

A "CRISPY" TALE
ON MICROALGAL GENOME EDITING:
featuring *Nannochloropsis oceanica* IMET1

MIHRIS IBNU SALEEM NADUTHODI

Propositions

1. Improved influx of carbon and efficient generation of reducing power is vital to improve microalgal productivities.
(this thesis)
2. Simultaneous engineering at multiple levels of metabolism will be required to produce commercially feasible biofuels and green chemicals from microalgae.
(this thesis)
3. Microbial production of biofuels are not limited by the availability of genetic tools, but by the motivation to develop ambitious engineering strategies to exploit their potential.
4. Duration of PhD programs need not be same for all projects as novelty of a species and its doubling time could impact the final quality of a thesis in a fixed time frame.
5. Scientific literacy among politicians is crucial towards societal acceptance of scientific advances.
6. Collaborations between scientists and artists are essential for disseminating complex scientific details to a broader audience.

Propositions belonging to the thesis, entitled

A "CRISPY" tale on microalgal genome editing: featuring *Nannochloropsis oceanica* IMET1

Mihris Ibnu Saleem Naduthodi
Wageningen,
13 December 2021

A “CRISPY” tale on microalgal
genome editing: featuring
Nannochloropsis oceanica IMET1

Mihris Ibnu Saleem Naduthodi

Thesis committee

Promotors

Prof. Dr John van der Oost
Personal chair, Laboratory of Microbiology
Wageningen University & Research

Prof. Dr Maria J. Barbosa
Personal chair, Bioprocess Engineering
Wageningen University & Research

Co-promotor

Dr Sarah D'Adamo
Assistant Professor, Bioprocess Engineering
Wageningen University & Research

Other members

Prof. Dr Luisa Trindade - Wageningen University & Research
Dr Rob Brown – Synthetic Genomics, La Jolla [CA], USA
Prof. Dr Luke Mackinder – University of York, UK
Dr Lolke Sijtsma – Wageningen University & Research

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrotechnology, Nutrition and Health Sciences)

A “CRISPY” tale on microalgal
genome editing: featuring
Nannochloropsis oceanica IMET1

THESIS

submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus,

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Monday 13 December 2021

at 11 a.m. in the Aula

A "CRISPY" tale on microalgal genome editing: featuring *Nannochloropsis oceanica* IMET1
248 Pages

PhD thesis, Wageningen University, Wageningen, NL (2021)
With references, with summary in English

ISBN 978-94-6395-918-6

DOI <https://doi.org/10.18174/555308>

Table of contents

Abbreviations	6
CHAPTER 1 General Introduction	9
CHAPTER 2 Progress of CRISPR-Cas based genome editing in photosynthetic microbes	21
CHAPTER 3 CRISPR–Cas ribonucleoprotein mediated homology-directed repair for efficient targeted genome editing in microalgae <i>Nannochloropsis oceanica</i> IMET1	35
CHAPTER 4 A comprehensive genome engineering toolbox for microalgae <i>Nannochloropsis oceanica</i> based on CRISPR-Cas systems	61
CHAPTER 5 High-throughput insertional mutagenesis reveals novel targets for enhancing lipid accumulation in <i>Nannochloropsis oceanica</i>	97
CHAPTER 6 Heterologous expression of EPA specific Lysophosphatidic acid acyltransferase in <i>Nannochloropsis oceanica</i> increases the EPA content in Triglycerides	141
CHAPTER 7 Synthetic biology approaches to enhance microalgal productivity	159
CHAPTER 8 General Discussion	185
References	203
Thesis Summary	236
About the Author	240
Overview of completed training activities	241
List of Publications	242
Acknowledgements	243

Abbreviations

ACCase: Acetyl-CoA carboxylase	EDTA: Ethylenediaminetetraacetic acid
ACBP: Acetyl-CoA binding proteins	ER: Endoplasmic reticulum
ACP: Acyl carrier protein	EMP: Emden-Meyerhof-Parnas
ARA: Arachidonic acid	FA: fatty acids
ARS: Autonomous replicating sequence	FAD: Fatty acid desaturase
ASW: Artificial sea water	FFA: Free fatty acids
ABCD: ATP binding cassette domain	FAS: Fatty acid synthesis
ALE: Adaptive laboratory evolution	FACS: Fluorescence activated cell sorting
BE: Base editing	FW: Forward
BnLPAAT: <i>Brassica napus</i> LPAAT	HR: Homologous recombination
CACT: Carnitine	FPLC: Fast protein liquid chromatography
acylcarnitine translocase	Fructose 1,6-bisphosphatase
CBB: Calvin-Benson-Bassham	PGK: Phosphoglycerate kinase
CCM: Carbon concentrating mechanism	GPT: glucose-6-phosphate translocator
CETCH: Crotonyl-CoA/ethylmalonyl-CoA/ hydroxybutyryl-CoA	GPAT: glycerol-3-phosphate acyltransferase
Chl: Chlorophyll	G3P: Glyceraldehyde-3-phosphate
cTP: Chloroplast targeting peptide	GAPDH: Glyceraldehyde 3-phosphate dehydrogenase
Cas: CRISPR associated proteins	GDH: Glycolate dehydrogenase
CEN: Centromere	GCL: Glyoxylate carboxyligase
CRISPR- Clustered regularly interspaced short palindromic repeats	GMO: Genetically modified organism
CRISPRi: CRISPR interference	HDR: Homology directed repair
crRNA: CRISPR RNA	HH: Hammer head
CDW: cell dry weight	HDV: Hepatitis delta virus
cPCR: Colony PCR	HLM: High lipid producing mutants
CDS: Coding DNA sequence	IC: Insertion cassette
CaMV: Cauliflower Mosaic Virus	IR: Infrared
DCW: Dry cell weight	IPTG: Isopropyl β -d-1- thiogalactopyranoside
dCas9: dead Cas9	IEX: Ion exchange
DSB: Double stranded breaks	KRAB: Kruppel associated box
dsDNA: Double stranded DNA	LPAAT: Lysophosphatidic acid acyltransferase
DR: Direct repeat	LPA: Lysophosphatidic acid
DGDG: Digalactosyldiacylglycerol	LDSP: Lipid droplet surface protein, and
DAG: Diacylglycerol	LC-PUFA: Long chain polyunsaturated fatty acids
DHA: Docosahexanoic acid	LL: Low light
DGAT: DAG acyltransferase	ML: Medium light
DGTS: Diacylglycerol-trimethyl-homo- serine	mTP: Mitochondrion targeting peptide
ETC: Electron transport chain	MGDG: Monogalactosyldiacylglycerol
EF: Elongation factor	MES: Microbial electrosynthesis
EPA: Eicosapentaenoic acid	

MOG: Malonyl-CoA/oxaloacetate/
glyoxylate
MAT: Malonyl-CoA ACP transacylase
MS: Malate synthase
MUFA: Mono unsaturated fatty acids
NHEJ: Non-homologous end joining
NPQ: Non-photochemical quenching
NR: Nitrate reductase
NL: Neutral lipids
Nlux: Nano luciferase
nCas9: Nickase Cas9
OD: Optical density
TAG: Triacylglycerol
UEP: Ubiquitin extension protein
PAM: Protospacer adjacent motif
PCR: Polymerised chain reaction
PE: Prime editing
pegRNA: Prime editing guide RNA
PBS: Primer binding site
PE: Photosynthetic efficiency
PS: Photosystem
PE: Phosphatidylethanolamine
VCP: Violaxanthin chlorophyll-binding
proteins
PL: Polar lipids
Ribi: Ribosomal subunit
TALEN: Transcription activator like effector
nuclease
PUFA: Polyunsaturated fatty acids
PGM: Phosphoglycerate mutase
PDC: Pyruvate decarboxylase
PFK: Phosphofructokinase
PG: Phosphoglycolate
PEP: Phosphoenolpyruvate
PAR: Photosynthetically active radiation
RNP: Ribonucleoproteins
RV: Reverse
ROS: Reactive oxygen species
RT: Reverse transcription
REN: Restriction endonuclease
SQDG: Sulfoquinovosyl diacylglycerol
SFA: Saturated fatty acids
ssDNA: Single stranded DNA
SQD2: Sulphoquinovosyltransferase
sgRNA: single guide RNA
SA: Signal anchor

SP: Signal peptide
TE: Thioesterase
TF: Transcription factor
TFA: Total fatty acids
tracrRNA: Trans-activating RNA
TSR: Tartronic-semialdehyde reductase
TLC: Thin layer chromatography
TCA: Tricarboxylic acid cycle
UGDH: UDP-glucose dehydrogenase
UTR: Untranslated region
UV: Ultraviolet
YFP: Yellow fluorescent protein
ZeoR: Zeocin resistance
ZFN: Zinc finger nuclease,

CHAPTER 1



General Introduction



Microalgae: food, feed and much more

During a Trans-Saharan expedition conducted in 1964-1965 by Belgium Ministry [1], botanist Prof. Leonard found interesting greenish cakes called “*dihe*” in markets of the country Chad. He later confirmed that this cake has been consumed in the region for centuries and was made of blue-green algae *Spirulina* [1]. Microalgae are predominantly photosynthetic microorganisms that use inorganic carbon (CO₂) and sunlight available in their native environment to multiply. Studies into the 16th-century chroniclers revealed the consumption of microalgae to the 1300s [2]. The inhabitants of Tenochtitlan (present Mexico City), the capital of the Aztec Empire, ate a sun-dried blue coloured muddy substance called “*Tecuitlatl*” that tasted like cheese and was inferred to be blue-green algae by researchers. In the modern era, traditional delicacies based on microalgae are still consumed in Japan, China, Mongolia, Tartaria and South America. They are also used as dietary components in various South-East-Asian countries [3, 4]. These findings indicate the timeless use of microalgae as a food source in the ancient and modern world.

In the 1940s, in addition to obtaining microalgae from natural environments, advancements were made in their mass production in open ponds and specialized photobioreactors [3], leading to successful commercialization of microalgae as a sustainable food supplement. Additionally, various microalgal strains were also used as feed for animals and aquaculture [5]. Further studies on microalgae revealed their ability to naturally produce valuable bioactive compounds, thereby expanding the application of microalgae into nutraceutical industries. The diverse bioactive compounds produced from microalgae are summarized in Table 1.

Table 1. Various products obtained from microalgae

Bioactive compounds	Microalgae	Reference
Vitamin B12, Niacin, Folate and Vitamin B2	<i>Chlorella</i> , <i>Spirulina</i>	[3, 6]
Vitamin E	<i>Euglena gracilis</i> , <i>Dunaliella tertiolecta</i> and <i>Tertraselmis suecica</i>	[7]
Beta-carotene	<i>Dunaliella</i>	[8]
Astaxanthin	<i>Haematococcus</i> , <i>Chlorella zofingiensis</i> , <i>Chlorococcum</i> sp.	[9-11]
Lutein	<i>Chlorella</i> sp, <i>Scenedesmus</i> sp and <i>Muriellopsis</i> sp	[9, 12]
Zeaxanthin	<i>Dunaliella</i> , <i>Synechococcus</i> , <i>Chlorella pyrenoidosa</i> , <i>Acaryochloris marina</i> , <i>Prochlorococcus marinus</i> , <i>Prochlorothrix hollandica</i> and <i>Chlorella saccharophila</i>	[13]
Canthaxanthin	<i>Chlorella zofingiensis</i>	[14]
Fucoxanthin	<i>Mallomonas</i> sp. and <i>Phaeodactylum tricornutum</i>	[15, 16]
Phytoene	<i>Dunaliella salina</i> , <i>Dunaliella bardawil</i> and <i>Chlorococcum</i> sp.	[17-19]
Phytofluene	<i>Dunaliella bardawil</i>	[20]
Violaxanthin and Antheraxanthin	<i>Chlorella</i>	[21]
Echinenone	<i>Botryococcus braunii</i>	[22]

Bioactive compounds	Microalgae	Reference
Phycobiliproteins (Phycocyanin, Phycoerythrin and Allophycocyanin)	<i>Spirulina</i> and <i>Porphyridium</i>	[23]
Chlorophyll A	<i>Aphanizomenon flos-aquae</i>	[24]
PUFAs (EPA, DHA, linoleic acid, oleic acid and lauric acid)	<i>Phaeodactylum</i> , <i>Nannochloropsis</i> , <i>Schizochytrium</i> , <i>Isochrysis</i> and <i>Pavlova</i>	[25]
Polysaccharides	<i>Porphyridium</i> and <i>Nostoc</i>	[26]
Sterols	<i>Dunaliella</i>	[27]
Lipids	<i>Nostoc</i> , <i>Ulkenia</i> and <i>Nannochloropsis</i>	[3, 28]

Microalgae: a sustainable source of lipids and fatty acids

In addition to high-value nutraceuticals, some microalgae accumulates lipids in considerable amounts that can be converted into biofuels via processes such as transesterification in a microalgal biorefinery (Figure 1) [29-32]. Under optimal growth conditions, microalgal metabolism is channeled towards the production of fatty acids (FAs). The precursor fatty acids (Palmitic acid and Oleic acid) undergo multiple rounds of chain elongation and desaturation, resulting in the accumulation of different types of fatty acids in microalgae. These FAs primarily undergo esterification with glycerol to form the membrane lipids. Glycosylglycerides accumulated in the chloroplast and phosphoglycerides present in the plasma membrane and endoplasmic reticulum are the major membrane lipids present in microalgae [31, 33-36]. However, many microalgae switch their lipid metabolism under sub-optimal and divert the carbon flux towards the accumulation of neutral lipids, mainly triacylglycerol (TAG).

In contrast to the structural role of membrane lipids, TAG serves as a carbon and energy storage and can be an alternative to oil crops such as palm oil, without leading to deforestation. In contrast to terrestrial oil crops, microalgae can be cultivated in non-arable lands without competing with food crops and have higher biomass productivities [36]. Due to these characteristics, microalgae were also studied as a potential source for sustainable production of biofuels [36-39].

Nevertheless, the high cost of production associated with the phototrophic cultivation of microalgae has been a significant bottleneck in commercializing microalgal biofuels. The stress induction to triggering the TAG accumulation in microalgae adversely affects biomass production and reduces the overall productivity [40-42]. These bottlenecks can be resolved by improving the microalgal photosynthetic efficiency and maximizing the partitioning of carbon into lipids without affecting the growth. A summary of all the major metabolic engineering studies in microalgae focused on improving the TAG accumulation is described in Chapter 5. Despite decades of research, a commercially feasible TAG producing microalgal strain is yet to be developed. Earlier engineering studies that resulted in improved lipid production were achieved in model microalgal strains such as *Chlamydomonas reinhardtii*, *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* [43-48]. However, the lipid productivities of these organisms were very low to achieve commercial feasibility. This observation diverted the focus towards oleaginous microalgal

strains such as *Nannochloropsis* spp that has six-fold higher lipid production compared to the model strains reaching up to 60% of its cells dry weight (CDW) [49, 50]

Table 2. Lipid contents of selected microalgae

Species	Lipid Content (% of CDW)	Reference
<i>Nannochloropsis</i>	40-60	[51]
<i>Chlamydomonas reinhardtii</i>	20-30	[52]
<i>Isochrysis</i>	25-33	[53]
<i>Dunaliella salina</i>	23	[54]
<i>Phaeodactylum tricornutum</i>	20-30	[55]
<i>Spirulina platensis</i>	7-8	[56]
<i>Tetraselmis obliquus</i>	14-16	[57]

***Nannochloropsis* spp.**

The genus *Nannochloropsis* was termed by Hibberd in 1981, and these microalgae are mainly found in aquatic environments, including saline, fresh and brackish waters [58-60]. They belong to the phylum of Heterokontophyte or Stramenopiles and are further classified as Eustigmatophytes. Eustigmatophytes are unicellular microalgae that are coccoid in shape and have a yellow-green appearance. *Nannochloropsis* constitutes the smallest organisms (2-8µm in diameter) in this class and are considered "sibling species" as it is challenging to differentiate them morphologically. Until now, seven variants (*N. australis*, *N. gaditana*, *N. granulata*, *N. limnetica*, *N. oceanica*, *N. oculata* and *N. salina*) of this genus are identified and differentiated by 18s rRNA sequencing [59, 60].

The marine strains of this genus have a reported specific growth rate that ranges between 0.11 d⁻¹ and 0.21 d⁻¹ [61]. Under optimal conditions, the lipids constitute only between 5% and 20% of the CDW in *Nannochloropsis*. However, under sub-optimal conditions, the lipid contents are between 37% and 60% of the CDW, with TAGs specifically accounting for up to 50% of the CDW. The lipid productivity of 158 mg/L/day and TAG accumulation of 1.67 g/L in these strains made them stand out from other microalgal strains [36, 61, 62]. Additionally, genetic accessibility that facilitates the development of genetic toolbox and the availability of large-scale outdoor cultivation techniques makes them potential microalgal strains for large scale production of biofuels [50, 51, 63]. Furthermore, *Nannochloropsis* strains have a unique fatty acid profile that accumulates large amounts of omega-3 fatty acids, specifically Eicosapentaenoic acid (EPA) [64, 65]. Omega-3 fatty acids are used as a food supplement to treat cardiovascular diseases and various other medical conditions [66]. They are one of the most prescribed food supplements worldwide with an exponential growth in market demand [67]. The omega-3 fatty acids are conventionally obtained from marine fish, which are unable to synthesize them, but obtain them by consuming microalgae, the primary producers of these fatty acids. In addition to direct human consumption, they are also used in aquafeeds to improve the omega-3 fatty acid

content in farmed fish [68]. The increasing demands for omega-3 fatty acids resulted in unsustainable wild fish capture, leading to a significant decline of the ocean wildlife population [68]. Moreover, the ocean pollution has raised concern regarding the quality of the wild fish for human consumption [68].

Microalgae being the primary producers of omega-3 fatty acids, could be used as a sustainable alternative for wild fish in this regard [28]. *Nannochloropsis* strains accumulate up to 12% of the cell dry weight (CDW) with EPA indicating its potential as a sustainable source for omega-3 fatty acid production [69]. Physiologically, in contrast to the TAG accumulation, the EPA production in *Nannochloropsis* was reported to be higher under optimal conditions that do not retard any growth properties [70]. Hence, *Nannochloropsis* could naturally be a more suitable strain for commercial EPA production.

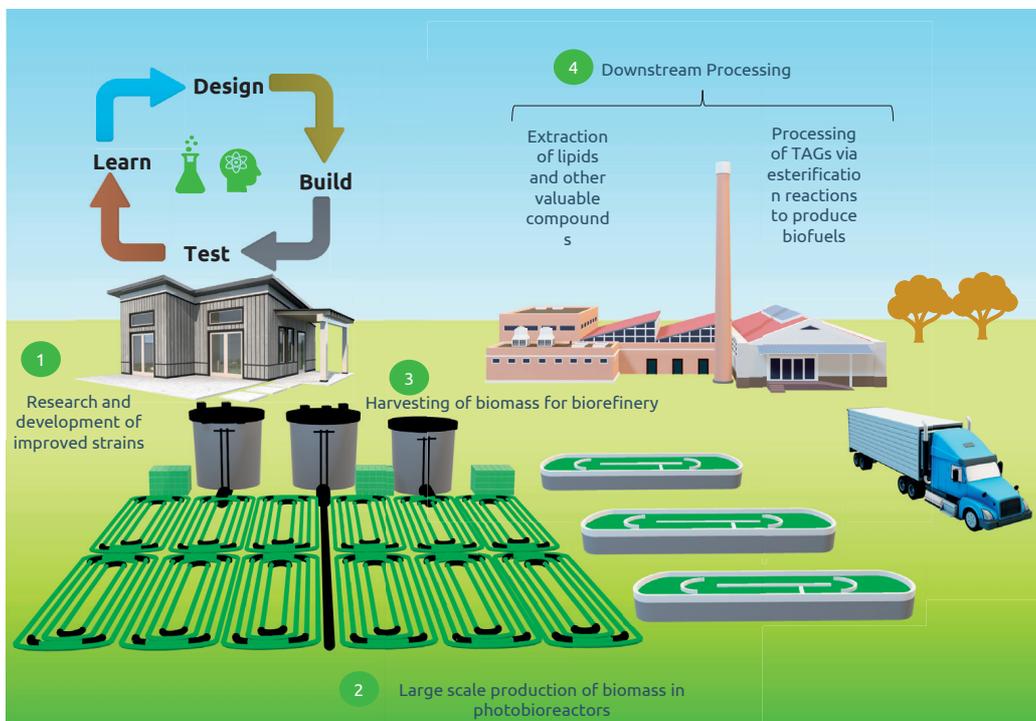


Figure 1. A Microalgal biorefinery

Genome assembly to engineering *Nannochloropsis* spp.

The unique metabolism of *Nannochloropsis* that accumulate valuable lipids and fatty acids has generated scientific interest to understand the metabolic pathways in this genus. Generating a complete and annotated genome sequence is the basis of drafting a metabolic map. Genome assemblies of various strains within the genus of *Nannochloropsis* are reported, such as *N. gaditana* CCMP526 [49], *N. gaditana* B-31 [71], *N. oceanica* CCMP1779 [63], *N. oceanica* IMET1, *N. salina* CCMP537, *N. oculata* CCMP525 and *N. granulata* CCMP529 [72]. These studies showed that *Nannochloropsis* contains 7,000 to 11,000 genes on its approximately 30 mega-bases long genomes segregated as multiple chromosomes (30 chromosomes for *N. gaditana* B-31 and 22 for *N. oceanica* IMET1) [71, 72]. The genome assembly and annotation of genes were followed by transcriptomic studies to further elucidate the metabolic pathways. The transcriptome of the strains was studied under various (nutrient starvation [63, 73, 74], light conditions [75, 76] and diverse growth phases [49]) conditions to reveal the genes that are associated with metabolism and their response to environmental conditions.

The knowledge on the metabolic pathways and their regulation suggested compelling metabolic engineering strategies to improve the production of lipids and nutraceuticals from *Nannochloropsis*. Consequently, these developments initiated the establishment of the genetic toolbox for *Nannochloropsis*. A reliable transformation protocol, resistance markers and regulatory elements to drive the expression of transgenes are the foundation of developing a genetic toolbox.

Among various transformation protocols, electroporation has been the widely used method to deliver DNA and proteins into *Nannochloropsis* due to its high efficiency and user-friendly protocol [49, 63, 77, 78]. Initially, linear DNAs were transformed for random integration into the host genome and to obtain stable expression of heterologous proteins. Recently, circular DNA based episomal plasmids were also developed for transforming *Nannochloropsis* strains [79]. Various antibiotics and respective resistance markers (Zeocin, Blastidicin and Hygromycin) for selecting the transformants of *Nannochloropsis* were identified along with the development of transformation protocols [49, 63, 78, 80]. Several endogenous promoters were also characterized to express the transgenes in *Nannochloropsis* such as the promoters driving the expression of ubiquitin extension protein (UEP), lipid droplet surface protein (LDSP), elongation factor (EF), violaxanthin chlorophyll-binding proteins (VCP), ribosomal subunits (Ribi) and beta-tubulin [49, 63, 78, 81].

These developments enabled the initial metabolic engineering studies in *Nannochloropsis* that required overexpression or heterologous expression of specific genes. Next, the genetic toolbox of *Nannochloropsis* was further elaborated by adding techniques that facilitated targeted gene disruption. Homologous recombination-based target gene disruption has been a straightforward approach for generating specific knockout and knock-ins in *Nannochloropsis* [78]. Targeted gene disruption using the engineered nucleases such as Zinc Finger Nucleases and Transcription Activator Like Effector Nuclease (TALEN) was reported for model microalgal species *Chlamydomonas reinhardtii* and

Phaeodactylum tricornutum [48, 82]. However, these technologies are laborious, expensive and complicated and were only recently reported in non-model microalgal strains, including *Nannochloropsis* [83].

Generating knockouts could be a lethal approach while targeting essential genes. In that case, gene repression could be an alternative technique that generates mutants with a moderate phenotype [80]. To this end, RNA interference has been a widely used gene suppression tool in microalgae [84-87]. This technique was also adapted and applied in *Nannochloropsis* strains [80, 88-90].

CRISPR-Cas systems

The clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated (Cas) protein systems emerged as a universal genome editing tool that simplified the process of genetic engineering across organisms. Fundamentally, CRISPR-Cas system is an adaptive immune system present in approximately 45% and 83% of sequenced bacteria and archaea respectively [91]. The CRISPR array contains short fragments of phage DNAs called spacers that are acquired during previous infections. This array is transcribed and processed to generate short RNA molecules that guides the Cas effector proteins towards the phage DNAs during secondary infection and destroy them by specifically introducing double stranded breaks (DSBs). This specificity of DSB induction on phages is based on the spacer sequences on the RNA molecule. Altering the spacer sequence and developing synthetic guide RNA molecules, facilitates the adaptation of this system to specifically target other DNA molecules. In 2013, the Cas endonuclease protein from *Streptococcus pyogenes* termed as SpCas9 was used for engineering prokaryotic and eukaryotic genomes [92-96]. Additionally, catalytically inactive variant of this Cas9 protein was used for transcriptional repression and activation of the target genes [97-99]. Since, multiple variants of Cas endonuclease were isolated and characterized from various species such as *Staphylococcus aureus* (SaCas9), *Streptococcus thermophilus* (StCas9), *Neisseria meningitidis* (NmCas9) and *Campylobacter jejuni* (CjCas9). Recently, Cas9 proteins isolated from thermophilic bacteria *Geobacillus stearothermophilus* (GeoCas9) and *Geobacillus thermodenitrificans* (ThermoCas9) with improved thermostability was used for efficiently engineering thermophiles [100, 101].

The Cas12a endonucleases were recently identified as a distinct type of Cas effector proteins more suited for genome engineering, compared to Cas9 [102]. Relatively smaller size and the ability to process the CRISPR array simplifies the multiplexed genome editing using Cas12a [103]. Other distinct features of Cas12a compared to Cas9 is described in chapter 2. Predominantly, three Cas12a variants from *Francisella novicida* (FnCas12a), *Lachnospiraceae bacterium* (LbCas12a) and *Acidaminococcus* spp (AsCas12a) are used for genome editing in both prokaryotes and eukaryotes. All these Cas12a variants and SpCas9 were used for generating targeted mutants in *Nannochloropsis* as described in Chapter 3 and 4. Later on, the extraordinary ability of Cas proteins to specifically bind double stranded DNA was creatively exploited to develop state of the art genome engineering techniques such as base-editing, prime editing and CRISPR-on/off. An attempt to implement the

prime-editing strategy in *Nannochloropsis* is reported in Chapter 4. However, further optimizations are required to effectively use this system in *Nannochloropsis*.

Nannochloropsis strains have the required characteristics to become a platform for lipid production. The high cost associated with present phototrophic cultivation technologies restricts the commercial production of lipids from these microalgae. Extensive and innovative metabolic engineering such as improving the carbon fixation and energy transfer at the upstream part of the metabolism to structured channelling of carbon towards lipid synthesis without limiting the growth will be required to establish *Nannochloropsis* as a commercial lipid production platform. Even though the proposed strategy is ambitious, the genome editing tools developed in this thesis enable extensive metabolic engineering in *Nannochloropsis*.

Thesis Outline

As described in **Chapter 1**, realizing the potential of microalgae to accumulate valuable bioactive compounds has drawn commercial and scientific attention to these organisms. *Nannochloropsis* stand out from other microalgal strains due to its ability to accumulate large amounts of lipids and omega-3 fatty acids. Nevertheless, the current productivity of lipids are not sufficient to achieve commercial feasibility. The available genetic tools limited the metabolic engineering studies to improve lipid productivity. This thesis describes the development of CRISPR-Cas based genetic tools that can facilitate extensive genome engineering in *Nannochloropsis*.

Chapter 2 reviews the development of CRISPR-Cas based genome editing tools for microalgae, diatoms, and cyanobacteria. The CRISPR-Cas system was first reported in microalga *Chlamydomonas* in 2016. We discuss the bottlenecks associated with implementing this genome editing tool in microalgae and suggest strategies to overcome them.

Chapter 3 describes the development of a CRISPR-Cas ribonucleoprotein (RNP) based genome editing strategy for *Nannochloropsis*. Homologous recombination-based genome engineering efficiency in *Nannochloropsis* was dependent on factors such as gene loci in the genome and the cell cycle phase during transformation. Thus, the generation of mutants using this approach can be laborious and unpredictable. The introduction of DSB using Cas proteins at the target site improved the HR in the presence of the editing template and simplified the generation of mutants. We compared various Cas proteins, such as the most widely used SpCas9 and three variants of Cas12a proteins, to identify the best Cas protein for editing the *Nannochloropsis* genome.

The mutants generated by this approach harboured an antibiotic resistance cassette at the target site. **Chapter 4** addresses this bottleneck by developing a genome editing technique that removes the antibiotic resistance cassette from the target site based on fluorescence-activated cell sorting (FACS). Moreover, we developed an episomal plasmid-mediated CRISPR system that employs the Cas12a protein to generate indels and knockout the target gene efficiently. Subsequently, the capability of Cas12a protein to mature the CRISPR array was utilized to perform multiplexed targeting and induce indels at three loci in a single transformation. The multiplexed genome editing using Cas12a processed CRISPR array is reported for the first time in microalgae and significantly simplifies the high throughput genome engineering in *Nannochloropsis*. This technique can also be easily adapted to other microalgal strains. Furthermore, a CRISPRi system based on dead variants of Cas9 and Cas12a proteins were used for efficiently downregulating the target gene expression. These developments reported in Chapter 3 and 4 set forth a comprehensive genome editing toolbox for *Nannochloropsis*. These advancements can facilitate the fundamental studies into the metabolism of *Nannochloropsis* to produce commercially relevant mutants.

Identifying novel target genes that can be engineered to improve the lipid accumulation is crucial for elucidating the microalgal lipid metabolism. Generating a random mutant

library, followed by isolation and genotypic characterisation of high lipid producing mutants could reveal novel genes associated with lipid production. **Chapter 5** describes this approach in *Nannochloropsis oceanica* to obtain various high lipid producing strains. The transformants were screened for high lipid mutants by staining the lipids and selection by FACS. The genotypic characterization of the mutants revealed novel genes associated with lipid production in *Nannochloropsis oceanica*.

In addition to lipids, *Nannochloropsis* is a potential platform for producing PUFAs such as EPA. In *Nannochloropsis*, EPA is mainly found at the sn-3 position of TAG. In **Chapter 6**, we attempt to improve the EPA content in TAG by expressing an EPA specific lysophosphatidic acid acyltransferases (*LPAAT*) gene from plant *Brassica napus*. This study analyses and quantifies the various lipid classes and profiles of the BnLPAAT expressed *N. oceanica* strains.

Albeit many studies, a commercially relevant strain for microalgal biofuel production is yet to be developed. **Chapter 7** critically assess the present and past metabolic engineering strategies and describes novel synthetic biology approaches to improve microalgal productivities. We highlight the importance of engineering the upstream metabolism to generate commercially relevant microalgal strains. The study proposes innovative approaches to improve the carbon assimilation and generation of reducing power in microalgae to improve their overall productivities.

Chapter 8 summarizes the advancements made in this thesis with possible future research lines that can improve the genetic toolbox for microalgae and strategies that can improve the microalgal productivity. We have also discussed the unique metabolism of *Nannochloropsis oceanica* focussing on future studies and strategies to improve the TAG and EPA production.

CHAPTER 2

2



Progress of CRISPR-Cas based genome editing in photosynthetic microbes

Mihris Ibnu Saleem Naduthodi^{1,2}, Maria J Barbosa², John van der Oost*¹

¹Laboratory of Microbiology, Wageningen University and Research, Wageningen, Netherlands.

²Bioprocess Engineering, AlgaePARC, Wageningen University and Research, Wageningen, Netherlands.

*Corresponding Author

*Chapter adapted from publication:
Biotechnol J. 2018 Sep;13(9):e1700591. doi: 10.1002/biot.201700591*

Abstract

The carbon footprint caused by unsustainable development and its environmental and economic impact has become a major concern in the past few decades. Photosynthetic microbes such as microalgae and cyanobacteria are capable of accumulating value-added compounds from carbon dioxide and have been regarded as environmentally friendly alternatives to reduce the usage of fossil fuels, thereby contributing to reducing the carbon footprint. This light-driven generation of green chemicals and biofuels has triggered the research for metabolic engineering of these photosynthetic microbes. CRISPR-Cas systems are successfully implemented across a wide range of prokaryotic and eukaryotic species for efficient genome editing. However, the inception of this genome editing tool in microalgal and cyanobacterial species took off rather slowly due to various complications. In this review, we elaborate on the established CRISPR-Cas based genome editing in various microalgal and cyanobacterial species. The complications associated with CRISPR-Cas based genome editing in these species are addressed along with possible strategies to overcome these issues. It is anticipated that in the near future this will result in improving and expanding the microalgal and cyanobacterial genome engineering toolbox.

Keywords

Cas12a (Cpf1); Cas9; CRISPR-Cas; Genome editing; microalgae; cyanobacteria.

Introduction

The exploitation of non-renewable energy sources to meet the perpetual requirement of increasing human population has resulted in their rapid depletion and a steady rise in price. Moreover, their uncontrolled usage has resulted in an elevated CO₂ concentration in the atmosphere, resulting in global warming and associated problems. In this scenario, it is of utmost importance to find environmentally friendly alternatives to meet the demands of the developing world. The potential of deploying photosynthetic microbes such as microalgae and cyanobacteria for production of next generation fuels is being studied extensively [104]. The ability to fix CO₂ and convert it into value added compounds without competing with food and feed crops make these green microbes promising biofuel producing organisms from an environmental perspective [105]. However, studies on microalgae to exploit their complete potential for commercial application has not proceeded fast mainly because of the lack of industrial strains. Efficient genome editing tools for microalgae are still lacking. Conventional homologous recombination-based gene editing was reported in genera like *Nannochloropsis* and *Ostreococcus*. The Zinc Finger Nucleases (ZFN) were applied in *Chlamydomonas reinhardtii* for achieving targeted gene editing, while Transcription Activator Like Effector Nucleases (TALEN) were implemented in *Phaeodactylum tricornutum* [78, 106-108]. Nevertheless, the labour intensiveness, complexity and high cost for application was a bottleneck for implementing ZFN and TALEN for genome engineering. Since 2014, CRISPR-Cas based genome editing has been reported in various microalgal species, which will be the prime focus of this review [109].

The genome editing of both prokaryotes and eukaryotes has been simplified with the introduction of RNA guided nucleases of CRISPR-Cas systems (Clustered regularly interspaced short palindromic repeats and associated proteins)[92, 110-112]{Charpentier, 2014, Harnessing CRISPR-Cas9 immunity for genetic engineering}. These endonucleases use short CRISPR-derived RNA guides to target complementary DNA. After recognition of a protospacer adjacent motif (PAM) sequence positioned next to the target sequence, the Cas nuclease (Cas9, Cpf1/Cas12a) introduces double strand DNA breaks (DSBs). Mutating the catalytic active site of Cas9 protein results in a dead Cas9 (dCas9) [113] that has been implemented for CRISPRinterference (CRISPRi) by stably binding, for instance to the promoter region of a target gene, and thereby downregulating the transcription of a target gene [97, 113]. The CRISPR systems are divided into two major classes based on their architecture: class I systems (types I, III and IV) form multi subunit protein crRNA-binding nuclease complexes, and class II system (type II,V and VI) consist of a single guide-binding nuclease protein [114]. Initially, the Cas9 from *Streptococcus pyogenes* (SpCas9, archetype of the type II CRISPR-Cas system) and related Cas9 variants have been widely applied for genome engineering. Recently, Cas12a (Cpf1) of the type V system is also gaining global attention for genome editing in various species [102, 115-123]. Cas12a is an interesting alternative tool for genome engineering as it has distinct features compared to Cas9 (Figure 1), such as (i) Cas12a uses a single crRNA instead of a set of crRNA and trans-activating crRNA (tracrRNA) in Cas9 that is synthetically fused as a single guide RNA (sgRNA) that is at least twice as long as the crRNA guide used by Cas12a, (ii) Cas12a has been demonstrated to catalyze the maturation of its own crRNA which allows for efficient multiplex genome editing [122, 124], while maturation of crRNA:tracrRNA complex of Cas9 relies on processing by the non-Cas ribonuclease RNase

III, (iii) Cas12a uses a T-rich PAM (5'-TTTN-3') upstream the protospacer in contrast to the downstream located G-rich PAM (5'-NGG-3') that is recognized by Cas9 [102], (iv) Cas12a generates staggered ends with 5 nucleotide overhangs in the target DNA, compared to the blunt end cleavage by Cas9 [102, 125], (v) Cas12a has a single nuclease domain (RuvC) that cleaves 18-23 base pairs downstream from the PAM-proximal seed sequence, whereas cleavage by the two nuclease domains of Cas9 (RuvC and HNH) occurs within the seed 3 base pairs upstream its PAM (Figure 1).

These mechanistic differences may have important practical advantages, such as directly using small (42-66 nt), commercially produced crRNA guides, and the potential to perform multiplex genome editing owing to the self-maturation of a precursor crRNA guide. In addition, it has been proposed that the fact that Cas12a cleavage occurs outside its seed region (Figure 1) [102], may enhance the frequency of HDR as the seed sequence is not destroyed by Non-Homologous End Joining (NHEJ) in eukaryotes, and hence cleavage could continue after NHEJ repair until HDR is successful (substituting essential nucleotides in PAM and/or seed). This review elaborates on the application of CRISPR-Cas based genome engineering in microalgae and cyanobacteria and provides insights on strategies to enhance efficiency of the CRISPR-Cas tools for generating targeted mutants in these microbes.

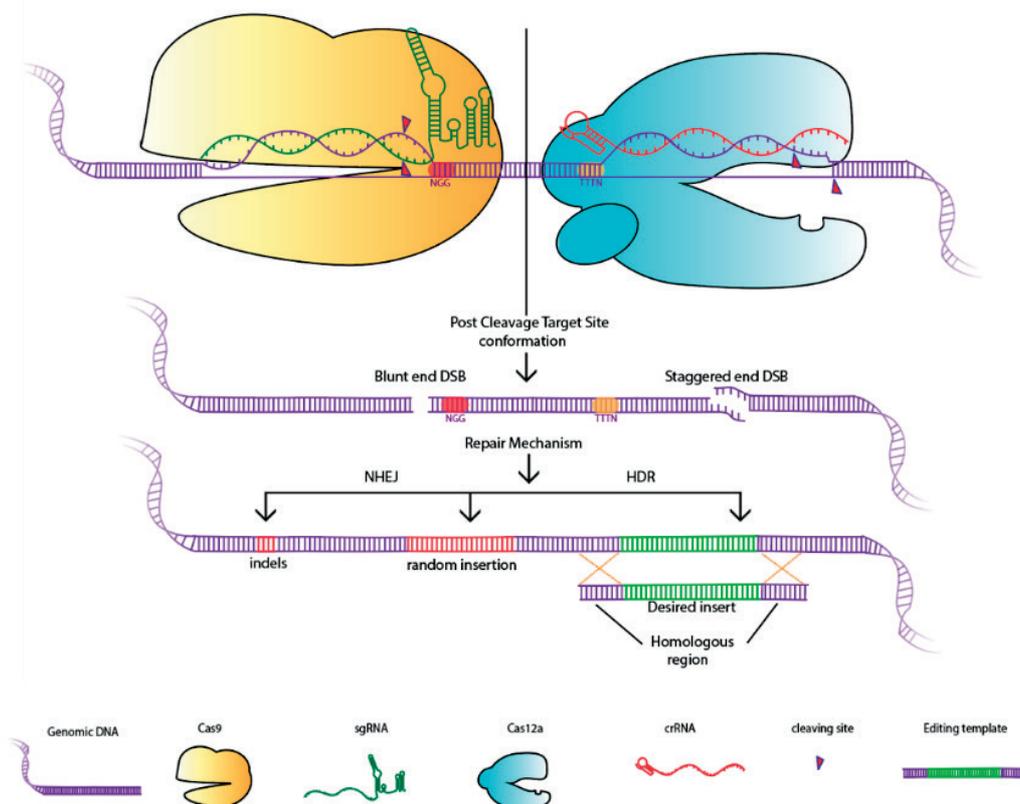


Figure 1. Schematic representation of DSB generation by Cas9 and Cas12a, and repair mechanisms present in eukaryotes for restoring the DSB.

Chlamydomonas reinhardtii

The completely sequenced and annotated genome, the genetic accessibility, the ease of generating and screening mutants owing to its haploid nature and decades of research on cellular and molecular level has made *Chlamydomonas reinhardtii* an exceptional model organism in the field of micro-algal research [126]. The first CRISPR-Cas9 based genome editing in microalgae was reported in *C. reinhardtii* [109]. Plasmid-based delivery of a codon optimized SpCas9 gene and specific sgRNAs to *Chlamydomonas reinhardtii* has resulted in transient expression and targeting restriction sites of various exogenous genes [109]. The exogenous target gene as well as the cas9 and sgRNA genes were delivered into the host on the same plasmid. After transformation, the target loci were PCR amplified after which the obtained amplicons were subjected to restriction digestion with the enzyme cleaving the target site. Introduction of indels at the target site by NHEJ upon Cas9 cleavage was hypothesized to result in the fragments' resistance to the specific restriction digestion. Sequencing of the fragments resistant to this restriction digestion revealed indels in the region of Cas9 cut site, confirming Cas9 nuclease activity. However, an experiment involving multiple transformations and more than 10^9 cells for targeting an endogenous gene (*FKB12*) by Cas9 resulted in a single mutant colony, demonstrated a rather low efficiency of this approach for generating frame-shift mutants. The cytotoxicity of Cas9 in *C. reinhardtii* was inferred for this problem as Western blot analysis failed to detect even minute levels of either Cas9 or dCas9 [109]. Recently, Jiang et al. managed to enhance the efficiency of plasmid-based Cas9 activity in *C. reinhardtii* by implementing a hybrid version of Cas9 [127]. In this work, a gene-within-a-gene approach was used in which an artificial intron sequence harboring a sgRNA was incorporated into the Cas9 coding sequence. The activity of this system was validated by restoring the abolished reading frame in an exogenously supplied mutant antibiotic resistance gene. Indeed, some recombinants gained the capacity to grow on plates with the corresponding antibiotic, indicating Cas9 cleavage and repair by non-homologous end joining (NHEJ) [127]. The targeting of the hybrid Cas9 system was reported to be 50 times more efficient compared to the previous version, as targeting the same gene with this approach yielded 13 mutants among 4×10^8 cells transformed [109, 127]. However, the observation that none of the mutants obtained by the hybrid-Cas9 approach harboured an intact Cas9 gene, strongly suggested toxicity of Cas9 [127]. In the same study, it has also been tested to combine homologous recombination and Cas9 nuclease activity to achieve precise, error-less gene editing. The transformation with the hybrid-Cas9 and short single stranded DNA repair fragments with single nucleotide mismatches at the Cas9 target site resulted in an average of 6 correct mutants upon when two different genes were targeted independently, whereas no mutants were obtained when the homologous ssDNA was transformed alone [127]. This relatively low homologous recombination frequencies in *C. reinhardtii* (in the absence of Cas9) were in agreement with earlier observations [128]. Also, the induction of DSBs has been reported to enhance the homologous recombination efficiency in various eukaryotes which could have occurred in *C. reinhardtii* facilitating the formation of these 6 expected mutants [92, 94, 129-131].

The plasmid-based expression of two Cas9 variants (spCas9 from *Streptococcus pyogenes* and saCas9 from *Staphylococcus aureus*) in *C. reinhardtii* and their efficiency in producing indel mutations was compared by Greiner et al [132]. In the experiment targeting the *PSY1*

gene by both Cas9 variants independently, application of SaCas9 yielded 9% mutants among the entire antibiotic resistant transformants obtained, outperforming SpCas9 which yielded 3% mutants. The efficiency of targeting by SaCas9 was further improved to 16% by recovering the cells after transformation for one day at 33°C and another day at 22°C before transferring to the antibiotic media plate. The expression of the SaCas9-encoding gene in this experiment was controlled by the HSP70A promoter that typically upregulates the expression of downstream genes after short heat shocks (typically at 40 °C) [133]. Greiner et al [132] suggests that the enhanced efficiency in this case could be due to the high Cas9 expression from HSP70A promoter at the elevated recovery temperature (33 °C). Also, SaCas9 derived from the mesophilic host *Staphylococcus aureus* with optimal growth temperature ranging from 37°C to 40°C [134, 135] might have its ideal activity at higher temperatures, thereby contributing to the improved efficiency. Four photoreceptor genes were successfully targeted by this approach using SaCas9 and, contradictory to SpCas9, intact sequences of the SaCas9-encoding gene were detected from engineered strains, most likely suggesting its reduced toxicity [132].

Kao et al. performed CRISPRi and observed decaying mRNA levels of Sp_dCas9 upon sub-culturing of strains up to 7 generations [136]. Linear phenotypic changes were also observed with gradual recovery of the expression of targeted genes in subsequent cultures, suggesting decreased expression of the dCas9 gene in the course of this experiment [136]. These results differ from the conclusions of Jiang et al., as successful CRISPRi by Kao et al. indicated effective Cas9 expression. In both the studies Cas9 proteins were expressed under same promoter and terminator (Cauliflower Mosaic Virus, CMV-35S). Nevertheless, the Cas9 gene sequences used in the two studies varied up to 14% indicating the effect of codon harmonization on gene expression in *C. reinhardtii*.

As an alternative delivery method, the CRISPR-Cas nuclease proteins purified from (bacterial) production systems, loaded with appropriate guides in vitro, after which the obtained ribonucleoprotein (RNP) complexes can be transformed to the host of interest. The fact that bacterial production systems are used, circumvent codon optimization for each to-be-edited host. An additional advantage may be that off-target problems are reduced because of a limited half-life of the RNP complex and saturation of the Cas protein with pre-loaded guide RNAs. Indeed, Cas9-RNP delivery to *C. reinhardtii* has substantially resolved issues with plasmid-based Cas9 genome editing [137, 138]. Efficiency of about 1% in generating indels at chromosomal target sites was observed by this approach, even though it varied considerably depending on both the gene targeted and the guide RNA sequence used [137-139]. Co-transformation of antibiotic resistance genes along with RNP to enhance the selection of transformants resulted in NHEJ-based knock in of antibiotic resistance genes at the target sites [137]. This opens up opportunities to achieve targeted knock in of gene of interest either by NHEJ or by Homology Directed Repair (HDR). Realizing the limited efficiency of the plasmid-based approach, Greiner et al. also adopted the RNP-based approach for targeting genes which upon disruption could result in non-selectable phenotypes, such as disruption of photoreceptor genes [132]. To facilitate HDR and achieve errorless DNA repair at an RNP target site, various donor species (plasmids, dsDNA and ssDNA) with 30 bps upstream and downstream homologous flanks were transformed along with the RNP complexes. The linear dsDNA template efficiently

integrated to the cleavage site of RNP by NHEJ, whereas the ssDNA template recombined less efficiently but with high chances of flawless HDR. The RNP along with various DNA editing templates were used to disrupt up to eight genes in *C. reinhardtii*, and again the efficiency of achieving each mutants appeared to be gene/guide dependent [132]. The Cas12a RNP was also used for demonstrating its efficacy in *C. reinhardtii* [140]. The transformation of Cas12a RNP alone targeting *FKB12* gene depicted targeting efficiency of 0.02% similar to Cas9 RNP. However, remarkable improvement in editing efficiency was observed when ssDNA was co-transformed along with Cas12a RNA for HDR resulting in 29% transformants and 10% errorless HDR mutants [140]. Apart from *FKB12*, three other genes were also targeted with the same approach for obtaining knockout mutants with an efficiency of 0.5 to 16% and errorless HDR based knockouts with an efficiency of 0.1-10% [140].

***Nannochloropsis* spp.**

Nannochloropsis spp belonging to the class Eustigmatophyceae are considered as model micro-algal strains primarily due to their simple and small genome organization (genome size approximately 30Mb compared to 120Mb of *C. reinhardtii*) [71, 141-145]. The industrial relevance for these organisms arise from their potential to accumulate large amounts of tri-acylglycerol (TAG) and poly-unsaturated fatty acids (PUFAs) along with tolerance to wide environmental conditions, such as temperature and light variations [145]. Vector driven CRISPR-Cas9 based genome editing has been reported in *Nannochloropsis oceanica* and *Nannochloropsis gaditana* [146, 147]. In *N. oceanica*, the frequency of indel mutants generated by Cas9 nuclease targeting the nitrate reductase gene was found to be 1.2% by next generation sequencing. Correspondingly, screening 300 colonies from the transformants resulted in only 2 confirmed mutants indicating the very low efficiency of Cas9 in the strains [146]. Even though reverse transcriptase PCR indicated stable transcripts of Cas9 gene, Western blot analysis could not validate the presence of Cas9 protein indicating the toxicity or feeble expression of Cas9 as observed in *C. reinhardtii* [146]. In *C. reinhardtii* there were knock in events observed at the Cas9 cleavage site after NHEJ repair, this was not revealed by deep sequencing analysis of *N. oceanica* mutants [137, 138, 146]. Instead, up to 0.3% of the sequenced PCR amplicons of the target site after transformation did not align with the wild type target site sequence, which might be random pieces of DNA or the cassette used for transformation inserted at the target site during NHEJ repair. The cytotoxic effects of vector driven expression of Cas9 was not observed in *N. gaditana* cultures, thereby favoring efficient targeted genome editing in the host [147]. Ajjawi et al. initially developed a *N. gaditana* strain named Ng-Cas9+, that successfully expressed the Cas9 protein from an integrated codon harmonized gene. Co-transformation of an sgRNA guide targeting the gene of interest and a hygromycin resistance cassette to strain Ng-Cas9+ and subsequent selection for growth on hygromycin plates resulted in colonies harboring the hygromycin resistance cassette inserted at the Cas9 cleavage site [147]. This strategy was used to independently knockdown 18 genes in *N. gaditana* in search for the negative regulators of lipid accumulation. Attenuating the transcription factor Zn_2Cys_6 , a homolog of fungal $Zn(II)_2Cys_6$ DNA binding domain protein using this engineering approach, doubled the lipid production in *N. gaditana* [147]. Colony PCR screening of transformants that appeared on hygromycin plates revealed that the efficiency of obtaining the 18 mutant genes ranged from 6-78% [147].

Diatoms

Extending from the equator to regions covered with ice, diatoms thrive in a wide range of marine ecosystems, contributing substantially to global photosynthesis [148]. Because of their natural capability to accumulate value added compounds along with other environmental and nanotechnology applications, diatoms are considered industrially important microbes [148, 149]. The development of CRISPR-Cas based tools for diatoms were aimed towards developing relatively cost effective and easy genome editing tools for studying fundamental diatom biology via reverse genetics as a basis for obtaining improved strains for biotechnology applications [150]. To date, CRISPR-Cas systems has been reported to be successfully applied in diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* [150, 151]. In *P. tricornutum*, a codon optimized Cas9 and sgRNA modules were transformed into the host on the same vector without a selectable marker. However, a pAF6 plasmid with zeocin resistance was co-transformed along with Cas9 and sgRNA vector to validate the transformation. Analysis of the transformants by high resonance melting analysis combined with sequencing indicated a mutation frequency of up to 31% for sgRNA targeting the gene *CpSRP54* (chloroplast signal recognition particle 54). Nymark et al. also reports to have achieved a mutation frequency of 25-63% while targeting two other genes in the same species. As diatoms are diploid organism, NHEJ repair of a Cas9-dependent DSB in one allele, may lead to HDR in the other allele, resulting in identical bi-allelic mutants [150]. Presence of mixed mutants have been observed in some of the colonies which was hypothesized to be due to the occurrence of cell division prior to initial mutation event [150].

Unlike the electroporation transformation used in *Chlamydomonas* and *Nannochloropsis*, biolistic bombardment was used for transformation of diatoms which could fragment the vectors resulting in incomplete introduction of Cas9 gene into the host [150-152]. Also, co-transformation of selectable and non-selectable vectors has resulted in 40% of the transformants to harbour only the plasmid with selectable marker [153]. To partially resolve these issues, in *T. pseudonana*, the human codon optimized cas9 gene reported to work in plants [154-156] and 2 sgRNAs targeting 37 nucleotides apart on the urease gene were transformed on a plasmid also harbouring an antibiotic resistance marker [151]. After transformation by micro-particle bombardment, only 12% (4/33) of the obtained colonies harboured an intact cas9 gene; all the colonies in which the cas9 gene was present did have mutations at the target site [151]. Out of the 4 colonies (M1, M2, M3 and M4) screened by PCR, the M4 colony was found to be a clean mutant with 37nt deletion between the two target sites. The M2 and M3 colonies showed the presence of both wild type and mutants with 37nt deletion which was confirmed to be mosaic colonies upon screening the sub-clones. Sequence analysis of the M1 colony revealed the presence of a mono-allelic 4 bp deletion at one of the sgRNA cleavage sites [151]. Overall, the fact that the majority of the mutants were bi-allelic with the designed 37bp deletion [151], indicates efficient Cas9 editing in *T. pseudonana*. The promising Cas9 activity in *T. pseudonana* was applied to achieve efficient HDR and efficiently obtain designed mutants [157]. Co-transforming the Cas9 and sgRNA on one plasmid with an editing template harbouring >500bp upstream and downstream homologous sequences and an antibiotic resistance marker between the homologous flanks successfully resulted in transformant colonies with 85% HDR

efficiency [157]. The efficient HR along with Cas9 nuclease obtained by this approach is a promising step towards high throughput genome engineering of *T. pseudonana*.

Cyanobacteria

The tractability of a prokaryotic organism combined with the photosynthetic capability of eukaryotic microalgae, promises cyanobacteria as a potential cell factory. The efficient conversion of solar energy into biomass, and requirement of low levels of carbon dioxide compared to eukaryotic microalgae result in high growth rates. Moreover, the transformation efficiency, sometimes exploiting natural competence, further elevates the possibilities of evolving selected cyanobacterial strains into platform organisms for producing biofuels and green chemicals [158, 159]. The application of CRISPR-Cas9 based genome engineering in cyanobacteria was first reported for *Synechocystis* sp. PCC 6803 where the inactive dCas9 was implemented for target gene downregulation [160]. Apart from successful single knockdown of genes coding for GFP, polyhydroxyalkanoate synthase and ADP-glucose phosphorylase with varying efficiencies, simultaneous knockdown of up to 4 genes coding for putative aldehyde dehydrogenases/ reductases were also achieved in this species [160]. CRISPRi-mediated gene downregulation has also been reported for *Synechococcus elongatus* PCC 7942 where the expression of an exogenous eYFP gene has been suppressed down to 1% of the control, and the expression of an endogenous gene *glgC* down to 6% [161]. Two genes encoding subunits of the succinate dehydrogenase (*sdhA* and *sdhB*) were also down-regulated to 19% and 33% compared to control levels resulting in an approximate 12.5% increase in succinate production [161]. Once the activity of CRISPRi in *Synechococcus* sp. PCC 7002 was confirmed by downregulating the heterologous YFP expression to 0.02% of the control, native genes encoding subunits of the Phycobilisome (*cpcB*) and the Carboxysome (*ccmK1*) were successfully downregulated [162]. The moderate repression of the glutamine synthetase I (*glnA*) gene by CRISPRi doubled the lactate production in strain PCC 7002, through a series of metabolic processes enhancing the flux of carbon towards pyruvate [162]. In *Anabaena* sp. PCC 7120, the *glnA* gene was downregulated by 80% using the CRISPRi, resulting in accumulation of ammonium [163]. The *devH* gene essential for heterocyst development in PCC 7120 was also repressed using CRISPRi, completely abolishing the growth under nitrogen fixing conditions [163]. The cytotoxic effects of Cas9 protein was observed in strains UTEX 2973 and PCC 7942 of the species *Synechococcus elongatus* [164, 165]. The constitutive expression of the Cas9 protein was found to be toxic in UTEX 2973 even at the minimal levels; this problem was overcome by expressing the Cas9 protein from a vector that cannot replicate at optimal temperature of *Synechococcus*, and thereby facilitating Cas9 expression only during a limited period after transformation. Absence of NHEJ in most prokaryotes in the event of DSB caused by Cas9 is generally lethal, unless a HDR based repair template is provided. Therefore, an upstream and downstream homologous regions of the target gene were incorporated into the plasmid, allowing for homologous recombination-based removal of a target gene, and counter-selecting of wild type cells by Cas9 [165]. This approach resulted in successful knock out of *nblA* gene involved in degradation of photosystem-associated proteins with 100% efficiency [165, 166]. In strain PCC 7942, the activity of Cas9 in inducing DSB resulting in cell death was confirmed upon observing proportional reduction in transformants with increase in the dosage of plasmid harboring Cas9 and sgRNA targeting the host genome for

transformation. In line with the observation made in other organisms, the HDR was found to be enhanced in PCC 7942 by the induction of DSB by Cas9 [92, 94, 129-131, 164]. This Cas9-assisted HDR was used for metabolic engineering of PCC 7942 by knocking in genes coding for phosphoenolpyruvate carboxylase and citrate synthase to enhance the carbon flux towards oxidative pathway of Tricarboxylic acid (TCA) cycle. Combining these knock-ins with the knock-out of the glucose-1-phosphate adenylyl transferase gene to block the conversion of glucose to glycogen, to increase the carbon flux towards the glycolysis, resulted in an 11-fold increase in succinate titer [164].

The toxicity of Cas9 was the bottleneck that restricted wide application of this system in cyanobacteria for genome engineering [165]. Replacement of Cas9 with Cas12a considerably solved these problems [167]. Cas12a has been successfully employed for obtaining marker-less mutants of various cyanobacterial species including *Synechococcus elongatus* UTEX 2973, *Synechocystis* sp. PCC 6803 and *Anabaena* sp. PCC 7120 [167]. Homologous flanks in the upstream and downstream part of the target region (both 1kb) were introduced into the same vector expressing the Cas12a and a crRNA guide was used for HDR based generation of mutants followed by counter selection with Cas12a. Sanger sequencing and PCR of the target site followed by phenotypic characterization confirmed the mutants with knock-ins, knock-outs and point mutations. In *Synechococcus*, a point mutation in the codon of the *psbA1* gene (S246A) was achieved among 25% of the colonies screened initially; re-streaking the 8 colonies thrice in media with antibiotic to maintain the plasmid resulted in 75% of the colonies having the desired genotype [167]. A similar observation was made during knocking in the YFP gene, where initially 20% of the colonies were segregated mutants and re-streaking the colonies twice on appropriate antibiotic plates yielded 60% mutant colonies. Also, the deletion of *nblA* gene resulted in 90% segregated mutants after 3 generations of re-streaking [167]. In *Synechocystis*, the efficiency of obtaining *nblA* gene deletion was comparatively reduced (45%), probably due to the high ploidy level in this species (up to 50 chromosomes per cell), and to the presence of two adjacent copies of the target gene on each chromosome [168]. Introduction of a point mutation in the *isiA* gene and a gene insertion by replacement of *nblA* gene with YFP was achieved with very high efficiency in *Synechocystis* where 85% of the obtained colonies were segregated mutants [167]. In *Anabaena*, deletion of first 400 bps of the *nifH* gene yielded the designed mutants in 60% of the colonies. Likewise, point mutation of the *nifD* gene and replacement of *nifH* gene with YFP resulted in 60% segregated mutants [167].

Conclusions

Genome editing in a wide variety of species has been simplified with efficient and successful application of the CRISPR-Cas technology. However, the cytotoxic effects of the Cas9 nuclease has been a hurdle in exploiting the complete potential of this system in at least some of the microalgal species [109, 127, 132, 146, 151, 164, 165]. In *Chlamydomonas*, the Cas9 was inferred to be toxic in transformants based on unsuccessful detection of even weak expression of the (intact) Cas9/dCas9-encoding genes [109]. The molecular basis of this cytotoxic effect of Cas9 has been proposed to be off target cleavage of host

genome. Nevertheless, this does not explain observations on inhibitory effects of dCas9, a catalytically inactive Cas9 variant. In another study, however, the successful expression of dCas9 was achieved under the same promoter and terminator for downregulation of various genes [136]. Whereas the former study used a *Chlamydomonas* codon-optimized dCas9-encoding gene, the latter study used a 15% different *Zea mays* codon-optimized sequence [109, 136]. This observation indicates the importance of factors such as codon harmonization to improve functional protein production. Adding to the proposed Cas9 cytotoxicity, the diminished expression of Cas9 due to yet unknown factors might be responsible for the feeble efficiency of the tool as observed in *C. reinhardtii* and *N. oceanica*. Moreover, the presence of introns and their role in regulating gene expressions in eukaryotes could also be detrimental, e.g. the genome of *Chlamydomonas* contains about 8.5 introns per gene [169]. Similarly, genomes from micro-algae such as *Nannochloropsis*, *Phaeodactylum* and *Thalassiosira* contain 1.7, 0.8 and 1.5 introns per gene, respectively [71, 170, 171]. The introduction of introns into the coding sequence of some exogenous proteins has been reported to considerably enhance the expression in *Chlamydomonas* [172]; hence, the introduction of introns could contribute to improved functional Cas9 production. Also, fusing an antibiotic resistance gene to a heterologous gene via a self-cleaving 2A peptide sequence and subsequent selection on antibiotic plates has resulted in transformants with up to 100-fold increased levels of heterologous protein production [173]. This strategy could be employed for alleviating Cas9 expression and thereby increase the efficiency of generating mutants.

Table 1. Overview of the native genes edited in Photosynthetic microbes using CRISPR-Cas systems. Abbreviations: RNP: Ribonucleoprotein; HDR: Homology Directed Repair.

Species	Strategy for Cas protein delivery	Outcome of the target gene	Modification at the target site	Target genes	Ref.
<i>Chlamydomonas reinhardtii</i>	Plasmid	Knockout	indels	<i>FKB12, PSY1, Chr2, COP1/2, COP5, PHOT</i>	[109, 132]
			HDR	<i>ARG7, ALS</i>	[127]
	RNP	Knockout	indels	<i>MAA7, CpSRP43, ChlM, CpFTSY, ZEP, PHT7</i>	[137-139, 168]
			HDR	<i>aCRY, COP1/2, COP5, PHOT, UVR8, VGCC, MAT3, KU80, POLQ</i>	[132]
Plasmid	Downregulation	none	<i>PEPC1</i>	[136]	
<i>Nannochloropsis oceanica</i>	Plasmid	Knockout	indels	<i>NR; g7988</i>	[146]
<i>Nannochloropsis gaditana</i>	Plasmid	Disruption	NHEJ based insertion of antibiotic resistant gene cassette	18 different putative transcriptional regulators.	[147]
<i>Phaeodactylum tricornutum</i>	Plasmid	Knockout	indels	<i>CpSRP54</i>	[150]

Species	Strategy for Cas protein delivery	Outcome of the target gene	Modification at the target site	Target genes	Ref.
<i>Thalassiosira pseudonana</i>	Plasmid	Knockout	Deletion of a 37 bps fragment in the coding region of the gene by simultaneous targeting of 2 regions on the same gene	<i>URE</i>	[157]
		Disruption	HDR based targeted insertion of antibiotic resistant gene cassette	<i>Silacidin</i>	[157]
<i>Synechocystis</i> sp. PCC 6803	Plasmid	Downregulation	none	<i>phaE</i> , <i>glgC</i> and 4 putative aldehyde reductases/dehydrogenases	[160]
		Knockout	HDR	<i>nblA</i> and <i>isiA</i>	[167]
<i>Synechococcus elongatus</i> PCC 7942	Plasmid	Downregulation	none	<i>glgC</i> , <i>sdhA</i> and <i>sdhB</i>	[161]
		Knockout	HDR	<i>glgC</i>	[164]
		Knock in	HDR	<i>gltA</i> and <i>ppc</i>	
<i>Synechococcus</i> sp. PCC 7002	Plasmid	Downregulation	none	<i>cpcB</i> , <i>ccmK1</i> and <i>glnA</i>	[162]
<i>Anabaena</i> sp. PCC 7120	Plasmid	Downregulation	none	<i>glnA</i> and <i>devH</i>	[163]
		Deletion	HDR	<i>nifH</i> and <i>nifD</i>	[167]
<i>Synechococcus elongatus</i> UTEX 2973	Plasmid	Knockout	HDR	<i>nblA</i> and <i>psbA1</i>	[165, 167]

Because the efficiency of implementing plasmid-based Cas9 for genome editing in microalgae was minimal, a breakthrough has been the delivery of RNPs by electroporation [132, 137, 138]. However, in cyanobacteria the toxicity problem of Cas9 was solved substantially by replacing it with Cas12a which at least in some cases has resulted in successful genome editing [167]. The ability of Cas12a to process its own crRNA guides has been exploited for efficient multiplex gene editing in various species [122, 174]. These results indicate the possibilities of expanding Cas12a into the micro-algae for efficient and high-throughput genome engineering. Apart from the SpCas9 and Cas12a other variants of CRISPR systems which could possibly emerge as genome editing tools were reviewed recently [175, 176]. The improved performance of SaCas9 compared to SpCas9 in *C. reinhardtii* [132] also indicates the possibilities of applying different unexplored variants of Cas9 and Cas12a to assess their functionality and toxicity effects. The recent characterization of thermostable Cas9 variants may provide a platform for genome engineering of cyanobacterial species adapted to extreme conditions [177, 178]. It is important to realize that the potential of CRISPR systems is not confined to the widely used Cas9, and that extending engineering studies of model and non-model strains by using distinct Cas nuclease variants could further expand the toolbox for genome engineering of photosynthetic microorganisms and revolutionizing their development as industry-relevant cell factories.

Acknowledgement

This work is supported by the project titled "*Microalgae As a Green source from Nutritional Ingredients for Food/Feed and Ingredients for Cosmetics by cost effective New Technologies*" [MAGNIFICENT], funded by Bio-based Industries Joint Technology Initiative under European Commission [project ID: 745754]

CHAPTER 3

3

CRISPR–Cas ribonucleoprotein mediated homology-directed repair for efficient targeted genome editing in microalgae *Nannochloropsis oceanica* IMET1

Mihris Ibnu Saleem Naduthodi^{1 2}, Prarthana Mohanraju¹, Christian Südfeld², Sarah D'Adamo², Maria J Barbosa², John van der Oost*¹

¹Laboratory of Microbiology, Wageningen University and Research, Wageningen, Netherlands.

²Bioprocess Engineering, AlgaePARC, Wageningen University and Research, Wageningen, Netherlands.

*Corresponding Author

Chapter adapted from publication:
Biotechnol Biofuels 12, 66 (2019). <https://doi.org/10.1186/s13068-019-1401-3>

Abstract

Microalgae are considered as a sustainable feedstock for the production of biofuels and other value-added compounds. In particular, *Nannochloropsis* spp stand out from other microalgal species due to their capabilities to accumulate both triacylglycerol (TAG) and polyunsaturated fatty acids (PUFAs). However, the commercialization of microalgae-derived products is primarily hindered by the high production costs compared to less sustainable alternatives. Efficient genome editing techniques leading to effective metabolic engineering could result in strains with enhanced productivities of interesting metabolites and thereby reduce the production costs. Competent CRISPR-based genome editing techniques have been reported in several microalgal species, and only very recently in *Nannochloropsis* spp (2017). All the reported CRISPR-Cas-based systems in *Nannochloropsis* spp rely on plasmid-borne constitutive expression of Cas9 and a specific guide, combined with repair of double-stranded breaks (DSB) by non-homologous end joining (NHEJ) for the target gene knockout.

In this study, we report for the first time an alternative approach for CRISPR-Cas-mediated genome editing in *Nannochloropsis* sp; the Cas ribonucleoproteins (RNP) and an editing template were directly delivered into microalgal cells via electroporation, making Cas expression dispensable and homology directed repair (HDR) possible with high efficiency. Apart from widely used SpCas9, Cas12a variants from three different bacterium were used for this approach. We observed that FnCas12a from *Francisella novicida* generated HDR based targeted mutants with highest efficiency (up to 93% mutants among transformants) while AsCas12a from *Acidaminococcus* sp. resulted in the lowest efficiency. We initially show that the native homologous recombination (HR) system in *N. oceanica* IMET1 is not efficient for easy isolation of targeted mutants by HR. Cas9/sgRNA RNP delivery greatly enhanced HR at the target site, generating around 70% of positive mutant lines. We show that the delivery of Cas RNP by electroporation can be an alternative approach to the presently reported plasmid-based Cas9 method for generating mutants of *N. oceanica*. The co-delivery of Cas-RNPs along with a dsDNA repair template efficiently enhanced HR at the target site, resulting in a remarkable higher percentage of positive mutant lines.

Keywords

Microalgae, *Nannochloropsis*, CRISPR, Cas9, Cas12a, Ribonucleoproteins, genome editing, homologous recombination, homology directed repair.

Background

Microalgae have the ability to combine photosynthesis with high lipid and biomass productivities. Moreover, understanding the potential of these solar powered cell factories in food, feed and bioenergy production has elevated their interest in biotechnological research. Among other microalgal species, *Nannochloropsis* spp. has attracted considerable attention owing to its oleaginous nature. The natural ability to accumulate 60% and 12% of the cell dry weight with lipids or triacylglycerol (TAG) and omega-3 poly unsaturated fatty acid (PUFA), respectively, under defined conditions has brought these genera to the spotlight of scientific exploration [69] [62]. The consumption of omega-3 fatty acids is beneficiary for human health as they prevent heart and aging related diseases [68], while TAG can be used as a precursor for biodiesel production [179]. Currently, the main source of omega-3 PUFA for human consumption is seafood [180]. However, the contamination of seafood due to severe ocean pollution is a major concern along with the reduction in global fish population. Furthermore, the exploitation of fossil fuels and the environmental impacts of a petroleum-based society are crucial issues to be addressed. In this context, developing *Nannochloropsis* spp as efficient production systems can contribute to establishing a sustainable bio-economy.

In spite of the apparent advantages, the major bottleneck in employing *Nannochloropsis* spp as an economically viable production platform for TAG and PUFA is the high cost of production [39]. One of the ways to overcome this barrier is to deploy metabolic engineering for improving the oil productivity of microalgal strains. To this end, efficient genome editing tools leading to competent metabolic engineering strategies need to be developed. Genome editing by Clustered Regularly Interspaced Short Palindromic Repeats and the CRISPR associated proteins (CRISPR-Cas) has been successfully implemented in cells of a wide range of organisms, including microalgae [79, 181-185]. The specific DNA double stranded breaks (DSBs) generated by CRISPR RNA (crRNA)-guided Cas nucleases is repaired by either of the two pathways: non-homologous end joining (NHEJ) or homology directed repair (HDR) [186-188]. The NHEJ is generally associated with the introduction of insertions and/or deletions (indels) of varying lengths at the DSB site, often leading to the disruption of the reading frame of the target gene. The HDR pathway results in a precise insertion or deletion at the DSB site by homologous recombination [111, 112, 189].

Since 2017, a plasmid-driven Cas9-based approach has been implemented in *Nannochloropsis* spp for marker-less genome editing [79, 182]. The Cas9 system has also been effectively implemented for metabolic engineering of *Nannochloropsis gaditana* to enhance the lipid production [147]. Verruto *et al* combined Cas9 with Cre recombinase to recycle the limited antibiotic resistance marker along with gene stacking in *N. gaditana* [182]. Recently, non-transgenic and marker-free gene disruption was achieved in *Nannochloropsis oceanica* using an episomal CRISPR-Cas system [79].

So far, all studies on genome editing in *Nannochloropsis* spp have relied on the plasmid based expression of Cas9 combined with NHEJ-mediated target repair [79, 146, 147, 182]. One-step delivery using Cas-ribonucleoprotein (Cas-RNP) complex harbouring pre-assembled Cas effector protein and targeting single guide RNA (sgRNA) has been

applied for genome editing in microalgae *Chlamydomonas reinhardtii* and *Phaeodactylum tricorutum* [132, 137, 138, 140, 183, 185]. The advantages of using the Cas-RNP delivery approach include (i) the dispensability of a codon harmonized Cas gene, as well as promoters and terminators that are active in host of interest, (ii) circumvention of the addition of ribonucleases/ribozymes required to obtain precise processing of sgRNA transcripts to facilitate the loading of Cas9 *in vivo*, and (iii) reduced off-target and cytotoxic effects as Cas protein is saturated with sgRNA and is transiently active before it is degraded by endogenous proteases in cells [190-192].

Homologous recombination (HR) based genome editing has been reported in *Nannochloropsis spp* in 2011 [78]. However, further studies did not report any application of this technique for genome editing in *Nannochloropsis spp*. We here report the development of a Cas-RNP-mediated HDR approach for genome editing in *N. oceanica* IMET1. We enhanced the native homologous recombination system of *N. oceanica* IMET1 with targeted DSB induction by Cas9 protein to facilitate HDR and obtain knockout mutants of the nitrate reductase (*NR*) gene. CRISPR-Cas12a (previously known as Cpf1), a relatively new Cas enzyme is a dual nuclease involved in crRNA processing, target-site recognition and DSB induction [102]. As Cas9 and Cas12a proteins recognize different PAM motifs, they can be used to target distinct sites. Thus, establishing the *in vivo* activity of Cas12a in *Nannochloropsis spp* expands the options of available target sites for the introduction of DSB and obtain desired mutants. Also, this could pave way for establishing plasmid based Cas12a systems in *Nannochloropsis spp* for multiplexed genome editing. We also compared the efficiencies of Cas12a protein variants from *Lachnospiraceae bacterium* ND2006 (LbCas12a), *Acidaminococcus sp.* BV3L6 (AsCas12a) and *Francisella novicida* (FnCas12a) with Cas9 from *Streptococcus pyogenes* in generating HDR based NR mutants in order to elucidate the optimal enzyme variant for generating targeted mutants using the Cas RNP based approach in *Nannochloropsis spp*.

Results

Homologous recombination and antibiotic-based selection for the generation of nitrate reductase (*NR*) knockouts

The nitrate reductase (*NR*) gene has been used as the target for genome editing studies in *N. oceanica* as the mutants can be easily characterized through replica plating, based on their ability to grow on ammonia containing media and their inability to grow on media containing nitrate as the sole nitrogen source [78, 79, 146]. To generate a *NR* gene knockout mutant in *N. oceanica* IMET1, we transformed cells with a vector that harbours a zeocin antibiotic resistance cassette with ~1 kb flanks homologous to the upstream and downstream ends of the *NR* gene locus for deletion. The 1 kb HR flanks were designed to replace 430 bp of 4.1 kb long *NR* gene with a zeocin resistance cassette. After electroporation, *N. oceanica* IMET1 cells that were successfully transformed with the PCR-amplified HR editing template were selected on solid nutrient media with ammonia and 2 µg/mL zeocin. The zeocin resistant colonies were screened for targeted *NR* gene knockout via PCR. For this, we designed primers (Check FW and Check RV) binding outside the upstream and downstream of the HR flanks on the host genome (Fig. 1b). The expected

PCR amplicon size of the mutant colonies generated by HDR and wild-type colonies was 4337 bp and 2591 bp, respectively.

We screened a total of 63 colonies from three independent transformation experiments with colony PCR (cPCR). For 90% colonies we observed amplicon bands corresponding to the size of the expected wild type while 10% colonies showed both wild type and mutant bands. The 10% colonies containing both the PCR amplicons are most likely the result of mutant and wild type cells at close proximity upon plating. This occurrence of 10% colonies indicates the presence of HR in *N. oceanica* IMET1. However, this Cas-independent approach is not efficient enough to facilitate routine generation of HR mutants without further optimisation.

Cas9 RNP mediated HDR for NR mutant generation in *N. oceanica* IMET1

The DSB induction by endonucleases has been reported to enhance the efficiency of gene replacements by HDR in various eukaryotic cell lines [111, 127, 193-196]. The CRISPR-Cas mediated HDR has been reported in the microalgal species *Chlamydomonas reinhardtii* and the diatom *Thalassiosira pseudonana* [127, 132, 157]. However, this strategy has not yet been applied in *Nannochloropsis* spp. We hypothesized that induction of DSB using Cas9 would increase the HR efficiency in *N. oceanica*. Previous studies have shown that a plasmid-borne SpCas9/sgRNA complex can accurately generate DSBs *in vivo* in *N. oceanica* resulting in NHEJ mediated target gene disruption [79]. Hence, this protein was selected for exploring an RNP-based transformation approach in this species. We randomly designed 2 sgRNAs (NR-1 and NR-2), targeting the *NR* gene and transcribed them *in vitro*. After assessing the *in vitro* nuclease activity of preassembled Cas9 RNP, it was delivered along with HR editing template into *N. oceanica* IMET1 cells by electroporation (Fig 1a). The frequency of NR mutants arising from Cas9 mediated HDR was evaluated by quantifying the percentage of expected mutants among the zeocin resistant colonies obtained after transformation.

From three independent transformation experiments using Cas9 NR-1 and Cas9 NR-2 RNPs, a total of 32 and 35 colonies were screened respectively by colony PCR. The cells transfected with the Cas9/sgRNA (NR-1), we observed 12% colonies with a wild type *NR* amplicon, 15% colonies indicating the presence of both cell types, and 71% *NR* mutant colonies (Fig. 2a). As for the cells transfected with the Cas9 NR-2, 57% colonies were found to contain the wild type *NR* gene, while colonies with both amplicons (wild type and *NR* mutant) and *NR* mutant colonies reduced to 8% and 34%, respectively (Fig. 2a). HDR-based gene insertion was confirmed by sequencing 4 mutant PCR amplicons, revealing the anticipated targeted insertion of the editing template at the *NR* locus. Furthermore, the phenotypic difference between the *NR* mutants and the wild-type cells were demonstrated by the bleaching of *NR* mutant cells when streaked on F/2 agar plates with NaNO_3 as the sole nitrogen source, but not on plates with NH_4Cl (Fig. 2c).

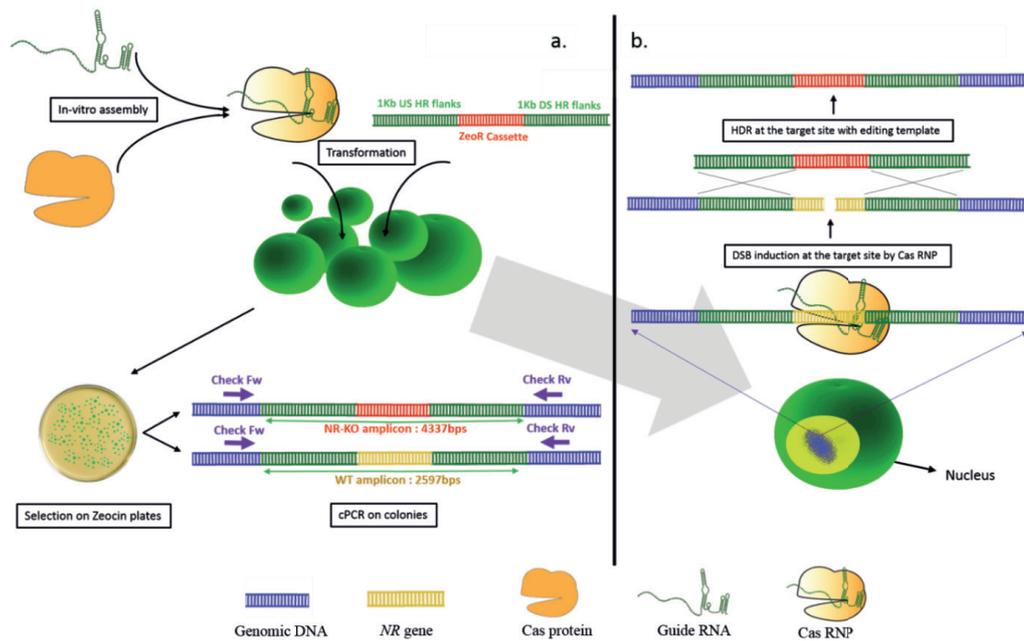


Figure 1. Approach; (a) The purified Cas protein is assembled with guide RNA to form the active RNP complex which is transformed along with the editing template into the competent *Nannochloropsis oceanica* IMET1 cells. The transformed cells are selected on ASW-NB plated with zeocin and ammonia and the mutants are screened by cPCR. (b) The Cas RNP introduces DSB at the target site which is repaired by HDR in presence of the editing template resulting in targeted mutants.

Cas12a RNP-mediated HDR for *NR* mutant generation in *N. oceanica* IMET1

To expand the genetic toolbox of *N. oceanica* IMET1 and to increase the availability of targeting sites for DSB induction, we then set out to test the efficiency of HDR using different Cas12a protein variants [*Lachnospiraceae* bacterium ND2006 (LbCas12a), *Acidaminococcus* sp. BV3L6 (AsCas12a) and *Francisella novicida* (FnCas12a)] for generating *NR* knockout mutants in *N. oceanica*. To this end, we designed three random crRNAs (*NR*-1, -2 and -3) that target the *NR* gene and were obtained as RNA molecules from Sigma Aldrich. Cas12a is guided to the target site by shorter crRNA molecules (43nt) compared to the sgRNAs (120nt) required for Cas9 proteins [102]. Each Cas12a protein was combined with one of the three guides, leading to a combination of 9 different Cas12a RNPs: FnCas12a *NR*-1, 2, 3; LbCas12a *NR*-1, 2, 3; AsCas12a *NR*-1, 2, 3. The activity of all the different Cas12a RNPs were validated *in vitro* (Figure S1) and were transfected along with the HR editing template into *N. oceanica* cells by electroporation. After four weeks, colonies were screened by cPCR.

In case of the *N. oceanica* cells transfected with FnCas12a RNPs and the HR editing template, 30 colonies per Cas12a RNP guide combination were screened. Of the 30 colonies screened, 47% of the transformants were *NR* mutants. The cells transfected with FnCas12a-*NR*2 and HR editing template resulted in 93% *NR* mutants, while the rest of the

transformants indicated both wild type and mutant cells in them. In the case of the cells transfected with FnCas12a-3, only 20% were *NR* mutants (Fig. 2a).

For the cells transfected with LbCas12a NR1, 28 colonies were screened among which 78% were *NR* mutants while 14% were wild type. Among 28 colonies screened for the cells transfected with LbCas12a NR-2, we observed 71% mutant colonies. Upon screening 23 colonies from the cells transfected with LbCas12a NR-3, we were unable to obtain any *NR* mutant colonies (0%). However, 65% of them indicated the presence of both wild type and mutant cells and the rest of the transformants indicated wild type colonies (Fig. 2a).

We screened 27 colonies for the cells transfected with AsCas12a NR-1 and observed that 70% of them were wild type and only 3% of them were a *NR* mutant. In the case of the cells transfected with AsCas12a NR-2, out of the 30 colonies screened no mutant colonies were observed (0%). Upon screening 25 colonies from the cells transfected with AsCas12a NR-3, only 4% *NR* mutants were obtained (Fig. 2a).

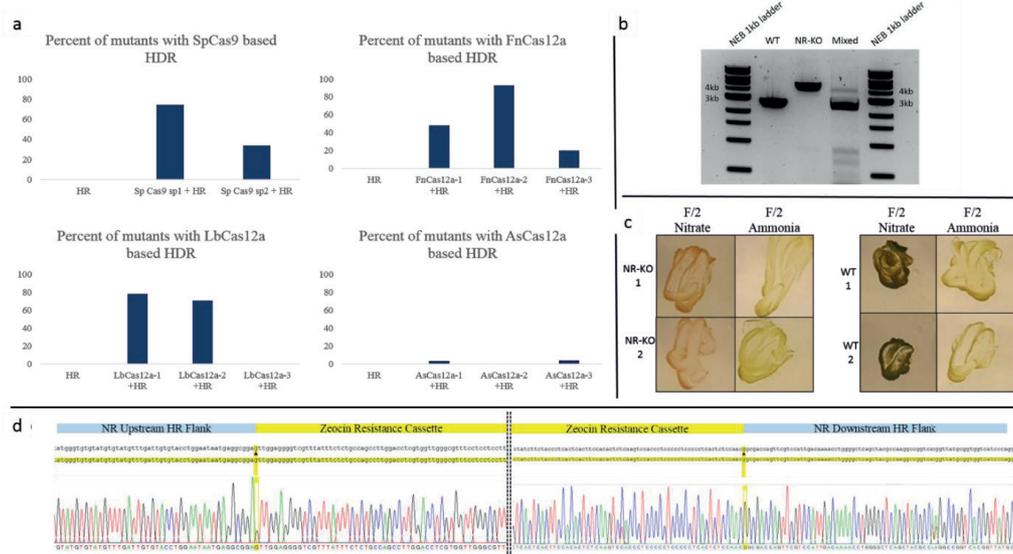


Figure 2. (a) Percent of mutants obtained with various RNP based HDR. (b) Gel electrophoresis image of the WT, NR-KO and mixed colonies upon colony PCR. (c) Phenotypic characterization of NR mutants, the NR mutants were observed to bleach in media with nitrate as the sole nitrogen source while they grew similar to WT in media with ammonia as the nitrogen source (The re-streak plates for characterization included in the supplementary data S13). (d) Sequencing result of a NR-KO mutant showing precise integration of the Zeocin resistance cassette in to the host genome knocking out the NR gene.

Discussion

Homologous recombination and antibiotic resistance-based selection in the absence of artificially induced DSBs has been demonstrated to be highly efficient in generating genetic mutants of *Nannochloropsis oceanica* [78]. Our results partly agree with this observation where efficient HR is showcased in *N. oceanica* IMET1. The presence of 10% mixed colonies in our experiment indicates that the HR system is active in this species but we were not able to isolate any clean *NR* mutants by this approach. Although it might be possible to obtain isolated mutants using an HR-based approach, laborious screening of large number of colonies would be required. HR-based gene targeting has been reported in microalgal strains such as *Chlamydomonas reinhardtii* and *Schizochytrium* albeit it with low efficiencies (1/40 – 1/2000) of plasmid integration into the host genome [197-200].

