UNIVERSITY OF CALIFORNIA, SAN DIEGO

Characterization of carbon metabolism in an oleaginous diatom suitable for commercial production

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

in

Marine Biology

by

Jessica C. Traller

Committee in charge:

Mark Hildebrand, Chair
Andrew E. Allen
Eric E. Allen
Nigel Crawford
Brian Palenik
Matteo Pellegrini

2017

The Dissertation of Jessica C. Traller is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2017

DEDICATION

This dissertation is dedicated to my wonderful parents, Carol and Dave Traller, for their ongoing support and unwavering optimism in me. I am forever grateful to have been shown such admirable examples of work ethic and intellect from my father, and perseverance and ambition from my trail-blazing mother. I am proud to be their daughter.

I would also like to dedicate this dissertation to my doctoral advisor, Dr. Mark Hildebrand. His passion for the pursuit of scientific discovery, his patience, and understanding has allowed me to fulfill a lifelong goal of obtaining a Doctor of Philosophy. I cannot thank him enough for his role in my life as a mentor and friend.

TABLE OF CONTENTS

LIST OF SUPPLEMENTAL FILES

Additional File 1:	OrthoMCL and functional annotation, methylome DarkHorse, subcellular targeting
Additional File 2:	Periplastid predicted genes
Additional File 3:	Triose phosphate transporters

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

It truly takes a village to obtain a Ph.D. First and foremost I would like to acknowledge my advisor, mentor, and friend Dr. Mark Hildebrand. Mark is one of the most patient, optimistic, and supportive people I know. There is no question in my mind I would not be here at the end of this academic journey were it not for Mark's support and belief in me. I owe him a lifetime of gratitude for his support.

I would also like to thank my doctoral committee, Drs. Andrew Allen, Eric Allen, Nigel Crawford, Brian Palenik, and Matteo Pellegrini for their valuable advice and guidance during my dissertation. Their insight and support particularly during the annotation of the *Cyclotella* genome made what seemed like an impossible task possible. I am especially grateful to Dr. Matteo Pellegrini, Dr. Sabeeha Merchant, and members of their labs, Drs. Shawn Cokus, David Lopez, Sean Gallaher, among others, for the numerous Skype sessions, endless clarifications of complex bioinformatic vernacular, and fruitful scientific discussions. The *Cyclotella* genome project with UCLA was a 6-7 year (!?) project that not only was an essential aspect of my dissertation, but was instrumental to my academic development and personal growth. I am proud to have these inspirational scientists as collaborators.

To my al-gals, Dr. Sarah Smith, Raffie Abbriano Burke, and Sarah Lerch... three women whom are among the most intelligent, daring, and inspirational women I have met. Sarah S, you have been as wonderful of a mentor as you have a friend and confidant. Sarah L, I could not have asked for a better officemate and hold our many 'door-closed' personal and scientific discussions near and dear to my heart. Raffoli, you

have been the best scientific partner, conference buddy, and friend that I could have ever asked for. My relationships with these three individuals are the some of the best examples of support, warmth, and friendship that I have ever experienced. To other members of the Hildebrand lab, past and present, especially Roshan Shrestha, Olia Gaidarenko, Eva Sanchez Alvarez, Aubrey Davis, Daniel Yee, Emily Trentacoste, and Orna Cook, thank you for all of your help and lifesaving lab advice over the years.

I have been part of the SIO community for almost a decade, entering when I was just a wee young college kid and becoming the (forever) young scientist and woman I am proud to be today. As an SIO student, I am especially lucky to have been surrounded by so many encouraging and intellectual voices. Thank you to Dr. Bianca Brahamsha, Dr. Vic Vacquier, Dr. John McCrow, Dr. Martin Tresguerres, Dr. Mike Latz, Dr. Dominick Mendola, Dr. Peter Franks. I am also thankful for the extremely helpful administration at the SIO graduate office and Hubbs Hall, especially Gilbert Bretado, Adam Petersen, Denise Darling, the Maureens (McCormick and McGreevy), Sam Chin, James Pollock, and Yvvone Bohan. Additionally, I'd like to thank fellow members and friends of the algae biotechnology community particularly Juergen Polle, Wendy Groves, Steve Mayfield, Rachel Diner, Niu Du, Flip McCarthy, Dave Hazlebeck, Tiffany Cannis, Aga Pinowska, Rodney Corpuz, and all my friends at Global Algae Innovations in Kauai.

Thank you to all the SIO friends, especially my 2011 cohort and 3rd floor Hubbsters, for making me feel a little less alone in this crazy thing called graduate school. To my life support outside of SIO- Audrey, Heidi, Cassie, Elena, Chuckie, Yana, Christy, Christina, Ashley, Cec, Nika, Libs, Kasey, Katrina, Nastash- you guys

have kept me sane, kept me laughing, and are the best description of lifelong amigos. Let's play!

To my family, especially Jordan, Casey, Bob, and Mama and Papá Ojeda, thank you for your continued belief in me. To my nephew and nieces, Hank, Maggie, and Camila, may I suggest you become a scientist when you grow up? The work is hard but the discoveries are extraordinary.

I had the absolute fortune to meet my Juan and only during my time as a graduate student. Juan, it is no doubt that I am on solid ground today because of you. Your patience, support, and love have allowed me to finish something that has been on the bucket list for quite some time. I cannot wait to begin this new chapter in my life with you by my side.

Lastly, to my Mommy and Daddy, I hope I have made you proud. You deserve the best of me, and I am indebted to your encouragement, love, and kindness. Thank you for shaping me into everything I am today.

MATERIAL PUBLISHED/PREPARED FOR PUBLICATION IN THE DISSERTATION

Chapter 1, in full is a reprint of the material as it appears in *Algal Research* 2013. Traller, Jesse C., and Hildebrand, Mark. "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(3), 244-252. 2013. The dissertation author was the primary researcher and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in Biotechnology for Biofuels 2016. Traller, Jesse C, Shawn J Cokus, David A Lopez, Olga Gaidarenko, Sarah R Smith, John P McCrow, Sean D Gallaher, Sheila Podell, Michael Thompson, Orna Cook, Marco Morselli, Artur Jaroszewicz, Eric E Allen, Andrew E Allen, Sabeeha S Merchant, Matteo Pellegrini, Mark Hildebrand. "Genome and Methylome of the Oleaginous Diatom *Cyclotella Cryptica* Reveal Genetic Flexibility Toward a High Lipid Phenotype," *Biotechnology for Biofuels,* 9(1), 258. 2016. The dissertation author was the primary researcher and author of this paper.

Chapter 3, in full is material currently being prepared for submission for publication. Traller, Jesse; Hildebrand, Mark. "Comparison of the transcript level and physiological response in the diatom *Cyclotella cryptica* under two distinct macronutrient limitations." Ms. Traller was the principal researcher/author on this paper.

VITA

2008 Bachelor of Science, University of California, San Diego
2016 Master of Science, University of California, San Diego
2017 Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Traller, J. C., Cokus, S. J., Lopez, D. A., Gaidarenko, O., Smith, S. R., McCrow, J. P., et al. (2016). Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. *Biotechnology for Biofuels*, 9(1), 258.

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

Hildebrand, M., Davis, A., Abbriano, R., Pugsley, H. R., **Traller, J. C.,** Smith, S. R., et al. (2015). Applications of Imaging Flow Cytometry for Microalgae. In N. S. Barteneva & I. A. Vorobjev (Eds.), *Imaging Flow Cytometry* (pp. 47–67). New York, NY: Springer New York.

Traller, J. C., & 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*(3), 244–252.

Hildebrand, M., Abbriano, R. M., Polle, J. E., **Traller, J. C.,** Trentacoste, E. M., Smith, S. R., & Davis, A. K. (2013). Metabolic and cellular organization in evolutionarily diverse microalgae as related to biofuels production. *Current Opinion in Chemical Biology*, *17*(3), 506–514.

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

ABSTRACT OF THE DISSERTATION

Characterization of carbon metabolism in an oleaginous diatom suitable for commercial production

by

Jessica C. Traller

Doctor of Philosophy in Marine Biology

University of California, San Diego, 2017

Mark Hildebrand, Chair

The realization of the environmental and economic impact that occurs from the release of carbon dioxide derived from fossil fuel has spurred technologies focused on renewable alternative energy production. The production of liquid based biofuel from microalgae has been one of the proposed solutions due to microalgae's high productivity per acre of land, its ability to accumulate large amounts of the precursor to biofuel, triacylglycerol, and its dependence on carbon dioxide, allowing for

sequestration of anthropogenic CO_2 from the atmosphere. The promise of microalgae as a feedstock for the large-scale commercial production of biofuel is impressive, however, there are significant advances in productivity still needed in order to make the solution economically and industrially feasible. For instance, the mechanisms that control triacylglycerol accumulation and cellular division in microalgae are still largely unknown. Moreover, most studies on these processes have been focused on model organisms for basic research, not on species that are industrially relevant. The immense diversity of algae as well as their complex evolutionary history indicates that molecular mechanisms controlling primary metabolism are not necessarily ubiquitous. Therefore, elucidation of carbon metabolism in a candidate species for commercial production is key for the biological comprehension within algal biofuel production. The main objective of this dissertation was to characterize primary carbon metabolism in the diatom *Cyclotella cryptica,* a species of microalgae with promise for large-scale industrial production. Chapter 1 investigated the neutral lipid accumulation response of *C. cryptica* through the comparison of two macronutrient starvation using imaging flow cytometry. Chapters 2 and 3 applied high-throughput sequencing technologies to identify the molecular and metabolic controls of triacylglycerol accumulation. The data presented in this dissertation will hopefully move this species and others closer to the economic feasibility of commercial scale algal biofuel by characterizing key areas of metabolism pertinent to triacylglycerol formation, providing genomic datasets necessary for the development of genetic engineering tools, and establishing a model for primary carbon metabolism in an industrially relevant species.

INTRODUCTION

The release of carbon dioxide and other greenhouse gases into the atmosphere by fossil fuel consumption has undoubtedly caused shifts in our global climate (Ramanathan and Carmichael 2008; Solomon et al. 2010; Change 2014). There is little argument within the scientific community that if left ignored, anthropogenic climate change will have adverse socioeconomic and environmental consequences in our future. There have been several proposed solutions to combat climate change that involve easing our dependence on fossil fuels and transitioning into sustainable energy practices. Biofuel production is one these proposed solutions, a technology that converts biomass and oil derived from organic material such as plants, animal waste, and microalgae into liquid fuel.

Microalgae are unicellular, mostly photosynthetic organisms that hold a lot of promise as a sustainable feedstock for biofuel production. They have the ability to be grown on non-arable land, utilize non-potable water, and through chemical conversion, microalgae can be directly fed into existing liquid diesel-based infrastructure, in contrast to bioethanol or biohydrogen (Sheehan et al. 1998; Georgianna and Mayfield 2012). Microalgae are also intriguing as a feedstock for biofuel because of their ability to accumulate substantial amounts of the neutral lipid triacylglycerol (TAG); which is often induced at the onset of an environmental stress, such as nutrient starvation. The stress impairs the cell's ability to divide and as a result the carbon and energy that would have been used for cellular growth is instead collected into large lipid droplets, which can account for up to 20-50% of the dry weight (Hu et al. 2008). The main fatty acids found in algae-derived triacylglycerol are chemically similar to the hydrocarbons of fossil fuel; in fact, fossil fuel itself is partly derived from ancient algal blooms (Falkowski et al. 2004). Consequently, TAG can be extracted out of the cell and transesterified into biofuel

using existing technologies. In order for the renewable algal biofuel industry to be economically viable and competitive with the current energy market there needs to be major advances in improvement of productivity. This productivity could be improved in two different ways: by increasing the biomass- the cellular density per acre, or by increasing the amount of product, in this case triacylglycerol, within the algal cell.

In addition to TAG utilization, microalgae have also been proposed for high protein animal feed. Like algal biofuels, switching food-producing animals such as cattle from a corn to microalgae dietary source could also lower carbon emissions, due to the algae's high yield per acre of land (Georgianna and Mayfield 2012). The nutritional characteristics of microalgae with regard to protein, carbohydrate, and lipid are comparable to conventional feed, and contain a relatively balanced distribution of amino acids (Lum et al. 2013; Yaakob et al. 2014). Current models have proposed microalgae to be harvested for both animal-feed and biofuel production, with the lipid allotted for biofuel and the protein 'waste' processed for animal feed. This model could aid in making the overall process economically viable by producing two distinct products. In addition to animal-feed, there are particular algal species that naturally produce a number of high value nutraceutical products, such as such as astaxanthin and lutein, further highlighting the economic potential of these organisms (Kobayashi et al. 1991; Cordero et al. 2011).

From an environmental context, microalgae are among the most important organisms on the planet; responsible for nearly half of the planetary oxygen generated, forming the basis of the oceanic food web, and participating as key players in global biogeochemical cycles. The importance of microalgae in the environment, as well as their

potential as an alternative energy and food source, demonstrates the need to understand the underlying basis of productivity of these organisms in a biological and molecular context.

The oil crises of the early and mid 1970's in the United States led to the creation of the Department of Energy and the Aquatic Species Program (ASP), which provided foundational and groundbreaking research in algae. The ASP spanned from 1978-1996 and carried out extensive research pertaining to applied algal biology, algal production systems, and natural resource availability for large-scale cultivation. Over its existence, the ASP conducted extensive screening across the algal tree of life through collection and evaluation of over 3,000 strains of algae. Species were screened based on their abilities to: i) accumulate abundant triacylglycerol ii) grow under variable environmental conditions (pH, salinity, temperature), and iii) grow at a rate that is productive and sustainable in outdoor pond cultivation. Of the various algae tested, a few genera stood out from the rest, including Chlorophytes, Chrysophytes, and diatoms. The latter, diatoms (Bacillariophyta), are a class of microalgae within the superphylum Heterokonta, known for their siliceous cell walls, and require silicon as a macronutrient. They are also known for their ecological success in the global oceans, responsible for an estimated 40% of the oceanic primary productivity (Field et al. 1998). Because of their high productivity, diatoms accounted for ~60% of the species in the top fifty candidates of biofuel-production organisms isolated during the ASP (Sheehan et al. 1998; Hildebrand et al. 2012).

One of those species featured in the ASP was the marine diatom *Cyclotella cryptica*. *C. cryptica,* along with another diatom were the first chlorophyll-c containing

algae to undergo stable nuclear transformation, a major milestone in genetic engineering of algae (Dunahay, Jarvis, & Roessler, 1995). In addition to the molecular work, *C. cryptica* also was shown to successfully grow in outdoor pond systems, with productivity levels averaging to 29.7 g AFDW/m^2 day in a 48m^2 outdoor pond (Laws et al. 1988). *C. cryptica* was shown to be a good accumulator of lipid (Roessler 1988a; Sheehan et al. 1998), with the ability to grow in under a wide range of conditions (Lewin and Lewin 1960; Reimann et al. 1963), and at a rate potentially suitable for production scale (Sheehan et al. 1998). More recently, in a survey of 175 strains of microalgae, *C. cryptica* performed in the top 10 for protein, nitrogen assimilation, and eicosapentaenoic acid production (Slocombe et al. 2015).

Despite the promising studies that were conducted during the ASP, applied research on microalgae, specifically diatoms, for large-scale biofuel/bioproduct cultivation nearly stopped after the termination of the ASP in 1996. Research funded during the ASP furthered the potential of microalgal-based biofuel, particularly in molecular engineering of algae and technological development of large-scale algal production systems (Dunahay et al. 1995; Sheehan et al. 1998). However, it was evident during the ASP and is still is today, that there remains a severe knowledge gap in our comprehension of the biological mechanisms that control growth and carbon metabolism in algae. Part of this stems from the fact that the diversity of microalgae is immense and often underappreciated; with an estimated number of species ranging from 30,000 to more than 1 million (Guiry 2012). Such diversity at the taxonomical level infers diversity at the morphological, metabolic, and molecular level. This has been demonstrated in recent reviews focused on diversity of carbon metabolism and organelle

compartmentation in algae (Smith et al. 2012; Hildebrand et al. 2013). Furthermore, cost and labor associated with physiological, biochemical, and molecular research can often force scientists to narrow their focus toward model organisms. While conclusions drawn from experimentation on model systems are critical for advancing our broad understanding of biology, there are significant limitations to how those findings would relate to actual production conditions. Ultimately, characterization of metabolic processes in a promising production organism such as *C. cryptica*, may provide the explicit knowledge needed to optimize productivity and develop a large-scale renewable algal industry that is economically viable, commercially competitive, and perhaps most excitingly, help mitigate anthropogenic carbon dioxide released into the atmosphere.

Fortunately, in the last twenty years, there have been monumental advances in high-throughput technologies, especially in the cost and feasibility of genomic and transcriptomic sequencing, which allow the opportunity to obtain valuable metabolic information in a relatively short amount of time. In applied algal microbiology, this has led to the sequencing of several candidate biofuel feedstock species across different evolutionary lineages, including two other genera highlighted in the ASP, *Chlorella* and *Nannochloropsis* (Blanc et al. 2010; Radakovits et al. 2012; Vieler et al. 2012; Gao et al. 2014). There have also been developments in diatom genomic sequencing, including *Fistulifera solaris,* an oleaginous pennate diatom also with potential as a biodiesel feedstock (Tanaka et al. 2015).

The increase in genomic data on microalgae not only is advantageous toward applied research as highlighted above, but also has allowed for sophisticated comparative analysis of core metabolic processes, such as glycolysis/gluconeogenesis, nitrogen

assimilation, and fatty acid biosynthesis (Radakovits et al. 2012; Smith et al. 2012; Bender et al. 2014). Such studies help to identify conserved regulatory elements and provide a frame of reference to understand carbon and energy flux in the cell. Furthermore, an essential component of the feasibility algal-derived biofuel on a commercial scale is the improvement in productivity, which can be done either by increasing biomass and/or amount of TAG per cell. Sequencing data can help to identify regulators associated with cellular division, metabolite production, and environmental sensing (Kim et al. 2015). The contribution of these regulators to productivity can then be evaluated further, using either genetic engineering (Gong et al. 2013; Trentacoste et al. 2013; Manandhar-Shrestha and Hildebrand 2015), or mutagenesis (Cordero et al. 2011; Manandhar-Shrestha and Hildebrand 2013). Additionally, sequencing data can be used for the creation of recombinant protein constructs, produced in algae, which can open the door to an extremely high-value market product (Rasala and Mayfield 2014; Delalat et al. 2015).

Given the potential of *Cyclotella cryptica* as a true production organism, the lack of '-omic' information and clarification of metabolic processes related to productivity inhibits its further development. This thesis seeks to use high-throughput technologies to describe the cellular mechanisms surrounding carbon and energy metabolism in the industrially relevant species *C. cryptica.* Existing genomic data from the last decade allowed an in-depth comparison of these metabolic processes amongst different species of algae. From a broad perspective, these analyses will elucidate particular regulatory nodes conserved in microalgae, with focus on areas of metabolism relevant to the biofuel and bioproduct industry. In addition to using a systems biology approach, physiological

and biochemical analyses using newly developed technologies were also conducted, which remains crucial to ground truth hypotheses generated from sequencing data. Ultimately, it is my goal that the research conducted during this dissertation enables further facilitates the use of *C. cryptica* in a production system and adds to the biological understanding of the large-scale industrial cultivation of microalgae.

Chapter 1 focused on the physiological characterization of TAG accumulation in *C. cryptica* using high-throughput imaging flow cytometry. During the ASP it was demonstrated that *C. cryptica* could accumulate lipid under silicon deprivation (Roessler 1988a; Sheehan et al. 1998). The TAG accumulation was due to a concomitant increase in the proportion of newly fixed carbon partitioned into lipid as well as an increase in the conversion of other carbon pools, primarily the storage carbohydrate chrysolaminarin, into lipid (Roessler 1988b). The work conducted largely by Dr. Paul Roessler during the ASP in *C. cryptica* was foundational, but was further characterized with modern day high-throughput technologies, and an understanding of lipid accumulation as it relates to the cell cycle and other nutrient starvations. Using high-throughput imaging flow cytometry, physiological measurements of TAG, chlorophyll fluorescence, and cell cycle stage, were conducted over two distinct short and long term nutrient starvations, silicon and nitrogen. Under both nutrient starvations, *C. cryptica* accumulated substantial levels of TAG, which occurred in a multi-phase manner, with a slower rate in the early stages of nutrient depletion followed by a faster rate in the later stages. There were also differences between the two conditions with regards to onset of lipid accumulation, chlorophyll content, cell size, and intracellular heterogeneity. The cell-to-cell variability and the distinct subpopulations identified using imaging flow cytometry highlighted the

complexity that is overlooked when using bulk measurements, which could result in misconceptions of the biological system. This study was the first of its kind to use imaging flow cytometry in a microalgae species, and characterized in a high-throughput manner lipid accumulation in *C. cryptica* (Traller and Hildebrand 2013). The intriguing nuances in the physiological response to nutrient stress and TAG accumulation highlighted in chapter 1 propelled the need to investigate *C. cryptica* at a molecular level.

Chapters 2 and 3 set out to accomplish the above goal. Chapter 2 investigated the nuclear genome and methylome in *C. cryptica.* In this study, it was determined that the *C. cryptica* genome does possess additional transcribed isozymes encoding for rate limiting enzymatic steps, which could bolster the flux of carbon into triacylglycerol. These genes include the beta-glucanses, involved in carbohydrate breakdown, a chloroplast targeted pyruvate kinase, and the only committed step in triacylglycerol biosynthesis, diacylglycerol acyltransferase. Chapter 2 has the potential to propel further research with *C. cryptica* because it provides fundamental genomic data for scientists to substantiate physiological, molecular, and biochemical experiments (Traller et al. 2016). The scientific findings substantiated the role of subcellular organization in carbon flux in the cell, identified the first putative heterotrophic hexose transporter in diatoms, and provided insight on diatom diversity and evolution. Additionally, annotation of the genome in an organism that has been demonstrated as a good accumulator of TAG provided the opportunity to identity the genetic components that are common to oleaginous organisms. These identified components can then either be target sites of genetic manipulation, or further biochemically probed to modify productivity. Sequencing data of this candidate production organism is instrumental in the development of molecular constructs that can

improve productivity or in biotechnological applications such as recombinant protein production.

The third and final chapter of this thesis compared the transcriptomic profiles of cells exposed to nitrogen or silicon deprivation across early and late stages of lipid accumulation. To the best of our knowledge, this is the first documented gene expression study focused on a comparison of different macronutrient starvations in a microalgal species. Transcriptome analysis between the two conditions across a time series allowed for the discrimination of the nutrient starvation response from the lipid response to further identify key enzymes involved in TAG accumulation. Intriguingly, carbon and nitrogen metabolism, particularly in the mitochondrion, was vastly shifted under the two nutrient starvations. A significant finding from this chapter was the upregulation of mitochondrial beta-oxidation transcripts in the late stages of silicon starvation, indicating a concomitant increase in fatty acid breakdown and triacylglycerol, which suggests a preference shift in carbon storage from chrysolaminarin to lipid in the early to late stages of starvation, respectively. Further elucidation on the nuances in gene expression between the early and late stages of nutrient starvation could allow for testable hypotheses to explain the different rates of TAG accumulation identified in chapter 1.

Overall, this thesis provided critical genomic datasets as well as previously undocumented phenomena associated with TAG accumulation, carbon metabolism, and regulatory components in a candidate organism for commercial production. The data not only is applicable toward the development of *Cyclotella cryptica* for large-scale industrial cultivation, but also will allow for significant advances in our understanding of diatom metabolism and physiology. The advent of affordable high throughput genomic

sequencing allowed for the opportunity to obtain data in a non-model system that could not previously be possible. The findings described in this thesis demonstrate a wealth of information and significant insight into the molecular underpinnings of triacylglycerol accumulation, which hopefully will bring us closer to the feasibility of commercial algal cultivation as a means to effectively mitigate anthropogenic climate change and contribute to the solution for our energy needs.

REFERENCES FOR THE INTRODUCTION

Bender SJ, Durkin CA, Berthiaume CT, Morales RL, Armbrust EV (2014) Transcriptional responses of three model diatoms to nitrate limitation of growth. Frontiers in Marine Science 1:1-15.

Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, Kuo A, Lindquist E, Lucas S, Pangilinan J, Polle J, Salamov A, Terry A, Yamada T, Dunigan DD, Grigoriev IV, Claverie J-M, Van Etten JL (2010) The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. Plant Cell 22:2943–2955.

Change IPOC (2014) Climate Change 2014–Impacts, Adaptation and Vulnerability: Regional Aspects.

Cordero BF, Obraztsova I, Couso I, Leon R, Vargas MA, Rodriguez H (2011) Enhancement of Lutein Production in Chlorella sorokiniana (Chorophyta) by Improvement of Culture Conditions and Random Mutagenesis. Marine Drugs 2013, Vol 11, Pages 4662-4697 9:1607–1624.

Delalat B, Sheppard VC, Ghaemi SR, Rao S, Prestidge CA, McPhee G, Rogers M-L, Donoghue JF, Pillay V, Johns TG, ger NKO, Voelcker NH (2015) Targeted drug delivery using genetically engineered diatom biosilica. Nature Communications 6:1–11.

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

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJR (2004) The Evolution of Modern Eukaryotic Phytoplankton. Science 305:354–360.

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science 281:237–240.

Gao C, Wang Y, Shen Y, Yan D, He X, Dai J, Wu Q (2014) Oil accumulation mechanisms of the oleaginous microalga *Chlorella protothecoides* revealed through its genome, transcriptomes, and proteomes. BMC Genomics 15:582.

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

Gong Y, Zhang J, Guo X, Wan X, Liang Z, Hu CJ, Jiang M (2013) Identification and characterization of PtDGAT2B, an acyltransferase of the DGAT2 acyl-Coenzyme A: Diacylglycerol acyltransferase family in the diatom *Phaeodactylum tricornutum*. FEBS Letters 587:481–487.

Guiry MD (2012) How many species of algae are there? Phycol 48:1057–1063.

Hildebrand M, Abbriano RM, Polle JE, 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 17:506–514.

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

Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. Plant J 54:621–639.

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.

Kobayashi M, Kakizono T, Nagai S (1991) Astaxanthin production by a green alga, *Haematococcus pluvialis* accompanied with morphological changes in acetate media. Journal of Fermentation and Bioengineering 71:335–339.

Laws EA, Taguchi S, Hirata J, Pang L (1988) Optimization of microalgal production in a shallow outdoor flume. Biotechnol Bioeng 32:140–147.

Lewin JC, Lewin RA (1960) Auxotrophy and heterotrophy in marine littoral diatoms. Canadian Journal of Microbiology. 6:127-137.

Lum KK, Kim J, Lei XG (2013) Dual potential of microalgae as a sustainable biofuel feedstock and animal feed. Journal of Animal Science and Biotechnology 2013 4:1 4:53.

Manandhar-Shrestha K, Hildebrand M (2015) Characterization and manipulation of a DGAT2 from the diatom *Thalassiosira pseudonana*: Improved TAG accumulation without detriment to growth, and implications for chloroplast TAG accumulation. Algal Research 12:239–248.

Manandhar-Shrestha K, Hildebrand M (2013) Development of flow cytometric procedures for the efficient isolation of improved lipid accumulation mutants in a *Chlorella sp*. microalga. J Appl Phycol 25:1643–1651.

Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, Boore JL, Posewitz MC (2012) Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. Nature Communications 3:686–.

Ramanathan V, Carmichael G (2008) Global and regional climate changes due to black carbon. Nature Geoscience 1:221–227.

Rasala BA, Mayfield SP (2014) Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. Photosynth Res 123:227–239.

Reimann BEF, Lewin JMC, Guillard RRL (1963) *Cyclotella cryptica*, a New Brackish-Water Diatom Species 1. Phycologia 3:75–84.

Roessler PG (1988a) Effects of silicon deficiency on lipid composition and metabolism in the diatom *Cyclotella cryptica.* J Phycol 24:394–400.

Roessler PG (1988b) Changes in the activities of various lipid and carbohydrate biosynthetic enzymes in the diatom *Cyclotella cryptica* in response to silicon deficiency. Archives of Biochemistry and Biophysics 267:521–528.

Sheehan J, Dunahay T, Benemann J, Roessler P (1998) A look back at the US Department of Energy's Aquatic Species Program: Biodiesel from algae.

Slocombe SP, Zhang Q, Ross M, Anderson A, Thomas NJ, Lapresa Á, Rad-Menéndez C, Campbell CN, Black KD, Stanley MS, Day JG (2015) Unlocking nature's treasure-chest: screening for oleaginous algae. Sci Rep 5:9844 EP —17.

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.

Solomon S, Daniel JS, Sanford TJ, Murphy DM, Plattner G-K, Knutti R, Friedlingstein P (2010) Persistence of climate changes due to a range of greenhouse gases. Proceedings of the National Academy of Sciences 107:18354–18359.

Tanaka T, Maeda Y, Veluchamy A, Tanaka M, Abida H, Maréchal E, Bowler C, Muto M, Sunaga Y, Tanaka M, Yoshino T, Taniguchi T, Fukuda Y, Nemoto M, Matsumoto M, Wong PS, Aburatani S, Fujibuchi W (2015) Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. Plant Cell 27:162–176.

Traller JC, Cokus SJ, Lopez DA, Gaidarenko O, Smith SR, McCrow JP, Gallaher SD, Podell S, Thompson M, Cook O, Morselli M, Jaroszewicz A, Allen EE, Allen AE, Merchant SS, Pellegrini M, Hildebrand M (2016) Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. Biotechnol Biofuels 9:258.

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.

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 110:19748–19753.

Vieler A, Wu G, Tsai C-H, Bullard B, Cornish AJ, Harvey C, Reca I-B, Thornburg C, Achawanantakun R, Buehl CJ, Campbell MS, Cavalier D, Childs KL, Clark TJ, Deshpande R, Erickson E, Ferguson AA, Handee W, Kong Q, Li X, Liu B, Lundback S, Peng C, Roston RL, Sanjaya, Simpson JP, TerBush A, Warakanont J, Zäuner S, Farre EM, Hegg EL, Jiang N, Kuo M-H, Lu Y, Niyogi KK, Ohlrogge J, Osteryoung KW, Shachar-Hill Y, Sears BB, Sun Y, Takahashi H, Yandell M, Shiu S-H, Benning C (2012) Genome, Functional Gene Annotation, and Nuclear Transformation of the Heterokont Oleaginous Alga *Nannochloropsis oceanica* CCMP1779. PLoS Genet 8:e1003064.

Yaakob Z, Ali E, Zainal A, Mohamad M, Takriff MS (2014) An overview: biomolecules from microalgae for animal feed and aquaculture. J Biol Res (Thessalon) 21:6.

CHAPTER 1

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 (2013) 244–252

Contents lists available at SciVerse ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

High throughput imaging to the diatom *Cyclotella cryptica* demonstrates substantial cell-to-cell variability in the rate and extent of triacylglycerol accumulation

Jesse C. Traller, Mark Hildebrand *

Marine Biology Research Division, Scripps Institution of Oceanography, UCSD, La Jolla, CA, United States

ARTICLE INFO

Article history:
Received 19 October 2012
Received in revised form 15 March 2013
Accepted 22 March 2013
Available online 23 April 2013

Keywords:
Diatom
Imaging flow cytometry
Triacylglycerol
Nutrient starvation
Cell-to-cell variability
BODIPY

ABSTRACT

In microalgal cultures, most analyses of cellular processes are done in bulk, on the entire population of cells. Information gained from this is representative of the mean; however, it obscures the richness of cell-to-cell variation, which is a well-documented phenomenon. Using imaging flow cytometry, we evaluate changes in triacylglycerol (TAG) content and chlorophyll resulting from silicon or nitrogen deprivation in the diatom *Cyclotella cryptica*. This approach allows detailed interrogation of large numbers of individual cells and reveals cell-to-cell variation. This study demonstrates several previously undescribed phenomena related to TAG accumulation in microalgae. First, the rate of TAG accumulation varies over time, with a faster rate occurring at the latter stage of the process, resulting in hyperaccumulation in which the majority of the cell volume is comprised of lipid droplets. In *C. cryptica* and other diatoms this hyperaccumulation occurs strictly under autotrophic conditions. Second, there are distinct responses to silicon or nitrogen limitation, and variation within a given type of limitation treatment. Under most conditions there is a large spread in the population when measuring either chlorophyll or TAG. Heterogeneity within the total population indicates that caution should be taken in interpreting bulk measurements for a variety of variables (TAG, transcript, protein, metabolites, etc.) related to cellular responses. However, a potential means to couple subpopulation-level responses with bulk analysis approaches is described, which could take advantage of the nuances observed during the TAG accumulation process.

1. Introduction

The development of microalgal biofuels is gaining momentum, as economic and environmental factors drive the need for renewable and sustainable sources of liquid transportation fuels. Many algal species synthesize abundant amounts of neutral lipids suitable for conversion to high energy density liquid fuels, producing 30–40% dry weight in lipids, with exceptional species reaching levels of 50–70% [1–4]. A number of analyses indicate that unicellular algae produce more biomass and oil than plants by a large margin [5]. Microalgae can also be grown on non-arable land in non-potable, saline water, reducing competition for resources for human needs. Currently lacking, however, is the infrastructure to efficiently grow algae on a large scale, and at the cellular level, an understanding of the basic mechanisms underlying controllable and reproducible lipid accumulation. In terms of the economics of production, the research focus has largely been to increase the amount of fuel precursor molecule; however, it may not be appreciated that the *rate* at which the molecule accumulates will also have a strong impact on the cost.

An ideal production organism will produce abundant neutral lipid or triacylglycerol (TAG) yet still rapidly grow to high biomass. In the natural environment, diatoms (Bacillariophyceae) often outcompete other classes of algae for nutrients and growth and are among the most productive and environmentally-flexible eukaryotic microalgae on the planet [6]. Extensive screening in the Aquatic Species Program (ASP) identified diatoms as some of the best candidate organisms for lipid-based biofuels production [3]. The centric diatom *Cyclotella cryptica* was defined as a potential model organism with high lipid accumulation ability [7,8], and with consistent productivity levels of 20 $g \cdot m^{-2} \cdot d^{-1}$ in outdoor ponds [9].

One distinguishing feature of diatoms is their cell wall, which is a composite of silica and organic material [10–12]. Most diatoms have an absolute requirement for silicon (Si) in order to grow [13], and require it to the same extent as nitrogen [14]. Depletion of silicon in diatom cultures leads to cessation of cell division and consequentially accumulation of organic carbon in the form of TAG in some species [4,8,15]. Silicon is not directly involved in many aspects of cellular metabolism [16] and intracellular stores of Si are low [13], which is in contrast to nitrogen, a commonly used limiting nutrient for TAG accumulation in microalgae [3,17,18]. Nitrogen metabolism is tightly

* Corresponding author at: 9500 Gilman Dr., La Jolla, CA 92093-0202, United States. Tel.: +1 828 822 0167.
E-mail address: mhildebrand@ucsd.edu (M. Hildebrand).

2211-9264/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.algal.2013.03.003

correlated with central carbon metabolism and is required in numerous processes essential to cell growth; including photosynthesis, DNA replication, and protein synthesis. A comparison between silicon and nitrogen starvation may elucidate distinct regulatory processes controlling TAG accumulation and enable identification of the core essential processes distinct from the secondary effects of nutrient stress. An understanding of these processes is a crucial step toward increasing productivity.

One method taken to understand the cell's behavior during TAG formation has been to look at the significant changes in expression of genes and proteins. Transcriptomic and proteomic studies have helped advance the field by seeking to provide a comprehensive view of the basic metabolic pathways involved [18–20]. Recent analysis revealed that there are substantial differences in fundamental carbon metabolism pathways among diatom species and between diatoms and other classes of algae [21], suggesting that generalizations should be avoided. At an even more fundamental level, the basic process of TAG accumulation has been under-investigated. We are not aware of studies that document what rates of TAG accumulation are possible in an algal cell, or studies that show whether different rates occur under different conditions. Baseline measurements during the entire process of TAG accumulation will be essential toward estimating how much improvement is needed in order to implement a successful production system. Methods have been established to generate such data, including gravimetric analysis to obtain total lipid content, and spectrofluorometry using fluorescent lipid probes Nile Red or BODIPY [3,22–25]. Fluorescent methods have consistently been shown in diverse classes of microalgae (including diatoms) to provide an accurate representation of TAG content compared with gravimetric methods ([26–28], Trentacoste et al., manuscript in prep.).

While the previously mentioned analyses are crucial, data generated from gravimetric, spectrofluorometric, transcriptomic and proteomic approaches is averaged over the entire cell population, and researchers must take caution when drawing conclusions about what occurs or what is possible in an individual cell when the data is in reality averaged over millions of cells per mL. A large shift of a particular parameter in a small percentage of the population could inaccurately skew the perception of a response. For example, if 10% of the population shows a 25-fold induction of a particular transcript and the remaining population shows no induction, a change on the order of 2.5-fold will be registered when averaged over the entire population, but in reality, no change has occurred in 90% of the cells. Similar analogies can be made for gravimetric and spectrofluorometric analyses. Examples of heterogeneity between different cells in a population have recently come to the forefront with the advent of single cell analysis; indicating that intrapopulation variability appears to be common throughout all branches of life, including microalgae [29–32]. Aside from issues stemming from cellular variability, many 'omic' studies compare only two conditions: a 'before' or uninduced sample, and an 'after' or induced sample, and include multiple replicates in an effort to constrain a typical response. This type of experimental design is informative but only to a certain extent because it is possible that transient metabolic shifts will be missed due to insufficient coverage of the process, or that genuine variability could occur but be discounted because it is less frequent. Such metabolic shifts could be due to epigenetic or other factors. Data collected in a time series (experiments which span the entire accumulation of lipid versus just the induction of such) will create a more extensive picture of metabolic shifts occurring in the cell, and provide a better opportunity to observe interesting transient phenomenon.

