Electronic Thesis and Dissertation Repository

10-14-2020 2:00 PM

# Development of the Model Marine Diatom Phaeodactylum tricornutum for Synthetic Biology Applications

Samuel S. Slattery, *The University of Western Ontario*

Supervisor: Edgell, David R., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biochemistry
© Samuel S. Slattery 2020

# Abstract

Harnessing organisms for protein and chemical production is useful to the scientific community and has applications in the fuel, food, and pharmaceutical industries. Biological systems commonly used for industrial chemical production include yeast and bacteria due to their fast growth rates and potential for high product yields. However, biologically active proteins, such as for human therapeutics, usually require production in mammalian and insect systems that are prohibitively expensive to grow at scale. Recently, photoautotrophic microalgae have emerged as promising platforms, as some species can be grown quickly and inexpensively at large scales and have the potential to produce biologically active proteins that mimic those produced in mammalian systems. Thus, they combine advantages from several traditional biological systems, but require further genetic and biotechnological development for their full potential to be unlocked.

Ideal biological production systems generally possess a variety of genetic tools to enable foreign gene expression, and genome editing systems to create desired genetic modifications. Here I present a robust gene editing system, novel genetic tools, and valuable strains for the microalga *Phaeodactylum tricornutum*. First, I identify novel endogenous regulatory elements that can drive expression of foreign genes in *P. tricornutum*, and design a plasmid-based Cas9 gene editing system for this species. Next, I demonstrate the utility of these tools with the generation of auxotrophic *P. tricornutum* strains through the genetic knockout of key enzymes in the uracil and histidine biosynthesis pathways. I complement the phenotype of the auxotrophs by introducing intact versions of these genes on replicating plasmids, and demonstrate that these genes function as selective markers for transformation of their respective auxotrophic strain. Finally, I highlight the potential of these tools by creating a *P. tricornutum* expression system for the production of SARS-CoV-2 antigens. This will potentially address the need for a cheap, scalable source of serologically-active antigens for population-wide serological testing to combat the SARS-CoV-2 pandemic, and this system can be rapidly adapted to tackle future pandemics. I hope that these novel tools and strains will broaden the potential applications of *P. tricornutum* for industrial production of high-value products, and further the study of diatom biology.

**Keywords:** *Phaeodactylum tricornutum*, microalgae, genetic tools, Cas9 gene editing, auxotrophs, synthetic biology, production strains.

# Lay Summary

There is a growing demand for methods to produce food, drugs, and fuels cheaply and in an environmentally sustainable way. One promising solution is to exploit organisms through genetic engineering to synthesize these products for us. Yeast, bacteria, and mammalian cell cultures are commonly used for this purpose but aren't suitable for all applications, such as large-scale production of pharmaceutical compounds that require special modifications to be functional. Microalgae are promising alternative platforms for these applications as many microalgal species are photosynthetic, making them inexpensive to grow at large scales, and several species have been deemed safe to eat. Some microalgae have also shown the potential to perform the special modifications that are required to produce functional pharmaceuticals. However, many promising microalgae species lack the proper development and tools required to engineer them for these purposes. Here, I develop genetic tools and strains of the marine microalga *Phaeodactylum tricornutum* to enable industrial-scale production of high-value products in this species.

First I create novel plasmid-based expression tools, and a CRISPR/Cas9 gene-editing system for *P. tricornutum*. I then use these tools to generate auxotrophic strains of this species by knocking out key genes in the *P. tricornutum* uracil and histidine biosynthesis pathways. I show that the auxotroph phenotypes can be complemented by plasmid-encoded intact versions of these genes, providing an antibiotic-free plasmid selection system. Finally, I use these tools and strains to engineer *P. tricornutum* to produce the SARS-CoV-2 spike protein to potentially address the need for a cheap, abundant source of viral proteins required for serological testing. This expression system can be easily adapted to address future pandemics, or produce human therapeutics. The tools and strains generated here will enable industrial-scale production of valuable products in *P. tricornutum*, and further the study of diatom biology.

# Co-Authorship Statement

For the peer-reviewed publications presented in Chapter 2 and 3, and unpublished work presented in Chapter 4, Samuel Slattery performed the research with exceptions noted below. Samuel Slattery and David Edgell conceived and designed the experiments and analyzed the data. Samuel Slattery and David Edgell wrote a majority of the manuscripts.

Chapter 2 - Bogumil Karas helped conceive the project and assisted in design and data analysis. Andrew Diamond and Isabel Desgagné-Penix designed the vanillin biosynthetic pathway. Jasmine Therrien, Teah Jazey, Kyle Lee, and Zachary Klassen built the vanillin biosynthetic plasmid and screened for vanillin biosynthetic plasmid stability. Jeremy Lant helped design the Cas9 expression plasmids. Helen Wang helped perform the urease knockout experiments. Bogumil Karas helped perform promoter screening assays.

Chapter 3 - Bogumil Karas helped conceive the project and assisted in design and data analysis. Daniel Giguere helped design, perform, and analyze data for the targeted long-read sequence experiments. Csanad Kocsis helped perform the PtPRA-PH/CH knockout experiments. Helen Wang helped perform the PtUMPS and PtPRA-PH/CH knockout experiments. Bradley Urquhart helped design and perform the PtUMPS knockout metabolomic characterization experiments.

Chapter 4 - Bogumil Karas and Greg Gloor helped conceive the project and assisted in design and data analysis. Jeremy Lant and Mallory Frederick helped perform Western blots. Daniel Giguere helped design, perform and analyze data for the RNASeq experiments. Emily Stuckless performed and analyzed data for the growth curve assays and glycerol conjugation experiments. Emily Stuckless helped perform plasmid stability assays.

# Dedication

This thesis is dedicated to my lab-mates, mentors, friends, and family members who have given me the support and encouragement I needed to succeed.

*Any sufficiently advanced technology is indistinguishable from magic.*

ARTHUR C. CLARKE

# Acknowledgments

# Table of Contents

# List of Figures

# List of Tables

# List of Appendices

# List of Abbreviations, Symbols, and Nomenclature

| | |
|---|---|
| $\mu$E | micro Einstein |
| 40SRPS8 | 40S ribosomal protein S8 |
| 4CL | 4-hydroxycinnamoyl-CoA ligase |
| 5-FOA | 5-fluoroorotic acid |
| 6xHis-tag | 6-histidine-tag |
| aa | amino acid |
| ade2 | phosphoribosylaminoimidazole carboxylase |
| ALA | alpha-linolenic acid |
| ApE | A plasmid Editor |
| ARG7 | argininosuccinate lyase |
| ARSH4 | autonomous replication sequence |
| ATCC | american type culture collection |
| BF | biopharmaceutical |
| BLAST | basic local alignment search tool |
| bp | base pair |
| BRIG | BLAST ring image generator |
| BSA | bovine serum albumin |
| C3H | p-coumarate 3-hydroxylase |
| C4H | trans-cinnamate 4-hydroxylase |
| CAH | CEN6-ARSH4-HIS3 |
| Cas9 | CRISPR-associated protein 9 |
| cat | chloramphenicol acetyltransferase |
| CCAP | culture collection of algae and protozoa |
| cDNA | complementary DNA |
| CEN6 | chromosome 6 centromere |
| CIP1 | replication-associated protein of a *Chaetoceros lorenzianus*-infecting DNA virus |
| COMT | caffeic acid 3-O-methyltransferase |
| CRISPR | clustered regularly interspaced short palindromic repeats |
| CRISPRi | CRISPR interference |
| Ct | cycle threshold |
| dCas9 | nuclease-dead cas9 |
| DHA | docosahexaenoic acid |
| DNA | deoxyribonucleic acid |
| DSB | double-strand break |
| DTT | dithiothreitol |
| ECL | enhanced chemiluminescence |
| EDTA | ethylenediaminetetraacetic acid |
| EF-1$\alpha$ | elongation factor 1 alpha |
| EPA | eicosapentanoic acid |
| ESI | electrospray ionization |

| | |
|---|---|
| EST | expressed sequence tag |
| EtOH | ethanol |
| FBAC2 | fructose-bisphosphate aldolase |
| FcpA | fucoxanthin chlorophyll a/c binding protein A |
| FcpB | fucoxanthin chlorophyll a/c binding protein B |
| FcpC | fucoxanthin chlorophyll a/c binding protein C |
| FcpD | fucoxanthin chlorophyll a/c binding protein D |
| FcpF | fucoxanthin chlorophyll a/c binding protein F |
| *g* | standard gravity |
| GAT | glyphosate acetyltransferase |
| gDNA | genomic DNA |
| GFP | green fluorescence protein |
| GRAS | generally recognized as safe |
| H4-1B | histone 4 1B |
| HASP1 | highly abundant secreted protein 1 |
| HA-tag | hemagglutinin-tag |
| HCT | hydroxycinnamoyl transferase |
| HDR | homology directed repair |
| his3 | imidazoleglycerol-phosphate dehydratase |
| HIS4 | PtPRA-PH/CH |
| hr | hour |
| HR | homologous recombination |
| HRP | horseradish peroxidase |
| INDEL | insertion or deletion of bases |
| IS | internal standard |
| kb | kilobase |
| kDa | kilodalton |
| KEGG | kyoto encyclopedia of genes and genomes |
| LB | lysogeny broth |
| ligIV | DNA ligase IV |
| MAA | mycosporin-like amino acid |
| MAT$\alpha$ | mating type alpha |
| Mb | megabase |
| met14 | adenylyl-sulfate kinase |
| min | minute |
| MPX | multiplex |
| mRNA | messenger RNA |
| N | nucleotide |
| Na$_2$EDTA | sodium ethylenediaminetetraacetic acid |
| *nat* | nourseothricin acetyltransferase gene |
| NHEJ | non-homologous end joining |
| Ni-NTA | nickel-nitrilotriacetic acid |
| NIT1 | nitrate reductase |
| N-linked | asparagine-linked |
| NLS | nuclear localization signal |

| | |
|---|---|
| NP | nitrate phosphate |
| NrsR | nourseothricin resistance |
| NTC | nourseothricin |
| ODC | orotidine-5'-phosphate decarboxylase |
| OMP | orotidine monophosphate |
| OPRT | orotate phosphoribosyl transferase |
| ORF | open reading frame |
| oriT | origin of transfer |
| P2A | porcine teschovirus-1 2A |
| PAGE | polyacrylamide gel electrophoresis |
| PAL | phenylalanine ammonia lyase |
| PAM | protospacer adjacent motif |
| PCR | polymerase chain reaction |
| pds | phytoene desaturase |
| PEG | polyethylene glycol |
| PES | polyethersulfone |
| PMSF | phenylmethylsulfonyl fluoride |
| Pt | *Phaeodactylum tricornutum* |
| PtAPT | *Phaeodactylum tricornutum* adenine phosphoribosyl transferase |
| PtI3GPS-PRAI | *Phaeodactylum tricornutum* bifunctional indole-3-glycerol-phosphate synthase and phosphoribosylanthranilate isomerase |
| PtIGPS | *Phaeodactylum tricornutum* imidazole glycerol phosphate synthase |
| PTM | post-translational modification |
| PtPRA-PH/CH | *Phaeodactylum tricornutum* bifunctional phosphoribosyl-ATP pyrophosphohydrolase and phosphoribosyl-AMP cyclohydrolase |
| PtUMPS | *Phaeodactylum tricornutum* uridine monophosphate synthase |
| PUFA | polyunsaturated fatty acid |
| PVDF | polyvinylidene difluoride |
| qPCR | quantitative PCR |
| RBCMT | ribulose-1,5 bisphosphate carboxylase/oxygenase small subunit N-methyltransferase |
| RBD | receptor-binding domain |
| RNA | ribonucleic acid |
| RNAi | RNA interference |
| RNAseq | RNA-sequencing |
| RPS4 | 40S ribosomal protein S4 |
| rRNA | ribosomal RNA |
| RT-qPCR | quantitative real-time PCR |
| s | second |
| SAR | Stramenopiles, Alveolates, and Rhizaria |
| SARS-CoV-2 | severe acute respiratory syndrome coronavirus 2 |
| SDS | sodium dodecyl sulfate |
| sgRNA | single guide RNA |
| *Sh ble* | bleomycin resistance protein gene |
| SNP | single-nucleotide polymorphism |

| | |
|---|---|
| SOC | super optimal broth with catabolite repression |
| T2A | thosea asigna virus 2A |
| T7EI | T7 endonuclease I |
| TAL | tyrosine ammonia lyase |
| TALEN | transcription activator-like effector nuclease |
| TBS | tris-buffered saline |
| TE | tris EDTA |
| TevCas9 | Tev CRISPR-associated protein 9 |
| TPM | transcripts per kilobase million |
| Tris-HCl | tris(hydroxymethyl)aminomethane hydrochloride |
| trp1 | N-(5'-phosphoribosyl)anthranilate isomerase |
| TUFA | elongation factor Tu |
| UMP | uridine monophosphate |
| UPLC | ultra performance liquid chromatography |
| ura3 | orotidine 5-phosphate decarboxylase |
| UV | ultraviolet |
| VAN | vanillin synthase |
| w/v | weight/volume |
| w/w | weight/weight |
| WT | wild type |
| YPD | yeast extract-peptone-dextrose |
| Zeo | zeocin |
| ZFN | zinc-finger nuclease |

# Chapter 1

# Introduction

## 1.1   The biodiversity of algae

Eukaryotic algae are a collection of diverse photosynthetic organisms with members found in three taxonomic supergroups including plantae, excavates, and SAR (stramenopiles, alveolates, and rhizaria), as well as the haptophytes and cryptophytes groups, with most lineages within the plantae and SAR supergroups (Fig. 1.1) (1–3). The plantae supergroup contains algae species that exclusively arose from primary endosymbiosis events where a photosynthetic cyanobacteria was engulfed by a heterotrophic eukaryote, followed by most of the cyanobacterial genome being transferred to the host nucleus (4). These endosymbiosis events gave rise to green and red algae, glaucophytes, and higher plants. Algae that are found in the SAR and excavates supergroups arose from secondary endosymbiosis events involving phagocytosis of a primary red or green algae, where the nucleus of the engulfed primary algae was largely reduced or lost completely (5). The SAR supergroup includes the lineages stramenopiles, alveolates and rhizaria. The stramenopiles, also called heterokonts, encompass diatoms and brown algae which are characterized by their golden-brown or brownish-green colour. These arose through secondary endosymbiosis of a primary red algae species and include some of the most abundant algal species (5). Algae are generally categorized as either red (Rhodophyta), green (Chlorophyta), or brown (Phaeophyta), and classified by size as macroalgae or microalgae. Macroalgae, such as seaweed, are large, multi-cellular organisms visible with the naked eye, while microalgae are microscopic, unicellular organisms.

Microalgae have adapted to live in a wide range of conditions. They have been found in saltwater, freshwater and terrestrial habitats, in extremes of hot and cold, varying mineral compositions
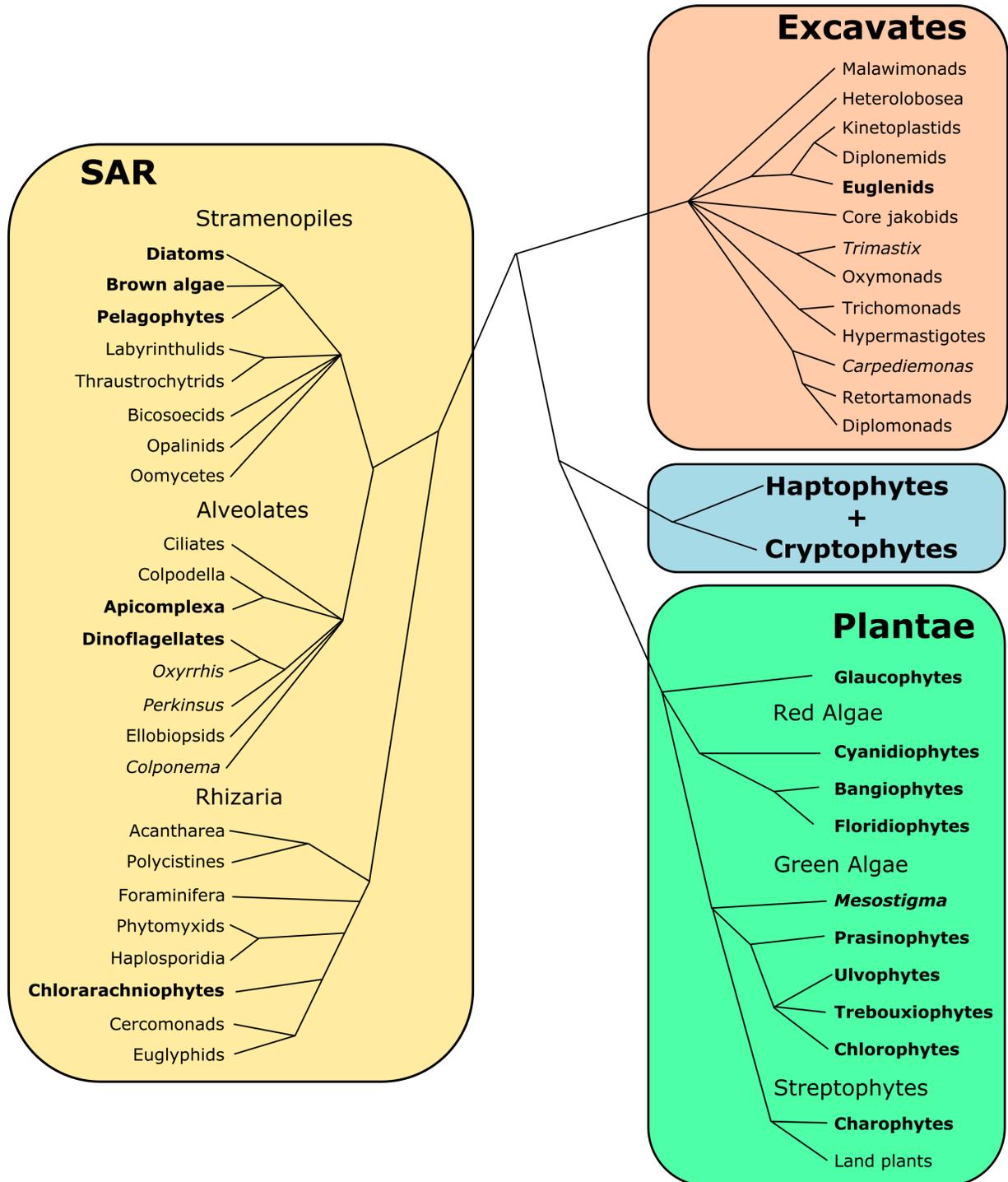
**Figure 1.1: A possible evolutionary tree of eukaryotic algae.** Diagram depicting supergroups in which algae lineages (**bolded**) are found, and the phylogenetic links that connect them.

and pH values, very low and very high light, and even industrial wastewater (6). While some microalgae are heterotrophs, most are photoautotrophs capable of synthesizing their own food from inorganic substances such as $CO_2$ and salts, using light as an energy source (7). A subset of these are mixotrophic, able to photosynthesize and acquire exogenous organic nutrients (8).

Microalgae are extremely abundant, being found in almost every natural environment, and play an important role in the global ecosystem (9). They are responsible for more than half of all atmospheric oxygen production and are potent carbon sequestering organisms (10). Additionally, microalgae are integral members of the global food chain with diatoms being a particularly important group, performing roughly one-fifth of all photosynthesis on earth (11). As a result, diatoms produce a huge abundance of organic carbon and are the foundation of aquatic ecosystem food chains worldwide.

## 1.2   Industrial applications of microalgae

Due to their extensive genetic diversity, microalgae represent an invaluable resource of unique and potentially useful biomolecules that make them very attractive for bioprospecting. They produce a wide range of natural products, including lipids, proteins, polysaccharides, pigments, vitamins, bioactive compounds, and antioxidants (12). Additionally, many microalgae species have high growth rates even under minimal photosynthetic conditions. These characteristics and others make them ideal for large scale production of natural products. As a result, they have become the focus of many industrial efforts to engineer organisms for commercial chemical biosynthesis, driven in part by the initiatives to use microalgae as a renewable source of biofuels (13, 14). However, it is believed that economically viable biofuel production from microalgae can only be possible if coupled with the production of other high-value co-products (15). To date, microalgae have been used for the biosynthesis of products including biofuels, cosmetics, food, and biopharmaceuticals (Fig. 1.2) (12, 16–21).

### 1.2.1   Biofuels

As the global demand for fuel energy increases, fossil fuels are rapidly being depleted and the need for alternative sources of fuels is becoming a major concern. One promising alternative is biofu-
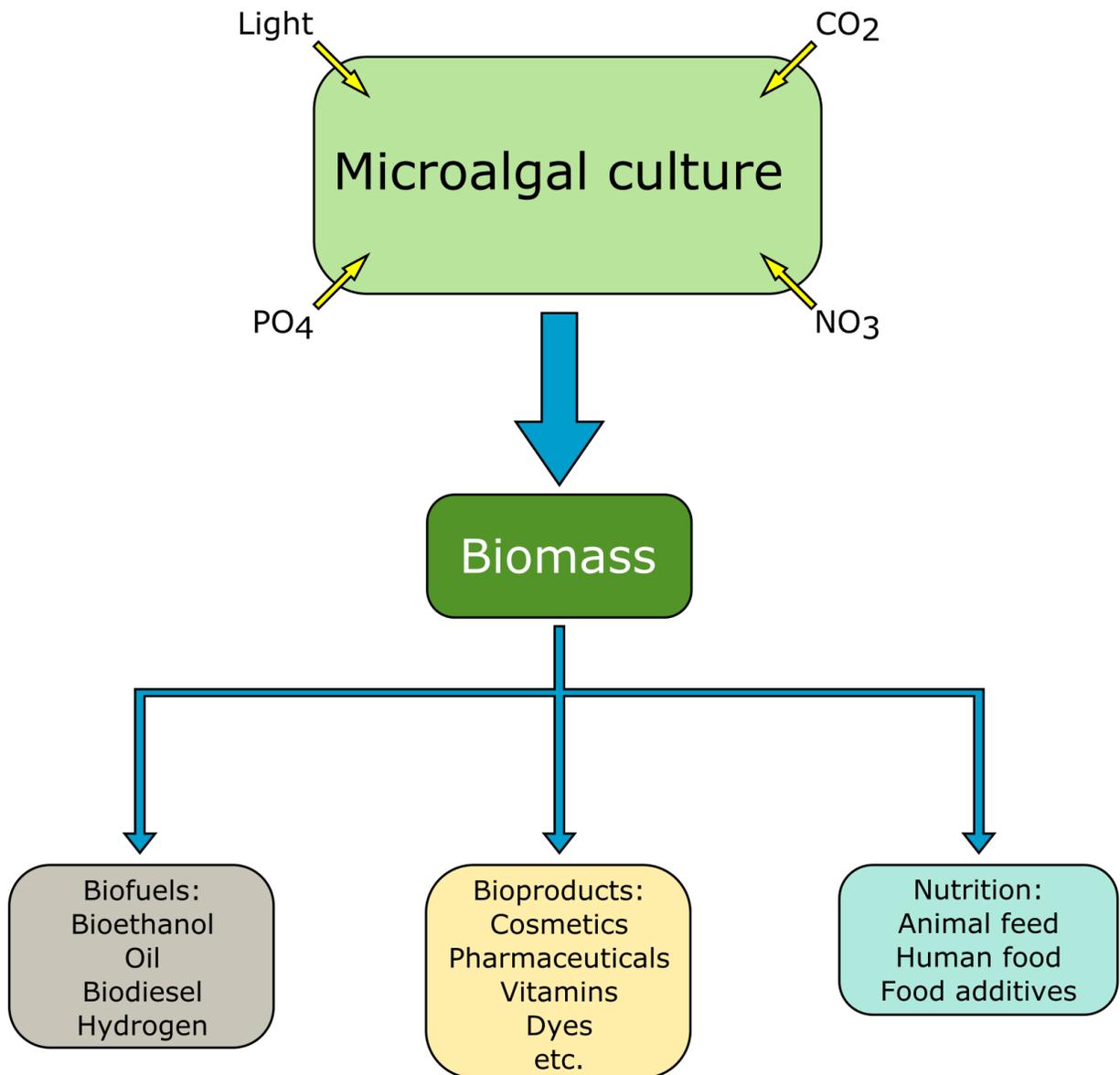
**Figure 1.2: Industrial applications of microalgae.** Microalgae can be cultured to produce biomass for direct consumption, refinement into biofuels, or extraction of valuable bioproducts produced naturally or through genetic modification.

els, fuels derived from living matter. The first generation of biofuels are fuels created using food crops such as corn, wheat, sugar cane, and rapeseed (22). These crops generally have high sugar or starch content and are converted into biofuels such as ethanol and biodiesel through fermentation processes. Second generation biofuels are those made from feedstocks such as food waste (e.g. waste vegetable oil), and crops that are not typically food crops, including lignocellulosic agriculture like switchgrass (23). Because they are not intended as food crops, they are usually grown on marginal land and do not compete with food crops for agricultural space, but generally require more complicated biofuel conversion processes than first generation feedstocks.

Third generation biofuels are those derived from algae. Microalgae are rapidly gaining interest as a feedstock for biofuel production because of their potential for high lipid production which outcompetes even the best first and second generation feedstocks (24). Microalgae have several advantages over traditional feedstocks, but are limited by a number of factors that prevent them from becoming the preferred method of biofuel production. Unlike agricultural biofuel feedstocks, microalgae are capable of year round production, resulting in much higher biofuel yields, and do not require herbicides or pesticides, reducing environmental pollution (25). Microalgae also need less water than terrestrial crops despite growing in an aqueous environment, and do not require cultivation on agricultural land, and thus do not compete for land use with food crops (26). Some microalgae have rapid growth rates and high lipid content reaching up to nearly 80% of dry-cell weight, such as in case of *Botryococcus braunii* (27), further increasing their biofuel production potential. As an added benefit, microalgae produce organic compounds that can be converted into valuable co-products. After biofuel extraction, their biomass can be fermented for further biofuel production, or used as fertilizer or feed for agriculture (24).

Photoautotrophic microalgae require five main inputs for sustained growth; light, water, carbon dioxide, nitrogen, and phosphorous, which all increase input costs and impact their sustainability. This can be partially offset by growing the microalgae in wastewater, which can meet some of their nitrogen and phosphorous needs (27). Their $CO_2$ requirement can be addressed by cultivating them in photobioreactors that can be set up to feed off of industrial waste $CO_2$, which has the added benefit of reducing industrial carbon emissions (27). In addition, it has been shown that growth under stressful conditions, such as nitrogen deprivation, can greatly increase oil production in some microalgae species (28). This has the potential to offset nitrogen requirements without

impacting biofuel yield.

Despite these approaches, microalgae still require a large amount of water, nitrogen, and phosphorous which can only be partially fulfilled by growth in industrial wastewater and will ultimately require extensive use of fertilizers. It is predicted that biofuel production from microalgae will need to be coupled with valuable co-product production to make them economically feasible (23). In addition, the infrastructure to support large-scale microalgae-based biofuel production is not fully established, partially due to the fact that there are currently no microalgae species productive enough to offset the cost of initial setup and continued maintenance and input costs of microalgae culturing (29). Overcoming these issues will likely require genetic engineering of promising microalgae species to create a balance between high lipid production, fast growth rate, a suitable lipid profile for biofuel processing, and production of high value co-products.

## 1.2.2 Cosmetics

Cosmetics are agents used to clean, protect, perfume, and improve or modify the appearance of skin, hair, nails, or teeth (30). Skin moisturizers, makeup, and toothpaste are common examples of cosmetics that are composed of a variety of ingredients, many of which can be derived from natural sources such as pigments and scents from plants, and lubricants like squalene from shark liver (19). Cosmetic ingredients can also be chemically synthesized if the natural source is scarce or expensive, or if there is no known natural source of a particular compound. However, for some traditionally derived cosmetic ingredients, harvesting them from plant and animal sources is potentially time consuming and can be too inefficient to meet market demands, while the chemical synthesis approach can be difficult and expensive. In addition, there is an increasing demand for naturally sourced cosmetic ingredients (30). As a result, cosmetic manufacturers have begun investigating alternative natural sources of cosmetic compounds, with one of the most promising being microalgae.

Microalgae are an attractive source of cosmetic compounds as they are generally easier to cultivate than many of the traditional plant sources of cosmetic ingredients. Additionally, certain species of microalgae either produce useful quantities of desired cosmetic ingredients naturally, or can be engineered to produce them. In fact, the ingredients of many cosmetics, including cer-

tain pigments and dyes, photoprotectants, moisturizers, thickeners, lubricants, and anti-ageing compounds, are already being sourced from microalgae (19). For example, phycocyanin, phycoerythrin, and allophycocyanin are pigments in the phycobilin family that are unique to algae and cyanobacteria. These pigments are harvested from species such as *Spirulina*, *Porphyridium*, *Rhodella* and *Galdieria* to be used in cosmetics and other applications (31, 32). Porphyra-334, asterina-330, and palythinol are examples of mycosporin-like amino acids (MAAs) which show promise as photoprotectant compounds due to their ability to absorb ultraviolet (UV) radiation and disperse it as heat energy (33). MAAs are produced by a number of organisms including algae and cyanobacteria, with some MAAs already on the market for use in the anti-photo-ageing cosmetic Helioguard 365 (34), while others are being investigated for use in sunscreens and UV-protective biomaterials (33). Polysaccharides from microalgae such as those found in the genera *Chlorella* and *Porphyridium* can act as moisturizing and gelling/thickening agents for cosmetics, much like the polysaccharides derived from macroalgae (35). Squalene is a compound harvested from shark livers to be used as a lubricant in cosmetics, skin care products, and hair conditioning agents. Squalene has been shown to be produced at appreciable levels (20% dry-cell weight) in the microalga *Aurantiochytrium mangrovei*, providing an alternative, sustainable source for this high-value compound (19). These examples of the many compounds that can be derived from microalgae for use in cosmetic formulations highlight the potential of microalgae as a source of valuable products.

### 1.2.3    Food and dietary supplements

Algae are an excellent natural source of carbohydrates, lipids, proteins, and other nutrients. In fact, algae have been used for nutritional purposes for much of recorded history (36). The *Porphyra* seaweeds have been consumed by almost every culture that has had access to them for thousands of years, and is most commonly recognized as the edible seaweed nori used in sushi preparation (37). The ancient peoples of Chad and Mexico consumed *Spirulina* and *Aphanizomenon* as a food source (38), and ~2000 years ago *Nostoc* was consumed as a means to survive famine in China (39). The consumption of *Nostoc flagelliforme* became so popular that it continues to be consumed by Chinese peoples today.

While macro- and microalgae have been exploited as a food source for much of human history,

the purposeful culturing of microalgae in specialized facilities is a relatively modern practice. This is becoming more popular as biotechnology and food industries show greater interest in microalgae as a means to cheaply produce valuable nutritional products. The first examples of large-scale commercial microalgae culturing for food production were with *Chlorella* in Japan in the 1960s, *Spirulina* in Mexico in the 1970s, and *Dunaliella salina* in Australia in the 1980s (19). Today, *Spirulina* and *Chlorella* are the two most popular microalgae on the market as a human food source (40). *Spirulina* is a nutrient-rich species of microalgae for both human and animal consumption, and is the most abundantly produced microalgae for food supplements. It has a high amino acid content, is a good source of A and B vitamins, and is a source of linolenic acid, an essential nutrient for human health (41). *Spirulina* also produces the phycocyanin commercially known as "lima blue", which is harvested for use as a colorant in foods such as soft drinks, candies, and even dairy products (20, 42). Likewise, *Chlorella* has been used to make many food products including soups, noodles, bread, cookies, ice cream, and soy sauce (43).

One of the most popular food products derived from microalgae are polyunsaturated fatty acids (PUFAs). Animals, including humans, lack the ability to synthesize PUFAs greater than 18 carbons in length and must obtain them from their diet (44). These include the omega-3 fatty acids eicosapentanoic acid (EPA), docosahexaenoic acid (DHA) and alpha-linolenic acid (ALA). A common dietary source of omega-3 fatty acids for much of the world are fish, which obtain them from consuming microalgae (45). However, many populations do not have access to a stable supply of fish, and harvesting omega-3 fatty acids from fish for use as dietary supplements has many issues and is ultimately unsustainable (46). As some microalgal species (e.g. diatoms) are particularly rich in these fatty acids, they are regarded as a viable alternative source of dietary PUFAs (20, 47). Currently, microalgal species like *Crypthecodinium cohnii*, *Schizochytrium*, and *Ulkenia* are being cultivated for their natural accumulation of omega-3 fatty acids (20), while attempts are being made to increase or alter the PUFA content of species like *Arthrospira* and *Porphyridium* for dietary supplements (20, 47). In addition, the diatom *Phaeodactylum tricornutum* naturally produces EPA and small amounts of DHA, and some groups have had success modifying this species to increase or alter its omega-3 fatty acid content (48–50). Microalgae have even been added to foods whole to enrich them with essential fatty acids and other nutrients, such as the addition of *Isochrysis galbana* biomass to biscuits (51). Microalgae are an excellent source of food and the purposeful

culturing of microalgae for food and food supplement production is only becoming more popular. Thus, it is likely that industrially farmed microalgae will become a sustainable food source and dietary staple for many parts of the world.

## 1.2.4   Biopharmaceuticals

Among the most promising applications for microalgae is the production of biopharmaceuticals (BFs) and health related products. Biopharmaceuticals are biological compounds like antibodies, hormones, and vaccines that have applications in the treatment or prevention of disease. The compounds can be naturally occurring, synthesized chemically, or derived from an engineered organism such as bacteria, yeast, mammalian or insect cells that have been genetically modified to produce the desired compound (52). For the latter, the choice of host organism depends on the type and complexity of the biomolecule, and each host comes with unique advantages and disadvantages. For example, protein BFs that require post-translational modifications (PTMs) like N-linked glycosylation are commonly produced in mammalian culture systems rather than in bacteria, as bacteria are incapable of performing these modifications. In contrast, simpler drugs like some peptide hormones can be reliably produced in bacteria much more cheaply than in mammalian cells (53). Some of the issues faced when producing BFs in mammalian cells include the risk of contamination with prions and human-infecting viruses (54, 55), and the prohibitive cost to scale up drug production (56). These issues have prompted a search for novel host organisms suitable for complex BF production, with many groups turning to microalgae as an attractive alternative platform (57–60).

The production of biopharmaceuticals in microalgae has a number of distinct advantages over traditional systems (7, 42, 61). Most microalgae species are photoautotrophs, allowing them to be cultured relatively inexpensively in specialized photobioreactors or large open ponds. Additionally, many microalgal species also have relatively fast growth rates compared to mammalian culture systems. These characteristics allow for massive scalability, a major advantage over mammalian and insect cell culture systems. Eukaryotic microalgae are capable of post-translationally modifying proteins, with a range of possible PTMs available depending on the host microalgal species (62–66). Many microalgae have GRAS (generally recognized as safe) status (67). Furthermore,

as microalgae do not require a reduced carbon source in their growth media, they don't incur the same risk of bacterial, fungal, or viral contamination as mammalian cell cultures. These qualities make downstream purification of products simpler and cheaper than other biological platforms.

Some of the earliest examples of biopharmaceutical production from microalgae were using the model green microalga *Chlamydomonas reinhardtii* (68). Since then, *Chlamydomonas* has been used to produce antibodies, growth factors, hormones, vaccines, and other valuable BFs (69–73). While *Chlamydomonas* might be the most studied and have the largest genetic toolkit for transgene expression, other microalgal species have also been investigated for BF production. In 2012, the model diatom *P. tricornutum* was successfully engineered to produce and secrete functional human antibodies against the Hepatitis B Virus surface protein (74). The microalgae *Chlorella sp.* and *D. salina* have been used to produce human growth hormone and Hepatitis B surface antigen, respectively (75, 76). While not manufactured for human use, another notable example is the production of fish growth hormone in the microlga *Nannochloropsis oculata* (77). These examples serve to highlight the potential of microalgae as hosts for the production of functional biopharmaceuticals and provide a glimpse into the future of sustainable large-scale drug production.

## 1.3   Modification and improvement of algal strains

There are many promising industrial applications for microalgae, and while some species naturally produce desirable high-value compounds (78), the success of many applications requires the modification of suitable microalgal strains. Techniques ranging from alteration of growth conditions to complex genome engineering can be used to increase growth rates or tolerance to different growth conditions, increase yields of natural products, or enable the production of desired proteins and metabolic products not natively produced by the species of interest (79). In addition to high value product biosynthesis, these techniques are also generally useful for the study of microalgal biology. Most modification strategies are applicable to a wide variety of microalgal species. Although the extent and complexity of modification techniques possible for an organism of interest depends in part on the availability of genetic tools for that species.

The earliest strategies developed for modification of microalgal strains involve altering growth media composition and growth conditions to increase or alter natural metabolite levels, or using

random mutagenesis through physical or chemical means followed by selection and screening to generate a strain with a desired trait. For example, both nitrogen limitation (80–83) and chemical mutagenesis (84, 85) have been employed to increase or alter the total fatty acid content of the model diatom *P. tricornutum* for potential applications in biofuel and cosmetics production. While these basic techniques can be simple and effective, this study will focus on targeted modification strategies.

Targeted approaches to microalgal modification involve methods such as targeted mutagenesis, gene knockout, gene overexpression and knockdown, and metabolic pathway engineering to generate novel microalgal strains. The rapid development of genome sequencing technologies and resources such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (86) have enabled targeted gene knockout and insertion, prediction and modeling of genetic pathways, and the rational design or redesign of metabolic pathways (87). Currently, genome sequences are available for several industrially-relevant microalgal species including *C. reinhardtii*, *P. tricornutum*, *Nannochloropsis gaditana*, *Chlorella sp.*, and others (79). Depending on the host species and the desired high-value product, success may require the application of one or more targeted modification strategies.

### 1.3.1  Gene expression strategies

One of the goals of modification of microalgal strains is the expression of foreign genes in a chosen organism. The ability to efficiently overexpress genes of interest in a host species is a powerful tool that allows for the production of proteins such as selective markers, metabolic enzymes, high-value peptides, and even targeted nucleases to be used as part of a genome editing strategy. Expression of foreign genes generally requires promoter and terminator elements that are recognized by the transcriptional machinery of the host species to drive transgene expression. A common strategy is to use endogenous promoters derived from genes that are abundantly expressed in the host species, such as the *C. reinhardtii* RbcS2 and HSP70A promoters (88–90), or the nitrate-inducible nitrate reductase promoters for *C. reinhardtii* and *P. tricornutum* (91–93). A related strategy, known as promoter trapping or gene trapping, functions by delivering a reporter gene or genetic selectable marker as a promoterless DNA construct that is designed to be randomly integrated into the genome

of the host organism (94). This strategy relies on the DNA construct integrating into a position downstream of an endogenous promoter that will drive expression of the reporter/marker gene. Other expression options include using promoters from a closely related species or from a virus that is known to infect the species of interest, such as the *Cylindrotheca fusiformis FcpA* promoter or the *Chaetoceros lorenzianus*-infecting DNA virus *CIP1* promoter, both of which function to drive gene expression in *P. tricornutum* (95, 96). Another consideration for achieving efficient gene expression is codon optimization. In some cases, optimizing codon usage of a transgene to match the codon usage preference of the host organism can greatly improve protein production (97, 98).

While gene overexpression is useful for many applications, some modification strategies require the ability to downregulate expression of endogenous genes in the species under study, without creating a genetic knockout. This can be accomplished using RNA interference (RNAi), a gene silencing tool which has been developed for some organisms, including microalgae species such as *P. tricornutum*, *C. reinhardtii*, and *Nannochloropsis oceanica* (99–101). Targeted RNAi knockdown is useful for species that possess functional RNAi machinery and is a way to post-transcriptionally regulate gene expression by using RNA molecules to cause the cleavage and degradation of a target mRNA transcript. CRISPR interference (CRISPRi) is a more recently developed method to downregulate gene expression and has been successfully used in microalgae like *C. reinhardtii* and *Synechococcus* (102–104). CRISPRi is a transcriptional regulation system which uses the catalytically inactive DNA binding protein dCas9 to block transcription of a target gene (105). As CRISPRi does not require the host species to possess native RNAi machinery, it is applicable to a wider range of organisms than RNAi.

## 1.3.2 Transformation and selection methods

Another requirement for microalgal modification is a delivery method to transform microalgae with a desired DNA construct. A major barrier to the delivery of DNA constructs to microalgae is the presence of the algal cell wall, which prevents passage of DNA through the cell membrane. As a result, many delivery methods require the generation of microalgal protoplasts, cell-wall deficient algal cells that must be generated using specialized proteases or polysaccharide-degrading enzymes

before transformation can be performed (106–111).

DNA constructs can either be delivered as linear DNA molecules or as circular plasmids. When transformed into a host organism, linear DNA can be transiently expressed or integrated into the host chromosome to be maintained and replicated along with the host genome. Circular plasmids can be replicating or non-replicating, the latter of which will be expressed transiently upon transformation and lost during cell division. Replicating plasmids in eukaryotic microalgae require the identification of endogenous or exogenous replication signals and centromeres recognized by the host cellular machinery that are added to a plasmid to create a stably replicating and partitioning episome. Replicating plasmids have been developed for a number of microalgal species including *C. reinhardtii*, *P. tricornutum*, and *Porphyridium purpureum* (112–114).

Both linear and circular DNA constructs can be delivered using a variety of transformation methods, several of which are generally applicable to a wide range of species. One of the simplest methods for gene delivery is by agitating algal protoplasts with glass beads in the presence of polyethylene glycol (PEG) and a DNA construct (91). Electroporation is another DNA delivery method which uses an electric pulse to open pores in the cell membrane to allow passage of DNA into the cell (115). Electroporation has been used to successfully transform several microalgae species including *P. tricornutum*, *C. reinhardtii*, *Dunaliella tertiolecta*, *Chlorella vulgaris*, *Scenedesmus obliquus*, *Neochloris oleoabundans*, and *Nannochloropsis sp.* (116–123). Biolistic transformation, also known as micro-particle bombardment, is the most-frequently used transformation method for microalgae (124–129). This method uses DNA-coated gold or tungsten microparticles fired at high speed into the target organism by pressurized gasses (e.g. helium). The DNA then elutes from the microparticle and either expresses transiently, becomes integrated into the chromosome, or replicates autonomously as an episome, depending on the construct design. Biolistic transformation can also be used to deliver purified proteins to the microalgae of interest in a similar fashion, as was demonstrated in *P. tricornutum* (130). A common plant transformation method which uses the soil bacterium *Agrobacterium tumefaciens* to deliver DNA to plant cells has also been successfully adapted for use in some microalgae species including *C. reinhardtii*, *Haematococcus pluvialis*, *Schizochytrium sp.*, and *Isochrysis sp.* (131–135). Additionally, a recent transformation method has been developed for the marine diatoms *P. tricornutum* and *Thalassiosira pseudonana* which uses conjugation from bacteria such as *Escherichia coli* to maintain

stably replicating episomal plasmids in the nucleus of the transformed microalga (113, 136).

One consideration common to nearly all microalgal transformation methods is the need for a selective marker to select for successful transformants. The most common selectable markers used for microalgae are antibiotic and herbicide resistance genes. Examples of popular antibiotic resistance markers for microalgae transformation are the zeocin resistance gene *ble*, and the neomycin resistance gene *nptII* (137, 138). Common herbicide resistance markers include the glyphosate and phytoene resistance genes, *GAT* and *pds* (137, 138). Another type of selectable marker that has seen use in microalgae are those based on metabolic genes. A microalgae that lacks, or is engineered to be deficient in, a key metabolic enzyme in an essential biosynthetic pathway can be transformed with the gene encoding that enzyme to complement the deficiency and be selected for on media lacking that nutrient. Some examples in microalgae are the *ARG7* gene, used to transform a *C. reinhardtii* arginine auxotroph, and *NIT1* (nitrate reductase), used to transform a variety of microalgal *NIT1* knockout strains (91, 124, 126, 139, 140).

### 1.3.3 Gene-editing technologies

The advent of gene-editing technologies, namely targeted nucleases allowing precise gene editing, is one of the most potent enablers of microalgal modification and opens up a wide array of industrial applications for genetically engineered microalgae. Engineered nucleases that have been successfully adapted for gene editing in microalgae include zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) (141–148). ZFNs and TALENs are fusions of the FokI restriction endonuclease with a zinc finger or transcription activator-like effector DNA-binding platform (149, 150). These DNA-binding platforms are comprised of repeating protein subunits that rely on protein-DNA contacts for targeting, making them difficult to construct. In addition, FokI functions as a dimer, requiring a pair of ZFNs or TALENs positioned on either strand of the target site to produce a double-strand break (DSB). In contrast, the CRISPR/Cas9 system is an RNA-guided DNA endonuclease that functions as a monomer to produce DSBs (151). Targeting is determined by RNA-DNA hybridization, where an expressed single guide RNA (sgRNA) complexes with the Cas9 protein and directs it to cleave DNA at the site of

DNA-binding which must be immediately upstream of short DNA sequence called a protospacer adjacent motif (PAM) for cleavage. Retargeting Cas9 simply requires expressing a sgRNA that hybridizes specifically to the target site of interest. In addition, multiple sgRNAs can be expressed simultaneously, allowing multiple editing events at different locations in the genome (152).

Metabolic engineering of a microalgal strain is often required to improve the production levels of the desired protein or chemical compound, and gene-editing tools are a simple and effective way to achieve this. Engineered nucleases like ZFNs, TALENs, and CRISPR/Cas9 can be used to stimulate a range of genetic modifications including gene disruption, deletion, insertion, and replacement (153). To achieve this, engineered nucleases are delivered to algal cells, either encoded on DNA constructs (154), or as assembled proteins (130). Once in the nucleus, these reagents can cleave DNA at a targeted site, generating a DSB that is repaired using the hosts endogenous DSB-repair machinery, which can be exploited for genome engineering purposes (Fig. 1.3). The most commonly exploited DSB-repair mechanism is the error-prone non-homologous end joining (NHEJ) repair pathway, which functions to ligate together blunt DNA ends. During this repair process, end-processing enzymes may add or remove bases from the free DNA ends, resulting in the deletion or insertion of short sequences of DNA (INDELs) at the cleavage site. This can be exploited to generate gene knockouts by introducing frameshift mutations, or by deleting residues critical for structure or function of a protein of interest. It can also be used to create knock-ins by using NHEJ to ligate a linear DNA construct into the DSB site.

Homologous recombination (HR) is another DSB-repair pathway that can be exploited for genome editing purposes. The HR repair mechanism uses a DNA sequence homologous to the region of the DSB to repair the break in an error-free manner. This function can be exploited to delete, insert or replace a chosen DNA sequence. This is achieved by co-transforming the engineered nuclease with a DNA construct, flanked by sequences of DNA homologous to the site of the DSB. The flanking DNA targets the DNA construct to be recombined into the site of the DSB, and can be designed to disrupt or replace the target DNA if desired. Unfortunately, this method has proven to be difficult as some microalgal species display very low frequencies of homologous recombination (155–158). Efficient homologous recombination has only been observed in a small number of algal species, including *P. tricornutum*, *Cyanidioschyzon merolae* and *Nannochloropsis sp.* (123, 155, 159). However, progress is being made towards improving the rates of HR in
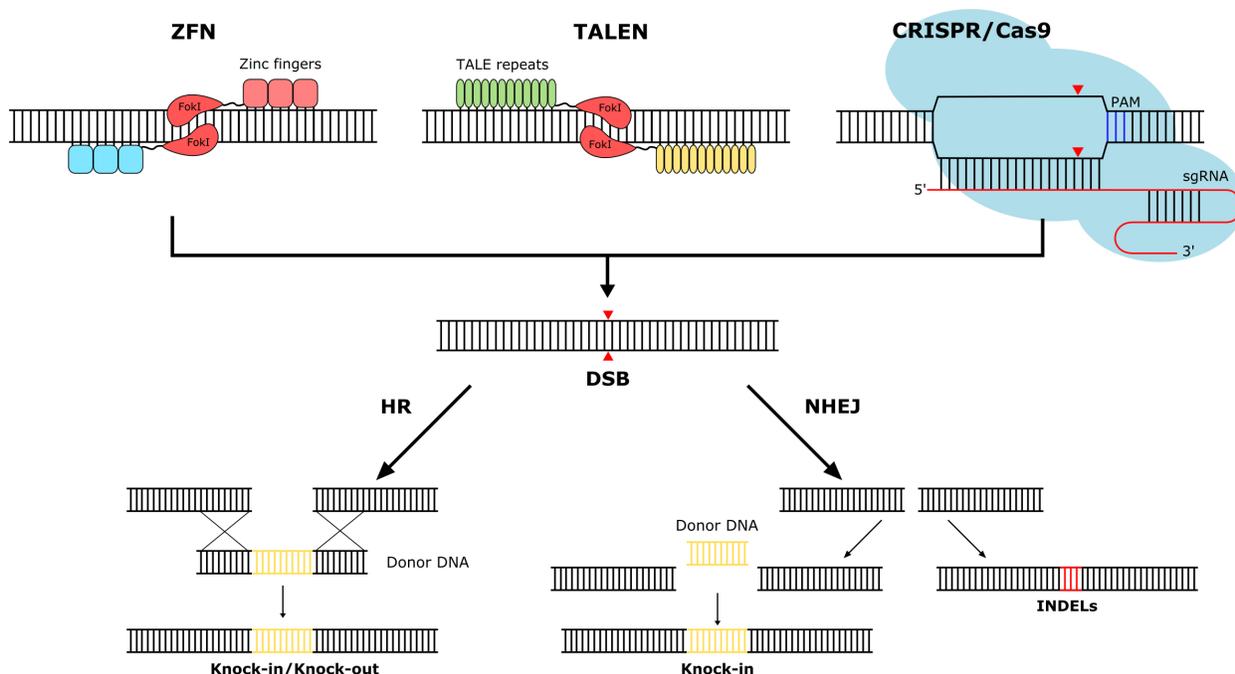
**Figure 1.3: Genome modification strategies using engineered nucleases.** A Zinc finger nuclease (ZFN), transcription activator-like effector nuclease (TALEN), or Cas9 nuclease can be used to generate a targeted DNA double-strand break (DSB). INDELs can be produced when the DSB is repaired by error-prone non-homologous end joining (NHEJ). Repair by homologous recombination (HR) can remove or replace a sequence of DNA to produce knock-outs. Knock-in events can be produced by either NHEJ or HR repair when donor DNA (yellow) is provided.

certain microalgae using strategies like knockdown of NHEJ-pathway proteins to bias repair events towards HR (160).

## 1.4 *Phaeodactylum tricornutum* is an attractive species for synthetic biology applications

The model marine diatom *P. tricornutum* is a photoautotrophic eukaryotic microalga. It is the second diatom species to have its genome fully sequenced and is one of the most well-studied diatoms to date. It has been observed to exhibit three morphotypes: oval, fusiform, and triradiate (Fig. 1.4), with fusiform being the predominant form (161). It has a genome size of ~27.4 Mb with over 10,000 predicted genes (162). Like other closely-related diatoms, *P. tricornutum* has the rare ability to metabolize silicates, which it incorporates into its silicified cell wall. Unlike most other diatom species, it is able to be grown in media lacking silicon (161).

Synthetic biology is the design or redesign of organisms or biological systems for new pur-
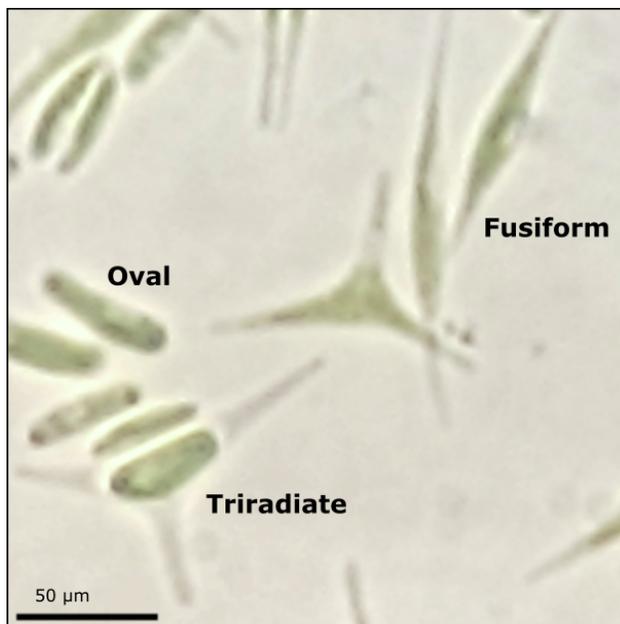
**Figure 1.4: The three morphologies of *Phaeodactylum tricornutum*.** *P. tricornutum* has been observed in three morphotypes: oval, fusiform, and triradiate.

poses (163). *P. tricornutum* has a number of characteristics that make it attractive as a platform for synthetic biology applications including the industrial production of value-added products. As a photoautotrophic organism, *P. tricornutum* is simple and inexpensive to grow, even at industrial scales, and has a relatively fast growth rate. It is a marine microalga and can be grown in filtered seawater. Thus, it does not require a fresh water supply for cultivation. *P. tricornutum* can accumulate high levels of fatty acids (164), making it of particular interest for biofuel production. It is also an abundant source of the PUFA EPA, though not at a level to be commercially viable without modification to improve yields (48). It has been demonstrated that *P. tricornutum* can perform protein post-translational modifications that result in functional proteins that resemble those made in mammalian cell systems (58, 165). Specifically, *P. tricornutum* possesses a glycosylation pathway that is very similar to the pathway found in human cells (62), including an important Fucosyltransferase (FuT) required to produce fucosylated N-Glycans (63). This is signficant as this enzyme is not present in other popular bioproduction species such as *Saccharomycaes cerevisae*. While *P. tricornutum* lacks two enzymes, present in humans, that are involved in the synthesis of core N-linked glycosylation events (ALG10 and MOGS), the N-glycan product of this pathway matches that produced in human cells (62).

In addition to the innate characteristics, *P. tricornutum* also has a currently limited but useful

foundation of genetic tools and resources that enable strain engineering. There are a number of promoter elements characterized for this species, including the *FcpA* and *FcpB* promoters, as well as the inducible nitrate reductase promoter (92, 93, 128, 166). It has well-established transformation methods, including electroporation and microparticle bombardment, as well as a number of reporter genes and selectable markers (93, 97, 116, 128, 167–169). A novel transformation method has been developed for *P. tricornutum* which uses conjugation with *E. coli* to establish stably replicating nuclear episomes, allowing stable gene expression without relying on genomic integration (113). The full genome sequence (162) and expressed sequence tag (EST) database (170, 171) enables the use of genome editing and manipulation tools such as TALENs, CRISPR/Cas9, and RNAi (100, 130, 144, 155, 172, 173). For these reasons *P. tricornutum* has been the platform of choice for various high-value product biosynthesis efforts such as the production of PUFAs, mono- and triterpenoids, and human antibodies (49, 58, 165, 174–178).

While it is the most well-studied diatom, it suffers from a lack of genetic tools that prevent it from becoming an ideal platform for synthetic biology applications. The few promoters available for transgene expression are inadequate for the design and construction of complex metabolic pathways. The need to use expensive antibiotics to maintain selection of transformants significantly hinders its use as an industrial-scale production platform. While gene-editing technologies have been adapted for use in *P. tricornutum*, delivery relies on biolistic transformation of expression constructs that often integrate randomly into the genome (144, 155), making it difficult to remove the engineered nuclease and causing unknown negative effects on genomic stability. Thus, there is a need to develop more sophisticated gene editing strategies for this organism, and to generate a wider range of genetic tools to allow for more complex metabolic engineering in *P. tricornutum*.

## 1.5   Scope of the thesis: Genetic tools to enable synthetic biology applications in *P. tricornutum*

This thesis focusses on the study and manipulation of *P. tricornutum* with the goal to generate novel genetic tools and strains, and to characterize *P. tricornutum* biology to aid in the development of this species for synthetic biology applications. The model diatom *P. tricornutum* has the potential

to be a powerful platform for production of high-value compounds such as food products and functional biopharmaceuticals due to its attractive natural characteristics and a solid foundation of genetic tools and laboratory methods available for this species. At the outset of my doctoral work, I hypothesized that *P. tricornutum* could be genetically modified to produce such high-value compounds. However, for this to be accomplished, *P. tricornutum* required further genetic and biotechnological development. Using the recently developed conjugation-based transformation method, I was able to rapidly generate and characterize a variety of genetic tools for modification of this organism, and subsequently engineer novel strains of *P. tricornutum* for industrial applications and the study of diatom biology.

My first goal was to expand the *P. tricornutum* genetic toolbox to enable the design and construction of more complex synthetic pathways that could be used to produce industrially-relevant products. As described in chapter two, I characterized a number of endogenous promoters, terminators, and introns that allow for the expression of exogenous genes in *P. tricornutum*. Using these tools I was able to construct a replicating plasmid-based CRISPR/Cas9 gene editing tool that I validated by generating a genetic knockout of the *P. tricornutum* urease gene. These tools were further used to construct a proof-of-principle vanillin biosynthesis pathway that I demonstrated could be maintained in *P. tricornutum* over several generations. In chapter 3, I created auxotrophic strains of *P. tricornutum* by using the plasmid-based CRISPR/Cas9 tool (Chapter 2) to knock out essential genes in the histidine and uracil metabolic pathways, and demonstrated that the auxotrophic phenotypes could be complemented with plasmid-encoded copies of the intact genes. I also provided evidence of large deletions and loss-of-heterozygosity stimulated by CRISPR/Cas9 editing using Nanopore long-read sequencing. In chapter 4, I successfully engineered a strain of *P. tricornutum* that produces the SARS-CoV-2 spike antigen. This strain was developed with the goal of addressing the need for population-wide serological testing. By developing these tools, I hope to transform *P. tricornutum* into a powerful platform for synthetic biology applications and further the study of diatom biology.

# 1.6 References

[1] Keeling, P. J.; Burger, G.; Durnford, D. G.; Lang, B. F.; Lee, R. W.; Pearlman, R. E.; Roger, A. J.; Gray, M. W. The tree of eukaryotes. *Trends in Ecology & Evolution* **2005**, *20*, 670–676.

[2] Burki, F.; Shalchian-Tabrizi, K.; Pawlowski, J. Phylogenomics reveals a new 'mega-group'including most photosynthetic eukaryotes. *Biology Letters* **2008**, *4*, 366–369.

[3] Burki, F.; Kaplan, M.; Tikhonenkov, D. V.; Zlatogursky, V.; Minh, B. Q.; Radaykina, L. V.; Smirnov, A.; Mylnikov, A. P.; Keeling, P. J. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proceedings of the Royal Society B: Biological Sciences* **2016**, *283*, 2015–2802.

[4] Reyes-Prieto, A.; Hackett, J. D.; Soares, M. B.; Bonaldo, M. F.; Bhattacharya, D. Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Current Biology* **2006**, *16*, 2320–2325.

[5] Armbrust, E. V. The life of diatoms in the world's oceans. *Nature* **2009**, *459*, 185–192.

[6] Larkum, A. W.; Ross, I. L.; Kruse, O.; Hankamer, B. Selection, breeding and engineering of microalgae for bioenergy and biofuel production. *Trends in Biotechnology* **2012**, *30*, 198–205.

[7] Walker, T. L.; Purton, S.; Becker, D. K.; Collet, C. Microalgae as bioreactors. *Plant Cell Reports* **2005**, *24*, 629–641.

[8] Burkholder, J. M.; Glibert, P. M.; Skelton, H. M. Mixotrophy, a major mode of nutrition for harmful algal species in eutrophic waters. *Harmful Algae* **2008**, *8*, 77–93.

[9] Williams, P. J. l. B.; Laurens, L. M. Microalgae as biodiesel & biomass feedstocks: review & analysis of the biochemistry, energetics & economics. *Energy & Environmental Science* **2010**, *3*, 554–590.

[10] Chapman, R. L. Algae: the world's most important "plants"—an introduction. *Mitigation and Adaptation Strategies for Global Change* **2013**, *18*, 5–12.

[11] Amin, S. A.; Parker, M. S.; Armbrust, E. V. Interactions between diatoms and bacteria. *Microbiology and Molecular Biology Reviews* **2012**, *76*, 667–684.