As an alternative to HR for disrupting the target genes, the availability of endonucleases with tuneable sequence specificity for the introduction of DSB at specific sites and repair through NHEJ or HDR has been an important milestone in the field of genome editing [201]. The first endonucleases that were used for genome editing were Zinc Finger Nucleases (ZFNs) and Transcription Activator Like Effector Nucleases (TALENs), but since 2013 has been dominated by the CRISPR-based nucleases for its cheap, precise and simpler approach [181, 202]

CRISPR-Cas9 based genome editing in microalgae was primarily reported in *C. reinhardtii* with a very low efficiency in generating mutants through NHEJ mediated disruption of the target gene [109]. The initial reports on *Nannochloropsis spp* also reported similar outcome, where approximately 300 colonies were screened to obtain 2 expected mutants [146]. However, an efficient system based on Cas9 targeting and NHEJ mediated insertion of antibiotic cassette was reported in *N. gaditana* for successful disruption of up to 18 different putative transcriptional regulators [147]. In this approach, the Cas9 protein was expressed from a genome integrated plasmid that harboured a codon harmonized *Cas9* gene. The efficiency of obtaining positive knockouts in this approach varied between 6-78% depending on the target genes [147]. The same research team combined this Cas9 expression plasmid along with a Cre recombinase expression cassette under the control of an inducible regulator to generate markerless knockouts and recycle the antibiotic resistance for further gene knockout in the same strain [182]. Adapting the CEN/ARS6 region from *Saccharomyces cerevisiae* to develop a circular self-replicating vector active in *N. oceanica* facilitated the development of marker-free non-transgenic mutants [79]. After obtaining the expected mutants through targeted gene disruption by DSB and NHEJ repair, the mutants were grown on media without antibiotic to clear the circular plasmid from the host, thereby resulting in a non-transgenic mutant [79].

All reports to date on Cas9-based genome editing strategies in *Nannochloropsis spp* have exploited the plasmid-based expression of Cas9 endonuclease combined with NHEJ repair systems for disruption of target genes. Our efforts to attain mutants through HR and antibiotic selection was not completely successful, but at least we obtained evidence for a functional HR system in *N. oceanica* IMET1. In the light of recently reported efficient Cas9 activity in *Nannochloropsis spp* [79, 147, 182] and earlier reports on enhanced HR

efficiency upon DSB induction [127, 132, 196, 203, 204], we applied the Cas9 based DSB induction to enhance the HR in *N. oceanica* IMET1 and thereby obtain targeted *NR* mutants.

As an alternative to the present approaches in *Nannochloropsis spp*, we used the Cas9 RNP to induce the DSB. Using a Cas RNP has some noteworthy advantages over the plasmid-based expression of Cas proteins in the host species. The off-target problem of Cas9 and strategies to overcome this issue is widely studied [205]. The transient presence of Cas RNPs in the host upon transformation reduces the risk of off-target effects compared to long-term, constitutive expression of Cas proteins from a plasmid. Moreover, the saturation of Cas protein prior to transformation with appropriate guides, reduces the risk of Cas protein binding to unwanted sgRNA-like molecules leading to unspecific cleavage events. The possibility to assay the activity of RNP complex *in vitro* prior to tedious *in vivo* experiments reduces the risk of experiment failure adding to the advantages of this approach. Furthermore, the possibility to avoid the requirement of gene regulators for Cas expression and RNA processing systems for precise sgRNA production *in vivo* further increases the simplicity of this approach.

Table 1. Key differences between the Cas9 and Cas12a CRISPR proteins.

Feature	Cas9	Cas12a (Cpf1)
<i>CRISPR enzyme</i>	<ul style="list-style-type: none"> Class 2, Type II-B Molecular weight: ~164 kDa Endonuclease domains: RuvC and HNH 	<ul style="list-style-type: none"> Class 2, Type V-A Molecular weight: ~158 kDa Endonuclease domains: RuvC only
<i>PAM</i>	3' NGG	5' TTTV
<i>guideRNA</i>	~120 nt, two RNA sequences crRNA and tracrRNA are synthetically fused to form the single guide RNA (sgRNA)	~43 nt, only crRNA is required
<i>crRNA maturation</i>	RNase III dependent	RNase III independent, autonomous processing
<i>DSB cleavage site</i>	<ul style="list-style-type: none"> 3 bases upstream of the protospacer sequence PAM site often destroyed during genome editing 	<ul style="list-style-type: none"> 18-23 bases downstream of the PAM PAM site may be preserved after genome editing
<i>DSB cleavage mechanism</i>	Blunt end cuts	5-nt staggered end cut distal to the PAM

The co-transformation of SpCas9 RNP targeting the *NR* gene along with the HR editing template rapidly generated mutants by HDR. The observation of up to 74% mutants among the screened antibiotic resistant colonies indicates the excessive increase in HR upon using the SpCas9 RNP for DSB induction. Based on the positive results obtained with SpCas9, we decided to try to expand the genome editing toolbox of *N. oceanica* even further by assessing the efficiency of Cas12a variants in generating similar mutants. We have summarized the key differences of Cas9 and Cas12a in Table 1. As the Cas12a endonuclease introduce DSB 18-23 bp away from the PAM, the NHEJ repair might not disturb the PAM and seed sequence, thereby allowing Cas12a to cleave the same site again until HDR is successful. According to this hypothesis, Cas12a variants might exhibit even higher efficiencies than Cas9 in generating mutants through HDR.

Among the three different Cas12a variants we tested (FnCas12a, LbCas12a and AsCas12a), a considerable variation in efficiencies were observed in generating mutants. Apart from the variations based on Cas12a protein, the randomly selected guide sequences used for targeting also made a difference. FnCas12a and LbCas12a were considerably active in generating mutants along with the guide sequences 1 and 2. FnCas12a marked the highest efficiency with 93% mutants among the 30 antibiotic resistant colonies screened. Irrespective of the guide sequence used for targeting, we did not observe efficient mutant generation when AsCas12a was used. Similarly, the guide sequence 3 exhibited inadequate activity in producing mutant colonies regardless of the Cas12a variant used for targeting. The diminished activity of AsCas12a in comparison to LbCas12a and FnCas12a in planta has been reported in various studies [115, 121, 140, 206-210]. It has been reported that the *in vivo* activity of Cas12a proteins were enhanced at higher temperatures (34°C) and this effect was remarkable in AsCas12a [125]. This explains the diminished activity of AsCas12a in microalgae and plants grown at lower temperatures (<28°C) [125]. The sgRNA dependent variation in Cas9 cleavage has been reported in previous studies [211-214]. Correspondingly, we observe major variations in efficiency of mutant generation by the same Cas12a or Cas9 with different guide sequences. In mammalian cells, high throughput screening of sgRNAs was performed to develop a predictive model for effective sgRNA design [215]. Similar studies should be performed in microalgae to allow predicting effective sgRNAs for different target genes *in silico* to achieve even more efficient CRISPR based genome editing.

The major bottleneck of our approach is the presence of an antibiotic resistance marker that allows for efficient selection of the desired mutants. This marker could be removed by applying the Cre-Lox approach as has been demonstrated in to work in *N. gaditana* [182]. However, this system still leaves a minor scar (34nt) at the target site. We have shown that HDR is very efficient in *Nannochloropsis spp.* Markerless mutant strains could thus be obtained by 2 subsequent rounds of transformation with alternating positive and negative selection. The RNP-based delivery approach could also be used for multiplexed genome editing as shown in *P. tricornutum* [185]. Co-transformation of multiple RNPs targeting different genomic loci along with multiple editing templates could facilitate HDR-based multiplexing. However, without a positive selection system based on toxicity as was demonstrated in *P. tricornutum* [185], we might require multiple antibiotic resistance markers to select mutants. Moreover, using RNPs for genome editing has also been considered to attain a non-GMO label in absence of transgenic sequences [216-218]. In this context, our approach could be developed towards achieving non-GMO knockout mutants with the help of high throughput selection tools such as FACS.

Conclusion

Understanding the potential of developing industrially relevant *N. oceanica* has accelerated the studies on genome editing in this species. Alternative to the presently available strategies, we demonstrate the possibility of implementing Cas RNPs to drive targeted genome editing in combination with HDR in *N. oceanica*. The efficient HDR could be effectively used for precise knockout and in-frame knock-in of genes. A previous

study reported the efficient generation of mutants by homologous recombination and antibiotic based selection in this species [78]. Even though we were unable to confirm the results of this study, we witnessed the presence of an efficient HR system in the species. Co-transformation of Cas9 RNPs targeting the NR gene along with an HR editing template substantially enhanced the portion of mutants among antibiotic resistant colonies. This indicates the reinforcing effect of DSB for inducing HR at the target sites through HDR. In addition to establishing a Cas9 RNP-based transformation protocol in *N. oceanica*, we also demonstrate the activity of 3 different Cas12a variants. We observed that FnCas12a performed the best in generating mutants while AsCas12a was the weakest. LbCas12a exhibited an efficiency similar to Cas9 as observed in previous studies in *C. reinhardtii* [140]. The AsCas12a has reported to work remarkably well in mammalian cells while its activity is substantially lower in ectothermic species as their *in vivo* activity is diminished under lower temperature conditions (<28°C) [125]. Even though the mutants that we generated contain an antibiotic resistance marker, future studies will investigate selection systems for attaining marker-free, non-GMO, multiplexed mutants of *Nannochloropsis* spp. using high-throughput technology and insights generated within this research.

Methods

Strain and growth conditions

N. oceanica IMET1 was kindly provided by prof. Jian Xu (Qingdao Institute for Bioenergy and Bioprocess Technology, Chinese Academy of Sciences). Microalgal cultures were grown in artificial sea water (ASW) media composed of 24.50 g/L Sodium chloride, 3.20 g/L Sodium sulphate, 0.80 g/L Calcium chloride di-hydrate, 0.85 g/L Potassium sulphate and 9.80 g/L Magnesium chloride hexahydrate. The ASW media was further enriched with commercial “nutribloom plus” (ASW-NB) obtained from Necton (Olhão, Portugal) or F/2 nutrients (ASW-F/2). When culturing with CO₂ supplementation (5% CO₂), the media was provided with 3 g/L Sodium bicarbonate and while culturing without CO₂ 4.77 g/L HEPES was added to ASW and pH was set to 8.0 prior to autoclaving and nutrient supplementation. Ammonium chloride with a final concentration of 12 mM was used in the media for NR mutants.

Plasmid construction

The homology repair template was designed to drive the removal of the nitrate reductase activity and in turn confer antibiotic resistance against Zeocin upon engineered cells. It carried the bleomycin resistance gene from *Streptoalloteichus hindustanus* (*shble*, GenBank accession number A31898.1) under the control of an endogenous promoter and terminator. The Violaxanthin Chlorophyll a Binding Protein Precursor (VCP) promoter region including the translation initiation ATG codon and the first intron of the VCP gene was amplified from genomic DNA using Q5 DNA polymerase (NEB) with the primers VCP FW and VCP RV. This promoter exhibited strong constitutive expression in transcriptomic studies [143] and a similar construct has been reported to drive *shble* expression in *N. oceanica* [219]. The *shble* gene was amplified from the plasmid pPtPuc3 (addgene #62863) which was a gift from Hamilton Smith. The gene was amplified with primer set BLE FW and BLE RV. The 3'UTR and transcriptional terminator of the alpha tubulin gene were previously

used in *N. oceanica* IMET1 [146]. We amplified them with the primer set A-TUB FW and A-TUB RV. We chose to add a second transcriptional terminator in reverse orientation, which was amplified from the 3'-region of the *Clp* protease proteolytic subunit gene with primers CLP FW and CLP RV. The fragments were assembled using the Gibson assembly technique (NEBuilder® HiFi DNA Assembly Master Mix) with 25 nucleotides overlap. The construct was introduced into the MCS of the cloning vector pUC19 (GenBank accession number M77789.2), which was linearized via PCR with primer set pUC FW and pUC RV. The correct assembly of these fragments in the construct pNIM14 was verified via PCR and sequencing.

In a second cloning step, ~1 kb homologous flanks for the knockout of the *NR* gene were added to both sides of the antibiotic resistance cassette. The upstream homologous flank was amplified from host genomic DNA with the primer set NR US Fw and NR US Rv and the downstream flank with primers NR DS Fw and NR DS Rv. The antibiotic resistance cassette was amplified by PCR with primer sets ZeoR Fw and ZeoR Rv harbouring overlaps to the US and DS HR flanks respectively. The pUC19 plasmid was linearized with primers pUC19 Fw and pUC19 Rv harbouring the overlaps to align with the HR flanks and antibiotic resistance cassette by Gibson assembly technique. After assembling the PCR products by NEBuilder® HiFi DNA Assembly Master Mix, the circular product was transformed into Dh5 alpha *E. coli* and the alignment was validated from single colonies by PCR and sequencing. The Genbank file of the vector developed is included in the Supplementary data. Once the construct was sequence-verified, the linear DNA for transformation of microalgal cells was PCR amplified from the construct with primers NR US Fw and NR DS Rv.

Transformation of *N. oceanica*

For transformation we followed the protocol developed by Vieler *et al.* The linearized vector, carrier DNA and RNPs were added into this 0.2 mL of 5×10^8 cells ml^{-1} concentrated culture and electroporation was performed using a 2mm cuvette and following pulse settings: 2000 V, 600 Ohm resistance and 50 μF capacitance. After electroporation, the cells were transferred to 5 mL ASW in 15 mL Falcon tubes supplemented with nutribloom and ammonium chloride (12 mM) and left under continuous light of $50 \mu\text{mol m}^{-2}\text{s}^{-1}$ without mixing for recovery (2 days).

The linear PCR product for transformation into *N. oceanica* IMET1 was amplified from the circular plasmids obtained from *Escherichia coli*. Primers NR US FW and NR DS RV were used for amplifying the linear vector. 3 μg of linear vector DNA and 30 μg of Salmon Sperm DNA were used for transformation. For transformation with RNP complex, we pre-assembled purified Cas protein with appropriate guide sequences in an equimolar ratio (10 μM) and incubated at 37 °C for 10 minutes to form the Cas RNPs. 30 μL of 10 μM RNPs were transformed along with 3 μg editing template and 30 μg carrier DNA (Salmon sperm DNA).

The transformants were plated on ASW-NB (1% Agar), 1.2 mM Ammonium chloride and 2 $\mu\text{g}/\text{mL}$ zeocin. The plates were incubated under continuous light of $50 \mu\text{mol m}^{-2}\text{s}^{-1}$ for 3-4 weeks until colonies appeared.

sgRNA production

The sgRNA for targeting the *NR* gene with SpCas9 was produced by *in vitro* transcription of the corresponding DNA template with HiScribe T7 RNA synthesis from NEB. Two targeting spacers (sp1: 5'-GCCGGCGCAGACAAGAGTGA-3' & sp2: 5'-AACCTCCTTGCGGCGATTGC-3') were selected from the *NR* gene based on the PAM sequence 5'-NGG-3'. Following forwards primers were designed for adding the spacer sequence onto the sgRNA loop by PCR: Sp1 sgRNA FW and Sp2 sgRNA FW with common reverse primer sgRNA RV. After obtaining the PCR fragment with spacer sequence and sgRNA loop, a subsequent PCR was performed to attach the T7 promoter sequence 5'-TAATACGACTCACTATAGG-3' to the 5' end of the fragment. Following forward primers were used for adding the T7 promoter; T7 sp1 Fw and T7 sp2 Fw with the reverse primer sgRNA RV. The same reverse primer used in the previous PCR was used for above PCR reactions to obtain the final DNA fragment for *in vitro* transcription. The final DNA template sequence used for *in vitro* transcription is provided in the Supplementary table.

The *in vitro* transcription was performed according to the protocol provided by the supplier. The sgRNA was gel purified according to the protocol developed by Anders and Jinek (2014) [220] and the activity was confirmed through the following *in vitro* assay. The sequences of all guide RNAs and crRNAs are included in the supplementary table.

Cas proteins expression and purification

E. coli codon optimized *spcas9*, *E. coli* codon harmonized *fnCas12a*, and human codon optimized *ascas12a* and *lbcas12a* genes were cloned into a bacterial expression vector [6-His-TEV-Cas12a, a pET-based vector that was a gift from Scott Gradia (Addgene plasmid # 29653)]. One litre of LB growth media with 100 µg/mL ampicillin was inoculated with 10 mL overnight culture Rosetta (DE3) (EMD Millipore) cells containing the Cas12a expression construct. Growth media plus inoculant was grown at 37 °C until the cell density reached 0.5 OD₆₀₀, then the temperature was decreased to 20 °C. Growth was continued until OD₆₀₀ reached 0.6 when a final concentration of 500 µM IPTG was added to induce Cas12a expression. The culture was induced for 14-18 hours before harvesting cells and freezing them at -20 °C until purification.

Cell paste was suspended in 20 mL of Lysis Buffer (50 mM NaH₂PO₄ pH 8, 500mM NaCl, 1 mM 2-Mercaptoethanol, 10 mM imidazole) supplemented with protease inhibitors (Roche complete, EDTA-free) and lysozyme. Once homogenized, cells were lysed by sonication (Bandelin Sonoplus) and then centrifuged at 16,000xg for 1 hour at 4 °C to clear the lysate. The lysate was filtered through 0.22 micron filters (Mdi membrane technologies) and applied to a nickel column (Histrap HP, GE lifesciences), washed and then eluted with 250 mM imidazole. Fractions containing protein of the expected size were pooled and dialyzed overnight into the dialysis buffer (250 mM KCl, 20mM HEPES/KOH, 1 mM DTT). After dialysis, sample was diluted 1:1 in 10 mM HEPES/KOH pH 8.0, and loaded on a heparin FF column pre-equilibrated in IEX-A buffer (150 mM KCl, 20 mM HEPES/KOH pH 8). Column was washed with IEX-A and then eluted with a gradient of IEX-C (2 M KCl, 20 mM HEPES/KOH pH 8.0). The sample was concentrated to 700 µL prior to loading on a gel filtration column (HiLoad 16/600 Superdex 200) via FPLC (AKTA Pure). Fractions from gel filtration were analyzed by SDS-PAGE; fractions containing the Cas12a1 were pooled and

concentrated to 200 μ L (50 mM Tris-HCl pH 8.0, 2 mM DTT, 5% glycerol, 500 mM NaCl) and either used directly for biochemical and transfection assays, or frozen at -80°C for storage.

***In vitro* assays of RNP complex**

To obtain the RNP complex, equimolar amount (1 μ M) of Cas protein and sgRNA were incubated in a sterile 1.5 mL eppendorf tube at 37°C for 15 minutes along with 3 μ L 10X Cas buffer (NEBuffer 3.1) and milliQ water to a final volume of 30 μ L. The 10X Cas buffer is composed of 1 M NaCl, 500 mM Tris-HCl, 100 mM MgCl_2 , and 1 mg/ mL BSA at pH 7.9. After 15 minutes incubation, 100 ng of the template DNA was added into the reaction mixture and incubated again for 1 hour. The target DNA was amplified from the host genomic DNA with Q5 (NEB) PCR by using primers NR Fw: 5'-GTGGTGCCTAGTCGGAATGG-3' and NR Rv: 5'-GTCGGCCAATCCAGTTCGTGTC-3'. The template DNA was 1207 bp long and the digestion with Cas9 could result in fragments of size 183 and 1022 bp with sp1 and 200 and 1005 bp with sp2. Similarly, the cleavage with Cas12a spacer 1 generates fragments of size 314 and 898 bp, spacer 2 can produce fragments of size 443 and 774 bp while spacer 3 produces fragments of size 458 and 750 bp.

The activity of the RNP complex for cleaving the target DNA was assayed at 37°C and 25°C to assess the activity of RNP at the optimum temperature of Cas protein and temperature for *in vivo* experiment respectively. A description of the *in vitro* cleavage assay can be found in the supplementary data.

Colony PCR on transformants

Colony PCR was performed with the Phire Plant Direct PCR Master Mix (Thermo). Single colonies were taken with a 10 μ L inoculation loops and re-streaked to a fresh plate and incubated at same conditions. Also, another portion of the same single colony was resuspended in 20 μ L dilution buffer (obtained with the Phire Plant Direct PCR Master Mix kit) in an 8-strip PCR tubes. After incubating the sample for 15 minutes at room temperature, the sample was centrifuged for 30 seconds. 2 μ L of the supernatant from the dilution buffer-colony mixture was used as template for the Phire Plant Direct PCR. Following primer set was used for screening the antibiotic resistant colonies for mutants: Check FW: GTGCTTGATGCGGACGACAG, Check RV: AAAGCGCACGACGCAATGG .

Declarations

Authors' contributions

M.N, M.B and J.v.d.O designed the experiment. M.N, P.M and C.S executed the experiments. S.D, M.B and J.v.d.O supervised the project. M.N wrote the manuscript with input from all authors.

Competing interests

Author's declare no financial or commercial conflict of interests.

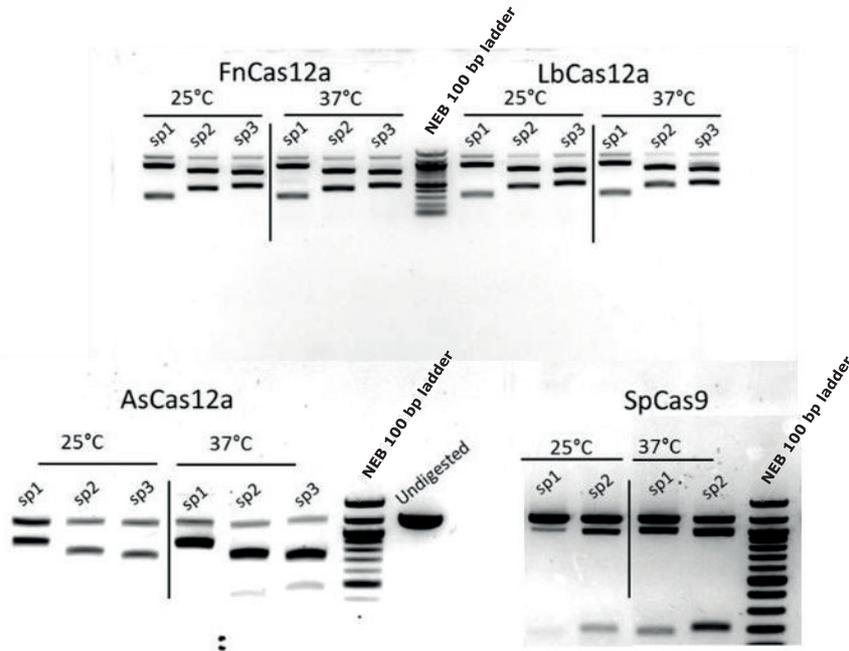
Availability of supporting data

All data generated or analysed during this study are included in this published article (and its Additional files).

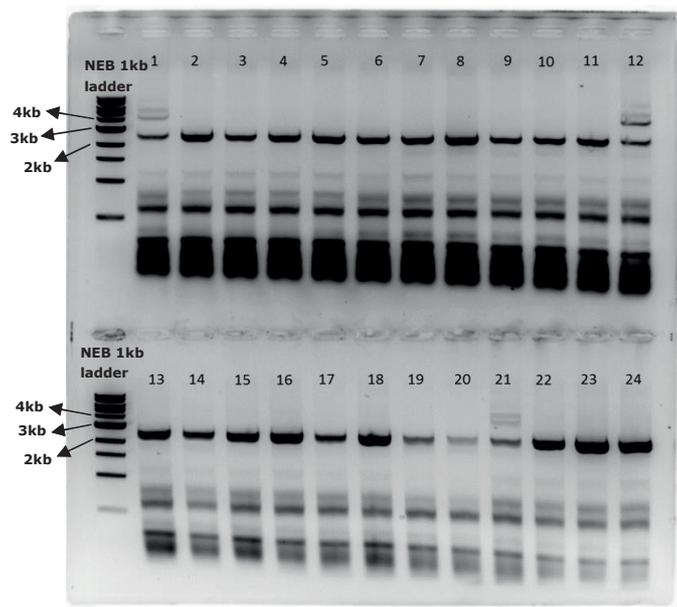
Funding

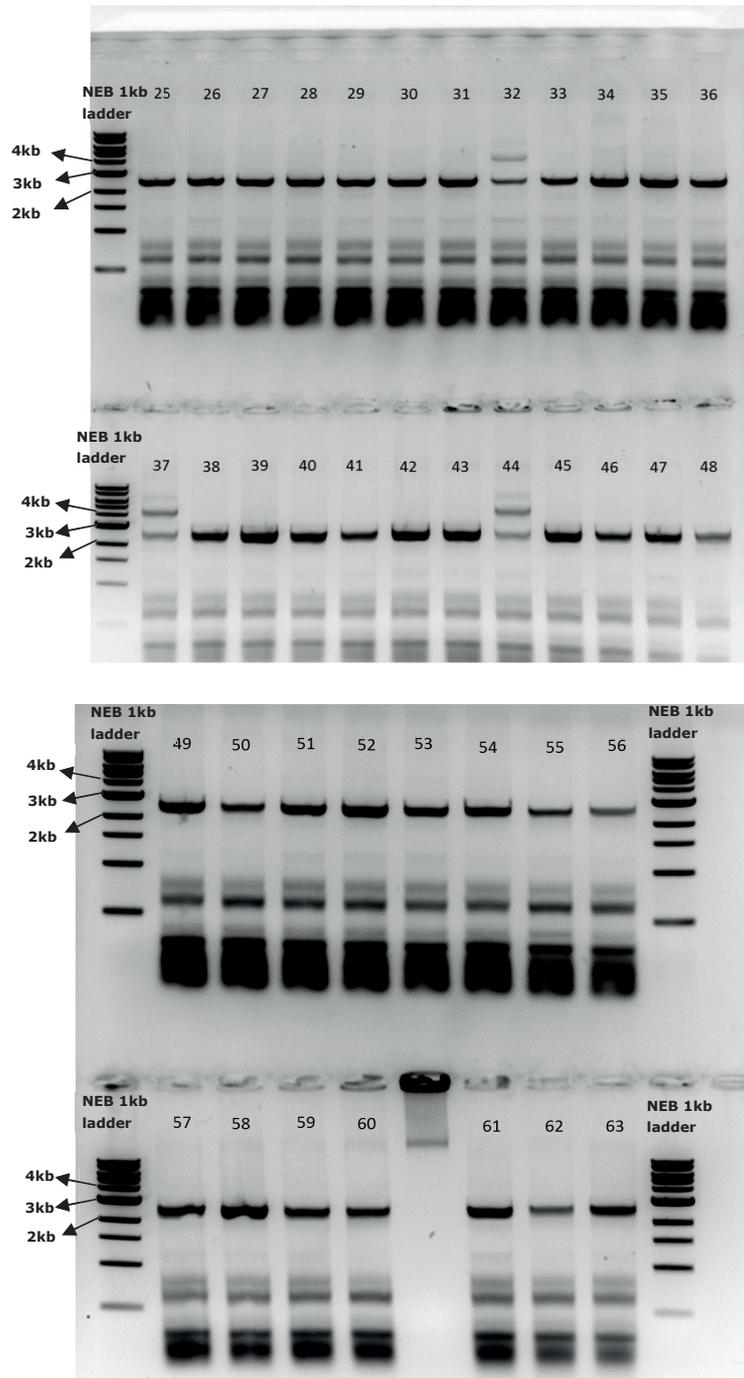
This work is supported by the project titled “Microalgae As a Green source from Nutritional Ingredients for Food/Feed and Ingredients for Cosmetics by cost effective New Technologies” [MAGNIFICENT], funded by Bio-based Industries Joint Technology Initiative under European Commission [project ID: 745754]

Supplementary data



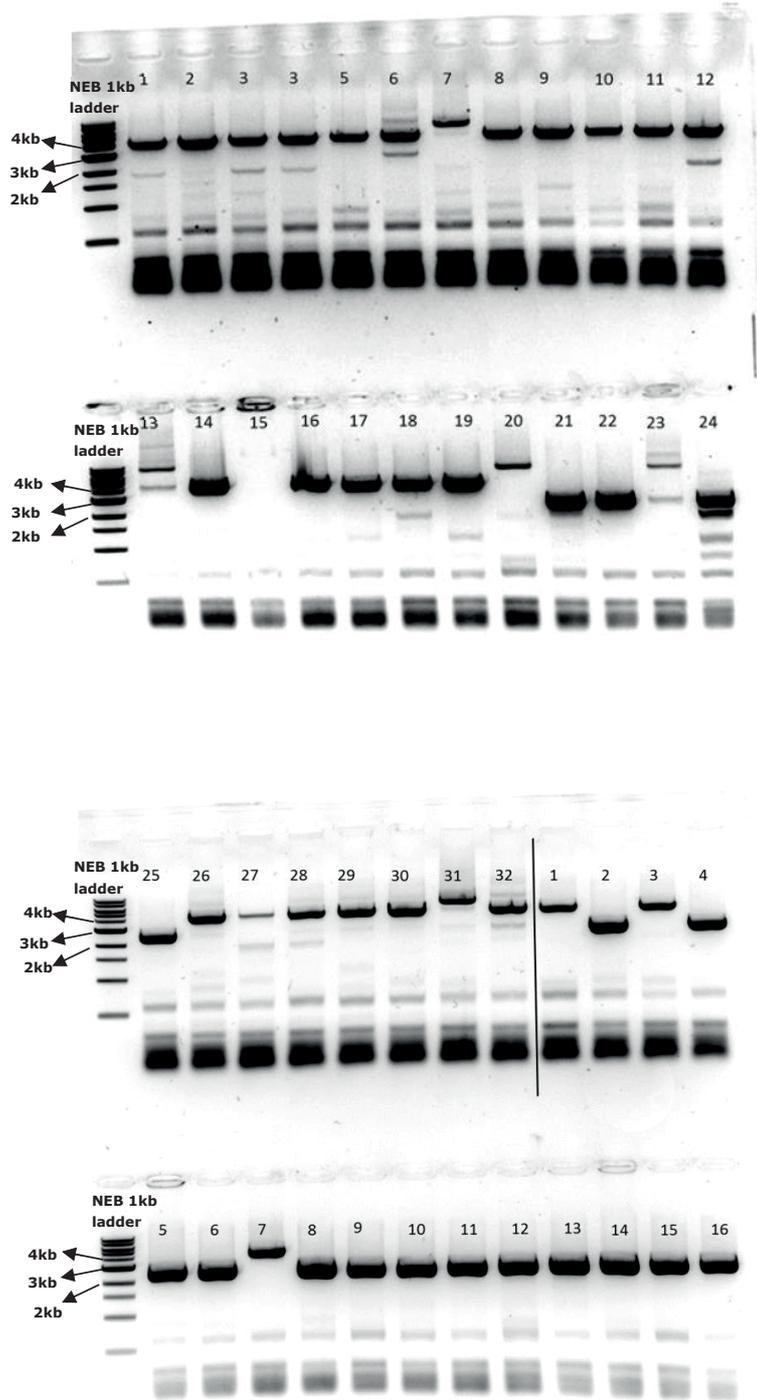
S1. *In-vitro* assays of various RNP complexes used in the study. The assay was performed at both 25°C and 37°C. The Cas12a variants are predicted to cleave the target fragment (1207 bps) into 2 smaller fragments of respective sizes based on the guide sequence; sp1: 314 bps and 893bps, sp2: 443 bps and 774 bps, sp3: 458 bps and 750bps. The Cas9 RNP with sp1 upon cleavage produced fragments of size 183 bps and 1022 bps while sp2 produced bands of size 200 bps and 1005 bps.

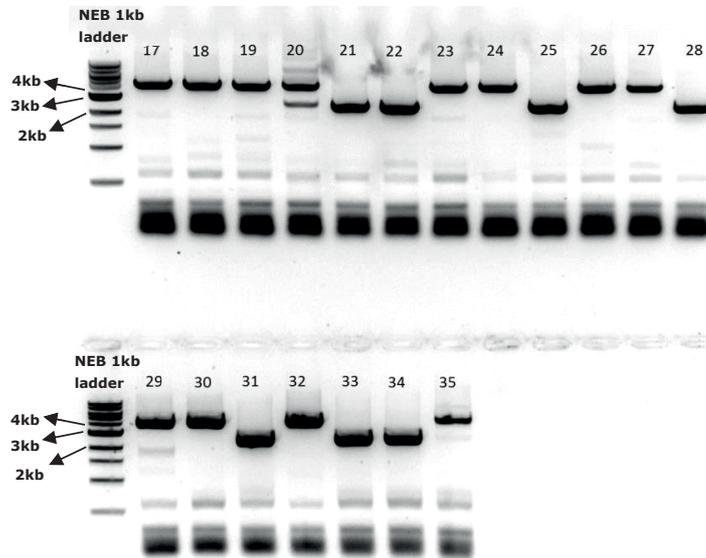




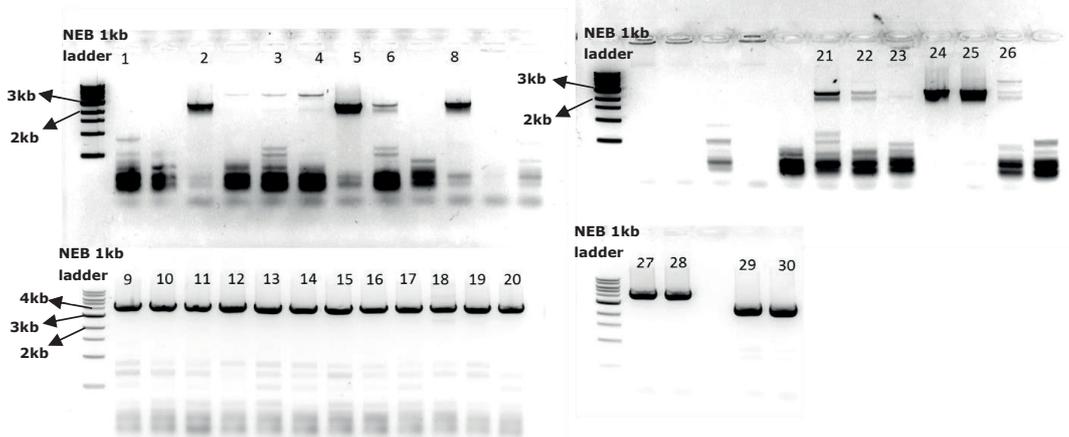
3

S2. cPCR results of 1kb HR transformants. A total of 63 zeocin resistant colonies were screened to check the presence of mutant colonies. The mutant colonies generates an amplicon of size 4337 bps while the wild type results in 2591 bps long fragment upon PCR. 57 colonies resulted in WT amplicon while 6 colonies indicated both mutant and WT bands.

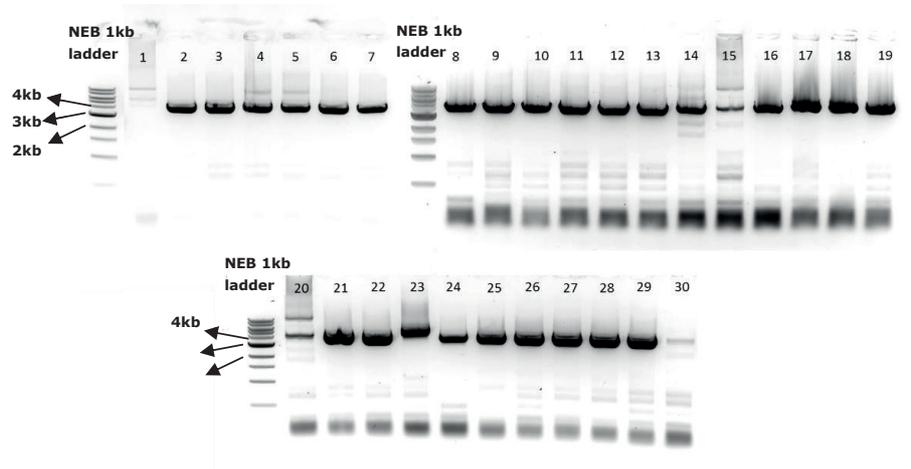




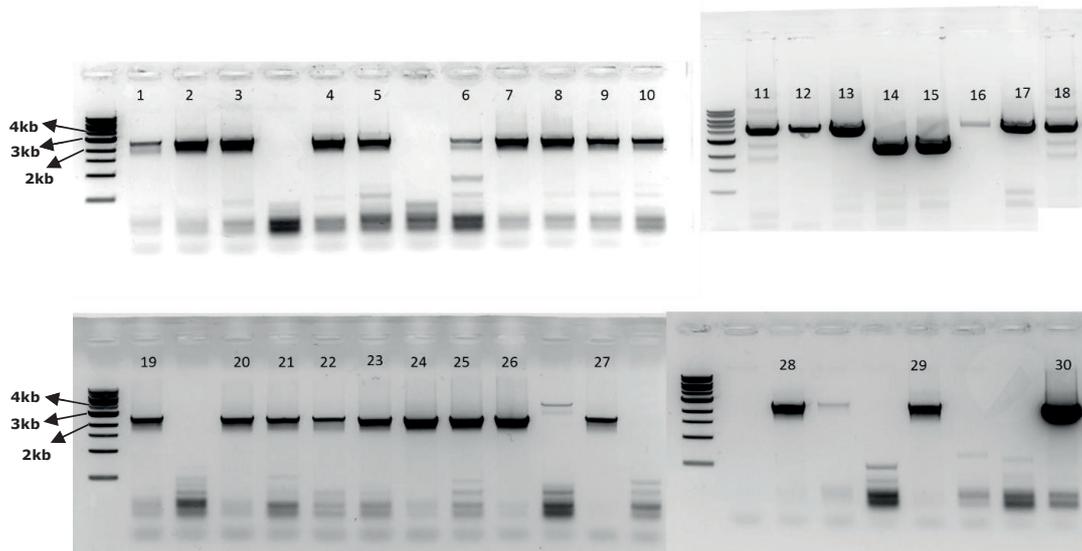
S3. cPCR results of Cas9 RNP based HDR. First 32 colonies are from transformants targeted with RNP harbouring guide 1 and next 35 colonies with guide 2. 22 out of 32 colonies screened for the NR targeting Cas RNP with guide 1 depicted mutant bands. For RNP with guide 2, among the 35 colonies screened 11 colonies were confirmed for mutant colonies.



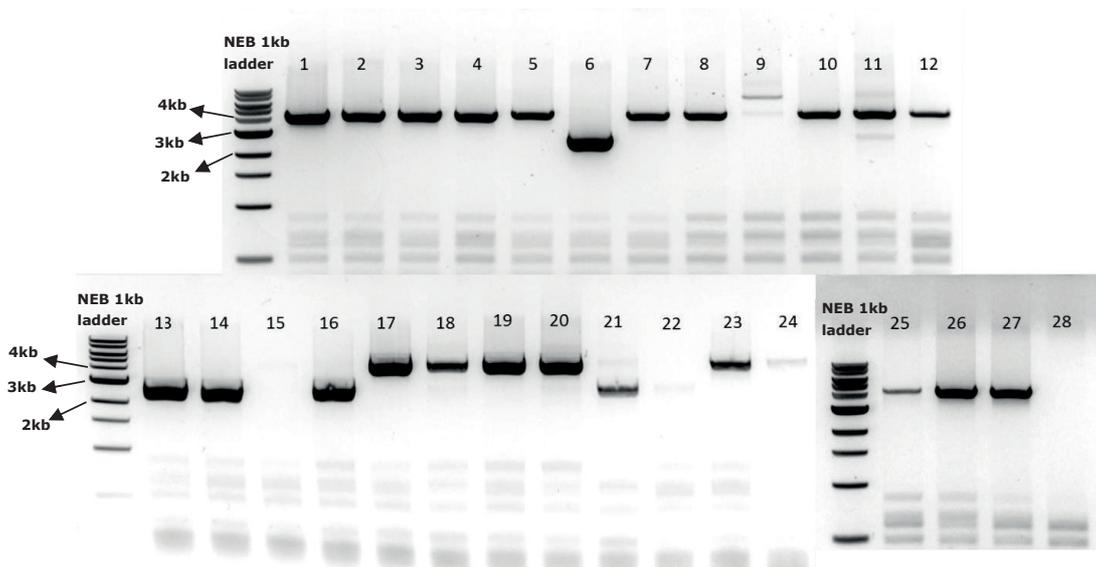
S4. cPCR results of FnCas12a-1 RNP based HDR. 30 colonies were screened and we observed that around 13 colonies produced the mutant band size of 4337bps.



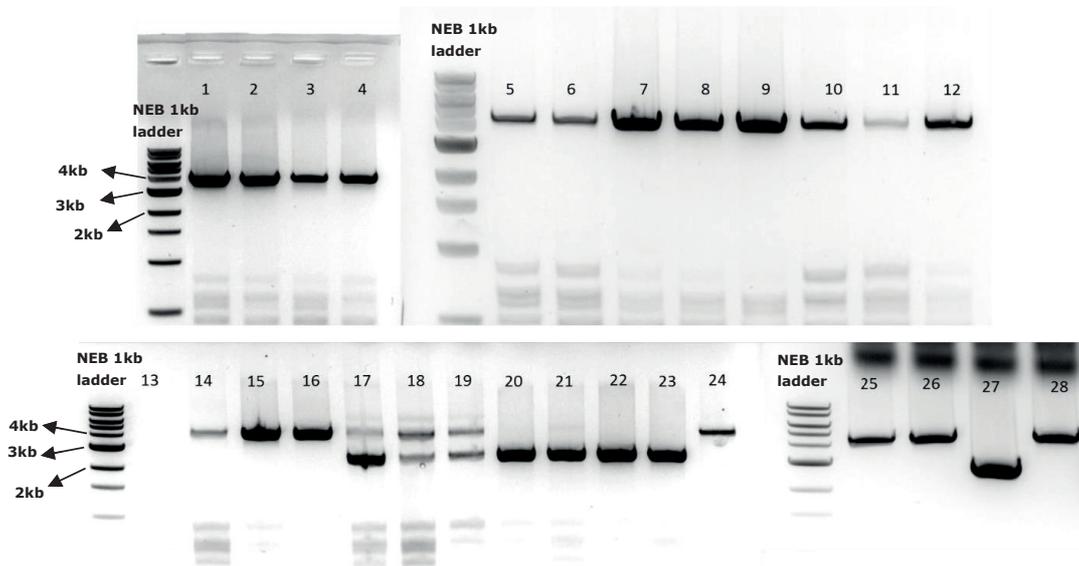
S5. cPCR results of FnCas12a-2 RNP based HDR. Out of the 30 colonies screened, 26 colonies were found to be the mutant colonies with the PCR band size of 4337 bps while the other 4 colonies indicated both the mutant and wild type (2591bps) bands



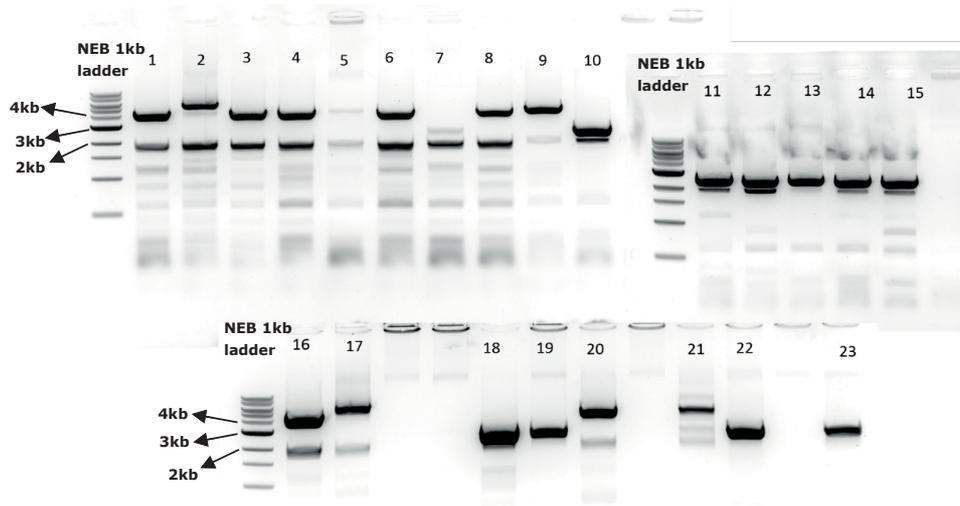
S6. cPCR results of FnCas12a-3 RNP based HDR. Only 6 colonies yielded mutant bands (4337bps) with this RNP among the 30 colonies screened



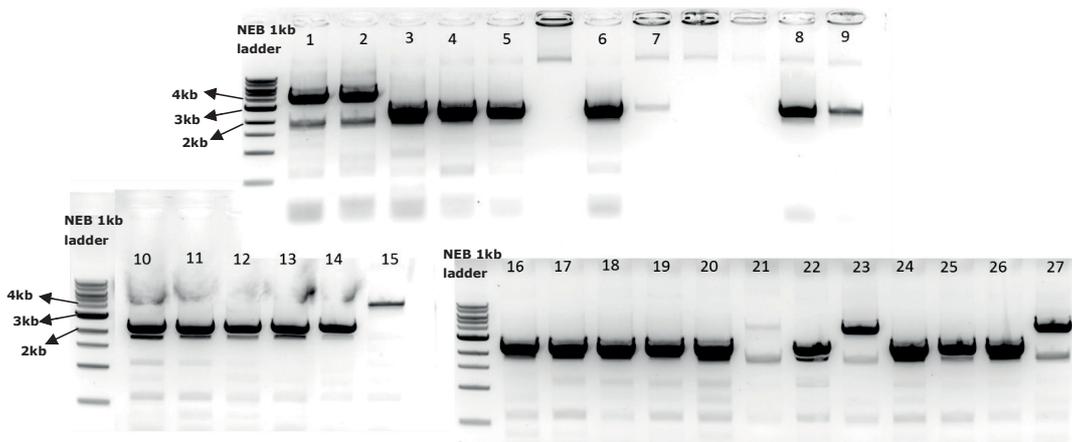
S7. cPCR results of LbCas12a-1 RNP based HDR. (cPCR 15,22 and 28 was repeated and we observed 22 was NR-KO while the other two were WT). 22 colonies among the 28 screened colonies indicated the mutant band size.



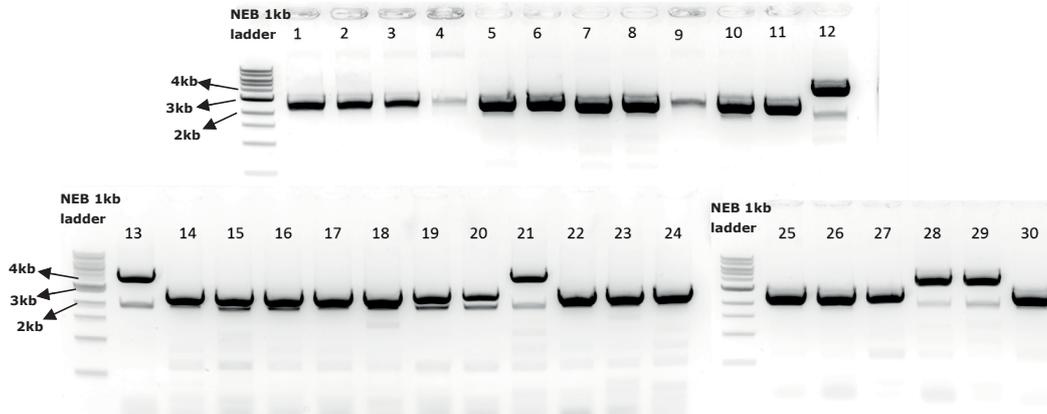
S8. cPCR results of LbCas12a-2 RNP based HDR. 19 colonies out of 28 colonies indicated the mutant bands.



S9. cPCR results of LbCas12a-3 RNP based HDR. No mutant colonies were observed

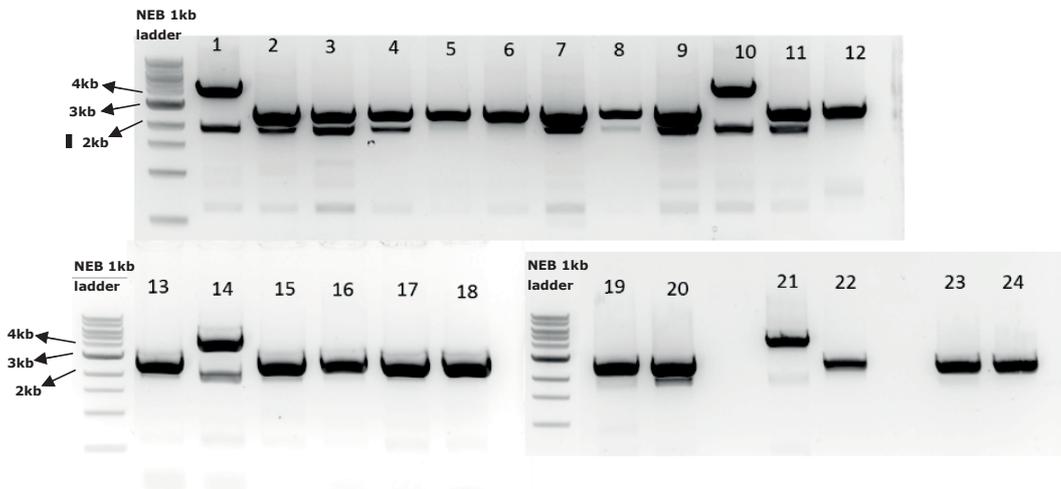


S10. cPCR results of AsCas12a-1 RNP based HDR. Obtained a single NR mutant colony among the 27 colonies screened

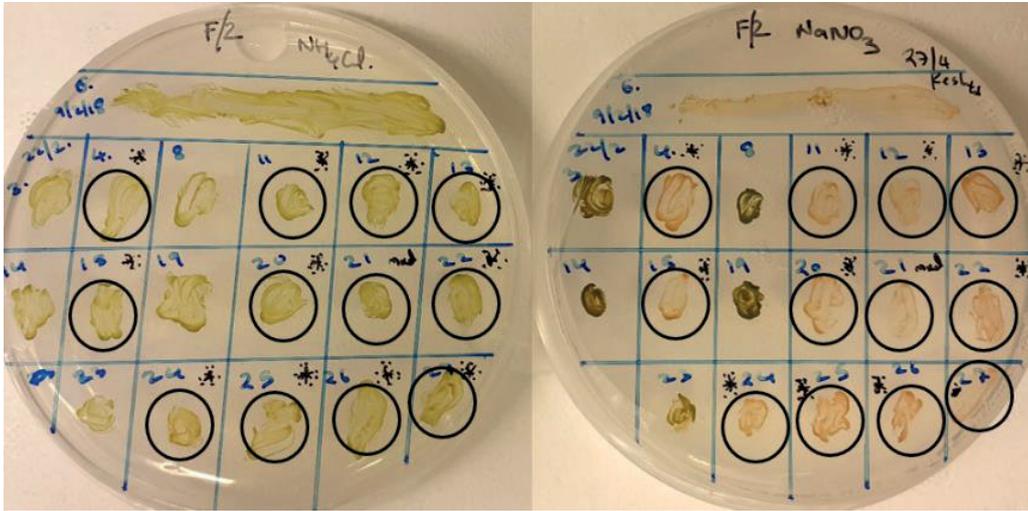


S11. cPCR results of AsCas12a-2 RNP based HDR. No mutant colonies obtained among the 30 colonies screened.

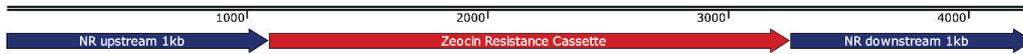
3



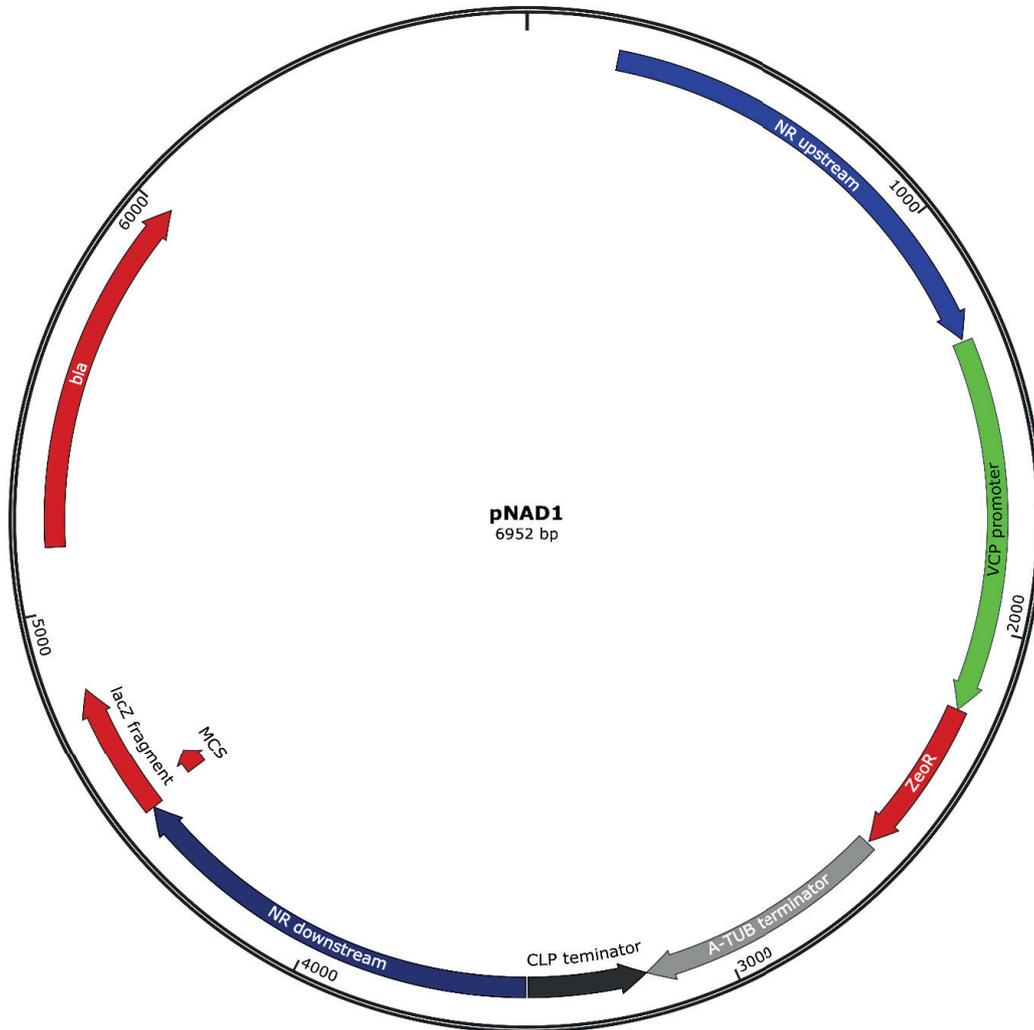
S12. cPCR results of AsCas12a-3 RNP based HDR. 1 mutant colony indicating the mutant band size of 4337bps was observed.



S13. WT and NR-KO (circled) mutants re-streaked on to F/2 media with Ammonia (left) or nitrate (right) as the sole nitrogen source. The mutants were found to bleach on media with nitrate as the sole nitrogen source while they grew well on the plate with ammonia as the nitrogen source. The image was taken after 4 weeks upon re-streaking the colonies into fresh F/2 plates with appropriate N₂ source.



S14. Map of the linear DNA harbouring the HR flanks for NR gene deletion and zeocin resistance cassette.



3

S15. Map of the plasmid used for amplifying the linear DNA harbouring the HR flanks for NR gene deletion and zeocin resistance cassette. The linear DNA was amplified by PCR using primers that bind at the 5' end of upstream HR flank and 3' end of the downstream HR flank. Genbank file of the above plasmid is provided below.

CHAPTER 4



A comprehensive genome engineering toolbox for microalgae *Nannochloropsis oceanica* based on CRISPR-Cas systems

Mihris Ibnu Saleem Naduthodi^{1,2}, Christian Südfeld², Emmanouil Klimis Avitzigiannis¹, Nicola Trevisan¹, Eduard van Lith¹, Javier Alcaide Sancho¹, Sarah D'Adamo², Maria Barbosa² and John van der Oost^{*1}

¹Laboratory of Microbiology, Wageningen University and Research, Wageningen, Netherlands.

²Bioprocess Engineering, AlgaePARC, Wageningen University and Research, Wageningen, Netherlands.

*Corresponding Author

Adapted from publication ACS Synth. Biol. (2021).
<https://doi.org/10.1021/acssynbio.1c00329>

Abstract

Microalgae can produce industrially-relevant metabolites using atmospheric CO₂ and sunlight as carbon and energy sources, respectively. Developing molecular tools for high-throughput genome engineering could accelerate the generation of tailored strains with improved traits. To this end, we developed a genome editing strategy based on Cas12a Ribonucleoproteins (RNPs) and homology directed repair (HDR) to generate scarless and markerless mutants of the microalga *Nannochloropsis oceanica*. We also developed an episomal plasmid-based Cas12a system for efficiently introducing indels at the target site. Additionally, we exploited the ability of Cas12a to process an associated CRISPR array to perform multiplexed genome engineering. We efficiently targeted 3 sites in the host genome in a single transformation, thereby making a major step towards high throughput genome engineering in microalgae. Furthermore, a CRISPR interference (CRISPRi) tool based on Cas9 and Cas12a was developed for effective downregulation of target genes. We observed up to 85% reduction in the transcript levels upon performing CRISPRi with dCas9 in *N. oceanica*. Overall, these developments substantially accelerate genome engineering efforts in *N. oceanica*, and potentially provide a general toolbox for improving other microalgal strains.

Keywords

Nannochloropsis, microalgae, genome editing, CRISPR-Cas, gene silencing, ribonucleoproteins, Cas9, Cas12a.

Introduction

The concept of microbial biofuel production has received major attention since it was realized that the exploitation of fossil fuels has become a threat to life on earth. Microalgae were already identified as a promising platform for sustainable production of biofuels in the 1950s [36-39]. The possibility to achieve higher photosynthetic efficiency in microalgae (10%) compared to that of land plants (~5%) and with algal lipid accumulation reaching up to 60% of their cell dry weight (CDW) make them suitable candidates for biofuel production [179, 221-225]. Microalgae belonging to the genus *Nannochloropsis* have attracted scientific and industrial attention in the last decade, as they can accumulate lipids up to 60% of CDW under certain conditions, including high-value omega-3-polyunsaturated fatty acids (PUFA) [50, 226].

Developing convenient genome editing tools is key for elucidating relevant details of the microalgal metabolism, and for developing targeted variants with desired traits. CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated proteins) based genome editing has been effectively implemented in *Nannochloropsis* spp. [79, 80, 146, 182, 227, 228]. A type II Cas endonuclease from *Streptococcus pyogenes* named SpCas9 is the predominantly used Cas variant for non-homologous end joining (NHEJ) mediated generation of gene disruption mutants [229, 230]. To this end, the Cas9 gene was either integrated into the host genome or into an episomal plasmid. The advantage of an episomal system is the possibility of generating non-transgenic mutants, as markerless deletions can be made, and the plasmid is generally cured upon relieve of antibiotic pressure [79, 228]. However, the dependence on ribozyme systems for precise sgRNA generation for Cas9-based editing [95], requires relatively complex and laborious cloning procedures to develop the plasmids for this approach.

Using delivery of Cas ribonucleoproteins (RNPs) as an alternative to plasmid transfection, could overcome the limitations posed by the plasmid-based systems. Almost a decade ago, homologous recombination (HR)-based insertion of an antibiotic resistance gene was reported as a tool for precise gene insertion or targeted gene disruption in *Nannochloropsis* spp [78]. In our previous study, we improved the HR efficiency in *Nannochloropsis oceanica* by using Cas RNPs for generating double strand breaks (DSBs) at the DNA target site that were subsequently restored by the host's homology directed repair (HDR) system [227]. However, the mutants developed by this approach did harbor an antibiotic resistance cassette at the target site. This outcome is a drawback for generating multiple knockouts as the available antibiotic resistance genes are limited for *Nannochloropsis* spp [231].

Here, we describe further development of the RNP-based strategy for generating scarless mutants in *N. oceanica* using Fluorescence-Activated Cell Sorting (FACS). Additionally, we now developed a plasmid-based genome editing system employing *Francisella novicida* Cas12a (FnCas12a) [102, 103]. As such, we did overcome the dependence on ribozyme-based guide maturation systems by exploiting the ability of Cas12a to auto-process the precursor CRISPR RNA (pre-crRNA) to mature crRNA guides, as a way to simplify the plasmid development. The latter feature of Cas12a also allowed us to develop a multiplexed targeting system that can simultaneously produce multiple knockouts in *N. oceanica* in a

single transformation. Moreover, we also report an effective CRISPR interference (CRISPRi) tool for downregulating target gene expression using Cas9 and Cas12a proteins. Finally, we attempt the novel prime editing technique using nickase Cas9 (nCas9) in *N. oceanica* IMET1 to develop precise mutants without inducing DSBs in the host genome.

Results

Lb/FnCas12a RNP-based markerless mutant generation and FACS-based selection

Developing non-transgenic mutants using RNPs can have widespread applications in microalgal biotechnology. To this end, we set out to utilize the phenotypic variation of transformants expressing a fluorescence marker integrated at the target site followed by a FACS mediated negative selection to generate markerless microalgal mutants.

The first step towards exploiting FACS for markerless genome editing was to knock out the target gene by replacing it with a fluorescence marker gene. To this end, the zeocin resistance (ZeoR) gene (*shble*, GenBank accession number A31898.1) described previously [227] was linked, via the 30 amino acid long P2A peptide sequence to the gene encoding the DsRed variant *tdTomato* [81, 232] (Figure 1A). This selection marker construct was inserted between HR flanks designed to precisely knock-out the *Nitrate reductase* (*NR*) gene. The *NR* gene was selected as target because characterizing *NR* mutants is easily achieved by testing for their loss of ability to grow on nitrate [79, 146] (Figure 2D). The RNP used for transformation was composed of a sgRNA targeting the *NR* gene and the *Lachnospiraceae* bacterium Cas12a (LbCas12a) protein. After transformation of the RNP and the linear marker fragment, obtained colonies were screened by PCR to check for predicted insertion of the editing template at the target site. Among 19 colonies, two colonies showed the desired target gene substitution. The mutant strains were analyzed by flow cytometry to confirm *tdTomato* expression. Fluorescence emission of mutants was easily separable from the wild type cells (Figure S1).

In the second step, we aimed to remove the ZeoR-*tdTomato* marker cassette to produce a marker-less nitrate reductase mutant of *N. oceanica*. To this end, we generated two RNP complexes, consisting of *Francisella novicida* Cas12a (FnCas12a) and two crRNAs for targeting different positions of the *tdTomato* gene. The ZeoR/*tdTomato*-expressing parental strain was transformed with the selection cassette targeting RNPs and the editing template that contains the 1Kb upstream and downstream HR flanks joined together (Fig. 1B). The transformed cells were recovered for six days to allow the degradation of the *tdTomato* protein. Next, the recovered transformants that contained a mix of non-edited parental (high *tdTomato* fluorescence) and edited (low *tdTomato* fluorescence) were sorted by FACS (Figure 1A, S1). From two independent transformations for each RNPs, the cells were sorted by FACS in order to obtain *tdTomato*-free clones, that were screened by colony PCR. 87 colonies for RNP with spacer 1, and 91 colonies for RNA with spacer 2 were sorted out as *tdTomato* less clones. We observed that all the colonies obtained via RNP with spacer 1 were false positives and did not contain any markerless mutants. However, for RNP harbouring spacer 2, 4/91 colonies showed the expected band size of markerless mutants (Figure 1C). Sequencing of PCR products confirmed the complete deletion of

the *tdTomato* and *Zeocin* resistance cassette from the host genome, thus indicating that markerless *Nitrate reductase* gene knockouts were successfully obtained.

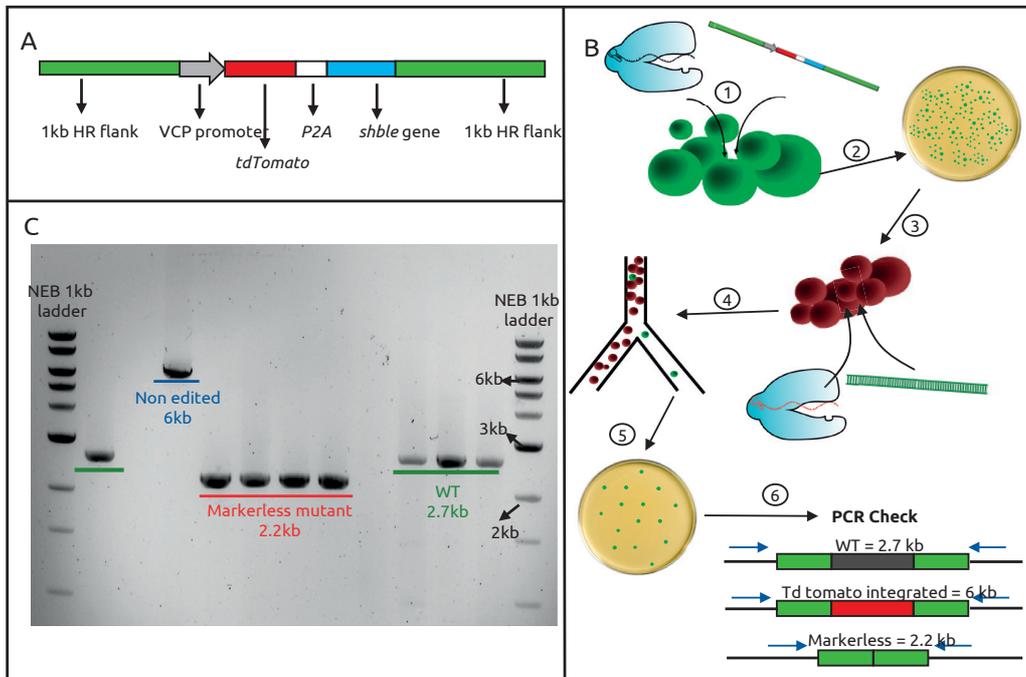


Figure 1. RNP-based genome editing strategy for generating markerless mutants. A) Editing template for introducing *tdTomato* fluorescence gene B) RNP-based approach for generation of scarless mutants: 1. The transformation of editing plasmid and RNP targeting the nitrate reductase gene, 2. Selection on antibiotic plates, 3. Selection and scaling up of mutants harboring designed integration of editing plasmid and *tdTomato* expression, 4. Transformation of *tdTomato* expressing mutants with editing plasmid and RNP targeting the *tdTomato* gene, 5. Sorting of cells that lack the *tdTomato* expression and growing on solid media without antibiotic selection, 6. PCR on obtained colonies to identify the scarless and markerless mutants. C) PCR analysis showing the various genotypes obtained after the sorting.

CRISPR-Cas-based genome editing in microalgae was first reported with feeble efficiencies in *Chlamydomonas reinhardtii* in 2014 [109]. Since then, various strategies were implemented in multiple microalgal strains to improve editing efficiencies or to perform compelling genetic engineering studies [233, 234]. However, in most studies the Cas9 gene was incorporated into the host genome, resulting in transgenic mutants [80, 96, 146, 182, 235, 236]. RNP-based approaches circumvented this bottleneck and developed markerless mutants by disrupting the target site with indels [138, 183, 227, 237, 238]. The strategy demonstrated in this study further optimize the RNP based genome engineering technique in microalgae by generating markerless and scarless mutants with desired mutations at the target site.

Codon harmonization and expression of *fnCas12a* in *N. oceanica*

The RNP-based approach described above required two rounds of transformation and selection via FACS to develop markerless mutants. Moreover, high cost of purified Cas protein and requirement of screening via FACS could make this approach inaccessible to wider microalgal researchers. Thus, to further accelerate the generation of mutants and to simplify the screening process, we set out to develop an episomal plasmid-based Cas12a system for genome editing in *N. oceanica*. We adapted a recently-developed episomal system for *N. oceanica* to develop our Cas12a-based genome editing system [79]. This system harbors a Cas9 gene sequence and ribozyme-mediated production of guide RNAs and it was used for deleting a 100 kb fragment in *Nannochloropsis oceanica* [228]. However, using ribozymes for precise generation of guide RNAs complicates plasmid production (especially in case of multiplexed genome editing) and limits the usefulness of this system for high-throughput application. Developing a Cas12a episomal plasmid system based on this chassis could expand the genetic toolbox for *Nannochloropsis* and address the bottlenecks associated with the Cas9 system. We opted to use the FnCas12a variant due to its improved activity in *N. oceanica* compared to LbCas12a and AsCas12a [227].

The *fnCas12a* gene sequence was codon harmonized for expression in *N. oceanica* using the online Galaxy codon harmonizer tool [239]. To the 3'end of the harmonized *fnCas12a* gene, an in-frame fusion was made with a gene encoding the reporter protein Nano Luciferase (Nlux) to allow validation of Cas12a expression in transformants [79, 240]. Expression of the fusion protein was regulated by the bidirectional Ribi promoter in the resulting plasmid pNOC_nfnCas12a-NLux. Wild type *N. oceanica* was transformed with this plasmid and transformants were selected for antibiotic resistance. Colonies were screened for Nlux activity, but no luminescence was detected in any of the colonies.

Human codon optimized Cas9 was previously shown to be effective in *N. oceanica* [79]. Assuming the codon harmonized *fnCas12a* gene sequence was the system's bottleneck, we expressed human codon-optimized *fnCas12a* and *lbCas12a* to identify the best expressed Cas gene sequences for *N. oceanica*. The plasmids pNOC_hfnCas12a-Nlux (encoding the humanized *fnCas12a*-Nlux fusion protein) and pNOC_hlbCas12a-Nlux (encoding the humanized *lbCas12a*-Nlux) were transformed, selected using the ZeoR marker, and screened for Nlux activity. In this case, the Nlux activity was observed for both variants and the activity of the fnCas12a-Nlux fusion protein was higher than the lbCas12a-Nlux (Figure S2). This observation indicates that the codon harmonization can be detrimental for expression of certain genes and warrants further investigation. Apparently, humanized *Cas12a* sequences are suitable for expression and genome editing in *N. oceanica*. Humanized *Cas9* gene sequence was also previously reported to be efficiently expressed in *N. oceanica* [79]. The humanized *fnCas12a* gene was selected for the next round of experiments owing to its improved expression and previously validated high activity in *N. oceanica*.

Episomal *fnCas12a* system introduces indels at the target site

Next, we set out to use the humanized *fnCas12a* as a genome editing tool for *N. oceanica* by generating indels at a target site. To this end, along with the *fnCas12a*, we used a full-length CRISPR RNA (crRNA) with 36-nucleotide (nt) direct repeat (DR) followed by

a 25-nt long spacer sequence targeting the gene of interest. We selected three spacer sequences targeting the exon of the *NR* gene (crRNA NR1, crRNA NR2 and crRNA NR3). The bidirectional Ribi promoter regulated the transcription of Cas12a and the crRNA (Figure 2A). To facilitate precise crRNA formation from mRNA molecules, we added the Hammerhead (HH) and Hepatitis Delta Virus (HDV) ribozymes on either end of the crRNA sequences [79, 241, 242].

The plasmids pNOC_hfnCas12a-Nlux_crRNA NR1/ NR2 and NR3 were transformed into wild type *N. oceanica*, and 96 colonies per plasmid were selected and screened for Nlux activity. 72, 55 and 75% of the colonies from plasmids pNOC_hfnCas12a-Nlux_NR1/ NR2 and NR3, respectively, showed Nlux activity. Screening of these colonies for indels at the target site by PCR and sequencing revealed that 82, 89 and 20% of the Nlux-positive colonies for plasmids harbouring crRNA 1, 2 and 3, respectively, showed indels at the target sites (Figure 2C) (Figure S6, S7, S8). When removing the Nlux tag from the *fnCas12a* gene and screening transformant colonies directly by PCR, indel formation efficiencies were 40%, 30% and 5% for pNOC_hfnCas12a_crRNA NR1, NR2 and NR3, respectively (Figure 2C).

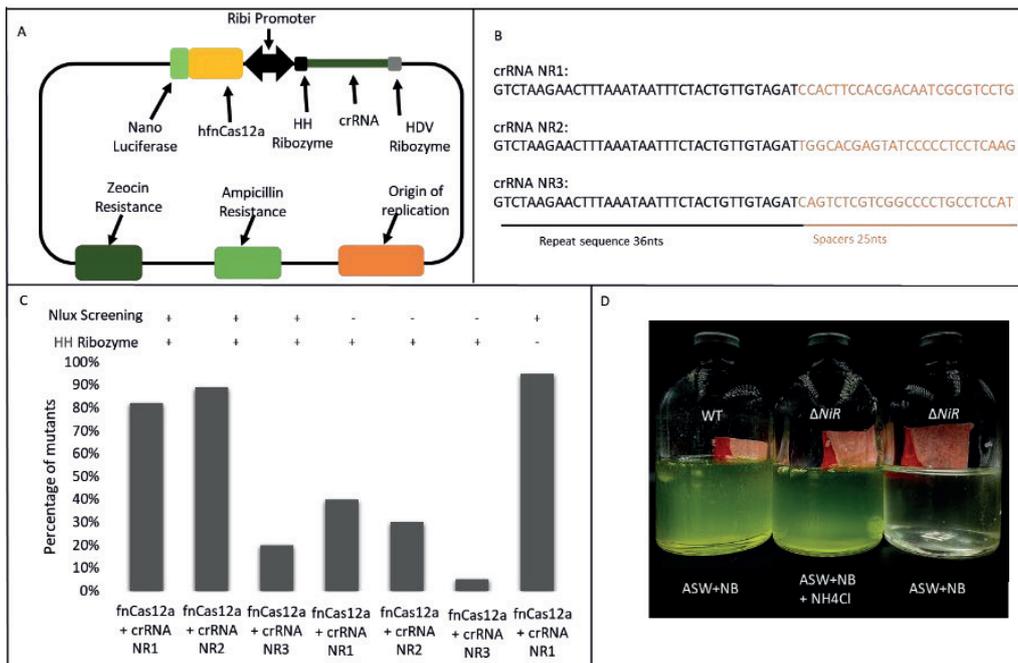


Figure 2. Plasmid-based Cas12a efficiently introduce DSB at target sites in *N. oceanica*. A) Schematic map of the episomal plasmid system used for generating NHEJ based indel mutations in *N. oceanica*. B) The various CRISPR RNA sequences targeting different positions of the nitrate reductase gene. C) The efficiencies of obtaining indel mutants upon using various plasmids. D) Phenotypic assay of nitrate reductase mutants shows the inability of mutants to grow in media without ammonia.

This 1-step approach to generate markerless mutants is substantially simpler than the 2-step procedure using negative fluorescence selection with FACS. However, it cannot easily be used for knocking out multiple genes at the same time due to complicated cloning of multiple HH ribozyme sequences. Therefore, we next tested whether the innate crRNA processing ability of Cas12a can be used to improve the plasmid-based gene knockout system by simplifying the cloning procedure. To assess the crRNA processing capability of fnCas12a in *N. oceanica*, we removed the HH ribozyme from the plasmid pNOC_hfnCas12a_crRNA NR1 to develop the plasmid pNOC_hfnCas12a_Nlux_(-HH)crRNA NR1. We achieved an editing efficiency of 95% with this plasmid (Figure 2C), indicating effective self-processing of crRNA by Cas12a enzyme in *N. oceanica*. This removal of HH ribozyme further simplifies the introduction of spacers into Cas12a plasmids for genome editing in this microalga.

Multiplexed genome editing in *N. oceanica* using fnCas12a

The efficient processing of crRNA by Cas12a can be exploited for multiplexed genome editing in *N. oceanica* [103, 243]. To this end, we introduced a CRISPR array in the crRNA locus of the targeting plasmid pNOC_hfnCas12a_crRNA NR1. The CRISPR array consisted of three 36-nt DR sequences interspaced by the spacer sequences NR1 and NR2, respectively (Figure 3A). This plasmid was developed by the CRATES assembly, a simplified method for generating CRISPR arrays for multiplexed genome engineering [244]. The NR1 and NR2 target sites were 375 nucleotides apart in the *NR* gene. Simultaneous targeting at both loci could either result in deletion of the region between the targets or introduce indels at both target sites. We transformed wild type *N. oceanica* with the multiplexing plasmid pNOC_hfnCas12a_Nlux_NR1_NR2, and selected colonies with Nlux activity for screening by PCR. We observed a reduction (approx. 375 base pairs) in amplicon size for 11/40 (27%) colonies, indicating efficient multiplexed targeting and deletion of the region between the target sites (Figure S9). Sequencing of PCR products confirmed this (Figure S10, S11) and showed that an additional 4 colonies had indels at both target loci, resulting in a total of 37% multiplexing efficiency. We further observed that 14/40 (35%) colonies carried mutations at the NR2 site, whereas no mutants had indels only at the NR1 site.

To further expand the multiplexing potential, we added an additional spacer to the CRISPR array to target three regions in a single transformation (pNOC_hfnCas12a_Nlux_NR1_NR2_LPT1) (Figure 3B). The additional spacer targeted the *LPAT1* gene, as successful non-lethal knock out of this gene has been demonstrated in *N. oceanica* previously [83]. After transformation and selecting 48 Nlux-positive colonies, we obtained 7/50 (14%) colonies that harboured mutations at all 3 target sites. 2/50 colonies (4%) showed different combinations with mutation at two target sites, while 6/50 (13%) colonies had a mutation at either one of the target sites. These results indicate the potential of using Cas12a for generating multiplexed knock-out mutants in microalgae. The mother plasmid we developed (pNOC_hfnCas12a_Nlux_CRATES) harbours type 2S restriction sites at the CRISPR array locus, which can be used for swift generation of multiplexing constructs, either by CRATES or Gibson assembly.

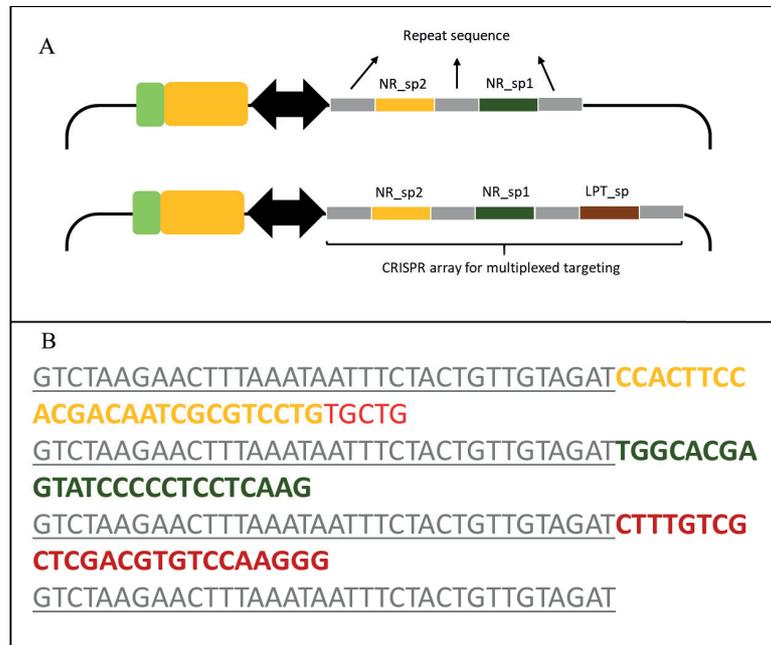


Figure 3. Multiplexing plasmids and the CRISPR array. A) The CRISPR array used for multiplexed genome engineering in *N. oceanica*. The spacers are targeting the nitrate reductase gene are indicated in yellow and green. The dark red colored spacer targets the *LPAT1* gene and was used for multiplexed editing at 3 loci in a single transformation. B) The CRISPR array sequence of the multiplexing plasmid targeting 3 regions. The colors used for the nucleotides corresponds to the annotation in panel A. The 5 nucleotides depicted in red (TGCTG) are the extensions used for the CRATES assembly to develop the plasmid.

CRISPRi mediated gene downregulation

Next to the aforementioned gene disruption approaches, the development of an efficient transcription repressor system would be very useful for accelerating fundamental studies in microalgae. The specific silencing of gene expression by CRISPRi has been successfully implemented in a wide range of organisms, including the microalga *Chlamydomonas reinhardtii* [97, 245-247]. A catalytically inactive Cas9 tagged with a transcriptional repressor domain has shown 94% gene silencing efficiency in *Chlamydomonas* [247]. We aimed to identify the efficiency of FnCas12a and SpCas9 in downregulating gene expression in *Nannochloropsis oceanica*. To this end, we developed catalytically inactive variants of SpCas9 by introducing two mutations D10A and H840 [97], and catalytically inactive fnCas12a by introducing mutations E1006A and D917A [102, 248, 249]. The deadCas (dCas) encoding genes were introduced in the aforementioned episomal system and replaced the ZeoR cassette with a blasticidin resistance cassette to establish the CRISPRi plasmids for *N. oceanica*.

Ideal target genes for this endeavour should facilitate simple quantification of gene expression. However, the wild type *N. oceanica* did not harbour a gene that met our requirement. For that reason, we used a previously constructed host strain in which the

gene encoding tdTomato was used to knockout the *NR*-encoding gene (Figure 1A, first step). We developed dCas9 and dCas12a episomal plasmid systems with spacers targeting two regions of the tdTomato CDS (Figure 5A). Plasmids expressing either Cas proteins together with a non-targeting spacer sequence were used as controls. After transformation of the tdTomato-expressing (tdTomato⁺) parental strain, colonies were selected on solid media with ammonia and blasticidin. Three colonies with active Nlux expression per construct were grown in liquid media to exponential phase and tdTomato fluorescence levels were quantified by plate reader. However, none of our transformants exhibited a significant reduction in tdTomato fluorescence levels (Figure S12). This outcome indicates that expression of dCas9 and dCas12a proteins might not be an effective CRISPRi tool for microalgae.

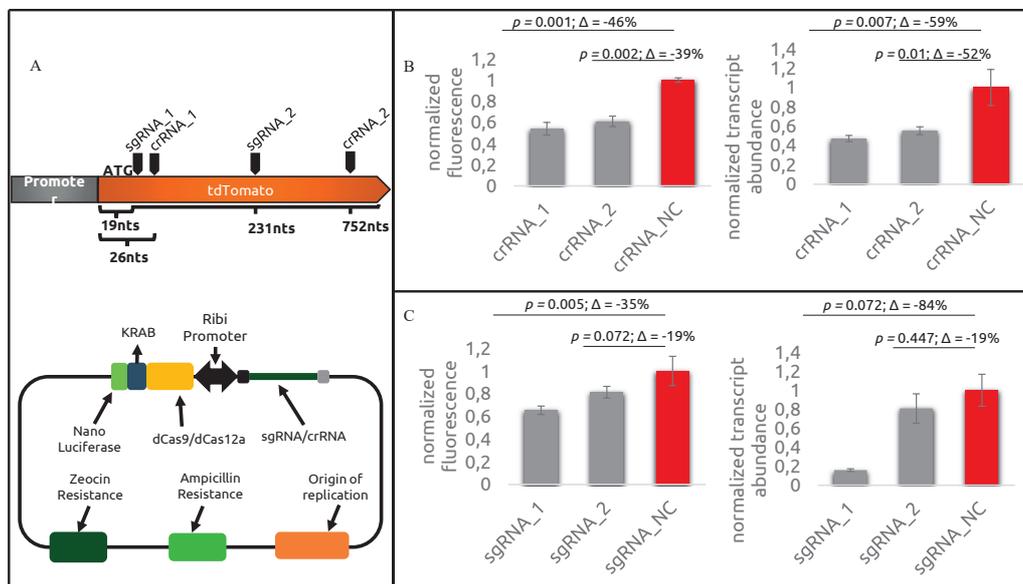


Figure 4. Silencing by CRISPR interference (CRISPRi). A) The map of dCas-KRAB plasmids for CRISPRi in *N. oceanica* and the distance of different target sites from the start codon in the CDS of tdTomato gene. B) Fluorescence and mRNA levels (qPCR) of various dCas12a-KRAB plasmids targeting the tdTomato gene. C) Fluorescence and mRNA levels (qPCR) of various dCas9-KRAB plasmids targeting the tdTomato gene. Fluorescence is normalized to OD and the fluorescence value of NC. Fluorescence (/transcript) values were normalized to the mean value of the negative control. Error bars denote the standard error of the mean (N>=2). Group means were compared for significant difference to the control using Tukey's HSD test. Relative differences compared to the control are given (Δ).

Attaching transcription repressor domains (TRD) to dCas proteins has been reported to enhance the CRISPRi efficiency in various organisms [98, 250-253]. Among various TRDs available, we selected KRuppel Associated Box (KRAB) for tagging dCas proteins as this variant domain was active in *Chlamydomonas* [247]. Next, we developed the same plasmids as in the previous experiment but with the KRAB domain sequence attached to the dCas-encoding sequences at their 3'-termini (Figure 4A). After transformation, selection

on antibiotic and Nlux screening, Nlux positive colonies were screened for tdTomato fluorescence during exponential growth phase. Various tdTomato downregulation levels were observed for cultures harbouring plasmids with spacers targeting the tdTomato gene (Figure 4B and 4C). Reduction in fluorescence levels was observed for both dCas9-KRAB and dCas12a-KRAB. To further validate this observation, RNA isolation followed by cDNA synthesis and qPCR were performed (Figure 4B and 4C), which revealed up to 85% reduction in transcript levels for spacers targeting the beginning of the CDS with dCas9-KRAB. The highest repression with dCas12a-KRAB was also observed for spacer targeting the 5'-end of the CDS. Weaker repression was observed for spacers targeting loci further downstream the start codon of the target gene. These observations are in line with previous studies showing reduced CRISPRi upon targeting more towards the 3'-end of the CDS and higher efficiencies when targeting close to the transcription/translation start site [97, 99].

Here, we report substantial improvements to the genome editing toolbox for microalgae that address most of the bottlenecks associated with the present CRISPR-Cas tools. Combining the low risk of off-target effects when using transiently present Cas RNP with the ability to generate scarless mutants could be exploited for the sequential generation of non-transgenic mutants with multiple alterations in the genome. However, further optimizations are required to improve the efficiencies of obtaining markerless mutants via FACS. The development of plasmid-based systems that employ Cas12a is a valuable addition to the existing Cas9 system, as it expands the available target sites in the A-T rich regions of the host genome for microalgal genome editing. Furthermore, efficient multiplexing and the elimination of the need for ribozymes compared to Cas9 systems, strengthens the position of Cas12a as enzyme of choice for genetic engineering of *Nannochloropsis* spp and other microalgae. Finally, the development of an efficient CRISPRi system completes the essential tools required for extensive microalgal genome engineering. Expanding the genome editing toolbox with improved techniques will accelerate the development of microalgae variants for fundamental and applied purposes [254].

