# UC San Diego
## UC San Diego Electronic Theses and Dissertations

**Title**

Uncoupling nitrogen limitation and lipid accumulation in the marine diatom Phaeodactylum tricornutum by CRISPR-Cas9 genetic engineering

**Permalink**

https://escholarship.org/uc/item/33b708vz

**Author**

Moosburner, Mark

**Publication Date**

2021

**Supplemental Material**

https://escholarship.org/uc/item/33b708vz#supplemental

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO


Uncoupling nitrogen limitation and lipid accumulation in the marine diatom *Phaeodactylum tricornutum* by CRISPR-Cas9 genetic engineering


A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy


in


Marine Biology


by


Mark Moosburner


Committee in charge:

      Andrew E. Allen, Chair
      Lena Gerwick
      William Gerwick
      Prashant Mali
      Stephen Mayfield


2021

The dissertation of Mark Moosburner is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

TABLE OF CONTENTS

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AF | Auto-fluorescence |
| amiRNA | Artificial micro ribonucleic acid |
| ASW | Artificial seawater |
| ATP | Adenosine triphosphate |
| cDNA | Copy deoxyribonucleic acid |
| CFP | Cyan fluorescent protein |
| CoA | Coenzyme A |
| CRISPR | Clustered regularly interspaced short palindromic repeats |
| DNA | Deoxyribonucleic acid |
| DSB | Double-stranded break |
| eGFP | Enhanced green fluorescent protein |
| ELISA | Enzyme-linked immunosorbent assay |
| FAME | Fatty acid methyl ester |
| FP | Fluorescent protein |
| GA | Gibson assembly |
| GG | Golden gate assembly |
| GMO | Genetically modified organism |
| GOI | Gene of interest |
| GS-GOGAT | Glutamine synthetase - glutamate synthase cycle |

| | |
|---|---|
| GUS | β-glucuronidase |
| HDR | Homology directed repair |
| HNS | Highly-nitrate sensitive |
| HR | Homologous recombination |
| HRMC | High-resolution melt curve |
| HVM | High-valued molecule |
| KO | Knockout |
| LacZ | Lactose operon |
| LB | Luria broth |
| LUC | Luciferase |
| miRNA | Micro ribonucleic acid |
| MMEJ | Microhomology-mediated end joining |
| MN | Mega-nuclease |
| mRNA | messenger ribonucleic acid |
| NHEJ | Non-homologous end joining |
| NPQ | Non-photochemical quenching |
| ORF | Open reading frame |
| PAM | Protospacer adjacent motif |
| PCR | Polymerase chain reaction |
| Pol-III | Ribonucleic acid polymerase III class promoter |

| | |
|---|---|
| PUFA | Polyunsaturated fatty acid |
| qPCR | Quantitative polymerase chain reaction |
| RNA | Ribonucleic acid |
| RNAi | Ribonucleic acid interference |
| RNP | Ribonucleoprotein complex |
| TAG | Triglyceride |
| TALEN | Transcription activator-like effector nuclease |
| TCA | Citric acid cycle |
| TIDE | Tracking indels by deconvolution |
| TOC | Total organic carbon |
| TON | Total organic nitrogen |
| TOPO | Topoisomerase cloning |
| sgRNA | Single-guide ribonucleic acid |
| StLDP | Stamenopile lipid droplet protein |
| UV | Ultraviolet |
| WT | Wild-type |
| YFP | Yellow fluorescent protein |

# LIST OF TABLES

# LIST OF FIGURES

# LIST  SUPPLEMENTAL TABLE AND INFORMATION

LIST OF SUPPLEMENTAL FIGURES

experiment for Pt_WT and Pt_GS2-KO cell lines. (b) Gel image to compare amplicon sizes of the GS2 gene from Pt_WT (1363-bp) and Pt_GS2-KO (1170-bp) cell lines

Supplemental Figure 2.10: Growth profiles of Pt_WT and Pt_GS2-KO during the sampling of TOC/TON, FAME,  and RNAseq on media  supplemented with (a) ammonium, (b) nitrate, (c) glutamine, and (d) no nitrogen (N-free).

Supplemental Figure 2.11: Instantaneous photosynthetic efficiency measured by Fv/Fm of Pt_WT and Pt_GS2-KO during the sampling of TOC/TON, FAME, and RNAseq on media supplemented with (a) ammonium, (b) nitrate, (c) glutamine, and (d) no nitrogen (N-free).

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Andrew E. Allen for his support as the chair of my committee. His guidance has proven invaluable through formulating research projects, providing financial and resource stability, affording me a breadth of academic opportunities, and allowing me intellectual freedom to pursue my research interests.

I would also like to acknowledge James McCarthy and Sarah R. Smith of the J. Craig Venter Institute for sharing their knowledge of diatom ecology, cell biology, and metabolism with me. They were always available to me when I needed an idea sounding board or help parsing complex concepts in cell biology and genetics.

I would also like to acknowledge Vince Bielinski of the J. Craig Venter Institute for his sharing with me his expertise in algal molecular biology laboratory techniques.

Section 1.1 of Chapter 1, in part, is material in revision to be published as it may appear in The Molecular Life of Diatoms, Chapter 26 (2021). Moosburner, M., Allen, A. E., & Daboussi, F. (2021) Genetic Engineering in Marine Diatoms: Current Practices and Emerging Technologies. In A. Falciatore, T. Mock (eds.), *The Molecular Life of Diatoms*. Chapter 26. Springer Press, London, UK. In Revision. Mark Moosburner was the primary author and principal investigator of this manuscript. Fazya Daboussi was the secondary author of this manuscript. All co-authors contributed to the revision and editing of the manuscript.

Section 1.2 of Chapter 1 is, in full, a reprint of the manuscript as it appears in Frontiers in microbiology, 2020. Moosburner, M. A., Gholami, P., McCarthy, J. K., Tan, M., Bielinski, V. A., & Allen, A. E. (2020). Multiplexed knockouts in the model diatom Phaeodactylum by episomal delivery of a selectable Cas9. *Frontiers in microbiology, 11*, 5. Mark Moosburner was the primary author and principal investigator of this manuscript.

Chapter 2, in part, is currently being prepared for submission for publication of the material. Moosburner, M., Gholami, P., Gurr, J., Du, N., McCarthy, J., Smith, S. A., Allen, A. E. (2021) Glutamine synthetase 2 knockout in the diatom Phaeodactylum tricornutum accumulates lipids early without compromising growth. In Preparation. Mark Moosburner was the primary author and principal investigator of this manuscript.

VITA

2011            Bachelor of Science, Boston University

2012-2014      Research Assistant, Harvard Medical School

2021            Doctor of Philosophy, University of California San Diego


PUBLICATIONS

Moosburner, M. A., Gholami, P., McCarthy, J. K., Tan, M., Bielinski, V. A., & Allen, A. E. (2020). Multiplexed knockouts in the model diatom Phaeodactylum by episomal delivery of a selectable Cas9. *Frontiers in microbiology, 11*, 5.

Moosburner, M., Allen, A. E., & Daboussi, F. (2021) Genetic Engineering in Marine Diatoms: Current Practices and Emerging Technologies. In A. Falciatore, T. Mock (eds.), *The Molecular Life of Diatoms.* Chapter 26. Springer Press, London, UK. In Revision.


FIELDS OF STUDY

Major Field: Marine Biology

      Studies in Cellular and Molecular Biology
      Professor Andrew E. Allen

ABSTRACT OF THE DISSERTATION


Uncoupling nitrogen limitation and lipid accumulation in the marine diatom *Phaeodactylum tricornutum* by CRISPR-Cas9 genetic engineering


by


Mark Moosburner


Doctor of Philosophy in Marine Biology


University of California San Diego, 2021


Professor Andrew E. Allen, Chair


Characterizing the function of a single gene can be achieved by investigating how the gene and its protein product act in a cell or by observing how a cell responds when the gene is removed or knocked out. The ladder, called reverse functional genetics, has been greatly bolstered by the development and implementation of genetic engineering tools. Most recently,

CRISPR-Cas9 genetic engineering has enabled functional genetics in a wide variety of organisms due to its ease of use and superior efficiency in gene mutagenesis compared to older technologies.

Functional genetics studies in marine diatoms, photoautotrophic eukaryotic algae, have too been aided by the development of CRISPR-Cas9 genetic engineering. This study optimized the CRISPR-Cas9 tool in the pennate diatom *Phaeodactylum tricornutum* for bacterial-mediated conjugation cell transformation methodology. Conjugation in diatoms occurs by the exchange of an artificial chromosome or episome from *E. coli* to *Phaeodactylum.* Here, a cloning scheme was developed to permit the construction of CRISPR-Cas9 episomes that can target multiple genomic loci simultaneously.

Nitrogen metabolism of marine diatoms contained unique features that gives them a competitive  advantage in oceanic food webs. One feature is the compartmentalization of two glutamine-glutamate metabolic cycles (GS-GOGAT) within the chloroplast and mitochondria, respectively. This study utilized the CRISPR-Cas9 episome tool to functionally analyze a key GS-GOGAT enzyme, glutamine synthetase 2 (GS2), in *Phaeodactylum* by producing mutant cell lines devoid of GS2. It was found that while cell growth was not compromised, GS2 mutants accumulated lipids, a precursor to biodiesel.

Lastly, the CRISPR-Cas9 episome was used to investigate this phenomenon of lipid accumulation in a nitrogen-gene mutant. To do so, a knockout library of nitrogen metabolism genes was produced. Also, the library episomes were fashioned with a novel lipid body fluorescent protein reporter that, by microscopy and flow cytometry based detection methods, was indicative of lipid production in *Phaeodactylum* mutants.

Here, nitrogen limitation and lipid biosynthesis was uncoupled in a *Phaeodactylum* GS2 mutant. The expansion to high-throughput mutagenesis of a nitrogen gene library will support research into diatom biofuels.

# Chapter 1

# Genetic Engineering in Marine Diatoms: Current Practices and Emerging Technologies

Mark Moosburner

**Abstract of Chapter 1**

Diatom ecology, physiology, and biotechnology research has recently been bolstered by the development of genetic engineering practices. The whole genome sequencing of two diatom species, *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*, has greatly aided in shaping these into laboratory models by allowing researchers to characterize gene-expression regulatory elements that are essential tools for synthetic biology and genetic engineering technologies. Tools that have been previously developed for other model organisms (yeast, plants, animals) have been quickly adapted for diatom research like gene over-expression, gene silencing, and targeted mutagenesis tools such as CRISPR-Cas9. These molecular tools rely on nucleic acid transformation methods to deliver the genetic payload to the diatom, be it particle bombardment or bacterial-conjugation. Now, more diatom species that may have alternative ecological relevance in diatom biology like cold-adaptation or a benthic life-cycle are entering the genetic engineering fold. The breadth of molecular and genetic engineering tools are reviewed here.

Finally, one development in diatom targeted mutagenesis using episomal-based conjugation transformation to deliver CRISPR-Cas9 components to *Phaeodactylum tricornutum* is outlined. The published manuscript, Moosburner et al. (2020), describes the cloning methodology for producing CRISPR-Cas9 episomes for multiplexed mutagenesis, validation methods for mutagenesis, and the first report of homology-directed repair CRISPR-Cas9 gene editing in a diatom.

**Introduction of Chapter 1**

Genome editing is a powerful approach to manipulate the genetic heritage of cells by introducing genetic modifications at specific genomic loci, including precisely targeted deletions or insertions. Such manipulations enable mechanistic investigation of complex biological systems and result in new insights, ideas, and strategies for improving organismal performances or integrating new features. Although genome editing in microalgae emerged more recently in comparison with classic model organisms such as yeasts, bacteria, plants or mammalian cells, more development has occurred at a rapid pace, particularly for diatoms. The diatom research community benefited from the many advances established in mammalian and plant systems and subsequently  established proof of concept for targeted gene insertion and inactivation (Daboussi et al. 2014, Weyman et al., 2015). Such strategies were rapidly deployed in the diatom research community thanks, in part, to robust platforms for community development and information sharing (e.g., Gordon and Betty Moore Foundation initiative in Experimental Model Systems for microbial eukaryotes, the biennial international Molecular Life of Diatoms congress, and Innovative Training Networks). More recently, notable improvements have been obtained in order to increase genome editing efficiency and decrease uncontrolled effects such as long-term background expression of the nuclease or random insertion of all or part of the plasmid DNA into the diatom genome.

To explore the physiological properties and exploit the full potential of diatoms, genetic approaches have been developed during the past three decades. There are two ways to link the sequence and function of a specific gene: forward and reverse genetics. While forward (also called classical) genetics aims to identify the sequence change that underlies a specific mutant phenotype, reverse genetics tries to gain insight into the underlying function by targeting one

specific gene. Implementation of both practices have been bolstered by molecular tools that enable the examination of individual genes and metabolic pathways. Researchers working on diatoms, which are relatively new laboratory organisms compared to traditional model organisms like plants, animals, and yeast, have largely adapted and improved upon molecular tools from these classical model organisms to accomplish functional genetic investigations.

DNA delivery methods for targeting nuclear and plastid genomes as well as reverse genetics methodology enabling the overexpression of endogenous or heterologous genes and silencing of specific genes in diatoms are now available (for review: Huang & Daboussi, 2017). However, it has become crucial to develop new tools that facilitate modification of the diatom genome in a site-specific manner. Such tools have been established in the recent years through development and use of site-specific nucleases suitable for targeted genome engineering or genome editing (For reviews: Kroth et al. 2018, Huang & Daboussi 2017). This chapter first provides a description of traditional genetic manipulation methods and targeted genome engineering approaches that have significantly contributed to improved understanding of basic cellular functions and metabolic processes in diatoms. Subsequently limitations associated with the use of such molecular manipulation as well as recent developments to circumvent them are discussed. Finally, we offer perspective on new challenges related to genome engineering and how such approaches might shape the future of diatom research.

## 1.1 The Molecular and Genetic Toolkit for Diatoms

Whole genome sequencing for a number of diatom species has generated a breadth of information that provides further insight into the roles diatoms play in ecology, biogeochemistry, and cellular metabolism. Nucleotide-level information for all of the genetic components (protein coding sequences, promoter, transcriptional effectors, etc.) in diatom genomes made available the information necessary to develop the tools required to conduct genetic manipulations. In practice, a central aspect of functional genetics requires, modifying the transcriptional activity of a single gene in order to infer the function of that gene and its product (i.e., protein). To accomplish this task in a diatom cell, a number of laboratory methods were developed to (1) introduce foreign nucleic acids, (2) transcriptionally express genes of interest and quantify the cellular impacts, and (3) manipulate endogenous gene targets within a diatom's genome (reviews: Huang & Daboussi 2017, Faktorova et al. 2020, Falciatore et al. 2020).

### 1.1.1 Genetic Transformations in Diatoms

Genetic transformation refers to the introduction of foreign nucleic acids followed by functionality of the introduced material within the cell. A number of previously established genetic transformation methods have been repurposed for diatom cells including micro-particle bombardment (i.e. biolistics), electroporation, and bacterial-mediated conjugation. Table 1.1 outlines the attributes for and disadvantages to using the available transformation methods in diatoms.

**Table 1.1** Summarization of attributes and disadvantages of biolistics, electroporation, and conjugation in diatoms. Reference articles for attributes and disadvantages can be found in the text.

| | Attributes | Disadvantages |
|---|---|---|
| **Biolistics**<br>* 1-50 transformants per $10^8$ cells | - Robust method that works for many diatoms<br>- Widely adopted throughout diatom research community<br>- Can transform multiple plasmids at once<br>- Delivery of ribonucleoprotein (RNP) complexes | - Need a specialized and costly bombardment equipment (BioRad PDS-1000)<br>- Random and multiple integrations, poor reproducibility<br>- Cell damage leads to long time lag in colony formation |
| **Conjugation**<br>* 100-1000 transformants per $10^8$ cells | - No integration into the genome<br>- Reproducible transgene expression between cell lines<br>- Removal of episome after transgene expression<br>- Semi-high throughput<br>- Shorter time to produce transgenic colonies | - Cloning of large (>15kb) episomal plasmid<br>- Potential loss of episomal genetic elements |
| **Electroporation**<br>*50-400 transformants per $10^8$ cells | - Fast protocol time | - Random and multiple integrations<br>- Expensive electroporation equipment |

### 1.1.1.1 Biolistics

Biolistics was first successfully demonstrated by delivering nucleic acids to live onion plant cells. To transform via biolistics, nucleic acids were loaded onto heavy-metal beads, typically tungsten and gold, and propelled at cells plated on solid growth medium at high velocity. The first development of biolistics transformation in two diatom species *Cyclotella cryptica* and *Navicula saprophila* marked a major turning point for diatom molecular biology research (Dunahay et al. 1995) and has since been utilized in numerous other diatom species (*Phaeodactylum tricornutum*, *Thalassiosira weissflogii*, *Cylindrotheca fusiformis*, *Thalassiosira*

*pseudonana, Chaetoceros gracilis, Amphora coffeaeformis, Fistulifera solaris, Pseudonitzschia multistrata, Pseudonitzschia arenysensis, Skeletomena marinoi,* and *Fragilariopsis cylindrus*) (refer to Faktorova et al. 2020 for a review on transformation in diatoms).

To accomplish biolistics in a diatom cell, a few technical criteria are necessary. First, the domestication of a diatom monoculture is necessary to avoid bacterial contamination and potential competitors and grazers. The culture must have the ability to live and grow on a solid growth medium at high cell density. The diatom culture is seeded on the surface of the solid growth medium prior to biolistics. Following transformation, the diatom cells need to recover from the impacts of the beads and are incubated in low to zero light conditions for 48 hours before cell line selection.

The preparation of the nucleic acid payload on the projectile beads needs to be considered. Nucleic acids such as linear and circular DNA are typically hybridized to tungsten beads (0.5μm - 17.5μm diameter) while RNA and proteins have been shown to hybridize to gold beads more effectively (Serif et al. 2018). Beads of smaller sizes (0.5μm - 0.7μm) are most effective in diatoms because they do not cause excessive cell damage and lethality (Dunahay et al. 1995, Apt et al. 1996).

For diatoms, biolistics transformation has advantages compared to other methods. One advantage is that the development of biolistics in diatoms was aided by past experiences and optimization for biolistic transformation methods in plant cells. Biolistics has also bolstered development of transformation by direct delivery of RNA and proteins directly, which is particularly helpful for work with diatom species that lack genetic information like promoters (Serif et al. 2018). Biolistics also conveniently allows for simultaneous delivery for multiple plasmids.

Biolistics in diatoms also has disadvantages. Compared to bacterial-mediated conjugation the transformation efficiency is 10-100X low (see Table 1.1 for efficiencies). Also, the transformation of DNA into diatoms by biolistics always results in the integration of the DNA randomly throughout the nuclear genome (George et al. 2020). Integration of DNA at random genomic loci could interrupt genetic functionality such a protein coding gene and would knock-out (KO) the gene unintentionally. Further, it is unpredictable how many copies of the transformed DNA are integrated. The uncertainty in copy number and integration loci could lead to poor reproducibility as well as loss or gain in cellular functions that can be interpreted erroneously.

### 1.1.1.2 Bacterial-mediated Conjugation

Delivery of DNA in the form of an artificial chromosome or episome to diatoms through bacterial conjugation was developed for the marine diatoms *Phaeodactylum* and *Thalassiosira pseudonana* (Karas et al. 2015). Bacterial conjugation involves the direct cell-to-cell contact of an *Escherichia coli (E. coli)* cell, harboring both a conjugative plasmid (pTA-MOB) and cargo plasmid (episome), and a diatom cell. The pTA-MOB plasmid was redesigned from an RK2 plasmid (Strand et al. 2014) to allow the formation of molecular bridges via cell-to-cell pili, while permitting transfer of an episomal plasmid that contained an origin of transfer element (oriT), from the *E. coli* cell to the *Phaeodactylum* cell. After successful transfer, the episome is not integrated into the nuclear genome but rather maintained as an additional chromosome. A centromeric element, CEN6-ARSH4-HIS3, is included on the episome that is necessary for nuclear episomal maintenance and efficient segregation during cell division (Karas et al. 2015).

Interestingly, conjugated cell lines are able to lose the episome after successive culture passages in the absence of antibiotics (Diner et al. 2016, Sharma et al. 2018).

Since its inception, conjugation has become more commonly utilized among diatom researchers. There are numerous advantages to conjugation over biolistics besides significantly increased transformation efficiency. Principally, the random integration of payload DNA that exists in biolistics transformation is avoided when using conjugation. Therefore, conjugation provides researchers the ability to control copy number within the diatom cell and expression profiles of delivered transgenes; a pillar of synthetic biology and genetic engineering. While this degree of control does ultimately result in improved reproducibility, episomal transgene expression and plasmid stability can vary between transformants and therefore episomal transformations require care in selection of cell lines with well-maintained episomal plasmids over multiple generations (Sharma et al. 2018, George et al. 2020). Additionally, ability to lose an episome after transgene activity by antibiotic removal is an attractive feature for generation of mutant diatoms followed by removal of the transgene following mutagenesis. Lastly, bacterial-mediated conjugation requires basic molecular biology and microbiology techniques including experience in plasmid design and cloning and ability to culture *E. coli* and diatom cell lines. The lack of expensive equipment such as a particle delivery system for biolistics and a gene-pulse system for electroporation makes conjugation more easily attainable for researchers.

### 1.1.1.3 Electroporation

Electroporation is an electrical current driven genetic transformation method that has also been adapted for delivery of nucleic acids to eukaryotic cell lines. To date, three diatom species, *Phaeodactylum* (Miyahara et al. 2013), *Chaetoceros gracilis* (Ifuku et al. 2015), and

*Skeletomena marinoi* (Johansson et al. 2019) have been reported being transformed using electroporation, including for introduction of a conjugation episome (Karas et al. 2015). Compared to the well-established biolistics approach and recently adapted conjugation, electroporation in diatoms has not gained in popularity amongst diatom researchers as a transformation technique due to relatively low transformation efficiencies, particular in comparison to efficiency reached using conjugation. Also, while biolistics and electroporation transformations result in similar transformation efficiencies, biolistics was widely adopted by diatom laboratories prior to development of electroporation.

### 1.1.2 Heterologous Gene Expression

Genetic engineering, as a practice, relies on the ability to introduce and express a gene or genes in a cell line in order to encode and confer a desired function. Heterologous gene expression refers to expression of a gene of a different relation to the cell. The heterologous gene may be an additional or modified copy of an endogenous gene or an exogenous gene that functions within the host cell. To express a heterologous gene or a transgene in a diatom cell line, researchers again looked to the nucleotide-level information generated by whole genome sequencing to find diatom-specific gene-expression components and tools.

### 1.1.2.1 Gene promoters and terminators

Like any gene expression system in an organism, diatom transcriptional elements such as promoters and terminators were identified and characterized to achieve heterologous gene expression. A promoter contains sequence-motifs that permit binding of proteins that control the transcriptional rate of the gene. In all cells, the enzyme RNA polymerase synthesizes a copy of a

gene (DNA) into RNA by following the gene sequence from it's promoter to it's terminator. With a functional promoter and terminator pair, RNA polymerase can in theory transcribe any gene that resides in the middle. Therefore, to express heterologous genes in diatom cell lines, diatom-specific promoters and terminators are required.

The first example of heterologous gene expression in diatoms was conducted in *Cyclotella cryptica* and *Navicula saprophila* (Dunahay et al. 1995). Though this study occurred prior to genome sequencing, putative promoter and terminator sequences that drive expression of the acyl-CoA carboxylase gene in *Cyclotella cryptica* were identified and used. The promoter and terminator were cloned into a vector to express an *E. coli* neomycin phosphotransferase II gene (NptII) that confers resistance to the antibiotic G418. Here, three components of the genetic engineering toolbox were demonstrated in diatoms; genetic transformation, heterologous gene expression, and antibiotic resistance.

In the past 25 years, diatom genetic engineering has become a staple practice in diatom research. Notably, the diatoms *Phaeodactylum* and *Thalassiosira pseudonana* have emerged as model diatom species for the development of molecular tools. Heterologous gene expression was harnessed to successfully transform the diatoms *Phaeodactylum* (Apt et al. 1996) and *Thalassiosira pseudonana* (Poulsen et al. 2006). Since, many promoters have been characterized and used for genetic engineering in these two species because of the availability of high quality sequenced genomes and assemblies (Armbrust et al. 2004, Bowler et al. 2008).

Promoters are classified into two functional groups: constitutive promoters that are continuously or nearly continuously  expressing the gene product and inducible promoters that switch gene expression on or off in response to an external cue. The first studies of heterologous gene expression in diatom used constitutive promoters to continuously drive expression of the

transgene. In both *Phaeodactylum* and *Thalassiosira pseudonana*, a class of endogenous promoters were characterized that drive the expression of genes in the fucoxanthin chlorophyll-a/c-binding protein (Fcp) class (Apt et al. 1996, Poulsen et al. 2006). Fcp promoters are nearly constitutive promoters that most strongly drive expression during illuminated conditions but, in practical terms, are always "on" in the absence or prolonged darkness. Multiple constitutive and endogenous promoters from *Phaeodactylum* have been functionally confirmed, including promoters from a histone gene (H4), an elongation factor II gene (Ef2), five Fcp promoters, (FcpA, B, C, E, F), and a V-ATPase C gene promoter. Additional, new stable and constitutive endogenous promoters from *Phaeodactylum* have also been found using multi-experiment transcriptomic data sets that were directly compared to the activity of FcpB when driving expression of YFP from an episomal vector (Windhagauer et al. 2021). Lastly, five promoters that were shown to function in *Phaeodactylum* were cloned from two diatom-infecting, single-stranded DNA viruses with one promoter, CIP1, driving higher rates of gene expression compared to genomic promoters like FcpA. Similarly, viral promoters such as the cytomegalovirus (CMV) promoter have been shown to function in *Thalassiosira pseudonana*. See Table 4 in Falciatore et al. (2020) for review of all promoters used, to date, in *Phaeodactylum* and *Thalassiosira pseudonana*.

Inducible promoters provide researchers the ability to turn gene expression on and off. In diatoms, inducible gene expression may be useful for the controlled expression of toxic proteins or for expression of reporter genes during specific cell conditions. The inducible promoter pNR (nitrate reductase gene promoter) has been characterized to increase gene expression in the presence of nitrate and to repress gene expression in the presence of ammonium in the diatoms *Cylindrotheca fusiformis, Thalassiosira pseudonana,* and *Phaeodactylum* (Poulsen & Kroger

2005, Poulsen et al. 2006, Niu et al. 2011). Additional promoters for *Phaeodactylum* have been characterized to induce gene expression in conditions of high-$CO_2$ (pCA1), iron limitation (Fbp1, Fld, and Isi1), phosphate limitation (pPhAP1), and lipid biosynthesis (pDGAT1) (for review: Falciatore et al. 2020). Two inducible promoters for silicon depletion responsiveness have been characterized for *Thalassiosira pseudonana* (TpSIT1 and TpSIT2) as well as one for *Cyclotella cryptica* (CcSIT1) (Shrestha &Hildebrand 2017).

Some examples of the diatom promoters described above were initially cloned from one diatom and then transformed into a second, such as pNR from *Cylindrotheca fusiformis* transformed into *Phaeodactylum*. Whether endogenous promoters for heterologous promoters are more effective may depend on the species and the promoter at hand.

### 1.1.2.2 Selectable markers

The selection of successfully transformed diatom cells after transformation depends on the effectiveness of a selectable marker introduced with the nucleic acid payload. In general, antibiotic selection is essential in providing selective pressure to a cell culture that results in the survival or death of cells depending on viable introduction of a gene that confers resistance (selectable marker). Antibiotics are also commonly used to maintain clonal diatom cell lines post-transformation (Huang & Daboussi 2017; Falciatore et al. 2020).

In addition to NptII, five additional antibiotic selectable marker genes have been developed for *Phaeodactylum* including Shble (phleomycin resistance), Nat (nourseothricin resistance), Bsr (blasticidin-S resistance), Cat (chloramphenicol resistance), and Sat (streptothricin resistance) (Falciatore et al. 2020). *Phaeodactylum* has emerged as a commonly-used and popular laboratory diatom due to the availability of a high quality

sequenced genome, well-developed molecular tools, and comparative ease-of-use compared to other pennate diatom species. The number of selectable markers developed for *Phaeodactylum* can be partially attributed to the amount of molecular-based research accomplished with the species. More functional antibiotics add versatility and options to molecular experiments in *Phaeodactylum*. Nonetheless, *Phaeodactylum* is not the only diatom species with multiple functional antibiotics. *Thalassiosira pseudonana* has been shown to harbor resistance to nourseothricin using the Nat gene and to phleomycin with the Shble gene (Falciatore et al. 2020). Other diatoms have only one demonstrated antibiotic resistance marker such as *Fistulifera solaris* (G418 resistance, Muto et al. 2015), *Cyclotella cryptica* (G418 resistance, Dunahay et al. 1995), *Navicula saprophila* (G418 resistance, Dunahay et al. 1995), *Cylindrotheca fusiformis* (phleomycin resistance, Fischer et al. 1999), *Chaetoceros gracilis* (nourseothricin resistance, Ifuku et al. 2015), *Amphora coffeaeformis* (nourseothricin resistance, Buhmann et al. 2014), *Skeletomena marinoi* (zeocin resistance, Johannson et al. 2019) *Pseudonitzschia arenysensis* and *Pseudonitzschia multistrata* (phleomycin resistance, Sabatino et al. 2015).

Chemical-based selectable markers have recently been applied in Serif et al. (2018) for plasmid-free targeted mutagenesis where the inactivation of the UMPS and APT genes in *Phaeodactylum* results in resistance to 5-fluoroorotic acid and 2-fluoroadenine, respectively. This represents the first trans-gene free applications of a selectable marker in a diatom. Similarly, an additional antibiotic-free selectable marker was developed for *Phaeodactylum* that confers resistance to the herbicide norfluorason in conjunction with a point mutation of the phytoene desaturase I gene (Taparia et al. 2019).

### 1.1.2.3 Reporter genes

The ability to select for transformed diatom cells is necessary for generating and maintaining clonal cell lines. Reporter genes, however, provide quantitative and qualitative subcellular information about the gene, promoter, or protein being investigated (refer to Falciatore et al. 2020 for review of reporter genes in *Phaeodactylum* and *Thalassiosira pseudonana*). As mentioned previously, the GUS reporter gene system was used for antibiotic-free cell selection (Falciatore et al. 1999). The GUS enzyme hybridizes the β-glucuronidase substrate and X-gluc by oxidation dimerization to produce a blue reporter molecule. The intensity of blue or amount of reporter molecule is directly proportional to the strength of the promoter driving the expression of GUS: a stronger promoter produces a more blue cell and a weaker promoter produces a less blue cell. Falciatore et al. (1999) also co-transformed an antibiotic resistant gene and the firefly luciferase LUC gene into *Phaeodactylum* and produced luminescent diatom cells detectable with a standard plate reader.

Fluorescent proteins are a versatile reporter that can provide detailed information about protein localization and temporal protein dynamics. Poulsen and Kroger 2005 expanded fluorescent proteins as reporter proteins using eGFP to determine the timing and intensity of mRNA and protein induction by the pNR promoter in *Cylindrotheca fusiformis*. Fluorescent proteins, such as eGFP, CFP, YFP, and mCherry, have since been regularly used in diatom transformation (Falciatore et al 2020). Now, many diatoms are able to express fluorescent proteins as reporter molecules including *Cylindrotheca fusiformis* (Poulsen et al. 2005), *Phaeodactylum* (De Riso et al. 2009), *Thalassiosira pseudonana* (Poulsen et al. 2006), *Pseudonitzschia arenysensis* (Sabatino et al. 2015), *Chaetoceros gracilis* (Ifuku et al. 2015), *Amphora coffeaeformis* (Buhmann et al. 2014), and *Fistulifera solaris* (Muto et al 2015).

### 1.1.2.4 Molecular cloning

Molecular cloning is the joining of multiple linear DNA fragments into a circular DNA plasmid to form functional genetic units. Functional genetic units can be gene expression cassettes, protein binding sites, centromeric regions, etc. The DNA plasmid is then introduced into and harbored within a host cell. A cell that successfully hosts the plasmid can be isolated as a clonal cell line. Dunahay et al. (1995) first functionalized a plasmid and the genetic unit (antibiotic resistance gene expression cassette for NptII) to produce clonal diatom cell lines resistant to the G418 antibiotic. In this study restriction enzyme, ligation-based cloning was used to stitch linear DNA into plasmids.

Later, a two-plasmid diatom cloning method was developed based on Gateway cloning methodology (Walhout et al. 2000). Topoisomerase-based cloning, known as Gateway cloning, was utilized to clone a gene into a pENT vector (Invitrogen, Carlsbad, CA) that led to the recombination-based cloning of the *Phaeodactylum*-specific destination vector, pDEST, that were built with a variety of promoters and fluorescent protein reporter genes for the expression of tagged-transgenes (Siaut et al. 2007). A gateway-based cloning for *Thalassiosira pseudonana* expression vector (pTpDEST-C'YFP) has also been developed that accepts recombination from pENT vectors (Karas et al. 2015).

Molecular cloning was made easier and faster in almost all laboratory organisms with the development of Golden-Gate (GG) (Engler et al. 2008) and Gibson Assembly (GA) cloning (Gibson et al. 2009). GG cloning is similar to restriction enzyme cloning but utilizes typeII-S restriction enzyme that cleaves DNA away from its binding site and leaves a 4-nt overhang at each 5' and 3' fragment end that can be designed by the researcher and used to orderly stitch a

plasmid together. For example in diatoms, a CRISPR-Cas9 plasmid was assembled using GG because multiple gene expression cassettes with the same promoter and terminator were cloned simultaneously into one plasmid for *Thalassiosira pseudonana* and *Phaeodactylum* (Hopes et al. 2016, Moosburner et al. 2020). GA cloning relies on producing homologous 5' and 3' ends of the DNA fragments via polymerase-chain reaction (PCR). Compared to GG, GA cloning is more commonly used in diatom cloning experiments due to its relative simplicity in primer design and isothermal assembly protocol. In diatoms, GA cloning has been used to assemble plasmids for a variety of functions such as fluorescent protein tags (McQuaid et al. 2018) and genome engineering (Daboussi et al. 2014, Weyman et al. 2015, Nymark et al. 2016, Stukenberg et al. 2018). Alternatively to GA assembly, NEB HiFi Assembly method was used to construct CRISPR-Cas9 episomes (Sharma et al. 2018)

Recently, a type-IIS-based cloning method was developed to aid in multi-level assembly for eukaryotic synthetic gene pathways called Universal Loop Assembly (uLoop) (Pollak et al. 2020). Here, uLoop produced a library of *Phaeodactylum* episomes of multiple fluorescent proteins with different promoters and signal peptides.

For all cloning methods mentioned above, competent *E. coli* cells were used to harbor assembled plasmids. Competent yeast cells can also be used to assemble plasmids and more notably large plasmids. In Karas et al. (2013), large chromosome fragments (> 150kb) from *Phaeodactylum* were successfully cloned in *S. cerevisiae* using a yeast origin or replication before being returned to the diatom host.

### 1.1.3 Genetic engineering

The development of molecular cloning methods and characterization of functional genetic elements in diatoms (heterologous gene expression, promoters, etc.) has led researchers to control the expression of endogenous and exogenous genes for specific purposes. The ability to change the genetic output of a cell is known as genetic engineering. Two approaches to genetic engineering experiments are forward genetics and reverse genetics.

Forward genetics consists of generating random mutations by using DNA mutagens (UV rays, ionizing radiation, ethidium bromide etc.) or by inserting exogenous DNA with a selection marker. Then, it is necessary to analyze a large number of cells in order to first find a specific phenotype and then identify the gene mutation associated with the phenotype. Although this approach has been widely used *in Chlamydomonas reinhardtii* to isolate mutants affected in many different biological processes such as photosynthesis and photooxidative stress, nitrogen metabolism, lipid production, carbon metabolism or $H_2$ production, it has not been widely adopted in diatoms. One reason is because they are diploid which complicates forward genetics applications. Diatom screening methods for induced mutations by nuclease activity is reviewed below in section 1.4 of this chapter.

Most studies have used DNA mutagens to increase the productivity of compounds (e.g., lipids, carotenoids) but have not linked the observed phenotype to a particular genotype. Only few studies have reported identification of mutations induced by DNA mutagens and linked them to a specific phenotype such as the mutation in the UMPS gene in *Phaeodactylum* conferring resistance to 5-fluoroorotic acid (Sakaguchi et al. 2011) and the mutation in a phytoene desaturase gene in *Phaeodactylum* conferring resistance to the herbicidal agent norflurazon (Taparia et al. 2019). Centromeric elements in *Phaeodactylum* were also found using a forward

genetics approach by introducing episomal vectors with large chromosomal fragments from *Phaeodactylum* and selecting for successful episomal maintenance (Diner et al. 2017).

Reverse genetics experiments aim to change the transcription or translation of a single gene (gene over-expression, gene silencing, gene mutagenesis) and can result in a cellular phenotype to be experimentally examined. Both forward and reverse genetics practices can result in functional characterization of a single gene. In diatoms, reverse genetics practices have been preferentially implemented due to the existence of many reverse genetics tools in other laboratory organisms that were translated to diatom genetic engineering. This chapter will focus exclusively on the current state and promising prospects of reverse genetics.

### 1.1.3.1 Gene over-expression

One method to investigate the function of a single gene is to transform a cell with and express an additional copy of the gene. Expressing an additional copy of the gene can result in a cellular phenotype that may be attributed to the function of the gene. Furthermore, by expressing a gene using a non-native promoter, one can increase the expression of the gene and decouple the introduced gene from the genetic regulation of the endogenous promoter and terminator. By this technique, a researcher can increase the expression of a gene beyond the endogenous levels, otherwise known as over-expression.

In diatoms, Dunahay et al. (1995) first reported the introduction of a second copy of the native ACCase gene in *Cyclotella cryptica* with the goal of increasing the production of the ACCase protein and increasing intercellular lipid content. Though this study did not report increased lipid content, it did report the increase in ACCase protein production in *Cyclotella cryptica* cell lines. Apt et al. (1996) characterized the gene expression profile of five fcp

promoters. Presently, FcpB is a commonly-used diatom promoter that has reported high-levels of transgene expression in *Phaeodactylum* (Falciatore et al. 1999, Siaut et al. 2007) and *Thalassiosira pseudonana* (Poulsen et al. 2006).

Nonetheless, gene over-expression should be limited to the increased expression of the introduced additional gene copy. First, the gene-of-interest (GOI) is cloned from the diatom or other host via PCR. The GOI sequence can then be cloned into an appropriate vector for over-expression that often is regulated by a strong promoter. Siaut et al. (2007) created the first diatom cloning kit for gene over-expression in *Phaeodactylum* with the vector pDEST-OX that drove the expression of the GOI using the FcpB promoter and FcpA terminator. Subsequently, other promoters have also been used in over-expression experiments (refer to section 1.2.1. Gene promoters and terminators). Gene expression vectors have also been developed for *Thalassiosira pseudonana* for constitutive and inducible gene expression (Poulsen et al. 2006, Shrestha and Hildebrand 2017). In Shrestha & Hildebrand (2017), recombinant protein production was increased in *Thalassiosira pseudonana* when the N-terminal sequence of the transgene was optimized and when the cells were exposed to photosynthetic inhibitors.

Validation of gene over-expression can be accomplished using multiple molecular methods. First, diatom clones with the correct over-expression cassette can be determined by colony PCR and sequencing of the cassette in multiple diatom clones. The expression profile of the over-expression cassette can be quantified at the transcriptional level by reverse-transcriptase-quantitative-PCR by designing specific primers to PCR amplify the cDNA sequence of the transgene. The expression profile can also be quantified at the protein level by Western Blot and, more effectively, ELISA assays. For all quantification methods, the expression

level of the transgene is compared to the expression level of a wild-type diatom (Siaut et al. 2007).

### 1.1.3.2 Gene silencing

Functional genetics studies have been greatly aided by gene silencing technologies like antisense RNA, RNA interference (RNAi), and silencing micro RNAs (miRNAs) that inhibits the translation of a specific coding transcript or promotes mRNA degradation and results in a decrease of protein production (Cerutti & Casas-Mollano 2006). A cellular phenotype can arise and may be attributed to the decreased protein levels and may infer the function of the GOI. Like gene over-expression determination, the resulting depression of mRNA translation can be quantified by protein-level assays.

Antisense technology functions by expressing a DNA construct so the transcribed RNA product complements a specific mRNA that inhibits translation of the mRNA. Expression cassettes with common diatom promoters are used to express complementary, or antisense, DNA. Silencing RNA technology was first demonstrated in the model *Phaeodactylum* by knocking down the GUS reporter and the endogenous cryptochrome gene involved in blue light acclimation using antisense RNA technology (De Riso et al. 2009).

RNAi technology accomplishes the same task of depressing the translation of a specific mRNA product as antisense technology but differs mechanistically. In RNAi experiments, a double-stranded RNA product is transcribed that is processed into single-stranded RNA by the RISC protein which allows the complementary binding to a mRNA product that activates the DICER protein to degrade the mRNA product. Here, the double-strand RNA is produced by the expression of inverted repeats that complement each other after transcription. Like antisense

DNA, the RNAi inverted repeats must be specifically designed to only complement one mRNA product. The degradation of mRNA results in a gene knock-down that can be quantified by protein-level assays. De Riso et al. (2009) first used RNAi in *Phaeodactylum* to knock-down the GUS reporter and cryptochrome gene. RNAi has been further adapted for utilization in *Thalassiosira pseudonana* (Trentacoste et al. 2013). RNAi has also been used to obtain fundamental insights into molecular mechanisms that control defining aspects of diatom metabolism and physiology, including dissipation of excess light energy through non photochemical quenching (Bailleul et al., 2010), the diatom urea cycle (Allen et al. 2011), and cell cycle progression in response to light sensing (Huysman et al. 2013).

Lastly, the RNA silencing toolkit was expanded by the development of artificial micro-RNAs (amiRNAs) to knock down the phytoene synthase gene and demonstrated its function in carotenoid production in *Phaeodactylum*  (Kaur et al. 2015).


### 1.1.3.3 Targeted gene mutagenesis

In recent decades, in vivo genome editing methods based on the use of site-specific nucleases have been developed for various organisms. Three classes of sequence-specific nucleases, Meganucleases (MNs), Transcriptional Activator-Like Effectors Nucleases (TALENs), Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR-Cas9), have been used to induce targeted modifications in diatom genomes (Table 1.2). These nucleases have been derived from natural molecules and functionally adapted to create powerful molecular scissors suitable for introduction of double-strand DNA-breaks (DSB) at precise genomic loci (For a review, see Huang & Daboussi 2017). Meganucleases are engineered for genome editing by introducing a small variation into their DNA-binding domain in order to allow customizable

MNs able to recognize and cleave specific targeted sequences. TALENs comprise a non-specific Fok1 endonuclease domain fused to a customizable DNA-binding domain. The TALE DNA-binding domain is composed of 14-24 repeated units of 33-35 amino acids nearly identical to each other, except for two polymorphic amino acids which are responsible for specific recognition of a single targeted nucleotide. Based on these features, a TALE protein can be customized as a DNA-binding domain able to recognize any nucleotide sequence. Finally, the CRISPR-Cas9 system is based on the homing endonuclease activity and designer small RNA molecules, called single-guide RNAs (sgRNAs), which are derived from bacterial and archaea systems for adaptive immunity against invading nucleic acids. The first Cas9 system engineered as a gene editing tool included a Cas9 endonuclease which contains two nucleases domains, RuvC and HNH, able to produce a blunt-ended DSB at the target DNA locus. It additionally included a sgRNA which carries a sequence of 20 nucleotides that defines the genomic target to be modified. Thus, unlike MNs and TALENs for which it is necessary to design new proteins to recognize each new target site, the target site recognition in the CRISPR-Cas9 system is entirely sgRNA-mediated and makes CRISPR-Cas9 a more flexible and user-friendly platform for genome editing. Refer to Table 1.2 for gene editing and mutagenesis efficiencies for each type of molecular scissor.

**Table 1.2** Targeted mutagenesis capabilities in the diatoms *Phaeodactylum* and *Thalassiosira pseudonana*. Each row represents a study where column two shows the targeted mutagenesis tool used, column three shows the genetic transformation used (RNP: ribonucleoprotein complex), column four shows the application used (mutagenesis: single gene NHEJ-mediated; multiplex mutagenesis: multiple gene NHEJ-mediated; HDR editing: homology-directed repair editing), column five shows the mutagenesis frequency achieved as a percentage of transformed cells that contain bi-allelic mutations of the target gene(s), and column six shows the reference article.

| Species | Molecular Scissor | Genetic transformation | Application | Mutagenesis Frequency | References |
|---|---|---|---|---|---|
| *Phaeodactylum tricornutum* | Meganuclease | Biolistics | Mutagenesis | 0.032 - 6% | Daboussi et al. 2014 |
| | TALEN | Biolistics | Mutagenesis | 7 - 56% | Daboussi et al. 2014 |
| | TALEN | Biolistics | HDR editing | 18% | Weyman et al. 2015 |
| | CRISPR- Cas9 | Biolistics | Mutagenesis | 31% | Nymark et al. 2016 |
| | | Biolistics | HDR editing | 17% | Moosburner et al. 2020 |
| | | Episome | Mutagenesis | 60% | Slattery et al. 2018 |
| | | Episome | Multiplex mutagenesis | 25 - 48% | Moosburner et al. 2020 |
| | | RNP | Multiplex mutagenesis | 65 - 100% | Serif et al. 2018 |
| *Thalassiosira pseudonana* | CRISPR-Cas9 | Biolistics | Mutagenesis | 61.5% | Hopes et al. 2016 |

Researchers faced notable challenges when first implementing gene editing tools in diatoms. First, it is difficult to deliver DNA through the complex diatom cell wall with a biolistics-based transformation (refer to section 1.1.1. Biolistics). Second, diatom genomes are diploid and it is therefore necessary to inactivate both target gene alleles to functionally KO a gene.

In order to effectively harness modern genome editing methodology, it is important to understand what types of genome modifications result from the introduction of DSBs. One major mechanism involved in the repair of a DSB is non-homologous end-joining (NHEJ), which

directly ligates the ends of cut DNA. The ligation is often imprecise and introduces deletions or insertions of nucleotides at the break site, leading to consideration of NHEJ as an error-prone mechanism. Secondly, homology-directed repair (HDR) drives gene insertion or gene replacement using exogenous DNA templates with homology to the target locus by harnessing the innate cellular homologous recombination (HR) DNA repair mechanism. The spontaneous HR mechanism is known to occur at various frequencies that are organism specific (Heyer et al. 2010). However, it has been shown that HR frequency can be increased by 1000-fold following the introduction of a DSB by a programmable nuclease in the presence of a DNA template (Richardson et al. 1998). It appears that introduction of a DSB by specific nucleases can lead to the inactivation of a target gene via NHEJ or HDR.

### 1.1.3.3.1 Mega-nucleases

The first proof of concept for targeted and stable modifications of the *Phaeodactylum* genome resulted from work conducted at the Cellectis SA Company (Daboussi et al. 2014; Duchateau & Daboussi 2013). Given the low transformation efficiency achieved in diatoms by the biolistics method, MN-encoding constructs, which can be delivered as a single chain molecule, were initially introduced. With this approach, it is possible to assess gene inactivation by NHEJ using two MNs targeting endogenous genes. Diatom cells were co-transformed with a MN-encoding plasmid and a selection cassette plasmid expressing the Nat gene conferring resistance to nourseothricin. Targeted mutagenesis induced by MN was 0.032%, which is very low (Duchateau & Daboussi 2014). In order to increase the efficiency of targeted mutagenesis, a plasmid encoding the DNA processing enzyme scTREX2 exonuclease, which was previously shown to increase targeted mutagenesis frequency up to 10-fold in mammalian cells and plants

(Delacote et al. 2013), was co-delivered with the MN-encoding plasmid and the plasmid carrying the *Nat* resistance gene. Sequencing results provided clear evidence that MNs in combination with scTREX2 greatly increased frequencies of targeted mutagenesis up to 6% (Daboussi et al. 2014).

### 1.1.3.3.2 TALEN

The production of MNs was time-consuming and not sufficiently flexible compared to TALENs, which have been shown to be highly efficient in various organisms. The simple one-to-one correspondence between the TALE repeats and the corresponding DNA sequence makes the process of assembling repeat arrays to recognize specific DNA sequences straightforward. To assess the capacity of TALENs to efficiently induce targeted mutagenesis, seven TALENs targeting genes potentially affecting lipid metabolism were designed, synthesized, and delivered into *Phaeodactylum* cells (Daboussi et al. 2014). This study demonstrated the robustness of the protocol established as all seven genes were independently mutated using TALEN technology.

Several technical improvements have been made in order to make the design easier, the TALEN synthesis more cost-effective, and the genome editing more efficient (Serif et al. 2017, Kroth et al. 2018). Using this protocol, KO strains for blue light aureochrome photoreceptors were generated in *Phaeodactylum* (Mann et al. 2017).

The co-delivery of site-specific nucleases and a donor DNA containing sequences homologous to those flanking the DSB site and a selectable marker has been shown to greatly increase the chance of integration of transgenes in plants or mammalian cells. The first study on HDR-mediated genome editing was conducted in *Phaeodactylum* using MNs and a donor

template which consisted of two homologous regions of 750 bp each with sequence identity to the targeted locus, flanking a 29-bp insertion (Daboussi et al. 2014). A new DNA template design was implemented by Weyman et al. (2015) based on the presence of an antibiotic resistance cassette flanked by sequences homologous to both sides of the TALEN target cut site. This study aimed to elucidate the role of the urease gene product in diatoms. Both the Urease-TALEN plasmid and the KO cassette, flanked by homologous sequence to the target site (donor template), were co-delivered in *Phaeodactylum* cells. While TALENs are less costly and time-consuming to engineer relative to MNs, their use has been limited to a few research groups. Despite this TALENs have been instrumental for generating KO strains for key enzymes in diatom iron and nitrogen metabolism, including phytotransferrin and nitrate reductase, that have resulted in major insights into diatom iron and nitrogen metabolism (McCarthy et al., 2017, McQuaid et al., 2018). TALENs have also been used to inactivate a thioesterase gene in *Phaeodactylum* that inhibited beta-oxidation and illuminated the lipid accumulation capabilities of mutant *Phaeodactylum* (Hao et al. 2018).

### 1.1.3.3.3 CRISPR-Cas9

Due to its ease of use, multi-functionality, and molecular efficiency, CRISPR-Cas has emerged as a robust genetic engineering tool for many model organisms. CRISPR (clustered regularly interspaced short palindromic repeats) are genetic elements naturally found in many bacteria and some archaea species that orchestrate multi-enzyme immune defenses against invading foreign nucleic acids. Six CRISPR bacterial immune systems have been reported, however the type-II system has been preferentially optimized for genome engineering due to the simplicity of design and implementation. The CRISPR-associated protein 9 (Cas9) was

characterized as an homing endonuclease capable of generating targeted blunt double-stranded breaks (DSBs) guided to a nucleic acid target by a single-guide RNA (sgRNA) in mammalian cells (Cong et al. 2013). The sgRNA is a 100-nt RNA molecule that contains a 20-nt spacer sequence at the 5' end that is designed to complement the genetic target sequence. To change genetic targets, one simply reprograms the spacer sequence. The discovery of Cas9 represented a monumental finding that has revolutionized genome engineering and expanded targeted manipulations to the epigenome and transcriptome (Mali et al. 2013).

Since its inception, CRISPR-Cas9 has targeted the genome of a vast range of organisms, including marine diatoms. CRISPR-Cas9 targeted mutagenesis was first reported in *Phaeodactylum* (Nymark et al. 2016) and shortly later in *Thalassiosira pseudonana* (Hopes et al. 2016). These studies utilized the NHEJ-mediated mutagenesis pathway to inactivate targeted genes and proved Cas9 functionality and directed mutagenesis. CRISPR-Cas9 targeted mutagenesis also proved easier to implement than previously established MN and TALEN methodologies. For one, only the sgRNA molecule requires reprogramming to direct Cas9 DSBs compared to the synthesis of large proteins required for MNs and TALENs.

Implementation of CRISPR-Cas9 targeted mutagenesis was first accomplished in *Phaeodactylum* by delivering a single plasmid encoding both Cas9 and sgRNA using biolistics transformation (Nymark et al. 2016) with a second plasmid containing a phleomycin resistance gene, *Shble*. Cas9 was codon-optimized for *Phaeodactylum* and expressed using the *FcpB* promoter. Though codon optimized here, a human codon-optimized Cas9 was shown to function with similar efficiencies in *Thalassiosira pseudonana* (Hopes et al. 2016) and *Phaeodactylum* (Slattery et al. 2018, Moosburner et al. 2020). The sgRNA, unlike Cas9 and all protein-coding genes, required expression by a RNA-polymerase III class promoter (pol-III). Nymark et al.

(2016) found a *Phaeodactylum* pol-III promoter that contained homology to the well-characterized human U6 promoter. An additional pol-III promoter for *Phaeodactylum* named sRNAi was also characterized to drive sgRNA expression (Moosburner et al. 2020). Biolistics-based Cas9 delivery into *Phaeodactylum* was further optimized by delivering Cas9 and a sgRNA on a single plasmid (Stukenberg et al. 2018). Here, Cas9 activity was also inducible using the promoter pNR.

