Quantifying the response of diatom nitrogen metabolism to environmental changes

Sara Jane Bender

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2013

Reading Committee:
E. Virginia Armbrust, Chair
Jody W. Deming
Anitra Ingalls

Program Authorized to Offer Degree:
School of Oceanography

University of Washington

Abstract

Quantifying the response of diatom nitrogen metabolism to environmental changes

Sara Jane Bender

Chair of Supervisory Committee:
Dr. E. Virginia Armbrust
Oceanography

Abstract

Diatoms are microscopic unicellular organisms that fix as much carbon annually as all terrestrial rainforests combined and, through nitrogen uptake and carbon drawdown, link global carbon and nitrogen cycles. As the most diverse group of marine phytoplankton, species-specific metabolic differences influence diatom distributions and abundance, and impact nitrogen availability in the marine environment. This thesis investigates the molecular underpinnings of diatom metabolism by focusing on the importance of nitrogen in controlling diatom growth in the marine environment. First, the role of the urea cycle in diatom nitrogen metabolism was investigated by monitoring changes in transcript abundance patterns for key nitrogen metabolism genes in *Thalassiosira pseudonana* grown under different nitrogen sources and light intensities on a light:dark cycle. Findings integrate the urea cycle into diatom metabolism, through

glutamine utilization and urea production, and demonstrate the effects of light and nitrogen source on the flow of nitrogen in a diatom cell. Second, the responses of three evolutionarily diverse model diatoms (*Thalassiosira pseudonana*, *Fragilariopsis cylindrus* and *Pseudo-nitzschia multiseries*) to the environmentally relevant condition of nitrate starvation were examined to test the hypothesis that changes in nitrogen availability elicit uniform metabolic responses across all diatoms. Greater transcriptional and functional similarities between the two pennate diatoms compared to the bipolar centric diatom form fundamental differences in how each group has evolved to respond to its environment. Each diatom appears to maintain its own cell-wide response to nitrogen availability that is activated via finely tuned transcriptional regulation. Third, to understand how these diverse transcriptional responses affect diatom community composition, metatranscriptional patterns were determined for diatom communities sampled along Line P in the Northeast Pacific Ocean. Phylogenetic analyses revealed that individual members of the diatom community responded to the same bulk chemical environment in vastly different ways. In the field, these species-specific differences likely play an important role in determining why one diatom species blooms over another. These complex transcript abundance patterns provide insight into the molecular mechanisms that underlie diatom distributions in the marine environment; future studies can use these findings as a platform for detecting and predicting the metabolic responses of individual species and taxonomic groups within a natural diatom assemblage.

"If I have seen further, it is by standing on the shoulders of giants."

-Isaac Newton

Acknowledgements

I would like to begin by thanking my fearless advisor, Dr. E. Virginia Armbrust. You have served many roles, from advisor to teacher to therapist, and each one has gently guided me through this incredible experience. I am in constant awe of your positive attitude, sharp scientific eye, and unwavering poise. How you are able to transition seamlessly between projects in the matter of seconds will forever be a mystery to me. I have looked forward to our meetings and conversations, knowing that your support, calming words, creativity, and sense of humor would get me through whatever obstacle I was facing at the time. Thank you for all of your guidance, for being a stalwart role model, and for teaching me how to do "good science."

I would also like to thank my committee members: Dr. Jody W. Deming, Dr. Anitra Ingalls, Dr. Steven Robertson, and Dr. Gabrielle Rocap. I have come to enjoy our bi-annual committee meetings, and I have always been touched by the thoughtfulness and genuine interest each of you has expressed in my research project and in my future career plans. I will always look back on my time at the University of Washington with positive memories.

This research would not have been possible without my collaborators: Dr. Micaela Schnitzler-Parker, Dr. Colleen A. Durkin, Chris Berthiaume, Rhonda L. Marohl-Morales, and Bryndan P. Durham. Also, thank you to Dr. Julie Koester, whose support guided me through each project. Additional technical help for these projects was provided by Jeff Bowman, Claire Ellis, Vaughn Iverson, Dr. Robin Kodner, Dr. Adrian Marchetti, Irina Oleinikov, Rita Peterson, Meghan Schatz, David M. Schruth, and Dr. Betsy Welsh. I would also like to thank my labmates past and present, particularly Dr. Shady Amin, Dr. Kate Hubbard, Dr. Thomas Mock, Dr.

Mikelle Rasmussen, Dr. Francois Ribalet, and Jessica Silver, all of whom have mentored me, inspired me, and helped me through so many experiences. Thank you to all of the members of the Center for Environmental Genomics, including Michael Carlson, Gwenn Hennon, Katie Marshall, Jacyln Saunders, and Helena van Tol, for providing such a positive and fun working environment. Additional thanks to the 2006 "First Years" for becoming my first family in Seattle: Nicholas Beaird, Alison Gray, Vaughn Iverson, Emily Nahas-Reistetter, and Kelsey van der Elst. And, finally, special thanks to Dr. Karen Chan, Dr. Bonnie Chiang, Dr. Eric Collins, Dr. John Kirkpatrick, and Dr. Sally Warner for their friendship.

Thank you to my funding sources: The Gordon and Betty Moore Foundation, as well as to the UW School of Oceanography and the National Science Foundation grant to the Center for Coastal Margin Observation and Prediction (CMOP) program. Many thanks to the crews of the R/V Thomas G. Thompson, R/V Wecoma, R/V New Horizon, and the R/V Western Flyer for allowing me to flex my seagoing skills on multiple wonderful cruises. I would also like to acknowledge Dr. Colleen A. Durkin and Dr. Adrian Marchetti, who taught me nearly everything I know about life at sea. I now have a new appreciation for duct tape.

To my early mentors, Dr. Thomas Bibby and Dr. Kay Bidle, who sent me out to sea, and opened the door to the field of biological oceanography, thank you. I am also eternally grateful for the 2011 C-MORE Summer Course at the University of Hawaii and Drs. David Karl and Matthew Church for bringing together an incredible group of scientists, including Team members Dr. Thomas Weber, Bryndan Durham, Kerry Whittaker, Santhiska Pather, and my partner-in-crime, Andreas Krupke.

I would also like to thank my Seattle family for making this city my home, along with my right coast family: John E. Bender, Tess I. Bender, Elaina D'Orto Jack, Ric Jack, Maureen

Wright, and Allison Gutknecht, my editor and best friend. Finally, thank you to my parents, John S. Bender, Lorraine Kulik-Bender, Ernest D'Orto and Jane A. D'Orto for supporting me every step of the way.

Dedication

This body of research is dedicated to the life and memories made on 403 Morris Boulevard, Beach Haven West, Manahawkin, New Jersey, and to my incredible and resilient family -- we made it through the storm, and we are stronger because of it.

Table of Contents

List of Figures

List of Tables

Introduction

Marine phytoplankton are a vital component of Earth's ecosystems; they account for < 1% of Earth's biomass, and are responsible for half of the oxygen produced on the planet at any given time (Falkowski 1994; Field et al. 1998). They are globally distributed and sustain life across diverse aquatic ecosystems by supplying organic matter to the microbial loop, as well as to higher trophic levels (Falkowski & Raven 2007). Marine phytoplankton have been the subject of observation for hundreds of years, even finding their way into the diaries of Charles Darwin (Darwin 1988) who remarked that plankton net tows offered him "hours of amusement & work." Marine phytoplankton are comprised of diverse taxonomic groups that resulted from a primary endosymbiotic event to form the red, green and glaucophyte lineages, and a secondary endosymbiosis that formed the cryptophytes, haptophytes and the stramenopiles (Yoon et al. 2004). Each group evolved differing metabolic capabilities and survival strategies to thrive in specialized environmental niches. The diatoms (found within the stramenopile lineage) are arguably the most globally important marine phytoplankton group because of their critical role in sequestering carbon, in structuring/sustaining productive food webs and in linking the carbon and nitrogen cycles (Armbrust 2009).

Diatoms drift through the oceans in microscopic glass houses, composed primarily of silica, that give them awe-inspiring visual appeal (Armbrust 2009). They constitute 20% of global primary production via carbon fixation, comparable to all the world's rainforests combined (Field et al. 1998; Nelson et al. 1995). Seasonal diatom blooms typically occur in coastal environments after an upwelling event delivers nitrate to surface waters (Wilkerson & Dugdale 1987). Following bloom formation, diatom organic matter may become entrained in the microbial loop through viral infection, autocatalytic cell death, or grazing (Bidle & Falkowski

2004). Diatom blooms also serve as nutrition for higher trophic levels, including zooplankton; following consumption, diatoms may sink as cell aggregates or in zooplankton fecal pellets, fueling the biological pump, which mediates organic carbon export below ocean surface waters (Smetacek 1999; Longhurst & Harrison 1989). Diatom frustules make them heavier than most other phytoplankton, and consequently, diatoms are associated with the highest rates of particle export; when diatom cells sink from the upper water column and become buried through sedimentation, they can sequester carbon from the atmosphere for hundreds to millions of years (Armbrust 2009). For these reasons, diatoms directly connect atmospheric carbon dioxide concentrations to the global carbon cycle (Smetacek 1999). With ongoing planetary change there comes a degree of urgency to determining the factors that control diatom growth and their presence, abundance, and distribution; the effects of a warming planet and an acidifying ocean on the proliferation of these important organisms remain theoretical.

Diatoms evolved ~ 200 million years ago (Sims et al. 2006). Their presence in the ocean along expanding continental margins contributed to an increase in organic matter export, increased planetary oxidation, and permanently altered biogeochemical cycles, including macro- and micro-nutrient concentrations (Falkowski et al. 2004; 2005). As diatoms increased in presence and abundance, and substantially impacted the marine environment, the marine environment, in turn, impacted the evolution of the diatoms. Diatoms were able to outcompete other phytoplankton groups through the evolution of finely tuned metabolic features that allowed them to thrive in turbulent environments (Katz et al. 2004). The success of diatoms in present-day oceans is a result of their ability to respond and acclimate to environmental changes, particularly in response to ephemeral nitrogen inputs (Smetacek 2001; Koester et al. 2012).

Diatoms possess differing nitrogen physiologies; growth rates vary across species and may differ by up to an order of magnitude between closely related diatoms (Auro & Cochlan 2013; Antia et al. 1991). Generally, when nitrate is supplied, diatoms exhibit a multi-phasic response to nitrate uptake that includes saturating uptake kinetics at low nitrogen concentrations and linear uptake at high nitrogen concentrations (Collos et al. 2005; Lomas & Glibert 2000). High-affinity transporters have low saturation constants and are most efficient when nitrogen concentrations are low (Miller et al. 2007; Song & Ward 2007). Low-affinity transporters may be used for luxury uptake in instances when nitrogen sources are abundant. In addition to nutritional nitrogen uptake, diatoms are hypothesized to exhibit luxury nitrogen uptake and store it as a way to take advantage of transitory and localized nitrate sources (Lomas & Glibert 1999b; 1999a; Tozzi et al. 2004). Nitrogen is stored in large vacuoles, which may constitute up to 40% of the cell volume and allow the cell to divide multiple times before needing to replenish its nutrients (Falkowski et al. 2004; Raven 1987). Internal nitrogen concentrations can span several orders of magnitude and provide an additional source of nitrogen when the external environment becomes nitrogen-limited (Lomas & Glibert 2000).

While diatoms are linked to the marine nitrogen cycle through new and recycled production (e.g. nitrate and ammonium uptake) and account for all new nitrate-based production in some oceanic regimes (e.g. eastern equatorial Pacific Ocean; Dugdale & Wilkerson 1998), they have also developed less conventional metabolic strategies for nitrogen utilization that connect them to other facets of the nitrogen cycle. Diatoms may undergo dissimilatory nitrate reduction (DNRA), an anaerobic process in which nitrate is converted to ammonium via a nitrite intermediate (Kamp et al. 2011; Peters & Thomas 1996; Burgin & Hamilton 2007). This pathway is hypothesized to allow benthic diatoms to form a dormant stage in dark anoxic

sediments until oxygen and nitrate are resupplied to the system (Kamp et al. 2011). The presence of this pathway in diatoms is surprising given that diatoms are traditionally associated with growth in the nitrate-replete oxic waters of the euphotic zone. Dissimilatory nitrate reduction in diatoms makes them a source of reduced nitrogen (ammonium) for other organisms and, in anoxic sections of the water column, could link the diatoms to the annamox pathway through ammonium (Kamp et al. 2011; Mulder et al. 1995). Diatoms also produce nitric oxide, a cell-to-cell signaling molecule activated by cellular stress (Vardi et al. 2006). In the field, nitric oxide is a by-product of both bacterial nitrification and denitrification, but the demise of diatom blooms may also produce a significant source of this free radical in the upper water column as diatom cells respond to (e.g.) decreased nutrient availability or predation, releasing this infochemical into surrounding waters. Nitric oxide production by diatoms would likely increase surface water concentrations of nitric oxide and its rate of exchange with the atmosphere (Vardi et al. 2008; Ward 2003; Zafiriou et al. 1980). Finally, in oligotrophic waters, diatom symbioses with nitrogen-fixing cyanobacteria (diazotrophs) significantly impact surface water biogeochemistry and are hypothesized to benefit both organisms: the symbiont supplies its host with fixed nitrogen, and the diatom supplies its symbiont with fixed carbon (Foster & Zehr 2006). Foster et al. (2011) demonstrated that diatoms are able to increase nitrogen fixation rates in their diazotrophic symbionts by more than 400%, a remarkable feat with direct implications for current estimates of nitrogen fixation. The above nitrogen-intensive processes link the diatoms to the marine nitrogen cycle, yet the underlying metabolic pathways associated with these processes are largely unknown.

Over the past several decades, various methodologies have been applied to study diatom nitrogen metabolism, including the use of stable-isotope ^{15}N uptake measurements, intracellular

nutrient ratios, enzyme assays and photosynthetic parameters. ^{15}N uptake measurements have served as a proxy for the nitrogen uptake potential of diatom assemblages, as well as for determining how much nitrogen is available for diatom growth (Dugdale & Goering 1967; MacIsaac & Dugdale 1972). Similarly, bulk chlorophyll analysis, inorganic nutrient measurements and particulate organic carbon/nitrogen ratios have provided snapshots of the diatom biomass across different nutrient concentrations (Dugdale et al. 1990). The presence of a given form of nitrogen in the environment does not mean that it can be used for diatom growth, for not all diatoms can grow on the same nitrogen substrate (Collos & Slawyk 1976). Similarly, uptake of available nitrogen does not always directly translate into nitrogen assimilation, as demonstrated by Collos (1980), who showed that some diatom species exhibit uncoupled nitrogen uptake and nitrogen assimilation through luxury nitrogen uptake. Generalizations on diatom metabolism enable modelers to predict *when* diatoms will bloom, whereas information on species differences in metabolism may enable oceanographers to predict *which* diatom species will bloom.

The use of molecular techniques has provided new insight into the nitrogen metabolism of these important organisms and the metabolic similarities and differences among diatom groups. The sequencing of the first diatom genome, from *Thalassiosira pseudonana*, confirmed the presence of integral metabolic pathways including photosynthesis and carbon fixation, but also revealed the presence of unsuspected pathways such as a complete urea cycle (Armbrust et al. 2004). The urea cycle is hypothesized to play a critical role in how the cell utilizes recycled nitrogen (Allen et al. 2011); it is also influenced by changes in light levels, which link the urea cycle to cellular energy management (Chapter 1, Bender et al. 2012). Subsequent sequencing projects, including the *Phaeodactylum tricornutum*, *Fragilariopsis cylindrus*, *Pseudo-nitzschia*

multiseries, and *Thalassiosira oceanica* genomes, have revealed the unique nature of diatom metabolism with the discovery of apparent diatom-specific genes, genes thus far found only in these datasets (Bowler et al. 2008; Lommer et al. 2012). Additionally, the presence of a majority of genes that encode for proteins of unknown function implies that many of the basic aspects of diatom metabolism are not currently modeled, and that critical components of diatom metabolism may be overlooked in the field because we cannot assign function to expressed genes when sampling natural diatom assemblages (Maheswari et al. 2009; Tirichine & Bowler 2011). Large-scale efforts to generate transcript sequences for a diverse range of diatoms (Marine Microbiology Initiative-supported Marine Microbial Eukaryote Transcriptome Sequencing Project) provide the opportunity to further explore the diversity of this group in the laboratory and in the field.

Presently, more than 100,000 diatom species are estimated to exist on Earth (Kooistra et al. 2007), categorized in four primary groups – the radial and bi/multipolar centric diatoms and the more recently diverged raphid and araphid pennate diatoms. Comparisons across the two, bipolar centric diatom genomes (*T. pseudonana* and *T. oceanica*), and the three, raphid pennate diatom genomes (*P. tricornutum*, *F. cylindrus*, and *P. multiseries*) indicate that the diatoms possess both a shared- and species-specific genetic repertoire (Bowler et al. 2008; Kroth et al. 2008; Smith et al. 2012; Lommer et al. 2012). Generally, metabolic pathways connected to nitrogen metabolism are shared across the groups, but differences in gene copy number and in the localization of proteins involved in these key pathways vary widely across species (Chapter 1, Bender et al. 2012; Allen et al. 2011). Additionally, horizontal gene transfer events, non-synonymous mutations, and the prevalence of numerous transposable elements in each genome are hypothesized to contribute to genotypic and phenotypic diversity (Bowler et al. 2008;

Koester et al. 2012). Given this genomic diversity, understanding the metabolic potential of the entire diatom community requires more knowledge of the metabolic capabilities of its individual members. Transcriptional tools enable deeper interrogation of diatom metabolism; changes in gene transcripts provide early indicators for changes in cell physiology.

This thesis seeks to understand how nitrogen controls diatom metabolism in the laboratory and in the field by testing the hypothesis that *diatoms exhibit a shared, core metabolic response to changes in nitrogen source and availability*. The results have implications for understanding how diatoms take up varied nitrogen sources in their environment, and how nitrogen source and availability affects cellular metabolism across diverse diatom species. In Chapter 1 (Bender et al. 2012), the role of the urea cycle in diatom nitrogen metabolism is investigated to test the null hypotheses that (1) *urea cycle transcript abundances are not affected by changes in diatom culturing conditions*, and (2) *the urea cycle is not connected to diatom nitrogen metabolism*. This work incorporates the urea cycle into the bigger picture of nitrogen metabolism in diatoms, through glutamine consumption and urea production, and demonstrates the effect of nitrogen source and light levels on this important metabolic pathway. Chapter 2 broadens our understanding of diatom nitrogen metabolism by examining the transcriptional responses of three diatoms (*Thalassiosira pseudonana*, *Fragilariopsis cylindrus*, and *Pseudo-nitzschia multiseries*) to the onset of nitrate starvation to test the hypothesis that (1) *diatoms exhibit similar transcript abundance patterns for critical, conserved nitrogen metabolism pathways in response to the onset of nitrate starvation*. Findings highlight the diverse metabolic capabilities of *T. pseudonana*, *F. cylindrus* and *P. multiseries* and demonstrate the importance of considering specific-specific metabolic responses to better understand and ultimately predict which diatoms will bloom in the marine environment. Chapter 3 assesses the nitrogen status of a

natural diatom assemblage using metatranscriptomics and tests the hypothesis that (1) *transcript abundances from a diatom community exhibit uniform expression patterns as the community transitions from a coastal nutrient-replete environment into high nitrate low chlorophyll waters.* Findings from Chapter 3 highlight the transcriptional heterogeneity among closely related diatoms to changes in both nitrogen and iron availability. While all diatoms in an assemblage are growing in the same bulk chemical environment, each species exhibits a different metabolic response because of its unique physiology. Results from this thesis provide a better of understanding of the fundamental components of diatom metabolism through laboratory experiments and provide a snapshot of diatom metabolism in the field. This work can be applied towards efforts to model diatom community dynamics by distinguishing the fraction of the community response that is shared across all diatom taxonomic groups from the fraction that is driven by one or two dominant diatom species. Accurately modeling and predicting when diatoms will bloom and, more importantly, which diatom species will bloom under a given set of environmental conditions, are critical to understanding the impact of a bloom on the marine environment, for diatoms differ markedly in morphology, nutrient uptake rates, and sinking rates (Durkin et al. 2013), all attributes that significantly impact carbon and nitrogen flux in the marine environment.

References for Introduction

Allen AE, Dupont CL, Obornik M, Horak A, Nunes-Nesi A, McCrow JP, et al. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature 473:203–209.

Antia N, Harrison PJ, Oliveira L. (1991). The role of dissolved organic nitrogen in phytoplankton nutrition, cell biology and ecology. Phycologia 30:1–89.

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

Armbrust EV, Berges JAJ, Bowler CC, Green BRB, Martinez DD, Putnam NH, et al. (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306:79–86.

Auro ME, Cochlan WP. (2013). Nitrogen utilization and toxin production by two diatoms of the *Pseudo-nitzschia pseudodelicatissima* complex: *P. cuspidata* and *P. fryxelliana*. J Phycol 49:156–169.

Bender SJ, Parker MS, Armbrust EV. (2012). Coupled effects of light and nitrogen source on the urea cycle and nitrogen metabolism over a diel cycle in the marine diatom *Thalassiosira pseudonana*. Protist 163:232–251.

Bidle KD, Falkowski PG. (2004). Cell death in planktonic, photosynthetic microorganisms. Nat Rev Microbiol 2:643–655.

Bowler CC, Allen AEA, Badger JHJ, Grimwood JJ, Jabbari KK, Kuo AA, et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456:239–244.

Burgin AJ, Hamilton SK. (2007). Have we overemphasized the role of denitrification in aquatic ecosystems? A review of nitrate removal pathways. Frontiers Ecol Environ 5:89–96.

Collos Y. (1980). Transient situations in nitrate assimilation by marine diatoms .1. changes in uptake parameters during nitrogen starvation. Limnol Oceanogr 25:1075–1081.

Collos Y, Slawyk G. (1976). Significance of cellular nitrate content in natural-populations of marine phytoplankton growing in shipboard cultures. Mar Biol 34:27–32.

Collos Y, Vaquer A, Souchu P. (2005). Acclimation of nitrate uptake by phytoplankton to high substrate levels. J Phycol 41:466–478.

Darwin C. (1988). Charles Darwin's Beagle Diary (1831-1836). Cambridge University Press.

Dugdale R, Goering J. (1967). Uptake of new and regenerated forms of nitrogen primary productivity. Limnol Oceanogr 12:196–206.

Dugdale R, Wilkerson F. (1998). Silicate regulation of new production in the equatorial Pacific

upwelling. Nature 391:270–273.

Dugdale R, Wilkerson F, Morel A. (1990). Realization of new production in coastal upwelling areas: a means to compare relative performance. Limnol Oceanogr 45:822–829.

Durkin CA, Bender SJ, Chan KYK, Gaessner K, Grunbaum D, Armbrust EV. (2013). Silicic acid supplied to coastal diatom communities influences cellular silicification and the potential export of carbon. Limnol Oceanogr 58(5):1707–26.

Falkowski PG. (1994). The role of phytoplankton photosynthesis in global biogeochemical cycles. Photosynth Res 39:235–258.

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield OM, et al. (2004). The evolution of modern eukaryotic phytoplankton. Science 305:354–360.

Falkowski PG, Katz ME, Milligan AJ, Fennel K, Cramer BS, Aubry MP, et al. (2005). The rise of oxygen over the past 205 million years and the evolution of large placental mammals. Science 309:2202–2204.

Falkowski PG, Raven JA. (2007). Aquatic Photosynthesis. 2nd ed. Princeton Univ Press.

Field C, Behrenfeld MJ, Randerson J, Falkowski PG. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. Science 281:237–240.

Foster RA, Kuypers MMM, Vagner T, Paerl RW, Musat N, Zehr JP. (2011). Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses. ISME J 5:1484–1493.

Foster RA, Zehr JP. (2006). Characterization of diatom–cyanobacteria symbioses on the basis of nifH, hetR and 16S rRNA sequences. Environ Microbiol 8:1913–1925.

Kamp A, de Beer D, Nitsch JL, Lavik G, Stief P. (2011). Diatoms respire nitrate to survive dark and anoxic conditions. Proc Natl Acad Sci 108:5649–5654.

Katz ME, Finkel ZV, Grzebyk D, Knoll AH, Falkowski PG. (2004). Evolutionary trajectories and biogeochemical impacts of marine eukaryotic phytoplankton. Annu Rev Ecol Syst 35:523–556.

Koester JA, Swanson WJ, Armbrust EV. (2012). Positive selection within a diatom species acts on putative protein interactions and transcriptional regulation. Mol Biol Evol 30:422–434.

Kooistra W, Gersonde R, Medlin LK, Mann DG. (2007). The origin and evolution of the diatoms: their adaptation to a planktonic existence. In:Evolution of Primary Producers in the Sea, Falkowski, PG & Knoll, AH, eds. Elsevier Academic Press: Amsterdam, pp. 210–225.

Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, et al. (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. PLOS One 3:e1426.

Lomas M, Glibert PM. (2000). Comparisons of nitrate uptake, storage, and reduction in marine diatoms and flagellates. J Phycol 36:903–913.

Lomas M, Glibert PM. (1999a). Interactions between NH_4^+ and NO_3^- uptake and assimilation: comparison of diatoms and dinoflagellates at several growth temperatures. Mar Biol 133:541–551.

Lomas M, Glibert PM. (1999b). Temperature regulation of nitrate uptake: A novel hypothesis about nitrate uptake and reduction in cool-water diatoms. Limnol Oceanogr 44:556–572.

Lommer M, Specht M, Roy A-S, Kraemer L, Andreson R, Gutowska MA, et al. (2012). Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. Genome Biol 13:R66.

Longhurst AR, Harrison WG. (1989). The biological pump - profiles of plankton production and consumption in the upper ocean. Progr Oceanogr 22:47–123.

MacIsaac J, Dugdale R. (1972). Interactions of light and inorganic nitrogen in controlling nitrogen uptake in sea. Deep-Sea Res Pt I 19:209–232.

Maheswari U, Mock T, Armbrust EV, Bowler C. (2009). Update of the Diatom EST Database: a new tool for digital transcriptomics. Nucleic Acids Res 37:D1001–D1005.

Miller AJ, Fan X, Orsel M, Smith SJ, Wells DM. (2007). Nitrate transport and signaling. J Exp Biol 58:2297–2306.

Mulder A, Graaf AA, Robertson LA, Kuenen JG. (1995). Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. FEMS Microbiol Ecol 16:177–184.

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

Peters E, Thomas DN. (1996). Prolonged nitrate exhaustion and diatom mortality: a comparison of polar and temperate *Thalassiosira* species. J Plankton Res 18:953–968.

Raven JA. (1987). The role of vacuoles. New Phytol 106:357–422.

Sims PA, Mann DG, Medlin LK. (2006). Evolution of the diatoms: insights from fossil, biological and molecular data. Phycologia 45:361–402.