[12] Khan, M. I.; Shin, J. H.; Kim, J. D. The promising future of microalgae: current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. *Microbial Cell Factories* **2018**, *17*, 36.

[13] Milano, J.; Ong, H. C.; Masjuki, H.; Chong, W.; Lam, M. K.; Loh, P. K.; Vellayan, V. Microalgae biofuels as an alternative to fossil fuel for power generation. *Renewable and Sustainable Energy Reviews* **2016**, *58*, 180–197.

[14] Chisti, Y. Biodiesel from microalgae. *Biotechnology Advances* **2007**, *25*, 294–306.

[15] Ono, E.; Cuello, J. Feasibility assessment of microalgal carbon dioxide sequestration technology with photobioreactor and solar collector. *Biosystems Engineering* **2006**, *95*, 597–606.

[16] Pierobon, S.; Cheng, X.; Graham, P.; Nguyen, B.; Karakolis, E.; Sinton, D. Emerging microalgae technology: a review. *Sustainable Energy & Fuels* **2018**, *2*, 13–38.

[17] Pulz, O.; Gross, W. Valuable products from biotechnology of microalgae. *Applied Microbiology and Biotechnology* **2004**, *65*, 635–648.

[18] Koller, M.; Muhr, A.; Braunegg, G. Microalgae as versatile cellular factories for valued products. *Algal Research* **2014**, *6*, 52–63.

[19] Borowitzka, M. A. High-value products from microalgae—their development and commercialisation. *Journal of Applied Phycology* **2013**, *25*, 743–756.

[20] Spolaore, P.; Joannis-Cassan, C.; Duran, E.; Isambert, A. Commercial applications of microalgae. *Journal of Bioscience and Bioengineering* **2006**, *101*, 87–96.

[21] Priyadarshani, I.; Rath, B. Commercial and industrial applications of micro algae–A review. *Journal of Algal Biomass Utilization* **2012**, *3*, 89–100.

[22] Naik, S. N.; Goud, V. V.; Rout, P. K.; Dalai, A. K. Production of first and second generation biofuels: a comprehensive review. *Renewable and Sustainable Energy Reviews* **2010**, *14*, 578–597.

[23] Carriquiry, M. A.; Du, X.; Timilsina, G. R. Second generation biofuels: Economics and policies. *Energy Policy* **2011**, *39*, 4222–4234.

[24] Brennan, L.; Owende, P. Biofuels from microalgae—a review of technologies for production, processing, and extractions of biofuels and co-products. *Renewable and Sustainable Energy Reviews* **2010**, *14*, 557–577.

[25] Raut, N.; Al-Balushi, T.; Panwar, S.; Vaidya, R.; Shinde, G. Microalgal biofuel. *Biofuels-Status and Perspective* **2015**, *6*, 101–140.

[26] Wijffels, R. H.; Barbosa, M. J. An outlook on microalgal biofuels. *Science* **2010**, *329*, 796–799.

[27] Schenk, P. M.; Thomas-Hall, S. R.; Stephens, E.; Marx, U. C.; Mussgnug, J. H.; Posten, C.; Kruse, O.; Hankamer, B. Second generation biofuels: high-efficiency microalgae for biodiesel production. *Bioenergy Research* **2008**, *1*, 20–43.

[28] Markou, G.; Nerantzis, E. Microalgae for high-value compounds and biofuels production: a review with focus on cultivation under stress conditions. *Biotechnology Advances* **2013**, *31*, 1532–1542.

[29] Singh, J.; Gu, S. Commercialization potential of microalgae for biofuels production. *Renewable and Sustainable Energy Reviews* **2010**, *14*, 2596–2610.

[30] Wang, H.-M. D.; Chen, C.-C.; Huynh, P.; Chang, J.-S. Exploring the potential of using algae in cosmetics. *Bioresource Technology* **2015**, *184*, 355–362.

[31] Eriksen, N. T. Production of phycocyanin—a pigment with applications in biology, biotechnology, foods and medicine. *Applied Microbiology and Biotechnology* **2008**, *80*, 1–14.

[32] Sekar, S.; Chandramohan, M. Phycobiliproteins as a commodity: trends in applied research, patents and commercialization. *Journal of Applied Phycology* **2008**, *20*, 113–136.

[33] Chrapusta, E.; Kaminski, A.; Duchnik, K.; Bober, B.; Adamski, M.; Bialczyk, J. Mycosporine-like amino acids: Potential health and beauty ingredients. *Marine Drugs* **2017**, *15*, 326.

[34] Schmid, D.; Schürch, C.; Zülli, F. Mycosporine-like amino acids from red algae protect against premature skin-aging. *Euro Cosmetics* **2006**, *9*, 1–4.

[35] Mourelle, M. L.; Gómez, C. P.; Legido, J. L. The potential use of marine microalgae and cyanobacteria in cosmetics and thalassotherapy. *Cosmetics* **2017**, *4*, 46.

[36] Wells, M. L.; Potin, P.; Craigie, J. S.; Raven, J. A.; Merchant, S. S.; Helliwell, K. E.; Smith, A. G.; Camire, M. E.; Brawley, S. H. Algae as nutritional and functional food sources: revisiting our understanding. *Journal of Applied Phycology* **2017**, *29*, 949–982.

[37] Mumford Jr, T. Porphyra as food: cultivation and economics. *Algae and Human Affairs* **1988**,

[38] Milledge, J. J. Commercial application of microalgae other than as biofuels: a brief review. *Reviews in Environmental Science and Bio/Technology* **2011**, *10*, 31–41.

[39] Gao, K. Chinese studies on the edible blue-green alga, Nostoc flagelliforme: a review. *Journal of Applied Phycology* **1998**, *10*, 37–49.

[40] Andrade, L.; Andrade, C.; Dias, M.; Nascimento, C.; Mendes, M. Chlorella and Spirulina Microalgae as Sources of Functional Foods. *Nutraceuticals, and Food Supplements* **2018**, 45–58.

[41] Vonshak, A.; Richmond, A. Mass production of the blue-green alga Spirulina: an overview. *Biomass* **1988**, *15*, 233–247.

[42] Raja, R.; Hemaiswarya, S.; Kumar, N. A.; Sridhar, S.; Rengasamy, R. A perspective on the biotechnological potential of microalgae. *Critical Reviews in Microbiology* **2008**, *34*, 77–88.

[43] Sathasivam, R.; Radhakrishnan, R.; Hashem, A.; Abd_Allah, E. F. Microalgae metabolites: A rich source for food and medicine. *Saudi Journal of Biological Sciences* **2019**, *26*, 709–722.

[44] Simopoulos, A. Human requirement for N-3 polyunsaturated fatty acids. *Poultry Science* **2000**, *79*, 961–970.

[45] Ryckebosch, E.; Bruneel, C.; Muylaert, K.; Foubert, I. Microalgae as an alternative source of omega-3 long chain polyunsaturated fatty acids. *Lipid Technology* **2012**, *24*, 128–130.

[46] Ji, X.-J.; Ren, L.-J.; Huang, H. Omega-3 biotechnology: a green and sustainable process for omega-3 fatty acids production. *Frontiers in Bioengineering and Biotechnology* **2015**, *3*, 158.

[47] Borowitzka, M. Biotechnological & environmental applications of microalgae. *Biotechnological & Environmental Applications of Microalgae.[Online] Murdoch University* **2006**,

[48] Yongmanitchai, W.; Ward, O. P. Growth of and omega-3 fatty acid production by Phaeodactylum tricornutum under different culture conditions. *Applied and Environmental Microbiology* **1991**, *57*, 419–425.

[49] Hamilton, M. L.; Haslam, R. P.; Napier, J. A.; Sayanova, O. Metabolic engineering of Phaeodactylum tricornutum for the enhanced accumulation of omega-3 long chain polyunsaturated fatty acids. *Metabolic Engineering* **2014**, *22*, 3–9.

[50] Garcí, M. C.; Sevilla, J. F.; Fernández, F. A.; Grima, E. M.; Camacho, F. G. Mixotrophic growth of Phaeodactylum tricornutum on glycerol: growth rate and fatty acid profile. *Journal of Applied Phycology* **2000**, *12*, 239–248.

[51] Gouveia, L.; Coutinho, C.; Mendonça, E.; Batista, A. P.; Sousa, I.; Bandarra, N. M.; Raymundo, A. Functional biscuits with PUFA-$\omega$3 from Isochrysis galbana. *Journal of the Science of Food and Agriculture* **2008**, *88*, 891–896.

[52] Berlec, A.; Štrukelj, B. Current state and recent advances in biopharmaceutical production in Escherichia coli, yeasts and mammalian cells. *Journal of Industrial Microbiology & Biotechnology* **2013**, *40*, 257–274.

[53] Moks, T.; Abrahmsen, L.; Holmgren, E.; Bilich, M.; Olsson, A.; Pohl, G.; Sterky, C.; Hultberg, H.; Josephson, S. Expression of human insulin-like growth factor I in bacteria: use of optimized gene fusion vectors to facilitate protein purification. *Biochemistry* **1987**, *26*, 5239–5244.

[54] Scott, M. R.; Will, R.; Ironside, J.; Nguyen, H.-O. B.; Tremblay, P.; DeArmond, S. J.; Prusiner, S. B. Compelling transgenetic evidence for transmission of bovine spongiform encephalopathy prions to humans. *Proceedings of the National Academy of Sciences* **1999**, *96*, 15137–15142.

[55] Jayme, D. W.; Smith, S. R. Media formulation options and manufacturing process controls to safeguard against introduction of animal origin contaminants in animal cell culture. *Cytotechnology* **2000**, *33*, 27–36.

[56] Chu, L.; Robinson, D. K. Industrial choices for protein production by large-scale cell culture. *Current Opinion in Biotechnology* **2001**, *12*, 180–187.

[57] Borowitzka, M. A. Microalgae as sources of pharmaceuticals and other biologically active compounds. *Journal of Applied Phycology* **1995**, *7*, 3–15.

[58] Hempel, F.; Lau, J.; Klingl, A.; Maier, U. G. Algae as protein factories: expression of a human antibody and the respective antigen in the diatom Phaeodactylum tricornutum. *PLoS ONE* **2011**, *6*, e28424.

[59] Schwartz, R. E.; Hirsch, C. F.; Sesin, D. F.; Flor, J. E.; Chartrain, M.; Fromtling, R. E.; Harris, G. H.; Salvatore, M. J.; Liesch, J. M.; Yudin, K. Pharmaceuticals from cultured algae. *Journal of Industrial Microbiology* **1990**, *5*, 113–123.

[60] Cannell, R. J. Algae as a source of biologically active products. *Pesticide Science* **1993**, *39*, 147–153.

[61] Potvin, G.; Zhang, Z. Strategies for high-level recombinant protein expression in transgenic microalgae: a review. *Biotechnology Advances* **2010**, *28*, 910–918.

[62] Mathieu-Rivet, E.; Kiefer-Meyer, M.-C.; Vanier, G.; Ovide, C.; Burel, C.; Lerouge, P.; Bardor, M. Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Frontiers in Plant Science* **2014**, *5*, 359.

[63] Baïet, B.; Burel, C.; Saint-Jean, B.; Louvet, R.; Menu-Bouaouiche, L.; Kiefer-Meyer, M.-C.; Mathieu-Rivet, E.; Lefebvre, T.; Castel, H.; Carlier, A., et al. N-glycans of Phaeodactylum tricornutum diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme. *Journal of Biological Chemistry* **2011**, *286*, 6152–6164.

[64] Levy-Ontman, O.; Arad, S.; Harvey, D. J.; Parsons, T. B.; Fairbanks, A.; Tekoah, Y. Unique N-glycan moieties of the 66-kDa cell wall glycoprotein from the red microalga Porphyridium sp. *Journal of Biological Chemistry* **2011**, *286*, 21340–21352.

[65] Mathieu-Rivet, E.; Scholz, M.; Arias, C.; Dardelle, F.; Schulze, S.; Le Mauff, F.; Teo, G.; Hochmal, A. K.; Blanco-Rivero, A.; Loutelier-Bourhis, C., et al. Exploring the N-glycosylation pathway in Chlamydomonas reinhardtii unravels novel complex structures. *Molecular & Cellular Proteomics* **2013**, *12*, 3160–3183.

[66] Levy-Ontman, O.; Fisher, M.; Shotland, Y.; Weinstein, Y.; Tekoah, Y.; Arad, S. M. Genes involved in the endoplasmic reticulum N-glycosylation pathway of the red microalga Porphyridium sp.: a bioinformatic study. *International Journal of Molecular Sciences* **2014**, *15*, 2305–2326.

[67] Jha, D.; Jain, V.; Sharma, B.; Kant, A.; Garlapati, V. K. Microalgae-based Pharmaceuticals and Nutraceuticals: An Emerging Field with Immense Market Potential. *ChemBioEng Reviews* **2017**, *4*, 257–272.

[68] Bañuelos-Hernández, B.; Beltrán-López, J. I.; Rosales-Mendoza, S. *Handbook of Marine Microalgae*; Elsevier, 2015; pp 281–298.

[69] Specht, E. A.; Mayfield, S. P. Algae-based oral recombinant vaccines. *Frontiers in Microbiology* **2014**, *5*, 60.

[70] Rosales-Mendoza, S. Future directions for the development of Chlamydomonas-based vaccines. *Expert Review of Vaccines* **2013**, *12*, 1011–1019.

[71] Rasala, B. A.; Mayfield, S. P. Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. *Photosynthesis Research* **2015**, *123*, 227–239.

[72] Barolo, L.; Abbriano, R. M.; Commault, A. S.; George, J.; Kahlke, T.; Fabris, M.; Padula, M. P.; Lopez, A.; Ralph, P. J.; Pernice, M. Perspectives for Glyco-Engineering of Recombinant Biopharmaceuticals from Microalgae. *Cells* **2020**, *9*, 633.

[73] Specht, E.; Miyake-Stoner, S.; Mayfield, S. Micro-algae come of age as a platform for recombinant protein production. *Biotechnology Letters* **2010**, *32*, 1373–1383.

[74] Hempel, F.; Maier, U. G. An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. *Microbial Cell Factories* **2012**, *11*, 1–6.

[75] Hawkins, R. L.; Nakamura, M. Expression of human growth hormone by the eukaryotic alga, Chlorella. *Current Microbiology* **1999**, *38*, 335–341.

[76] Geng, D.; Wang, Y.; Wang, P.; Li, W.; Sun, Y. Stable expression of hepatitis B surface antigen gene in Dunaliella salina (Chlorophyta). *Journal of Applied Phycology* **2003**, *15*, 451–456.

[77] Chen, H. L.; Li, S. S.; Huang, R.; Tsai, H.-J. Conditional production of a functional fish growth hormone in the transgenic line of nannochloropsis oculata (Eustigmatophyceae) 1. *Journal of Phycology* **2008**, *44*, 768–776.

[78] Sasso, S.; Pohnert, G.; Lohr, M.; Mittag, M.; Hertweck, C. Microalgae in the postgenomic era: a blooming reservoir for new natural products. *FEMS Microbiology Reviews* **2012**, *36*, 761–785.

[79] Fu, W.; Chaiboonchoe, A.; Khraiwesh, B.; Nelson, D. R.; Al-Khairy, D.; Mystikou, A.; Alzahmi, A.; Salehi-Ashtiani, K. Algal cell factories: approaches, applications, and potentials. *Marine Drugs* **2016**, *14*, 225.

[80] Alipanah, L.; Rohloff, J.; Winge, P.; Bones, A. M.; Brembu, T. Whole-cell response to nitrogen deprivation in the diatom Phaeodactylum tricornutum. *Journal of Experimental Botany* **2015**, *66*, 6281–6296.

[81] Wang, X.; Fosse, H. K.; Li, K.; Chauton, M. S.; Vadstein, O.; Reitan, K. I. Influence of nitrogen limitation on lipid accumulation and EPA and DHA content in four marine microalgae for possible use in aquafeed. *Frontiers in Marine Science* **2019**, *6*, 95.

[82] Yodsuwan, N.; Sawayama, S.; Sirisansaneeyakul, S. Effect of nitrogen concentration on growth, lipid production and fatty acid profiles of the marine diatom Phaeodactylum tricornutum. *Agriculture and Natural Resources* **2017**, *51*, 190–197.

[83] Rodolfi, L.; Biondi, N.; Guccione, A.; Bassi, N.; D'Ottavio, M.; Arganaraz, G.; Tredici, M. R. Oil and eicosapentaenoic acid production by the diatom Phaeodactylum tricornutum cultivated outdoors in Green Wall Panel (GWP®) reactors. *Biotechnology and Bioengineering* **2017**, *114*, 2204–2210.

[84] Yi, Z.; Su, Y.; Xu, M.; Bergmann, A.; Ingthorsson, S.; Rolfsson, O.; Salehi-Ashtiani, K.; Brynjolfsson, S.; Fu, W. Chemical mutagenesis and fluorescence-based high-throughput screening for enhanced accumulation of carotenoids in a model marine diatom Phaeodactylum tricornutum. *Marine Drugs* **2018**, *16*.

[85] Kaur, S. Genetic and biotechnological development of the pennate marine diatom Phaeodactylum tricornutum for high-value bioproducts and carbon bio-mitigation. Ph.D. thesis, 2014.

[86] Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **2000**, *28*, 27–30.

[87] Salehi-Ashtiani, K.; Koussa, J.; Dohai, B. S.; Chaiboonchoe, A.; Cai, H.; Dougherty, K. A.; Nelson, D. R.; Jijakli, K.; Khraiwesh, B. *Biomass and biofuels from microalgae*; Springer, 2015; pp 173–189.

[88] Stevens, D. R.; Purton, S.; Rochaix, J.-D. The bacterial phleomycin resistance geneble as a dominant selectable marker inChlamydomonas. *Molecular and General Genetics MGG* **1996**, *251*, 23–30.

[89] Lumbreras, V.; Stevens, D. R.; Purton, S. Efficient foreign gene expression in Chlamydomonas reinhardtii mediated by an endogenous intron. *Plant Journal* **1998**, *14*, 441–447.

[90] Schroda, M.; Blöcker, D.; Beck, C. F. The HSP70A promoter as a tool for the improved expression of transgenes in Chlamydomonas. *The Plant Journal* **2000**, *21*, 121–131.

[91] Kindle, K. L.; Schnell, R. A.; Fernández, E.; Lefebvre, P. A. Stable nuclear transformation of Chlamydomonas using the Chlamydomonas gene for nitrate reductase. *The Journal of Cell Biology* **1989**, *109*, 2589–2601.

[92] Chu, L.; Ewe, D.; Bártulos, C. R.; Kroth, P. G.; Gruber, A. Rapid induction of GFP expression by the nitrate reductase promoter in the diatom Phaeodactylum tricornutum. *PeerJ* **2016**, *4*, e2344.

[93] Niu, Y. F.; Yang, Z. K.; Zhang, M. H.; Zhu, C. C.; Yang, W. D.; Liu, J. S.; Li, H. Y. Transformation of diatom Phaeodactylum tricornutum by electroporation and establishment of inducible selection marker. *BioTechniques* **2012**, *52*, 1–3.

[94] Vila, M.; Díaz-Santos, E.; De la Vega, M.; Rodríguez, H.; Vargas, Á.; León, R. Promoter trapping in microalgae using the antibiotic paromomycin as selective agent. *Marine Drugs* **2012**, *10*, 2749–2765.

[95] Miyagawa, A.; Okami, T.; Kira, N.; Yamaguchi, H.; Ohnishi, K.; Adachi, M. Research note: High efficiency transformation of the diatom Phaeodactylum tricornutum with a promoter from the diatom Cylindrotheca fusiformis. *Phycological Research* **2009**, *57*, 142–146.

[96] Kadono, T.; Miyagawa-Yamaguchi, A.; Kira, N.; Tomaru, Y.; Okami, T.; Yoshimatsu, T.; Hou, L.; Ohama, T.; Fukunaga, K.; Okauchi, M.; Yamaguchi, H.; Ohnishi, K.; Falciatore, A.; Adachi, M. Characterization of marine diatom-infecting virus promoters in the model diatom Phaeodactylum tricornutum. *Scientific Reports* **2015**, *5*, 1–13.

[97] Zaslavskaia, L. A.; Casey Lippmeier, J.; Kroth, P. G.; Grossman, A. R.; Apt, K. E. Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a variety of selectable marker and reporter genes. *Journal of Phycology* **2000**, *36*, 379–386.

[98] Franklin, S.; Ngo, B.; Efuet, E.; Mayfield, S. P. Development of a GFP reporter gene for Chlamydomonas reinhardtii chloroplast. *The Plant Journal* **2002**, *30*, 733–744.

[99] Wei, L.; Xin, Y.; Wang, Q.; Yang, J.; Hu, H.; Xu, J. RNA i-based targeted gene knockdown in the model oleaginous microalgae Nannochloropsis oceanica. *The Plant Journal* **2017**, *89*, 1236–1250.

[100] De Riso, V.; Raniello, R.; Maumus, F.; Rogato, A.; Bowler, C.; Falciatore, A. Gene silencing in the marine diatom Phaeodactylum tricornutum. *Nucleic Acids Research* **2009**, *37*.

[101] Kim, E.-J.; Cerutti, H. *Methods in Cell Biology*; Elsevier, 2009; Vol. 93; pp 99–110.

[102] Kao, P.-H.; Ng, I.-S. CRISPRi mediated phosphoenolpyruvate carboxylase regulation to enhance the production of lipid in Chlamydomonas reinhardtii. *Bioresource Technology* **2017**, *245*, 1527–1537.

[103] Gordon, G. C.; Korosh, T. C.; Cameron, J. C.; Markley, A. L.; Begemann, M. B.; Pfleger, B. F. CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium Synechococcus sp. strain PCC 7002. *Metabolic Engineering* **2016**, *38*, 170–179.

[104] Huang, C.-H.; Shen, C. R.; Li, H.; Sung, L.-Y.; Wu, M.-Y.; Hu, Y.-C. CRISPR interference (CRISPRi) for gene regulation and succinate production in cyanobacterium S. elongatus PCC 7942. *Microbial Cell Factories* **2016**, *15*, 1–11.

[105] Larson, M. H.; Gilbert, L. A.; Wang, X.; Lim, W. A.; Weissman, J. S.; Qi, L. S. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols* **2013**, *8*, 2180–2196.

[106] JAENICKE, L.; KUHNE, W.; SPESSERT, R.; WAHLE, U.; WAFFENSCHMIDT, S. Cell-wall lytic enzymes (autolysins) of Chlamydomonas reinhardtii are (hydroxy) proline-specific proteases. *European Journal of Biochemistry* **1987**, *170*, 485–491.

[107] Imam, S. H.; Snell, W. J. Degradation of the framework of the Chlamydomonas cell wall by proteases present in a commercially available alpha-amylase preparation. *Applied and Environmental Microbiology* **1987**, *53*, 1701–1704.

[108] Takeda, H. Sugar composition of the cell wall and the taxonomy of Chlorella (Chlorophyceae) 1. *Journal of Phycology* **1991**, *27*, 224–232.

[109] Braun, E.; Aach, H. Enzymatic degradation of the cell wall of Chlorella. *Planta* **1975**, *126*, 181–185.

[110] Yamada, T.; Sakaguchi, K. Comparative studies onChlorella cell walls: Induction of protoplast formation. *Archives of Microbiology* **1982**, *132*, 10–13.

[111] Afi, L.; Metzger, P.; Largeau, C.; Connan, J.; Berkaloff, C.; Rousseau, B. Bacterial degradation of green microalgae: incubation of Chlorella emersonii and Chlorella vulgaris with Pseudomonas oleovorans and Flavobacterium aquatile. *Organic Geochemistry* **1996**, *25*, 117–130.

[112] Rochaix, J.-D.; Van Dillewijn, J.; Rahire, M. Construction and characterization of autonomously replicating plasmids in the green unicellular alga Chlamydomonas reinhardii. *Cell* **1984**, *36*, 925–931.

[113] Karas, B. J.; Diner, R. E.; Lefebvre, S. C.; McQuaid, J.; Phillips, A. P.; Noddings, C. M.; Brunson, J. K.; Valas, R. E.; Deerinck, T. J.; Jablanovic, J., et al. Designer diatom episomes delivered by bacterial conjugation. *Nature Communications* **2015**, *6*, 1–10.

[114] Li, Z.; Bock, R. Replication of bacterial plasmids in the nucleus of the red alga Porphyridium purpureum. *Nature Communications* **2018**, *9*, 1–8.

[115] Tsong, T. Y. *Electroporation and electrofusion in cell biology*; Springer, 1989; pp 149–163.

[116] Zhang, C.; Hu, H. High-efficiency nuclear transformation of the diatom Phaeodactylum tricornutum by electroporation. *Marine Genomics* **2014**, *16*, 63–66.

[117] Guo, S.-L.; Zhao, X.-Q.; Tang, Y.; Wan, C.; Alam, M. A.; Ho, S.-H.; Bai, F.-W.; Chang, J.-S. Establishment of an efficient genetic transformation system in Scenedesmus obliquus. *Journal of Biotechnology* **2013**, *163*, 61–68.

[118] Chow, K.-C.; Tung, W. Electrotransformation of Chlorella vulgaris. *Plant Cell Reports* **1999**, *18*, 778–780.

[119] Walker, T. L.; Becker, D. K.; Dale, J. L.; Collet, C. Towards the development of a nuclear transformation system for Dunaliella tertiolecta. *Journal of Applied Phycology* **2005**, *17*, 363–368.

[120] Chungjatupornchai, W.; Kitraksa, P.; Fa-aroonsawat, S. Stable nuclear transformation of the oleaginous microalga Neochloris oleoabundans by electroporation. *Journal of Applied Phycology* **2016**, *28*, 191–199.

[121] Brown, L. E.; Sprecher, S.; Keller, L. Introduction of exogenous DNA into Chlamydomonas reinhardtii by electroporation. *Molecular and Cellular Biology* **1991**, *11*, 2328–2332.

[122] Shimogawara, K.; Fujiwara, S.; Grossman, A.; Usuda, H. High-efficiency transformation of Chlamydomonas reinhardtii by electroporation. *Genetics* **1998**, *148*, 1821–1828.

[123] Kilian, O.; Benemann, C. S.; Niyogi, K. K.; Vick, B. High-efficiency homologous recombination in the oil-producing alga Nannochloropsis sp. *Proceedings of the National Academy of Sciences* **2011**, *108*, 21265–21269.

[124] Schiedlmeier, B.; Schmitt, R.; Müller, W.; Kirk, M. M.; Gruber, H.; Mages, W.; Kirk, D. L. Nuclear transformation of Volvox carteri. *Proceedings of the National Academy of Sciences* **1994**, *91*, 5080–5084.

[125] Jarvis, E. E.; Brown, L. M. Transient expression of firefly luciferase in protoplasts of the green alga Chlorella ellipsoidea. *Current Genetics* **1991**, *19*, 317–321.

[126] Dawson, H. N.; Burlingame, R.; Cannons, A. C. Stable transformation of Chlorella: rescue of nitrate reductase-deficient mutants with the nitrate reductase gene. *Current Microbiology* **1997**, *35*, 356–362.

[127] Dunahay, T. G.; Jarvis, E. E.; Roessler, P. G. Genetic transformation of the diatoms Cyclotella cryptica and Navicula saprophila. *Journal of Phycology* **1995**, *31*, 1004–1012.

[128] Apt, K. E.; Kroth-Pancic, P. G.; Grossman, A. R. Stable nuclear transformation of the diatom Phaeodactylum tricornutum. *Molecular and General Genetics* **1996**, *252*, 572–579.

[129] Zaslavskaia, L. A.; Lippmeier, J. C.; Shih, C.; Ehrhardt, D.; Grossman, A. R.; Apt, K. E. Trophic Conversion of an Obligate Photoautotrophic Organism Through Metabolic Engineering. *Science* **2001**, *292*, 2073–2075.

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

[131] Kathiresan, S.; Chandrashekar, A.; Ravishankar, G.; Sarada, R. Regulation of astaxanthin and its intermediates through cloning and genetic transformation of $\beta$-carotene ketolase in Haematococcus pluvialis. *Journal of Biotechnology* **2015**, *196*, 33–41.

[132] Cheng, R.; Ma, R.; Li, K.; Rong, H.; Lin, X.; Wang, Z.; Yang, S.; Ma, Y. Agrobacterium tumefaciens mediated transformation of marine microalgae Schizochytrium. *Microbiological Research* **2012**, *167*, 179–186.

[133] Prasad, B.; Vadakedath, N.; Jeong, H.-J.; General, T.; Cho, M.-G.; Lein, W. Agrobacterium tumefaciens-mediated genetic transformation of haptophytes (Isochrysis species). *Applied Microbiology and Biotechnology* **2014**, *98*, 8629–8639.

[134] Pratheesh, P.; Vineetha, M.; Kurup, G. M. An efficient protocol for the Agrobacterium-mediated genetic transformation of microalga Chlamydomonas reinhardtii. *Molecular Biotechnology* **2014**, *56*, 507–515.

[135] Kumar, S. V.; Misquitta, R. W.; Reddy, V. S.; Rao, B. J.; Rajam, M. V. Genetic transformation of the green alga—Chlamydomonas reinhardtii by Agrobacterium tumefaciens. *Plant Science* **2004**, *166*, 731–738.

[136] Brumwell, S. L.; MacLeod, M. R.; Huang, T.; Cochrane, R. R.; Meaney, R. S.; Zamani, M.; Matysiakiewicz, O.; Dan, K. N.; Janakirama, P.; Edgell, D. R., et al. Designer Sinorhizobium meliloti strains and multi-functional vectors enable direct inter-kingdom DNA transfer. *PLoS ONE* **2019**, *14*, e0206781.

[137] Bashir, K. M. I.; Kim, M.-S.; Stahl, U.; Cho, M.-G. Microalgae engineering toolbox: selectable and screenable markers. *Biotechnology and Bioprocess Engineering* **2016**, *21*, 224–235.

[138] Doron, L.; Segal, N.; Shapira, M. Transgene Expression in Microalgae-From Tools to Applications. *Frontiers in Plant Science* **2016**, *7*, 505.

[139] Debuchy, R.; Purton, S.; Rochaix, J.-D. The argininosuccinate lyase gene of Chlamydomonas reinhardtii: an important tool for nuclear transformation and for correlating the genetic and molecular maps of the ARG7 locus. *The EMBO Journal* **1989**, *8*, 2803–2809.

[140] Sun, Y.; Gao, X.; Li, Q.; Zhang, Q.; Xu, Z. Functional complementation of a nitrate reductase defective mutant of a green alga Dunaliella viridis by introducing the nitrate reductase gene. *Gene* **2006**, *377*, 140–149.

[141] Kim, H.; Kim, J. S. A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics* **2014**, *15*, 321–334.

[142] Sizova, I.; Greiner, A.; Awasthi, M.; Kateriya, S.; Hegemann, P. Nuclear gene targeting in Chlamydomonas using engineered zinc-finger nucleases. *Plant Journal* **2013**, *73*, 873–882.

[143] Daboussi, F.; Leduc, S.; Maréchal, A.; Dubois, G.; Guyot, V.; Perez-Michaut, C.; Amato, A.; Falciatore, A.; Juillerat, A.; Beurdeley, M.; Voytas, D. F.; Cavarec, L.; Duchateau, P. Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology. *Nature Communications* **2014**, *5*, 3831.

[144] Nymark, M.; Sharma, A. K.; Sparstad, T.; Bones, A. M.; Winge, P. A CRISPR/Cas9 system adapted for gene editing in marine algae. *Scientific Reports* **2016**, *6*, 24951.

[145] Jiang, W.; Brueggeman, A. J.; Horken, K. M.; Plucinak, T. M.; Weeks, D. P. Successful transient expression of Cas9 and single guide RNA genes in Chlamydomonas reinhardtii. *Eukaryotic Cell* **2014**, *13*, 1465–1469.

[146] Baek, K.; Kim, D. H.; Jeong, J.; Sim, S. J.; Melis, A.; Kim, J.-S.; Jin, E.; Bae, S. DNA-free two-gene knockout in Chlamydomonas reinhardtii via CRISPR-Cas9 ribonucleoproteins. *Scientific Reports* **2016**, *6*.

[147] Shin, S.-E.; Lim, J.-M.; Koh, H. G.; Kim, E. K.; Kang, N. K.; Jeon, S.; Kwon, S.; Shin, W.-S.; Lee, B.; Hwangbo, K., et al. CRISPR/Cas9-induced knockout and knock-in mutations in Chlamydomonas reinhardtii. *Scientific Reports* **2016**, *6*, 1–15.

[148] Wang, Q.; Lu, Y.; Xin, Y.; Wei, L.; Huang, S.; Xu, J. Genome editing of model oleaginous microalgae Nannochloropsis spp. by CRISPR/Cas9. *The Plant Journal* **2016**, *88*, 1071–1081.

[149] Kim, Y.-G.; Cha, J.; Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences* **1996**, *93*, 1156–1160.

[150] Christian, M.; Cermak, T.; Doyle, E. L.; Schmidt, C.; Zhang, F.; Hummel, A.; Bogdanove, A. J.; Voytas, D. F. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **2010**, *186*, 757–761.

[151] Jiang, F.; Doudna, J. A. CRISPR–Cas9 structures and mechanisms. *Annual Review of Biophysics* **2017**, *46*, 505–529.

[152] Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A., et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013**, *339*, 819–823.

[153] Zhang, Y.-T.; Jiang, J.-Y.; Shi, T.-Q.; Sun, X.-M.; Zhao, Q.-Y.; Huang, H.; Ren, L.-J. Application of the CRISPR/Cas system for genome editing in microalgae. *Applied Microbiology and Biotechnology* **2019**, *103*, 3239–3248.

[154] Slattery, S. S.; Diamond, A.; Wang, H.; Therrien, J. A.; Lant, J. T.; Jazey, T.; Lee, K.; Klassen, Z.; Desgagné-Penix, I.; Karas, B. J.; Edgell, D. R. An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum. *ACS Synthetic Biology* **2018**, *7*, 328–338.

[155] Weyman, P. D.; Beeri, K.; Lefebvre, S. C.; Rivera, J.; Mccarthy, J. K.; Heuberger, A. L.; Peers, G.; Allen, A. E.; Dupont, C. L. Inactivation of Phaeodactylum tricornutum urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnology Journal* **2015**, *13*, 460–470.

[156] Sodeinde, O. A.; Kindle, K. L. Homologous recombination in the nuclear genome of Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences* **1993**, *90*, 9199–9203.

[157] Nelson, J.; Lefebvre, P. A. Targeted disruption of the NIT8 gene in Chlamydomonas reinhardtii. *Molecular and Cellular Biology* **1995**, *15*, 5762–5769.

[158] Plecenikova, A.; Mages, W.; Andrésson, Ó. S.; Hrossova, D.; Valuchova, S.; Vlcek, D.; Slaninova, M. Studies on recombination processes in two Chlamydomonas reinhardtii endogenous genes, NIT1 and ARG7. *Protist* **2013**, *164*, 570–582.

[159] Minoda, A.; Sakagami, R.; Yagisawa, F.; Kuroiwa, T.; Tanaka, K. Improvement of culture conditions and evidence for nuclear transformation by homologous recombination in a red alga, Cyanidioschyzon merolae 10D. *Plant and Cell Physiology* **2004**, *45*, 667–671.

[160] Angstenberger, M.; Krischer, J.; Aktas, O.; Buchel, C. Knock-Down of a ligIV Homologue Enables DNA Integration via Homologous Recombination in the Marine Diatom Phaeodactylum tricornutum. *ACS Synthetic Biology* **2018**, *8*, 57–69.

[161] Martino, A. D.; Meichenin, A.; Shi, J.; Pan, K.; Bowler, C. Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions. *Journal of Phycology* **2007**, *43*, 992–1009.

[162] Bowler, C. et al. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **2008**, *456*, 239–44.

[163] Khalil, A. S.; Collins, J. J. Synthetic Biology : Applications Come of Age. **2010**, *11*, 367–379.

[164] Xue, J.; Niu, Y. F.; Huang, T.; Yang, W. D.; Liu, J. S.; Li, H. Y. Genetic improvement of the microalga Phaeodactylum tricornutum for boosting neutral lipid accumulation. *Metabolic Engineering* **2015**, *27*, 1–9.

[165] Vanier, G.; Hempel, F.; Chan, P.; Rodamer, M.; Vaudry, D.; Maier, U. G.; Lerouge, P.; Bardor, M. Biochemical characterization of human anti-hepatitis B monoclonal antibody produced in the microalgae Phaeodactylum tricornutum. *PLoS One* **2015**, *10*, e0139282.

[166] Devaki, B.; Grossman, A. R. Characterization of gene clusters encoding the fucoxanthin chlorophyll proteins of the diatom phaeodactylum tricornutum. *Nucleic Acids Research* **1993**, *21*, 4458–4466.

[167] Kira, N.; Ohnishi, K.; Miyagawa-Yamaguchi, A.; Kadono, T.; Adachi, M. Nuclear transformation of the diatom Phaeodactylum tricornutum using PCR-amplified DNA fragments by microparticle bombardment. *Marine Genomics* **2016**, *25*, 49–56.

[168] Miyahara, M.; Aoi, M.; Inoue-Kashino, N.; Kashino, Y.; Ifuku, K. Highly Efficient Transformation of the Diatom Phaeodactylum tricornutum by Multi-Pulse Electroporation. *Bioscience, Biotechnology, and Biochemistry* **2013**, *77*, 874–876.

[169] Falciatore, A.; Casotti, R.; Leblanc, C.; Abrescia, C.; Bowler, C. Transformation of Nonselectable Reporter Genes in Marine Diatoms. *Marine Biotechnology (New York, N.Y.)* **1999**, *1*, 239–251.

[170] Maheswari, U.; Montsant, A.; Goll, J.; Krishnasamy, S.; Rajyashri, K.; Patell, V. M.; Bowler, C. The diatom EST database. *Nucleic Acids Research* **2005**, *33*, D344–D347.

[171] Montsant, A.; Maheswari, U.; Bowler, C.; Lopez, P. J. Diatomics: toward diatom functional genomics. *Journal of Nanoscience and Nanotechnology* **2005**, *5*, 5–14.

[172] Stukenberg, D. D.; Zauner, S. S.; Dell'Aquila, G.; Maier, U. G. Optimizing CRISPR/Cas9 for the diatom Phaeodactylum tricornutum. *Frontiers in Plant Science* **2018**, *9*, 740.

[173] Serif, M.; Lepetit, B.; Weißert, K.; Kroth, P.; Rio Bartulos, C. A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom Phaeodactylum tricornutum. *Algal Research* **2017**, *23*, 186–195.

[174] Hamilton, M. L.; Powers, S.; Napier, J. A.; Sayanova, O. Heterotrophic production of omega-3 long-chain polyunsaturated fatty acids by trophically converted marine diatom phaeodactylum tricornutum. *Marine Drugs* **2016**, *14*.

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

[176] Wang, X.; Liu, Y.-H.; Wei, W.; Zhou, X.; Yuan, W.; Balamurugan, S.; Hao, T.-B.; Yang, W.-D.; Liu, J.-S.; Li, H.-Y. Enrichment of long-chain polyunsaturated fatty acids by coordinated expression of multiple metabolic nodes in the oleaginous microalga Phaeodactylum tricornutum. *Journal of Agricultural and Food Chemistry* **2017**, *65*, 7713–7720.

[177] D'Adamo, S.; Schiano di Visconte, G.; Lowe, G.; Szaub-Newton, J.; Beacham, T.; Landels, A.; Allen, M. J.; Spicer, A.; Matthijs, M. Engineering the unicellular alga Phaeodactylum tricornutum for high-value plant triterpenoid production. *Plant Biotechnology Journal* **2019**, *17*, 75–87.

[178] Fabris, M.; Abbriano, R. M.; Pernice, M.; Sutherland, D. L.; Commault, A. S.; Hall, C. C.; Labeeuw, L.; McCauley, J. I.; Kuzhiuparambil, U.; Ray, P., et al. Emerging Technologies in Algal Biotechnology: Toward the Establishment of a Sustainable, Algae-Based Bioeconomy. *Frontiers in Plant Science* **2020**, *11*.

# Chapter 2

# An expanded plasmid-based genetic toolbox enables Cas9 genome editing and stable maintenance of synthetic pathways in *Phaeodactylum tricornutum*

The work presented in this chapter is reprinted (adapted) with permission (Appendix A) from:

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

Copyright (2020) American Chemical Society

## 2.1   Introduction

Diatoms are a diverse and widespread group of eukaryotic microalgae that play an important role in many aquatic ecosystems through their capacity to fix atmospheric $CO_2$ (1). Diatoms produce more than 20% of global oxygen and possess a rare ability to metabolize silicates, which they incorporate into their cell walls (2, 3). These characteristics make diatoms attractive organisms for biotechnological purposes, with potential applications in the food, cosmetics, and pharmaceutical industries. Diatoms are also an abundant source of long chain fatty acids (4–6) and are particularly relevant to the biofuel industry. The sequencing of the *P. tricornutum* and *T. pseudonana* genomes in 2008 and 2004, respectively (7, 8), and the development of selectable markers and endogenous

regulatory sequences for foreign gene expression has further enabled biotechnological applications (9, 10).

A major barrier to biotechnological applications in diatoms is the low efficiency of DNA transformation (11, 12). A commonly used method for transformation of *P. tricornutum* is microparticle bombardment using DNA-coated microparticles, termed biolistic transformation (12, 13). However, the efficiency of biolistic transformation of *P. tricornutum* is variable, making it unsuitable for applications such as high-throughput screening of plasmid-based libraries. Moreover, selecting biolistic transformants relies on random genomic integration of an antibiotic resistance marker, with possible variable expression levels and potentially undesirable phenotypic effects for downstream applications (14). Targeted genomic integration of transformed DNA is possible by flanking the DNA insert with sequences homologous to the desired site of integration. Genomic integration occurs at very low frequency unless a double-strand break was made at the integration site using a targeted endonuclease (15).

Recently, a novel transformation method was established using conjugation from *E. coli* to maintain stably replicating plasmids in the nucleus of *P. tricornutum* (16–19). Utilizing this new transformation method, it is possible to express heterologous genes in *P. tricornutum* without the need for genomic integration. However, only a few transcriptional promoter elements and selectable markers have been identified and characterized (20–25), limiting the genetic tools available for the study of diatom biology and assembly of complex synthetic pathways. Here, we identified additional endogenous promoters, terminators and introns from *P. tricornutum* that regulated expression of plasmid-based selectable markers. We use these tools to develop a simple and robust plasmid-based Cas9 editing system, demonstrating efficient editing of the urease gene and subsequent curing of the Cas9 plasmid from edited strains. We also show that *P. tricornutum* can stably replicate two plasmids simultaneously, and constructed and replicated over a four-month period a plasmid encoding eight genes involved in vanillin biosynthesis.

## 2.2 Materials and Methods

### 2.2.1 Microbial strains and growth conditions

*S. cerevisiae* VL6-48 (ATCC MYA-3666: *MATα his3-Δ200 trp1-Δ1 ura3-52 lys2 ade2-1 met14 cir^0^*) was grown in YPD medium or complete minimal medium lacking histidine (Teknova) supplemented with 60 mg $l^{-1}$ adenine sulfate. Complete minimal media used for spheroplast transformation contained 1 M sorbitol. *E. coli* (Epi300, Epicentre) was grown in Luria Broth (LB) supplemented with appropriate antibiotics (chloramphenicol (25 mg $l^{-1}$) or kanamycin (50 mg $l^{-1}$) or ampicillin (50 mg $l^{-1}$) or gentamicin (20 mg $l^{-1}$)). *P. tricornutum* (Culture Collection of Algae and Protozoa CCAP 1055/1) was grown in L1 medium without silica at 18°C under cool white fluorescent lights (75 $\mu$E $m^{-2}$ $s^{-1}$) and a photoperiod of 16 h light:8 h dark.

*P. tricornutum* liquid L1 media consisted of 1 l aquil salts, 2 ml nitrate phosphate (NP) stock, 1 ml L1 trace metals stock, 0.5 ml f/2 vitamin solution, and was filter sterilized through a 0.2-$\mu$m filter. For agar plates, equal parts sterilized liquid L1 medium and autoclaved 2% agar were combined and poured into petri dishes. To make aquil salts, two separate solutions of 2× anhydrous and hydrous salts were mixed. Anhydrous salts consisted of (in 500 ml $H_2O$) 24.5 g NaCl, 4.09 g $Na_2SO_4$, 0.7 g KCl, 0.2 g $NaHCO_3$, 0.1 g KBr, 0.03 g $H_3BO_3$ (or 3 ml of 10 mg $ml^{-1}$ stock), 0.003 g NaF (or 300 $\mu$l of 10 mg $ml^{-1}$ stock). Hydrous salts consisted of 11.1 g $MgCl_2 \cdot 6H_2O$ and 1.54 g $CaCl_2$ resuspended in 500 ml $H_2O$. The NP stock was made in 100 ml $H_2O$ and consisted of 37.5 g $NaNO_3$ and 2.5 g $NaH_2PO_4 \cdot H_2O$. The L1 trace metal stock solution was made by mixing $FeCl_3 \cdot 6H_2O$ 3.15 g, $Na_2EDTA \cdot 2H_2O$ 4.36 g, $CuSO_4 \cdot 5H_2O$ (9.8 g $l^{-1}$ $dH_2O$) 0.25 ml, $Na_2MoO_4 \cdot 2H_2O$ (6.3 g $l^{-1}$ $dH_2O$) 3.0 ml, $ZnSO_4 \cdot 7H_2O$ (22.0 g $l^{-1}$ $dH_2O$) 1.0 ml, $CoCl_2 \cdot 6H_2O$ (10.0 g $l^{-1}$ $dH_2O$) 1.0 ml, $MnCl_2 \cdot 4H_2O$ (180.0 g $l^{-1}$ $dH_2O$) 1.0 ml, $H_2SeO_3$ (1.3 g $l^{-1}$ $dH_2O$) 1.0 ml, $NiSO_4 \cdot 6H_2O$ (2.7 g $l^{-1}$ $dH_2O$) 1.0 ml, $Na_3VO_4$ (1.84 g $l^{-1}$ $dH_2O$) 1.0 ml, $K_2CrO_4$ (1.94 g $l^{-1}$ $dH_2O$) 1.0 ml in 1L $H_2O$. The F/2 vitamin stock solution was made by mixing 200 mg powder $l^{-1}$ thiamine-HCl, 10 ml $l^{-1}$ of a 0.1 g $l^{-1}$ biotin stock, and 1 ml $l^{-1}$ of a cyanocobalamin 1 g $l^{-1}$ stock.

## 2.2.2 Transfer of DNA to *P. tricornutum via* conjugation from *E. coli*

Conjugation protocols were adapted from Karas and colleagues (16). Briefly, liquid cultures (250 $\mu$l) of *P. tricornutum* were adjusted to density of $1.0 \times 10^8$ cells ml$^{-1}$ using counts from a hemocytometer, plated on $^1/_2 \times$ L1 1% agar plates and grown for four days. L1 media (1 ml) was added to the plate and cells were scraped and the concentration adjusted to $5.0 \times 10^8$ cells ml$^{-1}$. *E. coli* cultures (50 ml) were grown at 37°C to A$_{600}$ of 0.8–1.0, centrifuged for 10 mins at 3,000 $\times$ *g* and resuspended in 500 $\mu$l of SOC media. Conjugation was initiated by mixing 200 $\mu$l of *P. tricornutum* and 200 $\mu$l of *E. coli* cells. The cell mixture was plated on $^1/_2 \times$ L1 5% LB 1% agar plates, incubated for 90 mins at 30°C in the dark, and then moved to 18°C in the light and grown for 2 days. After 2 days, 1.5 ml of L1 media was added to the plates, the cells scraped, and 300 $\mu$l (20%) plated on $^1/_2 \times$ L1 1% agar plates supplemented with Zeocin™ 50 mg l$^{-1}$ or nourseothricin 200 mg l$^{-1}$. Colonies appeared after 10–14 days incubation at 18°C with light.

## 2.2.3 Plasmid DNA isolation

Plasmid DNA was isolated from *E. coli* using the BioBasic miniprep kit. Plasmid DNA was isolated from all other species using a modified alkaline lysis protocol. For DNA isolation from *S. cerevisiae*, cells were pelleted at 3,000 $\times$ *g* for 5 mins and resuspended in 250 $\mu$l resuspension buffer, which contained 240 $\mu$l P1 (Qiagen), 0.25 $\mu$l of 14 M $\beta$-mercaptoethanol, and 10 $\mu$l zymolyase solution (200 mg zymolyase 20 T (USB), 9 ml H$_2$O, 1 ml 1 M Tris pH7.5, 10 ml 50% glycerol) and incubated at 37°C for 60 mins. For DNA isolation from *P. tricornutum*, cultures (10–20 ml) were harvested during exponential growth phase, pelleted at 4,000 $\times$ *g* for 5 mins, and resuspended in 250 $\mu$l resuspension buffer consisting of 235 $\mu$l P1 (Qiagen), 5 $\mu$l hemicellulose 100 mg ml$^{-1}$, 5 $\mu$l of lysozyme 25 mg ml$^{-1}$, and 5 $\mu$l zymolyase solution and incubated at 37°C for 30 mins. Next, 250 $\mu$l of lysis buffer P2 (Qiagen) was added, followed by 250 $\mu$l of neutralization buffer P3 (Qiagen) and centrifugation at 16,000 $\times$ *g* for 10 mins. The supernatant was transferred to a clean tube, 750 $\mu$l isopropanol was added, and the samples centrifuged at 16,000 $\times$ *g* for 10 mins. A 70% EtOH wash was performed, centrifuged at 16,000 $\times$ *g* for 5 mins, and pellets briefly dried, resuspended in 50–100 $\mu$l of TE buffer, and incubated at 37°C for 30–60 mins to dissolve.

## 2.2.4 Plasmid construction

Plasmids (Supplementary Table B.6) were constructed using a modified yeast assembly (26, 27). Plasmids pPtGE10–17 were made from pPtPuc3 (16) by replacing the *FcpF* promoter with PCR fragments consisting of 500 bp upstream of the start codon of the *P. tricornutum EF1α*, *40SRPS8*, *H4-1B*, *γ-Tubulin*, *RBCMT*, *FcpB*, *FcpC*, or *FcpD* genes. Plasmids pPtGE18–25 were made from pPtGE10–17 by replacing the *FcpA* terminator with PCR fragments consisting of 500 bp downstream of the stop codon of the gene corresponding to each promoter described above. Plasmids pPtGE26–29 were made by assembling the *nat* or *cat* ORF between the *H4-1B* or *FcpD* promoter and *FcpA* terminator and inserting them immediately downstream of the *CEN6-ARSH4-HIS3* element in pPtPuc3. Plasmid pPtGE30 was created from p0521s by replacing the *FcpF* promoter (driving *Sh ble*) by the *FcpD* promoter. In addition, a ClaI restriction site was added immediately upstream of the *FcpD* promoter, and the sequence of the I-SceI site downstream of the *URA3* element was corrected. Plasmid pPtGE31 was made by replacing the *Sh ble* ORF in pPtGE30 with a *nat* ORF. Plasmid pPtGE32 was made by replacing the *Sh ble* ORF in pPtGE30 by a *Sh ble* ORF linked to a *nat* ORF by a T2A self-cleaving peptide linker. Plasmid pPtGE33 was made by replacing the *URA3* element in pPtGE30 by a cassette containing the *FcpD* promoter and terminator driving GFP linked to mCherry by a T2A self-cleaving peptide linker. Plasmid pPtGE34 was made by assembling the oriT, *Sh ble*, and *CEN6-ARSH4-HIS3* elements from pPtGE19 into pKS diaCas9_sgRNA (Addgene #74923) (28) upstream of the *FcpB* promoter. The *Sh ble* gene driven by the *40SRPS8* promoter and terminator from pPtGE19 was chosen because these elements do not contain BsaI sites that would interfere with sgRNA cloning. Plasmid pPtGE35 was made by cloning the first 510 bp (170 aa) of the I-TevI homing endonuclease and a GGSGGS linker immediately upstream of Cas9 in pPtGE34. To facilitate this, the NLS and HA-tag were moved from the N-terminus of Cas9 to the C-terminus. I-TevI variants K135R+N140S, V117F+S134G, or S134G were used depending on the sequence requirements upstream of the sgRNA target site. Using golden gate assembly (29), sgRNAs targeting different regions of the *P. tricornutum* urease gene were cloned into the BsaI sites positioned between the *P. tricornutum U6* promoter and terminator in pPtGE34 and pPtGE35. Plasmid constructs were confirmed by Sanger sequencing at the London Regional Genomics Facility. The plasmid (p0521s-V) encoding components of a vanillin biosyn-

thesis pathway was made by assembling genes linked in pairs by 2A self-cleaving peptides, flanked by *P. tricornutum* promoters and terminators, between the I-CeuI and I-SceI sites in p0521s (16). The pairs are as follows: phenylalanine ammonia lyase (PAL) and trans-cinnamate 4-hydroxylase (C4H) with the H41B promoter and terminator, tyrosine ammonia lyase (TAL) and p-coumarate 3-hydroxylase (C3H) with the *40SRPS8* promoter and terminator, hydroxycinnamoyl transferase (HCT) and 4-hydroxycinnamoyl-CoA ligase (4CL) with the *EF1α* promoter and terminator, and caffeic acid 3-O-methyltransferase (COMT) and vanillin synthase (VAN) with the *FcpD* promoter and terminator. The plasmid was assembled with a *URA3* yeast marker positioned between the *40SRPS8* terminator and the *EF1α* promoter to assist in selection of correctly assembled plasmids. Two clones of p0521s-V were completely sequenced as described below.

## 2.2.5   Western blot

*P. tricornutum* cultures (5 ml) were harvested during exponential growth phase and pelleted at $4,000 \times g$ for 5 mins. The cell pellet was resuspended in 300 $\mu$l of sodium dodecyl sulphate (SDS) sample loading buffer (60 mM Tris-HCl, pH 6.8, 2% SDS [w/v], 0.01% bromophenol blue [w/v], 1% $\beta$-mercaptoethanol [v/v]) and boiled for 5 mins. Samples (10 $\mu$l) were resolved on 12% polyacrylamide gels and electroblotted to a polyvinylidene difluoride (PVDF) membrane. Membranes were blocked in 5% milk for 1 hr, then incubated in 5% milk at 4°C overnight with primary antibody (1:500 dilution mouse anti-GFP, Sigma) followed by 1 hr incubation with a 1:5000 dilution of secondary antibody (goat anti-mouse-HRP, Jackson ImmunoResearch). Blots were developed using Clarity ECL Western Blotting Substrate (BioRad).

## 2.2.6   RNA isolation and cDNA synthesis

To isolate total RNA from *P. tricornutum*, cultures (10 ml) were harvested during exponential growth phase, pelleted at $4,000 \times g$ for 5 mins, supernatant was removed, and pellets were flash frozen in liquid nitrogen. Frozen pellets were ground in liquid nitrogen and total RNA was extracted using an RNeasy Plant Mini Kit (Qiagen). RNA samples were treated with 3U of TURBO DNase (Invitrogen), supplemented with 15U of HhaI (NEB), in 50 $\mu$l reactions for 45 mins at 37°C, then Ethylenediaminetetraacetic acid (EDTA) was added to a final concentration of 15 mM

and samples were heat inactivated at 75°C for 15 mins. Reverse transcription of 2 $\mu$g of total RNA from all samples was performed with High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Reactions where the reverse transcriptase was not added were included for all samples to be used as genomic and plasmid DNA controls during the quantitative real-time PCR (RT-qPCR) analyses. The cDNA samples were diluted to 30 ng $\mu$l$^{-1}$ with TE buffer and stored at -20°C.

### 2.2.7   Quantitative PCR experiments

Quantitative PCR analysis was performed using a ViiA 7 Real-Time PCR System (Applied Biosystems) with SYBR Select Master Mix (Applied Biosystems). To perform copy number qPCR experiments, total DNA was extracted from *P. tricornutum* cells containing plasmids p0521s and pPtGE31. Standard curves were performed for *Sh ble* and *nat* using serial dilutions of purified p0521s or pPtGE31 plasmid, respectively. Copy number qPCR reactions consisted of master mix diluted to 1×, purified plasmid serial dilutions or *P. tricornutum* plasmid extracts diluted 1:10000, and 400 nM primers in a 10 $\mu$l total volume. Forward and reverse primers used are listed in Supplementary Table B.5. The reactions were cycled under the following conditions: 50°C for 2 mins, 95°C for 2 mins, 40 cycles of 95°C for 15 s, 60°C for 15 s, then 72°C for 1 min during which data were collected. Ct values were calculated by QuantStudio™ Real-Time PCR software, plotted as a function of number of template molecules and fit to a logarithmic trend line. Curves were linear over at least four orders of magnitude. Plasmid samples extracted from two *P. tricornutum* strains maintaining both plasmids were tested, each with at least three technical replicates. Ct values resulting from experimental samples were used to calculate the number of molecules of template molecule in the qPCR reaction. To perform relative gene expression qPCR experiments, total RNA was extracted from *P. tricornutum* cells containing pPtPuc3, pPtGE11, pPtGE12, pPtGE13, pPtGE16, pPtGE19, pPtGE20, pPtGE21, or pPtGE14 and converted to cDNA. Relative gene expression qPCR reactions consisted of master mix diluted to 1×, 30 ng cDNA generated from *P. tricornutum* total RNA, and 400 nM primers in 10 $\mu$l total volume. Forward and reverse primers used are listed in Supplementary Table B.5. The reactions were cycled under the following conditions: 50°C for 2 mins, 95°C for 2 mins, 40 cycles of 95°C for 15 s, 60°C for 15 s, then 72°C for 1 min

during which data were collected. Three biological replicates were analyzed with at least three technical replicates for each. The *P. tricornutum Actin16* gene was used as an endogenous control and *Sh ble* under control of the *FcpF* promoter and *FcpA* terminator was used as a reference gene for relative quantitation. Ct values were calculated by QuantStudio™ Real-Time PCR software and the relative expression of *Sh ble* for each sample was calculated using double delta Ct analysis.

## 2.2.8   Generation of urease knockouts using Cas9 and TevCas9

Plasmids pPtGE34 or pPtGE35, containing no guide RNA or sgRNA#1, sgRNA#2, sgRNA#3, or sgRNA#4 for the *P. tricornutum* urease gene, were conjugated from *E. coli* to *P. tricornutum* and exconjugants were selected on Zeocin™-containing media. Ten colonies from each conjugation were resuspended in TE buffer and flash frozen at -80°C followed by heating at 95°C to lyse cells and extract genomic DNA. The genomic target site of each sgRNA in *P. tricornutum* was amplified by PCR and the products were analyzed by T7EI assay as follows; PCR products were denatured at 95°C for 5 mins, slowly cooled to 50°C, and flash frozen at -20°C for 2 mins. PCR products (250 ng) were incubated with 2U of T7EI (NEB) in 1× NEBuffer 2 for 15 mins at 37°C and analyzed by agarose gel electrophoresis. Colonies that showed editing by T7EI assay were grown in liquid culture supplemented with Zeocin™ for 2 weeks and serial dilutions were plated on selective media to isolate sub-clones. Sub-clones were then screened for homozygous urease knockout phenotypes by replica streaking on L1 media containing 2.1 mM urea or 4.2 mM nitrate. Streaks were grown for 5 days before visual identification of phenotypes. Sub-clones that were identified as urease knockouts were resuspended in TE buffer and flash frozen at -80°C followed by heating at 95°C to lyse cells and extract genomic DNA, then sgRNA target sites were PCR amplified. Sanger sequencing of PCR products was performed at the London Regional Genomics Facility to identify the type and length of indels generated. Stable bi-allelic urease knockout mutant lines were then grown in nonselective L1 media for 1 week to cure them of plasmids before plating to obtain single colonies. Resulting colonies were replica streaked onto nonselective and Zeocin™-containing media to identify colonies which had successfully been cured of the plasmid.