Towards prime editing in *N. oceanica* IMET1

As observed in CRISPRi experiments, attaching various protein domains to Cas proteins can repurpose Cas for many applications in addition to inducing DSDBs. The recent development of base editing and prime editing technologies, that facilitate precise engineering of nucleotides in a programmable manner in the absence of a separate editing template are beautiful examples of these applications [255-259]. Base editing (BE) can efficiently perform four transition mutations such as A→G, T→C, C→T and G→A without inducing DSDBs [255, 256]. Prime editing (PE) is a new addition to the existing genome engineering toolbox that facilitates all the possible base to base conversions in addition to targeted insertions and deletions [259]. PE uses a nickase Cas9 tagged with an engineered reverse transcriptase (RT) enzyme and a prime editing guide RNA (pegRNA). The pegRNA is a prolonged version of sgRNA that, apart from the targeting guide at the 5' end, contains a 3' extension that consists of (i) a short primer (PBS, that base pairs to the displaced DNA strand) and (2) the genetic information to edit the target site with precise insertions, deletions or substitutions. The targeted nickase activity by the PE fusion

protein on the PAM containing strand releases the 3'-hydroxyl group, which hybridizes with the PBS at the 3' end of the pegRNA. The PBS consist of 11 to 15nts and is designed to base pair with the displaced, cleaved non-target strand. The PBS-bound part of the cleaved strand serves as a primer for the reverse transcription. Towards the 5' side of PBS on the pegRNA, the RT template harbours the required edit-encoding extension followed by few nucleotides complementary to the opposite strand to facilitate the hybridization between the edited and non-edited strands. Thus, the 3' end of the nicked strand will harbour the required edit while the 5' end remains the same. Although the unedited 5' end will be thermodynamically favoured for the hybridization, these 5' flaps are cleaved specifically by endonucleases such as FEN1 [260]. This facilitates the hybridization of edited 3' end to the opposite strand and subsequent copying of the edited information to the complimentary strand to permanently alter the genetic sequence at the target site (Figure 5).

The optimized prime editor 2 (PE2, containing a fusion of nCas9 (H840A) and a pentamutant of M-MLV RT), was used for demonstrating this technique in *N. oceanica* IMET1 [259]. As the nCas9 sequence in the PE2 plasmid was 81% identical to the previously used Cas9 sequence (CRISPRi), the nCas9 and M-MLV RT was directly used for developing the PE plasmids for IMET1. Additionally, the PE2 was tagged with the Nlux reporter to validate the expression of the fusion proteins. Initially, three plasmids with pegRNAs designed to introduce a 4bps deletion, 3bps insertion and 3bps substitution on the nitrate reductase gene were developed using the episomal plasmid backbone. Nevertheless, Nlux activity was not detected in any of the colonies transformed with these plasmids nor any editing was observed at the target site.

Assuming the outcome was due to the codon bias of the Cas9 sequence in the PE2 plasmid, this Cas9 sequence was replaced by the humanized wild type Cas9 sequence. The hCas9 fused with M-MLV RT and Nlux was introduced in the plasmid with pegRNA designed for 4nts deletion at the target site to validate the expression of the fusion protein complex. 48 colonies were screened for Nlux activity from the transformants, and 8 colonies with significant luminescence was further screened for edits at the target site by sequencing. Among these 8 colonies, 3 colonies indicated clean WT bands and 4 colonies had inconclusive sequencing results. Interestingly, one of these 8 colonies showed a clear 4 nucleotide deletion at the target site that was identical to the edit introduced in the pegRNA. The low efficiency to generate indels or prime editing-like indels could be due to the long fusion protein (WT Cas9-RT-Nlux) or the pegRNA. To assess this hypothesis, Nlux was removed from the plasmid to simplify the fusion protein. Twenty colonies were screened for indels at the target site, and 2 colonies had indels (not prime edited) at the target site while 16 colonies were WT and the other 2 colonies produced inconclusive results. As the efficiency of this system was also very low, the pegRNA was replaced by sgRNA with the same spacer sequence on the previously used plasmid. This plasmid generated indels at the target site with 50% efficiency (10/20), suggesting that the pegRNA could be responsible for the reduced efficiency.

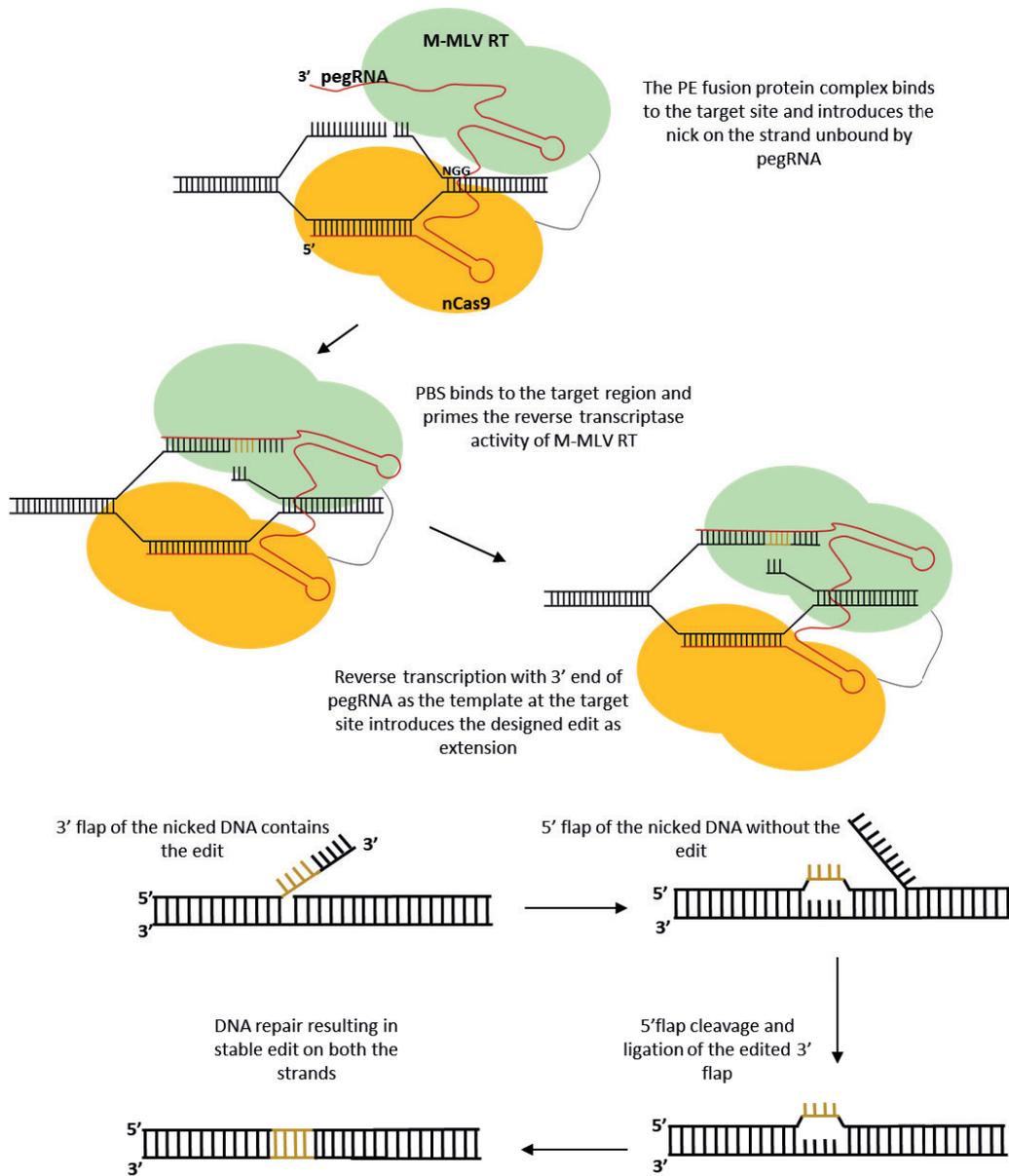


Figure 5. Prime editing mechanism to produce specific edits in the targeted DNA sequence.

Next, a H840A mutation was introduced to develop the nickase version of hCas9 (hnCas9) and tagged with the M-MLV RT protein to produce the PE fusion protein in *N. oceanica* IMET1. The three pegRNAs (4nts deletion, 3nts insertion and 3nts substitution) developed in the initial experiment were tested along with the PE fusion protein to produce prime edited mutants in *N. oceanica* IMET1. For each plasmid, 96 colonies were screened for edits at the target site. However, no colonies had indels nor prime edited-like indels. To increase the possibility of selecting colonies with active PE fusion protein expression, we

introduced the Nlux at the 5' end of M-MLV RT and selected Nlux active colonies after transformation for screening indels at the target site. From 72 colonies tested for Nlux activity, 20 colonies with Nlux activity were selected for checking the target site for indels or prime editing. The outcome of this approach was also similar to the previous strategy, as no mutants were observed for any of the plasmids.

Replacing the sgRNA with pegRNA reduced the targeting efficiency from 50% to 10% with hCas9-RT fusion protein. Using pegRNA along with hnCas9-RT did not produce any indels nor prime editing at the target site. However, the isolation of a possible prime edited mutant while using hCas9-RT-Nlux fusion protein along with the pegRNA indicate promising activity of the system in IMET1. The reduced activity of pegRNA in combination with hnCas9 could be the bottleneck in isolating prime editing mutants. Introducing an additional nick on the non-edited strand using sgRNA, few nucleotides away from the pegRNA target site is reported to enhance the efficiency of prime editing [259]. In this regard, future studies should check the possibility of using WT Cas9 for performing prime editing. In our experiments, extensive optimization of the pegRNA was not performed. The variation in the length of PBS was reported to have site specific effects on the efficiency of generating prime edited mutants [259]. Future studies should also determine the optimal PBS length for the selected target site. Additionally, codon harmonizing the RT gene sequence is a potential strategy that could increase the expression of fusion protein and thereby improve the efficiency of prime editing in IMET1.

To conclude, prime editing technology still is in an early stage of development. In a few years this technique may take-off as a potential alternative to HDR in microalgae and other organisms. The advances made in developing a prime editing system for *N. oceanica* IMET1 confirms the expression and targeting capability of the fusion protein complex. However, the reduced targeting efficiency while using the pegRNA is a major bottleneck in establishing this technique in IMET1. Future studies should focus on optimizing the pegRNAs to maximize their targeting efficiency and also compare the WT and nickase versions of Cas9 proteins in producing prime edited mutants. The next generation of this technology would be using Cas12a in place of Cas9 to perform multiplexed prime editing. These advancements will facilitate the generation of multiplexed mutants with designed edits at the target sites.

Table 1. Results of prime editing experiments in IMET1. (X- Not tested).

	pegRNA	sgRNA	Remarks
PE2-Nlux		No mutants	PE2 was not expressed in IMET1
hCas9-RT-Nlux	Obtained 1 prime edited-like colony from 48 screened colonies		X Fusion protein expression was validated by Nlux assay
hCas9-RT	Two colonies with indels from 20 screened colonies	Ten colonies with indels from 20 screened colonies	pegRNA reduces the targeting efficiency of hCas9
hnCas9-RT	No mutants		X pegRNA with hnCas9-RT is inefficient in generating indels or prime edited mutants
hnCas9-RT-Nlux	No mutants		X

Plasmid loss assay in mutant strains of *N. oceanica* IMET1

The major advantage of using an episomal plasmid system was the possibility to generate non-transgenic mutants. The mutants were reported to be cured of the episomal plasmid upon removal of antibiotic pressure [79]. To this end, a *NR* mutant developed in this study was cultured in antibiotic-less liquid media for three passes. Subsequently, we observed that the mutant lost its antibiotic resistance phenotype, and colony PCR was performed with plasmid specific primers to validate the plasmid loss (Figure 6). The PCRs generated multiple amplicons, but the length of these amplicons did not correspond to the amplification of plasmid DNA. Moreover, similar amplicons were observed upon amplifying a WT colony with the same primer sets. These observations indicate the loss of plasmids from the mutant strains and thus producing a non-transgenic *NR* mutant of IMET1.

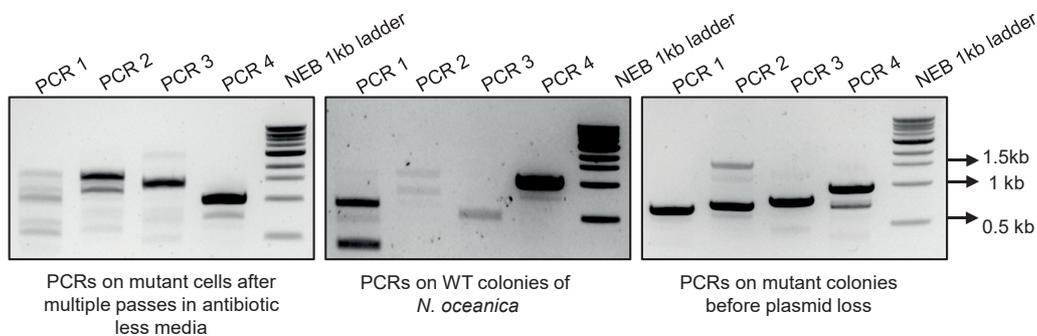


Figure 6. Colony PCR to validate the plasmid removal from cells cultured in antibiotic less media. The PCRs 1, 2 and 3 uses primer set (BG25541:BG25542, BG25543:BG25544 and BG25545:BG25546) that specifically amplifies the Cas12a sequence on the episomal plasmid. PCR 4 is a positive control that has primers specific (BG24321:BG24322) to the genomic DNA and amplifies downstream to the targeted *NR* gene. PCR1, 2 and 3 are expected to amplify fragments of sizes 670 nts, 727 nts and 810 nts, respectively from the episomal plasmid. PCR 4 is designed to amplify a 1 kb long fragment as shown in the gel on the right end. The bands observed upon amplifying the mutant without antibiotic resistance and the WT lack these bands indicating the absence of episomal plasmid.

Materials and Methods

Strain and Media Composition

The *N. oceanica* IMET1 was kindly provided by Prof. Jian Xu from QIBET, Chinese Academy of Sciences. The strains were cultured in artificial seawater, as previously reported [227], and supplemented with commercially available “nutribloom plus” (ASW-NB) provided by Necton (Olhão, Portugal). The liquid microalgae were grown with 5% CO₂ and 3 g/L Sodium Bicarbonate as the buffer to maintain the pH at 7.9 - 8.0. For maintenance, recovery after transformation, and selection on antibiotics, the cultures were grown in light incubators at 25 °C with a constant light intensity of 50 μmol m⁻² s⁻¹. For growing cells without CO₂, 4.77 g/L HEPES was added to the media and pH was set to 8. For nitrate reductase knockouts, a final concentration of 12 mM ammonium chloride was added to

the ASW-NB media. Antibiotics zeocin (10 µg/ml) or blasticidin (100 µg/ml) were used to select and cultivate transformants.

Plasmid construction

All the plasmids in the study were assembled using Gibson assembly (2x HiFi DNA assembly master mix from New England Biolabs, NEB) with the exception of the multiplexing plasmid that was cloned by CRATES assembly (Supplementary File) [244]. The details on plasmid construction and the GenBank files of plasmids used in the study are provided in the supplementary file. The Cas12a sequences obtained from plasmids pY004 (pcDNA3.1-hFnCpf1) and pY016 (pcDNA3.1-hLbCpf1) were developed in the group of Feng Zhang (Addgene plasmid # 69976 and #69988) [102]. The Cas9 sequence was a gift from Eva Farre [79]. The fragments required for the cloning procedures were amplified by PCR using Q5 High Fidelity 2x Master Mix (NEB), and the primers for PCR were obtained from Integrated DNA Technologies and their sequences are tabulated in Table S1.

Transformation and selection

A previously described electroporation protocol [63] was adapted for transforming the *N. oceanica* IMET1 strains. The microalgal cells were inoculated to an OD₇₅₀ of 0.08 and grown under medium light conditions (50 µmol m⁻² s⁻¹) for 2 days. For each transformation, 50 mL of culture was harvested at mid-exponential growth phase (OD₇₅₀ of 0.5) by centrifugation (2000 X g, 10 min at 4 °C). Cell pellets were washed twice and resuspended in 0.2 mL ice-cold 375 mM D-Sorbitol to a final concentration of 5 x 10⁸ cells/mL and transferred to a pre-cooled 2 mm electroporation cuvette. 3 µg of plasmid and 30 µg of salmon sperm DNA were added to the resuspended cells and mixed by gently flicking the cuvettes. The transformation was performed using a Bio-Rad GenePulser II with 11 kV cm⁻¹ field strength, 50 µF capacitance, and 600 Ω shunt. After transformation, the cells were quickly transferred to 5 mL of fresh media supplemented with NB and ammonia for 48 hours recovery under medium light conditions. Subsequently, the cells were pelleted and plated on solid media with NB, ammonia, and respective antibiotics. Colonies were observed between 3-6 weeks after the plating.

For the RNP based transformations, the Cas proteins were desalted at 4 °C to remove the excess salt present in the Cas protein storage buffer. The desalting was performed by multiple rounds of concentration (Amicon Ultra -0.5mL, Ultracel - 30K) and resuspension of the Cas protein in 10mM HEPES and 12.5% glycerol solution. Subsequently, the RNPs were assembled by mixing equimolar amount (10 µM) of Cas protein and the crRNA (obtained from IDT) at 37 °C for 15 minutes in an eppendorf tube. From the assembled RNP complex, 30 µL of sample was used for each transformation.

N. oceanica colony PCR

N. oceanica colonies were resuspended in 20 µL dilution buffer (Plant Direct PCR Master Mix, Thermo Scientific). PCR was performed using 2X Q5 PCR Master Mix (NEB) with 2 µL of the microalgal suspension per reaction.

Fluorescence-activated cell sorting of transformants

After transformation, *NR* mutants expressing tdTomato were recovered for 2 days in 5 ml ASW-NB and ammonia under medium light conditions. Cells were transferred to 50 mL of media and incubated for 3 more days. Subsequently, non-tdTomato-expressing cells were sorted using a Sony SH800S cell sorter (Sony Biotechnology Inc.). Cultures were diluted to an OD_{750} of 0.2, analyzed at sample pressure 4 and singlet events with suitable chlorophyll a content were selected as described previously [261]. The tdTomato fluorescence of events falling into the “Alive” gate was analyzed in the FL2-A channel at 585 ± 15 nm using a detector gain of 45% (Figure S1). Cells with an FL2-A signal of 10-250 (gate “LowRed”) were sorted into ASW media, plated on solid media and grown for 2 weeks.

Fluorescence and Luminescence assay

The Nlux activity was screened by resuspending 10 μ L of a *N. oceanica* colony in 100 μ L ASW media in a 96-wells F-bottom white Lumitrac plate (Greiner). The Nano-Glo substrate (Promega) was diluted 10,000 times in ASW media and 100 μ L was added to the resuspended culture. Luminescence was measured with following specifications on a BioTek Synergy Mx plate reader; delay: 100 ms, read height: 1 mm, Gain: 135, integration time: 30 ms.

The fluorescence measurement was performed using 150 μ L of exponentially grown cultures in a 96-well F-bottom dark Lumitrac plate (Greiner). The tdTomato fluorescence was measured with the following settings: Endpoint, Excitation: 555/9.0, Emission: 605/20.0, Gain: 150, read height: 8 mm. The fluorescence and luminescence measurements were obtained using the plate reader Synergy Mx from Biotek.

RNA isolation

RNA isolation was performed by Trizol-Phenol-Chloroform extraction. Microalgal cultures were grown to exponential phase and 25 ml of culture was harvested by centrifugation (2000 x g, 15 min, 4 °C). The supernatants were removed immediately, and pellets were resuspended in 0.5-1 mL Trizol (TRI Reagent from Sigma Life Science). Samples were vortexed and incubated at room temperature for 5 min. 200 μ L of chloroform was added and the samples were mixed by inverting tubes 5-10 times. Again, the samples were incubated for 5 min at room temperature and centrifuged at 12000 x g for 10 min. The upper phase was collected in a new tube and 20 μ L of 3 M sodium acetate and an equal volume of ice-cold isopropanol was added. The samples were vortexed for 10 seconds and centrifuged at maximum speed for 10 min to precipitate the DNA. The top phase was gently removed and centrifuged again for 2 min to remove all the liquid phase. The pellet obtained was incubated for another 3 min at room temperature. Subsequently, the pellet was resuspended in 5 μ L DNase I (NEB), 10 μ L DNase I reaction Buffer (NEB) and 85 μ L milliQ water. The samples were incubated at 37 °C for 15 min. 10 μ L of 3 M sodium acetate was added to the samples and 1 ml of -20 °C ethanol to precipitate the RNA. The mixture was vortexed and centrifuged for 5 min at maximum speed. Washing was repeated with 300 μ L of 70% ethanol. After completely removing the supernatant and drying the tube at room temperature, the pellets were resuspended in 40 μ L RNase free milliQ water. The samples were confirmed to be free of DNA contamination by performing PCR specific to

the housekeeping gene *elongation factor 1* (EEF1A2) (*N. oceanica* IMET1_NO22G01440). DNase treatment and RNA precipitation were repeated when necessary.

Transcript quantification by RT-qPCR

cDNA was generated from isolated RNA using SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions, using 500 ng of RNA as template and gene specific primers complementary towards the Td tomato and EEF1A2 CDS (Td Tomato: BG23326/BG23327, BG23328/BG23329 and BG23330/BG23331; EEF1A2: BG23365/BG23366, BG23367/BG23368). qPCR was performed using iQ SYBR Green Supermix (Biorad) in a Bio-Rad CFX96 real time PCR system. 1 μ l of undiluted cDNA was used as template for qPCR. Three sets of primers were used, amplifying different regions of approximately 250 nucleotides length of the tdTomato gene. For the control, two sets of primers were used, amplifying different regions of approximately 250 nucleotides length from the housekeeping gene *elongation factor 1*. The relative repression in the transcript levels was analyzed by the $-\Delta\Delta C_t$ method.

Acknowledgments

The project is funded by the project 'Microalgae as a Green source from Nutritional Ingredients for Food/Feed and Ingredients for Cosmetics by Cost-Effective New Technologies' (MAGNIFICENT), funded by the Bio-based Industries Joint Technology Initiative under the EU Horizon 2020 Research and Innovation Program (project 745754)

Supplementary materials

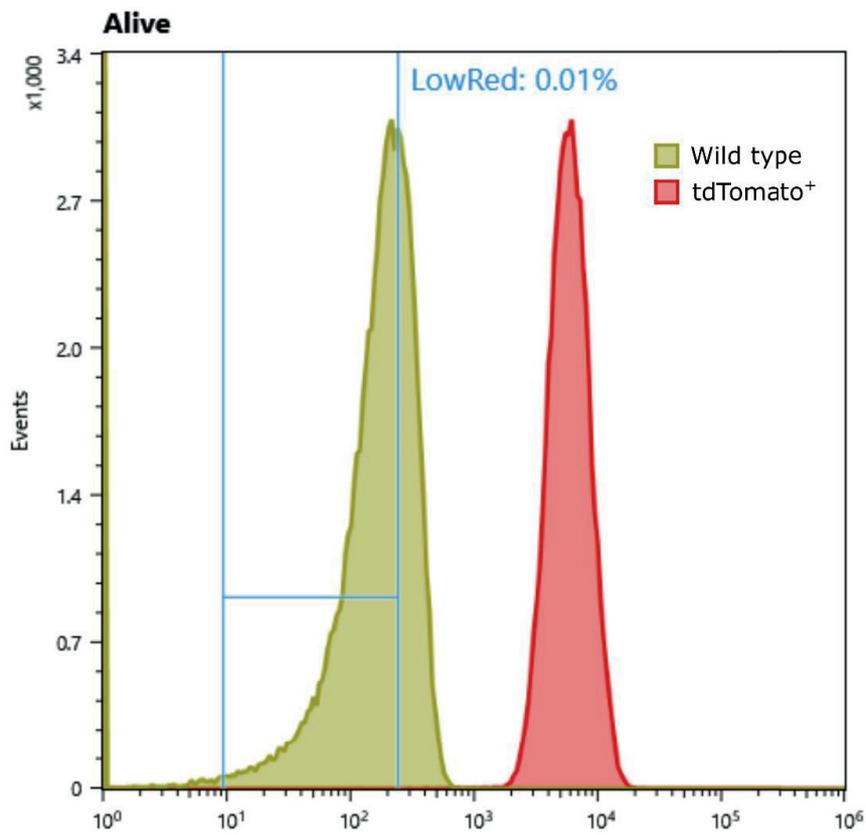


Figure S1. Sorting of scarless tdTomato mutants after transformation with RNPs and editing template. The Histograms show the distribution of a wild type culture and of the tdTomato-expressing (tdTomato+) parental strain. The two blue vertical lines delimit the sort gate ("LowRed"), i.e. the range of cells that were considered as non-tdTomato-expressing mutant cells after transformation. About 0.01% of total events in the parental gate ("Alive") fell into the sort gate.

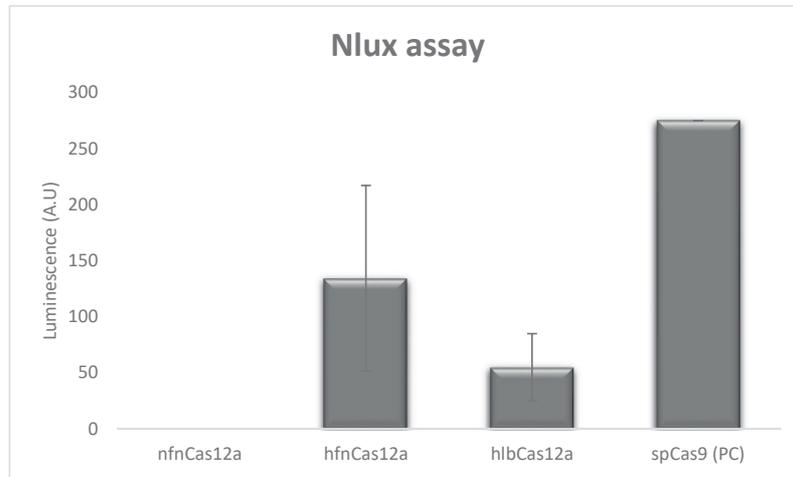


Figure S2. Nlux assay of colonies harboring different Cas12a variants. 10 colonies from each transformation were resuspended in 150 μ L of ASW media and diluted with equal amount of ASW containing 10,000X diluted Nano-Glo substrate. SpCas9 was used as a positive control (PC) as this variant was reported to be efficiently expressed in *N. oceanica*. The samples were quickly measured for luminescence on a plate reader with following specifications; Gain : 155, Read height : 1mm, Delay : 100msec. The standard deviation is from n=10 for each Cas variants used in the study.

Table S1. Codon usage table of *Nannochloropsis gaditana* CCMP52 used to generate the codon harmonized fnCas12a gene for *N. oceanica* IMET1

TGG	12413	CAG	20520	ATG	21767	CCG	13524
AGC	15334	GCC	34612	CCC	20283	TAT	6926
TAC	15727	CTC	23401	TAA	608	TAG	629
CGA	9316	CAA	14237	GAT	15956	TCG	13109
AAC	14772	AAA	16545	TCA	6576	TTA	4523
TTT	13495	TGT	5040	AGT	6563	TTG	22319
GGT	12017	GAC	29726	CTA	5054	GTG	29315
TGA	989	ACT	7120	ACA	8889	AGA	5481
ATA	4742	GGC	25428	TGC	9442	CTG	25622
CGT	8450	CAC	13672	AAT	10168	GCA	15312
GGG	25734	TTC	20686	GGA	17596	ATC	18976
GTC	21706	CCT	14275	ACG	16519	ATT	10121
CAT	8884	AAG	26487	CTT	13243	GAG	43562
CCA	8280	GCT	15096	GAA	22681	TCC	20913
ACC	16392	CGG	16607	AGG	9855	GCG	26155
TCT	10336	CGC	15759	GTA	6583	GTT	7994

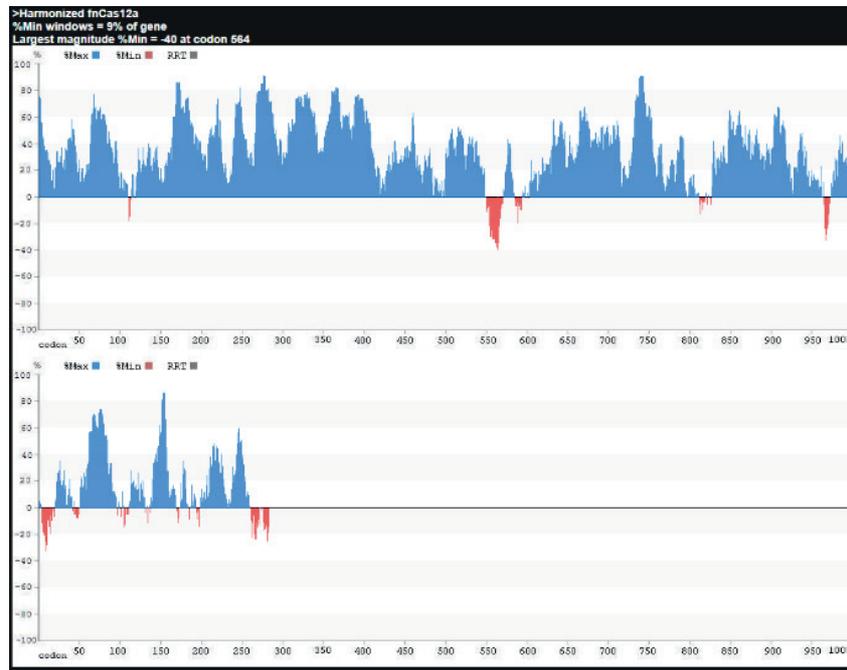


Figure S3. Rare codon calculator image for *nfnCas12a* with respect to the codon frequency table of *Nannochloropsis gaditana* (Table S1).

4

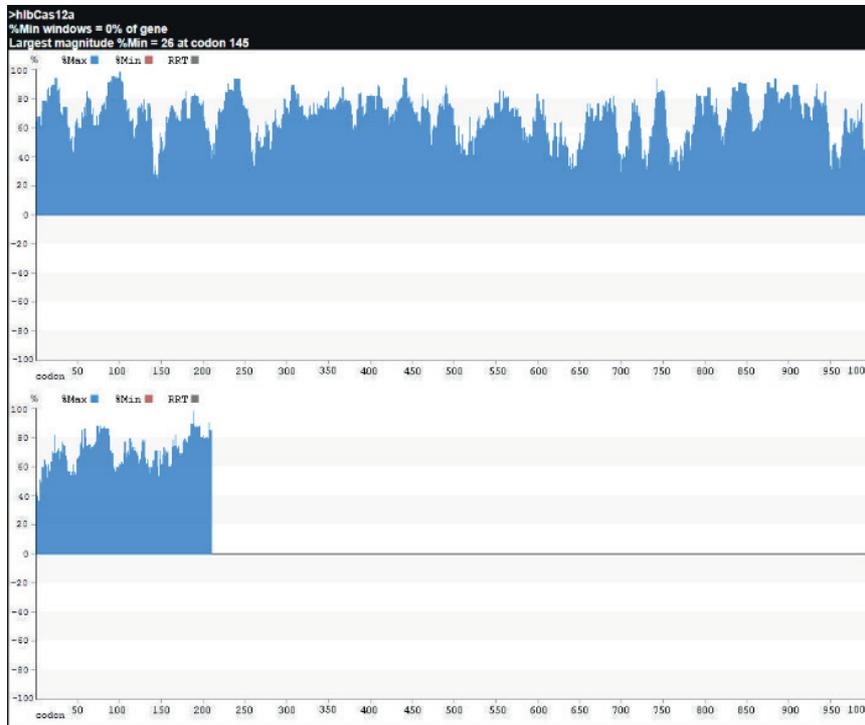


Figure S4. Rare codon calculator image for *hnbCas12a* with respect to the codon frequency table of *Nannochloropsis gaditana* provided above.

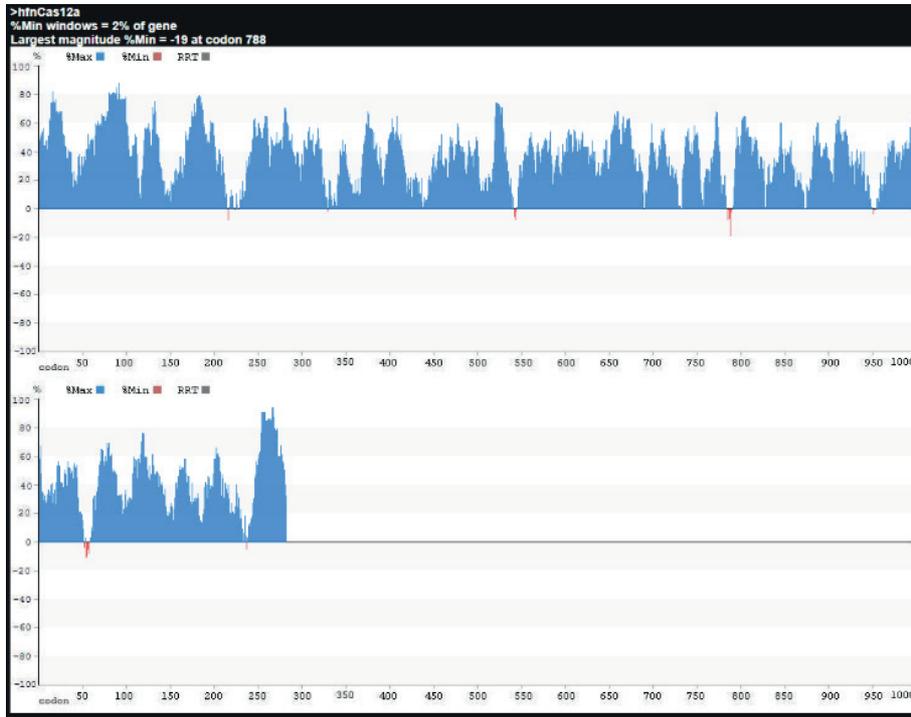


Figure S5. Rare codon calculator image for *hfnCas12a* with respect to the codon frequency table of *Nannochloropsis gaditana* provided above.

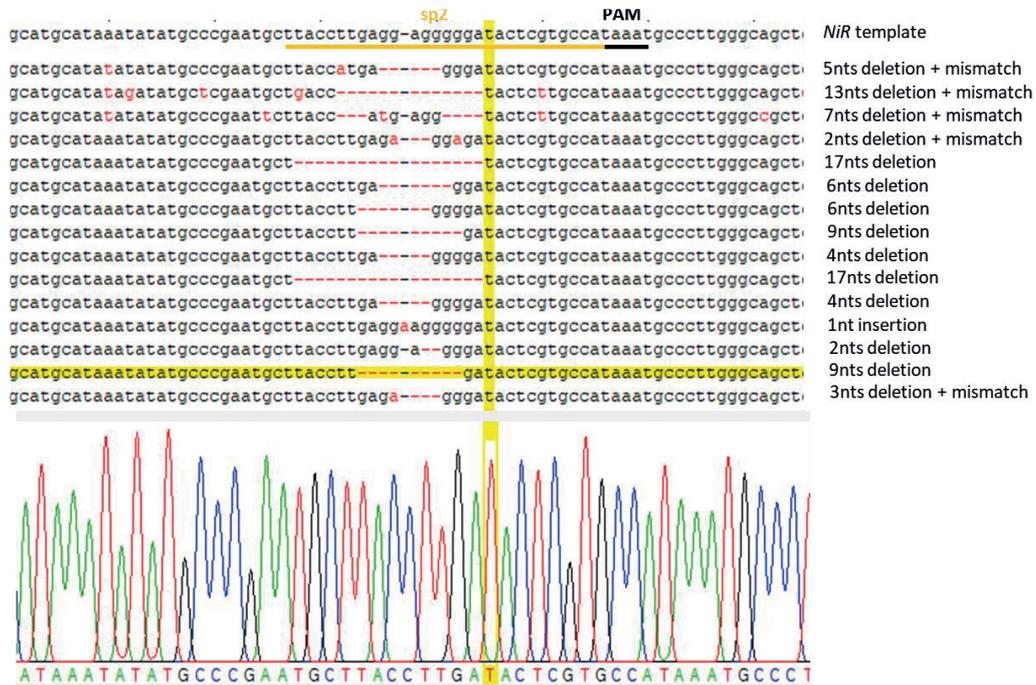


Figure S7. Sequence Alignment. Various indels observed upon targeting the spacer 1 using pNOC_hfnCas12a-Nlux_crRNA NR2 (Primers used for check PCR and sequencing BG23522:BG23523)

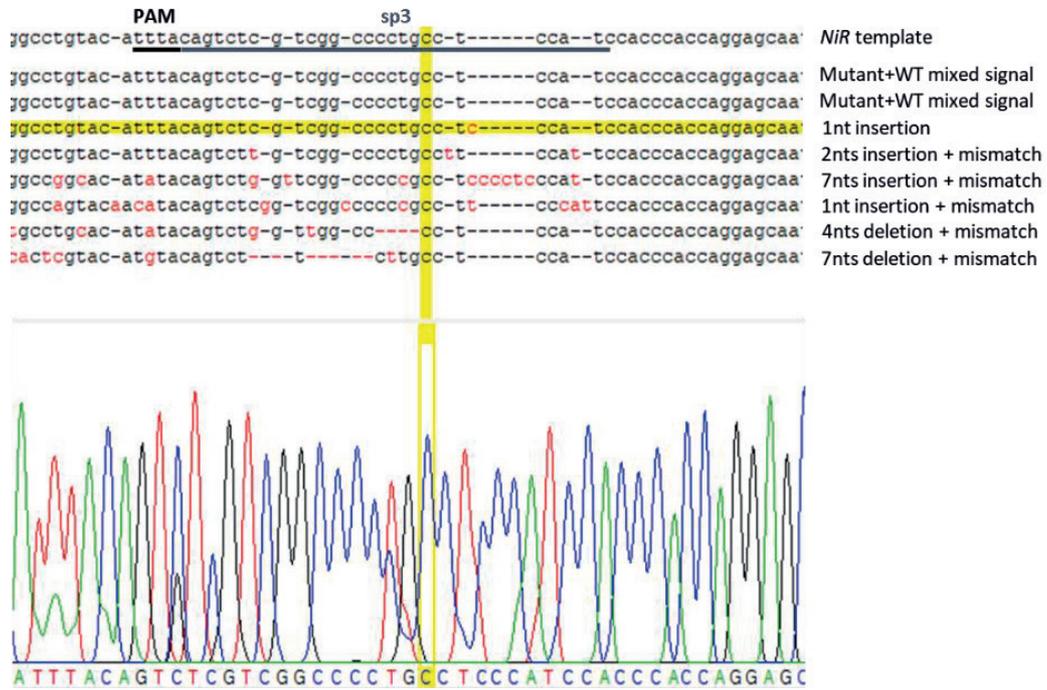


Figure S8. Sequence Alignment. Various indels observed upon targeting the spacer 1 using pNOC_hfnCas12a-NLux_crRNA NR3 (Primers used for check PCR and sequencing)

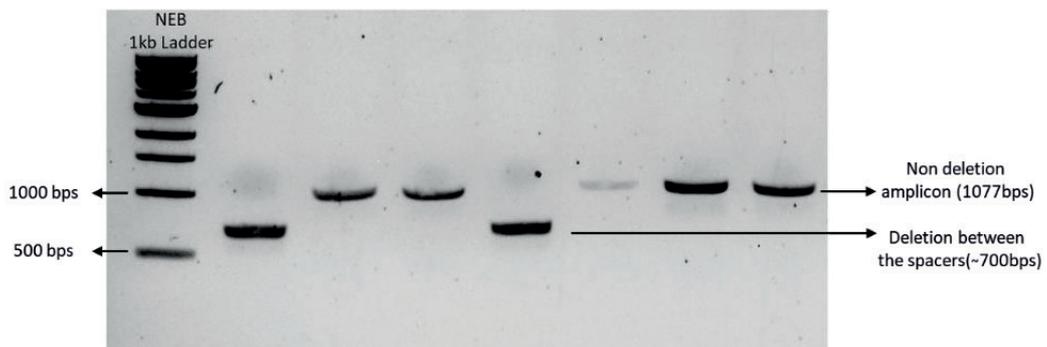


Figure S9. PCR amplicons of colonies edited by Multiplexing (pNOC_hfnCas12a_NLux_NR1_NR2) at 2 different sites ~375bps apart on the *NR* gene. The PCR was performed using primer sets BG21587 and BG21588. Mutations in the *LPAT1* were checked by performing cPCR of colonies and sequencing using primer sets BG24229:BG24230.

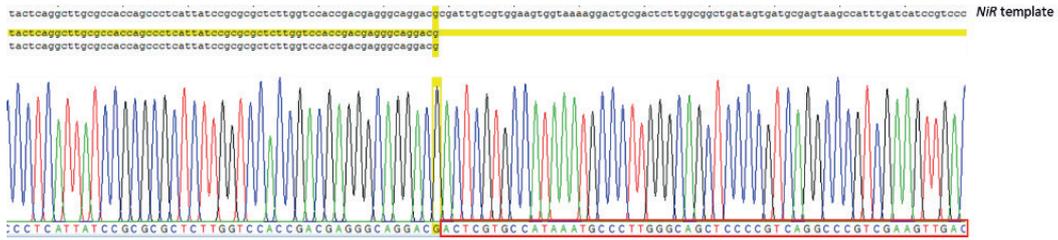


Figure S10. Sequence alignment of the amplicon with reduced length from figure S9. The non-aligned part corresponds to the region after the targeting site by spacer 2.



Figure S11. Inverted sequence of the *NR* gene: The highlighted regions show the target sites of multiplexing. The red font shows the region deleted by multiplexing as confirmed by the sequencing

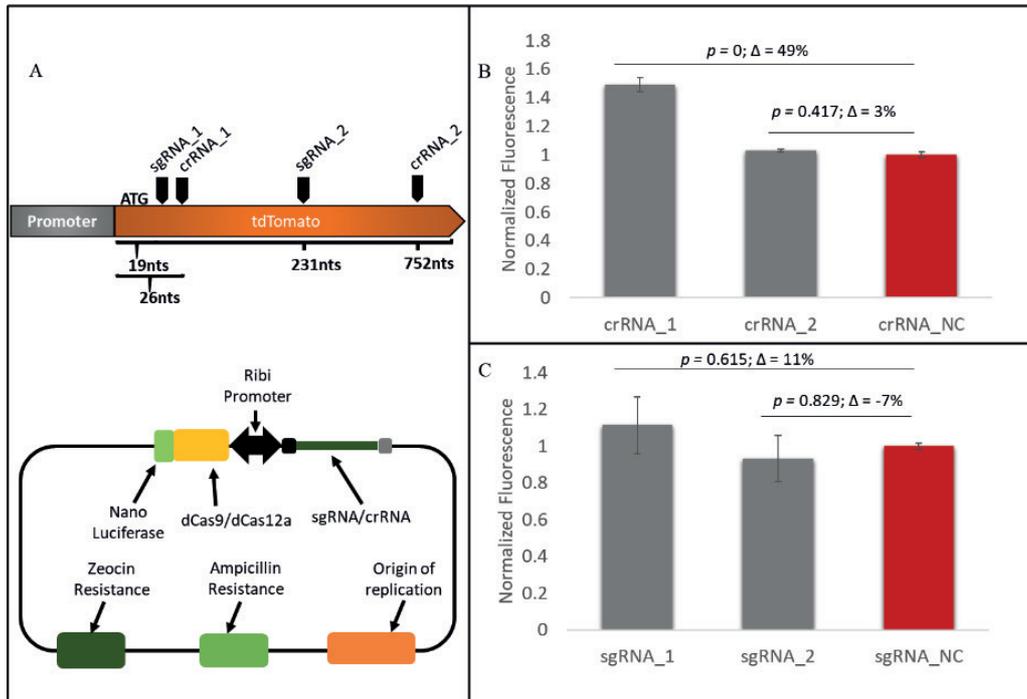


Figure S12. CRISPRi with dCas proteins in absence of KRAB. A) The map of dCas plasmids used for CRISPRi in *N. oceanica* with a schematic representation of various targeting loci in the *tdtomato* gene. B) Fluorescence levels of various dCas12a plasmids targeting the *tdTomato* gene. C) Fluorescence levels of various dCas9 plasmids targeting the *tdTomato* gene. Fluorescence is normalized to OD and to the fluorescence value of NC. The error bars indicate standard deviation based on ≥ 2 biological replicates.

Supplementary Table S2. Primers used in the study

Primer Code	Sequence 5'-3'
BG19114	GTAAACACCATGTTAACGGACCCGCTGCCGGACCCGTTATTTCTATTCTGGACAAACTC
BG19115	CATGCCCAAGAAAAAGCGGAAGGTGAGCATCTACCAGGAGTTC
BG19110	CACCTTCGGCTTTTTCTTGG
BG19111	GTCCGTTAACATGGTGTACTC
BG18129	ATGGCCTCTAGTGATCAG
BG18171	TCTTTCCAATTGTCAGACCTTC
BG15616	ATGGTGTTACTCTCGAGG
BG19114	GTAAACACCATGTTAACGGACCCGCTGCCGGACCCGTTATTTCTATTCTGGACAAACTC
BG19324	AACTTCGGCGACAAGG
BG19325	CCTTGTCGCCGAAGTTCTTG
BG19326	TGAGCATTGACCGAGGAGAG
BG19327	GGGTATAGTAGGCCAGATGC
BG19328	GAGGACATCTGCGGATTAG
BG19329	TAATCCGCAGGATGTCCTC
BG19330	CACAGCAGATCGCTCCAAAG
BG19331	GAGCGATCTGCTGTGAATG
BG19332	CCTCCATCATCTACCGAATC
BG19333	CCAGGAACTTTGGCAGATTG
BG19334	AGTCAACCACTCTCCTAATC
BG19335	CGGTGTCATGGCTTTCTTC
BG18169	TCGTGATGGTTGTTTGAATG
BG18170	TTCGTTCCACCACATGCCC
BG18173	AAAATTACGCTTAAAGCCGG
BG18174	CATCCTTTCTTTCCATGG
BG19112	GAGAGTAAACACCATGTTAACGGACCCGCTGCCGGACCCGCTTACGCTGGTCTG
BG19113	CATGCCCAAGAAAAAGCGGAAGGTGAGCAAGCTGGAGAAGTTACAAACTG
BG19336	CTCGCACAGACGCCGATTAC
BG19337	GTAATCGGCGTCTGTGCGAG
BG19338	TGTGATCGGCATCGATAGGG
BG19339	ATCGATGCCGATCACATAGG
BG19340	ATCGCCCTTCTGAATGTGC
BG19341	TGGATGGCCTACTATAACCC
BG19342	ACGCCGATCTGTCTGTGGTG
BG19343	CATCCACCTTCTGGATGATG
BG19344	GGACGCCATCTTTGATAAGC
BG19345	CAAAGATGGCGTCCACCTTC
BG18169	TCGTGATGGTTGTTTGAATG
BG18170	TTCGTTCCACCACATGCCC
BG18173	AAAATTACGCTTAAAGCCGG
BG18174	CATCCTTTCTTTCCATGG
BG19800	TTAGACCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGCTAAGAACTTTAAATAATTC- TACTGTTGTAGATCCACTTCCACGACAATCGCGTCTG
BG19801	CAGGACGCGATTGTCTGTGGAAGTGGATCTACAACAGTAGAAATTATTTAAAGTTCTTAGACGAC- GAGCTTACTCGTTTCGCTCCTACGGACTCATCAGGTCTAA
BG19802	TTAGACCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGCTAAGAACTTTAAATAATTC- TACTGTTGTAGATTGGCACGAGTATCCCCCTCTCAAG

Primer Code	Sequence 5'-3'
BG19803	CTTGAGGAGGGGATACTCGTGCCAATCTACAACAGTAGAAATTATTTAAAGTTCTTAGACGAC-GAGCTTACTCGTTTCGTCCTCACGGACTCATCAGGTCTAA
BG19804	TTAGACCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGCTAAGAACTTTAAATAATTTCTACTGTTGTAGATCAGTCTCGTCGGCCCTGCCTCCAT
BG19805	ATGGAGGCAGGGGCGACGAGACTGATCTACAACAGTAGAAATTATTTAAAGTTCTTAGACGAC-GAGCTTACTCGTTTCGTCCTCACGGACTCATCAGGTCTAA
BG19806	GCTCGAGGGTTGCGTGTGTATC
BG19807	GTCCTCACGGACTCATCAGGTCTAAGAAGAGCCAAGGCCTGCTCTTC
BG19808	CCACTTCCACGACAATCGCGTCTGGGCCGGCATGGTCCC
BG19809	AAACAGAAGAGGTGATATCGTCTCCTGCAGGCATGCAAGCTGATCC
BG19810	GCAGGAGACGATATCACCTCTTC
BG19811	ACAGATACACACGCAACCCTCGAGCATGCCAAGAAAAAGCGGAAG
BG20187	TGGCACGAGTATCCCCCTCTCAAGGGCCGGCATGGTCCC
BG20188	CAGTCTCGTCGGCCCTGCCTCCATGGCCGGCATGGTCCC
BG20508	GCACGTGGTGAACGTCATCTGGCATACT
BG20509	TGCAAAAACCTACAATTGCATTATACTCAATAACTAATTTGCAATTTT
BG20510	GTATAATGCAATTGTAGTTTTTGACAGATCTGAATTTGGGTTTAAACGTGGGAGATTC
BG20511	CAATCAGCTGTTGCCCGTCTC
BG20512	CAGTGAGACGGGCAACAGCTGATTGCCCTTAC
BG20513	AGTATGCCAGATGACGTTACCACGTCGAATGCTCAGAATATGAACATCATTAGCTTTTTT
BG20547	CGTAGTCGGGCACGTCGTAGGGGTAGTTATTTCTATTCTGGACAACTCGAAGTATTC
BG20548	CCATCCAACCCACATACAC
BG20549	TTTCTGTGTATGTGGGTTGGGATGGGTGGCTGTTTG
BG20550	TACCCCTACGACGTGCCCGACTAC
BG21575	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAG
BG21576	GTGGGAGTACGAGTGTG
BG21577	ACACACAACACTCGTACTCCACTTTC
BG21587	AACCCTGCCCGTCACTCTTG
BG21588	GGGTACTCCTCACTGCTCAG
BG21578	TAGAAATTATTTAAAGTTCTTAGACGAAGAGCCAAGGCCTGCTCTTC
BG21589	TTAGACCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGCTAAGAACTTTAAATAATTTCTACTGTTGTAGATGGATCCCATGACGCTAGTATCCAGCTGG
BG21590	CCAGCTGGATACTAGCGTCATGGGATCCATCTACAACAGTAGAAATTATTTAAAGTTCTTAGACGAC-GAGCTTACTCGTTTCGTCCTCACGGACTCATCAGGTCTAA
BG21591	TCCCATGACGCTAGTATCCAGCTGGGGCCGGCATGGTCCC
BG22022	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATCCACTTCCACGACAATCGCGTCTGTGCTG
BG22023	CAGGACGCGATTGTCGTGGAAGTGGATCTACAACAGTAGAAATTATTTAAAGTTCTTAGACGAAGA
BG22024	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATTGGCAGGATATCCCCCTCTCAAGGTCTA
BG22025	CTTGAGGAGGGGATACTCGTGCCAATCTACAACAGTAGAAATTATTTAAAGTTCTTAGACCAGCA
BG23245	CTTTGTCGCTCGACGTGTCCAAGGGGTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATGGTACCAT-GGGAAAGAAAGGATGAG
BG23326	GCGCATGAACTCTTTGATGACGG
BG23327	GCCACTACCTGGTGGAGTTCAAG
BG23328	GAAGCGCATGAACTCTTTGATGACGG
BG23329	CACTACCTGGTGGAGTTCAAGACC
BG23330	GATGACGGCCATGTTGTTGTCC
BG23331	GGCCACTACCTGGTGGAGTTC
BG23365	CACGTGAACCTGGTCTGATTGG

Primer Code	Sequence 5'-3'
BG23366	GTAGTACTTGGGAGACTCGAACTCC
BG23367	GGTACCTCCCAGGCTGATGTTG
BG23368	CGATCATGTTGTCGCCATTCCAG
BG23246	CCCTTGACACGTCGAGCGACAAAGATCTACAACAGTAGAAATTATTTAAAGTTCTTAGACCTTGAG- GAGGGGGATACTCGTG
BG23247	GGAGAATGAGTACCGCGAGGAAGTAGTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATGGTACCAT- GGGAAAGAAAGGATGAG
BG23248	TACTTCCTCGCGGTACTCATTCTCCATCTACAACAGTAGAAATTATTTAAAGTTCTTAGACCTTGAGGAG- GGGGATACTCGTG
BG23522	GTGCTTGATGCGGACGACAG
BG23523	AAAGCGCACGACGCAATGG
BG24229	CGAAACGCTCTTGCGGATTG
BG24230	GTAATCATGCAGTCGCCCACTTAC

Table S3. Details of guide RNAs used in the study

Guide RNAs used in the study	Sequence 5'-3'	Comments
NR sp (RNP)	AAUUUCUACUGUUGUAGAUUGGCACGAGUUAUCCCCCUCCUCAA	The RNA sequences used for RNP studies. The red font indicate the spacer sequences
Td Tomato sp1 (RNP)	AAUUUCUACUGUUGUAGAUUGGCACGAGUUAUCCCCCUCCUCAA	
Td Tomato sp2 (RNP)	AAUUUCUACUGUUGUAGAUUGGCACGAGUUAUCCCCCUCCUCAA	
crRNA NR sp1	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATCCACTTCCACGACAATCGGGTCTCTG	The spacer sequences used for targeting the Nitrate reductase gene. The red font indicate the 36nt repeat sequence used in the crRNA
crRNA NR sp2	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATGGACGAGTATCCCCCTCCTCAAG	
crRNA NR sp3	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATCAGTCTCGTGGCCCTGCCTCCAT	
LPT sp (Multiplexing)	GGAGATGAGTACCGCGGAAGTA	The spacer sequence for targeting the LPAT1 gene for multiplexing
Multiplexing CRISPR array NR1_NR2	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATCCACTTCCACGACAATCGGGTCTCTGTTGTAGATTTAAATAATTTCTACTGTTGTAGAT	The red font indicate the 36nt repeat and the black font indicate the spacer sequences sp1 and sp2 respectively
Multiplexing CRISPR array NR1_NR2_LPT1	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATCCACTTCCACGACAATCGGGTCTCTGTTGTAGATTTAAATAATTTCTACTGTTGTAGATGGAGATGATCCCGGAGAGTAGCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	
CRISPRi Td Tomato sgRNA sp1	GCGCATGAACCTTTGATGATGTTTTAGAGCTAGAAATAGCAAGTTAAATAAGGCAGTCCCGTTATCAACTTTGAAAAAGTGGCACCGATCGGTGCTTTT	The guide RNA sequence targeting the Td Tomato for CRISPRi using dCas9. The red font indicate the sgRNA scaffold and black font is the spacer sequence
CRISPRi Td Tomato sgRNA sp2	GCGGTGCTTACGTACGCTGTTTTAGAGCTAGAAATAGCAAGTTAAATAAGGCAGTCCCGTTATCAACTTTGAAAAAGTGGCACCGATCGGTGCTTTT	
CRISPRi Td Tomato crRNA sp1	AATTTCTACTGTTGTAGATGACCTCCTCGCCCTTGCTCACCA	The crRNA for targeting Td Tomato for CRISPRi using dCas12a. Red font is the 19nt repeat followed by the spacer sequences
CRISPRi Td Tomato crRNA sp2	AATTTCTACTGTTGTAGATGACGCCCATGTTGTTGCTCTCG	

Table S4. Cas12a genes sequences used in the study

<i>fnCas12a</i> codon harmonized for <i>Nannochlropsis</i>	<p>ATGTCCATCTACCAGGAGTTCGTGAACAAGTACAGCTTGAGCAAGACGCTTCGGTTCGAAT- TGATACCCCAAGGGAAGACCCTTGAGAATATTAAGGCACGGGGTAACTTTGGACGAC- GAAAAGCGGCCAAGGATTATAAGAAAGCCAAGCAGATTATCGACAAGTACCACCAATTCTT- CATTGAAGAAATTTGAGCTCAGTGTGCATCTCAGAGGACTTGTTCAGAAATTAACCCGACGTG- TACTTCAAGCTTAAGAAAAGCGACGACGACAACCTTCAGAAGGACTTCAAGAGCGCAAAGGA- CACTATTA AAAAGCAGATTTCCGAGTACATTAAGATTCCGAAAAGTTCAAAAACCTTATTAACCA- GAATCTTATAGACGCCAAGAAGGGTCAGGAATCCGACTTGATCCTTTGGCTTAAACAGTCCAAAG- ACAACGGGATTGAGCTTTTCAAGGCTAACAGCGACATAACCGACATTGACGAAGCTTTGGAGAT- TATAAAGTCTTCAAGGGGTGGACCACGTA CTCAAAGGGTCCACGAGAACCGGAAGAACGT- GTACAGCTCAAACGACATCCCTACCTCCATCATCTACAGAATTGTGCGACGACA ACTTACCTAAGT- TCCTTGAGAACAAGGCCAAATACGAAAGCTTGAAGGATAAGGCCCCGAGGCCATTAATAC- GAGCAGATCAAGAAGGACTTAGCAGAGGAACTTACTTTGACATCGATTATAAGACCTCCGAG- GTGAACAGCGGGTGTCTCCCTTGACGAGGTGTTGCAAATTCGAAATTTCAACAAC TACCT- TAACCAGAGCGGGATCACGAAGTTCAACACGATCATCGGGGGGAAGTTCTGTAACGGGGGA- GAACACCAAACGGAAGGGGATTAACGAGTACATTAACCTTTATTCCAACAGATTAACGACAAG- ACCTAAAGAAGTACAAGATGAGCGTGTGTTCAAACAGATCTTGAGCGACACCGAGTCCAAGT- CCTTCGTCATCGACAAATTGGAGGACGACAGCGACGTCGTGACCACTATGCGAGAGTTCTAC- GAACAGATTGCAGCCTTCAAGACCGTCGAGGAGAAGTCCATCAAGGAGACCTTTCTTTGTT- GTTGACGACTTGAAGGCCAGAAGCTTGACTTAAGCAAGATCTACTTCAAGAACGACAAGTC- CCTTACGGACCTTTCCAGCAGGTGTTGACGACTACAGCGTGATCGGGACCGCTGTCCTTGAG- TACTTACGACGAGATTGCACCTAAGAACCTTGACAATCTTAGCAAAAAGGAACAGGAGTT- GATTGCTAAGAAGACGGAGAAGGCAAAGTATTTGTCCTTGAGACGATTAACCTTGCTTTGGAG- GAGTTCAACAAACACCGGGACATTGACAAGCAATGCAGATTCGAGGAGATTCTTGCAAATTT- CGTGCCATCCCAATGATTTTCGACGAGATTGCCAGATAAAGGATAACTTAGCACAAATTTCCA- TAAAGTACCAGAACCAGGGGAAGAAGGATCTTCTCAGGCCAGCGTGAGGACGACGTGAAGG- CCATAAAAGACCTTTTGACCAGACGAAACAACCTATTGCACAAGCTTAAGATTTTCCACATCAGC- CAATCCGAGGACAAAGCAAACATCTTGGATAAAGACGAACTTCTACCTTGCTTGAAGAAT- GTTATTTGAACTTGCTAACATTGTACCTTTTACAATAAGATCCGGAATTACATTACGAGAAAC- CCTACAGCGACGAAAAGTTCAAATAACTTCAAAAATTCACGTTAGCCAACGGGTGGGA- CAAGAACAAGGAACCTGATAACACTGCAATCTTGTTTATAAAGGACGACAAGTACTACCTAGG- GGTAATGAACAAAAAGAACATAAAGATTTTCGACGACAAGGCCATAAAGGAGAACAAAGGGT- GAAGGGTACAAGAAGATCGTGTAACAAGCTTTTGCCTGGTGCAAACAAGATGTTGCCTAAAGT- GTTTTTCTCCGCAAGTCCATTAAGTTCTACAACCCTAGCGAGGACATTCTTCGGATTTCGGAAC- CACTCAACCCACACCAAGAACGGGAGCCCTCAGAAGGGTTACGAGAAGTTTCAATTAACATC- GAGGACTGTAGAAAGTTCATTGACTTACAAGCAATCCATTAGCAAACACCCAGAATGGAAGG- ACTTCGGTTTCCGGTTCCTCGACACGACGGGTACAACCTCATTGACGAGTTCTACCGGGAGGT- GGAGAACCAGGGTTATAAGCTTACGTTGAGAACATTTCCGAATCATAATCGACTCAGTCGT- GAACCAAGGGAAGTTATATCTTTTTAGATATAACAAGGACTTCTCCGCTACTCAAAGGGTA- GACCAACCTTACACGTTGTACTGGAAGGCTCTATTGACGAAACGGAACCTTCAGGACGTAGT- GTACAAACTTAACGGGGAAGCAGAACTTTTCTACGAAAGCAGTCCATTCTAAGAAGATAACG- CATCCCGCAAGGAAGCAATTGCCAACAAAGAATAAGGACAACCCTAAGAAGGAAAGCGTGTTC- GAGTACGACTTGATAAAGGACAAGAGATTCACGGAGGACAAATCTTTTTCCATTGCCCTATCA- CCATAAATTTCAAGTCCAGCGGTGCCAACAAATTAACGACGAGATAA ACTTGTACTTAAGGA- GAAGGCAAACGACGTGCACATTTGAGCATTGACCGGGGGGAGCGGCCTTGGCCTATTACA- CGTTAGTCGACGGGAAGGGTAACATAATAAAGCAGGACACGTTTAAATAATCGGGAACGACCG- GATGAAGACCAATTATCACGACAAACTTGCCGCAATTGAAAAGGACAGAGACTCCGCCAGAAAG- GATTGGAAGAAAATTAACAAATAAAGGAAATGAAGGAAGTTACCTTTCCCAAGTCGTGCAC- GAGATTGCCAAACTTGTGATTGAATAACAACGCCATCGTAGTGTTCGAAGACTTGA ACTTCGGT- TTCAAGCGGGGTAGATTTAAAGTCGAAAAACAAGTATACCAGAAATTTGGAGAAGATGCTTATC- GAAAAGCTTAATTACCTTGTTTAAAGGACAACGAATTCGACAAGACGGGTGGTACTTCCGGG- CCTACCAACTTACCGACCTTTGAAAACGTTCAAGAAAATGGGGAAGCAGACCGGGATCATATAT- TACGTCCCGCCGGGTTCACGTCCAAGATCTGCCCTGTCACGGGTTTCGTCAACCAATTGTACCC- TAAATACGAGAGCGTATCAAAGTCCAGGAATTTTCAGCAAATTCGATAAAATCTGCTACAATCT- TGACAAAGTTACTTCAATTCAGCTTCGACTACAAGAATTTCCGGGATAAAGCCGCTAAGGG- TAAATGGACGATTGCCTCATTGCGTAGCCGGTTAATCAATTTCCGGAACCTCCGACAAGAACCA- CAACTGGGACACGAGAGAGGTGTACCCACGAAGGAATTAGAAAAGTTACTTAAGGACTACT- CCATAGAGTACGGTACGGTGAGTGACATAAAGGCAGCCATCTGTGGGGAATCAGATAAGAAAT- TCTTCGCCAAACTTACGAGCGTACTTAACACGATATTGCAGATGAGAAATTTCCAAGACCGGG- ACGGAATTTGACTACTTCTCTCCCTTCGACAGCTCAACGCTA ACTTTTTCGACTCAAGA- CAAGTCCCAAGAACATGCCTCAGGACGCCGACGCTAACGGGGCCTACCACATCGGTCTTAAG- GGGCTAATGCTTCTGGGAGAATAAAGAACACCAGGAAGGTAAGAAGCTAAACTTAGTGA- TAAAGAACGAGGAATACTTCAATTTGTACAAAACAGAAACAATTGA</p>
--	---

human
optimized
fnCas12a

ATGAGCATCTACCAGGAGTTCGTCAACAAGTATTCAGTGAAGACTGCGGTTTCGAGCT-
GATCCCACAGGGCAAGACTGGAGAATCAAGGCCGAGGCCTGATTCTGGACGATGA-
GAAGCGGCAAAAGACTATAAGAAAGCCAAGCAGATCATTGATAAATACCACCAGTCTTTATC-
GAGGAAATCTGAGTCCGTGTGCATCAGTGAGGATCTGTGCAGAATTAAGTCTGAGCTGACT-
TCAAGCTGAAGAAGAGCGACGATGACAACCTGCAGAAGGACTTCAAGTCCGCCAAGGACAC-
CATCAAGAAACAGATTAGCGAGTACATCAAGGACTCCGAAAAGTTTAAAAATCTGTTCAACCA-
GAATCTGATCGATGCTAAGAAAGGCCAGGAGTCCGACCTGATCTGTGGCTGAAACAGTCTAAGG-
ACAATGGGATTGAAGTGTCAAGGCTAATCCGATATCACTGATATTGACGAGGCACTGGAAAT-
CATCAAGAGCTTCAAGGGATGGACCACATACTTTAAAGGCTTCCACGAGAACCAGCAAGAACGT-
GTACTCCAGCAACGACATTCCACTCCATCATCTACCGAATCGTCGATGACAATCTGCCAAAGT-
TCCTGGAGAACAAGGCCAAATATGAATCTCTGAAGGACAAAGCTCCCGAGGCAATTAATTAC-
GAACAGATCAAGAAAGATCTGGCTGAGGAACTGACATTGATATCGACTATAAGACTAGCGAG-
GTGAACCCAGAGGGTCTTTTCCCTGGACGAGGTGTTTGAATCGCCAATTTCAACAATTACCT-
GAACCAAGTCCGGCATTACTAAATCAATACCATCTTGGCGGGAAGTTTGTGAACGGGGAGAAT-
ACCAAGCGCAAGGGAATTAACGAATACATCAATCTGTATAGCCAGCAGATCAACGACAAAACTCT-
GAAGAAATACAAGATGTCTGTGCTGTCAAACAGATCCTGAGTGATACCGAGTCCAAGTCTT-
TTGTCATTGATAAAGTGAAGATGACTCAGACGTGGTCACTACCATGACAGAGCTTTTATGAG-
CAGATCGCCGCTTTCAAGACAGTGGAGGAAAAATCTATTAAGGAACTCTGAGTCTGTCTGTT-
GATGACCTGAAAGCCAGAAAGCTGGACCTGAGTAAGATCTACTTCAAAAACGATAAAGTCTG-
ACAGACCTGTACAGCAGGTGTTTGTGACTTATCCGTGATTGGGACCCGCTCTGGAGTACA-
TTACACAGCAGATCGCTCAAAGAACCTGGATAATCCCTTAAGAAAGAGCAGGAACTGATCGC-
TAAGAAAACCGAGAAGGCAAAATATCTGAGTCTGGAAACAATTAAGCTGGCACTGGAGGAGT-
TCAACAAGCACAGGGATATTGACAAAAGTCCGCTTTGAGGAAATCTGGCCAACCTTCGCAG-
CCATCCCATGATTTTTGATGAGATCGCCAGAACAAGACAATCTGGCTCAGATCAGTATTAAG-
TACCAGAACCAGGGCAAGAAAGACCTGTGCAGGCTTACGAGAAAGATGACGTGAAAGCCAT-
CAAGGATCTGTGGACCAGACCAACATCTGCTGCACAAGCTGAAAATCTTCCATATTAGTCACT-
CAGAGGATAAGGCTAATATCTGGATAAAGACGAACACTTCTACCTGGTGTTCGAGGAATGT-
TACTTCGAGCTGGCAAACATTGTCCCTGTATAACAAGATTAGGAATACATCACAGAGAAG-
CCTTACTCTGACGAGAAGTTAAACTGAACTTCGAAAATAGTACCCTGGCCAACGGGTGGGA-
TAAGAACAAGGAGCCTGACAACACAGTATCTGTTCATCAAGGATGACAAGTACTATCTGG-
GAGTGATGAATAAGAAAAACAATAAGATCTTCGATGACAAAGCCATTAAGGAGAACAAGGG-
GAAGGATACAAGAAAATCGTGTATAAGCTGTGCCCCGGCGCAAATAAGATGCTGCCTAAGGT-
GTTCTTACGCGCAAGAGTATCAAATCTACAACCCATCCGAGGACATCTGCGGATTAGAAAT-
CACTCAACACATACTAAGAACGGGAGCCCCAGAAAGGATATGAGAAATTTGAGTTCACATC-
GAGGATTGCAGGAAGTTTATTGACTTCTACAAGCAGAGCATCTCAAACACCCTGAATGGAAG-
GATTTTGGCTTCCGTTTTCCGACACACAGAGATATAACTCTATCGACGAGTCTACCGCAGGT-
GGAAAATCAGGGGTATAAGCTGACTTTTGAACATTTCTGAAAGTTACATCGACAGCGTGGT-
CAATCAGGGAAAGCTGTACTGTTCCAGATCTATAACAAGATTTTTTCAGATACAGCAAGGG-
CAGACCAAACCTGCATACACTGTACTGGAAGCCCTGTTTCGATGAGAGGAATCTGCAGGACGT-
GGTCTATAAAGTGAACGGAGAGGCCGAAGTGTTTTACCAGGAGCAGTCTATCTAAGAAAAT-
CACTACCCAGCTAAGGAGGCCATCGCTAACAAGAACAAGGACAATCTAAGAAAGAGAG-
CGTGTTCGAATCAGATCTGATTAAGGACAAGCGGTTACCGAAGATAAGTTCTTTTCCATTGT-
CCAATCACCATTAACCTCAAGTCAAGCGGCGTAACAAGTTCAACGACGAGATCAATCTGCT-
GCTGAAGGAAAAAGCAACGATGTGCACATCTGAGCATTGACCGAGGAGAGCGGCATCTGG-
CCTACTATACCCTGGTGGATGGCAAAGGGAAATATTAAGCAGGATACATCAACATATTGG-
CAATGACCGGATGAAAACCACTACCAGATAAAGTGGCTGCAATCGAGAAGGATAGAGACT-
CAGCTAGGAAGGACTGGAAGAAAATCAACAACATTAAGGAGATGAAGGAAGGCTATCTGAGC-
CAGGTGGTCCATGAGATTGCAAAGCTGGTCATCGAATAACAATGCCATTGTGGTTCGAGGATCT-
GAACTTCGGCTTTAAGAGGGGGCGCTTTAAGGTGGAAAAACAGGTCTATCAGAAGCTGGA-
GAAAATGCTGATCGAAAAGCTGAATTACCTGGTGTAAAGATAACGAGTTCGACAAGACCGGAG-
GCGTCTGAGAGCCTACCAGCTGACAGCTCCCTTTGAAACTTTCAAGAAAATGGGAAAACAG-
ACAGGCATCATCTACTATGTCCAGCCGATCACTTCCAAGATCTGCCCGTACCCGGCTTTGT-
CAACCAGCTGTACCTAAATATGAGTCAGTGAGCAAGTCCCAGGAATTTTTACGCAAGTTTCA-
TAAGATCTGTTATAATCTGGACAAGGGTACTTCGAGTTTTCTTCGATTACAAGAACTTCGGCG-
ACAAGGCCGTAAGGGGAAATGGACCATTGCTCCTTCGGATCTCGCCTGATCAACTTTTCAAAT-
TCCGATAAAAACCACAATTTGGACACTAGGGAGGTGTACCCAACCAAGGAGCTGGAAAAGCT-
GCTGAAAGACTACTCTATCGAGTATGGACATGGCGAATGCATCAAGGCAGCCATCTGTGGCA-
GAGTGATAAGAAATTTTTCGCAAGCTGACCTCAGTGTGAATACAATCCTGCAGATGCGGAACT-
CAAAGACCGGGACAGAACTGGACTATCTGATTAGCCCCGTGGCTGATGTCAACGGAACTTCT-
TCGACAGCAGACAGGCACCCAAAAATATGCCTCAGGATGCAGACGCCAACGGGGCTTACCACA-
TCGGGCTGAAGGACTGATGTCTGGCCGGATCAAGAACAATCAGGAGGGGAAGAAGCT-
GAACCTGGTCAATTAAGAACGAGGAATACTTCGAGTTTGTCCAGAATAGAAATAACTGA

human
optimized
IbCas12a

ATGAGCAAGCTGGAGAAGTTTACAACTGCTACTCCCTGTCTAAGACCCTGAGGTTCAAGGCCA-
TCCCTGTGGGCAAGACCCAGGAGAATCGACAATAAGCGGCTGCTGGTGGAGGACGAGAAGA-
GAGCCGAGGATTATAAGGGCGTGAAGAAGCTGCTGGATCGCTACTATCTGTCTTTTATCAACG-
ACGTGCTGCACAGCATCAAGCTGAAGAATCTGAACAATTACATCAGCCTGTTCCGGAAGAAAAC-
CAGAACCGAGAAGGAGAATAAGGAGCTGGAGAACCCTGGAGATCAATCTGCGGAAGGAGAT-
CGCCAAGGCCTTCAAGGGCAACGAGGGCTACAAGTCCCTGTTTAAAGAAGGATATCATCGAGA-
CAATCCTGCCAGAGTTCCTGGACGATAAGGACGAGATCGCCCTGGTGAACAGCTTCAATGGCT-
TTACCACAGCCTTACCCTGGCTTCTTTGATAACAGAGAGAATATGTTTTCCGAGGAGGCCAAGAG-
CACATCCATCGCCTTCAAGGTGATCAACGAGAATCTGACCCGCTACATCTTAATATGGACATCT-
TCGAGAAGGTGGACGCCATCTTTGATAAGCACGAGGTGCAGGAGATCAAGGAGAAGATCCT-
GAACAGCGACTATGATGTGGAGGATTTCTTTGAGGGCGAGTCTTTAACTTTGTGCTGACACAG-
GAGGGCATCGACGTGATAACGCCATCATCGGCGGCTTCGTGACCGAGAGCGGCGAGAAGAT-
CAAGGGCCTGAACGAGTACATCAACCTGTATAATCAGAAAACCAAGCAGAAGCTGCCTAAGT-
TTAAGCCACTGTATAAGCAGGTGCTGAGCGATCGGGAGTCTTGAGCTTCTACGGCGAGGGC-
TATACATCCGATGAGGAGGTGCTGGAGGTGTTTAGAAAACACCCTGAACAAGAACAGCGAGATCT-
TCAGTCCATCAAGAAGCTGGAGAAGCTGTTCAAGAATTTGACGAGTACTCTAGCGCCGGCA-
TCTTTGTGAAGAACGGCCCCGCCATCAGCACAAATCTCCAAGGATATCTTCGGCGAGTGGAAAGT-
GATCCGGGACAAGTGGAAATGCCGAGTATGACGATATCCACTGAAGAAGAAGGCCGTGGTGAC-
CGAGAAGTACGAGGACGATCGGAGAAAAGTCTTCAAGAAGATCGGCTCCTTTCTCTGGAG-
CAGCTGCAGGAGTACGCCGACGCCGATCTGTCTGTGGTGGAGAAGCTGAAGGAGATCATCATC-
CAGAAGGTGGATGAGATCTACAAGGTGATGGCTCCTCTGAGAAGCTGTTCCGACGCCGATTTT-
GTGCTGGAGAAGAGCCTGAAGAAGAACGACGCCGTGGTGGCCATCATGAAGGACCTGTGGAT-
TCTGTGAAGAGCTTCGAGAATTACATCAAGGCCTCTTTGGCGAGGGCAAGGAGACAAACAG-
GGACGAGTCTTATGCGGATTTTGTGCTGGCTACGACATCCTGCTGAAGGTGGACACATC-
TACGATGCCATCCGCAATTATGTGACCCAGAAGCCCTACTCTAAGGATAAGTTCAGCTGTATTT-
TCAGAACCTCAGTTCATGGGCGGCTGGGACAAGGATAAGGAGACAGACTATCGGGCCACCA-
TCCTGAGATACGGCTCCAAGTACTATCTGGCCATCATGGATAAGAAGTACGCCAAGTGCCTGCA-
GAAGATCGACAAGGACGATGTGAACGGCAATTACGAGAAGATCAACTATAAGCTGTGCCCGG-
CCCTAATAAGATGCTGCCAAAGGTGTTCTTTTCAAGAAGTGGATGGCCACTATAACCCAGC-
GAGGACATCCAGAAGATCTACAAGAATGGCACATTCAGAAGGGCGATATGTTAACTGAATG-
ACTGTCAAGCTGATCGACTTCTTTAAGGATAGCATCTCCCGTATCCAAAGTGGTCCATGCC-
TACGATTTCAACTTTTCTGAGACAGAGAAGTATAAGGACATCGCCGGCTTTTACAGAGAGGTG-
GAGGAGCAGGGCTATAAGGTGAGCTTCGAGTCTGCCAGCAAGAAGGAGGTGGATAAGCTGGT-
GAGGAGGGCAAGCTGTATATGTTCCAGATCTATAACAAGGACTTTTCCGATAAGTCTCACGGCA-
CACCCAATCTGCACACCATGTACTTCAAGCTGCTGTTTGACGAGAACAATCAGGACAGATCAG-
GCTGAGCGGAGGAGCAGAGCTGTTTATGAGCGCGCCCTCCCTGAAGAAGGAGGAGCTGGTGGT-
GCACCCAGCCAACTCCCCTATCGCCAAAGAATCCAGATAATCCCAAGAAAACCAACCCCTGT-
CCTACGAGTGTATAAGGATAAGAGTCTTCTGAGGACAGTACGAGCTGCACATCCCAATCCG-
CATCAATAAGTGCCCAAGAACATCTTCAAGATCAATACAGAGGTGCGCGTGTCTGAAGCACG-
ACGATAACCCCTATGTGATCGGCATCGATAGGGGCGAGCGCAATCTGCTGTATATCGTGGTGGTGG-
ACGGCAAGGGCAACATCGTGGAGCAGTATCCCTGAACGAGATCATCAACAACCTTCAACGGCAT-
CAGGATCAAGACAGATTACCACTCTCTGCTGGACAAGAAGGAGAAGGAGAGGTTTCGAGGCCG-
CCAGAACTGGACCTCCATCGAGAATATAAGGAGCTGAAGGCCGGCTATATCTCTCAGGTGGT-
CACAAGATCTGCGAGCTGGTGGAGAAGTACGATGCCGTGATCGCCCTGGAGGACCTGAACTCTG-
GCTTTAAGAATAGCCGCGTGAAGGTGGAGAAGCAGGTGATCAGAAGTTCGAGAAGATGTGAT-
CGATAAGCTGAAC TACATGGTGGACAAGAAGTCTAATCCTTGTGCAACAGGCGGCGCCCT-
GAAGGGCTATCAGATCACAATAAGTTCGAGAGCTTTAAGTCCATGTCTACCCAGAACGGCTT-
CATCTTTTACATCCCTGGCTGGCTGACATCCAAGATCGATCCATCTACCGGCTTTGTGAACCTGCT-
GAAAACCAAGTATACCAGCATCGCCGATTCCAAGAAGTTCATCAGCTCCTTTGACAGGATCATG-
TACGTGCCGAGGAGGATCTGTTCGAGTTTGGCCCTGGACTATAAGAACTTCTCTCGCACAGACG-
CCGATTACATCAAGAAGTGAAGCTGTACTCTACGGCAACCCGGATCAGAATCTCCGGAATCC-
TAAGAAGAACAACGTGTTGACTGGGAGGAGGTGTGCCTGACCAGCGCCTATAAGGAGCTGT-
TCAACAAGTACGGCATCAATTATCAGCAGGGCGATATCAGAGCCCTGCTGTGCGAGCAGTCCG-
ACAAGGCCTTCTACTCTAGCTTTATGGCCCTGATGAGCCTGATGCTGCAGATGCGGAACAGCAT-
CACAGGCCGACCGACGTGGATTTTCTGATCAGCCCTGTGAAGAACTCCGACGGCATCTTCTAC-
GATAGCCGGAATATGAGGCCAGGAGAATGCCATCCTGCCAAAGAACGCCGACGCCAATGG-
CGCTATAACATCGCCAGAAGGTGCTGTGGCCATCGGCCAGTTCAAGAAGGCCGAGCAAG-
GAAGCTGGATAAGGTGAAGATCGCCATCTTAACAAGGAGTGGCTGGAGTACGCCAGACGAG-
CGTGAAGCACTGA

Table S5. Plasmid details

Lab name	Publication name	Details
pNAD 11	pNOC_hfnCas12a-Nlux	pNOC episomal plasmid harboring the humanized <i>fnCas12a</i> gene sequence tagged with Nlux
pNAD 12	pNOC_hlbCas12a-Nlux	pNOC episomal plasmid harboring the humanized <i>lbCas12a</i> gene sequence tagged with Nlux
pNAD 13	pNOC_hfnCas12a-Nlux_crRNA NR1	pNOC episomal plasmid harboring the humanized <i>fnCas12a</i> gene sequence tagged with Nlux and crRNA targeting the Nitrate reductase gene with spacer 1
pNAD 14	pNOC_hfnCas12a-Nlux_crRNA NR2	pNOC episomal plasmid harboring the humanized <i>fnCas12a</i> gene sequence tagged with Nlux and crRNA targeting the Nitrate reductase gene with spacer 2
pNAD 15	pNOC_hfnCas12a-Nlux_crRNA NR3	pNOC episomal plasmid harboring the humanized <i>fnCas12a</i> gene sequence tagged with Nlux and crRNA targeting the Nitrate reductase gene with spacer 3
pNAD 17	pNOC_hfnCas12a_crRNA NR1	pNOC episomal plasmid harboring the humanized <i>fnCas12a</i> gene sequence without Nlux tag and crRNA targeting the Nitrate reductase gene with spacer 1
pNAD 18	pNOC_hfnCas12a_crRNA NR2	pNOC episomal plasmid harboring the humanized <i>fnCas12a</i> gene sequence without Nlux tag and crRNA targeting the Nitrate reductase gene with spacer 2
pNAD 19	pNOC_hfnCas12a_crRNA NR3	pNOC episomal plasmid harboring the humanized <i>fnCas12a</i> gene sequence without Nlux tag and crRNA targeting the Nitrate reductase gene with spacer 3
pNAD 20	pNOC_hfnCas12a-Nlux_(-HH)crRNA NR1	pNOC episomal plasmid harboring the humanized <i>fnCas12a</i> gene sequence tagged with Nlux and crRNA targeting the Nitrate reductase gene with spacer 1 without the hammerhead ribozyme
pNAD 22	pNOC_hfnCas12a_CRATES	pNOC episomal plasmid harboring the humanized <i>fnCas12a</i> gene sequence tagged with Nlux and type IIS restriction site for endonuclease BaeI at the crRNA region that facilitates the generation of multiplexing constructs by CRATES assembly
pNAD 24	pNOC_hfnCas12a_NR1_NR2	pNOC episomal plasmid harboring the humanized <i>fnCas12a</i> gene sequence tagged with Nlux and CRISPR array with spacers 1 and 2 separated by full length 36 nts long repeat sequence. This plasmid was developed by CRATES assembly.
pNAD 60	pNOC_hfnCas12a_NR1_NR2_LPT1	pNOC episomal plasmid harboring the humanized <i>fnCas12a</i> gene sequence tagged with Nlux and CRISPR array with spacers 1 and 2 targeting the nitrate reductase gene and one spacer targeting the LPAAT gene. All spacers are separated by full length 36 nts long repeat sequence.
pNAD 44	pNOC_dfnCas12a_crRNA Td1	pNOC episomal plasmid harboring the dead version of humanized <i>fnCas12a</i> gene sequence tagged with Nlux and crRNA targeting the fluorescent gene <i>tdtomato</i> with spacer 1
pNAD 46	pNOC_dfnCas12a_crRNA Td2	pNOC episomal plasmid harboring the dead version of humanized <i>fnCas12a</i> gene sequence tagged with Nlux and crRNA targeting the fluorescent gene <i>tdtomato</i> with spacer 2
pNAD 48	pNOC_dfnCas12a_crRNA NC	pNOC episomal plasmid harboring the dead version of humanized <i>fnCas12a</i> gene sequence tagged with Nlux and spacer sequence is replaced by the type IIS restriction site for endonuclease BaeI that can be used for introducing new spacers and the plasmid was used as a negative control for the experiment
pNAD 404	pNOC_dCas9_sgRNA Td1	pNOC episomal plasmid harboring the dead version of humanized <i>spCas9</i> gene sequence tagged with Nlux and crRNA targeting the fluorescent gene <i>tdtomato</i> with spacer 1
pNAD 401	pNOC_dCas9_sgRNA Td2	pNOC episomal plasmid harboring the dead version of humanized <i>spCas9</i> gene sequence tagged with Nlux and crRNA targeting the fluorescent gene <i>tdtomato</i> with spacer 2

Lab name	Publication name	Details
pNAD 403	pNOC_dCas9_ sgRNA NC	pNOC episomal plasmid harboring the dead version of humanized <i>spCas9</i> gene sequence tagged with Nlux and spacer sequence is replaced by the type IIS restriction site for endonuclease BaeI and was used as a negative control for the experiment
pNAD 50	pNOC_dfnCas12a- KRAB_crRNA Td1	pNOC episomal plasmid harboring the dead version of humanized <i>fnCas12a</i> gene sequence tagged to the KRAB domain and Nlux and crRNA targeting the fluorescent gene <i>tdtomato</i> with spacer 1
pNAD 52	pNOC_dfnCas12a- KRAB_crRNA Td2	pNOC episomal plasmid harboring the dead version of humanized <i>fnCas12a</i> gene sequence tagged to the KRAB domain and Nlux and crRNA targeting the fluorescent gene <i>tdtomato</i> with spacer 2
pNAD 54	pNOC_dfnCas12a- KRAB_crRNA NC	pNOC episomal plasmid harboring the dead version of humanized <i>fnCas12a</i> gene sequence tagged to the KRAB domain and Nlux and spacer sequence is replaced by the type IIS restriction site for endonuclease BaeI that can be used for introducing new spacers and the plasmid was used as a negative control for the experiment
pNAD 504	pNOC_dCas9- KRAB_sgRNA Td1	pNOC episomal plasmid harboring the dead version of humanized <i>spCas9</i> gene sequence tagged to the KRAB domain and Nlux and crRNA targeting the fluorescent gene <i>tdtomato</i> with spacer 1
pNAD 501	pNOC_dCas9- KRAB_sgRNA Td2	pNOC episomal plasmid harboring the dead version of humanized <i>spCas9</i> gene sequence tagged to the KRAB domain and Nlux and crRNA targeting the fluorescent gene <i>tdtomato</i> with spacer 2
pNAD 503	pNOC_dCas9- KRAB_sgRNA NC	pNOC episomal plasmid harboring the dead version of humanized <i>spCas9</i> gene sequence tagged to the KRAB domain and Nlux and spacer sequence is replaced by the type IIS restriction site for endonuclease BaeI and was used as a negative control for the experiment



CHAPTER 5

5

High-throughput insertional mutagenesis reveals novel targets for enhancing lipid accumulation in *Nannochloropsis oceanica*

Christian Südfeld², Michal Hubáček, Daniel Rodrigues Figueiredo, Mihris Ibnu Saleem Naduthodi^{1 2}, John van der Oost¹, Rene Wijffels, Maria Barbosa², and Sarah D'Adamo^{*2},

¹Laboratory of Microbiology, Wageningen University and Research, Wageningen, Netherlands.