Smetacek V. (2001). A watery arms race. Nature 411:745.

Smetacek V. (1999). Diatoms and the ocean carbon cycle. Protist 150:25–32.

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

Song B, Ward BB. (2007). Molecular cloning and characterization of high-affinity nitrate transporters in marine phytoplankton. J Phycol 43:542–552.

Tirichine L, Bowler C. (2011). Decoding algal genomes: tracing back the history of photosynthetic life on Earth. Plant J 66:45–67.

Tozzi S, Schofield OM, Falkowski PG. (2004). Historical climate change and ocean turbulence as selective agents for two key phytoplankton functional groups. Mar Ecol Prog Ser 274:123–132.

Vardi A, Bidle KD, Kwityn C, Hirsh DJ, Thompson SM, Callow JA, et al. (2008). A diatom gene regulating nitric-oxide signaling and susceptibility to diatom-derived aldehydes. Curr Biol 18:895–899.

Vardi AA, Formiggini FF, Casotti RR, De Martino AA, Ribalet FF, Miralto AA, et al. (2006). A stress surveillance system based on calcium and nitric oxide in marine diatoms. PLOS Biol 4:e60.

Ward BB. (2003). Significance of anaerobic ammonium oxidation in the ocean. Trends Microbiol 11:408–410.

Wilkerson F, Dugdale R. (1987). The use of large shipboard barrels and drifters to study the effects of coastal upwelling on phytoplankton dynamics. Limnol Oceanogr 32:368–382.

Yoon H, Hackett J, Ciniglia C, Pinto G, Bhattacharya D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol 21:809–818.

Zafiriou OCO, McFarland MM, Bromund RHR. (1980). Nitric oxide in seawater. Science 207:637–639.

Chapter 1

Coupled effects of light and nitrogen source on the urea cycle and nitrogen metabolism over a diel cycle in the marine diatom Thalassiosira pseudonana

Bender SJ, Parker MSP, and Armbrust EV. (2012). Protist. 163: 232 – 51.

1.1 Abstract

Diatoms are photoautotrophic organisms capable of growing on a variety of inorganic and organic nitrogen sources. Discovery of a complete urea cycle in diatoms was surprising, as this pathway commonly functions in heterotrophic organisms to rid cells of waste nitrogen. To determine how the urea cycle is integrated into cellular nitrogen metabolism and energy management, the centric diatom *Thalassiosira pseudonana* was maintained in semi-continuous batch cultures on nitrate, ammonium, or urea as the sole nitrogen source, under a 16:8 light:dark cycle and at light intensities that were low, saturating, or high for growth. Steady-state transcript levels were determined for genes encoding enzymes linked to the urea cycle, urea hydrolysis, glutamine synthesis, pyrimidine synthesis, photorespiration, and energy storage. Transcript abundances were significantly affected by nitrogen source, light intensity and a diel cycle. The impact of N source on differential transcript accumulation was most apparent under the highest light intensity. Models of cellular metabolism under high light were developed based on changes in transcript abundance and predicted enzyme localizations. We hypothesize that the urea cycle is integrated into nitrogen metabolism through its connection to glutamine and in the eventual production of urea. These findings have important implications for nitrogen flow in the cell over diel cycles at surface ocean irradiances.

1.2 Key Words

Carbamoyl phosphate synthetase; glutamine synthetase; nitrogen metabolism; photorespiration; urea cycle; urease.

1.3 Introduction

Diatoms are a key group of marine phytoplankton that are dominant in nutrient-rich coastal regions, and contribute an estimated 40% of total oceanic primary productivity (Nelson et al. 1995). Diatom blooms commonly occur in regions where nitrogen (N) source is variable and they possess a suite of N-related transporters and enzymes (Allen 2005; Armbrust et al. 2004; Hildebrand 2005; Hildebrand & Dahlin 2000) and utilize a variety of inorganic (e.g. nitrate, NO_3^-; ammonium, NH_4^+) and organic (e.g. urea; amino acids) N sources for growth. Diatoms exhibit their fastest growth rates on reduced forms of N such as NH_4^+ or urea (Dortch 1990; Dortch et al. 1991; Peers et al. 2000; Syrett 1981), in part due to the low energetic costs associated with assimilation of these forms (Hildebrand 2005).

Analysis of whole genome sequences of marine diatoms is providing new insights into mechanisms underlying the biogeochemical roles of these organisms. One of the more surprising outcomes was identification of genes required for a complete urea cycle i*n Thalassiosira pseudonan*a (Armbrust et al. 2004). This pathway was subsequently identified in the genomes of the diatom*s Phaeodactylum tricornutu*m (Bowler et al. 2008) an*d Fragilariopsis cylindru*s (http://genome.jgi- psf.org/Fracy1/Fracy1.home.html), and there is now evidence for a complete cycle in additional members of the chromaveolates (e.g. *Emiliania huxley*ii; http://genome.jgi-psf.org/Emihu1/Emihu1.home.html) (Allen et al. 2011). Recent studies suggest that plants also

possess the enzymes necessary for a complete urea cycle; however, localization of key enzymes differs from that of heterotrophs and diatoms (Gaufichon et al. 2010; Taylor et al. 2010).

In heterotrophs, the initial and rate-limiting step in the urea cycle occurs within mitochondria. Carbamoyl phosphate synthetase (CPS) catalyzes a 2-step ligation of 2 molecules of ATP, bicarbonate and NH_4^+ to form carbamoyl phosphate (P) (Beevers & Storey 1976; Tatibana & Shigesad 1972). One of two isoforms of CPS (CPSI or CPSIII) is used in the heterotrophic urea cycle, depending on the organism. A third isoform, CPSII, is involved in pyrimidine synthesis in the cytosol and utilizes glutamine. In higher plants, a plastid-localized CPSII serves a dual role in pyridimine synthesis and the production of carbamoyl-P (Slocum 2005). CPSIII and CPSII use the amide group from glutamine as the primary N donor, whereas CPSI requires NH_4^+ as its primary substrate (Holden et al. 1998; Hong et al. 1994). In diatoms, two novel forms of CPS have been identified: unCPS utilizes NH_4^+ in the mitochondria as part of the urea cycle and pgCPS2 utilizes glutamine in the cytosol (Allen et al. 2011).

Detection of the urea cycle, including a mitochondria-targeted unCPS, in diatoms was unexpected as this pathway commonly functions in heterotrophic organisms to rid cells of waste NH_4^+, which diatoms can use as a sole N source for growth. Several hypotheses may explain the functional role of the urea cycle in diatoms, including temporary energy storage through formation of creatine-P, recycling of N through the production of amino acids, including arginine required for polyamine synthesis, the formation of a urea by-product, or as a sink for the photorespiration-generated NH_4^+ (Allen et al. 2006; 2011; Armbrust et al. 2004; Bowler et al. 2008; Vardi et al. 2008). Detection of all urea cycle enzymes in *T. pseudonana* under a variety of laboratory conditions (Nunn et al. 2009), as well as the presence of many diatom urea cycle transcripts in a field metatranscriptome, suggest that this pathway plays a central role in diatom

metabolism (Marchetti et al. 2012). Furthermore, changes in urea cycle intermediates in *P. tricornutum* in response to N availability have implicated the urea cycle in C and N redistribution via the tricarboxylic acid cycle (TCA) and the glutamine synthetase/glutamate synthase (GS-GOGAT) pathway (Allen et al. 2011).

The N required for entry into the urea cycle may be generated from NH_4^+ or glutamine via the GS-GOGAT pathway (Zehr & Falkowski 1988). In *T. pseudonana*, three isoforms of GS have been identified: GSI, GSII and GSIII (Armbrust et al. 2004; Takabayashi et al. 2005). Transcriptional data is currently available for the genes encoding GSII and GSIII (Brown et al. 2009; Parker & Armbrust 2005; Takabayashi et al. 2005). GSII acts within the plastid where it utilizes the NH_4^+ derived from reduction of NO_3^- (Brown et al. 2009). GSIII was hypothesized to function in the cytosol to assimilate the NH_4^+ taken up directly by the cells in *T. pseudonana* (Brown et al. 2009), although direct evidence for this localization is lacking. A combination of *in silico* analyses and confocal microscopy using GFP labeling with *P. tricornutum* suggests that GSIII is targeted to the mitochondria in diatoms (Siaut et al. 2007).

To date, several studies have provided the framework for utilizing transcript abundances as a tool for elucidating the complex interactions between N metabolism, N assimilation and changing irradiance in diatoms (Granum et al. 2009; Hildebrand 2005; Hildebrand & Dahlin 2000; Kang et al. 2009; Kroth et al. 2008; Mock et al. 2008; Parker & Armbrust 2005; Parker et al. 2004). Previous work has identified transcriptional regulation in a variety of organisms for the genes examined in this study: from bacteria to complex metazoans e.g. *CPSIII* (Pierard et al. 1980); *CPSII* (Denis-Duphil 1989); *URE* (urease) reviewed in (Mobley et al. 1995); *GSII* (Takabayashi et al. 2005); *GDCT* (Parker et al. 2004; Parker & Armbrust 2005); and *CK* (creatine kinase; Jaynes et al. 1986). Post-transcriptional modification may also affect

downstream processes (e.g. Poulsen et al. 2006); however, changes in transcript abundance provide a first snapshot of the cell's response to its environment.

Our experimental design incorporated multi-way ANOVAs to explicitly test interactions between N source and light on cell physiology and gene expression. *In silico* analysis and quantitative reverse transcriptase PCR were used to identify connections between the urea cycle and other pathways integral to cell metabolism. A model of N flow in *T. pseudonana* was developed from this data that includes protein localization and suggests that the urea cycle plays a critical role in both N metabolism and energy balance in the cell.

1.4 Methods

Culture conditions

Axenic cultures of *Thalassiosira pseudonana* (Hustedt) Hasle et Heimdal (Provasoli-Guillard National Center for Culture of Marine Phytoplankton, CCMP 1335) were maintained in semi-continuous batch cultures (Brand et al. 1981) on a 16:8 light:dark cycle at 13^{o}C in artificial seawater amended with f/2 concentrations of silicate, phosphate, vitamins and trace metals (Berges et al. 2001). Cultures were acclimated to growth at three light intensities: 50 μmol photons m^{-2} s^{-1} (LL), 190 μmol photons m^{-2} s^{-1} (SL), or 400 μmol photons m^{-2} s^{-1} (HL) and three N sources: 882 μM NO_3^- or NH_4^+ or 441 μM urea [$(NH_2)_2CO$].

Growth rates were determined by daily monitoring of chlorophyll (chl) *a* fluorescence using a 10-AU fluorometer (Turner Designs, Sunnyvale, CA, USA). Cultures were considered acclimated when the slopes of three consecutive growth curves were not significantly different from one another (ANCOVA; Zar 1999). Growth rates were calculated based on the three acclimated growth curves. The R software (2.10.1 GUI 1.31; http://www.R-project.org) was used

to determine significant differences across growth rates by running a two-way ANOVA (p = 0.05). To ensure cultures were axenic, they were tested for the presence of bacteria prior to the growth experiments by transferring an aliquot to TM media (Provasoli-Guillard recipe) and by viewing Sybr Green (Invitrogen, Carlsbad, CA, USA)- stained cells on a Nikon i80 microscope.

Triplicate cultures maintained in exponential growth in 25 mL media were combined before transfer to three 2 L containers of fresh media. For the HL treatments, triplicates were pooled, inoculated into a 1 L bottle of fresh media, and then transferred to the three 2 L containers of f/2. A 10 mL aliquot of culture was removed from the 2 L experimental bottles for measurements of chl *a* fluorescence and photochemical yield of photosystem II (F_v/F_m) with a Phyto-PAM Phytoplankton Analyzer (Walz, Effeltrich, Germany). A two-way ANOVA (p = 0.05) was used to test for differences in F_v/F_m values. Triplicate 2 L cultures were maintained at each light intensity until mid-exponential phase. One liter was collected from each culture during exponential phase at 3 h into the light and again at 3 h into the dark, and cells from both time points were gently vacuum filtered onto 0.4 μm HTTP filters (Millipore Corp., Billerica, MA, USA). Filters were flash frozen in liquid N and stored at −80^{o}C prior to RNA isolation (see below). At the time of cell harvesting, approximately 50mL of sample were syringe-filtered (0.2 μm) into Falcon tubes, and stored at −20^{o}C for nutrient analysis. Nutrients were analysed at the University of Washington Marine Chemistry Laboratory on a Technicon AutoAnalyser II following the methods of UNESCO (1994). Urea concentrations were determined using the diacetyl monoxime method (Price & Harrison 1987). Cultures were monitored for chl *a* fluorescence several days post-harvest to ensure that the culture was in mid-exponential when filtered.

RNA extractions

RNA was extracted from frozen cells on filters using the RNAqueous-4PCR kit (Ambion, Austin, TX, USA). RNA samples were incubated with 4 units of RNAse-free *DNAse*I for 1 h at 37^{o}C before inactivation with DNase Inactivation Reagent (RNAqueous-4PCR, Ambion). DNA contamination was tested using *Actin* primers in a quantitative PCR (Bidle and Bender 2008). Additional *DNAse*I treatments were performed if necessary to ensure that the RNA was DNA-free. RNA samples were quantified using a NanoDrop Spectrophotometer (ND-1000; Thermo Fisher Scientific, Wilmington, DE, USA). Fifty-two ng of purified RNA were reverse-transcribed in triplicate cDNA reactions using the Superscript III First-Strand Synthesis Reaction System for RT-PCR with oligo $(dT)_{20}$ primers (Invitrogen) and diluted 4-fold to a final volume of 84 μL.

Gene identification and targeting information

The *T. pseudonana* gene models (v3.0; http://genome.jgi-psf.org/Thaps3/Thaps3.home.html) were used to identify the eight genes encoding: carbamoyl phosphate synthetase (unCPS, pgCPSII), glutamine synthetase (GSI, GSII, GSIII), urease (URE), glycine decarboxylase T-protein (GDCT), and creatine kinase (CK). Predicted *T. pseudonana* proteins were aligned to *P. tricornutum* full-length proteins and to sequences in GenBank using the Geneious Pro 5.0.3 software (2010 Biomatters Ltd.; http://www.geneious.com) to determine whether the predicted *T. pseudonana* proteins were full-length. Several gene models required additional extension: the genes encoding the predicted proteins unCPS (40323), GSIII (270138), URE (30193) and CK (263946), as well as numerous N-based transporters and enzymes. These genes were extended on the 5′ end to the first methionine downstream of a stop codon. The intron location on the 5′ end of GSIII was detected by alignment to *Chaetoceros compressus*

GSIII (BAG39455.1) and from available expressed sequence tag libraries (ESTs; jgi.doe.gov) and RNAseq data from a SOLiD: Next Generation Sequencing (Applied Biosystems, Carlsbad, CA) of *T. pseudonana* transcripts. The gene models encoding CK (263946), as well as UREG (accessory protein to URE, 263354), nitrite reductase (262125), a urea transporter (264305), and three NH_4^+ transporters (14096, 36263, 13996), also required extension on the 3′ end to the first in-frame stop codon. An AMT upstream of 268226, not previously identified by JGI gene models, was identified by tBLASTn. For two NH_4^+ transporters (268226 and 13996), intron locations were modified based on alignment to full-length genes and cDNA of NH_4^+ transporters from the diatom *Cylindrotheca fusiformis* (Hildebrand 2005) and the location of canonical GT-AG intron splice junctions. The gene encoding the GSI protein (261531) was extended upstream of the original JGI model and a potential fusion transcript was detected consisting of GSI and a transcript encoding a putative cysteine hydrolase. Predicted full-length sequences for all genes are given in supplemental Table 1.2.

SignalP was used to predict signal peptides, ChloroP (v1.1) was used to predict plastid target sequences, and MitoProt II v1.101 was used to predict mitochondrial targeting (Bendtsen et al. 2004; Claros & Vincens 1996; Emanuelsson et al. 1999). The logo plots of Gruber (Gruber et al. 2007) were used to help identify possible variants of the canonical chloroplast targeted signal peptide cleavage site (ASA-FAP) found in diatoms and cryptophytes. If SignalP returned an HMM probability of a signal peptide of 0.9 or greater and positive neural networks scores, the predicted signal peptide was trimmed and submitted to ChloroP. If the SignalP HMM probability was less than 0.9, the full-length predicted protein sequence was submitted to MitoProt II, which returns a probability of mitochondrial targeting. When targeting could not be determined definitively (probabilities < 0.9) with SignalP, ChloroP or MitoProt, WoLF PSORT (Horton et

al. 2007) was used on the Animal setting. If a signal peptide was detected in WoLF PSORT, the signal peptide was trimmed and the analysis was run on the Plant setting. EST libraries and RNAseq runs were used to confirm transcription of the full-length gene sequences.

For tree-building full-length predicted protein sequences encoded by *GSI*, *GSII*, and *GSIII* were used to query Genbank with BLAST using blastp (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The top ten hits for each gene were chosen, in addition to select sequences from previously published studies representing known GS isoforms (Robertson & Tartar 2006), and any homologous sequences in the two other diatoms with whole genome sequence (*Phaeodactylum tricornutum* and *Fragilariopsis cylindrus*, www.jgi.doe.gov). Furthermore, each of the three predicted GS isofroms was used to search known crystal structures using the CPHmodels 3.0 Server (http://www.cbs.dtu.dk/services/CPHmodels/), which relies on both alignment profiles as well as secondary structure predictions. The crystal structures with the strongest similarity to each predicted GS and all the above-mentioned homologous sequences were aligned using Geneious with the MUSCLE alignment tool (Edgar 2004) on default settings. Major gaps and over-hanging ends were trimmed (including the large upstream regions of *GSI* and *fluG*, a homolog to *GSI*), and maximum likelihood phylogenetic trees were run with PhyML (packaged as a Geneious plugin) using the WAG substitution model and 100 bootstrapped replicates. Four substitution categories were used and the gamma distribution was estimated.

Quantitative PCR (qPCR)

PCR primers were designed for the genes encoding the following enzymes: unCPS, GSIII, pgCPSII, GSI, URE, and CK. Primer sequences for the genes encoding GSII and GDCT were from Parker and Armbrust (2005). Three housekeeping genes *ACT1* (Actin1, protein ID 25772), *ACT4* (Actin4, protein ID 269504), *40s s11* (protein ID 31084) or total RNA (totRNA)

were used to normalize gene expression. Primer sequences for *ACT1* and *ACT4* were from Bidle and Bender (2008) and Durkin et al. (2009). Amplicon lengths for qPCR varied between 100 bp and 345 bp and spanned the exon portion of *unCPS*, *GSI*, *GSIII*, *pgCPSII*, *URE* and *CK* (Table 1.1).

Standards for qPCR were made by amplifying genomic DNA with the above PCR primers. PCR product was electrophoresed through a 3% agarose gel to ensure that fragments of the proper size were amplified. PCR product was purified using the High Pure PCR Product Purification Kit (Roche Applied Science, Indianapolis, IN, USA), cloned into the pCR2.1 TOPO vector and transformed into TOP10-competent *Escherichia coli* cells. Plasmid DNA was extracted from PCR-positive clones using the QIAprep Miniprep Kit and quantified. One μg of circular plasmid DNA was linearized using 25 units of *Spe*I with the exception of *GSI*, which was linearized with *Xba*I (NewEngland Biolabs, Ipswich, MA, USA). Linearized plasmid DNA (lpDNA) was quantified and run on a 1% agarose gel to confirm linearization. Amplified lpDNA was sequenced at the High-Throughput Genomics Unit (Dept. of Genomics, University of Washington; http://www.htSEQ.org) and analysed with Sequencher v.4.6 (Gene Codes Corporation, Ann Arbor, MI, USA).

For qPCR, 15 μL of SYBR Green IQ Supermix (Bio-Rad, Hercules, CA, USA), 800nM primers (final concentration), 7.2 μL of water and 3 μL of template were combined in a final volume of 30 μL. The qPCR consisted of 95^{o}C for 3 min, followed by 45 cycles of 95^{o}C for 10 sec, 60^{o}C for 30 sec, 72^{o}C for 50 sec followed by 80 cycles starting at 55^{o}C (up to 90^{o}C) for 10 sec per cycle to determine the amplicon melting temperature. All qPCRs were run on an iCycler iQ Real-Time PCR Detection System (Bio-Rad). Three technical replicates were run for each gene. Seven standards were used and ranged from 10^7 to 10^1 copies L^{-1}; standards for each gene

were also run in triplicate. All gene sets included a no-template control. Dissociation curves were examined post-run to ensure a single fragment was amplified.

qPCR analyses were conducted using the iCycle iQ Real-Time Detection System Software v.3.0 (Bio-Rad). Transcript abundance for each unknown was determined based on standard curves. The LinRegPCR program was used to compare amplification efficiencies for all qPCRs (Ramakers et al. 2003). An F-test was run to test for variation within each plate and a t-test was used to test against significant differences among standards and unknowns ($p = 0.05$; Zar 1999). Differences in transcript abundance for a given gene across treatments were tested with a two-way ANOVA ($p = 0.05$; R package). A two-way ANOVA with repeated measures tested for significant differences between transcript abundances in the light and the dark.

Normalization of gene transcript abundance

Various means of normalizing transcript abundances across treatments were examined. Gene normalization differences were tested using a two-way ANOVA ($p = 0.05$) and a post-hoc Tukey's test for pair-wise comparisons. No significant differences ($p > 0.05$) in relative transcript abundances were detected when *unCPS*, *URE* or *GSII* were normalized to either totRNA or to the transcript abundance of *ACT1*, *ACT4* or *40s s11* for the HL cultures during the light phase; however, significant differences were observed in the dark phase, depending on the normalization gene used. In addition, transcript abundances for *ACT1* and *ACT4* varied significantly when different light conditions were tested ($p < 0.05$). Subsequently, transcript abundances for all genes were normalized to totRNA to avoid diel and light-dependent variability introduced by the housekeeping genes.

To test for the reproducibility of the reverse transcription reactions, pools of cDNA were generated from three separate reverse transcriptase reactions and transcript abundances were

determined for *URE*. No significant differences ($p > 0.0020$) were detected for the triplicate reverse-transcription reactions.

1.5 Results

Growth rate comparisons

Growth rates were significantly different based on the interaction of N source and light intensity (two-way ANOVA, $p < 0.05$; Table 1.1; supplemental Table 1.1). The 50 μmol photons m^{-2} s^{-1} irradiance was defined as low light (LL), the 190 μmol photons m^{-2} s^{-1} irradiance as saturating light (SL) and the 400 μmol photons m^{-2} s^{-1} as high light (HL) for growth. The F_v/F_m also varied significantly based on the interaction of N source and light intensity in both the light and the dark ($p < 0.05$; Table 1.1; supplemental Table 1.1). There were no significant differences between F_v/F_m measured in the light compared to F_v/F_m measured in the dark. For subsequent comparisons, the control treatment was defined as cells grown on NO_3^- as a sole N source under SL conditions during the light phase.

At time of harvest, trace levels of NH_4^+ and NO_3^- (< 0.5 μM), respectively, were detected in the media with either NO_3^- or NH_4^+ as the sole N source; trace urea (< 0.5 μM) was also in these media. The urea-only media contained trace levels (0.47−0.79 μM) of NO_3^- and elevated levels (160–243 μM) of NH_4^+. The NH_4^+ concentrations in all of the urea treatments were about 25% of the NH_4^+ concentrations detected in the NH_4^+-only treatments, and were elevated relative to previous urea-grown cultures (data not shown). Phosphate concentrations ranged from 16 to 32 μM and silicic acid concentrations ranged between 18 and 85 μM at time of harvest, confirming that growth of the cultures was not limited by nutrient availability.

Predicted localization of proteins involved in nitrogen metabolism

To better understand the flow of N through the metabolic pathways targeted by our proxy genes, a model linking the urea cycle to N assimilation in *T. pseudonana* was built using predicted localization of key enzymes and transporters (Figure 1.1; supplemental Table 1.2) based on genome analysis, re-evaluation of predicted gene models (http://genome.jgi-psf.org/Thaps3/ Thaps3.home.html), and determination of predicted gene structure using EST and RNAseq data.

Carbamoyl phosphate synthetase. The *T. pseudonana* genome is predicted to encode two *CPS*: protein IDs 40323 and 24248, which were both previously annotated as CPSIII (http:// genome.jgi-psf.org/Thaps3/Thaps3.home.html). Upstream extension of the gene model encoding protein ID 40323 increased the length of the open reading frame (ORF) and identified a predicted mitochondrial targeting sequence, consistent with the role of this CPS in the first mitochondrial-localized step of the urea cycle (Lindley et al. 2007). The gene encoding protein ID 24248 lacked targeting motifs suggesting it is localized in the cytosol, consistent with the role of this CPS in pyrimidine biosynthesis (supplemental Table 1.2). Protein 40323 and protein 24248 were designated as unCPS and pgCPSII, respectively, based on the phylogenetic analysis of Allen et al. (2011); unCPS utilizes NH_4^+ as its preferred substrate, whereas pgCPSII utilizes glutamine (Allen et al. 2011).

Glutamine synthetases. Three genes were detected that are predicted to encode GS (protein IDs 261531, 26051 and 270138), which converts NH_4^+ and glutamate to glutamine (supplemental Table 1.2). The gene model encoding protein ID 261531 was extended in-frame at the 5′-end resulting in an ORF of 2292 bp that appears to generate a fusion between a gene encoding GS (COG0174, e-value: 8.6×10^{-68}) and one encoding a cysteine hydrolase (cd00431, e-value: 1.4 x

10^{-25}), possibly isochorismatase (pfam 00857, e-value: 1.0 x 10^{-24}) (NCBI Conserved Domains Database (Marchler-Bauer et al. 2009). Transcription of the full-length sequence was confirmed by examining RNAseq-based transcriptome data (supplemental Figure 1.1A) and the region spanning the RNASeq coverage drop was successfully sequenced using cDNA (supplemental Figure 1.1B), confirming the presence of the full-length ORF in the mRNA pool. The gene model encoding protein ID 26051 (GSII) did not require extension, and SignalP identified strong support for a signal peptide. A potential transit peptide was detected by WoLF PSORT, and the first 15 amino acids following the cleavage site are serine/threonine rich, which is consistent with the presence of a plastid target peptide (Nassoury & Morse 2005) and corroborates previous annotations of this protein as plastid localized (Ghoshroy et al. 2010; Robertson & Tartar 2006; Robertson et al. 1999). The gene model encoding protein ID 270138 (GSIII) was extended in-frame at the 5′-end, and a predicted mitochondrial-targeted peptide sequence was detected, suggesting that this protein is localized to the mitochondria. A maximum likelihood tree displayed strong bootstrap support for the clustering of each *T. pseudonana* glutamine synthetase with other known GSI, GSII and GSIII sequences (supplemental Figure 1.2).

