9-18-2023 11:00 AM

Characterization of DNA Regulatory Elements of the Highly Abundant Secreted Protein 1 (HASP1) Promoter and Analysis of Transcriptomic Responses Under Phosphate Depletion in the Diatom Phaeodactylum tricornutum

Mahsa Farmanbar, *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 Master of Science degree in Biochemistry
© Mahsa Farmanbar 2023

Abstract

Phosphorus (P) is essential for all life. The bioavailability of phosphorus in oceans impacts diatoms like *Phaeodactylum tricornutum (P. tricornutum)*. *P. tricornutum* is extensively used in transcriptomic studies to understand pathways involved in P-acquisition. However, activation mechanisms and roles of regulatory elements in P-acquisition responses remains unclear. Here, I deleted predicted phosphate regulatory sites in *P.tricornutum's HASP1* promoter to create different *HASP1*-eGFP constructs. Under P-depletion, two constructs showed increased eGFP secretion. Additionally, a *HASP1* knockout strain was grown under different phosphorus sources to determine if *HASP1* is a phytase. Cells grown in full phosphorus utilized organic sources of phosphorus and cells grown under P-depletion used inorganic sources. *HASP1* knockout cells were also used to understand transcriptional responses under phosphorus depletion using RNAseq. This revealed the loss of *HASP1* was not essential to cell survival. Understanding diatom responses under P-starvation will contribute to improving *P. tricornutum* for synthetic biology applications.

Keywords: *Phaeodactylum tricornutum,* phosphorus, phosphate starvation, promoter induction, regulatory elements, regulation, gene knockout, synthetic biology

Summary for Lay Audience

Phosphorus (P) is an essential element for all life forms. The amount of phosphorus in marine environments often fluctuates which impacts the ecology and physiology of marine phytoplankton living in these environments, specifically diatoms. Scientists are curious to understand how certain species living in these environments like the diatom, *Phaeodactylum tricornutum (P. tricornutum)*, can regulate internal systems to cope with adverse environmental conditions. One method is called a phosphate starvation response (or PSR). The process involves activating genes responsible for phosphate acquisition. Much of what we know about this phenomenon is through studies that look at what genes are turned up or down in response to phosphorus depletion. The downfall is we do not have a comprehensive understanding of how P-acquisition responses are activated in response to phosphate depletion, as well as localization and function of proteins involved. Thankfully, due to recent technological advances, we have the tools required to genetically manipulate regions within sections of DNA responsible for turning genes on or off to understand how genes are expressed.

In this study, I use a portion of the *P. tricornutum* genetic code, called the Highly Abundant Secreted Protein 1 (*HASP1*) promoter, which contains 4 regions predicted to be responsible for the activation of a phosphate starvation response. I create a series of deletions in the predicted sites and design novel *P. tricornutum* strains that are grown in various low-phosphorus media. 2 out of 3 mutant constructs showed an upregulation in promoter activity compared to normal *P. tricornutum* strains. As a result, there was an increase of protein secretion. These results introduce novel expression systems for the production of new proteins. Furthermore, the growth of the *HASP1* knockout strain was assessed under different phosphorus sources to observe phytase-like potential. Results showed the phosphorus source utilized was dependent on phosphorus concentration. Lastly, responses in the *HASP1* knockout revealed many genes were activated in response to nutrient depletion over time. Using new strategies to assess phosphate regulation in diatoms holds promising results in developing novel expression systems that may be used for biotechnological purposes.

Co-Authorship Statement

Dr. Samuel S. Slattery constructed the pSS10 plasmid containing the wildtype *HASP1* promoter and signal peptide driving eGFP. Mahsa Farmanbar cloned mutant promoters into the pSS10 plasmid. Dr. Slattery also conducted the 17-day experiment for RNAseq analysis and 5'RACE on wildtype *P. tricornutum* cells.

Emily E. Stuckless generated the CRISPR-Cas9 *HASP1* gene knockout. M. Farmanbar conjugated the pDMI-2 and gene rescue plasmids into the *HASP1 KO*.

Tyler S. Browne organized the raw count data from the RNAseq experiment completed by Dr. Slattery. M. Farmanbar used the count data to complete transcriptome analysis.

Dedication

I'd like to dedicate this thesis to all the young females wanting to pursue an education in STEM.

"Don't let anyone rob you of your imagination, your creativity,
or your curiosity. It's your place in the world; it's your life. Go on
and do all you can with it and make it the life
you want to live."

- *Mae Jemison.*

Acknowledgments

The journey that has been graduate school would not be complete without the people who accompanied me along the way.

To Mom and Dad: I would not be here if it wasn't for your support and love. You both left your home country to provide me and Mahshad a life where we could choose who we wanted to be and the freedom to do so. I love you.

To Mahshad: Thank you for being my rock throughout this life. You've always been my #1 cheerleader and continue to encourage me to challenge myself every day. I am amazed by your dedication and commitment to every project or task you take on and your continuous strength through life's toughest obstacles. I couldn't ask for a better older sister. You will forever be my role model.

To Max: Thank you for being the best furry friend a girl could ever wish for. You've been there through the toughest of times during my educational journey and I can't wait to see what the future holds for us.

To the C100 girls: I am so grateful to have met you both. Each of you brought something special to my life and for that I am ever grateful. Thank-you for the endless laughs and encouragements. You girls are some of the most hardworking and strongest individuals I have met, and I have no doubt you will all have great success in your future endeavors.

To the past and current members of the Edgell Lab: You have all been so helpful and supportive during my time in the lab. I couldn't have asked for a better group of individuals to spend countless hours (and weekends) with.

To the many other friends I made along the way, specifically from the O'Donoghue/Heinemann and Karas Lab: Thanks for always having an open ear, providing constructive feedback, and putting up with my silly shenanigans.

To my advisory committee: Drs. Greg Gloor and Bogumil Karas, your guidance and support throughout my graduate studies has been greatly appreciated. Thank you for the meaningful insights and thoughtful discussions during meetings.

And lastly, to the person who made all this possible, my supervisor, Dr. David Edgell. Thank you for taking a chance on me. Your encouragement and support during times where I didn't believe in myself truly pushed me to keep going. Your feedback has allowed me to enhance my critical thinking skills and to become a better scientist (and to read more). Your passion

and curiosity for science is contagious and I am forever grateful to have been a part of a group of ambitious and intelligent scientists.

Table of Contents

List of Tables

List of Figures

Appendix A: Supplementary Figures

List of Abbreviations, Symbols, and Nomenclature

°C	Degree(s) Celsius
$\mu E\ m^{-2}\ s^{-1}$	Microeinsteins per second per square meter
μL	Microliter(s)
μM	Micromolar
A_{600}	Absorbance at wavelength 600nm
A_{670}	Absorbance at wavelength 670nm
A_{750}	Absorbance at wavelength 750nm
AP	Alkaline phosphatase
bp	Base pair(s)
CCAP	Culture Collection of Algae and Protozoa
Cm	Chloramphenicol Antibiotic
CmR	Chloramphenicol resistances
DIP	Dissolved Inorganic Phosphorus
DNA	Deoxyribonucleic acid
DOP	Dissolved Organic Phosphorus
E. coli	*Escherichia coli*
eGFP	Enhanced Green Fluorescent Protein
FCP	Fucoxanthin Chlorophyll a/c binding Protein
fcpA	Fucoxanthin chlorophyll a/c binding protein A

g	g-force or relative centrifugal force (rcf)
Gm	Gentamycin Antibiotic
GmR	Gentamycin Resistance
HASP1	Highly Abundant Secreted Protein 1
HIS	Histidine
IP_x	Myo-inositol phosphates (x=1-6)
kDa	Kilodalton
KO	Knockout
½ x L1	Liquid saltwater media for *P. tricornutum* cells (contains 1/2 amount of salt)
L1	Liquid saltwater media for *P. tricornutum* cells
LB	Luria-Bertani Broth
LHCF	Light Harvesting Complex containing Fucoxanthin.
Mb	Megabase pair(s)
mg	Milligram(s)
mL	Milliliter(s)
MT1	Mutant 1 *HASP1* Promoter
MT2	Mutant 2 *HASP1* Promoter
MT3	Mutant 3 *HASP1* Promoter
Na_3PO_4	Sodium Phosphate
Nat	Nourseothricin antibiotic

NEB	New England Biolabs
nm	Nanometer(s)
NP-WT	*P. tricornutum* cells containing no plasmid.
NR	Nitrate Reductase
P1BS	Phosphate starvation response Binding Sites
P	Phosphorus
P_i	Orthophosphate Anion
PA	Phytic Acid
PCR	Polymerase Chain Reaction
pSS10	Plasmid containing wildtype *HASP1* promoter.
PSR	Phosphate Starvation Response
P. tricornutum	*Phaeodactylum tricornutum*
PtPSR	*P. tricornutum* Starvation Response
RBD	Receptor Binding Domain
SDS	Sodium Dodecyl Sulfate
SOC	Super Optimal broth with Catabolite repression
SP	Signal peptide
TSS	Transcriptional Start Site
WT	Wild type

List of Appendices

<div align="center">Chapter 1</div>

1   Introduction

1.1   Phosphorus as an Essential Element

Phosphorus (P) is a vital element of all life. It is the building block and a key component to many molecules like nucleic acids, adenosine triphosphates (ATP) and phospholipids [1,2,3]. Phosphorus is not only an essential element for healthy bones and teeth in humans, but it is often found in fertilizers to aid with agricultural crops [3]. Importantly, phosphorus supports the growth of photosynthetic organisms like phytoplankton which lie at the base of the marine food chain [3]. Phytoplankton, specifically diatoms, account for 40% of primary production (production of organic compounds from carbon dioxide) [4-6]. Thus, the bioavailability of phosphorus can greatly impact primary production rates in the ocean affecting species distribution and the structure of the marine ecosystem [7-11]. In nature, phosphorus is found in the form, phosphate (or orthophosphate, PO_4^{3-}), a free form of phosphorus. Orthophosphate in oceans can be written as an equation to denote average photosynthesis [7]:

$$106CO_2 + 16HNO_3 + H_3PO_4 + 122H_2O + \text{trace elements and vitamins } \textbf{--light} \rightarrow$$
$$C_{106}H_{263}O_{110}N_{16}P + 138O_2.$$
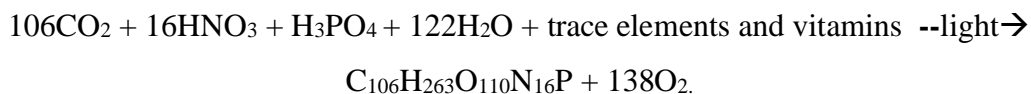
Unlike other molecules, P cannot be directly fixed from the atmosphere and can be considered the ultimate limiting macronutrient which poses a concern for many species in the ocean, specifically phytoplankton [12].

1.1.1   *Marine Phytoplankton Can Adapt to Fluctuating Phosphorus Levels*

Diatoms are unicellular photosynthetic microalgae and are among the most ecologically successful phytoplankton in oceans. Their presence in the ecosystem is of utmost importance, as they support both marine and terrestrial environments through their contribution to the global carbon cycle, annually generating an equivalent amount of organic carbon as the combined output of all terrestrial rainforests [2,4,7,8]. Diatom growth can be limited by a variety of nutrients such as nitrogen (N), silicon (Si), iron (Fe)

and phosphorus [4,13]. With the fast recycling of phosphorus in marine environments, the ecological success of diatoms becomes dependent on their resilience and capacity to respond to adverse environmental conditions through the regulation of genetic pathways responsible for phosphate acquisition. While diatoms demonstrate to be poor competitors in phosphate-deprived environments compared to flagellates, they possess the ability to endure these adverse conditions [2,9]. In regions with well-mixed coastal areas and upwelling regions, where nutrients are restored at optimal concentrations, diatoms frequently regain their dominance [2,10]. Of interest is determining how diatoms regulate their physiology in response to P fluctuations in the environment [11]. To cope with P fluctuations, one strategy that diatoms use is a phosphate starvation response (PSR) to increase the expression of genes responsible for phosphate acquisition [13]. As phosphorus becomes depleted, a signaling cascade will occur, recruiting transcription factors to regulatory sites in promoters of genes responsible for phosphate acquisition (i.e., phosphate transporters and alkaline phosphatases) leading to promoter induction and an increase in P-uptake (Figure 1.1) [1, 4, 14-19].
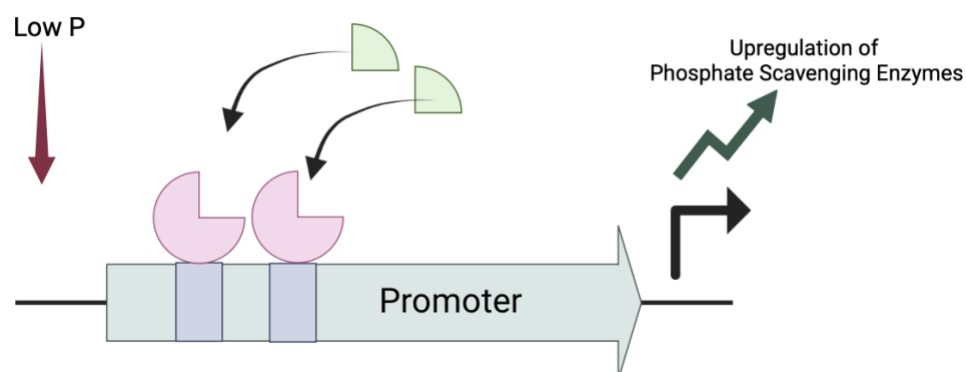
Figure 1-1: Low phosphorus recruits transcription factors to the promoters of phosphate regulatory genes in diatoms. Under low phosphorus conditions, inducers (green) bind to activators (pink) at specific binding sites (blue) which are responsible for the initiation of transcription for genes involved in phosphate acquisition.

In addition to the upregulation of scavenging enzymes to increase P-uptake, several studies have shown that during P-limitation diatoms will modify other metabolic pathways [4]. For example, during P-depletion in the diatom *Thalassiosira pseudonana (T. pseudonana)* and brown alga *Aureococcus anophagefferens (A. anophagefferens)* enzymes related to glycolysis and translation were upregulated [20, 21]. In other cases, diatoms will replace phospholipids with non-phosphorus membranes like sulfolipids, as seen in *T. pseudonana*, to reduce the demand of P [22,23]. In *Phaeodactylum tricornutum (P. tricornutum),* P-stress resulted in an increase of lipids, specifically triacylglycerols (TAG), which play a role in lipid biosynthesis [24].

Much of what is currently known about phosphate acquisition in diatoms is from transcriptomic studies which provide important information on gene expression in response to phosphate depletion [1,2,4]. However, these studies do not provide a comprehensive understanding of how P-acquisition responses are activated in response to phosphate depletion, as well as localization and function of phosphate regulatory elements and proteins.

1.1.2 *Dissolved Organic Phosphorus as a P-source for Diatoms Under Phosphorus Stress*

The source of phosphorus taken up in diatoms is dependent on environmental conditions and demand of P. The most common soluble form of P taken up by diatoms is orthophosphate anions (P_i) [4]. In addition to P_i, under P-deplete conditions, diatoms can also use dissolved organic phosphorus (or DOP) [4]. However, unlike inorganic phosphorus sources, DOP cannot be assimilated directly in diatoms, and so, diatoms possess other scavenging enzymes like phosphodiesterases and nucleotidases which aid in the utilization of DOP [4,17-18]. Prior research has shown phytoplankton can utilize various forms of DOP to sustain growth in low-phosphate environments [19, 20]. However, the efficiency in utilizing phosphate esters (ATP) and phosphonates (e.g., glyphosate), the primary categories of DOP, may differ among species [25-27].

Myo-inositol phosphates (IP_x, where $x = 1\text{-}6$) constitute a group of organic phosphorus compounds that are widely distributed among organisms and frequently found within marine ecosystems [25,28-29]. Among these, IP_6 (*myo*-inositol-1,2,3,4,5,6-

hexakisphosphate), referred to as phytic acid (PA) or phytate when in a salt form, is the most prevalent (Figure 1-2) [25,28]. Phytic acid holds a significant role in a variety of cellular processes (i.e., nutrient uptake and stress signal transduction) but the bioavailability of PA to phytoplankton is not well characterized [25,30]. The diatom *P. tricornutum* is one species of phytoplankton that can be used to investigate the induction of the phosphorus starvation response and potential sources of phosphorus for phosphate acquisition.
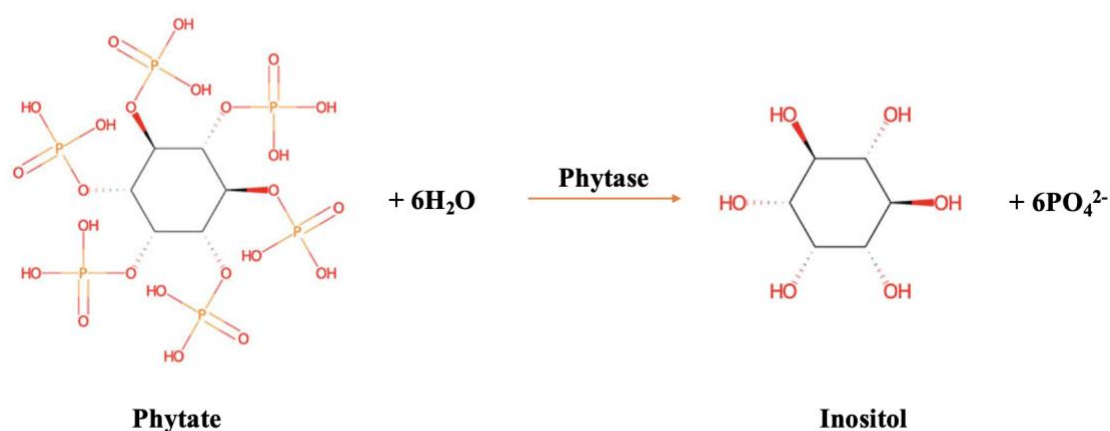
Figure 1-2: Chemical activity of phytase. Phytase will hydrolyze phosphomonoester bonds of phytate molecule to form inositol phosphates.

1.2   *Phaeodactylum tricornutum* as a Model for Synthetic Biology Applications

Phaeodactylum tricornutum (P. tricornutum) is a unicellular, photoautotrophic microalgae and is an attractive model organism for studying phosphorus limitations. *P. tricornutum* exists in three morphotypes, oval, fusiform and triradiate, with the fusiform morphotype being the most common variety (Figure 1-3) [31]. Its genome was the second diatom genome to be fully sequenced, making it one of the most extensively studied diatom organisms to date. Currently, it's genome is predicted to be ~27.4Mb in size and contains over 10 000 predicted genes [32,33]. Interestingly, *P. tricornutum* can metabolize silicates that can be incorporated into its silicified cell wall. However, unlike

other diatoms, it can grow in silicon-free media [27]. In a laboratory setting, the species is highly resilient as it can grow in defined medium, in liquid or on solid media plates. *P. tricornutum* has be shown to dominate and outcompete other algal species in outdoor mass-culture systems, tolerating high pH and low light conditions [34].

In recent years, research communities are changing directions and exploring the organisms using multiple interdisplinary approaches to broaden the understanding of diatom biology, ecology, and biotechnology [34]. With a diverse molecular toolbox, the organism can undergo genetic manipulation such as gene silencing, editing, overexpression and has a variety of plasmid delivery methods established [34-38]. These characteristics make the diatom suitable for applications in synthetic biology.

Synthetic biology is a field of science in which organisms are redesigned to have useful purposes by engineering them to have new abilities [39]. Researchers and industries around the world are using the power of nature to develop novel solutions in medicine,
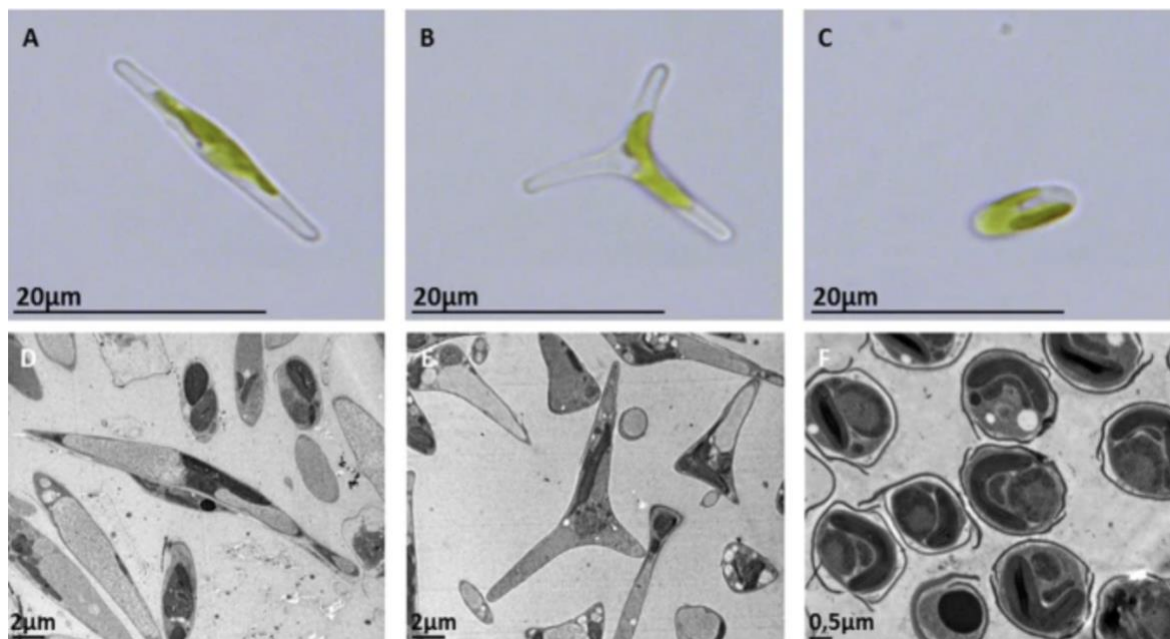
Figure 1-3: Various morphotypes of *P. tricornutum* cells. (A) Fusiform, (B) Triradiate (C) Oval. Image adapted from: Ovide et al., 2018.

manufacturing, and agriculture. Under cellular stress, the diatom has shown to accumulate large amounts of lipids which, for the petroleum industry, has become a great

source for manufacturing biofuels [40]. Furthermore, *P. tricornutum* cells have been genetically engineered to produce pharmaceutically relevant protein (i.e., human IgG antibodies) [35 ,41]. Endogenous, inducible, promoter systems within the organism have also been extensively studied to produce cost-effective recombinant proteins [35,42].

1.2.1   *Endogenous Promoters for Synthetic Biology*

Diatoms have garnered significant interest because of their remarkable ability to thrive in various environmental conditions, likely stemming from their intricate evolutionary history [2,32]. Exploring their metabolic processes is crucial for understanding their adaptive capabilities and is a fundamental step for metabolic manipulation. Decades of extensive research has culminated in the creation of genome engineering techniques essential for such investigations. Diatoms hold immense significance in the realm of industrial biotechnology, owing to their abundance in natural lipids and carotenoids, as well as their capacity to generate recombinant proteins that cater to a wide range of markets [38,42].

Among molecular tools are several endogenous gene promoters for expressing transgenes in diatoms. The most widely used endogenous promoter is the fucoxanthin chlorophyll a/c binding protein (FCP) gene promoters (now referred to has the light-harvesting complex containing fucoxanthin, *Lhcf*) [35, 42]. These promoters are constitutive and capable of driving constant and high levels of transgene expression. LHCF is a part of a family of proteins involved in photosynthesis called light harvesting complexes (LHC). Present in centric and pennate diatoms are other LHC families, LHCR and LHCX [42]. In *P. tricornutum*, there are 17 LHCF genes, 14 LHCR genes and 4 LHCX genes [42]. Although these promoters have been used in a variety of applications, their reliance on light makes them insufficient for examining transgenes in all growth phases. The promoter of the elongation factor 2 (*ef2)* gene, another constitutive promoter, has shown expression levels greater than those seen with *Lchf2* in *P. tricornutum,* however, the diversity of terminators compatible with such promoters remain limited [42].

Inducible promoters have become an attractive source for transgene expression due to their capability of switching expression on and off in an efficient and reversible manner

[35,42]. One of the most widely used is the nitrate reductase (NR) gene promoter [35,42]. The promoter has been used for heterologous expression of human antibodies. Induction of the promoter occurs when provided an exogenous nitrogen source (nitrate) as its sole source and the presence of ammonium ions will lead to its deactivation [35,42]. In a recent study, the promoter has shown residual activity in the absence of nitrate, leading to low level of gene expression.

The discovery of the Highly Abundant Secreted Protein 1 (*HASP1*) and its promoter opened new possibilities in the search for a suitable inducible promoter system to produce a variety of recombinant proteins [35].

1.3 *HASP1*: A Story About Phosphate Regulation, Loss (of Function) and Regulatory Networks

The fluctuations of phosphorus in marine environments has guided much of the discoveries in *P. tricornutum* regarding the species adaptions to nutrient depletion. As previously mentioned, much of what we know about the phosphate starvation response is from transcriptomic studies, assessing the regulation of genes responsible for phosphate acquisition. These studies provide great insight on regulatory networks; however, they fail to determine the localization and regulatory elements involved at the transcriptional level. The *HASP1* promoter system provides novel insights on phosphate regulation under P-deprivation.

1.3.1 *HASP1 Promoter for Expression of Recombinant Proteins*

In 2019, a group of scientists sought out to identify a novel expression system to produce a cost-effective platform in recombinant protein production. Notably, they wanted a system in which proteins were secreted into the culture medium, eliminating the step of intracellular protein purification [35]. Through liquid chromatography tandem mass spectrometry (LC-MS/MS), they identified hundreds of proteins secreted at high levels into the supernatant during stationary phase of the diatom *P. tricornutum*. The Highly Abundant Secreted Protein 1 (*HASP1*) was the most abundant and further investigations were done to compare its function to other commonly used expression systems [35]. The researchers noted the *HASP1* promoter and signal peptide (SP) drove GFP expression

much stronger in all growth phases in comparison to the well-known *fcpA* promoter, which was less active during stationary. This was the beginning of a novel inducible expression system to produce recombinant proteins in *P. tricornutum*.

The COVID-19 pandemic increased the need for rapid serological testing. In our lab, the novel *HASP1* promoter and SP were used to create a novel construct [44]. In conjunction, they were cloned to express the receptor binding domain (RBD) of the SARS-CoV-2 protein. Wildtype *HASP1*-RBD constructs were able to overexpress RBD protein under phosphate-limiting conditions, suggesting the promoter was induced under phosphate-depletion [44]. The novel inducible promoter and SP were determined to be a scalable system for pandemic and other diagnostics. Despite these novel findings, a few questions were left unanswered.

1.3.2   *Thesis Objectives*

The *HASP1* protein is suggested to be phytase-like; when phosphorus levels are abundantly low, the induction of the promoter will increase the number of proteins responsible for the break-down of phytates as a phosphorus source. Within the *HASP1* promoter are 4 predicted phosphate starvation response binding sites (P1BS-like regions). These regions are suggested to play a role in the activation of the phosphate-starvation signaling. The induction of the *HASP1* promoter system under phosphate depletion allows us to investigate the localization and potential regulatory elements that are responsible for the phosphorus starvation response (PSR).

THESIS OBJECTIVE ONE: Identify potential regulatory elements within the *HASP1* promoter responsible for the phosphate-starvation response.

