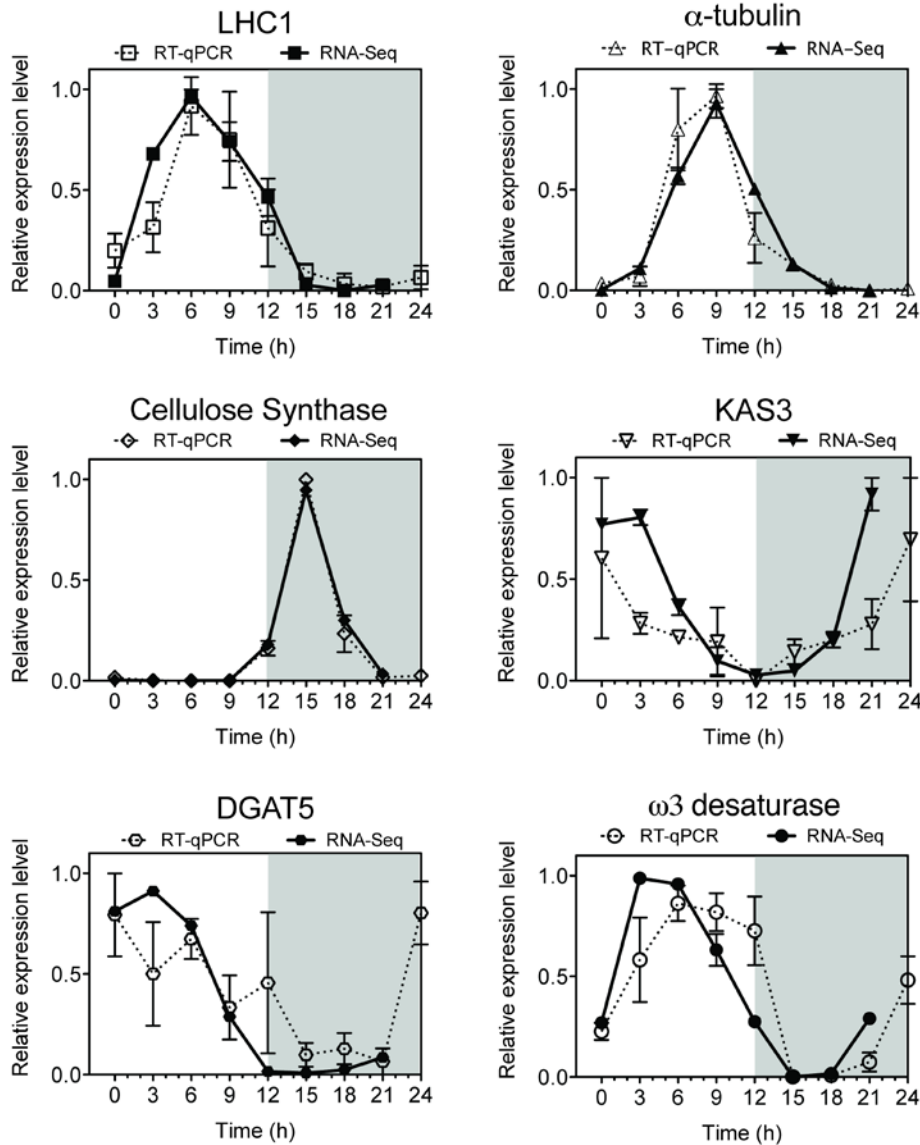
Figure 2.3. Confirmation of RNA-Seq measurements by reverse transcription quantitative PCR.
Cells were grown in light:dark cycles and collected at the indicated times. RNA levels determined by RT-qPCR were normalized against the elongation factor (10181) reference gene. RNA-seq and RT-qPCR expression values were then normalized between 0 and 1 and represent the mean of 2 (RNA-Seq) or 2-3 (RT-qPCR) biological replicates plotted with SEM or range. *LHC1, LIGHT HARVESTING COMPLEX I* (8367); *α-TUBULIN* (4716); *CS, CELLULOSE SYNTHASE* (5780); *KAS3, β-KETOACYL-ACP SYNTHASE* (2094); DGAT5, *DIACYLGLYCEROL ACYLTRANSFERASE 5* (3915); *ω3, ω3 DESATURASE* (6416). Shaded areas represent dark periods.
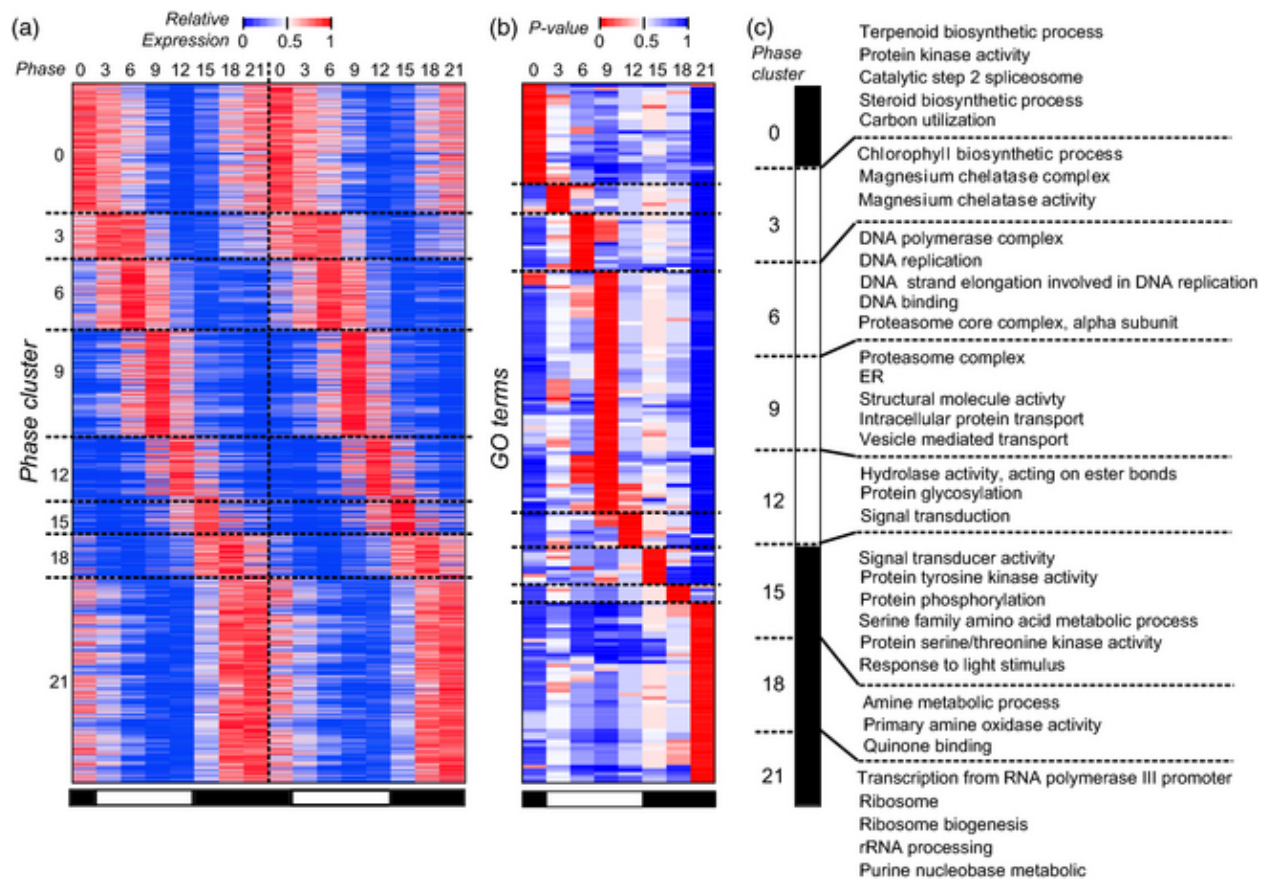
Figure 2.4. Gene expression phase and GO terms enriched in cyclically expressed genes. (a) The relative expression of each cycling gene (each row) across a 24-hour period (columns). The two biological replicates were plotted as consecutive days. Genes were assigned to phase clusters based on the predicted peak expression. Genes in each phase cluster were ordered using hierarchical clustering. The white and black bars below indicate samples from the light and the dark periods, respectively. FPKM values were normalized between 0 and 1. (b) The Fisher Exact Test p-value (Benjamini-Hochberg adjusted) of GO term (row) enrichment for genes cycling in each phase cluster (column). GO terms were ordered along the y-axis according to their first enriched phase and clustered within each enriched phase. (c) The most enriched GO terms in each phase with a p-value <0.005.
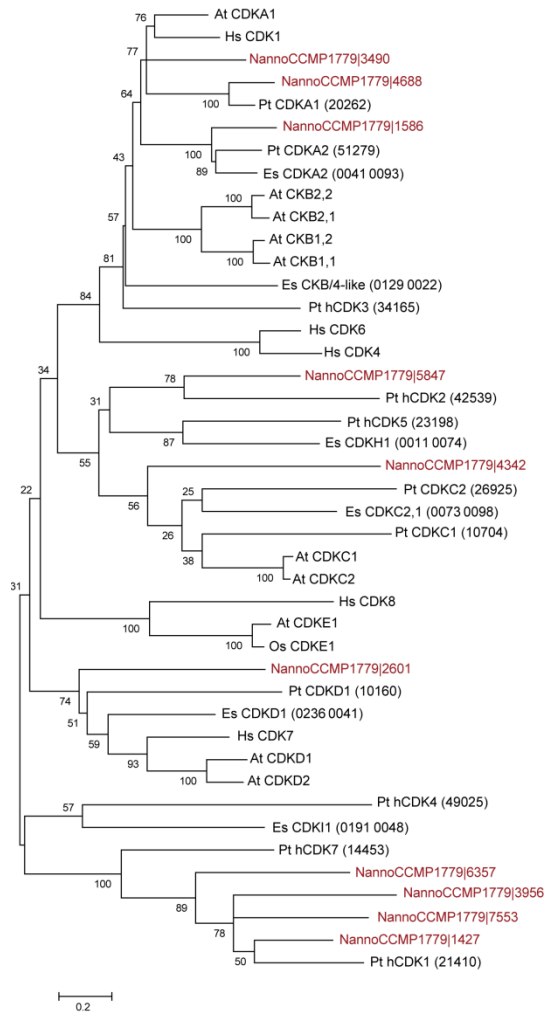
Figure 2.5. Phylogenetic analysis of *N. oceanica* CCMP1779 CDK-related proteins.
The phylogeny was inferred using the Maximum Likelihood method based on the General
Reverse Transcriptase + Freq. model (*198*). Branch bootstrap values (500 replicates) are
indicated. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining
method to a matrix of pairwise distances estimated using a JTT model (*199*). The branch lengths
are proportional to the number of substitutions per site. All positions with less than 95 % site
coverage were eliminated. Evolutionary analyses were conducted in MEGA5 (*193*). The ID
numbers for the diatom *P. tricornutum* (Pt) proteins are according to the annotated genomes at
JGI-release 2 as described in Huysman et al., (2010) (*51*). The names and ID numbers for the *E.
siliculosus* (Es) proteins are from Bothwell et al. (2010) (*200*). *Arabidopsis thaliana* (At), *Oryza
sativa* (Os) and *Homo sapiens* (Hs) sequences were retrieved from UniProt.
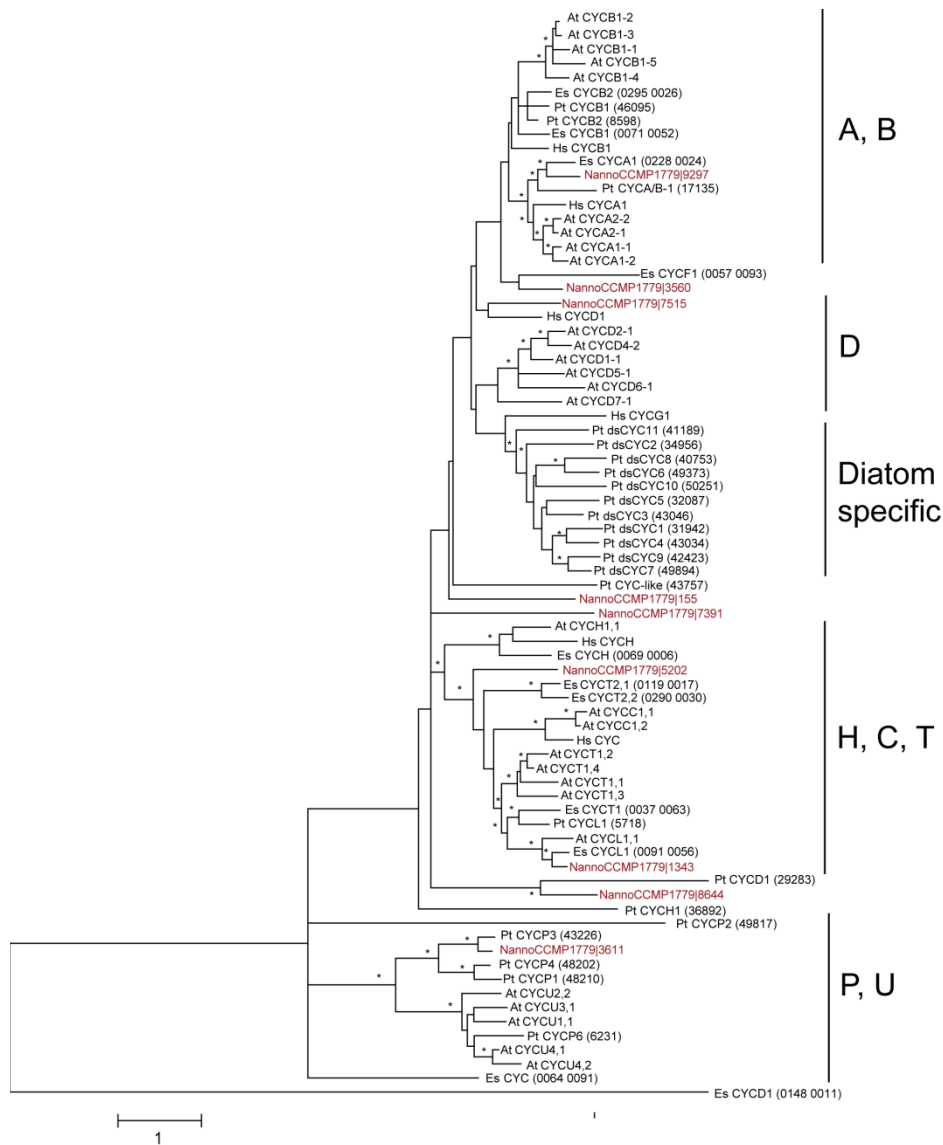
Figure 2.6. Phylogenetic analysis of *N. oceanica* CCMP1779 cyclin related proteins.
The phylogeny was inferred by using the Maximum Likelihood method based on the Whelan
And Goldman model (*201*). Initial tree(s) were generated and alignment positions filtered as in
Figure 2.5. Branches with bootstrap support (500 replicates) values higher than 50% are
indicated by (*). The branch lengths are proportional to the number of substitutions per site. All
positions with less than 95% site coverage were eliminated. Evolutionary analyses were
conducted in MEGA5(*193*). The names and ID numbers were obtained from the same sources as
detailed in Figure 2.5.
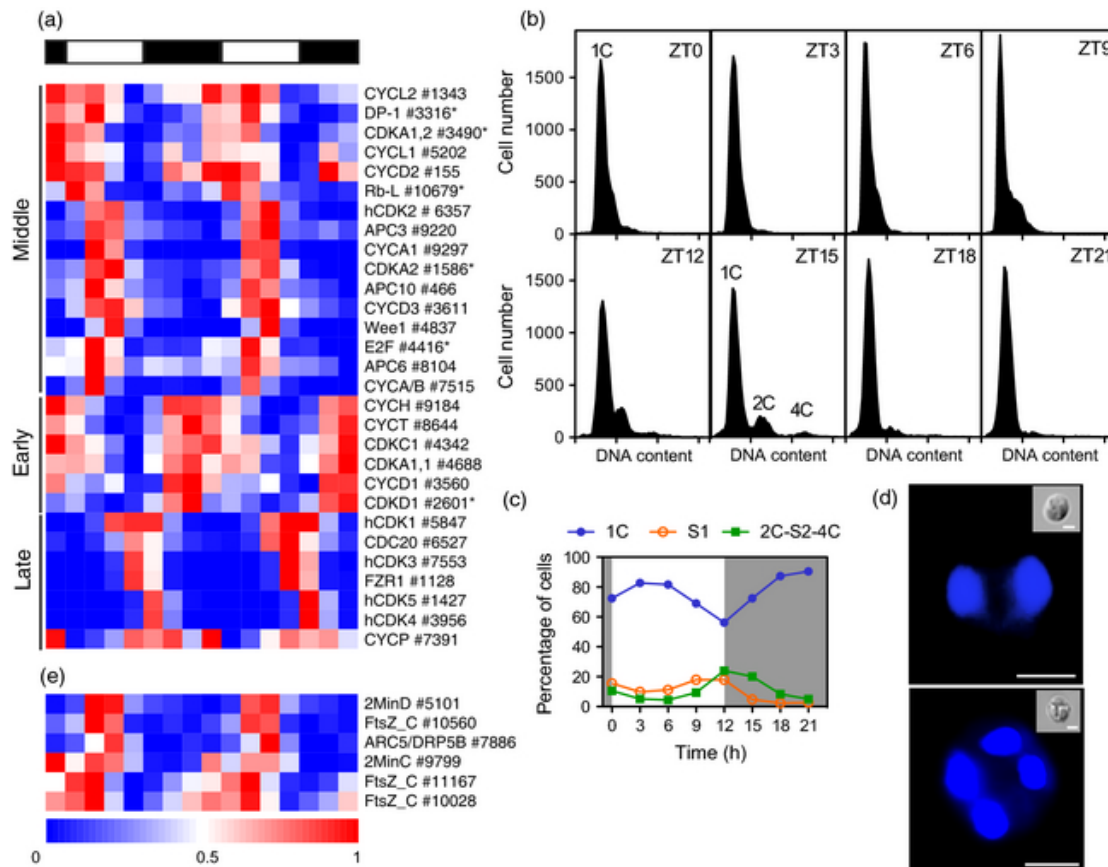
Figure 2.7. Cell cycle progression in light:dark cycles.
(a) Heatmap displaying relative gene expression levels of genes involved in cell division in *N. oceanica* CCMP1779. Early, middle and late, relate to phases in the cell cycle. (b) DNA content per cell measured as DAPI fluorescence using flow cytometry at different times of day (ZT, hours after lights on). (c) Estimation of the percentage of cells in 1C (unreplicated haploid genome), S1 (first round of DNA synthesis) and 2C (replicated haploid genome)- S2 (second round of DNA synthesis)- 4C (cells with four haploid genomes). (d) Images of DAPI stained *N. oceanica* CCMP1779 cells harvested at ZT18, containing two or four nuclei. Three-dimensional optical sections were acquired at 0.45 μm step intervals. The final images represent projection of an image stack. The inserts are the corresponding bright field images. Bars=1.5 μm. (e) Organelle division genes. For (a) and (e) FPKM values were normalized between 0 and 1 and used for hierarchical cluster analysis implemented in R. The two biological replicates were plotted as consecutive days to visualize variability. Numbers indicate the digits of the gene IDs. Descriptions and expression values can be found in Data 2.2. (*) Proteins discussed in the text.
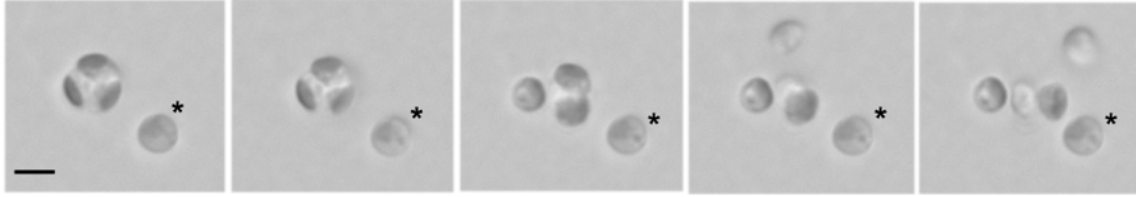
Figure 2.8. Time lapse images of a *N. oceanica* CCMP1779 cell undergoing division into four daughter cells.
Bright field microscopy using a Leica DMRA2 microscope and a 100x objective. Scale bar = 3 μm. * Non-dividing cell.
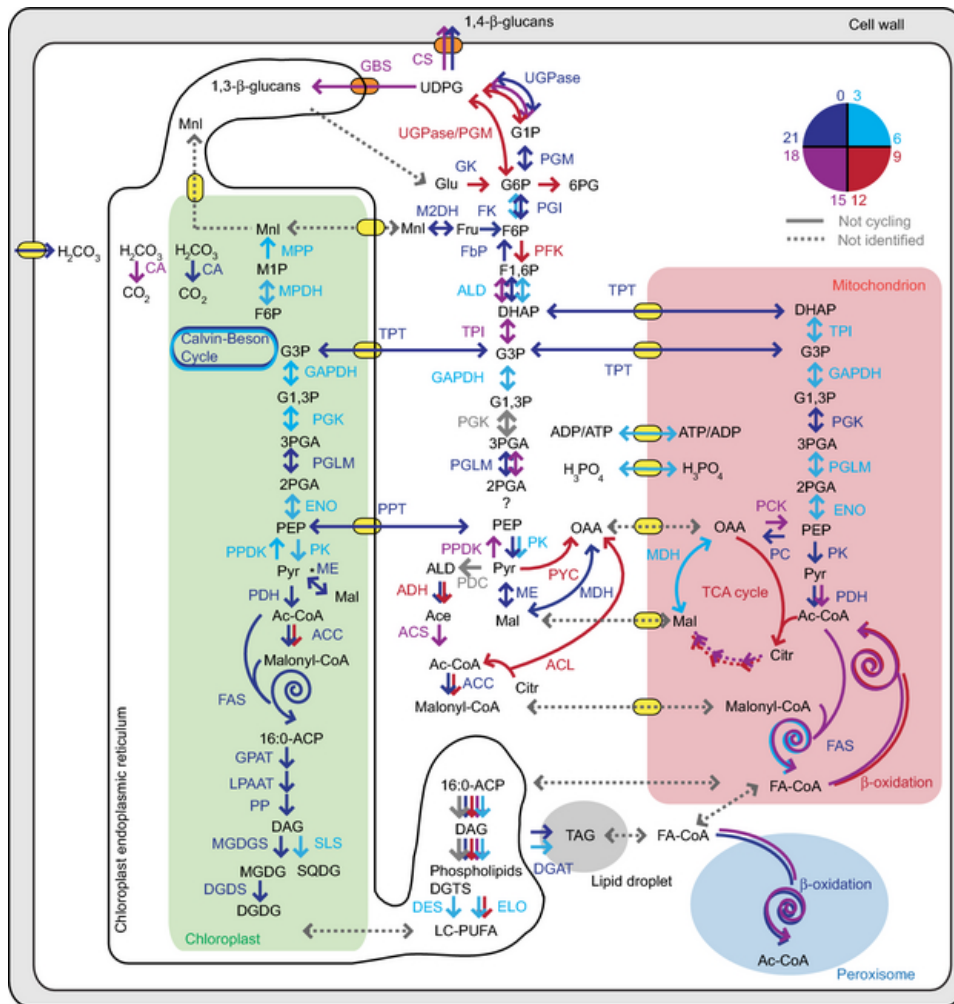
Figure 2.9. Transcription regulation of genes involved in central carbon metabolism in *N. oceanica* CCMP1779 under diel conditions.

Arrow and enzyme name colors indicate phase of expression as determined visually. Parallel arrows with different colors indicate multiple genes expressed at different times of day encoding for one enzyme. Expression data can be found in Data 2.2. Subcellular localization was based on HECTAR analysis. In addition, subcellular analysis for a few proteins was also performed using PredPlantPTS1 and MitoProtII. Enzyme abbreviations: ACC, acetyl co-A carboxylase; ACL, ATP citrate lyase; ACS, acetyl-coA synthetase; ADH, aldehyde dehydrogenase; ALD, aldolase; CA, carbonic anhydrase; CS, cellulose synthase; DES, desaturase; DGDS, digalactosyl-1,2,sn-diacyl-glycerol synthase; ELO, elongase; ENO, enolase; FAS, fatty acid synthase; FbP, fructose 1,6-bisphosphase; FK, fructokinase; GAPDH, glyceraldehyde 3-phosphase dehydrogenase; GBS, β-1,3-glucan synthase ; GPAT, glycerol-3-phosphate acyltransferase; LPAAT, lysophosphatidic acid acyltransferase; M2HD, mannitol 2-dehydrogenase; MDH, malate dehydrogenase; ME, malic enzyme; MGDGS, monogalactosyl-1,2sn-diacylglycerol synthase; MPDH, mannitol-1-phosphate dehydrogenase; MPP, manitol-1-phosphatase; PC, phosphoenolpyruvate carboxylase; PCK, pyruvate carboxykinase; PDC, pyruvate decarboxylase; PDH, pyruvate dehydrogenase; PFK, phosphofructokinase; PGI, phosphoglucoisomerase; PGK, phosphoglycerate kinase;

Figure 2.9 (cont'd)

PGLM, phosphoglycerase mutase; PGLM, phosphoglycerate mutase; PGM, phosphoglucomutase; PK, pyruvate kinase; PP, phosphatidate phosphatase; PPDK, pyruvate phosphate dikinase; PYC, pyruvate carboxylase; SLS, sulfolipid synthase; TPI, triose-phosphase isomerase; TPT, triosephosphate transporter; UGPase, UDP-glucose pyrophosphylase. Metabolite abbreviations: 16:0-ACP; 16:0 fatty acid; 2PGA, 2-phosphoglycerate; 3PGA, 3-phosphoglycerate; 6PG, 6-phosphogluconate; Ac-CoA, acetyl Co-A; Ace, acetate; Citr, citrate; DAG, diacylglycerol phosphate; DGDG, digalactosyldiacylglycerol; DGTS, diacylglyceryltrimethylhomoserine; DHAP, dihydroxyacetone phosphate; F1,6P, fructose 1,6-bisphosphate; F6P, fructose 6-phosphate; FA, fatty acid; Fru, fructose; G1,3-P, glycerase 1,3-bisphosphate; G1P, glucose 1-phosphate; G3P, glycerate 3-phosphate; G6P, glucose 6-phosphate; Glu, glucose; LC-PUFA, long chain polyunsaturated fatty acids; M1P, mannitol 1-phosphate; Mal, malate; MGDG, monogalactosyldiacylglycerol; Mnl, mannitol; OAA, oxaloacetate; PEP, phosphoenolpyruvate; Pyr, pyruvate; SQDG, sulfoquinovsyldiacylglycerol; TAG, triacylglycerol; UDPG, UDP-glucose.
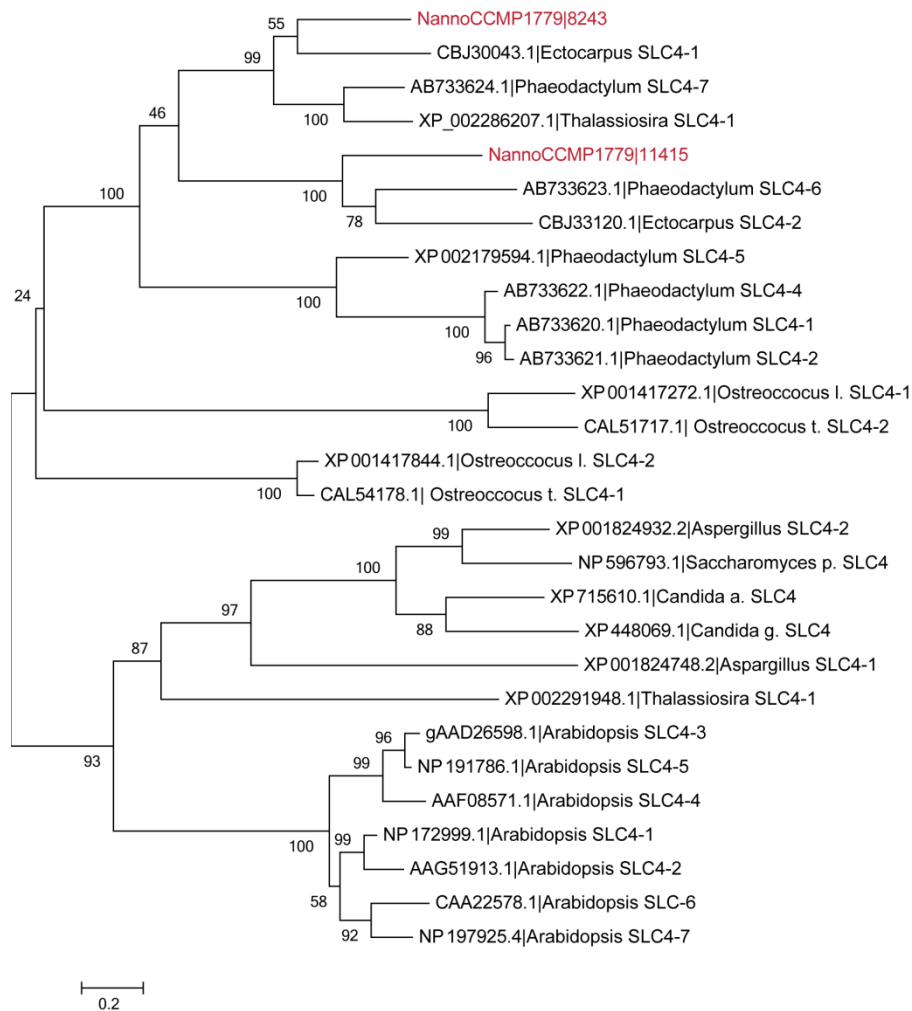
Figure 2.10. Phylogenetic analysis of *N. oceanica* CCMP1779 SLC4 related proteins.
The phylogeny was inferred as in Figure 2.5. Branch bootstrap values (500 replicates) are
indicated. Initial tree(s) for the heuristic search were obtained automatically by applying
Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT
model, and then selecting the topology with superior log likelihood value. A discrete Gamma
distribution was used to model evolutionary rate differences among sites. The rate variation
model allowed for some sites to be evolutionarily invariable. The branch lengths are proportional
to the number of substitutions per site. All positions with less than 95% site coverage were
eliminated. Evolutionary analyses were conducted in MEGA5 (*193*). Identifiers are from
Genbank with the exception of *N. oceanica* CCMP1779 sequences.
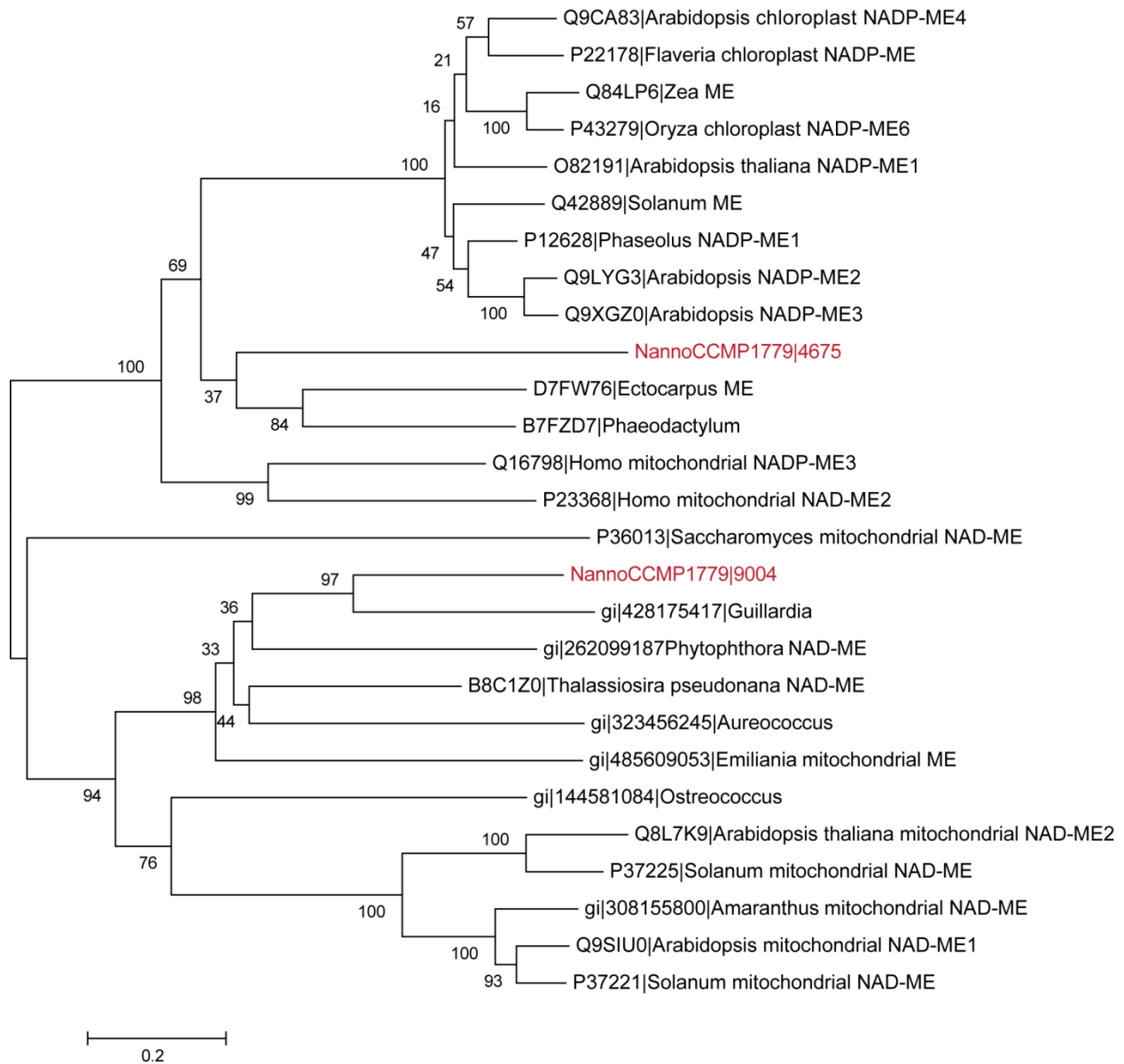
Figure 2.11. Phylogenetic analysis of *N. oceanica* CCMP1779 malic enzyme related proteins.

The phylogeny was inferred as in Figure 2.5. Branch bootstrap values (500 replicates) are indicated. The search for the initial trees and the modeling of evolutionary rate differences were performed as described in Figure 2.5. The branch lengths are proportional to the number of substitutions per site. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA5. Protein identifiers are from UniProt or NCBI gi sequence identifiers with the exception of *N. oceanica* CCMP1779 sequences.