An ideal situation would be to apply approaches in which a global bulk analysis will still effectively capture the ongoing dynamic and sometimes variable processes within individual cells. In this study, we characterize the TAG accumulation process in the diatom *C. cryptica*, using bulk culture analyses combined with the application of imaging flow cytometry to enable evaluation of cell-to-cell variation. A statistically sound, comprehensive view of the TAG accumulation process resulted, which allowed for bulk analyses to be related to both subpopulations and individual cells within the same sample. We also compare the variability in response between nitrogen and silicon limitation. The data reveal significant variability and substantially different rates of TAG accumulation in different subpopulations of cells, but suggest that correlations can still be made with bulk sampling procedures. Rates are substantially higher at some stages than others, which could have positive implications for improving strains for production.

2. Materials and methods

2.1. Culture conditions

50 mL working stocks of clonal *C. cryptica* (CCMP 332, National Center for Marine Algae and Microbiota; http://ccmp.bigelow.org/) were maintained under axenic conditions in artificial seawater (ASW) medium [33] under a light:dark cycle of 12:12 at 18 °C and 150 μmol·m^{-2}·s^{-1} light. Cells to be used for experiments were transferred to fresh ASW and grown under continuous light to mid-exponential phase (approximately 2.5 × 10^5–5.5 × 10^5 cells·mL^{-1}) and placed on a rotating shaker before transferring to 2.5 L ASW.

2.2. Nutrient limitation experiments

Once the cells were inoculated into 2.5 L cultures, they were bubbled with air and mixed using a magnetic stir plate at a speed of 250 rpm in continuous light. For nutrient comparison experiments, cells were grown to mid-exponential phase in nutrient replete ASW (for approximately 4 days). To transfer cells into nutrient deplete conditions, cells were harvested via centrifugation in 250 mL sterile centrifuge bottles for 15 min at 3100 ×*g*. The concentrated cells were washed once with ASW lacking silicon or nitrogen, depending on the condition of starvation. The washed cells were then placed in 1 L nutrient deplete ASW (Si-ASW or N-ASW) and monitored on a daily basis. For consistency, both nutrient limitation experiments were conducted in 2 L polycarbonate bottles to prevent silicon contamination from glass vessels.

2.3. Sampling

Quadruplicate samples (10 mL each) were taken from cultures for cell counts and flow cytometry. Cells were concentrated by centrifugation for 6 min at 3200 ×*g*. Cells were transferred to microfuge tubes, pelleted, the supernatant was removed and the cell pellet kept frozen at −20 °C until analysis. Cells were analyzed within 2 weeks of sampling to avoid cellular damage from long term storage.

Cells were counted in duplicate or quadruplicate using a Neubauer hemocytometer to generate data for the growth curves.

2.4. Fluorescent staining

Cells monitored for lipid were stained using the fluorescent dye 4, 4-difluoro-4-bora-3a, 4a-diaza-s-indacene (BODIPY 493/503, Molecular Probes). Cells were resuspended in 500 μl of 2.3% NaCl and then stained by adding 1.3 μL of a 1 mg ml^{-1} BODIPY stock dissolved in DMSO and incubated in the dark on ice for a minimum of 10 min. Cells were processed within 2 h of staining while the dye characteristics remained stable. For the purpose of this study, we chose the fluorescent dye BODIPY as a proxy for lipid content in cells. It has been found that BODIPY is more photostable than Nile Red, and its narrow green emission spectrum allows for simple comparison to chlorophyll autofluorescence without spill over [24,25].

2.5. Microscopy

Live cells were stained with BODIPY and observed with a Zeiss Axio Observer Z1 inverted microscope equipped with a Zeiss ApoTome and a Zeiss AxioCam MRm camera. Non-fluorescent images were taken using differential interference contrast (DIC). The filter sets used for fluorescent imaging were Zeiss #16: Excitation BP 485/20 nm, Dichromatic mirror FT 510 nm, Emission LP 515 nm for chlorophyll autofluorescence, and Zeiss #38HE Excitation BP 470/40 nm, Dichromatic mirror FT 495 nm, Emission BP 525/50 nm for BODIPY. Images shown in Fig. 2 were captured using the 63× objective with an oil immersion. Images shown in Fig. S3 were captured using the 40× objective.

2.6. Cell cycle analysis

10 mL of live cells were concentrated down and resuspended in 10 mL100% ice-cold methanol and stored at 4 °C until analysis. Methanol was removed and cells were washed three times in Tris–EDTA (TE) buffer (pH 8). After the third wash cells were resuspended in 1 mL TE and treated with Ribonuclease A (0.3 mg ml^{-1} final concentration, Sigma). Samples were then incubated at 37 °C for 45 min. After incubation, 10 μL of 100× SYBR Green I nucleic acid gel stain (10,000× concentrate in DMSO, Molecular Probes) stock dissolved in DMSO was added. Cells were stained for a minimum of 10 min in the dark on ice. Cells were analyzed using the Becton Dickinson In-flux sorting cytometer (BD Biosciences, San Jose, CA). Data were analyzed using FlowJo cytometry software (Tree Star Inc., Ashland, OR).

2.7. Imaging flow cytometer analysis

Frozen cells were processed using the protocol mentioned above for analysis on the ImageStreamX (Amnis Corp., Seattle, WA) imaging flow cytometer. At least 10,000 events were collected for each time point using the INSPIRE software package (Amnis Corp.). To minimize collection of data on debris, cell classifier parameters were set to capture images with a minimum and maximum area of 20 μm^2 and 200 μm^2 respectively. The side scatter image was collected in channel 6 (Ex 785 nm, Em 745–800 nm), BODIPY fluorescence in channel 2 (Ex 488 nm, Em 470–560 nm), chlorophyll autofluorescence in channel 5 (Ex 488 nm, Em 660–720 nm), and bright field in channel 4. Because chlorophyll autofluorescence oversaturates at the lowest 488 nm laser power, two neutral density filters (0.6 and 1.0) were added to block a portion of the light from the laser. The 488 nm laser was set at an intensity of 25 mW and the side scatter and bright field 785 nm laser to 2.85 mW.

Post-acquisition spectral compensation and data analysis were performed using the IDEAS software package provided by Amnis Corp. To correct for crossover fluorescence of each fluorochrome into other channels, a compensation matrix was created as described in [34]. Single, in-focus cells were enriched by using the guided analysis features 'focus' and 'single cell' in IDEAS. After elimination of debris, 5000–10,000 cells were analyzed per sample.

3. Results

3.1. Time course experiments demonstrate distinct differences between the early and late stages of TAG formation and between silicon and nitrogen deplete cell cultures

Exponentially growing, nutrient replete cultures were harvested by centrifugation and transferred into nutrient deplete (Si or N) ASW (see Materials and methods, Sections 2.1–2.2). After cells were transferred to Si-ASW, growth immediately ceased (Fig. 1a). It has been reported in the diatom *Cylindrotheca fusiformis* that DNA synthesis and cell division immediately ceases in Si-starved cultures [33]. The Si-starvation response presented here was comparable to other studies with *C. cryptica* as well as other diatom species ([35] and Traller, unpublished data) — highlighting the robust nature of this response. In both experiments, cell cycle arrest was predominantly in the G2 phase (Fig. S1). During Si-starvation, chlorophyll concentration per cell increased continually (Fig. 1b), which is a physiological response to Si deprivation (Glé and Smith, unpublished results and [36]). In terms of TAG accumulation as measured by BODIPY fluorescence, there was an initial phase of lower TAG accumulation rate, followed by a rapid rate phase, then followed by a lower rate as the process appeared to plateau (Fig. 1c). The initial phase spanned 0–48 h. The faster rate occurred at approximately 48–72 h, and saturated from 72 to 96 h. The pattern of accumulation was similar in both experiments, but the extent of fold change relative to 0 h was different (Fig. 1c).

In contrast to Si-limitation, nitrogen deprivation produced a more variable response in growth and chlorophyll levels, but a similar response in terms of TAG accumulation (Fig. 1d–f). In the replicate N-ASW cultures in Fig. 1d, in one case (N1) growth ceased within 24 h, but in the other (N2), cells continued to grow for 2–3 days. The cell cycle arrest points differed in these two experiments — in N1 arrest was predominantly in the G2 phase of the cell cycle and in N2, the G1 phase (Fig. S1). In both experiments, chlorophyll levels decreased over time (Fig. 1e), however this was slightly delayed in N1 (Fig. 1e). Although the amount of BODIPY fluorescence differed in the two experiments the TAG accumulation pattern was similar (Fig. 1f). Both patterns exhibited an increasing rate until 48 h, then either a steady rate or decrease, followed by the highest rate.

Overall, these experiments showed a general trend of TAG accumulation in *C. cryptica* in three phases: an early, slower phase, a later, rapid phase, and a plateau phase (the latter specific for Si limitation, Figs. 1c, f, S2). The average increase in rate between the first and second phases was 3.5 fold and 3 fold for Si and N, respectively. The substantial difference in rate between the early and late stages of nutrient depletion suggests differences in metabolic processes involved in TAG accumulation.

3.2. Fluorescent microscopy of cells during TAG accumulation demonstrate a hyperaccumulation of lipid droplets

Fluorescent and differential interference contrast (DIC) images were captured of cells throughout the time course of Si and N limitation (Fig. 2). Especially evident at the later time points is an abundant amount of TAG in the form of large lipid droplets. Cells in both Si1 and N1 experiments at 96 h contain large droplets that comprise the majority of the cell volume (Fig. 2e, j). In some cases the lipid bodies became so large that they escaped from the cell (Fig. S3). Because of the increased rate of TAG accumulation during the later time points (Fig. 1c and f) and the large volume of lipid droplets in the cell (Figs. 2, S3), we defined this later stage as *hyperaccumulation*.

3.3. Characterization of TAG accumulation in C. cryptica using imaging flow cytometry

3.3.1. Imaging flow cytometry analysis reveals subpopulation level differences within the total population

It is difficult to obtain rigorous statistical analyses using standard microscopy approaches, due to limitations in sampling size. We analyzed the TAG accumulation process in a more statistically robust manner by conducting high-throughput imaging analysis of cells using an ImageStreamX imaging flow cytometer. This instrument has the capability of performing scatter-plot analyses similar to traditional flow cytometers as well as high resolution microscopy of each cell, enabling the interrogation of a large number of individual cells for TAG and chlorophyll content.

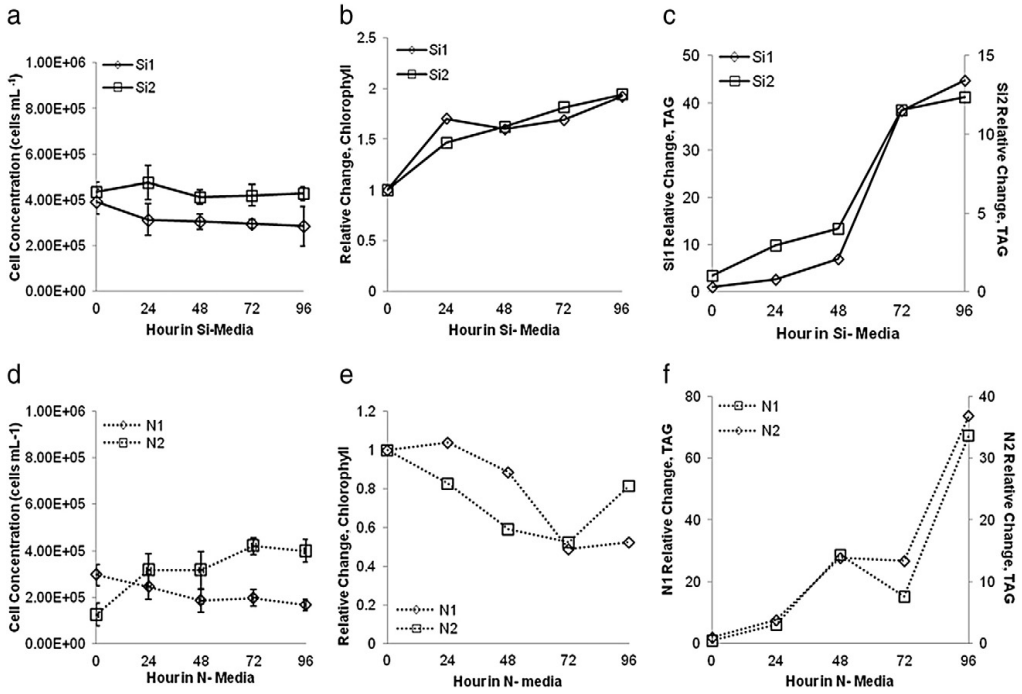


Fig. 1. Cell count and fold change in chlorophyll and BODIPY fluorescence over time in nutrient deplete media (Si-media, solid lines; N-media, dotted lines). (a, d) Cells transferred from nutrient replete ASW to nutrient deplete ASW exhibit cessation of growth, except in the case of N2, where growth continued for approximately 2 days. Error bars indicate standard deviation. (b, e) Relative changes in chlorophyll fluorescence relative to t = 0. (c, f) Relative change in BODIPY fluorescence over time in Si-media (c) or N-media (f). X-axis for all graphs is the time (in hours) the cells are placed in nutrient deplete ASW. Data generated are from two independent biological replicates.

Using the ImageStream, six subpopulations within the total sample population were identified (Fig. 3). Each subpopulation was defined by "gates" that were generated using the fluorescence intensity values of chlorophyll and TAG (using BODIPY) on a log scale and the morphological changes observed in the images (Fig. 3c). The gates enabled assessment of distinct shifts in neutral lipid and chlorophyll content and appearance.

The first two subpopulations, 'zero lipid (ZL)' and 'diffuse staining (DS)', were comprised of cells containing little or no TAG. Cells categorized in the ZL subgroup (Fig. 3a, b, magenta gate) had high chlorophyll fluorescence and no visible BODIPY fluorescence. The second subpopulation, DS, contained cells with a small amount of BODIPY fluorescence, which appeared to be initial TAG formation based on microscopy (Figs. 2a, f, 3c). Cultures grown under exponential,

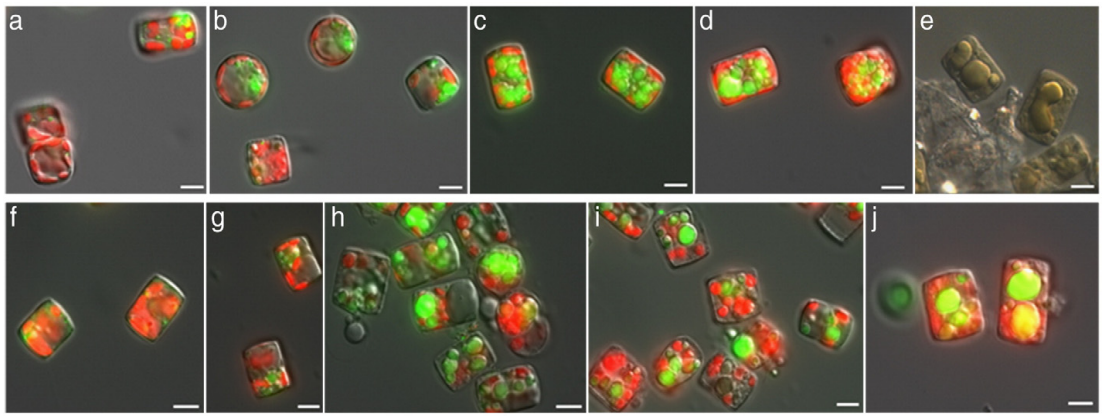


Fig. 2. The accumulation of lipid in *C. cryptica* over time as based on high resolution microscopy. Images are composites of BODIPY fluorescence (falsely colored green) and chlorophyll autofluorescence (falsely colored red), or are solely differential interference contrast (DIC). (a–e) Cells in silicon deplete medium (Si1 experiment) at hours 0, 24, 48, 72, 96 respectively. (f–j) Cells in nitrogen deplete medium (N1 experiment) at hours 0, 24, 28, 72, 96. (a–d, f–j). Merged images of BODIPY, chlorophyll, and DIC. (e) DIC image of silicon deprived cells at 96 h; demonstrating that droplets can be visualized without a lipid probe in the later stages. Scale bars 5 μm.

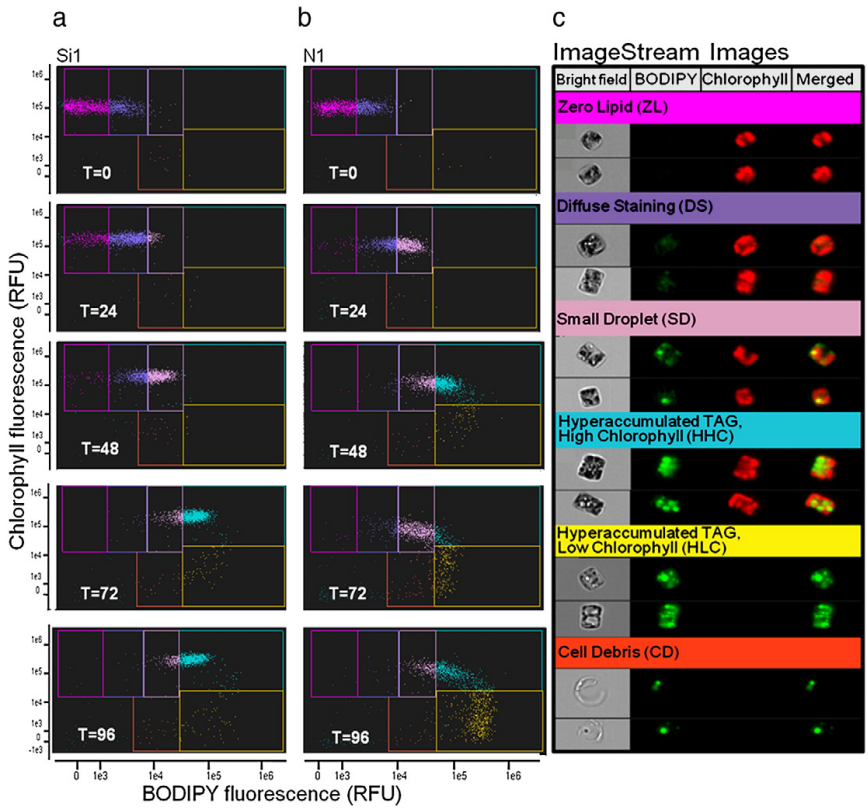


Fig. 3. Imaging cytometry analysis reveals six subpopulations from silicon (a) and nitrogen (b) starvation experiments. (a) Silicon starvation over 0, 24, 48, 72, and 96 h; (b) Nitrogen starvation over 0, 24, 48, 72, 96 h. (c) Representative cells, two for each subpopulation, color of heading corresponds to a particular gated subpopulation shown in the cytograms. Each cell shown has four images: bright field, BODIPY fluorescence (falsely colored green), chlorophyll autofluorescence (falsely colored red), and a merged image of BODIPY and chlorophyll. The cells depicted here are the mean for the subpopulation.

nutrient replete conditions exhibited high percentages of cells in these two subgroups; 98% and 99% for the selected silicon and nitrogen experiments respectively (Fig. 3a, b). At t = 24 and t = 48 (Fig. 3) another subpopulation was evident in cells that contained 'small droplets' and was named accordingly (SD, third subpopulation, lavender).

The next subpopulation (blue gate) was 'hyperaccumulated TAG with high chlorophyll (HHC)' in which large lipid droplets were visible in cells still containing substantial chlorophyll. The fifth subpopulation (yellow gate) was 'hyperaccumulated TAG with low chlorophyll (HLC)' in which large lipid droplets were present but chlorophyll was dramatically decreased or absent. These subpopulations became predominant in the later stages of nutrient depletion, developing at approximately 48 h into the starvation response. The final subpopulation detected was defined as 'cell debris (CD)' (orange gate). Cells falling into this subgroup were presumed dead, as denoted by empty silica frustules with residual lipid (Fig. 3c).

3.3.2. The dynamics of TAG accumulation revealed by imaging cytometry

A template defining the six subpopulations of TAG accumulation in *C. cryptica* defined was applied to scatter plots of time course experiments as a method of determining shifts within the total population of cells over time (Fig. 3). Fluctuations in the lipid and chlorophyll content of the total population can be more readily visualized by an animation comparing the different scatter-plots (Fig. S4). Following the time course of TAG accumulation, we see a distinct shift in subpopulations ranging from the initial ZL and DS populations, followed by SD, then HHC, HLC, and CD (Figs. 3, S4). Observations from these and many other experiments suggest that this general progression is typical not only of diatoms, but other classes of microalgae (Abbriano, Sanchez-Alvarez, Smith, Trentacoste, Hildebrand, unpublished data).

The gating approach allows evaluation of subpopulation transitions between distinct stages in the TAG accumulation process. In Fig. 4, we compare changes in the percentage of cells in six subpopulations with the mean TAG and chlorophyll fluorescence over time in the Si1 and N1 limitation experiments shown in Fig. 3. Fig. 4b shows the progression of the mean (BODIPY and chlorophyll) for the entire culture for Si-as previously shown (Fig. 1c). At the start of the experiment, the predominant subpopulations were ZL and DS, which subsequently decreased over time (Fig. 4a). By 48 h in Si-ASW, SD became the predominant subpopulation and by 72 h HHC dominated (Fig. 4a). For N-cells, the subpopulation shift occurred 24 h earlier than Si (Fig. 4c). There was a correlation between the predominance of hyperaccumulated cells (cells belonging in either HHC or HLC subgroups) and the increased rate of TAG accumulation, indicating that these cells not only contained more TAG, but they accumulated it at a faster rate. Under nitrogen limitation, ZL and DS populations decreased and there was a correlation between increased TAG accumulation rate and the predominance of the HHC population; similar to Si- cultures (Fig. 4c and d). This was further evident at T = 72 in N1 where there was a loss in both TAG and chlorophyll fluorescence

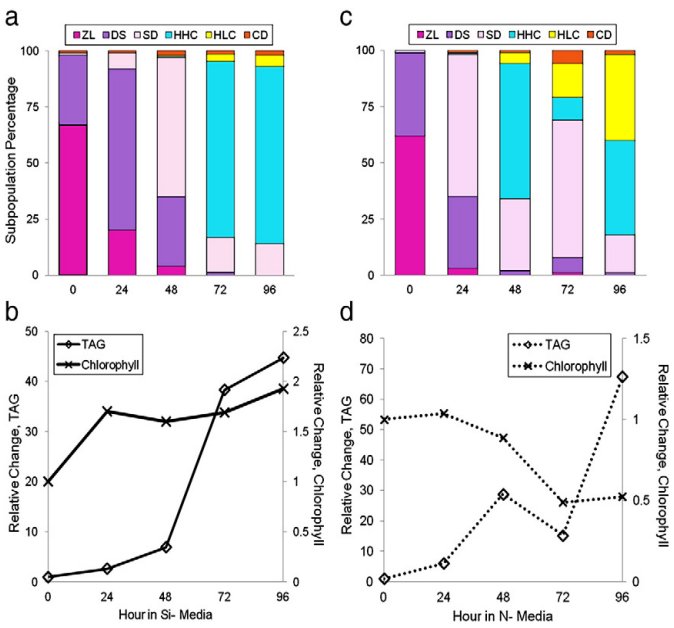


Fig. 4. Relative distribution of cells within the different subpopulations plays a role in the rate of TAG accumulation. (a, c) Stacked bar graphs indicate the relative percentage of cells in each subpopulation. Colors of each bar correspond to the color of each gate as shown in Fig. 3; (a) silicon, Si1; (c) nitrogen, N1; (b) fold change in TAG as quantified by BODIPY and fold change in chlorophyll autofluorescence over silicon depletion and (d) nitrogen depletion. Changes in the rate of TAG accumulation correspond to shifts into the hyperaccumulated lipid with high chlorophyll subpopulation (HHC, blue bar). Due to the loss in chlorophyll, N-cells also experienced an increase in the hyperaccumulated lipid, low chlorophyll subpopulation (HLC, yellow bar).

and subsequent decrease in HHC; the rate then increased dramatically from 72 to 96 h, consistent with an increase in the HHC population (Fig. 4c, d, and animation shown in Fig. S4). A distinction between N and Si is that the HLC subpopulation becomes substantial in N, where it does not in Si within the hours observed. This is a manifestation of the consistent loss of chlorophyll under N limitation (Fig. 1e).

3.4. Intrapopulation heterogeneity

It is evident from the plots in Fig. 3 that there is a spread in the population(s) with regard to TAG content at any given time. In some cases the spread is quite substantial; consider N limitation at 96 h, whereas in other cases the population is tight, such as N limitation at 24 h (Fig. 3b). Control experiments and repeated observations, combined with the ability to observe lipid droplets without BODIPY staining (Fig. 2e, Fig. S3), as well as kinetic data in other classes of algae with less permeable cell walls than diatoms [37], indicate that these differences are not due to staining artifacts resulting from differences in cell permeability.

The ability to image large numbers of cells is essential to monitor population heterogeneity. In Fig. 5 we show the effect of varying the number of cells imaged on the quality of the data. Different numbers of randomly selected cells were included in the analysis of chlorophyll and BODIPY fluorescence during N limitation. Both analyses showed that an accurate representation of the actual distribution was not apparent until at least 450 cells were included (Fig. 5). The means also did not approach their accurate values until this point. The mean chlorophyll fluorescence decreased with more cells, but increased for BODIPY — a manifestation of the misrepresentation of the population with a smaller sample size. We also observed a decrease in the standard error with increasing number of analyzed cells, providing an accurate representation of the mean (Fig. 5), in spite of relatively large spreads in individual values for cells in the population. The benefits of interrogating large numbers of cells during a time course are shown in Fig. 6, where we plot histograms of BODIPY and chlorophyll fluorescence at three time points during the N1 limitation experiment. The designation of two subpopulations (ZL and DS) differing in TAG content at t = 0, which were defined by visual observation of the cells, is confirmed in the histogram. By 48 h, the populations merged, but by 96 h, two subpopulations are again evident. For chlorophyll, we observe one predominant subpopulation until 48 h, and then a substantial amount of chlorophyll degradation occurs, generating a new subpopulation and heterogeneity in the total culture (Fig. 6). Detection of such heterogeneity would be extremely challenging in a bulk measurement.

4. Discussion

The application of imaging flow cytometry to a time course analysis documented several dynamic and variable aspects of the TAG accumulation process resulting from silicon or nitrogen limitation in the diatom *C. cryptica*. Cell-to-cell variability in a clonal population of cells is a well-established phenomenon [29–32], but can be overlooked when applying bulk analysis approaches. A bulk measurement deals only with the mean of a population, but interrogation of large numbers of individual cells enables finer resolution of the diversity of processes that constitute the mean. Traditional flow cytometry can distinguish subpopulations, but imaging flow cytometry provides additional information in the context of monitoring a visible change in the cells such as the accumulation of lipid droplets. The variability documented in this study emphasizes the importance of understanding the individual cell's response to TAG accumulation as opposed to the total population. It is unlikely that gene and enzyme activity in a cell that contains large levels of TAG is the same as that of a cell with little TAG. While bulk analyses of lipid are important for productivity

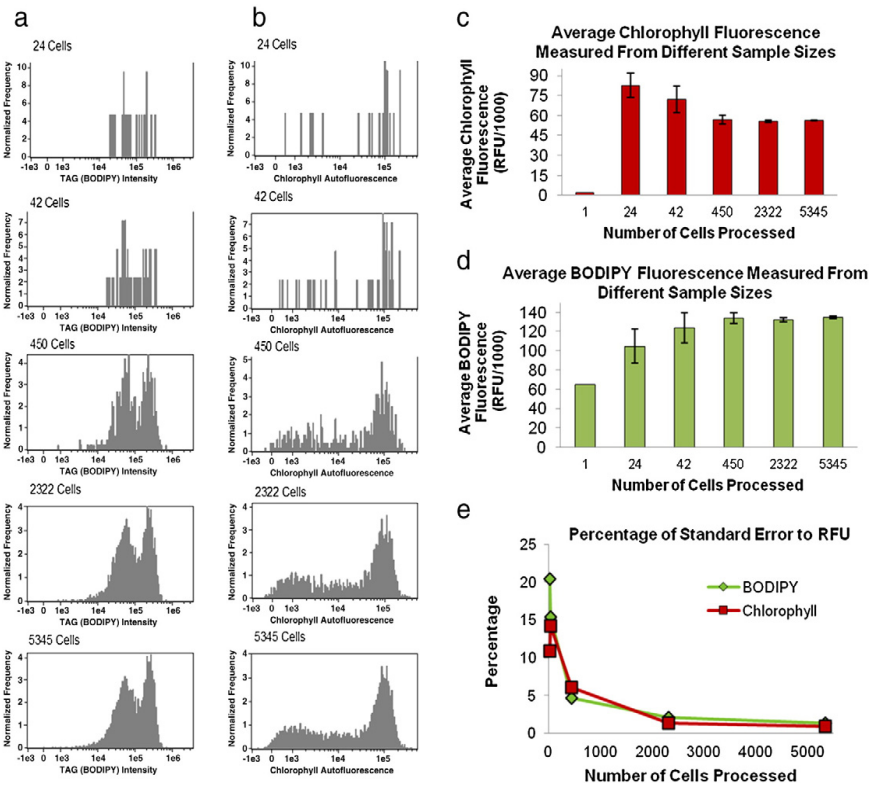


Fig. 5. Sample size number plays role in accuracy of imaging cytometry data. To demonstrate discrepancies and issues that arise when using microscopy or averaging over total populations, the IDEAS software template was applied to randomly chosen cells from sample sizes of 1, 24, 42, 450, 2322, and 5345 cells. The day four N-ASW (N1) BODIPY and chlorophyll histograms were plotted in a, b respectively. It was not until 450 cells were processed did the pattern emerge that was similar to the highest sample size. When comparing the n = 5345 to the n = 24, there is a significant difference in both the TAG and chlorophyll fluorescence; this indicates that there is large variation on the individual cell level in TAG and chlorophyll content in nitrogen deprived cells which is further confirmed by the bimodal distribution in both TAG and chlorophyll as shown in the large sample size histograms.

measurements, assessment of lipid accumulation on an individual cell basis will be crucial toward understanding the actual cellular processes involved in accumulation.

This is not the first study documenting subpopulation variability under nutrient stress in microalgae. Flow cytometric results from the Aquatic Species Program documented two subpopulations based on TAG (Nile Red) and chlorophyll fluorescence in *Isochrysis* under N deprivation [3]. The bimodal distribution of lipid and chlorophyll found in both a haptophyte and a diatom (this study) implies that this phenomenon is more common in microalgae than is appreciated. Additionally, other studies have examined photosynthetic efficiency (Fv/Fm) and activity (^{14}C bicarbonate fixation rate) and found substantial variability amongst individual cells in diatoms and dinophyceae [20,21].

Cell-to-cell variability in a clonal population could result from several factors. In *E. coli*, there are differences in the growth rates of the daughter cells from parent cells that appear to divide symmetrically — daughters that inherited the "older half" of the cell tended to have slower growth rates than the daughter with the newly synthesized cell wall [31]. One possible explanation for differences in the growth rate of a clonal population could be epigenetic factors. Recent work has documented the same phenomenon in diatoms, in which daughter cells that inherited more parental material exhibited increased division rates [38]. Factors other than unequal distribution of material were also implicated [38], suggesting that epigenetic differences could play a role. A recent analysis highlights the potential of epigenetic variation in diatoms supporting this concept [39]. Combining the differential fitness of daughter cells with the fact that cells of different ages constitute the population (only 50% of the population is represented in the most recent generation in any culture) provides reasonable explanations for why large differences in metabolic capacity can occur under otherwise carefully controlled culture conditions. The general large spread in BODIPY and chlorophyll fluorescence under the culture conditions used here (Fig. 3) is evidence for this in relation to TAG levels.

The ultrastructure of lipid droplet formation has not been extensively examined in microalgae, however recent work in *Chlamydomonas reinhardtii* demonstrates a similar trend as documented here (Figs. 2 and 3c) [40–43]. Actively growing cells contain little or no fluorescence diagnostic of lipid droplets, and when droplets are present, they are small. Under limitation conditions, droplets increase in size and number [41,43], however under most conditions, they remain well defined in size and separated from each other. In *C. cryptica*, as droplets accumulate, they become more heterogeneous in size, and very large droplets become common (Fig. 2). In *C. reinhardtii*, only in the presence of acetate do lipid bodies become large enough to overwhelm the volume of the cell [41]. In contrast, *C. cryptica* (Figs. 2 and 3) and other diatoms (unpublished observations) are capable of hyperaccumulation under autotrophic conditions with no additional organic carbon source.

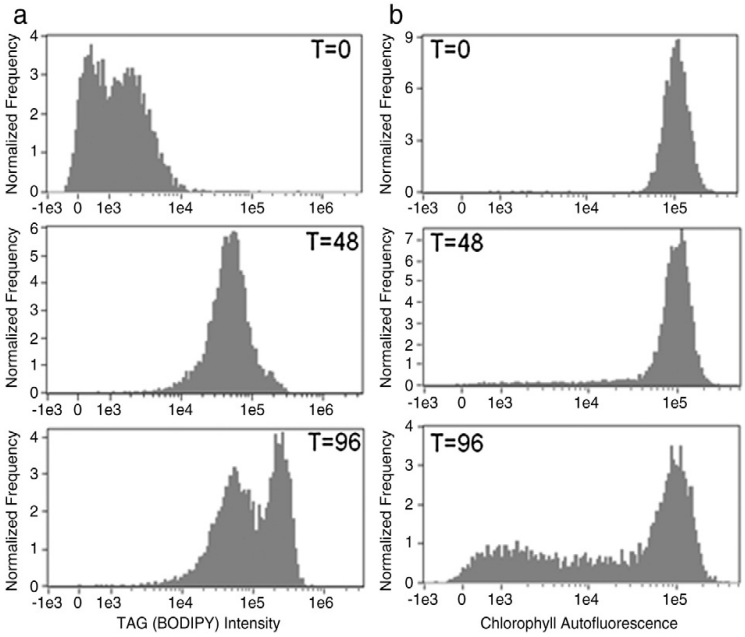


Fig. 6. Variance in BODIPY intensity and chlorophyll autofluorescence intensity under nitrogen deprivation (N1 experiment). (a) Histogram depicting TAG levels (measured using BODIPY) over 0, 48 and 96 h in N-ASW. (b) Histogram depicting chlorophyll autofluorescence intensity over 0, 48, and 96 h in N-ASW.

In the silicon limitation experiments, even though there is a consistent pattern of growth cessation and chlorophyll increase (Fig. 1a and b), the timing and extent of TAG accumulation varies (Fig. 1c). There is no strict correlation between cell cycle arrest (Fig. S1) or cessation of growth and TAG accumulation (Fig. 1). This may indicate that even though cell division immediately ceases, there is a slow progression to an arrest point in the cell cycle prior to an additional division occurring. We have observed a different response in *Thalassiosira pseudonana*, where cell cycle arrest correlates with growth arrest and TAG accumulation (Smith et al. in preparation).

The cellular response during TAG accumulation is more variable under N limitation than Si limitation in *C. cryptica*. For instance, the cell cycle arrest point for nitrogen limitation differed between replicates, with N1 arresting in G2 and N2 in G1 (Fig. S1). Based on previous investigations, differences in cell cycle arrest points under nitrogen free conditions are feasible. Nitrogen limitation has been found to cause an increase of cells in either G1 or G2 [16,44]. The data for both N and Si limitation do not support the concept that arrest at a particular cell cycle stage is required for TAG accumulation. This is in contrast to the dinoflagellate *Crypthecodinium cohnii*, where a commitment point for fatty acid biosynthesis in the late G1 phase of the cell cycle was demonstrated [45]. An interesting observation with the N limitation experiments was the consistent decrease or lack of increase in TAG accumulation at 72 h relative to 48 h. We cannot at present demonstrate why this occurred, but could speculate that some intracellular store of nitrogen became available. Regardless of the reason, this response highlights the idea that the TAG accumulation response is not linear; it is transient throughout time, an observation that would be missed with a 'before-and-after' analysis. In our experiments, during the course of Si limitation, the spread in the total population becomes tighter (Fig. 3a); whereas under N limitation, it becomes broader, with highly distinct subpopulations being generated at the later time point (Figs. 3b and 5). A major difference between the two limitations is the response of chlorophyll, which generally increases under Si limitation, but decreases under N limitation and generates a more highly populated HLC subpopulation at the later time point. Intracellular conditions in the HLC and HHC subpopulations in this experiment should be expected to be substantially different.

Common between both conditions were distinct phases of TAG accumulation: an early phase, and a faster phase during which hyperaccumulation occurred. Based on the substantially different rates, we hypothesize that the early and hyperaccumulation phases result from two distinct metabolic processes. These different processes are not related to the specific nutrient limitation, implying that they come into play after the initial triggering event for TAG accumulation. There are especially obvious advantages to understanding the underlying metabolic processes involved in hyperaccumulation, because this would be a first step in determining whether it could be triggered earlier to reduce the time required for maximum TAG accumulation. As demonstrated visually (Fig. 2), the amount of TAG that *C. cryptica* is capable of accumulating is substantial, as observed by microscopy of large lipid droplets within the cell (Figs. 2 and S3), but most of the increase occurs in one day after a 2-day period of less accumulation (Figs. 1 and 4).

As mentioned earlier, most analysis approaches are bulk measurements that generate data representative of the mean of a population, which in some cases can obscure the full richness of a response, or in the worst case, misrepresent it. An ideal experimental approach would allow evaluation of subpopulations of cells undergoing similar processes. Flow cytometric sorting enables separation of virtually any subpopulations, but only a relatively small number of cells can be sorted. Procedures such as whole transcriptome amplification [46] can address the problem of lack of material, but can also generate biases in transcript abundance. The application of imaging cytometry suggests a reasonable alternative. By defining subpopulations according to phenotypic differences (Fig. 3), and then following the progression of cells in a time course experiment, sequential times in which the majority of the population shifts between subpopulations can be identified. For example, comparing 48 and 72 hour time points in Si1 (Figs. 3a,

4a) we see that the majority of the cells shift from SD to HHC. A bulk analysis on these two samples should be fairly specific for changes occurring during this transition. The relevance of some shifts may not be obvious when only measuring the mean, for example in both S1 and N1 comparing 0 with 24 h (Fig. 3), major shifts occur according to phenotype which are not reflected in the bulk mean measurement (Fig. 4). The approaches developed in this report should enable more informed choices of relevance to specific cellular phenomena for analysis.