This first objective was to understand how the phosphate starvation response is activated and what role these potential regulatory elements may play in cell signaling. To tackle this objective, 3 different groups of deletions were created in the predicted P1BS-like regions of the promoter. These mutant promoters were cloned into a plasmid containing a reporter gene to determine promoter function through protein expression and secretion. Since the promoter is induced under low phosphate conditions, the 3 different constructs

were grown in various concentrations of phosphate as well as iron. Our previous findings have shown that a decrease in iron can also increase protein secretion into the extracellular space [43].

A fundamental understanding of localization and regulatory elements involved in the promoter will allow future scientist to fine-tune the system and control recombinant protein production. Furthermore, the manipulation of the *HASP1* promoter may uncover a novel system with enhanced function (i.e., protein expression) for synthetic biological applications.

It was hypothesized that if predicted regulatory elements were involved in the activation of the phosphorus starvation response, then deletions in these regions would cause a change in transcriptional regulation compared to wildtype.

THESIS OBJECTIVE TWO: Understanding *HASP1* function and the characterization of other potential regulatory pathways responsible for phosphate acquisition using a knockout strain.

The function of *HASP1* is not well-characterized. The predicted function is a phytase, breaking down phytates as an organic phosphorus source. As previously mentioned, under limited phosphorus conditions, diatoms will use organic sources of phosphorus. Recent studies have shown the diatom *P. tricornutum* can use phytic acid when other sources are unavailable [25].

Given this information, I sought out to determine if *HASP1* proteins do behave as phytases. To do this, a *HASP1* knockout (KO) strain was created and grown alongside wildtype in various concentrations of inorganic and organic phosphorus. Growth was analyzed for a total of 28 days.

It was hypothesized that if *HASP1* was an important phytase, under phosphorus depletion, the knockout strain would have decreased phytase activity compared to wildtype, affecting cellular growth.

Additionally, the *HASP1* KO was grown alongside wildtype cells for 17 days and RNAseq was conducted to assess transcriptional responses as phosphorus depleted in medium. The purpose of this study was to analyze the fundamental importance of *HASP1* regarding other networks that may be responsible for phosphate acquisition under cellular stress.

It was hypothesized that under nutrient depletion, if *HASP1* was an important phytase, it's knockout would cause changes in transcriptional regulation. As a result, there would be an upregulation of genes involved in phosphate regulation.

Chapter 2

2   Characterizing Potential Phosphate Regulatory Motifs in the *HASP1* Promoter

2.1   Introduction

Due to their ease of culturing and diverse molecular tool kits, diatoms remain an appealing model for a wide range of disciplines at low costs [14]. In the field of synthetic biology, diatoms are currently being designed and used to produce biofuels, pharmaceuticals, and cosmetics [34, 44, 45].

To metabolically engineer diatoms for useful purposes, an optimized expression system with an appropriate promoter that can drive transgene expression at high levels is necessary [35, 42, 43, 46]. Inducible promoters are an attractive tool for transgene expression due the specificity it provides in controlling gene expression under certain conditions.

The identification of the novel *HASP1* promoter and signal peptide, brought insight on the possibility of a new expression system and how the phosphate starvation response may be activated. This was further investigated by members in our lab, which led to not only the discovery of a simple, scalable system to produce virus antigens for diagnostics but more insight to promoter induction under phosphate starvation.

We know that diatoms have developed mechanisms that will enhance their longevity in poor environmental conditions. However, we do not know how and what allows for the induction of these genes. The *HASP1* promoter is an ideal candidate to begin answering some of these questions.

The 643 base-pair (bp) promoter is predicted to have approximately 4 phosphate-starvation responsive elements (P1BS-like, Figure 2-1). P1BS elements are regulatory motifs often involved in Pi-signaling (consensus sequence: GNATATNC) [47]. These predictions are based off work conducted by Erdene-Ochir et al., using an *in silico* analysis with PlantPAN2.0 and PlantCARE [35]. Plant PAN2.0 conducts "plant promoter

analysis and reconstructing transcriptional regulatory networks" [48] while PlantCARE looks at cis-acting regulatory elements within plant promoters [49]. Although plants and a diatom are far removed on an evolutionary scale, this provides some framework in determining phosphate regulatory elements.

Here, I worked to determine whether the predicted P1BS-like regions within the *HASP1* promoter are phosphate regulatory motifs required for promoter induction under P-deprivation. By creating multiple deletions within the P1BS-like regions, I tested promoter activity using *HASP1*-eGFP constructs grown under various phosphate and phosphate plus iron conditions. The following work presented provides a basis of understanding phosphate regulation in *P. tricornutum* while developing novel expression systems with the potential of using these systems to create protein factories.

-643   -600
↓   ↓
gga ttg ata gtg aaa acc tta ttc att gtc aga gct taa gcc ggt ctg gtc tat ctt tcc act gtc aaa

cag ctc ttg att gtc gcc cgc gcg aaa ata gta gca cta act gta act tca aaa tac aaa atg ttc tct

                        -480
                        ↓
gtt acc ata cag tga atg taa ctt tcg aat tga cag tat tag tag tcg tat tga cag tga ggc acg ccc

ctc aat gtg cga ggt gga aaa tat acc agc atg aca atg aat ctt gga gat tc ttt gct gtc atc aag

        -360
        ↓
att cac cgc caa atc ttc agg aac cta tca cgt cca cag gcg atg tta att ctt gag tcg tca aaa caa

                                                -240
                                                ↓
agt cct gtc cta cct gta gaa gtt gac agc gag caa ttg tat gca aac ttc tga ctt atg tta taa taa

catt aaa ggt aat taa gta tct tca att tgg cat ttt gtc act gtc agt ccg ttc cga caa tat agg tag att

        -150
        ↓
tgg aat gaa tct ttt cta tgc tgc tgc gaa tct tgt aca cct ttg agg ccg tag att ctg tcc gac gaa

                        -45
                        ↓
gcg ata att att gca aaa tac atg gac tca tta ttt tga ttc gat ttc ttt ttg gta tcc gac tcg aaa aga

            -1
            ↓
tcc atc acg gcg agc

Figure 2-1: The 643bp Highly Abundant Secreted Protein 1 (*HASP1*) promoter in *P. tricornutum*. Nucleotides highlighted in blue represent predicted P1BS-like regions.

2.2   Methods

2.2.1   *Plasmid Design and Construction*

The first goal of my thesis is to assess the function of predicted P1BS-like regions in the *HASP1* promoter. To do this, 3 distinct mutants were constructed. The *HASP1* promoter is predicted to contain 4 P1BS-like regions. The first mutant, MT1-pSS10, contained deletions at P1BS-like regions spanning -417 through -409 and -390 through -383 of the 643bp promoter (Figure 2-1, 2-2B) The second mutant, MT2-pSS10, contained deletions at P1BS-like regions spanning –214 through -206 and -106 through -99 Figure 2-1, 2-2B). The third mutant, MT3-pSS10, contained deletions of all four regions.

Promoter sequences containing mutations were ordered as custom gBlock sequences from Integrated DNA Technologies (IDT, 2023) and cloned into the pSS10 plasmid, replacing the wildtype *HASP1* promoter using the yeast assembly method detailed in Noskov et al., 2012 (Figure 2-2A, Table B-1,2) [50, 51]. All novel constructs contained the mutant *HASP1* promoter and its signal peptide driving eGFP, followed by the *P. tricornutum 40SRPS8* (40S ribosomal protein 8) terminator downstream of the coding region (Figure 2-2A,B). The resultant medium containing yeast colonies were collected, DNA was extracted and transformed into EPI300 *E. coli* cells. DNA from single *E. coli* colonies was extracted using the Monarch® Plasmid Miniprep kit (New England Biolabs, T1010L). To determine correct assembly of plasmids, a restriction enzyme digest was performed, and DNA was sent for analysis by Sanger sequencing at the London Regional Genomics Center.

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

A heat shock transformation was conducted on correct colonies into the pTA-MOB *E. coli* strain for conjugation. pTA-MOB contains the cellular machinery required for bacterial conjugation into *P. tricornutum* cells. The day before conjugation, the resultant colonies were grown in 5mL of LB media containing chloramphenicol (25mg L^{-1}) and gentamycin (20 mg L^{-1}) antibiotics.

Conjugations were performed and adapted from Karas et al. [36]. In short, liquid cultures of *P. tricornutum* (250μL, Culture Collection of Algae and Protozoa CCAP

1055/1) were adjusted to a density of 1.0×10^8 cells mL^{-1} using counts from a hemocytometer. Cells were plated on ½ x L1 1% agar plates containing no antibiotics and grown for four days at 18°C under cool white, fluorescent lights (75 µE m^{-2}s^{-1}) and a photoperiod of 16h light: 8h dark. After 4 days, 1.5mL of L1 media was added to plates and then scraped. The concentration is then adjusted to 5.0×10^8 cells mL^{-1}. *E.coli* cultures containing plasmids for conjugation were grown in 50mL LB media at 37°C to an A$_{600}$ of 0.8-1.0, centrifuged for 10 minutes at 3000xg and resuspended in 500µL SOC media.

To initiate conjugation, 200µL of *P. tricornutum* and *E. coli* cells were combined and plated on ½ x L1 x 5% LB 1% agar plates, incubated at 30°C for 90 minutes and then moved to 18°C and grown for 2 days. After 2 days, plates were scraped using 1.5mL L1 media and 300µL is plated on ½ x L1 1% agar plates containing nourseothricin antibiotic (100mg L^{-1}) and grown for 1-2 weeks at 18°C.

2.2.3 *PCR Screen*

Resultant *P. tricornutum* colonies were screened for the presence of the mutant *HASP1* promoter using the Thermo Scientific Phire Direct Plant PCR Master Mix (Thermo Scientific^{TM}, F160S) kit. To determine the presence of a mutant *HASP1* promoter, PCR was performed using forward primer overlapping 3' end of HIS terminator and 5' end of *HASP1* promoter (DE5241) and reverse primer located at *HASP1* signal peptide (DE5242). The predicted product for MT1 and MT2 is 711bp and for MT3 is 695bp (Figure A-1, Table B-1).

2.2.4 *Sample Preparation*

After determining successful colonies, the colonies were streaked onto ½ x L1 1% agar plates containing nourseothricin antibiotic (100mg L^{-1}) and grown for ~2 weeks at 18°C under cool white, fluorescent light (75 µE m^{-2}s^{-1}) and a photoperiod of 16h light: 8h dark. Streaks were then transferred to liquid cultures containing ½ x L1 media with nourseothricin (100mg L^{-1}) and grown to a concentration of 1.0×10^6 cells mL^{-1} or greater for ~2 weeks at the same condition previously mentioned.

Cultures were counted under a light microscope to determine cell density after 2 weeks. An aliquot from each culture (wildtype *P. tricornutum*, MT-pSS10, pSS10) was pelleted at 3000 x g, 4°C for 10 minutes. The cell pellet was washed twice by resuspension in 10mL of L1 containing no phosphate or iron. Final resuspension was done in L1 media containing no phosphate or iron.

Three replicates were created for each condition by passaging $1.0x10^6$ cells mL^{-1} of the prepared culture into 15mL of L1 media containing no phosphate or iron. For cells containing the mutant promoter or pSS10, phosphate and iron stock solutions were used to adjust cultures to 20% phosphate, 10% phosphate, 5% phosphate, 20% phosphate+iron, 10% phosphate+iron and 5% phosphate+ iron. Additionally, pSS10 and wildtype *P. tricornutum* were supplemented to 100% phosphate and iron. Plasmid-containing cultures were supplemented with nourseothricin (150mg L^{-1}). Wildtype *P. tricornutum* containing no plasmid (NP-WT) was supplemented with chloramphenicol (25mg L^{-1}). Cultures were grown for a total of 28 days. All conditions and constructs used in the experiments are outlined in Table 2-1.

Table 2-1: All cultures and conditions used in this study. All conditions were done in triplicate.

Construct	Conditions
Wildtype *P. tricornutum* *(no plasmid)*	100% phosphate and iron 20, 10, 5% phosphate 20, 10, 5% phosphate + iron
pSS10 (Wildtype *HASP1* promoter)	100% phosphate and iron 20, 10, 5% phosphate 20, 10, 5% phosphate + iron
Mutant 1 (MT1)	20, 10, 5% phosphate 20, 10, 5% phosphate + iron
Mutant 2 (MT2)	20, 10, 5% phosphate 20, 10, 5% phosphate + iron
Mutant 3 (MT3)	20, 10, 5% phosphate 20, 10, 5% phosphate + iron

2.2.5 *Determining Cell Density and eGFP Fluorescence in Supernatant*

Cell density was determined by aliquoting 100 µL of whole cell culture into a clear 96- well plate and measuring absorbance at 670nm (A_{670}) and 750nm (A_{750}) using the Biotek Synergy H1 plate reader every two days. Absorbance values were fitted to a standard curve to determine cells mL^{-1}.

To determine eGFP fluorescence in supernatant, 700µL of whole cell cultures were pelleted at 18 000 x g for 15 minutes. 200µL of supernatant was aliquoted into clear bottom 96-well plates. Fluorescence was measured at an excitation of 475nm and emission of 515nm. Wildtype (auto)fluorescence was subtracted from final reads.

2.2.6 *Western Blots of Supernatant and Cell Lysate*

Samples for western blots were taken on days 0,4,8,16, 24 and 28. To prepare samples for western blotting, 700µL of whole cell cultures were pelleted down at 18 000 x g for 15 minutes. For supernatant: 70µL of supernatant was added to 35µL of 3xSDS dye (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] β-mercaptoethanol). For cell lysates, pellets were resuspended in 50µL of 3xSDS dye. All samples were boil-lysed at 95°C for 10 minutes.

Samples were then loaded on 15% SDS-PAGE gels. Volume to be loaded was normalized to cells mL^{-1} determined by absorbance readings. 5µL of prepared (5ng $µL^{-1}$ in 3xSDS dye) commercially purchased recombinant GFP protein (abcam, ab84191) was used as a positive control. Gels were transferred to polyvinylidene difluoride (PVDF) membranes using the Trans-Blot Turbo Transfer System (BioRad, Hercules, CA, USA). Following transfer, membranes were incubated for 1 hour at room temperature in blocking buffer solution (3% bovine serum albumin (BSA), 0.1% Tween-20, 1$\times$ TBS) and then incubated overnight at 4°C in anti-eGFP primary antibody (Invitrogen, A-6455) at a final dilution of 1:2500. In the morning, membranes were washed for 10 minutes(x3) in wash solution (1% BSA, 0.1% Tween-20, 1$\times$ TBS), followed by a 2-hour incubation in a 1:5000 dilution (in wash solution) of anti-rabbit horseradish peroxidase-linked secondary antibody (Sigma, GENA9340). After, membranes were washed in 1xTBS with

0.1% Tween-20 for 10 minutes(x3). The final wash was completed in 1x TBS for 10 minutes. To image blots, Clarity Western ECL Substrate (BioRad, #1705061) was applied following manufacturers' instructions and then imaged using the ChemiDoc^{TM} Imaging System (BioRad, 12003153). Western blots were quantified using ImageLab software.

2.2.7 *Statistical Analysis*

Microsoft Excel (version 16.71) was used for all statistical analysis. All data were presented as mean ± standard error of mean (SEM). For experiments comparing the average cell density of mutant and pSS10 cells, the average of 3 biological replicates for each condition (mutant and pSS10 cells) was calculated. Each mutant promoter was compared to pSS10 on days 8, 16 and 28 using an unpaired one-tailed Student's *t* test. A p value of ≤ 0.05 was considered statistically significant. The same analysis was done for experiments comparing average eGFP fluorescence in the supernatant between mutant and pSS10 cells.

2.3 Results

2.3.1 *Mutations within the HASP1 promoter do not affect cellular growth.*

The first step was to determine if the new mutations within the *HASP1* promoter affect cellular growth. Cells containing mutant *HASP1* promoters and wildtype *HASP1* (pSS10) were subject to 6 different conditions while wildtype *P. tricornutum* cells with no plasmid (NP-WT) were supplemented with 100% phosphate and iron (Figure 2-2C). *P. tricornutum* cells in our lab are routinely cultured in L1 media that contains ten times the amount of phosphate versus other formulations (3.62×10^{-4} M versus 3.62×10^{-5} M) as it promotes higher cell densities. Thus, using an appropriate amount of phosphate that would sustain cellular growth and induce promoter function was necessary. Previous work in our lab has shown 5% of normal phosphate levels (18µM) allows for cellular growth and promoter induction [43]. Additionally, other vital components such as nitrogen and iron were tested. 5% iron in conjunction with phosphate also showed promoter induction [43]. Based on these results, 5% phosphate (and iron) were chosen as the "lower" bounds and 20% as the arbitrary "higher" concentration. Lastly, similar growth rates between the mutants and wildtype are important as other downstream results could be due to other stressors which are limited by growth.
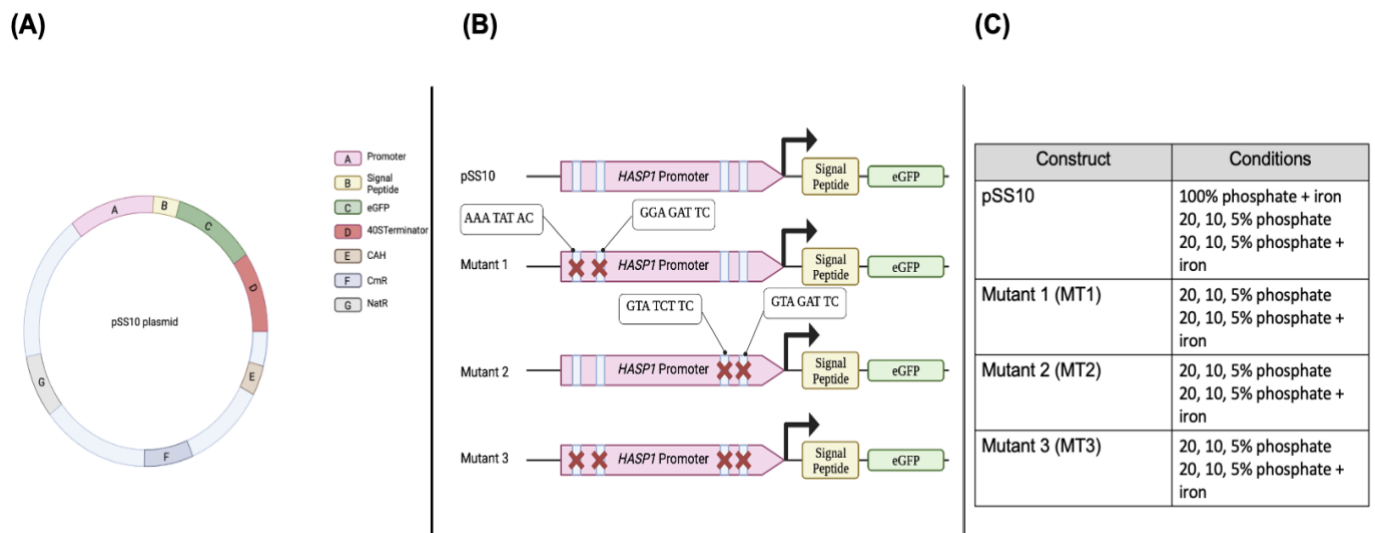
Figure 2-2: All plasmid-containing constructs and conditions used in this study. (A) pSS10 plasmid. CAH denotes CEN6-ARSH4-HIS3, CmR denotes chloramphenicol resistance and NatR denotes nourseothricin resistance. (B) Four *HASP1* promoter constructs driving eGFP. Blue rectangles denote predicted P1BS regions within the promoter. Red crosses denote deletions. (C) Table shows the different concentrations of phosphorous and iron each construct was grown in. Each condition was done in triplicate.

To measure cellular growth, absorbance readings were taken every 2 days (for a total of 28) using a plate reader and then cells mL^{-1} were determined by fitting absorbance readings to a standard curve.

In 10% phosphate conditions, there was a significant increase in average cell density compared to pSS10 on day 28 (Figure 2-3A, $p \leq 0.05$). These findings were also seen in 10% phosphate and iron conditions (Figure 2-3B, $p \leq 0.05$). Furthermore, in 20% phosphate and iron conditions, MT2 had a significant decrease in cell density compared to pSS10 on day 28 (Figure 3.2B, $p < 0.05$). No other significant changes were observed. This suggests that mutations within the *HASP1* promoter may affect cellular growth in various concentrations of phosphate and iron.
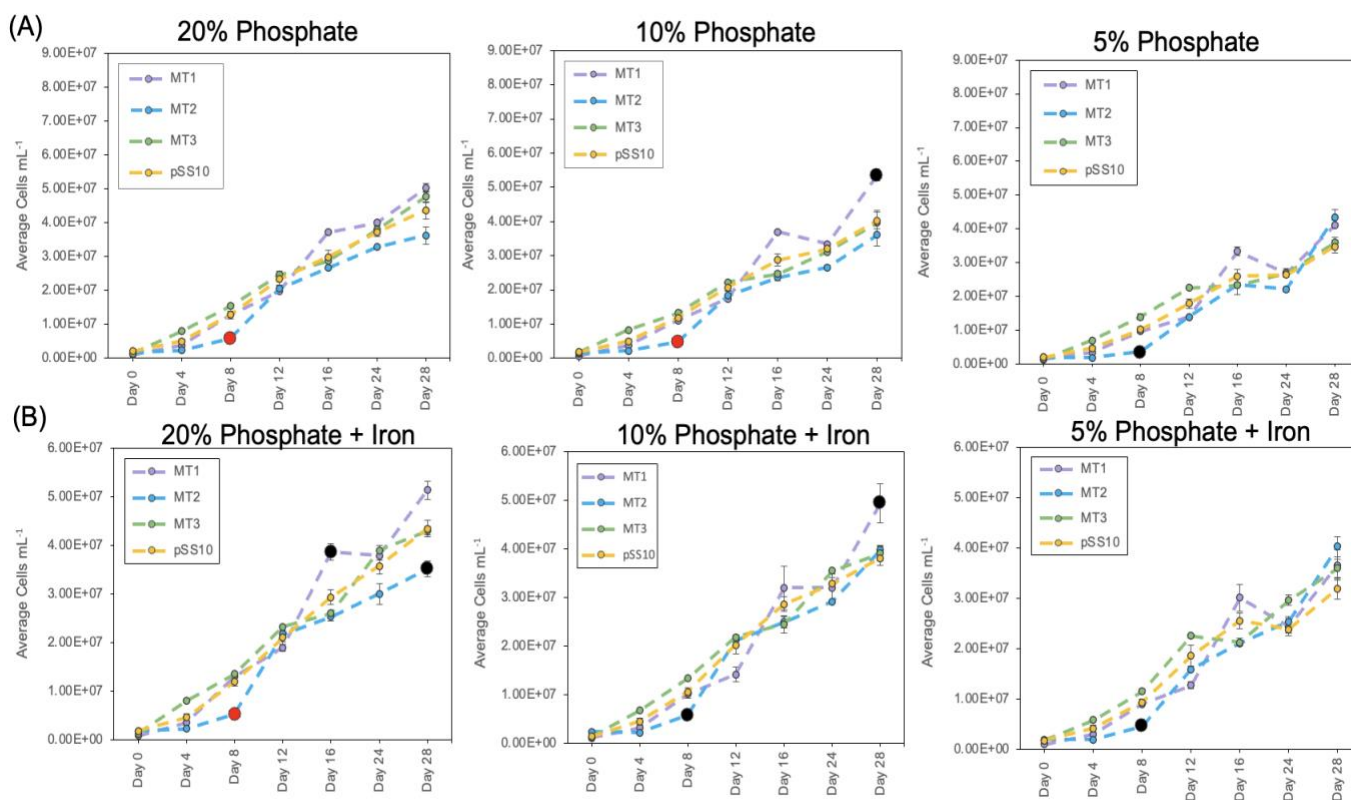
Figure 2-3: Cell Density of cultures containing mutant promoters compared to pSS10 under nutrient depletion for 28 days. (A) Cell density of cultures grown in 20%, 10% and 5% phosphate (left to right). (B) Cell density of cultures grown in 20%, 10% and 5% phosphate and iron (left to right). Absorbance was taken at 670nm. Each point denotes mean cells mL^{-1} ± SEM of three biological replicates. An unpaired two-sample *t* test was computed for each mutant promoter compared to pSS10 on days 8, 16 and 28. Black ($p \leq 0.05$) and red points ($p \leq 0.01$) denote statistically significant increase (or decrease) in cell density compared to pSS10 cells.

2.3.2 *eGFP Secretion in MT1 and MT3 increases over time and in low phosphorus and iron conditions.*

The *HASP1* promoter and signal peptide work to efficiently secrete protein into the supernatant. Previous research had shown eGFP secretion in low phosphate and iron conditions [43]. It was predicted that if the deleted regions play a role in activating the phosphate starvation response, promoter activity would be deregulated in response to phosphate depletion (i.e., a decrease or an increase in protein secretion).

eGFP fluorescence in the supernatant was significantly greater in MT1 and MT3 on day 28 compared to pSS10 (WT) and MT2 in all conditions (Figure 2-4A, $p \leq 0.05$). This suggests MT1 and MT3 can express and secrete eGFP to a greater extent compared to MT2 and pSS10. Additionally, as phosphate (and iron) depleted over time, MT1 and MT3 eGFP fluorescence in the supernatant increased. The increase in promoter activity over time in the two mutants is consistent with previous research: a decrease in phosphate results in promoter induction [43].
Interestingly, this is not seen in MT2. MT2 decreased in protein secretion compared to the other mutants and pSS10 (Figure 2-4). Furthermore, in 10% concentrations of phosphate, eGFP secretion in the supernatant of MT2 was significantly lower than pSS10 on day 28 ($p \leq 0.05$). This is also the case in 5% concentrations of phosphate and iron, there was a significant decrease in eGFP secretion in the supernatant of MT2 compared to pSS10 by the end of the experiment (Figure 2-4B, $p \leq 0.05$).
The decrease in protein secretion in MT2 suggests that the deletion made within the promoter system may have decreased the rate at which transcription was initiated and as result impaired expression of eGFP. Overall, these results suggest changes in transcriptional regulation.
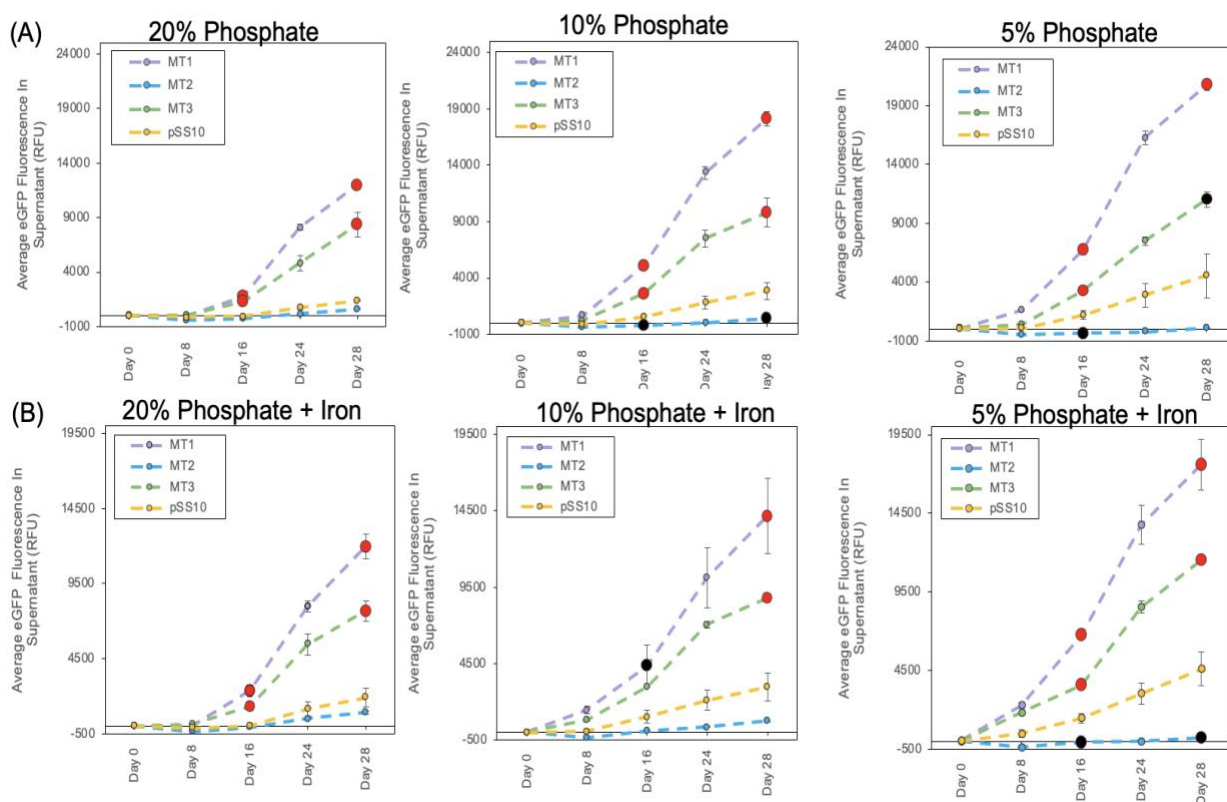
Figure 2-4: Average eGFP Fluorescence in Supernatant (RFU) of constructs containing mutant promoters compared to pSS10 under nutrient depletion for 28 days. (A) Fluorescence in supernatant of constructs grown in 20%, 10% and 5% phosphate. (B) Fluorescence in supernatant of constructs grown in 20%, 10%, and 5% phosphate and iron. Fluorescence was measured at an excitation of 475nm and emission of 515nm. Each point denotes mean eGFP fluorescence in supernatant $\pm$ SEM of three biological replicates. An unpaired two-sample t test was computed for each mutant promoter compared to pSS10 on days 8, 16 and 28. Black ($p \leq 0.05$) and red points ($p \leq 0.01$) denote a statistically significant increase (or decrease) in eGFP fluorescence in the supernatant compared to pSS10 cells.

