UNIVERSITY OF CALIFORNIA, SAN DIEGO

Insights into the molecular regulation of growth and carbon flux in marine diatoms

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

in

Marine Biology

by

Raffaela Mikiko Abbriano Burke

Committee in charge:

Mark Hildebrand, Chair
Andrew Allen
Michael Burkart
Brian Palenik
Martin Tresguerres

2017

The Dissertation of Raffaela Mikiko Abbriano Burke is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2017

DEDICATION

To my first teachers in life -

My grandmother, Elsie Hiraoka, who taught by example the values of independence and continual self-improvement.

My grandfather, Kazunori Hiraoka, who taught me to express my creativity and to appreciate all the beautiful and interesting things in this world.

My mother, Karen Abbriano, who has always encouraged me to keep on exploring.

EPIGRAPH

“This is a simple thing to say, but the profound feeling of it made a Jesus, a St. Augustine, a St. Francis, a Roger Bacon, a Charles Darwin, and an Einstein. Each of them in his own tempo and with his own voice discovered and reaffirmed with astonishment the knowledge that all things are one thing and that one thing is all things - plankton, a shimmering phosphorescence on the sea and the spinning planets and an expanding universe, all bound together by the elastic string of time. It is advisable to look from the tide pool to the stars and then back to the tide pool again.”

John Steinbeck

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SUPPLEMENTAL FILES

Supplemental File 1-1. Spreadsheet containing primer sequences used to validate intron retention events by RT-PCR.

Supplemental File 1-2. Spreadsheet containing gene expression data (DESeq2 normalized values) and annotation information for the silicon starvation timecourse.

Supplemental File 1-3. Spreadsheet containing gene expression data (DESeq2 normalized values) and annotation information for the synchrony timecourse.

Supplemental File 1-4. Spreadsheet containing a list of AUGUSTUS genes potentially derived from horizontal gene transfer as predicted by the DarkHorse program.

Supplemental File 1-5. Spreadsheet containing a list of AUGUSTUS genes with significant intron retention as detected by the rMATS program.

Supplemental Figure 2-1. Comparison of mitochondrial and peroxisomal labeling in *T. pseudonana.*

Supplemental Figure 2-2. Phylogenetic analysis of the 2-hydroxyacid dehydrogenase proteins.

Supplemental Figure 2-3. Phylogenetic tree of NAD(P) binding sites in malic enzyme proteins.

Supplemental Figure 2-4. Extension of *T. pseudonana* gene model Thaps3_34030 based on Illumina RNAseq data.

Supplemental File 2-5. Manually curated protein sequences in FASTA format.

Supplemental Table 2-1. Intracellular targeting predictions for malic enzyme genes in sequenced diatom genomes.

Supplemental Table 2-2. Primer sequences used to create expression constructs.

Supplemental Figure 3-1. RNAseq read coverage supporting modified gene models for PFK2/F2BP proteins.

Supplemental Figure 3-2. Construct maps for genetic manipulation of PFK2/F2BP proteins in *T. pseudonana.*

Supplemental Figure 3-3. Fluorescent images of PFK2-2 overexpression lines and GFP controls.

Supplemental File 3-1. Spreadsheet containing quantitative FAME profile data for PFK2-2 overexpression lines and wild-type control.

ACKNOWLEDGEMENTS

First and foremost, I would like to express deep gratitude to Dr. Mark Hildebrand for his guidance and support throughout this process. Mark is a dedicated and patient mentor who always finds the time to give to his students. I am grateful for the level of independence he granted me in his lab, which allowed me to pursue a diversity of projects. His enthusiasm for science is admirable as well as contagious, and our office brainstorming sessions are among the highlights of my time at SIO. His model of hard work and commitment to research is something that I will carry with me for the rest of my career. I would also like to thank the members of my committee, Dr. Andrew Allen, Dr. Michael Burkart, Dr. Brian Palenik, and Dr. Martin Tresguerres, for their thoughtful insights and unique perspectives that have greatly expanded my horizons. I am especially grateful for their willingness to provide access to their laboratory's resources and expertise, which have been critical to the completion of this thesis. Additionally, I would like to thank other faculty mentors at UCSD/SIO and JCVI that have offered valuable guidance and advice: Dr. Bianca Brahamsha, Dr. Eric Allen, Dr. Maria Vernet, Dr. Lena Gerwick, Dr. Terry Gaasterland, Dr. Peter Franks, and Dr. Philip Weyman.

My graduate funding was provided in part by the Department of Energy Office of Science Graduate Fellowship. A special thanks to Dr. Ping Ge and the ORISE staff for managing the program and organizing annual meetings for fellowship recipients. A special thanks to the donors of the Robert Buzzelli Fellowship, Jeff Graham Fellowship,

Krinsk Research Advancement Fellowship, and the Ralph Lewin Fellowship who helped to support me financially in the last year of my Ph.D.

To my coauthors and collaborators, it has been a pleasure to work with and learn from you. A special thanks to the Pellegrini and Merchant labs at UCLA (Dr. David Lopez, Dr. Jing Lu, Dr. Roberto Spreafico, Dr. Matteo Pellegrini, Dr. Sean Gallaher), who contributed significantly to the RNA sequencing and bioinformatic analysis. Thank you to Dr. Joris Beld, who took the time to teach a biologist some chemistry and assisted with FAME analysis for multiple projects. Thank you to Dr. Sheila Podell for sharing your time and computational expertise. Thank you to Dr. Nurcan Vardar for her assistance with the growth experiments and biochemical analyses. I would also like to thank Dr. Justin Ashworth, Dr. Jacob Valenzuela, and collaborators at the Institute of Systems Biology for their work on past and continuing research projects.

To the Hildebrand lab (past and present), thank you for being an amazing group of people to work with. I've truly enjoyed my time at SIO because of you. A special thanks to Aubrey Davis, Roshan Shrestha, and Orna Cook for willingness to mentor me on countless molecular techniques and for helping me to become a better biologist. Thank you to Jesse Traller, Sarah Lerch, Sarah Smith, Emily Trentacoste, Eva Sanchez Alvarez, Olia Gaidarenko, Daniel Yee, Corinne Sathoff, and Maitreyi Nagarkar (office mate and honorary lab member) for all the feedback, support, and friendship.

Thank you to my friends at SIO, and especially to the 2010 cohort, for all the laughs along the way. Special thanks to my SIO twin Lani Gleason, Neal Arakawa, and Corey Jew (the other one) for all the good food and adventures. Thanks to Trevor Joyce and Noah Ben-Aderet for their friendship and accurate surf forecasts. Additional thanks

to Doug Krause, Charles Perretti, Tessa Pierce, Stephanie Snyder, Amanda Netburn, and the SIO beach volleyball crew. Thank you to Rachel Morrison (we miss you), and all the members of Rachel Morrison's Maniacs intramural softball team.

Thank you to my family for their continual support throughout the years. To Katha Burke, Ashlee Moreno, Jim and Viola Burke, Pat and Bob Wood, the Anderson family (Mike, Jess, Hanalei, Vivee, Maizie, and Harvi) and the Leonard family (Samson, Dante, Theodore), thank you for being such a strong source of love and support throughout this dissertation process. To my sisters Gianna and Chiara, thank you for being the people I can always count on for anything. To my mom Karen, I know I wouldn't be where I am now without the sacrifices you made to make sure I had every opportunity to succeed. You are an amazing woman, and I will always be grateful for your belief in me.

Finally, thank you to my husband Joel, who deserves special recognition for his continual love and understanding through busy days, long nights, and stressful times. Joel, you have been my anchor through this entire process and my gratitude for all the ways that you have enriched my life cannot be overstated. Thank you for everything and I am so excited to start the next chapter of our lives together.

The introduction, in part, appears in Algal Research 2012. Smith, SR; Abbriano, R.; Hildebrand, M. The dissertation author contributed to the Algal Research manuscript of which Sarah Smith is the primary author.

The introduction, in part, appears in Current Opinion in Chemical Biology 2013. Hildebrand, M.; Abbriano, R.; Polle, J.; Traller, J.; Trentacoste, E.; Smith, S.; and Davis,

A. The dissertation author is the primary author of the material in the introduction; Mark Hildebrand is the primary author of the published manuscript.

Chapter 1, in part, is currently being prepared for submission for publication of the material in 2017. Abbriano, R.; Spreafico, R.; Lopez, D.; Lu, J.; Traller, J.; Shrestha, R.; Tesson, B.; Podell, S.; Hovde, B.; Allen, E.; Pellegrini, M.; and Hildebrand, M. The dissertation author was the primary investigator and author of this material.

Chapter 1, in part, appears in New Phytologist 2016. Smith, S., Glé, C., Abbriano, R., Traller, J., Davis, A., Vernet, M., Trentacoste, E., Allen, A., and Hildebrand, M. The dissertation author is the primary investigator of the data reported in Chapter 1; Sarah Smith is the primary author of the published manuscript.

Chapter 2, in full, is a reprint of the material as it appears in Protist in 2016. Davis, A.; Abbriano, R.; Smith, S.R.; Hildebrand, M. The dissertation author was the co-primary investigator and author of this material.

Chapter 3, in part, is currently being prepared for submission for publication of the material in 2017. Abbriano, R.; Vardar, N.; Yee, D.; Hildebrand, M. The dissertation author was the primary investigator and author of this material.

Chapter 3, in part, appears in Algal Research 2012. Smith, S., Abbriano, R., and Hildebrand, M. The dissertation author is the primary investigator of the data reported in Chapter 3; Sarah Smith is the primary author of the published manuscript.

VITA

2008 Bachelor of Arts with Honors, University of San Diego

2014 Master of Science, University of California, San Diego

2017 Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Abbriano, R., Spreafico, R., Lopez, D., Lu, J., Traller, J., Shrestha, R., Tesson, B., Podell, S., Hovde, B., Allen, E., Pellegrini, M., and Hildebrand, M. RNA sequencing improves gene models and reveals alternative splicing in the model diatom *Thalassiosira pseudonana*. In prep.

Abbriano, R., Vardar, N., Yee, D., and Hildebrand, M. Regulation of growth and carbon partitioning by phosphofructo-2-kinase in the marine diatom *Thalassiosira pseudonana*. In prep.

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

Davis, A.[*], **Abbriano, R.**[*], and Hildebrand, M. (2016). Clarification of photorespiratory processes and the role of malic enzyme in diatoms. Protist, 168: 134-153. ([*]authors contributed equally)

Smith, S., Glé, C., **Abbriano, R.**, Traller, J., Davis, A., Vernet, M., Trentacoste, E., Allen, A., and Hildebrand, M. (2016). Transcript level coordination of carbon pathways during silicon-starvation induced lipid accumulation in the diatom *Thalassiosira pseudonana*. New Phytol, 210: 890-904.

Beld, J., **Abbriano, R.**, Finzel, K., Hildebrand, M., and Burkart, M. (2016). Probing fatty acid metabolism in bacteria, cyanobacteria, green microalgae, and diatoms with natural and unnatural fatty acids. Mol Biosyst 12(4): 1299-1312.

Hildebrand, M., Davis, A., **Abbriano, R.**, Pugsley, H., Traller, J., Smith, S., Shrestha, R., Cook, O., Sánchez-Alvarez, E.L., Manandhar-Shrestha, K., and Alderete, B. (2015). Applications of imaging flow cytometry for microalgae. In Barteneva, N.S. and Vorobjev, I.A., (eds.) Imaging Flow Cytometry (pp. 47-69) New York, NY: Springer New York.

Hildebrand, M., **Abbriano, R.**, Polle, J., Traller, J., Trentacoste, E., Smith, S., Davis, A. (2013). Metabolic and cellular organization in evolutionarily diverse microalgae as related to biofuels production. Curr Opin Chem Biol 17(3): 506-514.

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

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

Abbriano, R., Carranza, M., Hogle, S., Levin, R., Netburn, A., Seto, K., Snyder, S. Franks, P. (2011). Deepwater Horizon oil spill: a review of the planktonic response. Oceanography 24 (3): 294-301.

ABSTRACT OF THE DISSERTATION

Insights into the molecular regulation of growth and carbon flux in marine diatoms

by

Raffaela Abbriano Burke

Doctor of Philosophy in Marine Biology

University of California, San Diego, 2017

Mark Hildebrand, Chair

Diatoms are a highly productive group of phytoplankton collectively responsible for up to 40% of the annual organic carbon production in the world's oceans. Photosynthetic carbon fixation by diatoms sustains diverse marine ecosystems and contributes significantly to the global carbon cycle. The natural productivity of diatoms, coupled with high hydrocarbon yield, makes this group an attractive option for the

production of next-generation biofuels and other valuable bioproducts. The need for a renewable alternative to fossil fuels has recently become apparent due to the environmental and political costs of petroleum production, including the inevitability of peak oil, the costs of obtaining foreign fuel sources, and rising atmospheric CO_2 levels that exacerbate climate change. However, an increased understanding of the molecular mechanisms that control intracellular carbon partitioning in diatom cells is critical to develop successful strategies to optimize future production strains.

Our current understanding of the organization and regulation of carbon metabolism in diatoms relies heavily on genomic information from a few model species. However, recent advancements in 'omics' analyses are expanding our understanding the dynamics of diatom metabolism based on gene expression, protein abundance, and metabolite levels. Despite increasing accessibility to these strategies, much remains to be understood about the molecular mechanisms and their regulation that underlie diatom growth productivity. Therefore, the broad objective of my dissertation is to gain further insight into the controls of intracellular carbon assimilation, partitioning, and storage in diatoms. This will be accomplished using several different approaches: a bioinformatics refinement of gene models and analysis of splicing during diatom cell cycle arrest and progression (Chapter 1), an investigation into the organization of proteins involved in photorespiratory metabolism (Chapter 2), and an investigation of the influence of a regulatory protein (6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase) on intracellular carbon partitioning (Chapter 3).

INTRODUCTION

I.1 Diatom productivity and global significance

Marine diatoms are ubiquitous unicellular microalgae that constitute a dominant fraction of eukaryotic phytoplankton in the modern oceans. Diatoms are collectively responsible for up to 40% of the annual marine organic carbon production (Nelson et al. 1995) and are major carriers of carbon to the deep ocean (Armbrust 2009). Diatoms tend to dominate phytoplankton communities in coastal regions, where they can bloom rapidly in response to nutrient influxes from upwelling events. In addition, diatoms are of significant ecological importance, as primary production from diatoms support a variety of diverse marine ecosystems.

I.2 Diatoms as photosynthetic cell factories

Microalgae offer a potential means to sustainably produce food, energy, and industrial feedstocks and are an attractive alternative to traditional plant, fungal or animal cell culture-based systems for production of valuable bioproducts such as vaccines, antibodies, specialty oils, and novel carotenoids (Spolaore et al. 2006; Hallmann 2007). Advantages of microalgal culture include rapid photosynthetic growth, inexpensive culture medium, and high biomass productivity, all of which can translate to a significant reduction in production costs at scale (Georgianna and Mayfield 2012).

The rapid growth rates and high biomass productivity of some marine diatom species identified them as particularly suitable options for certain industrial and biotechnological applications. The value of diatoms for the production of economically

important compounds such as omega-3 fatty acids and carotenoids has long been recognized, and many species are also capable of producing valuable hydrocarbons that are suitable for fuel production (Fig. I-1; Nelson et al. 1995; Hildebrand et al. 2012). The suitability of several diatom strains for biofuel feedstock was outlined in the pioneering biofuel research conducted during the DOE funded Aquatic Species Program (Sheehan et al. 1998). However, large scale production of renewable biofuels from diatoms or other microalgae faces considerable challenges, and genetic engineering of strains will likely be an important component in the future development of microalgal agriculture (Hildebrand et al. 2012; Georgianna and Mayfield 2012). An increased understanding of the controls on intracellular carbon partitioning is critical to develop successful metabolic engineering strategies to increase bioproduct yields in diatoms.

I.2 Diatom evolution and metabolism

The evolution of modern phytoplankton is punctuated by major endosymbiotic events that convey large-scale genetic changes to a host organism. The capability to perform oxygenic photosynthesis originated in cyanobacteria and was transferred to the eukaryotic lineage approximately 1.5 billion years ago (Armbrust 2009). In this process, the cyanobacterial cell was appropriated as new membrane-bound organelle within the host cell known as the plastid (Fig. I-2), and genes from the cyanobacterial symbiont were transferred to the host nucleus (Timmis et al. 2004). This primary endosymbiotic event gave rise to the photosynthetic algal lineages, including glaucophytes, red algae, and green algae, the latter of which eventually gave rise to vascular plants (Falkowski 2004).

Diatoms belong to a taxonomic group known as the Stramenopiles, which arose through a secondary endosymbiotic event that occurred approximately 1 billion years ago (Fig. I-2) and involved the engulfment and retention of a red algal cell by a heterotrophic eukaryote (Armbrust 2009). Evidence for this secondary endosymbiotic event includes the presence of genes of red algal origin in the nuclear genome of diatoms (Bowler et al. 2008) and a distinctive organization of the chloroplast membrane (McFadden 2001).

The genetic merger that occurred following the secondary endosymbiotic event created a mosaic genome comprised of genes from various sources (Falkowski 2004; Bowler et al. 2008) and a distinctive organization of metabolism in diatoms relative to other photosynthetic eukaryotes (Armbrust et al. 2004; Kroth et al. 2008; Allen et al. 2011; Smith et al. 2012). In addition, diatom genomes have been further shaped by horizontal gene transfer, gene replacement and retargeting, selective loss or expansion of gene families, and a rapid evolutionary rate (Bowler et al. 2008). These genetic modifications over time have altered the metabolic capabilities and routes for intracellular carbon flux in diatom cells (Figure I-2). Major organizational features that affect the routes of carbon flux in diatoms are detailed below.

I.3.1 Increased compartmentation

Diatom plastids are surrounded by four membranes (Fig. I-3); the inner two are the chloroplast inner and outer membranes typical of all photosynthetic eukaryotes, the third is the periplastid membrane (the relic plasma membrane of the red algal endosymbiont), and the outermost membrane, commonly called the ‘chloroplast ER’, is continuous with the endoplasmic reticulum (Bolte et al. 2009; Keeling 2009). The

periplastid and outer chloroplast membranes delineate the periplastid compartment (PPC), which adds additional structural complexity around diatom chloroplasts. The PPC has a distinct proteome (Moog et al. 2011) that includes carbonic anhydrases (Moog et al. 2011; Tachibana et al. 2011), and therefore is thought to have a role in the acquisition of inorganic carbon (Edward Lee and Kugrens 1998).

The multiple membranes around the diatom chloroplast necessitate the use of unique multipartite N-terminal signals consisting of both a signal and a transit peptide for complete transfer of proteins in to the chloroplast (Apt et al. 2002; Kilian and Kroth 2005; Gruber et al. 2007; Bolte et al. 2009; Hempel et al. 2009). Metabolite exchange between the chloroplast stoma and the cytosol requires an additional set of transporters, which is an additional consideration when considering routes of carbon flux in diatoms as transport may be an important mode of regulating substrate availability (Smith et al. 2012).

I.3.2 Organization of central carbon metabolism

The primary pathways that are involved in the partitioning of carbon (the distribution of photosynthetic products throughout the cell) are glycolysis and gluconeogenesis, which work in opposing directions to utilize or store carbon intermediates depending on the metabolic and energetic demands of the cell. Glycolysis catabolizes glucose to produce pyruvate and ATP, while gluconeogenesis converts pyruvate to glucose to supply carbohydrate biosynthesis pathways. While glycolysis and gluconeogenesis share many of the same enzymes that are regulated by mass action, carbon flux through glycolysis/gluconeogenesis is strictly regulated at metabolic

checkpoints that control the directionality of carbon flux through these pathways (Plaxton 1996). In addition to their roles in carbohydrate metabolism, glycolytic enzymes are necessary to generate precursors for fatty acid, isoprenoid, amino acid, and nucleotide biosynthesis and are key enzymes in carbon fixation (Lunn 2006).

The Embden-Meyerhof-Parnas (EMP) version of glycolysis (the canonical pathway widespread in eukaryotes) is typically considered to be a cytosolic pathway, however, photosynthetic eukaryotes contain at least a partial glycolytic pathway in the plastid (Ginger et al. 2010). In both chlorophytes and diatoms, EMP glycolysis typically not complete in either the plastid or the cytosol (Fig. I-3), necessitating the transport of glycolytic intermediates across plastid membranes (Ginger et al. 2010; Terashima et al. 2011; Smith et al. 2012). Diatoms have additional EMP glycolysis capabilities in the mitochondria (Fig. I-3; Liaud et al. 2000; Kroth et al. 2008), an extremely unique arrangement that is known to occur only in diatoms and non-photosynthetic oomycetes (Ginger et al. 2010). One potential advantage of this metabolic arrangement may be the ability to produce the reducing equivalents needed to feed oxidative phosphorylation directly in the mitochondria, eliminating the need to import them from the cytosol (Smith et al. 2012). In addition, recent work had identified a Entner–Doudoroff (ED) glycolytic pathway, a prokaryotic variant of glycolysis, that may allow for the complete catabolism of C6 compounds to pyruvate in diatom mitochondria (Fabris et al. 2012).

Another major organizational difference in diatom metabolism is the localization of the oxidative pentose phosphate pathway (OPP), which supplies ribose-5-phosphate for *de novo* nucleotide biosynthesis with the concomitant production of NADPH. In green algae and higher plants, this pathway is co-localized to the plastid with the

reductive pentose phosphate pathway (Calvin-Benson cycle enzymes), necessitating strict light-responsive regulation of enzyme activity to avoid futile cycling (Ruelland and Miginiac-Maslow 1999). In diatoms, OPP and nucleotide biosynthesis occur in the cytosol, potentially liberating OPP from strict temporal regulation (Fig I-2; Michels et al. 2005; Kroth et al. 2008; Gruber et al. 2009). The cytosolic OPP may serve to provision the mitochondrial ED pathway, which utilizes the OPP intermediate 6-phosphogluconate. Reducing equivalents generated via cytosolic OPP in diatoms may be utilized in fatty acid biosynthesis, but would have to be imported into the plastid.

The organization of carbohydrate storage and pathways related to carbohydrate biosynthesis are also unique in diatoms. Eukaryotic phytoplankton utilize a diversity of strategies pertaining to the storage of intracellular carbohydrate, which have important metabolic and energetic implications for the cell (Hildebrand et al. 2013). Diatoms store a soluble ß-(1,3)-linked glucan called chrysolaminarin in a cytoplasmic vacuole (Fig. I-3; Chiovitti et al. 2004), whereas green algae store carbohydrate as insoluble starch granules within the plastid (Deschamps et al. 2008). The localization of carbohydrate storage in the plastid puts restrictions on when carbohydrate breakdown can occur in order to avoid competition with carbon fixation (Deschamps et al. 2008) and may explain differences in temporal usage of carbohydrates to fuel growth and division (Hildebrand et al. 2017).

I.4 Functional genomics and molecular engineering in diatoms

The suite of genetic tools available for the engineering of microalgal strains is comparatively limited relative to other plant and microbial systems, which has previously impeded their use as industrial production strains. However, significant progress has been

made in the development of genetic tool for diatoms, including the adoption of the Gateway cloning system (Siaut et al. 2007), expansion of available selectable markers and reporter genes (Zaslavskaia et al. 2000; Poulsen et al. 2006; Shrestha et al. 2013), use of constitutive and inducible promoters (Dunahay et al. 1995; Kilian and Kroth 2005; Poulsen 2005; Sakaue et al. 2008; Hildebrand 2008; Russo et al. 2015), gene knockdown by antisense RNA and RNAi (De Riso et al. 2009), and genome editing/gene knockout using TALENs (Weyman et al. 2015) and CRISPR Cas9 (Nymark et al. 2016).

The development of new tools has been expedited by the publication of several diatom genomes, including the centric diatom *Thalassiosira pseudonana* (Armbrust et al. 2004) and pennate diatom *Phaeodactylum tricornutum* (Bowler et al. 2008). Subsequent transcriptome studies have been critical to the connection of diatom genome sequences with gene function, and have resulted in numerous insights about how diatoms respond to changes in environmental factors such as light, temperature, CO_2, and nutrient status (Mock et al. 2008; Allen et al. 2008; Sapriel et al. 2009; Maheswari et al. 2010; Shrestha et al. 2012; Thamatrakoln et al. 2012; Ashworth et al. 2013; Bender et al. 2014; Smith et al. 2016b; Smith et al. 2016a).

The insights garnered through increased accessibility of high-throughput, next generation data has informed several successful efforts to engineer diatom metabolism for biotechnological purposes, including increased production of biofuel-relevant lipids (Radakovits et al. 2011; Trentacoste and Shrestha 2013), nutritional oils (Hamilton et al. 2014; Cook and Hildebrand 2016), and recombinant proteins (Hempel et al. 2011). Recent technological advances in diatom biotechnology, including the design of an extra-chromosomal replicating vector that enables stable transgene expression in diatoms and

improved transformation efficiency via bacterial conjugation with *E. coli* (Karas et al. 2015), will enable rapid advances in the functional characterization of diatom genes. The increasing availability of high-throughput data and the expansion of the molecular tool kit positions diatoms to become an important photosynthetic system for the heterologous expression of biochemical pathways.

I.5 Dissertation contents

The chapters in this dissertation investigate knowledge gaps in the molecular regulation of diatom growth and metabolism using three diverse approaches, 1) a global, bioinformatics analysis of gene models and alternative splicing, 2) detailed analysis of metabolic pathway organization using *in silico* and *in vivo* validation of gene targeting, and 3) functional characterization of a regulatory protein (6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase) by reverse genetics. Detailed summaries for each chapter are included below.

Chapter 1 describes a collaborative project to leverage transcriptomic data collected during experimental variation in silicon availability to improve gene model annotation in *T. pseudonana*. High-throughput RNA sequencing was used to inform *ab initio* gene annotation programs, permitting prediction of novel or modified gene boundaries with high coverage at single base pair resolution. We report 13,399 predicted genes in the *T. pseudonana* genome, over half of which are modified from the current reference annotation (Thaps3 filtered models, Joint Genome Institute) including 1,883 genes with unique loci. 37.5% of genes in the unique model set are differentially expressed in response to both silicon starvation and silicon readdition experiments,

representing a set of novel genes that include DNA-binding proteins, putative cell wall associated proteins, and cyclins that are likely related to silicon availability and cell cycle processes. In addition, Chapter 1 includes the first assessment of alternative splicing (AS) in the *T. pseudonana* transcriptome. We found AS is dominated by intron retention (IR), a mechanism related to the fine-tuning of gene expression rather than an increase in proteomic diversity. Approximately 3% of genes in the genome were significantly regulated by changes in IR in response to changes in environmental silicon availability. Gene model improvements, coupled with analysis of alternative splicing in *T. pseudonana*, aides in the interpretation of future transcriptomic datasets and contributes to the development of more accurate metabolic models in diatoms.

Chapter 2 is an in-depth investigation into enzymes related to photorespiratory and glyoxylate metabolism in *T. pseudonana.* Photorespiration represents a major route of carbon flux in photosynthetic cells, and the organization and extent of photorespiratory metabolism in diatoms is currently unresolved. In Chapter 2, we combined *in silico* sequence analysis, transcript expression, and *in vivo* localization of proteins putatively involved in these processes to update the model for the organization of C2 photorespiratory metabolism in *T. pseudonana* and to propose a mechanism for their operation under silicon starvation-induced cell cycle arrest. Our analysis provides support for distinctive aspects of photorespiratory metabolism in diatoms. These features include the mitochondrial localization of the C2 photorespiratory pathway, including a unique glycolate-oxidizing enzyme (GOX2), as well as the absence of a glycerate kinase needed to return processed glycolate to the chloroplast. In addition, we propose upregulation of the glyoxylate cycle can operate in coordination with mitochondrial malic enzyme to

serve as an alternative source of pyruvate in the mitochondria under silicon starvation conditions.

Chapter 3 focuses on the characterization of carbon flux regulation by 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatases (PFK2/F2BP) in *T. pseudonana*. PFK2/F2BP is responsible for intracellular levels of the signaling molecule fructose-2,6-bisphosphate, which is known to regulate glycolytic flux in other eukaryotes. In Chapter 3, we provide evidence for a unique diatom PFK2/F2BP with a distinct functional role in the control of intracellular carbon partitioning. Overexpression resulted in higher levels of neutral lipids and proteins and less carbohydrate, showing that alteration of carbon partitioning between glycolysis and gluconeogenesis affects the metabolic fate of fixed carbon in diatoms, and that protein and lipid accumulation can be elicited in diatoms in the absence of environmental cues (such as nutrient limitation). Higher glycolytic activity also resulted in a reduced growth rate and extension of the G1 phase of the cell cycle, suggesting that redirection of central carbon flux in diatoms can feed back on cell cycle progression.

ACKNOWLEDGEMENTS:

The introduction, in part, appears in Algal Research 2012. Smith, SR; Abbriano, R.; Hildebrand, M. The dissertation author contributed to the Algal Research manuscript of which Sarah Smith is the primary author.

The introduction, in part, appears in Current Opinion in Chemical Biology 2013. Hildebrand, M.; Abbriano, R.; Polle, J.; Traller, J.; Trentacoste, E.; Smith, S.; and Davis, A. The dissertation author is the primary author of the material in the introduction; Mark

Hildebrand is the primary author of the Current Opinion in Chemical Biology manuscript.

FIGURES

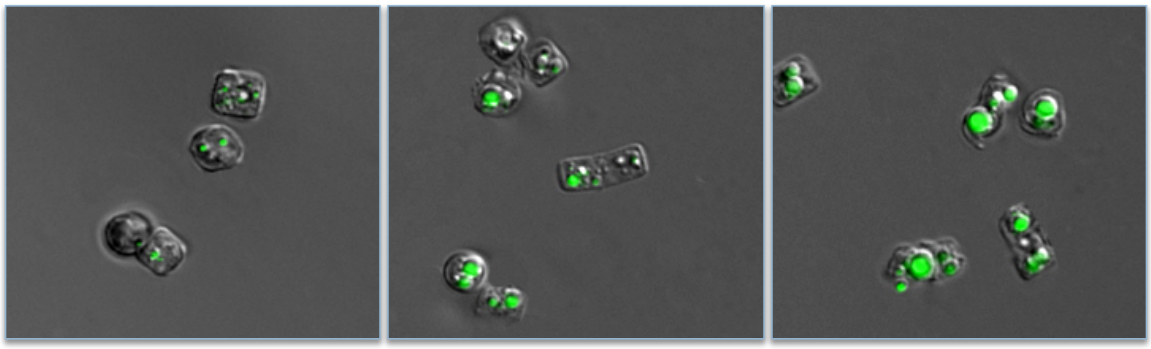

Figure I-1. *T. pseudonana* cells stained with the fluorescent, lipophilic dye BODIPY over a six day timecoure of phosphate limitation. Artificial green color indicates staining of neutral lipids.

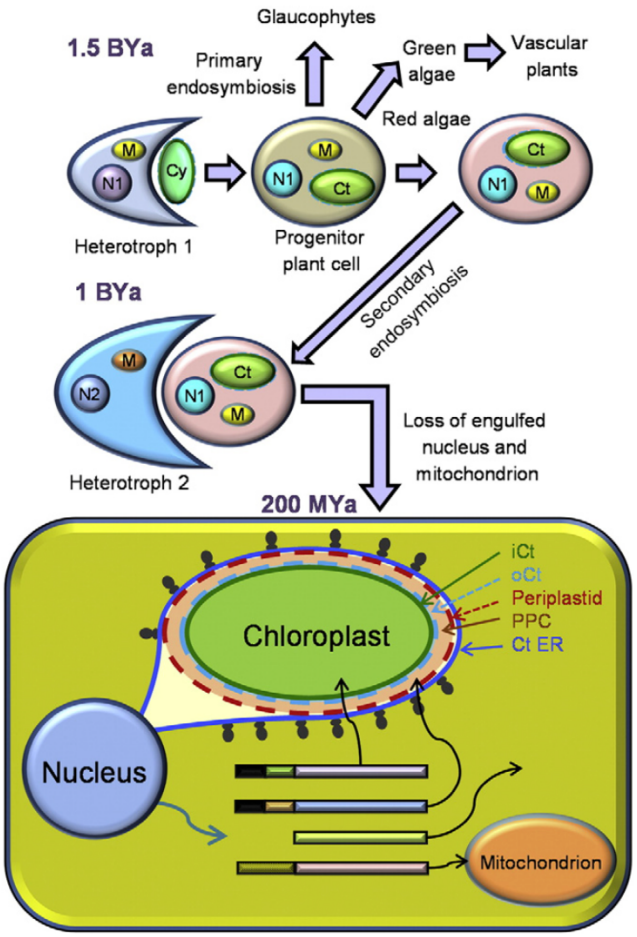

Figure I-2. The primary and secondary endosymbiotic events that gave rise to modern diatoms. The upper panel shows the evolutionary progression leading to the heterokonts. Organelles are colored to denote different origins and labeled as such (N = nucleus, M = mitochondria, Ct = chloroplast, Cy = cyanobacterium). The lower panel is a diagram of diatom intracellular compartmentation, with an emphasis on the chloroplast and associated extra membranes relative to the progenitor plant cell. Colored bars represent proteins with different leader sequences targeting them to different intracellular locations. The inner chloroplast membrane (iCt), outer chloroplast membrane (oCt), periplastid membrane, periplastid compartment (PPC), and chloroplast ER (ct ER) are labeled.

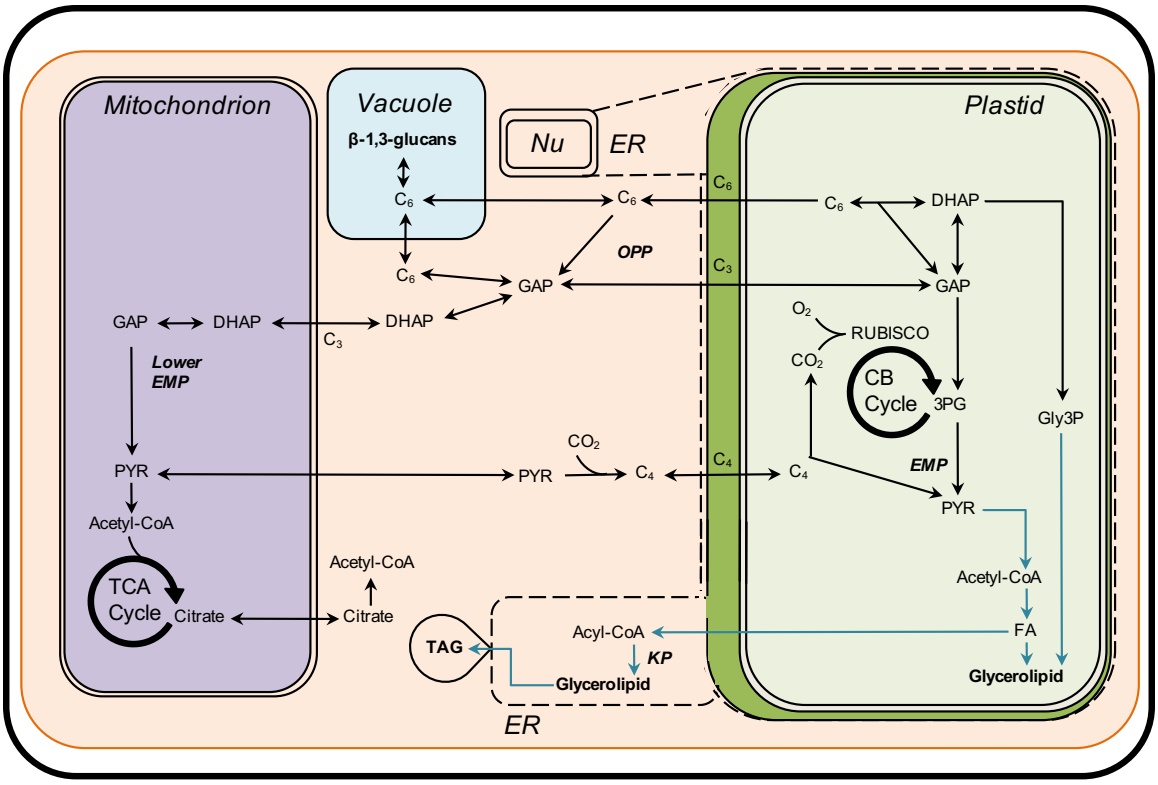

Figure I-3. Overview of the organization of central carbon metabolism in *T. pseudonana*. Organelles: ER = endoplasmic reticulum; Nu = nucleus. Pathways: CB = Calvin–Benson; EMP = Emden–Meyerhoff–Parnas glycolysis; KP = Kennedy pathway; OPP = oxidative pentose phosphate; TCA = citric acid cycle. Metabolism: 3PG = 3-phosphoglycerate; DHAP = dihydroxyacetone phosphate; FA = fatty acid; GAP = glyceraldehyde-3-phosphate; Gly3P = glycerol-3-phosphate; PYR = pyruvate; RuBisCO = ribulose-bisphosphate carboxylase oxygenase; TAG = triacylglycerol. C3, 4, 6 refer to carbon skeletons containing the specified number of carbon molecules.