In all CRISPR-Cas9 experiments, spacer sequences must be designed to efficiently guide Cas9 to the genetic target (protospacer) for cleavage and to avoid guiding Cas9 to additional target sequences known as off-targets. In addition to the spacer sequence, a protospacer-adjacent-motif (PAM) sequence of 5'-NGG-3' that is directly downstream of the protospacer and on the opposite stand must also be satisfied for Cas9 cleavage. The NCBI-BLAST portal can provide the means to search for how many genomic loci can satisfy a protospacer-PAM design. Multiple methods have been developed to search for Cas9 target sequences (custom Perl script, Nymark et al. 2016; PhytoCRISPR-Ex website, Rastogi et al. 2016; CRISPOR, Serif et al. 2018; Benchling software, Stukenberg et al. 2018). Typically, if the goal of Cas9-based mutagenesis is to produce a gene KO, designing a sgRNA to target the 5' end of the gene is a strategy to introduce mutations at the beginning of the coding sequence that causes a codon frame-shift for the entire downstream portion of the gene target.

After sgRNA design, the sgRNA expression cassette, Cas9 expression cassette, and selectable marker is cloned into appropriate vectors for transformation. Nymark et al. (2016) transformed two plasmids into *Phaeodactylum* for targeted mutagenesis using biolistics: the pKS plasmid containing Cas9 and sgRNA expression cassette and the pAF6 (Falciatore et al. 1999) that contains a phleomycin resistance gene cassette. Also, Cas9 and the sgRNA cassette have

been cloned into the episomal plasmids PtpUC1 (Slattery et al. 2018) and PtpBR322 (Moosburner et al. 2020).

Alternative strategies to biolistics-based CRISPR-Cas9 mutagenesis in diatoms have been made. Off-target effects are always a concern in Cas9 experiments despite the development for in-silico programs like PhytoCRISP-Ex for sgRNA design. One strategy taken was the development of Cas9-nickase (D10A) target mutagenesis in *Thalassiosira pseudonana* (Nawaly et al. 2020). Here, two sgRNAs were designed to target an omega-carbonic anhydrase gene where two nicks were made at loci on opposite strands and produced NHEJ-mediated mutations between the nicks. Here, Cas9 nickase off-targeting was very ineffective because the likelihood of dual targeting at loci close together was very low.

### 1.1.3.4 Optimizing the delivery of CRISPR-Cas9

Although these tools enable exciting new possibilities to further understand the biology of diatoms, as well as to manipulate metabolic pathways for biotechnological applications, there are a number of challenges to be addressed in order to ensure accurate and safe genome editing. Until 2018, genome editing in diatoms has been mediated by biolistic transformation using a nuclease encoding plasmid and a plasmid carrying an antibiotic resistance gene. Plasmids stably integrate at random sites within the genome and can result in several potential problems: (i) the long-term expression of the nuclease: (ii) random integration of the delivered DNA into the genome; (iii) the transformation efficiency for biolistics remains relatively low (Table 1.1), (iv) marker genes cannot be eliminated through outcrossing since diatoms lack a haploid state like other microalgae. Thus, alternative genome editing strategies have been developed to circumvent

part or all these issues like episomal and ribonucleoprotein (RNP) delivery of CRISPR-Cas9 components.

### 1.1.3.4.1 Episomal CRISPR-Cas9

CRISPR-Cas9 gene mutagenesis in *Phaeodactylum* has been adapted for episomal expression by conjugation transformation (Slattery et al. 2018, Sharma et al. 2018, Moosburner et al. 2020). As an initial proof of concept, Slattery et al. (2018) targeted the *Phaeodactylum* urease gene using both a Cas9 episome and a TevCas9 dual nuclease episome. As mentioned above, sgRNA cloning in this study  was accomplished using two inverted BsaI restriction enzyme sites for GG cloning of spacer sequences targeting the urease gene. NEB HiFi Assembly was also used to clone spacer sequences into a Cas9 episome for *Phaeodactylum* (Sharma et al. 2018). In both studies, a low-copy version of the episomal vector, pPtPUC3 (Karas et al. 2015), was used to clone Cas9 and spacer sequences. Subsequently, a medium-copy episomal vector, pPtpBR322 (Diner et al. 2017), was used for hierarchical GG assembly cloning of multiple sgRNA expression cassettes (Moosburner et al. 2020). Here, an sgRNA expression vector encoded two inverted BbsI restriction enzyme sites flanking an *E. coli* LacZ expression cassette to permit blue-white colony screening where white colonies contained the cloned spacer sequence after GG assembly. The pPtpBR322 episome was used to clone Cas9 and an *E. coli* red-fluorescent-protein (RFP) expression cassette for negative selection that was flanked by two inverted BsaI sites. Next, one or more sgRNA expression cassettes and a LacZ cassette were cloned by GG to replace the RFP cassette to permit blue-red colony screening where *E. coli* with correctly assembled episomes were blue (Moosburner et al. 2020).

One caveat to episomal expression of any transgene, including Cas9, is that genetic

elements harbored on an episomal plasmid that are not selected for may be lost after conjugation (Karas et al. 2015). To circumvent this problem for CRISPR-Cas9 mutagenesis in *Phaeodactylum*, the Cas9 and Shble coding sequences were transcriptionally fused by P2A, a self-cleaving peptide linker, that ensured selection of Cas9 activity in all transformed cells resistant to phleomycin (Moosburner et al. 2020). This method was further implemented to obtain important insights into diatom siderophore acquisition by targeting ferric reductases in *Phaeodactylum* (Coale et al., 2019)

Lastly, it was demonstrated that a Cas9-episome can be "cured" from the diatom mutant cell line (Diner et al. 2016, Sharma et al. 2018). Episomes are retained in transformed cell lines by antibiotic selection. However, Sharma et al. (2018) showed that the episomal vector can be eliminated from diatom cells by removing the selection pressure. Such ability for scarless excision of gene editing and selection marker machinery could have major benefits for introduction of gene edits into diatoms without inclusion of foreign DNA and associated GMO designation.

### 1.1.3.4.2 Ribonucleoprotein delivery of Cas9-sgRNA

Another strategy to avoid off-targeting and to produce mutant diatom lines without transgenes was the delivery of ribo-nuceoprotein (RNP) complexes to *Phaeodactylum* (Serif et al. 2018). Delivering Cas9-sgRNA RNP complexes by biolistics resulted in the transient exposure to targeted mutagenesis and reduced the risk of off-target effects. Also, by not delivering transgenes for Cas9 or sgRNA, diatom mutant cell lines were established that cannot be classified as genetically-modified organisms; a sought-after goal in industrial and commercial microbial products. Serif et al. (2018) demonstrated the ability to deliver  RNP  without using a

plasmid encoding antibiotic resistance. To achieve this goal, they developed a strategy relying on simultaneous co-delivery of multiple RNP complexes, one targeting an endogenous gene for which inactivation confers positive selection (APT or UMPS) and the others targeting the blue-light photoreceptor Aureochrome1A gene in *Phaeodactylum*. Several dozen strains mutated in the APT or UMPS genes were generated and 65% to 100% of them showed mutations in the Aureochrome1A gene. This methodology was confirmed by creating strains in which all three genes were inactivated simultaneously (three-gene knockout) without introducing selection markers or DNA into the cells.

These results represent considerable advancements in targeted mutagenesis in diatoms in several key aspects: (i) improvement in the safety of genome engineering methodologies since vector components are not randomly integrated into the genome; (ii) transient Cas9 nuclease exposure to cells which limits potential off-targets; (iii) identification and validation of two new positive selection markers which avoid the use of antibiotic resistance gene cassettes.

### 1.1.3.4.3 Limitations of CRISPR-Cas9 in diatoms

Cas9-mediated HDR of the nitrate reductase gene in *Phaeodactylum* has been reported (Moosburner et al. 2020). Here, biolistics delivery of a Cas9 plasmid and a sgRNA plasmid was accompanied by a HDR-template plasmid. The HDR template plasmid, pKO-NR, was designed akin to the plasmid pTALEN-Ble (Weyman et al. 2014) where an Shble expression cassette was flanked by 1-kb homology arms complementary to the flanking genomic regions around the sgRNA cleavage locus. A stop-codon that prematurely ended nitrate reductase gene transcription also preceded the Shble cassette. This is similar to the TALEN-mediated HDR of the urease gene (Weyman et al. 2015). Nevertheless, random integration of the delivered transgenes persisted

here. HDR editing via Cas9 has yet to be reported using conjugation or RNP delivery, which would eliminate random integration. Regardless of the success of precise Cas9-based gene editing in *Phaeodactylum* (Moosburner et al. 2020), the random integration of transgenes still persisted. The development of HDR gene editing via conjugation or RNP delivery would aid in achieving the goal of producing precise gene edited cell lines without the genomic scars of random integration but has yet to be achieved.

While conjugation transformation increased the diatom transformation efficiency compared to biolistics and electroporation (Table 1.1), conjugation has only been reported in two diatom species, *Phaeodactylum* and *Thalassiosira pseudonana*, compared to ten species for biolistics (Faktorova et al. 2020). The episomal plasmid relies on a centromeric region (*CEN6-ARSH4-HIS3*) for cellular maintenance and stability. This region was isolated in *Phaeodactylum* through time consuming 'episome rescue' experiments (Karas et al. 2015). Interestingly, however, the yeast-derived *CEN6-ARSH4-HIS3* sequences are effective diatom sequences for episome maintenance in *Phaeodactylum* and *Thalassiosira pseudonana*, (Diner et al. 2017). It remains to be seen whether or not the yeast element will work effectively in other diatom species, or if additional centrometeric elements have to be found and characterized. If additional species-specific centromeric elements have to be found and characterized and a diatom researcher wishes to step away from *Phaeodactylum* and *Thalassiosira pseudonana*, they are limited to using biolistics for the time being. Furthermore, the episomal plasmid may limit 'episome rescue' and centromere isolation experiments that require large transgene cassettes that exceed 50kb (Karas et al. 2015). Also, with an increasingly larger episome, the loss of episomal element may increase the difficulty in producing a diatom transformant with a correctly maintained episome (Karas et al. 2015, Moosburner et al. 2020).

Lastly, RNP delivers transient Cas9-sgRNA. To employ diatoms as a synthetic biology platform, the ability to KO genes and introduce synthetic gene pathways is necessary. This could be accomplished by combining RNP and episomal delivery but not using RNP alone. Therefore, conjugation appears to be a more versatile synthetic biology platform for basic research applications in order to employ Cas9-based mutagenesis and synthetic gene pathways simultaneously.

**1.1.4 Diatom Cell Line Screening**

The genetic transformation of diatom cells via biolistics, electroporation, and conjugation typically delivers antibiotic resistance genes in order to select for diatom cell clones by antibiotic selective pressure. The transformed cells are spread on solid agar plates supplemented with appropriate antibiotic concentrations. Most of the cells spread on the plates die however the ones that recieved and express the antibiotic resistance genes live and form visible colonies between 10 and 30 days, depending on the transformation method. After genetic transformation of transgenes to diatom cells, the selection of which diatom cells line to evaluate is critical, particularly after targeted gene mutagenesis. It has been reported that colonies formed on the plates directly from transformation are not always clonal and can be mosaic, as in the resulting colony contains more than one diatom cell, each potentially containing unique allelic mutations induced by a targeting nuclease (MNs, TALENs, or Cas) (Weyman et al. 2015, Daboussi et al. 2014, Huang & Daboussi 2014, Sharma et al. 2018, Slattery et al. 2018, Serif et al. 2018). In order to bypass mosaicism in a diatom mutant, isolation and validation of clonal diatom cell lines has been achieved by subcloning methodology. A proposed cell line screening and selecting workflow following targeted mutagenesis in diatoms is shown in Figure 1.1.
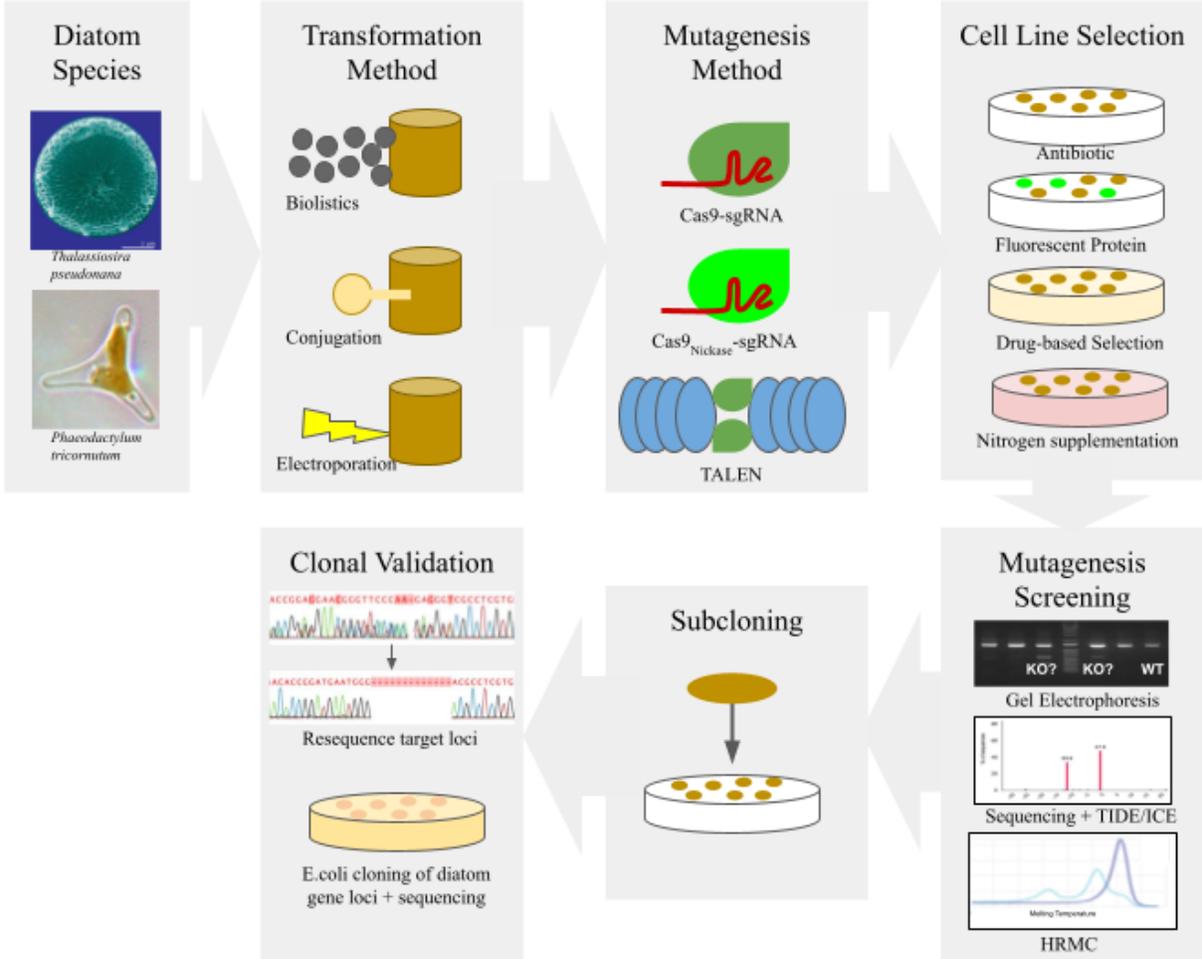
**Figure 1.1** A proposed workflow for targeted mutagenesis and cell line screening in diatoms. Two diatom species, *Thalassiosira pseudonana* and *Phaeodactylum* are highlighted as they have been used to generate most molecular biology and genetics data compared to other species. First, choose which of the three transformation methods (Table 1.1 and 1.2) to use. Next, choose how to target the gene of interest (CRISPR-Cas9, Cas9-Nickase, or TALENs). Decided on how to select for transformants (antibiotics, fluorescent protein, drug-based positive selection, or nitrogen supplementation growth assay). Lastly, determine whether the cell lines have a mutation at the target locus by PCR, sequencing, TIDE/ICE (tracking indels by deconvolution / interference of CRISPR edits), or HRMC. It is always recommended to subclone cell lines of interest and resequence diatom target loci or clone target loci into *E. coli* vector followed by sequencing to validate clonality.

### 1.1.4.1 Subcloning diatom transformants

To subclone a diatom colony, the colony is picked from the selection agar plate and resuspended in a small volume of liquid growth medium. The liquid culture is then re-plated on solid agar plates supplemented with antibiotics. Again, the resulting colonies are picked and validated for clonality. Typically, one round of subcloning is sufficient to produce clonal diatom cell lines as long as the colonies are not physically close (Weyman et al 2014, Nymark et al. 2016). Serial dilutions of the initial resuspended colony can usually result in spatially distant colonies.

Also, subcloning extends the time of targeted nuclease activity and can continue to mutate alleles if sufficient mutagenesis is not achieved during the first round of colony formation. This is due to the fact that DSBs do not systematically induce mutations. Indeed, it is observed that mutation frequency varies between recovered colonies and that colonies can comprise a mixed population of cells with or without mutations. This phenomenon suggested that mutagenesis does not necessarily occur within the initially selected cell, but rather can occur during subsequent cell divisions. Consequently, an additional subcloning step of isolating a single cell using limiting dilution methods is required to produce monoclonal transgenic cell lines. Thus, a sub-cloning step is necessary to identify clones mutated on both alleles (bi-allelic mutation). In the case where the mutation is present only on one allele (mono-allelic mutation), there is potential for new mutations after the subcloning step because the cell is still expressing the nuclease and is potentially still able to cut an unmutated target (wild type allele). Subcloning was and still is a recommended step in targeted mutagenesis experiments in diatoms to simply ensure clonal cell lines are produced (Figure 1.1).

**1.1.4.2 Targeted mutagenesis determination**

This problem of producing clonal diatom cell lines became most apparent when screening for genetic mutations induced by TALEN and CRISPR-Cas9 targeted mutagenesis (Weyman et al. 2014, Daboussi et al. 2014). Diatoms have a diploid genome; two copies of each chromosome and each gene. To successfully produce a biallelic gene KO, mutations have to be made on each copy, respectively. When screening for mutation in diatom transformants, it is expected that either one sequence mutation (homozygous mutagenesis; both alleles contain the same mutation) or two sequence mutations (heterozygous mutagenesis, the alleles contain unique mutations) will arise. Nevertheless, most diatom studies where targeted mutagenesis was done reported that some colonies contained more than two mutation types. This indicates that more than two gene allele mutations exist within the colonies and therefore the colonies were mosaic.

One method to validate that a diatom colony has a KO genotype is to sequence the gene from the diatom genome. In practice, the GOI is PCR amplified from either purified genomic DNA or directly from diatom cells. Paired-end Sanger sequencing is then performed to reveal the gene allele sequences that are either mutated or not and compared to that of a wild-type gene sequence. If a homozygous mutation is made at both gene alleles (in a truly clonal cell line) then the sequencing electropherogram will contain only one peak for every nucleotide and the paired sequencing run should reflect that. However, if a heterozygous mutation is made (in a clonal cell line) then the sequencing electropherogram will contain more than one peak for the nucleotides downstream of the predicted Cas9 cleavage site. In this instance, the cell line may be clonal but Sanger sequencing alone is not sufficient enough to make such a determination. For a mosaic colony, the resulting electropherogram may contain more than two peaks per nucleotide after the Cas9 cleavage site. Using Sanger sequencing, therefore, cannot alone distinguish between

heterozygous mutations in a clonal cell line and a mosaic colony and further validation is needed. Nymark et al. (2016) highlights the variety of electropherograms of sequenced loci produced by Cas9-mediated mutagenesis of the CpSRP54 gene in *Phaeodactylum* (Figure 1 in Nymark et al. 2016).

One method to resolve this issue is to PCR amplify the GOI and clone the sequences into a vector for *E. coli* cloning and transformation. The resulting *E. coli* colonies will only contain one vector each that harbors one allele's sequence. Again, Sanger sequencing can be used to sequence the GOI but the resulting electropherograms will only result in one peak per nucleotide. By sequencing multiple *E. coli* colonies, the genotype of the diatom colony in question can be determined as homozygous (one sequence in the *E. coli* colonies), heterozygous (two sequences in the *E. coli* colonies) or mosaic (more than two sequences in the *E. coli* colonies).

A second method to determine mutation genotypes is using the T7-endonuclease I (T7E1) assay (Huang et al. 2012, Bloom et al. 2017). The PCR amplified GOI here is first subjected to a denaturing and annealing cycle. A slow annealing cycle is used to force mismatched base-pairs to form which would occur if the resulting genotype is heterozygous. The T7E1 enzyme specifically cleaves mismatched dsDNA and therefore a heterozygous amplicon would be cleavage. Gel electrophoresis is then used to size separate the DNA products where T7E1-cleaved DNA would be shorter than uncleaved DNA. In diatoms, T7E1 has been used to identify Cas9-mediated heterozygous mutants in individual diatom colonies (Slattery et al. 2018) and also to compared the efficiency of sgRNAs by Cas9-mediated mutagenesis for all resulting diatom colonies (Moosburner et al. 2020). Unlike *E. coli* subcloning of alleles, this method cannot distinguish between heterozygous genotypes and mosaic genotypes and diatom subcloning is necessary before performing the T7E1 assay.

A third diatom genotyping method is high-resolution melt curve (HRMC) analysis that uses qPCR methodology to compare wild-type sequences to mutant sequences at single-nucleotide resolution. Here, GOI amplicons are stained with a saturating DNA stain and slowly denatured. A qPCR machine can distinguish between the temperature at which each nucleotide (A, T, C, and G) denatures and can therefore determine if a wild-type sequence is different from a mutant sequence, even one that has a single nucleotide polymorphism. In diatoms, HRMC analysis has been used to identify Cas9-mediated mutant genotypes after biolistics (Nymark et al. 2016) and after conjugation (Sharma et al. 2018).

Lastly, web-based software has been specifically designed to deconvolve electropherograms that contain more than one peak per nucleotide resulting from targeted mutagenesis. Two software packages, TIDE (Tracking Indel by DEconvolution[1], Brinkman et al. 2014) and ICE (Interference of CRISPR Edits[2]) are publicly available. Both packages can be used to separate electropherogram peaks to determine how many sequences (one for homozygous, two for heterozygous, three for mosaic) are represented. In diatoms, TIDE analysis has been used as a screening tool to identify homozygous and heterozygous mutants prior to colony subcloning (Sentmanat et al. 2018, Moosburner et al. 2020).

Certain genes in a diatom's genome have particular functions that, when mutated, arrest a specific metabolic function that leads to cell death and can be used for negative selection to screen for successful targeted mutagenesis events. For one, the nitrate reductase gene is the sole protein in many diatom species that reduces nitrate to nitrite. When mutated or knocked out, nitrate cannot be metabolized and the cell dies if nitrate is the only nitrogenous source supplemented in the growth medium (McCarthy et al. 2017). By induced targeted mutagenesis of

---

[1] https://tide.nki.nl/
[2] https://ice.synthego.com/

the nitrate reductase gene, a simple phenotype screen can be administered to the diatom cell line to determine if the nitrate reductase enzyme is active or not. In Phaeodactylum, this nitrate growth screen was used to show cell death on nitrate and cell viability on ammonium after TALEN-mediated (McCarthy et al. 2017) and Cas9-mediated (Moosburner et al. 2020) mutagenesis of nitrate reductase. A second diatom gene, urease, has also been used as a genomic target to screen for cell viability. Urease, similarly, is the sole enzyme responsible for the hydrolysis of urea to produce ammonium for amino acid biosynthesis. If urease is mutated or KOed, the diatom cell diatom when supplemented with urea as the only nitrogenous source in the growth medium. Urease mutagenesis was harnessed for the proof-of-concept for targeted mutagenesis in *Phaeodactylum* for TALEN-based targeting (Weyman et al. 2014) and in *Thalassiosira pseudonana* for Cas9-based targeting (Hopes et al. 2016).

Though not a nitrogen-based assay, Serif et al. 2018 also utilized the mutagenesis of PtUMP in *Phaeodactylum* that confers resistance to 5-fluoroorotic acid as a positive selection method. A Cas9-sgRNA RNP for APT was co-delivered with a second RNP for a target gene and showed 65% to 100% cell lines resistant to 5-fluoroorotic acid also contained targeted mutation of the second gene target. Here, a positive selection method based on the selective pressure of a herbicidal drug can be an attractive alternative to nitrogen-based screening mentioned above.  All things considered, a mosaic diatom colony where all alleles present have mutations in either the nitrate reductase, urease gene, or APT can still produce the KO phenotype mentioned above. Before administering these phenotypic screens, it is recommended to subclone the diatom colonies.

### 1.1.5 Future Directions and Perspectives

Advances in genetic engineering in diatoms have been rapid and effective. The development of genome editing marked a turning point for the diatom community. Researchers now have the tools they need to increase their knowledge of underlying metabolic and regulatory mechanisms in these potentially valuable organisms. Furthermore, extension of these tools to other diatom species will be of great use for basic and applied research. While this work seemed difficult to imagine a few years ago because of the difficulty of transforming some species or the need to identify transcriptional regulation elements, such as promoters and terminators for the expression of nucleases, hope has emerged with development of new genetic editing methodologies. Successful delivery of nuclease via an episomal system removes limitations, such as poor efficiency and random integration compared with biolistics transformation (Table 1.1). Also, the delivery of CRISPR-Cas9 and ribonucleoproteins offers the possibility to edit diatom species that can be cultured in the lab and that have an available genomic sequence.

While reverse genetics techniques developed to date have greatly aided in advancing foundational knowledge about key molecular mechanisms that underlie diatom biology and ecology, some key limitations exist that make various applications difficult. Most diatom studies have used one of two species, either *Phaeodactylum* or *Thalassiosira pseudonana*. While both species are easy to grow and maintain in the laboratory, there are key aspects of diatom biology, including sexual reproduction and interaction with viruses, that cannot be addressed with these models. The extent to which these models adequately represent the breadth of pennate and centric diatom biology remains unclear. Nevertheless, they continue to prove capable of providing the diatom community with key information on the evolutionary origins and metabolic

adaptations that led to ecological success. Importantly, with at least 30,000 extant diatom species, the majority have not been cultured in a laboratory setting let alone be genetically transformed and modified. With a wide diversity of species and potential metabolic features, a lack of established resources and protocols for genetic engineering may hinder studies into specialized features of diatom biology such as high-lipid content *(Cyclotella cryptica,; Fistulifera solaris)*, benthic life-style (*Fragilariopsis cylindrus*, *Seminavis robusta*), cold-adaptation (*Fragilariopsis cylindrus,*), chain formation (*Skeletomena marinoi*), and toxin biosynthesis (*Pseudonitzschia multiseries*) (for review on diatom features, see Falciatore et al. 2020). Examples of hurdles that may exist for such species include the ability to be cultured axenically (without bacteria), to be selected on solid agar plates supplemented with nutrients, to grow to high cell density, or requirement to undergo a sexual cycle. Other hurdles include good quality genomic and transcriptomic data, assembly, and annotation of the development of basic genetic engineering protocols (transformation, promoters, terminators, selection markers). The U.S. Department of Energy Joint Genome Institute's upcoming project to sequence 100 diatom genomes and transcriptomes will certainly contribute to the advancement of the field of diatom genetic engineering.

Bacterial-conjugation occurs in the oceanic environment between diatoms and bacteria commonly. It is a naturally-occurring genetic exchange process that could be expanded as a technology to diatoms that have not yet been successfully transformed. The development of conjugation as a transformation method in diatoms is an exciting alternative to biolistics for a number of reasons (Table 1.1). In addition to being important ecological players in contemporary oceanic food webs, diatoms have specific metabolic features that may potentially be exploited for biotechnology and synthetic biology applications. Conjugation offers reproducibility and

stability in transgene expression compared to biolistics (Table 1.1) that is a useful and sometimes necessary feature in synthetic biology application development (George et al. 2020). The episomal plasmid has also been shown to be a suitable chassis for expressing multi-gene synthetic pathways to add additional metabolic outputs like vanillin biosynthesis (Slattery et al. 2018) and geraniol (Fabris et al. 2020) in *Phaeodactylum*. More diatom laboratories are adopting conjugation to show proof-of-concept synthetic biology applications like these. Hopefully, conjugation can be applied to other diatom species that have endogenous features (see section 1.3.4.3 Limitations of CRISPR-Cas9 in diatoms) that are appealing for basic-research applications. Producing compounds using an episomal vector for industrial applications has yet to be attempted. Currently, other gene delivery technologies like biolistics and RNP have been used in industrial settings for transgene expression in microalgae. Conjugation may be a powerful tool for commercial purposes in the future but it is a relatively new technology and has yet to be accepted as a platform for industrial synthetic biology applications.

Diatoms can now be considered key players in biotechnology based on several criteria: the abundance of marketable bioproducts (lipids, carotenoids), the cost-effectiveness of industrial processes involving diatoms, performance in large-scale cultivation, and their robustness to harsh environmental conditions, and the potential to improve production or create new compounds through genetic engineering approaches. This last criterion is crucial to reduce the production cost for mass markets (biofuel, food, feed, green chemistry) and to create product diversity for high value markets (health or cosmetic). Although it is not always easy to find the names of companies already exploiting diatoms at an industrial scale, we can cite the companies: BarAlgae, Algatec, Solabia, Greentech, Astaxa GmbH whom employ diatom such as *Phaeodactylum*, *Chaetoceros gracilis* and *Odontella aurita*. For the production of biofuels, the

industry is still at the research and development stage. However, several initiatives have been launched to accelerate the transition to industrial application. An example is the partnership between the oil group Total, the Grenoble physiology laboratory, the Qingdao Chinese Institute of Bioenergy and Bioprocess Technology (QIBEBT), Wageningen University (European Magnifique project) and the University of Qatar. A second is a partnership between the oil group Exxon Mobil and Synthetic Genomics Incorporated (La Jolla, CA).

Although only a few studies have reported the use of genome editing tools to reshape the metabolic pathways of diatoms, we can expect a substantial increase in such activities in the near future due to the pace of development in recent years (Daboussi et al. 2014, Karas et al. 2015, Nymark et al. 2016, Nawaly et al. 2020). A central issue in metabolic engineering consists of identifying the cellular processes that control and limit the production flux. To overcome this issue, it is imperative to shift from the modification of specific pathways to the manipulation of metabolism in its entirety (the concept of "systems metabolic engineering") (Becker & Wittmann 2018). Thus, the accumulation of genomic, transcriptomic, metabolomics, and fluxomic data associated with development of low-cost and easy-to-use genome editing technologies will ensure rapid and continuous progress toward fully understanding the metabolic potential of diatoms.

An additional aspect, still tentatively studied, is the ability for diatoms to create new products for industry through heterologous gene expression. Several proof of concept studies have been reported, including production of a human IgG antibody against the Hepatitis B virus surface protein (Hempel & Maier 2012), a bioplastic poly-3-hydroxybutyrate (Hempel et al. 2011), an engineering PETase enzyme (Moog et al. 2015), of vanillin (Slattery et al. 2018), and monoterpenoids (Fabris et al. 2020). However, in order to avoid bioproduction problems

encountered in other organisms, such as the instability to express heterologous metabolic pathways or production of toxic bioproducts, several developments will have to be made. Here, we choose to focus on three major axes of research for the future. First, the identification of "safe harbor" integration loci, which enable stable and predictable transgene expression, without impacting neighboring genes, will be a major asset towards achieving the goal of improving strain stability and to meeting industrial specifications. Nonetheless, such loci would be useless if we are unable to integrate transgenes into them. The integration of transgenes into diatoms has occurred randomly and in many copies, it is imperative to circumvent this issue. For that, a second axis will be to play on DSB repair mechanisms in order to silence or inactivate the NHEJ controlling the random integration. Finally, the third axis concerns the development of fine-tuning gene expression toolbox enables to optimize synthetic gene networks and avoid metabolic pathways imbalances. Such genetic circuits could then be integrated into safe harbor loci or implemented through artificial chromosomes adapted for transmission by conjugation.

## 1.1.6 Funding Acknowledgements

**1.2 Multiplexed Knockouts in the Model Diatom *Phaeodactylum* by Episomal Delivery of a Selectable Cas9 (Moosburner et al. 2020)**

### 1.2.1 Abstract

Marine diatoms are eukaryotic microalgae that play significant ecological and biogeochemical roles in oceans. They also have significant potential as organismal platforms for exploitation to address biotechnological and industrial goals. In order to address both modes of research, sophisticated molecular and genetic tools are required. We presented here new and improved methodologies for introducing CRISPR-Cas9 to the model diatom *Phaeodactylum tricornutum* cells and a streamlined protocol for genotyping mutant cell lines with previously unknown phenotypes. First, bacterial-conjugation was optimized for the delivery of Cas9 by transcriptionally fusing Cas9 to a selectable marker by the 2A peptide. An episome cloning strategy using both negative and positive selection was developed to streamline CRISPR-episome assembly. Next, cell line picking and genotyping strategies, that utilize manual sequencing curation, TIDE sequencing analysis, and a T7 endonuclease assay, were developed to shorten the time required to generate mutants. Following this new experimental pipeline, both single-gene and two-gene knockout cell lines were generated at mutagenesis efficiencies of 48% and 25%, respectively. Lastly, a protocol for precise gene insertions via CRISPR-Cas9 targeting was developed using particle-bombardment transformation methods. Overall, the novel Cas9 episome design and improved genotyping methods presented here allow for quick and easy genotyping and isolation of *Phaeodactylum* mutant cell lines (less than 3 weeks) without relying on a known phenotype to screen for mutants.

**1.2.2 Introduction**

Marine diatoms are of biotechnological significance for their capacity to assimilate large amounts of carbon and nitrogen, divide rapidly (Cermeño et al. 2011), thrive at high cell densities, and fix carbon via photosynthesis. Diatoms shunt nutrients toward metabolic processes that produce industrially valuable products such as biodiesel precursors (triglycerides a.k.a. TAGs), polyunsaturated fatty acids (PUFAs), and photosynthetic pigments (Pulz & Gross 2004; Vinayak et al. 2015). Regardless of their innate ability to produce these high-valued molecules (HVMs), increasing production beyond their native capabilities is desirable for reducing production costs, and achieving sustainability goals in industrial agriculture applications (Gao 2018). Utilization of genetic engineering technologies could shift metabolic outputs of marine diatoms toward these HVMs and improve commercial production where algae have been suggested as biological platforms for industrial feedstock and nutraceutical production (Vinayak et al. 2015). Nevertheless, the development and optimization of such genetic engineering technologies is necessary before they can be safely, easily, and robustly employed.

Genetic engineering technologies, particularly the versatile CRISPR-Cas9 toolbox, have been used for numerous applications in a vast array of organisms (Mali et al. 2013b; Sanders and Joung 2014; Wang et al. 2016). CRISPR (clustered regularly interspaced short palindromic repeats), a foreign nucleic acid immunity system, is encoded in the genomes of most bacteria and some archaeal species (Deltcheva et al. 2011). The type-II CRISPR system from the bacterium *Streptococcus pyogenes* was first adapted for gene editing using the Cas9 enzyme, which induces blunt-ended double-stranded breaks (DSBs) when guided to a genetic target by a chimeric RNA molecule of the tracr-RNA and cr-RNA called a single-guided RNA (sgRNA) (Cong et al. 2013; Jinek et al. 2013; Mali et al. 2013a). Together Cas9 and an sgRNA form a ribo-endonuclease

complex *in vivo* in many cell types to edit or mutate genomic targets. In the bacterial immunity system, the sgRNA targeting sequence, a string of ~20 nucleotides (nt), preserves the sequence of a previous immune response, called spacer acquisition (McGinn and Marraffini, 2019); whereas in gene editing applications the nucleotide sequence of 20-nt spacer is designed by the researcher to target a specific gene/target, followed by interrogation and cleavage of a nucleic acid target by the Cas9 endonuclease. The sgRNA guides Cas9 to the target by forming complementary base pairing with the 20 nt at the 5′ end of the sgRNA, called the spacer. Once a DSB is induced, the blunt ended breaks are repaired by one of two native DNA repair mechanisms: homology-directed repair (HDR) or non-homologous end joining (NHEJ). HDR relies on a homologous donor to repair the DSB, typically a sister chromosome in eukaryotes, by homologous recombination (LaFave and Sekelsky 2009). Endogenous HDR can be hijacked in order to introduce, at the DSB, an exogenous donor sequence, containing a user-designed mutation that will result in an edited target locus. For NHEJ, ligation of the blunt ends repairs the DSB, however, this process is error-prone and leads to random insertions or deletions of nucleotides of unpredictable sizes called indels (Davis & Chen 2013).

CRISPR-Cas9 has been employed in both prokaryotic and eukaryotic organisms but eukaryotes pose particular challenges for CRISPR-Cas9 experiments as most have multiple copies of their chromosomes. Multiple copies of chromosomes provide two or more target loci for Cas9. For instance, in a diploid organism, two individual DNA repair events must occur to produce homozygous or heterozygous mutations at the target loci. For NHEJ-mediated mutagenesis (where no homologous donor is introduced), a heterozygous genotype arises after two separate Cas9 cleavage events and NHEJ-mediated repair that results in two distinct indels at the target loci. A homozygous genotype can also arise when one indel-containing target locus is

used as a homologous donor for the second target loci (located on the sister chromosome of the first target), thereby producing an identical indel mutation. The unpredictability of NHEJ-mediated mutagenesis in diploid organisms therefore requires highly sensitive genotyping methodologies. Many genotyping methods have been used to genotype diploid organisms; among the most common are high-resolution melt curve analysis (HRMC) (Nymark et al. 2016), T7 endonuclease I assay (Guschin et al. 2010; Cho et al. 2013), high-throughput next generation sequencing (Zhou et al. 2014), and manual curation of sequencing reads.

Marine diatom species have been subjected to CRISPR-Cas9 mutagenesis including the model strains: the pennate, *Phaeodactylum tricornutum* (Nymark et al. 2016; Serif et al. 2018; Sharma et al. 2018; Slattery et al. 2018; Stukenberg et al. 2018) and the centric *Thalassiosira pseudonana* (Hopes et al. 2016). To date, NHEJ-mediated mutagenesis has been observed only in *Phaeodactylum* (Nymark et al. 2016; Serif et al. 2018; Slattery et al. 2018) by using one of two genetic transformation methods, micro-particle bombardment or bacterial conjugation, to introduce CRISPR-Cas9. Despite the successes of utilizing CRISPR-Cas9 in generating mutant diatom cell lines, genotyping the generated cell lines has been difficult. *Phaeodactylum*, a diploid organism, requires high-sensitivity screening methods to genotype multiple cell lines in a high-throughput manner. Genotyping methods such as high-throughput melt curve (HRMC) analysis and the T7 endonuclease I assay are commonly used to evaluate mutant genotypes in multi-ploidy organisms. To date, HRMC analysis and CRISPR Activity Factor (CAF) analysis have been used for *Phaeodactylum* genotyping experiments (Nymark et al. 2016; Sharma et al. 2018; Stukenberg et al. 2018). Similarly to CAF, TIDE (Tracking Indels by DEconvolution) sequencing analysis, an open-source software package, deconvolves raw sequencing reads to reveal indels in diploid or multiploidy organisms (Brinkman et al. 2014; Serif et al. 2018).

Allele-specific PCR amplification has been used for genotyping *Phaeodactylum* mutants without the need for subcloning (Serif et al. 2018).

Here, TIDE sequencing analysis and manual sequencing curation is used to streamline the genotyping pipeline from colony formation to mutant cell line validation. Both single-gene and two-gene mutant cell lines were produced utilizing conjugation to introduce a selectable-Cas9 and either one or two sgRNAs to *Phaedocatylum*. Also, the T7 endonuclease I assay has also been adapted to quantify the mutagenesis efficiency of individual sgRNAs throughout a population of transformed *Phaeodactylum* prior to genotyping individual colonies.

A new selectable-Cas9 episome design was used that permits antibiotic selection for the Cas9 gene with the goal of both increasing mutagenesis efficiency and potentially avoiding mixed genotypes forming in generated colonies after conjugation. The P2A self-cleaving peptide was used to transcriptionally fuse Cas9 and the *Phaeodactylum* antibiotic resistant gene, Shble (Kim et al. 2011). By doing so, Cas9 and Shble are co-transcribed under the same promoter followed by cleavage of the translated products, essentially selecting for Cas9 via antibiotic treatment. Lastly, a robust golden-gate assembly based method was developed to visualize positive and negative *E. Coli* clones that harbor ready-to-conjugate episomes. Overall, the methods presented here allow for easy Cas9-sgRNA episome assembly and streamlined genotyping methodology and isolation of *Phaeodactylum* mutant cell lines in less than 3 weeks without relying on a known phenotype to screen for mutants.

### 1.2.3 Materials and Methods

#### 1.2.3.1 Cell culturing

A reference strain of *Phaeodactylum* (CCAP-1055/1) was used in all experiments. *Phaeodactylum* was grown at 18°C under white fluorescent lights (50 µE m–2 s–1) and subjected to a diel growth cycle (14 h:10 h; light:dark). Culture medium was artificial seawater (ASW) supplemented with trace metals, essential vitamins, 55 µM of NaPO4, and 880 µM of the appropriate nitrogen source (NaNO$_3$, NH$_4$Cl, or Urea). Cultures, when not used in growth assays, were maintained with chloramphenicol antibiotic (10 mg/L) to keep cultures bacteria-free. Mutant *Phaeodactylum* strains were supplemented with either phleomycin (50 mg/L), zeocin (50 mg/L), or nourseothricin (200 mg/L).

*Escherichia coli* (TOP-10, Life Technologies, Carlsbad, CA) was used for all molecular cloning purposes. The cultures were grown on Luria-Bertani broth or agar and supplemented with the following antibiotics when necessary: ampicillin (100 mg/L), carbenicillin (100 mg/L), tetracycline-HCl (10 mg/L), gentamicin (20 mg/L), zeocin (25 mg/L).

#### 1.2.3.2 Molecular biology

Plasmid construction was performed with reference to Sambrook and Russell (2006) unless stated otherwise. Multiple DNA polymerases were used for cloning purposes including Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, MA, United States), AccuPrime Taq High-Fidelity DNA polymerase (Thermo Fisher Scientific), OneTaq 2X Master Mix with standard buffer (New England Biolabs, Ipswich, MA, United States), and Phire

Plant Direct PCR Master Mix (Thermo Fisher Scientific). Enzymatic components to make a 2X GA master mix that included Phusion HF DNA polymerase, T5 exonuclease (Thermo Fisher Scientific), and DNA Taq Ligase (Thermo Fisher Scientific) were purchased separately and mixed in lab. Enzymatic components for Golden-Gate cloning were also purchased separately but only mixed on the day of cloning. They included the type-II restriction enzymes BsaI-HFv2 and BbsI (New England Biolabs) and a T4 DNA Ligase (Thermo Fisher Scientific).

Primer sequences can be found in the Supplemental Information 1.

### 1.2.3.2.1 CRISPR-Cas9 genetic components

A human-codon optimized Cas9 open reading frame (ORF) containing three nucleus-localization signals (NLS) was produced by the Feng Zhang laboratory (Broad Institute, Cambridge, MA, United States) and purchased from addgene.com [pSpCas9(BB)-2A-Puro (PX459) V2.0, ID: 62988, Ran et al. 2013). The Cas9 ORF was cloned into a pUC19 backbone using GA cloning (Gibson et al. 2009) using primers Cas9-2A-shble_1 and P2A_1. For constitutive expression of Cas9, the FcpB/FcpA (Fucoxanthin chlorophyll-a/-binding protein B promoter/Fucoxanthin chlorophyll-a/c-binding protein A terminator) (Bailleul et al. 2010) promoter/terminator pair were chosen (FcpB-Cas9).

The HDR-donor plasmid used in Weyman et al. (2014) for TALEN mutagenesis was adapted to a HDR-donor plasmid for targeting nitrate reductase (pKO-NR). pKO-NR contained 1 kb homology arms flanking an expression cassette coding for the Shble gene that conferred resistance to the antibiotic phleomycin. The cassette inserted both a Shble cassette and a STOP codon upstream of the Shble promoter (Fucoxanthin chlorophyll-a/c-binding protein F, FcpF).

This was in frame with the targeted coding gene. The primers and plasmid sequence information can be found in Weyman et al. (2014).

The sRNAi promoter (chr2: 28,038-29,083) and terminator (chr2: 29,124-29,424) was amplified from *Phaeodactylum* genomic DNA (promoter primers = sRNAi1_1 and sRNAi1_2; terminator primers = sRNAi_3 and sRNAi_4). For Golden Gate cloning, the sRNAi terminator was shortened by PCR to 112 bp to eliminate an internal BsaI restriction site. The sRNAi promoter and terminator was cloned from *Phaeodactylum* genomic DNA at the locus of the most highly expressed small, non-coding, intergenic RNA called sRNAi (Rogato et al. 2014; Bielinski et al. 2017). The sequences for the sRNAi promoter and terminator can be found in the Supplemental Information 1.

The sgRNA backbone, which consists of the genetic fusion of the crRNA and tracrRNA (Jinek et al. 2012), was amplified from the same Addgene plasmid that contained Cas9 (primers = sRNAgU1_F and sRNAgU1_R). An sgRNA expression cassette for *Phaeodactylum* (sRNAi_gBbsI) was constructed using a 4-piece GA approach into an XbaI-digested pUC19 plasmid containing a bacterial expression cassette for the bla gene conferring resistance to ampicillin/carbenicillin. The resulting plasmid was then PCR amplified at the 3′ end of the sRNAi and the 5′ end of the sgRNA backbone (primers = gRNA_1 and gRNA_2). This amplicon served as a linearized vector for a 2-piece GA procedure for a spacer insert that contains two flanking BbsI restriction sites (Supplemental Figure 1.1). The 100 bp GA insert was constructed by ligating two 60 bp oligos with homology to the sRNAi promoter and sgRNA backbone. The resulting plasmid, sRNAi_g2XBbsI, was used for BbsI-based GG cloning of CRISPR spacers (Supplemental Figure 1.1b).

For all Cas9 target genes, spacer sequences were obtained for nitrate reductase (Phatr3_J54983), glutamine synthetase II (GS-2) (Gene ID: Phatr3_J51092), and chloroplastic glutamate synthase (cGOGAT) (Gene ID: Phatr3_J24739) using the Phyto CRISP-Ex online package (Rastogi et al. 2016) that passed the built-in criteria of minimizing off-target Cas9 activity.

A GG cloning protocol was modified from the sgRNA expression cassette supplied by Feng Zhang laboratory. In place of the BbsI-containing spacer sequence within the sgRNA expression plasmid, a LacZ bacterial expression cassette with flanking BbsI sites was cloned (primers = LacZ_GG_1 and LacZ_GG_2) into the sRNAi expression cassette using GA. The *Phaeodactylum* sgRNA expression plasmid previously produced (sRNAi_gBbsI) was PCR amplified (primers = LacZ_GG_3 and LacZ_GG_4) and the amplicon was gel extracted and used as a GA vector. The sgRNA cloning vector (sRNAi_gLacZ) was the assembled plasmid. This plasmid was used in all future sgRNA expression cassette cloning and assembly experiments as a cloning vector (Supplemental Figure 1.1b). An appropriate GG insert for sRNAi_gLacZ was assembled by ligating two 56-nt oligos together that are complementary to each other. The oligo primer sequences for all sgRNA target sequences can be found in Supplemental Information 1.

Detailed protocols for spacer cloning into the sRNAi_gLacZ plasmid (name GG1 cloning) are available at protocols.io[3].

---

[3] https://www.protocols.io/view/gg1-sgrna-cloning-for-phaeodactylum-tricornutum-4abgsan

### 1.2.3.2.2 Cas9-episomal plasmid cloning

The *Phaeodactylum* episomal plasmid pPtPBR1 was obtained from Rachel Diner (Diner et al. 2016). First, one BsaI restriction site within the bla coding region of pPtPBR1 was removed by PCR amplification (primers = pBR322-Amp_1 and pBR322-Amp_2) and reassembly to produce plasmid pPtPBR1 (BsaI-). pPtPBR1 was then used as a template for GA cloning of a Cas9 ORF and a GG cloning BsaI acceptor site for sgRNA expression cassette cloning (Supplemental Figure 1.2). Cas9 was PCR amplified twice using one forward primer (Cas9-2A-Shble_1) and two reverse primers (P2A_1 and P2A_2) so to add GA homology to the FcpF promoter (5′ of Cas9) and the Shble ORF (3′ of Cas9). The double PCR reaction added a GSG-2A peptide DNA coding sequence to transcriptionally fuse hCas9 and Shble (Supplemental Figure 1.2, GSG-P2A nucleotide sequence is in Supplemental Information 1). Next, a red-fluorescent protein bacterial expression cassette (mRFP) was obtained and PCR amplified (primers = Cas9-2A_Shble-RFP_1 and 2) to produce GG acceptor sites. The remaining two GA pieces were PCR amplified from pPtPBR1 (BsaI-) to insert two GG cloning BsaI sites between the bla and tet expression cassettes and flanking mRFP, one using primers Cas9-2A-Shble_2 and 3 (5′ – Shble – Pt_Centromere – pMB1 – bla – 3′) and two using primer Cas9-2A-Shble_6 and 7 (5′ – tet – OriT – FcpF – 3′). After cloning and *E. coli* transformation, the correctly assembled episome, pBR_Cas9-2A-Shble, 2XBsaI-mRFP(AT) (Supplemental Figure 1.3a), was sequence verified.

Next, primers were designed to PCR amplify up to six sgRNA expression cassettes (cloning and construction described above) and to add BsaI-restriction and cleavage sites to the 5′ and 3′ ends of the sRNAi promoter/terminator. GG cloning is an optimal assembly platform when cloning more than one sgRNA expression cassette. For inserting one sRNAi_sgRNA

cassette, primers GG-gRNA1-F and sgRNA(GG)-2 were used. For two sgRNAs, primers GG-gRNA1-F and R were used for sgRNA #1 and primers GG-gRNA3-F and sgRNA(GG)-2 were used for sgRNA #2. Between the two sgRNAs was cloned a LacZ expression cassette, which, when induced with IPTG activates β-galactidose that digests Xgal to produce blue *E. coli* colonies (Supplemental Figure 1.3c). The sgRNA expression cassette(s) and LacZ amplicons was mixed at double the molar concentration of the BsaI-digested pBR_Cas9-2A-Shble, 2XBsaI-mRFP(AT) vector and carried through the GG cloning reaction with BsaI-HF and a T4-ligase (Supplemental Figure 1.1c). The final sgRNA-LacZ-sgRNA episomal plasmid map can be seen in Supplemental Figure 1.3b.

Detailed protocols for sgRNA cloning into the Cas9-2A-Shble epsiome (named GG2 cloning) is available at protocols.io[4].