2.3.3   *Relative eGFP Concentration in Cell Lysate Increases in Mutant Promoters.*

To determine eGFP concentration in cells, cultures were pelleted and the supernatant and cell lysates were resuspended in 3xSDS dye. Following resuspension, samples were boil-lysed for 10 minutes and resolved onto a 15% SDS-PAGE gel for western blot analysis. A commercially purchased recombinant eGFP was used as a positive control.

Analyzing relative eGFP concentrations within the cell was another method for determining protein expression in case the secretory signal was lost and no fluorescence could be detected in the supernatant.

Mutant 1

Induction of the MT1 promoter occurred sooner than pSS10 in all conditions (Figure 2-5,6). On day 16, eGFP was expressed in MT1 and pSS10, however the relative eGFP concentration in the mutant was greater (Figure 2-5). Ratio of eGFP expression of MT1:pSS10 showed the same trend: eGFP expression in MT1 was greater than pSS10. Although the ratio plots show a decrease in expression, all ratios of MT:pSS10 are >1. This is due to the lack of expression on day 8 in pSS10, resulting in a large ratio of MT:pSS10 and then a smaller ratio once pSS10 is expressed.

In phosphate and iron conditions, eGFP expression in MT1 occurred sooner than pSS10. By day 16, eGFP expression was seen in both pSS10 and MT1 (Figure 2-6). Additionally, the ratio of eGFP expression of MT1:pSS10 showed that MT1 expression was greater throughout the 28-day study.

Overall, these results suggest the deletions within mutant 1 have caused changes in transcriptional regulation as MT1 was able to drive eGFP expression more than pSS10 in the same conditions.
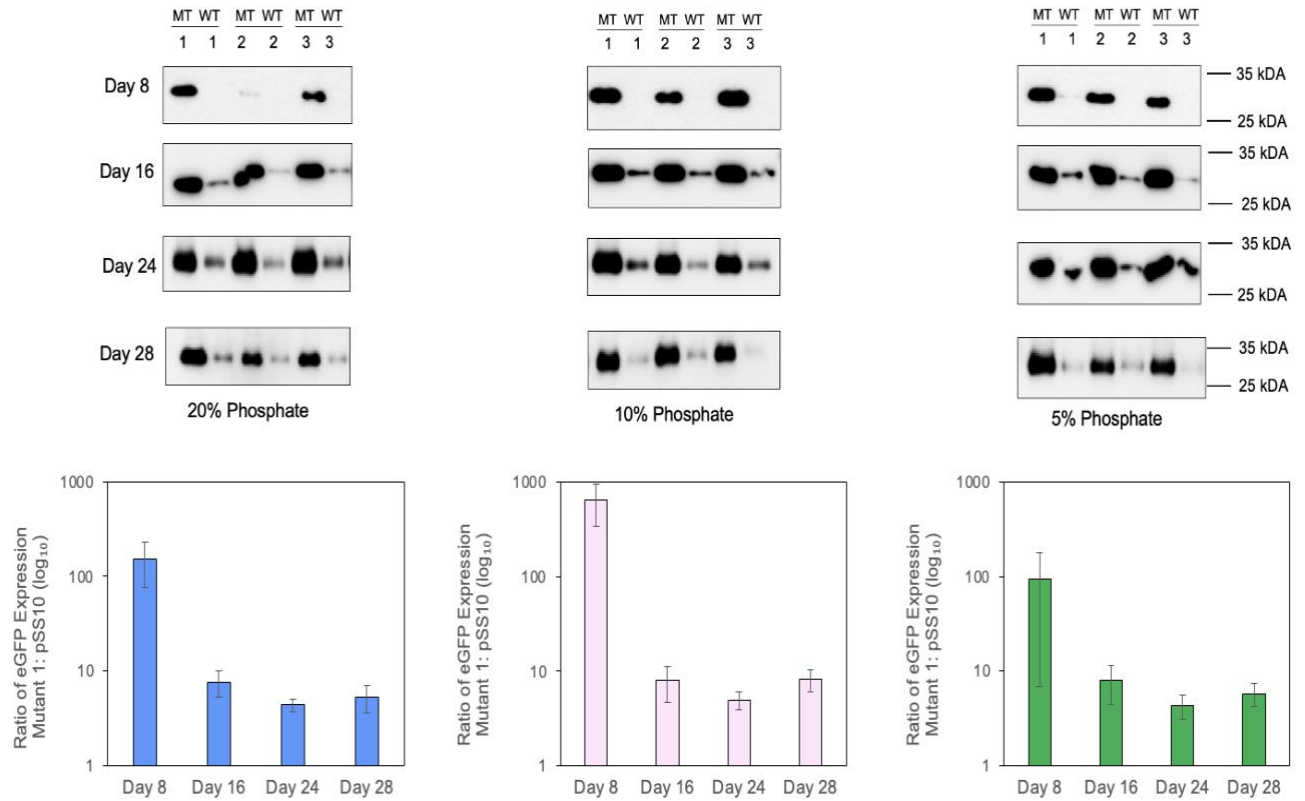
Figure 2-5: Western blots comparing relative eGFP expression in MT1 and pSS10 under different phosphate conditions (20-, 10-, 5%) for a duration of 28 days. Numbers above blots denote experimental replicates. Plots show ratio of eGFP expression in mutant 1 and pSS10, where a ratio of 1 = no difference, >1 is an increase in eGFP expression in MT1 and <1 is a decrease in eGFP expression in MT1. Error bars denote standard error of MT1: pSS10. Y-axis is a logarithmic scale (log10).
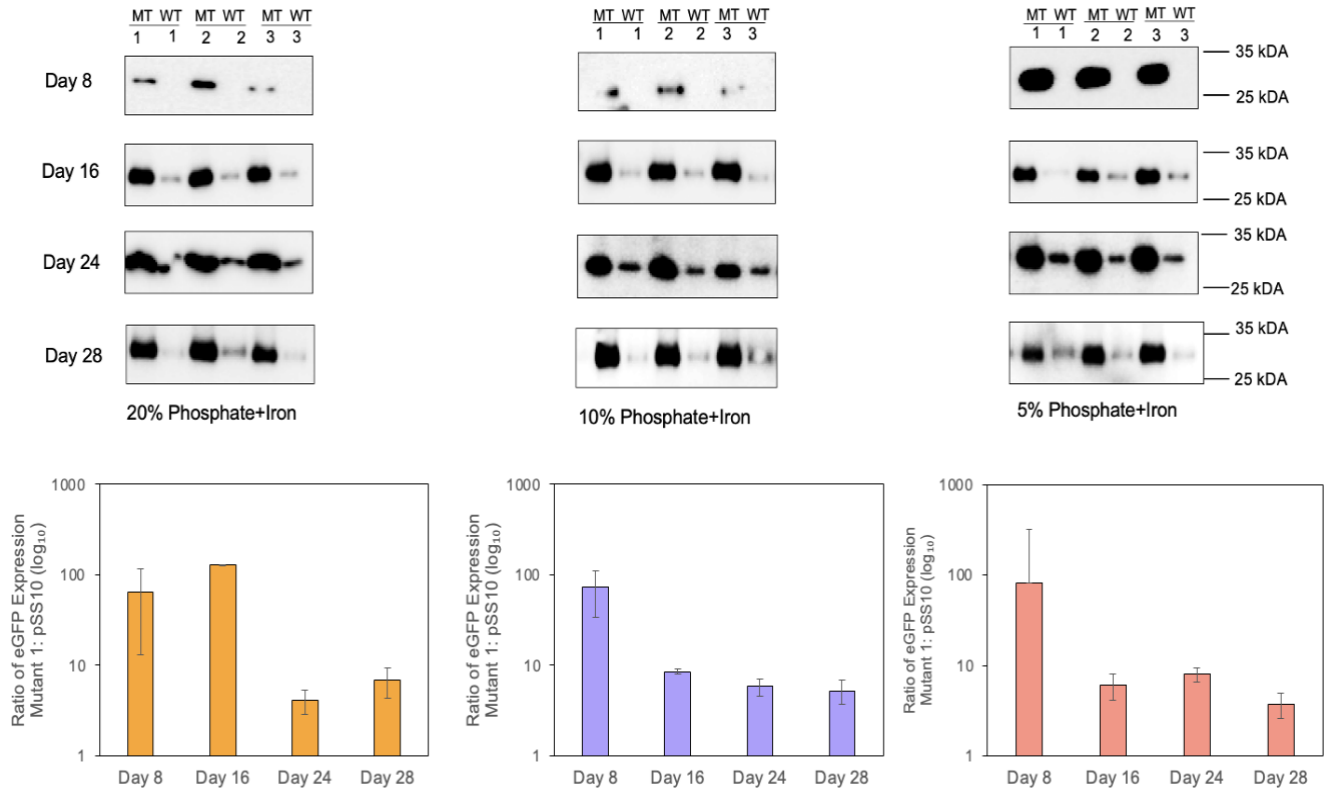
Figure 2-6: Western blots comparing relative eGFP expression in MT1 and pSS10 under different phosphate and iron conditions (20-, 10-, 5%) for a duration of 28 days. Numbers above blots denote experimental replicates. Plots show ratio of eGFP expression in mutant 1 and pSS10, where a ratio of 1 = no difference, >1 is an increase in eGFP expression in MT1 and <1 is a decrease in eGFP expression in MT1. Error bars denote standard error of MT1:pSS10. Y-axis is a logarithmic scale (log10).

Mutant 2

In phosphate and phosphate+iron conditions, the mutant and wildtype promoters were induced on day 8 except for in 20% conditions (Figure 2-7). In the phosphate only condition, eGFP was expressed by day 16 in 20% (Figure 2-7). Looking at the ratio of eGFP expression between mutant2:pSS10, the expression of eGFP in the mutant was greater than wildtype during the initial timepoints of the experiment for 10% and 5% phosphate conditions. Expression became relatively equivalent by day 28. However, in 20% phosphate, the ratio of eGFP expression increased in mutant 2 compared to pSS10.
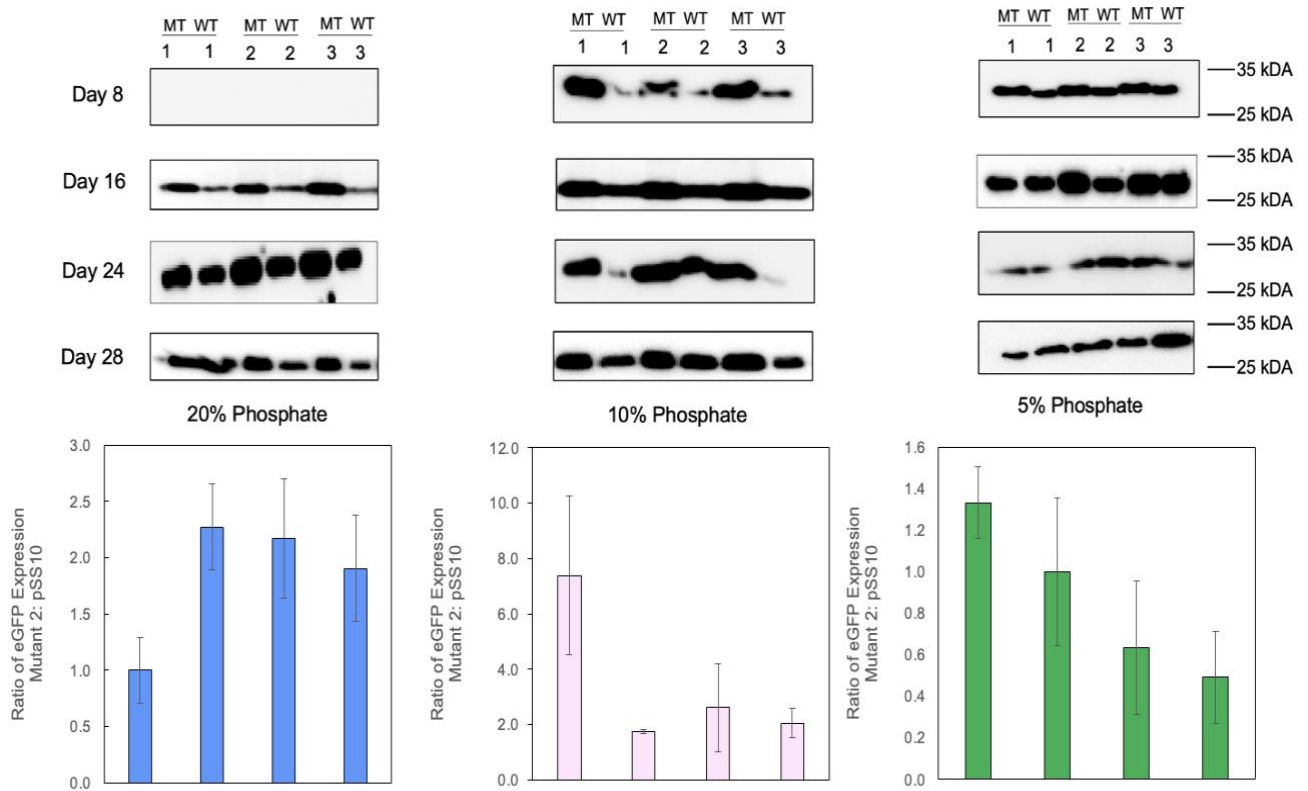
Figure 2-7: Western blots comparing relative eGFP expression in MT2 and pSS10 under different phosphate conditions (20-, 10-, 5%) for a duration of 28 days. Numbers above blots denote experimental replicates. Plots show ratio of eGFP expression in mutant 2 and pSS10, where a ratio of 1 = no difference, >1 is an increase in eGFP expression in MT2 and <1 is a decrease in eGFP expression in MT2.

In phosphate and iron conditions, eGFP expression in MT2 and wildtype occurred in 10% and 5% conditions on day 8 (Figure 2-8). In 20%, eGFP expression occurred in MT2 and wildtype by day 24. Based on the eGFP expression ratio of MT2:pSS10, in 10% and 5% phosphate and iron, eGFP expression was similar throughout the 28-days. In 20% phosphate and iron, eGFP expression increased over the course of 28-days.

Overall, the MT2 promoter could express eGFP sooner than pSS10. However, overtime, the amount of eGFP was similar. These results suggest the deletions within mutant 2 have

caused changes in transcriptional regulation as MT2 was able to drive expression sooner in lower concentrations of phosphate and iron than pSS10 in the same conditions.
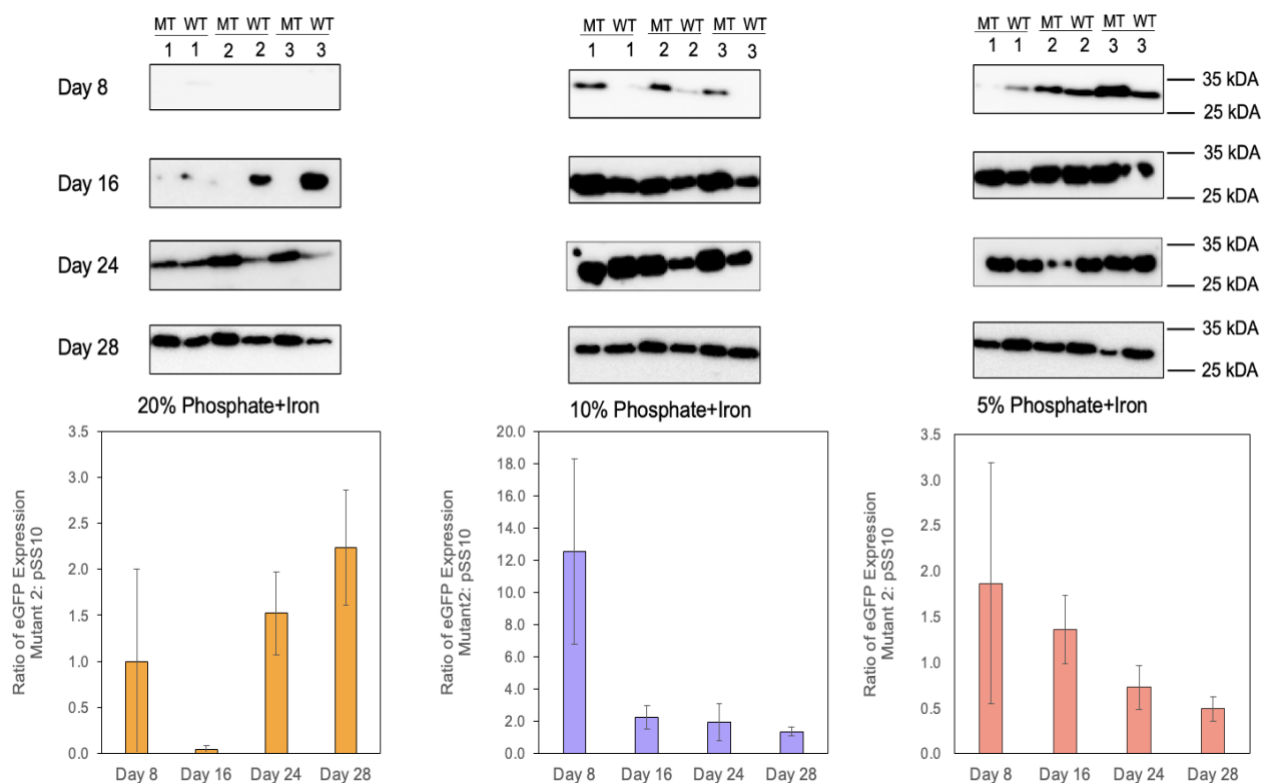
Figure 2-8: Western blots comparing relative eGFP expression in MT2 and pSS10 under different phosphate and iron conditions (20-, 10-, 5%) for a duration of 28 days. Numbers above blots denote experimental replicates. Plots show ratio of eGFP expression in mutant 2 and pSS10, where a ratio of 1 = no difference, >1 is an increase in eGFP expression in MT2 and <1 is a decrease in eGFP expression in MT2.

Mutant 3

In the phosphate-only condition, induction of the mutant 3 promoter occurred at the same time as pSS10 on day 16, except in 5% phosphate, where both promoters are induced on day 8 (Figure 2-9). eGFP expression ratios of MT3:pSS10 demonstrate the MT3 promoter can express eGFP more than pSS10 over time.

In phosphate and iron conditions, the mutant 3 promoter was induced sooner than wildtype on day 8 and by day 16, both promoters were induced (Figure 2-10). The eGFP

expression ratios of MT3:pSS10 in phosphate and iron conditions show that over time, the amount of eGFP expressed is similar in 10% and 5% phosphate and iron conditions.

Overall, these results suggest the deletions within the mutant 3 promoter have caused changes in transcriptional regulation and can drive eGFP expression sooner than wildtype in the same conditions.
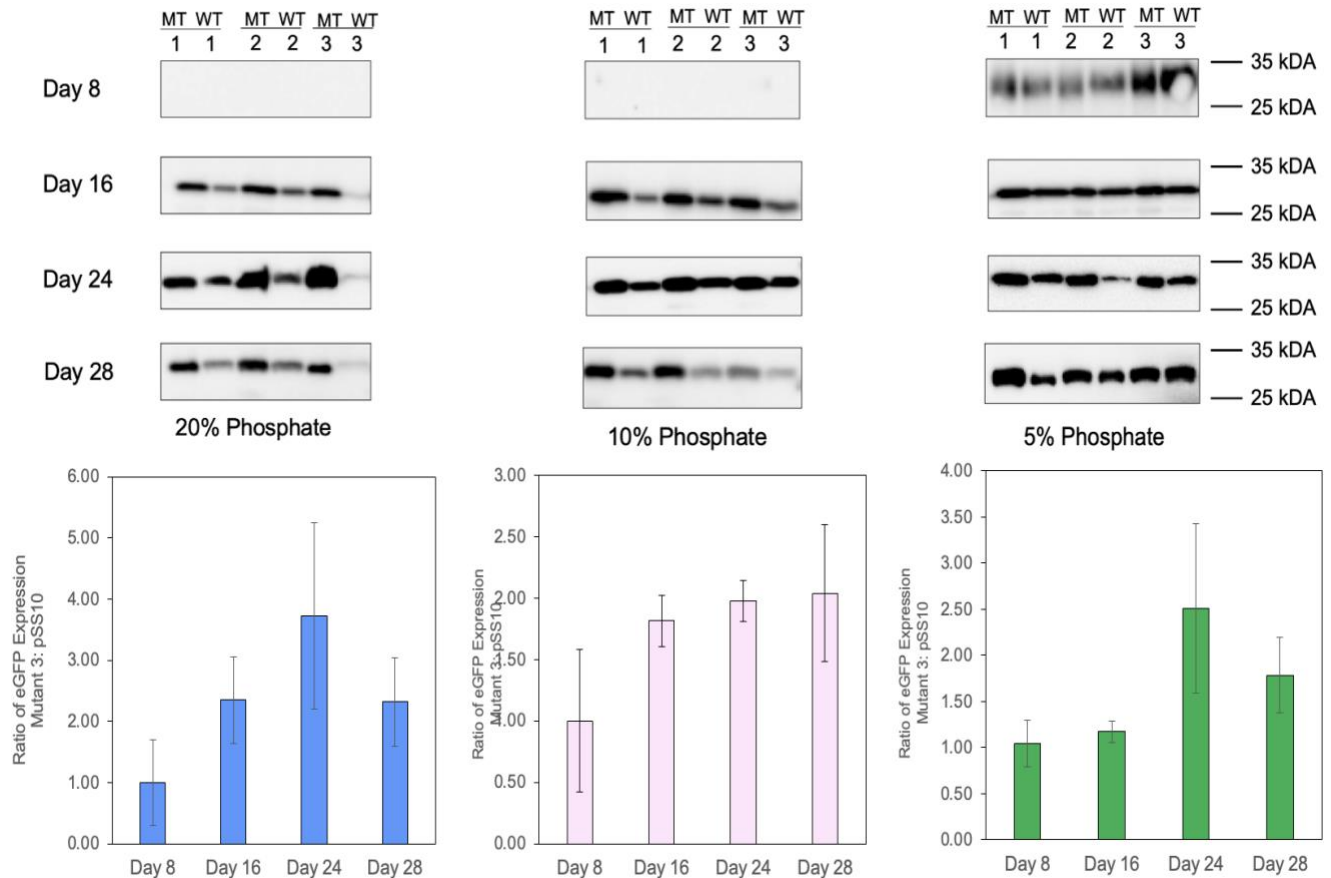
Figure 2-9: Western blots comparing relative eGFP expression in MT3 and pSS10 under different phosphate conditions (20-, 10-, 5%) for a duration of 28 days. Numbers above blots denote experimental replicates. Plots show ratio of eGFP expression in mutant 3 and pSS10, where a ratio of 1 = no difference, >1 is an increase in eGFP expression in MT3 and <1 is a decrease in eGFP expression in MT3.
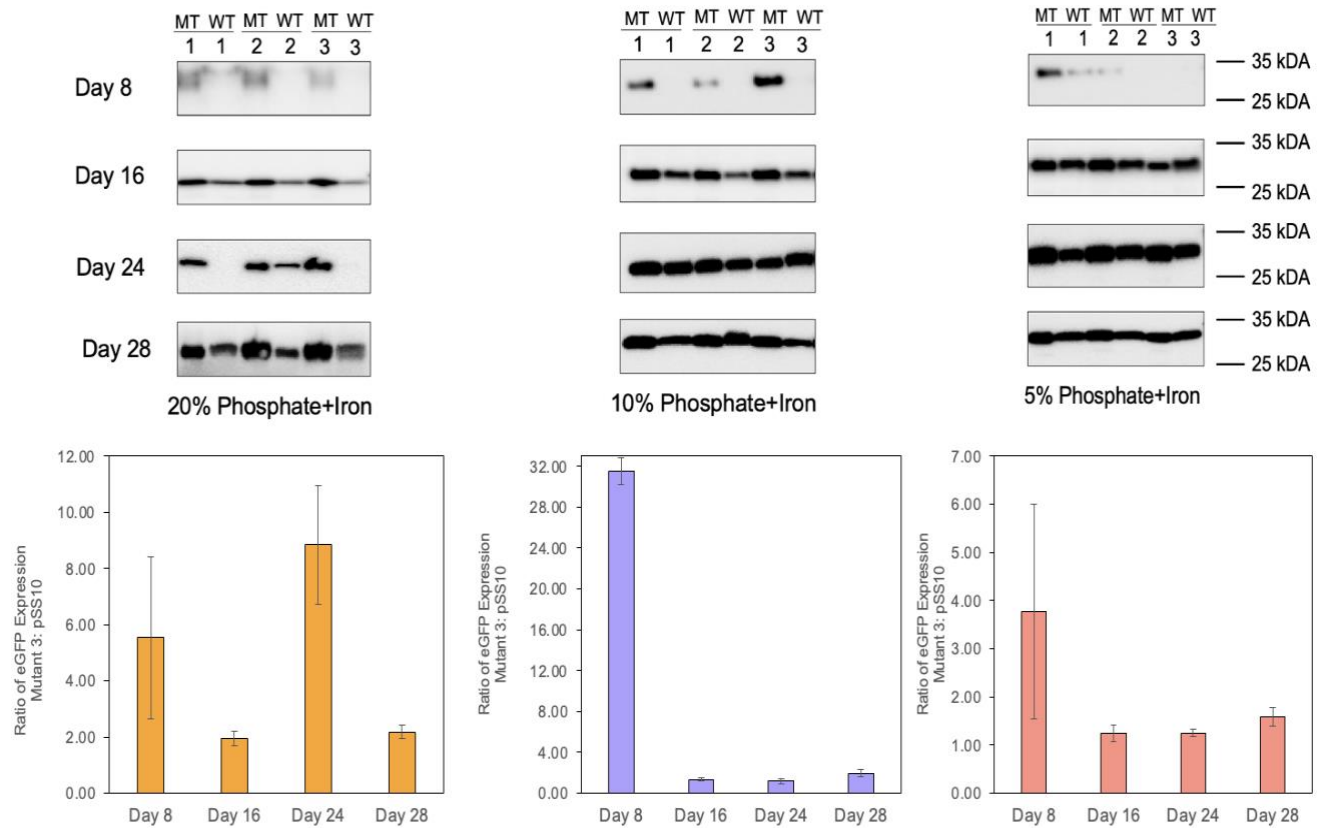
Figure 2-10: Western blots comparing relative eGFP expression in MT3 and pSS10 under different phosphate and iron conditions (20-, 10-, 5%) for a duration of 28 days. Numbers above blots denote experimental replicates. Plots show ratio of eGFP expression in mutant 3 and pSS10, where a ratio of 1 = no difference, >1 is an increase in eGFP expression in MT3 and <1 is a decrease in eGFP expression in MT3.