²Bioprocess Engineering, AlgaePARC, Wageningen University and Research, Wageningen, Netherlands.

*Corresponding Author

Adapted from publication Metabolic Engineering (2021).
<https://doi.org/10.1016/j.ymben.2021.04.012>

Abstract

The microalga *Nannochloropsis oceanica* is considered a promising platform for the sustainable production of high-value lipids and biofuel feedstocks. However, current lipid yields of *N. oceanica* are too low for economic feasibility. Gaining fundamental insights into the lipid metabolism of *N. oceanica* could open up various possibilities for the optimization of this species through genetic engineering. Therefore, the aim of this study was to discover novel genes associated with an elevated neutral lipid content. We constructed an insertional mutagenesis library of *N. oceanica*, selected high lipid mutants by five rounds of fluorescence-activated cell sorting, and identified disrupted genes using a novel implementation of a rapid genotyping procedure. One particularly promising mutant (HLM23) was disrupted in a putative APETALA2-like transcription factor gene. HLM23 showed a 40%-increased neutral lipid content, increased photosynthetic performance, and no growth impairment. Furthermore, transcriptome analysis revealed an upregulation of genes related to plastidial fatty acid biosynthesis, glycolysis and the Calvin-Benson-Bassham cycle in HLM23. Insights gained in this work can be used in future genetic engineering strategies for increased lipid productivity of *Nannochloropsis*.

Introduction

Microalgae have recently emerged as a promising platform for sustainable production of lipids, pigments and other bioactive compounds. In the last decade, the mostly marine microalgal genus *Nannochloropsis* has received considerable scientific and industrial attention as a suitable candidate for the production of high-value lipids and biofuel feedstocks [51, 262-265]. *Nannochloropsis oceanica*, in particular, exhibits robust outdoor growth at relatively high growth rates and it can reach neutral lipid contents of up to 50% of the dry cell weight [61, 70, 74]. However, these exceptionally high neutral lipid contents are only reached when *N. oceanica* is exposed to a stress condition that is unfavorable for cell division, for instance nitrogen (N) deficiency [50]. This poses an obstacle for achieving economically viable lipid productivities because biomass production rates are impaired under N deficiency. One way to reach high neutral lipid production without impairing growth is by genetic engineering [254], but progress is hampered by limited functional genome annotations, as research into the polyphyletic group of microalgae is relatively young [266-268]. About 50% of all predicted genes of the model strain *N. oceanica* IMET1 lack a functional annotation because the majority of the predicted proteins do not display sufficient sequence similarities with known proteins [63, 72]. This poses a serious limitation for targeted modification of the metabolic network of this organism.

One well-established way to investigate the metabolic functions of unknown proteins in an unbiased manner *in vivo* is the use of forward genetic screens [269, 270]. This approach couples random mutagenesis to a phenotypic screening for traits of interest. After this screening, mutants with the desired characteristics are analyzed genetically and causative mutations are identified. This allows the researcher to map a phenotype and metabolic function to a protein. Typical mutagenesis strategies include chemical- or radiation-induced mutagenesis which lead to nucleotide substitutions in the host DNA. These mutations can induce changes in primary protein sequences and gene expression patterns [271, 272]. An advantage of these mutagenesis strategies is a desirably broad phenotypic diversity. However, these strategies rely on time-consuming and costly genotyping procedures, such as whole genome resequencing [273]. Moreover, detecting the causative mutations for a selected trait can be difficult due to several nucleotide substitutions in multiple different genes.

An alternative mutagenesis approach is insertional mutagenesis, which is the introduction of a foreign DNA molecule, called an insertion cassette (IC) into the genome of an organism at random positions [274, 275]. This approach has been the gold standard for elucidating gene functions in the context of forward genetics for a variety of organisms including microalgae [276-278]. The insertion of the IC can be facilitated by enzymes such as transposases, or it can occur through nonhomologous end joining during DNA repair events at the site of double-stranded breaks. If the IC is inserted into a gene, the encoded protein will be either truncated or knocked out and thereby lose its metabolic function. The insertion site in strains with interesting characteristics can be identified through modified PCR procedures such as genome walking [279] or RESDA-PCR [280]. Although this obviates the need for whole genome resequencing, the commonly employed methodologies for tracing insertions are time-consuming. Moreover, they rely on unpredictable factors,

such as the presence of a specific restriction endonuclease (REN) recognition sequence in close proximity to the insertion site. Therefore, these methodologies will likely be too cumbersome to unravel the functions of the myriad of unknown genes in *N. oceanica*. An elegant variation of insertional mutagenesis was first reported for *Escherichia coli* [281] and later adapted for *Chlamydomonas reinhardtii* [57, 278]. It involves using an IC that carries recognition sequences for a type IIS REN at both of its termini. Because this type of enzyme cuts in a defined distance outside of its recognition site, the IC can be precisely extracted together with a short defined stretch of adjacent genomic DNA from the genome of a transformed strain. This alleviates any uncertainty during genotyping PCRs.

Phenotypical screenings in forward genetics of microbial organisms often include high-throughput single cell methodologies such as fluorescence-activated cell sorting (FACS), which has gained an outstanding importance in the field of microalgal strain development [282, 283]. Based on flow cytometry, this technique allows quantification of cellular characteristics such as size and chlorophyll content through the light scattering or fluorescent properties of different cell constituents. Fluorescent probes such as BODIPY dye can be utilized to further quantify non-auto fluorescent components of cells, for example lipid bodies. Lipid-rich microalgal strains have successfully been isolated by FACS for a variety of microalgal genera such as *Chlamydomonas* (Terashima et al. 2015), *Dunaliella* [284], *Chlorococcum* [285] and *Nannochloropsis* [286] [287]. Recently, first studies have shown that coupling insertional mutagenesis and FACS methodology can lead to the isolation of high lipid-producing *Nannochloropsis* strains [288, 289]. This has opened the door to sophisticated forward genetic screen that will aid in unravelling gene functions in this organism at a larger scale. These screenings would greatly benefit from building upon a straightforward and reliable genotyping procedure.

The goal of this study was to identify novel genes that are associated with increased lipid accumulation phenotypes in *N. oceanica*. To achieve this, we developed a forward genetic screen based on insertional mutagenesis, featuring high-throughput phenotype screening using FACS, and a rapid and improved genotyping procedure using the type IIS REN Mmel.

Results and Discussion

Construction and screening of a random insertional mutant library

We created a random insertional mutant library of *N. oceanica* by transforming 120 independent samples with an insertion cassette (IC). This IC (Fig. 1A) contains a zeocin resistance gene (zeo^R) driven by the endogenous promoter P_{VCP} [146, 227]. The IC further carries two transcriptional terminators in head-to-head orientation to ensure that genes at the insertion site will be knocked out or truncated regardless of cassette orientation. Terminal barcodes were added along with Mmel restriction endonuclease (REN) recognition sites, to enable rapid tracing of the insertion site by cassette PCR. Microalgal transformants were selected for resistance to Zeocin on agar plates, and ~25,000 colonies were pooled into a single flask to create the insertional mutant library (Fig. 1B).

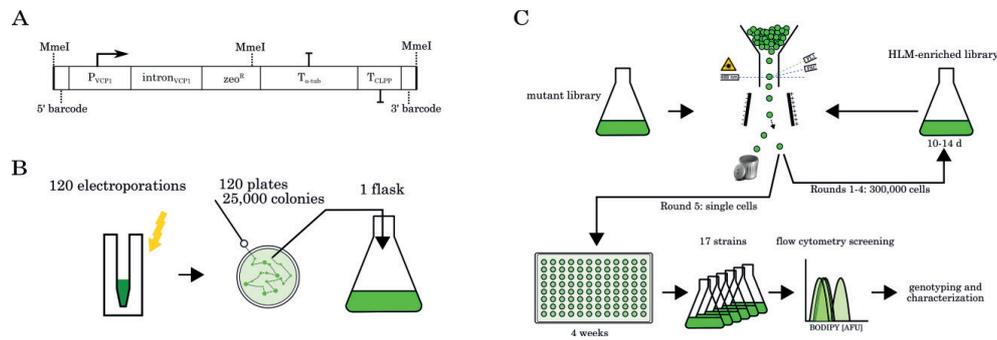


Figure 1. Creation and screening of an insertional mutant library. (A) Schematic of the IC used to transform *N. oceanica*. The resistance gene zeo^R is driven by the endogenous VCP gene promoter and flanked by 2 transcriptional terminators in head-to-head orientation. Terminal barcode sequences and recognition sequences for the REN MmeI are needed for genotyping. (B) Wild type *N. oceanica* cells were transformed by electroporation by using 120 μg of DNA in 120 separate reactions. Transformants were selected on zeocin-containing agar plates and $\sim 25,000$ colonies were pooled into a single flask for screening. (C) Screening pipeline for enrichment and isolation of high lipid mutant (HLM) cells. The initial mutant library was harvested during exponential growth phase, stained with neutral lipid dye BODIPY, and 300,000 cells were sorted into a new flask by gating for high BODIPY fluorescence (Fig. S1). The HLM-enriched library was recovered for 10-14 d before the next sorting step. The library was enriched for HLM cells 4x before the fifth and final sorting during which strains were isolated by deposition of single cells onto agar plates. 17 colonies were selected for expansion and NL content screening by BODIPY staining and flow cytometry.

The mutant library was repeatedly screened and enriched for high lipid mutant (HLM) strains by iterative sorting using FACS for 5x in total (Fig. 1C). Consecutive rounds of sorting can help to enrich a population of cells within a random mutant population [285]. Screening for HLM strains was achieved using the fluorescent BODIPY dye, which specifically and quantitatively stains neutral lipids (NLs) without affecting cell viability [261]. The mutant library was grown to mid-exponential stage, stained with BODIPY, and cells with high NL content were selected by applying a 2-dimensional “high-NL” gate in the channels capturing BODIPY fluorescence and forward scatter (Fig. S1). The forward scatter was taken into account as a proxy for cell size to ensure that smaller HLM cells would be sorted as well. These cells may have only average absolute BODIPY fluorescence but higher BODIPY fluorescence than most cells of similar size. In the fifth round of sorting, single cells were deposited onto an agar plate to isolate strains from the enriched library.

Screening and characterization of isolated high lipid mutant strains

After strain isolation, 17 colonies were further screened for NL content cell^{-1} to confirm that isolated strains were higher in NL content and to select candidates for in-depth analyses. Cultures were grown to mid exponential growth phase, stained with BODIPY and analyzed by flow cytometry. The medians of the BODIPY fluorescence distributions were used as an estimation of the average NL content cell^{-1} ($N \geq 2$, Fig. 2). Almost all strains showed an increase in average BODIPY fluorescence levels. Five strains with particularly high BODIPY fluorescence (highlighted red in Fig. 2) were selected for physiological and biochemical characterization.

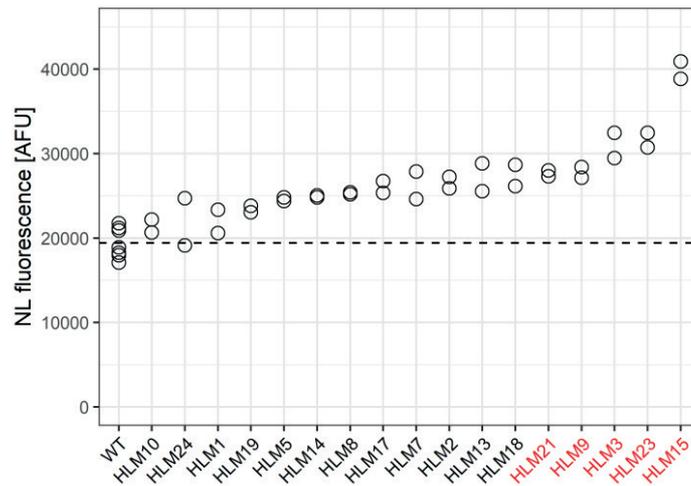


Figure 2. Cellular NL content of 17 mutant strains compared to the wild type. Cellular NL contents were quantified using lipophilic fluorophore BODIPY and flow cytometry. Each point represents the median of single cell BODIPY fluorescence emission for an independent culture. The mean BODIPY fluorescence of wild type cultures is indicated by a dashed line. Mutant strains that were chosen for further analysis are highlighted red.

Characterization of promising strains

We compared the growth of five transformant strains and wild type in a batch cultivation with diurnal light cycles at $600 \mu\text{mol m}^{-2} \text{s}^{-1}$ illumination intensity, using a parallel screening photobioreactor. All strains grew exponentially for 3 d and continued to grow linearly thereafter (Fig. 3A, $N \geq 3$). HLM15, frequently collapsed after >3 d of cultivation and showed a 21% reduction in maximum specific growth rate compared to the wild type (Fig. 3B, Tab. S1). HLM9 showed a 10% reduction in maximum specific growth rate compared to the wild type, whereas growth rate of the remaining mutants was unaffected. HLM21 and HLM23 showed a higher maximum quantum efficiency of photosystem II (PSII) photochemistry (F_v/F_m), compared to the wild type and HLM15 (Fig. 3C, Tab. S1). We observed that all HLM strains had larger cells compared to the wild type, and all but HLM9 showed increased chlorophyll a autofluorescence per cell (Fig. S2).

We quantified the neutral lipid (NL) and polar lipid (PL) contents and characterized the fatty acid composition in exponentially growing cultures of all strains (Fig. 4, $N \geq 3$). NL contents DCW^{-1} were increased by 40% in HLM9 ($\text{mean } (M) = 0.302\text{gg}^{-1}$, $\text{standard deviation } (SD) = 0.017$) and HLM23 ($M = 0.302\text{gg}^{-1}$, $SD = 0.016$) and by 58% in HLM15 ($M = 0.341\text{gg}^{-1}$, $SD = 0.047$) (Fig. 4A, Tab. S1) relative to the wild type ($M = 0.215\text{gg}^{-1}$, $SD = 0.014$). Increased NL accumulation in these strains did not occur at the expense of PL contents, which were similar for all strains (Fig. 4B, Tab. S1).

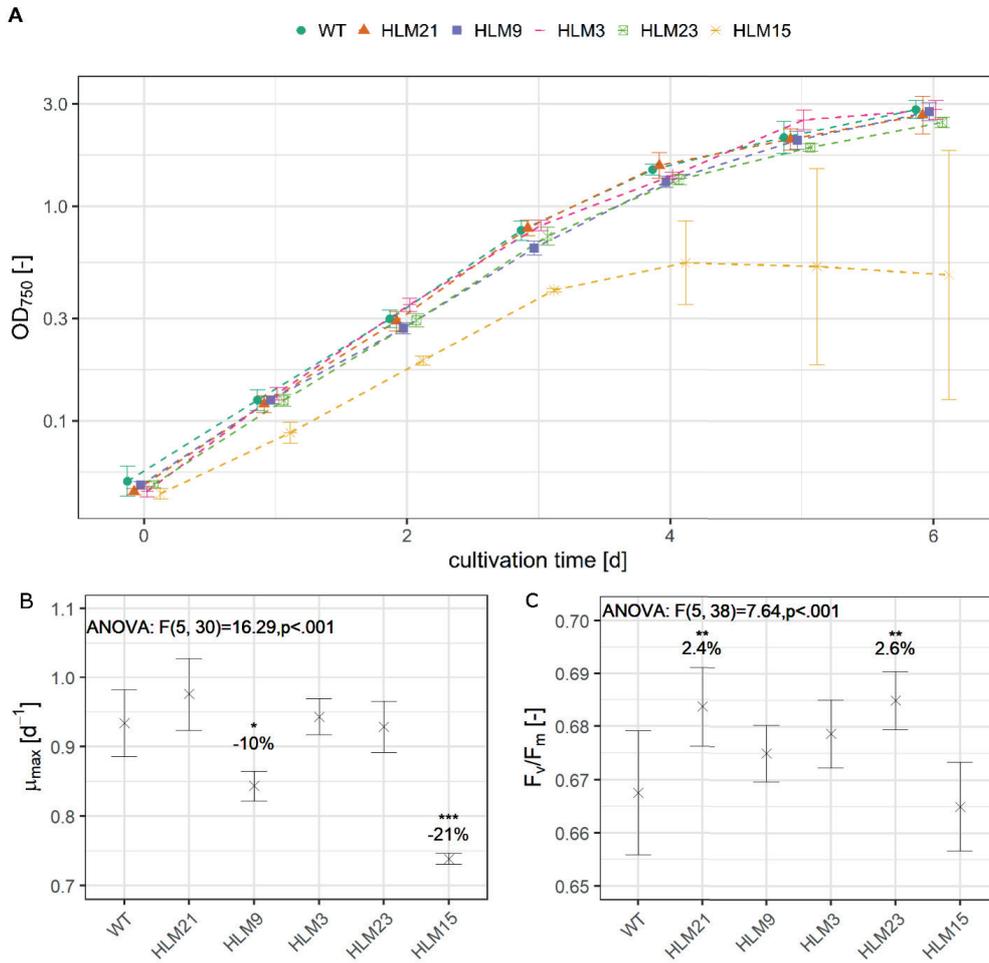


Figure 3. Physiological characterization of selected transformant lines. (A) Growth curves of five transformants and wild type on a logarithmically-scaled y-axis ($N \geq 3$). All cultures showed exponential growth for 3 d, until they reached an OD_{750} of ~ 0.8 . HLM15 cultures were prone to collapsing after this period whereas all other strains displayed linear growth until they reached an OD_{750} of $\sim 5-6$ (graph truncated). Final biomass concentrations were comparable between cultures. Points are dodged horizontally for better distinction between samples. (B) Maximum specific growth rates (μ_{max}) of all analyzed strains were calculated for the exponential growth phase. The cross and errorbars denote the mean \pm SD ($N \geq 3$). Growth rates of HLM3, HLM21, HLM23 were similar to the wild type, whereas growth rates of HLM9 and HLM15 were decreased by 10 and 21% respectively. (C) Maximum quantum efficiency of PSII photochemistry (F_v/F_m) was comparable between the wild type and HLM3, HLM9 and HLM15. Increased photosynthetic performance was seen for HLM21 and HLM23. (B-C) Significant changes relative to the wild type are indicated above each group. Significance levels were assessed by Tukey's HSD test in case of a significant ANOVA outcome. (*): $p < 0.05$; (**): $p < 0.01$; (***) : $p < 0.001$.

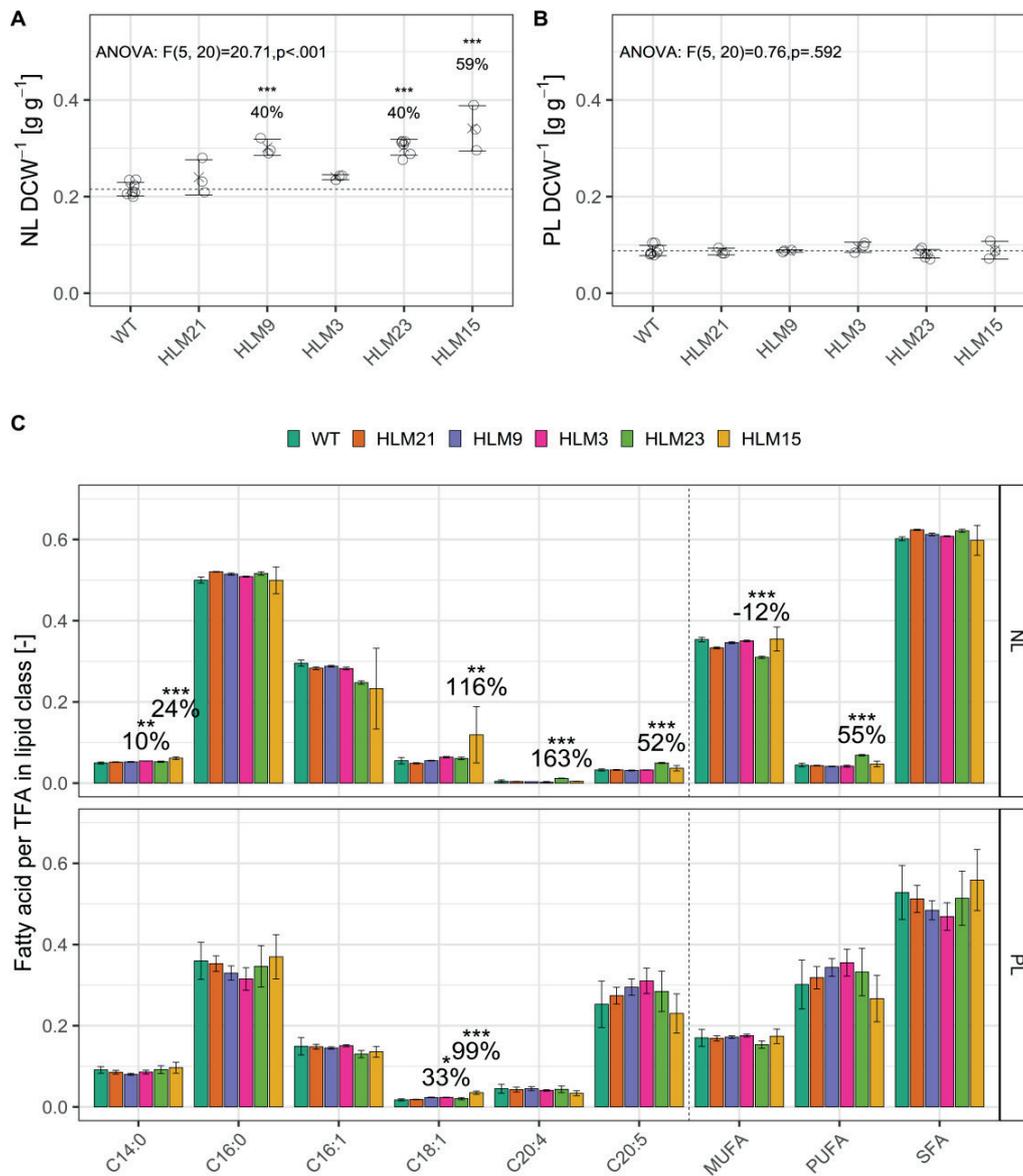


Figure 4. Biochemical characterization of selected transformant lines. (A,B) The amount of neutral (A) and polar (B) lipids DCW⁻¹. The superimposed cross and error bars denote the mean \pm SD. Changes in lipid content relative to the mean wild type lipid content (dashed lines) and significance levels are given above each group. All strains had either increased or unchanged NL content DCW⁻¹ compared to the wild type. Two-way ANOVA did not suggest a main effect of microalgal strain on PL content DCW⁻¹. (C) Fatty acid composition of neutral lipids (NL) and polar lipids (PL) are shown for selected transformants. Bars and error bars denote the mean \pm SD of $N \geq 3$ biological replicates. Changes relative to the wild type are shown for significant increases or decreases. (A-C) Significance levels were assessed by Tukey's HSD test. (*): $p < 0.05$; (**): $p < 0.01$; (***): $p < 0.001$. MUFA: monounsaturated fatty acids. PUFA: polyunsaturated fatty acids. SFA: saturated fatty acids.

In addition to the increase in NL content DCW⁻¹, HLM23 and HLM15 showed an altered fatty acid composition (Fig. 4C). HLM23 had a 12% decreased fraction of monounsaturated fatty acids (MUFAs) in NLs, concomitant with a 55% increase in polyunsaturated fatty acids (PUFAs). The increased PUFA content was mainly due to an increase in the fraction of C20:4 n6 and C20:5 n3, known as eicosapentaenoic acid (EPA). EPA is considered a high-value molecule because it is an essential dietary compound and has potential in disease prevention [290, 291]. In *Nannochloropsis*, EPA is mainly found in lipid classes such as MGDG or DGDG that are associated with the thylakoid membranes of the chloroplast [292], whereas the fraction of EPA in NLs is negligible under favorable growth conditions. However, translocation of EPA to NLs can be observed during nitrogen stress in wild type *Nannochloropsis* [293].

HLM15 showed increases of 24 and 116% for C14:0 and C18:1 (oleic acid), respectively, in the NL fraction. Oleic acid was further increased by 99% in the PL fraction, compared to the wild type. Oleic acid is a suitable candidate for enrichment in biodiesel due to its chemical properties [294].

Tracing of cassette insertion sites in selected transformants

We traced the insertion sites in the nuclear genome of the five *N. oceanica* HLM mutants discussed above, by using a novel implementation of Mmel-based version of cassette PCR, which was inspired by the previously reported ChlMmeSeq technique [278, 295]. The procedure produces Mmel-restriction fragments that are ligated to an adapter molecule and amplified by cassette PCR (Fig. S3). Using a single REN and obtaining fragments with a constant length offer a substantial improvement over other versions of cassette PCR that rely e.g. on the presence of REN recognition sequences in the genomic DNA around the insertion site [279, 280].

The previously reported method [278] was altered to eliminate spurious amplification of unspecific templates (Fig. S3). These modifications include: 1) the use of a non-symmetrical adapter (Fig. S3D, blue), inspired by vectorette PCR [296], which prevents background noise due to amplification of genomic DNA fragments that have adapter molecules ligated to both ends; 2) An amine modification at the 3'-end of the shorter adapter strand that prevents the formation of an AP-template strand by extension of the adapter on DNA fragments without a barcode; 3) Use of synthetic barcode sequences at the cassette termini, to allow design of cassette-specific primers that have low similarity with endogenous genome sequences; 4) Nested PCR protocols (Fig. S3F) with touchdown temperature settings [297, 298] in both PCR cycles that further increase the stringency of target amplification. Using our adapted version of Mmel-based insertion site tracing, we were able to map the insertion sites of HLM21, HLM9, HLM3, HLM23 and HLM15 to five distinct genomic loci (Fig. 5). Putative insertion sites were confirmed by PCR with primers (arrows with half arrowheads in Fig. 5) complementary to the genome sequence of *N. oceanica* near the insertion sites.

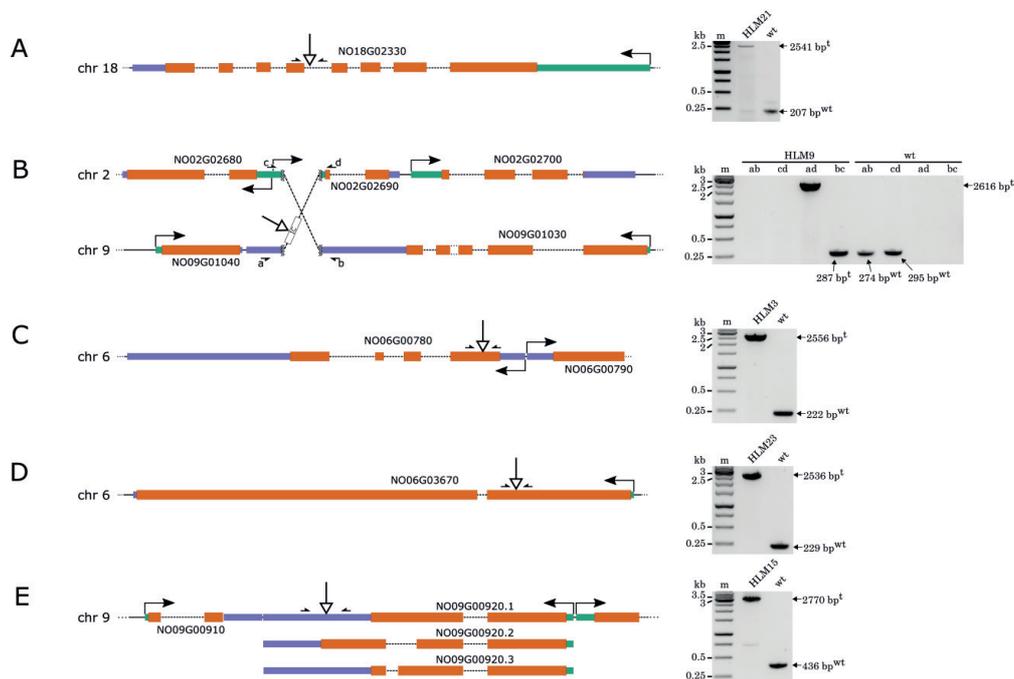


Figure 5. Schematic of insertion sites in selected transformant strains and PCR verification of tracing. (A) Insertion in HLM21 occurred inside of the fourth intron of *NO18G02330*. (B) Insertion in HLM9 occurred during reciprocal translocation of chromosomes 2 and 9. The IC (white-filled box labelled IC) was ligated between the two shorter fragments. The 5'-UTR of *NO02G02690* and the 3'-UTR of *NO09G01030* were disrupted during chromosomal rearrangement. Exon three of *NO09G03010* was shortened for clarity. a-d denote PCR primers that were used in combinations a+b (ab), c+d, a+d and b+c for verification of rearrangement. (C) Cassette insertion in HLM3 occurred inside of the first exon of *NO06G00780*. (D) Cassette insertion in HLM23 occurred inside of the first exon of *NO06G03670*. (E) Cassette insertion in HLM15 occurred in *NO09G00920*. The last exon of isoform 2 and the 3'-UTR of isoforms 1 and 3 were disrupted by the insertion. Gene exons are depicted in red, putative 5' and 3'-UTR sequences in green and blue respectively. Dashed lines denote introns. Solid kinked arrows show putative transcription start sites. Arrows with white-filled arrowheads show the position of cassette insertion. Arrows with solid, half arrowheads show primer binding sites during PCR reactions that were done to verify correct tracing. Kilobase (kb) sizes of marker (m) bands are indicated on the left side of gel images. Markings on the right and bottom side of gel images indicate the expected band size for transformant (t) and wild type (wt) samples.

Cassette insertion in HLM21 occurred in a tetratricopeptide repeat domain-containing protein

In HLM21, the insertion cassette was integrated inside intron 4 of *NO18G02330* (Fig. 5A). Due to the transcriptional terminators present in the insertion cassette, the exons 5-8 of *NO18G02330* are likely not transcribed in HLM21, resulting in a loss of the 190 C-terminal amino acids in the corresponding protein. The uncharacterized *NO18G02330* has insufficient similarity to any known protein to predict its metabolic function. We bioinformatically analyzed the *NO18G02330* primary protein sequence to predict protein localization and identify conserved domains. The protein does not contain any predicted target peptide, and it contains a tetratricopeptide repeat (TPR) domain. TPRs are found in proteins involved in a variety of metabolic functions, such as cell cycle coordination and

transcriptional regulation [299], protein folding [300], transport across membranes [301-303], membrane assembly [304], vesicle fusion [305] and bio-mineralization [306]. The TPR domain usually serves as an interface for interactions between proteins and multiprotein complex assemblies [307]. Thus, it is likely that NO18G02330 interacts with one or more proteins in *N. oceanica* and that this interaction is disrupted in HLM21. More detailed studies are necessary to verify this hypothesis, for instance by screening for possible interaction partners with the well-established yeast two-hybrid system [308, 309].

Cassette insertion in HLM9 occurred during chromosomal rearrangement

In HLM9 we traced the IC insertion to two different chromosomes for the 5' and 3' sequencing reactions. PCR analyses showed that a reciprocal translocation t(2;9) had occurred after double-stranded breaks at position 739,793 of chromosome 2 (total length 1,674,082 bp) and position 349,677 of chromosome 9 (total length 1,290,044 bp). The longer fragments of both chromosomes were each ligated to one side of the IC (Fig. 5B), creating a new, larger chromosome, named here derivative chromosome 9 (total length 1,876,990 bp). The two shorter fragments were ligated to each other creating a new, shorter chromosome, named here derivative chromosome 2 (total length 1,089,470 bp). The chromosome number of a derivative chromosome would normally be assigned based on the chromosome number that donated the centromere, but the positions of centromeres are not known for *N. oceanica*. The designated derivative chromosome numbers are therefore putative. Whereas chromosomal translocations can affect global gene expression in an unpredictable manner, e.g. in mammalian cells [310], they have a more concrete effect on expression of the genes located at the breakage, for instance by causing gene knockouts or knockdowns or by creating novel fusion genes.

In HLM9, the putative promoter region and/or the 5'-UTR of *NO02G02690* and the 3'-UTR of *NO09G01030* are disrupted by the translocation and IC insertion (Fig. 5B). According to transcriptomic analyses [74, 311, 312], *NO02G02690* is a low expressed gene encoding a protein of unknown function that has no homologues outside of the *Nannochloropsis* genus. Orthologues in *N. gaditana* (*Naga_100026g34*, E value = 2×10^{-16} , 55% sequence identity on 55% coverage using BLASTp) and *N. salina* (*NSK_001029*, E value = 3×10^{-11} , 60% sequence identity on 41% coverage using BLASTp) are hypothetical proteins. *NO09G01030* encodes a fungal-type Zn(II)₂Cys₆ DNA-binding, Zinc finger (ZF) domain protein, which has high amino acid sequence similarity with the *Nannochloropsis gaditana* transcription factor *Naga_100104g18* (E value = 0.0, 65.59% identity on 72% sequence coverage using BLASTp, 44.46% identity on full length sequences using CLUSTALΩ). *Naga_100104g18* was previously identified as a control hub for lipid content in *N. gaditana* [80]. Ajjawi and co-workers found that a knockdown of *Naga_100104g18* increased lipid content by ~2-fold and resulted in a 5-15% reduction in growth compared to the wild type. Similarly, HLM9 displayed a 9% reduction in growth and 40% increase in NL content compared to the wild type in our experiments, which suggests that *NO09G01030* and *Naga_100104g18* may be functional homologues and that *NO09G01030* expression may be decreased in HLM9. In their study, Ajjawi and co-workers inserted an antibiotic resistance cassette into the 3'-UTR of *Naga_100104g18*, which reduced transcript abundance by 50%. This drastic decrease exemplifies the importance of an intact 3'-UTR for gene expression in *Nannochloropsis*. Ajjawi and colleagues inserted the antibiotic resistance cassette at a

position 30 nucleotides downstream of the translational STOP codon, which replaced ~98% of the exceptionally long 3'-UTR of *Naga_100104g18* with heterologous sequence. Opposed to this, the truncation of the 3'-UTR of *NO09G01030* in HLM9 occurred much further away from the translational STOP codon of the gene, leaving ~70% of the endogenous 3'-UTR sequence intact (Fig. 5B).

Cassette insertion in HLM3 occurred in a Zn(II)-finger type protein of unknown function

Cassette insertion in HLM3 disrupted *NO06G00780* in the first exon (Fig. 5C). *NO06G00780* is a relatively short protein (190 AAs) without predicted subcellular compartmentalization that carries a ZF-HIT domain at its C-terminus. Proteins containing this domain are mostly nuclear proteins associated with gene regulation, chromatin remodeling or pre-snoRNP assembly in multimodal RNA/protein complexes [313]. Their precise functions, however, remain largely unknown. *NO06G00780* has moderate-high similarity (E value $\geq 2 \times 10^{-41}$) with a variety of proteins mostly from filamentous protists belonging to the subphylum Oomycota. Oomycota is a sister-group to the *Nannochloropsis*-containing subphylum Ochrophyta, and consists of non-photosynthetic organisms including the extensively researched parasite *Phytophthora infestans* [314, 315]. A few of these proteins are putatively classified as Ca^{2+} :cation antiporters such as the *Phytophthora palmivora* protein POM74018.1 (E value = 2×10^{-35} , 43% sequence identity on 94% coverage using BLASTp). However, none of the putative homologues that were identified by BLASTp analysis are functionally characterized.

Cassette insertion in HLM15 occurred in a DnaJ (HSP40) gene

The cassette insertion in HLM15 occurred in *NO09G00920*, which encodes 3 isoforms of a DnaJ type B protein. DnaJ proteins, otherwise known as HSP40, are part of a molecular chaperone system that is present in all eukaryotic cells. This chaperone system, consisting of HSP40, HSP70 and nucleotide exchange factors, plays an essential role in protein homeostasis by governing protein folding, unfolding, translation, degradation, translocation and protein disaggregation [316-318]. DnaJ/HSP40 family members are obligate co-chaperones to HSP70 as they deliver substrate to HSP70 and stimulate its ATPase activity [319, 320]. As molecular chaperones, DnaJ/HSP40 proteins are pivotal for protein homeostasis especially under stress conditions. DnaJ loss of function mutants of *E. coli* e.g. fail to grow at elevated and reduced temperatures but show less drastic phenotypes at optimal growth conditions [321]. Severe growth defects, temperature sensitivity or lethal phenotypes were also observed for knockouts of different cytosolic DnaJ variants in *Saccharomyces cerevisiae* [322]. In *Arabidopsis thaliana*, the knockout of a nuclear-encoded chloroplastic DnaJ family member induced altered photosynthetic behaviour and triggered a global stress response in transgenic plants including expression of genes related to ROS detoxification. A chloroplastic DnaJ protein was also disrupted in a chemically mutagenized *N. gaditana* strain that showed a light-dependent increase in lipid content of up to ~60% DCW⁻¹ and a changed proton motive force at the thylakoid membrane [323]. However, this *N. gaditana* strain carried mutations in 233 additional genes, which prevents drawing of reliable conclusions on a direct link between the DnaJ mutation and the observed phenotype.

Based on the high functional conservation of DnaJ proteins, *NO09G00920* likely aids protein homeostasis in *N. oceanica*, and it may play a crucial role in stress response. A functional impairment of *NO09G00920* in HLM15 would explain the 21% decrease in growth rate observed for this strain, as substantial growth defects were also observed for DnaJ mutant strains of other organisms. A function of *NO09G00920* in stress response is further suggested by HLM15's propensity for collapsing after prolonged incubation under high light conditions (Fig. 3A). Generally, high light conditions cause an overreduction of the photosynthetic electron transport chain in phototrophs. This leads to the formation of reactive oxygen species (ROS) [324], which can oxidize proteins. Oxidized proteins must be degraded by the cell to prevent toxic effects [325]. An impaired activity of *NO09G00920* in HLM15 could pejorate the strain's ability to cope with protein aggregation or unfolded protein stress, which offers a potential explanation for its decrepit growth and increased high light susceptibility. Unfolded protein stress is further linked to lipid biosynthesis. Mitigation of unfolded protein stress involves activation of the unfolded protein response (UPR) pathway that is tied to TAG production and lipid droplet formation in *S. cerevisiae* [326], animal cells [327], higher plants [328] and microalgae [329]. The increased NL content and decreased growth rate of HLM15 may therefore be the consequence of an elevated level of unfolded protein stress due to the decreased chaperone activity. Similarly, *C. reinhardtii* mutants with an impaired UPR grew slower and contained more TAG than the parental strain when treated with the unfolded protein stress inducer tunicamycin [329].

According to the representative version of the *N. oceanica* IMET1 genome [330], *NO09G00920* is expressed in 3 different splice variants (*NO09G00920.1*, *NO09G00920.2* and *NO09G00920.3*) in *N. oceanica*. The existence of *NO09G00920.2* is strongly supported by high similarity of this protein with known DnaJ proteins of other organisms over the full amino acid sequence length. However, the C-termini of the predicted *NO09G00920* isoforms 1 and 3 show no similarities with any known protein in the NCBI database, suggesting that they may be misannotated. *NO09G00920* contains an N-terminal J domain, which is responsible for the interaction of DnaJ/HSP40 proteins with HSP70, whereas the C terminus of the protein may be responsible for substrate specificity and functional diversity [331]. Depending on the splice variant, the insertion of the IC in HLM15 happened in either the 3'-UTR (for the putative isoforms *NO09G00920.1* and *NO09G00920.3*), or, in case of *NO09G00920.2*, inside the last putative exon of the CDS, 30 nucleotides away from the translational STOP codon. This results in a substitution of the 10 C-terminal amino acids (AAs) of *NO09G00920.2* with the dipeptide Trp-Ile. It is unclear whether this mutation would have a detrimental effect on the activity of the 413 AA-containing protein, but a conserved domain of unknown function (DUF1977) is present at the very C-terminus of isoform *NO09G00920.2* and this domain is truncated by the insertion. The 3' end of the IC further replaces the original 3'-UTR of *NO09G00920.2* and the biggest part of the original 3'-UTR of *NO09G00920.1* and *NO09G00920.3* in HLM15 (Fig. 5E). Gene expression may be severely affected for all putative isoforms because the 3'-UTR plays a crucial role in modulating gene expression by regulating transcript localization, stability and translational efficiency [332]. The substantial increase in NL content in HLM15 suggests that DnaJ proteins are potential targets for genetically engineering increased lipid productivities in microalgae. Despite this, we must point out that interfering with mechanisms that ensure protein homeostasis can negatively influence biomass

productivity and culture stability under stress conditions, which is a disadvantage for industrial production processes.

Cassette insertion in HLM23 occurred in an APETALA2-like transcription factor

In HLM23, IC insertion occurred in the first exon of *NO06G03670*. We identified conserved domains in *NO06G03670* by querying the conserved domain database [333]. *NO06G03670* contains two AP2 domains (Fig. 6A), which are conserved domains involved in DNA binding in the plant superfamily of transcription factors AP2/EREBP, and an SNF2 domain and adjacent tandem chromo domains that are together found in ATP-dependent chromatin remodelers. Moreover, an N-terminal plant homeodomain (PHD) type Zn-finger may aid in interaction with histones or other proteins. In HLM23, *NO06G03670* is rendered non-functional because the protein is truncated after the PHD domain and therefore misses the domains responsible for DNA-binding and chromatin remodeling.

In a computational study, *NO06G03670* has been identified as one of three putative orthologues of the transcription factor AtWRI1, which is related to lipid metabolism in *A. thaliana* [334]. This gene encodes a transcription factor belonging to the family of APETALA2-like proteins from higher plants and falls into the transcription factor superfamily AP2/EREBP [335]. In plants, AP2/EREBP family members are reported to be involved in stress response, sugar metabolism, hormone signaling, and the coordination of developmental processes such as floral organ identity determination and seed germination [336-338]. In their pioneering work, Hu and co-workers identified numerous transcription factor genes and transcription factor binding sites in *Nannochloropsis* by *in silico* analyses, and constructed a preliminary global regulatory network [334]. *NO06G03670* was predicted to modulate transcription of 74 and 69 genes positively and negatively respectively, by binding to the DNA motifs 5'-CGCGCCAW-3', 5'-TCCGCCCM-3' and 5'-GCCSATCC-3' in promoter sequences. These motifs are enriched in genes related to chromosome organization, protein modification and cofactor metabolic processes, respectively. Among the genes predicted to be regulated by *NO06G03670* are several genes related to fatty acid biosynthesis.

Table 1 summarizes the phenotypes and genotypes associated with the HLMs that were investigated more closely in this study.

Table 1. Summary of transformant pheno- and genotypes. All values are increases relative to the wild type strain for a one-step batch cultivation. The NL productivity increase is a projection based on the maximum specific growth rate (μ_{\max}) and NL content during exponential growth phase. Significance levels in parentheses were calculated by Tukey's HSD test. (ns): not significant, $p > 0.05$; (*): $p < 0.05$; (**): $p < 0.01$; (***): $p < 0.001$.

Strain	μ_{\max}	NL DCW ⁻¹	NL productivity	Affected gene	Protein type
HLM21	6% (ns)	12% (ns)	19%	NO18G02330	unknown function
HLM9	-10% (*)	40% (***)	26%	NO09G01030/ NO02G02690	Zn(II) ₂ Cys ₆ fungal type TF
HLM3	1% (ns)	12% (ns)	13%	NO06G00780	Zn finger type
HLM23	-1% (ns)	40% (***)	39%	NO06G03670	APETALA2-like TF
HLM15	-21% (***)	59% (***)	26%	NO09G00920	DnaJ type II

Targeted knockout of *NO06G03670* in wild type *N. oceanica* confirms the HLM23 phenotype

Because mutant HLM23 showed the highest potential for lipid-overproduction in the here presented experiments (Tab. 1), we selected this strain to further confirm the connection between its genotype and phenotype. To do so, we employed a plasmid-based CRISPR-Cas mutagenesis strategy, which was recently developed for *Nannochloropsis* [79]. Using this strategy, genes can be knocked out in a highly controlled manner, with the help of a plasmid that encodes a constitutively-expressed Cas12a protein and a crRNA sequence. Upon plasmid delivery into cells, this system creates double-stranded breaks in a 5'-proximal gene exon. Repair of double-stranded breaks by nonhomologous end joining can lead to frameshift mutations that can be identified by DNA sequencing [339]. Frameshift mutations are considered a simple and robust strategy to produce targeted genetic knockouts [340].

By employing this approach, we designed three crRNAs with protospacer sequences (sp1, sp2 and sp3) that are complementary to different positions of the first exon of *NO06G03670*. The positions were chosen upstream of the IC insertion site in HLM23 (Fig. 6A), to ensure that the resulting truncated *NO06G03670* proteins would not contain any additional domains compared to the truncated *NO06G03670* of HLM23. The protospacer sequences sp1, sp2 and sp3 were incorporated into a Cas12a-encoding plasmid, generating pTKO-sp1, pTKO-sp2 and pTKO-sp3, respectively. The three plasmids were used to transform wild type *N. oceanica*, and ten mutant colonies per construct were screened by PCR and sequencing (Fig. 6A). 0/10, 2/10 and 0/10 colonies with nucleotide mutations were recovered for pTKO-sp1, pTKO-sp2 and pTKO-sp3, respectively. The pTKO-sp2 colonies three (pTKO-sp2-C3) and seven (pTKO-sp2-C7) showed a deletion of *NO06G03670* nucleotides 1781-1784 (CCAA). This deletion had occurred around the position expected for Cas12a-mediated cleavage, corresponding to nucleotides 18-21 of protospacer sp2 [102]. The four nucleotide deletion causes a shift of the genetic reading frame, rendering *NO06G03670* non-functional in pTKO-sp2-C3 and pTKO-sp2-C7.

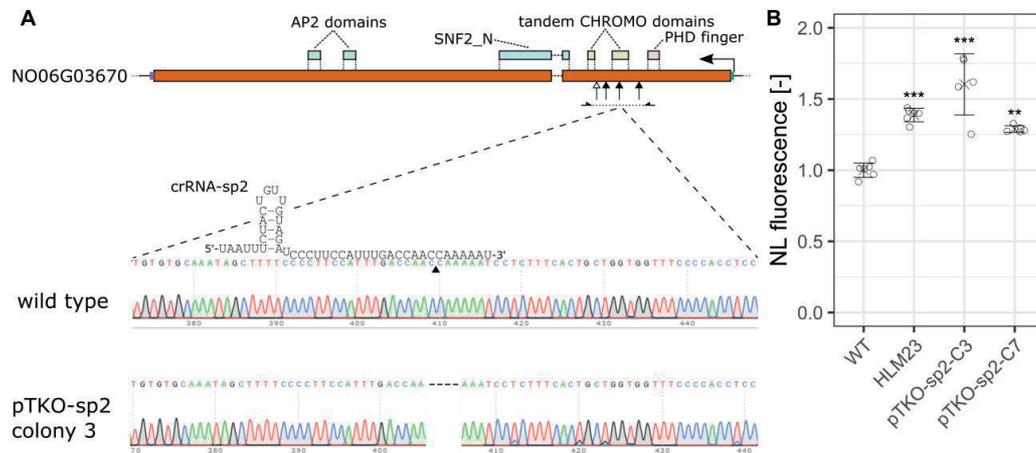


Figure 6. CRISPR/Cas-mediated knockout of *NO06G03670* in wild type *N. oceanica* re-instates the HLM23 phenotype. (A) Schematic of the domain architecture of *NO06G03670* and illustration of knockout approach using Cas12a. The insertion site (white-headed arrow) of the IC in HLM23 lies between tandem CHROMO domains within the first exon of the gene. An SNF2_N and AP2 domains that are required for recognition of target DNA sequences by the transcription factor, are located between the insertion site and the C-terminal end of the protein and are not present in the truncated *NO06G03670* of HLM23. Three crRNA protospacers were designed to target the first *NO06G03670* exon at different positions (black-headed arrows) between the HLM23 insertion site and the translation start site. The homology between crRNA-sp2 and the target region of wild type *NO06G03670* nucleotide sequence is shown exemplarily. pTKO-sp2 colonies 3 and 7 had a four nucleotide deletion at the expected position for Cas12a-mediated cleavage (solid black triangle), which causes a shift of the genetic reading frame. Chromatograms illustrate Sanger sequencing data of PCR products obtained from genomic DNA of the wild type strains and pTKO-sp2 colony 3, using primers located in the first exon of *NO06G03670* (solid black half-headed arrows). (B) NL contents of HLM23 and two CRISPR/Cas-induced *NO06G03670* knockout mutants were increased similarly relative to the wild type. NL contents were measured by BODIPY staining and flow cytometry. Medians of the single cell BODIPY fluorescence distributions were normalized to the average wild type value. Crosses and error bars denote mean and SD ($N \geq 5$). Significance levels were assessed by Tukey's HSD test. (*): $p < 0.05$; (**): $p < 0.01$; (***): $p < 0.001$.

We examined the NL content of the confirmed *NO06G03670* knockout mutant strains pTKO-sp2-C3 and pTKO-sp2-C7. When grown at high light conditions, the strains had NL contents similar to that of HLM23, and significantly higher than that of wild type controls (Fig. 6B). This confirms that the high lipid phenotype of HLM23 is connected to the knockout of *NO06G03670*, and underpins the relevance of this gene as a control hub for growth-associated NL production in *N. oceanica*. The connection between genotype and phenotype of other HLMs could be investigated accordingly during follow-up research.

Knockout of *NO06G03670* causes upregulation of metabolic pathways that are associated with fatty acid biosynthesis

In a computational study involving analysis of conserved transcription factor binding site motifs, Hu and colleagues predicted *NO06G03670* to attenuate the expression of genes related to fatty acid biosynthesis (FAB) and long chain fatty acyl-CoA synthetases (LC-FACS). LC-FACS activate fatty acids for different metabolic processes such as trafficking, lipid assembly or beta oxidation [341]. In order to understand why the knockout of

NO06G03670 leads to an increase in lipid content, we analyzed gene expression in HLM23 by mRNA sequencing technology. Out of 9578 expressed genes, 1530 and 1495 genes were transcriptionally up- and downregulated, respectively, in HLM23, relative to the wild type (Fig. S4).

Fatty acid biosynthesis genes are transcriptionally upregulated in HLM23

Plants and microalgae possess a prokaryotically-derived type II FAB pathway in their chloroplasts, which utilizes discrete, monofunctional enzymes to synthesize straight-chain fatty acids (Fig. 7). Out of 12 FAB genes that putatively encode these monofunctional enzymes, 10 were transcriptionally upregulated ($p_{\text{adjust}} < 0.05$) in HLM23 and two were not differentially expressed, compared to the wild type (Fig. 7). Most of the upregulated proteins were predicted to contain a chloroplast targeting peptide (cTP) or signal peptide (SP) by one or multiple *in silico* prediction tools and may thus localize in the plastid. We also identified two acyl carrier protein (ACP) genes in *N. oceanica*, *NO29G00550* and *NO30G00840*, predicted to localize in the chloroplast and mitochondrion respectively. As carriers of the growing acyl chains, ACPs fulfill a central role in FAB and it has been shown that they are much stronger expressed than other FAB-related genes in plants [342]. In HLM23, *NO30G00840* transcript abundance was >10-times higher than that of other FAB-associated genes and 1.22- \log_2 fold ($p = 6.37 \times 10^{-12}$) higher than in the wild type (Tab. S2), ranking the gene at position 28 of the strongest expressed genes of the entire transcriptome dataset. Expression of the putative mitochondrial ACP isoform *NO29G00550* was similar in HLM23 and the wild type with much lower expression than the likely chloroplastic counterpart.

The first committed and rate-limiting step of FAB is the carboxylation of acetyl-CoA, carried out by acetyl-CoA carboxylase (ACCase, EC# 6.4.1.2) [343]. We identified two isoforms for ACCase: *NO21G02220* and *NO09G00220*. In particular, isoform *NO09G00220* shows high similarity with ACCases from other microalgae. Moreover, it had the highest transcript abundance among all FAB-associated enzymes in the wild type strain and it showed strong upregulation in HLM23, with transcript levels being 1.72- \log_2 fold ($p = 5.21 \times 10^{-18}$) increased relative to the wild type. Expression of putative chloroplastic candidates for the other FAB enzymes was moderately to strongly increased in HLM23, ranging from a 0.66- \log_2 fold ($p = 2.16 \times 10^{-3}$) increase for beta-hydroxyacyl-[acyl-carrier-protein] dehydratase (EC# 4.2.1.59) to a 1.59- \log_2 fold ($p = 5.14 \times 10^{-17}$) increase for beta-ketoacyl-[acyl-carrier-protein] synthase II (EC# 2.3.1.41/2.3.1.179).

The primary end products of plastidial FAB are C16:0 and C18:0 fatty acyl-ACPs [341], which can further be desaturated by stearoyl-ACP $\Delta 9$ desaturases. In *Nannochloropsis*, only a single $\Delta 9$ desaturase, *NO02G01510*, has been identified. *NO02G01510* expression was decreased by 173% ($p = 1.43 \times 10^{-3}$) in HLM23 compared to the wild type, which might explain the decreased fraction of MUFA in the neutral lipids of this mutant. In a previous study, Dong and co-workers found decreased *NO02G01510* protein levels during N-starvation [344], but other studies have shown that the gene is upregulated at the transcriptional level during N-stress [74, 345, 346].

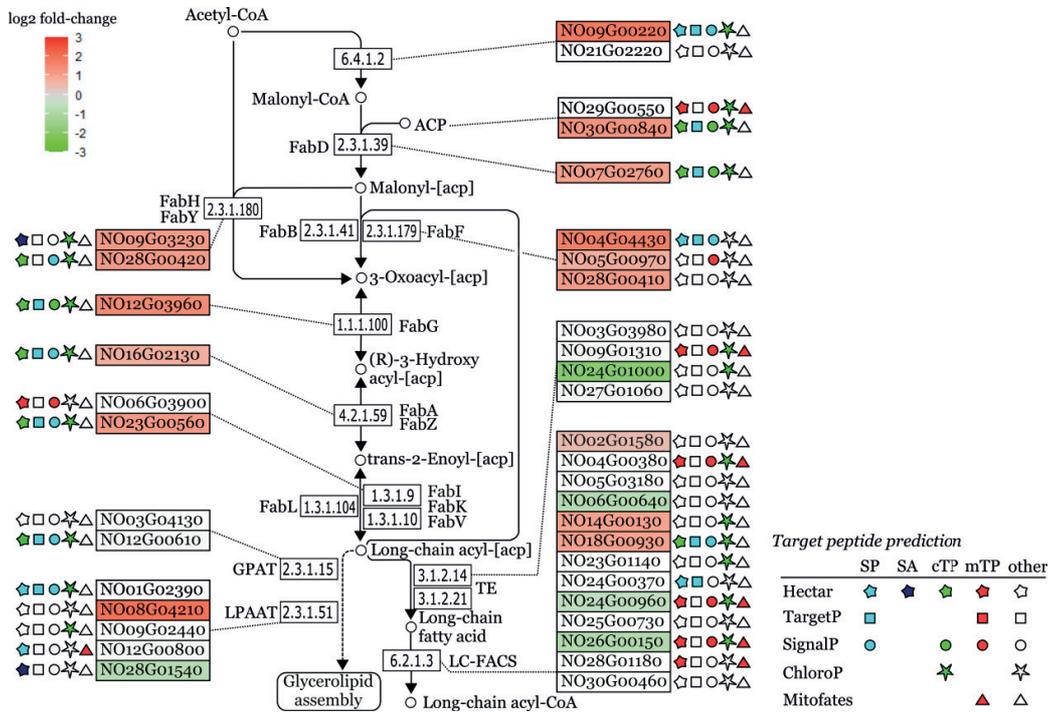


Figure 7. Differential expression of genes encoding putative fatty acid biosynthesis enzymes in HLM23. The pathway map illustrates the enzymatic conversion of acetyl-CoA to fatty acids, and transfer of fatty acyl chains to glycerol-3-phosphate or coenzyme A. The genes encoding the enzymes that putatively catalyze these reactions in *N. oceanica* are shown in solid boxes, with the exception of NO29G00550 and NO30G00840 which encode putative ACPs. Genes with a significant ($p_{adjust} < 0.05$) differential expression in HLM23 relative to the wild type are colored according to the \log_2 fold change in the mutant strain. Symbols next to boxes represent the subcellular localization, predicted by five in silico tools. Almost all genes putatively involved in fatty acid biosynthesis were transcriptionally upregulated in HLM23. Most of these genes encode proteins that may contain a chloroplast targeting peptide (cTP) or a signal peptide (SP), suggesting that plastidic fatty acid biosynthesis is enhanced at the transcriptional level in HLM23. TE: thioesterase; SP: signal peptide; SA: signal anchor; cTP: chloroplast targeting peptide; mTP: mitochondrion targeting peptide. The pathway map was adapted from the KEGG PATHWAY database (Kanehisa and Goto 2000).

Fatty acyl-ACPs produced in FAB can directly be used by chloroplastic isoforms of glycerol-3-phosphate acyltransferase (GPAT, EC# 2.3.1.15) or lysophosphatidic acid acyltransferase (LPAAT, EC# 2.3.1.51) for glycerolipid assembly [347]. The putative GPAT enzymes of *N. oceanica* were not differentially expressed in HLM23, and expression of LPAAT isoforms NO28G01540 was decreased by 170% compared to the wild type ($p=1.25e-02$). Expression of NO08G04210, which is the only *Nannochloropsis* LPAAT with no homologue in *A. thaliana* [74], was increased by 1.93- \log_2 fold ($p=2.13e-09$).

In addition to GPAT and LPAAT, another enzyme that processes fatty acyl-ACPs is thioesterase, which hydrolyzes the thioester bond between ACP and the fatty acyl chain (EC# 3.1.2.21/3.1.2.14), releasing free fatty acids. Four *N. oceanica* enzymes (NO03G03980, NO09G01310, NO24G01000 and NO27G01060) are predicted to contain a thioesterase domain, but only the putatively mitochondrial NO09G01310 shows a high degree of similarity to known thioesterases. None of the four proteins were predicted to contain a SP or cTP by more than one prediction tool. Therefore, the plastidic thioesterase isoform of *N. oceanica* needs to be identified by biochemical analysis.

Free fatty acids are converted to acyl-CoA at the outer plastid envelope and in other subcellular compartments by long chain fatty acyl-CoA synthetase (LC-FACS) [348]. Out of the 13 putative LC-FACs of *N. oceanica*, three were up- and three were downregulated in HLM23. In particular, *NO18G00930* transcript was increased 0.94- \log_2 fold ($p=3.27e-04$) with a predicted chloroplastic localization. *NO14G00130* transcript levels were 0.95- \log_2 fold ($p=6.78e-04$) higher in HLM23 than in the wild type, however, subcellular localization and metabolic function of *NO14G00130* are unclear. Previously, Li and co-workers have shown that expression of *NO14G00130* gradually increases during N-starvation experiments, as well as for another LC-FACS isoform-encoding gene, *NO05G03180* [74]. However, in HLM23, *NO05G03180* was not differentially expressed compared to the wild type. *NO05G03180* is predicted to localize in the peroxisome, thus it may be responsible for the activation of fatty acids for beta oxidation.

After activation by LC-FACS, fatty acyl-CoA is sequestered by acyl-CoA binding proteins (ACBPs), that aid in transport to the ER for fatty acid modifications and lipid assembly [349]. Out of the four putative ACBPs of *N. oceanica* (*NO16G02490*, *NO19G00260*, *NO23G00150*, *NO24G01960*), *NO19G00260* and *NO24G01960* transcript levels were similar in HLM23 and the wild type (Tab. S5), while *NO23G00150* expression was undetectable, which matches previous reports [74]. *NO16G02490* expression was increased 1.63- \log_2 fold ($p=3.55e-06$) in HLM23, putatively implicating this ACBP isoform in sequestration and transport of excess fatty acyl-CoA produced by the mutant.

Supply of fatty acid building blocks may be enhanced in HLM23

The transcriptional upregulation of multiple chloroplastic FAB proteins including ACCase in HLM23 may well explain the increased lipid content observed for this mutant. To meet an increased demand of FAB substrates, the key fatty acid building block acetyl-CoA needs to be supplied at sufficient rates. The photosynthetically produced glyceraldehyde-3-phosphate (G3P), which is the primary product of the Calvin-Benson-Bassham (CBB) cycle, can either be used for production of hexoses through gluconeogenesis, or it can be converted to pyruvate and then acetyl-CoA through the "lower part" of glycolysis (i.e. the reactions from G3P to pyruvate). We mapped all the enzymes putatively involved in this part of central carbon metabolism (Fig. 8).

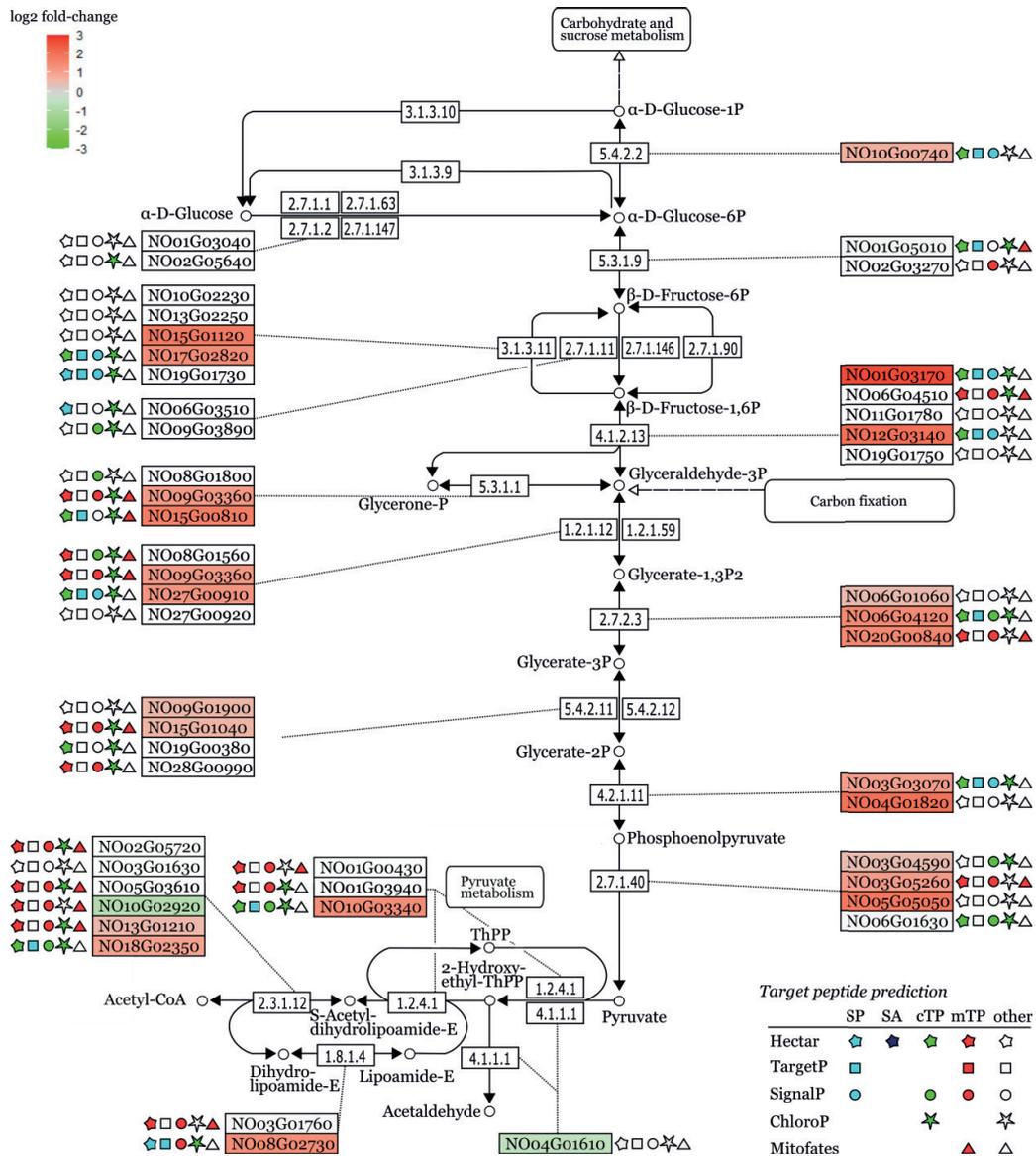


Figure 8. Differential expression of genes encoding enzymes putatively involved in glycolysis, gluconeogenesis and pyruvate metabolism in HLM23. The schematic illustrates the main reactions of the Embden-Meyerhof-Parnas (EMP) pathway, and of pyruvate metabolism. Genes involved in both pathways were either transcriptionally upregulated or not differentially expressed in HLM23 relative to the wild type, with the exceptions of NO04G01610 and NO10G02920. Among the upregulated genes, subcellular localization is predominantly predicted to be chloroplastic. Most EMP pathway enzymes can catalyze reactions in either the glycolytic or gluconeogenic direction. The pathway map was adapted from the KEGG PATHWAY database (Kanehisa and Goto 2000).

HLM23 showed transcriptional upregulation for multiple isoforms of all enzymes that participate in the conversion of G3P to pyruvate (Fig. 8; Tab. S3). In *N. oceanica*, the lower part of glycolysis is likely present in the cytosol, chloroplast and mitochondrion, whereas the upper part may only be fully functional in the chloroplast [75]. Enzymes of the lower part of glycolysis were previously shown to be co-regulated with fatty acid biosynthesis genes in *N. oceanica*, suggesting functional cooperation [75]. The rate-limiting step of the lower part of glycolysis is catalyzed by glyceraldehyde 3-phosphate dehydrogenase (GAPDH, EC#1.2.1.12) [350, 351], which has four predicted isoforms in *N. oceanica*. Among them, the putative mitochondrial isoform *NO09G03360* and the putative chloroplastic isoform *NO27G00910* had a 1.1- \log_2 fold ($p=9.52e-08$) and 1.17- \log_2 fold ($p=3.22e-08$) increased transcript abundance, respectively. In agreement with this, several isoforms of all other enzymes involved in the lower part of glycolysis were also upregulated in HLM23. For each enzyme, at least one of the upregulated isoforms was predicted to localize in the chloroplast by one or multiple prediction tools, with the exception of phosphoglycerate mutase (PGM, EC# 5.4.2.11). At the same time, previous studies have shown that the reaction catalyzed by PGM does not pose a bottleneck for glycolysis among various organisms [351].

Glycolysis and FAB are bridged by a multi-enzyme complex called the pyruvate dehydrogenase complex (PDHC) that converts pyruvate into acetyl-CoA. PDHC consists of the subunits E1 (EC# 1.2.4.1), E2 (EC# 2.3.1.12) and E3 (EC# 1.8.4.1) which are present in the mitochondrion and chloroplast of *N. oceanica* (Fig. 8). The three putative chloroplastic E1, E2 and E3 subunits *NO10G03340*, *NO18G02350* and *NO08G02730* were upregulated 1.32- \log_2 fold ($p=3.11e-14$), 0.94- \log_2 fold ($p=5.92e-06$) and 1.3- \log_2 fold ($p=3.47e-09$), respectively, in HLM23. Expression of putative mitochondrial PDHC subunits, on the other hand, was mainly unchanged. The mitochondrial PDHC provides acetyl-CoA not only as substrate for mitochondrial FAB, but more importantly for the citric acid cycle. Pyruvate can further be converted to acetyl-CoA via the PDHC-bypass. However, the highest expressed acetyl-CoA synthetase of this pathway lacks co-expression with FAB genes during diel cycles [75], suggesting that the PDHC-bypass likely does not contribute to acetyl-CoA supply for lipid biosynthesis in *N. oceanica*. The single copy pyruvate decarboxylase (PDC, EC# 4.1.1.1) *NO04G01610* that channels pyruvate into this bypass is transcriptionally downregulated in HLM23 ($p=2.77e-02$), suggesting that this pathway is not responsible for the increased fatty acid content of the mutant. Expression of putative isoforms of the other PDHC-bypass enzymes aldehyde dehydrogenase and acetate-CoA ligase was also decreased in HLM23 (Tab. S3).

(EC# 2.7.1.19) in HLM23 (Fig. 9, Tab. S4). Expression levels of ribulose 1,5-bisphosphate carboxylase (RuBisCO, EC# 4.1.1.39), a key enzyme of the CBB cycle, were not assessable in our experiments because in *Nannochloropsis* both enzyme subunits are encoded in the chloroplast genome. As chloroplast transcripts do not contain poly(A) tails, the RuBisCO transcripts were not analyzed during mRNA sequencing. Transcription of a CbbX protein homolog, which is an activator of red-type RuBisCO protein [353], was increased 0.56- \log_2 fold ($p=2.24e-02$) in HLM23.

The upregulation of genes encoding enzymes involved in the CBB cycle, central carbon metabolism and chloroplastic FAB in HLM23 may be related to this mutant's altered ability to cope with higher light irradiation ($600 \mu\text{mol m}^{-2} \text{s}^{-1}$), which was reflected in an increased F_v/F_m under these conditions (Fig. 3C). Accordingly, when cultures were grown at lower irradiance conditions (150 or $350 \mu\text{mol m}^{-2} \text{s}^{-1}$), F_v/F_m of HLM23 and the wild type were similar and the differences in NL contents were less pronounced (Fig. S5). Previously, Yang and colleagues have shown that overexpression of chloroplastic FBPase in the green microalga *Chlorella vulgaris* caused an increase in photosynthetic quantum yield and oxygen evolution rates [354], showcasing a connection between expression levels of CBB cycle enzymes and photosynthetic performance. Future studies might shed light on the ability of HLM23 to cope with the adverse effects associated with high light conditions, and if there is a link to the transcriptional upregulation of metabolic pathways that directly utilize photosynthetic assimilates.

In line with the elevated F_v/F_m , HLM23 showed upregulation of several proteins that are directly or indirectly associated with the light reactions of photosynthesis (Tab. S5). Transcript levels of NO24G02290, encoding a chlorophyll a-b binding protein of the light harvesting complex, were increased 0.46- \log_2 fold ($p=4.29e-02$) in HLM23 compared to the wild type. Expression of four out of six putative ferredoxin proteins was increased 1.1-3.45- \log_2 fold ($p<8.78e-04$) and the only known ferredoxin NADP(+) oxidoreductase was upregulated 1.15- \log_2 fold ($p=3.07e-11$) in HLM23 (Tab. S5). Additionally, the only isoform of the highly expressed PSII manganese-stabilizing protein PsbO was upregulated 0.89- \log_2 fold ($p=2.09e-07$). PsbO aids in rapid turnovers of the oxygen evolving reactions [355], and is required for high photosynthetic quantum yields in *A. thaliana* [356]. Expression of two magnesium chelatase (EC# 6.6.1.1) subunits ChlD and ChlH was increased 1.31- \log_2 fold ($p=3.54e-07$) and 2.02- \log_2 fold ($p=7.44e-25$) respectively in HLM23 (Tab. S5). Magnesium chelatase catalyzes the insertion of a magnesium ion into protoporphyrin IX, which is the first committed step of chlorophyll biosynthesis (Masuda 2008). The third subunit, ChlI is encoded in the chloroplast genome of *N. oceanica*, and its expression was therefore not quantified.

Intriguingly, a putative glucose-6-phosphate/ P_i translocator (GPT) was upregulated 3.54- \log_2 fold ($p=7.56e-06$) in HLM23 (Tab. S5). In *A. thaliana*, a GPT knockout is associated with decreased starch synthesis under high light conditions (Dyson et al. 2015). It was suggested that high cytosolic triose-phosphate levels can induce expression of GPT, which promotes the import of glucose-6-phosphate (G6P) from the cytosol to the chloroplast (Weise et al. 2019). Imported G6P can then fuel either carbohydrate synthesis or the chloroplastic G6P shunt, which corresponds to the oxidative phase of the pentose phosphate pathway and

may have a stabilizing effect on photosynthesis [357]. The function of GPT in microalgae, however, remains obscure. Chloroplastic isoforms of the G6P shunt enzymes, such as 6-phosphogluconolactonase, were transcriptionally downregulated by up to -1.03-log_2 fold ($p=1.12\text{e-}03$) in HLM23, suggesting that the upregulation of GPT in this mutant has a different reason than contributing to activity of this pathway.

Lastly, it should be noted that multiple putative and confirmed transcriptional regulators were differentially expressed in HLM23. For instance, expression of the NobZIP1 transcription factor NO19G01720 was increased 1.05-log_2 fold ($p=3.16\text{e-}07$). Recently, Li and co-workers found that an *N. oceanica* mutant overexpressing a NO19G01720 homologue had an increased lipid content, which the authors linked to a decreased expression of UDP-glucose dehydrogenase (UGDH), a key enzyme in cell wall carbohydrate polymer metabolism [358]. However, despite the upregulation of NO19G01720, expression of the UGDH gene NO14G01750 was not decreased, but increased 0.7-log_2 fold ($p=4.46\text{e-}03$) in HLM23, suggesting that NO14G01750 expression could be modulated by multiple transcriptional regulators and not only by NO19G01720. This also indicates that lipid contents of HLM23 could potentially be further improved by silencing the gene expression of UGDH in this strain, for instance through RNA interference, as previously demonstrated for wild type *N. oceanica* [358].

In conclusion, the knockout of the putative APETALA2-like transcription factor NO06G03670 in HLM23 resulted in increased NL content and F_v/F_m , without affecting growth rate compared to the wild type. This phenotype was accompanied by a transcriptional upregulation of key components of several chloroplast-located metabolic pathways, including sugar metabolism, the conversion of pyruvate to acetyl-CoA, fatty acid biosynthesis and the CBB cycle. However, the exact cause of the alteration in the lipid content of *N. oceanica* due to the NO06G03670 knockout remains to be elucidated. In this context, a loss-of function mutant of APETALA2 in *A. thaliana* showed changes in the ratio of hexose to sucrose during seed development, suggesting that APETALA2 may control seed mass by affecting sugar metabolism, in plants [337]. Because lipid biosynthesis from sugars and related substrates is a preferentially anabolic activity in the oleaginous *N. oceanica* [76], NO06G03670 may be involved in directly or indirectly regulating sugar metabolism and trafficking, which in turn could affect photosynthetic carbon assimilation and partitioning, and consequently lipid content. In this regard, the increased F_v/F_m of HLM23 is intriguing and future studies are needed to elucidate the role of the APETALA2-like transcription factor in the regulatory network of *N. oceanica*. Engineering of microalgal strains with an increased photosynthetic efficiency is a difficult task that has thus far been met with limited success. However, commercialization of microalgal-derived bulk commodities will heavily rely on an increased photosynthetic performance of microalgal strains [42].

Conclusion

In this study, we developed a novel forward genetic screen for *N. oceanica* based on insertional mutagenesis and FACS. We isolated multiple lipid-overproducing mutant strains by iteratively sorting an insertional mutant library five times. An optimized genotyping methodology based on the type IIS restriction endonuclease Mmel allowed us to identify the causative mutations in five mutants, which showed 12-59% increased NL contents DCW⁻¹ relative to the wild type. The genes that were disrupted in these strains present potential targets for genetically engineering *Nannochloropsis* for an increased lipid productivity. Highest NL contents were found for strain HLM15, which is likely incapable of sustaining protein homeostasis due to a disruption of a DnaJ protein. The 59% increase in NL DCW⁻¹ that we found for this mutant was accompanied by a 21% decrease in growth rate and susceptibility to high light. Two of the genes that were affected in the other mutant strains encode transcription factors, which highlights the importance of this type of protein as a master switch for metabolic regulation. One of these transcription factors, *NO09G01030*, is a homologue of a transcription factor of *N. gaditana* that was previously shown to be associated with lipid metabolism [80]. The other one, *NO06G03670*, was predicted to upregulate genes involved in protein homeostasis and downregulate chloroplastic genes associated with fatty acid biosynthesis and carbon flux to acetyl-CoA [334], which we confirmed by mRNA sequencing methodology. *NO06G03670* knockout strain HLM23 had a NL content of 30.2% DCW⁻¹ and increased photosynthetic quantum yield under favorable biomass production conditions, and growth was not impaired. Accordingly, projected NL productivities of one-step batch cultivations with HLM23 are 39% higher compared to productivities when using wild type *N. oceanica*. The insights gained in this study increase the genetic knowledge about an industrially attractive microalga and will be useful to formulate new hypotheses for the genetic engineering of *Nannochloropsis* strains with an increased NL productivity. While our findings significantly contribute to an improved understanding of *N. oceanica*, more studies are needed to elucidate the functions of the many genes that remain unknown. In this regard, our optimized Mmel-based genotyping methodology lays the foundation of future screenings for not only lipid accumulation but any desired phenotype.

Materials and Methods

Strains and cultivation conditions

The microalgal strain used in this study, *N. oceanica* IMET1 was a kind gift by prof. Jian Xu (Qingdao Institute for Bioenergy and Bioprocess Technology, Chinese Academy of Sciences). Microalgae were cultured in artificial sea water (ASW, 419.23 mM NaCl, 22.53 mM Na₂SO₄, 5.42 mM CaCl₂, 4.88 mM K₂SO₄, 48.21 mM MgCl₂ and 20 mM HEPES, pH 8), supplemented with 2 ml l⁻¹ of Nutribloom plus (Necton, Portugal) growth media (ASW-NB) at 25 °C. For maintenance purposes and after fluorescence-activated cell sorting steps, cultures were incubated on an orbital shaker operated at 90 rpm in a climate-controlled chamber, aerated with ambient air and illuminated with warm-white fluorescent light bulbs at 50 μmol m⁻² s⁻¹ light intensity ("LL" conditions). An HT Multitron Pro (Infors Benelux, Netherlands) orbital shaker unit with 0.2% CO₂-enriched air was used

for cultivation at $150 \mu\text{mol m}^{-2} \text{s}^{-1}$ illumination intensity ("ML" conditions) at 120 rpm shaking frequency. Isolated high lipid mutant (HLM) strains were cultivated in 50 ml shake flasks at $600 \mu\text{mol m}^{-2} \text{s}^{-1}$ ("HL" conditions) in an Algem HT24 photobioreactor (Algenuity, UK) that was placed inside an HT Multitron Pro operated at conditions as described above. All experiments were carried out with a 16:8 h diurnal light cycle.

For cultivation on solid medium, ASW was supplemented with 1% (w/v) of agar (Merck) before autoclaving, cooled to $60 \text{ }^\circ\text{C}$ and mixed with 2 ml l^{-1} of nutribloom plus and $5 \mu\text{g ml}^{-1}$ Zeocin before distribution into petri dishes for solidification. Algae-containing plates were incubated at $25 \text{ }^\circ\text{C}$ in ambient air under warm-white fluorescent light bulbs at an intensity of $60\text{--}80 \mu\text{mol m}^{-2} \text{s}^{-1}$ with a 16:8 h diurnal cycle.

Construction of insertion cassette DNA

The insertion cassette (IC) carried the bleomycin resistance gene *shble* (GenBank accession number A31898.1) from *Streptoalloteichus hindustanus*, under control of the strong endogenous promoter of the violaxanthin chlorophyll a binding protein precursor (VCP) gene, including the Kozak sequence and first genetic intron. Two transcriptional terminators belonging to the α -tubulin and *Clp* protease proteolytic subunit genes were added in head-to-head orientation downstream of *shble*, to safeguard transcriptional termination of genes at the insertion site regardless of IC orientation. Assembly of these elements into the pUC19 derivative plasmid pNIM14 was previously described [227]. Using pNIM14 as a template, we constructed pNIM21 by adding terminal MmeI restriction endonuclease recognition sequences to facilitate precise excision of the IC from genomic DNA during insertion site tracing. We further inserted ~ 100 nucleotides of artificial barcodes between these MmeI recognition sequences and the VCP promoter and *Clp* terminator elements respectively (Fig. 1A). This allowed us to design barcode-complementary cassette PCR primers for insertion site tracing, which reduced spurious primer annealing to endogenous genome sequences. The barcode sequences were inserted into pNIM14 as PCR primer overhangs and assembled using Gibson assembly technique (NEBuilder HiFi DNA Assembly Master Mix, NEB #E2621) with 22 nucleotides overlap. The first assembly fragment was PCR-amplified using oligonucleotides and with 1 ng of Mlsl-linearized pNIM14 as template and an annealing temperature (T_a) of $65 \text{ }^\circ\text{C}$. The second assembly fragment was PCR-amplified using oligonucleotides and with 1 ng of Scal-linearized pNIM14 as template and a T_a of $69 \text{ }^\circ\text{C}$. Both PCRs were carried out using Q5 DNA polymerase (NEB #M0492), $0.25 \mu\text{M}$ primers and the following cycling scheme: 30 s at $98 \text{ }^\circ\text{C}$, 30 cycles (8 s at $98 \text{ }^\circ\text{C}$, 10 s at T_a , 90 s at $72 \text{ }^\circ\text{C}$), 3 min at $72 \text{ }^\circ\text{C}$. Linearization of pNIM14 using Mlsl (Thermo Fisher Scientific #FD1214) or Scal (Thermo Fisher Scientific #FD0434) was carried out according to the manufacturer's instructions. PCR products were purified using spin columns (Zymo Research #D4004) and used in Gibson assembly according to the manufacturer's instructions. Correct assembly of pNIM21 was verified by PCR and sequencing. The IC was amplified from 1 ng SchI-linearized (Thermo Fisher Scientific #FD1374) pNIM21 with Q5 DNA polymerase, using $0.5 \mu\text{M}$ of 5'-GTTGGAATCTTTCAACACGCTAGG-3' and 5'-GTTGGATTTGATACTGTTGTTGTGG-3' with the following cycling scheme: 30 s at $98 \text{ }^\circ\text{C}$, 35 cycles (5 s at $98 \text{ }^\circ\text{C}$, 85 s at $72 \text{ }^\circ\text{C}$), 5 min at $72 \text{ }^\circ\text{C}$. The IC was purified (Thermo Fisher Scientific #K0832) according to the manufacturer's

instructions, using the *PCR cleanup* version of the protocol, with the exception that nuclease-free (NF) H₂O was used for elution of DNA from the column.

Transformation of *N. oceanica*

N. oceanica IMET1 was transformed using electroporation similar to the protocol described by Vieler and colleagues [63]. Briefly, microalgal cells were cultivated at ML conditions and harvested during mid-exponential growth stage by centrifugation (2,500 x g, 5 min, 4 °C). Cell pellets were washed thrice with 4 °C-cold 375 mM sorbitol and resuspended at 2.5 x 10⁹ cells ml⁻¹. 200 µl of cell suspension was mixed with 1 µg of IC DNA in pre-cooled 2 mm electroporation cuvettes and rapidly pulsed using a Bio-Rad GenePulser II, set to exponential pulse decay with 11 kV cm⁻¹ electric field strength, 50 µF capacitance and 600 Ω resistance. After the ~25 ms pulse, cells were immediately transferred to 5 ml of 22 °C ASW-NB and recovered for 24 h at dim light without agitation. Subsequently, cells were pelleted (2,500 x g, 10 min), resuspended in a small amount of supernatant and plated onto ASW-NB plates containing 5 µg ml⁻¹ Zeocin and incubated as described previously for 4 weeks, before plates were flushed with ASW-NB containing 5 µg ml⁻¹ of Zeocin, and pooled in a 250 ml shake flask, yielding the insertional mutant library.

Staining of intracellular neutral lipids with BODIPY

Neutral lipids (NLs) were visualized for flow cytometry and FACS using the fluorescent dye BODIPY 505/515 (Invitrogen, #D3921) as previously described [261]. Microalgae were stained at an OD₇₅₀ of 0.3–0.5 with 6% (v/v) of DMSO and 1.2 µg ml⁻¹ BODIPY 505/515. After addition of the dye, cell suspensions were vortexed for 5 s and incubated in the dark for 15 min before being subjected to flow cytometry/FACS.

Flow cytometry analysis

BODIPY fluorescence emission was quantified by flow cytometry analysis using an SH800S (Sony Biotechnology, USA) cell sorter, equipped with a 488 nm laser and a 70 µm nozzle microfluidic chip. Detector wavelengths for different channels were: 488 nm (forward scatter, gain 2); 488 nm (side scatter, gain 22%); 510 ± 10 nm (BODIPY, gain 29%); 720 ± 30 nm (chlorophyll a autofluorescence, gain 40%). For standard fluorescence analyses, a minimum of 50,000 events were screened per sample. Gating was applied as previously described [261] to remove background noise. For statistical data treatment, only events in the final gate (“living cells”) were considered.

FACS sorting for enriching high lipid mutants in the mutant library

Cell sorting was carried out using the same instrument and operational settings as during regular flow cytometry analysis. Cells were sorted directly into sterile 250 ml Erlenmeyer flasks, filled with 50 ml ASW-NB. Sorting was terminated at 300,000 sorted events in ‘Ultra purity’ sort mode of the SH800S software. Gating was applied as indicated previously [261] to filter out non-vital cells and background noise, with an additional final gate for high-NL cells (Fig. S1). Sorted cultures were recovered under LL conditions for 10 days (d). New cultures were inoculated to an OD₇₅₀ of 0.1 and acclimated to ML conditions for 3 d. Finally, cultures were diluted to an OD₇₅₀ of 0.1, incubated for 2 more days under ML conditions and subjected to the next round of sorting. ML conditions were employed during HLM enrichment and isolation by FACS, because they are sufficient to induce NL accumulation,

but less stressful for cells than HL conditions. Therefore, using ML conditions prevented loss of high-light sensitive mutant strains during the library enrichment for HLMs.

Strain isolation from enriched libraries using FACS

After 4 rounds of sorting, HLM-enriched libraries were recovered under LL conditions, set to an OD_{750} of 0.1 and acclimated to ML conditions for 3 d. New cultures were inoculated to an OD_{750} of 0.1 and grown under ML conditions for an additional 2 d before strains were isolated from the pooled libraries by depositing single cells onto agar plates using a high-NL gate as described above, and the 'Single cell' sort mode of the SH800S software.