The results of this study have revealed several previously undescribed phenomena related to TAG accumulation. One is the change in rate of TAG accumulation over time, which is especially relevant for biofuels production because much higher rates were shown to be possible than under the initial conditions. Coupled to this, we document a hyperaccumulation of TAG under strictly autotrophic conditions, which we have also seen in other diatom species (unpublished observations). Differences in the responses occur even under carefully controlled culture conditions, which we suggest may be due to epigenetic variations. Similarly, under most conditions, there is a large spread in the population when measuring either chlorophyll or BODIPY, which could also relate to epigenetic variations. These variations document the richness of the TAG accumulation response, and such variations may be more likely to occur under less controlled outdoor conditions. Finally, a potential means to couple subpopulation-level responses with bulk analysis approaches has been described, which could take advantage of the nuances observed during the TAG accumulation process.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.algal.2013.03.003.

Acknowledgments

Funding of this project was provided in part by AFOSR grant FA9550-08-1-0178, DOE grant DE-EE0001222 and NSF grant CBET-0903712. The authors would like to acknowledge Raffaela Abbriano and Emily Trentacoste for their advice and helpful comments to the manuscript.

References

[1] Y. Chisti, Biodiesel from microalgae, Biotechnology Advances 25 (2007) 294–306.
[2] Y. Chisti, Biodiesel from microalgae beats bioethanol, Trends in Biotechnology 26 (2008) 1–6.
[3] J. Sheehan, T. Dunahay, J. Benemann, P. Roessler, A look back at the US department of energy's aquatic species program: biodiesel from algae, 1998.
[4] N. Shifrin, S. Chisholm, Phytoplankton lipids: interspecific differences and effects of nitrate, silicate and light-dark cycles, Journal of Phycology 17 (1981) 374–384.
[5] P. Pienkos, A. Darzins, The promise and challenges of microalgal-derived biofuels, Biofuels 3 (2009) 431–440.
[6] M. Hildebrand, A.K. Davis, S.R. Smith, J.C. Traller, R. Abbriano, The place of diatoms in the biofuels industry, Biofuels 3 (2012) 221–240.
[7] D. Werner, Die Kieselsäure im Stoffwechsel von *Cyclotella cryptica* Reimann, Lewin und Guillard, Archives of Microbiology 55 (1966) 278–308.
[8] P. Roessler, Effects of silicon deficiency on lipid composition and metabolism in the diatom *Cyclotella cryptica*, Journal of Phycology 24 (1988) 394–400.
[9] M. Huesemann, T. Hausmann, R. Bartha, Biomass productivities in wild type and pigment mutant of *Cyclotella* sp. (diatom), Applied Biochemistry 3 (2009) 507–526.
[10] N. Kröger, C. Bergsdorf, M. Sumper, A new calcium binding glycoprotein family constitutes a major diatom cell wall component, The EMBO Journal 3 (1994) 4676–4683.
[11] N. Kröger, M. Sumper, Diatom cell wall proteins and the cell biology of silica biomineralization, Protist 149 (1998) 213–219.
[12] M. Hildebrand, Diatoms, biomineralization processes, and genomics, Chemical Reviews 108 (2008) 4855–4874.
[13] V. Martin-Jézéquel, M. Hildebrand, Silicon metabolism in diatoms: implications for growth, Journal of Phycology 36 (2000) 821–840.
[14] M. Brzezinski, The Si:C:N ratio of marine diatoms: interspecific variability and the effect of some environmental variables, Journal of Phycology 21 (1985) 347–357.
[15] E. Yu, F. Zendejas, P. Lane, S. Gaucher, Triacylglycerol accumulation and profiling in the model diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (Baccilariophyceae) during starvation, Journal of Applied Phycology 21 (2009) 669–681.
[16] P. Claquin, V. Martin-Jézéquel, Uncoupling of silicon compared with carbon and nitrogen metabolisms and the role of the cell cycle in continuous cultures of *Thalassiosira pseudonana* (Bacillariophyceae) under light, nitrogen, and phosphorous control, Journal of Phycology 38 (2002) 922–930.
[17] Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances, The Plant Journal 54 (2008) 621–639.
[18] M.T. Guarnieri, A. Nag, S.L. Smolinski, al Darzins, M. Seibert, P.T. Pienkos, Examination of triacylglycerol biosynthetic pathways via de novo transcriptomic and proteomic analyses in an unsequenced microalga, PLoS One 6 (2011) e25851.
[19] A. Jamers, R. Blust, W. Coen, Omics in algae: paving the way for a systems biological understanding of algal stress phenomena? Aquatic Toxicology 92 (2009) 114–121.
[20] N. Hockin, T. Mock, F. Mulholland, S. Kopriva, The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants, Plant Physiology 158 (2012) 299–312.
[21] S. Smith, R. Abbriano, M. Hildebrand, Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways, Algal Research 1 (2012) 2–16.
[22] K. Cooksey, J. Guckert, S. Williams, Fluorometric determination of the neutral lipid content of microalgal cells using Nile Red, Journal of Microbiological Methods 6 (1987) 333–345.
[23] W. Chen, C. Zhang, L. Song, M. Sommerfeld, A high throughput Nile Red method for quantitative measurement of neutral lipids in microalgae, Journal of Microbiological Methods 77 (2009) 41–47.
[24] T. Govender, L. Ramanna, I. Rawat, F. Bux, BODIPY staining, an alternative to the Nile Red fluorescence method for the evaluation of intracellular lipids in microalgae, Bioresource Technology 114 (2012) 507–511.
[25] M.S. Cooper, W.R. Hardin, T.W. Petersen, R.A. Cattolico, Visualizing "green oil" in live algal cells, Journal of Bioscience and Bioengineering 109 (2010) 198–201.
[26] E. Bertozzini, L. Galluzzi, A. Penna, M. Magnani, Application of the standard addition method for the absolute quantification of neutral lipids in microalgae using Nile red, Journal of Microbiological (2011) 17–23.
[27] H. Guzmán, A. Valido, K. Presmanes, L. Duarte, Quick estimation of intraspecific variation of fatty acid composition in *Dunaliella salina* using flow cytometry and Nile Red, Journal of Applied Phycology 24 (2012) 1237–1243.
[28] D. Xu, Z. Gao, F. Li, X. Fan, X. Zhang, N. Ye, et al., Detection and quantitation of lipid in the microalga *Tetraselmis subcordiformis* (Wille) Butcher with BODIPY 505/515 staining, Bioresource Technology 127 (2013) 386–390.
[29] N. Toriello, E. Douglas, N. Thaitrong, Integrated microfluidic bioprocessor for single-cell gene expression analysis, Proceedings of the National Academy of Sciences 105 (2008) 20173–20178.
[30] D. Wang, S. Bodovitz, Single cell analysis: the new frontier in "omics", Trends in Biotechnology 28 (2010) 281–290.
[31] E.J. Stewart, R. Madden, G. Paul, F. Taddei, Aging and death in an organism that reproduces by morphologically symmetric division, PLoS Biology 3 (2005) e45.
[32] E. Voronova, L. Il'ash, S. Pogosyan, A. Ulanova, D. Matorin, M. Cho, A. Rubin, Intrapopulation heterogeneity of the fluorescence parameters of the marine plankton alga *Thalassiosira weissflogii* at various nitrogen levels, Microbiology 78 (2009) 469–478.
[33] W.M. Darley, B.E. Volcani, Role of silicon in diatom metabolism, Experimental Cell Research 58 (1969) 334–342.
[34] P. Beum, M. Lindorfer, B. Hall, T. George, Quantitative analysis of protein co-localization on B cells opsonized with rituximab and complement using the ImageStream multispectral imaging flow cytometer, Journal of Immunological Methods 317 (2006) 90–99.
[35] M. Brzezinski, R. Olson, S. Chisholm, Silicon availability and cell-cycle progression in marine diatoms, Marine Ecology Progress Series 67 (1990) 83–96.
[36] F. Healey, J. Coombs, B. Volcani, Changes in pigment content of the diatom *Navicula pelliculosa* (Bréb.) Hilse in silicon-starvation synchrony, Archives of Microbiology 59 (1967) 131–142.
[37] L. Brennan, A. Fernández, A. Mostaert, P. Owende, Enhancement of BODIPY$^{505/515}$ lipid fluorescence method for applications in biofuel-directed microalgae production, Journal of Microbiological Methods 90 (2012) 137–143.
[38] S. Laney, R. Olson, H. Sosik, Diatoms favor their younger daughters, Limnology and Oceanography 57 (2012) 1572–1578.
[39] F. Maumus, P. Rabinowicz, C. Bowler, Stemming epigenetics in marine stramenopiles, Current Genomics 12 (2011) 357–370.
[40] E. Moellering, C. Benning, RNA interference silencing of a major lipid droplet protein affects lipid droplet size in *Chlamydomonas reinhardtii*, Eukaryotic Cell 9 (2010) 97–106.
[41] C. Goodson, R. Roth, Z.T. Wang, U. Goodenough, Structural correlates of cytoplasmic and chloroplast lipid body synthesis in *Chlamydomonas reinhardtii* and stimulation of lipid body production with acetate boost, Eukaryotic Cell 10 (2011) 1592–1606.
[42] M. Siaut, S. Cuiné, C. Cagnon, B. Fessler, Oil accumulation in the model green alga *Chlamydomonas reinhardtii*: characterization, variability between common laboratory strains and relationship with starch reserves, BMC Biotechnology 11 (2011) 1–15.
[43] Z. Wang, N. Ullrich, S. Joo, S. Waffenschmidt, Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtii*, Eukaryotic Cell 8 (2009) 1856–1868.
[44] D. Vaulot, F. Partensky, Cell cycle distributions of prochlorophytes in the North Western Mediterranean Sea, Deep Sea Research Part A Oceanographic Research Papers 39 (1992) 727–742.
[45] A. Kwok, J. Wong, Lipid biosynthesis and its coordination with cell cycle progression, Plant and Cell Physiology 46 (2005) 1973–1986.
[46] C. Peano, M. Severgnini, I. Cifola, Transcriptome amplification methods in gene expression profiling, Expert Review Of Molecular Archives, Neoplasia 8 (2006) 153–162.

CHAPTER 1 SUPPLEMENTAL MATERIAL

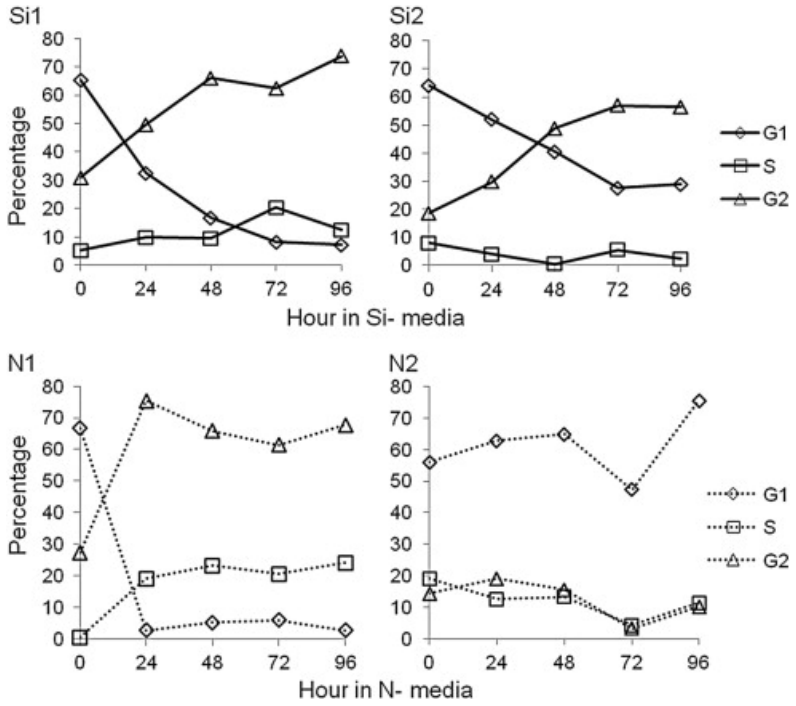


Figure S1: Cell cycle analysis of nutrient depletion experiments. Top row, silicon limitation. Bottom row, nitrogen limitation. The percentage of cells arrested at each cell cycle stage (G1, S, or G2) is plotted.

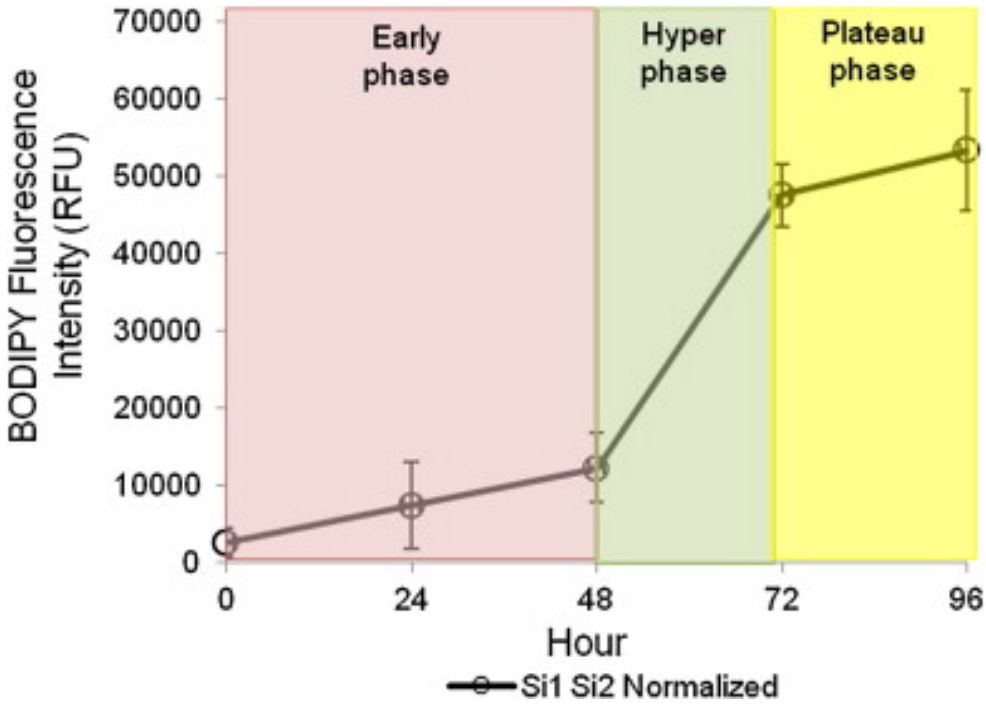


Figure S2: Normalized data depicting three different phases of TAG accumulation under silicon deprivation in *C. cryptica*: early, hyper, and plateau phases. Normalization was done by averaging the values over all time points for Si1 and Si2, then adjusting the values of individual time points by that factor.

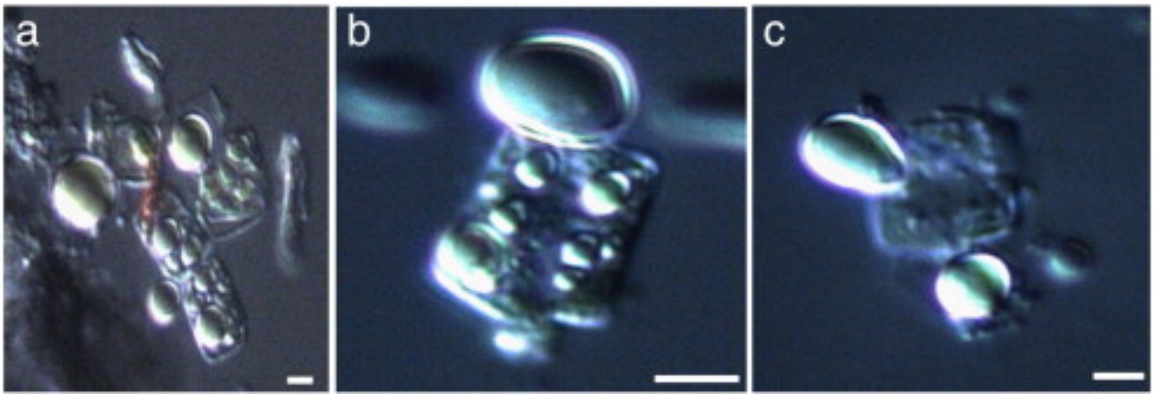


Figure S3: High resolution microscopy depicts lipid bodies that have escaped from the cell. (a) 144 h in N-ASW. (b, c) 96 h in N-ASW. Similar images were captured under silicon deprivation (data not shown).

1.8 ACKNOWLEDGEMENTS

Chapter 1, in full, is a reprint of the material as it appears in Algal Research 2013. Traller, Jesse C. and Hildebrand, Mark. "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(3), 244-252. 2013. The dissertation author was the primary researcher and author of this paper.

CHAPTER 2

Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype

Traller *et al. Biotechnol Biofuels (2016) 9:258*
DOI 10.1186/s13068-016-0670-3

Biotechnology for Biofuels

RESEARCH Open Access

Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype

Jesse C. Traller[1], Shawn J. Cokus[2], David A. Lopez[2], Olga Gaidarenko[1], Sarah R. Smith[1,3], John P. McCrow[3], Sean D. Gallaher[2,4], Sheila Podell[1], Michael Thompson[2], Orna Cook[1], Marco Morselli[2], Artur Jaroszewicz[2], Eric E. Allen[1], Andrew E. Allen[1,3], Sabeeha S. Merchant[4], Matteo Pellegrini[2] and Mark Hildebrand[1*]

Abstract

Background: Improvement in the performance of eukaryotic microalgae for biofuel and bioproduct production is largely dependent on characterization of metabolic mechanisms within the cell. The marine diatom *Cyclotella cryptica,* which was originally identified in the Aquatic Species Program, is a promising strain of microalgae for large-scale production of biofuel and bioproducts, such as omega-3 fatty acids.

Results: We sequenced the nuclear genome and methylome of this oleaginous diatom to identify the genetic traits that enable substantial accumulation of triacylglycerol. The genome is comprised of highly methylated repetitive sequence, which does not significantly change under silicon starved lipid induction, and data further suggests the primary role of DNA methylation is to suppress DNA transposition. Annotation of pivotal glycolytic, lipid metabolism, and carbohydrate degradation processes reveal an expanded enzyme repertoire in *C. cryptica* that would allow for an increased metabolic capacity toward triacylglycerol production. Identification of previously unidentified genes, including those involved in carbon transport and chitin metabolism, provide potential targets for genetic manipulation of carbon flux to further increase its lipid phenotype. New genetic tools were developed, bringing this organism on a par with other microalgae in terms of genetic manipulation and characterization approaches.

Conclusions: Functional annotation and detailed cross-species comparison of key carbon rich processes in *C. cryptica* highlights the importance of enzymatic subcellular compartmentation for regulation of carbon flux, which is often overlooked in photosynthetic microeukaryotes. The availability of the genome sequence, as well as advanced genetic manipulation tools enable further development of this organism for deployment in large-scale production systems.

Keywords: Diatom, Genome sequence, *Cyclotella cryptica*, Algae biofuel, Carbon metabolism, DNA methylation

Background

Global environmental changes are happening at an increasingly rapid rate, and development of technologies to alleviate negative outcomes is urgently needed. One proposed solution is producing energy from biofuel, which is renewable and has fewer detrimental effects than the use of fossil fuels. A potential feedstock for biofuel production, microalgae, has garnered interest because of its high productivity. The overall requirement for land to grow algae to sustain the United States' fuel supply is predicted to be relatively small (<4% of the total land mass) compared to other crops [1].

Developing technologies have increased the promise of algal biofuel to meet energy needs [2, 3]. These new advances were preceded by the US Department of Energy funded Aquatic Species Program (ASP) which produced pioneering work in aquaculture and large-scale

*Correspondence: mhildebrand@ucsd.edu
[1] Scripps Institution of Oceanography, University California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0202, USA
Full list of author information is available at the end of the article

production of biodiesel from microalgae [4]. During the ASP, algal species across the tree of life were assessed for their ability to accumulate abundant triacylglycerol (a precursor to biofuel), to grow under variable environmental conditions such as pH, salinity and temperature, and to grow at a productive and sustainable rate in outdoor raceway ponds [4]. Several genera stood out from the rest, including chlorophytes, chrysophytes, and diatoms. Diatoms (Bacillariophyta), which are naturally highly productive, accounted for ~60% of the top-performing species in a recommended list of biofuel production organisms produced by the ASP [4].

Cyclotella cryptica (Fig. 1), a brackish water diatom isolated from Martha's Vineyard, Massachusetts, was identified in the ASP as a top species for large-scale biofuel production. *Cyclotella cryptica* has been shown to be an excellent accumulator of lipid ([5, 6]; Fig. 1), is euryhaline, enabling flexibility in cultivation conditions [5–7], and can grow outdoors at levels between 20.0 and 29.7 g Ash Free Dry Weight (AFDW) m^2/day in a 2.8 and 48 m^2 pond, respectively [8, 9]. During the ASP, *C. cryptica* was used to investigate the lipid accumulation response during starvation for silicon, a macronutrient required by diatoms to synthesize their silicified cell walls, as well as to understand the properties of key enzymes involved in flux of carbon into lipid [10, 11]. More recently, based on a survey of 175 different microalgal strains, *C. cryptica* was selected as a top candidate for omega-3 fatty acid production, which is highly desirable for pharmaceutical and aquaculture applications, as well as a top producer of protein and nitrogen, suitable for agricultural feed [5]. In addition to having traits suited for commercial production, *C. cryptica* was the first chlorophyll-*c* containing algae to undergo stable nuclear transformation, a milestone in algal genetic engineering [12]. That study was the first of many significant advances in diatom genetic engineering, including determining the subcellular localization of proteins using GFP fusions, RNAi and antisense knockdowns, CRISPR, transcription activator-like effector nucleases (TALEN), and plasmid delivery via conjugation using an artificial episome [13–18].

Although *C. cryptica* and other microalgal species have excellent native productivity characteristics, cost analyses [19] indicate that further improvements are necessary to make algal biofuel production economically competitive with fossil fuels. Genetically based approaches are required to establish robust improved-productivity phenotypes. Both random mutagenesis and directed genetic manipulation can be used to accomplish this. The latter requires knowledge of an organism's genome sequence, and such information has successfully been used to identify targets for genetic manipulation in diatoms to

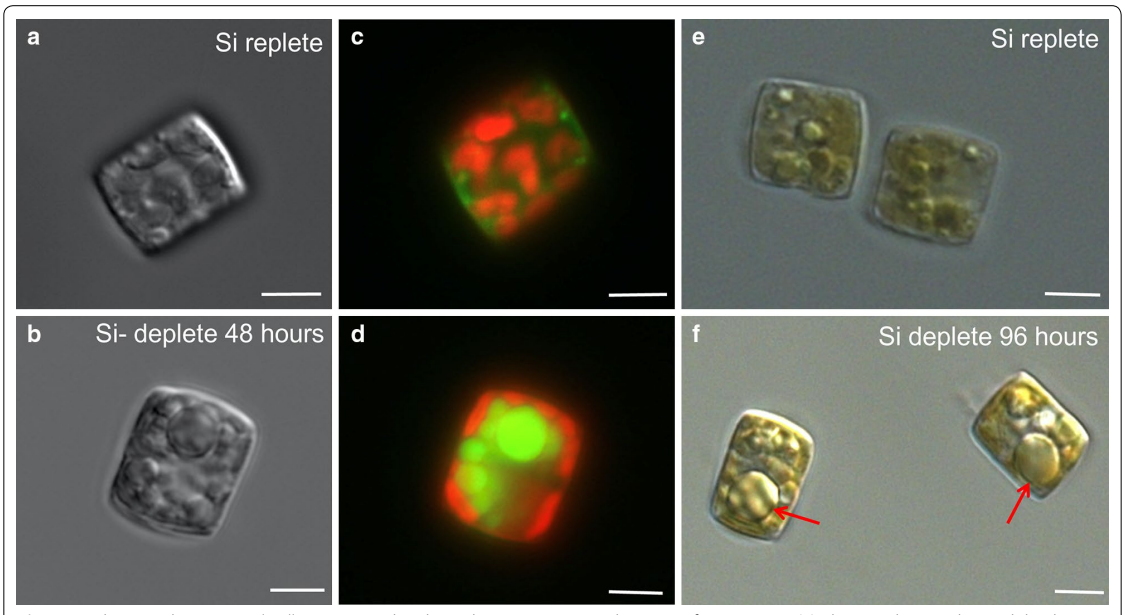


Fig. 1 Lipid accumulation in *Cyclotella cryptica* under silicon deprivation. Grayscale image of *C. cryptica* in (**a**) silicon replete medium, 0 h lipid uninduced, or (**b**) 48 h silicon deplete, lipid induced. **c**, **d** Respective composite images of chlorophyll autofluorescence (*red*) and fluorescent lipophilic dye BODIPY (*green*). **e**, **f** Differential interference contrast image of silicon replete (**e**) and 96 h silicon deplete, lipid induced, with *red arrows* identifying lipid droplets (**f**). *Scale bars* 5 μm

improve lipid productivity [14, 17, 20, 21]. To identify appropriate gene targets for manipulation, a thorough understanding of the enzymes involved in central carbon metabolism is required, including the number of isozymes that catalyze each chemical reaction and the compartment-specific localization of enzymes or enzymatic processes within the cell. It is especially important to consider organellar compartmentation of metabolic processes in a diatom cell, which, because of diatoms' secondary endosymbiotic origin, contain additional compartments relative to the green algae. This includes the periplastid compartment surrounding the chloroplast and endoplasmic reticulum that also surrounds the periplastid compartment around the chloroplast [22–24]. Studying the diversity of core processes in carbon metabolism, such as glycolysis and fatty acid biosynthesis, within different lineages of algae [25, 26] as well as more closely within species of the same lineage [27–29], may allow researchers to address why certain species are better suited for biofuel production than others. Differences in the organization of primary carbon metabolism between species likely reflect differences in efficiencies in processing carbon, which relates to factors controlling their productivity. Understanding the dissimilarities amongst microalgae will help elucidate what constitutes an optimized biofuel/bioproduct production system, and enable a production species to be manipulated to create the desired product in the most efficient manner. The availability of additional algal genome sequences has enabled a more thorough comparison of these diverse polyphyletic organisms to identify metabolic steps that may influence the organism's productivity characteristics. These comparisons are also essential to understanding the complex evolutionary history and ecology of microalgae.

To bring the promising characteristics of *C. cryptica* in line with current day approaches to improve productivity, we sequenced the nuclear, chloroplast, and mitochondrial genomes, performed a detailed in silico analysis of the core metabolic processes involved in or competing with lipid production and compared these to other diatom genomes, in particular the closely related *Thalassiosira pseudonana*. To investigate whether epigenetic factors influence lipid accumulation or primary metabolism in general, bisulfite sequencing was performed to examine the methylome of cells under silicon replete and silicon-deplete, lipid induced conditions. We further developed *C. cryptica* by application of genetic tools including fluorescent protein tagging and the use of an inducible promoter using genetic constructs derived from *T. pseudonana*. Genome sequence and data generated from this analysis provides a foundation to further improve this species for large-scale biofuel/bioproducts production and provide insight into central carbon metabolism in diatoms.

Results and discussion

Genome sequence determination and assembly

Three libraries with different average insert lengths were prepared from purified *C. cryptica* genomic DNA (Additional file 1: Additional methods). These were sequenced as paired-end 76-mer + 76-mer reads on an Illumina GA-IIx 120-tiles/lane run. Two genomic DNA mate pair libraries (aiming for 10 K nucleotide effective inserts) were prepared and run by Illumina service on a 48-tile/lane v3 HiSeq flow cell. The paired-end and mate pair libraries contributed ~23.4 G nucleotides (nt) and ~57.4 G nt, respectively, for a total of ~80.8 G nt. The main genome assembly was performed with an ABySS 1.3.1 single end, paired-end, mate-pair pipeline. There were 116,817 genomic contigs and the N50 value was 11,951 bp (Additional file 1: Figure S1).

In addition to the genome assembly, transcriptomes (13 samples) were generated for *C. cryptica* using RNAseq data under silicon limitation and nitrogen limitation (Additional file 1: Additional methods). Estimates of mRNA abundance were calculated for each gene model in terms of fragments per kbp of transcript per million mapped reads (FPKMs) to identify relative transcript levels for all genes across all experimental conditions. Gene expression patterns will be analyzed in a subsequent study.

Genome statistics and gene model prediction

The estimated size of the haploid genome of *C. cryptica* was 161.7 Mbp, which is substantially larger than the 31 Mbp genome of *T. pseudonana*, a closely related model centric diatom. Similar to *T. pseudonana*, *C. cryptica* primarily exists in its vegetative state as diploid, and rarely undergoes sexual reproduction in controlled culture conditions. Multiple gene model prediction pipelines, including AUGUSTUS and MAKER, were evaluated to estimate the number of genes (Additional file 1: Table S1, Additional methods). The quality of predicted gene models was assessed by comparing *C. cryptica* RNAseq data and against manually curated *T. pseudonana* (version 3) gene models, which are supported by RNAseq and EST data [30–32]. MAKER predicted the fewest genes (Table 1; Additional file 1: Table S1), yet most of these models were supported by transcript data, and therefore, considered high confidence. However, intron and exon boundaries were improved in gene models from AUGUSTUS trained on the de novo *C. cryptica* RNA assembly. Despite these improvements, AUGUSTUS predicted a large proportion (30%) of fragmented, short gene models unsupported by transcript data (sum FPKM = 0.00, Table 1; Additional file 1: Table S1), which were presumed misscalls and removed from subsequent analysis. To leverage the strengths of the different predictors

Table 1 Genomic features in *Cyclotella cryptica* and *Thalassiosira pseudonana*

Statistics	*C. cryptica*		*T. pseudonana*
Cell size	8 × 10 µm		4 × 5 µm
Nuclear genome size	161.7 Mbp (GC 43%)		31 Mbp (GC 47%)
Repeatome (%)	53		2
Classified repeats (%)	13		ND
Unclassified repeats (%)	40		ND
Gene models	**High confidence**	**AUGUSTUS models**	***T. pseudonana* Joint Genome Institute models**
Percent coding DNA	10.0	19.2	66.3
Gene density (genes/Mbp)	50	131	379
Total models	8133	21,121	11,776
Average model length (bp)	1986	1471	1746
Average number of exons per gene	2.95	2.18	2.54
Average exon length (bp)	599	608	613
Average number of introns per gene	1.95	1.18	1.5
Average intron length (bp)	115	125	125
Chloroplast genome size (bp)	129,320		128,813
Total chloroplast models	132		127
Mitochondrial genome size (bp)	58,021		43,827
Total mitochondrial models	35		35

T. pseudonana data from [30, 34]

(boundary accuracy vs. expression-supported models), final 'high-confidence' gene models for *C. cryptica* are the set of AUGUSTUS gene models overlapping a MAKER prediction and assigned non-zero FPKM values (8133 genes). The total number of genes in *C. cryptica* per haploid genome, including the high confidence gene subset and all other AUGUSTUS models with RNAseq support was 21,121.

Gene density in *C. cryptica* was 3.4× lower than the closely related model species *T. pseudonana* with 19% coding DNA compared to 66% (Table 1). While the total number of gene models in *C. cryptica* was nearly double that of *T. pseudonana*, it did not scale with the increase in genome size; the *C. cryptica* genome was 5.2× larger than *T. pseudonana*. This generally correlated with the logarithmic relationship between gene content and genome size in microeukaryotes as shown in [33]. The low gene density in *C. cryptica* can be attributed to a large quantity of repeat sequence, 53% of the genome. The majority of repeat sequence was detected using RepeatModeler, a program to identify de novo repeat families. RepeatModeler predicted 314,059 unclassified repeat elements totaling to 40% of the genome (Additional file 1: Table S2b). In addition to novel repeat families, 8.6% of the genome is comprised of long terminal repeat elements, characteristic of Class I retrotransposons (Additional file 1: Tables S2a, S2b), which are common to diatoms and are hypothesized to contribute to the diversity and ecological success of these organisms in the oceans [31, 34].

Cyclotella cryptica gene model repertoire

To determine the underlying basis for *C. cryptica's* genetic characteristics relative to other diatoms, we defined orthologs and compared molecular divergence between *C. cryptica* and five other diatoms based on OrthoMCL, 18S phylogeny, and reciprocal best BLAST hit (RBH) analysis (Fig. 2; Additional file 1: Figure S2). *Thalassiosira pseudonana* and *C. cryptica* were the most closely related centric diatoms based on 18S and percent identity of RBH pairs, with 5498 shared pairs between them and 63.6% average percent identity across all RBH pairs (Additional file 1: Figure S2).

OrthoMCL uses a Markov Cluster algorithm to cluster putative orthologs across species and paralogs within species and is a powerful tool for comparative genomics and functional genome annotation [35]. Just under half of the genes have an orthologous match to *T. pseudonana* (9877 genes, 47%, Fig. 2a). By comparing all diatom genomes, we found 5008 genes (24% of total genes) were only found in *C. cryptica*, while 2699 genes (13%) shared a match to at least one other diatom but were absent in *T. pseudonana,* presumably due to gene loss (Fig. 2a, b). In addition, 17% of genes in *C. cryptica* are expanded orthologs to a gene(s) in *T. pseudonana*, which have resulted from either paralogous

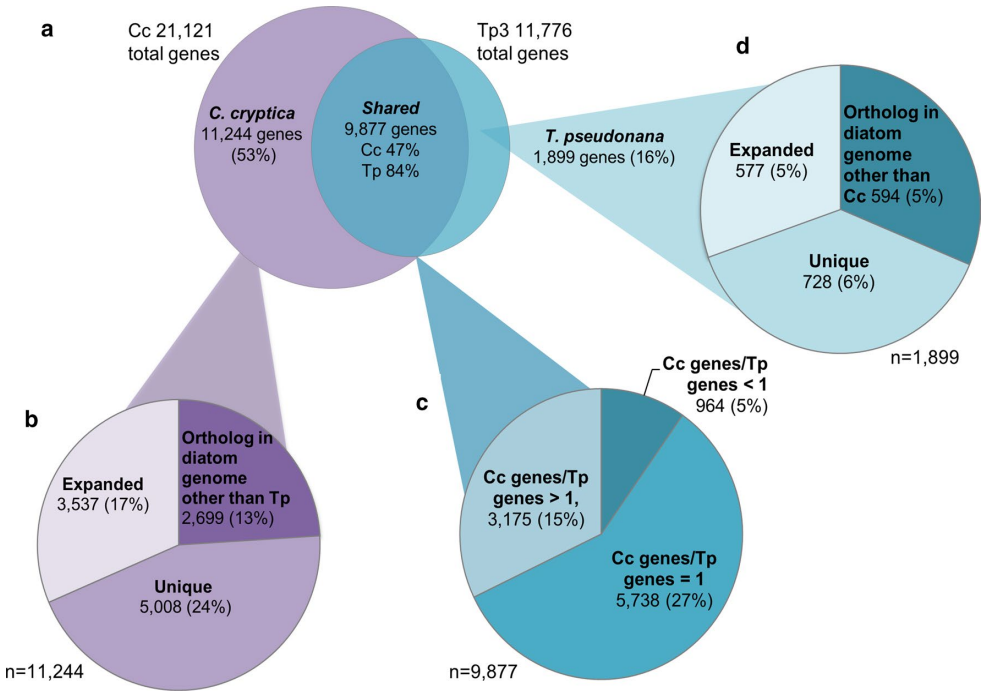


Fig. 2 *Cyclotella cryptica* genetic components. Percentages reflect group relative to the total number of gene models in the species. **a** Comparison of genes between *C. cryptica* and *T. pseudonana* (JGI version 3). Over half the genes in *C. cryptica* (Cc) are not found in *T. pseudonana* (Tp). **b** Classification of *C. cryptica* specific genes ($n = 11,244$) relative to other diatom genomes. **c** Relative copy numbers of genes comparing *C. cryptica* and *T. pseudonana* ($n = 9877$). Percentages shown are based on the Cc total gene models. **d** Classification of *T. pseudonana* specific genes relative to other diatom genomes ($n = 1899$)

duplication or horizontal gene transfer in *C. cryptica,* or gene loss in *T. pseudonana* (Fig. 2b). 5738 genes were found in orthologous clusters where a gene in *C. cryptica* was present in equal copy numbers to genes in *T. pseudonana,* accounting for 27% of the total genes in *C. cryptica* (Fig. 2a, c). Based on the OrthoMCL data with these diatom genomes, the increase in gene content in *C. cryptica* relative to *T. pseudonana* is predominantly composed of unique genes not found in any other species, comprising 24% of the gene content. The proportion of unique genes in *T. pseudonana* was significantly less at 6% (Fig. 2b, d).

Genes unique to *C. cryptica* and not detected in the other diatoms contained minimal functional annotation based on KEGG, KOG, GO, Pfam, and TIGRfam; however, there was a notable enrichment in genes related to transposon processes, such as transposable element domains integrase (Pfam accession PF00665, 33 predicted genes), plant transposon gene (PF04827, 12 predicted genes), and reverse transcriptase (PF07727, 71 predicted genes). In contrast, the *T. pseudonana* genome contained only four genes with Integrase annotation, and one reverse transcriptase. These domains are indicators of mobile DNA elements, which are often responsible for genome size expansion. Given that *C. cryptica* and *T. pseudonana* have speciated relatively recently [36], the large amount of repetitive sequence with additional unique genes with transposable element annotation suggest that either *C. cryptica* has undergone a recent dramatic genome expansion event, more than doubling the amount of DNA and/or *T. pseudonana* has undergone a substantial genome deletion of repetitive sequence. The former scenario is favored due to the high number of *C. cryptica* unique genes with RNAseq support (Fig. 2b).

In addition to genes encoding mobile DNA element proteins, which were more abundant than those found in all other diatoms we analyzed, there was also an expansion of genes encoding proteins putatively involved in cell adhesion, signaling, and transport. Orthologous cluster (clust_77) contained 85 genes in *C. cryptica* with an enrichment of the discoidin domain, which is a glycoprotein putatively involved in carbohydrate binding and cytoskeletal organization [37]. In addition, relative to *T. pseudonana*, *C. cryptica* has two additional silicon transporters (clust_209, six copies total, including one partial

sequence), and additional ATPases (clust_43, 67 copies) and phosphate and nitrate transporters (clust_248-NRT, Additional file 2). Since the *C. cryptica* cell is approximately twice as large in volume as *T. pseudonana*, it is reasonable to suggest that *C. cryptica* requires additional proteins at the cellular surface to compensate for the lower surface to volume ratio. Another possibility is that the expanded gene repertoire could allow for diversification function and metabolite flux. In addition to a nitrate transporter, *C. cryptica* possesses an additional plastid nitrate/formate transporter (g21971.t1), which may explain its high capacity for nitrogen assimilation, as shown in [5].