70
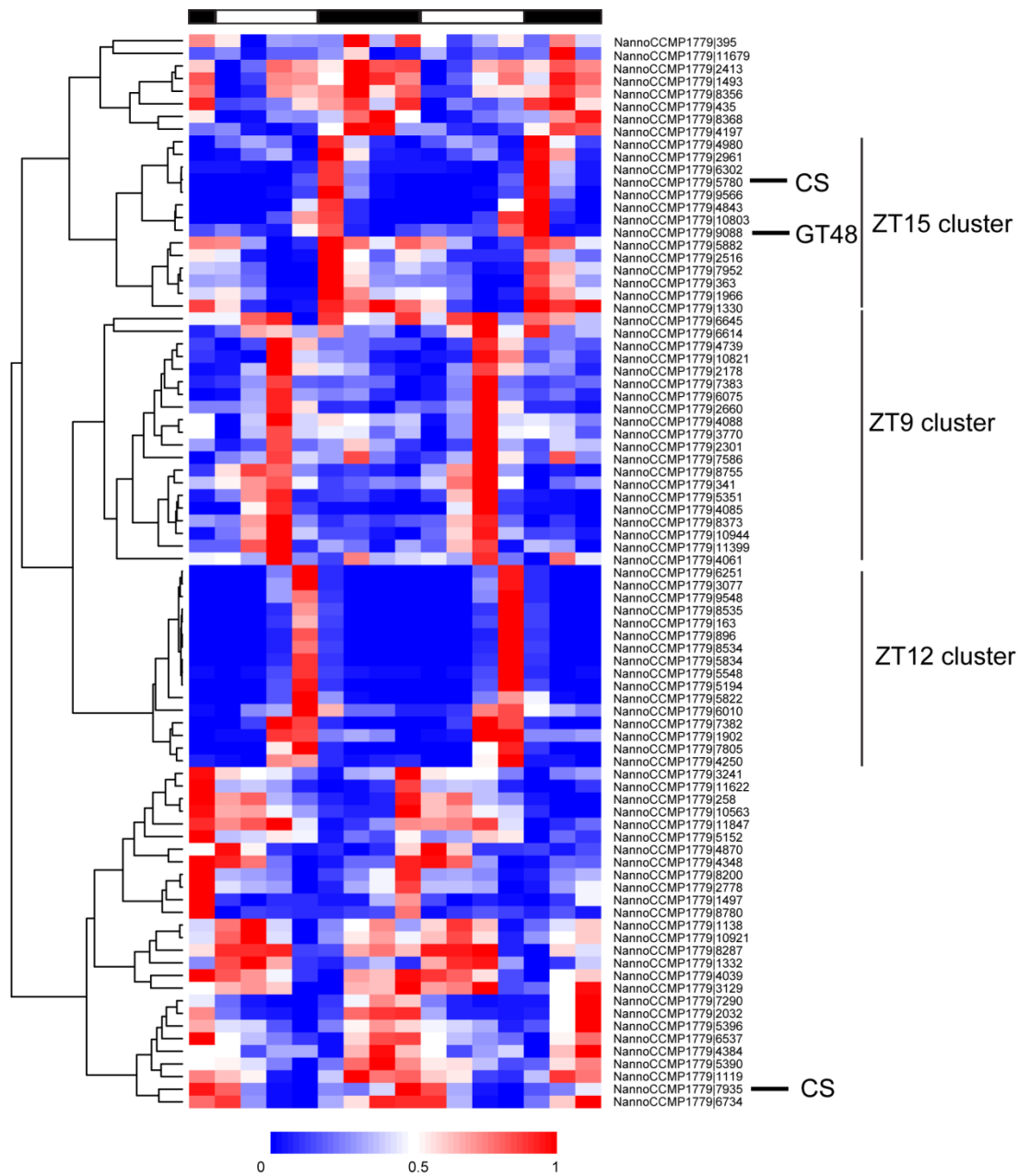
Figure 2.12. Heatmap displaying relative expression levels of putative glycosyl transferases and glycosyl hydrolases encoding genes.
Expression values were analyzed as described in Figure 2.7. Row descriptions indicate gene IDs. CS, cellulose synthase; GT48, glycosyl transferase 48 family gene.
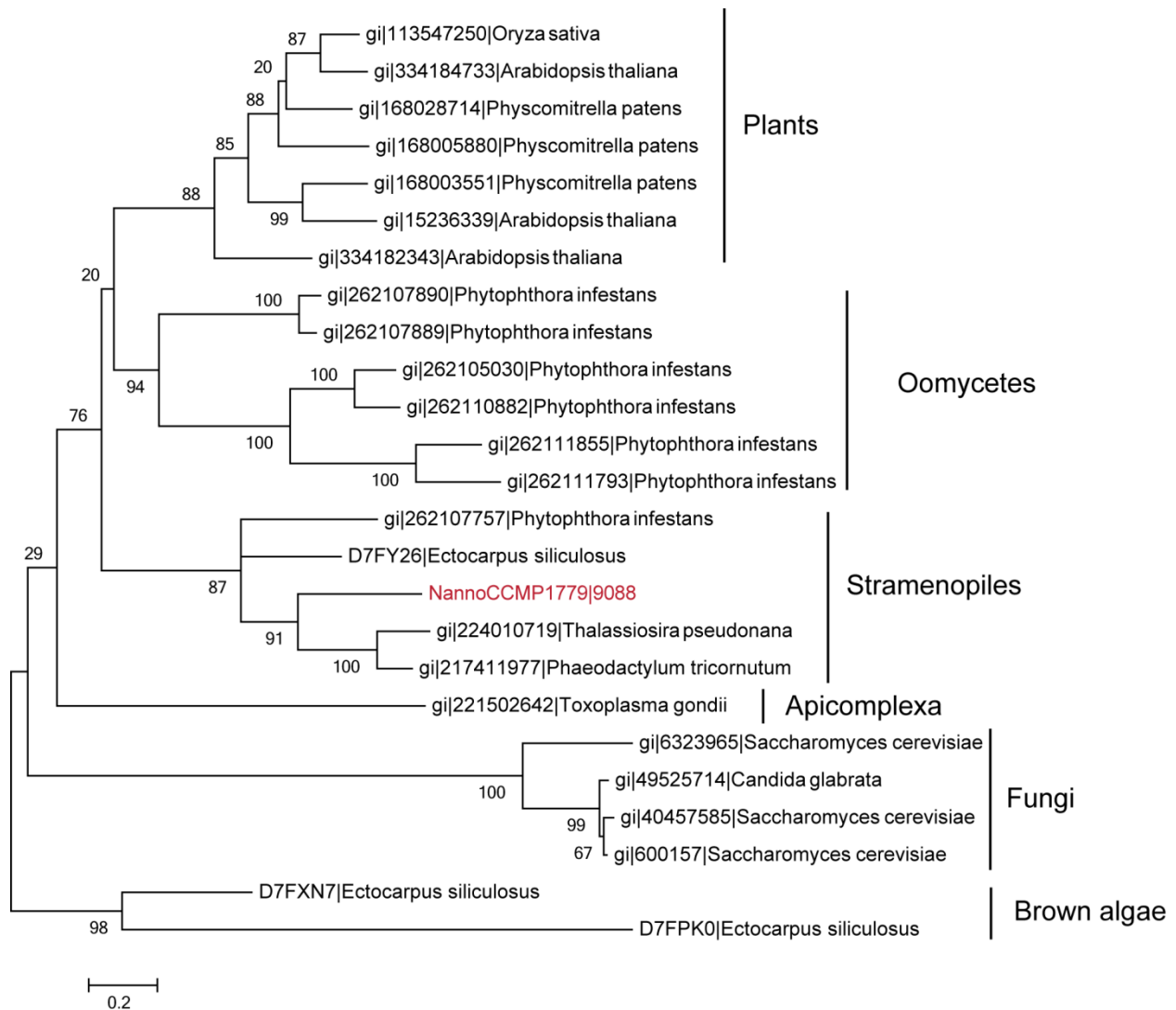
Figure 2.13. Phylogenetic analysis of the *N. oceanica* CCMP1779 β-1,3-glucan synthase related protein.
The phylogeny was inferred as described for Figure 2.6. Branch bootstrap values (500 replicates) are indicated. The search for the initial trees and the modeling of evolutionary rate differences were performed as described in Figure 2.6. The branch lengths are proportional to the number of substitutions per site. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA5 (*193*). Protein identifiers are NCBI gi sequence identifiers or UniProt accession numbers for the *Ectocarpus* sequences with the exception of *N. oceanica* CCMP1779 sequences.
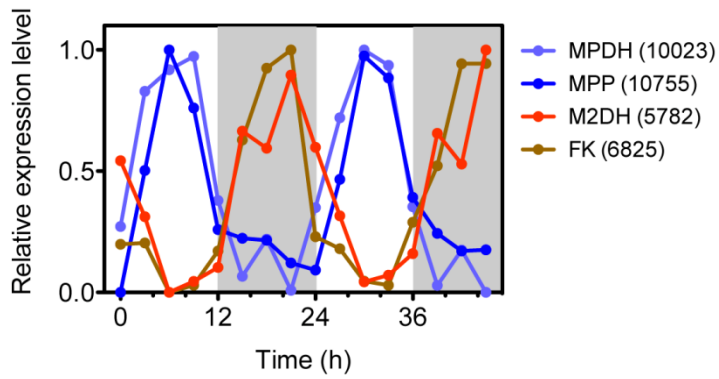
Figure 2.14. Expression of genes potentially involved in the mannitol cycle.
FPKM values were normalized between 0 and 1. MPDH, mannitol 1-phosphate dehydrogenase; MPP, mannitol 1-phosphate phosphatase; M2DH, mannitol 2-dehydrogenase; FK, fructokinase. Numbers indicate the digits of the gene IDs.
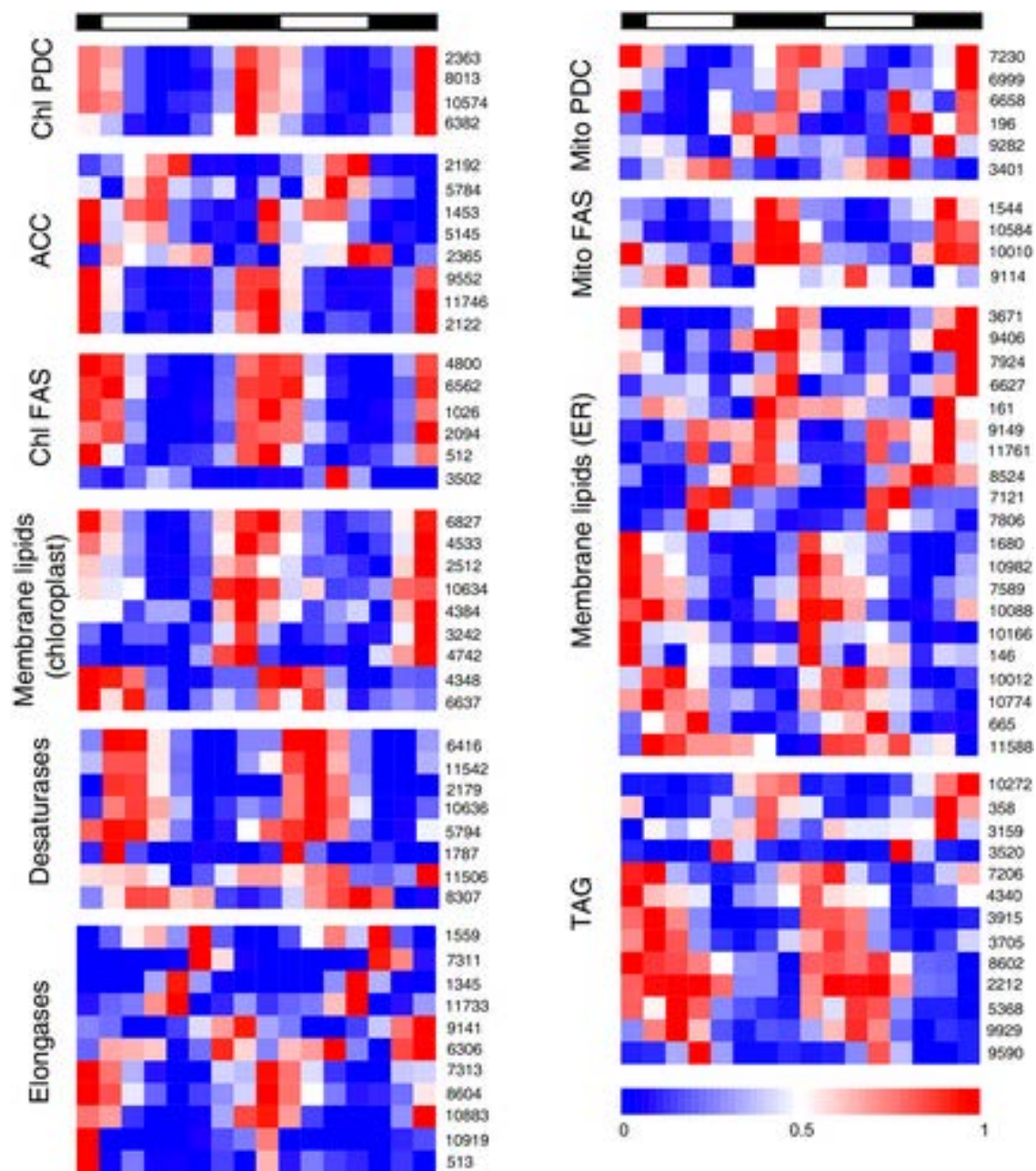
Figure 2.15. Gene expression of genes involved in lipid biosynthesis under light:dark cycles. Expression values were analyzed as described in Figure 2.7. Descriptions and expression values can be found in Data 2.2. Chl, chloroplast; PDC, pyruvate decarboxylase comples; FAS, fatty acid synthase; Mito, mitochondria; ACC, acetyl-CoA carboxylase components; ER, endoplasmatic reticulum; TAG, triacylglycerol. Numbers indicate digits in the gene IDs.
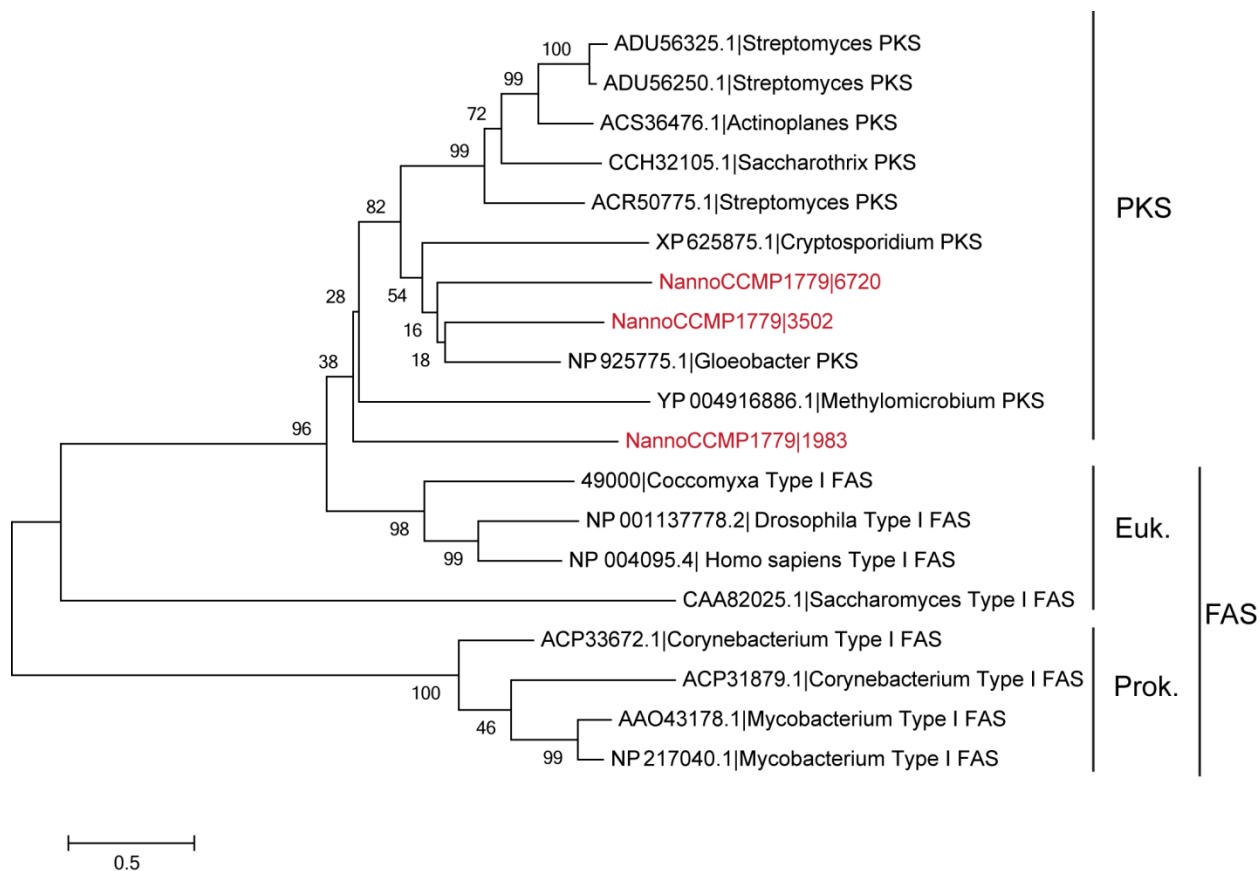
Figure 2.16. Phylogenetic analysis of the *N. oceanica* CCMP1779 type I FAS-like genes.
The phylogeny was inferred as in Figure 2.6. Branch bootstrap values (500 replicates) are
indicated. The search for the initial trees and the modeling of evolutionary rate differences were
performed as described in Figure 2.6. The tree is drawn to scale, with branch lengths measured in
the number of substitutions per site. All positions with less than 95% site coverage were
eliminated. Evolutionary analyses were conducted in MEGA5 (*193*). Identifiers are from
Genbank with the exception of *N. oceanica* CCMP1779 sequences, and the Coccomyxa
sequence (JGI Phytozome *Coccomyxa subellipsoidea* C-169 v2.0). PKS, polyketide synthases;
FAS, type I fatty acid synthases; Euk, eukaryotes; Prok, prokaryotes.
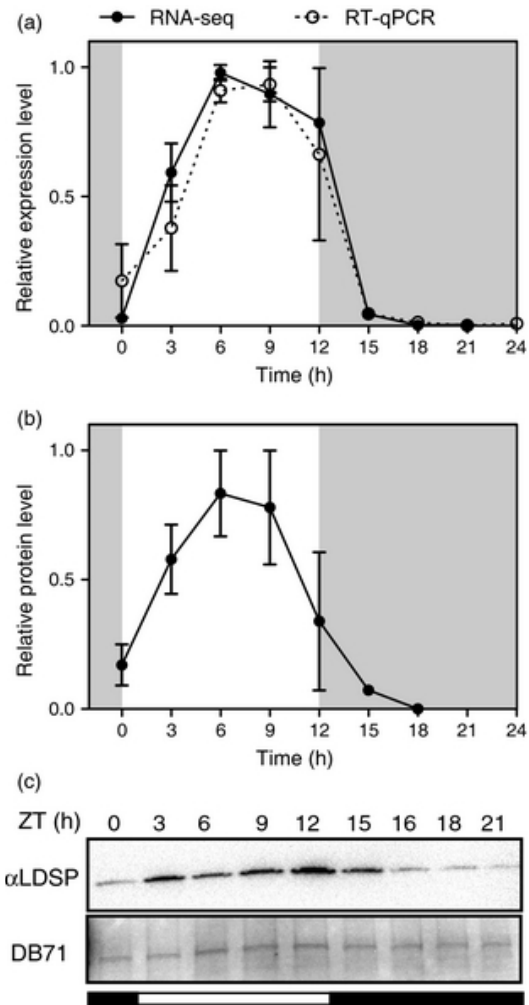
Figure 2.17. LDSP expression under light:dark cycles.
(a) LDSP RNA levels. RNA levels determined by RT-qPCR were normalized against the elongation factor (10181) reference gene. RNA-seq and RT-qPCR expression values were then normalized between 0 and 1 and represent the mean of 2 (RNA-Seq) or 2-3 (RT-qPCR) biological replicates plotted with SEM or range. (b) Quantification of LDSP protein content. Values are the average ± range of two biological replicates. (c) Immunodetection of LDSP protein levels under diel conditions using an anti-LDSP (αLDSP) antibody. Proteins extracted from one million cells were loaded on each well. Direct-Blue 71 (DB71) staining was used as a total protein stain. Shaded areas indicate dark periods.
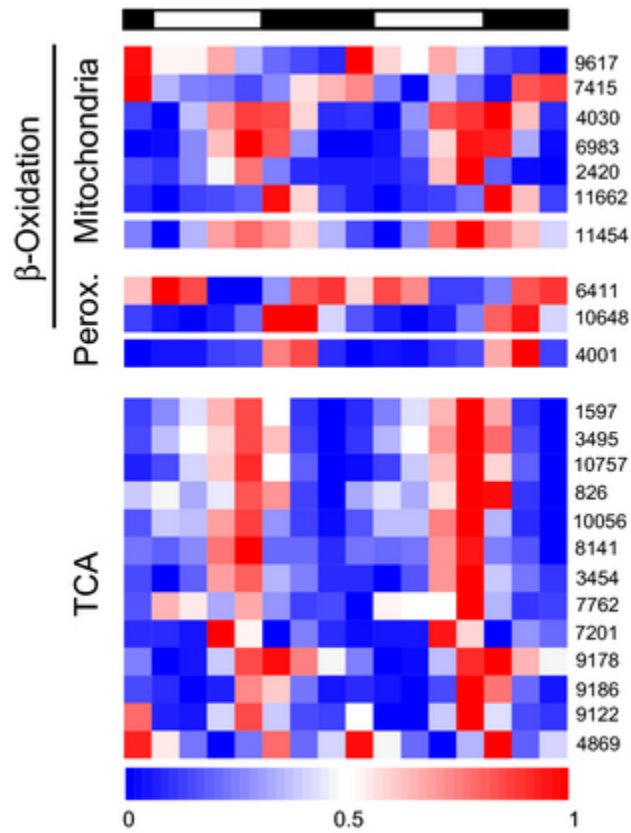
Figure 2.18. Gene expression of genes involved in fatty acid degradation under light:dark cycles.
Expression values were analyzed as described in Figure 2.7. Descriptions and expression values can be found in Data 2.2. Perox., peroxisome; TCA, tricarboxylic acid cycle. Numbers indicate digits of gene IDs.
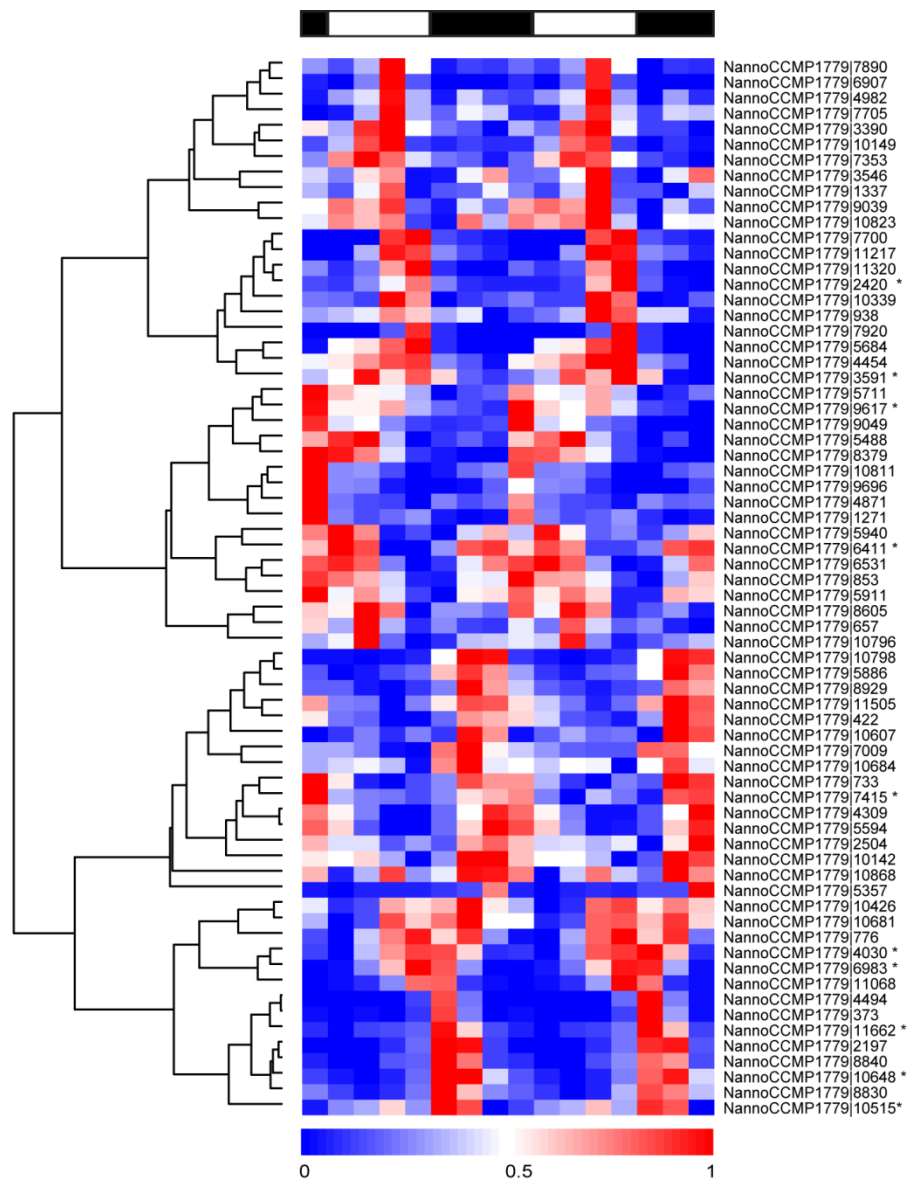
Figure 2.19. Heatmap displaying relative expression levels of genes involved in the lipid degradation.
Expression values were analyzed as described in Figure 2.7. Labels indicate gene IDs. * Indicate genes potentially involved in β-oxidation. Expression values can be found in Data 2.2.
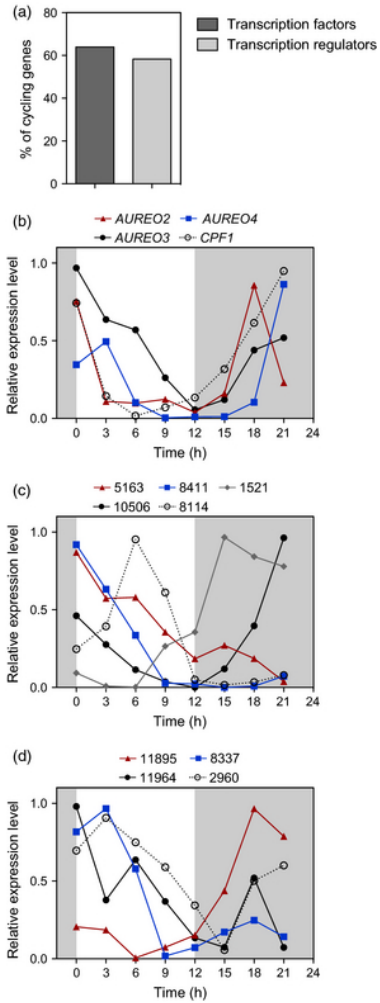
Figure 2.20. Cyclic expression of transcription regulators.
(a) Percent of cycling transcription regulators under light:dark cycles as determined by COSPOT and DFT. Relative expression level of blue light receptors (b), Sigma70-like factors (c) and mTERFs (d). FPKM values were normalized between 0 and 1. The values represent the average of two biological replicates. Shaded region represent dark periods. In (c) and (d) labels indicate the digits of the gene IDs.
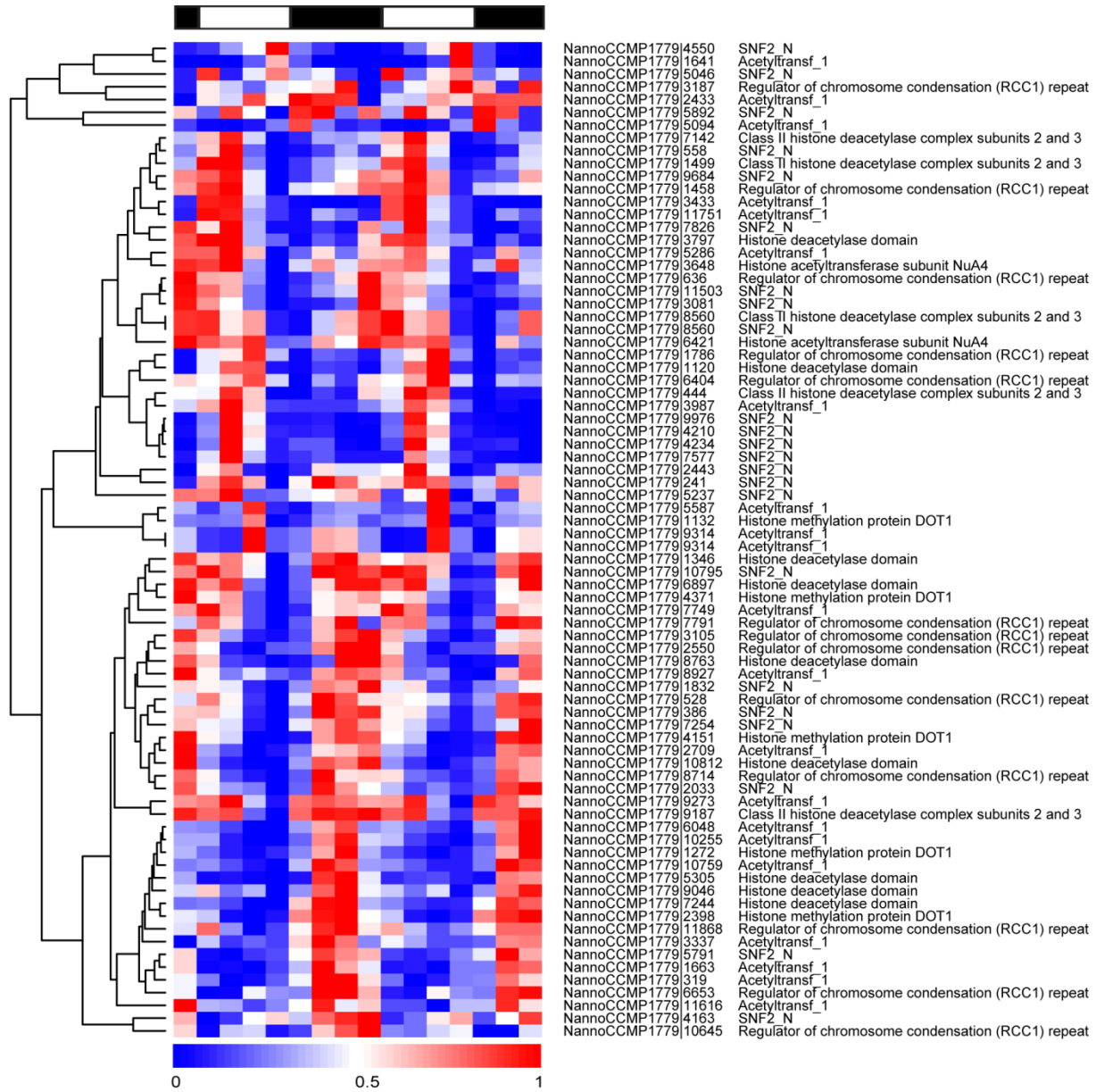
Figure 2.21. Heatmap displaying relative expression levels of genes involved in chromatin modification.
Expression values were analyzed as described in Figure 2.7. Labels indicate gene IDs, descriptions indicate conserved domains identified using HMMER (*180*).
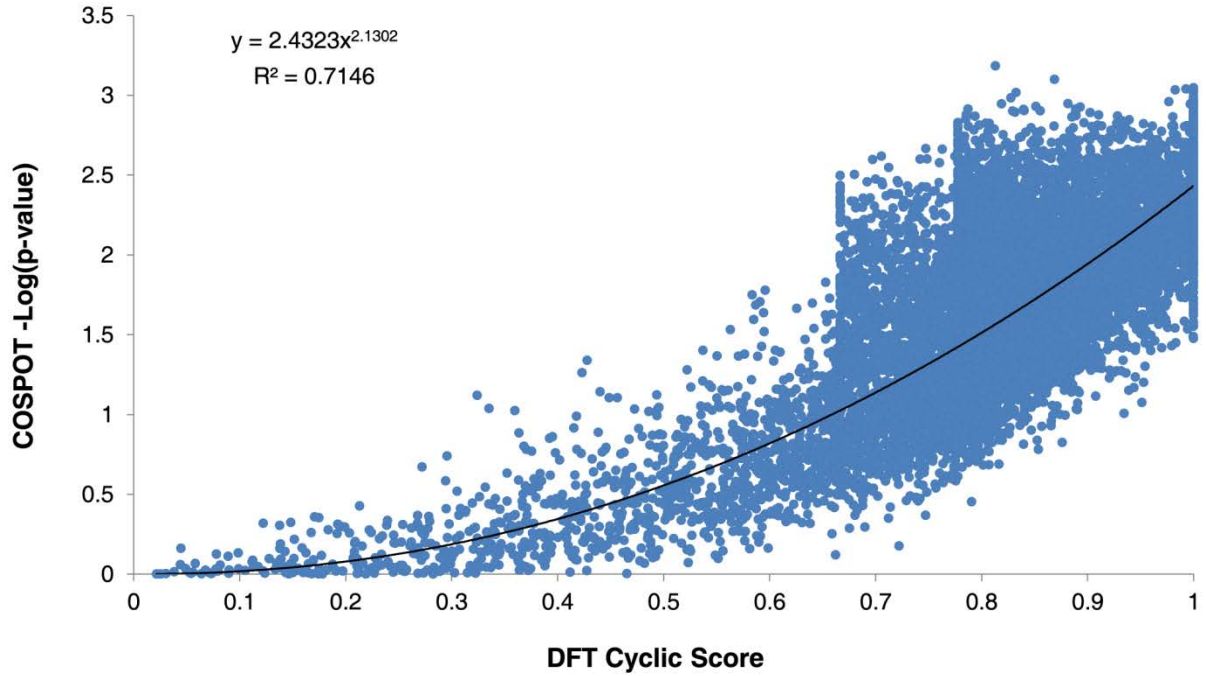
Figure 2.22. The relationship between the cyclic score derived from the DFT and negative log transformed p-value from COSPOT.
A power-law trendline is plotted against the data (equation in upper left hand corner). The results of the parallel analysis in *C. reinhardtii* show a similar power law relation (y = 2.4871x$^{2.1549}$) between COSPOT and the DFT with R^2 > 0.70.

Table 2.1. Prediction of cyclic gene expression.
The number of cyclic genes predicted by COSPOT, the DFT, and the combination of both methods, as well as the period mean and phase peaks for each set of predictions. A gene was called "cyclic" if its expression vector had a COPSOT p-value < 0.02 or a cyclic score > 0.800, which is equivalent to the p-value of 0.02 in a population of randomized expression vectors derived from the *N. oceanica* CCMP1779 FPKM data set.

Method	# of Genes	Period	Phase Peaks
COSPOT	5513	24.36 (+/ 1.07)	0, -11
DFT	6254	24.24 (+/ 1.20)	0, +9
Combined	7402	24.25 (+/ 1.20)	0, +9

Table 2.2. Primers used for RT-qPCR in Chapter 2.
The target gene is listed in the first column, primer name in the second column, and the primer sequence in the third column.

Gene ID	Primer name	Sequence
NannoCCMP1779|5780	CS qpcr F2+	GGCCTATAACTCGACCTTGC
	CS qpcr R2	CATGACAGGCACCGTCTTAC
NannoCCMP1779|4716	Tua qpcr F+	CATGGCCTGCTGCCTCATG
	Tua qpcr R	TTGATGCCGCACTTGAAGC
NannoCCMP1779|4188	LDSP qpcr F2+	ACAAGGCCCCTAAGGTAAAC
	LDSP qpcr R2	ATGCAATGAGCGCAAACGAC
NannoCCMP1779|10181	Ef qpcr F2+	TGGTGGAGTCCTTCACTGAG
	Ef qpcr R2	CTTGCCCTTGACATCGACTG
NannoCCMP1779|6416	W3 qpcr F2+	GCTCTCCTATGTGGATTGGG
	W3 qpcr R2	GTTGGTCTTCCTTCTCGGAC
NannoCCMP1779|8367	LHC1 qpcr F+	GAGGCGCTCAAGTACCAGAA
	LHC1 qpcr R	ACTGGCTCTTCCACAGAAGC
NannoCCMP1779|2094	KAS3 qpcr F+	GGAGACTTCGGATGAATGGA
	KAS3 qpcr R	CCACGCATGTAATCGTCATC
NannoCCMP1779|3915	DGAT5 qpcr F+	CAGCAAACTGTTCAATGTGG
	DGAT5 qpcr R	TAGTAAAGCTCCTCGACCTT

Appendix 2.2. Chapter 2 datasets.

Supplemental datasets are available online from:

http://onlinelibrary.wiley.com/doi/10.1111/tpj.12944/full

Data 2.1. Expression of *N. oceanica CCMP1779* under light:dark cycles and prediction of cyclic gene expression.

Data 2.2. Expression patterns and estimation of subcellular localization of manually annotated genes.

Chapter 3. 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

ABSTRACT

Nannochloropsis oceanica is an oleaginous microalga rich in $\omega 3$ long-chain polyunsaturated fatty acids (LC-PUFAs) content, in the form of eicosapentaenoic acid (EPA). We identified the enzymes involved in LC-PUFA biosynthesis in *N. oceanica* CCMP1779 and generated multigene expression vectors aiming at increasing LC-PUFA content *in vivo*. We isolated the cDNAs encoding four fatty acid desaturases (FAD) and determined their function by heterologous expression in *S. cerevisiae*. To increase the expression of multiple fatty acid desaturases in *N. oceanica* CCMP1779 we developed a genetic engineering toolkit that includes an endogenous bidirectional promoter and optimized peptide bond skipping 2A peptides. The toolkit also includes multiple epitopes for tagged fusion protein production and two antibiotic resistance genes. We applied this toolkit, towards building a gene stacking system for *N. oceanica* that consists of two vector series, pNOC-OX and pNOC-stacked. These tools for genetic engineering were employed to test the effects of the overproduction of one, two or three desaturase encoding cDNAs in *N. oceanica* CCMP1779 and prove the feasibility of gene stacking in this genetically tractable oleaginous microalga. All FAD overexpressing lines had considerable increases in the proportion of LC-PUFAs, with the overexpression of *Δ12* and *Δ5* FAD encoding sequences leading to an increase in the final $\omega 3$ product, EPA.

INTRODUCTION

Long-chain polyunsaturated fatty acids (LC-PUFAs) are hydrocarbon acyl chains (18-22 carbons) containing multiple *cis* double bonds. The double bond closest to the methyl (ω,

omega-) end of a fatty acid, usually 3 or 6 carbons, differentiates LC-PUFAs into $\omega3$ or $\omega6$ classes, respectively. Evidence suggests consuming LC-PUFAs in adequate amounts with a balanced ratio among LC-PUFAs is essential for human physical and mental health (*202*), while current diets are often low in $\omega3$ LC-PUFAs.

Currently seafood is the major source of $\omega3$ LC-PUFAs, particularly EPA (20:5; number of carbons:number of double bonds) and docosahexaenoic acid (DHA, 22:6). However overfishing, habitat destruction, and pollution have reduced wild fish stocks, while increasing human populations drive a record demand for LC-PUFA nutrients (*203, 204*). Though fish are rich in $\omega3$ fatty acids, the majority of LC-PUFAs acyl chains originate at the base of the food chain in marine microalgae (*205*) (*205, 206*). Aquaculture of microalgae is an emerging source of LC-PUFAs for fish farming or direct human consumption, and understanding the biosynthetic pathways and factors that influence LC-PUFA metabolism in algae will enable increased production from these organisms (*2*).

One promising source of LC-PUFAs are oleaginous microalgae of the *Nannochloropsis* genus, which have a high EPA content (*177, 207, 208*). In *Nannochloropsis* species, EPA is associated with membrane lipids, particularly monogalactosyl diacylglycerol (MDGD) and the betaine lipid,diacylglyceryl-*N,N,N*-trimethylhomoserine (DGTS) (*11, 54, 209, 210*). Despite the research interest and economic implications, the EPA biosynthetic pathway in *Nannochloropsis* species is only partially characterized. Gene annotation and ^{14}C-acetate pulse-chase labeling suggest that EPA biosynthesis occurs via the $\omega6$ desaturation and elongation pathway and is localizedin the endoplasmic reticulum, with the phospholipids, phosphatidylcholine (PC) and phosphatidylethanolamine (PE), as substrates for fatty acid desaturases (FADs) (*11, 91, 209*).

Genome annotation identified putative components of the EPA biosynthetic pathway including $\Delta 9$, $\Delta 12$, $\Delta 6$, $\Delta 5$ and $\omega 3$ FADs, and a $\Delta 6$ fatty acid elongase (FAE) in *N. oceanica* CCMP1779 (*11*). Only the *N. oceanica* $\Delta 6$ and $\Delta 12$ FADs have been isolated and expressed in *S. cerevisiae* to corroborate their predicted functions (*14, 91*).

Genetic engineering can increase the production of LC-PUFAs in plants (*211-213*), fungi (*124, 214, 215*), and microalgae (*102, 216-219*). For example, the overexpression of the $\Delta 12$ FAD encoding cDNA using a stress inducible promoter in *N. oceanica* caused an increase of the 18:2 fatty acid content of PC during normal growth and diversion of 18:2 into triacylglycerol (TAG) following nitrogen deprivation (*91*). Furthermore, $\omega 3$ LC-PUFA production has been introduced into oleaginous yeast and plants by multiple gene expression systems (*124, 220, 221*). The effective diversion of metabolism into LC-PUFA production is likely to require co-expression of multiple genes encoding enzymes that catalyze different desaturation and elongation steps (*2, 222, 223*).

Metabolic engineering of multigene pathways has been developed using a variety of strategies (*224, 225*). Commonly, each transgene is placed under the regulation of separate promoters and terminators, and single gene expression plasmids are independently introduced into the host or assembled into a larger multigene expression plasmid (*224*). However, other strategies exist to express multiple genes. For example, bidirectional promoters allow compact assembly of coregulated gene pairs and have been utilized in a number of transgenic strategies, including in *Nannochloropsis* species (*14, 24*).

An obstacle to compact multigene expression systems in eukaryotes is the requirement of each transgene to be encoded in individual mRNAs. However, introducing short viral 2A peptide coding sequences (*226, 227*), prevents the formation of a peptide bond during translation,

allowing multiple proteins to be encoded by a single mRNA molecule, which has increased the efficiency of multiple coding sequence expression in eukaryotic cells.Advantages of using 2A peptides include: their compact size of 20-60 amino acids (aa), high efficiency in most eukaryotes tested to date, low toxicity, and stoichiometric co-expression of linked proteins (*100, 226-229*). However, the efficiency of peptide bond skipping varies depending on the 2A peptide and the host combination, therefore, optimization for the respective host is needed for the approach to become practical.

Using state-of-the-art technology, we developed vectors that combinea highly active unidirectional promoter (EF) or a bidirectional promoter (Ribi) with a variety of reporter proteins and epitope tags for optimized transgene expression in *N. oceanica* CCMP1779. Moreover, we optimized a viral 2A peptide for polycistronic expression to generate a multigene expression system for this oleaginous microalga. This vector toolkit was successfully used to manipulate the EPA biosynthesis pathwayin *N. oceanica* CCMP1779.

RESULTS

Genes encoding enzymes of the eicosapentaenoic acid biosynthesis pathway are co-expressed

The identification of *ω6* intermediates ($20:4^{\Delta5,\Delta8,\Delta11,\Delta14}$) (Figure 3.1a) points to the presence of an *ω6* pathway for EPA biosynthesis in *N. oceanica* CCMP1779 (Figure 3.1b). We isolated the cDNAs for the 5 FADs and a putative *Δ6* fatty acid elongase (FAE) proposed to be involved in LC-PUFA biosynthesis in *N. oceanica* CCMP1779 (Figure 3.1b). These cDNA sequences served to update the gene models of the *Δ9*, *Δ12*, *Δ6*, *Δ5*, *ω3* FADs, and *Δ6* FAE. Their updated cDNA sequences were deposited at NCBI with the accession numbers KY214449, KY214450,

KY214451, KY214453, KY214454, and KY214452 respectively. Using the corrected gene models, we observed that the FAD genes are highly co-expressed under light:dark conditions with a maximum expression 6 hours after dawn (Figure 3.1c).

Computational tools and manual examination were used to predict functional domains, subcellular localization, and transmembrane sequences in these proteins in support of their initial functional annotations (Figure 3.2). The FADs contain typical fatty acid desaturase domains, in particular the crucial three histidine boxes for coordinating a diiron center in the active site (*230, 231*). Moreover, the *Δ6* and *Δ5*FADs contain a cytochrome b5 domain that donates electrons for desaturation as observed for front end desaturases of eukaryotic origin, as well as glutamine substitutions in the third histidine box characteristic of front end desaturases (*232-234*).

The *Δ6* FAE contains an ELO family elongase domain, which is involved in very long chain fatty acid elongation and sphingolipid formation (*235, 236*). Further evidence of a possible elongase function is provided by the presence of FLHXYHH and MYSYY motifs characteristic of *Δ6* and *Δ5* fatty acid elongases; the first is positioned with an upstream glutamine characteristic of PUFA elongases, and the latter is typically found in microalga *Δ6* and *Δ5* fatty acid elongases (*234, 237, 238*) (Figure 3.2).

All the *N. oceanica* CCMP1779 EPA biosynthesis proteins are predicted to contain membrane spanning domains by TMHMM2 (*239*). The heterokont specific subcellular localization algorithm HECTAR (*36*) predicted endoplasmic reticulum localization signals for the *Δ12* and *Δ6* FADs, and *Δ6* FAE (Figure 3.2).

N. oceanica CCMP1779 FADs catalyze the production of LC-PUFAs in yeast

In order to test the biological activity of the putative FADs and FAE from *N. oceanica* CCMP1779, the pathway was reconstituted in the heterologous host *Saccharomyces cerevisiae* using a gene stacking strategy (Figure 3.3a). *S. cerevisiae* contains a single *Δ9* FAD which produces 16:1^{Δ9} and 18:1^{Δ9}(*240*). Therefore, to generate EPA from the endogenous 18:1^{Δ9} it is necessary to introduce four additional desaturases and an elongase. Towards this end we co-expressed cDNAs encoding the *Δ12*, *Δ6* and *Δ5*, *ω3* FADs and the *Δ6* FAE under the control of galactose inducible promoters in the *S. cerevisiae* strain InvSc1 (Figure 3.3a).

The presence of these constructs in the absence of galactose induction or the introduction of a *LacZ* expression control vector (Sc-LacZ strain) did not alter the cellular fatty acid composition of *S. cerevisiae* (Figure 3.3b, Figure 3.4a, Table 3.1). Induction of the *Δ12* and *Δ6* FADs led to the production of 18:2^{Δ9,Δ12}(1.8% of total fatty acids) and 18:3^{Δ6,Δ9,Δ12} (4.1%) in the Sc-Δ12+Δ6 strain (Figure 3.2b). The putative *Δ6* FAE was introduced in combination with *Δ12* and *Δ6* FADs resulting in elongation of 18:3^{Δ6,Δ9,Δ12} to 20:3^{Δ8,Δ11,Δ14} (di-homo gamma linolenicacid, DGLA) in the Sc-Δ12+Δ6+E6 strain (Figure 3.3b). Finally, co-expression of the *Δ12*, *Δ6*, *Δ5*, *ω3* FADs and the *Δ6* FAE in the Sc-Δ12+Δ6+E6+Δ5+ω3 strain resulted in the formation of EPA (20:5^{Δ5,Δ8,Δ11,Δ14,Δ17}) (0.1%) (Figure 3.3b, Table 3.1). Reconstruction of this LC-PUFA pathway in *S. cerevisiae* documented the combined functionality of the proteins as predicted from their respective sequences.

Interestingly, the 20:4 fatty acid produced by Sc-Δ12+Δ6+E6+Δ5+ω3, 20:4^{Δ8,Δ11,Δ14,Δ17} differed from the major 20:4^{Δ5,Δ8,Δ11,Δ14} found in *Nannochloropsis* (*11*)(Figure 3.1a, Figure 3.3b). This difference could be due to the *ω3* FAD acting on 20:3, preceding the *Δ5* reaction when present in *S. cerevisiae*. To test this hypothesis, we expressed the *Δ5* and *ω3* FADs individually

and in combination in yeast to generate the strains Sc-Δ5+ω3, Sc-Δ5, and Sc-ω3. When Sc-Δ5+ω3 was grown in the presence of exogenous $20:3^{Δ8,Δ11,Δ14}$, this strain was able to produce $20:4^{Δ8,Δ11,Δ14,Δ17}$ and EPA ($20:5^{Δ5,Δ8,Δ11,Δ14,Δ17}$) establishing the activity of the final two desaturases (Figure 3.4a, Table 3.1). To further examine the substrate specificity of the *Δ5* and *ω3* FADs, yeast expressing the individual FADs were supplied $20:3^{Δ8,Δ11,Δ14}$ and either $20:4^{Δ8,Δ11,Δ14,Δ17}$ or $20:4^{Δ5,Δ8,Δ11,Δ14}$, individually or in combination (Figure 3.4b-d, Table 3.1). When supplied with $20:3^{Δ8,Δ11,Δ14}$, Sc-Δ5 produced $20:4^{Δ5,Δ8,Δ11,Δ14}$ and Sc-ω3 produced $20:4^{Δ8,Δ11,Δ14,Δ17}$ (Figure 3.4b, Table 3.1). Supplying Sc-Δ5 and Sc-ω3 with $20:4^{Δ8,Δ11,Δ14,Δ17}$ and $20:4^{Δ5,Δ8,Δ11,Δ14}$ respectively, resulted in the production of EPA ($20:5^{Δ5,Δ8,Δ11,Δ14,Δ17}$) (Figure 3.4c, Table 3.1). Sc-Δ5 supplied with $20:3^{Δ8,Δ11,Δ14}$ and $20:4^{Δ8,Δ11,Δ14,Δ17}$ resulted in the respective products $20:4^{Δ5,Δ8,Δ11,Δ14}$ and EPA ($20:5^{Δ5,Δ8,Δ11,Δ14,Δ17}$) (Figure 3.4d, Table 3.1). Supplying Sc-ω3 with $20:3^{Δ8,Δ11,Δ14}$ and $20:4^{Δ5,Δ8,Δ11,Δ14}$ resulted in production of $20:4^{Δ8,Δ11,Δ14,Δ17}$ and EPA ($20:5^{Δ5,Δ8,Δ11,Δ14,Δ17}$) (Figure 3.4d, Table 3.1). These experiments revealed that the *Δ5* FAD can accept either *ω3* ($20:4^{Δ8,Δ11,Δ14,Δ17}$) or *ω6* ($20:3^{Δ8,Δ11,Δ14}$) substrates, and the *ω3* FAD can accept either $20:3^{Δ8,Δ11,Δ14}$ or $20:4^{Δ5,Δ8,Δ11,Δ14}$ substrates when produced in yeast.

A vector toolkit for multigene expression in *Nannochloropsis* species.