Characterization of isolated HLM strains

Four weeks after HLM strain isolation by FACS, all colonies showed identical morphology and comparable colony diameter on agar. Therefore, the first 17 viable colonies were picked from agar plates and expanded to shake flasks, which were incubated at LL conditions in duplicates. After ~1 week, fresh cultures were inoculated to an OD_{750} of 0.07, and incubated in an Algem HT24 photobioreactor (Algenuity, UK) operated at $350 \mu\text{mol m}^{-2} \text{s}^{-1}$ for 3 d. Subsequently, cultures were diluted to an OD_{750} of 0.07 and incubated at $600 \mu\text{mol m}^{-2} \text{s}^{-1}$ for 14 d. Growth curves were obtained by measuring the OD_{750} daily, 1 h before onset of the dark phase of the diurnal light cycle. Maximum specific growth rate μ_{max}^{750} was obtained as the slope of a linear model that was fitted on logarithmically-transformed OD_{750} values, obtained within the first 72 h of the cultivation. For biochemical characterization, cultures were harvested 1 h before onset of the dark phase on the second day of cultivation at $600 \mu\text{mol m}^{-2} \text{s}^{-1}$. HL conditions were employed to improve translatability of findings to outdoor production systems in future studies and to maximize the relevance of findings. Microalgal suspensions were centrifuged ($4,000 \times g$, 10 min) and resuspended in 2 ml of ASW and 600 μl of the suspension was subjected to lipid quantification as described below. The dry cell weight (DCW) concentration of the suspension was measured as previously described (Zhu and Lee 1997) with 0.5 M ammonium formate as wash buffer. Cell count of the suspension was determined using a Beckman Coulter Multisizer 3 (Beckman Coulter Inc., USA, 50 μm orifice).

Assessment of photosynthetic performance

The maximum efficiency of photosystem II photochemistry was assessed using *in vivo* chlorophyll fluorescence with an AquaPen AP 100-C (Photon System Instruments, Czech Republic) handheld fluorometer. Samples were dark adapted for 20 min prior to analysis according to the manufacturer's protocol.

Identification of cassette insertion sites

Cassette PCR was carried out employing a modified version of procedures reported by Goodman and co-workers (Goodman et al. 2009) and Zhang and co-workers [278]. As described above, the IC carried terminal recognition sites for the type IIS restriction enzyme MmeI. MmeI induces a 2 nucleotide staggered-end cut 20-21 nucleotides outside of its recognition site. The recognition sites were oriented in a way that the MmeI would cut in the flanking genomic DNA of transformant strains, facilitating 'excision' of the ends of the IC together with 20-21 nucleotides of flanking sequences. For this, genomic DNA was extracted from exponentially growing cultures. Briefly, $\sim 6 \times 10^7$ cells were pelleted

(5000 × g, 3 min) and resuspended in 150 µl of lysis buffer as described by Daboussi and colleagues [48]. The suspension was vortexed for 60 s at high speed, incubated overnight at 22 °C, again vortexed for 60 s and pelleted (5000 × g, 5 min). The supernatant was purified using the FavorPrep column purification kit (Bio-Connect B.V. #FAGDC001) according to the manufacturer's instructions, except that NF H₂O was used for eluting DNA from the column. DNA concentration was determined using a NanoDrop device, and 350 ng was digested with 0.16 µl Mmel (NEB #R0637) at 37 °C for 10 min according to the manufacturer's instructions. Integrity of genomic DNA and successful digestion were verified by agarose gel electrophoresis. Double stranded DNA adapters were assembled by mixing 100 µM stock solutions of oligonucleotides and 5'-P-CTCCACGTTACCC-NH₃-3' (5'-phosphate-modified, 3'-amine modified), heating the mixture for 5 min at 93 °C followed by slow cooling to 22 °C (~1 h). The bottom strand of the adapter was designed with a 5'-phosphate and 3'-NH₃ modification to increase ligation efficiency and to prevent unspecific amplification during the PCR steps. The top strand was designed to have a NN-3' overhang after adapter formation to facilitate annealing to the NN-3' overhang of Mmel-digested DNA fragments. Roughly 13 µg of Mmel-digested DNA was mixed with 4.8 µl of 50 µM adapters and subjected to ligation using 2.5 U of T4 DNA ligase (Thermo Fisher Scientific #EL0014) in an 8 µl reaction for 60 min at 22 °C. The reaction was stopped by heat inactivation of the enzyme and the solution was diluted 1:5 with NF H₂O. To amplify the target sequence over the genomic background, 1 µl of the mixture was used as template for nested PCR (2 rounds total) with Taq polymerase (Thermo Fisher Scientific #K1081) using 0.25 µM of primers specific for the adapter and for either the 5'-end or the 3'-end of the IC respectively (two separate nested PCRs per strain). The first PCR iteration was carried out using oligonucleotide 5'-TAGGCAGGCGGGCACCTCGCGTTAGTG-3' together with either 5'-CGGAGATGCGTACCGTAAGGTGGCATTAGC-3' or 5'-CTCGGCTGGTATCTGAGGAGTAGCCACAC-3' in the reactions specific for the 5'- and 3'-end of the IC respectively. For increased specificity, PCRs were run with touchdown settings, as follows: 5 min at 95 °C, 7 cycles (30 s at 95 °C, 45 s at 72 °C), 32 cycles (30 s at 95 °C, 15 s at T_a , 30 s at 72 °C), 10 min at 72 °C. T_a was 69.2 and 70.6 °C for the 5'- and 3'-end specific reactions respectively. PCR products were diluted 1:50 with NF H₂O and subjected to a second round of PCR with nested primers, namely 5'-GTTAGTGCTGGGTCTAGGCGCTCTGG-3' together with either 5'-AGCCAGACATCATCCATCGCCTCTGATCG-3' or 5'-CGAGGAAGTAGACTGTTGCACGTTGGCGAT-3' for the 5'- and 3'-end specific reactions respectively. Cycling schemes were similar to the first round of PCR, but cycle number was 28 instead of 32. T_a during these 28 cycles was 70.3 and 68.9 °C for the 5'- and 3'-end specific reactions respectively. PCR products were purified and sequenced in order to reveal the nucleotide sequence surrounding the IC. The genomic position of IC integration was then determined by using the flanking genomic nucleotide sequences in BLASTn queries against the *N. oceanica* IMET1 genome version 2 [330], using standard settings, except changing the E value to 0.01. The overall procedure is summarized in Figure S3.

Quantification of neutral and polar lipids via GC-FID

Neutral and polar lipid contents were determined using a modified version of the protocol described by Remmers et al. [359]. 600 µl of an algal suspension with known cell concentration and DCW concentration was subjected to fatty acid extraction, separation into neutral and polar acyl lipids, methylation and quantification. The

suspension was freeze-dried in a lysing matrix (#116914050-CF, MP Biomedicals) and subjected to mechanical cell disruption and lipid extraction with a chloroform:methanol (1:1.25) solution containing the 2 internal standards tripentadecanoin (T4257; Sigma-Aldrich) and 1,2-didecanoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (840434, Avanti Polar Lipids Inc.) at a known concentration. Polar and apolar lipids were separated using Sep-Pak Vac silica cartridges (6 cc, 1000 mg; Waters). Neutral lipids were eluted with a hexane:diethylether (7:1 v/v) solution, and polar lipids using methanol:acetone:hexane (2:2:1 v/v). Solvents were evaporated under N₂ gas and lipid fractions were methylated in 5% (v/v) H₂SO₄-containing MeOH (1 h, 100 °C), extracted with hexane and analyzed by gas chromatography (GC-FID). Fatty acids were quantified based on the relative responses of individual fatty acids compared to the signal of the internal standard and normalized to the DCW or cell concentration.

Targeted knockout of *NO06G03670*

Targeted gene knockouts were produced using CRISPR-Cas technology as previously described [79]. Three plasmids pTKO-sp1, pTKO-sp2 and pTKO-sp3 were generated that encode an FnCas12a variant [102] and three different CRISPR RNAs (crRNAs) crRNA1, crRNA2 and crRNA3, containing three different protospacer sequences (sp1, sp2 and sp3, respectively), targeting the first exon of the gene *NO06G03670* (Fig. 6A). The FnCas12a was obtained from plasmid pY004 (pcDNA3.1-hFnCpf1) which was a gift from Feng Zhang (Addgene plasmid # 69976 ; <http://n2t.net/addgene:69976> ; RRID:Addgene_69976). The protospacers were designed using CHOPCHOP v3 [360], selecting the software options *CRISPR/Cpf1* or *CasX*, and *sgRNA length without PAM: 25*, and the 5'-PAM sequence TTTN. Targets that were predicted as highly efficient were curated to select an appropriately positioned 5'-PAM sequence TTTV (V = A/C/G), and restriction-cloned into the vector pNAD_Bael as follows. The vector (2 µg) was digested for 60 min using 10 U Bael (NEB #R0613) according to the manufacturer's instructions, and the linear product was purified from a 0.5% (w/v) agarose gel after 90 min of electrophoresis at 100 V (Thermo Fisher Scientific #K0832). Inserts containing sp1, sp2 and sp3 sequences were synthesized as single-stranded oligonucleotides and annealed by mixing 100 µM solutions in equal volumes, heating at 95 °C for 5 min, and slowly cooling the suspension to 22 °C (~1 h). The pTKO-sp1 insert was created by annealing 5'-ACCACCCTCCCCCTTGACGTTACTTGGCCG-3' and 5'-AAGTAACGTC AAGGGGAGGGTGGTATCTA-3', the pTKO-sp2 insert was created by annealing 5'-CCCTTCCATTTGACCAACCAAAAATGGCCG-3' and 5'-ATTTTTGGTTGGTCAAATGGAAGGGATCTA-3', and the pTKO-sp3 insert was created by annealing 5'-GTTGAGGCGCATCGGGTGATTGCCTGGCCG-3' and 5'-AGGCAATCACCCGATGCGCCTCAACATCTA-3'. Then, 40 ng of linearized vector was ligated with 50 pmol of double-stranded insert using 1.25 Weiss U of T4 DNA ligase (Thermo Fisher Scientific #EL0011) at 22 °C for 60 min. Correct assembly of recombinant plasmids was verified by sequencing.

The derived plasmids pTKO-sp1, pTKO-sp2 and pTKO-sp3 were transformed into wild type *N. oceanica* as described above, but using 3 µg of circular plasmid instead of linear DNA, and in the presence of 30 µg denatured salmon sperm DNA (Thermo Fisher Scientific #AM9680) per transformation reaction before applying the electroporation pulse. The electroporation mixture was plated on Zeocin-containing agar for 4 weeks, after which

transformants were re-streaked on fresh Zeocin-containing agar. After 1 week, colonies were transferred to 48 well microtiter plates with Zeocin-containing ASW-NB medium. After 1 week at ML conditions, 1.5 μ l of culture was used as template in 30 μ l PCR reactions, using Phire polymerase (Thermo Fisher Scientific #F160) with the following thermocycling conditions: 5 min at 98 °C, 35 cycles (5 s at 98 °C, 5 s at 61.2 °C, 60 s at 72 °C), and 3 min at 72 °C. Oligonucleotides 5'-ACTTCACAAGGTATCTCACCTCATC-3' and 5'-CTAAGGGGAGTGGAAGAAAAAG-3' were used as PCR primers at 0.25 μ M. PCR products were purified and analyzed using Sanger sequencing at MacroGen Europe B.V. (Amsterdam, Netherlands).

mRNA sequencing

For transcriptomic analysis, microalgal cultures were inoculated in ASW-NB at an OD_{750} of 0.1 and cultivated in 50 ml shake flasks at $600 \mu\text{mol m}^{-2} \text{s}^{-1}$ in an Algem HT24 photobioreactor (Algenuity, UK) operated as described in section 3.1. After 3 d, 22 ml of culture was harvested (4,255 x g, 10 min, 4 °C) and cell pellets were snap-frozen. RNA extraction, mRNA library preparation and sequencing, transcript mapping and quantification were performed by Novogene (UK) Company Limited (N = 3). RNA was extracted from cell pellets using the RNeasy Plus Universal Mini Kit (Qiagen #73404) according to the manufacturer's instructions. RNA purity was verified on 1% agarose gels and using a NanoPhotometer spectrophotometer (IMPLEN, CA, USA). RNA integrity and quantity was assessed using 1% agarose gels and the Bioanalyzer 2100 RNA 6000 Nano Kit (Agilent Technologies, CA, USA). A total of 1 μ g RNA was used for RNA-seq library preparation with NEBNext Ultra™ RNA Library Prep Kit for Illumina (NEB, USA) following manufacturer's instructions. Briefly, poly-T oligonucleotides attached to magnetic beads were used to purify mRNA from total RNA samples. mRNA was fragmented at elevated temperatures in NEBNext First Strand Synthesis Reaction Buffer (5X) and first strand cDNA was reverse transcribed using random hexamers and M-MuLV reverse transcriptase (RNase H-). Second strand cDNA was synthesized using DNA Polymerase I and RNase H. Blunt ends were created by exonuclease/polymerase treatment. 3'-DNA ends were adenylated to allow for ligation of NEBNext adaptors, containing a hairpin loop structure. Fragments of ~150-200 nucleotides length were purified using the AMPure XP system (Beckman Coulter, Beverly, USA). Hairpin loops were opened by treatment of size-selected cDNA fragments with 3 μ l USER enzyme (NEB, USA) at 37 °C for 15 min, followed by heat inactivation of the enzyme at 95 °C for 5 min. Linear double-stranded cDNA fragments were subjected to PCR using Phusion High-Fidelity DNA polymerase with Index primers or Universal PCR primers provided for the NEBNext Ultra™ RNA Library Prep system. Barcoded PCR products were purified using the AMPure XP system, and the library quality was checked using an Agilent Bioanalyzer 2100. Index-coded samples were clustered using a cBot Cluster Generation System with the PE Cluster Kit cBot-HS (Illumina) according to the manufacturer's recommendations. After this, libraries were sequenced on an Illumina Novaseq 6000 platform and paired-end reads were generated. FASTQ raw reads were preprocessed using fastp [361] to remove reads of low quality and reads containing poly-N or adapter sequences.

The *N. oceanica* IMET1v2 reference genome and gene model annotation were retrieved from <ftp://nandesyn.single-cell.cn/pub/> [330]. Cleaned paired-end reads were mapped to the reference genome using HISAT2 [362]. Mapped reads were quantified using the

featureCounts package [363] of R statistical software (R Core Team 2018). FPKM values were calculated for gene count comparisons within a sample. For differential expression analysis, gene counts were normalized and compared for \log_2 fold change and statistical significant difference between sample groups using the DESeq2 package [364] of the R Bioconductor project [365]. The false discovery rate was controlled by adjusting the p-values using the Benjamini and Hochberg correction [366]. Genes were considered differentially expressed between sample groups for $p_{\text{adjust}} \leq 0.05$.

Gene candidates associated with specific metabolic pathways were retrieved from the *Nannochloropsis* biochemical pathway database NannoCyc [330]. Functional gene affiliations in this database are predicted using PathwayTools [367]. Gene candidates for different biochemical pathways were then pooled with non-redundant genes of the same functions identified in previous studies [63, 74, 75, 334, 344]. Gene sets were manually curated by analyzing the encoded amino acid sequences for similarities with functionally annotated proteins in the *Non-redundant protein sequences* database of the NCBI [368], using the protein-protein BLAST feature operated with standard algorithm parameters. Proteins were further analyzed for conserved domains and domain architecture using the NCBI's conserved domain database [369]. Subcellular localization of encoded proteins was analyzed using the *in silico* prediction tools HECTAR 1.3 [370], SignalP 5.0 [371], TargetP 2.0 [372], ChloroP 1.1 [373], MitoFates 1.2 [374] and for certain genes using PredPlantPTS1 [375]. \log_2 fold changes, p_{adjust} values and localization predictions for genes from different biochemical pathways are listed in tables 3.1–3.4.

Statistical data treatment

R statistical software (R Core Team 2018) was used for all data processing and statistical tests. Two-way ANOVA was employed to test for significant main effects. In case of a significant ANOVA outcome, Tukey's HSD test was used to compare the means of multiple groups with each other. Differences were considered significant in case of $p < 0.05$. When only two groups were compared, Welch's two-tailed t-test was employed instead of Tukey's HSD test. For quantitative flow cytometry analyses, we considered the median of the fluorescence distribution of gated events in a specific channel (FL1-A for measuring BODIPY fluorescence, FL5-A for measuring chlorophyll a autofluorescence) as representative value for one sample.

Data availability

The RNA-Seq data produced in this study are available at the National Center for Biotechnology Information Gene Expression Omnibus GSE167058.

Acknowledgements

This work was part of the Netherlands Organisation of Scientific Research (NWO) Buildings Blocks of Life programme (grant number 737.016.007), and of the MAGNIFICENT project, funded by the Bio-Based Industries Joint Undertaking under the European Union's Horizon 2020 research and innovation program (grant number 745754).

Credit author contribution statement

CS: Conceptualization, Methodology, Investigation, Validation, Software, Formal analysis, Writing - original draft, Writing - review & editing. **MH:** Conceptualization, Methodology, Investigation. **DRF:** Conceptualization, Methodology, Investigation. **MISN:** Conceptualization, Methodology. **JVDO:** Writing - review & editing. **SD:** Conceptualization, Project administration, Supervision, Writing - review & editing. **RHW:** Project administration, Supervision, Writing - review & editing, Funding acquisition. **MJB:** Project administration, Supervision, Writing - review & editing.

Supplementary material

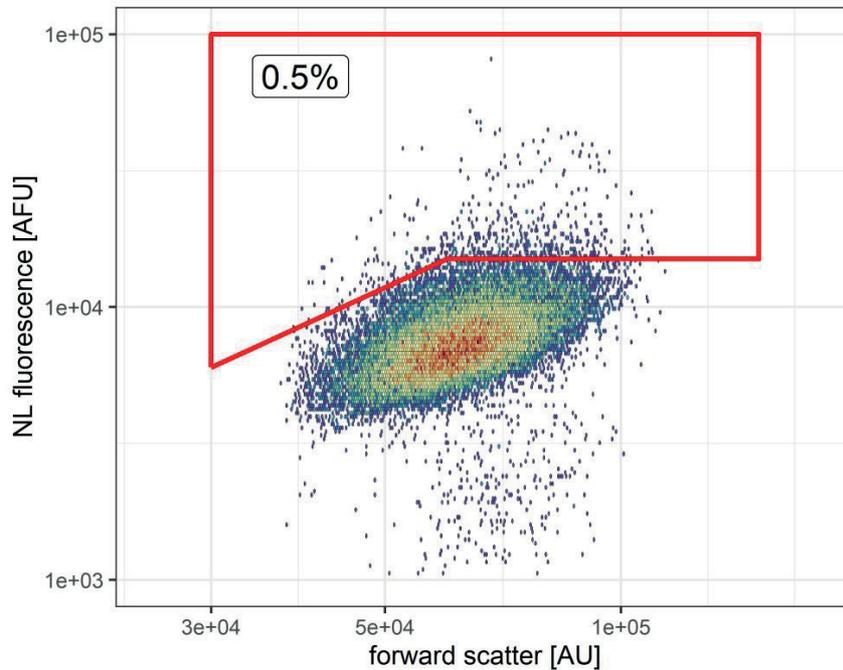


Figure S1. Sorting gate used for enrichment and isolation of HLM cells. A transitional sort gate (red polygon) was constructed on a 2-dimensional hexagonally binned density plot of the FL1 and forward scatter channels, capturing BODIPY fluorescence emission and cell size respectively. The gate was re-positioned each sorting round, to always capture ~0.5% of the total events in the parental gate (living cells, [261]). Hexagons are pseudo-colored according to the density of event counts.

Strain	NL	PL	μ_{\max}	F_v/F_m
WT	0.215±0.014	0.088±0.011	0.934±0.048	0.667±0.012
HLM21	0.240±0.036	0.086±0.007	0.992±0.065	0.684±0.007**
HLM9	0.302±0.017***	0.087±0.002	0.843±0.022*	0.679±0.006
HLM3	0.240±0.005	0.095±0.010	0.943±0.026	0.675±0.005
HLM23	0.302±0.016***	0.082±0.009	0.929±0.037	0.685±0.005**
HLM15	0.341±0.047***	0.089±0.018	0.739±0.008***	0.665±0.008

Table S1. Lipid content, growth rate and photosynthetic performance of selected HLMs. Mean±SD are given for $N \geq 3$ biological replicates. NL: neutral lipids; PL: polar lipids. Significant differences compared to the wild type (WT) group were assessed by Tukey's HSD test. (*): $p < 0.05$; (**): $p < 0.01$; (***): $p < 0.001$.

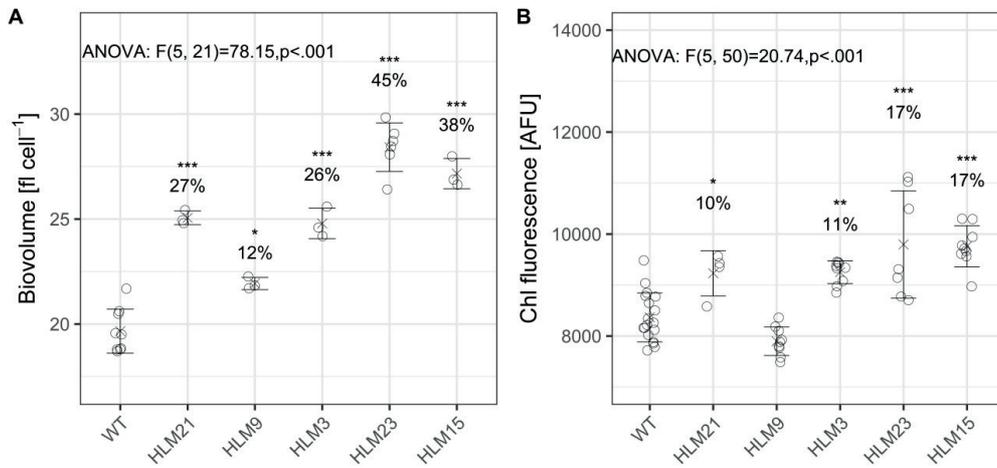


Figure S2. Physiological characterization of selected transformant lines. (A) Average cellular biovolume of transformant strains measured by particle size analysis ($N \geq 3$). The superimposed cross and errorbars denote the mean \pm SD. All transformant strains had bigger cells than the wild type, ranging from a relative 12% increase for HLM9 to 45% for HLM23. (B) Single cell chlorophyll (chl) fluorescence emission measured by flow cytometry analysis. HLM21, HLM3, HLM23 and HLM15 show increases in average chl content cell⁻¹. Significant changes relative to the wild type reference are given above each group. Significance levels were assessed by Tukey's HSD test following a significant ANOVA outcome. (*): $p < 0.05$; (**): $p < 0.01$; (***): $p < 0.001$.

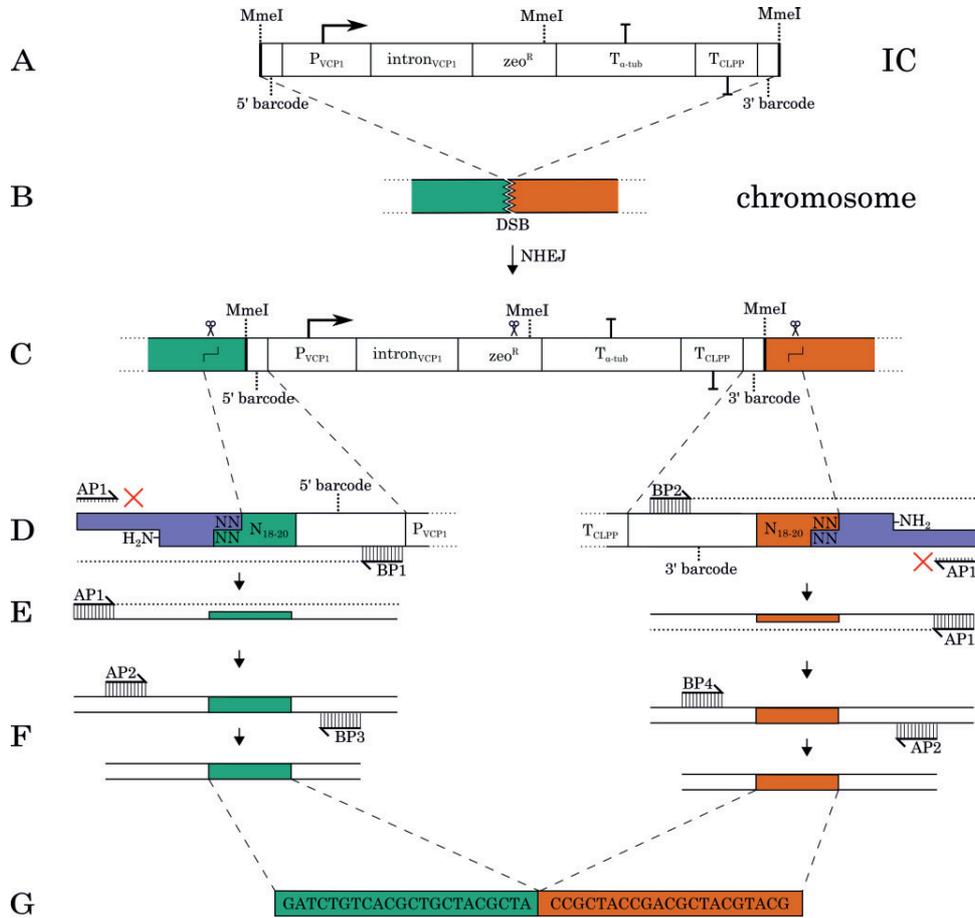


Figure S3. Genotyping procedure based on cassette PCR. (A) The IC with terminal MmeI recognition sequences is inserted into a double stranded break (DSB) at a random site of an algal chromosome (B) via nonhomologous end joining (NHEJ). Genomic mutant DNA (C) is digested with MmeI, giving rise to fragments containing the barcode sequences of the IC and 18-20 nucleotides flanking genomic DNA (green/red). (D) Fragments are ligated to asymmetric adapters (blue) and amplified by PCR (D-E) using barcode-specific primers (BPs) and adapter primers (APs). AP1 cannot anneal to the adapter but only to a template strand that has to be synthesized through extension of BP1/BP2 in the first cycle of PCR. (F) PCR with nested primers increases specificity of the procedure. (G) Sequencing of PCR products together with BLAST analysis allows the identification of the IC insertion site.

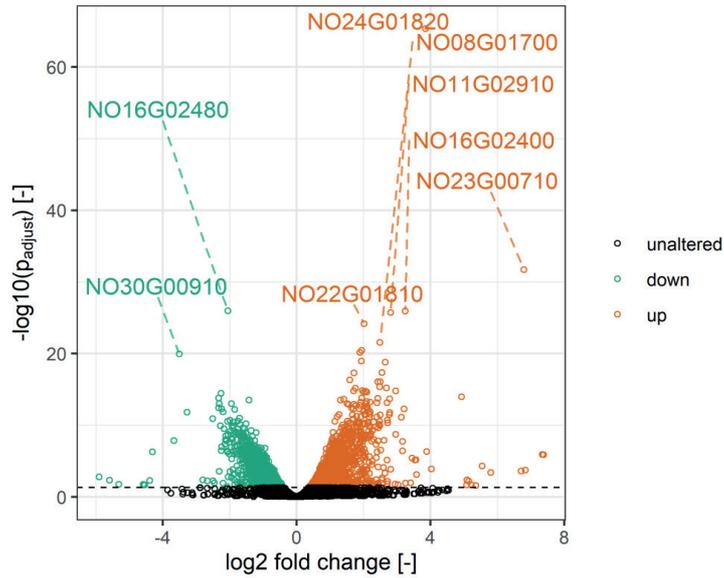


Figure S4. Visualization of differential gene expression in HLM23. The volcano plot illustrates the statistical significance and magnitude of the changes in transcript abundance for HLM23 relative to the wild type. Each dot represents a different gene, colored according to the \log_2 fold change for transcripts with $p_{\text{adjust}} \leq 0.05$. Genes with a high statistical significance and a high absolute \log_2 fold change are labelled with name.

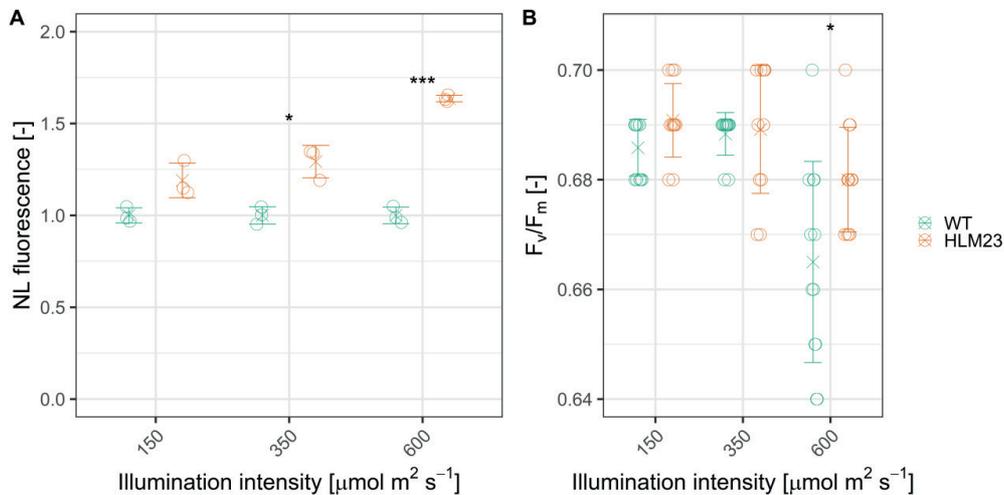


Figure S5. Relative neutral lipid content and photosynthetic quantum yield of HLM23 grown at different light intensities. Single cell neutral lipid content of exponentially growing cultures was estimated by BODIPY staining and flow cytometry. Median single cell fluorescence of $N = 3$ replicate cultures was normalized to the average wild type value. Increases are more pronounced for higher illumination intensities. F_v/F_m values were evaluated for $N = 12$ cultures during exponential growth phase. Statistical significance levels were assessed by two-tailed t-test and are given above each group. (*): $p \leq 0.05$; (**): $p < 0.01$; (***): $p < 0.001$. Error bars denote the standard deviation of the mean.

Gene	EC#	Gene ID	log ₂ fold change	p adjust	Hectar	SignalP	TargetP	ChloroP	Mitofates
LC-FACS	6.2.1.3	NO02G01580	0.47	1.95e-02	other	other	other	-	-
		NO04G00380	-0.01	9.85e-01	mTP	other	mTP	cTP	mTP
		NO05G03180	-0.36	8.47e-02	other	other	other	-	-
		NO06G00640	-0.65	8.98e-03	other	other	other	-	-
		NO14G00130	0.95	6.78e-04	other	other	other	cTP	-
		NO18G00930	0.94	3.27e-04	cTP	SP	SP	cTP	-
		NO23G01140	-0.15	6.23e-01	other	other	other	cTP	-
		NO24G00370	-0.01	9.85e-01	SP	SP	other	-	-
		NO24G00960	-0.48	2.55e-02	mTP	other	mTP	cTP	mTP
		NO25G00730	0.5	4.57e-01	other	other	other	-	-
		NO26G00150	-0.82	4.52e-04	mTP	other	mTP	cTP	mTP
		NO28G01180	0.06	8.78e-01	mTP	other	other	-	mTP
		NO30G00460	0.24	3.96e-01	other	other	other	-	-

Table S3. Differential expression and subcellular localization prediction of genes involved in glycolysis/gluconeogenesis and pyruvate metabolism in HLM23 relative to the wild type

Gene	EC#	Gene ID	log ₂ fold change	p adjust	Hectar	SignalP	TargetP	ChloroP	Mitofates
glyceraldehyde-3-phosphate dehydrogenase	1.2.1.12	NO08G01560	0.33	1.16e-01	mTP	other	cTP	cTP	mTP
		NO09G03360	1.1	9.52e-08	mTP	other	mTP	cTP	mTP
		NO27G00910	1.17	3.22e-08	cTP	SP	SP	cTP	-
glucokinase	2.7.1.1/2.7.1.2	NO27G00920	0.2	3.13e-01	other	other	other	-	-
		NO01G03040	-0.36	1.19e-01	other	other	other	-	-
6-phosphofructokinase	2.7.1.11	NO06G03510	0	9.92e-01	SP	other	cTP	cTP	-
		NO09G03890	0.62	8.69e-02	other	other	cTP	cTP	-
		NO03G04590	0.48	1.08e-02	other	other	cTP	cTP	-
		NO03G05260	1.02	5.40e-04	mTP	other	mTP	-	mTP
		NO05G05050	1.77	1.86e-13	other	other	other	-	-
phosphoglycerate kinase	2.7.2.3	NO06G01630		NA	cTP	other	cTP	cTP	-
		NO06G01060	0.65	3.41e-03	other	other	other	-	-
		NO06G04120	1.4	6.81e-08	cTP	SP	cTP	cTP	-
fructose-bisphosphatase	3.1.3.11	NO20G00840	1.49	1.93e-07	mTP	other	mTP	-	mTP
		NO10G02230	-0.07	7.87e-01	other	other	other	-	-
		NO13G02250	-0.07	8.74e-01	other	other	other	-	-
		NO15G01120	1.72	1.15e-13	other	other	other	-	-
		NO17G02820	1.39	2.11e-09	cTP	SP	SP	cTP	-
fructose 1,6-bisphosphate aldolase	4.1.2.13	NO19G01730	0.31	2.80e-01	SP	SP	SP	cTP	-
		NO01G03170	2.55	4.50e-18	cTP	SP	SP	cTP	-
		NO06G04510	-0.43	4.99e-02	mTP	other	mTP	cTP	mTP
		NO11G01780	0.34	2.19e-01	other	other	other	-	-
		NO12G03140	1.94	1.09e-19	cTP	SP	SP	-	-
phosphopyruvate hydratase	4.2.1.11	NO19G01750	0.04	8.75e-01	other	other	other	-	-
		NO03G03070	0.97	6.53e-06	cTP	SP	SP	cTP	-
		NO04G01820	1.95	3.74e-21	other	other	cTP	cTP	-
triose-phosphate isomerase	5.3.1.1	NO08G01800	0.41	1.44e-01	other	other	cTP	-	-
		NO09G03360	1.1	9.52e-08	mTP	other	mTP	cTP	mTP
		NO15G00810	1.7	1.36e-11	cTP	SP	other	cTP	mTP

Gene	EC#	Gene ID	log ₂ fold change	p adjust	Hectar	SignalP	TargetP	ChloroP	Mitofates
glucose-6-phosphate isomerase	5.3.1.9	NO01G05010	0.08	8.10e-01	cTP	SP	other	cTP	mTP
		NO02G03270	-0.11	6.61e-01	SP	other	mTP	-	-
		NO10G01100	0.63	8.37e-01	other	other	other	cTP	-
phosphoglycerate mutase	5.4.2.11	NO09G01900	0.56	1.29e-02	other	other	other	-	-
		NO15G01040	0.51	3.03e-02	mTP	other	mTP	cTP	mTP
		NO19G00380	-0.31	1.83e-01	cTP	other	other	cTP	-
		NO28G00990	1.14	5.66e-02	mTP	other	mTP	cTP	-
		NO01G00430	0.19	3.90e-01	mTP	other	mTP	-	mTP
PDHC E1	1.2.4.1	NO01G03940	0.29	2.21e-01	mTP	other	mTP	cTP	-
		NO02G05720	0.07	7.88e-01	mTP	other	mTP	cTP	mTP
		NO03G01760	-0.07	8.22e-01	mTP	other	mTP	-	mTP
		NO05G03610	0.43	1.06e-01	mTP	other	mTP	cTP	mTP
		NO10G03340	1.32	3.11e-14	cTP	SP	cTP	-	-
		NO02G02010	0.47	7.36e-02	mTP	other	mTP	-	mTP
		NO03G01760	-0.07	8.22e-01	mTP	other	mTP	-	mTP
		NO08G02730	1.3	3.47e-09	other	other	other	-	-
		NO26G00730	0.27	2.62e-01	other	other	other	cTP	-
		NO02G05720	0.07	7.88e-01	mTP	other	mTP	cTP	mTP
PDHC E2	2.3.1.12	NO03G01630	0.23	3.31e-01	other	other	other	-	-
		NO05G03610	0.43	1.06e-01	mTP	other	mTP	cTP	mTP
		NO13G01210	0.51	5.02e-02	mTP	other	mTP	cTP	mTP
		NO18G02350	0.94	5.92e-06	cTP	SP	cTP	-	-
		NO12G01110	-0.17	4.66e-01	mTP	other	mTP	-	mTP
		NO14G03070	0.18	4.29e-01	other	other	other	-	-
		NO20G01100	-0.48	3.78e-02	SP	other	other	-	-
		NO04G01610	-0.47	2.77e-02	other	other	other	-	-
		NO01G04110	-0.69	4.85e-02	other	other	other	-	-
		NO13G03340	-0.01	9.73e-01	cTP	other	SP	cTP	-
aldehyde dehydrogenase	1.2.1.3								
pyruvate decarboxylase	4.1.1.1								
acetate-CoA ligase	6.2.1.1								

Table S4. Differential expression and subcellular localization prediction of genes involved in the Calvin–Benson–Bassham cycle in HLM23 relative to the wild type

Gene	EC#	Gene ID	log ₂ fold change	p adjust	Hectar	SignalP	TargetP	ChloroP	Mitofates
glyceraldehyde-3-phosphate dehydrogenase	1.2.1.12	NO08G01560	0.33	1.16e-01	mTP	other	cTP	cTP	mTP
		NO09G03360	1.1	9.52e-08	mTP	other	mTP	cTP	mTP
		NO27G00910	1.17	3.22e-08	cTP	SP	SP	cTP	-
transketolase	2.2.1.1	NO27G00920	0.2	3.13e-01	other	other	other	-	-
		NO08G02220	1.45	1.08e-08	cTP	other	cTP	cTP	-
		NO14G02050	-0.34	1.88e-01	other	other	other	-	-
phosphoribulokinase	2.7.1.19	NO13G02830	2.1	5.17e-14	cTP	SP	SP	mTP	
phosphoglycerate kinase	2.7.2.3	NO06G01060	0.65	3.41e-03	other	other	other	-	-
		NO06G04120	1.4	6.81e-08	cTP	SP	cTP	cTP	-
		NO20G00840	1.49	1.93e-07	mTP	other	mTP	-	mTP
fructose-bisphosphatase	3.1.3.11/3.1.3.37	NO10G02230	-0.07	7.87e-01	other	other	other	-	-
		NO13G02250	-0.07	8.74e-01	other	other	other	-	-
		NO15G01120	1.72	1.15e-13	other	other	other	-	-
fructose 1,6-bisphosphate aldolase	4.1.2.13	NO17G02820	1.39	2.11e-09	cTP	SP	SP	cTP	-
		NO19G01730	0.31	2.80e-01	SP	SP	SP	cTP	-
		NO01G03170	2.55	4.50e-18	cTP	SP	SP	cTP	-
		NO06G04510	-0.43	4.99e-02	mTP	other	mTP	cTP	mTP
		NO11G01780	0.34	2.19e-01	other	other	other	-	-
		NO12G03140	1.94	1.09e-19	cTP	SP	SP	-	-
ribulose-phosphate 3-epimerase	5.1.3.1	NO19G01750	0.04	8.75e-01	other	other	other	-	-
		NO11G02000	0.85	6.15e-04	other	other	other	-	-
		NO11G02910	2.81	1.94e-26	other	other	other	-	-
triose-phosphate isomerase	5.3.1.1	NO19G01060	-0.13	6.45e-01	other	other	other	-	-
		NO08G01800	0.41	1.44e-01	other	other	cTP	-	-
		NO09G03360	1.1	9.52e-08	mTP	other	mTP	cTP	mTP
ribose-5-phosphate isomerase	5.3.1.6	NO15G00810	1.7	1.36e-11	cTP	SP	other	cTP	mTP
		NO08G03720	0.82	3.76e-05	SP	SP	cTP	cTP	mTP

Table S5. Differential expression and subcellular localization prediction of other relevant genes in HLM23 relative to the wild type

Gene	EC#	Gene ID	log ₂ fold change	p adjust	Hectar	SignalP	TargetP	ChloroP	Mitofates
acyl-CoA-binding protein	-	NO16G02490	1.63	3.55e-06	other	other	other	-	-
		NO19G00260	0.3	1.84e-01	other	other	other	-	mTP
		NO23G00150		NA	other	other	other	-	-
ferredoxin		NO24G01960	-0.38	1.62e-01	other	other	other	-	-
		NO03G01380	1.29	5.76e-05	mTP	other	mTP	-	mTP
		NO05G02270	1.1	8.78e-04	SP	SP	SP	-	-
		NO09G00110	3.45	3.85e-06	cTP	SP	SP	-	-
		NO12G01630	-0.36	3.12e-01	SP	SP	SP	-	-
		NO12G03710	0.04	8.86e-01	mTP	other	mTP	cTP	mTP
		NO24G00440	1.2	2.93e-06	SP	SP	SP	-	-
		NO24G02290	0.46	4.29e-02	SA	other	other	cTP	-
		NO24G01680	0.89	2.09e-07	SP	SP	cTP	cTP	-
		chlorophyll a-b binding protein							
oxygen-evolving enhancer protein 1 (PsbO)	1.10.3.9								
ferredoxin:NADP(+) oxidoreductase	1.18.1.2	NO16G02580	1.15	3.07e-11	cTP	SP	cTP	-	
magnesium-chelatase subunit ChlD	6.6.1.1	NO16G00200	1.31	3.54e-07	SP	other	SP	-	
magnesium-chelatase subunit ChlH		NO22G01810	2.02	7.44e-25	SA	other	cTP	-	
phosphogluconate dehydrogenase	1.1.1.44	NO24G01300	-0.44	1.78e-01	SA	other	cTP	-	
glucose-6-phosphate dehydrogenase	1.1.1.49	NO04G01320	-0.4	8.70e-02	other	other	-	-	
6-phosphogluconolactonase	3.1.1.31	NO10G01570	-0.5	6.56e-02	SP	SP	-	-	
glucose-6-phosphate/P _i translocator	-	NO12G01420	-0.5	3.90e-02	other	other	cTP	-	
acyl-[acyl-carrier-protein] desaturase	1.14.19.2	NO25G00510	-1.03	1.12e-03	other	other	-	-	
		NO02G01930	3.54	7.56e-06	SP	SP	mTP	cTP	
		NO02G01510	-0.73	1.43e-03	other	other	other	cTP	



CHAPTER 6

Heterologous expression of EPA specific Lysophosphatidic acid acyltransferase in *Nannochloropsis oceanica* increases the EPA content in Triglycerides

Mihris Ibnu Saleem Naduthodi^{1 2}, Narcis Ferrer Ledo², Sarah D'Adamo², John van der Oost¹ and Maria Barbosa²

¹Laboratory of Microbiology, Wageningen University and Research, Wageningen, Netherlands.

²Bioprocess Engineering, AlgaePARC, Wageningen University and Research, Wageningen, Netherlands.

Abstract

Eicosapentaenoic acid (EPA) is a valuable metabolite with various health benefits when accumulated in the lipid triglycerides (TAG). Because the major source of this fatty acid is wild fish oil, the increasing demand for EPA has endangered the wild fish population. Importantly, it is not fish but rather microalgae that are the primary producers of these fatty acids. Among microalgae, *Nannochloropsis oceanica* stands out as they accumulate large amounts of EPA. Thus, *N. oceanica* is considered a sustainable alternative to fish oils as a source of EPA. In this study, a plant-derived gene encoding an EPA-specific lysophosphatidic acid acyltransferase was heterologously expressed in *N. oceanica*, aiming at enhancing the level of EPA accumulation in TAGs. LPAAT catalyses the conversion of lysophosphatidic acid into phosphatidic acid by esterifying the sn-2 position of the glycerol backbone and subsequently produces EPA rich diglycerides and triglycerides in *Nannochloropsis*. Under nitrogen starved growth conditions, the EPA content in the total fatty acids as well as in the TAGs of an BnLPAAT expressing strain was improved by 28% and 38%, respectively. Although the EPA content in polar lipids was not improved in selected LPAAT overproducing clones, concentrations of medium chain fatty acids in this fraction were elevated. We conclude that heterologous expression of EPA-specific LPAAT can improve the EPA content in *N. oceanica*. As LPAAT specifically esterifies the sn-2 position of the glycerol backbone to produce diglycerides, further validations are required to confirm the desired EPA enrichment at the sn-2 position of TAG molecules.

Introduction

Long-chain polyunsaturated fatty acids (LC-PUFA) such as Eicosapentaenoic acid (EPA) are essential fatty acids in aquaculture applications [68]. Also in human nutrition health benefits have been demonstrated such as prevention of diabetes, cardiovascular diseases and Alzheimer's [376, 377]. The wide application of this omega-3 fatty acid in food, feed, and nutraceutical industries has steeply increased its global demand [378]. Fish oil is presently the main dietary source of these fatty acids [377]. The increasing demand for omega-3 fatty acids has fuelled the unsustainable capture of wild fish, potentially risking the extinction of various marine species [68]. Marine fish does not have the capacity to synthesize these fatty acids itself, but it rather acquires them by consuming microalgae, the primary producers of LC-PUFAs [28]. Thus, extracting these fatty acids from microalgae could be a sustainable alternative to obtaining them from wild fish. *Nannochloropsis* sp. are identified as potential strains for sustainable EPA production as they can accumulate EPA up to 12% of their cell dry weight (CDW) [69].

The biosynthesis of LC-PUFAs in microalgae starts with the conversion of acetyl-CoA to malonyl-CoA by enzyme acetyl-CoA carboxylase. The malonyl-CoA is further converted to malonyl-ACP and then used as a substrate by enzyme KASIII to form palmitoyl-ACP (C16:0-ACP). The C16:0-ACP undergoes multiple rounds of elongation and desaturation to form various fatty acid species in microalgae. These fatty acid profiles are highly variable between species [379]. In this regard, *Nannochloropsis* sp. stands out from other organisms as they accumulate valuable EPA in large amounts. These fatty acids are esterified with glycerol to form polar lipids (PLs) that constitute the membranes of various organelles or neutral lipids (NLs) such as triglycerides (TAG) [379].

The TAGs act as carriers of EPA and the nutritional value of TAG molecules varies based on the position of EPA or LC-PUFAs on TAG [380]. After human consumption, the fatty acids at the sn-1 and sn-3 positions are cleaved during TAG digestion, resulting in the formation of monoacylglycerol (MAG) (sn-2) and free fatty acids in the stomach. The free fatty acids are absorbed by the intestine, while the MAGs (sn-2) undergo esterification at the sn-1 and sn-3 positions with endogenous FAs by retaining the fatty acid at the sn2 position [381]. Therefore, TAG molecules with EPA accumulated at the sn-2 position could be nutritionally relevant compared to TAG with EPA at sn-1 or sn-3 positions [376]. The TAG assembly in *Nannochloropsis* is facilitated via the Kennedy pathway and PDAT mediated conversion of diacylglycerol (DAG) to TAG. In the Kennedy pathway, the glycerol-3 phosphate is esterified at the sn-1 position by glycerol-3-phosphate acyltransferase (GPAT) to form Lysophosphatidic acid (LPA) [382]. Subsequently, LPA undergoes esterification at the sn-2 position by Lysophosphatidic acid acyltransferase (LPAAT) to form phosphatidic acid, which is converted to DAG by phosphatidic acid phosphatase (PAP) [383]. The DAG undergoes esterification at the sn-3 positions by enzyme diacylglycerol acyltransferase (DGAT) to form TAG or they are converted to various PLs such as Monogalactosyldiacylglycerol (MGDG) and Sulfoquinovosyl diacylglycerol (SQDG) by enzymes MGDG synthase and UDP-sulphoquinovose:DAG sulphoquinovosyltransferase, respectively [384].

In *Nannochloropsis oceanica*, EPA is predominantly accumulated at the sn-3 position of TAG molecules [74]. Thus, esterification of the sn-2 position of TAG with EPA could improve the nutritional value of TAG molecules produced and the EPA production in *Nannochloropsis* [376]. To this end, an EPA-specific LPAAT enzyme from the plant *Brassica napus* was expressed in *Nannochloropsis oceanica*, and the obtained strains expressing the heterologous protein were characterized. The main objective of this study was to illustrate the effect of EPA-specific BnLPAAT expression in the fatty acid content of *Nannochloropsis*. Our study shows that BnLPAAT expression could improve the EPA fraction in the total fatty acid (TFA) and TAG under nitrogen stress.

Material and Methods

Media and growth conditions of *N. oceanica*

The *N. oceanica* IMET 1 was grown under 16:8 day-night cycles and a constant illumination intensity of 130 $\mu\text{mol}/\text{m}^2/\text{s}$ at 25°C and 0.2% CO_2 in an Infors incubator. The cultures were grown in an artificial sea water (ASW) solution that contains 419 mM sodium chloride, 22.53 mM sodium sulphate, 5.42 mM calcium chloride dihydrate, 4.88 mM potassium sulphate, 48.21 mM magnesium chloride hexahydrate and 20mM HEPES. The pH of ASW was adjusted to 7.9 and the media was autoclaved. The sterilized media was enriched by 500X concentrated nutrient solution that contains 2 M NaNO_3 , 10 mM KH_2PO_4 , 1 mM ZnCl_2 , 1 mM ZnSO_4 , 1 mM $\text{MnCl}_2 \cdot 2\text{H}_2\text{O}$, 0.1 mM $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 0.1 mM $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, 0.1 mM $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 26.4 mM EDTA, 2 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 20mM $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$, 35 mg/l Thiamine, 5 mg/l Biotin and 3 mg/l Vitamin B_{12} . For nitrogen starved conditions, the same nutrient solution without nitrate was supplied to the media. To assess the growth characteristics of the mutants under nitrogen stress, the cultures were initially grown to an exponential growth phase under optimal conditions and subsequently, the cultures were harvested, washed, and resuspended in equal volume of nitrogen less media.

Developing the BnLPAAT expressing plasmid

To express the BnLPAAT in *N. oceanica*, the BnLPAAT gene sequence was initially codon harmonized [239] for *Nannochloropsis*. The codon harmonized gene sequence was ordered as synthetic dsDNA fragment from Twist bioscience. The gene sequence was PCR amplified using primer sequences with overhangs that overlaps to the native lipid droplet surface protein (LDSP) promoter on the 5' end and the Nano luciferase (Nlux) protein on the 3' end. The Cauliflower mosaic virus CaMV 35 S terminator was PCR amplified using primers that harboured overlaps to the Nlux protein on the 5' end and the previously reported zeocin resistance on the 3' end [227]. The zeocin resistance cassette was amplified along by PCR along with a pUC19 backbone and all the fragments were assembled using 2X Hifi DNA assembly master mix from New England Biolabs (NEB). The assembled products were cloned in chemically competent Dh5-alpha *Escherichia coli* cells and selected on ampicillin. Multiple colonies were independently scaled up in liquid LB with ampicillin and the plasmids were isolated using Gene-Jet Plasmid Miniprep Kit from Thermo Scientific. The plasmids were initially screened for proper assembly by restriction digestion and then the BnLPAAT and Zeocin resistance cassettes were sequence verified from Macrogen Sequencing. The correct plasmids were used as template and the BnLPAAT-

Zeocin Cassettes were amplified by PCR to obtain the linear plasmid for transformation in *N. oceanica*.

Transformation of *N. oceanica*

The liquid culture of *N. oceanica* was inoculated into a fresh media with a starting OD₇₅₀ value of 0.06 to 0.08. The cultures were grown with constant medium light conditions until they reach an OD₇₅₀ value of 0.5. At this point, 50ml of the culture was harvested and centrifuged at 4°C at 2000 g for 10 mins. The supernatant was discarded, and the pellet was resuspended in 25 mL 375 mM D-Sorbitol solution that was pre-cooled on ice for 1 hour. The cells were again centrifuged, and the pellets were washed again in D-Sorbitol solution. After 3 rounds of washing, the pellets were resuspended in 200 µL cold D-Sorbitol solution. 30 µg of Salmon sperm DNA (Invitrogen) and 3 µg of our linear plasmid was added to the resuspended culture and gently mixed by pipetting. The suspension was then transferred to a pre-cooled 2mm electroporation cuvette and transformed by providing an exponential pulse at specifications: 2200V, 600 ohm and 50 µF. The transformed cultures were quickly transferred into 5mL media and incubated under continuous medium light condition for 2 days. After the recovery period, the samples were centrifuged, and the cell pellets were plated on solid media with 10 µg/ml of zeocin antibiotic and incubated under constant light for approximately 4 weeks to obtain the transformed colonies.

Nano-luciferase assay

The colonies were resuspended in fresh media on 24 well plates and grown under continuous light for few days until the cultures start growing exponentially. 150 µL of the cultures were transferred into a new 96 well plate and equal volume of the media with 10,000X diluted Nano-luciferase substrate was added to the cultures to measure the luminescence on a SynergyMx plate reader from BioTek.

Lipid extraction and quantification from *N. oceanica*

Under optimal growth conditions (OD₇₅₀ ~0.5), the lipids were extracted from mutants and WT during the exponential growth phase. To extract the lipids under nitrogen stress, the cultures were initially grown to the exponential phase and then harvested and resuspended in nitrogen-less media. The lipids were extracted after 6 days of starvation. From the cultures, 50 mL was used for extracting the lipids and 30 mL was used for measuring the dry weight. The 50 mL culture was centrifuged for 5 minutes at 1200 g and 4°C to obtain the cell pellet (For the nitrogen starved cultures or cultures with high lipid content, longer centrifugation step will be required to pellet the entire culture). The supernatant was completely removed into a clean tube on ice and gently resuspend the pellet in 150-200 µL of the supernatant media on ice by pipetting. The resuspended culture was carefully transferred to a bead beater tube and the centrifuge tube was washed again by 100-150 µL of milliQ water and transferred to the bead beater tube. The bead beater tube was centrifuged for 60 seconds at maximum RPM in room temperature and the tubes were stored at -20°C.

To the tube with the sample, 1 mL of Chloroform: Methanol (2: 2.5) solution was added and run on a bead beater from Bertin technologies with specifications: 8 x 60 sec with 120 sec pause in between at 2500 rpm. The sample from the bead beater was transferred to

a heat resistant 10 mL glass tube. The bead beater tubes were washed three times using 1 mL of chloroform:methanol solution and all the solutions after each wash were collected along with the sample in the glass tube. The sample was vortexed for 5 seconds and incubated 10 minutes in a sonicator bath at 80 kHz. To the sonicated sample, 2.5 mL of 50 mM Tris-HCL (with 1M NaCl at pH 7.5) was added and the sonication was repeated. After a 5-minute centrifugation at 2500 rpm, the lower phase was transferred using a glass pipette into a fresh tube and 1 mL of chloroform was added to the old tube, vortexed and sonicated again. The old tube was centrifuged, and the lower phase was pooled in the tube along with the first chloroform fraction. The washing step with chloroform was repeated two more times and finally the chloroform was evaporated using nitrogen gas in the fresh collection tube to obtain the lipid extracts from *N. oceanica*.

The dried lipid extract was resuspended in 1 mL chloroform and 50 μ L of the resuspended sample was transferred to a fresh glass tube for quantifying the lipid extracts. The 950 μ L of the sample was dried using nitrogen gas stream and stored at -20°C . The 50 μ L sample was also dried using nitrogen gas stream and 10 μ L of internal standard that contains C15:0 FFA (in 2:1 Chloroform: Methanol solution) was added. To this sample, 3 mL of methanol and 5% sulphuric acid was added and heated in a water bath at 100°C for 1 hour. The sample was cooled to room temperature, then, 3 mL milliQ water and 1 mL hexane were added to the sample. After a quick vortex and mixing on a test tube rotator for 15 minutes, the samples were centrifuged at 15°C and 2500 rpm. The hexane phase on the top was transferred to GC vials and the samples were analysed using Gas chromatograph.

Thin Layer Chromatography (TLC) for neutral lipid (NL) and polar lipid (PL) separation

The silica plates for TLC were activated at 100°C for 1 hour in a hot air oven. The migration solvents for neutral lipid TLC were made using 70 mL hexane, 30 mL diethyl ether and 1 mL acetic acid. Similarly, 65 mL chloroform, 25 mL methanol and 4 mL MilliQ water were combined to make the migration solvent for the polar lipid TLC. The lipid extract was resuspended in 950 μ L chloroform and based on the quantification data, approximately the volume corresponding to 400 μ g of lipid extract was transferred to a new tube and dried using nitrogen gas. The lipid extract stock was also dried and stored at -20°C . The dried 400 μ g lipid extract was resuspended in 40 μ L chloroform and gently added on a specific spot on the activated and labelled TLC plate using a glass Pasteur pipette. The TLC plate with the samples are placed inside a migration box until the migration solvent (separate for each lipid class) reached approximately 2cm from the top edge of the TLC plate. The TLC plates are dried and stained by spraying the Primulin (0.5% primulin in 8:2 acetone: milliQ water) solution on to the plate. The plates are again dried and the spots corresponding to each lipid class are visualized under UV light (366 nm).

The individual spots corresponding to different lipid classes were scratched from the TLC plate and collected separately in a fresh glass tube. Each sample were methylated separately as mentioned above and analysed on GC to obtain the fatty acid profiles of each lipid class [385, 386].

Results and Discussion

Codon harmonized BnLPAAT expression in *N. oceanica*

Overexpression of LPAAT has been reported to enhance the lipid content in microalgae such as *Chlamydomonas reinhardtii*, *Phaeodactylum tricornutum*, *Neochloris oleabundans* and *Chlorella* sp. [387-391]. However, overexpression of LPAAT to tailor the accumulation of specific fatty acids in TAGs was seldom reported in microalgae. We hypothesized the heterologous expression of EPA-specific LPAAT from *B. napus* could enhance the EPA content in TAGs or even improve the total lipid content in the mutants as a similar strategy was successfully implemented in plant *Nicotiana benthamiana* to improve the docosahexaenoic acid (DHA) content in TAG and lipids by overexpressing DHA specific LPAAT gene [376]. As TAG accumulation in *Nannochloropsis* is triggered under nitrogen starvation, to channel more EPA into TAGs, we selected a native lipid droplet surface protein (LDSP) promoter that was proposed to be highly active under nitrogen stress. Thus, LDSP promoter and 35S terminator were used for driving the expression of a codon harmonized *BnLPAAT* gene in *N. oceanica* IMET1 [81, 239].

The major *de novo* TAG synthesis in eukaryotes is assumed to be localized in the endoplasmic reticulum (ER) [392, 393]. The BnLPAAT protein was also predicted (PSORT II and WoLF PSORT) to localize in the ER and hence additional localization signals were not tagged onto the protein sequence [394]. The coding DNA sequence of the *BnLPAAT* gene was tagged with NanoLuciferase (Nlux) to validate the protein expression. Thus, the 3D structure of the protein was predicted (Phyre2) to confirm that the Nlux tag at the C' terminal does not adversely affect the protein structure [395].

The linear DNA harbouring the codon harmonized *BnLPAAT* and antibiotic resistance cassette was transformed into *N. oceanica* IMET1. Two colonies (mutant A and B) with high Nlux activity were selected from the transformants for characterizing the BnLPAAT expressing strains.

Characterization of BnLPAAT expressing lines

BnLPAAT expressing strain has reduced growth and higher lipid content under nitrogen stress

The selected strains were initially characterized for their growth properties under optimal and nitrogen starved conditions. Under optimal conditions, the selected transformants, and the wild type (WT) cultures had similar growth characteristics (Figure 1). Under nitrogen starved conditions, we observed that the mutants A and B had slower growth rates and lower biomass accumulation compared to the WT after 6 days of nitrogen starvation (Figure 1). The native LDSP promoter is proposed to have higher activity under stress in *Nannochloropsis* [396]. Thus, it is possible that, under optimal conditions, the BnLPAAT expression might not be detrimental for the growth parameters of *Nannochloropsis*. On the other hand, the improved protein production under stress in the mutants could be hindering their growth properties. In line with the other studies, a 12.8 % increase in the lipid content of mutant B (13.02 g. gDW⁻¹) under nitrogen starved condition compared to the WT (11.88 g. gDW⁻¹) (Figure S1) [387, 389, 390]. However, the lipid contents in the mutant strains were lower compared to the WT under optimal conditions.

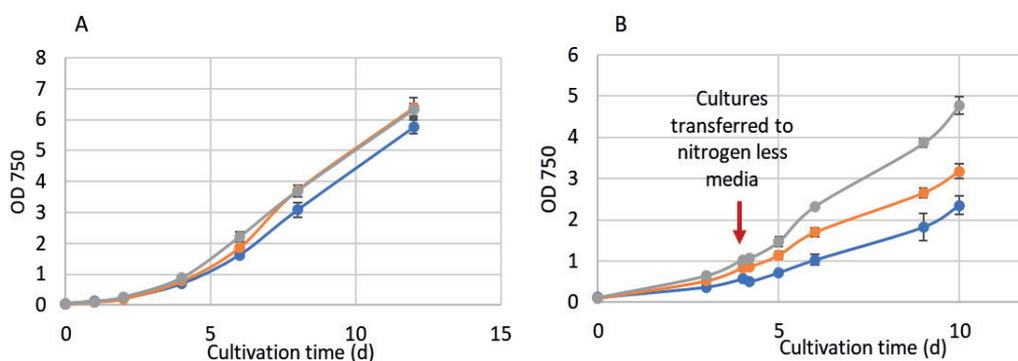


Figure 1. Growth curve of mutant strains expressing BnLPAAT under (A) optimal and (B) nitrogen starved conditions. The mean value at each time-point and the standard deviation is calculated from sample size N=3.

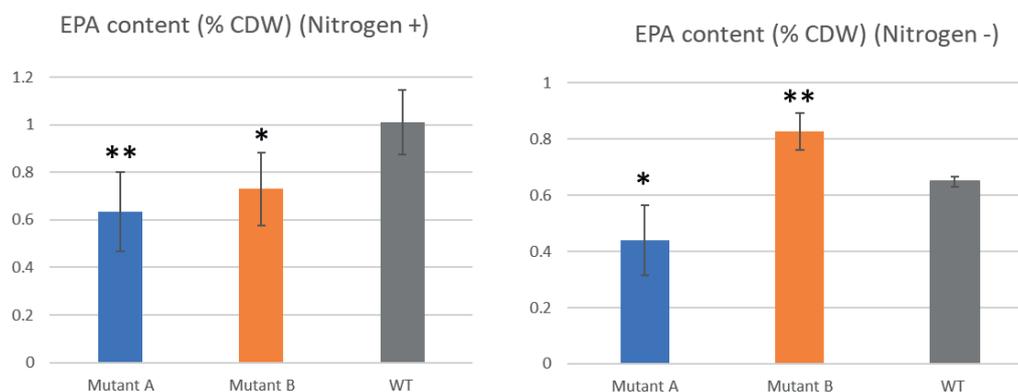


Figure 2. EPA content in the BnLPAAT expressing strains and WT under optimal (left) and nitrogen starved (right) conditions. Error bars denote the standard error of the mean (N=3). The statistical significance of each sample was measured using two sample t test. * Indicate that the p value < 0.1 and ** indicate the p value < 0.04

Similarly, the EPA content in the total fatty acids was lower for the mutants under optimal conditions, while the EPA content in the total cell dry weight (CDW) of mutant B under nitrogen starved condition was increased by 28% (Figure 2). Under these conditions, the precursor of EPA, Arachidonic acid (ARA, C20:4) was also improved by 60% in the mutant B indicating an improved flux towards EPA production in *BnLPAAT* expressing mutants (Table S1). Furthermore, the fraction of C14:0 was also increased by 48% in the mutant B under nitrogen starved conditions (Table S1). C16 and C16:1 were the major fatty acids produced by *N. oceanica* and the fraction of these fatty acids in the TFA doubled under nitrogen stress in all the strains. The C16:1 content in the TFA of mutant B was increased by 21% under nitrogen starved condition compared to the WT (Table S1). Similar variations in the fatty acid profiles were also reported in *P. tricornutum* where endogenous LPAAT overexpression improved the fatty acids with chain length C14 to C20 [389]. Under

nitrogen starved conditions, mutant A depicted significantly slower growth rate, reduced lipid and EPA accumulation compared to mutant B and WT.

Table 1. Fatty acid profile in Triglycerides (TAGs) of BnLPAAT expressing mutants and WT IMET1 strains. The average value and standard deviation are based on sample size $N \geq 2$.

Fatty acid	Optimal condition					
	Mutant A		Mutant B		WT	
	mg/g.CDW	S.D	mg/g.CDW	S.D	mg/g.CDW	S.D
C14:0	1,85	0,05	1,54	0,12	1,38	0,09
C16:0	12,01	0,65	8,37	0,94	9,76	1,02
C16:1	13,35	0,70	8,50	0,76	9,09	0,81
C18:0	0,39	0,05	0,28	0,01	0,33	0,00
C18:1	1,49	0,08	0,85	0,09	1,08	0,03
C18:2	0,32	0,01	0,15	0,01	0,20	0,03
C20:4-n6	0,00	0,00	0,00	0,00	0,00	0,00
C20:5-n3	0,49	0,02	0,19	0,07	0,33	0,03

Fatty acid	Nitrogen starved					
	Mutant A		Mutant B		WT	
	mg/g.CDW	S.D	mg/g.CDW	S.D	mg/g.CDW	S.D
C14:0	6,94	0,62	9,06	0,12	6,00	0,10
C16:0	42,51	2,20	38,74	0,57	39,59	0,62
C16:1	31,76	0,04	31,13	0,38	26,29	0,84
C18:0	1,11	0,01	0,95	0,03	0,92	0,04
C18:1	7,55	0,02	6,79	0,13	7,40	0,10
C18:2	0,93	0,03	0,97	0,05	0,98	0,04
C20:4-n6	0,95	0,06	1,50	0,07	0,81	0,02
C20:5-n3	2,38	0,05	4,14	0,12	3,00	0,08

Neutral lipid classes and fatty acid profiles

To further assess the effect of BnLPAAT expression in neutral lipids (NLs), various NL classes (TAG, DAG, and FFA) were separated, and their fatty acid profiles were analysed. The NLs were approximately 3 to 5 times higher (Table S2) in all the strains under nitrogen starved conditions with TAGs being the dominant lipid class. Under optimal conditions, TAGs constituted around 85% of the total NLs and this fraction was improved to 95% under stress (Table S3). The fraction of DAGs in the NLs were similar under stress and optimal conditions for mutants and WT. Nevertheless, the fraction of FFAs (free fatty acids) in NLs were reduced under stress from roughly 15% to 2.5% of NLs in all the strains (Table S3).

Table 2. Fatty acid profile in diglycerides (DAGs) of BnLPAAT expressing mutants and WT IMET1 strains. The average value and standard deviation are based on sample size $N \geq 2$.

Fatty acid	Optimal condition					
	Mutant A		Mutant B		WT	
	mg/g.CDW	S.D	mg/g.CDW	S.D	mg/g.CDW	S.D
C14:0	0,04	0,01	0,05	0,01	0,04	0,00
C16:0	0,12	0,03	0,14	0,00	0,13	0,02
C16:1	0,05	0,05	0,08	0,00	0,08	0,02
C18:0	0,00	0,00	0,00	0,00	0,00	0,00
C18:1	0,00	0,00	0,00	0,00	0,00	0,00
C18:2	0,00	0,00	0,00	0,00	0,00	0,00
C20:4-n6	0,00	0,00	0,00	0,00	0,00	0,00
C20:5-n3	0,00	0,00	0,00	0,00	0,00	0,00

Fatty acid	Nitrogen starved					
	Mutant A		Mutant B		WT	
	mg/g.CDW	S.D	mg/g.CDW	S.D	mg/g.CDW	S.D
C14:0	0,20	0,00	0,32	0,02	0,33	0,05
C16:0	0,50	0,04	0,81	0,01	0,97	0,12
C16:1	0,39	0,03	0,69	0,03	0,70	0,10
C18:0	0,00	0,00	0,00	0,00	0,00	0,00
C18:1	0,00	0,00	0,00	0,00	0,13	0,01
C18:2	0,00	0,00	0,00	0,00	0,00	0,00
C20:4-n6	0,00	0,00	0,00	0,00	0,05	0,03
C20:5-n3	0,21	0,04	0,49	0,04	0,83	0,20

Under optimal conditions, the EPA content in TAGs were lower in mutant B compared to the WT. However, under nitrogen starved conditions, the fraction of EPA in TAGs of mutant B was improved by 38%. In DAGs, EPA was only observed in the WT under stress constituting 0.05 mg/g.CDW. In the FFA, mutant B and WT accumulated similar amounts of EPA under optimal and nitrogen starved conditions (Table S4). The precursor of EPA, ARA was observed only in the FFAs of mutant B and WT during optimal growth conditions, while under nitrogen stress, ARA was only present in the TAGs. Interestingly, compared to the WT, ARA content in mutant A and B was 17% and 85% higher, respectively. These observations are in line with previous studies that showed LC-PUFAs are stored in TAGs under stress and are used for developing membrane lipids as nitrogen is replenished in the media [293].

Polar lipid classes and fatty acid profiles

As the BnLPAAT directly influence the formation of DAGs, the fatty acid profiles of PLs can also be affected by the expression of this enzyme in *Nannochloropsis*. TLC separated various polar lipid classes and the fatty acid contents were analysed from both WT and mutant strains. The polar lipids were separated by a simplified 1D TLC procedure. Thus, the separation of polar lipid classes on the chromatogram was not optimal for differentiating the lipid classes. Nevertheless, the total fatty acid contents from the polar lipid classes were determined by this procedure.

EPA mainly constituted the PLs under optimal growth conditions in both the mutants and the WT strains. However, mutants did not have improved EPA content in the PLs under optimal nor stressed conditions. The mutant A accumulated significantly lower EPA fractions than the WT, while the mutant B accumulated comparable EPA fractions in the PLs (Table 3).

C16 and C16:1 constituted a significant part of PLs in *N. oceanica* and under optimal conditions, the fraction of these fatty acids in the PLs were similar in mutants and WT. However, under nitrogen starved conditions, C16:0 and C16:1 were doubled, in the PLs of mutant A compared to WT.

Table 3. Fatty acid profile in the polar lipids of BnLPAAT expressing mutants and WT IMET1 strains. The average value and standard deviation are based on sample size $N \geq 2$.

Fatty acid	Optimal condition					
	Mutant A		Mutant B		WT	
	mg/g.CDW	SD	mg/g.CDW	SD	mg/g.CDW	SD
C14:0	2,71	0,67	3,58	0,70	3,29	0,37
C16:0	12,64	4,00	12,47	3,07	13,16	0,91
C16:1	12,07	3,75	11,94	3,27	11,59	0,20
C18:1	2,62	0,57	2,25	0,43	2,26	0,14
C18:2	1,05	0,08	1,41	0,24	1,55	0,12
C20:4	1,20	0,02	1,58	0,56	2,08	0,31
C20:5	7,18	0,92	9,38	1,29	11,81	0,94

Fatty acid	Nitrogen starved					
	Mutant A		Mutant B		WT	
	mg/g.CDW	SD	mg/g.CDW	SD	mg/g.CDW	SD
C14:0	2,97	0,30	1,72	0,31	1,52	0,54
C16:0	16,56	0,69	5,51	1,11	7,89	3,47
C16:1	10,66	0,33	3,96	0,82	4,80	2,29
C18:1	4,07	0,52	2,89	0,25	3,14	0,58
C18:2	0,50	0,07	0,43	0,02	0,43	0,14
C20:4	0,76	0,26	1,41	0,04	1,23	0,10
C20:5	1,81	0,24	3,72	0,07	3,32	0,62

In *Nannochloropsis*, the PLs such as MGDG and DGTS contains EPA in the sn2 position while TAGs contain EPA predominantly at the sn3 position [74]. Considering DAG as the precursor for all these lipid classes, DAGs with EPA in the sn2 position may maybe specifically incorporated in PLs at least under optimal conditions. The high EPA content in PLs under the optimal condition, as observed in our study also indicate high demand for EPA rich DAG for PL assembly. Suppose this is the case, to accumulate TAGs with EPA at the sn2 position, DGATs with high substrate specificity towards DAG molecules with EPA at sn2 position should be co-expressed with BnLPAAT. To this end, DGAT enzymes from

strains that produce TAGs with EPA in the sn2 position would be a better choice than overexpressing native DGAT genes of IMET1. The lipidomic analysis of *Nannochloropsis* sp. PJ12 shows the presence of TAG species with EPA in all the 3 positions [397]. Thus, the DGAT enzymes from this strain could be potential candidates to channel EPA rich DAGs to TAGs in IMET1. On the other hand, monogalactosyldiacylglycerol synthase (MGDG synthase) and sulfoquinovosyltransferase (SQD2) catalyzes the conversion of DAG into polar lipids MGDG and sulfoquinovosyldiacylglycerol (SQDG), respectively. Characterization of these enzymes in *N. oceanica* IMET1 could reveal their specificity towards various DAG species. Downregulation of MGDG synthase and SQD2 with high specificity towards EPA rich DAG molecules could also be a strategy to improve the conversion of these DAG molecules into TAG.

The EPA could accumulate in TAGs via *de novo* fatty acid synthesis or reshuffling of the fatty acids between other lipid classes [293]. The *de novo* EPA is synthesized in the ER using PLs such as 1,2-dipalmitoyl-sn-glycero-3-O-4'-(N,N,N-trimethyl)-homoserine (DGTS) and Phosphatidylethanolamine (PE) as the carriers [398]. Our PL fatty acid profile indicate that MGDG, DGTS and digalactosyldiacylglycerol (DGDG) could be the major pools of EPA under stress. Strategies to translocate these EPA from PLs to TAG under nitrogen stress could also be an interesting approach to enhance EPA content in NLs. In *Chlamydomonas*, a galactoglycerolipid lipase coded by the gene *PGD1* was reported to be responsible for partial translocation of plastid lipids to TAG under nitrogen starved conditions [399]. Identifying potential enzymes that could hydrolyze the EPA rich PLs under stress to release the EPA for TAG assembly could improve EPA rich TAG molecules. Although the nitrogen stress is reported to translocate EPA from PL into NL in modest amounts in *Nannochloropsis*, a considerable amount of EPA is also degraded at similar rates [293]. In this regard, transcriptomic analysis to identify the enzymes that are upregulated under nitrogen stress and catalyze the breakdown of fatty acids could indicate potential targets for further metabolic engineering. Knocking down these enzymes could be a straightforward approach to reduce the EPA degradation under nitrogen stress.

Conclusion

EPA is a commercially-relevant fatty acid that can be sustainably produced from microalgae *N. oceanica*. Even though EPA is predominantly synthesized under optimal conditions, they are accumulated in TAGs under nitrogen stress. TAG molecules act as the carriers of LC-PUFAs in the diet, and EPA-rich TAG species are attractive food supplements [400]. Thus, developing TAGs with tailored EPA or LC-PUFA content could have interesting applications in the nutraceutical industries [400]. In this regard, the native DGAT genes from *N. oceanica* were recently characterized to reveal the specificity of each DGAT gene towards various fatty acids. The expression of the DGAT genes was demonstrated to be controlled to tailor the fatty acid species at the sn-3 position of TAG [400]. Similarly, six *LPAAT* coding genes are annotated in *N. oceanica* [74]. Characterizing these native *LPAAT* proteins and elucidating their affinity towards fatty acids could further improve the production of tailored TAG molecules in *Nannochloropsis* and other relevant strains. As a shortcut to this approach and to improve the EPA content in TAGs, we expressed in *N. oceanica* an EPA specific *LPAAT* protein from the plant *B. napus* [376].

The BnLPAAT expression in *N. oceanica* was validated by checking the expression of the Nlux tag linked at the C' terminal of the protein. The mutant strains had reduced growth and biomass accumulation under nitrogen starvation compared to the WT. Nevertheless, the EPA content in the TFA and TAG of mutant B was improved by 28% and 38%, respectively, under nitrogen stress. In the PLs, the EPA contents were not improved in mutants under any conditions. However, the concentration of medium-chain fatty acids C16 and C16:1 was increased in mutant A under nitrogen-starved conditions. Further studies will be performed to evaluate the substrate specificity of BnLPAAT to confirm these variations are dependent on the heterologous expression of this protein.

Our study shows that the BnLPAAT expression in *N. oceanica* could improve the EPA content in TAG under stress. Nevertheless, the major TAG species produced during BnLPAAT expression should be characterized to confirm that the EPA has been introduced into the sn-2 position, resulting in improved TAG content. Moreover, it is evident that the EPA molecules are synthesized in the PLs and DAG molecules with EPA could be primarily channeled towards PLs. Therefore, co-expression of DGAT molecules with high affinity towards EPA rich DAGs and strategies to translocate EPA in PLs to TAGs during nitrogen starvation could further improve the EPA contents in TAG.

Acknowledgments

The project is funded by the project 'Microalgae as a Green source from Nutritional Ingredients for Food/Feed and Ingredients for Cosmetics by Cost-Effective New Technologies' (MAGNIFICENT), funded by the Bio-based Industries Joint Technology Initiative under the EU Horizon 2020 Research and Innovation Program (project 745754)

Supplementary data

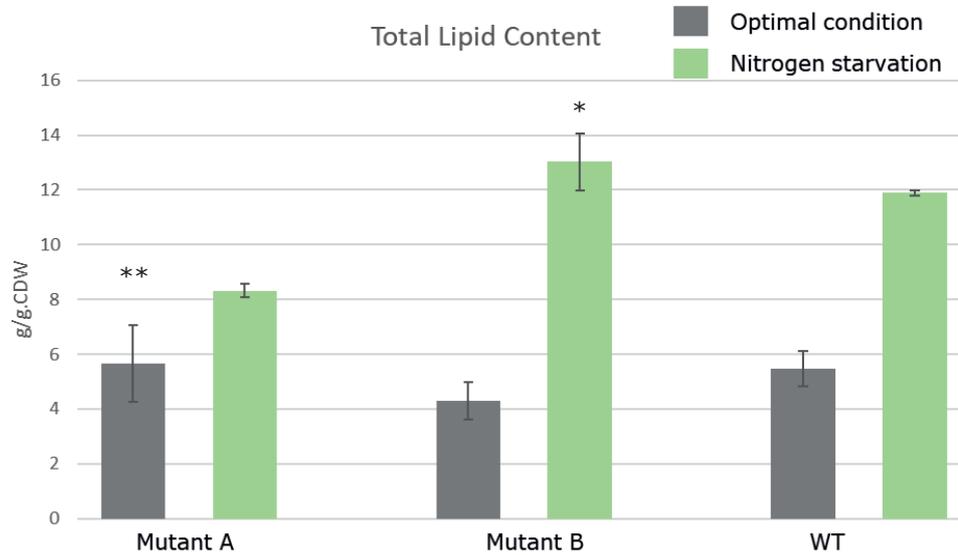


Figure S1. Total lipid content in the BnLPAAT expressing mutants and WT IMET1 strains under optimal and nitrogen starved growth conditions. The values and SD are based on $N \geq 2$. The p-value of each sample group was tested for statistical significant with the WT using two sample t-test to confirm the p-value is < 0.1 . (* p-value < 0.2 , ** p-value > 0.5 (not significant)).

Table S1. Fatty acid content in the TFA or total lipid isolated from the mutants expressing BnLPAAT and WT strains of *N. oceanica* IMET1. The average values and SD are based on sample size $N \geq 2$.

Fatty acid	Optimal condition					
	Mutant A		Mutant B		WT	
	mg/g.CDW	S.D	mg/g.CDW	S.D	mg/g.CDW	S.D
C14:0	4,05	0,54	3,76	0,11	4,10	0,17
C16:0	21,92	3,85	14,42	1,18	18,64	0,29
C16:1	22,53	3,17	14,22	0,69	16,65	0,56
C18:0	1,13	0,18	0,62	0,03	0,79	0,03
C18:1	3,61	0,55	2,12	0,14	2,64	0,05
C18:2	1,30	0,21	1,18	0,01	1,64	0,09
C20:4-n6	1,69	0,42	1,76	0,32	2,26	0,09
C20:5-n3	7,36	1,05	8,27	0,85	11,03	0,43

Fatty acid	Nitrogen starved					
	Mutant A		Mutant B		WT	
	mg/g.CDW	S.D	mg/g.CDW	S.D	mg/g.CDW	S.D
C14:0	5,74	0,81	11,70	0,98	7,90	0,08
C16:0	36,56	0,20	51,96	4,40	53,78	0,66
C16:1	25,90	1,31	40,33	3,17	33,78	0,18
C18:0	1,02	0,01	1,30	0,10	1,20	0,01
C18:1	7,55	0,31	10,49	0,75	10,98	0,17
C18:2	0,95	0,08	1,53	0,12	1,57	0,02
C20:4-n6	1,36	0,10	3,22	0,18	2,05	0,07
C20:5-n3	3,51	0,09	8,27	0,65	6,49	0,18

Table S2. Neutral lipid (NL) and Polar lipid (PL) classes in the BnLPAAT expressing mutant and WT strains of IMET1. The value and SD are based on $N \geq 2$.

Lipid Class	Optimal condition					
	Mutant A		Mutant B		WT	
	g/g.CDW	S.D	g/g.CDW	S.D	g/g.CDW	S.D
NL	3,02%	0,66	2,36%	0,26	2,81%	0,17
PL	3,50%	1,25	4,45%	0,95	4,76%	0,34
Nitrogen starved						
NL	9,85%	0,16	10,39%	0,52	9,15%	0,17
PL	4,16%	0,46	2,76%	0,95	2,33%	0,75

Table S3. Neutral lipid (NL) class content (Triglycerides, diglycerides, and free fatty acids) in the BnLPAAT expressing mutant and WT strains of IMET1 under optimal and nitrogen starved conditions. The value and SD are based on N \geq 2.

Fatty acid	Optimal condition						
	Mutant A		Mutant B		WT		
	mg/g.CDW	S.D	mg/g.CDW	S.D	mg/g.CDW	S.D	
TAG	30,65	1,60	19,39	1,98	22,20	1,73	
DAG	0,12	0,00	0,27	0,02	0,28	0,06	
FFA	3,86	0,24	3,90	0,80	4,48	0,36	
Fatty acid	Nitrogen starved						
	TAG	95,02	1,47	99,90	4,10	85,83	1,79
	DAG	1,31	0,11	2,70	0,46	3,03	0,50
	FFA	2,20	0,01	3,73	0,19	2,08	0,15

Table S4. Fatty acid profile (FFA) of the free fatty acids obtained from the TLC of Neutral Lipids in the BnLPAAT expressing mutant and WT strains of IMET1 under optimal and nitrogen starved conditions. The value and SD are based on N \geq 2.

Fatty acid	Optimal condition					
	Mutant A		Mutant B		WT	
	mg/g.CDW	S.D	mg/g.CDW	S.D	mg/g.CDW	S.D
C14:0	0,41	0,04	0,45	0,01	0,47	0,04
C16:0	1,49	0,15	1,16	0,04	1,20	0,12
C16:1	1,12	0,05	1,08	0,05	1,21	0,14
C18:0	0,49	0,05	0,41	0,05	0,30	0,02
C18:1	0,00	0,00	0,19	0,02	0,18	0,03
C18:2	0,00	0,00	0,00	0,00	0,00	0,00
C20:4-n6	0,00	0,00	0,15	0,00	0,17	0,01
C20:5-n3	0,12	0,12	0,94	0,31	0,82	0,07

Fatty acid	Nitrogen starved					
	Mutant A		Mutant B		WT	
	mg/g.CDW	S.D	mg/g.CDW	S.D	mg/g.CDW	S.D
C14:0	0,17	0,01	0,33	0,01	0,14	0,01
C16:0	0,97	0,04	1,53	0,04	0,90	0,07
C16:1	0,65	0,04	1,28	0,02	0,67	0,04
C18:0	0,21	0,01	0,14	0,01	0,12	0,02
C18:1	0,16	0,01	0,20	0,01	0,13	0,01
C18:2	0,00	0,00	0,00	0,00	0,00	0,00
C20:4-n6	0,00	0,00	0,00	0,00	0,00	0,00
C20:5-n3	0,00	0,00	0,12	0,00	0,11	0,02



CHAPTER 7



Synthetic biology approaches to enhance microalgal productivity

Mihris Ibnu Saleem Naduthodi ^{*1 2}, Nico J. Claassens^{*1}, Sarah D'Adamo², John van der Oost¹ and Maria Barbosa²

¹Laboratory of Microbiology, Wageningen University and Research, Wageningen, Netherlands.

²Bioprocess Engineering, AlgaePARC, Wageningen University and Research, Wageningen, Netherlands.

*Contributed equally

Adapted from publication Trends in Biotechnology (2021).
<https://doi.org/10.1016/j.tibtech.2020.12.010>

Abstract

The major bottleneck in commercializing biofuels and other commodities produced by microalgae is the high cost associated with phototrophic cultivation. Improving microalgal productivities could be a solution to this problem. Synthetic biology methods have recently been used to engineer the downstream production pathways in several microalgal strains. However, engineering upstream photosynthetic and carbon fixation metabolism to enhance growth, productivity, and yield has barely been explored in microalgae. We describe strategies to improve the generation of reducing power from light, as well as to improve the assimilation of CO₂ by either the native Calvin cycle or synthetic alternatives. Overall, we are optimistic that recent technological advances will prompt long-awaited breakthroughs in microalgal research.

Highlights

- The high cost of microalgal cultivation has hindered the exploitation of their sustainable benefits for producing green chemicals and biomass. Nevertheless, recent advances in the field of synthetic biology could help overcome the associated bottlenecks.
- Improving the reducing power generation and carbon influx will be crucial to attain overall improvement in the microalgal productivities.
- Enhancing the light absorption along with techniques to swiftly channel the electrons through electron transport chain (ETC) could enhance the reducing power generation
- Calvin-Benson-Bassham (CBB) cycle might not be the best CO₂ fixation pathway. Other natural and synthetic pathways could outperform CBB cycle. But, implementing these entire pathways into new hosts will be very challenging.
- Mixotrophic cultivation and microbial electrosynthesis (MES) could be implemented as an additional source of energy and carbon to improve the microalgal productivities.

Glossary

2A signal peptide: These are specific amino acid sequences that are introduced in between the coding DNA sequences of 2 genes expressed from a single promoter. These sequences induce a ribosomal skipping during[401] the translation of the mRNA leading to expression of multiple proteins from a single mRNA sequences. These sequences can be effectively used for expressing multiple protein from a single mRNA when there are limitations in the availability of regulatory elements for a particular species.

CrCAO: *Chlamydomonas reinhardtii* Chlorophyllide a oxygenase is the enzyme that is responsible for the production of chlorophyll b. The enzyme uses the precursor of

chlorophyll a (chlorophyllide a) as the substrate for the reaction to produce chlorophyllide b. Chlorophyllide b is further converted to chlorophyll b by another enzyme chlorophyll synthase.