### 1.2.3.3 *Phaeodactylum* genetic transformations

#### 1.2.3.3.1 Bacterial-mediated conjugation

Bacterial-mediated conjugation was used to introduce CRISPR components to *Phaeodactylum*. Diner et al., 2016 provided detailed methodology for introducing episomal plasmids to *Phaeodactylum* in a high-throughput manner. For all transformations, 100 µL of dense *Phaeodactylum* (1e8 cells/mL) was plated on conjugation-based solid agar medium ($NH_4/NO_3$-ASW, 1% agar, 5% LB in $NH_4$-supplemented ASW) in 6-well cell culture plates and incubated for 4 days under light (18:6) and at 18°C. Prior to transformation,

---

[4] https://www.protocols.io/view/gg2-crispr-cas9-episome-cloning-using-red-blue-scr-4acgsaw

pBR_Cas9-2A-Shble, sRNAi-sgRNA episomes were transformed into recipient *E. coli* harboring the conjugation plasmid pTA-MOB (Karas et al. 2015; Diner et al. 2016) and selected on agar plates for both the cargo and conjugation plasmids (Figure 1.2). For a transformation control (positive for colonies not Cas9 cleavage), the episome pBR_Cas9-2A-Shble, 2XBsaI_gRFP1 was used where Cas9 expression would still occur without an accompanying sgRNA. For a negative control, a similar episomal plasmid was built harboring the NAT gene that confers resistance to nourseothricin and not phleomycin. Under phleomycin selection, the transformed cells would die. Transformation efficiency for the delivery of Cas9-2A-Shble:sgRNA episomes was 500–750 transformants per 1e8 phleomycin-resistant cells.
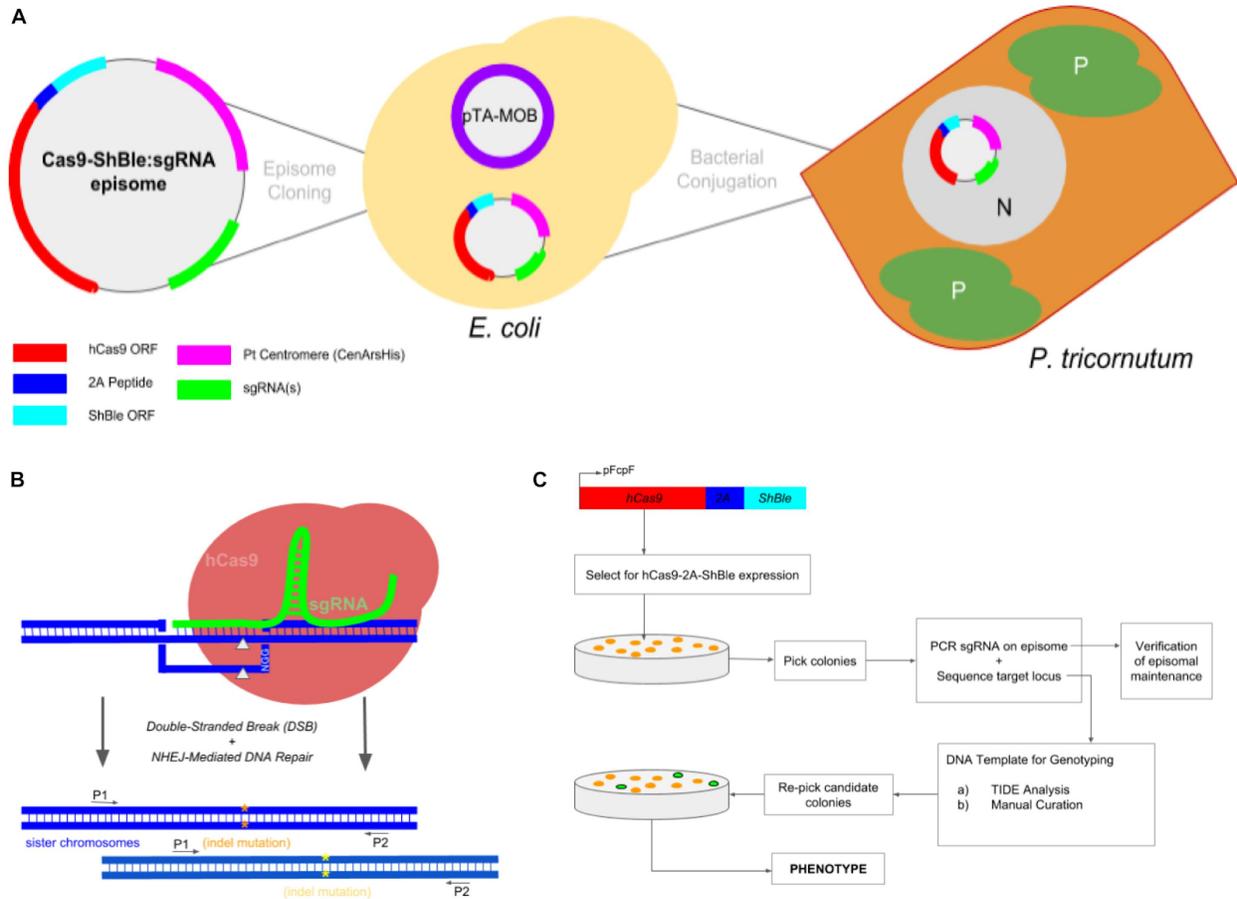
**Figure 1.2** Experimental workflow to produce *Phaeodactylum* mutants via conjugation-based delivery of a CRISPR-Cas9 episome. **(a)** A cartoon schematic of Cas9-Shble:sgRNA episome cloning to *E. coli* cells harboring the pTA-MOB plasmid and followed by bacterial-conjugation transformation of the episome to *Phaeodactylum*. *Phaeodactylum* cells maintain and replicate the Cas9-Shble:sgRNA episome as their native chromosomes in the nucleus (N). The diatom plastids (P) and cytoplasm (C) are labeled. Here, Cas9 is transcriptionally fused to Shble and can therefore be selected for after transformation with the antibiotic phleomycin. **(b)** Episomal expression of Cas9 and sgRNA(s) and target mutagenesis of a diploid organism. The sgRNA(s) guides Cas9 to genomic target(s), and then Cas9 induces a double-stranded break (gray arrows). *Phaeodactylum*, being a diploid eukaryote, will contain two distinct NHEJ-mediated mutations (orange and yellow asterisks) between the sister chromosomes. **(c)** Genotyping and cell line picking workflow after colony selection for Cas9-Shble. A "double picking" strategy was used by which cell lines were picked and target locus sequenced using two primers. After genotyping, the colony can be re-picked for further cell line analysis.

The day before transformation, 4 random colonies from the selection plates were used to inoculate 3 mL of LB containing selection (Amp, Tet, Gent). After growing overnight, 200 μL of culture was used to inoculate 12.5 mL of LB. The culture was grown for 3–4 h, until an $OD_{600}$ nm of 0.8–1 was reached, pelleted, and suspended in 100μL of SOC media. Then, the *E. coli* was pipetted into a corresponding well containing *Phaeodactylum* and mixed either by a spreading loop or by rotating the 6-well plate such that the *E. coli* covered the entire lawn of *Phaeodactylum*. After *E. coli* was added to all wells and allowed to dry under a vacuum PCR hood, the 6-well plate was placed in the dark at 30°C for 90 min. The plate was then placed back at 18°C and exposed to light for 48 h.

Finally, each well was scraped using a sterile cell spreader, collected in a 2 mL epitube then re-plated on 100 mm selection plates containing the appropriate selective media. The selection plates were then allowed to grow for 10–14 days or until colonies appeared. For the T7 endonuclease I assay reaction, half of the scraped *Phaeodactylum* cell lines (~400 μL) was resuspended in 200 mL $NH_4$-supplemented ASW media supplemented with phleomycin and chloramphenicol. The cells were then grown under light and at 18°C for 4 days, passed into 200 mL fresh media, grown for an additional 3 days, then finally pelleted and flash frozen.

Detailed protocols for bacterial-conjugation in *Phaeodactylum* are publicly available at protocols.io[5].

### 1.2.3.3.2 Micro-particle bombardment (biolistics)

Micro-particle bombardment genetic transformation (Falciatore et al. 1999) was used to introduce CRISPR components to *Phaeodactylum*. First, *Phaeodactylum* was pelleted during

---

[5] https://www.protocols.io/view/crispr-cas9-episome-conjugation-into-phaeodactylum-5pvg5n6

exponential growth in liquid culture ($NO_3^-$-supplemented ASW), suspended, and plated onto agar growth plates at a cell concentration of 3e8 cells/mL (400 µL). Next, 24 µg of DNA (8 µg each plasmid; FcpB-Cas9, sRNAi-gNR-B, pKO-NR) were hybridized to tungsten beads and introduced to the plated *Phaeodactylum* at high velocity using the PDS-1000/He Biolistic® Particle Delivery System (Bio-Rad, Hercules, CA). The transformed *Phaeodactylum* plates were then allowed to recover for 48 h at 18°C and in constant darkness. After recovery, *Phaeodactylum* was re-plated on agar plates containing $NH_4$-supplemented ASW with phleomycin to select for the presence of the pKO-NR plasmid. The selection plates were then grown under normal conditions for 21–28 day, or until *Phaeodactylum* colonies formed and were visible. Experimental controls were also performed where either FcpB-Cas9 and pKO-NR were co-transformed or pKO-NR was transformed alone.

### 1.2.3.4 T7 Endonuclease I Assay

The T7 endonuclease I assay (New England Biolabs) was used to quantify the *in vivo* cleavage efficiency for multiple clonal *Phaeodactylum* populations for each sgRNA target (per manufacturer's methods[6]). Frozen cell pellets obtained following liquid selection (as described above) were slowly thawed on ice. Genomic DNA was then extracted from the pelleted cells using the Plant DNAzol Reagent (Thermo Fisher Scientific). The product-supplemented protocol was used; however, the pulverization step, the first step, was skipped because diatom pellets can be easily resuspended. Genomic DNA was then extracted from the cell pellet. 200 µL of genomic DNA at concentration ranging from 200 to 500 ng/uL was recovered. A total of 200 ng of genomic DNA was used as a PCR template to start the T7 endonuclease I assay. Six separate

---

[6] https://www.neb.com/protocols/2014/08/11/determining-genome-targeting-efficiency-using-t7-endonuclease-i

amplifications were performed (only 25X amplification cycles) and pooled prior to the T7 assay. For the NR locus, primers NR-gene-1 and NR-KO-2∗ were used. This amplification product was then re-amplified with nested primers NR-HRM-A and NR-HRM-B. The nested-amplicons were then PCR cleaned and concentrated and used as a DNA input for the heteroduplex hybridization. This input is referred to as "Amplicon" in Figure 1.3 and the hybridized heteroduplex is referred to as "Heteroduplex." T7 endonuclease I was then added to the hybridized product to cleave all heteroduplex DNA. All reaction conditions and efficiency calculations (as shown in Figure 1.3) were supplied by New England Biolabs. Genome editing efficiency was determined using the Agilent 4200 TapeStation bioanalyzer and Agilent High-Sensitivity D1000 screen chip and assay. The fraction cleaved value was determined by examining the peak intensities of the high-sensitivity gel of the T7 product compared to the uncut Heteroduplex product. Fraction cleaved was then an input for the following equation to calculate the percentage of gene modification:

$$\% \text{ gene modification} = 100*(1- (1-\text{fraction cleaved})^{1/2}).$$
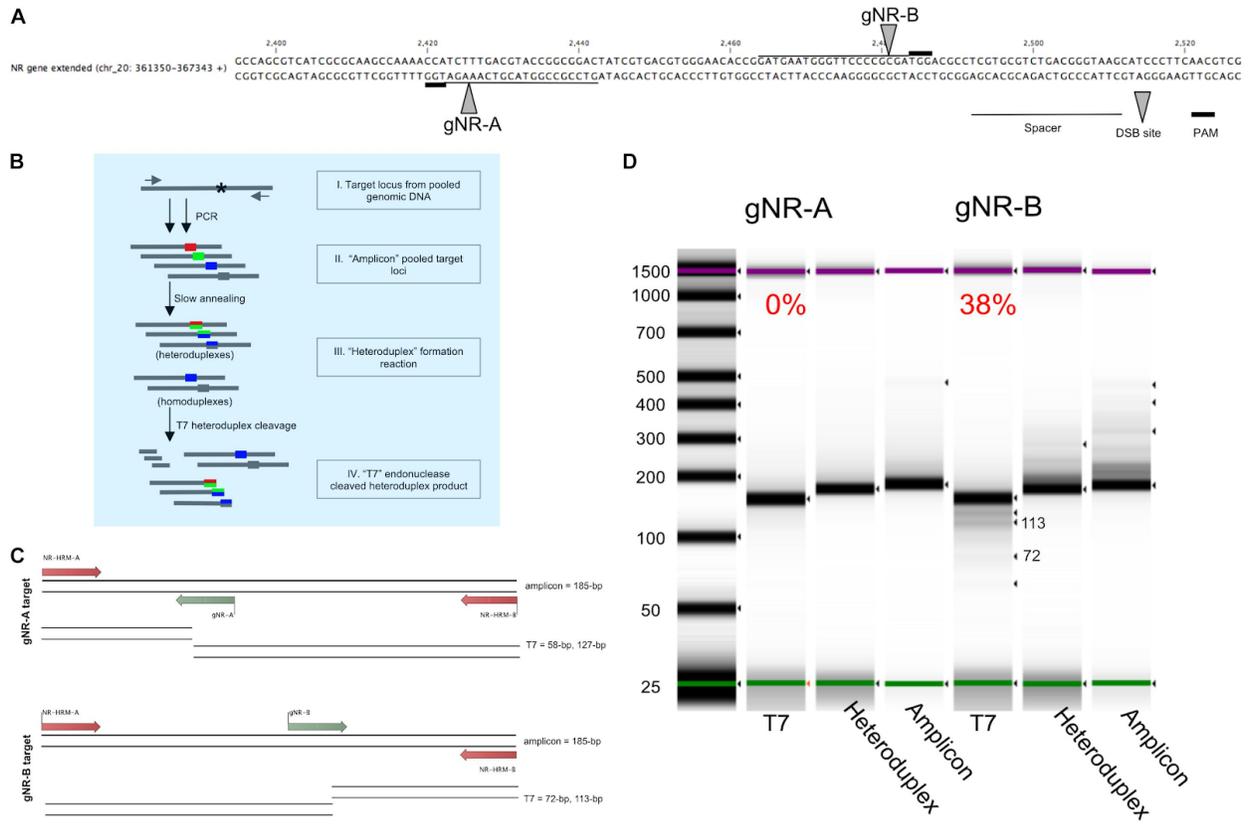
**Figure 1.3** Pooled-diatom transformant T7 assay screen for *in vivo* Cas9 activity to compare the mutagenesis efficiency between two sgRNAs designed to target the nitrate reductase (Phatr3_J54983) gene. **(a)** Two sgRNAs were designed within the *Phaeodactylum* genomic region displayed. The spacer sequence (line), PAM site (bold line) and Cas9 cut site (grew arrow) are defined for sgRNA gNR-A and gNR-B. **(b)** The T7 heteroduplex assay was used to quantify the efficiency of *in vivo* Cas9 activity for each sgRNA for a population of pooled-*Phaeodactylum* transformants. This chart demonstrates how the T7 heteroduplex assay can produce and cleave mismatched amplicon products(I). The pool of target loci amplicons (II) is denatured and slowly annealed to form mismatched double-stranded DNA, or heteroduplexes (III). The heteroduplexes are then digested by T7 endonuclease I (New England Biolabs). Heteroduplexes are subsequently cleaved by T7 while homoduplexes remain full length. **(c)** The expected T7 cleavage pattern for gNR-A and gNR-B associated amplicons. The size of the "amplicon" input is 185-bp. For gNR-A, the expected "T7" products are ~185-bp (homoduplex), 127-bp and 58-bp. For gNR-B, ~185-bp (homoduplex), 113-bp and 72-bp. **(d)** The three products, "amplicon," "heteroduplex," and "T7" were analyzed for size on a 4200 TapeStation System (Agilent, Santa Clara, CA, United States). The T7 cleavage efficiency is displayed above each "T7" lane in the gel that represents the mutagenesis efficiency of each sgRNA for all *Phaeodactylum* cell lines. Each band detected by the software is indicated by a black arrow. The expected "T7" 113-bp and 72-bp bands are labeled.

**1.2.3.5 Genotyping *Phaeodactylum* exconjugants**

*Phaeodactylum* cells selected on solid media were grown for 10–14 day or until visible colonies formed. Colonies were picked with pipette tips directly into 5μL of dilution buffer supplied with the Direct Plant PCR Kit (Thermo Fisher Scientific) and incubated at RT for 15 min. The buffered colony was then pipette mixed and 2μL were used as a DNA template for two PCR reactions. The first, common for all sgRNAs, used primers Ars_Seq and V2_insert_2 (Supplemental Figure 1.4a, black arrows) to amplify from the Pt_Centromere to and through the sgRNA expression cassette to check for correct episomal maintenance and sgRNA presence (Supplemental Figure 1.4b). For nitrate reductase, primers NR-gene-1 and NR-KO-2∗ were used to amplify a 641 bp region around the target loci for gNR-A and gNR-B. The Direct Plant PCR Kit was used for both PCR reactions for 40 cycles. The resulting PCR amplicons were used for genotyping.

**1.2.3.6 TIDE sequencing analysis**

TIDE (Tracking of Indels by DEconvolution, Brinkman et al. 2014) sequence analysis is an online-based software that analyzes raw sequencing data to find one or multiple indels mutations at a Cas9 target site. For the purposes of this study, TIDE was used to analyze indels resulting from Cas9 targeting of nitrate reductase in *Phaeodactylum* by sgRNAs gNR-A and gNR-B (Figure 1.4). An example of the TIDE analysis output for a wild-type sequence is shown in Figure 3A. 80 colonies for gNR-A and 64 colonies for gNR-B were picked directly into PCR mix (protocol described above). 200 ng of the PCR amplicons were then directly sequenced

using two primers, NR-HRM-A and NR-gene-2, for paired-end reads (Figure 1.3a). *Phaeodactylum* transformed with the Cas9-2A-Shble vector without an sgRNA was used as a control. TIDE analysis requires the input of (1) the nuclease species used (SpCas9), (2) the 20 nt spacer target, (3) an ABI (.abi) file for a reference sequence (control Cas9-2A-Shble), and (4) an ABI (.abi) file for a sample sequence. The sample sequence is compared to the control sequence for significant differences between the sequence electropherograms. Certain criteria were identified and selected for when analyzing the TIDE output. First, a total efficiency value (in percentage) of 50% was chosen because it signified that the target locus was mutated at a high rate. Second, the *p*-value of each output peak must be <0.001. Third, only colonies with distinct genotypes, based on the criteria of only 1 or 2 predominant indels each with greater than 10% of the total sequencing reads, were selected as candidate mutant lines. Lastly, for the cell lines that passed each of these criteria, results were verified by conducting TIDE analysis of the paired-end read of the same sample. If both paired-end TIDE analyses pass the criteria and are matched indels, the sample was chosen as a mutant cell line candidate. Each sequencing read was manually analyzed, individually, using the open-sourced TIDE web-tool. Genotyping results for gNR-A (Supplemental Table 1.1) and gNR-B (Supplemental Table 1.2) are available.

**Figure 1.4** Genotyping NR-KO mutant cell lines by colony sequencing and TIDE sequence analysis. **(a)** The sequencing alignment paired with a TIDE chart associated with a wild-type *Phaeodactylum* NR sequence. The gNR-B spacer and PAM are represented by the gray arrow and the Cas9 cut site by the black arrow. Below is the sequencing read, the associated electropherogram, and the TIDE result chart. The TIDE chart is a bar plot where the *x*-axis represents the predicted indel mutation (0, wild-type, negative values, nucleotide deletion; positive values, nucleotide insertion) and the *y*-axis represents the percentage the specific mutation is of the entire sequence. A bar shaded pink indicates a *p*-value < 0.001 for that predicted indel. Here, there is one pink bar at position "0" (wild-type) that represents 97.1% of the entire sequence (labeled above the bar). **(b)** Six NR-KO mutant cell line genotypes were assessed the same as wild-type. The wild-type sequencing read was used as a reference input for TIDE analysis to compare potential mutant sequence genotypes. NR-KO mutant genotyping was performed for all colonies using both manual sequencing curation and TIDE analysis (Supplemental Table 1.2) and candidate NR-KO cell lines were chosen for phenotype analysis. Here, six NR-KO candidates and their associated genotyping results are shown. Cell lines 20, 21, and 56 display a single peak in the negative values that indicates a homozygous deletion genotype. Cell lines 27, 43, and 51 display two peaks in the negative values that indicate two deletions or a heterozygous deletion genotype. **(c)** The six candidate NR-KO cell lines were validated to have lost nitrate reductase function when grown in liquid medium supplemented with nitrate. Compared to growth on urea (left panel), the NR-KO mutant could not grow on nitrate (right panel). Two wild-type cell lines are blue and the six NR-KO cell lines are in red. All lines were grown in three technical replicates.

66
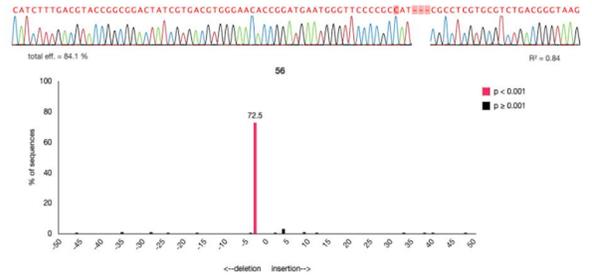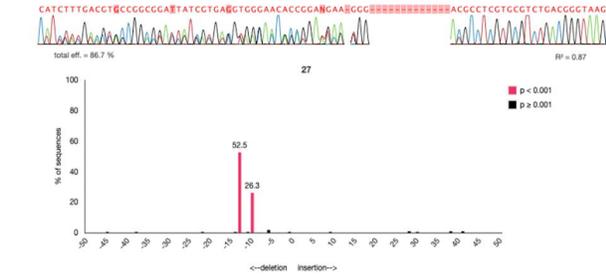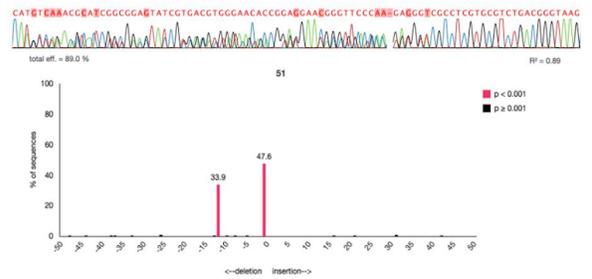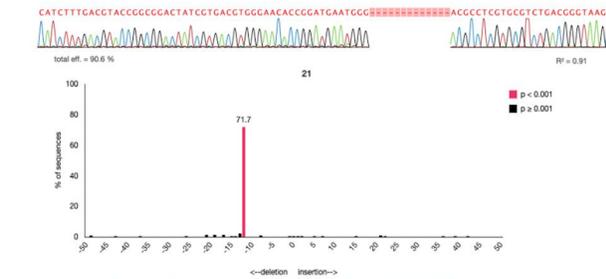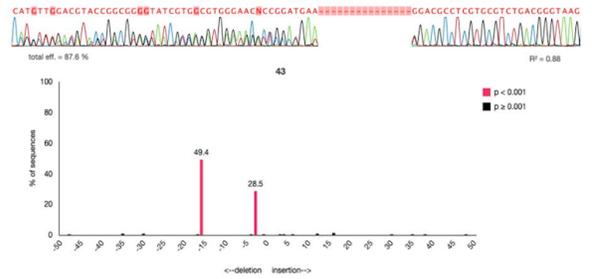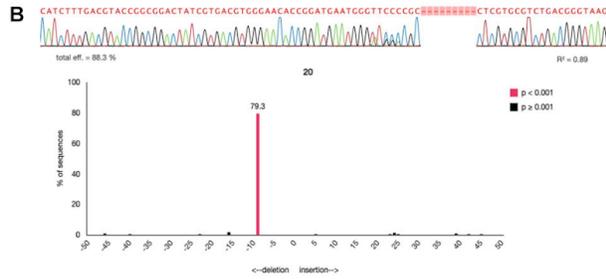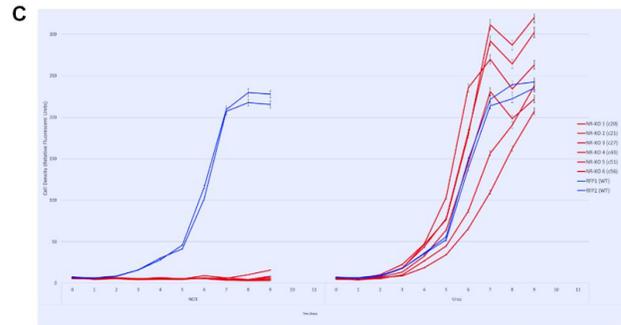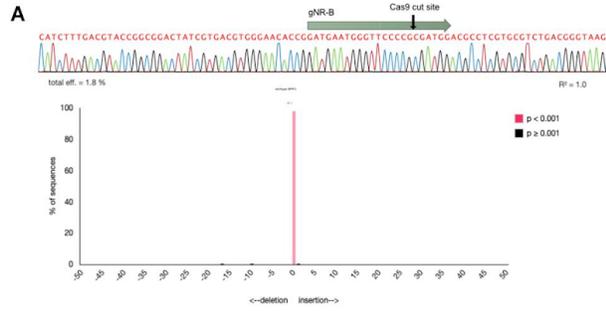
### 1.2.3.7 Nitrate Phenotype Assay

NR-KO *Phaeodactylum* cell lines were identified by TIDE and subjected to the nitrate phenotype assay where *Phaeodactylum* mutants of nitrate reductase are grown on media supplemented with nitrate and ammonium or urea, separately. The same nitrate assay was used for mutant cell lines produced via micro-particle bombardment. Mutants with a dysfunctional NR gene are unable to grow on nitrate media while proliferating on ammonium. First, all 40 colonies (for each sgRNA) were re-picked into liquid media supplemented with ammonium. The cultures (including four control cultures) were grown under normal conditions until mid-log phase and then passed into nitrate and ammonium media at a concentration of 1e4 cells/mL. Cell density was measured daily (2:00pm PST) for 4 days and growth between the two conditions were compared. The cell lines that could grow on ammonium-supplemented ASW and could not grow on nitrate-supplemented ASW were considered NR mutant cell lines.

### 1.2.4 Results

### 1.2.4.1 CRISPR-Cas9 Mutagenesis Workflow

The general workflow from episomal exchange to *Phaeodactylum* colony genotyping is shown in Figure 1.2. The *Phaeodactylum* episome was built with a transcriptionally fused Cas9 and Shble, an antibiotic selectable marker gene, expression cassette that allows Cas9 to be selected for after transformation with the eukaryote antibiotic phleomycin. Additionally, a red-fluorescent protein bacterial expression cassette (mRFP) was cloned into the episome with flanking BsaI restriction enzyme sites that, when digested, leave unique 4-bp overhangs for

golden-gate cloning of sgRNA constructs (Supplemental Figure 1.1). A hierarchical cloning strategy was used to build and clone one *Phaeodactylum* sgRNA expression cassette in two golden-gate assembly steps (Supplemental Figure 1.2). Negative selection against RFP was used for cloning an individual sgRNA expression cassette where negative *E. coli* colonies were visibly red. For cloning more than one sgRNA, a LacZ bacterial expression cassette was cloned with and between the sgRNAs and colors *E. coli* colonies blue in the presence of IPTG and Xgal. The red-white-blue screening method helped expedite episome cloning in general (Supplemental Figure 1.3).

After construction and sequence verification, a Cas9-Shble:sgRNA episome was transformed into electro-competent *E. coli* cells harboring a pta-MOB plasmid (Karas et al. 2015; Diner et al. 2016) that has been shown to be required for efficient conjugation exchange of the episome to *Phaeodactylum* (Figure 1.2a). The episome then resides in the *Phaeodactylum* nucleus, separate from the *Phaeodactylum* chromosomes as an artificial chromosome, where expression of Cas9 and sgRNA(s) occurs. Like many CRISPR-Cas9 gene targeting experiments in organisms with diploid genomes, the Cas9-sgRNA ribo-endonuclease complex is guided to both gene alleles where distinct NHEJ-mediated mutations may occur (Figure 1.2b). While a heterozygous genotype (two NHEJ-mediated mutations) was expected, a homozygous genotype was also expected where one mutated allele can serve as an HDR donor template to repair the second allele (HDR) or where micro-homology loci upstream and downstream of the cut site anneal following strand re-sectioning, referred to as MMEJ (Microhomology-Mediated End Joining; Wang and Xu 2017).

Since both hetero- and homozygous mutations were expected, the genotyping workflow was conceived such that resulting colonies could be screened following selection for

Cas9-2A-Shble expression (Figure 1.2c). Colonies were first picked and added directly to PCR mix to amplify the sgRNA expression cassette on the episome harbored within *Phaeodactylum* (Supplemental Figure 1.4). Only colonies that contained a correctly sized amplicon were chosen for genotyping. While Cas9 expression was not verified for each individual colony, a Cas9-Venus (YFP) fusion was visualized by confocal microscopy to ensure localization in the *Phaeodactylum* nucleus (Supplemental Figure 1.5).

The methods described above were used for the following two applications: single-gene mutagenesis and two-gene mutagenesis.

### 1.2.4.2 Single-gene mutagenesis

Two sgRNAs were individually cloned into the Cas9-Shble episome by golden-gate assembly (cloned episome product can be visualized in Supplemental Figure 1.2a). The sgRNAs were designed to mutate the nitrate reductase (NR) gene in *Phaeodactylum*. NR was picked as a genomic target because the function of NR has been well studied in *Phaeodactylum* and NR knockout cell mutants exhibit an easily screenable growth phenotype of cell death when supplemented with nitrate as a sole nitrogen source (McCarthy et al. 2017). The two sgRNAs were designed to target regions 55-bp apart so that the same amplification and sequencing primers could be used for genotyping efforts.

**1.2.4.2.1 Comparative sgRNA-effectiveness assessment**

To quantify the mutagenesis efficiency for both sgRNAs population-wide, the T7 endonuclease assay was used (Figure 1.3). While the T7 assay has been typically used to genotype a clonal cell line, this method was useful to quantify the mutagenesis efficiency of all cells after transformation and phleomycin-selection of Cas9-2A-Shble. This was done with the intention to identify sgRNAs that may produce low rates of mutagenesis and subsequently discard them. The sgRNAs were designed to target the NR locus 55-bp apart (Figure 1.3a). This allowed both targeted populations (one for gNR-A and one for gNR-B) to be subjected to the same T7 assay parameters (Figure 1.3b). The T7 enzyme cleaves mismatched nucleotides and should cut at Cas9 target loci with NHEJ-mediated mutations. The expected band sizes after T7 digestion was 127-bp and 58-bp for gNR-A and 113-bp and 72-bp for gNR-B (Figure 1.3c). Homoduplexes, annealed bands that perfectly match during "heteroduplex" formation (Figure 1.3b), was expected to be observed in the T7 migration gel because biallelic knockouts that arise from MMEJ-mediated repair may, in theory, anneal to each other and avoid T7 cleavage. It was observed that gNR-A insufficiently targeted the NR locus (0% mutagenesis efficiency) while gNR-B mutated the NR locus at 38% (Figure 1.3d).

By employing this method prior to colony genotyping, sgRNAs with low rates of mutagenesis throughout the population can be ignored which saves on reagent costs and labor. Henceforth, gNR-A was discarded from future targeting experiments due to its low mutagenesis efficiency compared to gNR-B.

### 1.2.4.2.2 Picking and Genotyping NR-KO Mutants

TIDE sequence analysis was used to genotype *Phaeodactylum* cell lines (Figure 1.4). To do so, each sequence of the gNR-B target locus was first curated and analyzed manually. Next, one of two paired-end sequencing files were entered into the TIDE software and compared to a wild-type sequence for each colony. TIDE outputs a bar plot with predicted indel (insertion or deletion) mutations based on the inputted electropherogram compared to the wild-type. In Figure 1.4a, a wild-type NR sequence read is shown above the TIDE result plot that compares wild-type to wild-type. The TIDE plot displays one bar at *x*-axis position "zero" and with a *y*-axis value of 97.1% (displayed above bar). A pink colored bar also indicates the predicted indel was statistically significant ($p < 0.001$). This result indicates that there is a zero-nucleotide indel mutation at the predicted Cas9 cut site that comprises 97.1% of the total sequencing reads. Clearly, the TIDE plot validates the sequencing read as a wild-type genotype.

Figure 1.4b shows six NR knockout mutant genotypes for cell lines 20, 21, 27, 43, 52, and 56. Cell line 20, for example, appears to have a bi-allelic mutation that is homozygous because the TIDE chart shows one pink peak that corresponds to a deletion mutation of 9-bp. Cell line 43, for example, appears to have a bi-allelic mutation that is heterozygous because the TIDE charts shows two pink peaks corresponding to a deletion genotype of 3-bp (28.5%) and 15-bp (49.4%). All six NR-KO genotypes display either homozygous or heterozygous genotypes by their TIDE charts. They also lack any peak at the "zero" position that would indicate that there was a wild-type contamination. Furthermore, cell lines with TIDE charts with only 1 or 2 indel peaks are shown here; additional peaks would indicate that the cell lines were not clonal and need to be re-streaked to find a clonal mutant.

Manual sequencing curation and TIDE analysis was performed for 80 gNR-A colonies and 64 gNR-B colonies. 2/80 (2.5%) gNR-A colonies and 26/64 (40%) gNR-B colonies contained a Cas9-mediated mutation. Here, the Cas9-sgRNA mutagenesis efficiencies of 0% and 28% calculated using T7 (Figure 1.3) reflects the efficiencies observed using TIDE.

Lastly, the nitrate growth assay was used to correlate the TIDE genotyping efforts to the expected NR mutant phenotype. The six mutant cell lines identified using TIDE did not grow on nitrate media compared to growth on ammonium media (Figure 1.4c).

### 1.2.4.3 Two-gene mutagenesis

The Cas9-2A-Shble episome was assembled to harbor two sgRNA, one targeting glutamine synthetase 2 (GS-2, Gene ID: Phatr3_J51092) and one targeting a chloroplast-localized glutamate synthase (cGOGAT, Phatr3_J24739). The episome was assembled by Golden Gate cloning and built to harbor both sgRNA expression cassette flanking a LacZ bacterial expression cassette (Figure 1.5a and Supplemental Figure 1.3).

**Figure 1.5.** Two-gene mutagenesis of the genomic target glutamine synthetase 2 (GS-2, Phatr3_J51092) and the chloroplast localized glutamate synthase (cGOGAT, Phatr3_J24739). **(a)** The assembly Cas9-Shble:sgRNA episome plasmid map with two sgRNA expression cassettes (g51092-4 and g24739-B, green arrows). **(b)** A total of 28 phleomycin-resistant colonies and 4 wild-type cell lines were genotyped for mutations at both target loci. The screening was performed iteratively where all 32 cell lines (mutants and wild-types) were first screened for mutation at the g51092-4 target locus then candidate GS-2-KO cell lines were screened for mutation at the g24739-B locus. 6/28 cell lines exhibited Cas9 activity in 51092 and 4/6 GS-2 mutants exhibited Cas9 activity in 24739. The genotyping results for all cell lines was performed this way (Supplemental Tables 1.3, 1.4). **(c)** Four knock-out cell lines (3, 4, 10, and 15) and their associated TIDE charts show the diversity of mutagenesis at both loci.

### 1.2.4.3.1 TIDE genotyping of two-gene mutants

A total of 28 transformed *Phaeodactylum* colonies and 4 wild-type colonies were picked and genotyped using manual sequencing curation and TIDE sequencing analysis (Figure 1.5). The genotyping was performed successively. First, all 28 colonies were screened for mutation at the GS-2 target locus. Second, GS-2 KO candidate colonies were screened for mutation at the cGOGAT target locus. Therefore, genotyping information is not provided for all 28 colonies for the cGOGAT target locus. The genotyping results using manual curation and TIDE analysis for each colony is provided in Supplemental Table 1.3 (GS-2) and Supplemental Table 1.4 (cGOGAT).

A total of 6/28 (21%) colonies contained a Cas9-mediated mutation(s) at the GS-2 target locus and 4/6 GS-2 mutant candidates also contained Cas9-mediated mutations at the cGOGAT target locus (Figure 1.5b). Figure 4C shows the TIDE plots for each double-knock-out cell line (3, 4, 10, and 15) for both target loci (GS-2 on the left, cGOGAT on the right). Some mutations appear bi-allelic and homozygous, such as cell line 15 at the GS-2 locus (48-bp deletion) and cell line 3 at the cGOGAT locus (24-bp deletion). Only cell line 10 at locus cGOGAT displayed a bi-allelic, heterozygous mutation of a 15-bp deletion and 12-bp deletion. The remaining genotypes were not as clearly defined and have multiple (more than two) peaks in their TIDE chart and/or they contain a trace of a wild-type sequence (both examples are shown in cell line 4 at the cGOGAT locus).

### 1.2.4.4 HDR-mediated gene editing of NR by biolistics

Six cell lines were produced via HDR-mediated gene editing of the NR gene in *Phaeodactylum* (Figure 1.6). It should be noted that, although attempted, HDR-mediated mutagenesis was not observed when the HDR-donor was encoded on the Cas9-Shble episome (results not shown). All six cell lines exhibited the genotype (Figure 1.6b) and phenotype (Figure 1.6c) of having an HDR-mediated insertion of the Shble expression cassette at the gNR-B target locus. An efficiency of 17% for producing a bi-allelic, HDR-mediated Pt_NR mutant was achieved.
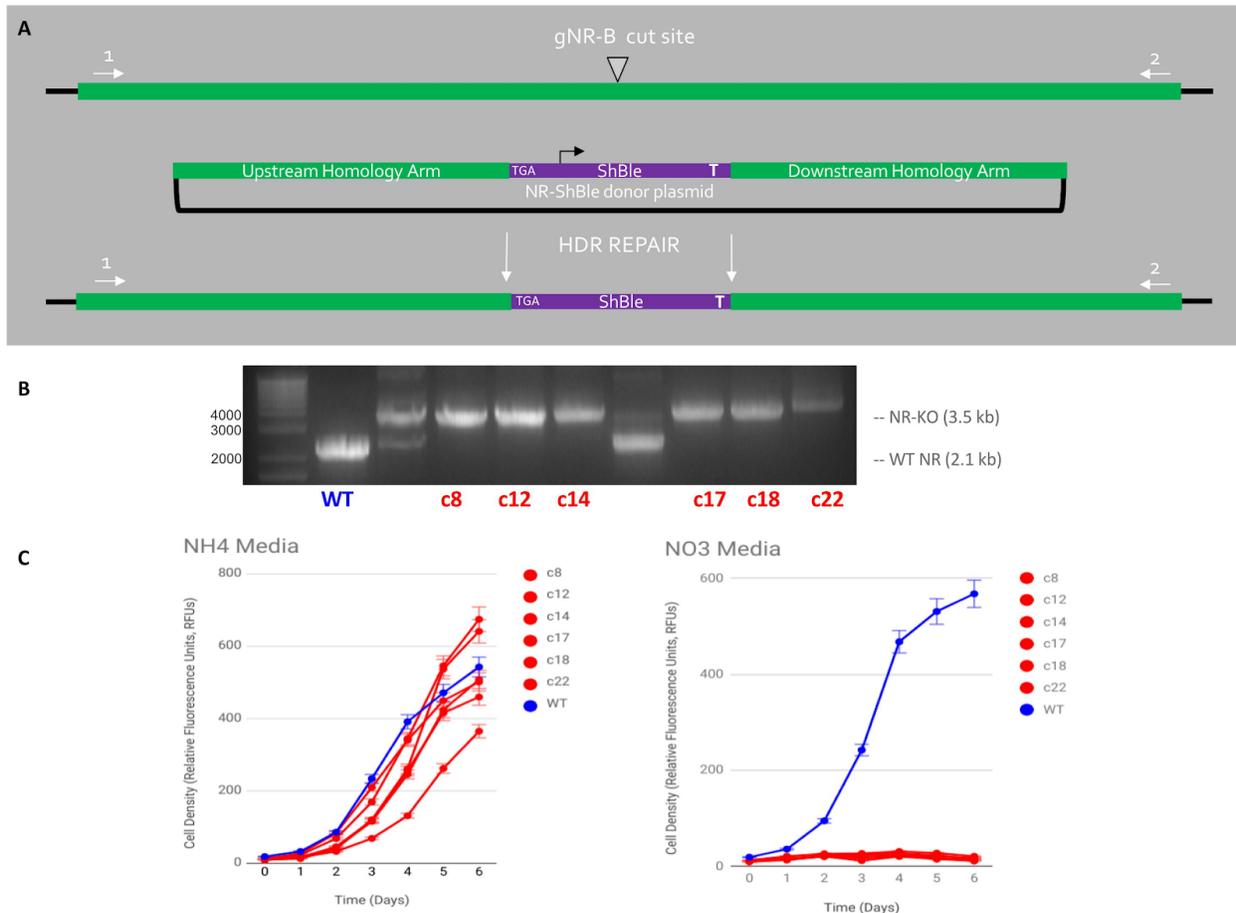
**Figure 1.6.** HDR-mediated CRISPR-Cas9 targeted insertion of Shble gene via micro-particle bombardment genetic transformation. **(a)** An Shble expression cassette with an upstream premature STOP codon was precisely inserted at the gNR-B DSB site Cas9, gNR-B, and a NR-Shble donor plasmid were delivered to *Phaeodactylum* cells on separate plasmids by micro-particle bombardment. The plasmids are presumably integrated within genomic DNA and the CRISPR components are expressed from *Phaeodactylum* chromosomes. **(b)** An electrophoresis gel shows migrated amplicon from six mutant target loci compared to a wild-type NR locus. A single band, rather than two or more, indicated a bi-allelic HDR-mediated edit of the gNR-B target. **(c)** All six mutant cell lines could not grow on nitrate-supplemented media compared to wild-type but could grow on ammonium-supplemented media, a confirmation of the predicted NR-KO mutant phenotype. All cell lines were grown in three technical replicates.

## 1.2.5 Discussion

CRISPR-Cas9 mutagenesis was performed in the model diatom *Phaeodactylum* specifically to streamline the use of bacterial-conjugation transformation. First, the episomal delivery system for Cas9 was re-engineered so Cas9 could be selected for by phleomycin-antibiotic pressure. The 2A peptide transcriptionally fused Cas9 and the selective

gene, Shble, to ensure Cas9 could be selected for after conjugation. Previous attempts to deliver Cas9 in an expression cassette separate from the Shble cassette resulted in low mutagenesis efficiency when colonies were not initially screened for a predicted phenotype, such as NR phenotype. Although not reported here, previously published experiments that delivered a non-selectable Cas9 via bacterial conjugation in *Phaeodactylum* resulted in mutagenesis efficiencies ranging from 25 to 33% using one sgRNA target (Sharma et al. 2018). Here, mutagenesis efficiencies ranged from 2 to 40% when targeting the nitrate reductase gene. Despite reporting a higher mutagenesis efficiency when using gNR-B, mutagenesis efficiencies cannot be directly compared when targeting two different genomic loci.

It has been observed that only 10% of transformed colonies contained a full Cas9 expression cassette, drastically lowering the mutagenesis efficiency (results not shown). It has been previously reported that the *Phaeodactylum* episome, naked, only retains its full plasmid sequence after transfer from *E. coli* to *Phaeodactylum* at 30% (Karas et al. 2015; Diner et al. 2016). Such a Cas9 episome design would not be ideal for targets without a previously known phenotype. The 2A peptide proved valuable in increasing the efficiency of Cas9 delivery to *Phaeodactylum* and in identifying mutant cell lines prior to checking for a phenotype response. Due to a transcriptional fusion to Shble, it was presumed that the full Cas9 coding region was retained in all transformed *Phaeodactylum* transformants, though PCR amplification of the Cas9 ORF was not performed post-conjugation.

Although not thoroughly quantified, selection for Cas9 seems to alleviate previously reported problems of producing non-clonal *Phaeodactylum* that led to inevitable re-streaking, waiting for clonal cell lines to appear, and re-picking (Weyman et al. 2014). TIDE sequence analysis provided a predictive genotype for each cell line that can be interpreted as homozygous,

heterozygous, or mixed genotype. For gNR-B, 26/64 cell lines exhibited Cas9 activity while 6 exhibited either a homo- or heterozygous genotype by TIDE analysis (Figure 1.4c), which can be interpreted to mean that those cell lines were clonal. However, 20/64 lines that passed the same efficiency criteria (in methods) contained more than two peaks in the TIDE chart or contained a trace wild-type sequence and therefore they were not considered clonal. Clonality was inherently tested for, also, when the NR-KO cell lines were subjected to a nitrate growth assay. If a wild-type cell contaminated the culture there would be growth on nitrate. Nevertheless, none of the six NR-KO cell lines grew on nitrate.

While TIDE sequence analysis is a useful tool to quickly screen raw sequencing reads for Cas9-induced indels. Here, TIDE was useful in identifying cell lines that contained targeted Cas9 activity and in inferring genotype, however, it is not recommended to use TIDE to confirm a cell lines genotype. Rather, the gold standard of sub-cloning the target locus into a TOPO vector followed by sequencing is recommended after TIDE analysis. It should be noted that TIDE analysis is complementary to CRISPR-Cas9 activity factor (CAF) analysis that was also designed to deconvolve diploid genotypes after CRISPR-Cas9 targeting (Stukenberg et al. 2018).

Re-streaking *Phaeodactylum* mutant is also recommended after TIDE analysis. The TIDE chart generated for gNR-B cell lines (Figure 1.4) suggest that the genotypes are either homozygous or heterozygous and that the cell lines are clonal. Nevertheless, TIDE is not the gold-standard for genotyping and assessment of clonality. As shown in Figure 1.5, not all TIDE plots indicate clonality and therefore re-streaking would be necessary to isolate double-knockout mutant cell lines. All supplemental tables outline the genotyping efforts performed in this study. The last column for each table indicates the predicted genotype of the colony and suggests whether or not streaking the culture is recommended.

Lastly, the new Cas9 episome design decreased the time to produce and validate mutant cell lines. The time to produce visible colonies was 10–14 days, genotyping was 3 days, and phenotype screening was 10 days after re-picking. In total, the production of 6 *Phaeodactylum* mutants for the NR gene took 3–4 weeks compared to 4–8 weeks when using particle bombardment (Sharma et al. 2018). In a rapidly developing research landscape, optimization of time to produce mutant cell lines should not be a time-limiting step when there are questions regarding biological processes, ecological relevance, or biotechnology utility is an important goal.

## 1.2.6 Data Availability Statement

The datasets generated for this study can be found in the MM Frontiers 2-27-29 Data Availability Document[7].

Plasmid designed and mentioned in this manuscript are publicly available for purchase at the plasmid repository addgene.com.

## 1.2.7 Author Contributions

---

[7] ttps://docs. google.com/spreadsheets/d/1snw_Jao9sdCDy2Y7cC3tP33cNzZ0 xTxOQTdHUEaARME/edit#gid=2069237253

MM was responsible for all of the molecular cloning, genetic transformations, genotyping assays, and writing. VB helped in plasmid and episome design, specifically for the conjugation protocols and implementation of the 2A peptide, and other intellectual contributions. PG performed the phenotype assays. MT designed golden gate primers for the episome. JM helped in daily tasks and manuscript writing. AA contributed to intellectual design and scope of the study. All authors contributed to the editing and feedback for this manuscript.

### 1.2.8 Funding Acknowledgement

### 1.2.9 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### 1.2.10 Supplemental Material

The Supplemental Material for this article can be found in the Supplemental Figure File, Supplemental Information and Table file, and online[8].

## 1.3 Acknowledgements of Chapter 1

Section 1 of Chapter 1, in part, is material in revision to be published as it may appear in The Molecular Life of Diatoms, Chapter 26 (2021). Moosburner, M., Allen, A. E., & Daboussi, F. (2021) Genetic Engineering in Marine Diatoms: Current Practices and Emerging Technologies. In A. Falciatore, T. Mock (eds.), *The Molecular Life of Diatoms.* Chapter 26. Springer Press, London, UK. In Revision. Mark Moosburner was the primary author and principal investigator of this manuscript. Fazya Daboussi was the secondary author of this manuscript. Andrew. E. Allen was the third author of this manuscript. All members contributed to revision and editing of the manuscript.

Section 2 of Chapter 1 is, in full, a reprint of the manuscript as it appears in Frontiers in microbiology, 2020. Moosburner, M. A., Gholami, P., McCarthy, J. K., Tan, M., Bielinski, V. A., & Allen, A. E. (2020). Multiplexed knockouts in the model diatom Phaeodactylum by episomal delivery of a selectable Cas9. *Frontiers in microbiology, 11*, 5. Mark Moosburner was the primary author and principal investigator of this manuscript.

---

[8] https://www.frontiersin.org/articles/10.3389/fmicb.2020.00005/full#supplementary-material

## 1.4 References of Chapter 1

Allen, A. E., C. L. Dupont, M. Oborník, A. Horák, A. Nunes-Nesi, J. P. McCrow, H. Zheng, D. A. Johnson, H. Hu, A. R. Fernie and C. Bowler (2011). "Evolution and metabolic significance of the urea cycle in photosynthetic diatoms." Nature 473(7346): 203-207.

Apt, K. E., A. R. Grossman and P. G. Kroth-Pancic (1996). "Stable nuclear transformation of the diatomPhaeodactylum tricornutum." Molecular and General Genetics MGG **252**(5): 572-579.

Armbrust, E. V., J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kröger, W. W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson and D. S. Rokhsar (2004). "The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism." Science **306**(5693): 79-86.

Bailleul, B., A. Rogato, A. De Martino, S. Coesel, P. Cardol, C. Bowler, A. Falciatore and G. Finazzi (2010). "An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light." Proceedings of the National Academy of Sciences **107**(42): 18214-18219.

Becker, J. and C. Wittmann (2018). "From systems biology to metabolically engineered cells—an omics perspective on the development of industrial microbes." Current opinion in microbiology **45**: 180-188.

Bielinski, V. A., T. M. Bolt, C. L. Dupont and P. D. Weyman (2017). Episomal tools for RNAi in the diatom Phaeodactylum tricornutum, PeerJ Preprints.

Bowler, C., A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, R. P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. A. Berges, C. Brownlee, J. P. Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. Huysman, B. D. Jenkins, K. Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kroger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin-Jezequel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M. P. Oudot-Le Secq, C. Napoli, M. Obornik, M. S. Parker, J. L. Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. A. Rynearson, J. Schmutz,

H. Shapiro, M. Siaut, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van de Peer and I. V. Grigoriev (2008). "The Phaeodactylum genome reveals the evolutionary history of diatom genomes." Nature **456**(7219): 239-244.

Brinkman, E. K., T. Chen, M. Amendola and B. van Steensel (2014). "Easy quantitative assessment of genome editing by sequence trace decomposition." Nucleic acids research **42**(22): e168-e168.

Cermeño, P., J.-B. Lee, K. Wyman, O. Schofield and P. G. Falkowski (2011). "Competitive dynamics in two species of marine phytoplankton under non-equilibrium conditions." Marine ecology progress series **429**: 19-28.

Cerutti, H. and J. A. Casas-Mollano (2006). "On the origin and functions of RNA-mediated silencing: from protists to man." Current genetics **50**(2): 81-99.

Cho, S. W., S. Kim, J. M. Kim and J.-S. Kim (2013). "Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease." Nature biotechnology **31**(3): 230-232.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science **339**(6121): 819-823.

Daboussi, F., S. Leduc, A. Marechal, G. Dubois, V. Guyot, C. Perez-Michaut, A. Amato, A. Falciatore, A. Juillerat, M. Beurdeley, D. F. Voytas, L. Cavarec and P. Duchateau (2014). "Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology." Nat Commun **5**: 3831.

Daboussi, F., T. J. Stoddard and F. Zhang (2015). Engineering Meganuclease for Precise Plant Genome Modification. Advances in New Technology for Targeted Modification of Plant Genomes**:** 21-38.

Davis, A. J. and D. J. Chen (2013). "DNA double strand break repair via non-homologous end-joining." Translational cancer research **2**(3): 130.

De Riso, V., R. Raniello, F. Maumus, A. Rogato, C. Bowler and A. Falciatore (2009). "Gene silencing in the marine diatom Phaeodactylum tricornutum." Nucleic Acids Res **37**(14): e96.

Delacote, F., C. Perez, V. Guyot, M. Duhamel, C. Rochon, N. Ollivier, R. Macmaster, G. H. Silva, F. Paques, F. Daboussi and P. Duchateau (2013). "High frequency targeted mutagenesis using engineered endonucleases and DNA-end processing enzymes." PLoS One **8**(1): e53217.

Deltcheva, E., K. Chylinski, C. M. Sharma, K. Gonzales, Y. Chao, Z. A. Pirzada, M. R. Eckert, J. Vogel and E. Charpentier (2011). "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Nature **471**(7340): 602-607.

Diner, R. E., V. A. Bielinski, C. L. Dupont, A. E. Allen and P. D. Weyman (2016). "Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods." Front Bioeng Biotechnol **4**: 65.

Diner, R. E., C. M. Noddings, N. C. Lian, A. K. Kang, J. B. McQuaid, J. Jablanovic, J. L. Espinoza, N. A. Nguyen, M. A. Anzelmatti, Jr., J. Jansson, V. A. Bielinski, B. J. Karas, C. L. Dupont, A. E. Allen and P. D. Weyman (2017). "Diatom centromeres suggest a mechanism for nuclear DNA acquisition." Proc Natl Acad Sci U S A **114**(29): E6015-E6024.

Duchateau, P. D., F. Daboussi (2014). Method for Targeted Modification of Algae Genomes. Patent Cooperation Treaty. W. I. P. Organization. France. WO2014076571A9.

Dunahay, T. G., E. E. Jarvis and P. G. Roessler (1995). "Genetic transformation of the diatoms Cyclotella cryptica and Navicula saprophila." Journal of Phycology **31**(6): 1004-1012.

Engler, C., R. Kandzia and S. Marillonnet (2008). "A one pot, one step, precision cloning method with high throughput capability." PLoS One **3**(11): e3647.

Fabris, M., J. George, U. Kuzhiumparambil, C. A. Lawson, A. C. Jaramillo-Madrid, R. M. Abbriano, C. E. Vickers and P. Ralph (2020). "Extrachromosomal Genetic Engineering of the Marine Diatom Phaeodactylum tricornutum Enables the Heterologous Production of Monoterpenoids." ACS Synth Biol **9**(3): 598-612.

Faktorova, D., R. E. R. Nisbet, J. A. Fernandez Robledo, E. Casacuberta, L. Sudek, A. E. Allen, M. Ares, Jr., C. Areste, C. Balestreri, A. C. Barbrook, P. Beardslee, S. Bender, D. S. Booth, F. Y. Bouget, C. Bowler, S. A. Breglia, C. Brownlee, G. Burger, H. Cerutti, R. Cesaroni, M. A. Chiurillo, T. Clemente, D. B. Coles, J. L. Collier, E. C. Cooney, K. Coyne, R. Docampo, C. L. Dupont, V. Edgcomb, E. Einarsson, P. A. Elustondo, F. Federici, V. Freire-Beneitez, N. J. Freyria, K. Fukuda, P. A. Garcia, P. R. Girguis, F. Gomaa, S. G. Gornik, J. Guo, V. Hampl, Y. Hanawa, E. R. Haro-Contreras, E. Hehenberger, A. Highfield, Y. Hirakawa, A.