REFERENCES

Allen AE, Dupont CL, Obornik M, Horák A, Nunes-Nesi A, McCrow JP, Zheng H, Johnson DA, Hu H, Fernie AR, Bowler C (2011) Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature 473:203–207.

Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C (2008) Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. Proceedings of the National Academy of Sciences 105:10438–10443.

Apt KE, Zaslavkaia L, Lippmeier JC, Lang M, Kilian O, Wetherbee R, Grossman AR, Kroth PG (2002) In vivo characterization of diatom multipartite plastid targeting signals. J Cell Sci 115:4061–4069.

Armbrust EV (2009) The life of diatoms in the world's oceans. Nature 459:185–192.

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306:79–86.

Ashworth J, Coesel S, Lee A, Armbrust EV, Orellana MV, Baliga NS (2013) Genome-wide diel growth state transitions in the diatom *Thalassiosira pseudonana*. Proceedings of the National Academy of Sciences 110: 7518-7523.

Bender SJ, Durkin CA, Berthiaume CT, Morales RL, Armbrust EV (2014) Transcriptional responses of three model diatoms to nitrate limitation of growth. Front Microbio 1:1–15.

Bolte K, Bullmann L, Hempel F, Bozarth A, Zauner S, Maier U-G (2009) Protein targeting into secondary plastids. Journal of Eukaryotic Microbiology 56:9–15.

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret FED, Berges JA, Brownlee C, Cadoret J-P, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet JER, Haruta M, Huysman MJJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, ger NKO, Kroth PG, La Roche J, Lindquist E, Lommer M, quel VERMNJEZE, Lopez PJ, Lucas S, Mangogna M, McGinnis K,

Medlin LK, Montsant A, Le Secq M-PON, Napoli C, Obornik M, Parker MS, Petit J-L, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, Dassow von P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456:239–244.

Chiovitti A, Molino P, Crawford SA, Teng R, Spurck T, Wetherbee R (2004) The glucans extracted with warm water from diatoms are mainly derived from intracellular chrysolaminaran and not extracellular polysaccharides. European Journal of Phycology 39:117–128.

Cook O, Hildebrand M (2016) Enhancing LC-PUFA production in Thalassiosira pseudonana by overexpressing the endogenous fatty acid elongase genes. J Appl Phycol 28:897–905.

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

Deschamps P, Haferkamp I, d'Hulst C, Neuhaus HE, Ball SG (2008) The relocation of starch metabolism to chloroplasts: when, why and how. Trends in Plant Science 13:574–582.

Dunahay TG, Jarvis EE, Roessler PG (1995) Genetic transformation of the diatoms *Cyclotella cyrptica* and *Navicula saprophila*. Journal of Phycology 31:1004–1012.

Edward Lee R, Kugrens P (1998) Hypothesis: the ecological advantage of chloroplast ER – the ability to outcompete at low dissolved CO_2 concentrations. Protist 149:341–345.

Fabris M, Matthijs M, Rombauts S, Vyverman W, Goossens A, Baart GJE (2012) The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. The Plant Journal 70:1004–1014.

Falkowski PG (2004) The Evolution of modern eukaryotic phytoplankton. Science 305:354–360.

Georgianna DR, Mayfield SP (2012) Exploiting diversity and synthetic biology for the production of algal biofuels. Nature 488:329–335.

Ginger ML, McFadden GI, Michels PAM (2010) Rewiring and regulation of cross-compartmentalized metabolism in protists. Philosophical Transactions of the Royal Society B: Biological Sciences 365:831–845.

Gruber A, Vugrinec S, Hempel F, Gould SB, Maier U-G, Kroth PG (2007) Protein targeting into complex diatom plastids: functional characterisation of a specific

targeting motif. Plant Molecular Biology 64:519–530.

Gruber A, Weber T, Bártulos CR, Vugrinec S, Kroth PG (2009) Intracellular distribution of the reductive and oxidative pentose phosphate pathways in two diatoms. J Basic Microbiol 49:58–72.

Hallmann A (2007) Algal transgenics and biotechnology. Transgenic Plant Journal 1:81-98.

Hamilton ML, Haslam RP, Napier JA, Sayanova O (2014) Metabolic engineering of *Phaeodactylum tricornutum* for the enhanced accumulation of omega-3 long chain polyunsaturated fatty acids. Metabolic Engineering 22:3–9.

Hempel F, Bullmann L, Lau J, Zauner S, Maier UG (2009) ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. Molecular Biology and Evolution 26:1781–1790.

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:e28424.

Hildebrand M (2008) Development of new genetic manipulation tools for metabolic engineering of diatoms. Air Force Office of Scientific Research Final Report.

Hildebrand M, Abbriano RM, Polle J, Traller JC, Trentacoste EM, Smith SR, Davis AK (2013) Metabolic and cellular organization in evolutionarily diverse microalgae as related to biofuels production. Current Opinion in Chemical Biology 17:1-9.

Hildebrand M, Davis AK, Smith SR, Traller JC (2012) The place of diatoms in the biofuels industry. Biofuels 3:221–240.

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

Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips APR, Noddings CM, Brunson JK, Valas RE, Deerinck TJ, Jablanovic J, Gillard JTF, Beeri K, Ellisman MH, Glass JI, Hutchison CA III, Smith HO, Venter JC, Allen AE, Dupont CL, Weyman PD (2015) Designer diatom episomes delivered by bacterial conjugation. Nature Communications 6:6925.

Keeling PJ (2009) Chromalveolates and the evolution of plastids by secondary endosymbiosis. Journal of Eukaryotic Microbiology 56:1–8.

Kilian O, Kroth P (2005) Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. The Plant Journal 41:175–183.

Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, Stanley MS, Kaplan A, Caron L, Weber T, Maheswari U, Armbrust EV, Bowler C (2008) A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. Plos One 3:e1426.

Liaud M, Lichtl C, Apt K, Martin W (2000) Compartment-specific isoforms of TPI and GAPDH are imported into diatom mitochondria as a fusion protein: evidence in favor of a mitochondrial origin of the eukaryotic glycolytic pathway. Mol Biol Evol 17:213-223.

Lunn JE (2006) Compartmentation in plant metabolism. Journal of Experimental Botany 58:35–47.

Maheswari U, Jabbari K, Petit J-L, Porcel BM, Allen AE, Cadoret J-P, De Martino A, Heijde M, Kaas R, La Roche J, Lopez PJ, Martin-Jezequel V, Meichenin A, Mock T, Schnitzler Parker M, Vardi A, Armbrust EV, Weissenbach J, Katinka M, Bowler C (2010) Digital expression profiling of novel diatom transcripts provides insight into their biological functions. Genome Biol 11:R85.

McFadden G (2001) Primary and secondary endosymbiosis and the origin of plastids. Journal of Phycology 37:951–959.

Michels A, Wedel N, Kroth P (2005) Diatom plastids possess a phosphoribulokinase with an altered regulation and no oxidative pentose phosphate pathway. Plant Physiol 137:911–920.

Mock T, Samanta MP, Iverson V, Berthiaume C, Robison M, Holtermann K, Durkin C, BonDurant SS, Richmond K, Rodesch M, Kallas T, Huttlin EL, Cerrina F, Sussmann MR, Armbrust EV (2008) Whole-genome expression profiling of the marine diatom Thalassiosira pseudonana identifies genes involved in silicon bioprocesses. Proceedings of the National Academy of Sciences 105:1579–1584.

Moog D, Stork S, Zauner S, Maier UG (2011) In silico and in vivo investigations of proteins of a minimized eukaryotic cytoplasm. Genome Biology and Evolution 3:375–382.

Nelson DM, Tréguer P, Brzezinski MA, Leynaert A, Quéguiner B (1995) Production and dissolution of biogenic silica in the ocean - revised global estimates, comparison with regional data and relationship to biogenic sedimentation. Global Biogeochem Cy 9:359–372.

Nymark M, Sharma AK, Sparstad T, Bones AM, Winge P (2016) A CRISPR/Cas9 system adapted for gene editing in marine algae. Sci Rep 6:24951.

Plaxton W (1996) The organization and regulation of plant glycolysis. Annu Rev Plant Phys 47:185–214.

Poulsen N (2005) A new molecular tool for transgenic diatoms. FEBS Journal 272:3413-3423.

Poulsen N, Chesley PM, Kroger N (2006) Molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (Bacillariophyceae). Journal of Phycology 42:1059–1065.

Radakovits R, Eduafo PM, Posewitz MC (2011) Genetic engineering of fatty acid chain length in *Phaeodactylum tricornutum*. Metabolic Engineering 13:89–95.

Ruelland E, Miginiac-Maslow M (1999) Regulation of chloroplast enzyme activities by thioredoxins: activation or relief from inhibition? Trends in Plant Science 4:136–141.

Russo MT, Annunziata R, Sanges R, Ferrante MI, Falciatore A (2015) The upstream regulatory sequence of the light harvesting complex Lhcf2 gene of the marine diatom *Phaeodactylum tricornutum* enhances transcription in an orientation- and distance-independent fashion. Mar Genomics 1–11.

Sakaue K, Harada H, Matsuda Y (2008) Development of gene expression system in a marine diatom using viral promoters of a wide variety of origin. Physiol Plantarum 133:59–67.

Sapriel G, Quinet M, Heijde M, Jourdren L, Tanty V, Luo G, Le Crom S, Lopez PJ (2009) Genome-wide transcriptome analyses of silicon metabolism in *Phaeodactylum tricornutum* reveal the multilevel regulation of silicic acid transporters. Plos One 4:e7458.

Sheehan J, Dunahay TG, Beneman JR, Roessler PG (1998) Look back at the U. S. Department of Energy's Aquatic Species Program: Biodiesel from Algae; Close-Out Report. 1–328.

Shrestha RP, Haerizadeh F, Hildebrand M (2013) Molecular Genetic Manipulation of Microalgae: Principles and Applications. John Wiley & Sons, Ltd, Oxford, UK

Shrestha RP, Tesson B, Norden-Krichmar T, Federowicz S, Hildebrand M, Allen AE (2012) Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. BMC Genomics 13:499.

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

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

Smith SR, Gillard JTF, Kustka AB, McCrow JP, Badger JH, Zheng H, New AM, Dupont CL, Obata T, Fernie AR, Allen AE (2016a) Transcriptional orchestration of the

global cellular response of a model pennate diatom to diel light cycling under iron limitation. PLoS Genet 12:e1006490.

Smith SR, Glé C, Abbriano RM, Traller JC, Davis A, Trentacoste E, Vernet M, Allen AE, Hildebrand M (2016b) Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. New Phytol 210:890–904.

Spolaore P, Joannis-Cassan C, Duran E, Isambert A (2006) Commercial applications of microalgae. J Biosci Bioeng 101:87–96.

Tachibana M, Allen AE, Kikutani S, Endo Y, Bowler C, Matsuda Y (2011) Localization of putative carbonic anhydrases in two marine diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. Photosynth Res 109:205-221.

Terashima M, Specht M, Hippler M (2011) The chloroplast proteome: a survey from the *Chlamydomonas reinhardtii* perspective with a focus on distinctive features. Curr Genet 57:151–168.

Thamatrakoln K, Korenovska O, Niheu AK, Bidle KD (2012) Whole-genome expression analysis reveals a role for death-related genes in stress acclimation of the diatom *Thalassiosira pseudonana*. Environ Microbiol 14:67–81.

Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5:123–135.

Trentacoste EM, Shrestha RP, Smith, SR, Glé, C, Hartmann, AC, Hildebrand, M, Gerwick, WH (2013) Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. Proceedings of the National Academy of Sciences 110: 19748-53.

Weyman PD, Beeri K, Lefebvre SC, Rivera J, McCarthy JK, Heuberger AL, Peers G, Allen AE, Dupont CL (2015) Inactivation of *Phaeodactylum tricornutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. Plant Biotechnol J 13:460–470.

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

CHAPTER 1:

RNA sequencing improves gene models and reveals alternative splicing in the marine diatom *Thalassiosira pseudonana*

1.1 ABSTRACT

The publication of the *Thalassiosira pseudonana* genome as a model photosynthetic eukaryote greatly expanded our understanding of the genetic basis for the unique aspects of diatom biology. However, many of the gene model predictions are incorrect, obscuring important structural information. We combined existing models with evidence from high coverage RNAseq datasets to inform new gene model predictions using the AUGUSTUS algorithm. Over half of the 13,399 genes predicted are modified from the reference genome, including revised gene boundaries, novel splice sites, and unique loci containing 1,883 previously unannotated genes. In addition, we present the first assessment of alternative splicing (AS) in the *T. pseudonana* transcriptome. We find AS is dominated by intron retention (IR), and approximately 3% of genes are significantly regulated by changes in IR during rapid changes in silicon availability.

1.2 INTRODUCTION

Diatoms are ubiquitous eukaryotic microalgae that influence biogeochemical cycles on a global scale. Annual diatom primary productivity can account for approximately 20% of Earth's total photosynthetic output, and contribute up to 40% of the organic carbon produced in the oceans (Nelson et al. 1995). A distinctive feature of most diatoms is a siliceous cell wall, or frustule, which imposes a requirement for silicon for growth and cell division (Brzezinski et al. 1990). The high productivity of diatoms makes them attractive candidates for biotechnology, as rapid growth and natural production of valuable compounds (such as proteins and lipids) are ideal characteristics for applications such as biofuel or therapeutic protein production (Hempel et al. 2011;

Hildebrand et al. 2012). In addition, diatom biosilica may have applications in the development of designer nanomaterials (Gordon et al. 2009), as vaccine adjuvants (Nazmi et al. 2016), and in targeted drug delivery (Delalat et al. 2015).

Thalassiosira pseudonana is a coastal centric diatom with has historically served as an important model organism for studying diatom ecology and physiology. Due to its relatively small genome size (34 Mbp), it was chosen as the subject for the first diatom genome sequencing project (Armbrust et al. 2004). Gene model prediction and annotation is a computationally challenging aspect of a genome sequencing effort that is essential to link genome sequencing to the functional potential of predicted genes. The original *T. pseudonana* gene annotation relied heavily on homology-based approaches, with additional support of relatively limited EST evidence (Armbrust et al. 2004). Since the sequencing of the *T. pseudonana* genome in 2004, there have been additional efforts to optimize the gene model catalog using evidence from microarray expression (Mock et al. 2008) as well as upstream/downstream model extension (Gruber et al. 2015). Eukaryotic gene prediction has continued to evolve to accommodate new data, such as high-throughput RNA sequencing (RNAseq). In addition to providing information pertaining to gene expression, RNAseq is a relatively inexpensive approach that provides high-coverage structural information about the transcriptome (Sultan et al. 2008; Trapnell et al. 2010; Trapnell et al. 2012), such as evidence for better delimitation of exons, splice sites and alternatively spliced exons (Yandell and Ence 2012). For example, recent RNAseq studies have revealed additional transcriptome complexity and expanded previous estimates for alternative splicing in both metazoans and plants (Sultan et al. 2008; Pan et al. 2008; Filichkin et al. 2010; Trapnell et al. 2010; Marquez et al. 2012). RNAseq

evidence can be useful when paired with *ab initio* gene prediction programs to improve the annotations of protein-coding genes, especially for non-model or less studied genomes that may contain a higher proportion of non-homologous genes (Li et al. 2011; Zhao et al. 2013). Gene model improvement will be key for understanding the organization of metabolic pathways and for functional studies of diatom genes, as well as a foundation for studies into the complexity diatom transcriptome.

The alternative splicing of introns contributes to transcriptome structure and complexity in eukaryotic organisms though the production of multiple mRNAs from a single gene. Alternative splicing can also produce protein isoforms with different functional characteristics, change the subcellular localization of the protein, or alter protein stability (Reddy 2007). Therefore, alternative splicing has the potential to be a major generator of proteome diversity in eukaryotes. In plants, functional studies have begun to reveal that alternative splicing can influence a variety of metabolic and developmental processes, including (but not limited to) photosynthetic carbon fixation, defense responses, flowering time, and grain quality (Reddy 2007).

Several types of alternative splicing exist, including exon skipping, splicing of mutually exclusive exons, use of alternative donor/acceptor sites, intron retention, and trans-splicing. The extent of alternative splicing can vary greatly among organisms, and fundamental differences exist between metazoans and plants. For example, it is estimated that up to 95% of multiexon genes in humans undergo some form of alternative splicing (Pan et al. 2008), while estimates in the model plant *Arabidopsis thaliana* range from approximately 40 to 60% (Filichkin et al. 2010; Marquez et al. 2012). Furthermore, the dominant form of alternative splicing in humans is exon skipping (Sultan et al. 2008),

while the major form of alternative splicing in plants is intron retention (Ner-Gaon et al. 2004; Marquez et al. 2012). Global analyses of alternative splicing in unicellular photosynthetic eukaryotes are less prevalent, but information from a few representative species is beginning to emerge. In the unicellular green alga *Chlamydomonas reinhardtii*, the extent of alternative splicing was found to be low (~3%) and the dominant form of alternative splicing (as in plants) is intron retention (Labadorf et al. 2010). However, this pattern does not hold for all descendants of the green algal lineage. The chlorarachniophyte *Bigelowiella natans,* a secondary endosymbiont derived from the incorporation of a green-algal ancestor into a eukaryotic host cell, exhibits complex splicing throughout the transcriptome, including higher levels of exon skipping than observed in most metazoan tissues (Curtis et al. 2012). No comprehensive analysis has been done to investigate alternative splicing in diatoms, although there is evidence that it occurs in specific cases. For example, RT-PCR was used to show that the marine diatom *Chaetoceros compressum* transcribes two mRNAs from the heat stress-responsive gene HI-5 that are differently expressed at elevated temperatures (Kinoshita et al. 2001). In addition, cloning of the cDNAs derived from the DGAT1 gene in the pennate diatom *Phaeodactylum tricornutum* revealed two separate transcripts, the longer of which is likely a result of intron retention (Guihéneuf et al. 2011).

In this study, we leverage an extensive RNAseq dataset from two complementary experiments related to silicon availability and cell division in *T. pseudonana*: 1) silicon starvation, which immediately halts cell cycle progression in siliceous diatom species (Martin-Jézéquel et al. 2000; Smith et al. 2016b), and 2) readdition of silicon following starvation, which reinitiates cell cycle progression in a synchronous fashion (Darley and

Volcani 1969; Hildebrand et al. 2007; Shrestha et al. 2012). This data was used to further refine existing gene model 3' and 5' boundaries and to identify previously unannotated gene models. In addition, we investigate splicing patterns over silicon-induced cell cycle changes, and demonstrate silicon-responsive changes in the extent of alternative splicing in *T. pseudonana*.

1.3 METHODS

1.3.1 Growth conditions

Axenic cultures of *Thalassiosira pseudonana* CCMP1335 were grown in artificial seawater medium (NEPC, http://www3.botany.ubc.ca/cccm/NEPCC/esaw.html) at 18°C under continuous light (150 μmol m^{-2} s^{-1}). For silicon starvation experiments, 20L cultures were grown to a concentration of approximately $4x10^5$ cells ml^{-1}. The culture was sampled in nutrient replete conditions prior to harvesting for silicon starvation. The remaining cells were harvested by filtration, rinsed 1X in silicate-free medium, and resuspended at the same cell density. Cultures were stirred and bubbled with air under constant continuous light. Sampling was conducted at 0, 4, 8, 12, 18, and 24-hour time points following inoculation into silicon-free NEPC medium. For synchrony experiments, axenic cultures of *Thalassiosira pseudonana* (CCMP1335) were synchronized as previously described (Hildebrand et al. 2007; Shrestha et al. 2012). Exponentially-grown cells were harvested by centrifugation, rinsed once and placed in silicon-free medium. After 24 hours of Si starvation, rhodamine 123 (R123) and 200 μM sodium silicate were added to the culture. Samples were collected prior to and then every hour after silicate addition. Aliquots of R123-labeled cells were visualized under a fluorescent microscope

to monitor synchronous cell cycle progression. Experiments for both silicon starvation and synchrony were conducted in biological duplicate.

1.3.2 RNA processing and sequencing

Previously frozen cell pellets containing 3-4e^8 cells were used for RNA isolation with TRI Reagent as previously described (Hildebrand and Dahlin 2000). Quality control of RNA samples and library preparation was done as described in (Smith et al. 2016b) using the Illumina TruSeq mRNA Sample Prep kit (Illumina). Bar-coded libraries were then pooled (10 or 11 per pool), and sequenced on an Illumina HiSeq 2000 sequencer with 100 bp paired end reads.

1.3.3 Read alignment, assembly, and differential expression analysis

Reads from individual samples were demultiplexed based on a perfect barcode match using a custom perl script, and TopHat (v2.0.6) running Bowtie 2 (version 2.0.2) was used for strand-specific mapping of RNAseq reads to the *T. pseudonana* reference genome assembly obtained from the Joint Genome Institute (JGI): (http://genome.jgi.doe.gov/Thaps3/Thaps3.download.ftp.html). The genome assembly used consisted of repeat-masked finished chromosomes (Thaps3), "unmapped sequence" (Thaps3_bd) and organelle sequences. Mapping parameters were adjusted for diatom gene structure, with 40 bp and 2000 bp for minimum and maximum intron lengths, respectively. Cufflinks (v.2.2.1) was used to assemble transcripts, and the assembly was used for gene model prediction. Raw counts to AUGUSTUS generated gene models were generated using htseq-count (HTSeq 0.6.1p1). Normalized counted and differential expression analysis was done using the DESeq2_1.6.3 package. Adjusted p-values (padj)

were calculated using the likelihood ratio test (LRT) with deviance between biological replicates as the reduced model (dds = DESeq(dds, test = "LRT", reduced = ~replicate)).

1.3.4 Bioinformatic gene model prediction

Gene model predictions were generated from several pipelines as follows: (1) FGENESH using the built-in diatom training set; (2) web-based AUGUSTUS trained on the *de novo* RNA assembly; and (3) MAKER with FGENESH, AUGUSTUS, GeneMarkES analyses enabled. All prediction software was run using default settings except where noted. Several genes were selected where intron/exon boundaries were well characterized in *T. pseudonana* and used to test the accuracy of the gene model predictions. Cuffcompare v.2.2.1 (within the Cufflinks package) with default parameters was used to compare JGI and AUGUSTUS gene model predictions.

Functional annotation of unique AUGUTUS genes was performed with the BLAST2GO plug-in the Geneious software v.R8 (Kearse et al. 2012). Bioinformatic tools used to predict subcellular localization include web-based SignalP 3.0 (Nielsen and Krogh 1998; Bendtsen et al. 2004), ASAFAP (Gruber et al. 2007), and HECTAR (Gschloessl et al. 2008). The DarkHorse program version 1.5 was used to evaluate the phylogenetic distance between proteins from AUGUSTUS prediction to sequences in NCBI GenBank nr database (accessed October 16, 2015) at species and class level phylogenetic granularity. BLASTP alignments to GenBank nr sequences were required to cover at least 70% of total query length and have e-value scores of $1e^{-5}$ or lower for inclusion in this analysis.

1.3.5 Splicing analysis

2x100bp reads were hard-trimmed with Trimmomatic 2x82bp. Trimmed reads were aligned with STAR 2.4.1d using ENCODE parameters as per developer's manual, with soft clipping disabled. Equal-length reads and soft-clipping-free mappings are a requirement of rMATS. Unmapped reads and alignments with MAPQ lower than 30 were removed using samtools. Expression levels were estimated using htseq-count. Genomic tracks were generated using the bam2wig script from the RSeQC package.

Alignments were used as input to rMATS 3.0.9 to detect alternative splicing events. Because rMATS 3.0.9 did not support strand-specific data, samtools was used to split alignments between positive and negative strand before being processed by rMATS. Results from positive and negative strands were merged back into a single result table, and the global False Discovery Rate (FDR) was recomputed using R v3.1. To enable detection of all possible intron retention (RI) events, the Augustus-generated GTF file was pre-processed using a custom Python script provided by the rMATS developers. Analysis and visualization of results were performed using the R statistical language.

1.3.6 RT-PCR confirmation of intron retention

2L cultures of axenic *T. pseudonana* were grown to exponential phase ($1.4e^6$ cells ml^{-1}) for silicon starvation and synchrony experiments in biological duplicate. Cultures were harvested by centrifugation (4,000 xg), washed 1X with silicon-free medium, and resuspended in silicon-free ASW (Darley and Volcani 1969) at approximately the same cell density. After 24 hours, silicon was added back to the culture at a final concentration of 200 μM to reinitiate cell cycle progression. RNA sampling was performed as previously described on 200 ml aliquots of culture. RNA was extracted from *T.*

pseudonana pellets using TRI Reagent as described previously (Hildebrand and Dahlin 2000). Contaminating DNA was removed with the TURBO DNA-free kit (Ambion) and first-strand cDNA synthesis was performed using Superscript III reverse transcriptase (Thermo Fisher Scientific) with oligo (dT) primers. RT-PCR targets were chosen by screening for genes with at least one intron with a significant intron retention event and one intron that is completely spliced (as a control for possible DNA contamination). Primers were designed to flank the individual introns (Supplemental File 1-1) and cDNA was amplified by PCR using GoTaq DNA polymerase (Promega) and the following parameters: initial denaturation (92°C, 2 min), denaturation (92°C, 30 sec), annealing (59°C, 30 sec), elongation (72°C, 50 sec), 36 cycles. Intron presence/absence was determined by fragment size in a 2% agarose gel.

1.4 RESULTS AND DISCUSSION

1.4.1 RNA sequencing of silicon starvation and synchrony

Since silicon availability is integral to cell cycle processes in diatoms, we performed high-throughput RNA sequencing on timecourse experiments following silicon starvation and readdition in *T. pseudonana* cultures to obtain a comprehensive view of gene expression over the entire cell cycle. Cultures were sampled over a period of 26 hours for silicon starvation (7 samples) and 9 hours for readdition (10 samples) in biological duplicate, resulting in a total of 34 individual libraries prepared for 100+100 paired-end sequencing. Longer, paired reads were chosen to increase the accuracy of the splice variant predictions. Sequencing resulted in an average of 10.8 million pairs of reads per timepoint (Fig. 1-1). On average, 79.4% of these initial reads mapped uniquely

to the *T. pseudonana* genome, which equates to ~58X coverage of the genome or ~150X coverage of the estimated transcriptome.

1.4.2 RNAseq improves gene model annotation in T. pseudonana

RNAseq read alignments were processed through the gene prediction programs AUGUSTUS (Stanke and Morgenstern 2005), Fgenesh (Salamov and Solovyev 2000), and MAKER (Cantarel et al. 2008), and basic gene structure parameters were compared to the current filtered annotation for *T. pseudonana* (Table 1-1) available through the Joint Genome Institute (Thaps3). FGENESH predicts a total of 10,904 gene models, which is short of, but closest to, the JGI prediction. MAKER also predicted fewer models, but by a much larger margin (43.5% decrease). AUGUSTUS predicts 13,399 gene models (a 17.6% increase). However, it is within 55 genes of the prediction from 13,344 genes in the optimized gene catalog generated by extension and filtering of all existing *T. pseudonana* gene model predictions (Gruber et al. 2015). The mean gene length, number of exons per gene, and number of introns per gene are substantially higher (>20%) in the Fgenesh and MAKER predictions (Fig. 1-2, Table 1-1). Manual evaluation of these models revealed that merging of neighboring, but independent, gene models was a common issue that artificially increased the average gene length and decreased the total number of predicted genes. Excessive merging of genes can be a persistent problem in high density genomes with shorter intergenic distances (Yandell and Ence 2012), such as is the case in *T. pseudonana*. In addition, manual comparisons of the predictions to our RNAseq read coverage suggested that the gene architecture for AUGUSTUS models was well supported by the transcript data, while both Fgenesh and MAKER models commonly miscalled gene and intron/exon boundaries.

Based on our initial assessment of gene model quality, we selected the AUGUSTUS prediction for further comparison to the existing JGI gene models. A common error in the current JGI models is the truncation of models at the 5' end of the gene, eliminating important subcellular targeting information that is often encoded in the N-terminal portion of the protein. To test for gene model completeness, we searched for the presence of an ATG start codon at the 5' end of predicted gene models. The percentage of models that begin with an ATG was substantially higher in the AUGUSTUS predictions (99.9%) compared to the JGI models (83.2%). In addition, we subjected the models to several gene targeting prediction programs appropriate for special protein targeting considerations in heterokont algae: ASAFfind using SignalP 3.0 signal peptide predictions, and HECTAR. Diatoms possess secondary plastids, and a bipartite N-terminal sequence consisting of both a signal peptide and a transit peptide is required for protein import into the chloroplast (Gruber et al. 2015). The percentage of total proteins with signal peptides predicted by SignalP 3.0 was 22.2%, 21.9%, and 18.2%, for AUGUSTUS, optimized models (Gruber et al. 2015), and JGI models, respectively. In addition, the AUGUSTUS proteome had a higher proportion of models with chloroplast targeting (12.4% of models with chloroplast prediction; 8.0% with high confidence) compared to the JGI proteins (9.4% of models with chloroplast prediction; 5.6% with high confidence) based on ASAFind targeting predictions (Fig. 1-3A). HECTAR also predicted a higher percentage of signal peptides (13.7% vs. 12.0%), chloroplast targeting (7.0% vs. 4.8%) and mitochondrial targeting (3.8% vs. 2.7%) in AUGUSTUS models relative to the JGI models (Fig. 1-3B). Thus, our RNAseq-guided

gene model extension revealed additional 5' targeting information missed in previous annotation efforts, which should improve our knowledge of the organellar proteomes.

A major limitation of the AUGUSTUS results for *T. pseudonana* was a lack of UTR regions, which could not be reliably predicted due to relatively low sequence coverage at both ends of sequenced transcripts. Both the RNAseq alignment files and the Cufflinks transcriptome assemblies are useful resources help estimate UTR boundaries when viewed alongside gene model predictions. The RNAseq alignment files, Cufflinks transcriptome assemblies, and AUGUSTUS gene models are available for download or web browsing through Greenhouse, an omics database for algal feedstocks hosted by Los Alamos National Lab (https://greenhouse.lanl.gov/greenhouse/organisms/).

1.4.3 Computational prediction of previously unannotated gene models

The program Cuffcompare was used to compare the loci of our AUGUSTUS dataset to existing JGI models (Trapnell et al. 2012). The results indicate that 48% of AUGUSTUS models are identical to Thaps3, while 38% are modified (overlapping loci with different gene boundaries). An additional 14% (1,883 genes) are predicted to be unique to the AUGUSTUS set. The majority of the unique gene set have no overlap with existing gene models (strict set; 1,300 genes), while the remainder are predicted to fall entirely within Thaps3 intronic regions or occur on the strand opposite the reference gene (inclusive set; 583 genes) (Fig. 1-4, Supplemental Files 1-2, 1-3).

Previous work leveraging microarray, EST, and proteomic data under multiple environmental conditions has provided support for as many as 3,470 transcriptional units that likely correspond to genes not included in the JGI Thaps3 dataset (Mock et al. 2008). Using this approach, it was estimated that the total number of genes in *T. pseudonana*

could be as high as 14,860, which is higher than our estimate by 1,461 genes. Potential reasons for the higher prediction in the Mock set may be transcript data from other conditions (e.g. nitrogen starvation), as well as the inclusion of non-coding TUs or other regions that did not meet the AUGUSTUS gene prediction criteria. The average gene length in the unique AUGUSTUS set (1,089 bp) is approximately twice the average TU length (564 bp), suggesting that the RNAseq-informed gene predictions likely provided support for the extension of the TUs. However, the average gene length in the unique set is shorter on average than the entire JGI or the AUGUSTUS gene model sets (27% and 36% shorter, respectively), which may indicate that previous prediction methods may have been biased against the prediction of shorter models (Fig. 1-5). Analysis of base mean gene expression (averaged over all experimental conditions) does not suggest a bias due to expression level, as genes from the unique set represent a wide range of expression (Fig. 1-6).

Functional annotation of the unique gene set revealed that 74% of the unique AUGUSTUS sequences hit other sequences in the NBCI database (E-value cut-off = $1e^{-3}$). In addition, 1141 genes (61%) could be assigned an Interpro ID and/or a GO annotation (Fig. 1-7), supporting a biological role for these genes. The majority of genes with GO annotations belong to functional categories related to cellular metabolism, including biosynthetic processes and other pathways related to the metabolic processing of organic carbon and nitrogen-containing compounds (Fig. 1-S1), suggesting that some of the newly predicted genes will be useful in completing metabolic maps and allow for more comprehensive metabolic modeling efforts.

The DarkHorse program was used to identify genes without significant similarity to sequences in the NCBI Genbank nr database (Podell 2007), and are therefore likely to be specific to *T. pseudonana*. The unique AUGUTSUS set contained 661 *T. pseudonana*-specific genes (35.2% of unique set). The prevalence of *T. pseudonana*-specific genes in the entire AUGUSTUS prediction is 14%, suggesting that the unique set is enriched for genes that are not shared in other diatom genomes, and likely escaped detection by past homology-based annotation efforts. In addition, DarkHorse analysis identifies genes of potentially derived from horizontal gene transfer. We found that there is no enrichment of genes with putative bacterial origin in the unique gene set (0.8% genes in the unique set versus 1.7% overall). We identified a total of 16 genes of potential viral or bacterial origin in the unique gene set (Supplemental file 1-4), including genes with similarity to bacterial cell wall-anchored and PKD-domain containing proteins (g1315, g1971, g2521, g2596, g9508), which play a role in adhesion and/or cell surface interactions (Jing et al. 2002). Some of these proteins also have integrin domains (g8366, 8367), which are cell surface receptors that recruit intracellular protein complexes that interact with the actin cytoskeleton and activate internal signal transduction pathways (van der Flier and Sonnenberg 2001). In addition, a putative tryptophan-rich sensory protein (g962), which plays a role in transmembrane signaling, was identified. Taken together, we identify a complement of newly annotated genes of putative bacterial origin that suggest significant prokaryotic influence on cell-surface processes in *T. pseudonana*.

1.4.4 Cell cycle-related expression of previously unannotated genes

The global physiological and transcriptional responses to early-stage silicon starvation and synchronous cell division for JGI Thaps3 genes have previously been

described in detail (Shrestha et al. 2012; Smith et al. 2016b). Analysis of sample-to-sample distance in silicon starvation experiments show major transcriptional changes in response to the transition from silicon-replete (exp) to silicon-free (0h) medium, followed by another shift from 0-4 hours (Fig. 1-8A). For synchrony experiments, sample distances reflected progression through the cell cycle, with a return to initial conditions by the end of the timecourse (Fig. 1-8B). Gene-specific expression data is included in Supplemental Files 1-2 and 1-3.

A high proportion of genes in the unique gene set were differentially expressed (DESeq2 Likelihood ratio test, adjusted p-value ≤ 0.05) through silicon starvation and cell cycle progression (878 genes and 1,129 genes in starvation and synchrony, respectively). Of these, most (707 genes) are significantly differentially expressed during both silicon starvation and cell cycle progression, and likely represent previously unannotated genes related to silicon availability and cell cycle processes. Several genes from the unique set are among the 35 most significantly differentially expressed genes in both datasets (Fig. 1-9). These include several proteins that contain the SMC_prok_B multi-domain (g10431, g3719, g6888, g558). Structural maintenance of chromosome (SMC) proteins are DNA-binding proteins found in both bacteria and eukaryotes, and are involved in processes related to cell division such as chromatid cohesion and chromosome segregation (Hirano 2006). In addition, a frustulin-like protein (g432) similar to other diatom cell surface proteins (Kröger et al. 1996) containing a Ricin-like beta-trefoil multi-domain (carbohydrate binding) was significantly differentially expressed in both datasets. The expression patterns of the SMC proteins and g432 during silicon starvation are strikingly similar, with a sharp decrease in transcript abundance

following the transition from silicon-replete to silicon-free medium (Fig. 1-10), suggesting reduced importance during stress induced cell cycle arrest. For the synchrony experiments, transcripts for SMC proteins all peak at 4h following silicon addition, coinciding with the timing of early mitotic processes such as chromosome cohesion and condensation. Expression of the SMC proteins dramatically decreases as the cell progresses towards the metaphase-anaphase transition in preparation for chromosome segregation (Hirano 2000). Expression for g432 peaks an hour later, during late mitosis (Fig. 1-10).