To facilitate the co-expression of multiple coding sequences and characterization of the respective enzymes, we generated a set of vectors to overexpress multiple FADs in *N. oceanica* (Figure 3.5a). For single transcript expression vectors we chose the elongation factor promoter (EFpro) (*NannoCCMP1779_10181*) due to its high constitutive activity during light:dark cycles (Figure 3.5b). We generated a vector series (pNOC-OX) encoding a variety of reporter proteins and epitope tags, including hemagglutinin (HA), Green Fluorescent Protein (eGFP), Cyan

90

Fluorescent Protein (Cerulean variant), Yellow Fluorescent Protein (Venus variant), and the ultra-bright codon optimized (Table 3.2) NanoLuciferase (Nlux) (*97*), flanked by glycine-serine-glycine encoded linkers and a set of compatible restriction sites (AscI/HpaI, and MluI/NruI) which enable translational fusion of the epitope tags to either the C or N terminus of the targeted protein (Figure 3.5a).

To identify candidate bidirectional promoters in *N. oceanica*, a custom python script (Data 3.1) was used to find diverging gene pairs that co-expressed (Table 3.3). We selected the intergenic region between two ribosomal subunits as a promising candidate bidirectional promoter (Ribi) due to a high degree of co-expression and moderate expression levels throughout the light:dark cycle of the respective gene pair (*NannoCCMP1779_9669, NannoCCMP1779_9668*) (Figure 3.5b). The Ribi promoter (after modification to remove MluI and NruI recognition sites, Figure 3.6) was assembled with the best performing selection marker P2A cassette coding sequences (BleR-P2A(60) and HygR-P2A(60)) and the toolkit epitope tag coding sequences to generate the pNOC-stacked vector series (Figure 3.5a).

To test the newly developed vectors we transformed *N. oceanica* CCMP1779 with pNOC-OX-CFP, pNOC-OX-Nlux, and pNOC-stacked-Nlux. Production of CFP and Nlux was confirmed in selected strains by immunoblotting with α-GFP and α-HA antibodies respectively (Figure 3.5c, Figure 3.7a). In order to assess the activity of the selected promoters, transformants of *N. oceanica* with pNOC-OX-Nlux and pNOC-stacked-Nlux were screened for their luminescence signals. To quantitatively compare Nlux reporter signal from each promoter, luminescence from an equal number of cells of pNOC-OX-Nlux and pNOC-stacked-Nlux transformants was measured. Reflecting the higher activity level of EFpro than Ribi, the

luminescence of Nlux in pNOC-OX-NLux was greater than in pNOC-stacked-Nlux lines (Figure 3.7b).

Viral derived 2A peptides are used for polycistronic expression of multiple transgenes in eukaryotes and are widely used to tie resistance markers to the production of target proteins. We first determined the ribosomal skipping efficiency of the three mostly commonly used 2A peptides of ~20 aa, designated 2A peptide(amino acid length). The F2A(24), T2A(18), and P2A(19) coding sequences were appended to the zeocin/bleomycin (BleR) and/or hygromycin (HygR) resistance marker genes followed by insertion of the HA-tagged firefly luciferase (Flux) coding sequence (Figure 3.8a). Introduction of these constructs into *N. oceanica* CCMP1779 resulted in zeocin or hygromycin resistant colonies, of which those with high luciferase activity were selected for further study. Immunoblotting detected full-length BleR-2A-Flux protein, and only small amounts of released Flux in the first round of screening (Figure 3.8b, Figure 3.9). Ribosomal skipping efficiency was less than 10% for F2A, and T2A sequences, while P2A had an efficiency of 10-30% (Figure 3.9c). In order to increase ribosomal skipping efficiency the N-terminal F2A peptide encoding sequence was extended to 58 aa, and the P2A sequence to 30 aa, 45 aa, and 60 aa. These changes enhanced the ribosomal skipping efficiency for F2A(58) to ~20%, for P2A(30) to ~40%, for P2A(45) to ~50%, and for P2A(60) to >50% (Figure 3.8b, Figure 3.9). Based on these results, the extended P2A sequence was selected as the most promising 2A peptide for use in *N. oceanica* CCMP1779.

Overexpression of EPA biosynthesis genes in *N. oceanica* CCMP1779

We first generated lines expressing the *Δ9*, *Δ12*, and *Δ5* FADs under the control of the EF promoter in *N. oceanica* CCMP1779 (designated DOX9, DOX12, and DOX5 lines, respectively)

(Figure 3.10a). *Δ9* FAD coding sequence was cloned into pNOC-P2A(30) (Figure 3.8) to generate the pNOC-DOX9 vector. The *Δ12* and *Δ5* FAD coding sequences were placed in overexpression vectors with C-terminal epitope tags (pNOC-DOX12-HA, pNOC-DOX12-CFP, pNOC-DOX5). Lines with changes in their fatty acid profiles were selected for further studies. The DOX12 and DOX5 lines produced appropriately sized proteins (Figure 3.10b). However, we did not detect full-length HA-*Δ9* FAD in DOX9 lines but only several small molecular weight polypeptide products (Figure 3.10b), which could be due to degradation of the tagged protein. Confocal microscopy of DOX5 and DOX12 lines revealed that the desaturase CFP fusion proteins have an ER subcellular location, as indicated by overlap with an ER specific marker dye, supporting the protein annotation and predicted location (Figure 3.11).

In order to test whether overproduction of more than one FAD would improve LC-PUFA accumulation in *N. oceanica* CCMP1779, we transformed wild-type and the DOX5 strain B3 with a desaturase stacking vector, containing the HA-*Δ9* FAD and *Δ12* FAD-HA coding regions under the control of the Ribi bidirectional promoter (pNOC-stacked-DOX9+12) (Figure 3.10a,c). DOX9+12 and DOX5+9+12 lines with changes in their fatty acid profile were selected for further analyses. Immunoblotting of selected lines detected full-length *Δ12* FAD-HA and HA-*Δ9* FAD peptides in all lines, and *Δ5* FAD-CFP in the triple FAD overexpression lines (Figure 3.10d).

To determine the level of overexpression, mRNA levels were quantified by qPCR (Figure 3.12a). Lines transformed with the pNOC-OX vectors (Figure 3.5a), resulted in expression up to 3.5 fold higher than the wild-type, while DOX9 overexpression using pNOC-P2A(30) led to~8fold increase in mRNA. In the DOX9+12 lines, the *Δ12* FAD mRNA content was 10-30

93

times that of the wild-type and displayed large differences between the lines, while the *Δ9* FAD

expression was increased 2-4 fold.

Increase in LC-PUFA content in FAD overexpressing lines

Fatty acid profiling of wild-type, empty vector controls (EV), and DOX lines showed that an

increase in FAD production altered fatty acid proportions (Table 3.4). While EV controls did not

cause alterations to the fatty acid profile compared to the wild-type, the overexpression of *Δ9*

FAD led to a small increase in the mol percent of its product $18:1^{\Delta9}$ and the overexpression of the

Δ12 or *Δ5* FADs resulted in a higher LC-PUFA fraction. We did not observe a further increase in

LC-PUFAs in the stacking lines, DOX9+12 and DOX5+9+12, and these lines had a similar ~

25% increase in EPA ($20:5^{\Delta5,\Delta8,\Delta11,\Delta14,\Delta17}$)and 35% increase in LC-PUFAs mol ratio as the single

FAD overexpressing lines (Table 3.4).

To further assess the impact of FAD overproduction, we measured cell growth over

several days (Figure 3.12b). The growth rate of EV, DOX5, DOX12, DOX9+12, and

DOX5+9+12 lines, but not DOX9, was increased with respect to the wild-type (Figure 3.13).

This effect could be related to the decreased average cell sizes of these lines (Figure 3.12c). Total

cellular fatty acid content per cell was likewise decreased in DOX5, DOX12, DOX9+12, and

DOX5+9+12 lines, while EV and DOX9 lines were unaffected (Table 3.5).

DISCUSSION

Expanding transgenic techniques in *N. oceanica*

A strong biotechnological interest has driven development of transgenic tools for

Nannochloropsis species in recent years (*11, 13, 14, 19, 79*). However, a comprehensive and

modular protein production toolkit vector set allowing the engineering of convenient epitope tagged fusion proteins was lacking. Such a toolkit with multiple genetic reporters, selection markers, two additional promoters, and several strategies for multi-gene expression is now available.

We tested two new promoters for protein overproduction in *N. oceanica* CCMP1779. To enhance expression of single transgenes in *Nannochloropsis* species, engineers have used several endogenous promoters including the lipid droplet surface protein (LDSP) (*11, 91*) β-tubulin, heat shock protein 70, and ubiquitin extension promoters (UEP) (*13, 79*). We tested an elongation factor promoter that displays high and stable expression throughout light:dark cycles (*15, 45*), and has been utilized in diverse organisms (*241, 242*), including diatoms (*243*) . A variety of promoters allow gene expression at different levels and in response to different environmental conditions. Furthermore, repeated use of transgenic elements can lead to genetic instability and as gene stacking techniques mature in *Nannochloropsis* species multiple promoters are desirable to modulate the expression of many genes.

Bidirectional promoters are extremely useful for multi-gene expression. The endogenous bidirectional promoter (VCP) between a pair of violaxanthin/chlorophyll a-binding proteins (*NannoCCMP1779|4698* and *NannoCCMP1779|4699*) has been used to drive resistance genes (*14*) and has been paired with coding sequences for fluorescent proteins with subcellular location tags (*24*). We have tested the ribosomal component S15 and S12 bidirectional promoter (Ribi) and identified several additional bidirectional promoters that may also be suitable for transgenic expression including histones, other VCPs, and nitrate reductase (Table 3.3). These bidirectional promoters can be used for selection of high target transgene expression or for reducing potential silencing effects of transgenes by linking transgene production to a resistance marker.

Transgenic techniques in a number of eukaryotes have exploited 2A peptides, which enable a single transcript to encode multiple discrete protein products. A variety of 2A peptides have been identified and exploited with differing levels of efficiency in a number of eukaryotes, including yeast (*244*), animals (*226, 228*), plants (*245*), insects (*246*), and algae (*101, 247*). Applying this approach to a heterokont, we have successfully developed an extended P2A sequence that is efficient for peptide bond skipping in *N. oceanica* CCMP1779. When placed between a resistance gene and firefly luciferase reporter, we obtained skipping efficiency of greater than >60%. Moreover, the overproduction of the *Δ9* FAD coding sequence downstream of BleR-P2A(30) resulted in non-significant increase in $18:1^{Δ9}$, indicating enzymes produced as a P2A mediated fusion with a resistance gene are functional in *N. oceanica* CCMP1779. Our extensions of the P2A sequence showed diminishing increases in cleavage efficiency suggesting an optimal length has been found (Figure 3.9). Therefore, the P2A peptide is a promising tool for producing discrete proteins from a single reading frame in *N. oceanica* CCMP1779, and further mutational studies of the P2A sequences could potentially yield higher performing variants. We anticipate that the development of the pNOC-OX and pNOC-stacked expression vectors will facilitate research in *Nannochloropsis* species and lay the foundation for combinatorial genetic engineering in an oleaginous microalga chassis for synthetic biology.

Characterization of the EPA biosynthetic pathway

As an example for multi-gene metabolic engineering in *N. oceanica* CCMP1779, we identified and characterized the four fatty acid desaturases and one elongase involved in the production of the LC-PUFA, EPA (20:5). Heterologous expression in *S. cerevisiae* confirmed the predicted biochemical function of each gene in the pathway, leading to EPA production without the supply

of external fatty acids (Figure 3.3 , Table 3.1). We also showed that the $\Delta5$ and $\omega3$ FADs can both use $\omega3$ or $\omega6$ fatty acids when reconstituted in yeast (Figure 3.4). Similarly, *N. oceanica $\Delta6$* FAD expressed in *S. cerevisiae* was able to process $\omega3$ or $\omega6$ fatty acids when exogenously supplied (*248*), suggesting substrate preference may be determined by glycerolipid acyl carriers.

We used our gene stacking vector toolset (Figure 3.5) to increase expression levels of single or multiple endogenous genes involved in EPA synthesis. Strains overexpressing one, two, or three FAD encoding genes displayed elevated fractions of LC-PUFAs, notably EPA (Table 3.4). However, we did not observe a further increase in LC-PUFAs in the lines overproducing more than one enzyme. The isolation of the EPA biosynthetic genes enables further studies into regulation of the pathway and provides tools useful for manipulation of LC-PUFA production in *N. oceanica* and heterologous hosts.

Metabolic engineering for increased EPA content in *N. oceanica*

The overexpression of single $\Delta5$, or $\Delta12$ FADs led to an approximate 25% increase in EPA mol ratio (Table 3.4), in line with observations in other heterokonts. The overproduction of an endogenous $\Delta5$ FAD in *P. tricornutum* under the control of a fucoxanthin chlorophyll a/c binding protein gene promoter led to a 58% increase in EPA (*217*). Overproduction of an elongase involved in LC-PUFA production in *T. pseudonana* led to a 40% increase in EPA (*219*). It has been previously shown that in *N. oceanica* CCMP1779 the inducible overproduction of the $\Delta12$ FAD using the LDSP promoter produced a 50-75% increase in the $20:4^{\Delta5,\Delta8,\Delta11,\Delta14}$ mol ratio during nitrogen deprivation and stationary phase (*91*).

The FAD genes in *N. oceanica* CCMP1779 are co-expressed under diurnal light:dark cycles during exponential growth (Figure 3.1) (*45*). Therefore, we had initially hypothesized that

high expression of multiple enzymes is necessary to maintain flux through the pathway and further increase end product concentration. However, the overproduction of the multiple desaturase proteins in the DOX9+12 and DOX5+9+12 lines did not have an additive effect. These results suggest that there might be a limit to PUFA content in *N. oceanica* CCMP1779 before cell physiology is affected under the growth conditions tested.

The cell division rate of DOX lines was not reduced when compared to wild-type cells. However, decreases in total cellular fatty acids of DOX lines expressing the *Δ12* or *Δ5* FAD coding sequences indicates that the physiology of *N. oceanica* is compromised by desaturase overproduction (Table 3.5). In *N. oceanica* EPA is an endogenous major LC-PUFA that is associated with polar lipids of plastidic membranes and likely has a functional role in the photosynthetic membrane (*249*). Typically, in response to environmental conditions, membrane saturation is altered to maintain membrane fluidity, and LC-PUFAs are increased in membranes during lower temperatures and high-light. Interference with native membrane composition may have deleterious effects that limit the accumulation of EPA beyond certain levels.

A higher relative increase in some LC-PUFAs has been achieved when the strategy has been to introduce novel LC-PUFAs (*124, 220*) or to elevate the level of minor LC-PUFAs, initially present in small amounts (*102*). For example, the heterologous expression of both a *Δ6* FAD and a *Δ5* FAE coding sequence in *P. tricornutum* led to a stronger DHA accumulation than in the single overexpressing lines (*102*). However, wild-type *P. tricornutum* contains only trace amounts of DHA and in these transgenic lines the accumulation of DHA correlated with a strong decrease in EPA levels, indicating that these transgenics increased partitioning towards DHA but not an overall increase in flux through the LC-PUFA pathway. These observations support the hypothesis of a limit of LC-PUFA imposed on the cell under specific growth conditions.

In addition to possible negative effects on algal physiology, it is likely that other steps in LC-PUFA biosynthesis, such as fatty acid elongation or *Δ6* or *ω3* desaturation, are rate limiting and need further enhancement to increase LC-PUFA content. Moreover, it is possible that increased LC-PUFA turnover by β-oxidation is involved in maintaining a balanced LC-PUFAs content (*124, 250*). Downregulation of the native EPA biosynthetic pathway at a transcriptional or post-transcriptional level may also compensate for the overproduction of FADs described here.

Strategies for sequestration of LC-PUFAs in TAG could be an option to overproduce LC-PUFAs without compromising cell physiology (*91*). Although *N. oceanica* contains little LC-PUFAs in TAG under normal conditions, the content of LC-PUFAs in TAG increases following cellular stresses (*176, 210*). We observed a decrease 16:0 and 16:1 per cell in the DOX lines (Table 3.5), however, the overexpression of a *Δ12* FAD using a stress inducible promoter (*91*) did not cause such a decrease, indicating that the timing of gene expression is likely to be important for minimizing negative effects of enhanced LC-PUFA content by maybe sequestering LC-PUFAs in TAG. Diacylglycerol acyltransferases (DGATs) and phosphatidylglycerol acyltransferase (PDATs) that have preferences for 20C fatty acids also offer potential tools for channeling EPA to the TAG stores (*251, 252*). *N. oceanica* contains 12 DGATS and two PDATs that are likely to have different substrate preferences, including for LC-PUFAs (*31*). Studies into the functional effects of altering the fatty acid profile as well as identifying compensating internal forms of regulation are needed to identify strategies for further accumulation of LC-PUFAs.

EXPERIMENTAL PROCEDURES

Growth conditions

Axenic cultures were grown in shaking flasks of f/2 medium under 100 µm s^{-1} m^{-2} white light, at 22° C and 120 rpm. For protein, metabolite and gene expression analyses, cells were grown under constant light and samples were collected from mid log cultures (30 10^6 cells ml^{-1}).

Cloning of *N. oceanica* CCMP1779 EPA pathway genes

Axenic cultures were under 12:12 light:dark cycle at 22° C (*45*). Cell counts and cell size measurements were obtained using a Coulter Counter Z2 (Beckman Coulter) using a profile with a range of 1.8-3.6 µm. *N. oceanica* CCMP1779 at midlog phase was used for RNA isolation as described previously (*45*). First strand DNA synthesis was accomplished using SuperScript III with oligo dT (NEB). cDNAs were amplified using primers shown in Table 3.6 and Q5 polymerase (NEB), blunt cloned into pCR-Blunt (Thermo Scientific) and sequenced.

Yeast transformation and expression

EPA pathway genes were cloned into yeast expression vectors containing galactose inducible promoters. The elongase PCR product was integrated into pYES2.1-topo (Invitrogen). Desaturases were amplified with the addition of C-terminal 6X histidine tails and restriction sites for integration into pESC-his and pESC-leu (Agilent). The PCR product was digested with the noted restriction enzymes (Table 3.6) and ligated into the yeast expression vectors. InvSc1 yeast cells (*253*) were transformed with the expression vectors using the Frozen-EZ Yeast Transformation II Kit (Zymo Research) and selected on SC (ClonTech) medium with proper drop-out auxotrophy selection. Several colonies from each transformation were selected for

further experimentation and were grown in 5 ml of SC overnight at 30° C. The overnight cultures were collected by centrifugation at 1000 x G for 5 minutes, thoroughly decanted, resuspended in 5 ml of SC without sugar, histidine, leucine, and uracil and OD600 was measured in duplicate. For the zero-hour time point fatty acid analysis 0.5 ml culture was collected by centrifugation at 13,000 x G, decanted and frozen in liquid nitrogen. Flasks of 5 ml SC 2% galactose were inoculated at 0.4 OD600 and grown at 20° C and 24 hour and 48-hour time points were collected. For substrate feeding 0.1% NP-40 (Sigma-Aldrich), as a detergent, and 0.5 mM free fatty acids (Santa Cruz Biotechnology) in glucose and galactose SC medium were included. Fatty acid analysis with washed cell pellets and 5 μg of pentadecanoic acid internal standard were conducted. LC-PUFA authentic standards were prepared in a separate reaction for confirmation of LC-PUFA running times. Fatty acid methyl ester preparation and extraction was performed as described previously (Liu et al., 2013).

Protein sequence analysis

Protein sequences were generated by translation of the obtained cDNA using canonical codon usage. To identify functional domains, the protein sequences were submitted to CDD BLAST (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). To identify localization signals, protein sequences were submitted to HECTAR (http://webtools.sb-roscoff.fr/) (*36*). To identify transmembrane domains, protein sequences were submitted to TMHMM2 (http://www.cbs.dtu.dk/services/TMHMM/).

Identification of bidirectional promoters in *N. oceanica* CCMP1779

A custom python script was used to assemble the coding regions of each gene as determined in the genome assembly and annotation of *N. oceanica* CCMP1779 (*11*), and only genes with start and stop codons were added to the final list. These coding regions were assessed with the CUSP function of the EMBOSS program (Table 3.2). A custom python script was used to identify diverging gene pairs with intergenic regions of <1500 base pairs, with suitable gene expression during light:dark cycles (*45*). Putative gene pairs were manually examined to determine functional annotation (Table 3.3).

Construction of *Nannochloropsis* expression vectors

The elongation factor promoter was amplified with the primers EFpro GW F+ and EFpro GW R- listed in Table 3.6 and inserted into the pENTR gateway entry vector (Invitrogen), sequenced, and transferred to pNoc-Dlux by a LR clonase (Invitrogen) reaction. The EF promoter was placed 5' of the codon optimized firefly luciferase reporter coding sequence (DNA 2.0) in a *Nannochloropsis* transformation vector (pNOC-Dlux), and transformants with luciferase activity were confirmed by immunoblotting detected the presence of the HA tagged protein (Figure 3.9). Subsequent modifications included removal of the luciferase reporter coding sequence by digestion with AscI and SacI and replacement with an HA tag coding sequence and a multicloning site. Primers for insertion of the reporter genes with glycine-serine-glycine linkers as listed in table 3.6 were used to amplify, fluorescent protein cerulean, venus, GFP, and NanoLuciferase coding sequences, which were then blunt cloned, sequenced and inserted into the HpaI and MluI sites. A codon optimized NanoLuciferase sequence was synthesized (IDT).

The 2A peptide coding sequences were appended to the 3' end of the zeocin (BleR) and hygromycin (HygR) resistance genes by site directed mutagenesis PCR of resistance vectors, pNOC-401 and pNOC-411, using Q5 polymerase and the primers listed in Table 3.6. For elongation of the 2A peptide coding sequences, plasmids containing shorter variants were used as a template for site directed mutagenesis PCR. To insert a firefly luciferase (with c-terminal HA tag) coding sequence 3' of the BleR-2A coding sequences, the 2A vector was digested with MluI/SacI and the firefly luciferase released from pNOC-Dlux by digestion with AscI/SacI. Ligations yielded pNOC-411-2A-Flux constructs that were introduced into *N. oceanica* to assess 2A peptide skipping efficiency. A new multicloning site containing PspomI/XbaI/AfeI/stop codon/KpnI recognition sequences followed by the heat shock terminator was inserted after the 2A peptide coding sequence. The 2A resistance cassette was inserted into the pNOC-OX-NLux vector in the opposite orientation using the restriction sites XhoI/PsiI. The Ribi promoter was amplified from the *N. oceanica* genome, subcloned, subjected to two round of site directed mutagenesis to remove the MluI and NruI sites (Figure 3.6). The modified promoter was inserted between the GFP and 2A coding sequences to generate the vectors pNOC-Stacked-HygR-NLux and pNOC-Stacked-BleR-NLux.

The coding and terminator regions of the Δ9 FAD gene were PCR amplified from *N. oceanica* genomic DNA with the addition of an N-terminal HA tag and the restriction sites MluI and EcorV, NsiI using the primers D9 mlui HA start F+ and D9 ecorv nsi R- listed in table 3.6. The PCR product and the vector pNOC-411-P2A(30) was digested with MluI and NsiI and ligated together. To overexpress the desaturases coding sequences in the pNOC-OX vector series, desaturase coding regions were PCR amplified with MluI/HpaI sites and ligated to the AscI and HpaI sites of pNOC-OX. To assemble a desaturase stacking vector the Δ12-HA-LDSP

terminator sequence was PCR amplified to add the restriction sites EcoRI and PsiI, for digestion

and ligation into the corresponding sites of pNOC-DOX9. The bidirectional ribosomal

component promoter was PCR amplified from the *N. oceanica* genome with the restriction sites

EcoRI and XhoI and ligated between the two desaturase expression cassettes. All of the vectors

for expression in *S. cerevisiae* and *N. oceanica* are listed in Table 3.7 and the complete annotated

sequences are included below. Complete vector seqences are available in Data 3.2.

Nannochloropsis transformation

Vectors were linearized by restriction digestion, and purified and concentrated by ethanol

precipitation. *N. oceanica* CCMP1779 transformation was performed according to the method of

Vieler et al., (*11*) with 3 µg of vector DNA, with 30 µg carrier DNA (Invitrogen UltraPure™

Salmon Sperm DNA Solution). Transformed cells were allowed to recover for 48 hours then

plated in top agar with the respective selection. After 3-4 weeks, individual colonies were

resuspended in 100 µl f/2. From each transformation ~20 colonies were screened for increased

LC-PUFA content, and 2-3 colonies identified as positive.

Nannochloropsis luminescence assays

N. oceanica CCMP1779 culture was mixed with f/2 supplemented with either firefly luciferin

(Gold Biotech) or NanoLuciferase substrate (Promega), at a final volume of 200 µl, with 500 µM

firefly luciferin or 10,000X dilutions of NanoLuciferase substrate. For normalized

measurements, 1 million *N. oceanica* cells were used. Luminescence was measured with a

Centro XS3 LB960 luminometer (Berthold Technologies) over a 0.3 sec exposure.

Expression analysis in *N. oceanica* CCMP1779

For protein expression, frozen pellets from 5 ml culture were ball milled in 2 ml tubes (30Hz, 2 minutes) with a TissueLyser II (Qiagen). After addition of protein extraction buffer (100mM Tris pH 8.0, 2mM PMSF, 2% B-mercaptoethanol, 4% SDS) the sample was heated to for 3 min (60° C for FADs, 80° C for other proteins), centrifuged at 13,000 x G for 3 minutes, and the supernatant was transferred to new tube. Protein content was determined using the RCDC assay (Bio-Rad) and equal quantities of protein were loaded for SDS-PAGE. Proteins were transferred to PVDF membranes (Bio-Rad) overnight at 4° C. Blots were blocked in TBST with 5% milk for 1 hour at room temperature and washed 6 times with TBST. For GFP detection we used α-GFP antibody (Abcam ab5450) 1:1000 in TBST with 5% BSA for 1 hour, and a secondary donkey α-goat HRP antibody (Santa Cruz sc-2020) 1:10000 in TBST with 5% milk. For HA detection, we used α-HA-HRP antibody solution (Roche 3F10) at 1:1000 in TBST with 5% milk for 1 hour. Signals were detected using clarity chemiluminescence reagent (Bio-Rad). Band quantification was conducted using Image Lab software (Bio-Rad).

RNA isolation, cDNA synthesis and real time PCR were performed as described previously (*45*). Real time PCR primers were checked for efficiency and specificity. The delta-delta Ct method was used to determine gene expression relative to the gene encoding the actin related protein (*ACTR*) *NannoCCMP1779_1845*.

Fatty acid methyl ester extractions in *N. oceanica*

Cells (2 ml) were collected by filtration through GF/C filters, and filters stored in screw top tubes at -80° C. Fatty acid methyl esters (FAMEs) extraction and analysis was carried out as described previously (*210*).

Confocal microscopy

Cerulean detection in transformed *N. oceanica* CCMP1779 was carried out with Olympus Spectral FV1000 microscope (Olympus, Japan) at the excitation wavelength of 435 nm (a diode laser). For endoplasmic reticulum labeling 50 nM DiOC6 (Sigma-Aldrich) in f/2 medium was used. Cells were labeled directly before microscopic analysis. An argon (488 nm) laser was used for DiOC6 excitation. Chloroplast autofluorescence was excited by using a solid state (515 nm) laser. CLSM figures represent Z-series images composed using the Olympus FluoView FV1000 confocal microscope software (Olympus).

APPENDICES

Appendix 3.1. Chapter 3 figures and tables.
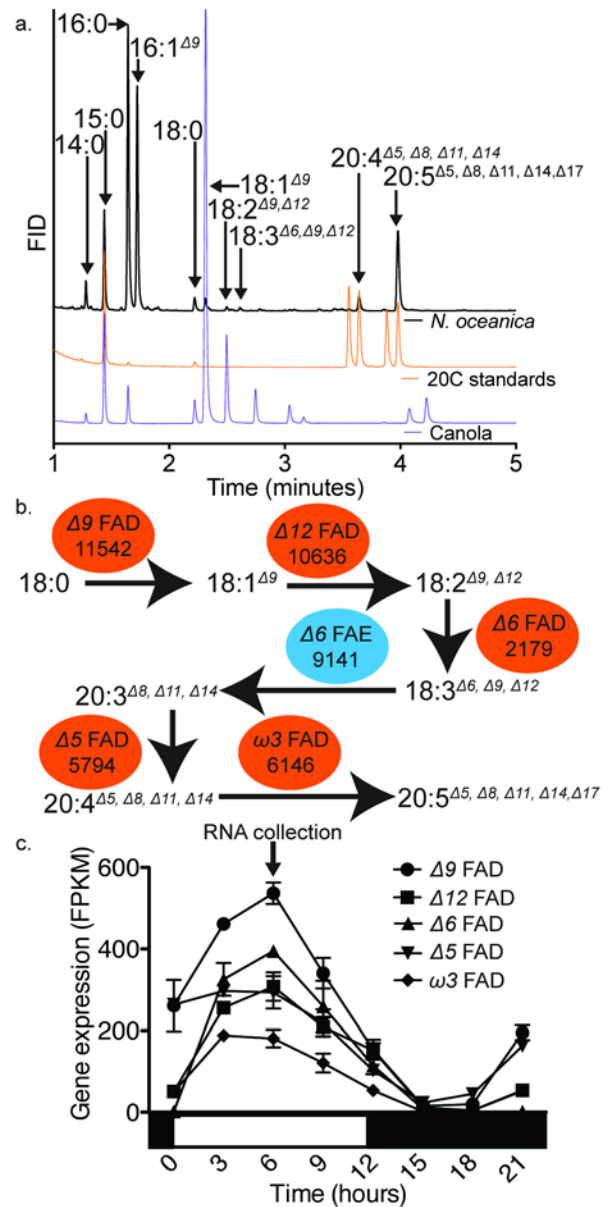
Figure 3.1. EPA biosynthetic pathway identification in *N. oceanica* CCMP1779.
(a) Identification of the fatty acid species by comparison with canola oil and 20 carbon LC-PUFAs standards (20:3$^{\Delta8,\Delta11,\Delta14}$, di-homo gamma linolenic acid; 20:4$^{\Delta5,\Delta8,\Delta11,\Delta14}$, $\omega6$ eicosatetranoic acid; 20:4$^{\Delta8,\Delta11,\Delta14,\Delta17}$, $\omega3$ eicosatetraenoic acid; and 20:5$^{\Delta5,\Delta8,\Delta11,\Delta14,\Delta17}$, eicosapentaenoic acid). (b) The EPA biosynthetic pathway includes five desaturases (red ovals, FADs) and an elongase (blue oval, FAE). Gene IDs are shortened from the *NannoCCMP1779_#* format. (c) Expression of the EPA biosynthetic pathway encoding genes during a light:dark cycle. The arrow indicates the time of cell harvest for cDNA isolation of EPA biosynthetic pathway genes. Gene expression was calculated using a previous experiment (*45*) and the corrected gene annotation. Values are average ± range of two independent cultures.
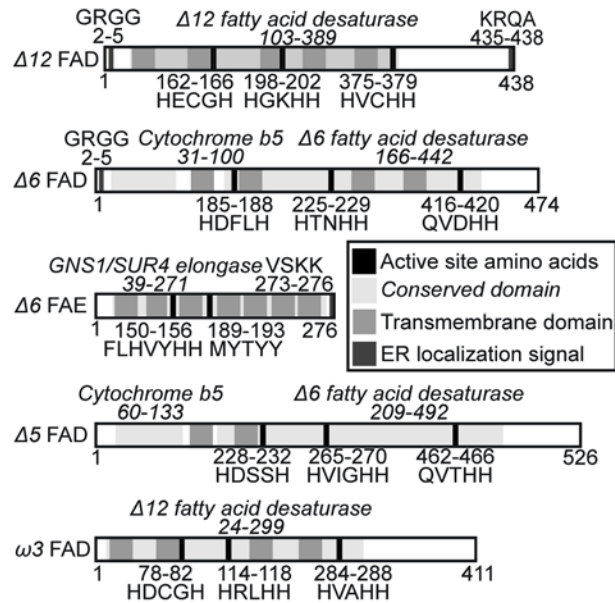
Figure 3.2. Computational annotation of protein sequences of isolated EPA biosynthetic genes.
Protein sequences were generated based upon isolated cDNA sequences. Active site amino acids and conserved domains identified using conserved domain BLAST. ER localization signals identified with HECTAR. Transmembrane domains predicted using TMHMM2.
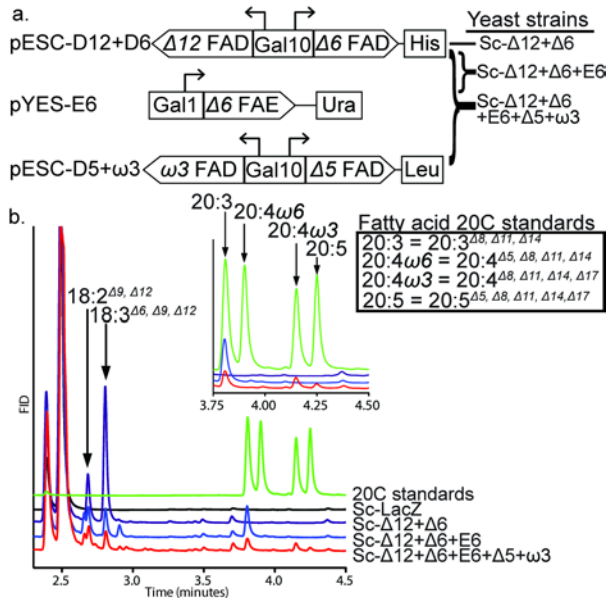
Figure 3.3. Galactose inducible expression of the EPA pathway genes in *S. cerevisiae*.
(a) The stacking strategy for EPA production in *S. cerevisiae* used three plasmids, containing different auxotrophic markers. Dual expression of FADs (pESC-D12+D6, and pESC-D5+W3) was achieved by a bidirectional Gal10 promoter, and expression of the FAE was under control of the Gal1 promoter (pYES-E6). (Co)transformation of vectors yielded *S. cerevisiae* strains Sc-Δ12+Δ6, Sc-Δ12+Δ6+E6, and Sc-Δ12+Δ6+E6+Δ5+ω3. (b) Representative GC-FID fatty acid profiles of Sc-LacZ negative control and FAD expressing yeast strains 48 hours post galactose induction.
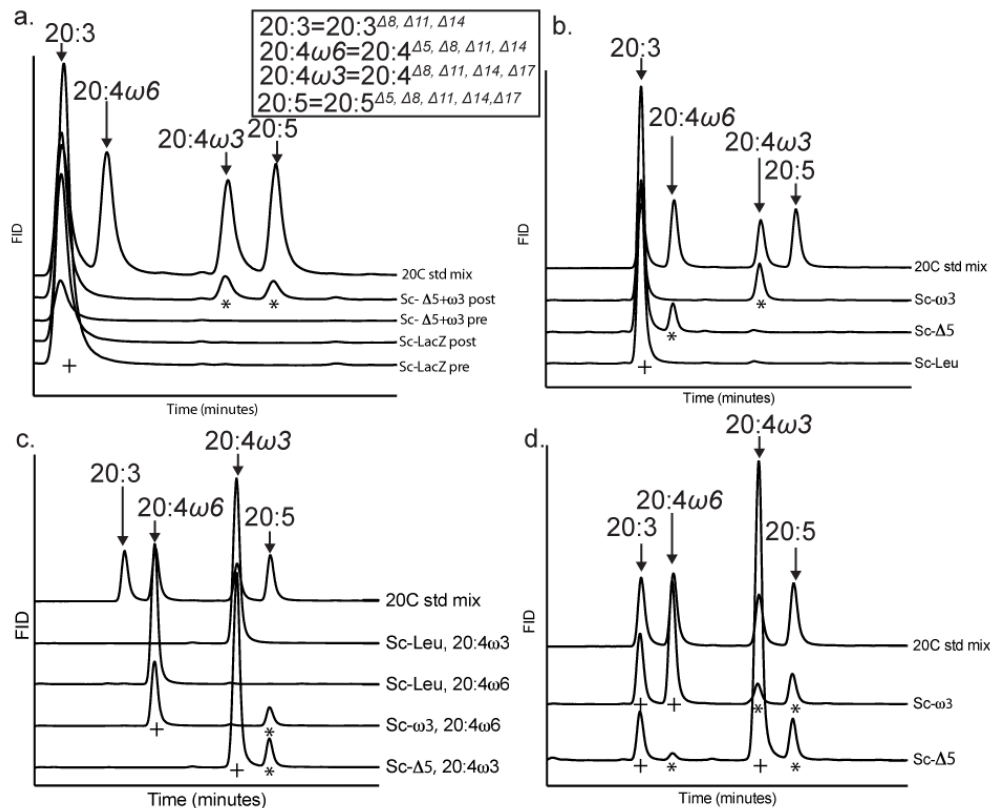
Figure 3.4. Functional characterization of EPA biosynthesis enzymes in *S. cerevisae*. Representative GC-FID fatty acid profiles. In the chromatograms, fatty acids supplied in the media are indicated with a (+), and fatty acids produced by yeast are indicated with a (*) underneath the respective peaks. 20C std mix, indicates standards of 20 carbon LC-PUFAs. (a) Yeast exogenously supplied with $20:3^{\Delta 8,\Delta 11,\Delta 14}$ and expressing either the $\Delta 5$ and $\omega 3$ FAD (Sc-$\Delta 5+\omega 3$), or the LacZ control gene (Sc-LacZ) before (pre) or after (post) 48 h gene expression induction by galactose. (b) Yeast exogenously supplied with $20:3^{\Delta 8,\Delta 11,\Delta 14}$ and expressing either $\omega 3$ FAD (Sc-$\omega 3$), $\Delta 5$ FAD (Sc-$\Delta 5$), or the negative control (Sc-Leu) after 48 h gene expression induction by galactose. (c) Yeast exogenously supplied with either $20:4\omega 3$ ($20:4^{\Delta 8,\Delta 11,\Delta 14,\Delta 17}$) or $20:4\omega 6$ ($20:4^{\Delta 5,\Delta 8,\Delta 11,\Delta 14}$), and expressing either $\omega 3$ FAD (Sc-$\omega 3$), $\Delta 5$ FAD (Sc-$\Delta 5$), or the negative control (Sc-Leu) after 48 h gene expression induction by galactose. (d) Yeast expressing either $\omega 3$ FAD (Sc-$\omega 3$) or $\Delta 5$ FAD (Sc-$\Delta 5$) and exogenously supplied with $20:4\omega 3$ ($20:4^{\Delta 8,\Delta 11,\Delta 14,\Delta 17}$) and $20:4\omega 6$ ($20:4^{\Delta 5,\Delta 8,\Delta 11,\Delta 14}$) respectively after 48 h gene expression induction by galactose.
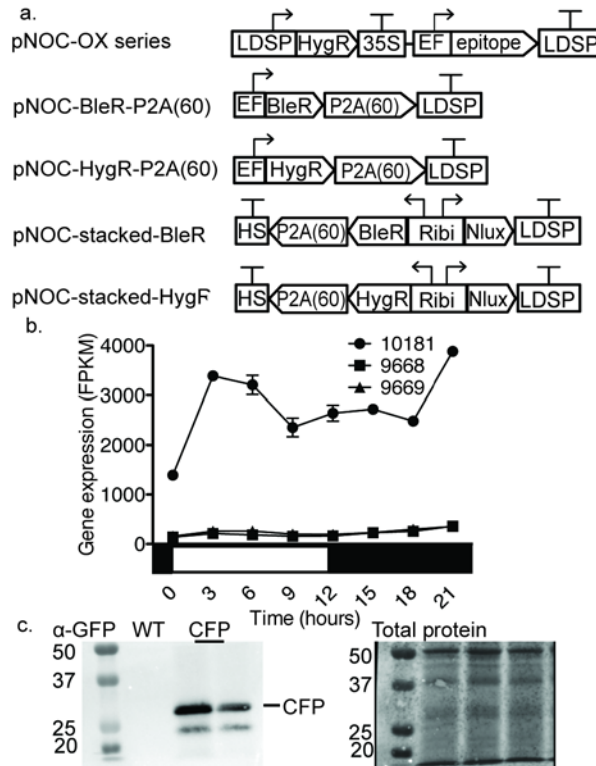
Figure 3.5. Assembly of native promoters, terminators, and a range of reporters to generate a transgenic expression toolkit for *N. oceanica* CCMP1779.

(a) The pNOC-OX vector series contains a series of reporters under the control the EF promoter with the LDSP terminator, and a hygromycin resistance gene under the control of the LDSP promoter and 35S terminator. Epitopes include the cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), green fluorescent protein (GFP) and NanoLuciferase (Nlux), and hemagglutinin peptide (HA) encoding sequences. P2A(60) is an extended 2A peptide coding sequence placed 3' of the zeocin resistance gene (BleR) or hygromycin resistance gene (HygR) for bicistronic expression by ribosomal skipping. The pNOC-stacked vectors utilize a bidirectional promoter (Ribi) for co-expression of a reporter and resistance gene with 2A peptide coding sequence and multicloning site followed by a heat shock terminator (HS). (b) RNA expression of endogenous genes corresponding to the promoters, *NannoCCMP1779_10181* (EFpro), and the gene pair *NannoCCMP1779_9669* and *NannoCCMP1779_9669* (Ribi promoter) under light:dark cycles (data from (*45*)). (c) Transgenic protein confirmation by immunoblot of pNOC-OX-CFP transformants detected with α-GFP. Total protein was stained using the dye DB71.