### 2.2.9  Whole plasmid sequencing and genotyping

Plasmids encoding the vanillin synthesis pathway were propagated for 1 or 4 months, and DNA isolated from *P. tricornutum* as described and transformed into *E. coli* for cloning. Plasmids were extracted from *E. coli*, linearized by digestion with I-CeuI, electrophoresed on a 0.8% agarose gel, stained with ethidium bromide, and analyzed on an AlphaImager™ 3400 (Alpha Innotech). Plasmids were also used as template for multiplex PCR designed to amplify the *Sh ble* and *nat* ORFs using primers DE3214, DE3215, DE3216, and DE3217. Plasmids were completely sequenced by Nicole Stange Thomann and the team at CCIB DNA Core (Massachusetts General Hospital, *via* Science Exchange). To display sequence coverage and SNPs, each FASTQ file was aligned to the plasmid reference sequence using BOWTIE2 on the Galaxy server (https://cpt.tamu.edu/galaxy-pub/), and the bam files converted to sam format using SAMTools. A plasmid Editor (ApE) and BLAST Ring Image Generator (BRIG) (30) were used to draw the plasmid map and display sequence coverage.

## 2.3  Results

### 2.3.1  Endogenous *P. tricornutum* promoters drive exogenous gene expression

To characterize additional *P. tricornutum* endogenous promoters and terminators for expression of exogenous genes, we identified putative promoter and terminator regions from the *P. tricornutum* genome sequence that are predicted to be highly expressed based on their function (Supplementary Table B.1. Supplementary Table B.7). The promoters with their corresponding native terminators, or with the previously described *FcpA* terminator (31, 32), were cloned up- and downstream of the *Sh ble* coding region, respectively (Figure 2.1A). The resultant plasmids (named pPtGE10 through pPtGE25) were conjugated to *P. tricornutum* from *E. coli* and the efficiency of promoter/terminator pairings was determined by the number of *P. tricornutum* colonies that formed on Zeocin™-containing media after 14-days incubation (Figure 2.1B). As a positive control, we used the previously described *FcpF* promoter and *FcpA* terminator that were cloned up- and downstream of the *Sh ble* gene, respectively (31–34). Similar assays have been used to test genetic

elements in microalgae (20, 32). We found that all promoter/terminator pairs produced colonies but the $\gamma$-tubulin and *FcpC* promoters with their native terminators yielded an 11- and 7.4-fold increase in Zeocin™-resistant colonies than when the promoters were paired with the *FcpA* terminator (Figure 2.1B,C). Other combinations of promoter/terminator pairs exhibited smaller changes in the appearance of Zeocin™-resistant colonies (Figure 2.1B,C). To extend the results of the plating assays, we used quantitative reverse transcription PCR (RT-qPCR) to measure the expression levels of 4 promoters (*40SRPS8*, *H4-1B*, $\gamma$-tubulin, and *FcpC*), finding that the differences in expression levels when each promoter was paired with its native or *FcpA* terminator correlated with the results of the plating assays (Figure 2.1D).

### 2.3.2   *P. tricornutum* introns allow expression of foreign genes in *P. tricornutum* and not in *E. coli*

In addition to promoter and terminator elements, we also tested the function of *P. tricornutum* nuclear (spliceosomal) introns embedded within exogenous genes (Supplementary Table B.7). We first cloned an intron from the *FBAC2* gene at position 1072 of the MegaTev genome-editing nuclease (35), and showed by RT-PCR that the intron was correctly spliced in *P. tricornutum* (Figure 2.2a). We next inserted the *FBAC2* intron, and introns from the *TUFA* and *RPS4* genes, into the MscI (position 1) or SexAI (position 2) restriction sites in the Zeocin™ resistance gene *Sh ble* on plasmid p0521s (Figure 2.3a). The MscI and SexAI sites of the *ble* gene have previously been used to test the effects of endogenous introns on gene expression in the microalga *C. reinhardtii* (36). Plating of *P. tricornutum* exconjugants harboring each of these plasmids on Zeocin™-containing media reproducibly yielded similar colony numbers to an intronless *Sh ble* control, whereas a *Sh ble* construct with the RPS4-1 intron lacking the predicted splice donor and acceptor sites did not grow on Zeocin™-containing plates (Figure 2.3b). We also demonstrated that the intron-containing constructs prevented growth of *E. coli* on Zeocin™-containing plates, presumably because the introns cannot be spliced from the *Sh ble* transcript (Figure 2.2b).
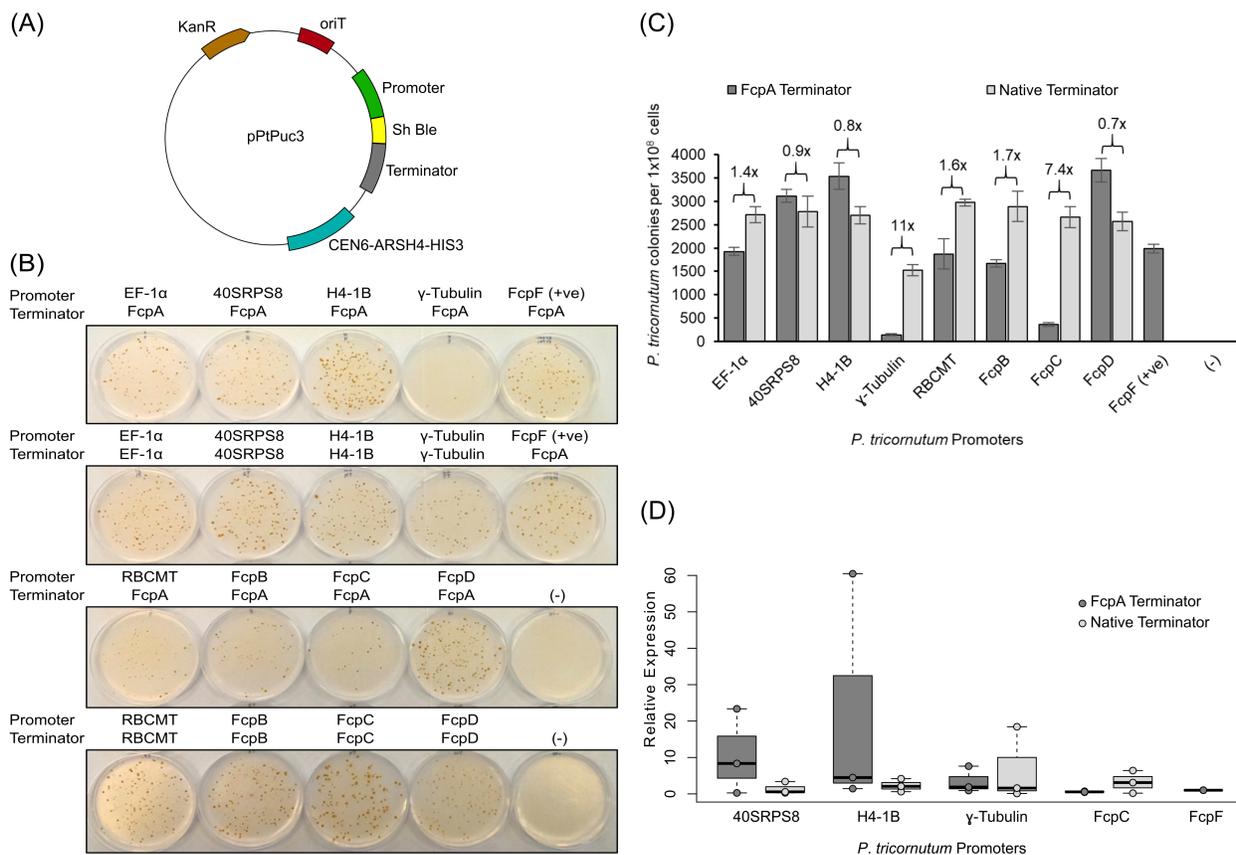
**Figure 2.1: Effect of *P. tricornutum* promoters and terminators on the efficiency of the *Sh ble* marker.** (A) Diagram of promoter-terminator assay constructs. (B) *P. tricornutum* transformants selected on L1 agar media supplemented with Zeocin™. (C) Bar graph of *P. tricornutum* colony formation for each of the promoter/terminator pairs. Values listed above each data set indicate the fold difference in Zeocin™-resistant colonies obtained from promoters paired with their native terminator compared to the *FcpA* terminator. Dark gray and light gray bars indicate promoters paired with the *FcpA* terminator or their native terminator, respectively. The *FcpF* promoter with *FcpA* terminator driving *Sh ble* was used as a positive control (+). Wild-type *P. tricornutum* was used as a negative control (-). Error bars represent standard error of the mean for at least three replicates. (D) Boxplot of relative expression of *Sh ble* determined by RT-qPCR. Expression was normalized to *Sh ble* driven by the *FcpF* promoter and *FcpA* terminator. Solid bars represent the median values. The lower and upper boundaries of the boxes represent the first and third quartiles, respectively. The whiskers represent the highest and lowest values.
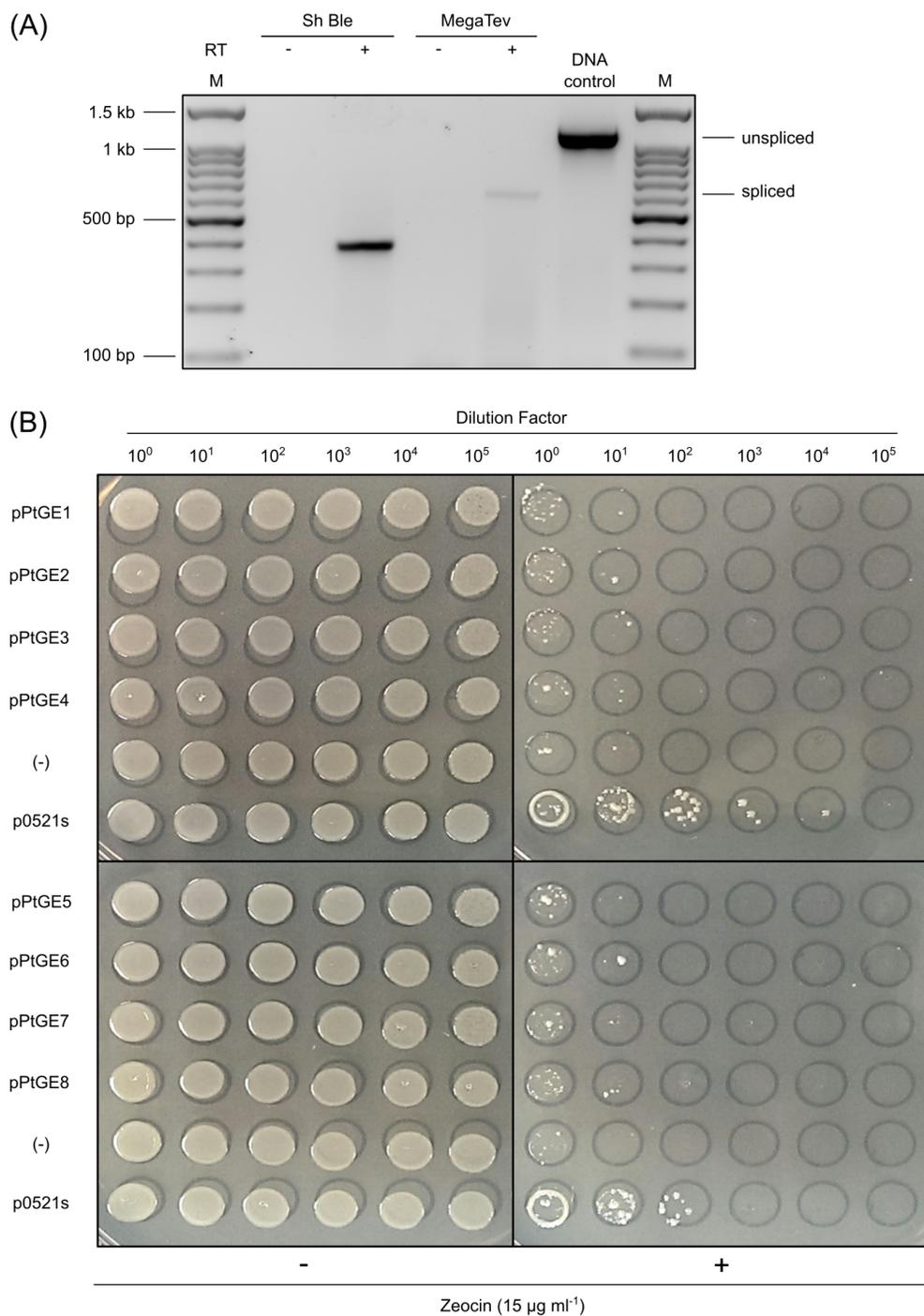
**Figure 2.2: Effect of *P. tricornutum* introns on foreign gene expression.** (A) RT-PCR detection of the *P. tricornutum* FBAC2-1 intron splicing from the MegaTev endonuclease ORF. The *Sh ble* construct was not interrupted by an intron. M, 100 bp ladder. (B) Introns in either MscI (top) or SexAI (bottom) position of the *Sh ble* ORF prevented *E. coli* growth on media supplemented with zeocin. Colonies of strains not resistant to zeocin that form on zeocin plates at the $10^0$ and $10^1$ dilutions are due to spontaneous resistance.
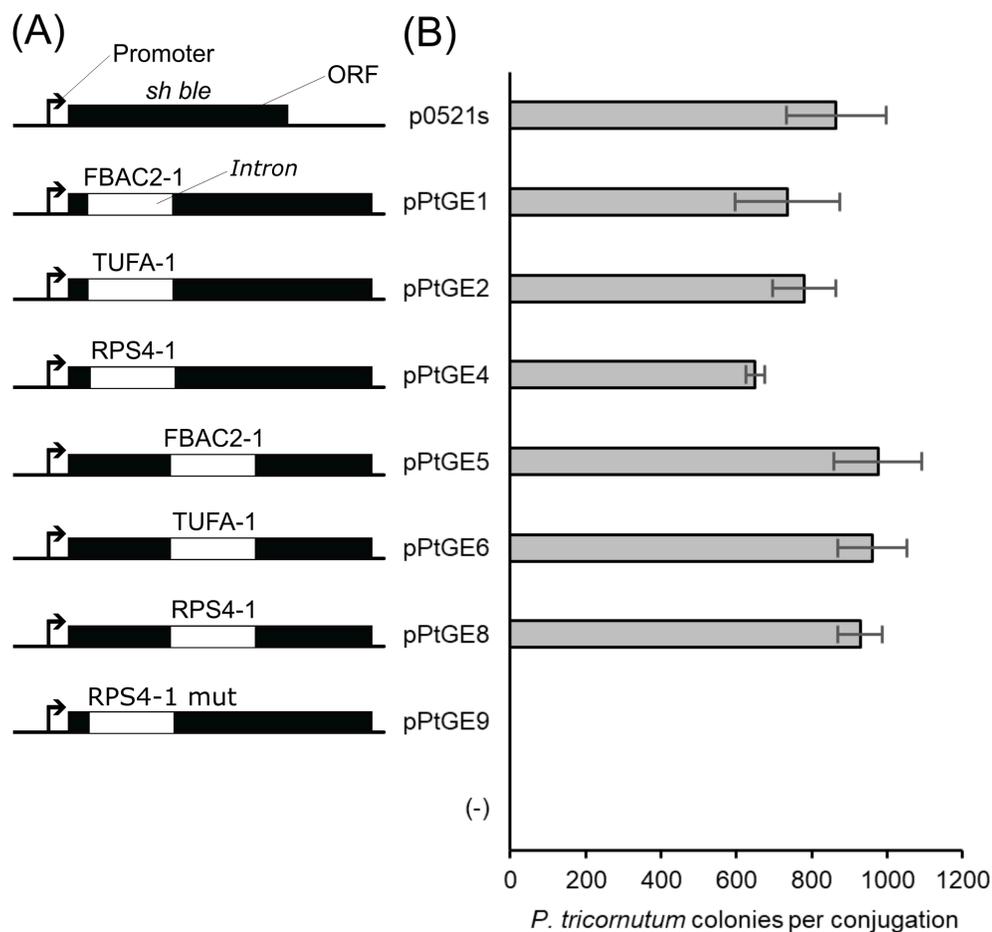
**Figure 2.3: Effect of *P. tricornutum* introns on colony formation.** (A) Diagram of intron assay constructs. FBAC2-1, fructose-1,6-bisphosphate aldolase intron 1; TUFA-1, elongation factor Tu intron 1; RPS4-1, ribosomal protein S4 intron 1; RPS4-1 mut, intron lacking predicted splice donor and acceptor sites; (-), mock conjugation negative control. (B) Bar graph of *P. tricornutum* colony counts for each construct on zeocin-containing plates. Values represent the average colonies obtained from conjugation with each plasmid. Error bars show standard error of the mean for at least three replicates.

### 2.3.3  The *Sh ble* and *nat* antibiotic resistance markers allow selection of replicating plasmids in *P. tricornutum*

When designing plasmid-based systems for expression in eukaryotic systems, a selectable marker (typically an antibiotic resistance gene) is necessary to isolate transformed cells. Several antibiotic resistance genes, including but not limited to those for nourseothricin and chloramphenicol, have been used as selectable markers when integrated into the nuclear genome of *P. tricornutum* (9, 11, 37, 38), but have not yet been shown to function in replicating plasmids. We therefore cloned the *Sh ble* gene (conferring resistance to Zeocin™) with either the *nat* or *cat* genes (conferring resistance to nourseothricin and chloramphenicol (39, 40)) onto a replicating plasmid (Figure 2.4A). The *nat* and *cat* resistance genes were cloned downstream of the *H4-1B* or *FcpD* promoter, and paired with the corresponding native terminator, as shown in Figure 2.4A. Plasmids were conjugated to *P. tricornutum* and selected on Zeocin™-containing plates. Individual colonies harboring the different plasmids were grown in liquid media and dilutions were spotted onto agar plates supplemented with different concentrations of Zeocin™, nourseothricin or chloramphenicol. As shown in Figure 2.4B, all plasmids promoted growth on Zeocin™-containing plates regardless of the promoter used and the concentration of antibiotic in the media. We observed that the *cat* marker was much less effective at promoting chloramphenicol resistance than previously reported(11), which could be due to specific genomic location integration or integration of multiple copies of the *cat* gene (12, 14), or the possibility that we used different promoter/terminator pairs to drive expression of the *cat* gene than in previous studies (11). Regardless, including multiple copies of the *cat* gene or enhancing expression with stronger promoter/terminator pairings may be required for stringent resistance in plasmid-based systems. Interestingly, we observed that chloramphenicol, nourseothricin, and Zeocin™ selection was more effective when the amount of aquil salts used in media preparation was reduced to one quarter (half that in typical $^1/_2$ × L1 plates) when re-plating 10% of scraped cells onto selective media (Figure 2.4B and Figure 2.5). This effect has been previously observed where a decrease in the concentration of seawater from 100% to 50% in plates increased the sensitivity of *P. tricornutum* to Zeocin™ and phleomycin (32). We also demonstrated antibiotic resistance on appropriate plates when the *nat* and *Sh ble* genes were separated by the sequence for an in-frame 2A peptide linker that would self-cleave upon translation

(41), liberating mature *nat* and *Sh ble* polypeptides (Figure 2.6a). Function of the self-cleaving 2A peptide was also demonstrated using a GFP-2A-mCherry reporter system (Figure 2.6b and c).
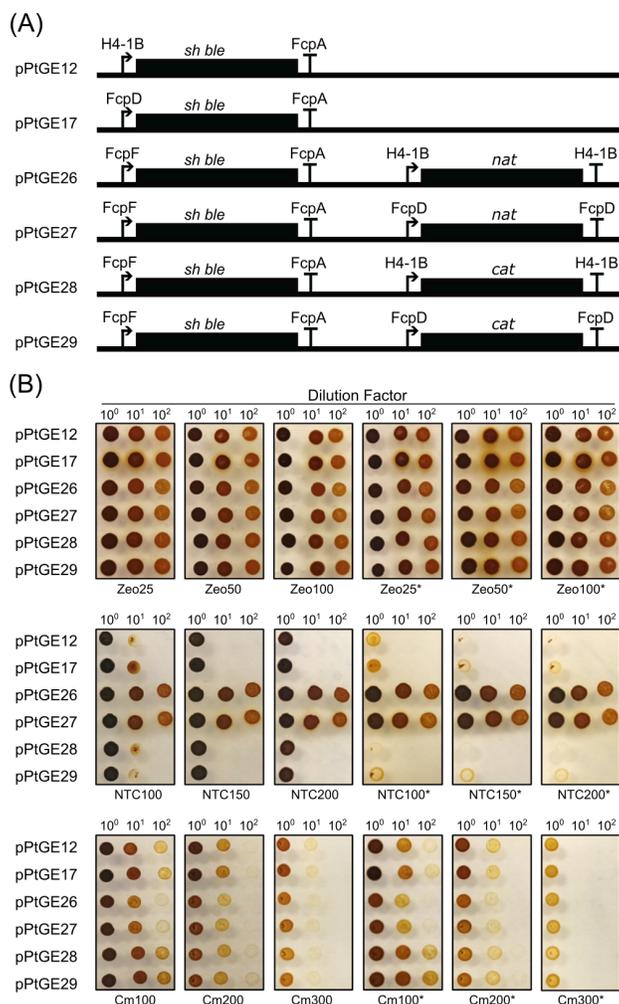


**Figure 2.4: Effect of plasmid-based selectable markers on antibiotic resistance in *P. tricornutum*.** (A) Diagram of selective marker constructs. (B) Representative images of *P. tricornutum* exconjugants plated on L1 media supplemented with Zeocin™ (zeo), nourseothricin (NTC), or chloramphenicol (cm). Antibiotic concentrations listed below each panel are in $\mu$g mL$^{-1}$. Antibiotic concentrations with asterisks (*) indicate plates made with 1/4 × aquil salts concentration (half that of typical 1/2 × L1 plates).

To further compare the effectiveness of the Zeocin™ and nourseothricin resistance genes as selective markers for conjugation to *P. tricornutum*, we cloned the *Sh ble* or *nat* coding region between the *FcpD* promoter and *FcpA* terminator and separately assembled them into p0521s to generate pPtGE30 and pPtGE31, respectively. Plasmids pPtGE30 or pPtGE31 were moved into *P. tricornutum via* conjugation with *E. coli*, and exconjugants were isolated on media supplemented with Zeocin™ or nourseothricin, respectively (Figure 2.7a). *P. tricornutum* colonies selected on
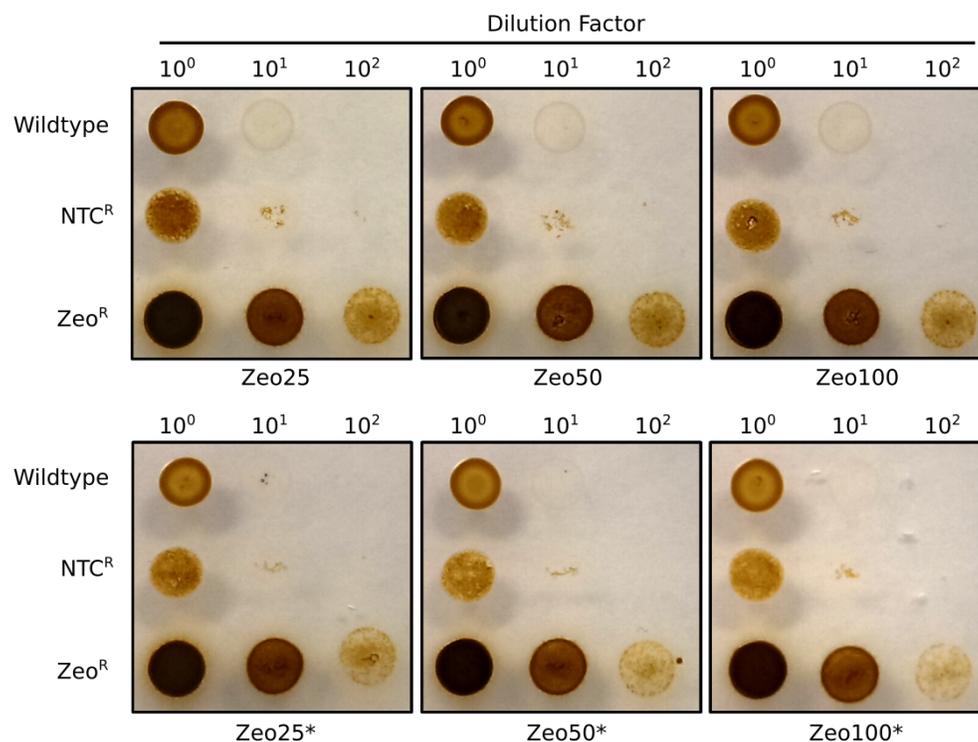
**Figure 2.5: Effect of varying antibiotic and salt concentration on the *Sh ble* marker in *P. tricornutum*.** *P. tricornutum* conjugated with marker plasmids were first selected on appropriate antibiotic, single colonies were isolated, grown in selective liquid media, and adjusted to the same concentration before spotting onto L1 media containing different concentrations of Zeocin™ (Zeo). Antibiotic concentrations listed below each panel are in $\mu$g ml$^{-1}$. Antibiotic concentrations with asterisks (*) indicate plates made with $^1/_4$ aquil salts concentration (half that of typical $^1/_2 \times$ L1 plates).
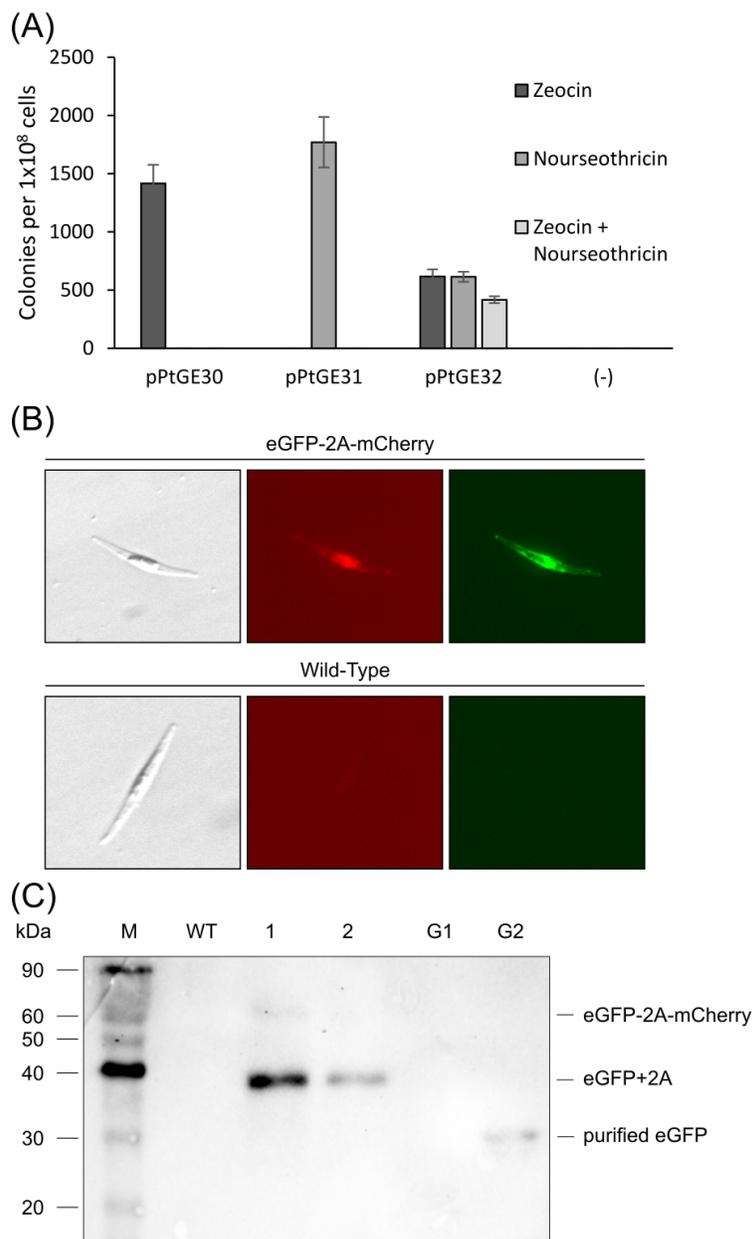
**Figure 2.6: The T2A self-cleaving peptide linker promotes co-expression of proteins in *P. tricornutum*.** (A) The *Sh ble* ORF between the *FcpD* promoter and *FcpA* terminator in pPtGE30 was replaced by a *Sh ble* ORF linked to a *nat* ORF by a T2A self-cleaving peptide linker to generate plasmid pPtGE32. *P. tricornutum* transformed with pPtGE32 survives on media containing Zeocin™ and nourseothricin. (B) The *URA3* element in pPtGE30 was replaced by GFP linked to mCherry by a T2A self-cleaving peptide linker, driven by the *FcpD* promoter and terminator, to generate plasmid pPtGE33. GFP and mCherry fluorescence was detected in *P. tricornutum* transformed with pPtGE33. (C) Western blot of GFP extracted from *P. tricornutum* strains containing pPtGE33. M, protein ladder; WT, wild-type *P. tricornutum* whole cell extract; 1 and 2, whole cell extracts of two *P. tricornutum* clones containing pPtGE33; G1, 3 ng of purified eGFP; G2, 30 ng of purified eGFP.

nourseothricin were visible after 7 days, two days earlier than Zeocin™-resistant colonies, possibly because nourseothricin is more efficient at killing off the untransformed *P. tricornutum* after conjugation (Figure 2.7b). These results indicate that the nourseothricin resistance gene, *nat*, is an effective plasmid-based marker for selection of *P. tricornutum* exconjugants.
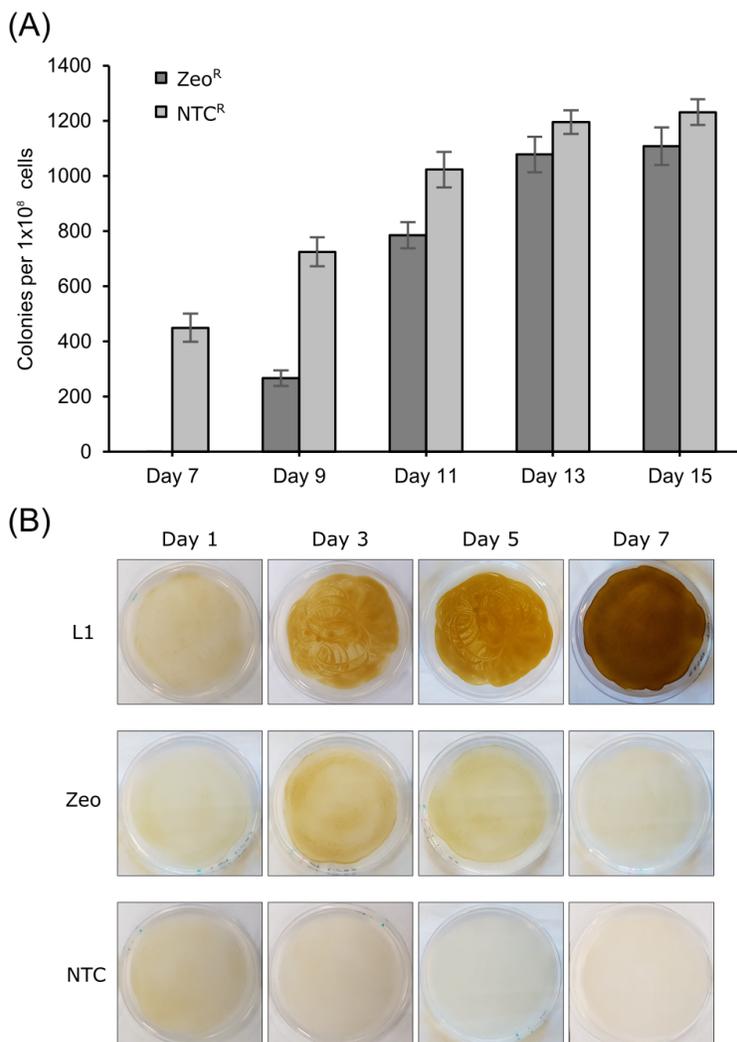


**Figure 2.7: Comparison of *Sh ble* and *nat* as selectable markers for *P. tricornutum* conjugation.** (A) Bar plot shows number of *P. tricornutum* colonies obtained over 15 days from conjugations with pPtGE30 (ZeoR) or pPtGE31 (NTCR) plasmids. Error bars show standard error of the mean for six replicates. (B) Plates containing 100 $\mu$g ml$^{-1}$ nourseothricin (NTC) and are more effective at selecting against wild-type *P. tricornutum* than plates containing 50 $\mu$g ml$^{-1}$ Zeocin™ (Zeo). Wild-type *P. tricornutum* was plated on nonselective media (L1) as a control.

### 2.3.4   Simultaneous maintenance of multiple plasmids in *P. tricornutum*

Stable replication of multiple plasmids in the same organism is useful for probing basic biological processes such as protein-protein and protein-DNA interactions (42–44), or for use as reporter readouts of gene expression (45). Multiple plasmids also offers advantages in the design, construction, and testing of synthetic pathways (46–48). To investigate whether *P. tricornutum* can stably maintain two plasmids (each using identical centromere sequences), we conjugated p0521s encoding the *Sh ble* gene from *E. coli* to *P. tricornutum*, followed by conjugation of pPtGE31 encoding the *nat* marker to *P. tricornutum* containing p0521s (Supplementary Figure B.1). As shown in Figure 2.8A, *P. tricornutum* cells harboring both plasmids survived on media containing either Zeocin™ or nourseothricin, and on media with both antibiotics. To demonstrate that both p0521s and pPtGE31 were present in exconjugants, plasmid DNA isolated from individual *P. tricornutum* colonies selected on plates containing nourseothricin and Zeocin™ was independently transformed into *E. coli*. Here, the plasmid DNA is expected to be a mixture of p0521s and pPtGE31. Because the efficiency of co-transformation is low, and p0521s and pPtGE31 possess the same bacterial origin of replication, individual *E. coli* transformants would contain either p0521s and pPtGE31, but not both. Plasmids isolated from 20 *E. coli* transformants showed no obvious signs of rearrangements as judged by diagnostic restriction digest (Figure 2.8B). To determine the identity of each plasmid, we used a multiplex PCR strategy with primers specific for the *nat* and *Sh ble* genes, and found that p0521s (*Sh ble*) was approximately 3-fold more abundant than pPtGE31 (*nat*) (Figure 2.8C, Supplementary Table B.2). We also used qPCR to assess plasmid copy number from total DNA extracts of two *P. tricornutum* clones, each containing p0521s and pPtGE31, and found that the copy number of p0521s to pPtGE31 was approximately 3:1 in both clones (Figure 2.9), agreeing with the multiplex PCR results in Figure 2.8C. This ratio of p0521s to pPtGE31 is higher than the 1:1 ratio expected if there is no replication bias of the plasmid in *P. tricornutum* or *E. coli*, but nonetheless demonstrates that *P. tricornutum* can be sequentially conjugated with *E. coli* to stably replicate two independent plasmids expressing different antibiotic resistance genes.
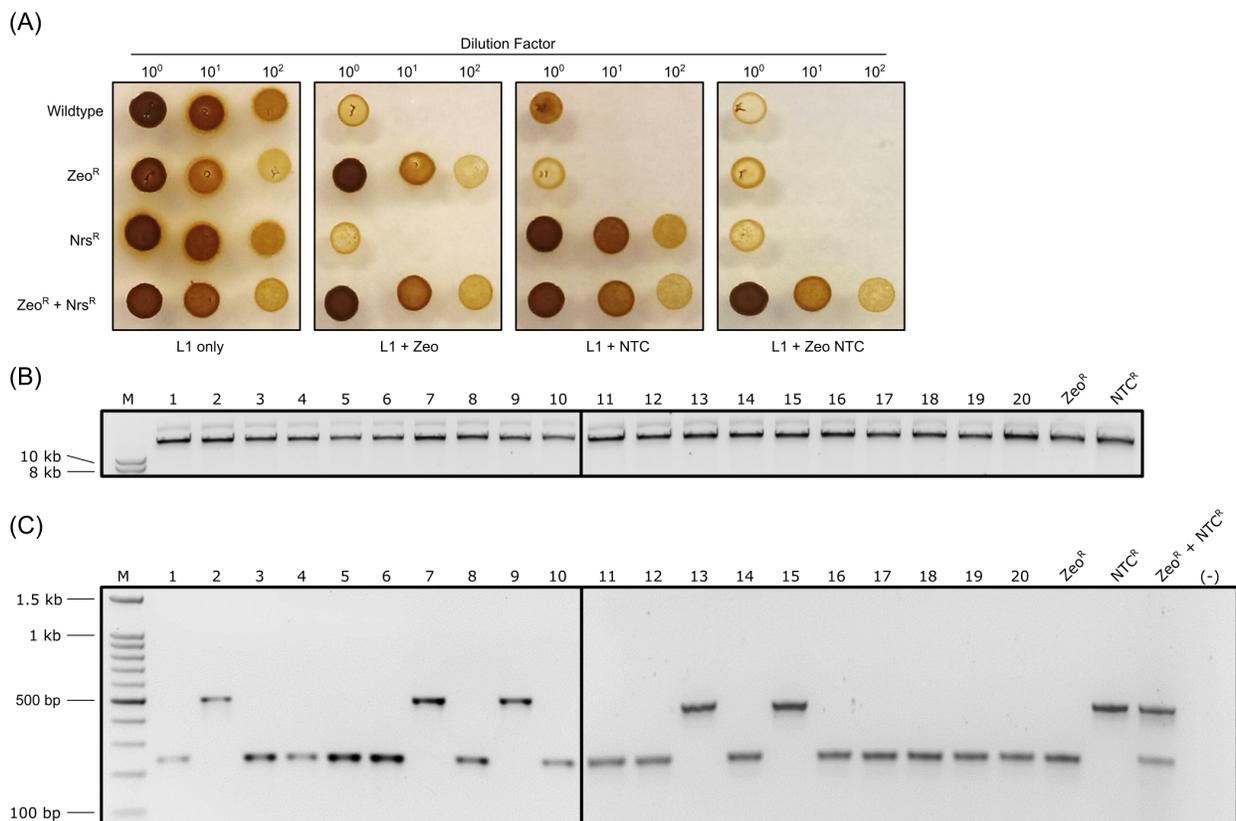
**Figure 2.8: Stable maintenance of two plasmids in *P. tricornutum* imparts dual antibiotic resistance.** (A) *P. tricornutum* possessing both p0521s (ZeoR) and pPtGE31 (NTCR) survive on media containing Zeocin™ (Zeo) and nourseothricin (NTC) antibiotics. (B) Agarose gel of plasmids recovered from one *P. tricornutum* two-plasmid exconjugant and linearized by digestion with I-CeuI. M, 1 kb ladder. (C) Agarose gel of multiplex PCR products showing the identity of 20 plasmids recovered from one *P. tricornutum* two-plasmid exconjugant. M, 100 bp ladder.
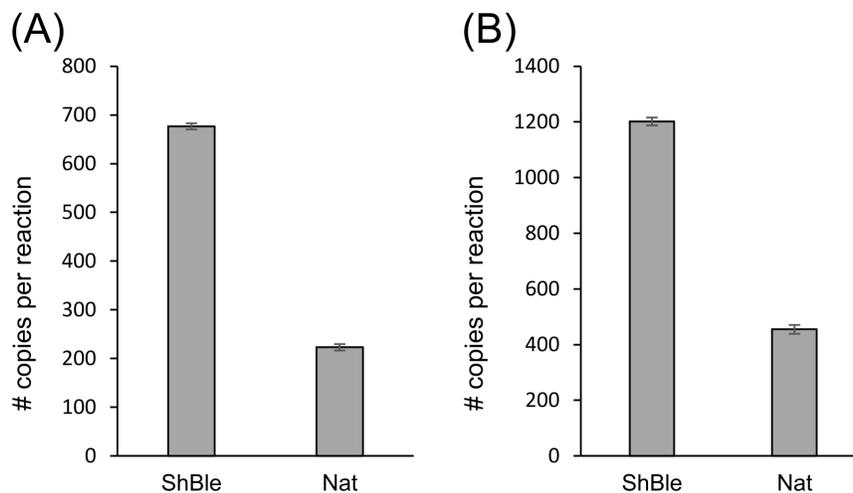


**Figure 2.9: Copy number of two plasmids in *P. tricornutum* DNA extracts determined by qPCR.** Plasmids were extracted from two *P. tricornutum* clones containing p0521s and pPtGE31, diluted 1:10000, and used as template for qPCR reactions. (A) Copy number of Shble and Nat for clone 1. (B) Copy number of ShBle and Nat for clone 2. ShBle, the *Sh ble* gene on p0521s; Nat, the *nat* gene on pPtGE31. Error bars show standard error of the mean for three replicates.

### 2.3.5 CRISPR/Cas9 gene editing in *P. tricornutum* using a stably replicating plasmid-based system

Encouraged by our results and previous demonstration of stable replication of plasmids expressing exogenous genes in *P. tricornutum* (16), we developed a simple and robust conjugative plasmid-based protocol for genome editing using the Cas9 protein derived from the *Streptococcus pyogenes* type II CRISPR system (49), and a TevCas9 dual nuclease variant (50) (Figure 2.10A). As proof-of-principle, we designed 4 different single guideRNAs (sgRNAs) against the *P. tricornutum* urease gene. Each sgRNA, designed to take into account the 5'-CNNNG-3' cleavage motif targeting requirements of the TevCas9 dual nuclease (50), could be used for gene editing with Cas9 or TevCas9. After conjugative delivery and selection of exconjugants, editing efficiency was determined by T7 endonuclease I (T7EI) mismatch cleavage assays on fragments of the urease gene directly amplified from *P. tricornutum* colonies by PCR. We observed a 10–60% editing rate depending on the sgRNA and nuclease used (Figure 2.10B). To determine if the *P. tricornutum* colonies positive by the T7EI assay were monoallelic or mosaics, sub-clones were isolated and analyzed in two ways. First, dilutions of each sub-clone were plated on agar-containing plates with nitrate or urea as a nitrogen source, with the expectation that clones with a knockout of the urease gene (or homozygous knockout) would be unable to grow on plates with urea as the sole nitrogen source (Figure 2.10C). This screen identified 15 sub-clones with a urea-deficient growth phenotype. Second, the urease gene of the 15 sub-clones identified by this phenotypic screen were analyzed by sequencing of PCR products. In 12 of 15 sub-clones analyzed, a single genotype was evident from the sequencing reads (Figure 2.10D), with different length deletions observed as would be expected by mutagenic non-homologous end-joining repair. Interestingly, 3 sub-clones each possessed two different genotypes, suggesting different repair events at each copy of the urease gene. In 9 of 18 sequences, Cas9 or TevCas9 editing created a -2 frameshift that would result in a truncated or non-functional urease protein, 6 of 18 sequences revealed a -1 frameshift, 2 of 18 revealed in frame deletions that disrupted urease function, and a single sequence revealed a +1 frameshift. Performing long-range PCR of the full urease ORF of the knockouts, followed by sequencing of the PCR products, revealed that 7 of the 12 strains that were originally thought to contain homozygous mutations actually possessed large deletions in the second copy of the urease gene which could

not be detected using our previous screening method (Figure 2.11). The high proportion of strains with large deletions as a result of NHEJ repair of Cas9 cleavage events suggests that this may be a common DSB repair outcome in *P. tricornutum*.
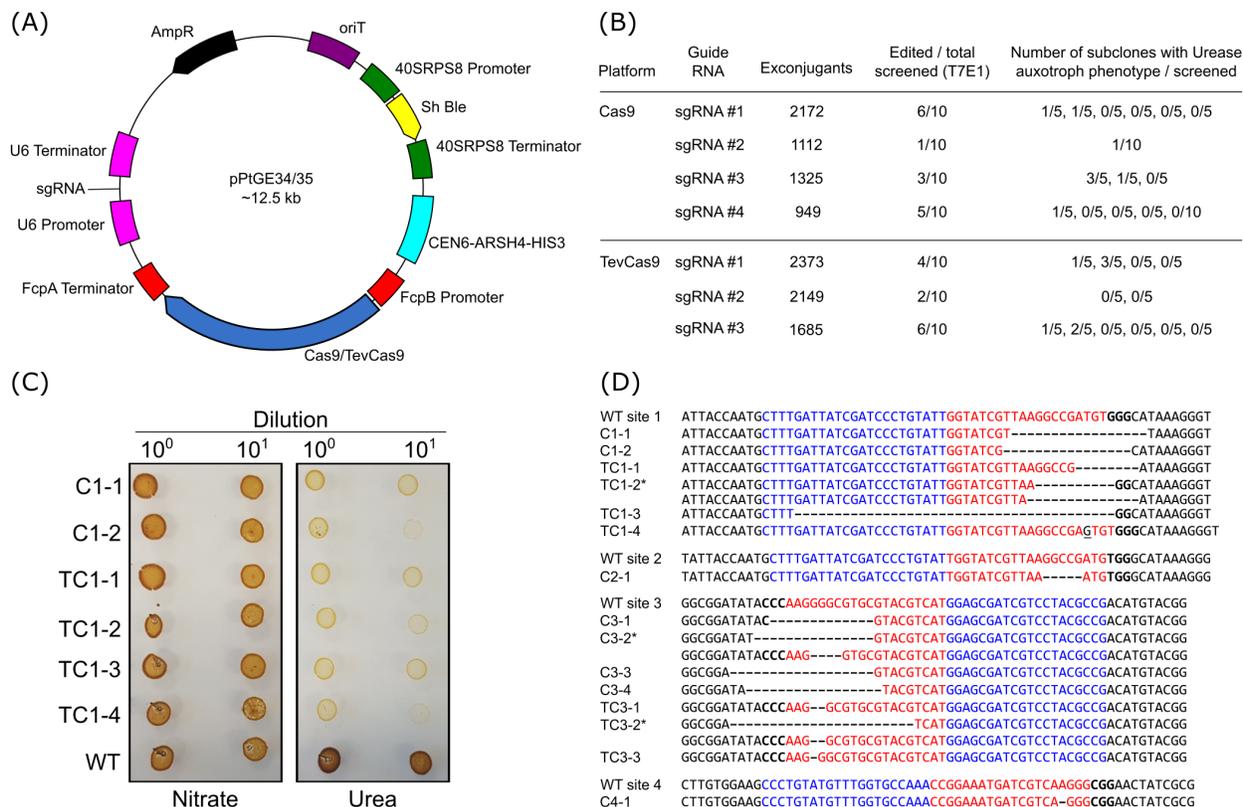


**Figure 2.10: Editing of the *P. tricornutum* urease gene by Cas9 and TevCas9.** (A) Diagram of pPtGE34 and pPtGE35. (B) Table of experimental workflow and number of urease knockouts generated by Cas9 or TevCas9 with each sgRNA. Exconjugants were screened by T7EI assay to identify editing at the urease sgRNA target site. Colonies that showed editing were diluted in L1 media and plated to obtain subclones. Five or ten subclones from each were subsequently screened for urease knockout phenotypes by streaking onto media containing nitrate or urea. (C) Representative images of *P. tricornutum* urease knockout strains plated on L1 media supplemented with nitrate or urea. Biallelic urease knockouts are unable to survive on plates containing urea as the sole nitrogen source. (D) Sequences of urease sgRNA target sites in *P. tricornutum* urease knockout strains. Red text indicates the sgRNA target sequence, blue text indicates the Tev target sequence, bold text indicates the PAM motif, underlined text indicates insertions, and dashes (-) indicate deletions. Asterisks (*) next to strain names indicate subclones that contain two different genotypes, suggesting different repair events at each copy of the urease gene.

Previous groups showed Cas9-mediated editing of diatoms using biolistic bombardment with non-replicating plasmids expressing Cas9 (28, 51). One drawback of this methodology is that the plasmids are stably integrated into the genome, with (as of yet) no mechanism to prevent long-term Cas9 expression. To demonstrate that we could cure the Cas9 or TevCas9 expressing plasmids from urease-edited *P. tricornutum*, we grew sub-clones under non-selective conditions (media lacking
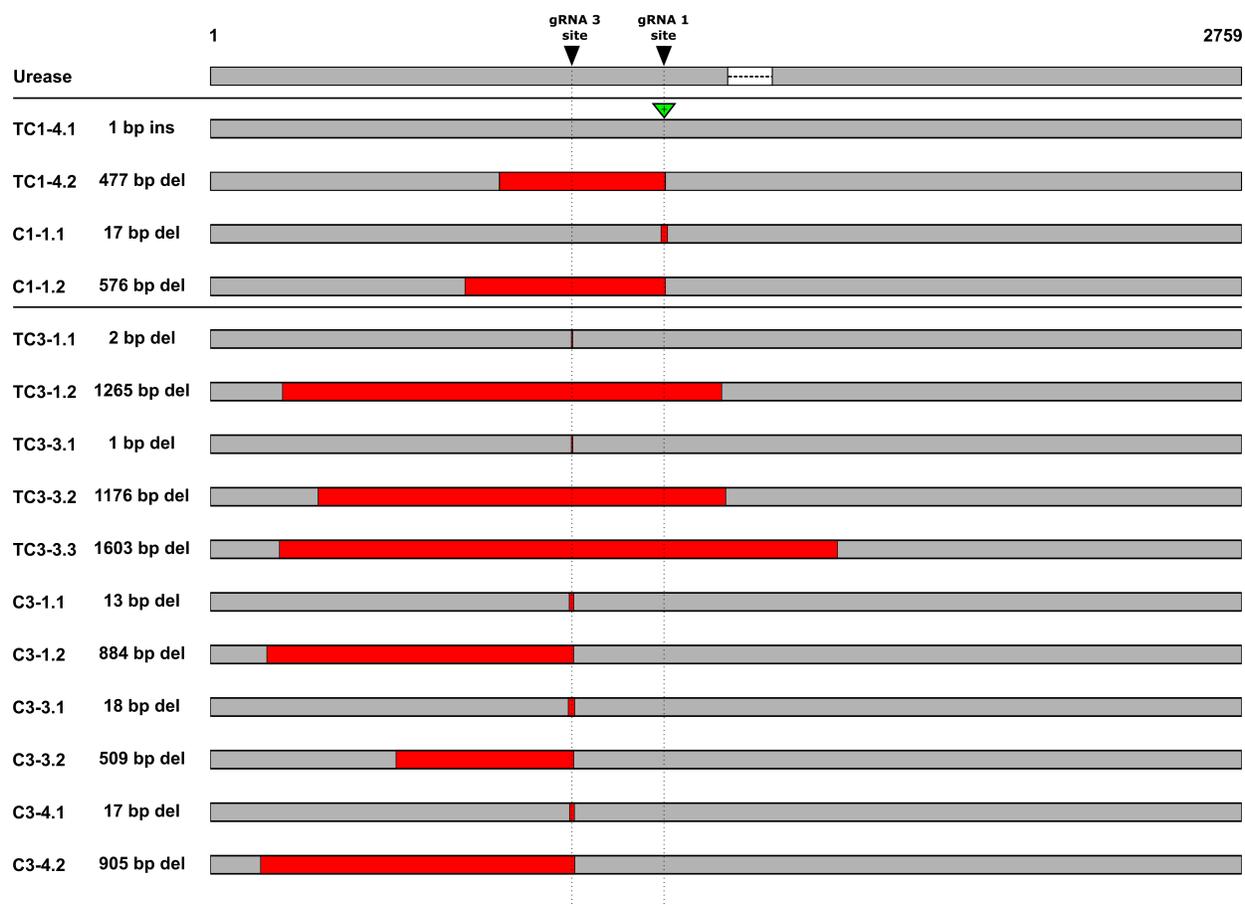
**Figure 2.11: Large deletions in the _P. tricornutum_ urease gene revealed by sequencing.** Graphical map of the position and extent of indels for each of the urease knockouts relative to the wild-type urease gene (shown at top). Red rectangles indicate nucleotide deletions, green triangles indicate nucleotide insertions, and white rectangles with dashed lines represent introns.

Zeocin™) for 9 generations, and streaked colonies onto plates containing or lacking Zeocin™. As shown in Figure 2.12, 3 of 18 sub-clones displayed phenotypes consistent with Zeocin™ sensitivity. To provide further evidence for plasmid loss, multiplex PCR was successful for a genomic target (actin), but not a plasmid-based target (*Sh ble*) for colonies that showed Zeocin™ sensitivity (Figure 2.12b).
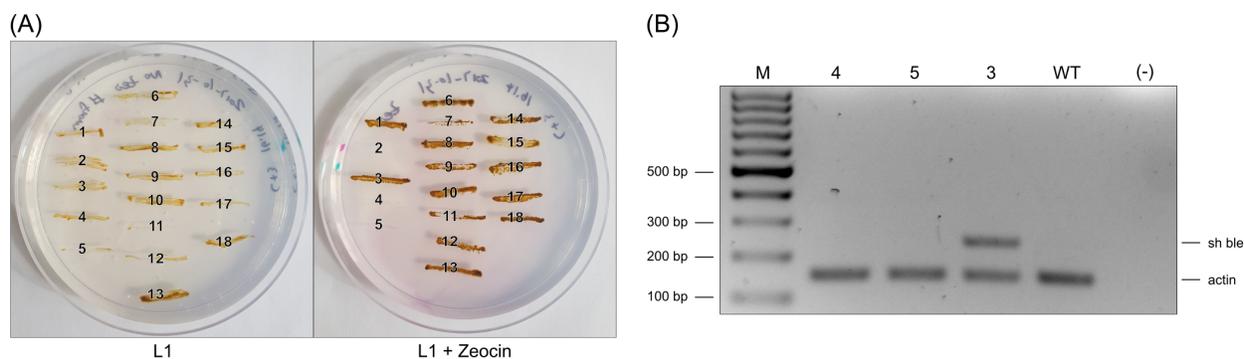


**Figure 2.12: Isolation of urease knockout clones cured of the pPtGE34 plasmid.** (A) Representative image of plasmid-cured *P. tricornutum* urease knockout strains. *P. tricornutum* containing stable bi-allelic knockouts of the urease gene confirmed by sequencing and phenotypic analysis were grown in nonselective media for 1 week before plating to obtain single colonies. Colonies were replica streaked onto Zeocin™ (L1 + Zeocin) and nonselective (L1) plates to identify clones that were cured of the pPtGE34 plasmid (colonies 2, 4, and 5). (B) Multiplex PCR to detect *Sh ble* and actin in urease knockout colonies. Urease knockout colonies identified as Zeocin™ sensitive were lysed in TE and used as template in a multiplex PCR to screen for the presence of the pPtGE34 plasmid. A colony resistant to Zeocin™ (colony 3) and wild-type *P. tricornutum* were used as controls. M, 100 bp ladder; 4, colony 4 from panel (a); 5, colony 5 from panel (a); 3, colony 3 from panel (a); WT, wild-type *P. tricornutum*; (-), no template control.

### 2.3.6 Stable replication of a plasmid-based vanillin biosynthesis pathway in

*P. tricornutum*

To highlight the synthetic biology potential of our plasmid-based tools, we next tested whether *P. tricornutum* can maintain a putative synthetic metabolic pathway on a replicating plasmid over an extended growth period. We assembled components of a vanillin biosynthetic pathway (52) in the p0521s plasmid (creating p0521s-V) (Figure 2.13A). The pathway consists of eight genes derived from *Zea mays* (corn), *Arabidopsis thaliana* (mouse-ear cress), and *Vanilla planifolia* (vanilla orchid) under control of *P. tricornutum* promoters and terminators. The pathway is designed to metabolize phenylalanine and tyrosine, through a series of intermediates, into ferulic acid (53). The eight genes were organized into four pairs that mimics proposed steps in the vanillin biosynthesis pathway (52). Each gene pair was cloned in-frame with an intervening 2A self-cleaving peptide (41) to coordinately express each protein.

As a precursor to detailed analyses of mRNA expression, protein stability, and metabolic intermediates, we first examined long-term stability and replication fidelity of the plasmids in *P. tricornutum*, as faithful long-term maintenance would be necessary for robust synthetic output. The pathway was assembled in yeast and moved to *E. coli* where they were screened for correctly assembled clones. From 54 clones screened, 15 correct clones were identified as assayed by multiplex PCR screen. Six of these were analytically digested which identified a total of four correct clones. These four individual clones (called V1, V2, V3, V4) were conjugated from *E. coli* into *P. tricornutum*, then two exconjugants from each transformation were selected and grown in liquid culture. Next, plasmids were recovered in *E. coli* after ~1 and 4 months (40 and 120 days) of serial culture passage in *P. tricornutum*, representing approximately 50 and 160 *P. tricornutum* generations, respectively (Supplementary Table B.3). Multiplex PCR of plasmids recovered from *P. tricornutum* cultures after ~1 and 4 months showed that all 8 metabolic genes were present (Figure 2.13C), and the recovered plasmids showed no major rearrangements by diagnostic restriction digest (Figure 2.13D). To more accurately assess replication fidelity in *P. tricornutum*, we completely sequenced two plasmid clones after yeast assembly and initial cloning in *E. coli* (V1.0 and V3.0), and after ~1 month (V1.1 and V3.1) and 4 months (V1.4 and V3.4) of propagation in *P. tricornutum* (Figure 2.13E). We identified several point mutations that differ between the ini-

tial sequences of clones V1.0 and V3.0 and that possibly occurred during yeast assembly (Figure 2.13E). Additionally, we found several point mutations that appeared in each clone after ~1 month of passage in *P. tricornutum*, but no further mutations had accumulated by the 4-month endpoint (Figure 2.13E, Supplementary Table B.4). This finding suggests that mutations in V1.1 and V3.1 occurred during or shortly after conjugation, but no new mutations accumulated once the plasmids were stably replicating in *P. tricornutum*. The majority of mutations were in the non-coding regions of the plasmid, including the plasmid backbone and promoter and terminator elements. However, some point mutations were identified in coding regions of the genes encoding enzymes of the biosynthetic pathway. Notably, a 1-bp insertion at nucleotide position 812 was identified in the 4-hydroxycinnamoyl-CoA ligase (4CL) coding region in the 1- and 4-month time points of V3 (V3.1 and V3.4). This 1-bp insertion results in a +1 reading frame shift creating a premature stop codon at position 855 of the coding region (numbered according to the wild-type sequence) that results in a truncated protein missing the last 281 amino acid residues.
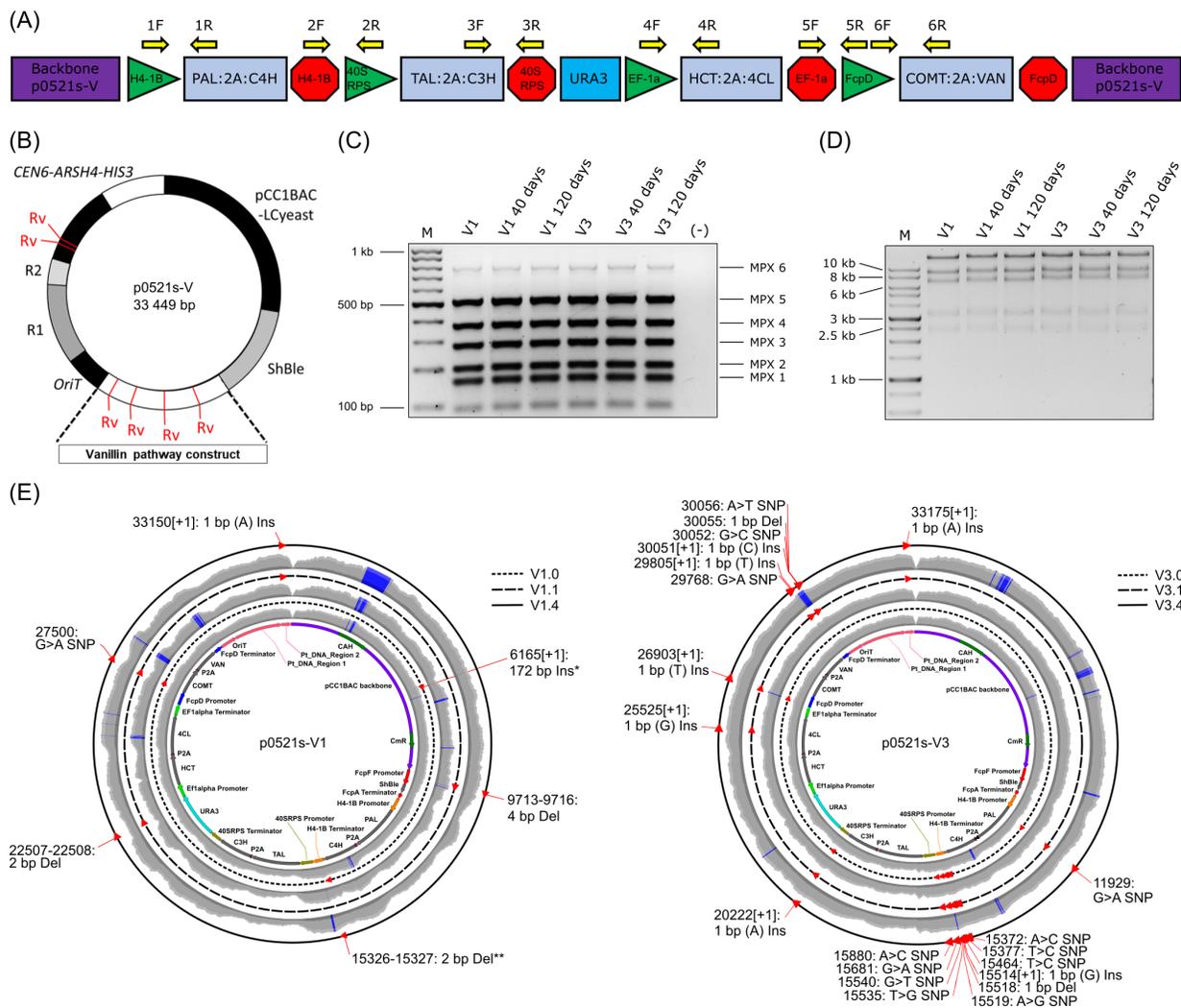
**Figure 2.13: Stable maintenance of a plasmid-based biosynthetic pathway in *P. tricornutum*.** (A) Diagram of the putative vanillin biosynthetic pathway (not to scale). Yellow arrows represent multiplex PCR primer pair locations. PAL, phenylalanine ammonia lyase; C4H, *trans*-cinnamate 4-hydroxylase; TAL, tyrosine ammonia lyase; C3H, p-coumarate 3-hydroxylase; HCT, hydroxycinnamoyl transferase; 4CL, 4-hydroxycinnamoyl-CoA ligase; COMT, caffeic acid 3-O-methyltransferase; VAN, vanillin synthase. 2A, P2A self-cleaving peptide. Promoters are indicated by right-facing green arrows, and terminators by red hexagons. (B) Diagram of p0521s-V. Rv; EcoRV sites. (C) Agarose gel electrophoresis of multiplex PCR showing presence of metabolic genes in vanillin biosynthesis plasmids recovered from *P. tricornutum*. V1, p0521s-V clone 1; and V3, p0521s-V clone 3. MPX 1–6, multiplex PCR products generated by the corresponding primer pairs shown in panel (a). M, 100 bp ladder (D) Agarose gel electrophoresis of vanillin biosynthesis plasmids recovered from *P. tricornutum* and digested with EcoRV. M, 1 kb ladder. (E) Plasmid maps depicting sequencing coverage and positions of SNPs. Constructs were fully sequenced after initial assembly (V1.0 and V3.0) and after ~1 month (V1.1 and V3.1) and ~4 months (V1.4 and V3.4) of serial passage in *P. tricornutum*. The inner ring is the reference plasmid map, and outer rings indicate the plasmid sequence for each time point. Red triangles indicate positions of mutation relative to the reference sequence. Sequencing coverage for each construct is shown in gray, with blue shading indicating regions where sequencing reads exceed a set maximum: V1.0 = 1000 reads; V1.1 = 400 reads; V1.4 = 1200 reads; V3.0 = 900 reads; V3.1 = 900 reads; V3.4 = 1100 reads. *172 bp insertion was identified in a region of repetitive sequence in V1.0 but was not present in later time points and is likely a sequencing contig assembly error. **2 bp deletion was identified in V1.0 and V1.4 but was not present in V1.1 (sequence coverage at these positions in V1.1 is 0).