Hopes, C. J. Howe, I. Hu, J. Ibanez, N. A. T. Irwin, Y. Ishii, N. E. Janowicz, A. C. Jones, A. Kachale, K. Fujimura-Kamada, B. Kaur, J. Z. Kaye, E. Kazana, P. J. Keeling, N. King, L. A. Klobutcher, N. Lander, I. Lassadi, Z. Li, S. Lin, J. C. Lozano, F. Luan, S. Maruyama, T. Matute, C. Miceli, J. Minagawa, M. Moosburner, S. R. Najle, D. Nanjappa, I. C. Nimmo, L. Noble, A. M. G. Novak Vanclova, M. Nowacki, I. Nunez, A. Pain, A. Piersanti, S. Pucciarelli, J. Pyrih, J. S. Rest, M. Rius, D. Robertson, A. Ruaud, I. Ruiz-Trillo, M. A. Sigg, P. A. Silver, C. H. Slamovits, G. Jason Smith, B. N. Sprecher, R. Stern, E. C. Swart, A. D. Tsaousis, L. Tsypin, A. Turkewitz, J. Turnsek, M. Valach, V. Verge, P. von Dassow, T. von der Haar, R. F. Waller, L. Wang, X. Wen, G. Wheeler, A. Woods, H. Zhang, T. Mock, A. Z. Worden and J. Lukes (2020). "Genetic tool development in marine protists: emerging model organisms for experimental cell biology." Nat Methods **17**(5): 481-494.

Falciatore, A., R. Casotti, C. Leblanc, C. Abrescia and C. Bowler (1999). "Transformation of nonselectable reporter genes in marine diatoms." Marine Biotechnology **1**(3): 239-251.

Falciatore, A., M. Jaubert, J.-P. Bouly, B. Bailleul and T. Mock (2020). "Diatom Molecular Research Comes of Age: Model Species for Studying Phytoplankton Biology and Diversity." The Plant Cell **32**(3): 547-572.

Fischer, H., I. Robl, M. Sumper and N. Kröger (1999). "Targeting and covalent modification of cell wall and membrane proteins heterologously expressed in the diatom Cylindrotheca fusiformis (Bacillariophyceae)." Journal of Phycology **35**(1): 113-120.

Gao, C. (2018). "The future of CRISPR technologies in agriculture." Nature Reviews Molecular Cell Biology **19**(5): 275-276.

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

Gibson, D. G., L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison III and H. O. Smith (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases." Nat Methods **6**(5): 343-345.

Guschin, D. Y., A. J. Waite, G. E. Katibah, J. C. Miller, M. C. Holmes and E. J. Rebar (2010). A rapid and general assay for monitoring endogenous gene modification. Engineered zinc finger proteins, Springer**:** 247-256.

Hempel, F., A. S. Bozarth, N. Lindenkamp, A. Klingl, S. Zauner, U. Linne, A. Steinbüchel and U. G. Maier (2011). "Microalgae as bioreactors for bioplastic production." Microbial Cell Factories **10**(1): 1-6.

Hempel, F. and U. G. Maier (2012). "An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency." Microbial cell factories **11**(1): 1-6.

Heyer, W.-D., K. T. Ehmsen and J. Liu (2010). "Regulation of homologous recombination in eukaryotes." Annual review of genetics **44**: 113-139.

Hopes, A., V. Nekrasov, S. Kamoun and T. Mock (2016). "Editing of the urease gene by CRISPR-Cas in the diatom Thalassiosira pseudonana." Plant Methods **12**: 49.

Huang, W. and F. Daboussi (2017). "Genetic and metabolic engineering in diatoms." Philosophical Transactions of the Royal Society B: Biological Sciences **372**(1728): 20160411.

Huysman, M. J., A. E. Fortunato, M. Matthijs, B. S. Costa, R. Vanderhaeghen, H. Van den Daele, M. Sachse, D. Inzé, C. Bowler and P. G. Kroth (2013). "AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (Phaeodactylum tricornutum)." The Plant Cell **25**(1): 215-228.

Ifuku, K., D. Yan, M. Miyahara, N. Inoue-Kashino, Y. Y. Yamamoto and Y. Kashino (2015). "A stable and efficient nuclear transformation system for the diatom Chaetoceros gracilis." Photosynth Res **123**(2): 203-211.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier (2012). "A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity." science **337**(6096): 816-821.

Jinek, M., A. East, A. Cheng, S. Lin, E. Ma and J. Doudna (2013). "RNA-programmed genome editing in human cells." elife **2**: e00471.

Johansson, O. N., M. Topel, M. I. M. Pinder, O. Kourtchenko, A. Blomberg, A. Godhe and A. K. Clarke (2019). "Skeletonema marinoi as a new genetic model for marine chain-forming diatoms." Sci Rep **9**(1): 5391.

Karas, B. J., R. E. Diner, S. C. Lefebvre, J. McQuaid, A. P. Phillips, C. M. Noddings, J. K. Brunson, R. E. Valas, T. J. Deerinck, J. Jablanovic, J. T. Gillard, K. Beeri, M. H. Ellisman, J. I. Glass, C. A. Hutchison, 3rd, H. O. Smith, J. C. Venter, A. E. Allen, C. L. Dupont and P.

D. Weyman (2015). "Designer diatom episomes delivered by bacterial conjugation." <u>Nat Commun</u> **6**: 6925.

Karas, B. J., B. Molparia, J. Jablanovic, W. J. Hermann, Y.-C. Lin, C. L. Dupont, C. Tagwerker, I. T. Yonemoto, V. N. Noskov, R.-Y. Chuang, A. E. Allen, J. I. Glass, C. A. Hutchison, H. O. Smith, J. C. Venter and P. D. Weyman (2013). "Assembly of eukaryotic algal chromosomes in yeast." <u>Journal of Biological Engineering</u> **7**(1): 30.

Kaur, S. and C. Spillane (2015). "Reduction in carotenoid levels in the marine diatom Phaeodactylum tricornutum by artificial microRNAs targeted against the endogenous phytoene synthase gene." <u>Mar Biotechnol (NY)</u> **17**(1): 1-7.

Kim, J. H., S.-R. Lee, L.-H. Li, H.-J. Park, J.-H. Park, K. Y. Lee, M.-K. Kim, B. A. Shin and S.-Y. Choi (2011). "High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice." <u>PloS one</u> **6**(4): e18556.

Kroth, P. G., A. M. Bones, F. Daboussi, M. I. Ferrante, M. Jaubert, M. Kolot, M. Nymark, C. Rio Bartulos, A. Ritter, M. T. Russo, M. Serif, P. Winge and A. Falciatore (2018). "Genome editing in diatoms: achievements and goals." <u>Plant Cell Rep</u> **37**(10): 1401-1408.

LaFave, M. C. and J. Sekelsky (2009). "Mitotic recombination: why? when? how? where?" <u>PLoS Genet</u> **5**(3): e1000411.

Mali, P., K. M. Esvelt and G. M. Church (2013). "Cas9 as a versatile tool for engineering biology." <u>Nature methods</u> **10**(10): 957-963.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville and G. M. Church (2013). "RNA-guided human genome engineering via Cas9." <u>Science</u> **339**(6121): 823-826.

McCarthy, J. K., S. R. Smith, J. P. McCrow, M. Tan, H. Zheng, K. Beeri, R. Roth, C. Lichtle, U. Goodenough, C. P. Bowler, C. L. Dupont and A. E. Allen (2017). "Nitrate Reductase Knockout Uncouples Nitrate Transport from Nitrate Assimilation and Drives Repartitioning of Carbon Flux in a Model Pennate Diatom." <u>Plant Cell</u> **29**(8): 2047-2070.

McGinn, J. and L. A. Marraffini (2019). "Molecular mechanisms of CRISPR–Cas spacer acquisition." <u>Nature Reviews Microbiology</u> **17**(1): 7-12.

McQuaid, J. B., A. B. Kustka, M. Obornik, A. Horak, J. P. McCrow, B. J. Karas, H. Zheng, T. Kindeberg, A. J. Andersson, K. A. Barbeau and A. E. Allen (2018). "Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms." <u>Nature</u> **555**(7697): 534-537.

Miyahara, M., M. Aoi, N. Inoue-Kashino, Y. Kashino and K. Ifuku (2013). "Highly efficient transformation of the diatom Phaeodactylum tricornutum by multi-pulse electroporation." <u>Biosci Biotechnol Biochem</u> **77**(4): 874-876.

Moog, D., S. A. Rensing, J. M. Archibald, U. G. Maier and K. K. Ullrich (2015). "Localization and evolution of putative triose phosphate translocators in the diatom Phaeodactylum tricornutum." <u>Genome biology and evolution</u> **7**(11): 2955-2969.

Moosburner, M. A., P. Gholami, J. K. McCarthy, M. Tan, V. A. Bielinski and A. E. Allen (2020). "Multiplexed Knockouts in the Model Diatom Phaeodactylum by Episomal Delivery of a Selectable Cas9." <u>Front Microbiol</u> **11**: 5.

Muto, M., M. Tanaka, Y. Liang, T. Yoshino, M. Matsumoto and T. Tanaka (2015). "Enhancement of glycerol metabolism in the oleaginous marine diatom Fistulifera solaris JPCC DA0580 to improve triacylglycerol productivity." <u>Biotechnol Biofuels</u> **8**(1): 4.

Nawaly, H., Y. Tsuji and Y. Matsuda (2020). "Rapid and precise genome editing in a marine diatom, Thalassiosira pseudonana by Cas9 nickase (D10A)." <u>Algal Research</u> **47**: 101855.

Niu, Y. F., M. H. Zhang, W. H. Xie, J. N. Li, Y. F. Gao, W. D. Yang, J. S. Liu and H. Y. Li (2011). "A new inducible expression system in a transformed green alga, Chlorella vulgaris." <u>Genet Mol Res</u> **10**(4): 3427-3434.

Nymark, M., A. K. Sharma, T. Sparstad, A. M. Bones and P. Winge (2016). "A CRISPR/Cas9 system adapted for gene editing in marine algae." <u>Sci Rep</u> **6**: 24951.

Pollak, B., T. Matute, I. Nunez, A. Cerda, C. Lopez, V. Vargas, A. Kan, V. Bielinski, P. von Dassow, C. L. Dupont and F. Federici (2020). "Universal loop assembly: open, efficient and cross-kingdom DNA fabrication." <u>Synth Biol (Oxf)</u> **5**(1): ysaa001.

Poulsen, N., P. M. Chesley and N. Kröger (2006). "Molecular Genetic Manipulation of the Diatom Thalassiosira Pseudonana (Bacillariophyceae)." <u>Journal of Phycology</u> **42**(5): 1059-1065.

Poulsen, N. and N. Kroger (2005). "A new molecular tool for transgenic diatoms: control of mRNA and protein biosynthesis by an inducible promoter-terminator cassette." FEBS J 272(13): 3413-3423.

Pulz, O. and W. Gross (2004). "Valuable products from biotechnology of microalgae." Applied microbiology and biotechnology 65(6): 635-648.

Ran, F. A., P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott and F. Zhang (2013). "Genome engineering using the CRISPR-Cas9 system." Nature protocols 8(11): 2281-2308.

Rastogi, A., O. Murik, C. Bowler and L. Tirichine (2016). "PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing." BMC bioinformatics 17(1): 1-4.

Richardson, C., M. E. Moynahan and M. Jasin (1998). "Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations." Genes & development 12(24): 3831-3842.

Rogato, A., H. Richard, A. Sarazin, B. Voss, S. C. Navarro, R. Champeimont, L. Navarro, A. Carbone, W. R. Hess and A. Falciatore (2014). "The diversity of small non-coding RNAs in the diatom Phaeodactylum tricornutum." BMC genomics 15(1): 1-20.

Sabatino, V., M. T. Russo, S. Patil, G. d'Ippolito, A. Fontana and M. I. Ferrante (2015). "Establishment of Genetic Transformation in the Sexually Reproducing Diatoms Pseudo-nitzschia multistriata and Pseudo-nitzschia arenysensis and Inheritance of the Transgene." Mar Biotechnol (NY) 17(4): 452-462.

Sakaguchi, T., K. Nakajima and Y. Matsuda (2011). "Identification of the UMP synthase gene by establishment of uracil auxotrophic mutants and the phenotypic complementation system in the marine diatom Phaeodactylum tricornutum." Plant Physiol 156(1): 78-89.

Sambrook, J. and D. W. Russell (2006). The condensed protocols from molecular cloning: a laboratory manual. (No. Sirsi) i9780879697723).

Sander, J. D. and J. K. Joung (2014). "CRISPR-Cas systems for editing, regulating and targeting genomes." Nature biotechnology 32(4): 347-355.

Sentmanat, M. F., S. T. Peters, C. P. Florian, J. P. Connelly and S. M. Pruett-Miller (2018). "A survey of validation strategies for CRISPR-Cas9 editing." Scientific reports 8(1): 1-8.

Serif, M., G. Dubois, A. L. Finoux, M. A. Teste, D. Jallet and F. Daboussi (2018). "One-step generation of multiple gene knock-outs in the diatom Phaeodactylum tricornutum by DNA-free genome editing." <u>Nat Commun</u> **9**(1): 3924.

Serif, M., B. Lepetit, K. Weißert, P. G. Kroth and C. Rio Bartulos (2017). "A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom Phaeodactylum tricornutum." <u>Algal Research</u> **23**: 186-195.

Shalem, O., N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. S. Mikkelsen, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench and F. Zhang (2014). "Genome-scale CRISPR-Cas9 knockout screening in human cells." <u>Science</u> **343**(6166): 84-87.

Sharma, A. K., M. Nymark, T. Sparstad, A. M. Bones and P. Winge (2018). "Transgene-free genome editing in marine algae by bacterial conjugation - comparison with biolistic CRISPR/Cas9 transformation." <u>Sci Rep</u> **8**(1): 14401.

Shemesh, Z., S. Leu, I. Khozin-Goldberg, S. Didi-Cohen, A. Zarka and S. Boussiba (2016). "Inducible expression of Haematococcus oil globule protein in the diatom Phaeodactylum tricornutum : Association with lipid droplets and enhancement of TAG accumulation under nitrogen starvation." <u>Algal Research</u> **18**: 321-331.

Shrestha, R. P. and M. Hildebrand (2017). "Development of a silicon limitation inducible expression system for recombinant protein production in the centric diatoms Thalassiosira pseudonana and Cyclotella cryptica." <u>Microb Cell Fact</u> **16**(1): 145.

Siaut, M., M. Heijde, M. Mangogna, A. Montsant, S. Coesel, A. Allen, A. Manfredonia, A. Falciatore and C. Bowler (2007). "Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum." <u>Gene</u> **406**(1-2): 23-35.

Slattery, S. S., A. Diamond, H. Wang, J. A. Therrien, J. T. Lant, T. Jazey, K. Lee, Z. Klassen, I. Desgagne-Penix, B. J. Karas and D. R. Edgell (2018). "An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum." <u>ACS Synth Biol</u> **7**(2): 328-338.

Strand, T. A., R. Lale, K. F. Degnes, M. Lando and S. Valla (2014). "A new and improved host-independent plasmid system for RK2-based conjugal transfer." <u>PLoS One</u> **9**(3): e90372.

Stukenberg, D., S. Zauner, G. Dell'Aquila and U. G. Maier (2018). "Optimizing CRISPR/Cas9 for the diatom Phaeodactylum tricornutum." <u>Frontiers in plant science</u> **9**: 740.

Taparia, Y., A. Zarka, S. Leu, R. Zarivach, S. Boussiba and I. Khozin-Goldberg (2019). "A novel endogenous selection marker for the diatom Phaeodactylum tricornutum based on a unique mutation in phytoene desaturase 1." Sci Rep **9**(1): 8217.

Trentacoste, E. M., R. P. Shrestha, S. R. Smith, C. Gle, A. C. Hartmann, M. Hildebrand and W. H. Gerwick (2013). "Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth." Proc Natl Acad Sci U S A **110**(49): 19748-19753.

Vinayak, V., K. M. Manoylov, H. Gateau, V. Blanckaert, J. Hérault, G. Pencreac'h, J. Marchand, R. Gordon and B. Schoefs (2015). "Diatom milking: a review and new approaches." Marine drugs **13**(5): 2629-2665.

Walhout, A. J., G. F. Temple, M. A. Brasch, J. L. Hartley, M. A. Lorson, S. van den Heuvel and M. Vidal (2000). [34] GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes. Methods in enzymology, Elsevier. **328:** 575-IN577.

Wang, H., M. La Russa and L. S. Qi (2016). "CRISPR/Cas9 in genome editing and beyond." Annual review of biochemistry **85**: 227-264.

Wang, H. and X. Xu (2017). "Microhomology-mediated end joining: new players join the team." Cell & bioscience **7**(1): 1-6.

Weyman, P. D., K. Beeri, S. C. Lefebvre, J. Rivera, J. K. McCarthy, A. L. Heuberger, G. Peers, A. E. Allen and C. L. Dupont (2015). "Inactivation of Phaeodactylum tricornutum urease gene using transcription activator-like effector nuclease-based targeted mutagenesis." Plant Biotechnol J **13**(4): 460-470.

Zhou, Y., S. Zhu, C. Cai, P. Yuan, C. Li, Y. Huang and W. Wei (2014). "High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells." Nature **509**(7501): 487-491.

# Chapter 2

# Nitrate assimilation in diatoms and the uncoupling from the lipid stress response by the mutagenesis of a chloroplastic glutamine synthetase gene in *Phaeodactylum tricornutum*

Mark Moosburner

## 2.1 Introduction

Photosynthetic algae produce industrially valuable primary and secondary metabolites such as nutritional fatty acids, food additives, cosmetic pigments, antioxidants, pharmaceuticals, and precursor molecules to diesel fuel (Pulz & Gross 2004, Vinayak, et al. 2015). Advancements in genomic systems biology and molecular/genetic tools present unique opportunities to produce sustainable and economical bio-products in a carbon-neutral context. Marine diatoms, eukaryotic photo-autotrophic algae, are of particular interest due to their ecological proliferation in contemporary oceans and their exceptional metabolic properties. Their impact on the global carbon cycle is extraordinary, contributing up to 45% of the estimated 60 gigatons of carbon fixed in the ocean per year via photosynthetic processes (Nelson et al. 1995). Diatoms also influence marine foods webs immensely by forming large algae blooms in reaction to influxes of nitrate from coastal upwelling regions where diatoms have typically been shown to out-compete other species for nitrate (Savidge et al. 1995, Falkowski et al. 2007, Cermeño et al. 2011). Their hallmark feature is their silicon shell, called a frustule, that is a mesoporous surface produced intracellularly upon cell division that has nanotechnology applications in bio-compatible drug delivery (Delalat et al., 2015) and nano-optics (Ferrera et al. 2014). The bio-precipitation of the frustule makes the diatom the most influential organism in the biogeochemical cycling of silica on Earth (Nelson et al. 1995). They also contain unique metabolic processes, such as a fully-functional urea cycle (Allen, et al. 2011), resulting from two sequential secondary endosymbiotic events that formed the bacillariophyceae class. Due to their ecological dominance and biotechnology potential, the diatom genomes of *Phaeodactylum tricornutum* (Bowler et al. 2008) and *Thalassiosira pseudonana* (Armbrust et al. 2004) were sequenced and paved the way for explorations into diatom genomics and metabolism.

## 2.2 Lipids in diatoms

Marine diatoms hold great promise in biotechnological applications due to their ability to fix large amounts of carbon dioxide and grow rapidly and densely in laboratory culture conditions. It has been theorized that cultivated algae can produce biofuel 6-200X higher than oilseed crop per hectare (Hu et al. 2008, Ramachandra et al. 2009, Vinayak et al. 2015). Additionally, algae raceway ponds and translucent "grow bags" can technically grow anywhere (where light and temperature permits) and do not require in-ground acreage like terrestrial plants.

Diatoms cannot produce biofuel directly but do produce biofuel precursors in the form of storage carbohydrates and neutral lipids. The primary neutral lipids produced by diatoms are triglycerides (TAG). Through a transesterification process, fatty-acid ester derivatives can be produced from TAGs and directly used as fuel in diesel engines. TAG species can also be valuable nutritional supplements. Poly-unsaturated fatty acids (PUFAs) are the primary TAG components of the model diatom *Phaeodactylum*, mainly eicosapentaenoic acid (EPA, C20:5$\omega$3) and docosahexaenoic acid (DHA, C22:6 $\omega$3). EPA is an omega-3 fatty acid that is the main ingredient in fish oil, which has been theorized to have beneficial cardiovascular effects. There are increasing efforts to enhance the cellular capacity to produce lipids by manipulating diatom lipid metabolism directly. A 2011 study aimed to improve the cold flow properties of *Phaeodactylum* fatty acids by shortening fatty acid chain lengths. They accomplished this by expressing two heterologous Acyl-ACP thioesterases that produced the short fatty acids lauric acid (C12:0) and myristic acid (C14:0) (Radakovitz et al. 2011). In a 2013 study, the pivotal TAG catalysis enzyme, diacylglycerol acyltransferase type II (DGAT2), was overexpressed in *Phaeodactylum* and increased neutral lipid content by 35% and EPA by 76% (Niu, et al. 2013). Reverse genetic approaches have also been successful in increasing neutral lipids. For example,

the expression of a lipase enzyme in the diatom *Thalassiosira pseudonana* was down-regulated by silencing RNAs that produced cell lines with 4.2-fold more lipids than wild-type without compromising culture growth (Trentacoste et al. 2013). In 2014, TALEN technology was used to directly mutate a UDP-glucose pyrophosphorylase gene in *Phaeodactylum* that resulted in 45-fold increase in intracellular TAG (Daboussi et al. 2014). Lastly, in 2018, TALENs were also used to target a Hotdog-fold thioesterase gene that decreased beta-oxidation and increased TAG by 1.7-fold (Hao et al. 2018). Currently, CRISPR-Cas9 targeted mutagenesis approaches are quickly being developed for multiple species of diatoms like *Phaeodactylum* (Nymark et al. 2016, Slattery et al. 2018, Serif et al. 2018, Moosburner et al. 2020), and *Thalassiosira pseudonana* (Hopes et al. 2016). The ease to implement Cas9-based mutagenesis in these diatoms opens up future research avenues to manipulate lipid metabolism directly or indirectly in order to boost TAG production.

## 2.3 Lipid stress response in diatoms

Diatoms naturally produce a vast number of high-valued molecules (HVM) that are of industrial interest. Specifically, neutral lipids and TAGs are diatom by-products produced naturally after cellular exposure to environmental stress. The lipid stress response is most notable when diatoms are starved of silicon and nitrogen (Eizadora et al. 2009, Alipanah et al. 2015, Levitan et al. 2015b, Smith et al. 2016, Longsworth et al. 2016). Though lipid accumulation caused by stressful conditions is unfortunately paired with decreased cell division. Diatoms direct lipid synthesis towards the storage of neutral lipids and away from the esterification of glycerol-based lipids and phospholipids used for plasma membrane biosynthesis, which could be a possible explanation for lipid accumulation during vegetative growth when nutrients are

limited (Hu et al. 2008). A long standing goal of diatom research has been to produce a lipid accumulation cell response without impacting cell growth by manipulating nitrogen metabolism. McCarthy et al. 2017 found in a nitrate reductase knockout *Phaeodactylum* cell line that lipids accumulate when in growth medium supplemented with nitrate. Not surprisingly, the cells could not reduce the supplemented nitrate and ceased to grow. Here, cell growth was severely impaired but the study was the first example of a lipid response during nutrient-replete conditions in a diatom. Regardless, the goal of lipid synthesis during cell proliferation was not achieved. Interestingly, it was also observed that *Phaeodactylum* mutants store excess nitrate in their vacuoles and physically dilate. Similar to when carbon is stored in lipid droplets within a *Phaeodactylum* cell during nitrogen stress, *Phaeodactylum* also displays the ability to store nitrogen when the nitrate assimilation pathway is disrupted (McCarthy et al. 2017).

Carbon and nitrogen metabolism are intrinsically linked in diatoms. The relationship is highlighted by the physiological response to accumulate lipid during nitrogen-depleted conditions. There have been recent efforts to characterize nitrogen-related and -responsive genes in *Phaeodactylum* and *Thalassiosira pseudonana* towards understanding nitrogen metabolism generally and identifying nitrogen gene candidates associated with TAG biosynthesis (Allen et al. 2011, Hildebrand et al. 2012, Levitan et al. 2015b, Alipanah et al. 2015, Smith et al. 2016, McCarthy et al. 2017). In general, *Phaeodactylum* coordinates the repartitioning of carbon, specifically towards TAGs, during nitrogen-stressed conditions. Nitrogen-responsive genes are some of the most highly expressed genes in *Phaeodactylum* and especially those involved in nitrogen assimilation into amino acids through the GS-GOGAT cycle (Levitan et al. 2015b) and by a putative alanine-aspartate shuttle (Smith et al. 2019). Specifically, the most up-regulated genes in response to nitrogen depletion are nitrate and urea assimilation genes, GS-GOGAT

genes, pyruvate dehydrogenase genes, and genes involved in the citric acid cycle (Levitan et al. 2015b, McCarthy et al. 2017). As expected, genes associated with photosynthesis including glycolysis, gluconeogenesis, and fucoxanthin chlorophyll a/c binding proteins are the most down-regulated. Interestingly, about a third of genes associated with lipid biosynthesis are down-regulated, about 60% do not change expression levels, and only 8% of genes, and specifically DGAT2, are up-regulated during nitrogen starvation (Levitan et al. 2015b). The gene expression regulation in Levitan et al. (2015b) was also observed in a NR-KO mutant in nitrate growth conditions and most notably the up-regulation of DGAT2, the protein that catalyzes the last reaction in TAG biosynthesis (McCarthy et al. 2017). With DGAT2, a few other TAG regulator protein were also up-regulated in an NR-KO mutant including the protein StLDP (alternatively named MLDP in McCarthy et al. 2017, Phatr3_J48859) that has been identified as the most abundant protein associated with lipid droplets in *Phaeodactylum* (Yoneda et al. 2016). It has been speculated that during environmentally stressful conditions diatoms recycle membrane phospholipids into neutral lipids rather than increase the production of new ones. Transcriptomic analysis supports this notion since most of the lipid biosynthesis genes are not differentially regulated during nitrogen stress (Levitan et al. 2015b) and since a phospholipase gene is up-regulated in a NR-KO cell line that recycles phospholipid disassembly (McCarthy et al. 2017).

## 2.4 Nitrogen metabolism in *Phaeodactylum*

The first step of nitrate assimilation in *Phaeodactylum* is the transport of nitrate into the cell primarily by nitrate transporter 2 (NRT2, Phatr3_J26029, Phatr3_J40691) followed by the reduction of nitrate by nitrate reductase (NR, Phatr3_J54983). Diatoms take up nitrate in excess

and store it in the vacuole or reduce it to nitrite. Nitrate storage seems necessary for diatoms to actively proliferate in oceanic conditions with variable nitrate availability (McCarthy et al. 2017). Nitrite can either be excreted into the environment or further reduced to ammonium by one of two nitrite reductase enzymes (Fd-NiR, Phatr3_J12092; NAD(P)H-NiR, Phatr3_EGO2286). Ammonium can then be directly incorporated into the amino acid glutamine by glutamine synthetase (GS) and used for protein production or as a substrate for glutamate synthesis.

The nitrate assimilation pathway is shunted towards the chloroplast yet greatly influences cellular carbon and nitrogen partitioning between organelles (Figure 2.1). The GS enzyme is part of the GS-GOGAT cycle that coordinates the interconversion of glutamine and glutamate with the enzyme glutamine oxoglutarate aminotransferase or glutamate synthase (GOGAT). GS dephosphorylates ATP to ligate an amino group to glutamate, producing glutamine, ADP, and a hydrogen ion. GOGAT, in turn, synthesizes glutamate using glutamine, 2-oxoglutarate, and a reductant species. Which reductant species depends on which of two GS-GOGAT cycles, chloroplast or mitochondria localized, is utilized. The chloroplast GOGAT (cGOGAT, Phatr3_J24739) reduces oxidized iron while the mitochondria GOGAT (mGOGAT, large subunit, Phatr3_J51214; small subunit, Phatr3_J20342) reduces NAD(P)H. The two GS-GOGAT hubs are also differentially regulated according to a diel light cycle, where both GOGAT enzyme and the chloroplast GS (GS2, Phatr3_J51092) are transcriptionally up-regulated during the day and the mitochondria GS (GS3, Phatr3_J22357) is transcriptionally up-regulated at night and independent of iron availability (Smith et al. 2016). GS3, contrary to GS2 and GOGAT expression, is not coordinated with the regulation of nitrogen assimilation genes NR, NiR, a nitrate transporter (NRT, Phatr3_J26029), and a chloroplast formate/nitrite transporter (FNT,

Phatr3_J13076) (Smith et al 2016, Smith et al. 2019). The transcriptional coordination of GS gene expressions has also been reported in the diatom *Thalassiosira pseudonana* (Brown et al. 2009). Furthermore, flux between GS-GOGAT-related metabolites, and specifically the catabolism of glutamate, increased when limited for nitrate (Levitan et al. 2015b). An iron-influenced coordination of the nitrate assimilation pathway in the chloroplast has been observed (Allen et al. 2008), but is contradicted by a more recent *Phaeodactylum* transcriptome analysis (Smith et al. 2016).

**Figure 2.1.** *Phaeodactylum* metabolic network highlighting nitrogen metabolism including two GS-GOGAT cycles (blue boxes), the putative ornithine-glutamine shunt (purple boxes), the putative aspartate-alanine shuttle (red boxes), the citric acid cycle (light grey box), and the urea cycle (dark grey). Enzymes are colored red and italicized. Metabolites and molecular substrates are black and bolded. The chloroplast (green box), the mitochondria (yellow box), and the cytosol (orange box) are shown. Carbon-based metabolic processes are shown in grey objects with a black border and include glycolysis, fatty acid biosynthesis, and TAG biosynthesis. Enzyme abbreviations used: ACO, aconitase; ACOAT, acetylornithine aminotransferase; AGK, acetylglutamate kinase; AGPR, N-acetyl-gamma-glutamyl-phosphate reductase; AKD, alpha-ketoglutarate dehydrogenase; ALT1, alanine aminotransferase 1; ALT2, alanine aminotransferase 2; ARG, arginase; ASL, argininosuccinate lyase; AST1, aspartate aminotransferase 1; AST2, aspartate aminotransferase 2; ASuS, argininosuccinate synthase; cGOGAT, glutamate synthase (Fd); CS, citrate synthase; FUM, fumarase; GACT, glutamate N-acetyltransferase; GDH, glutamate dehydrogenase; GLS, glutamate semialdehyde degradation (spontaneous); GS2, glutamine synthetase 2; GS3, glutamine synthetase 3; ICD, isocitrate dehydrogenase; MD, malate dehydrogenase; mGOGAT, glutamate synthase (NAD(P)H); NAGS, N-acetyl-glutamate synthase; NiR, nitrite reductase; NR, nitrate reductase; NRT, nitrate transporter; OAT, ornithine aminotransferase; OTC, ornithine carbamoyltransferase; P5CD, 1-pyrroline-5-carboxylate dehydrogenase; PC, pyruvate carboxylase; PD, pyruvate dehydrogenase; SCS, succinyl-CoA synthase; SD, succinic dehydrogenase; unCPS, carbamoyl phosphate synthase; URE, urease. Metabolite abbreviations used: 1P5C, (S)-1-pyrroline-5-carboxylate; 2OG, 2-oxoglutarate; A-CoA, acetyl-CoA; ACO, cis-aconitate; ACG, N-acetyl-L-glutamate; AGS, arginosuccinate; ALN, alanine; ARG, arginine; ASP, aspartate; CT, citrate; CIT, citrulline; CP, carbamoyl phosphate; FUM, fumarate; GL5A, glutamate 5-semialdehyde; GLN, glutamine; GLU, glutamate; ICT, isocitrate; MAL, malate; NAO, N-acetylornithine; NAGP, N-acetyl-L-glutamate 5-phosphate; NG5S, N-acetyl-L-glutamate 5-semialdehyde; NH4, ammonium; NO2, nitrite; NO3, nitrate; OAA, oxaloacetate; ORN, ornithine; PYR, pyruvate; SUC, succinate; SUCA, succinyl-CoA; UREA; urea.
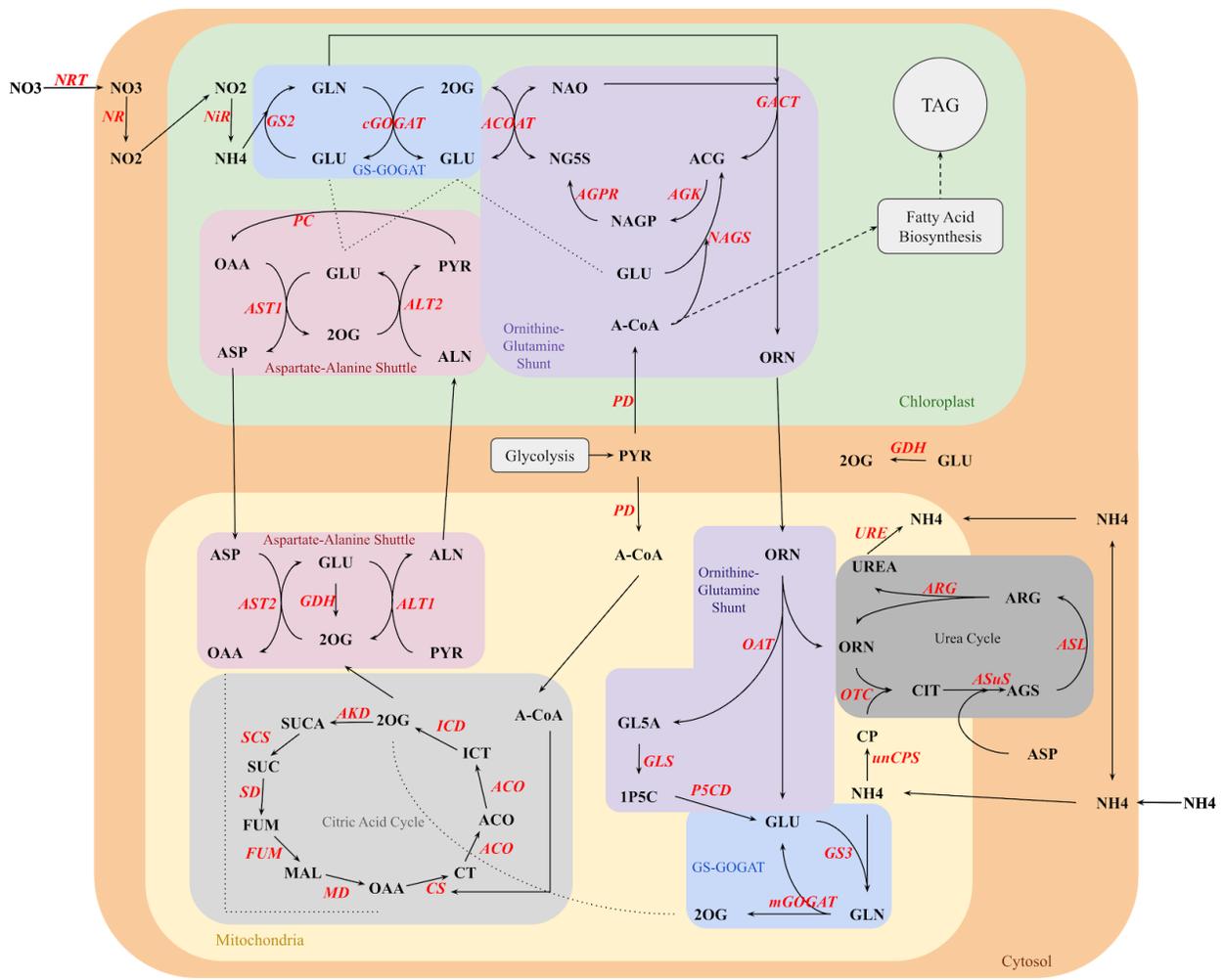
Two fully compartmentalized GS-GOGAT systems are a unique metabolic adaptation only found in diatoms (Figure 2.1, GS-GOGAT cycles). Vascular plants contain a fully GS-GOGAT cycle within the chloroplast while other GS-GOGAT-related enzymes are localized to the cytosol, mitochondria, and peroxisome. In higher plant species such as tobacco (Masclaux-Daubresse et al. 2006), *Lotus japonicus* (Betti et al. 2014) and a model plant *Arabidopsis thaliana* (Lancien et al. 2000, Ferreira et al. 2019) a full GS-GOGAT cycle is located in the chloroplast and a GS species, referred to as GS1, in the cytosol without a paired GOGAT enzyme. Green algae have been reported to mimic the GS-GOGAT intercellular organization as plants and have multiple GS1 isozymes in their cytosol (Sanz-Luque et al. 2015). Evolutionary divergence of diatoms from plants and green algae has been shown in the context of GS enzymes found in the diatom *Skeletonema costatum* (Robertson & Tartar 2006). It is also speculated that diatoms may have an evolutionarily cosmopolitan nitrate assimilation pathway that consists of enzymes from oomycete (theoretical secondary endosymbiotic host) and red algal (endosymbiont) lineages (Smith et al. 2019). Here, *Phaeodactylum* NR and GS2 enzymes align most similarly with a red algal endosymbiont where other assimilation enzymes like a nitrate transporter and NAD(P)H-NiR enzyme originate from a heterotrophic oomycete host (Smith et al. 2019). This mix-and-match pathway appears to be evolutionarily advantageous due to the dominance of diatoms amongst other phytoplankton in marine ecosystems. Further speculation surrounds the origin of GS3 in the mitochondria. GS3 may most closely align with a bacterial GS, typically denoted GlnA, and has been theorized to have been acquired in diatoms through horizontal gene transfer after the secondary endosymbiotic convergence where GS2 was acquired (Allen et al. 2011, Smith et al. 2019). This is not unique to GS3 since hundreds of diatom genes originate from a bacterial source (Bowler et al. 2008).

Unlike other photosynthetic organisms, diatoms have acquired an additional nitrogen assimilation pathway through the mitochondria GS-GOGAT that can impact amino acid biosynthesis and carbon skeleton generation via interactions with tricarboxylic acid (TCA) cycle metabolites (Figure 2.1, Citric Acid Cycle) (Smith et al. 2019). 2-oxoglutarate (2OG) and glutamate are the key carbon skeletons that interact between the TCA and GS-GOGAT cycles in plants, green algae, and diatoms. In the GS-GOGAT cycle, an amino group is transferred to 2-OG from glutamine by GOGAT to produce two glutamate molecules. 2-OG is produced by glutamate dehydrogenase (GDH) and isocitrate dehydrogenase (ICD). The regulation of GS-GOGAT is intrinsically impacted by GDH activity that produces 2-OG from the reduction of glutamate (in the cytosol and mitochondria). ICD activity also impacts 2-OG production but in only the mitochondria via the TCA cycle. The TCA cycle is a metabolic hub that connects glycolysis and protein production. By consuming acetyl-CoA, the TCA cycle produces NADH reductant that is used in the electron transport chain to produce ATP. Along with 2-OG, oxaloacetate (OAA) is another amino acid precursor metabolite produced in the TCA cycle. OAA and 2OG are also enzymatically connected by aspartate aminotransferase (AST) that converts OAA to 2OG in conditions of high glutamate concentrations (Allen et al. 2011, Smith et al. 2019). 2OG production impacts the TCA cycle directly, however glutamate indirectly impacts the levels of 2OG and OAA and could influence the energetic homeostasis of the cell.

Energetic exchanges between the chloroplast and mitochondria may also be occurring by the inter-organelle exchange of alanine and aspartate (Figure 2.1, Aspartate-Alanine Shuttles). In *Phaeodactylum*, the chloroplast contains one pair of AST and alanine aminotransferase (ALT) enzymes (AST1, ALT2) while the mitochondria contains a second pair (AST2, ALT1). On nitrate supplemented growth conditions, AST1 condenses OAA and glutamate to produce glutamate and

aspartate in the chloroplast. Aspartate is then transported to the mitochondria where it is deaminated by AST2 to produce glutamate. On urea supplemented growth, ALT1 condenses pyruvate and glutamate to produce 2OG and alanine in the mitochondria. Alanine is then transported to the chloroplast and deaminated by ALT2 to produce glutamate and pyruvate. Effectively, amino moieties are exchanged between the two organelles either on oxaloacetate (chloroplast to mitochondria) or pyruvate (mitochondria to chloroplast) carbon skeletons (Smith et al. 2019). In both scenarios, the exchange of aspartate and alanine effectively resupplies glutamate to the respective organelles. Glutamate, therefore, may be a significant signaling molecule that connects nitrogen assimilation and the TCA cycle. In plants, glutamate is the central metabolite that regulates amino acid production and the primary signaling molecule for nitrogen metabolism (Forde & Lea 2007, Nunes-Nesi et al. 2010). In animals, glutamate is the most abundant neurotransmitter and a precursor for the neuronal inhibitor, GABA. Like plants, diatoms nitrogen metabolism may be dictated by glutamate production.

Other enzymes potentially impact glutamate levels, too. N-acetylglutamate synthase (NAGS) in the chloroplast converts glutamate and acetyl-CoA into N-acetylglutamate (ACG), the start to one of two pathways for chloroplastic ornithine production (Figure 2.1, Ornithine-Glutamine Shunts). Acetylornithine aminotransferase (ACOAT) in the chloroplast converts glutamate and N-acetyl-L-glutamate 5-semialdehyde (NG5S) to 2-OG and N-acetyl-L-ornithine (NAO). GACT then condenses N-acetyl-L-ornithine and glutamine to produce N-acetylglutamate and ornithine. In the mitochondria, ornithine can be used as a substrate in the urea cycle or deaminated by ornithine aminotransferase (OAT) to produce glutamate 5-semialdehyde and glutamate. The conversion between glutamate and ornithine in both the chloroplast and mitochondria hints at organelle cross-talk. There is precedence in

suggesting a mitochondria-chloroplast crosstalk signaling pathway reliant on a proposed glutamine-ornithine shunt in *Phaeodactylum* (Levering et al. 2016, Smith, et al. 2016). Glutamine and ornithine contain the same number of carbon (five) and nitrogen (two) atoms. It has been bioinformatically predicted that increased reductant and reactive oxygen species in the chloroplast could drive signaling to the mitochondria in the form of ornithine (Levering et al. 2016). The production of ornithine acts as an electron sink in the chloroplast during high-light stress by utilizing four electrons produced by the Calvin-Benson carbon fixation processes. The chloroplastic GS-GOGAT cycle not only reduces reductants (incorporates ammonium and glutamate to make glutamine) but also generates new glutamate. The production of glutamine, glutamate, ornithine, and the activity of GS-GOGAT in general could influence metabolic flux between the two energetic hubs of a *Phaeodactylum* cell but has yet to be experimentally validated.

The experimental and bioinformatically predicted evidence of nitrogen metabolism in *Phaeodactylum* points to the GS-GOGAT paired cycles as an important metabolic hub for repackaging and redistributing carbon and nitrogenous compounds within and between the chloroplast and mitochondria. Figure 2.1 displays a nitrogen metabolic network that combines the GS-GOGAT cycles, the predicted ornithine-glutamine shunt (Levering et al. 2016), and the aspartate-alanine shuttle (Smith et al. 2019). Since ammonium is both a cellular reductant species and a product of nitrogen assimilation in diatoms, GS and GOGAT enzymes are particularly important in nitrogen usage and carbon partitioning and are potential gene candidates to characterize further for their role in nitrogen stress-related lipid accumulation in *Phaeodactylum*. In particular, GS2 serves as the receiver enzyme in the nitrate assimilation pathways of ammonium ions in the chloroplast. It's response to an influx of nitrate into the cell is also strong

and strongly coordinated with other enzymes like nitrate reductase and nitrate transporters (Smith et al. 2019). Targeting GS2 for functional genetics experiments may provide insights into the nitrate assimilation pathway generally and insights into the physiological and transcriptional consequences of losing GS2 function. A GS2-knockout cell line may also prove to be a good model to test how effective GS3 is as a substitute for glutamine synthetase activity, if at all. Lastly, a GS2 mutant may display nitrogen deficient characteristics, like an NR mutant did (McCarthy et al. 2017), such as vacuole dilation or early onset lipid biosynthesis.

## 2.5 Functional genetics of GS-GOGAT genes in *Phaeodactylum*

To begin investigating the function of GS-GOGAT enzymes, CRISPR-Cas9 mutant cell lines were produced for individual genes and pairs of GS-GOGAT genes. There are four genes that were targeted for mutagenesis; GS2 (Phatr3_J51092), GS3 (Phatr3_J22357), cGOGAT (Phatr3_J24739), and mGOGAT large subunit (Phatr3_J51214). Mutant cell lines were produced for each individual gene. Also, the entire GS-GOGAT cycle to the chloroplast and to the mitochondria, respectively, were functionally mutated. This involved producing two-gene knockout cell lines for the chloroplastic GS-GOGAT (GS2 and cGOGAT) and two-gene knockout cell lines for the mitochondria GS-GOGAT (GS3 and mGOGAT). All CRISPR-Cas9 mutant cell lines were generated by episomal CRISPR-Cas9 mutagenesis methods outlined in Moosburner et al. (2020).

### 2.5.1 GS-GOGAT mutant growth characterization

Phenotypic growth assays on nitrate-supplemented media were performed for all six mutant cell lines (Figure 2.2). Pt_GS2-KO (Figure 2.2a), Pt_cGOGAT-KO (Figure 2.2b), and Pt_GS2, cGOGAT-KO (Figure 2.2c) cell lines grew normally compared to wild-type

*Phaeodactylum*. For Pt_GS2-KO, normal growth on nitrate media was unexpected since the nitrate assimilation pathway shunts its reduced product, ammonium, exclusively to the chloroplast where GS2 resides. This finding was the first piece of evidence that a reconfiguration of nitrate assimilation, and specifically the incorporation of ammonium into amino acids, in a GS2 mutant was happening to retain normal growth. It was even more surprising to find that the functional inactivation of the entire chloroplastic GS-GOGAT cycle (Pt_GS2, cGOGAT-KO mutant) also retained normal growth on nitrate media. Seemingly, the mitochondria GS-GOGAT cycle can compensate for a lack of chloroplast GS-GOGAT activity in nitrate growth conditions. Similar growth compensations in GS2 mutants have been reported in the barley plant (Wallsgrove et al. 1987), the legume plant *Lotus japonicus* (Betti et al. 2006, Pérez-Delgado et al. 2013), and a model flowering plant *Arabidopsis thaliana* (Ferreira et al. 2019). In these studies the plant growth rate was retained compared to wild-type, however physiological consequences were found in the mutants such as decreased plant size, decreased growth during photorespiratory conditions, and elevated ammonia levels in leafs. At the transcript level, *Lotus japonicus* and *Arabidopsis thaliana* mutants deficient of GS2 was compensated by an up-regulation of a cytosolic GS gene (GLN1) (Pérez-Delgado et al. 2013, Ferriera et al. 2019).

As for the mitochondria GS-GOGAT mutants, only Pt_GS3-KO cell lines grew normally compared to wild-type (Figure 2.2d). Pt_mGOGAT-KO growth decreased compared to wild-type (Figure 2.2e) and Pt_GS3, mGOGAT-KO growth was even further inhibited compared to Pt_mGOGAT-KO and wild-type (Figure 2.2f). The functional knockout of mGOGAT appears to negatively impact cell growth on nitrate media and the functional knockout of the entire mitochondrial GS-GOGAT cycle appears to negatively impact growth even greater. These

findings support the theory that the mitochondrial GS-GOGAT cycle is important for maintaining cellular proliferation in nitrate growth conditions.

The compensatory mechanisms that retain growth in Pt_GS2-KO mutants could also lead to potential physiological consequences as seen in plants. Though, unlike plants, diatoms do not contain cytosolic GS enzymes and therefore transcriptional compensation would have to occur by other means like GS3 or GDH activity. Considering that the environmental stress response in diatoms is to produce lipids, a GS2 mutant may increase its intracellular lipid biosynthesis while retaining robust growth. The following draft manuscript outlines the experimental findings that show increased lipid accumulation in normally growing GS2 mutants.
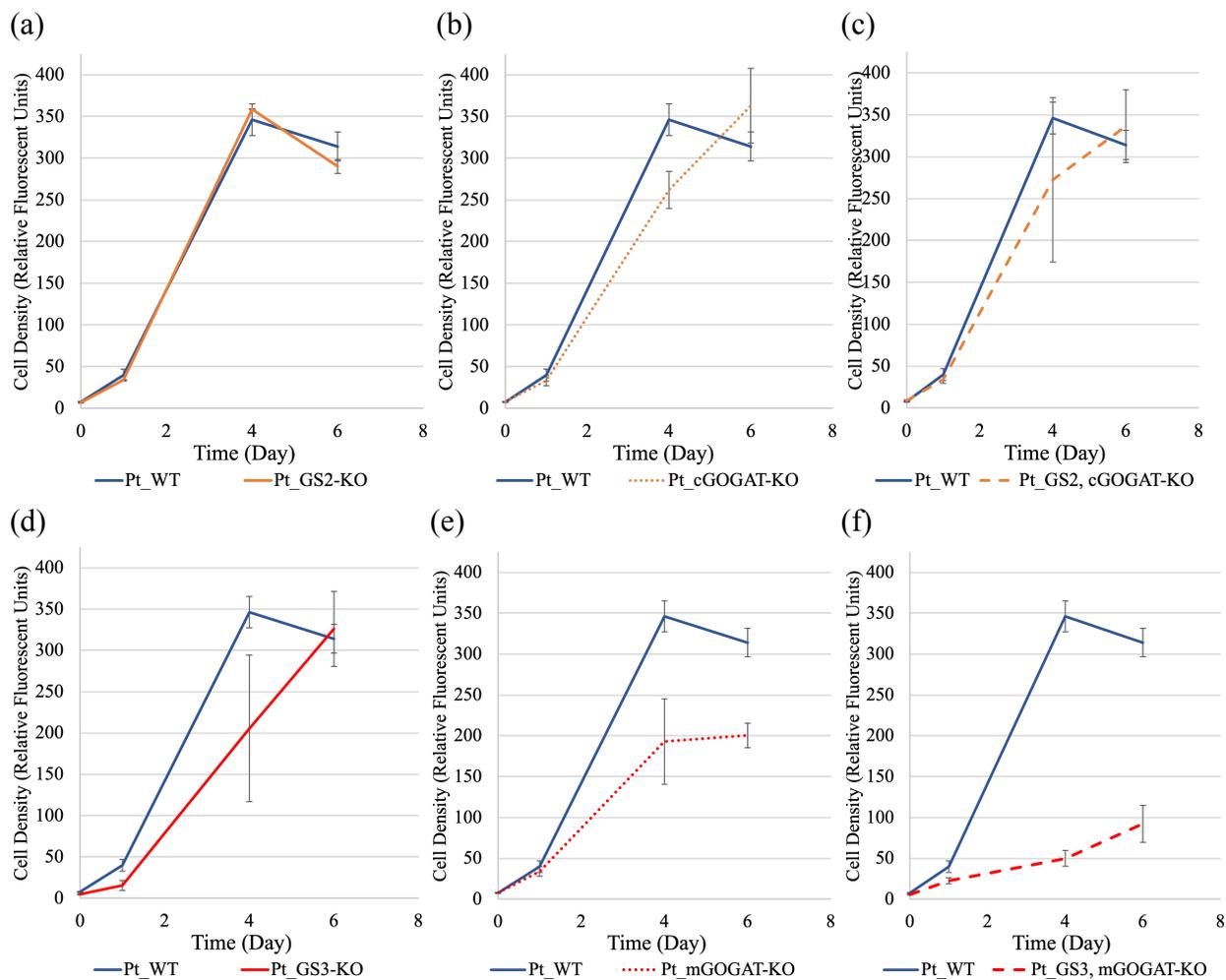
**Figure 2.2** Growth plots for GS-GOGAT single gene knockout cell lines and two-gene knockout cell lines grown on nitrate media. Growth profiles are compared to a wild-type *Phaeodactylum* cell lines (blue line) for **(a)** glutamine synthetase 2 knockout cell lines (Pt_GS2-KO, solid orange line), **(b)** chloroplastic glutamate synthase knockout cell lines (Pt_cGOGAT-KO, dotted orange line), **(c)** two-gene knockout cell lines for GS2 and cGOGAT (Pt_GS2, cGOGAT-KO, dashed orange line), **(d)** glutamine synthetase 3 knockout cell lines (Pt_GS3-KO, solid red line), **(e)** mitochondria glutamate synthase knockout cell lines (Pt_mGOGAT-KO, dotted red line), and **(f)** two-gene knockout cell lines for GS3 and mGOGAT (Pt_GS3, mGOGAT-KO, dashed red line). Errors bars represent the standard deviation between two cell lines and three technical replicates per cell line.

**2.6 Glutamine synthetase 2 knockout in the diatom *Phaeodactylum tricornutum* accumulates lipids early without compromising growth**

### 2.6.1 Abstract

Marine diatoms are known to outcompete other phytoplankton for nitrogen-based metabolites in oceanic environments due to unique metabolic features. Unlike other microalgae, diatoms have a silicified cell wall, a urea cycle, and two GS-GOGAT cycles. In fact, having two GS-GOGAT cycles is a metabolic feature not seen in any other photosynthetic organism. One GS-GOGAT cycle is contained within the diatom chloroplast while the second GS-GOGAT cycle resides in the mitochondria. In ammonium growth conditions, either glutamine synthetase 2 (GS2) in the chloroplast or glutamine synthetase 3 (GS3) in the mitochondria can condense ammonium into amino acids. During nitrate-supplemented growth, the nitrate assimilation pathway sends reduced nitrate exclusively to the chloroplast to condense ammonium into amino acids by GS2. In plants, it has been reported that mutants deficient in GS2 can be transcriptionally compensated by other glutamine synthetase enzymes in the cytosol. Diatoms do not have a cytosolic glutamine synthetase. Here, the function of GS2 was analyzed in the diatom *Phaeodactylum tricornutum* mutants deficient in GS2 (Pt_GS2-KO) to see how the mutant cell lines metabolically compensate. Like in plants, the cell growth of Pt_GS2-KO was not negatively impacted compared to wild-type. Physiological measurements like nitrate uptake and carbon uptake were assessed but a phenotype of increased lipid accumulation was the most apparent in the mutant. Lipid body staining, total carbon analysis, and fatty acid analysis was performed and revealed that Pt_GS2-KO cell lines exhibited a doubling of lipids and triglycerides during nitrate growth conditions compared to wild-type. The combined

characteristics of normal growth and increased lipid content for Pt_GS2-KO cell lines are attractive features for the production of biodiesel from microalgae.