MT1 v MT2 v MT3

Each mutant promoter was individually tested against the wildtype. Although this information provides us a comparative breakdown of how the deletions impacted eGFP expression in comparison to the wildtype, it was also important to compare changes in eGFP expression for each promoter alongside one another.

The three mutant promoters and wildtype (on day 28) were resolved on a 15% SDS-gel and a western blot was completed. The same number of cells were loaded (1.0×10^6 cells mL^{-1}) to provide the most accurate comparison.

In 10% phosphate, MT1 and MT3 had greater eGFP expression compared to wildtype and MT2 (Figure 2-11A). The amount of eGFP present in mutant 2 was similar to wildtype except in replicate 1, where there was more eGFP expression.

Panel B shows a slightly different trend. Similar eGFP expression was seen in mutants 1 and 3 compared to wildtype, however mutant 2 showed less eGFP expression compared to the other three promoters (Figure 2-11B). This suggests that there have been changes in transcriptional regulation and perhaps changes in iron regulation in mutant 2.
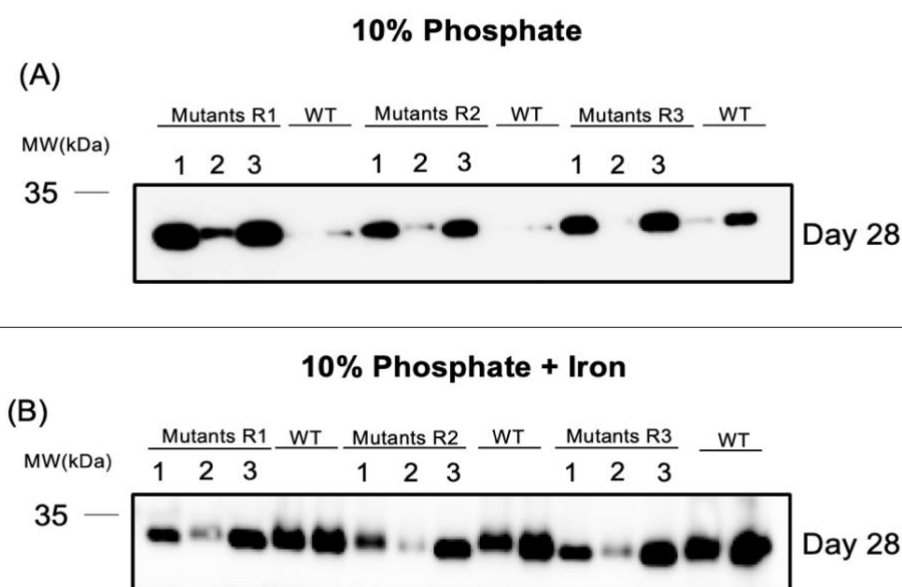
Figure 2-11: Western blots comparing mutant promoters to wildtype on day 28.
(A) Comparison of mutants to wildtype in 10% phosphate. (B) Comparison of mutants to wildtype in 10% phosphate and iron. 1.0×10^6 cells mL^{-1} loaded in each lane.

2.4   Discussion

Phosphorus is an essential element for all life forms. In marine environments phosphate levels are often fluctuating which effects the physiology and ecology of many aquatic organisms, specifically diatoms. As a response, diatoms will regulate the expression of genes responsible for phosphate acquisition.

The *HASP1* promoter system was predicted to contain 4 P1BS-like regions which may be involved in the activation of the phosphate starvation response. To understand how the promoter is induced under phosphate starvation, I created 3 specific mutations within the P1BS-like regions of the promoter and tested their function in response to phosphate starvation by expressing eGFP under the regulation of the promoter and signal peptide.

Cultures were grown under 6 different conditions for a duration of 28 days where samples were taken to assess growth and promoter function.

Compared to wildtype pSS10, all mutant promoters showed similar growth. In some conditions the mutant promoters had significantly higher cell densities ($p \leq 0.05$) however, such changes were seen towards the end of the study. Overall, these results suggests that mutations within the promoters did not affect growth (Figure 2-3).

Relative eGFP fluorescence in the supernatant provided information about eGFP expression under the regulation of the mutant promoters. Under low phosphorus conditions, the *HASP1* promoter should be induced, resulting in the transcription of its signal peptide and eGFP resulting in protein secretion so long as the mutations do not impact promoter induction. In mutants 1 and 3, eGFP expression was significantly greater than wildtype in all conditions ($p \leq 0.05$), which suggests that there have been changes in transcriptional regulation which allowed for the increase in protein secretion (Figure 2-4). Alternatively, the mutant 2 promoter showed a decrease in eGFP expression compared to wildtype. In fact, the amount of eGFP being secreted was the same as background fluorescence seen in NP-WT. A possible explanation for the decrease in protein secretion could be the nature of where the deletions were made in mutant 2 and as a result impacted initiation of transcription and protein secretion. At the time of publication, the

HASP1 promoter's transcriptional start site (TSS) was only a prediction based on *in silico* analysis of the promoter. A rapid amplification of cDNA ends (5'RACE) was conducted in our lab on the *HASP1* locus to determine the potential TSS location. This method is often used to determine unknown sequences at the 5' (or 3') end of mRNA by converting mRNA to cDNA using reverse transcription. This information is important as it could provide insight on why eGFP expression and secretion may have been limited in MT2 conditions (i.e., the deletions may have impacted the TSS). Products from the 5'RACE were resolved on a 1% agarose gel (Figure 2-12). Two colonies from the band at ~400bp were sent for Sangar sequencing. Sequencing results showed that alignment occurred upstream of the predicted transcriptional start site and initiator sequence (Figure 2-13), suggesting that the TSS was further upstream. Unfortunately, due to time constraints the band at ~500bp could not be sequenced. However, these results indicated that there is a possibility that the TSS is further upstream, which would suggest deletions created in MT2 would be deletions in the TSS, possibly impacting protein expression and secretion. It is important to note that MT3 had the same deletions as MT2, which brings up the question of why was MT3 expression increased and not MT2? One possibility is that the 4 deletions truncated promoter regulation (i.e., no repressors bound to regulatory sites) and as a result, the promoter was no longer responsive to phosphate concentrations. Subsequent tests with MT3 should be completed.
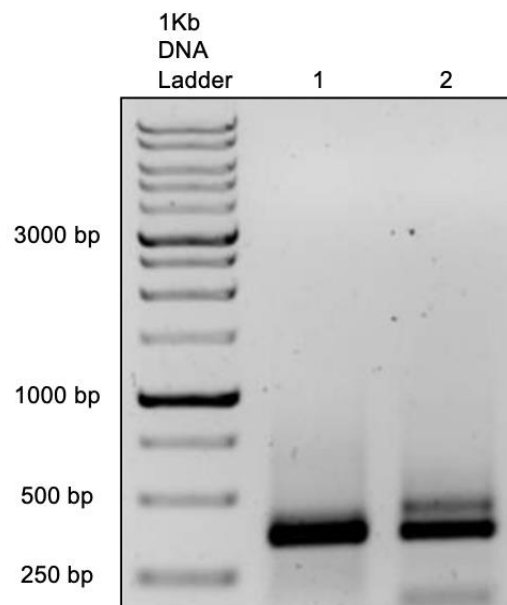
Figure 2-12: Nested PCR of 5'RACE for HASP1 transcriptional start site. Lane 1 contains wildtype P. tricornutum cells from Day 0 of RNAseq experiment conducted in chapter 3. Lane 2 contains wildtype P.tricornutum cells from Day 4 of RNAseq experiment conducted in chapter 3.
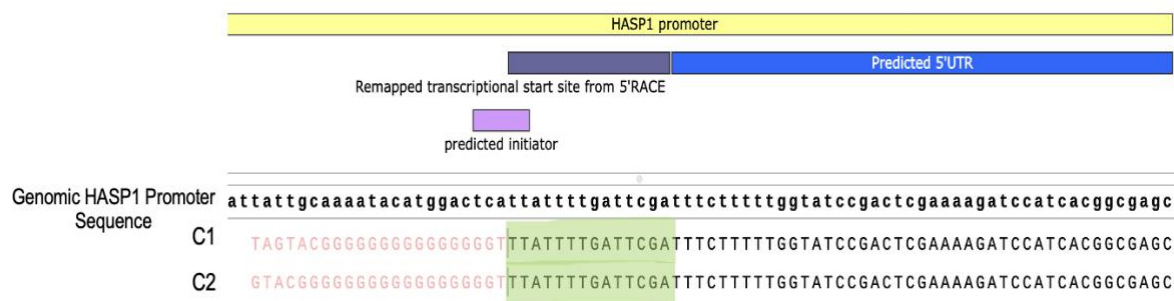
Figure 2-13: Sequencing results from HASP1 promoter 5'RACE experiment. Genomic HASP1 promoter sequence is labelled at the top. C1 and C2 are two clones that were sent for Sangar sequencing of HASP1 promoter. Red nucleotides denote part of pJet vector used for blunt cloning in 5'RACE which do not align to HASP1 locus. Predicted transcriptional start site and initiator sequence from Erdene-Ochir et al., 2019 is shown in blue and purple, respectively. Sequences highlighted in green denote remapped transcriptional start site from 5'RACE experiment.

The western blots of cell lysates comparing the expression of mutant 2 to the other mutant promoters and wildtype show that although expression is limited compared to the other mutant promoters, it is still functional. Based on where one of the mutant 2 deletions are located (-106 through -99) from the predicted 5'untranslated region (5'UTR) and ~40bp upstream of the initiator sequence, this may have impacted the initiation of transcription and downstream secretion (Figure 2-14). Furthermore, the *in silico* analysis done by Erdene-Ochir et al., did not provide any information about other eukaryotic promoter elements, for example, a TATA-box. I used an open online software called YAPP (http://www.bioinformatics.org/yapp/cgi-bin/yapp.cgi) to try and determine other regulatory elements. Although there are many alternative websites that may be used to assess eukaryotic promoter elements, many of them required the user to pick a specific species to align the sequence to. YAPP looked for general trends in the sequence provided. Based on the *HASP1* promoter sequence, there was only 1 potential TATA-box (Figure 2-15). A TATA-box can be found to be several 100bp's upstream of a transcriptional start site. Thus, there is a possibility that the TATA-box is located at this

position and the original predicted TSS is located -37 (Figure 2-14). Additionally, if the TATA-box was located at this site, deletion of P1BS #4 in MT2 could impact the initiation of transcription and protein expression (Figure 2-14,15). It is important to note that not all eukaryotic promoters contain a TATA-box and that the *HASP1* promoter may be one of these promoters.
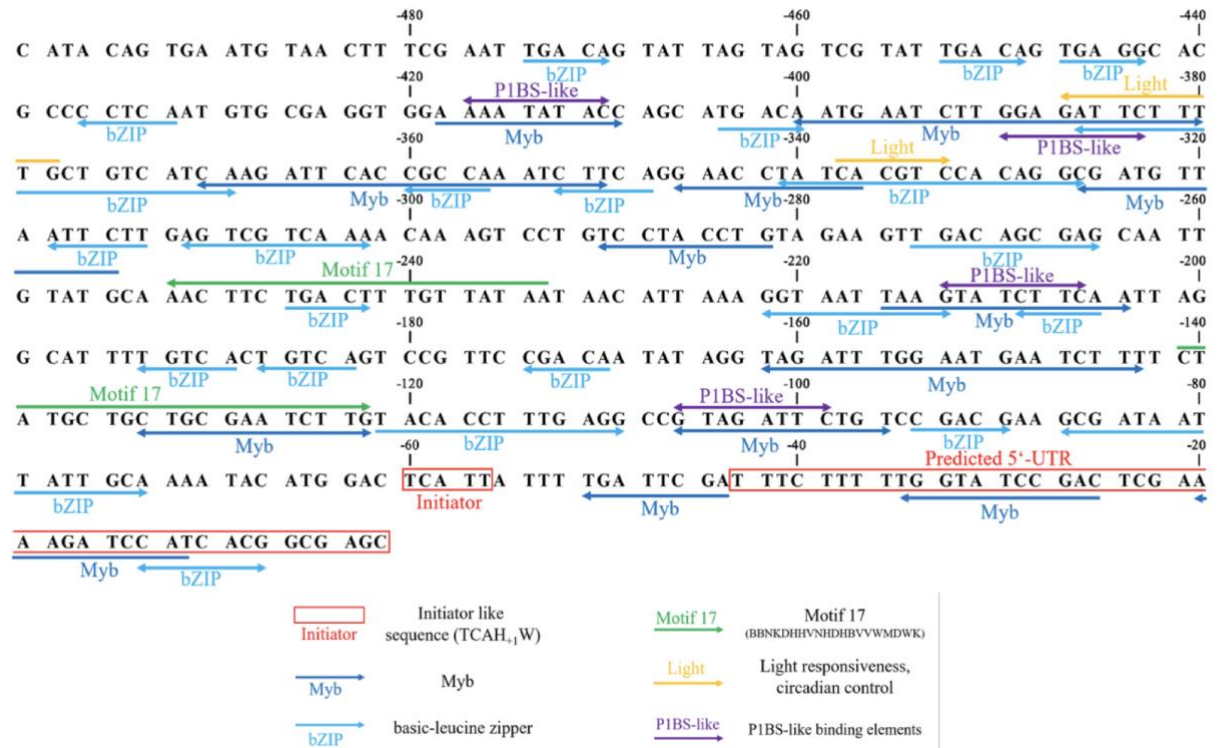
Figure 2-14: Initial prediction of the *HASP1* promoter sequence. Figure obtained from Erdene-Ochir et al., 2019.
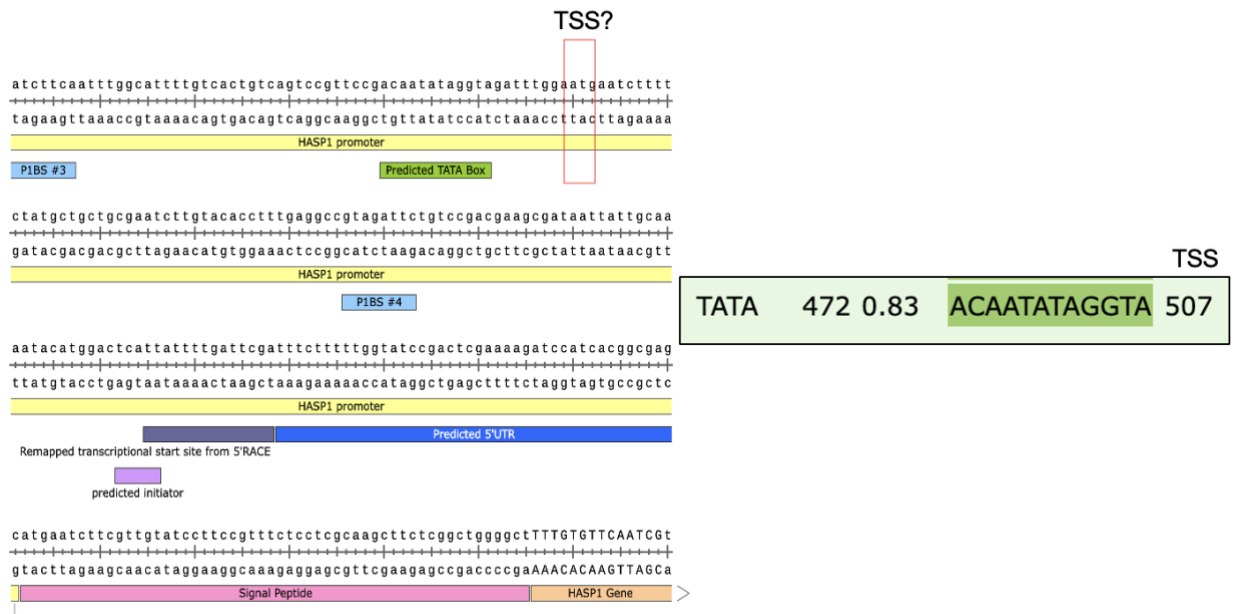
Figure 2-15: Predicted TATA-box using YAPP software. The predicted TATA-box is highlighted in green and the predicted TSS with a TATA-box at this position is denoted by a red box. Light blue boxes denote P1BS motifs deleted in mutant 2. Light purple box denotes the predicted initiator, dark purple box denotes remapped transcriptional start site from 5'RACE and dark blue boxes denote predicted 5'UTR from Erdene-Ochier et al., 2019.

Transcription factors responsive to phosphate depletion within other organisms have been found to act as a positive inducer and bind as a dimer via its MYB domain to the P1BS motif within the promoters of crucial Pi-responsive genes (Figure 2-16) [47, 52].

The mutant 1 promoter showed an increase in eGFP expression compared to wildtype and mutant 2. This may suggest that these promoter elements are important for promoter regulation under P-stress as their absence caused an increase in promoter induction, potentially revealing the possibility of a repressor binding to said regions (Figure 2-17).
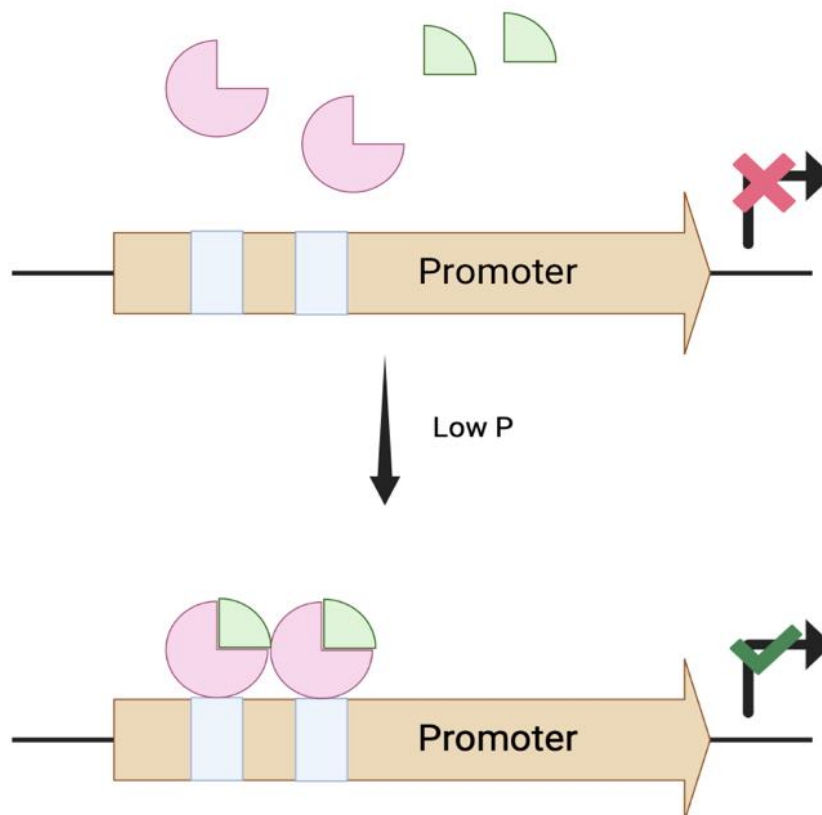
Figure 2-16: Positive promoter induction. The depletion of phosphate recruits inducers (green) to bind to the activators (pink) at specific phosphorus binding sites (blue). As a result, expression of genes responsible for phosphate acquisition occurs.

On the contrary, when P1BS regions closest to the predicted TSS are deleted (MT2) there is a large reduction in protein expression. In this scenario, P1BS regions further upstream (MT1) are present, again suggesting the importance of MT1 P1BS-like elements in transcriptional regulation. Furthermore, the reduction in protein expression in MT2 could be a result of the presence of MT1 if a repressor is bound.

Mutant 3 is an interesting case, as it contains deletions in all four P1BS-like regions. eGFP expression and secretion still occurs which may suggest the promoter is transiently turned on under any condition. The 4 deletions may have created a conformational change, allowing the promoter to be accessible to RNA polymerase. To support this idea, future work should assess MT3 function in 100% phosphate and iron conditions and determine when promoter induction occurs. If secretion occurs earlier, this may suggest dysregulation.

To conclude, this work only provides a small glimpse into promoter induction in response to phosphate starvation. Specifically, the mutant 1 promoter serves as a potential candidate for a novel inducible expression system to produce high yield recombinant proteins as its induction occurs early on and the ratio of protein being expressed throughout the experiment compared to wildtype remained the highest.
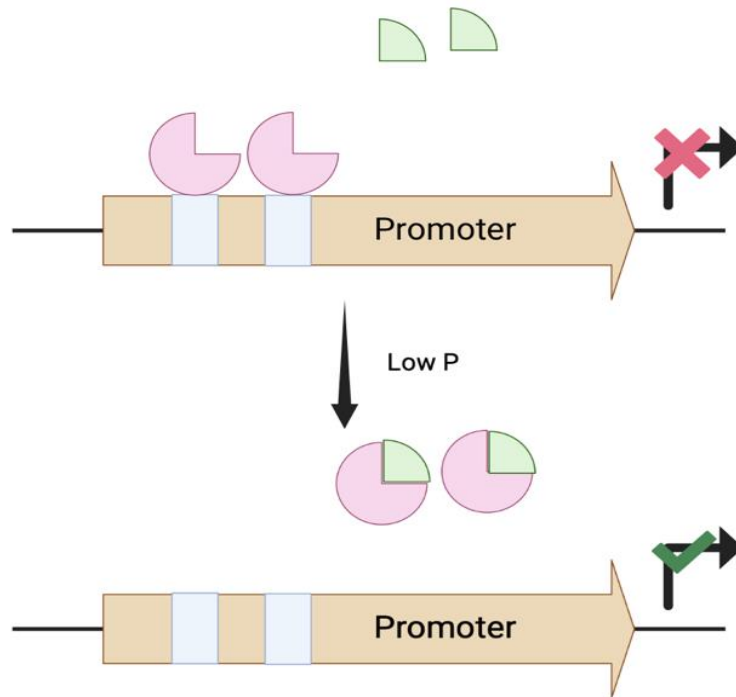
Figure 2-17: Derepression of a promoter. The depletion of phosphorous recruits the inducer (green) to the repressors (pink) at specific phosphorous binding sites. As a result, expression of genes responsible for phosphate acquisition occurs.

Chapter 3

3 Assessing *HASP1* phytase-like potential in different phosphorus sources and transcriptomic responses in response to phosphate depletion in a *HASP1* knockout strain

3.1 Introduction

Diatoms have acquired many genes and metabolic pathways through secondary endosymbiosis and horizontal gene transfer from bacteria that allows them to survive under adverse environmental conditions [4,53]. Due to its fast cycling, phosphorus is often limited in many areas of the ocean [4, 13]. Orthophosphate anions are the most common soluble inorganic form of phosphate that is taken up by diatoms. Under nutrient deprivation, like phosphate depletion, there is upregulation of phosphate transporters and alkaline phosphatases (AP) which are associated with phosphate starvation [4,6,54]. In such instances, diatoms can also use dissolved organic phosphorus (DOP). DOP cannot be assimilated directly by diatoms. Thus, they can use scavenging enzymes (i.e., alkaline phosphatases, phosphodiesterases, nucleotidases) to acquire a phosphorus source [17,18]. Myo-inositol phosphates are a group of organic phosphorus compounds widely dispersed in organisms and often found in marine ecosystems. The most abundant is IP_6, known as phytic acid (PA, or phytate in salt form) [25, 55]. Phytic acid plays an important role in cellular regulation and the transformation of different P storage forms.

The *HASP1* protein is predicted to be phytase-like, as it contains a phytase-like domain at its c-terminal end. Under low-phosphate conditions, phytases will break down phytic acid (DOP) as a phosphorus source [35]. PA was recently used in a study to determine the underlying metabolic mechanisms in *P. tricornutum* [25]. Their results showed that *P. tricornutum* can utilize PA as a P nutrient source but less efficiently and by a different mechanism compared to other sources of DOPs [25, 30].

Here, I worked to determine *HASP1* protein function by using a gene (KO) construct to assess cellular growth under different phosphate sources; phytic acid and sodium phosphate (DOP vs DIP, respectively). Additionally, two gene rescue plasmids were constructed and evaluated in the same conditions as a cross comparison if any phenotypic

changes were observed in the KO. Lastly, a transcriptomic study was conducted using the KO alongside wild type to determine which genes may be differentially regulated in response to phosphate depletion. Understanding protein function can provide a glimpse of how the phosphate starvation is activated and other potential key players responsible for phosphate acquisition.

3.2   Methods

3.2.1   *Plasmid Design and Construction*

Two rescue plasmids were created to determine any potential phenotypic changes that may have occurred in the KO. To create the rescue plasmids, the *HASP1* promoter, signal peptide, gene and terminator were cloned into the pDMI-2 backbone vector. The two different constructs differ in that one has the *HASP1* gene with introns (INT) and the other is cDNA only. The two different gene sequences were ordered from IDT (IDT, 2023) as gBlocks (Table B-3). Fragments were cloned into the backbone using a yeast assembly method (Figure 3-1) [50,51]. The resultant media containing yeast colonies was collected and DNA was extracted and transformed into EPI300 *E. coli* cells. DNA from single *E. coli* colonies was extracted using the Monarch® Plasmid Miniprep kit (New England Biolabs, T1010L). To determine correct assembly of plasmids, a restriction enzyme digest was performed, and DNA was sent for analysis of the *HASP1* locus by Sanger sequencing at the London Regional Genomics Center.
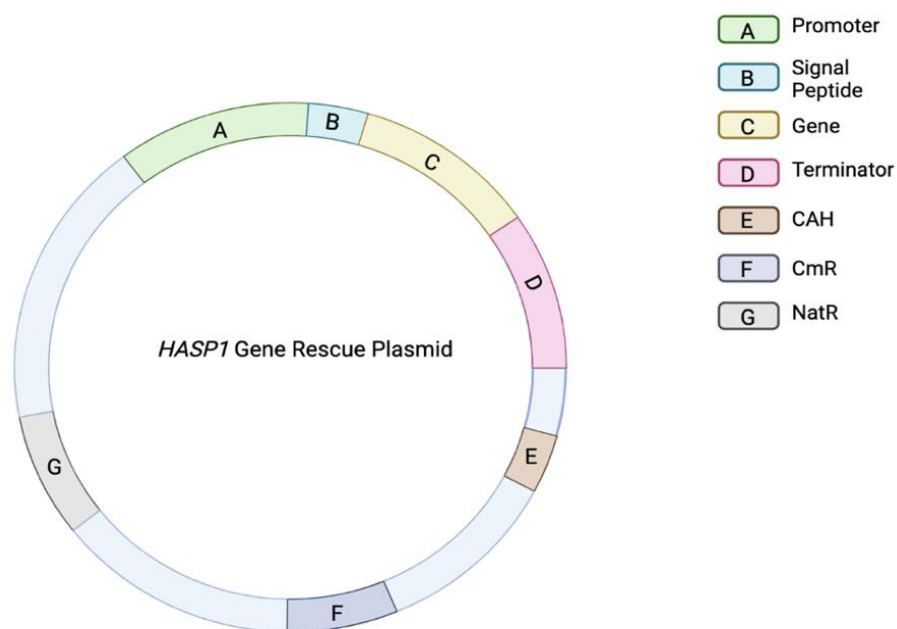
Figure 3-1: *HASP1* **gene rescue plasmid.** CAH denotes CEN6-ARSH4-HIS3, CmR denotes chloramphenicol resistance and NatR denotes nourseothricin resistance.