112

a. >nanno_989:5127..6310

tctgtggatggaggagggaaggaggggggggagtgagtgagtcgagagaacgacgactggggcaaagagaggatgacgccactgcaacaaaa
gacagggaagaatagaggctgttggaaaacaggaacacagtgaagaatagaggctgcagggtaacaggcacatttatcgcagagaggcgggcg
gcagggggcgatacgtaaagagacccctaaacttctccgcgtcatccgttcacgctacctccttccctcccctttactcccattgatcaccaaggcagcgc
ccgcagccaccagcagccaccgtcctcagtgtcacagcta**gacgcgt**ccccatcaggcagcagaccaccagaatatgcgctttcgcttcccatcat
atccctccctcctcccccttctattcctcatcatgcatcatttctttctccttgctgcatgcccctcttcttaccttggctaagatctaccggaggaagaagtctt
caacgatagataaatgcctgtatcgtgatggttgtttgaatggtggaagtgagccaggtattgagaacgaggggcgcaagggtgcgtggaagaagg
cagggagaggcgaggcgaaggagaagggcatgtggtgggaacgaagcctccacaccacagccacccatgcttttctctgttcgagttcccgggtt
ctactctcacacactacacgtgcacagaaaacattcacgcacaacagagatacagagggatccgtcgagatggtctacacgtcctaatgtcccctgct
ttcgaagcccagagccatcgctttctgtttttatccattcgcaacacatcctgttc**tcgcga**cctctcctcttcctggccaccatcgcttctccgtgccatgc
cctgtacatccttccctccctcgttgtctaccacactaatcatgcgctctgcagcagacactcggctactatttgtgtcgtacggattgaggggagggag
gggaagggaggcaagaagtaaatcggtggtcaacaacagacgcggtctccccaccctccctcattcctttcccctcctataatgagcgaagaaagcc
atggacaccgagtcacccactctcctaatccaccaccaccagggcactaacggactcacgctcaccaaactcggtaacaatatccactgcacacaga
tacacacgcaacc

b. >Ribi sdm

tctgtggatggaggagggaaggaggggggggagtgagtgagtcgagagaacgacgactggggcaaagagaggatgacgccactgcaacaaaa
gacagggaagaatagaggctgttggaaaacaggaacacagtgaagaatagaggctgcagggtaacaggcacatttatcgcagagaggcgggcg
gcagggggcgatacgtaaagagacccctaaacttctccgcgtcatccgttcacgctacctccttccctcccctttactcccattgatcaccaaggcagcgc
ccgcagccaccagcagccaccgtcctcagtgtcacagcta**gac...t**ccccatcaggcagcagaccaccagaatatgcgctttcgcttcccatcatat
ccctccctcctcccccttctattcctcatcatgcatcatttctttctccttgctgcatgcccctcttcttaccttggctaagatctaccggaggaagaagtcttca
acgatagataaatgcctgtatcgtgatggttgtttgaatggtggaagtgagccaggtattgagaacgaggggcgcaagggtgcgtggaagaaggca
gggagaggcgaggcgaaggagaagggcatgtggtgggaacgaagcctccacaccacagccacccatgcttttctctgttcgagttcccgggttcta
ctctcacacactacacgtgcacagaaaacattcacgcacaacagagatacagagggatccgtcgagatggtctacacgtcctaatgtcccctgctttc
gaagcccagagccatcgctttctgtttttatccattcgcaacacatcctgttc**tcAcga**cctctcctcttcctggccaccatcgcttctccgtgccatgccc
tgtacatccttccctccctcgttgtctaccacactaatcatgcgctctgcagcagacactcggctactatttgtgtcgtacggattgaggggagggaggg
gaagggaggcaagaagtaaatcggtggtcaacaacagacgcggtctccccaccctccctcattcctttcccctcctataatgagcgaagaaagccat
ggacaccgagtcacccactctcctaatccaccaccaccagggcactaacggactcacgctcaccaaactcggtaacaatatccactgcacacagata
cacacgcaacc

Figure 3.6. Modification of the Ribi promoter to remove restriction sites.
(a) The genomic sequence of the intergenic region between the bidirectional gene pair, *NannoCCMP1779_9669* and *NannoCCMP1779_9669*, is shown and the restriction sites MluI and NruI are bolded and underlined. (b) The Ribi promoter after site directed mutagenesis. The modified restriction sites are bolded and underlined, with nucleotide deletions represented by periods and alterations shown in upper case.
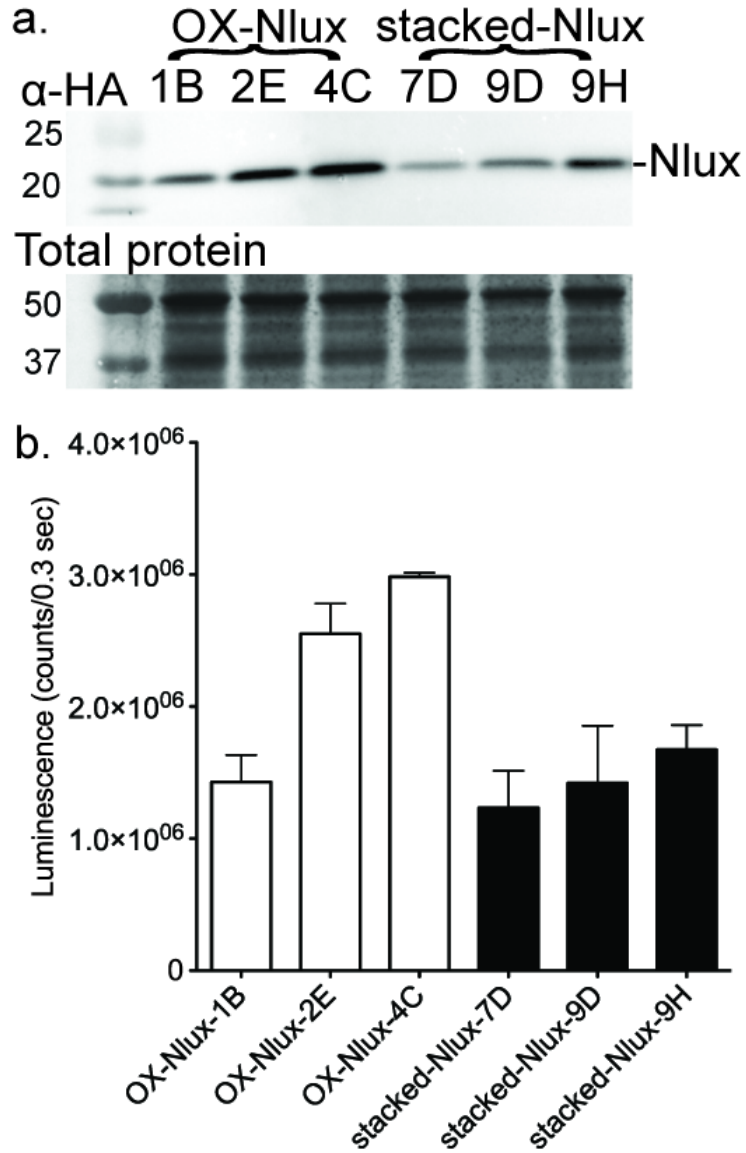
Figure 3.7. Assessment *N. oceanica* promoters' strength using NanoLuciferase.
(a) Immunoblotting with α-HA of *N. oceanica* lines transformed with pNOC-OX-Nlux and pNOC-Stacked-Nlux detects the Nlux-HA protein. Total protein stained using the dye DB71. (b) Normalized luminescence signal of *N. oceanica* lines transformed with pNOC-OX-Nlux and pNOC-Stacked-Nlux (average ± range, 2 technical replicates).
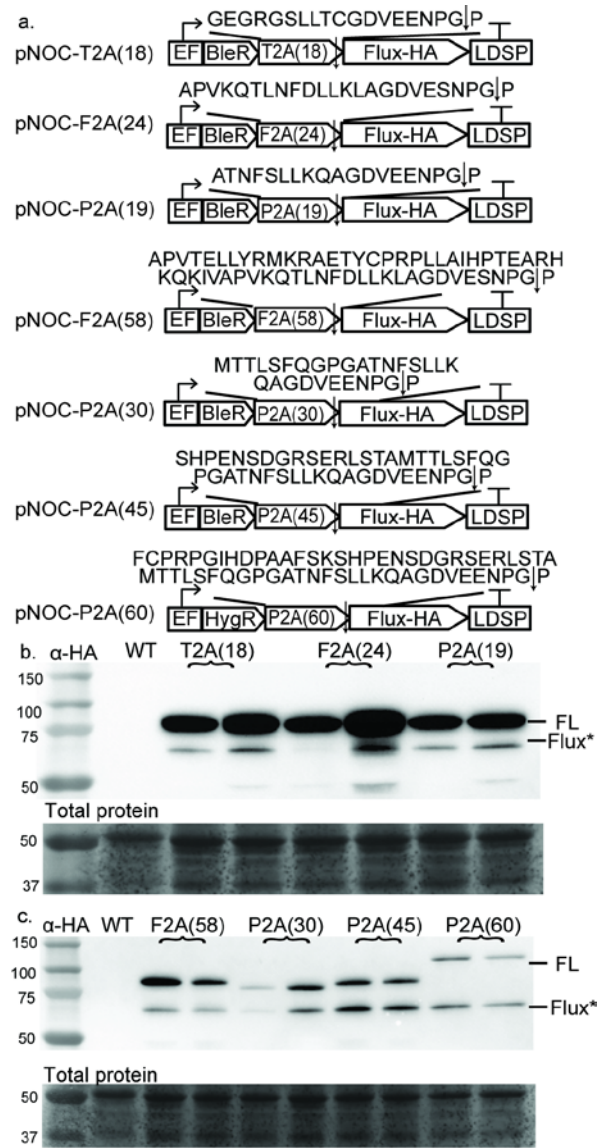
Figure 3.8. Optimization of 2A peptides ribosomal skipping efficiency in *N. oceanica* CCMP1779.
(a) Schematic of pNOC-2A vector series containing a zeocin resistance coding sequence (BleR) followed by a 2A peptide coding sequence and a multicloning site. Numbers in parentheses correspond to numbers of amino acids of the 2A peptide incorporated, and the encoded amino acid sequence is above (arrow indicates skipping site). For assessing function in *N. oceanica* CCMP1779 a coding sequence for firefly luciferase (Flux) with a C-terminal HA tag was inserted downstream of the 2A peptide. (b-c) Immunoblotting with α-HA antibody of lines transformed with the pNOC-2A vector series expressing full-length BleR-2A-Flux (FL) and Flux (Flux*). Total protein stained using DB71.
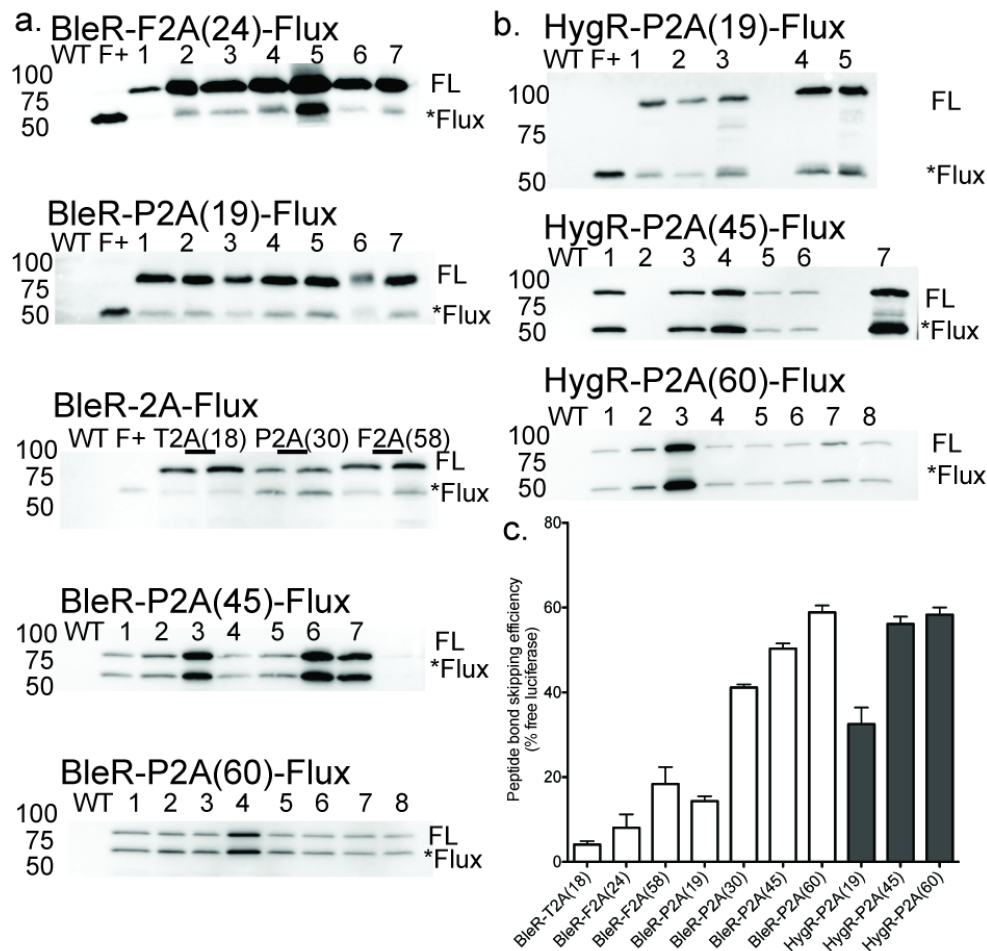
Figure 3.9. N-terminal extended 2A peptide screening for increased ribosomal skipping efficiency.

(a) Numbers in parentheses correspond to numbers of amino acids of the 2A peptide. Immunoblotting with α-HA detects full-length (FL) and released firefly luciferase (*Flux). Assesment of transformants producing BleR linked to Flux by a variety of 2A peptides: BleR-F2A(24)-Flux, BleR-P2A(19)-Flux, BleR-T2A(18)-Flux, BleR-P2A(30)-Flux, BleR-F2A(58)-Flux, BleR-P2A(45)-Flux, and BleR-P2A(60)-Flux. (b) Assesment of transformants producing HygR linked to Flux by different length P2A peptides: HygR-P2A(19)-Flux, HygR-P2A(45)-Flux, and HygR-P2A(60)-Flux. (c) Peptide bond skipping quantified using densitometric measurements of FL and Flux quantities and efficiency determined with the equation (*Flux/(FL+*Flux))×100; (average ± SEM, n=2-8).

116
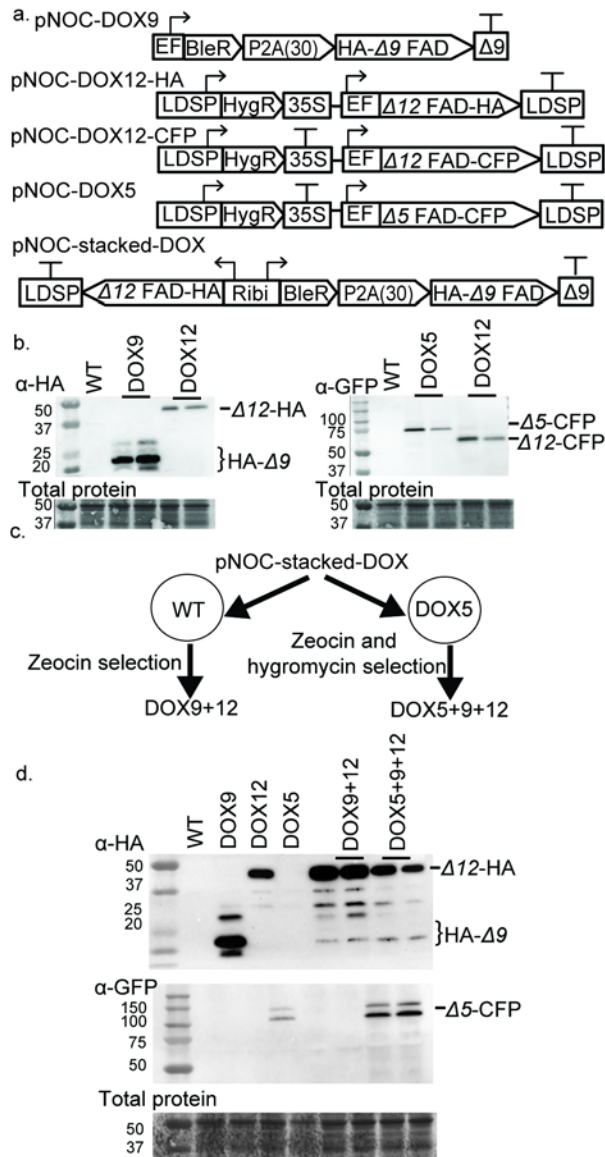
Figure 3.10. Vectors for FAD overproduction in *N. oceanica* CCMP1779.
(a) Schematics of pNOC-DOX vectors for overexpression of *N. oceanica* FADs. The *Δ9* FAD coding sequence was expressed 3' of the BleR-P2A(30) coding cassette in the vector pNOC-DOX9. Expression of *Δ5* or *Δ12* FAD coding sequences by a pNOC-OX vector with a CFP or HA epitope tag. The stacking vector (pNOC-stacked-DOX) uses the Ribi promoter to co-express the coding sequences for BleR-P2A(30)-*Δ9* FAD, and *Δ12* FAD. (b) Immunoblotting *N. oceanica* single desaturase DOX lines using α-HA or α-GFP antibodies. Total protein was stained using DB71. (c) Gene stacking strategy for *N. oceanica* by use of bicistronic pNOC-stacked-DOX and sequential introduction of vectors with different selection markers. (d) Immunoblotting of stacked DOX9+12 and DOX5+9+12 lines using α-HA or α-GFP antibodies. Total protein was stained using the dye DB71.

117
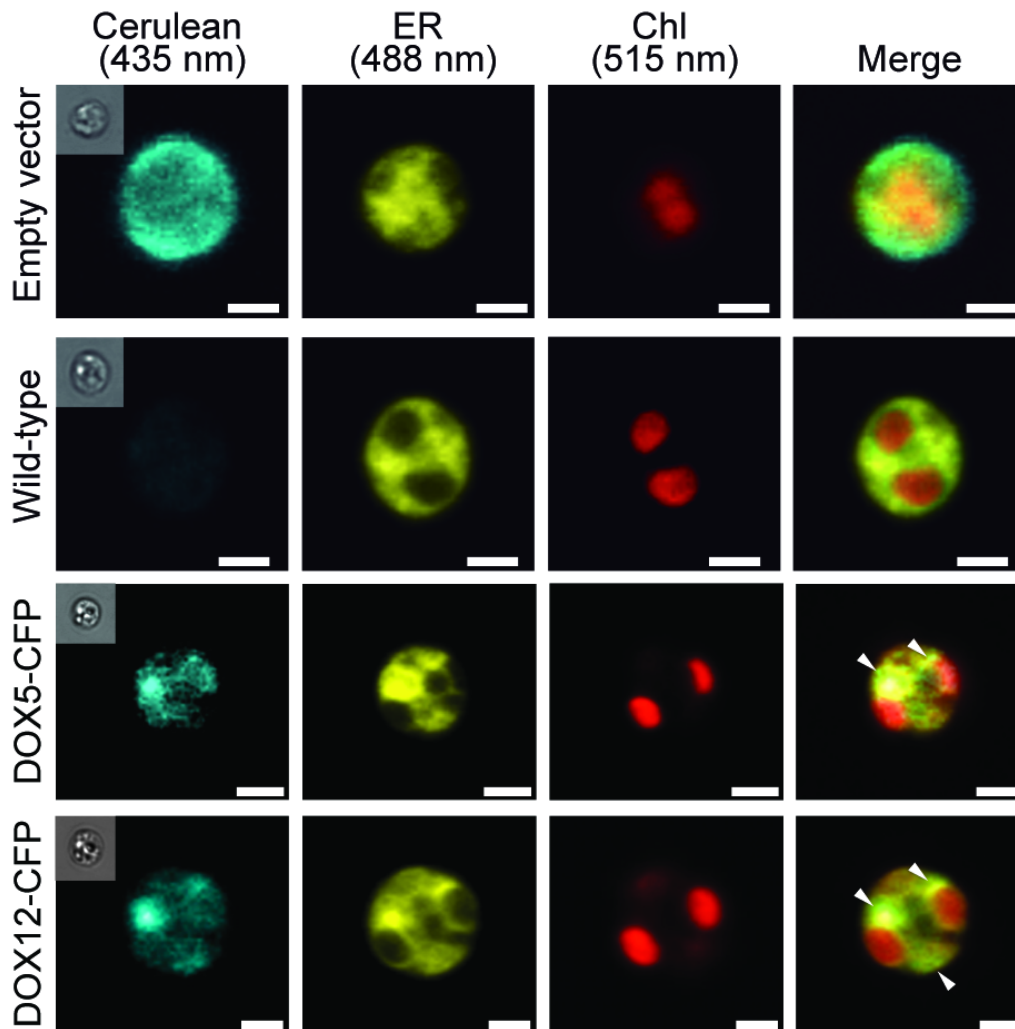
Figure 3.11. CLSM analysis of *N. oceanica* CCMP1779 wild type, and empty vector and CFP-desaturase overexpressing (DOX) transformants.
Cells were examined for the presence of cerulean fluorescence. Staining with an ER specific fluorescent dye and chlorophyll autofluorescence were compared to the cerulean signal. Arrowheads indicate co-localization of ER and cerulean fluorescence. Bars = 1.5 μm.
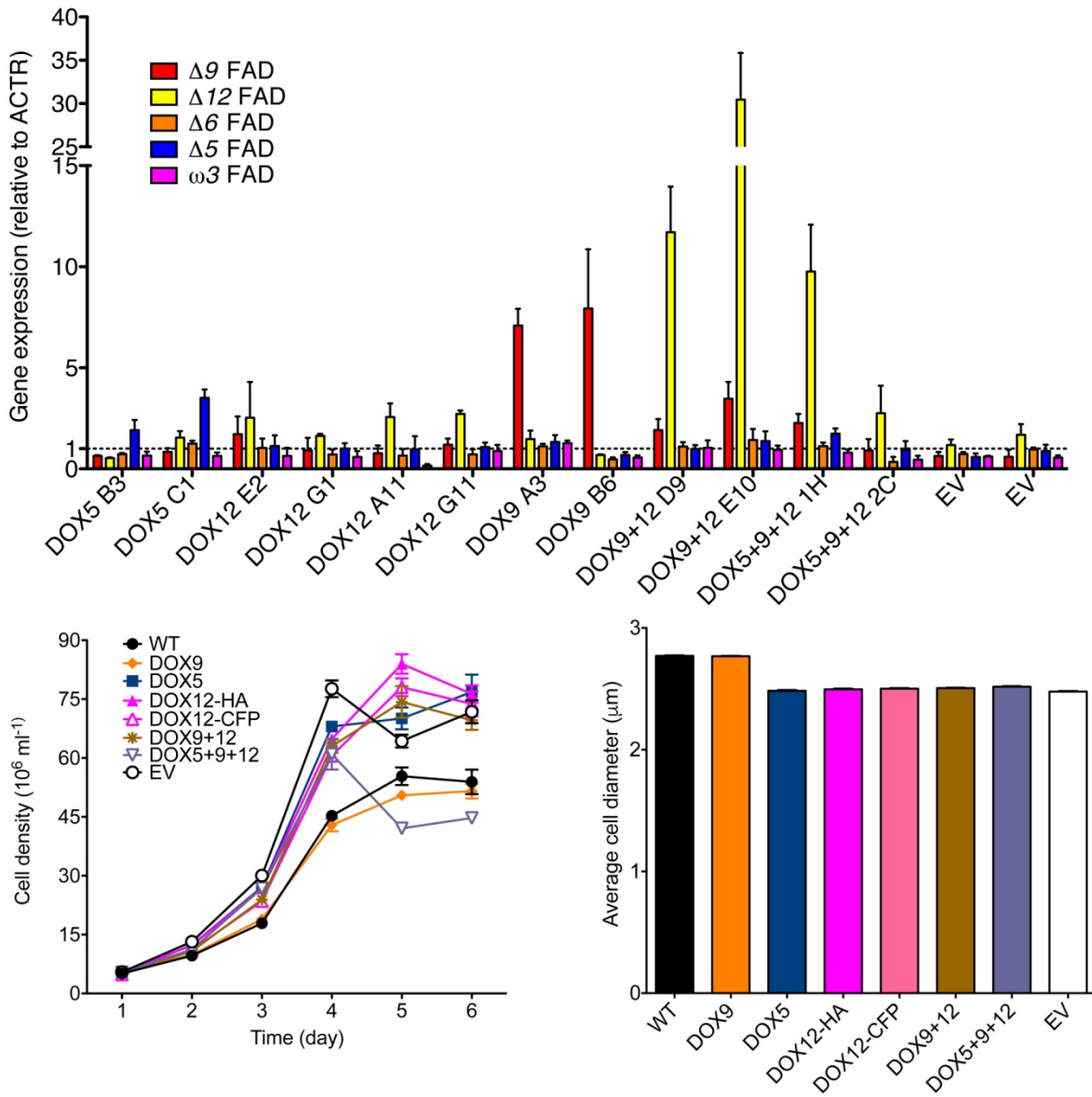
Figure 3.12. Desaturase overproduction alters the fatty acid profile of *N. oceanica* CCMP1779.
(a) Gene expression of FADs measured by qPCR using the *ACTR* gene as control (average ± SEM of 3 independent cultures). (b) Cell growth of DOX lines during five days under constant light (average ± SEM of 4 independent cultures). (c) Cell diameter of DOX lines (average ± SEM of 4 independent cultures). WT, wild-type; EV, empty vector control.
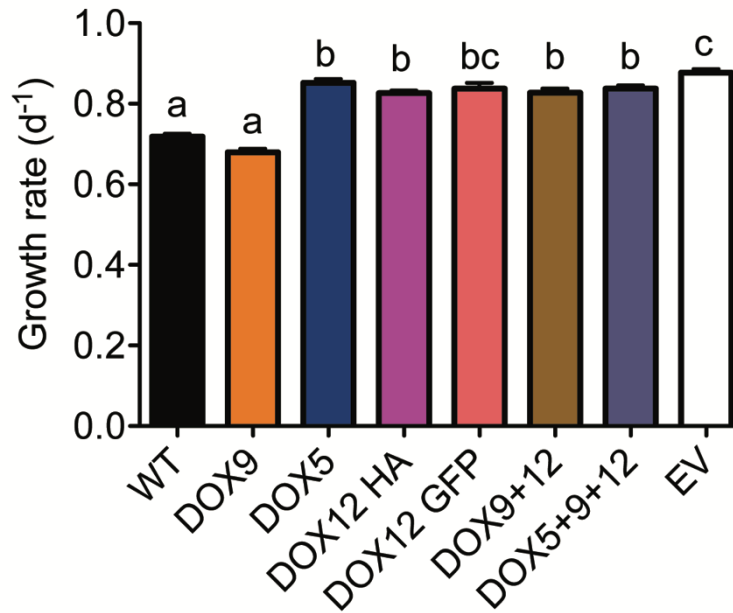
Figure 3.13. Growth rates of *N. oceanica* CCMP1779 DOX lines during exponential growth.
Growth rates were estimated using the first four days of the growth curves shown in Figure 3.12b (average +/- SEM, 8-12 independent cultures from 2-3 lines). Cell counts were natural log transformed and growth rates calculated using linear regression. Values statistically different are labeled with different letters ($p < 0.05$; ANOVA followed by Bonferroni's post hoc test).

Table 3.1. Fatty acid mole percentage of *S. cerevisiae* strains.
Yeast strains and timepoints with replicates are listed in the first column. Exogenous fatty acid when supplied is listed in the second column. Fatty acid mole percentage is determined by converting mass of each fatty acid class to moles, and normalizing the mole content of each fatty acid class against the total moles of all fatty acids (average ± standard error). Fatty acids not detected are represented by n.d..

Strain and timepoint	Substrate supplied	16:0	16:1	18:0	18:1	18:2	18:3ω6	20:3	20:4ω6	20:4ω3	20:5
Sc-LacZ 0hr (n=2)		28.71 +/- 0.0437	47.72 +/- 0.3081	5.64 +/- 0.0826	17.93 +/- 0.2692	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Sc-LacZ 24hr (n=2)		18.51 +/- 0.6455	52.74 +/- 0.7295	4.12 +/- 0.1101	24.62 +/- 0.1942	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Sc-LacZ 48hr (n=2)		19.66 +/- 0.0805	49.44 +/- 0.4685	4.77 +/- 0.0951	26.13 +/- 0.2928	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Sc-Δ12+Δ6 0hr (n=3)		25.56 +/- 1.2474	47.21 +/- 0.1305	5.7 +/- 0.083	21.54 +/- 1.1312	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Sc-Δ12+Δ6 24hr (n=3)		21.83 +/- 0.0178	47.36 +/- 0.0785	5.15 +/- 0.0314	19.86 +/- 0.1178	1.4 +/- 0.0517	4.41 +/- 0.092	n.d.	n.d.	n.d.	n.d.
Sc-Δ12+Δ6 48hr (n=3)		25.04 +/- 0.4096	40.62 +/- 0.3604	6.24 +/- 0.0899	22.17 +/- 0.1482	1.82 +/- 0.1425	4.1 +/- 0.1791	n.d.	n.d.	n.d.	n.d.
Sc-Δ12+Δ6+E6 0hr (n=2)		28.75 +/- 0.5478	48.56 +/- 1.8096	5.13 +/- 0.5659	17.56 +/- 0.696	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Sc-Δ12+Δ6+E6 24hr (n=2)		19.48 +/- 0.1286	50.8 +/- 0.4153	5.1 +/- 0.1362	21.17 +/- 0.4623	0.88 +/- 0.021	1.35 +/- 0.0309	1.23 +/- 0.0026	n.d.	n.d.	n.d.
Sc-Δ12+Δ6+E6 48hr (n=2)		22.7 +/- 0.5734	43.27 +/- 0.7372	6.42 +/- 0.2541	23.35 +/- 0.064	1.41 +/- 0.0504	1.31 +/- 0.0475	1.55 +/- 0.0292	n.d.	n.d.	n.d.
Sc-Δ12+Δ6+E6+Δ5+ω3 0hr (n=3)		25.0 +/- 1.2938	50.54 +/- 1.1552	4.8 +/- 0.2115	19.66 +/- 0.9521	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Sc-Δ12+Δ6+E6+Δ5+ω3 24hr (n=3)		18.55 +/- 0.1181	51.37 +/- 0.1766	4.99 +/- 0.0922	23.97 +/- 0.4621	0.39 +/- 0.2547	0.48 +/- 0.0572	0.25 +/- 0.0586	n.d.	n.d.	n.d.
Sc-Δ12+Δ6+E6+Δ5+ω3 48hr (n=3)		18.59 +/- 0.1782	45.77 +/- 0.5259	6.21 +/- 0.2244	26.33 +/- 0.2078	1.4 +/- 0.1521	0.74 +/- 0.0047	0.55 +/- 0.0371	n.d.	0.3 +/- 0.0013	0.11 +/- 0.0039

Table 3.1 (cont'd)

Sc-LacZ 0hr (n=2)	20:3	27.98 +/- 0.1938	30.08 +/- 2.9067	4.52 +/- 0.0903	9.85 +/- 1.3444	n.d.	n.d.	27.58 +/- 4.1476	n.d.	n.d.	n.d.
Sc-LacZ 24hr (n=2)	20:3	18.12 +/- 0.1678	47.28 +/- 0.29	3.88 +/- 0.0302	21.79 +/- 0.5658	n.d.	n.d.	8.93 +/- 0.7182	n.d.	n.d.	n.d.
Sc-LacZ 48hr (n=2)	20:3	22.14 +/- 0.0269	43.77 +/- 0.0804	4.57 +/- 0.0472	24.63 +/- 0.0782	n.d.	n.d.	4.89 +/- 0.0226	n.d.	n.d.	n.d.
Sc-Δ5+ω3 0hr (n=4)	20:3	26.84 +/- 0.6474	45.58 +/- 0.1301	5.24 +/- 0.0125	17.89 +/- 0.1804	n.d.	n.d.	4.44 +/- 0.3245	n.d.	n.d.	n.d.
Sc-Δ5+ω3 24hr (n=4)	20:3	19.58 +/- 0.133	47.27 +/- 0.2142	4.66 +/- 0.1392	22.95 +/- 0.6476	n.d.	n.d.	4.01 +/- 0.2516	n.d.	0.92 +/- 0.0854	0.62 +/- 0.1026
Sc-Δ5+ω3 48hr (n=4)	20:3	20.68 +/- 0.1245	42.18 +/- 0.4403	5.51 +/- 0.1087	26.85 +/- 0.2623	n.d.	n.d.	4.11 +/- 0.0245	n.d.	0.4 +/- 0.0171	0.26 +/- 0.0137
Sc-Leu 0hr (n=2)	20:3	21.62 +/- 0.3112	37.6 +/- 0.1856	5.58 +/- 0.0529	21.91 +/- 0.4474	n.d.	n.d.	13.09 +/- 0.3574	n.d.	0.19 +/- 0.0173	n.d.
Sc-Leu 24hr (n=2)	20:3	18.76 +/- 0.0367	42.33 +/- 0.8341	4.89 +/- 0.0294	22.53 +/- 0.1811	n.d.	n.d.	11.49 +/- 0.6604	n.d.	n.d.	n.d.
Sc-Leu 48hr (n=2)	20:3	18.47 +/- 0.0384	41.69 +/- 0.2594	5.25 +/- 0.2512	23.85 +/- 0.8267	n.d.	n.d.	10.73 +/- 0.78	n.d.	n.d.	n.d.
Sc-Leu 0hr (n=2)	20:4ω6	23.41 +/- 0.8118	35.01 +/- 0.4783	6.01 +/- 0.0155	22.29 +/- 0.9076	n.d.	n.d.	n.d.	13.09 +/- 0.5289	0.18 +/- 0.0297	n.d.
Sc-Leu 24hr (n=2)	20:4ω6	19.51 +/- 0.3211	40.15 +/- 0.8664	5.28 +/- 0.1969	25.04 +/- 0.1745	n.d.	n.d.	n.d.	9.95 +/- 0.5157	0.07 +/- 0.0072	n.d.
Sc-Leu 48hr (n=2)	20:4ω6	18.28 +/- 0.5734	40.87 +/- 0.4416	5.23 +/- 0.2064	26.08 +/- 0.038	n.d.	n.d.	n.d.	9.54 +/- 0.3003	n.d.	n.d.
Sc-Leu 0hr (n=1)	20:4ω3	21.75	27.02	6.66	19.34	n.d.	n.d.	n.d.	n.d.	25.24	n.d.
Sc-Leu 24hr (n=1)	20:4ω3	19.15	32.72	5.31	22	n.d.	n.d.	n.d.	n.d.	20.83	n.d.
Sc-Leu 48hr (n=1)	20:4ω3	17.22	32.96	5.94	23.66	n.d.	n.d.	n.d.	n.d.	20.21	n.d.
Sc-Δ5 0hr (n=3)	20:3	20.8 +/- 0.2822	35.4 +/- 0.161	5.76 +/- 0.0353	22.75 +/- 0.5027	n.d.	n.d.	15.04 +/- 0.2402	n.d.	0.24 +/- 0.1283	n.d.
Sc-Δ5 24hr (n=3)	20:3	17.98 +/- 0.0081	41.34 +/- 0.9836	5.12 +/- 0.0799	23.84 +/- 0.6283	n.d.	n.d.	10.36 +/- 1.5669	1.36 +/- 0.0818	n.d.	n.d.
Sc-Δ5 48hr (n=3)	20:3	17.99 +/-	39.71 +/-	5.72 +/-	24.89 +/-	n.d.	n.d.	10.39 +/-	1.3 +/-	n.d.	n.d.

Table 3.1 (cont'd)

		0.1227	0.9696	0.1621	0.8541			1.9407	0.0892		
Sc-Δ5 0hr (n=3)	20:4ω3	20.78 +/- 0.7178	26.12 +/- 0.1373	6.68 +/- 0.0833	19.78 +/- 0.7863	n.d.	n.d.	0.26 +/- 0.0348	n.d.	26.38 +/- 0.2179	n.d.
Sc-Δ5 24hr (n=3)	20:4ω3	20.51 +/- 0.1	29.52 +/- 0.2052	6.32 +/- 0.0839	21.22 +/- 0.1081	n.d.	n.d.	n.d.	n.d.	19.59 +/- 0.1188	2.83 +/- 0.0265
Sc-Δ5 48hr (n=3)	20:4ω3	16.87 +/- 0.124	32.74 +/- 0.3841	6.04 +/- 0.1991	25.39 +/- 0.4848	n.d.	n.d.	n.d.	n.d.	16.85 +/- 0.2848	2.11 +/- 0.024
Sc-Δ5 0hr (n=2)	20:3 + 20:4ω3	21.23 +/- 1.0009	24.82 +/- 0.0068	6.63 +/- 0.0876	18.63 +/- 1.1196	n.d.	n.d.	8.32 +/- 0.3738	n.d.	20.38 +/- 0.1606	n.d.
Sc-Δ5 24hr (n=2)	20:3 + 20:4ω3	20.53 +/- 0.1962	28.73 +/- 0.9679	7.26 +/- 0.9888	21.4 +/- 0.1983	n.d.	n.d.	2.42 +/- 0.1677	0.36 +/- 0.0236	17.03 +/- 0.1943	2.28 +/- 0.02
Sc-Δ5 48hr (n=2)	20:3 + 20:4ω3	16.88 +/- 0.2146	32.37 +/- 0.1615	5.85 +/- 0.051	24.73 +/- 0.1856	n.d.	n.d.	2.49 +/- 0.1484	0.25 +/- 0.0145	15.64 +/- 0.3764	1.78 +/- 0.0281
Sc-ω3 0hr (n=3)	20:3	20.85 +/- 0.1576	36.85 +/- 0.2822	5.78 +/- 0.0231	22.84 +/- 0.2797	n.d.	n.d.	13.52 +/- 0.4101	n.d.	0.15 +/- 0.1527	n.d.
Sc-ω3 24hr (n=3)	20:3	18.9 +/- 0.7494	43.99 +/- 0.3518	5.4 +/- 0.0276	22.54 +/- 1.13	n.d.	n.d.	6.79 +/- 0.4948	n.d.	2.38 +/- 0.2214	n.d.
Sc-ω3 48hr (n=3)	20:3	17.99 +/- 0.029	41.93 +/- 0.3393	6.14 +/- 0.1451	25.22 +/- 0.2175	n.d.	n.d.	6.78 +/- 0.1037	n.d.	1.94 +/- 0.022	n.d.
Sc-ω3 0hr (n=3)	20:4ω6	22.99 +/- 0.1278	34.62 +/- 0.163	6.24 +/- 0.0229	22.63 +/- 0.1817	n.d.	n.d.	0.23 +/- 0.0145	13.14 +/- 0.1967	0.14 +/- 0.003	n.d.
Sc-ω3 24hr (n=3)	20:4ω6	20.6 +/- 1.1558	39.72 +/- 0.6478	5.55 +/- 0.1178	22.02 +/- 1.479	n.d.	n.d.	n.d.	9.48 +/- 0.732	n.d.	2.64 +/- 0.2762
Sc-ω3 48hr (n=3)	20:4ω6	18.36 +/- 0.2632	39.42 +/- 0.4584	5.93 +/- 0.1357	25.09 +/- 0.3808	n.d.	n.d.	n.d.	9.11 +/- 0.1571	n.d.	2.09 +/- 0.1326
Sc-ω3 0hr (n=3)	20:3 + 20:4ω6	22.45 +/- 0.1346	32.8 +/- 0.075	6.09 +/- 0.004	21.55 +/- 0.1546	n.d.	n.d.	7.98 +/- 0.2614	8.97 +/- 0.2458	0.15 +/- 0.0043	n.d.
Sc-ω3 24hr (n=3)	20:3 + 20:4ω6	21.83 +/- 0.5563	37.86 +/- 0.5847	5.85 +/- 0.0551	20.74 +/- 0.5772	n.d.	n.d.	3.55 +/- 0.4969	7.25 +/- 0.8004	1.12 +/- 0.0685	1.81 +/- 0.2391
Sc-ω3 48hr (n=3)	20:3 + 20:4ω6	18.12 +/- 0.2614	38.12 +/- 0.2251	5.94 +/- 0.0163	25.48 +/- 0.1351	n.d.	n.d.	3.54 +/- 0.6561	6.88 +/- 0.2697	0.69 +/- 0.0909	1.24 +/- 0.0962

Table 3.2. Codon usage of *N. oceanica* CCMP1779.
DNA codons (and the corresponding amino acid) from the coding regions of *N. oceanica* were computated using the CUSP function of EMBOSS. Total usage of each codon is displayed in column 3. Frequency refers to the expected occurrence of each codon for a random 1000 amino acid protein. Fraction is the relative usage between synonymous codons for each amino acid.

Codon	AA	# in CDS	Frquency	Fraction
GCA	A	37448	18.577	0.195
GCC	A	63080	31.292	0.329
GCG	A	57333	28.441	0.299
GCT	A	33942	16.837	0.177
TGT	C	12131	6.018	0.432
TGC	C	15928	7.901	0.568
GAT	D	43756	21.706	0.441
GAC	D	55406	27.485	0.559
GAG	E	103188	51.188	0.694
GAA	E	45410	22.526	0.306
TTT	F	35522	17.621	0.5
TTC	F	35467	17.594	0.5
GGT	G	26541	13.166	0.15
GGG	G	62294	30.902	0.351
GGA	G	41466	20.57	0.234
GGC	G	47132	23.381	0.266
CAT	H	20110	9.976	0.448
CAC	H	24792	12.298	0.552
ATC	I	37199	18.453	0.478
ATA	I	10216	5.068	0.131
ATT	I	30401	15.081	0.391
AAG	K	75091	37.25	0.714
AAA	K	30136	14.949	0.286
CTT	L	22678	11.25	0.117
CTG	L	47550	23.588	0.245
CTA	L	11493	5.701	0.059
CTC	L	39175	19.433	0.202
TTA	L	12010	5.958	0.062
TTG	L	60902	30.211	0.314
ATG	M	51143	25.37	1
AAC	N	29344	14.557	0.52
AAT	N	27071	13.429	0.48
CCT	P	25477	12.638	0.258

Table 3.2 (cont'd)

CCG	P	25713	12.755	0.26
CCA	P	15181	7.531	0.154
CCC	P	32469	16.107	0.329
CAA	Q	28943	14.358	0.38
CAG	Q	47211	23.42	0.62
AGG	R	26519	13.155	0.209
AGA	R	11067	5.49	0.087
CGA	R	17006	8.436	0.134
CGC	R	26259	13.026	0.207
CGG	R	29741	14.753	0.234
CGT	R	16435	8.153	0.129
AGC	S	33287	16.513	0.233
AGT	S	19055	9.453	0.133
TCG	S	27308	13.547	0.191
TCA	S	12477	6.189	0.087
TCC	S	32748	16.245	0.229
TCT	S	17960	8.909	0.126
TAG	Stop	1602	0.795	0.348
TGA	Stop	1535	0.761	0.333
TAA	Stop	1473	0.731	0.32
ACT	T	16188	8.03	0.154
ACA	T	19127	9.488	0.182
ACG	T	37173	18.44	0.354
ACC	T	32493	16.119	0.31
GTA	V	16373	8.122	0.114
GTC	V	37871	18.787	0.264
GTG	V	70785	35.114	0.494
GTT	V	18173	9.015	0.127
TGG	W	24709	12.257	1
TAT	Y	19599	9.722	0.407
TAC	Y	28550	14.163	0.593

Table 3.3. Bidirectional gene pairs.
Diverging genes with suitable expression (column 2) based upon light:dark RNA-Seq are listed with their contig (column 1). Maximum, minimum, fold change, and Pearson's correlation based on the light:dark RNA-seq is displayed in columns 3-6. Functional annotation of genes are listed in column 7.

Contig	Gene number (NannoCCMP1779_#)	Light:Dark minimum expression (FPKM)	Light:Dark maximum expression (FPKM)	Fold change	Co-expression	Annotation
nanno_913	210	65	268	4.15	0.8	Histone
nanno_913	197	180	293	1.63		Histone
nanno_989	9669	162	362	2.24	0.94	Ribosmal 40S S15
nanno_989	9668	145	366	2.53		Ribosomal 60S l21
nanno_84	7418	10	476	45.88	-0.7	Ribosomal 60S component
nanno_84	7414	83	315	3.81		Transaldose
nanno_750	7642	421	1736	4.12	0.47	Ribosomal component
nanno_750	7640	306	548	1.79		Mitochondrial NADH oxireductase
nanno_1064	10564	448	2453	5.47	0.99	VCP4
nanno_1064	10562	829	4471	5.39		VCP5
nanno_4100	4699	173	6760	39.16	0.99	VCP1
nanno_4100	4698	507	9277	18.3		VCP2
Nanno_476	6455	160	232	1.44	0.13	Histone
Nanno_476	6451	281	568	2.02		Histone
Nanno_913	194	457	801	1.75	-0.25	Histone
Nanno_913	212	16	41	2.5		Histone
Nanno_1361	443	4	111	25.67	1	Nitrate transporter
Nanno_1361	438	8	858	107.93		Nitrate reductase

Table 3.4. Fatty acid mole percent of *N. oceanica* CCMP1779 strains.