### 2.6.2 Introduction

The production of biodiesel from photosynthetic microorganisms at industrial scales has been a sought-after alternative to petroleum-based fuels for decades (Hu et al. 2008, Scott et al. 2010, Wijffels & Barbosa 2010, Jones & Mayfield 2012, Khan & Fu 2020). To date, vascular plants like palm and jatropha fruit plants have been successfully cultivated for biodiesel at production scale (Mishra & Goswami 2017, Gutierrez et al. 2017). Compared to that of land plants, unicellular algal species are more efficient at partitioning carbon towards lipids per hectare of land use however industrial applications of this have yet to be realized (Hildebrand et al. 2012, Jamwal et al. 2020). Microalgae synthesize lipids in the form of triglycerides (TAG), a precursor molecule to biodiesel. Currently, implementing large-scale cultures for TAG biosynthesis requires biological improvements to cellular metabolism. First, the photosynthetic efficiency of microalgae to convert $CO_2$ to biomass in the form of lipids needs improvement. Instances of photosystem optimization have been attempted such as the manipulation of light harvesting antennae, improvements to energy dissipation, and the engineering of RuBisCO to enhance carbon assimilation (Stephenson et al. 2011, Vecchi et al. 2020). Second, a major hurdle to overcome is that nutrient limitation inhibits the robust growth of algal cells and prevents the continuous production of TAG. Historically, TAG biogenesis has been induced when cells are starved of nutrients or when cells have depleted the available nutrients in the culture medium (Zienkiewicz et al. 2016). Nutrient limitation, specifically nitrogen depletion, and the physiological response of the cells to accumulate lipid is well-known to where nitrogen depletion

is commonly used to "trigger" lipid biosynthesis (Boyle et al. 2012, Ajjawi et al. 2017). For the commercial production of microalgal lipids, enabling lipid production should not come at a cost to cell growth and nitrogen depletion inherently inhibits growth (Berges et al. 1996, Hildebrand et al. 2012).

Marine diatoms, like other eukaryotic photosynthetic microbes, can synthesize commercially valuable bioproducts like lipids (Vinayak et al. 2015). Similar to other algae, they increase lipid production under conditions of nutrient depletion though diatoms have a different metabolism compared to green algae and other phototrophic organisms. For one, they biomineralize a silica(Si)-based outer membrane called a frustule. When Si limited, the cells cannot build frustules and stop dividing which leads to lipid accumulation (Hu et al. 2008, Hildebrand et al. 2012, Shrestha et al. 2012). For another, diatoms, in relation to green algae and plants, have unique nitrogen metabolism features such as an exosymbiont-derived ornithine-urea cycle (Allen et al. 2011) and two fully compartmentalized GS-GOGAT cycles (Smith et al. 2019). These features may contribute to their ability to outcompete other microbes for nitrogenous compounds in oceanic food webs (Dugdale et al. 1998, Falkowski et al. 2004, Falkowski et al. 2007). In relation to nitrogen metabolism, Si metabolism seems to only impact frustule formation and doesn't directly or indirectly impact other cellular processes (Hildebrand et al. 2012). On the other hand,  the complexity of nitrogen metabolism and its impacts on other cellular processes (Levitan et al. 2015b, Smith et al. 2019) may lend opportunities to drive carbon partitioning towards lipid biogenesis indirectly by manipulating nitrogen-related genes.

As previously mentioned, the molecular architecture of GS-GOGAT cycles in diatoms is different from other photosynthetic organisms.  First, unlike green algae and plants, diatoms lack a cytosolic glutamine synthetase (GS). *Chlamydomonas reinhardtii* contains four GS isozymes;

two localized to the chloroplast (GS2 type) and two localized to the cytosol (GS1 type) (Sanz-Luque et al. 2015). Only the two chloroplast localized GS2 enzymes have a corresponding GOGAT enzyme that fully compartmentalizes the GS-GOGAT cycle in the chloroplast of green algae. Second, similar to green algae, diatoms contain a GS2 enzyme and its GOGAT pair (iron-dependent) in the chloroplast. Unlike greens, diatoms lack a cytosolic GS1 but do have a second GS enzyme, GS3, localized to the mitochondria. Thirdly, as in the chloroplast, diatoms have a corresponding GOGAT (NAD(P)H-dependent) in the mitochondria that pairs with GS3. Thus, unlike green algae and land plants, diatoms contain two compartmentalized GS-GOGAT cycles; a chloroplastic GS2-GOGAT(Fd) and a mitochondrial GS3-GOGAT(NAD(P)H).

The transcriptional regulation of GS2 during nitrate replete and deplete conditions indicates that GS2 is an important enzyme for nitrate assimilation for the production of amino acids (Levitan et al. 2015a, Smith et al. 2019). Conversely, GS3 may be important during conditions when nitrogen is limiting (Smith et al. 2019). GS2 and GS3 are transcriptionally coordinated such that GS2 responds to an influx of nitrate and GS3 responds to limiting conditions when diatoms have to scavenge the media for nitrogen after nitrate from the media is depleted (McCarthy et al. 2017). In coastal environments where up-welling of deep waters enriches sun-lit surface waters with nutrients, it can be speculated that GS2 responds to up-welling when nitrate may be replete and GS3 responds after massive cell growth depletes the waters of nitrate. In general, GS2 and GS3 seem to coordinate their gene expression depending on the status of nitrogen availability.

A diatom's alternative nitrogen features, like dual GS-GOGAT cycles, compared to other algal species results in a unique transcriptional regulation during times of nitrogen stress and hence during lipid biogenesis. Efforts to decouple nutrient limitation and lipid induction in

diatoms have been done in two ways. First, researchers have explored the transcriptional dynamics of nitrogen-related genes in response to nitrogen depletion (Smith et al. 2012, Levitan et al. 2015b, Alipanah et al. 2015, Longworth et al. 2016). Doing so has led to the identification of key genes and pathways that are differentially regulated during lipid biogenesis like nitrate assimilation, the pyruvate dehydrogenase complex, glutamate and glutamine metabolism, the TCA cycle, and the urea cycle (Levitan et al. 2015b). Nitrogen limitation has been used as an conditional indicator to facilitate in the finding of key genetic components like a novel RING-domain transcription factors (Matthijs et al. 2016), a master TCA cycle regulator, bZIP14 (Matthijs et al. 2017), and a class of nitrogen starvation inducible promoters (Adler-Agnon et al. 2018). Second, genetic engineering approaches have been pursued to improve lipid biosynthesis beyond wild-type capabilities in the diatoms *Thalassiosira pseudonana* (Trentacoste et al. 2013) and *Phaeodactylum tricornutum* (Daboussi et al. 2014 , Levitan et al. 2015a, McCarthy et al. 2017, Hao et al. 2018).

The identification of nitrogen-related genes that, when made dysfunctional, induce lipids without compromising growth has yet to be accomplished. For example, in a *Phaeodactylum tricornutum* mutant cell line McCarthy et al. (2017) reported an induction of lipids when the nitrate reductase (NR) gene was functionally inactive and cells were cultured on nitrate media. The cells lacked the ability to reduce nitrate and incorporate nitrogen into new amino acids via the GS-GOGAT cycle. Therefore, despite there being an increase in lipids in a nitrogen gene knockout cell line, growth was severely diminished regardless of nitrogen availability (McCarthy et al. 2017). NR is a single-copy gene in *Phaeodactylum* and the sole enzyme responsible for the beginning of nitrate assimilation. Nonetheless, cell death on nitrate media was expected.

The production of lipids in a diatom should not come at the cost of a decrease in cell growth if to be applied in industrial settings. The innate redundancy of dual GS-GOGAT cycles in diatoms may be an opening to potentially perturb nitrate assimilation to trigger lipid biogenesis. The decoupling of nitrogen depletion and lipid biogenesis by genetic engineering approaches would contribute to the growing knowledge of diatom biofuel production. In this study, a Cas9-mediated GS2 knockout cell line in the model pennate diatom *Phaeodactylum tricornutum* was produced in order to elucidate the physiological consequences of the mutant when grown on nitrate. Without a functional GS2 enzyme, the mutant cell line exhibited no growth defects on any nitrogen source. Interestingly, growth on nitrate was not inhibited but led to the biosynthesis of more TAG in the mutant than wild-type compared to growth on other nitrogen supplements. Here, it was demonstrated that the episomal delivery of CRISPR-Cas9 components to *Phaeodactylum* can induce lipid biosynthesis by functionally inactivating the nitrate assimilation-adjacent GS2 gene without compromising growth.

### 2.6.3 Results

#### 2.6.3.1 GS2 gene expression and localization in wild-type *Phaeodactylum*

In *Phaeodactylum*, nitrate assimilation metabolites are exclusively shunted towards the chloroplast and utilized by the chloroplast GS-GOGAT cycle (Figure 2.3a). Once nitrate is transported into the cell, nitrate reductase (NR) reduces nitrate to nitrite. Nitrite is further reduced by nitrite reductase (NiR), of which there are two versions that are both localized to the chloroplast. Due to the predicted localization of NiR, ammonium derived from nitrate is localized to the chloroplast where GS2 is localized.

**Figure 2.3.** Metabolism, localization, and regulation information for glutamine synthetase 2 in *Phaeodactylum*. **(a)** Schematic of nitrate assimilation and GS-GOGAT cycles in *Phaeodactylum*. Metabolites are represented by grew boxes ($NO_3^-$=nitrate, $NO_2^-$=nitrite, $NH_4^+$=ammonium, GLN=glutamine, GLU=glutamate, 2-OG=2-oxoglutarate). Chloroplast (green), cytosol (blue), and mitochondria (orange) localized protein are represented by ovals (NR=nitrate reductase (Phatr3_J54983), NiR=nitrite reductase (Phatr3_J12902, Phatr3_J13154, Phatr3_J9538), GS2=glutamine synthetase 2 (Phatr3_J51092), GS3=glutamine synthetase 3 (Phatr3_J22357), cGOGAT(Fd)=glutamate synthase (Phatr3_J24739), mGOGAT(NADPH)=glutamate synthase (Phatr3_J51214), GDH=glutamate dehydrogenase (Phatr3_J45239, Phatr3_J13951, Phatr3_J50971), unCPS=carbamoyl phosphatase (Phatr3_J24195)). **(b)** Subcellular localization of Venus (YFP) fused to chloroplastic glutamine synthetase 2 in *Phaeodactylum*. DIC bright field image (top); chloroplast autofluorescence (top middle); YFP-GS2 fluorescence (bottom middle); merge (bottom). Scale bar is 5um. **(c)** Transcript expression and protein levels of Pt_GS2 and Pt_GS3 in reaction to a "spike" of nitrate to the media (black), during N-free conditions (grey), during NH4 growth (orange), during nitrite growth (green), and during nitrate growth (blue). Solid lines are transcription data and dotted lines are protein data. Data for this was mined from the N-short experiment conducted in Smith et al. (2019).
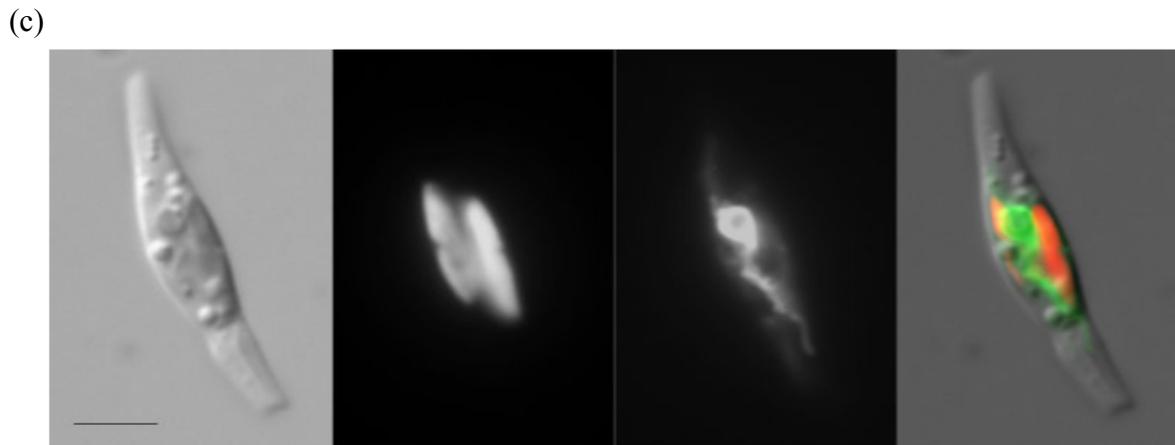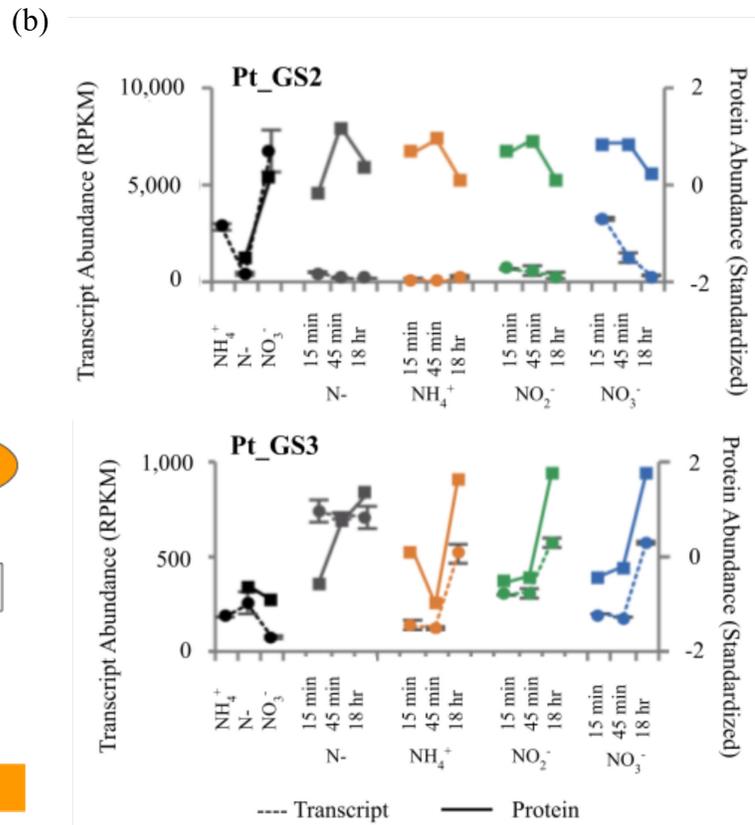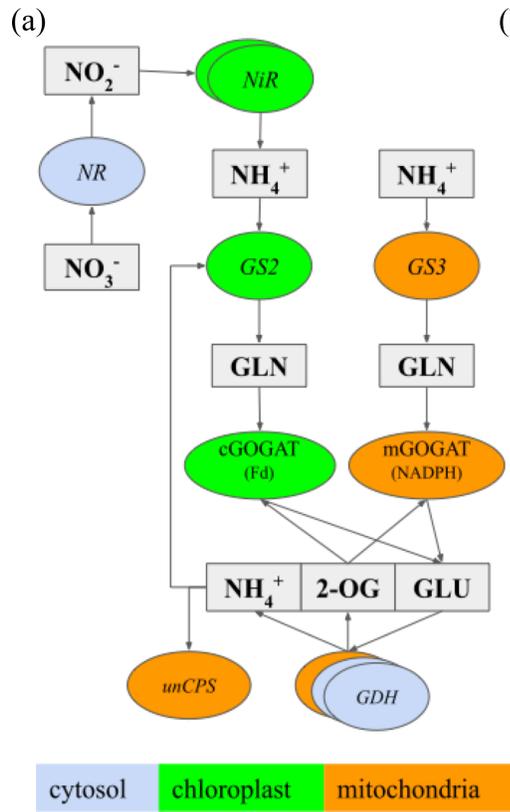
GS2 was further validated to be localized to the chloroplast by subcellular co-localization of GS2 and a yellow fluorescent protein (YFP) (Figure 2.3b). First, GS2 contains a putative chloroplast signal peptide and diatom predicted ASA-F peptide cleavage site at the 17-AA position (SignalP 4.1, www.cbs.dtu.dk). Second, an episomal expression vector was constructed to harbor a YFP-GS2 expression cassette using the *Phaeodactylum* NR promoter (pNR) and terminator sequence for transcriptional regulation (Supplemental Figure 2.1). pNR only transcribes its gene product in the presence of nitrate and suppressed expression in the absence of nitrate (Chu et al. 2018). Transgenic *Phaeodactylum* cell lines with the YFP-GS2 episome were produced and first grown on ammonium media then switched to growth on nitrate media and imaged for YFP fluorescence. YFP-GS2 was not localized to the entire chloroplast but rather was localized to the chloroplast endoplasmic reticulum. This localization was in line with the previously reported localization of GS2 during constitutive gene expression (Siaut et al. 2007) and also for the localization of the carbonic anhydrase 3 enzyme in *Phaeodactylum* (Tachibana et al. 2011).

The coordination of amino acid production by either GS2 or GS3 in diatom is strongly correlated with the availability of nitrate in the media. Since nitrate assimilation drives the localization of reduced nitrogen in the form of ammonium to the chloroplast, GS2 is thought to be the primary GS enzyme responsible for the condensation of glutamate and ammonium to produce glutamine. In nitrate media, it was reported in the diatom *Skeletomena costatum* that GS2 is transcriptionally upregulated compared to growth on ammonium media where GS3 transcription remains constant (Takabayashi et al. 2005). A transcript-level analysis of the nitrate response in *Phaeodactylum* was conducted in Smith et al. 2019 by deploying a nitrate "spike-in" experiment where nitrate was added to cultures in nitrogen-free media and transcriptional

regulation was observed for 15 min, 45 min, and 18 hours after the nitrate addition. Among the most highly upregulated genes in response to nitrate included NR, a nitrite reductase (NiR), and GS2 (Smith et al. 2019). The transcriptional response of GS2 was fast where high levels of transcription occurred 15 and 45 minutes after nitrate addition and was reduced after 18 hours (Figure 2.3c). GS3, in contrast, exhibited an opposite response where transcription was downregulated in response to an addition of nitrate and was increased after 18 hours (Figure 2.3c). Interestingly, the transcriptional regulation for GS2 was matched in protein abundance only during nitrate spike conditions (black lines, Figure 2.3c) and during nitrate growth (blue lines, Figure 2.3c). During growth on nitrogen-free, ammonium, and nitrite, GS2 protein abundance was diminished in comparison to nitrate conditions regardless of the transcriptional responses being relatively equivalent. The coordination of GS2 and GS3 gene expression was also observed during extended growth time on nitrate media where GS2 was highly upregulated in reaction to an addition of nitrate and GS3 transcription was increased incrementally throughout growth (Supplemental Figure 2.2). The two GOGAT genes have almost identical transcriptional profiles but differ in magnitude (Supplemental Figure 2.2). In wild-type *Phaeodactylum,* it was reported that GS2 transcription increased in a NR knockout cell line when grown on nitrate media while GS3 transcription was unchanged compared to wild-type (McCarthy et al. 2017). Interestingly, in *Thalassiosira pseudonana* GS3 was shown to be transcriptionally upregulated when nitrate levels were depleted in the growth media while GS2 remained constant (Hockin et al. 2012).

### 2.6.3.2 Generation of GS2 knockout cell lines

The three CRISPR-Cas9 targets were designed to mutate the second exon of the GS2 (Phatr3_J51092) genomic gene sequence (chr.20:363,436-366,254(+)) (Figure 2.4a). The location of the target loci were designed to mutate the GS2 gene sequence upstream (gGS2-1) and near the 5' end (gGS2-2 and gGS2-3) of a glutamine synthetase catalytic domain (InterPro Accession Number: IPR008146). The predicted Cas9 cut sites for each sgRNA are 27-nt, 211-nt, and 283-nt from the 5' end of the second exon of Pt_GS2 for gGS2-1, gGS2-2, and gGS2-3, respectively.
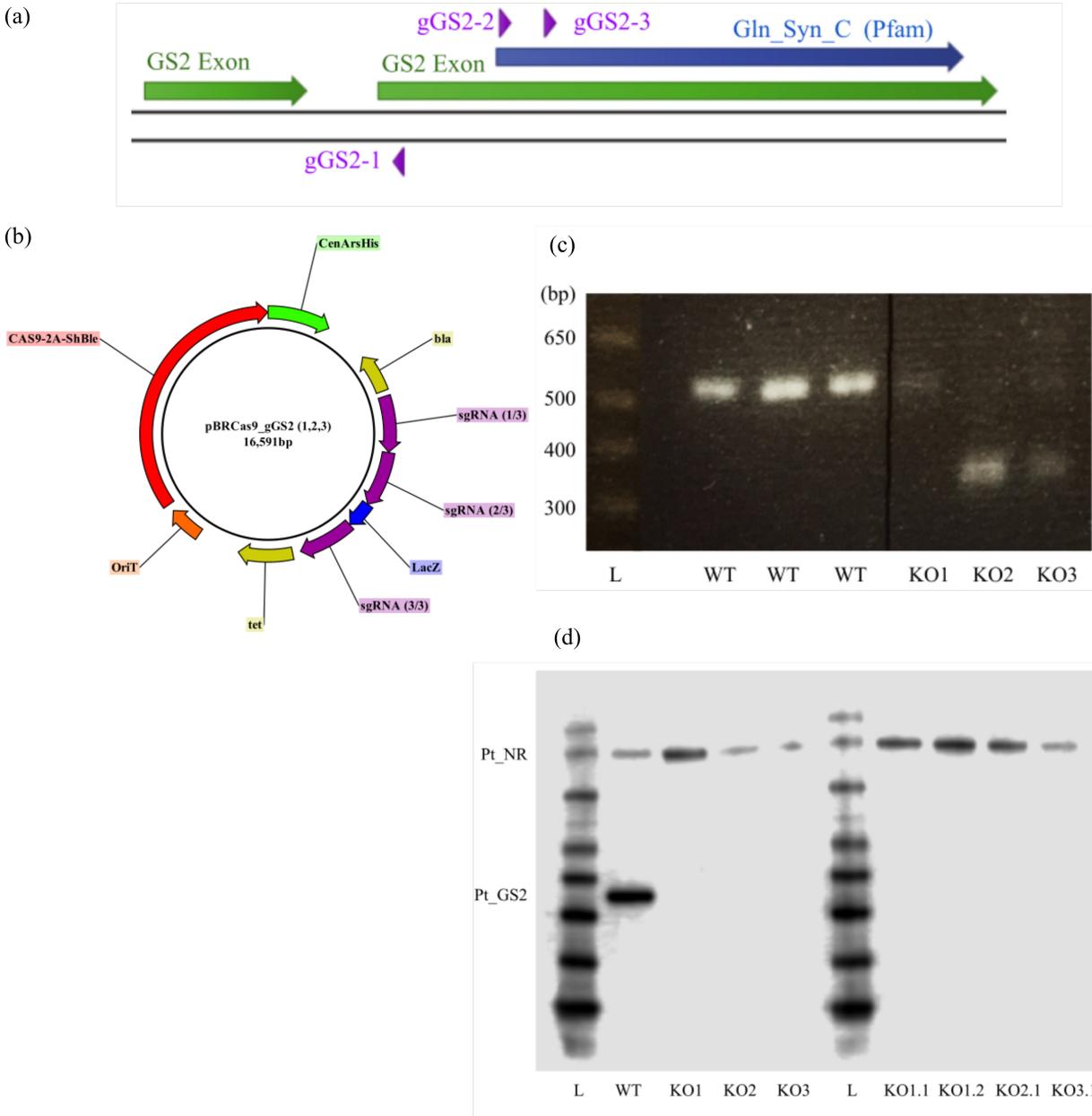
**Figure 2.4** Generation of mutant *Phaeodactylum* cell lines of glutamine synthetase 2. **(a)** Diagram of the genomic DNA locus of GS2 (Gene ID: Phatr3_J51092) with its two exons (green) and glutamine synthase catalytic domain (blue) shown. Three Cas9-sgRNA target loci (gGS2-1, gGS2-2, gGS2-3, purple) were designed to mutate GS2. **(b)** Plasmid map of an episomal vector harboring the three sgRNA expression constructs for gGS2-1, gGS2-2, and gGS2-3. **(c)** Gel picture comparing the amplicon of three wild-type GS2 genes (519-bp) and the amplicon of three Cas9-mediated mutant GS2 genes. **(d)** Protein gel of wild-type and Pt_GS2-KO cell lines before (left) and after sub-cloning (right). Antibodies to detect the presence of the nitrate reductase protein (NR, 901-AA, 108.1 kDa) and the GS2 protein (GS2, 416-AA, 49.9 kDa) were both used.

Pt_GS2-KO cell lines were produced using an episomal CRISPR-Cas9 system to simultaneously deliver the three sgRNA expression cassettes and a selectable-Cas9 to *Phaeodactylum* (Figure 2.4b). Conjugation transformation of the pBRCas9_gGS2(1,2,3) resulted in 200-300 *Phaeodactylum* colonies resistant to the antibiotic phleomycin. 36 colonies were picked and screened via PCR with a wild-type control for mutation in the Pt_GS2 genomic locus. 8 colonies exhibited a Pt_GS2 genotype different from wild-type and three were chosen for further investigation (Figure 2.4c). To examine whether the three candidate cell lines contained functional knockouts of Pt_GS2, a protein gel was run (Figure 2.4d). The three candidates were compared to a wild-type control for the presence of the Pt_GS2 and Pt_NR (nitrate reductase) enzymes. The protein gel shows that all three candidate cell lines do not produce Pt_GS2 enzymes and do produce Pt_NR. The three Pt_GS2-KO cell lines were then subcloned to ensure clonality of the Pt_GS2-KO cell lines. Two subclones of Pt_GS2-KO 1 and one subclone of Pt_GS2-KO2 and Pt_GS2-KO3 were examined for the presence of the Pt_GS2 and Pt_NR enzymes by a protein gel (Figure 2.4d). All subclones did not contain Pt_GS2 and did contain Pt_NR in their protein fractions.

### 2.6.3.3 Pt_GS2-KO nitrate media drawdown

Pt_GS2 is predicted to be the primary GS enzyme in *Phaeodactylum* to assimilate nitrogen into glutamine when grown on nitrate media. Nitrate drawdown from the growth medium was then analyzed for Pt_WT and Pt_GS2-KO cell lines. Two growth conditions were examined; when cells were pre-conditioned on nitrate media and remained on nitrate media during drawdown examination ($NO_3^-$ to $NO_3^-$) and when cells were pre-conditioned on ammonium media and switched to nitrate media during drawdown examination ($NH_4^+$ to $NO_3^-$).

The nitrogen switch was performed due to the transcriptional up-regulation observed for GS2 when wild-type cells were pre-conditioned on ammonium and spiked for nitrate (Figure 2.5b, Supplemental Figure 2.3). During the nitrate drawdown experiments on both conditions, the growth profiles for Pt_WT and Pt_GS2-KO cell lines were comparable and showed no significant differences (Supplemental Figure 2.3). Upon initial observations, the nitrate drawdown for both conditions was also comparable for both conditions throughout growth (Figure 2.5a,b). After further examination, it was found that the nitrate uptake rate per cell in the Pt_GS2KO cell line was significantly elevated between hours 24 and 48 for $NO_3^-$ to $NO_3^-$ by about 4.7 fmol/cell (Figure 2.5c) and for the $NH_4^+$ to $NO_3^-$ condition by 2.5 fmol/cell (Figure 2.5d). For the remaining time periods (48 - 72hrs and 72 - 96hrs) the nitrate uptake rate per cell was even between Pt_WT and Pt_GS2-KO cell lines for both conditions.
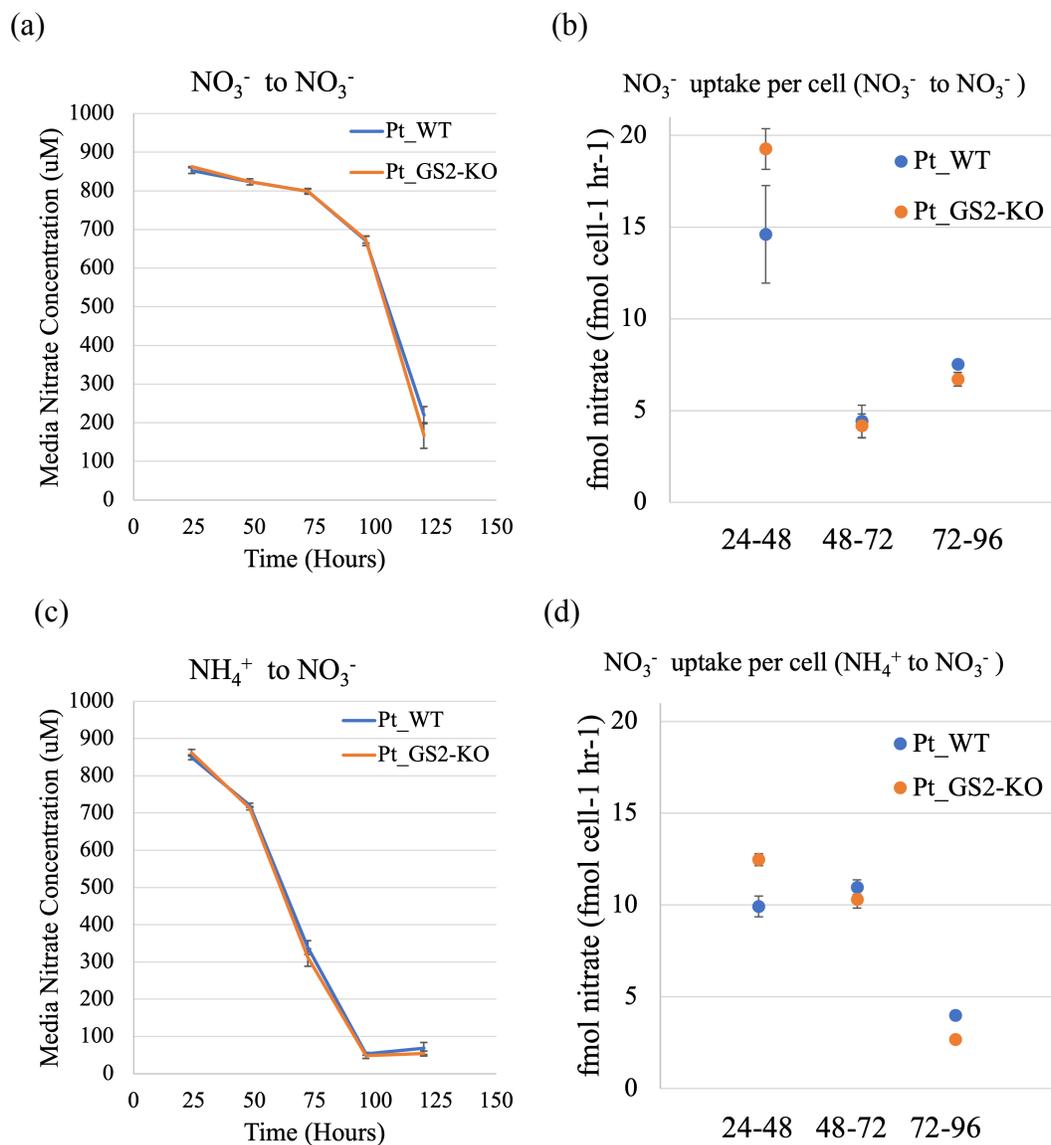
**Figure 2.5** Nitrate drawdown from the media for Pt_WT (blue) and Pt_GS2-KO (orange) cell lines. **(a)** Nitrate drawdown plot for when cells were pre-conditioned on nitrate media. **(b)** Nitrate drawdown plot for when cells were pre-conditioned on ammonium media and switched to nitrate media. **(c)** Nitrate uptake per cell for cells pre-conditioned on nitrate media for three 24 hour growth periods (24-48 hours, 48-72 hours, 72-96 hours). **(d)** Nitrate uptake per cell for cells pre-conditioned on ammonium media and switched to nitrate media for three 24 hour growth periods (24-48 hours, 48-72 hours, 72-96 hours).

### 2.6.3.4 Carbon uptake, $O_2$ evolution, and non-photochemical quenching

Growth impairment was not apparent for Pt_GS2-KO cell lines, however examination of the carbon dioxide uptake and oxygen evolution in Pt_GS2-KO cell lines indicates that the cellular efficiency to assimilate carbon via photosynthesis may be impair on short time scales. During a 30 minute experiment, Pt_WT cell lines were able to assimilate 1.6X $CO_2$ compared to Pt_GS2-KO (Figure 2.6a). Likewise, oxygen evolution was also elevated in Pt_WT cell lines by 1.7X (Figure 2.6b). Conversely, differences in non-photochemical quenching (NPQ) for Pt_WT and Pt_GS2-KO cell lines were negligible (Figure 2.6c). This indicates that, regardless of diminished carbon uptake and oxygen evolution in Pt_GS2-KO cell lines, the photosystems' ability to quench excess heat when exposed to supersaturated light is not diminished and it is reasonable to conclude that the photosystems are not necessarily damaged. Regardless of the observed decrease in carbon uptake and oxygen evolution in Pt_GS2-KO cell lines, the comparable growth profile over longer time-scales (0 - 120 hours, Supplemental Figure 2.3) between Pt_WT and Pt_GS2-KO cell lines indicates that carbon uptakes rates are not diminished for times greater than observed here.

(a)

**CO$_2$ Instantaneous Uptake**

(b)

**Oxygen Evolution**

(c)

**Non-photochemical Quenching (NPQ)**
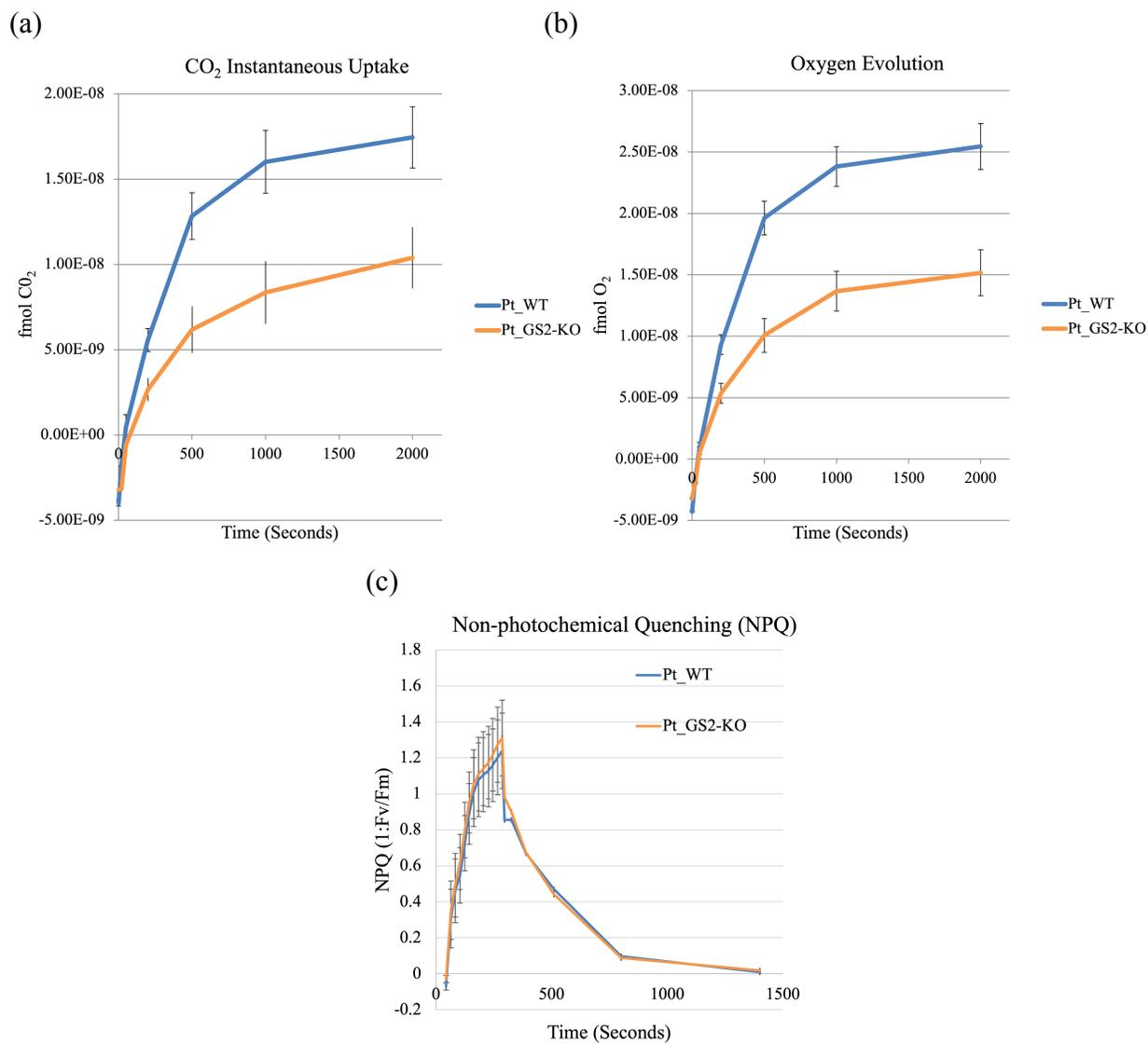
**Figure 2.6** Short-term measurement profiles for **(a)** carbon dioxide uptake rate, **(b)** oxygen evolution, and **(c)** non-photochemical quenching (NPQ) for cell lines Pt_WT (blue) and Pt_GS2-KO (orange).

### 2.6.3.5 Lipid analysis by lipid body staining

The slight physiological changes in short-term carbon uptake and oxygen evolution and the observed increase in nitrate uptake during lag and early exponential growth phases indicated that cellular homeostasis in Pt_GS2-KO may be altered. One common phenomenon observed in *Phaeodactylum* and other microalgae is the accumulation of lipids during conditions of stress. To test whether Pt_GS2-KO cell lines were exhibiting intercellular stress and accumulating lipids, Pt_WT and Pt_GS2-KO cells were stained with the lipid body stain BODIPY 493/503 and imaged by confocal microscopy (Figure 2.7). It was observed that less lipids accumulated in Pt_WT cell lines (Figure 2.7a,b) than what was observed in Pt_GS2-KO cell lines (Figure 2.7c,d) in both ammonium and nitrate growth conditions. It was also observed that lipids accumulated as soon as 48 hours and through 96 hours in Pt_GS2-KO cell lines. Lipids were not detectable in Pt_WT cell lines until 72 hours and 96 hours yet still at a much lower qualitative level.

**Figure 2.7** Lipid body staining of Pt_WT and Pt_GS2-KO cell lines after 48, 72, and 96 hours of growth. Microscope images for all time points are merged with chlorophyll autofluorescence (red), BODIPY 493/503 stain (green), and the bright-field image. The last column merges only the chlorophyll autofluorescence and BODIPY staining images. Lipid body staining and imaging was done for **(a)** Pt_WT on ammonium media, **(b)** Pt_WT on nitrate media, **(c)** Pt_GS2-KO on ammonium media, and **(d)** Pt_GS2-KO on nitrate media. **(e)** Lipid phenotype induction rescue with glutamine supplementation. Pt_GS2-KO cell lines were grown on nitrate media for 72 hours and BODIPY stained (left). The cells were washed with N-free media and grown on nitrate and glutamine media for 72 hours and BODIPY stained (middle). The cells were then washed with N-free media and grown on nitrate media for 72 hours and BODIPY stained. Chlorophyll autofluorescence (red); BODIPY 493/503 stain (green).
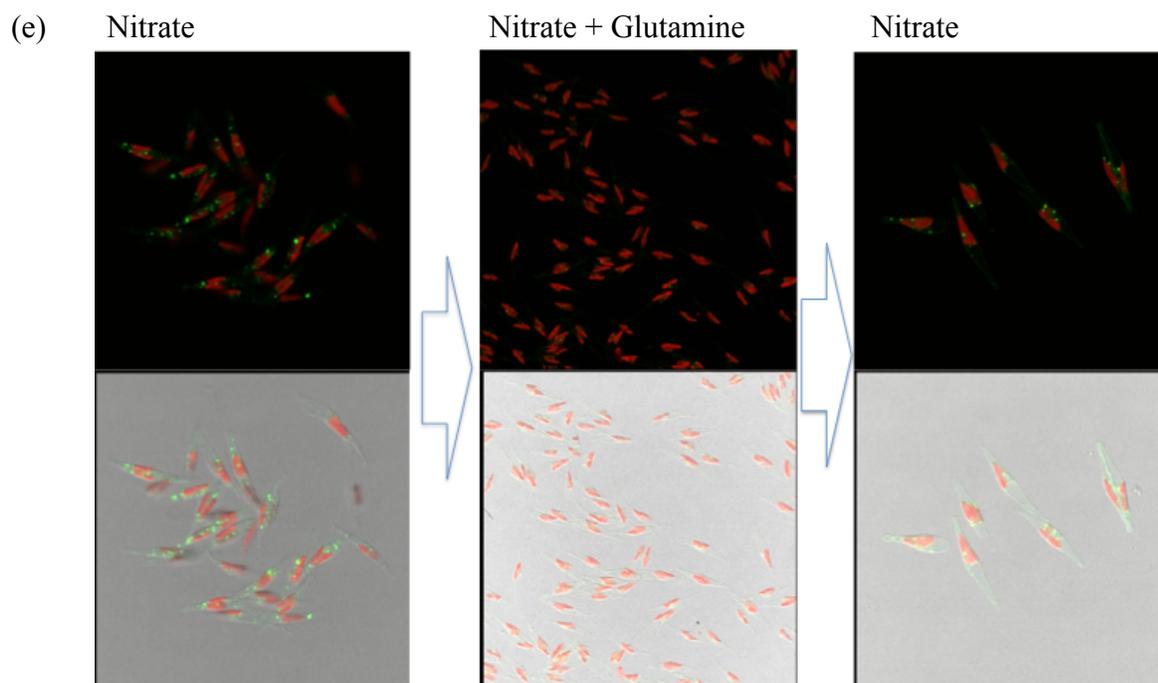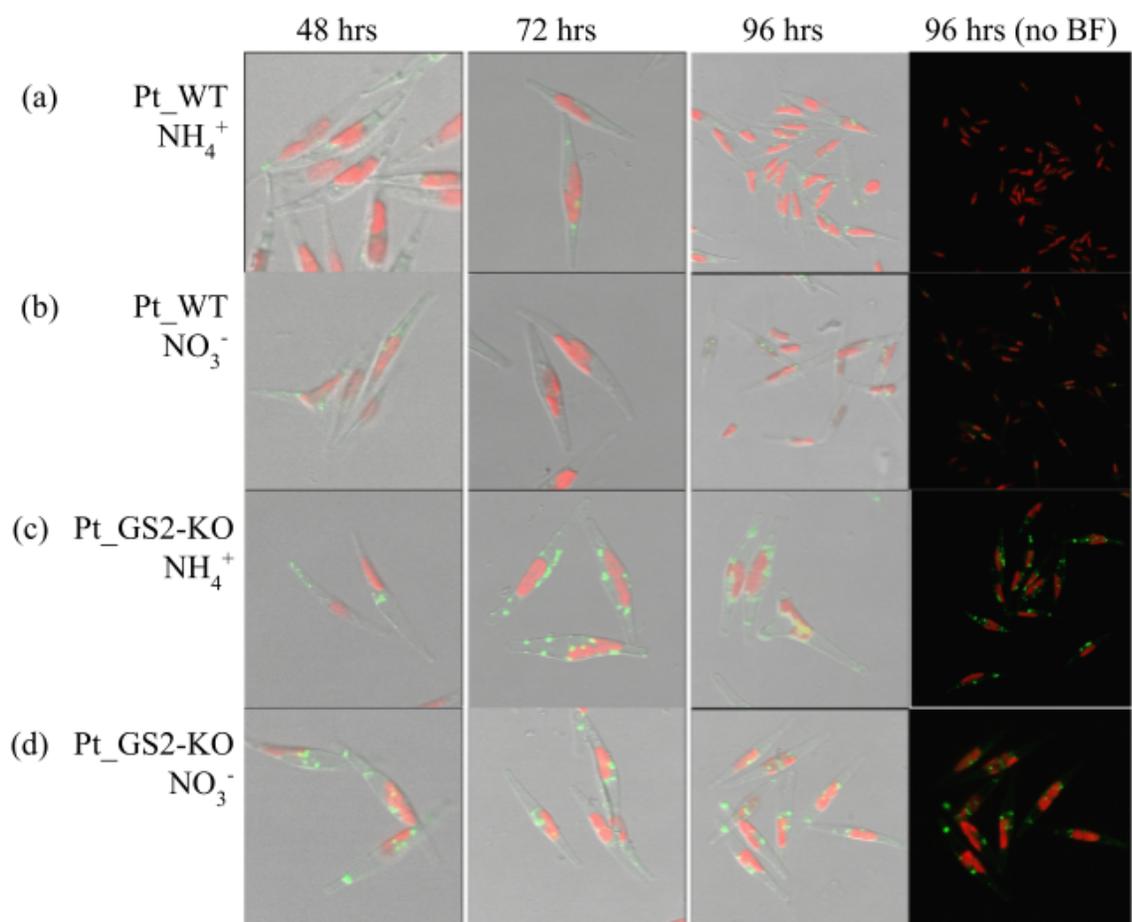
The observed lipid accumulation phenotype for Pt_GS2-KO cell lines was attempted to be "rescued" or abolished by supplementing nitrate grown cells with glutamine (Figure 2.7e). Since glutamine biosynthesis may be impaired, the direct addition of glutamine to the growth medium may skip glutamine synthetase in the GS-GOGAT cycle and give glutamate synthase (GOGAT) the ability to condense glutamine and 2-oxoglutarate to produce glutamate. After 72 hours, lipid accumulation was observed by BODIPY staining in Pt_GS2-KO cell lines. After addition of nitrate and glutamine and another 72 hours, lipids were not observed. After the cells were switched back to nitrate-only media and grown for 72 hours, lipids were observed again.

### 2.6.3.6 Total carbon and fatty acid sampling

### 2.6.3.6.1 TOC/TON analysis

Total organic carbon (TOC) and total organic nitrogen (TON) analysis was performed for each condition and time-point in the sampling experiment (Supplemental Figure 2.4 and Supplemental Figure 2.5, respectively). In general, both cell lines contained equivalent amounts of TON for all time-points and between each culture condition (Supplemental Figure 2.5). TOC, though, varied between the cell lines for growth conditions on ammonium, nitrate, and glutamine at the 108 hour time-point (Figure 2.8). At 108 hours, Pt_WT accumulated more TOC per volume than Pt_GS2-KO cell lines cultured on ammonium media (Figure 2.8a) and on glutamine media (Figure 2.8c). The opposite effect persisted when cultured on nitrate media where Pt_GS2-KO cell lines exhibited more TOC per volume than Pt_WT cell lines at 108 hours (Figure 2.8b). The TOC:TON ratio at the 108 hour time-point was also calculated (Figure 2.8d). During nitrate growth, TOC:TON was elevated for Pt_GS2-KO compared to Pt_WT but not

significantly so. During glutamine growth, TOC:TN was significantly elevated for Pt_WT compared to Pt_GS2-KO.
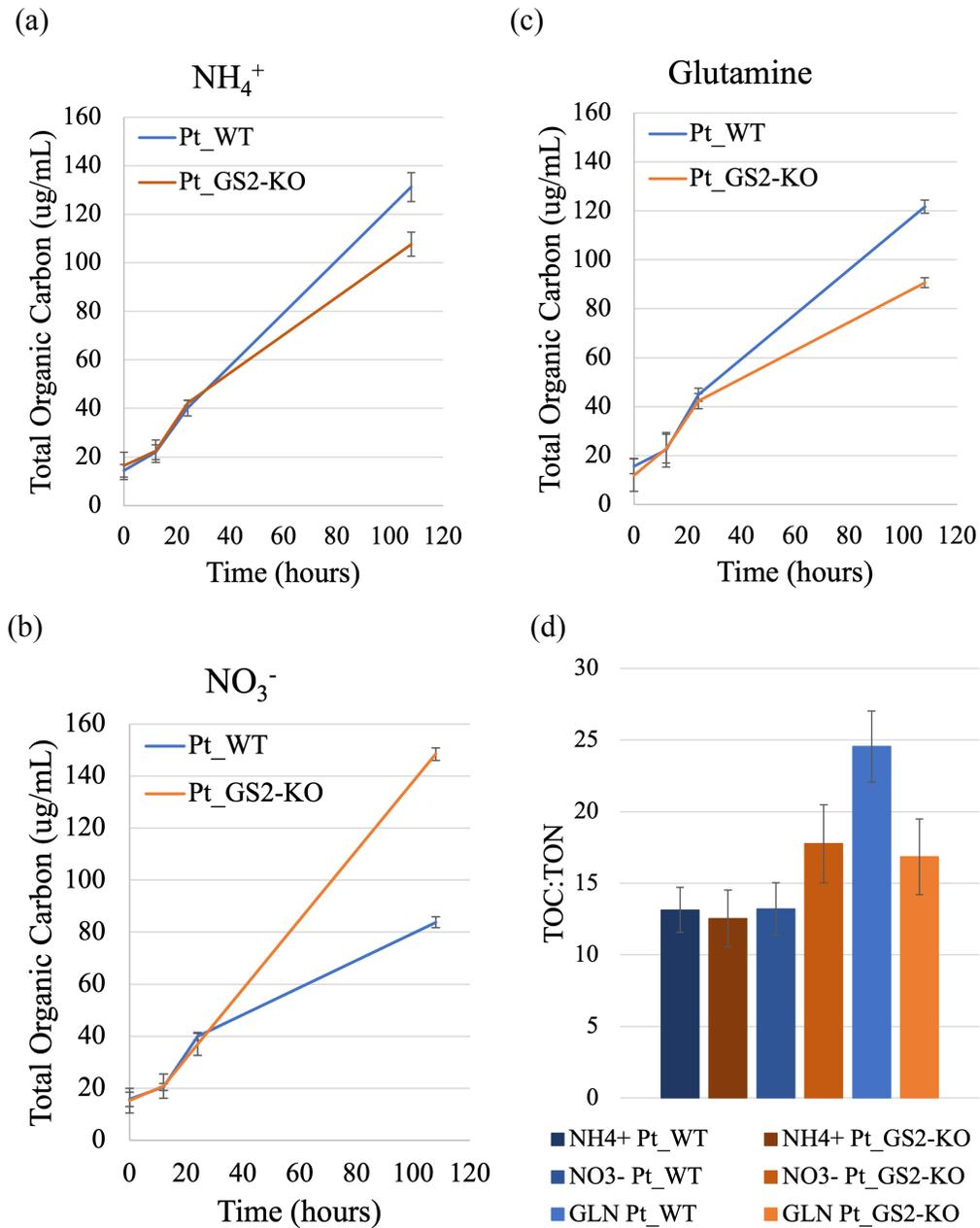


**Figure 2.8** Total organic carbon for Pt_WT (blue) and Pt_GS2-KO (orange) cell lines for each sampling point (0hrs, 12hrs, 24hrs, and 108hrs) during conditions of growth on **(a)** ammonium media, **(b)** nitrate media, and **(c)** glutamine media. **(d)** A histogram of the TOC:TON ratio at the 108hr sampling point. All Pt_WT conditions are shown in shades of blue and all Pt_GS2-KO conditions are shown in shades of orange.