Three predicted proteins (protein IDs 264918, 262236, 13922) were identified in *T. pseudonana* that displayed homology (e-values < 1 x 10^{-27}) to *Arabidopsis* CAT8, a known glutamine transporter (Yang et al. 2010); however, plastid-targeting was not evident with our *in silico* analyses. We detected other examples of putative amino acid transporters in the *T. pseudonana* genome with clear plastid targeting (e.g. protein IDs 23142, 20889); however, the specificity of these transporters is unclear (data not shown).

Urea metabolism. A single urease (URE: protein ID 30193) that hydrolyzes urea into NH_4^+ and CO_2 is encoded by the *T. pseudonana* genome (supplemental Table 1.2). No evidence for

organelle targeting was detected for this enzyme. Two genes encoding predicted urease accessory proteins (UREG: protein ID 263354; URED/UREF fusion: protein ID 1586) were identified. The gene models were extended in frame at the 5′- and 3′-ends. The resulting predicted urease accessory proteins appear to be targeted to the mitochondria, suggesting that URE is also localized to the mitochondria. This targeting is further supported by evidence that one urea transporter (protein ID 23417) is targeted to the mitochondria (supplemental Table 1.2).

Connected metabolic pathways. Biosynthesis of creatine requires the precursor arginine, an amino acid generated solely via the urea cycle in animals. Phosphorylation of creatine by creatine kinase (CK) creates high-energy P-bonds in the form of creatine-P. Creatine-P can be used to phosphorylate ADP to ATP, thus serving as an energy buffer in the cell. The *CK* gene model (protein ID 263946) was extended in-frame at both the 5′ and 3′ ends (supplemental Table 1.2). Although support for a plastid target peptide is weak using ChloroP, the presence of a signal peptide, as well as WoLF PSORT plastid prediction support, suggests that the protein may be localized to the plastid. No genes encoding obvious creatine transporters were detected. The glycine decarboxylase T-subunit gene model (*GDCT*, protein ID 36208) did not require extension and this protein is predicted to localize to the mitochondria.

Additional N-based enzymes. Previous studies identified one NO_2^- transporter, three NO_3^- transporters, seven NH_4^+ transporters, and three urea transporters encoded by the *T. pseudonana* genome (Allen 2005; Armbrust et al. 2004). Predicted localization for these transporters, as well as NO_3^- and NO_2^- reductases and glutamate synthases are modeled in Figure 1.1 and listed in supplemental Table 1.2.

Impact of nitrogen and light on transcription of genes associated with the urea cycle and N assimilation

Quantitative PCR was used to compare transcript abundances for the genes encoding the following enzymes: unCPS, URE, GSIII, GSII, pgCPSII, GSI, GDCT, and CK (Table 1.2). Gene transcripts were normalized to total RNA (totRNA), rather than housekeeping genes because of a demonstrated impact of light on housekeeping transcript abundances. With few exceptions, transcript abundances were significantly affected by the interaction between N source and light, as well as the presence of a light:dark cycle ($p < 0.05$; Table 1.3; supplemental Table 1.1).

The enzyme unCPS catalyzes formation of carbamoyl-P in the mitochondria during the initial step of the urea cycle. Relative abundance of *unCPS* transcripts varied significantly with N source and light intensity ($p < 0.05$; Figure 1.2A; Table 1.3). Under HL conditions, transcript abundances for *unCPS* in the NO_3^- and urea treatments increased by 8-fold and 12-fold, respectively, in the light, relative to the NO_3^- SL control treatment, and by 12-fold and 8-fold, respectively, in the dark. Transcript abundances between the light and the dark phases were not significantly different ($p = 0.060$) across N sources, although there were small increases in transcript abundances in all treatments, except for the urea HL treatment, in the dark.

Urea is either taken up from the environment or generated as a by-product of the urea cycle. The *T. pseudonana* URE is hypothesized to hydrolyze the urea into CO_2 and NH_4^+ within the mitochondria. Transcript abundance for *URE* varied significantly based on light and N source ($p < 0.05$; Table 1.3). *URE* transcripts were detected regardless of N source suggesting that transcription of *URE* is not regulated solely by uptake of exogenous urea. Transcript abundances were significantly greater during the dark for all N sources ($p < 0.05$; Figure 1.2B). During both the light and dark, cells grown under HL irradiances exhibited the greatest transcript abundances.

There was a 4-fold increase in the NO_3^- HL treatment relative to the control, but transcript abundance did not change by more than ± 2-fold for the NH_4^+ or urea treatments.

Transcript abundance for *GSIII* was determined to test whether the N required by unCPS is linked to the activity of GSIII and generation of glutamine within the mitochondria. Transcript abundances differed significantly between the light and dark ($p < 0.05$; Table 1.3), and they were an order of magnitude lower during the light than during the dark, regardless of N source or light intensity (Figure 1.2C). In the dark, *GSIII* transcript abundances were affected by N source or light intensity, but not by the inter- action between the two ($p = 0.11$). The greatest increases in transcript abundances relative to the control occurred during the dark with 15- to 30-fold increases in the HL treatments.

Glutamine is also generated by the enzyme GSII localized to the plastid. Transcript abundances for *GSII* differed significantly between the light and dark ($p < 0.05$; Table 1.3), and in direct contrast to the *GSIII* patterns, *GSII* transcripts were an order of magnitude lower in the dark than the light, regardless of N source or light intensity (Figure 1.2D). Relative transcripts for *GSII* increased the most during the light when grown under NO_3^- HL or urea HL conditions relative to the control, a trend similar to that observed with *unCPS*. In the light, transcript abundances for all NH_4^+ treatments, as well as the urea LL and urea SL treatments, were equal to or up to 4-fold lower than those for the control.

When transcript abundances for *URE* were compared to *unCPS*, abundances were weakly correlated ($r^2 \leq 0.500$) in both the light ($r^2 = 0.410$) and in the dark ($r^2 = 0.504$), $p < 0.05$; Figure 1.3A). Stronger positive correlations were found between *unCPS* and the mitochondrial-localized *GSIII* during the dark ($r^2 = 0.515$, $p = 3.2 \times 10^{-10}$; Figure 1.3B) and between *unCPS* and the plastid-localized *GSII* during the light ($r^2 = 0.600$, $p = 2.2 \times 10^{-16}$; Figure 1.3C). Weak

correlations were obtained when *URE* transcripts were compared to *GSIII* in the light ($r^2 = 0.367$, $p = 4.2 \times 10^{-09}$) and the dark ($r^2 = 0.319$, $p = 1.9 \times 10^{-11}$), as well as *GSII* in the light ($r^2 = 0.437$, $p = 4.0 \times 10^{-08}$) and dark ($r^2 = 0.213$, $p = 1.5 \times 10^{-05}$) (data not shown).

T. pseudonana possesses a second CPS isoform (pgCPSII), localized to the cytosol. This enzyme catalyzes the formation of carbamoyl-P, which serves as the precursor of pyrimidine biosynthesis. *pgCPSII* transcript abundances varied significantly based on N source and light intensity in the light and dark ($p < 0.05$; Figure 1.4A; Table 1.3). Transcript abundance for *pgCPSII* was significantly different ($p < 0.05$) between the light and the dark, with up to an 8-fold increase in transcript abundance in the dark. In the light, transcript abundance for the urea HL condition was 2- to 3-fold higher than the NO_3^- HL, respectively. In the dark, the NO_3^- HL and NH_4^+ HL treatments displayed 2- to 6-fold increases in transcript abundance, whereas the urea HL treatment increased by only 50% between the light and dark.

Glutamine may also be produced by GSI in the cytosol of the cell. Transcript abundances for *GSI* were the greatest in cultures grown under the SL and HL irradiances for all N sources in the dark and were about 2-fold greater than in the light (Figure 1.4B). Although light intensity and N source affected transcript abundances, a statistical interaction between the two was only significant in the dark ($p < 0.05$). When *GSI* was compared to *pgCPSII*, transcript abundances were weakly correlated in the dark ($r^2 = 0.431$, $p < 0.05$; Figure 1.5). There were no correlations ($r^2 < 0.010$) for *pgCPSII* with *GSII* or *GSIII* in the light and dark.

<u>Additional pathways connected to the urea cycle</u>

Photorespiration generates CO_2 and NH_4^+ through the coordinated conversion of glycine into serine by the enzymes serine hydroxymethyltransferase and the multiprotein enzyme glycine decarboxylase (GDC). *GDCT* encodes the T-protein component of GDC found in the

mitochondria in *T. pseudonana*, where unCPS is located (Parker & Armbrust 2005). Transcript abundances for *GDCT* displayed a significant interaction between N source and light intensity ($p < 0.05$; Table 1.3). The NO_3^- HL, NO_3^- LL and urea HL, urea LL treatments were 3−4-fold greater than the control (Figure 1.6A). There were no significant differences in transcript abundances between the light and the dark ($p = 0.05$), and transcript abundances stayed the same or decreased in the dark, with the greatest decrease occurring in the NO_3^- and urea treatments. There was a weak correlation between *GDCT* and *unCPS* transcript abundances in the light ($r^2 = 0.227$, $p < 0.05$; data not shown).

Transcript abundance for *CK* varied based on N source and light intensity ($p < 0.05$; Table 1.3). SL treatments exhibited the greatest transcript abundances, which were 2-fold higher than the control (Figure 1.6B). There were no significant differences in transcript abundances in the light and the dark ($p = 0.660$). A weak correlation between *CK* and *unCPS* transcript abundances was observed in the light ($r^2 = 0.273$, $p < 0.05$; data not shown).

1.6 Discussion

Our results demonstrate that light intensity and N source differentially affect diurnal transcript accumulation of genes involved in the key metabolic pathways of the urea cycle (*unCPS*); urea hydrolysis (*URE*); N assimilation (*GSI*; *GSII*; *GSIII*); pyrimidine biosynthesis (*pgCPSII*); photorespiration (*GDCT*); and the formation of energy-storage compound creatine-P (*CK*). With few exceptions, the highest light treatment yielded the most transcripts for all genes and conditions surveyed; under high light, the greatest differences in transcript abundance among the N sources were also documented.

We identified a strong influence of the light:dark cycle on transcript abundances for genes associated with N metabolism. Transcripts encoding proteins targeted to the mitochondria were higher in the dark whereas transcripts encoding proteins targeted to the plastid were higher in the light. Our sampling scheme consisted of collecting two discrete samples: one sample in the light and one in the dark. This approach results in a conservative estimate of the impact of the diel cycle because of the potential to miss the peak in transcript abundance, although based on previous studies (Brown et al. 2009), we know that peaks in transcript abundance were likely captured for at least two of our genes (*GSII* and *GSIII*). Here, we summarize our findings for N assimilation and allocation in diatom cells by assuming that changes in transcript abundances reflect changes in flux through these pathways (Figure 1.7).

The influence of a diel cycle on nitrogen flow into the urea cycle

Diatoms, like all photoautotrophic organisms, are dependent upon sunlight for life. It is not surprising, then, that their core physiology is strongly impacted by the diurnal cycle. Light energy is harvested and converted to ATP and reductants in the plastid during the day, while respiration generates ATP and reductants through catabolism of storage compounds in the mitochondria at night. The tight coupling between C and N metabolism in photoautotrophs suggests a similar pattern of diel regulation exists for N including rates of uptake and assimilation (Berges et al. 1995; Needoba & Harrison 2004; Vanlerberghe et al. 1992; Vergara et al. 1998; Vincent 1992).

Transcript abundances for *unCPS* served as a proxy for N entry into the urea cycle. The observed correlation between transcript abundance for *unCPS* and *GSIII* in the dark and *unCPS* and *GSII* in the light, regardless of N source, suggests an intimate connection between glutamine production and the urea cycle; glutamine likely serves as an intermediate to deliver the required

amide group to the unCPS enzyme (Allen et al. 2011; Hong et al. 1994; Lawson et al. 1996; Lindley et al. 2007). In the light, glutamine appears to be generated in the plastid via GSII whereas in the dark, glutamine appears to be generated in the mitochondria via GSIII. The clearest interpretation of these results is that a transporter shuttles glutamine from the plastid to the mitochondria to fuel unCPS activity in the light. Glutamine transporters have been identified in both mammalian cells (Chaudhry et al. 2002; Indiveri et al. 1998) and plant cells (Yang et al. 2010). We identified three proteins in *T. pseudonana* with similarity to CAT8, a cationic amino acid transmembrane transporter recently shown to transport glutamine in *Arabidopsis* root tips (Yang et al. 2010), although *in silico* analyses did not predict plastid targeting for these proteins. We did find examples of putative amino acid transporters with clear plastid targeting. We include hypothetical glutamine transporters (GluT) (Figure 1.1) in our N models to connect the plastid-generated glutamine to the mitochondrial-localized enzymes of the urea cycle.

In the light, cells grown on either NO_3^- (Figure 1.7A) or urea (Figure 1.7B) accumulate the most transcripts for *unCPS*, and thus have the greatest assumed urea cycle activity. The NH_4^+-grown cells display lower *unCPS* transcript accumulation, suggesting N acquired from exogenous NH_4^+ does not appreciably impact the urea cycle in the light (Figure 1.7C). These results corroborate earlier findings of Lomas (2004) that NO_3^- and urea-grown cells may exhibit similar N physiology distinct from cells grown on NH_4^+. The observed difference between cells grown on different N sources likely results from the organellar compartmentalization of N pathways. When cells are grown on NO_3^-, the presumed exclusive reduction of N from NO_3^- into NH_4^+ within the plastid results in a high flux of NH_4^+ into plastid-localized glutamine biosynthesis and subsequent transport to the mitochondria. In contrast, cells grown on NH_4^+ appear to have a relatively low flux of NH_4^+ from the cytosol to the plastid, which results in low levels of plastid

GSII transcripts and glutamine biosynthesis. This may help to explain the low levels of *unCPS* transcript accumulation in the light for cells grown on NH_4^+.

Cells grown on urea as a sole N source hydrolyze the urea to NH_4^+ and CO_2 in the mitochondria. *URE* transcript abundances were significantly lower during the day than at night, regardless of N source, suggesting that most URE activity, and thus NH_4^+ generation from urea, occurs at night. We expected that during the day, the flux of NH_4^+ to the plastid in urea- grown cells would be as low as for NH_4^+ grown cells leading to similarly low levels of glutamine biosynthesis. Instead, transcript abundances for *unCPS* and *GSII* in the urea-grown cells were comparable to those of NO_3^--grown cells, suggesting that when grown on urea, cells transport NH_4^+ to the plastid during the day to generate glutamine, some of which is transported to the mitochondria to drive the urea cycle.

Our results suggest a disconnect between the low levels of NH_4^+ generated in the light via URE activity and the observed high levels of *GSII* transcript abundance. One possible explanation is that the relatively high concentrations of NH_4^+ contamination in the urea media were transported to the plastid during the day. However, this explanation requires that cells cycle N differently when grown on NH_4^+ alone than when grown on NH_4^+ and urea together based on the differences in *GSII* transcript abundance in the light. Alternatively, some of the NH_4^+ generated via URE activity during the night may be stored, perhaps in the vacuole, and transported to the plastid during the day.

In the dark, glutamine biosynthesis appears to occur predominately in the mitochondria, regardless of N source (Figure 1.7D-F). When cells are grown on NO_3^-, the NH_4^+ produced in the plastid at night is transported to the mitochondria for incorporation into glutamine. When cells

are grown on urea, the NH_4^+ generated from hydrolysis of urea at night remains within the mitochondria for incorporation into glutamine. When cells are grown on NH_4^+, a portion of the NH_4^+ is transported directly from the cytosol to the mitochondria. These hypothesized models for N movement require a plastid-targeted NH_4^+ transporter, which we could not definitively identify via *in silico* analyses. For the purposes of our model, we placed an NH_4^+ transporter (AMT, protein ID 13996) on the plastid membrane. Although the SignalP score did not meet our cutoff, it was high (0.721) and cleavage at the canonical motif (A-FA) resulted in detection of a putative transit peptide using ChloroP (score = 0.520). Since transporters are targeted to membranes, it is expected that the signal and/or targeting peptides for transporters may be different from proteins that function in the interior of plastids or thylakoids. Additional work examining any conserved motifs among transporters or targeting proteins using *in vivo* techniques are necessary to confirm the localization of the seven *T. pseudonana* NH_4^+ transporters (Figure 1.1).

The fate of urea and re-distribution of nitrogen

Our evidence that the urea cycle functions in both the light and the dark, as well as the potentially reduced URE activity in the light, provides several possible scenarios for urea cycle N flux. Previous studies have documented the constitutive expression of the URE enzyme with variations in the magnitude of URE activity determined by N source, as well as diatom species (Fan et al. 2003; Lomas 2004; Peers et al. 2000; Solomon et al. 2010). Although we observed weak relationships between *GSI*/*GSII*/*GSIII* and *URE* transcript abundances, we can neither confirm nor negate a relationship among these enzymes. Given this uncertainty, we propose two additional scenarios to explain the observed patterns in light:dark *URE* expression in *T. pseudonana*.

In the first scenario, *URE* transcript accumulation is driven by the light: dark phase, in a manner similar to what has been hypothesized here for *GSII* and *GSIII*. In this instance, the main by-product of the urea cycle is urea itself and consequently internal urea concentrations will increase during the day when URE activity is lowest and decrease at night, when URE activity is greatest. A preferential storage of urea in the vacuole during the day is predicted to enhance buoyancy based on measurements of cellular density and osmolyte concentrations (Boyd & Gradmann 2002; Raven 1987). This could be a strong advantage to a heavy diatom cell that needs to be near the surface during the day to photosynthesize.

In the second scenario, *URE* transcript accumulation is driven by substrate concentration. In this scenario, less urea is generated by the cycle during the day because cycle intermediates, most obviously fumarate (TCA cycle) and arginine, are channeled into peripheral cellular pathways. Arginine is required for proline synthesis, cell signaling (e.g. nitric oxide), and the formation of agmatine (Wu & Morris 1998). Arginine also serves as a precursor for long-chain polyamine synthesis, required for diatom frustule formation and cell division (Armbrust et al. 2004). In light: dark synchronized cultures, *T. pseudonana* increases protein synthesis at the onset of the light phase following arrest in G1 at the end of the dark phase (Hildebrand & Dahlin 2000). Although our cultures were not synchronized, they were acclimated to growth on a light: dark cycle, and thus, we would expect arginine demand to increase at the beginning of the light phase along with protein synthesis. As the cell cycle progresses, the demand for arginine would decline, allowing the urea cycle to generate more urea and less intermediates. This process would ultimately lead to an increase in urea accumulation in the mitochondria, followed by an increase in URE activity in the dark. Of the above two scenarios, URE activity is likely controlled by the

availability of urea substrate given the documented connection between the urea cycle and amino acid synthesis (Allen et al. 2011).

Nitrogen incorporation in the cytosol

pgCPSII is required for catalysis of the first step of pyrimidine biosynthesis in the cytosol and requires glutamine as a source of N. Whereas *unCPS* is transcribed throughout the light: dark cycle, *pgCPSII* is affected by the presence of a diel cycle with a greater accumulation of transcripts in the dark. The differences in the *unCPS* and *pgCPSII* transcripts indicate that N allocation to these pathways is not uniform. Similar to the correlations observed between *unCPS* and *GSIII* in the dark and to *GSII* in the light, there appears to be an association between *pgCPSII* and *GSI*. *GSI* transcripts were significantly higher in the dark, and we observed a correlation between *GSI* and *pgCPSII* in the dark. This relationship provides additional evidence for glutamine partitioning in the cell. In the light, the majority of glutamine synthesis occurs in the plastid (via GSII), whereas at night, glutamine synthesis occurs in both the mitochondria (GSIII) and the cytosol (GSI). GSIII-derived glutamine is then incorporated into processes underway in the mitochondria (e.g. urea cycle), whereas GSI-derived glutamine is utilized via cytosolic pathways (e.g. pyrimidine synthesis).

Role of light intensity on nitrogen and energy management pathways

Two potential links exist between the urea cycle and cellular energy management: photorespiration and production of creatine-P. The multi-subunit enzyme GDC is localized to the mitochondria and links photorespiration to N metabolism by participating in the release of NH_4^+ from glycine. It has been hypothesized that the NH_4^+ generated by GDC in the mitochondria during photorespiration may either feed the urea cycle by providing a substrate for unCPS (Allen

et al. 2006) or be transported back to the plastid for incorporation into glutamine by GSII (Brown et al. 2009; Parker & Armbrust 2005).

However, in our experiments, no obvious similarities were seen between the transcript abundance patterns for *GDCT* and either *unCPS* or *GSII*. The timing of our transcript data for *GDCT* corresponds with the lowest transcript abundance observed for the gene encoding another subunit of this complex, *GDCP* (glycine decarboxylase P-subunit) at a similar time point in the light phase. Granum et al. (2009) documented the greatest *GDCP* transcript abundances at the end of the light phase. Thus, additional timescale experiments are necessary to test these alternative hypotheses.

CK is involved in the interconversion between ATP and creatine-P. In animals, creatine-P balances energy fluctuations in the cell and is often localized to regions of high ATP and reductant production such as the mitochondria (Wallimann et al. 1992; 1998). In *T. pseudonana*, CK appears localized to the plastid rather than the mitochondria and cytosolic space, as in mammalian cells (Ellington & Hines 1991; Farres et al. 2002; Wallimann et al. 1998). Creatine culture amendments have been hypothesized to buffer oscillations in the ATP pools generated during photosynthesis in a species of dinoflagellate (Roenneberg & Morse 1994; Roenneberg et al. 1988). CK has not been reported in green plants; however, it has been engineered into transgenic tobacco as a mechanism for increasing stress tolerance and energy buffering (Farres et al. 2002). As with *GDCT*, *CK* transcript abundance was significantly affected by both N source and irradiance. Interestingly, when *T. pseudonana* is grown on NH_4^+ as a sole N source, the greatest *CK* transcript accumulation occurs in cells grown at a moderate irradiance. These results suggest that N is diverted to other urea cycle products at low and high light intensities, and correlates with the low level of urea cycle activity during the day at high light in cells grown on

NH_4^+. In the dark, the greatest transcript accumulation occurred in urea-grown cells, perhaps indicating that the majority of N entering the urea cycle under these conditions is diverted to energy storage rather than to arginine synthesis.

Our results highlight the strong influence of a diel cycle, light intensity and N source on the flow of N in a diatom cell, and begin to incorporate these processes into the broader context of diatom metabolism in the marine environment through cell models for N flow. When cells experience high daily irradiance, N is partitioned between the plastid during the day and the mitochondria at night with variations based on a particular N source. We can now update our view of N metabolism in the cell to include light: dark processes such as the role of the urea cycle in producing urea and supplying regenerated N to the cell. Thus, a better understanding of the metabolic reaction of diatoms to N sources and light conditions common in the surface ocean emerges from our data.

1.7 Acknowledgments

The authors thank Dr. Julie Koester for her insightful comments on this manuscript, helpful discussions and assistance with statistical applications. We would also like to acknowledge Chris Berthiaume for his assistance with the RNAseq transcriptome data, Dave Schruth for help with the R software package and Claire Ellis for her assistance with gene sequencing efforts. We also appreciate feedback from two anonymous reviewers.

1.8 References

Allen AE. (2005). Beyond sequence homology: Redundant ammonium transporters in a marine diatom are not functionally equivalent. J Phycol 41:4–6.

Allen AE, Dupont CL, Obornik M, Horak A, Nunes-Nesi A, McCrow JP, et al. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature 473:203–209.

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

Armbrust EV, Berges JAJ, Bowler CC, Green BRB, Martinez DD, Putnam NH, et al. (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306:79–86.

Beevers L, Storey R. (1976). Glutamate synthetase in developing cotyledons of *Piseum sativum*. Plant Physiol 57:862–866.

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

Berges JA, Cochlan WP, Harrison PJ. (1995). Laboratory and field responses of algal nitrate reductase to diel periodicity in irradiance, nitrate exhaustion, and the presence of ammonium. Mar Ecol Prog Ser 124:259–269.

Berges JA, Franklin DJ, Harrison PJ. (2001). Evolution of an artificial seawater medium: improvements in enriched seawater, artificial water over the last two decades. J Phycol 37:1138–1145.

Bidle KD, Bender SJ. (2008). Iron starvation and culture age activate metacaspases and programmed cell death in the marine diatom *Thalassiosira pseudonana*. Eukaryot Cell 7:223–236.

Bowler CC, Allen AEA, Badger JHJ, Grimwood JJ, Jabbari KK, Kuo AA, et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456:239–244.

Boyd C, Gradmann D. (2002). Impact of osmolytes on buoyancy of marine phytoplankton. Mar Biol 141:605–618.

Brand LE, Guillard RRL, Murphy LS. (1981). A method for the rapid and precise determination of acclimated phytoplankton reproduction rates. J Plankton Res 3:193–201.

Brown KL, Twing KI, Robertson DL. (2009). Unraveling the regulation of nitrogen assimilation in the diatom *Thalassiosira pseudonana* (Bacillariophyceae): Diurnal variations in transcript levels for five genes involved in nitrogen assimilation. J Phycol 45:413–426.

Chaudhry FA, Reimer RJ, Edwards RH. (2002). The glutamine commute: take the N line and transfer to the A. J Cell Biol 157:349–355.

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

Denis-Duphil M. (1989). Pyrimidine biosynthesis in *Saccharomyces cerevisiae*: the ura2 cluster gene, its multifunctional enzyme product, and other structural or regulatory genes involved in de novo UMP synthesis. Biochem Cell Biol 67:612–631.

Dortch Q. (1990). The interaction between ammonium and nitrate uptake in phytoplankton. Mar Ecol Prog Ser 61:183–201.

Dortch Q, Thompson P, Harrison PJ. (1991). Short-term interaction between nitrate and ammonium uptake in *Thalassiosira pseudonana*- effect of preconditioning nitrogen source and growth rate. Mar Biol 110:183–193.

Durkin CA, Mock T, Armbrust EV. (2009). Chitin in diatoms and its association with the cell wall. Eukaryot Cell 8:1038–1050.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

Ellington W, Hines A. (1991). Mitochondrial activities of phosphagen kinases are not widely distributed in the invertebrates. Biol Bull 180:505–507.

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

Fan C, Glibert PM, Alexander JA, Lomas M. (2003). Characterization of urease activity in three marine phytoplankton species, *Aureococcus anophagefferens*, *Prorocentrum minimum*, and *Thalassiosira weissflogii*. Mar Biol 142:949–958.

Farres J, Holmberg N, Schlattner U, Bailey J, Wallimann T, Kallio P. (2002). Expressing creatine kinase in transgenic tobacco - a first step towards introducing an energy buffering system in plants. Transgenic Res 11:49–59.

Gaufichon L, Reisdorf-Cren M, Rothstein SJ, Chardon F, Suzuki A. (2010). Biological functions of asparagine synthetase in plants. Plant Sci 179:141–153.