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

A heat shock transformation was conducted on correct colonies into the pTA-MOB *E. coli* strain for conjugation. pTA-MOB contains the cellular machinery required for bacterial conjugation into *P. tricornutum* cells. The day before conjugations, the resultant colonies were grown in 5mL of LB media containing chloramphenicol (25mg L^{-1}) and gentamycin (20 mg L^{-1}) antibiotics.

Conjugations were performed and adapted from Karas et al. [36]. It is important to note rescue plasmids were conjugated into the *HASP1* KO strain instead of wildtype to "rescue" any phenotypic changes that may have occurred (Table 3-1). Additionally, the pDMI-2 plasmid was conjugated into the wildtype and KO strain, so all constructs contained a plasmid and contained the same antibiotic resistance (Table 3-1). In short, liquid cultures of *P. tricornutum* containing the *HASP1* gene KO and wildtype (250μL), were adjusted to a density of 1.0×10^8 cells mL^{-1} using counts from a hemocytometer. Cells were plated on ½ x L1 1% agar plates containing no antibiotics and grown for four days at 18°C under cool white, fluorescent lights (75 μE m^{-2}s^{-1})

and a photoperiod of 16h light: 8h dark. After 4 days, 1.5mL of L1 media was added to plates and then scraped. The concentration was then adjusted to 5.0×10^8 cells mL^{-1}. *E.coli* cultures containing plasmids for conjugation were grown in 50mL LB media at 37°C to A$_{600}$ of 0.8-1.0, centrifuged for 10 minutes at 3000xg and resuspended in 500µL SOC media.

To initiate conjugation, 200µL of *P. tricornutum* (wildtype and KO) and *E. coli* cells were combined and plated on ½ x L1 x 5% LB 1% agar plates, incubated at 30°C for 90 minutes and then moved to 18°C and grown for 2 days. After 2 days, plates were scraped using 1.5mL L1 media and 300µL is plated on ½ x L1 1% agar plates containing nourseothricin antibiotic (100mg L^{-1}) and grown for 1-2 weeks at 18°C. **Table 3-1** shows all plasmid containing constructs.

Table 3-1: *P. tricornutum* cultures containing plasmids.

P. tricornutum strain	Plasmid
Wildtype *P. tricornutum*	pDMI-2
HASP1 KO	pDMI-2
HASP1 KO	cDNA + pDMI-2
HASP1 KO	INT+pDMI-2

3.2.3   *PCR Screen*

Resultant *P. tricornutum* colonies were screened for the presence of the mutant *HASP1* locus or nourseothricin resistance marker using the Thermo Scientific Phire Direct Plant PCR Master Mix (Thermo Scientific^{TM}, F160S) kit. For rescue plasmids containing *HASP1* and its terminator, PCR was performed on 6 colonies for each construct using forward primers overlapping 3' end of the signal peptide and 5' end of *HASP1* (DE6778) and reverse primers overlapping the 5' end of the origin of transfer (OriT) and 3' end of

HASP1 terminator (DE6712, Table B-1). The predicted product for colonies containing cDNA are 2898bp and colonies containing introns are 3062bp (Figure A-2).

For wildtype and KO cells containing pDMI-2 plasmids, forward primers overlapping 3' end of FcpD promoter and 5' end of the nourseothricin resistance marker (DE3074) and reverse primers overlapping 3' end of nourseothricin resistance marker (DE3072) were used (Table B-1).

Correct colonies, C1 and I3, plasmids with genes containing cDNA and introns, respectively, and wildtype and KO colonies with pDMI-2 were streaked onto ½ x L1 1% agar plates containing nourseothricin antibiotic (100mg L^{-1}) and grown for ~2 weeks at 18°C under cool white, fluorescent light (75 μE m^{-2}s^{-1}) and a photoperiod of 16h light: 8h dark. Streaks were then transferred to liquid cultures containing ½ x L1 media with nourseothricin (100mg L^{-1}) and grown to a concentration of 1.0×10^6 cells mL^{-1} or greater for ~2 weeks at the same condition previously mentioned.

3.2.4 *HASP1* Gene Knockout

HASP guide 1 (Table B-1) was cloned into the pSS29 vector (NR-cas9 with pACYC origin) and conjugated into *P. tricornutum* as previously described. Exconjugants were selected for on plates containing 100mg L^{-1} of Zeocin and streaked onto new plates for further screening.

A PCR screen was conducted using the Thermo Scientific Phire Direct Plant PCR Master Mix (Thermo Scientific^{TM}, F160S) kit. Forward primer (DE5481) and reverse primer (DE5482) were used to verify insertion of guide 1 (Table B-1). The predicted size was 515bp (Figure A-3).

After determining correct colonies, a T7 assay to evaluate gene editing was completed as follows: 2.5μL of sample PCR product and 2.5μL of wildtype PCR product were combined and incubated at 95°C for 5 minutes, then slow cooled to 50°C and flash frozen at -20°C. Next, 1.5μL of NEB2 buffer, 0.2μL of NEB T7E1 (New England Biolabs,

M0302S) and 8.3μL of ddH2O were added to each sample. Samples were incubated for 20 minutes at 37°C. Entire reactions were loaded onto an agarose gel (Figure A-4).

Colonies that had positive T7 hits were re-streaked onto plates containing 100mg L^{-1} of Zeocin to get single colonies and then re-streaked again.

A PCR (20uL reaction) was conducted for subclones of HASP1 g1 and g2 followed by a T7 assay (mixed PCR products of T7 edited and WT PCR were mixed at a 50:50 ratio, Figure A-5). Finally, positive subclones from T7 assay were sent for Sanger sequencing at the London Regional Genomics Center. Subclone 1.1.1 was found to have a homozygous 20 base-pair deletion (Figure A-6).

3.2.5   *Sample Preparation*

Cultures with correct colonies were transferred to liquid cultures containing ½ x L1 media with nourseothricin (100mg L^{-1}) and grown to a concentration of 1.0×10^6 cells mL^{-1} or greater for ~2 weeks. Our L1 media contains 362μM of sodium phosphate (Na$_3$PO$_4$), an inorganic source of phosphate. Phytic acid (PA) was used as the organic phosphate source for this study and was commercially purchased from Millipore Sigma (Sigma-Aldrich, 593648-250ML). Cultures were counted under a light microscope to determine cell density after 2-weeks. 1.0×10^6 cells mL^{-1} of each culture (Table 3-2) were pelleted down at 3000 x g, 4°C for 10 minutes.

The cell pellet was washed by resuspension in 10mL of L1 containing no phosphate. Final resuspension was done in phosphate free L1. Three replicates were made for each condition by passaging 1.0×10^6 cells mL^{-1} of the prepared culture into 15mL of L1 media containing no phosphate. For this study, the amount of phosphorus had to be equal between Na$_3$PO$_4$ and PA conditions. Na$_3$PO$_4$ and PA stock solutions were used to adjust cultures to conditions outlined in Table 3.2. All cultures were supplemented with nourseothricin (150mg L^{-1}). Cultures were grown for a total of 28 days at 18°C under cool white, fluorescent light (75 μE m^{-2}s^{-1}) and a photoperiod of 16h light: 8h dark.

Table 3-2: Cultures and conditions used in this study.

P. tricornutum Culture	Conditions
Wildtype + pDMI-2	100% Na_3PO_4, 100% PA
	5% Na_3PO_4, 5% PA
HASP1 KO + pDMI-2	100% Na_3PO_4, 100% PA
	5% Na_3PO_4, 5% PA
C1	100% Na_3PO_4, 100% PA
	5% Na_3PO_4, 5% PA
I3	100% Na_3PO_4, 100% PA
	5% Na_3PO_4, 5% PA
pSS10	100% Na_3PO_4, 100% PA
	5% Na_3PO_4, 5% PA

3.2.6 *Determining Cell Density and eGFP Fluorescence in Supernatant*

Cell density was determined by aliquoting 100 μL of whole cell culture into a clear 96- well plate and measuring absorbance at 670nm (A_{670}) and 750nm (A_{750}) using the Biotek Synergy H1 plate reader every two days. Absorbance values were fitted to a standard curve to determine cells mL^{-1}.

To determine eGFP fluorescence in supernatant, 700μL of whole cell cultures were pelleted at 18 000 x g for 15 minutes. 200μL of supernatant was aliquoted into clear bottom 96-well plates. Fluorescence was measured at an excitation of 475nm and emission of 515nm. Wildtype (auto)fluorescence was subtracted from reads. Fluorescence readings were taken every four days and was only done on pSS10 cultures to determine if the phosphate source effected secretion.

3.2.7 *RNA isolation of P. tricornutum cells for RNAseq*

To prepare cells, 200mL of *HASP1* KO *P. tricornutum* cells and 200mL (5x50mL) of wildtype *P. tricornutum* cells were grown to a density of 1×10^7 cells mL^{-1}. On day 0 of experiment, 20mL of sample cultures were taken for analysis. Another 30mL of culture was passaged into 270mL of fresh L1 media in 1-liter flasks (x5 replicates) to a final density of 1×10^6 cells mL^{-1}. From the 20mL of culture, 2×10^8 cells mL^{-1} were pelleted for RNA extraction. On days 4, 10 and 17, 50mL of sample were taken for analysis and 2×10^8 cells mL^{-1} were pelleted for RNA extraction.

RNA extraction protocol is as follows: *P. tricornutum* cells were flash frozen in liquid nitrogen and ground up using a pastel and mortar that has been washed in liquid nitrogen to prevent cells from thawing and stored at -80°C. 50-100mg of frozen ground *P. tricornutum* was measured into a clean Eppendorf tube. RNA is then extracted using the Monarch® total RNA miniprep kit protocol for "Tough-to-Lyse Samples" (New England Biolabs, T2010S). RNAseq was performed at the London Reginal Genomics Center using Illumina NextSeq high output single end 75 run.

3.2.8 *Statistical Analysis*

For all experiments Microsoft Excel (version 16.71) was used. For experiments assessing cell density of various constructs grown in different concentrations of phytic acid or sodium phosphate data was presented as mean ± standard error of mean (SEM). The average of 3 biological replicates was calculated. To determine statistically significant changes between cells grown in PA or NaPO$_4$, an unpaired one-tailed Student's *t* test was conducted on days 8, 16 and 28. A p value of ≤ 0.05 was considered statistically significant.

For experiments assessing eGFP secretion under different sources of phosphorous, all conditions were compared to pSS10 grown in 5% sodium phosphate using an unpaired one-tailed Student's *t* test for days 8, 16 and 28. A p value of ≤ 0.05 was considered statistically significant. Data was presented as mean ± standard error of mean (SEM) of three biological replicates.

To determine differentially expressed genes, reads were trimmed using Trimmomatic version 0.36 with options LEADING: 10 TRAILING: 10 [56]. Processed reads were then mapped to *P. tricornutum* genome and plasmid reference sequences using Hisat2 version 2.2.0 [57]. Htseq-count version 0.13.5 was used to count the number of reads mapping to each annotated feature within the genome and plasmid sequences [58].

Analysis of count data was carried out with the ALDEx2 package from Bioconductor using R Studio (version 3.0.1). A model matrix for a vector of conditions and count data were used to generate 16 Monte Carlo instances of the Dirichlet distribution for each sample (aldex.clr). Each instance was converted using a log-ratio transformation. A general linearized model (glm) was used to calculate the expected values for each coefficient of a glm model from the clr data. Data were adjusted for a false discovery rate using the Benjamini–Hochberg procedure and an adjusted p-value of $p < 0.01$ was considered statistically significant.

3.3   Results

3.3.1   *Verification of HASP1 KO*

To determine whether the gene had been successfully knocked out, 6μL of wildtype *P. tricornutum* and KO cells originally grown in full L1 media were filter sterilized into fresh tubes. Cultures were pelleted at 6000 x g for 5 minutes to pellet cells. Supernatant was then transferred into a 6mL 30kDa spin column and centrifuged for 45 minutes at 4000 x g. The concentrated supernatant was resolved on an SDS-PAGE gel (8%) and stained with coomassie blue for visualization.

Figure 3-2 shows a successful knockout of the *HASP1* gene. Furthermore, the upregulation of another protein was seen in lane 2 at ~85kDa. The band at 85kDa was isolated from the SDS-PAGE gel and sent for MALDI-mass spectrometry (MALDI-MS) at the London Regional Proteomics Centre. The results indicated the band was a predicted AP (PhoA, PHATRDRAFT_49678).
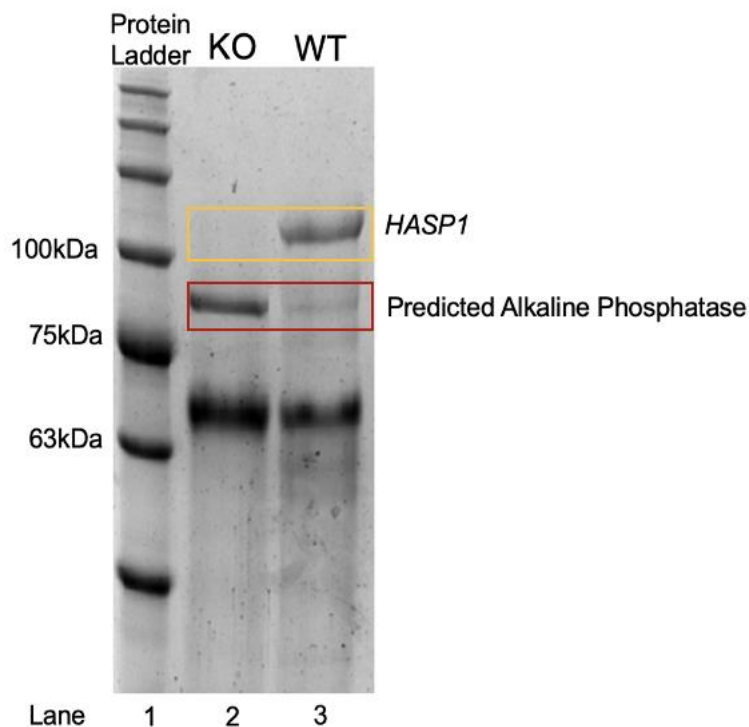
Figure 3-2: : Concentrated supernatant of wildtype (WT) and *HASP1* gene knockout (KO) strains resolved on 8% SDS-PAGE gel stained with coomassie blue. Lane 1 contains a protein ladder, lane 2 contains the *P. tricornutum* knockout cells and lane 3 contains wildtype *P. tricornutum* cells. Yellow box denotes location of *HASP1*. Red box shows location of predicted alkaline phosphatase at ~85kDa (PhoA, PHATRDRAFT_49678).

3.3.2   *Cell Density may be Dependent on the Source and Concentration of Phosphorus.*

P. tricornutum cells were grown in various concentrations of organic (PA) and inorganic (Na$_3$PO$_4$) phosphorus to determine if the source and concentration of phosphorus affected the growth of cells. Table 3-3 briefly outlines all constructs used in the study.

Table 3-3: All constructs and their description used in this study.

P. tricornutum Construct	Description
Wildtype (WT)	Wildtype *P. tricornutum* cells which contain the pDMI-2 plasmid
HASP1 KO	*P. tricornutum* cells which contain *HASP1* gene knockout and pDMI-2 plasmid
C1	*HASP1* rescue plasmid in *HASP1* KO cells. This construct contains cDNA of *HASP1* gene.
I3	*HASP1* gene rescue plasmid in *HASP1* KO cells. This construct contains introns in *HASP1* gene.
pSS10	Plasmid contains *HASP1* promoter and signal peptide driving eGFP.

All cells grown in 100% PA had a higher cell density compared to cells grown in 100% Na_3PO_4 (Figure 3-3). Notably, *HASP1* KO cells in 100% PA had a lower average cell density compared to WT cells in 100% PA (Figure 3-3A,B). In all cells (except C2 and I2), there was a significant increase in average cell density on days 8 and 16 when cells were grown in PA (Figure 3-3, $p \leq 0.05$). However, by day 28, there was no significant difference between cells grown in PA versus Na_3PO_4 ($p > 0.05$). These results suggest *P. tricornutum* cells have various methods of scavenging for phosphorus which may be beneficial during early stages of cellular growth.
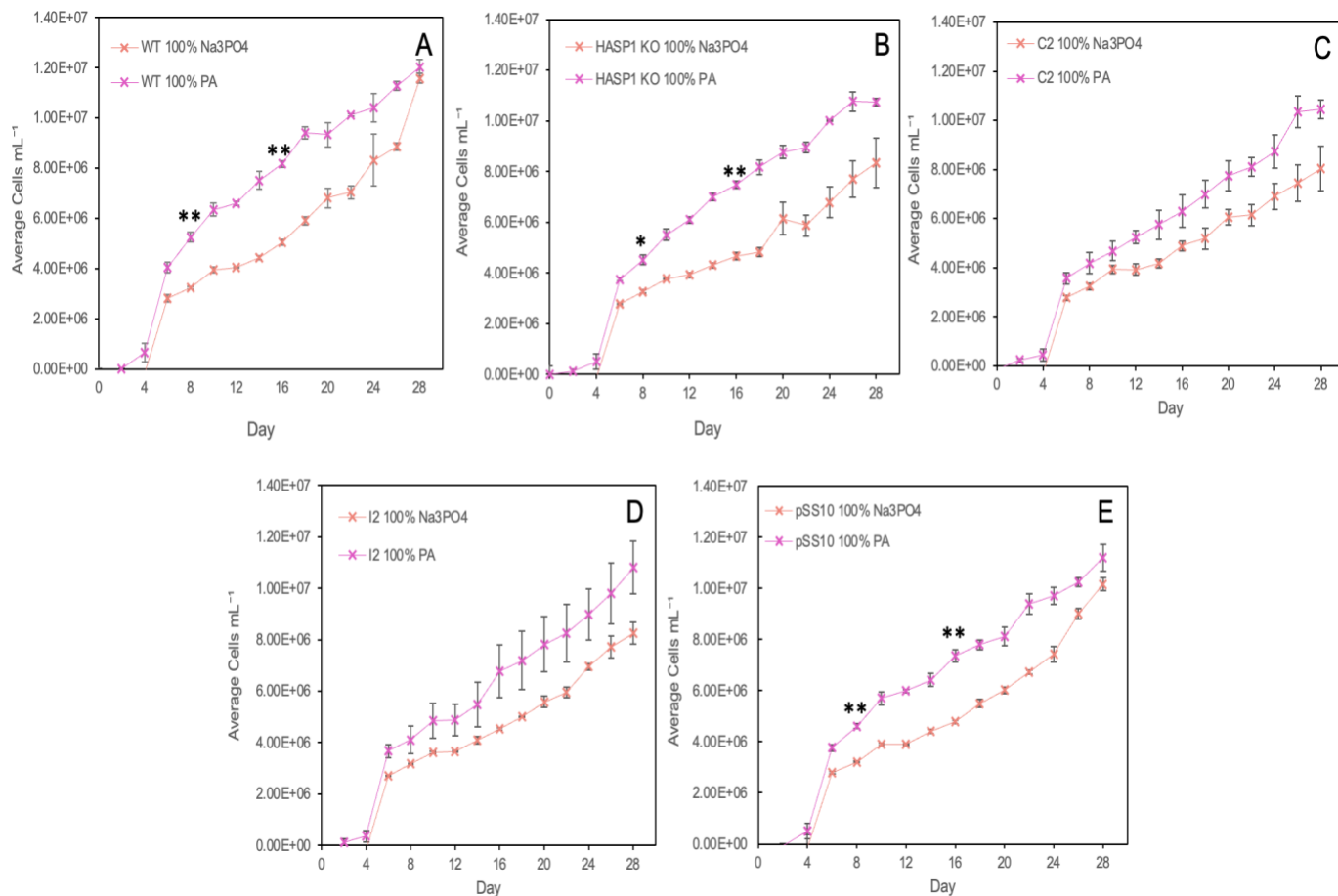
Figure 3-3: Growth curve of *P. tricornutum* cells grown in 100% sodium phosphate (Na₃PO₄) or phytic acid (PA) for a duration of 28 days. (A) Wildtype *P. tricornutum* cells, (B) *P. tricornutum* cells with *HASP1* KO, (C) *HASP1* KO cells with C1 rescue plasmid (D) *HASP1* KO cells with I3 rescue plasmid, (E) *P. tricornutum* cells with pSS10 plasmid. Pink denotes cells grown in PA and orange denotes cells grown in Na₃PO₄. Absorbance was taken at 670nm. Each point denotes mean cells mL^{-1} ± SEM of three biological replicates. An unpaired two-sample *t* test was computed between the same constructs grown in different phosphorous sources on days 8, 16 and 28. * denotes a p≤0.05 and ** denotes a p≤0.01.

Cells grown in 5% Na₃PO₄ had a higher cell density compared to cells grown in 5% PA (Figure 3-4). The average cell density of cells grown in Na₃PO₄ was significantly higher on days 8 and 16 than cells grown in PA (P≤0.01), except for wildtype cells which were significantly higher on day 8. These results suggest that under nutrient deprivation, *P.*

tricornutum cells will use inorganic sources of phosphorus, however, like in 100% phosphorus conditions, the source of phosphate may be crucial during the first two weeks of cellular growth.

Compared to cells grown in 100% phosphorus conditions, the cells grown in 5% conditions had a decrease in cell density. This suggests that under nutrient stress, *P. tricornutum* cells will allocate their energy to other important cellular processes rather than cell growth to ensure survival.
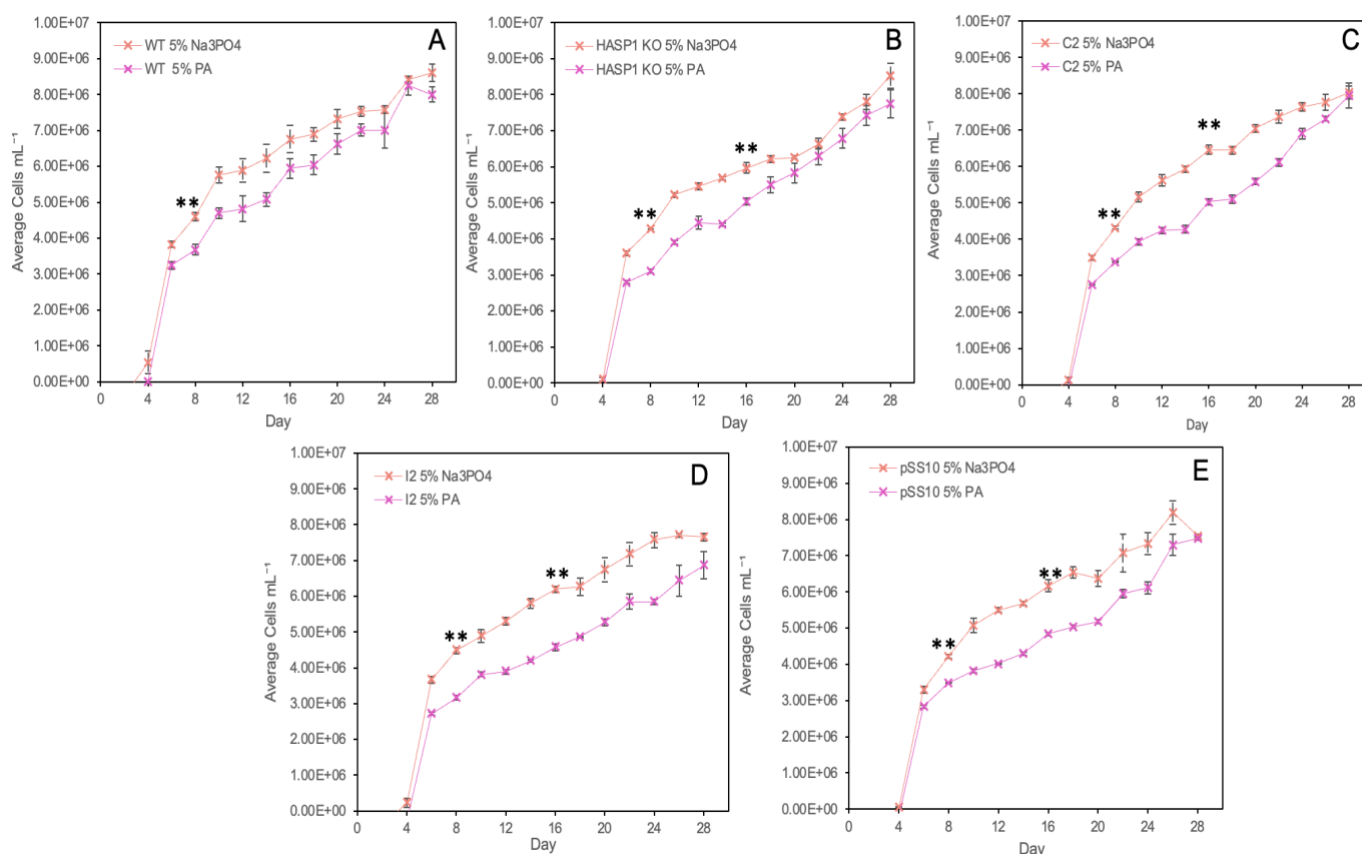
Figure 3-4: Growth curve of *P. tricornutum* cells grown in 5% sodium phosphate (Na₃PO₄) or phytic acid (PA) for a duration of 28-days. (A) Wildtype *P. tricornutum* cells, (B) *P. tricornutum* cells with *HASP1* KO, (C) *HASP1* KO cells with C1 rescue plasmid (D) *HASP1* KO cells with I3 rescue plasmid, (E) *P. tricornutum* cells with pSS10 plasmid. Pink denotes cells grown in PA and orange denotes cells grown in Na₃PO₄. Absorbance was taken at 670nm. Each point denotes mean cells mL^{-1} ± SEM of three biological replicates. An unpaired two-sample *t* test was computed between the same constructs grown in different phosphorous sources on days 8, 16 and 28. * denotes a p≤0.05 and ** denotes a p≤0.01.