Fatty acid mole percentage was determined by normalizing the mole content of each fatty acid class against the total moles of all fatty acids. Average ± SEM of 4 independent cultures. Bold indicates a significant difference with both empty vector (EV) controls (ANOVA and Bonferroni's post hoc test, $p<0.05$). $16:1^{\Delta9}$; $18:1^{\Delta9}$; $18:2^{\Delta9,\Delta12}$; $20:4^{\Delta5,\Delta8,\Delta11,\Delta14}$; $20:5^{\Delta5,\Delta8,\Delta11,\Delta14,\Delta17}$.

Line	14:00	16:00	16:01	18:00	18:01	18:02	20:04	20:05	LCPUFA
WT1	4.13 +/- 0.35	37.83 +/- 1.55	32.34 +/- 0.72	1.77 +/- 0.33	1.94 +/- 0.10	0.90 +/- 0.16	3.66 +/- 0.42	17.36 +/- 1.08	22.01 +/- 1.61
WT2	3.84 +/- 0.32	38.28 +/- 1.66	33.40 +/- 0.55	1.49 +/- 0.14	1.98 +/- 0.07	0.80 +/- 0.11	3.41 +/- 0.32	16.80 +/- 1.23	21.01 +/- 1.59
DOX5 B3	6.17 +/- 0.39	**34.35 +/- 1.49**	**27.87 +/- 0.14**	1.41 +/- 0.20	1.43 +/- 0.32	1.62 +/- 0.18	5.83 +/- 0.22	**21.27 +/- 1.27**	**28.79 +/- 1.69**
DOX5 C1	6.45 +/- 0.22	**34.33 +/- 1.36**	**26.43 +/- 0.51**	1.33 +/- 0.24	1.27 +/- 0.15	1.24 +/- 0.12	5.21 +/- 0.11	**23.52 +/- 1.60**	**30.21 +/- 1.86**
DOX12 A11	6.02 +/- 0.21	35.79 +/- 1.51	**26.43 +/- 0.21**	1.19 +/- 0.18	0.82 +/- 0.22	1.27 +/- 0.25	5.27 +/- 0.24	**23.04 +/- 1.11**	**29.76 +/- 1.62**
DOX12 G11	6.30 +/- 0.11	35.93 +/- 1.55	**27.03 +/- 0.44**	1.23 +/- 0.09	1.02 +/- 0.12	1.38 +/- 0.18	5.53 +/- 0.21	**21.27 +/- 1.28**	**28.50 +/- 1.73**
DOX9 A3	3.49 +/- 0.04	34.85 +/- 0.91	33.49 +/- 0.35	1.22 +/- 0.13	3.08 +/- 0.18	1.30 +/- 0.12	4.61 +/- 0.18	17.85 +/- 0.92	**23.87 +/- 1.21**
DOX9 B6	3.42 +/- 0.09	**33.65 +/- 1.02**	35.25 +/- 0.09	1.21 +/- 0.13	3.77 +/- 0.32	1.30 +/- 0.06	4.97 +/- 0.20	16.43 +/- 1.11	**22.70 +/- 1.30**
DOX9+12 9D	6.25 +/- 0.30	34.88 +/- 0.69	**26.76 +/- 0.34**	1.13 +/- 0.18	1.70 +/- 0.17	1.93 +/- 0.15	5.82 +/- 0.13	**21.53 +/- 0.62**	**29.29 +/- 0.76**
DOX9+12 10E	6.55 +/- 0.22	35.72 +/- 0.96	**26.39 +/- 0.18**	0.98 +/- 0.34	0.89 +/- 0.32	2.05 +/- 0.23	5.93 +/- 0.20	**21.51 +/- 0.96**	**29.48 +/- 1.33**
DOX5+9+12 1H	6.25 +/- 0.31	34.91 +/- 1.46	**27.61 +/- 0.49**	1.19 +/- 0.23	1.35 +/- 0.13	1.64 +/- 0.24	5.83 +/- 0.29	**21.11 +/- 1.03**	**28.71 +/- 1.63**
DOX5+9+12 2C	6.66 +/- 0.32	**33.43 +/- 1.40**	**27.04 +/- 0.13**	1.21 +/- 0.25	1.51 +/- 0.16	2.18 +/- 0.27	6.37 +/- 0.31	**21.47 +/- 0.92**	**30.15 +/- 1.46**
EV F1	3.94 +/- 0.13	38.09 +/- 1.21	35.69 +/- 0.21	1.46 +/- 0.12	2.06 +/- 0.05	0.59 +/- 0.06	2.93 +/- 0.19	15.24 +/- 0.90	18.76 +/- 1.15
EV G2	3.62 +/- 0.06	38.24 +/- 1.58	35.81 +/- 0.26	1.59 +/- 0.15	1.98 +/- 0.05	0.41 +/- 0.10	2.79 +/- 0.22	15.57 +/- 1.15	18.76 +/- 1.45

Table 3.5. Cellular fatty acid content of *N. oceanica* CCMP1779 strains.
Average ± SEM of 4 independent cultures. Bold indicates a significant difference with both empty vector (EV) controls (ANOVA and Bonferroni's post hoc test, $p < 0.05$). Fatty acids descriptions as in Table 3.4.

Line	14:0	16:0	16:1	18:0	18:1	18:2	20:4	20:5	Total
WT1	14.9 +/- 0.8	157.2 +/- 18.4	133.1 +/- 13.2	8.4 +/- 2.3	8.8 +/- 1.0	3.9 +/- 0.4	17.4 +/- 0.8	83.7 +/- 5.0	427.8 +/- 35.4
WT2	15.8 +/- 0.7	180.4 +/- 20.0	155.4 +/- 13.0	7.9 +/- 1.3	10.3 +/- 1.1	4.0 +/- 0.4	18.6 +/- 1.0	91.9 +/- 6.0	484.3 +/- 36.4
DOX5 B3	17.7 +/- 2.3	**110.3 +/- 10.7**	**88.5 +/- 7.4**	5.0 +/- 0.8	5.1 +/- 1.4	5.6 +/- 0.7	22.1 +/- 1.7	79.6 +/- 5.7	**334.0 +/- 26.1**
DOX5 C1	18.8 +/- 1.3	**112.7 +/- 8.1**	**86.0 +/- 4.8**	4.8 +/- 0.9	4.5 +/- 0.5	4.4 +/- 0.5	20.3 +/- 1.0	91.4 +/- 9.5	**344.0 +/- 20.8**
DOX12 A11	18.5 +/- 0.8	**123.9 +/- 7.1**	**90.7 +/- 2.6**	4.5 +/- 0.7	3.2 +/- 0.9	4.8 +/- 0.9	21.6 +/- 0.9	93.7 +/- 3.3	**361.6 +/- 8.6**
DOX12 G11	16.0 +/- 0.9	**103.6 +/- 10.6**	**76.8 +/- 5.4**	4.0 +/- 0.5	3.3 +/- 0.6	4.2 +/- 0.4	18.7 +/- 0.7	71.1 +/- 2.8	**298.7 +/- 19.7**
DOX9 A3	18.0 +/- 1.6	202.0 +/- 20.7	191.6 +/- 15.3	8.0 +/- 1.4	19.8 +/- 2.7	8.1 +/- 0.9	31.5 +/- 2.5	121.2 +/- 9.6	601.0 +/- 52.0
DOX9 B6	16.0 +/- 2.2	176.1 +/- 24.6	182.2 +/- 23.0	7.3 +/- 1.8	21.8 +/- 3.5	7.3 +/- 0.8	30.3 +/- 2.5	100.2 +/- 11.9	541.2 +/- 68.2
DOX9+12 9D	16.6 +/- 1.3	**104.1 +/- 7.0**	**79.2 +/- 4.7**	3.7 +/- 0.7	5.7 +/- 0.8	6.3 +/- 0.6	20.6 +/- 1.2	75.5 +/- 3.7	**311.7 +/- 17.7**
DOX9+12 10E	17.1 +/- 1.9	**105.6 +/- 12.6**	**77.0 +/- 7.9**	3.4 +/- 1.2	3.1 +/- 1.2	6.4 +/- 0.3	20.5 +/- 1.4	73.7 +/- 4.8	**306.9 +/- 29.8**
DOX5+9+12 1h	16.8 +/- 01.0	**107.1 +/- 12.5**	**83.2 +/- 6.5**	4.2 +/- 1.2	4.6 +/- 0.8	5.4 +/- 0.7	20.9 +/- 1.5	74.9 +/- 3.4	**317.4 +/- 25.6**
DOX5+9+12 2C	18.7 +/- 1.3	**97.9 +/- 11.6**	**81.8 +/- 8.3**	4.5 +/- 1.1	6.0 +/- 1.1	8.4 +/- 1.0	24.5 +/- 2.9	75.6 +/- 6.3	**317.3 +/- 32.0**
EV F1	17.3 +/- 0.7	188.2 +/- 6.6	174.8 +/- 1.7	8.0 +/- 0.6	11.2 +/- 0.3	3.2 +/- 0.3	17.2 +/- 1.1	88.7 +/- 5.0	508.6 +/- 4.5
EV G2	17.2 +/- 1.0	204.4 +/- 18.5	188.7 +/- 10.0	9.3 +/- 01.0	11.6 +/- 0.8	2.3 +/- 0.5	17.4 +/- 0.8	96.6 +/- 3.3	547.6 +/- 29.4

Table 3.6. Primers used in Chapter 3.

Primer sequences are provided 5' to 3'. Restriction sites used in cloning are underlined, while restriction sites in the primer are listed in column 3. Small features that were introduced into constructs by the primer are listed in column 4.

Primer name	Primer sequence	Restriction sites	Features incorporated
cDNA isolation			
D12 10636 F+	Atgggacgcggcggtgag		
D12 10636 R-	cggccctctagaggtggtgatggtgatgatgtgcccgctgcttgtagaatac	XbaI	His tag
D6 2179 F+	Acactggcggccgcaaaaaaatgtctggacgcggtggcgagcgg	NotI	S. cerevisiae kozak sequence
D6 2179 R-	cggccctctagaggtggtgatggtgatgatgcatggcggggaaatcggcc	XbaI	His tag
D5 5794 F+	Gcctccccagaacgac		
D5 5794 R-	Ctaacccatgtgcacctcc		
W3 6416 F+	Atggttgagcaaacgttaccg		
W3 126 R-	Cggaggggatgatgaacg		
E6 9141 F+	Gccgccgcccttcttg		
E6 9141 R-	Gccggtttccaagaaggct		
Yeast expression			
D12 Sc F+	Aaccccggatccaaaaaatgtctggacgcggcggtgag	BamHI	S. cerevisiae kozak sequence
D12 Sc R-	cggccctctagaggtggtgatggtgatgatgtgcccgctgcttgtagaatac	XbaI	His tag
E6 9141 ScTopo F+	Atgtctgccgccgcccttcttg		
E6 9141 ScTopo R-	Gccggtttccaagaaggct		
D5 Sc F+	Gcggccgcaaaaaaatgtctcctcccagaacgac	NotI	S. cerevisiae kozak sequence
D5 Sc R-	Gagctcctagtggtgatggtgatgatgacccatgtgcacctcc	SacI	His tag
W3 Sc F+	Gggcccaaaaaaatgtctgttgagcaaacgttaccgac	ApaI	S. cerevisiae kozak sequence
W3 Sc R-	cggccctctagaggtggtgatggtgatgatgcggaggggatgatgaacg	XbaI	His tag

Table 3.6 (cont'd)

Toolkit epitopes			
OX Hafillin+	cgcgccctcgttaacatgtacccctatgacgtgccggactacgccacgcgtgagtcgcgatgagct	HpaI	HA epitope
OX HA fillin -	catcgcgactcacgcgtggcgtagtccggcacgtcataggggtacatgttaacgaggg	MluI, NruI	HA epitope
venus HM GSG OX F+	Gttaaccaattggggagcggcatggtgagcaagggcgag	HpaI, MfeI	Glycine-serine-glycine
venus GSG SM OX R-	Acgcgtaggcctgccgcttcctttgtacaactcatccatcccaagc	StuI, MluI	Glycine-serine-glycine
Cerulean HM GSG OX F+	Gttaaccaattggggagcggcatggtgtccaagggcgag	HpaI, MfeI	Glycine-serine-glycine
Cerulean GSG SM OX R-	Acgcgtaggcctgccgcttcccttgtacagctcgtccatgcc	StuI, MluI	Glycine-serine-glycine
nocox EGFP mfe gsg F+	Caattggggagcggcatggtgagcaagggcgag	MfeI	Glycine-serine-glycine
nocox EGFP gsg stui R-	Aggcctgccgcttcccttgtacagctcgtccatgc	StuI	Glycine-serine-glycine
nocox Nlux mfe gsg F+	Caattggggagcggcatggtgtttactctcgaggacttc	MfeI	Glycine-serine-glycine
nocox Nlux gsg stui R-	Aggcctgccgcttcccgcgtagtcgggcacgtc	StuI	Glycine-serine-glycine
Nannochloropsis promoters			
EFbidi asci T clai F+	Ccatccacagaatcgattggcgcgccctcgtt	AscI, ClaI	
EFbidi xho R-	Atggatctcgagggttgcgtgtgtatc	XhoI	
EF bidi pro ecori clai F+	Gaattcatcgattctgtggatggagggagg	EcorI, ClaI	
EF bidi pro xhoi R-	Ctcgagggttgcgtgtgtatctgtg	XhoI	
Efpro GW F+	cacctatagctacatggtagctag,		
Efpro GW R-	Tgttacgaagtgagggttgag		

130

Table 3.6 (cont'd)

NoEF1a	Cttggccatggtcagctg<u>ctcgag</u>tgttacgaagtgagggttg	XhoI	
NoEF1a	Cctttttacggttcctggcc<u>tctagag</u>tatagctacatggtagc	XbaI	
Ribi sdm nru F+	Gaggagaggtcgtgagaacaggatg		
Ribi sdm nru R-	Ttcctggccaccatcgcttc		
Ribi sdm mlui F+	Tccccatcaggcagcagac		
Ribi sdm mlui R-	Gtctagctgtgacactgaggac		
Desaturase overexpression cloning			
D9 mlui HA start :	<u>acgcgt</u>tatccgtacgatgtccctgactatgcgatggtcttccagctcgc	MluI	HA epitope
D9 ecorv nsi R-	<u>Atgcat</u>g<u>atatc</u>cagatttccagtccttggtc	EcorV, NsiI	
D12 claI xhoi F+	<u>Ctcgag</u>at<u>cgat</u>gggacgcggcggtgag	XhoI, ClaI	
D12 nrui R+	<u>Tcgcga</u>ggcgtagtccggcacgtc	NruI	
D5 OX mluI F+	<u>Acgcgt</u>atgcctccccagaacgacgctg	MluI	
D5 OX nrui R-	<u>Tcgcga</u>acccatgtgcacctccgc	NruI	
D6 OX mluI F+	<u>Acgcgt</u>atgggacgcggtggcgagc	MluI	
D6 OX nrui R-	<u>Tcgcga</u>catggcggggaaatcggc	NruI	
D12 OX mluI F+	<u>Acgcgt</u>atgggacgcggcggtgagaag	MluI	
D12 OX nrui R-	<u>Tcgcga</u>tgcccgctgcttgtagaatacc	NruI	
w3 OX mluI F+	<u>Acgcgt</u>atggttgagcaaacgttaccg	MluI	
w3 OX nrui R-	<u>Tcgcga</u>cggaggggatgatgaacg	NruI	
2A site directed mutagenesis			
F2A 58 F+	Ccggcgcgcccgtgaccgagttgctttaccg	AscI	
F2A 58 R-	Gcgttctagagggcccccgggttggactc	PspomI, XbaI	

131

Table 3.6 (cont'd)

T2A F+	Ccttgctgacgtgcgggggacgtggaggagaac		
T2A R-	Aaccccgcccctctccactagtgccgcatgc	SphI	
401 p2A F+	gtccgagggcaaaggaagcatgcggcactagtggagcgacgaacttctccttgctga	SphI, SpeI	
4-1 p2A R-	gagctcacgtacgcgttctagagggccccgggttctcctccacgtccccggcctgct	SacI, MluI, XbaI	
4-1 p2A R+	agcaggccggggacgtggaggagaacccggggccctctagaacgcgtacgtgagctc	NheI	
401 F2A F+:	gtccgagggcaaaggaaggctccggcgcgcccgtgaagcagacgctgaacttcgacctgc	AscI	
401 p2A F-	tcagcaaggagaagttcgtcgctccactagtgccgcatgcttcctttgccctcggac	SphI, SpeI	
4-1 p2A R+	agcaggccggggacgtggaggagaacccggggccctctagaacgcgtacgtgagctc	SacI, MluI, XbaI	
401 F2A F-:	gcaggtcgaagttcagcgtctgcttcacgggcgcgccggagccttcctttgccctcggac	AscI	
P2a 30 aa F+	Ttccaggggccgggggcgacgaacttctccttg		
P2a 30 aa R+	Gcatgcggcactagtggaatgacgaccctgtcc	SphI	
p2a 45aa sdm F+	cgcagcgaacgcctgagcaccgcg atgacgaccctgtccttcca	SphI, SpeI	
p2a 45aa sdm R-	Gccatcgctgttttccggatggcttccactagtgccgcatgc	SphI, SpeI	
p2a 60aa sdm F+	tgatccggcggcgtttagcaaa agccatccggaaaacagcg	SphI, SpeI	
p2a 60aa sdm R-	Tgaatgccgggggcgcgggcaaaatccactagtgccgcatgc	SphI, SpeI	
N. oceanica qPCR			
ActRP4 1845 qPCR F2+	Gatgtggaggataacgaccc		
ActRP4 1845 qPCR R2-	Aactctttccgcacgtccac		
D9 qpcr F2+	Agcagttcaacccgaccaag		
D9 qpcr R2-	Tcatgctcttcccattcgcc		
D12 qpcr F2+	Caagatgccgttctaccacg		
D12 qpcr R2-	Gcccgctgcttgtagaatac		
D6 qpcr F2+	Gttggtgcagtggttctgtg		
D6 qpcr R2-	Ccctcccacatattcgtctc		

Table 3.6 (cont'd)

D5 qpcr F2+	Ggactacgcccacaataacg		
D5 qpcr R2-	Gtcacaaaatcgggcaggac		
w3 qpcr F2+	Gctctcctatgtggattggg		
w3 qpcr R2-	Gttggtcttccttctcggac		

Table 3.7. Constructs generated in Chapter 3.
Plasmids generated for yeast and *Nannochloropsis* expression. Restriction sites within the same multiple MCS are separated by slashes and different multiple cloning sites (MCS) separated by commas. Addgene IDs when applicable are listed in column 4.

Construct	Host selection	MCS	Addgene ID
Yeast expression vectors			
pESC-leu-D5+w3	leucine	-	98121
pESC-his-D12+D6	histidine	-	98120
pYES-E6	uracil	-	98117
Nannochloropsis toolkit			
Selection casettes			
pNOC-411	zeocin	NheI/MluI/SacI	98853
pNOC-401	hygromycin	NheI/MluI/SacI	98852
2A peptides			
pNOC-411-F2A(24)	zeocin	XbaI/MluI/SacI/MfeI	
pNOC-411-T2A(18)	zeocin	XbaI/MluI/SacI/MfeI	
pNOC-411-P2A(19)	zeocin	XbaI/MluI/SacI/MfeI	
pNOC-411-P2A(30)	zeocin	XbaI/MluI/SacI/MfeI	
pNOC-411-P2A(45)	zeocin	XbaI/MluI/SacI/MfeI	
pNOC-411-P2A(60)	zeocin	XbaI/MluI/SacI/MfeI	98843
pNOC-411-F2A(58)	zeocin	XbaI/MluI/SacI/MfeI	
pNOC-401-P2A(19)	hygromycin	XbaI/MluI/SacI/MfeI	
pNOC-401-P2A(29)	hygromycin	XbaI/MluI/SacI/MfeI	
pNOC-401-P2A(45)	hygromycin	XbaI/MluI/SacI/MfeI	
pNOC-401-P2A(60)	hygromycin	XbaI/MluI/SacI/MfeI	98850
pNOC-401-P2A(60)-Hsterm	hygromycin	PspomI/XbaI/AfeI/KpnI	98859
pNOC-411-P2A(60)-Hsterm	zeocin	PspomI/XbaI/AfeI/KpnI	98860
pNOC-411-F2A(24)-Flux	zeocin	XbaI/MluI/SacI/MfeI	
pNOC-411-T2A(18)-Flux	zeocin	XbaI/MluI/SacI/MfeI	
pNOC-411-P2A(19)-Flux	zeocin	XbaI/MluI/SacI/MfeI	
pNOC-411-P2A(29)-Flux	zeocin	XbaI/MluI/SacI/MfeI	

Table 3.7 (cont'd)

pNOC-411-P2A(45)-Flux	zeocin	XbaI/MluI/SacI/MfeI	
pNOC-411-P2A(60)-Flux	zeocin	XbaI/MluI/SacI/MfeI	98844
pNOC-411-F2A(58)-Flux	zeocin	XbaI/MluI/SacI/MfeI	
pNOC-401-P2A(19)-Flux	hygromycin	XbaI/MluI/SacI/MfeI	
pNOC-401-P2A(29)-Flux	hygromycin	XbaI/MluI/SacI/MfeI	
pNOC-401-P2A(45)-Flux	hygromycin	XbaI/MluI/SacI/MfeI	
pNOC-401-P2A(60)-Flux	hygromycin	XbaI/MluI/SacI/MfeI	98851
Epitopes			
pNOC-Dlux	hygromycin	-	
pNOC-Efpro-Dlux	hygromycin	-	
pNOC-OX-HA	hygromycin	AscI/HpaI,MluI/NruI/SacI	98114
pNOC-OX-CFP	hygromycin	AscI/HpaI/MfeI, StuI/MluI/NruI/SacI	98111
pNOC-OX-YFP	hygromycin	AscI/HpaI/MfeI, StuI/MluI/NruI/SacI	98112
pNOC-OX-GFP	hygromycin	AscI/HpaI/MfeI, StuI/MluI/NruI/SacI	98110
pNOC-OX-Nlux	hygromycin	AscI/HpaI/MfeI, StuI/MluI/NruI/SacI	98113
pNOC-stacked-HygR-Nlux	hygromycin	AscI/HpaI/MfeI, StuI/MluI/NruI/SacI, PspomI/XbaI/AfeI/KpnI	98842
pNOC-stacked-BleR-GFP	zeocin	AscI/HpaI/MfeI, StuI/MluI/NruI/SacI, PspomI/XbaI/AfeI/KpnI	98116
Desaturase overexpression			
pNOC-DOX5-CFP	hygromycin	-	98869
pNOC-DOX12-CFP	hygromycin	-	98870
pNOC-DOX12-HA	hygromycin	-	98871
pNOC-DOX9	zeocin	-	98873
pNOC-stacked-DOX9+12	zeocin	-	98872

Appendix 3.2. Chapter 3 datasets.

Supplemental datasets are available online from:
http://onlinelibrary.wiley.com/doi/10.1111/pbi.12772/full

Data 3.1 Script used for the identification of bidirectional promoters.

Data 3.2 Description of vector construction and sequences of vectors generated in this study.

Chapter 4. Non-transgenic marker-free gene disruption by an episomal CRISPR system in the oleaginous microalga, *Nannochloropsis oceanica* CCMP1779

ABSTRACT

Utilization of microalgae has been hampered by limited tools for creating loss-of-function mutants. Furthermore, modified strains for deployment into the field must be free of antibiotic resistance genes and face significantly less regulatory hurdles if they are transgene free. The oleaginous microalga, *Nannochloropsis oceanica* CCMP1779, is an emerging model for microalga lipid metabolism. We present a one-vector episomal CRISPR/Cas9 system for *N. oceanica* that enables the generation of marker-free mutant lines. The CEN/ARS6 region from *Saccharomyces cerevisiae* was included in the vector to facilitate its maintenance as circular extrachromosal DNA. The vector utilizes a bidirectional promoter to produce both Cas9 and a ribozyme flanked sgRNA. This system efficiently generates targeted mutations, and allows the loss of episomal DNA after the removal of selection pressure, resulting in marker-free non-transgenic engineered lines. To test this system, we disrupted the nitrate reductase gene (*NR*) and subsequently removed the CRISPR episome to generate non-transgenic marker-free nitrate reductase knockout lines (NR-KO).

INTRODUCTION

Microalgae are some of the most productive biomass sources on Earth and can generate many high-value bioproducts such as omega-3 fatty acids, carotenoids, and unusual polysaccharides (*5*). One genus of microalgae that has distinguished itself for productivity and genetic engineering potential is *Nannochloropsis* (*11, 15, 254*). Genetic engineering of microalgae has been hampered by a limited ability to disrupt genes and by the challenges of generating transformed algae that meet biocontainment requirements for open pond production (*255*).

Recently, the use of cisgenic elements, marker-free strategies, and DNA free genome editing have allowed the generation of modified organisms that lack antibiotic selection markers (*89, 125, 256*) or even any heterologous DNA (*112, 257*). Such organisms satisfy the criteria for non-transgenic organisms, making deployment of modified strains into open systems feasible.

Gene targeting in microalgae has been performed by homologous recombination (*14*), TALENs (*121, 126, 258*), and CRISPR/Cas9 (*17, 27, 109, 112, 259-261*). So far, gene disruption by these methods in microalgae has required the integration of a transgenic selection marker into the genome. A recently described CRISPR/Cas9 system for *N. oceanica* IMET had low efficiency (~1-4%), possibly due to low targeting efficiency of the single guide RNA (sgRNA) produced from a protein coding gene promoter (*27*). Another CRISPR/Cas9 system for *N. gaditana* utilized exogenously synthesized sgRNAs and required two selection markers for targeted disruption (*17*). Efficient CRISPR/Cas9 based mutagenesis requires a sgRNA without modified ends or extraneous sequences for efficient interaction with the Cas9 nuclease and DNA target. Therefore,sgRNAs flanked by the hammerhead (HH) and hepatitis delta virus (HDV) self-cleaving ribozymes are used to generate precise 5' and 3' ends (*257, 262*), and enable the production of sgRNAs under the control of protein-coding gene promoters (transcribed by RNA polymerase II).

Episomes (extrachromosomal nuclear DNA that is usually circular) occur naturally in some red algae (*263*), and diatoms are capable of maintaining synthetic circular episomes containing a low GC (guanine cytosine) centromere-like region (*40, 42, 85*). In diatoms, an extrachromosomal episomewith a centromere and autonomous replication sequence fusion developed from *Saccharomyces cerevisiae* (CEN/ARS) is replicated and segregated into daughter cells (*40, 42*). Furthermore, this episomal system is able to produce proteins that can be

localized throughout the cell and results in more uniform transgene expression levels between independent transformantscompared to integrated vectors, possibly by avoiding integration site-specific effects (*42*). While circular episomes are an effective expression platform in diatoms, without selection pressure, they are lost over several generations providing a way to "cure" cells of them (*85*).

We developed a CRISPR/Cas9 system that can be expressed from an episome in the heterokont microalgae *Nannochloropsis oceanica* CCMP1779. The subsequent removal of the episome after the mutation was produced generates marker-free non-transgenic gene disruption mutants that can be modified repeatedly.

RESULTS

Development of a CRISPR system for *N. oceanica*

The linearized pNOC-CRISPR-GFP vector, containing a Cas9-GFP expression cassette and a hygromycin resistance cassette (*11*), generated Hygromycin B resistant transformants when introduced into *N. oceanica* (Figure 4.1A). To establish a highly efficient CRISPR/Cas9 gene editing system in *N. oceanica* we confirmed the nuclear localization of a Cas9-GFP nuclease with C' and N' terminal SV40 nuclear localization signals (NLS) by confocal microscopy. Cas9-GFP proteins co-localized with a DNA stain (4', 6-Diamidino-2-Phenylindole, DAPI), demonstrating their nuclear localization in *N. oceanica* (Figure 4.1B, Figure 4.2). To facilitate Cas9 detection *in vivo*, we then generated a Cas9 nuclease fused to the ultra-bright NanoLuciferase reporter protein with an HA tag (Nlux-HA) in the vector pNOC-CRISPR. Nlux has high luminescence activity (*97*), and accordingly *N. oceanica* transformants can be screened for the presence of Cas9-Nlux-HA fusion in a 96-well plate using a small number of cells (*15*).

Immunoblotting confirmed the production of Cas9-Nlux-HA (186.2 kDa) and Cas9-GFP

proteins (191.8 kDa) (Figure 4.1C-D) in transgenic integrated empty-vector CRISPR *N. oceanica*

lines (iEV).

The pNOC-CRISPR vector contains the *Cas9-Nlux-HA* and a gRNA scaffold with a 3'

fusion to the HDV ribozyme under the control of the Ribi bidirectional promoter (*257, 262*). To

facilitate cleavage precisely at the 5' end of a sgRNA sequence, a hammerhead ribozyme (HH)

specific to the guide sequence is introduced during sgRNA generation (Figure 4.3). Strategies for

cloning sgRNAs are described in the Methods section.

Use of an episomal CRISPR system to disrupt the nitrate reductase gene

We next designed an episome-based CRISPR/Cas9 system, by incorporating the CEN/ARS

region from *S. cerevisiae* into the pNOC-CRISPR-Nlux vector, generating the vector pNOC-

ARS-CRISPR (Figure 4.4A). To test this system, we targeted the nitrate reductase gene (*NR*) for

the known phenotype of NR disruption lines (NR-KO) (Figure 4.4A) (*125, 127, 264*).

Nannochloropsis NR-KO lines can grow on ammonium (NH_4) but are not able to sustain growth

when provided with only nitrate (NO_3) as a nitrogen source (*14, 27*). Two sgRNAs flanked by

self-cleaving ribozymes (sgNR1 and sgNR2) were used to independently target two sites at the

5' end of the *NR* gene. We transformed the pNOC-ARS-CRISPR-sgNR1 and pNOC-ARS-

CRISPR-sgNR2 episomes into *N. oceanica* and generated the NR1-KO and NR2-KO lines

respectively. Introduction of the circular episomal CRISPR plasmids into *N. oceanica* resulted in

Hygromycin B resistant lines, which were screened for Cas9-Nlux-HA signal (Figure 4.5A-B).

Lines displaying luminescence were examined for mutations in the *NR* gene by PCR

amplification of the genomic locus and Sanger sequencing (Figure 4.5). We found that 47% of

lines with Nlux signal had frame-shift mutations and 13% had in-frame mutations, while 10% had no mutation and 30% resulted in low-quality sequencing outcomes (Figure 4.5). The low-quality sequencing reactions often had good quality chromatographs up to the target site and then became unreadable. This observation suggests that a heterogenous population of mutants was recovered from single colonies, possibly due to mutagenesis occurring after plating. We selected *N. oceanica* mutants with small deletions (1D, 1H), a larger (47 bp) deletion (4G), and a small insertion (B12) as frame-shifted null mutants (NR-KO lines), and the B7 line with a 3 bp deletion as an in-frame mutant (NR-IF) for further analyses (Figure 4.4B, Figure 4.5D-E). Immunoblotting detected the Cas9-Nlux-HA protein in all the lines indicating that proteins can be produced from an episomal DNA in *N. oceanica* (Figure 4.4C). To test for the loss of NR activity, cell growth was analyzed in liquid cultures after transfer from NH_4 to NO_3 containing medium. The frame-shifted NR-KO mutants could not grow, while the wild type (WT), the iEV lines, and the in-frame mutant B7 grew in NO_3 containing liquid medium (Figure 4.4D). This indicates frame-shifts introduced into the coding sequence of targeted genes ablates the function of the resulting protein, while in-frame mutations may still produce functional proteins.

To determine if the episomes were maintained as circular DNA, we conducted an episome rescue experiment by transforming *Escherichia coli* with DNA isolated from *N. oceanica* episome carrying lines (NR-KOs). WT and iEV lines were used as negative controls. *E. coli* transformants were only obtained from untreated DNA isolated from NR-KO lines, but not from untreated DNA of iEV lines, or the WT strain (Figure 4.4E). To further confirm that the plasmids rescued from *N. oceanica* NR-KO lines were circular, an equal amount of DNA from the NR-KO lines was subjected to endonuclease ClaI restriction digest and/or treated with Exonuclease V, which acts on free DNA ends *41*). Due to the ClaI site in the pNOC-ARS-

CRISPR plasmid, ClaI treatment of DNA isolated from NR-KO lines strongly decreased the number of *E. coli* transformants (Figure 4.4E). Although treatment with Exonuclease V resulted in a reduction of *E. coli* transfomants compared to mock treatments, a combined endonuclease and exonuclease treatment resulted in no colonies being recovered (Figure 4.4E). Restriction fragment analysis of plasmids recovered from the episome rescue indicated that the episomes were faithfully maintained in *N. oceanica* (Figure 4.6). Restriction enzyme recognition sites by the CEN/ARS region (AgeI), in the *Cas9* and *HygR* genes (EcoRI), and backbone (NotI), indicated that the episome had no apparent insertions or deletions (Figure 4.6A-B). Sanger sequencing of the recovered episomes further confirmed their authenticity (Figure 4.6C-D).

We performed southern blots to further characterize the CRISPR episomal lines (Figure 4.7). As a control, we used iEV lines containing the genomic integrated pNOC-CRISPR vector. DNA was isolated from NR-KO, iEV and WT cells and was then digested with SacI and hybridized with either a HygR or an AmpR gene probe (Figure 4.7A). In all the NR-KO lines tested, both probes generated one fragment of ~13,000 bp that matched the episome CRISPR vector size (Figure 4.7B). However, the iEV lines generated different size fragments when using the HygR probe and only one line had a detectable fragment when the AmpR probe was used (Figure 4.7B). The AmpR probe hybridizes close to the AseI cut site used for linearizing the pNOC-CRISPR vector before transformation to generate the iEV lines. Therefore, the absence of a band in AmpR probed iEV DNA indicates that the AmpR was partially lost during integration or that the fragment is too small to be detected in the current experiment. None of the WT lines had a detectable signal in these assays. This experiment further confirmed the presence of a circular DNA molecule in the CRISPR episomal lines.

Removal of the CRISPR episomes from *N. oceanica* mutants

After generating the NR-KO lines we sought to remove the episome to form marker-less non-transgenic mutants. Nlux signal is a convenient measure for Cas9 presence in the transformed lines and to monitor for the loss of the episomes. Four NR-KO lines were grown for 10 days without selection in liquid medium and plated on solid medium to generate independent colonies. The colonies were screened for Nlux luminescence, followed by a PCR test for the presence of the episome and a control genomic locus (Figure 4.8A). After growth in non-selective medium, the NR-KO lines had various levels of Nlux signal and 25-75% of the lines had low luminescence signal (Figure 4.8B). To confirm that these cured low Nlux luminescent lines had lost the episome, we conducted PCR for the *Cas9* gene contained in the episome, and the *NR* gene as a genomic positive control (Figure 4.9A-B). PCR detected the *NR* gene in all lines, while the Cas9 coding sequence was only present in the episomal lines but not in the cured lines (Figure 4.9B). Furthermore, growth tests were used to demonstrate the different phenotypes of the generated lines. First, the cured lines that had lost the episome were unable to grow on Hygromycin B containing medium, while the episome carrying parental lines were Hygromycin B resistant (Figure 4.9C). Second, growth on solid medium containing either NO_3 or NH_4 revealed a chlorotic phenotype of the episome carrying and cured NR-KO lines only when grown on NO_3 (Figure 4.9C) (*14*).

DISCUSSION

We have developed a system to produce marker-less non-transgenic gene knockout strains that not only allows for the generation of lines with multiple modifications but also facilitates the transfer of strains developed in the lab to the field. Our CRISPR/Cas9 components have been optimized for efficient detection, with a Cas9-Nlux-HA reporter fusion that has high signal to

noise ratio and can be screened in a 96-well plate format soon after transformant isolation. Furthermore, the Cas9-Nlux-HA and sgRNAare coregulated by a bidirectional promoter, thus linking Cas9 production to sgRNA expression (Figure 4.1A) (*15*). The pNOC-CRISPR vector series contains unique restriction sites between the elements making it an ideal platform for further development of CRISPR/Cas9 as a system for transcriptional reprogramming and other DNA targeting techniques.

The use of self-cleaving ribozymes facilitates the flexible expression sgRNAs. The ribozyme flanked sgRNAs can be expressed from different types of promoters, in our case a bidirectional promoter, but also possibly from conditional promoters. Our final episomal vector, pNOC-ARS-CRISPR-v2, contains a multi-cloning site with a pair of BspQI type IIS restriction sites for scarless insertion of a guide sequence, for one-step cloning (Figure 4.10A-B).Self-cleaving sequences for expression of the sgRNA could be used for the production of multiple sgRNAs from a single transcript (*22*). Production of multiple sgRNAs can be used to target multiple genes for simultaneous disruption, used in pairs for deletion of a region between two targeted sites on a chromosome, or in pairs along with a nickase Cas9 variant for enhanced specificity.

Transgenic expression in algae is most often performed by integrating a constructed DNA (construct) into the genome. However, this can result in site-specific effects, or the disruption of genes at the insertion site. Therefore, episomes are an emerging expression platform in microalgae that avoids these issues. In diatoms, effective centromere sequences have the simple requirements of a GC content <30%, a length of over >500 bp, with long stretches of A-T sequences (*40, 42, 85*) . These centromere-like regions interact with the centromeric histone protein (CENH3) likely facilitating segregation during cell division (*40*). However, the

molecular processes for replication of synthetic episomes in algae are yet to be determined. While these first-generation synthetic episomes can robustly express transgenes, they are unstable and lost from the population when selection pressure is not applied. These characteristics are ideal for transient CRISPR/Cas9 gene disruption, making it easy to screen transformants for mutations in the target site, followed by curing of the episome (*256, 257*). The cured mutants, thus contain a scar in the target site but lack an antibiotic resistance gene or any other transgenic elements, and can be used for subsequent modifications and do not need further biocontainment strategies. This technique will aid in the development of chassis strains for biotechnology, and makes marker-free genomic integration or endogenous locus tagging of expression constructs foreseeable.

To make the generated non-transgenic NR-KO mutant available for wide use, the NR1-KO 4G-5A strain is deposited with NCMA. To make the high-efficiency, CRISPR/Cas9 dual component pNOC-CRISPR vectors widely available, the plasmids developed in this study are deposited with Addgene (Table 4.1).

EXPERIMENTAL PROCEDURES

Strains and growth conditions

N. oceanica CCMP1779 was used in all experiments. Cells were grown in f/2 medium under constant 100 µmol m^{-2} s^{-1} light at 22º C, on a shaker set at 120 rpm. F/2 medium with 2.5 mM KNO_3 or 2.5 mM $NaNH_4$ was used. Hygromycin B (Sigma-Aldrich) was utilized at a concentration of 100 µg/ml. Cell concentrations were measured with a Z2 Coulter Counter (Beckman Coulter) with a range of 1.8-3.6 µM. To determine cell growth rates, *N. oceanica* strains were inoculated to 5 10^6 cells/ml and cell density measured every 24 hours. For

assessment of antibiotic resistance, growth after plating 30,000 cells on solid medium was monitored for 1 month.

CRISPR plasmid construction

All enzymes were obtained from NEB, unless otherwise noted. Vector maps were produced using SnapGene (www.Snapgene.com), and sequence views generated with Benchling (www.Benchling.com). Primers used in this study are listed in Table 4.2.

Generation of pNOC-CRISPR-GFP: The *Cas9* gene was PCR amplified from hCas9 (AddGene 41815) with the primers Cas9 NLS F+ and Cas9 linker R-, and *GFP* was amplified from pOCS-Ole-LDSP (*3*) with the primers GFP linker F+ and GFP NLS R- (Table 4.2). The vector backbone of pNOC 411 (containing the EF promoter (*NannoCCMP1779_10181*) and LDSP terminator (*NannoCCMP1779_418*8)) (*11*) was amplified with the primers 411 Cas9 F- and 411 GFP R+, and fused by Gibson Assembly (*265*) with Cas9 and GFP fragments to form pNOC-Cas9-GFP. A Hygromycin B resistance cassette, composed of the LDSP promoter, the *HygR* gene, and the 35S terminator was PCR amplified from pSEL100 with the primers LDSP NsiI F+ and LDSP BspQI R-. This PCR product and the vector pNOC-Cas9-GFP were digested with BspQI and NsiI and ligated to form pNOC-Cas9-GFP-HygC. The Ribi promoter (between *NannoCCMP1779_9669* and *NannoCCMP1779_9668*) and the terminator of the *N. oceanica* cellulose synthase gene (*CS*, *NannoCCMP1779_5780*) were PCR amplified from genomic DNA with the primer pair Ribi CRISPR F+ and Ribi CRISPR R-, and CSterm sgRNA F+ and CSterm sgRNA R- respectively, and were combined with the XhoI/AgeI digested pNOC-Cas9-GFP backbone by Gibson Assembly, to form pNOC-stacked-Cas9-GFP. A smaller section of the CS terminator that did not contain BspQI sites was PCR amplified CSterm sgRNA F+ and CSterm

sgRNA R2-, and used to replace the CS terminator. The gRNA scaffold with a HDV 3' ribozyme sequence was synthesized (Integrated DNA Technologies) and placed into the ClaI and KpnI sites in pNOC-stacked-Cas9-GFP to form pNOC-CRISPR-GFP.

Generation of pNOC-CRISPR: A C-terminal section of *Cas9* was amplified with the addition of an HA tag-SV40 NLS by overlap extension PCR using the primers Cas9 3' AscI F+, Cas9 3' RE R1-, and Cas9 3' RE R2-, and after digestion with AscI/MfeI ligated with AscI/MfeI digested pNOC-CRISPR-GFP to form pNOC-CRISPR-HA. *N. oceanica* codon optimized *Nlux-HA* was PCR amplified from pNOC-OX-Nlux (*15*) with primers Nlux HpaI F+ and Nlux NheI R-. This PCR product and the vector pNOC-CRISPR-HA were digested with HpaI/NheI and ligated to form pNOC-CRISPR.

Generation of pNOC-ARS-CRISPR: The CEN/ARS region was PCR amplified from pDEST22 (Thermofisher) with the primers CEN/ARS F+ and CEN/ARS R-, and digested with AgeI. The vector pNOC-CRISPR was digested with AgeI, phosphatase (CIP) treated, and ligated with the ARS region to form the vector pNOC-ARS-CRISPR.

Design and cloning of *NR* sgRNAs: Target sequences were chosen by identifying NGG motifs at the 5' end of the *NR* coding sequence. The potential target sequences were submitted to the *N. oceanica* CCMP1779 BLAST (https://genome.jgi.doe.gov/pages/blast-query.jsf?db=Nanoce1779) to identify mismatches and nonspecific targets sharing PAM sequences were rejected. Corresponding guide sequences with a HH ribozyme sequence were appended to the 5' end of gRNA scaffold by PCR site directed mutagenesis (*257, 262*). For this process a fragment of the pNOC-CRISPR including the gRNA scaffold, the CSterminator, the *AmpR* resistance gene and the origin of replication was amplified with the primers sgHH NR-1 ClaI F+ and sgHH NR-1 R-, and sgHH NR-2 ClaI F+ and sgHH NR-2 R-, for sgNR1 and sgNR2

respectively (Figures 4.3). These primers contain the HH ribozyme sequence that includes a section complementary to the guide sequence (Figure 4.3A). After PCR amplification, the product was ligated and cloned into *E. coli* to generate the sgHH-target plasmid (Figure 4.3B). The ClaI/KpnI fragment of sgHH-target was transferred to pNOC-ARS-CRISPR digested with ClaI/KpnI.