In previous studies, it has been demonstrated that up to 5% of diatom genes were proposed to arise via horizontal gene transfer [31]. We used the bioinformatic tool DarkHorse [38] to evaluate the contribution of horizontal gene transfer to gene expansion in *C. cryptica* (Additional file 2, DarkHorse). There were 16 unique genes of unknown function but matching viral sequences, however, most of these genes contained low levels of transcript and were not full length. Further analysis identified 312 genes (1.47% of total genes) with best matches to bacterial sequences, with 137 of those specific for Proteobacteria and 20 of cyanobacterial origin, a similar taxonomic grouping as found in [31]. Among these, there was an enrichment in genes belonging in the KOG class 'Secondary metabolites biosynthesis, transport and catabolism', with 6 of 312 (1.92%) versus 69 of 21,121 (0.33%) overall. In addition to horizontally acquired genes identified by DarkHorse, many of the predicted proteins contained no matches to any previously sequenced organism (4175, 20% of genes), which suggest that they were either acquired from an uncharacterized organism, have rapidly evolved in *C. cryptica*, or this diatom lineage is significantly unrepresented in sequencing databases. While it is possible that some of these unique genes may be erroneously predicted, several proteins did contain at least one Pfam domain. Overall, the number of genes acquired horizontally appears to be similar to that found in previous studies in diatoms with smaller genome sizes, and the increase in gene number in *C. cryptica* is primarily due to the prevalence of unknown genes, similar to findings from OrthoMCL (Fig. 2b; Additional file 2). While the annotation of these genes is limited, given the high number of repetitive sequence and transposable elements, we hypothesize that many of these unique genes are cryptic ORFs created from transposition events, similar to what has been detected in higher plants [39, 40]. It must be noted that many candidate foreign genes identified using the DarkHorse analysis may also be ancestral genes of diatoms, acquired vertically or via endosymbiosis. Examples of these are genes matching to organisms within 'SAR excluding diatoms' (the subgroup containing Stramenopiles, Alveolates, and Rhizaria) and Excavata; these genes are found in *C. cryptica* but appear to have been lost in the other diatoms included in this analysis.

Methylation of the genome

DNA methylation can play a role in altering genome size and content by silencing genomic regions, and can have more dynamic effects in regulating expression of particular genes. The significant difference in genome size between *C. cryptica* with *T. pseudonana* and high levels of repetitive sequence prompted us to ask whether DNA methylation played a role in gene regulation and silencing of mobile DNA elements. Data from the pennate diatom *Phaeodactylum tricornutum* indicated that DNA methylation profiles correlate with transcript levels in some genes controlling nitrogen metabolism [41], therefore, an examination of dynamic changes in methylation in *C. cryptica* was also undertaken. To determine the relationship between DNA methylation and silicon limitation, which, similar to nitrogen, is an essential macronutrient to diatoms, we performed whole genome bisulfite sequencing on cells under silicon replete conditions (0 h), and after 48-h silicon depletion. In addition to silicon-limitation associated phenomena, a dramatic induction of triacylglycerol (TAG) levels occurs after 48 h of starvation (Fig. 1b, d; [4, 42]).

The global per-cytosine (CG, CHH, CHG motifs, where H is any nucleotide but G) methylation level averaged in both conditions of the *C. cryptica* genome was 61% (Fig. 3a), the highest amount of DNA methylation in a diatom genome observed to date. Under both experimental conditions, methylation of cytosine sites was remarkably similar and bimodal (Additional file 1: Figure S3), suggesting that *C. cryptica* does not undergo substantial temporal shifts in overall methylation.

We found that in silicon replete, lipid uninduced (0 h) and 48 h silicon depletion, lipid induced conditions, methylation patterns across the whole genome were almost identical, with 98.7% correlation between the two conditions (Additional file 1: Figure S4). While [34, 41] detected shifts in DNA methylation under nutrient (nitrate) stress in *P. tricornutum*, studies in green microalgae suggest that dramatic changes in DNA methylation tend to occur primarily during cellular division [43, 44]. In *C. cryptica*, cell growth ceases almost immediately after silicon starvation, while under N limitation, cell cycle progression will still proceed until full arrest occurs, and thus the immediate cessation of cell cycle under silicon deprivation may preclude changes in methylation.

Consistent with studies in other diatoms [41, 45], methylation is predominantly at CpG dinucleotides, with low levels of CHH and CHG methylation (Fig. 3a).

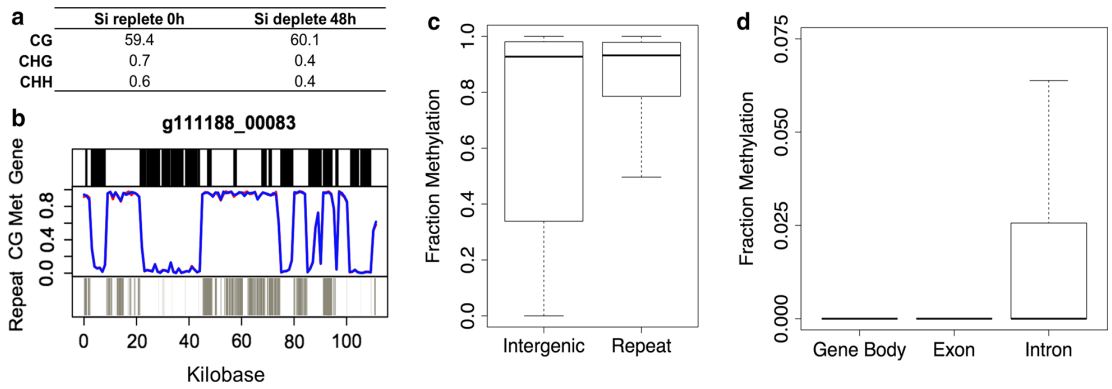


Fig. 3 DNA methylation in *Cyclotella cryptica*. **a** Percent genomic methylation under both conditions according to cytosine motif (denoted at *left*, H = A, T, or C nucleotides). **b** Fraction methylation across genomic contig g111188_00083. *Blue* silicon replete, *red* silicon deplete 48 h. Augustus V3 models depicted in *black*, Repeat Modeler data depicted in *gray*. **c** Fraction methylation across intergenic, repeat, and **d** gene body, exon, intron regions. Outliers in **c** and **d** have been removed

By comparison, the nuclear genome of *T. pseudonana* is 2.57% methylated, and *Fragilariopsis cylindrus*, with a genome of 81 Mbp, is 8.63% methylated [45]. Based on this study and previous methylome studies in other microalgae, there appears to be no strict correlation between overall methylation and genome size, as well as no pattern of DNA methylation within an algal class; for example, the global per-cytosine methylation of the chlorophytes *Chlorella NC64A* and *Chlamydomonas reinhardtii* are 82.65 and <2% with genome sizes of 42 and 120 Mbp, respectively [44, 46, 47].

To assess what genomic features are methylated in *C. cryptica*, AUGUSTUS gene models and RepeatModeler data were overlaid against methylation levels across the largest assembled genomic contigs for both conditions. The majority of highly methylated regions aligned with repeat sequence (Fig. 3b; Additional file 1: Figure S5). This is consistent with other methylation studies which hypothesize that DNA methylation inhibits transposable element expansion and is commonly found across repeat regions [41, 48]. These highly repetitive/methylated regions were substantial, spanning as large as 30 kb (Fig. 3b; Additional file 1: Figure S5). Additionally, there were similarly sized hypomethylated regions that were gene-rich, and contained essentially no methylation. While the pattern of hypermethylated repetitive sequence and hypomethylated genic sequence is common, the striking large-scale binary distribution of genomic methylation has not been shown before in an algal species or other organisms and this unique genomic architecture may likely have influence on higher order chromatin structure in *C. cryptica*.

Methylation was minimal over gene sequences. There was a slight difference in the fraction of methylation in introns (3.23% on average) versus exons (4.14%) (Fig. 3c, d), a trend that is in agreement with that found in other organisms, including *P. tricornutum* [41]. Across the gene body, methylation gradually increased towards the 3′ direction and was lower 500 bp upstream compared to downstream (Additional file 1: Figure S6). Notably, there were higher levels of repetitive sequence in the 5′ upstream, hypomethylated region (Additional file 1: Figure S6), suggesting that these repeat elements are under a different control mechanism compared to the majority of repeats in the genome.

For genes with sufficient coverage to define a methylation status, 77% (averaged between both experimental conditions) were unmethylated and 23% were methylated (Additional file 1: Table S3). Methylated genes had an average methylation of 88% and had lower transcript levels with ~500 average FPKM across all transcriptomes. Unmethylated genes had an average methylation of 7% and had significantly higher levels of transcript (average ~4500 FPKM, Additional file 1: Table S3).

There was no linear correlation between fraction methylation and FPKM in either condition, however, three general populations appeared (Additional file 1: Figure S7). The first consisted of genes with very little transcript abundance (ranging from 0 to 10 FPKM), almost all of which (95%) were methylated (Additional file 1: Figure S7a, box i). The second population, similar to findings from [41], comprised the majority of genes which contained moderate to high transcript abundance and were largely unmethylated (Additional file 1: Figure S7b).

Lastly, and somewhat surprisingly, there appeared to be several genes with moderate to high FPKM levels with high levels of methylation (average FPKM values ranging from 4600 to 161,000, Additional file 1: Figure S7b, box iii, Additional file 2). Functional annotation of these genes indicated enrichment in genes with transporter activity and/or an interaction with the cell surface. These included a nitrate transporter (g14234.t1), silicon transporter (g21035.t1), sodium-dicarboxylate symporter (g11380.t1, g10492.t1), sulfate transporter (g6964.t1), and a bacterially derived integrin (g18136.t1). In addition, genes involved in nitrogen metabolism, including the above nitrate transporter, as well as nitrate reductase (g22809.t1) and glutamate synthetase (g14060.t1) were also in this subpopulation. Highly methylated, highly expressed genes may be indicative of a unique role DNA methylation plays in these specific cellular functions. Methylation status of these genes did not change under our experimental conditions. Additionally, the absence of DNA methylation based silencing in these genes may indicate other epigenetic regulatory mechanisms that enable active transcription are at play. An example is histone tail acetylation, which has been shown to have a strong activating effect in *P. tricornutum* regardless of methylation status [49].

Investigations in *P. tricornutum* identified methylation shifts in nitrate metabolism genes under depletion conditions [41, 49], indicating that DNA methylation may play a regulatory role in nitrate metabolism in diatoms. We identified no significant change in methylation of genes pertaining to silicon metabolism or lipid accumulation under our conditions, instead there was a globally significant correlation of methylation between the two conditions. This data suggests that DNA methylation does not control the cell's silicon status response in these conditions as is reported with nitrate metabolism in *P. tricornutum.* One explanation for these differences could be the fact that nitrogen stress on diatoms can cause severe cellular damage and detrimental effects, such as chlorosis and DNA and protein degradation [28, 50], whereas silicon starvation does not produce as severe phenotypes [32, 42, 51].

We also determined whether there was a relation between the methylation status of genes and their evolutionary origin. We applied the DarkHorse analysis to identify possible horizontally transferred genes, and identified 1723 genes (41% of total methylated genes) that contained a top BLAST hit to diatoms, however, in comparing the proportion of diatom-derived genes within the subset of methylated genes, this number is significantly depleted relative to the proportion of diatom-shared genes in the whole genome (72% of the total, Fig. 4a). There was enrichment in methylated genes with no known taxonomic match ('unknown,' 1956 genes, 47% of total methylated genes versus 20% of all genes, Fig. 4b). There was also enrichment of methylated Opisthokont genes (Fungi, Metazoan, Choanoflagellida, Ichthyosporea), which were primarily annotated for transposon processes (Fig. 4a; Additional file 2). Within the methylated subset of genes, there was a slightly higher proportion of foreign genes (i.e., viral, bacterial) as well as genes ancestral to the cell, but not core to diatom function, indicated by loss in other diatom genomes (groups SAR excluding diatoms, Archaeplastida, Fig. 4b).

In general, methylated genes in *C. cryptica* are not highly expressed and appear to be non-essential under our conditions. Furthermore, there appears to be a slight preference to methylate genes that are horizontally acquired, perhaps as a silencing strategy to prevent potential detrimental effects of the foreign gene. Alternatively, it could also indicate that the mechanism for foreign gene insertion in eukaryotic microalgae, which is largely unknown, is more likely to occur in hypermethylated regions than in hypomethylated regions of the genome (Fig. 4b).

Lastly, using the same definitions of gene methylation as in Additional file 1: Additional methods, we compared methylation levels of genes in *T. pseudonana* from [5, 45] to the 0 h methylome of *C. cryptica*. Only 139 genes (1% of the total) in *T. pseudonana* were methylated and out of 24 *T. pseudonana* methylated genes with orthologs in *C. cryptica*, only five were methylated in both diatoms. Shared methylated genes were annotated as reverse transcriptases (g23181.t1, g15433.t1), and transposase IS4 (g9451.t1). The significantly low number of genes with shared methylation status in the two diatoms suggests a conserved role for methylation in these centric diatoms to mask and silence mobile DNA elements. Furthermore, while DNA methylation does not appear to temporally regulate genes in *C. cryptica* under our conditions, the comparison of these two diatom methylomes showed that both were associated with transposable element silencing. It has been demonstrated in higher plants that transposable element silencing has effects on the expression of nearby genes [52]. The methylation of the few genes found in hypermethylated regions (Fig. 3b) in *C. cryptica* may have resulted from the methylation of a nearby transposable element insertion and any resulting changes in gene expression could have downstream effects on the metabolism and function of that organism.

Predicted subcellular localization of enzymes

An overarching theme in the study of diatom metabolism is the diversity of intracellular localization of enzymes in different organelles: the mitochondrion, cytosol, chloroplast, ER, and the periplastid compartment [22–24, 26,

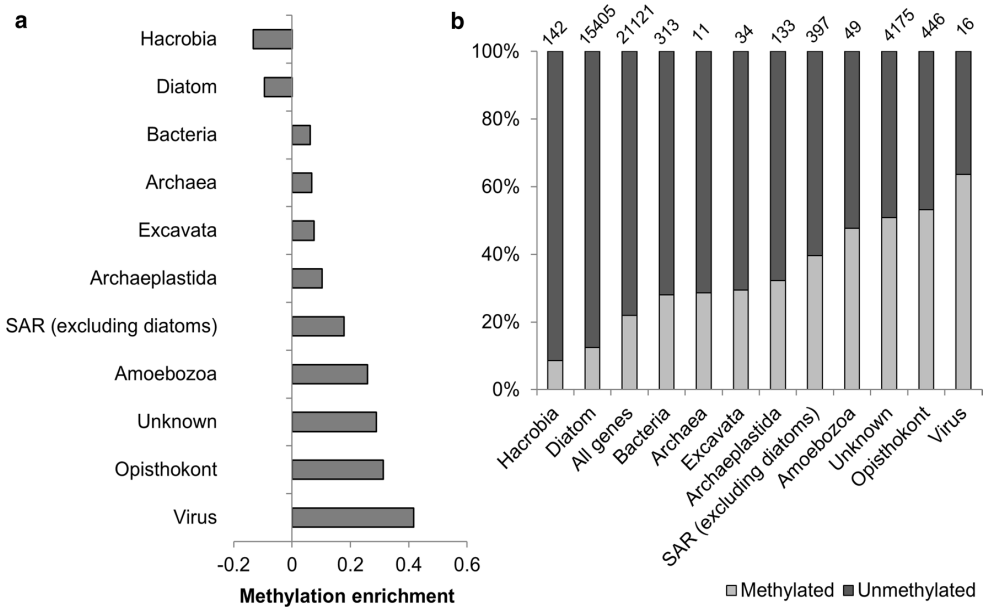


Fig. 4 Methylation based on evolutionary origin of genes from DarkHorse analysis. **a** Enrichment analysis of methylated genes given their taxonomic group. Enrichment of methylation within a taxonomic group was determined by subtracting the proportion of genes methylated within a taxonomic group from the proportion of genes methylated across the *C. cryptica* genome. Groups with numbers below 0 have a lower proportion of methylated genes in the taxonomic group compared to the proportion of this group in the overall genome. **b** Percent of methylated and unmethylated genes relative to taxonomic group. *Top numbers* are the total number of genes from that group in the genome

27, 53]. The latter of these compartments is unique to photosynthetic eukaryotes that have undergone a secondary (or greater) endosymbiotic event. In addition, specific to heterokonts, the ER is extended and surrounds the chloroplast. Using in silico tools as a means to map the location of metabolic processes in the cell provides a clearer description of metabolic function and transport of metabolites and proteins in and out of these organelles and information for accurate modeling. We used SignalP 3.0, TargetP, Predotar, ChloroP, HECTAR, and ASAFind to predict which proteins are putatively targeted to the organelles in *C. cryptica* (Fig. 5; Additional file 1: Additional methods; [54–58]). The first four in silico organellar prediction programs are developed for higher plants and other eukaryotes, but we have found their collective use to be suitable for some basic understanding of metabolic compartmentation in diatoms, particularly when overlaid with coordinate expression patterns of particular genes [27, 32]. The latter two programs, HECTAR and ASAFind are specific for heterokonts, and take advantage of the conserved amino acid sequence motif (ASA-FAP) which is part of the N-terminal bipartite signal peptide that is unique to these genera of algae. Given sequence manipulations [22–24] that have identified specific

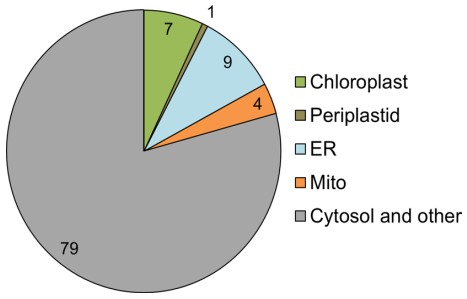


Fig. 5 In silico targeting predictions of all nuclear gene models in *C. cryptica*. Percentages are listed in the chart. See also Additional file 2, subcellular targeting. targeting predictions used for this specific analysis were SignalP 3.0, TargetP, ChloroP, HECTAR, Predotar, and ASAFind [43, 44, 55–58]

amino acid substitutions within that conserved motif resulting in altered organellar targeting of an enzyme, it is feasible to use these prediction programs to hypothesize which proteins are partitioned into the chloroplast versus periplastid compartment. Because of the uniqueness of the periplastid compartment, we focused our

analysis on proteins targeted there. Targeting of enzymes to other compartments will be described in the context of the metabolic processes analysis below. We identified a total of 120 proteins with targeted prediction to the periplastid compartment (Additional file 3).

Moog and colleagues [24] characterized the predicted proteome of the periplastid compartment (PPC) in *P. tricornutum* and concluded that some important cellular functions occurred there, specifically with regard to transport and protein import processes into the chloroplast, but that very few protein components of housekeeping biochemical pathways were present. Using direct localization approaches, they also documented that only 55% of proteins predicted to be targeted to the PPC actually were confirmed to be there, highlighting the poor state of predictive programs, perhaps due to transmembrane domains near the N-terminus, or mis-predicted signal cleavage sites.

Our analysis generally corroborated the findings of [24], with some exceptions, and potential new findings (Additional file 3). We identified one protein (g12899.t1) with an ANTH domain which may be putatively involved in clathrin assembly, as well as a dynamin, which is likely involved in chloroplast division (g1668.t1), suggesting that components of clathrin-mediated vesicular trafficking may be present in the PPC. Another protein, g23203.t1, is a member of the S2P/M50 family of regulated intramembrane proteolysis proteases, which use proteolytic activity within the membrane to transfer information across as a means to integrate gene expression with physiologic stresses occurring in another cellular compartment [59]. Mechanisms of cross-talk between the cytoplasm and chloroplast of diatoms have not been well addressed in the literature, but of necessity, would require the involvement of the PPC.

Lastly, phosphoenolpyruvate carboxylase (PEPC—g7839.t1) and several carbonic anhydrases were predicted to be PPC targeted, consistent with a mechanism to recapture CO_2 lost from the chloroplast by inefficient carbon fixation assimilation and incorporate it into oxaloacetate. Experimental validation using genetic constructs to confirm localization of predicted enzymes in the PPC and other organelles is needed to confirm in silico analyses. Nevertheless, the prediction programs, particularly on a proteome with accurate prediction of correct N-termini based on RNAseq data, offer a starting point to further characterize the significance of the PPC in diatoms.

Comparative analysis of glycolysis, gluconeogenesis, and the pyruvate hub

Glycolysis, the catabolism of glucose to produce pyruvate, and the reverse pathway gluconeogenesis, are core metabolic pathways which process photosynthetically fixed carbon into compounds and energy for use by the cell. Pyruvate metabolism is involved in the distribution of carbon to different cellular processes, which include biosynthesis of compounds and energy generation. While genes and predicted subcellular localization of steps in these processes were generally conserved between the oleaginous *C. cryptica* and relatively non-oleaginous *T. psuedonana,* there were several important differences in the architecture of the pyruvate hub in three different intracellular locations.

Cyclotella cryptica has an additional copy of plastid-localized pyruvate kinase (Figs. 6, 7i), a key glycolytic regulatory enzyme that catalyzes the unidirectional conversion of phosphoenolpyruvate (PEP) and ADP to pyruvate and ATP. In addition, the gene encoding plastid-localized PEP synthase (PEPS) which catalyzes the reciprocal reaction to produce PEP from pyruvate, is heavily methylated (96% fraction methylation) with low transcript (average FPKM 0.61) abundance. This suggests that under the conditions tested, PEPS may be silenced by methylation (Fig. 7ii), and can be assumed to be non-functional. In contrast, the plastid PEPS gene is homologous to the single copy of PEPS found in *T. pseudonana*, which shows moderate levels of expression and is unmethylated (0.19% across gene body) under similar experimental conditions [32, 45].

In *C. cryptica,* there are two isozymes of PEPS localized in the cytosol not found in *T. pseudonana* (Figs. 6, 7iii). Generation of PEP from pyruvate in the cytosol instead of the plastid would separate a reaction that utilizes pyruvate from a reaction that produces pyruvate, potentially improving the efficiency of pyruvate utilization for processes such as fatty acid synthesis in the plastid of *C. cryptica*. The methylation of PEPS, along with additional copies of PEPS in the cytosol might further allow for an increased degree of fine-scale regulation of this important node of metabolism. The extra copy of plastid-localized PK in *C. cryptica* relative to *T. pseudonana* could indicate a greater potential for pyruvate production in *C. cryptica*, possibly translating into a greater fatty acid synthesis capability (Fig. 7).

There were also differences in the mitochondrial pyruvate hub between the two species. *T. pseudonana* lacks a mitochrondrial-localized pyruvate carboxylase (PYC), which converts pyruvate to oxaloacetate (OAA), yet possesses a cytosolic copy, whereas *C. cryptica* has a mitochondrial but not cytoplasmic PYC (Figs. 6, 7iv). This suggests that *C. cryptica* can produce OAA via pyruvate in the mitochondria, but *T. pseudonana* would have to generate it elsewhere then import into mitochondria, via a malate–aspartate shuttle. In addition to the differential requirement for transport, metabolic drivers for possible

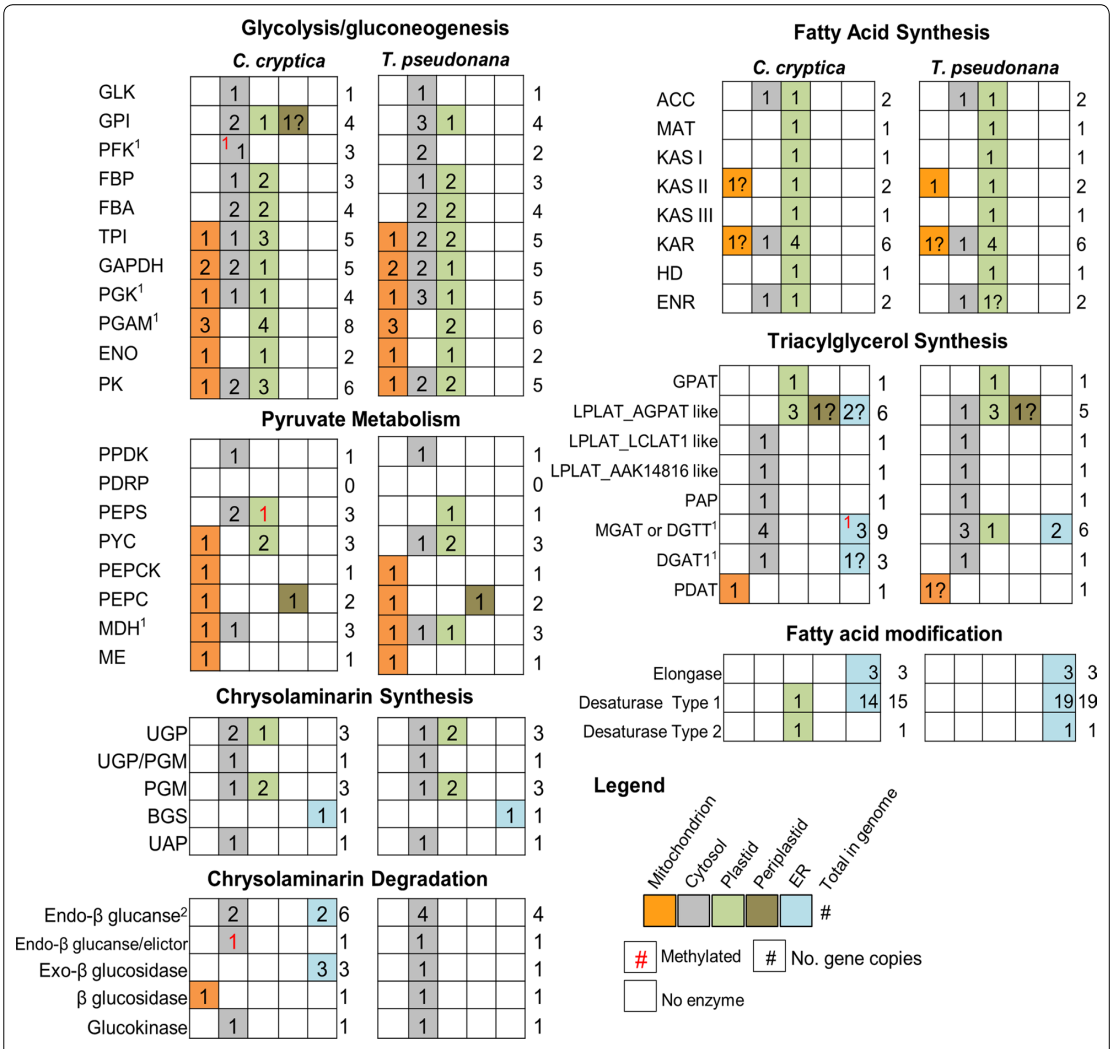


Fig. 6 Comparative analysis of key carbon metabolic pathways between *C. cryptica* and *T. pseudonana*. *Different colored boxes* denote subcellular location of a given enzyme based on bioinformatic targeting predictions. *Number* within the *box* indicates how many enzyme copies are targeted to that location. *Numbers* to the *right of boxes* are total gene copy number found in the genome. *Question marks* indicate weak targeting and/or inconsistencies between targeting prediction programs. *Superscript numbers* on enzymes indicate the number of genes found with partial sequence in *C. cryptica* at the N-terminus, thereby not containing targeting prediction. *Red numbers* are methylated genes. In the case of MGAT or DGTT, one of the four ER predicted genes is methylated. *AAK14816-like* putative glycerol-3-phosphate acyltransferase, *ACC* acetyl-CoA carboxylase, *AGPAT* 1-acyl-glycerol-3-phosphate acyltransferase, *BGS* 1,3 β-glucan synthase, *DGAT* diacylglycerol acyltransferase, *DGTT* diacylglycerol acyltransferase type 2, *ENO* enolase, *ENR* enoyl-ACP reductase, *FBA* fructose bisphosphate aldolase, *FBP* fructose 1,6 bisphosphatase, *GAPDH* glyceraldehyde 3-phosphate dehydrogenase, *GLK* glucokinase, *GPI* glucose-6-phosphate isomerase, *GPAT* glycerol-3-phosphate acyltransferase, *HD* 3-hydroxyacyl-ACP dehydratase, *KAR* 3-ketoacyl-ACP reductase, *KAS* 3-ketoacyl-ACP synthase, *LCLAT1* lysocardiolipin acyltransferase 1, *LPLAT* lysophospholipid acyltransferase, *MAT* malonyl-CoA-ACP transacylase, *MDH* malate dehydrogenase, *ME* malic enzyme, *MGAT* monoacylglycerol acyltransferase, *PAP* phosphatidic acid phosphatase, *PYC* pyruvate carboxylase, *PDRP* pyruvate phosphate dikinase regulatory protein, *PEPC* phosphoenolpyruvate carboxylase, *PEPCK* phosphoenolpyruvate carboxykinase, *PEPS* phosphoenolpyruvate synthase, *PFK* phosphofructokinase, *PGAM* phosphoglycerate mutase, *PGK* phosphoglycerate kinase, *PGM* phosphoglucomutase, *PK* pyruvate kinase, *PPDK* pyruvate phosphate dikinase, *TPI* triose phosphate isomerase, *UAP* UDP-*N*-acetylglucosamine pyrophosphorylase, *UGP* UTP-glucose-1-phosphate uridylyltransferase

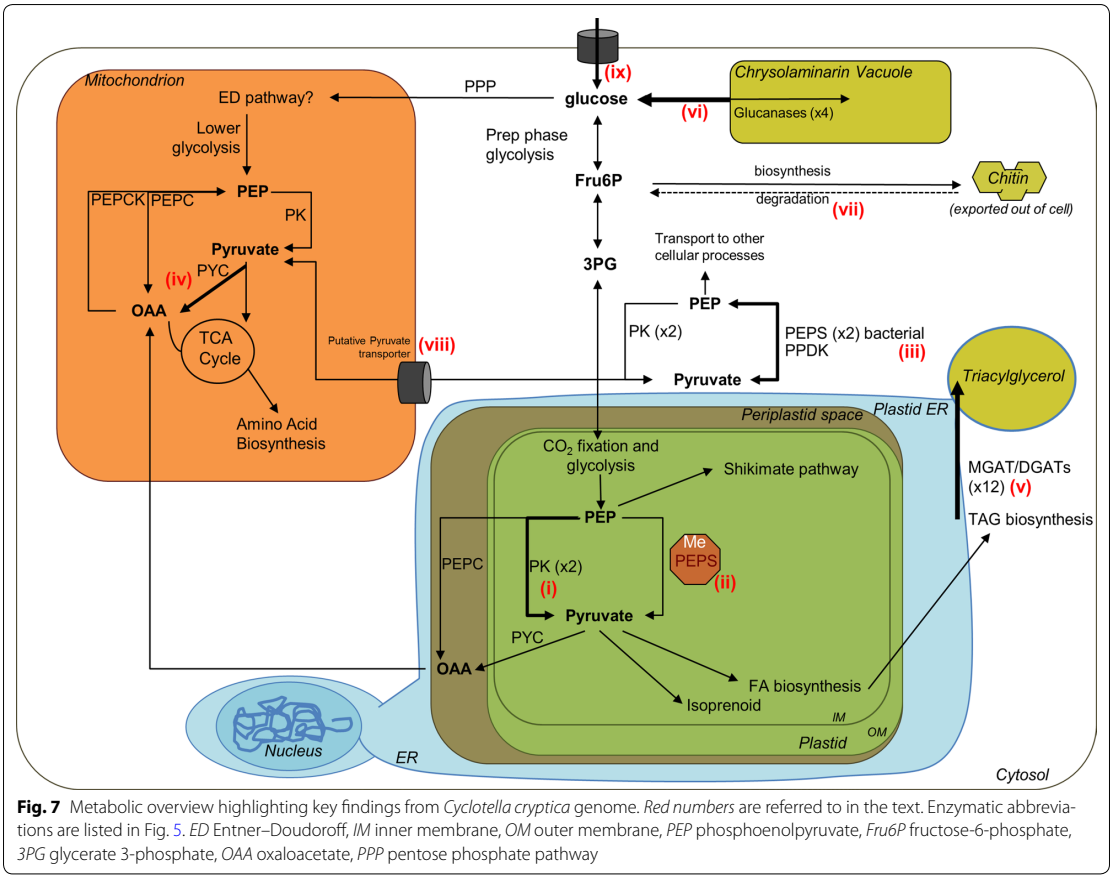


Fig. 7 Metabolic overview highlighting key findings from *Cyclotella cryptica* genome. *Red numbers* are referred to in the text. Enzymatic abbreviations are listed in Fig. 5. *ED* Entner–Doudoroff, *IM* inner membrane, *OM* outer membrane, *PEP* phosphoenolpyruvate, *Fru6P* fructose-6-phosphate, *3PG* glycerate 3-phosphate, *OAA* oxaloacetate, *PPP* pentose phosphate pathway

mitochondrial gluconeogenesis may differ in the two species, since in *C. cryptica,* gluconeogenesis could be initiated within the mitochondrion by conversion of pyruvate to PEP by the combined action of PYC and PEPCK.

Lipid metabolism

The genetic basis for enhanced lipid productivity in *C. cryptica* was investigated by identifying genes for fatty acid and TAG biosynthesis, as well as fatty acid modification biosynthesis, and comparing this repertoire with that found in *T. pseudonana*. An identical suite of plastid-localized essential fatty acid biosynthesis orthologs from *T. pseudonana* were found in *C. cryptica* (Fig. 6). There were differences in the TAG biosynthesis enzyme inventory between the two diatoms, such as the presence of several additional genes found in *C. cryptica* and subcellular localization predictions (Fig. 6). An additional three genes possessing an LPLAT_MGAT-like domain found in either MGAT or DGTT enzymes were present in *C. cryptica*. One of these putative MGAT/DGTT (g3706.t1) had highly abundant transcripts (average FPKM of 23,833). Another putative MGAT/DGTT (g21947.t1) is a homolog to the *P. tricornutum* DGTT2b, which has been tested as a target for genetic manipulation to increase TAG content [60, 61]. There were also two additional copies of DGAT1: one with predicted targeting to the ER (g23184.t1), and the other (g9565.t1) was a partial sequence, hence targeting could not be determined. The additional copies of these enzymes in *C. cryptica* may indicate an increased ability to generate TAG and/or allow for finer control over TAG production in the case that these enzymes differ in their specificity, regulation or kinetic properties (Fig. 7v).

Omega-3 long chain polyunsaturated fatty acids (LC-PUFAs) are a high-value product that is of interest for nutritional purposes [5]. *Cyclotella cryptica* synthesizes abundant LC-PUFAs including over 16% eicosapentaenoic acid (EPA) and 4% docosahexaenoic acid (DHA) of the total fatty acids; generally a higher level than in other algal species [5]. We investigated if the higher capacity to produce these biotechnologically relevant fatty acids

was due to an increase gene number within the fatty acid elongases and desaturases, which encode for the enzymes that catalyze the biosynthesis of LC-PUFAs. Overall, elongases and desaturases potentially involved in LC-PUFA synthesis were highly conserved between *C. cryptica* and *T. pseudonana* (Fig. 6). Although *C. cryptica* contained fewer copies of type 1 desaturases, it had all the types including Δ5, Δ6, Δ8, Δ9, Δ11 and Δ12. Both diatoms contain a single copy of a Δ4 desaturase that catalyzes the addition of the sixth double bond to docosapentaenoic acid (22:5) chain to make DHA (C22:6) [20]. The existence of the same types of elongases and desaturases in both diatoms suggest that the greater ability of *C. cryptica* to produce EPA compared to *T. pseudonana* is not due to an increase in isozymes in the genome and could be more dependent on the flux of carbon from upstream processes into this pathway.

Carbohydrate biosynthesis and degradation

Chrysolaminarin, the main form of storage carbohydrate in diatoms, is a water soluble β-(1-3) linked glucan, and is stored outside of the plastid in a large vacuole [36, 62]. In diatoms, genes involved in biosynthesis and degradation of chrysolaminarin have been annotated, and the biochemistry of key enzymatic steps has been characterized [10, 53, 63]. The gene repertoire for chrysolaminarin biosynthesis enzymes is highly conserved between *T. pseudonana* and *C. cryptica* in terms of sequence similarity and predicted targeting (Fig. 6). Like other stramenopiles, *C. cryptica* possesses a fused cytosolic phosphoglucomutase/UTP-glucose-1-phosphate uridyltransferase, which carries out the first two-enzymatic steps toward chrysolaminarin production [53]. Both *T. pseudonana* and *C. cryptica* also possess three individual copies of phosphoglucomutase and UTP glucose-1-phosphate uridyltransferase similarly distributed between the plastid and cytosol. The metabolic significance of the fused gene and multiple single gene copies leading to chrysolaminarin production is unknown, yet might provide metabolic flexibility by creating several routes toward synthesis that differ in organellar location (plastid vs. cytosol), carbon transporter specificity out of the plastid (Glu-6P versus UDP-glucose), and enhanced efficiency or substrate channeling (fused protein versus single reaction proteins). While these enzymatic steps have an intriguing metabolic flexibility, the final step in chrysolaminarin production, catalyzed by a beta-glucan synthase, does not. The β-glucan synthase is present as a single copy and highly conserved in all sequenced diatom genomes.

In contrast to the chrysolaminarin biosynthetic pathway, genes coding for enzymes involved in the degradation of chrysolaminarin were poorly conserved between *T. pseudonana* and *C. cryptica* with an average low percent identity between homologs (51%). The *C. cryptica* genome encodes four additional enzymes putatively involved in chrysolaminarin breakdown compared to *T. pseudonana* (two additional endoglucanases, two additional exoglucosidases, Figs. 6, 7vi). Of the three ER predicted exoglucosidases, two were found only in *C. cryptica,* presumably recently acquired by horizontal transfer. The most highly expressed exoglucosidase in *C. cryptica* (g19489.t1) is unique, with low similarity compared with those of higher plants and fungi. In addition to the two-glycosyl hydrolase domains responsible for carbohydrate breakdown, the gene also contains two fascin-like domains, putatively involved in actin crosslinking, which implies a positioning component to its function. The increased number of chrysolaminarin degradation enzymes in *C. cryptica* compared to *T. pseudonana* may indicate a greater capacity to break down β-1,3 glucans, allowing for more carbon substrate to be used for the production of triacylglycerol or other carbon products.