## 2.4   Discussion

The model diatom *P. tricornutum* is an attractive species for synthetic biology applications, but requires further development to be suitable for industrial production applications. Toward this, we have developed novel genetic tools and an efficient gene editing system for *P. tricornutum* that will enable more complex genetic engineering of this species. The novel regulatory elements described here are important for three reasons. First, we have expanded the repertoire of promoter and terminator elements that function to express an exogenous gene in *P. tricornutum*. Second, our data imply that pairing promoters with different combinations of terminators could be used to modulate expression of individual genes, as might be required in a synthetic metabolic pathway. Third, the strategy of using a eukaryotic spliceosomal intron to limit or prevent expression in *E. coli* would be useful when cloning *P. tricornutum* genes or during construction of synthetic pathways where very low levels of expression could result in toxicity or plasmid instability in *E. coli*. In addition, we have shown that Zeocin™ and nourseothricin can be used for selection of plasmid based systems, and suggest a cost-saving measure in that less antibiotic can be used for selection of *P. tricornutum* exconjugants than what is currently used for selection of integrated copies of the *nat* gene (15).

Examples of Cas9 gene editing in *P. tricornutum* previously relied on genomic integration of the expression construct (28, 54, 55). This results in prolonged exposure to the nuclease and potentially increases the risk of off-target effects. We have demonstrated simple and robust editing of the *P. tricornutum* genome using a conjugative plasmid system that can be cured after editing to avoid complications related to long-term Cas9 expression.

Using several genetic regulatory elements described in this study, we created a plasmid-based multi-gene biosynthetic pathway for expression in *P. tricornutum*. Tracking plasmid stability over 160 generations revealed evidence that suggests mutations occurred during initial conjugation, but no new mutations were accumulated during plasmid propagation. These results highlight the importance of validating plasmid constructs by whole-plasmid sequencing after transformation before assaying for biosynthetic activity. Furthermore, our data show that while mutations can occur during or shortly after conjugation, plasmids encoding large biosynthetic pathways can be stably replicated in *P. tricornutum* over extended periods.

In summary, we have described the development and testing of several plasmid-based tools for expression and regulation of genes in *P. tricornutum*. The addition of a multi-plasmid system with new genetic elements will facilitate heterologous gene expression, study of protein-protein interactions, reporter readouts of gene expression, and stable maintenance of plasmid-based synthetic metabolic pathways in this species. When combined with the simple and robust plasmid-based Cas9 genome-editing systems described here, we anticipate that our tools will open the door for construction of designer *P. tricornutum* knockout strains for synthetic biology applications, and for the study of the basic biology of this important species.

## 2.5   References

[1] Norton, T. A.; Melkonian, M.; Andersen, R. A. Algal biodiversity. *Phycologia* **1996**, *35*, 308–326.

[2] Field, C. B. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **1998**, *281*, 237–240.

[3] Martin-Jezequel, V.; Hildebrand, M.; Brzezinski, M. A. Silicon metabolism in diatoms: Implications for growth. *Journal of Phycology* **2000**, *36*, 821–840.

[4] Bozarth, A.; Maier, U. G.; Zauner, S. Diatoms in biotechnology: Modern tools and applications. 2009.

[5] Scaife, M. A.; Smith, A. G. Towards developing algal synthetic biology. *Biochemical Society Transactions* **2016**, *44*, 716–22.

[6] Yongmanitchai, W.; Ward, O. P. Growth of and omega-3 fatty acid production by Phaeodactylum tricornutum under different culture conditions. *Applied and Environmental Microbiology* **1991**, *57*, 419–425.

[7] Bowler, C. et al. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **2008**, *456*, 239–44.

[8] Armbrust, E. V. et al. The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism. *Science* **2004**, *306*, 79–86.

[9] Zaslavskaia, L. A.; Casey Lippmeier, J.; Kroth, P. G.; Grossman, A. R.; Apt, K. E. Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a variety of selectable marker and reporter genes. *Journal of Phycology* **2000**, *36*, 379–386.

[10] Siaut, M.; Heijde, M.; Mangogna, M.; Montsant, A.; Coesel, S.; Allen, A.; Manfredonia, A.; Falciatore, A.; Bowler, C. Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum. *Gene* **2007**, *406*, 23–35.

[11] Niu, Y. F.; Yang, Z. K.; Zhang, M. H.; Zhu, C. C.; Yang, W. D.; Liu, J. S.; Li, H. Y. Transformation of diatom Phaeodactylum tricornutum by electroporation and establishment of inducible selection marker. *BioTechniques* **2012**, *52*, 1–3.

[12] Apt, K. E.; Kroth-Pancic, P. G.; Grossman, A. R. Stable nuclear transformation of the diatom Phaeodactylum tricornutum. *Molecular and General Genetics* **1996**, *252*, 572–579.

[13] Kira, N.; Ohnishi, K.; Miyagawa-Yamaguchi, A.; Kadono, T.; Adachi, M. Nuclear transformation of the diatom Phaeodactylum tricornutum using PCR-amplified DNA fragments by microparticle bombardment. *Marine Genomics* **2016**, *25*, 49–56.

[14] Taylor, N. J.; Fauquet, C. M. Microparticle bombardment as a tool in plant science and agricultural biotechnology. *DNA and Cell Biology* **2002**, *21*, 963–977.

[15] Daboussi, F.; Leduc, S.; Maréchal, A.; Dubois, G.; Guyot, V.; Perez-Michaut, C.; Amato, A.; Falciatore, A.; Juillerat, A.; Beurdeley, M.; Voytas, D. F.; Cavarec, L.; Duchateau, P. Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology. *Nature Communications* **2014**, *5*, 3831.

[16] Karas, B. J.; Diner, R. E.; Lefebvre, S. C.; McQuaid, J.; Phillips, A. P.; Noddings, C. M.; Brunson, J. K.; Valas, R. E.; Deerinck, T. J.; Jablanovic, J., et al. Designer diatom episomes delivered by bacterial conjugation. *Nature Communications* **2015**, *6*, 1–10.

[17] Diner, R. E.; Bielinski, V. A.; Dupont, C.; Allen, A. E.; Weyman, P. D. Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-based DNA Delivery Methods. *Frontiers in Bioengineering and Biotechnology* **2016**, *4*, 65.

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

[19] Strand, T. A.; Lale, R.; Degnes, K. F.; Lando, M.; Valla, S. A new and improved host-independent plasmid system for RK2-based conjugal transfer. *PLoS ONE* **2014**, *9*, 1–6.

[20] Erdene-Ochir, E.; Shin, B. K.; Huda, M. N.; Kim, D. H.; Lee, E. H.; Song, D. G.; Kim, Y. M.; Kim, S. M.; Pan, C. H. Cloning of a novel endogenous promoter for foreign gene expression in Phaeodactylum tricornutum. *Applied Biological Chemistry* **2016**, 1–7.

[21] Martino, A. D.; Meichenin, A.; Shi, J.; Pan, K.; Bowler, C. Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions. *Journal of Phycology* **2007**, *43*, 992–1009.

[22] Miyagawa, A.; Okami, T.; Kira, N.; Yamaguchi, H.; Ohnishi, K.; Adachi, M. Research note: High efficiency transformation of the diatom Phaeodactylum tricornutum with a promoter from the diatom Cylindrotheca fusiformis. *Phycological Research* **2009**, *57*, 142–146.

[23] Poulsen, N.; Kröger, N. A new molecular tool for transgenic diatoms: Control of mRNA and protein biosynthesis by an inducible promoter-terminator cassette. *FEBS Journal* **2005**, *272*, 3413–3423.

[24] Seo, S.; Jeon, H.; Hwang, S.; Jin, E.; Chang, K. S. Development of a new constitutive expression system for the transformation of the diatom Phaeodactylum tricornutum. *Algal Research* **2015**, *11*, 50–54.

[25] Kadono, T.; Miyagawa-Yamaguchi, A.; Kira, N.; Tomaru, Y.; Okami, T.; Yoshimatsu, T.; Hou, L.; Ohama, T.; Fukunaga, K.; Okauchi, M.; Yamaguchi, H.; Ohnishi, K.; Falciatore, A.; Adachi, M. Characterization of marine diatom-infecting virus promoters in the model diatom Phaeodactylum tricornutum. *Scientific Reports* **2015**, *5*, 1–13.

[26] Noskov, V. N.; Karas, B. J.; Young, L.; Chuang, R. Y.; Gibson, D. G.; Lin, Y. C.; Stam, J.; Yonemoto, I. T.; Suzuki, Y.; Andrews-Pfannkoch, C.; Glass, J. I.; Smith, H. O.; Hutchison, C. A.; Venter, J. C.; Weyman, P. D. Assembly of large, high G+C bacterial DNA fragments in yeast. *ACS Synthetic Biology* **2012**, *1*, 267–273.

[27] Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. a.; Smith, H. O.; Iii, C. A. H.; America, N. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **2009**, *6*, 343–5.

[28] Nymark, M.; Sharma, A. K.; Sparstad, T.; Bones, A. M.; Winge, P. A CRISPR/Cas9 system adapted for gene editing in marine algae. *Scientific Reports* **2016**, *6*, 24951.

[29] Engler, C.; Kandzia, R.; Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **2008**, *3*, e3647.

[30] Alikhan, N.-F.; Petty, N. K.; Ben Zakour, N. L.; Beatson, S. A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **2011**, *12*, 402.

[31] Devaki, B.; Grossman, A. R. Characterization of gene clusters encoding the fucoxanthin chlorophyll proteins of the diatom phaeodactylum tricornutum. *Nucleic Acids Research* **1993**, *21*, 4458–4466.

[32] Falciatore, A.; Casotti, R.; Leblanc, C.; Abrescia, C.; Bowler, C. Transformation of Nonselectable Reporter Genes in Marine Diatoms. *Marine Biotechnology (New York, N.Y.)* **1999**, *1*, 239–251.

[33] Coesel, S.; Mangogna, M.; Ishikawa, T.; Heijde, M.; Rogato, A.; Finazzi, G.; Todo, T.; Bowler, C.; Falciatore, A. Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. *EMBO Reports* **2009**, *10*, 655–661.

[34] Hopkinson, B. M.; Dupont, C. L.; Allen, A. E.; Morel, F. M. M. Efficiency of the $CO_2$-concentrating mechanism of diatoms. *Proceedings of the National Academy of Sciences* **2011**, *108*, 3830–3837.

[35] Wolfs, J. M.; Dasilva, M.; Meister, S. E.; Wang, X.; Schild-Poulter, C.; Edgell, D. R. MegaT-evs: Single-chain dual nucleases for efficient gene disruption. *Nucleic Acids Research* **2014**, *42*, 8816–8829.

[36] Lumbreras, V.; Stevens, D. R.; Purton, S. Efficient foreign gene expression in Chlamydomonas reinhardtii mediated by an endogenous intron. *Plant Journal* **1998**, *14*, 441–447.

[37] Zhang, C.; Hu, H. High-efficiency nuclear transformation of the diatom Phaeodactylum tricornutum by electroporation. *Marine Genomics* **2014**, *16*, 63–66.

[38] Miyahara, M.; Aoi, M.; Inoue-Kashino, N.; Kashino, Y.; Ifuku, K. Highly Efficient Transformation of the Diatom Phaeodactylum tricornutum by Multi-Pulse Electroporation. *Bioscience, Biotechnology, and Biochemistry* **2013**, *77*, 874–876.

[39] Krügel, H.; Fiedler, G.; Smith, C.; Baumberg, S. Sequence and transcriptional analysis of the nourseothricin acetyltransferase-encoding gene nat1 from Streptomyces noursei. *Gene* **1993**, *127*, 127–131.

[40] Shaw, W. V.; Packman, L. C.; Burleigh, B. D.; Dell, A.; Morris, H. R.; Hartley, B. S. Primary structure of a chloramphenicol acetyltransferase specified by R plasmids. 1979.

[41] Radcliffe, P. A.; Mitrophanous, K. A. Multiple gene products from a single vector: 'self-cleaving' 2A peptides. *Gene Therapy* **2004**, *11*, 1673–1674.

[42] Sobhanifar, S. Yeast Two Hybrid Assay: A Fishing Tale. *BioTeach Journal* **2003**, *1*, 81–88.

[43] Hu, C. D.; Grinberg, A. V.; Kerppola, T. K. Visualization of protein interactions in living cells using bimolecular fluorescence complementation (BiFC) analysis. *Curr Protoc Protein Sci* **2005**, *Chapter 19*, Unit 19 10.

[44] Sun, N.; Zhao, H. A two-plasmid bacterial selection system for characterization and engineering of homing endonucleases. *Methods in Molecular Biology* **2014**, *1123*, 87–96.

[45] Nacken, V.; Achstetter, T.; Degryse, E. Probing the limits of expression levels by varying promoter strength and plasmid copy number in Saccharomyces cerevisiae. *Gene* **1996**, *175*, 253–260.

[46] DeJong, J. M.; Liu, Y.; Bollon, A. P.; Long, R. M.; Jennewein, S.; Williams, D.; Croteau, R. B. Genetic engineering of taxol biosynthetic genes in Saccharomyces cerevisiae. *Biotechnology and Bioengineering* **2006**, *93*, 212–224.

[47] Schmidt-Dannert, C.; Umeno, D.; Arnold, F. H. Molecular breeding of carotenoid biosynthetic pathways. *Nature Biotechnology* **2000**, *18*, 750–753.

[48] Beyer, P.; Al-Babili, S.; Ye, X.; Lucca, P.; Schaub, P.; Welsch, R.; Potrykus, I. Golden Rice: introducing the beta-carotene biosynthesis pathway into rice endosperm by genetic engineering to defeat vitamin A deficiency. *The Journal of Nutrition* **2002**, *132*, 506S–510S.

[49] Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptice Bacterial Immunity. *Science (New York, N.Y.)* **2012**, *337*, 816–822.

[50] Wolfs, J. M.; Hamilton, T. A.; Lant, J. T.; Laforet, M.; Zhang, J.; Salemi, L. M.; Gloor, G. B.; Schild-Poulter, C.; Edgell, D. R. Biasing genome-editing events toward precise length deletions with an RNA-guided TevCas9 dual nuclease. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 14988–14993.

[51] Hopes, A.; Nekrasov, V.; Kamoun, S.; Mock, T. Editing of the urease gene by CRISPR-Cas in the diatom Thalassiosira pseudonana. *Plant Methods* **2016**, *12*, 49.

[52] Gallage, N. J.; Møller, B. L. Vanillin-bioconversion and bioengineering of the most popular plant flavor and its de novo biosynthesis in the vanilla orchid. 2015.

[53] Yang, H.; Barros-Rios, J.; Kourteva, G.; Rao, X.; Chen, F.; Shen, H.; Liu, C.; Podstolski, A.; Belanger, F.; Havkin-Frenkel, D.; Dixon, R. A. A re-evaluation of the final step of vanillin biosynthesis in the orchid Vanilla planifolia. *Phytochemistry* **2017**, *139*, 33–46.

[54] Stukenberg, D. D.; Zauner, S. S.; Dell'Aquila, G.; Maier, U. G. Optimizing CRISPR/Cas9 for the diatom Phaeodactylum tricornutum. *Frontiers in Plant Science* **2018**, *9*, 740.

[55] Allorent, G.; Guglielmino, E.; Giustini, C.; Courtois, F. *Plastids*; Springer, 2018; pp 367–378.

# Chapter 3

# Plasmid-based complementation of large deletions in *Phaeodactylum tricornutum* biosynthetic genes generated by Cas9 editing

The work presented in this chapter is reprinted (adapted) with permission (Appendix A) from:

S. S. Slattery, H. Wang, D. J. Giguere, C. Kocsis, B. L. Urquhart, B. J. Karas, 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.

Copyright (2020) Springer Nature

## 3.1   Introduction

Photoautotrophic microalgae and cyanobacteria are emerging as alternative platforms for synthetic biology applications (1, 2). One microalgae species of interest is the diploid marine diatom *P. tricornutum*. A variety of plasmid-based genetic tools have been developed for *P. tricornutum* that facilitate basic molecular manipulations and expression of complex synthetic pathways (3–6). We, and others, have developed plasmid-based and DNA-free CRISPR (clustered regularly interspaced palindromic repeats) reagents for targeted chromosome editing in *P. tricornutum* and related diatoms using the Cas9 protein (CRISPR-associated protein 9) (6–11). *P. tricornutum* is diploid, meaning that Cas9-edited cells must be carefully screened to determine if knockouts are monoallelic or biallelic and exhibit loss of heterozygosity. These plasmid-based tools and

synthetic pathways are currently maintained by available antibiotic-based selections, including Zeocin™, phleomycin, nourseothricin, and blasticidin-S and their resistance genes, *Sh ble*, *nat*, and *bsr* (12–15). Antibiotic-based selections can be prohibitively expensive for maintaining large-scale cultures and are problematic for applications such as the biosynthesis of products intended for human consumption (16–18).

A viable alternative to antibiotics is the use of auxotrophic selective markers which require a strain engineered to have a loss of function mutation in a key enzyme of an essential biosynthetic pathway. Examples of commonly used auxotrophic strains in industrial and academic labs include uracil, histidine, and tryptophan auxotrophs (19–21). Two approaches have been taken to generate *P. tricornutum* auxotrophs. First, uracil-requiring mutants were generated by random mutagenesis that resulted in the identification of the bi-functional uridine monophosphate synthase (PtUMPS) gene predicted to catalyze the conversion of orotate into uridine monophosphate (UMP) (22). Biolistic transformation and chromosomal integration of the PtUMPS gene rescued the uracil-requiring phenotype. Second, Cas9 was used to knockout the PtUMPS gene to create uracil auxotrophs and the PtAPT gene encoding a predicted adenine phosphoribosyl transferase to create adenine auxotrophs (7). However, direct selection of these auxotrophs *via* transformation with the corresponding complementation marker has not been explored and the generation of additional auxotrophic strains would facilitate development of new plasmid-based complementation markers.

Here, we used a plasmid-based editing strategy to generate knockouts in the uracil, histidine, and tryptophan biosynthesis pathways of *P. tricornutum* and show for the first time that plasmid-based copies of the intact PtUMPS and PtPRA-PH/CH genes can complement the uracil- and histidine-requiring phenotypes, respectively. Individual auxotrophic strains are characterized by loss of heterozygosity at the edited alleles, and Nanopore sequencing of the edited population reveals large, heterogeneous deletions up to ∼ 2.7 kb. The uracil and histidine auxotrophs and their respective complementation markers are a potential alternative to antibiotic-based selection of plasmids in *P. tricornutum*. Our results also suggest a simple methodology to cure plasmids from uracil auxotrophs to enable strain and genome engineering.

## 3.2   Methods

### 3.2.1   Microbial strains and growth conditions

*S. cerevisiae* VL6-48 (ATCC MYA-3666: *MATα his3-δ200 trp1-δ1 ura3-52 lys2 ade2-1 met14 °*) was grown as described in section 1.2.1. *E. coli* (Epi300, Epicenter) was grown as described in section 1.2.1. *P. tricornutum* (Culture Collection of Algae and Protozoa CCAP 1055/1) was grown in L1 medium without silica, with or without uracil (50 mg l$^{-1}$) or histidine (200 mg l$^{-1}$) or 5-FOA (100 mg l$^{-1}$), supplemented with appropriate antibiotics Zeocin™ (50 mg l$^{-1}$) or nourseothricin (100 mg l$^{-1}$), at 18°C under cool white fluorescent lights (75 $\mu$E m$^{-2}$ s$^{-1}$) and a photoperiod of 16 h light:8 h dark. L1 media supplemented with nourseothricin contained half the normal amount of aquil salts. *P. tricornutum* auxotroph genotypes are as follows. Mutations in PtUMPS are described in reference to the chromosome 6 sequence (GenBank: CM000609.1), and mutations for PtPRA-PH/CH are in reference to the chromosome 3 sequence (GenBank: CP001142.1). Mutations described for each gene are listed for allele 1 followed by allele 2, and numbered beginning from the first nucleotide of the start codon for simplicity. Genotypes of auxotroph strains generated in this study are listed in Supplementary Table C.4.

### 3.2.2   Transfer of DNA to *P. tricornutum via* conjugation from *E. coli*

Conjugations were performed as described in section 1.2.2. Exconjugants were selected on $^1/_2$ × L1 1% agar plates supplemented with Zeocin™ 50 mg l$^{-1}$ or nourseothricin 100 mg l$^{-1}$. Colonies appeared after 7–14 days incubation at 18°C with light.

### 3.2.3   Plasmid design and construction

All plasmids (Supplementary Table C.5) were constructed using a modified yeast assembly protocol (23, 24). Plasmids pPtUMPSA1 and pPtUMPSA2 were made from pPtGE31 (6) by replacing the *URA3* element with a PCR fragment consisting of PtUMPS allele 1 or 2 with ~1 kb up- and downstream of the PtUMPS ORF, amplified from *P. tricornutum* genomic DNA (oligonucleotides are listed in Supplementary Table C.6). Plasmids pPtUMPScA1 and pPtUMPScA2 were made from pPtUMPSA1 and pPtUMPSA2 by replacing the PtUMPS ORF with a PCR fragment con-

sisting of PtUMPS allele 1 or 2 amplified from *P. tricornutum* cDNA. Plasmid pPtUMPS40S was made from pPtGE31 by replacing the *URA3* element with a cassette consisting of PCR fragments of the *40SRPS8* promoter and terminator (6) flanking a PCR fragment of the PtUMPS allele 1 ORF amplified from *P. tricornutum* genomic DNA. Plasmid pPtPRAPHCH was made from pPtGE31 (6) by replacing the *URA3* element with a PCR fragment consisting of PtPRA-PH/CH with ~1 kb up- and downstream of the PtPRA-PH/CH ORF, amplified from *P. tricornutum* genomic DNA. Using Golden Gate assembly (25), sgRNAs targeting different regions of the PtUMPS and PtPRA-PH/CH genes were cloned into the BsaI sites positioned between the *P. tricornutum U6* promoter and ter- minator in pPtGE34 and pPtGE35. Plasmid constructs were confirmed by Sanger sequencing at the London Regional Genomics Facility.

### 3.2.4   Generation of PtUMPS and PtPRA-PH/CH knockouts using Cas9 and TevCas9

Plasmids pPtGE34 or pPtGE35, containing no guide RNA or sgRNA.UMPS.1944, sgRNA.UMPS.1646, sgRNA.UMPS.157, sgRNA.UMPS.311 for the PtUMPS gene, or sgRNA.PRAPHCH.929 or sgRNA.PRAPHCH.120 for the PtPRA-PH/CH gene, were conjugated from *E. coli* to *P. tricornutum* and exconjugants were selected on Zeocin™-containing media, sup- plemented with uracil or histidine as appropriate (26). Ten colonies from each conjugation were resuspended in TE buffer and flash frozen at -80°C followed by heating at 95°C to lyse cells and extract genomic DNA. The genomic target site of each sgRNA in *P. tricornutum* was amplified by PCR and the products were analyzed by T7EI assay as follows; PCR products were denatured at 95°C for 5 mins, slowly cooled to 50°C, and flash frozen at -20°C for 2 mins. PCR products (250 ng) were incubated with 2U of T7EI (NEB) in 1× NEBuffer 2 for 15 mins at 37°C and an- alyzed by agarose gel electrophoresis. Colonies that showed editing by T7EI assay were grown in liquid culture supplemented with Zeocin™ and uracil or histidine as appropriate for 2 weeks and serial dilutions were plated on selective media with uracil or histidine to isolate sub-clones. Sub-clones were then screened for homozygous PtUMPS or PtPRA-PH/CH knockout phenotypes by replica streaking on minimal L1 media and L1 media supplemented with uracil or histidine as appropriate. Streaks were grown for 5 days before visual identification of phenotypes. Sub-clones

that were identified as phenotypic knockouts were resuspended in TE buffer and flash frozen at -80°C followed by heating at 95°C to lyse cells and extract genomic DNA, then sgRNA target sites were PCR amplified. Sanger sequencing of PCR products was performed at the London Regional Genomics Facility to identify the type and length of indels generated. Stable bi-allelic PtUMPS or PtPRA-PH/CH knockout mutant lines were then grown in nonselective L1 media supplemented with uracil or histidine for 1 week to cure them of plasmids before plating to obtain single colonies. Resulting colonies were replica streaked onto nonselective and Zeocin™-containing media supplemented with uracil or histidine to identify colonies which had successfully been cured of the plasmid.

### 3.2.5   Spot plating *P. tricornutum*

Cultures of *P. tricornutum* were adjusted to $1 \times 10^6$ cells ml$^{-1}$ and serially diluted 2× three times. For uracil auxotrophs, 10 $\mu$L of each adjusted culture and dilutions were spot plated onto minimal L1 media and L1 media supplemented with uracil (50 mg l$^{-1}$), 5-FOA (100 mg l$^{-1}$), or both. For histidine auxotrophs, 10 $\mu$L of each adjusted culture and dilutions were spot plated onto minimal L1 media and L1 media supplemented with histidine (200 mg l$^{-1}$). Plates were incubated at 18°C under cool white fluorescent lights (75 $\mu$E m$^{-2}$ s$^{-1}$) and a photoperiod of 16 h light:8 h dark for 7–10 days.

### 3.2.6   Measuring *P. tricornutum* growth rates

Growth was measured in a Multiskan Go microplate spectrophotometer. Cultures of each strain (WT, ΔUMPS1, ΔUMPS1 + pPtUMPS40S, ΔUMPS1 + pPtUMPSA1, ΔUMPS1 + pPtUMPSA2, ΔUMPS1 + pPtUMPScA1, ΔUMPS1 + pPtUMPScA2, ΔUMPS2 + pPtUMPSA1, ΔUMPS2 + pPtUMPSA2, ΔUMPS2 + pPtUMPScA1, ΔUMPS2 + pPtUMPScA2) were adjusted to $5 \times 10^5$ cells ml$^{-1}$ in L1 media with and without supplemented uracil (50 mg l$^{-1}$), 5-FOA (100 mg l$^{-1}$), or both. Two hundred microliters of each adjusted culture was added to three wells (technical replicates) of a 96-well microplate. The 96-well microplates were incubated at 18°C under cool white fluorescent lights (75 $\mu$E m$^{-2}$ s$^{-1}$) and a photoperiod of 16 h light:8 h dark for 10 days, and absorbance at 670 nm (A$_{670}$) was measured every 24 hrs. The 96-well microplates were shaken

briefly to resuspend any settled cells prior to absorbance measurements. Note that the $A_{670}$ was not adjusted for path length and light scattering from the microplate lid and is therefore not directly comparable to optical density readings measured in a standard cuvette.

### 3.2.7 *P. tricornutum* metabolite extraction

Cultures of *P. tricornutum* (wild-type, ΔUMPS1, and ΔUMPS2) were grown with and without uracil supplementation and harvested during exponential phase as follows (Note: The ΔUMPS1 and ΔUMPS2 cultures were first grown with uracil supplementation, then switched to minimal L1 media for 1 week prior to harvesting). Cultures ($\sim 1 \times 10^9$ cells) were pelleted by centrifugation at $4000 \times g$ for 10 mins and washed by resuspending in fresh L1 media. Cells were pelleted again, resuspended in a small volume ($\sim$5 ml) of L1 media, and transferred to a clean 10 ml syringe (without needle) with the exit plugged by parafilm. The syringe was placed, tip-down, into a clean 50 ml falcon tube and the cells were pelleted as above. The supernatant was removed and the pellet was slowly ejected from the syringe into a pre-chilled mortar containing liquid nitrogen. The frozen cells were ground to a fine powder and then transferred to a clean pre-weighed 1.5 ml Eppendorf tube, suspended half way in liquid nitrogen. Being careful to keep samples frozen, 50 mg of frozen ground powder was weighed out into a new clean 1.5 ml Eppendorf tube, pre-cooled in liquid nitrogen, and 250 $\mu$L of cold extraction buffer with internal standard (IS) (80% methanol in MilliQ water, 125 $\mu$M $^{15}N_2$-uracil) was added. The IS was added to the samples to compensate for losses that might occur during preparation of the samples and loss of sensitivity attributable to quenching of the signal by coeluting compounds. Samples were then homogenized by vigorous vortexing for 30 s in 10 s intervals, between which samples are kept on ice for $\sim$30 s. Homogenized samples were then spun down at 4°C for 10 mins at 20,000 $\times$ g. The supernatant was transferred to a new clean 1.5 ml Eppendorf tube and spun down at 20,000 $\times$ g for 5 mins at 4°C. The supernatant was again transferred to a new clean 1.5 ml Eppendorf tube and kept at 4°C overnight prior to LC-MS analysis.

### 3.2.8 Chromatographic separation and mass spectrometry

Metabolites were separated at 45°C on a Waters Acquity HSS T3 column [2.1 × 100 mm, 1.8 $\mu$m] in a Waters ACQUITY UPLC I-Class system (Waters, Milford, MA). Solvent A consisted of water and solvent B consisted of methanol, both containing 0.1% formic acid. Elution was performed by use of a linear gradient, at a flow rate of 0.3 ml min$^{-1}$, as follows: 0–2 mins, 100% solvent A to 90% solvent B; 2.01 mins, 100% solvent A to recondition the column. A Waters Xevo G2-S quadrupole time of flight mass spectrometer was operated in negative electrospray ionization (ESI) in resolution mode. The capillary voltage was set to 1.0 kV, the source temperature was 150°C, desolvation temperature was 600°C, the cone gas was 50 l hr$^{-1}$ and the desolvation gas was 1000 l hr$^{-1}$. Leucine enkephalin was infused as the lock mass with a scan time of 0.3 s every 10 s, and three scans were averaged. Linearity and detection limits for each compound were established by injection of calibration mixtures with different concentrations (0, 1, 2, 4, 8, 16, 31.25, 62.5, 125, 250, and 500 $\mu$mol l$^{-1}$). Stable-isotope-labeled uracil ($^{15}$N$_2$-uracil) was used as the IS. The concentration of each analyte was determined by use of the slope and intercept of the calibration curve that was obtained from a least-squares regression for the analyte/IS peak-area ratio vs the concentration of the analyte in the calibration mixture.

### 3.2.9 *P. tricornutum* DNA extraction and targeted long-read sequencing

Plasmids pPtGE34 or pPtGE35, containing sgRNAs targeting the PtUMPS, PtUrease, or PtI3GPS-PRAI gene were conjugated from *E. coli* to *P. tricornutum* and exconjugants were selected on Zeocin™-containing media, supplemented with uracil or tryptophan (100 mg l$^{-1}$) as appropriate. For each transformation, colonies (~1000) were scraped and pooled in liquid L1 media and genomic DNA was extracted using a modified akaline lysis protocol as follows: Cells were pelleted at 4000 × g for 5 mins, and resuspended in 250 $\mu$L resuspension buffer consisting of 235 $\mu$L P1 (Qiagen), 5 $\mu$L hemicellulose 100 mg ml$^{-1}$, 5 $\mu$L of lysozyme 25 mg ml$^{-1}$, and 5 $\mu$L zymolyase solution (200 mg zymolyase 20 T (USB), 9 ml H$_2$O, 1 ml 1 M Tris pH 7.5, 10 ml 50% glycerol) and incubated at 37 °C for 30 mins. Next, 250 $\mu$L of lysis buffer P2 (Qiagen) was added, followed by 250 $\mu$L of neutralization buffer P3 (Qiagen) and centrifugation at 16000 × g for 10 mins. The supernatant was transferred to a clean tube, 750 $\mu$L isopropanol was added, and the samples cen-

trifuged at $16000 \times g$ for 10 mins. A 70% EtOH wash was performed, centrifuged at $16000 \times g$ for 5 mins, and pellets briefly dried, resuspended in 50–100 $\mu$L of TE buffer, and incubated at 37°C for 30–60 mins.

The sgRNA target site regions were PCR amplified from sgRNA transformant genomic DNA samples, as well as a wild-type sample, with PrimeStar GXL polymerase (Takara) using primers positioned ~3 kb up- and downstream of the target site (Supplementary Table C.6). PCR products were purified and DNA libraries were prepared, barcoded, and pooled using an Oxford Nanopore Ligation Sequencing Kit (SQK LSK109) and Native Barcoding Expansion 1–12 (EXP-NBD104) kit according to manufacturers protocol with the following modification: all reactions were scaled down to half the recommended volume and the end prep incubation times were extended to 15 mins at 20°C and 15 mins at 65°C. The pooled library was then loaded on to a MinION R9.4.1 flowcell and sequenced.

### 3.2.10   Targeted long-read sequencing analysis

After sequencing on an R9.4.1 flowcell, base calling was performed using GPU Guppy with the high accuracy configuration file version 3.4.4 (https://community.nanoporetech.com). Reads in each barcode were filtered using NanoFilt (27) for a minimum average read quality score of 10 and a minimum read length of 2000, mapped using minimap2 (28) and filtered for reads that map to within 100 bases of each end of the reference sequence (the unedited 6 kb PCR product sequence) to remove short fragments. The filtered reads were mapped using minimap2 (parameters: -ax map-ont) and outputted in sam format, then converted to bam, sorted, and indexed using samtools (29). The per-base coverage depth for each barcode was calculated using Mosdepth (30). All plots were created in R using the ggplot2 package (31).

## 3.3   Results

### 3.3.1   Identification of Cas9 targets in biosynthetic pathway genes

We examined the KEGG predictions (32, 33) based on the genome sequence of *P. tricornutum* to identify genes in the uracil and histidine biosynthetic pathways for Cas9 editing. We focused

on these two pathways as uracil and histidine auxotrophy, and counter-selection strategies are commonly used in other model organisms. This approach identified the previously described bi-functional PtUMPS gene that is predicted to catalyze two steps in the uracil pathway — conversion of orotate to orotidine monophosphate (OMP), and conversion of OMP to UMP (Fig. 3.1 and Fig. 3.2A) (22). Proteins that are orthologs of characterized enzymes involved in histidine biosynthesis were also identified (Fig. 3.3 and Fig. 3.4A ). The PHATR_3140 gene, hereafter called PtPRA-PH/CH, encodes a predicted bifunctional protein that shares sequence similarity with the bacterial protein HisIE, and its *Arabidopsis* counterpart HISN2 (34, 35). These proteins possess two functional domains that are homologous to the phosphoribosyl-ATP pyrophosphohydrolase (PRA-PH) and phosphoribosyl-AMP cyclohydrolase (PRA-CH) enzymes, respectively. PRA-PH and PRA-CH, alone or as a bifunctional protein, are predicted to catalyze two successive steps that occur early in the histidine biosynthesis pathway (Fig. 3.3 and Fig. 3.4A). The PtIGPS gene encoding imidazole glycerol phosphate synthase (a HIS3 homolog) was found to be a duplicated gene in the *P. tricornutum* genome assembly and thus not prioritized as a Cas9 target.

We also identified the PtI3GPS-PRAI gene as a potential target as it encodes a predicted bi-functional enzyme that is a fusion of indole-3-glycerol-phosphate synthase (I3GPS) and phospho-ribosylanthranilate isomerase (PRAI), and would catalyze two successive steps in the tryptophan biosynthesis pathway (Fig. 3.5).

To confirm the genomic target sites, we PCR-amplified and sequenced the PtUMPS and PtPRA-PH/CH genes of the *P. tricornutum* CCAP 1055/1 strain used in our laboratory. Two distinct alleles for both the PtUMPS and PtPRA-PH/CH genes were identified. Seven single-nucleotide polymorphisms (SNPs) in the PtUMPS alleles result in amino acid substitutions that differentiate the two alleles from each other and from the published *P. tricornutum* genome (Supplementary Table C.1). All substitutions are located in non-conserved regions of the PtUMPS protein (Fig. 3.6). Similarly, an A to G mutation at base position 1205 in allele 2 of the PtPRA-PH/CH gene was identified (Supplementary Table C.2). This transversion converts a highly conserved glutamate to a glycine in the catalytic site of the PRA-PH domain. The impact of these substitutions on PtUMPS and PtPRA-PH/CH function is unknown.
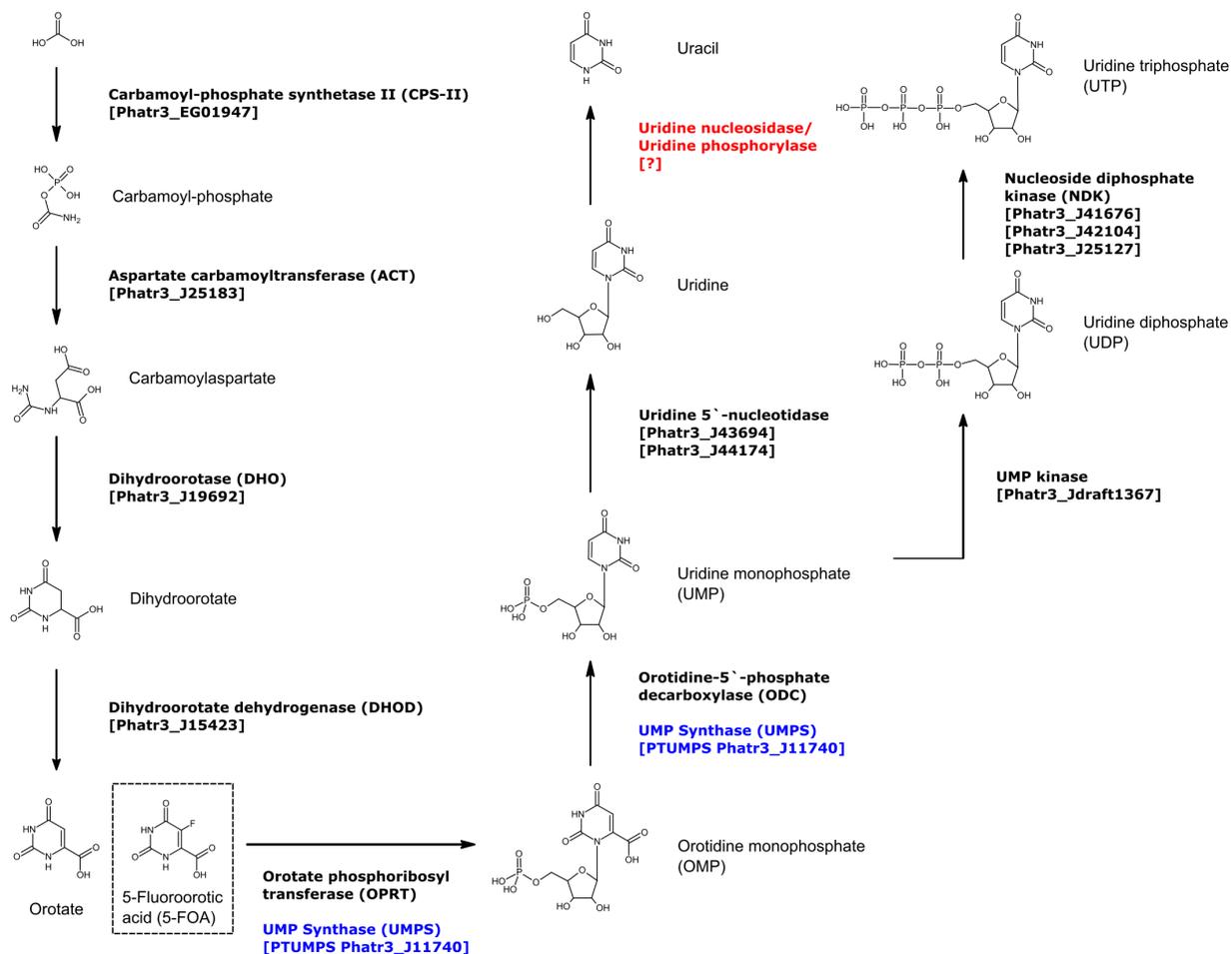
**Figure 3.1: The predicted *P. tricornutum* uracil biosynthesis pathway.** The diagram depicts the biosynthesis pathway for conversion of carbonic acid to uracil and uridine triphosphate, with the PtUMPS enzyme highlighted in blue. The competitive inhibitor, 5-Fluoroorotic acid (5-FOA), is also shown in a hashed box at the position where it enters the pathway. No enzymes sharing significant sequence similarity with uridine nucleosidase or uridine phosphorylase (highlighted in red) were identified in *P. tricornutum* by NCBI and EnsemblProtists BLAST queries. Abbreviated names for molecules and enzymes are indicated in parentheses, and the predicted corresponding *P. tricornutum* gene names are indicated in square brackets.
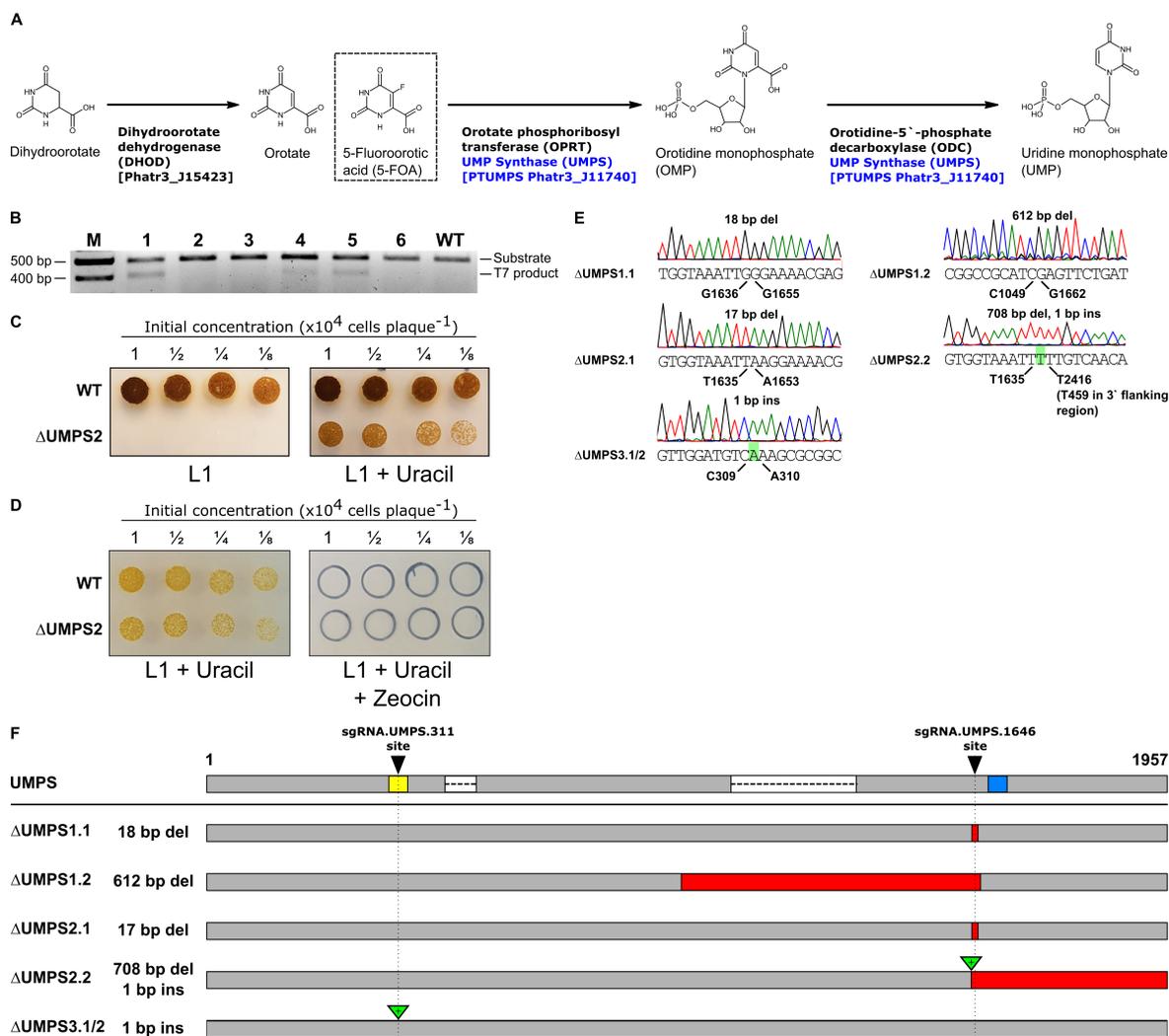
**Figure 3.2: CRISPR-generated knockouts in the predicted *P. tricornutum* uracil biosynthesis pathway.** (A) A portion of the predicted *P. tricornutum* biosynthesis pathway for conversion of carbonic acid to uracil and uridine triphosphate, with the PtUMPS enzyme highlighted in blue. The competitive inhibitor, 5-fluoroorotic acid (5-FOA), is shown in a dashed box at the position where it enters the pathway. Abbreviated names for molecules and enzymes are indicated in parentheses, and the predicted corresponding *P. tricornutum* gene names are indicated in square brackets. (B) Example image of T7EI editing assay to screen exconjugants for potential editing events in the PtUMPS gene. Substrate indicates PtUMPS gene fragments amplified by the PCR, while T7 product indicates exconjugants with evidence of Cas9 editing. WT, wild-type *P. tricornutum* genomic DNA used in the T7EI editing assay. M, 100 bp ladder with sizes indicated in basepairs (bp). (C) Example of phenotypic screening of one PtUMPS knockout strain (ΔUMPS2) plated on L1 alone or L1 supplemented with uracil at the indicated dilution of initial concentration. (D) Example of screening for loss of the zeocin-resistant Cas9 editing plasmid in a ΔUMPS2 knockout strain by plating on L1 supplemented with uracil or L1 supplemented with uracil and Zeocin™. (E) Sanger sequencing traces of characterized PtUMPS knockouts with the position (below trace) and type of insertion or deletion (above trace) indicated for each allele of the three strains. (F) Graphical map of the position and extent of indels for each of the three PtUMPS knockouts relative to the wild-type UMPS gene (shown at top). Red rectangles indicate nucleotide deletions, green triangles indicate nucleotide insertions, the yellow and blue rectangles on the WT gene indicate the position of the PtUMPS active sites (orotate phosphoribosyl transferase and orotidine-5'-phosphate decarboxylase), and the white rectangles with dashed lines represent introns.
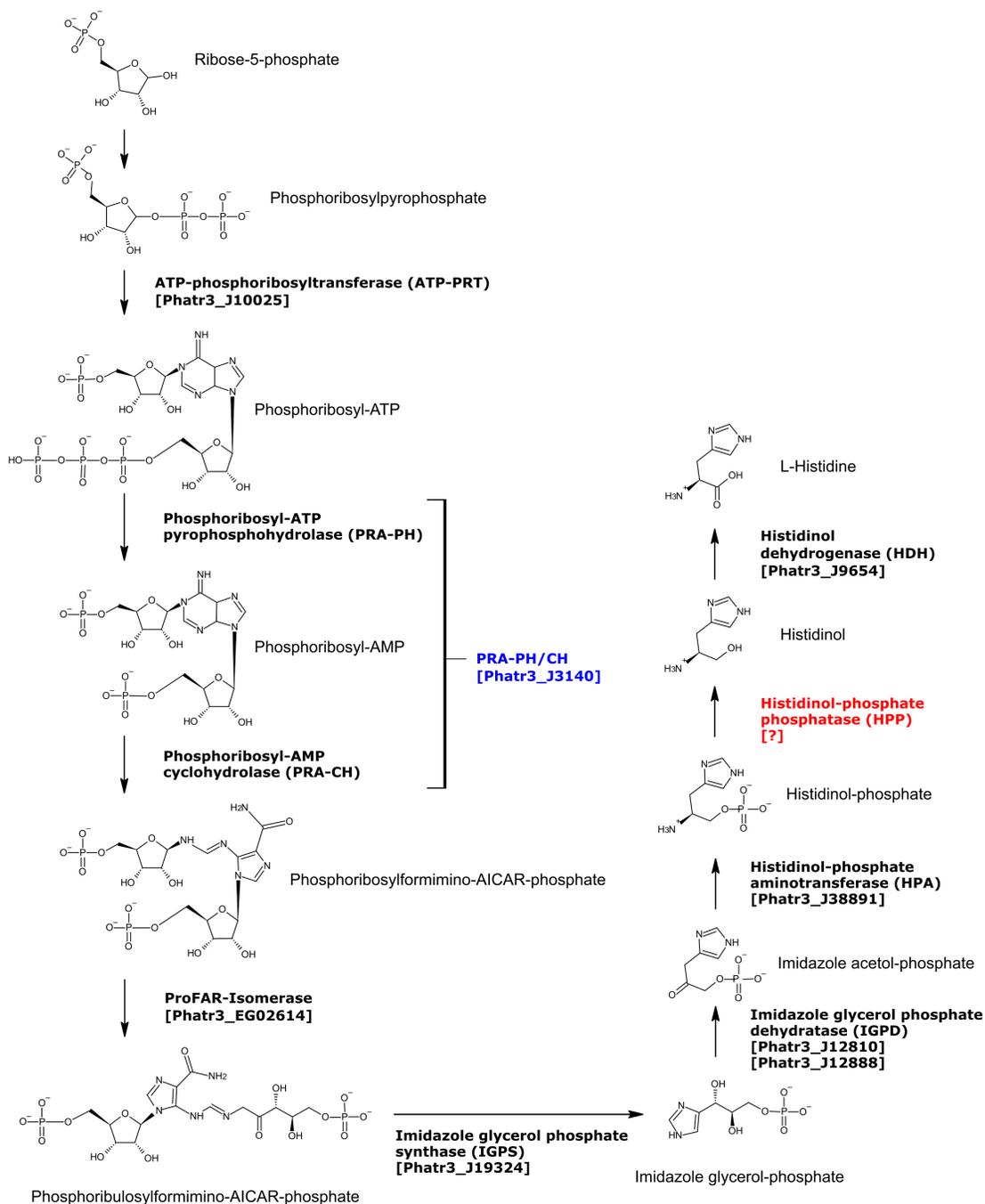
**Figure 3.3: The predicted *P. tricornutum* histidine biosynthesis pathway.** The diagram depicts the biosynthesis pathway for conversion of ribose-5-phosphate to L-histidine, with the PtPRA-PH/CH enzyme highlighted in blue. No enzymes sharing significant homology with HPP (highlighted in red) were identified in *P. tricornutum* by NCBI and EnsemblProtists BLAST queries. Abbreviated names for each enzyme are indicated in parentheses, and the predicted corresponding *P. tricornutum* gene names are indicated in square brackets.
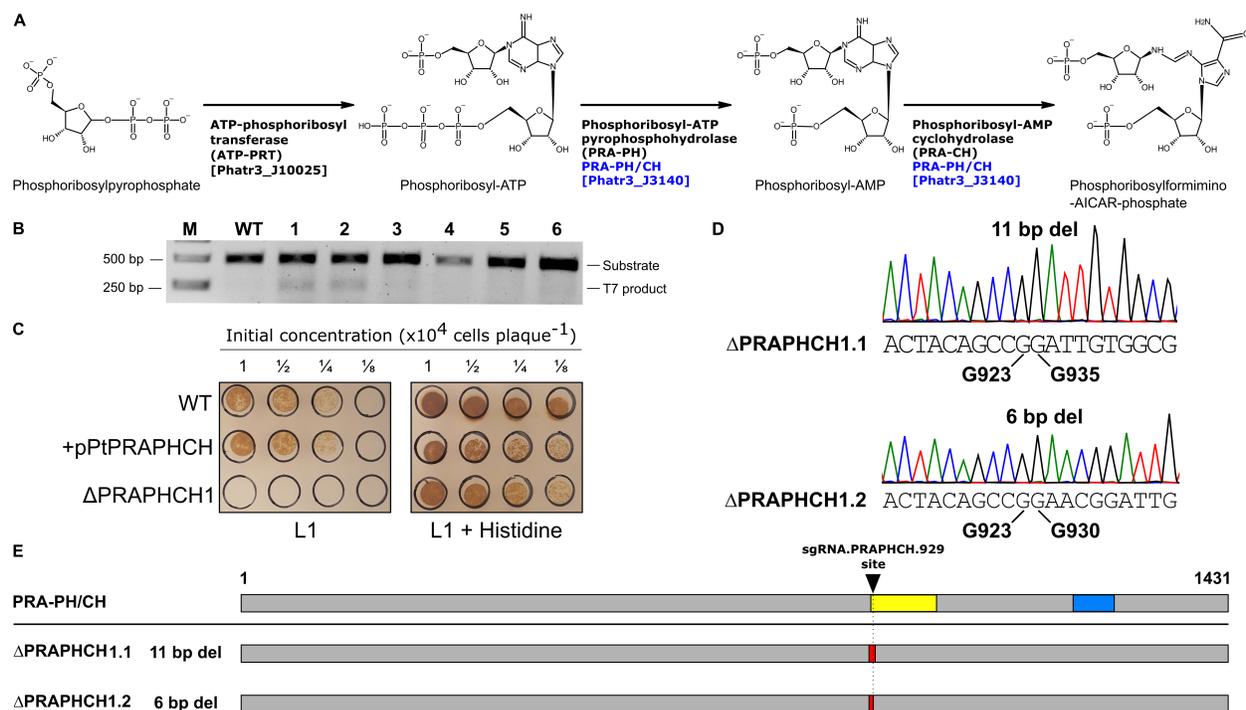
**Figure 3.4: CRISPR-generated knockouts in the predicted *P. tricornutum* histidine biosynthesis pathway.** (A) A portion of the predicted biosynthesis pathway for conversion of ribose-5-phosphate to L-histidine, with the bi-functional PtPRA-PH/CH enzyme highlighted in blue. Abbreviated names for each enzyme are indicated in parentheses, and the predicted corresponding *P. tricornutum* gene names are indicated in square brackets. (B) Example image of T7EI editing assay to screen exconjugants for potential editing events in the PtPRA-PH/CH gene. Substrate indicates PtPRA-PH/CH gene fragments amplified by the PCR, while T7 product indicates exconjugants with evidence of Cas9 editing. WT, wild-type *P. tricornutum* genomic DNA used in the T7EI editing assay. M, 1 kb ladder with sizes indicated in basepairs (bp). (C) Example of phenotypic screening of one PtPRA-PH/CH knockout strain (ΔPtPRAPHCH1) transformed with or without the complementing PRA-PH/CH plasmid (pPtPRAPHCH) on L1 solid media alone or L1 supplemented with histidine at the indicated dilution of initial concentration. WT, wild-type *P. tricornutum* strain. (D) Sanger sequencing traces of characterized PtPRA-PH/CH knockouts with the position (below trace) and type of insertion or deletion (above trace) indicated for each allele. (E) Graphical map of the position and extent of indels for PtPRA-PH/CH knockout relative to the wild-type PtPRA-PH/CH gene (shown at top). Red rectangles indicate nucleotide deletions, while the yellow and blue rectangles on the WT gene indicated the position of the PRA-PH and PRA-CH active sites.
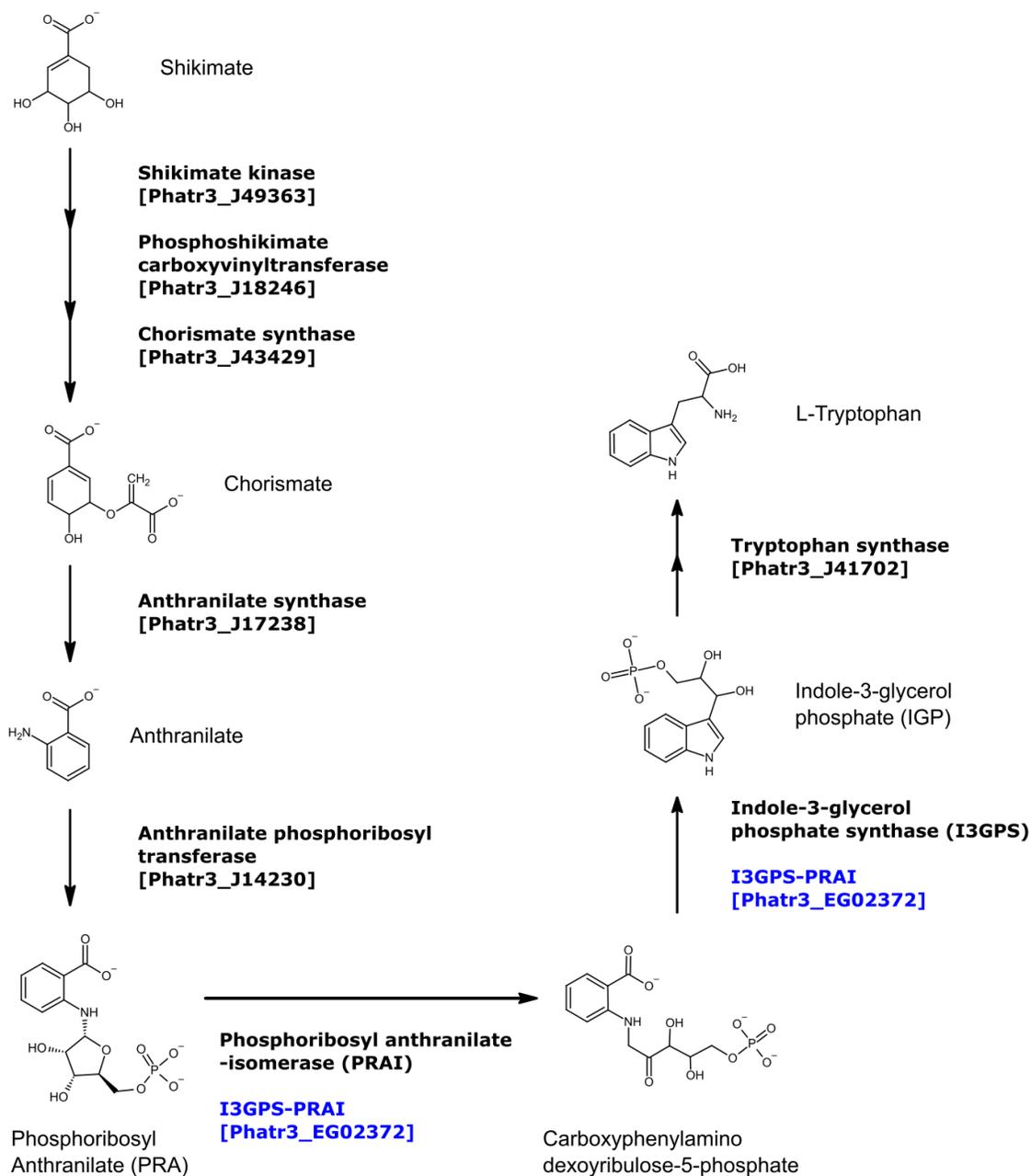
**Figure 3.5: The predicted *P. tricornutum* tryptophan biosynthesis pathway.** The diagram depicts the biosynthesis pathway for conversion of shikimate to L-tryptophan, with the PtI3GPS-PRAI enzyme highlighted in blue. Abbreviated names for molecules and enzymes are indicated in parentheses, and the predicted corresponding *P. tricornutum* gene names are indicated in square brackets.
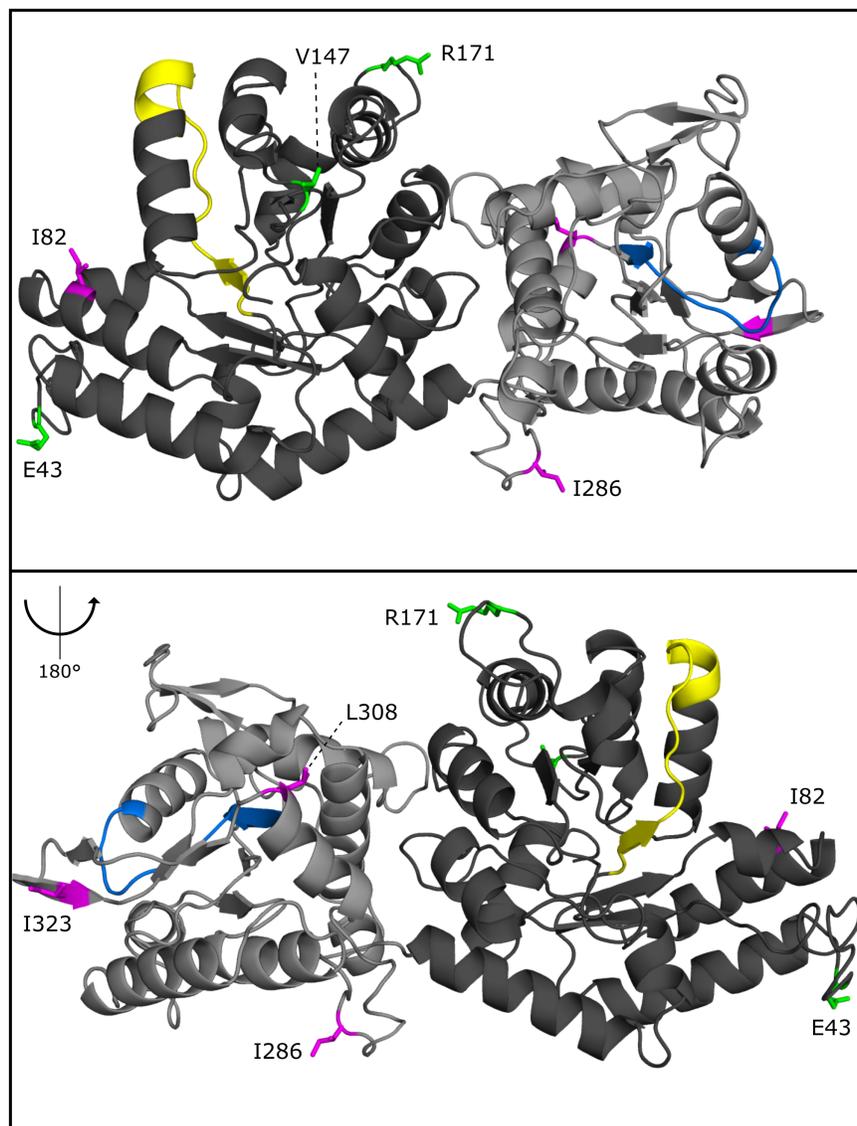
**Figure 3.6: Predicted PtUMPS structure indicating the positions of the ODC (dark grey) and OPRT (light grey) domains.** Regions containing conserved active site residues for the ODC and OPRT domains are indicated in yellow and blue, respectively. Residue substitutions that differentiate the two alleles are labeled and highlighted in green (allele 1) and magenta (allele 2). Folding prediction was modeled using the PHYRE2 Protein Fold Recognition Server.