Generation of pNOC-ARS-CRISPR-v2: This smaller vector was generated to allow HH-guide insertion on the gRNA scaffold using BspQI restriction sites. To remove the BspQI site originally present in the CEN/ARS region, the backbone of pDEST22 (Thermofisher) was amplified with the primers Crispr compact F+ and Crispr compact R-, religated, and amplified with CEN/ARS -bspq sdm F+ and CEN/ARS -bspq sdm R. This vector backbone and the CRISPR cassette (the *Cas9* and sgRNA expression cassette under the control of the Ribi promoter, and CS and LDSP terminators) from pNOC-ARS-CRISPR were digested with NotI/AgeI and ligated together to form pNOC-ARS-CRISPR-compact. A 35S terminator was amplified from pSELECT100 (*3*) with the primers 35S terminator F+ and 35S terminator R- and digested with NotI/PmeI and ligated into pNOC-ARS-CRISPR-compact NotI/PmeI sites to form pNOC-ARS-CRISPR-35S. The Hygromycin B resistance cassette from pNOC-ARS-CRISPR, that includes the LDSP promoter and HygR gene, was amplified with the primers LDSPpro F+ and HygR R-, digested with NotI/NheI and ligated into the NotI/SpeI sites of pNOC-ARS-CRISPR-35S to form pNOC-ARS-CRISPR-v2.

N. oceanica transformation

Integrating vectors were digested with AseI and concentrated by ethanol precipitation. Transformations were conducted as previously described (*15*), with 3 µg of DNA and 30 µg of

blocking DNA (Ultrapure salmon sperm DNA - Invitrogen). Transformants were plated using the top agar method on 100 µg/ml Hygromycin B and grown under 100 µmol m^{-2} s^{-1} for 3 weeks. Individual lines were transferred to 500 µl of f/2 with 100 µg/ml Hygromycin B (GoldBio) in a 96 deep well plate (Evergreen Biotech). Cultures were then maintained on solid plates.

NanoLuciferase luminescence assays

For Nlux activity screening, 100 µl of cell culture was transferred to a luminescence plate, and 100 µl of f/2 containing Nanoglo substrate (Promega) at a 10,000X dilution. For normalized luminescence measurements, 1.5 million cells per well were transferred to the 96-well luminometer plate, the volume was adjusted to 100 µl with f/2 medium, and 100 µl of f/2 containing Nanoglo substrate (Promega) at a 10,000X dilution was added. Luminescence was measured after 180 sec delay with a 0.3 sec exposure using a Centro XS3 LB960 (Berthold).

N. oceanica **colony PCR**

A 1X Q5 buffered solution (NEB) with 1 µl of *N. oceanica* culture per 10 µl of required sample was boiled (100º C) for 10 min. A 10 µl PCR master mix was added to the 10 µl of sample and PCR conducted according to the manufacturer's suggestions (NEB). The primers NR F+ and NR R- were used to amplify the NR gene (Table 4.2).

Episomal DNA isolation from *N. oceanica*

For DNA isolation 10 ml of culture at mid-log phase (~ 3 10^7 cells ml^{-1}) was collected and frozen. The protocol of Karas et al. (*41*) was utilized with some modifications. Briefly, for each reaction 235 µl Qiagen P1 solution was combined with 5 µl lysozyme (25 mg/ml) (Sigma-

Aldrich), 2.5 µl macerozyme (100 mg/ml)(Yakult Pharmaceuticals), 2.5 µl zymolyase (10

mg/ml) (Zymo Research), and 5 µl cellulase (100mg/ml)(Yakult Pharmaceuticals) and used to

resuspend the cell pellets. The resuspended cell culture was incubated at 37º C for 30 min. After

addition of the P2 solution, and S3 solution (Qiagen), the cell debris was pelleted at 14000 x G

for 10 min. Supernatant was aspirated and combined with an equal volume of isopropanol, mixed

and centrifuged at 14,000 x G at 4º C for 10 min. After decanting the pellet was washed with 750

µl 70% ethanol, centrifuged, decanted, and dried. The DNA was resuspended in 40 µl TE, and

the concentration determined using a NanoDrop Lite (Thermo Scientific).

Episome rescue

Equal amounts of DNA isolated from NR-KO, iEV, and WT lines were used for *E. coli*

transformations. For enzyme treatment of the DNA extracted from NR-KO lines, 2 µg of DNA

was treated with 10 units of Exonuclease V (NEB) and/or 10 units of ClaI (NEB) in a 20 µl

reaction for 1 hour at 37º C. Reactions were heat inactivated at 75º C for 30 min. An equal

amount of DNA (500 ng) from enzyme treated samples, mock treated, and untreated samples

were used to transform *E. coli* (DH5α subcloning efficiency, NEB). Resulting colonies were

counted and random colonies selected for plasmid isolation. Resulting plasmids and control

pNOC-ARS-CRISPR-sgNR2 DNA, were digested with AgeI, NotI, and EcorI (NEB) according

to manufacturer's instructions (NEB), and separated on a 0.9% agarose gel.

Episome curing

Episome containing NR-KO lines were grown in liquid 2.5 mM $NaNH_4$ f/2 without Hygromycin

B for 10 days. Cultures were diluted to 3,000 cells per ml, and 3,000 cells were plated on 2.5

mM NaNH$_4$ f/2. Single colonies developed over 3 weeks, and 48 colonies were isolated and grown in 2.5 mM NaNH$_4$ f/2. After 1 week, the Nlux luminescence was measured and low Nlux signal lines were further screened by colony PCR. Primers used were NR F+, NR R-, and epi Cas9 F+and epi Cas9 R- (Table 4.2).

Immunoblotting

Immunoblotting was conducted as described previously (*2*). Protein extract (50 µg) was separated on 8-10% SDS-polyacrylamide gels. After transfer to PVDF membranes Cas9-GFP was detected with α-GFP (Abcam ab5450) 1:1,000 in TBST with 5% BSA followed by donkey α-goat-HRP (Santa Cruz sc-2020) 1:10,000 in TBST with 5% milk, and Cas9-Nlux-HA with α-HA-HRP (Roche 3F10) 1:1,000 in TBST with 5% milk. Detection was conducted by chemilumiscence with Femto substrate (ThermoFisher). After detection, total protein was stained with Direct Blue 71 (*197*).

Confocal microscopy

Confocal microscopy was performed using an inverted Olympus FluoView1000 confocal laser scanning microscope (Olympus Corporation, USA). Cells grown in f/2 medium were stained with 4',6-Diamidino-2-Phenylindole (DAPI, 358/461 - Life Technologies) with a final concentration of 0.2 µg/µl and observed using a 100x UPlanSApo oil objective (N.A. 1.4). DAPI fluorochromes were excited using a 405 nm blue diode laser, and the emission signals were filtered using the BA430-470 band pass filter. Cas9-GFP was excited using an argon 488 nm laser, and the emission signals from GFP were filtered using a BA500-530 band pass filter. Post-imaging analyses were performed using Olympus FluoView1000 software.

Southern-blot analysis

We used a procedure described previously (*3*). Briefly, *N. oceanica* DNA was isolated by CTAB and 20 µg was digested with SacI and was separated on an agarose gel (0.9%, 70 Volts overnight), followed by blotting to a Hybond Nylon membrane overnight (GE Health Care). Subsequent hybridization and detection was performed with a digoxigenin (DIG) labeling and detection kit according to the manufacturer's instructions (Roche Applied Sciences). Two probes were amplified by PCR, one for the HygR gene (with the primers HygR probe F+ and HygR probe R-) and one for the AmpR gene (AmpR probe F+ and AmpR probe R-), and were used for hybridization in PefectHyb Plus hybridization buffer (Sigma-Aldrich).

APPENDIX
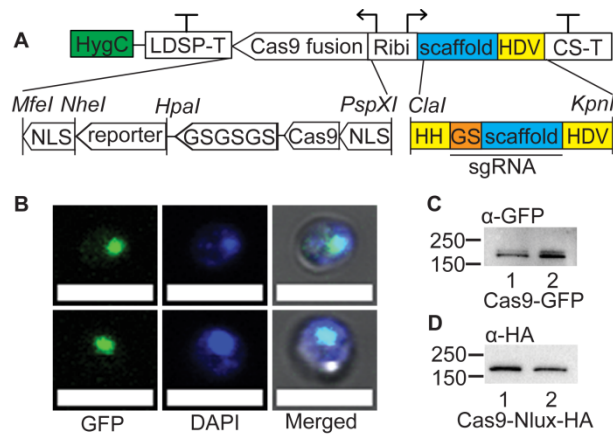
Figure 4.1. A one-vector CRISPR system for gene disruption in *N. oceanica*.
(A) The pNOC-CRISPR vector series includes a Hygromycin B resistance cassette (HygC, green). A bidirectional promoter (Ribi) drives the transcription of the Cas9-reporter fusion and gRNA scaffold (scaffold, blue), along with the LDSP and CS terminators (LDSP-T, CS-T) respectively. Cas9 is fused to either the GFP or Nlux (with HA tag) reporters by a 3x glycine-serine linker (GSGSGS) and contains SV40 nuclear localization signals on the N' and C' termini (NLS). The gRNA scaffold and the 3' self-cleaving HDV ribozyme is integrated into the vector. The 5' hammerhead ribozyme (HH) specific for each guide sequence (GS, orange) is fused to the gRNA scaffold (scaffold) to form a sgRNA. Ribozymes are highlighted in yellow. Unique restriction sites are shown with an upwards line and the name in italics. (B) Confocal analysis of DAPI nuclear staining, Cas9-GFP signal, and merged brightfield, DAPI and GFP signal in *N. oceanica* cells. Scale bar of 5 µM. (C) Immunoblotting with an α-GFP antibody detected the Cas9-GFP produced in *N. oceanica* transformed with pNOC-CRISPR-GFP. (D) Immunoblotting with an α-HA antibody detected the appropriately sized Cas9-Nlux-HA in *N. oceanica* transformed with pNOC-CRISPR. For (C) and (D) numbers on the left of immunoblots indicate size markers (KDa).
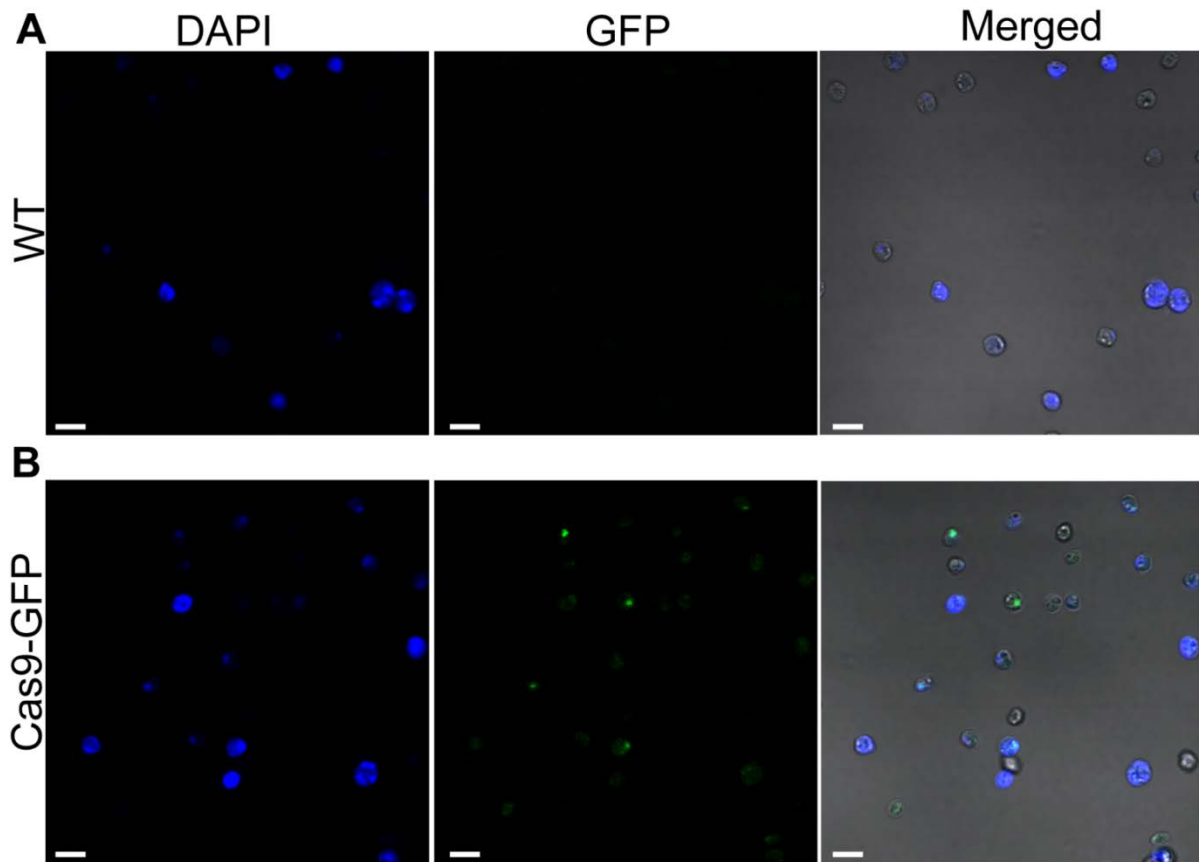
Figure 4.2. Confocal microscopy of Cas9-GFP expressing *N. oceanica* with the nucleus stained by DAPI.
Scale bars of 5 µm. (A) Wild-type (WT) cultures with DAPI nuclear stain, GFP signal, and merged view of DAPI, GFP and brightfield signals. (B) Tranformants producing Cas9-GFP with DAPI nuclear stain, GFP signal, and merged view of DAPI, GFP and brightfield signal.
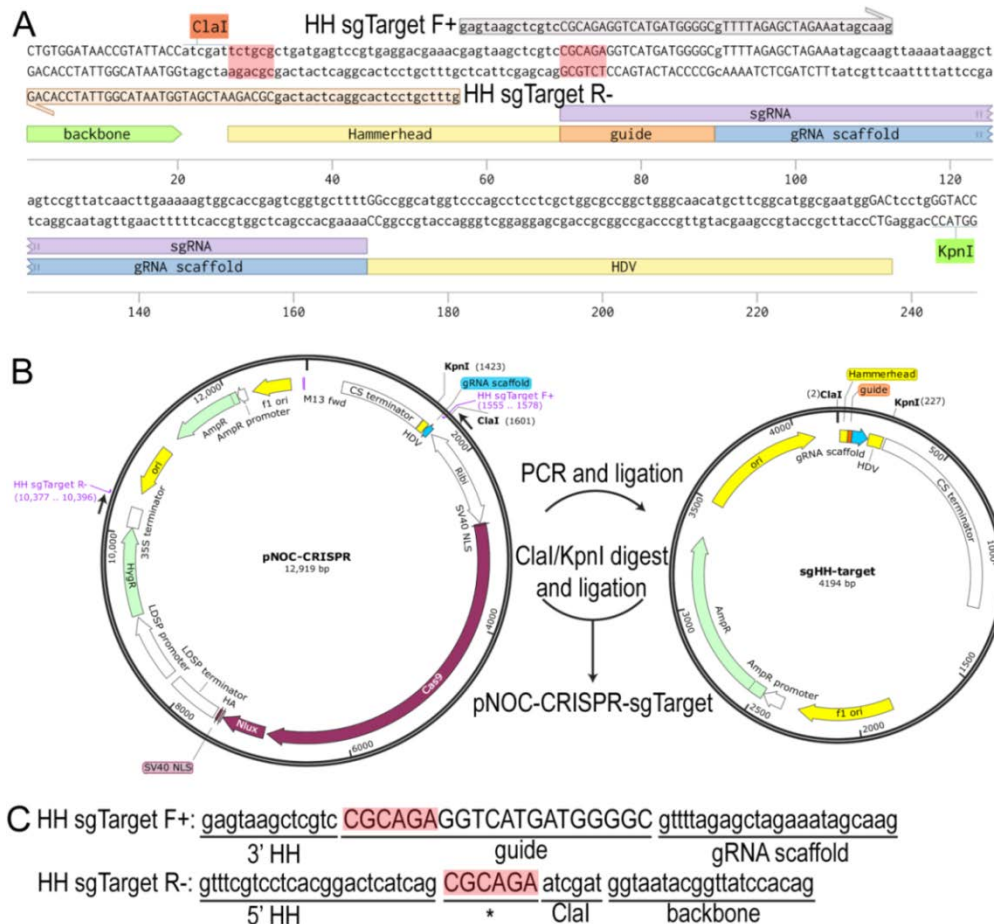
Figure 4.3. Cloning strategies for the generation of a ribozyme-sgRNA.
(A) Example of a ribozyme (highlighted in yellow) flanked sgRNA (highlighted in purple) with unique ClaI and KpnI restriction sites indicated. The hammerhead ribozyme is customized to each guide sequence (highlighted in orange) with a reverse complement of the first six basepairs (highlighted in red). The 80 bp gRNA scaffold (highlighted in blue) has a HDV ribozyme at the 3' end. Primers used to clone the hammerhead-guide sequence, HH sgTarget F+ and HH sgTarget R-, are indicated as grey and orange arrows respectively. (B) Subcloning strategy to generate ribozyme flanked sgRNA in pNOC-CRISPR series. The HH sgTarget F+ primer, anneals at the 5' end of the gRNA scaffold (blue) on pNOC-CRISPR vectors and includes a 5' extension containing the 20 bp guide sequence and 3' end of the HH sequence. The HH sgTarget R- primer anneals to a region in the vector backbone downstream of the hygromycin resistance gene (HygR, green) and includes a 5' extension containing the 5' end of the HH ribozyme sequence, 6 bp specific to the guide sequence for loop formation and a ClaI site. The vector map indicates the location of primer annealing. After amplification, ligation, and transformation into *E. coli*, a subcloning vector (sgHH-target) containing the HH-guide-sgRNA-HDV sequence is recovered. The ClaI/KpnI digested fragment of the sgHH-target vector is then ligated with ClaI/KpnI digested pNOC-CRISPR vector to form the final pNOC-CRISPR-sgTarget vector. (C) Sequences of generic sgTarget F+ and HH sgTarget R- primers. The 20 bp guide sequence is placed in the guide region (shown in uppercase) of a HH sgTarget F+ primer, and the first 6 bp of the guide sequence (uppercase, highlighted in red) placed in the asterisk region (*) of the HH sgTarget R- primer.

156
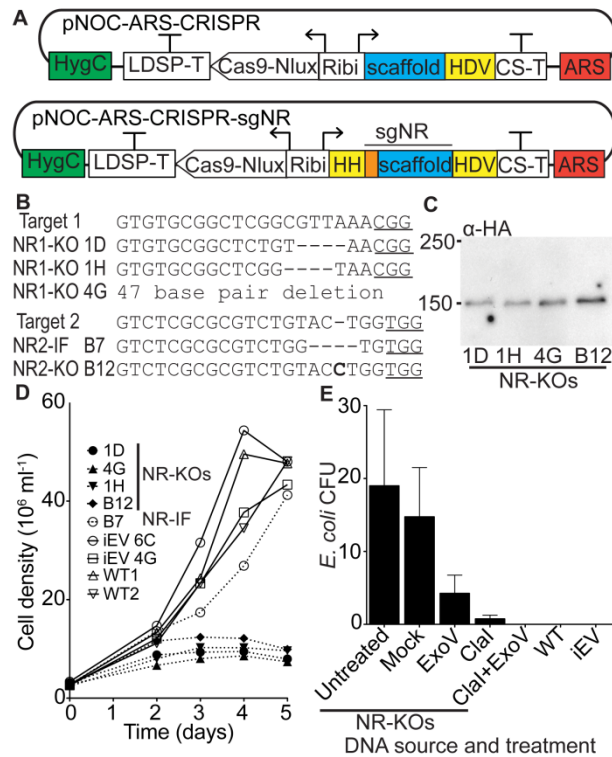
Figure 4.4. Development of an episomal CRISPR system.
(A) The *S. cerevisiae* CEN/ARS6 region (ARS, red) was included in the pNOC-ARS-CRISPR construct for episomal maintenance. Guide sequences (orange) for nitrate reductase targets, with a 5' hammerhead ribozyme (HH), were fused to the gRNA scaffold to form the NR sgRNAs (sgNR). Ribozymes are highlighted in yellow. The sgNR1 and sgNR2 were added to pNOC-ARS-CRISPR to form pNOC-ARS-CRISPR-sgNR1 and pNOC-ARS-CRISPR-sgNR2, respectively. *N. oceanica* is transformed with circular episomal CRISPR constructs. (B) Mutations in the two target sites in the *NR* genomic locus (Target 1 and Target 2) of *N. oceanica* transformed with the respective pNOC-ARS-CRISPR-sgNR construct. Mutant lines are identified by 96-well plate location (Figure 4.5). Deleted nucleotides are represented with dashes and inserted nucleotides are shown in bold. Protospacer adjacent motifs (PAM sites) are underlined. (C) Immunoblot using an α-HA antibody of *N. oceanica* NR1-KO and NR2-KO lines producing Cas9-Nlux-HA from the CRISPR episome. Numbers on the left indicate size markers (KDa). (D) Growth curves after transfer from NH_4 to NO_3 containing medium of NR-KO frame-shifted lines (1D, 1H, 4G, B12) and NR2-IF B7 in-frame line, empty vector integrated CRISPR control lines (iEV), and wild-type (WT). (E) Episome rescue by *E. coli* transformation using equal quantities of DNA isolated from NR-KO lines, iEV, and WT *N. oceanica* lines. Values are the average colonies generated ± SE (NR-KOs n = 4 independent lines, WT n = 2 biological replicates, and iEV n = 2 independent lines). Equal quantities of DNA from NR-KO lines after treatment were used for *E. coli* transformation, and the resulting colonies counted (n = 4 independent lines). Exonuclease V (ExoV), ClaI endonuclease (ClaI), and ClaI endonuclease with Exonuclease V (ClaI+ExoV).

157
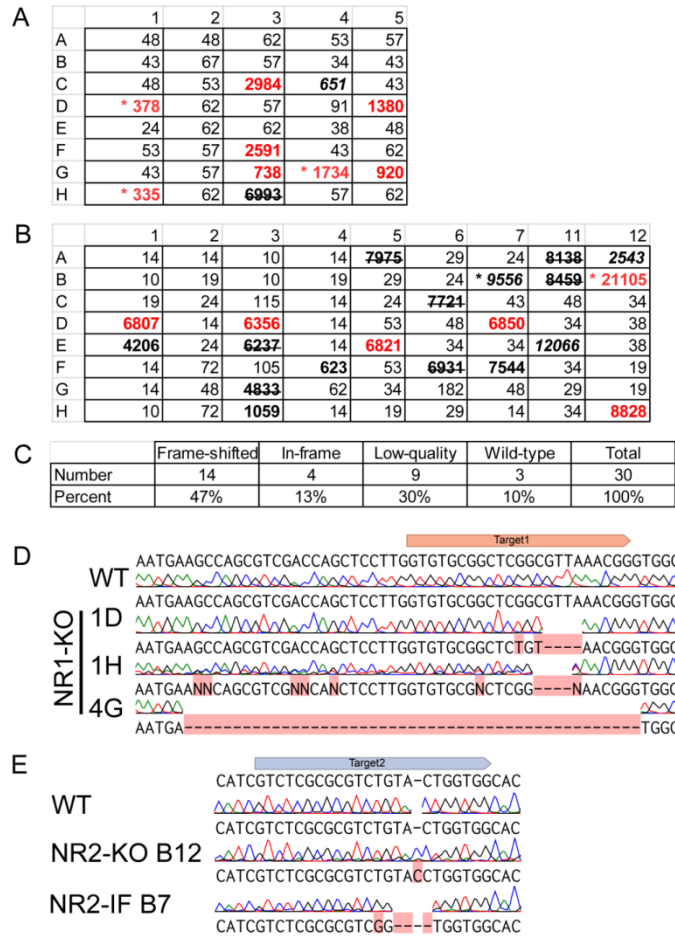
Figure 4.5. Identification of NR knockout mutants by CRISPR/Cas9.
(A) Representative 96-well Nlux luminescence screen of NR1-KO lines. The PCR products of the *NR* genomic locus from bolded wells were submitted for Sanger sequencing. Frame-shift mutations are red, in-frame mutations italicized, colonies returning poor quality sequences are indicated by strikethroughs, wild-type sequences are unmodified (bolded), and lines selected are marked with an asterisks (*). (B) Representative 96-well Nlux luminescence screen of NR2-KO lines. Lines are highlighted as described in panel A. (C) Rates of sequencing results of NR1-KO and NR2-KO lines. (D) Chromatographs of target1 region from wild-type and NR1-KO lines (identified by plate location). (E) Chromatographs of target2 region from wild-type, NR2-KO, and NR2-IF line (identified by plate location).

158
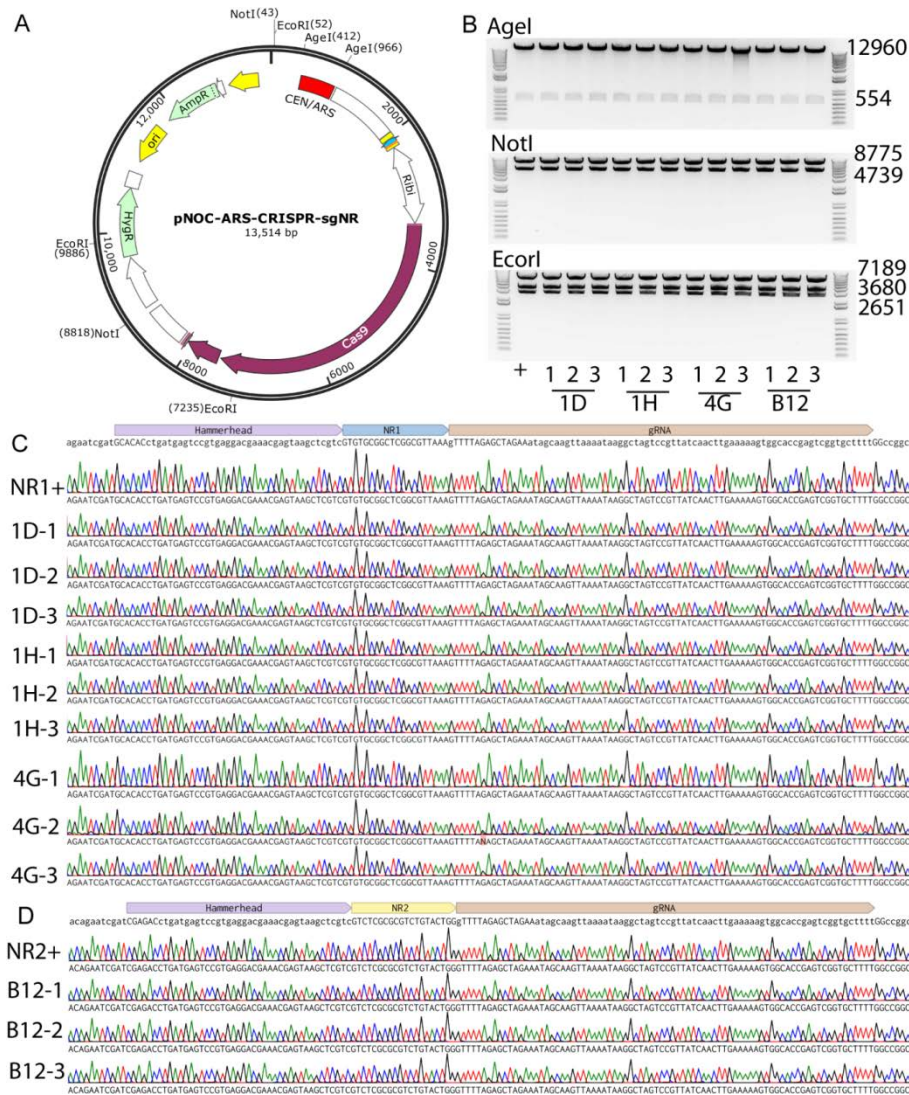
Figure 4.6. Verification of rescued episomes.
(A) Plasmid map indicating restriction sites used for restriction fragment analysis. (B) Triplicate recovered plasmids from each NR-KO episomal line were digested with AgeI, NotI, and EcorI endonucleases. The positive control plasmid (+), pNOC-ARS-CRISPR-sgNR2, was used. Restriction fragments were visualized after separation by agarose (0.9%) gel electrophoresis using a 1 kb+ size ladder (Invitrogen). (C) Sanger sequencing of the sgRNA region of the recovered episomes from NR1-KO lines aligned to pNOC-ARS-CRISPR-sgNR1. (D) Sanger sequencing of the sgRNA region of the recovered episomes from the NR2-KO line aligned to pNOC-ARS-CRISPR-sgNR2.
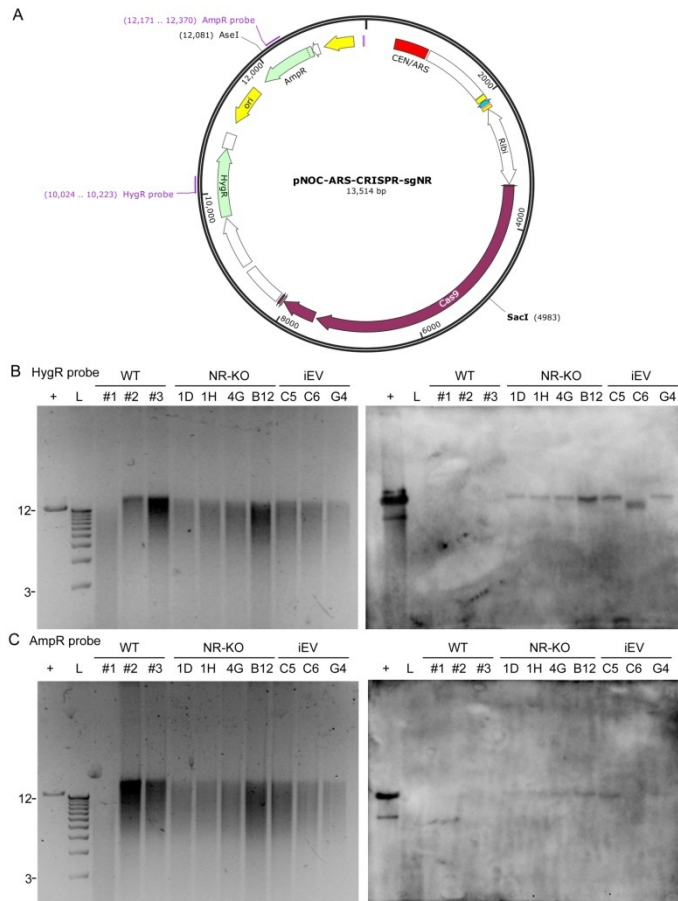
Figure 4.7. Southern blot analysis of the episomal and integrated empty-vector CRISPR mutants.

(A) Plasmid map showing the pNOC-ARS-CRISPR-sgNR vectors used for transformation. DNA was digested SacI site prior to analysis. HygR and AmpR regions were used as probes. The AseI site was used to linearize the plasmid for integrated empty-vector controls. (B) and (C) Genomic DNA digested with SacI was separated with a 0.9% agarose gel. Left panels show the stained agarose gels and right panels show the hybridization signals. The ladder location of 12 kb and 3 kb bands are noted. The blot contains the plasmid pNOC-ARS-CRISPR-sgNR2 as a positive control (+), a 1 kb+ DNA ladder (L), three wild-type DNA (WT) samples as negative controls, DNA from episomal mutant lines (NR-KO), and DNA from integrated empty-vector lines (iEV). HygR and AmpR probes were used in (B) and (C), respectively.

B

NR1-KO 1D (columns 1–6) · **NR1-KO 1H** (columns 7–12)

	1	2	3	4	5	6	7	8	9	10	11	12
A	3195	**5044**	2802	4622	3286	1293	1485	704	53	91	19	**19**
B	1844	556	2088	2294	38	38	115	**43**	354	268	647	38
C	1274	125	1207	474	1025	129	350	125	398	34	29	**1485**
D	541	484	738	**91**	445	436	77	177	53	24	29	556
E	2496	2500	1763	2979	**1001**	1044	1356	**2678**	134	101	38	43
F	1025	647	4244	704	1284	263	680	29	747	833	43	48
G	86	**91**	326	115	153	431	675	38	848	29	139	**38**
H	220	216	163	331	29	**34**	216	273	**53**	762	**34**	522
Low luminescence			25%				Low luminescence		58%			

NR1-KO G4 (columns 1–6) · **NR2-KO B12** (columns 7–12)

	1	2	3	4	5	6	7	8	9	10	11	12
A	**4876**	86	53	77	**72**	77	**43**	369	53	**62**	72	364
B	**34**	273	144	704	627	81	158	**48**	77	163	211	1997
C	163	43	96	556	4182	57	38	57	125	77	239	48
D	29	153	43	48	96	1154	211	307	340	57	57	62
E	24	77	38	**34**	757	43	297	48	72	72	91	**2620**
F	53	57	38	43	105	881	216	1174	158	**623**	1073	484
G	38	**38**	57	**1466**	896	163	91	86	129	273	67	**43**
H	560	182	43	158	53	96	575	129	546	96	924	120
Low luminescence			75%				Low luminescence		60%			

Figure 4.8. Curing episomes from NR-KO lines.
(A) Strategy for the generation of episome-less NR-KO lines. NR-KO mutants carrying the episome were grown without selection for 10 days, and then plated on medium without selection. Single colonies were isolated in 96-well plates and screened for luminescence signal, followed by PCR to detect the *Cas9* gene. A PCR against the genomic sequence of the *NR* gene was used as genomic locus control. (B) Nlux luminescence screen of independent colonies after curing procedure of NR-KO lines. A cutoff of 300 counts/0.3 seconds was used to define low luminescent lines. The percentage of lines falling below the threshold are displayed beneath the raw measurements. Bolded wells were screened by PCR and lines in red were selected as putative "cured" lines (C) Colony PCR for *Cas9* (episomal marker) and *NR* (genomic marker) genes separated on 0.9% agarose gels. 1 Kb+ ladder (L), wild-type (WT), episome carrying parental lines were used as positive control (+), 4 low-Nlux signal, and 2 high-Nlux recovered colonies (identified by location in the 96-well growth plate) were tested. Final cured lines are highlighted in red. The ladder location for 1.65, 1.0, 0.85, and 0.65 kb bands are noted on the left side of each gel.
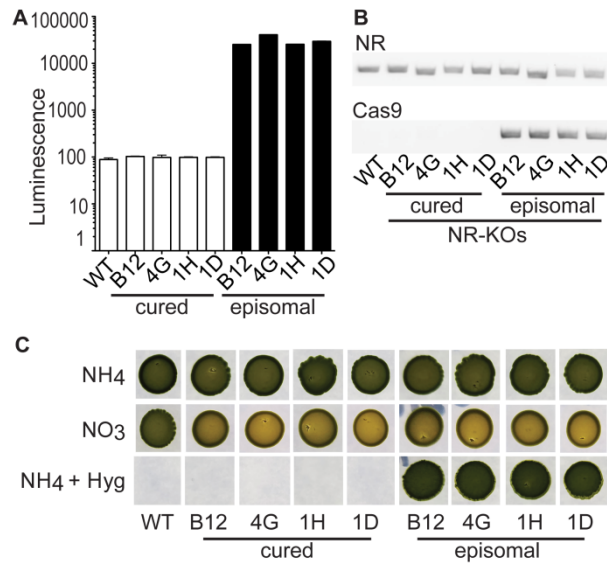
Figure 4.9. Generation of marker-free non-transgenic mutants by episomal removal (curing).
(A) Luminescence from equal number of cells of WT, and NR-KO lines either containing the episome or cured of the episome. (B) PCR for detection of a positive control *NR* genomic locus and the Cas9 regions on the episome conducted on the same DNA extract obtained from WT and, NR-KO episomal and cured lines. (C) Plating of an equal number of cells of WT, NR-KO and NR-KO cured lines on NH_4, NO_3, and NH_4 with Hygromycin B on f/2 solid medium after 1 month of growth.
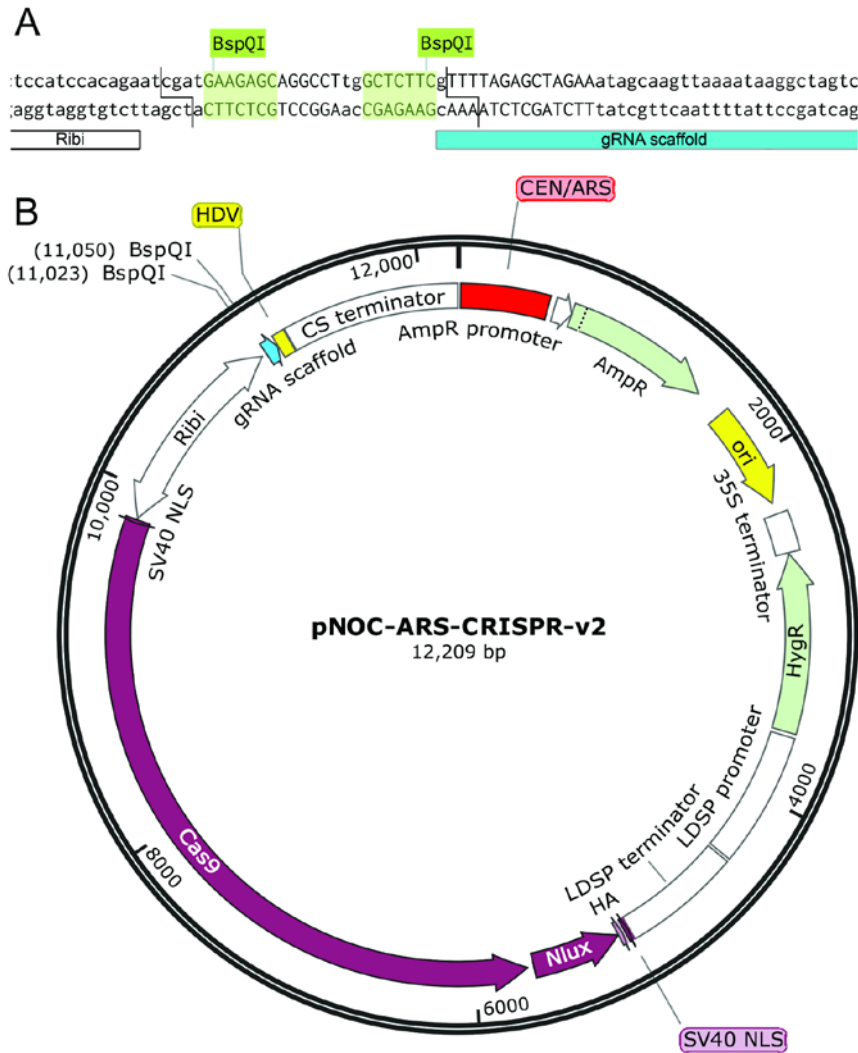
Figure 4.10. A one-vector CRISPR system for scarless cloning of guide sequences.
(A) The multicloning site for guide sequence insertion contains a pair of BspQI sites for removal of the MCS and generation of sticky ends on the Ribi promoter and gRNA scaffold. (B) Map of pNOC-ARS-CRISPR-v2.

Table 4.1. Materials generated in Chapter 4.
Vectors and *Nannochloropsis* strains generated for the study are listed in the first column and if deposited with third-party repository their identification number is listed in the second column

Vectors	Addgene
pNOC-Cas9-GFP	
pNOC-Cas9-GFP-HygC	
pNOC-stacked-Cas9-GFP	
pNOC-CRISPR-GFP	100009
pNOC-CRISPR-HA	99370
pNOC-CRISPR	100008
pNOC-ARS-CRISPR	100011
pNOC-ARS-CRISPR-sgNR-1	100010
pNOC-ARS-CRISPR-sgNR-2	102915
pNOC-ARS-CRISPR-compact	99369
pNOC-CRISPR-35S	99861
pNOC-ARS-CRISPR-v2	99863
Nannochloropsis strains	NCMA
NR1-KO 1D	
NR1-KO 1H	
NR1-KO 4G	
NR1-KO B12	
NR1-KO 1D-6H	
NR1-KO 1H-12A	
NR1-KO 4G-5A	
NR1-KO B12-12G	

.

Table 4.2. Primers used in Chapter 4.
The primer name is listed on the left followed by the oligo sequence. Restriction enzymes used for cloning are noted in the third column and is underlines in the sequence. Features created or altered by the oligo are listed in the fourth column and bolded in the sequence.