Chitin biosynthesis and degradation

Chitin, or poly β-(1,4) linked *N*-acetylglucosamine, is the most abundant polymer in the ocean, and second most on the planet [64, 65]. The exoskeleton of arthropods and insects is made of chitin, and chitin is also produced by many diatoms where it has been proposed to control buoyancy and/or be involved in silica cell wall synthesis [66–68]. Chitin is estimated to comprise over 30% of the mass of the cell in some diatom species [69], which is a significant reservoir for cellular carbon that could be shunted into other pathways to facilitate productivity in a controlled cultivation environment. However, little is known about the cellular role of chitin, the biosynthetic pathway is poorly characterized, and annotation of relevant genes in diatoms is incomplete. In *C. cryptica*, most enzymes involved in synthesis and degradation of chitin are predicted to be cytosolic or ER targeted (Fig. 8)—the latter may indicate extracellular secretion, plasma membrane localization, or an association with the silica deposition vesicle.

Chitin fibrils are apparently located between the silica frustule and plasma membrane [69], yet the intracellular site of chitin production is unknown, therefore we investigated the targeting of enzymes predicted to be involved in this process for additional insight. The first and rate-limiting enzyme of the chitin synthase complex is glucosamine fructose 6-P aminotransferase, of which *C. cryptica* contains four putative copies (Fig. 8a). Three of those copies are predicted by several programs to have mitochondrial targeting, perhaps suggesting direct utilization of mitochondrial-produced glutamine from the glutamate–glutamine cycle. However, fructose-6P

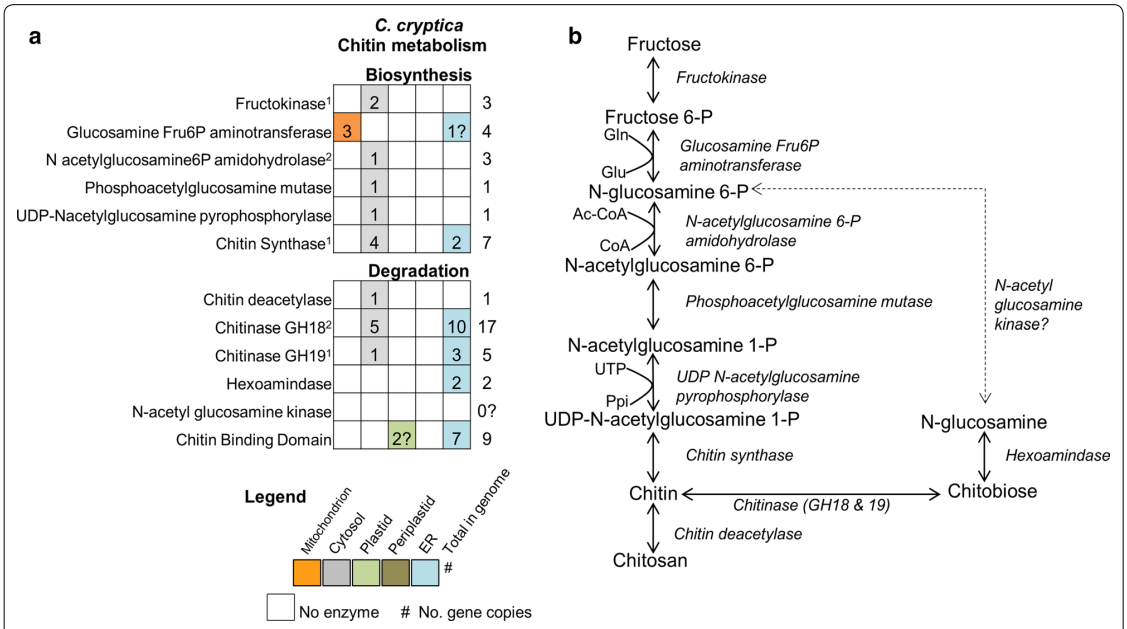


Fig. 8 Chitin metabolism in *C. cryptica*. **a** Genes and their predicted subcellular localization. Organizational schematic as described in Fig. 6. **b** Hypothetical chitin biosynthesis and degradation pathway in *C. cryptica*. *Dashed line* indicates that the enzyme is not found in *C. cryptica*

is not produced in the mitochondria and would need to be transported there. *Cyclotella cryptica* GFATs all also have less well supported ER signal peptides, which may also indicate targeting through the ER to a different cellular location. In both *T. pseudonana* and *C. cryptica* there are three steps in the pathway represented by a single isozyme, all with cytosolic targeting prediction. The final step of chitin biosynthesis polymerizes the *N*-acetyl-D-glucosamine molecules by chitin synthase (CS) to produce chitin. There are five full copies of CS, which contain the chitin_synth_C domain, and two partial sequences in *C. cryptica* (Fig. 8a). These Division 2 chitin synthases have been found only within *Thalassiosirales* diatoms, of which *C. cryptica* is a member, and were distinct from the Division 1 enzymes found in the pennate diatoms [70, 71]. Since *Thalassiosirales* all secrete chitin fibrils, whereas pennate species do not; these chitin synthases could be involved with fibril formation.

Cyclotella cryptica has a nearly complete chitin degradation pathway including a substantial number of chitinases, 22 in total, which perform the first step in chitin degradation (Fig. 8). Similar numbers were reported in *T. pseudonana* [68]. Most of the chitinases are shared within the *Thalassiosirales* group and have similarity to either bacterial (13 copies) or fungal (7 copies) chitinases. It is unknown when or why diatoms would break down chitin, but the number of chitinase genes combined with their general transcript abundance in *C. cryptica*, suggests that they play an important if not as yet elucidated role in cellular processes. Four chitinases also contain one or two peritrophin A domains. In insects and other multicellular organisms, peritrophin is a protein embedded within a chitinous membrane (the peritrophin matrix) which separates digested food from the midgut epithelium where it is hypothesized to aid in digestion, protect against pathogens, and provide mechanical support [72]. In *T. weissflogii*, freeze fracture images of the space between the plasma membrane and silica frustule revealed structures with the appearance of chitin fibrils [69], which raises the possibility that peritrophin-containing chitinases in *C. cryptica* may be found in a similar matrix as an organic component of the cell wall. This might suggest that a dynamic processing of chitin, involving synthesis and degradation, occurs in this location. In addition to the chitinases, other chitin binding genes in *C. cryptica* possessed Peritrophin A domains which have also been identified in *T. pseudonana* [73].

Two recently duplicated gene copies encoding the subsequent step in chitin degradation, the breakdown of chitobiose catalyzed by hexosaminidase, are ER targeted in *C. cryptica* (g14774.t1, g22243.t1). A homolog could not be identified in *T. pseudonana*. Furthermore, even with a manual search, we were unable to identify a kinase that phosphorylates the acetyl glucosamine molecule to

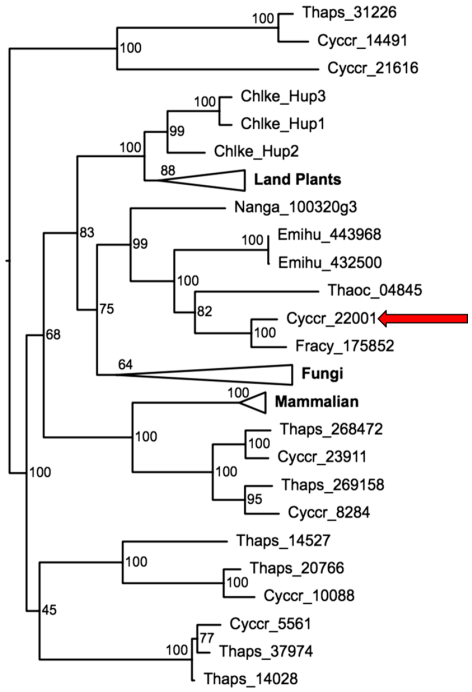


Fig. 9 Phylogenetic analysis of hexose transporters in *C. cryptica (Cyccr)* and *T. pseudonana (Thaps)*. All genes contain PF00083 Sugar (and other) transporter domain. Highlighted gene is hypothetical heterotrophic glucose transporter in *C. cryptica* and with nearest matches to that transporter found in the diatoms *T. oceanica, F. cylindrus,* the haptophyte *E. huxleyi,* and stramenopile *N. gaditana.* Phylogeny was generated using default parameters in RAxML_GUI v1.3 and is mid-point rooted. Abbreviations and accession numbers are found in Additional file 4

production because it allows investigation of cellular metabolism and enables enhancement of the species' characteristics to increase production efficiency. While successful nuclear transformation of *C. cryptica* was achieved during the ASP [12], there has been little molecular work since. To improve the tools available for *C. cryptica,* we further adapted the nuclear transformation procedure, tested two *T. pseudonana* expression vectors and constructed one using *C. cryptica* sequences (Additional file 1: Additional methods). Two vectors allow for constitutively high expression, one of which uses native *C. cryptica* ribosomal protein L41 promoter and terminator sequences, while the second vector was derived from the *T. pseudonana fcp* gene [13]. The third vector, utilizing the *T. pseudonana* nitrate reductase (NR) promoter and terminator enables conditional expression, being induced upon transfer of nitrogen source from ammonia to nitrate [85]. Imaging of GFP under control of the NR promoter in *C. cryptica* grown in nitrate and ammonium is shown in Fig. 10. The larger size of *C. cryptica* (8–10 µm) compared to *T. pseudonana* (4–5 µm) and *P. tricornutum* (4–5 µm) may facilitate microscopy for identification of subcellular localization of enzymes. While vectors with native promoters and native sequence are important in some contexts, the successful utilization of pre-existing molecular tools from *T. pseudonana* in *C. cryptica* reduces effort and could facilitate development of specific applications.

Conclusions

Determination of the genome sequence of *C. cryptica* elucidates metabolic features that underlie its high productivity and lipid accumulation ability, and the application of state-of-the-art approaches to further develop this species featured in the Aquatic Species Program for biofuel/bioproduct production [4]. Diatoms in general have attractive characteristics for large-scale biofuel/bioproduct production. In addition to their innate high productivity, their distinctive lipid induction stimulus by silicon starvation leads to rapid and substantial increases in triacylglycerol levels [6, 26, 42] without detrimental effects on cellular protein content that would typically occur under nitrogen starvation conditions used in other microalgae. Thus, they could serve as both a biofuel precursor and protein source in a production system.

A dominant feature of the *C. cryptica* genome is the presence of large regions of repeat sequences encoding a variety of types of mobile genetic elements (Additional file 1: Table S2) which are highly methylated, presumably as a means to inhibit expansion of these elements. In a global sense, in the conditions tested, methylation does not appear to play a significant role in regulating metabolic gene expression, however, there are notable and important exceptions in the examples of highly methylated yet highly expressed genes (Additional file 1: Figure S7), and the chloroplast-localized PEPS gene which is highly methylated and not expressed (Fig. 7ii). There are a few, yet significant, differences in gene content and metabolic topology when comparing *C. cryptica* and *T. pseudonana*, which may underlie the former's high productivity. In particular, pyruvate metabolism is substantially different, with *C. cryptica* potentially being able to process pyruvate with greater efficiency because of compartmental separation of distinct reaction steps comparing the chloroplast and cytoplasm, and intercompartmental processing in the mitochondria, avoiding a transport step (Figs. 6, 7). Additional isozymes involved in TAG synthesis and carbohydrate breakdown are present in *C. cryptica*, which may improve carbon flux into TAG synthesis (Figs. 6, 7v, vi). Characterization of these metabolic differences, and those involved in other potential carbon sinks such as chitin, identifies a variety of gene

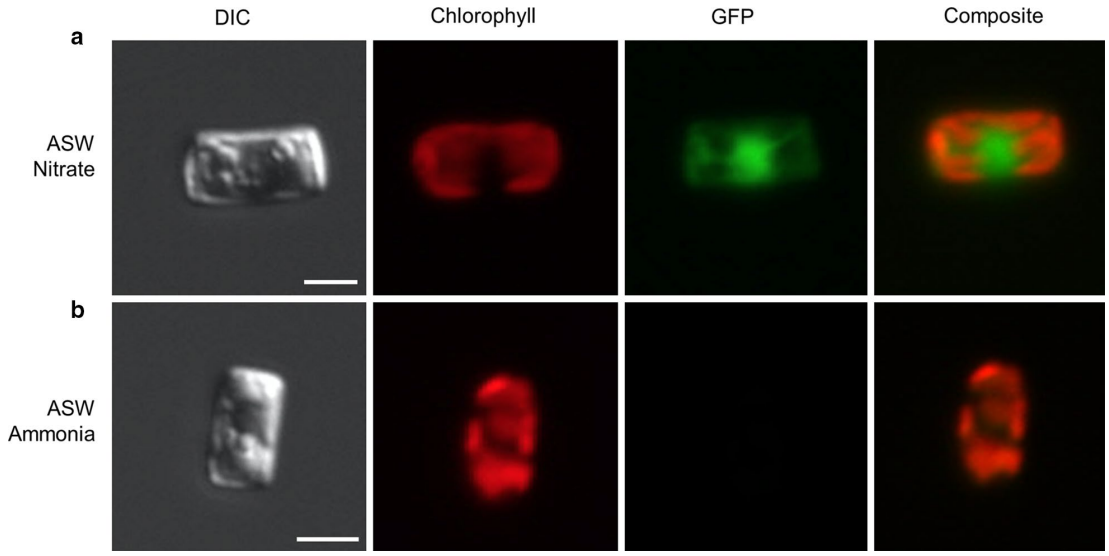


Fig. 10 Genetic manipulation in *C. cryptica*. Conditional expression using the nitrate reductase promoter. **a** GFP expression in ASW media, with nitrate as the nitrogen source. **b** No GFP expression in repressive modified ASW media, with ammonia as the nitrogen source. Chlorophyll and GFP are falsely colored *red* and *green*, respectively. Composite images are 3D reconstructions. *Scale bars* 5 µm

targets to manipulate to directly test these hypotheses, and to understand how the differences in carbon flux may affect productivity, with a goal of ultimately improving productivity. The heterotrophic capability of *C. cryptica* may further aid in increasing productivity, either by enabling heterotrophic production or reducing respiratory losses at night from photosynthetically-fixed carbon. The application of state-of-the-art genetic tools to *C. cryptica* (Fig. 10), and continued development of such, will enable reprogramming of the metabolism otherwise adapted for environmental survival to improve productivity in the relatively controlled conditions for large-scale production. By bringing the promising productivity characteristics of *C. cryptica* in line with current day approaches to improve productivity, its potential as a model biofuel/bioproduct organism can be maximized.

Methods

Source of genetic material

Genomic DNA for genome and methylome sequencing was isolated from a clonal culture of *C. cryptica*, CCMP332. This strain of *C. cryptica* was originally isolated in June of 1956 from the West Tisbury Great Pond (41.355° N, 70.655° W) in MA, USA and maintained and by the National Center for Marine Algae and Microbiota (NCMA, formerly Provasoli–Guillard National Centre for Culture of Marine Phytoplankton CCMP). CCMP332 has been maintained in the lab since 2008 under a 12:12 light–dark cycle at 18 °C with an illumination of 150 µmol/m^2 s light. For genomic DNA isolation, *C. cryptica* was plated onto artificial seawater (ASW) agar plates [86] under the presence of 1:1000 dilution of penicillin/streptomycin (Gibco, catalog no. 15140-122) to minimize bacterial contamination. A single colony was picked and grown in liquid ASW for scale up. 1 L liquid cultures were mixed using a magnetic stir plate (250 rpm), bubbled with air, and grown in continuous light (150 µmol/m^2 s) for 3 days to early stationary phase (~1.2 × 10^6 cells/mL) before harvesting for DNA isolation. To isolate DNA for methylome analysis, a 2.5 L culture was grown using above conditions to mid exponential phase (5 × 10^5 cells/mL). On the third day, 300 mL of liquid culture was harvested for the $T = 0$, silicon replete time point for genomic DNA isolation. 1500 mL of cells was harvested by centrifugation and placed into ASW deprived of silica (Si-ASW) as described in [42]. Cells were monitored for 48 h in Si-ASW before harvesting 300 mL DNA for Si-$T = 48$ methylome. Cells were stained for 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene (BODIPY 493/503, Molecular Probes) as described in [42] and imaged using a Zeiss Axio Observer Z1 Inverted Microscope at a 63× objective. Cell growth was monitored using a Neubauer hemocytometer with counts done in duplicate or quadruplicate.

Functional annotation

Functional annotation of genes from *C. cryptica* and the five other reference diatom genomes was performed

using PhyloDB, a comprehensive database of proteins at JCVI. KEGG, KO, KO Pathway, EC annotations were assigned using TimeLogic® Tera-BLAST™ algorithm (Active Motif Inc., Carlsbad, CA, USA), *e* value threshold 1e−5; Pfam/TIGRfam using Hmmer v 3.1b2 (http://hmmer.org), and transmembrane domains with TMHMM 2.0c [87]. Genes from all six genomes were clustered into orthologous groups or larger gene families using MCL [88] with inflation parameter = 2.0. Perl scripts were developed to aggregate cluster annotations based on shared functional content of genes. Methods for prediction of lateral gene transfer, RBH pairs, phylogenetic analysis, and subcellular localization prediction are described in Additional file 1.

Additional files

Additional file 1. Additional methods and figures. **Figure S1.** (a) *Cyclotella cryptica* nuclear assembly overview broken down by minimum contig size (Kbp) versus total Mbp of all contigs (left) and minimum contig size versus the total number of contigs (right) and (b) genomic sequencing data. **Table S1.** Statistics from different gene model prediction pipelines. The pipelines are: (1) FGENESH gene predictions, Diatom training set, (2) Augustus gene predictions V1, *Chlamydomonas* training set, no RNAseq data, (3) Augustus gene predictions V2, FGENESH100 training set, RNAseq data (4) Augustus gene predictions V3, 'self' trained, RNAseq data (5) MAKER gene predictions, (Augustus self trained + FGENESH + GeneMarkES). Details are presented in Additional methods. Data presented includes all predicted models regardless of read counts from RNAseq data and prior to removing apparent duplicate contigs (Additional methods, Genome Library Construction and Sequencing). **Table S2.** Repeat sequences in *C. cryptica* identified using (a) RepBase data and (b) RepeatModeler data. **Figure S2.** Phylogenetic comparison of diatom species with sequenced genomes. (a) 18S sequence comparison from [91], accession numbers are listed in Additional file 4. (b) Reciprocal best blast hits comparison. *C. cryptica* and *T. pseudonana* are the most similarly related centric diatoms with available genomic data. **Figure S3.** Per-cytosine fraction methylation. (a) 0 h, silicon replete sample (b) 48 h, silicon-deplete sample. All sites shown have a read coverage greater than or equal to 4. **Figure S4.** Per-site methylation differences between silicon replete and deplete conditions. (a) Histogram showing absolute value of the difference between 48 and 0 h. (b) Scatter plot showing the slope between fraction methylation of genes when comparing the two conditions is linear, outliers are apparent. **Figure S5.** Fraction methylation across the eight largest contigs in *C. cryptica*. Blue is 0 h silicon replete sample, red is 48 h silicon-deplete sample. AUGUSTUS V3 models (labeled "Gene") are depicted in black while repeat modeler data are depicted in gray. X-axis is number of kilobase pairs in the contig. **Figure S6.** Binned repeats (a) and fraction methylation across gene body (b) in *C. cryptica*. **Table S3.** Summary of gene methylation (AUGUSTUS V3 models) in both experimental conditions. **Figure S7.** Gene methylation relative to gene expression. (a) Three populations emerged (red boxes i–iii) when comparing average fraction methylation to Log2 FPKM. (b) Gene count binned according to FPKM and shaded according to methylation status. Line depicts the proportion of methylated genes per bin. Data shown is for silicon replete, 0 h condition. **Figure S8.** Vector map for ribosomal protein L41 construct for *C. cryptica*.

Additional file 2. OrthoMCL and functional annotation, methylome, DarkHorse, subcellular targeting. ND (no data) indicates data not sufficient enough to determine a value.

Additional file 3. Periplastid predicted genes.

Additional file 4. Triose phosphate transporters, accession numbers for sequences used in Fig. 9 and Additional file 1: Figure S2.

Abbreviations

AAK14816-like: putative glycerol-3-phosphate acyltransferase; ACC: acetyl-CoA carboxylase; AGPAT: 1-acyl-glycerol-3-phosphate acyltransferase; ASP: Aquatic Species Program; ASW: artificial Seawater; BGS: 1,3 β-glucan synthase; DGAT: diacylglycerol acyltransferase; DGTT: diacylglycerol acyltransferase type 2; ED: Entner Doudoroff; ENO: enolase; ENR: enoyl-ACP reductase; FBA: fructose bisphosphate aldolase; FBP: fructose 1,6 bisphosphatase; Fru6P: fructose-6-phosphate; GAPDH: glyceraldehyde 3-phosphate dehydrogenase; 3PG: glycerate 3-phosphate; GLK: glucokinase; GPI: glucose-6-phosphate isomerase; GPAT: glycerol-3-phosphate acyltransferase; HD: 3-hydroxyacyl-ACP dehydratase; KAR: 3-ketoacyl-ACP reductase; KAS: 3-ketoacyl-ACP synthase; LCLAT1: lysocardiolipin acyltransferase 1; LPLAT: lysophospholipid acyltransferase; MAT: malonyl-CoA-ACP transacylase; MDH: malate dehydrogenase; ME: malic enzyme; MGAT: monoacylglycerol acyltransferase; NR: nitrate reductase; OAA: oxaloacetate; PAP: phosphatidic acid phosphatase; PDRP: pyruvate phosphate dikinase regulatory protein; PEP: phosphoenolpyruvate; PEPC: phosphoenolpyruvate carboxylase; PEPCK: phosphoenolpyruvate carboxykinase; PEPS: phosphoenolpyruvate synthase; PFK: phosphofructokinase; PGAM: phosphoglycerate mutase; PGK: phosphoglycerate kinase; PGM: phosphoglucomutase; PK: pyruvate kinase; PPC: periplastid compartment; PPDK: pyruvate phosphate dikinase; PPP: pentose phosphate pathway; PYC: pyruvate carboxylase; RBH: reciprocal best BLAST hit; SAR: Stramenopile, Alveolata, Rhizaria supergroup; TAG: triacylglycerol; TPI: triose phosphate isomerase; UAP: UDP-*N*-acetylglucosamine pyrophosphorylase; UGP: UTP-glucose-1-phosphate uridylyltransferase.

Authors' contributions

JCT, SDG, MM processed genomic material and assembled DNA and RNA libraries. SJC and DAL assembled the genome, methylomes, and performed genome annotation pipelines. MT, AJ, SP provided additional bioinformatics support. JCT, OC and MH performed further annotation of metabolic processes. OG performed genetic engineering experiments. SRS, JPM provided OrthoMCL and annotation data. EEA, AEA, SSM, MP, and MH provided financial support. JCT, SJC, DAL, SSM, MP, and MH conceived the study, oversaw the project and wrote the manuscript. All authors read and approved the final manuscript.

Author details

[1] Scripps Institution of Oceanography, University California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0202, USA. [2] Institute for Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA. [3] J. Craig Venter Institute, 4120 Capricorn Lane, La Jolla, CA 92037, USA. [4] Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA.

Acknowledgements

The authors would like to thank Dr. Roshan Shrestha for his advice and assistance on genetic manipulation experiments.

Competing interests

The authors declare that they have no competing interests.

Availability of supporting data

The dataset(s) supporting the results of this article are available in the Additional files 1, 2, 3, 4 and the UCSC Genome Browser [89]. Links to downloadable files including nuclear, mitochondrial, and chloroplast genomes, RepeatModeler, gene models, methylome, and RNAseq expression data can be found at the Downloads page of [90]. The assembly used for this study is cycCry0dot2 dated February 15, 2013.

Funding

Support for JCT, OG, OC, and MH was provided in part by the Air Force Office of Scientific Research Grant FA9550-08-1-0178, Department of Energy Grants DE-SC0012556 and DE-EE0001222 and National Science Foundation Grant CBET-0903712. SJC was supported by the Broad Stem Cell Research Center. Support for DAL is from NIH Training Grant in Genomic Analysis and Interpretation T32HG002536. Support for SDG and SSM was provided to SSM by US Department of Energy Grant DE-FC02-02ER63421. MT was supported by P01 GM099134. Support for SP, and EEA was provided by US National Science Foundation (MCB-1149552). Support for SRS, JPM and AEA was provided to AEA by the Gordon and Betty Moore Foundation (GBMF3828 and GBMF5006),

the US Department of Energy (DE-SC0008593) and the National Science Foundation (OCE-1136477). MM was supported by the Whitcome Training grant. AJ was supported by R01 AI065617.

Received: 8 July 2016 Accepted: 15 November 2016
Published online: 25 November 2016

References

1. Georgianna DR, Mayfield SP. Exploiting diversity and synthetic biology for the production of algal biofuels. Nature. 2012;488:329–35.
2. Chisti Y. Biodiesel from microalgae. Biotechnol Adv. 2007;25:294–306.
3. Mata TM, Martins AA, Caetano NS. Microalgae for biodiesel production and other applications: a review. Renew Sustain Energy Rev. 2010;14:217–32.
4. Sheehan J, Dunahay T, Benemann J, Roessler P. A look back at the US Department of Energy's aquatic species program: biodiesel from algae. 1998. p. 328. http://www.nrel.gov/docs/legosti/fy98/24190.pdf. Accessed 21 Nov 2016.
5. Slocombe SP, Zhang Q, Ross M, Anderson A, Thomas NJ, Lapresa Á, et al. Unlocking nature's treasure-chest: screening for oleaginous algae. Sci Rep. 2015;5:1–17.
6. Roessler PG. Effects of silicon deficiency on lipid composition and metabolism in the diatom *Cyclotella cryptica*. J Phycol. 1988;24:394–400.
7. Reimann BEF, Lewin JMC, Guillard RRL. *Cyclotella cryptica*, a new brackish-water diatom species. Phycologia. 1963;3:75–84.
8. Huesemann MH, Hausmann TS, Bartha R, Aksoy M, Weissman JC, Benemann JR. Biomass productivities in wild type and pigment mutant of *Cyclotella* sp. (diatom). Appl Biochem Biotechnol. 2008;157:507–26.
9. Laws EA, Taguchi S, Hirata J, Pang L. Optimization of microalgal production in a shallow outdoor flume. Biotechnol Bioeng. 1988;32:140–7.
10. Roessler PG. Changes in the activities of various lipid and carbohydrate biosynthetic enzymes in the diatom *Cyclotella cryptica* in response to silicon deficiency. Arch Biochem Biophys. 1988;267:521–8.
11. Roessler PG. UDP glucose pyrophosphorylase activity in the diatom *Cyclotella cryptica.* Pathway of chrysolaminarin biosynthesis. J Phycol. 1987;23:494–8.
12. Dunahay TG, Jarvis EE, Roessler PG. Genetic transformation of the diatoms *Cyclotella cryptica* and *Navicula saprophila*. J Phycol. 1995;31:1004–12.
13. Shrestha RP, Hildebrand M. Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses. Eukaryot Cell. 2014;14:29–40.
14. Trentacoste EM, Shrestha RP, Smith SR, Glé C, Hartmann AC, Hildebrand M, et al. Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. Proc Natl Acad Sci. 2013;110:19748–53.
15. Nymark M, Sharma AK, Sparstad T, Bones AM, Winge P. A CRISPR/Cas9 system adapted for gene editing in marine algae. Sci Rep. 2016;6:24951.
16. Delalat B, Sheppard VC, Ghaemi SR, Rao S, Prestidge CA, McPhee G, et al. Targeted drug delivery using genetically engineered diatom biosilica. Nat Commun. 2015;6:1–11.
17. Daboussi F, Leduc S, Maréchal A, Dubois G, Guyot V, Perez-Michaut C, et al. Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology. Nat Commun. 2014;5:3831.
18. Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips APR, Noddings CM, et al. Designer diatom episomes delivered by bacterial conjugation. Nat Commun. 2015;6:6925.
19. Davis R, Aden A, Pienkos PT. Techno-economic analysis of autotrophic microalgae for fuel production. Appl Energy. 2011;88:3524–31.
20. Cook O, Hildebrand M. Enhancing LC-PUFA production in *Thalassiosira pseudonana* by overexpressing the endogenous fatty acid elongase genes. J Appl Phycol. 2015;28:897–905.
21. Manandhar-Shrestha K, Hildebrand M. Characterization and manipulation of a DGAT2 from the diatom *Thalassiosira pseudonana*: improved TAG accumulation without detriment to growth, and implications for chloroplast TAG accumulation. Algal Res. 2015;12:239–48.
22. Kilian O, Kroth PG. Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. Plant J. 2005;41:175–83.
23. Gruber A, Vugrinec S, Hempel F, Gould SB, Maier U-G, Kroth PG. Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. Plant Mol Biol. 2007;64:519–30.
24. Moog D, Stork S, Zauner S, Maier U-G. In silico and in vivo investigations of proteins of a minimized eukaryotic cytoplasm. Genome Biol Evol. 2011;3:375–82.
25. Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, Boore JL, et al. Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. Nat Commun. 2012;3:686.
26. Hildebrand M, Abbriano RM, Polle JE, Traller JC, Trentacoste EM, Smith SR, et al. Metabolic and cellular organization in evolutionarily diverse microalgae as related to biofuels production. Curr Opin Chem Biol. 2013;17:506–14.
27. Smith SR, Abbriano RM, Hildebrand M. Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. Algal Res. 2012;1:2–16.
28. Bender SJ, Durkin CA, Berthiaume CT, Morales RL, Armbrust EV. Transcriptional responses of three model diatoms to nitrate limitation of growth. Front Mar Sci. 2014;1:1–15.
29. Tanaka T, Maeda Y, Veluchamy A, Tanaka M, Abida H, Maréchal E, et al. Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. Plant Cell. 2015. doi:10.1105/tpc.114.135194.
30. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, et al. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science. 2004;306:79–86.
31. Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, et al. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature. 2008;456:239–44.
32. Smith SR, Glé C, Abbriano RM, Traller JC, Davis A, Trentacoste E, et al. Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. New Phytol. 2016;210:810–904.
33. Hou Y, Lin S. Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: gene content estimation for dinoflagellate genomes. PLoS ONE. 2009;4:e6978.
34. Maumus F, Allen AE, Mhiri C, Hu H, Jabbari K, Vardi A, et al. Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. BMC Genom. 2009;10:624.
35. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178–89.
36. Alverson AJ, Beszteri B, Julius ML, Theriot EC. The model marine diatom *Thalassiosira pseudonana* likely descended from a freshwater ancestor in the genus *Cyclotella*. BMC Evol Biol. 2011;11:125.
37. Alexander S, Sydow LM, Wessels D, Soll DR. Discoidin proteins of Dictyostelium are necessary for normal cytoskeletal organization and cellular morphology during aggregation. Differentiation. 1992;51:149–61.
38. Podell S, Gaasterland T. DarkHorse: a method for genome-wide prediction of horizontal gene transfer. Genome Biol. 2007;8:R16.
39. Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet. 2005;37:997–1002.
40. Bennetzen JL, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol. 2014;65:505–30.
41. Veluchamy A, Lin X, Maumus F, Rivarola M, Bhavsar J, Creasy T, et al. Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricornutum*. Nat Commun. 2013;4:2091.
42. Traller JC, Hildebrand M. High throughput imaging to the diatom *Cyclotella cryptica* demonstrates substantial cell-to-cell variability in the rate and extent of triacylglycerol accumulation. Algal Res. 2013;2:244–52.
43. Umen JG. Chloroplast DNA methylation and inheritance in Chlamydomonas. Genes Dev. 2001;15:2585–97.
44. Lopez DA, Hamaji T, Kropat J, De Hoff P, Morselli M, Rubbi L, et al. Dynamic changes in the transcriptome and methylome of *Chlamydomonas reinhardtii* throughout its life cycle. Plant Physiol. 2015. doi:10.1104/pp.15.00861.
45. Huff JT, Zilberman D. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. Cell. 2014;156:1286–97.
46. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, et al. From the cover: conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci. 2010;107:8689–94.

47. Rastogi A, Lin X, Lombard B, Loew D, Tirichine L. Probing the evolutionary history of epigenetic mechanisms: what can we learn from marine diatoms. Genetics. 2015;2:173–91.
48. Kato M, Miura A, Bender J, Jacobsen SE, Kakutani T. Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. Curr Biol. 2003;13:421–6.
49. Veluchamy A, Rastogi A, Lin X, Lombard B, Murik O, Thomas Y, et al. An integrative analysis of post-translational histone modifications in the marine diatom *Phaeodactylum tricornutum*. Genome Biol. 2015;16:1–18.
50. Hockin NL, Mock T, Mulholland F, Kopriva S, Malin G. The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants. Plant Physiol. 2012;158:299–312.
51. Claquin P, Jézéquel VM, Kromkamp JC, Veldhuis M, Kraay G. Uncoupling of silicon compared with carbon and nitrogen metabolism and the role of the cell cycle in continuous cultures of *Thalassiosira pseudonana* (Bacillariophyceae) under light, nitrogen, and phosphorus control. J Phycol. 2002;38:922–30.
52. Hollister JD, Gaut BS. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. 2009;19:1419–28.
53. Kroth PG, Chiovitti A, Gruber A, Martin Jézéquel V, Mock T, Parker MS, et al. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. PLoS ONE. 2008;3:e1426.
54. Dyrløv Bendtsen J, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 2004;340:783–95.
55. Small I, Peeters N, Legeai F, Lurin C. Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics. 2004;4:1581–90.
56. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc. 2007;2:953–71.
57. Gschloessl B, Guermeur Y, Cock JM. HECTAR: a method to predict subcellular targeting in heterokonts. BMC Bioinform. 2008;9:393.
58. Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. Plant J. 2015;81:519–28.
59. Brown MS, Ye J, Rawson RB, Goldstein JL. Regulated intramembrane proteolysis. Cell. 2000;100:391–8.
60. Gong Y, Zhang J, Guo X, Wan X, Liang Z, Hu CJ, et al. Identification and characterization of PtDGAT2B, an acyltransferase of the DGAT2 acyl-coenzyme A: diacylglycerol acyltransferase family in the diatom *Phaeodactylum tricornutu*m. FEBS Lett. 2013;587:481–7.
61. Chungjatupornchai W, Watcharawipas A. Diacylglycerol acyltransferase type 2 cDNA from the oleaginous microalga *Neochloris oleoabundans*: cloning and functional characterization. J Appl Phycol. 2014;27:1499–507.
62. Chiovitti A, Molino P, Crawford SA, Teng R, Spurck T, Wetherbee R. The glucans extracted with warm water from diatoms are mainly derived from intracellular chrysolaminarin and not extracellular polysaccharides. Eur J Phycol. 2004;39:117–28.
63. Chauton MS, Winge P, Brembu T, Vadstein O, Bones AM. Gene regulation of carbon fixation, storage, and utilization in the diatom *Phaeodactylum tricornutum* acclimated to light/dark cycles. Plant Physiol. 2013;161:1034–48.
64. Aluwihare LI, Repeta DJ, Pantoja S, Johnson CG. Two chemically distinct pools of organic nitrogen accumulate in the ocean. Science. 2005;308:1007–10.
65. Ruiz-Herrera J, Ruiz-Medrano R. Chitin biosynthesis in fungi. In: Arora DK, Bridge PD, Bhatnagar D, editors. Handbook of fungal biotechnology. New York: Marcel Dekker Inc; 2004. p. 315–30.
66. Brunner E, Richthammer P, Ehrlich H, Paasch S, Simon P, Ueberlein S, et al. Chitin-based organic networks: an integral part of cell wall biosilica in the diatom *Thalassiosira pseudonana*. Angew Chem Int Ed. 2009;48:9724–7.
67. Walsby AE, Xypolyta A. The form resistance of chitan fibres attached to the cells of *Thalassiosira fluviatilis* Hustedt. Br Phycol J. 1977;12:215–23.
68. Durkin CA, Mock T, Armbrust EV. Chitin in diatoms and its association with the cell wall. Eukaryot Cell. 2009;8:1038–50.
69. Smucker RA. Chitin primary production. Biochem Syst Ecol. 1991;19:357–69.
70. Ruiz-Herrera J, González-Prieto JM, Ruiz-Medrano R. Evolution and phylogenetic relationships of chitin synthases from yeasts and fungi. FEMS Yeast Res. 2002;1:247–56.
71. Morozov AA, Likhoshway YV. Evolutionary history of the chitin synthases of eukaryotes. Glycobiology. 2016;26:635–9.
72. Hegedus D, Erlandson M, Gillott C, Toprak U. New insights into peritrophic matrix synthesis, architecture, and function. Annu Rev Entomol. 2009;54:285–302.
73. Davis AK, Palenik B. Characterization of a modular, cell-surface protein and identification of a new gene family in the diatom *Thalassiosira pseudonana*. Protist. 2008;159:195–207.
74. Moog D, Rensing SA, Archibald JM, Maier U-G, Ullrich KK. Localization and evolution of putative triose phosphate translocators in the diatom *Phaeodactylum tricornutum*. Genome Biol Evol. 2015;7:2955–69.
75. Furumoto T, Yamaguchi T, Ohshima-Ichie Y, Nakamura M, Tsuchida-Iwata Y, Shimamura M, et al. A plastidial sodium-dependent pyruvate transporter. Nature. 2011;476:472–5.
76. Bricker DK, Taylor EB, Schell JC, Orsak T. A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, Drosophila, and humans. Science. 2012;337:96–100.
77. Herzig S, Raemy E, Montessuit S, Veuthey JL. Identification and functional expression of the mitochondrial pyruvate carrier. Science. 2012;337:93–6.
78. Lewin JC, Lewin RA. Auxotrophy and heterotrophy in marine littoral diatoms. Can J Microbiol. 1960;6:127–37.
79. White AW. Growth of two facultatively heterotrophic marine centric diatoms. J Phycol. 1974;10:292–300.
80. Pahl SL, Lewis DM, Chen F, King KD. Growth dynamics and the proximate biochemical composition and fatty acid profile of the heterotrophically grown diatom *Cyclotella cryptica*. J Appl Phycol. 2009;22:165–71.
81. Pahl SL, Lewis DM, Chen F, King KD. Heterotrophic growth and nutritional aspects of the diatom *Cyclotella cryptica* (Bacillariophyceae): effect of some environmental factors. J Biosci Bioeng. 2010;109:235–9.
82. Perez-Garcia O, Escalante FME, de-Bashan LE, Bashan Y. Heterotrophic cultures of microalgae: metabolism and potential products. Water Res. 2011;45:11–36.
83. Zaslavskaia LA, Lippmeier JC, Shih C, Ehrhardt D, Grossman AR, Apt KE. Trophic conversion of an obligate photoautotrophic organism through metabolic engineering. Science. 2001;292:2073–5.
84. Kruckeberg AL. The hexose transporter family of *Saccharomyces cerevisiae*. Arch Microbiol. 1996;166:283–92.
85. Poulsen N, Chesley PM, Kröger N. Molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (Bacillariophyceae). J Phycol. 2006;42:1059–65.
86. Darley WM, Volcani BE. Role of silicon in diatom metabolism. Exp Cell Res. 1969;58:334–42.
87. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567–80.
88. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002;30:1575–84.
89. *C. cryptica* genome browser. http://genomes.mcdb.ucla.edu/cgi-bin/hgGateway. Accessed 21 Nov 2016.
90. *Cyclotella cryptica* landing page. http://genomes.mcdb.ucla.edu/Cyclotella/download.html. Accessed 21 Nov 2016.
91. Damste JSS. The rise of the rhizosolenid diatoms. Science. 2004;304:584–7.