Overall, wildtype cells grown in the same conditions as KO cells had a greater overall average cell density. This may suggest cellular energy is being redirected to the expression of other proteins needed for phosphate scavenging rather than cell growth.
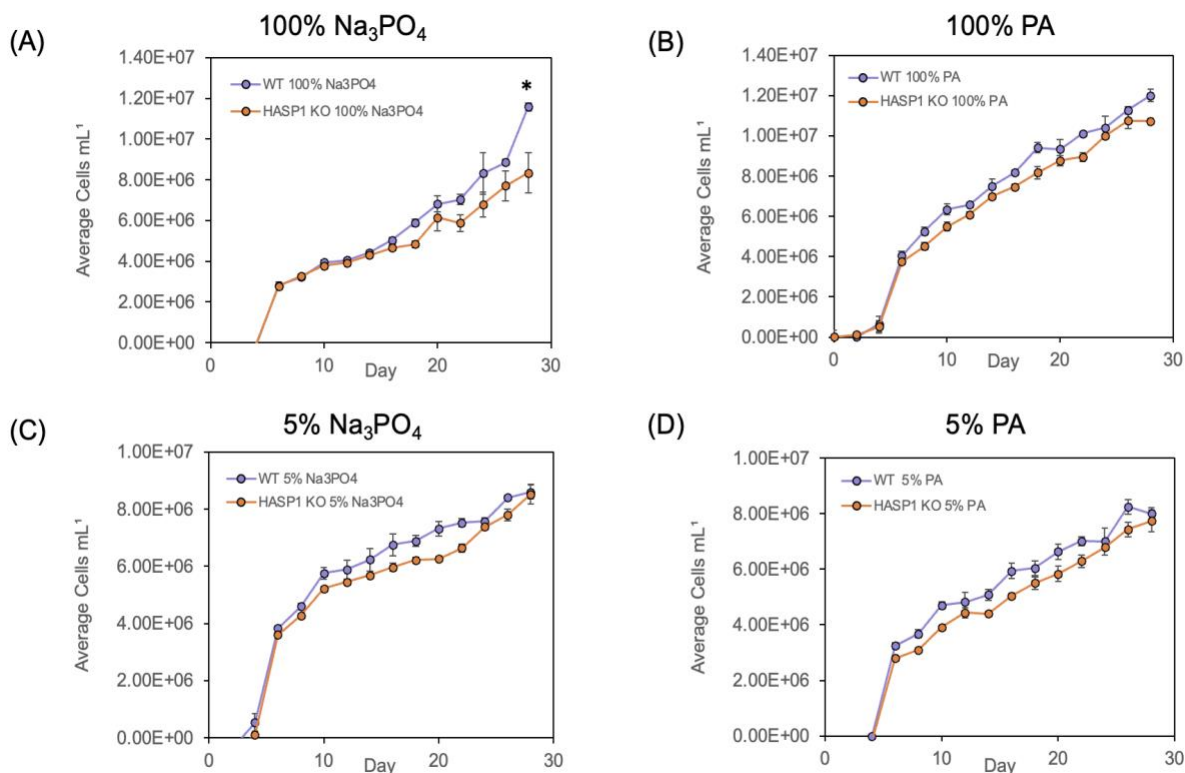
Figure 3-5: Comparison of average cell density between WT and *HASP1* KO cells grown in various concentrations of sodium phosphate and phytic acid. (A) 100% Na₃PO₄, (B) 100% PA, (C) 5% Na₃PO₄, (D) 5% PA. Absorbance was taken at 670nm. Each point denotes mean cells mL^{-1} ± SEM of three biological replicates. An unpaired two-sample *t* test was computed between wildtype and KO cells grown under the same condition on days 8, 16 and 28. * denotes a statistically significant change in average cell density (p≤0.05).

Following the 28 day experiment, supernatants of wildtype and KO *P.tricornutum* cells were concentrated 100-fold to assess the upregulation of genes in response to different concentrations and sources of phosphorus. 10µL of each sample was resolved on an 8% SDS-PAGE gel and stained in coomassie blue. Cells grown in low concentrations of phosphorus (5%) show an increase of protein secretion in the supernatant (Figure 3-6). These results suggest that under phosphate depletion there was changes in the regulation of genes responsible for phosphate acquisition.
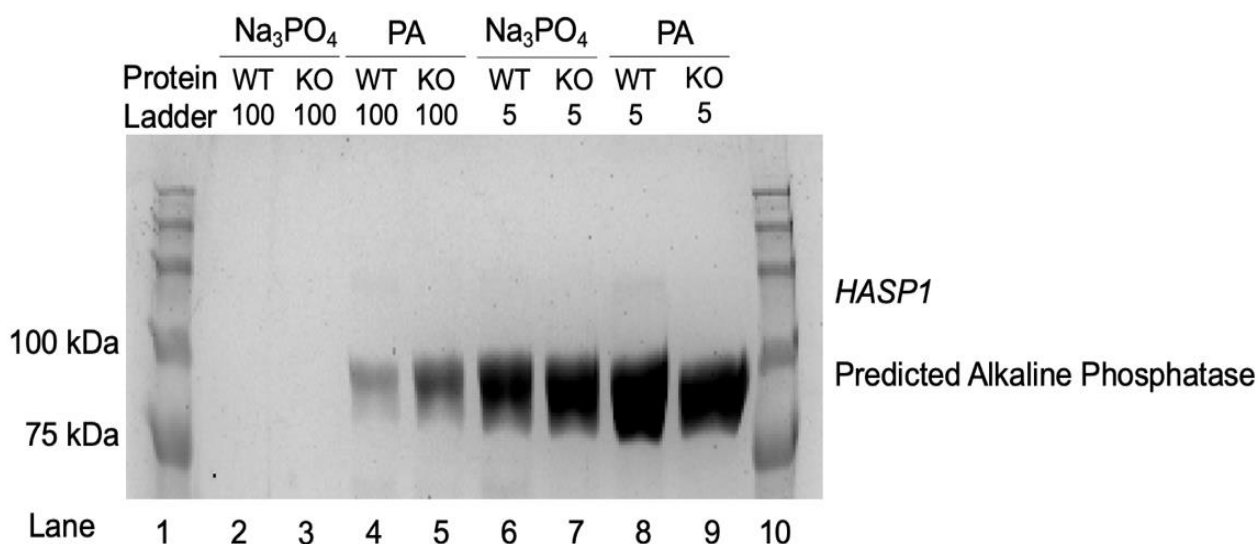
Figure 3-6: Concentrated supernatant of wildtype and knockout *P. tricornutum* cells grown in various concentrations and sources of phosphorus following a 28 day growth assay. Lane 1 contains BLUelf pre-stained protein ladder. Cells grown in 100% and 5% concentrations of phosphorous are denoted by 100 and 5, respectively. WT denotes wild type cells, KO denotes cells containing a *HASP1* gene knockout, Na₃PO₄ denotes cells grown in sodium phosphate and PA denotes cells grown in phytic acid.

3.3.3   *eGFP Secretion Increases in Phytic Acid Conditions*

The pSS10 plasmid contains a wildtype *HASP1* promoter and signal peptide driving eGFP expression. Cells containing pSS10 were grown under various concentrations of

PA and Na₃PO₄ to determine whether the source and concentration of phosphorus was important for protein secretion. All conditions were compared to pSS10 grown in 5% sodium phosphate as our previous work has shown enhanced eGFP secretion in the supernatant in 5% sodium phosphate conditions.

pSS10 grown in 5% PA had 2x more eGFP expression compared to cells grown in 5% Na₃PO₄ by day 28 (p<0.01). Furthermore, eGFP expression was higher in cells grown in 100% PA compared to 100% and 5% Na₃PO₄ (Figure 3-7). pSS10 cells grown in 100% sodium phosphate had significantly lower eGFP expression compared to 5% sodium phosphate (p<0.01). Lastly, cells grown in 5% PA had an increase in eGFP expression compared to cells grown in 100% Na₃PO₄. These results suggest PA may enable protein secretion more than Na₃PO₄.
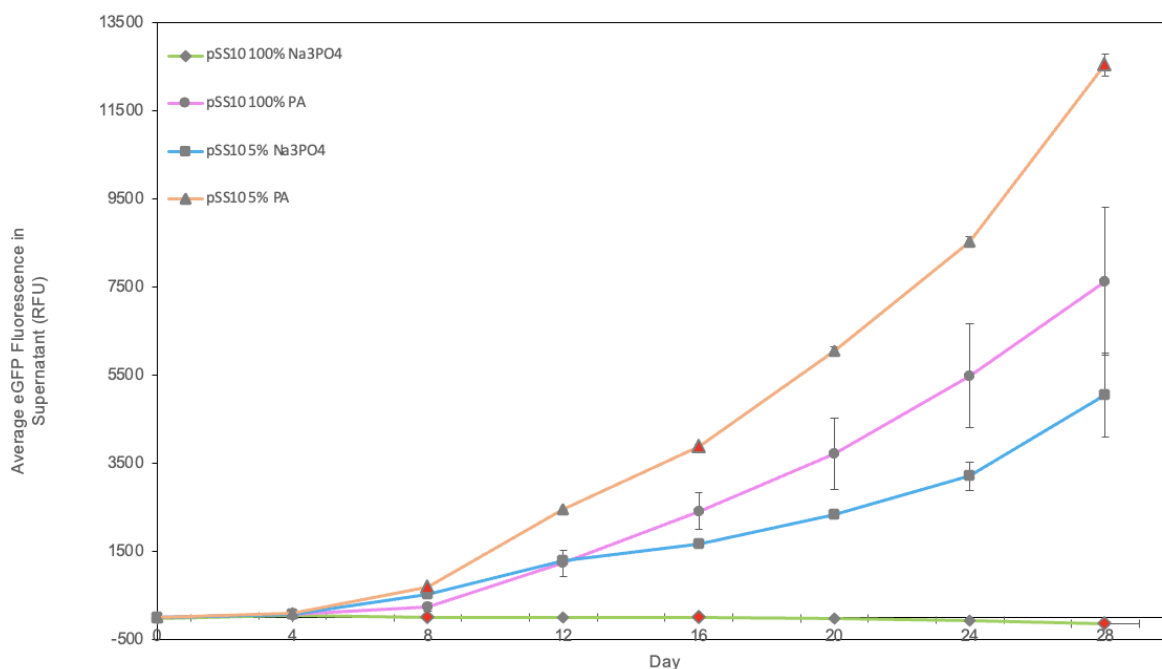
Figure 3-7: Average eGFP fluorescence in supernatant of pSS10 cells grown in different sources and concentrations of phosphorus for 28 days. Each point denotes average eGFP fluorescence ± SEM of three biological replicates. An unpaired two-sample *t* test was computed between cells grown in 5% Na₃PO₄ and other cells grown in different concentrations of sodium phosphate or phytic acid on days 8, 16 and 28. Red points denote a statistical significance of p≤0.01.

3.3.4 *RNAseq Data Analysis*

RNAseq analysis was conducted to determine other genes which may be expressed under phosphate depletion in the absence *HASP1*. Wild type and *HASP1* KO *P. tricornutum* cells were grown in full L1 media for a total of 17 days. Five replicates were used for KO and WT cells. Samples for RNAseq were taken as phosphorus depleted on days 0, 4, 10 and 17.

An effect plot was used to show the difference versus the pooled standard deviation (dispersion) of KO cells during the 17 day experiment (Figure 3-8A). These results suggested differences in differential gene expression are due to changes in cellular growth and nutrient depletion after 17 days.

Figure 3-8A shows the symmetry of the data shown in Figure 3-8B. Lastly a volcano plot was used to look at difference versus adjusted p-values (Figure 3-8C). The volcano plot shows the magnitude of change and allows for quick identification of genes that may be of biological interest. Values below 0 on the x-axis represent the downregulation of genes in WT compared to KO cells. Many of the differentially expressed genes were in involved in metabolic processes. A list of some of the differentially expressed genes can be found in Table 3-4.
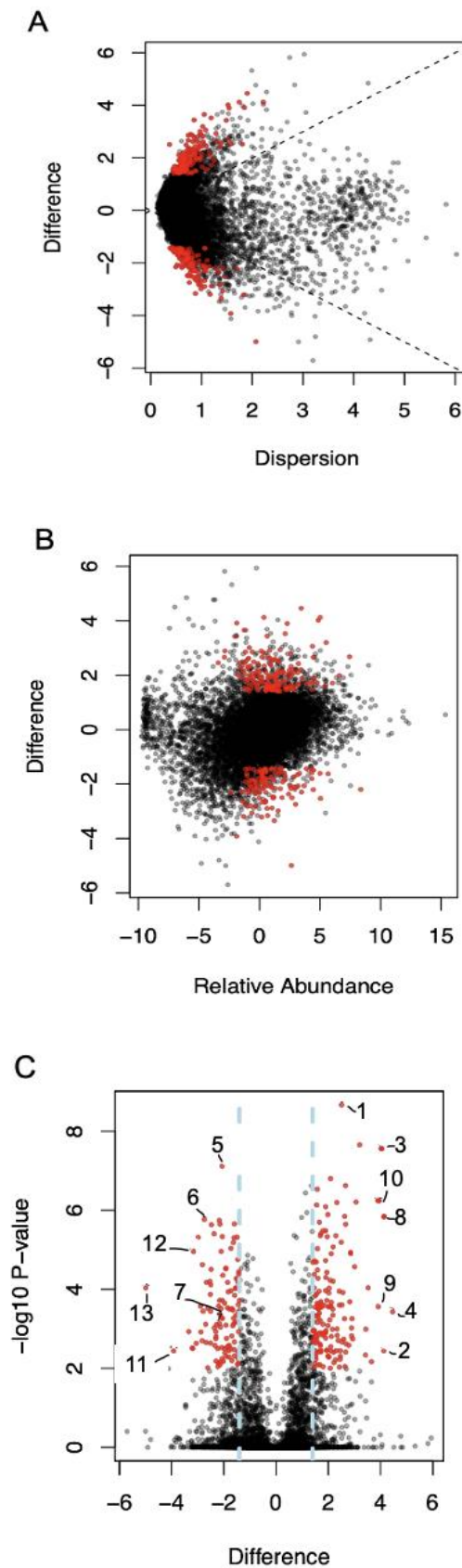
Figure 3-8: Transcriptional changes in the *HASP1* KO versus wildtype after 17 days of gradual phosphate depletion. (A) An effect plot (MW), difference versus dispersion (pooled standard deviation) of day 0 versus day 4 + 17. Dotted lines denote an effect size of +/- 1. (B) Bland-Altmann (MA) plot, difference versus mean value of all samples (relative abundance). (C) Volcano plot, difference versus $\log_{10}$ p-value (change in expression). Red dots denote statistical significance, p<0.01**.**

Table 3-4: List of gene names and protein function for differentially expressed genes.

Number	ID	Protein
1	PHATRDRAFT_48322	Unknown
2	PHATRDRAFT_35158	Heat Shock Protein 20 (HSP20)
3	PHATRDRAFT_54019	Heat Shock Protein 70 (HSP70)
4	PHATRDRAFT_28056	5-methyltetrahydro-pteroyltriglutamate-homocysteine S-methyltransferase
5	PHATRDRAFT_44902	Aminoacyl-transfer RNA synthetases class-II family profile domain-containing protein
6	PHATR_44091	Right-handed beta helix domain-containing protein
7	PHATRDRAFT_44900	GP-PDE domain-containing protein
8	PHATRDRAFT_55215	Heat Shock Protein 90 (HSP90)
9	PHATR_36981	Heat Shock Protein 20 (HSP20)
10	PHATRDRAFT_54257	SPX domain-containing protein
11	PHATRDRAFT_20779	Transaldolase
12	PHATRDRAFT_29283	Unknown
13	PHATRDRAFT_45944	Unknown

3.4   Discussion

As phosphorus is often limited in many parts of the ocean, diatoms have adapted to such changes through the upregulation of phosphate transporters and AP. In ideal conditions, diatoms will use orthophosphate anions, a soluble inorganic form of phosphate. However, more often, due to the fast recycling of inorganic phosphates, organic phosphates are indirectly assimilated.

A recent study has shown the diatom, *P. tricornutum,* can use phytic acid under phosphate deprivation [25]. *HASP1* is predicted to be phytase-like, suggesting it's protein will break down phytic acid as a phosphorus source. To determine phytase-like potential, a gene knockout was created and assessed under various concentrations of sodium phosphate, Na_3PO_4, an inorganic source of phosphorus used in our lab, and phytic acid (PA), an organic source of phosphorus.

The successful knockout of the gene resulted in the upregulation of another AP, PhoA (Figure 3-2). Biologically this makes sense; if the *HASP1* gene is phytase-like, we would expect the upregulation of another phytase to compensate for its loss.

All cells grown in 100% PA had a greater average cell density in comparison to cells grown in 100% sodium phosphate. Furthermore, *HASP1* KO cells had a significant increase in average cell density in 100% PA compared to the average cell density of KO cells grown in 100% Na_3PO_4 on day 16 (Figure 3-3B). Since the KO cells were able to grow to a larger cell density, this suggests there are other phytases that can compensate for the loss of *HASP1*. Since *HASP1* was found to be the most abundantly secreted protein in *P. tricornutum* [35], the assumption is that it must be an important protein for the cell. However, these results suggest *HASP1* may not be a crucial phytase under phosphate starvation as cells grew similar wildtype cells (Figure 3-5). Furthermore, the most statistically significant changes between cells grown in 100% sodium phosphate versus phytic acid occurred during the earlier stages of cellular growth. By day 28, no statistically significant changes were seen. This may also suggest that the source of phosphorous may be crucial during the early stages of cellular growth.

When cells were grown under low phosphorus conditions, cells grown in 5% sodium phosphate grew better than cells grown in 5% PA. In conditions where nutrients are scarce, it would be energetically costly to assimilate phosphorus indirectly. However, KO cells grown in 5% PA were still able to grow to a substantial cell density which implies other phytases are working to break down phosphorus and again suggests *HASP1* may not be a critical phytase under phosphate starvation. Again, the most statistically significant changes in cellular growth were early on (days 8 and 16, figure 3-4), which suggests that even under cellular stress, the source of phosphorous initially available may impact growth of the cells however by stationary phase, the source of phosphorous is not critical to cell survival.

Overall, *HASP1* KO cells grown in PA (100% or 5%) had a lower average cell density compared to WT cells in the same conditions (Figure 3-5). This may be due to changes in gene expression in knockout cells, where energy is being redirected to the expression of other proteins (i.e., phosphate scavengers), impeding cellular growth.

The concentrated supernatant of both wildtype and *HASP1* cells show an increase in protein secretion when grown in low concentrations of phosphorus (Figure 3-6). Additionally, the predicted AP seems to be upregulated in most conditions. The decrease in phosphorus has increased the expression of other phosphate scavenging enzymes. In *HASP1* KO conditions, there is a notable upregulation in AP compared to wildtype cells, suggesting a compensation for the loss of phytase activity in KO cells. Unexpectedly, in 100% Na_3PO_4 no protein is being expressed. On day 28, phosphorous in 100% conditions should be almost depleted and *HASP1* should be expressed (Figure 3-2). This unexpected result may be attributed to the volume of supernatant concentrated. Further studies should evaluate the effect of an AP KO on *HASP1* expression and furthermore a double knockout of AP and *HASP1* to assess how these mutants effect growth.

As noted, the induction of the *HASP1* promoter occurs under phosphate depletion. During the duration of the study, we saw an increase of protein secretion in the supernatant as phosphate became depleted (Figure 3-7). Phytic acid increased eGFP expression and secretion into the supernatant compared to Na_3PO_4 in both conditions (100% and 5%).

There was a significant increase of eGFP in the supernatant of 5% PA cells compared to 5% sodium phosphate from day 8. Since *HASP1* is predicted to function as a phytase, the induction of the promoter and increase of protein secretion in PA conditions suggests a phytase-like function.

There are multiple ways to observe differential expression between two conditions. For this study, we were interested in differential expression between KO and WT cells. The first analysis was an effect plot (MW, Figure 3-8A). MW plots shows the symmetry seen in a Bland-Altmann (MA) plot (Figure 3-8B) and the relationship between p-values and differences (difference between the means of two groups) seen in a volcano plot (Figure 3-9C). MA plots fail to describe how large differences are relative to the underlying variability. Volcano plots seemingly show biological significance; however, the p-values do not correlate to effect size [59].

For simplicity, we will first focus on the effect plot as it encompasses both the MA and volcano plots. Our effect plot showed only a small subset of genes had a small difference and a smaller dispersion. Statistically significant points are denoted in red (FDR < 0.01), however not all the points are "real" changes (Figure 3-8A). Points with greater differences and larger dispersions are more likely to be real. After evaluating these points, transcripts which were upregulated were involved in secondary metabolism and cell stress response. Interestingly, there were no signatures of P-stress response (i.e., an increase of sodium/phosphate transporters or lipids). More interestingly, we did not see the upregulation of the AP which was previously shown. This can be attributed to the study design. Although the focus of the study was phosphate depletion, cells were grown in full L1 media which contains 100% of the nutrient cells need to grow. However, on day 17, phosphate is not the only nutrient that has become depleted. Other nutrients such as iron and nitrogen are also limited. The upregulation of genes involved in secondary metabolism and cell stress makes sense as the cells are responding to the depletion of many nutrients. Future studies should control for other nutrient stressors by supplementing cells with L1 media without phosphate every few days. This ensures the only nutrient becoming depleted is phosphate. Furthermore, the concentration of phosphate routinely used in our lab is 10 times the concentration used in other studies. As

a result, the time it takes for phosphate to become depleted is longer (~1 month). Lastly, cells that show signatures of P-stress are often in stationary phase. Cells grown to 17 days are unlikely to be late stationary and will likely not exhibit P-stress responses. A volcano plot is a great tool to visualize transcriptional fold-changes between two conditions. The volcano plot shows the upregulation of multiple heat shock proteins (HSPs, Figure 3-8C). HSPs are often expressed during normal growth processes but are also expressed during various stress conditions (i.e., environmental changes, temperature) [19]. The expression of HSPs makes sense as the environmental conditions by day 17 have changed (depletion of phosphorus, iron, and other nutrients). Interestingly, genes #7 and #10 (Table 3-3) have both been identified to be strongly upregulated genes under P-stress [14]. GP-PDE proteins are associated with phospholipid biosynthesis and degradation and SPX-domain proteins are involved in the maintenance of inorganic phosphorus homeostasis, specifically transport, storage, and signal transduction of inorganic phosphorus [61,62]. Much of the changes in gene expression were the same between KO and WT and the changes that did occur were due to nutrient depletion over time – not the loss of the predicted phytase. This means cells can compensate for the loss of *HASP1* and that its loss is not detrimental to the cell and other metabolic pathways. This has also been seen in this study, as KO cells were still capable of growing to a substantial cell density during nutrient depletion.

Moreover, many studies have conducted transcriptomic studies assessing transcriptional changes under phosphorus depletion in *P. tricornutum* cells [1, 4, 12, 63-65]. For example, Yang et al., conducted a systems-level analysis of metabolic responses in *P. tricornutum* to phosphorus stress. The researchers conducted RNA-seq on P-deplete diatom cells in early stationary phase and then validated RNA-seq data with qPCR [4]. One gene, PHATRDRAFT_40433, which encodes a phosphate transporter was upregulated 4.92 ($\log_2$)-fold and 4.81 ($\log_2$)-fold under P-deprivation from qPCR and RNA-seq analysis, respectively. This same gene was shown to be upregulated 5.8 ($\log_2$)-fold in a study done by Alipanah et al., [4, 12, 63, 65]. Using DESeq2 analysis (data not shown), I compared KO cells on day 4 versus day 17 and found PHATRDRAFT_40433 was upregulated 1.8($\log_2$)-fold (P<0.01). In comparison to other studies showing the upregulation of this gene in response to phosphorus depletion, the fold-change is lower.

This could be attributed to the cells not having reached stationary phase (i.e., still phosphorus present) by day 17. Additionally, the AP (PHATRDRAFT_49678) that we have seen to be upregulated under phosphate stress in this study (Figure 3-2, 3-6) and many other studies [4, 12 14 , 25, 63, 65] was not upregulated in the RNAseq data. Again, this may be attributed to the cells not reaching stationary phase and the media the cells were grown in. Another RNAseq (or qPCR) experiment assessing phosphate depletion over the duration of 30 days should be considered to validate these results.

Overall, the results of this study outline some interesting information about *HASP1*. *P. tricornutum* cells can use different sources of phosphorus, specifically, sodium phosphate and phytic acid and furthermore, the *HASP1* promoter is most sensitive to PA levels versus Na_3PO_4, suggesting phytase capabilities. Lastly, *HASP1,* although important to phosphate acquisition during phosphate stress, it is not essential for cellular function. In conclusion, these studies demonstrate the role of *HASP1* in the regulation of phosphate acquisition.

4   Concluding Remarks and Future Directions

The following section provides a brief summary of the work conducted in the previous two chapters, some study limitations, and future directions.

4.1   Summary of Work

Many studies have focused on assessing phosphorus starvation responses in marine diatoms. These studies are important because diatoms are crucial contributors to the marine food chain. They are responsible for approximately 40% of primary productivity in the ocean and are large contributors to the biological carbon pump [53,63]. Much of what is known about diatom survival under nutrient deprivation is through transcriptomic studies – although very informative, how these responses are activated, and the localization of proteins involved are still unknown.

Furthermore, transcriptomic studies have allowed researchers to use diatoms for diverse biotechnological applications through the understanding of metabolic processes and the development of various molecular tools [42,63]. The model *P. tricornutum* has been widely studied, making it a great candidate for such research as its genome has been entirely sequenced, has databases of expressed sequence tags (ESTs), methods for transformation, electroporation and bacterial conjugation, and lastly genome-editing tools like TALEN and CRISPR/Cas9 technologies [43, 66-76]. Of great importance for synthetic biology purposes is the development of expression systems for heterologous gene expression. To create an optimized system, a suitable promoter that can drive high yields of recombinant protein is important to develop a cost-effective system but also key for metabolic engineering through gene regulation [42].

Many endogenous promoters have been identified in the diatom *P. tricornutum*, however the discovery of the Highly Abundant Secreted Protein 1 (*HASP1*) promoter and signal peptide brought new prospects in creating a novel expression system for various applications. The promoter and signal peptide have been identified to effectively secrete protein under phosphate deprivation [35, 43]. However, how the promoter is induced and what proteins may be involved remain unknown.

I sought out to determine potential phosphate regulatory elements with the *HASP1* promoter. By creating various deletions in predicted P1BS-like regions in the promoter, which have previously been shown to be involved in P-starvation signaling. The promoter and signal peptides were fused to a reporter, eGFP, to determine promoter function under various concentrations of phosphate and iron. To summarize, under phosphate depletion, deletions within the promoter furthest away from the predicted transcriptional start site (TSS), MT1, were capable of driving eGFP expression and secretion more so than wildtype and deletions closer to the predicted TSS, MT2. Furthermore, MT3, which consisted of deletions in all 4 predicted sites, showed similar trends as seen in MT1, but to a lesser extent. These results suggest there may be repressors bound to phosphate regulatory sites within the *HASP1* promoter and sequences upstream of the TSS are important for promoter induction under phosphorous stress.

The second half of my thesis was to identify the phytase-like potential of *HASP1* and other transcriptional regulators under phosphorus starvation. First, a gene knockout (KO) was created and grown alongside wildtype and two gene rescue plasmids. These constructs were grown in Na_3PO_4, an inorganic source of phosphorus and phytic acid, an organic source of phosphorus. Each source of phosphorus was grown in 100% and 5% concentrations. To summarize, when grown in 100% concentrations, the cells grew better in phytic acid compared to sodium phosphate. This makes sense as the cell isn't under cellular stress and if phytase-like, it would preferentially use phytic acid. However, under cellular stress, cells preferred sodium phosphate, which could be a result of phytic acid being too energetically costly to breakdown. A pSS10 construct (wildtype *HASP1* promoter and SP fused to eGFP) was also used to determine promoter function under the two various forms of phosphorus. Cells grown in high and low concentrations of phytic acid (100% and 5%, respectively) were able to secrete more protein, alluding to the promoter potentially being more responsive to phytic acid. Lastly, transcriptional analysis of wildtype and *HASP1* KO *P. tricornutum* cells suggested under nutrient depletion, many genes responsible for secondary metabolism and cell stress response were upregulated. However, the loss of *HASP1* was not detrimental to the cell, as other genes were upregulated to compensate for its loss, suggesting *HASP1* may not be essential for cell survival under nutrient deprivation.