## 3.3.2 Cas9 and TevCas9 editing of auxotrophic genes is characterized by loss of heterozygosity

To generate knockouts in uracil and histidine biosynthetic genes, we designed and individually cloned Cas9 and TevCas9 single guide RNAs (sgRNAs) against different sites in the PtUMPS, PtPRA-PH/CH, and PtI3GPS-PRAI genes (Table 3.1 and Fig. 3.7, Fig. 3.8, Fig. 3.9). The TevCas9 nuclease is a dual nuclease that generates a 33–38 base pair deletion between the I-TevI (Tev) and Cas9 cut sites (36). The targeting requirements for a TevCas9 nuclease are an I-TevI 5′-CNNNG-3′ cleavage motif positioned ~15–18 base pairs upstream of the 5′ end of the sgRNA binding site. The Cas9 or TevCas9 editing plasmids were moved into *P. tricornutum* by bacterial conjugation and exconjugants selected on Zeocin™-containing media.

**Table 3.1: Summary of sgRNAs used for Cas9 and TevCas9 editing.**

| Target | Platform | Guide RNA | Exconjugants edited / total screened (T7E1) | Number of subclones with auxotroph phenotype / screened |
|---|---|---|---|---|
| UMPS | Cas9 | sgRNA.UMPS.1944 | 0/10 | N/A |
|  |  | sgRNA.UMPS.1646 | 0/10 | N/A |
|  |  | sgRNA.UMPS.157 | 0/10 | N/A |
|  |  | sgRNA.UMPS.311 | 4/10 | 1/35 |
|  | TevCas9 | sgRNA.UMPS.1944 | 0/10 | N/A |
|  |  | sgRNA.UMPS.1646 | 4/10 | 2/35 |
|  |  | sgRNA.UMPS.157 | 0/10 | N/A |
| PRA-PH/CH | Cas9 | sgRNA.PRAPHCH.929 | 2/6 | 1/28 |
|  |  | sgRNA.PRAPHCH.120 | 0/6 | N/A |
|  |  | sgRNA.PRAPHCH.1000 | 1/6 | 0/28 |
|  | TevCas9 | sgRNA.PRAPHCH.929 | 3/6 | 0/28 |
|  |  | sgRNA.PRAPHCH.120 | 0/6 | N/A |
| I3GPS-PRAI | Cas9 | sgRNA.IGPSPRAI.244 | 0/10 | N/A |
|  | TevCas9 | sgRNA.IGPSPRAI.244 | 3/10 | 1/35 |

We first assessed editing by screening *P. tricornutum* exconjugants by T7 endonuclease I (T7EI) mismatch cleavage assays on PCR products amplified from each target gene (Fig. 3.2B and 3.4B, and Table 3.1). This assay identified 6 sgRNAs with detectable editing rates based on screening of exconjugants. Colonies that showed editing were diluted, plated to obtain subclones, and subsequently screened for the corresponding auxotrophic phenotype on solid media with and without auxotrophic supplement (uracil or histidine) (Fig. 3.2C and 3.4C). To cure the Cas9-editing plasmids, knockout strains were grown without Zeocin™ selection for 1 week, and dilutions were plated to obtain single colonies. Colonies were streaked onto L1 plates with and without Zeocin™ to screen for plasmid loss. A representative image demonstrating Zeocin™ sensitivity due to Cas9-editing plasmid loss is shown in Fig. 3.2D. For knockout of the PtUMPS gene, we further charac-

**Figure 3.7: The PtUMPS coding sequence and gRNA target sites.** The coding regions for the ODC (yellow) and OPRT (blue) catalytic residues, and TevCas9 target sites (light blue) are highlighted. Residues that differ between the two alleles are indicated by red boxes.

ATGTCGAGTTTTCCGTATCCCGCCTTTCTCCACCACCGTCCGGTGGACGCCGCGGCGCAGCGTGCCGCACTCGCGGCCGTTGGTCGTTGCCGCATCGTCCCTACATCGTCCTTTCCT**GGG**     120

M  S  S  F  P  Y  P  A  F  L  H  H  R  P  V  D  A  A  A  Q  R  A  A  L  A  A  V  G  R  C  R  I  V  P  T  S  S  [F]  P  G
1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

**GCCGCTGACTTTCGCGAGATTGTCCAAATGAACGTGG**CACAAGCACACGTCGGACTCACGGACGAGACCGACGGTGCGGATCAGGAAACGCTCCAGACCACACGCAAGGCCCGGATGAAA     240

-sgRNA.PRAPHCH.120—■
A  A  D  F  R  E  I  V  Q  M  N  V  A  Q  A  H  V  G  L  T  D  E  T  D  G  A  D  [Q]  E  T  L  Q  T  T  R  K  A  R  M  K
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80

CAGGGCATCGCCATGGCCAAGACGTATTTCGTCGAAACCTTACAGTTGGATGTCACGGTCGTGGATGTGGAAGTCGTTTTGGATGCTCGGGATGTGATTCAATCGGACGGCTGTCTCGCG     360

Q  G  I  A  M  A  K  T  Y  [F]  V  E  T  L  Q  L  D  V  T  V  V  D  V  E  V  V  L  D  A  R  D  V  I  Q  S  D  G  C  L  A
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120

AGTGGACTCCTCGATGCGGGTTGTGACAAAATTGTCGTGGACGGCACCAATCGGGACGCTTTTGAGACTGCACAAATTCCGAAAGCGAGACTGATGGGCCACTTTGTGTATCCCCATACC     480

S  G  L  L  D  A  G  C  D  K  I  V  V  D  G  [T]  N  R  D  A  [F]  E  T  A  Q  I  P  K  A  R  L  M  G  H  F  V  Y  P  H  T
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160

GATTCACGAACCCGAAACGGGAGTCTCGCCGTGGCCTTTGCCGAAGTCCAACCCTTTTGCGATTCCGTCAGTGTCGAGCTATCGAATCCTGAAGCCGTCACGGAAAGTGTCATCCAGGAA     600

D  S  R  T  R  N  G  S  L  A  V  A  F  A  E  V  Q  P  F  C  D  S  V  S  V  E  L  S  N  P  E  A  V  T  E  S  V  I  Q  E
161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200

ATTGTGCAAGCGGCAAAAGCCCAGCAATTGCAAATTTCAATCGTCTACAATCCCAACGACTACGAGAATCCTGGTGATCACGCCGCCATGGTTGGTCGCGTATGCCAGAGCTTCCCAGAA     720

I  V  Q  A  A  K  A  [Q]  Q  L  Q  I  S  I  V  Y  N  P  N  D  Y  E  N  P  G  D  H  A  A  M  V  G  R  V  C  Q  S  F  P  E
201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240

GGGCAAGTCCGGTTGGCCTTGACCGATCCGACCGCCTCGCAACTCGGACACGCCTACGCGGCGTGTCTCAAAACCGACCGGGACGACGGTCTCTTTACCACCGTCGTCTGTACTCGCAAC     840

G  Q  V  R  L  A  L  T  D  P  T  A  S  Q  L  G  H  A  Y  A  A  C  L  K  T  D  R  D  D  G  L  F  [T]  T  V  V  C  T  R  N
241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280

**PRA-CH**

AATGAAGCACTGGGGCTTGTCTACTCCTCCACGTCGTCTATTGTCGCCG**CATTGGAATGTGGTCGGGGTGTCTAC**TACAGCCGTTCCCGGAACGGATTGTGGCGCAAGGGCGACACGTCC     960

sgRNA.PRAPHCH.929——■
N  E  A  L  G  L  V  Y  S  S  T  S  S  I  V  A  A  L  E  C  G  R  G  V  Y  Y  S  R  S  R  N  G  L  W  R  K  G  D  T  S
281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320

GGACACTTTCAGGTACTGCACCGGCTCGACAAAGACTGTGACGGGGAT GCTTTGCGGGTGACGGTCACTCAAAAAGGCACCCCTGCCGCATTTTGCCACTTGAATACTCTAACGTGCTGG     1080

■—sgRNA.PRAPHCH.1000——■
G  H  F  Q  V  L  H  R  L  D  K  D  C  D  G  D  A  L  R  V  T  V  T  Q  K  G  T  P  A  A  F  C  H  L  N  T  L  T  C  W
321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360

GGTACGTCGAGAGGCTTGCGACATTTGGAAGAAACGCTGCAAGACCGACTCGTGTCGGCTCCGGAAGGATCCTACACGAAACGTCTGTTTCAAGACGAGACACTGTTGCGGGACAAGCTG     1200

G  T  S  R  G  L  R  H  L  E  E  T  L  Q  D  R  L  V  S  A  P  E  G  S  Y  T  K  R  L  F  Q  D  E  T  L  L  R  D  K  L
361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400

**PRA-PH**

GTC**GAAGAAGCGCAGGAATTGTCGGAAGCCGATACGAAGGCACACGTGGCGGAAGAACTGGCGGAT**GTCCTCTACTTTGCCATGGTGAGGGCGGTCAAGGCTGGAGTCAGTTTGGATGAT     1320

V  [E]  E  A  Q  E  L  S  E  A  D  T  K  A  H  V  A  E  E  L  A  D  V  L  Y  F  A  M  V  R  A  V  K  A  G  [V]  S  L  D  D
401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440

GCCGTCGCAGAGCTAGATTCTAGGGCGCGAAAAGTGACGCGACGAAAAGGCGATAGTAAAGCCTTTCGAATTGCAGCGGGTGAAGCTATTTTGAAAAAAGCGCAGTCGTAG

A  V  A  E  L  D  S  R  A  R  K  V  T  R  R  K  G  D  S  K  A  F  R  I  A  A  G  E  A  I  L  K  K  A  Q  S
441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477

**Figure 3.8: The PtPRA-PH/CH coding sequence and gRNA target sites.** The coding regions for the PRA-PH (blue) and PRA-CH (yellow) catalytic residues, and TevCas9 target sites (light blue) are highlighted. Residues that differ between the two alleles are indicated by red boxes.

**Figure 3.9: The PtI3GPS-PRAI coding sequence and gRNA target sites.** The coding regions for the I3GPS (blue) and PRAI (yellow) catalytic residues are highlighted. Residues that differ between the two alleles are indicated by red boxes.

terized 3 subclones with a uracil-requiring phenotype to determine if the knockouts were monoallelic or biallelic. Because the two PtUMPS alleles of *P. tricornutum* possessed SNPs relative to each other, we were able to map allele-specific editing events (Fig. 3.2E and 3.2F). Two of the strains, ΔUMPS1 and ΔUMPS2, were biallelic and exhibited loss of heterozygosity with one allele possessing a small deletion (< 20 bps) and the other allele possessing a large deletion (> 610 bp). The third characterized subclone, ΔUMPS3, was monoallelic and possessed a homozygous 1-bp insertion. For the PtPRA-PH/CH knockouts that generated a histidine-requiring phenotype (Fig. 3.4B and 3.4C), targeted sequencing of one subclone revealed a biallelic genotype with an 11-bp deletion in one allele and a 6-bp deletion in the second allele (Fig. 3.4D and 3.4E).

The types of deletions observed in the uracil- and histidine-auxotrophs are consistent with heterogeneous editing events resulting in loss of heterozygosity (37–39). To extend these observations, we used Nanopore sequencing to better assess the spectrum of large deletions that are often overlooked in Cas9-editing studies. In addition to the two sgRNAs that showed robust editing on the PtUMPS gene, we examined deletion events in exconjugants with sgRNAs targeted to the PtUREASE gene (6) and the PtI3GPS-PRAI gene. For each experiment, ~1000 exconjugants were pooled and a ~6 kb PCR product generated for each of the target genes with the predicted Cas9 or TevCas9 target sites in the middle of the amplicon. We focused our attention on deletions > 50 bp as these deletions are typically under-reported in targeted amplicon sequencing. We noted a drop in Nanopore read coverage centered around the predicted sgRNA target sites for products amplified from Cas9 and TevCas9 editing experiments (black dots) as compared to read coverage for control experiments (orange dots), consistent with editing at those sites (Fig. 3.10, *left* panels). Mapping the deletion start and end points revealed that most deletions were centered on the Cas9 or TevCas9 target site (Fig. 3.10, *right* panels), with deletions extending up to 2700 bp (Fig. 3.10, *centre* panel). The mean deletion length for editing events examined by Nanopore sequencing and > 50 bp was 1735±719 bp for Cas9 and 2006±633 bp for TevCas9.

Collectively, this data shows that Cas9 or TevCas9 editing of biosynthetic genes can readily generate *P. tricornutum* auxotrophs that can be identified by phenotypic or genetic screens. Moreover, our data agree with a growing body of evidence revealing that Cas9 editing (and TevCas9 editing here) generates large deletions that would typically be missed unless screening strategies are explicitly designed to look for loss of heterozygosity.
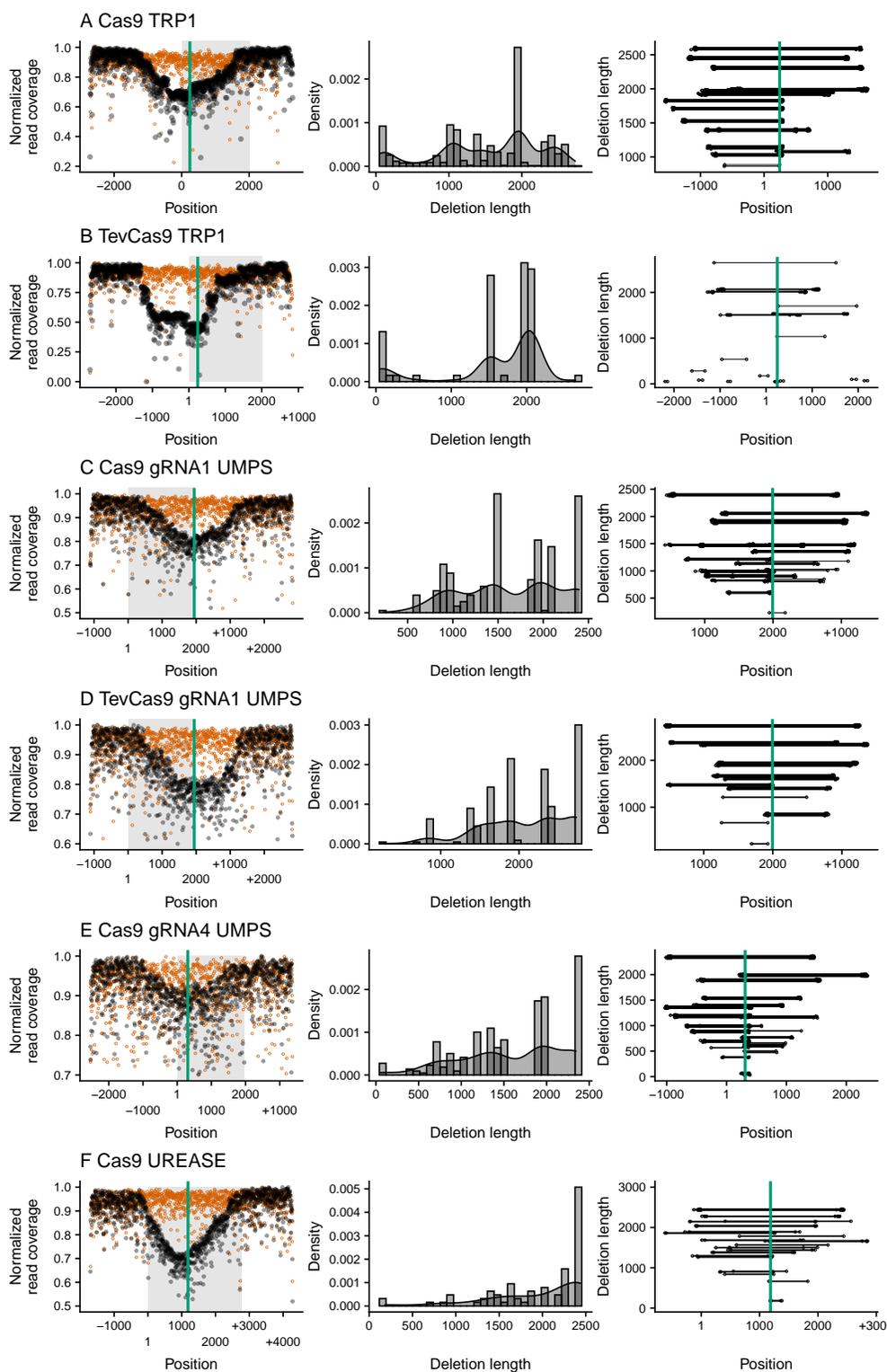
Figure 3.10: Caption continued on following page.

**Figure 3.10: Large deletions in edited *P. tricornutum* metabolic genes captured by Nanopore amplicon sequencing.** For each panel A–E, the name of the target gene as well as the editing enzyme are indicated. The *leftmost* plot shows normalized read coverage averaged over a 5-bp window for the edited sample (black dots) and the wild-type sample (orange dots) relative to the position in PCR amplicon. Numbering on the x-axis is relative to the ATG start codon for each gene, with sequence upstream indicated by a minus (-) symbol and sequence downstream indicated by a plus (+) symbol. The green vertical line indicates the Cas9 or TevCas9 cleavage site, while the shaded rectangle indicates the ORF. The *middle* plot is a density plot of deletions > 50 bp. The *rightmost* plot shows the length and position of deletions > 50 bp relative to their position in the PCR amplicon, with numbering of the x-axis as in the *leftmost* panel. Each horizontal line indicates a mapped deletion event. Deletions are ordered from longest to smallest. The green line indicates the Cas9 or TevCas9 cleavage site.

### 3.3.3 Phenotypic and metabolomic characterization of the PtUMPS knockouts

Two uracil-requiring auxotrophs (ΔUMPS1 and ΔUMPS2) were selected for further characterization by first spot plating onto L1 media with and without uracil and 5-FOA (Fig. 3.11A). The PtUMPS knockout strains were only able to survive in the presence of uracil supplementation. Additionally, the knockouts survived on 5-FOA concentrations that fully inhibited the growth of wild-type *P. tricornutum* (Fig. 3.11A). This is consistent with phenotypes previously observed for *P. tricornutum* UMPS knockouts (7, 22). There was a slight growth advantage of ΔUMPS1 over ΔUMPS2 on media supplemented with both 5-FOA and uracil, but not on media containing uracil alone. To compare if the observed phenotypes were consistent across solid and liquid media, we monitored the growth of these strains over 10 days in liquid media (Fig. 3.11B) and found that the growth rates were consistent with those observed on solid media, with one notable difference (Fig. 3.12, Fig. 3.13, and Supplementary Table C.3). The growth advantage of ΔUMPS1 over ΔUMPS2 observed on solid media supplemented with both 5-FOA and uracil was not replicated in liquid media as the generation times for ΔUMPS1 and ΔUMPS2 were very similar (~24 and ~22 hrs, respectively).

To investigate the impact of PtUMPS knockouts on uracil metabolism, we performed targeted metabolomics on the UMPS substrate orotate using LC-MS in wild-type and knockouts strains (Fig. 3.11C). We focused on characterizing the orotate intermediate in the uracil pathway (Fig. 3.4A and 3.11C) predicting that there should be an increase of orotate in knockout strains relative to wild type. We were unable to detect orotate in the ΔUMPS1 strain in the absence of uracil supplementation (-uracil), or in the ΔUMPS2 strain in either the -uracil or +uracil condition. A
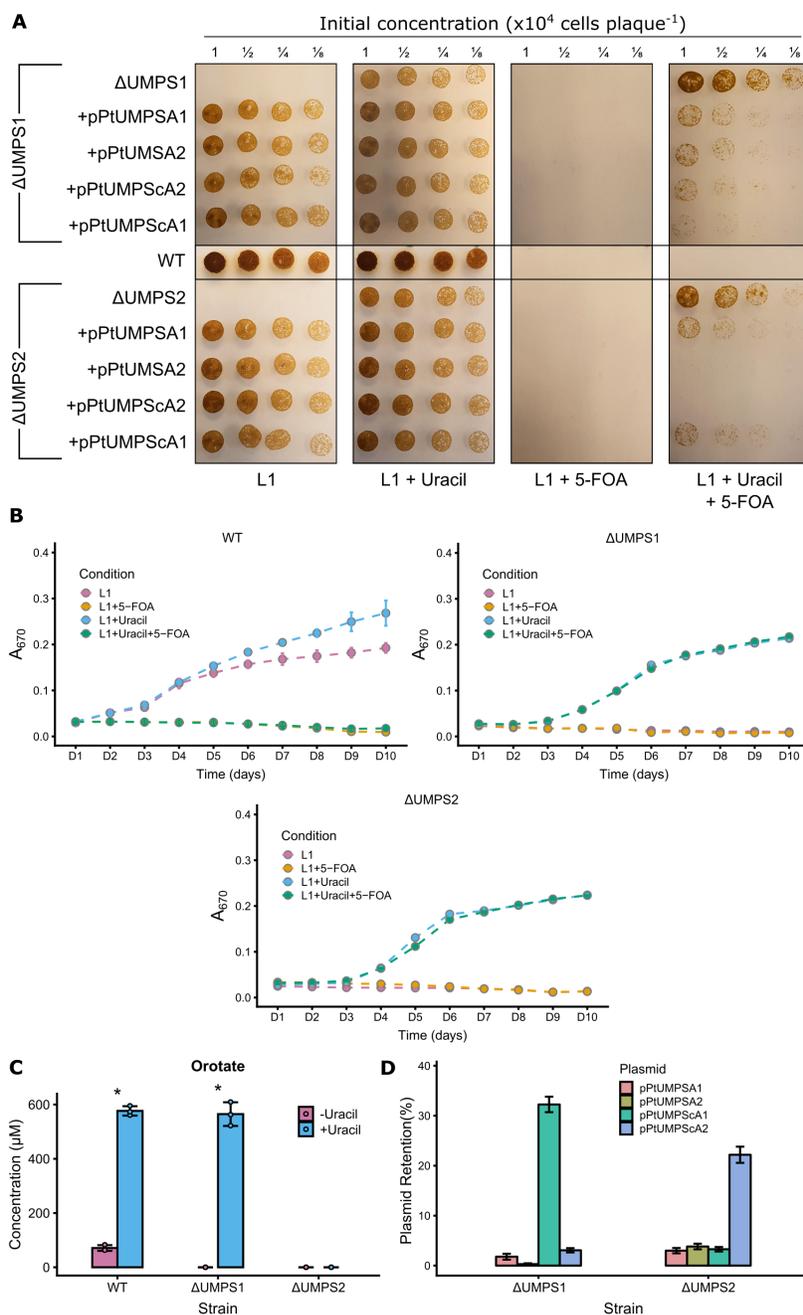
Figure 3.11: **Phenotypic and metabolomic characterization of PtUMPS knockouts.** (A) Spot plating assays of wild type (WT), ΔUMPS1 and ΔUMPS2 strains on L1 solid media alone, L1 supplemented with uracil, L1 supplemented with 5-FOA, or L1 supplemented with both uracil and 5-FOA. Indicated dilutions are relative to the initial concentration. (B) Liquid growth curves of wild type (WT), ΔUMPS1 and ΔUMPS2 strains in L1 liquid media alone, or supplemented with uracil or 5-FOA or both. Data points are the mean of three independent replicates, with error bars representing the standard error of the mean. (C) Orotate concentrations were measured by LC-MS from cultures grown with and without uracil supplementation. Bars represent mean values and error bars represent standard deviation for three biological replicates. Individual data points are represented as colored dots. Statistical confidence level was calculated by one-sided t test. $p < 0.001$ is indicated by an asterisk. (D) Bar graph showing percent plasmid retention in the ΔUMPS1 and ΔUMPS2 strains harbouring various PtUMPS constructs after 14 days of outgrowth. Bars represent the mean ratio of colonies on selective L1 + nourseothricin versus non-selective L1 plates from three independent replicates, with error bars representing the standard error of the mean.
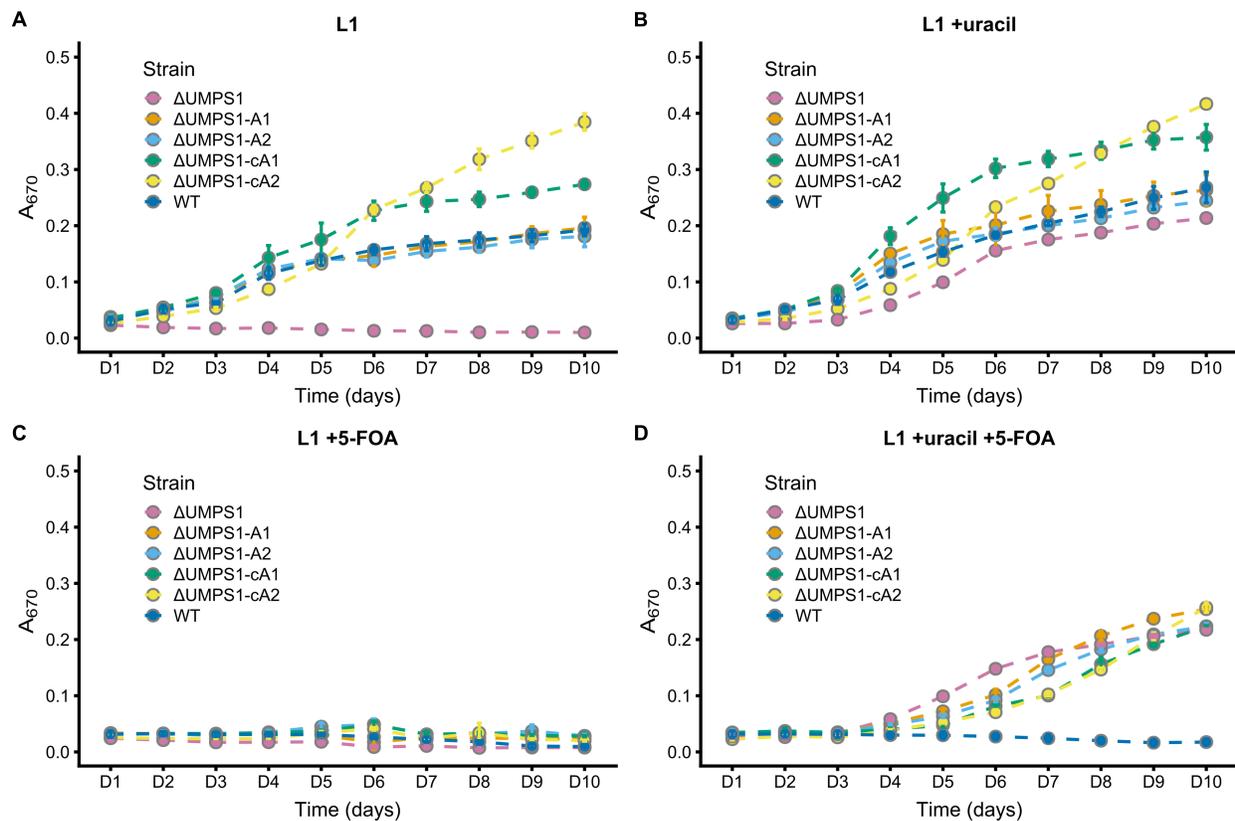
**Figure 3.12: Assaying ΔUMPS1 knockout and complemented growth rates in L1 media supplemented with uracil, 5-FOA, or both.** A biallelic PtUMPS mutant (ΔUMPS1) is unable to grow in L1 media without uracil supplementation. Reintroducing the gene on a stably replicating plasmid restores the WT phenotype in L1 media. (A) Growth rates in L1 media. (B) Growth rates in L1 media supplemented with uracil. (C) Growth rates in L1 media supplemented with 5-FOA. (D) Growth rates in L1 media supplemented with uracil and 5-FOA. WT, Wild-type *P. tricornutum*; ΔUMPS1, PtUMPS knockout strain 1; ΔUMPS1-A1, ΔUMPS1 possessing pPtUMPSA1; ΔUMPS1-A2, ΔUMPS1 possessing pPtUMPSA2; ΔUMPS1-cA1, ΔUMPS1 possessing pPtUMPScA1; ΔUMPS1-cA2, ΔUMPS1 possessing pPtUMPScA2. Points represent mean values and error bars represent standard deviation for three replicates.

**Figure 3.13: Assaying ΔUMPS2 knockout and complemented growth rates in L1 media supplemented with uracil, 5-FOA, or both.** A biallelic PtUMPS mutant (ΔUMPS2) is unable to grow in L1 media without uracil supplementation. Reintroducing the gene on a stably replicating plasmid restores the WT phenotype in L1 media. (A) Growth rates in L1 media. (B) Growth rates in L1 media supplemented with uracil. (C) Growth rates in L1 media supplemented with 5-FOA. (D) Growth rates in L1 media supplemented with uracil and 5-FOA. WT, Wild-type *P. tricornutum*; ΔUMPS2, PtUMPS knockout strain 1; ΔUMPS2-A1, ΔUMPS2 possessing pPtUMPSA1; ΔUMPS2-A2, ΔUMPS2 possessing pPtUMPSA2; ΔUMPS2-cA1, ΔUMPS2 possessing pPtUMPScA1; ΔUMPS2-cA2, ΔUMPS2 possessing pPtUMPScA2. Points represent mean values and error bars represent standard deviation for three replicates.

~6-fold increase of cellular orotate levels was observed in the wild-type strain when L1 media was supplemented with uracil (+uracil) as compared to minimal L1 media (-uracil) (Fig. 3.11C). Interestingly, when the ΔUMPS1 strain was grown with uracil supplementation we detected orotate at levels similar to those observed in the wild-type strain grown with uracil. This result suggests that allele 1 in the ΔUMPS1 knockout strain (with an 18-bp in-frame deletion) retains UMPS activity that behaves similarly to the wild-type strain. In contrast, the ΔUMPS2 strain has two out-of-frame deletions that likely abolish ODC and OPRT activity. We speculate that undetectable levels of orotate in the ΔUMPS2 strain may be because it is diverted to another biosynthetic pathway.

### 3.3.4 Plasmid complementation of the uracil and histidine auxotrophs

Plasmid-based complementation of *P. tricornutum* auxotrophs would validate that the Cas9-editing event was the cause of the auxotrophic phenotype, as well as providing alternatives to antibiotic-based selection methods to maintain episomal vectors. We first examined complementation of the uracil-requiring phenotype by cloning both gDNA and cDNA versions of each PtUMPS allele with the native promoter and terminator into the nourseothricin-resistant pPtGE31 expression plasmid (6). These plasmids were designated pPtUMPSA1, pPtUMPSA2, pPtUMPScA1, and pPtUMPScA2 (Supplementary Table C.3) and moved into the ΔUMPS1 and ΔUMPS2 strains *via* conjugation. Exconjugants were spot-plated onto solid L1 media with and without uracil and 5-FOA supplementation (Fig. 3.11A). All complemented strains grew on minimal L1 media, while the uncomplemented knockouts did not, confirming expression of the UMPS gene from the pPtGE31 plasmid. No strain grew on 5-FOA alone. Unexpectedly, some of the complemented strains survived on plates supplemented with both 5-FOA and uracil. For example, when ΔUMPS2 was transformed with either of the allele 1 complementation plasmids (pPtUMPSA1 and pPtUMPScA1), clear resistance to 5-FOA in the presence of uracil was observed. The phenotypes observed on solid media were consistent with those observed when the strains were grown in liquid media with similar media supplementation (Fig. 3.12 and Fig. 3.13).

The growth phenotype of the ΔUMPS1 and ΔUMPS2 strains in media supplemented with uracil and 5-FOA could be explained by counter-selection against the plasmid carrying an intact PtUMPS gene that would metabolize 5-FOA to a toxic intermediate. We thus tested for plasmid loss in the

complemented strains by plating the ΔUMPS1 and ΔUMPS2 strains carrying different expression plasmids on solid L1 with and without nourseothricin after 14 days of growth. As shown in Fig. 3.11D, plasmid retention, as measured by the ratio of colonies on L1 plus nourseothricin versus L1 plates, was severely reduced in all strains, ranging from ~1% to ~33%. This observation could explain why colonies readily appeared on L1 media supplemented with 5-FOA and uracil and suggest that curing of plasmids carrying the PtUMPS gene from PtUMPS knockout strains is a simple matter of growth on the appropriate media.

Similarly, we were able to complement the histidine-requiring phenotype by cloning a wild-type copy of the PtPRA-PH/CH gene into an expression vector, and transforming the plasmid into the ΔPRAPHCH1 strain by conjugation. The ΔPRAPHCH1 strain with the complementing plasmid grew on both solid L1 media with and without histidine supplementation, whereas the ΔPRAPHCH1 strain without the complementing plasmid only grew on L1 media with histidine supplementation (Fig. 3.4C).

## 3.4   Discussion

The available tools for genetic manipulation of *P. tricornutum* and other diatoms have grown substantially in recent years, including the adaptation of TALEN and Cas9 genome-editing nucleases for targeted knockouts as well as plasmid-based and DNA-free methods to deliver the nucleases to cells (6–9, 11, 40, 41). Applications of genome-editing nucleases in *P. tricornutum* at this point have mostly been to generate gene knockouts, with a few examples of reporter construct knock-ins. Generation of gene knockouts by Cas9 or other editing enzymes relies on non-homologous end-joining repair pathways (42), homologs of which are predicted to occur in the *P. tricornutum* genome (32, 33). A recent study used antisense RNA to knockdown a predicted DNA ligase IV homolog (*ligIV*) in *P. tricornutum* resulting in an increased rate of homologous recombination of a reporter construct (43). What is not yet known for *P. tricornutum* is the balance between NHEJ and homology directed repair (HDR) pathways that process endonuclease-introduced double-strand breaks (see for example (44)). Examination of Cas9 or TALEN-edited sites in *P. tricornutum* revealed small nucleotide insertions or deletions localized near the editing site that are consistent with NHEJ repair events. It is becoming increasingly apparent that repair of Cas9-edited sites

result in heterogenous alleles often characterized by both small and large deletions (for example, Fig. 3.2D)(6, 8, 45). Repair events leading to large indels are often missed by experimental strategies that examine repair outcomes localized around the editing site. In contrast, large deletions visible by long-read sequencing methodologies and our Nanopore data indicate that Cas9 and Tev-Cas9 editing events result in deletions up to ~2.7 kb in length. Cas9 editing with a single sgRNA in *P. tricornutum* could achieve the same goal as the paired Cas9 nickase strategy to specifically introduce large deletions (46), and may be complementary with recently developed methods to multiplex sgRNAs on Cas9-editing plasmids for *P. tricornutum* (10). Regardless, a better understanding of DNA repair pathways that operate on Cas9-introduced double-strand breaks will better inform strategies to bias repair events depending on the experimental goal.

The creation of auxotrophic strains of *P. tricornutum* with plasmid based rather than chromosomally integrated complementation markers is critical for a number of reasons. Auxotrophic strains expand the available selection schemes beyond traditional antibiotic markers and provide a facile method for strain cataloging and validation. Antibiotic-free selection is also an advantage when *P. tricornutum* is used for production of human therapeutics. In the case of uracil auxotrophs, complementing plasmids can be cured (or counter selected) by simple inclusion of 5-FOA and uracil in the growth media. We have previously shown that plasmids are lost from *P. tricornutum* by passaging cultures over multiple days in the absence of antibiotic selection required for maintenance of the plasmid (5). However, the counter selection method by 5-FOA and uracil supplementation is more rapid and requires screening significantly fewer colonies to confirm plasmid loss. The ability to rapidly cure plasmids will be of tremendous value to prevent prolonged expression of Cas9 and possible toxicity issues during strain engineering, to cure incompatible plasmids, or to cure reporter or expression plasmids under distinct growth conditions. We also envision that rapid curing of plasmids would allow recycling of a limited number of selection markers for serial transformations needed for strain construction or genomic engineering.

## 3.5   References

[1] Scaife, M. A.; Smith, A. G. Towards developing algal synthetic biology. *Biochemical Society Transactions* **2016**, *44*, 716–22.

[2] Wang, B.; Wang, J.; Meldrum, D. R. Application of synthetic biology in cyanobacteria and algae. *Frontiers in Microbiology* **2012**, *3*, 1–15.

[3] Yongmanitchai, W.; Ward, O. P. Growth of and omega-3 fatty acid production by Phaeodactylum tricornutum under different culture conditions. *Applied and Environmental Microbiology* **1991**, *57*, 419–425.

[4] Bozarth, A.; Maier, U. G.; Zauner, S. Diatoms in biotechnology: Modern tools and applications. 2009.

[5] Karas, B. J.; Diner, R. E.; Lefebvre, S. C.; McQuaid, J.; Phillips, A. P.; Noddings, C. M.; Brunson, J. K.; Valas, R. E.; Deerinck, T. J.; Jablanovic, J., et al. Designer diatom episomes delivered by bacterial conjugation. *Nature Communications* **2015**, *6*, 1–10.

[6] Slattery, S. S.; Diamond, A.; Wang, H.; Therrien, J. A.; Lant, J. T.; Jazey, T.; Lee, K.; Klassen, Z.; Desgagné-Penix, I.; Karas, B. J.; Edgell, D. R. An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum. *ACS Synthetic Biology* **2018**, *7*, 328–338.

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

[8] Stukenberg, D. D.; Zauner, S. S.; Dell'Aquila, G.; Maier, U. G. Optimizing CRISPR/Cas9 for the diatom Phaeodactylum tricornutum. *Frontiers in Plant Science* **2018**, *9*, 740.

[9] Sharma, A. K.; Nymark, M.; Sparstad, T.; Bones, A. M.; Winge, P. Transgene-free genome editing in marine algae by bacterial conjugation–comparison with biolistic CRISPR/Cas9 transformation. *Scientific Reports* **2018**, *8*, 14401.

[10] Moosburner, M. A.; Gholami, P.; McCarthy, J. K.; Tan, M.; Bielinski, V. A.; Allen, A. E. Multiplexed knockouts in the model diatom Phaeodactylum by episomal delivery of a selectable Cas9. *Frontiers in Microbiology* **2020**, *11*.

[11] Hopes, A.; Nekrasov, V.; Kamoun, S.; Mock, T. Editing of the urease gene by CRISPR-Cas in the diatom Thalassiosira pseudonana. *Plant Methods* **2016**, *12*, 49.

[12] Gatignol, A.; Durand, H.; Tiraby, G. Bleomycin resistance conferred by a drug-binding protein. *FEBS Letters* **1988**, *230*, 171–5.

[13] Zaslavskaia, L. A.; Casey Lippmeier, J.; Kroth, P. G.; Grossman, A. R.; Apt, K. E. Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a variety of selectable marker and reporter genes. *Journal of Phycology* **2000**, *36*, 379–386.

[14] Krügel, H.; Fiedler, G.; Smith, C.; Baumberg, S. Sequence and transcriptional analysis of the nourseothricin acetyltransferase-encoding gene nat1 from Streptomyces noursei. *Gene* **1993**, *127*, 127–131.

[15] Buck, J. M.; Bártulos, C. R.; Gruber, A.; Kroth, P. G. Blasticidin-S deaminase, a new selection marker for genetic transformation of the diatom Phaeodactylum tricornutum. *PeerJ* **2018**, *6*, e5884.

[16] Chassy, B. M. Food safety evaluation of crops produced through biotechnology. *Journal of the American College of Nutrition* **2002**, *21*, 166S–173S.

[17] Romero, J.; Feijoó, C. G.; Navarrete, P. Antibiotics in aquaculture–use, abuse and alternatives. *Health and Environment in Aquaculture* **2012**, *159*.

[18] Peterbauer, C.; Maischberger, T.; Haltrich, D. Food-grade gene expression in lactic acid bacteria. *Biotechnology Journal* **2011**, *6*, 1147–1161.

[19] Pronk, J. T. Auxotrophic Yeast Strains in Fundamental and Applied Research. *Applied and Environmental Microbiology* **2002**, *68*, 2095–2100.

[20] Newman, E.; Miller, B.; Colebrook, L.; Walker, C. A mutation in Escherichia coli K-12 results in a requirement for thiamine and a decrease in L-serine deaminase activity. *Journal of Bacteriology* **1985**, *161*, 272–276.

[21] Goldstein, A.; Goldstein, D. B.; Brown, B. J.; Chou, S.-C. Amino acid starvation in an Escherichia coli auxotroph: I. Effects on protein and nucleic acid synthesis and on cell division. *Biochimica et Biophysica Acta* **1959**, *36*, 163–172.

[22] Sakaguchi, T.; Nakajima, K.; Matsuda, Y. Identification of the UMP Synthase Gene by Establishment of Uracil Auxotrophic Mutants and the Phenotypic Complementation System in the Marine. *Plant Physiology* **2011**, *156*, 78–89.

[23] Noskov, V. N.; Karas, B. J.; Young, L.; Chuang, R. Y.; Gibson, D. G.; Lin, Y. C.; Stam, J.; Yonemoto, I. T.; Suzuki, Y.; Andrews-Pfannkoch, C.; Glass, J. I.; Smith, H. O.; Hutchison, C. A.; Venter, J. C.; Weyman, P. D. Assembly of large, high G+C bacterial DNA fragments in yeast. *ACS Synthetic Biology* **2012**, *1*, 267–273.

[24] Gibson, D. G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Research* **2009**, *37*, 6984–6990.

[25] Engler, C.; Kandzia, R.; Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **2008**, *3*, e3647.

[26] Wang, H.; Slattery, S.; Karas, B.; Edgell, D. Delivery of the Cas9 or TevCas9 System into Phaeodactylum tricornutum via Conjugation of Plasmids from a Bacterial Donor. *Bio-Protocol* **2018**, *8*, e2974.

[27] De Coster, W.; D'Hert, S.; Schultz, D. T.; Cruts, M.; Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **2018**, *34*, 2666–2669.

[28] Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100.

[29] Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079.

[30] Pedersen, B. S.; Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **2018**, *34*, 867–868.

[31] Wickham, H. *ggplot2: elegant graphics for data analysis*; Springer, 2016.

[32] Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **2000**, *28*, 27–30.

[33] Kanehisa, M.; Sato, Y.; Furumichi, M.; Morishima, K.; Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Research* **2019**, *47*, D590–D595.

[34] Chiariotti, L.; Alifano, P.; Carlomagno, M. S.; Bruni, C. B. Nucleotide sequence of the Escherichia coli hisD gene and of the Escherichia coli and Salmonella typhimurium hisIE region. *Molecular and General Genetics MGG* **1986**, *203*, 382–388.

[35] Ingle, R. A. Histidine Biosynthesis. *Arabidopsis Book* **2011**, 1–9.

[36] Wolfs, J. M.; Hamilton, T. A.; Lant, J. T.; Laforet, M.; Zhang, J.; Salemi, L. M.; Gloor, G. B.; Schild-Poulter, C.; Edgell, D. R. Biasing genome-editing events toward precise length deletions with an RNA-guided TevCas9 dual nuclease. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 14988–14993.

[37] Russo, M. T.; Cigliano, R. A.; Sanseverino, W.; Ferrante, M. I. Assessment of genomic changes in a CRISPR/Cas9 Phaeodactylum tricornutum mutant through whole genome resequencing. *PeerJ* **2018**, *6*, e5507.

[38] Kosicki, M.; Tomberg, K.; Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology* **2018**, *36*, 765–771.

[39] Gorter de Vries, A. R.; Couwenberg, L. G.; van den Broek, M.; De La Torre Cortés, P.; ter Horst, J.; Pronk, J. T.; Daran, J.-M. G. Allele-specific genome editing using CRISPR–Cas9 is associated with loss of heterozygosity in diploid yeast. *Nucleic Acids Research* **2019**, *47*, 1362–1372.

[40] Weyman, P. D.; Beeri, K.; Lefebvre, S. C.; Rivera, J.; Mccarthy, J. K.; Heuberger, A. L.; Peers, G.; Allen, A. E.; Dupont, C. L. Inactivation of Phaeodactylum tricornutum urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnology Journal* **2015**, *13*, 460–470.

[41] Daboussi, F.; Leduc, S.; Maréchal, A.; Dubois, G.; Guyot, V.; Perez-Michaut, C.; Amato, A.; Falciatore, A.; Juillerat, A.; Beurdeley, M.; Voytas, D. F.; Cavarec, L.; Duchateau, P. Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology. *Nature Communications* **2014**, *5*, 3831.

[42] Yeh, C. D.; Richardson, C. D.; Corn, J. E. Advances in genome editing through control of DNA repair pathways. *Nature Cell Biology* **2019**, 1–11.

[43] Angstenberger, M.; Krischer, J.; Aktas, O.; Buchel, C. Knock-Down of a ligIV Homologue Enables DNA Integration via Homologous Recombination in the Marine Diatom Phaeodactylum tricornutum. *ACS Synthetic Biology* **2018**, *8*, 57–69.

[44] Certo, M. T.; Ryu, B. Y.; Annis, J. E.; Garibov, M.; Jarjour, J.; Rawlings, D. J.; Scharenberg, A. M. Tracking genome engineering outcome at individual DNA breakpoints. *Nature Methods* **2011**, *8*, 671.

[45] Serif, M.; Lepetit, B.; Weißert, K.; Kroth, P.; Rio Bartulos, C. A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom Phaeodactylum tricornutum. *Algal Research* **2017**, *23*, 186–195.

[46] Mali, P.; Aach, J.; Stranges, P. B.; Esvelt, K. M.; Moosburner, M.; Kosuri, S.; Yang, L.; Church, G. M. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology* **2013**, *31*, 833.

# Chapter 4

# A plasmid-based system for expression and secretion of the SARS-CoV-2 spike protein in the model diatom *Phaeodactylum tricornutum*

## 4.1   Introduction

The spread of the SARS-CoV-2 virus (Severe acute respiratory syndrome coronavirus 2) is currently causing a global pandemic with major impacts on healthcare systems and economies. Rapid, accurate testing is required to help fight the spread and reduce the burden associated with this pandemic. There are two main methods of testing: nucleic acid tests that allow for the detection of active infections, and serological assays that screen blood samples for the presence of antibodies against the virus. While serological assays are not suitable for detecting active infections, they can determine who has been previously exposed to the virus and are now potentially immune or resistant to reinfection. This type of testing is critical, especially for front line healthcare workers, as it identifies individuals who are at low risk from the virus and may be safe to return to work. A recent study has provided evidence that reinfection does not occur in primates who have developed antibodies against the virus (1). Individuals who have been identified to possess antibodies against the virus are also candidates for serum/plasma donation to develop antiviral therapies. Thus, large scale serological testing will be a major advantage in the fight against the SARS-CoV-2 virus.

One of the main barriers to development of serological assays has been the availability of a cheap, reliable source of serologically-reactive SARS-CoV-2 proteins. Serological assays require

purified viral antigens that are used to react with and thereby indicate the presence of antibodies against the virus in blood samples. One of the most promising viral antigens is the SARS-CoV-2 spike protein (2, 3). Coronavirus spike proteins, and in particular their receptor-binding domain (RBD), have been show to be highly immunogenic and antibodies against the spike protein and its RBD domain are commonly found in individuals who have survived infections with previous coronaviruses (4). They are therefore a good candidate for use in serological assays against this most recent coronavirus. Production of viral antigens is currently achieved in engineered mammalian and insect culture due to their ability to properly post-translationally modify the antigens, namely N-linked glycosylation. Glycosylation of the spike protein has been shown to be a requirement for proper antigen recognition by neutralizing antibodies (5). Production of viral antigens in these systems is viable for small-scale production, but is prohibitively expensive for large-scale production, especially at the levels that would be required for population-wide serological testing. These systems also require specialized equipment and complicated supply chains, biosafety certifications, and can have low protein yields. Thus, there is a need for a production system that is cheap, scalable, and has the ability to produce serologically-reactive viral antigens.

The model diatom *P. tricornutum* has the potential to be an efficient, cost-effective biological platform for large-scale viral antigen production. *P. tricornutum* is a photoautotrophic marine microalga with a fast growth rate which can be grown relatively cheaply even at industrial-scales. It has a variety of genetic tools for genome modification and overexpression of transgenic proteins available (6, 7). *P. tricornutum* possesses a glycosylation pathway that is similar to that of humans and is therefore likely to properly glycosylate expressed viral antigens (8, 9). Additionally, production of biologically active, glycosylated proteins has been demonstrated in *P. tricornutum* engineered to express the hepatitis B virus surface antigen and human antibodies against the antigen (10, 11). Secretion of viral antigens from *P. tricornutum* would mimic mammalian expression systems and facilitate downstream purification. This strategy is promising as secretion of transgenic proteins using endogenous and exogenous signal peptides has been previously demonstrated in *P. tricornutum* (11, 12). In addition, *P. tricornutum* secretes very few proteins, potentially simplifying downstream purification of secreted viral antigens. Here, we report the use of a plasmid-based system to overexpress and secrete the SARS-CoV-2 spike protein and RBD domain in *P. tricornutum*. This system will aid in the production of viral antigens to facilitate population-wide

serological testing to combat the current SARS-CoV-2 pandemic, and can be rapidly repurposed to tackle future pandemics.

## 4.2    Materials and Methods

### 4.2.1    Microbial strains and growth conditions

*S. cerevisiae* VL6-48 (ATCC MYA-3666: *MATα his3-δ200 trp1-δ1 ura3-52 lys2 ade2-1 met14 °*) was grown as described in section 1.2.1. *Escherichia coli* (Epi300, Epicenter) was grown as described in section 1.2.1. *P. tricornutum* (Culture Collection of Algae and Protozoa CCAP 1055/1) was grown in L1 medium without silica, with or without histidine (200 mg l$^{-1}$), supplemented with appropriate antibiotics Zeocin$^{TM}$ (50 mg l$^{-1}$) or nourseothricin (100 mg l$^{-1}$), at 18°C under cool white fluorescent lights (75 $\mu$E m$^{-2}$ s$^{-1}$) and a photoperiod of 16 h light:8 h dark. L1 media supplemented with nourseothricin contained half the normal amount of aquil salts.

### 4.2.2    Plasmid design and construction

All plasmids (Supplementary Table D.1) were constructed using a modified yeast assembly protocol (13, 14). We cloned versions of the SARS-CoV-2 spike protein gene into the *E. coli / P. tricornutum* pDMI-2 shuttle plasmid, also called pPtGE31 (6). Primers are listed in Supplementary Table D.2. We obtained SARS-CoV-2 expression plasmids from the Krammer lab in NY city. These served as templates for PCR amplification of the human codon-optimized spike and receptor-binding domain coding regions for cloning into *P. tricornutum* expression plasmids. We also ordered synthetic constructs corresponding to the full-length spike protein gene and RBD from IDT-DNA that were codon-optimized for *P. tricornutum*. The nine initial constructs are listed in Supplementary Table D.1 and schematics can be seen in Figure 4.1. The constructs differed in the following ways: codon-optimization for human or *P. tricornutum*, full-length or RBD of the spike protein, version 1 or version 2 of the *P. tricornutum HASP1* promoter (originating from two homologous *P. tricornutum* chromosomes). We also cloned a version of the full-length construct with two proline stabilizing mutations and a mutation of the furin cleavage site, in addition to a stabilizing trimerization motif. The constructs were made with a promoter from the *P. tri-*

*cornutum 40SRPS8* (40S ribosomal protein S8) gene or from the *HASP1* gene (highly abundant secreted protein 1). All constructs contained the *40SRPS8* terminator. All constructs also included a HIS4 (PtPRA-PH/CH) expression cassette from the pPtPRAPHCH plasmid (7) for selection and maintenance in a *P. tricornutum* histidine auxotroph strain. Plasmids encoding the *P. tricornutum* codon-optimized versions had the *HASP1* promoter and the HASP1 secretory signal peptide, while plasmids encoding the human codon-optimized versions used the *40SRPS8* promoter and the native spike protein secretory signal. The plasmid constructs were assembled in yeast by co-transforming linear DNA fragments of the pDMI-2 plasmid backbone, the HIS4 expression cassette, the *HASP1* or *40SRPS8* promoter, and the spike or RBD protein gene. Resultant yeast colonies were pooled, DNA extracted, and transformed into *E. coli*. Single *E. coli* colonies were grown and plasmid DNA isolated using standard plasmid mini-prep kits. Correct assembly was confirmed by restriction enzyme digests, by whole-plasmid sequencing using a Minion sequencer from Nanopore, and by sequencing of the spike protein expression cassette at the London Regional Genomics Centre.

### 4.2.3  Transfer of DNA to *P. tricornutum via* conjugation from *E. coli*

Conjugations were performed as previously described (6, 15). Briefly, liquid cultures (250 $\mu$l) of *P. tricornutum* were adjusted to a density of $1.0 \times 10^8$ cells ml$^{-1}$ using counts from a hemocytometer, plated on $^1/_2 \times$ L1 1% agar plates with or without histidine (200 mg l$^{-1}$), and grown for four days. L1 media (1.5 ml) was added to the plate and cells were scraped and the concentration was adjusted to $5.0 \times 10^8$ cells ml$^{-1}$. *E. coli* cultures (50 ml) were grown at 37°C to A$_{600}$ of 0.8–1.0, centrifuged for 10 mins at $3,000 \times g$ and resuspended in 500 $\mu$l of SOC media. Conjugation was initiated by mixing 200 $\mu$l of *P. tricornutum* and 200 $\mu$l of *E. coli* cells. The cell mixture was plated on $^1/_2 \times$ L1 5% LB 1% agar plates, incubated for 90 mins at 30°C in the dark, and then moved to 18°C in the light and grown for 2 days. After 2 days, L1 media (1.5 ml) was added to the plates, the cells scraped, and 300 $\mu$l (20%) plated on $^1/_2 \times$ L1 1% agar plates supplemented with Zeocin$^{TM}$ 50 mg l$^{-1}$ or nourseothricin 100 mg l$^{-1}$. Colonies appeared after 7–14 days incubation at 18°C with light.

### 4.2.4 Measuring *P. tricornutum* growth rates

Growth was measured in a Multiskan Go microplate spectrophotometer. Wild-type *P. tricornutum* was adjusted to $5 \times 10^5$ cells ml$^{-1}$ in L1 media with and without the following additives or conditions. L1 plus 0.05 M, 0.1 M, or 0.2 M glycerol; L1 plus 5%, 10%, 25%, or 50% (v/v) final concentration Luria Broth (LB); L1 plus 5%, 10%, 25%, or 50% (v/v) final concentration Soper optimal broth with catabolite repression (SOC); L1 supplemented with tryptone (1 g l$^{-1}$); L1 supplemented with yeast extract (0.5 g l$^{-1}$); L1 supplemented with yeast extract and tryptone; L1 prepared with 10x diluted nitrate phosphate stock buffer. Two hundred microliters of each adjusted culture was added to three wells (technical replicates) of a 96-well microplate. The 96-well microplates were incubated at 18°C under cool white fluorescent lights (75 $\mu$E m$^{-2}$ s$^{-1}$) and a photoperiod of 16 h light:8 h dark for 10 days, and absorbance at 670 nm (A$_{670}$) was measured every 24 hrs. The 96-well microplates were shaken briefly to resuspend any settled cells prior to absorbance measurements. Note that the A$_{670}$ was not adjusted for path length and light scattering from the microplate lid and is therefore not directly comparable to optical density readings measured in a standard cuvette.