CHAPTER 2 SUPPLEMENTAL MATERIAL

Supplemental Methods for Traller et al "Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype"

DNA extraction and purification

Liquid cell cultures concentrated either by filtration using a 3.0µm polycarbonate filter (DNA for genome sequencing) or by centrifugation in 50mL conical tubes at 4000 xg for 8 minutes (methylome DNA) using in an Eppendorf 5810R centrifuge. DNA for genome and bisulfite sequencing was purified using CsCl as described in [1]. To remove RNA contamination, DNA for bisulfite sequencing was treated with 10mg/mL stock of RNase A for 15 minutes at 37ºC.

Genome Library Construction and Sequencing

Three libraries were prepared following Illumina's standard genomic DNA paired end construction: "PE-short" (post-assembly empirical insert lengths ~90-235 nucleotides (nt) exclusive of adapters [preparation target 180-230 nt], with mode 123 nt), "PE-medium" (~155-330 nt [preparation target 230- 330 nt], mode 221 nt), and "PE-long" (~225-460 nt [preparation target 330-430 nt] with mode 305 nt) for ~58% of inserts, with the balance in a second mode ~60-225 nt peaking at ~105 nt). These were sequenced as paired end 76-mer + 76-mer reads on an Illumina GA-IIx 120-tiles/lane run ("TP003") at the UCLA BSCRC Core Sequencing facility, using two dedicated lanes for each of PE-short and PE-medium, three dedicated lanes for PE-long, a dedicated PhiX control lane for RTA image analysis autocalibration, and spiking in ≈1% Illumina PhiX into each non-control lane. Two genomic DNA mate pair libraries (aiming for 10K nt effective inserts) were prepared and run by Illumina service on a 48-tile/lane v3 HiSeq flow cell, each library on

a single dedicated lane: “MP-short” (effective ~2,100-3,320 nt with mode ~2,625 nt for ~58% of inserts, with the balance in a PE-orientation [rather than MP-orientation] second mode ~170-480 nt peaking at ~205 nt), and “MP-long” (~1,740-2,730 nt with mode ~2,260 nt for ~63% of inserts, with the balance in a PE-oriented second mode ~176-535 nt peaking at ~223 nt). These were physically sequenced as paired end 101-mer + 10-mer index + 101-mer, with the index reads and the last base of each main end discarded (in the usual way so that the last retained base has bidirectional RTA phasing corrections). The number of raw read pairs for PE-short/medium/long/control is found in Additional File 1, Figure S1b. Only read pairs with no ‘N’ basecalls were retained; due to the pattern of ‘N’ basecalls in PE-long, the first four bases of each end of its lanes were discarded before this filter (and subsequent uses of this library).

The number of raw read pairs for MP-short and long libraries was 142,455,072 and 154,107,079, respectively, and only RTA PF=1-passing pairs were retained (Additional File 1, Figure S1b). Although not used as a filter, relative to PF=1 read pairs, the ‘N’-free read pairs for MP-short/long were ~98.9% / 98.6%.

The PE and MP libraries contributed ~23.4G nt and ~57.4G nt, respectively, for a total of ~80.8G nt (≈461x coverage of a 175Mbp genome; for 65-mers: ~3.4G and ~20.7G, total ~24.0G and ≈137x).

Using many iterations of a variety of standard and *ad hoc* assemblers and alignment tools, with extensive inspection of intermediate stages and judgment calls made by hand, read pairs from PE-short/medium/long and MP-short/long were used to construct high-quality best assemblies from the available data for the chloroplast (“chrC”) and mitochondrial (“chrM”) genomes in *C. cryptica*. In each case, a single complete circular

sequence of pure A/C/G/T's without gaps was formed (chrC 129,320 bp, chrM 58,021 bp). This was greatly assisted by the presence in NCBI of related genomes: KJ958480.1 for *Cyclotella* strain L04_2 chrC (~96% identity; also useful: KJ958481.1 for *Cyclotella* strain WC03_2 chrC), and NC_007405.1 for *Thalassiosira pseudonana* chrM (more distant; even on homologous stretches, overall percent identity ≈80%). The *C. cryptica* chrM is estimated with 17,880 bp (~31%) being 120 exact copies of the 149 bp sequence

TTATCGGCCTCAAATCAAGCAGTGTTTAAGCTGGAAT

CTATCGGCCTCAAATCGAAACAGTGTTTTAGCCTGAAT

TTATCGGCCTCAAATCAAGCAGTGTTTAAGCTGGAAT

CTATCGGCCTCAAATCGAAACAGTGTTTTGCCTGAAT (which is itself four approximate tandem copies of a smaller unit). Given current data, this region cannot be completely resolved, and there is likely additional variation here, and the included number of copies is an estimate informed in part by depth of coverage relative to other chrM sequence.

The main genome assembly was performed with an ABySS 1.3.1 SE+PE+MP pipeline using k=65-mers, t=65, q=3, e=3, E=0, c=3, m=30, p=0.9, no scaffolding at PopBubbles, s=200, n=10, overlap min=5 with scaffolding and join masking at simple repeats, SimpleGraph d=6 and scaffolding, greedy MergePaths, a=4, and abyss-scaffold min-gap=100. The SE stage used PE-short+medium+long, MP-short+long, the PE stage was applied to PE-short+medium+long, and the MP stage was applied to MP-short+long. The assembly was taken as the final scaffolds of nt length $\geq 130 = 2k$. Based on alignments, scaffolds apparently consisting of PhiX, chrC, or chrM were removed.

Genome annotation revealed several nearly identical genomic contigs with nearly identical fold coverage, which appear to be due to a failure of the assembler to collapse identical contigs into one contig. In an effort to resolve this artifact from the assembler, we performed all versus all BLASTn for each genomic contig against the whole genome and removed contigs which contained a >95% threshold for query coverage and nucleotide identity. This reduced the total number of contigs from 199,501 to 116,817 (41.3% reduction) and total genome size from 182,854,974bp to 161,759,242bp (excluding the mitochondrial and chloroplast genomes).

Notation for contigs is in the form of gXXXXXX_YYYYY, where XXXXXX gives contig length and YYYYY gives the approximate average coverage for that contig. Contigs beginning with a 'g' are genomic sequences and contigs beginning with an 'r' are mRNA sequences.

DNA Bisulfite sequencing and analysis

1 µg of *C. cryptica* nuclear DNA for bisulfite treatment was resuspended in 50 µl of EB buffer (QIAGEN) and sonicated in AFA-fiber microTubes using a Covaris S2 machine (Duty Cycle = 10%; Intensity = 5; Cycles/Burst = 200; for 6 minutes) to obtain 100-300 bp fragments. The DNA was then subjected to End-Repair, A-tailing and Adapter Ligation using Illumina TruSeq DNA Sample Prep kit v2 according to manufacturer's instructions. The Adapter-ligated DNA was purified and size-selected using AMPure XP beads. DNA was then bisulfite treated using EpiTect kit (QIAGEN) using the following conversion protocol: 95°C 5min, 60°C 25min, 95°C 5min, 60°C 85min, 95°C 5min, 60°C 175min, 95°C 5min, 60°C 25min, 95°C 5min, 60°C 85min,

95°C 5min, 60°C 175min. Bisulfite-treated DNA was then desulphonated according to manufacturer's protocol ("Small Amount of Fragmented DNA" variant) and DNA eluted twice with EB. Converted DNA was amplified with MyTaq 2x mix (Bioline): 98°C 2 min; 12 cycles of 98°C 15 sec, 60°C 30 sec, 72°C 30 sec; 72°C 5 min. Amplified DNA was diluted to 10 nM and sequenced using Illumina HiSeq2000 (100 single end reads).

Bisulfite converted reads were inspected for sequencing quality using FastQC 0.10.1. Reads passing the Illumina quality filter ('PF' value equal to 1) were retained and aligned to the genome assembly using BS-Seeker2 in local alignment mode with Bowtie2 as the aligner [2]. For the purpose of this study, a 'methylated' base pair is defined as a cytosine which has >= 4 methylated reads, similar to as described in [3]. Therefore an 'unmethylated' cytosine is that which has >= 4 unmethylated reads. Those cytosines which do not have at least 4 aligned reads from the bisulfite sequencing are considered to not have enough data sufficiently conclude whether that site is methylated or not. Genes that are 'methylated' are those which are defined as having >=50% of callable CG sites (the most common motif for methylation) containing a 'methylated' cytosine. Unmethylated genes are <50%.

RNA sequencing

Total RNA was purified from cultures of *C. cryptica* in log-phase growth under conditions of either silicon starvation or nitrogen starvation. For RNA isolation, 750mL of liquid cell culture for each time point was harvested and treated with 20mg/mL cycloheximide and concentrated by filtration. Cells were stored in -80ºC prior to extraction. Total RNA was extracted according to [4,5].

Five μg of total RNA per sample was treated with Turbo DNase (Ambion) for 30 min at 37°C according to the manufacturer's instructions in order to remove any contaminating DNA. The resulting RNA was purified by ethanol precipitation, and RNA quality was evaluated on a BioAnalyzer RNA Nano kit (Agilent). RNAseq libraries were prepared using the Illumina TruSeq mRNA Sample Prep kit (Illumina) according to manufacturer's protocols (Rev. A). Sequencing was performed at the UCLA Broad Stem Cell Research Center sequencing core on a HiSeq 2000 sequencer (Illumina) using a mixture of 50+50 nt paired end reads and 100 nt single end reads. 11 libraries were sequenced on a single lane each, 2 libraries were multiplexed onto a 1 lane. Pooled raw reads were demultiplexed and all reads mapped to transcriptome data, with remaining unmapped reads mapped to reference *Cyclotella* genome version cycCry0dot2 using TopHat 2.0.4 allowing for two mismatches, reporting only unique mappings [6]. Bam files were processed through HTSeq 0.5.3 using the "intersection-nonempty" method which assigns a read to a gene only if the read overlaps with only one gene in its entirety. Single-end and paired-end data were combined and all data was run through DESeq 1.8.3 to obtain FPKM counts [7]. A combination of 50+50 nt paired end reads and 100 nt single end reads were pooled to facilitate accurate mapping of the genes.

Genome Annotation

Gene model predictions were generated from several pipelines as follows: (1) FGENESH using the built-in diatom training set; (2) standalone AUGUSTUS using the built-in '*Chlamydomonas reinhardtii'* training parameters; (3) standalone AUGUSTUS using the 100 longest FGENESH predictions as a training set; (4) web-based

AUGUSTUS trained on the de-novo RNA assembly; and (5) MAKER with FGENESH, AUGUSTUS, GeneMarkES analyses enabled [8-11]. 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. AUGUSTUS predictions from set (4) that overlap MAKER predictions from set (5) constitute the 'high-confidence' collection of gene predictions. For gene structure, it was determined that (4) more accurately predicted start/stop sites and intron/exon boundaries. Gene set (4 and 5) were functionally annotated [12].

Repetitive elements were identified using RepeatMasker 4.0.1 with RMBLASTN 2.2.27+, using the Bacillariophyta repeat library from RepBase and default settings. Additional repeat masking by RepeatMasker was performed using a custom de-novo library generated using RepeatModeler 1.0.7 with the NCBI BLAST engine.

The diatom genomes used in Figure 9 and S2, OrthoMCL, and RBH analyses are *T. pseudonana* v3.0 [13], *P. tricornutum* v2.0 [14], *F. cylindrus* v1.0 [15], *P. multiseries* v1.0 [16], and *T. oceanica* v.3.0 (NCBI accession numbers JP288099-JP2977110, http://www.ncbi.nlm.nih.gov)

For reciprocal best BLAST hit (RBH) analysis, gene models from diatom genomes (listed above) were aligned using BLASTp 2.2.28+ [17], with an e-value cutoff score of 1e-5 and a query coverage of at least 70%. RBH pairs were detected using python script [18].

Predicted proteins from *T. pseudonana* and *C. cryptica* genomes were evaluated for phylogenetic relatedness to sequences in NCBI GenBank nr (accessed October 16, 2015)

using the DarkHorse program version 1.5 with a threshold filter setting of 0.1 [24]. 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 better for inclusion in this analysis.

For phylogenetic analysis in Figure 9 and S2, and to further investigate proteins of interest, amino acid sequences were aligned using MUSCLE with 10 maximum number of iterations and default parameters [19]. Trees were generated using default parameters in RAxML_GUI v1.3 for 100 iterations and visualized using FigTree v1.4.0 [20]. For each tree shown, bootstrap values are listed and have been midpoint rooted.

Subcellular Localization Prediction

All open reading frames were analyzed using several computational tools for predicting the likely location of the protein product within the cell. Nucpred [21] and NetNES [22] provided estimations of nuclear localization for each putative protein. PREDOTAR [23], PSORT [24] , HECTAR [25], CELLO [26] , PredAlgo [27], SignalP 3.0 [28] TargetP, ChloroP [29] and ASAFind [30] provided possible localizations to mitochondria, chloroplast, endoplasmic reticulum, or plastid. The presence of N-terminal signal peptides and transmembrane regions were assessed using PHOBIUS [31,32], the SignalP 3.0 portion of TargetP and PredAlgo. Specific sequences and predicted cleavage sites for signal peptides were taken from SignalP 3.0. All open reading frames from *C. cryptica* were submitted to web-servers in an automated fashion using scripts for html or webmail submission.

Methods for targeting predictions shown in Figure 4 are as follows: for plastid targeting, all proteins with predicted positive ASAfind plastid targeting were surveyed for

proper cleavage site using SignalP 3.0 [28, 30]. Predicted SignalP 3.0 cleavage sites were then defined by the guidelines addressed in Figure 5 of [33] and split into canonical plastid (AF, GF, and SF cleavage sites), noncanonical plastid (AW, AY, AL, GW, GY, GL, SW, SY, SL), periplastid and other (AH, AI, AM, AR, AE, AG, SH, SI, SM, SR, SE, SG, GH, GI, GM, GR, GE, GG). Periplastid predicted cleavage sites which did not contain a positive ChloroP prediction were then categorized as ER/secreted. Any negative plastid ASAfind result but SignalP 3.0 positive with a predicted periplastid cleavage site, and positive ChloroP were also classified as periplastid. ER/secreted proteins were then defined as proteins with negative plastid ASAfind prediction, positive SignalP 3.0, with a cleavage site not one that is specified in [33].

Mitochondrially-targeted proteins were classified according to any of the following prediction combinations for HECTAR, Predotar, and TargetP (listed in respective order): Type II, Mito, Mito; SigP, Mito, Mito; Mito, Plastid, Mito; Mito, Mito, SigP; Mito, Mito, Mito; Mito, Mito, Plastid; Mito, Mito, no prediction; Mito, no prediction, Mito; no prediction, Mito, Mito. MitoProt was also used for all predicted proteins shown in Figure 5-8 [34]. All other proteins that did not fall into the above criterion were classified as 'cytosol and other.'

Vector Construction and Diatom Transformation

The vector utilizing the *C. cryptica* rpL41 promoter and terminator sequences was assembled using the GeneArt® Seamless Cloning and Assembly Kit (Invitrogen), with a Gateway™ vector (Invitrogen) as the backbone. The construction of the *T. pseudonana fcp* [35] and nitrate reductase (NR) [36,37] vectors is described elsewhere.

nitrate as the nitrogen source, which would repress the expression of GFP under the NR promoter [36]. The cultures were allowed to reach exponential phase and imaged using Zeiss Axio Observer Z1 Inverted Microscope (Zeiss Axioscope, Carl Zeiss Microimaging Inc., USA). The filter set for fluorescent imaging was Emission LP 515 nm for chlorophyll autofluorescence, and Zeiss #38HE Excitation BP 470/40 nm, Dichromatic mirror FT 495 nm, Emission BP 525/50 nm for GFP.

1. Jacobs JD, Ludwig JR, Hildebrand M, Kukel A, Feng T-Y, Ord RW, et al. Characterization of two circular plasmids from the marine diatom Cylindrotheca fusiformis: plasmids hybridize to chloroplast and nuclear DNA. Molec. Gen. Genet. 1992;233:302–10.

2. Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics. 2013;14:774.

3. Lopez DA, Hamaji T, Kropat J, De Hoff P, Morselli M, Rubbi L, et al. Dynamic changes in the transcriptome and methylome of Chlamydomonas reinhardtii throughout its life cycle. plantphysiol.org. American Society of Plant Biologists; 2015;169:00861.2015–743.

4. Smith SR, Glé C, Abbriano RM, Traller JC, Davis A, Trentacoste E, et al. Transcript level coordination of carbon pathways during silicon starvation‐induced lipid accumulation in the diatom Thalassiosira pseudonana. New Phytol. 2016;210:810-904.

5. Hildebrand M, Dahlin K. Nitrate transporter genes from the diatom Cylindrotheca fusiformis (Bacillariophyceae): mRNA levels controlled by nitrogen source and by the cell cycle. J. Phycol. 2000;36:702-13.

6. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.

7. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

8. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 2005;33:W465–7.

9. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, et al. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2007;18:188–96.

10. Salamov AA. Ab initio Gene Finding in Drosophila Genomic DNA. Genome Res. 2000;10:516–22.

11. Borodovsky M, Lomsadze A. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. Curr Protoc Bioinformatics. 2011;Chapter 4:Unit4.6.1–10.

12. Lopez D, Casero D, Cokus SJ, Merchant SS, Pellegrini M. Algal Functional Annotation Tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. BMC Bioinformatics. 2011;12:282.

13. Joint Genome Institute. http://genome.jgi.doe.gov/Thaps3/Thaps3.home.html

14. Joint Genome Institute. http://genome.jgi.doe.gov/Phatr2/Phatr2.home.html

15. Joint Genome Institute. http://genome.jgi.doe.gov/Fracy1/Fracy1.home.html

16. Joint Genome Institute. http://genome.jgi.doe.gov/Psemu1/Psemu1.home.html

17. Altschul SF, Gish W, Miller W, Myers EW. Basic local alignment search tool. J. Mol. Biol. 1990;215:403-10.

18. Qin H. hongqin/Simple-reciprocal-best-blast-hit-pairs.

19. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004;32:1792–7.

20. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22:2688–90.

21. Brameier M, Krings A, MacCallum RM. NucPred Predicting nuclear localization of proteins. Bioinformatics. 2007;23:1159–60.

22. La Cour T, Kiemer L, Mølgaard A, Gupta R, Skriver K, Brunak S. Analysis and prediction of leucine-rich nuclear export signals. Protein Engineering Design and Selection. 2004;17:527–36.

23. Small I, Peeters N, Legeai F, Lurin C. Predotar: A tool for rapidly screening proteomes forN-terminal targeting sequences. Proteomics. 2004;4:1581–90.

24. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35:W585–7.

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

26. Yu C-S, Lin CJ, Hwang J-K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Science. Cold Spring Harbor Laboratory Press; 2004;13:1402–6.

27. Tardif M, Atteia A, Specht M, Cogne G, Rolland N, Brugiere S, et al. PredAlgo: A New Subcellular Localization Prediction Tool Dedicated to Green Algae. Mol Biol Evol. 2012;29:3625–39.

28. Dyrløv Bendtsen J, Nielsen H, Heijne von G, Brunak S. Improved Prediction of Signal Peptides: SignalP 3.0. J Mol Biol. 2004;340:783–95.

29. Emanuelsson O, Brunak S, Heijne von G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc. 2007;2:953–71.

30. Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. Plant J. 2015;81:519–28.

31. Laws EA, Taguchi S, Hirata J, Pang L. Optimization of microalgal production in a shallow outdoor flume. Biotechnol. Bioeng. 1988;32:140–7.

32. Käll L, Krogh A, Sonnhammer E. A Combined Transmembrane Topology and Signal Peptide Prediction Method. J Mol Biol. 2004;338:1027–36.

33. Gruber A, Vugrinec S, Hempel F, Gould SB, Maier U-G, Kroth PG. Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. Plant Mol Biol. 2007;64:519–30.

34. Claros MG, Vincens P. Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem. 1996.

35. Shrestha RP, Hildebrand M. Evidence for a Regulatory Role of Diatom Silicon Transporters in Cellular Silicon Responses. Eukaryot Cell. 2014;14:29–40.

36. Poulsen N, Chesley PM, Kröger N. Molecular genetic manipulation of the diatom Thalassiosira pseudonana (Bacillariophyceae). J. Phycol. 2006;42:1059–65.

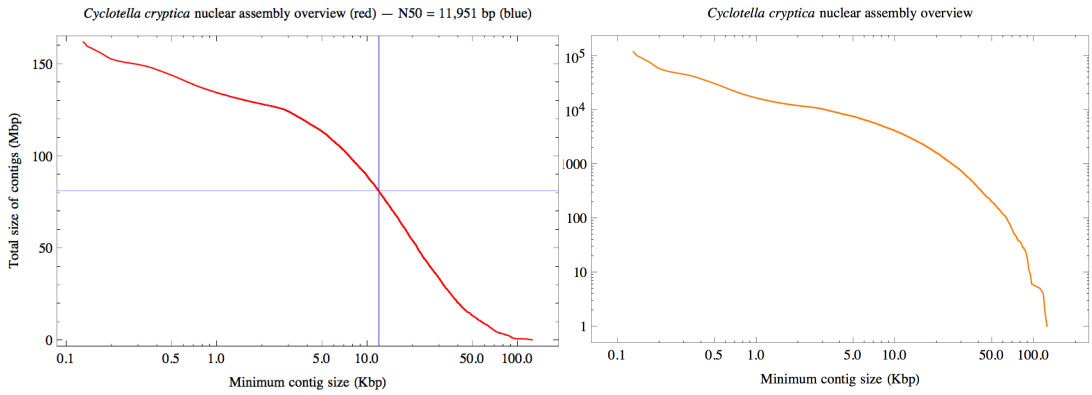
37. Trentacoste EM, Shrestha RP, Smith SR, Glé C, Hartmann AC, Hildebrand M, et al. Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. Proc Nat Acad Sci. 2013;110:19748–53.

326 38. Darley WM, Volcani BE. Role of silicon in diatom metabolism. Exp Cell Res.
327 1969;58:334–42.

328

Figure S1: (a) *Cyclotella cryptica* nuclear assembly overview broken down by minimum contig size (Kbp) versus total Mbp of all contigs (left) and minimum contig size versus the total number of contigs (right) and (b) genomic sequencing data.

a.



b.

Genomic Libraries	Number of Raw Reads for Paired End Libraries			Percentage passing pairs				Percentage passing pairs after removal of pairs with 'N'
	Lane 1	Lane 2	Lane 3 Total Reads	Lane 1	Lane 2	Lane 3	Average	Average
Paired End Short	65270867	17592293	82863160	4.8	57		71.1	73.0
Paired End Medium	65333130	66235838	131568968	68.5	69.5		60.0	79.6
Paired End Long	69147985	69578501	65179466 203905952	27.8	34	65.6	42.5	57.9
Paired End Control	30301553			92.8				99.5
Mate Pair Short	142455072			96.9				98.9
Mate Pair Long	154107079			96.7				98.6

Supplementary table S1. Statistics from different gene model prediction pipelines. The pipelines are: (1) FGENESH Gene predictions, Diatom training set, (2) Augustus Gene predictions V1, *Chlamydomonas* training set, no RNAseq data, (3) Augustus Gene predictions V2, FGENESH100 training set, RNAseq data (4) Augustus Gene predictions V3, 'self' trained, RNAseq data (5) MAKER Gene predictions, (Augustus self trained + FGENESH + GeneMarkES). Details are presented in Supplementary Methods. Data presented includes all predicted models regardless of read counts from RNAseq data and prior to removing apparent duplicate contigs (Additional File 1: Supplementary Methods, Genome Library Construction and Sequencing).

	Gene Model Prediction Pipeline				
	CcFgenesh	**CcAugustusV1**	**CcAugustusV2**	**CcAugustusV3**	**CcMAKERV1**
Total Models	50,288	6,295	24,819	**33,682**	**9,049**
Average Model Length (bp)	1,561.8	926.1	1,746.2	**1,265**	**1,999.7**
Average Exons	3.72	1.22	2.65	**1.95**	**3.69**
Avg. Exon Length (bp)	247.4	115.8	561.5	**585.6**	**457.6**
Avg. # Introns	2.72	0.22	1.65	**0.95**	**2.69**
Avg. Intron Length (bp)	129.8	542	153.9	**128.1**	**128.1**

Table S2: Repeat sequences in *C. cryptica* identified using (a) RepBase data and (b) RepeatModeler Data.

a.

Repeat sequences in *Cyclotella cryptica* using RepBase data
Sequences: 199501
total length: 182854974 bp (174198679 bp excl N/X-runs)
GC level: 43.01%
Bases masked: 7037708 bp (3.85 %)

	Number of elements*	Length occupied (bp)	Percentage of sequence
Retroelements	12264	4694900	2.57
SINEs:	0	0	0
Penelope	13	1650	0
LINEs:	58	4854	0
CRE/SLACS	0	0	0
L2/CR1/Rex	0	0	0
R1/LOA/Jockey	0	0	0
R2/R4/NeSL	0	0	0
RTE/Bov-B	2	100	0
L1/CIN4	0	0	0
LTR elements:	12206	4690046	2.56
BEL/Pao	0	0	0
Ty1/Copia	2901	1432661	0.78
Gypsy/DIRS1	9305	3257385	1.78
Retroviral	0	0	0
DNA transposons	346	113716	0.06
hobo-Activator	0	0	0
Tc1-IS630-Pogo	0	0	0
En-Spm	0	0	0
MuDR-IS905	0	0	0
PiggyBac	2	103	0
Tourist/Harbinger	305	105959	0.06
Other (Mirage, P-element, Transib)	0	0	0
Rolling-circles	0	0	0
Unclassified:	28	6197	0
Total interspersed repeats:		4814813	2.63
Small RNA:	142	17166	0.01
Satellites:	0	0	0
Simple repeats:	32133	1928189	1.05
Low complexity:	3624	294614	0.16

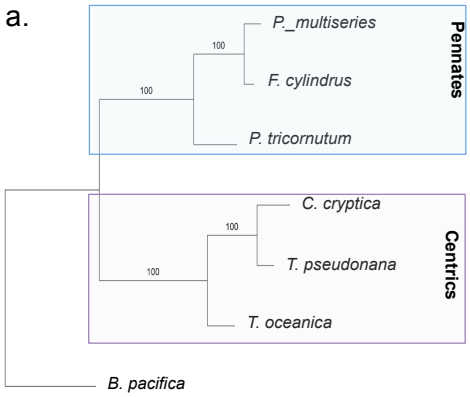
b.

Repeat sequences in *Cyclotella cryptica* using RepeatModeler data
Sequences: 199501
Total length: 182854974 bp (174198679 bp excl N/X-runs)
GC level: 43.01 %
Bases masked: 98288109 bp (53.75 %)

	Number of elements*	Length occupied (bp)	Percentage of sequence
SINEs:	0	0	0
ALUs	0	0	0
MIRs	0	0	0
LINEs:	2626	1877047	1.03
LINE1	0	0	0
LINE2	0	0	0
L3/CR1	0	0	0
LTR elements:	43176	15802661	8.64
ERVL	0	0	0
ERVL-MaLRs	0	0	0
ERV_classI	0	0	0
ERV_classII	0	0	0
DNA elements:	15402	5899536	3.23
hAT-Charlie	0	0	0
TcMar-Tigger	155	44023	0.02
Unclassified:	314059	73067346	39.96
Total interspersed repeats:		96646590	52.85
Small RNA:	0	0	0
Satellites:	0	0	0
Simple repeats:	23628	1861243	1.02
Low complexity:	1672	129011	0.07

Figure S2: Phylogenetic comparison of diatom species with sequenced genomes. (a) 18S sequence comparison from [89], accession numbers are listed in Additional File 4 (b) Reciprocal best blast hits comparison. *C. cryptica* and *T. pseudonana* are the most similarly related centric diatoms with available genomic data.

a.



b.

	Total Reciprocal BLAST Hits	Average % Identity
C. cryptica/T. pseudonana	5498	63.6
C. cryptica/T. oceanica	4443	56.4
T. pseudonana/T. oceanica	3592	59.2
P. tricornutum/F. cylindrus	7060	55.5
P. tricornutum/P. multiseries	2878	54.2
F. cylindrus/P. multiseries	7773	66.0

Figure S3: Per-cytosine fraction methylation. (a) 0 hour, silicon-replete sample (b) 48 hour, silicon-deplete sample. All sites shown have a read coverage greater than or equal to 4.

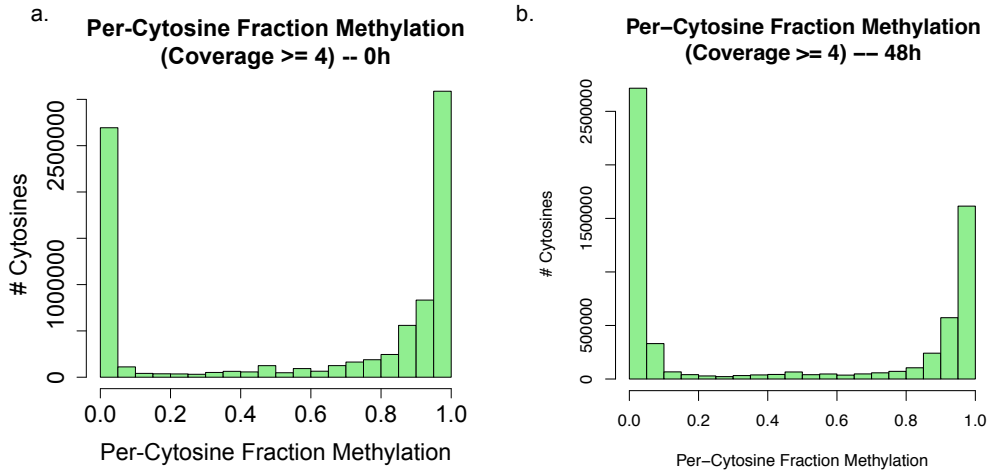
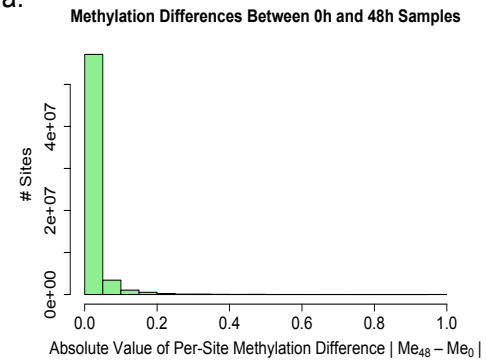


Figure S4: Per-site methylation differences between Silicon replete and deplete conditions. (a) Absolute value of the difference between 48h and 0h. (b) Slope between fraction methylation of genes in when comparing the two conditions is linear, outliers are apparent.

a.



b.

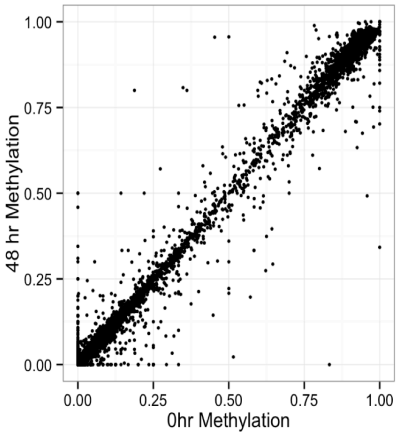


Figure S5: Per-site methylation differences between silicon replete and deplete conditions. (a) Histogram showing absolute value of the difference between 48h and 0h. (b) Scatter plot showing the slope between fraction methylation of genes in when comparing the two conditions is linear, outliers are apparent.

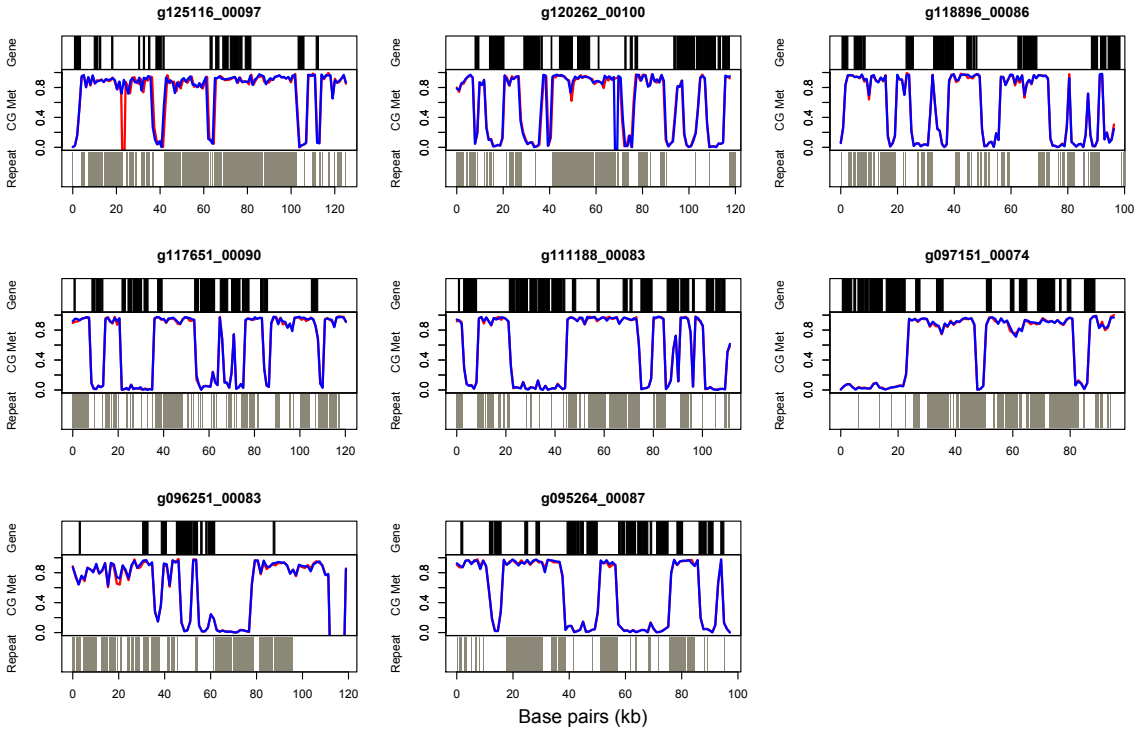


Figure S6: Binned repeats (a) and fraction methylation across gene body (b) in *C. cryptica.*

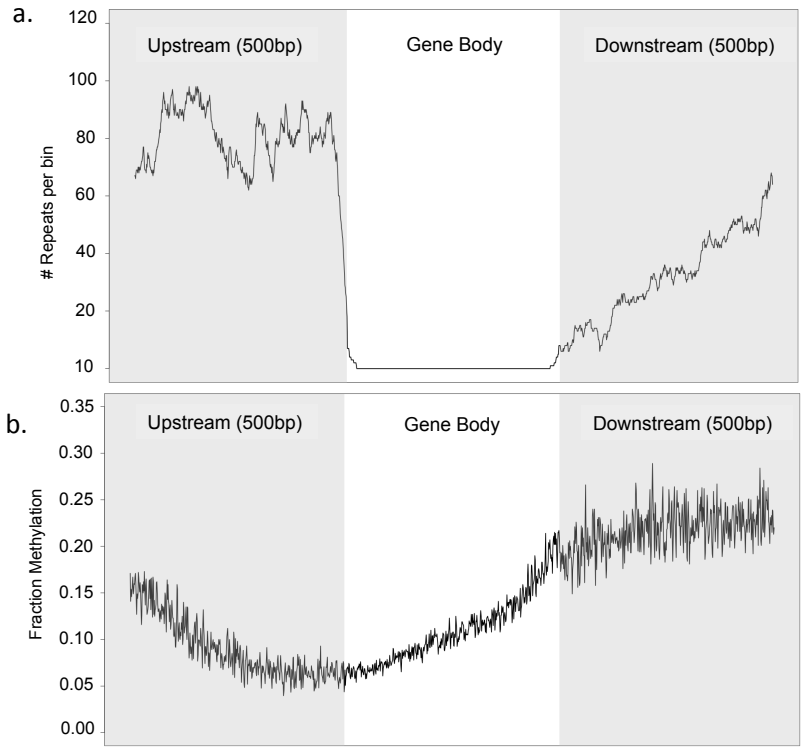


Table S3: Summary of gene methylation (AUGUSTUS V3 models) in both experimental conditions.