CRISPR-Cas: Clustered Regularly Interspaced Palindromic Repeats and CRISPR associated proteins is an immune system present in bacteria and archaea that protects against invading viruses and bacteriophages. This immune system incorporates short fragments of viral DNA into the host genome called CRISPR array. Multiple short viral fragments are separated by a recurring repeat sequence and the entire CRISPR array is transcribed and cleaved into single CRISPR RNAs by host machinery. The single CRISPR RNA and the Cas protein forms a complex and this complex introduces a precise double stranded DNA break (DSDB) when it encounters the viral DNA with the same sequence present in the CRISPR RNA. This ability of Cas proteins to introduce precise DSDB was exploited to develop the genome editing tools that are presently used across various organisms.

Gene stacking: The process of expressing multiple heterologous proteins in a single host is termed gene stacking. Recycling the antibiotic resistance for selection, using multiple antibiotics in a single or multiple rounds of transformation, implementing 2A peptides for multi-cistronic expression or bi-directional promoters are some of the strategies to advocate the gene stacking in microalgae.

Photosynthetic efficiency (PE): During the process of photosynthesis, the light energy is converted into chemical energy to fuel host metabolism. PE is usually determined as the percentage of incident light energy that is converted to chemical energy for biomass production. In nature, only 50% of total solar radiation belongs to the photosynthetically active radiation (PAR), in other words the fraction of total solar radiation that can be used by photosynthetic organisms to perform photosynthesis. The theoretical maximum of PE is reported to be 11%, whereas the maximum overall PE is reported to be 3–6% in plants and microalgae.

Towards Improving Upstream Photosynthetic and CO₂ Fixation Metabolism

Research on microalgae as a platform for biofuel production dates back to the 1980s [402]. Despite decades of research, the cost of microalgal biofuel is still considerably higher than the biofuel produced from traditional agricultural crops. The expensive cultivation of microalgae in large-scale photobioreactors has been a bottleneck for economically feasible production of low-value commodities. Hence, microalgae are currently only used to produce high-value compounds such as pigments and fatty acids, focusing on niche markets with limited competition [403, 404]. However, the massive deforestation and other environmental consequences associated with the production of biofuels and biomass from traditional crops urges alternative sustainable production methods. Hence, research to realize the full potential of microalgae for sustainable production of biomass and products remains an urgent topic. Current genetic engineering strategies to improve microalgal productivity of lipids, which are the precursors for biofuel production, have often resulted in reduced biomass production and/or growth rate. At least to some extent this is because the supply of carbon or energy is insufficient to meet the increased demands of engineered strains (Figure 1) [80, 401, 405]. Thus, to overcome carbon or energy shortage, further major improvements in microalgal productivities will require extensive rewiring of the upstream metabolic pathways such as carbon fixation, as well as the generation of ATP and reducing power by photosynthesis. Some of these strategies have been recently proposed and partly implemented in plants (reviewed in [406-408]). These strategies require varying levels of genetic engineering interventions ranging from up- or downregulation or knockout of one or a few native genes to the introduction of complete, complex non-native systems such as synthetic photosystems or synthetic carbon fixation pathways. Recent improvements in the synthetic biology toolbox for microalgae, for example CRISPR/Cas (see Glossary) technology (Box 1) [233], now allow similar explorations in microalgae, potentially leading to large improvements in their productivity.

In this review we elaborate on synthetic biology-based approaches for enhancing the light-driven generation of reducing power and CO₂ fixation to improve overall microalgal productivity. Moreover, we discuss potential 'mixotrophy' strategies to utilize inorganic and organic molecules, as a complement or alternative to light and CO₂, as energy and carbon sources to enhance microalgal productivities.

Improving the Generation of Reducing Power from Light

Limiting the losses in the conversion of solar energy during photosynthesis could provide additional energy for host metabolism. Some of the solar energy absorbed by the photosystems is lost via a process termed non-photochemical quenching (NPQ), which is influenced by the lag in the adaptation of photosystem antenna sizes to the light conditions (Box 2). For these reasons, it has been proposed to reduce the antenna size by genetic engineering as a strategy to reduce the NPQ and improve photosynthetic efficiency (PE). Truncation of light-harvesting antenna and disruption of the chloroplast signal recognition particle (CpSRP) protein have been the primary targets for reducing

antenna size. CpSRP is responsible for translocating light-harvesting proteins to the thylakoid membrane of the chloroplast. Truncation of antennas led to improved PE and biomass productivity in several microalgal species (Table S2). Contrasting observations were also made in some cases where reduction in antenna size did not make a difference or adversely affected the PE [409, 410]. A more refined strategy to optimize light capture and PE would be to fine-tune the content of chlorophyll pigments (Chl). However, determining the optimal values of pigment content, especially for cultures grown under fluctuating light conditions, is impractical and could detrimentally affect engineered strains with adapted pigment and antennae composition in some conditions [411].

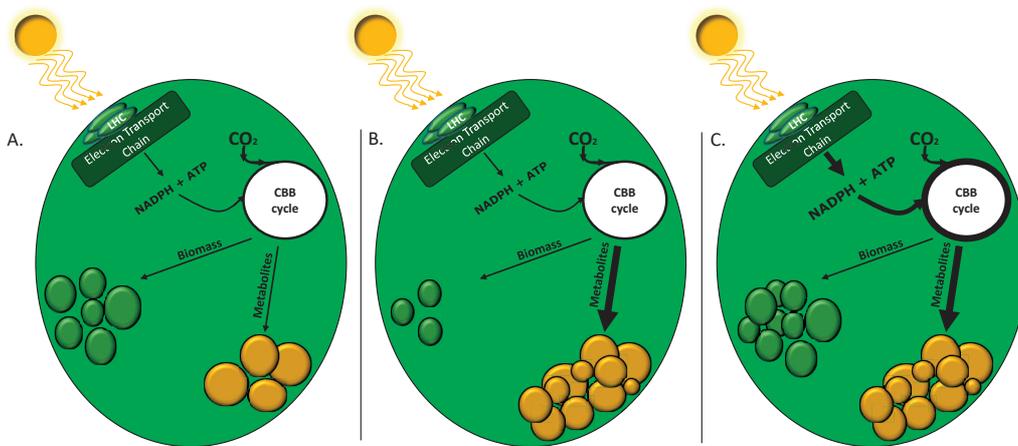


Figure 1. Schematic representation of algal carbon partitioning. (A) The Carbon dioxide fixed via photosynthesis is the sole source of carbon in photoautotrophs and the natural metabolism split the carbon into production of metabolites such as polyunsaturated fatty acids (PUFAs) and Lipids and towards cell growth or biomass production. (B) Conventional metabolic engineering approaches to enhance the production of metabolites focus on overexpressing the genes in the production pathway of these metabolites, mostly resulting in an improvement in the metabolite titer and reduction in the growth of microalgae. (C) To enhance the production of metabolites without limiting the growth, the generation of reducing power and carbon fixation via photosynthesis should be improved. This could enhance the influx of carbon in microalgae and engineering these strains that has improved carbon fixation and reducing power generation to produce more metabolites could have improved production of metabolites without limiting the growth or biomass production.

An alternative strategy to improve PE in microalgae would be to introduce heterologous photosystems that have a potentially higher light-harvesting efficiency or a wider absorption spectrum. Plant and microalgal photosystems only absorb light in the visible range from 400 nm to 700 nm, whereas ~50% of the available solar radiation falls outside that range [412]. Some of the non-absorbed light is emitted in the more energetic part of the spectrum (UV, <400 nm), whereas most of the non-captured light is in the IR spectrum (>700 nm). Although photons >700 nm are less energetic, and are insufficiently energetic to perform water-splitting, photons up to 1100 nm are still sufficiently energetic to drive the generation of a typical proton motive force of 200 mV and generate ATP [413]. Several bacterial photosystems found in nature have higher light-harvesting efficiencies and broader absorption spectra. Components of these bacterial photosystems could

potentially be harnessed to improve PE. Complete rewiring of eukaryotic oxygenic photosynthesis to improve its efficiency has been proposed before [412, 414]; however, rebuilding the highly complex multi-subunit photosystems may be highly ambitious. Nevertheless, replacement of some parts of native photosystems by heterologous counterparts has already been demonstrated in microalgae.

A study along these lines replaced the D1 subunit in photosystem II (PSII) of *Chlamydomonas reinhardtii* by its counterpart from the cyanobacterium *Synechococcus elongatus*. The mutants showed that the photochemistry of the PSII system could be improved for some variants under particular conditions (low or high light) [415]. Other studies showed that heterologous expression of PSII subunits D1 and PsbH from plants in the cyanobacterium *Synechocystis* did not improve the PE of the engineered strains [416-420]. To date, swapping of PSII subunits has not improved PE in cyanobacteria and microalgae under photobioreactor conditions. One reason could be that replacement of single proteins within the native photosystem complexes is not sufficient because efficient photosystem operation requires specific interactions between multiple proteins in the photosystem. In another study, a larger set of six core proteins of PSII was replaced by the complete set of PSII core proteins from the microalgae *Scenedesmus obliquus* or *Volvox carteri*, or from the same strain of *C. reinhardtii* as a control. Complementation of the deleted PSII proteins successfully reconstituted 85%, 55%, and 53% of the photosynthetic activity upon expression of the core proteins from *C. reinhardtii*, *S. obliquus*, and *V. carteri*, respectively [421]. Although functional heterologous replacement of PSII was demonstrated, improved PE in microalgae has not yet been realized by PSII subunit engineering, and this will likely require more extensive rewiring and optimization of photosystems.

Other than introducing heterologous photosystem complexes, introducing the biosynthetic pathway for non-native pigments may increase the absorption range and PE. In plants, heterologous expression of photosynthetic proteins and pigment biosynthesis pathways that were lost during evolution resulted in improved photosynthesis, growth, and stress tolerance [422]. A similar strategy was recently followed for the improvement of PE in the microalga *Nannochloropsis salina*. The chlorophyllide a oxygenase (CrCAO) gene from *C. reinhardtii* was introduced to produce chlorophyll b (Chl b), which is not native to *N. salina*. This strategy resulted in a 26% higher cell number, 31% improved cell dry weight, and an 8% increase in total lipid content of *N. salina* under medium-light conditions after 12 days [423, 424]. The production of Chl b in *N. salina* indicates the potential of producing non-native pigments in microalgae.

Following this approach, the chlorophyll f synthase (ChIF) recently discovered in some cyanobacteria, that produces the far-red light-absorbing chlorophyll f (Chl f), could be an interesting candidate for expression in microalgae [425]. This can lead to an expansion of the light absorption spectrum into the IR wavelengths. Recent insights have revealed that ChIF, which is highly homologous to the D1 subunit of PSII, probably forms a heterodimer with the D2 subunit of PSII to synthesize Chl f [426]. Remarkably, it was shown that, by engineering chimeric D1/ChIF proteins in *Synechocystis sp.* PCC 6803, which natively does not harbor Chl f pigments, the engineered strain could synthesize Chl f pigments [426]. The chimera strategy could potentially be interesting to enable Chl f synthesis in

microalgae. However, further engineering will probably be necessary to achieve functional interaction of Chl f with the photosystems. In addition, UV-absorbing pigments found in Antarctic plants could further broaden the light spectrum accessible by microalgae upon heterologous expression, and possibly improve biomass or product yield [427]. However, the genes involved in the biosynthesis of these pigments remain to be elucidated.

Apart from improving the photosynthetic machinery, potential improvements could be achieved in the machinery that channels the photosynthetically excited electrons via various redox reactions to generate ATP: the electron transport chain (ETC) (Box 2). To improve ETC reactions, the rate-limiting protein complexes in the ETC should be identified for the species of interest, and the entire subunit could be replaced with a faster variant to improve electron flow and reduce NPQ [428]. Characterization of the ETC complexes in fast-growing microalgal and cyanobacterial strains could potentially reveal catalytically superior variants of these complexes. Taken together, inspired by the aforementioned studies, the synthetic biology tools could be exploited for stepwise improvement of the entire light reactions of photosynthesis, which could potentially dramatically enhance the microalgal productivity.

Improving Carbon Fixation

Overcoming the Limitations of Natural Carbon Fixation Systems in Microalgae

Microalgae assimilate CO₂ in the Calvin–Benson–Bassham (CBB) cycle via the carboxylating enzyme RuBisCO (ribulose-1,5-bisphosphate carboxylase-oxygenase) that has some shortcomings (Box 2), resulting in a relatively low rate and efficiency of carbon fixation. Multiple genetic engineering strategies have been implemented in various microalgal strains to partly overcome these shortcomings [408]. One strategy is to simply increase the levels of active RuBisCO enzyme to increase the overall carboxylation activity, assuming that the levels of functional protein are limiting growth. This strategy was reported in the microalgal strain *Nannochloropsis oceanica* where overexpression of native RuBisCO activase improved both growth rate (32%) and biomass accumulation (46%) [429]. In the cyanobacterium *Synechocystis* PCC 6803, overexpression of native RuBisCO improved growth rate and biomass production by ~20% [429, 430].

Alternatively, native RuBisCO enzymes could be replaced by either superior natural or engineered RuBisCO variants. To select variants, it is important to consider their catalytic activity and their CO₂/O₂ specificity because RuBisCO variants with improved catalytic activity often have reduced CO₂ specificity, and vice versa [431]. This ‘trade-off’ is typically observed among natural RuBisCO variants [432]; in addition, directed engineering of RuBisCO to improve its performance only led to limited overall carboxylation improvements under ambient conditions [433–435]. Therefore, especially when microalgae are cultivated in atmospheric CO₂ conditions, this trade-off needs to be taken into account. However, under elevated CO₂ concentrations that are often applied in photobioreactors, the use of RuBisCO variants with increased catalytic rates (but with lower CO₂/O₂ specificity) may be attractive to increase productivity.

To harness the power of potentially superior RuBisCO variants, attempts were made to express natural or engineered RuBisCO proteins in RuBisCO deletion mutants of bacteria [436, 437] and plants [438, 439], as well as in the microalgal species *Chlamydomonas reinhardtii* [440]. However, so far these studies led to only limited improvements in growth, probably because of the modest number of variants screened and challenges in heterologous expression for many multimeric (type I) RuBisCO variants that often require activases or chaperone proteins. Pyrenoids – the microalgal sub-compartments in which RuBisCO is often located to facilitate higher local CO₂ concentrations – are another complication facing the expression of heterologous RuBisCO in microalgae. It was recently determined that the pyrenoid compartment assembles by specific binding of pyrenoid-associated proteins to RuBisCO [441]. This assembly is likely disrupted when replacing a native RuBisCO by a heterologous candidate. However, the binding regions of the RuBisCO variant or of the pyrenoid-associated proteins could be engineered to support pyrenoid formation with a heterologous RuBisCO.

A potentially promising type I RuBisCO variant was discovered in the red alga *Griffithsia monilis* that has an unusually high CO₂/O₂ specificity factor of 167, from two- to fourfold higher than the values typically reported for C3 plants, green algae, and cyanobacteria [432, 442-444]. However, the requirement for post-translational modifications of the *G. monilis* RuBisCO proteins and for the assembly of multiple subunits of this RuBisCO enzyme may explain its poor heterologous expression in tobacco plants [444]. Recent developments indicate that co-expression of chaperones that facilitate the assembly of specific RuBisCO enzymes could enable the heterologous expression of these proteins [445]. Therefore, developing techniques to assemble the red alga RuBisCO in microalgae could be an interesting approach towards enhancing CO₂ fixation.

By contrast, the simpler homodimeric type II RuBisCO enzymes may be a better choice for heterologous expression. Metagenomic studies have revealed numerous RuBisCO-like proteins from non-domesticated prokaryotic and eukaryotic organisms [446, 447]. Recently, 33 000 putative RuBisCO proteins were identified by this approach, from which 143 type II and type II/III RuBisCOs were selected and expressed in *Escherichia coli* and screened by in vitro assays to obtain catalytically superior variants [448]. The most active variant was a type II variant obtained from the bacterium *Gallionella* sp., which exhibited an eightfold improved turnover number compared to typical values reported for type I RuBisCO from green algae and plants [443]. However, the low affinity for CO₂ of this bacterial RuBisCO (K_m of 276 μM, compared to 39 μM in *Chlamydomonas* and 14 μM in plants) reduces the carboxylation efficiency compared to plants under normal atmospheric conditions [443, 448]. Nevertheless, this protein could be a suitable candidate for culturing microalgae in reactors with CO₂ supplementation.

In addition to RuBisCO, it is important to identify other rate-limiting steps in the CBB cycle to achieve optimal CO₂ fixation. The CBB cycle reactions catalyzed by sedoheptulose bisphosphatase (SBP) and fructose bisphosphate aldolase (FBA), in addition to RuBisCO, are reported to be rate-limiting steps in plants and cyanobacteria [449, 450]. Overexpression of other CBB enzymes was reported to enhance photosynthesis and growth rate in the cyanobacteria *Synechocystis*, *Synechococcus*, and *Anabaena* spp. [430, 449, 451].

Respective overexpression of the cyanobacterial SBP and FBA in the microalgae *Chlorella vulgaris* and *Euglena gracilis* improved their biomass production by 1.2- and twofold, respectively [354, 452]. By contrast, overexpression of native fructose 1,6-bisphosphatase in *C. reinhardtii* adversely affected growth rate and biomass production [453]. Obtaining improved variants of these enzymes via directed evolution and metagenomic studies should also be considered, alongside improvements to RuBisCO, as a means to increase the CO₂ fixation via CBB cycle.

Decreasing Carbon Loss via Photorespiration

As described in Box 3, the oxygenation activity of RuBisCO, in addition to competing with the carbon fixation reaction, generates the toxic compound 2-phosphoglycolate (2-PG), especially under atmospheric CO₂ concentrations. Hence, engineering more efficient alternative photorespiration pathways for recycling 2-PG could be a strong synthetic biology approach to tackle the losses of photorespiration in microalgae (Figure 2). Such approaches were already successfully demonstrated in plants, but not yet in microalgae. For example, the more energy-efficient glycerate photorespiration pathway, that is present in some bacteria [454, 455], was introduced into *Arabidopsis thaliana* (arabidopsis) chloroplasts. The introduction of five genes encoding this pathway improved arabidopsis biomass and growth rate [456]. The malate cycle is an alternative photorespiration pathway that could be considered [455]. This pathway leads to complete decarboxylation of 2-PG, but could increase CO₂ concentrations within RuBisCO-containing chloroplasts, and was therefore engineered into some plant species. This malate cycle improved biomass yield and photosynthetic rates in both arabidopsis [457] and tobacco [458], in the latter case in field studies PE and biomass production were impressively increased by 17% and 24%, respectively.

The alternatives discussed above for natural photorespiration pathways still release CO₂. Following a more ground-breaking approach, different synthetic pathways were proposed based on engineered enzymes that could lead to CO₂-neutral photorespiration. Such promising pathways, primarily based on the reduction of glycolate to glycolaldehyde via glycolyl-CoA with engineered enzymes (glycolyl-CoA synthetase and glycolyl-CoA reductase or glycolyl-CoA carboxylase), were so far only demonstrated in vitro [459, 460]. However, these promising designs for synthetic photorespiration await in vivo testing in microalgae and other photosynthetic organisms.

Like many other photosynthetic organisms, microalgae have naturally evolved carbon concentrating mechanisms (CCMs) to reduce the effects on photorespiration. For microalgae, this CCM involves multiple steps such as bicarbonate uptake, carbonic anhydrase-mediated interconversion of bicarbonate and CO₂, and localization of RuBisCO in pyrenoid compartments within the microalgal chloroplast [461, 462]. Because CCMs involve multiple steps and pathways, the engineering strategies for improving this system are in the early stages. Overexpressing a bicarbonate transporter from *C. reinhardtii* improved the growth rate and biomass accumulation in *N. salina* by increasing the intracellular inorganic carbon concentration. However, the overexpression of the same gene in *C. reinhardtii* did not yield significant improvements [463, 464].

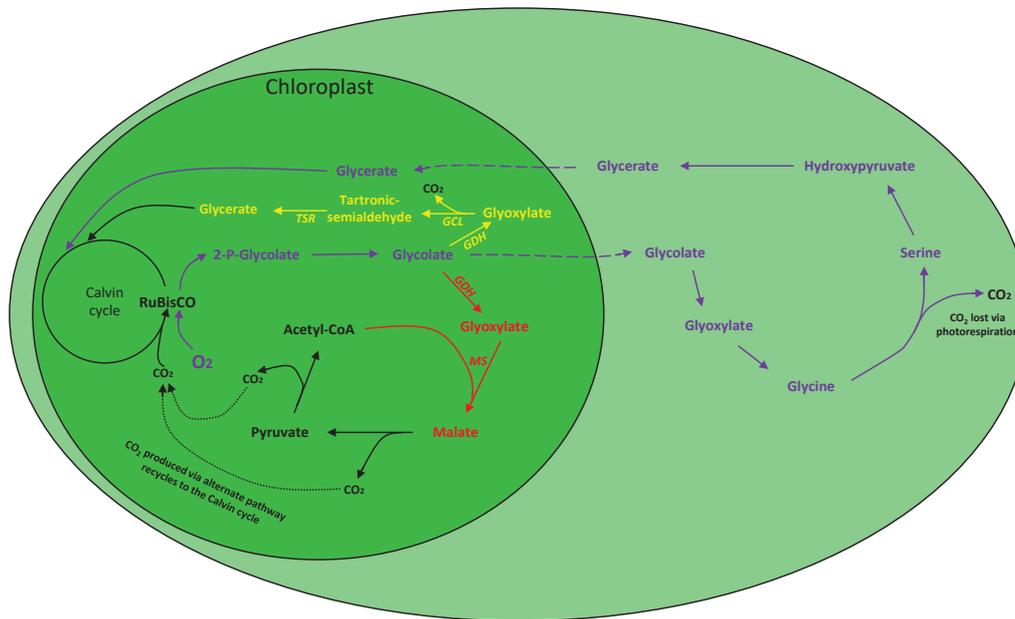


Figure 2. Natural and synthetic alternative photorespiratory pathways: The pathway in purple indicates the natural photorespiratory pathway that transports glycolate outside the chloroplast and convert it back into glycerate before transporting it back into chloroplast for the Calvin cycle. During this process carbon dioxide is released resulting in a net loss of fixed carbon. Two synthetic alternative photorespiratory pathways demonstrated in C₃ plants to improve the biomass are indicated in yellow and red. The pathway in yellow implements 3 enzymes from the *E. coli* glycolate pathway i.e. GDH: glycolate dehydrogenase, GCL: glyoxylate carboxyligase and TSR: Tartronic-semialdehyde reductase. The pathway in red implements 2 enzymes i.e. GDH from *C. reinhardtii* and MS: malate synthase from *Cucurbita maxima*. In the synthetic pathways, the carbon dioxide is released back in the chloroplast and this can be again taken up by the RuBisCO for the Calvin cycle thereby minimizing the carbon loss via photorespiration.

Synthetic Carbon Fixation

Engineering an alternative and more efficient CO₂ fixation pathway is a promising and challenging, radical strategy to overcome the shortcomings of the RuBisCO enzyme, the CBB cycle, and photorespiration in microalgae (Table 1 and Box 4).

Some of the alternative carbon fixation pathways, including the highly ATP-efficient Wood–Ljungdahl pathway, include oxygen-sensitive enzymes, making them poorly compatible with oxygenic photosynthesis in microalgae. However, there are also aerotolerant natural carbon fixation pathways, of which the reductive glycine pathway (via serine) is probably the most attractive and ATP-efficient. In this pathway, CO₂ is first reduced to formate, which in subsequent steps is activated and further carboxylated to glycine, that can next be converted to serine and pyruvate, only consuming 2 ATP/pyruvate. The two carboxylation reactions in this pathway are both thermodynamically reversible, meaning that this pathway can only function under elevated CO₂ concentrations, which is often the case in photobioreactor setups. It has been recently demonstrated that this linear pathway can be engineered into heterologous hosts, at least for the assimilation of formate. The

pathway has been successfully implemented in the bacteria *E. coli* and *Cupriavidus necator* by using a combination of modular engineering and adaptive laboratory evolution [465, 466]. The core part of the reductive glycine pathway, the production of glycine from formate, was also demonstrated to be functional in the yeast *S. cerevisiae* [467]. Although this pathway has not yet been introduced in photosynthetic organisms for the fixation of CO₂, the ubiquitous nature of all the enzymes involved and its simple linear nature make it a promising pathway to be also introduced into microalgae.

Moreover, the 3-hydroxypropionate bi-cycle and the 3-hydroxypropionate-4-hydroxybutyrate pathways can also operate under aerobic conditions. Both pathways fix CO₂ with propionyl-CoA carboxylase and acetyl-CoA carboxylase, which use the more soluble HCO₃⁻ instead of CO₂ as carbon source and may have advantageous kinetics over RuBisCO [468]. However, both systems require the generation of relatively complex and circular pathways involving many enzymatic steps, thus complicating their engineering, and especially in organisms with more limited genetic toolboxes such as microalgae. Although some modules of both pathways were successfully engineered into prokaryotic hosts, engineering the full pathways and associated autotrophic growth has not yet been achieved in vivo [469, 470].

In addition to the natural CO₂ fixation pathways, synthetic pathways can be designed based on the large collection of characterized natural enzymes and potential engineered enzymes to develop an efficient CO₂ fixation route (Box 4). Some of these pathways have been successfully characterized in vitro, but their full implementation in vivo, even in bacterial hosts, remains to be demonstrated. Altogether, even though the implementation of full synthetic CO₂ fixation pathways in microalgae may seem to be a far-fetched goal, impressive recent progress in engineering functional CO₂ pathways in several hosts is reassuring. Recently, genetic engineering approaches combined with adaptive laboratory evolution led to the establishment of a fully functional CBB cycle in the heterotrophic model bacterium *E. coli* [471], as well as in the yeast *Pichia pastoris* [472] that has a less advanced genetic toolbox. In both cases the introduction of the CBB cycle only required the addition of a limited number (3–8) of heterologous genes. However, this work, as well as the previously mentioned engineering of the reductive glycine pathway for formate assimilation, demonstrates the feasibility of the engineering of complete autotrophic pathways. Nevertheless, a major challenge in achieving this goal, especially for microalgae, is the major metabolic engineering effort that will be necessary to express and localize the high number of, mostly foreign, enzymes involved. However, recent developments in the genome-editing toolbox of microalgae are very promising. The characterization of inducible promoters, the 2A signal peptide-based system for multi-cistronic expression of heterologous proteins, gene stacking techniques based on bidirectional promoters, the identification of resistance to multiple antibiotics, Cre recombinase-based marker recycling, and, last but not least, CRISPR-associated nucleases for precise genome editing (Box 1), are some recent developments that could be exploited for expressing a large number of heterologous proteins in microalgal strains [81, 182, 231, 473-476]. We believe that the growing knowledge about efficient and transplantable CO₂ fixation pathways, together with state-of-the-art techniques for heterologous protein expression and for adaptive evolution, should be combined to develop microalgal strains with improved CO₂ fixation ability.

Mixotrophic Routes for Carbon and Energy Source

Cofeeding of Organic Substrates

In addition to cultivating microalgae under photoautotrophic conditions, some microalgae can also be grown heterotrophically owing to their capacity to utilize exogenous organic carbon sources. The concomitant use of light/CO₂ and organic carbon sources is termed mixotrophy. Mixotrophic cultivation of microalgae can lead to improved growth rates in comparison to phototrophic growth owing to the presence of additional carbon and energy sources [477]. However, mixotrophy does not consistently improve lipid or biomass yield across different microalgal species [478]. For example, photoautotrophic cultivation of marine *Chlorella* sp., *Nannochloropsis* sp., and *Chlorella vulgaris* UTEX 259 yielded higher lipid content compared to heterotrophic and mixotrophic cultivation [479, 480]. These studies indicate that improvement of biomass or product yield in microalgae by mixotrophy will be species- and product-dependent.

Mixotrophic cultivation could also be a potential strategy to eliminate the energy-intensive process of aeration in photobioreactors that is required for CO₂ supply and oxygen release from the system [481]. The CO₂ required for photosynthesis can be generated by heterotrophic metabolism whereas the O₂ requirement for heterotrophic growth can be met by oxygenic photosynthesis, thereby creating an internal gas circulation to maximize the yield from substrate and avoid adverse oxygen accumulation [482]. An innovative mixotrophic cultivation strategy was recently reported to double photobioreactor productivity by completely eliminating the gas-liquid transfer of CO₂ and O₂ [481]. In this case, the CO₂ generated by heterotrophic metabolism of acetic acid was employed to fuel the CO₂ requirement of photosynthesis. In addition, phototrophic oxygen production maintained the dissolved oxygen concentration (DOC) for the heterotrophic metabolism of acetic acid. The heterotrophic biomass yield observed in this study was 0.50 (C-molbiomass per C-molsubstrate), which is close to the theoretical maximum of 0.7 for aerobic heterotrophic organisms. The overall mixotrophic biomass productivity was almost twofold higher than for purely autotrophic growth in the same reactor setup [481].

In practice, the applicability of mixotrophy is limited to microalgal strains that can take up and degrade particular organic carbon sources. This limitation could be resolved by the expression of carbon transporters and missing enzymes for organic carbon metabolism in the host species. Expression of human glucose transporter GLUT1 in *Phaeodactylum tricornutum* improved the cell concentration by fivefold under mixotrophic conditions [483]. In addition, *C. reinhardtii* expressing the glucose transporter Hup1 from *Chlorella kessleri*, when grown on media supplemented with 100 mM glucose in absence of light, doubled the cell density in 12 h and then remained stationary throughout the experiment, whereas the wild-type strain did not grow in dark [484]. This strategy could be employed for the trophic conversion and mixotrophic cultivation of commercially relevant microalgae. From a sustainability perspective, however, the advantage of direct conversion of CO₂ and sunlight as sustainable resources may be compromised by the reliance on organic (plant-based) carbon sources.

Direct or Shuttle-Based Cofeeding of Electrons

As an alternative or complement to light energy and organic substrates, electricity generated by renewable energy sources has been proposed as a sustainable biotechnological feedstock, and this could also be extended to microalgae [485]. Solar/photovoltaic (PV) cells can mediate more efficient conversions than biological water-splitting photosystems. Specifically, state-of-the-art PV systems can attain a solar-to-electricity efficiency of 20%, and can be combined with electrolysis to generate hydrogen at an overall solar to chemical energy conversion of 11%. By comparison, the maximum conversion efficiency of solar energy to biomass for microalgae is 3% when grown in outdoor bioreactors, and 1% for crop plants [412].

Feeding with (renewable) electricity for bioproduction has already been extensively explored in bacteria, and is mostly referred to as microbial electrosynthesis (MES) [486, 487]. The uptake of external electrons from electrodes is achieved by microbes such as *Shewanella oneidensis* that are naturally capable of extracellular electron transfer (EET) [488]. EET is facilitated via various types of mechanisms that transfer the external electron into the ETC of the host species [489].

Alternatively, mediator molecules such as hydrogen or formate could facilitate the transfer of electrons from cathodes to non-electroactive microbes [490, 491]. A proof of principle for mediator-based MES was demonstrated in the hydrogen-oxidizing bacterium *C. necator* by using hydrogen as the mediator of electron transfer. This study demonstrated biomass production with an overall solar energy to biomass conversion efficiency of 9.7%, compared to 3% and 6% for microalgae grown at pilot scale outdoors and under indoor laboratory conditions, respectively [492, 493].

In microalgae, MES could be a potential strategy to enhance carbon influx and reducing power generation. Direct electron uptake, or uptake of an electron mediator molecule such as hydrogen or formate, could improve NAD(P)H generation in microalgae. In the case of formate, the CO₂ released by formate oxidation in microalgae to generate NAD(P)H could also be fixed via the CBB cycle (Figure 3). Similar to the use of organic substrates, this system could alleviate the problems associated with light limitation or could be an additional energy source for culturing the microalgae at night. Although MES has solely focused on selected bacterial strains, the potential of this promising technology could be expanded to unicellular microalgae and cyanobacteria to improve the productivities of biomass and/or green chemicals.

Electricity-driven cultivation may already have been achieved in microalgae during pulsed electric field treatment – however, this was primarily to prevent bacterial contamination in algal cultures or to stimulate product release. For the microalgal and cyanobacterial species *C. vulgaris* and *Arthrospira plantensis*, respectively, this technique was reported to enhance cell growth, but it was not effective for *C. reinhardtii* [494, 495]. Various assumptions have been made regarding the underlying mechanism that results in improved growth, but this remains elusive [496]. *C. vulgaris* and *A. plantensis* might be naturally able to transfer electrons via mechanisms such as EET, and these may be lacking in *C. reinhardtii*. Further studies could open up interesting possibilities by

identifying microalgal species for which MES could be applied to improve the production of biomass or green chemicals. In addition to the direct supply of electrons, enzymatic conversion of dihydrogen could provide an interesting electron source. However, eukaryotic hydrogenase enzymes prefer the reduction reaction in which dihydrogen is produced, whereas prokaryotic hydrogenases preferentially oxidize the dihydrogen [497-499]. Implementing cyanobacterial hydrogenases that can directly uptake dihydrogen and utilize it as an electron source in microalgae could provide an additional source of electrons.

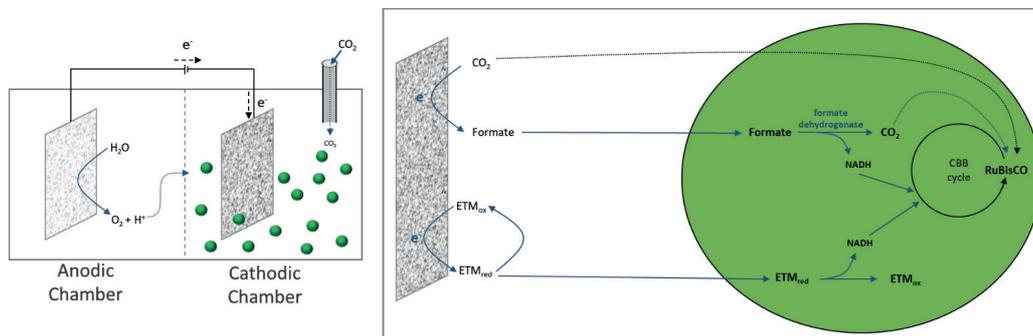


Figure 3. Conceptual MES in microalgae: The blue lines indicate the pathways involved in MES based generation of NADH and carbon dioxide. During the process, the water split at the anodic chamber releases the H⁺ ions and electrons that reduces the extracellular electron transfer mediators (ETM) in the cathodic chamber. The ETMs could directly enter the host species and undergo oxidation to release an additional NADH. Carbon dioxide that is electrochemically reduced to formate at the cathode could also be used as an electron donor if formate dehydrogenase is expressed in the host to achieve the MES via formate based pathway. The oxidation of formate releasing carbon dioxide could also be fixed into the Calvin cycle.

Concluding Remarks

The commercial production of microalgal products is hindered by the high cost of their phototrophic cultivation. Improving the influx of carbon as well as the generation of reducing power, while minimizing losses in their conversion after entering the host metabolism, should be the prime focus in tackling this issue. Host-specific studies on the CBB cycle and the identification of rate-limiting steps should stimulate the design of engineering strategies to improve CO₂ fixation capacity. Recent advances in metagenomics-based discoveries, in CRISPR-based genome editing, and in adaptive evolution strategies could be exploited to develop natural or synthetic CO₂ fixation pathways that will lead to microalgal hosts with improved productivity. Given the rapid development of the genome-editing toolbox for several microalgal strains, the challenging task of expressing the large number of enzymes involved in these pathways could be realized. Improving CO₂ fixation will require additional reducing power, and this can be produced either via alternative sources such as organic carbon and electricity or by improving the light reactions of photosynthesis. Expanding the spectrum of light available for photosynthesis

together with the swift flow of electrons via the ETC will be instrumental for attaining an optimal generation of reducing power.

Genetic components for improving the productivity of slowly growing model producer microalgae could be potentially obtained from fast-growing microalgae. By contrast, developing reliable genome-editing tools for these fast-growing strains could help in engineering them to exploit their improved photosynthesis for the production of interesting commodities [500]. *Picochlorum celerii* is an interesting fast-growing strain for this approach. This strain is reported to have a 10-fold faster doubling time (<2 h) compared to most of model microalgal species [501]. In addition, its three- to ninefold improved photosynthetic rates, the ability of the strain to withstand the variation in light intensities, and its resistance to photoinhibition indicate that it harbors interesting light reaction machineries, and potentially also interesting CBB enzyme variants [501]. Characterization of the light-harvesting complex (LHC) and ETC of this strain could pave the way for heterologous expression of these systems in other model microalgal organisms to improve the PE.

Taken together, the development of microalgal organisms as a major platform for green chemical production is challenging but feasible. Application of a single strategy (such as improving the CBB cycle or ETC, or reducing photorespiration), as has been done in the past, is probably insufficient to address current inefficiencies. To this end, the diverse strategies discussed in this review will need to be combined together in a model microalgal strain (Figure 4). Given the spectacular progress that has recently been made in the fields of synthetic biology and genetic engineering, crucial steps can now be taken towards engineering microalgae as a sustainable platform for the production of biomass and green chemicals.

CO ₂ fixing pathway	Natural / Synthetic (Aerobic/ Anaerobic)	Total number of reactions	ATP required per pyruvate produced	Carboxylating enzyme	Reference
Calvin cycle	Natural (Aerobic)	10	7	RuBisCO	[468, 502]
3-hydroxypropionate cycle	Natural (Aerobic)	16	7	Acetyl-CoA carboxylase and Propionyl-CoA carboxylase	[503]
Wood-Ljungdahl pathway	Natural (Anaerobic)	8	<1	Formate dehydrogenase and CO methylating acetyl-CoA synthase	[504]
Reductive TCA cycle	Natural (Anaerobic)	9	1-2	2-oxoglutarate synthase and Isocitrate dehydrogenase	[468, 505]
Dicarboxylate/4-hydroxybutyrate cycle	Natural	14	5	Pyruvate synthase and PEP carboxylase	[468, 506]
3-hydroxypropionate/4-hydroxybutyrate cycle	Natural (Aerobic)	16	5	Acetyl-CoA carboxylase and Propionyl-CoA carboxylase	[468, 507]
Reductive glycine pathway (via serine)	Natural (Aerobic)	7	2	Formate dehydrogenase and glycine synthase	(Sánchez-Andrea et al. 2020)
Reductive glycine pathway (via glycine reductase)	Natural (Anaerobic)	8	1-2	Formate dehydrogenase and glycine synthase	(Sánchez-Andrea et al. 2020)
C-4 glyoxylate cycle (MOG pathway)	Synthetic (Aerobic)	14	10	PEP carboxylase	[468]
		13	6	Pyruvate carboxylase	
CETCH 5.4	Synthetic (Aerobic)	13	12	Crotonyl-CoA carboxylase / reductase and Propionyl-CoA carboxylase	[508]
CETCH 7.0	Synthetic (Aerobic)	14	6-9	Crotonyl-CoA carboxylase / reductase and Propionyl-CoA carboxylase	[509]

Table 1: CO₂ fixation pathways. Abbreviations: CETCH, crotonyl-CoA/ethylmalonyl-CoA/hydroxybutyryl-CoA; MOG, malonyl-CoA/oxaloacetate/glyoxylate; PEP, phosphoenol pyruvate; TCA, tricarboxylic acid. For the production of 1 pyruvate via the CETCH 5.4 and glyoxylate carboxylase pathway, 2 ATP are required directly and 4 NADPH are additionally consumed for the production of pyruvate, which are not required in any of the other pathways [only 5 NAD(P)H or other types of reducing equivalents are necessary to reduce 3 CO₂ to 1 pyruvate]. These extra electrons are lost in the two oxidation reactions in the CETCH 5.4 pathway. Assuming a P/O ratio of 2.5, these 4 NADPH equivalents add an additional indirect cost of 10 ATP equivalents. For the production of 1 pyruvate via the CETCH 6.0/7.0 and glyoxylate carboxylase pathways, 4 ATP are required directly and 2 NADPH are additionally consumed for the production of pyruvate. The extra required electrons are lost in an oxidation reaction in the CETCH 6.0 pathway (methylsuccinyl-CoA oxidase). Assuming a P/O ratio of 2.5, these extra 2 NADPH equivalents add an additional cost of 5 ATP equivalents. However, in CETCH 7.0 the methylsuccinyl-CoA oxidase is replaced by methylsuccinyl-CoA dehydrogenase that can donate the electrons to ubiquinone, hence some additional ATP may be regenerated in an electron transport chain. We assume in this case that only one ATP is 'lost' by the partial oxidation of NADPH, hence only 2 extra ATP are consumed indirectly.

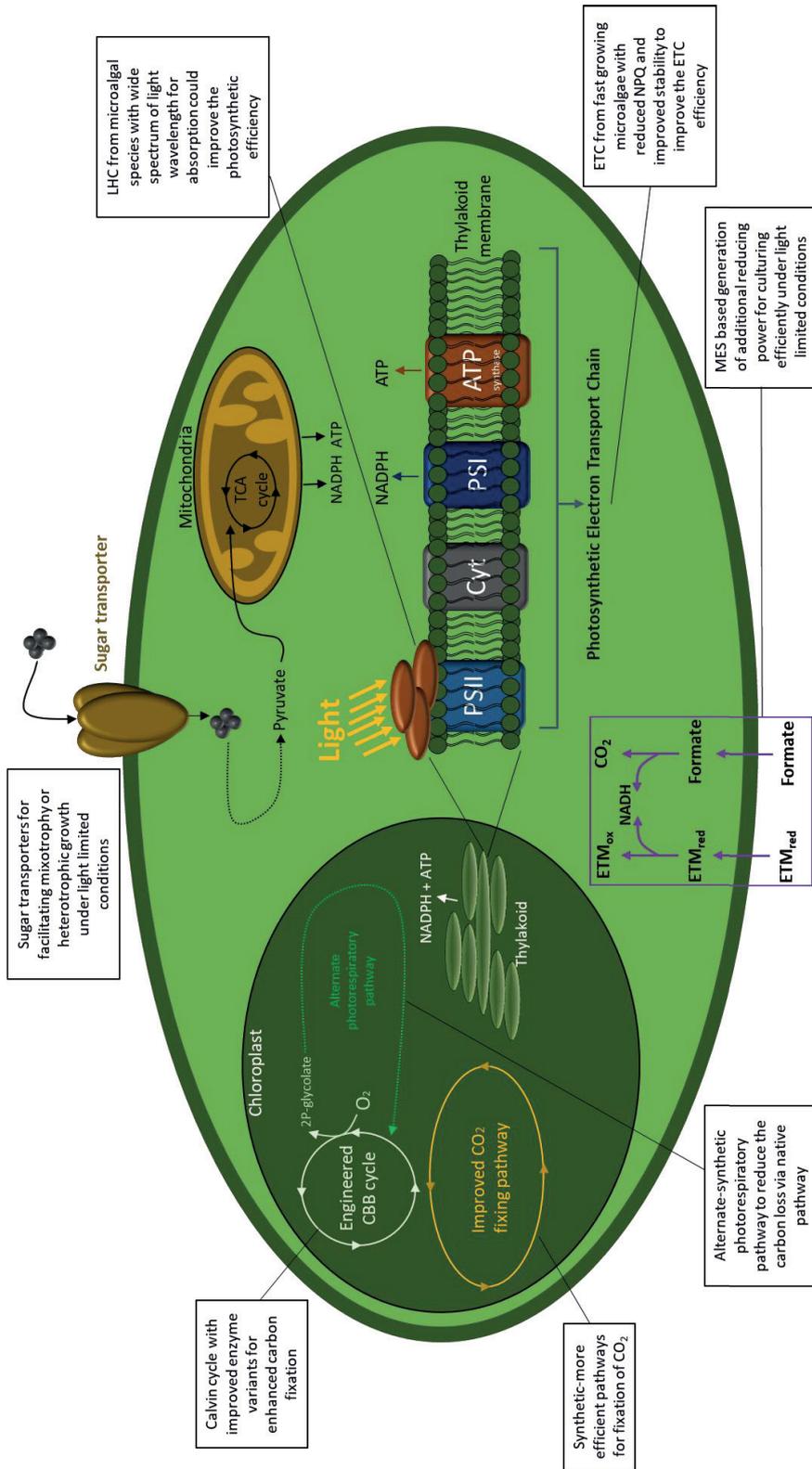


Figure 4. Various synthetic biology approaches that could be implemented in microalgae to enhance the productivity of biomass or metabolites.

Box 1. Development of the Genome-Editing Toolbox for Microalgae

The genome editing toolbox for microalgae consist of various nuclease enzymes such as Zinc Finger nucleases (ZFNs), transcriptional activator like effector nucleases (TALENs) and clustered regularly interspaced palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins. Before the emergence of CRISPR-Cas systems, ZFNs and TALENs were widely used in microalgae for generating mutants. The DNA binding domains of zinc finger proteins were combined with the endonuclease domain *FokI* to develop the synthetic nucleases (ZFNs) for generating the precise double stranded DNA breaks (DSDBs) [510]. The DNA binding domain of transcription activator like effector protein along with the *FokI* comprised the TALENs [511]. These 2 nucleases were successfully implemented in various microalgal organisms for generating mutants [512, 513]. However, the laborious modification and design of these proteins for each target site and high off-target events limited the reliability of using these systems for gene editing [234].

The CRISPR-Cas system emerged as a solution to the bottlenecks associated with the ZFNs and TALENs. The CRISPR system constitute the natural bacterial and archaeal immune system against bacteriophages and viruses [514]. Proteins of the CRISPR immune systems have the potential to cleave and degrade the invading alien DNA. The nuclease proteins are targeted to the alien DNA via CRISPR-derived RNA molecules (crRNA) [515]. By using the Cas nucleases and synthetic crRNA guides, these features of the immune system were exploited to develop a high-precision system for introducing double strand DNA breaks at specific sites, allowing either for disrupting genes or for specific recombination events. The CRISPR based genome editing was first demonstrated in human and mouse cells in 2013 [92, 95]. Since then, this tool has revolutionized the field of genome editing as it was successfully implemented across various organisms for precise generation of mutants. The CRISPR based gene editing in microalgae was first reported in *Chlamydomonas reinhardtii* in 2014 with a very low efficiency of yielding 1 targeted in mutant after transforming 10^9 cells [109]. However, optimization of the system in later years improved the editing efficiencies and expanded its application to other microalgal species. Moreover, catalytically inactive variants of the Cas proteins called the deadCas (dCas) were used for gene downregulation (silencing). The CRISPR-Cas based genome editing in microalgae is summarized in the Supplementary Table S1.

Box 2. Photosynthesis and the non-photochemical quenching (NPQ) response in microalgae

Conversion of solar energy into chemical energy sustains the majority of biological life on earth. In photosynthetic organisms the solar energy is used by photosystem II (PSII) to split 2 molecules of water into dioxygen, 4 electrons and 4 H⁺ molecules, while photosystem I (PSI) further uses solar energy to energize electrons towards the energy-level of the general cellular reducing equivalents NADPH [516, 517]. The excitation of chlorophyll a (Chl a) molecules in the light harvesting complex (LHC) of PSII by absorption of sunlight releases the energized electrons to enter the electron transport chain (ETC). These electrons pass through ETC by reducing various cofactors and PSI, to eventually generate NADPH and a H⁺ gradient over the thylakoid membranes. The latter gradient is used to drive the ATP synthase pump to generate ATP. This chain of reactions to produce the chemical energy from solar energy is facilitated by multiple enzymes in the ETC [518, 519]. The energization of the electrons by capturing the photosynthetically active radiation (PAR) from sunlight and rate of electron transfer in the ETC are important factors that determine the efficiency of this process. However, only 40% of the incident sunlight (100nm-1mm) belongs to the PAR (380-700nm) thereby limiting the spectrum of available sunlight that could be exploited for photosynthesis. Additionally, not all the excited electrons enter the ETC due to the lag in transfer of electrons between various complexes in the system. In this case, the excited electrons are de-excited by dissipating this excess energy as heat. This process is termed as non-photochemical quenching (NPQ) and results in loss of absorbed solar energy. NPQ is a protection mechanism that harmlessly de-excites the non-transferred electrons to prevent the generation of reactive oxygen species (ROS) [520-522]. Limiting the acquisition of photons under high light condition is a natural strategy to avoid the over-excitation of electrons leading to NPQ. To this end, photosynthetic organisms generally regulate the size of the light harvesting antennae based on the available light intensities. These antennae are usually composed of Chl a and Chl b molecules linked to pigment complexes that absorb and transfer the light energy towards the reaction centres of PSI and PSII. Under low light conditions the antenna size is increased to maximize light capture and vice versa. However, the response of this antenna size regulation is species specific and relatively slow resulting in harmful light capture under sudden variation from low to high light conditions such as found in dense cultures grown in photobioreactors, resulting in NPQ. This energy loss via NPQ is detrimental for achieving optimum photosynthetic efficiency (PE), and results in reduced biomass productivity.

Box 3. Calvin-Benson-Bassham cycle and other natural CO₂ fixing pathways

The CO₂ fixation in microalgae is primarily driven by the RuBisCO (D-ribulose-1,5-bisphosphate carboxylase/oxygenase) enzyme, as part of the Calvin-Benson-Bassham (CBB) CO₂ fixation cycle [523]. There are four forms of the RuBisCO proteins among which the Type I is the most abundantly found and harbors 8 large and 8 small subunits (L₈S₈). The type II RuBisCO found in some bacteria and dinoflagellates is structurally simpler as it comprises a dimer of large subunit (L₂) and thus could be a suitable candidate for heterologous expression. The type III RuBisCO generally found in archaea comprises a dimer of large subunit as seen in type II or as multiple dimers attached to form a multi-subunit complex ((L₂)₅) [524]. The type I RuBisCO enzymes are known to have relatively low catalytic rates ($k_{cat} < 10 \text{ s}^{-1}$) compared to other CO₂ fixing enzymes [525]. Furthermore, RuBisCO displays a 'wasteful' side-activity with oxygen in the presence of relatively high O₂ concentrations versus CO₂ (e.g. in ambient air with 20% O₂ and only 0.04% CO₂). This oxygenase reaction leads to the formation of 2-phosphoglycolate, which has to be detoxified or recycled into the Calvin cycle via photorespiration pathways. Both in plants and microalgae, 2-phosphoglycolate is recycled into the Calvin cycle via the C₂-cycle [526]. However, this C₂ cycle requires ATP input and releases some of the fixed CO₂ as well as fixed NH₃. Losses related to photorespiration have been reported to reduce photosynthetic efficiency in plants by up to 50% [527]. A study on microalgal cultivation in a photobioreactor, even in the presence of elevated CO₂, reported that the photorespirations could lead to 66% loss of fixed carbon, especially as in the bioreactor dissolved oxygen concentrations get very high due to photosynthesis [528].

In addition to the CBB cycle harbouring RuBisCO, there are six alternative natural CO₂ fixing pathways: (i) the reductive tricarboxylic acid cycle, (ii) the Wood-Ljungdahl pathway, (iii) the 3-hydroxypropionate bicycle, (iv) the 3-hydroxypropionate-4-hydroxybutyrate cycle, and (v) Dicarboxylate-4-hydroxybutyrate cycle and the recently discovered (vi) reductive glycine pathway [502, 503, 505-507, 529-532]. Among these six pathways, the CBB cycle is the least efficient CO₂ fixing route in terms of ATP consumption, as 7 molecules of ATP consumed to generate 1 molecule of pyruvate, on top of the aforementioned limitations of the carboxylating RuBisCO enzymes (Table 1) [523]. Other natural pathways consume much less ATP and the carboxylating enzymes in these pathways are much more efficient than RuBisCO.

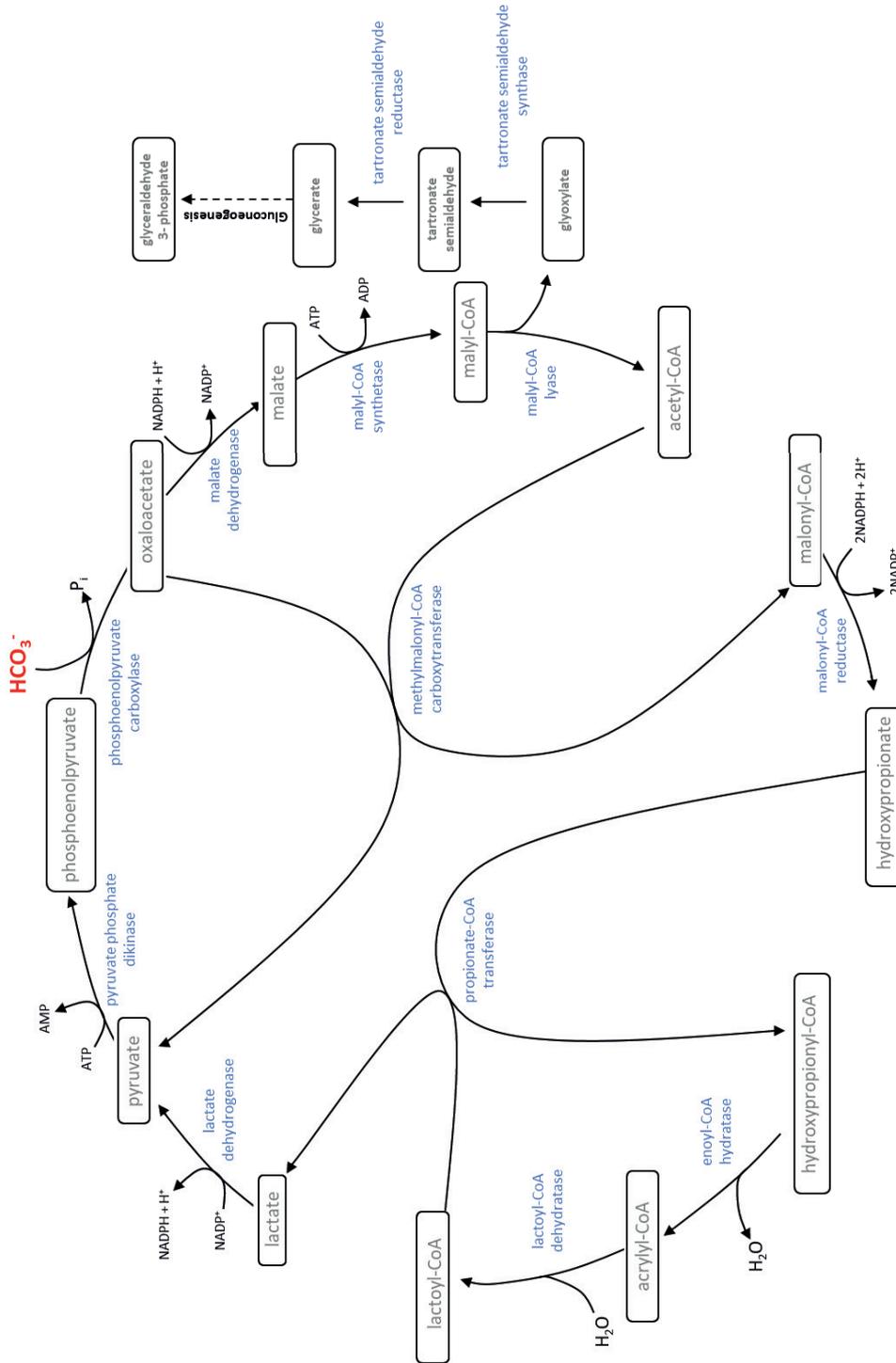
Box 4. Synthetic CO₂ fixing pathways

Based on the enormous repertoire of natural enzymes as well potential engineered enzymes, in addition to natural CO₂ fixing pathways, synthetic CO₂ fixation pathways can be envisioned. A systematic exploration of these possibilities based on around 5000 natural enzymes led to the design and analysis of a large collection of synthetic pathways [468]. These pathways were also analysed *in silico* in a genome-scale metabolic model of the microalgae *C. reinhardtii*, to estimate their potential for improving growth versus the CBB cycle. This analysis identified a promising family of pathways termed the malonyl-CoA-oxaloacetate-glyoxylate (MOG) pathways, which could operate under aerobic, ambient CO₂ conditions and is predicted to outperform the CBB cycle (Supplementary Figure S1). A MOG pathway employing the kinetically superior carboxylating enzyme phosphoenolpyruvate (PEP) carboxylase was proposed to have a 2-3-fold higher CO₂/O₂ specificity compared to RuBisCO in the CBB cycle. However, it is good to note that this pathway involves some enzymes for which only natural thermophilic enzyme variants are known [468]. So far, only the section of the pathway, converting malate to acetyl-CoA via the 'reverse glyoxylate shunt' has been realized by metabolic engineering in *E. coli* and the cyanobacterium *Synechococcus elongatus* [533, 534].

Another well-known hallmark of synthetic CO₂ fixation research is the crotonyl-CoA/ethylmalonyl-CoA/hydroxybutyryl-CoA (CETCH) cycle (Supplementary Figure S2), which includes some engineered enzymes. This synthetic CO₂ fixation route employs the enoyl-CoA carboxylases/reductases as carboxylase, which has a 2-4 fold higher catalytic rate than RuBisCO and is unreactive to oxygen as a substrate [508]. This aerotolerant pathway was demonstrated and optimized so far only *in vitro*. For the first version of the CETCH cycle, 12 candidate enzymes were sequentially added to an *in vitro* reaction mixture containing the cofactors and intermediates to prove that the cycle was functional. This cycle of reactions were optimized in multiple steps to improve the CO₂ fixation efficiency by 20 fold. This improved version of pathway termed as CETCH 5.4 was capable of fixing CO₂ at a rate 5 fold more efficient than the CBB cycle *in vitro*. CETCH 5.4 comprised of 13 core reactions and in total 17 proteins originating from 9 organisms [508]. As a proof of concept, Miller and co-workers recently demonstrated that the CETCH pathway potentially is compatible with the photosynthetic machinery and hence can be driven by light energy when combined with the thylakoid membranes isolated from spinach. However, the interactions between enzymes in CETCH 5.4 and the thylakoid membranes resulted in some negative interactions and reduced productivity resulting in an inactive system. These bottlenecks were addressed by addition and replacement of a few enzymes in the pathway giving rise to CETCH versions 6.0 and 7.0 [509]. Even though the CETCH cycle may have some attractive features for implementation *in vivo*, initial engineering, even in well-accessible bacterial hosts remains to be demonstrated.

Acknowledgements

The authors thank Arren Bar-Even (1980–2020) for inspiring discussions and groundbreaking ideas on improving carbon fixation metabolism. M.N. is funded by the project 'Microalgae as a Green source from Nutritional Ingredients for Food/Feed and Ingredients for Cosmetics by Cost-Effective New Technologies' (MAGNIFICENT), funded by the Bio-based Industries Joint Technology Initiative under the EU Horizon 2020 Research and Innovation Program (project 745754). N.J.C. is funded by a Dutch Organization of Science (NWO) Veni Grant (VI.Veni.192.156).



Supplementary Figure S1. A MOG pathway based on C4-glyoxylate cycle utilizing the most efficient carboxylating enzyme phosphoenolpyruvate carboxylase as the only carboxylating reaction. Glyoxylate, the product of the reaction could be converted to glyceroldehyde 3 phosphate.

Supplementary table 1. CRISPR based genome editings reported in microalgae

Organism	Editing technique	Targeted gene	Efficiency	Reference
<i>C. reinhardtii</i>	Plasmid based Cas9 expression	<i>FKB12</i>	10 ⁻⁹	[109]
<i>C. reinhardtii</i>	Cas9 Ribonucleoproteins	<i>MAA7, CoSRP43</i> and <i>ChlM</i>	Up to 40%	[237]
<i>C. reinhardtii</i>	Cas9 Ribonucleoproteins	<i>CpFTSY</i> and <i>ZEP</i>	0.56%	[138]
<i>C. reinhardtii</i>	dCas9-KRAB expression from plasmid	<i>CrPEPC1</i> gene downregulation	94%	[247]
<i>C. reinhardtii</i>	Cas9 Ribonucleoproteins	<i>ZEP</i>		[183]
<i>P. tricornutum</i>	Plasmid based Cas9 expression	<i>CpSRP54</i>	Up to 63%	[535]
<i>T. pseudonana</i>	Plasmid based Cas9 expression	<i>UreA</i>	61.5%	[151]
<i>N. oceanica</i>	Plasmid based Cas9 expression	<i>NarB</i>	1%	[146]
<i>N. oceanica</i>	Plasmid based Cas9 expression	<i>NarB</i>	Up to 90%	[79]
<i>N. oceanica</i>	Ribonucleoprotein : Cas9 and Cas12a	<i>NarB</i> ; HDR based deleteion	Up to 92%	[227]
<i>N. gaditana</i>	Plasmid based Cas9 expression	18 different transcriptional regulator	Up to 78%	[80]
<i>Tetraselmis spp</i>	Ribonucleoprotein	<i>AGP</i>	0.1%	[238]
<i>C. reinhardtii</i>	Cas9 Ribonucleoproteins	<i>ELT</i>	13%	[536]
<i>C. vulgaris</i>	Plasmid based Cas9 expression	<i>Fad3</i>	Approx. 30%	[537]
<i>P.celeris</i>	Ribonucleoprotein	<i>NarB</i> and <i>CRITSO</i>	6% and 50%	[500]

Supplementary Table 2. Light harvesting antenna truncation in microalgae resulting in improved productivity

Organism	Gene targeted	PSII size variation	Remarks	Reference
<i>C. reinhardtii</i>	<i>tla1</i>	50% reduction	Improvement in the productivity at high light intensities	[538]
<i>C. reinhardtii</i>	<i>tla2</i>	35% reduction	Productivity improved at high density and high light intensity	[539]
<i>C. reinhardtii</i>	<i>tla3</i>	60% reduction	Photosynthetic productivity improved by 100% at high light intensities	[540]
<i>C. reinhardtii</i>	<i>tla4</i>	27% reduction	15% higher growth compared to WT	[541]
<i>C. reinhardtii</i>	<i>Stm6GlcT7</i>	Up to 17% reduction	53% improvement in the exponential growth rate	[542]
<i>Chlamydomonas perigranulata</i>	<i>LHC-1</i>	15% reduction	Improved photosynthetic productivity at high cell density	[543]
<i>Chlorella vulgaris</i>	<i>CpSRP43</i>	Reduced (value not specified)	Growth rate improved by 44.5% at 400 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$	[410, 544]
<i>Chlorella sorokiniana</i>	<i>TAM</i>	45% reduction	Biomass productivity improved by 32% at light intensities of 450 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$	[545]

CHAPTER 8



General Discussion





Requirement and potential of genetic engineering

Genetic engineering can be described as a process that alters the native genome architecture by dedicated insertion, deletion, or substitution of nucleotides, genes, or large genome fragments. Recent genetic engineering strategies even include techniques that do not interfere with the native genome sequence, but rather introduce epigenetic modifications that are stable for many generations resulting in knockdown phenotypes [546]. In search of better traits, humans have been performing the traditional breeding techniques for thousands of years in plants and animals by crossing different varieties of the same species with valuable characteristics. This technique could be considered a traditional form of genetic engineering as the offspring is designed to harbour selected traits from parent strains. The generation of high-yielding crop varieties by breeding paved the way towards the “Green Revolution” in the 1960s, saving billions of people from starvation. Continuous breeding in plants and domesticated animals for many years has resulted in almost all the food we eat to be genetically modified. However, the process of breeding and selecting the strains with desirable traits is laborious and time-consuming. The direct delivery of foreign genetic material into a host strain and to propagate it for various generations could be marked as the first steps towards advanced genetic engineering. This was achieved in the early 1970s where recombinant DNA technology was used to generate circular plasmids based on phage DNA fragments [547], and subsequent application of these tools to introduce antibiotic resistance genes in *Escherichia coli* bacterium to develop the first transgenic organism [547-549]. Afterwards, this technology was applied in mouse blastocysts to develop the first transgenic animal [550]. Another breakthrough in the field of genetic engineering was the production of human insulin in *E. coli* in the early 1980s. This technology replaced the administration of insulin in humans that was extracted from the pancreas of bovines and porcine [551]. Soon after, the genetic engineering technology was implemented in plants and adopted in various crops to provide traits that simplify farming and improve the quality of harvest [552]. Recent studies show that the adoption of this technology has remarkably increased the profit for farmers by reducing the use of pesticides and improving the crop yield [552]. These milestones achieved since the development of this technology half a century ago, indicate its huge potential in revolutionizing the fields of fundamental biological science as well as biotechnology.

In microalgae, genetic engineering was first implemented in the 1980s where a chloroplast transformation technique was reported in *Chlamydomonas reinhardtii* using high-velocity tungsten microprojectile bombardment [553]. This development can be considered the first milestone of genetic engineering in microalgae. During the early 1990s, a nuclear transformation protocol based on agitation of glass beads was reported for cell wall deficient *C. reinhardtii* [554]. Soon after, a nuclear transformation protocol for *C. reinhardtii* with intact cell wall was also developed. This technique was based on electroporation or agitation of cells with plasmid DNA and silicon carbide whiskers [554, 555]. The development of *Agrobacterium*-mediated gene transfer further improved the transformation efficiencies in *Chlamydomonas* by five-fold [556]. In 2011, electroporation based transformation of linear DNA was reported for *Nannochloropsis* to knockout target genes by homologous recombination [78]. The development of transformation techniques

and antibiotic resistance cassettes for swift selection of transformants paved the way towards implementing novel genetic engineering approaches in microalgae. In 2013, a synthetic Zinc Finger Nuclease (ZFN) was used in *Chlamydomonas* to generate targeted knockouts [82]. A similar technology using a synthetic Transcription Activator Like Effector Nuclease (TALEN) was used for generating knockouts in *Phaeodactylum tricornutum* in 2014 [48]. During the same year, the revolutionary CRISPR-Cas system was reported to be transiently expressed in *C. reinhardtii* [109]. And, in 2016, CRISPR-Cas mediated knockouts were reported for the first time in *Chlamydomonas* and *Nannochloropsis* [138, 146, 237]. These advancements could be considered as the major milestones in the field of microalgal genetic engineering.

Cas ribonucleoproteins for engineering *Nannochloropsis oceanica* IMET1:

The first peer-reviewed studies on implementing CRISPR-Cas systems in microalgae for generating targeted knockouts had feeble efficiencies in generating mutants [109, 146, 236]. In some studies, this outcome was attributed to the cytotoxic effect of the Cas9 enzyme in the host organism as the plasmid-based expression of Cas9 enzyme with lack of stringent inducible promoters resulted in the constitutive expression of this enzyme and possible off-target activity [109, 236]. Transformation of Cas9 ribonucleoproteins (RNPs) could address this concern as it can provide a stable and transient Cas9 presence in the host cell for inducing the double-stranded breaks (DSB). Moreover, Cas9-RNP based genome editing has an established protocol as it was widely used for editing plants and mammalian cell lines [557, 558]. This approach substantially improved the efficiency of generating mutants in *Chlamydomonas* [138, 183, 237]. We adapted this approach to develop an RNP-based genome editing toolbox for *N. oceanica* in combination with homology-directed repair (HDR). The HDR allowed precise generation of mutants with high efficiencies in contrast to induction of uncontrolled indels as demonstrated in *Chlamydomonas*. Moreover, selecting knockouts based on indels were laborious when targeted genes did not produce selectable phenotypes, whereas the HDR based approach simplified the selection of mutants based on antibiotic resistance as reported in Chapter 3.

Development of this genome editing tool was followed by the investigation of strategies to improve the lipid accumulation in *N. oceanica* IMET1. Overexpression of rate-limiting steps in the *de novo* production pathways is a straightforward approach to enhance the accumulation of final metabolite. In this regard, various studies have reported overexpression of acetyl-CoA carboxylase that triggers the fatty acid synthesis in microalgae without significant improvements in the lipid content [558-561]. Thus, this step was ruled out to be a rate-limiting step in fatty acid synthesis (FAS). In *Nannochloropsis*, the second step in the FAS pathway was upregulated by overexpression of malonyl-CoA acyl carrier protein trans-acylase (MCAT) to increase the lipid content by 36% [562]. Furthermore, overexpression of diacylglycerol acyltransferase 2 (DGAT2) enzyme that catalyses the final step in triglyceride (TAG) synthesis in the Kennedy pathway increased the TAG content by 129% in *Nannochloropsis* [563]. Thus, independent or co-overexpression of

specific enzymes in the FAS pathway and lipid assembly pathway could improve the lipid accumulation in *Nannochloropsis*.

Apart from the overexpression of the lipid production pathway, blocking the lipid degradation pathway such as β -oxidation is reported to enhance the lipid content [564]. In microalgae, β -oxidation occurs either in peroxisomes and/or mitochondria depending on the species and is the major pathway that breaks down fatty acids to release acetyl-CoA that fuels the TCA cycle to generate ATP [565]. The process begins with the activation of fatty acids into fatty acyl CoA in the cytoplasm and their transport into peroxisome or mitochondria by the transporters ATP binding cassette protein (ABCD1) or carnitine acylcarnitine translocase (CACT), respectively [566]. The fatty acyl CoA in the mitochondria or peroxisome undergoes multiple cycles of four reactions (Figure 1), that completely breaks down the fatty acyl CoA into acetyl-CoA. The acyl-CoA is initially oxidized to form enoyl-CoA, followed by hydration of this molecule to form 3-hydroxy-acyl-CoA, which is then dehydrogenated to form 3-keto-acyl-CoA. Finally, 3-keto-acyl-CoA undergoes thiolytic cleavage to release an acetyl-CoA molecule and acyl-CoA with chain length reduced by two carbons. The shorter acyl-CoA molecule enters another round of fatty acid oxidation and this cycle continues until the entire acyl-CoA is converted into acetyl-CoA (Figure 1) [564, 565].

We opted to execute the second approach of blocking the β -oxidation as the RNP based genome editing in *Nannochloropsis* was efficient in generating targeted knockout mutants. Based on literature, we attempted to knockout multiple genes associated with peroxisomal and mitochondrial β -oxidation in *Nannochloropsis*. The mitochondrial transporter protein CACT, peroxisomal biogenesis factor 10 protein (PEX10), and the acyl CoA oxidase were the selected target genes. PEX10 facilitates the formation of matrix proteins in peroxisome and disruption of the gene coding for this protein in yeast *Yarrowia lipolytica* increased the EPA accumulation. Acyl CoA oxidase catalyses the first step of β -oxidation where the fatty acyl-CoA is oxidized to enoyl-CoA. Simultaneous knockout of acyl CoA oxidase and overexpression of DGAT in *Saccharomyces cerevisiae* resulted in 286% increased lipid accumulation, rendering this target as a potential knockout candidate in *Nannochloropsis* to improve lipid accumulation [567]. Nevertheless, even after multiple rounds of transformations, we were unable to isolate knockout mutants of CACT, PEX10 or acyl CoA oxidase. Interestingly, recent studies on the transcriptomics of *Nannochloropsis* sp. showed that the genes involved in the β -oxidation were stably expressed regardless of stress induction [397]. Thus, the acetyl-CoA generation via this process could be inevitable in the metabolism and any interference in this pathway could be lethal for *Nannochloropsis*.

Plasmid based genome editing tools for *Nannochloropsis oceanica* IMET1:

The Cas RNP-mediated generation of mutants appeared to be a straightforward approach for easy adaptation in multiple organisms. However, from a laboratory perspective, the application of this system had a few drawbacks such as the requirement of highly purified Cas proteins, generation of guide RNAs (gRNA) for different target genes, desalting and

assembly of Cas RNPs from their storage buffer just before transformation, and varying transformation efficiency due to the interference of Cas storage buffers on electroporation. On the other hand, the transformations using plasmids in *Nannochloropsis* are based on a user-friendly protocol and are relatively well reproducible. Even though plasmid-based genome engineering techniques were established for *Nannochloropsis*, the integration of the plasmid in the genome always resulted in transgenic mutants [80, 146, 182]. The development of an episomal plasmid system that can be cured from transformed *N. oceanica*, opened up possibilities to use a plasmid-based approach to generate non-transgenic mutants [79].

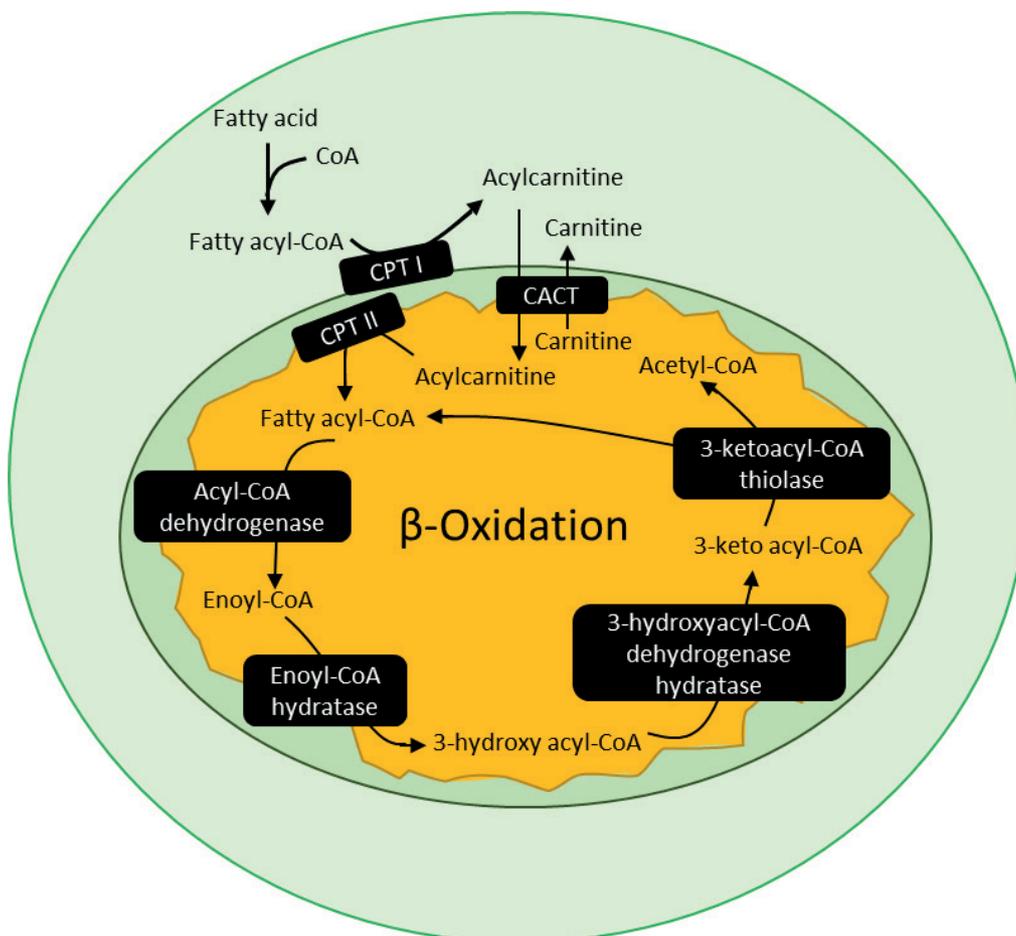


Figure 1. Mitochondrial beta-oxidation in microalgae. The acyl-CoA dehydrogenase or acyl-CoA oxidase and CACT were the targets for generating knockout in *N. oceanica* IMET1. CPT is the carnitine palmitoyl transferase adds a carnitine group to the fatty acyl-CoA to form acylcarnitine, which is transferred by the CACT transporter across the mitochondrial membranes.

Cas9 enzyme was the predominant CRISPR protein used for developing genetic tools in microalgae due to its early establishment compared to other Cas nuclease variants [233]. As an alternative enzyme, Cas12a was initially demonstrated in *Chlamydomonas* where LbCas12a RNPs induced indels and HDR to generate targeted mutants [140]. A plasmid-based Cas12a system was also demonstrated in the cyanobacterium *Anabaena* PCC 7120 [568]. Nevertheless, these studies did not exploit the full potential of the Cas12a enzyme to establish this Cas variant as a preferred genome editing tool in microalgae or cyanobacteria. The crRNA biogenesis by Cas12a is the major advantage of using this enzyme over Cas9. In most of the type II CRISPR systems including Cas9, the crRNA maturation, after base pairing with an additional RNA (tracrRNA), is performed by the host protein RNase III. Contrary to this, the unique endoribonuclease and endonuclease activity of Cas12a process the pre-crRNA to produce mature crRNAs in the host organism [124, 569]. This feature allows the dispensability of complex ribozyme systems in the plasmid-based approach to producing the crRNA and allows for swift multiplexed genome editing targeted host organisms [103]. Moreover, an *in vivo* comparative study of Cas9 and Cas12a showed that Cas12a results in lower off-targets in the host rendering this enzyme as a preferred variant over Cas9 [570, 571].

Our efforts to develop Cas12a as a principal genome editing tool for *Nannochloropsis* is described in Chapters 3 and 4. We demonstrated plasmid and RNP mediated strategies using Cas12a as an efficient tool for generating targeted mutants in *Nannochloropsis*. The possibility to remove ribozymes for producing the crRNA simplifies the cloning procedure and facilitates multiplexed genome editing with high efficiencies. Moreover, we demonstrated CRISPR interference (CRISPRi) using dead Cas12a and Cas9 variants in *N. oceanica* for the first time. We attempted prime editing using a nickase Cas9 in *N. oceanica* but failed to isolate any prime edited mutants. The expression of prime editing protein complex (nickase Cas9 and reverse transcriptase complex) was validated in our host, but the design of the prime editing guide RNA (pegRNA) in forming the RNP complex in the host and guiding the Cas protein to the target probably needs to be optimized.

Even though the tool development resulted in a comprehensive genome editing toolbox for *N. oceanica*, we encountered multiple obstacles during the process. The expression of Cas proteins in *Nannochloropsis* was one of the major bottlenecks. Initially, we codon harmonized the *fnCas12a* gene sequence for expressing in *N. oceanica* and failed, while a human codon optimized Cas9 sequence was successfully expressed for generating targeted indels. Similarly, the initial prime editor system harboured a Cas9 sequence that was approximately 81% identical to the human optimized Cas9 sequence and failed to express in *N. oceanica*. These observations indicate stringent control of protein expression based on codon bias in *N. oceanica*. Initial studies mentioned that Cas9 cytotoxicity is responsible for reduced efficiency in generating mutants in *Chlamydomonas*, while these studies did not validate the Cas9 expression in the host [109]. Thus, the expression of Cas9 could have been the reason behind reduced efficiency rather than the cytotoxic effects of Cas9, as recent studies efficiently generate mutants with the same enzyme [236]. In-depth research is required to understand the regulation of gene expression in *Nannochloropsis* based on codon bias to overcome this bottleneck in the future.

The copy number of transformed episomal plasmid in *Nannochloropsis* and its effect on the heterologous protein expression was another bottleneck that we encountered during the development of CRISPRi system. Our first approach to demonstrate the CRISPRi was based on a single plasmid system that expressed both the Td tomato and dCas protein. However, the characterization of transformants by quantifying the Td tomato fluorescence from different transformations harbouring plasmids that targeted the Cas proteins to Td tomato gene sequence and non-targeting negative control provided inconclusive results. This could have been due to variations in the copy number of plasmids in the transformants. The transformants harbouring non-targeting plasmids might not be comparable to transformants with plasmids targeting the Td tomato if the plasmid copy number is lower in the former case and vice-versa. However, our second approach with Td tomato being stably expressed from the genomic DNA resolved this issue and resulted in statistically significant downregulation of target gene using both dCas9 and dCas12a. Additionally, we also tested a type V-U1 Cas12 nuclease presently termed as MmuCas12m, that was isolated from *Mycolicibacterium mucogenicum* for gene downregulation in *Nannochloropsis* (Wu, Mohanraju et al. unpublished). This variant was interesting due to its smaller size compared to other established Cas variants and its presently hypothesized ability to cleave the RNA. Even though the expression of this Cas variant was validated in *Nannochloropsis*, an efficient CRISPRi was not observed using this variant, not even while tagging the protein with the KRAB domain (data not shown).

Further elaboration of the genetic toolbox for *Nannochloropsis*:

With the afore-mentioned developments, the genome editing toolbox of *Nannochloropsis* could be the most advanced compared to other microalgal strains. The application of Cas12a to perform multiplexed genome editing and RNPs to generate scarless mutants based on HDR are not reported in any other microalgae. Nevertheless, further optimizations can improve the reliability of these genome editing tools. The constitutive expression of Cas proteins in the plasmid-based approach could result in cytotoxic or off-target effects [572]. Thus, driving the Cas proteins under the control of an inducible promoter may improve the system. In *Nannochloropsis*, only the native nitrate reductase promoter is characterized as an inducible promoter [473]. This promoter has low activity while using ammonia as the nitrogen source in the media and can be induced by switching the nitrogen source to nitrate [473]. Using this promoter to drive the Cas proteins in *Nannochloropsis* could reduce possible off-target effects.

Presently, the plasmid-based CRISPR tools for *Nannochloropsis* generates indels at the target site [79, 80]. Optimising these tools to efficiently produce HDR-based mutants could be a valuable addition to the toolbox as this system facilitates precise deletions, insertions, and substitutions at the target site. As HDR is reported to be efficient in *Nannochloropsis*, introducing the homologous recombination (HR) flanks with required mutations (Figure 2) in the same episomal plasmid expressing the Cas protein and crRNA could facilitate this approach. However, experiments will be required in defining the optimum length of the HR flanks and the position of these flanks on the episomal plasmid to achieve maximum

efficiency of HDR. This approach could be an alternative to the prime editing as both systems yield precise alterations of the targeted loci.

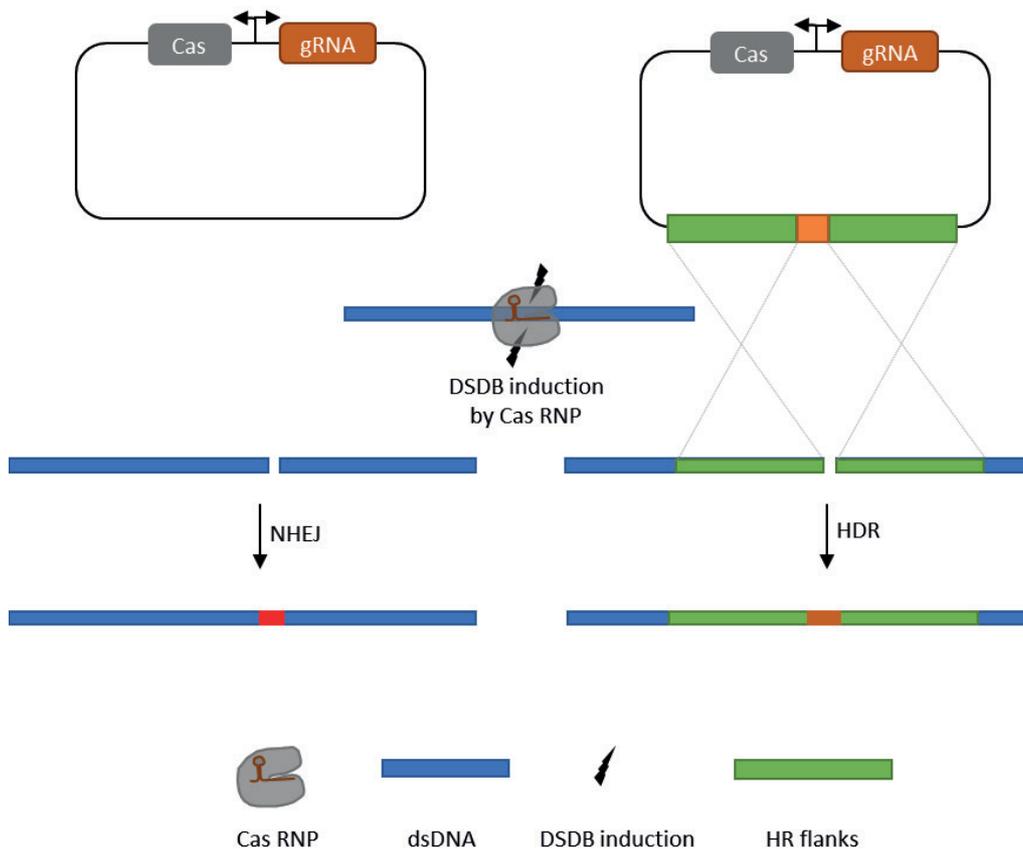


Figure 2. Episomal plasmid based approach for generation of indels and precise generation of mutants based on homology directed repair (HDR). In absence of the homologous recombination flanks, double stranded DNA break (DSDB) induction is repaired by non-homologous end joining resulting in mutants with indels at the target site (Left). In presence of homologous recombination flanks on the episomal plasmid, homology directed repair can facilitate precise generation of mutants.

The episomal plasmid-based CRISPRi tool requires stable deadCas (dCas) protein expression to achieve consistent downregulation of the target gene. In this case, integrating the dCas and the guide RNA into the genomic DNA would be better than using an episomal plasmid system as scaling up of this phenotype for large scale production of metabolites will be expensive due to the requirement of antibiotics to maintain the episomal plasmid in the host. Also, these mutants will be transgenic. Recently developed CRISPRoff technology could address this issue by generating non-transgenic strains with stable downregulation of the target gene [546]. The CRISPRoff technology alters the epigenetic memory of the host genome at a precise location and maintains this alteration for more than 450 generations. This is achieved by fusing the dCas protein with methyltransferase enzymes

Dnmt3A and Dnmt3L, in addition to the zinc finger protein 10 (ZNF-10) KRAB domain. The fusion protein was targeted to the gene of interest (GOI) using the guide RNA resulting in the methylation of the targeted region and subsequent downregulation in the expression of the GOI [546]. Even though this epigenetic modification was stable for hundreds of generations, knockout of the DNA methylation maintenance enzyme DNMT1 in the host resulted in reactivation of silenced genes, indicating the flexibility of the approach in reversing the modification [546]. The CRISPRoff technology has only been demonstrated in mammalian cell lines but might have huge potential in various organisms including microalgae.