### 2.6.3.6.2 Fatty acid analysis

Fatty acid methyl ester (FAME) analysis was performed to calculate the total fatty acids in the form of triglycerides (TAG) and the fatty acid species compositional profile. Compared to all growth conditions, there was a significant difference in TAG per TOC and TAG per cell only during nitrate growth conditions (Figure 2.9, Supplemental Figure 2.6). Specifically for nitrate conditions, there was significantly more TAG for Pt_GS2-KO at 12 hours (1.4X more) and 108 hours (2X more) (Figure 2.9a). The most abundant fatty acid species (14:0, 16:0, 16:1, 16:2,12, 16:3, and 20:5) were compared between Pt_WT and Pt_GS2-KO at 12, 24, and 108 hours (Figure 2.9b). At 12 hours, Pt_GS2-KO contained significantly more TAG and specifically within the fatty acid species with molecular structures of 16:0 and 16:1. At 24 hours, though seemingly elevated, there were not significantly elevated fatty acid species between the two cell lines. At 108 hours, Pt_GS2-KO contained significantly more TAG within the fatty acid species with molecular structures of 14:0, 16:0, 16:1, and 20:5.
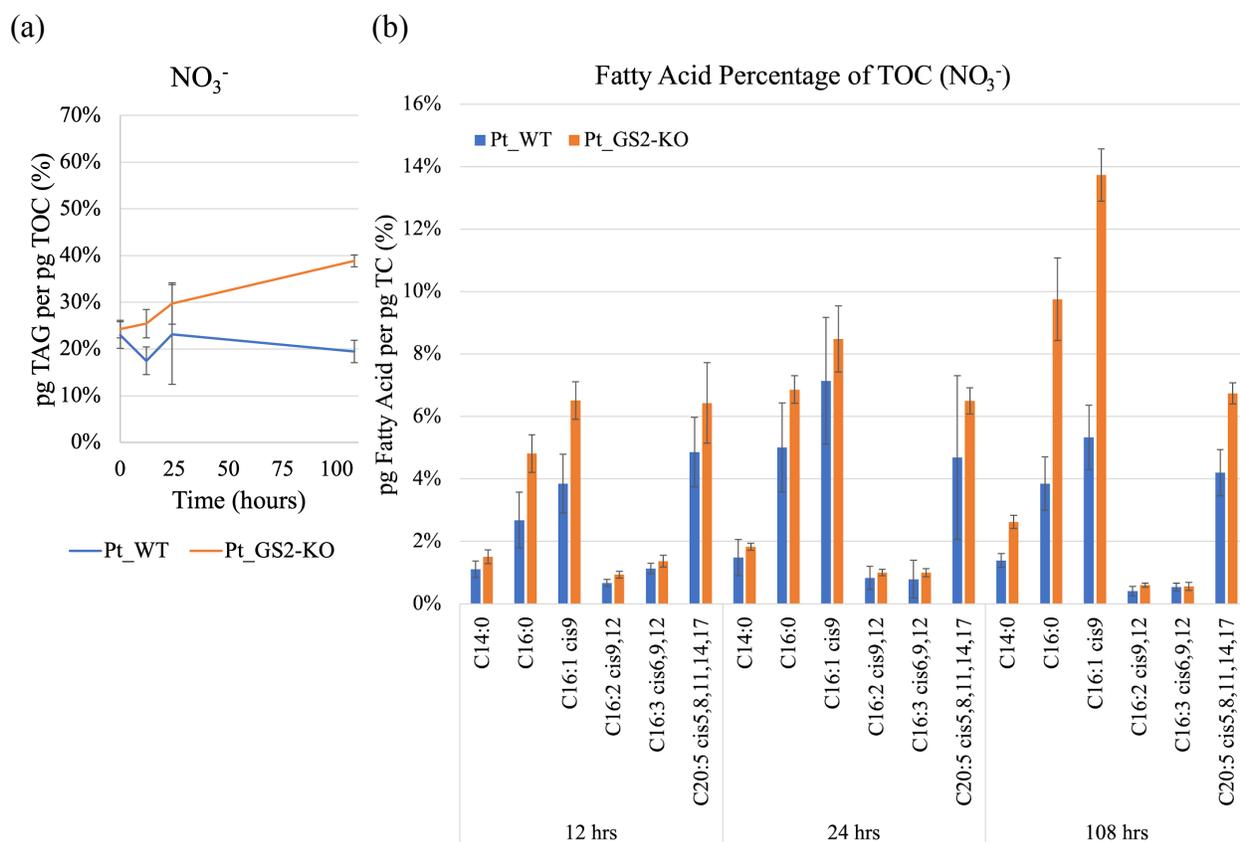
**Figure 2.9** Fatty acid profile of Pt_WT (blue) and Pt_GS2-KO (orange) during nitrate growth. **(a)** Percentage TAG of TOC for all time-points during nitrate growth. **(b)** Quantitative analysis of fatty acid species in cell lines Pt_WT and Pt_GS2-KO during nitrate growth. Percentages of fatty acid species of total organic carbon for time points 12hrs, 24hrs, and 108hrs are calculated using three biological replicates with error bars by standard deviation.

The accumulation of TAG was also analyzed over continuous growth conditions for Pt_WT, Pt_GS2-KO, and Pt_GS3-KO, a knockout cell line for the mitochondrial glutamine synthetase GS3. The cells were cultured so nitrogen limitation never occurred and so that the cells never entered the stationary phase of growth. By maintaining these conditions, a lipid-stress response was avoided. The cells were inoculated into nitrate media at 1e6 cells/mL and cultured for 3 days. Then, the cultures were diluted back to 1e6 cells/mL. Everyday, FAME analysis was performed and revealed that Pt_GS2-KO accumulates more TAG per cell at 24 hours and 72 hours compared to Pt_WT and Pt_GS3-KO cell lines (Supplemental Figure 2.7).

The composition of fatty acids (molecular percentage of fatty acid species per total TAG) was compared between Pt_WT and Pt_GS2-KO cell lines during nitrate growth for time-points 12, 24, and 108 hours (Supplemental Figure 2.8). At 12 hours, the only fatty acid that was elevated for Pt_GS2-KO cell lines was 16:1 (Supplemental Figure 2.8a). For Pt_WT at 12 hours, the composition of fatty acids 16:3, 20:5, and 24:0 was greater than Pt_GS2-KO. At 24 hours, 16:1 was greater for Pt_WT and 20:5 was greater for Pt_gGS2-KO (Supplemental Figure 2.8b). Finally,  at 108 hours, 16:0 and 16:1 were significantly greater in Pt_GS2-KO (Supplemental Figure 2.8c).

Similar to what was observed for the glutamine addition and lipid body staining experiment (Figure 2.7), glutamine supplementation reversed the increase of fatty acid accumulation as observed for ammonium and nitrate conditions at 108 hours (Figure 2.10). On ammonium, Pt_GS2-KO accumulated 1.75X more 16:0 and 1.65X more 16:1. On nitrate, the ratios increase further by 2.5X and 2.6X, respectively. On glutamine, the ratios increase for Pt_WT by 1.4X and 1.5X, respectively. Therefore, glutamine supplementation seems to suppress TAG biogenesis for Pt_GS2-KO compared to ammonium and nitrate supplementation.
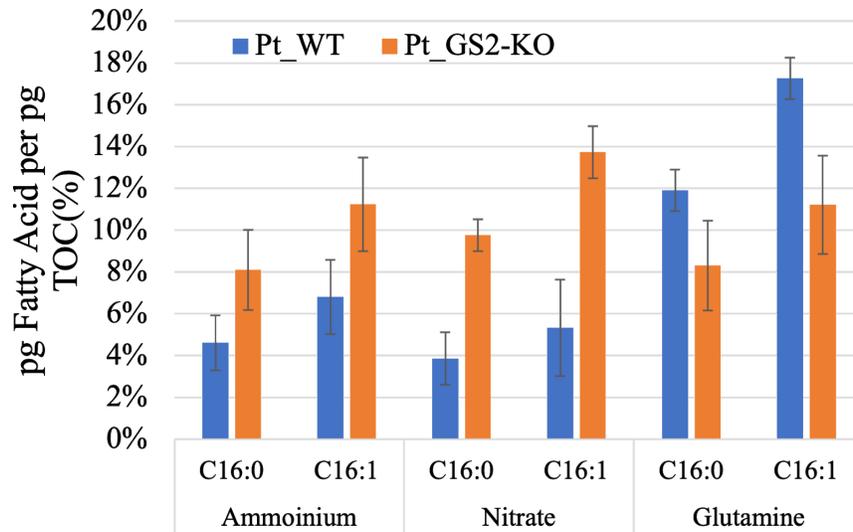
**Figure 2.10** Percentage of fatty acids species 16:0 and 16:1 of total carbon for cell lines Pt_WT (blue) and Pt_GS2-KO (orange) at 108 hours growth. Histogram displays the fatty acids per TOC for growth conditions on ammonium, nitrate, and glutamine supplemented media.

### 2.6.4 Discussion

Here, the CRISPR-Cas9 episomal delivery system developed by Moosburner et al. (2020) was used to produce GS2 knockout cell lines in *Phaeodactylum tricornutum* by saturation mutagenesis. Growth of the Pt_GS2-KO cell lines did not diminish compared to wild-type for any culture conditions and nitrogen sources attempted. Before Pt_GS2-KO was assessed for growth, it was hypothesized that cell growth would be inhibited when grown on nitrate media. The molecular architecture of nitrate assimilation in *Phaeodactylum* indicates that nitrite, the product of nitrate reduction, is reduced in the chloroplast since two NiR isoenzymes are chloroplastic localized. The reduced product of nitrite is ammonium which, in wild-type cells, is incorporated into amino acids by GS2 and the chloroplastic GOGAT. Without a functional GS2, it was thought that ammonium would need to be re-localized to alternative nitrogen sinks in order to retain normal growth. Ammonium could be re-localized to the mitochondria for either

condensation to glutamine by GS3 of for condensation to carbamoyl phosphate by carbamoyl phosphate synthase, or to the cytosol for condensation of 2-oxoglutarate and ammonium to glutamate by glutamate dehydrogenase. It was thought that the re-allocation of ammonium would be energetically costly to a GS2 mutant and growth would be inhibited. Nevertheless, growth was not inhibited for Pt_GS2-KO compared to Pt_WT on nitrate media, ammonium media, and glutamine media.

A retention of normal growth may have been expected for a GS2 mutant. In plants GS2 mutants, normal growth rates occurred but with physiological consequences (Wallsgrove et al. 1987, Betti et al. 2006, Ferreira et al. 2019). In barley, a mutated GS2 isoenzyme resulted in increased ammonia in the leaves. Plus these mutants could not grow on photorespiratory growth conditions (Wallsgrove et al. 1987). Similarly, GS2 mutants in *Lotus japonicus* were identified by the inability to grow during photorespiration (Betti et al. 2006). Lastly, GS2 mutants of *Arabidopsis thaliana* were able to grow during photorespiration but grew to a smaller size and with decreased salt tolerance than wild-type (Ferreira et al. 2019). In plant GS2 mutants, a common side-effect is the increase in intracellular ammonia that causes chlorosis and activates the transcriptional increase of GS1 (Pérez-Delgado et al. 2013, Ferreira et al. 2019). Supposedly, the same effect of increased ammonia occurs in the Pt_GS2-KO but without noticeable chlorotic impacts. High-light conditions that plague plant mutants have been shown to enhance photorespiration in  wild-type diatoms (Schnitzler Parker et al. 2004, Parker et al. 2005). However, Pt_GS2-KO did not exhibit any growth differences in growth on high-light and low-light conditions (data not shown).  In Pt_GS2-KO, instead of GS1 compensation, GS3 may be the compensatory enzyme when cells are deficient in GS2.

With growth being retained for Pt_GS2-KO, it was not surprising that nitrate drawdown was also not inhibited. Rather, the cells were able to transport more nitrate from the media between 24 and 48 hours after growth began for continuous nitrate growth (Figure 2.5c) and when cells were switched from ammonium to nitrate (Figure 2.5d). The nitrate drawdown thereafter equalized between Pt_GS2-KO and Pt_WT. Here, nitrate drawdown was temporarily increased for Pt_GS2-KO in the lag to early exponential phase of growth. With the typically nitrate assimilation pathway being shunted in Pt_GS2-KO by lack of a functional GS2 enzyme, the cells may have perceived either a decrease of chloroplastic glutamine/glutamate levels or ammonium (due to re-allocation away from the chloroplast to mitochondria GS3) as a signal to replenish the chloroplast with more nitrate. Although, by the 48 hour to 72 hour growth period the excess nitrate replenishment halted because metabolite homeostasis was reached, indicative by the comparable growth and nitrate uptake rates between Pt_GS2-KO and Pt_WT.

Carbon uptake and oxygen evolution were also impacted in Pt_GS2-KO cell lines. Here, carbon uptake and oxygen evolution was negatively impacted at short time-scales (thirty minutes) and at low cell concentrations (1e4 cells/mL). *Phaeodactylum* typically divides once per 24 hours. Since the cells were not actively dividing in this experiment, the carbon uptake and oxygen evolution measurements can be considered instantaneous measurements. Also, for the MIMs system here, a cell concentration of 1e6 cells/mL would cause light attenuation in the culture and would add a layer of complexity to the measurements. Considering that growth and TOC was not impacted on nitrate growth, it can be assumed that carbon uptake and oxygen evolution was also not impacted at longer time scales (hours to days). If a diminished carbon uptake rate persisted throughout growth for Pt_GS2-KO, then cell density TOC per volume (Supplemental Figure 2.8) would have been negatively impacted.

Upon these initial findings, metabolic homeostasis was hypothesized to be altered in Pt_GS2-KO cell lines since the cells must reconstitute its nitrate assimilation pathway away from GS2 condensation of ammonium. Nitrogen depletion and limitation was known to induce TAG biogenesis in *Phaeodactylum* (Levitan et al. 2015a, Levitan et al. 2015b, Alipanah et al. 2015, McCarthy et al. 2017) and therefore lipid detection methodologies were used to assess if an altered nitrate assimilation pathway could trigger TAG biosynthesis. Lipid body staining revealed that Pt_GS2-KO cell lines accumulate lipids during early exponential phases of growth on nitrate and ammonium media where Pt_WT accumulated lipids later during stationary phases of growth (Figure 2.7). Another way TAG biogenesis occurs in *Phaeodactylum* is when cells deplete nitrogen from the media and reach a stationary phase of growth. Both Pt_GS2-KO and Pt_WT cell lines accumulated lipids during the stationary phase. Furthermore, the addition of glutamine to the growth media reverse lipid accumulation in Pt_GS2-KO (Figure 2.7, Figure 2.10). Here, glutamine supplementation seems to account for the lack of active GS2 enzyme and reverses the increased lipid content in Pt_GS2-KO cells. Glutamine supplementation skips the enzymatic step of producing it by GS2 and GS3 and directly supplies substrate for GOGAT enzymes to produce glutamate. By doing so, cellular homeostasis was purportedly reestablished in Pt_GS2-KO cell lines.

The extent to which Pt_GS2-KO accumulated lipids during early exponential and late exponential/stationary phases of growth were analysed on multiple nitrogen sources. A growth experiment was performed (Supplemental Figure 2.4) that was sampled simultaneously for TOC/TON and lipid analysis by FAME. Sampling occurred at 12 and 24 hours and those time points were considered to be lag phases (12 hours) and early exponential phases (24 hours). Sampling was also performed during late exponential/stationary phases (108 hours). The phases

of growth were demarcated by the position in the growth curve (Supplemental Figure 2.5) and by the reported relative health of the photosynthetic apparatus via measurements of Fv/Fm (Supplemental Figure 2.6), a commonly used method to indicate growth phases (McCarthy et al. 2017).

During lag and early exponential growth, TOC and TON levels were equivalent between nitrogen sources and cell lines. At 108 hours, the TOC per mL values differed. In ammonium and glutamine media, Pt_WT contained more TOC. In nitrate media, Pt_GS2-KO contained more TOC (Figure 2.8). At 108 hours, Pt_GS2-KO had a higher cell density than Pt_WT. It should be noted that TOC per cell at 108 hours was only slightly elevated but not significantly. Regardless of an insignificant increase of TOC per cell in Pt_GS2-KO, the proportion of TAG per TOC was significantly elevated during lag phase (12 hours) and during late exponential phase (108 hours) for Pt_GS2-KO on nitrate media (Supplemental Figure 2.9, Figure 2.9). Therefore, regardless of the comparable TOC levels and growth metrics between the two cell lines, Pt_GS2-KO increased the cellular TAG content compared to Pt_WT cells on nitrate media.

The most abundant fatty acids in *Phaeodactylum* are 14:0, 16:0, 16:1, and 20:5 (Dolch et al. 2015). During the 12 hours time-point, only 16:0 and 16:1 were increased in Pt_GS2-KO cell lines in abundance. At 108 hours, the abundance of 14:0, 16:0, 16:1, and 20:5 fatty acids were all greater in Pt_GS2-KO cell lines. (Figure 2.9b). Being that overall TAG was elevated at these two time points, this was not surprising.

The compositional profile of the fatty acids were also altered in Pt_GS2-KO during nitrate growth. At 12 hours, 16:0 and 16:1 fatty acid species were elevated while 20:5 was diminished (Supplemental Figure 2.8a). At 108 hours, only 16:0 and 16:1 were elevated (Supplemental Figure 2.8c). The synthesis of 16:0 and 16:1 occurs in the chloroplast stroma

while 20:5 is synthesized by elongases and desaturases in the endomembrane space (Dolch et al. 2015, Hao et al. 2018). This seems to indicate that the production of fatty acids within the stroma (16:0 and 16:1) increased in reaction to nitrate growth in Pt_GS2-KO cell lines while the activity of elongation and desaturation factors that synthesize 20:5 remained unchanged.

The reversal of TAG accumulation in Pt_GS2-KO due to glutamine supplementation was not surprising (Figure 2.7, Figure 2.10). Being that GS2 is not present in the mutant cell line, direct supplementation of glutamine aided in dampening the impact of a dysfunctional GS-GOGAT cycle in the chloroplast. This finding points to the fact that a lack of glutamine in the chloroplast causes TAG biogenesis to increase. Glutamine is condensed with 2-OG to produce glutamate by the GOGAT enzyme. Perhaps, glutamine supplementation resupplies chloroplastic GOGAT in order to produce sufficient glutamate molecules. Thus, diminished glutamine and glutamate levels in the chloroplast could have been the root cause of cellular stress in Pt_GS2-KO to trigger increased TAG biogenesis.

TAG biosynthesis in *Phaeodactylum* mutants deficient in GS2 was enhanced in nitrate growth conditions without a compromise in cell growth. Prior genetic engineering applications to improve lipid biosynthesis in diatom targeted lipid metabolism genes (Trentacoste et al. 2013, Hao et al. 2018). Here, lipid biogenesis was enhanced indirectly by targeting a nitrogen metabolism gene. The carbon repartitioning that occurs to increase lipids in response to a lack of GS2 is the first example of nitrogen-gene genetic engineering for the purpose of lipid biogenesis in a diatom. Though other nitrogen genes may not appear as obvious targets, this example alludes at the possibility that others may have a similar impact as GS2.

GS-GOGAT cycles are not the only redundant nitrogen metabolic hubs in *Phaeodactylum*. Two theoretical metabolic shunts, the ornithine-glutamine shunt (Levering et al.

2016) and the aspartate-alanine shuttle (Smith et al. 2019), reportedly trade metabolites between the chloroplast and the mitochondria. The redundancy of enzymes involved in the chloroplast and mitochondria sections of these shunts may be poignant targets for mutagenesis to indirectly increase lipid biosynthesis. Though just a hypothesis, these unique nitrogen metabolic features may lend credence to further explore the potential for diatoms to be industrial producers of lipids and biodiesel precursors.

### 2.6.5 Methods and materials

#### 2.6.5.1 Strain and cell culturing information

A reference strain of *Phaeodactylum tricornutum* (CCAP-1055/1) was used for the genetic transformation of episomal plasmids. *Phaeodactylum* was grown at 18°C under white fluorescent lights (50 µE m–2 s–1) and subjected to a diel growth cycle (14 h:10 h; light:dark) for all culturing experiments. Culture medium was artificial sea water (ASW) supplemented with trace metals, essential vitamins, 55 µM of NaPO$_4$, and 880 µM of the appropriate nitrogen source (NaNO$_3$, NH$_4$Cl, or L-glutamine). Cultures, when not used in growth assays, were maintained with chloramphenicol antibiotic (10 mg/L) or Provasoli's antibiotics cocktail to keep cultures bacteria-free. Transformed *Phaeodactylum* strains were supplemented with either phleomycin (50 mg/L) or zeocin (50 mg/L) to maintain clonality.

*Escherichia coli* (TOP-10, Life Technologies, Carlsbad, CA) was used for all plasmid cloning. The cultures were grown on Luria-Bertani broth or agar and supplemented with the following antibiotics when necessary: ampicillin (100 mg/L), carbenicillin (100 mg/L), tetracycline-HCl (10 mg/L), gentamicin (20 mg/L), zeocin (25 mg/L). A strain of *Escherichia coli* (epi300, Lucigen Corporation, Middleton, WI) that contained the plasmid pTA-MOB for

bacterial conjugation was used for electro-competent genetic transformation of episomal plasmids (Karas et al. 2015).

### 2.6.5.2 Molecular Biology

Plasmid cloning was performed in reference to Sambrook and Russell (2006) unless stated otherwise. The DNA polymerases were used for cloning purposes include Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, MA, United States), AccuPrime Taq High-Fidelity DNA polymerase (Thermo Fisher Scientific), OneTaq 2X Master Mix with standard buffer (New England Biolabs, Ipswich, MA, United States), and Phire Plant Direct PCR Master Mix (Thermo Fisher Scientific). Enzymes to make a 2X GA master mix include Phusion HF DNA polymerase, T5 exonuclease (Thermo Fisher Scientific), and DNA Taq Ligase (Thermo Fisher Scientific) and were purchased separately and mixed with 5X isothermal buffer. Enzymes for GG cloning were also purchased separately but only mixed on the day of cloning. They included the type-II restriction enzymes BsaI-HFv2 and BbsI-HF (New England Biolabs) and a T4 DNA Ligase (Thermo Fisher Scientific) and were mixed with 10mM DTT, 10mM ATP, and 10X CutSmart Buffer.

The episomal plasmid map used for the subcellular localization of YFP-GS2 in *Phaeodactylum* can be viewed in Supplemental Figure 1. To construct the episomal plasmid, the YFP-GS2 expression construct was first cloned into a pUC19 vector backbone using GA. First, the promoter region upstream (region) and the terminator region downstream (region) of the nitrate reductase gene (Phatr3_J54983*)* were PCR amplified from genomic *Phaeodactylum* DNA to include GA cloning homology regions. The GS2 coding sequence was PCR amplified from Phaeodactylum cDNA to include GA cloning homology regions. Lastly, a YFP (Venus) was PCR

amplified from a YFP expression cassette used in McQuaid et al. 2018 to include GA cloning homology regions. A five piece GA reaction was performed to produce the plasmid pUC19_NR-YFP-GS2. The full expression cassette was then PCR amplified to include GA cloning homology regions for episome cloning. The *Phaeodactylum* episomal plasmid PtpBR1 (Diner et al. 2016, Moosburner et al. 2020) was PCR amplified twice to set up a three-piece GA. The first episomal fragment was amplified to span the region ShBle-CenArsHis-Bla and the second was amplified to span the region Tet-OriT. The three PCR products, two from PtpBR1 and NR-YFP-GS2, were assembled using GA into epi300 (pTA-MOB) electrocompetent *E. coli*. Correct clones, verified using a combination of colony PCR, restriction digest assay, and Sanger sequencing, were conjugation transformation ready.

The episomal plasmid map for pBRCas9-gGS2(1,2,3) that was used for saturation mutagenesis of the GS2 gene in *Phaeodactylum* can be viewed in Figure 2.4.

First, sgRNA target sequences, known as spacers, were first identified. A web-based software (CRISP-Ex[9], Rastogi et al. 2016) to find compatible Cas9 target loci (spacer and protospacer adjacent motif (PAM)) CRISP-Ex was used to identify three spacer sequences that guide Cas9 to target the 5' end of the second exon of GS2. The genomic sequence for the second exon of GS2 was extracted from the *Phaeodactylum* genome browser[10]. Three target loci that did not have off-targets within the *Phaeodactylum* genome were chosen. The spacer sequences without the PAM were then assembled into GG compatible DNA inserts for GG1 cloning (Moosburner et al. 2020).

The GG assembly of the episome pBRCas9_gGS2(1,2,3) followed the cloning protocols detailed in Moosburner et al. 2020.

---

[9] http://www.phytocrispex.biologie.ens.fr/
[10] http://protists.ensembl.org/

### 2.6.5.3 Bacterial-mediated conjugation transformation

Bacterial-mediated conjugation was used to introduce episomal plasmids to *Phaeodactylum*. For all transformations, 100 μL of dense *Phaeodactylum* (1e8 cells/mL) was plated on conjugation-based solid agar medium ($NH_4/NO_3$-ASW, 1% agar, 5% LB in $NH_4$-supplemented ASW) in 6-well cell culture plates and incubated for 4 days under light (18:6) and at 18°C. A transformation control (positive for colonies not Cas9 cleavage), the episome pBR_Cas9-2A-shble, 2XBsaI_gRFP1, was used where Cas9 expression would still occur without an accompanying sgRNA (Moosburner et al. 2020). A negative control episomal plasmid was built harboring the NAT gene that confers resistance to nourseothricin. Upon phleomycin selection, the transformed cells would not survive. Transformation efficiency for the conjugation-mediated delivery of episomes was 500–750 transformants per 1e8 phleomycin-resistant cells.

The day before transformation, 3 mL of LB containing selection (Amp, Tet, Gent) was inoculated with E. coli from a glycerol stock of a correct episomal clone. After growing overnight, 250 uL of culture was used to inoculate 12.5 mL of LB with the same antibiotics. The culture was grown for 3–4 hours or until an $OD_{600}$ of 0.8–1 was reached, pelleted, and suspended in 90μL of SOC media. Then, the *E. coli* was pipetted into a corresponding well containing *Phaeodactylum* and mixed by rotating the 6-well plate such that the *E. coli* covered the all of *Phaeodactylum* cells. After *E. coli* was added to all wells and allowed to dry under a vacuum PCR hood, the 6-well plate was incubated in the dark at 30°C for 90 minutes. The plate was then incubated at 18°C and exposed to light for 48 hours.

Finally, each well was scraped with a sterile cell spreader using 1mL ASW media and collected in a 2 mL Epitube then re-plated on 100 mm selection plates containing phleomycin. The selection plates were then allowed to grow for 10–14 days or until colonies appeared. Colonies for both episomal vectors were directly picked into Phire PCR mix for colony PCR verification. For Phaeodactylum transformed with pBR_NR-YFP-GS2, the episomal vector expression cassette for YFP-GS2 was PCR amplified. For *Phaeodactylum* transformed with pBR_GS2(1,2,3), the GS2 gene within genomic DNA was PCR amplified.

### 2.6.5.4 Fluorescent protein localization

*Phaeodactylum* cell lines transformed with the episome pBR_NR-YFP-GS2 were imaged by confocal microscopy (Figure 2.3c). Cell lines were first cultured on ammonium media until mid-exponential phase, pelleted, and resuspended in nitrate media. Three days later, cells were imaged. The images were collected using a 100X objective. The autofluorescence of the cells chlorophyll was imaged using a band-pass filter between 650nm and 720nm. The fluorescence of YFP-GS2 was imaged using a band-pass filter between 520-nm and 540-nm. Images were captured at 100 Hz speed and a resolution of 1024x1024 pixels.

### 2.6.5.5 Cas9 mutagenesis screening and genotyping

As stated previously, *Phaeodactylum* colonies were picked directly off selection plates and into Phire PCR mix to screen for targeted mutagenesis of GS2. Since multiple sgRNAs were transformed into each selected cell, a deletion of 256-bp (between gGS2-1 and gGS2-3), 184-bp (between gGS2-1 and gGS2-2), or 72-bp (between gGS2-2 and gGS2-3) was expected. The expected deletions would also be visible by gel electrophoresis. It was also reasonably expected

that indel mutations caused by NHEJ-mediated double-stranded break repair at the individual target sites were expected, too. 36 transformed *Phaeodactylum* colonies were PCR amplified using primers (GS2_KO_Seq3 and GS2_KO_Seq4) that flank all three sgRNA target loci. The PCR amplicons were run on a gel and Sanger sequenced. Three colonies were then re-picked off selection plates and resuspended in liquid ASW media supplemented with phleomycin and Provasoli's antibiotic cocktail.

### 2.6.5.6 Western blot analysis

The three *Phaeodactylum* cell lines that were picked into liquid media were cultured until mid-exponential phase and 50μL of cells were pelleted and flash frozen in liquid nitrogen for protein analysis. Next day, the cell pellets were resuspended in 100μL of protein lysis buffer and mixed. The pellets were sonicated on high setting for 5 minutes (30 seconds on, 60 seconds off. Cell debris was then pelleted by centrifugation at 13,500 rpm for 45 minutes. The protein layer (top) was collected into a fresh tube.

A coomassie Bradford Assay was performed to quantify the protein concentrations. A standard curve was generated using serial dilutions of bovine serum albumin (BSA). Proteins were quantified using three replicates.

The protein gel was run using a NuPAGE (8-12%) Novex Bis-Tris Mini gel. 10ug of total protein was run per sample. After, the gel was run and transferred from the gel to imaging paper. Blocking agent was administered for 10-12 hours overnight. The next morning, a primary antibody solution containing the GS2 and NR protein antibody specific for *Phaeodactylum* was administered for 6-8 hours. Next, the secondary antibody was administered for 50 minutes.

Lastly, a chemiluminescent stain was pipetted on top of the whole gel membrane paper for 5 minutes to bind to the secondary antibody. The paper membrane was then imaged.

Pt_GS2-KO cell line #2 was used for the following physiological growth experiments due to its clear excision of 184-bp in the gel and no presence of GS2 protein in the protein fraction compared to Pt_WT.

### 2.6.5.7 $CO_2$ uptake, oxygen evolution, and NPQ measurements

Pt_GS2-KO were measured for carbon uptake, oxygen evolution, and non-photochemical quenching (NPQ) measurements simultaneously using a custom membrane inlet mass spectrometer (MIMs) system that recorded fine-scale pH oscillations (Du et al. 2018). Pt_GS2-KO and Pt_WT cell lines were diluted in nitrate media (880μM) to 1000 cells/mL to limit light attenuation that may occur in more dense cultures. One mL of cells was used in the MIMs system. The time-scale of the measurements was 30 minutes maximum and therefore it can be assumed that cell doubling did not occur during the measurements. Light intensity during the experiment was 200μmole quanta m-2 s-1.

### 2.6.5.8 Nitrate drawdown

Nitrate drawdown experiments were performed comparing Pt_GS2-KO and Pt_WT. Cells were first pre-conditioned on either ammonium or nitrate media until mid-exponential phase, pelleted, and resuspended in nitrate media at a cell concentration of 1e6 cells/mL. Samples were collected daily for cell density, Fv/Fm, and ultraviolet-visible (UV-vis) spectroscopy. UV-vis spectroscopy was used to measure the nitrate concentration in cell-free media where nitrate absorbs light at 220nm. Cells were removed from the sample medium by filtering one mL of cell

culture through a 0.22μm membrane filter disc. Three samples were collected for each time point and UV-vis measurements were performed three times for each time point.

### 2.6.5.9 Pt_GS2-KO cell culturing and sampling

For TOC/TON, FAME, and RNA-seq sampling, a single growth experiment was set up (Supplemental Figure 2.9a). Before the start of the experiment, a PCR was performed to ensure the genotype of Pt_GS2-KO cell line 2 was retained (Supplemental Figure 2.9b). For all growth conditions and cell lines, cells were cultured in four liter cylindrical bottles and bubbled with filtered air. Cells were cultured on a 14:10 diel growth cycle at a light intensity of 300μmole quanta m-2 s-1.

Pt_GS2-KO and Pt_WT were grown in the four liters of ammonium media for 48 hours, each. Four two liter cultures were prepared without nitrogen supplementation and pre-chilled at 18°C. From the four liter cultures, the cell density was calculated and 2e6 cells were pelleted four times. The cells were then resuspended in the chilled two liter cultures and incubated for two hours (N-free wash). After two hours, nitrogen sources were added to the cultures (nitrate, 880μM, ammonium 880μM, L-glutamine, 440μM) and cultured for an additional 156 hours.

Sampling time points were performed at 0 hours (immediately before nitrogen addition), 12 hours, 24 hours, and 108 hours. Therefore, the 0 hour time-point is considered the N-free pre-condition. At each time-point, TOC/TON, FAME, and RNAseq samples were collected simultaneously. The sampling regime was conducted three times and data was analyzed using all three replicates.

Throughout the sampling experiments, the cell density of each cell line and condition was monitored and no significant differences in growth was observed between Pt_WT and

149

Pt_GS2-KO cell lines for any of the nitrogen conditions (Supplemental Figure 2.10). Instantaneous photosynthetic efficiency was also measured along the growth profiles by measuring Fv/Fm (Supplemental Figure 2.11).

### 2.6.5.10 TOC and TON

At each time-point, 23mL of culture was collected and filtered on 0.22µM disc filters. The supernatant was discarded to rid of any organic carbon or nitrogen that was not taken up by the cells. The filters were inserted into 1.5mL epitubes and flash frozen. The filters were then slowly thawed on dry ice and resuspended in 20mL of dH$_2$O in 25mL glass vials suitable for TOC/TON analysis. TOC/TON analysis was performed using a Shimadzu TOC-L Laboratory Total Organic Carbon Analyzer fashioned with a TNM-L Total Nitrogen Unit for simultaneous TOC and TON measurements (Shimadzu Scientific Instruments, Columbia, Maryland).

### 2.6.5.11 FAME analysis

At each time-point, 100mL of culture was collected for fatty acid methyl ester (FAME) analysis. Cells were centrifuged for 15 minutes at 3000xg, resuspended in residual media, pelleted in 1.5mL epitubes for one minute at 3000xg, and flash frozen. Frozen pellets were then freeze dried using a lyophilizer overnight to remove all water. The dried pellets were then transferred to 4mL glass vials suitable for FAME analysis. For the continuous growth experiments, 25mL of culture was collected for FAME analysis.

Lipid content of microalgal cultures was determined through the conversion of acyl lipid species into their FAME derivatives and subsequent analysis via GC-FID. Briefly, odd-chain lipid species were added to the freeze-dried algal biomass of each sample as quantitative and

qualitative internal standards. The samples were treated with a solution of boron trifluoride in methanol and subjected to microwave-assisted transesterification (Anton Paar Multiwave Pro). The transesterification reaction was quenched with aqueous potassium hydroxide and FAMEs were phase separated via a liquid-liquid partition into heptanes. Chromatographic separation and detection of FAMEs in the heptanes layer was performed on an Agilent 6890 GC system fitted with a DB-FFAP capillary column. Hydrogen was used as a carrier gas.

### 2.6.6 Author Contributions

This manuscript was primarily authored by Mark Moosburner. Pardis Ghlomai aided in the generation of GS-GOGAT mutants, generating growth curves for the mutants, maintaining the cell lines in culture, and other molecular biology and cell biology tasks. Josh Gurr performed FAME analysis and produced raw FAME results. Du Niu conducted the carbon uptake, oxygen evolution, and NPQ measurements and supplied the raw data. James McCarthy and Sarah R. Smith were invaluable in providing transcriptomic data for GS-GOGAT genes and in providing knowledge regarding nitrogen metabolism in diatoms in general. Andrew E. Allen was the principal investigator for the study.

### 2.7 Acknowledgements for Chapter 2

Chapter 2, in part, is currently being prepared for submission for publication of the material. Moosburner, M., Gholami, P., Gurr, J., Du, N., McCarthy, J., Smith, S. A., Allen, A. E. (2021) Glutamine synthetase 2 knockout in the diatom Phaeodactylum tricornutum accumulates lipids early without compromising growth. In Preparation. Mark Moosburner was the primary author and principal investigator of this manuscript.

## 2.8 References for Chapter 2

Adler-Agnon, Z., S. Leu, A. Zarka, S. Boussiba and I. Khozin-Goldberg (2018). "Novel promoters for constitutive and inducible expression of transgenes in the diatom Phaeodactylum tricornutum under varied nitrate availability." Journal of Applied Phycology **30**(5): 2763-2772.

Ajjawi, I., J. Verruto, M. Aqui, L. B. Soriaga, J. Coppersmith, K. Kwok, L. Peach, E. Orchard, R. Kalb, W. Xu, T. J. Carlson, K. Francis, K. Konigsfeld, J. Bartalis, A. Schultz, W. Lambert, A. S. Schwartz, R. Brown and E. R. Moellering (2017). "Lipid production in Nannochloropsis gaditana is doubled by decreasing expression of a single transcriptional regulator." Nature Biotechnology **35**(7): 647-652.

Alipanah, L., J. Rohloff, P. Winge, A. M. Bones and T. Brembu (2015). "Whole-cell response to nitrogen deprivation in the diatom Phaeodactylum tricornutum." Journal of experimental botany **66**(20): 6281-6296.

Allen, A. E., C. L. Dupont, M. Oborník, A. Horák, A. Nunes-Nesi, J. P. McCrow, H. Zheng, D. A. Johnson, H. Hu, A. R. Fernie and C. Bowler (2011). "Evolution and metabolic significance of the urea cycle in photosynthetic diatoms." Nature **473**(7346): 203-207.

Allen, A. E., J. LaRoche, U. Maheswari, M. Lommer, N. Schauer, P. J. Lopez, G. Finazzi, A. R. Fernie and C. Bowler (2008). "Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation." Proceedings of the National Academy of Sciences **105**(30): 10438-10443.

Armbrust, E. V., J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kröger, W. W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson and D. S. Rokhsar (2004). "The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism." Science **306**(5693): 79-86.

Berges, J. A., D. O. Charlebois, D. C. Mauzerall and P. G. Falkowski (1996). "Differential effects of nitrogen limitation on photosynthetic efficiency of photosystems I and II in microalgae." Plant Physiology **110**(2): 689-696.

Betti, M., T. Arcondéguy and A. J. Márquez (2006). "Molecular analysis of two mutants from Lotus japonicus deficient in plastidic glutamine synthetase: functional properties of purified GLN2 enzymes." Planta **224**(5): 1068-1079.

Betti, M., M. García-Calderón, C. M. Pérez-Delgado, A. Credali, P. Pal'ove-Balang, G. Estivill, M. Repčák, J. M. Vega, F. Galván and A. J. Márquez (2014). "Reassimilation of ammonium in Lotus japonicus." Journal of experimental botany **65**(19): 5557-5566.

Bowler, C., A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, R. P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. A. Berges, C. Brownlee, J. P. Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. Huysman, B. D. Jenkins, K. Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kroger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin-Jezequel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M. P. Oudot-Le Secq, C. Napoli, M. Obornik, M. S. Parker, J. L. Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. A. Rynearson, J. Schmutz, H. Shapiro, M. Siaut, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van de Peer and I. V. Grigoriev (2008). "The Phaeodactylum genome reveals the evolutionary history of diatom genomes." Nature **456**(7219): 239-244.

Boyle, N. R., M. D. Page, B. Liu, I. K. Blaby, D. Casero, J. Kropat, S. J. Cokus, A. Hong-Hermesdorf, J. Shaw, S. J. Karpowicz, S. D. Gallaher, S. Johnson, C. Benning, M. Pellegrini, A. Grossman and S. S. Merchant (2012). "Three Acyltransferases and Nitrogen-responsive Regulator Are Implicated in Nitrogen Starvation-induced Triacylglycerol Accumulation in Chlamydomonas*." Journal of Biological Chemistry **287**(19): 15811-15825.

Brown, K. L., K. I. Twing and D. L. Robertson (2009). "UNRAVELING THE REGULATION OF NITROGEN ASSIMILATION IN THE MARINE DIATOM THALASSIOSIRA PSEUDONANA (BACILLARIOPHYCEAE): DIURNAL VARIATIONS IN TRANSCRIPT LEVELS FOR FIVE GENES INVOLVED IN NITROGEN ASSIMILATION 1." Journal of phycology **45**(2): 413-426.

Cermeño, P., J.-B. Lee, K. Wyman, O. Schofield and P. G. Falkowski (2011). "Competitive dynamics in two species of marine phytoplankton under non-equilibrium conditions." Marine ecology progress series **429**: 19-28.

Chu, L., D. Ewe, C. R. Bártulos, P. G. Kroth and A. Gruber (2016). "Rapid induction of GFP expression by the nitrate reductase promoter in the diatom Phaeodactylum tricornutum." PeerJ **4**: e2344.

Daboussi, F., S. Leduc, A. Marechal, G. Dubois, V. Guyot, C. Perez-Michaut, A. Amato, A. Falciatore, A. Juillerat, M. Beurdeley, D. F. Voytas, L. Cavarec and P. Duchateau (2014). "Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology." Nat Commun **5**: 3831.

Delalat, B., V. C. Sheppard, S. Rasi Ghaemi, S. Rao, C. A. Prestidge, G. McPhee, M.-L. Rogers, J. F. Donoghue, V. Pillay, T. G. Johns, N. Kröger and N. H. Voelcker (2015). "Targeted drug delivery using genetically engineered diatom biosilica." Nature Communications **6**(1): 8791.

Diner, R. E., V. A. Bielinski, C. L. Dupont, A. E. Allen and P. D. Weyman (2016). "Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods." Front Bioeng Biotechnol **4**: 65.

Dolch, L.-J. and E. Maréchal (2015). "Inventory of fatty acid desaturases in the pennate diatom Phaeodactylum tricornutum." Marine drugs **13**(3): 1317-1339.

Du, N., P. Gholami, D. I. Kline, C. L. DuPont, A. G. Dickson, D. Mendola, T. Martz, A. E. Allen and B. G. Mitchell (2018). "Simultaneous quantum yield measurements of carbon uptake and oxygen evolution in microalgal cultures." PloS one **13**(6): e0199125.

Dugdale, R. C. and F. P. Wilkerson (1998). "Silicate regulation of new production in the equatorial Pacific upwelling." Nature **391**(6664): 270-273.

Eizadora, T. Y., F. J. Zendejas, P. D. Lane, S. Gaucher, B. A. Simmons and T. W. Lane (2009). "Triacylglycerol accumulation and profiling in the model diatoms Thalassiosira pseudonana and Phaeodactylum tricornutum (Bacillariophyceae) during starvation." Journal of Applied Phycology **21**(6): 669-681.

Falkowski, P. G., M. E. Katz, A. H. Knoll, A. Quigg, J. A. Raven, O. Schofield and F. Taylor (2004). "The evolution of modern eukaryotic phytoplankton." Science **305**(5682): 354-360.

Falkowski, P. G. and M. J. Oliver (2007). "Mix and match: how climate selects phytoplankton." Nature reviews microbiology **5**(10): 813-819.

Ferrara, M. A., P. Dardano, L. De Stefano, I. Rea, G. Coppola, I. Rendina, R. Congestri, A. Antonucci, M. De Stefano and E. De Tommasi (2014). "Optical properties of diatom nanostructured biosilica in Arachnoidiscus sp: micro-optics from mother nature." PLoS One **9**(7): e103750.

Ferreira, S., E. Moreira, I. Amorim, C. Santos and P. Melo (2019). "Arabidopsis thaliana mutants devoid of chloroplast glutamine synthetase (GS2) have non-lethal phenotype under photorespiratory conditions." Plant Physiology and Biochemistry **144**: 365-374.

Forde, B. G. and P. J. Lea (2007). "Glutamate in plants: metabolism, regulation, and signalling." Journal of experimental botany **58**(9): 2339-2358.

Gutiérrez, C. D. B., D. L. R. Serna, C. A. C. Alzate and C. Ariel (2017). "A comprehensive review on the implementation of the biorefinery concept in biodiesel production plants." Biofuel Research Journal **4**(3): 691.

Hao, X., L. Luo, J. Jouhet, F. Rébeillé, E. Maréchal, H. Hu, Y. Pan, X. Tan, Z. Chen, L. You, H. Chen, F. Wei and Y. Gong (2018). "Enhanced triacylglycerol production in the diatom Phaeodactylum tricornutum by inactivation of a Hotdog-fold thioesterase gene using TALEN-based targeted mutagenesis." Biotechnology for Biofuels **11**(1): 312.

Hildebrand, M., A. K. Davis, S. R. Smith, J. C. Traller and R. Abbriano (2012). "The place of diatoms in the biofuels industry." Biofuels **3**(2): 221-240.

Hockin, N. L., T. Mock, F. Mulholland, S. Kopriva and G. Malin (2012). "The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants." Plant physiology **158**(1): 299-312.

Hopes, A., V. Nekrasov, S. Kamoun and T. Mock (2016). "Editing of the urease gene by CRISPR-Cas in the diatom Thalassiosira pseudonana." Plant Methods **12**: 49.

Hu, Q., M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert and A. Darzins (2008). "Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances." The plant journal **54**(4): 621-639.

Jamwal, V. L., N. Kapoor and S. G. Gandhi (2020). Biotechnology of Biofuels: Historical Overview, Business Outlook and Future Perspectives. Biotechnology Business-Concept to Delivery, Springer**:** 109-127.

Jones, C. S. and S. P. Mayfield (2012). "Algae biofuels: versatility for the future of bioenergy." Current opinion in biotechnology **23**(3): 346-351.

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

Khan, S. and P. Fu (2020). "Biotechnological perspectives on algae: a viable option for next generation biofuels." Current opinion in biotechnology **62**: 146-152.

Lancien, M., P. Gadal and M. Hodges (2000). "Enzyme redundancy and the importance of 2-oxoglutarate in higher plant ammonium assimilation." Plant Physiology **123**(3): 817-824.

Levering, J., J. Broddrick, C. L. Dupont, G. Peers, K. Beeri, J. Mayers, A. A. Gallina, A. E. Allen, B. O. Palsson and K. Zengler (2016). "Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom." PloS one **11**(5): e0155038.

Levitan, O., J. Dinamarca, E. Zelzion, M. Y. Gorbunov and P. G. Falkowski (2015). "An RNA interference knock−down of nitrate reductase enhances lipid biosynthesis in the diatom Phaeodactylum tricornutum." The Plant Journal **84**(5): 963-973.

Levitan, O., J. Dinamarca, E. Zelzion, D. S. Lun, L. T. Guerra, M. K. Kim, J. Kim, B. A. Van Mooy, D. Bhattacharya and P. G. Falkowski (2015). "Remodeling of intermediate metabolism in the diatom Phaeodactylum tricornutum under nitrogen stress." Proceedings of the National Academy of Sciences **112**(2): 412-417.

Longworth, J., D. Wu, M. Huete-Ortega, P. C. Wright and S. Vaidyanathan (2016). "Proteome response of Phaeodactylum tricornutum, during lipid accumulation induced by nitrogen depletion." Algal research **18**: 213-224.

Masclaux-Daubresse, C., M. Reisdorf-Cren, K. Pageau, M. Lelandais, O. Grandjean, J. Kronenberger, M.-H. Valadier, M. Feraud, T. Jouglet and A. Suzuki (2006). "Glutamine synthetase-glutamate synthase pathway and glutamate dehydrogenase play distinct roles in the sink-source nitrogen cycle in tobacco." Plant physiology **140**(2): 444-456.

Matthijs, M., M. Fabris, S. Broos, W. Vyverman and A. Goossens (2016). "Profiling of the early nitrogen stress response in the diatom Phaeodactylum tricornutum reveals a novel family of RING-domain transcription factors." Plant physiology **170**(1): 489-498.

Matthijs, M., M. Fabris, T. Obata, I. Foubert, J. M. Franco−Zorrilla, R. Solano, A. R. Fernie, W. Vyverman and A. Goossens (2017). "The transcription factor bZIP14 regulates the TCA cycle in the diatom Phaeodactylum tricornutum." The EMBO journal **36**(11): 1559-1576.

McCarthy, J. K., S. R. Smith, J. P. McCrow, M. Tan, H. Zheng, K. Beeri, R. Roth, C. Lichtle, U. Goodenough, C. P. Bowler, C. L. Dupont and A. E. Allen (2017). "Nitrate Reductase Knockout Uncouples Nitrate Transport from Nitrate Assimilation and Drives Repartitioning of Carbon Flux in a Model Pennate Diatom." The Plant Cell **29**(8): 2047-2070.

Mishra, V. K. and R. Goswami (2018). "A review of production, properties and advantages of biodiesel." Biofuels **9**(2): 273-289.

Moosburner, M. A., P. Gholami, J. K. McCarthy, M. Tan, V. A. Bielinski and A. E. Allen (2020). "Multiplexed Knockouts in the Model Diatom Phaeodactylum by Episomal Delivery of a Selectable Cas9." Front Microbiol **11**: 5.

Nelson, D. M., P. Tréguer, M. A. Brzezinski, A. Leynaert and B. Quéguiner (1995). "Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation." Global Biogeochemical Cycles **9**(3): 359-372.

Niu, Y.-F., M.-H. Zhang, D.-W. Li, W.-D. Yang, J.-S. Liu, W.-B. Bai and H.-Y. Li (2013). "Improvement of neutral lipid and polyunsaturated fatty acid biosynthesis by overexpressing a type 2 diacylglycerol acyltransferase in marine diatom Phaeodactylum tricornutum." Marine drugs **11**(11): 4558-4569.

Nunes-Nesi, A., A. R. Fernie and M. Stitt (2010). "Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions." Molecular plant **3**(6): 973-996.

Nymark, M., A. K. Sharma, T. Sparstad, A. M. Bones and P. Winge (2016). "A CRISPR/Cas9 system adapted for gene editing in marine algae." Scientific reports **6**(1): 1-6.

Parker, M. S. and E. V. Armbrust (2005). "Synergistic Effects of Light, Temperature, and Nitrogen Source on Transcription of Genes for Carbon and Nitrogen Metabolism in the Centric Diatom Thalassiosira Pseudonana (Bacillariophyceae) 1." Journal of Phycology **41**(6): 1142-1153.

Pérez-Delgado, C. M., M. García-Calderón, D. H. Sánchez, M. K. Udvardi, J. Kopka, A. J. Márquez and M. Betti (2013). "Transcriptomic and metabolic changes associated with photorespiratory ammonium accumulation in the model legume Lotus japonicus." Plant Physiology **162**(4): 1834-1848.

Pulz, O. and W. Gross (2004). "Valuable products from biotechnology of microalgae." Applied microbiology and biotechnology **65**(6): 635-648.

Radakovits, R., P. M. Eduafo and M. C. Posewitz (2011). "Genetic engineering of fatty acid chain length in Phaeodactylum tricornutum." Metabolic engineering **13**(1): 89-95.

Ramachandra, T., D. M. Mahapatra and R. Gordon (2009). "Milking diatoms for sustainable energy: biochemical engineering versus gasoline-secreting diatom solar panels." Industrial & Engineering Chemistry Research **48**(19): 8769-8788.

Rastogi, A., O. Murik, C. Bowler and L. Tirichine (2016). "PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing." BMC bioinformatics **17**(1): 1-4.

Robertson, D. L. and A. Tartar (2006). "Evolution of glutamine synthetase in heterokonts: evidence for endosymbiotic gene transfer and the early evolution of photosynthesis." Molecular biology and evolution **23**(5): 1048-1055.

Sambrook, J. and D. W. Russell (2006). The condensed protocols from molecular cloning: a laboratory manual. (No. Sirsi) i9780879697723).

Sanz-Luque, E., A. Chamizo-Ampudia, A. Llamas, A. Galvan and E. Fernandez (2015). "Understanding nitrate assimilation and its regulation in microalgae." Frontiers in plant science **6**: 899.

Savidge, G., D. Harbour, L. Gilpin and P. Boyd (1995). "Phytoplankton distributions and production in the Bellingshausen Sea, Austral spring 1992." Deep Sea Research Part II: Topical Studies in Oceanography **42**(4-5): 1201-1224.

Schnitzler Parker, M., E. V. Armbrust, J. Piovia−Scott and R. G. Keil (2004). "INDUCTION OF PHOTORESPIRATION BY LIGHT IN THE CENTRIC DIATOM THALASSIOSIRA WEISSFLOGII (BACILLARIOPHYCEAE): MOLECULAR CHARACTERIZATION AND PHYSIOLOGICAL CONSEQUENCES 1." Journal of Phycology **40**(3): 557-567.

Scott, S. A., M. P. Davey, J. S. Dennis, I. Horst, C. J. Howe, D. J. Lea-Smith and A. G. Smith (2010). "Biodiesel from algae: challenges and prospects." Current opinion in biotechnology **21**(3): 277-286.

Serif, M., G. Dubois, A. L. Finoux, M. A. Teste, D. Jallet and F. Daboussi (2018). "One-step generation of multiple gene knock-outs in the diatom Phaeodactylum tricornutum by DNA-free genome editing." Nat Commun **9**(1): 3924.

Shrestha, R. P., B. Tesson, T. Norden-Krichmar, S. Federowicz, M. Hildebrand and A. E. Allen (2012). "Whole transcriptome analysis of the silicon response of the diatom Thalassiosira pseudonana." BMC genomics **13**(1): 1-16.

Siaut, M., M. Heijde, M. Mangogna, A. Montsant, S. Coesel, A. Allen, A. Manfredonia, A. Falciatore and C. Bowler (2007). "Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum." Gene **406**(1-2): 23-35.

Slattery, S. S., A. Diamond, H. Wang, J. A. Therrien, J. T. Lant, T. Jazey, K. Lee, Z. Klassen, I. Desgagne-Penix, B. J. Karas and D. R. Edgell (2018). "An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum." ACS Synth Biol **7**(2): 328-338.

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

Smith, S. R., C. L. Dupont, J. K. McCarthy, J. T. Broddrick, M. Oborník, A. Horák, Z. Füssy, J. Cihlář, S. Kleessen, H. Zheng, J. P. McCrow, K. K. Hixson, W. L. Araújo, A. Nunes-Nesi, A. Fernie, Z. Nikoloski, B. O. Palsson and A. E. Allen (2019). "Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom." Nature Communications **10**(1): 4552.

Smith, S. R., J. T. Gillard, A. B. Kustka, J. P. McCrow, J. H. Badger, H. Zheng, A. M. New, C. L. Dupont, T. Obata, A. R. Fernie and A. E. Allen (2016). "Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation." PLoS genetics **12**(12): e1006490.

Stephenson, P. G., C. M. Moore, M. J. Terry, M. V. Zubkov and T. S. Bibby (2011). "Improving photosynthesis for algal biofuels: toward a green revolution." Trends in biotechnology **29**(12): 615-623.