The unique gene set also included seven putative cyclin sequences (g7761, g9962, g268, g13056, g6714, g6526, g2512). Cyclins are proteins that form complexes with cyclin-dependent kinases to regulate aspects of cell cycle progression (Inzé and De Veylder 2006). Some cyclins are known to participate in alternate roles, such as transcriptional activation or splicing (Coqueret 2002; Kitsios et al. 2008). Cyclins have been expanded in diatom genomes (Bowler et al. 2008), and up to 48 other putative cyclin genes have been previously identified in *T. pseudonana* (Montsant et al. 2007), many of which are diatom-specific cyclins (dsCYC). It has been proposed that some dsCYCs may act as integrators of environmental signals, and some have been demonstrated to be transcriptionally responsive to different environmental cues, such as light, phosphate, iron, or silicon levels (Sapriel et al. 2009; Huysman et al. 2010; Valenzuela et al. 2012; Smith et al. 2016a). Notably, g7761 is highly expressed in our datasets and its expression is responsive to silicon availability (Fig. 1-11). The transcript was significantly upregulated during the silicon starvation timecourse, and is downregulated within 1 hr of silicon readdition (Fig. 1-11). Transcription also peaked at

5h during the cell cycle timecourse, coinciding with valve formation and late mitosis (Fig. 1-11). The increase in g7761 may also be tied to silicon availability in this case, as the biological requirement for silicon increases during valve formation.

1.4.4 Extent of alternative splicing in the T. pseudonana transcriptome

We used the rMATS program to detect significant instances of alternative splicing from the alignment of RNAseq reads. Significant alternative splicing events (FDR <= 0.05) were determined from pairwise sample comparisons. Our results show that intron retention (IR) (i.e. presence of unspliced introns, Fig. 1-12), is the major form of AS that occurs in the *T. pseudonana* transcriptome (99.4% of all AS events). IR is also the most prevalent form of splicing that occurs in plants (Ner-Gaon et al. 2004; Marquez et al. 2012), green algae (Labadorf et al. 2010) and yeast (Inzé and De Veylder 2006; Brogna and Wen 2009). The dominance of IR is not universal in all photosynthetic eukaryotes; the chlorarachniophyte *Bigelowiella natans* transcriptome is characterized by complex splicing, including high levels of exon skipping (Curtis et al. 2012). A total of 821 significant intron retention events were found in 397 different genes (Supplemental File 1-5), which represents 3.5% of genes in the genome. This percentage is similar to the extent of splicing (3%) detected in *C. reinhardtii* (Labadorf et al. 2010). This should be considered a conservative estimate of the AS that occurs in *T. pseudonana*, as it only includes significant changes in splicing in our silicon and synchrony timecourses. Also, rMATS would not consider an IR event to be significant if the calculated ratio between retained and spliced introns remained the same between two timepoints. Therefore, it is likely that other significant IR events would be detected given different experimental conditions.

The majority of detected splicing events occur once and are specific to a particular timepoint (Fig. 1-13), suggesting intron retention is a dynamic phenomenon and that the extent of intron retention for a particular intron can change on relatively rapid timescales (hours or less). The distribution of intron retention suggests that the extent of IR is condition-dependent, with a notable trend towards more significant IR events as the silicon starvation timecourse progresses (Fig. 1-14A). Stress is known to alter splicing patterns in plant genes (Palusa et al. 2007; Tanabe et al. 2007), which may suggest that some IR may not be a controlled by specific regulation by rather by a general reduction of spliceosome efficiency under unfavorable conditions. However, we also observe sample-specific peaks in IR during the synchrony timecourse that do not correlate with environmental stress, suggesting a more regulated mechanism (Fig. 1-14B). In addition, we observed condition-dependent changes in the extent of IR coordinated with changes in gene expression. For example, the chitinase (g8889) has a distinctive peak at 3 hours during the synchrony timecourse (Fig. 1-12). The pattern and extent of IR in this gene corresponds to expression, with an observed release of IR as the expression of the gene increases, suggesting a regulated process in some cases.

We also examined the relationship between gene expression and significant intron retention events to examine whether IR detection could be an artifact of high expression or large changes in expression (rapid increases in transcription allows less time for complete splicing). However, we found no correlation between transcript abundance and intron retention (Fig. 1-S2) or between fold change expression and intron retention (Fig. 1-15). We further validated intron retention in select genes by RT-PCR (Fig. 1-16). Criterion for gene selection included genes with 2 or more exons, at least one of which

was detected to have significant IR, and one of which did not (as a control for possible genomic DNA contamination in cDNA library). In both examples, we observed the presence and abundance of the IR transcript corresponded to the rMATs prediction (Fig. 1-16).

Intron retention can regulate gene expression by determining the amount of functional transcript available for translation. Most intron retention events likely introduce premature stop codons, making these unproductive transcripts targets for mRNA quality control pathways such as nonsense mediated decay (NMD). NMD is a translation-coupled RNA surveillance mechanism that degrades transcripts containing premature stop codons, thereby eliminating the production of potentially problematic truncated proteins (Brogna and Wen 2009). The core NMD machinery consists of three evolutionarily conserved proteins, *UPF1*, *UPF2*, and *UPF3*, which are thought to occur universally in all eukaryotes (Brogna and Wen 2009). A survey of the *T. pseudonana* genome revealed homologs for *UPF1* (g191) and *UPF2* (g4659), the former of which was found in the new AUGUSTUS gene model set. However, a homolog to *UPF3* was not found in *T. pseudonana* or in any other Bacillariophyceae sequence in the NCBI database, casting doubt that canonical NMD may occur in diatoms. However, there is some experimental support for *UPF3*-independent mechanisms for NMD in other organisms (Chan et al. 2007).

Alternatively, the presence of IR in diatoms may be a means to fine-tune gene expression, independent of a translation-based mechanism such as NMD. For example, the deletion of *UPF1* was shown to not affect the prevalence of IR in the majority of transcripts in the yeast *Cryptococcus neoformans*, suggesting that the majority (99%) of

IR events were not regulated by NMD (Gonzalez-Hilarion et al. 2016). Rather, the authors propose a mechanism by which environmental factors regulate splicing efficiency, with unproductive transcripts retained and degraded inside of the nucleus.

1.5 CONCLUSIONS

Advances in sequencing technology have enabled the continual improvement of early genome sequencing projects. We used alignments of high-throughput RNA sequencing data to inform a gene reannotation effort for the model diatom *T. pseudonana.* The newly predicted gene set contains 13,399 genes, including 38% with modified gene structure and 14% previously unpredicted genes with no structural overlap with current models. Within the previously unpredicted genes, we find gene annotations from a diversity of metabolic and cell processes, including several cell wall-associated genes of putative prokaryotic origin. In addition, a high proportion of newly predicted genes are transcriptionally responsive to silicon availability and cell cycle progression, including SMC and cyclin proteins, that may provide further insight into cell cycle regulation and related processes in diatoms. Additionally, we present the first survey of alternatively splicing in *T. pseudonana*, and identify intron retention as the dominant alternative splicing phenomenon. This result is contrast to the extensive exon skipping observed in another algal secondary endosymbiont, *B. natans*, perhaps reflecting regulatory differences related to the secondary host organism. Significant changes in intron retention during our experimental conditions are relatively rare (~3% of transcripts) and occur over relatively rapid timescales (within hours), suggesting a dynamic regulation of splicing. Interestingly, a major component of NMD (a widespread

and conserved RNA surveillance mechanism in eukaryotes) is missing in diatom genomes, suggesting a modified or alternative mechanism for the regulation of functional transcript availability.

1.6 ACKNOWLEDGEMENTS

This research was supported by a Department of Energy Office of Science Graduate Fellowship Program (DOE SCGF), made possible in part by the American Recovery and Reinvestment Act of 2009, administered by ORISE-ORAU under contract no. DE-AC05-06OR23100. This research was also made possible by the Robert Buzzelli Fellowship, Jeff Graham Fellowship, Krinsk Research Advancement Fellowship, and the Ralph Lewin Fellowship at the Scripps Institution of Oceanography.

Chapter 1, in part, is currently being prepared for submission for publication of the material in 2017. Abbriano, R.; Spreafico, R.; Lopez, D.; Lu, J.; Traller, J.; Shrestha, R.; Tesson, B.; Podell, S.; Hovde, B.; Allen, E.; Pellegrini, M.; and Hildebrand, M. The dissertation author was the primary investigator and author of this material. Chapter 1, in part, is included in New Phytologist in 2016. Smith, S.; Glé, C.; Abbriano, R.; Traller, J.; Davis, A.; Vernet, M.; Trentacoste, E.; Allen, A.; and Hildebrand, M. The dissertation author is the primary investigator of the data reported in Chapter 1; Sarah Smith is the primary investigator of the New Phytologist manuscript.

TABLES

Table 1-1. Gene model statistics comparing Thaps3 filtered models (JGI), FGENESH, MAKER and AUGUSTUS gene model predictions. Values that deviate from JGI by more than 20% are highlighted in red.

	Number of models	Mean gene length (bp)	Mean intron length (bp)	Mean exon length (bp)	Number of exons per gene	Number of introns per gene
JGI	11,390	1,745	123	613	2.5	1.5
FGENESH	10,904	2,570	126	681	3.3	2.3
MAKER	6,433	2,638	210	672	4.5	3.5
AUGUSTUS	13,399	1,503	124	646	2.7	1.7

FIGURES

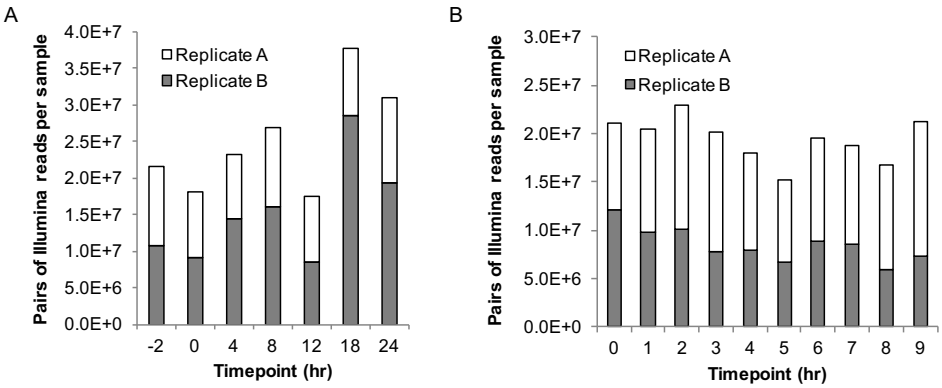

Figure 1-1. Total number of paired reads per sample from A) silicon starvation (cell cycle arrest) and B) synchronous cell cycle progression experiments. Stacked bar graph represents combined RNAseq read pairs from two individual biological replicates.

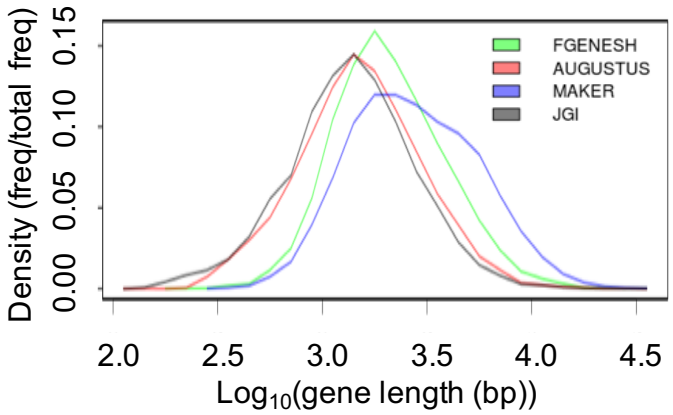

Figure 1-2. Gene length distributions for JGI, FGENESH, MAKER and AUGUSTUS gene model predictions.

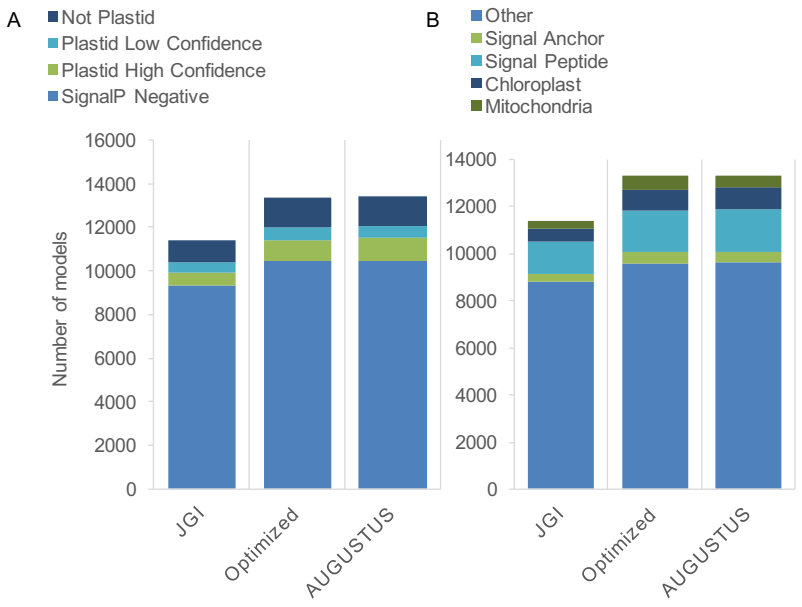

Figure 1-3. Intracellular targeting analysis for JGI (Thaps3 filtered models), optimized model set from Gruber et al. 2015, and AUGUSTUS gene models. A) Analysis of secondary plastid targeting signals by ASAFind using SignalP 3.0 output (Gruber et al. 2015). B) Analysis of intracellular targeting signals by HECTAR.

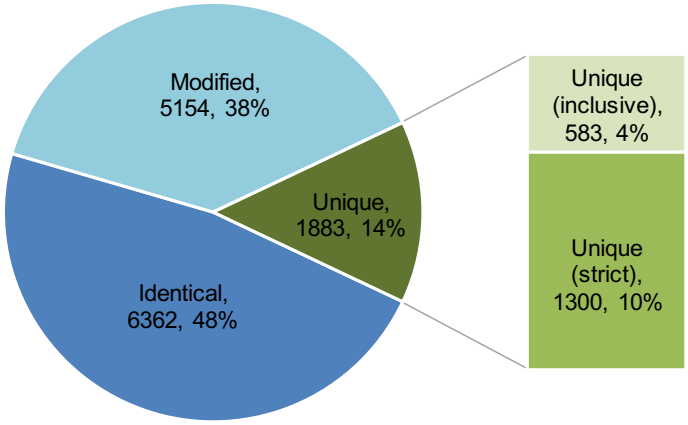

Figure 1-4. Comparison of JGI and AUGUTSUS gene model structure by Cuffcompare. The modified set includes gene model extensions and modifications of intron/exon boundaries. The unique set includes models with no overlap with existing models (strict) as well as models that fall within JGI intronic regions or overlap on the opposite strand (inclusive).

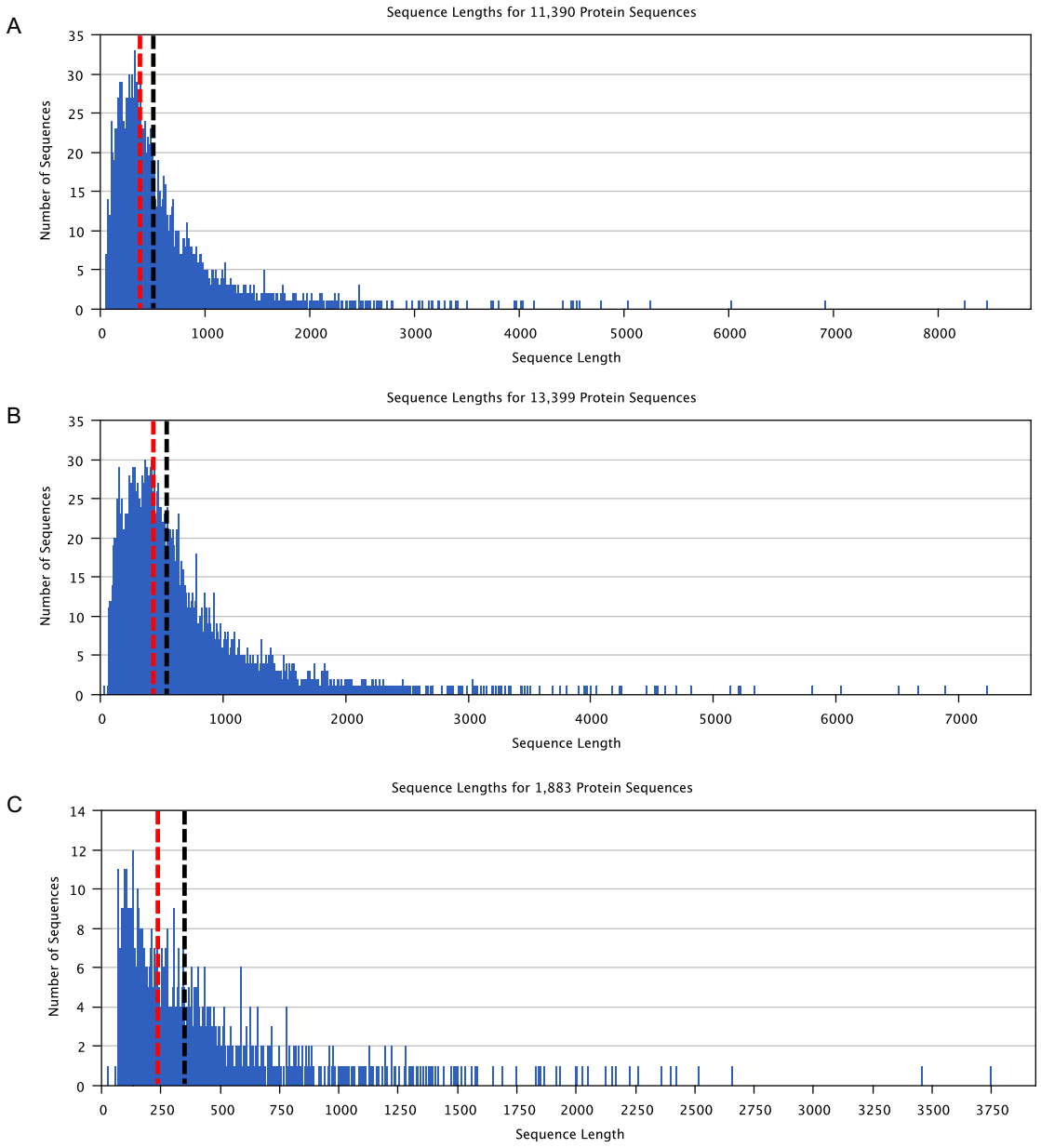

Figure 1-5. Protein length distributions for A) JGI (Thaps3 filtered models) (11,390 proteins), B) the entire AUGUSTUS prediction (13,399 proteins), and C) previously unannotated gene model predictions (unique models; 1,883 proteins). Red dashed lines represent the median protein sequence length; black dashed lines represent the mean protein sequence length.

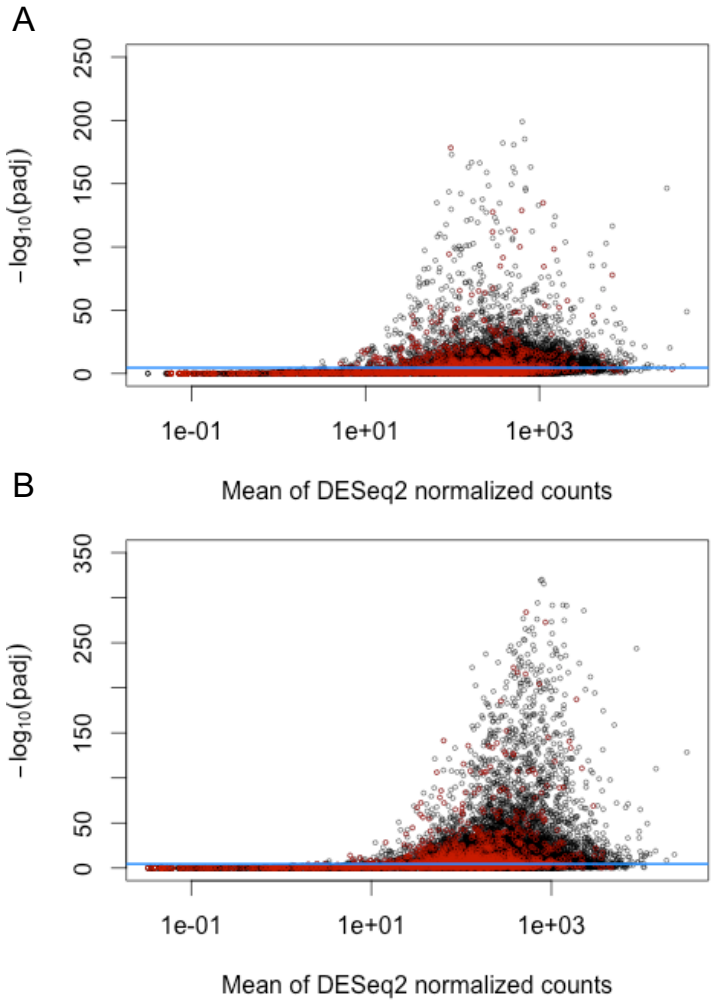

Figure 1-6. Distribution of baseMean counts (average gene expression) of all genes in A) starvation experiment, and B) synchrony experiment. Red color indicates genes belonging to the unique AUGUSTUS set. Blue line indicates significance cutoff based on adjusted p-value of 0.01.

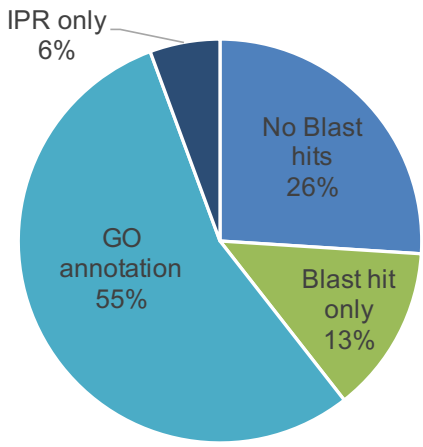

Figure 1-7. Functional annotations of the unique set of AUGUSTUS genes not found in the filtered JGI (Thaps3 filtered models) dataset.

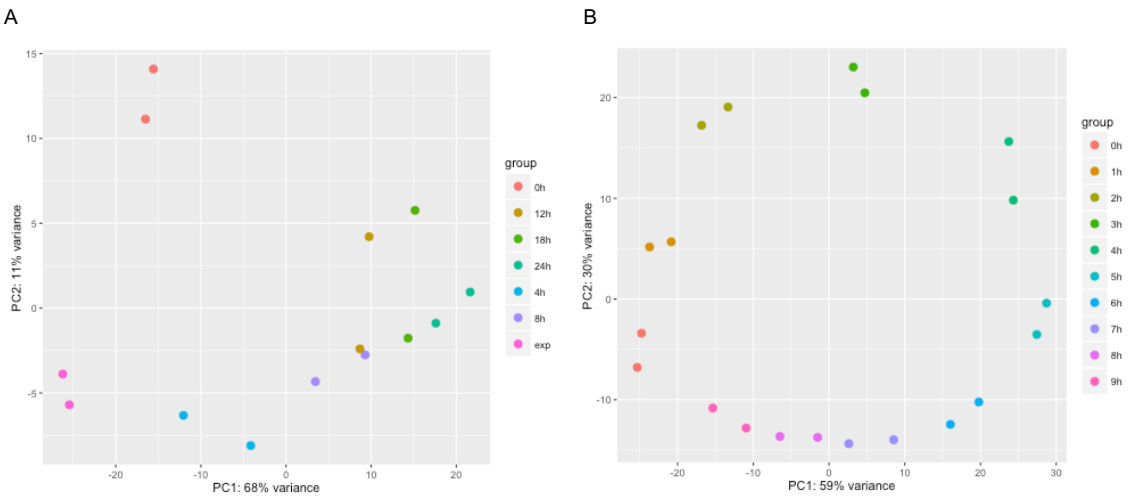

Figure 1-8. PCA plots of sample distances among RNAseq timepoints from A) silicon starvation experiments and B) synchrony experiments. Samples of the same color represent biological replicates (n=2).

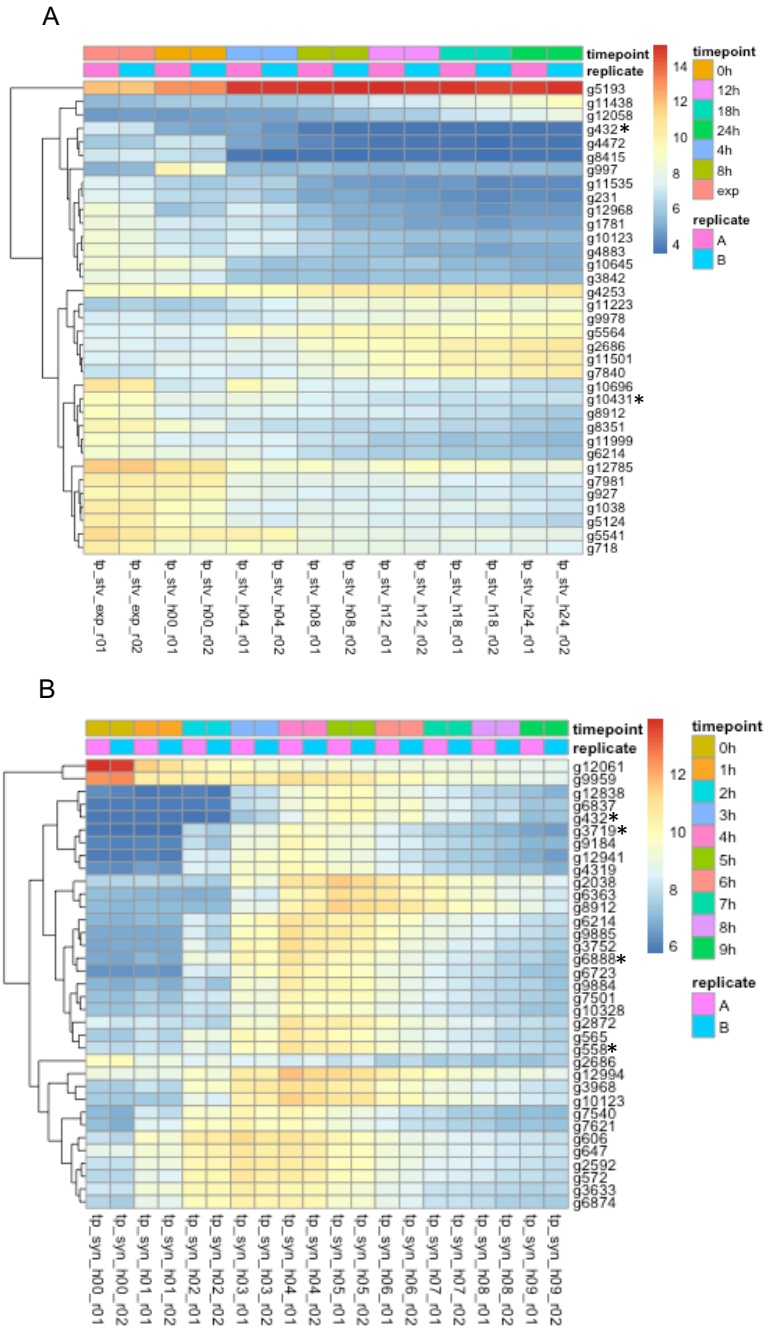

Figure 1-9. Heatmaps filtered for top 35 most significantly changing genes as determined by DESeq2 analysis in A) silicon starvation and B) synchrony datasets. (*) indicates genes belonging to the unique AUGUSTUS gene set with functional annotations as structural maintenance of chromosomes (SMC) proteins (g10431, g3719, g6888, g558) or frustulin-like proteins (g432).

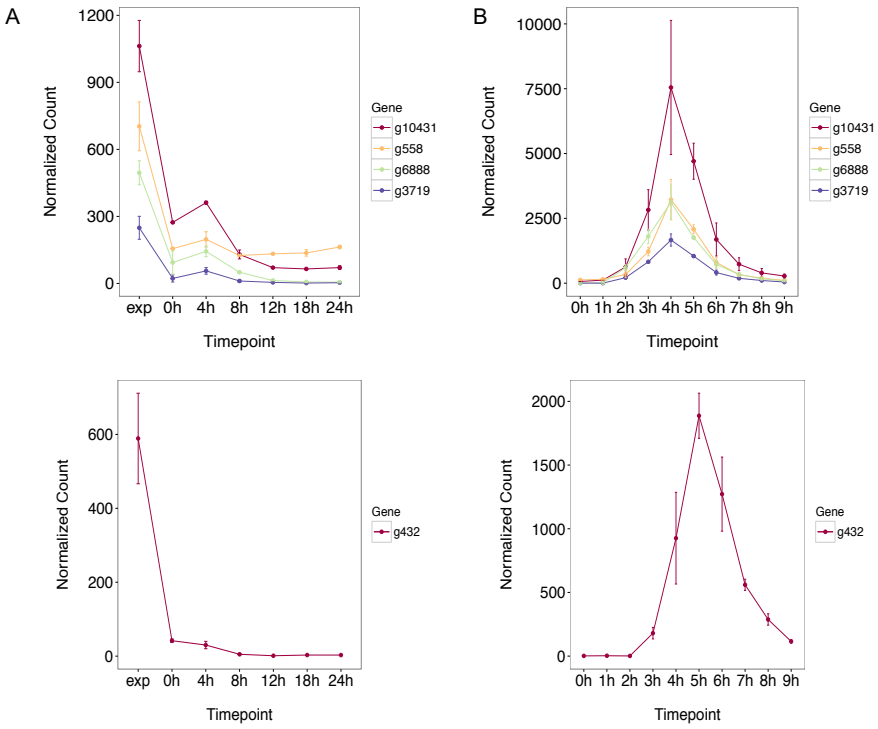

Figure 1-10. Gene expression patterns for SMC proteins (g10431, g3719, g6888, g558) and frustulin-like protein (g432) during A) silicon starvation and B) synchrony experiments.

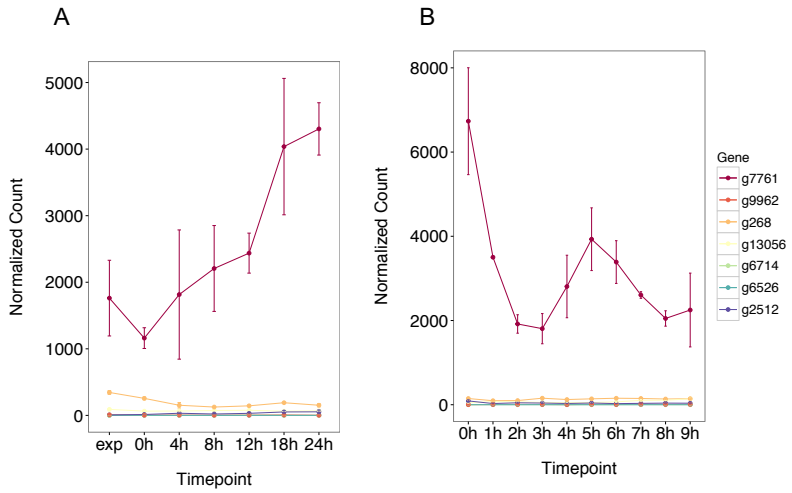

Figure 1-11. Gene expression patterns for putative cyclins found in the unique AUGUTUS gene set (g7761, g9962, g268, g13056, g6714, g6526, g2512) during A) silicon starvation and B) synchrony experiments.

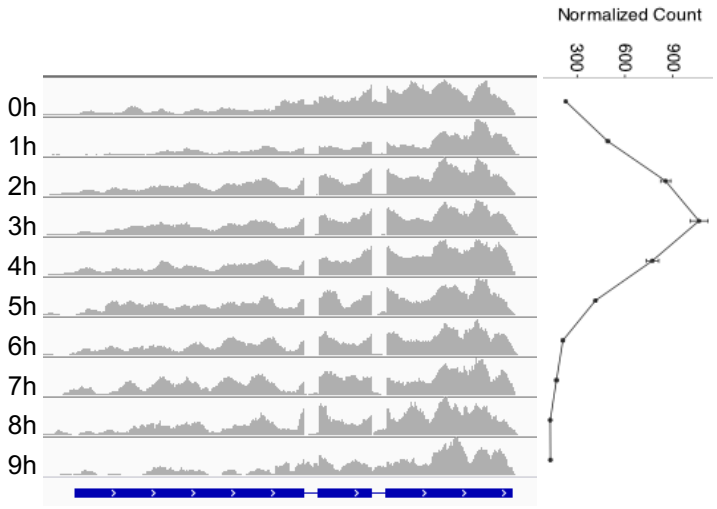

Figure 1-12. Pattern of transcript expression and intron retention during the synchrony timecourse for a putative chitinase (g8889) depicting a release of intron retention as transcript expression increases from 0-3 hours. RNAseq read coverage (grey) is autoscaled to track height.

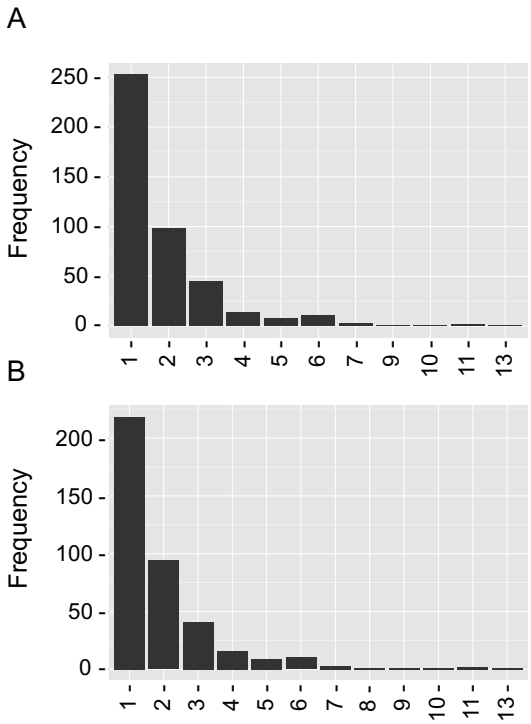

Figure 1-13. Occurrence of A) significant intron retention events and B) genes with significant intron retention in all experimental timepoints. The majority of significant events are detected once in the experimental conditions tested.

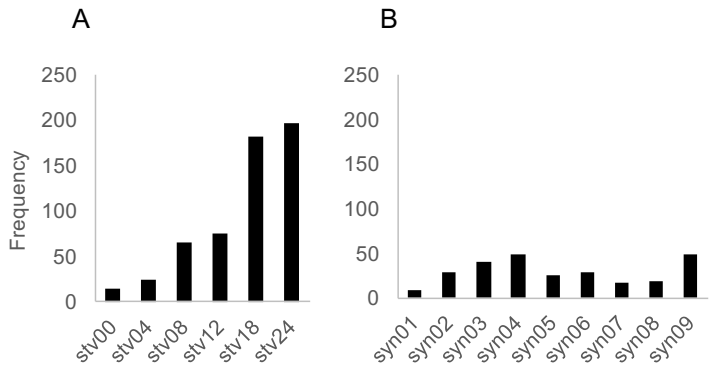

Figure 1-14. Frequency of significant intron retention events throughout the A) silicon starvation and B) synchrony timecourses.