Another interesting application of Cas9 demonstrated in *Nannochloropsis* was the deletion of approximately a 100kb long fragment from the telomeric end of the chromosome [228]. The successful deletion of long fragments in *Nannochloropsis* indicates exciting possibilities in developing a minimal genome of *Nannochloropsis* by a top-down approach as demonstrated in bacterium *Mycoplasma mycoides* [573]. In eukaryotes, unknown and non-essential regions occupy up to 70% of the genome [574]. The presence of such mysterious regions containing transposons or insertion sequence elements could continuously mutate the genome, thereby increasing the genome complexity and instability [575]. Developing a minimal genome was an ambitious and fascinating concept to address this issue where the host strain only harbours the essential genes required to multiply under optimal conditions [576]. These minimal cells are hypothesized to be optimal platform strains for producing heterologous proteins and metabolites as cellular energy used for transcription and translation of a large portion of the genome can be redirected towards heterologously expressed genes or pathways [577]. Moreover, developing such strains can provide in-depth insights into the metabolism, structure, and function of a cell. Targeted deletion of large fragments of the genome would be required to develop a minimal genome in *Nannochloropsis*. Even though Cas9 was used to this end in *Nannochloropsis*, developing a Cascade-Cas3 system for this species could accelerate the development of a minimal *Nannochloropsis* genome. Cas3 belong to the most prevalent Type 1 (Class 1) CRISPR-Cas system [515, 578] and in contrast to Cas9 (Class 2, type 2) and Cas12a (Class 2, type 5) proteins, the type 1 system works as a multi-subunit complex called Cascade (CRISPR associated complex for antiviral defence). The Cascade identifies and forms the heteroduplex between the crRNA and the matching dsDNA with an optimal protospacer adjacent motif (PAM) [579]. The Cas3 protein harbouring the helicase and nuclease activity is recruited towards the dsDNA bound Cascade complex to initially nick the non-targeted DNA strand. Subsequently, the helicase activity of Cas3 unwinds the dsDNA and degrades this ssDNA in a 3'-5' direction [580]. Approximately 837 kilobases were deleted from the genome of bacteria *Pseudomonas aeruginosa* and more than 200 kilobases in human cells using the Cascade-Cas3 system. The Cascade-Cas3 system is also reported to be more efficient than Cas9 and Cas12a in generating large deletions [581]. Thus, developing a Cascade-Cas3 CRISPR system for *Nannochloropsis* could accelerate the research towards the development of a minimal genome.

Random mutagenesis; an alternative to targeted engineering in *Nannochloropsis*:

The development of targeted genome editing tools can be futile unless the genetic composition or metabolic pathway in the host species is well defined. Thus, the application of these genetic tools in non-model species will be limited to the knowledge on their metabolism and omics data. In this regard, random mutagenesis can be an interesting, un-biased alternative to targeted genetic engineering to produce mutants with desirable characteristics. This approach exposes the microbes to various physical (radiations: UV, X-ray, or γ -ray) or chemical mutagens (intercalating agents such as ethidium bromide, base analogues that trigger base transitions such as 5- bromouracil, non-alkylating and alkylating agents) that randomly alters the genomic DNA of the host organism [582]. Additionally, insertional mutagenesis based on the transformation of linear antibiotic resistance cassettes that randomly integrate into the genomic DNA is also used to drive random mutagenesis [582]. Subsequently, the cells exposed to mutagens are screened for the desired phenotype. We have demonstrated insertional mutagenesis-based isolation of high lipid producing mutants in Chapter 5. This technique was also demonstrated in *Chlamydomonas* and *Nannochloropsis salina* to characterize various genes associated with photosynthesis and lipid production, respectively [583] [289]. Random mutagenesis based on physical and chemical mutagens is widely used in microalgae to improve the lipid yields [582]. Moreover, the genotypic characterization of these high lipid producing mutants could provide novel insights into the lipid metabolism in microalgae. In addition to improving the lipid yields, robust microalgal strains with tolerance towards high light intensity and temperature were developed using random mutagenesis. Chemically mutated *Chlorella* strains had approximately 43% higher cell dry weight compared to WT when cultivated outdoors at higher temperatures [584]. Similarly, chemically mutated *Chlorella vulgaris* was able to grow at high light intensities with improved biomass productivity due to its reduced photosynthetic pigment content and truncated antenna [410]. On the other hand, high light tolerant *Desmodesmus* sp. overexpressed most of the genes associated with light-harvesting. The lipid content in this mutant was higher compared to the WT, which, based on transcriptomic data, was attributed to the reallocation of membrane lipids to TAG [585]. These results indicate that, in obtaining a specific phenotype such as high light intensity tolerance, the strategies can be different based on the host species. Thus, adapting the same targeted engineering strategies across various species might not produce expected phenotypes.

Similar to random mutagenesis, adaptive laboratory evolution (ALE) is another potential technique that could yield mutants with desired characteristics. In this technique, the strain of choice is cultured under specific stress conditions such as high temperature or high light intensity for long durations that results in the outgrowth of cells that are adapted to the stress from the overall population [582]. This approach also results in the accumulation of multiple mutations in the genome in contrast to targeted engineering and could provide novel insights into the metabolism and function of the strain. Various microalgal strains have been subjected to ALE to improve the CDW and lipid yield [582]. Nevertheless, this technology is yet to be implemented in *Nannochloropsis*.

The techniques of random mutagenesis and ALE could be the best approach to characterize novel strains and establish a library of mutants with different phenotypes. Characterization of these strains could be used to establish or update the knowledge on the metabolism and function of the studied organism. This approach can also accelerate the development of a genome-wide metabolic model, that in turn could provide a basis for a targeted engineering approach. Moreover, the mutants developed by these approaches are generally considered safe with no legal or regulatory restrictions to cultivate them outdoor, in contrast to the GMO regulations applied for mutants developed by targeted engineering [586].

Countries such as the USA, Canada and Argentina do not regulate organisms, such as plants, animals or microalgae, as genetically modified organisms (GMO) if they are free of foreign DNA [587]. This rule has been the basis of developing DNA free strategies that use CRISPR RNPs for developing mutants in plants, mammalian cell lines and microalgae [227, 588, 589]. However, the regulations are still stringent in Europe where all organisms that have undergone the process of mutagenesis and harbour a genome that cannot naturally occur are regulated as GMOs. However, an exemption to this is the classical, random mutagenesis techniques that has been conventionally used to develop mutant crop varieties [590]. This declaration from the European Court of Justice has been a burden on biotech industries using the CRISPR systems to market their products as it requires additional permissions and labelling adding up to the financial cost [591]. Moreover, this restriction will be a limitation in obtaining the fruit of this efficient technology that can produce highly desirable and profitable mutants. Thus, global discussions and efforts are ongoing, trying to provide transparent and unbiased knowledge on this technology to the public. Creating awareness among the politicians and lawmakers about the developments that help mitigate the ethical concerns such as off-target effects in GMOs hopefully will reframe the present guidelines [592, 593].

An outlook on unique metabolism of *Nannochloropsis*.

Nannochloropsis has been a genus of choice for scientific studies owing to its natural capability to accumulate commercially relevant eicosapentaenoic acid (EPA) and triglycerides (TAG) in high amounts compared to other model microalgal strains. The majority of the metabolic engineering studies on *Nannochloropsis* also focussed on strategies to optimize these natural pathways (Figure 3) to further improve their productivities. The conversion of acetyl-CoA into malonyl-CoA by the enzyme acetyl-CoA carboxylase (ACCase) can be considered as the first step in the *de novo* fatty acid synthesis (FAS) in microalgae. Subsequently, malonyl-CoA is converted to malonyl-ACP by the enzyme malonyl-CoA ACP transacylase (MAT). Next, malonyl-ACP undergoes multiple rounds of condensation, two steps of reductions and dehydration to form the saturated fatty acids (SFA) (Figure 3). C16:0 is the most abundant SFA observed in *Nannochloropsis* followed by C14:0 and C18:0. The SFAs can be converted to mono-unsaturated fatty acids (MUFAs), C14:1, C16:1 or C18:1 by the enzyme FAD9 desaturase. These fatty acids are activated by the addition of a CoA molecule using the acyl-CoA synthetase. Subsequently, these fatty acyl-CoAs are transferred to the glycerol backbone by releasing the attached

CoA molecule to synthesise various lipid molecules. These fatty acids attached to the glycerol backbone will be further desaturated and elongated to produce long-chain polyunsaturated fatty (LC-PUFAs) acids such as C20:4 and the commercially relevant C20:5 or EPA. The LC-PUFAs released from the glycerol backbone forms the membrane of organelles or they can also be recycled into lipids via the Kennedy pathway.

Kennedy pathway triggers the glycerolipid synthesis, where the enzyme Glycerol 3-phosphate (GPAT) transfers the fatty acyl-CoAs into the sn1 position of the glycerol backbone to produce lysophosphatidic acid (LPA). LPA is further acylated using fatty acyl-CoAs at the sn2 position by the enzyme LPA acyltransferase (LPAAT) to form phosphatidic acid (PA), which is converted to diacylglycerol (DAG) by the enzyme phosphatidic acid phosphatase (PAP). DAG serve as the precursor for the synthesis of various polar lipids such as monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG) and Sulfoquinovosyl diacylglycerol (SQDG). The final step in the Kennedy pathway is catalysed by the enzyme DAG acyltransferase (DGAT) to produce triglycerides (TAG).

Metabolic engineering studies mainly followed the overexpression of enzymes in the FAS pathway and LC-PUFA synthesis pathway to achieve modest improvements in the lipid and EPA content in *Nannochloropsis* [562, 594]. *N. oceanica* CCMP1779 is reported to have 12 variants of the enzyme DGAT [595]. The overexpression of the DGAT enzyme (*NoDGT5*) resulted in 1.75 fold increase in the lipid content [595], and the overexpression of *DGAT1A* resulted in a 2.4 fold increased TAG content in *N. oceanica* [90]. These strategies that improve the lipid content were associated with reduced growth rates [595]. This has been a bottleneck in improving the overall productivity of lipids in *Nannochloropsis*. In this regard, engineering the transcription factors has provided promising results. Overexpression of a basic leucine zipper (bZIP) transcription factor (TF) involved in stress responses and lipid metabolism in *N. salina* increased the expression levels of enzymes in the FAS and Kennedy pathway. This approach increased the TAG content by 88% under stress with similar improvement in growth rates [596]. As nitrogen starvation induces lipid production in *Nannochloropsis*, the TFs with reduced expression during this growth condition was hypothesized to be negative regulators of lipid production [80]. Knocking out Zn (II)2Cys6 TF, *ZnCys*, that was downregulated under nitrogen stress resulted in a 175% increase in the total fatty acid content with a reduced growth rate in *N. gaditana* [80]. However, knocking down the expression of this TF increased the total fatty acid content by 100% without inhibiting the growth rate [80]. The characterization of high lipid producing mutants developed by random insertional mutagenesis in *N. oceanica* also indicated the disruption of the gene coding for *NoZnCys* [261]. Another high lipid producing mutant of *N. oceanica* with approximately 40% improved lipid content had disruption of APETALA2-like TF. This mutant had improved the activity of enzymes in the glycolysis and FAS pathways, as well as the and Calvin-Benson-Bassham cycle [597]. These studies indicate that the overexpression or knockout of a single gene or few genes might not be sufficient to improve the lipid productivity in *Nannochloropsis*. On the other hand, TF engineering can indirectly rewire the entire metabolism leading to favourable phenotypes. Fundamental studies into TFs to elucidate their role in metabolism can be crucial to exploiting this technique for metabolic engineering [596]. Without appropriate metabolic insights, random approach may be a better option.

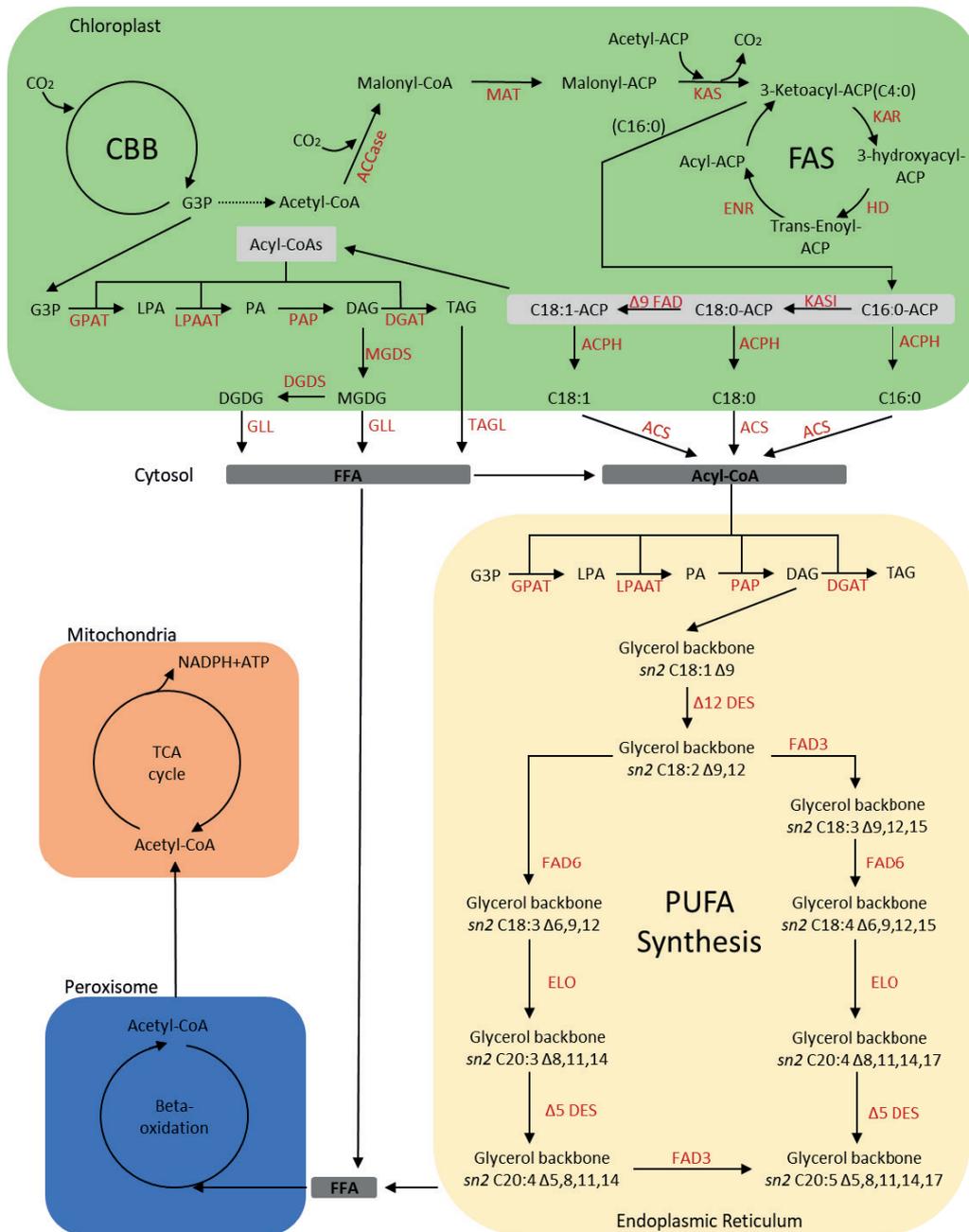


Figure 3. Schematic lipid biosynthesis pathway in microalgae. Abbreviations; acetyl-CoA carboxylase (ACCase), ACP hydrolase (ACPH), acyl-CoA synthase (ACS), Calvin Benson Bassham Cycle (CBB), DGDG synthase (DGDS), enoyl-ACP reductase (ENR), elongase (ELO) fatty acid desaturase (FAD), free fatty acids (FFA), glycerolipid lipase (GLL), glyceraldehyde-3 phosphate (G3P), beta-ketoacyl-ACP synthase I (KASI), 3-hydroxyacyl-ACP dehydratase (HD), 3-ketoacyl-ACP synthase (KAS), 3-ketoacyl reductase (KAR), malonyl-CoA:ACP transacylase (MAT), MGDG synthase (MGDS) and TAG lipase (TAGL).

The limitation of the carbon and energy in the cell limits the microalgal productivities. The strategies described in Chapter 7 to improve the upstream part of metabolism could overcome this bottleneck and generate strains capable of producing value-added chemicals. We have also discussed various strategies to improve the carbon fixation in microalgae by engineering the rate-limiting Rubisco enzyme in Chapter 7. Rubisco in unicellular photosynthetic organisms is reported to have higher catalytic rates compared to Rubisco enzymes in higher plants [431]. On the other hand, evolution has provided the Rubisco enzymes from higher plants with improved specificity towards carbon dioxide to limit the oxygenation via this enzyme [431, 432]. The carbon concentrating mechanism is a natural strategy that concentrates carbon dioxide around Rubisco to improve the carbon fixation [441]. However, this mechanism is not observed in majority of the higher plants that uses C₃ photosynthesis and some of the microalgal strains [598]. Recent studies have been successful in elucidating the mechanism of carbon concentration in microalgae such as *Chlamydomonas*. The implementation of this technique is hypothesized to improve the productivity in plants up to 60% [599]. The pyrenoid, a phase-separated organelle facilitates the carbon concentration mechanism in *Chlamydomonas* [598]. Fundamental research into the pyrenoid formation was successful in elucidating the structure and proteins involved in the formation of this organelle [598]. Recently, a proto-pyrenoid system was synthesized in plants by expressing various proteins from microalgae involved in pyrenoid formation [600, 601]. Further developments in this research line could establish a functional pyrenoid system in plants to improve the carbon fixation. Implementing this research line in microalgae such as *Nannochloropsis oceanica* that lacks this organelle could be a promising approach to improve microalgal productivities.

Coming to the downstream part of metabolism, the strategies should focus on improving the flux towards the product of interest or introduce heterologous pathways to produce novel compounds in the host strain. In this regard, *Nannochloropsis* is a potential strain to produce TAG molecules with a tailored fatty acid profile. The TAG molecules with SFAs are favourable for biofuel production while the TAGs with PUFAs are valuable for producing nutraceuticals [602]. The diverse fatty acid profile of *Nannochloropsis* comprise of SFAs, MUFAs and PUFAs with TAG being a dominant lipid class under nitrogen starvation. Therefore, *Nannochloropsis* can be an optimal strain for tailoring the TAG species with varying fatty acid profiles. The overexpression of EPA specific LPAAT as mentioned in Chapter 6 was streamlined to channel more EPA into the sn₂ position of TAG and thereby improve the *de novo* EPA production and accumulation in TAG. This approach increased the lipid content and EPA accumulation in TAG under nitrogen starved conditions in *Nannochloropsis*. However, further studies are required to confirm the enrichment of EPA specifically at the sn₂ position. Developing a synthetic Kennedy pathway optimized for producing TAG molecules with specific fatty acid profile and can be the best approach for producing tailored TAGs. Identifying the substrate specificity of acyltransferase enzymes in the Kennedy pathway from various organisms could facilitate the development of a synthetic Kennedy pathway in *Nannochloropsis* tailored for producing the TAG molecules with desired fatty acids.

Concluding Remarks

The development of high throughput genome engineering tools is inevitable for elucidating the metabolism, for development of metabolic models to identify potential target sites for metabolic engineering, and finally for performing these metabolic engineering in the host strain. The development of CRISPR-Cas systems has substantially accelerated the fundamental research into the metabolism of microalgae as well as in other organisms. As reviewed in Chapter 2, the CRISPR-Cas systems were first reported in microalgae in 2014 and since then various techniques and tools based on Cas proteins were developed for multiple microalgal strains.

The research described in this thesis aimed at developing a CRISPR-Cas based genetic toolbox that can generate markerless mutants with high efficiencies. As presented in Chapters 3 and 4, we initially focussed on developing an RNP based approach for generating HDR based precise mutants in *Nannochloropsis*. We observed that Cas12a could be a better Cas variant for engineering *Nannochloropsis*. We optimized the RNP based approach for isolating scarless and markerless mutants as these mutants are accepted as non-GMOs in some countries [587]. In Chapter 4, we further elaborated the genetic toolbox by developing a plasmid-based system using Cas12a for targeted gene knockout. The ability of Cas12a to process the CRISPR array was exploited to generate multiplexed mutants by targeting three loci on the genome in a single transformation. Furthermore, a CRISPR interference system based on dead Cas proteins was developed to efficiently downregulate the target genes in *Nannochloropsis*.

The second part of the thesis focussed on the metabolic engineering of *Nannochloropsis* to improve lipid accumulation and EPA production. To this end, as described in Chapter 5, random insertional mutagenesis was performed to create a library of *N. oceanica* mutants. High throughput selection based on FACS isolated multiple mutants with increased lipid production without affecting the growth rates. The tracking of the insertional cassette revealed various potential targets and novel insights into lipid overproduction. The mutant named HLM23 with disruption in the APETALA2-like TF had the most improved lipid productivity and a Cas12a based knockout of this protein in WT *N. oceanica* IMET1 confirmed the high lipid producing phenotype. Subsequently, to increase the EPA production and EPA accumulation in TAG, an EPA specific LPAAT was heterologously expressed in *N. oceanica* IMET1. A modest improvement in the EPA and lipid content was observed in *N. oceanica* IMET1 with increased EPA content in the TAG fraction of neutral lipids under nitrogen stress. However, EPA was not increased under optimal growth conditions and further research is required to confirm the accumulation of EPA at the sn2 position of TAG. Co-expression of other acyltransferase enzymes in the Kennedy pathway with high specificity towards EPA and EPA rich glycerolipids could be an interesting approach to channel more EPA into TAGs. The current metabolic engineering strategies to improve the microalgal productivities are limited by factors such as available carbon and reducing power. Thus, future studies should focus on improving the upstream part of the metabolism. In this regard, a detailed review on potential strategies to improve microalgal productivity are discussed in Chapter 7. The strategies to improve the upstream metabolism will require the implementation of entire pathways or major rewiring of the

native pathways. Nevertheless, with the established genetic toolbox that consists of high throughput genome engineering techniques, these ambitious engineering projects are feasible.

To conclude, the advancements reported in this thesis developed high throughput CRISPR-Cas mediated genome editing approaches for *Nannochloropsis oceanica*, with the potential to use (directly, or after adaptation) in other genetically accessible microalgal strains. The metabolic engineering studies have provided novel insights into the lipid production pathway and facilitated the development of mutants with increased lipid productivity and EPA content. Moreover, the bottlenecks of present metabolic engineering strategies were pointed out to stimulate the development of alternative approaches to further improve the productivities of microalgae.



References

- [1] Léonard, J., *Carte du degré d'exploration floristique de l'Afrique au sud du Sahara: Map of the extent of floristic exploration in Africa south of the Sahara*, Instituto Botanico dell'Università 1965.
- [2] Farrar, W., Tecuitlatl; a glimpse of Aztec food technology. *Nature* 1966, 211, 341-342.
- [3] García, J. L., de Vicente, M., Galán, B., Microalgae, old sustainable food and fashion nutraceuticals. *Microbial biotechnology* 2017, 10, 1017-1024.
- [4] Mall, A., Sobotta, J., Huber, C., Tschirner, C., *et al.*, Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. *Science* 2018, 359, 563-567.
- [5] Sousa, I., Gouveia, L., Batista, A. P., Raymundo, A., Bandarra, N. M., Microalgae in novel food products. *Food chemistry research developments* 2008, 75-112.
- [6] Edelmann, M., Aalto, S., Chamlagain, B., Kariluoto, S., Piironen, V., Riboflavin, niacin, folate and vitamin B12 in commercial microalgae powders. *Journal of Food Composition and Analysis* 2019, 82, 103226.
- [7] Carballo-Cárdenas, E. C., Tuan, P. M., Janssen, M., Wijffels, R. H., Vitamin E (α -tocopherol) production by the marine microalgae *Dunaliella tertiolecta* and *Tetraselmis suecica* in batch cultivation. *Biomolecular Engineering* 2003, 20, 139-147.
- [8] Pourkarimi, S., Hallajisani, A., Alizadehdakhel, A., Golzary, A., Factors affecting production of beta-carotene from *Dunaliella salina* microalgae. *Biocatalysis and Agricultural Biotechnology* 2020, 101771.
- [9] Del Campo, J., Rodríguez, H., Moreno, J., Vargas, M., *et al.*, Accumulation of astaxanthin and lutein in *Chlorella zofingiensis* (Chlorophyta). *Applied microbiology and biotechnology* 2004, 64, 848-854.
- [10] Zhao, L., Chen, G., Zhao, G., Hu, X., Optimization of microwave-assisted extraction of astaxanthin from *Haematococcus pluvialis* by response surface methodology and antioxidant activities of the extracts. *Separation Science and Technology* 2009, 44, 243-262.
- [11] Rodríguez-Sáiz, M., de la Fuente, J. L., Barredo, J. L., *Xanthophyllomyces dendrorhous* for the industrial production of astaxanthin. *Applied microbiology and biotechnology* 2010, 88, 645-658.
- [12] Del Campo, J. A., García-González, M., Guerrero, M. G., Outdoor cultivation of microalgae for carotenoid production: current state and perspectives. *Applied microbiology and biotechnology* 2007, 74, 1163-1174.
- [13] Zhang, Y., Liu, Z., Sun, J., Xue, C., Mao, X., Biotechnological production of zeaxanthin by microorganisms. *Trends in Food Science & Technology* 2018, 71, 225-234.
- [14] Gong, M., Bassi, A., Carotenoids from microalgae: A review of recent developments. *Biotechnology Advances* 2016, 34, 1396-1412.
- [15] Kim, S. M., Jung, Y.-J., Kwon, O.-N., Cha, K. H., *et al.*, A potential commercial source of fucoxanthin extracted from the microalga *Phaeodactylum tricorutum*. *Applied biochemistry and biotechnology* 2012, 166, 1843-1855.
- [16] Petrushkina, M., Gusev, E., Sorokin, B., Zotko, N., *et al.*, Fucoxanthin production by heterokont microalgae. *Algal Research* 2017, 24, 387-393.
- [17] Laje, K., Seger, M., Dungan, B., Cooke, P., *et al.*, Phytoene accumulation in the novel microalga *Chlorococcum* sp. using the pigment synthesis inhibitor fluridone. *Marine drugs* 2019, 17, 187.

- [18] León, R., Vila, M., Hernánz, D., Vilchez, C., Production of phytoene by herbicide-treated microalgae *Dunaliella bardawil* in two-phase systems. *Biotechnology and bioengineering* 2005, 92, 695-701.
- [19] Srinivasan, R., Babu, S., Gothandam, K., Accumulation of phytoene, a colorless carotenoid by inhibition of phytoene desaturase (PDS) gene in *Dunaliella salina* V-101. *Bioresource technology* 2017, 242, 311-318.
- [20] Meléndez-Martínez, A. J., Mapelli-Brahm, P., Benítez-González, A., Stinco, C. M., A comprehensive review on the colorless carotenoids phytoene and phytofluene. *Archives of Biochemistry and Biophysics* 2015, 572, 188-200.
- [21] Soontornchaiboon, W., Joo, S. S., Kim, S. M., Anti-inflammatory effects of violaxanthin isolated from microalga *Chlorella ellipsoidea* in RAW 264.7 macrophages. *Biological and Pharmaceutical Bulletin* 2012, 35, 1137-1144.
- [22] Abrahamsson, V., Rodríguez-Meizoso, I., Turner, C., Determination of carotenoids in microalgae using supercritical fluid extraction and chromatography. *Journal of chromatography A* 2012, 1250, 63-68.
- [23] Sonani, R. R., Rastogi, R. P., Patel, R., Madamwar, D., Recent advances in production, purification and applications of phycobiliproteins. *World journal of biological chemistry* 2016, 7, 100.
- [24] Bishop, W., Zubeck, H., Evaluation of microalgae for use as nutraceuticals and nutritional supplements. *J Nutr Food Sci* 2012, 2, 1-6.
- [25] Adarme-Vega, T. C., Thomas-Hall, S. R., Schenk, P. M., Towards sustainable sources for omega-3 fatty acids production. *Current opinion in biotechnology* 2014, 26, 14-18.
- [26] Delattre, C., Pierre, G., Laroche, C., Michaud, P., Production, extraction and characterization of microalgal and cyanobacterial exopolysaccharides. *Biotechnology advances* 2016, 34, 1159-1179.
- [27] Luo, X., Su, P., Zhang, W., Advances in microalgae-derived phytosterols for functional food and pharmaceutical applications. *Marine drugs* 2015, 13, 4231-4254.
- [28] Olofsson, M., Lamela, T., Nilsson, E., Bergé, J. P., *et al.*, Seasonal variation of lipids and fatty acids of the microalgae *Nannochloropsis oculata* grown in outdoor large-scale photobioreactors. *Energies* 2012, 5, 1577-1592.
- [29] Guschina, I. A., Harwood, J. L., Lipids and lipid metabolism in eukaryotic algae. *Progress in lipid research* 2006, 45, 160-186.
- [30] Thompson Jr, G. A., Lipids and membrane function in green algae. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism* 1996, 1302, 17-45.
- [31] Wada, H., Murata, N., Lipids in photosynthesis. *Essential and* 1998.
- [32] Chew, K. W., Chia, S. R., Show, P. L., Ling, T. C., Chang, J.-s., Biofuels from microbial lipids, *Bioreactors for Microbial Biomass and Energy Conversion*, Springer 2018, pp. 359-388.
- [33] Guckert, J. B., Cooksey, K. E., Triglyceride accumulation and fatty acid profile changes in *Chlorella* (Chlorophyta) during high Ph-induced cell cycle Inhibition 1. *Journal of Phycology* 1990, 26, 72-79.
- [34] Harwood, J., Lipids in Photosynthesis: Structure, Function and Genetics, *Membrane Lipids in Algae*, Kluwer Academic Publisher Netherlands 1998.
- [35] Pohl, P., Zurheide, F., New York: Walter de Gruyter 1979.
- [36] Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., *et al.*, Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *The plant journal* 2008, 54, 621-639.

- [37] Goncalves, E. C., Wilkie, A. C., Kirst, M., Rathinasabapathi, B., Metabolic regulation of triacylglycerol accumulation in the green algae: identification of potential targets for engineering to improve oil yield. *Plant biotechnology journal* 2016, 14, 1649-1660.
- [38] Vasudevan, V., Stratton, R. W., Pearlson, M. N., Jersey, G. R., *et al.*, Environmental performance of algal biofuel technology options. *Environmental science & technology* 2012, 46, 2451-2459.
- [39] Wijffels, R. H., Barbosa, M. J., An outlook on microalgal biofuels. *Science* 2010, 329, 796-799.
- [40] Remmers, I. M., Wijffels, R. H., Barbosa, M. J., Lamers, P. P., Can we approach theoretical lipid yields in microalgae? *Trends in biotechnology* 2018, 36, 265-276.
- [41] Klok, A., Lamers, P., Martens, D., Draaisma, R., Wijffels, R., Edible oils from microalgae: insights in TAG accumulation. *Trends in biotechnology* 2014, 32, 521-528.
- [42] Benvenuti, G., Ruiz, J., Lamers, P. P., Bosma, R., *et al.*, Towards microalgal triglycerides in the commodity markets. *Biotechnology for biofuels* 2017, 10, 1-10.
- [43] Ngan, C. Y., Wong, C.-H., Choi, C., Yoshinaga, Y., *et al.*, Lineage-specific chromatin signatures reveal a regulator of lipid metabolism in microalgae. *Nature Plants* 2015, 1, 1-12.
- [44] Goold, H. D., Nguyen, H. M., Kong, F., Beyly-Adriano, A., *et al.*, Whole genome re-sequencing identifies a quantitative trait locus repressing carbon reserve accumulation during optimal growth in *Chlamydomonas reinhardtii*. *Scientific reports* 2016, 6, 1-12.
- [45] Schulz-Raffelt, M., Chochois, V., Auroy, P., Cuiné, S., *et al.*, Hyper-accumulation of starch and oil in a *Chlamydomonas* mutant affected in a plant-specific DYRK kinase. *Biotechnology for biofuels* 2016, 9, 1-12.
- [46] Trentacoste, E. M., Shrestha, R. P., Smith, S. R., Glé, C., *et al.*, Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *Proceedings of the National Academy of Sciences* 2013, 110, 19748-19753.
- [47] Levitan, O., Dinamarca, J., Zelzion, E., Gorbunov, M. Y., Falkowski, P. G., An RNA interference knock-down of nitrate reductase enhances lipid biosynthesis in the diatom *Phaeodactylum tricorutum*. *The Plant Journal* 2015, 84, 963-973.
- [48] Daboussi, F., Leduc, S., Maréchal, A., Dubois, G., *et al.*, Genome engineering empowers the diatom *Phaeodactylum tricorutum* for biotechnology. *Nature communications* 2014, 5, 1-7.
- [49] Radakovits, R., Jinkerson, R. E., Fuerstenberg, S. I., Tae, H., *et al.*, Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. *Nature communications* 2012, 3, 1-11.
- [50] Rodolfi, L., Chini Zittelli, G., Bassi, N., Padovani, G., *et al.*, Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnology and bioengineering* 2009, 102, 100-112.
- [51] Ma, X.-N., Chen, T.-P., Yang, B., Liu, J., Chen, F., Lipid production from *Nannochloropsis*. *Marine drugs* 2016, 14, 61.
- [52] Darwish, R., Gedi, M. A., Eakpetch, P., Assaye, H., *et al.*, *Chlamydomonas reinhardtii* Is a Potential Food Supplement with the Capacity to Outperform *Chlorella* and *Spirulina*. *Applied Sciences* 2020, 10, 6736.
- [53] Bougaran, G., Rouxel, C., Dubois, N., Kaas, R., *et al.*, Enhancement of neutral lipid productivity in the microalga *Isochrysis affinis Galbana* (T-Iso) by a mutation-selection procedure. *Biotechnology and bioengineering* 2012, 109, 2737-2745.
- [54] Vanitha, A., Narayan, M., Murthy, K., Ravishankar, G., Comparative study of lipid composition of two halotolerant alga, *Dunaliella bardawil* and *Dunaliella salina*. *International journal of food sciences and nutrition* 2007, 58, 373-382.

- [55] Valenzuela, J., Mazurie, A., Carlson, R. P., Gerlach, R., *et al.*, Potential role of multiple carbon fixation pathways during lipid accumulation in *Phaeodactylum tricornutum*. *Biotechnology for biofuels* 2012, 5, 1-17.
- [56] Narayan, M., Manoj, G., Vatchravelu, K., Bhagyalakshmi, N., Mahadevaswamy, M., Utilization of glycerol as carbon source on the growth, pigment and lipid production in *Spirulina platensis*. *International journal of food sciences and nutrition* 2005, 56, 521-528.
- [57] Li, L., Cui, J., Liu, Q., Ding, Y., Liu, J., Screening and phylogenetic analysis of lipid-rich microalgae. *Algal Research* 2015, 11, 381-386.
- [58] Hibberd, D., Notes on the taxonomy and nomenclature of the algal classes Eustigmatophyceae and Tribophyceae (synonym Xanthophyceae). *Botanical journal of the linnean society* 1981, 82, 93-119.
- [59] Andersen, R. A., Brett, R. W., Potter, D., Sexton, J. P., Phylogeny of the Eustigmatophyceae based upon 18S rDNA, with emphasis on Nannochloropsis. *Protist* 1998, 149, 61-74.
- [60] Ott, D. W., Oldham-Ott, C. K., Rybalka, N., Friedl, T., Xanthophyte, eustigmatophyte, and raphidophyte algae, *Freshwater Algae of North America*, Elsevier 2015, pp. 485-536.
- [61] Ma, Y., Wang, Z., Yu, C., Yin, Y., Zhou, G., Evaluation of the potential of 9 Nannochloropsis strains for biodiesel production. *Bioresource technology* 2014, 167, 503-509.
- [62] Doan, T. T. Y., Sivaloganathan, B., Obbard, J. P., Screening of marine microalgae for biodiesel feedstock. *Biomass and Bioenergy* 2011, 35, 2534-2544.
- [63] Vieler, A., Wu, G., Tsai, C.-H., Bullard, B., *et al.*, Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. *PLoS Genet* 2012, 8, e1003064.
- [64] Hoffmann, M., Marxen, K., Schulz, R., Vanselow, K. H., TFA and EPA productivities of *Nannochloropsis salina* influenced by temperature and nitrate stimuli in turbidostatic controlled experiments. *Marine drugs* 2010, 8, 2526-2545.
- [65] Chen, C.-Y., Chen, Y.-C., Huang, H.-C., Huang, C.-C., *et al.*, Engineering strategies for enhancing the production of eicosapentaenoic acid (EPA) from an isolated microalga *Nannochloropsis oceanica* CY2. *Bioresource technology* 2013, 147, 160-167.
- [66] Mohebi-Nejad, A., Bikdeli, B., Omega-3 supplements and cardiovascular diseases. *Tanaffos* 2014, 13, 6.
- [67] Schultz, H., Retail omega-3s sales to hit \$34.7 billion in 2016, report predicts. 2012.[cited 2013 September 17]. Available on: <http://www.nutraingredients-usa.com/Markets/Retail-omega-3s-sales-to-hit-34.7-billion-in-2016-report-predicts>.
- [68] Martins, D. A., Custódio, L., Barreira, L., Pereira, H., *et al.*, Alternative sources of n-3 long-chain polyunsaturated fatty acids in marine microalgae. *Marine drugs* 2013, 11, 2259-2281.
- [69] Sharma, K., Schenk, P. M., Rapid induction of omega-3 fatty acids (EPA) in *Nannochloropsis* sp. by UV-C radiation. *Biotechnology and bioengineering* 2015, 112, 1243-1249.
- [70] Meng, Y., Jiang, J., Wang, H., Cao, X., *et al.*, The characteristics of TAG and EPA accumulation in *Nannochloropsis oceanica* IMET1 under different nitrogen supply regimes. *Bioresource technology* 2015, 179, 483-489.
- [71] Carpinelli, E. C., Telatin, A., Vitulo, N., Forcato, C., *et al.*, Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion. *Molecular plant* 2014, 7, 323-335.
- [72] Wang, D., Ning, K., Li, J., Hu, J., *et al.*, *Nannochloropsis* genomes reveal evolution of microalgal oleaginous traits. *PLoS Genet* 2014, 10, e1004094.

- [73] Mühlroth, A., Winge, P., El Assimi, A., Jouhet, J., *et al.*, Mechanisms of phosphorus acquisition and lipid class remodeling under P limitation in a marine microalga. *Plant Physiology* 2017, 175, 1543-1559.
- [74] Li, J., Han, D., Wang, D., Ning, K., *et al.*, Choreography of transcriptomes and lipidomes of *Nannochloropsis* reveals the mechanisms of oil synthesis in microalgae. *The Plant Cell* 2014, 26, 1645-1665.
- [75] Poliner, E., Panchy, N., Newton, L., Wu, G., *et al.*, Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP 1779 under light/dark cycles. *The Plant Journal* 2015, 83, 1097-1113.
- [76] Alboresi, A., Perin, G., Vitulo, N., Diretto, G., *et al.*, Light remodels lipid biosynthesis in *Nannochloropsis gaditana* by modulating carbon partitioning between organelles. *Plant Physiology* 2016, 171, 2468-2482.
- [77] Cha, T.-S., Chen, C.-F., Yee, W., Aziz, A., Loh, S.-H., Cinnamic acid, coumarin and vanillin: Alternative phenolic compounds for efficient *Agrobacterium*-mediated transformation of the unicellular green alga, *Nannochloropsis* sp. *Journal of microbiological methods* 2011, 84, 430-434.
- [78] Kilian, O., Benemann, C. S., Niyogi, K. K., Vick, B., High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proceedings of the National Academy of Sciences* 2011, 108, 21265-21269.
- [79] Poliner, E., Takeuchi, T., Du, Z.-Y., Benning, C., Farré, E. M., Nontransgenic marker-free gene disruption by an episomal CRISPR system in the oleaginous microalga, *Nannochloropsis oceanica* CCMP1779. *ACS synthetic biology* 2018, 7, 962-968.
- [80] Ajjawi, I., Verruto, J., Aqui, M., Soriaga, L. B., *et al.*, Lipid production in *Nannochloropsis gaditana* is doubled by decreasing expression of a single transcriptional regulator. *Nature biotechnology* 2017, 35, 647-652.
- [81] Poliner, E., Pulman, J. A., Zienkiewicz, K., Childs, K., *et al.*, A toolkit for *Nannochloropsis oceanica* CCMP 1779 enables gene stacking and genetic engineering of the eicosapentaenoic acid pathway for enhanced long-chain polyunsaturated fatty acid production. *Plant biotechnology journal* 2018, 16, 298-309.
- [82] Sizova, I., Greiner, A., Awasthi, M., Kateriya, S., Hegemann, P., Nuclear gene targeting in *Chlamydomonas* using engineered zinc-finger nucleases. *The Plant Journal* 2013, 73, 873-882.
- [83] Kurita, T., Moroi, K., Iwai, M., Okazaki, K., *et al.*, Efficient and multiplexable genome editing using Platinum TALENs in oleaginous microalga, *Nannochloropsis oceanica* NIES-2145. *Genes to Cells* 2020, 25, 695-702.
- [84] Moellering, E. R., Benning, C., RNA interference silencing of a major lipid droplet protein affects lipid droplet size in *Chlamydomonas reinhardtii*. *Eukaryotic cell* 2010, 9, 97-106.
- [85] Rohr, J., Sarkar, N., Balenger, S., Jeong, B. r., Cerutti, H., Tandem inverted repeat system for selection of effective transgenic RNAi strains in *Chlamydomonas*. *The Plant Journal* 2004, 40, 611-621.
- [86] Hildebrand, M., Manandhar-Shrestha, K., Abbriano, R., Effects of chrysolaminarin synthase knockdown in the diatom *Thalassiosira pseudonana*: Implications of reduced carbohydrate storage relative to green algae. *Algal research* 2017, 23, 66-77.
- [87] Costa, B. S., Sachse, M., Jungandreas, A., Bartulos, C. R., *et al.*, Aureochrome 1a is involved in the photoacclimation of the diatom *Phaeodactylum tricornutum*. *PLoS one* 2013, 8, e74451.

- [88] Ma, X., Yao, L., Yang, B., Lee, Y. K., *et al.*, RNAi-mediated silencing of a pyruvate dehydrogenase kinase enhances triacylglycerol biosynthesis in the oleaginous marine alga *Nannochloropsis salina*. *Scientific reports* 2017, 7, 1-12.
- [89] Wei, L., Xin, Y., Wang, Q., Yang, J., *et al.*, RNA i-based targeted gene knockdown in the model oleaginous microalgae *Nannochloropsis oceanica*. *The Plant Journal* 2017, 89, 1236-1250.
- [90] Wei, H., Shi, Y., Ma, X., Pan, Y., *et al.*, A type-I diacylglycerol acyltransferase modulates triacylglycerol biosynthesis and fatty acid composition in the oleaginous microalga, *Nannochloropsis oceanica*. *Biotechnology for biofuels* 2017, 10, 1-18.
- [91] Grissa, I., Vergnaud, G., Pourcel, C., The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC bioinformatics* 2007, 8, 1-10.
- [92] Mali, P., Yang, L., Esvelt, K. M., Aach, J., *et al.*, RNA-guided human genome engineering via Cas9. *Science* 2013, 339, 823-826.
- [93] Jinek, M., East, A., Cheng, A., Lin, S., *et al.*, RNA-programmed genome editing in human cells. *elife* 2013, 2, e00471.
- [94] Jiang, W., Bikard, D., Cox, D., Zhang, F., Marraffini, L. A., RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* 2013, 31, 233-239.
- [95] Cong, L., Ran, F. A., Cox, D., Lin, S., *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013, 339, 819-823.
- [96] Jiang, W., Zhou, H., Bi, H., Fromm, M., *et al.*, Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic acids research* 2013, 41, e188-e188.
- [97] Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., *et al.*, CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature protocols* 2013, 8, 2180-2196.
- [98] Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., *et al.*, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 2013, 154, 442-451.
- [99] Bikard, D., Jiang, W., Samai, P., Hochschild, A., *et al.*, Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research* 2013, 41, 7429-7437.
- [100] Harrington, L. B., Paez-Espino, D., Staahl, B. T., Chen, J. S., *et al.*, A thermostable Cas9 with increased lifetime in human plasma. *Nature communications* 2017, 8, 1-8.
- [101] Mougias, I., Mohanraju, P., Bosma, E. F., Vrouwe, V., *et al.*, Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nature communications* 2017, 8, 1-11.
- [102] Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., *et al.*, Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 2015, 163, 759-771.
- [103] Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., *et al.*, Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nature biotechnology* 2017, 35, 31-34.
- [104] Wijffels, R. H., Barbosa, M. J., Eppink, M. H., Microalgae for the production of bulk chemicals and biofuels. *Biofuels, Bioproducts and Biorefining* 2010, 4, 287-295.
- [105] Wang, B., Li, Y., Wu, N., Lan, C. Q., CO₂ bio-mitigation using microalgae. *Applied microbiology and biotechnology* 2008, 79, 707-718.
- [106] Sizova, I., Greiner, A., Awasthi, M., Kateriya, S., Hegemann, P., Nuclear gene targeting in *Chlamydomonas* using engineered zinc-finger nucleases. *The Plant Journal* 2013, 73, 873-882.
- [107] Daboussi, F., Leduc, S., Maréchal, A., Dubois, G., *et al.*, Genome engineering empowers the diatom *Phaeodactylum tricoratum* for biotechnology. *Nature communications* 2014, 5, 3831.

- [108] Lozano, J. C., Schatt, P., Botebol, H., Vergé, V., *et al.*, Efficient gene targeting and removal of foreign DNA by homologous recombination in the picoeukaryote *Ostreococcus*. *The Plant Journal* 2014, 78, 1073-1083.
- [109] Jiang, W., Brueggeman, A. J., Horken, K. M., Plucinak, T. M., Weeks, D. P., Successful transient expression of Cas9 and single guide RNA genes in *Chlamydomonas reinhardtii*. *Eukaryotic cell* 2014, 13, 1465-1469.
- [110] Charpentier, E., Marraffini, L. A., Harnessing CRISPR-Cas9 immunity for genetic engineering. *Current opinion in microbiology* 2014, 19, 114-119.
- [111] Doudna, J. A., Charpentier, E., The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014, 346, 1258096.
- [112] Hsu, P. D., Lander, E. S., Zhang, F., Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014, 157, 1262-1278.
- [113] Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., *et al.*, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 2013, 152, 1173-1183.
- [114] Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., *et al.*, Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 2016, 353, aad5147.
- [115] Endo, A., Masafumi, M., Kaya, H., Toki, S., Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from *Francisella novicida*. *Scientific reports* 2016, 6, 38169.
- [116] Hur, J. K., Kim, K., Been, K. W., Baek, G., *et al.*, Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nature biotechnology* 2016, 34, 807-808.
- [117] Kim, Y., Cheong, S.-A., Lee, J. G., Lee, S.-W., *et al.*, Generation of knockout mice by Cpf1-mediated gene targeting. *Nature biotechnology* 2016, 34, 808-810.
- [118] Kim, D., Kim, J., Hur, J. K., Been, K. W., *et al.*, Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nature biotechnology* 2016.
- [119] Kleinstiver, B. P., Tsai, S. Q., Prew, M. S., Nguyen, N. T., *et al.*, Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nature biotechnology* 2016, 34, 869-874.
- [120] Tóth, E., Weinhardt, N., Bencsura, P., Huszár, K., *et al.*, Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells. *Biology direct* 2016, 11, 46.
- [121] Xu, R., Qin, R., Li, H., Li, D., *et al.*, Generation of targeted mutant rice using a CRISPR-Cpf1 system. *Plant biotechnology journal* 2017, 15, 713-717.
- [122] Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., *et al.*, Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nature biotechnology* 2017, 35, 31-34.
- [123] Verwaal, R., Buiting-Wiessenhaan, N., Dalhuijsen, S., Roubos, J. A., CRISPR/Cpf1 enables fast and simple genome editing of *Saccharomyces cerevisiae*. *Yeast* 2017.
- [124] Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., Charpentier, E., The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* 2016, 532, 517-521.
- [125] Moreno-Mateos, M. A., Fernandez, J. P., Rouet, R., Vejnar, C. E., *et al.*, CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nature communications* 2017, 8, 2024.
- [126] Harris, E. H., *Chlamydomonas* as a model organism. *Annual review of plant biology* 2001, 52, 363-406.
- [127] Jiang, W. Z., Weeks, D. P., A gene-within-a-gene Cas9/sgRNA hybrid construct enables gene editing and gene replacement strategies in *Chlamydomonas reinhardtii*. *Algal Research* 2017.

- [128] Kindle, K. L., Schnell, R. A., Fernández, E., Lefebvre, P. A., Stable nuclear transformation of *Chlamydomonas* using the *Chlamydomonas* gene for nitrate reductase. *The Journal of cell biology* 1989, *109*, 2589-2601.
- [129] Li, H. L., Fujimoto, N., Sasakawa, N., Shirai, S., *et al.*, Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. *Stem cell reports* 2015, *4*, 143-154.
- [130] Pyne, M. E., Moo-Young, M., Chung, D. A., Chou, C. P., Coupling the CRISPR/Cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. *Applied and environmental microbiology* 2015, *81*, 5103-5114.
- [131] Turan, S., Zehe, C., Kuehle, J., Qiao, J., Bode, J., Recombinase-mediated cassette exchange (RMCE)—a rapidly-expanding toolbox for targeted genomic modifications. *Gene* 2013, *515*, 1-27.
- [132] Greiner, A., Kelterborn, S., Evers, H., Kreimer, G., *et al.*, Targeting of photoreceptor genes in *Chlamydomonas reinhardtii* via zinc-finger nucleases and CRISPR/Cas9. *The Plant Cell* 2017, tpc.00659.02017.
- [133] Schroda, M., Blöcker, D., Beck, C. F., The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. *The plant journal* 2000, *21*, 121-131.
- [134] Halpin-Dohnalek, M. I., Marth, E. H., *Staphylococcus aureus*: production of extracellular compounds and behavior in foods—a review. *Journal of food protection* 1989, *52*, 267-282.
- [135] Zangerl, P., Asperger, H., Media used in the detection and enumeration of *Staphylococcus aureus*. *Progress in Industrial Microbiology* 2003, *37*, 91-110.
- [136] Kao, P.-H., Ng, I.-S., CRISPRi mediated phosphoenolpyruvate carboxylase regulation to enhance the production of lipid in *Chlamydomonas reinhardtii*. *Bioresource Technology* 2017.
- [137] Shin, S.-E., Lim, J.-M., Koh, H. G., Kim, E. K., *et al.*, CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*. *Scientific reports* 2016, *6*, 27810.
- [138] Baek, K., Kim, D. H., Jeong, J., Sim, S. J., *et al.*, DNA-free two-gene knockout in *Chlamydomonas reinhardtii* via CRISPR-Cas9 ribonucleoproteins. *Scientific Reports* 2016, *6*, 30620.
- [139] Baek, K., Yu, J., Jeong, J., Sim, S. J., *et al.*, Photoautotrophic production of macular pigment in a *Chlamydomonas reinhardtii* strain generated by using DNA-free CRISPR-Cas9 RNP-mediated mutagenesis. *Biotechnology and bioengineering* 2017.
- [140] Ferenczi, A., Pyott, D. E., Xipnitou, A., Molnar, A., Efficient targeted DNA editing and replacement in *Chlamydomonas reinhardtii* using Cpf1 ribonucleoproteins and single-stranded DNA. *Proceedings of the National Academy of Sciences* 2017, *114*, 13567-13572.
- [141] Pan, K., Qin, J., Li, S., Dai, W., *et al.*, Nuclear monoploidy and asexual propagation of *Nannochloropsis oceanica* (Eustigmatophyceae) as revealed by its genome sequence. *Journal of phycology* 2011, *47*, 1425-1432.
- [142] Radakovits, R., Jinkerson, R. E., Fuerstenberg, S. I., Tae, H., *et al.*, Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. *Nature communications* 2012, *3*, 686.
- [143] Vieler, A., Wu, G., Tsai, C.-H., Bullard, B., *et al.*, Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. *PLoS Genetics* 2012, *8*, e1003064.
- [144] Wei, L., Huang, X., Huang, Z., Zhou, Z., Orthogonal test design for optimization of lipid accumulation and lipid property in *Nannochloropsis oculata* for biodiesel production. *Bioresource technology* 2013, *147*, 534-538.

- [145] Wang, D., Ning, K., Li, J., Hu, J., *et al.*, Nannochloropsis genomes reveal evolution of microalgal oleaginous traits. *PLoS genetics* 2014, 10, e1004094.
- [146] Wang, Q., Lu, Y., Xin, Y., Wei, L., *et al.*, Genome editing of model oleaginous microalgae *Nannochloropsis* spp. by CRISPR/Cas9. *The Plant Journal* 2016, 88, 1071-1081.
- [147] Ajjawi, I., Verruto, J., Aqui, M., Soriaga, L. B., *et al.*, Lipid production in *Nannochloropsis gaditana* is doubled by decreasing expression of a single transcriptional regulator. *Nature biotechnology* 2017, 35, 647.
- [148] Bozarth, A., Maier, U.-G., Zauner, S., Diatoms in biotechnology: modern tools and applications. *Applied microbiology and biotechnology* 2009, 82, 195-201.
- [149] Lebeau, T., Robert, J.-M., Diatom cultivation and biotechnologically relevant products. Part I: Cultivation at various scales. *Applied Microbiology and Biotechnology* 2003, 60, 612-623.
- [150] Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., Winge, P., A CRISPR/Cas9 system adapted for gene editing in marine algae. *Scientific reports* 2016, 6.
- [151] Hopes, A., Nekrasov, V., Kamoun, S., Mock, T., Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant methods* 2016, 12, 49.
- [152] Jacobs, T. B., LaFayette, P. R., Schmitz, R. J., Parrott, W. A., Targeted genome modifications in soybean with CRISPR/Cas9. *BMC biotechnology* 2015, 15, 16.
- [153] Falcatore, A., Casotti, R., Leblanc, C., Abrescia, C., Bowler, C., Transformation of nonselectable reporter genes in marine diatoms. *Marine Biotechnology* 1999, 1, 239-251.
- [154] Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D., Kamoun, S., Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nature biotechnology* 2013, 31, 691-693.
- [155] Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Nekrasov, V., Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant methods* 2013, 9, 39.
- [156] Brooks, C., Nekrasov, V., Lippman, Z. B., Van Eck, J., Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant physiology* 2014, 166, 1292-1297.
- [157] Belshaw, N., Grouneva, I., Aram, L., Gal, A., *et al.*, Efficient CRISPR/Cas-mediated homologous recombination in the model diatom *Thalassiosira pseudonana*. *bioRxiv* 2017, 215582.
- [158] Yu, J., Liberton, M., Cliften, P. F., Head, R. D., *et al.*, *Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO₂. *Scientific reports* 2015, 5.
- [159] Parmar, A., Singh, N. K., Pandey, A., Gnansounou, E., Madamwar, D., Cyanobacteria and microalgae: a positive prospect for biofuels. *Bioresource technology* 2011, 102, 10163-10172.
- [160] Yao, L., Cengic, I., Anfelt, J., Hudson, E. P., Multiple gene repression in cyanobacteria using CRISPRi. *ACS synthetic biology* 2015, 5, 207-212.
- [161] Huang, C.-H., Shen, C. R., Li, H., Sung, L.-Y., *et al.*, CRISPR interference (CRISPRi) for gene regulation and succinate production in cyanobacterium *S. elongatus* PCC 7942. *Microbial cell factories* 2016, 15, 196.
- [162] Gordon, G. C., Korosh, T. C., Cameron, J. C., Markley, A. L., *et al.*, CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Metabolic engineering* 2016, 38, 170-179.
- [163] Higo, A., Isu, A., Fukaya, Y., Ehira, S., Hisabori, T., Application of CRISPR interference for metabolic engineering of the heterocyst-forming multicellular cyanobacterium *Anabaena* sp. PCC 7120. *Plant and Cell Physiology* 2017.

- [164] Li, H., Shen, C. R., Huang, C.-H., Sung, L.-Y., *et al.*, CRISPR-Cas9 for the genome engineering of cyanobacteria and succinate production. *Metabolic engineering* 2016, 38, 293-302.
- [165] Wendt, K. E., Ungerer, J., Cobb, R. E., Zhao, H., Pakrasi, H. B., CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacterium *Synechococcus elongatus* UTEX 2973. *Microbial cell factories* 2016, 15, 115.
- [166] Collier, J. L., Grossman, A., A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. *The EMBO journal* 1994, 13, 1039.
- [167] Ungerer, J., Pakrasi, H. B., Cpf1 is a versatile tool for CRISPR genome editing across diverse species of cyanobacteria. *Scientific reports* 2016, 6, 39681.
- [168] Zerulla, K., Ludt, K., Soppa, J., The ploidy level of *Synechocystis* sp. PCC 6803 is highly variable and is influenced by growth phase and by chemical and physical external parameters. *Microbiology* 2016, 162, 730-739.
- [169] Jąkalski, M., Takeshita, K., Deblieck, M., Koyanagi, K. O., *et al.*, Comparative genomic analysis of retrogene repertoire in two green algae *Volvox carterii* and *Chlamydomonas reinhardtii*. *Biology direct* 2016, 11, 35.
- [170] Roy, S. W., Penny, D., A very high fraction of unique intron positions in the intron-rich diatom *Thalassiosira pseudonana* indicates widespread intron gain. *Molecular biology and evolution* 2007, 24, 1447-1457.
- [171] Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., *et al.*, The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 2008, 456, 239-244.
- [172] Lumberras, V., Stevens, D. R., Purton, S., Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. *The Plant Journal* 1998, 14, 441-447.
- [173] Rasala, B. A., Lee, P. A., Shen, Z., Briggs, S. P., *et al.*, Robust expression and secretion of Xylanase 1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. *PLoS one* 2012, 7, e43349.
- [174] Wang, M., Mao, Y., Lu, Y., Tao, X., Zhu, J.-k., Multiplex gene editing in rice using the CRISPR-Cpf1 system. *Molecular Plant* 2017.
- [175] Murovec, J., Pirc, Ž., Yang, B., New variants of CRISPR RNA guided genome editing enzymes. *Plant Biotechnology Journal* 2017.
- [176] Nakade, S., Yamamoto, T., Sakuma, T., Cas9, Cpf1 and C2c1/2/3-What's next? *Bioengineered* 2017, 1-9.
- [177] Harrington, L. B., Paez-Espino, D., Staahl, B. T., Chen, J. S., *et al.*, A thermostable Cas9 with increased lifetime in human plasma. *Nature Communications* 2017, 8, 1424.
- [178] Mougias, I., Mohanraju, P., Bosma, E. F., Vrouwe, V., *et al.*, Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nature communications* 2017, 8, 1647.
- [179] Chisti, Y., Biodiesel from microalgae. *Biotechnology advances* 2007, 25, 294-306.
- [180] Wen, Z.-Y., Chen, F., Heterotrophic production of eicosapentaenoic acid by microalgae. *Biotechnology advances* 2003, 21, 273-294.
- [181] Naduthodi, M. I. S., Barbosa, M. J., van der Oost, J., Progress of CRISPR-Cas based genome editing in Photosynthetic microbes. *Biotechnology journal* 2018, 1700591.
- [182] Verruto, J., Francis, K., Wang, Y., Low, M. C., *et al.*, Unrestrained markerless trait stacking in *Nannochloropsis gaditana* through combined genome editing and marker recycling technologies. *Proceedings of the National Academy of Sciences* 2018, 115, E7015-E7022.

- [183] Baek, K., Yu, J., Jeong, J., Sim, S. J., *et al.*, Photoautotrophic production of macular pigment in a *Chlamydomonas reinhardtii* strain generated by using DNA-free CRISPR-Cas9 RNP-mediated mutagenesis. *Biotechnology and bioengineering* 2018, 115, 719-728.
- [184] Shin, Y. S., Jeong, J., Nguyen, T. H. T., Kim, J. Y. H., *et al.*, Targeted knockout of phospholipase A2 to increase lipid productivity in *Chlamydomonas reinhardtii* for biodiesel production. *Bioresource Technology* 2018.
- [185] Serif, M., Dubois, G., Finoux, A.-L., Teste, M.-A., *et al.*, One-step generation of multiple gene knock-outs in the diatom *Phaeodactylum tricornutum* by DNA-free genome editing. *Nature communications* 2018, 9, 3924.
- [186] Chapman, J. R., Taylor, M. R., Boulton, S. J., Playing the end game: DNA double-strand break repair pathway choice. *Molecular cell* 2012, 47, 497-510.
- [187] Wyman, C., Kanaar, R., DNA Double-Strand Break Repair: All's Well that Ends Well. *Annual Review of Genetics* 2006, 40, 363-383.
- [188] Symington, L. S., Gautier, J., Double-strand break end resection and repair pathway choice. *Annual review of genetics* 2011, 45, 247-271.
- [189] Sander, J. D., Joung, J. K., CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* 2014, 32, 347.
- [190] Gao, X., Tao, Y., Lamas, V., Huang, M., *et al.*, Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* 2018, 553, 217.
- [191] Staahl, B. T., Benekareddy, M., Coulon-Bainier, C., Banfal, A. A., *et al.*, Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nature biotechnology* 2017, 35, 431.
- [192] Yin, H., Xue, W., Chen, S., Bogorad, R. L., *et al.*, Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nature biotechnology* 2014, 32, 551.
- [193] Jiang, W., Marraffini, L. A., CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. *Annual Review of Microbiology* 2015, 69, 209-228.
- [194] Li, J.-F., Norville, J. E., Aach, J., McCormack, M., *et al.*, Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nature biotechnology* 2013, 31, 688.
- [195] Svitashv, S., Young, J., Schwartz, C., Gao, H., *et al.*, Targeted mutagenesis, precise gene editing and site-specific gene insertion in maize using Cas9 and guide RNA. *Plant physiology* 2015, pp. 00793.02015.
- [196] Rouet, P., Smih, F., Jasin, M., Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences* 1994, 91, 6064-6068.
- [197] Sodeinde, O. A., Kindle, K. L., Homologous recombination in the nuclear genome of *Chlamydomonas reinhardtii*. *Proceedings of the National Academy of Sciences* 1993, 90, 9199-9203.
- [198] Nelson, J., Lefebvre, P. A., Targeted disruption of the NIT8 gene in *Chlamydomonas reinhardtii*. *Molecular and cellular biology* 1995, 15, 5762-5769.
- [199] Zorin, B., Lu, Y., Sizova, I., Hegemann, P., Nuclear gene targeting in *Chlamydomonas* as exemplified by disruption of the PHOT gene. *Gene* 2009, 432, 91-96.
- [200] Cheng, R.-B., Lin, X.-Z., Wang, Z.-K., Yang, S.-J., *et al.*, Establishment of a transgene expression system for the marine microalga *Schizochytrium* by 18S rDNA-targeted homologous recombination. *World Journal of Microbiology and Biotechnology* 2011, 27, 737-741.

- [201] Gaj, T., Gersbach, C. A., Barbas III, C. F., ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology* 2013, 31, 397-405.
- [202] Wu, W. Y., Lebbink, J. H., Kanaar, R., Geijsen, N., Van Der Oost, J., Genome editing by natural and engineered CRISPR-associated nucleases. *Nature chemical biology* 2018, 14, 642.
- [203] Bollag, R. J., Watdman, A., Liskay, R. M., Homologous recombination in mammalian cells. *Annual review of genetics* 1989, 23, 199-225.
- [204] Bibikova, M., Beumer, K., Trautman, J. K., Carroll, D., Enhancing gene targeting with designed zinc finger nucleases. *Science* 2003, 300, 764-764.
- [205] Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., Yang, S.-H., Off-target effects in CRISPR/Cas9-mediated genome engineering. *Molecular Therapy-Nucleic Acids* 2015, 4.
- [206] Tang, X., Lowder, L. G., Zhang, T., Malzahn, A. A., *et al.*, A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nature plants* 2017, 3, 17018.
- [207] Hu, X., Wang, C., Liu, Q., Fu, Y., Wang, K., Targeted mutagenesis in rice using CRISPR-Cpf1 system. *Journal of Genetics and Genomics* 2017, 44, 71-73.
- [208] Kim, H., Kim, S.-T., Ryu, J., Kang, B.-C., *et al.*, CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nature communications* 2017, 8, 14406.
- [209] Zhong, Z., Zhang, Y., You, Q., Tang, X., *et al.*, Plant genome editing using FnCpf1 and LbCpf1 nucleases at redefined and altered PAM sites. *Mol Plant* 2018.
- [210] Begemann, M. B., Gray, B. N., January, E., Gordon, G. C., *et al.*, Precise insertion and guided editing of higher plant genomes using Cpf1 CRISPR nucleases. *Scientific reports* 2017, 7, 11606.
- [211] Cui, L., Bikard, D., Consequences of Cas9 cleavage in the chromosome of Escherichia coli. *Nucleic acids research* 2016, 44, 4243-4251.
- [212] Cong, L., Ran, F. A., Cox, D., Lin, S., *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013, 1231143.
- [213] Wang, T., Wei, J. J., Sabatini, D. M., Lander, E. S., Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 2014, 343, 80-84.
- [214] Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., *et al.*, Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 2014, 343, 84-87.
- [215] Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., *et al.*, Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature biotechnology* 2014, 32, 1262.
- [216] Ledford, H., US regulation misses some GM crops. *Nature* 2013, 500, 389-390.
- [217] Waltz, E., Nature Publishing Group 2012.
- [218] Jones, H. D., Regulatory uncertainty over genome editing. *Nat. Plants* 2015, 1, 10.1038.
- [219] Kilian, O., Vick, B., Google Patents 2015.
- [220] Anders, C., Jinek, M., In vitro enzymology of Cas9, *Methods in enzymology*, Elsevier 2014, pp. 1-20.
- [221] Chisti, Y., Biodiesel from microalgae beats bioethanol. *Trends in biotechnology* 2008, 26, 126-131.
- [222] Edmundson, S. J., Wilkie, A. C., Landfill leachate—a water and nutrient resource for algae-based biofuels. *Environmental technology* 2013, 34, 1849-1857.
- [223] Farooq, W., Suh, W. I., Park, M. S., Yang, J.-W., Water use and its recycling in microalgae cultivation for biofuel application. *Bioresource technology* 2015, 184, 73-81.
- [224] Costa, J. A. V., De Morais, M. G., The role of biochemical engineering in the production of biofuels from microalgae. *Bioresource technology* 2011, 102, 2-9.

- [225] Smith, V. H., Sturm, B. S., Denoyelles, F. J., Billings, S. A., The ecology of algal biodiesel production. *Trends in ecology & evolution* 2010, 25, 301-309.
- [226] Sukenik, A., Beardall, J., Kromkamp, J. C., Kopecký, J., *et al.*, Photosynthetic performance of outdoor Nannochloropsis mass cultures under a wide range of environmental conditions. *Aquatic Microbial Ecology* 2009, 56, 297-308.
- [227] Naduthodi, M. I. S., Mohanraju, P., Südfeld, C., D'Adamo, S., *et al.*, CRISPR–Cas ribonucleoprotein mediated homology-directed repair for efficient targeted genome editing in microalgae Nannochloropsis oceanica IMET1. *Biotechnology for biofuels* 2019, 12, 1-11.
- [228] Wang, Q., Gong, Y., He, Y., Xin, Y., *et al.*, Genome engineering of Nannochloropsis with hundred-kilobase fragment deletions by Cas9 cleavages. *The Plant Journal* 2021.
- [229] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., *et al.*, A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science* 2012, 337, 816-821.
- [230] Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nature protocols* 2013, 8, 2281-2308.
- [231] Poliner, E., Clark, E., Cummings, C., Benning, C., Farre, E. M., A high-capacity gene stacking toolkit for the oleaginous microalga, Nannochloropsis oceanica CCMP1779. *Algal Research* 2020, 45, 101664.
- [232] Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., *et al.*, Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature biotechnology* 2004, 22, 1567-1572.
- [233] Naduthodi, M. I. S., Barbosa, M. J., van der Oost, J., Progress of CRISPR-Cas based genome editing in photosynthetic microbes. *Biotechnology Journal* 2018, 13, 1700591.
- [234] Zhang, Y.-T., Jiang, J.-Y., Shi, T.-Q., Sun, X.-M., *et al.*, Application of the CRISPR/Cas system for genome editing in microalgae. *Applied microbiology and biotechnology* 2019, 103, 3239-3248.
- [235] Stukenberg, D., Zauner, S., Dell'Aquila, G., Maier, U. G., Optimizing CRISPR/Cas9 for the diatom *Phaeodactylum tricornutum*. *Frontiers in plant science* 2018, 9, 740.
- [236] Jiang, W. Z., Weeks, D. P., A gene-within-a-gene Cas9/sgRNA hybrid construct enables gene editing and gene replacement strategies in *Chlamydomonas reinhardtii*. *Algal research* 2017, 26, 474-480.
- [237] Shin, S.-E., Lim, J.-M., Koh, H. G., Kim, E. K., *et al.*, CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*. *Scientific Reports* 2016, 6, 1-15.
- [238] Chang, K. S., Kim, J., Park, H., Hong, S.-J., *et al.*, Enhanced lipid productivity in AGP knockout marine microalga *Tetraselmis* sp. using a DNA-free CRISPR-Cas9 RNP method. *Bioresource technology* 2020, 303, 122932.
- [239] Claassens, N. J., Siliakus, M. F., Spaans, S. K., Creutzburg, S. C., *et al.*, Improving heterologous membrane protein production in *Escherichia coli* by combining transcriptional tuning and codon usage algorithms. *PLoS one* 2017, 12, e0184355.
- [240] Hall, M. P., Unch, J., Binkowski, B. F., Valley, M. P., *et al.*, Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS chemical biology* 2012, 7, 1848-1857.
- [241] Zhang, W.-W., Matlashewski, G., CRISPR-Cas9-mediated genome editing in *Leishmania donovani*. *MBio* 2015, 6.
- [242] Weninger, A., Hatzl, A.-M., Schmid, C., Vogl, T., Glieder, A., Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. *Journal of biotechnology* 2016, 235, 139-149.

- [243] Adiego-Pérez, B., Randazzo, P., Daran, J. M., Verwaal, R., *et al.*, Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS microbiology letters* 2019, 366, fnz086.
- [244] Liao, C., Ttofali, F., Slotkowski, R. A., Denny, S. R., *et al.*, Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. *Nature communications* 2019, 10, 1-14.
- [245] Yao, L., Cengic, I., Anfelt, J., Hudson, E. P., Multiple gene repression in cyanobacteria using CRISPRi. *ACS synthetic biology* 2016, 5, 207-212.
- [246] Huang, C.-H., Shen, C. R., Li, H., Sung, L.-Y., *et al.*, CRISPR interference (CRISPRi) for gene regulation and succinate production in cyanobacterium *S. elongatus* PCC 7942. *Microbial cell factories* 2016, 15, 1-11.
- [247] Kao, P.-H., Ng, I.-S., CRISPRi mediated phosphoenolpyruvate carboxylase regulation to enhance the production of lipid in *Chlamydomonas reinhardtii*. *Bioresource technology* 2017, 245, 1527-1537.
- [248] Zhang, X., Wang, J., Cheng, Q., Zheng, X., *et al.*, Multiplex gene regulation by CRISPR-ddCpf1. *Cell discovery* 2017, 3, 1-9.
- [249] Choi, S. Y., Woo, H. M., CRISPRi-dCas12a: A dCas12a-mediated CRISPR interference for repression of multiple genes and metabolic engineering in cyanobacteria. *ACS synthetic biology* 2020, 9, 2351-2361.
- [250] Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., *et al.*, Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 2014, 159, 647-661.
- [251] Thakore, P. I., Black, J. B., Hilton, I. B., Gersbach, C. A., Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nature methods* 2016, 13, 127-137.
- [252] Konermann, S., Brigham, M. D., Trevino, A. E., Hsu, P. D., *et al.*, Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 2013, 500, 472-476.
- [253] Yeo, N. C., Chavez, A., Lance-Byrne, A., Chan, Y., *et al.*, An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nature methods* 2018, 15, 611-616.
- [254] Naduthodi, M. I. S., Claassens, N. J., D'Adamo, S., van der Oost, J., Barbosa, M. J., Synthetic Biology Approaches To Enhance Microalgal Productivity. *Trends in Biotechnology* 2021.
- [255] Rees, H. A., Liu, D. R., Base editing: precision chemistry on the genome and transcriptome of living cells. *Nature reviews genetics* 2018, 19, 770-788.
- [256] Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., Liu, D. R., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 2016, 533, 420-424.
- [257] Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., *et al.*, Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 2017, 551, 464-471.
- [258] Gao, X., Tao, Y., Lamas, V., Huang, M., *et al.*, Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* 2018, 553, 217-221.
- [259] Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., *et al.*, Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 2019, 576, 149-157.
- [260] Liu, Y., Kao, H.-I., Bambara, R. A., Flap endonuclease 1: a central component of DNA metabolism. *Annual review of biochemistry* 2004, 73, 589-615.
- [261] Südfeld, C., Hubáček, M., D'Adamo, S., Wijffels, R. H., Barbosa, M. J., Optimization of high-throughput lipid screening of the microalga *Nannochloropsis oceanica* using BODIPY 505/515. *Algal Research* 2021, 53, 102138.
- [262] Liu, J., Song, Y., Qiu, W., Oleaginous microalgae *Nannochloropsis* as a new model for biofuel production: review & analysis. *Renewable and Sustainable Energy Reviews* 2017, 72, 154-162.

- [263] Zittelli, G. C., Lavista, F., Bastianini, A., Rodolfi, L., *et al.*, Production of eicosapentaenoic acid by *Nannochloropsis* sp. cultures in outdoor tubular photobioreactors, *Progress in Industrial Microbiology*, Elsevier 1999, pp. 299-312.
- [264] Archambault, S., Downes, C. M. S., Van Voorhies, W., Erickson, C. A., Lammers, P., *Nannochloropsis* sp. algae for use as biofuel: Analyzing a translog production function using data from multiple sites in the southwestern United States. *Algal Research* 2014, 6, 124-131.
- [265] San Pedro, A., González-López, C., Acién, F., Molina-Grima, E., Outdoor pilot production of *Nannochloropsis gaditana*: Influence of culture parameters and lipid production rates in raceway ponds. *Algal research* 2015, 8, 205-213.
- [266] Rumin, J., Nicolau, E., Gonçalves de Oliveira Junior, R., Fuentes-Grünwald, C., *et al.*, A bibliometric analysis of microalgae research in the world, Europe, and the European Atlantic area. *Marine drugs* 2020, 18, 79.
- [267] Blaby-Haas, C. E., Merchant, S. S., Comparative and functional algal genomics. *Annual review of plant biology* 2019, 70, 605-638.
- [268] Hanschen, E. R., Starkenburg, S. R., The state of algal genome quality and diversity. *Algal Research* 2020, 50, 101968.
- [269] Moresco, E. M. Y., Li, X., Beutler, B., Going forward with genetics: recent technological advances and forward genetics in mice. *The American journal of pathology* 2013, 182, 1462-1473.
- [270] Beutler, B., Jiang, Z., Georgel, P., Crozat, K., *et al.*, Genetic analysis of host resistance: Toll-like receptor signaling and immunity at large. *Annu. Rev. Immunol.* 2006, 24, 353-389.
- [271] Stadler, L. J., Genetic effects of X-rays in maize. *Proceedings of the National Academy of Sciences of the United States of America* 1928, 14, 69.
- [272] Auerbach, C., Robson, J., Production of mutations by allyl isothiocyanate. *Nature* 1944, 154, 81-81.
- [273] Shelake, R. M., Pramanik, D., Kim, J.-Y., Evolution of plant mutagenesis tools: a shifting paradigm from random to targeted genome editing. *Plant Biotechnology Reports* 2019, 13, 423-445.
- [274] Austin, C. P., Battey, J. F., Bradley, A., Bucan, M., *et al.*, The knockout mouse project. *Nature genetics* 2004, 36, 921.
- [275] Ram, H., Soni, P., Salvi, P., Gandass, N., *et al.*, Insertional mutagenesis approaches and their use in rice for functional genomics. *Plants* 2019, 8, 310.
- [276] Greco, R., Ouwerkerk, P. B., Sallaud, C., Kohli, A., *et al.*, Transposon insertional mutagenesis in rice. *Plant physiology* 2001, 125, 1175-1177.
- [277] Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., *et al.*, Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 2003, 301, 653-657.
- [278] Zhang, R., Patena, W., Armbruster, U., Gang, S. S., *et al.*, High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. *The Plant Cell* 2014, 26, 1398-1409.
- [279] Siebert, P. D., Chenchik, A., Kellogg, D. E., Lukyanov, K. A., Lukyanov, S. A., An improved PCR method for walking in uncloned genomic DNA. *Nucleic acids research* 1995, 23, 1087.
- [280] González-Ballester, D., De Montaigu, A., Galván, A., Fernández, E., Restriction enzyme site-directed amplification PCR: a tool to identify regions flanking a marker DNA. *Analytical biochemistry* 2005, 340, 330-335.
- [281] Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., *et al.*, Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell host & microbe* 2009, 6, 279-289.

- [282] Hyka, P., Lickova, S., Přibyl, P., Melzoch, K., Kovar, K., Flow cytometry for the development of biotechnological processes with microalgae. *Biotechnology advances* 2013, 31, 2-16.
- [283] Pereira, H., Schulze, P. S., Schüller, L. M., Santos, T., *et al.*, Fluorescence activated cell-sorting principles and applications in microalgal biotechnology. *Algal research* 2018, 30, 113-120.
- [284] Yao, L., Shen, H., Wang, N., Tatlay, J., *et al.*, Elevated acetyl-CoA by amino acid recycling fuels microalgal neutral lipid accumulation in exponential growth phase for biofuel production. *Plant biotechnology journal* 2017, 15, 497-509.
- [285] Cabanelas, I. T. D., Van Der Zwart, M., Kleinegris, D. M., Wijffels, R. H., Barbosa, M. J., Sorting cells of the microalga *Chlorococum littorale* with increased triacylglycerol productivity. *Biotechnology for biofuels* 2016, 9, 1-12.
- [286] Doan, T. T. Y., Obbard, J. P., Enhanced intracellular lipid in *Nannochloropsis* sp. via random mutagenesis and flow cytometric cell sorting. *Algal Research* 2012, 1, 17-21.
- [287] Wang, S., Zhang, L., Yang, G., Han, J., *et al.*, Breeding 3 elite strains of *Nannochloropsis oceanica* by nitrosoguanidine mutagenesis and robust screening. *Algal research* 2016, 19, 104-108.
- [288] Osorio, H., Jara, C., Fuenzalida, K., Rey-Jurado, E., Vásquez, M., High-efficiency nuclear transformation of the microalgae *Nannochloropsis oceanica* using Tn5 Transposome for the generation of altered lipid accumulation phenotypes. *Biotechnology for biofuels* 2019, 12, 1-12.
- [289] Ryu, A. J., Kang, N. K., Jeon, S., Hur, D. H., *et al.*, Development and characterization of a *Nannochloropsis* mutant with simultaneously enhanced growth and lipid production. *Biotechnology for biofuels* 2020, 13, 1-14.
- [290] Calder, P. C., n- 3 Polyunsaturated fatty acids, inflammation, and inflammatory diseases. *The American journal of clinical nutrition* 2006, 83, 1505S-1519S.
- [291] Mozaffarian, D., Wu, J. H., Omega-3 fatty acids and cardiovascular disease: effects on risk factors, molecular pathways, and clinical events. *Journal of the American College of Cardiology* 2011, 58, 2047-2067.
- [292] Dolch, L.-J., Rak, C., Perin, G., Tourcier, G., *et al.*, A palmitic acid elongase affects eicosapentaenoic acid and plastidial monogalactosyldiacylglycerol levels in *Nannochloropsis*. *Plant Physiology* 2017, 173, 742-759.
- [293] Janssen, J. H., Lamers, P. P., de Vos, R. C., Wijffels, R. H., Barbosa, M. J., Translocation and de novo synthesis of eicosapentaenoic acid (EPA) during nitrogen starvation in *Nannochloropsis gaditana*. *Algal Research* 2019, 37, 138-144.
- [294] Knothe, G., "Designer" biodiesel: optimizing fatty ester composition to improve fuel properties. *Energy & Fuels* 2008, 22, 1358-1364.
- [295] Lu, Y., Gan, Q., Iwai, M., Alboresi, A., *et al.*, Role of an ancient light-harvesting protein of PSI in light absorption and photoprotection. *Nature communications* 2021, 12, 1-10.
- [296] Arnold, C., Hodgson, I., Vectors PCR: a novel approach to genomic walking. *Genome Research* 1991, 1, 39-42.
- [297] Don, R., Cox, P., Wainwright, B., Baker, K., Mattick, J., 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic acids research* 1991, 19, 4008.
- [298] Zhang, Z., Gurr, S. J., Walking into the unknown: a 'step down' PCR-based technique leading to the direct sequence analysis of flanking genomic DNA. *Gene* 2000, 253, 145-150.
- [299] Hirano, T., Kinoshita, N., Morikawa, K., Yanagida, M., Snap helix with knob and hole: essential repeats in *S. pombe* nuclear protein nuc2+. *Cell* 1990, 60, 319-328.
- [300] Lamb, J. R., Tugendreich, S., Hieter, P., Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends in biochemical sciences* 1995, 20, 257-259.

- [301] Brocard, C., Hartig, A., Peroxisome targeting signal 1: is it really a simple tripeptide? *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 2006, 1763, 1565-1573.
- [302] Baker, M. J., Frazier, A. E., Gulbis, J. M., Ryan, M. T., Mitochondrial protein-import machinery: correlating structure with function. *Trends in cell biology* 2007, 17, 456-464.
- [303] Fransen, M., Amery, L., Hartig, A., Brees, C., et al., Comparison of the PTS1- and Rab8b-binding properties of Pex5p and Pex5Rp/TRIP8b. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 2008, 1783, 864-873.
- [304] Gatsos, X., Perry, A. J., Anwari, K., Dolezal, P., et al., Protein secretion and outer membrane assembly in Alphaproteobacteria. *FEMS microbiology reviews* 2008, 32, 995-1009.
- [305] Young, J. C., Barral, J. M., Hartl, F. U., More than folding: localized functions of cytosolic chaperones. *Trends in biochemical sciences* 2003, 28, 541-547.
- [306] Zeytuni, N., Ozyamak, E., Ben-Harush, K., Davidov, G., et al., Self-recognition mechanism of MamA, a magnetosome-associated TPR-containing protein, promotes complex assembly. *Proceedings of the National Academy of Sciences* 2011, 108, E480-E487.
- [307] Zeytuni, N., Zarivach, R., Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* 2012, 20, 397-405.
- [308] Cowell, I. G., Yeast two-hybrid library screening, *cDNA Library Protocols*, Springer 1997, pp. 185-202.
- [309] Erfelink, M.-L., Ribeiro, B., Perassolo, M., Pauwels, L., et al., A user-friendly platform for yeast two-hybrid library screening using next generation sequencing. *PLoS One* 2018, 13, e0201270.
- [310] Harewood, L., Schütz, F., Boyle, S., Perry, P., et al., The effect of translocation-induced nuclear reorganization on gene expression. *Genome research* 2010, 20, 554-564.
- [311] Shi, Y., Liu, M., Ding, W., Liu, J., Novel insights into phosphorus deprivation boosted lipid synthesis in the marine alga *nannochloropsis oceanica* without compromising biomass production. *Journal of Agricultural and Food Chemistry* 2020, 68, 11488-11502.
- [312] Wei, L., El Hajjami, M., Shen, C., You, W., et al., Transcriptomic and proteomic responses to very low CO₂ suggest multiple carbon concentrating mechanisms in *Nannochloropsis oceanica*. *Biotechnology for biofuels* 2019, 12, 1-21.
- [313] Bragantini, B., Tiotiu, D., Rothé, B., Saliou, J.-M., et al., Functional and structural insights of the zinc-finger HIT protein family members involved in box C/D snoRNP biogenesis. *Journal of molecular biology* 2016, 428, 2488-2506.
- [314] Cavalier-Smith, T., Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. *Protoplasma* 2018, 255, 297-357.
- [315] Derelle, R., López-García, P., Timpano, H., Moreira, D., A phylogenomic framework to study the diversity and evolution of stramenopiles (= heterokonts). *Molecular biology and evolution* 2016, 33, 2890-2898.
- [316] Youker, R. T., Brodsky, J. L., Regulation of Hsp70 function: hsp40 co-chaperones and nucleotide exchange factors, *Cell stress proteins*, Springer 2007, pp. 209-227.
- [317] Young, J. C., Mechanisms of the Hsp70 chaperone system. *Biochemistry and Cell Biology* 2010, 88, 291-300.
- [318] Ahmad, A., Bhattacharya, A., McDonald, R. A., Cordes, M., et al., Heat shock protein 70 kDa chaperone/DnaJ cochaperone complex employs an unusual dynamic interface. *Proceedings of the National Academy of Sciences* 2011, 108, 18966-18971.
- [319] Qiu, X.-B., Shao, Y.-M., Miao, S., Wang, L., The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cellular and Molecular Life Sciences CMLS* 2006, 63, 2560-2570.

- [320] Ghazaei, C., Role and mechanism of the Hsp70 molecular chaperone machines in bacterial pathogens. *Journal of Medical Microbiology* 2017, 66, 259-265.
- [321] Genevaux, P., Schwager, F., Georgopoulos, C., Kelley, W. L., Scanning mutagenesis identifies amino acid residues essential for the in vivo activity of the Escherichia coli DnaJ (Hsp40) J-domain. *Genetics* 2002, 162, 1045-1053.
- [322] Sahi, C., Craig, E. A., Network of general and specialty J protein chaperones of the yeast cytosol. *Proceedings of the National Academy of Sciences* 2007, 104, 7163-7168.
- [323] Cecchin, M., Berteotti, S., Paltrinieri, S., Vigliante, I., et al., Improved lipid productivity in Nannochloropsis gaditana in nitrogen-replete conditions by selection of pale green mutants. *Biotechnology for biofuels* 2020, 13, 1-14.
- [324] Mittler, R., Oxidative stress, antioxidants and stress tolerance. *Trends in plant science* 2002, 7, 405-410.
- [325] Xiong, Y., Contento, A. L., Nguyen, P. Q., Bassham, D. C., Degradation of oxidized proteins by autophagy during oxidative stress in Arabidopsis. *Plant physiology* 2007, 143, 291-299.
- [326] Fei, W., Wang, H., Fu, X., Bielby, C., Yang, H., Conditions of endoplasmic reticulum stress stimulate lipid droplet formation in Saccharomyces cerevisiae. *Biochemical Journal* 2009, 424, 61-67.
- [327] Basseri, S., Austin, R. C., Endoplasmic reticulum stress and lipid metabolism: mechanisms and therapeutic potential. *Biochemistry research international* 2012, 2012.
- [328] Shank, K. J., Su, P., Brglez, I., Boss, W. F., et al., Induction of lipid metabolic enzymes during the endoplasmic reticulum stress response in plants. *Plant physiology* 2001, 126, 267-277.
- [329] Yamaoka, Y., Shin, S., Choi, B. Y., Kim, H., et al., The bZIP1 transcription factor regulates lipid remodeling and contributes to ER stress management in Chlamydomonas reinhardtii. *The Plant Cell* 2019, 31, 1127-1140.
- [330] Gong, Y., Kang, N. K., Kim, Y. U., Wang, Z., et al., The NanDeSyn database for Nannochloropsis systems and synthetic biology. *The Plant Journal* 2020, 104, 1736-1745.
- [331] Vos, M. J., Hageman, J., Carra, S., Kampinga, H. H., Structural and functional diversities between members of the human HSPB, HSPH, HSPA, and DNAJ chaperone families. *Biochemistry* 2008, 47, 7001-7011.
- [332] Huang, Z., Teeling, E. C., ExUTR: a novel pipeline for large-scale prediction of 3' UTR sequences from NGS data. *BMC genomics* 2017, 18, 1-11.
- [333] Lu, R.-M., Hwang, Y.-C., Liu, I.-J., Lee, C.-C., et al., Development of therapeutic antibodies for the treatment of diseases. *Journal of biomedical science* 2020, 27, 1-30.
- [334] Hu, J., Wang, D., Li, J., Jing, G., et al., Genome-wide identification of transcription factors and transcription-factor binding sites in oleaginous microalgae Nannochloropsis. *Scientific reports* 2014, 4, 1-11.
- [335] Kong, Q., Yuan, L., Ma, W., WRINKLED1, a "Master Regulator" in transcriptional control of plant oil biosynthesis. *Plants* 2019, 8, 238.
- [336] Riechmann, J. L., Meyerowitz, E. M., The AP2/EREBP family of plant transcription factors. *Biological chemistry* 1998, 379, 633-646.
- [337] Ohto, M.-a., Fischer, R. L., Goldberg, R. B., Nakamura, K., Harada, J. J., Control of seed mass by APETALA2. *Proceedings of the National Academy of Sciences* 2005, 102, 3123-3128.
- [338] Xie, Z., Nolan, T. M., Jiang, H., Yin, Y., AP2/ERF transcription factor regulatory networks in hormone and abiotic stress responses in Arabidopsis. *Frontiers in plant science* 2019, 10, 228.
- [339] Seol, J.-H., Shim, E. Y., Lee, S. E., Microhomology-mediated end joining: Good, bad and ugly. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 2018, 809, 81-87.