Tachibana, M., A. E. Allen, S. Kikutani, Y. Endo, C. Bowler and Y. Matsuda (2011). "Localization of putative carbonic anhydrases in two marine diatoms, Phaeodactylum tricornutum and Thalassiosira pseudonana." Photosynthesis Research **109**(1): 205-221.

Takabayashi, M., F. P. Wilkerson and D. Robertson (2005). "Response of glutamine synthetase gene transcription and enzyme activity to external nitrogen sources in the diatom skeletonema costatum (bacillariophyceae) 1." Journal of Phycology **41**(1): 84-94.

Trentacoste, E. M., R. P. Shrestha, S. R. Smith, C. Glé, A. C. Hartmann, M. Hildebrand and W. H. Gerwick (2013). "Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth." Proceedings of the National Academy of Sciences **110**(49): 19748-19753.

Vecchi, V., S. Barera, R. Bassi and L. Dall'Osto (2020). "Potential and challenges of improving photosynthesis in algae." Plants **9**(1): 67.

Vinayak, V., K. M. Manoylov, H. Gateau, V. Blanckaert, J. Hérault, G. Pencréac'h, J. Marchand, R. Gordon and B. Schoefs (2015). "Diatom Milking: A Review and New Approaches. Marine Drugs 2015, 13, 2629–2665." Marine drugs **13**(12): 7301.

Wallsgrove, R. M., J. C. Turner, N. P. Hall, A. C. Kendall and S. W. Bright (1987). "Barley mutants lacking chloroplast glutamine synthetase—biochemical and genetic analysis." Plant Physiology **83**(1): 155-158.

Wijffels, R. H. and M. J. Barbosa (2010). "An outlook on microalgal biofuels." Science **329**(5993): 796-799.

Yoneda, K., M. Yoshida, I. Suzuki and M. M. Watanabe (2016). "Identification of a major lipid droplet protein in a marine diatom Phaeodactylum tricornutum." Plant and Cell Physiology **57**(2): 397-406.

Zienkiewicz, K., Z.-Y. Du, W. Ma, K. Vollheyde and C. Benning (2016). "Stress-induced neutral lipid biosynthesis in microalgae—molecular, cellular and physiological insights." Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids **1861**(9): 1269-1281.

# Chapter 3


# CRISPR-Cas9 Nitrogen gene knockout library using a fluorescent protein lipid body reporter as an indicator of lipid biogenesis


Mark Moosburner

## 3.1 Introduction of Chapter 3

CRISPR-Cas9 revolutionized functional genetics in a plethora of species due to the ease of producing targeted mutations and precise site-specific gene edits. The targeting of single genes and characterizing their roles involved in a cellular phenotype or organismal trait has led to amazing findings such as the ability to reverse sickle cell anemia by changing a single nucleotide in blood cells in humans (Frangoul et al. 2021, Esrick et al. 2021). The potential to treat diseases in humans using Cas9-based therapies has sparked a new age in the medical interventions via functional genetics.

Diatom biology, too, has benefited from CRISPR-Cas9 functional genetics. Recent insights like the siderophore-based acquisition of iron from the environment (Coale et al. 2019) and the diel-based regulation of lipid β-oxidation (Jallet et al. 2020) have been bolstered by single-gene Cas9 mutagenesis in the diatom *Phaeodactylum tricornutum.*

Mutating one gene is often predicated on previously gained knowledge about the gene and its inferred function in cellular metabolism. That knowledge may be derived from studies into transcriptomics, proteomics, metabolomics, or physiological investigatory experiments that predict the function of a gene or genes. Nevertheless, targeting one gene by reverse genetics techniques is the gold-standard to investigate the function of said gene, whether it be by using zinc-finger nucleases, TALENs, or CRISPR-Cas9 technologies. By mutating a gene, and therefore producing a cell deficient in its functional protein product, the consequences to cellular metabolism and gene transcription can be analyzed in a number of ways in order to validate the predicted function of the gene and to possibly discover new functions. For instance, McCarthy et al. (2017) sought to knock out the nitrate reductase (NR) gene in *Phaeodactylum* and was already knowledgeable about the function in the context of nitrate assimilation. It was known that NR

was the single protein responsible for the reduction of nitrate to nitrite, that the regulation of the NR gene was under the control of a nitrate-inducible promoter, and that it was one of the most highly upregulated genes in the genome when the cells were grown on nitrate. After producing mutant cell lines deficient in NR, it was discovered that the *Phaeodactylum* vacuole dilated with excess nitrate being taken up by the cell. The diatom mutant retained the ability to uptake nitrate but could not reduce it. Thus, a new function of NR was discovered which was that it aided in the regulation of vacuolar nitrate uptake and release.

Targeted mutagenesis of one gene can validate predicted functions and reveal new functions of the gene. This is called reverse genetics. Forward genetics, on the other hand, is the experimental process where multiple genetic loci are mutated simultaneously, often at random, and a phenotypic assay is implemented to screen a mutant cell population for individual mutant cells that fit a specific phenotypic characteristic. Forward genetic screens have been used to develop novel genetic selection methods in *Phaeodactylum* (Sakaguchi et al. 2011, Taparia et al. 2019). For example, Taparia et al. (2019) exposed a *Phaeodactylum* culture to UV irradiation which damages double stranded DNA at random. The mutant cell population was then cultured with the bleaching herbicide norflurazon and the surviving cells' genomes were analyzed for mutations. It was revealed that an amino acid substitution of serine to arginine at Gly290 in the native phytoene desaturase-1 (PDS1) gene confers resistance to norflurazon. Similarly, random genome-wide mutagenesis was performed using the alkylation agent, N-ethyl-N-nitrosourea, in *Phaeodactylum* cells (Sakaguchi et al. 2011). This led to the identification of mutant cells that can survive on 5-fluoroorotic acid only when a uridine-5'-monophosphate synthase gene (UMPS) was knocked out. This discovery also bolstered experiments using Cas9-sgRNA ribonucleoprotein (RNP) complexes to co-target a gene of interest and UMPS in order to select

for mutants using 5-fluoroorotic acid instead of an antibiotic since an antibiotic resistant gene cassette was not introduced (Serif et al. 2018). These studies demonstrated that forward genetics can be used in *Phaeodactylum* to find and characterize individual genes that, when mutated, conferred resistance to lethal chemical agents.

In bacteria and eukaryotic organisms with a haploid life cycle, transposon-mediated mutagenesis and sequencing (TnSeq) has become a gold-standard in forward genetics screens (Shields & Jensen 2019, Fenster & Eckert 2021). By allowing transposons, mobile genetic elements, to recombine randomly throughout a cell's genome, mutant cell lines can be produced and the genomic loci where the transposon landed can be easily found by sequencing. The transposon has a known genetic sequence that can be used to sequenced outward into the 5' and 3' neighboring genomic sequence regions. With this technology, one transposon is introduced per cell in a culture and, thus, the transposon mobilizes to one genomic loci per cell. UV irradiation or chemical based mutagenesis can potentially mutate multiple loci per cell.

Random, genome-wide, mutagenesis is particularly beneficial when a live-or-dead screening assay is used. However, if the survival of cells is not used as a screening methodology, finding a mutant cell line that confers a specific phenotype becomes increasingly difficult. Without a live-or-dead assay, one would have to resort to screening many mutant cell lines and ones that have mutations in non-coding regions of the genome that wouldn't produce a phenotype whatsoever.

The ability to specifically target multiple genes simultaneously without relying on random mutagenesis, but to only target protein-coding gene loci, circumvents the difficulties of random mutagenesis screens. The CRISPR-Cas9 genetic engineering tool has demonstrated the capabilities to target multiple genes in a cell culture (but only one per cell) more easily than other

targeted mutagenesis tools like zinc finger nucleases and TALENs. Most notably, changing the genomic target using CRISPR-Cas9 only requires modulating the sgRNA spacer sequence, a 20-nt region. The same Cas9 enzyme can utilize all iterations of a sgRNA that retain the sgRNA backbone. TALENs, on the other hand, require the re-engineering of the entire protein to change genomic targets. One example where TALEN technology was chosen instead of CRISPR-Cas9 was to generate a knockout library of microRNA-encoding genes in human cells (Kim et al. 2013). microRNA gene sequences are short and may not contain the 5'-NGG-3' trinucleotide protospacer-adjacent motif (PAM) requirement for Cas9-sgRNA targeting. TALENs were chosen since they do not have any sequence targeting requirements and have more flexibility when targeting short sequences. Nevertheless, manufacturing TALEN constructs are time-consuming and laborious and typically not a preferred or commonly used method used to generate gene knockout libraries.

Applications of CRISPR-Cas9 knockout libraries have been demonstrated in a wide variety of organisms such as humans (Shalem et al. 2014, Hart et al. 2017), plants (Bao et al. 2019), mice (Chen et al. 2015), and yeast (Lian et al. 2018). Exceptionally, CRISPR-Cas9 knockout libraries have been heavily applied in cancer research towards the identification of new cancer drugs and oncogenic genes in cultured cell lines (Moses et al. 2019) and in animal models (Chow & Chen 2018). In agriculture applications, CRISPR-Cas9 screens have shown the potential to boost desirable crop traits in maize (Liu et al. 2020) and for studying C4 photosynthesis in green foxtail (Weiss et al. 2020) by deploying high throughput screening methodologies.

Currently, there are no reports of using CRISPR-Cas9 knockout libraries in any algal species. Though, in the green algae *Chlamydomonas reinhardtii*, a genome-wide insertional

mutagenesis platform, which randomly integrates in the genome, has been demonstrated to identify genes required for photosynthesis (Li et al. 2019). Here, high-throughput insertional mutagenesis and high throughput sequencing methodologies were converged to reveal 303 unique genes required for photosynthesis. The inserted sequence contained a randomized barcode sequence that allowed for each mutant to be tracked within a culture of pooled mutants. The pooled mutant culture was then grown in phototrophic conditions (light and $CO_2$ only) and heterotrophic conditions (dark and acetate) and the prevalence of individual barcodes were compared between the two conditions to reveal mutants that could not grow in the light but could grow in the dark. Despite not having to rely on a live-or-dead screen, the pooled mutant culture contained many mutants where the insert did not land within a protein-coding gene sequence. Certainly, the ability to specifically target all protein-coding genes using CRISPR-Cas9 would have greatly reduced the number of mutants to screen. Unfortunately, the expression of the Cas9 endonuclease in *Chlamydomonas reinhardtii* has proven very difficult. Success has only been obtained in cells stripped of their cell wall and while using customized electroporation wave-forms for genetic transformations (Greiner et al. 2017, Park et al. 2020). Interference CRISPR (CRISPRi) transcriptional repression has been reported in *Chlamydomonas reinhardtii* but relies on the expression and localization of a nuclease-null Cas9 (Kao et al. 2017). Recently, the combination of delivering Cas9-sgRNA RNP complexes and a homologous DNA donor template was utilized for site-specific Cas9-based insertional mutagenesis in *Chlamydomonas reinhardtii* (Kim et al. 2020, Picariello et al. 2020). These new developments have the potential to be optimized for a genome-wide RNP-mediated knockout library in the future.

A CRISPR-Cas9 knockout library has yet to be reported in a diatom species. However, unlike *Chlamydomonas reinhardtii*, the plasmid-based expression of Cas9 and sgRNA in diatoms

is straight forward and reproducible (Nymark et al. 2016, Hopes et al. 2016, Slattery et al. 2018, Sharma et al. 2018, Moosburner et al. 2020). In this chapter, the design and deployment of a CRISPR-Cas9 knockout library targeting nitrogen metabolism genes in the diatom *Phaeodactylum tricornutum* is described and results are discussed. To construct the library, the episomal plasmid harboring a selectable Cas9 expression cassette developed in Moosburner et al. (2020) was used. Also, to screen for mutants with elevated lipid content, a fluorescent-protein lipid body reporter was designed and characterized.

## 3.2 Rationale for nitrogen metabolism gene targeting

As outlined in chapter two, nitrogen metabolism has been extensively studied in marine diatoms. Nitrogen utilization drives bottom-up primary productivity in marine food webs where diatoms are major microbial players. Nitrogen depletion also triggers a lipid accumulation stress response that has implications towards the generation of sustainable biofuels from microalgae (Hildebrand et al. 2012, Vinayak et al. 2015). Specifically, nitrate assimilation genes are among the most differentially expressed genes in response to nitrate starvation (Levitan et al. 2015, Alipanah et al. 2015). These genes include nitrate transporters genes, nitrate reductase (NR), nitrite reductase (NiR)*,* and glutamine synthetase 2 (GS2). Furthermore, a set of 61 nitrate sensitive genes, dubbed highly nitrate sensitive (HNS), were characterized and reported in Smith et al. (2019) as the most transcriptionally upregulated on nitrate compared to other nitrogen sources plus as highly upregulated in response to a "spike" of nitrate to the media after 15 minutes. The HNS gene set also contained nitrate assimilation genes including NR and GS2. It was shown in chapter two that a GS2 mutant accumulated lipids on nitrate media without compromising growth. Robust growth during lipid accumulation characteristics are highly

attractive for biofuels applications (Berges et al. 1996, Hildebrand et al. 2012, Ajjawi et al. 2017). The rationale to target both genes that are differentially expressed during nitrate depletion and nitrate sensitive genes (HNS genes) is to find additional *Phaeodactylum* mutants that have growth and lipid accumulation characteristic akin to a GS2 mutant when grown on nitrate media.

The gene targets in the nitrogen gene (N-gene) knockout library consist of HNS genes, nitrate assimilation genes, GS-GOGAT genes, urea cycle genes, putative ornithine-glutamine shunt genes (Levering et al. 2016), putative aspartate-alanine shuttle genes (Smith et al. 2019), and citric-acid cycle (TCA) genes (Table 3.1). The annotation of HNS genes and the cellular processes they are involved in for nitric oxide metabolism, redox, transport, post-transcriptional regulation, post-translational regulation, weak annotations, and hypothetical are the same as the ones reported in Smith et al (2019), Figure 2. There are a total of 136 genes in the N-gene library.

**Table 3.1** N-gene knockout library gene list. For each of the 136 genes, the cellular process it's associated with (colored columns), the gene ID (Phatr3 gene ID's), common name, and functional annotation are given. The HNS genes described in Smith et al. (2019) are highlighted in yellow. For visualization, the table is partition into three sections.

**Table 1**

| Cellular Process | Gene ID | Name | Functional Annotation |
|---|---|---|---|
| Nitrate Assimilation | 54983 | NR | Nitrate reductase |
| | 21961 | NitK | Nitrate and nitrite sensing kinase |
| | 40691 | NRT | Nitrate transporter |
| | 26029 | NIT1 | Nitrate transporter |
| | 12902 | NiR1 | Nitrite reductase (Fd) |
| | 9538 | NiR2 | Nitrite reductase (Fd) |
| | 13076 | NIRT | Nitrite transporter |
| | EG01952 | VacT | Vacuolar transporter |
| | 15625 | | Molybdenum cofactorsynthesis C |
| | 50045 | NirA | Nitrate sensitive transcription factor |
| GS-GOGAT | 51092 | GS2 | Glutamine synthetase 2 (chloroplast) |
| | 22357 | GLNA | Glutamine synthetase 3 (mitochondria) |
| | 24739 | cGOGAT | Glutamate synthase (chloroplast) |
| | 51214 | mGOGAT | Glutamate synthase (mitochondria) |
| | 45239 | GDHB | Glutamate dehydrogenase |
| | 13951 | GDHA | Glutamate dehydrogenase |
| | 50971 | GDHC | Glutamate dehydrogenase |
| Urea cycle | 29702 | URE | Urease |
| | 24195 | unCPS | Carbamoyl phosphate synthase |
| | 39528 | CPS2 | Carbamoyl phosphate synthase |
| | 38509 | ARG | Arginase |
| | 36139 | FUM | Fumarase |
| | 34526 | | Arginosuccinate lyase |
| | 21116 | | Arginosuccinate synthase |
| | 30514 | OTC | Ornithine transcarboxylase |
| | 24238 | | Carbamate kinase |
| | 27877 | AT | Ammonium transporter |
| | 36398 | | Arginine deaminase |
| Ornithine - Glutamine Shunt | 10902 | NAGS1 | N-acetyl glutamate synthase |
| | 44492 | NAGS2 | N-acetyl glutamate synthase |
| | 12762 | GACT | Glutamate acetyltransferase |
| | 3969 | AGK | Acetylglutamate kinase |
| | 36913 | AGPR | Acetyl-glutamyl phosphate reductase |
| | 50577 | ACOAT | 1 Acetylornithine aminotransferase |
| | 36347 | ACOAT | 2 Acetylornithine aminotransferase |
| | 19762 | ACOAT | 3 Acetylornithine aminotransferase |
| | 27726 | OAT | Ornithin aminotransferase |
| | 5175 | OT | Ornithine transporter |
| | 15083 | ACY1 | Aminoacylase metalloexopeptidase |
| | 12762 | GNAT | Glutamate N-acetyltransferase |
| | 9278 | | Glutaminase |
| Aspartate - Alanine Shuttle | 34010 | ANT1 | Alanine aminotransferase(chloro plast) |
| | EG02631 | ANT2 | Alanine aminotransferase (mitochondria) |
| | 23059 | AAT1 | Aspartate aminotransferase (chloroplast) |
| | 23871 | AAT2 | Aspartate aminotransferase (mitochondria) |

**Table 2**

| Cellular Process | Gene ID | Name | Functional Annotation |
|---|---|---|---|
| Citric Acid Cycle | 30519 | | Pyruvate carboxylase |
| | 10356 | | Succinate dehydrogenase |
| | 11021 | | Pyruvate dehydrogenase |
| | 14762 | ICD | Isocitrate dehydrogenase |
| | 29016 | OGD | 2-oxoglutarate dehydrogenase |
| | 19708 | FUM | Fumarase |
| | 16069 | | Dihydrolipyl dehydrogenase |
| | 13894 | | Dihydrolipollysine-residue acetyltransferase |
| | 17401 | | Dihydroliollysine-residue acetyltransferase |
| | 20183 | | Pyruvate dehydrogenase |
| | 26432 | | Dihydrolipoyl dehydrogenase |
| | 26921 | | Succinate-CoA ligase (ADP forming) |
| | 30113 | | Dihydrolipoyl dehydrogenase |
| | 30415 | CS | Citrate synthase |
| | 38009 | | Dihydrolipollysine-residue acetyltransferase |
| | 40430 | | Dihydrolipollysine-residue acetyltransferase |
| | 41812 | | Succinate dehydrogenase (ubiquinone) |
| | 42015 | | Succinate-CoA ligase (GDP forming) |
| | 42398 | | Malate dehydrogenase |
| | 45428 | | Dihydrolipoyl dehydrogenase |
| | 47448 | | Succinate-CoA ligase (GDP forming) |
| | 49253 | | Dihydrolipoyl dehydrogenase |
| | 49339 | | Pyruvate decarboxylase |
| | 51720 | | Succinate dehydrogenase (ubiquinone) |
| | 52539 | | Succinate dehydrogenase (ubiquinone) |
| | 54477 | | ATP citrate synthase |
| | 54834 | | Malate dehydrogenase |
| | 55018 | | Phosphoenolpyruvate carboxykinase (ATP) |
| | 55035 | | Pyruvate dehydrogenase |
| Nitric Oxide Metabolism | 45621 | NOD | Nitric oxide dioxygenase |
| | 37957 | THB | Bacterial-like globin |
| Metabolism | 16615 | | Polyprenyl synthetase |
| | 43490 | | Retinal pigment epithelial membrane protein |
| | 37532 | | Polyketide cyclase/dehydrase and lipid transporter |
| Redox | 45758 | | NAD(PH-cytochrome P450 reductase |
| | 8683 | | Methioninsulfoxide reductase |
| Transport | 50071 | | Transmembrane amino acid transporter |
| | 50408 | | Mitochondria carrier protein |
| | 45965 | | EamA-like transporter family |
| Post - transcriptional regulation | 50624 | | Cleavage and polyadenylation specificity protein |
| | EG01412 | | OCRE/G-patch |
| | 11337 | | Pumilio-family RNA-binding repeat |
| Post - translational regulation | 16036 | | Tyrosine kinase |
| | 48446 | | Histidine acid phosphatase |
| | 31822 | | NmrA-like family |
| | 44704 | | N-terminal asparagine amidohydrolase |
| | 47107 | | Rhomboid protease |

**Table 3**

| Cellular Process | Gene ID | Name | Functional Annotation |
|---|---|---|---|
| Weak annotations | 45283 | | DDE superfamily endonuclease |
| | 44836 | | FYVE zinc finger |
| | 44839 | | FYVE zinc finger |
| | 38623 | | PAP fibrilin |
| | 49215 | | Von willebrand factor type A |
| | 49280 | | RING finger domain |
| | 50373 | | Methyltransferase |
| | 44835 | | Zinc finger, C3HC4 type(RING) |
| | 44838 | | Zinc finger, C3HC4 type(RING) |
| | 7736 | | DUF572 |
| | 55004 | | Glucoamylase |
| | 5651 | | 14-3-3-like protein, tryosine 3-monooxygenase |
| | EG00664 | | Molybdopterin converting factor subunit 1 |
| | 8706 | | Fructose-2,6-biphosphatase |
| | 46237 | | Globin |
| | 40435 | | Ribosomalsubunit 14 |
| | 15229 | | Serine/threonin protein kinase |
| | 12925 | | Uroporphyrin methyltransferase |
| Hypothetical | EG01376 | | Hypothetical protein |
| | 43348 | | Hypothetical protein |
| | 44510 | | Hypothetical protein |
| | 43503 | | Hypothetical protein |
| | 47780 | | Hypothetical protein |
| | 39656 | | Hypothetical protein |
| | 49287 | | Hypothetical protein |
| | 49236 | | Hypothetical protein |
| | 45632 | | Hypothetical protein |
| | 42546 | | Hypothetical protein |
| | 43822 | | Hypothetical protein |
| | EG02484 | | Hypothetical protein |
| | EG01953 | | Hypothetical protein |
| | EG02286 | | Hypothetical protein |
| | EG02531 | | Hypothetical protein |
| | EG00939 | | Hypothetical protein |
| | 46671 | | Hypothetical protein |
| | 39769 | | Hypothetical protein |
| | 46511 | | Hypothetical protein |
| | 48137 | | Hypothetical protein |
| | EG01504 | | Hypothetical protein |
| | 36843 | | Hypothetical protein |
| | 47892 | | Hypothetical protein |
| | 49238 | | Hypothetical protein |
| | 47900 | | Hypothetical protein |
| | 50103 | | Hypothetical protein |

**3.3 Development of fluorescent protein lipid body reporter**

A fluorescent protein (FP) lipid reporter system was developed in order to screen a culture of pooled *Phaeodactylum* mutants for high lipid producing cells in a high-throughput manner. FP expression is a commonly used practice in cell biology to characterize subcellular protein localization and gene promoter activity. In diatoms and other oleaginous microalgae, the histochemical cell stains BODIPY and Nile Red are regularly used for the localization to and fluorescence of lipid bodies (Rumin et al. 2015). They are useful indicators and microscopy stains for cells that accumulate lipids however cell staining methods require time and labor. A lipid-body FP reporter was previously developed in *Phaeodactylum* cells that used a eGFP fused to a lipid body protein native to a species of *Haematococcus* (Shemesh et al. 2016). The heterologous expression of the protein (HOGP) conferred localization to oil droplets in *Phaeodactylum* during nitrogen starvation when expression of eGFP-HOGP was driven by the native *Phaeodactylum* promoter of the diacylglycerol acyltransferase 1 (DGAT1) gene. DGAT condenses acyl-CoA and diacylglycerol to produce triglycerides (TAG) and is transcriptionally upregulated during nitrogen depletion conditions (Guiheneuf et al. 2011, Shemesh et al. 2016).

Though the eGFP-HOGP reporter would have been useful here in implementing the N-gene knockout library and screening for lipid producers, the genetic construct was never obtained due to proprietary licensing. Nevertheless, a new TAG reporter was developed here using similar expression and localization principles demonstrated in Shemesh et al. (2016).

A lipid droplet protein, named the stramenopile lipid droplet protein (StLDP) was previously characterized as the most abundant protein associated with lipid droplets in *Phaeodactylum* (Yoneda et al. 2016). StLDP was shown to be associated with lipid droplets in a lipid fraction extracted from a nitrogen starved *Phaeodactylum* culture and contains both

hydrophobic and hydrophilic domain structures in its peptide sequences, indicating that it may be bound to the surface of lipid bodies. A follow-up study functionally characterized StLDP by the heterologous expression of eGFP-StLDP and using the promoter, FcpA, that constitutively expresses its gene product (Yoneda et al. 2018). It was speculated that the overexpression of StLDP would enhance lipid body formation, however it was only reported that eGFP-StLDP was associated with lipid bodies and could not increase the TAG content in transgenic *Phaeodactylum* cell lines (Yoneda et al. 2018).

For this study, StLDP was used as a TAG reporter since (1) eGFP-StLDP was shown to localize to lipid bodies and (2) the heterologous expression of StLDP did not alter the TAG content of the cells. The expression of StLDP was also driven by the DGAT1 promoter, as done in Shemesh et al. 2016 to express eGFP-HOGP, in order to synchronize TAG biosynthesis with the expression of the reporter. Furthermore, an mCherry FP was fused to the N-terminal of StLDP. mCherry was chosen as a FP reporter instead of eGPF due to the excitation and emission spectrum of mCherry in relation to the auto-fluorescence (AF) of photosynthetic chlorophyll in *Phaeodactylum* cells. Both eGFP and chlorophyll AF are strongly excited by a blue-spectrum laser (488nm laser). A yellow/green laser (561nm laser) strongly excites mCherry fluorescence while producing a dampened chlorophyll AF signal relative to a blue laser (Takahashi 2019). Also, applying optical filters to distinguish chlorophyll AF from mCherry emission in microscopy and flow cytometry applications can be accomplished since an mCherry signal with emissions spectra peaking at 610nm can be distinguished from chlorophyll AF using specific band-pass filters since chlorophyll AF has maximum absorption emission at 650nm and higher.

An episomal plasmid was cloned with the expression cassette for mCherry-StLDP. The DGAT1 promoter was used to drive expression of mCherry-StLDP fusion genes in order to express the mCherry reporter only during TAG formation. Transgenic *Phaeodactylum* cell lines that contained mCherry-StLDP were subjected to nitrogen starvation in order to induce mCherry-StLDP expression (Figure 3.1). mCherry fluorescence was detected by confocal microscopy. First, the cells were grown in nitrogen replete media (artificial seawater supplemented with nitrate) for 72 hours. The cells were then imaged and after 72 hours no mCherry was detected (Figure 3.1a). The cells were then pelleted and resuspended in nitrogen-free (N-free) media. Little to none of the cells after 24 hours and 48 hours of N-free growth produced detectable mCherry fluorescence. After 72 hours of nitrogen starvation, mCherry fluorescence was detected in the majority of cells (Figure 3.1b). It was clear that the cells also started to accumulate lipid droplets. mCherry fluorescence continued to be detectable through 96 hours (Figure 3.1c) and 120 hours (Figure 3.1d) of N-free conditions.
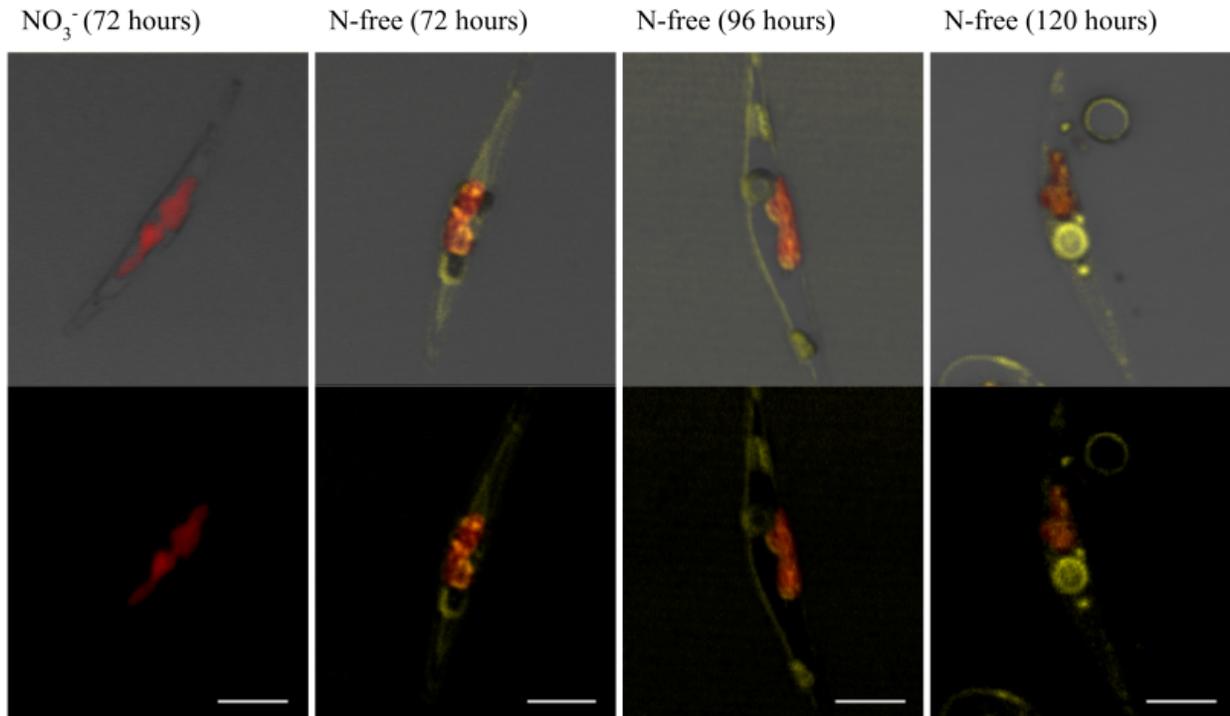
**Figure 3.1** Transgenic wild-type *Phaeodactylum* cell line expressing mCherry-StLDP during nitrogen starvation. Cells were imaged after **(a)** 72 hours of growth in nitrogen replete conditions and during nitrogen-free conditions **(b)** after 72 hours , **(c)** 96 hours, and **(d)** 120 hours. In all images, the red signal is from the chlorophyll autofluorescence and the yellow is from mCherry-StLDP fluorescence. The top panels are merged images of fluorescence and brightfield and the bottom panels are images of fluorescence without brightfield. Scale bars are 7.5µm. Images were produced using a confocal microscope wielding a blue-spectrum laser at 514nm.

The localization of mCherry-StLDP was also apparent from the microscopy images. mCherry-StLDP was localized both to the outer membrane of lipid droplets and to the plasma membrane of the cell. Lipid droplet localization was visualized by a fluorescent ring surrounding the lipid droplets inside and outside of the cell (Figure 3.1d). Localization to the plasma membrane was most obvious during early stages of nitrogen starvation (Figure 3.1b,c). In general, lipid accumulation increases over time during nitrogen starvation conditions for *Phaeodactylum* cells (McCarthy et al. 2017). Perhaps here, cells accumulated less lipids during early nitrogen starvation conditions (1-2 days) compared to later conditions (3 or more days) and mCherry-StLPD preferentially localized to the plasma membrane. However, discrimiation

between the plasma membrane and lipid droplet localization of mCherry-StLDP during certain conditions could not be discernible for certain. Yoneda et al. (2016) reported that StLDP was the most abundant protein in the lipid fraction of nitrogen starved cells. The lipid fraction was isolated from intact cells and the protein analysis did not account for lipids retained within cells. Therefore, StLDP may be the most abundant lipid droplet protein amongst extracellular lipids and not necessarily of all lipid droplets in a *Phaeodactylum* culture. This may give reason to why mCherry-StLDP localized to the plasma membrane because StLDP may localize to both intracellular lipids and lipid droplets that are secreted into the media. Localization to extracellular lipid droplets was visualized after 120 hours of nitrogen starvation conditions (Figure 3.1d). Nevertheless, Yoneda et al. (2016) also found that StLDP contained both hydrophobic and hydrophilic protein domains and therefore could potentially be localized to the plasma membrane where a hydrophobic region exists between the bilayer of phospholipids. It also appears that StLDP was localized to the plasma membrane and lipid droplets when tagged with eGFP and expressed by the FcpA promoter (Yoneda et al. 2018).

## 3.4 Validation of lipid body reporter in a nitrate reductase knockout cell line

The CRISPR-Cas9 episomal cloning vector (pBRCas9_gLacZ) from Moosburner et al. (2020) was used to clone the expression cassette for mCherry-StLDP. The vector contains a LacZ for cloning selection flanked by two BsaI restriction enzymes for GG cloning of sgRNA expression cassettes (Moosburner et al. 2020). The mCherry-StLDP expression cassette, with the *Phaeodactylum* DGAT1 promoter (pDGAT1), was cloning downstream of LacZ using GA cloning to produce a new sgRNA cloning vector, pBRCas9_StLDP.

To clone sgRNA expression cassettes, GG assembly cloning was used for a three-piece assembly of the pBRCas9_StLDP vector, a sgRNA expression cassette, and a kanamycin resistance gene cassette. Hence, 100% of colonies that survived cloning when supplemented with kanamycin also contained the sgRNA expression cassette (Figure 3.2)
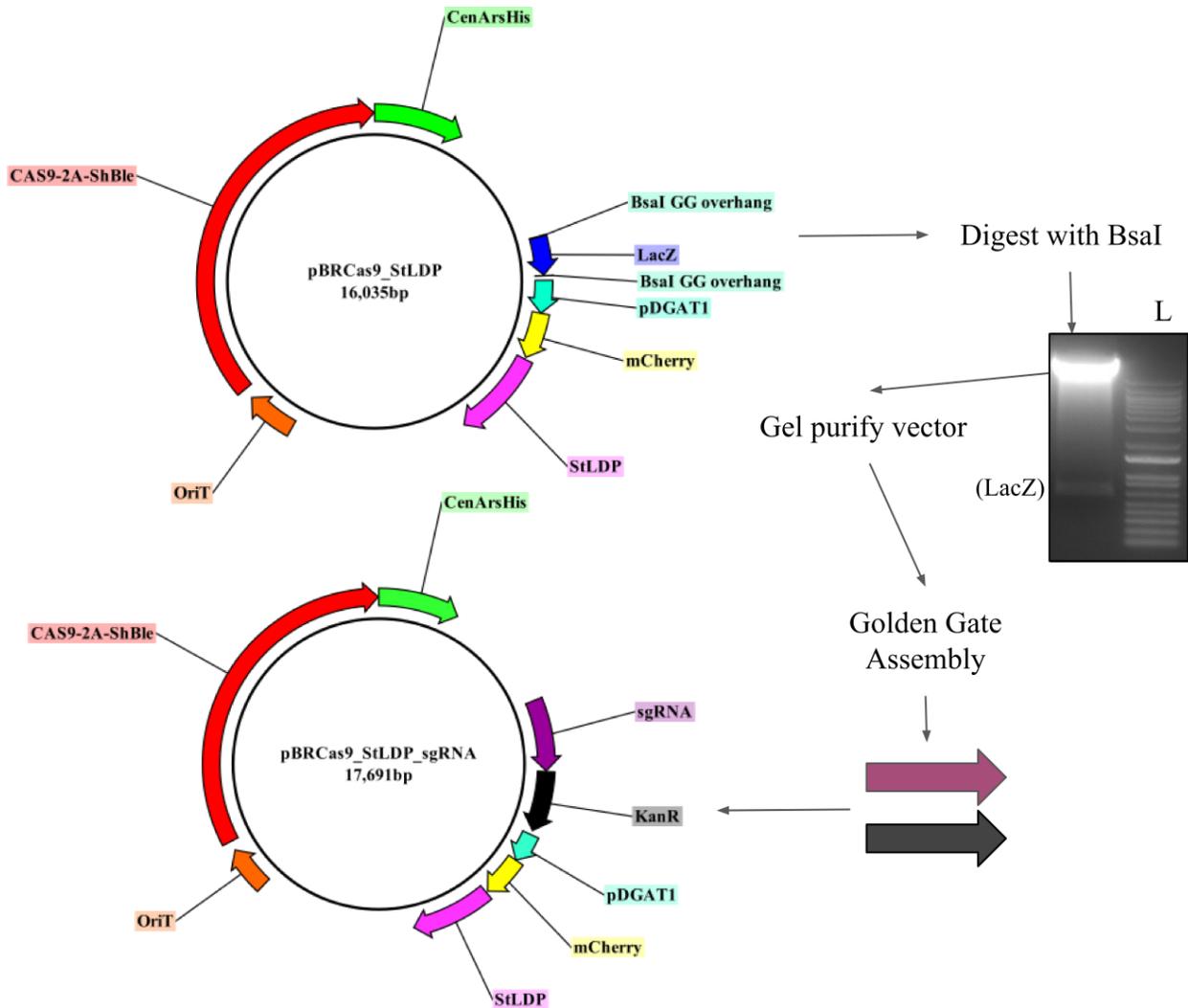


**Figure 3.2** Episome plasmid maps for the cloning vector pBRCas9_StLDP and the cloned plasmid after GG assembly, pBRCas9_StLDP_sgRNA. The process for GG cloning is shown where the vector is pre-digested with BsaI-HFv2 restriction enzyme and the larger band (~15kb) is gel purified. The excised LacZ band can also be seen in the gel. A three-piece assembly occurs with the purified vector, a kanamycin resistant gene (black), and a sgRNA expression cassette (purple). The cloning vector and constructed episomal plasmid both contain a Cas9-ShBle expression cassette (red), CenArsHis (green), origin of transfer (orange), and a mCherry (yellow) - StLDP (pink) expression cassette where pDGAT1 (light blue) is the promoter.

A proof-of-concept experiment was designed to test the mCherry-StLDP lipid body detection system in a mutant *Phaeodactylum* cell line deficient of NR. Much has been previously reported about NR knockout cell lines (McCarthy et al. 2017, Moosburner et al. 2020). Most notably that Pt_NR-KO cell lines die when supplemented with nitrate and that their inability to assimilate nitrate leads to lipid biogenesis. A sgRNA expression cassette for NR (gNR-B, from Moosburner et al. (2020)) was cloned into pBRCas9_StLDP and the resulting episome was named pBRCas9_StLDP_gNR-B.

Conjugation of pBRCas9_StLDP_gNR-B resulted in hundreds of transformed *Phaeodactylum* colonies. 16 colonies were screened for mutations at the gNR-B target site by colony PCR. Three candidate cell lines were then further analyzed for mutations at the target locus by Sanger sequencing and for the expected nitrate growth phenotype of cell death. All three candidate cell lines exhibited NHEJ-mediated mutagenesis at the gNR-B target site and also could not grow on nitrate media compared to growth on ammonium media.

Pt_NR-KO cell lines were then subjected to a nitrate growth assay (growth on ammonium media and nitrate media) and imaged daily by confocal microscopy for mCherry-StLDP localization (Figure 3.3). First, the Pt_NR-KO cell lines could not grow on nitrate media compared to ammonium media (Figure 3.3a). In ammonium growth conditions, mCherry fluorescence was not detected until 120 hours when cells were entering the stationary phase of growth and depleted the media of ammonium (Figure 3.3b). In nitrate growth conditions, mCherry fluorescence was detected after 24 hours and throughout the experimental time-frame (Figure 3.3c). This finding was expected since lipids were elevated in a Pt_NR-KO cell line compared to wild-type after 24 to 36 hours in nitrate growth conditions (McCarthy et al. 2017). This experiment also demonstrated that pDGAT1 can increase expression of

mCherry-StLDP in a Pt_NR-KO cell line when nitrate is readily available to the cell. Therefore, pDGAT1 gene expression is not exclusive to conditions of nitrogen starvation but actively up-regulates its gene product during nitrogen stress, too.
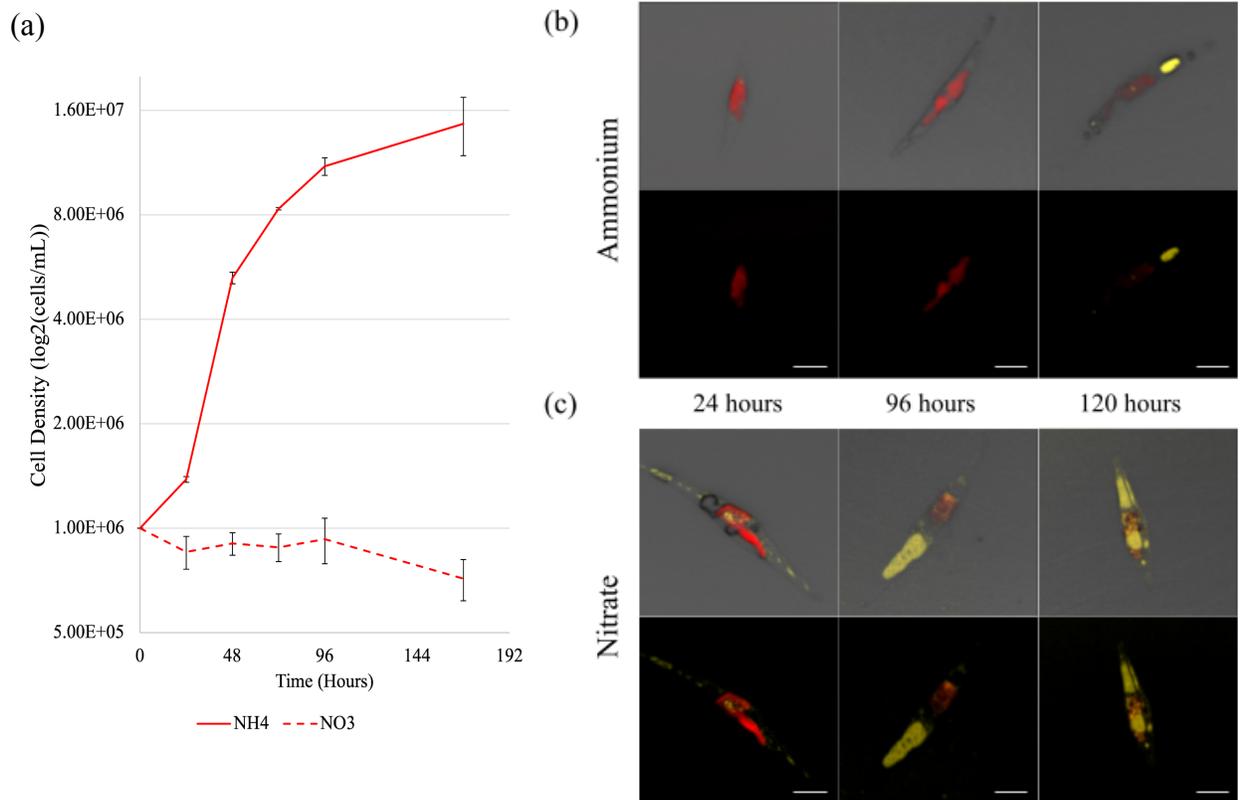


**Figure 3.3** Transgenic nitrate reductase mutant cell line expressing mCherry-StLDP during nitrate growth assay. **(a)** Growth profile of a *Phaeodactylum* NR-KO cell line on ammonium (solid red line) and on nitrate (dotted red line) media. Confocal microscopy images NR-KO on **(b)** ammonium media and **(c)** nitrate media after 24 hours, 96 hours, and 120 hours of growth. For both sets of images, the top panel shows the merged image of brightfield and fluorescence and the top panel shows only fluorescence. Red fluorescence is from the chlorophyll autofluorescence and yellow fluorescence is from mCherry-StLDP. Scale bars are 7.5μm. Images were produced using a confocal microscope wielding a blue-spectrum laser at 514nm.

mCherry fluorescence was quantified for the population of Pt_NR-KO cell lines in ammonium and nitrate condition by flow cytometry (Figure 3.4a). Again, the Pt_NR-KO cell line was subjected to nitrate growth assay and was grown in ammonium media and nitrate media. Pt_NR-KO grew normally on ammonium media and could not grow on nitrate media. Flow

cytometry was applied to detect mCherry fluorescence using a yellow/green laser (561nm) and chlorophyll AF using a blue laser (488nm). After 8 days, nitrate growth conditions were expected to increase mCherry fluorescence in the Pt_NR-KO cell line population-wide, as was observed by microscopy-based methods (Figure 3.3). Two gates in the flow cytometry dot plots were used to count the population of cells that contain an elevated mCherry fluorescence signal. The "P1" gate was used to count cells that displayed very high mCherry fluorescence and the "mCherry-StLDP" gate used to count the cells that display a detectable mCherry signal but not as high as "P1". Therefore, the cells that fell within "P1" were considered to contain a large amount of lipids and the cells that fell within "mCherry-StLDP" were considered to contain lipids, generally. All cells that fell within the "P1" also fell within the "mCherry-StLDP" gate. The flow  cytometry workflow and sample dot plot generated from flow cytometry data shows the "P1" (orange) and "mCherry-StLDP" (pink) populations (Figure 3.4b).

**Figure 3.4** Flow cytometry detection of mCherry-StLDP fluorescence in a nitrate reductase mutant cell line. **(a)** Schematic of the flow cytometry workflow to detect mCherry-StLDP fluorescence. The expression vector, pDGAT1_mCherry-StLDP, should only be expressed during nitrate limited conditions. NR-KO cell lines were first culture on ammonium media then passed into ammonium and nitrate conditions, separately. mCherry fluorescence excited by a green-spectrum laser daily over 8 days of growth. **(b)** Representative dot plots from flow cytometry runs for ammonium (NH4) and nitrate (NO3) conditions. The top two plots show the "P1" gate. The bottom two plots display events colored pink that are from the "mCherry-StLDP" gate. The other dot colors represent the following: blue = alive *Phaeodactylum* cells, red = all mCherry *Phaeodactylum* cells, green = medium mCherry *Phaeodactylum* cells. **(c)** The percentage of cells over time in a culture (a) within the "P1" gate that represents cells with a very high level of mCherry signal and **(d)** within the "mCherry-StLDP" gate that represents the cells that contain a high mCherry, (Figure 3.4 caption, Continued) respectively. For both, the solid line represents cells cultured in ammonium media (NH4) and the dotted line represents cells cultured in nitrate media (NO3). Two biological replicates and at least 10,000 events were measured for each time point. **(e)** Ratio of nitrate (NO3) to ammonium (NH4) populations for "P1" gated cells (orange) and for "mCherry-StLDP" gated cells.
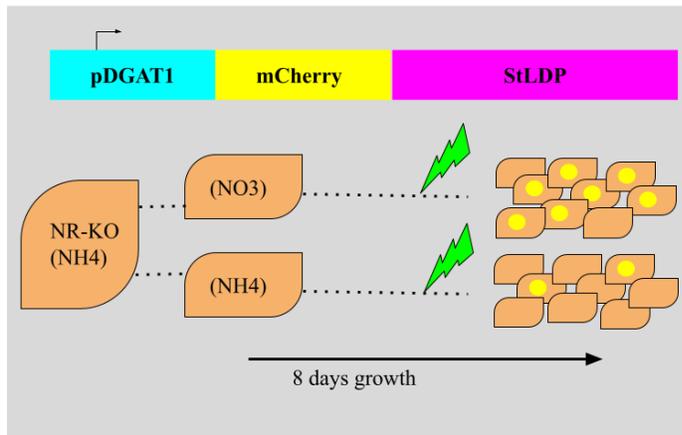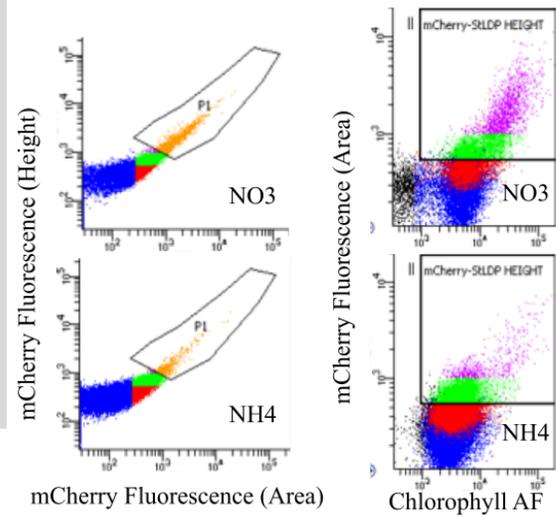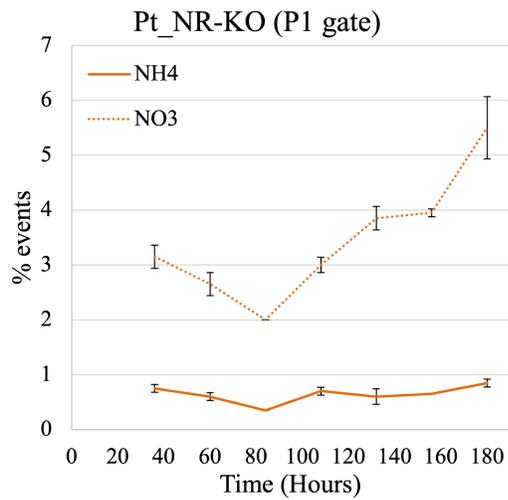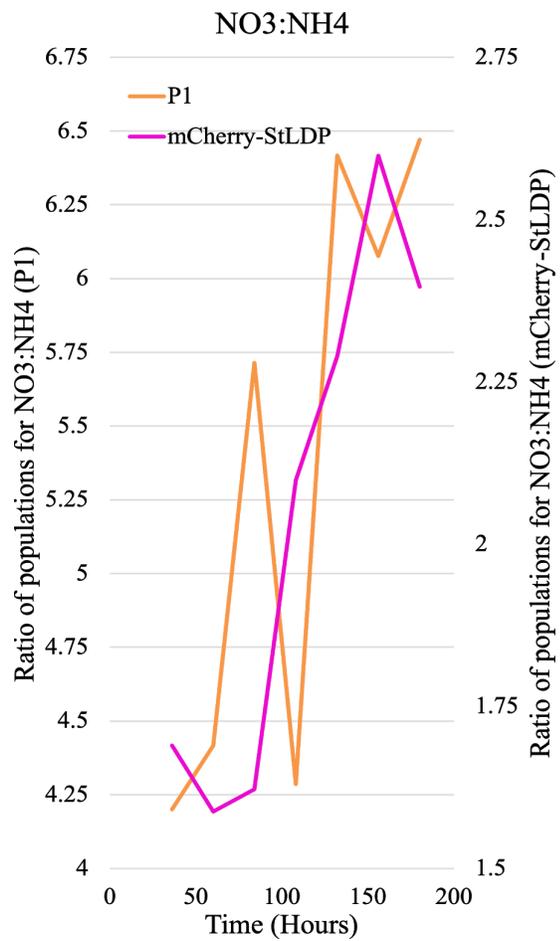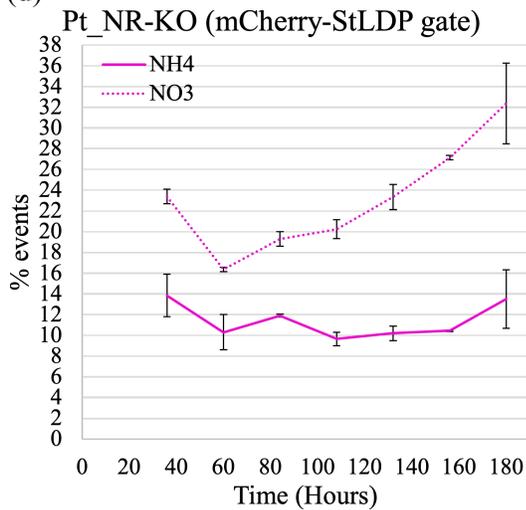
Overall, Pt_NR-KO cell lines exhibited a higher proportion of mCherry reporter signal population-wide when cultured on nitrate media (Figure 3.4). For the "P1" populations, the percentage of Pt_NR-KO cells for nitrate-growth increased incrementally from 2.0% at 84 hours to 3.0% (108 hours), 3.85% (132 hours), 3.95% (156 hours), and finally to 5.5% at 180 hours (Figure 3.4c). The percentage of "P1" cells did not substantially increase throughout the experiment for Pt_NR-KO cells cultured on ammonium. At most, the Pt_NR-KO cells contained 5.5-fold more cells in the "P1" population on nitrate compared to ammonium at 180 hours. At the least, Pt_NR-KO cells on nitrate contained only 1.2-fold more cells in "P1" than ammonium at 36 hours.

For the "mCherry-StLDP" gated populations, the percentage of Pt_NR-KO cells on nitrate also increased incrementally from 16.4% at 60 hours to 19.3% (84 hours), 20.3% (108 hours), 23.4% (132 hours), 27.2% (156 horus), and finally to 32.4% at 180 hours (Figure 3.4d). Compared to nitrate growth, ammonium growth conditions did not substantially increase the percentage of Pt_NR-KO cells in the "mCherry-StLDP" population throughout the experiment. The ratio of the "mCherry-StLDP' population, like in the "P1" population, increased for nitrate cells. At the most, Pt_NR-KO cells on nitrate contained 2.6-fold more cells in "mCherry-StLDP" than on ammonium at 156 hours. At the least, Pt_NR-KO cells contained 1.6-fold more cells on nitrate compared to ammonium at 60 hours.