Silicon Replete 0hr					
	Count	**Percent total of all genes**	**Percent total Excluding genes with insufficient coverage**	**Average Fraction Methylation**	**Average FPKM**
Methylated genes	4170	20	22	0.87	536
Unmethylated genes	14866	70	78	0.07	4593
No Methylation Data	2085	10	-	ND	1767
Total genes	21121			0.24	3520
Silicon Deplete 48hr					
	Count	**Percent total of all genes**	**Percent total Excluding genes with insufficient coverage**	**Average Fraction Methylation**	**Average FPKM**
Methylated genes	2627	12	24	0.88	523
Unmethylated genes	8453	40	76	0.07	4589
No Methylation Data	10041	48	-	ND	3641
Total genes	21121			0.26	3408

	Methylated 48hr	**Unmethylated 48hr**
Methylated 0hr	2584	31
Unmethylated 0hr	17	8165

Figure S7: Gene methylation relative to gene expression. (a) 3 populations emerged (Red boxes i-iii) when comparing average fraction methylation to Log2 FPKM. (b) Gene count binned according to FPKM and shaded according to methylation status. Line depicts the proportion of methylated genes per bin. Data shown is for silicon replete, 0 hour condition.

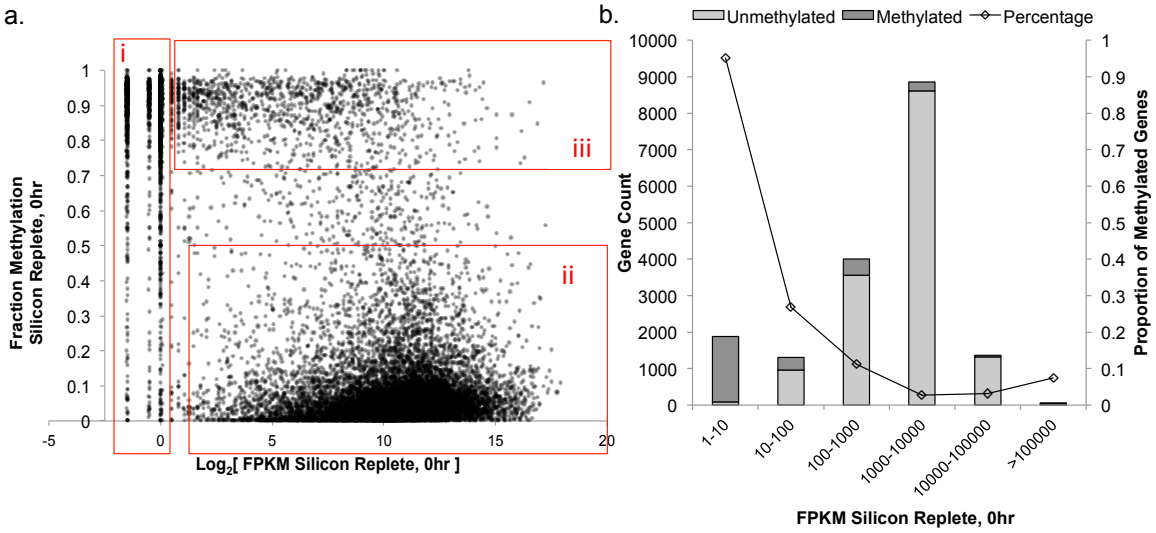
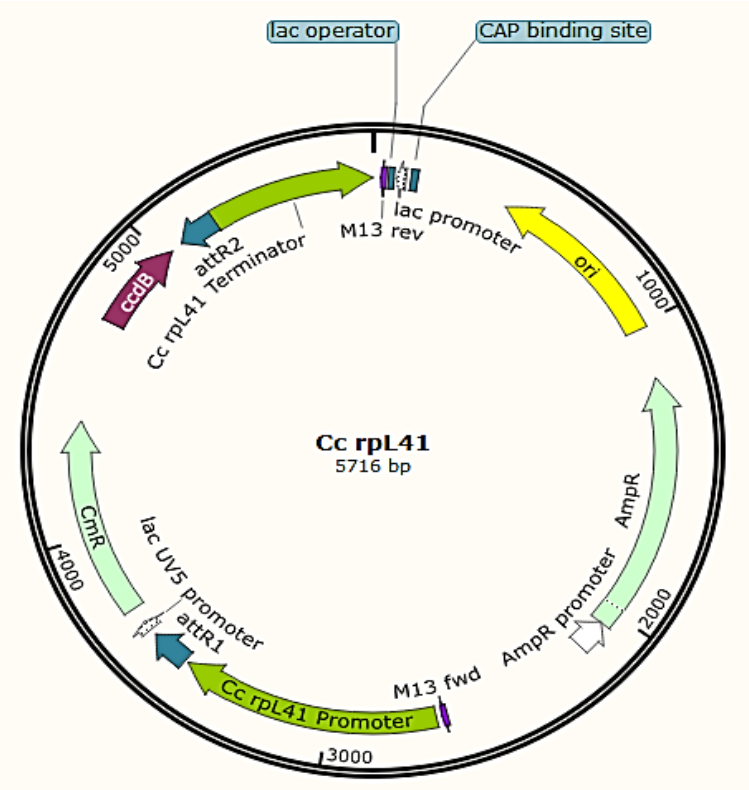


Figure S8: Vector map for rpL41 construct for *C. cryptica.*



2.8 ACKNOWLEDGEMENTS

Chapter 2, in full, is a reprint of the material as it appears in Biotechnology for Biofuels 2016. Traller, Jesse C, Shawn J Cokus, David A Lopez, Olga Gaidarenko, Sarah R Smith, John P McCrow, Sean D Gallaher, Sheila Podell, Michael Thompson, Orna Cook, Marco Morselli, Artur Jaroszewicz, Eric E Allen, Andrew E Allen, Sabeeha S Merchant, Matteo Pellegrini, Mark Hildebrand. "Genome and Methylome of the Oleaginous Diatom *Cyclotella Cryptica* Reveal Genetic Flexibility Toward a High Lipid Phenotype," *Biotechnology for Biofuels,* 9(1), 258. 2016. The dissertation author was the primary researcher and author of this paper.

CHAPTER 3

Comparison of the transcript level and physiological response in the diatom *Cyclotella cryptica* under two distinct macronutrient limitations

3.1 ABSTRACT

Due to its ability to accumulate substantial amounts of triacylglycerol during nutrient starvation, the marine diatom *Cyclotella cryptica* holds promise as a candidate species for large-scale cultivation for biofuel. However, the molecular and genetic mechanisms that control lipid accumulation and the cellular response to nutrient starvation are largely unknown. To aid in elucidating the metabolic processes involved in lipid accumulation in *C. cryptica,* we determined the transcriptomic profiles of cells undergoing two distinct macronutrient starvations, nitrogen and silicon, over the course of 96 hours. Comparison of the two nutrient starvations revealed fundamental differences in metabolite processing in the mitochondrion, with silicon-deprived cells utilizing beta-oxidation and the glyoxylate cycle, and nitrogen-deprived cells preferring mitochondrial glycolysis and the TCA cycle. Dramatic transcriptional shifts in triacylglycerol synthesis and chrysolaminarin degradation genes indicated lipid accumulation is more likely due to a repartitioning of carbon from carbohydrate to triacylglycerol, rather than *de novo* fatty acid synthesis from newly formed photosynthate. We further substantiated this observation with physiological measurements on carbohydrate, lipid, and photosynthetic efficiency. Overall, this data identifies the repartitioning of carbon and energy molecules during nutrient-induced cell cycle arrest and provides insight into the distinct mechanisms of triacylglycerol accumulation in this oleaginous diatom.

3.2 INTRODUCTION

The advent of high throughput sequencing technologies, in particular transcriptome analysis, has helped to understand global cellular metabolism as it responds to environmental stimuli such as light-dark oscillations, nutrient stress, and temperature. All across the algal tree of life there have been several gene expression studies focusing on the cell's response to nutrient starvation. Commonly, cell cycle arrest from nutrient starvation leads to lipid accumulation, and analysis of transcriptomic data has resulted in significant progress to pinpoint key gene(s) controlling lipid accumulation (Schmollinger et al. 2014; Tanaka et al. 2015; Alipanah et al. 2015; Hovde et al. 2015; Smith et al. 2016a). Furthermore, hypotheses generated from these studies have aided in successful attempts to increase lipid accumulation by genetically engineering diatoms to over-express or knock-down genes involved in either triacylglycerol biosynthesis or competing pathways (Guihéneuf et al. 2011; Trentacoste et al. 2013; Ge et al. 2014; Manandhar-Shrestha and Hildebrand 2015; Yang et al. 2016). Even without the use of genetic engineering tools, transcriptomic datasets allow for the elucidation of the molecular mechanisms controlling physiological processes and provide the researcher with a broad view of global metabolism in the cell; as opposed to physiological studies that provide an in-depth perspective on a specific area of metabolism. Therefore, when transcript level studies are combined with physiological measurements, there is the opportunity to more fully understand the biological system for which we seek to control.

The sequencing and annotation of the nuclear, chloroplast, and mitochondrial genomes in the diatom *Cyclotella cryptica* as highlighted in Chapter 2 opened the door

for further ‘omic studies investigating the lipid accumulation response in this species. While there have been several gene expression studies focusing on nutrient stress and lipid accumulation in eukaryotic microalgae, to the best of our knowledge, there have been no transcriptomic studies comparing two different nutrient starvations to one another. A transcript level comparison of lipid induction under different nutrient stresses may highlight the conserved metabolic processes responsible for lipid accumulation as well as elucidate unique strategies the cell has developed in order to adapt to distinct environmental stimuli. An intriguing feature of diatoms specifically is their requirement for silicon at macronutrient concentrations, for cell wall synthesis. As a result, silicon starved diatoms arrest for growth and induce lipid accumulation, with potentially with fewer negative secondary effects such as chlorosis and protein degradation that is often observed during nitrogen depletion (Traller and Hildebrand 2013; Yang et al. 2013; Bender et al. 2014). In an effort to understand metabolic changes involved in lipid accumulation in *C. cryptica*, and more broadly in microalgae, we performed transcript level analysis of cells separately undergoing silicon and nitrogen depletion and compared the two conditions over time, from early (0-24 hour) to late stages (24-96 hours) of nutrient depletion. This data was substantiated with physiological measurements such as photosynthetic efficiency, cell cycle progression, and absolute carbohydrate and lipid quantification.

3.3 MATERIALS AND METHODS

3.3.1 Experimental conditions

1 liter cultures of *Cyclotella cryptica* CCMP332 were grown at 18ºC in artificial seawater (ASW) medium (Darley and Volcani 1969). Cultures were bubbled with air,

mixed using a magnetic stir plate at 250rpm in continuous light (150mol $\mu m^{-2}s^{-1}$) for 4 days to exponential phase before transfer to 19L ASW. Cultures were then grown under the conditions above until exponential phase (8.26×10^5 and 3.57×10^5 cells/mL for experiments 1 and 2 respectively). Cells were harvested by centrifugation for 8 minutes at 3700g and then washed twice in either ASW without silicic acid or nitrate, depending on the condition. Harvested cells were split evenly into two 20L polycarbonate carboys containing Si deplete ASW (Si-ASW) or N deplete ASW (N-ASW). Nutrient deplete cultures were cultured as described above.

Cell concentration was monitored using the Muse Cell Analyzer (EMD Millipore). For photochemical quantum yield of photosystem II (F_v/F_m), cells were dark-adapted for >20 minutes and measured using WALZ WATER-PAM fluorimeter.

3.3.2 Flow cytometric analysis

For lipid and chlorophyll imaging flow cytometry, 10mL of diatom culture was harvested by centrifugation (3700xg, 5 minutes) and cell pellets were kept frozen at -20ºC until analysis. For fluorescence microscopy, 1mL of cell culture was fixed in 1% glutaraldehyde and stored at 4ºC until analysis. Frozen cell pellets were resuspended in ASW and stained with the lipophilic fluorescent dye 4, 4-difluoro-4-bora-3a, 4a-diaza-s-indacene (BODIPY 493/503, Molecular Probes) and processed according to (Traller and Hildebrand 2013) using the ImageStreamX (Amnis Corp). ImageStreamX parameters and post-processing were conducted according to Chapter 1.

For cell cycle analysis, 10mL of cell culture was harvested via centrifugation. Cell pellets were resuspended in ice-cold methanol for pigment extraction and stored at

4ºC until analysis. DNA was stained using SYBR green for cell cycle analysis according to methods in Chapter 1 and analyzed using the Becton Dickinson Influx sorting cytometer (BD Biosciences). Cytometric data were analyzed using FCS Express (De Novo Software) using the software's cell cycle model.

For soluble carbohydrate analysis, 10mL of cell culture was harvested via centrifugation and washed once with 2.3% NaCl. Carbohydrate assay was performed according to (Hildebrand et al. 2017) modified for smaller volumes. Analysis was done in biological and technical duplicate. Carbohydrate experiments were performed separately from RNAseq experiments.

3.3.3 RNA isolation and sequencing

For transcriptomic analysis, at each time point, 800mL of cells were treated with 20µg/mL cycloheximide, harvested by centrifugation, and stored at -80ºC. RNA isolation procedures were followed according to (Hildebrand and Dahlin 2000; Smith et al. 2016b).

For further purification, RNA pellets were chloroform extracted, precipitated with 5M ammonium acetate, washed with ethanol and resuspended in 50µL DEPC water. RNA samples were quality tested using Agilent 4200 TapeStation (Agilent Technologies). Sequencing libraries were prepared according to the Illumina TruSeq Stranded mRNA protocol (Illumina), which selects for poly-adenylated sequences, reducing the amount of total rRNA. All 28 bar-coded libraries were pooled together, and sequenced over two lanes, in technical replicate, using the Illumina HiSeq4000 with 50bp single-end reads. RNA bioanalyzer runs, library construction, and sequencing were

performed at the University California San Diego Institute for Genomic Medicine Genomics Center.

3.3.4 Post sequencing processing and gene expression analysis

Sequencing data was checked for quality and adapter contamination using FastQC (Babraham Bioinformatics). Raw reads were aligned using TopHat, version 2.1.0 on Galaxy platform (Kim et al. 2013; Blankenberg et al. 2014; Afgan et al. 2016). Reads were mapped to the *C. cryptica* genome (Traller et al. 2016), with the maximum read alignment distance = 0, FR first strand, final read mismatches = 2, intron minimum = 40bp, maximum=2000, using *C. cryptica* AUGUSTUS gene models (Traller et al. 2016) to inform mapping, and default parameters otherwise. On average, 94.5% of reads mapped to the *C. cryptica* genome. Technical replicates were combined using BamTools 0.0.2 (Barnett et al. 2011). Raw counts were obtained using HTSeq 0.6.0 on the Galaxy platform (Anders et al. 2015), intersection nonempty mode, stranded = reverse, with a minimum alignment quality of 10. Differential expression and normalized counts were obtained using DEseq2 (Anders and Huber 2010; Love et al. 2014), using the likelihood ratio test, with the reduced design formulas ~replicate, ~replicate + time, and ~replicate + condition. Genes with counts less than 10 across all time points were discarded from analysis.

Log_2 fold change expression values were calculated from the average fold change from experiments 1 and 2 relative to replete conditions. The replete time point was defined as the cells prior to harvest in nutrient free media and is also identified as the -2 h sample, with -2 hour indicating 2 hours prior to the beginning of the nutrient

limitation. These $\log_2$ fold change expression values were then clustered using *k*-means clustering in Genesis 1.1.7 (Sturn et al. 2002). Annotation of genes was derived from (Traller et al. 2016). Heatmaps and PCA plots were generated using the pheatmap and ggplot package in R respectively (Team 2014) .

3.4 RESULTS

3.4.1 Physiological data

As highlighted in chapter 1, when cells were placed in Si-ASW, cell growth immediately ceased. Contrastingly, cells in nitrate deplete medium continued to grow for roughly 48 hours. Growth arrest initiated earlier in N- experiment 2, whereas N-experiment 1 continued to grow, but at a slower rate than replete cells (Figure 3.1A). Variance in growth arrest for cells under nitrate depletion has been demonstrated before (Sheehan et al. 1998; Voronova et al. 2009; Traller and Hildebrand 2013) and could be attributed to differences in vacuolar nitrogen reserves from cell-to-cell (Stolte and Riegman 1995). Cell cycle analysis demonstrated that cells were primarily arrested in the G2 phase during Si starvation, whereas nitrogen starvation arrested in G1 (Figure 3.2). Cell cycle data of the nitrogen experiments further indicated that cell cycle progression ceased in experiment 2 earlier than experiment 1, with the majority of cells in the G1 phase at 24 hours versus 72 for replicate 1 (Figure 3.2A, C). Interestingly, experiment 2 for both silicon and nitrogen limitation experiments arrested earlier than experiment 1. Cells in experiment 2 were harvested and transferred into nutrient replete

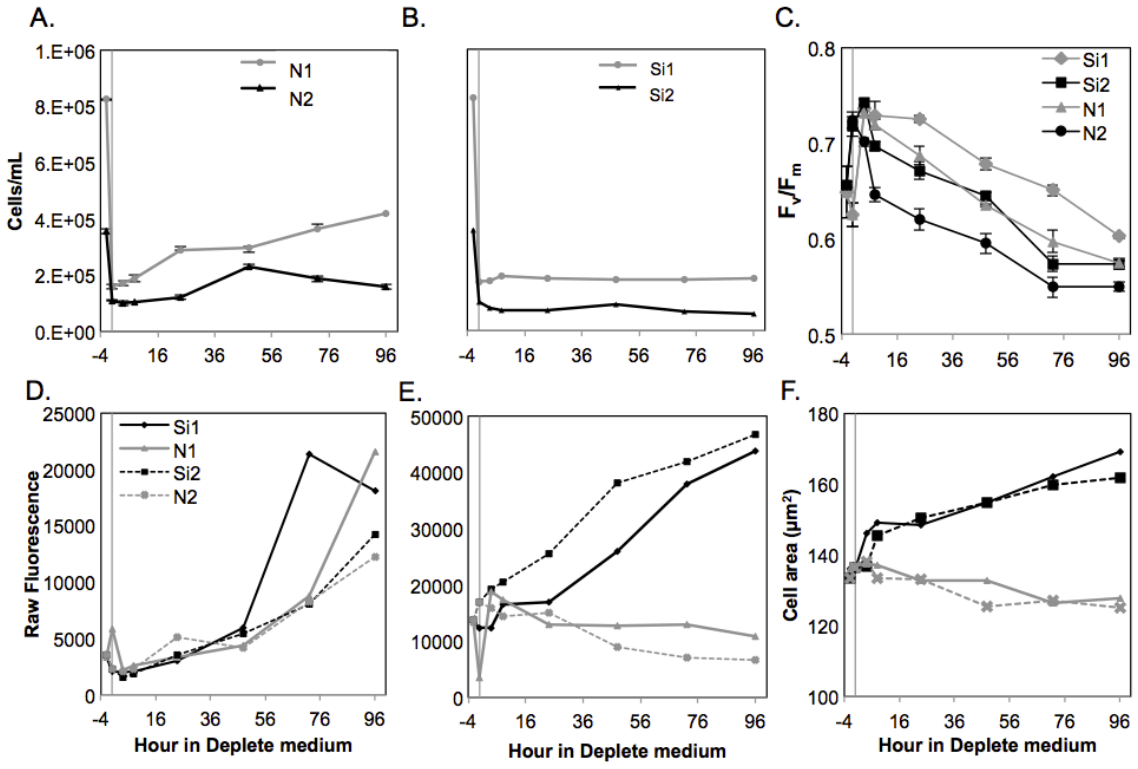


Figure 3.1: Physiological measurement of RNAseq experiments during nutrient depletion. A. Cell growth, nitrogen depletion. B. Cell growth, silicon depletion. C. Photochemical quantum yield as measured by Fv/Fm. D. Relative TAG levels as measured by the lipophilic dye BODIPY. E. Chlorophyll auto fluorescence. F. Cell area measured using the brightfield channel of the imaging flow cytometer. Grey vertical line indicates 0 hour after transfer into nutrient deplete medium. Standard deviation of technical replicates are shown in A-C.

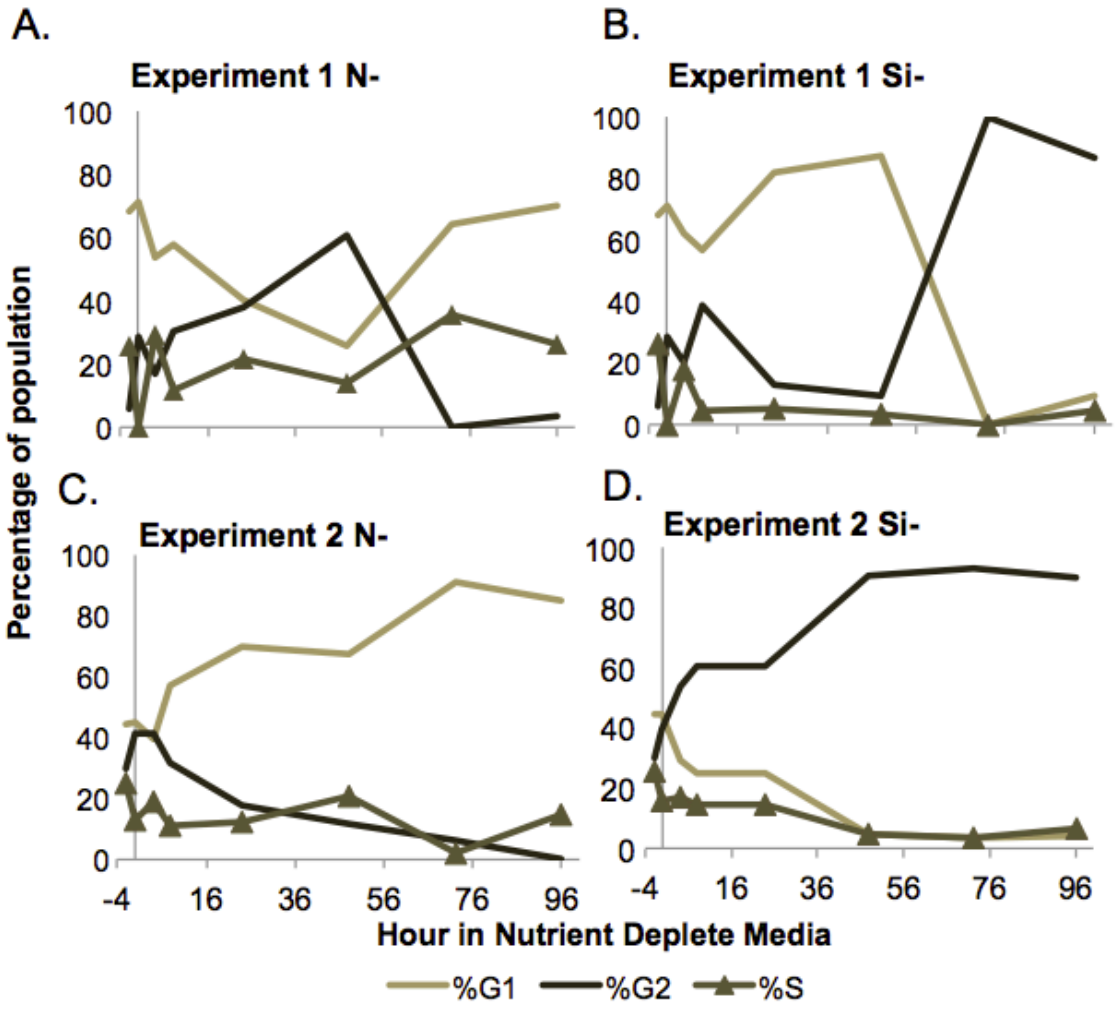


Figure 3.2: Cell cycle analysis on RNAseq replicate experiments. Cell cycle arrest point differs from silicon and nitrogen, with nitrogen cultures arresting in G1 (A, C) and silicon cultures arresting in G2 (B, D). Timing of cell cycle arrest was specific to each experiment indicating that it is more dependent on cell concentration rather than nutrient depletion.

media at a slightly lower cell density than in the exponential phase (Figure 3.1A, B), suggesting that growth and culture phase rather than nutrient stress is more of a factor in the timing of cell cycle arrest.

Under both conditions, photosynthetic activity as measured by Fv/Fm peaked at 4 hours before decreasing, 13.9% and 9.8% under nitrogen and silicon depletion respectively when compared to exponential, nutrient replete conditions (Figure 3.1C). The slight increase at four hours is likely attributed to the increase in light intensity due to the slightly lower culture density after harvest.

Analysis using imaging flow cytometry indicated that despite differences in cell cycle arrest points, all cultures accumulated neutral lipid. Fold change in the BODIPY fluorescence comparing 96 hours starvation to exponential phase was 4.7x and 5.0x in Si and N deplete cultures respectively, with a higher rate of lipid accumulation in the later stages of nutrient stress, roughly 48-96 hours (Figure 3.1D). Chlorophyll fluorescence per cell was dramatically different between silicon and nitrogen deplete cultures, with Si- cultures increasing in fluorescence 3-fold and N- cultures decreasing 20% relative to exponential phase (Figure 3.1E). Changes in chlorophyll fluorescence under either nitrogen or silicon stress have been well documented in *C. cryptica* as well as other diatoms and highlight an important physiological difference between the two methods of nutrient starved lipid induction (Traller and Hildebrand 2013; Bender et al. 2014; Smith et al. 2016b).

There were also notable differences in cell size between the two different starvations (Figure 3.1F). Based cell area measurements from the ImageStream, silicon starved cells were 29% larger than nitrogen-starved cells, and 21% larger than cells in

nutrient replete conditions, while nitrogen deplete cells were 8% smaller compared to replete. This phenomenon may be attributed to the differences in arrest stages for each stress, as cells arrested for growth in G2 phase are about to divide and have already duplicated DNA (Darley and Volcani 1969). Normalization of chlorophyll fluorescence by cell size indicated that the increase in chlorophyll fluorescence in Si cultures is not solely due to increase in cell volume.

3.4.2 RNAseq global assessment

Despite the slight differences in the cell concentration at harvest, gene expression patterns replicated well between experiments (Pearson Correlation Coefficient = .85 for Si-, .90 for N-). Principal component analysis (PCA) of rlog normalized data clustered samples more similar according to condition (Si or N) as well as time (Figure 3.3). Exponential T= -2 h, as well as the early stages of nutrient stress (0-8 hours) were more similar between the two nutrient stresses in contrast to the late stages of starvation (24-96h) which showed condition specific clustering (Figure 3.3).

Differentially expressed genes were determined using the likelihood ratio test in the DEseq2.0 package and changing the design formula to identify genes that change significantly (p-adjusted value < 0.05) across the time series and/or by condition (N or Si). Modifications of the design formula helped to identify significant gene expression patterns, regardless of the condition, as well as patterns that were significant between conditions, that insignificantly change over time. After removing genes with low read count across all samples, 54% (7322/13580) of genes were differentially expressed between the two conditions, while 43% (5839) of genes significantly changed over the

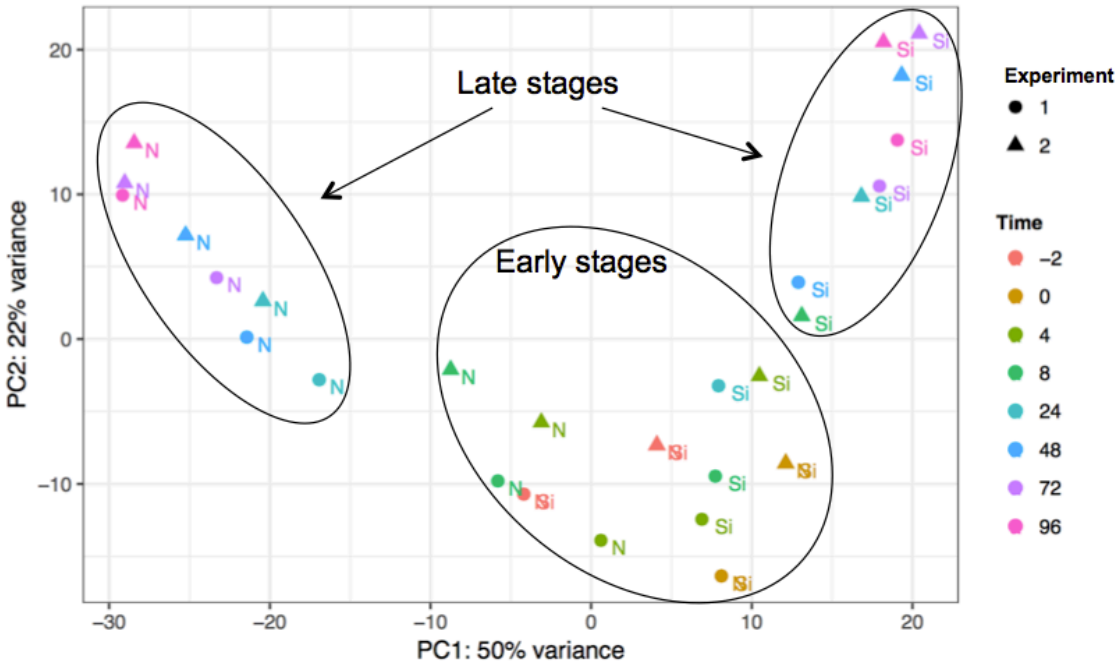


Figure 3.3: Principal component analysis of rlog normalized RNAseq samples from biological replicate experiments. Shape of point depicts experiment and color depicts time point.

time-series, but did not significantly change between conditions. Of these, 2997 genes (22%), or 14% of the whole genome, were considered differentially expressed between both condition and time. Genes expression patterns that were not significant under both design formulas were characterized as either 'time-specific', meaning expression patterns were similar under both nitrogen and silicon but varied with time, or 'condition-specific;' expression patterns with insignificant variance over time, but significantly differed between silicon and nitrogen starvations.

We performed functional KOG class enrichment analysis on differentially expressed genes in the time-specific, condition-specific, or both time and condition subsets, as well as those that have a greater than 1x difference in Log_2 fold change relative to the other nutrient starvation (Figure 3.4). Overall, differentially expressed genes across both time and condition were slightly enriched in processes pertaining to amino acid and inorganic ion metabolism, DNA replication and repair, and energy production and conversion. Under the nitrogen stress-induced subset, there was an enrichment of genes involved in amino acid metabolism and transporter activity (Figure 3.4, i). These genes were specific to the aspartate/arginosuccinate shunt, and the chloroplast GS/GOGAT cycle. There was also a slight enrichment in genes with amino acid KOG class identification in the silicon starvation-induced subset, two examples of genes in this category were arginase and spermidine synthase, enzymes that catalyze the precursors to long chain polyamines which aid in silica polymerization (Allen et al. 2011). Inorganic ion metabolism was enriched in the nitrogen stress-induced subset including two ammonium transporters and a ferredoxin dependent nitrate reductase, which indicates a nitrogen starvation response.

KOG Class	N > Si	Si > N	Time Specific	Condition Specific	Time + Condition
Amino acid transport and metabolism	6.5% (i)	1.3%	0.4%	-0.3%	2.4%
Carbohydrate transport and metabolism	-1.4%	0.4%	-0.7%	0.3%	0.6%
Cell cycle control, cell division, chromosome partitioning	-0.4%	0.0%	-0.4%	0.1%	0.7%
Cell motility	-0.1%	0.6%	0.1%	0.0%	-0.1%
Cell wall/membrane/envelope biogenesis	-0.2%	1.2%	-0.2%	-0.2%	-0.3%
Chromatin structure and dynamics	-0.8%	-1.2%	-1.0%	0.6%	-0.3%
Coenzyme transport and metabolism	-0.4%	1.2%	-0.2%	-0.1%	0.7%
Cytoskeleton	-0.5%	1.0%	0.2%	0.1%	-0.9%
Defense mechanisms	0.8%	-0.3%	0.2%	0.0%	0.2%
Energy production and conversion	-0.4%	2.9%	0.2%	0.2%	1.2%
Extracellular structures	-0.1%	0.6%	0.1%	0.0%	-0.1%
Function unknown	1.5%	-0.9%	0.5%	-0.3%	-1.0%
General function prediction only	6.2%	0.5%	-0.3%	-0.5%	1.5%
Inorganic ion transport and metabolism	5.5% (i)	0.6%	-0.4%	-0.4%	2.1%
Intracellular trafficking, secretion, and vesicular transport	-2.4%	-0.1%	0.2%	0.4%	-1.4%
Lipid transport and metabolism	-2.7%	2.2%	0.0%	0.4%	0.4%
Nuclear structure	0.0%	-0.6%	-0.2%	0.0%	0.1%
Nucleotide transport and metabolism	0.1%	-0.3%	0.1%	0.5%	-0.1%
Posttranslational modification, protein turnover, chaperones	-2.8%	-6.6%	0.1%	1.8%	-4.2%
Replication, recombination and repair	-2.7%	0.9%	0.5%	0.4%	1.8%
RNA processing and modification	-0.1%	-4.6%	2.0%	-0.7%	-1.5%
Secondary metabolites biosynthesis, transport and catabolism	-0.2%	2.7%	-0.9%	0.3%	0.7%
Signal transduction mechanisms	-1.3%	1.6%	-1.7%	-0.6%	1.2%
Transcription	-4.7%	0.0%	-1.3%	-0.4%	-0.6%
Translation, ribosomal structure and biogenesis	0.5%	-3.2%	2.7%	-1.5%	-3.3%
Total KOG class genes	169	148	524	933	583

Figure 3.4: Enrichment analysis based on KOG class gene annotation. Percentages reflect the difference in the proportion of genes within the KOG category, comparing the subset listed (N > Si, Si > N, Time Specific, Condition Specific, Time + Condition) to the total number of KOG annotated genes in the RNAseq dataset. Enriched KOG categories contain positive values and depleted category values are negative. Purple color indicates an enrichment, orange color indicates a depletion, with the richness of the color depending on the extent of enrichment or depletion. White numerals in parenthesis are highlighted in the text. Subsets: N> Si, genes with greater than 1x difference the average Log_2 fold change across all time points comparing nitrogen starvation versus silicon starvation; Si> N, genes with greater than 1x difference the average Log_2 fold change across all time points comparing silicon starvation versus nitrogen starvation; Time Specific; genes that significantly changed (p-adjusted value <0.05) at some point in the time-series; Condition Specific, genes that significantly changed between the two nutrient conditions; Time + Condition, genes that significantly changed in both time and condition subsets. Total number of genes below are only those with KOG class annotation.

There was an enrichment of genes upregulated at the transcript-level in the silicon-starvation subset and downregulated under nitrogen-starvation that were annotated for lipid transport and metabolism and energy production and conversion. Genes in these classes under the silicon starvation subset were part of either mitochondrial beta-oxidation or mitochondrial carrier proteins, including the acylcarnitine/carnitine trasnslocase. This data indicates that the most highly differentiated gene expression patterns between the two nutrient stresses encoded for processes related to nitrogen assimilation and mitochondrial localized pathways such as beta-oxidation and the urea cycle.

It is well documented that cells undergoing a specific nutrient stress will increase transcripts encoding for transporters which import the required element in an effort to scavenge (Hildebrand et al. 1998; Allen et al. 2006; Bender et al. 2014; Smith et al. 2016b). This was prevalent in our experiments, as two silicon transporters (SITs) were among the most upregulated genes in the Si starvation experiments (Figure 3.5A). In the case of g10780.t1, which has also been proven to be an excellent promoter/terminator for expression constructs, there was a 15.5 $\log_2$ fold increase from silicon replete conditions to the zero hour, indicating that this gene undergoes rapid and substantial transcription-level change, within 2 hours (Figure 3.5A). Similarly, genes encoding for nitrate transporters exhibited significant transcript-level increases over time under nitrogen starvation, particularly in the late stages (24-96 hours) (Figure 3.5B). Under this condition, there was also an increase of transcripts encoding urea and ammonium transporters, particularly from 24-96 hours (Figure 3.5B). Interestingly, while cellular growth between the two nitrogen limiting biological duplicates was different (Figure

3.1A), genes involved in nitrogen uptake were well correlated with one another (median PCC= 0.65), suggesting a decoupling of nitrogen sensing mechanisms and cellular growth in diatoms.

Similar to other diatom gene expression studies focusing on nitrate stress, we also observed general decrease over time in chlorophyll and fucoxanthin binding proteins, chlorophyll biosynthesis, and carbon fixation under nitrogen deprivation (Figures 3.6 and 3.7B). Broadly, gene expression peaks from replete to 0 hours, likely due to shifts in light intensity that occur during harvest, then decrease, with transcript levels lower than replete at 8 hours. The down regulation of these processes is indicative of nitrogen stress in the cell and relates to the decrease in global protein content and mild chlorosis phenotype as depicted in figures 3.1C, D as well as in studies in other microalgae (Allen et al. 2011; Hockin et al. 2012; Bender et al. 2014; Schmollinger et al. 2014). During silicon starvation, transcripts for photosynthetic genes did not exhibit as significant of a decrease; chlorophyll a/b binding proteins are slightly decreased in the later stages of limitation, but to a much lesser extent than nitrogen limitation (Figure 3.6 i). Additionally, genes encoding enzymes of the Calvin Benson Cycle had more temporal expression level changes under silicon starvation compared to nitrogen limitation, under which transcripts generally flat-lined after 24 hours (Figure 3.7). Furthermore, in both conditions transcripts encoding for the triose phosphate/phosphoenolpyruvate transporters which export carbon skeletons produced from carbon fixation out of the plastid, also decreased, albeit more so under N- (Figure 3.6). This data as well as the photosynthetic measurements in Figure 3.1C and the increase in chlorophyll

fluorescence under silicon starvation coordinated with decrease under nitrogen starvation (Figure 3.1E, {Smith:2016ed, Traller:2013ch}) demonstrate that in the late stages under both conditions cells experienced a reduction in their photosynthetic capacity, but to a much more significant degree under nitrate depletion versus silicon starvation.