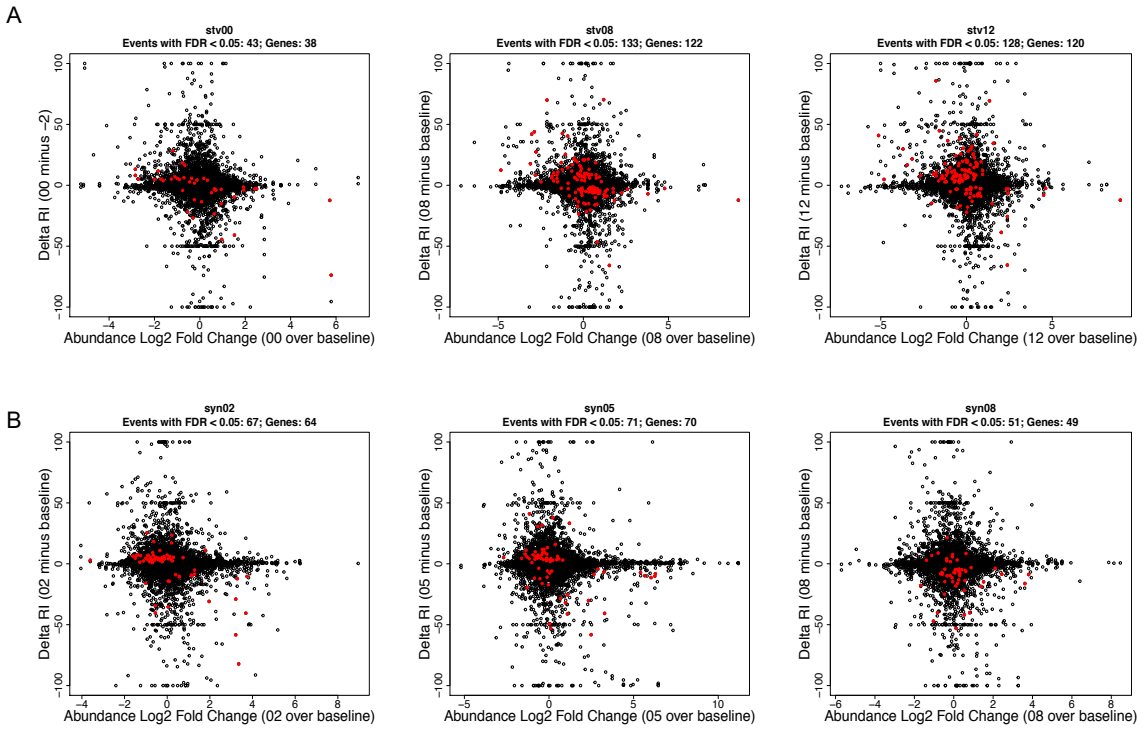

Figure 1-15. Relationship between gene expression (log2 fold change) and intron retention (delta RI) in representative timepoints from A) silicon starvation and B) synchrony experiments. Events with statistically significant changes in intron retentions are represented as red points.

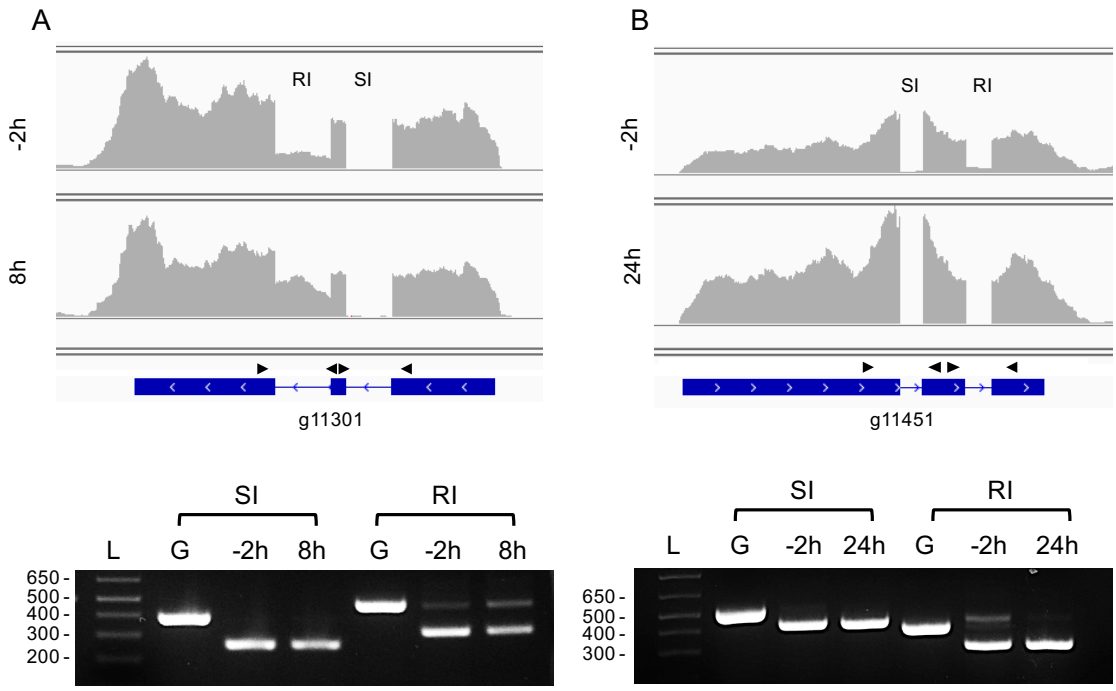

Figure 1-16. Validation of intron retention in two representative genes: A) g11301 and B) g11451. The top panel is a genome browser view with RNAseq read coverage from two timepoints where intron retention is predicted to be statistically significant. The gene model is depicted in blue, with black arrows indicating approximate location of the primers used in the RT-PCR. The bottom panel includes gel images demonstrating intron retention (L=ladder, G= genomic DNA, SI = spliced intron/control, RI= retained intron).

1.7 REFERENCES

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306:79–86.

Bendtsen J, Nielsen H, Heijne von G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795.

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret FED, Berges JA, Brownlee C, Cadoret J-P, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet JER, Haruta M, Huysman MJJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, ger NKO, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin-Jézéquel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Oudot-Le Secq M, Napoli C, Obornik M, Parker MS, Petit J-L, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, Dassow von P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456:239–244.

Brogna S, Wen J (2009) Nonsense-mediated mRNA decay (NMD) mechanisms. Nature Structural & Molecular Biology 16:107–113.

Brzezinski MA, Olson RJ, Chisholm SW (1990) Silicon availability and cell-cycle progression in marine diatoms. Mar Ecol Prog Ser 67:83–96.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research 18:188–196.

Chan WK, Huang L, Gudikote JP, Chang YF, Imam JS, MacLean JA, Wilkinson MF (2007) An alternative branch of the nonsense-mediated decay pathway. The EMBO Journal 26:1820–1830.

Coqueret O (2002) Linking cyclins to transcriptional control. Gene 299:35–55.

Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, Hopkins JF, Kuo A, Rensing SA, Schmutz J, Symeonidi A, Elias M, Eveleigh RJM, Herman EK, Klute MJ, Nakayama T, Obornik M, Reyes-Prieto A, Armbrust EV, Aves SJ, Beiko RG, Coutinho P, Dacks JB, Durnford DG, Fast NM, Green BR, Grisdale CJ, Hempel F, Henrissat B, Höppner MP, Ishida K-I, Kim E, Kořený L, Kroth PG, Liu Y, Malik S-B, Maier U-G, McRose D, Mock T, Neilson JAD, Onodera NT, Poole AM, Pritham EJ, Richards TA, Rocap G, Roy SW, Sarai C, Schaack S, Shirato S, Slamovits CH, Spencer DF, Suzuki S, Worden AZ, Zauner S, Barry K, Bell C, Bharti AK, Crow JA, Grimwood J, Kramer R, Lindquist E, Lucas S, Salamov A, McFadden GI, Lane CE, Keeling PJ, Gray MW, Grigoriev IV, Archibald JM (2012) Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 492:59-65.

Darley WM, Volcani BE (1969) Role of silicon in diatom metabolism: a silicon requirement for deoxyribonucleic acid synthesis in diatom *Cylindrotheca fusiformis Reimann and Lewin*. Exp Cell Res 58:334-342.

Delalat B, Sheppard VC, Ghaemi SR, Rao S, Prestidge CA, McPhee G, Rogers M-L, Donoghue JF, Pillay V, Johns TG, Kroger N, Voelcker NH (2015) Targeted drug delivery using genetically engineered diatom biosilica. Nature Communications 6:8791.

Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong WK, Mockler TC (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome Research 20:45–58.

Gonzalez-Hilarion S, Paulet D, Lee K-T, Hon C-C, Lechat P, Mogensen E, Moyrand F, Proux C, Barboux R, Bussotti G, Hwang J, Coppee J-Y, Bahn Y-S, Janbon G (2016) Intron retention-dependent gene regulation in *Cryptococcus neoformans*. Sci Rep 6:32252.

Gordon R, Losic D, Tiffany MA, Nagy SS, Sterrenburg FAS (2009) The glass menagerie: diatoms for novel applications in nanotechnology. Trends in Biotechnology 27:116–127.

Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T (2015) Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. The Plant Journal 81:519–528.

Gruber A, Vugrinec S, Hempel F, Gould SB, Maier UG, Kroth PG (2007) Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. Plant Molecular Biology 64:519–530.

Gschloessl B, Guermeur Y, Cock JM (2008) HECTAR: A method to predict subcellular targeting in heterokonts. BMC Bioinformatics 9:393.

Guihéneuf F, Leu S, Zarka A, Goldberg IK, Khalilov I (2011) Cloning and molecular characterization of a novel acyl-CoA:diacylglycerol acyltransferase 1-like gene (PtDGAT1) from the diatom *Phaeodactylum tricornutum*. FEBS Journal 278:3651-3666.

Hempel F, Lau J, Klingl A, Maier UG (2011) Algae as protein factories: expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. Plos One 6:e28424.

Hildebrand M, Dahlin K (2000) Nitrate transporter genes from the diatom *Cylindrotheca fusiformis* (Bacillariophyceae): mRNA levels controlled by nitrogen source and by the cell cycle. Journal of Phycology 36:702–713.

Hildebrand M, Davis AK, Smith SR, Traller JC (2012) The place of diatoms in the biofuels industry. Biofuels 3:221–240.

Hildebrand M, Frigeri LG, Davis AK (2007) Synchronized growth of *Thalassiosira pseudonana* (Bacillariophyceae) provides novel insights into cell wall synthesis processes in relation to the cell cycle. Journal of Phycology 43:730–740.

Hirano T (2006) At the heart of the chromosome: SMC proteins in action. Nature Reviews Molecular Cell Biology 7:311–322.

Hirano T (2000) Chromosome cohesion, condensation, and separation. Annu Rev Biochem 69:115–144.

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

Inzé D, De Veylder L (2006) Cell cycle regulation in plant development. Annu Rev Genet 40:77-105.

Jing H, Takagi J, Liu JH, Lindgren S, Zhang R-G, Joachimiak A, Wang JH, Springer TA (2002) Archaeal surface layer proteins contain β Propeller, PKD, and β Helix domains and are related to metazoan cell surface proteins. Structure 10:1453–1464.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Kinoshita S, Kaneko G, Lee JH, Kikuchi K, Yamada H, Hara T, Itoh Y, Watabe S (2001) A novel heat stress-responsive gene in the marine diatom *Chaetoceros compressum* encoding two types of transcripts, a trypsin-like protease and its related protein, by alternative RNA splicing. Eur J Biochem 268:4599–4609.

Kitsios G, Alexiou KG, Bush M, Shaw P, Doonan JH (2008) A cyclin-dependent protein kinase, CDKC2, colocalizes with and modulates the distribution of spliceosomal components in *Arabidopsis*. The Plant Journal 54:220–235.

Kröger N, Bergsdorf C, Sumper M (1996) Frustulins: Domain conservation in a protein family associated with diatom cell walls. Eur J Biochem 239:259–264.

Labadorf A, Link A, Rogers MF, Thomas J, Reddy ASN, Ben-Hur A (2010) Genome-wide analysis of alternative splicing in *Chlamydomonas reinhardtii*. BMC Genomics 11:114.

Li Z, Zhang Z, Yan P, Huang S, Fei Z, Lin K (2011) RNA-Seq improves annotation of protein-coding genes in the cucumber genome. BMC Genomics 12:540.

Martin-Jézéquel V, Hildebrand M, Brzezinski, MA (2000) Silicon metabolism in diatoms: implications for growth. J. Phycol 36:821-840.

Marquez Y, Brown JWS, Simpson C, Barta A, Kalyna M (2012) Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. Genome Research 22:1184–1195.

Mock T, Samanta MP, Iverson V, Berthiaume C, Robison M, Holtermann K, Durkin C, BonDurant SS, Richmond K, Rodesch M, Kallas T, Huttlin EL, Cerrina F, Sussmann MR, Armbrust EV (2008) Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. P Natl Acad Sci Usa 105:1579–1584.

Montsant A, Allen AE, Coesel S, De Martino A, Falciatore A, Mangogna M, Siaut M, Heijde M, Jabbari K, Maheswari U, Rayko E, Vardi A, Apt KE, Berges JA, Chiovitti A, Davis AK, Thamatrakoln K, Hadi MZ, Lane TW, Lippmeier JC, Martinez D, Parker MS, Pazour GJ, Saito MA, Rokhsar DS, Armbrust EV, Bowler C (2007) Identification and comparative genomic analysis of signaling and regulatory components in the diatom *Thalassiosira pseudonana*. Journal of Phycology 43:585–604.

Nazmi A, Hauck R, Davis A, Hildebrand M (2017) Diatoms and diatomaceous earth as novel poultry vaccine adjuvants. Poultry Science 96:288-294.

Nelson DM, Tréguer P, Brzezinski MA, Leynaert A, Quéguiner B (1995) Production and dissolution of biogenic silica in the ocean - revised global estimates, comparison with regional data and relationship to biogenic sedimentation. Global Biogeochem Cy 9:359–372.

Ner-Gaon H, Halachmi R, Savaldi-Goldstein S, Rubin E, Ophir R, Fluhr R (2004) Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*. The Plant Journal 39:877–885.

Nielsen H, Krogh A (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. ISMB-98 Proceedings.

Palusa SG, Ali GS, Reddy ASN (2007) Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: regulation by hormones and stresses. The Plant Journal 49:1091–1107.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40:1413–1415.

Podell S, Gaasterland, T (2007) DarkHorse: a method for genome-wide prediction of horizontal gene transfer. Genome Biology 8:R16.

Reddy ASN (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol 58:267–294.

Salamov AA, Solovyev VV (2000) *Ab initio* gene finding in *Drosophila* genomic DNA. Genome Research 10:516–522.

Sapriel G, Quinet M, Heijde M, Jourdren L, Tanty V, Luo G, Le Crom S, Lopez PJ (2009) Genome-wide transcriptome analyses of silicon metabolism in *Phaeodactylum tricornutum* reveal the multilevel regulation of silicic acid transporters. Plos One 4:e7458.

Shrestha RP, Tesson B, Norden-Krichmar T, Federowicz S, Hildebrand M, Allen AE (2012) Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. BMC Genomics 13:499.

Smith SR, Gillard JTF, Kustka AB, McCrow JP, Badger JH, Zheng H, New AM, Dupont CL, Obata T, Fernie AR, Allen AE (2016a) Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. PLoS Genet 12:e1006490.

Smith SR, Glé C, Abbriano R, Traller JC, Davis A, Trentacoste E, Vernet M, Allen AE, Hildebrand M (2016b) Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. New Phytol 210:890-904.

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research 33:W465–W467.

Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keeffe S, Haas S, Vingron M, Lehrach H, Yaspo ML (2008) A global view of gene activity and alternative

splicing by deep sequencing of the human transcriptome. Science 321:956–960.

Tanabe N, Yoshimura K, Kimura A, Yabuta Y, Shigeoka S (2007) Differential expression of alternatively spliced mRNAs of *Arabidopsis* SR protein homologs, atSR30 and atSR45a, in response to environmental stress. Plant and Cell Physiology 48:1036–1049.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515.

Valenzuela J, Mazurie A, Carlson RP, Gerlach R, Cooksey KE, Peyton BM, Fields MW (2012) Potential role of multiple carbon fixation pathways during lipid accumulation in *Phaeodactylum tricornutum*. Biotechnology for Biofuels 5:40.

van der Flier A, Sonnenberg A (2001) Function and interactions of integrins. Cell and Tissue Research 305:285–298.

Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. Nat Rev Genet 13:329–342.

Zhao C, Waalwijk C, de Wit PJGM, Tang D, van der Lee T (2013) RNA-seq analysis reveals new gene models and alternative splicing in the fungal pathogen *Fusarium graminearum*. BMC Genomics 14:21.

CHAPTER 2:

Clarification of photorespiratory processes and the role of malic enzyme in diatoms

Protist, Vol. 168, 134–153, xx 2017
http://www.elsevier.de/protis
Published online date 21 October 2016

Protist

ORIGINAL PAPER

Clarification of Photorespiratory Processes and the Role of Malic Enzyme in Diatoms

Aubrey Davis[a,1], **Raffaela Abbriano**[a,1], **Sarah R. Smith**[b,c], **and Mark Hildebrand**[a,2]

[a]Marine Biology Research Division, Scripps Institution of Oceanography, UC San Diego, La Jolla, CA, U.S.A.
[b]Integrative Oceanography Division, Scripps Institution of Oceanography, UC San Diego, La Jolla, CA, U.S.A.
[c]J. Craig Venter Institute, La Jolla, CA, U.S.A.

Submitted May 3, 2016; Accepted October 8, 2016
Monitoring Editor: Saul Purton

Evidence suggests that diatom photorespiratory metabolism is distinct from other photosynthetic eukaryotes in that there may be at least two routes for the metabolism of the photorespiratory metabolite glycolate. One occurs primarily in the mitochondria and is similar to the C2 photorespiratory pathway, and the other processes glycolate through the peroxisomal glyoxylate cycle. Genomic analysis has identified the presence of key genes required for glycolate oxidation, the glyoxylate cycle, and malate metabolism, however, predictions of intracellular localization can be ambiguous and require verification. This knowledge gap leads to uncertainties surrounding how these individual pathways operate, either together or independently, to process photorespiratory intermediates under different environmental conditions. Here, we combine in silico sequence analysis, in vivo protein localization techniques and gene expression patterns to investigate key enzymes potentially involved in photorespiratory metabolism in the model diatom *Thalassiosira pseudonana*. We demonstrate the peroxisomal localization of isocitrate lyase and the mitochondrial localization of malic enzyme and a glycolate oxidase. Based on these analyses, we propose an updated model for photorespiratory metabolism in *T. pseudonana*, as well as a mechanism by which C2 photorespiratory metabolism and its associated pathways may operate during silicon starvation and growth arrest.

Key words: Diatom; photorespiration; glyoxylate cycle; malic enzyme; *Thalassiosira pseudonana*.

Introduction

Diatoms are a ubiquitous and highly productive group of eukaryotic phytoplankton that contribute significantly to the global carbon cycle. Diatoms are estimated to fix 10 billion metric tons of inorganic carbon annually, which is at least a quarter of the carbon fixed each year in the ocean (Granum et al. 2005). Photosynthetic carbon fixation by diatoms sustains diverse marine ecosystems and contributes significantly to total export production (Sarthou et al. 2005). The high productivity that drives these large-scale processes is ulti-

[1]These authors contributed equally to this work.
[2]Corresponding author; fax +1 858 534 7313
e-mail mhildebrand@ucsd.edu (M. Hildebrand).

http://dx.doi.org/10.1016/j.protis.2016.10.005
1434-4610/

mately connected to the efficiency of carbon fixation and the metabolism of fixed carbon within diatom cells. Photorespiration can represent the second largest route for carbon flux in photosynthetic organisms (Bauwe et al. 2010; Pick et al. 2013). Photorespiration occurs when RubisCO catalyzes a reaction with O_2 rather than CO_2, resulting in reduced photosynthetic carbon fixation (Eisenhut et al. 2008). The metabolic by-product of photorespiration, 2-phosphoglycolate, cannot be used directly in biosynthetic pathways and additional energy-requiring steps are necessary to recover this carbon in a usable form. Although active carbon concentrating mechanisms can maintain a higher concentration of CO_2 within the proximity of RubisCO (Reinfelder 2011; Reinfelder et al. 2004), experimental evidence suggests that diatoms must contend with photorespiratory by-products under certain environmental conditions (Schnitzler Parker et al. 2004). The existence of photorespiration in diatoms is supported by extracellular glycolate release and transcript-level induction of key enzymes involved in glycolate metabolism in response to high light exposure (Parker and Armbrust 2005; Schnitzler Parker et al. 2004). Adaptation to high light led to less glycolate production, indicating that either photorespiration was reduced or that efficient recycling pathways were in operation (Schnitzler Parker et al. 2004). In plants, 2-phosphoglycolate recycling is accomplished through a series of reactions known as the photorespiratory C2 cycle, which is distributed among the chloroplasts, leaf peroxisomes and mitochondria (Bauwe et al. 2012). However, the specific pathways utilized by diatoms to metabolize photorespiratory by-products are not yet fully understood, and represent a significant knowledge gap in terms of building accurate metabolic models for carbon metabolism in these important primary producers.

While many of the enzymes involved in the classical C2 photorespiratory pathway have been identified in diatom genomes (Kroth et al. 2008), some discrepancies exist between past experimental data and genomic information. The first step in processing photorespiratory glycolate is to convert it to glyoxylate, which can be accomplished by either glycolate dehydrogenases (GDH) or by glycolate oxidases (GOX). Biochemical studies suggest that diatoms utilize glycolate dehydrogenases (GDH) to oxidize glycolate (Suzuki et al. 1991; Winkler and Stabenau 1992), however only putative glycolate oxidases (GOX) have been identified in diatom genomes (Kroth et al. 2008). It is also unclear how similar diatom C2 pathways are to those most commonly described in other organisms. For example, glycerate kinase, the enzyme that catalyzes final step of the C2 cycle in plants, was not originally identified in diatoms (Kroth et al. 2008), although a candidate gene has been proposed (Fabris et al. 2012). Beyond this uncertainty, additional experimental evidence suggests that diatom photorespiratory metabolism may not be limited to the steps of the classical C2 cycle. For example, glycolate oxidation has been detected in both diatom peroxisomes and mitochondria, providing support for at least two possible fates for photorespiratory glycolate (Winkler and Stabenau 1995). Furthermore, reducing CO_2 concentration and increasing available light did not increase the activity of particular canonical C2 pathway enzymes in the diatom *Nitzschia laevis* (Winkler and Stabenau 1995), again supporting the possibility of a distinct photorespiratory mechanism from that commonly described.

Metabolism via the peroxisomal glyoxylate cycle has been proposed in the literature to be an alternate route to recycle photorespiratory intermediates (Fig. 1), and enzymes in this pathway are responsive to conditions that induce photorespiration in diatoms (Winkler and Stabenau 1995). The glyoxylate cycle enzymes isocitrate lyase (ICL) and malate synthase (MS) serve as a metabolic bypass to the TCA cycle, but MS can also combine photorespiratory glyoxylate with acetyl-CoA to produce malate, which then re-enters the TCA cycle in the mitochondria (Igamberdiev and Lea 2002; Winkler and Stabenau 1992) or is converted to pyruvate by malic enzyme (ME) (Sweetlove et al. 2010). This alternate fate for photorespiratory metabolites implicates malate as a potentially important compound linking the photorespiratory processes in the peroxisome with mitochondrial metabolism (Igamberdiev and Lea 2002). Another proposed pathway for photorespiratory metabolism based on research in the diatom *Cylindrotheca fusiformis* suggested that glycolate may be metabolized via enzymes derived from a bacterial glycerate pathway using a glyoxylate carboligase and tartronate semialdehyde reductase to produce glycerate (Paul and Volcani 1976). However, genome annotation efforts have not been able to identify key enzymes involved in this pathway with confidence (Kroth et al. 2008).

Photorespiration serves an additional purpose beyond carbon recovery, as the high energy requirement of phosphoglycolate recycling also makes this pathway an important means to dissipate excess energy and prevent reaction center damage (Wingler et al. 2000; Igamberdiev et al.

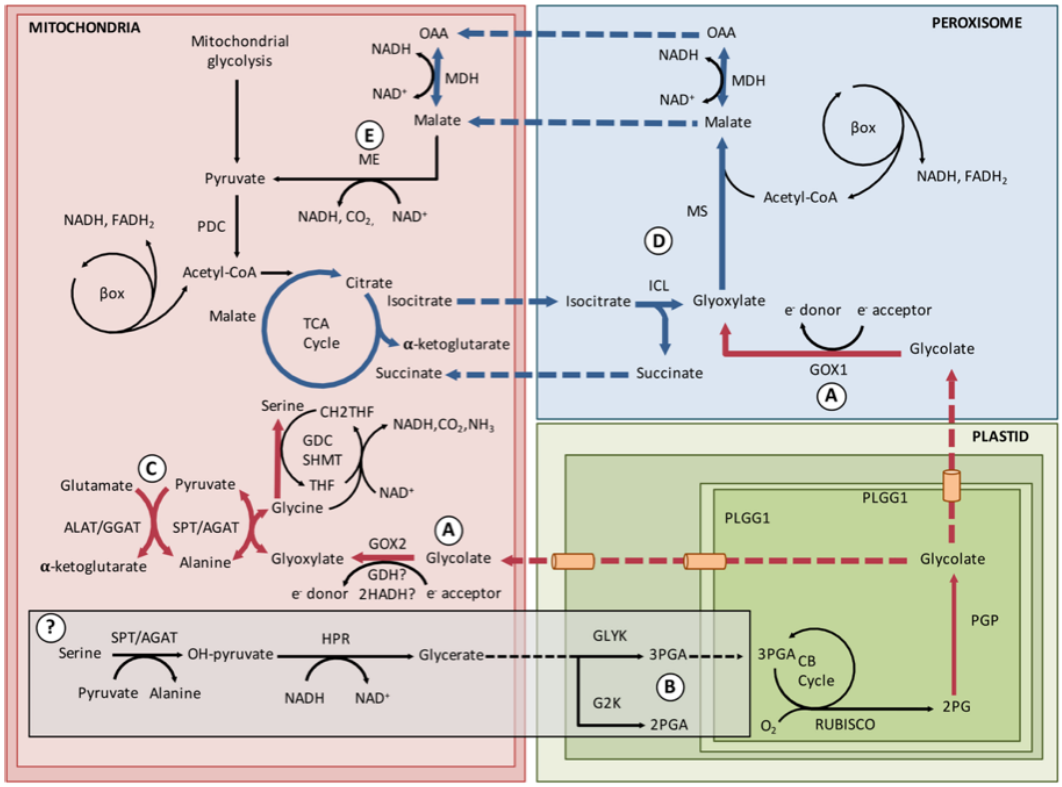

Figure 1. Proposed metabolic map based on in silico analysis and select *in vivo* validation. Map features glyoxylate cycle in blue and the C2 photorespiratory pathway in pink. Dashed lines indicate uncharacterized transport processes. Circled letters indicate metabolic processes featured in the discussion: **A**) Glycolate oxidation in the peroxisome and the mitochondria; **B**) Completion of the C2 pathway by a glycerate kinase; **C**) transamination of glyoxylate; **D**) Peroxisomal isocitrate lyase and malate synthase; **E**) Decarboxylation of malate in the mitochondria by malic enzyme;**?)** Grey box contains previously proposed reactions unsupported by our analysis. **Enzyme abbreviations**: 2HADH, 2-hydroxyacid dehydrogenase; ALAT/GGAT, alanine aminotransferase/glutamate glyoxylate aminotransferase; GDH, glycolate dehydrogenase; G2K, glycerate-2-kinase; GOX, glycolate oxidase, HPR, hydroxypyruvate reductase; ICL, isocitrate lyase; MDH, malate dehydrogenase; ME, malic enzyme; MS, malate synthase; PDC, pyruvate dehydrogenase complex; PGP, phosphoglycolate phosphatase; PLGG1, plastidic glycolate glycerate transporter; SHMT; SPT/AGAT, serine-pyruvate aminotransferase/alanine-glyoxylate aminotransferase.

2001). At high light intensities, reducing power generated by the light reactions of photosynthesis exceeds the capacity of carbon assimilating enzymes to regenerate $NADP^+$ (Wilhelm and Jakob 2011). Stress conditions that further inhibit carbon fixation can exacerbate the imbalance, resulting in overreduction of the photosynthetic electron transport chain, photoinhibition, and the production of reactive oxygen species (Wingler et al. 2000). Photorespiratory metabolism is one of several alternative electron pathways (Wilhelm and Jakob 2011) that serve to prevent overreduction of photosynthetic electron transport components, since photorespiratory intermediates transported out of the chloroplast carry energy to be distributed to other organelles. Strong evidence from inhibitor studies in *Phaeodactylum tricornutum* support a role for mitochondrial electron transport processes in dissipating excess reducing power from diatom chloroplasts (Bailleul et al. 2015), but the molecules that facilitate this interaction remain to be identified. Additionally, photorespiratory intermediates are substantially integrated with other aspects of central carbon and nitrogen metabolism (Bauwe

et al. 2012), including the production of metabolites that can further boost antioxidant production (Wingler et al. 2000). The role of photorespiration in alleviating photooxidative stress may be of particular importance in diatoms, as it has been shown that an extreme light shift is sufficient to induce a photorespiratory response (Allen 2005; Schnitzler Parker et al. 2004).

We previously reported on physiological and transcript level changes in response to silicon starvation in *T. pseudonana* (Smith et al. 2016). Silicon is a requirement for cell wall synthesis in most diatom species, and therefore necessary for cell division. During silicon starvation, cell division halts and fatty acid and triglyceride (TAG) synthesis are induced (Smith et al. 2016; Yu et al. 2009), signaling considerable changes in intracellular carbon partitioning. We found that the first 8 hours following transfer into silicon-free medium is characterized by initial photodamage to the PSII reaction centers, followed by a sustained decrease in the maximal rate of carbon fixation (Smith et al. 2016). In addition, we observed a large increase in the de-epoxidation state of the xanthophyll pool, indicating increased xanthophyll-dependent non-photochemical quenching (NPQ) to reduce electron flow and protect PSII from photooxidative damage (Smith et al. 2016). The increase in NPQ suggests the need for alternative electron cycling pathways such as photorespiration under these experimental conditions. Silicon starvation potentially offers advantages over some other experimental conditions used to induce photorespiratory responses because most cellular metabolic processes are minimally affected by the lack of silicon. In addition, silicon-starvation induced photorespiration does not involve an artificial shift in CO_2 levels. While we recognize the limitations of using transcript data as a proxy to interpret shifts in metabolic processes, ample evidence exists in the literature to demonstrate that when coordinated and consistent changes in transcript levels occur at several steps in a pathway, it is most likely indicative of a metabolic shift (Fernie et al. 2002; Lancien et al. 1999; Smith et al. 2016; Sweetlove et al. 2010). Given the lack of clarity regarding photorespiratory carbon cycling in diatoms, investigation of the transcriptional response under silicon limitation may provide new insights into this process.

Using a combination of in silico analysis, transcript expression analysis, and in vivo localization to assess metabolic potential, our goal in this study was to clarify and update a model for the organization and operation of photorespiratory metabolism in diatoms by investigating the response during silicon starvation in *T. pseudonana*. Based on the existing model, we chose to focus on the following questions: 1) What are the potential routes for glycolate oxidation and transport in *T. pseudonana*? 2) Does evidence support a complete C2 cycle with carbon recycled back to the plastid? 3) What are the respective roles of the mitochondria and the peroxisome in terms of the metabolism of photorespiratory intermediates? 4) Could malic enzyme play an auxiliary role in photorespiratory metabolism? By combining our data and analysis with data available in the literature, we generate an updated model (Fig. 1) that incorporates previous experimental and analytical approaches into a self-consistent picture of photorespiratory C2 metabolism in diatoms, and identifies aspects in need of further clarification.

Results and Discussion

Photophysiology and Transcript Changes

Transfer of *T. pseudonana* to silicon-deplete medium at half of its original culture density induced light stress, as evidenced by photophysiology data (Smith et al. 2016). We analyzed the whole genome microarray dataset from Smith et al. (2016) and determined that transcripts for a subset of genes potentially involved in photorespiratory C2 metabolism were upregulated in accordance with the timing of photostress (Fig. 2 . These genes included previously identified markers for high light induction of photorespiration in diatoms such as GDC-T and PGP (Parker and Armbrust 2005; Schnitzler Parker et al. 2004). The photophysiology data combined with the transcript data are consistent with induction of photorespiration under these experimental conditions.

We also observed highly coordinated changes in photorespiratory transcripts (Fig. 2), linking genes under what are likely common regulatory processes. Co-expression of many genes related to photorespiratory metabolism due to transcript regulatory sequences called CO_2-cAMP-Responsive Elements (CCREs) has been previously discussed (Hennon et al. 2015; Ohno et al. 2012). However, not all putative photorespiratory genes share the same expression pattern (Fig. 2), suggesting that multiple regulatory factors may come into play in the transcript-level response. Three distinct expression clusters containing multiple genes were identified (Fig. 2A-C), and the genes in each cluster fell into functional categories that will be described in detail below. Transcripts for three other genes did not cluster with the rest (Fig. 2D). All

Table 1. In silico targeting predictions for proteins associated with C2 photorespiratory metabolism and the glyoxylate cycle. Proteins are organized based on predicted or demonstrated localization to A) mitochondria, B) peroxisome, and C) chloroplast. Protein models were extended when appropriate using RNAseq evidence to obtain the most accurate targeting prediction. Predictions were made with programs available through the ExPASy Bioinformatics Resource Portal (http://www.expasy.org/tools/), with the exception of ASAFind, which is hosted by the Rocap Lab (http://rocaplab.ocean.washington.edu/tools/asafind) and is specific for secondary endosymbionts of the red algal lineage.

A) Mitochondria

Protein	Thaps3_PID	SignalP 3.0	ChloroP	ASAFind	Mitoprot	PTS1
GOX2	3353	–	N, 0.451	–	0.287	–
GDH	42965	–	Y, 0.565	–	0.312	–
GDC-H	28521	–	Y, 0.535	–	0.993	–
GDC-T	36208	–	Y, 0.507	–	0.999	–
GDC-P	39799	SNA-LL	N, 0.479	Yes, low	0.78	–
GDC-L	36716	–	Y, 0.555	–	0.486	–
SHMT	269942	–	Y, 0.520	–	0.955	–
ALAT/GGAT	412	–	Y, 0.517	–	0.949	–
AGAT/SPT	22208	–	Y, 0.583	–	0.997	–
2HADH	2846	–	N, 0.469	–	0.969	–
MDH	20726	–	Y, 0.575	–	0.9616	–
ME	34030	–	Y, 0.506	–	0.996	–

B) Peroxisome

Protein	Thaps3_PID	SignalP 3.0	ChloroP	ASAFind	Mitoprot	PTS1
GOX1	406	–	N, 0.453	–	0.43	SKL
CAT	11615	–	N, 0.447	–	0.026	SRL
MS	26293	–	N, 0.434	–	0.036	SKL
ICL	35523	–	N, 0.449	–	0.471	–
MDH	41425	–	N, 0.442	–	0.762	SRI

C) Chloroplast

Protein	Thaps3_PID	SignalP 3.0	ChloroP	ASAFind	Mitoprot	PTS1
PGP	25544	TTA-FS	Y, 0.588	Yes, high	0.914	–
PLGG1	9352	SVA-LV	Y, 0.563	Yes, low	0.828	–
	14106	IVA-TT	Y, 0.512	No	0.904	–
GK	261750	SSA-SS	Y, 0.544	No	0.968	–

genes included in subsequent analysis are listed in Table 1.

Multiple Routes for Glycolate Oxidation in Diatoms

The pathway that determines the metabolic fate of photorespiratory glycolate remains unresolved in diatoms. Our analysis, coupled with previous biochemical analyses, suggests the presence of a single glycolate oxidizing activity in the peroxisome (GOX1), and potentially two genes with glycolate oxidizing activity (GOX2 and a previously unidentified GDH), with the possibility of a third (2-HADH – see section on "Processing of photorespiratory glyoxylate in the mitochondria" below) in the mitochondria (Fig. 1).