Ghoshroy S, Binder M, Tartar A, Robertson DL. (2010). Molecular evolution of glutamine synthetase II: Phylogenetic evidence of a non-endosymbiotic gene transfer event early in plant evolution. Bmc Evol Biol 10:198.

Granum E, Roberts K, Raven JA, Leegood RC. (2009). Primary carbon and nitrogen metabolic gene expression in the diatom *Thalassiosira pseudonana* (Bacillariophyceae): Diel periodicity and effects of inorganic carbon and nitrogen. J Phycol 45:1083–1092.

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

Hildebrand M. (2005). Cloning and functional characterization of ammonium transporters from the marine diatom *Cylindrotheca fusiformis* (Bacillariophyceae). J Phycol 41:105–113.

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.

Holden HM, Thoden JB, Raushel FM. (1998). Carbamoyl phosphate synthetase: a tunnel runs through it. Curr Opin Struc Biol 8:679–685.

Hong J, Salo WL, Lusty C, Anderson PM. (1994). Carbamyl phosphate synthetase III, an evolutionary intermediate in the transition between glutamine-dependent and ammonia-dependent carbamyl phosphate synthetases. J Mol Biol 243:131–140.

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier C, et al. (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Res 35:W585–W587.

Indiveri C, Abruzzo G, Stipani I, Palmieri F. (1998). Identification and purification of the reconstitutively active glutamine carrier from rat kidney mitochondria. Biochem J 333(Pt 2):285–290.

Jaynes JB, Chamberlain JS, Buskin JN, Johnson JE, Hauschka SD. (1986). Transcriptional regulation of the muscle creatine kinase gene and regulated expression in transfected mouse myoblasts. Mol Cell Biol 6:2855–2864.

Kang L-K, Hwang S-PL, Lin H-J, Chen P-C, Chang J. (2009). Establishment of minimal and maximal transcript levels for nitrate transporter genes for detecting nitrogen deficiency in the marine phytoplankton *Isochrysis galbana* (Prymnesiophyceae) and *Thalassiosira pseudonana* (Bacillariophyceae). J Phycol 45:864–872.

Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, et al. (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. PLOS One 3:e1426.

Lawson FS, Charlebois RL, Dillon JA. (1996). Phylogenetic analysis of carbamoylphosphate synthetase genes: complex evolutionary history includes an internal duplication within a gene which can root the tree of life. Mol Biol Evol 13:970–977.

Lindley TE, Laberge T, Hall A, Hewett-Emmett D, Walsh PJ, Anderson PM. (2007). Sequence, expression and evolutionary relationships of carbamoyl phosphate synthetase I in the toad *Xenopus laevis*. J Exp Zool Part A 307A:163–175.

Lomas M. (2004). Nitrate reductase and urease enzyme activity in the marine diatom *Thalassiosira weissflogii* (Bacillariophyceae): Interactions among nitrogen substrates. Mar

Biol 144:37–44.

Marchetti A, Schruth DM, Durkin CA, Parker MS, Kodner RB, Berthiaume CT, et al. (2012). Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. Proc Natl Acad Sci 109:1–9.

Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, et al. (2009). CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res 37:D205–10.

Mobley H, Island M, Hausinger R. (1995). Molecular biology of microbial ureases. Microbiol Rev 59:451–480.

Mock T, Samanta MP, Iverson V, Berthiaume C, Robison MM, Holtermann K, et al. (2008). Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. Proc Natl Acad Sci 105:1579–1584.

Nassoury N, Morse D. (2005). Protein targeting to the chloroplasts of photosynthetic eukaryotes: getting there is half the fun. Biochim Biophys Acta 1743:5–19.

Needoba JA, Harrison PJ. (2004). Influence of low light and a light: Dark cycle on NO_3^- uptake, intracellular NO_3^-, and nitrogen isotope fractionation by marine phytoplankton. J Phycol 40:505–516.

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

Nunn BL, Aker JR, Shaffer SA, Tsai S, Strzepek RF, Boyd PW, et al. (2009). Deciphering diatom biochemical pathways via whole-cell proteomics. Aquat Microb Ecol 55:241–253.

Parker MS, Armbrust EV. (2005). Synergistic effects of light, temperature, and nitrogen source on transcription of genes for carbon and nitrogen metabolism in the centric diatom *Thalassiosira pseudonana* (Bacillariophyceae). J Phycol 41:1142–1153.

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

Peers G, Milligan A, Harrison PJ. (2000). Assay optimization and regulation of urease activity in two marine diatoms. J Phycol 36:523–528.

Pierard A, Lissens W, Halleux P, Cunin R, Glansdorff N. (1980). Role of transcriptional regulation and enzyme inactivation in the synthesis of *Escherichia coli* carbamoyl phosphate synthase. J Bacteriol 141:382–385.

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

Price NM, Harrison PJ. (1987). Comparison of methods for the analysis of dissolved urea in seawater. Mar Biol 94:307–317.

Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 339:62–66.

Raven JA. (1987). The role of vacuoles. New Phytol 106:357–422.

Robertson DL, Smith GJ, Alberte R. (1999). Characterization of a cDNA encoding glutamine synthetase from the marine diatom *Skeletonema costatum* (Bacillariophyceae). J Phycol 35:786–797.

Robertson DL, Tartar A. (2006). Evolution of glutamine synthetase in heterokonts: Evidence for endosymbiotic gene transfer and the early evolution of photosynthesis. Mol Biol Evol 23:1048–1055.

Roenneberg T, Morse D. (1994). Cellular mechanisms of the internal clock of a unicellular organism. Science 81:343–349.

Roenneberg T, Nakamura H, Hastings J. (1988). Creatine accelerates the circadian clock in a unicellular alga. Nature 334:432–434.

Siaut MM, Heijde M, Mangogna M, Montsant A, Coesel S, Allen AE, et al. (2007). Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. Gene 406:23–35.

Slocum RD. (2005). Genes, enzymes and regulation of arginine biosynthesis in plants. Plant Physiol Biochem 43:729–745.

Solomon CM, Collier JL, Berg GM, Glibert PM. (2010). Role of urea in microbial metabolism in aquatic systems: a biochemical and molecular review. Aquat Microb Ecol 59:67–88.

Syrett P. (1981). Nitrogen metabolism of microalgae. Can B Fish Aquat Sci 182–210.

Takabayashi M, Wilkerson F, Robertson DL. (2005). Response of glutamine synthetase gene transcription and enzyme activity to external nitrogen sources in the diatom *Skeletonema costatum* (Bacillariophyceae). J Phycol 41:84–94.

Tatibana M, Shigesad K. (1972). Control of pyrimidine biosynthesis in mammalian tissues .5. Regulation of glutamine-dependent carbamyl phosphate synthetase- activation by 5-phosphoribosyl 1-pyrophosphate and inhibition by uridine triphosphate. J Biochem (Tokyo) 72:249–271.

Taylor NL, Howell KA, Heazlewood JL, Tan TYW, Narsai R, Huang S, et al. (2010). Analysis of the rice mitochondrial carrier family reveals anaerobic accumulation of a basic amino acid carrier involved in arginine metabolism during seed germination. Plant Physiol 154:691–704.

UNESCO. (1994). Protocols for the Joint Global Ocean Flux Study (JGOFS) core measurements. In:IOC Manual and Guides 29, IOC, eds. Intergovernmental Oceanographic

Commission: Paris, p. 170.

Vanlerberghe G, Huppe H, Vlossak K, Turpin D. (1992). Activation of respiration to support dark NO_3^- and NH_4^+ assimilation in the green alga *Selenastrum minutum*. Plant Physiol 99:495–500.

Vardi A, Bidle KD, Kwityn C, Hirsh DJ, Thompson SM, Callow JA, et al. (2008). A diatom gene regulating nitric-oxide signaling and susceptibility to diatom-derived aldehydes. Curr Biol 18:895–899.

Vergara J, Berges JA, Falkowski PG. (1998). Diel periodicity of nitrate reductase activity and protein levels in the marine diatom *Thalassiosira weissflogii* (Bacillariophyceae). J Phycol 34:952–961.

Vincent W. (1992). The daily pattern of nitrogen uptake by phytoplankton in dynamic mixed layer environments. Hydrobiologia 238:37–52.

Wallimann T, Dolder M, Schlattner U, Eder M, Hornemann T, O'Gorman E, et al. (1998). Some new aspects of creatine kinase (CK): compartmentation, structure, function and regulation for cellular and, mitochondrial bioenergetics and physiology. Biofactors 8:229–234.

Wallimann T, Wyss M, Brdiczka D, Nicolay K, Eppenberger H. (1992). Intracellular compartmentation, structure and function of creatine kinase isoenzymes in tissues with high and fluctuating energy demands- the phosphocreatine circuit for cellular energy homeostasis. Biochem J 281:21–40.

Wu G, Morris SM. (1998). Arginine metabolism: nitric oxide and beyond. Biochem J 336 :1–17.

Yang H, Bogner M, Stierhof Y-D, Ludewig U. (2010). H-independent glutamine transport in plant root tips. PLOS One 5:e8917.

Zar JH. (1999). Biostatistical Analysis. 4 ed. Prentice Hall: Upper Saddle River.

Zehr JP, Falkowski PG. (1988). Pathway of ammonium assimilation in a marine diatom determined with the radiotracer N^{13}. J Phycol 24:588–591.

Table 1.1 Treatments, growth rates (± SE), and F_v/F_m for *T. pseudonana* maintained in exponential growth at 13°C on a 16 h light:8 h dark cycle. F_v/F_m taken at time of culture harvest.

Treatment ID	Nitrogen Source	Light (μmol photons m^{-2} s^{-1})	Growth rate (d^{-1})	F_v/F_m Light	F_v/F_m Dark
NO_3^- LL	Nitrate	50	0.78 ± 0.01	0.67 ± 0.00	0.67 ± 0.00
NO_3^- SL		190	0.99 ± 0.02	0.65 ± 0.00	0.66 ± 0.00
NO_3^- HL		400	0.92 ± 0.08	0.63 ± 0.01	0.66 ± 0.00
NH_4^+ LL	Ammonium	50	0.92 ± 0.01	0.68 ± 0.00	0.67 ± 0.00
NH_4^+ SL		190	1.22 ± 0.03	0.67 ± 0.00	0.67 ± 0.00
NH_4^+ HL		400	1.05 ± 0.01	0.63 ± 0.01	0.64 ± 0.00
Urea LL	Urea	50	1.11 ± 0.01	0.69 ± 0.01	0.67 ± 0.00
Urea SL		190	1.34 ± 0.02	0.67 ± 0.00	0.66 ± 0.00
Urea HL		400	1.05 ± 0.02	0.66 ± 0.00	0.49 ± 0.01

Table 1.2 Primer sequences for genes encoding proteins of interest.

Primer ID	Protein (*Gene*)	Protein ID	Amplicon Length (bp)	Primer Sequence (5' – 3')
unCPS-F	Carbamoyl-phosphate synthetase III			CTCCGGACTTTGTAGAGGAAATTA
unCPS-R	(*unCPS*)	40323	167	ATTGATGACGGATTCAATAGGAGT
URE-F	Urease			TGGGACTCGTCATGTAAGCA
URE-R	(*URE*)	30193	121	CAAAGGATCGCACCACAATC
GSIII- F	Glutamine Synthetase III			ACATTGGGTTGGAGCAAGAG
GSIII- R	(*GSIII*)	270138	250	TACTGATTGGGAGCCACCTC
pgCPSII-F	Carbamoyl-phosphate synthetase II			CGAGTGATCCTTTCCGTGTT
pgCPSII-R	(*pgCPSII*)	24248	126	GCTTCGACACCACCTTCATA
GSI- F	Glutamine Synthetase I			TGAAAGTGCAGCTGGACAAC
GSI- R	(*GSI*)	261531	160	AGCCTGAGTCGTTGACGTTT
				***GACTCTTCGACCCGACTTTT**
			597[a]	***AGACTGACCCCAGAAGGTT**
CK-F	Creatine Kinase			TTGGTCGATTGCCTGGAT
CK-R	(*CK*)	263946	128	CTTTCCCAATCGTTCTGCAT
GSII- F	Glutamine Synthetase II			
GSII- R	(*GSII*)	26051	345	
GDCT- F	Glycine Decarboxylase T-subunit			
GDCT- R	(*GDCT*)	36208	322	

[a]Primer set used to clone the *GSI* fragment (bold) prior to qPCR amplification

Table 1.3 Analysis of Variance (ANOVA) and Repeated Measures statistics for transcript abundance data. Significant *p*-values ($p < 0.05$) in bold.

Gene	**Factors**	**Light Cycle** Two-way ANOVA			**Dark Cycle** Two-way ANOVA			**Light:Dark Cycle** Repeated Measures		
		Df	F	*p*	Df	F	*p*	Df	F	*p*
unCPS	Nitrogen Source x Light Intensity	4	19.230	**8.5×10^{-11}**	4	7.8957	**2.5×10^{-5}**	1	3.5689	6.1×10^{-2}
URE	Nitrogen Source x Light Intensity	4	8.9062	**6.7×10^{-6}**	4	3.453	**1.2×10^{-2}**	1	12.096	**6.5×10^{-4}**
GSIII	Nitrogen Source x Light Intensity	4	8.3406	**1.5×10^{-5}**	4	1.9619	1.1×10^{-1}[a]	1	33.041	**4.7×10^{-8}**
GSII	Nitrogen Source x Light Intensity	4	6.8743	**9.6×10^{-5}**	4	7.2866	**5.5×10^{-5}**	1	4.3786	**3.8×10^{-2}**
pgCPSII	Nitrogen Source x Light Intensity	4	10.702	**7.3×10^{-7}**	4	8.2236	**1.6×10^{-5}**	1	9.2499	**2.8×10^{-3}**
GSI	Nitrogen Source x Light Intensity	3	1.2044	3.2×10^{-1}[a]	3	2.8988	**4.2×10^{-2}**	1	30.747	**1.5×10^{-7}**
GDCT	Nitrogen Source x Light Intensity	4	10.4390	**1.1×10^{-6}**	4	8.1890	**1.7×10^{-5}**	1	0.3964	5.3×10^{-1}
CK	Nitrogen Source x Light Intensity	4	7.7558	**3.0×10^{-5}**	4	10.054	**1.6×10^{-6}**	1	0.1906	6.6×10^{-1}

[a]Refer to Supplemental Table 1.1 for full two-way ANOVA and repeated measures tables

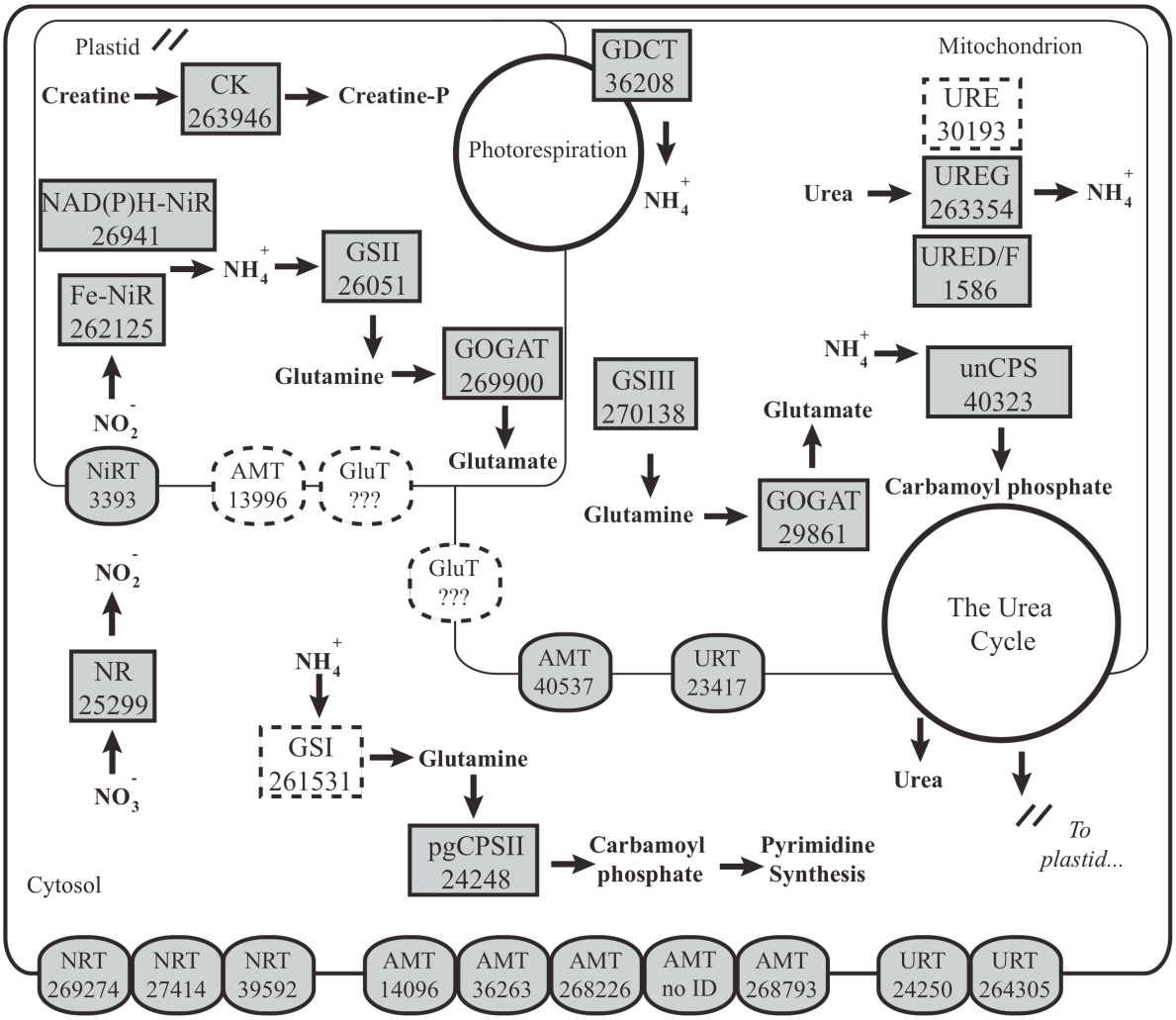


Figure 1.1 Predicted localization for proteins associated with the urea cycle and N metabolism in the cell. Protein name and ID are given for each protein of interest. Gray boxes, outlined in black are shown for proteins where hypothesized targeting is clear. White boxes, outlined with dashed lines are shown for proteins where targeting remains unclear. The ferredoxin-dependent NiR enzyme (protein ID 262125) is abbreviated as Fe-NiR.

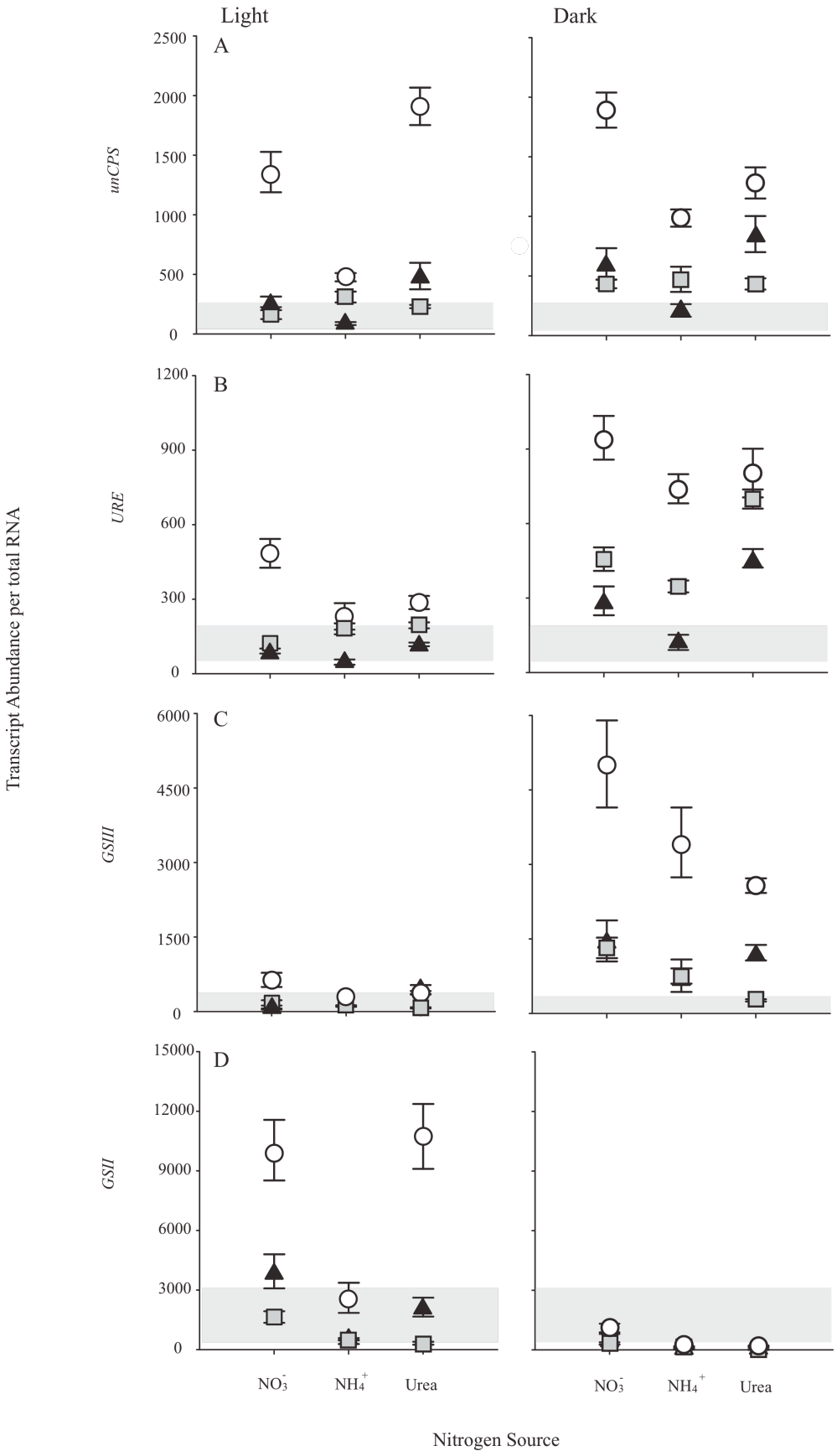

Light
Dark
A
B
C
D
unCPS
URE
GSIII
GSII
2500
2000
1500
1000
500
0
1200
900
600
300
6000
4500
3000
1500
15000
12000
9000
6000
3000
NO3-
NH4+
Urea
Transcript Abundance per total RNA
Nitrogen Source

Figure 1.2 Transcript abundance for *unCPS* (A), *URE* (B), *GSIII* (C), and *GSII* (D) during the light or dark phase. Cells were grown under NO_3^-, NH_4^+ or urea at low light (black triangles), saturating light (gray squares) or high light (white circles) conditions. The gray shaded area indicates ± two-fold change in transcript abundance relative to the NO_3^- SL light phase treatment. Error bars larger than symbol size represent ± standard error (SE) from three biological replicates and include three technical replicates per biological replicate. For *GSIII*, the mean of two biological replicates is shown for the NH_4^+ HL treatment in the light. Note differences in the y-axis scale for each gene.

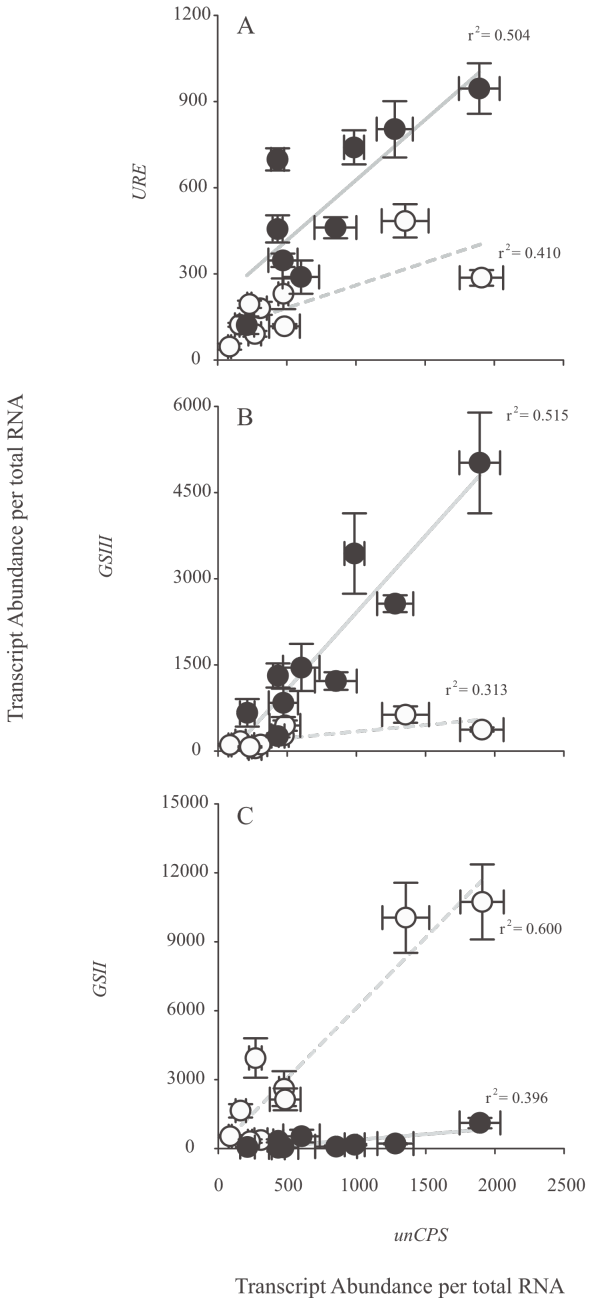


Figure 1.3 Transcript abundance for *URE* (A), *GSIII* (B) and *GSII* (C) versus *unCPS* during the light phase (white circles) or dark phase (black circles). Linear regressions are indicated by gray, dashed (light phase) or solid lines (dark phase).