Primer name	Primer sequence	Restriction site	Feature
CRISPR constructs			
Cas9 NLS F+	<u>ACTCGAGC</u>ATG**CCCAAGAAAAAGCGGAAGGTG**GACAAGAAGTACTCCATTG	PspXI	SV40 NLS
Cas9 linker R-	GGACCCGCTGCCGGACCCGTCAGCCCTGCTGTCTC		
GFP linker F+	GACGGGTCCGGCAGCGGGT<u>CCATGG</u>TGAGCAAGGGCG	NcoI	
GFP NLS R-	<u>CAATTG</u>TCA**GACCTTCCTCTTCTTTTTCGG**CTTGTACAGCTCGTCCATG	MfeI	SV40 NLS
411 Cas9 F-	CCGCTTTTTCTTGGGCATG<u>CTCGAG</u>TGTTACGAAGTG	PspXI	
411 GFP R+	**CCGAAAAAGAAGAGGAAGGTC**TGACAATTGGAAAGATCCAAGAGAG		SV40 NLS
LDSP HygR NsiI F+	ATG<u>CATACT</u>TGGAAGATGGAGTGGATGGAGG	NsiI	
LDSP BspQI R+-	<u>GCTCTTC</u>CAGCTGGTCACTGGATTTTGGTTTTAG	BspQI	
CSterm sgRNA F+	A<u>ATCGAT</u>AGGCCTTGGTACCATGGGAAAGAAAGGATGAGAA	ClaI	
CSterm sgRNA R-	GGTGCATGGGTTG<u>ACCGGT</u>CGAGGAGGCATTGTATTTAC	AgeI	
CSterm sgRNA R2-	GGGTTG<u>ACCGGT</u>GCAAAGCTGACGCCCTTTTC	AgeI	
Ribi CRISPR F+	CCATGGTACCAAGGCCT<u>ATCGAT</u>TCTGTGGATGGAGGG	ClaI	
Ribi CRISPR R-	GCTTTTTCTTGGGCAT<u>GCTCGAG</u>GGTTGCGTGTGTATCTGTGTG	PspXI	
Cas9 3' RE AscI F+	CTCTGACCAACTTGGGCG		
Cas9 3' RE R1-	GGG<u>CT**AGC**G</u>TAGTCCGGCACGTCATAGGGGTAGTTAACGGACCCGCTGCCG	NheI	HA tag

Table 4.2 (cont'd)

	GACC		
Cas9 3' RE R2-	CATCACAATTGTCAGACCTTCCTCTTCTTTTTCGGGCTAGCGTAGTCCGG	MfeI	
Nlux HpaI F+	GTTAACATGGTGTTTACTCTCGAGGAC	HpaI	
Nlux NheI R-	GCTAGCCACCATGGACGCGTAGTCGGGCACGTC	NheI	
CEN/ARS R-	GAGTTGACCGGTGGCCGGCCGGTTTCTTAGACGGATCGCTTG	AgeI	
CEN/ARS F+	CTTTGCACCGGTGTTTAAACACCTGGGTCCTTTTCATCACGTG	AgeI	
Crispr compact F+	CTTTGCACCGGTGGCACTTTTCGGGGAAATGTG	AgeI	
Crispr compact R-	GCCATGGCGGCCGCACTAGTAGTAGATGCCGACCGGAGTC	NotI	
CEN/ARS -bspq sdm F+	ACAAGATAAAAGGTAGTATTTGTTGGC		343 G to A
CEN/ARS -bspq sdm R-	ACTTCCTGCTCTCAGGTATTAATG		A to T
35S terminator F+	GCTTGAACTAGTAGTAGATG	SpeI	
35S terminator R-	ACATGTTTAAACCAGGTCACTGGATTTTGG	PmeI	
LDSPpro F+	ACGTAGGCGGCCGCGATGGAGTGGATGGAGGAGGAGG	NotI	
HygR R-	TCGGCAGCTAGCCTATTCCTTTGCCCTCGGAC	NheI	
Target cloning			
sgHH NR-1 ClaI F+	GTTTCGTCCTCACGGACTCATCAGGTGTGCATCGATGGTAATACGGTTATCCACAG	ClaI	
sgHH NR-1 R-	GAGTAAGCTCGTCGTGTGCGGCTCGGCGTTAAAGTTTTAGAGCTAGAAATAGCAAG		
sgHH NR-2 ClaI F+	GTTTCGTCCTCACGGACTCATCAGGTCTCGATCGATGGTAATACGGTTATCCACAG	ClaI	
sgHH NR-2 R-	GAGTAAGCTCGTCGTCTCGCGCGTCTGTACTGGGTTTTAGAGCTAGAAATAGCAAG		

Table 4.2 (cont'd)

Detection primers			
NR F+	CCCCCACTCAAGAGTAACTG		
NR R-	AATGGAGGTGCTCACTGCAC		
Cas9 epi F+	ACCATGCGCATGATGCCTAC		
Cas epi R-	CGTCCAGGACCTCCTTTGTAG		
Southern blot probes			
HygR probe F+	AGCCAGACGAGCGGGTTC		
HygR probe R-	CTTCTGCGGGCGATTTGT		
AmpR probe F+	TTTCCGTGTCGCCCTTAT		
AmpR probe R-	GCTCGTCGTTTGGTATGG		

167

Chapter 5. Concluding Remarks

Characterization of growth in light:dark cycles

I found that *N. oceanica* growth is coordinated under cycling light conditions, with cyclic changes in cell physiology, metabolite levels, and transcript abundance (Chapter 2)(*45*). During the day *N. oceanica* cells grow in size and divide early in the night (Chapter 2)(*45*). Cellular lipids in the form of polar lipids and TAG are accumulated during the day, as well as cellular sugars (Chapter 2)(*45*). On a dry-weight basis TAG and sugars oscillate, with accumulation during the day and decreases during the dark period, indicating these act as storage compounds during the dark period (Chapter 2)(*45*).

Measurements of global transcript abundance indicated coordination between transcription and metabolism during the light:dark cycle (Chapter 2)(*45*). A majority of genes (63.7%) were found to have phased expression with different processes active at different times of the day (Chapter 2)(*45*). Cluster analysis followed by analysis for GO term enrichment identified processes phased to different times during the day including DNA synthesis, photosynthesis, and translation (Chapter 2)(*45*). In the case of LDSP the transcript and protein abundance were correlated (Chapter 2)(*45*). The RNA-seq data is a useful resource for generating new hypotheses and informed the subsequent studies (Chapter 2)(*45*). Work to characterize whether the observed oscillations are regulated by a circadian clock under free-running conditions is continuing. Circadian bioluminescent reporter lines have been established and I have identified putative clock component genes for disruption by CRISPR/Cas9.

The genes of the EPA biosynthetic pathway are strongly co-expressed during the light:dark cycle, with an increase in transcript abundance starting at dawn and peaking midday (Chapter 2 and Chapter 3)(*15, 45*). This suggests the genes of the pathway may be coregulated

by light. The four LC-PUFA FADs and delta-6 FAE cDNAs were isolated, and the sequences analyzed and deposited with NCBI (Chapter 3)(*15*). Heterologous expression of these cDNAs in *S. cerevisiae* led to production of LC-PUFAs and ultimately EPA at 0.1% TFA (Chapter 3)(*15*). The successful reconstitution of the pathway in yeast provided proof of concept for further biotechnological studies focusing on these genes to realize their potential. Therefore, overexpression in the native host, *N. oceanica*, was tested for promoting alterations in the fatty acid profile. Overproduction of a single (delta-9, delta-12, delta-6) FAD increased LC-PUFA and EPA proportions by ~35% and ~25% of TFA, respectively (Chapter 3)(*15*). Stacking of dual and triple FAD overexpression cassettes resulted in lines with two or three overproduced FADs, however the fatty acid profile was not altered further compared to single FAD overexpression lines (Chapter 3)(*15*).

Thus the EPA pathway of *N. oceanica* could potentially be useful for the production of LC-PUFAs for human consumption. In order to make the pathway widely available for experimentation, the yeast expression vectors were deposited with Addgene. The limitation of EPA accumulation in *N. oceanica* could be due to the physiological effect of changing the fatty acid composition, and future work to channel EPA to TAG may lead to further enhancements. Investigations into the role of light regulation on the EPA pathway are ongoing.

Development of transgenic tools

In order to investigate the roles of specific genes, genome modification techniques to achieve the expression of transgenes and the inactivation of endogenous genes are necessary. To build a flexible high-capability toolkit, many elements are needed. The essential components for transgene expression include promoters, terminators, resistance markers, and reporter genes. The

addition of 2A peptides and self-cleaving ribozymes enable targeted breaks in polypeptides and oligoribonuleotides, respectively. Several endogenous promoters have been utilized for transgenic expression; with bidirectional promoters proving particularly useful due to their capacity to express two genes. Based on the RNA-Seq data generated in Chapter 2, the unidirectional EF promoter and the Ribi bidirectional promoter were identified, and tested for transgenic expression (Chapter 3)(*15*). Several reporters were implemented in *N. oceanica* with both fluorescent proteins (including green, yellow, and cyan variants), and luciferases (including firefly and NanoLuciferase) used (Chapter 3)(*15*). These reporters were employed to localize proteins and quantify transgene expression, respectively (Chapter 3)(*15*). Endogenous terminators for the CS, LDSP, and HS genes were utilized in the toolkit (Chapter 3)(*15*). For use of multiple constructs in one strain both HygR and BleR antibiotic resistance markers are available (Chapter 3). After screening several variants of 2A peptides, the P2A of 60 amino acids was found to be most effective at causing ribosomal skipping (Chapter 3)(*15*). These elements were integrated into several vectors for gene overexpression and stacking (Chapter 3)(*15*).

The aforementioned vector toolkit was deposited with Addgene for wide access by the scientific community and I am continuing to develop more advanced genetic engineering techniques. The luciferases reporter proteins are proving particularly useful for monitoring transgene expression, and I am testing additional reporters and biosensors. In ongoing work, I found several additional antibiotic selection agents and implemented resistance markers. I constructed destination and entry vectors to combine multiple stacking cassettes with the capability to produce up to six transgenes, and used the system to produce four transgenes from a single transformation. Ongoing work to characterize additional bidirectional promoters identified

170

addition of 2A peptides and self-cleaving ribozymes enable targeted breaks in polypeptides and oligoribonuleotides, respectively. Several endogenous promoters have been utilized for transgenic expression; with bidirectional promoters proving particularly useful due to their capacity to express two genes. Based on the RNA-Seq data generated in Chapter 2, the unidirectional EF promoter and the Ribi bidirectional promoter were identified, and tested for transgenic expression (Chapter 3)(*15*). Several reporters were implemented in *N. oceanica* with both fluorescent proteins (including green, yellow, and cyan variants), and luciferases (including firefly and NanoLuciferase) used (Chapter 3)(*15*). These reporters were employed to localize proteins and quantify transgene expression, respectively (Chapter 3)(*15*). Endogenous terminators for the CS, LDSP, and HS genes were utilized in the toolkit (Chapter 3)(*15*). For use of multiple constructs in one strain both HygR and BleR antibiotic resistance markers are available (Chapter 3). After screening several variants of 2A peptides, the P2A of 60 amino acids was found to be most effective at causing ribosomal skipping (Chapter 3)(*15*). These elements were integrated into several vectors for gene overexpression and stacking (Chapter 3)(*15*).

The aforementioned vector toolkit was deposited with Addgene for wide access by the scientific community and I am continuing to develop more advanced genetic engineering techniques. The luciferases reporter proteins are proving particularly useful for monitoring transgene expression, and I am testing additional reporters and biosensors. In ongoing work, I found several additional antibiotic selection agents and implemented resistance markers. I constructed destination and entry vectors to combine multiple stacking cassettes with the capability to produce up to six transgenes, and used the system to produce four transgenes from a single transformation. Ongoing work to characterize additional bidirectional promoters identified

in Chapter 3 is underway. For example, I found the bidirectional promoter regulating the nitrate reductase and transporter genes is conditionally expressed depending on the nitrogen source.

A system to disrupt a target gene is needed for robust model organisms and biotechnological chassis organisms. The CRISPR/Cas9 RNA-guided nuclease is a potent tool for genome modification and the dual components were optimized for *N. oceanica* as described in Chapter 4. A series of Cas9-reporter fusions were tested for expression and nuclear location, and single-guide RNAs were produced with flanking self-cleaving ribozymes for efficient gene targeting (Chapter 4). *N. oceanica* has the ability to utilize episomes for transgene expression, which was used as a platform for the CRISPR system (Chapter 4). After a mutation is produced and antibiotic selection pressure is removed, the construct will be lost, producing a strain that is marker and transgene free (Chapter 4). This will be important for future implementation of engineered but non-transgenic production strains in outdoor ponds. A nitrate reductase knockout (NR-KO) line was developed in this way, which has lost the ability to grow on nitrate (Chapter 4). This strain and the vectors utilized provide a facile new selection system and are deposited with public access repositories, to allow broad access to scientists. Due to the NR gene being required for growth on nitrate, work is ongoing to develop the NR-KO strain and the NR gene into an auxotrophic selection system by complementation. The developed vectors can potentially be used in marker-free genetic engineering and episomes as an artificial chromosome.

Additional RNA-seq and genome-wide datasets will enable charting of regulatory maps and identification of crucial enzymatic steps necessary for survival under dynamic environmental conditions or which control metabolic flux into bioproducts. The key regulators and enzymes will be targets for genetic engineering, in part using tools established in this study. A genetic engineering toolkit requires the ability go express genes at different levels and robust reporters

for characterization. The NR, VCP, and Ribi bidirectional promoters identified in this study express at different levels and under different conditions, so regulation of transgenes can be fine-tuned. Putative rate-limiting enzymes can now be overexpressed rapidly and in a combinatorial manner by gene stacking. It is possible transcriptional reprogramming may be superior to promoter swapping for altering gene expression, and this may be feasible by CRISPR transcriptional activators and repressors. Either way, transgene promoter design and transcriptional reprogramming techniques will be aided by a greater understanding of the TF: DNA motif interactions that underly transcriptional regulation. The FP and lux reporters integrated in our toolkit will enable high-throughput and high-resolution characterization of gene expression and function. For example, Nlux protein fusions can track protein production, be easily screened for, used as a BRET donor, or fused to a FP for localization. Epitope tagging of TFs enable several methods for studying DNA and protein interactions, such as chromatin immunoprecipitation DNA sequencing (CHIP-Seq) and, co-immunoprecipitation mass-spectrometry (CoIP-MS). Additionally, CRISPR chromatin affinity mass-spectrometry (CRISPR-ChAP-MS) is a newly developed technique using non-cutting Cas9 to isolate a genome sequence and the associated proteins, which are identified by MS (*266*). Thus the tools developed in this study make possible state-of-the-art techniques to unravel the molecular regulators of metabolism in *N. oceanica*.

APPENDIX

Chapter 1. is prepared as a review article for Plant Cell Reports titled "*Nannochloropsis oceanica* an open-access oleaginous microalga".

Chapter 2. Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1779 under light:dark cycles is reproduced from "Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1779 under light:dark cycles" published by The Plant Journal on Aug 20, 2015 (DOI: 10.1111/tpj.12944) with permission.

Chapter 3. 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 is reproduced from "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" published by Plant Biotechnology Journal on Jul 13, 2017 (DOI: 10.1111/pbi.12772) under a Creative Commons Attribution License (CC-BY).

Chapter 4. Non-transgenic marker-free gene disruption by an episomal CRISPR system in the oleaginous microalga, *Nannochloropsis oceanica* CCMP1779 has been submitted to ACS Synthetic Biology and is reproduced under the terms of the AMERICAN CHEMICAL SOCIETY JOURNAL PUBLISHING AGREEMENT.

REFERENCES

REFERENCES

1.   Zou, N., Zhang, C., Cohen, Z., and Richmond, A. (2000) Production of cell mass and eicosapentaenoic acid (EPA) in ultrahigh cell density cultures of *Nannochloropsis* sp. (Eustigmatophyceae), *European Journal of Phycology 35*, 127-133.

2.   Mühlroth, A., Li, K., Røkke, G., Winge, P., Olsen, Y., Hohmann-Marriott, M., Vadstein, O., and Bones, A. (2013) Pathways of Lipid Metabolism in Marine Algae, Co-Expression Network, Bottlenecks and Candidate Genes for Enhanced Production of EPA and DHA in Species of Chromista, *Marine Drugs 11*, 4662-4697.

3.   Yaakob, Z., Ali, E., Zainal, A., Mohamad, M., and Takriff, M. (2014) An overview: biomolecules from microalgae for animal feed and aquaculture, *Journal of Biological Research-Thessaloniki 21*, 6.

4.   Sheehan, J., Dunahay, T., Benemann, J., and Roessler, P. (1998) Look Back at the U.S. Department of Energy's Aquatic Species Program: Biodiesel from Algae; Close-Out Report.

5.   Chew, K. W., Yap, J. Y., Show, P. L., Suan, N. H., Juan, J. C., Ling, T. C., Lee, D.-J., and Chang, J.-S. (2017) Microalgae biorefinery: High value products perspectives, *Bioresource Technology 229*, 53-62.

6.   Rodolfi, L., Chini Zittelli, G., Bassi, N., Padovani, G., Biondi, N., Bonini, G., and Tredici, M. R. (2009) Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor, *Biotechnology and Bioengineering 102*, 100-112.

7.   Unkefer, C. J., Sayre, R. T., Magnuson, J. K., Anderson, D. B., Baxter, I., Blaby, I. K., Brown, J. K., Carleton, M., Cattolico, R. A., Dale, T., Devarenne, T. P., Downes, C. M., Dutcher, S. K., Fox, D. T., Goodenough, U., Jaworski, J., Holladay, J. E., Kramer, D. M., Koppisch, A. T., Lipton, M. S., Marrone, B. L., McCormick, M., Molnár, I., Mott, J. B., Ogden, K. L., Panisko, E. A., Pellegrini, M., Polle, J., Richardson, J. W., Sabarsky, M., Starkenburg, S. R., Stormo, G. D., Teshima, M., Twary, S. N., Unkefer, P. J., Yuan, J. S., and Olivares, J. A. (2017) Review of the algal biology program within the National Alliance for Advanced Biofuels and Bioproducts, *Algal Research 22*, 187-215.

8.   Meng, Y., Jiang, J., Wang, H., Cao, X., Xue, S., Yang, Q., and Wang, W. (2015) The characteristics of TAG and EPA accumulation in *Nannochloropsis oceanica* IMET1 under different nitrogen supply regimes, *Bioresource Technology 179*, 483-489.

9.   Jia, J., Han, D., Gerken, H. G., Li, Y., Sommerfeld, M., Hu, Q., and Xu, J. (2015) Molecular mechanisms for photosynthetic carbon partitioning into storage neutral lipids in *Nannochloropsis oceanica* under nitrogen-depletion conditions, *Algal Research 7*, 66-77.

10.   Xiao, Y., Zhang, J., Cui, J., Yao, X., Sun, Z., Feng, Y., and Cui, Q. (2015) Simultaneous accumulation of neutral lipids and biomass in *Nannochloropsis oceanica* IMET1 under high light intensity and nitrogen replete conditions, *Algal Research 11*, 55-62.

11.   Vieler, A., Wu, G., Tsai, C.-H., Bullard, B., Cornish, A. J., Harvey, C., Reca, I.-B., Thornburg, C., Achawanantakun, R., Buehl, C. J., Campbell, M. S., Cavalier, D., Childs, K. L., Clark, T. J., Deshpande, R., Erickson, E., Armenia Ferguson, A., Handee, W., Kong, Q., Li, X., Liu, B., Lundback, S., Peng, C., Roston, R. L., Sanjaya, Simpson, J. P., TerBush, A., Warakanont, J., Zäuner, S., Farre, E. M., Hegg, E. L., Jiang, N., Kuo, M.-H., Lu, Y., Niyogi, K. K., Ohlrogge, J., Osteryoung, K. W., Shachar-Hill, Y., Sears, B. B., Sun, Y., Takahashi, H., Yandell, M., Shiu, S.-H., and Benning, C. (2012) Genome, Functional Gene Annotation, and Nuclear Transformation of the Heterokont Oleaginous Alga *Nannochloropsis oceanica* CCMP1779, *PLoS Genet 8*, e1003064.

12.   Corteggiani Carpinelli, E., Telatin, A., Vitulo, N., Forcato, C., D'Angelo, M., Schiavon, R., Vezzi, A., Giacometti, G. M., Morosinotto, T., and Valle, G. (2013) Chromosome Scale Genome Assembly and Transcriptome Profiling of *Nannochloropsis gaditana* in Nitrogen Depletion, *Molecular Plant*.

13.   Radakovits, R., Jinkerson, R. E., Fuerstenberg, S. I., Tae, H., Settlage, R. E., Boore, J. L., and Posewitz, M. C. (2012) Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*, *Nat Commun 3*, 686.

14.   Kilian, O., Benemann, C. S. E., Niyogi, K. K., and Vick, B. (2011) High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp., *Proceedings of the National Academy of Sciences of the United States of America 108*, 21265-21269.

15.   Poliner, E., Pulman, J. A., Zienkiewicz, K., Childs, K., Benning, C., and Farre, E. M. (2017) 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 biotechnology journal*.

16.   Wei, L., Xin, Y., Wang, Q., Yang, J., Hu, H., and Xu, J. (2017) RNAi-based targeted gene knockdown in the model oleaginous microalgae *Nannochloropsis oceanica*, *The Plant Journal 89*, 1236-1250.

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

18.   Wang, D., Ning, K., Li, J., Hu, J., Han, D., Wang, H., Zeng, X., Jing, X., Zhou, Q., Su, X., Chang, X., Wang, A., Wang, W., Jia, J., Wei, L., Xin, Y., Qiao, Y., Huang, R., Chen, J., Han, B., Yoon, K., Hill, R. T., Zohar, Y., Chen, F., Hu, Q., and Xu, J. (2014) *Nannochloropsis* Genomes Reveal Evolution of Microalgal Oleaginous Traits, *PLoS Genetics 10*, e1004094.

19. Hu, J., Wang, D., Li, J., Jing, G., Ning, K., and Xu, J. (2014) Genome-wide identification of transcription factors and transcription-factor binding sites in oleaginous microalgae *Nannochloropsis*, *Scientific Reports 4*.

20. Andrianantoandro, E., Basu, S., Karig, D. K., and Weiss, R. (2006) Synthetic biology: new engineering rules for an emerging discipline, *Molecular Systems Biology 2*.

21. Agapakis, C. M. (2014) Designing Synthetic Biology, *ACS Synthetic Biology 3*, 121-128.

22. Li, J., Han, D., Wang, D., Ning, K., Jia, J., Wei, L., Jing, X., Huang, S., Chen, J., Li, Y., Hu, Q., and Xu, J. (2014) Choreography of Transcriptomes and Lipidomes of *Nannochloropsis* Reveals the Mechanisms of Oil Synthesis in Microalgae, *Plant Cell 26*, 1645-1665.

23. Jinkerson, R. E., Radakovits, R., and Posewitz, M. C. (2013) Genomic insights from the oleaginous model alga *Nannochloropsis gaditana*, *Bioengineered 4*, 37-43.

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

25. Nobusawa, T., Hori, K., Mori, H., Kurokawa, K., and Ohta, H. (2017) Differently localized lysophosphatidic acid acyltransferases crucial for triacylglycerol biosynthesis in the oleaginous alga *Nannochloropsis*, *The Plant Journal 90*, 547-559.

26. Dolch, L.-J., Rak, C., Perin, G., Tourcier, G., Broughton, R., Leterrier, M., Morosinotto, T., Tellier, F., Faure, J.-D., Falconet, D., Jouhet, J., Sayanova, O., Beaudoin, F., and Maréchal, E. (2017) A Palmitic Acid Elongase Affects Eicosapentaenoic Acid and Plastidial Monogalactosyldiacylglycerol Levels in *Nannochloropsis*, *Plant Physiology 173*, 742-759.

27. Wang, Q., Lu, Y., Xin, Y., Wei, L., Huang, S., and Xu, J. (2016) Genome editing of model oleaginous microalgae *Nannochloropsis* spp. by CRISPR/Cas9, *The Plant Journal 88*, 1071-1081.

28. Wei, L., Xin, Y., Wang, D., Jing, X., Zhou, Q., Su, X., Jia, J., Ning, K., Chen, F., Hu, Q., and Xu, J. (2013) *Nannochloropsis* plastid and mitochondrial phylogenomes reveal organelle diversification mechanism and intragenus phylotyping strategy in microalgae, *BMC Genomics 14*, 534.

29. Janouskovec, J., Horak, A., Obornik, M., Lukes, J., and Keeling, P. J. (2010) A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids, *Proceedings of the National Academy of Sciences 107*, 10949-10954.

30. Keeling, P. J. (2009) Chromalveolates and the Evolution of Plastids by Secondary Endosymbiosis, *Journal of Eukaryotic Microbiology 56*, 1-8.

31. Zienkiewicz, K., Zienkiewicz, A., Poliner, E., Du, Z.-Y., Vollheyde, K., Herrfurth, C., Marmon, S., Farré, E. M., Feussner, I., and Benning, C. (2017) *Nannochloropsis*, a rich source of diacylglycerol acyltransferases for engineering of triacylglycerol content in different hosts, *Biotechnology for Biofuels 10*.

32. Xin, Y., Lu, Y., Lee, Y.-Y., Wei, L., Jia, J., Wang, Q., Wang, D., Bai, F., Hu, H., Hu, Q., Liu, J., Li, Y., and Xu, J. (2017) Producing Designer Oils in Industrial Microalgae by Rational Modulation of Co-evolving Type-2 Diacylglycerol Acyltransferases, *Molecular Plant*.

33. Murakami, R., and Hashimoto, H. (2009) Unusual Nuclear Division in *Nannochloropsis oculata* (Eustigmatophyceae, Heterokonta) which May Ensure Faithful Transmission of Secondary Plastids, *Protist 160*, 41-49.

34. Kroth, P., and Strotmann, H. (1999) Diatom plastids: Secondary endocytobiosis, plastid genome and protein import, *Physiologia Plantarum 107*, 136-141.

35. Bolte, K., Bullmann, L., Hempel, F., Bozarth, A., Zauner, S., and Maier, U.-G. (2009) Protein Targeting into Secondary Plastids, *Journal of Eukaryotic Microbiology 56*, 9-15.

36. Gschloessl, B., Guermeur, Y., and Cock, J. M. (2008) HECTAR: a method to predict subcellular targeting in heterokonts, *BMC Bioinformatics 9*, 393.

37. Ma, X., Yao, L., Yang, B., Lee, Y. K., Chen, F., and Liu, J. (2017) RNAi-mediated silencing of a pyruvate dehydrogenase kinase enhances triacylglycerol biosynthesis in the oleaginous marine alga *Nannochloropsis salina*, *Scientific Reports 7*.

38. Loira, N., Mendoza, S., Paz Cortés, M., Rojas, N., Travisany, D., Genova, A. D., Gajardo, N., Ehrenfeld, N., and Maass, A. (2017) Reconstruction of the microalga *Nannochloropsis salina* genome-scale metabolic model with applications to lipid production, *BMC Systems Biology 11*.

39. Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J. A., Brownlee, C., Cadoret, J. P., Chiovitti, A., Choi, C. J., Coesel, S., De Martino, A., Detter, J. C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M. J. J., Jenkins, B. D., Jiroutova, K., Jorgensen, R. E., Joubert, Y., Kaplan, A., Kroger, N., Kroth, P. G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jezequel, V., Lopez, P. J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L. K., Montsant, A., Oudot-Le Secq, M. P., Napoli, C., Obornik, M., Parker, M. S., Petit, J. L., Porcel, B. M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T. A., Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M. R., Taylor, A. R., Vardi, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L. S., Rokhsar, D. S., Weissenbach, J., Armbrust, E. V., Green, B. R., Van De Peer, Y., and Grigoriev, I. V. (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes, *Nature 456*, 239-244.

40. Diner, R. E., Noddings, C. M., Lian, N. C., Kang, A. K., McQuaid, J. B., Jablanovic, J., Espinoza, J. L., Nguyen, N. A., Anzelmatti, M. A., Jansson, J., Bielinski, V. A., Karas, B. J., Dupont, C. L., Allen, A. E., and Weyman, P. D. (2017) Diatom centromeres suggest a mechanism for nuclear DNA acquisition, *Proceedings of the National Academy of Sciences 114*, E6015-E6024.

41. Schonknecht, G., Chen, W. H., Ternes, C. M., Barbier, G. G., Shrestha, R. P., Stanke, M., Brautigam, A., Baker, B. J., Banfield, J. F., Garavito, R. M., Carr, K., Wilkerson, C., Rensing, S. A., Gagneul, D., Dickenson, N. E., Oesterhelt, C., Lercher, M. J., and Weber, A. P. M. (2013) Gene Transfer from Bacteria and Archaea Facilitated Evolution of an Extremophilic Eukaryote, *Science 339*, 1207-1210.

42. Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P., Noddings, C. M., Brunson, J. K., Valas, R. E., Deerinck, T. J., Jablanovic, J., Gillard, J. T., Beeri, K., Ellisman, M. H., Glass, J. I., Hutchison, C. A., 3rd, Smith, H. O., Venter, J. C., Allen, A. E., Dupont, C. L., and Weyman, P. D. (2015) Designer diatom episomes delivered by bacterial conjugation, *Nat Commun 6*, 6925.

43. Starkenburg, S. R., Kwon, K. J., Jha, R. K., McKay, C., Jacobs, M., Chertkov, O., Twary, S., Rocap, G., and Cattolico, R. (2014) A pangenomic analysis of the *Nannochloropsis* organellar genomes reveals novel genetic variations in key metabolic genes, *BMC Genomics 15*, 212.

44. Mühlroth, A., Winge, P., Assimi, A. E., Jouhet, J., Marechal, E., Hohmann-Marriott, M. F., Vadstein, O., and Bones, A. M. (2017) Mechanisms of phosphorus acquisition and lipid class remodelling under P limitation in a marine microalga, *Plant Physiology*.

45. Poliner, E., Panchy, N., Newton, L., Wu, G., Lapinsky, A., Bullard, B., Zienkiewicz, A., Benning, C., Shiu, S. H., and Farre, E. M. (2015) Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1779 under light/dark cycles, *The Plant journal : for cell and molecular biology 83*, 1097-1113.

46. Alboresi, A., Perin, G., Vitulo, N., Diretto, G., Block, M. A., Jouhet, J., Meneghesso, A., Valle, G., Giuliano, G., Maréchal, E., and Morosinotto, T. (2016) Light Remodels Lipid Biosynthesis in *Nannochloropsis gaditana* by Modulating Carbon Partitioning Between Organelles, *Plant Physiology*, pp.00599.02016.

47. Tsai, C. H., Warakanont, J., Takeuchi, T., Sears, B. B., Moellering, E. R., and Benning, C. (2014) The protein Compromised Hydrolysis of Triacylglycerols 7 (CHT7) acts as a repressor of cellular quiescence in *Chlamydomonas*, *Proceedings of the National Academy of Sciences of the United States of America 111*, 15833-15838.

48. Miller, R., Wu, G., Deshpande, R. R., Vieler, A., Gartner, K., Li, X., Moellering, E. R., Zauner, S., Cornish, A. J., Liu, B., Bullard, B., Sears, B. B., Kuo, M. H., Hegg, E. L., Shachar-Hill, Y., Shiu, S. H., and Benning, C. (2010) Changes in transcript abundance in *Chlamydomonas reinhardtii* following nitrogen deprivation predict diversion of metabolism, *Plant Physiology 154*, 1737-1752.

49. Chauton, M. S., Winge, P., Brembu, T., Vadstein, O., and Bones, A. M. (2013) Gene Regulation of Carbon Fixation, Storage, and Utilization in the Diatom *Phaeodactylum tricornutum* Acclimated to Light/Dark Cycles, *Plant Physiology 161*, 1034-1048.

50. Huysman, M. J., Fortunato, A. E., Matthijs, M., Costa, B. S., Vanderhaeghen, R., Van den Daele, H., Sachse, M., Inze, D., Bowler, C., Kroth, P. G., Wilhelm, C., Falciatore, A., Vyverman, W., and De Veylder, L. (2013) AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (*Phaeodactylum tricornutum*), *Plant Cell 25*, 215-228.

51. Huysman, M. J., Martens, C., Vandepoele, K., Gillard, J., Rayko, E., Heijde, M., Bowler, C., Inze, D., Van de Peer, Y., De Veylder, L., and Vyverman, W. (2010) Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling, *Genome biology 11*, R17.

52. Ashworth, J., Coesel, S., Lee, A., Armbrust, E. V., Orellana, M. V., and Baliga, N. S. (2013) Genome-wide diel growth state transitions in the diatom *Thalassiosira pseudonana*, *Proceedings of the National Academy of Sciences 110*, 7518-7523.

53. Braun, R., Farre, E. M., Schurr, U., and Matsubara, S. (2014) Effects of light and circadian clock on growth and chlorophyll accumulation of *Nannochloropsis gaditana*, *Journal of phycology 50*, 515-525.

54. Sukenik, A., Carmeli, Y., and Berner, T. (1989) Regulation of Fatty Acid Composition by Irradiance Level in the Eustigmatophyte *Nannochloropsis* Sp., *Journal of phycology 25*, 686-692.

55. Sukenik, A., and Carmeli, Y. (1990) Lipid Synthesis and Fatty Acid Composition in *Nannochloropsis* Sp. (Eustigmatophyceae) Grown in a Light-Dark Cycle, *Journal of phycology 26*, 463-469.

56. Cao, S., Zhang, X., Xu, D., Fan, X., Mou, S., Wang, Y., Ye, N., and Wang, W. (2013) A transthylakoid proton gradient and inhibitors induce a non-photochemical fluorescence quenching in unicellular algae *Nannochloropsis* sp, *FEBS Letters 587*, 1310-1315.

57. Chukhutsina, V. U., Fristedt, R., Morosinotto, T., and Croce, R. (2017) Photoprotection strategies of the alga *Nannochloropsis gaditana*, *Biochimica et Biophysica Acta (BBA) - Bioenergetics 1858*, 544-552.

58. Alboresi, A., Le Quiniou, C., Yadav, S. K. N., Scholz, M., Meneghesso, A., Gerotto, C., Simionato, D., Hippler, M., Boekema, E. J., Croce, R., and Morosinotto, T. (2017) Conservation of core complex subunits shaped the structure and function of photosystem I in the secondary endosymbiont alga *Nannochloropsis gaditana*, *New Phytologist 213*, 714-726.

59. Umetani, I., Kunugi, M., Yokono, M., Takabayashi, A., and Tanaka, A. (2017) Evidence of the supercomplex organization of photosystem II and light-harvesting complexes in *Nannochloropsis granulata*, *Photosynthesis Research*.

60. Schneider, J. C., Livne, A., Sukenik, A., and Roessler, P. G. (1995) A mutant of *Nannochloropsis* deficient in eicosapentaenoic acid production, *Phytochemistry 40*, 807-814.

61. Schneider, J. C., and Roessler, P. (1994) Radiolabeling Studies Of Lipids And Fatty Acids In *Nannochloropsis* (Eustigmatophyceae), An Oleaginous Marine Alga1, *Journal of phycology 30*, 594-598.

62. Pal, D., Khozin-Goldberg, I., Didi-Cohen, S., Solovchenko, A., Batushansky, A., Kaye, Y., Sikron, N., Samani, T., Fait, A., and Boussiba, S. (2013) Growth, lipid production and metabolic adjustments in the euryhaline eustigmatophyte *Nannochloropsis oceanica* CCALA 804 in response to osmotic downshift, *Applied Microbiology and Biotechnology 97*, 8291-8306.

63. Hildebrand, M., Manandhar-Shrestha, K., and Abbriano, R. (2017) Effects of chrysolaminarin synthase knockdown in the diatom *Thalassiosira pseudonana*: Implications of reduced carbohydrate storage relative to green algae, *Algal Research 23*, 66-77.

64. Arnold, A. A., Genard, B., Zito, F., Tremblay, R., Warschawski, D. E., and Marcotte, I. (2015) Identification of lipid and saccharide constituents of whole microalgal cells by 13C solid-state NMR, *Biochimica et Biophysica Acta (BBA) - Biomembranes 1848*, 369-377.

65. Scholz, M. J., Weiss, T. L., Jinkerson, R. E., Jing, J., Roth, R., Goodenough, U., Posewitz, M. C., and Gerken, H. G. (2014) Ultrastructure and Composition of the *Nannochloropsis gaditana* Cell Wall, *Eukaryotic Cell 13*, 1450-1464.

66. Jeong, S. W., Nam, S. W., HwangBo, K., Jeong, W. J., Jeong, B.-r., Chang, Y. K., and Park, Y.-I. (2017) Transcriptional Regulation of Cellulose Biosynthesis during the Early Phase of Nitrogen Deprivation in *Nannochloropsis salina*, *Scientific Reports 7*.

67. Stormo, G. D. (2000) DNA binding sites: representation and discovery, *Bioinformatics 16*, 16-23.

68. Buitrago-Flórez, F. J., Restrepo, S., and Riaño-Pachón, D. M. (2014) Identification of Transcription Factor Genes and Their Correlation with the High Diversity of Stramenopiles, *PLoS ONE 9*, e111841.

69. Rayko, E., Maumus, F., Maheswari, U., Jabbari, K., and Bowler, C. (2010) Transcription factor families inferred from genome sequences of photosynthetic stramenopiles, *New Phytologist 188*, 52-66.

70. Thiriet-Rupert, S., Carrier, G., Chénais, B., Trottier, C., Bougaran, G., Cadoret, J.-P., Schoefs, B., and Saint-Jean, B. (2016) Transcription factors in microalgae: genome-wide prediction and comparative analysis, *BMC Genomics 17*.

71.   Wingender, E. (2008) The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation, *Brief. Bioinformatics 9*, 326-332.

72.   Kwon, S., Kang, N. K., Koh, H. G., Shin, S.-E., Lee, B., Jeong, B.-r., and Chang, Y. K. (2017) Enhancement of biomass and lipid productivity by overexpression of a bZIP transcription factor in *Nannochloropsis salina*: Engineering of *Nannochloropsis* with bZIP TF, *Biotechnology and Bioengineering*.

73.   Schellenberger Costa, B., Sachse, M., Jungandreas, A., Bartulos, C. R., Gruber, A., Jakob, T., Kroth, P. G., and Wilhelm, C. (2013) Aureochrome 1a Is Involved in the Photoacclimation of the Diatom *Phaeodactylum tricornutum*, *PLoS ONE 8*, e74451.

74.   Mann, M., Serif, M., Jakob, T., Kroth, P. G., and Wilhelm, C. (2017) PtAUREO1a and PtAUREO1b knockout mutants of the diatom *Phaeodactylum tricornutum* are blocked in photoacclimation to blue light, *Journal of Plant Physiology 217*, 44-48.

75.   Coesel, S., Mangogna, M., Ishikawa, T., Heijde, M., Rogato, A., Finazzi, G., Todo, T., Bowler, C., and Falciatore, A. (2009) Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity, *EMBO reports 10*, 655-661.

76.   Zou, Y., Wenzel, S., Müller, N., Prager, K., Jung, E.-M., Kothe, E., Kottke, T., and Mittag, M. (2017) An Animal-Like Cryptochrome Controls the *Chlamydomonas* Sexual Cycle, *Plant Physiology 174*, 1334-1347.

77.   Müller, N., Wenzel, S., Zou, Y., Künzel, S., Sasso, S., Weiß, D., Prager, K., Grossman, A., Kottke, T., and Mittag, M. (2017) A Plant Cryptochrome Controls Key Features of the *Chlamydomonas* Circadian Clock and Its Life Cycle, *Plant Physiology 174*, 185-201.

78.   Fortunato, A. E., Jaubert, M., Enomoto, G., Bouly, J.-P., Raniello, R., Thaler, M., Malviya, S., Bernardes, J. S., Rappaport, F., Gentili, B., Huysman, M. J. J., Carbone, A., Bowler, C., d'Alcalà, M. R., Ikeuchi, M., and Falciatore, A. (2016) Diatom Phytochromes Reveal the Existence of Far-Red-Light-Based Sensing in the Ocean, *Plant Cell 28*, 616-628.

79.   Kang, N. K., Choi, G.-G., Kim, E. K., Shin, S.-E., Jeon, S., Park, M. S., Jeong, K. J., Jeong, B.-r., Chang, Y. K., Yang, J.-W., and Lee, B. (2015) Heterologous overexpression of sfCherry fluorescent protein in *Nannochloropsis salina*, *Biotechnology Reports 8*, 10-15.

80.   Kang, N. K., Jeon, S., Kwon, S., Koh, H. G., Shin, S.-E., Lee, B., Choi, G.-G., Yang, J.-W., Jeong, B.-r., and Chang, Y. K. (2015) Effects of overexpression of a bHLH transcription factor on biomass and lipid production in *Nannochloropsis salina*, *Biotechnology for Biofuels 8*.

81.   Beacham, T. A., and Ali, S. T. (2016) Growth dependent silencing and resetting of DGA1 transgene in *Nannochloropsis salina*, *Algal Research 14*, 65-71.

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

83. Li, F., Gao, D., and Hu, H. (2014) High-efficiency nuclear transformation of the oleaginous marine *Nannochloropsis* species using PCR product, *Bioscience, Biotechnology, and Biochemistry 78*, 812-817.

84. Peter Dehoff, S. A., Leah Soriaga. (2014) Autonomous replication sequences and episomal DNA molecules, Synthetic Genomics, Inc., US9447422 B2.

85. Diner, R. E., Bielinski, V. A., Dupont, C. L., Allen, A. E., and Weyman, P. D. (2016) Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods, *Frontiers in bioengineering and biotechnology 4*, 65.

86. Lin, G., Wang, Y., Guo, L., Ding, H., Hu, Y., Liang, S., Zhang, Z., and Yang, G. (2017) Verification of mutagen function of Zeocin in *Nannochloropsis oceanica* through transcriptome analysis, *Journal of Ocean University of China 16*, 501-508.

87. Zaslavskaia, L. A., Lippmeier, J. C., Kroth, P. G., Grossman, A. R., and Apt, K. E. (2001) Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes, *Journal of phycology 36*, 379-386.

88. Huang, J., Liu, J., Li, Y., and Chen, F. (2008) Isolation and Characterization of the Phytoene Desaturase Gene as a Potential Selective Marker for Genetic Engineering of the Astaxanthin-Producing Green Alga *Chlorella Zofingiensis* (Chlorophyta), *Journal of phycology 44*, 684-690.

89. Cheah, Y. E., Albers, S. C., and Peebles, C. A. M. (2013) A novel counter-selection method for markerless genetic modification in *Synechocystis* sp. PCC 6803, *Biotechnology Progress 29*, 23-30.

90. Wei, H., Shi, Y., Ma, X., Pan, Y., Hu, H., Li, Y., Luo, M., Gerken, H., and Liu, J. (2017) A type-I diacylglycerol acyltransferase modulates triacylglycerol biosynthesis and fatty acid composition in the oleaginous microalga, *Nannochloropsis oceanica*, *Biotechnology for Biofuels 10*.

91. Kaye, Y., Grundman, O., Leu, S., Zarka, A., Zorin, B., Didi-Cohen, S., Khozin-Goldberg, I., and Boussiba, S. (2015) Metabolic engineering toward enhanced LC-PUFA biosynthesis in *Nannochloropsis oceanica*: Overexpression of endogenous $\Delta 12$ desaturase driven by stress-inducible promoter leads to enhanced deposition of polyunsaturated fatty acids in TAG, *Algal Research 11*, 387-398.

92. Peter Dehoff, L. S. (2014) Nannochloropsis kozak consensus sequence, Exxonmobil Research And Engineering Company, US9447422 B2.

93.   Gallie, D. R., Sleat, D. E., Watts, J. W., Turner, P. C., and Wilson, T. M. A. (1987) The 5'-leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts *in vitro* and *in vivo*, *Nucleic Acids Research 15*, 3257-3273.

94.   Xue, J., Niu, Y.-F., Huang, T., Yang, W.-D., Liu, J.-S., and Li, H.-Y. (2015) Genetic improvement of the microalga *Phaeodactylum tricornutum* for boosting neutral lipid accumulation, *Metabolic Engineering 27*, 1-9.

95.   Li, D.-W., Cen, S.-Y., Liu, Y.-H., Balamurugan, S., Zheng, X.-Y., Alimujiang, A., Yang, W.-D., Liu, J.-S., and Li, H.-Y. (2016) A type 2 diacylglycerol acyltransferase accelerates the triacylglycerol biosynthesis in heterokont oleaginous microalga *Nannochloropsis oceanica*, *Journal of Biotechnology 229*, 65-71.

96.   Gee, C. W., and Niyogi, K. K. (2017) The carbonic anhydrase CAH1 is an essential component of the carbon-concentrating mechanism in *Nannochloropsis oceanica*, *Proceedings of the National Academy of Sciences 114*, 4537-4542.

97.   Hall, M. P., Unch, J., Binkowski, B. F., Valley, M. P., Butler, B. L., Wood, M. G., Otto, P., Zimmerman, K., Vidugiris, G., Machleidt, T., Robers, M. B., Benink, H. A., Eggers, C. T., Slater, M. R., Meisenheimer, P. L., Klaubert, D. H., Fan, F., Encell, L. P., and Wood, K. V. (2012) Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate, *ACS chemical biology 7*, 1848-1857.

98.   Suzuki, K., Kimura, T., Shinoda, H., Bai, G., Daniels, M. J., Arai, Y., Nakano, M., and Nagai, T. (2016) Five colour variants of bright luminescent protein for real-time multicolour bioimaging, *Nat Commun 7*, 13718.

99.   Shih, C.-H., Chen, H.-Y., Lee, H.-C., and Tsai, H.-J. (2015) Purple Chromoprotein Gene Serves as a New Selection Marker for Transgenesis of the Microalga *Nannochloropsis oculata*, *PLOS ONE 10*, e0120780.