Glycolate oxidation can be catalyzed by either glycolate oxidase (GOX) or glycolate dehydrogenase (GDH). Characterized GOX enzymes use O_2 as an electron acceptor, generate H_2O_2, are insensitive to cyanide and carry out a secondary reaction with L-lactate (Frederick et al. 1973). In contrast, GDHs utilize NAD^+/$NADP^+$ as an electron acceptor, are inhibited by cyanide and carry out a secondary reaction with D-lactate (Frederick et al. 1973). Biochemical studies in whole cell and organelle fractions of diatoms have shown a lack of H_2O_2 production, consistent with diatoms using GDH to

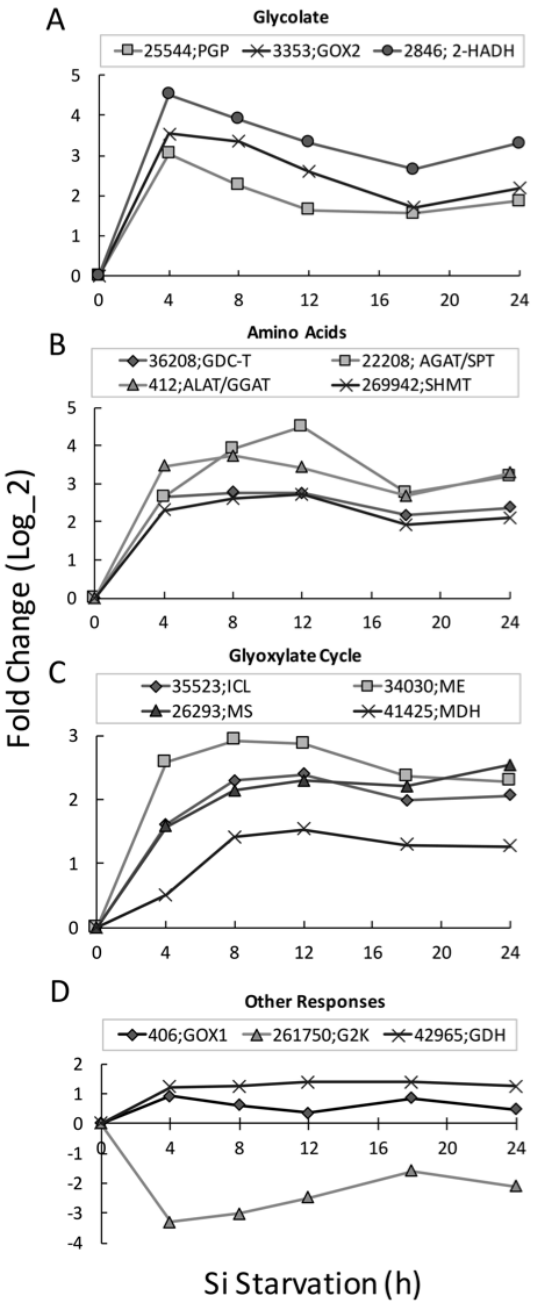

Figure 2. Expression of genes associated with photorespiratory processes and the glyoxylate cycle in *T. pseudonana* during the silicon starvation time-course. **A**) Expression of genes putatively involved in the production of and oxidation of photorespiratory glycolate; **B**) Expression of genes proposed to participate in the metabolism of photorespiratory intermediates and amino acid synthesis in diatoms; **C**) Expression of genes proposed to participate in the glyoxylate cycle in diatoms; **D**) Genes with expression patterns that are distinct from those involved in photorespiratory processes and the glyoxylate cycle. Clustering was done manually by normalizing transcript levels across each time point, then calculating the standard deviation of the mean for each point. The mean variation of the standard deviation in terms of percent for each point ranged from 3.7-13%.

generate glyoxylate in both the mitochondria and the peroxisome (Paul et al. 1975; Suzuki et al. 1991; Winkler and Stabenau 1992). However, lack of H_2O_2 production could be explained by efficient catalase activity, ascorbate peroxidase activity, or use of alternate electron acceptors (Mittler 2002; Paul et al. 1975; Paul and Volcani 1975; Suzuki et al. 1991), and insensitivity to cyanide and a preference by purified peroxisomes for oxidation of L-lactate is consistent with GOX activity (Suzuki et al. 1991; Winkler and Stabenau 1992). The conflicting biochemical evidence prompted a bioinformatic analysis of putative GDH and GOX genes from diatom genomes to attempt to clarify the enzymatic potential in each cellular compartment.

Two gene candidates for glycolate oxidation (GOX1 and GOX2) have previously been identified from genomic sequence (Kroth et al. 2008). To investigate the relationship of these sequences to other GOXs, we generated a maximum likelihood tree to place GOX1 and GOX2 among other characterized lactate oxidase (LOX) and GOX proteins (Fig. 3). The peroxisomal GOX1 protein, encoded by Thaps3_406, has a PTS1 C-terminal tripeptide targeting sequence (Kroth et al. 2008) (Table 1; Fig. 1A) and groups with other LOX/GOX proteins found in higher plants, green algae, and cyanobacteria (Fig. 3). This finding is similar to previous reports (Kern et al. 2011), and is consistent with the GOX-like biochemical characteristics observed in diatom peroxisomes (i.e. cyanide insensitivity and L-lactate preference). The GOX1 transcript is not coordinately expressed with other genes involved in metabolizing photorespiratory glycolate in transcriptome datasets (Fig. 2; Hennon et al. 2015), however, a substantial induction of peroxisomal glycolate oxidation activity was demonstrated under more severe photorespiratory stimulating conditions (Winkler and Stabenau 1995). We were unable to identify any other candidate genes encoding glycolate oxidation activity in the peroxisome,

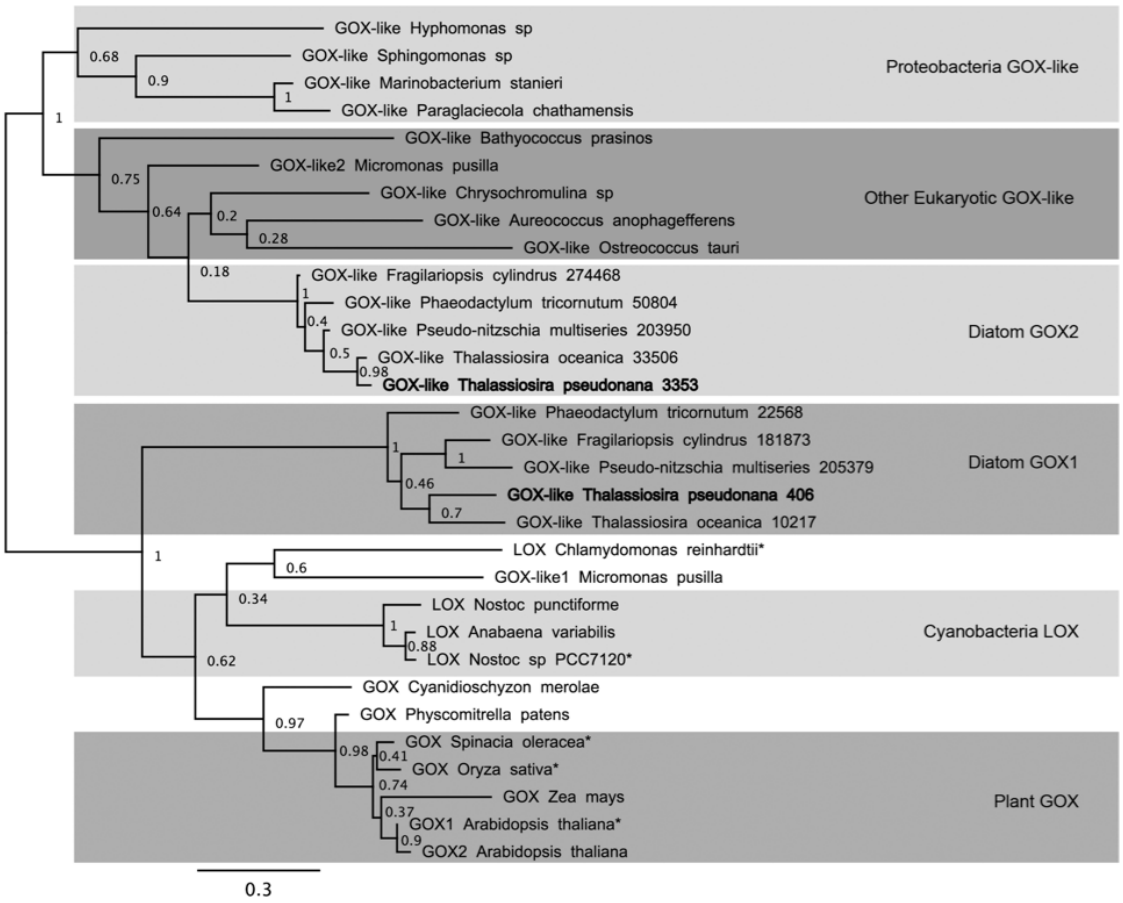

Figure 3. Phylogenetic analysis of glycolate oxidizing proteins found in proteobacteria, photosynthetic eukaryotes and cyanobacteria. 31 sequences were aligned using the MAFFT algorithm, and the MEGA6 software was used to estimate the maximum likelihood tree using 1000 bootstrap replicates. The distance scale (substitutions per site) is shown in the bottom left-hand corner. Enzymes that are included in the SWISS-Prot database with biochemical evidence to support their annotation are indicated with an asterisk. Diatom glycolate oxidase-like proteins (GOXs) form two distinct clades. In *T. pseudonana*, GOX1 and GOX2 are localized in peroxisomes and mitochondria, respectively.

suggesting that GOX1 is the gene responsible for this activity.

Diatom GOX2 sequences were highly divergent from diatom GOX1 (Fig. 3). GOX2 sequences, along with other GOX-like sequences from various eukaryotic algae, are more similar to GOX-like proteins in proteobacteria. None of the proteins within this group have been biochemically characterized, so it is possible that they may be distinct from other GOXs studied to date. The *in silico* targeting of GOX2 (Thaps3_3353) is ambiguous in *T. pseudonana* (Table 1), despite a clear mitochondrial targeting signal in the homolog of this gene in *P. tricornutum* (Kroth et al. 2008) (Fig. 1A). We clarified the localization of this protein in *T. pseudonana* by generating and expressing a C-terminal GFP fusion. GFP fluorescence confirmed that the GOX2 protein was localized to the mitochondria (see Supplementary Material Fig. S1 in the online version at DOI: 10.1016/j.protis.2016.10.005) as verified through co-localization with the mitochondrial citrate synthase (Fig. 4). Supporting a role in metabolizing phosphoglycolate, GOX2 transcript levels were co-induced with other known photorespiratory genes (Fig. 2; Hennon et al. 2015).

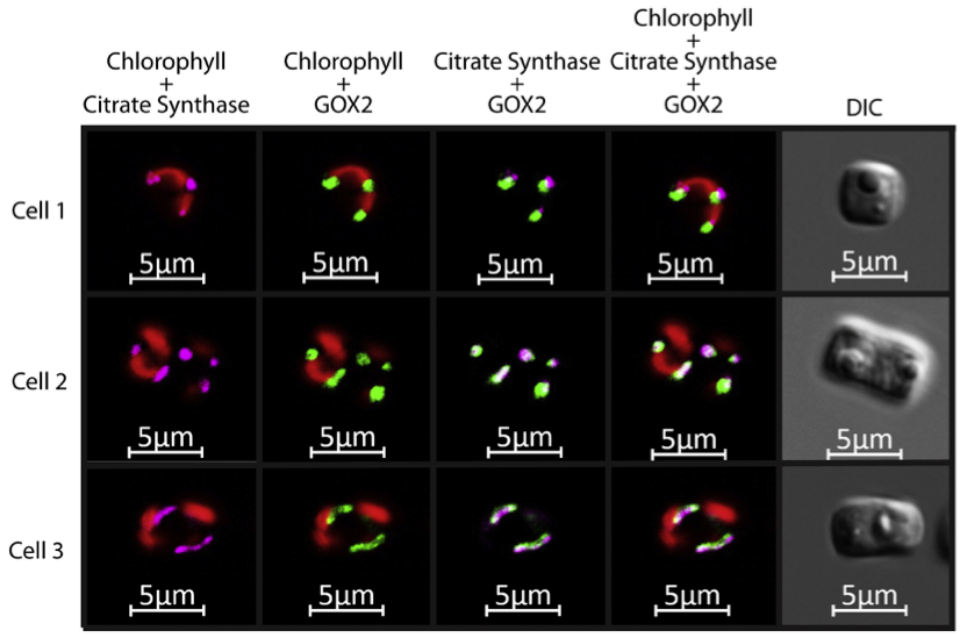

Figure 4. Colocalization of glycolate oxidase 2 (GOX2) with mitochondrial citrate synthase (CS). Representative *T. pseudonana* cells. False color images represent GFP in green, TagRFP in pink and chlorophyll autofluorescence in red. White demonstrates co-localization of the fluorescent proteins. Expression of GOX2 from the plasmid pTp FCP/Tp3353_gfp and CS from from pTp rpL41/Tp11411-TagRFP demonstrates that GOX2 is co-localized with mitochondrial CS. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Glycolate oxidizing enzymatic activity in isolated diatom mitochondria suggested a preference for a secondary reaction with D-lactate, consistent with the presence of a GDH (Winkler and Stabenau 1992), which prompted us to search the *T. pseudonana* genome for candidates responsible for this activity. A putative GDH gene (Thaps3_42965) containing a GlcD domain with sequence similarity to FAD-linked oxidases and D-lactate dehydrogenases (BLAST e-value $1e^{-101}$) that has not previously been suggested as a candidate for glycolate oxidation was found. Homologs for this gene were also identified in *P. tricornutum* and *Phytophthora sojae*. The gene product of Thaps3_42965 had a weak Mitoprot score (0.312), but was predicted to be targeted to mitochondria by TargetP (Emanuelsson et al. 2007), and the *P. tricornutum* homolog had high predicted mitochondrial targeting by both programs. Thus, the gene product is a candidate for the mitochondrial GDH activity previously characterized biochemically (Winkler and Stabenau 1992). Although induced in our conditions, the GDH-like protein was not tightly coexpressed with other photorespiratory genes as was GOX2 (Fig. 2). It also did not cluster with other photorespiratory genes in other transcriptomics data (Hennon et al. 2015). Ultimately, we are unable to definitively assign a lone glycolate-oxidizing candidate to the mitochondria of *T. pseudonana* based on the sequence content, gene expression, or localization information. However, evidence supports at least two candidates for this reaction (Thaps3_3353 and Thaps3_42965), and it is possible that either may operate in diatom mitochondria depending on the specific set of conditions.

Recycling of Photorespiratory Intermediates into the Chloroplast

Whether or not diatoms recycle photorespiratory intermediates completely back to the chloroplasts has been debated (Fabris et al. 2012; Igamberdiev and Lea 2002; Kern et al. 2011; Kroth et al. 2008; Winkler and Stabenau 1995) due to an apparent lack of a plastidial glycerate kinase (GLYK) (Kern et al. 2011; Kroth et al. 2008) (Fig. 1B). However, a more recent study using an orthology-based approach identified a GLYK candidate in *P. tricor-*

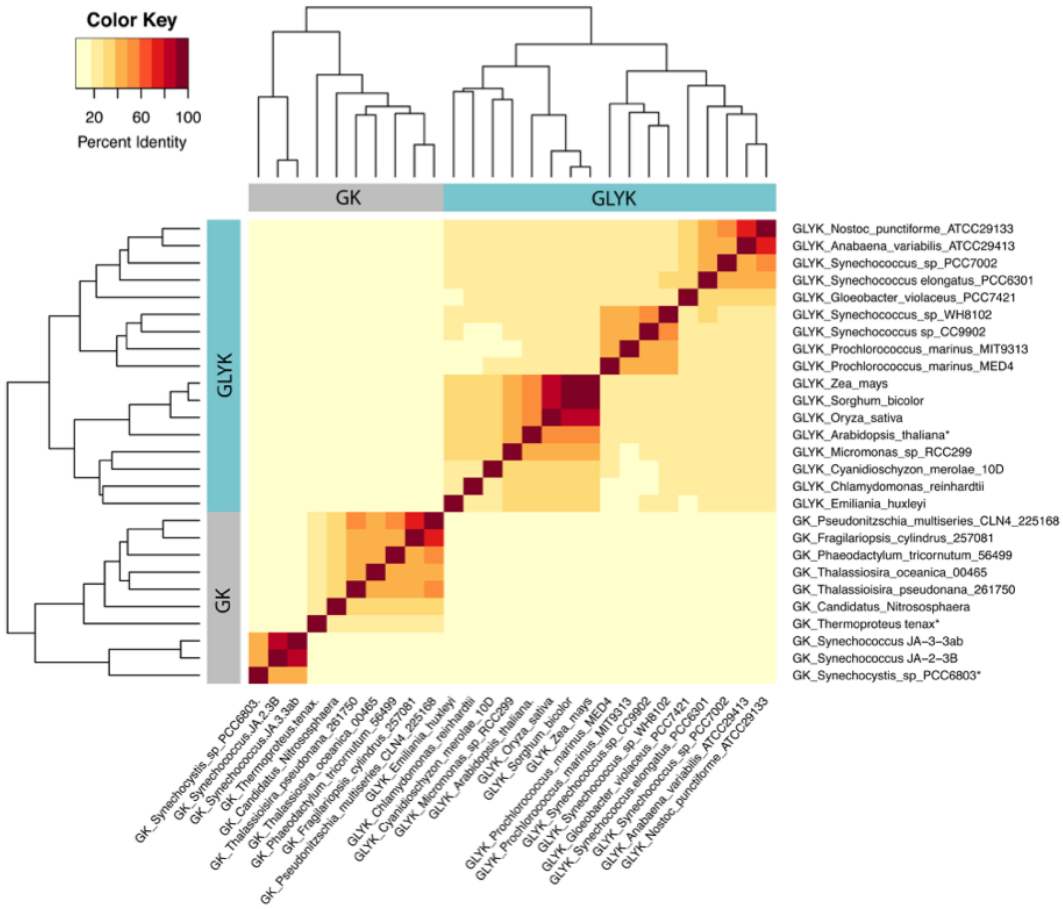

Figure 5. Heatmap representing calculated percent identity among 27 glycerate kinase (GLYK, 3-PGA forming, blue bar) and glycerate-2-kinase (GK, 2-PGA forming, gray bar) protein sequences. Biochemically characterized enzymes are noted with an asterisk. Diatom GK sequences are distinct from known GLYKs found in cyanobacteria, algae, and higher plants. Diatom GK sequences are most similar to archaeal GKs, such as the characterized sequence from *Thermoproteus tenax*, suggesting a unique origin for these enzymes. In addition, the lack of a GLYK sequence in diatom genomes supports the idea that the canonical C2 photorespiratory pathway is incomplete in diatoms. Hierarchical clustering of percent identity scores was done in R (http://www.R-project.org/) using the heatmap.2 function within the gplots package (http://CRAN.R-project.org/package=gplots). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

nutum (Fabris et al. 2012). We found that although the homolog of this putative GLYK in *T. pseudonana* (Thaps3_261750) is similar to other GLYK sequences, it has substantially higher sequence similarity to glycerate-2-kinases (GKs) that produce glycerate 2-phosphate (2PGA) rather than 3PGA (Bartsch et al. 2008), including a biochemically characterized GK from the archaea *Thermoproteus tenax* (Kehrer et al. 2007) (Fig. 5). The Thaps3_261750 protein has predicted periplastid targeting (Gould et al. 2006) (Table 1), and our analysis suggests the same for the *P. tricornutum* (Phatr2_54867) homolog. In addition, the transcript for the putative GK is substantially downregulated, which is inconsistent with the other genes in the C2 pathway (Fig. 2), implying a reduction of glycerate

processing by this enzyme under our experimental conditions. Taken together, the data does not support the idea that a GLYK returns photorespiratory carbon to diatom plastids.

Another route for reintroducing 3PGA back into the Calvin-Benson cycle is possible, which is consistent with the predicted location of enzymes. If a GK converts glycerate to 2PGA in the periplastid compartment, 2PGA could be transported into the chloroplast by an unidentified transporter (2PGA transporters have been characterized in bacteria) (Saier et al. 1975), then converted to 3PGA by a chloroplast-localized phosphoglycerate mutase (PGAM; (Thaps3_263397). Alternatively, a periplastid-localized PGAM (Thaps3_260919) could generate 3PGA which could be transported into the plastid. Arguing against these processes occurring under our experimental conditions is the fact that the GK transcript is highly downregulated.

Transport is another consideration in the completion of a C2 cycle. In higher plants, glycerate is transported into the plastid before it is converted into 3PGA by a GLYK. This import is mediated by a glycolate/glycerate transporter (PLGG1) in plastid membranes (Pick et al. 2013; Reumann and Weber 2006). We identified two genes with homology to plant PLGG1 (Thaps3_9352, Thaps3_14106) in *T. pseudonana* with predicted chloroplast and periplastid membrane targeting, respectively (Table 1). Although the PLGG1 transporter is commonly depicted as a glycolate/glycerate exchanger, it actually functions via a unidirectional proton symport mechanism and can recognize either substrate (as well as lactate and glyoxylate) and transport them in either direction (Young and McCarty 1993). Thus there is no a priori reason to correlate the presence of PLGG1 with a requirement for glycerate uptake into the chloroplast since the transporter may simply function as a glycolate exporter. The lack of predicted chloroplast targeted enzymes to process glycerate in the chloroplast is consistent with this possibility.

The lack of an identifiable GLYK candidate and downregulation of GK are consistent with diatoms using a distinct mechanism for recovering photorespiratory C2 rather than recycling it directly back into the Calvin-Benson cycle. This led us to investigate other possible routes for reprocessing carbon in photorespiratory metabolism in *T. pseudonana*.

Processing of Photorespiratory Glyoxylate in the Mitochondria

The metabolism of glyoxylate by C2 cycle enzymes is predicted to occur primarily in the mitochondria of diatoms (Kroth et al. 2008) (Table 1, Fig. 1). It has been proposed (Kroth et al. 2008) that mitochondrial glyoxylate is converted to glycine by a serine-pyruvate/alanine-glyoxylate aminotransferase (SPT/AGAT), followed by conversion to serine by stepwise reactions of glycine decarboxylase (GDC) and serine hydroxymethyltransferase (SHMT). Exhibiting substrate flexibility, the same SPT/AGAT is proposed to use serine to generate hydroxypyruvate (Fig. 1, gray box), which is further converted to glycerate by a hydroxypyruvate reductase (HPR). In this scheme, SPT/AGAT would compete with itself by both generating and utilizing serine, which may be inefficient. In addition, careful biochemical analysis failed to identify HPR activity in diatom mitochondria, even under conditions when other photorespiratory enzymes were detected (Winkler and Stabenau 1995). These issues, plus the possible lack of importance of glycerate in recycling photorespiratory carbon (discussion above), prompted us to consider alternative means of processing of glyoxylate in diatom mitochondria.

We identified an additional transaminase similar to alanine amintotransferase/glutamate glyoxylate amintotransferase (ALAT/GGAT, Thaps3_412) that exhibited similar transcript changes as other mitochondrial C2 pathway genes (Fig. 2B). The Thaps v.3 gene model is prematurely truncated, but N-terminal extension using RNAseq data revealed a mitochondrial signal peptide (Table 1). A similar ALAT with GGAT activity catalyzes the transamination of glyoxylate in the photorespiratory pathway of *Arabidopsis* (Liepman and Olsen 2003), and the presence of this enzyme is consistent with previous detection of GGAT activity in diatom mitochondria (Winkler and Stabenau 1995).

SPT/AGAT and ALAT/GGAT are both capable of carrying out multiple reactions with different substrates. The lack of HPR activity (Winkler and Stabenau 1995) suggests no need for the reaction by SPT/AGAT that converts glyoxylate and serine to hydroxypyruvate and glycine, therefore, we propose that this enzyme (Thaps3_22208) carries out its alternative reaction by converting glyoxylate and alanine to pyruvate and glycine (Fig. 1C). Induction of transcripts (Fig. 2B) for GDC-T (Thaps3_36208) and SHMT (Thaps3_269942) suggest that glycine is further converted to serine (Fig. 1). We propose that the ALAT/GGAT (Thaps3_412) catalyzes the conversion of pyruvate and glutamate to α-ketoglutarate and alanine. The transcript pattern for the ALAT/GGAT, GDC-T, SHMT, and SPT/AGAT are similar, with SPT/AGAT having a more prolonged upregulation than the others (Fig. 2B).

The end results of this proposed scheme are that glycine and serine would be the primary amino acids synthesized from photorespiratory C2, and that α-ketoglutarate levels would also be increased under photorespiratory conditions. Diatom labeling studies documented specific labeling of serine and glycine as a result of photorespiratory processes, and a substantial increase in labeling of α-ketoglutarate relative to non-photorespiratory conditions (Huang et al. 2015). Thus, the model is entirely consistent with labeling studies (Huang et al. 2015; Zheng et al. 2013).

This model and biochemical evidence (Winkler and Stabenau 1995) predict that hydroxypyruvate is not formed in the mitochondria via photorespiratory processes. A mitochondrially-targeted HPR was previously proposed (Kroth et al. 2008), but our analysis indicated that it was the previously discussed GK (Fig. 5). We identified another *T. pseudonana* gene, Thaps3_2846, with weak similarity to plant and human HPRs, but high similarity to bacterial 2-hydroxyacid dehydrogenases (2-HADHs). 2-HADHs can perform diverse functions in many aspects of metabolism, and previous phylogenetic analyses have demonstrated that members of the HADH family clade well according to substrate specificity (Fauvart et al. 2007). In an effort to better annotate the 2-HADH proteins in the *T. pseudonana* genome, we generated a sequence alignment that includes 2-HADH proteins that have experimentally verified substrate specificity. We observed that the diatom 2-HADH sequences, including Thaps_2846, fall within a large and diverse group containing glyoxylate/hydroxypyruvate reductases, 2-ketogluconate reductases, and phenylpyruvate reductases (see Supplementary Material Fig. S2 in the online version at DOI: 10.1016/j.protis.2016.10.005), thus its substrate specificity still remains unclear. However, comparison of conserved amino acids known to be functionally important based on the human HPR 3D structure indicate that two critical residues (W-141 and S-296) are missing in the diatom 2-HADHs (data not shown), arguing against HPR activity. The transcript pattern for Thaps3_2846 clusters tightly with PGP and GOX2 (Fig. 2), suggesting that the enzyme may instead play a role in glyoxylate reduction. In addition, these 2-HADHs have a preference for D-stereoisomers, which is consistent with the D-lactate preference in the mitochondria (Winkler and Stabenau 1992). Confirmation of the exact activities and role of Thaps3_2846 will require experimental validation.

Peroxisomal Processing of Glycolate Through the Glyoxylate Cycle

The glyoxylate cycle offers an alternate route to convert glycolate into useful biosynthetic precursors (Kunze et al. 2006). Glyoxylate is produced in the peroxisome by isocitrate lyase (ICL, Thaps3_35523) during β-oxidation, and can be produced from photorespiratory glycolate by GOX1, (Thaps3_406). Glyoxylate combined with acetyl-CoA forms malate by the activity of malate synthase (MS, Thaps3_26293) (Fig. 1D). Biochemical studies showed that MS and ICL enzyme activities are induced under conditions of high irradiance and low CO_2 availability, suggesting that the glyoxylate cycle is involved in processing glycolate in the peroxisome (Winkler and Stabenau 1995). Our transcript data show coordinate upregulation of MS, ICL and malate dehydrogenase (MDH, Thaps3_41425), but not GOX1, suggesting that the glyoxylate cycle is responsive to our conditions, but that the formation of glyoxylate via photorespiratory glycolate may not be a prominent contributor to glyoxylate formation. Alternatively, GOX1 may be constitutively expressed and not regulated at the transcript level. The GOX1 transcript response also does not cluster with the photorespiratory response in other datasets (Hennon et al. 2015). To substantiate the possible role of the glyoxylate cycle, we analyzed and experimentally verified the location of enzymes involved in it. MS sequences from both *T. pseudonana* and *P. tricornutum* were analyzed and found to contain canonical C-terminal PTS1 signals recognized by a Pex5 receptor protein that facilitates transport into the peroxisome (Rucktäschel et al. 2011). In isolated peroxisomes of the diatom *Fragilaria* ^{14}C-labeled malate is produced from ^{14}C-labeled glycolate, supporting the localization of this reaction (Winkler and Stabenau 1992). For these reasons, we used MS as a peroxisomal marker. *T. pseudonana* cells transformed with GFP-tagged MS produced labeling in two closely-associated spots per cell (Fig. 6) that were similar in size (≤1 μm) as peroxisomes described in other organisms (van den Bosch et al., 1992).

While significant efforts have advanced our understanding of the signals that direct proteins into secondarily-derived plastids of diatoms (Apt et al. 2002; Gruber et al. 2007, 2015), signals that dictate targeting to other cellular compartments, such as the peroxisome, are just beginning to emerge (Gonzalez et al. 2011). In plants, ICL is targeted to the peroxisome through PTS1, but ICL is known to be cytosolic in some fungi (Kunze et al. 2006). *T. pseudonana* ICL (Thaps3_35523)

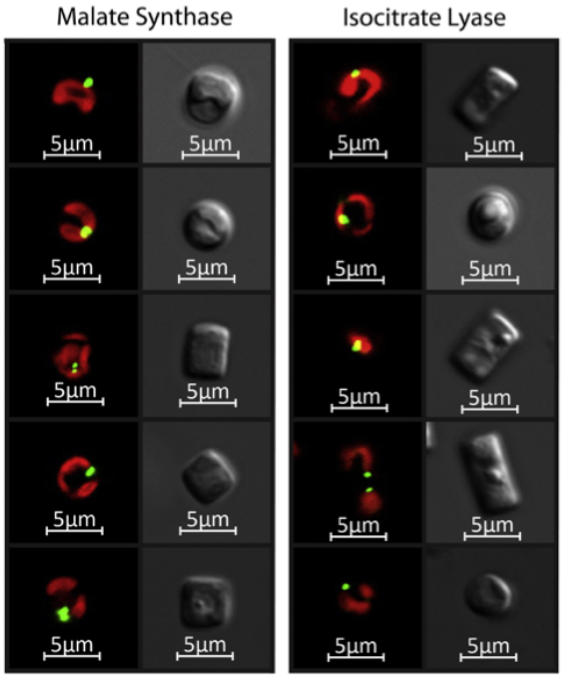

Figure 6. Localization of malate synthase and isocitrate lyase in *T. pseudonana*. Representative cells of *T. pseudonana* in different orientations. False color images represent GFP in green and chlorophyll autofluorescence in red. **A**) Expression of MS from plasmid pTp_native/Tp26293-bfloa1_gfp and ICL from plasmid pTp_FCP/Tp35523-gfp demonstrate peroxisomal localization that typically appeared as two distinct, but closely associated spots per cell that are less than or equal to 1 µm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

lacks the PTS1 sequence (Table 1), yet expression of GFP-fused ICL resulted in similar labeling as that observed for MS, supporting peroxisomal localization that was expected based on previous biochemical studies (Fig. 6) (Winkler and Stabenau 1995). Diatoms appear to lack alternate peroxisomal import systems that have been described in other organisms (i.e. PTS2 targeting) (Gonzalez et al. 2011). Therefore, the mechanism targeting ICL to diatom peroxisomes remains unclear and points to the possibility of the presence of an uncharacterized peroxisomal import signal within the ICL sequence. In support of this possibility, so-called non-PTS proteins have been described, including MS and ICL from *Candida albicans* (Piekarska et al. 2008) and ICL from castor bean (Parkes et al. 2003), that are imported to peroxisomes in a PTS1-independent but Pex5-dependent manner (Rucktäschel et al. 2011; van der Klei and Veenhuis 2006). The in vivo localization of ICL demonstrates the presence of key glyoxylate cycle enzymes in peroxisomes of *T. pseudonana*.

In higher plants, the glyoxylate cycle utilizes acetyl-CoA generated by the ß-oxidation of lipids, for example, during mobilization of storage lipids during seed germination or during senescence due to periods of prolonged dark exposure when endomembrane systems are degraded (Kunze et al. 2006). Therefore, the glyoxylate cycle may be an important mechanism for the repackaging of acetyl-CoA from oxidized lipids. The breakdown and repackaging of membrane lipids is a documented response in diatoms during nutrient-limited growth arrest and neutral lipid accumulation (Yang et al. 2013).

The Anapleurotic Role of Malic Enzyme

Malate or OAA generated by the glyoxylate cycle may be transported into the mitochondria and enter the TCA cycle directly (Igamberdiev and Lea 2002),

or be converted into pyruvate via the action of malic enzyme (Fig. 1E) (Plaxton and Podestá 2007; Sweetlove et al. 2010). Malic enzyme is a C4 decarboxylase with diverse functional roles in photosynthetic eukaryotes. Malic enzyme isozymes use either NAD^+ or $NADP^+$ as a cofactor and can be targeted to the cytosol, chloroplast, and/or the mitochondria. While a chloroplast-localized NADP-dependent ME would be more consistent with roles associated with C4 photosynthesis and/or fatty acid biosynthesis, a cytosolic or mitochondrial NAD-dependent ME is less likely to be involved in these pathways. Previous analyses of ME amino acid sequence suggests that NAD may be the preferred cofactor for diatom MEs (Kroth et al. 2008). To confirm this analysis, we compared NAD(P) binding sites from diatom MEs (see Supplementary Material Table S1 in the online version at DOI: 10.1016/j.protis.2016.10.005) and from several manually curated and biochemically validated MEs from the Uniprot database (see Supplementary Material Fig. S3 in the online version at DOI: 10.1016/j.protis.2016.10.005). Our analysis demonstrates that the *T. pseudonana* ME (Thaps3_34030) nucleotide binding sequence groups with other known NAD-specific binding sites, as do most of the other diatom sequences. A notable exception is a second *P. tricornutum* isoform (Pt_51970), which groups with NADP-dependent ME sequences found in plants, highlighting the possibility that this ME may serve a different functional role in P. tricornutum (Smith et al. 2012).

Conflicting information concerning the intracellular localization of diatom ME exists in the literature. Although the Thaps_34030 homolog found in *P. tricornutum* also has a mitochondrial targeting sequence (Kroth et al. 2008), experimental localization in *T. pseudonana* suggested this protein was cytosolic (Tanaka et al. 2014). With additional RNAseq support, we extended the N-terminus of the protein model based on RNAseq coverage to uncover a mitochondrial signal (see Supplementary Material Fig. S4 in the online version at DOI: 10.1016/j.protis.2016.10.005). We also generated a GFP fusion construct using native promoter/terminator regions for ME and observed colocalization to the mitochondria with the TCA cycle enzyme citrate synthase (CS; Thaps3_11411) (Fig. 7; see Supplementary Material Fig. S1 in the online version at DOI: 10.1016/j.protis.2016.10.005).

Both the similarity to NAD-dependent enzymes and the mitochondrial localization of ME suggests an anapleurotic role through the provision of pyruvate for the synthesis of TCA cycle intermediates. This is consistent with previous transcriptomic, proteomic, and enzyme activity data that suggest ME is not directly involved in a biochemical carbon concentration mechanism (Kustka et al. 2014; Clement et al. 2015). In addition, coordinate expression of ME with photorespiratory (Hennon et al. 2015) and glyoxylate cycle genes (Fig. 2C) suggest this enzyme may have an important supporting role in mitochondrial metabolism.

In plants, the activity of mitochondrial ME maintains TCA cycle function during conditions of low glycolytic flux and reduced pyruvate availability (Kromer 1995). Under such scenarios, ME generates an alternative source of pyruvate for acetyl-CoA formation and utilization in the TCA cycle (Sweetlove et al. 2010). This arrangement conveys some metabolic flexibility, as it allows for malate to serve as the sole substrate for the TCA cycle if necessary (Sweetlove et al. 2010). While the contribution of ME-derived pyruvate is typically small, flux through this pathway is significant under certain conditions, including during fatty acid accumulation in oilseed embryos (Plaxton and Podestá 2007; Schwender et al. 2004, 2006; Sweetlove et al. 2010). In our conditions, malate produced by the glyoxylate cycle may be routed to mitochondria where it can serve a similar role in supplementing the TCA cycle. Furthermore, the observed increase in glyoxylate cycle and ME transcripts during silicon starvation supports the idea that ME serves as a connection between peroxisomal metabolism and the repurposing of carbon for biosynthetic precursors. Increases in TAG and chlorophyll content are hallmarks of silicon starvation in *T. pseudonana* (Smith et al. 2016), and these carbon-intensive processes may reduce pyruvate availability for mitochondrial respiration and increase the flux of carbon skeletons (in the form of TCA cycle intermediates) out of this organelle. Malate, coupled with the activity of ME, can serve as an alternative source of organic carbon with the ability to support biosynthetic and respiratory processes of mitochondrial metabolism during silicon starvation in diatoms.