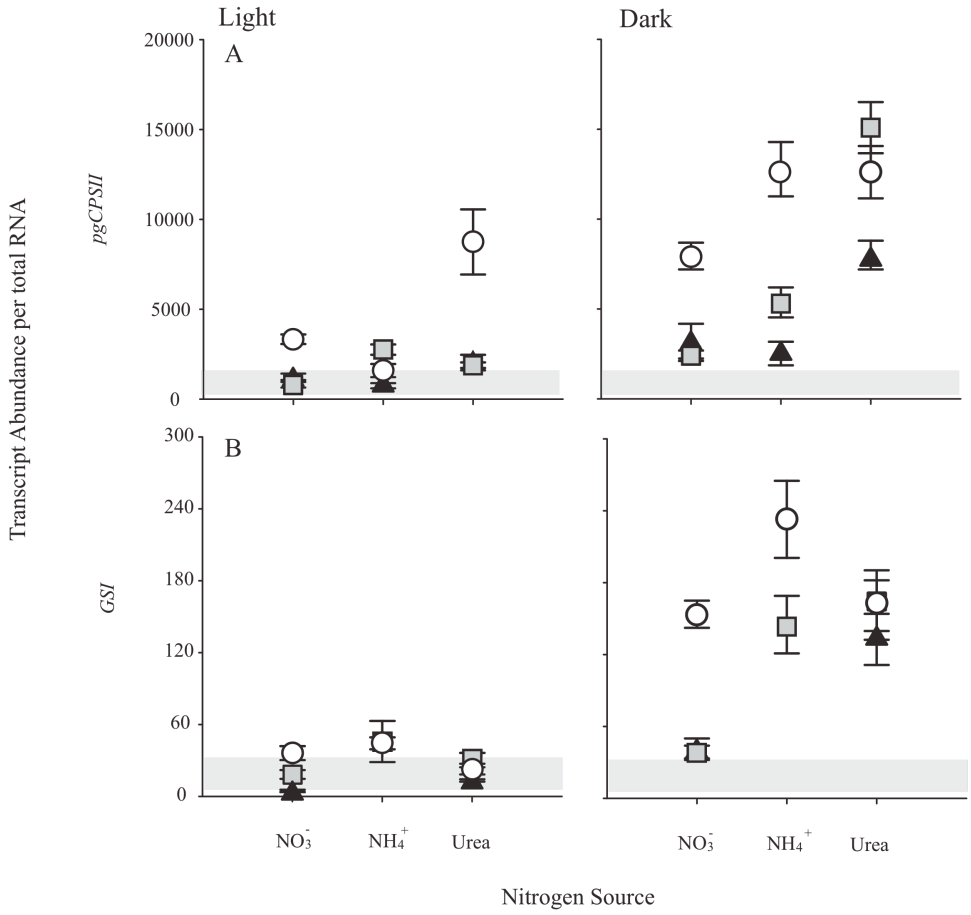


Figure 1.4 Transcript abundance for *pgCPSII* (A) and *GSI* (B) during the light or dark phase. Cells were grown under NO_3^-, NH_4^+ or urea at low light (black triangles), saturating light (gray squares) or high light (white circles) conditions. The gray shaded area indicates ± two-fold change in transcript abundance relative to the NO_3^- SL light phase treatment. Error bars larger than symbol size represent ± standard error (SE) from three biological replicates and include three technical replicates per biological replicate. For GSI, transcript abundances for NH_4^+ LL light phase and dark phase treatments are not included. Note differences in the y-axis scale for each gene.

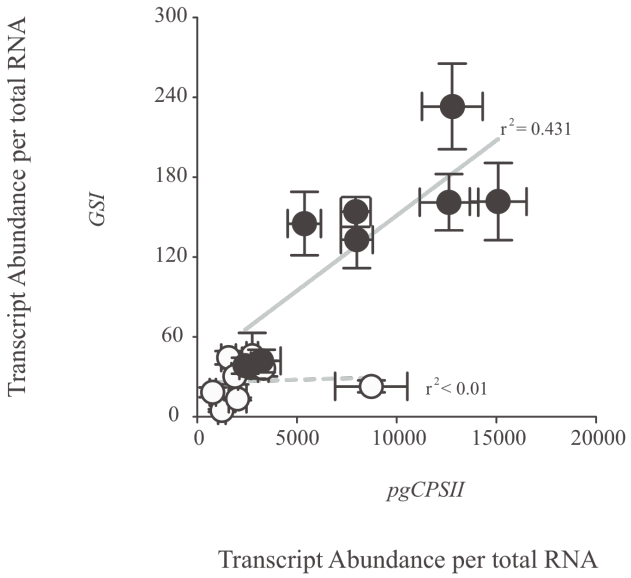


Figure 1.5 Transcript abundance for *GSI* versus *pgCPSII* during the light phase (white circles) or dark phase (black circles). Linear regressions are indicated by gray, dashed (light phase) or solid lines (dark phase).

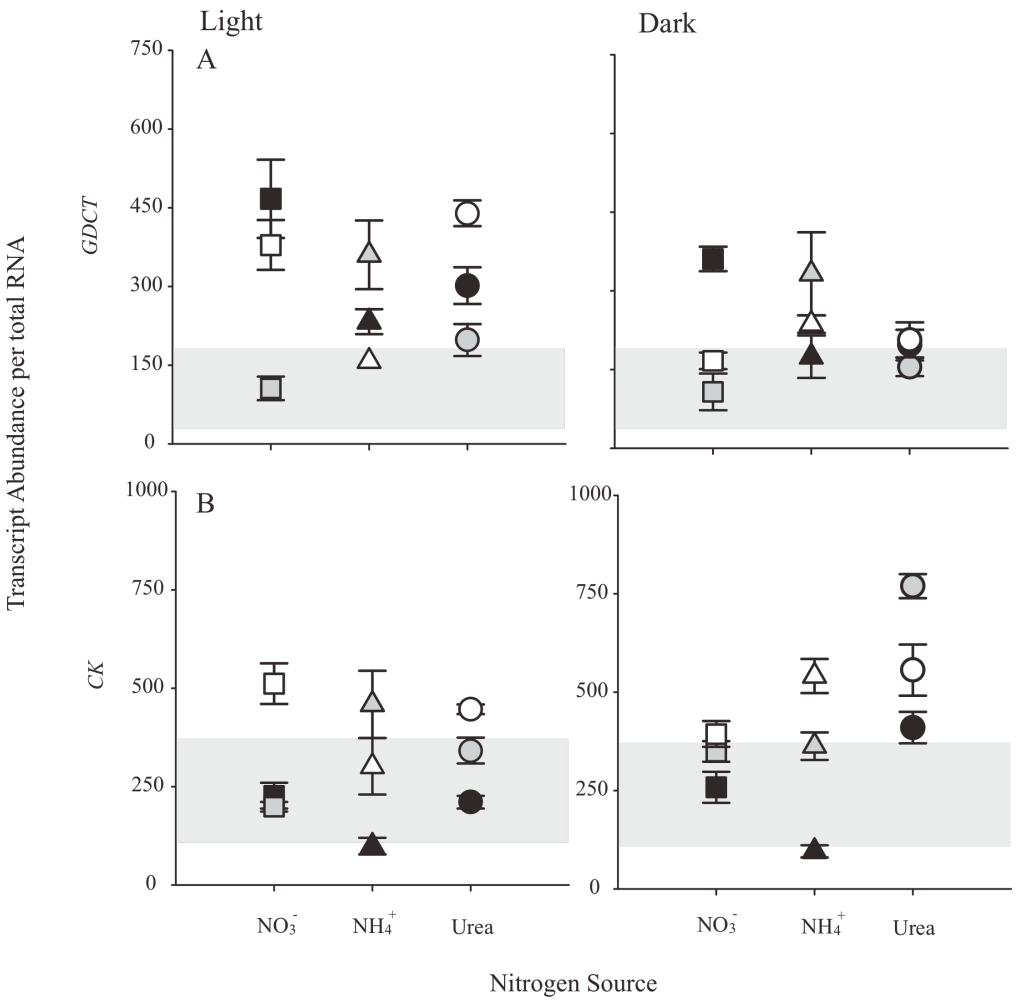


Figure 1.6 Transcript abundance for *GDCT* (A) and *CK* (B) during the light or dark phase. Cells were grown under NO_3^-, NH_4^+ or urea at low light (black triangles), saturating light (gray squares) or high light (white circles) conditions. The gray shaded area indicates ± two-fold change in transcript abundance relative to the NO_3^- SL light phase treatment. Error bars larger than symbol size represent ± standard error (SE) from three biological replicates and include three technical replicates per biological replicate. For *GDCT*, mean of two biological replicates are shown for the NH_4^+ HL treatment in the light. Note differences in the y-axis scale for the genes.

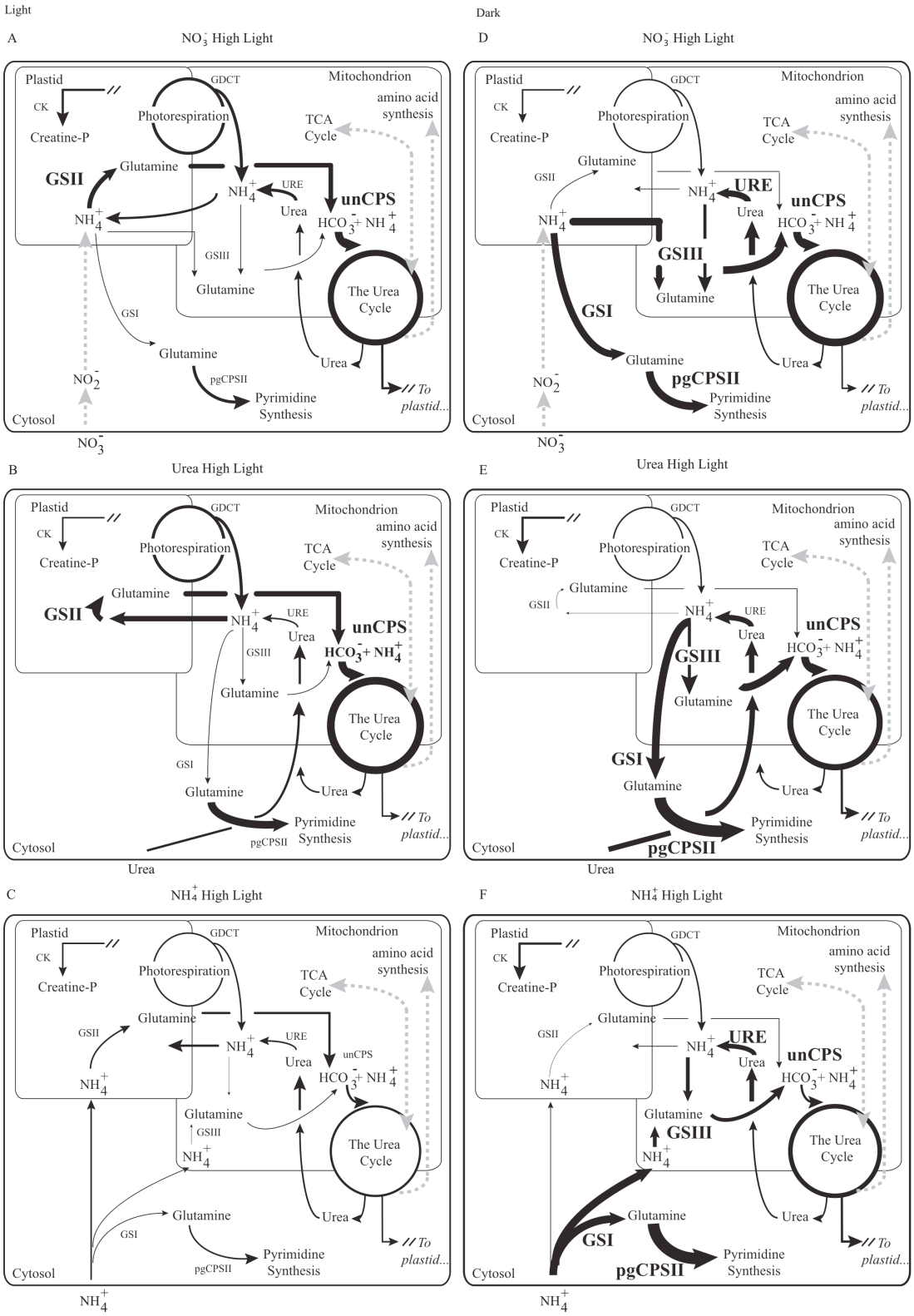
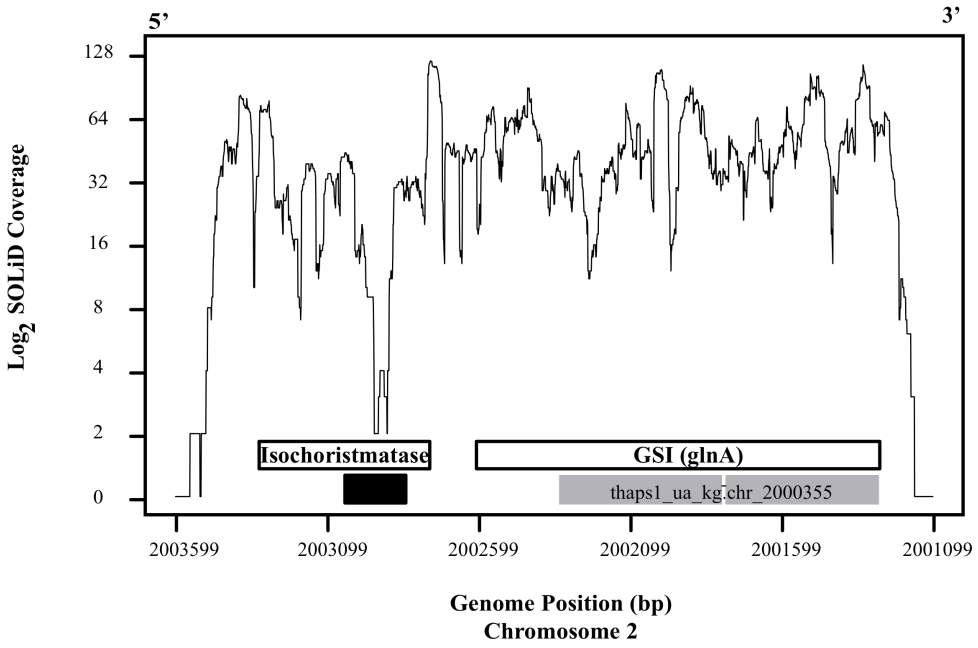

Light
A
NO_3^- High Light
Plastid
CK
Creatine-P
Photorespiration
GDCT
Mitochondrion
amino acid synthesis
TCA Cycle
GSII
Glutamine
NH_4^+
URE
Urea
unCPS
$HCO_3^- + NH_4^+$
GSIII
Glutamine
The Urea Cycle
GSI
Glutamine
Urea
NO_2^-
pgCPSII
Pyrimidine Synthesis
To plastid...
Cytosol
NO_3^-
Dark
D
NO_3^- High Light
B
Urea High Light
E
Urea High Light
C
NH_4^+ High Light
F
NH_4^+ High Light

Figure 1.7 A hypothetical model of the *T. pseudonana* cell showing the connection between N metabolism and cellular energy management pathways including their connection to the urea cycle in cells grown under NO_3^- (A, light phase; D, dark phase), Urea (B, light; E, dark) or NH_4^+ (C, light; F, dark) under HL. Proteins of interest are indicated with enzyme abbreviation. Substrates and products are also shown and their associated pathways vary based on the flow of N in the cell (thickness of lines). In conditions where N flow is predicted to be elevated (relative to the control), the associated enzymes are enlarged and in bold. Gray, dashed lines denote N metabolism pathways not examined in this study; the gray lines are not scaled to represent N flow.

Supplemental Table 1.1 Summary of 2-way ANOVA and repeated measures statistics conducted on growth rates, F_v/F_m and steady-state transcript levels for *unCPS*, *URE*, *GSIII*, *GSII*, *pgCPSII*, *GSI*, *GDCT*, and *CK*. Significantly different values are in bold ($p < 0.05$).
See attached .xls file.

Supplemental Table 1.2 Full-length predicted sequences for proteins connected to nitrogen metabolism in *T. pseudonana*, including urea cycle and related-enzymes. Protein IDs, targeting information (Signal P, Chloro P, MitoProtII, and WoLF PSORT), hypothesized localization information, and EST/RefSeq support are provided, and inform Figure 1.1.
See attached .xls file.

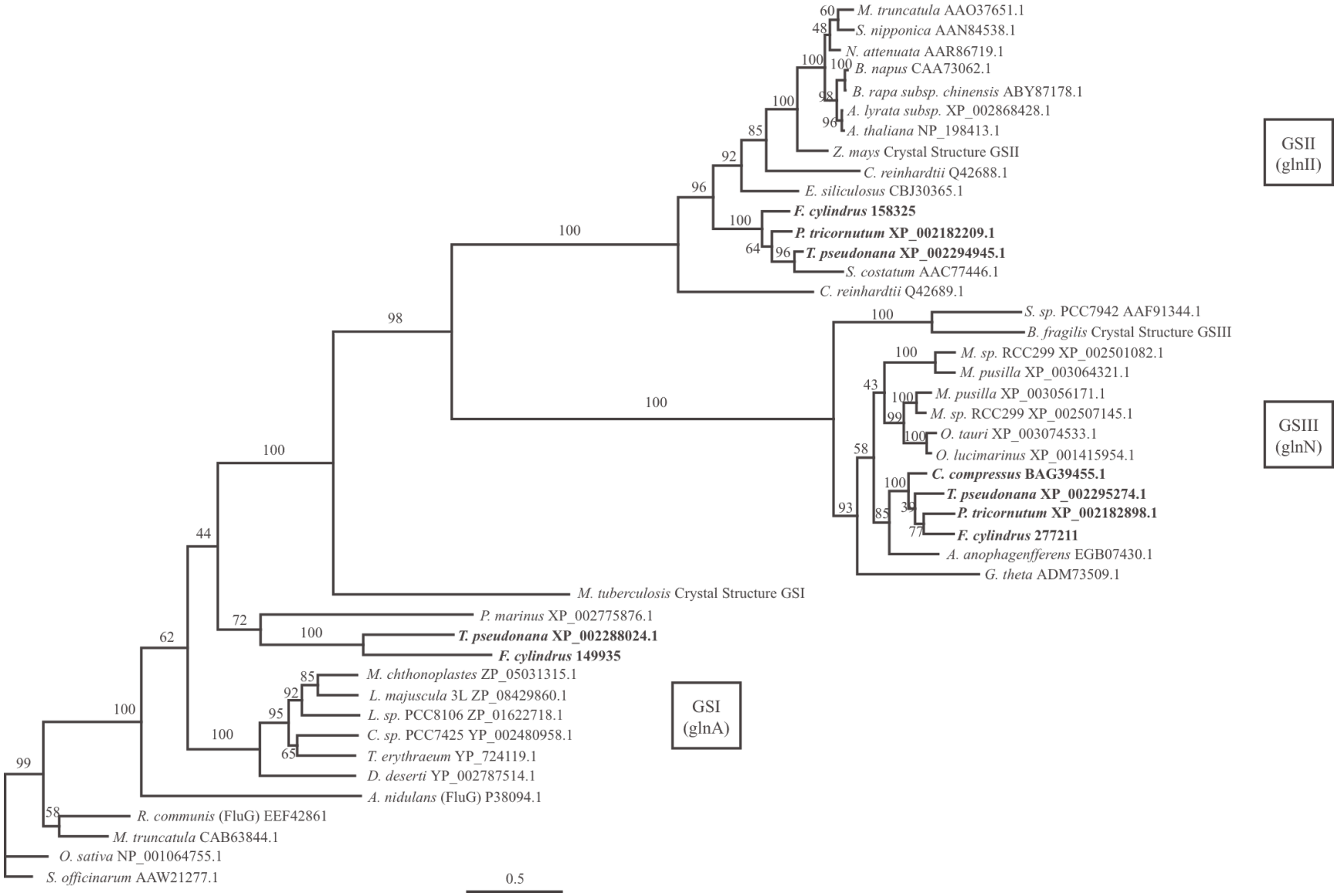
A



B

5' 2003035bp GCATCGTCCACCACTGTAACAAAGTATCCTAAATCAGCCGCGTCAC
GAACTGCGCTCTCGACGCACTGATCCGTCAATTGACCGCAAACAACCAATTGC
TCAACGTTGAGATTGCGTAGAATGTAATCCAAGTTGGTAGATTGAAAGACGGA
GCAAGATCTTTTGGGTAGACAAATATCCTTTCCTTGAATTGGAGCA 3' 2002838bp

Supplemental Figure 1.1 RNAseq read coverage for *GSI* (A) and *GSI* region sequenced in this study (B). A- Coverage of SOLiD reads from transcriptomes of *T. pseudonana* aligned to the *GSI* region of the *T. pseudonana* genome on a log-scale, y-axis: base pair position on chromosome 2, x-axis. Note that strand shown is the complimentary strand. Conserved domains (Conserved Domain Database; http://www.ncbi.nlm.nih.gov/Structure/cdd) enclosed by boxes for GSI (glnA) and isochorismatase. Original *T. pseudonana* gene model for GSI (version 3; thaps1_ua_kg.chr_2000355) spanning gray, shaded area. Solid, black box shows area of cDNA sequence from this study. B- Location of sequenced region given from 5' (base pair position 2003035) to 3' (2002838 base pair position) of the complimentary strand.


M. truncatula AAO37651.1
S. nipponica AAN84538.1
N. attenuata AAR86719.1
B. napus CAA73062.1
B. rapa subsp. chinensis ABY87178.1
A. lyrata subsp. XP_002868428.1
A. thaliana NP_198413.1
Z. mays Crystal Structure GSII
C. reinhardtii Q42688.1
E. siliculosus CBJ30365.1
F. cylindrus 158325
P. tricornutum XP_002182209.1
T. pseudonana XP_002294945.1
S. costatum AAC77446.1
C. reinhardtii Q42689.1
GSII (glnII)
S. sp. PCC7942 AAF91344.1
B. fragilis Crystal Structure GSIII
M. sp. RCC299 XP_002501082.1
M. pusilla XP_003064321.1
M. pusilla XP_003056171.1
M. sp. RCC299 XP_002507145.1
O. tauri XP_003074533.1
O. lucimarinus XP_001415954.1
C. compressus BAG39455.1
T. pseudonana XP_002295274.1
P. tricornutum XP_002182898.1
F. cylindrus 277211
A. anophagenfferens EGB07430.1
G. theta ADM73509.1
GSIII (glnN)
M. tuberculosis Crystal Structure GSI
P. marinus XP_002775876.1
T. pseudonana XP_002288024.1
F. cylindrus 149935
M. chthonoplastes ZP_05031315.1
L. majuscula 3L ZP_08429860.1
L. sp. PCC8106 ZP_01622718.1
C. sp. PCC7425 YP_002480958.1
T. erythraeum YP_724119.1
D. deserti YP_002787514.1
A. nidulans (FluG) P38094.1
GSI (glnA)
R. communis (FluG) EEF42861
M. truncatula CAB63844.1
O. sativa NP_001064755.1
S. officinarum AAW21277.1
0.5

Supplemental Figure 1.2 Maximum likelihood phylogenetic tree showing relatedness of the three GS gene isoforms in *T. pseudonana* and other organisms (see Methods). Organism names abbreviated in italics with accompanying accession number or protein id. For *R. communis* and *A. nidulans*, FluG protein sequence was used (indicated with parentheses). Full-length organism names listed below: *Saccharum officinarum*, *Oryza sativa*, *Medicago truncatula*, *Ricinus communis*, *Aspergillus nidulans*, *Deinococcus deserti*, *Trichodesmium erythraeum*, *Cyanothece sp.* PCC7425, *Lyngbya sp.* PCC 8106, *Lyngbya majuscula*, *Microcoleus chthonoplastes*, *Fragilariopsis cylindrus*, *Thalassiosira pseudonana*, *Prochlorococcus marinus*, *Mycobacterium tuberculosis*, *Guillardia theta*, *Aureococcus anophagefferens*, *Phaeodactylum tricornutum*, *Chaetoceros compressus*, *Ostreococcus lucimarinus*, *Ostreococcus tauri*, *Micromonas sp.* RCC299, *Micromonas pusilla*, *Bacteroides fragilis*, *Synechococcus* PCC7942, *Chlamydomonas reinhardtii*, *Skeletonema costatum*, *Ectocarpus siliculosus*, *Zea mays*, *Arabidopsis thaliana*, *Arabidopsis lyrata subspecies*, *Brassica rapa subspecies chinensis*, *Brassica napus*, *Nicotiana attenuata*, and *Stellaria nipponica*.

Chapter 2

Transcriptional responses of three evolutionarily diverse diatoms to nitrate starvation

2.1 Abstract

Diatoms are the most diverse group of phytoplankton in the ocean, yet despite their widely recognized influence on ocean ecosystems and global biogeochemistry, little is known about how this diversity affects their role in these large-scale processes. Here, we examined the transcriptional response of three diatoms – *Thalassiosira pseudonana*, *Fragilariopsis cylindrus*, and *Pseudo-nitzschia multiseries* – to the onset of nitrate starvation. The three species shared 5,583 clusters of orthologous genes based on OrthoMCL clustering, representing 30–54% of the predicted genes in each diatom. Less than 5% of genes within these shared clusters displayed the same transcriptional responses to the onset of nitrate starvation across the three diatoms, and orthologs involved in nitrogen uptake and assimilation were expressed differently across the three species. A majority of the genes shared between only the two pennate diatoms, as well as the non-orthologous genes in each species had minimal annotation information, but were often significantly differentially expressed under nitrate starvation, indicating their importance in diatom cellular metabolism. Although the specific transcriptional response to nitrate starvation varied among diatoms, changes in expression across the annotated gene transcripts suggested that each diatom shifted resource partitioning between carbon and nitrogen metabolism, a general physiological response that is shared across all diatoms. Apparent functional similarities between the pennate diatoms (e.g. greater expression of carbon fixation/carbohydrate metabolism genes) relative to the bipolar centric diatom (e.g. more highly expressed cell repair genes) during nitrate starvation also identified potential metabolic differences between the pennate and centric groups. Overall, each species exhibited a unique transcriptional response, highlighting the

diverse metabolic capabilities of *T. pseudonana*, *F. cylindrus* and *P. multiseries*. From these findings, we hypothesize that each diatom possesses a different transcriptional strategy for achieving a similar metabolic endpoint. Characterization of these finely tuned responses will enable better prediction of when diatoms will bloom in the marine environment, and what species will bloom under a given set of environmental factors.

2.2 Key Words

Diatoms; transcriptomics; nitrate starvation; orthologous genes

2.3 Introduction

Diatoms are a globally influential group of marine phytoplankton with an estimated 200,000 species on Earth (Kooistra et al. 2007). They first appeared in the fossil record about 190 million years ago (Sims et al. 2006) and subsequently diverged into four main groups – the radial and bi/mulitpolar centric diatoms and the more recently diverged raphid and araphid pennate diatoms (Kooistra et al. 2007). Globally, diatoms contribute an estimated 40% to marine primary production and are a major sink of fixed carbon out of marine surface waters (Kooistra et al. 2007; Dugdale & Wilkerson 1998; Falkowski et al. 2004). In coastal waters, diatoms constitute approximately 75% of primary production, forming the base of highly productive food webs (Sims et al. 2006; Nelson et al. 1995; Smetacek 1998). Diatom production, fueled by seasonal and episodic upwelling events, leads to extensive organic matter export in coastal regions: the growth and decline of diatom blooms provide an important link between the carbon and nitrogen cycles (Kooistra et al. 2007; Dugdale & Wilkerson 1998; Martin et al. 2011; Wilkerson et al. 2000). The continued success of diatoms is attributed to their ability to quickly

acclimate to changes in their environment and outcompete other phytoplankton for important nitrogen resources (Katz et al. 2004; Koester et al. 2012; Lomas & Glibert 2000; Tozzi et al. 2004).