- [340] Campenhout, C. V., Cabochette, P., Veillard, A.-C., Laczik, M., *et al.*, Guidelines for optimized gene knockout using CRISPR/Cas9. *BioTechniques* 2019, *66*, 295-302.
- [341] Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., *et al.*, Acyl-lipid metabolism. *The Arabidopsis book/American Society of Plant Biologists* 2013, *11*.
- [342] O'Hara, P., Slabas, A. R., Fawcett, T., Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis. *Plant Physiology* 2002, *129*, 310-320.
- [343] Majerus, P. W., Kilburn, E., Acetyl coenzyme A carboxylase: the roles of synthesis and degradation in regulation of enzyme levels in rat liver. *Journal of Biological Chemistry* 1969, *244*, 6254-6262.
- [344] Dong, H.-P., Williams, E., Wang, D.-z., Xie, Z.-X., *et al.*, Responses of *Nannochloropsis oceanica* IMET1 to long-term nitrogen starvation and recovery. *Plant physiology* 2013, *162*, 1110-1126.
- [345] Zienkiewicz, A., Zienkiewicz, K., Poliner, E., Pulman, J. A., *et al.*, The microalga *Nannochloropsis* during transition from quiescence to autotrophy in response to nitrogen availability. *Plant physiology* 2020, *182*, 819-839.
- [346] Janssen, J. H., Spoelder, J., Koehorst, J. J., Schaap, P. J., *et al.*, Time-dependent transcriptome profile of genes involved in triacylglycerol (TAG) and polyunsaturated fatty acid synthesis in *Nannochloropsis gaditana* during nitrogen starvation. *Journal of Applied Phycology* 2020, *32*, 1153-1164.
- [347] Joyard, J., Ferro, M., Masselon, C., Seigneurin-Berny, D., *et al.*, Chloroplast proteomics highlights the subcellular compartmentation of lipid metabolism. *Progress in lipid research* 2010, *49*, 128-158.
- [348] Koo, A. J., Ohlrogge, J. B., Pollard, M., On the export of fatty acids from the chloroplast. *Journal of Biological Chemistry* 2004, *279*, 16101-16110.
- [349] Knudsen, J., Mandrup, S., Rasmussen, J. T., Andreasen, P. H., *et al.*, The function of acyl-CoA-binding protein (ACBP)/diazepam binding inhibitor (DBI). *Molecular and cellular biochemistry* 1993, *123*, 129-138.
- [350] Shestov, A. A., Liu, X., Ser, Z., Cluntun, A. A., *et al.*, Quantitative determinants of aerobic glycolysis identify flux through the enzyme GAPDH as a limiting step. *elife* 2014, *3*, e03342.
- [351] Orlenko, A., Hermansen, R. A., Liberles, D. A., Flux control in glycolysis varies across the tree of life. *Journal of molecular evolution* 2016, *82*, 146-161.
- [352] Kanehisa, M., Goto, S., KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 2000, *28*, 27-30.
- [353] Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F. U., *et al.*, Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature* 2011, *479*, 194-199.
- [354] Yang, B., Liu, J., Ma, X., Guo, B., *et al.*, Genetic engineering of the Calvin cycle toward enhanced photosynthetic CO₂ fixation in microalgae. *Biotechnology for biofuels* 2017, *10*, 1-13.
- [355] Popelkova, H., Yocum, C. F., PsbO, the manganese-stabilizing protein: analysis of the structure-function relations that provide insights into its role in photosystem II. *Journal of Photochemistry and Photobiology B: Biology* 2011, *104*, 179-190.
- [356] Yi, X., McChargue, M., Laborde, S., Frankel, L. K., Bricker, T. M., The manganese-stabilizing protein is required for photosystem II assembly/stability and photoautotrophy in higher plants. *Journal of Biological Chemistry* 2005, *280*, 16170-16174.
- [357] Sharkey, T. D., Weise, S. E., The glucose 6-phosphate shunt around the Calvin-Benson cycle. *Journal of experimental botany* 2016, *67*, 4067-4077.

- [358] Li, D.-W., Balamurugan, S., Yang, Y.-F., Zheng, J.-W., *et al.*, Transcriptional regulation of microalgae for concurrent lipid overproduction and secretion. *Science advances* 2019, *5*, eaau3795.
- [359] Remmers, I. M., Hidalgo-Ulloa, A., Brandt, B., Evers, W. A., *et al.*, Continuous versus batch production of lipids in the microalgae *Acutodesmus obliquus*. *Bioresource technology* 2017, *244*, 1384-1392.
- [360] Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., *et al.*, CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic acids research* 2019, *47*, W171-W174.
- [361] Chen, S., Zhou, Y., Chen, Y., Gu, J., fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018, *34*, i884-i890.
- [362] Kim, D., Paggi, J. M., Park, C., Bennett, C., Salzberg, S. L., Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* 2019, *37*, 907-915.
- [363] Liao, Y., Smyth, G. K., Shi, W., featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014, *30*, 923-930.
- [364] Love, M. I., Huber, W., Anders, S., Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 2014, *15*, 1-21.
- [365] Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., *et al.*, Bioconductor: open software development for computational biology and bioinformatics. *Genome biology* 2004, *5*, 1-16.
- [366] Benjamini, Y., Hochberg, Y., Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 1995, *57*, 289-300.
- [367] Karp, P. D., Paley, S., Romero, P., The pathway tools software. *Bioinformatics* 2002, *18*, S225-S232.
- [368] Database resources of the national center for biotechnology information. *Nucleic acids research* 2018, *46*, D8-D13.
- [369] Marchler-Bauer, A., Bo, Y., Han, L., He, J., *et al.*, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic acids research* 2017, *45*, D200-D203.
- [370] Gschloessl, B., Guermeur, Y., Cock, J. M., HECTAR: a method to predict subcellular targeting in heterokonts. *BMC bioinformatics* 2008, *9*, 1-13.
- [371] Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., *et al.*, SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature biotechnology* 2019, *37*, 420-423.
- [372] Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., *et al.*, Detecting sequence signals in targeting peptides using deep learning. *Life science alliance* 2019, *2*.
- [373] Emanuelsson, O., Nielsen, H., Von Heijne, G., ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science* 1999, *8*, 978-984.
- [374] Fukasawa, Y., Tsuji, J., Fu, S.-C., Tomii, K., *et al.*, MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites*[S]. *Molecular & Cellular Proteomics* 2015, *14*, 1113-1126.
- [375] Lingner, T., Kataya, A. R., Antonicelli, G. E., Benichou, A., *et al.*, Identification of novel plant peroxisomal targeting signals by a combination of machine learning methods and in vivo subcellular targeting analyses. *The Plant Cell* 2011, *23*, 1556-1572.
- [376] Shrestha, P., Hussain, D., Mulder, R. J., Taylor, M. C., *et al.*, Increased DHA production in seed oil using a selective lysophosphatidic acid acyltransferase. *Frontiers in plant science* 2018, *9*, 1234.
- [377] Swanson, D., Block, R., Mousa, S. A., Omega-3 fatty acids EPA and DHA: health benefits throughout life. *Advances in nutrition* 2012, *3*, 1-7.

- [378] Tacon, A. G., Metian, M., Fishing for feed or fishing for food: increasing global competition for small pelagic forage fish. *Ambio* 2009, 294-302.
- [379] Mühlroth, A., Li, K., Røkke, G., Winge, P., *et al.*, 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* 2013, 11, 4662-4697.
- [380] Innis, S. M., Dietary triacylglycerol structure and its role in infant nutrition. *Advances in Nutrition* 2011, 2, 275-283.
- [381] Schuchardt, J. P., Hahn, A., Bioavailability of long-chain omega-3 fatty acids. *Prostaglandins, leukotrienes and essential fatty acids* 2013, 89, 1-8.
- [382] Yamashita, A., Hayashi, Y., Matsumoto, N., Nemoto-Sasaki, Y., *et al.*, Glycerophosphate/acylglycerophosphate acyltransferases. *Biology* 2014, 3, 801-830.
- [383] Bates, P. D., The significance of different diacylglycerol synthesis pathways on plant oil composition and bioengineering. *Frontiers in plant science* 2012, 3, 147.
- [384] Kennedy, E. P., Metabolism of lipides. *Annual review of biochemistry* 1957, 26, 119-148.
- [385] Gros, V., Jouhet, J., Quantitative assessment of the chloroplast lipidome, *Plastids*, Springer 2018, pp. 241-252.
- [386] Breuer, G., Evers, W. A., de Vree, J. H., Kleinegris, D. M., *et al.*, Analysis of fatty acid content and composition in microalgae. *Journal of visualized experiments: JoVE* 2013.
- [387] Wang, C., Li, Y., Lu, J., Deng, X., *et al.*, Effect of overexpression of LPAAT and GPD1 on lipid synthesis and composition in green microalga *Chlamydomonas reinhardtii*. *Journal of applied phycology* 2018, 30, 1711-1719.
- [388] Yamaoka, Y., Achard, D., Jang, S., Legéret, B., *et al.*, Identification of a *Chlamydomonas* plastidial 2-lysophosphatidic acid acyltransferase and its use to engineer microalgae with increased oil content. *Plant biotechnology journal* 2016, 14, 2158-2167.
- [389] LI, S., CHEN, Y., WANG, Y., YANG, X., *et al.*, Over-expression of LPAAT gene in *Phaeodactylum tricorutum* enhances fatty acid accumulation and increases fatty acid chain length. *Indian Journal of Pharmaceutical Sciences* 2020, 19-25.
- [390] Chungjatupornchai, W., Areerat, K., Fa-Aroonsawat, S., Increased triacylglycerol production in oleaginous microalga *Neochloris oleoabundans* by overexpression of plastidial lysophosphatidic acid acyltransferase. *Microbial cell factories* 2019, 18, 1-11.
- [391] Chien, L.-J., Hsu, T.-P., Huang, C.-C., Teng, K., Hsieh, H.-J., Novel codon-optimization genes encoded in *Chlorella* for triacylglycerol accumulation. *Energy Procedia* 2015, 75, 44-55.
- [392] Nobusawa, T., Hori, K., Mori, H., Kurokawa, K., Ohta, H., Differently localized lysophosphatidic acid acyltransferases crucial for triacylglycerol biosynthesis in the oleaginous alga *Nannochloropsis*. *The Plant Journal* 2017, 90, 547-559.
- [393] Murphy, D. J., The dynamic roles of intracellular lipid droplets: from archaea to mammals. *Protoplasma* 2012, 249, 541-585.
- [394] Horton, P., Park, K.-J., Obayashi, T., Fujita, N., *et al.*, WoLF PSORT: protein localization predictor. *Nucleic acids research* 2007, 35, W585-W587.
- [395] Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., Sternberg, M. J., The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* 2015, 10, 845-858.
- [396] Vieler, A., Brubaker, S. B., Vick, B., Benning, C., A lipid droplet protein of *Nannochloropsis* with functions partially analogous to plant oleosins. *Plant physiology* 2012, 158, 1562-1569.
- [397] Liang, J., Wen, F., Liu, J., Transcriptomic and lipidomic analysis of an EPA-containing *Nannochloropsis* sp. PJ12 in response to nitrogen deprivation. *Scientific reports* 2019, 9, 1-18.

- [398] Meng, Y., Cao, X., Yao, C., Xue, S., Yang, Q., Identification of the role of polar glycerolipids in lipid metabolism and their acyl attribution for TAG accumulation in *Nannochloropsis oceanica*. *Algal Research* 2017, 24, 122-129.
- [399] Li, X., Moellering, E. R., Liu, B., Johnny, C., *et al.*, A galactoglycerolipid lipase is required for triacylglycerol accumulation and survival following nitrogen deprivation in *Chlamydomonas reinhardtii*. *The Plant Cell* 2012, 24, 4670-4686.
- [400] Xin, Y., Shen, C., She, Y., Chen, H., *et al.*, Biosynthesis of triacylglycerol molecules with a tailored PUFA profile in industrial microalgae. *Molecular plant* 2019, 12, 474-488.
- [401] Beacham, T. A., Ali, S. T., Growth dependent silencing and resetting of DGA1 transgene in *Nannochloropsis salina*. *Algal research* 2016, 14, 65-71.
- [402] Huang, G., Chen, F., Wei, D., Zhang, X., Chen, G., Biodiesel production by microalgal biotechnology. *Applied energy* 2010, 87, 38-46.
- [403] Ruiz, J., Olivieri, G., de Vree, J., Bosma, R., *et al.*, Towards industrial products from microalgae. *Energy & Environmental Science* 2016, 9, 3036-3043.
- [404] Borowitzka, M. A., High-value products from microalgae—their development and commercialisation. *Journal of applied phycology* 2013, 25, 743-756.
- [405] Muñoz, C. F., Weusthuis, R. A., D'Adamo, S., Wijffels, R. H., Effect of Single and Combined Expression of Lysophosphatidic Acid Acyltransferase, Glycerol-3-Phosphate Acyltransferase, and Diacylglycerol Acyltransferase on Lipid Accumulation and Composition in *Neochloris oleoabundans*. *Frontiers in Plant Science* 2019, 10.
- [406] Batista-Silva, W., da Fonseca-Pereira, P., Martins, A. O., Zsögön, A., *et al.*, Engineering Improved Photosynthesis in the Era of Synthetic Biology. *Plant Communications* 2020, 1, 100032.
- [407] Weber, A. P., Bar-Even, A., Update: improving the efficiency of photosynthetic carbon reactions. *Plant physiology* 2019, 179, 803-812.
- [408] Vecchi, V., Barera, S., Bassi, R., Dall'Osto, L., Potential and Challenges of Improving Photosynthesis in Algae. *Plants* 2020, 9, 67.
- [409] Jin, E., Polle, J. E., Melis, A., Involvement of zeaxanthin and of the Cbr protein in the repair of photosystem II from photoinhibition in the green alga *Dunaliella salina*. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 2001, 1506, 244-259.
- [410] Shin, W.-S., Lee, B., Jeong, B.-r., Chang, Y. K., Kwon, J.-H., Truncated light-harvesting chlorophyll antenna size in *Chlorella vulgaris* improves biomass productivity. *Journal of Applied Phycology* 2016, 28, 3193-3202.
- [411] Ort, D. R., Zhu, X., Melis, A., Optimizing antenna size to maximize photosynthetic efficiency. *Plant physiology* 2011, 155, 79-85.
- [412] Blankenship, R. E., Tiede, D. M., Barber, J., Brudvig, G. W., *et al.*, Comparing photosynthetic and photovoltaic efficiencies and recognizing the potential for improvement. *science* 2011, 332, 805-809.
- [413] Claassens, N. J., Volpers, M., dos Santos, V. A. M., van der Oost, J., de Vos, W. M., Potential of proton-pumping rhodopsins: engineering photosystems into microorganisms. *Trends in biotechnology* 2013, 31, 633-642.
- [414] Ort, D. R., Merchant, S. S., Alric, J., Barkan, A., *et al.*, Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings of the national academy of sciences* 2015, 112, 8529-8536.
- [415] Vinyard, D. J., Gimpel, J., Ananyev, G. M., Mayfield, S. P., Dismukes, G. C., Engineered Photosystem II reaction centers optimize photochemistry versus photoprotection at different solar intensities. *Journal of the American Chemical Society* 2014, 136, 4048-4055.

- [416] Nixon, P. J., Rögner, M., Diner, B. A., Expression of a higher plant psbA gene in *Synechocystis* 6803 yields a functional hybrid photosystem II reaction center complex. *The Plant Cell* 1991, 3, 383-395.
- [417] Chiamonte, S., Giacometti, G. M., Bergantino, E., Construction and characterization of a functional mutant of *Synechocystis* 6803 harbouring a eukaryotic PSII-H subunit. *European journal of biochemistry* 1999, 260, 833-843.
- [418] Carpenter, S. D., Ohad, I., Vermaas, W. F., Analysis of chimeric spinach/cyanobacterial CP43 mutants of *Synechocystis* sp. PCC 6803: the chlorophyll-protein CP43 affects the water-splitting system of photosystem II. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1993, 1144, 204-212.
- [419] Vermaas, W. F., Shen, G., Ohad, I., Chimeric CP47 mutants of the cyanobacterium *Synechocystis* sp. PCC 6803 carrying spinach sequences: construction and function. *Photosynthesis research* 1996, 48, 147-162.
- [420] Viola, S., Rühle, T., Leister, D., A single vector-based strategy for marker-less gene replacement in *Synechocystis* sp. PCC 6803. *Microbial cell factories* 2014, 13, 1-12.
- [421] Gimpel, J. A., Nour-Eldin, H. H., Scranton, M. A., Li, D., Mayfield, S. P., Refactoring the six-gene photosystem II core in the chloroplast of the green algae *Chlamydomonas reinhardtii*. *ACS synthetic biology* 2016, 5, 589-596.
- [422] Chida, H., Nakazawa, A., Akazaki, H., Hirano, T., *et al.*, Expression of the algal cytochrome c 6 gene in *Arabidopsis* enhances photosynthesis and growth. *Plant and cell physiology* 2007, 48, 948-957.
- [423] Koh, H. G., Kang, N. K., Jeon, S., Shin, S.-E., *et al.*, Heterologous synthesis of chlorophyll b in *Nannochloropsis salina* enhances growth and lipid production by increasing photosynthetic efficiency. *Biotechnology for biofuels* 2019, 12, 122.
- [424] Koh, H. G., Ryu, A. J., Jeon, S., Jeong, K. J., *et al.*, Photosynthetic Improvement of Industrial Microalgae for Biomass and Biofuel Production, *Microbial Photosynthesis*, Springer 2020, pp. 285-317.
- [425] Ho, M.-Y., Shen, G., Canniffe, D. P., Zhao, C., Bryant, D. A., Light-dependent chlorophyll f synthase is a highly divergent paralog of PsbA of photosystem II. *Science* 2016, 353.
- [426] Trinugroho, J. P., Bečková, M., Shao, S., Yu, J., *et al.*, Chlorophyll f synthesis by a super-rogue photosystem II complex. *Nature Plants* 2020, 6, 238-244.
- [427] Post, A., Larkum, A., UV-absorbing pigments, photosynthesis and UV exposure in Antarctica: comparison of terrestrial and marine algae. *Aquatic Botany* 1993, 45, 231-243.
- [428] Hasan, S. S., Cramer, W. A., On rate limitations of electron transfer in the photosynthetic cytochrome b 6 f complex. *Physical Chemistry Chemical Physics* 2012, 14, 13853-13860.
- [429] Wei, L., Wang, Q., Xin, Y., Lu, Y., Xu, J., Enhancing photosynthetic biomass productivity of industrial oleaginous microalgae by overexpression of RuBisCO activase. *Algal research* 2017, 27, 366-375.
- [430] Liang, F., Lindblad, P., *Synechocystis* PCC 6803 overexpressing RuBisCO grow faster with increased photosynthesis. *Metabolic engineering communications* 2017, 4, 29-36.
- [431] Flamholz, A. I., Prywes, N., Moran, U., Davidi, D., *et al.*, Revisiting trade-offs between Rubisco kinetic parameters. *Biochemistry* 2019, 58, 3365-3376.
- [432] Badger, M. R., Bek, E. J., Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO₂ acquisition by the CBB cycle. *Journal of experimental botany* 2008, 59, 1525-1541.

- [433] Zhou, Y., Whitney, S., Directed evolution of an improved Rubisco; in vitro analyses to decipher fact from fiction. *International journal of molecular sciences* 2019, 20, 5019.
- [434] Wilson, R. H., Martin-Avila, E., Conlan, C., Whitney, S. M., An improved Escherichia coli screen for Rubisco identifies a protein–protein interface that can enhance CO₂-fixation kinetics. *Journal of Biological Chemistry* 2018, 293, 18-27.
- [435] Gomez-Fernandez, B. J., Garcia-Ruiz, E., Martin-Diaz, J., de Santos, P. G., *et al.*, Directed-in vitro-evolution of Precambrian and extant Rubiscos. *Scientific reports* 2018, 8, 1-11.
- [436] Satagopan, S., Tabita, F. R., Rubis CO selection using the vigorously aerobic and metabolically versatile bacterium *Ralstonia eutropha*. *The FEBS journal* 2016, 283, 2869-2880.
- [437] Satagopan, S., Huening, K. A., Tabita, F. R., Selection of Cyanobacterial (*Synechococcus* sp. Strain PCC 6301) RubisCO Variants with Improved Functional Properties That Confer Enhanced CO₂-Dependent Growth of *Rhodobacter capsulatus*, a Photosynthetic Bacterium. *MBio* 2019, 10, e01537-01519.
- [438] Atkinson, N., Leitão, N., Orr, D. J., Meyer, M. T., *et al.*, Rubisco small subunits from the unicellular green alga *Chlamydomonas* complement Rubisco-deficient mutants of *Arabidopsis*. *New Phytologist* 2017, 214, 655-667.
- [439] Fukayama, H., Koga, A., Hatanaka, T., Misoo, S., Small subunit of a cold-resistant plant, timothy, does not significantly alter the catalytic properties of Rubisco in transgenic rice. *Photosynthesis research* 2015, 124, 57-65.
- [440] Genkov, T., Meyer, M., Griffiths, H., Spreitzer, R. J., Functional hybrid rubisco enzymes with plant small subunits and algal large subunits engineered rbcS cDNA for expression in *Chlamydomonas*. *Journal of Biological Chemistry* 2010, 285, 19833-19841.
- [441] Meyer, M. T., Itakura, A. K., Patena, W., Wang, L., *et al.*, Assembly of the algal CO₂-fixing organelle, the pyrenoid, is guided by a Rubisco-binding motif. *Science advances* 2020, 6, eabd2408.
- [442] Jordan, D. B., Ogren, W. L., Species variation in kinetic properties of ribulose 1, 5-bisphosphate carboxylase/oxygenase. *Archives of Biochemistry and Biophysics* 1983, 227, 425-433.
- [443] Spreitzer, R. J., Peddi, S. R., Satagopan, S., Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco. *Proceedings of the National Academy of Sciences* 2005, 102, 17225-17230.
- [444] Lin, M. T., Hanson, M. R., Red algal Rubisco fails to accumulate in transplastomic tobacco expressing *Griffithsia monilis* RbcL and RbcS genes. *Plant direct* 2018, 2, e00045.
- [445] Aigner, H., Wilson, R., Bracher, A., Calisse, L., *et al.*, Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* 2017, 358, 1272-1278.
- [446] Zhu, G., Kurek, I., Liu, L., Engineering photosynthetic enzymes involved in CO₂-assimilation by gene shuffling, *The Chloroplast*, Springer 2010, pp. 307-322.
- [447] Varaljay, V. A., Satagopan, S., North, J. A., Witte, B., *et al.*, Functional metagenomic selection of ribulose 1, 5-bisphosphate carboxylase/oxygenase from uncultivated bacteria. *Environmental microbiology* 2016, 18, 1187-1199.
- [448] Davidi, D., Shamshoum, M., Guo, Z., Bar-On, Y. M., *et al.*, Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. *The EMBO Journal* 2020, e104081.
- [449] De Porcellinis, A. J., Nørgaard, H., Brey, L. M. F., Erstad, S. M., *et al.*, Overexpression of bifunctional fructose-1, 6-bisphosphatase/sedoheptulose-1, 7-bisphosphatase leads to enhanced photosynthesis and global reprogramming of carbon metabolism in *Synechococcus* sp. PCC 7002. *Metabolic engineering* 2018, 47, 170-183.

- [450] Janasch, M., Asplund-Samuelsson, J., Steuer, R., Hudson, E. P., Kinetic modeling of the Calvin cycle identifies flux control and stable metabolomes in *Synechocystis* carbon fixation. *Journal of experimental botany* 2019, 70, 973-983.
- [451] Liang, F., Lindblad, P., Effects of overexpressing photosynthetic carbon flux control enzymes in the cyanobacterium *Synechocystis* PCC 6803. *Metabolic engineering* 2016, 38, 56-64.
- [452] Ogawa, T., Tamoi, M., Kimura, A., Mine, A., *et al.*, Enhancement of photosynthetic capacity in *Euglena gracilis* by expression of cyanobacterial fructose-1, 6-/sedoheptulose-1, 7-bisphosphatase leads to increases in biomass and wax ester production. *Biotechnology for biofuels* 2015, 8, 80.
- [453] Dejtisakdi, W., Miller, S. M., Overexpression of Calvin cycle enzyme fructose 1,6-bisphosphatase in *Chlamydomonas reinhardtii* has a detrimental effect on growth. *Algal research* 2016, 14, 116-126.
- [454] Eisenhut, M., Ruth, W., Haimovich, M., Bauwe, H., *et al.*, The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiotically to plants. *Proceedings of the National Academy of Sciences* 2008, 105, 17199-17204.
- [455] Claassens, N. J., Scarinci, G., Fischer, A., Flamholz, A. I., *et al.*, Phosphoglycolate salvage in a chemolithoautotroph using the Calvin cycle. *Proceedings of the National Academy of Sciences* 2020, 117, 22452-22461.
- [456] Kebeish, R., Niessen, M., Thiruveedhi, K., Bari, R., *et al.*, Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*. *Nature biotechnology* 2007, 25, 593-599.
- [457] Maier, A., Fahnenstich, H., Von Caemmerer, S., Engqvist, M. K., *et al.*, Transgenic introduction of a glycolate oxidative cycle into *A. thaliana* chloroplasts leads to growth improvement. *Frontiers in Plant Science* 2012, 3, 38.
- [458] South, P. F., Cavanagh, A. P., Liu, H. W., Ort, D. R., Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science* 2019, 363.
- [459] Trudeau, D. L., Edlich-Muth, C., Zarzycki, J., Scheffen, M., *et al.*, Design and in vitro realization of carbon-conserving photorespiration. *Proceedings of the National Academy of Sciences* 2018, 115, E11455-E11464.
- [460] Scheffen, M., Marchal, D. G., Beneyton, T., Schuller, S. K., *et al.*, A new-to-nature carboxylation module to improve natural and synthetic CO₂ fixation. *Nature Catalysis* 2021, 4, 105-115.
- [461] Singh, S. K., Sundaram, S., Sinha, S., Rahman, M. A., Kapur, S., Recent advances in CO₂ uptake and fixation mechanism of cyanobacteria and microalgae. *Critical reviews in environmental science and technology* 2016, 46, 1297-1323.
- [462] Hennacy, J. H., Jonikas, M. C., Prospects for engineering biophysical CO₂ concentrating mechanisms into land plants to enhance yields. *Annual Review of Plant Biology* 2020, 71, 461-485.
- [463] Vikramathithan, J., Hwangbo, K., Lim, J.-M., Lim, K.-M., *et al.*, Overexpression of *Chlamydomonas reinhardtii* LCIA (CrLCIA) gene increases growth of *Nannochloropsis salina* CCMP1776. *Algal Research* 2020, 46, 101807.
- [464] Yamano, T., Sato, E., Iguchi, H., Fukuda, Y., Fukuzawa, H., Characterization of cooperative bicarbonate uptake into chloroplast stroma in the green alga *Chlamydomonas reinhardtii*. *Proceedings of the National Academy of Sciences* 2015, 112, 7315-7320.
- [465] Kim, S., Lindner, S. N., Aslan, S., Yishai, O., *et al.*, Growth of *E. coli* on formate and methanol via the reductive glycine pathway. *Nature chemical biology* 2020, 16, 538-545.

- [466] Claassens, N. J., Bordanaba-Florit, G., Cotton, C. A. R., De Maria, A., *et al.*, Replacing the Calvin cycle with the reductive glycine pathway in *Cupriavidus necator*. *Metabolic Engineering* 2020, 62, 30-41.
- [467] Gonzalez de la Cruz, J., Machens, F., Messerschmidt, K., Bar-Even, A., Core catalysis of the reductive glycine pathway demonstrated in yeast. *ACS synthetic biology* 2019, 8, 911-917.
- [468] Bar-Even, A., Noor, E., Lewis, N. E., Milo, R., Design and analysis of synthetic carbon fixation pathways. *Proceedings of the National Academy of Sciences* 2010, 107, 8889-8894.
- [469] d Mattozzi, M., Ziesack, M., Voges, M. J., Silver, P. A., Way, J. C., Expression of the sub-pathways of the *Chloroflexus aurantiacus* 3-hydroxypropionate carbon fixation bicycle in *E. coli*: Toward horizontal transfer of autotrophic growth. *Metabolic engineering* 2013, 16, 130-139.
- [470] Keller, M. W., Schut, G. J., Lipscomb, G. L., Menon, A. L., *et al.*, Exploiting microbial hyperthermophilicity to produce an industrial chemical, using hydrogen and carbon dioxide. *Proceedings of the National Academy of Sciences* 2013, 110, 5840-5845.
- [471] Gleizer, S., Ben-Nissan, R., Bar-On, Y. M., Antonovsky, N., *et al.*, Conversion of *Escherichia coli* to generate all biomass carbon from CO₂. *Cell* 2019, 179, 1255-1263. e1212.
- [472] Gassler, T., Sauer, M., Gasser, B., Egermeier, M., *et al.*, The industrial yeast *Pichia pastoris* is converted from a heterotroph into an autotroph capable of growth on CO₂. *Nature Biotechnology* 2020, 38, 210-216.
- [473] de Grahl, I., Rout, S. S., Maple-Grødem, J., Reumann, S., Development of a constitutive and an auto-inducible high-yield expression system for recombinant protein production in the microalga *Nannochloropsis oceanica*. *Applied Microbiology and Biotechnology* 2020, 1-14.
- [474] Fabris, M., George, J., Kuzhiumparambil, U., Lawson, C. A., *et al.*, Extrachromosomal genetic engineering of the marine diatom *Phaeodactylum tricornutum* enables the heterologous production of monoterpenoids. *ACS Synthetic Biology* 2020, 9, 598-612.
- [475] Li, D.-W., Xie, W.-H., Hao, T.-B., Cai, J.-X., *et al.*, Constitutive and chloroplast targeted expression of acetyl-CoA carboxylase in oleaginous microalgae elevates fatty acid biosynthesis. *Marine Biotechnology* 2018, 20, 566-572.
- [476] Gan, Q., Jiang, J., Han, X., Wang, S., Lu, Y., Engineering the chloroplast genome of oleaginous marine microalga *Nannochloropsis oceanica*. *Frontiers in plant science* 2018, 9, 439.
- [477] Gao, F., Yang, H.-L., Li, C., Peng, Y.-Y., *et al.*, Effect of organic carbon to nitrogen ratio in wastewater on growth, nutrient uptake and lipid accumulation of a mixotrophic microalgae *Chlorella* sp. *Bioresource technology* 2019, 282, 118-124.
- [478] Shen, X.-F., Hu, H., Ma, L.-L., Lam, P. K., *et al.*, FAMES production from *Scenedesmus obliquus* in autotrophic, heterotrophic and mixotrophic cultures under different nitrogen conditions. *Environmental Science: Water Research & Technology* 2018, 4, 461-468.
- [479] Cheirsilp, B., Torpee, S., Enhanced growth and lipid production of microalgae under mixotrophic culture condition: effect of light intensity, glucose concentration and fed-batch cultivation. *Bioresource technology* 2012, 110, 510-516.
- [480] Liang, Y., Sarkany, N., Cui, Y., Biomass and lipid productivities of *Chlorella vulgaris* under autotrophic, heterotrophic and mixotrophic growth conditions. *Biotechnology letters* 2009, 31, 1043-1049.
- [481] Abiusi, F., Wijffels, R. H., Janssen, M., Doubling of microalgae productivity by oxygen balanced mixotrophy. *ACS Sustainable Chemistry & Engineering* 2020, 8, 6065-6074.
- [482] Turon, V., Trably, E., Fouilland, E., Steyer, J.-P., Growth of *Chlorella sorokiniana* on a mixture of volatile fatty acids: The effects of light and temperature. *Bioresource technology* 2015, 198, 852-860.

- [483] Zaslavskaja, L., Lippmeier, J., Shih, C., Ehrhardt, D., *et al.*, Trophic conversion of an obligate photoautotrophic organism through metabolic engineering. *Science* 2001, 292, 2073-2075.
- [484] Doebbe, A., Rupprecht, J., Beckmann, J., Mussgnug, J. H., *et al.*, Functional integration of the HUP1 hexose symporter gene into the genome of *C. reinhardtii*: impacts on biological H₂ production. *Journal of Biotechnology* 2007, 131, 27-33.
- [485] Claassens, N. J., Sánchez-Andrea, I., Sousa, D. Z., Bar-Even, A., Towards sustainable feedstocks: A guide to electron donors for microbial carbon fixation. *Current opinion in biotechnology* 2018, 50, 195-205.
- [486] Harrington, T. D., Mohamed, A., Tran, V. N., Biria, S., *et al.*, Neutral red-mediated microbial electrosynthesis by *Escherichia coli*, *Klebsiella pneumoniae*, and *Zymomonas mobilis*. *Bioresource technology* 2015, 195, 57-65.
- [487] Tremblay, P.-L., Angenent, L. T., Zhang, T., Extracellular electron uptake: among autotrophs and mediated by surfaces. *Trends in biotechnology* 2017, 35, 360-371.
- [488] Nevin, K. P., Hensley, S. A., Franks, A. E., Summers, Z. M., *et al.*, Electrosynthesis of organic compounds from carbon dioxide is catalyzed by a diversity of acetogenic microorganisms. *Applied and environmental microbiology* 2011, 77, 2882-2886.
- [489] Shi, L., Dong, H., Reguera, G., Beyenal, H., *et al.*, Extracellular electron transfer mechanisms between microorganisms and minerals. *Nature Reviews Microbiology* 2016, 14, 651-662.
- [490] Claassens, N. J., Cotton, C. A., Kopljar, D., Bar-Even, A., Making quantitative sense of electromicrobial production. *Nature Catalysis* 2019, 2, 437-447.
- [491] Li, H., Opgenorth, P. H., Wernick, D. G., Rogers, S., *et al.*, Integrated electromicrobial conversion of CO₂ to higher alcohols. *Science* 2012, 335, 1596-1596.
- [492] Liu, C., Colón, B. C., Ziesack, M., Silver, P. A., Nocera, D. G., Water splitting–biosynthetic system with CO₂ reduction efficiencies exceeding photosynthesis. *Science* 2016, 352, 1210-1213.
- [493] Cuaresma, M., Janssen, M., Vilchez, C., Wijffels, R. H., Horizontal or vertical photobioreactors? How to improve microalgae photosynthetic efficiency. *Bioresource technology* 2011, 102, 5129-5137.
- [494] Gusbeth, C. A., Eing, C., Göttel, M., Frey, W., 2013 Abstracts IEEE international conference on plasma science (ICOPS), IEEE 2013, pp. 1-1.
- [495] Buchmann, L., Frey, W., Gusbeth, C., Ravaynia, P. S., Mathys, A., Effect of nanosecond pulsed electric field treatment on cell proliferation of microalgae. *Bioresource technology* 2019, 271, 402-408.
- [496] Buchmann, L., Böcker, L., Frey, W., Haberkorn, I., *et al.*, Energy input assessment for nanosecond pulsed electric field processing and its application in a case study with *Chlorella vulgaris*. *Innovative Food Science & Emerging Technologies* 2018, 47, 445-453.
- [497] HOWARTH, D. C., CODD, G. A., The uptake and production of molecular hydrogen by unicellular cyanobacteria. *Microbiology* 1985, 131, 1561-1569.
- [498] Bothe, H., Tennigkeit, J., Eisbrenner, G., The utilization of molecular hydrogen by the blue-green alga *Anabaena cylindrica*. *Archives of Microbiology* 1977, 114, 43-49.
- [499] Tamagnini, P., Axelsson, R., Lindberg, P., Oxelfelt, F., *et al.*, Hydrogenases and hydrogen metabolism of cyanobacteria. *Microbiology and molecular biology reviews : MMBR* 2002, 66, 1-20.
- [500] Krishnan, A., Cano, M., Burch, T. A., Weissman, J. C., Posewitz, M. C., Genome editing using Cas9-RNA ribonucleoprotein complexes in the high-productivity marine alga *Picochlorum celeri*. *Algal Research* 2020, 49, 101944.

- [501] Weissman, J. C., Likhogrud, M., Thomas, D. C., Fang, W., *et al.*, High-light selection produces a fast-growing *Picochlorum celeri*. *Algal Research* 2018, *36*, 17-28.
- [502] Calvin, M., Benson, A. A., The path of carbon in photosynthesis. 2008.
- [503] STRAUSS, G., FUCHS, G., Enzymes of a novel autotrophic CO₂ fixation pathway in the phototrophic bacterium *Chloroflexus aurantiacus*, the 3-hydroxypropionate cycle. *European Journal of Biochemistry* 1993, *215*, 633-643.
- [504] Ragsdale, S. W., The Eastern and Western branches of the Wood/Ljungdahl pathway: how the East and West were won. *Biofactors* 1997, *6*, 3-11.
- [505] Evans, M., Buchanan, B. B., Arnon, D. I., A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proceedings of the National Academy of Sciences of the United States of America* 1966, *55*, 928.
- [506] Huber, H., Gallenberger, M., Jahn, U., Eylert, E., *et al.*, A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum *Ignicoccus hospitalis*. *Proceedings of the National Academy of Sciences* 2008, *105*, 7851-7856.
- [507] Berg, I. A., Kockelkorn, D., Buckel, W., Fuchs, G., A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* 2007, *318*, 1782-1786.
- [508] Schwander, T., von Borzyskowski, L. S., Burgener, S., Cortina, N. S., Erb, T. J., A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* 2016, *354*, 900-904.
- [509] Miller, T. E., Beneyton, T., Schwander, T., Diehl, C., *et al.*, Light-powered CO₂ fixation in a chloroplast mimic with natural and synthetic parts. *Science* 2020, *368*, 649-654.
- [510] Kim, Y.-G., Cha, J., Chandrasegaran, S., Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences* 1996, *93*, 1156-1160.
- [511] Mak, A. N.-S., Bradley, P., Cernadas, R. A., Bogdanove, A. J., Stoddard, B. L., The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 2012, *335*, 716-719.
- [512] Serif, M., Lepetit, B., Weißert, K., Kroth, P. G., Bartulos, C. R., A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom *Phaeodactylum tricorutum*. *Algal research* 2017, *23*, 186-195.
- [513] Gao, H., Wright, D. A., Li, T., Wang, Y., *et al.*, TALE activation of endogenous genes in *Chlamydomonas reinhardtii*. *Algal Research* 2014, *5*, 52-60.
- [514] Barrangou, R., Fremaux, C., Deveau, H., Richards, M., *et al.*, CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 2007, *315*, 1709-1712.
- [515] Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., *et al.*, Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 2008, *321*, 960-964.
- [516] Barber, J., A mechanism for water splitting and oxygen production in photosynthesis. *Nature Plants* 2017, *3*, 1-5.
- [517] Sproviero, E. M., Gascón, J. A., McEvoy, J. P., Brudvig, G. W., Batista, V. S., Quantum mechanics/molecular mechanics study of the catalytic cycle of water splitting in photosystem II. *Journal of the American Chemical Society* 2008, *130*, 3428-3442.
- [518] Joliot, P., Joliot, A., Cyclic electron flow in C₃ plants. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 2006, *1757*, 362-368.
- [519] Hase, T., Schürmann, P., Knaff, D. B., The interaction of ferredoxin with ferredoxin-dependent enzymes, *Photosystem I*, Springer 2006, pp. 477-498.
- [520] Külheim, C., Ågren, J., Jansson, S., Rapid regulation of light harvesting and plant fitness in the field. *Science* 2002, *297*, 91-93.

- [521] Niyogi, K. K., Bjorkman, O., Grossman, A. R., Chlamydomonas xanthophyll cycle mutants identified by video imaging of chlorophyll fluorescence quenching. *The Plant Cell* 1997, 9, 1369-1380.
- [522] Müller, P., Li, X.-P., Niyogi, K. K., Non-photochemical quenching. A response to excess light energy. *Plant physiology* 2001, 125, 1558-1566.
- [523] Berg, I. A., Ecological aspects of the distribution of different autotrophic CO₂ fixation pathways. *Applied and environmental microbiology* 2011, 77, 1925-1936.
- [524] Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E., Scott, S. S., Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *Journal of experimental botany* 2008, 59, 1515-1524.
- [525] Cotton, C. A., Edlich-Muth, C., Bar-Even, A., Reinforcing carbon fixation: CO₂ reduction replacing and supporting carboxylation. *Current opinion in biotechnology* 2018, 49, 49-56.
- [526] Kern, R., Bauwe, H., Hagemann, M., Evolution of enzymes involved in the photorespiratory 2-phosphoglycolate cycle from cyanobacteria via algae toward plants. *Photosynthesis Research* 2011, 109, 103-114.
- [527] Bauwe, H., Hagemann, M., Fernie, A. R., Photorespiration: players, partners and origin. *Trends in plant science* 2010, 15, 330-336.
- [528] Toro-Huertas, E. I., Franco-Morgado, M., de Los Cobos Vasconcelos, D., González-Sánchez, A., Photorespiration in an outdoor alkaline open-photobioreactor used for biogas upgrading. *Science of The Total Environment* 2019, 667, 613-621.
- [529] Gong, F., Zhu, H., Zhang, Y., Li, Y., Biological carbon fixation: From natural to synthetic. *Journal of CO₂ Utilization* 2018, 28, 221-227.
- [530] Schulman, M., Parker, D., Ljungdahl, L. G., Wood, H. G., Total synthesis of acetate from CO₂. V. Determination by mass analysis of the different types of acetate formed from ¹³CO₂ by heterotrophic bacteria. *Journal of Bacteriology* 1972, 109, 633-644.
- [531] Kono, T., Mehrotra, S., Endo, C., Kizu, N., *et al.*, A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. *Nature communications* 2017, 8, 1-12.
- [532] Sánchez-Andrea, I., G., I.A., H., B., B., *et al.*, The reductive glycine pathway allows autotrophic growth of *Desulfovibrio desulfuricans*. *nature communications* 2020.
- [533] Yu, H., Li, X., Duchoud, F., Chuang, D. S., Liao, J. C., Augmenting the Calvin–Benson–Bassham cycle by a synthetic malyl-CoA-glycerate carbon fixation pathway. *Nature communications* 2018, 9, 1-10.
- [534] Mainguet, S. E., Gronenberg, L. S., Wong, S. S., Liao, J. C., A reverse glyoxylate shunt to build a non-native route from C₄ to C₂ in *Escherichia coli*. *Metabolic engineering* 2013, 19, 116-127.
- [535] Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., Winge, P., A CRISPR/Cas9 system adapted for gene editing in marine algae. *Scientific reports* 2016, 6, 24951.
- [536] Nguyen, T. H. T., Park, S., Jeong, J., Shin, Y. S., *et al.*, Enhancing lipid productivity by modulating lipid catabolism using the CRISPR-Cas9 system in *Chlamydomonas*. *Journal of Applied Phycology* 2020, 1-12.
- [537] Lin, W.-R., Ng, I.-S., Development of CRISPR/Cas9 system in *Chlorella vulgaris* FSP-E to enhance lipid accumulation. *Enzyme and microbial technology* 2020, 133, 109458.
- [538] Tetali, S. D., Mitra, M., Melis, A., Development of the light-harvesting chlorophyll antenna in the green alga *Chlamydomonas reinhardtii* is regulated by the novel Tla1 gene. *Planta* 2007, 225, 813-829.

- [539] Kirst, H., García-Cerdán, J. G., Zurbriggen, A., Melis, A., Assembly of the light-harvesting chlorophyll antenna in the green alga *Chlamydomonas reinhardtii* requires expression of the TLA2-CpFTSY gene. *Plant Physiology* 2012, *158*, 930-945.
- [540] Kirst, H., Garcia-Cerdan, J. G., Zurbriggen, A., Ruehle, T., Melis, A., Truncated photosystem chlorophyll antenna size in the green microalga *Chlamydomonas reinhardtii* upon deletion of the TLA3-CpSRP43 gene. *Plant physiology* 2012, *160*, 2251-2260.
- [541] Jeong, J., Baek, K., Kirst, H., Melis, A., Jin, E., Loss of CpSRP54 function leads to a truncated light-harvesting antenna size in *Chlamydomonas reinhardtii*. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 2017, *1858*, 45-55.
- [542] Beckmann, J., Lehr, F., Finazzi, G., Hankamer, B., *et al.*, Improvement of light to biomass conversion by de-regulation of light-harvesting protein translation in *Chlamydomonas reinhardtii*. *Journal of biotechnology* 2009, *142*, 70-77.
- [543] Nakajima, Y., Tsuzuki, M., Ueda, R., Improved productivity by reduction of the content of light-harvesting pigment in *Chlamydomonas perigranulata*. *Journal of applied phycology* 2001, *13*, 95-101.
- [544] Shin, W.-S., Lee, B., Kang, N. K., Kim, Y.-U., *et al.*, Complementation of a mutation in CpSRP43 causing partial truncation of light-harvesting chlorophyll antenna in *Chlorella vulgaris*. *Scientific reports* 2017, *7*, 1-11.
- [545] Cazzaniga, S., Dall'Osto, L., Szaub, J., Scibilia, L., *et al.*, Domestication of the green alga *Chlorella sorokiniana*: reduction of antenna size improves light-use efficiency in a photobioreactor. *Biotechnology for biofuels* 2014, *7*, 157.
- [546] Nuñez, J. K., Chen, J., Pommier, G. C., Cogan, J. Z., *et al.*, Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* 2021, *184*, 2503-2519. e2517.
- [547] Jackson, D. A., Symons, R. H., Berg, P., Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proceedings of the National Academy of Sciences* 1972, *69*, 2904-2909.
- [548] Cohen, S. N., Chang, A. C., Boyer, H. W., Helling, R. B., Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences* 1973, *70*, 3240-3244.
- [549] Chang, A. C., Cohen, S. N., Genome construction between bacterial species in vitro: replication and expression of *Staphylococcus* plasmid genes in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 1974, *71*, 1030-1034.
- [550] Jaenisch, R., Mintz, B., Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. *Proceedings of the national academy of sciences* 1974, *71*, 1250-1254.
- [551] Frank, B., Chance, R., Two routes for producing human insulin utilizing recombinant DNA technology. *MMW, Munchener medizinische Wochenschrift* 1983, S14-20.
- [552] Klümper, W., Qaim, M., A meta-analysis of the impacts of genetically modified crops. *PloS one* 2014, *9*, e111629.
- [553] Boynton, J. E., Gillham, N. W., Harris, E. H., Hosler, J. P., *et al.*, Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. *Science* 1988, *240*, 1534-1538.
- [554] Kindle, K. L., High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. *Proceedings of the National Academy of Sciences* 1990, *87*, 1228-1232.

- [555] Jeon, K., Suresh, A., Kim, Y.-C., Highly efficient molecular delivery into *Chlamydomonas reinhardtii* by electroporation. *Korean Journal of Chemical Engineering* 2013, 30, 1626-1630.
- [556] Kumar, S. V., Misquitta, R. W., Reddy, V. S., Rao, B. J., Rajam, M. V., Genetic transformation of the green alga—*Chlamydomonas reinhardtii* by *Agrobacterium tumefaciens*. *Plant Science* 2004, 166, 731-738.
- [557] Wada, N., Ueta, R., Osakabe, Y., Osakabe, K., Precision genome editing in plants: state-of-the-art in CRISPR/Cas9-based genome engineering. *BMC Plant Biology* 2020, 20, 1-12.
- [558] Hendel, A., Bak, R. O., Clark, J. T., Kennedy, A. B., *et al.*, Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nature biotechnology* 2015, 33, 985-989.
- [559] Gomma, A. E., Lee, S.-K., Sun, S. M., Yang, S. H., Chung, G., Improvement in oil production by increasing malonyl-CoA and glycerol-3-phosphate pools in *Scenedesmus quadricauda*. *Indian journal of microbiology* 2015, 55, 447-455.
- [560] Yan, J., Cheng, R., Lin, X., You, S., *et al.*, Overexpression of acetyl-CoA synthetase increased the biomass and fatty acid proportion in microalga *Schizochytrium*. *Applied microbiology and biotechnology* 2013, 97, 1933-1939.
- [561] Rengel, R., Smith, R. T., Haslam, R. P., Sayanova, O., *et al.*, Overexpression of acetyl-CoA synthetase (ACS) enhances the biosynthesis of neutral lipids and starch in the green microalga *Chlamydomonas reinhardtii*. *Algal Research* 2018, 31, 183-193.
- [562] Chen, J. W., Liu, W. J., Hu, D. X., Wang, X., *et al.*, Identification of a malonyl CoA-acyl carrier protein transacylase and its regulatory role in fatty acid biosynthesis in oleaginous microalga *Nannochloropsis oceanica*. *Biotechnology and applied biochemistry* 2017, 64, 620-626.
- [563] Li, D.-W., Cen, S.-Y., Liu, Y.-H., Balamurugan, S., *et al.*, A type 2 diacylglycerol acyltransferase accelerates the triacylglycerol biosynthesis in heterokont oleaginous microalga *Nannochloropsis oceanica*. *Journal of biotechnology* 2016, 229, 65-71.
- [564] Houten, S. M., Wanders, R. J., A general introduction to the biochemistry of mitochondrial fatty acid β -oxidation. *Journal of inherited metabolic disease* 2010, 33, 469-477.
- [565] Kong, F., Romero, I. T., Warakanont, J., Li-Beisson, Y., Lipid catabolism in microalgae. *New Phytologist* 2018, 218, 1340-1348.
- [566] Massart, J., Begriche, K., Buron, N., Porceddu, M., *et al.*, Drug-induced inhibition of mitochondrial fatty acid oxidation and steatosis. *Current Pathobiology Reports* 2013, 1, 147-157.
- [567] Runguphan, W., Keasling, J. D., Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metabolic engineering* 2014, 21, 103-113.
- [568] Niu, T.-C., Lin, G.-M., Xie, L.-R., Wang, Z.-Q., *et al.*, Expanding the potential of CRISPR-Cpf1-based genome editing technology in the cyanobacterium *Anabaena* PCC 7120. *ACS synthetic biology* 2018, 8, 170-180.
- [569] Swarts, D. C., van der Oost, J., Jinek, M., Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-Cas12a. *Molecular cell* 2017, 66, 221-233. e224.
- [570] Gao, L., Cox, D. B., Yan, W. X., Manteiga, J. C., *et al.*, Engineered Cpf1 variants with altered PAM specificities. *Nature biotechnology* 2017, 35, 789-792.
- [571] Kim, D., Kim, J., Hur, J. K., Been, K. W., *et al.*, Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nature biotechnology* 2016, 34, 863-868.
- [572] Zhang, J., Chen, L., Zhang, J., Wang, Y., Drug inducible CRISPR/Cas systems. *Computational and structural biotechnology journal* 2019, 17, 1171-1177.

- [573] Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., *et al.*, Creation of a bacterial cell controlled by a chemically synthesized genome. *science* 2010, 329, 52-56.
- [574] Wessler, S. R., Transposable elements and the evolution of eukaryotic genomes. *Proceedings of the National Academy of Sciences* 2006, 103, 17600-17601.
- [575] Park, M. K., Lee, S. H., Yang, K. S., Jung, S.-C., *et al.*, Enhancing recombinant protein production with an Escherichia coli host strain lacking insertion sequences. *Applied microbiology and biotechnology* 2014, 98, 6701-6713.
- [576] Koonin, E. V., How many genes can make a cell: the minimal-gene-set concept. *Annual review of genomics and human genetics* 2000, 1, 99-116.
- [577] Choe, D., Cho, S., Kim, S. C., Cho, B. K., Minimal genome: worthwhile or worthless efforts toward being smaller? *Biotechnology journal* 2016, 11, 199-211.
- [578] Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., *et al.*, Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 2016, 353.
- [579] Xiao, Y., Luo, M., Dolan, A. E., Liao, M., Ke, A., Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science* 2018, 361.
- [580] Gong, B., Shin, M., Sun, J., Jung, C.-H., *et al.*, Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proceedings of the National Academy of Sciences* 2014, 111, 16359-16364.
- [581] Morisaka, H., Yoshimi, K., Okuzaki, Y., Gee, P., *et al.*, CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. *Nature communications* 2019, 10, 1-13.
- [582] Arora, N., Yen, H.-W., Philippidis, G. P., Harnessing the power of mutagenesis and adaptive laboratory evolution for high lipid production by oleaginous microalgae and yeasts. *Sustainability* 2020, 12, 5125.
- [583] Dent, R. M., Sharifi, M. N., Malnoë, A., Haglund, C., *et al.*, Large-scale insertional mutagenesis of Chlamydomonas supports phylogenomic functional prediction of photosynthetic genes and analysis of classical acetate-requiring mutants. *The Plant Journal* 2015, 82, 337-351.
- [584] Ong, S.-C., Kao, C.-Y., Chiu, S.-Y., Tsai, M.-T., Lin, C.-S., Characterization of the thermal-tolerant mutants of Chlorella sp. with high growth rate and application in outdoor photobioreactor cultivation. *Bioresource technology* 2010, 101, 2880-2883.
- [585] He, M., Song, H., Chen, W., Zhang, Y., *et al.*, Comparative transcriptome analysis of wild type and an oleaginous mutant strain of Desmodesmus sp. reveals a unique reprogramming of lipid metabolism under high light. *Journal of Applied Phycology* 2019, 31, 2895-2910.
- [586] Arora, N., Philippidis, G. P., Microalgae strain improvement strategies: random mutagenesis and adaptive laboratory evolution. *Trends in Plant Science* 2021.
- [587] Waltz, E., With a free pass, CRISPR-edited plants reach market in record time. *Nature biotechnology* 2018, 36, 6-8.
- [588] Li, C., Li, W., Zhou, Z., Chen, H., *et al.*, A new rice breeding method: CRISPR/Cas9 system editing of the Xa13 promoter to cultivate transgene-free bacterial blight-resistant rice. *Plant biotechnology journal* 2020, 18, 313.
- [589] Liang, Z., Chen, K., Li, T., Zhang, Y., *et al.*, 2017.
- [590] Voigt, B., Münichsdorfer, A., Regulation of genome editing in plant biotechnology: European Union, *Regulation of Genome Editing in Plant Biotechnology*, Springer 2019, pp. 137-238.
- [591] Medvedieva, M., Blume, Y. B., Legal regulation of plant genome editing with the CRISPR/Cas9 technology as an example. *Cytology and Genetics* 2018, 52, 204-212.

- [592] Zhang, D., Hussain, A., Manghwar, H., Xie, K., *et al.*, Genome editing with the CRISPR-Cas system: an art, ethics and global regulatory perspective. *Plant biotechnology journal* 2020, 18, 1651-1669.
- [593] European Union, Official website of European Union 2018, p. 8.
- [594] Poliner, E., Farré, E. M., Benning, C., Advanced genetic tools enable synthetic biology in the oleaginous microalgae *Nannochloropsis* sp. *Plant cell reports* 2018, 37, 1383-1399.
- [595] Zienkiewicz, K., Zienkiewicz, A., Poliner, E., Du, Z.-Y., *et al.*, *Nannochloropsis*, a rich source of diacylglycerol acyltransferases for engineering of triacylglycerol content in different hosts. *Biotechnology for biofuels* 2017, 10, 1-20.
- [596] Kwon, S., Kang, N. K., Koh, H. G., Shin, S. E., *et al.*, Enhancement of biomass and lipid productivity by overexpression of a bZIP transcription factor in *Nannochloropsis salina*. *Biotechnology and bioengineering* 2018, 115, 331-340.
- [597] Südfeld, C., Hubáček, M., Figueiredo, D., Naduthodi, M. I., *et al.*, High-throughput insertional mutagenesis reveals novel targets for enhancing lipid accumulation in *Nannochloropsis oceanica*. *Metabolic Engineering* 2021, 66, 239-258.
- [598] Barrett, J., Girr, P., Mackinder, L. C., Pyrenoids: CO₂-fixing phase separated liquid organelles. *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research* 2021, 118949.
- [599] Long, S. P., Burgess, S., Causton, I., Redesigning crop photosynthesis. *Sustaining Global Food Security: The Nexus of Science and Policy* 2019, 128.
- [600] Atkinson, N., Mao, Y., Chan, K. X., McCormick, A. J., Condensation of Rubisco into a proto-pyrenoid in higher plant chloroplasts. *Nature communications* 2020, 11, 1-9.
- [601] Long, B. M., Hee, W. Y., Sharwood, R. E., Rae, B. D., *et al.*, Carboxysome encapsulation of the CO₂-fixing enzyme Rubisco in tobacco chloroplasts. *Nature communications* 2018, 9, 1-14.
- [602] Aslam, A., Thomas-Hall, S. R., Manzoor, M., Jabeen, F., *et al.*, Mixed microalgae consortia growth under higher concentration of CO₂ from unfiltered coal fired flue gas: fatty acid profiling and biodiesel production. *Journal of Photochemistry and Photobiology B: Biology* 2018, 179, 126-133.

Thesis Summary

Global warming has been a major threat to sustaining life on earth, and people all over the world have been looking for strategies to address this challenge. Uncontrolled burning of fossil fuels is releasing tons of greenhouse gases into the atmosphere, and this has been a major cause of the global warming phenomenon. The idea of biofuels was established as a sustainable alternative to fossil fuels. The first-generation biofuels produced from food and feed crops (e.g. starch-containing corn cobs), reduced the amount of greenhouse gas released into the atmosphere, but implied competition with food production crops. For obtaining second-generation biofuels, microbes are exploited that are able to convert inedible biomass into fuels and green chemicals. Even though this addressed the drawback of first-generation biofuels, large area of land will be required to produce the biomass to feed these microbes, that could otherwise be used for food/feed production. The third-generation biofuels are based on processes that utilize photosynthetic microbes such as microalgae to produce biofuels from sunlight and atmospheric carbon dioxide as the carbon source. This third-generation technology requires comparatively small land area and does not compete with food/feed production since they can be cultivated on barren lands with photobioreactors. For this reason, this technology is considered more efficient and sustainable than the first- and second-generation biofuel production technologies. However, the efficacy of this technology relies on the strains used and cultivation techniques to maximize productivity. **Chapter 1** describes the transformation of microalgae from being a staple food in some parts of the world into platform strains for producing valuable compounds. Among various nutraceuticals and chemicals derived from by microalgae, this thesis focuses on the production of lipids as the precursor for biofuel production, and of polyunsaturated fatty acid (EPA) with well-established health benefits, by microalgae. In this regard, *Nannochloropsis* spp stands out from other microalgal strains as they can accumulate up to 60% and 10% of their cell dry weight with lipids and EPA, respectively. These characteristics of *Nannochloropsis* spp have led to the sequencing of several genomes, the analyses of transcriptomic data under different conditions, and to the development of transformation techniques for this genus. However, these species had primitive genome editing tools and required extensive development in this field to further exploit their potential as microbial cell factories. The CRISPR-Cas technology emerged as a revolutionary, generic genome-editing tool in 2013. Until recently, research on microalgal organisms was hampered by the lack of efficient genetic tools.

Chapter 2 provides a detailed review on developing CRISPR-Cas technology as a genome editing tool in various microalgal and cyanobacterial strains. CRISPR-Cas based genome editing was first reported in *Chlamydomonas reinhardtii*. However, this initial study reported feeble efficiencies using a plasmid-based approach and attributed this outcome to the cytotoxic effects of Cas protein in *Chlamydomonas*. The development of transformation strategies for Cas ribonucleoprotein (RNP) in *Chlamydomonas* facilitated transient Cas existence in the host and produced mutants with high efficiencies. Additionally, co-transformation of editing templates along with the Cas proteins resulted in producing precise mutants via homology-directed repair (HDR). In *Nannochloropsis*,

a genome integrated plasmid was used for constitutively producing Cas9 to generate indels or integrate an antibiotic resistance gene to disrupt the target gene. Nevertheless, the mutants produced in *Nannochloropsis* harboured the integration cassette resulting in transgenic mutants and the lack of regulatory elements limited the application of CRISPR-Cas in this genus.

To address some of the bottlenecks associated with the CRISPR-Cas toolbox for *Nannochloropsis*, as described in Chapter 2, a Cas RNP-based genome editing system was developed for *Nannochloropsis oceanica* in **Chapter 3**, which can also be adapted for other species in this genus. In contrast to the conventional approach that generated uncontrolled indels to disrupt the target gene, a homologous recombination-based approach was employed to generate precise mutants. To this end, the homologous recombination (HR) efficiency of the species was verified to identify that this mechanism was ineffective in generating mutants. However, co-transformation of a Cas9 RNP targeting the gene along with the HR template designed to delete the target gene facilitated homology-directed repair (HDR) at the target site and produced precise mutants with efficiencies exceeding 70%. Additionally, multiple variants of the Cas proteins such as FnCas12a, LbCas12a and AsCas12a were tested by the same approach to identify the best Cas variant for genome editing in *N. oceanica*. In this experiment, FnCas12a emerged as the best Cas variant by generating HDR based mutants with 92% efficiency while AsCas12a had very low efficiency. LbCas12a and Cas9 generated mutants with similar efficiencies. However, the generation of HDR based mutants by this approach resulted in the integration of an antibiotic resistance cassette at the target site resulting in transgenic mutants.

To address this bottleneck and to generate markerless and scarless mutants using the Cas RNPs and HDR, a new approach was defined in **Chapter 4**. A dual Cas RNP transformation protocol that initially integrated an antibiotic resistance cassette along with a fluorescence marker by HDR at the target site was followed by a second round of transformation using RNPs targeting the fluorescence gene with the same HR template without any nucleotides between the upstream and downstream flanks. The cells lacking the fluorescence were selected using fluorescent activated cell sorting (FACS). This strategy resulted in scarless and markerless nitrate reductase mutants of *N. oceanica*. Even though this technique was successful in generating non-transgenic mutants, the RNP based approach required a laborious transformation protocol and highly purified Cas proteins. In a previous study, an episomal plasmid-based system harbouring Cas9 was demonstrated to generate non-transgenic mutants in *N. oceanica*. This system was adapted to develop an episomal plasmid-based Cas12a system for *N. oceanica*. A codon-harmonized fnCas12a sequence was initially tested and failed to express in *N. oceanica*. Subsequently, a human-optimized fnCas12a sequence was successfully expressed and efficiently used to generate indels in *N. oceanica*. The ribozyme systems used for producing the sgRNAs for Cas9 was dispensable while using Cas12a, simplifying the cloning procedures and facilitating the use of CRISPR arrays for multiplexing. This advantage was exploited to perform the first-ever multiplexed targeting in *N. oceanica* using Cas12a by introducing indels at three different loci in the genome in a single transformation. Furthermore, the episomal plasmid system was adapted to develop a CRISPRi tool for *N. oceanica*. Dead or catalytically inactive versions of Cas9 and fnCas12a proteins were tested to downregulate the expression of

a fluorescence gene expressed from the host genome. A dead version of Cas9 tagged along with a transcriptional repressor domain KRAB targeting the beginning of the gene sequence reduced the transcript levels by 85% showing a proof of concept of this system in *N. oceanica*. It was observed that the KRAB domain was required for efficient silencing. Also, Cas9 was a better tool for CRISPRi in *N. oceanica* compared to Cas12a. These advances reported in Chapters 3 and 4 could be summarized as the development of a comprehensive genome editing toolbox for microalgae *N. oceanica*.

The development of the genome editing toolbox was followed by strategies to gain novel insights into the microalgal lipid metabolism and subsequently perform targeted metabolic engineering to increase lipid production. In **Chapter 5**, a random insertional mutagenesis library of *N. oceanica* was developed by enriching high lipid producing mutants by five rounds of FACS. The insertional cassette for generating the mutagenesis was flanked by recognition sequences for a type IIS restriction endonuclease that cleaves few nucleotides away from the recognition site. This property was used to map the integration sites in the genome from the mutants with valuable phenotypes. Among various high lipid producing mutants, disruption in the APETALA2-like transcription factor improved the lipid productivity by 39%. Subsequently, a targeted mutant of the gene coding for APETALA2-like TF was generated using the Cas12a episomal plasmid system to confirm the phenotype.

In addition to producing large quantities of triglycerides (TAG), the production of eicosapentaenoic acid (EPA) is another unique feature of *N. oceanica*. Improving the EPA content and/or accumulating more EPA in the triglycerides (TAG) are valuable outcomes if achievable in *N. oceanica*. EPA was mainly observed in the sn3 position of the glycerol molecule to form TAG in *N. oceanica*. The characterization of various endogenous enzymes involved in the esterification of sn3 position of TAG molecules revealed multiple enzymes with high affinity towards EPA. The lysophosphatidic acid acyltransferase (LPAAT) enzyme catalyses the esterification of the sn2 position of the glycerol molecule. In **Chapter 6**, overexpression of an EPA specific LPAAT gene from the plant *Brassica napus* in *N. oceanica* was hypothesized to channel more EPA into the sn2 position of a glycerol backbone and improve the EPA content in TAG and total fatty acids (TFA). The BnLPAAT overexpression did not improve the lipid nor EPA content during optimal growth conditions. However, under nitrogen starved conditions, the EPA content in the TFA and TAG was increased by 28% and 38%, respectively. The improved EPA content indicates the potential of this approach to accumulate more EPA in TAGs. Nevertheless, future studies should co-overexpress EPA specific DGAT enzyme that esterifies the sn3 position of a glycerol backbone to further improve the EPA accumulation in TAGs.

Microalgae have been a sustainable source for commercially producing high-value compounds such as pigments, or whole-biomass to be used as nutraceuticals. Nevertheless, a photoautotrophic microalgal strain capable of producing lipids with high productivities that can reduce the cost of production and compete with non-sustainable sources is yet to be developed. The trade-off between lipid accumulation and growth has been a major bottleneck in improving the lipid productivities in microalgae. Limited availability of carbon and reducing power in the host strain upon triggering lipid production could

be attributed to the reduced lipid productivity. To achieve economically feasible biofuel production from microalgae, genome engineering manoeuvres will be required, ranging from simultaneously manipulating selected genes to rewiring or introducing entire pathways. And the currently developed CRISPR toolbox for microalgae can facilitate these explorations in microalgae. **Chapter 7** reviews various strategies to improve the generation of reducing power from light and carbon fixation via natural and synthetic routes in microalgae. Moreover, alternate routes other than the natural pathways are proposed to improve the microalgal productivities. The results and observations from all the chapters are summarized and discussed in **Chapter 8**. Next steps to further improve the genetic toolbox are briefly discussed along with potential strategies to engineer *N. oceanica* as a platform strain for producing EPA and TAGs.

About the Author

Mihris Ibnu Saleem Naduthodi was born on the 11th of May 1992 in the district of Malappuram, situated in the southern Indian state of Kerala. He holds a bachelor's degree in biotechnology and biochemical engineering from Sree Chitra Thirunal College of Engineering, Trivandrum. Subsequently, Mihris moved to the Netherlands in 2015 to pursue his Master's in cellular and molecular biotechnology from Wageningen University and Research. During the first year of his Master's, he was fascinated by the development of CRISPR-Cas proteins as genome editing tools. This led him to the Bacterial genetics lab chaired by Prof. John van der Oost in the Microbiology department of the university to perform his master thesis. Mihris was supervised by Dr. Ioannis Mougiakos and Prof. Richard van Kranenburg for his thesis, where he developed thermoCas9 as a genetic tool for gene silencing in thermophilic bacteria *Bacillus smithii*. In parallel, he also worked on engineering the metabolism of *B. smithii* to improve succinate production. His work was selected as the best student thesis in Wageningen university for the academic year 2017.



After completing his Master's, Mihris was appointed as a PhD candidate under the supervision of Prof. John van der Oost, Prof. Maria Barbosa and Dr Sarah D'Adamo, where he developed a comprehensive genetic toolbox for microalgae *Nannochloropsis oceanica*. He developed a Cas ribonucleoprotein-based approach to develop markerless and scarless mutants in *N. oceanica*. Furthermore, he developed a plasmid-based system that facilitates multiplexed genome editing and efficient gene silencing in *N. oceanica*. Most of his work can be found in this thesis. Currently, he works as a Research Associate in the biology department at the University of York, UK.

Overview of completed training activities

Discipline specific activities

Microalgae Process Design: from cells to photobioreactors, VLAG, Wageningen, 2017
ICRC conference, Bielefeld University, Germany, 2017
Microbiology Centennial Symposium, WUR, Wageningen, 2017
Symposium "What is Life", WUR, Wageningen, 2018
Algae Europe, EABA, Amsterdam, 2018
National Biotechnology Conference, NBV, Ede, 2019
Metabolic Engineering and Systems Biology, Chalmers University, Gothenburg, Sweden, 2019
Biofuels Conference, Frontiers meetings, Rome, Italy, 2019

General Courses

VLAG PhD Week, VLAG, Baarlo, The Netherlands, 2018
Project and time management, Wageningen Graduate School (WGS) Wageningen, 2018
Poster and Pitching, WGS, Wageningen, 2018
Supervising MSc and BSc students, WGS, Wageningen, 2018
Career perspective, WGS, Wageningen, 2021

Other activities

Preparation of research proposal, Department of Microbiology, WUR, 2017
PhD study trip to San Diego, USA, Department of BPE, WUR, 2018
Weekly group meeting, Bacterial genetics group, MIB, (2017-2021)
MAGNIFICENT consortium meeting, Department of BPE, WUR, Arnhem, 2017
MAGNIFICENT consortium meeting, Necton, Portugal, 2018
MAGNIFICENT consortium meeting, MadeBiotech/NatureXtracts, Madeira, Portugal, 2018
MAGNIFICENT consortium meeting, Algosolis, Nantes, France, 2019
MAGNIFICENT consortium meeting, CMP, Lisbon, Portugal, 2019
MAGNIFICENT consortium meetings, Department of BPE, WUR, 2020/2021

List of Publications

Naduthodi, Mihris Ibnu Saleem, Christian Südfeld, Nicola Trevisan, Emmanouil Avitzigiannis, Eduard van Lith, Javier Alcaide Sancho, Sarah D'Adamo, Maria J. Barbosa, and John Van Der Oost. "A comprehensive genome editing toolbox for microalgae *Nannochloropsis oceanica* IMET1." *ACS Synth. Biol.* (2021). <https://doi.org/10.1021/acssynbio.1c00329>

Muñoz, Camilo F., Christian Südfeld, **Mihris IS Naduthodi**, Ruud A. Weusthuis, Maria J. Barbosa, René H. Wijffels, and Sarah D'Adamo. "Genetic engineering of microalgae for enhanced lipid production." *Biotechnology Advances* (2021): 107836.

Südfeld, Christian, Michal Hubáček, Daniel Figueiredo, **Mihris IS Naduthodi**, John Van Der Oost, René H. Wijffels, Maria J. Barbosa, and Sarah D'Adamo. "High-throughput insertional mutagenesis reveals novel targets for enhancing lipid accumulation in *Nannochloropsis oceanica*." *Metabolic Engineering* 66 (2021): 239-258.

Naduthodi, Mihris Ibnu Saleem, Nico J. Claassens, Sarah D'Adamo, John van der Oost, and Maria J. Barbosa. "Synthetic biology approaches to enhance microalgal productivity." *Trends in Biotechnology* (2021).

Naduthodi, Mihris Ibnu Saleem, Prarthana Mohanraju, Christian Südfeld, Sarah D'Adamo, Maria J. Barbosa, and John Van Der Oost. "CRISPR-Cas ribonucleoprotein mediated homology-directed repair for efficient targeted genome editing in microalgae *Nannochloropsis oceanica* IMET1." *Biotechnology for biofuels* 12, no. 1 (2019): 1-11.

Naduthodi, Mihris Ibnu Saleem, Maria J. Barbosa, and John van der Oost. "Progress of CRISPR-Cas based genome editing in photosynthetic microbes." *Biotechnology journal* 13, no. 9 (2018): 1700591.

Mougiakos, Ioannis, Prarthana Mohanraju, Elleke F. Bosma, Valentijn Vrouwe, Max Finger Bou, **Mihris IS Naduthodi**, Alex Gussak, Rudolf BL Brinkman, Richard Van Kranenburg, and John Van Der Oost. "Characterizing a thermostable Cas9 for bacterial genome editing and silencing." *Nature communications* 8, no. 1 (2017): 1-11.

Acknowledgements

To do a PhD has been my dream since my Bachelor's. When I am getting close to this dream, *Alhamdulillah*, I am happy that I had an incredible journey throughout my time in Wageningen. I could not have asked for a better place to do my PhD, and it would not have been fruitful if not for the support from some of the fantastic people that I would like to thank in this chapter.

John, thank you for being an excellent mentor and supervisor in the last four years. It was an honour to complete my PhD under your supervision and thank you for giving me this opportunity. I was always fascinated by how you approach the bottlenecks in experiments and the ideas you deliver during our brainstorming sessions. Moreover, the atmosphere in Bacgen lab that encourages PhDs to express novel ideas out of our expertise and support to execute them is rare in academia. I hope you can keep inspiring the new generation of scientists and make incredible advances on the topics that you pursue.

Maria, we first met during the course bioprocess design in my first year of MSc. You mentored our team of 6-7 people, and we were all pleased by the way you supported us throughout the one-month project. After that, I never thought our paths would cross again. And to my luck, in another year, you were supervising me for my PhD thesis. Good things happen when we are at the right place at the right time. I believe I was at that right place (with John) when the MAGNIFICENT grant came to you, and you wanted someone with CRISPR experience to work on our beloved Nannochloropsis. Thank you for being the most supportive and caring supervisor anyone can have, and I wish you all success and happiness in all your endeavours. I still must learn your secret of sleeping very few hours during the MAGNIFICENT meetings and being super energetic the next day.

Sarah, you joined my supervision team when we were moving towards developing plasmid-based tools for Nannochloropsis. At that point, it was great to have someone to approach with ample experience in the molecular aspects of microalgal biotechnology. If not for your expertise in this field, I might have made some decisions in the project that could have taken me in the wrong direction, and I would not have completed my project as expected. Thank you for all your time, patience, and valuable advice throughout the last three years. I believe the new research lines you have started are auspicious. I hope to see some excellent works published in the near future.

Dr Mougiakos et al., thank you for all the support, patience, motivation, scolding, supervision, and recommendation (to John) that made me a PhD candidate in Bacgen. I know that I was not the best thesis student a supervisor wished to have when I started working with you. I am happy that I was able to change that impression about myself very soon. I am glad that I could work under your supervision for almost a year, where I had an exponential learning curve in my laboratory skills. You will be an asset for any research lab, and I wish you all success in your life.

Nico, you never forget to say hi when you see me working alone in the office or the lab. I had a great time writing the review paper with you. I am sure that without your input and suggestions, it would have been tough for me to get it done. Also, thank you for being my occasional weekend supervisor in the lab during my MSc thesis.

The beginnings of new steps can be overwhelming. I had a good start for my PhD, and it was primarily due to the support from a few fantastic colleagues. **Tom van de Weijer**, I am thankful for your help setting up the incubators and light chambers for growing microalgae in our lab. It was great to work with you, and you were severely missed in Bacgen when you moved to your new job. **Prarthana**, I had no clue about protein expression and purification before my PhD. When I designed my first project to use the RNPs, your involvement in the project accelerated the production of proteins and resulted in my first research article as the first author.

The technical staff of MIB, the backstage works that you do reflect how well the MIB department has been doing in delivering exciting publications. I am thankful to **Rob** and **Guus** for all their efforts to make our lab an excellent place to work. Thank you for keeping all the orders on time and ensuring the lab workers get everything they need.

The MAGNIFICENT boys and the lady manager (**Narcis, August, Calvin, Iago** and **Marta**). What a wonderful time we had the last four years. We might have been the luckiest PhD candidates in terms of travelling to exotic locations. I enjoyed all these trips, and it was an excellent opportunity to get to know all of you very well. **Narcis**, thank you for all the help, patience, and support throughout the LPAT works. It was constructive to have you around during the experiments and your guidance throughout my time in BPE was incredibly valuable.

I did not have a lot of students to supervise during my PhD. But, I was lucky to have some intelligent and hardworking students. **Manos (Emmanouil)**, you were the first student I managed myself and one of the humblest people I have seen. **Pepijn**, I know the project you worked on was challenging, and I respect your persistence to keep going until the end. **Nicola**, you are intelligent, hardworking, and one of the coolest people to work with. It was nice to see you as a colleague later on. **Eduard** and **Javier**, you both worked during the pandemic, and I was happy to see you both complete the thesis with good results. I would like to wholeheartedly thank all my students for all your hard work and inputs that helped me take this thesis forward.

I always thought that I had the best office space on our entire floor. I had a fantastic view from my space, and I was right next to the lab as well. Thanks to my office mates **Jeroen, Joyshree, Costas, Ioannis** and **Eric** for making the office room the best to work. **Jeroen**, I have always had great chats with you, and I admire your knowledge on diverse topics. I hope you will have a great time in Sweden with your new research group.

The Bacgen group, I could not have asked for any other group of talents to work with during my PhD. I always wonder about the diverse research we all do, and somehow, John has managed to put us together where we complement each other's work.

Wen, thanks for the initiatives you took to organise most of the lab stuff, including the successful science hour that everyone enjoys. It was always fun to have you around in the lab. **Sjoerd**, I admire the knowledge and lab skills that you possess. I always came to you when I had a problem that I could not solve, especially regarding the guide designs of CRISPR, and you always had a way of solving these problems in a simple way. I have learned a lot from you in the past four years and wish you all success in your career. **Belen**, you are one of the most helpful people in the lab, and I have always enjoyed working in your company. **James, Mr Dykstra** (or Dijkstra), you seem to be a silent person at first, but an amiable and cheerful person to be around. Thanks for all the squash games we had. It was a lot of fun, and I hope you will beat me the next time if we get a chance to play (☺). **Mr Swartjes**, it was fun to have you around in the lab, and I always enjoyed the small jokes you made in the lab. **Stijn** and **Jurre**, I admire the hard work you both put in the work you do, and I firmly believe the Scope Biosciences is going to be a proud venture for all of us at Bacgen and MIB. **Joep**, we have been together in the Bacgen since our Master thesis, it was nice to have you along and thank you for being a great lab mate. I wish you all success in completing your PhD.

Despoina, you are the most hardworking person I have come across. I am sure that all your effort and determination will take you a long way in your career. **Evgenious**, it was nice to see your promotion from student to a PhD candidate in Bacgen. I hope your superb ideas in collaboration with physical chemistry become valuable publications soon. **Lorenzo**, you are another bright talent who came to our lab as a student and was promoted to a PhD candidate. It was a pleasure to have you around as a colleague, and thank you for all the good times in the lab. **Carina**, during my free time, I occasionally search about the presentation you gave during the science hour on sleep cycles and lucid dreaming. This subject has been fascinating for me to read and listen to out of office hours. I hope your work on gene drives will be a great success. **Thijs**, I have always found your ideas and the projects you do very futuristic. In a way, the kind of research you do encouraged me to push myself forward and think of ambitious ideas. The collaboration between you and Costas is a good mix of talents, and I hope all your grants and thoughts result in fruitful outcomes.

Max, Agnes and **Valentijn**, if it was not for you guys and the fun we had in the lab, my Master thesis life might have become dull. Thanks for all the fun, learnings and camaraderie we shared during the phase of our MSc.

Camilo, thanks for all the support and advice during the early phase of my PhD thesis. It was great to have someone in the lab with experience in the molecular part of microalgae. Moreover, thanks for inviting me and making me a part of the review that I thoroughly enjoyed working with you. Wendy, thanks for all the help you provided during the lipid analysis.

Eric Poliner, I believe that I met you at the right time. The pNOC plasmids that you developed have been the backbone of my PhD thesis. Thank you for all the advice and meetings we had, along with the Nanno round table that helped me take my project

forward. I hope the development of Cas12a and CRISPRi will be exciting additions to the Nannochloropsis toolbox along with your Cas9 system. I also show my gratitude to all the members of the Nanno round table meetings for the advice on laboratory bottlenecks associated with working on our beloved strain *Nannochloropsis oceanica*.

Having a chance to play cricket during the weekends were the best stress buster during my time as a PhD in Wageningen. It was nice to meet and get to know many fellow Indians during our games. Also, our venue for the games evolved a lot in the past six years, starting from the parking slot where we presently have the Unilever building to the grass field in the running tracks and finally to the football ground. From seniors in the group **Simha, Nitin, Sandeep, and Soumya** to the latest additions such as **Raghav, Rohan, Nithin, Vedant** and **Kalyan**, thank you for all the fun and healthy competitions we had during the past few years. **Shimlal** and **Jonathan**, you were the only people in Wageningen to whom I could speak in Malayalam. I had a great time whenever we all were together, either with cricket or fantastic food.

Christian, without you in our research team, I would have struggled a lot in developing the Cas systems for our strain. It was a great fortune for me that you started your thesis a few months before me and had already developed the transformation protocol for *Nannochloropsis*. The suggestions and brainstorming during our meetings have been very fruitful, and I am happy that we both could complement each other's work in the past four years. I am proud to have you as one of my paranymphs, and I am sad that I physically missed being present at your PhD defence. But, I am sure we will have a great time during my defence in December.

Costas or Mr Patinios, I had a great time being by your side in the office in the past four years. My first meeting with you was on the basketball court almost six years back during the beginning of my Master, and I never thought that we would become office mates and great friends in the future. Our trip together during the PhD week and to Gothenburg were all lot of fun and will be great memories. You are one of the most open and calm people in the lab. I admire your passion for science and the courage to pursue the ambitious projects and ideas that come your way. It will be an honour to have you as one of my paranymphs. I hope all your hard work paves the way towards the fantastic future you deserve. And, thanks for all the tricky jokes that you come up with in the meetings. Some of them that I could make out was hilarious and cheered up everyone in the meetings.

Bajikka (Baji Krishna), thank you for being the person that lifted me through all the troubles that I had during the time of my bachelors. If not for the care and advice that came from your side, I would have had a hard time completing my bachelors, and I might not have even ended up in Wageningen. I hope you can guide and inspire the young biotechnology students at SCT for a long time.

Now it is time for the family, and I have no clue where to start. **Uppappa** and **Ummamma**, you both have been a source of inspiration. All your prayers and support throughout my life have taken me where I am and will be in the future. **Shaju Maama** and **Maamama**, I am very sure that without your support, my studies in the Netherlands would not have been

as easy as it went. I will always be in debt to the invaluable support that I had from both of you. **Mishab** and **Shumaima**, thanks for all the wishes, prayers and support throughout the past years and for bringing the bundle of joy of our family **Hesther** and **Hemin**, to this world. **Mariah**, I was not always the sweet brother you expected me to be, but I think it was better that I remained a strict one, and I thank you for all the unconditional love, prayers, and wishes you gave me. **Miqu**, you were always the quiet one at home, but for me, you were just a phone call away whenever I needed something to be sorted at home. Thank you for being the amazing brother that you are. I am also grateful to all my **appaapas**, **aamaamas** and **cousins** for all the good wishes and support throughout my PhD journey.

Uppa, I always wonder about the confidence and enthusiasm you had in sending me to Wageningen even when it was not that affordable. You were always there, a phone call away when I had any problem, and you always made sure I was safe and secured. You also made sure that I never had any pressures that could divert my focus from my studies. Even though I have missed your and Umma's physical presence on various occasions here in Wageningen, I was happy to see how proud you both were with all the small and big milestones I achieved throughout my time in Wageningen. I hope I can make you both even more happy and proud forever.

Umma, no one will ever be worried about my well being as you are. When I write an examination or submit a paper for publication, you have more tension and concern regarding the outcome than I ever had. Whenever I feel low, a phone call with you would cheer me up. The promise that things will get better, the motivation to keep going and assuring that I am always in your prayers makes me stronger after any bad day. You always wanted to visit Wageningen, where I lived for more than six years. I hope In Sha Allah we will make that trip soon.

Last but not least, I am thankful to my wife, **Shijisha**, who had a late start in being part of my life. It was fantastic to have you beside me through all the thick and thin times of my PhD. I appreciate all the sacrifices you had to make to ensure that I had a smooth ride in completing my works. The last few months of PhD were not easy, and I am glad that you were with me in Wageningen during this challenging time as a pillar of support. I admire your passion for your studies. I wonder how you study the whole night and then manage to get back to other chores after a few hours of sleep. I hope that all your hard work be rewarded by having multiple opportunities to make a career in the field you like. I am confident that without your support in preparing me for the interview and making my CV, I would not have landed my current job in York. I love you for everything that you are, and thank you for making me confident about life. I am excited to have this beautiful life with you by my side and hope to keep you happy and cheerful forever. I am also grateful to **Uppa**, **Umma**, **Shijas** and **Sania** for welcoming me to the family and supporting me throughout my PhD.

The research described in this thesis was financially supported by the MAGNIFICENT project funded by the Bio Based Industries Joint Undertaking (BBI-JU) under the European Union's Horizon 2020 (H2020) research and innovation programme under grant agreement No. 745754 to Prof. Dr Maria J. Barbosa.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Ines Vilalva

Printed by ProefschriftMaken