4.2   Study Limitations

Despite the promising results of creating novel expression systems by the deletion of predicted P1BS-like regions, the predictions of where these motifs were located are based on analysis from plant promoter sequences. This suggests that there is a possibility for other regulatory regions within the promoter essential for promoter induction. Based on the 5'RACE data, it can be inferred that the predicted transcriptional start site from the original *HASP1* paper [35] may lay somewhere upstream. To confirm this, sequencing of the band seen at ~500bp should be completed to determine if the sequence aligns with where any of the mutant 2 deletions were. Furthermore, different combinations of deletions within the predicted motifs could have been constructed to determine other variations and potential differences in regulation. Due to the recent finding of the *HASP1* promoter, other limitations are the lack of specific antibodies that could be used for chromatin immunoprecipitation (ChIP) analysis. Many studies assessing regulatory elements of promoter systems used ChIP to identify proteins binding to regulatory motifs.

Another limitation across the studies was controlling for other variables. Although phosphorus depletion was our biggest concern, as time goes on other nutrients are also becoming depleted within the media, which was seen in the transcriptomic study. Transcriptomic responses were not specific to phosphorus depletion, but other metabolic processes related to cellular stress.

Lastly, the RNAseq data consisted of samples taken on day 0, 4 and 17. Cells on day 17 have not reached stationary phase. *HASP1*-driven expression has been shown to be highest in late stationary (~1 month) [43]. In the future, another analysis where time points are taken from days 0, 8, 16, 28 should be considered. Based on my research and previous studies [43], most changes in expression are seen within this time frame. Letting the cells grow for 28 days, ensures cells have reached stationary to late stationary phase where the amount of phosphorus is close to depletion. This would theoretically result in a strong induction of stress response genes involved in phosphate depletion given that other nutrients are supplemented. The amount of phosphate used in our "100%" conditions as previously mentioned is magnitudes higher than other studies, which means for phosphate to become substantially low, it would take a relatively long time. Lastly, the

upregulation of the predicted alkaline phosphatase under the *HASP1* KO condition was not seen during the RNAseq experiment, which may also suggest the culture was not completely under phosphorus stress.

4.3   Future Directions

To conclude, the availability of diatom sequences and the variety of genetic tools accessible allows us to manipulate and explore the complexity of these organisms. This research only provides a glimpse into the complexity that is the regulatory network and the ecological success of these organisms. This research provided more context about how newly discovered promoters can be used to diversify research across various disciplines and the importance of understanding regulation in any expression system. Although the *HASP1* promoter holds a promising future for synthetic biological purposes other endogenous promoters found in marine phytoplankton are still left uncharacterized and it would be worthwhile to investigate their potential.

Future work involving the *HASP1* promoter should focus on early promoter induction to minimize the duration for full promoter function and whether protein secretion can be optimized. Additionally, a long-term study assessing transcriptional changes in the *HASP1* KO could provide more insight on *HASP1* function and its importance at the proteomic level.

In conclusion, diatoms hold a promising future. Current advance in research and technology, will enable rapid, cost-efficient engineering of *P. tricornutum* and increase its potential for various biotechnological applications and further promote diatom research and development.

5   References

[1] Yang, Z. K., Zheng, J. W., Niu, Y. F., Yang, W. D., Liu, J. S., & Li, H. Y. (2014). Systems-level analysis of the metabolic responses of the diatom Phaeodactylum tricornutum to phosphorus stress. *Environmental microbiology*, *16*(6), 1793–1807.

[2] Cruz de Carvalho, M. H., Sun, H. X., Bowler, C., & Chua, N. H. (2016). Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. *The New phytologist*, *210*(2), 497–510.

[3] White, A., & Dyhrman, S. (2013). The marine phosphorus cycle. *Frontiers in microbiology*, *4*, 105.

[4] Alipanah, L., Winge, P., Rohloff, J., Najafi, J., Brembu, T., & Bones, A. M. (2018). Molecular adaptations to phosphorus deprivation and comparison with nitrogen deprivation responses in the diatom Phaeodactylum tricornutum. *PloS one*, *13*(2), e0193335.

[5] Falkowski, P. G., Barber, R. T., & Smetacek, V., V. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science (New York, N.Y.)*, *281*(5374), 200–207.

[6] Paytan, A., & McLaughlin, K. (2007). The oceanic phosphorus cycle. *Chemical reviews*, *107*(2), 563–576.

[7] Nelson D. M., Treguer, P., Brzezinsk, M, A., Leynaert, A., & Queguiner, B. (1995). Production and dissolution of biogenic silica in the ocean – revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles, 9*, 359–372.

[8] Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science (New York, N.Y.)*, *281*(5374), 237–240.

[9] Egge,J, K. (1998). Are diatoms poor competitors at low phosphate concentrations? *Journal of Marine Systems, 16*, 191–198.

[10] Morel, F. M., & Price, N. M. (2003). The biogeochemical cycles of trace metals in the oceans. *Science (New York, N.Y.)*, *300*(5621), 944–947.

[11] Dell'Aquila, G., & Maier, U. G. (2020). Specific acclimations to phosphorus limitation in the marine diatom Phaeodactylum tricornutum. *Biological chemistry*, *401*(12), 1495–1501.

[12] Dell'Aquila, G., Zauner, S., Heimerl, T., Kahnt, J., Samel-Gondesen, V., Runge, S., Hempel, F., & Maier, U. G. (2020). Mobilization and Cellular Distribution of Phosphate in the Diatom *Phaeodactylum tricornutum*. *Frontiers in plant science*, *11*, 579.

[13] Litchman, E., Klausmeier, C, A., Miller, J. R., Schofield, O. M., & Falkowski , P. G. (2006). Multi-nutrient, multi-group model of present and future oceanic phytoplankton communities. *Biogeosciences, 3*, 585–606.

[14] Kumar Sharma, A., Mühlroth, A., Jouhet, J., Maréchal, E., Alipanah, L., Kissen, R., Brembu, T., Bones, A. M., & Winge, P. (2020). The Myb-like transcription factor phosphorus starvation response (PtPSR) controls conditional P acquisition and remodelling in marine microalgae. *The New phytologist*, *225*(6), 2380–2395.

[15] Donald, K. M., Scanlan, D. J., Carr, N. G., Mann, N. H., & Joint, I. (1997). Comparative phosphorus nutrition of the marine cyanobacterium Synechococcus WH7803 and the marine diatom Thalassiosira weissflogii. *Journal of Plankton Research*, *19,* 1793-1813.

[16] Shimogawara, K., Wykoff, D. D., Usuda, H., & Grossman, A.R. (1999). Chlamydomonas reinhardtii mutants abnormal in their responses to phosphorus deprivation. *Plant Physiology, 120,* 685-694.

[17] Flynn, K. J., O¨pik, H., & Syrett, P. J. (1986). Localization of the alkaline phosphatase and 5'-nucleotidase activities of the diatom Phaeodactylum tricornutum. *Journal of General Microbiology*, *133*, 289–298.

[18]Yamaguchi, H., Arisaka, H., Otsuka, N., & Tomaru, Y. (2014). Utilization of phosphate diesters by phosphodiesterase- producing marine diatoms. *Journal of Plankton Research, 36,* 281–285.

[19] Lin, H. Y., Shih, C. Y., Liu, H. C., Chang, J., Chen, Y. L., Chen, Y. R., Lin, H. T., Chang, Y. Y., Hsu, C. H., & Lin, H. J. (2013). Identification and characterization of an extracellular alkaline phosphatase in the marine diatom Phaeodactylum tricornutum. *Marine biotechnology (New York, N.Y.)*, *15*(4), 425–436.

[20] Dyhrman, S. T., Jenkins, B. D., Rynearson, T. A., Saito, M. A., Mercier, M. L., Alexander, H., Whitney, L. P., Drzewianowski, A., Bulygin, V. V., Bertrand, E. M., Wu, Z., Benitez-Nelson, C., & Heithoff, A. (2012). The transcriptome and proteome of the diatom Thalassiosira pseudonana reveal a diverse phosphorus stress response. *PloS one*, *7*(3), e33768.

[21] Wurch, L. L., Bertrand, E. M., Saito, M. A., Van Mooy, B. A., & Dyhrman, S. T. (2011). Proteome changes driven by phosphorus deficiency and recovery in the brown tide-forming alga Aureococcus anophagefferens. *PloS one*, *6*(12), e28949.

[22]] Martin, P., Van Mooy, B. A., Heithoff, A., & Dyhrman, S. T. (2011). Phosphorus supply drives rapid turnover of membrane phospholipids in the diatom Thalassiosira pseudonana. *The ISME journal*, *5*(6), 1057–1060.

[23] Van Mooy, B. A., Fredricks, H. F., Pedler, B. E., Dyhrman, S. T., Karl, D. M., Koblízek, M., Lomas, M. W., Mincer, T. J., Moore, L. R., Moutin, T., Rappé, M. S., & Webb, E. A. (2009). Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature*, *458*(7234), 69–72.

[24] Reitan, K. I., Rainuzzo, J. R., & Olsen, Y. (1994). Effect of nutrient limitation on fatty acid and lipid content of marine microalgae. *Journal of Phycology, 30,* 972–979.

[25] Li, J., Zhang, K., Lin, X., Li, L., & Lin, S. (2022). Phytate as a Phosphorus Nutrient with Impacts on Iron Stress-Related Gene Expression for Phytoplankton: Insights from the Diatom *Phaeodactylum tricornutum*. *Applied and environmental microbiology*, *88*(2), e0209721.

[26] Lin, S., Litaker, R. W., & Sunda, W. G. (2016). Phosphorus physiological ecology and molecular mechanisms in marine phytoplankton. *Journal of phycology*, *52*(1), 10–36.

[27] Duhamel, S., Diaz, J. M., Adams, J. C., Djaoudi, K., Steck, V., & Waggoner, E.M. (2021). Phosphorus as an integral component of global marine biogeochemistry. *Nature Geoscience, 14,* 359-368.

[28] Turner, B. L., Papházy, M. J., Haygarth, P. M., & McKelvie, I. D. (2002). Inositol phosphates in the environment. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *357*(1420), 449–469.

[29] Turner, B.L., & Weckström, K. (2009). Phytate as a novel phosphorus-specific paleo-indicator in aquatic sediments. *Journal of Paleolimnology, 42,* 391–400.

[30] Richardson, B., & Corcoran, A. A. (2015). Use of dissolved inorganic and organic phosphorus by axenic and nonaxenic clones of Karenia brevis and Karenia mikimotoi. *Harmful algae*, *48*, 30–36.

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

[32] Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., … Grigoriev, I. V. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature*, *456*(7219), 239–244.

[33] Giguere, D. J., Bahcheli, A. T., Slattery, S. S., Patel, R. R., Browne, T. S., Flatley, M., Karas, B. J., Edgell, D. R., & Gloor, G. B. (2022). Telomere-to-telomere genome assembly of *Phaeodactylum tricornutum*. *PeerJ*, *10*, e13607.

[34] Butler, T., Kapoore, R. V., & Vaidyanathan, S. (2020). Phaeodactylum tricornutum: A Diatom Cell Factory. *Trends in biotechnology*, *38*(6), 606–622.

[35] Erdene-Ochir, E., Shin, B. K., Kwon, B., Jung, C., & Pan, C. H. (2019). Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from Phaeodactylum tricornutum. *Scientific reports*, *9*(1), 9941.

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

[37] 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), 13879.

[38] Huang, W., & Daboussi, F. (2017). Genetic and metabolic engineering in diatoms. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *372*(1728), 20160411.

[39] Benner, S. A., & Sismour, A. M. (2005). Synthetic biology. *Nature reviews. Genetics*, *6*(7), 533–543.

[40] Alishah Aratboni, H., Rafiei, N., Garcia-Granados, R., Alemzadeh, A., & Morones-Ramírez, J. R. (2019). Biomass and lipid induction strategies in microalgae for biofuel production and other applications. *Microbial cell factories*, *18*(1), 178.

[41] Hempel, F., Lau, J., Klingl, A., & Maier, U. G. (2011). Algae as protein factories: expression of a human antibody and the respective antigen in the diatom Phaeodactylum tricornutum. *PloS one*, *6*(12), e28424.

[42] Kadono, T., Tomaru, Y., Suzuki, K., Yamada, K., & Adachi, M. (2020). The possibility of using marine diatom-infecting viral promoters for the engineering of marine diatoms. *Plant science : an international journal of experimental plant biology*, *296*, 110475.

[43] Slattery, S. S., Giguere, D. J., Stuckless, E. E., Shrestha, A., Briere, L. K., Galbraith, A., Reaume, S., Boyko, X., Say, H. H., Browne, T. S., Frederick, M. I., Lant, J. T., Heinemann, I. U., O'Donoghue, P., Dsouza, L., Martin, S., Howard, P., Jedeszko, C., Ali, K., Styba, G., … Edgell, D. R. (2022). Phosphate-regulated expression of the SARS-CoV-2 receptor-binding domain in the diatom Phaeodactylum tricornutum for pandemic diagnostics. *Scientific reports*, *12*(1), 7010.

[44] Yarkent, C., Gürlek, C., & Oncel, S.S. (2020). Potential of microalgal compounds in trending natural cosmetics: A review. *Sustainable Chemistry and Pharmacy, 17,* 1000304.

[45] Khavari, F., Saidijam, M., Taheri, M., & Nouri, F. (2021). Microalgae: therapeutic potentials and applications. *Molecular biology reports*, *48*(5), 4757–4765.

[46] Lin, H., Yen, S., Kuo, P., Chung, C., Yeh, K., Huang, C., Chang, J., & Lin, H. (2017). Alkaline phosphatase promoter as an efficient driving element for exogenic recombinant in the marine diatom Phaeodactylum tricornutum. *Algal Research*, *23*, 58-65.

[47] Zhao, Q., Luo, Z., Chen, J., Jia, H., Ai, P., Chen, A., Li, Y., & Xu, G., (2021). Characterization of two cis-acting elements, P1BS and W-box, in the regulation of OsPT6 responsive to phosphors deficiency. *Plant Growth Regulation, 93*, 303-310.

[48] Chow, C. N., Zheng, H. Q., Wu, N. Y., Chien, C. H., Huang, H. D., Lee, T. Y., Chiang-Hsieh, Y. F., Hou, P. F., Yang, T. Y., & Chang, W. C. (2016). PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic acids research*, *44*(D1), D1154–D1160.

[49] Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P., & Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic acids research*, *30*(1), 325–327.

[50] 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., 3rd, Venter, J. C., & Weyman, P. D. (2012). Assembly of large, high G+C bacterial DNA fragments in yeast. *ACS synthetic biology*, *1*(7), 267–273.

[51] Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods, 6*(5), 343–345.

[52] Sobkowiak, L., Bielewicz, D., Malecka, E. M., Jakobsen, I., Albrechtsen, M., Szweykowska-Kulinska, Z., & Pacak, A. (2012). The Role of the P1BS Element Containing Promoter-Driven Genes in Pi Transport and Homeostasis in Plants. *Frontiers in plant science*, *3*, 58.

[53] Bowler, C., Vardi, A., & Allen, A. E. (2010). Oceanographic and biogeochemical insights from diatom genomes. *Annual review of marine science*, *2*, 333–365.

[54] Dyhrman, S.T., J.W. Ammerman, & B.A.S. Van Mooy. (2007). Microbes and the marine phosphorus cycle. *Oceanography*, *20*(2), 110–116.

[55] Kremer, C., Torres, J., Bianchi, A., Savastano, M., & Bazzicalupi, C. (2020). Myo-inositol hexakisphosphate: Coordinative versatility of natural product. *Coordination Chemistry Reviews, 419,* 213403.

[56] Program: Trimmomatic v 0.36 http://www.usadellab.org/cms/?page=trimmomatic Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

[57] Program: Hisat2 v 2.2.0 https://www.nature.com/articles/nmeth.3317 Kim, D., Langmead, B. & Salzberg, S. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods, 12*, 357–360.

[58] Program: Htseq-count v 2.2.0 https://htseq.readthedocs.io/en/master/ Putri, G.H., Anders, S., Pyl, P.T., Pimanda, J.E., & Zanini, F. (2022). Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics*, *38*(10), 2943–2945.

[59] Gloor, G.B., Macklaim, J.M., & Fernandes, A.D. (2016). Displaying Variation in Large Datasets: Plotting a Visual Summary of Effect Sizes. *Journal of Computational and Graphical Statistics*, *25*:3, 971-979.

[60] Dubey, A., Prajapati, K. S., Swamy, M., & Pachauri, V. (2015). Heat shock proteins: a therapeutic target worth to consider. *Veterinary world*, *8*(1), 46–51.

[61] Secco, D., Wang, C., Arpat, B.A., Wang, Z., Poirier, Y., Tyerman, S.D., Wu, P., Shou, H., & Whelan, J. (2012). The emerging importance of the SPX domain-containing proteins in phosphate homeostasis. *New Phytology*, *193*(4), 842-851.

[63] Alipanah, L., Winge, P., Rohloff, J., Najafi, J., Brembu, T., & Bones, A. M. (2018). Molecular adaptations to phosphorus deprivation and comparison with nitrogen deprivation responses in the diatom Phaeodactylum tricornutum. *PloS one*, *13*(2), e0193335.

[62] Liu, N., Shang, W., Li, C., Jia, L., Wang, X., Xing, G., & Zheng, W. (2018). Evolution of the *SPX* gene family in plants and its role in the response mechanism to phosphorus stress. *Open biology*, *8*(1), 170231.

[64] Osborn, H. L., & Hook, S. E. (2013). Using transcriptomic profiles in the diatom Phaeodactylum tricornutum to identify and prioritize stressors. *Aquatic toxicology (Amsterdam, Netherlands)*, *138-139*, 12–25.

[65] Cruz de Carvalho, M. H., Sun, H. X., Bowler, C., & Chua, N. H. (2016). Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. *The New phytologist*, *210*(2), 497–510.

[66] Apt, K. E., Kroth-Pancic, P. G., & Grossman, A. R. (1996). Stable nuclear transformation of the diatom Phaeodactylum tricornutum. *Molecular & general genetics : MGG*, *252*(5), 572–579.

[67] Falciatore, A., Casotti, R., Leblanc, C., Abrescia, C., & Bowler, C. (1999). Transformation of Nonselectable Reporter Genes in Marine Diatoms. *Marine biotechnology (New York, N.Y.)*, *1*(3), 239–251.

[68] Niu, Y. F., Yang, Z. K., Zhang, M. H., Zhu, C. C., Yang, W. D., Liu, J. S., & Li, H. Y. (2012). Transformation of diatom Phaeodactylum tricornutum by electroporation and establishment of inducible selection marker. *BioTechniques*, *52*(6), 10.2144/000113881.

[69] Ifuku, K., Yan, D., Miyahara, M., Inoue-Kashino, N., Yamamoto, Y. Y., & Kashino, Y. (2015). A stable and efficient nuclear transformation system for the diatom Chaetoceros gracilis. *Photosynthesis research*, *123*(2), 203–211.

[70] Muto, M., Fukuda, Y., Nemoto, M., Yoshino, T., Matsunaga, T., & Tanaka, T. (2013). Establishment of a genetic transformation system for the marine pennate diatom Fistulifera sp. strain JPCC DA0580--a high triglyceride producer. *Marine biotechnology (New York, N.Y.)*, *15*(1), 48–55.

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

[72] Velmurugan, N., & Deka, D. (2018). Transformation techniques for metabolic engineering of diatoms and haptophytes: current state and prospects. *Applied microbiology and biotechnology*, *102*(10), 4255–4267.

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

[74] Huang, W., & Daboussi, F. (2017). Genetic and metabolic engineering in diatoms. *Philosophical Transactions of the Royal Society B: Biological Science, 372*, 20160411.

[75] Hopes, A., Nekrasov, V., Kamoun, S., & Mock, T. (2016). Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant methods*, *12*, 49.

[76] Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., & Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Scientific reports*, *6*, 24951.

Appendices

Appendix A: Supplementary Figures
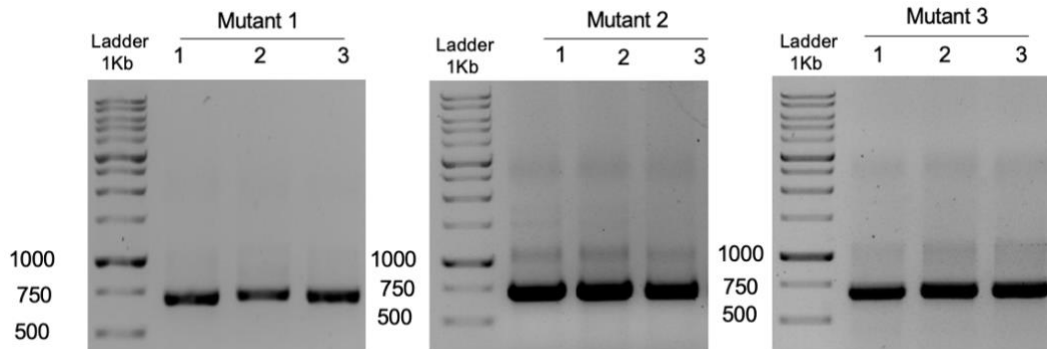
Figure A-1: PCR screen of P. tricornutum colonies after conjugation. The predicted product size for mutant 1 and mutant 2 is 711bp and for mutant 3 is 695bp.
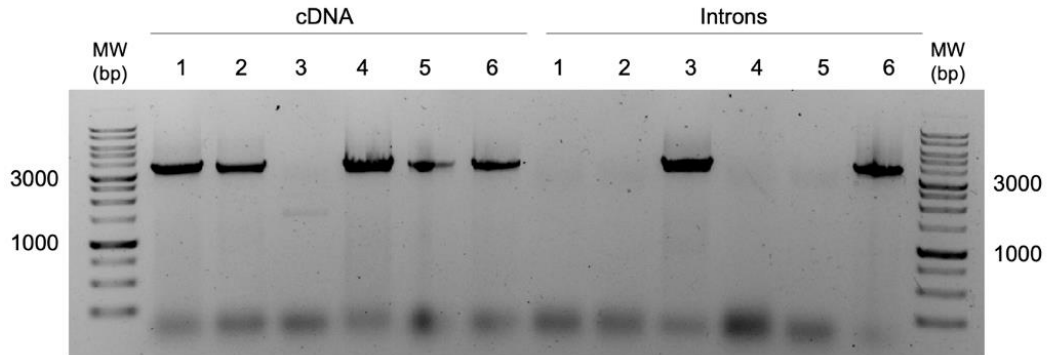
Figure A-2: PCR screen of P. tricornutum colonies after conjugation. The predicted product size for colonies containing cDNA are 2898bp and for colonies containing introns are 3062bp
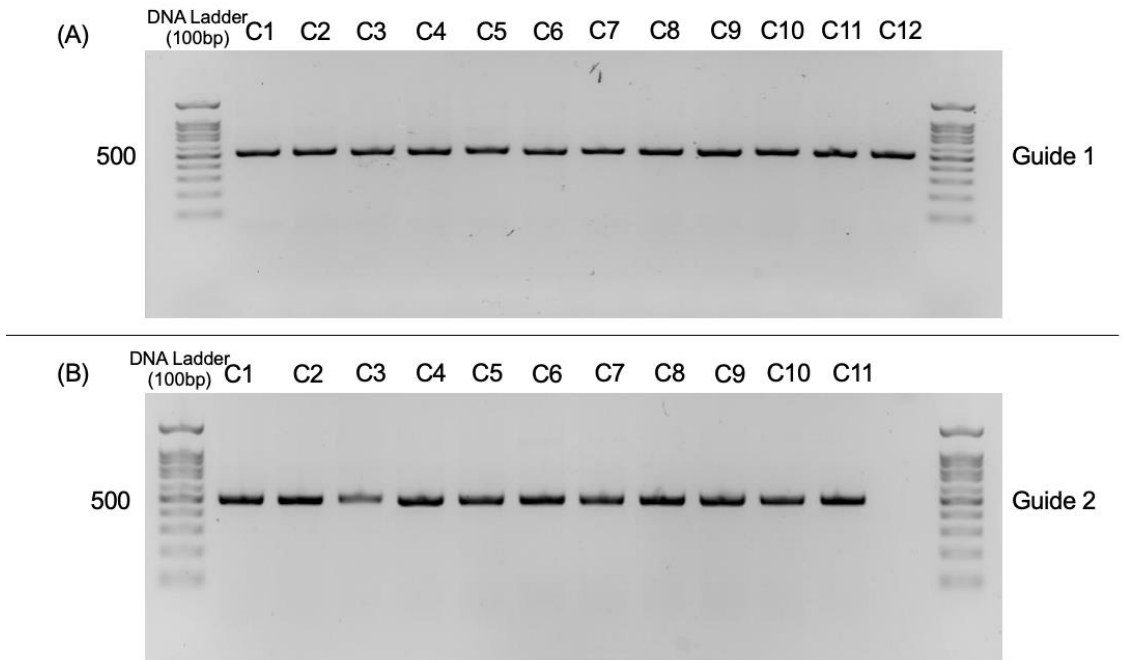
Figure A-3: PCR Screen of *P. tricornutum* colonies after conjugation. (A) PCR for pSS29 + HASP1 g1, exconjugants 1-12. (B) PCR for pSS29 + HASP1 g2, exconjugants 1-11.
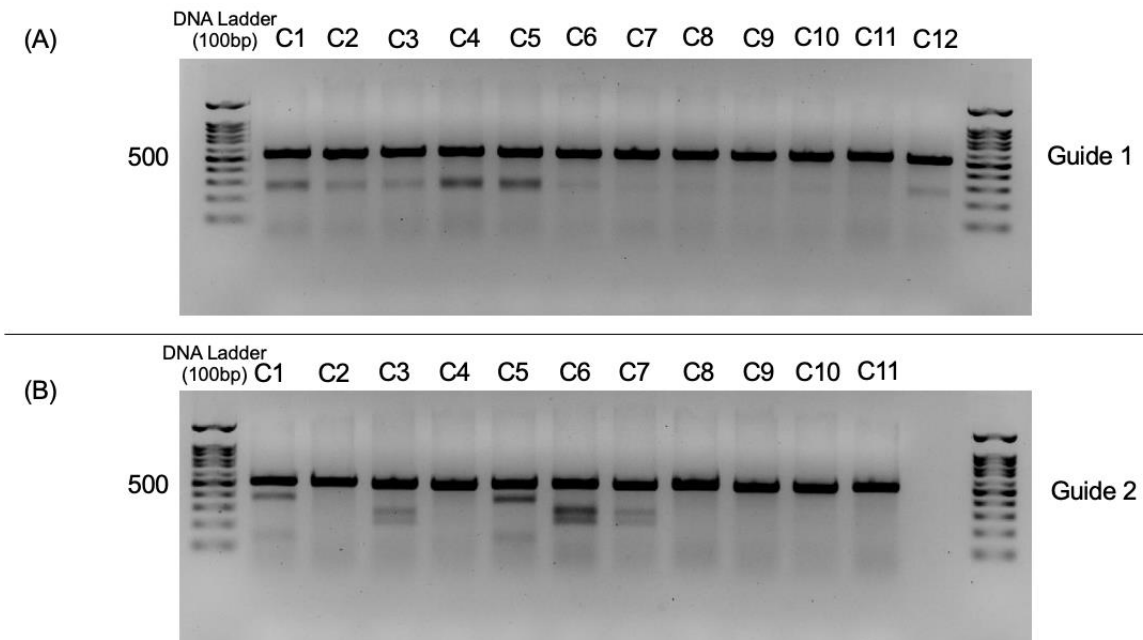
Figure A-4: T7 Endonuclease I Edited *P. tricornutum* cells. (A) T7 edited cells of pSS29 + HASP1 g1, exconjugants 1-12. (B) T7 edited cells of pSS29 + HASP1 g2, exconjugants 1-11.
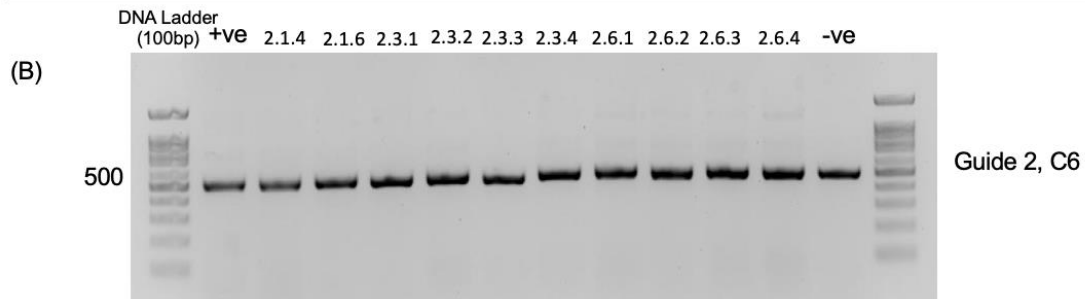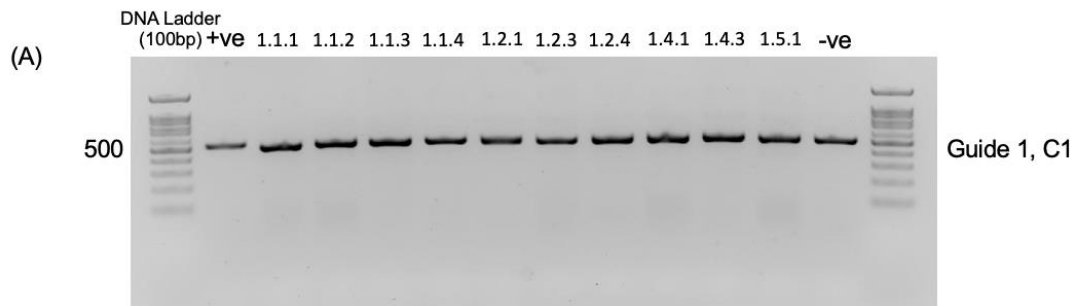
Figure A-5: PCR Screen of *P. tricornutum* subclones from edited exconjugants. (A) positive control (+ve) is exconjugant of guide 1, colony 1. Negative control (-ve) is guide 1 wildtype (B) positive control (+ve) is exconjugant of guide 2, colony 6. Negative control (-ve) is guide 2 wildtype.
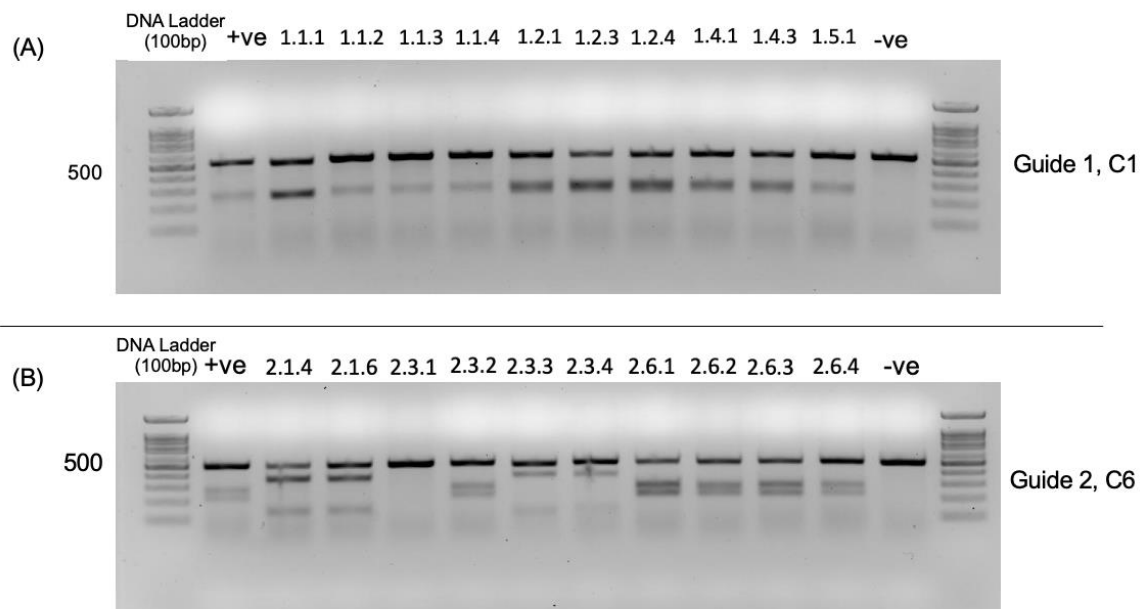
Figure A-6: T7 Screen of *P. tricornutum* subclones from edited exconjugants. (A) positive control (+ve) is exconjugant of guide 1, colony 1. Negative control (-ve) is guide 1 wildtype (B) positive control (+ve) is exconjugant of guide 2, colony 6. Negative control (-ve) is guide 2 wildtype.