Nitrogen uptake and assimilation in diatoms is directly connected to their carbon metabolism. Nitrogen assimilation requires ATP and reductants derived from photosynthesis (Falkowski & Stone 1975) and carbon skeletons derived from glycolysis (Behrenfeld et al. 2008; Hockin et al. 2012; Turpin 1991). When diatoms are starved for nitrate, protein levels and cellular nitrogen concentrations drop and carbon to nitrogen ratios increase (Claquin et al. 2002; Harrison et al. 1990; LaRoche et al. 1993), triggering increased intracellular nitrogen recycling and storage of carbon-rich compounds, such as lipids (Allen et al. 2011; Hockin et al. 2012; Palmucci et al. 2011). Ultimately, a decrease in chlorophyll pigments results in cellular chlorosis (Kolber et al. 1988) and leads to a reduction in photosynthetic capacity, which in turn negatively impacts energy conversion, carbon fixation, and cell growth (Behrenfeld et al. 2008).

Insights into the molecular responses of diatoms to nitrogen starvation come largely from whole genome comparisons of the model diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*; both organisms have publicly available whole genome sequences (Armbrust et al. 2004; Bowler et al. 2008). A suite of similarities and differences in carbon and nitrogen metabolism have been detected between *T. pseudonana* and *P. tricornutum*. Differences include the predicted localization and number of isoforms of carbonic anhydrases and the localization and number of bicarbonate transporters, both of which are required for the carbon-concentrating mechanisms that enhance delivery of carbon dioxide to the carbon fixation enzyme, ribulose bis-phosphate carboxylase (Kroth et al. 2008). There are also differences in the presence and localization of some key genes involved in glycolysis and gluconeogenesis, which may have

downstream implications for diatom nitrogen metabolism (Montsant et al. 2005; Smith et al. 2012). The increased availability of energy and carbon stores from glycolysis and the TCA cycle is hypothesized to fuel the urea cycle with the subsequent production of compounds containing recycled nitrogen, such as urea and arginine; many of the key enzymes involved in connecting these pathways to one another and shuttling substrates through the cell have yet to be identified (Allen et al. 2006; 2011; Chapter 1, Bender et al. 2012; Hockin et al. 2012; Smith et al. 2012). Comparative genomic approaches have highlighted the diverse genetic potential among the diatoms and hinted at differing adaptive strategies for survival within the group (Smith et al. 2012).

Comparisons of transcriptional studies among diatoms have also indicated that the underlying mechanisms of the nitrate starvation response may vary among species. A relatively small proportion of orthologs are differentially transcribed by both *P. tricornutum* and *T. pseudonana* in response to nitrate starvation (Maheswari et al. 2010). This subset includes genes with unknown functions, nitrogen metabolism genes, and genes involved in cell regulatory functions (Maheswari et al. 2010). In both diatoms, transcript abundances are lower for genes involved in nitrogen assimilation (e.g. nitrate/nitrite reductase) and for the first step of the urea cycle (e.g. carbamoyl phosphate synthetase) when the cells are starved for nitrogen compared to nutrient-replete growth (Allen et al. 2011; Ashworth et al. 2013; Hockin et al. 2012; Mock et al. 2008). Transcripts for nitrate transporters are expressed under both nitrate-replete and nitrate-starved growth in both diatoms, whereas transcripts ammonium transporters transcripts increase under nitrate-starved growth in the absence of ammonium only in *P. tricornutum* (Ashworth et al. 2013; Maheswari et al. 2010). Differences in expression of urea cycle genes, as well as the gene that encodes for urease, have also been observed between the two diatoms; urease gene

expression is higher in *T. pseudonana* and lower in *P. tricornutum* under nitrate starvation versus nutrient-replete conditions (Allen et al. 2011; Hockin et al. 2012). To date, no study has compared transcriptomes of diatoms grown under the same controlled, nitrate-starved laboratory conditions.

We conducted a comparative transcriptional analysis of the bipolar centric diatom *T. pseudonana* and two raphid pennate diatoms, *F. cylindrus* and *Pseudo-nitzschia multiseries*, under nitrate starvation. These species were chosen because they are frequently identified in environmental samples, each diatom has a sequenced genome, and the bipolar centric diatom is evolutionarily distant from the two raphid pennate diatoms. Our goal was to identify those aspects of the starvation response that are uniform across diatoms and those aspects that appear to be exclusive to each diatom species. Findings from this study have implications for how diatom metabolism is modeled in the marine environment and how we classify and generalize the metabolic response of the most diverse group of marine phytoplankton.

2.4 Methods

Culture acclimation

Thalassiosira pseudonana (Hustedt) Hasle et Heimdal (Provasoli-Guillard National Center for Marine Algae and Microbiota, CCMP 1335), *Fragilariopsis cylindrus* (Provasoli-Guillard National Center for Marine Algae and Microbiota, CCMP 1102), and *Pseudo-nitzschia multiseries* (S. Bates, Fisheries and Oceans Canada; CLN 17) were maintained without bubbling in semi-continuous batch cultures under continuous light (100 μmol photons m^{-2} s^{-1}) in modified artificial seawater (Berges et al. 2001) with f/2 concentrations of nutrients (Guillard & Ryther 1962). Growth of non-axenic triplicate cultures of *T. pseudonana* at 20°C, *F. cylindrus* at 4°C

and *P. multiseries* at 13°C were monitored with a 10-AU fluorometer (Turner). Cultures were considered acclimated to the growth conditions when the growth rates of three consecutive transfers were not significantly different from one another (ANCOVA, Brand et al. 1981).

Growth experiments

Growth experiments for each diatom were conducted in 10 L (*T. pseudonana* and *F. cylindrus*) or 4 L (*P. multiseries*) of artificial seawater media (Berges et al. 2001); all cultures were bubbled with sterile filtered air. *Pseudo-nitzschia multiseries* was also stirred on a stir plate. As in Durkin et al. (2012), each diatom was grown in triplicate batch cultures with nutrient-replete media (882 μM $NaNO_3$, 106 μM Na_2SiO_3, 36.2 μM NaH_2PO_4) or low nitrate media (55 μM $NaNO_3$, 212 μM Na_2SiO_3, 72.4 μM NaH_2PO_4). Photosynthetic yield of photosystem II (F_v/F_m) was monitored with a PhytoPAM fluorometer (Waltz). The onset of stationary phase due to nitrate starvation was determined when chlorophyll *a* fluorescence of a culture no longer increased exponentially and F_v/F_m values decreased. Particulate organic carbon (POC) and particulate organic nitrogen (PON) were collected by filtering 50–200 mL of culture onto pre-combusted (450°C for 4.5 h) 25 mm Whatman GF/F filters. Samples were analyzed for POC and PON on an elemental analyzer (Elementar) at the University of California Davis Stable Isotope Facility.

Between 4.5 and 1.5 L (half of the original culture volume) of the experimental cultures were filtered onto 0.8 μm polycarbonate filters (Millipore) at the onset of stationary phase (nitrate starvation) or during mid-exponential growth (nutrient-replete). Filtered cells were immediately flash frozen in liquid nitrogen and stored at −80°C until later RNA extraction (see below). Approximately 50 mL of sample were syringe-filtered (0.2 μm) into Falcon tubes at the onset of nitrate starvation or during mid-exponential growth for each replicate and stored at

−20°C for later quantification of dissolved nutrients. Dissolved nutrient concentrations in *P. multiseries* nutrient-replete treatments were taken from a separate but identical, set of experiments conducted one week prior. All nutrients were analyzed at the University of Washington Marine Chemistry Laboratory on a Technicon AutoAnalyser II following the methods of UNESCO (1994). Depletion of nitrate was confirmed by removing two 50 mL aliquots from the remaining nitrate-starved cultures and adding 1764 μM $NaNO_3$ to aliquot one, no nutrients to aliquot two, and 1764 μM $NaNO_3$ to the remaining nitrate-starved large-volume cultures. All samples were monitored for at least two more days to confirm that chlorophyll *a* fluorescence increased after the addition of nitrate to aliquot one and the large-volume cultures. Significant differences in growth rate, F_v/F_m, and POC/PON were determined using paired T-tests with a Bonferroni correction for multiple comparisons.

RNA extraction, SOLiD library prep, and SOLiD sequencing

RNA was extracted from frozen cells on filters using the Totally RNA extraction kit (Invitrogen). The RNA was incubated with DNAse I (Ambion) at 37°C for 2 h and purified by DNAse inactivation reagent (Ambion). RNA from the control and nitrate-starved treatments was fragmented enzymatically to an average size of 100–200 bp with RNAse III, according to the SOLiD Total RNA-Seq Kit protocol. SOLiD-specific adaptors were ligated to mRNA ends and the mRNA was reverse transcribed; the resulting cDNA was run on a Novex 6% TBE-Urea gel 1.0mM (Invitrogen). cDNA of 150–250 bp was excised from a gel and PCR-amplified for 15 cycles following the methods in the SOLiD Total RNA-Seq Kit protocol. Amplified cDNA was purified and concentrated using the PureLink PCR Micro Kit (Invitrogen). Each sample was quantified by qPCR using the SOLiD specific TaqMan assay before dilution to 250 pM for emulsion PCR (completed on the EZ Bead System from Life Technologies). For each sample,

700 million beads were loaded onto a slide and run on an Applied Biosystems (AB) SOLiD sequencer version 4. Predicted read length was 50 base pairs.

Sequence read cleanup and alignments

Sequence reads were trimmed based on quality score using trimfastq (SEAStAR, http://armbrustlab.ocean.washington.edu/seastar). The appropriate read error probability cutoff ("-p") was determined based on a subsample of reads trimmed at 10 equally spaced cutoff values. The cutoff value that yielded the highest number of bases aligned was chosen. Trimmed reads with a length shorter than 30 colorspace transitions ("-l") or an entropy value less than 3.0 ("-e") were discarded. Trimmed reads were aligned with BWA (v.0.5.9) to gene models for *Thalassiosira pseudonana* (http://genome.jgi.doe.gov/Thaps3/Thaps3.home.html), *Fragilariopsis cylindrus* (http://genome.jgi.doe.gov/Fracy1/Fracy1.home.html) or *Pseudo-nitzschia multiseries* (http://genome.jgi.doe.gov/Psemu1/Psemu1.home.html). Parameters for "bwa aln" were "-n .001 -c -l 18 -k 2", and parameters for "bwa samse" were "-n 500000" (Heng Li & Durbin 2009). The number of sequence reads that aligned to gene models were calculated from the resulting SAM alignment files using ref_select as part of the SEAStAR package.

Differential expression analysis, annotations and OrthoMCL gene clustering

Orthologous genes were clustered using the Identification of Ortholog Groups for Eukaryotic Genomes (OrthoMCL) software and default settings (Li et al. 2003). Potential function of predicted proteins was based on a translated query sequence search (BLASTX v2.2.27+, e-value 1 x 10^{-5}) to proteins within the Kyoto Encyclopedia of Genes and Genomes (KEGG) 2012-05-09 database (Altschul et al. 1990; Kanehisa et al. 2012; Kanehisa & Goto 2000). The KEGG module annotation for the protein with the lowest e-value was assigned to all members of a cluster. The KEGG Orthology (KO) definition and KEGG pathways associated

with the best module annotation were also documented. In those instances where KEGG annotations differed between orthologs within a cluster, the annotations were categorized as "multiple pathways." KEGG annotations were hierarchically clustered using the heatmap.2 function in R on default settings to calculate the distance matrix (R gplots package). Transcribed genes that clustered within only one diatom or between the two pennate diatoms (and not between one pennate diatom and *T. pseudonana*) were translated and compared to a protein database of *P. tricornutum* (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html), and 21 EST databases (Moore Foundation Marine Microbiology Initiative-supported Marine Microbial Eukaryote Transcriptome Sequencing Project, National Center for Genome Resources) using BLASTX with an e-value cutoff of 1×10^{-5} (supplemental Table 2.1).

Transcriptomes from biological triplicates (N = 3) for each condition and species were analyzed with the R software package edgeR from Bioconductor (Robinson et al. 2010). An estimate of common dispersion across the three biological replicates within each treatment was calculated to account for differences in sequence reads mapped per gene (gene counts) relative to total number of reads in a sample (library size) (Robinson & Smyth 2007). From here, pairwise comparisons of each gene across treatments were made with the exactTest function. A false discovery rate (fdr) of 0.01 was used to account for Type I errors when testing for significant differences across treatments; all p values ($p < 0.01$) provided have been fdr-corrected. Up-regulated genes exhibit greater transcript abundances in response to nitrate starvation, and down-regulated genes exhibit greater transcript abundances under nutrient-replete growth.

2.5 Results

Orthologous and non-orthologous gene cluster comparisons

A total of 49,170 genes are predicted to be encoded by the three diatom genomes (*T. pseudonana*, 11,393 predicted genes; *F. cylindrus*, 18,077 predicted genes; *P. multiseries*, 19,703 predicted genes). Of this total, 18,077 (37%) orthologous genes were shared amongst the three diatoms and grouped into 5,583 clusters (Figure 2.1). This core set of genes represented 54% (6,098 genes) of the *T. pseudonana* genome, 33% (5,982 genes) of the *F. cylindrus* genome, and 30% (5,997 genes) of the *P. multiseries* genome. Most core orthologous gene clusters contained only one gene copy per diatom; a minority of clusters were composed of multi-copy gene families (supplemental Figure 2.1A). The two pennate diatoms, *F. cylindrus* and *P. multiseries*, shared 3,839 clusters (4,639 and 5,105 genes, respectively) that had no orthologs in the *T. pseudonana* genome; a minority of these clusters contained more than one ortholog representative per diatom (Figure 2.1; supplemental Figure 2.1B). Each of these genes were present in at least one bipolar centric diatom other than *T. pseudonana* (supplemental Table 2.1) suggesting they were present in the pennate bipolar ancestor and were lost during the emergence of *T. pseudonana*.

Forty percent of the genes in each diatom did not cluster with genes from the two other diatoms and were defined as non-orthologous, although all of these genes displayed sequence similarities (e-value 1 x 10^{-5}) to transcripts from other diatoms (supplemental Table 2.1). A majority of these genes were present as single copies, with a few clusters that could consist of several tens of genes in *T. pseudonana* and *F. cylindrus* and several hundreds of genes in *P. multiseries* (supplemental Figure 2.1C). About 1,600 of the *T. pseudonana* non-orthologous genes were not found in the eight queried datasets of pennate diatoms (supplemental Table 2.1) suggesting that these genes are of centric diatom origin, and were lost during the emergence of pennate lineages. One non-orthologous gene (Tp1457), identified in *T. pseudonana* and encoding

a hypothetical protein, was present in all of the queried centric diatoms but in none of the pennate diatoms.

Transcriptional patterns under nitrate starvation

To identify the transcriptional response to nitrogen starvation, the three diatom species were either maintained under optimal growth conditions or allowed to grow until nitrate was depleted. Exponential growth rates under nutrient-replete and optimal temperature conditions were 0.87 ± 0.19 day^{-1} for *T. pseudonana*, 0.34 ± 0.04 day^{-1} for *F. cylindrus*, and 0.60 ± 0.12 day^{-1} for *P. multiseries* (supplemental Table 2.2). The photochemical yield of photosystem II (F_v/F_m) was 0.68 for *T. pseudonana* and *P. multiseries* and 0.52 for *F. cylindrus* at time of harvest under replete growth. When cells were nitrate-starved, F_v/F_m at time of harvest ranged between 0.47 and 0.40, depending on the diatom, and F_v/F_m was significantly different between the *F. cylindrus* replete and nitrate-starved cultures. The ratio of particulate organic carbon to nitrogen (POC/PON) was lowest in the *F. cylindrus* nitrate-starved cultures, and was highest in the *P. multiseries* nitrate-starved cultures; differences in POC/PON were also observed between replete growth and nitrate-starved growth for each diatom species (supplemental Table 2.2). Nitrate concentrations in the nitrate-starved cultures were 2–3 orders of magnitude less than the nitrate concentrations of the nutrient-replete cultures and an order of magnitude less than the phosphate and silicic acid concentrations of either the nutrient-replete or nitrate-starved cultures (supplemental Table 2.2). When nitrate was added back to the nitrate-starved cultures, F_v/F_m increased to near nutrient-replete values (data not shown).

SOLiD-sequenced transcriptomes were generated from triplicate nutrient-replete controls and triplicate nitrate-starved treatment cultures of *T. pseudonana*, *F. cylindrus* and *P. multiseries*. Quality-filtered read counts ranged from 5 x 10^7 (*F. cylindrus*) to 1.5 x 10^8 (*P. multiseries*) reads

per sample; 20–80% of the filtered reads (for a given replicate) aligned to the model genes. The remaining reads aligned to unmodeled regions of each organism's genome. The lowest number of aligned reads was from one *F. cylindrus* nutrient-replete treatment (1.0×10^7 reads). Compared to the total number of predicted modeled genes, the vast majority of genes in each diatom were transcribed under nutrient-replete or nitrate-starved conditions (Table 2.1). About half the genes that were differentially expressed under nitrate starvation were in the orthologous core group (1,016–2,106 genes) (Table 2.1; supplemental Table 2.3), with 1.5–2.5 times more of these genes down-regulated than up-regulated (Table 2.1; Figure 2.2A). *Thalassiosira pseudonana* had the greatest range of differential expression followed by *P. multiseries*, whereas differential expression of genes encoded by *F. cylindrus* fell within a smaller range of values (Figure 2.2A). Core genes with significant differential expression in one diatom did not necessarily exhibit the same differential expression patterns in the other diatoms, contributing to the observed variability in expression.

About a third of the differentially expressed orthologous core genes had a KEGG module annotation (Table 2.1). When the presence and abundance of KEGG annotations were compared across the normalized diatom transcriptomes, three groups emerged. One group contained KEGG annotations for transcripts more abundant under nitrate starvation in all three diatoms. The other two groups contained KEGG annotations for transcripts more abundant under nutrient-replete conditions: one consisted of *T. pseudonana* and *P. multiseries* annotations (Figure 2.2B) and the other consisted of *F. cylindrus* annotations. Carbon-related genes (e.g. central carbohydrate metabolism and carbon fixation) constituted the largest portion of annotations in the nitrate-starved subset; these genes were also abundant in the nutrient-replete subset. Proportionally, a larger increase in nitrogen-related KEGG annotations (e.g. nitrogen metabolism) between the

nutrient-replete and nitrate-starved subsets was observed, compared to changes in carbon-related annotations between the two subsets. Additional KEGG annotations in the up-regulated nitrate-starved group were for RNA and protein processing, amino acid metabolism, and transporters, all with potential links to cell repair and nitrogen recycling. The most abundant KEGG annotations in the down-regulated nutrient-replete groups were for carbohydrate metabolism, fatty acid metabolism, carbon fixation, and cofactor and vitamin biosynthesis. Overall, there were 30–50% more KEGG annotations associated with nutrient-replete growth.

A small subset (9–19%) of differentially expressed core genes shared the same differential expression patterns across all three diatoms, such that all genes within a given cluster were either up- or down-regulated in response to nitrate starvation (Figure 2.3). We define the expression patterns of these genes as the shared core response to nitrate starvation (Table 2.1). Genes within twenty-four orthologous clusters were all up-regulated. Eleven clusters had an associated KEGG module annotation that suggested involvement in the formation, degradation and transport of nitrogen-rich compounds and included a uroporphyrin methyltransferase connected to porphyrin and chlorophyll metabolism, an ABC transporter, and genes involved in serine biosynthesis, leucine degradation, cell repair and protein processing (supplemental Table 2.3). Genes within 160 orthologous clusters were all down-regulated. Sixty-three clusters had a KEGG annotation that suggested involvement in porphyrin and chlorophyll metabolism, glycolysis, fatty acid biosynthesis, terpenoid backbone biosynthesis, the pentose phosphate pathway, photosynthesis and carbon fixation. The largest gene cluster in the down-regulated core subset contained six ferredoxin genes required for electron transport. These results indicated a general enhancement of transcription of genes involved in nitrogen processing and a reduction of

transcription of genes involved in carbon metabolism and electron transport in response to nitrate starvation.

A subset of clusters was differentially expressed under nitrate starvation and not under silicic acid starvation (data not shown). This subset was classified as the "nitrate only" response and constituted 52–90% of core genes that were significantly differentially expressed in each species (*T. pseudonana*, 1,102 genes; *F. cylindrus*, 915 genes; *P. multiseries*, 1,047) (Table 2.1). The transcriptional patterns of this subset mirrored what was seen with the overall core response: an up-regulation of genes involved in nitrogen metabolism and a down-regulation of genes involved in carbon metabolism. Four genes identified as nitrate-only in the *T. pseudonana*, *F. cylindrus*, and *P. multiseries* transcriptomes were also identified as nitrate-only in a separate study by (Mock et al. 2008), who included in their comparison iron limitation, low temperature and high pH conditions. Three of the gene clusters encode hypothetical proteins with no obvious annotations (Tp268343, Fc164230, Pm303116; Tp4888, Fc206263, Pm284060; Tp6551, Fc227588, Pm315644); the fourth gene cluster is connected to carotenoid biosynthesis (Tp264039, Fc169705, Pm242952).

Expression of nitrogen metabolism genes

We examined in more detail the transcription patterns of genes known to be required for nitrogen uptake and assimilation but not present within the core shared response to nitrogen starvation. The first step in nitrogen assimilation is the transport of exogenous nitrogen (e.g. nitrate, ammonia, or urea) into the cell or reduced nitrite into the plastid for further reduction. There are multiple gene copies of each type of transporter gene (except for the nitrite transporter) (Figure 2.4). Each diatom encodes four to six ammonium transporters, represented by three orthologous clusters. *T. pseudonana* down-regulated transcription of three of the six ammonium

transporter genes in response to nitrate starvation, whereas *F. cylindrus* and *P. multiseries* up-regulated transcription of three (of five) and two (of four) genes, respectively, in response to nitrate starvation ($p < 0.05$; Figure 2.4). Nitrate-starved cultures of *T. pseudonana* and *F. cylindrus* up-regulated one (of three) and two (of four) urea transporter genes, respectively; these genes were not differentially regulated by nitrate-starved cultures of *P. multiseries*. *Thalassiosira pseudonana* and *F. cylindrus* each encode three, and *P. multiseries* encodes four nitrate transporters that group into three orthologous clusters. The nitrate transporter genes all up-regulated in response to nitrate starvation, except for one down-regulated copy in *P. multiseries*. The nitrate reductase gene in *T. pseudonana* (one copy) and *P. multiseries* (one copy) were both up-regulated, whereas the two *F. cylindrus* copies did not exhibit significant differential expression. The reduced nitrite is transported to the plastid via a single nitrite transporter. Only the *F. cylindrus* transporter gene was down-regulated in response to nitrate starvation. Each diatom possesses two groups of nitrite reductases: Fe-dependent and NAD(P)H-dependent isoforms. Both types of nitrite reductase genes encoded by *T. pseudonana* and *P. multiseries* were up-regulated in response to nitrate starvation and one (of three) copies encoded by *F. cylindrus* (Figure 2.4).

Once in a reduced form, nitrogen is assimilated into the cell through the GS-GOGAT (glutamine synthetase and glutamate synthetase) cycle. Each diatom has two copies of glutamate synthetase that form one cluster; at least one gene copy per diatom was up-regulated. The glutamine synthetase genes formed three orthologous gene clusters (one for each isoform: GSI, GSII and GSIII). None of genes within the GSI cluster encoded by the three diatoms were transcribed differently in response to nitrate starvation. The GSII genes were down-regulated by nitrate-starved *T. pseudonana* and *F. cylindrus* cells, and were not differentially regulated by

nitrate-starved *P. multiseries* cells. The GSIII gene encoded by *T. pseudonana* was down-regulated and the GSIII genes encoded by *F. cylindrus* and *P. multiseries* were up-regulated.

We also explored the possible role of the urea cycle and urea degradation in the diatom response to nitrate starvation. One cluster of carbamoyl phosphate synthetase (*CPS*) genes included both isoforms, one involved in pyrimidine synthesis (*pgCPSII*) and the other in the urea cycle (*unCPS*). The urea cycle *CPS* (*unCPS*) was down-regulated by nitrate-starved *T. pseudonana* (protein ID 40323) and *F. cylindrus* (protein ID 169332). Nitrate-starved *P. multiseries* up-regulated both *CPS* isoforms. In subsequent urea cycle steps, the only gene with differential expression (down-regulation) was the ornithine carbamoyltransferase gene in *T. pseudonana*. Mixed expression patterns were observed for urease (*URE*) and its accessory genes (*UREf* and *UREg*), which formed three separate orthologous gene clusters.

Expression patterns of pennate-clustered genes

We examined transcriptional patterns of the genes shared between only the pennate diatoms to further explore potential group-specific responses to nitrate starvation. About 16–30% of the 3,839 gene clusters shared between the pennate diatoms were significantly differentially expressed under nitrate starvation (Table 2.2). Those with the greatest differential expression had no KEGG module annotations (Figure 2.5A). When the small fraction of predicted proteins with a KEGG module annotation were clustered hierarchically, the KEGG annotations clustered into a nitrate-starved (up-regulated) and a nutrient-replete (down-regulated) group for both diatoms (Figure 2.5B). The nitrate-starved subset contained more annotations associated with two-component regulatory system and RNA processing/polymerase, as well as nitrogen metabolism (e.g. lysine metabolism, polyamine biosynthesis, cysteine, and methionine metabolism). The nutrient-replete subset contained more annotations associated with central carbohydrate

metabolism, carbon fixation and terpenoid biosynthesis, as well as fatty acid metabolism and amino acid metabolism. About 160 genes in both diatoms displayed a shared response to nitrate starvation (all up- or down-regulated in a given cluster). Nearly half of the up-regulated genes (N = 7 genes) with an annotation were involved in nucleotide and amino acid metabolism (e.g. polyamine synthesis, lysine metabolism) (Figure 2.5C). In the down-regulated fraction, thirteen out of nineteen genes with KEGG annotations were associated with carbohydrate and lipid metabolism. This subset included genes connected to glycolysis, gluconeogenesis, the Calvin Cycle, and isoprenoid biosynthesis. The largest difference in expression of shared pennate genes between *F. cylindrus* and *P. multiseries* was observed in the down-regulated subset. The majority of pennate-clustered genes were differentially transcribed under nitrate starvation only (data not shown).