The similarities in chlorophyll biosynthetic gene expression patterns between nitrogen and silicon prompted us to question the dramatic increase in auto fluorescence observed under silicon starvation. Smith and colleagues attributed the increase in chlorophyll fluorescence to chloroplast replication and biosynthesis (Smith et al. 2016a). Interestingly some of the most dramatic differences in gene expression patterns between silicon and nitrogen were transcripts encoding nitrogen assimilation processes, particularly the chloroplast localized GS/GOGAT cycle, an important metabolic intersection of carbon and nitrogen metabolism (Allen et al. 2011), which increased in N- and decreased in Si-. Glutamate, a key metabolite in the GS/GOGAT cycle, is also the precursor molecule for chlorophyll biosynthesis, and it is possible that the increase in pigment content in diatoms during silicon stress may also be a reallocation of glutamate from GS/GOGAT cycle, toward chlorophyll biosynthesis. Since the cells are no longer dividing, there is less of a need to produce organic nitrogenous compounds that would have been directed for growth.

3.4.3 Mitochondrial metabolic reprogramming during nutrient starvation

Diatoms are unique from other classes of algae in that they have a complete lower half of glycolysis localized in the mitochondria, which is hypothesized to be the primary supply of carbon and energy into the Citric Acid (TCA) Cycle and cellular

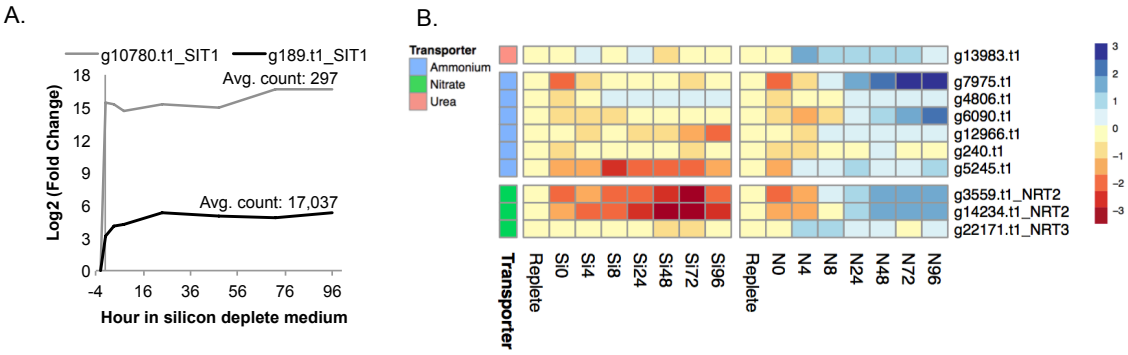


Figure 3.5: Substantial transcript-level changes in nutrient transporters. A. Gene expression patterns of two silicon transporters during silicon starvation with the average normalized count across the time series experiment. B. Heatmap depicting changes in Log_2 Fold change relative to the replete condition of nitrogen transporters. Largest transcript-level changes were observed in the nitrate transporters.

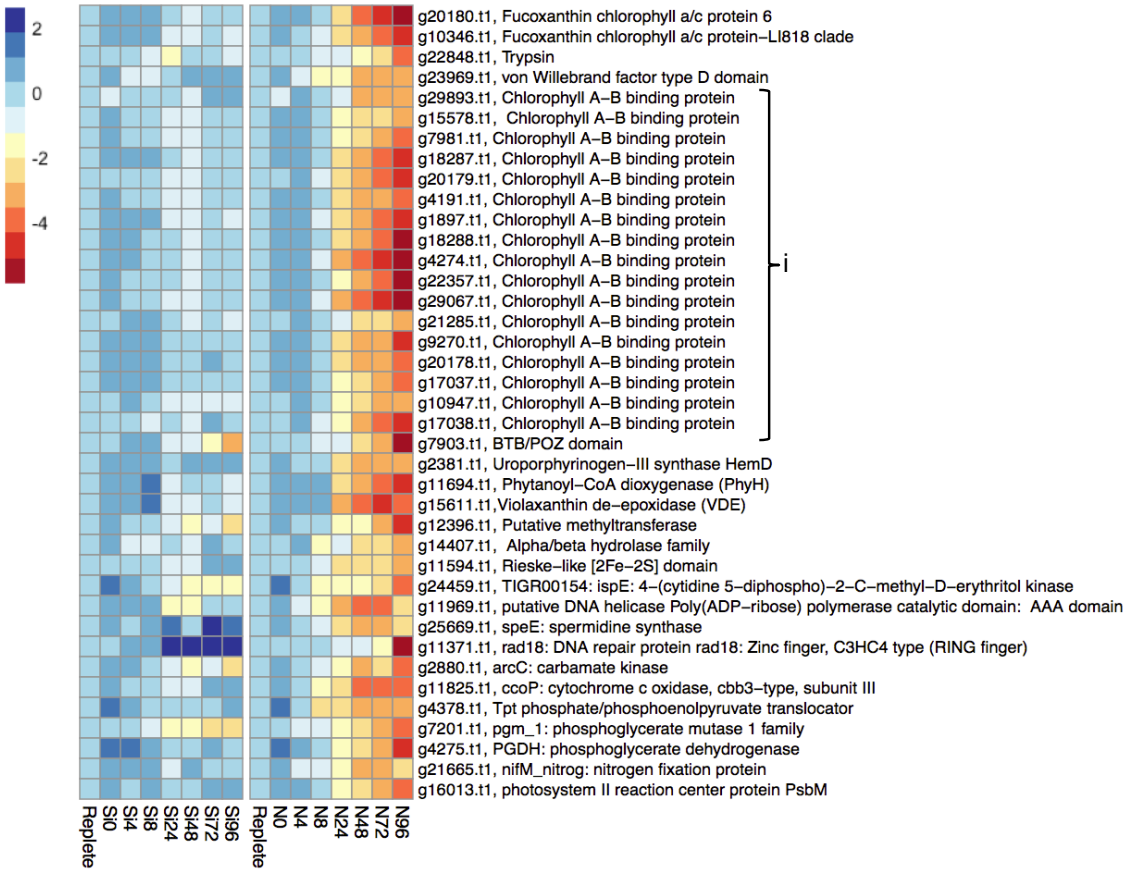


Figure 3.6: Transcript level changes of genes belonging to nitrogen starvation cluster 11 belonging primarily to chloroplast processes. Log_2 Fold Change relative to replete condition depicts significant decrease in transcripts in the later hours of nitrogen starvation, and moderate decrease of transcripts under silicon deprivation. Bracketed genes labeled i are referred in the text.

respiration (Kroth et al. 2008; Smith et al. 2012; Smith et al. 2016b; Smith et al. 2016a). In *C. cryptica* mitochondrial glycolytic gene expression patterns were distinctly different between silicon and nitrogen starvation. Transcripts under both conditions decreased to a minimum at 8 hours, but after 8 hours, glycolytic genes increased under nitrogen starvation and decreased under silicon starvation (Figure 3.8). Phosphoenolpyruvate carboxylase, which catalyzes the unidirectional carboxylation of phosphoenolpyruvate to oxaloacetate, was coordinately expressed with mitochondrial triose phosophate isomerase-glyceraldehyde-3-phosphate dehydrogenase (TPI-GAPDH) and enolase under nitrogen depletion (Figure 3.8). The coordination of mitochondrial glycolysis with the pyruvate hub indicates that during nitrogen stress, the mitochondrion is functioning to provide carbon skeletons and energy for the TCA cycle, genes of which also exhibited similar expression patterns (Figure 3.8B). Interestingly, under nitrogen depletion, there was co-expression and upregulation of a chloroplast localized triose phosphate isomerase, TPI2A (Figure 3.7B, 3.8B), which was not observed under silicon depletion, suggesting that this enzyme may aid in supplying the mitochondrion with glyceraldehyde-3-phosphate to sustain mitochondrial glycolysis. Under silicon deprivation, significant increase at the transcript-level of mitochondrial beta-oxidation, isocitrate lyase, and malate synthase indicated that beta-oxidation and glyoxylate cycle enzymes provide energy and carbon skeletons rather than mitochondrial glycolysis and the TCA cycle. The glyoxylate cycle, which partially occurs in the peroxisome, creates a non-cyclic TCA cycle, circumventing decarboxylation and NADH producing steps (Davis et al. 2017). Transcripts for genes encoding for key enzymes in the mitochondrial pyruvate hub, including pyruvate carboxylase, phosphoenolpyruvate carboxykinase, and

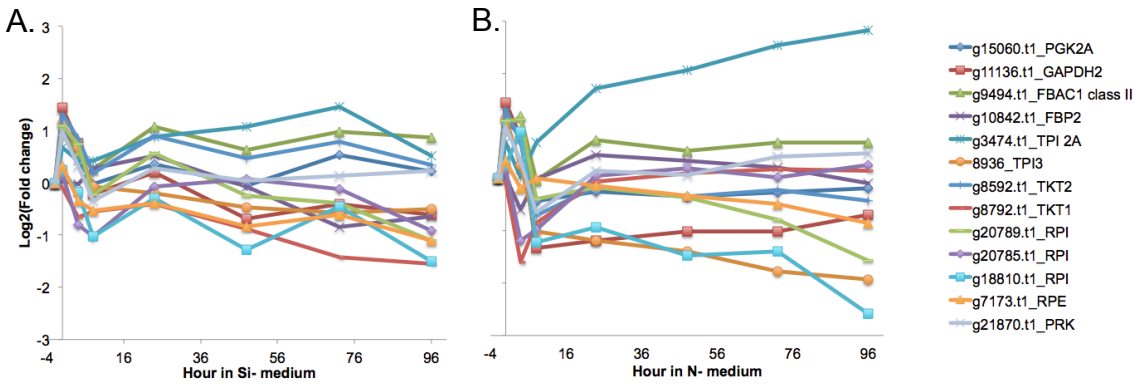


Figure 3.7: Transcript level changes for genes putatively involved in the Calvin Benson cycle. A, silicon starvation; B, nitrogen starvation.

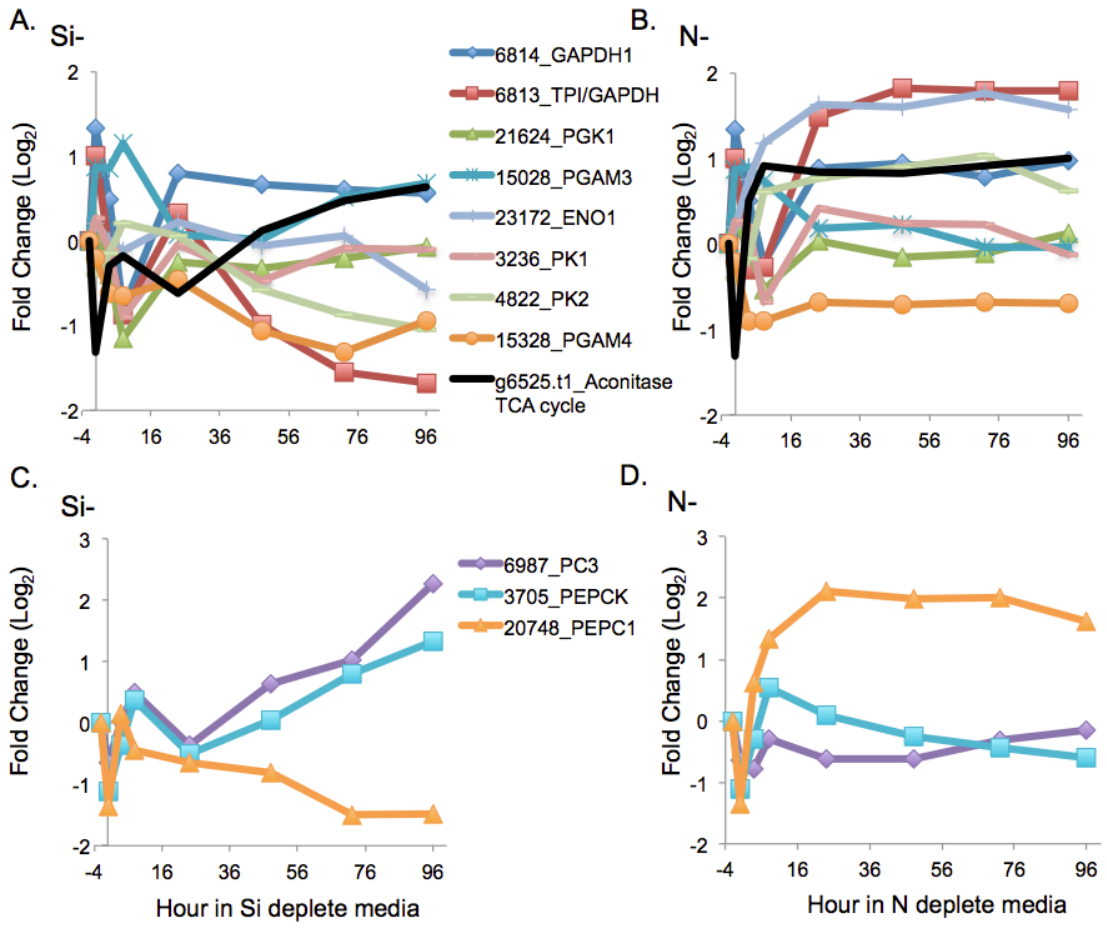


Figure 3.8: Transcript level changes of mitochondrial targeted genes encoding for glycolysis and pyruvate metabolism. Glycolytic genes expression patterns during silicon starvation (A) and nitrogen starvation (B). Gene expression patterns of genes involved in pyruvate metabolism during silicon starvation (C) and nitrogen starvation (D).

malate dehydrogenase also increased under silicon stress. This data demonstrates that substantial differences in mitochondrial carbon and energy production occur under nitrogen and silicon stress. Implications of which still remain unclear.

3.4.4 Fluctuations in carbohydrate content and gene expression

We quantified soluble carbohydrate, primarily chrysolaminarin, levels on a per cell basis using a colorimetric assay developed in (Chiovitti et al. 2004; Hildebrand et al. 2017). Under silicon deprivation, beta 1,3 glucan (chrysolaminarin) accumulation fluctuated temporally; peaking at 24 hours before decreasing to near similar levels as replete conditions by 96 hours (Figure 3.9A). Alternatively under nitrogen stress, beta 1,3 glucan levels increased at a slower rate, with no observable peak at 24 hours. Under both conditions, genes encoding for beta 1,3 glucan biosynthesis were not coordinately expressed, and expression varied between conditions, suggesting that carbohydrate biosynthesis is more subject to post transcriptional, mass action, or allosteric regulation or that degradation rather than biosynthesis is a more important element in carbohydrate accumulation in the cell. Transcripts encoding for carbohydrate breakdown genes significantly increased over the time course under both conditions, two genes under silicon stress and three under nitrogen. The large increase in glucanase transcripts over the nitrogen depletion time-course, more than over the silicon starvation time course, may explain the lower carbohydrate per cell comparatively.

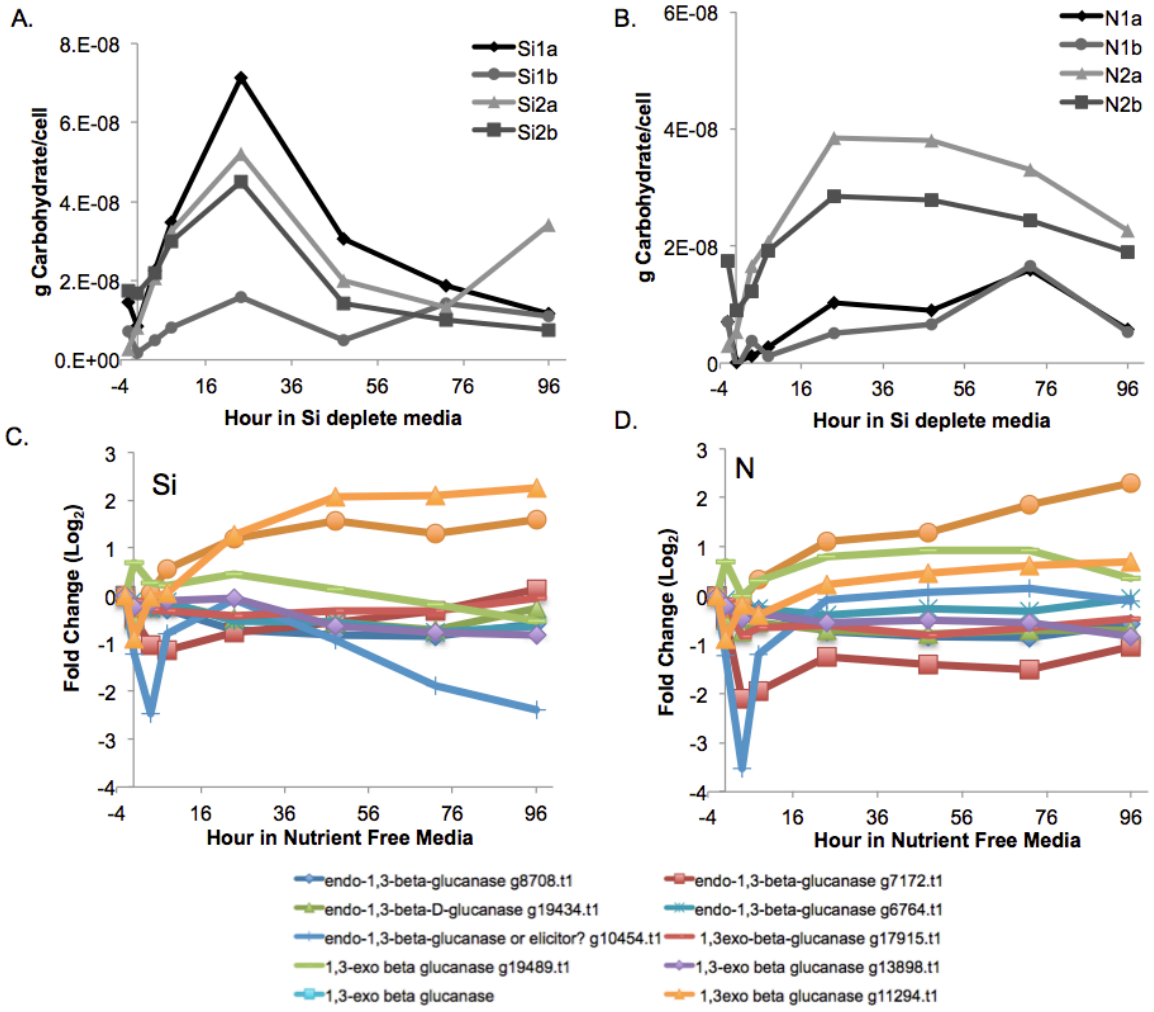


Figure 3.9: Chrysolaminarin degradation gene expression patterns as it relates to total soluble carbohydrate content. Carbohydrate assay were run under similar experimental conditions as cultures for RNAseq. Grams soluble carbohydrate normalized per cell over the time in silicon (A) and nitrogen (B) deplete medium. Gene expression patterns from RNAseq experiments under silicon (C) and nitrogen (D) deprivation of endo- and exo- 1,3 beta glucanases, which are responsible for enzymatic breakdown of the main carbohydrate in *C. cryptica,* chrysolaminarin.

3.4.5 Fatty acid and triacylglycerol biosynthetic processes

Genes for *de novo* fatty acid biosynthesis, a process that occurs in the chloroplast, showed similar expression patterns to Calvin-Benson cycle genes, broadly decreasing under nitrogen starvation, and moderately increasing during silicon starvation (Figure 3.10A, B). Transcripts for the cytoplasmic Acetyl-CoA carboxylase (ACCase), which is essential for very long chain fatty acid (VLCFA) and flavonoid biosynthesis (>20 carbon, (Baud et al. 2003)) is strongly decreased under both conditions, suggesting down regulation of VLCFA production. Transcripts encoding the chloroplast ACCase did not significantly change over time, and actually slightly decreased under silicon deprivation (Figure 3.10A), indicating that under nutrient starvation, transcript-level regulation of ACCase does not play a large role in lipid accumulation. This data is in alignment with previous results that did not show a significant increase in lipid content in *C. cryptica* transformant cell lines engineered to overexpress ACCase transcripts (Sheehan et al. 1998).

Annotation of the *C. cryptica* genome led to the discovery of an expansion in genes encoding for acyl-CoA monoacylglycerol acyltransferase (MGAT) and acyl-CoA diacylglycerol acyltransferase (DGAT), enzymes which catalyze consecutive steps in triacylglycerol biosynthesis, with DGAT being the only committed step (Traller et al. 2016). Gene expression patterns for MGAT, DGAT1, and Type 2 DGAT (DGTT) were diverse and varied over time, especially under silicon deprivation (Figure 3.10C, D) suggesting that each isozyme may have evolved to be utilized under different physiological conditions. The most abundant is for DGTT, g3706.t1, which does not share a homolog in other diatoms, is significantly up regulated at the transcript level

under nitrogen deprivation, and only moderately up regulated under silicon deprivation, peaking at 48 hours. Conversely, g9591.t1, encoding DGTT has moderate levels of transcript, which increased 3x (log_2 fold change) under silicon deprivation, peaking at 72 hours (Figure 3.10C, D).

Overall, under both nutrient stresses, transcription-level changes in fatty acid biosynthesis are moderate, suggesting that in *C. cryptica,* induction of de novo fatty acids synthesis enzymes is not a required to substantially increase carbon flux though that pathway. Significant increases in specific MGAT and DGAT enzymes, which is distinct depending on the nutrient stress, demonstrates the importance of these enzymes in nutrient starved lipid accumulation and indicates fundamental differences in the triacylglycerol biosynthetic enzymatic machinery between silicon and nitrogen starvation.

3.5 DISCUSSION

We compared over time the transcript profiles of two distinct macronutrient limitations, silicon and nitrogen, in the oleaginious diatom *Cyclotella cryptica.* While there are distinct physiological differences comparing the two limitations, both conditions create a substantial accumulation of neutral lipids, which can be utilized for biofuel production. Comparison between the two conditions allowed for the opportunity to detangle the nutrient stress response from the lipid accumulation response in *C. cryptica* in an effort to understand the key enzymes responsible for lipid droplet accumulation in microalgae.

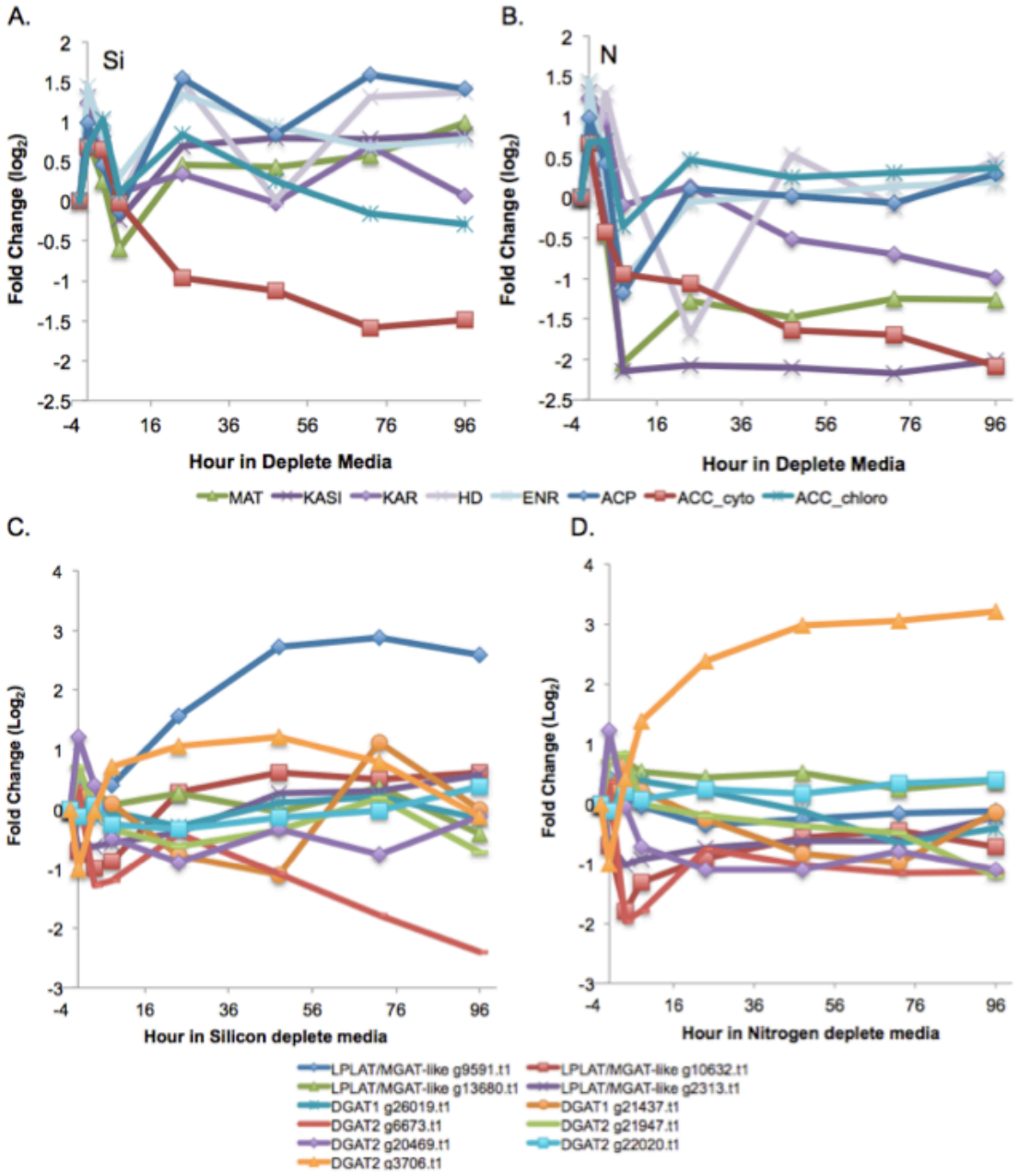


Figure 3.10: Transcript level changes of genes involved fatty acid and TAG biosynthesis. Fatty acid biosynthesis transcripts under silicon (A) and nitrogen (B) deprivation. C, D: Gene expression patterns of putative mono- and di-acylglycerol acyltransferases abbreviated MGAT and DGAT, respectively. C, Silicon starvation; D, Nitrogen starvation.

Most lipid-induced nutrient limitation or starvation transcript studies sample under replete, exponential conditions as a control, and compare with deplete or limiting conditions, collecting data from only one or two instances during the limitation. In this particular experimental design, conclusions are drawn from fixed conditions, which miss subtle shifts in gene expression as well as underestimate the temporal regulation. In our data, gene expression patterns in silicon-deprived cells were highly dynamic; particularly genes involved in carbon fixation and cytosolic glycolysis, indicating that there are important physiological and molecular shifts across the time course with regards to carbon flux that will likely factor into the global cellular response to nutrient starvation. The time course sampling therefore allowed to identify different transcript phases that might attribute to the differences in rates TAG accumulation between the early and late stages of nutrient starvation.

It was apparent under both conditions that there were two distinct phases in gene expression over the time course, with the first being coordinated transcription shifts resulting from harvest of cells from replete to deplete conditions from 0 to 8 hours. At the time of harvest, cells undergo large shifts in light intensity, including a brief period during centrifugation where the cells are in the dark, followed by resuspension at a more dilute density, in the light – effectively increasing light exposure per cell. Additionally, cells are immediately washed twice with medium deplete of either silicon or nitrogen. These highly fluctuating abiotic shifts cause dramatic transcriptional change, demonstrated in gene expression patterns shown in figure 3.8, 3.9, and 3.10C, D. PCA analysis (Figure 3.3) also revealed that the 4 and 8 hour samples regardless of the condition were more similar to the replete T-2 time points than they were to the 0 hour

time points. Smith and colleagues (Smith et al. 2016b) observed a similar transcriptional shift in chloroplast targeted genes during silicon limitation in the closely related *T. pseudonana* and attributed it to an enzymatic need to populate newly dividing chloroplasts. Differences between the pre-harvest T= -2 h and the T=0 h time point, observed under both nutrient limitations, as well as the transcript coordination of genes targeted to other cellular locations besides the chloroplast, such as the mitochondria and cytosol, suggest that this transcriptional response is more likely an harvest effect than due to increase in total population of chloroplasts. This method of experimentation may be argued to not be ideal because the cells may undergo unnecessary stress that could limit production of desired product. However, this method of experimentation can also be seen as a good strategy for production systems and experimental design because it causes cell synchronization and creates a reference point for onset of nutrient starvation. Given the level of heterogeneity that was documented in chapter 1, synchronization of cells is essential for understanding temporal changes that may occur in cellular physiology that occurs during nutrient stress.

The second transcript level response observed in both conditions occurred from 48-96 hours and appeared to have been a chronic response to nutrient stress. Under nitrogen stress conditions, this was characterized by decrease of photosynthetic processes, and shutdown for chloroplast function, including fatty acid and chlorophyll biosynthesis, similar to findings depicted in other diatoms and microalgae. Cells under silicon deprivation also experienced decreases in chloroplast processes, but to a much lesser extent. The genes with the most significant transcript-level decrease in the later stages of silicon deprivation encoded for processes pertaining to in cobalamin

biosynthesis, and nitrogen metabolism, which suggest a shutdown of processes pertaining to growth and division. Contrastingly though, we did observe a significant increase in genes putatively involved in biosynthesis of compounds needed in cell wall formation, such as chitin and long chain polyamines, indicating that the cell may could be synthesizing components of the cell wall to prepare for when silicon is available again.

Due to the difference in the function of each macronutrient in the cell, there are distinct differences with regards to energetic shuttling and carbon flux between the two different conditions. The substantial transcript increase of mitochondrial beta oxidation in the late stages of silicon deprivation was surprising, but the coordinated expression of genes involved in the peroxisomal glyoxylate cycle suggest that the cell has developed a unique, multi-organelle path toward energy production through the breakdown of lipids and non-cyclic flow through the TCA cycle. It is not yet known why the cell may be doing this, but one hypothesis is that it is a mechanism to help lower the C:N ratio, which would have increased due to lipid accumulation. The decrease of amino acid biosynthesis and nitrogen assimilation processes and increase of beta-oxidation in our data suggest that under silicon deprivation, *Cyclotella cryptica* experiences a metabolic metamorphosis focused on cellular homeostasis and survival rather than division and growth. The depletion of carbohydrate after 24 hours may also be connected to this, as the cell increases glucanase transcripts to breakdown carbohydrate that would normally be utilized for cellular division, and increases carbon flux toward lipid droplets, which can be further utilized for energy production in the cell.

Cessation of the cell division under nitrogen depletion is not as robust as silicon, presumably due to intracellular nitrogen stores that vary from cell-to-cell depending on specific stage in the cell cycle at that moment. Cells under external nitrogen depletion still continue carbon flux through mitochondrial glycolysis and the TCA cycle, despite the limiting nitrogen reserves. This, combined with the observation of the increase in transcripts encoding for nitrogen scavenging mechanisms, create dramatic differences in mitochondrial metabolism compared to silicon. The scavenging effect under nitrogen depletion likely causes damage to essential carbon fixation and photosynthetic processes, which is demonstrated in the transcriptional decrease of genes. Under silicon deprivation, based on gene expression patterns, photosynthetic measurements, and increase in chlorophyll pigmentation, cellular damage appears to be less prevalent, which is likely attributed to the fact that silicon is only used for cell wall synthesis. These transcript level differences between nitrogen and silicon starvation in *C. cryptica* highlight the decoupling of silicon metabolism from carbon metabolism in diatoms, which is in contrast to the intertwining of nitrogen and carbon metabolism (Claquin et al. 2002). This advantage of a non-carbon based substrate for cell walls may contribute to the ecological success of diatoms in the open oceans. Cells undergoing silicon starvation appear to have sophisticated mechanisms to adapt and cope with the stress, which is not the case with nitrogen starvation. These differences between the two nutrient starvations should be noted when deciding how to increase TAG in a production system.

While there lie fundamental differences in carbon metabolism and energy production between the two different nutrient starvations, we did observe shared transcript increases in key triacylglycerol biosynthetic and chrysolaminarin breakdown

genes under both conditions. Interestingly, however, the specific isozymes in these processes were different between silicon and nitrogen starvation. While *C. cryptica* has a complex arsenal of enzymes that regulate triacyglycerol in the cell, overall, the data indicates that the lipid accumulation phenotype common to both nutrient starvations is primarily due to a reallocation of carbon from carbohydrate to triacylglycerol.

3.6 CONCLUSIONS

The RNAseq study of silicon and nitrogen depletion in *Cyclotella cryptica* revealed dramatic transcript level shifts in carbon and energy production, particularly in genes involved in pyruvate metabolism, glyoxylate cycle, and the TCA cycle. The data also highlighted preference of key regulatory isozymes, such as DGAT, in each nutrient stress. Altogether the data indicate that the despite the lipid accumulation phenotype common to both nutrient starvations, primary carbon metabolism is unique to each stress, likely because of the entanglement of nitrogen with carbon metabolism and lack thereof with silicon. Additionally, over the long 96 hour time course experiment, gene expression patterns vary, highlighting the importance in time-course evaluation of physiological processes associated with a production system.

3.7 ACKNOWLEDGEMENTS

Chapter 3, in full is material currently being prepared for submission for publication. Traller, Jesse; Hildebrand, Mark. “Comparison of the transcript level and physiological response in the diatom *Cyclotella cryptica* under two distinct macronutrient limitations.” Ms. Traller was the principal researcher/author on this paper.

3.8 REFERENCES

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Kuster Von G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Research 44:W3–W10.

Alipanah L, Rohloff J, Winge P, Bones AM, Brembu T (2015) Whole-cell response to nitrogen deprivation in the diatom *Phaeodactylum tricornutum*. Journal of Experimental Botany 66:6281–6296.

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, Vardi A, Bowler C (2006) An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. Current Opinion in Plant Biology 9:264–273.

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biology 11:R106.

Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169.

Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT (2011) BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27:1691–1692.

Baud S, Guyon V, Kronenberger J, Wuillème S, Miquel M, Caboche M, Lepiniec L, Rochat C (2003) Multifunctional acetyl-CoA carboxylase 1 is essential for very long chain fatty acid elongation and embryo development in *Arabidopsis*. Plant J 33:75–86.

Bender SJ, Durkin CA, Berthiaume CT, Morales RL, Armbrust EV (2014) Transcriptional responses of three model diatoms to nitrate limitation of growth. Frontiers in Marine Science 1:4.

Blankenberg D, Kuster Von G, Bouvier E, Baker D, Afgan E, Stoler N, Galaxy Team, Taylor J, Nekrutenko A (2014) Dissemination of scientific software with Galaxy ToolShed. Genome Biol 15:403.

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.

Claquin P, Jézéquel VM, Kromkamp JC, Veldhuis M, Kraay G (2002) Uncoupling of silicon compared with carbon and nitrogen metabolism and the role of the cell cycle in continuous cultures of *Thalassiosira pseudonana* (Bacillariophyceae) under light, nitrogen, and phosphorous control. J Phycol 38:922–930.

Darley WM, Volcani BE (1969) Role of silicon in diatom metabolism. Experimental Cell Research 58:334–342.

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.

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.

Guihéneuf F, Leu S, Zarka A, Khozin-Goldberg I, Khalilov I, Boussiba S (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.

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.

Hildebrand M, Dahlin K, Volcani BE (1998) Characterization of a silicon transporter gene family in *Cylindrotheca fusiformis*: sequences, expression analysis, and identification of homologs in other diatoms. Molecular and General Genetics MGG 260:480–486.

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.

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.

Hovde BT, Deodato CR, Hunsperger HM, Ryken SA, Yost W, Jha RK, Patterson J, Monnat RJ Jr, Barlow SB, Starkenburg SR, Cattolico RA (2015) Genome Sequence and Transcriptome Analyses of *Chrysochromulina tobin* : Metabolic Tools for Enhanced Algal Fitness in the Prominent Order Prymnesiales (Haptophyceae). PLoS Genet 11:e1005469.

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:R36.

Kroth PG, Chiovitti A, Gruber A, Martin Jézéquel 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.

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:31.

Manandhar-Shrestha K, Hildebrand M (2015) Characterization and manipulation of a DGAT2 from the diatom *Thalassiosira pseudonana*: Improved TAG accumulation without detriment to growth, and implications for chloroplast TAG accumulation. Algal Research 12:239–248.

Schmollinger S, Mühlhaus T, Boyle NR, Blaby IK, Casero D, Mettler T, Moseley JL, Kropat J, Sommer F, Strenkert D, Hemme D, Pellegrini M, Grossman AR, Stitt M, Schroda M, Merchant SS (2014) Nitrogen-Sparing Mechanisms in *Chlamydomonas* Affect the Transcriptome, the Proteome, and Photosynthetic Metabolism. Plant Cell 26:1410–1435.

Sheehan J, Dunahay T, Benemann J, Roessler P (1998) A look back at the US Department of Energy's Aquatic Species Program: Biodiesel from algae.

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.

Stolte W, Riegman R (1995) Effect of phytoplankton cell size on transient-state nitrate and ammonium uptake kinetics. Microbiology 141:1221–1229.

Sturn A, Quackenbush J, Trajanoski Z (2002) Genesis: cluster analysis of microarray data. Bioinformatics 18:207–208.

Tanaka T, Maeda Y, Veluchamy A, Tanaka M, Abida H, Maréchal E, Bowler C, Muto M, Sunaga Y, Tanaka M, Yoshino T, Taniguchi T, Fukuda Y, Nemoto M, Matsumoto M, Wong PS, Aburatani S, Fujibuchi W (2015) Oil accumulation by the

oleaginous diatom Fistulifera solaris as revealed by the genome and transcriptome. Plant Cell 27:162–176.

Team RC (2014) The R project for statistical computing. Available at www. R-project. org/. Accessed October

Traller JC, Cokus SJ, Lopez DA, Gaidarenko O, Smith SR, McCrow JP, Gallaher SD, Podell S, Thompson M, Cook O, Morselli M, Jaroszewicz A, Allen EE, Allen AE, Merchant SS, Pellegrini M, Hildebrand M (2016) Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. Biotechnol Biofuels 9:258.

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.

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 110:19748–19753.

Voronova EN, Il'ash LV, Pogosyan SI, Ulanova AY (2009) Intrapopulation heterogeneity of the fluorescence parameters of the marine plankton alga *Thalassiosira weissflogii* at various nitrogen levels. Microbiology 78:469-478.

Yang J, Pan Y, Bowler C, Zhang L, Hu H (2016) Knockdown of phosphoenolpyruvate carboxykinase increases carbon flux to lipid synthesis in *Phaeodactylum tricornutum*. Biotechnol Biofuels 6:1-14.

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 6:67.