100.   Sharma, P., Yan, F., Doronina, V. A., Escuin-Ordinas, H., Ryan, M. D., and Brown, J. D. (2012) 2A peptides provide distinct solutions to driving stop-carry on translational recoding, *Nucleic Acids Research 40*, 3143-3151.

101.   Plucinak, T. M., Horken, K. M., Jiang, W., Fostvedt, J., Nguyen, S. T., and Weeks, D. P. (2015) Improved and versatile viral 2A platforms for dependable and inducible high-level expression of dicistronic nuclear genes in *Chlamydomonas reinhardtii*, *The Plant Journal 82*, 717-729.

102.   Hamilton, M. L., Haslam, R. P., Napier, J. A., and Sayanova, O. (2014) Metabolic engineering of *Phaeodactylum tricornutum* for the enhanced accumulation of omega-3 long chain polyunsaturated fatty acids, *Metabolic Engineering 22*, 3-9.

103.   Jia, B., Zheng, Y., Xiao, K., Wu, M., Lei, Y., Huang, Y., and Hu, Z. (2016) A vector for multiple gene co-expression in *Chlamydomonas reinhardtii*, *Algal Research 20*, 53-56.

104. Moellering, E. R., and Benning, C. (2010) RNA Interference Silencing of a Major Lipid Droplet Protein Affects Lipid Droplet Size in *Chlamydomonas reinhardtii*, *Eukaryotic Cell 9*, 97-106.

105. Rohr, J., Sarkar, N., Balenger, S., Jeong, B.-r., and Cerutti, H. (2004) Tandem inverted repeat system for selection of effective transgenic RNAi strains in *Chlamydomonas*: Tandem inverted repeat system for efficient RNAi, *The Plant Journal 40*, 611-621.

106. De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C., and Falciatore, A. (2009) Gene silencing in the marine diatom *Phaeodactylum tricornutum*, *Nucleic Acids Research 37*, e96-e96.

107. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems, *Science 339*, 819-823.

108. Ran, F. A., Hsu, Patrick D., Lin, C.-Y., Gootenberg, Jonathan S., Konermann, S., Trevino, A. E., Scott, David A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013) Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity, *Cell 154*, 1380-1389.

109. Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016) A CRISPR/Cas9 system adapted for gene editing in marine algae, *Scientific Reports 6*, 24951.

110. Cermak, T., Curtin, S. J., Gil-Humanes, J., Čegan, R., Kono, T. J. Y., Konečná, E., Belanto, J. J., Starker, C. G., Mathre, J. W., Greenstein, R. L., and Voytas, D. F. (2017) A multi-purpose toolkit to enable advanced genome engineering in plants, *Plant Cell*, tpc.00922.02016.

111. Xie, Y., Wang, D., Lan, F., Wei, G., Ni, T., Chai, R., Liu, D., Hu, S., Li, M., Li, D., Wang, H., and Wang, Y. (2017) An episomal vector-based CRISPR/Cas9 system for highly efficient gene knockout in human pluripotent stem cells, *Scientific Reports 7*.

112. Baek, K., Kim, D. H., Jeong, J., Sim, S. J., Melis, A., Kim, J.-S., Jin, E., and Bae, S. (2016) DNA-free two-gene knockout in *Chlamydomonas reinhardtii* via CRISPR-Cas9 ribonucleoproteins, *Scientific Reports 6*, 30620.

113. Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., and Joung, J. K. (2016) High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects, *Nature 529*, 490-495.

114. Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., and Zhang, F. (2016) Rationally engineered Cas9 nucleases with improved specificity, *Science 351*, 84-88.

115. Li, X., Zhang, R., Patena, W., Gang, S. S., Blum, S. R., Ivanova, N., Yue, R., Robertson, J. M., Lefebvre, P. A., Fitz-Gibbon, S. T., Grossman, A. R., and Jonikas, M. C. (2016)

An Indexed, Mapped Mutant Library Enables Reverse Genetics Studies of Biological Processes in *Chlamydomonas reinhardtii*, *Plant Cell 28*, 367-387.

116.   Perin, G., Bellan, A., Segalla, A., Meneghesso, A., Alboresi, A., and Morosinotto, T. (2015) Generation of random mutants to improve light-use efficiency of *Nannochloropsis gaditana* cultures for biofuel production, *Biotechnology for Biofuels 8*, 161.

117.   Iwai, M., Hori, K., Sasaki-Sekimoto, Y., Shimojima, M., and Ohta, H. (2015) Manipulation of oil synthesis in *Nannochloropsis* strain NIES-2145 with a phosphorus starvation–inducible promoter from *Chlamydomonas reinhardtii*, *Frontiers in Microbiology 6*.

118.   Tian, J., Zheng, M., Yang, G., Zheng, L., Chen, J., and Yang, B. (2013) Cloning and stress-responding expression analysis of malonyl CoA-acyl carrier protein transacylase gene of *Nannochloropsis gaditana*, *Gene 530*, 33-38.

119.   Chen, J.-W., Liu, W.-J., Hu, D.-X., Wang, X., Balamurugan, S., Alimujiang, A., Yang, W.-D., Liu, J.-S., and Li, H.-Y. (2017) Identification of a malonyl CoA-acyl carrier protein transacylase and its regulatory role in fatty acid biosynthesis in oleaginous microalga *Nannochloropsis oceanica*: *Nannochloropsis* MCAT, *Biotechnology and Applied Biochemistry*.

120.   Kang, N. K., Kim, E. K., Kim, Y. U., Lee, B., Jeong, W.-J., Jeong, B.-r., and Chang, Y. K. (2017) Increased lipid production by heterologous expression of AtWRI1 transcription factor in *Nannochloropsis salina*, *Biotechnology for Biofuels 10*.

121.   Daboussi, F., Leduc, S., Maréchal, A., Dubois, G., Guyot, V., Perez-Michaut, C., Amato, A., Falciatore, A., Juillerat, A., Beurdeley, M., Voytas, D. F., Cavarec, L., and Duchateau, P. (2014) Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology, *Nature Communications 5*.

122.   Work, V. H., Radakovits, R., Jinkerson, R. E., Meuser, J. E., Elliott, L. G., Vinyard, D. J., Laurens, L. M. L., Dismukes, G. C., and Posewitz, M. C. (2010) Increased Lipid Accumulation in the *Chlamydomonas reinhardtii* sta7-10 Starchless Isoamylase Mutant and Increased Carbohydrate Synthesis in Complemented Strains, *Eukaryotic Cell 9*, 1251-1261.

123.   Trentacoste, E. M., Shrestha, R. P., Smith, S. R., Gle, C., Hartmann, A. C., Hildebrand, M., and Gerwick, W. H. (2013) Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth, *Proceedings of the National Academy of Sciences 110*, 19748-19753.

124.   Xue, Z., Sharpe, P. L., Hong, S.-P., Yadav, N. S., Xie, D., Short, D. R., Damude, H. G., Rupert, R. A., Seip, J. E., Wang, J., Pollak, D. W., Bostick, M. W., Bosak, M. D., Macool, D. J., Hollerbach, D. H., Zhang, H., Arcilla, D. M., Bledsoe, S. A., Croker, K., McCord, E. F., Tyreus, B. D., Jackson, E. N., and Zhu, Q. (2013) Production of omega-3 eicosapentaenoic acid by metabolic engineering of *Yarrowia lipolytica*, *Nature Biotechnology 31*, 734-740.

125. Kindle, K. L., Schnell, R. A., Fernández, E., and Lefebvre, P. A. (1989) Stable nuclear transformation of Chlamydomonas using the Chlamydomonas gene for nitrate reductase, *J. Cell Biol. 109*, 2589-2601.

126. Diner, R. E., Schwenck, S. M., McCrow, J. P., Zheng, H., and Allen, A. E. (2016) Genetic Manipulation of Competition for Nitrate between Heterotrophic Bacteria and Diatoms, *Frontiers in Microbiology 7*.

127. McCarthy, J. K., Smith, S. R., McCrow, J. P., Tan, M., Zheng, H., Beeri, K., Roth, R., Lichtle, C., Goodenough, U., Bowler, C. P., Dupont, C. L., and Allen, A. E. (2017) Nitrate Reductase Knockout Uncouples Nitrate Transport from Nitrate Assimilation and Drives Repartitioning of Carbon Flux in a Model Pennate Diatom, *Plant Cell 29*, 2047-2070.

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

129. Ma, X., Pan, K., Zhang, L., Zhu, B., Yang, G., and Zhang, X. (2016) Genetic transformation of *Nannochloropsis oculata* with a bacterial phleomycin resistance gene as dominant selective marker, *Journal of Ocean University of China 15*, 351-356.

130. Chen, H. L., Li, S. S., Huang, R., and Tsai, H. J. (2008) Conditional Production of a Functional Fish Growth Hormone in the Transgenic Line of *Nannochloropsis Oculata* (Eustigmatophyceae), *Journal of phycology 44*, 768-776.

131. Wei, L., Wang, Q., Xin, Y., Lu, Y., and Xu, J. (2017) Enhancing photosynthetic biomass productivity of industrial oleaginous microalgae by overexpression of RuBisCO activase, *Algal Research 27*, 366-375.

132. Matsuo, T., and Ishiura, M. (2011) Chlamydomonas reinhardtii as a new model system for studying the molecular basis of the circadian clock, *FEBS Letters 585*, 1495-1502.

133. Nikaido, S. S., and Johnson, C. H. (2000) Daily and circadian variation in survival from ultraviolet radiation in *Chlamydomonas reinhardtii*, *Photochemistry and photobiology 71*, 758-765.

134. Lemaire, S. D., Hours, M., Gerard-Hirne, C., Trouabal, A., Roche, O., and Jacquot, J.-P. (1999) Analysis of light/dark synchronization of cell-wall-less *Chlamydomonas reinhardtii* (Chlorophyta) cells by flow cytometry, *European Journal of Phycology 34*, 279-286.

135. McClung, C. R. (2011) The Genetics of Plant Clocks, *Adv Genet 74*, 105-139.

136. Dodd, A. N., Salathia, N., Hall, A., Kevei, E., Toth, R., Nagy, F., Hibberd, J. M., Millar, A. J., and Webb, A. A. (2005) Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage, *Science 309*, 630-633.

137. Ouyang, Y., Andersson, C. R., Kondo, T., Golden, S. S., and Johnson, C. H. (1998) Resonating circadian clocks enhance fitness in cyanobacteria, *Proc Natl Acad Sci U S A 95*, 8660-8664.

138. Litchman, E., Klausmeier, C. A., and Bossard, P. (2004) Phytoplankton nutrient competition under dynamic light regimes, *Limnol Oceanogr 49*, 1457-1462.

139. Adarme-Vega, T. C., Lim, D. K., Timmins, M., Vernen, F., Li, Y., and Schenk, P. M. (2012) Microalgal biofactories: a promising approach towards sustainable omega-3 fatty acid production, *Microbial cell factories 11*, 96.

140. Hu, H., and Gao, K. (2003) Optimization of growth and fatty acid composition of a unicellular marine picoplankton, *Nannochloropsis* sp., with enriched carbon sources, *Biotechnology letters 25*, 421-425.

141. Xu, F., Cai, Z. L., Cong, W., and Ouyang, F. (2004) Growth and fatty acid composition of *Nannochloropsis* sp. grown mixotrophically in fed-batch culture, *Biotechnology letters 26*, 1319-1322.

142. Van Vooren, G., Le Grand, F., Legrand, J., Cuine, S., Peltier, G., and Pruvost, J. (2012) Investigation of fatty acids accumulation in *Nannochloropsis oculata* for biodiesel application, *Bioresource Technology 124*, 421-432.

143. Zheng, M., Tian, J., Yang, G., Zheng, L., Chen, G., Chen, J., and Wang, B. (2013) Transcriptome sequencing, annotation and expression analysis of *Nannochloropsis* sp. at different growth phases, *Gene 523*, 117-121.

144. Panchy, N., Wu, G., Newton, L., Tsai, C. H., Chen, J., Benning, C., Farre, E. M., and Shiu, S. H. (2014) Prevalence, evolution, and cis-regulation of diel transcription in *Chlamydomonas reinhardtii*, *G3 4*, 2461-2471.

145. Filichkin, S. A., Breton, G., Priest, H. D., Dharmawardhana, P., Jaiswal, P., Fox, S. E., Michael, T. P., Chory, J., Kay, S. A., and Mockler, T. C. (2011) Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules, *PLoS One 6*, e16907.

146. Monnier, A., Liverani, S., Bouvet, R., Jesson, B., Smith, J., Mosser, J., Corellou, F., and Bouget, F.-Y. (2010) Orchestrated transcription of biological processes in the marine picoeukaryote *Ostreococcus* exposed to light/dark cycles, *BMC Genomics 11*, 192.

147. Farre, E. M. (2012) The regulation of plant growth by the circadian clock, *Plant Biology 14*, 401-410.

148.   Putt, M., and Prézelin, B. B. (1988) Diel Periodicity Of Photosynthesis And Cell Division Compared In *Thalassiosira Weissflogii* (Bacillariophyceae), *Journal of phycology 24*, 315-324.

149.   Moulager, M., Monnier, A., Jesson, B., Bouvet, R., Mosser, J., Schwartz, C., Garnier, L., Corellou, F., and Bouget, F. Y. (2007) Light-dependent regulation of cell division in *Ostreococcus*: evidence for a major transcriptional input, *Plant Physiology 144*, 1360-1369.

150.   Asato, Y. (2003) Toward an understanding of cell growth and the cell division cycle of unicellular photoautotrophic cyanobacteria, *Cellular and molecular life sciences : CMLS 60*, 663-687.

151.   Higashi, Y., and Seki, H. (2000) Ecological adaptation and acclimatization of natural freshwater phytoplankters with a nutrient gradient, *Environmental pollution 109*, 311-320.

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

153.   Michael, T. P., Mockler, T. C., Breton, G., McEntee, C., Byer, A., Trout, J. D., Hazen, S. P., Shen, R., Priest, H. D., Sullivan, C. M., Givan, S. A., Yanovsky, M., Hong, F., Kay, S. A., and Chory, J. (2008) Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules, *PLoS Genet 4*, e14.

154.   Panda, S., Antoch, M. P., Miller, B. H., Su, A. I., Schook, A. B., Straume, M., Schultz, P. G., Kay, S. A., Takahashi, J. S., and Hogenesch, J. B. (2002) Coordinated transcription of key pathways in the mouse by the circadian clock, *Cell 109*, 307-320.

155.   Blasing, O. E. (2005) Sugars and Circadian Regulation Make Major Contributions to the Global Regulation of Diurnal Gene Expression in Arabidopsis, *Plant Cell 17*, 3257-3281.

156.   Zinser, E. R., Lindell, D., Johnson, Z. I., Futschik, M. E., Steglich, C., Coleman, M. L., Wright, M. A., Rector, T., Steen, R., McNulty, N., Thompson, L. R., and Chisholm, S. W. (2009) Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *prochlorococcus*, *PLoS One 4*, e5135.

157.   Shi, T., Ilikchyan, I., Rabouille, S., and Zehr, J. P. (2010) Genome-wide analysis of diel gene expression in the unicellular N(2)-fixing cyanobacterium *Crocosphaera watsonii* WH 8501, *The ISME journal 4*, 621-632.

158.   Ottesen, E. A., Young, C. R., Gifford, S. M., Eppley, J. M., Marin, R., 3rd, Schuster, S. C., Scholin, C. A., and DeLong, E. F. (2014) Ocean microbes. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages, *Science 345*, 207-212.

159. Izawa, T., Mihara, M., Suzuki, Y., Gupta, M., Itoh, H., Nagano, A. J., Motoyama, R., Sawada, Y., Yano, M., Hirai, M. Y., Makino, A., and Nagamura, Y. (2011) Os-GIGANTEA confers robust diurnal rhythms on the global transcriptome of rice in the field, *Plant Cell 23*, 1741-1755.

160. Waldbauer, J. R., Rodrigue, S., Coleman, M. L., and Chisholm, S. W. (2012) Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle, *PLoS One 7*, e43432.

161. Gillard, J., Devos, V., Huysman, M. J., De Veylder, L., D'Hondt, S., Martens, C., Vanormelingen, P., Vannerum, K., Sabbe, K., Chepurnov, V. A., Inze, D., Vuylsteke, M., and Vyverman, W. (2008) Physiological and transcriptomic evidence for a close coupling between chloroplast ontogeny and cell cycle progression in the pennate diatom *Seminavis robusta*, *Plant Physiol 148*, 1394-1411.

162. Setlikova, E., Setlik, I., Kupper, H., Kasalicky, V., and Prasil, O. (2005) The photosynthesis of individual algal cells during the cell cycle of *Scenedesmus quadricauda* studied by chlorophyll fluorescence kinetic microscopy, *Photosynthesis Research 84*, 113-120.

163. Claudio, P. P., Tonini, T., and Giordano, A. (2002) The retinoblastoma family: twins or distant cousins?, *Genome biology 3*, reviews3012.

164. Nakajima, K., Tanaka, A., and Matsuda, Y. (2013) SLC4 family transporters in a marine diatom directly pump bicarbonate from seawater, *Proceedings of the National Academy of Sciences of the United States of America 110*, 1767-1772.

165. Tanaka, R., Kikutani, S., Mahardika, A., and Matsuda, Y. (2014) Localization of enzymes relating to C4 organic acid metabolisms in the marine diatom, *Thalassiosira pseudonana*, *Photosynthesis Research 121*, 251-263.

166. Ratledge, C. (2014) The role of malic enzyme as the provider of NADPH in oleaginous microorganisms: a reappraisal and unsolved problems, *Biotechnology letters 36*, 1557-1568.

167. Johnson, X., and Alric, J. (2013) Central carbon metabolism and electron transport in *Chlamydomonas reinhardtii*: metabolic constraints for carbon partitioning between oil and starch, *Eukaryotic Cell 12*, 776-793.

168. Haimovich-Dayan, M., Garfinkel, N., Ewe, D., Marcus, Y., Gruber, A., Wagner, H., Kroth, P. G., and Kaplan, A. (2013) The role of C4 metabolism in the marine diatom *Phaeodactylum tricornutum*, *The New phytologist 197*, 177-185.

169. Rousvoal, S., Groisillier, A., Dittami, S. M., Michel, G., Boyen, C., and Tonon, T. (2011) Mannitol-1-phosphate dehydrogenase activity in *Ectocarpus siliculosus*, a key role for mannitol synthesis in brown algae, *Planta 233*, 261-273.

170. Groisillier, A., Shao, Z., Michel, G., Goulitquer, S., Bonin, P., Krahulec, S., Nidetzky, B., Duan, D., Boyen, C., and Tonon, T. (2014) Mannitol metabolism in brown algae involves a new phosphatase family, *Journal of experimental botany 65*, 559-570.

171. Gravot, A., Dittami, S. M., Rousvoal, S., Lugan, R., Eggert, A., Collen, J., Boyen, C., Bouchereau, A., and Tonon, T. (2010) Diurnal oscillations of metabolite abundances and gene analysis provide new insights into central metabolic processes of the brown alga *Ectocarpus siliculosus*, *The New phytologist 188*, 98-110.

172. Smith, S. R., Abbriano, R. M., and Hildebrand, M. (2012) Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways, *Algal Research 1*, 2-16.

173. Takahashi, H., McCaffery, J. M., Irizarry, R. A., and Boeke, J. D. (2006) Nucleocytosolic acetyl-coenzyme a synthetase is required for histone acetylation and global transcription, *Molecular cell 23*, 207-217.

174. Roessler, P. G., and Ohlrogge, J. B. (1993) Cloning and characterization of the gene that encodes acetyl-coenzyme A carboxylase in the alga *Cyclotella cryptica*, *The Journal of biological chemistry 268*, 19254-19259.

175. Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., Baud, S., Bird, D., Debono, A., Durrett, T. P., Franke, R. B., Graham, I. A., Katayama, K., Kelly, A. A., Larson, T., Markham, J. E., Miquel, M., Molina, I., Nishida, I., Rowland, O., Samuels, L., Schmid, K. M., Wada, H., Welti, R., Xu, C., Zallot, R., and Ohlrogge, J. (2013) Acyl-lipid metabolism, *The Arabidopsis book / American Society of Plant Biologists 11*, e0161.

176. Vieler, A., Brubaker, S. B., Vick, B., and Benning, C. (2012) A Lipid Droplet Protein of *Nannochloropsis* with Functions Partially Analogous to Plant Oleosins, *Plant Physiology 158*, 1562-1569.

177. Krienitz, L., and Wirth, M. (2006) The high content of polyunsaturated fatty acids in *Nannochloropsis limnetica* (Eustigmatophyceae) and its implication for food web interactions, freshwater aquaculture and biotechnology, *Limnologica - Ecology and Management of Inland Waters 36*, 204-210.

178. Li, X., Moellering, E. R., Liu, B., Johnny, C., Fedewa, M., Sears, B. B., Kuo, M. H., and Benning, C. (2012) A galactoglycerolipid lipase is required for triacylglycerol accumulation and survival following nitrogen deprivation in *Chlamydomonas reinhardtii*, *Plant Cell 24*, 4670-4686.

179. Zentner, G. E., and Henikoff, S. (2013) Regulation of nucleosome dynamics by histone modifications, *Nature structural & molecular biology 20*, 259-266.

180. Eddy, S. R. (2011) Accelerated Profile HMM Searches, *Plos Comput Biol 7*.

181. Koike, N., Yoo, S. H., Huang, H. C., Kumar, V., Lee, C., Kim, T. K., and Takahashi, J. S. (2012) Transcriptional architecture and chromatin landscape of the core circadian clock in mammals, *Science 338*, 349-354.

182. Noordally, Z. B., Ishii, K., Atkins, K. A., Wetherill, S. J., Kusakina, J., Walton, E. J., Kato, M., Azuma, M., Tanaka, K., Hanaoka, M., and Dodd, A. N. (2013) Circadian control of chloroplast transcription by a nuclear-encoded timing signal, *Science 339*, 1316-1319.

183. Roberti, M., Polosa, P. L., Bruni, F., Manzari, C., Deceglie, S., Gadaleta, M. N., and Cantatore, P. (2009) The MTERF family proteins: mitochondrial transcription regulators and beyond, *Biochimica et biophysica acta 1787*, 303-311.

184. Bligh, E. G., and Dyer, W. J. (1959) A rapid method of total lipid extraction and purification, *Canadian journal of biochemistry and physiology 37*, 911-917.

185. Vieler, A., Wilhelm, C., Goss, R., Süß, R., and Schiller, J. (2007) The lipid composition of the unicellular green alga *Chlamydomonas reinhardtii* and the diatom *Cyclotella meneghiniana* investigated by MALDI-TOF MS and TLC, *Chemistry and Physics of Lipids 150*, 143-155.

186. Cavalier, D. M., Lerouxel, O., Neumetzler, L., Yamauchi, K., Reinecke, A., Freshour, G., Zabotina, O. A., Hahn, M. G., Burgert, I., Pauly, M., Raikhel, N. V., and Keegstra, K. (2008) Disrupting two *Arabidopsis thaliana* xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component, *Plant Cell 20*, 1519-1537.

187. Trapnell, C., Pachter, L., and Salzberg, S. L. (2009) TopHat: discovering splice junctions with RNA-Seq, *Bioinformatics 25*, 1105-1111.

188. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, *Nat Biotechnol 28*, 511-515.

189. Edgar, R., Domrachev, M., and Lash, A. E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, *Nucleic Acids Res 30*, 207-210.

190. Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics 26*, 139-140.

191. Cock, J. M., Sterck, L., Rouze, P., Scornet, D., Allen, A. E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J. M., Badger, J. H., Beszteri, B., Billiau, K., Bonnet, E., Bothwell, J. H., Bowler, C., Boyen, C., Brownlee, C., Carrano, C. J., Charrier, B., Cho, G. Y., Coelho, S. M., Collen, J., Corre, E., Da Silva, C., Delage, L., Delaroque, N., Dittami, S. M., Doulbeau, S., Elias, M., Farnham, G., Gachon, C. M. M., Gschloessl, B., Heesch, S., Jabbari, K., Jubin, C., Kawai, H., Kimura, K., Kloareg, B., Kupper, F. C., Lang, D., Le Bail, A., Leblanc, C., Lerouge, P., Lohr, M., Lopez, P. J., Martens, C.,

Maumus, F., Michel, G., Miranda-Saavedra, D., Morales, J., Moreau, H., Motomura, T., Nagasato, C., Napoli, C. A., Nelson, D. R., Nyvall-Collen, P., Peters, A. F., Pommier, C., Potin, P., Poulain, J., Quesneville, H., Read, B., Rensing, S. A., Ritter, A., Rousvoal, S., Samanta, M., Samson, G., Schroeder, D. C., Segurens, B., Strittmatter, M., Tonon, T., Tregear, J. W., Valentin, K., von Dassow, P., Yamagishi, T., Van de Peer, Y., and Wincker, P. (2010) The Ectocarpus genome and the independent evolution of multicellularity in brown algae, *Nature 465*, 617-621.

192.   Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res 32*, 1792-1797.

193.   Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods, *Mol Biol Evol 28*, 2731-2739.

194.   Reumann, S., Buchwald, D., and Lingner, T. (2012) PredPlantPTS1: A Web Server for the Prediction of Plant Peroxisomal Proteins, *Frontiers in plant science 3*, 194.

195.   Claros, M. G., and Vincens, P. (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences, *European journal of biochemistry / FEBS 241*, 779-786.

196.   Gruber, A., Rocap, G., Kroth, P. G., Armbrust, E. V., and Mock, T. (2015) Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage, *The Plant journal : for cell and molecular biology 81*, 519-528.

197.   Hee-Youn Hong, G.-S. Y. J.-K. C. (2000) Direct Blue 71 staining of proteins bound to blotting membranes, *Electrophoresis 21*, 841-845.

198.   Dimmic, M. W., Rest, J. S., Mindell, D. P., and Goldstein, R. A. (2002) rtREV: an amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny, *Journal of molecular evolution 55*, 65-73.

199.   Jones, D. T., Taylor Wr Fau - Thornton, J. M., and Thornton, J. M. The rapid generation of mutation data matrices from protein sequences.

200.   Bothwell, J. H. F., Marie, D., Peters, A. F., Cock, J. M., and Coelho, S. M. (2014) Cell cycles and endocycles in the model brown seaweed, *Ectocarpus siliculosus*, *Plant Signaling & Behavior 5*, 1473-1475.

201.   Whelan, S., and Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach.

202.   Chen, G.-C., Yang, J., Eggersdorfer, M., Zhang, W., and Qin, L.-Q. (2016) N-3 long-chain polyunsaturated fatty acids and risk of all-cause mortality among general populations: a meta-analysis, *Scientific Reports 6*, 28165.

203. Betancor, M. B., Sprague, M., Usher, S., Sayanova, O., Campbell, P. J., Napier, J. A., and Tocher, D. R. (2015) A nutritionally-enhanced oil from transgenic *Camelina sativa* effectively replaces fish oil as a source of eicosapentaenoic acid for fish, *Sci Rep 5*, 8104.

204. Petrie, J. R., and Singh, S. P. (2011) Expanding the docosahexaenoic acid food web for sustainable production: engineering lower plant pathways into higher plants, *AoB PLANTS 2011*, plr011.

205. Doughman, S. D., Krupanidhi, S., and Sanjeevi, C. B. (2007) Omega-3 fatty acids for nutrition and medicine: considering microalgae oil as a vegetarian source of EPA and DHA, *Curr Diabetes Rev 3*, 198-203.

206. Martins, D., Custódio, L., Barreira, L., Pereira, H., Ben-Hamadou, R., Varela, J., and Abu-Salah, K. (2013) Alternative Sources of n-3 Long-Chain Polyunsaturated Fatty Acids in Marine Microalgae, *Marine Drugs 11*, 2259-2281.

207. Chini Zittelli, G., Lavista, F., Bastianini, A., Rodolfi, L., Vincenzini, M., and Tredici, M. R. (1999) Production of eicosapentaenoic acid by *Nannochloropsis* sp. cultures in outdoor tubular photobioreactors, *Journal of Biotechnology 70*, 299-312.

208. Spolaore, P., Joannis-Cassan, C., Duran, E., and Isambert, A. (2006) Commercial applications of microalgae, *Journal of Bioscience and Bioengineering 101*, 87-96.

209. Sukenik, A., Schneider, J. C., Roessler, P. G., Livne, A., Berner, T., Kolber, Z., Wyman, K., Prasil, O., and Falkowski, P. G. (1998) Photosynthetic Characterization of a Mutant of *Nannochloropsis* Deficient in the Synthesis of Eicosapentaenoic Acid, *Israel Journal of Plant Sciences 46*, 101-108.

210. Liu, B., Vieler, A., Li, C., Daniel Jones, A., and Benning, C. (2013) Triacylglycerol profiling of microalgae *Chlamydomonas reinhardtii* and *Nannochloropsis oceanica*, *Bioresource Technology 146*, 310-316.

211. Qi, B., Fraser, T., Mugford, S., Dobson, G., Sayanova, O., Butler, J., Napier, J. A., Stobart, A. K., and Lazarus, C. M. (2004) Production of very long chain polyunsaturated omega-3 and omega-6 fatty acids in plants, *Nature Biotechnology 22*, 739-745.

212. Hong, H. (2002) High-Level Production of gamma -Linolenic Acid in *Brassica juncea* Using a Delta 6 Desaturase from *Pythium irregulare*, *PLANT PHYSIOLOGY 129*, 354-362.

213. Wu, G., Truksa, M., Datla, N., Vrinten, P., Bauer, J., Zank, T., Cirpus, P., Heinz, E., and Qiu, X. (2005) Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants, *Nat Biotechnol 23*, 1013-1017.

214. Okuda, T., Ando, A., Negoro, H., Muratsubaki, T., Kikukawa, H., Sakamoto, T., Sakuradani, E., Shimizu, S., and Ogawa, J. (2015) Eicosapentaenoic acid (EPA) production by an oleaginous fungus *Mortierella alpina* expressing heterologous the Δ17-

desaturase gene under ordinary temperature, *European Journal of Lipid Science and Technology 117*, 1919-1927.

215. Tavares, S., Grotkjaer, T., Obsen, T., Haslam, R. P., Napier, J. A., and Gunnarsson, N. (2011) Metabolic Engineering of *Saccharomyces cerevisiae* for Production of Eicosapentaenoic Acid, Using a Novel 5-Desaturase from *Paramecium tetraurelia*, *Applied and Environmental Microbiology 77*, 1854-1861.

216. Hwangbo, K., Ahn, J.-W., Lim, J.-M., Park, Y.-I., Liu, J. R., and Jeong, W.-J. (2014) Overexpression of stearoyl-ACP desaturase enhances accumulations of oleic acid in the green alga *Chlamydomonas reinhardtii*, *Plant Biotechnology Reports 8*, 135-142.

217. Peng, K.-T., Zheng, C.-N., Xue, J., Chen, X.-Y., Yang, W.-D., Liu, J.-S., Bai, W., and Li, H.-Y. (2014) Delta 5 Fatty Acid Desaturase Upregulates the Synthesis of Polyunsaturated Fatty Acids in the Marine Diatom *Phaeodactylum tricornutum*, *Journal of Agricultural and Food Chemistry 62*, 8773-8776.

218. Hildebrand, M., Abbriano, R. M., Polle, J. E. W., Traller, J. C., Trentacoste, E. M., Smith, S. R., and Davis, A. K. (2013) Metabolic and cellular organization in evolutionarily diverse microalgae as related to biofuels production, *Current Opinion in Chemical Biology 17*, 506-514.

219. Cook, O., and Hildebrand, M. (2015) Enhancing LC-PUFA production in *Thalassiosira pseudonana* by overexpressing the endogenous fatty acid elongase genes, *Journal of Applied Phycology 28*, 897-905.

220. Ruiz-Lopez, N., Haslam, R. P., Napier, J. A., and Sayanova, O. (2014) Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop, *The Plant journal : for cell and molecular biology 77*, 198-208.

221. Ruiz-Lopez, N., Sayanova, O., Napier, J. A., and Haslam, R. P. (2012) Metabolic engineering of the omega-3 long chain polyunsaturated fatty acid biosynthetic pathway into transgenic plants, *Journal of experimental botany 63*, 2397-2410.

222. Metz, J. G. (2001) Production of Polyunsaturated Fatty Acids by Polyketide Synthases in Both Prokaryotes and Eukaryotes, *Science 293*, 290-293.

223. Cahoon, E. B., Shockey, J. M., Dietrich, C. R., Gidda, S. K., Mullen, R. T., and Dyer, J. M. (2007) Engineering oilseeds for sustainable production of industrial and nutritional feedstocks: solving bottlenecks in fatty acid flux, *Current Opinion in Plant Biology 10*, 236-244.

224. Halpin, C. (2005) Gene stacking in transgenic plants - the challenge for 21st century plant biotechnology: The challenge of transgene and trait stacking, *Plant biotechnology journal 3*, 141-155.

225. Naqvi, S., Farré, G., Sanahuja, G., Capell, T., Zhu, C., and Christou, P. (2010) When more is better: multigene engineering in plants, *Trends in Plant Science 15*, 48-56.

226. Szymczak, A. L., Workman, C. J., Wang, Y., Vignali, K. M., Dilioglou, S., Vanin, E. F., and Vignali, D. A. A. (2004) Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide–based retroviral vector, *Nature Biotechnology 22*, 589-594.

227. Szymczak-Workman, A. L., Vignali, K. M., and Vignali, D. A. A. (2012) Design and Construction of 2A Peptide-Linked Multicistronic Vectors, *Cold Spring Harbor Protocols 2012*, pdb.ip067876-pdb.ip067876.

228. Kim, J. H., Lee, S.-R., Li, L.-H., Park, H.-J., Park, J.-H., Lee, K. Y., Kim, M.-K., Shin, B. A., and Choi, S.-Y. (2011) High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice, *PLoS ONE 6*, e18556.

229. de Felipe, P., Hughes, L. E., Ryan, M. D., and Brown, J. D. (2003) Co-translational, Intraribosomal Cleavage of Polypeptides by the Foot-and-mouth Disease Virus 2A Peptide, *Journal of Biological Chemistry 278*, 11441-11448.

230. López Alonso, D., García-Maroto, F., Rodríguez-Ruiz, J., Garrido, J. A., and Vilches, M. A. (2003) Evolution of the membrane-bound fatty acid desaturases, *Biochemical Systematics and Ecology 31*, 1111-1124.

231. Broun, P. (1998) Catalytic Plasticity of Fatty Acid Modification Enzymes Underlying Chemical Diversity of Plant Lipids, *Science 282*, 1315-1317.

232. Domergue, F., Lerchl, J., Zähringer, U., and Heinz, E. (2002) Cloning and functional characterization of *Phaeodactylum tricornutum* front-end desaturases involved in eicosapentaenoic acid biosynthesis: Δ5- and Δ6-fatty acid desaturases from diatom, *European Journal of Biochemistry 269*, 4105-4113.

233. Meesapyodsuk, D., and Qiu, X. (2012) The Front-end Desaturase: Structure, Function, Evolution and Biotechnological Use, *Lipids 47*, 227-237.

234. Hashimoto, K., Yoshizawa, A. C., Okuda, S., Kuma, K., Goto, S., and Kanehisa, M. (2008) The repertoire of desaturases and elongases reveals fatty acid variations in 56 eukaryotic genomes, *The Journal of Lipid Research 49*, 183-191.

235. Tvrdik, P., Westerberg, R., Silve, S., Asadi, A., Jakobsson, A., Cannon, B., Loison, G., and Jacobsson, A. (2000) Role of a new mammalian gene family in the biosynthesis of very long chain fatty acids and sphingolipids, *J. Cell Biol. 149*, 707-718.

236. Oh, C. S., Toke, D. A., Mandala, S., and Martin, C. E. (1997) ELO2 and ELO3, homologues of the *Saccharomyces cerevisiae* ELO1 gene, function in fatty acid elongation and are required for sphingolipid formation, *The Journal of biological chemistry 272*, 17376-17384.

237. Jiang, M., Guo, B., Wan, X., Gong, Y., Zhang, Y., and Hu, C. (2014) Isolation and Characterization of the Diatom Phaeodactylum Δ5-Elongase Gene for Transgenic LC-PUFA Production in *Pichia pastoris*, *Marine Drugs 12*, 1317-1334.

238. Yu, S. Y., Li, H., Tong, M., Ouyang, L. L., and Zhou, Z. G. (2012) Identification of a Δ6 fatty acid elongase gene for arachidonic acid biosynthesis localized to the endoplasmic reticulum in the green microalga *Myrmecia incisa Reisigl*, *Gene 493*, 219-227.

239. Moller, S., Croning, M. D. R., and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions, *Bioinformatics 17*, 646-653.

240. Stukey, J. E., McDonough, V. M., and Martin, C. E. (1990) The OLE1 gene of *Saccharomyces cerevisiae* encodes the delta 9 fatty acid desaturase and can be functionally replaced by the rat stearoyl-CoA desaturase gene, *The Journal of biological chemistry 265*, 20144-20149.

241. Gill, D. R., Smyth, S. E., Goddard, C. A., Pringle, I. A., Higgins, C. F., Colledge, W. H., and Hyde, S. C. (2001) Increased persistence of lung gene expression using plasmids containing the ubiquitin C or elongation factor 1alpha promoter, *Gene Ther. 8*, 1539-1546.

242. Kim, D. W., Uetsuki, T., Kaziro, Y., Yamaguchi, N., and Sugano, S. (1990) Use of the human elongation factor 1α promoter as a versatile and efficient expression system, *Gene 91*, 217-223.

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

244. Doronina, V. A., Wu, C., de Felipe, P., Sachs, M. S., Ryan, M. D., and Brown, J. D. (2008) Site-specific release of nascent chains from ribosomes at a sense codon, *Molecular and cellular biology 28*, 4227-4239.

245. Burén, S., Ortega-Villasante, C., Ötvös, K., Samuelsson, G., Bakó, L., and Villarejo, A. (2012) Use of the Foot-and-Mouth Disease Virus 2A Peptide Co-Expression System to Study Intracellular Protein Trafficking in Arabidopsis, *PLoS ONE 7*, e51973.

246. Wang, Y., Wang, F., Wang, R., Zhao, P., and Xia, Q. (2015) 2A self-cleaving peptide-based multi-gene expression system in the silkworm *Bombyx mori*, *Scientific Reports 5*, 16273.

247. Rasala, B. A., Lee, P. A., Shen, Z., Briggs, S. P., Mendez, M., and Mayfield, S. P. (2012) Robust Expression and Secretion of Xylanase1 in *Chlamydomonas reinhardtii* by Fusion to a Selection Gene and Processing with the FMDV 2A Peptide, *PLoS ONE 7*, e43349.

248. Ma, X., Yu, J., Zhu, B., Pan, K., Pan, J., and Yang, G. (2011) Cloning and characterization of a delta-6 desaturase encoding gene from *Nannochloropsis oculata*, *Chinese Journal of Oceanology and Limnology 29*, 290-296.

249. Valentine, R. C., and Valentine, D. L. (2004) Omega-3 fatty acids in cellular membranes: a unified concept, *Progress in Lipid Research 43*, 383-402.

250. Moire, L. (2004) Impact of Unusual Fatty Acid Synthesis on Futile Cycling through -Oxidation and on Gene Expression in Transgenic Plants, *Plant Physiology 134*, 432-442.

251. Manandhar-Shrestha, K., and Hildebrand, M. (2015) Characterization and manipulation of a DGAT2 from the diatom *Thalassiosira pseudonana*: Improved TAG accumulation without detriment to growth, and implications for chloroplast TAG accumulation, *Algal Research 12*, 239-248.

252. Xu, J., Kazachkov, M., Jia, Y., Zheng, Z., and Zou, J. (2013) Expression of a type 2 diacylglycerol acyltransferase from *Thalassiosira pseudonana* in yeast leads to incorporation of docosahexaenoic acid β-oxidation intermediates into triacylglycerol, *FEBS Journal 280*, 6162-6172.

253. Kajiwara, S., Shirai, A., Fujii, T., Toguri, T., Nakamura, K., and Ohtaguchi, K. (1996) Polyunsaturated fatty acid biosynthesis in *Saccharomyces cerevisiae*: expression of ethanol tolerance and the FAD2 gene from *Arabidopsis thaliana*, *Applied and Environmental Microbiology 62*, 4309-4313.

254. Rodolfi, L., Chini Zittelli, G., Bassi, N., Padovani, G., Biondi, N., Bonini, G., and Tredici, M. R. (2009) Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor, *Biotechnology and bioengineering 102*, 100-112.

255. Kumar, S. (2015) *GM Algae for Biofuel Production: Biosafety and Risk Assessment*, Vol. 9.

256. Lu, J., Tong, Y., Pan, J., Yang, Y., Liu, Q., Tan, X., Zhao, S., Qin, L., and Chen, X. (2016) A redesigned CRISPR/Cas9 system for marker-free genome editing in *Plasmodium falciparum*, *Parasites & Vectors 9*.

257. Weninger, A., Hatzl, A.-M., Schmid, C., Vogl, T., and Glieder, A. (2016) Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*, *Journal of Biotechnology 235*, 139-149.

258. Weyman, P. D., Beeri, K., Lefebvre, S. C., Rivera, J., McCarthy, J. K., Heuberger, A. L., Peers, G., Allen, A. E., and Dupont, C. L. (2015) Inactivation of *Phaeodactylum tricornutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis, *Plant biotechnology journal 13*, 460-470.

259. Jiang, W., Brueggeman, A. J., Horken, K. M., Plucinak, T. M., and Weeks, D. P. (2014) Successful Transient Expression of Cas9 and Single Guide RNA Genes in *Chlamydomonas reinhardtii*, *Eukaryotic Cell 13*, 1465-1469.

260. Shin, S.-E., Lim, J.-M., Koh, H. G., Kim, E. K., Kang, N. K., Jeon, S., Kwon, S., Shin, W.-S., Lee, B., Hwangbo, K., Kim, J., Ye, S. H., Yun, J.-Y., Seo, H., Oh, H.-M., Kim, K.-J., Kim, J.-S., Jeong, W.-J., Chang, Y. K., and Jeong, B.-r. (2016) CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*, *Scientific Reports 6*, 27810.

261.   Hopes, A., Nekrasov, V., Kamoun, S., and Mock, T. (2016) Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*, *Plant Methods 12*.

262.   Zhang, W.-W., and Matlashewski, G. (2015) CRISPR-Cas9-Mediated Genome Editing in *Leishmania donovani*, *mBio 6*, e00861-00815.

263.   Moon, D. A., and Goff, L. J. (1997) Molecular characterization of two large DNA plasmids in the red alga *Porphyra pulchra*, *Curr. Genet. 32*, 132-138.

264.   Fernández, E., Schnell, R., Ranum, L. P., Hussey, S. C., Silflow, C. D., and Lefebvre, P. A. (1989) Isolation and characterization of the nitrate reductase structural gene of *Chlamydomonas reinhardtii*, *Proceedings of the National Academy of Sciences of the United States of America 86*, 6449-6453.

265.   Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nature methods 6*, 343-345.

266.   Waldrip, Z. J., Byrum, S. D., Storey, A. J., Gao, J., Byrd, A. K., Mackintosh, S. G., Wahls, W. P., Taverna, S. D., Raney, K. D., and Tackett, A. J. (2014) A CRISPR-based approach for proteomic analysis of a single genomic locus, *Epigenetics 9*, 1207-1211.