Expression patterns of non-orthologous genes

Forty-two percent (1,981 genes; *T. pseudonana*), 17% (1,178 genes; *F. cylindrus*), and 18% (1,542 genes; *P. multiseries*) of the non-orthologous genes were significantly differentially expressed (Figure 2.6A). Approximately half these genes were categorized as "nitrate only" because they were not significantly differentially expressed under silicic acid starvation (Table 2.3). Additionally, tens of genes in each diatom were transcribed only under nitrate starvation, with no detected transcription under nutrient-replete conditions (Figure 2.6A). Eleven nitrate only genes from the *T. pseudonana* transcriptomes were also identified as nitrate-only by (Mock et al. 2008). Nine of the 11 genes had minimal annotation information (Tp 10945, 22640, 22671, 24923, 3614, 6048, 7491, 9958, 4002); the other two genes were involved in two-component sensing (osmoregulation; Tp 269238), and fatty acid biosynthesis (Tp 21299), respectively. Few of the non-orthologous genes (7%, *T. pseudonana*; 5%, *F. cylindrus*; 3%, *P. multiseries*) had an

assigned KEGG module annotation. None of the most highly transcribed genes under either condition had KEGG annotations (Figure 2.6A).

When the presence and abundance of KEGG annotations of non-orthologous genes was compared across the normalized diatom transcriptomes, the pennate diatoms (*F. cylindrus* and *P. multiseries*) clustered separately from *T. pseudonana* (Figure 2.6B). Hierarchical clustering was driven by the low number of annotated genes. The up-regulated nitrate-starved genes displayed the greatest annotation differences among the three diatoms. In *T. pseudonana*, these genes were involved in ubiquitin system, cell repair, cysteine and methionine metabolism, and polyamine biosynthesis; a spermidine synthase required for the formation of nitrogen-rich polyamines was up-regulated 1.6-fold. Most of the nitrate-starved genes in *F. cylindrus* were involved in carbon fixation, fatty acid metabolism, and central carbohydrate metabolism. *Pseudo-nitzschia multiseries* had the fewest KEGG module annotations in both subsets. Nitrate-starved genes included a 2-oxoglutarate dehydrogenase (TCA cycle), a possible ferredoxin and an RNA polymerase, and the largest group of nitrate-starved, annotated genes were involved in RNA processing. All three diatoms had greater transcript abundances for several genes connected to the TCA cycle under nitrate-starved growth. Among the down-regulated nutrient-replete genes, comparable expression patterns to the orthologous genes involved in carbon metabolism (glycolysis, lipid metabolism, and the pentose phosphate pathway) were observed.

Overall, transcription patterns differed between the centric and the pennate non-orthologous gene subsets: *T. pseudonana* differentially transcribed a greater number of genes involved in lipid metabolism, two-component regulatory systems, and general cellular repair and information processing (e.g. RNA processing, spliceosome), whereas the pennate diatoms differentially transcribed a greater number of genes involved in carbon fixation and carbohydrate

metabolism. These centric versus pennate functional differences were similar, but not identical to the orthologous core response (Figure 2.2B). The response to nitrate starvation by *T. pseudonana* involved more orthologous genes connected to fatty acid and lipid metabolism and two-component regulatory systems, whereas the response by the pennate diatoms involved more nitrogen-related gene representatives (e.g. proteasome, lysine metabolism).

2.6 Discussion

Diatoms form massive blooms in coastal environments due to their superior ability to respond to changes in nitrogen availability. We compared the transcriptional responses of *T. pseudonana*, *F. cylindrus*, and *P. multiseries* at the onset of nitrate starvation; the experiment was designed to capture the response of cells at the end of a bloom scenario when nitrate is depleted from seawater, preventing further cell division. The three diatom species examined have many genes in common, yet the transcriptional snapshots of the three diatoms under nutrient-replete and nitrate-starved growth did not return a large number of orthologous genes with shared differential expression. Each diatom appears to maintain its own cell-wide transcriptional response to nitrogen availability. Yet similar functional responses among the three diatoms, created by many different individual genes, also indicate that diatoms may share a pathway-level response to changes in nitrogen availability. A shared functional response was especially evident between the two pennate diatoms compared to the bipolar centric. This study highlights fundamental differences in how individual diatom species and groups of diatoms have evolved to respond to their environment, and it provides insight into the mechanisms that allow each species to inhabit its own environmental niche that is distinct from all other diatoms.

The over-representation of carbon-related pathways in both the up- and down-regulated core, pennate and non-orthologous subsets may indicate that separate components of the same pathway are differentially regulated within and across diatom species. In previous work on the transcriptional and proteomic response of *T. pseudonana* to nitrate starvation, gene transcripts of the gene that encodes carbamoyl phosphate synthetase (first reaction of the urea cycle) were lower under nitrate starvation, yet protein concentrations for subsequent urea cycle enzymes were elevated under nitrate starvation (Hockin et al. 2012). A similar decoupling among urea cycle genes was observed in *P. tricornutum* (Allen et al. 2011). Differential regulation within a metabolic pathway may be the result of post-transcriptional and post-translational modification for select genes or proteins, respectively, or it may be attributed to a lag in how nitrogen and carbon are partitioned in the cell, with the first enzymatic reaction of a pathway responding to changes in nutrient availability earlier than downstream reactions (Hockin et al. 2012; Poulsen et al. 2006).

Despite all of the variability in transcriptional regulation, pathway-level changes were detected, and provide valuable insight into the physiological responses among these three species. We hypothesize that distantly related genes encode diverged proteins that nonetheless carry out similar metabolic functions in the diatoms. This hypothesis was first proposed by Thompson et al. (2011), who conducted comparative transcriptomics on two closely-related ecotypes of *Prochloroccocus* and observed similar functional responses, but different transcriptional responses. In our study, we observed an overall increase in potential nitrogen recycling pathways, and a general decrease in cellular metabolism, including a down-regulation of photosynthesis and carbon-related pathways, demonstrating the known connection between carbon and nitrogen metabolism in the diatoms (Turpin 1991). A number of up-regulated genes

were also connected to central carbohydrate metabolism, supporting the hypothesis that carbon skeletons supplied through the TCA cycle are used to re-assimilate nitrogen generated by protein degradation; this metabolic response allows the cell to shift cellular efforts towards the generation of essential nitrogen compounds to combat cellular stress (Claquin et al. 2002; Harrison et al. 1990; Hockin et al. 2012; LaRoche et al. 1993; Tolonen et al. 2006).

In response to nitrate starvation, the pennate diatoms differentially expressed more carbon metabolism genes (e.g. carbon fixation genes and central carbohydrate metabolism genes), whereas *T. pseudonana* metabolism was directed toward cell repair processes, lipid metabolism, and regulatory components. Differences in how carbon and nitrogen metabolism are regulated in the cell may reflect inherent physiological differences between the two groups, and one evolutionary mechanism for these physiological differences could be the genome-wide expansions or deletions of carbon and nitrogen metabolism-related gene families (Tolonen et al. 2006; Smith et al. 2012). Previous work has identified expanded gene families in the pennates. For example, two additional orthologs of the chloroplast-targeted fructose 1,6 bisphosphatase (FBP, connected to gluconeogenesis and the Calvin cycle) identified in *F. cylindrus* and in *P. tricornutum*, but not in *T. pseudonana*, are hypothesized to confer a cellular advantage by providing the pennate diatoms with more flexibility (via more genes) in how they can react to environmental triggers (Martens et al. 2008; Smith et al. 2012). The expansion or reduction of gene families may produce gene copies with diverse transcriptional regulation and cellular localization of protein products. The extent of the genetic flexibility created by gene families may affect where and when nutrients are moved through the cell, ultimately affecting the flow of carbon and nitrogen in the cell (Lommer et al. 2012; Smith et al. 2012).

The ammonium transporter family provided an intriguing example of an expanded gene family (relative to other transporter families) that also exhibited a conserved response between *F. cylindrus* and *P. multiseries*. The three *T. pseudonana* ammonium transporters with significant expression were down-regulated in response to nitrate starvation, as observed in a previous study (Ashworth et al. 2013). In contrast, the five pennate ammonium transporters with significant differential expression were all up-regulated in response to nitrate starvation. The ammonium transporters of pennate diatoms *Cylindrotheca fusiformis* and *Phaeodactylum tricornutum* also had the highest transcript levels during nitrate starvation (Hildebrand 2005; Maheswari et al. 2010), suggesting that this could be a conserved response among pennate diatoms. Diatoms encode at least twice as many ammonium transporters as urea and nitrate transporters (Allen 2005). The expansion of the ammonium transporter gene family, as well as potential transcriptional patterns that appear to follow diatom phylogeny, may provide insight into an underlying regulatory network that controls the cell-wide transcriptional responses observed in this study.

We hypothesize that the pennate diatoms transcribe their ammonium transporters as part of a larger regulatory response that differs from the bipolar centric diatom. Pennate-specific responses among carbon-related genes (e.g. the expanded gene family for fructose 1,6 bisphosphatase, FBP) indicate that the regulatory network likely encompasses more than just nitrogen-related genes (Smith et al. 2012). Compared to *T. pseudonana*, *F. cylindrus* and *P. multiseries* transcribed all FBP orthologs at a higher level during nutrient-replete growth. In FBP clusters where there was also a *T. pseudonana* gene ortholog, the pennates shared a transcriptional response that was different from *T. pseudonana*. Ashworth et al. (2013) demonstrated a cell-wide transition to nitrate starvation in *T. pseudonana* that involved co-

expression of hundreds of genes in response to an environmental stimuli. A similar nitrogen network proposed in *Prochlorococcus* involves the transcriptional regulator NtcA, which has been shown to regulate expression not only of key nitrogen genes, but also genes related to photosynthesis (Su et al. 2005; Tolonen et al. 2006). In the diatoms, this regulatory mechanism may coordinate multiple pathways (and multiple orthologs) in a different manner in the pennate diatoms, compared to *T. pseudonana*, when nitrogen becomes limiting in the marine environment. Coordinated expression patterns for the ammonium transporters and the FBP genes in the two pennate diatoms support this hypothesis, and warrant additional studies that investigate similarities and differences in transcriptional regulation between the two groups.

The majority of genes in the shared core response, as well as in the pennate group and within the non-orthologous group, did not have a KEGG module annotation and could not be assigned a cell function, and yet they appear play a critical role in how diatoms perceive and respond to their environment given their significant differential expression patterns. Several of these genes were highly expressed under nitrate starvation, but not during the nutrient-replete condition (and vice versa); for some genes, differential expression ranged up to ±10 fold. These genes with unknown functions may be involved in unidentified nitrogen-related pathways in diatoms; they may represent a set of genes specific to the diatom lineage that play a critical role in how the cells respond to nitrogen availability. As an example, diatom cells contain a vacuole for storing nitrate that can be used during periods of low environmental nitrogen, but the genes connected to the formation and maintenance of this vacuole have not been identified (Allen et al. 2006; Armbrust et al. 2004; Katz et al. 2004; Raven 1987). Additionally, species within the *Pseudo-nitzschia* genus produce the nitrogen-rich toxic compound domoic acid, but the genetic basis of its synthesis is unknown (Auro & Cochlan 2013). Given the ecological importance of the

nitrate storage vacuole and the human health concerns associated with domoic acid, targeted studies of these nitrogen-responsive genes with unknown function could lead to discoveries about diatom nitrogen metabolism of both social and ecological importance.

The evolutionary trajectory of diatoms, including their origin via secondary endosymbiosis, has given them complex patterns of inheritance (Armbrust 2009; Bowler 2010; Allen et al. 2006), leading to unique metabolic features (e.g. complete urea cycle; Armbrust et al. 2004). Our findings provide evidence that gene family expansions and losses influence the whole-genome transcriptional responses of each of these diatom species to nitrate starvation. We hypothesize that each diatom possesses a different transcriptional strategy for achieving a similar metabolic endpoint. These individualized cell-states are likely governed by regulatory networks that make the response of each diatom to its environment unique (Ashworth et al. 2013). Additional laboratory studies on representative diatom lineages will help to identify robust evolutionary patterns in gene transcription among groups of diatoms. Transcriptional similarities and differences among the diatoms may provide insight into the evolutionary trajectories and niche differentiation within this diverse group of primary producers to determine whether or not the marked diversity in gene-specific responses translates directly to significantly different ecological consequences.

2.7 Acknowledgments

The authors acknowledge Franziska Lutz and Tiffany Truong for assistance with laboratory culturing, Dr. Betsy Welsh for assistance with analyses conducted in R, and Dr. Jody W. Deming and Dr. Anitra Ingalls for input on the manuscript. This research was funded by a Gordon and Betty Moore Foundation Marine Microbiology Investigator Award to E. Virginia

Armbrust. MMETSP Samples in supplemental Table 2.1 were sequenced at the National Center for Genome Resources through Grant #2637 from the Gordon and Betty Moore Foundation.

2.8 References

Allen AE. (2005). Beyond sequence homology: Redundant ammonium transporters in a marine diatom are not functionally equivalent. J Phycol 41:4–6.

Allen AE, Dupont CL, Obornik M, Horak A, Nunes-Nesi A, McCrow JP, et al. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature 473:203–209.

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

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. J Mol Biol 215:403–410.

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

Armbrust EV, Berges JAJ, Bowler CC, Green BRB, Martinez DD, Putnam NH, et al. (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306:79–86.

Ashworth J, Coesel S, Lee A, Armbrust EV, Orellana MV, Baliga NS. (2013). Genome-wide diel growth state transitions in the diatom *Thalassiosira pseudonana*. Proc Natl Acad Sci in press.

Auro ME, Cochlan WP. (2013). Nitrogen utilization and toxin production by two diatoms of the *Pseudo-nitzschia pseudodelicatissima* complex: *P. cuspidata* and *P. fryxelliana*. J Phycol 49:156–169.

Behrenfeld MJ, Halsey KH, Milligan AJ. (2008). Evolved physiological responses of phytoplankton to their integrated growth environment. Philos T R Soc B 363:2687–2703.

Bender SJ, Parker MS, Armbrust EV. (2012). Coupled effects of light and nitrogen source on the urea cycle and nitrogen metabolism over a diel cycle in the marine diatom *Thalassiosira pseudonana*. Protist 163:232–251.

Berges JA, Franklin DJ, Harrison PJ. (2001). Evolution of an artificial seawater medium: improvements in enriched seawater, artificial water over the last two decades. J Phycol 37:1138–1145.

Bowler C. (2010). Genomics-enabled approaches for revealing the molecular secrets of marine diatoms. Febs J 277:8–8.

Bowler CC, Allen AEA, Badger JHJ, Grimwood JJ, Jabbari KK, Kuo AA, et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456:239–244.

Brand LE, Guillard RRL, Murphy LS. (1981). A method for the rapid and precise determination of acclimated phytoplankton reproduction rates. J Plankton Res 3:193–201.

Claquin P, Martin-Jezequel V, Kromkamp J, Veldhuis M, Kraay G. (2002). 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 phosphorus control. J Phycol 38:922–930.

Dugdale R, Wilkerson F. (1998). Silicate regulation of new production in the equatorial Pacific upwelling. Nature 391:270–273.

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield OM, et al. (2004). The evolution of modern eukaryotic phytoplankton. Science 305:354–360.

Falkowski PG, Stone D. (1975). Nitrate uptake in marine phytoplankton- Energy sources and interaction with carbon fixation. Mar Biol 32:77–84.

Guillard RR, Ryther JH. (1962). Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (cleve) Gran. Can J Microbiol 8:229–239.

Harrison PJ, Thompson P, Calderwood GA. (1990). Effects of nutrient and light limitation on the biochemical composition of phytoplankton. J Appl Phycol 2:45–56.

Hildebrand M. (2005). Cloning and functional characterization of ammonium transporters from the marine diatom *Cylindrotheca fusiformis* (Bacillariophyceae). J Phycol 41:105–113.

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.

Kanehisa M, Goto S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40:D109–14.

Katz ME, Finkel ZV, Grzebyk D, Knoll AH, Falkowski PG. (2004). Evolutionary trajectories and biogeochemical impacts of marine eukaryotic phytoplankton. Annu Rev Ecol Syst 35:523–556.

Koester JA, Swanson WJ, Armbrust EV. (2012). Positive selection within a diatom species acts on putative protein interactions and transcriptional regulation. Mol Biol Evol 30:422–434.

Kolber ZS, Zehr JP, Falkowski PG. (1988). Effects pf growth, irradiance and nitrogen limitation on photosynthetic energy conversion in photosystem II. Plant Physiol 88:923–929.

Kooistra W, Gersonde R, Medlin LK, Mann DG. (2007). The origin and evolution of the diatoms: their adaptation to a planktonic existence. In:Evolution of Primary Producers in the

Sea, Falkowski, PG & Knoll, AH, eds. Elsevier Academic Press: Amsterdam, pp. 210–225.

Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, et al. (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. PLOS One 3:e1426.

LaRoche J, Geider RJ, Graziano LM, Murray H, Lewis K. (1993). Induction of specific proteins in eukaryotic algae grown under iron-deficient, phosphorus-deficient, or nitrogen-deficient conditions. J Phycol 29:767–777.

Li H, Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Li L, Stoockert C, Roos D. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189.

Lomas M, Glibert PM. (2000). Comparisons of nitrate uptake, storage, and reduction in marine diatoms and flagellates. J Phycol 36:903–913.

Lommer M, Specht M, Roy A-S, Kraemer L, Andreson R, Gutowska MA, et al. (2012). Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. Genome Biol 13:R66.

Maheswari U, Jabbari K, Petit J-L, Porcel BM, Allen AE, Cadoret J-P, et al. (2010). Digital expression profiling of novel diatom transcripts provides insight into their biological functions. Genome Biol 11:R85.

Martens C, Vandepoele K, Van de Peer YY. (2008). Whole-genome analysis reveals molecular innovations and evolutionary transitions in chromalveolate species. Proc Natl Acad Sci 105:3427–3432.

Martin P, Lampitt RS, Jane Perry M, Sanders R, Lee C, D'Asaro E. (2011). Export and mesopelagic particle flux during a North Atlantic spring diatom bloom. Deep-Sea Res Pt I 58:338–349.

Mock T, Samanta MP, Iverson V, Berthiaume C, Robison MM, Holtermann K, et al. (2008). Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. Proc Natl Acad Sci 105:1579–1584.

Montsant A, Jabbari K, Maheswari U, Bowler C. (2005). Comparative genomics of the pennate diatom *Phaeodactylum tricornutum*. Plant Physiol 137:500–513.

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

Palmucci M, Ratti S, Giordano M. (2011). Ecological and evolutionary implications of carbon allocation in marine phytoplankton as a function of nitrogen availability: A fourier transform

infrared spectroscopy approach. J Phycol 47:313–323.

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

Raven JA. (1987). The role of vacuoles. New Phytol 106:357–422.

Robinson MD, McCarthy DJ, Smyth GK. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

Robinson MD, Smyth GK. (2007). Moderated statistical tests for assessing differences in tag abundance. Bioinformatics 23:2881–2887.

Sims PA, Mann DG, Medlin LK. (2006). Evolution of the diatoms: insights from fossil, biological and molecular data. Phycologia 45:361–402.

Smetacek V. (1998). Biological oceanography: diatoms and the silicate factor. Nature 39:224–225.

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

Su ZZ, Olman VV, Mao FF, Xu YY. (2005). Comparative genomics analysis of NtcA regulons in cyanobacteria: regulation of nitrogen assimilation and its coupling to photosynthesis. Nucleic Acids Res 33:5156–5171.

Tolonen ACA, Aach JJ, Lindell DD, Johnson ZIZ, Rector TT, Steen RR, et al. (2006). Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. Mol Syst Biol 2:53–53.

Tozzi S, Schofield OM, Falkowski PG. (2004). Historical climate change and ocean turbulence as selective agents for two key phytoplankton functional groups. Mar Ecol Prog Ser 274:123–132.

Turpin D. (1991). Effects of inorganic N availability on algal photosynthesis and carbon metabolism. J Phycol 27:14–20.

UNESCO. (1994). Protocols for the Joint Global Ocean Flux Study (JGOFS) core measurements. In:IOC Manual and Guides 29, IOC, eds. Intergovernmental Oceanographic Commission: Paris, p. 170.

Wilkerson F, Dugdale R, Kudela RM, Chavez FP. (2000). Biomass and productivity in Monterey Bay, California: contribution of the large phytoplankton. Deep-Sea Res Pt II 47:1003–1022.

Table 2.1 Genomic and transcriptomic comparisons of orthology and expression among *T. pseudonana* (*Tp*), *F. cylindrus* (*Fc*), and *P. multiseries* (*Pm*), including genes with differential expression under nitrate starvation ($p < 0.01$).

	Modeled genes	Expressed genes[a]	Differentially expressed genes[b]	Differentially expressed core genes[c]		Differentially expressed core genes with KEGG annotations[d]		Core genes with shared differential expression[e]		Genes differentially expressed under nitrate starvation only[f]
				UP	DOWN	UP	DOWN	UP 24 clusters	DOWN 160 clusters	
Tp	11,390	11,324	4,315	787	1,319	246	437	27	168	1,102
		(99)	(38)	(45)		(32)		(9)		(52)
Fc	18,077	17,386	2,960	285	731	87	335	25	166	915
		(96)	(16)	(14)		(42)		(19)		(90)
Pm	19,703	16,839	4,768	737	1,070	277	359	24	168	1,047
		(86)	(24)	(22)		(35)		(11)		(58)

For each diatom, values in parentheses represent the percentage of [a]modeled genes with expression, [b]modeled genes with differential expression, [c]differentially expressed genes that are present in the core group, and differentially expressed core genes that have [d]KEGG annotations, [e]shared differential expression, or [f]were expressed under nitrate starvation only

Table 2.2 Shared response (differential expression; $p < 0.01$) of pennate diatoms *F. cylindrus* (*Fc*), and *P. multiseries* (*Pm*) to the onset of nitrate starvation.

	Differentially expressed pennate genes		% of total differentially expressed genes	Differentially expressed pennate genes with KEGG annotations		% of differentially expressed pennate genes	Pennate genes with shared differential expression		% of differentially expressed pennate genes
	UP	DOWN		UP	DOWN		UP 74 clusters	DOWN 81 clusters	
Fc	400	291	16%	58	29	13%	78	84	23%
Pm	473	889	30%	61	89	11%	74	84	12%

Table 2.3 Distinct species responses (differential expression; $p < 0.01$) of *T. pseudonana* (*Tp*), *F. cylindrus* (*Fc*), and *P. multiseries* (*Pm*) at the onset of nitrate starvation.

	Differentially expressed non-orthologous genes		% of total differentially expressed genes	Differentially expressed non-orthologous genes with KEGG annotations		% of differentially expressed non-orthologous genes	Differentially expressed non-orthologous under nitrate starvation only	% of differentially expressed non-orthologous genes
	UP	DOWN		UP	DOWN			
Tp	984	997	42%	73	67	7%	900	45%
Fc	661	517	17%	34	27	5%	790	67%
Pm	702	840	18%	19	32	3%	843	55%

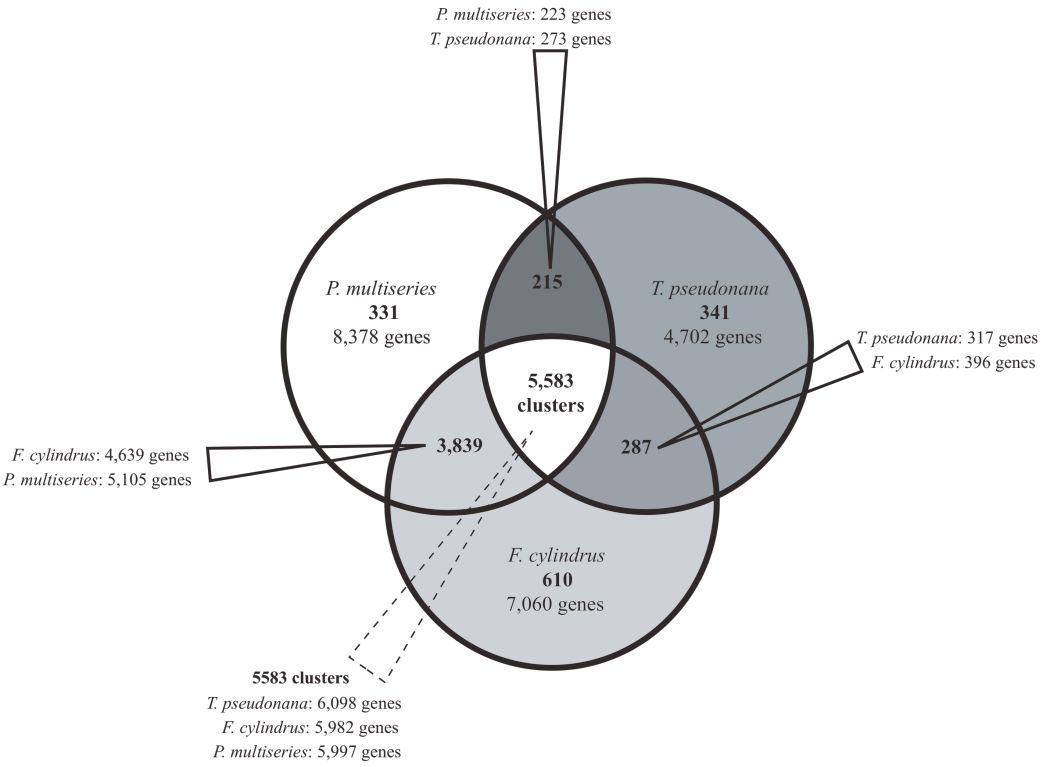
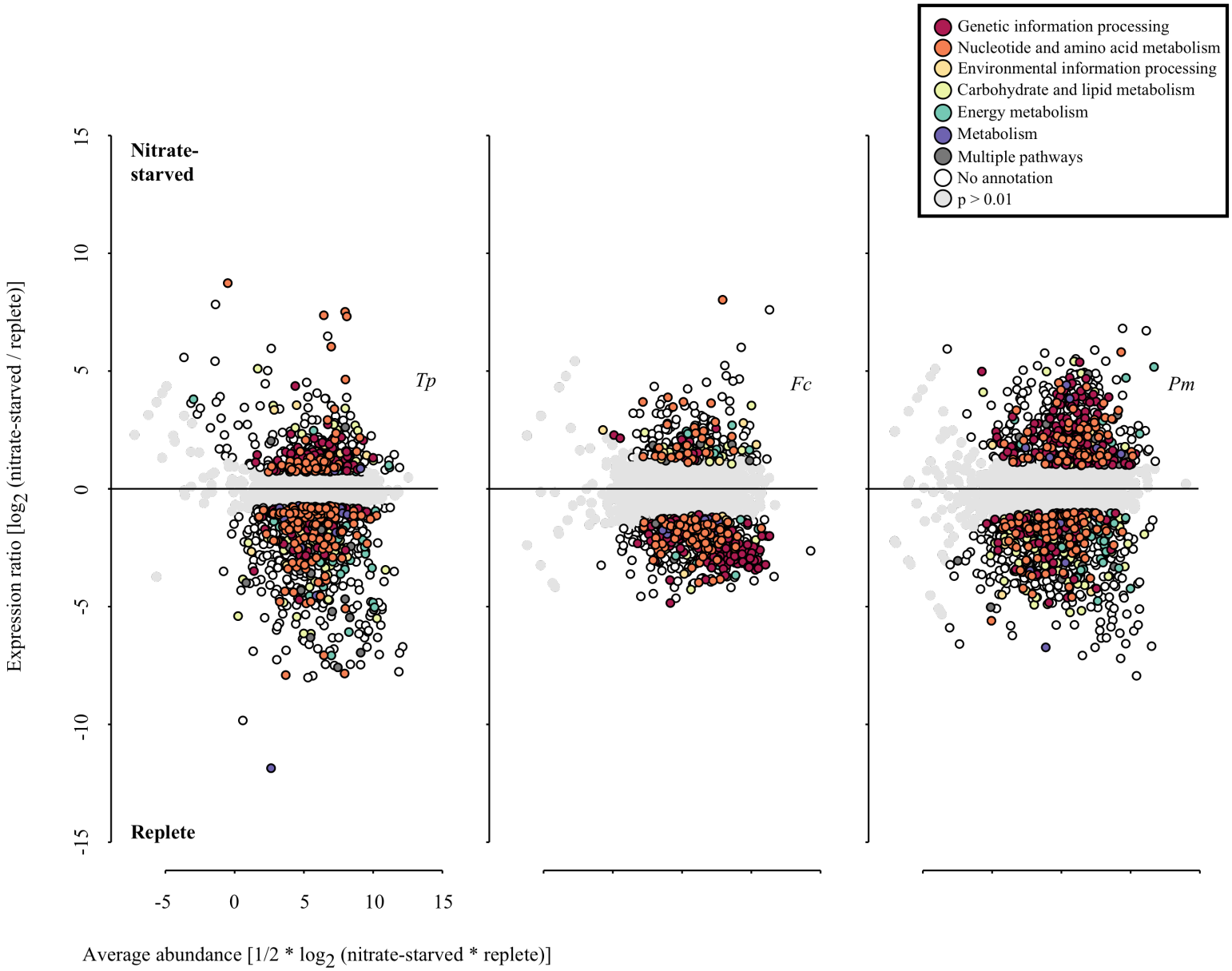


Figure 2.1 Venn diagram of modeled genes from *T. pseudonana*, *F. cylindrus* and *P. multiseries*. OrthoMCL-identified homologous gene clusters (in bold) and the number of genes within each cluster below or adjacent.