Implications of the Model

Physiological data (Smith et al. 2016), coupled with an induction of photorespiratory transcripts, suggests that photorespiration may be induced as a mechanism to manage an imbalance in cellular reducing potential in our experimental conditions. Recent evidence suggests that diatoms maintain favorable ATP/NADPH ratios in the chloroplast through the transfer of reducing equivalents to mito-

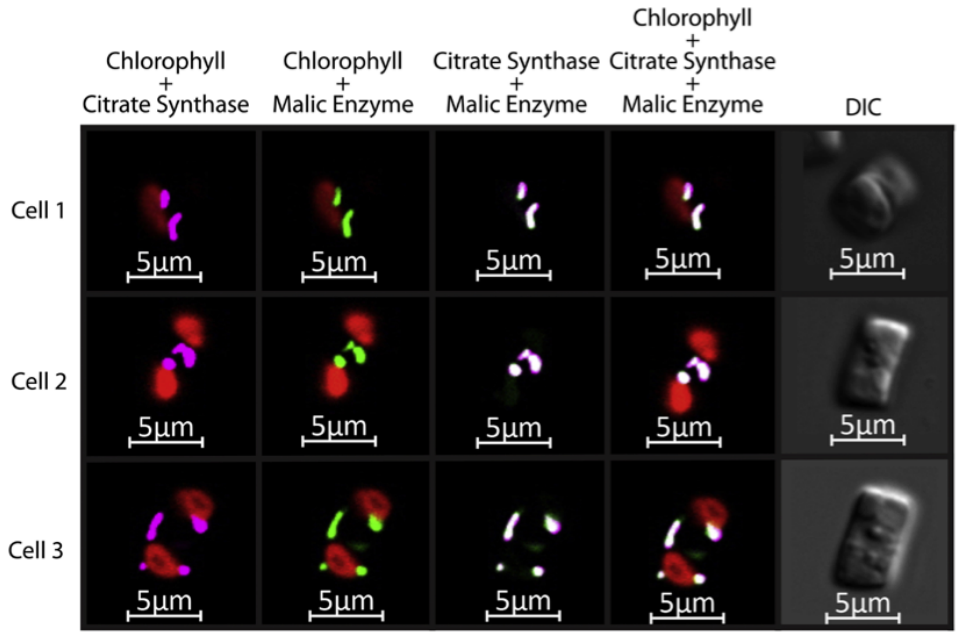

Figure 7. Colocalization of malic enzyme (ME) with mitochondrial citrate synthase (CS). Representative *T. pseudonana* cells. False color images represent GFP in green, TagRFP in pink and chlorophyll autofluorescence in red. White demonstrates co-localization of the fluorescent proteins. Expression of ME from the plasmid pTp Native/Tp5100-gfp and CS from pTp rpL41/Tp11411-tag_rfp demonstrate that ME is co-localized with mitochondrial CS. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

chondria, where reductive potential can be used to drive ATP production or be dissipated in the electron transport chain via transfer to O_2 (Allen et al. 2006, 2008; Bailleul et al. 2015; Hockin et al. 2011; Kim et al. 2015; Prihoda et al. 2012). The latter process involves the activity of alternative oxidase (AOX), which transfers electrons directly to molecular oxygen without contributing to ATP synthesis (Bailleul et al. 2015; Vanlerberghe 2013). In plants, AOX is subject to many layers of regulation including redox regulation and activation by pyruvate and photorespiratory intermediates (Vanlerberghe 2013), suggesting the possibility that the buildup of these metabolites in diatom mitochondria may also signal AOX that the cellular reducing status needs to be adjusted. The details of a mechanism for transferring reducing potential between the chloroplast and the mitochondria remain unclear (Bailleul et al. 2015), although dihydroxyacetone phosphate (DHAP) was recently put forth as a possible candidate carrier molecule (Kim et al. 2015). Given our identification of three possible candidates for glycolate oxidation in the mitochondria of *T. pseudonana* (Thaps3_3353, Thaps3_42965, Thaps3_2846), including confirmation of mitochondrial localization for GOX2 (Fig. 4, see Supplementary Material Fig. S1A in the online version at DOI: 10.1016/j.protis.2016.10.005), we suggest glycolate may also participate in the transport of reducing potential and carbon skeletons from the chloroplast to the mitochondria under certain conditions (Fig. 1A), and may also support active energy dissipating mechanisms (such as AOX) within the mitochondria. Consistent with this was the transcript-level induction (1.5-fold) of a mitochondrial AOX (Thaps3_42992) in our experimental conditions.

Our analysis does not support the presence of a GLYK or a convincing HPR in diatom genomes (Fig. 5), and is in agreement with previous assertions that the C2 cycle is incomplete (Igamberdiev and Lea 2002; Kroth et al. 2008). The implica-

tion of this finding is that photorespiratory carbon is repurposed initially into other metabolic processes, rather than being returned directly to the Calvin-Benson cycle. Instead, we propose that the transaminases SPT/AGAT (Thaps3_22208) and ALAT/GGAT (Thaps3_412) work in conjunction to produce precursors for amino acid biosynthesis (Fig. 1C). This alternative approach simultaneously allows for the dissipation of energy and the repurposing of carbon skeletons within the same organelle and may be especially advantageous under extreme conditions when the rate of carbon fixation (and therefore the requirement for 3PGA in the plastid) is low.

Additionally, we suggest that peroxisomal processes may be interconnected with photorespiratory metabolism in diatoms, and this multicompartmentation may afford some flexibility in responding to different environmental conditions. For example, mitochondrial photorespiratory processes permitted cells to satisfy carbon backbone demand for reassimilation of nitrogen released from intracellular stores (Huang et al. 2015). Consistent with this was down-regulation of enzymes involved in the glyoxylate cycle during nitrogen limitation (Yang et al. 2013). Alternatively, when diatoms were cultivated in low CO_2 conditions, biochemical studies found that glycolate oxidation activity was substantially increased in the peroxisome as were glyoxylate cycle enzymes (Winkler and Stabenau 1995). This suggests carbon-limited cells salvage photorespiratory glycolate and convert it into malate, which can be metabolized through other pathways, minimizing carbon and energy losses associated with the C2 cycle (Winkler and Stabenau 1992). The differential biochemical response in the mitochondria and peroxisome to moderate versus severe photorespiration (Winkler and Stabenau 1995) coupled with transcriptome analysis (Hennon et al. 2015) (Fig. 2) suggest that primary processing of photorespiratory carbon occurs in the mitochondrion, and that peroxisomal processing can contribute as an "overflow valve" when conditions are more severe. Collectively these findings provide a framework through which we can interpret the response of diatoms during silicon starvation-induced growth arrest.

In our conditions, we observe induction of transcripts likely involved in mitochondrial glycolate oxidation (Fig. 2A), including a mitochondrial GOX2 (Fig. 4), concomitant with a high light shift and the onset of cell cycle arrest. The transport and subsequent oxidation of glycolate in the mitochondria may be one strategy to mitigate damage from overreduction of the chloroplast. Additionally, we observe an increase in the transcripts of genes that likely participate in the conversion of glyoxylate to glycine, serine, and TCA cycle intermediates (Fig. 2B). Due to a lack of cell division, cellular energy demands are reduced, facilitating a noncyclic mode for the TCA cycle that feeds isocitrate into the glyoxylate cycle. The induction of peroxisomal glyoxylate cycle enzymes (Figs 2C, 6; see Supplementary Material Fig. S1B in the online version at DOI: 10.1016/j.protis.2016.10.005) may participate in salvaging carbon produced during ß-oxidation and membrane remodeling that occurs during lipid accumulation in diatoms (Yang et al. 2013). This situation also permits decarboxylating and NADH producing steps of the TCA cycle to be bypassed, thereby limiting carbon loss and excess energy production. The entry of carbon into the mitochondria as malate permits cells to utilize carbon from non-glycolytic sources to provision the TCA cycle. This is accomplished through the action of a mitochondrial ME (Fig. 7; see Supplementary Material Fig. S1A in the online version at DOI: 10.1016/j.protis.2016.10.005), which supplies the pyruvate needed to replenish carbon skeletons via the partial TCA cycle (Hanning and Heldt 1993; Sweetlove et al. 2010; Turpin et al. 1988). In this way, our model for silicon starvation suggests that both mitochondrial photorespiratory processes and the peroxisomal glyoxylate cycle contribute carbon for biosynthetic purposes, unlike the recycling of photorespiratory carbon back into the Calvin-Benson cycle that has been described in plants.

The model presented in Figure 1 places our results in the context of earlier research and clarifies areas of ambiguity from previous models. Taken together, our data supports the idea that photorespiratory intermediates processed in the mitochondria are routed into amino acid metabolism rather than back into the chloroplast. In addition, we suggest that malate derived from the peroxisomal glyoxylate cycle may support the operation of the TCA cycle in certain conditions through the action of malic enzyme. This organizational scheme is a step towards understanding these unique carbon processing pathways in diatoms and serves to provide a framework for future investigation.

Methods

Cultivation: Cultivation for expression analysis was performed as previously described (Smith et al. 2016). Briefly, axenic 8L cultures of *Thalassiosira pseudonana* (CCMP1335) were grown in artificial seawater medium

(ESAW, http://www3.botany.ubc.ca/cccm/NEPCC/esaw.html) at 18 °C under continuous light (150 μmol m^{-2} s^{-1}) to a concentration of approximately 1×10^6 cells ml^{-1} and then harvested by centrifugation for 12 min at 3100 × *g* and placed in 8 L of silicic acid free medium in a polycarbonate bottle at a concentration of approximately 5×10^5 cells ml^{-1}. Cultures were stirred and bubbled with air under constant continuous light. Cultures were sampled at 0, 4, 8, 12, 18, and 24-hour time points following inoculation into silicic acid free medium.

Transcriptomics: RNA for transcriptomic analysis was extracted from *T. pseudonana* cells during a 24-hour time course of silicon limitation. Details of the experimental design, RNA sequencing, and processing are described in Smith et al. (2016). Briefly, cells were treated with cycloheximide (final concentration, 20 μg ml^{-1}), harvested by filtration, pelleted and stored at −80 °C prior to total RNA isolation (Hildebrand and Dahlin 2000). RNA was processed either for hybridization to an Affymetrix GeneChip whole genome tiling array or processed for Illumina-based RNA-Seq.

Bioinformatic analyses: The genome of *Thalassiosira pseudonana* (http://genome.jgi-psf.org/Thaps3/Thaps3.home.html) was searched for genes involved in photorespiratory metabolism and carbon concentration based on BLAST similarity with previously annotated sequences available from the Uniprot database. RNAseq reads generated during duplicate silicon starvation experiments (described in Smith et al. 2016) were mapped against the *T. pseudonana* genome version 3 using TopHat2 (Kim et al. 2013). Alignment files of uniquely mapped reads in BAM format were used to create index files using the Bam to Bai function available through the Galaxy platform (Blankenberg et al. 2010; Giardine et al. 2005; Goecks et al. 2010). The *T. pseudonana* genome, existing version 3 gene models, and the mapped reads were visualized together on the Integrated Genomics Viewer (Robinson et al. 2011; Thorvaldsdóttir et al. 2012) to assess the reliability of each gene model. A FASTA file of evaluated protein sequences used in this study is available in the supplemental material (see Supplementary Material Fig. S5 in the online version at DOI: 10.1016/j.protis.2016.10.005). When necessary, the N-terminus of the gene was extended to the first in-frame methionine to recover any excluded targeting information. Several bioinformatic tools were used to predict subcellular localization, including SignalP (Bendtsen et al. 2004; Nielsen and Krogh 1998), ChloroP (Emanuelsson et al. 1999), ASAFind (Gruber et al. 2015), MitoProt (Claros and Vincens 1996), and PTS1 Predictor (Neuberger et al. 2003a,b). In diatoms, nuclear encoded proteins targeted to the chloroplast have a bipartite targeting signal necessary for import across several membrane systems (Apt et al. 2002). Proteins predicted to have a signal peptide (SignalP) and a chloroplast transit-peptide (ChloroP), but not meeting the sequence requirements for stromal import (ASAFind) were predicted to have periplastid localization. An assessment of the conservation of targeting predictions from available diatom sequences was attempted, and in many cases supported our observations from *P. tricornutum* and *T. pseudonana*. However, the variability in accurate gene model predictions, especially at the gene model boundaries, precludes confidence in drawing rigorous conclusions.

Phylogenetic analyses: Homologous sequences were selected by BLAST search of the NCBI nr protein database (http://www.ncbi.nlm.nih.gov). Sequences from model organisms that had been manually annotated and reviewed (submitted to the UniprotKB/Swiss-Prot database) were included as points of reference. Diatom protein sequences were obtained from the JGI genome browser and (http://genome.jgi-psf.org/cgi-bin/searchGM?db=Thaps3). Sequences were aligned using MAFFT with default parameters (Katoh et al. 2002) and manually trimmed to conserved functional domains in Geneious version 8.1.2 (http://www.geneious.com) (Kearse et al. 2012). Maximum likelihood trees were constructed from the alignment in the MEGA 6 software (Tamura et al. 2013) using the best fit model (LG + G) with complete gap deletion and 1,000 bootstrap replicates. For the ME tree, sequences were trimmed to the previously characterized nicotinamide-binding region (Scrutton et al. 1990) and a maximum likelihood tree was constructed using PHYML (Guindon et al. 2010) with 1000 bootstrap replicates. We were not able to produce a meaningful alignment between the GK and GLYK proteins, and therefore decided it was not appropriate to include them in the same phylogenetic tree. In order to represent the idea that GK and GLYK sequences are nonhomologous proteins, we generated a hierarchically clustered heatmap from the percent identity matrix using the gplots package (Warnes 2015) in R (R Core Team 2015).

Vector construction for protein localization: Destination and expression vectors were constructed as described in Shrestha and Hildebrand (2015) using MultiSite Gateway Technology (Life Technologies). Briefly, the constitutive expression vector pTpfcpGFP (Poulsen et al. 2006) was modified by cloning a Gateway frame B cassette upstream of e*gfp* creating a destination vector. In cases where higher expression levels were required for *in vivo* localization, another destination vector was created utilizing the regulatory elements that drive constitutive expression of the ribosomal protein L41 (Shrestha et al. 2013) instead of FCP. Native expression constructs were created by incorporating 1000 bp of the 5′ flanking region and 500 bp of the 3′ flanking region of a gene of interest into the pMHL_71 vector to provide the native regulatory elements as described in Shrestha and Hildebrand (2015). Primers used to amplify genes and regulatory elements for construct synthesis are listed in Supplementary Material Table S2 in the online version at DOI: 10.1016/j.protis.2016.10.005. Each expression vector was co-transformed with pMHL_9 expressing the *nat1* gene under the control of the acetyl coenzyme A carboxylase promoter, which confers resistance to the antibiotic nourseothricin.

Diatom transformation: The protocol for diatom transformation was adapted from Poulsen et al. (2006). Axenic exponential-phase wild type *T. pseudonana* cells (1×10^8 cells total) were pelleted (3,000 × *g* for 10 minutes) and plated on artificial seawater medium (ASW) 1.2% agar plates. M-17 tungsten particles were coated with 5 ug of plasmid DNA for each construct with the $CaCl_2$-spermidine method as per manufacturer's instructions (BIORAD, 165-2267). Each plate was bombarded twice with 2–5 mg of coated tungsten beads using the Biolistic DS-1000/He particle delivery system with a 1350 psi rupture disk. Following bombardment, cells were immediately resuspended in 10 ml of ASW medium and incubated for 24 hours under constant illumination (150 uE m^{-2} s^{-1}). The following day, the cell density was determined and a range of concentrations (1×10^6–5×10^6 cells) were plated on ASW agar plates containing the antibiotic nourseothricin at a final concentration of 100 ug ml^{-1}. The plates were incubated at 18 °C in continuous light for approximately 2 weeks. Individual colonies from the plates were isolated and screened by PCR and imaging flow cytometry to confirm nuclear insertion of the plasmid and for GFP expression, respectively.

Fluorescence microscopy: Diatom transformants were imaged with a Zeiss Axio Observer Z1 inverted microscope equipped with an ApoTome and a Zeiss AxioCam MRm camera (Carl Zeiss Microimaging, Inc., Thornwood, NY, USA). Non-

fluorescent images were taken using differential interference contrast (DIC). The filter sets used for fluorescent imaging were as follows: chlorophyll (Zeiss #16: Excitation BP 485/20 nm, Dichromatic mirror FT 510 nm, Emission LP 515 nm), GFP (Zeiss #38HE Excitation BP 470/40 nm, Dichromatic mirror FT 495 nm, Emission BP 525/50 nm), and RFP (Zeiss #43HE Excitation BP 550/25 nm, Dichromatic mirror FT 570 nm, Emission BP 605/70 nm). Images were acquired with 63x/1.4 objective oil immersion Plan-Apochromat objective and processed using Axiovision 4.7.2 software.

Acknowledgements

This research was funded by Department of Energy Grants DE-SC0012556 and DE-EE0001222 and National Science Foundation grant CBET-0903712. Additional support was provided through a Department of Energy Office of Science Graduate Fellowship Program (DOE SCGF), made possible in part by the American Recovery and Reinvestment Act of 2009, administered by ORISE-ORAU under contract no. DE-AC05-06OR23100.

References

Allen A (2005) Defining the molecular basis for energy balance in marine diatom under fluctuating environmental conditions. J Phycol **41**:1073–1076

Allen AE, Vardi A, Bowler C (2006) An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. Curr Opin Plant Biol **9**:264–273

Allen AE, LaRoche J, Maheswari U, Loomer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C (2008) Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. Proc Natl Acad Sci USA **105**:10438–10443

Apt KE, Zaslavkaia L, Lippmeier JC, Lang M, Kilian O, Wetherbee R, Grossman AR, Kroth PG (2002) In vivo characterization of diatom multipartite plastid targeting signals. J Cell Sci **115**:4061–4069

Bailleul B, Berne N, Murik O, Petroutsos D, Prihoda J, Tanaka A, Villanova V, Bligny R, Flori S, Falconet D, Krieger-Liskay A, Santabarbara S, Rappaport F, Joliot P, Tirichine L, Falkowski P, Cardol P, Bowler C, Finazzi C (2015) Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. Nature **524**:366–369

Bartsch O, Hagemann M, Bauwe H (2008) Only plant-type (GLYK) glycerate kinases produce d-glycerate 3-phosphate. FEBS Lett **582**:3025–3028

Bauwe H, Hagemann M, Fernie AR (2010) Photorespiration: players, partners and origin. Trends Plant Sci **15**:330–336

Bauwe H, Hagemann M, Kern R, Timm S (2012) Photorespiration has a dual origin and manifold links to central metabolism. Curr Opin Plant Biol **15**:269–275

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3. 0. J Mol Biol **340**:783–795

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J (2010) Galaxy: A web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol **89**, 19.10:19.10.1–19.10.21

Claros MG, Vincens P (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem **241**:779–786

Clement R, Dimnet L, Maberly SC, Gontero B (2015) The nature of the CO_2-concentrating mechanisms in a marine diatom, *Thalassiosira pseudonana*. New Phytol **209**:1417–1427

Eisenhut M, Ruth W, Haimovich M, Bauwe H, Kaplan A, Hagemann M (2008) The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. Proc Natl Acad Sci USA **105**:17199–17204

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP, and related tools. Nat Protoc **2**:953–971

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Prot Sci **8**:978–984

Fabris M, Matthijs M, Rombauts S, Vyverman W, Goossens A, Baart GJE (2012) The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. Plant J **70**:1004–1014

Fauvart M, Braeken K, Daniels R, Vos K, Ndayizeye M, Noben J, Robben J, Vanderleyden J, Michiels J (2007) Identification of a novel glyoxylate reductase supports phylogeny-based enzymatic substrate specificity prediction. Biochim Biophys Acta Prot Proteom **1774**:1092–1098

Fernie AR, Tiessen A, Stitt M, Willmitzer L, Geigenberger P (2002) Altered metabolic fluxes result from shifts in metabolite levels in sucrose phosphorylase-expressing potato tubers. Plant Cell Environ **25**:1219–1232

Frederick SE, Gruber PJ, Tolbert NE (1973) The occurrence of glycolate dehydrogenase and glycolate oxidase in green plants. Plant Physiol **52**:318–323

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A (2005) Galaxy: A platform for interactive large-scale genome analysis. Genome Res **15**:1451–1455

Goecks J, Nekrutenko A, Taylor J (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol **11**:1–13

Gonzalez NH, Felsner G, Schramm FD, Klingl A, Maier UG, Bolte K (2011) A single peroxisomal targeting signal mediates matrix protein import in diatoms. PLoS ONE **6**:e25316

Gould SB, Sommer MS, Hadfi K, Zauner S, Kroth PG, Maier U (2006) Protein targeting into the complex plastid of cryptophytes. J Mol Evol **62**:674–681

Granum E, Raven JA, Leegood RC (2005) How do marine diatoms fix 10 billion tonnes of inorganic carbon per year? Can J Bot **83**:898–908

Gruber A, Vugrinec S, Hempel F, Gould SB, Maier U, Kroth PG (2007) Protein targeting into complex diatom plastids: functional characterization of a specific targeting motif. Plant Mol Biol **64**:519–530

Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T (2015) Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. Plant J **81**:519–528

Guindon S, Dufayard J, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3. 0. Syst Biol **59**:307–321

Hanning I, Heldt HW (1993) On the function of mitochondrial metabolism during photosynthesis in spinach (*Spinacia oleracea* L.) leaves: Partitioning between respiration and export of redox equivalents and precursors for nitrate assimilation products. Plant Physiol **103**:1147–1154

Hennon GMM, Ashworth J, Groussman RD, Berthiaume C, Morales RL, Baliga NS, Orellana MV, Armbrust EV (2015) Diatom acclimation to elevated CO_2 via cAMP signalling and coordinated gene expression. Nat Clim Change **5**:761–765

Hildebrand M, Dahlin K (2000) Nitrate transporter genes from the diatom *Cylindrotheca fusiformis* (Bacillariophyceae): mRNA levels controlled by nitrogen source and by the cell cycle. J Phycol **36**:702–713

Hockin N, Mock T, Mulholland F (2011) The response of diatom central carbon metabolism to nitrogen starvation is different to that of green algae and higher plants. Plant Physiol **158**:299–312

Huang A, Liu L, Yang C, Wang G (2015) *Phaeodactylum tricornutum* photorespiration takes part in glycerol metabolism and is important for nitrogen-limited response. Biotechnol Biofuels **8**:1–16

Igamberdiev AU, Lea PJ (2002) The role of peroxisomes in the integration of metabolism and evolutionary diversity of photosynthetic organisms. Phytochemistry **60**:651–674

Igamberdiev AU, Bykova NV, Lea PJ, Gardestrom P (2001) The role of photorespiration in redox and energy balance of photosynthetic plant cells: A study with a barley mutant deficient in glycine decarboxylase. Physiologia Plantarum **111**: 427–438

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res **30**:3059–3066

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28**:1647–1649

Kehrer D, Ahmed H, Brinkmann H, Siebers B (2007) Glycerate kinase of the hyperthermophilic archaeon *Thermoproteus tenax*: new insights into the phylogenetic distribution and physiological role of members of the three different glycerate kinase classes. BMC Genomics **8**:1–18

Kern R, Bauwe H, Hagemann M (2011) Evolution of enzymes involved in the photorespiratory 2-phosphoglycolate cycle from cyanobacteria via algae toward plants. Photosyn Res **109**:103–114

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14**:1–13

Kim J, Fabris M, Baart G, Kim MK, Goossens A, Vyverman W, Falkowski PG, Lun DS (2015) Flux balance analysis of primary metabolism in the diatom *Phaeodactylum tricornutum*. Plant J **85**:161–176

Kromer S (1995) Respiration during photosynthesis. Annu Rev Plant Physiol Plant Mol Biol **46**:45–70

Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Schnitzler Parker M, Stanley MS, Kaplan A, Caron L, Weber T, Maheswari U, Armbrust EV, Bowler C (2008) A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. PloS ONE **3**:e1426

Kunze M, Pracharoenwattana I, Smith SM, Hartigm A (2006) A central role for the peroxisomal membrane in glyoxylate cycle function. Biochim Biophys Acta Mol Cell Res **1763**:1441–1452

Kustka AB, Milligan AJ, Zheng H, New AM, Gates C, Bidle KD, Reinfelder JR (2014) Low CO_2 results in a rearrangement of carbon metabolism to support C4 photosynthetic carbon assimilation in *Thalassiosira pseudonana*. New Phytol **204**:507–520

Lancien M, Ferrario-Mery S, Roux V, Bismuth E, Masclaux C, Hirel B, Gadal P, Hodges M (1999) Simultaneous expression of NAD-dependent isocitrate dehydrogenase and other Krebs cycle genes after nitrate resupply to short-term nitrogen starved tobacco. Plant Physiol **120**:717–725

Liepman AH, Olsen LJ (2003) Alanine aminotransferase homologs catalyze the glutamate:glyoxylate aminotransferase reaction in peroxisomes of *Arabidopsis*. Plant Physiol **131**:215–227

Mittler R (2002) Oxidative stress, antioxidants and stress tolerance. Trends Plant Sci **7**:405–410

Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F (2003a) Motif refinement of the peroxisomal targeting signal 1 and evaluation of taxon-specific differences. J Mol Biol **328**:567–579

Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F (2003b) Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. J Mol Biol **328**:581–592

Nielsen H, Krogh A (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. Proc Int Conf Intell Sys Mol Biol **6**:122–130

Ohno N, Inoue T, Yamashiki R, Nakajima K, Kitahara Y, Ishibashi M, Matsuda Y (2012) CO_2-cAMP-responsive cis-elements targeted by a transcription factor with CREB/ATF-like basic zipper domain in the marine diatom *Phaeodactylum tricornutum*. Plant Physiol **158**:499–513

Parker M, Armbrust EV (2005) Synergistic effects of light, temperature, and nitrogen source on transcription of

genes for carbon and nitrogen metabolism in the centric diatom *Thalassiosira pseudonana* (Bacillariophyceae). J Phycol **41**:1142–1153

Parkes JA, Langer S, Hartig A, Baker A (2003) PTS1-independent targeting of isocitrate lyase to peroxisomes requires the PTS1 receptor Pex5p. Mol Membrane Biol **20**:61–69

Paul JS, Volcani BE (1975) Photorespiration in diatoms III. Glycolate:cytochrome *c* reductase in the diatom *Cylindrotheca fusiformis*. Plant Sci Lett **5**:281–285

Paul JS, Sullivan CW, Volcani BE (1975) Photorespiration in diatoms III. Mitochondrial glycolate dehydrogenase in *Cylindrotheca fusiformis* and *Nitzschia alba*. Arch Biochem Biophys **169**:152–159

Pick TR, Bräutigam A, Schulz MA, Obata T, Fernie AR, Weber APM (2013) PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. Proc Natl Acad Sci USA **110**:3185–3190

Piekarska K, Hardy G, Mol E, van den Burg J, Strijbis K, van Roermund C, van den Berg M, Distel B (2008) The activity of the glyoxylate cycle in peroxisomes of *Candida albicans* depends on a functional β-oxidation pathway: evidence for reduced metabolite transport across the peroxisomal membrane. Microbiology **154**:3061–3072

Plaxton WC, Podestá FE (2007) The functional organization and control of plant respiration. Critical Rev Plant Sci **25**:159–198

Poulsen N, Chesley PM, Kröger N (2006) Molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (Bacillariophyceae). J Phycol **42**:1059–1065

Prihoda J, Tanaka A, de Paula WBM, Allen JF, Tirichine L, Bowler C (2012) Chloroplast-mitochondria cross-talk in diatoms. J Exp Bot **63**:1543–1557

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria http://www.R-project.org/

Reinfelder JR (2011) Carbon concentrating mechanisms in eukaryotic marine phytoplankton. Annu Rev Mar Sci **3**:291–315

Reinfelder JR, Milligan AJ, Morel FMM (2004) The role of the C4 pathway in carbon accumulation and fixation in a marine diatom. Plant Physiol **135**:2106–2111

Reumann S, Weber APM (2006) Plant peroxisomes respire in the light: Some gaps of the photorespiratory C2 cycle have become filled – others remain. Biochim Biophys Acta Mol Cell Res **1763**:1496–1510

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol **29**:24–26

Rucktäschel R, Girzalsky W, Erdmann R (2011) Protein import machineries of peroxisomes. Biochim Biophys Acta Biomembranes **1808**:892–900

Saier MH, Wentzel DL, Feucht BU, Judice JJ (1975) A transport system for phosphoenolpyruvate, 2-phosphoglycerate, and 3-phosphoglycerate in *Salmonella typhimurium*. J Biol Chem **250**:5089–5096

Sarthou G, Timmermans KR, Blain S, Tréguer P (2005) Growth physiology and fate of diatoms in the ocean: a review. J Sea Res **53**:25–42

Schnitzler Parker M, Armbrust EV, Piovia-Scott J, Keil RG (2004) Induction of photorespiration by light in the centric diatom *Thalassiosira weissflogii* (Bacillariophyceae): Molecular characterization and physiological consequences. J Phycol **40**:557–567

Schwender J, Ohlrogge J, Shachar-Hill Y (2004) Understanding flux in plant metabolic networks. Curr Opin Plant Biol **7**:309–317

Schwender J, Shachar-Hill Y, Ohlrogge JB (2006) Mitochondrial metabolism in developing embryos of *Brassica napus*. J Biol Chem **281**:34040–34047

Scrutton NS, Berry A, Perham RN (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. Nature **343**:38–43

Shrestha RP, Hildebrand M (2015) Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses. Eukaryot Cell **14**:29–40

Shrestha RP, Haerizadeh F, Hildebrand M (2013) Molecular Genetic Manipulation of Microalgae: Principles and Applications. In Richmond Am, Hu X (eds) Handbook of Microalgal Culture: Applied Phycology and Biotechnology, 2nd edn. John Wiley & Sons, Ltd., Oxford UK, pp 146–167

Smith SR, Glé C, Abbriano RM, Traller JC, Davis A, Trentacoste E, Vernet M, Allen AE, Hildebrand M (2016) Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. New Phytol **210**:890–904

Smith SR, Abbriano RM, Hildebrand M (2012) Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. Algal Res **1**:2–16

Suzuki K, Iwamoto K, Yokoyama S, Ikawa T (1991) Glycolate-oxidizing enzymes in algae. J Phycol **27**:492–498

Sweetlove LJ, Beard KFM, Nunes-Nesi A, Fernie AR, Ratcliffe RG (2010) Not just a circle: flux modes in the plant TCA cycle. Trends Plant Sci **15**:462–470

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6. 0. Mol Biol Evol **30**:2725–2729

Tanaka R, Kikutani S, Mahardika A, Matsuda Y (2014) Localization of enzymes relating to C4 organic acid metabolisms in the marine diatom, *Thalassiosira pseudonana*. Photosyn Res **121**:251–263

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2012) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings Bioinformatics **14**:178–192

Turpin DH, Elrifi IR, Birch DG, Weger HG, Holmes JJ (1988) Interactions between photosynthesis, respiration, and nitrogen assimilation in microalgae. Can J Bot **66**:2083–2097

van den Bosch H, Schutgens RBH, Wanders RJA, Tager JM (1992) Biochemistry of peroxisomes. Annu Rev Biochem **61**:157–197

van der Klei IJ, Veenhuis M (2006) PTS1-independent sorting of peroxisomal matrix proteins by Pex5p. Biochim Biophys Acta Mol Cell Res **1763**:1794–1800

Vanlerberghe G (2013) Alternative oxidase: a mitochondrial respiratory pathway to maintain metabolic and signaling homeostasis during abiotic and biotic stress in plants. Int J Mol Sci **14**:6805–6847

Warnes GR (2015) gplots: various R programming tools for plotting data. R package version 2.17.0. http://CRAN.R-project.org/package=gplots

Wingler A, Lea PJ, Quick WP, Leegood RC (2000) Photorespiration: metabolic pathways and their role in stress protection. Philos Trans R Soc B Biol Sci **355**:1517–1529

Winkler U, Stabenau H (1992) Compartmentation of Peroxisomal Enzymes in the Diatom *Fragilaria*. In Stabenau H (ed) Phylogenetic Changes in Peroxisomes of Algae — Phylogeny of Plant Peroxisomes. University Press, Oldenburg, pp 130–139

Winkler U, Stabenau H (1995) Isolation and characterization of peroxisomes from diatoms. Planta **195**:403–407

Yang ZK, Yang WD, Liu JS, Lu SH, Guan Y, Li HY (2013) Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. Biotechnol Biofuels **6**:1–67

Young XK, McCarty RE (1993) Assay of proton-coupled glycolate and D-glycerate vesicles by stopped-flow fluorescence. Plant Physiol **101**:793–799

Yu ET, Zendejas FJ, Lane PD, Gaucher S, Simmons BA, Lane TW (2009) Triacylglycerol accumulation and profiling in the model diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (Bacillariophyceae) during starvation. J Appl Phycol **21**:669–681

Zheng Y, Quinn AH, Sriram G (2013) Experimental evidence and isotopomer analysis of mixotrophic glucose metabolism in the marine diatom *Phaeodactylum tricornutum*. Microb Cell Factories **12**:109

Chapter 2, in full, is a reprint of the material as it appears in Protist in 2016. Davis, A.; Abbriano, R.; Smith, S.R.; Hildebrand, M. The dissertation author was the co-primary investigator and author of this material.

CHAPTER 3:

Regulation of growth and carbon partitioning by phosphofructo-2-kinase in the marine diatom *Thalassiosira pseudonana*

3.1 ABSTRACT

A unique variant of 6-phosphofructo-2-kinase/fructose 2,6 bisphosphatase (PF2K/F2BP), a regulator of glycolytic flux previously uncharacterized in single-cell photosynthetic eukaryotes, was identified in diatom genomes and overexpressed in *T. pseudonana.* Overexpression affected carbon partitioning, resulting in higher levels of neutral lipids and proteins and less carbohydrate, showing that alteration of carbon partitioning between glycolysis and gluconeogenesis affects the metabolic fate of fixed carbon in diatoms. Higher glycolytic activity correlated with reduced growth rate and an extension of the cell cycle G1 phase. These effects differ from other unicellular eukaryotes and suggest a unique role for PFK2/F2BP in diatoms. We also show that redirection of metabolic resources can elicit lipid and protein accumulation in the absence of an environmental trigger, such as nutrient limitation.

3.2 SIGNIFICANCE STATEMENT

The regulation of carbon partitioning in microalgae is key to understand mechanisms involved in their high productivity, and regulatory mechanisms controlling carbon flux towards different metabolic fates are currently not well understood. We present the first *in vivo* analysis of PFK2/F2BP enzymes, key regulators of eukaryotic carbon metabolism, in a single-celled photosynthetic eukaryote. The characterization of a unique PFK2/F2BP protein indicates a prominent role in controlling carbon partitioning and affecting cell cycle progression in diatoms.

3.3 INTRODUCTION

A primary decision point in central carbon metabolism is the second bypass of glycolysis, where carbon is unidirectionally shunted towards either glycolysis or gluconeogenesis (Fig. 3-1). Enzymes that regulate this metabolic checkpoint are allosterically controlled by fructose-2,6-bisphosphate (F2,6P), which plays a key role in the control of carbohydrate metabolism in eukaryotes (Okar and Lange 1999) by affecting phosphofructokinase (PFK) and fructose-1,6-bisphosphatase (FBP) (Fig. 3-1). The reciprocal regulation of these pathways prevents futile cycling of glycolytic intermediates and exerts control over carbon storage, cellular energetics and cell cycle progression (Okar et al. 2001; Moncada et al. 2012; Ros and Schulze 2013). This allosteric regulator is phosphorylated and dephosphorylated by a bifunctional enzyme, phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFK2/F2BP) (Rider et al. 2004). The N-terminal kinase domain of the PFK2/F2BP protein phosphorylates fructose 6-phosphate (F6P) to produce F2,6P, while the C-terminal phosphatase domain performs the reverse reaction.

The modulation of carbon flux at the second bypass can affect carbon partitioning into biosynthetic pathways (including carbohydrate, protein, and lipid synthesis), with larger consequences for growth and productivity. Manipulation of F2,6P levels in model eukaryotic systems affects carbon partitioning and cell proliferation. In mammals, overexpression of PFK2 increased glycolytic flux into lipids (Wu et al. 2001; Duran et al. 2008), and elevated F2,6P levels may maintain high glycolytic rates characteristic of cancer cells (Atsumi et al. 2002; Yalcin et al. 2009; Ros and Schulze 2013). In plants, PFK2/F2BP coordinates the rate of sucrose synthesis in the cytosol with carbon fixation

and plastidic starch production (Draborg et al. 2001; Cseke et al. 2003; Nielsen et al. 2004). Overall the role of PFK2/F2BP in multicellular organisms relates to controlling the rate of glycolysis to support cell proliferation. The situation is more complicated in heterotrophic unicellular eukaryotes such as yeast, where changes in F2,6P had little phenotypic effect or resulted defective cell division (Bolesm et al. 1996; Fernández de Mattos et al. 2008).