### 4.2.5 Plasmid stability assay

*P. tricornutum* harbouring different spike protein expression constructs were grown for ~28 days in liquid media. Total DNA was isolated and 1–5 $\mu$l was used to electroporate 40 $\mu$l of *E. coli*. Electroporations were recovered for 60 mins in SOC media at 37°C and then plated on LB plates supplemented with chloramphenicol (25 mg l$^{-1}$) and grown overnight at 37°C. For each electroporation, a minimum of 6 colonies were picked and grown overnight in 5 ml liquid LB with chloramphenicol (25 mg l$^{-1}$), plasmid DNA was isolated using a BioBasic miniprep kit, and digested with appropriate diagnostic restriction enzymes. Plasmids were scored as correct if the digest pattern matched the predicted pattern.

For direct PCR assays, dilutions of RBD protein expression cultures were plated onto L1 with nourseothricin (100 mg l$^{-1}$) and screened for the presence of the RBD gene using a Thermo Scientific Phire Plant Direct PCR Master Mix according to manufacturers instructions. PCR screens used a forward primer located in the *HASP1* promoter (DE5241) for *P. tricornutum* codon-optimized

constructs, or in the *40SRPS8* promoter (DE4130) for human codon-optimized constructs. Reverse primers were positioned inside the RBD domain coding regions (DE5323, DE5326). Screening primers are listed in Supplementary Table D.2.

## 4.2.6   RNAseq analysis

Total RNA was extracted from 50 ml cultures with an $OD_{670}$ of ~0.6–0.7 by first crushing the algal cells in liquid nitrogen as follows. Cultures were centrifuged at $3000 \times g$ for 10 mins at 4°C. The pellet was resuspended in 1 ml TE and added dropwise to a mortar (pre-cooled at -80°C) filled with liquid nitrogen. The frozen droplets were ground into a fine powder with a mortar and pestle, being careful to keep the cells from thawing by adding more liquid nitrogen when necessary. The frozen ground powder was transferred to a new clean 1.5 ml microfuge tube and stored at -80°C. RNA was extracted from 50–100 mg of frozen ground powder with the Monarch Total RNA Miniprep Kit (T2010S). The RNA was stored in TE pH 8.0 at -80°C until use. RNA was then sequenced with a $1 \times 150$ NextSeq 550 mid-output run after rRNA depletion with the Ribo-off Plant rRNA depletion kit. Reads were filtered using Trimmomatic v0.36 with the following settings: AVGQUAL:30 CROP:150 MINLEN:140. Reads counts were generated after mapping to the *P. tricornutum* genome and appropriate plasmid, and counts were generated using htseq-count (16). TPM (Transcripts Per Kilobase Million) values were calculated for the Actin16 (XP_002183295.1), HIS4, NrsR, and RBD genes in each sample. Wild type reads were independently mapped against the pDMI-2 PtRBD c14, pDMI-2 PtRBD c17, and pDMI-2 HsRBD plasmids and subsequent TPM calculations were performed (all plasmids mapped against gave identical values).

## 4.2.7   Protein extraction and purification

*P. tricornutum* cultures (50 ml) were harvested during exponential growth phase and pelleted at $3,000 \times g$ for 10 mins. Cell pellets were resuspended in 5 ml lysis buffer (50 mM Tris-HCl pH 7.4, 0.5 M NaCl, 1 mM PMSF, 1 mM DTT) and sonicated for 5 mins to lyse. Sonicated cells were centrifuged at $10,000 \times g$ for 10 mins at 4°C to pellet cell debris and supernatants were collected in a new tube and stored on ice before purification using a 1 ml GE Healthcare HisTrap

HP Ni-NTA column as follows. Columns were first washed with 10 column volumes of ddH2O, then equilibrated with 10 column volumes of binding/wash buffer (50 mM Tris-HCl pH 7.4, 0.5 M NaCl, 1 mM DTT). Supernatants from lysed cultures were run over equilibrated columns at a flow rate of 1 ml min$^{-1}$ and flowthrough was collected. Columns were washed with 10 ml binding/wash buffer and wash was collected. His-tagged proteins were eluted with 5 ml elution buffer (50 mM Tris-HCl pH 7.4, 0.5 M NaCl, 0.5 M imidazole, 1 mM DTT), collecting 0.5 ml fractions. Samples (20 $\mu$l) of lysis supernatant, flowthrough, wash, and elutions were mixed with 10 $\mu$l of SDS sample loading buffer (187.5 mM Tris-HCl (pH 6.8), 6% (w/v) SDS, 30% [v/v] glycerol, 150 mM DTT, 0.03% (w/v) bromophenol blue, 2% [v/v] $\beta$-mercaptoethanol) and boiled for 5 mins. Samples (10 $\mu$l) were resolved on standard SDS-polyacrylamide gels (15%). Bands were visualized with Coommassie Brilliant Blue and destained with a solution of 40% methanol, 10% acetic acid.

## 4.2.8   Western blots

Samples (20 $\mu$l) of lysis supernatant and Ni-NTA elution (above) were mixed with 10 $\mu$l of SDS sample loading buffer and boiled for 5 mins. Samples (10 $\mu$l) were resolved on standard SDS-polyacrylamide gels (15%) and electroblotted to a polyvinylidene difluoride (PVDF) membrane using a Trans-Blot Turbo Transfer System (BioRad, Hercules, CA, USA). Membranes were incubated for 1 hour in blocking solution (3% bovine serum albumin (BSA), 0.1% Tween 20, 1× TBS) before adding primary antibody at a 1:1000 final dilution (2019-nCoV Spike/RBD, Sino Biological, 40592-T62). Membranes were incubated overnight at 4°C, washed for 3 × 10 mins in washing solution (1% BSA, 0.1% Tween 20, 1× TBS), then incubated with anti-rabbit (Sigma, GENA9340) horse radish peroxidase-linked secondary antibody for 2 hr at 1:2000 final dilution. Membranes were then washed in 1× TBS with 0.1% Tween 20 for 3 × 10 mins, followed by one wash for 10 mins in 1× TBS. Blots were developed using Clarity ECL Western Blotting Substrate (BioRad) following the manufacturer's instructions and imaged with a ChemiDoc XRS+ System (Bio-Rad).

## 4.2.9   Mass spectrometry

Wild-type *P. tricornutum* and cultures (200 ml) harboring different SARS-CoV-2 overexpression plasmids were grown to a density of ~1 × 10$^7$ cells ml$^{-1}$. Cultures were centrifuged at 3000 × *g*

for 10 mins to pellet. Supernatants were collected and filtered through 0.2 $\mu$m vacuum filters to remove cells. Protease inhibitor cocktail (Sigma) was added to the supernatants to 1× concentration and supernatants were concentrated using Thermo Scientific Pierce™ Protein Concentrators (PES) with a 10 kDa molecular weight cutoff at 4°C to a final volume of ~0.3 ml. Protein concentrations were measured using a Bradford assay and ~30 $\mu$g of protein was loaded onto a 12% SDS-PAGE gel. Bands were visualized with Coommassie Brilliant Blue and destained with a solution of 40% methanol, 10% acetic acid. Bands were excised and placed in 1.5 ml microfuge tubes with 500 $\mu$l of 1% acetic acid. Mass spectrometry analysis and peptide identification by ms/ms was performed at the SPARC BioCentre, Sick Kids Hospital, University of Toronto. Peptide data was downloaded from the SPARC BioCentre server and analyzed by the Scaffold software package using a FASTA file of the *P. tricornutum* and SARS-CoV-2 proteomes. The protein threshold was set at 90% and the peptide threshold at a 1% false discovery rate.

## 4.3  Results

### 4.3.1  Design of SARS-CoV-2 antigen expression constructs

We designed nine plasmids overexpressing different versions of the SARS-CoV-2 spike/RBD protein gene, testing different promoter/terminator pairs and secretion signals (Fig. 4.1, Supplementary Table D.1). All constructs were designed to secrete the expressed proteins into the media, and contained nourseothricin resistance and *P. tricornutum* HIS4 (PtPRA-PH/CH) markers for selection in wild-type and histidine auxotroph (7) *P. tricornutum* strains, respectively. We designed *P. tricornutum* codon-optimized and human codon-optimized versions of the full length SARS-CoV-2 spike gene, full length with K986P and V987P substitutions, and soluble RBD domain for expression in our system. The human codon-optimized genes were subcloned from plasmids obtained from the Krammer laboratory (17). The *P. tricornutum* codon-optimized versions were driven by the *HASP1* promoter and included an N-terminal HASP1 secretion signal peptide (12). The human codon-optimized versions were driven by the *40SRPS8* promoter and included the native N-terminal spike protein secretion signal peptide. All constructs contained a C-terminal 6xHis tag to facilitate purification. Sequencing of the *P. tricornutum* codon-optimized constructs identi-

fied two unique sequences of the *HASP1* promoter. The *HASP1* promoter was PCR amplified from wild-type *P. tricornutum* genomic DNA for cloning into the expression constructs. As *P. tricornutum* is diploid, the two sequences identified correspond to different *HASP1* promoter sequences present on two homologous chromosomes. We chose to examining both versions of the *HASP1* promoter for each of the *P. tricornutum* codon-optimized constructs in the event that there were expression differences between them. Expression plasmids were transformed into wild-type and histidine auxotroph *P. tricornutum* strains *via* conjugation (15). Resulting transformants for each construct were pooled and grown in liquid L1 media with nourseothricin antibiotic selection for wild-type transformants, and L1 media without supplementation for histidine auxotroph transformants. All downstream analyses, unless otherwise stated, were performed on cultures of pooled wild-type exconjugants.

## 4.3.2   Optimization of *P. tricornutum* growth conditions

We tested a number of media supplements to maximize *P. tricornutum* growth rates for rapid screening of SARS-CoV-2 spike protein expression. *P. tricornutum* growth in 96-well microtitre plates was measured at $A_{670}$ nM over 17 days. As shown in Figure 4.2, addition of glycerol, tryptone, yeast extract, or a combination of tryptone and yeast extract to the media had the largest impacts on growth rate. These cultures also reached a higher density after 17 days than those grown with L1 media alone. We subsequently used L1 liquid media supplemented with 0.1 M glycerol for all growth experiments. We also tested the impact of glycerol on conjugation frequency of plasmids from *E. coli* to *P. tricornutum*. Bacterial conjugation is a simple and rapid method to move plasmids from common laboratory *E. coli* strains to *P. tricornutum*. The protocol involves growing *P. tricornutum* on L1 plates without antibiotic selection for 4 days before conjugation with *E. coli*. When 0.1 M glycerol was included in the initial L1 plates, conjugation frequency, as judged by the number of exconjugant *P. tricornutum* colonies on antibiotic containing plates, increased by 2.5 fold (Fig. 4.3A). Including glycerol in the selective antibiotic plates did not impact conjugation frequency but colonies were smaller than those on selective plates without glycerol (Fig. 4.3B). These data suggest that inclusion of glycerol in initial plates results in a higher conjugation frequency.
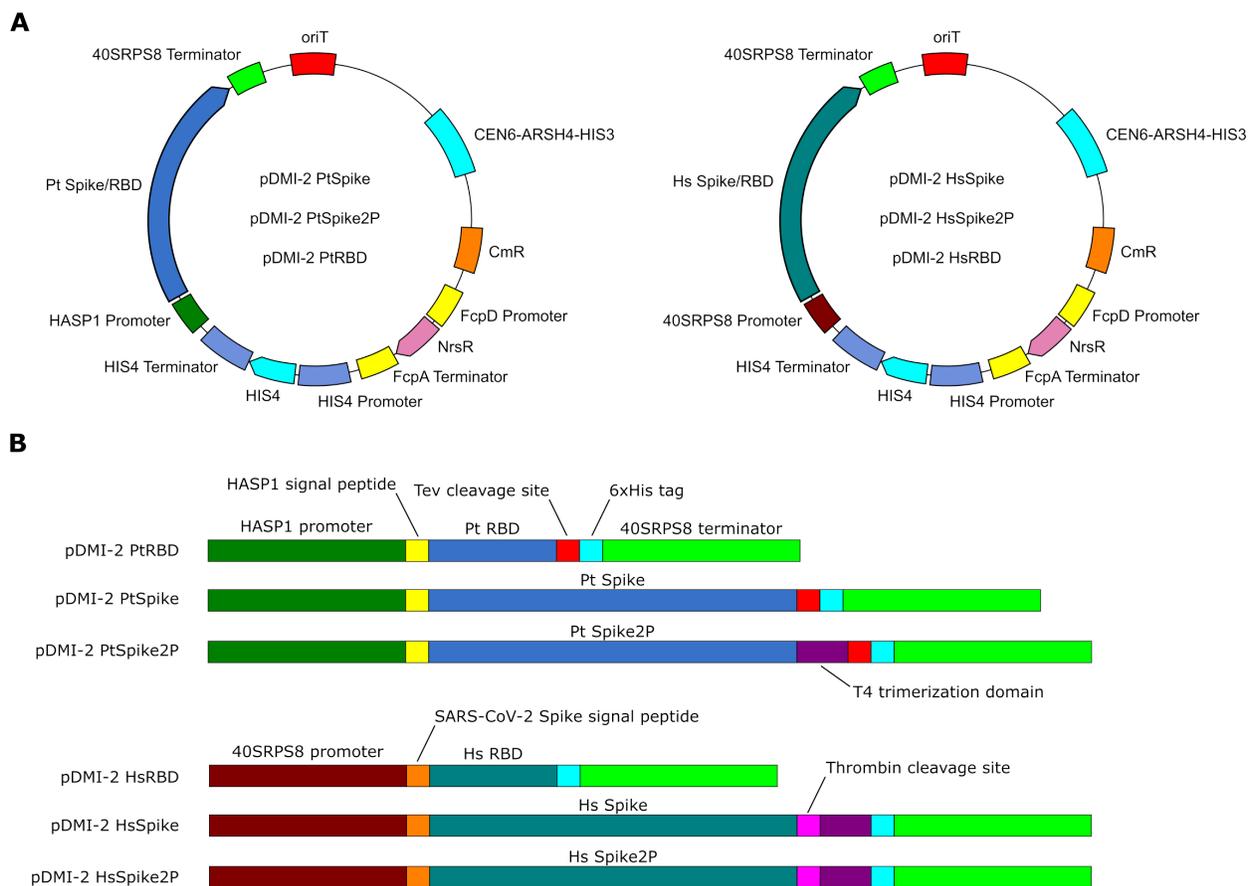
**Figure 4.1: Schematic of plasmid vectors (A) and expression constructs (B) for overexpression of *P. tricornutum* (Pt) and human (Hs) codon-optimized SARS-CoV-2 spike protein and receptor-binding domain (RBD).** CEN6-ARSH4-HIS3, yeast element for replication and selection; CmR, chloramphenicol resistance gene; *FcpD* promoter, *FcpA* terminator, regulatory elements from the *P. tricornutum* Fucoxanthin-chlorophyll a-c binding protein D and A genes, NrsR, nourseothricin resistance gene; HIS4, HIS4 (PtPRA-PH/CH) gene and regulatory elements from *P. tricornutum*; *HASP1* promoter, regulatory element from the *P. tricornutum HASP1* gene; *40SRPS8* promoter and terminator, regulatory elements from the *P. tricornutum 40SRPS8* gene; oriT, origin of conjugative DNA transfer.
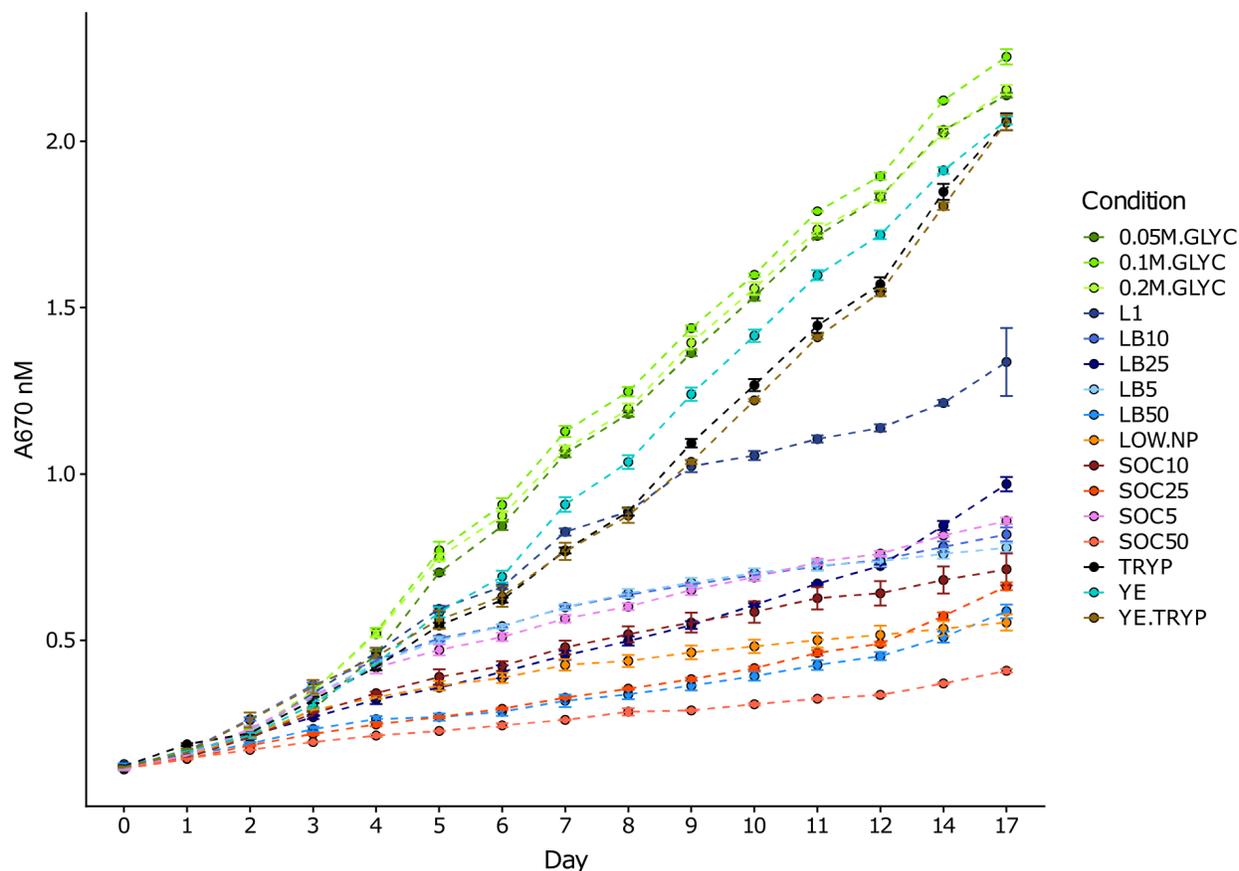
**Figure 4.2: Growth rates of wild-type *P. tricornutum* in L1 media supplemented with different additives.** 0.05M.GLYC, L1 plus 0.05 M glycerol; 0.1M.GLYC, L1 plus 0.1 M glycerol; 0.2M.GLYC, L1 plus 0.2 M glycerol; L1 - no additive; LB5, LB10, LB25, LB50, L1 plus 5%, 10%, 25% or 50% (vv) final concentration Luria Broth; SOC5, SOC10, SOC25, SOC50, L1 plus 5%, 10%, 25%, or 50% (vv) final concentration Super optimal broth with catabolite repression; TRYP, L1 supplemented with tryptone (1 g l$^{-1}$); YE, L1 supplemented with yeast extract (0.5 g l$^{-1}$); YE.TRYP, L1 supplemented with yeast extract and tryptone; LOW.NP, L1 prepared with 10x diluted nitrate phosphate stock buffer.
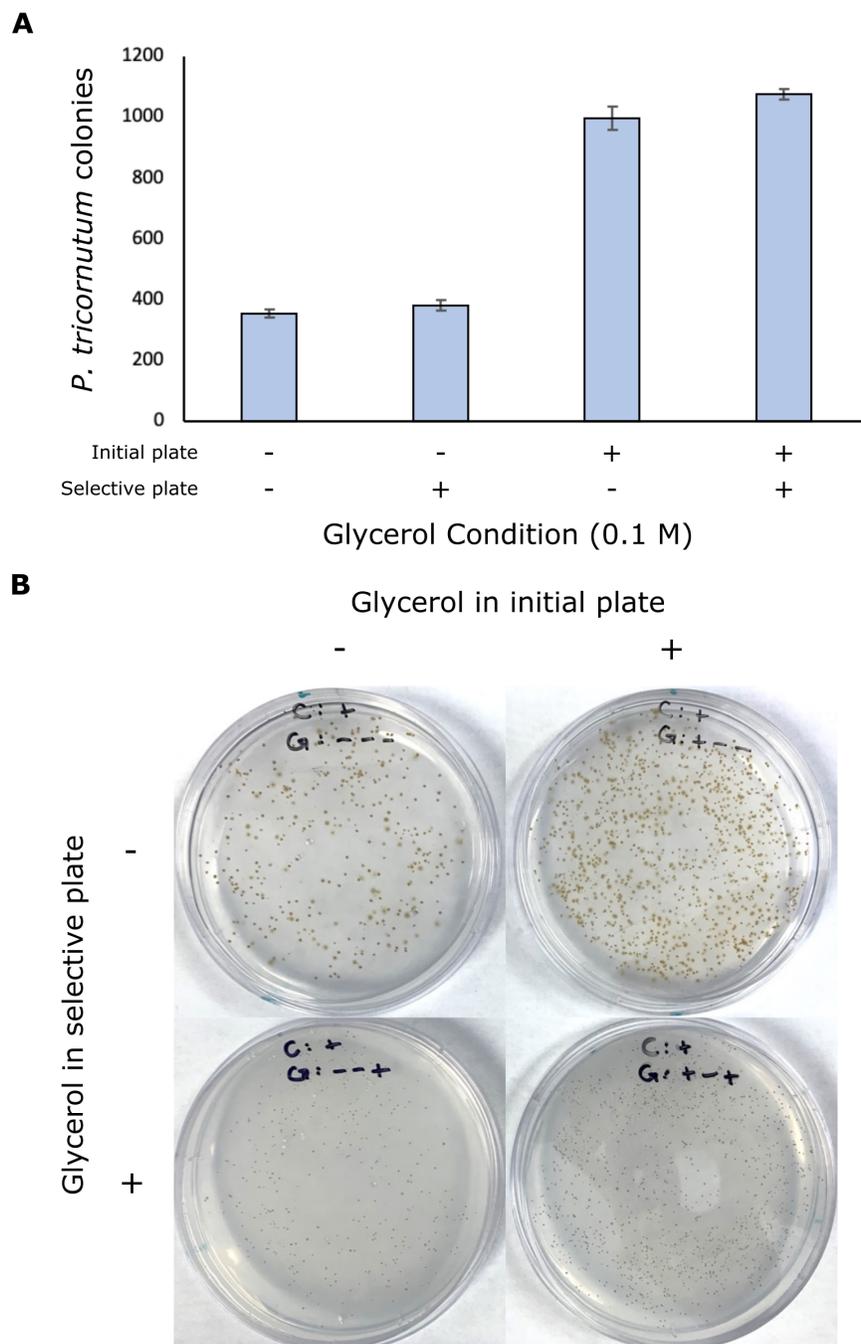
**Figure 4.3: Impact of glycerol on conjugation frequency.** (A) Bar plot of number of *P. tricornutum* exconjugants with 0.1 M glycerol added to L1 plates at various steps of the conjugation protocol. Data are mean of three experimental replicates with standard error of the mean. (B) Representative pictures of L1 plates from conjugations with glycerol added at the indicated steps.

### 4.3.3 Stability of overexpression plasmids in *P. tricornutum*

Stability of the overexpression plasmids in *P. tricornutum* was measured after 28 days growth by isolating total DNA from each *P. tricornutum* culture followed by transformation into *E. coli*. We examined plasmid DNA from at least 6 *E. coli* colonies for each transformation by diagnostic restriction digest (Table 4.1). Of the rearranged plasmids identified, there appeared no consistent stability pattern between the different spike protein expression plasmids, suggesting that rearrangements likely occurred during conjugation from *E. coli* to *P. tricornutum*, as previously observed (6), and did not occur in response to selective pressure to remove the expression construct. The exception was the *P. tricornutum* strain harbouring the pDMI-2 HsSpike plasmid (Fig. 4.1), where no correct clones were observed in 16 total colonies screened. It is important to note that all plasmids retained both the chloramphenicol and nourseothricin resistance genes required for maintenance in *E. coli* and wild-type *P. tricornutum*, respectively.

**Table 4.1: Summary of correct spike expression clones by diagnostic restriction digests.** RBD, receptor-binding domain; Hs, human codon-optimized gene construct, Pt, *P. tricornutum* codon-optimized gene construct; c, refers to clone number.

| Plasmid | Correct / Total Analyzed |
|---|---|
| pDMI-2 PtRBD c14 (*HASP1* v1) | 2/7 |
| pDMI-2 PtRBD c17 (*HASP1* v2) | 2/6 |
| pDMI-2 PtSpike c1 (*HASP1* v1) | 4/6 |
| pDMI-2 PtSpike c2 (*HASP1* v2) | 3/9 |
| pDMI-2 PtSpike2P c1 (*HASP1* v1) | 2/7 |
| pDMI-2 PtSpike2P c2 (*HASP1* v2) | 6/7 |
| pDMI-2 HsRBD | 2/8 |
| pDMI-2 HsSpike | 0/16 |
| pDMI-2 HsSpike2P | 5/8 |

Following this result, we opted to screen and isolate individual *P. tricornutum* clones harbouring correct expression plasmids. To do this we plated dilutions of all wild-type and histidine auxotroph expression cultures onto L1 media supplemented with nourseothricin to obtain single colonies. We began by screening for the presence of the RBD genes by direct PCR of colonies from all RBD expressing strains (Fig. 4.4). For the histidine auxotroph strains, all colonies screened for the human and *P. tricornutum* codon-optimized RBD genes were positive. For the wild-type strains, all colonies screened for the *P. tricornutum* codon-optimized RBD gene were positive,

but only 2 out of 6 colonies screened for the human codon-optimized RBD gene were positive. These results demonstrate that complementation of a histidine auxotroph with a plasmid-based HIS4 (PtPRA-PH/CH) gene is a viable method to stably maintain the RBD expression plasmid in *P. tricornutum* in the absence of antibiotic selection. Colonies that were positive by PCR screen were picked and grown in liquid media for further analysis. However, we have not been able to perform analyses on these clones at this time.
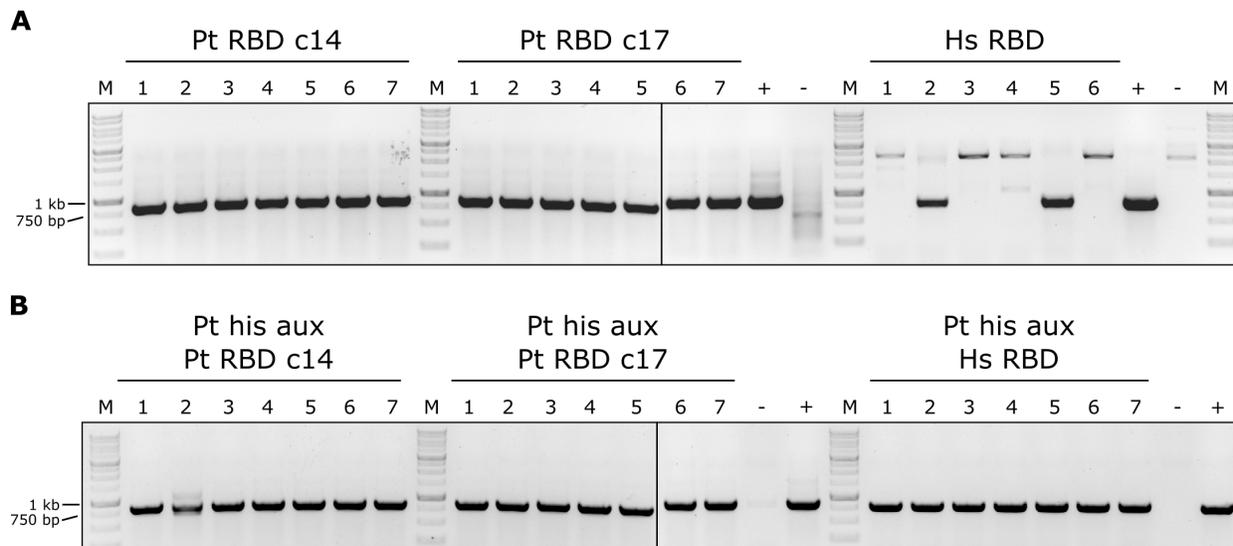


**Figure 4.4: PCR-based screen of *P. tricornutum* expression strain colonies.** Representative images of a PCR-based screen for the RBD gene in single colonies of wild-type (A) and histidine auxotroph (B) *P. tricornutum* strains harbouring RBD expression plasmids. -, no template control; +, PtRBD and HsRBD expression plasmid controls; c14, plasmid clone 14 encoding *HASP1* promoter sequence 1; c17, plasmid clone 17 encoding *HASP1* promoter sequence 2.

### 4.3.4 Analysis of full-length spike protein expression and secretion by mass spectrometry

To examine expression and secretion of full-length spike or RBD proteins, *P. tricornutum* harbouring expression plasmids were grown in 200 ml volumes, and culture supernatants were collected by pelleting cell mass by centrifugation. An aliquot of the cell mass was also taken for further analysis. The supernatants were used for (i) concentration with a spin concentrator with a 10 kDa molecular weight cutoff, (ii) direct application to a metal-ion affinity column for purification by the C-terminal 6xHis tag, or (iii) analysis without further concentration or purification. We found that

option (iii), analysis without further concentration or purification, yielded insufficient material for visualization by SDS-PAGE or Western blot analysis. All future analyses were done using concentrated or purified samples. Concentration of cell supernatants from 200 ml cultures typically gave ~1–2 mg ml$^{-1}$ of total protein. Approximately 30 ng of total protein was resolved on SDS-PAGE gels and visualized by Coomassie blue staining. Example gels are shown in Figure 4.5A. Of particular note are bands that are visible in supernatants from full-length spike overexpressing strains that are absent from wild-type control supernatants. Western blotting of these supernatants using a commercially available anti-spike antibody gave inconclusive results. We then excised the bands from these gels and used mass spectrometry to identify peptides present after in-gel digestion with trypsin. Unique peptides that map to the SARS-CoV-2 spike protein were identified in samples from all spike expression strains except the Pt Spike c1 strain (Table 4.2). No peptides that map to the SARS-CoV-2 spike protein were identified in any wild-type samples. As shown in Figure 4.5B and C, peptide hits map to regions across the SARS-CoV-2 spike protein for some samples, and are localized to specific regions of the protein for others. It should be noted that the excised bands that contained peptides derived from the SARS-CoV-2 spike protein were all smaller than the predicted size of the full-length spike protein (~140 kDa). These results suggest that while it appears the the full-length spike protein is being expressed and secreted into the media, it is likely being degraded.

### 4.3.5   Analysis of spike RBD domain expression by RNASeq and Western blot

Based on protein yields reported for mammalian and insect expression systems (17), we reasoned that the RBD domain would be more highly expressed in *P. tricornutum* than the full-length spike protein, and be less susceptible to proteolysis. We first analyzed expression of the RBD domain by RNASeq in cultures harbouring both *P. tricornutum* and human codon-optimized RBD expression constructs. Reads were mapped against the *P. tricornutum* genome and appropriate plasmids, and TPM (Transcripts Per Kilobase Million) values were calculated for the HIS4, NrsR, and RBD genes encoded on the plasmid, as well as the endogenous Actin16 gene. As shown in Table 4.3, TPM values for the RBD gene were several fold higher than those of the endogenous Actin16 gene
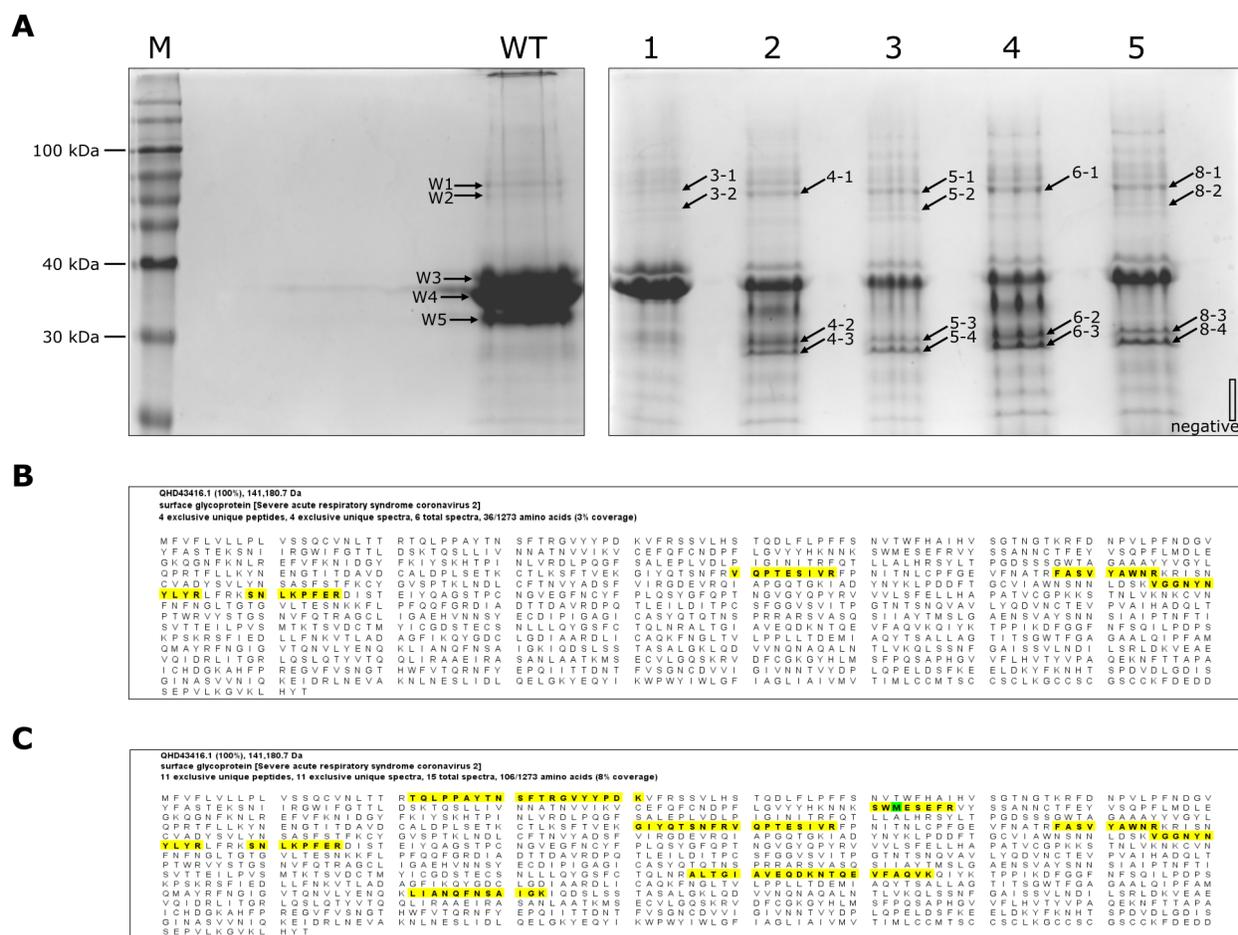
**Figure 4.5: Identification of SARS-CoV-2 spike proteins in supernatants of *P. tricornutum* expression strains.**
(A) SDS-PAGE gels of concentrated supernatants from wild type (WT) and 5 strains expressing full-length spike protein. Arrows indicate bands that were excised and sent for mass spectrometry analyses. M, markers; Lane 1, *P. tricornutum* codon-optimized full-length spike, clone1; Lane 2, *P. tricornutum* codon-optimized full-length spike, clone2; Lane3, *P. tricornutum* codon-optimized full-length spike with K986P and V987P mutations, clone1; Lane4, *P. tricornutum* codon-optimized full-length spike with K986P and V987P mutations, clone2; Lane 5, human codon-optimized full-length spike protein. (B and C) Screenshots of Scaffold software analysis of mass spectrometry data showing peptides from samples 5-4 (B) and 8-3 (C) that map to the SARS-CoV-2 spike protein sequence. Sample IDs are those shown in panel (A) above.

**Table 4.2: List of peptide hits to the SARS-CoV-2 spike protein identified in samples by mass spectrometry.** NA, not applicable (no peptide hits); Hs, human codon-optimized gene construct, Pt, *P. tricornutum* codon-optimized gene construct; c, refers to clone number. Sample names are those shown in Figure 4.5A. The protein threshold was set at 90% and the peptide threshold at a 1% false discovery rate.

| Strain | Sample Name | Unique Peptide Hits | Protein Identification Probability | Percent Sequence Coverage |
|---|---|---|---|---|
| Wild Type | W1 | 0 | NA | NA |
| | W2 | 0 | NA | NA |
| | W3 | 0 | NA | NA |
| | W4 | 0 | NA | NA |
| | W5 | 0 | NA | NA |
| Pt Spike c1 | 3-1 | 0 | NA | NA |
| | 3-2 | 0 | NA | NA |
| Pt Spike c2 | 4-1 | 0 | NA | NA |
| | 4-2 | 2 | 100% | 1.49% |
| | 4-3 | 4 | 100% | 2.83% |
| Pt Spike2P c1 | 5-1 | 0 | NA | NA |
| | 5-2 | 0 | NA | NA |
| | 5-3 | 3 | 100% | 2.20% |
| | 5-4 | 4 | 100% | 2.83% |
| Pt Spike2P c2 | 6-1 | 0 | NA | NA |
| | 6-2 | 3 | 100% | 2.20% |
| | 6-3 | 6 | 100% | 4.10% |
| Hs Spike | 8-1 | 7 | 100% | 5.73% |
| | 8-2 | 9 | 100% | 7.46% |
| | 8-3 | 11 | 100% | 8.33% |
| | 8-4 | 11 | 100% | 7.62% |

**Table 4.3: Expression of the RBD gene from plasmids measured by RNAseq analysis.** Reads were mapped to the *P. tricornutum* genome and appropriate RBD expression plasmid and TPM (Transcripts Per Kilobase Million) values were calculated for the Actin16, HIS4, NrsR, and RBD genes. Wild type reads mapped to each plasmid independently gave identical TPM values for the HIS4, NrsR, and RBD genes. WT, wild type; RBD, receptor-binding domain; Actin16, *P. tricornutum* Actin16 gene; HIS4, *P. tricornutum* HIS4 (PtPRA-PH/CH) gene; NrsR, nourseothricin resistance gene; Hs, human codon-optimized gene construct, Pt, *P. tricornutum* codon-optimized gene construct; c, refers to clone number.

| Strain | Gene | TPM |
|--------|------|-----|
| WT | RBD | 0 |
| | HIS4 | 1.1 |
| | NrsR | 0 |
| | Actin16 | 4.88 |
| Pt RBD c14 | RBD | 60.5 |
| | HIS4 | 9.92 |
| | NrsR | 119 |
| | Actin16 | 13.2 |
| Pt RBD c17 | RBD | 79.6 |
| | HIS4 | 10.8 |
| | NrsR | 44.2 |
| | Actin16 | 6.68 |
| Hs RBD | RBD | 30.8 |
| | HIS4 | 4.93 |
| | NrsR | 39.8 |
| | Actin16 | 10.0 |

and plasmid-encoded HIS4 genes in all RBD expression strains. No expression of the RBD was detected in wild-type cells. As the HIS4 gene is an endogenous gene in the *P. tricornutum* genome in addition to being encoded on expression plasmids, the reads observed for the HIS4 gene in wild type are expected.

We then proceeded to analyze RBD expression by Western blotting with an anti-RBD antibody on whole cell extracts. As shown in Figure 4.6, the RBD could be detected in whole cell extracts and in the elution from a nickel-based affinity column purification. No RBD protein was detected in wild-type cells. This result demonstrates expression of the RBD protein from *P. tricornutum* strains harbouring the RBD expression plasmid, and enrichment of the 6xHis tagged RBD domain from whole cell extracts can be achieved by nickel-affinity chromatography.
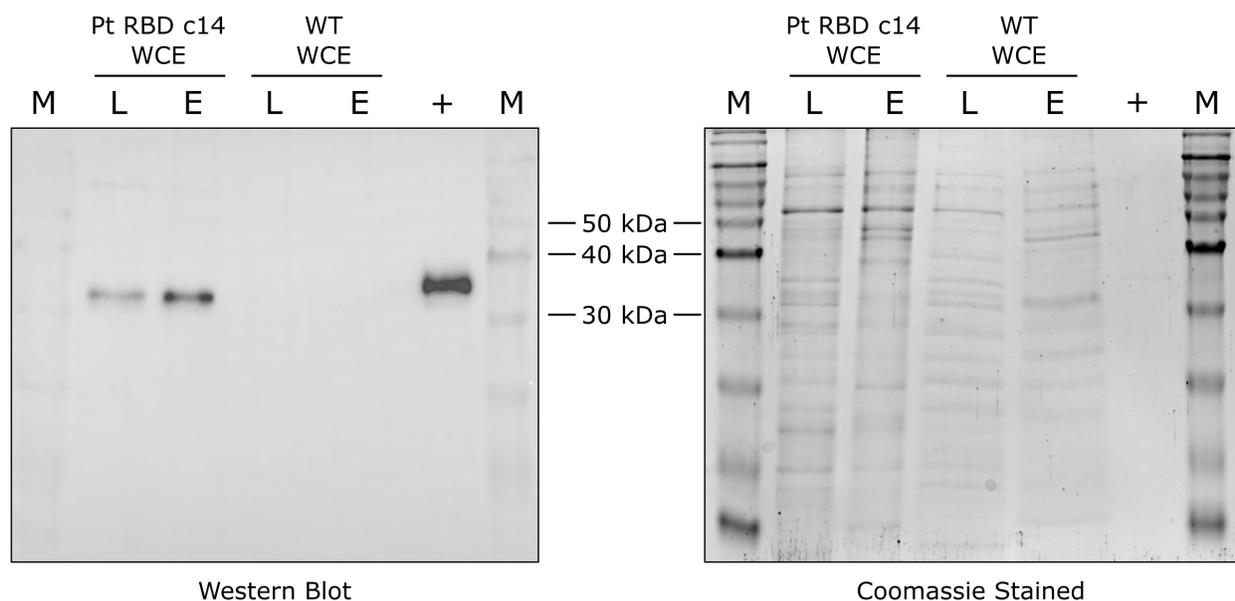


**Figure 4.6: Identification of SARS-CoV-2 spike RBD domain in *P. tricornutum* expression strains.** (Left) Western blot of whole cell extracts (WCE) before (L) enrichment over a nickel-affinity column (E) for a strain harbouring the *P. tricornutum* codon-optimized RBD clone 14 plasmid (Pt RBD c14) and for wild-type cells not harbouring any plasmid (WT). The (+) lane indicates 5 ng of commercially purified RBD. (Right) Coomassie stained replica of the gel that was used for Western blotting.

## 4.4   Discussion

Biological systems like yeast and bacteria are useful protein overexpression systems, but are not suitable for expression of biopharmaceutical and medical reagent proteins that require proper post-

translational modifications to be biologically active (18). Mammalian expression systems can address this but are prohibitively expensive to scale and therefore not appropriate for large-scale protein production. Photoautotrophic microalgae, like the model diatom *P. tricornutum*, can produce post-translationally modified proteins and be grown at scale rapidly and inexpensively. Thus, they have the potential to be a more cost-effective biological platform for industrial-scale production of medically relevant protein reagents. We demonstrate here that the SARS-CoV-2 spike proteins can be expressed and secreted by *P. tricornutum*, and show that *P. tricornutum*-produced SARS-CoV-2 spike receptor-binding domain (RBD) proteins are recognized by commercially available antibodies designed against a mammalian-produced RBD protein. These results suggest the *P. tricornutum*-produced RBD proteins are likely to be serologically-reactive in commercial test kits.

This expression system used strong constitutive promoters to drive expression of the spike and RBD domain proteins at laboratory scales (less than 200 ml). However, sustained overexpression of these proteins at industrial scales may have unforeseen negative effects on culture viability and protein yield. An inducible expression system, using promoters such as the nitrate-inducible nitrate reductase promoter (19–21) may be a more stable alternative and should be explored.

Maintenance of the spike protein expression plasmids in *P. tricornutum* relies on use of an antibiotic (typically zeocin or nourseothricin). While antibiotic selection is feasible in small-scale laboratory cultures, the cost of antibiotics in larger scale cultures will be prohibitive for industrial-scale production of the spike protein. To overcome this problem, we transformed all spike protein expression plasmids into a histidine auxotroph strain of *P. tricornutum* (7). To facilitate this, all spike expression plasmids included the *P. tricornutum* HIS4 marker for selection and maintenance of plasmids in L1 media without the addition of antibiotics. A PCR-based screen revealed that the RBD gene was intact on the expression plasmids in all RBD expression colonies screened, confirming that the HIS4 gene is a viable method to stably maintain the RBD expression plasmids in a *P. tricornutum* histidine auxotroph strain without the need for antibiotics.

Plasmid stability analyses performed on spike protein expression cultures produced seemingly conflicting results. While rearrangements were observed in a large proportion of the plasmids recovered from these cultures, the RBD coding region was present and unmutated in a majority of transformed cells. These results suggest that the rearrangements were random, likely occurring during conjugation, and not due to active selection against overexpression of the RBD domain.

Regardless, this result highlights the need for thorough analysis of plasmid stability in *P. tricornutum* protein overexpression strains, and identification/isolation of transformants with unmutated plasmids for downstream protein expression.

Mass spectrometry analysis of proteins isolated from expression culture media revealed that while full-length spike proteins were being expressed and secreted, they were ultimately being degraded. This is likely occurring *via* proteolysis, as several peptides derived from predicted proteases were also identified by the mass spectrometry analysis. The cultures that were harvested for mass spectrometry analysis were in late-log/stationary phase and might have been expressing proteases as a scavenging mechanism in response to nutrient starvation. Analysis of secreted proteins in cultures in earlier growth stages is needed to investigate this possibility. If proteases are determined to be an issue in all growth stages it may require the genetic knockout of these secreted proteases in order for stable expression and high yield of the full-length spike protein to be achieved.

In this study, we have engineered strains of *P. tricornutum* able to express and secrete SARS-CoV-2 spike proteins, and express SARS-CoV-2 spike receptor-binding domain proteins that mimic commercially available mammalian-produced receptor-binding domain proteins in regards to anti-RBD antibody affinity. This system will potentially allow for cost-effective large-scale production of viral antigens to facilitate widespread serological testing in the fight against the current SARS-CoV-2 pandemic. This system can also be rapidly repurposed to address future pandemics and produce other high-value bioproducts.

## 4.5   References

[1] Bao, L.; Deng, W.; Gao, H.; Xiao, C.; Liu, J.; Xue, J.; Lv, Q.; Liu, J.; Yu, P.; Xu, Y., et al. Reinfection could not occur in SARS-CoV-2 infected rhesus macaques. *BioRxiv* **2020**,

[2] Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367*, 1260–1263.

[3] Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **2020**,

[4] Berry, J. D.; Hay, K.; Rini, J. M.; Yu, M.; Wang, L.; Plummer, F. A.; Corbett, C. R.; An-

donov, A. Neutralizing epitopes of the SARS-CoV S-protein cluster independent of repertoire, antigen structure or mAb technology. *MAbs*. 2010; pp 53–66.

[5] Prabakaran, P.; Gan, J.; Feng, Y.; Zhu, Z.; Choudhry, V.; Xiao, X.; Ji, X.; Dimitrov, D. S. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *Journal of Biological Chemistry* **2006**, *281*, 15829–15836.

[6] Slattery, S. S.; Diamond, A.; Wang, H.; Therrien, J. A.; Lant, J. T.; Jazey, T.; Lee, K.; Klassen, Z.; Desgagné-Penix, I.; Karas, B. J.; Edgell, D. R. An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum. *ACS Synthetic Biology* **2018**, *7*, 328–338.

[7] Slattery, S. S.; Wang, H.; Giguere, D. J.; Kocsis, C.; Urquhart, B. L.; Karas, B. J.; Edgell, D. R. Plasmid-based complementation of large deletions in Phaeodactylum tricornutum biosynthetic genes generated by Cas9 editing. *Scientific Reports* **2020**, *10*, 1–12.

[8] Mathieu-Rivet, E.; Kiefer-Meyer, M.-C.; Vanier, G.; Ovide, C.; Burel, C.; Lerouge, P.; Bardor, M. Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Frontiers in Plant Science* **2014**, *5*, 359.

[9] Baïet, B.; Burel, C.; Saint-Jean, B.; Louvet, R.; Menu-Bouaouiche, L.; Kiefer-Meyer, M.-C.; Mathieu-Rivet, E.; Lefebvre, T.; Castel, H.; Carlier, A., et al. N-glycans of Phaeodactylum tricornutum diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme. *Journal of Biological Chemistry* **2011**, *286*, 6152–6164.

[10] Hempel, F.; Lau, J.; Klingl, A.; Maier, U. G. Algae as protein factories: expression of a human antibody and the respective antigen in the diatom Phaeodactylum tricornutum. *PLoS ONE* **2011**, *6*, e28424.

[11] Hempel, F.; Maier, U. G. An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. *Microbial Cell Factories* **2012**, *11*, 1–6.

[12] Erdene-Ochir, E.; Shin, B.-K.; Kwon, B.; Jung, C.; Pan, C.-H. Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from Phaeodactylum tricornutum. *Scientific Reports* **2019**, *9*, 1–10.

[13] Noskov, V. N.; Karas, B. J.; Young, L.; Chuang, R. Y.; Gibson, D. G.; Lin, Y. C.; Stam, J.; Yonemoto, I. T.; Suzuki, Y.; Andrews-Pfannkoch, C.; Glass, J. I.; Smith, H. O.; Hutchison, C. A.; Venter, J. C.; Weyman, P. D. Assembly of large, high G+C bacterial DNA fragments in yeast. *ACS Synthetic Biology* **2012**, *1*, 267–273.

[14] Gibson, D. G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Research* **2009**, *37*, 6984–6990.

[15] Karas, B. J.; Diner, R. E.; Lefebvre, S. C.; McQuaid, J.; Phillips, A. P.; Noddings, C. M.; Brunson, J. K.; Valas, R. E.; Deerinck, T. J.; Jablanovic, J., et al. Designer diatom episomes delivered by bacterial conjugation. *Nature Communications* **2015**, *6*, 1–10.

[16] Anders, S.; Pyl, P. T.; Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **2015**, *31*, 166–169.

[17] Amanat, F.; Stadlbauer, D.; Strohmeier, S.; Nguyen, T. H.; Chromikova, V.; McMahon, M.; Jiang, K.; Arunkumar, G. A.; Jurczyszak, D.; Polanco, J., et al. A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nature Medicine* **2020**, 1–4.

[18] Berlec, A.; Štrukelj, B. Current state and recent advances in biopharmaceutical production in Escherichia coli, yeasts and mammalian cells. *Journal of Industrial Microbiology & Biotechnology* **2013**, *40*, 257–274.

[19] Kindle, K. L.; Schnell, R. A.; Fernández, E.; Lefebvre, P. A. Stable nuclear transformation of Chlamydomonas using the Chlamydomonas gene for nitrate reductase. *The Journal of Cell Biology* **1989**, *109*, 2589–2601.

[20] Chu, L.; Ewe, D.; Bártulos, C. R.; Kroth, P. G.; Gruber, A. Rapid induction of GFP expression by the nitrate reductase promoter in the diatom Phaeodactylum tricornutum. *PeerJ* **2016**, *4*, e2344.

[21] Niu, Y. F.; Yang, Z. K.; Zhang, M. H.; Zhu, C. C.; Yang, W. D.; Liu, J. S.; Li, H. Y. Transformation of diatom Phaeodactylum tricornutum by electroporation and establishment of inducible selection marker. *BioTechniques* **2012**, *52*, 1–3.

# Chapter 5

# Discussion

The model marine diatom *P. tricornutum* is a promising species for biosynthesis of high-value products including biofuels, omega-3 fatty acids, and biopharmaceuticals. As *P. tricornutum* is photoautotrophic and has a relatively fast growth rate, it is able to be grown inexpensively at large scales. It is a prolific oil producer, being able to accumulate up to 60% dry-cell weight in lipids, making it an attractive species for biofuel production (1), and due to its high ratios of polyunsaturated fatty acids (PUFAs), it is being explored for omega-3 fatty acid production as well (2). Additionally, previous studies suggest *P. tricornutum* is capable of producing proteins with modifications that mimic those made in mammalian cells (3, 4), making it attractive for protein biopharmaceutical production. However, converting *P. tricornutum* into an ideal organism for biofuel, food, and pharmaceutical production requires further genetic and biotechnological development of this species.

Compared to more well-studied microalgae such as *C. reinhardtii*, *P. tricornutum* is relatively undeveloped as a bioproduction strain. As a result, it has limited genetic tools available for genome editing and synthetic pathway development. Consequently, this thesis focused on the biotechnological development of *P. tricornutum* to facilitate engineering of this species for industrial production and synthetic biology applications. We accomplished this by generating novel plasmid-based expression tools, a gene-editing system (5), and valuable strains for *P. tricornutum* (6). We demonstrated the utility of these tools and strains by engineering *P. tricornutum* to produce SARS-CoV-2 antigens to aid in the fight against the current global pandemic.

To improve *P. tricornutum*'s ability to express foreign genes, we identified new endogenous genetic regulatory elements that could drive expression of exogenous genes. These elements suc-

cessfully drove expression of antibiotic resistance markers and reporter genes, and were used to design and assemble a plasmid-based multi-gene biosynthesis pathway that was stably maintained in *P. tricornutum* for 160 generations. While we examined pathway stability, expression of the biosynthetic pathway was not investigated by us, but was to be performed by our collaborators (5). Examining expression of complex synthetic pathways would aid in understanding how artificial multi-gene pathways driven by endogenous promoters behave in this organism. Additionally, growth rates and viability of *P. tricornutum* strains harbouring a desired biosynthetic pathway should be examined to ensure that the pathway is not toxic or a significant metabolic burden, before moving to industrial scales. Further characterization of expression levels of these promoters during various growth stages and under different growth conditions should also be performed to better determine how they behave. This will be critical to understand for the successful design and implementation of future biosynthetic pathways.

To enable precise genome modifications we designed a stably replicating plasmid-based Cas9 gene-editing system for *P. tricornutum*. This system was used to successfully knock out the *P. tricornutum* urease gene and we demonstrated that edited strains could be easily cured of the Cas9 expression plasmid by culturing in the absence of antibiotic selection. Previous Cas9 editing systems relied on genomic integration of the expression construct (7–9), causing prolonged exposure to the nuclease and potentially increasing the risk of off-target cleavage effects. We predict the ability to cure our Cas9 editing system likely decreases the risk of off-target cleavage. A recent study has shown that replicating plasmid-based Cas9 editing systems did not produce mutations at predicted off-target sites (10), supporting our prediction.

Transformation and maintenance of episomal expression constructs in *P. tricornutum* is currently achieved using antibiotics that are prohibitively expensive for growing cultures large enough for industrial-scale chemical or protein production (11). To address this, we created auxotrophic *P. tricornutum* mutants using our Cas9 gene-editing system by knocking out key enzymes in the uracil and histidine biosynthesis pathways. We demonstrated that intact, plasmid-encoded versions of these genes could complement the auxotrophic phenotypes and be used as selective markers for transformation and maintenance of plasmids in these strains. Although these selection systems function robustly, further biological characterization of the auxotrophic strains and how their complementation markers function to maintain large biosynthetic pathways should be performed to

properly evaluate their utility as production strains. Generation of additional auxotrophic strains may also be beneficial, as different auxotrophs may grow more robustly when complemented, or be more suitable to certain biosynthesis applications than those described here.

We observed large deletions generated by Cas9 editing when creating the knockout strains described in chapters 2 and 3. These observations add to a body of evidence that Cas9 editing promotes large deletions which largely go unobserved due to the nature of the screens for Cas9 editing that are commonly used (7, 12, 13). For example, the Surveyor and T7EI assays are commonly used screens for editing which use endonucleases that cleave at mismatched DNA bases in DNA heteroduplexes (14). The substrate DNA for these assays are double-stranded DNA molecules generated by PCR amplifying a ~500 bp region centered around the Cas9 target site from edited and untreated cells, which are then hybridized together. Thus, the assays are only able to examine repair outcomes localized around the editing site. The long-range PCR and Nanopore long-read sequencing methods employed to screen for Cas9 editing events in this study were shown to be good alternatives to the traditional screening methods (15), and can be applied to identify large deletions generated by Cas9 editing in other biological systems. Our results are further supported by a recent study which used both long-range PCR and PacBio long-read sequencing to detect large deletions generated by Cas9 in mammalian cells (16). The large deletions detected here suggest that Cas9 editing with a single sgRNA in *P. tricornutum* could achieve the same result as paired Cas9 nickases that are designed to introduce large deletions (17), and may be an alternative to recently developed methods to multiplex sgRNAs on Cas9-editing plasmids for *P. tricornutum* (18).

To demonstrate the utility of the tools and strains described in chapters 2 and 3, we engineered a *P. tricornutum* strain that expresses the SARS-CoV-2 full-length spike and spike receptor-binding domain (RBD) antigens. These strains were designed to secrete antigens into the media, and while mass spectrometry confirmed full-length spike protein presence in culture supernatants, it appeared to be heavily degraded. The presence of proteases detected in culture supernatants suggests successful purification of secreted spike proteins may require creating secreted-protease-deficient *P. tricornutum* strains. We observed that the RBD antigen was robustly expressed intracellularly and could be purified by nickel-affinity chromatography. *P. tricornutum*-produced RBD proteins were also recognized by anti-RBD antibodies developed against an RBD antigen produced in hu-

man cells. This suggests they may be serologically-reactive, a requirement for use in serological tests. Further analysis of the RBD expression strains is required to determine if the antigen is being secreted and whether secreted RBD protein is recognized by antibodies produced in humans in response to SARS-CoV-2 infection. The RBD domain possesses two sites which undergo N-linked glycosylation in human cells (19). These may be differentially modified in *P. tricornutum* with unknown effects on antibody affinity, although previous studies suggest *P. tricornutum* may perform glycosylation similarly to mammalian systems (3, 4, 20, 21). Currently, SARS-CoV-2 antigens are being produced in mammalian and insect cell culture systems which are prohibitively expensive to scale for large-scale antigen production. *P. tricornutum* cultures are much cheaper to maintain and can easily be grown in industrial scales at a substantially lower cost than would be required for mammalian or insect cultures. If this system proves to produce serologically-reactive SARS-CoV-2 proteins, it will be an efficient and cost-effective method to produce antigens for serological tests. This system can also be rapidly adapted to address future pandemics and could be used to produce functional human therapeutics.

The tools and strains described here have many benefits for the future of this field. Only a small number of secretion signal peptides have been applied for transgenic protein production in *P. tricornutum* (4, 22), and to date no study has been published comparing the efficiency of secretion signal peptides for promoting the secretion of transgenic proteins from *P. tricornutum* into the growth media. The replicating-plasmids and tools developed here will enable rapid screening, not only of signal peptides, but of many genetic elements to further characterize the biology of this organism and expand its utility as a biosynthesis platform.

Efficient production of exogenous proteins in *P. tricornutum* may be hampered by a number of endogenous proteins produced by this species. For example, *P. tricornutum* abundantly secretes a number of proteins into the growth media (22), and produces proteases that may target transgenic proteins and reduce product yields (23, 24). The tools developed here will allow for knockout of these endogenous proteins and proteases to create strains that would enhance stability and simplify the downstream purification of secreted transgenic proteins.