In general, Pt_NR-KO cells contained a greater percentage of cells in both "P1" and "mCherry-StLDP" populations in nitrate growth conditions compared to ammonium growth conditions. This was expected since nitrate conditions induce nitrogen stress in Pt_NR-KO cell lines which causes lipid accumulation to increase. Though an incremental increase was observed in both "P1" and "mCherry-StLDP" gate populations, an increase was not observed throughout

the entire growth period. At 36 hours, the percentage of cells in nitrate were higher than at 84 hours and at 60 hours for "P1" and "mCherry-StLDP", respectively. The same was observed for ammonium conditions.

Further inspection of the population data shows that regardless of the percentage of populations for each condition, the ratio of nitrate to ammonium cell populations did increase throughout the growth period (Figure 3.4e). The ratios of nitrate to ammonium "P1" populations increased over time but were sporadic during the first 108 hours of growth. Conversely, the ratio of nitrate to ammonium "mCherry-StLDP" populations steadily increased over time. It is well known that Pt_NR-KO cell lines increase their lipid content over time when cultured on nitrate media (McCarthy et al. 2017) and thus the ratio of lipids in nitrate conditions compared to ammonium conditions should also increase over time. The nitrate to ammonium ratio for "mCherry-StLDP" appears to reflect the expected lipid accumulation effect for Pt_NR-KO cell lines. Therefore, compared to "P1", the "mCherry-StLDP" gate seems to be a better indicator for lipid accumulation within a *Phaeodactylum* culture.

## 3.5 Examining the lipid accumulation phenotype in a glutamine synthetase 2 mutant

*Phaeodactylum* mutants deficient in glutamine synthetase 2 (GS2) were produced using similar methods outlined previously for the generation of Pt_NR-KO cell lines. Pt_GS2-KO and a wild-type (Pt-WT) cell line used in the following experiments also contained the mCherry-StLDP lipid body reporter. The detection of the reporter fluorescence was conducted using flow cytometry methods as discussed previously.

Detection of mCherry-StLDP fluorescence in Pt_GS2-KO and Pt_WT cell lines was performed in three growth conditions. The cell lines were first pre-cultured in nitrate or

ammonium conditions. For the detection experiment, the cell lines were either passed from ammonium media to ammonium media (NH4 to NH4), from ammonium media to nitrate media (NH4 to NO3), or from nitrate media to nitrate media (NO3 to NO3) (Figure 3.5). For the NH4 to NH4 and NH4 to NO3 conditions, the Pt_NR-KO cell line with mCherry-StLDP reporter was included for comparison. Cell growth was not decreased in Pt_GS2-KO on any growth conditions compared to Pt_WT (Figure 3.5a). In fact, Pt_GS2-KO cell lines appeared to grow faster and at a higher density only for NH4 to NO3 conditions.

**Figure 3.5** Comparison of mCherry-StLDP reporter detection between NR-KO, GS2-KO, and wild-type *Phaeodactylum* cell lines in three culture conditions. For all plots, NR-KO cell lines are colored red, GS2-KO cell lines are colored green, and wild-type cell lines are colored purple. Also, the conditions are represented in all plots by the type of line plotted where. For cultures pre-conditioned on ammonium and then cultured on ammonium (NH4 to NH4) the lines are solid. For cultures pre-conditioned on ammonium and then cultured on nitrate (NH4 to NO3) the lines are dotted. For cultures pre-conditioned on nitrate and then cultured on nitrate (NO3 to NO3) the lines are dashed. **(a)** The cell density growth profiles comparing the three cell lines for the culture conditions NH4 to NH4 (left), NH4 to NO3 (middle) and NO3 to NO3 (right). **(b)** mCherry-StLDP reporter detection by flow cytometry for NH4 to NH4 (left), NH4 to NO3 (middle), and NO3 to NO3 (right) conditions and comparisons between cell lines. Flow cytometry events were gathered from the "mCherry-StLDP" gate. Two biological replicates and at least 10,000 events were measured for each time point.
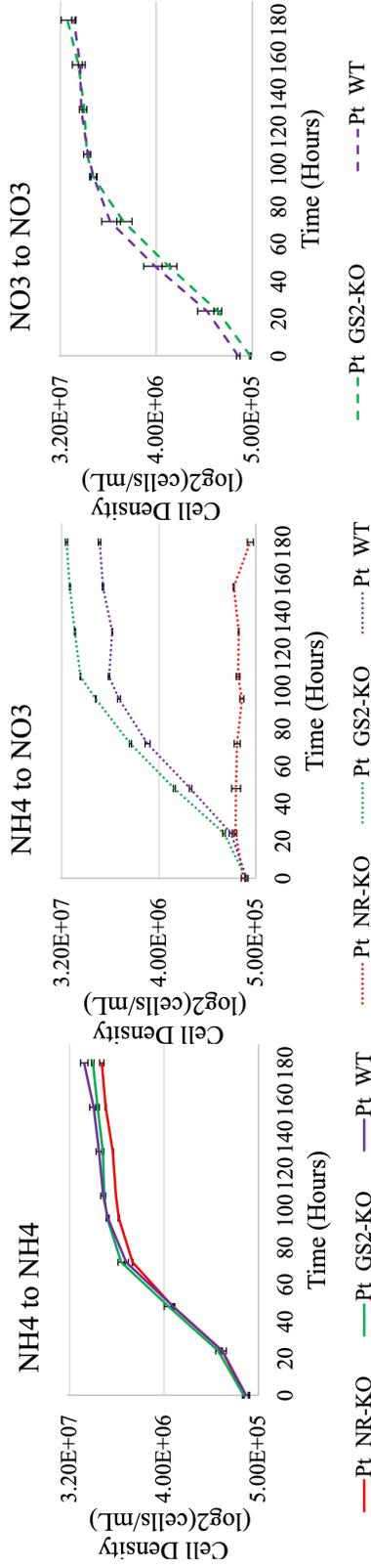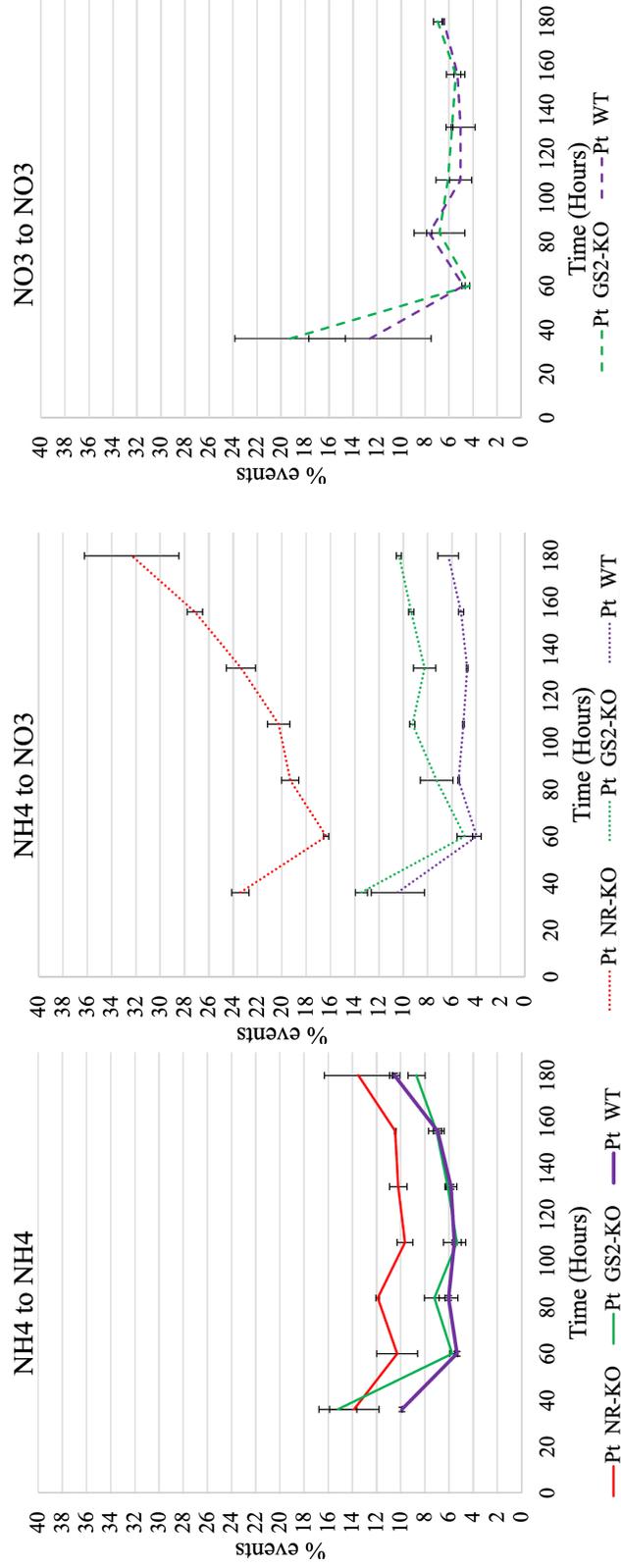
mCherry-StLDP fluorescence detection via flow cytometry was performed daily. The "mCherry-StLDP" gate was used to report the percentage of the cell population that contained detectable lipids by mCherry fluorescence (Figure 3.5b). In the NH4 to NH4 conditions, Pt_GS2-KO and Pt_WT cell lines contained comparable populations of "mCherry-StLDP" throughout growth. Similar findings were observed for the NO3 to NO3 condition. Differences in "mCherry-StLDP" populations were only observed in NH4 to NO3 conditions. After 84 hours, the percentage of cells increased in Pt_GS2-KO cell lines compared to Pt_WT. This increase was sustained throughout the remainder of growth. Though notable compared to Pt_WT, the percentage of "mCherry-StLDP" cells in Pt_GS2-KO did not reach the level reached for Pt_NR-KO. Nonetheless, Pt_GS2-KO cell lines can grow on nitrate media while Pt_NR-KO cannot. Pt_GS2-KO cell lines therefore appear to increase their lipid content compared to Pt_WT without experiencing nitrogen stress and compromising growth.

The increased lipid content in Pt_GS2-KO cell lines was only apparent when cells were switched from ammonium growth conditions to nitrate growth conditions. This observation reflects what was reported in a *Phaeodactylum* GS2 mutant in chapter two. The percentages of "mCherry-StLDP" populations were compared between two growth conditions, NH4 to NH4 compared to NH4 to NO3, for the three cell lines (Figure 3.6). The ratio between the two conditions (NH4:NO3) was calculated for all cell lines (Figure 3.6a). As observed before, the NO3:NH4 ratio increased overtime for Pt_NR-KO. In Pt_GS2-KO cell lines, the NO3:NH4 ratio peaked at 108 hours and remained above a ratio of 1 for the rest of the growth period. Interestingly, the NO3:NH4 ratio in Pt_WT cell lines was below 1 for most of the growth period which suggests that ammonium growth conditions cause a slight increase in lipid accumulation compared to that in nitrate growth conditions for wild-type *Phaeodactylum*. Nonetheless, the

conditional switch from ammonium growth to nitrate growth appears to "trigger" lipid accumulation in Pt_GS2_KO compared to Pt_WT.
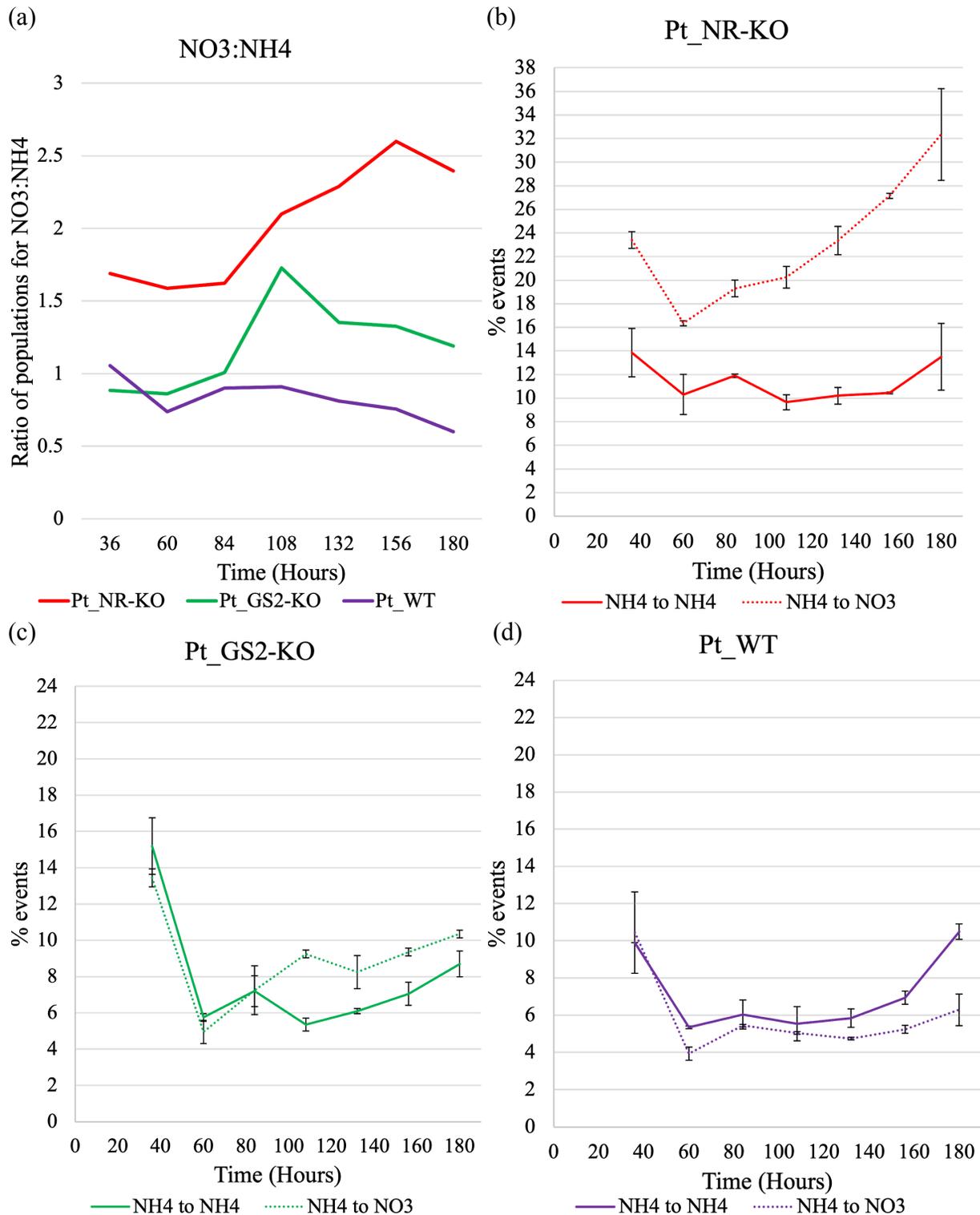
**Figure 3.6** The comparison of lipid producing populations between nitrate conditions and ammonium conditions for Pt_NR-KO, Pt_GS2-KO, and Pt_WT cell lines. **(a)** The ratio of nitrate to ammonium populations. The percentage of cells plotted for both conditions (NH4 to NH4, solid lines; NH4 to NO3, dotted lines) for **(b)** Pt_NR-KO cell lines, **(c)** Pt_GS2-KO cell lines, and **(d)** Pt_WT cell lines. Flow cytometry events were gathered from the "mCherry-StLDP" gate. Two biological replicates and at least 10,000 events were measured for each time point.

The comparison between the two conditions was also reported for Pt_NR-KO (Figure 3.6b), Pt_GS2-KO (Figure 3.6c), and Pt_WT (Figure 3.6d). For Pt_GS2-KO, the 36 hours time point appears to have an elevated percentage of "mCherry-StLDP" compared to Pt_WT for both conditions. However, at 36 hours the cells are beginning to exit the lag phase of growth and start to exponentially grow. During this growth phase, *Phaeodactylum* cells are not expected to produce intercellular lipids because they are not limited for nitrogen or any other nutrients. Though, McCarthy et al. (2017) reported that Pt_WT and Pt_NR-KO both began to increase their lipid content at 18 hours through 60 hours of growth on nitrate. Lipids may not accumulate in wild-type *Phaeodactylum* until late-exponential and stationary phases of growth but lipid biogenesis may begin earlier during exponential growth (McCarthy et al. 2017). This may explain why a greater percentage of cells were detected to contain mCherry fluorescence during early exponential phases. The pDGAT1 promoter drives the expression of mCherry-StLDP and is an active promoter during lipid biosynthesis. Also, mCherry-StLDP appeared to localize to lipid bodies and to the plasma membrane in *Phaeodactylum* (Figure 3.1). Thus, even without lipids accumulating, mCherry-StLDP may be localizing to the plasma membrane of cells during lag and exponential growth. In lag phase, cells do not divide as fast as in exponential growth. *Phaeodactylum*, at its fastest, doubles once per 24 hours. In cells that divide slower, mCherry-StLDP protein may be accumulating at greater abundances than in faster dividing cells when nutrients are not limiting. If so, mCherry fluorescence would be detectable in a higher percentage of cells during lag phase than during exponential phase.

## 3.6 Design and synthesis of nitrogen gene knockout library episomes

The mCherry-StLDP reporter, validated for high-throughput lipid detection via flow cytometry, will be a critical component in the design and implementation of a CRISPR-Cas9 knockout library. The lipid reporter will facilitate in finding mutant cells with elevated lipid content within a culture of pooled mutants. To generate a pool of mutants for all genes outlined in the N-gene knockout library (Table 1), a high-throughput batch cloning scheme was implemented to clone multiple sgRNA target loci simultaneously (Figure 3.7).
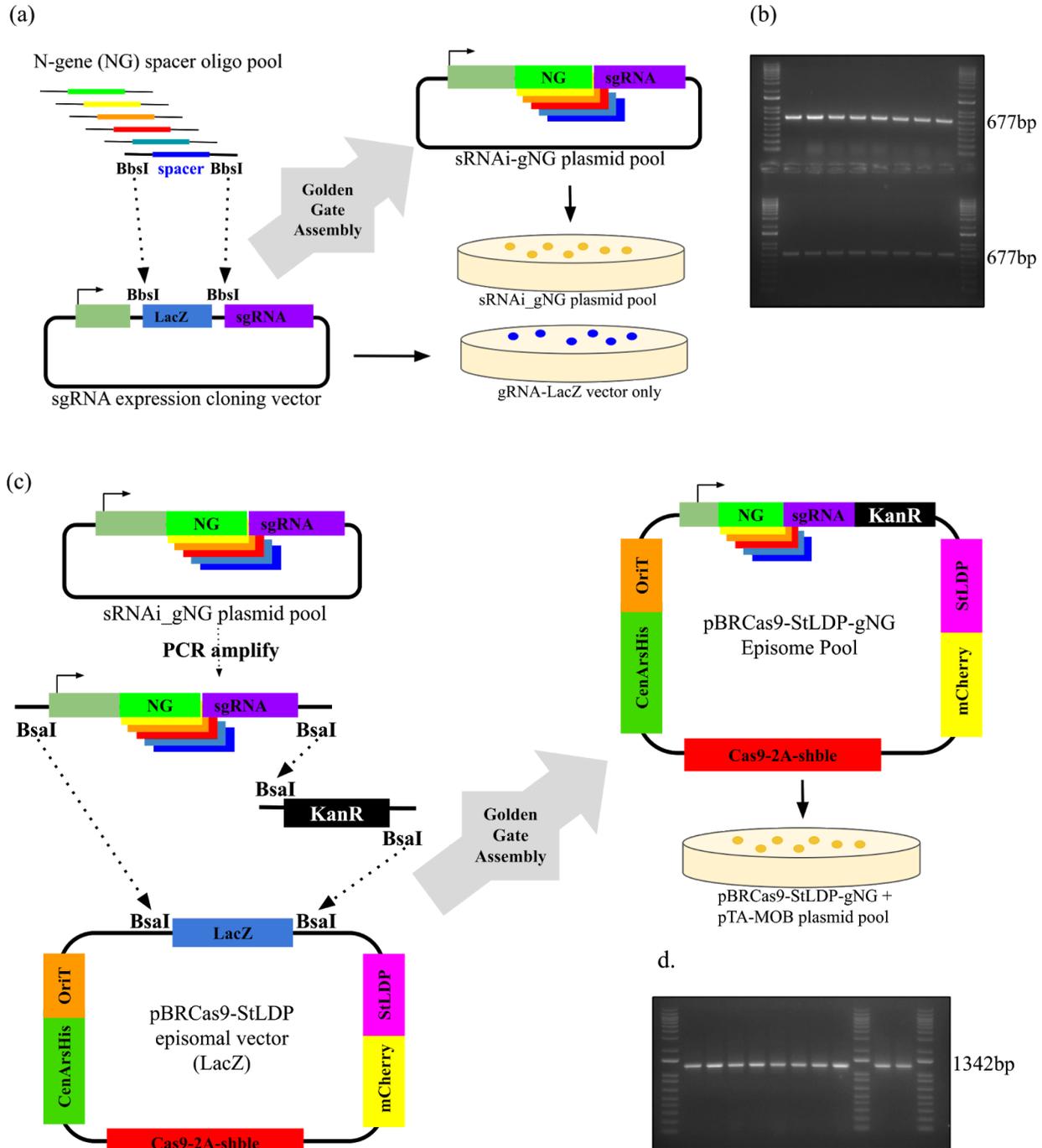
**Figure 3.7** Batch cloning scheme of N-gene (NG) spacer oligo library into a Cas9-StLDP episomal vector. First, the oligo library was cloned into a TOPO vector and the spacer and BbsI sites were amplified and pooled. **(a)** The N-gene spacer pool was the batch cloned into an sgRNA expression cloning plasmid that contained a LacZ cassette for negative selection (blue colonies). Correctly cloned plasmids produced a non-colored *E. coli* colony. **(b)** Randomly picked colonies were PCR amplified and run on a gel to ensure correct insert size. **(c)** The pooled sgRNA expression cassette (sRNAi_gNG) were PCR amplified to add BsaI restriction sites for Golden Gate cloning into the pBRCas9-StLDP episomal vector. The pooled sgRNA expression amplicons were cloned with a kanamycin resistance cassette (KanR) to replace the LacZ stuffer sequence within the episome. The episomal pool was transformed into cells that contained pTA-MOB. Correctly cloned episomes produced E. coli resistant to kanamycin and gentamicin. **(d)** Randomly picked colonies were PCR amplified and run on a gel to ensure correct insert size.

### 3.6.1 Spacer oligo pool design and cloning

First, a library of sgRNA target loci was produced. Plasmid-based CRISPR-Cas9 knockout libraries have been reported to contain at least three sgRNAs per gene to sufficiently to mutate all genes within a library but was also reported that six sgRNAs per gene are recommended for optimal targeting (Shemesh et al. 2014, Hart et al. 2017). Hence, six sgRNAs were designed per gene for a total of 816 unique sgRNA targets for all 136 genes in the N-gene library. Homing Cas9 to its target locus only requires changing the 20nt spacer sequence at the 5' end of the 102nt sgRNA. Here, a 56nt oligo insert was designed for Gibson assembly cloning into a sgRNA expression vector (Moosburner et al. 2020). Each oligo contained the same Gibson assembly overhangs at the 5' and 3' ends but differed within the 20nt spacer region. The 816 unique oligo sequences were generated for each target gene using the CRISP-Ex sgRNA target loci software specific for *Phaeodactylum* and other algal species (Rastogi et al. 2016). The oligo library was synthesized by Twist Biosciences (San Francisco, CA, USA) and the pooled oligo library was delivered in a single tube.

To efficiently clone all 816 oligos from a Twist Biosciences oligo pool, a cloning scheme was developed akin to the one reported previously for an spacer oligo pool for a CRISPR-Cas9 knockout library (Becker et al. 2020). The N-gene oligo pool was PCR amplified using 15X cycles to limit amplification bias. The oligo amplicon was then TOPO cloned using the Zero Blunt™ TOPO™ PCR Cloning Kit (Invitrogen). In summary, 2μL of PCR product was mixed with 2μL salt solution, 2μL water, and 1μL TOPO vector. One TOPO cloning reaction is sufficient to clone up to 5000 oligos per microliter thus one TOPO reaction was performed to clone the 816 N-gene oligos. All of the cloned TOPO plasmids were then precipitated by an isopropanol-based DNA precipitation method and resuspended in 2μL water. The 2uL of TOPO

plasmid was transformed by heat-shock into One Shot™ MAX Efficiency™ DH5α-T1R

Competent Cells (Invitrogen). All of the transformed cells were plated onto 245mm square plates

with LB agar supplemented with 50μg/mL kanamycin and grown overnight at 37°C. Next day,

thousands of colonies were scraped off the plates and the TOPO plasmids were extracted and

isolated using the NucleoBond Xtra Maxi Kit (MachereyNagel). Colony PCR was also

performed on 16 colonies to check for the correct insert size and all 16 were correct.

### 3.6.2 Pooled spacer cloning into sgRNA expression cassette

Next, the TOPO oligo pool was PCR amplified using only 25X cycles to prevent

amplification bias. The PCR product was run on a 2% agarose gel and then column purified. The

oligos were designed with BbsI restriction enzyme sites flanking the sgRNA spacer sequence.

TOPO cloning randomly orients the oligos but with BbsI restriction sites, the 4nt overhangs that

are generated by BbsI digestion are unique to either the 5' or 3' end. By instituting BbsI-based

Golden Gate (GG) assembly, the spacer sequences can be oriented correctly into the vector. A

sgRNA expression cassette vector (pUC_sRNAi_LacZ, from GG1 cloning in Moosburner et al.

2020) was used to clone the N-gene oligo pool using BbsI-based GG assembly (Figure 3.7a).

First, pUC_sRNAi_LacZ was predigested with BbsI-HF and BamHI-HF restriction enzymes.

BbsI-HF digestion produces 4nt overhangs that are complementary to the 4nt overhangs

generated in the N-gene oligos. BamHI-HF digestion cuts only within the LacZ operon sequence

which was done in order to prevent recombination of the LacZ operon insert into the cloning

vector during GG assembly. Second, GG assembly was performed four times, pooled, and

precipitated into 2μL water. The precipitate GG pool was then transformed by heat shock into

One Shot™ MAX Efficiency™ DH5α-T1R Competent Cells (Invitrogen). All of the transformed

cells were plated onto 245mm square plates with LB agar supplemented with 100µg/mL carbomycin, Xgal, and IPTG and grown overnight at 37ºC. Next day, thousands of colonies were scraped off the plates and TOPO plasmids were extracted and isolated using the NucleoBond Xtra Maxi Kit (MachereyNagel). Prior to plasmid isolation the 5-10 blue colonies were removed from the selection plate using a scalpel. Colony PCR was also performed on 16 colonies to check for the correct insert size and all 16 were correct (Figure 3.7b).

### 3.6.3 Pooled sRNAi_gNG cloning into Cas9-StLDP episomes

Next, the sgRNA expression plasmid pool (sRNAi_gNG) was PCR amplified using only 20X cycles to prevent amplification bias. The PCR product was run on a 1.2% agarose gel and gel purified. The PCR primers amplify sRNAi_gNG to add BsaI restriction enzyme site on the 5' and 3' ends for GG cloning into the pBCas9-StLDP episomal vector with a kanamycin resistance cassette (KanR). pBRCas9-StLDP was predigested with BsaI-HFv2 restriction enzyme and gel purified. A three-piece BsaI-based GG assembly was performed using the predigested pBRCas9-StLDP vector, KanR, and the sRNAi_gNG PCR product to produce the episomal plasmid pool, pBRCas9-StLDP-gNG. (Figure 3.7c) The GG reaction was performed four times, pooled, and precipitated into 2µL water. The precipitate product was then transformed by electroporation into epi300 cells (Epicenter) that contained a pTA-MOB plasmid for bacterial conjugation. The transformation was performed twice. All of the transformed cells were plated onto 245mm square plates with LB agar supplemented with 100µg/mL carbenicillin, 10µg/mL tetracycline-HCl, 25µg/mL kanamycin, and 20µg/mL gentamicin and grown overnight at 37ºC. Next day, thousands of colonies were scraped off the plates. Since the transformed colonies contain both an episomal plasmid and the pTA-MOB plasmid, conjugation transformation into

*Phaeodactylum* can be performed using the pooled *E. coli* colonies. Five 2mL glycerol stocks of the pooled colonies were frozen at -80ºC. Episomal plasmids were extracted and isolated from the remaining E. coli using the NucleoBond Xtra Maxi Kit (MachereyNagel). Colony PCR was also performed on 10 colonies to check for the correct insert size and all 10 were correct (Figure 3.7d).

### 3.6.4 Conjugation of pooled pBRCas9-StLDP-gNG episomes

Conjugation of the N-gene knockout library episomes was performed using all 2mL of a glycerol stock collected after pBR-StLDP-gNG cloning. The conjugation protocol was adapted from Karas et al. (2015). Four day before conjugation, 400µL of wild-type *Phaeodactylum* cells were spread onto seawater plates (0.8% agarose, $NH_4^+/NO_3^-$ artificial sea water) at a concentration of 2e8 cells/mL. The night before conjugation, the glycerol stock was inoculated into fresh LB media and grown overnight at 37ºC. The next morning, five flasks of 50mL LB media were inoculated with 2mL of overnight culture, each, and grown for three to four hours or until the cell density reached an optical density ($OD_{600nm}$) between 0.8 and 1.0. During the three to four hour *E. coli* outgrowth, the plated *Phaeodactylum* cells were scraped off the plates and resuspended in ammonium media. The cell density was calculated and the cells were then concentrated to a volume of 1000µL and at a density of 5e8 cells/mL. Once the *E. coli* outgrowth reached the optimal density, all five 50mL cultures were pelleted and resuspended in 200µL SOC media, each. Finally, 200µL of *E. coli* were mixed with 200µL of *Phaeodactylum* and plated onto conjugation plates (0.8% agarose, $NH_4^+/NO_3^-$ artificial sea water, 5% LB media). The cells were dried under a PCR hood and incubated at 30ºC for 90 minutes. After, the plated cells were incubated at 18ºC in 14:10 (light:dark) diel growth conditions for 48 hours.

After 48 hours, the cells were scraped off of all five plates and collected together. The cells were then plated onto two 245mm square selection plates (0.8% agarose, $NH_4^+/NO_3^-$ artificial sea water) supplemented with 100μg/mL zeocin and 10μg/mL chloramphenicol. Chloramphenicol was supplemented to kill *E. coli* from conjugation. Zeocin was used to select for transformed *Phaeodactylum* containing a pBRCas9-StLDP-gNG episome. *Phaeodactylum* transformants were grown on selection plates for 4 weeks. Thousands of colonies formed on the plates. All colonies were then scraped off the selection plates using nitrogen-free artificial seawater. The transformed *Phaeodactylum* culture was finally split into two and resuspended in 100mL of either ammonium and nitrate media.

## 3.7 Screening and sorting of nitrogen gene mutant library for high-lipid producers

The ability to screen *Phaeodactylum* cells by flow cytometry in order to detect mCherry-StLDP lipid body localization presents an easy and fast method for culture wide lipid-accumulation analysis. Specifically, the ability to screen thousands to millions of cells quickly would permit for high throughput screening of mCherry-StLDP expression and localization. Here, a strong signal for mCherry will be used as a proxy for high-lipid producing *Phaeodactylum* cells within a culture of pooled knockout cell lines. The flow cytometry gates that were designated during the NR-KO cell line experiment (Figure 3.4 and Figure 3.5) were used to screen for high lipid producing *Phaeodactylum* transformants in the N-gene knockout library.

Two outgrowth cultures were analyzed for lipid accumulation by mCherry fluorescence detection (Figure 3.8). Both ammonium and nitrate supplemented cultures were allowed to grow for 72 hours prior to flow cytometry detection to prevent nitrogen limitation. The ultimate goal in

screening the mutant library is to identify cells with elevated lipid content during nitrogen-replete conditions. Also, the N-gene library targets nitrate specific genes as identified in Smith et al. (2019) as well as other genes associated with nitrate growth conditions. A large increase in the "P1" population was not expected since both ammonium and nitrate cultures were sampled at 72 hours and in nitrogen replete conditions. Nevertheless, the nitrate condition exhibited a slightly elevated reporter signal in the "mCherry-StLDP" gate (15.9% of cells for nitrate growth and 14.6% of cells for ammonium growth, Figure 8). Single cells from the "P1" population were sorted into 100μL of ammonium-supplemented media in 96-well plates.
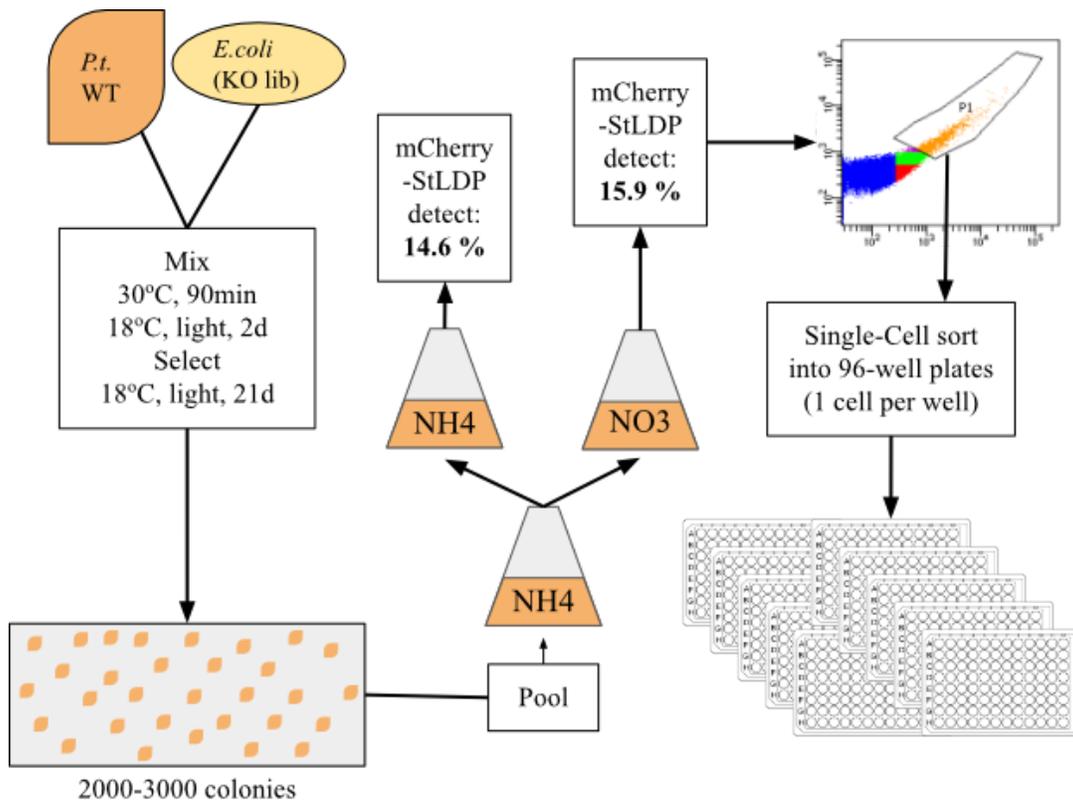


**Figure 3.8** Experiment workflow for delivering N-gene library to *Phaeodactylum (P.t.)* cell and single-cell sorting high-lipid producing cells by flow cytometry. The N-gene library (pBRCas9_StLDP_gNG) is contained within a *E. coli* and mixed with wild-type (WT) *Phaeodactylum* for conjugation transformation. Transformed *Phaeodactylum* cells are selected on large 245-mm plates. 2000 to 3000 colonies appeared on selection plates and were pooled into a single culture. The selection culture was split into two cultures, one grown on ammonium (NH4) media and a second on nitrate (NO3) media. After 72 hours growth, mCherry-StLDP reporter detection via flow cytometry was performed. Single cells were sorted from the "P1" gate from the nitrate culture into 96-well plates. 2304 single cells were sorted into ammonium and nitrate media.

**3.8 Future directions and discussion**

Presumably, single-cell sorting will result in the outgrowth of clonal *Phaeodactylum* cell lines. Traditional methods for producing clonal mutant cell lines utilized sub-cloning techniques like agar plate spreading and serial dilution of non-clonal cultures. Here, thousands of clonal cell lines were produced in less than one day using the fluorescence-activated cell sorting (FACS) methods. Each cell line should also contain a CRISPR-Cas9 episome that is unique to its sgRNA target sequence. After generation of the clonal cell lines, the sgRNA encoded on the episome will be sequenced for each cell line. Which sgRNAs are present in the sorted cell population will be an indication for 1) which gene knockouts were potentially produced and recovered within the sgRNA library and 2) which genes when mutated confer a lipid-accumulation phenotype when nitrate is available in the media.

It is possible that not all episomes, and thus not all sgRNA targets, in the N-gene library are within the sorted population. In fact, it is expected that some mutants will produce a lipid phenotype and some will not. Therefore, the sgRNAs that are accounted for within the sorted population should be skewed towards ones that result in a lipid-producing mutant cell line. The identified genes by sequencing the sgRNA target sequence will thus indicate that the genes, when mutated, may alter lipid biogenesis in *Phaeodactylum*.

The sorted mutant library could identify individual knockout cell lines as lipid producers but it may also aid in identifying metabolic hubs in nitrogen metabolism that impact lipid metabolism. Transcript-level analysis of *Phaeodactylum* during nitrogen starvation has revealed metabolic hubs that are among the most highly up-regulated genes that include genes involved in nitrate uptake and assimilation, glutamate and glutamine metabolism, the TCA cycle, and the urea cycle (Levitan et al. 2015). The genes that make up these metabolic hubs were included in

the N-gene library. These genes were highly upregulated during nitrogen starvation and thus during lipid accumulation. By including them in the N-gene library and screen for lipid accumulation, groups of genes may be identified that reveal entire metabolic hubs or features that impact lipid production. For instance, GS2 mutants accumulate lipids while retaining normal growth compared to wild-type. Normal growth also persists in mutants for cGOGAT and for the two-gene mutants of GS2 and cGOGAT (Chapter 2, Figure 2). Perhaps the entire chloroplast GS-GOGAT cycle is a metabolic hub that impacts lipid biogenesis without compromising growth. Nevertheless, this has yet to be explored in detail yet may be revealed through the analysis of the sorted mutant library.

Here, a high-throughput methodology was developed to identify multiple nitrogen-related genes simultaneously that may be involved in inducing lipid biogenesis. This method also produced a library of clonal cell lines with elevated lipid content. This library has the potential to be used to functionally evaluate multiple *Phaeodactylum* knockout cell lines for lipid accumulation. Unlike traditional functional genetics studies where the mutagenesis of one gene at a time is performed, this mutant library will provide *Phaeodactylum* researchers the ability to pick-and-choose mutant cell lines to investigate without having to rebuild a CRISPR-Cas9 mutagenesis tool, transformed cells, genotype, and finally screen for a lipid phenotype.

The application of producing a knockout library in *Phaeodactylum* provides the framework to produce additional knockout libraries for other gene sets or even all protein-coding genes. High-throughput functional genetics in diatoms could be useful beyond screening for lipid accumulation phenotypes. For instance, an all protein-coding knockout library could reveal all essential genes. Depending on the phenotypic screen implemented, a knockout library could help

answer a number of fundamental questions in diatom cell physiology, biotechnology, and ecology.

## 3.9 References

Ajjawi, I., J. Verruto, M. Aqui, L. B. Soriaga, J. Coppersmith, K. Kwok, L. Peach, E. Orchard, R. Kalb, W. Xu, T. J. Carlson, K. Francis, K. Konigsfeld, J. Bartalis, A. Schultz, W. Lambert, A. S. Schwartz, R. Brown and E. R. Moellering (2017). "Lipid production in Nannochloropsis gaditana is doubled by decreasing expression of a single transcriptional regulator." Nature Biotechnology **35**(7): 647-652.

Alipanah, L., J. Rohloff, P. Winge, A. M. Bones and T. Brembu (2015). "Whole-cell response to nitrogen deprivation in the diatom Phaeodactylum tricornutum." Journal of experimental botany **66**(20): 6281-6296.

Bao, A., D. J. Burritt, H. Chen, X. Zhou, D. Cao and L.-S. P. Tran (2019). "The CRISPR/Cas9 system and its applications in crop genome editing." Critical reviews in biotechnology **39**(3): 321-336.

Becker, M., H. Noll-Puchta, D. Amend, F. Nolte, C. Fuchs, I. Jeremias and C. J. Braun (2020). "CLUE: a bioinformatic and wet-lab pipeline for multiplexed cloning of custom sgRNA libraries." Nucleic acids research **48**(13): e78-e78.

Berges, J. A., D. O. Charlebois, D. C. Mauzerall and P. G. Falkowski (1996). "Differential effects of nitrogen limitation on photosynthetic efficiency of photosystems I and II in microalgae." Plant Physiology **110**(2): 689-696.

Chen, S., Neville E. Sanjana, K. Zheng, O. Shalem, K. Lee, X. Shi, David A. Scott, J. Song, Jen Q. Pan, R. Weissleder, H. Lee, F. Zhang and Phillip A. Sharp (2015). "Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis." Cell **160**(6): 1246-1260.

Chow, R. D. and S. Chen (2018). "Cancer CRISPR screens in vivo." Trends in cancer **4**(5): 349-358.

Coale, T. H., M. Moosburner, A. Horák, M. Oborník, K. A. Barbeau and A. E. Allen (2019). "Reduction-dependent siderophore assimilation in a model pennate diatom." <u>Proceedings of the National Academy of Sciences</u> **116**(47): 23609-23617.

Esrick, E. B., L. E. Lehmann, A. Biffi, M. Achebe, C. Brendel, M. F. Ciuculescu, H. Daley, B. MacKinnon, E. Morris, A. Federico, D. Abriss, K. Boardman, R. Khelladi, K. Shaw, H. Negre, O. Negre, S. Nikiforow, J. Ritz, S.-Y. Pai, W. B. London, C. Dansereau, M. M. Heeney, M. Armant, J. P. Manis and D. A. Williams (2020). "Post-Transcriptional Genetic Silencing of BCL11A to Treat Sickle Cell Disease." <u>New England Journal of Medicine</u> **384**(3): 205-215.

Fenster, J. A. and C. A. Eckert (2021). "High-Throughput Functional Genomics for Energy Production." <u>Current Opinion in Biotechnology</u> **67**: 7-14.

Frangoul, H., D. Altshuler, M. D. Cappellini, Y.-S. Chen, J. Domm, B. K. Eustace, J. Foell, J. de la Fuente, S. Grupp, R. Handgretinger, T. W. Ho, A. Kattamis, A. Kernytsky, J. Lekstrom-Himes, A. M. Li, F. Locatelli, M. Y. Mapara, M. de Montalembert, D. Rondelli, A. Sharma, S. Sheth, S. Soni, M. H. Steinberg, D. Wall, A. Yen and S. Corbacioglu (2020). "CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia." <u>New England Journal of Medicine</u> **384**(3): 252-260.

Greiner, A., S. Kelterborn, H. Evers, G. Kreimer, I. Sizova and P. Hegemann (2017). "Targeting of photoreceptor genes in Chlamydomonas reinhardtii via zinc-finger nucleases and CRISPR/Cas9." <u>The Plant Cell</u> **29**(10): 2498-2518.

Guihéneuf, F., S. Leu, A. Zarka, I. Khozin−Goldberg, I. Khalilov and S. Boussiba (2011). "Cloning and molecular characterization of a novel acyl−CoA: diacylglycerol acyltransferase 1−like gene (PtDGAT1) from the diatom Phaeodactylum tricornutum." <u>The FEBS journal</u> **278**(19): 3651-3666.

Hart, T., A. H. Y. Tong, K. Chan, J. Van Leeuwen, A. Seetharaman, M. Aregger, M. Chandrashekhar, N. Hustedt, S. Seth, A. Noonan, A. Habsid, O. Sizova, L. Nedyalkova, R. Climie, L. Tworzyanski, K. Lawson, M. A. Sartori, S. Alibeh, D. Tieu, S. Masud, P. Mero, A. Weiss, K. R. Brown, M. Usaj, M. Billmann, M. Rahman, M. Costanzo, C. L. Myers, B. J. Andrews, C. Boone, D. Durocher and J. Moffat (2017). "Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens." <u>G3 Genes|Genomes|Genetics</u> **7**(8): 2719-2727.

Hildebrand, M., A. K. Davis, S. R. Smith, J. C. Traller and R. Abbriano (2012). "The place of diatoms in the biofuels industry." <u>Biofuels</u> **3**(2): 221-240.

Hopes, A., V. Nekrasov, S. Kamoun and T. Mock (2016). "Editing of the urease gene by CRISPR-Cas in the diatom Thalassiosira pseudonana." Plant Methods **12**(1): 1-12.

Jallet, D., D. Xing, A. Hughes, M. Moosburner, M. P. Simmons, A. E. Allen and G. Peers (2020). "Mitochondrial fatty acid β−oxidation is required for storage−lipid catabolism in a marine diatom." New Phytologist **228**(3): 946-958.

Kao, P.-H. and I.-S. Ng (2017). "CRISPRi mediated phosphoenolpyruvate carboxylase regulation to enhance the production of lipid in Chlamydomonas reinhardtii." Bioresource technology **245**: 1527-1537.

Kim, J., S. Lee, K. Baek and E. Jin (2020). "Site-specific gene knock-out and on-site heterologous gene overexpression in Chlamydomonas reinhardtii via a CRISPR-Cas9-mediated knock-in method." Frontiers in plant science **11**: 306.

Kim, Y.-K., G. Wee, J. Park, J. Kim, D. Baek, J.-S. Kim and V. N. Kim (2013). "TALEN-based knockout library for human microRNAs." Nature structural & molecular biology **20**(12): 1458.

Levering, J., J. Broddrick, C. L. Dupont, G. Peers, K. Beeri, J. Mayers, A. A. Gallina, A. E. Allen, B. O. Palsson and K. Zengler (2016). "Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom." PloS one **11**(5): e0155038.

Levitan, O., J. Dinamarca, E. Zelzion, D. S. Lun, L. T. Guerra, M. K. Kim, J. Kim, B. A. Van Mooy, D. Bhattacharya and P. G. Falkowski (2015). "Remodeling of intermediate metabolism in the diatom Phaeodactylum tricornutum under nitrogen stress." Proceedings of the National Academy of Sciences **112**(2): 412-417.

Li, X., W. Patena, F. Fauser, R. E. Jinkerson, S. Saroussi, M. T. Meyer, N. Ivanova, J. M. Robertson, R. Yue, R. Zhang, J. Vilarrasa-Blasi, T. M. Wittkopp, S. Ramundo, S. R. Blum, A. Goh, M. Laudon, T. Srikumar, P. A. Lefebvre, A. R. Grossman and M. C. Jonikas (2019). "A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis." Nature Genetics **51**(4): 627-635.

Lian, J., M. HamediRad and H. Zhao (2018). "Advancing metabolic engineering of Saccharomyces cerevisiae using the CRISPR/Cas system." Biotechnology journal **13**(9): 1700601.

Liu, H.-J., L. Jian, J. Xu, Q. Zhang, M. Zhang, M. Jin, Y. Peng, J. Yan, B. Han, J. Liu, F. Gao, X. Liu, L. Huang, W. Wei, Y. Ding, X. Yang, Z. Li, M. Zhang, J. Sun, M. Bai, W. Song, H. Chen, X. a. Sun, W. Li, Y. Lu, Y. Liu, J. Zhao, Y. Qian, D. Jackson, A. R. Fernie and J. Yan (2020). "High-Throughput CRISPR/Cas9 Mutagenesis Streamlines Trait Gene Identification in Maize[OPEN]." The Plant Cell **32**(5): 1397-1413.

McCarthy, J. K., S. R. Smith, J. P. McCrow, M. Tan, H. Zheng, K. Beeri, R. Roth, C. Lichtle, U. Goodenough, C. P. Bowler, C. L. Dupont and A. E. Allen (2017). "Nitrate Reductase Knockout Uncouples Nitrate Transport from Nitrate Assimilation and Drives Repartitioning of Carbon Flux in a Model Pennate Diatom." The Plant Cell **29**(8): 2047-2070.

Moosburner, M. A., P. Gholami, J. K. McCarthy, M. Tan, V. A. Bielinski and A. E. Allen (2020). "Multiplexed knockouts in the model diatom phaeodactylum by episomal delivery of a selectable cas9." Frontiers in microbiology **11**: 5.

Moses, C., B. Garcia-Bloj, A. R. Harvey and P. Blancafort (2018). "Hallmarks of cancer: The CRISPR generation." European Journal of Cancer **93**: 10-18.

Nymark, M., A. K. Sharma, T. Sparstad, A. M. Bones and P. Winge (2016). "A CRISPR/Cas9 system adapted for gene editing in marine algae." Scientific reports **6**(1): 1-6.

Park, R. V., H. Asbury and S. M. Miller (2020). "Modification of a Chlamydomonas reinhardtii CRISPR/Cas9 transformation protocol for use with widely available electroporation equipment." MethodsX **7**: 100855.

Picariello, T., Y. Hou, T. Kubo, N. A. McNeill, H.-a. Yanagisawa, T. Oda and G. B. Witman (2020). "TIM, a targeted insertional mutagenesis method utilizing CRISPR/Cas9 in Chlamydomonas reinhardtii." PloS one **15**(5): e0232594.

Rastogi, A., O. Murik, C. Bowler and L. Tirichine (2016). "PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing." BMC bioinformatics **17**(1): 1-4.

Rumin, J., H. Bonnefond, B. Saint-Jean, C. Rouxel, A. Sciandra, O. Bernard, J.-P. Cadoret and G. Bougaran (2015). "The use of fluorescent Nile red and BODIPY for lipid measurement in microalgae." Biotechnology for biofuels **8**(1): 1-16.

Sakaguchi, T., K. Nakajima and Y. Matsuda (2011). "Identification of the UMP synthase gene by establishment of uracil auxotrophic mutants and the phenotypic complementation system in the marine diatom Phaeodactylum tricornutum." Plant physiology **156**(1): 78-89.

Serif, M., G. Dubois, A.-L. Finoux, M.-A. Teste, D. Jallet and F. Daboussi (2018). "One-step generation of multiple gene knock-outs in the diatom Phaeodactylum tricornutum by DNA-free genome editing." Nature communications **9**(1): 1-10.

Shalem, O., N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. S. Mikkelsen, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench and F. Zhang (2014). "Genome-scale CRISPR-Cas9 knockout screening in human cells." Science **343**(6166): 84-87.

Sharma, A. K., M. Nymark, T. Sparstad, A. M. Bones and P. Winge (2018). "Transgene-free genome editing in marine algae by bacterial conjugation–comparison with biolistic CRISPR/Cas9 transformation." Scientific reports **8**(1): 1-11.

Shemesh, Z., S. Leu, I. Khozin-Goldberg, S. Didi-Cohen, A. Zarka and S. Boussiba (2016). "Inducible expression of Haematococcus oil globule protein in the diatom Phaeodactylum tricornutum: association with lipid droplets and enhancement of TAG accumulation under nitrogen starvation." Algal research **18**: 321-331.

Shields, R. C. and P. A. Jensen (2019). "The bare necessities: Uncovering essential and condition−critical genes with transposon sequencing." Molecular oral microbiology **34**(2): 39-50.

Slattery, S. S., H. Wang, D. J. Giguere, C. Kocsis, B. L. Urquhart, B. J. Karas and D. R. Edgell (2020). "Plasmid-based complementation of large deletions in Phaeodactylum tricornutum biosynthetic genes generated by Cas9 editing." Scientific reports **10**(1): 1-12.

Smith, S. R., C. L. Dupont, J. K. McCarthy, J. T. Broddrick, M. Oborník, A. Horák, Z. Füssy, J. Cihlář, S. Kleessen, H. Zheng, J. P. McCrow, K. K. Hixson, W. L. Araújo, A. Nunes-Nesi, A. Fernie, Z. Nikoloski, B. O. Palsson and A. E. Allen (2019). "Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom." Nature Communications **10**(1): 4552.

Takahashi, T. (2019). "Routine management of microalgae using autofluorescence from chlorophyll." Molecules **24**(24): 4441.

Taparia, Y., A. Zarka, S. Leu, R. Zarivach, S. Boussiba and I. Khozin-Goldberg (2019). "A novel endogenous selection marker for the diatom Phaeodactylum tricornutum based on a unique mutation in phytoene desaturase 1." Scientific reports **9**(1): 1-12.

Vinayak, V., K. M. Manoylov, H. Gateau, V. Blanckaert, J. Hérault, G. Pencréac'h, J. Marchand, R. Gordon and B. Schoefs (2015). "Diatom Milking: A Review and New Approaches. Marine Drugs 2015, 13, 2629–2665." Marine drugs **13**(12): 7301.

Weiss, T., C. Wang, X. Kang, H. Zhao, M. Elena Gamo, C. G. Starker, P. A. Crisp, P. Zhou, N. M. Springer, D. F. Voytas and F. Zhang (2020). "Optimization of multiplexed CRISPR/Cas9 system for highly efficient genome editing in Setaria viridis." The Plant Journal **104**(3): 828-838.

Yoneda, K., M. Yoshida, I. Suzuki and M. M. Watanabe (2016). "Identification of a major lipid droplet protein in a marine diatom Phaeodactylum tricornutum." Plant and Cell Physiology **57**(2): 397-406.

Yoneda, K., M. Yoshida, I. Suzuki and M. M. Watanabe (2018). "Homologous expression of lipid droplet protein-enhanced neutral lipid accumulation in the marine diatom Phaeodactylum tricornutum." Journal of Applied Phycology **30**(5): 2793-2802.