Among the most highly productive organisms for converting fixed carbon to biomass are unicellular algae. Marine diatoms contribute ~40% of the total primary productivity in the modern oceans (Nelson et al. 1995), and thus are major players in the global carbon cycle. Diatoms are attractive candidates for use in biotechnology because they produce valuable carbon-based biopolymers such as lipid suitable for fuel production or other commercial applications (Hildebrand et al. 2012). In contrast to green algae, diatoms have a distinct evolutionary history involving a secondary endosymbiosis with a red alga (Falkowski 2004) and a distinctive organization of carbon metabolism (Kroth et al. 2008). This includes unique compartmentation of glycolysis/gluconeogenesis, pyruvate metabolism, and carbohydrate biosynthesis (Kroth et al. 2008; Smith et al. 2012). Diatoms store carbohydrates as chrysolaminarin (a soluble β-1,3-linked glucan) in a cytoplasmic vacuole (Fig. 3-1), instead of storing starch in the plastid. Diatom mitochondria contain the lower half of glycolysis for the production of energy and pyruvate (Kroth et al. 2008; Smith et al. 2012; Kim et al. 2015), an evolutionarily unique but conserved metabolic feature in diatoms (Fig. 3-1).

Nutrient limitation is commonly used to study carbon partitioning in diatoms, during which growth slows and biomolecules accumulate as trade-offs between

carbohydrate, lipid, and amino acid metabolism occur (Roessler 1988; Granum and Myklestad 2001; Hockin et al. 2012; Valenzuela et al. 2013; Yang et al. 2013; Ge et al. 2014; Levitan et al. 2015; Smith et al. 2016b; Longworth et al. 2016). In *Skeletonema costatum,* short-term nitrogen limitation mobilizes storage carbohydrate to produce amino acids in the dark (Granum and Myklestad 2001). In *Phaeodactylum tricornutum,* more severe nitrogen limitation triggers triacylglycerol (TAG) accumulation (Yu et al. 2009; Yang et al. 2013), derived in part from carbon repurposed from amino acids (Ge et al. 2014; Levitan et al. 2015). Silicon starvation stimulates TAG accumulation in most diatom species (Roessler 1988; Yu et al. 2009; Traller and Hildebrand 2013; Smith et al. 2016b). In *Cyclotella cryptica,* silicon starvation-induced lipid accumulation is due to both *de novo* lipid synthesis and redistribution of carbon from non-lipid compounds, such as carbohydrate, into lipids (Roessler 1988). Although the effects of nutrient limitation on growth and carbon metabolism are well documented, the mechanisms involved have not been elucidated. A potentially fruitful approach would be to genetically manipulate carbon flux independent of nutrient limitation, which would enable an independent evaluation of the effect of carbon flux on growth.

Despite the demonstrated importance of PFK2/F2BP in controlling carbon partitioning and productivity in other systems, microalgal PFK2/F2BP has not yet been characterized. Previous analysis suggested that diatoms possess two functionally distinct forms of PFK2/F2BP, with one form containing a critical amino acid substitution (Fig. 3-2) that suggests it functions solely as a kinase to produce fructose-2,6-bisphosphate and promote glycolysis (Smith et al. 2012). Using the marine diatom *Thalassiosira pseudonana*, we demonstrate for the first time in a microalgal system the effect of PFK2

controls the rate of glycolysis, with consequences for growth and productivity in terms of carbohydrate, protein, and lipid storage.

3.4 METHODS

3.4.1 Sequence analysis

T. pseudonana PFK2/F2BP sequences were obtained from the JGI genome browser (http://genome.jgi.-pdf.org/cgi-bin/searchGM?db=Thaps3) and gene models were manually adjusted based on RNAseq coverage (Fig. 3-S1). The models were used to search the NCBI database and a custom database containing diatom transcriptomes from the Marine Microbial Eukaryote Transcriptome Sequencing Project for PFK2/F2BP sequences. Truncated sequences containing only one of two conserved domains were excluded. Fifty-one amino acid sequences were manually curated and aligned using MAFFT with default parameters (Katoh et al. 2002) using Geneious (Kearse et al. 2012). MEGA5 was used to determine the best-fit substitution model (WAG+G) and to estimate the maximum likelihood tree using 1000 bootstrap replicates. FigTree software was used to generate the final graphical representation. Accession numbers sequences analyzed are included in Table S1.

3.4.2 Vector construction and diatom transformation

Destination and expression vectors were constructed using MultiSite Gateway Technology (Life Technologies) as described previously (Shrestha and Hildebrand 2015). The PFK2-2 overexpression construct was generated by cloning the manually adjusted Thaps3_109629 open reading frame upstream of e*gfp* in a modified version of the constitutive expression vector pTpfcpGFP (Fig. 3-S2) (Poulsen et al. 2006). Expression

vectors were cotransformed with pMHL_9, a vector expressing the *nat1* gene under the control of the acetyl coenzyme A carboxylase promoter. Antisense constructs were designed by amplifying 500 bp regions in antisense orientation to the Thaps3_14563 and Thaps3_109629 transcript sequences, and were assembled into individual Gateway destination vectors under the control of the native *T. pseudonana* fcp promoter and terminator sequences (Fig. S2A and B). Vectors were introduced into *T. pseudonana* cells using the Biolistic DS-1000/He particle delivery system as described (Davis et al. 2016).

3.4.3 Cultivation and sampling

Axenic cultures of T. pseudonana (CCMP 1335) were grown in ASW media (http://www3.botany.ubc.ca/cccm/NEPCC/esaw.html) at 20-22^{o}C under constant illumination at 150 μ mol m^{2} sec^{-1}. Initial screening of 50 mL cultures was done in 125 mL flasks grown on a rotary shaker. Larger scale experiments for biochemical characterization were done in 2L volume with magnetic stirring and aeration with sterile air. For synchrony experiments, exponential phase cultures were harvested by centrifugation (4,000xg for 7 minutes), rinsed once in silicon-free ASW medium, and resuspended in silicon-free medium at a density of approximately 1e^{6} cells ml^{-1}. Cell counts were performed using a Muse® Cell Analyzer (Millipore Corp., Billerica MA, USA).

3.4.4 Fluorescence microscopy

Diatom transformants expressing GFP were imaged with a Zeiss Axio Observer Z1 inverted microscope equipped with an ApoTome and a Zeiss AxioCam MRm camera (Carl Zeiss Microimaging, Inc., Thornwood, NY, USA) as described in (Davis et al. 2017).

PFK activity assay: PFK activity was quantified using the Phosphofructokinase Activity Colorimetric Assay Kit (Sigma-Aldrich, MAK093). Protein extracts from $2e^6$ *T. pseudonana* cells were diluted 2X in PFK Assay buffer and added to 50 µL Reaction Mix in a 96 well plate. All samples were run in technical duplicates alongside sample blanks to account for a NADH background signal. Known concentrations of NADH (0 (blank), 2, 4, 6, 8, and 10 nmole/well) were used as a standard curve. Measurements at 450 nm were taken using a SPECTRAmax M2 microplate reader (Molecular Devices, Sunnyvale, CA) and used to calculate PFK activity per manufacturer's instructions.

3.4.5 Carbohydrate assay and aniline blue staining

Quantification of total β-1,3-glucan in diatoms (Granum and Myklestad 2002) was performed on frozen pellets containing a total of 2-$5e^8$ *T. pseudonana* cells. Sample absorbance was measured at 485 nm and compared to a series of glucose standards. For aniline blue staining, cell pellets were resuspended in 0.1 M potassium phosphate buffer (pH 7.4). Cells were stained for 10 minutes in the dark with 1 mg • ml^{-1} aniline blue (Waterkeyn and Bienfait 2011). Stained samples were run on an ImageStream X (Millipore Corp. Billerica MA, USA) imaging flow cytometer using the 405 nm laser at 30 mW and a chlorophyll blocking filter. 10,000 cells per sample were analyzed using Amnis IDEAS™ software.

3.4.6 Protein assay

Protein concentrations were measured using the Biorad DC™ Protein Assay, based on the Lowry method for protein quantification. 5-10 mL of *T. pseudonana* culture (1-$3e^7$ cells) were harvested and stored at -20°C until analysis. Proteins were extracted and measured per manufacturer's instructions and measured in duplicate on a 96 well

plate reader at 750 nm. A series of BSA dilutions were used to generate a standard curve and calculate sample protein concentrations.

3.4.7 Neutral lipid and FAME

Neutral lipids were measured using the lipophilic fluorescent dyes BODIPY (493/503, Molecular Probes) and Nile Red (Cooksey et al. 1987). Nile Red was used instead of BODIPY in overexpression lines to avoid overlap with GFP fluorescence. Frozen cell pellets were resuspended in 0.5 mL 0.45 M NaCl and stained with either 1.3 µL of 1 mg/ml BODIPY stock or 3.1 µL of 250 mg/L Nile Red stock for 30 minutes in the dark. Stained samples were run on an ImageStream X (Millipore Corp. Billerica MA, USA) imaging flow cytometer excited at 488 nm using 0.6 and 1.0 neutral density filters and a chlorophyll blocking filter. 10,000 cells per sample were analyzed using Amnis IDEAS^{TM} software. FAME samples were collected and analyzed as described in (Cook and Hildebrand 2016).

3.4.8 Chlorophyll

Unstained cell samples were run on an ImageStreamX imaging flow cytometer excited at 488 nm at 50mW and using 0.6 and 1.0 neutral density filters. 10,000 individual cells per sample were analyzed using the Amnis IDEAS^{TM} software. For total extracted chlorophyll, 5-10 mL of *T. pseudonana* culture was filtered onto duplicate 25 mm glass fiber filters and immediately frozen at -80^{o}C. Filters were extracted in 100% methanol for 24 hours in the dark and extracted chlorophyll was measured at 420 nm excitation/670 nm emission in a Turner 10-AU fluorometer.

3.4.9 Photosynthetic parameters

Fv/Fm was measured on a WALZ WATER-PAM fluorometer. Samples were normalized to $5e^5$ cells ml^{-1} to limit shading within the cuvette, and allowed to dark acclimate for 10 minutes prior to measurement. The fluorometer was calibrated using ASW media as a blank.

Photosynthetic activity was calculated by oxygen evolution rates measured on an ALGInstruments environmental control chamber fitted with a YSI Clark-type electrode. Lighting in the chamber was calibrate with a 4-pi submersible spherical micro quantum sensor (WALZ). The chamber temperature was set to 20˚C and the electrode was calibrated in ASW media, a high baseline was recorded with unconditioned media while a zero baseline was recorded during bubbling with pure nitrogen gas. Cell samples were normalized to $5e^5$ cells/ml and dark acclimated for at least 10 minutes prior to measurement. Oxygen evolution was recorded under a light intensity of 1000 µmol photons m^{-2} sec^{-1}. Maximum oxygen evolution rates were calculated as the sum of the oxygen evolution rate during illumination and the respiration rate following illumination, and were normalized per cell.

3.4.10 Cell cycle

Cultures were grown to $1x10^6$ cells ml^{-1}, harvested at 4,000xg, rinsed once in silicon-free ASW medium, and resuspended at the same cell density in 50 ml of silicon-free ASW. After 24 hours, silicon was added back to the medium (200 uM final concentration). Samples (1 ml) were pelleted, resuspended in 1 ml 100% ice cold methanol, and stored at 20^{o}C. Samples were prepared by pelleting to remove methanol and rinsing 3X in TE buffer. Pellets were resuspended in 1 ml of TE and treated with RNase A (0.3 mg ml^{-1}) at 37^{o}C for 40 minutes. RNase treated cells were stained with

SYBR Green (10 ul of 100X SYBR Green in DMSO per 1 ml sample) for at least ten minutes on ice in the dark. Samples were run on a BD Influx flow cytometer with 488 nm excitation and SYBR Green fluorescence was monitored at 530 nm. SYBR Green fluorescence was used to estimate DNA content using FCS Express software.

3.5 RESULTS

3.5.1 Sequence analysis

Multiple PFK2/F2BP homologs are present in organisms descended from the red algal secondary endosymbiosis and all members of the stramenopile lineage have at least two homologs (Fig. 3-3). Other photosynthetic eukaryotes in Rhodophyta, Chlorophyta, and Viridiplantae possess only one PFK2/F2BP. PFK2/F2BP sequences belong to one of two major clades: one representing most major eukaryotic groups (Group 1; PFK2-1), and another smaller clade exclusive to organisms descended from a red algal endosymbiotic event (Group 2; PFK2-2). Group 1 and 2 sequences were highly divergent; the Group 1 *T. pseudonana* sequence (Thaps3_14563) is more similar to human PFK2/F2BP (36% identity, calculated by ClustalW2) than to the Group 2 *T. pseudonana* PFK2/F2BP (Thaps3_109629; 29% identity). The extant red algal species did not possess a Group 2 PFK2/F2BP, suggesting these genes may have originated in the eukaryotic host involved in the secondary endosymbiosis.

3.5.2 Antisense knockdown of PFK2/F2BPs

To evaluate the influence of PFK2-1 and PFK2-2 on carbon storage and growth, we applied an antisense knockdown approach. Six clonal transformant lines for each knockdown indicated no differences compared with wild-type in growth rate or

maximum cell density (Fig. 3-4A and C). Knockdown generated minimal differences in exponential phase neutral lipid content (< 25%) with no consistent trends in all transformant lines (Fig. 3-4B and D). Aniline blue fluorescence, as a proxy for β-1,3-glucan content (Waterkeyn and Bienfait 2011), was consistently higher (1.5X) in PFK2-2 knockdown lines than wild-type (Fig. 3-4D), suggesting a relationship between PFK2-2 transcript levels and intracellular carbohydrate content. We thus focused on further characterization on PFK2-2.

3.5.3 Effect of PFK2-2 overexpression

To further evaluate the effect of manipulation, we overexpressed PFK2-2 (Thaps3_109629) fused to GFP controlled by the fcp promoter (Fig. 3-5; Fig. 3-S2C). Eight lines incorporated the overexpression construct, but only two (109fcp1 and 109fcp2) resulted in visible GFP fluorescence, suggesting a suppression of overexpression in the other lines (Fig. 3-6A). Lines 109fcp1 and 109fcp2 had lower growth rates than wild-type or the lines with little or no GFP expression (Fig. 3-6A). To test whether the reduction in growth rate was due solely to overexpression of GFP, we generated overexpression lines containing only eGFP driven by the fcp promoter (Fig. 3-S2D), which were visually positive for GFP (Fig. 3-S3) but had similar growth rates to wild-type (Fig. 3-6B). Doubling times in 2L cultures for wild-type, 109fcp1, and 109fcp2 were 7.0 hours, 10.8 hours, and 9.7 hours, respectfully. Overexpression lines eventually reached a maximum cell density comparable to wild-type (4-6e^6 cells mL^{-1}) transitioning into stationary phase (Fig. 3-7A-C).

We monitored exponential phase phosphofructokinase (PFK) activity to test the ability of PFK2-2 overexpression to stimulate glycolysis. PFK catalyzes a rate-limiting

step of glycolysis, is a known target of F2,6P, and its activity relates to glycolytic flux (TeSlaa and Teitell 2014). The overexpression lines had 2.0-3.9 times higher PFK activity than wild-type (Fig. 3-8) and growth rate and PFK activity was significantly inversely correlated (p-value = $9e^{-6}$).

We characterized the effect of overexpression of PFK2-2 on biochemical composition during exponential phase ($1e^6$ cells mL^{-1}) and stationary phase ($4\text{-}5e^6$ cells mL^{-1}). In both exponential and stationary phase, total carbohydrate per cell was 1.5 to 1.7-fold lower in the overexpression lines relative to wild-type (Fig. 3-9A). Higher total protein was present in both overexpression lines (Fig. 3-9B), which was more pronounced during exponential phase (1.6-fold higher) than stationary phase (1.2-fold higher). Average neutral lipid content in all cultures increased 1.5-fold in the transition from exponential to stationary phase, but transformant lines had 2.5-fold more during exponential phase and 2-fold higher more in stationary phase than wild-type (Fig. 3-9C). FAME content (which includes FAMEs derived from membrane lipids) was 1.3-fold higher in overexpression lines relative to wild-type in both exponential and stationary phase (Fig. 3-9D, Supplemental File 3-1).

In some microalgae, carbohydrate biosynthesis functions as an energetic sink for ATP and NADPH generated through photosynthesis (Deschamps et al. 2008), therefore, we evaluated whether overexpression of PFK2-2 affected photosynthetic efficiency. No significant differences in extracted or *in vivo* chlorophyll fluorescence were found between wild-type and overexpression lines (Fig. 3-9E, Table 3-1). Quantum efficiency (F_v/F_m) and oxygen evolution were not significantly different among wild type

and transformant cultures (Table 3-1), suggesting minimal effects of PFK2-2 overexpression on photosynthetic processes in our growth conditions.

We evaluated the effect of PFK2-2 overexpression on cell cycle dynamics using silicon-starvation synchronized cultures (Hildebrand et al. 2007). DNA content was monitored following silicon readdition as the cells progressed through the cell cycle. As previously documented (Hildebrand et al. 2007), the majority of the wild-type culture arrested in G1 phase, and upon silicon replenishment, S phase occurred at 2-3 hours and G2+M phase at 3-5 hours (Fig. 3-10A). The bulk of the population re-entered G1 by 6 hours and remained in G1 for 5 more hours before transitioning back to S phase (Fig 3-10A). For 109fcp1, a lower proportion of the population arrested in G1 and a more rapid transition into S and G2+M occurred (Fig. 3-10B), suggesting arrest at the G1/S phase boundary. A small proportion of the population transitioned into S phase by 10 hours, but the majority of cells remained in G1 throughout the experiment (Fig. 3-10B). The 109fcp2 line poorly synchronized, and most cells remained in G1 throughout the experiment with minor S phase peaks at 5 and 10 hours (Fig. 3-10C), suggesting multiple arrest points resulting in sub-populations of cells progressing through the cell cycle at different times. Both sets of results suggest that PFK2-2 overexpression delayed cell cycle progression via a prolonged G1 phase.

3.6 DISCUSSION

3.6.1 Diatom PFK2/F2BP proteins are functionally distinct

The low level of amino acid conservation among the diatom PFK2/F2BPs (Fig. 3-3), a critical substitution in the PFK2-2 phosphatase domain (Fig. 3-2), and phenotypic

differences in knockdown experiments (Fig. 3-4) all support distinct functional roles for diatom PFK2/F2BPs. For PFK2-1, knockdown experiments do not support a role for this enzyme in the control of growth, glycolytic flux, or carbon partitioning. These results are similar to functional studies in *Saccharomyces cerevisiae* that demonstrated that yeast PFK2/F2BPs did not play an important role in carbon flux regulation, but may be involved in maintaining metabolic homeostasis (Müller et al. 1997). In contrast, our results indicate that PFK2-2 plays an important role in the regulation of carbon flux in diatoms.

3.6.2 PFK2-2 regulates carbon partitioning in T. pseudonana

Carbohydrate content was uniformly higher in six PFK2-2 knockdown lines relative to wild-type (Fig. 3-4), suggesting that PFK2-2 knockdown increased the availability of carbon precursors for chrysolaminarin biosynthesis. Overexpression lines displayed 3-fold higher PFK activity relative to wild-type, which was accompanied by reduced carbohydrate content, higher neutral lipid content, and higher protein content in both exponential and stationary phase (Fig. 3-9). Thus, manipulation of this upstream step of carbon partitioning can affect both carbohydrate storage and the downstream accumulation of protein and lipid storage pools.

The metabolic trade-offs between carbohydrates and lipids under nutrient limitation has been previously investigated in diatoms (Roessler 1988). We demonstrate that partitioning carbon flux away from carbohydrate biosynthesis enables increased neutral lipid and protein accumulation, even during growth. Experiments in *Skeletonema costatum* show that carbohydrates serve as a source for amino acid and protein biosynthesis during the dark or under carbon limitation conditions (Granum and

Myklestad 2001). Our data further show that this redistribution occurs during nutrient replete, illuminated conditions if flux into storage carbohydrates is restricted.

3.6.3 Physiological and metabolic implications of cytosolic carbohydrate biosynthesis

A major difference in central carbon metabolism in diatoms and green algae is the compartmentation of carbohydrate biosynthesis (Hildebrand et al. 2013). In model green algae, starch is produced and degraded in the chloroplast, requiring a high degree of temporal regulation to prevent competing glycolytic and gluconeogenic processes (Deschamps et al. 2008). As a consequence, starch biosynthesis is tied to photosynthetic processes and serves as a sink for photosynthetically-generated ATP and NADPH (Deschamps et al. 2008). In *Chlamydomonas reinhardtii*, the starchless *sta6* mutant has reduced growth and photosynthetic activity (Li et al. 2010; Work et al. 2010; Krishnan et al. 2015) due to allosteric down-regulation of carbon fixation, reduced NADPH re-oxidation, and consequently, a slowing of photosynthetic electron transport (Krishnan et al. 2015). Diatoms export carbon from the plastid to synthesize carbohydrates, which may avoid major metabolic feedbacks on photosynthesis. In contrast to starchless mutants, reduced carbohydrate content in PFK2-2 overexpression lines did not negatively impact quantum efficiency (F_v/F_m) or photosynthetic oxygen evolution, suggesting that any potential imbalance in $NADP^+$/NADPH ratio was contained in the cytosolic compartment. Build-up of precursor molecules, rather than increases in sugar/starch, lipid, and protein was observed in the *sta6* mutant, while we observed increased partitioning of carbon into both proteins and lipids in the diatom.

3.6.4 Metabolic control of growth and cell cycle progression

Cell cycle progression is controlled by complex regulatory networks that

coordinately regulate growth and metabolism in response to environmental cues. In diatoms, both light and external nutrient supply are important factors (Huysman et al. 2010). In particular, the G_1 phase has checkpoints sensitive to light, nitrogen, and silicon availability (Olson and Chisholm 1983; Olson et al. 1986; Vaulot et al. 1987; Brzezinski et al. 1990; Hildebrand et al. 2007). The reduced growth rate in PFK2-2 overexpression lines demonstrated for the first time that intracellular carbon partitioning can influence cell cycle progression in *T. pseudonana,* even in the absence of environmental forcing. This implies that changes in carbon flux through cytosolic glycolysis can serve as a signal to affect cell cycle processes in diatoms.

The negative relationship between PFK activity and growth rate (Fig. 3-8) was unexpected, as high rates of glycolytic flux are associated with cell proliferation in plants and mammals (Vander Heiden et al. 2009), and increased F2,6P activates glycolysis (Atsumi et al. 2002; Cordero-Espinoza and Hagen 2013), while lower F2,6P delays cell cycle progression (Perez et al. 2000). The relationship between glycolytic flux and growth in yeast is less consistent; in *Saccharomyces cerevisiae*, changes in F2,6P have modest effects on growth and cellular metabolism (Bolesm et al. 1996; Müller et al. 1997), but PFK2 overexpression in *S. pombe* caused cell cycle delay related due to impaired cytokinesis (Fernández de Mattos et al. 2008). Our cell cycle data (Fig. 3-10) suggest that a consequence of increased glycolytic flux is the elongation of G1. In other unicellular eukaryotes, biosynthetic reactions necessary for cell duplication occur during the G1 phase, and adequate accumulation of metabolites is intimately linked to G1/S transition (Moncada et al. 2012). Also, in yeast the G1/S transition is coordinated tightly with carbon status (Ewald et al. 2016; Zhao et al. 2016).

One explanation for the inverse relationship between growth and glycolysis in diatoms (Fig. 3-8) may be their unique arrangement of central carbon metabolism. Energetic crosstalk occurs between diatom plastids and mitochondria (Bailleul et al. 2015) and mitochondrial glycolysis is likely the major pathway used for energy production and to provision anabolic processes during growth and division in the light phase (Smith et al. 2016a). This differs from green algae, where starch stored in the plastid is broken down and used to fuel cell division (Vitova et al. 2015). The prominence of mitochondrial glycolysis in diatoms implies that the cytosolic pathway serves other functions, such as supplying carbon for carbohydrate synthesis and the oxidative pentose phosphate pathway (OPP) and repurposing carbon skeletons from carbohydrate breakdown. The reduced growth rate in response to PFK2-2 overexpression could result from a limit on the amount of glucose-6-phosphate precursors available for entry into the OPP, which produces precursors for nucleotide biosynthesis (Fig. 3-11). Both the OPP and nucleotide biosynthesis occur in the cytosol in diatoms (Kroth et al. 2008; Ast et al. 2009). Impaired nucleotide biosynthesis would have a negative impact on transcriptional waves accompany that accompany cell cycle transitions in eukaryotes (Bertoli et al. 2013), extending the time necessary to progress through the G_1/S boundary.

Our results demonstrate a fundamental relationship between growth and carbon metabolism in diatoms and supports a critical role for PFK2-2 in these processes. Nutrient limitation inhibits growth and causes significant shifts in carbon partitioning (Roessler 1988; Granum and Myklestad 2001; Yu et al. 2009; Hockin et al. 2012; Traller and Hildebrand 2013; Valenzuela et al. 2013; Yang et al. 2013; Ge et al. 2014; Levitan et al. 2015; Smith et al. 2016b; Longworth et al. 2016), but it was not known that changes in

carbon partitioning could feedback to affect growth. Elucidating the mechanisms by which this occurs in diatoms will clarify whether it will be possible to engineer carbon flux to enrich for specific biomolecules without negatively impacting biomass productivity.

Chapter 3, in part, is included in Algal Research in 2012. Smith, S., Abbriano, R., and Hildebrand, M. The dissertation author is the primary investigator of the data reported in Chapter 3; Sarah Smith is the primary investigator of the published manuscript. Chapter 3, in part, has been submitted for publication of the material as it may appear in Proceedings of the National Academy of Sciences of the United States of America in 2017. Abbriano, R.; Vardar, N.; Yee, D.; Hildebrand, M. The dissertation author was the primary investigator and author of this material.

3.7 ACKNOWLEDGEMENTS

We thank Roshan Shrestha, Joris Beld, and Jeffrey Mindrebo for their advice and technical assistance. This research is supported by AFOSR Grant No. FA9550-08-1-0718 as well as the Department of Energy Office of Science Graduate Fellowship Program (DOE SCGF), made possible in part by the American Recovery and Reinvestment Act of 2009, administered by ORISE-ORAU under contract no. DE-AC05-06OR23100. Research reported in this publication was also supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number T32GM067550. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Research was also supported by an international research fellowship from TUBITAK.

3.8 REFERENCES

Ast M, Gruber A, Schmitz-Esser S, Neuhaus HE, Kroth PG, Horn M, Haferkamp I (2009) Diatom plastids depend on nucleotide import from the cytosol. Proceedings of the National Academy of Sciences 106:3621–3626.

Atsumi T, Chesney J, Metz C, Leng L, Donnelly S, Makita Z, Mitchell R, Bucala R (2002) High expression of inducible 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (iPFK-2; PFKFB3) in human cancers. Cancer Research 62: 5881-5887.

Bailleul B, Berne N, Murik O, Petroutsos D, Prihoda J, Tanaka A, Villanova V, Bligny R, Flori S, Falconet D, Krieger-Liszkay A, Santabarbara S, Rappaport F, Joliot P, Tirichine L, Falkowski PG, Cardol P, Bowler C, Finazzi G (2015) Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. Nature 54:366–369.

Bertoli C, Skotheim JM, de Bruin RAM (2013) Control of cell cycle transcription during G1 and S phases. Nature Reviews Molecular Cell Biology 14:518–528.

Bolesm E, Göhlmann HWH, Zimmermann FK (1996) Cloning of a second gene encoding 6-phosphofructo-2-kinase in yeast, and characterization of mutant strains without fructose-2,6-bisphosphate. Mol Microbiol 20:65–76.

Brzezinski MA, Olson RJ, Chisholm SW (1990) Silicon availability and cell cycle progression in marine diatoms. Mar Ecol Prog Ser 67:83–96.

Cook O, Hildebrand M (2016) Enhancing LC-PUFA production in *Thalassiosira pseudonana* by overexpressing the endogenous fatty acid elongase genes. J Appl Phycol 28:897–905.

Cooksey KE, Guckert JB, Williams SA, Callis PR (1987) Fluorometric determination of the neutral lipid content of microalgal cells using Nile Red. Journal of Microbiological Methods 6:333–345.

Cordero-Espinoza L, Hagen T (2013) Increased concentrations of fructose 2,6-bisphosphate contribute to the Warburg Effect in phosphatase and tensin homolog (PTEN)-deficient cells. J Biol Chem 288:36020–36028.

Cseke C, Balogh A, Wong J, Buchanan B, Stitt M, Herzog B, Heldt H (2003) Fructose 2,6-bisphosphate - a regulator of carbon processing in leaves. Trends in Biochemical Sciences 9:533–535.

Davis A, Abbriano R, Smith SR, Hildebrand M (2017) Clarification of Photorespiratory Processes and the Role of Malic Enzyme in Diatoms. Protist 168:134–153.

Deschamps P, Haferkamp I, d'Hulst C, Neuhaus HE, Ball SG (2008) The relocation of starch metabolism to chloroplasts: when, why and how. Trends in Plant Science 13:574–582.

Draborg H, Villadsen D, Nielsen TH (2001) Transgenic *Arabidopsis* plants with decreased activity of frustose-6-phosphate,2-kinase/fructose-2,6-bisphosphatase have altered carbon partitioning. Plant Physiol 126:750–758.

Duran J, Navarro-Sabatè A, Pujol A, Perales JC, Manzano À, Obach M, Gómez M, Bartrons R (2008) Overexpression of ubiquitous 6-phosphofructo-2-kinase in the liver of transgenic mice results in weight gain. Biochemical and Biophysical Research Communications 365:291–297.

Ewald JC, Kuehne A, Zamboni N, Skotheim JM (2016) The yeast cyclin-dependent kinase routes carbon fluxes to fuel cell cycle progression. Mol Cell 61:532-545.

Falkowski PG (2004) The Evolution of Modern Eukaryotic Phytoplankton. Science 305:354–360.

Fernández de Mattos S, Alemany V, Aligué R, Tauler A (2008) Increase in Fru-2,6-P2 levels results in altered cell division in Schizosaccharomyces pombe. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1783:144–152.

Ge F, Huang W, Chen Z, Zhang C, Xiong Q, Bowler C, Yang J, Xu J, Hu H (2014) Methylcrotonyl-CoA carboxylase regulates triacylglycerol accumulation in the model diatom *Phaeodactylum tricornutum*. Plant Cell 26:1681–1697.

Granum E, Myklestad S (2002) A simple combined method for determination of beta-1,3-glucan and cell wall polysaccharides in diatoms. Hydrobiologia 477:155–161.

Granum E, Myklestad SM (2001) Mobilization of beta-1,3-glucan and biosynthesis of amino acids induced by NH_4^+ addition to N-limited cells of the marine diatom *Skeletonema costatum* (Bacillariophyceae). Journal of Phycology 37:772–782.

Hildebrand M, Abbriano RM, Polle J (2013) Metabolic and cellular organization in evolutionarily diverse microalgae as related to biofuels production. Current Opinion in Chemical Biology 16:1–9.

Hildebrand M, Davis AK, Smith SR, Traller JC (2012) The place of diatoms in the biofuels industry. Biofuels 3:221–240.

Hildebrand M, Frigeri LG, Davis AK (2007) Synchronized growth of *Thalassiosira pseudonana* (Bacillariophyceae) provides novel insights into cell wall synthesis processes in relation to the cell cycle. Journal of Phycology 43:730–740.

Hockin NL, Mock T, Mulholland F, Kopriva S, Malin G (2012) The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae

and higher plants. Plant Physiol 158:299–312.

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

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30:3059–3066.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Kim J, Fabris M, Baart G, Kim MK, Goossens A, Vyverman W, Falkowski PG, Lun DS (2015) Flux balance analysis of primary metabolism in the diatom *Phaeodactylum tricornutum*. The Plant Journal 85:161–176.

Krishnan A, Kumaraswamy GK, Vinyard DJ, Gu H, Ananyev G, Posewitz MC, Dismukes GC (2015) Metabolic and photosynthetic consequences of blocking starch biosynthesis in the green alga *Chlamydomonas reinhardtii* sta6 mutant. The Plant Journal 81:947–960.

Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, Stanley MS, Kaplan A, Caron L, Weber T, Maheswari U, Armbrust EV, Bowler C (2008) A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. Plos One 3:e1426.

Levitan O, Dinamarca J, Zelzion E, Lun DS, Guerra LT, Kim MK, Kim J, van Mooy BAS, Bhattacharya D, Falkowski PG (2015) Remodeling of intermediate metabolism in the diatom *Phaeodactylum tricornutum* under nitrogen stress. P Natl Acad Sci Usa 112:412–417.

Li Y, Han D, Hu G, Sommerfeld M, Hu Q (2010) Inhibition of starch synthesis results in overproduction of lipids in *Chlamydomonas reinhardtii*. Biotechnol Bioeng 107:258–268.

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

Moncada S, Higgs EA, Colombo SL (2012) Fulfilling the metabolic requirements for cell proliferation. Biochem J 446:1–7.

Müller S, Zimmermann FK, Boles E (1997) Mutant studies of phosphofructo-2-kinases

do not reveal an essential role of fructose-2, 6-bisphosphate in the regulation of carbon fluxes in yeast cells. Microbiology 143:3055–3061.

Nelson DM, Tréguer P, Brzezinski MA, Leynaert A, Quéguiner B (1995) Production and dissolution of biogenic silica in the ocean - revised global estimates, comparison with regional data and relationship to biogenic sedimentation. Global Biogeochem Cy 9:359–372.

Nielsen T, Rung J, Villadsen D (2004) Fructose-2,6-bisphosphate: a traffic signal in plant metabolism. Trends in Plant Science 9:556–563.

Okar D, Lange A (1999) Fructose-2,6-bisphosphate and control of carbohydrate metabolism in eukaryotes. Biofactors 10:1–14.

Okar DA, Lange AJ, Manzano À, Navarro-Sabatè A, Riera L, Bartrons R (2001) PFK-2/FBPase-2: maker and breaker of the essential biofactor fructose-2,6-bisphosphate. Trends in Biochemical Sciences 26:30–35.

Olson RJ, Chisholm SW (1983) Effects of photocycles and periodic ammonium supply on 3 marine phytoplankton species: cell division patterns. Journal of Phycology 19:522–528.

Olson RJ, Vaulot D, Chisholm SW (1986) Effects of environmental stresses on the cell cycle of two marine phytoplankton species. Plant Physiol 80:918–925.

Perez JX, Roig T, Manzano A, Dalmau M, Boada J, Ventura F, Rosa JL, Bermudez J, Bartrons R (2000) Overexpression of fructose 2,6-bisphosphatase decreases glycolysis and delays cell cycle progression. American Journal of Physiology-Cell Physiology 279:C1359–C1365.

Poulsen N, Chesley PM, Kroger N (2006) Molecular genetic manipulation of the diatom Thalassiosira pseudonana (Bacillariophyceae). Journal of Phycology 42:1059–1065.

Rider MH, Bertrand L, Vertommen D, Michels PA, Rousseau GG, Hue L (2004) 6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase: head-to-head with a bifunctional enzyme that controls glycolysis. Biochem J 381:561.

Roessler P (1988) Effects of silicon deficiency On lipid composition and metabolism in the diatom *Cyclotella cryptica*. J Phycol 24:394-400.

Ros S, Schulze A (2013) Balancing glycolytic flux: the role of 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatases in cancer metabolism. Cancer Metab 1:8–10.

Shrestha RP, Hildebrand M (2015) Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses. Eukaryotic Cell 14:29–40.

Smith SR, Abbriano RM, Hildebrand M (2012) Comparative analysis of diatom genomes

reveals substantial differences in the organization of carbon partitioning pathways. Algal Research 1:2–16.

Smith SR, Gillard JTF, Kustka AB, McCrow JP, Badger JH, Zheng H, New AM, Dupont CL, Obata T, Fernie AR, Allen AE (2016a) Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. PLoS Genet 12:e1006490.