The auxotrophs and complementation systems developed here eliminate the need for expensive antibiotics typically used for selection of transgenic *P. tricornutum* strains. Utilizing these auxotrophic strains for large-scale production would lower the input costs of protein and chemical

biosynthesis in this species, and open up a wide range of production applications to *P. tricornutum* that were previously not economically viable. Although *P. tricornutum* has Generally Recognized As Safe (GRAS) status, it is a photoautotroph which makes biocontainment an issue for certain laboratory and industrial applications of this species. We have shown that the auxotrophs described here do not survive in the absence of supplementation with uracil or histidine. Performing research in these auxotrophic backgrounds would allow for biocontainment of any transgenic *P. tricornutum* strains created. This will potentially enable research on hazardous chemical and protein products to be performed in *P. tricornutum* that would previously have not been possible.

This study is the first report of i) an episomal Cas9 gene-editing system for *P. tricornutum*, and ii) the selection and maintenance of episomal vectors in auxotrophic *P. tricornutum* strains using plasmid-encoded complementation markers. These genetic tools and strains will enable more rapid and more advanced engineering of *P. tricornutum* for protein and chemical biosynthesis applications, and will further the study of diatom biology.

# 5.1   References

[1] Xue, J.; Niu, Y. F.; Huang, T.; Yang, W. D.; Liu, J. S.; Li, H. Y. Genetic improvement of the microalga Phaeodactylum tricornutum for boosting neutral lipid accumulation. *Metabolic Engineering* **2015**, *27*, 1–9.

[2] Yongmanitchai, W.; Ward, O. P. Growth of and omega-3 fatty acid production by Phaeodactylum tricornutum under different culture conditions. *Applied and Environmental Microbiology* **1991**, *57*, 419–425.

[3] Hempel, F.; Lau, J.; Klingl, A.; Maier, U. G. Algae as protein factories: expression of a human antibody and the respective antigen in the diatom Phaeodactylum tricornutum. *PLoS ONE* **2011**, *6*, e28424.

[4] Hempel, F.; Maier, U. G. An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. *Microbial Cell Factories* **2012**, *11*, 1–6.

[5] Slattery, S. S.; Diamond, A.; Wang, H.; Therrien, J. A.; Lant, J. T.; Jazey, T.; Lee, K.; Klassen, Z.; Desgagné-Penix, I.; Karas, B. J.; Edgell, D. R. An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum. *ACS Synthetic Biology* **2018**, *7*, 328–338.

[6] Slattery, S. S.; Wang, H.; Giguere, D. J.; Kocsis, C.; Urquhart, B. L.; Karas, B. J.; Edgell, D. R. Plasmid-based complementation of large deletions in Phaeodactylum tricornutum biosynthetic genes generated by Cas9 editing. *Scientific Reports* **2020**, *10*, 1–12.

[7] Stukenberg, D. D.; Zauner, S. S.; Dell'Aquila, G.; Maier, U. G. Optimizing CRISPR/Cas9 for the diatom Phaeodactylum tricornutum. *Frontiers in Plant Science* **2018**, *9*, 740.

[8] Nymark, M.; Sharma, A. K.; Sparstad, T.; Bones, A. M.; Winge, P. A CRISPR/Cas9 system adapted for gene editing in marine algae. *Scientific Reports* **2016**, *6*, 24951.

[9] Allorent, G.; Guglielmino, E.; Giustini, C.; Courtois, F. *Plastids*; Springer, 2018; pp 367–378.

[10] Russo, M. T.; Cigliano, R. A.; Sanseverino, W.; Ferrante, M. I. Assessment of genomic changes in a CRISPR/Cas9 Phaeodactylum tricornutum mutant through whole genome re-sequencing. *PeerJ* **2018**, *6*, e5507.

[11] Karas, B. J.; Diner, R. E.; Lefebvre, S. C.; McQuaid, J.; Phillips, A. P.; Noddings, C. M.; Brunson, J. K.; Valas, R. E.; Deerinck, T. J.; Jablanovic, J., et al. Designer diatom episomes delivered by bacterial conjugation. *Nature Communications* **2015**, *6*, 1–10.

[12] Adikusuma, F.; Piltz, S.; Corbett, M. A.; Turvey, M.; McColl, S. R.; Helbig, K. J.; Beard, M. R.; Hughes, J.; Pomerantz, R. T.; Thomas, P. Q. Large deletions induced by Cas9 cleavage. *Nature* **2018**, *560*, E8–E9.

[13] Owens, D. D.; Caulder, A.; Frontera, V.; Harman, J. R.; Allan, A. J.; Bucakci, A.; Greder, L.; Codner, G. F.; Hublitz, P.; McHugh, P. J., et al. Microhomologies are prevalent at Cas9-induced larger deletions. *Nucleic Acids Research* **2019**, *47*, 7402–7417.

[14] Vouillot, L.; Thélie, A.; Pollet, N. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3: Genes, Genomes, Genetics* **2015**, *5*, 407–415.

[15] Sentmanat, M. F.; Peters, S. T.; Florian, C. P.; Connelly, J. P.; Pruett-Miller, S. M. A survey of validation strategies for CRISPR-Cas9 editing. *Scientific Reports* **2018**, *8*, 1–8.

[16] Kosicki, M.; Tomberg, K.; Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology* **2018**, *36*, 765–771.

[17] Mali, P.; Aach, J.; Stranges, P. B.; Esvelt, K. M.; Moosburner, M.; Kosuri, S.; Yang, L.; Church, G. M. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology* **2013**, *31*, 833.

[18] Moosburner, M. A.; Gholami, P.; McCarthy, J. K.; Tan, M.; Bielinski, V. A.; Allen, A. E. Multiplexed knockouts in the model diatom Phaeodactylum by episomal delivery of a selectable Cas9. *Frontiers in Microbiology* **2020**, *11*.

[19] Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**,

[20] Mathieu-Rivet, E.; Kiefer-Meyer, M.-C.; Vanier, G.; Ovide, C.; Burel, C.; Lerouge, P.; Bardor, M. Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Frontiers in Plant Science* **2014**, *5*, 359.

[21] Baïet, B.; Burel, C.; Saint-Jean, B.; Louvet, R.; Menu-Bouaouiche, L.; Kiefer-Meyer, M.-C.; Mathieu-Rivet, E.; Lefebvre, T.; Castel, H.; Carlier, A., et al. N-glycans of Phaeodactylum tricornutum diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme. *Journal of Biological Chemistry* **2011**, *286*, 6152–6164.

[22] Erdene-Ochir, E.; Shin, B.-K.; Kwon, B.; Jung, C.; Pan, C.-H. Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from Phaeodactylum tricornutum. *Scientific Reports* **2019**, *9*, 1–10.

[23] Allen, A. E.; LaRoche, J.; Maheswari, U.; Lommer, M.; Schauer, N.; Lopez, P. J.; Finazzi, G.; Fernie, A. R.; Bowler, C. Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. *Proceedings of the National Academy of Sciences* **2008**, *105*, 10438–10443.

[24] Matthijs, M.; Fabris, M.; Broos, S.; Vyverman, W.; Goossens, A. Profiling of the early nitrogen stress response in the diatom Phaeodactylum tricornutum reveals a novel family of RING-domain transcription factors. *Plant Physiology* **2016**, *170*, 489–498.

# Appendix A

# Copyright Permissions for Chapters 2 and 3

The reprint permissions for the following accepted articles are provided:

- Accepted Manuscript 1 details.

- Accepted Manuscript 2 details.

All other copyright material is under sole ownership by the author, including pre-prints and articles currently under submission or revision.

# A.1   Permission for Chapter 2

**Copyright Clearance Center**   RightsLink®

Home | Help | Email Support | Sign in | Create Account

**An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum**

ACS Publications
Most Trusted. Most Cited. Most Read.

**Author:** Samuel S. Slattery, Andrew Diamond, Helen Wang, et al

**Publication:** ACS Synthetic Biology

**Publisher:** American Chemical Society

**Date:** Feb 1, 2018

*Copyright © 2018, American Chemical Society*

# A.2    No copyright permission required for Chapter 3

**Copyright Clearance Center**

**RightsLink®**

**?**
Help

**✉**
Email Support

---

**SPRINGER NATURE**

**Plasmid-based complementation of large deletions in Phaeodactylum tricornutum biosynthetic genes generated by Cas9 editing**

**Author:** Samuel S. Slattery et al

**Publication:** Scientific Reports

**Publisher:** Springer Nature

**Date:** Aug 17, 2020

*Copyright © 2020, Springer Nature*

**Creative Commons**

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

To request permission for a type of use not listed, please contact Springer Nature

# Appendix B

# Supplemental information for Chapter 2

## B.1 Supplemental Figures
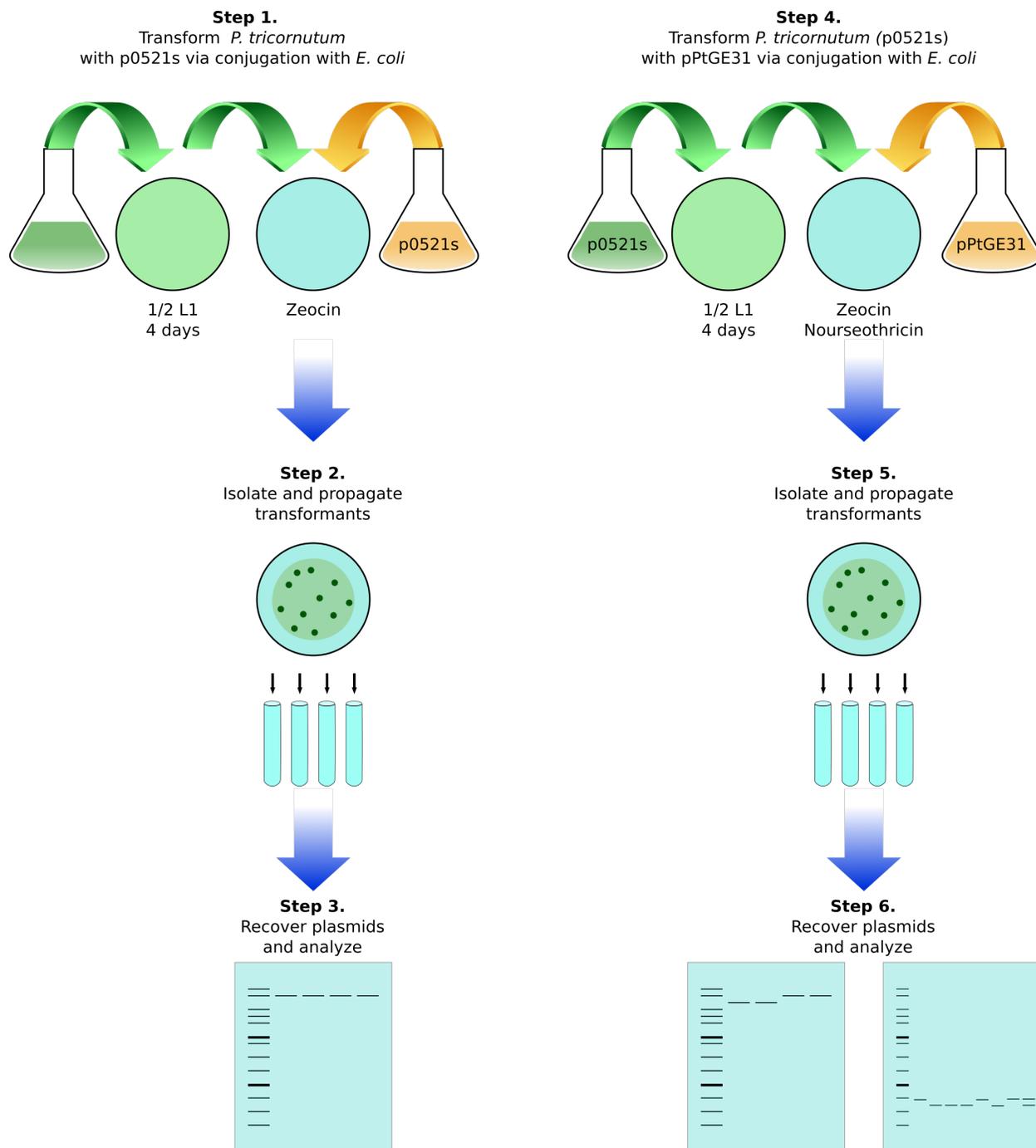
**Figure B.1: Flow chart of two-plasmid propagation experimental design.**

# B.2   Supplemental Tables

**Table B.1: Assaying promoters and terminators in *P. tricornutum*.** Values listed for each replicate represent total exconjugants (calculated by counting surviving colonies after plating 5% of total scraped cells and multiplying by 20.

| Promoter Terminator | Pt colonies per conjugation | | | | Standard Error of the Mean |
|---|---|---|---|---|---|
| | Rep. 1 | Rep. 2 | Rep. 3 | Average | |
| *EF1α* | 2000 | 1760 | 2020 | 1927 | 84 |
| *40SRPS8* | 2880 | 3100 | 3360 | 3113 | 139 |
| *H4-1B* | 4000 | 3020 | 3600 | 3540 | 284 |
| *γ-Tubulin* | 180 | 120 | 120 | 140 | 20 |
| *RBCMT* | 2500 | 1720 | 1400 | 1873 | 327 |
| *FcpB* | 1560 | 1620 | 1820 | 1667 | 79 |
| *FcpC* | 440 | 320 | 320 | 360 | 40 |
| *FcpD* | 4100 | 3680 | 3220 | 3667 | 254 |
| *EF1α* Term | 3000 | 2420 | 2720 | 2713 | 167 |
| *40SRPS8* Term | 3320 | 2840 | 2180 | 2780 | 330 |
| *H4-1B* Term | 3000 | 2360 | 2740 | 2700 | 186 |
| *γ-Tubulin* Term | 1640 | 1640 | 1280 | 1520 | 120 |
| *RBCMT* Term | 3060 | 3020 | 2840 | 2973 | 68 |
| *FcpB* Term | 3300 | 3140 | 2240 | 2893 | 330 |
| *FcpC* Term | 3100 | 2420 | 2460 | 2660 | 220 |
| *FcpD* Term | 2760 | 2180 | 2780 | 2573 | 197 |
| *FcpF* (+ve) | 1980 | 2160 | 1840 | 1993 | 92 |
| Mock conj. (-) | 0 | 0 | 0 | 0 | 0 |

**Table B.2: Two-plasmid propagation in *P. tricornutum*.** Plasmids p0521s and pPtGE31 were sequentially conjugated into *P. tricornutum*

| *P. tricornutum* strain | Plasmid rescue (Number of *E. coli* colonies obtained) | Number of *E. coli* colonies from which correct plasmids were isolated based on I-CeuI digest and multiplex PCR screen | Ratio of p0521s to pPtGE31 isolated from *P. tricornutum* clone |
|---|---|---|---|
| Pt + p0521s clone 1 | 113 | 3/3 | Not applicable |
| Pt + p0521s clone 2 | 0 | 0 | Not applicable |
| Pt + p0521s clone 3 | 0 | 0 | Not applicable |
| Pt + p0521s clone 4 | 149 | 3/3 | Not applicable |
| Pt + p0521s clone 5 | 84 | 2/3 | Not applicable |
| Pt + p0521s clone 1, pPtGE31 clone 1 (c1.1) | 44 | 20/20 | 14:6 |
| Pt + p0521s clone 1, pPtGE31 clone 2 (c1.2) | 55 | Not determined | Not determined |
| Pt + p0521s clone 1, pPtGE31 clone 3 (c1.3) | 29 | Not determined | Not determined |
| Pt + p0521s clone 1, pPtGE31 clone 4 (c1.4) | 29 | Not determined | Not determined |
| Pt + p0521s clone 1, pPtGE31 clone 5 (c1.5) | 47 | Not determined | Not determined |
| Pt + p0521s clone 4, pPtGE31 clone 1 (c4.1) | 72 | 20/20 | 15:5 |
| Pt + p0521s clone 4, pPtGE31 clone 2 (c4.2) | 68 | Not determined | Not determined |
| Pt + p0521s clone 4, pPtGE31 clone 3 (c4.3) | 48 | Not determined | Not determined |
| Pt + p0521s clone 4, pPtGE31 clone 4 (c4.4) | 128 | Not determined | Not determined |
| Pt + p0521s clone 4, pPtGE31 clone 5 (c4.5) | 155 | Not determined | Not determined |

**Table B.3: Vanillin biosynthesis plasmid assembly, propagation and recovery.**

| | V1 | | V7 | | V3 | | V4 | |
|---|---|---|---|---|---|---|---|---|
| Four *E. coli* colonies with correctly assembled pathway plasmid were identified. | | | | | | | | |
| Each correct plasmid was conjugated to *P. tricornutum*, and for each two colonies were isolated for further analysis | c1 | c2 | c1 | c2 | c1 | c2 | c1 | c2 |
| After 40 days of propagation, DNA was extracted from *P. tricornutum* cells and transformed into *E. coli*. Number of *E. coli* colonies obtained are listed. | 1 | 3 | 3 | 1 | 0 | 3 | 3 | 0 |
| Number of correct colonies as confirmed by MPX PCR and diagnostic digest over the number of colonies tested. | *1/1 | 0/3 | 0/3 | 0/1 | 0 | *3/3 | 0/3 | 0 |
| After an additional 80 days (total 120) of propagation DNA was extracted again from the same *P. tricornutum* cultures and transformed into *E. coli*. Number of *E. coli* colonies obtained are listed. | 700 | not tested | 33 | 0 | 0 | 44 | not tested | 0 |
| Number of correct colonies as confirmed by MPX PCR and diagnostic digest over the number of colonies tested. | *3/3 | | 0/3 | 0 | 0 | *3/3 | | 0 |

**Table B.4: Vanillin biosynthesis construct (p0521s-V) sequence analysis.** Constructs were fully sequenced after initial assembly (V1.0 and V3.0) and after 40 days (V1.1 and V3.1) and 120 days (V1.4 and V3.4) of serial passage in *P. tricornutum.*

| Ref Pos | Ref base | V10 | V11 | V14 | V30 | V31 | V34 | Type | Description | Feature affected | Effect |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6165 [+1] | - | 172 bp | - | - | - | - | - | Indel | V10 has a 172 bp insertion after position 6165. The area consists of repeats of "TCTGGGCCCACTGTTCCACTT GTATCGTCGGTCTGATAATCAG" and the 172 bp insertion corresponds to 4 additional repeats of this sequence. Therefore, it is likely a sequencing contig assembly error. Sequence coverage for V10 in this area is low but high in the adjacent regions. Sequence coverage in this area for V11 and V14 is high. | Backbone | Unknown |
| 9713 | G | G | - | - | G | G | G | Indel | V11 and V14 have a 4 bp deletion of positions 9713-9716. Sequence coverage at these positions is high for all assemblies. | *FcpF* Promoter | Unknown |
| 9714 | A | A | - | - | A | A | A | Indel | | | |
| 9715 | C | C | - | - | C | C | C | Indel | | | |
| 9716 | A | A | - | - | A | A | A | Indel | | | |
| 11929 | G | G | G | G | A | A | A | SNP | V30, V31, and V34 have a G to A transition at position 11929. This results in a silent mutation at R127 in the phenylalanine ammonia lyase (PAL) enzyme in the metabolic pathway. Sequence coverage at this position is high for all assemblies. | phenylalanine ammonia lyase (PAL) | Silent mutation (R127) |
| 15326 | A | - | A | - | A | A | A | Indel | V10 and V14 have a 2 bp deletion of positions 15326-15327. This causes a frameshift in the coding region of the trans cinnamate 4-hydroxylase (C4H) enzyme (505 a.a) of the metabolic pathway resulting in a normal protein from residues 1-503, followed by 14 new residues and a stop codon. Sequencing coverage for V10 and V14 is very high in this area. However, sequence coverage for V11 is 0 at these positions and very poor in the surrounding area. Therefore, it is likely that V11 also contains the deletion. | *trans*-cinnamate 4-hydroxylase (C4H) | Frameshift |
| 15327 | A | - | A | - | A | A | A | Indel | | | |
| 15372 | A | A | A | A | C | C | C | SNP | V30, V31, and V34 have an A to C transversion at position 15372. Sequence coverage at this position is high for all assemblies. | *H4-1B* Terminator | Unknown |
| 15377 | T | T | T | T | C | C | C | SNP | V30, V31, and V34 have a T to C transition at position 15377. Sequence coverage at this position is high for all assemblies. | *H4-1B* Terminator | Unknown |
| 15464 | T | T | T | T | C | C | C | SNP | V30, V31, and V34 have a T to C transition at position 15464. Sequence coverage at this position is high for all assemblies. | *H4-1B* Terminator | Unknown |
| 15514 [+1] | - | - | - | - | G | G | G | Indel | V30, V31, and V34 have a 1 bp G insertion after position 15514. Sequence coverage at this position is high for all assemblies. | *H4-1B* Terminator | Unknown |
| 15518 | A | A | A | A | - | - | - | Indel | V30, V31, and V34 have a 1 bp deletion of position 15518. Sequence coverage at this position is high for all assemblies. | *H4-1B* Terminator | Unknown |
| 15519 | A | A | A | A | G | G | G | SNP | V30, V31, and V34 have an A to G transition at position 15519. Sequence coverage at this position is high for all assemblies. | *H4-1B* Terminator | Unknown |
| 15535 | T | T | T | T | G | G | G | SNP | V30, V31, and V34 have a T to G transversion at position 15535. Sequence coverage at this position is high for all assemblies. | *H4-1B* Terminator | Unknown |
| 15540 | G | G | G | G | T | T | T | SNP | V30, V31, and V34 have a G to T transversion at position 15540. Sequence coverage at this position is high for all assemblies. | *H4-1B* Terminator | Unknown |
| 15681 | G | G | G | G | A | A | A | SNP | V30, V31, and V34 have a G to A transition at position 15681. Sequence coverage at this position is high for all assemblies. | *H4-1B* Terminator | Unknown |
| 15880 | A | A | A | A | C | C | C | SNP | V30, V31, and V34 have an A to C transversion at position 15880. Sequence coverage at this position is high for all assemblies. | *H4-1B* Terminator | Unknown |
| 20222 [+1] | - | - | - | - | A | A | A | Indel | V30, V31, and V34 have a 1 bp A insertion after position 20222. Sequence coverage at this position is high for all assemblies. | *40SRPS8* Terminator | Unknown |
| 22507 | C | C | - | - | C | C | C | Indel | V11 and V14 have a 2 bp deletion of positions 22507-22508. Sequence coverage at these positions is high for all assemblies. | URA3 region (downstream of URA3 ORF) | Unknown |
| 22508 | C | C | - | - | C | C | C | Indel | | | |
| 25525 [+1] | - | - | - | - | - | G | G | Indel | V31 and V34 have a 1 bp G insertion after position 25525. This causes a frameshift in position 812 of the coding region of the 4-hydroxycinnamoyl-CoA ligase (4CL) enzyme (566 a.a) in the metabolic pathway resulting in a normal protein from residues 1-271, followed by 14 residue substitutions and a premature stop codon which removes the last 281 residues. Sequence coverage is very high in this region and there is no ambiguity. | 4-hydroxycinnamoyl CoA ligase (4CL) | Frameshift |
| 26903 [+1] | - | - | - | - | T | T | T | Indel | V30, V31, and V34 have a 1 bp T insertion after position 26903. Sequence coverage at this position is high for all assemblies. | *EF1α* Terminator | Unknown |
| 27500 | G | A | A | A | G | G | G | SNP | V10, V11, and V14 have an G to A transition at position 27500. Sequence coverage at these positions is high for all assemblies. | *FcpD* Promoter | Unknown |
| 29768 | G | G | G | G | G | A | A | SNP | V31 and V34 have a G to A transition at position 29768. This results in a A355T mutation in the vanillin synthase (VAN) enzyme (356 a.a) of the metabolic pathway. Sequence coverage at this position is high for all assemblies. | vanillin synthase (VAN) | A355T |

| Position | | | | | | | | Type | Description | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 29805 [+1] | - | - | - | - | - | T | T | Indel | V31 and V34 have a 1 bp T insertion after position 29805. Sequence coverage at this position is high for all assemblies. | *FcpD* Terminator | Unknown |
| 30051 [+1] | - | - | - | - | - | C | C | Indel | V31 and V34 have a 1 bp C insertion after position 30051. Sequence coverage at this position is high for all assemblies. | *FcpD* Terminator | Unknown |
| 30052 | G | G | G | G | G | C | C | SNP | V31 and V34 have a G to C transversion at position 30052. Sequence coverage at this position is high for all assemblies. | *FcpD* Terminator | Unknown |
| 30055 | C | C | C | C | C | - | - | Indel | V31 and V34 have a 1 bp deletion at position 30055. Sequence coverage at this position is high for all assemblies. | *FcpD* Terminator | Unknown |
| 30056 | A | A | A | A | A | T | T | SNP | V31 and V34 have an A to T transversion at position 30056. Sequence coverage at this position is high for all assemblies. | *FcpD* Terminator | Unknown |
| 33150 [+1] | - | - | A | A | - | - | - | Indel | V11 and V14 have a 1 bp A insertion after position 33150. Sequence coverage at this position is high for all assemblies. | Backbone | Unknown |
| 33175 [+1] | - | - | - | - | - | A | A | Indel | V31 and V34 have a 1 bp A insertion after position 33175. Sequence coverage at this position is high for all assemblies. | Backbone | Unknown |

## Table B.5: List of oligonucleotides used in this study.

| Name | Sequence (5′ to 3′) | Description |
|---|---|---|
| DE2675 | AAAATTAACCCTCACTAAAGGGAACAAAAGCTGGTACCTAACGAAACGAATAGAAGCTCCC | Forward *EF1α* Promoter with pPtPuc3 homology |
| DE2676 | GTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGATTACGTGATATTATACGATACAA | Reverse *EF1α* Promoter with *Sh ble* homology |
| DE2677 | GAAATTAACCCTCACTAAAGGGAACAAAAGCTGGTACCTAACCCTGCGATAGACCTTTC | Forward *40SRPS8* Promoter with PtPuc3 homology |
| DE2678 | GGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGGTATTCTATTCTCTGATTCCTTC | Reverse *40SRPS8* Promoter with *Sh ble* homology |
| DE2681 | CGAAATTAACCCTCACTAAAGGGAACAAAAGCTGGTACCTAACGTCGGTCTCTTTCCCGG | Forward *H4-1B* Promoter with PtPuc3 homology |
| DE2682 | CGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGTTGGCTGTTGTTTGTTTTCG | Reverse *H4-1B* Promoter with *Sh ble* homology |
| DE2683 | TAACCCTCACTAAAGGGAACAAAAGCTGGTACCTAATTCGTTGATATTTTTATTCAAATG | Forward *γ-Tubulin* Promoter with PtPuc3 homology |
| DE2684 | GCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATATTGCAAGTAGAAGCGCAGC | Reverse *γ-Tubulin* Promoter with *Sh ble* homology |
| DE2687 | ATTAACCCTCACTAAAGGGAACAAAAGCTGGTACCTAAACACGGAGGATCTATCTACAGC | Forward *RBCMT* Promoter with PtPuc3 homology |
| DE2688 | GTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGGTAGGATCACGAAGAGATACTTTG | Reverse *RBCMT* Promoter with *Sh ble* homology |
| DE2689 | ATTAACCCTCACTAAAGGGAACAAAAGCTGGTACCTAAGAAACATACCTTCAGCGTCGTC | Forward *FcpB* Promoter with PtPuc3 homology |
| DE2690 | CGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATCTTGACTTCTGGCAACCGC | Reverse *FcpB* Promoter with *Sh ble* homology |
| DE2691 | AATTAACCCTCACTAAAGGGAACAAAAGCTGGTACCTAAACGTCGGTCTCTTTCCCGG | Forward *FcpC* Promoter with PtPuc3 homology |
| DE2692 | GTGAGCACCGGAACGGCACTGGTCAACTTGGCCATTTTGTCGAAATGTATATATGTGTTG | Reverse *FcpC* Promoter with *Sh ble* homology (for *FcpD* also) |
| DE2693 | TTAACCCTCACTAAAGGGAACAAAAGCTGGTACCTAAACTAGCTTGATTGGGGATATCTCG | Forward *FcpD* Promoter with PtPuc3 homology |
| D9_F | CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGAATGCGGGAGTGGACCGCGACG ATCCGTCCG | Forward *EF1α* Terminator with *Sh ble* homology |
| D9_R | TATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGAAGCAGTCGAAAATTCTTACCCAA ACTTAG | Reverse *EF1α* Terminator with pPtPuc3 homology |
| D18_F | CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGAAGATAAGAATATCTCATTGTGA ACATCTAA | Forward 40SRPS Terminator with *Sh ble* homology |
| D18_R | TATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGGGGGGTGGATAAAGAAGAAAGGGG TAGGGGT | Reverse 40RPS Terminator with pPtPuc3 homology |
| D14_Falt | CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGAGCTTGCGCTTGATCCTCGACTTT GTTGCTC | Forward *H4-1B* Terminator with *Sh ble* homology |
| D14_Ralt | TATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGAGCCGATTAGCAACAGAGAATGAA GCCAAC | Reverse *H4-1B* Terminator with pPtPuc3 homology |
| D13_F | CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGAAGCAAACTCATTATGATGCATG GGAGTGCG | Forward *γ-Tubulin* Terminator with *Sh ble* homology |
| D13_R | TATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGAATGGCGATGACTTACGAGGAGG GCGACCA | Reverse *γ-Tubulin* Terminator with pPtPuc3 homology |
| D16_F | CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGAAAATACAAATTCATGTACCTAA ACGATAGT | Forward *RBCMT* Terminator with *Sh ble* homology |
| D16_R | TATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGGGCACTACCGGTACGCCCAACACG ATCCAA | Reverse *RBCMT* Terminator with pPtPuc3 homology |
| D10_F | CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGATTTACTTGCTGGGTAGGCCGTTT CTGGAAT | Forward *FcpB* Terminator with *Sh ble* homology |
| D10_R | TATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGCACCTTTGGCAATTCGAATTTGTTT GATAG | Reverse *FcpB* Terminator with pPtPuc3 homology |
| D11_F | CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGATTTTGTTACATTTACTGACTTCA AGGAGTC | Forward *FcpC* Terminator with *Sh ble* homology |
| D11_R | TATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGATTGCAGCTTGTTGCAGAAGGATG TAGGCTC | Reverse *FcpC* Terminator with pPtPuc3 homology |
| D12_F | CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGATTTTGTTACATTTACTGACTTCA AGGAGTC | Forward *FcpD* Terminator with *Sh ble* homology |
| D12_R | TATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGGACGTTTCACTCTCGAGCACAGGT TTTTT | Reverse *FcpD* Terminator with pPtPuc3 homology |
| DE2580 | GACGACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGAGTAC GTACGCGTAACATATTG | Forward FBAC2-1 intron with *Sh ble* homology (SexAI) |
| DE2581 | TACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGCTG ATTGTGGAATCAAAGAA | Reverse FBAC2-1 intron with *Sh ble* homology (SexAI) |
| DE2534 | ACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGAGTACGTTT TGTTGTGGTCTATTGAC | Forward EIF6-1 intron with *Sh ble* homology (SexAI) |
| DE2535 | CTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGCTTCAAGA TGACAAAAAGCGTCAGA | Reverse EIF6-1 intron with *Sh ble* homology (SexAI) |
| DE2536 | ACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGAGTAAGAA TACTCATTCTTCGTCAAT | Forward RPS4-1 intron with *Sh ble* homology (SexAI) |
| DE2537 | CTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGCTATTAAA GGAGAAAAAGAATCAAC | Reverse RPS4-1 intron with *Sh ble* homology (SexAI) |
| DE2538 | ACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGAGTAAGACT CTGCAGGCTCCACGTGA | Forward TUFA-1 intron with *Sh ble* homology (SexAI) |
| DE2539 | CTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGCTAACAAG GATCAACAACAGGGAAG | Reverse TUFA-1 intron with *Sh ble* homology (SexAI) |
| DE2582 | AATCGAAAAATTAACCAAGTCGACGGTATCGATAATATTCTAGCTGAGGGTACCCATGGGTAC GTACGCGTAACATATTG | Forward FBAC2-1 intron with *Sh ble* homology (MscI) |
| DE2583 | ACTCGACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCT GATTGTGGAATCAAAGAA | Reverse FBAC2-1 intron with *Sh ble* homology (MscI) |

| Name | Sequence | Description |
|---|---|---|
| DE2540 | GAAAAATTAACCAAGTCGACGGTATCGATAATATTCTAGCTGAGGGTACCCATGGGTACGTTT TGTTGTGGTCTATTGAC | Forward EIF6-1 intron with *Sh ble* homology (MscI) |
| DE2541 | ACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCTTCAAG ATGACAAAAGCGTCAGA | Reverse EIF6-1 intron with *Sh ble* homology (MscI) |
| DE2542 | GAAAAATTAACCAAGTCGACGGTATCGATAATATTCTAGCTGAGGGTACCCATGGGTAAGAAT ACTCATTCTTCGTCAAT | Forward RPS4-1 intron with *Sh ble* homology (MscI) |
| DE2543 | ACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCTATTAA AGGAGAAAAAGAATCAAC | Reverse RPS4-1 intron with *Sh ble* homology (MscI) |
| DE2544 | GAAAAATTAACCAAGTCGACGGTATCGATAATATTCTAGCTGAGGGTACCCATGGGTAAGACT CTGCAGGCTCCACGTGA | Forward TUFA-1 intron with *Sh ble* homology (MscI) |
| DE2545 | ACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCTAACAA GGATCAACAACAGGGAAG | Reverse TUFA-1 intron with *Sh ble* homology (MscI) |
| D49_F | GTTTCACAGTCAGGAATAACACTAGCTCGTCTTCACGTAACTATAACGGTCCTAACGTCGGTCT CTTTCCCGGGAAACGG | Forward *H4-1B* Promoter with p0521s homology |
| D49-R | AACTAGTGGTGGTGGTGGTGGTGCACAGCCTAGGCACTCGTGGTGGTGGTGGTGGTGCATGT TGGCTGTTGTTTGTTTTC | Reverse *H4-1B* Promoter with PAL homology |
| D67_F | CTCTCGGCTACAACCGACAAAGTCTTTACAGTCAATACCGAAAACAAACAACAGCCAACATGC ACCACCACCACCACCA | Forward PAL-2A-C4H with *H4-1B* Promoter homology |
| D67_R | TGTATATAATCGTTCCAAGGTTTTAAAGACGAGCAACAAGTCGAGGATCAAGCGCAAGCTTA ACAGTTCCTTGGTTTCA | Reverse PAL-2A-C4H with *H4-1B* Terminator homology |
| D68_F | CAATTCAGCTTGCACATCCTTAACCACTCCATAATCGTTATGAAACCAAGGAACTGTTAAGCTTG CGCTTGATCCTCGAC | Forward *H4-1B* Terminator with PAL-2A-C4H homology |
| D68_R | ATACTTCTCACCCCTTTGTTTTCTTGGACTGCGTGAGTTTGGAAAAGGTCTATCGCAGGGAGCC GATTAGCAACAGAGAA | Reverse *H4-1B* Terminator with *40SRPS8* Promoter homology |
| D69_F | TCACTACATATAGATCCTTCACGGAGAAAAGTTGGCTTCATTCTCTGTTGCTAATCGGCTCCCTG CGATAGACCTTTTCC | Forward *40SRPS8* Promoter with *H4-1B* Terminator homology |
| D69_R | GTTCAGCGGGTCGCTCTCCACGATGGCGCCGTTGCCCGCGTGGTGGTGGTGGTGGTGCATGG TATTCTATTCTCTGATTC | Reverse *40SRPS8* Promoter with TAL-2A-C3H homology |
| D70_F | TCGTCGTCGCAATTGTCGTTGCGTTGATCTTGCACCGAAGGAATCAGAGAATAGAATACCATG CACCACCACCACCACCA | Forward TAL-2A-C3H with *40SRPS8* Promoter homology |
| D70_R | CTGTAAACAAAGAATAAAATTGGTAAATCATTAGATGTTCACAATGAGATATTCTTATCTTTAC ATATCGTAAGGCACGC | Reverse TAL-2A-C3H with *40SRPS8* Terminator homology |
| D71_F | GCGGTTGCAACGCCTCGGTTGCCTTCGGATCTGTACAAACGCGTGCCTTACGATATGTAAAGA TAAGAATATCTCATTGT | Forward *40SRPS8* Terminator with TAL-2A-C3H homology |
| D71_R | ATCCTTGAAGCTGTCCCTGATGGTCGTCATCTACCTGCCTGGACAGCATGGCCTGCAACGGGG GTGGATAAGAAGAAAG | Reverse *40SRPS8* Terminator with URA3 homology |
| D72_F | CCCTGAGCCATCCTCCCCACTCAGGAGATGACCCCTACCCCTTTCTTCTTTATCCACCCCCGTTGC AGGCCATGCTGTCC | Forward URA3 with *40SRPS8* Terminator homology |
| D72_R | CCGAAACCACCATGAACCTCCCAAACAACACCCGACCTCGGGGAGCTTCTATTCGTTTCGCGAT GAAACGAGAGAGGATG | Reverse URA3 with *EF1α* Promoter homology |
| D73_F | GTAACCGGGCATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGCGAA ACGAATAGAAGCTCCC | Forward *EF1α* Promoter with URA3 homology |
| D73_R | AGTGATTGGTGTCTCGGTGGCAGGCCGGACCATGGTGGAATCTCTGATGTTAATTTTCATGATT ACGTGATATTATACGA | Reverse *EF1α* Promoter with HCT-2A-4CL homology |
| D74_F | TAGTAGAGAAACTTCCATCGGATCGTATCATAATATTGTATCGTATAATATCACGTAATCATGA AAATTAACATCAGAGA | Forward HCT-2A-4CL with *EF1α* Promoter homology |
| D74_R | GCCCCTGTGATAGCACCTAGTATTGTTTTCCGGACGGATCGTCGCGGTCCACTCCCGCATTCAT AGTCTTGAAGAAACAG | Reverse HCT-2A-4CL with *EF1α* Terminator homology |
| D75_F | AAGATATTGAGAAAGGAACTAAAGAGAATTCTTACAAACTCTGTTTCTTCAAGACTATGAATGC GGGAGTGGACCGCGAC | Forward *EF1α* Terminator with HCT-2A-4CL homology |
| D75_R | AAGTACCCTAAAAGACATAGCACGCGACAAGCACGAGCGAGATATCCCAATCAAGCTAGTAA GCAGTCGAAAATTCTTAC | Reverse *EF1α* Terminator with *FcpD* Promoter homology |
| D76_F | AGTTTAAAGTAGGTGTATCTACAAACGGTCCTAAGTTTGGGTAAGAATTTTCGACTGCTTACTA GCTTGATTGGGATATC | Forward *FcpD* Promoter with *EF1α* Terminator homology |
| D76_R | CTCCTCGTCCACGTCCTTGGATCCATTTTGCTGCTGTTGGTGCTCCACCCATGTAGCCATTTTGT CGAAATGTATATATG | Reverse *FcpD* Promoter with COMT-2A-VAN homology |
| D77_F | AGTCCTCCCTCATAATCTCTGTTAGAGTTTACCAACAACACATATATACATTTCGACAAAATGGC TACATGGGTGGAGCA | Forward COMT-2A-VAN with *FcpD* Promoter homology |
| D77_R | TCCTGGAAACGACGGCAGTATCGATTCCTCGACTCCTTGAAGTCAGTAAATGTAACAAAACTA GTGGTGGTGGTGGTGGT | Reverse COMT-2A-VAN with *FcpD* Terminator homology |
| D50_F | ACATGCACCACCACCACCACCACGAGTGCctaGGCTGTGCACCACCACCACCACCACTAGTTTTG TTACATTTACTGACT | Forward *FcpD* Terminator with COMT-2A-VAN homology |
| D50-R | ATCGCTATAATGACCCCGAAGCAGGGTTATGCAGCGGAAGATCTATATTACCCTGTTATGACGT TTTCACTCTCGAGCAC | Reverse *FcpD* Terminator with p0521s homology |
| DE3214 | CAAGTTGACCAGTGCCGTTC | Forward *Sh ble* qPCR and MPX PCR screening primer |
| DE3215 | GGACACGACCTCCGACCAC | Reverse *Sh ble* qPCR and MPX PCR screening primer |
| DE3216 | ACTGGATGGGTCCTTCACC | Forward *nat* qPCR and MPX PCR screening primer |
| DE3217 | CAGGGCATGCTCATGTAGAG | Reverse *nat* qPCR and MPX PCR screening primer |
| DE2614 | AGGCAAAGCGTGGTGTTCTTA | Forward Actin16 qPCR and MPX screening primer |
| DE2615 | TCTGGGGAGCCTCAGTCAATA | Reverse Actin16 qPCR and MPX screening primer |
| DE3526 | tcgaGGTATCGTTAAGGCCGATGT | Top strand urease sgRNA#1 |
| DE3527 | aaacACATCGGCCTTAACGATACC | Bottom strand urease sgRNA#1 |
| DE3528 | tcgaTGGTATCGTTAAGGCCGATG | Top strand urease sgRNA#2 |
| DE3529 | aaacCATCGGCCTTAACGATACCA | Bottom strand urease sgRNA#2 and reverse urease sgRNA#3 target site primer |
| DE3530 | tcgaATGACGTACGCACGCCCCTT | Top strand urease sgRNA#3 and forward urease sgRNA#1 and sgRNA#2 target site primer |
| DE3531 | aaacAAGGGGCGTGCGTACGTCAT | Bottom strand urease sgRNA#3 |
| DE3532 | tcgaCCGGAAATGATCGTCAAGGG | Top strand urease sgRNA#4 |
| DE3533 | aaacCCCTTGACGATCATTTCCGG | Bottom strand urease sgRNA#4 |
| DE3713 | CAGTGCCCCCTCCAAACAT | Reverse urease sgRNA#1 and sgRNA#2 target site primer |
| DE3712 | CAAGGCGTTGTCCTTCGAC | Forward urease sgRNA#3 target site primer |
| DE3714 | TAGTCAAGCCATGGGCCG | Forward urease sgRNA#4 target site primer |
| DE3715 | CGTTCCAGACCATGTCTTTCTT | Reverse urease sgRNA#4 target site primer |
| MPX 1F | TCGTAGCGCTGTGAAGGTTA | Forward Vanillin pathway MPX PCR screening primer 1 |
| MPX 1R | GAATCGCTGTTGGTTGCAG | Reverse Vanillin pathway MPX PCR screening primer 1 |
| MPX 2F | GCAGTTGCTGCCAATCACTA | Forward Vanillin pathway MPX PCR screening primer 2 |
| MPX 2R | GTGCCCCTAGCTTCTTCACA | Reverse Vanillin pathway MPX PCR screening primer 2 |
| MPX 3F | AGGGACTAAACCGGAGGAGA | Forward Vanillin pathway MPX PCR screening primer 3 |
| MPX 3R | TCAATCTCGTGCTCATATTCAA | Reverse Vanillin pathway MPX PCR screening primer 3 |
| MPX 4F | GCTCTGGACCGATGAGTTG | Forward Vanillin pathway MPX PCR screening primer 4 |
| MPX 4R | CACTAGGGGTATGGAATCTGG | Reverse Vanillin pathway MPX PCR screening primer 4 |

MPX 5F    GATGCGAGAGCAAATGTCAA                          Forward Vanillin pathway MPX PCR screening primer 5
MPX 5R    GCGAATCGTTTTGTGTCAGA                          Reverse Vanillin pathway MPX PCR screening primer 5
MPX 6F    TCGGTTGACATCTGATTGGA                          Forward Vanillin pathway MPX PCR screening primer 6
MPX 6R    CCACCATTTGTCACTGCATC                          Reverse Vanillin pathway MPX PCR screening primer 6

**Table B.6: List of plasmids used in this study.**

| Plasmid | Description | Reference or source |
| --- | --- | --- |
| p0521s | *P. tricornutum* expression vector, ZeoR | Karas, et al., 2015 |
| pPtPuc3 | *P. tricornutum* expression vector, ZeoR | Karas, et al., 2015 |
| pPtGE1 | p0521s with FBAC2 intron 1 in *Sh ble* MscI site | This study |
| pPtGE2 | p0521s with TUFA intron 1 in *Sh ble* MscI site | This study |
| pPtGE3 | p0521s with EIF6 intron 1 in *Sh ble* MscI site | This study |
| pPtGE4 | p0521s with RPS4 intron 1 in *Sh ble* MscI site | This study |
| pPtGE5 | p0521s with FBAC2 intron 1 in *Sh ble* SexAI site | This study |
| pPtGE6 | p0521s with TUFA intron 1 in *Sh ble* SexAI site | This study |
| pPtGE7 | p0521s with EIF6 intron 1 in *Sh ble* SexAI site | This study |
| pPtGE8 | p0521s with RPS4 intron 1 in *Sh ble* SexAI site | This study |
| pPtGE9 | p0521s with mutated RPS4 intron 1 in *Sh ble* MscI site | This study |
| pPtGE10 | pPtPuc3 with *EF1α* Promoter driving *Sh ble* | This study |
| pPtGE11 | pPtPuc3 with *40SRPS8* Promoter driving *Sh ble* | This study |
| pPtGE12 | pPtPuc3 with *H4-1B* Promoter driving *Sh ble* | This study |
| pPtGE13 | pPtPuc3 with γ-*Tubulin* Promoter driving *Sh ble* | This study |
| pPtGE14 | pPtPuc3 with *RBCMT* Promoter driving *Sh ble* | This study |
| pPtGE15 | pPtPuc3 with *FcpB* Promoter driving *Sh ble* | This study |
| pPtGE16 | pPtPuc3 with *FcpC* Promoter driving *Sh ble* | This study |
| pPtGE17 | pPtPuc3 with *FcpD* Promoter driving *Sh ble* | This study |
| pPtGE18 | pPtPuc3 with *EF1α* Promoter and Terminator driving *Sh ble* | This study |
| pPtGE19 | pPtPuc3 with *40SRPS8* Promoter and Terminator driving *Sh ble* | This study |
| pPtGE20 | pPtPuc3 with *H4-1B* Promoter and Terminator driving *Sh ble* | This study |
| pPtGE21 | pPtPuc3 with γ-*Tubulin* Promoter and Terminator driving *Sh ble* | This study |
| pPtGE22 | pPtPuc3 with *RBCMT* Promoter and Terminator driving *Sh ble* | This study |
| pPtGE23 | pPtPuc3 with *FcpB* Promoter and Terminator driving *Sh ble* | This study |
| pPtGE24 | pPtPuc3 with *FcpC* Promoter and Terminator driving *Sh ble* | This study |
| pPtGE25 | pPtPuc3 with *FcpD* Promoter and Terminator driving *Sh ble* | This study |
| pPtGE26 | pPtPuc3 with *H4-1B* Promoter and *FcpA* Terminator driving *nat* | This study |
| pPtGE27 | pPtPuc3 with *FcpD* Promoter and *FcpA* Terminator driving *nat* | This study |
| pPtGE28 | pPtPuc3 with *H4-1B* Promoter and *FcpA* Terminator driving cat | This study |
| pPtGE29 | pPtPuc3 with *FcpD* Promoter and *FcpA* Terminator driving cat | This study |
| pPtGE30 | *P. tricornutum* expression vector, *FcpD* promoter and *FcpA* terminator driving *Sh ble* | This study |
| pPtGE31 | *P. tricornutum* expression vector, *FcpD* promoter and *FcpA* terminator driving *nat* | This study |
| pPtGE32 | *P. tricornutum* expression vector, *FcpD* promoter and *FcpA* terminator driving *Sh ble* and *nat* linked by a T2A self cleaving peptide linker | This study |
| pPtGE33 | *P. tricornutum* expression vector, *FcpD* promoter and *FcpA* terminator driving GFP and mCherry linked by a T2A self cleaving peptide linker | This study |
| PPtGE34 | *P. tricornutum* expression vector, *40SRPS8* promoter and terminator driving *Sh ble*, *FcpB* promoter and *FcpA* terminator driving Cas9 | This study |
| pPtGE35 | *P. tricornutum* expression vector, *40SRPS8* promoter and terminator driving *Sh ble*, *FcpB* promoter and *FcpA* terminator driving TevCas9 | This study |
| p0521s-V | *P. tricornutum* vanillin biosynthesis pathway plasmid | This study |

**Table B.7: List of genetic parts.**

| Name | Sequence | Description |
|------|----------|-------------|
| FBAC2-1 Intron | GTACGTACGCGTAACATATTGTAGCCAATTTGGTGTCGACGGCATGGTCTCGCAGG GAACGATAGAAAAACGTTGACACCTAGAAACGGGGGCTCTGGCACGGCAGCTCTC CACGGATTCTCTCGCAGTATTACACGGGCTATGCAGTGGACAGGGATACCAAACG TATGTTGGTGTCTTAATGTAAACTTTGCCCGTAAATTCCGTCCATATCGATCGAATC CTTACCGTCAAGGGGAGACCTCCAGTTCCCATGGTCGAGGGGCTTTCGTGGACCATC CCGCCGCAAGATCCATCCGCTGGTGTGACGTCGAGCAGCGCCGAACGGTGTCAGT GACGTGGCATCCCTCCCCCATCCACAGCAAACACGAGTAGTTTTGGTCGCGTTTAT ACCGCGCTCCAAACCCCAGAATTGGCCGTCGCGGTTTTCCGTACCGTTGGTCTCAC ACTGGTCCCGCGTTTTTTCTTTGATTCCACAATCAG | Fructose-1,6-bisphosphate aldolase intron 1 [20:568575-569056] |
| TUFA-1 Intron | GTAAGACTCTGCAGGCTCCACGTGAACGAGTACCTCGAACGGTATGGTACCGTCA CAATACACGGTTTCTGCCCTTGTTAGCTCACACGTTTGCTGTCCTTTCTACTCGTTC TTCCCTGTTGTTGATCCTTGTTAG | Elongation Factor Tu intron 1 [21:403882-404017] |
| EIF6-1 Intron | GTACGTTTTGTTGTGGTCTATTGACAACTGTAGAGTGCGTGAAGCATTTACATCTT GAAATGGTACATCTGACGCTTTTTGTCATCTTGAAG | Eukaryotic Translation Initiation Factor 6 intron 1 [1:704734-704825] |
| RPS4-1 Intron | GTAAGAATACTCATTCTTCGTCAATGAGATTGTTGAGTCTCTGATAGGAACCGAAA ATGTAGGAAGGAAGCGCTGGCAACTTTCTGATGAAAGATGTTTCTGATGAAAGAC GTTTGCCGTTGACAAACATCCGTCCCACGAAAGTAGTGTCGGGAAACGTTGGCTC ACTCGGTTGATTCTTTTTCTCCTTTAATAG | 40S Ribosomal Protein S4 intron 1 [1:607235-607430] |
| pEF1α | CGAAACGAATAGAAGCTCCCCGAGGTCGGGTGTTGTTTGGGAGGTTCATGGTGGT TTCGGTGTCGCTTGCTCGCTCGCTCGTTCGCAGTGACAGACAGTTCGTGAGACACG AGAACCGTTGGCGTCCGAGTTCGGGTGCCGCATTTCGTCGTCTCCACGATTCAATT CTTGCCCATCAGACGAGTCCCGAATTCCGTGACTCTGGATGCGATTTACTTTCTAA CTGTAAGCGAAACTCAACGATTCCGTACGTTGTTTTCTATTTTACAGTGAGTCTTCG ATACCACCGTACAACCATCGTTCGTGTACCGTCTGGTAGTCCCACGTGTCGACAAC GTGTGGCTCTGGACCGATGAGTTGTTTGCCGTTCGGAAACGAGCAGTACCAAGGA ATTCACAGAAACACAGCCCATGTAACACAACGACCGCGAATCGTTTCGGTGCTCT CGCTTCGCGTACGGGCGGGCGGTCCTCCCGAGCAGCGAGAGGAGTCCGCAGCGTC ATAGTTGCAATCCGGGCCCCCCTCGCGTTGTTCACTCTCTCGTCTAGTAGAGAAAC TTCCATCGGATCGTATCATAATATTGTATCGTATAATATCACGTAATC | Elongation Factor-1 alpha Promoter |
| p40SRPS8 | CCCTGCGATAGACCTTTTCCAAACTCACGCAGTCCAAGAAAACAAAGGGGTGAGA AGTATACGCACCTTTCGGTTTCGGCATAATTCTTAAACTCTTGTGGTCACTTTCTTG TGAAGAAGCTAGGGGCACTCGTTTTCCCTCAGAGCCTGCAAACACAAAATTCCTG CAGTCAATTGTCCCAACACTCGGCAAACCGTATGCGCAAGCAACGATGCGCAGAA GGCCGTGGATGGATGGCGACTCGCGATATGGCTTCTTGGGTCGCCAGTGTGGTAC GTCCGGCGTATGTCAATACGCGAATTCGGACGACTGGCATCTCTAGGAGGAGGAT TCCTTCTTTTTATGACATGTTTATTTTATATACATTGATGCTTTCCGACAGTCGGAAG TAATAAATGAATTTATTTCAAGACTACCTATACTCCTTTGACTTGTTCGACTAATCT TACCGCTTACTAAAATCTCGAAATCACGCTTGACCTCTCGCACGCAAATTTTTGCT GCTGGACGCTACGCACTCGGCCCAATTCTTCTCGGTCCTCGTCGTCGCAATTGTCG TTGCGTTGATCTTGCACCGAAGGAATCAGAGAATAGAATACC | 40S ribosomal protein S8 Terminator |
| pH4-1B | CGTCGGTCTCTTTCCCGGGAAACGGGTACACTCCTCCGCGCCAACAACATATTACT ACTACTACCAAGAACGTCCACGGCCTTGTCGTGCGTTACGCTCTCCCAACGCGTGC GGGGTAAATTACGTCTCGGTTTGCTAAGTAGCGCACAGCTAAATAGATGACCGTT ATTGTATTTAAGATCATTCAATATTGATTGTACTTTGCGTCAAACTGAAATT CCCTCGTACTAACGGTTAACCCGTCAACCCTAAGCGTTCGCCCAAAGTAGTCAACC GGGACACGCGAACCGACATTGGGCAGATCTTTCACAGACAGAAAACCATTTCCAA TCCAAATAAGCATGACTATTACACACCCATTCGTAGCGCGAGGACAAACTGATAG CTCCAACAAAATGCGCCAACATCGTACATTGTAAGAAGCTTACGGAACACTATGT ATGTAGAACCATACGAACAGCAACTAGTACTGGCCATCGAGCAGCGGTGACTCCC GGCTTTCGTAGCGCTGTGAAGGTTACACTCTCACAATTCGCTCTCGGCTACAACCG ACAAAGTCTTACTCACAGTCAATACCGAAAACAAACAACAGCCAAC | Histone H4 Promoter |
| pγ-Tubulin | TTCGTTGATATTTTTATTCAAATGTATCGGGAGGAGTAGAGGTTGATTAACTGTAA ACAATTTCCTATTTACTGTTAAGGACCAGCTGCTGCAGTAGGTATGGCCTATCCAC TAAACGCACTCACGGAACGCCTCGCGAAATTTACCCACGGCCAACTTACATTACC GCCTTTTGTGAATTGGAAACGCCGCATGATTCTCAAATGCGCAGAATTTCAAACGG TAGCTTGCGGTGGAGACTCGCTCATTGACAGTGAAACTACCTTGTGTCCTCGGATT TTCAGATATACCTATACAGTTCATGGCAAAATTTCGTTCATGAACGCACGTGATCC ATTGCTCGCGATTCCCGTTTTTGATTGTGAACGCGGGATTACATGCGTGCGGTGAC GGTAGTCCAGACACAGATATTTGCAATACCGGGCCCTTTTCACTACAGACCCTGTA GGGGTATGTTGACGAGAATGAACTCGCAGACTGCCAAAATCGCTTTGGCTGATCC CAAGTTTTGGCACTCCATCGTAATTTGTCATATTCCATACGGTAGCTTCGACTGAA TCCAGACAAACAATTTAGTCCAGCTGCGCTTCTACTTGCAAT | Tubulin gamma chain Promoter |

| | | |
|---|---|---|
| pRBCMT | ACACGGAGGATCTATCTACAGCAGCGATGAGGGCGCCCGAGAAAGAAAGAACGA TTGCCGTACTATTCTCTTTGACCTTTGGGCGCTCGCTCGTCTGTATCTTTGAAGCGACTGT TGGGGTCTCAGGGTCCAAAAAACAGAAACTGGATTGACAGTGTGTCTGGACCTTG TCGAACCTTACAGTTACATTACAGTTAATTGTCACTGTAAATAGTCTATCGCTGGA TTACGTCATCGCGTGACTGGGTGGGAATCCTTCTTGTTGACAGTGAATCTACGGTA TACTATTCCTTGGGCGCTTGTACTTGTGTCTCGAGATTGCCGACAGTGACGTCAAT TCGGCACCCACACCTTCCACCCGCCGAACCAAAATCAACAACACGAAGCACACGA CCGACCGACTGTACACGTGAAGGAGCAAACCATCGAACGAAAGGAGCCTTCCACG GACACAACCCGAAAGCTCGACACCCTTCACCCACGCAAAGTATCTCTTCGTGATCC TACC | Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit N-methyltransferase I Promoter |
| pFcpB | GAAACATACCTTCAGCGTCGTCTTCACTGTCACAGTCAACTGACAGTAATCTTTGG CCCGTAGAGGTTCGAAATTCAATCTATTAAATACAGCAGGATAAGACACAAGAGC GACATCCTGACATCAACTCCGTGAACAGCAAATCCTGGTTGAACACGTATCCTTTT GGGGGCCTCCAGCTACGACGCTCGCCCCAGCTGGGGCTTCCTTACTATACACAGCG CATATTTCGCGGTTGCCAGAAGTCAAG | Fucoxanthin-chlorophyll *a/c* binding protein B Promoter |
| pFcpC | GAGCACAAGAGGTGACAAAAGCCACCGGCTGGATCGCACTTCTCGGAATTTCCCC CCTACTATCAAACAAATTCGAATTGCCAAAGGTGAAGgGACTAACTGTAAATCCTg aTCAATCAAGGTCTCAATCAAGTACAATGGGCTACAATGATATTTAGATGGGAACA CAATGAAACaAATTGAAACTTCTACTGACAGGAGCGCAATTGACTTGTGTAGCTTT TCATGAGCACTTGATTGCTACCaATTGTGAACGGGATGGGGAAAGACTCGAAAAG GTGCATGCTTCCGATAATCTACTATATTTTCTAGAATCAAATAATATTTAAATGAA TGAGGTCCTCAGCGTACGTTAAGCCTACTTATTTAGAACGAGAAGTCAGACCGAG GGGTACTAAAATTCTAAGGGTTGAGAGGTATCTTGATTCCGGGTCTATGGAAGCCC ATCCTTGTTGAAGCTTGAACACGATCCTTGTGAAAGGCCGACGTTGCGCGAAAAA ACAGCCTGCCGATTTCTTTCCTTCTTTCTCGTCTCAACCTATATACTTTCATAATCTC TGTTAgAGTTTACCAACAACACATATATACATTTCGACAAA | Fucoxanthin-chlorophyll *a/c* binding protein C Promoter |
| pFcpD | ACTAGCTTGATTGGGATATCTCGCTCATGTTTGTCGCGTGCTATGTCTTTTTAGGTA CTTTGAACCTACGTTCGTACTTGTATAAATGATCATCGTATTATCGTTTTTCATCC GTCCAGCGCAAAATGCATTAGCAGCTAGTCCTAGCGTGCGGAGCTACCTGgACAG GTGCATGACGGATGCGTGTCCTTCAGTGAcTTTCTAATTAACAGTAACTTCTTTACT TATGTTTCAGTTTGTAAGAAGCGGGATTCGCTCGTCGGTTGACATCTGATTGGACT GCGTCGGCACaTGAAAACTACATTGTGAAATCTGCTAAAACTCCGGGTATCTCTGA CACAAAACGATTCGgCtTCgCAATTTCAACATTACGGTCAAGGCTAACGTATCTTTC TCGGTCAACTTCAGATTAtGCCGAtTAAATTGTCGTAGCTTTCAAGGCGTTTTGAGT ACTGCGGCAGTTGTTGAACCTGCAAGGAGAAGATCTCGACAACAGAATAaAGCGA AAAATGGGTCTCATGCACTAACACTCAGgCCTCCCTCATAATCTCTGTTtGAGTTTA CCAACAACACATATATACATTTCGACAAA | Fucoxanthin-chlorophyll *a/c* binding protein D Promoter |
| tEF1α | ATGCGGGAGTGGACCGCGACGATCCGtCCgGaAAAcAATACTAGGTGCTATCACAG GGGCGCGTTTTGGAGAGACGTTCTGCGGAAACACGAATTTAGAATACGTAACTAA CATATAAACTGGATAGCCCTCGCATCGGAACTTAGAATGTTCGCCTCAATTTTTAG TTTAGCGTGGAGCAGAGATACCTTTCCATTTGGCAAAATCTACCTTTCGTGAGGGA CATCTTGAGAAATAAGCGGACTTGTAGACTAGGACCGTGGTAACCTCCTCTCAATC TACCAATGTTGTCTGATTTCCGAGCCGCGCGGCTGAAAATCGTCTAGCACTTGGAT GCGAGAGCAAATGTCAAGTCCTGCTCTGTCCTGTTGGACGCTTTCCTCTCACCGCG AGAGGGCTTTCACTCGCGAAACACGTATTTCATATTCAAACTCTATGAAGTTTAAA GTAGATGTATCTACAAACGGTCCTAAGTTTGGGTAAGAATTTTCGACTGCAT | Elongation Factor-1 alpha Terminator |
| t40SRPS8 | AGATAAGAATATCTCATTGTGAACATCTATGATTTACCAATTTTATTCTTTGTTTAC AGTTAGACGCCAGTAATTGTGCTGTTTCTCTCAAGTCTGTGTCAATACAAACTACG AAACTTGGCAATTTTTCTCTTGAATATGAGCACGAGATTGAAACGCACAAAGGAA ATTAGTTTCCATCCTTTGACAAAGTTTGTTGCTGTTTAGAGAACAGATGTCAAAAT TAACGTGCCATGGAATTGAACCATGGTGCGCTATCCCCAAATCACGTCGTTTGACC TCGTCACATTTAAGGTATATCAAGCATATTCACTTATATCTTGACATCCTCCCTGCT TGATATTCCTTTGCCCTTGAGCCATCTTCCCCACACGCTGGAGATGACCCCTGCCC CTCCTTTCTTTATCCACCCGAGTAAGGTGTCGACAAGTCACTTTGCCCCTGAGCCA TCCTCCCCACTCAGGAGATGACCCCTACCCCTTTCTTCTTTATCCACCCC | 40S ribosomal protein S8 Terminator |
| tH4-1B | GCTTGCGCTTGATCCTCGACTTTGTTGCTCGTCTTTAAAACCTTGGAACGATTATAT ACAATCACCAACTCAAAAACCGGATTTTCTAAAATCCACCCAAACAACCAAAGAA AATACTTCTCATTGCATTTATGAATCACAGCAGACCTGCGTCCTTTTAAAACTTAG ATCCTGTTTTCTTATAAAAAACAGATCAAATTTTCTGGGAGTTCATTGACTCTGCC AGTCAGAATCAATCCTGCAGTAATTCTTTATTTACAGGTGAAAGTAAAAGAGAAT CCCAATTTTTTGCTTGTACATTAAGGTCCTCCTTACATTACAGCTAATTTCAAAATA AGATGAAGTTGGATTCGTGTCCTTTCATGGTGATGTGATATTCTGCACTATACCAA ACACCGTGAAATGTCAGCTAGAGCTTGTCATGAGGCAGTTGCTGCCAATCACTAC ATATAGATCCTTCACGGAGAAAAGTTGGCTTCATTCTCTGTTGCTAATCGGCT | Histone H4 Terminator |