Appendix B: Supplementary Tables

Table B-1: Oligonucleotides used in this study.

Name	Sequence (5' → 3')*	Description
DE5241	TTTCCCAGCTACTCGACGCATCA GGCGGGCGGATTGATAGTGAAAA CCTTATTCATTGTC	Forward primer, PCR amplification of mutant HASP1 promoters and colony screen.
DE5242	AGCCCCAGCCGAGAAG	Reverse primer, PCR amplification of mutant HASP1 promoters and colony screen.
DE5277	TGTATCCTTCCGTTTCTCCTCGCA AGCTTCTCGGCTGGGGCTATGGT GAGCAAGGGCGAG	Forward primer, PCR amplification of HASP1 SP, eGFP and CAH of pSS10 backbone for yeast assembly.
DE2742	AGGTCAAGTCCAGACTCCTGT	Reverse primer, PCR amplification of HASP1 SP, eGFP and CAH of pSS10 backbone for yeast assembly.
DE5692	GGCGCGTGTAAGTTACAGGC	Forward primer, PCR amplification of CAH, CmR and FcpA terminator of pSS10 backbone for yeast assembly.
DE5246	AGTTCGTGGGCCAAGAAACTGACG GCGCGTAAAACGCGAGAAACCAAA GCGGAGTGACTG	Reverse primer, PCR amplification of CAH, CmR and FcpA terminator of pSS10 backbone for yeast assembly.
DE5243	AAAAATTTAATTTTCATTAGTTGCAG TCACTCCGCTTTGGTTTCTCGCGTTT TACGCGCC	Forward primer, PCR amplification of FcpA terminator, HIS locus, and HASP1 promoter of pSS10 backbone for yeast assembly.

DE5244	ACCGGCTTAAGCTCTGACAATGAATA AGGTTTTCACTATCAATCCGCCCGCC TGATGCGT	Reverse primer, PCR amplification of FcpA terminator, HIS locus, and HASP1 promoter of pSS10 backbone for yeast assembly.
DE3074	AGTCCTCCCTCATAATCTCTGTTAG AGTTTACCAACAACACATATATACA TTTCGACAAAATGACCACTCTTGAC GACAC	Forward primer, PCR amplification of NatR.
DE3072	GATCTCCGAGGCCTCGGACCCGTCG GGCCGCGTCGGACCGGCGGTGTTGG TCGGCGTCGGTCAGGGGCAGGGCAT GCTCA	Reverse primer, PCR amplification of NatR for colony screen.
DE6778	CTTCCGTTTCTCCTCGCAAGCTTCTC GGCTGGGGCTTTTGTGTTCAATCGT GTGGCAACG	Fwd Primer, PCR amplification of *HASP1* and terminator for colony screen
DE6712	TATAATGACCCCGAAGCAGGGTTAT GCAGCGGAAGATTTTTATGGTCTCC ATTGCTGCGG	Rev Primer, PCR amplification of *HASP1* and terminator for colony screen
DE5481	CGTTGTATCCTTCCGTTTCTCC	Fwd Primer, PCR amplification of guide 1 for colony screen
DE5482	GAGCCAGGAGGCATCTGGGG	Rev Primer, PCR amplification of guide 2 for colony screen
DE5263	GGACTTGGAACAATCGATGT	Guide 1, PAM:AGG

Table B-1: gBlocks used in Chapter 2

Description	Sequence (5' → 3')
Mutant 1 Promoter	GGATTGATAGTGAAAACCTTATTCATTGTCAGAGCTTAAGCCGGTCTGGTCTATCTTTCCACTGTCAA ACAGCTCTTGATTGTCGCCCGCGCGAAAATAGTAGCACTAACTGTAACTTCAAAATACAAAATGTTC TCTGTTACCATACAGTGAATGTAACTTTCGAATTGACA GTATTAGTAGTCGTATTGACAGTGAGGCACGCCCCTCAATGTGCGAGGTGGACAGCATGACAATGAA TCTTTTTTGCTGTCATCAAGATTCACCGCCAAATCTTCAGGAACCTATCAC GTCCACAGGCGATGTTAATTCTTGAGTCGTCAAAACAAAGTCCTGTCCTACCTGTAGAAGTTGACAGC GAGCAATTGTATGCAAACTTCTGACTTATGTTATAATAACATTAAAGGTAATTA AGTATCTTCAATTTGGCATTTTGTCACTGTCAGTCCGTTCCGACAATATAGGTAGATTTGGAATGAATC TTTTCTATGCTGCTGCGAATCTTGTACACCTTTGAGGCCGTAGATTCTGTCCGACGA AGCGATAATTATTGCAAAATACATGGACTCATTATTTTGATTCGATTTCTTTTTGGTATCCGACTCGAA AAGATCCATCACGGCGAGC
Mutant 2 Promoter	GGATTGATAGTGAAAACCTTATTCATTGTCAGAGCTTAAGCCGGTCTGGTCTATCTTTCCACTGTCAAA CAGCTCTTGATTGTCGCCCGCGCGAAAATAGTAGCACTAACTGTAACTTCAAAATAC AAAATGTTCTCTGTTACCATACAGTGAATGTAACTTTCGAATTGACAGTATTAGTAGTCGTATTGACAG TGAGGCACGCCCCTCAATGTGCGAGGTGGAAAAATATACCAGCATGACAATGAATCTT GGAGATTCTTTTGCTGTCATCAAGATTCACCGCCAAATCTTCAGGAACCTATCACGTCCACAGGCGATG TTAATTCTTGAGTCGTCAAAACAAAGTCCTGTCCTACCTGTAGAAGTTGACAGCGAGCAATTGTATGCA AACTTCTGACTTATGTTATAATAACATTAAAGGTAATTAAAATTTGGCATTTTGTCACTGTC AGTCCGTTCCGACAATATAGGTAGATTTGGAATGAATCTTTTCTATGCTGCTGCGAATCTTGTACACCT TTGAGGCCTGTCCGACGAAGCGATAATTATTGCAAAATACATGGACTCATTATTTTGATTCGATTTCTT TTTGGTATCCGACTCGAAAAGATCCATCACGGCGAGC
Mutant 3 Promoter	GGATTGATAGTGAAAACCTTATTCATTGTCAGAGCTTAAGCCGGTCTGGTCTATCTTTCCACTGTCAAAC AGCTCTTGATTGTCGCCCGCGCGAAAATAGTAGCACTAACTGTAACTTCAAAATACAAAATGTTCTCTGT TACCATACAGTGAATGTAACTTTCGAATTGACAGTATTAGTAGTCGTATTGACAGTGAGGCACGCCCCTC AATGTGCGAGGTGGACAGCATGACAATGAATCTTTTTTGCTGTCATCAAGATTCACCGCCAAATCTTCAG GAACCTATCACGTCCACAGGCGATGTTAATTCTTGAGTCGTCAAAACAAAGTCCTGTCCTACCTGTAGAA GTTGACAGCGAGCAATTGTATGCAAACTTCTGACTTATGTTATAATAACATTAAA GGTAATTAAAATTTGGCATTTTGTCACTGTCAGTCCGTTCCGACAATATAGGTAGATTTGGAATGAATCTT TTCTATGCTGCTGCGAATCTTGTACACCTTTGAGGCCTGTCCGACGAAGCGATAATTATTGCAAAATACAT GGACTCATTATTTTGATTCGATTTCTTTTTGGTATCCGACTCGAAAAGATCCATCACGGCGAGC

Table B-2: gBlock sequences used in Chapter 3.

Description	Sequence (5' → 3')
HASP1 gene (cDNA) and terminator	TTTGTGTTCAATCGtGTGGCAACGTTCAATATTTGTACGCAGATTGATCCCACGTGTAACACCGACACCGAA GCAGTTGCCGAGATTGTTGATGCAtCACCTGACGGTAACAcTCTATTGTATACTGATGGCGTCGCTGGGAAC ATTGGCCTTGTTGACATTTcAGATCCTTCCAATCCTGTGGGACTTGGAACAATCGATGTAGGTGGTGAGCCT ACATCCGTCGCAGTGACTGGCGACGGCTTGTACGCTGTAGCAGCCGTGAACACTTCTCCTG ACTTTGTGAATGTCACCGGGAATTTGATTGTCATCGACATCGAATCCCAGTTAATTGTGGCCACCCTCCCTC TACCAGGTCAACCTGACTCACTTAAGGTTTCTCCTGACAATCAGTTCATCGCCATTGCTATAGAAAACGAAC GCGATGAAGATCTCGGGGACGGGGCCCCTCCCCAGATGCCTCCTGGCTCGTTGGTTGTTGTT GACAAGAGCGATACTGATCCAACAATGTGGTCGACGACTCTTGTCAACCTGACCGGACTCGACG

GCTTGCGATTCCCCGAGGATCCCGAGCCTGAGTTTGTCGACATCTCGACCGACAACGTCTGCGTGATCACTC
TTCAGGAAAACAACGGCATAGTTATCCTTGACCTGGCTTCTCTTAGTGTTTTGACTAGCTTTTCGGCTGGGGC
GGTGGCAGAAATCACTCAAATTGACGCCACTGAAGATGACATGATTTCGCAGACTGAAAGTCTTAACGA
CATCCTTCGCGAGCCCGACAGTGCTGTGTTCATtGGAACCGACTACTTCGCCACAGCCGATGAGGGTGACCTA
GATGGAGGGTCTCGCGGGTTTACAATTTTTGGTCTTGACGGCACCGTTGTTTACGAGTCC
GGAAACTTCTTGGAACACGCTaCTGCTCGCTACGGACACTACCCAGAAGGTCGTTCCGAGAACAAG
GGAAACGAGCCCGAGGGTATGACCTATTCAACTTTTGGTGACGAGAACCTTCTGTTTgTTCTATCCGAACGC
GCGAACACTGTCTACGTCTTTAACGTCGACCCGACAGATCCATCAAATGTTTCGTTTAAGCAAGTCCTTCCCA
GTGGTGTCGGTCCAGAGGGTGTCACGGCTATTCCAAGCAGAAATCTGTTAGCGGTCGCGTCGGAAGTGGATG
ATCGAGGTGCTGCCATTCGCAGCAGTATTACCATTTACGAATATGCCGAAGG
AACAGCCGCCTACCCTAcTCTAGTGTCAAACAACCGCGCGGATGGAaCTCCTATACCGTTTTCAGCCCTTTC
TGGACTgGCTGCTGCAGATGCGCCAGGAATAGGTGGTCAAGCCGGGATATTGTATTCTGTTGATGACTCATTC
TACTCGAAAAGCAGAATTTTCACGATCGATGTCACTTCCTTcCCTTACGCGATTaCTAATGAAATGCGTGTCAT
GGACTCCGACGGCGTCCTTGCGGCTGCTTTCCCCAACAGCACtCTTGTCAACGACGACATGACGGTCAACCTTG
ACCAAGAAGGAATTGCTGTTTCTCGGGAGGGCGGATATTGGATTGCCAACGAAGGTGCAGGTACA
GTGGGTGACATCGAGAATCCTGTCACCACTGCAAACTTTGTTGTTAAGGTGTCGACGGAGGGTGTA
ATAGAGACCGTAGTCACCTTACCtGAAGAGATAGATGCTATTCAAGTTCGCTTTGGCTTCGAAGGTATTGCTGA
GTACGGTGATTTCATTGTTGTCGCATTTCAACGCGCATGGGgAGAGGAGGCCAAT
CCGCGATTGGGGATCTACGATACAACTTCTAGTACATGGAAATTTGTCTTCTACCCCCTCGATACGCCCTCTTC
TCAGAACGGAGGATGGGTCGGTATTGGAGATCTTTCTCCAGTCGGCAATGGCGACTTTGTGGTCCTTGAGCGCG
ACAATCAAGGAGGACCTGATGCCGCCATCAAAAAGATATACAAAATaAGCCTTGGCTCGATGCTT
GAGGTCGAGGAAAGTACTACGGTCGAGAAGACTTTGGTGAGGGACGTCTTACCAGACCTTGCTGCGTCGAATG
GCAATATTTATGAGAAACTTGAGGGTCTTGCTGTGACTAGCGAAGGAGAAGTGTGGATCAACAACGACAACGA
TGGCGTTGACGACAACAGTGGGGAGCAGCAGCTTATTAGTCTGGGAATGATCGTCACCATTCCCGAGGGA
ACAGCAACTCCTATTGAAGACTCCCCGACTGtGTCGCCCGTGCAAGGAACTAGTATGGCCAATCGAGTAGGCC
ACGGAGGTGCTCTTCTC
GTCTTCGGCTTCTTTGCTTTATTGTGAGTATAGAAGGTCACGTAGCTAGAGAAAGAAAGAGTTAATCTCAAAA
AGAGTCAAGTCCACGAACTTCAAATAAACTATAAGCCTTCTTTCGTTGACCCATTGGCCTTCATC
CTTCCAATCAACACTCGATTCTGAACAGAAAATTGGACCGGGAATGATTCCTGCGATACTGTAAGCGATTTCC
CTACCTGATTCCTCGATGCCATTAAATCTTCGGCAGCTGCGCGGCGCATTGCGTTCAGTGCTCTAATCGCCCTA
TGCTGTCCACCAGATGAAGCTCCGGTTGAAAGAACCGAAACACGCCTAATATCGTCACGAAATTTTTCTAATG
ACTTTTGGTTGGCATCGCTGCGAAACATGAAATCAGCACGACGGGTCAAACACGCCAGAATCTGGTTTGACCT
GAAGCAGCAAGTAGATATACAGAACGAAGTCTGTGCTCGAAGAGCAATTTCTACAGCAGCTTCTGCAAGGCCT
CCGCAGCAATGGAGACCATAAAA

HASP1 gene (with introns) and terminator	TTTGTGTTCAATCGTGTGGCAACGTTCAATATTTGTACGCAGATTGATCCCACGTGTAACACCGACACCGAAGC AGTTGCCGAGATTGTTGATGCATCACCTGACGGTAACACTCTATTGTATACTGATGGCGTCGCTGGGAACATTG GCCTTGTTGACATTTCAGATCCTTCCAATCCTGTGGGACTTGGAACAATCGATGTAGGTGGTGAGCCTACATCC GTCGCAGTGACTGGCGACGGCTTGTACGCTGTAGCAGC CGTGAACACTTCTCCTGACTTTGTGAATGTCACCGGGAATTTGATTGTCATCGACATCGAATCCCAGTTAATTGT GGCCACCCTCCCTCTACCAGGTCAACCTGACTCACTTAAGG TTTCTCCTGACAATCAGTTCATCGCCATTGCTATAGAAAACGAACGCGATGAAGATCTCGGGGACGGGGCCCCT CCCCAGATGCCTCCTGGCTCGTTGGTTGTTGTTGACAAGAGCGATACTGATCCAACAATGTGGTCGACGACTCT TGTCAACCTGACCGGACTCGACGGCTTGCGATTCCCCGAGGATCCCGAGCCTGAGTTTGTCGACATCTCGACCGA CAACGTCTGCGTGATCACTCTTCAGGAAAACAACGGCATAGTTATCCTTGACCTGGCTTCTCTTAGTGTTTTGA CTAGCTTTTCGGCTGGGGCGGTGGCAGAAATCACTCAAATTGACGCCACTGAAGATGACATGATTTCGCAGACTG AAAGTCTTAACGACATCCTTCGCGAGCCCGACAGTGCTGTGTTCATTGGAACCGACTACTTCGCCACAGCCGATG AGGGTGACCTAGATGGAGGGTCTCGCGGGTTTACAATTTTTGGTCTTGACGGCACCGTTGTTTACGAGTCCGGAA ACTTCTTGGAACACGCTACTGCTCGCTACGGACACTACCCAGAAGGTCGTTCCGAGAACAAGGGAAACGAGCCC GAGGGTATGACCTATTCAACTTTTGGTGACGAGAACCTTCTGTTTGTTCTATCCGAACGCGCGAACACTGTCTACG TCTTTAACGTCGACCCGACAGATCCATCAAATGTTTCGTTTAAGCAAGTCCTTCCCAGTGGTGTCGGTCCAGAG GGTGTCACGGCTATTCCAAGCAGAAATCTGTTAGCGGTCGCGTCGGAAGTGGATGATCGAGGTGCTGCCATTCGC AGCAGTATTACCATTTACGAATATGCCGAAGGAACAGCCGCCTACCCTACTCTAGTGTCAAACAACC

GCGCGGATGGAACTCCTATACCGTTTTCAGCCCTTTCTGGACTGGCTGCTGCAGGTAAGTG
CACAGGCTCCCCTCTCTCTTCCCACATACTTACAGAGATATTTTCAAATCTCTGACTTTGTGTCTCTTTCGTGCATC
GCTAGATGCGCCAGGAATAGGTGGTCAAGCCGGGATATTGTATTCTGTTGATGACTCATTCTACTCGAAAAGCAG
AATTTTCACGATCGATGTCACTTCCTTCCCTTACGCGATTACTAATGAAATGCGTGTCATGGACTCCGACGGCGTC
CTTGCGGCTGCTTTCCCCAACAGCACTCTTGTCAACGACGACATGACGGTCAACC
TTGACCAAGAAGGAATTGCTGTTTCTCGGGAGGGCGGATATTGGATTGCCAACGAAGGTGCAGGTACAGTGG
GTGACATCGAGAATCCTGTCACCACTGCAAACTTTGTTGTTAAGGTGTCGACGGAGGGTGTAATAGAG
ACCGTAGTCACCTTACCTGAAGAGATAGATGCTATTCAAGTTCGCTTTGGCTTCGAAGGTATTGCTGAGTACGG
TGATTTCATTGTTGTCGCATTTCAACGCGCATGGGGAGAGGAGGCCAATCCGCGATTGGGGATCTACGATACAAC
TTCTAGTACATGGAAATTTGTCTTCTACCCCCTCGATACGCCCTCTTCTCAGAACGGAGGATGGGTCGGTATTGGA
GATCTTTCTCCAGTCGGCAATGGCGACTTTGTGGTCCTTGAGCGCGACAATCAAGGTA
AGACTTGCACGCTGCTGCGTGACACTTAAACTGATTCTCTTTCGCTCATTGTGTACTCAACTTTTACAGGAGGACC
TGATGCCGCCATCAAAAAGATATACAAAATAAGCCTTGGCTCGATGCTTGAGGTCGAGGAAAGTACTACGGTC
GAGAAGACTTTGGTGAGGGACGTCTTACCAGACCTTGCTGCGTCGAATGGCAATATTTATGA
GAAACTTGAGGGTCTTGCTGTGACTAGCGAAGGAGAAGTGTGGATCAACAACGACAACGATGGCGTTGACGACA
ACAGTGGGGAGCAGCAGCTTATTAGTCTGGGAATGATCGTCACCATTCCCGAGGGAACA
GCAACTCCTATTGAAGACTCCCCGACTGTGTCGCCCGTGCAAGGAACTAGTATGGCCAATCGAGTAGGCCACGGA
GGTGCTCTTCTCGTCTTCGGCTTCTTTGCTTTATTGTGAGTATAGAAGGTCACGTAGCTAGAGAAAGAAAGAGTTA
ATCTCAAAAAGAGTCAAGTCCACGAACTTCAAATAAACTATAAGCCTTCTTTCGTTG
ACCCATTGGCCTTCATCCTTCCAATCAACACTCGATTCTGAACAGAAAATTGGACCGGGAATG
ATTCCTGCGATACTGTAAGCGATTTCCCTACCTGATTCCTCGATGCCATTAAATCTTCGGCAGCTGCGCGGCGCAT
TGCGTTCAGTGCTCTAATCGCCCTATGCTGTCCACCAGATGAAGCTCCGGTTGAAAGAACCGAAACACGCCTAAT
ATCGTCACGAAATTTTTCTAATGACTTTTGGTTGGCATCGCTGCGAAACATGAAATCAGCACGACGGGTCAAACA
CGCCAGAATCTGGTTTGACCTGAAGCAGCAAGTAGATATACAGAACGAAGTCTGTGCTCGAAGAGCAATTT
CTACAGCAGCTTCTGCAAGGCCTCCGCAGCAATGGAGACCATAAAA

Curriculum Vitae

Name:	Mahsa Farmanbar
Post-secondary Education and Degrees:	The University of Western Ontario London, Ontario, Canada 2016-2021 HBSc. The University of Western Ontario London, Ontario, Canada 2021-2023 MSc.
Courses Taken:	BIOCHEM 9501: Scientific Writing BIOCHEM 9545Q: Bioinformatics I BIOCHEM 9516T: Cancer Biology
Honours and Awards:	Chair's Travel Award $500, 2023 Western Graduate Research Scholarship $1000 (2021, 2022) Dean's Honor List 2018-2021
Related Work Experience	Teaching Assistant The University of Western Ontario 2021-2023 Private Biochemistry Student Tutor 2021-2023 Internship Lactalis Canada, London, ON, Canada August 2019-2020
Presentations:	(Poster) London Health Research Day Title: Characterizing DNA regulatory elements of the Highly Abundant Secreted Protein 1 (*HASP1*) promoter in *Phaeodactylum tricornutum.* University of Western Ontario, London, ON, Canada June 2023

(Poster) International Conference on Algal Biomass, Biofuels and Bioproducts (AlgalBBB)
Title: Characterizing DNA regulatory elements of the Highly Abundant Secreted Protein 1 (*HASP1*) promoter in *Phaeodactylum tricornutum.*
Waikōloa Beach, Hawaii, USA
June 2023

(Talk) Biochemistry Graduate Research Spring Symposium
Title: *HASP1*: A Story About Phosphate Regulation, Loss (of Function) and Regulatory Networks
University of Western Ontario, London, ON, Canada
May 2023

(Poster) Harold Steward Research Day
Title: Phosphate Regulation in the diatom *Phaeodactylum tricornutum*
University of Western Ontario, London, ON, Canada
January 2022

Leadership and Extracurricular Activities:

Science Rendezvous Biochemistry Booth Leader
University of Western Ontario, London, ON, Canada
May 2022, 2023

Little Sisters in Science Mentorship Program
University of Western Ontario, London, ON, Canada
September 2021-2022