Smith SR, Glé C, Abbriano RM, Traller JC, Davis A, Trentacoste E, Vernet M, Allen AE, Hildebrand M (2016b) Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. New Phytol 210:890–904.

TeSlaa T, Teitell MA (2014) Techniques to monitor glycolysis. Methods in enzymology 542:91–114.

Traller JC, Hildebrand M (2013) High throughput imaging to the diatom *Cyclotella cryptica* demonstrates substantial cell-to-cell variability in the rate and extent of triacylglycerol accumulation. Algal Research 2:244-252.

Valenzuela J, Carlson RP, Gerlach R, Cooksey K, Peyton BM, Bothner B, Fields MW (2013) Nutrient resupplementation arrests bio-oil accumulation in *Phaeodactylum tricornutum*. Appl Microbiol Biotechnol 97:7049–7059.

Vander Heiden MG, Cantley LC, Thompson CB (2009) Understanding the Warburg Effect: the metabolic requirements of cell proliferation. Science 324:1029–1033.

Vaulot D, Olson RJ, Merkel S, Chisholm SW (1987) Cell-cycle response to nutrient starvation in two phytoplankton species, *Thalassiosira weissflogii* and *Hymenomonas carterae*. Marine Biology 95:625–630.

Vitova M, Bisova K, Kawano S, Zachleder V (2015) Accumulation of energy reserves in algae: From cell cycles to biotechnological applications. Biotechnol Adv 33:1204–1218.

Waterkeyn L, Bienfait A (2011) Localization and Role of B-1.3 Glucans in the Genera *Pinnularia*. La Cellule 74: 198–226.

Work VH, Radakovits R, Jinkerson RE, Meuser JE, Elliott LG, Vinyard DJ, Laurens LML, Dismukes GC, Posewitz MC (2010) Increased lipid accumulation in the *Chlamydomonas reinhardtii* sta7-10 starchless isoamylase mutant and increased carbohydrate synthesis in complemented strains. Eukaryotic Cell 9:1251–1261.

Wu C, Okar DA, Newgard CB, Lange AJ (2001) Overexpression of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase in mouse liver lowers blood glucose by suppressing hepatic glucose production. J Clin Invest 107:91–98.

Yalcin A, Clem BF, Simmons A, Lane A, Nelson K, Clem AL, Brock E, Siow D, Wattenberg B, Telang S, Chesney J (2009) Nuclear targeting of 6-phosphofructo-2-kinase (PFKFB3) increases proliferation via cyclin-dependent kinases. J Biol Chem 284:24223–24232.

Yang Z-K, Niu Y-F, Ma Y-H, Xue J, Zhang M-H, Yang W-D, Liu J-S, Lu S-H, Guan Y, Li H-Y (2013) Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. Biotechnol Biofuels 6:1–67.

Yu ET, Zendejas FJ, Lane PD, Gaucher S, Simmons BA, Lane TW (2009) Triacylglycerol accumulation and profiling in the model diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (Baccilariophyceae) during starvation. 21:669–681.

Zhao G, Chen Y, Carey L, Futcher B (2016) Cyclin-dependent kinase co-ordinates carbohydrate metabolism and cell cycle in *S. cerevisiae.* Mol Cell 62:546–557.

TABLES

Table 3-1. Measurements of photosynthetic parameters in wild-type and Group 2 overexpression lines.

Cell line	Chlorophyll fluorescence at 488 nm (RFU)	Quantum efficiency (F_v/F_m)	O_2 evolution (pmol/cell/hour)
WT	$6.36e^4$ +/- $1.78e^4$	0.695 +/- 0.022	0.415 +/- 0.049
109fcp1	$6.31e^4$ +/- $1.13e^4$	0.686 +/- 0.050	0.457 +/- 0.093
109fcp2	$6.18e^4$ +/- $6.22e^3$	0.674 +/- 0.030	0.358 +/- 0.115

FIGURES

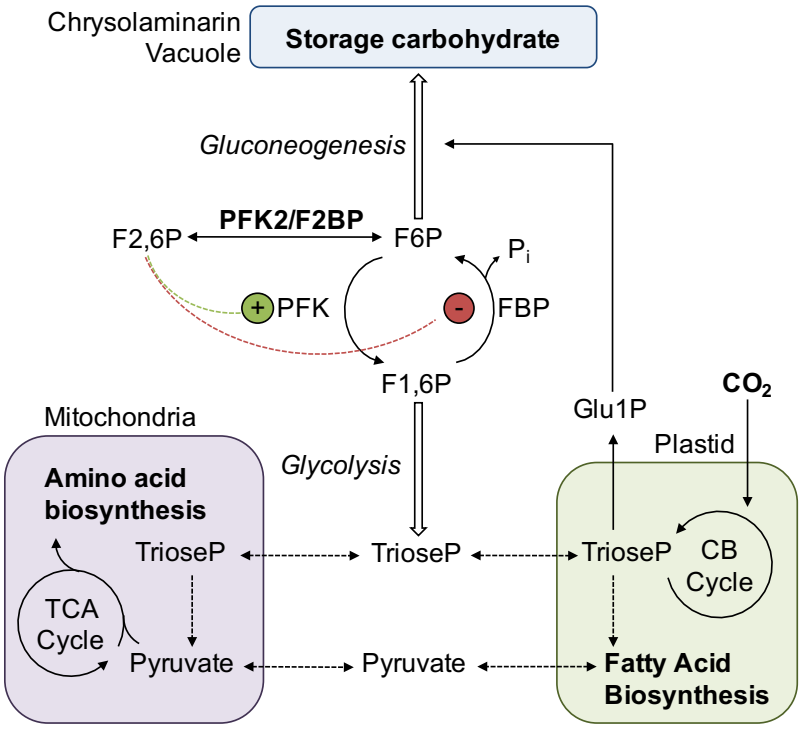

Figure 3-1. Model for the regulatory influence of phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFK2/F2BP) on carbon flux in diatoms. Key: F1,6P = fructose 1,6-bisphosphate; F2,6P = fructose 2,6-bisphosphate; F6P = fructose 6-phosphate; FBP = fructose-1,6-bisphosphatase; PFK = phosphofructokinase; Glu1P = glucose-1-phosphate. PFK2 activity produces more F2,6P, which allosterically activates PFK (promoting glycolysis) and inhibits FBP (reducing gluconeogenesis).

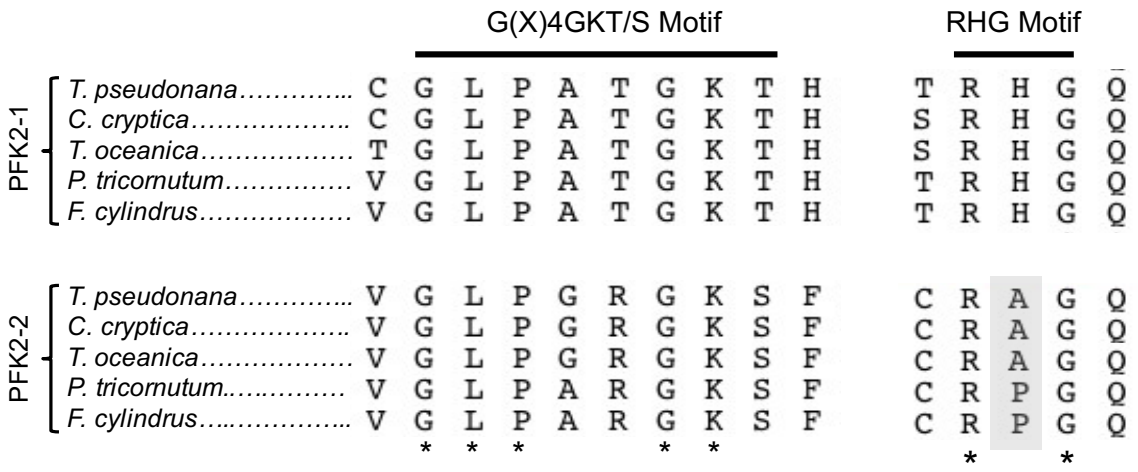

Figure 3-2. Multiple sequence alignment of conserved domains in six diatom PFK2/F2BP isozymes. The grey box indicates a substitution of a critical histidine residue in the F2BP domain that is required for phosphatase activity. The conserved substitution for the histidine in the RHG motif suggests that Group 2 enzymes function unidirectionally to promote carbon flux in the glycolytic direction.

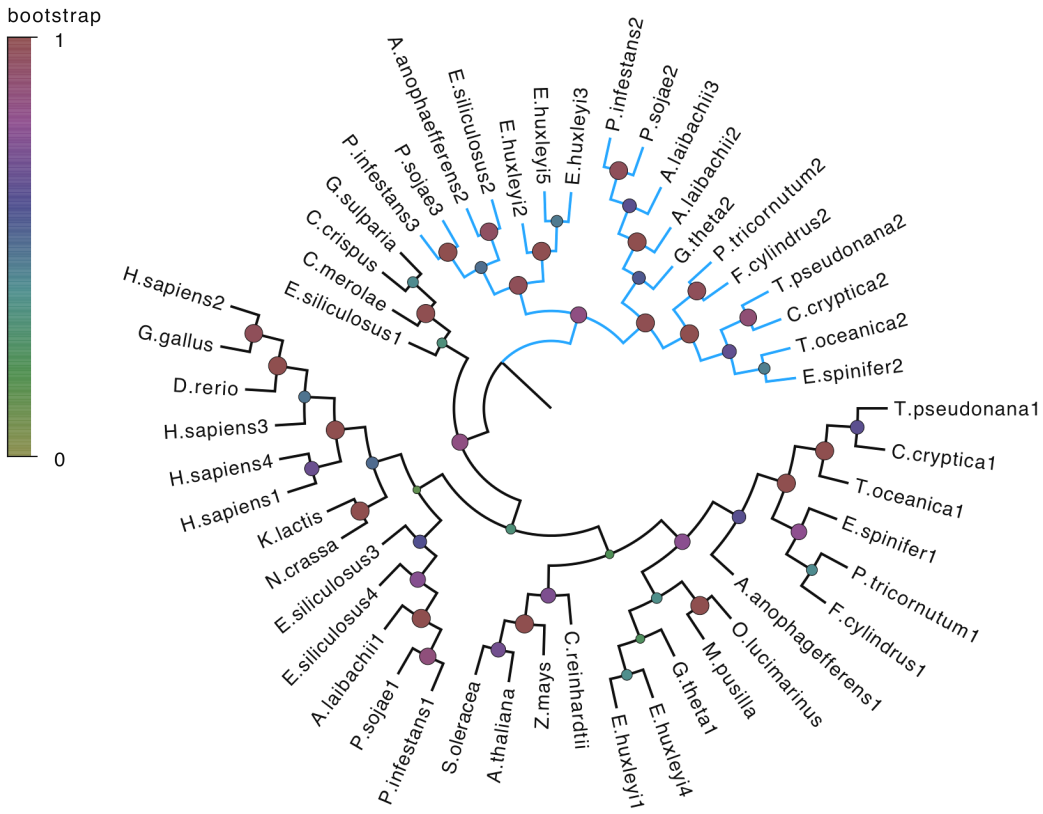

Figure 3-3. Phylogenetic relationships among 51 PFK2/F2BP sequences. Node labels (colored circles) are scaled to reflect bootstrap support. Group 1 PFK2/F2BP sequences (PFK2-1) are highlighted in grey; Group 2 PFK2/F2BPs sequences (PFK2-2) in blue.

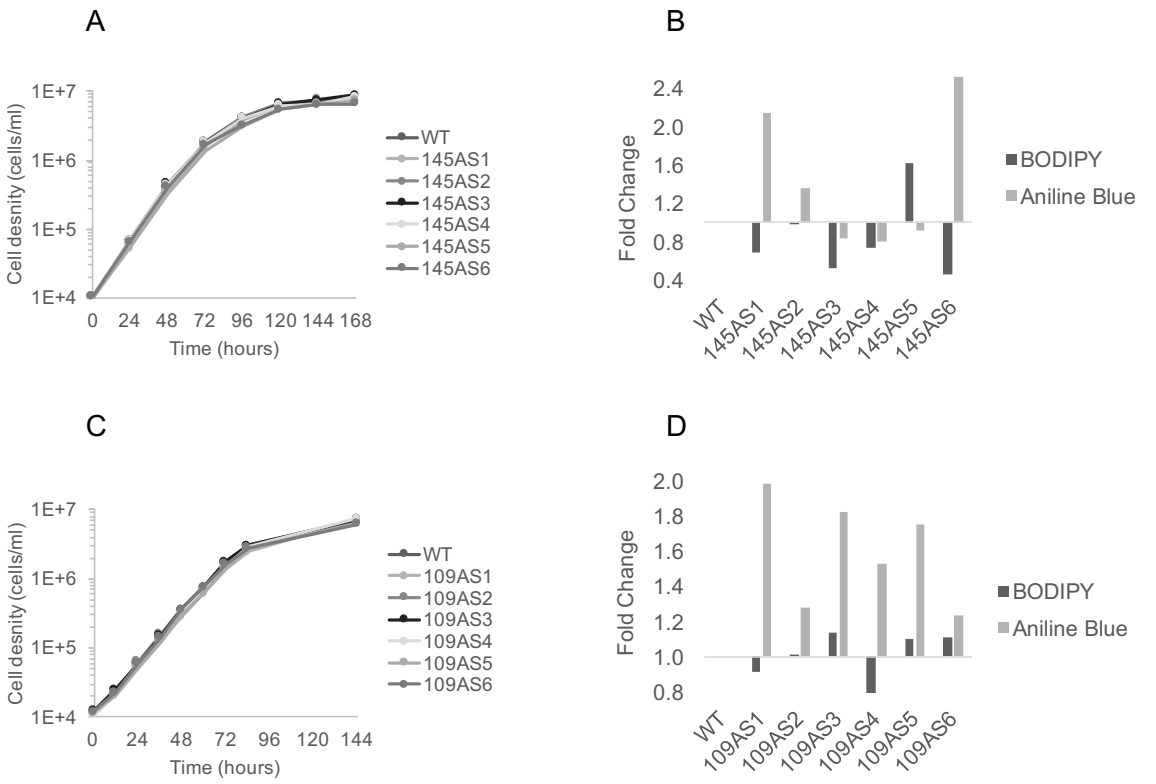

Figure 3-4. Screening of Group 1 (Thaps3_14563) and Group 2 (Thaps3_109629) antisense knockdown lines. A and C) Growth curves for Group 1 and 2 knockdown lines, respectively. B and D) Neutral lipid and β-1,3 glucan content in Group 1 and 2 knockdown lines relative to wild-type as measured by BODIPY and aniline blue fluorescence.

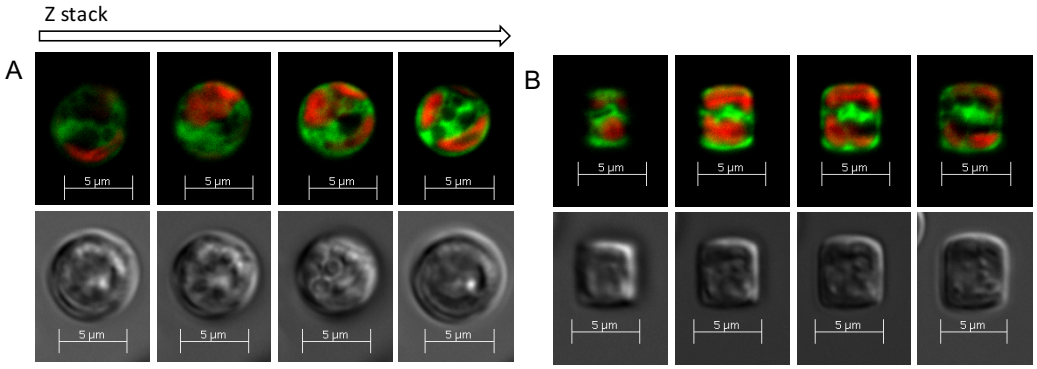

Figure 3-5. Z-stack micrographs of *T. pseudonana* cells in A) valve orientation and B) girdle orientation expressing the PFK2-2/GFP fusion protein in the cytosol. The top row of each panel contains false color fluorescent images where GFP is represented by green and chlorophyll autofluorescence is represented by red. The bottom row of each panel contains the corresponding DIC images.

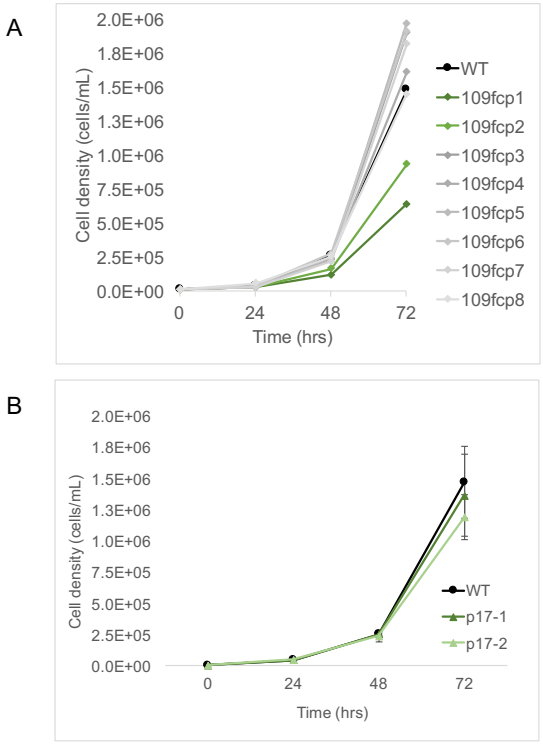

Figure 3-6. Growth curves for PFK2-2 overexpression lines and eGFP controls at 50 ml scale. Black = wild type, Gray = transgenic ines with no visual GFP fluorescence; Green = transgenic lines with visible GFP fluorescence. A) Comparison of PFK2-2 overexpression lines with wild type; only 109fcp1 and 109fcp2 are expressing eGFP and have reduced growth characteristics. B) Control transgenic lines expressing eGFP under the fcp promoter/terminator pair have comparable growth to wild type.

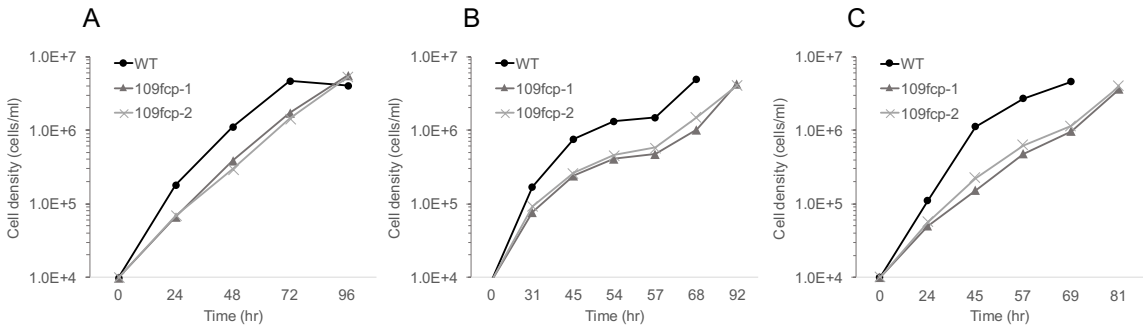

Figure 3-7. Growth curves for three independent biological replicates (A, B, and C) for wild-type, 109fcp1, and 109fcp2 lines.

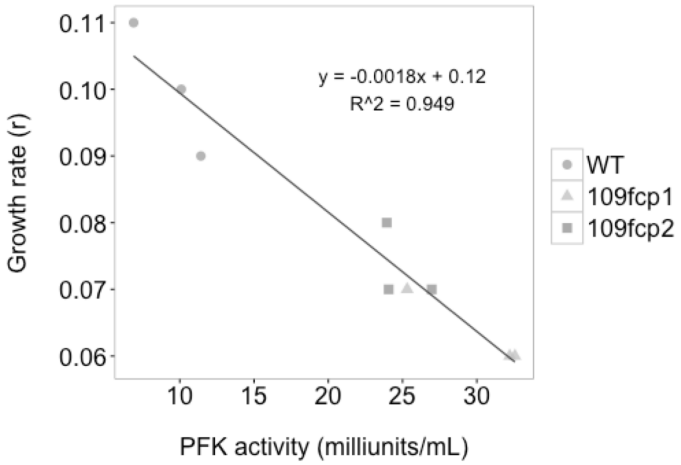

Figure 3-8. Relationship between growth rate and PFK activity in *T. pseudonana* wild-type and two PFK2-2 overexpression lines. Measurements were taken from three independent biological replicates (n=3).

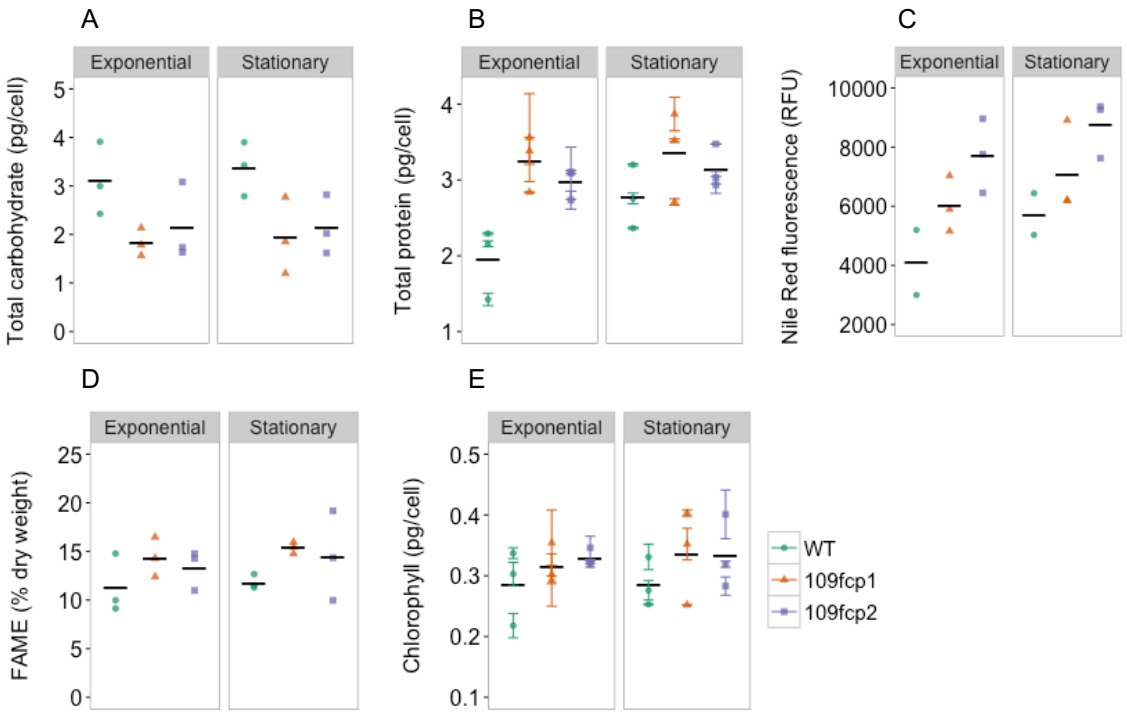

Figure 3-9. Biochemical composition of wild-type and Group 2 overexpression lines in exponential and stationary phase. A) total carbohydrate content; B) protein content; C) neutral lipid content quantified by Nile Red fluorescence; D) FAME content; E) chlorophyll content. Each point represents an independent biological replicate; error bars represent standard deviation of technical replicates.

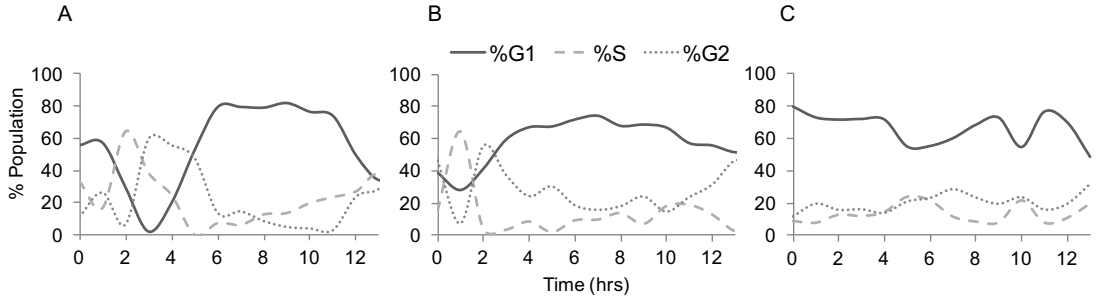

Figure 3-10. Progression of A) Wild-type, B) 109fcp1, and C) 109fcp2 *T. pseudonana* cultures through the cell cycle post-synchronization via silicon deprivation and readdition.

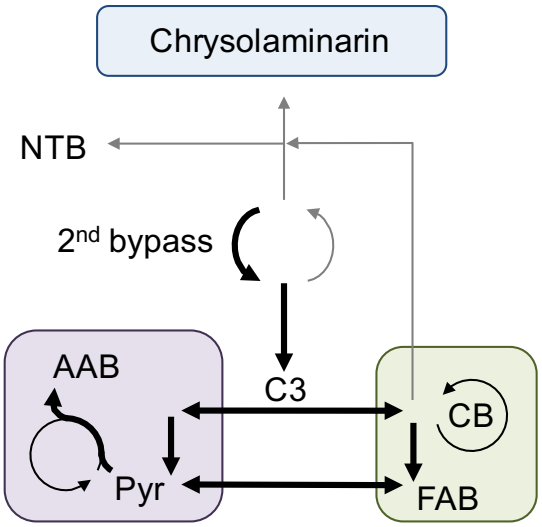

Figure 3-11. Schematic representation of the role of PFK2-2 in *T. pseudonana*. Increased PFK2-2 increases PFK activity and glycolytic flux at the second bypass, increasing carbon precursor (C3) availability for amino acid (AAB) and fatty acid biosynthesis (FAB). Concurrent reduction in gluconeogenic flux deprives chrysolaminarin and nucleotide biosynthesis (NTB).

CONCLUSIONS

Diatoms are a high-productivity photosynthetic system to produce valuable compounds such as lipids, proteins, vitamins, pigments, and antioxidants (Bozarth et al. 2009). Increasing the yield of these valuable bioproducts through metabolic engineering has become an active area of research in algal biology, and advances in the understanding of the organization and regulation of metabolism in diatom cells is critical to enable more strategic pathway engineering to produce a myriad of useful and interesting metabolites. The goal of this dissertation was to address specific knowledge gaps related to carbon metabolism and to gain insight into the relationship between carbon metabolism and growth in marine diatoms.

Work in Chapter 1 expanded the set of useful gene annotations in the model marine diatom *T. pseudonana*. As the first sequenced diatom genome (Armbrust et al. 2004), *T. pseudonana* has been considered a model marine diatom species. This improved gene set is a reference that can be used by other diatom researchers interested in understanding gene structure, gene regulation, and the organization of diatom metabolism. In addition, RNAseq reads provides information at single base pair resolution, opening the potential for future evaluation of allelic differences in the genome. While the analysis of the resulting models suggests a marked improvement in gene model quality, this effort would benefit from further manual refinement. In particular, UTR predictions could not be accurately predicted by AUGUSTUS. In addition, an initial manual assessment of 50 randomly chosen gene IDs suggests that approximately 24% of AUGUSTUS genes could be further refined manually. This would

be a relatively large effort beyond the scope of an individual lab. Manual annotation could be accomplished via a concerted effort on behalf of the diatom community using tools such as WebApollo (Lee et al. 2013), which is designed for collaborative genome annotation.

Extension of existing models revealed additional protein targeting information, improving the ability to predict the subcellular localization of proteins *in silico* and further refine metabolic models. An additional set of genes was identified (1,883 genes), many of which are differentially expressed and therefore may be important to cell cycle or silicon bioprocesses. These genes included several SMC_prok_B domain-containing proteins that are highly expressed during early mitosis. These are distinct from eukaryotic-type SMC proteins, which are also found in *T. pseudonana* (Patil et al. 2015). This raises questions related to whether these sets of SMC proteins work together or independently to facilitate chromosome condensation and cohesion in diatoms prior to cell division. In addition, SMC proteins can also participate in other cellular functions, such as recombination, DNA repair, and regulation of gene expression (Harvey et al. 2002). Proteins with these functions are of biotechnological interest because they may influence the efficiency of homologous recombination, which has recently been shown to occur in *P. tricorntum* (Daboussi et al. 2014), but has not been reported in centric diatoms. I also identified additional dsCYC proteins in *T. pseudonana*, including one that is highly responsive to silicon availability. While some components of cell cycle regulation have been elucidated in *P. tricornutum* (Huysman et al. 2013b), several questions remain unresolved. In particular, only one dsCYC has been functionally characterized (Huysman et al. 2013a). However, transcriptomic studies suggest that

dsCYCs may serve as signal integrators that relay information to CDKs (Huysman et al. 2013b). Additional functional characterization of these proteins, either through targeted reverse genetic approaches or through forward genetic screens, will be necessary to demonstrate the effect of SMC and dsCYC proteins on the regulation of the cell cycle.

Chapter 1 also includes the first genome-wide assessment of alternative splicing in diatoms, and demonstrated that alternative splicing events in *T. pseudonana* are dominated by intron retention. Approximately 3% of genes to were differentially alternatively spliced by rMATS analysis. However, this finding is specific to this dataset, and it is possible that additional splicing could be identified in other experimental conditions or with deeper sequencing. The biological significance of intron retention is debated in the literature (Ner-Gaon et al. 2004; Keren et al. 2010; Braunschweig et al. 2014), and it remains to be determined if and how this process may contribute to gene expression regulation in diatoms. In other organisms, splicing is controlled by splice sites and specific splicing regulatory elements (De Conti et al. 2013). A global assessment of intronic sequences in *T. pseudonana* would likely provide additional insights into the evolution of introns and mechanisms that affect splicing in diatoms.

Chapter 2 combines *in silico* sequence analysis (including extended gene models generated in Chapter 1), gene expression data, and *in vivo* protein localization to refine a metabolic model for C2 photorespiratory metabolism in *T. pseudonana*. Photorespiration can be a major route for carbon flux in photosynthetic cells, and may also be an important mechanism by which reductant is transferred between diatom chloroplasts and mitochondria in order to fine-tune photosynthetic carbon fixation under various irradiances (Bailleul et al. 2015). The analysis confirms the localization of a glycolate

oxidase enzyme (GOX2) to the mitochondria, where it may participate in the direct delivery of reductant from the plastid to the mitochondrial electron transport chain as proposed by Winkler and Stabenau (1995). A subsequent study demonstrated that the homolog of the *T. pseudonana* GOX2 enzyme in *P. tricornutum* can use alternate electron acceptors (Schmitz et al. 2017), further supporting this hypothesis. Further testing of this theory could be tested in the future by measuring the effect of GOX2 knockdown on glycolate accumulation, AOX activity, and photosynthesis.

Chapter 2 analysis did not support that significant glycolate oxidation occurred in the peroxisome. Rather, it supports the idea that transcriptional upregulation of the peroxisomal glyoxylate cycle works in conjunction with mitochondrial malic enzyme to provision the TCA cycle and/or photorespiratory metabolism with pyruvate. The analysis also did not support the presence of a complete C2 photorespiratory pathway that returns carbon to the plastid due to the absence of a glycerate kinase sequence. The apparent absence of glycerate kinase could be further confirmed by testing for glycerate kinase activity. Further validation of the entire photorespiratory model could be accomplished by genome-scale metabolic modeling, as has been done for *P. tricornutum* (Kim et al. 2015).

Chapter 3 demonstrated the importance of PFK2-2 on the regulation of glycolytic flux and cell cycle progression in *T. pseudonana*. The results showed that, especially during early growth stages, increased glycolytic flux enhanced both protein and lipid accumulation and reduced carbohydrate content. A decrease in growth rate was also observed in response to this shift in metabolic flux. While this manipulation provides insights into the balance between biosynthesis and growth in diatom cells, it is not a

particularly useful constitutive manipulation since PFK2 overexpression limits biomass production. However, one could potentially take advantage of this regulatory point in conjunction with other modifications by temporarily inducing higher glycolytic flux to conditionally increase substrate availability to other pathways.

The work related to the regulation of cytosolic glycolysis has also provided insight into the organization of diatom central metabolism and into how multiple glycolytic pathways in separate compartments might work together to contribute to diatom productivity. For example, Chapter 3 demonstrates that cytosolic glycolysis plays a large role in carbon partitioning and biomolecule accumulation, but there is no evidence that it is directly involved in energy generation. Expression data and other metabolic modeling (Kim et al. 2015; Smith et al. 2016b; Smith et al. 2016a) suggest that mitochondrial glycolysis appears to be main pathway involved in cellular energetics to fuel growth. Localization of carbohydrate biosynthesis to the cytosol decouples that process from both the plastid and the mitochondria, allowing for simultaneous accumulation of carbohydrates in the cytosol and the production of energy through glycolysis in the mitochondria. Due to this unique arrangement, diatoms may have some energetic advantages relative to other organisms that do not similarly compartmentalize their metabolism.

REFERENCES

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306:79–86.

Bailleul B, Berne N, Murik O, Petroutsos D, Prihoda J, Tanaka A, Villanova V, Bligny R, Flori S, Falconet D, Krieger-Liszkay A, Santabarbara S, Rappaport F, Joliot P, Tirichine L, Falkowski PG, Cardol P, Bowler C, Finazzi G (2015) Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. Nature 54:366–369.

Bozarth A, Maier U-G, Zauner S (2009) Diatoms in biotechnology: modern tools and applications. Appl Microbiol Biotechnol 82:195–201.

Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis T, Frey B, Irimia M, Blencowe BJ (2014) Widespread intron retention in mammals functionally tunes transcriptomes. Genome Research 24:1774–1786.

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

De Conti L, Baralle M, Buratti E (2013) Exon and intron definition in pre-mRNA splicing. WIREs RNA 4:49–60.

Harvey SH, Krien MJ, O'Connell MJ (2002) Structural maintenance of chromosomes (SMC) proteins, a family of conserved ATPases. Genome Biol 3:reviews3003.1-3003.5.

Huysman MJJ, Fortunato AE, Matthijs M, Costa BS, Vanderhaeghen R, Van den Daele H, Sachse M, Inzé D, Bowler C, Kroth PG, Wilhelm C, Falciatore A, Vyverman W, De Veylder L (2013a) AUREOCHROME1a-Mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (*Phaeodactylum tricornutum*). Plant Cell 25:215–228.

Huysman MJJ, Vyverman W, De Veylder L (2013b) Molecular regulation of the diatom cell cycle. Journal of Experimental Botany 65:2573–2584.

Keren H, Lev-Maor G, Ast G (2010) Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 11:345–355.

Kim J, Fabris M, Baart G, Kim MK, Goossens A, Vyverman W, Falkowski PG, Lun DS (2015) Flux balance analysis of primary metabolism in the diatom *Phaeodactylum tricornutum*. The Plant Journal 85:161–176.

Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE (2013) Web Apollo: a web-based genomic annotation editing platform. Genome Biol 14:R93.

Ner-Gaon H, Halachmi R, Savaldi-Goldstein S, Rubin E, Ophir R, Fluhr R (2004) Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*. The Plant Journal 39:877–885.

Patil S, Moeys S, Dassow von P, Huysman MJJ, Mapleson D, De Veylder L, Sanges R, Vyverman W, Montresor M, Ferrante MI (2015) Identification of the meiotic toolkit in diatoms and exploration of meiosis-specific SPO11 and RAD51 homologs in the sexual species *Pseudo-nitzschia multistriata* and *Seminavis robusta*. BMC Genomics 16:930.

Schmitz J, Srikanth NV, Hüdig M, Poschmann G, Lercher MJ, Maurino VG (2017) The ancestors of diatoms evolved a unique mitochondrial dehydrogenase to oxidize photorespiratory glycolate. Photosynth Res 132:183–196.

Smith SR, Gillard JTF, Kustka AB, McCrow JP, Badger JH, Zheng H, New AM, Dupont CL, Obata T, Fernie AR, Allen AE (2016a) Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. PLoS Genet 12:e1006490.

Smith SR, Glé C, Abbriano RM, Traller JC, Davis A, Trentacoste E, Vernet M, Allen AE, Hildebrand M (2016b) Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. New Phytol 210:890–904.

Winkler U, Stabenau H (1995) Isolation and characterization of peroxisomes from diatoms. Planta 195:403–407.