| | | |
|---|---|---|
| tγ-Tubulin | AGCAAACTCATTATGATGCATGGGAGTGCGACCGAGTTTCGAACGATGCTAACGA ACATTATTAGTGGAAGGCAGTCATTGTTGTATGCGTCAAAGTATATAATCAGACGG ACAGTAGTTCATTTTAAACTTTTGTTCGGAAAGCGTTGATCATTCATCGGGGAATC GCGCTAACGAGCAGTAATTGGAGTTGTAACTGCAGGCTCAACGCGTTTCTGGCTCC CGTGGGTTACCATAATGCTAACTATCATTCTTTATTTGCAGTATCACCAGTCCAAG AATTATTCGTGGTCATTTCTGATGCTACTGTGGAGAAGTGAGAGTAATTTCGCCGA CTTTGAAGTGAAGACGCGTCTCCGAACGTAGATTGTTGGTATTGTACCATTGAAGG AGAGTTTTTGAACTCAGGTTGTTGATCGTAATGTCCGCGACATCCTGGCGCACTAC ACGCCAATGACCATAACATGTCTTGGTCGCCCTCCTCGTAAGTCATCGCCATT | Tubulin gamma chain Terminator |
| tRBCMT | AAATACAAATTCATGTACCTAAACGATAGTATGGATGATGGGAGTAATTGCACTA TAATTGTAGAACCTTGTAAGAGGAAAAAATGATCTTACTGTGTTATTTCCTCTTGA AAGAATCTATGGATAAAATAAGAGAGGACGCTAGGTGGTAACATTCCGGCAAAAC ACTGGCGCCTAAATTTTTGCCGGAATCGTCAATTGCAACGGTTGTACCGGTGCTTT GTTTTAGTGTTTCGCCTTGCTACCCTTCAGGAACCGGTAACGAACACCTCCGCCAC CCCGGACCCGCTCTGTTTTAGTTTGAATGGACTTGCGCAAAGAATCCTTATCATTC GCCTGTGACCGGGTTCCTTCCTTGGCCAAATTTGCCAGAAAGGTTTCCCGATCAAT GCTAGGGGTGCTTCCGCCTCCACTTGCCGTGGGAGTGTTCCCACCGCTACTTGCCG TTGGAGTGCTATCACTATCGGAGCCTTGGATCGTGTTGGGCGTACCGGTAGTGCC | Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit N-methyltransferase I Terminator |
| tFcpB | TTTACTTGCTGGGTAGGCCGTTTCTGGAATAACATATTAGATTCTAACTGGTTCGA AGCATTGCGTTGCTGTAACATTCCCGTTCACAAAAATACAGAACAGTCTAGAAGTT CGCGACGACATAATTTTTCTCTTTAGGAGGCCGGGGTTGTAATTGTTCTAGGGCTG TTCCAATAGAGAAGATAAGATGATCAAACATACCAGCCGCGCTTGATTGGACGGA GTACGTTTGCATCAGCTATTTTTCAAAAGCGCTGCACGACGCACACTCTATGAACA CTTCAAGACTCTCAACGCAAGTGACAACCATCCTCTCCAAAAGGCTATCTTTCGGG GCACCTGTAATATAAAAAAAGCATGGCAGTGCATTCCATGCAAAAAATGTCTAATC TGGTTGGGTTTTAAAGTCCGTATCGAGCACAGAGGTGACAAAAGCCACCGGCTGG ATCGCACTTCTCGGAATTTCCCCCCTACTATCAAACAAATTCGAATTGCCAAAGGT G | Fucoxanthin-chlorophyll *a/c* binding protein B Terminator |
| tFcpC | TTTTGTTACATTGACTTCAAGGAGTCGAGGAATCGATACTGCCGTCGTTTCCAGGA TCCGAGGTTTCTATAGACTCTCTATAGACTCTGTTAACCTAATAGAATCAGACATA CCTCTCCTGCTATTTTGTTTTTATGAATTTGGCTTTTGCCTCTCTAGTCAGATTTGAA TGTTATTTTCCGCCAGGTGTGTTAGTCGGGCTCTCGTTTGAGTTACAAGAGGGATT GAGTGGCGAGGATTCACTCTAATGTAAATATGACTGTGAACAAAACTTTAAAATT ACTACGCATCTTCTTTGACTGTCAGATATTCGTCGGTGACAGCAGTCAATGCCTGC AAATTGTCCTCCTGGGTCGCAATTTGGTTTTGGATTGACCTGGTATGCATTATGAA GAAAAAAATTCGTTATTAGCCAACTGCCTAGCGTGCACATTGCATGGTTAGACCTC CTTGACGACTGTGAGCCTACATCCTTCTGCAACAAGCTGCAAT | Fucoxanthin-chlorophyll *a/c* binding protein C Terminator |
| tFcpD | TTTTGTTACATTTACTGACTTCAAGGAGTCGAGGAATCGATACTGCCGTCGTTTCC AGGATCCGAGGTTTCATAAACTCTGTTAACGTTATAGAAACAGACTTACCTCTCCT ACGCCATTCACGTAATATTCGCAATATGCTATTCTTCCTCTGAAGACCAGGTTTAT GTGCTGCCTGAAACTATTTCAATAAGTCAGCTGCACTTGCACAGGGTTTCACAAGG AAAGCGTGTCTTTTTTTCCAACGTAGGCGTCGCTTTCGTCTGACTCTTACTCTTACA TTCACAGCCAATACTTACAATTAGTAAAAAACCTGTGCTCGAGAGTGAAAACGTC | Fucoxanthin-chlorophyll *a/c* binding protein D Terminator |

| PAL | ATGCACCACCACCACCACCACGAGTGCGAGAACGGCCGCGGCGTTGCTGCAACCA<br>ACAGCGATTCCCTGTGCATGGCGACGCCCCGCGCCGACCCGCTTAACTGGGGCAA<br>GGCGGCGGAGGAGCTGATGGGCAGCCACCTGGACGAGGTCAAGCGGATGGTGGC<br>CGAGTACCGCCAACCCTTGGTGAAGATCGAGGGCGCCAGCCTCAGCATCGCGCAG<br>GTGGCCGCCGTGGCCACCGGCGCTGGCGAGGCCCGGGTCGAGCTGGACGAGTCTG<br>CCCGTAGCCGGGTCAAGGCAAGCAGCGACTGGGTCATGACCAGCATGATGAACGG<br>CACCGACAGCTACGGCGTCACCACCGGCTTCGGCGCCACGTCGCACAGGAGGACC<br>AAGGAGGGCGGCGCGCTCCAGAGGGAGCTCATCAGGTTCCTCAACGCCGGCGCCT<br>TCGGCACCGGCGCGGACGGCCACGTCCTGCCGGCCGAGACGACGCGCGCGGCCAT<br>GCTCGTCCGCATCAACACCCTCCTCCAGGGGTACTCTGGCATCCGCTTCGAGATCC<br>TCGAGGCGATCGTCAAGCTGCTCAACGCCAACGTCACGCCGTGCCTGCCGCTCCGC<br>GGCACGGTAACCGCGTCCGGCGACCTCGTGCCGCTCTCCTACATTGCGGGCCTCGT<br>CACCGGGCGTGAGAACTCTGTTGCGGTTGCCCCAGACGGCAGCAAGGTCAACGCC<br>GCCGAGGCATTCAAGATTGCCGGAATCCAGGGCGGCTTCTTCGAGCTGCAGCCCA<br>AGGAGGGTCTTGCGATGGTGAACGGCACGGCCGTGGGCTCTGGCCTTGCCTCCAC<br>TGTGCTCTTCGAGGCTAATATCCTTGCCATCCTCGCCGAGGTCCTGTCTGCCGTGTT<br>CTGCGAGGTCATGAACGGCAAGCCGGAGTACACTGACCACCTGACCCACAAGCTC<br>AAGCACCACCCGGGACAGATCGAGGCTGCTGCCATCATGGAGCACATCTTGGAGG<br>GCAGCTCCTACATGAAGCTTGCCAAGAAGCTCGGCGAGCTCGATCCCTTGATGAA<br>GCCCAAGCAGGACCGGTACGCGCTCCGCACGTCGCCGCAGTGGCTTGGCCCCCAG<br>ATTGAGGTTATCCGCGCCGCCACCAAGTCCATTGAGCGCGAGATCAACTCCGTCA<br>ACGACAACCCGCTCATCGATGTCGCCCGAAGCAAGGCTCTTCACGGTGGCAACTT<br>CCAGGGGCACGCCCATCGGGGTGTCCATGGACAACACCCGTCTCGCCATCGCAGCC<br>ATCGGCAAGCTCATGTTTGCGCAGTTCTCTGAGCTCGTCAACGACTACTACAACAA<br>CGGCTTGCCCTCCAACCTGTCCGGTGGGCGCAACCCCAGCTTGGACTACGGCTTTA<br>AGGGTGCCGAGATCGCCATGGCGTCCTACTGCTCCGAGCTGCAGTTCCTGGGGAA<br>CCCCGGTCACCAACCACGTGCAGAGCGCGGAGCAGCACAACCAGGACGTCAACTCA<br>CTGGGACTCATCTCCTCCAGGAAGACTGCTGAGGCCATCGAGATCCTAAAGCTCAT<br>GTCCTCGACGTTCCTGATCGCCCTGTGCCAGGCCGTGGACCTGCGCCACATCGAGG<br>AGAACGTCAAGAGCGCTGTCAAGAGCTGCGTGATGACGGTGGCGAAGAAGACCCT<br>GAGCACCAACTCCACCGGTGGTCTCCACGTCGCCCGCTTCTGCGAGAAGGACCTG<br>CTCCAGGAGATCGAGCGCGAGGCGGTGTTCGCGTATGCCGACGACCCCTGCAGCG<br>CTAACTACCCGCTGATGAAGAAGCTTCGCAACGTGCTCGTGGAGCGCGCCCTCGC<br>CAACGGCGCTGCCGAGTTCAACGCGGAGACATCCGTGTTCGCCAAGGTCGCCCAG<br>TTCGAGGAGGACCTGCGCGCGGCGCTGCCAAAGGCGGTGGAGGCTGCACGGGCG<br>GCTGTCGAGAACGGCACGGCAGGGATACCGAACAGAATCGCCGAGTGCCGCTCCT<br>ACCCGCTCTACCGCTTCGTGCGCGAGGAGCTCGGAGCCGTGTACCTCACCGGCGA<br>GAAGACGCGCTCTCCCGGCGAGGAGCTGAACAAGGTGCTCGTTGCCATCAACCAG<br>GGCAAGCACATCGACCCGTTGCTCGAGTGCCTCAAGGAGTGGAACGGCGAGCCAC<br>TGCCCATCTGCTAA | Phenylalanine ammonia lyase from *Zea mays* |

| C4H | ATGGACCTCCTCTTGCTGGAGAAGTCTCTAATCGCCGTCTTCGTGGCGGTGATTCT<br>CGCCACGGTGATTTCAAAGCTCCGCGGCAAGAAATTGAAGCTACCTCCAGGTCCT<br>ATACCAATTCCGATCTTCGGAAACTGGCTTCAAGTAGGAGATGACCTCAACCACC<br>GTAATCTCGTCGATTACGCTAAGAAATTCGGCGATCTCTTCCTCCTCCGTATGGGT<br>CAGCGTAACCTAGTCGTCGTCTCTTCACCGGATCTAACCAAGGAAGTGCTCCACAC<br>ACAAGGCGTTGAGTTTGGATCTAGAACGAGAAACGTCGTGTTCGACATTTTCACCG<br>GGAAAGGTCAAGATATGGTGTTCACTGTTTACGGCGAGCATTGGAGGAAGATGAG<br>AAGAATCATGACGGTTCCTTTCTTCACCAACAAAGTTGTTCAACAGAATCGTGAAG<br>GTTGGGAGTTTGAAGCAGCTAGTGTTGTTGAAGATGTTAAGAAGAATCCAGATTCT<br>GCTACGAAAGGAATCGTGTTGAGGAAACGTTTGCAATTGATGATGTATAACAATA<br>TGTTCCGTATCATGTTCGATAGAAGATTTGAGAGTGAGGATAGTCCTCTTTTCCTT<br>AGGCTTAAGGCTTTGAATGGTGAGAGAAGTCGATTAGCTCAGAGCTTTGAGTATA<br>ACTATGGAGATTTCATTCCTATCCTTAGACCATTCCTCAGAGGCTATTTGAAGATTT<br>GTCAAGATGTGAAAGATCGAAGAATCGCTCTTTTCAAGAAGTACTTTGTTGATGAG<br>AGGAAGCAAATTGCGAGTTCTAAGCCTACAGGTAGTGAAGGATTGAAATGTGCCA<br>TTGATCACATCCTTGAAGCTGAGCAGAAGGGAGAAATCAACGAGGACAATGTTCT<br>TTACATCGTCGAGAACATCAATGTCGCCGCGATTGAGACAACATTGTGGTCTATCG<br>AGTGGGGAATTGCAGAGCTAGTGAACCATCCTGAAATCCAGAGTAAGCTAAGGAA<br>CGAACTCGACACGGTTCTTGGACCGGGTGTGCAAGTCACCGAGCCTGATCTTCACA<br>AACTTCCATACCTTCAAGCTGTGGTTAAGGAGACTCTTCGTCTGAGAATGGCGATT<br>CCTCTCCTCGTGCCTCACATGAACCTCCATGATGCGAAGCTCGCTGGCTACGATAT<br>CCCAGCAGAAAAGCAAAATCCTTGTTAATGCTTGGTGGCTAGCAAACAACCCCAAC<br>AGCTGGAAGAAGCCTGAAGAGTTTAGACCAGAGAGGTTCTTTGAAGAAGAATCGC<br>ACGTGGAAGCTAACGGAAATGACTTCAGGTATGTGCCGTTTGGTGTTGGACGTAG<br>AAGCTGTCCCGGGATTATATTGGCATTACCTATTTTGGGGATCACCATTGGTAGGA<br>TGGTCCAGAACTTCGAGCTTCTTCCTCCTCCAGGACAGTCTAAAGTGGATACTAGT<br>GAGAAAGGTGGACAATTCAGCTTGCACATCCTTAACCACTCCATAATCGTTATGAA<br>ACCAAGGAACTGTTAA | *Trans*-cinnamate 4-hydroxylase from *Arabidopsis thaliana* |
| TAL | ATGCACCACCACCACCACCACGCGGGCAACGGCGCCATCGTGGAGAGCGACCCGC<br>TGAACTGGGGCGCGGCGGCGGCGGAGCTGGCCGGGAGCCACCTGGACGAGGTGA<br>AGCGCATGGTGGCGCAGGCCCGGCAGCCCGTGGTCAAGATCGAGGGCTCCACCCT<br>CCGCGTCGGCCAGGTGGCCGCCGTCGCCTCCGCCAAGGACGCGTCCGGCGTCGCC<br>GTCGAGCTCGACGAGGAGGCCCGCCCCCGCCTGGCCGCAGCAGCGAGTGGATCC<br>TCGACTGCATCGCCCACGGCGGCGACATCTACGGCGTCACCACCGGCTTCGGCGG<br>CACCTCCCACCGCCGCACCAAGGACGGGCCCGCGCTCCAGGTCGAGCTGCTCAGG<br>CATCTCAACGCCGGAATCTTCGGCACCGGCAGCGACGGGCACACGCTGCCGTCGG<br>AGGTCACCCGCGCGGCGATGCTGGTGCGCATCAACACCCTCCTCCAGGGCTACTCC<br>GGCATCCGCTTCGAGATCCTCGAGGCCATCACGAAGCTGCTCAACACCGGTGTCA<br>GCCCCTGCCTGCCGCTCCGGGGCACCATCACCGCGTCGGGCGACCTGGTCCCGCTC<br>TCCTACATCGCCGGCCTCATCACGGGCCGCCCCAACGCGCAGGCCGTCACCGTCG<br>ACGGAAGGAAGGTGGACGCCGCCGAGGCGTTCAAGATCGCCGGCATCGAGGGCG<br>GCTTCTTCAAGCTCAACCCCAAGGAGGGCCTCGCCATCGTCAACGGCACGTCCGTG<br>GGCTCCGCGCTCGCGGCCACCGTGATGTACGACGCCAACGTCCTGGCCGTCCTGTC<br>GGAGGTCCTGTCCGCCGTCTTCTGCGAGGTCATGAACGGCAAGCCCGAGTACACG<br>GACCACCTGACCCACAAGCTGAAGCACCACCCGGGGTCCATCGAGGCCGCGGCCA<br>TCATGGAGCACATCCTGGATGGCAGCTCCTTCATGAAGCAGGCCAAGAAGGTGAA<br>CGAGCTGGACCCGCTGCTGAAGCCCAAGCAGGACAGGTACGCGCTCCGCACGTCG<br>CCGCAGTGGCTGGGCCCCCAGATCGAGGTCATCCGCGCCGCCACCAAGTCCATCG<br>AGCGCGAGGTCAACTCCGTGAACGACAACCCGGTCATCGACGTCCACCGCGGCAA<br>GGCGCTGCACGGCGGCAACTTCCAGGGCACCCCCATCGGCGTGTCCATGGACAAC<br>GCCCGCCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCCGAGCT<br>CGTCAACGAGTTCTACAACAACGGGCTCACCTCCAACCTGGCCGGCAGCCGCAAC<br>CCCAGCCTGGACTACGGCTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTC<br>CGAGCTCCAGTACCTGGGCAACCCCATCACCAACCACGTGCAGAGCGCGGACGAG<br>CACAACCAGGACGTGAACTCCCTGGGCCTCGTCTCGGCCAGGAAGACCGCCGAGG<br>CGATCGACATCCTGAAGCTCATGTCGTCCACCTACATCGTGGCGCTGTGCCAGGCC<br>GTGGACCTGCGCCACCTCGAGGAGAACATCAAGGCGTCGGTGAAGAACACCGTGA<br>CCCAGGTGGCCAAGAAGGTGCTGACCATGAACCCCTCGGGCGAGCTCTCCAGCGC<br>CCGCTTCAGCGAGAAGGAGCTGATCAGCGCCATCGACCGCGAGGCCGTGTTCACG<br>TACGCGGAGGACGCGGCCAGCGCCAGCCTGCCGCTGATGCAGAAGCTGCGCGCCG<br>TGCTGGTGGACCACGCCCTCAGCAGCGGCGAGCGCGGAGCGGGAGCCCTCCGTGT<br>TCTCCAAGATCACCAGGTTCGAGGAGGAGCTCCGCGCGGTGCTGCCCCAGGAGGT<br>GGAGGCCGCCCGCGTGGCGTCGCCGAGGGCACCGCCCCCGTGGCGAACCGGATCG<br>CGGACAGCCGGTCGTTCCCGCTGTACCGCTTCGTGCGCGAGGAGCTCGGCTGCGTG<br>TTCCTGACCGGCGAGAGGCTCAAGTCCCCCGGCGAGGAGTGCAACAAGGTGTTCG<br>TCGGCATCAGCCAGGGCAAGCTCGTGGACCCCATGCTCGAGTGCCTCAAGGAGTG<br>GGACGGCAAGCCGCTGCCCATCAACATCAAGTAA | Tyrosine ammonia lyase from *Zea mays* |

| | | |
|---|---|---|
| C3H | ATGTCGTGGTTTCTAATAGCGGTGGCGACAATCGCCGCCGTCGTATCCTACAAGCT<br>AATCCAACGGCTAAGATACAAGTTCCCACCAGGCCCGCCGTCGTATCCTACAAGCT<br>GGTAACCTCTACGACATAAAACCGGTCCGGTTCAGATGTTACTACGAGTGGGCTC<br>AATCTTATGGACCAATCATATCGGTCTGGATCGGTTCAATTCTAAACGTGGTCGTA<br>TCTAGCGCCGAGCTAGCAAAAGAAGTTCTGAAAGAACACGACCAGAAACTCGCCG<br>ACCGGCACCGGAACAGATCGACGGAAGCATTTAGCCGCAACGGTCAGGATCTTAT<br>ATGGGCCGATTATGGGCCTCATTACGTGAAGGTGAGAAAAGTTTGCACGCTTGAG<br>CTCTTCACACCGAAACGACTCGAGTCTCTCAGACCTATCCGTGAAGATGAAGTCAC<br>CGCCATGGTTGAATCCGTCTTCAGAGACTGTAACCTTCCTGAAAACAGAGCAAAA<br>GGTTTACAACTGAGGAAGTACTTAGGAGCGGTTGCGTTCAACAACATAACGCGGC<br>TAGCCTTTGGGAAGCGTTTTATGAACGCTGAAGGTGTTGTGGACGAGCAAGGGCT<br>TGAGTTCAAGGCCATAGTATCCAACGGTCTGAAGCTAGGTGCTTCACTGTCAATAG<br>CTGAACACATCCCGTGGCTCAGGTGGATGTTTCCGGCTGATGAGAAGGCGTTTGCT<br>GAGCACGGGGCTCGTCGTGACCGCCTCACTCGAGCTATCATGGAGGAGCATACTT<br>TGGCCCGTCAAAAGTCTAGTGGAGCGAAACAGCATTTCGTTGATGCGTTGCTAAC<br>GTTGAAGGATCAGTATGATCTTAGTGAGGATACTATCATTGGTCTTCTATGGGATA<br>TGATCACGGCAGGGATGGACACGACAGCGATAACAGCGGAATGGGCGATGGCGG<br>AAATGATCAAGAATCCAAGAGTGCAACAAAAAGTGCAAGAAGAGTTCGACAGAG<br>TGGTTGGACTTGACCGGATCTTAACCGAGGCAGATTTCTCCCGCTTACCTTACTTG<br>CAATGCGTGGTGAAAGAGTCATTCAGGCTGCATCCTCCAACGCCTCTAATGCTACC<br>TCACCGAAGCAACGCAGATGTCAAGATCGGAGGCTATGATATTCCCAAAGGATCA<br>AACGTTCATGTGAATGTGTGGGCTGTGGCTAGAGACCCGGCTGTATGGAAAAATC<br>CATTTGAGTTTAGACCAGAGAGATTCTTGGAAGAAGATGTTGACATGAAGGGGTCA<br>TGATTTTAGGCTGCTTCCGTTTGGAGCTGGAAGACGGGTTTGTCCCGGTGCACAAC<br>TTGGTATCAATTTGGTAACTTCGATGATGAGTCATTTGCTTCACCATTTTGTTTGGA<br>CACCTCCTCAAGGGACTAAACCGGAGGAGATTGACATGTCTGAAAACCCTGGACT<br>CGTTACTTACATGCGTACCCCTGTGCAAGCGGTTGCAACGCCTCGGTTGCCTTCGG<br>ATCTGTACAAACGCGTGCCTTACGATATGTAA | *p*-coumarate 3-hydroxylase<br>from *Arabidopsis thaliana* |
| HCT | ATGAAAATTAACATCAGAGATTCCACCATGGTCCGGCCTGCCACCGAGACACCAA<br>TCACTAATCTTTGGAACTCCAACGTCGACCTTGTCATCCCCAGATTCCATACCCCT<br>AGTGTCTACTTCTACAGACCCACCGGCGCTTCCAATTTCTTTGACCCTCAGGTCAT<br>GAAGGAAGCTCTTTCCAAAGCCCTTGTCCCTTTTTACCCTATGGCTGGTCGCTTGA<br>AGAGAGACGATGATGGTCGTATTGAGATCGATTGTAACGGTGCTGGTGTTCTCTTC<br>GTTGTGGCTGATACTCCTTCTGTTATCGATGATTTTGGTGATTTTGCTCCTACCCTT<br>AATCTCCGTCAGCTTATTCCCGAAGTTGATCACTCCGCTGGCATTCACTCTTTCCCG<br>CTTCTCGTTTTGCAGGTGACTTTCTTTAAATGTGGGGGAGCTTCACTTGGGGTTGG<br>GATGCAACATCACGCGGCAGATGGTTTCTCTGGTCTTCATTTTATCAACACATGGT<br>CTGATATGGCTCGTGGTCTTGACCTAACCATTCCACCTTTCATTGATCGAACACTCC<br>TCCGAGCTAGGGACCCGCCACAGCCTGCTTTTCATCATGTTGAATATCAGCCTGCA<br>CCAAGTATGAAGATACCTCTTGATCCGTCTAAATCAGGACCTGAGAATACCACTGT<br>CTCTATATTCAAATTAACACGAGACCAGCTTGTTGCTCTTAAGGCGAAATCCAAGG<br>AGGATGGGAACACTGTCAGCTACAGCTCATACGAGATGTTGGCAGGGCATGTGTG<br>GAGATCAGTGGGAAAGGCGCGAGGGCTTCCAAACGACCAAGAGACGAAACTGTA<br>CATTGCAACTGATGGAAGGTCTAGACTACGTCCGCAGCTGCCTCCTGGTTACTTTG<br>GGAATGTGATATTCACTGCAACACCATTGGCTGTTGCAGGGGATTTGTTATCTAAG<br>CCAACATGGTATGCTGCAGGACAGATTCATGATTTCTTGGTTCGTATGGATGATAA<br>CTATCTGAGGTCAGCTCTTGACTACCTGGAGATGCAGCCTGATCTGTCAGCCCTTG<br>TCCGCGGTGCACATACCTACAAGTGCCCAAATTTGGGAATCACAAGCTGGGTTAG<br>ATTACCTATTTATGATGCAGACTTTGGTTGGGGTCGTCCTATCTTTATGGGACCTGG<br>TGGAATTCCATACGAGGGTTTGTCTTTTGTGCTACCAAGTCCTACTAATGATGGCA<br>GCTTATCCGTTGCCATTGCCCTCCAATCTGAACACATGAAACTGTTTGAGAAGTTT<br>TTGTTTGAGATATAA | Hydroxycinnamoyl transferase<br>from *Arabidopsis thaliana* |

| 4CL | ATGGCGGCGACACATCTTCACATACCGCCAAATCCTAAAACCCAAACGTCTCACC<br>AAAACCCTCCCTTTTGGTTCTCTTCTAAAACAGGAATCTACACCAGCAAATTCCCT<br>TCCCTACACCTCCCCGTCGACCCAAATCTCGACGCCGTCTCCGCTCTTTTCTCACAC<br>AAACACCACGGCGATACAGCGCTCATCGATTCCTTAACCGGGGTTCTCAATATCTCA<br>CACTGAGCTACAGATTATGGTTCAATCAATGGCGGCTGGGACCTATCACGTTTTAG<br>GTGTTCGTCAAGGTGACGTTGTATCACTCGTCTTGCCTAATTCCGTCTATTTTCCGA<br>TGATTTTCCTCTCTTTGATCTCGCTTGGTGCTATTGTTACTACCATGAATCCTTCGA<br>GTAGTTTAGGAGAGATTAAGAAGCAAGTTAGTGAGTGTAGTGTTGGATTAGCTTTT<br>ACTTCTACTGAAAACGTTGAGAAGCTGAGTTCTTTGGGGGGTTAGTGTGATTAGTGT<br>ATCTGAAAGTTACGATTTTGATTCGATTCGTATCGAAAACCCGAAGTTTTACTCCA<br>TTATGAAAGAAAGTTTTGGGTTTGTACCAAAACCGTTGATTAAGCAAGACGATGT<br>AGCTGCAATTATGTATTCATCTGGAACAACTGGAGCTAGTAAAGGAGTTTTGTTAA<br>CTCATAGGAATTTGATAGCATCAATGGAGCTTTTTGTGAGGTTTGAAGCTTCTCAG<br>TACGAATATCCGGGATCGAGTAATGTTTATCTAGCAGCTTTGCCTTTGTGCCATAT<br>CTACGGGTTATCACTCTTTGTGATGGGATTGTTGTCTCTGGGATCAACTATTGTTGT<br>TATGAAGAGGTTTGATGCTTCTGATGTTGTTAATGTAATTGAAAGGTTTAAGATCA<br>CTCATTTCCCGGTTGTTCCACCTATGTTGATGGCATTGACAAAGAAGGCAAAAGGA<br>GTTTGTGGTGAAGTGTTTAAAAGTTTGAAGCAAGTTTCTTCTGGAGCAGCTCCTTT<br>GAGTAGGAAGTTCATTGAAGATTTCCTTCAGACTCTTCCTCATGTCGATTTGATTC<br>AGGGATATGGAATGACTGAATCAACTGCAGTTGGAACACGGGGCTTCAACTCGGA<br>AAAGCTTAGTAGGTATTCTTCTGTGGGACTACTGGCTCCTAATATGCAAGCCAAGG<br>TAGTAGATTGGAGCTCTGGTTCTTTCCTTCCACCAGGAAACCGAGGAGAGCTCTGG<br>ATACAAGGTCCCGGTGTCATGAAAGGATACTTGAATAACCCAAAAGCAACTCAGA<br>TGAGTATAGTTGAAGATAGTTGGCTACGTACTGGGGATATTGCTTATTTTGATGAA<br>GATGGTTACTTGTTTATTGTTGACCGTATTAAGGAGATTATAAAGTACAAAGGCTT<br>TCAGATCGCTCCAGCAGATTTGGAGGCTGTTCTTGTTTCACATCCCTTGATTATCGA<br>CGCTGCAGTAACAGCTGCCCCAAATGAAGAATGTGGAGAGATTCCGGTAGCATTC<br>GTTGTCCGGAGACAAGAAACAACACTTTCAGAAGAAGATGTAATAAGCTATGTAG<br>CTTCTCAGGTTGCACCCTACAGGAAGGTGAGGAAAGTGGTGATGGTCAACTCTAT<br>ACCAAAATCTCCAACTGGGAAGATATTGAGAAAGGAACTAAAGAGAATTCTTACA<br>AACTCTGTTTCTTCAAGACTATGA | 4-hydroxycinnamoyl-CoA ligase from *Arabidopsis thaliana* |
| COMT | ATGGCTACATGGGTGGAGCACCAACAGCAGCAAAATGGATCCAAGGACGTGGAC<br>GAGGAGGCGTGCATGTACGCCATGCAGTTGTCGAGCATGGTCGTCCTCCCGATGA<br>CGCTTAGGGTAGCCGTCGAGCTCGGCATACTCGAACAAATCCAGGCCGGGGGCCC<br>AGATTCGTACCTTACTGCCGAGGATTTGGCGGCGAGGCTCGGCAACTCCAACCCCT<br>TAGCTCCGGTCATGATCGAGCGGATCCTGCGCCTGCTCACCAGCTACTCCATCCTT<br>AACTTCACCGACACCGTCGACGGGGAGGGTAGGACCGTCCGGAGCTACGGCGCGG<br>CGCATGTCTGCAAGTACCTGACTCCCAACCAGGACGGCGTCTCCATGGCGCCTCTC<br>GTCCTCATGAACACGGATAAGGTCCTTATGGAGAGCTGGTACCACATGAAGGATG<br>CAGTGACAAATGGTGGAATACCATTCAATCTAGCATATGGGATGACAGCTTTTGA<br>GTATCATGGGAAAGATCTAAGGTTTAATAAGGTGTTCAACGAGGGCATGAAGAAC<br>AACTCGATCATTATAACGAAGAAGATTTTAGAGAGATACAAAAGGTTTGAAGATG<br>TCAATGTTTTAATTGATGTTGGTGGTGGTGGAATTGGTGGAACTATCAGTATGATTACT<br>GCAAAGTATCCACATATACATGGGATTAATTTTGACCTTCCTCATGTTGTTTCTGA<br>AGCTCCACCTTTCCAAGGGGTAGAACATGTCGGTGGAAACATGTTTGAAAGTGTC<br>CCCATTGGTGATGCAATCTTCATAAAGTGGATTCTTCATGATTGGAGTGATGAGCA<br>TTGTTTGAAGCTCCTAAGAAATTGTGCAAAATCTTTACCTGACAAAGGAAAAGTCA<br>TAGTTGTGGAATGCATTCTTCCCGATGCACCTTTGGTGACGCCAGAGGCTGAAGGT<br>GTCTTTCATTTGGACATGATAATGTTGGCTCACAATCCTGGGGGAAAGGAGAGAA<br>CAAAGAAAGAGTTTAAAGAATTGGCTATGCTATCTGGTTTCTCTAATTTCAAGGCA<br>CTTTTTAGTTATGCTAATGTTTGGGTCATGGAATTCAACAAATAA | Caffeic acid 3-*O*-methyltransferase from *Vanilla planifolia* |

| | | |
|---|---|---|
| VAN | ATGGCAGCTAAGCTCCTCTTCTTCCTACTCTTCCTGGTCTCCGCCCTCTCCGTCGCG<br>CTCGCCGGTTTCGAAGAAGACAATCCAATCCGGTCCGTTACACAAAGGCCTGACT<br>CGATTGAGCCTGCCATCCTCGGCGTCCTTGGCAGTTGCCGCCACGCCTTCCACTTC<br>GCACGGTTCGCCCGCAGGTACGGGAAGAGCTACGGATCGGAGGAGGAGATCAAG<br>AAGAGGTTCGGGATCTTCGTGGAGAATCTAGCGTTTATCCGGTCCACTAATCGGAA<br>GGATCTGTCGTATACCCTAGGAATCAACCAATTCGCCGACCTGACCTGGGAGGAA<br>TTCCGGACCAATCGCCTTGGTGCGGCGCAGAACTGCTCGGCGACTGCGCATGGAA<br>ACCACCGGTTTGTCGATGGCGTGCTTCCTGTAACGAGGGATTGGAGGGAGCAAGG<br>GATAGTGAGCCCTGTAAAGGACCAAGGAAGCTGTGGATCTTGCTGGACTTTCAGT<br>ACTACTGGAGCACTAGAGGCTGCATATACACAGCTAACTGGAAAGAGCACATCAT<br>TATCTGAACAGCAACTTGTGGACTGTGCCTCAGCATTCAATAACTTTGGATGCAAT<br>GGAGGTTTGCCTTCCCAAGCCTTTGAATACGTTAAGTACAATGGAGGCATCGACAC<br>AGAACAGACTTATCCATACCTTGGTGTCAATGGTATCTGCAACTTCAAGCAGGAG<br>AATGTTGGTGTCAAGGTCATTGATTCGATAAACATCACCCTGGGTGCTGAGGATGA<br>GTTGAAGCATGCAGTGGGCTTGGTGCGTCCAGTTAGCGTTGCATTTGAGGTTGTGA<br>AAGGTTTCAATCTGTACAAGAAAGGTGTATACAGCAGTGACACCTGTGGAAGAGA<br>TCCAATGGATGTGAACCACGCAGTTCTTGCCGTCGGTTATGGAGTCGAGGACGGG<br>ATTCCTTATTGGCTCATCAAGAACTCATGGGGTACAAATTGGGGTGACAATGGCTA<br>CTTTAAGATGGAACTCGGCAAGAACATGTGTGGTGTTGCAACTTGCGCATCTTATC<br>CCATTGTGGCTGTGCACCACCACCACCACCACTAG | Vanillin synthase from *Vanilla planifolia* |

# Appendix C

# Supplemental information for Chapter 3

## C.1 Supplemental Tables

**Table C.1: Wild-type *P. tricornutum* UMPS sequence analysis.** Base positions are numbered relative to the first base of the start codon in the genomic PtUMPS sequence. SNPs highlighted in red were not present in the reference genome. SNPs located in intronic sequences were not included in this table.

| Base Position | Reference Genome | Allele 1 Base | Allele 2 Base | Allele 1 Residue | Allele 2 Residue |
|---|---|---|---|---|---|
| 102 | A/G | A | G | A34 | A34 |
| 127 | G/A | G | A | E43 | K43 |
| 237 | T/C | T | C | D79 | D79 |
| 246 | G/A | G | A | M82 | I82 |
| 519 | G | G | A | V147 | I147 |
| 542 | G/T | G | T | L154 | L154 |
| 591 | C/A | C | A | R171 | S171 |
| 599 | T/C | T | C | N173 | N173 |
| 938 | G/A | G | A | M286 | I286 |
| 1003 | A/T | A | T | Q308 | L308 |
| 1048 | C/T | C | T | T323 | I323 |
| 1930 | T/A | T | A | I510 | I510 |

**Table C.2: Wild-type *P. tricornutum* PRA-PH/CH sequence analysis.** Base positions are relative to the first base of the start codon in the genomic PtPRA-PH/CH sequence. SNPs highlighted in red were not present in the reference genome.

| Base Position | Reference Genome | Allele 1 Base | Allele 2 Base | Allele 1 Residue | Allele 2 Residue |
|---|---|---|---|---|---|
| 113 | T/A | T | A | F38 | Y38 |
| 203 | A/C | A | C | Q68 | P68 |
| 268 | C/T | C | T | L90 | F90 |
| 270 | C/T | C | T | L90 | F90 |
| 318 | T/C | T | C | V106 | V106 |
| 319 | T/C | T | C | L107 | L107 |
| 354 | C/T | C | T | C118 | C118 |
| 406 | A/G | A | G | T136 | A136 |
| 423 | T/G | T | G | F141 | L141 |
| 623 | A/T | A | T | Q208 | L208 |
| 732 | G/A | G | A | R244 | R244 |
| 817 | A/G | A | G | T273 | A273 |
| 825 | C/T | C | T | V275 | V275 |
| 892 | C/T | C | T | L298 | L298 |
| 978 | T/G | T | G | L326 | L326 |
| 1205 | A/G | A | G | E402 | G402 |
| 1306 | G/A | G | A | V436 | I436 |
| 1323 | C/T | C | T | A441 | A441 |
| 1350 | G/A | G | A | R450 | R450 |

**Table C.3: Growth rates of WT and UMPS complement strains in minimal L1 media.** Generation times represent the mean value ± standard deviation for three replicates.

| *P. tricornutum* strain | generation time (hours) |
|---|---|
| WT | 29.8±2.6 |
| ΔUMPS2 + pPtUMPSA1 | 27.9±2.8 |
| ΔUMPS2 + pPtUMPSA2 | 24.8±1.5 |
| ΔUMPS2 + pPtUMPScA1 | 28.5±0.9 |
| ΔUMPS2 + pPtUMPScA2 | 23.4±0.8 |
| ΔUMPS1 + pPtUMPSA1 | 26.5±2.7 |
| ΔUMPS1 + pPtUMPSA2 | 29.8±0.7 |
| ΔUMPS1 + pPtUMPScA1 | 26.5±4.0 |
| ΔUMPS1 + pPtUMPScA2 | 30.7±1.2 |

**Table C.4: List of *P. tricornutum* auxotroph genotypes.**

| Strain Name | Genotype | Description |
|---|---|---|
| ΔUMPS1 | PtUMPSg.[1637_1654del];[1050_1661del] | *P. tricornutum* uracil auxotroph strain 1 |
| ΔUMPS2 | PtUMPSg.[1636_1652del];[1636_2415delinsT] | *P. tricornutum* uracil auxotroph strain 2 |
| ΔUMPS3 | PtUMPSg.[309_310insA];[309_310insA] | *P. tricornutum* uracil auxotroph strain 3 |
| ΔPRAPHCH1 | PtPRA-PH/CHg.[924_934del];[924_929del] | *P. tricornutum* histidine auxotroph strain 1 |

**Table C.5: List of plasmids used in this study.**

| Plasmid | Description | Reference or Source |
|---|---|---|
| pPtGE31 | *P. tricornutum* expression vector | Slattery, et al., 2018 |
| pPtGE34 | *P. tricornutum* expression vector, *40SRPS8* promoter and terminator driving *Sh ble*, *FcpB* promoter and *FcpA* terminator driving Cas9 | Slattery, et al., 2018 |
| pPtGE35 | *P. tricornutum* expression vector, *40SRPS8* promoter and terminator driving *Sh ble*, *FcpB* promoter and *FcpA* terminator driving TevCas9 | Slattery, et al., 2018 |
| pPtUMPSA1 | pPtGE31 encoding PtUMPS allele 1 driven by the PtUMPS promoter and terminator | This study |
| pPtUMPSA2 | pPtGE31 encoding PtUMPS allele 2 driven by the PtUMPS promoter and terminator | This study |
| pPtUMPScA1 | pPtGE31 encoding PtUMPS allele 1 cDNA driven by the PtUMPS promoter and terminator | This study |
| pPtUMPScA2 | pPtGE31 encoding PtUMPS allele 2 cDNA driven by the PtUMPS promoter and terminator | This study |
| pPtPRAPHCH | pPtGE31 encoding PtPRA-PH/CH allele 1 driven by the PtPRA-PH/CH promoter and terminator | This study |
| pTA-Mob | Mobilization helper plasmid required for conjugation | Strand, et al., 2014 |

**Table C.6: List of oligonucleotides used in this study.**

| Name | Sequence (5' to 3') (Priming sequence for PCR) | Description |
|------|-----------------------------------------------|-------------|
| DE3644 | tcgaATTAAGTATCGAAACGAATA | Top strand sgRNA.UMPS.1944 for PtUMPS |
| DE3645 | aaacTATTCGTTTCGATACTTAAT | Bottom strand sgRNA.UMPS.1944 for PtUMPS |
| DE3646 | tcgaTAAATTGGTCGGGACTTCGT | Top strand sgRNA.UMPS.1646 for PtUMPS |
| DE3647 | aaacACGAAGTCCCGACCAATTTA | Bottom strand sgRNA.UMPS.1646 for PtUMPS |
| DE3648 | tcgaCGACCAACGTTTTGCAAA | Top strand sgRNA.UMPS.157 for PtUMPS |
| DE3649 | aaacTTTGCAAAACGTTGGTCG | Bottom strand sgRNA.UMPS.157 for PtUMPS |
| DE4236 | tcgagATTTTGTTGGATGTCAAGCG | Top strand sgRNA.UMPS.311 for PtUMPS |
| DE4237 | aaacCGCTTGACATCCAACAAAATc | Bottom strand sgRNA.UMPS.311 for PtUMPS |
| DE4162 | tcgagCTACTACAGCCGTTCCCGGAA | Top strand sgRNA.PRAPHCH.929 for PtPRA-PH/CH |
| DE4163 | aaacTTCCGGGAACGGCTGTAGTAGc | Bottom strand sgRNA.PRAPHCH.929 for PtPRA-PH/CH |
| DE4164 | tcgagTCGCGAAAGTCAGCGGCCCC | Top strand sgRNA.PRAPHCH.120 for PtPRA-PH/CH |
| DE4165 | aaacGGGGCCGCTGACTTTCGCGAc | Bottom strand sgRNA.PRAPHCH.120 for PtPRA-PH/CH |
| DE4166 | tcgagGGCTCGACAAAGACTGTGAC | Top strand sgRNA.PRAPHCH.1000 for PtPRA-PH/CH |
| DE4167 | aaacGTCACAGTCTTTGTCGAGCCc | Bottom strand sgRNA.PRAPHCH.1000 for PtPRA-PH/CH |
| DE3650 | tcgaGGACTCGGACAACCCCGATT | Top strand sgRNA.IGPSPRAI.244 for PtI3GPS-PRAI |
| DE3651 | aaacAATCGGGGTTGTCCGAGTCC | Bottom strand sgRNA.IGPSPRAI.244 for PtI3GPS-PRAI |
| DE3728 | GGCCTGGCAGCTTTATCAGTAG | Rev screening primer for sgRNA.UMPS.1944 site (pair with DE3646) |
| DE3726 | GGTCAACACCAATTCGCTG | Fwd screening primer for sgRNA.UMPS.1646 site |
| DE3727 | GAAGACTGACAATTGACACGACC | Rev screening primer for sgRNA.UMPS.1646 site |
| DE3724 | GACGAAGTGTCTACTCACAGACAGC | Fwd screening primer for sgRNA.UMPS.157 and sgRNA.UMPS.311 sites |
| DE3725 | TCCCATCAGTGGTGAAAGCG | Rev screening primer for sgRNA.UMPS.157 and sgRNA.UMPS.311 sites |
| DE4176 | ACGCCGCCATGGTTGGTC | Fwd sgRNA.PRAPHCH.929 screening primer |
| DE4177 | CTCGTCTTGAAACAGACGTTTCGTG | Rev sgRNA.PRAPHCH.929 screening primer |
| DE4174 | CCGAGTGATACTGTTCGCTTCG | Fwd sgRNA.PRAPHCH.120 screening primer |
| DE4175 | GGAGTCCACTCGCGAGAC | Rev sgRNA.PRAPHCH.120 screening primer |
| DE4178 | GACCGCCTCGCAACTCGG | Fwd sgRNA.PRAPHCH.1000 screening primer |
| DE4179 | CCACGTGTGCCTTCGTATCG | Rev sgRNA.PRAPHCH.1000 screening primer |
| DE4553 | AGCAGGGTTATGCAGCGGAAGATCTATATTAC CCTGTTAT**AAAGGCGGCGAATACTTCAT** | Fwd PtUMPS promoter with pPtGE31 homology |
| DE4552 | TGCAGTCACTCCGCTTTGGTTTCGTAACTATA ACGGTC**TCGATGACTGACAAAGGTATTC** | Rev PtUMPS terminator with pPtGE31 homology |
| DE4622 | TGCAGTCACTCCGCTTTGGTTTCGTAACTAT AACGGTC**CTCGCGTTTTACGCGCCGTC** | Fwd PtPRA-PH/CH promoter with pPtGE31 homology |
| DE4623 | AGCAGGGTTATGCAGCGGAAGATCTATATT ACCCTGTTAT**GCCCGCCTGATGCGTCG** | Rev PtPRA-PH/CH terminator with pPtGE31 homology |
| DE5178 | TCCTTTTCCAGTGCTTCGAT | Fwd primer to amplify ~6kb around sgRNA.IGPSPRAI.244 site |
| DE5179 | TACCTGCTGCATCAGCTTTG | Rev primer to amplify ~6kb around sgRNA.IGPSPRAI.244 site |
| DE5180 | TGGGTGTTGTGCTCTGCTAC | Fwd primer to amplify ~6kb around sgRNA.UMPS.1944 site |
| DE5181 | TCTCCAAAGCCCAATTTTTG | Rev primer to amplify ~6kb around sgRNA.UMPS.1944 site |
| DE5182 | ACTCAGCGTCACCTCCACTT | Fwd primer to amplify ~6kb around sgRNA.UMPS.311 site |
| DE5183 | GAATGCACCGATCACAACTG | Rev primer to amplify ~6kb around sgRNA.UMPS.311 site |
| DE5184 | TCGAGGGAAGAGGCTAGACA | Fwd primer to amplify ~6kb around PtUrease sgRNA.UREASE.1187 site |
| DE5185 | TTCCGTTGCATTGATGTTGT | Rev primer to amplify ~6kb around PtUrease sgRNA.UREASE.1187 site |

# Appendix D

# Supplemental information for Chapter 4

## D.1   Supplemental Tables

**Table D.1: List of plasmids used in this study.**

| Plasmid | Description | Reference or Source |
|---|---|---|
| pPtGE31 | *P. tricornutum* expression vector (also called pDMI-2) | Slattery et al., 2018 |
| pPtPRAPHCH | pPtGE31 encoding PtPRA-PH/CH allele 1 driven by the PtPRA-PH/CH promoter and terminator | Slattery et al., 2020 |
| pDMI-2 PtRBD c14 | pPtGE31 encoding *P. tricornutum* codon-optimized RBD driven by *HASP1* promoter version 1 | This study |
| pDMI-2 PtRBD c14 | pPtGE31 encoding *P. tricornutum* codon-optimized RBD driven by *HASP1* promoter version 2 | This study |
| pDMI-2 PtSpike c1 | pPtGE31 encoding *P. tricornutum* codon-optimized full-length Spike driven by *HASP1* promoter version 1 | This study |
| pDMI-2 PtSpike c2 | pPtGE31 encoding *P. tricornutum* codon-optimized full-length Spike driven by *HASP1* promoter version 2 | This study |
| pDMI-2 PtSpike2P c1 | pPtGE31 encoding *P. tricornutum* codon-optimized full-length Spike (K986P and V987P) driven by *HASP1* promoter version 1 | This study |
| pDMI-2 PtSpike2P c2 | pPtGE31 encoding *P. tricornutum* codon-optimized full-length Spike (K986P and V987P) driven by *HASP1* promoter version 2 | This study |
| pDMI-2 HsRBD | pPtGE31 encoding human codon-optimized RBD driven by 40SRPS8 promoter | This study |
| pDMI-2 HsSpike | pPtGE31 encoding human codon-optimized full-length Spike driven by 40SRPS8 promoter | This study |
| pDMI-2 HsSpike2P | pPtGE31 encoding human codon-optimized full-length Spike (K986P and V987P) driven by 40SRPS8 promoter | This study |
| pTA-Mob | Mobilization helper plasmid required for conjugation | Strand et al., 2014 |

**Table D.2: List of oligonucleotides used in this study.**

| Name | Sequence (5' to 3') (Priming sequence for PCR) | Description |
|---|---|---|
| DE3197 | ATCTTCCGCTGCATAACCC | Fwd pDMI-2 vector |
| DE3627 | ATGGATATACCGAAAAAATCGCTATAATGACCCCGAAGCAGGGTTATGCAGCGGAAGAT**GGGGTGGATAAAGAAGAAAGG** | Rev 40SRPS8 terminator with pDMI-2 vector homology |
| DE4130 | CCCTGCGATAGACCTTTTCC | Fwd 40SRPS8 Promoter |
| DE3061 | GGTATTCTATTCTCTGATTC | Rev 40SRPS8 Promoter |
| DE5239 | ATGAATCTTCGTTGTATCCTTCCG | Fwd PtSpike/RBD (universal) |
| DE5240 | TGGTAAATCATAGATGTTCACAATGAGATATTCTTATCT**TCAATGGTGATGGTGATGGTG** | Rev PtSpike/RBD (universal) with 40SRPS8 terminator homology |
| DE5241 | TTTCCCAGCTACTCGACGCATCAGGCGGGCGGATTG**ATAGTGAAAACCTTATTCATTGTC** | Fwd *HASP1* promoter and signal peptide with HIS4 cassette homology |
| DE5242 | AGCCCCAGCCGAGAAG | Rev *HASP1* promoter and signal peptide |
| DE5243 | AAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGGTTT**CTCGCGTTTTACGCGCC** | Fwd HIS4 cassette with pDMI-2 vector homology |
| DE5244 | ACCGGCTTAAGCTCTGACAATGAATAAGGTTTTCACTAT**CAATCCGCCCGCCTGATGCGT** | Rev HIS4 cassette with *HASP1* promoter homology |
| DE5245 | AGATAAGAATATCTCATTGTGAACA | Fwd 40SRPS8 terminator |
| DE5246 | AGTTCGTGGGCCAAGAAACTGACGGCGCGTAAAACGCGAG**AAACCAAAGCGGAGTGACTG** | Rev pDMI-2 vector with HIS4 cassette homology |
| DE5247 | GCGTTGATCTTGCACCGAAGGAATCAGAGAATAGAATACC**ATGTTCGTGTTTCTGGTGCT** | Fwd HsSpike/RBD (universal) with 40SRPS8 promoter homology |
| DE5248 | TAAATCATAGATGTTCACAATGAGATATTCTTATCT**TTATCAATGGTGATGGTGATGGTG** | Rev HsRBD with 40SRPS8 terminator homology |
| DE5249 | AATCATAGATGTTCACAATGAGATATTCTTATCT**TCATTAGTGATGATGATGATGATGTC** | Rev HsSpike with pDMI-2 vector homology |
| DE5250 | TTGTTTTCTTGGACTGCGTGAGTTTGGAAAAGGTCTATCGCAGGG**GCCCGCCTGATGCGT** | Rev HIS4 cassette with 40SRPS8 promoter homology |
| DE5251 | GCGGGATCTGCCTCAGG | Fwd HsSpike sequencing primer 1 |
| DE5252 | CCTGTACCGGCTGTTCCG | Fwd HsSpike sequencing primer 2 |
| DE5253 | GCCAGCCAGAGCATCATTG | Fwd HsSpike sequencing primer 3 |
| DE5254 | ATGTGCTGTACGAGAACCAG | Fwd HsSpike sequencing primer 4 |
| DE5255 | AGATTCTGTCCGACGAAGCG | Fwd PtSpike/RBD sequencing primer 1 |
| DE5256 | CTCCAAACACACTCCGATTAAC | Fwd PtSpike sequencing primer 2 |
| DE5257 | CTTGGATTCGAAAGTTGGAGG | Fwd PtSpike sequencing primer 3 |
| DE5258 | CCAGACTCAAACGAATTCGC | Fwd PtSpike sequencing primer 4 |
| DE5259 | GGCCTACCGTTTCAACGG | Fwd PtSpike sequencing primer 5 |
| DE5260 | ACCGTGTACGATCCGCTG | Fwd PtSpike sequencing primer 6 |
| DE5323 | ACGGAGTAATCGGCGACGC | Rev PtRBD screening primer |
| DE5326 | GAGTTGTACAGCACGGAGTAGTCG | Rev HsRBD screening primer |

# Curriculum Vitae

## SAMUEL S. SLATTERY
======================

---

## Education:

PhD Candidate, Biochemistry, 2014-current
    Western University, London, Ontario

Bachelor of Science, Honors Specialization in Genetics and Biochemistry, 2010-2014
    Western University, London, Ontario

## Publications:

**Slattery, S.S.**, Cholod, J., Stuckless, E., Wang, H., Olteanu, D., Rizvi, I., Karas, B.J., Edgell, D.R. (2020). Examining the Effects of a Ku70 Knockout on DNA Damage Repair in Phaeodactylum tricornutum. (In Preparation).

**Slattery, S.S.**, Wang, H., Giguere, D.J., Kocsis, C., Urquhart, B.L., Karas, B.J., Edgell, D.R. (2020). Plasmid-Based Complementation of Large Deletions in Phaeodactylum tricornutum Biosynthetic Genes Generated by Cas9 Editing. Scientific Reports. 10: 1-12.
https://doi.org/10.1038/s41598-020-70769-6

Wang, H., **Slattery, S.S.**, Karas, B.J., Edgell, D.R. (2018). Delivery of the Cas9 or TevCas9 System into Phaeodactylum tricornutum via Conjugation of Plasmids from a Bacterial Donor. Bio-protocol 8(16): e2974.
http://dx.doi.org/10.21769/BioProtoc.2974

**Slattery, S.S.**, Diamond, A., Wang, H., Therrien, J.A., Lant, J.T., Jazey, T., Lee, K., Klassen, Z., Desgagne-Penix, I., Karas, B.J., Edgell, D.R. (2018). An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum. ACS Synthetic Biology. 7(2): 328-338.
http://dx.doi.org/10.1021/acssynbio.7b00191

**Presentations:**

An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum. Annual Conference of the Canadian Society of Microbiology (CSM), Sherbrooke, Quebec. June, 2019 (abstract/poster)

An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum. Synthetic Biology Symposium 4.0 Conference, Waterloo, Ontario. May, 2019 (abstract/poster)

An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum. Synthetic Biology: Engineering Evolution, and Design (SEED) Conference, Scottsdale, Arizona. June, 2018 (abstract/poster)

An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum. Synthetic Biology Symposium 3.0 Conference, London, Ontario. May, 2018 (abstract/poster)

Plasmid-based tools for generating designer Phaeodactylum tricornutum strains. Synthetic Biology Symposium 2.0 Conference, London, Ontario. May, 2017 (abstract/poster)

Plasmid-based tools for developing designer Phaeodactylum tricornutum strains. London Health Research Day, London, Ontario. March, 2017 (invited speaker)

**Awards and Honors:**

| Award | Value (CDN) | Level | Location of Tenure | Period Held |
|-------|-------------|-------|--------------------|-------------|
| MITACS Accelerate Internship | $10,000 | National | Western University | 2020/05-2020/09 |
| Ontario Graduate Scholarship | $15,000 | Provincial | Western University | 2018/05-2019/05 |
| Synthetic Biology Symposium Top Poster | $200 | Provincial | Western University | 2019 |
| Synthetic Biology Symposium 2nd Place Poster | $200 | Provincial | Western University | 2018 |
| Synthetic Biology Symposium Top Poster | $250 | Provincial | Western University | 2017 |
| Departmental Symposium Top Oral Presentation | $50 | Institutional | Western University | 2017 |
| Western Graduate Research Scholarship | (tuition) | Institutional | Western University | 2014/09-2019/09 |
| Dean's Honour List | - | Institutional | Western University | 2012/09-2014/05 |

**Experience:**

Laboratory
　　Research Assistant, Western University, Department of Biochemistry 2014/05 - 2014/09 (Dr. Edgell)

Teaching Assistantships
　　Biochemistry 3381A, Western University 2018
　　Biochemistry 3380G, Western University 2018
　　Biochemistry 3380G, Western University 2017
　　Biochemistry 2280A, Western University 2017
　　Biochemistry 2280A, Western University 2016
　　Biochemistry 2280A, Western University 2016 Summer
　　Biochemistry 2280A, Western University 2015