A.


Genetic information processing
Nucleotide and amino acid metabolism
Environmental information processing
Carbohydrate and lipid metabolism
Energy metabolism
Metabolism
Multiple pathways
No annotation
p > 0.01
Nitrate-starved
Replete
Tp
Fc
Pm
Expression ratio [log2 (nitrate-starved / replete)]
Average abundance [1/2 * log2 (nitrate-starved * replete)]
15
10
5
0
-5
-10
-15
-5
0
5
10
15

B.

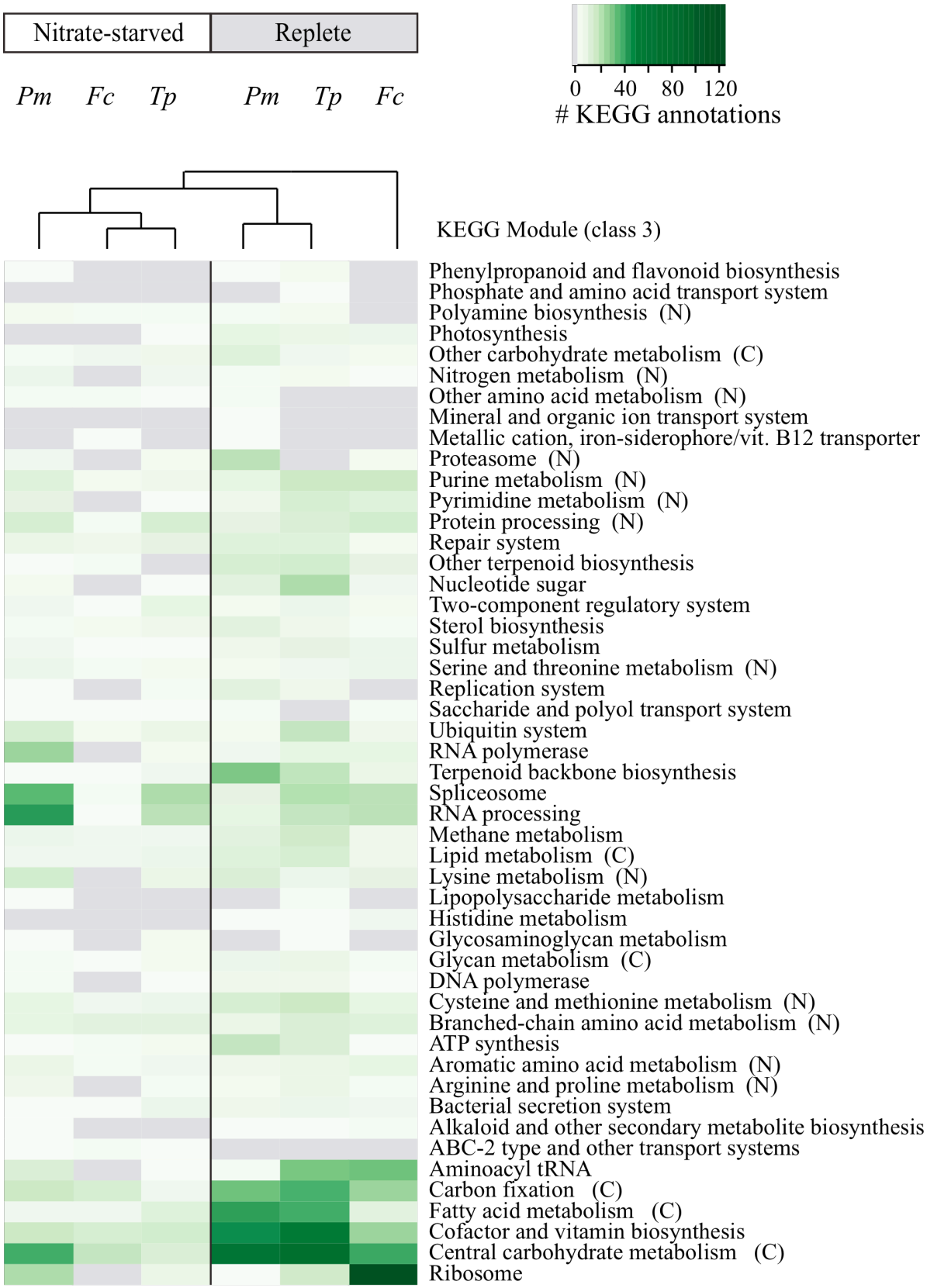

Nitrate-starved
Replete
Pm
Fc
Tp
Pm
Tp
Fc
0
40
80
120
KEGG annotations
KEGG Module (class 3)
Phenylpropanoid and flavonoid biosynthesis
Phosphate and amino acid transport system
Polyamine biosynthesis (N)
Photosynthesis
Other carbohydrate metabolism (C)
Nitrogen metabolism (N)
Other amino acid metabolism (N)
Mineral and organic ion transport system
Metallic cation, iron-siderophore/vit. B12 transporter
Proteasome (N)
Purine metabolism (N)
Pyrimidine metabolism (N)
Protein processing (N)
Repair system
Other terpenoid biosynthesis
Nucleotide sugar
Two-component regulatory system
Sterol biosynthesis
Sulfur metabolism
Serine and threonine metabolism (N)
Replication system
Saccharide and polyol transport system
Ubiquitin system
RNA polymerase
Terpenoid backbone biosynthesis
Spliceosome
RNA processing
Methane metabolism
Lipid metabolism (C)
Lysine metabolism (N)
Lipopolysaccharide metabolism
Histidine metabolism
Glycosaminoglycan metabolism
Glycan metabolism (C)
DNA polymerase
Cysteine and methionine metabolism (N)
Branched-chain amino acid metabolism (N)
ATP synthesis
Aromatic amino acid metabolism (N)
Arginine and proline metabolism (N)
Bacterial secretion system
Alkaloid and other secondary metabolite biosynthesis
ABC-2 type and other transport systems
Aminoacyl tRNA
Carbon fixation (C)
Fatty acid metabolism (C)
Cofactor and vitamin biosynthesis
Central carbohydrate metabolism (C)
Ribosome

Figure 2.2 Differential expression and KEGG annotations for orthologous genes. A. Ratio-average (RA) plots of core orthologous genes among *T. pseudonana* (*Tp*, left), *F. cylindrus* (*Fc*, center) and *P. multiseries* (*Pm*, right) under nitrate starvation relative to nutrient-replete control conditions. White circles indicate genes with significant differential expression ($p < 0.01$); gray circles indicate genes where expression does not differ significantly from control ($p > 0.01$). Genes with significant differential expression that mapped to a KEGG module (class 2) are highlighted in color. Positive differential expression values, nitrate-starved response; negative differential expression values, nutrient-replete response. B. Heatmap of KEGG annotations based on presence, absence, and abundance of annotations (KEGG class 3 modules). The darker green colors indicate a greater abundance of KEGG annotations; gray boxes, no KEGG annotation. KEGG modules associated with carbon (C) and nitrogen (N) metabolism are labeled accordingly. Treatments and KEGG modules are clustered hierarchically based on similarities and differences in KEGG annotation abundances. Abbreviations: *Tp* (*T. pseudonana*), *Fc* (*F. cylindrus*), and *Pm* (*P. multiseries*).

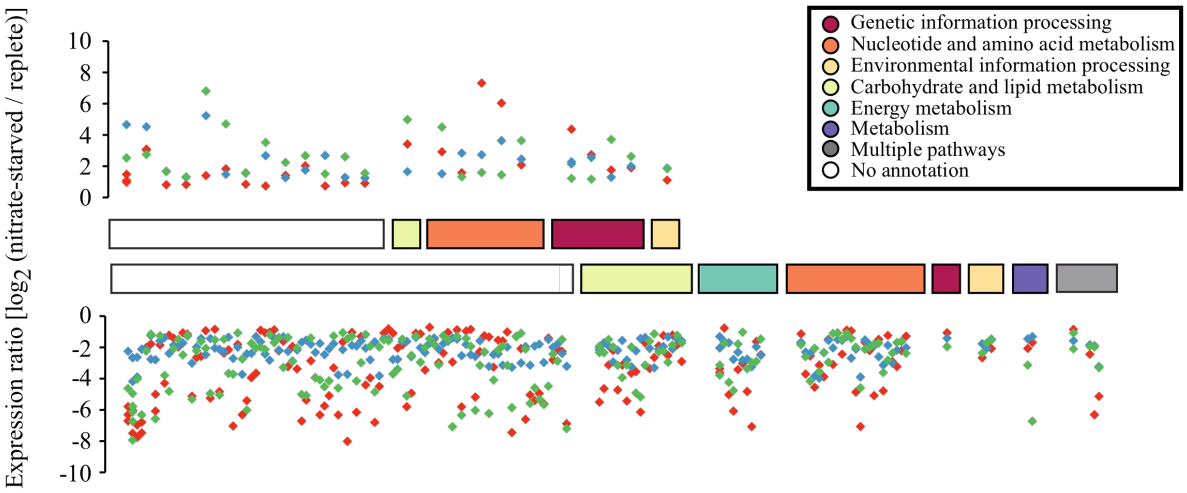


Figure 2.3 Scatterplot of core orthologous genes with shared significant differential patterns among the three diatoms. Orthologous genes from the same cluster are stacked vertically; colored diamonds correspond to each species (red, *T. pseudonana*; blue, *F. cylindrus*; green, *P. multiseries*). Stacked clusters are sorted according to their KEGG module class 2 ID (x-axis; colored bars). Clusters with no KEGG module ID are represented by the white bars. The length of colored bars along the x-axis is qualitatively proportional to the number of orthologous gene clusters represented within a KEGG module.

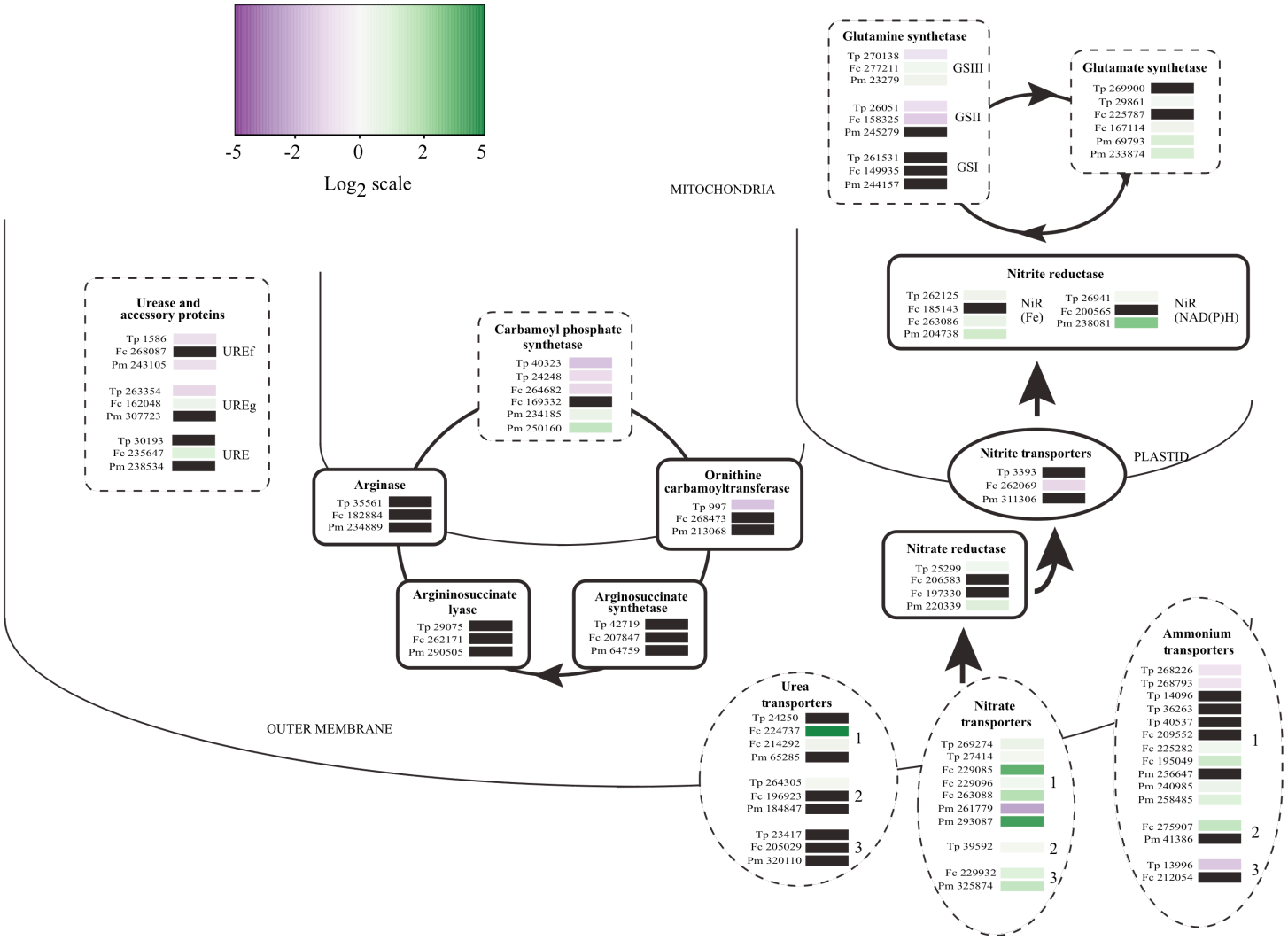
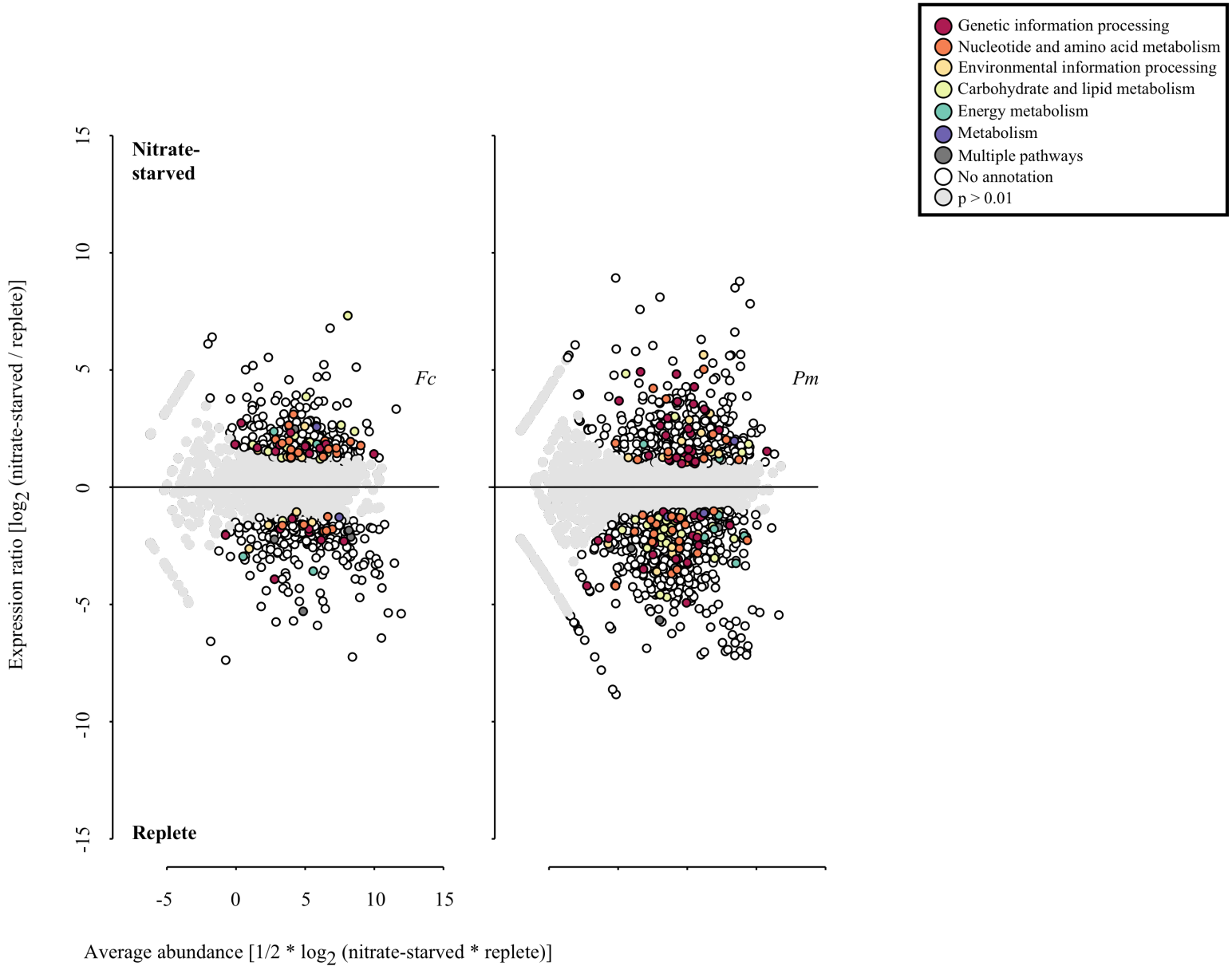

-5 -2 0 2 5
Log2 scale
MITOCHONDRIA
Glutamine synthetase
Tp 270138
Fc 277211
Pm 23279
GSIII
Tp 26051
Fc 158325
Pm 245279
GSII
Tp 261531
Fc 149935
Pm 244157
GSI
Glutamate synthetase
Tp 269900
Tp 29861
Fc 225787
Fc 167114
Pm 69793
Pm 233874
Nitrite reductase
Tp 262125
Fc 185143
Fc 263086
Pm 204738
NiR (Fe)
Tp 26941
Fc 200565
Pm 238081
NiR (NAD(P)H)
Urease and accessory proteins
Tp 1586
Fc 268087
Pm 243105
UREf
Tp 263354
Fc 162048
Pm 307723
UREg
Tp 30193
Fc 235647
Pm 238534
URE
Carbamoyl phosphate synthetase
Tp 40323
Tp 24248
Fc 264682
Fc 169332
Pm 234185
Pm 250160
Arginase
Tp 35561
Fc 182884
Pm 234889
Ornithine carbamoyltransferase
Tp 997
Fc 268473
Pm 213068
Nitrite transporters
Tp 3393
Fc 262069
Pm 311306
PLASTID
Nitrate reductase
Tp 25299
Fc 206583
Fc 197330
Pm 220339
Argininosuccinate lyase
Tp 29075
Fc 262171
Pm 290505
Arginosuccinate synthetase
Tp 42719
Fc 207847
Pm 64759
OUTER MEMBRANE
Urea transporters
Tp 24250
Fc 224737
Fc 214292
Pm 65285
1
Tp 264305
Fc 196923
Pm 184847
2
Tp 23417
Fc 205029
Pm 320110
3
Nitrate transporters
Tp 269274
Tp 27414
Fc 229085
Fc 229096
Fc 263088
Pm 261779
Pm 293087
1
Tp 39592
2
Fc 229932
Pm 325874
3
Ammonium transporters
Tp 268226
Tp 268793
Tp 14096
Tp 36263
Tp 40537
Fc 209552
Fc 225282
Fc 195049
Pm 256647
Pm 240985
Pm 258485
1
Fc 275907
Pm 41386
2
Tp 13996
Fc 212054
3

Figure 2.4 Heat map of the transcriptional response to nitrate starvation for known nitrogen-related genes in each diatom. The protein ID of genes encoded by each species (Tp = *T. pseudonana*; Fc = *F. cylindrus*; Pm = *P. multiseries*) is next to a color-coded bar of significant differential expression relative to a control treatment ($p < 0.05$). Black bars indicate genes without significant differential expression under nitrate-starved or nutrient-replete growth. Genes are grouped according to their respective isoforms; transporters are grouped according to their OrthoMCL clusters (indicated with a number).

A.


Genetic information processing
Nucleotide and amino acid metabolism
Environmental information processing
Carbohydrate and lipid metabolism
Energy metabolism
Metabolism
Multiple pathways
No annotation
p > 0.01
Nitrate-starved
Replete
Fc
Pm
Expression ratio [log2 (nitrate-starved / replete)]
Average abundance [1/2 * log2 (nitrate-starved * replete)]
15
10
5
0
-5
-10
-15

B.


Nitrate-starved
Replete
Pm
Fc
Fc
Pm
0 2 4 6 8 10
KEGG annotations
KEGG Module (class 3)
Sterol biosynthesis
Sulfur metabolism
Serine and threonine metabolism (N)
Saccharide and polyol transport system
Two-component regulatory system
Ubiquitin system
Terpenoid backbone biosynthesis
Spliceosome
RNA processing
Ribosome
RNA polymerase
Replication system
Purine metabolism (N)
Repair system
Protein processing (N)
Proteasome (N)
Polyamine biosynthesis (N)
Photosynthesis
Phosphate and amino acid transport system
Phenylpropanoid and flavonoid biosynthesis
Other terpenoid biosynthesis
Other carbohydrate metabolism (C)
Other amino acid metabolism (N)
Nucleotide sugar
Nitrogen metabolism (N)
Methane metabolism
Metallic cation, iron-siderophore/vit. B12 transporter
Lipopolysaccharide metabolism
Lysine metabolism (N)
Lipid metabolism (C)
Glycosaminoglycan metabolism
Fatty acid metabolism (C)
Glycan metabolism (C)
DNA polymerase
Cofactor and vitamin biosynthesis
Cysteine and methionine metabolism (N)
Carbon fixation (C)
Central carbohydrate metabolism (C)
Bacterial secretion system
ATP synthesis
Aromatic amino acid metabolism (N)
Aminoacyl tRNA
Arginine and proline metabolism (N)
Alkaloid and other secondary metabolite biosynthesis

C.

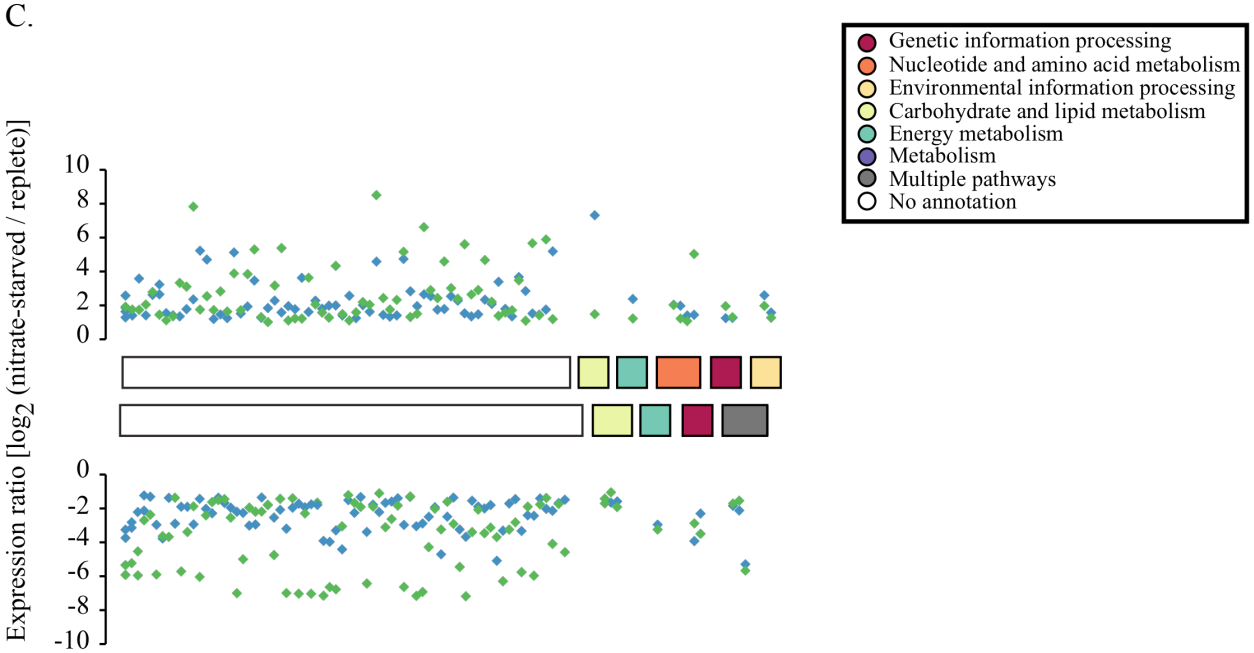
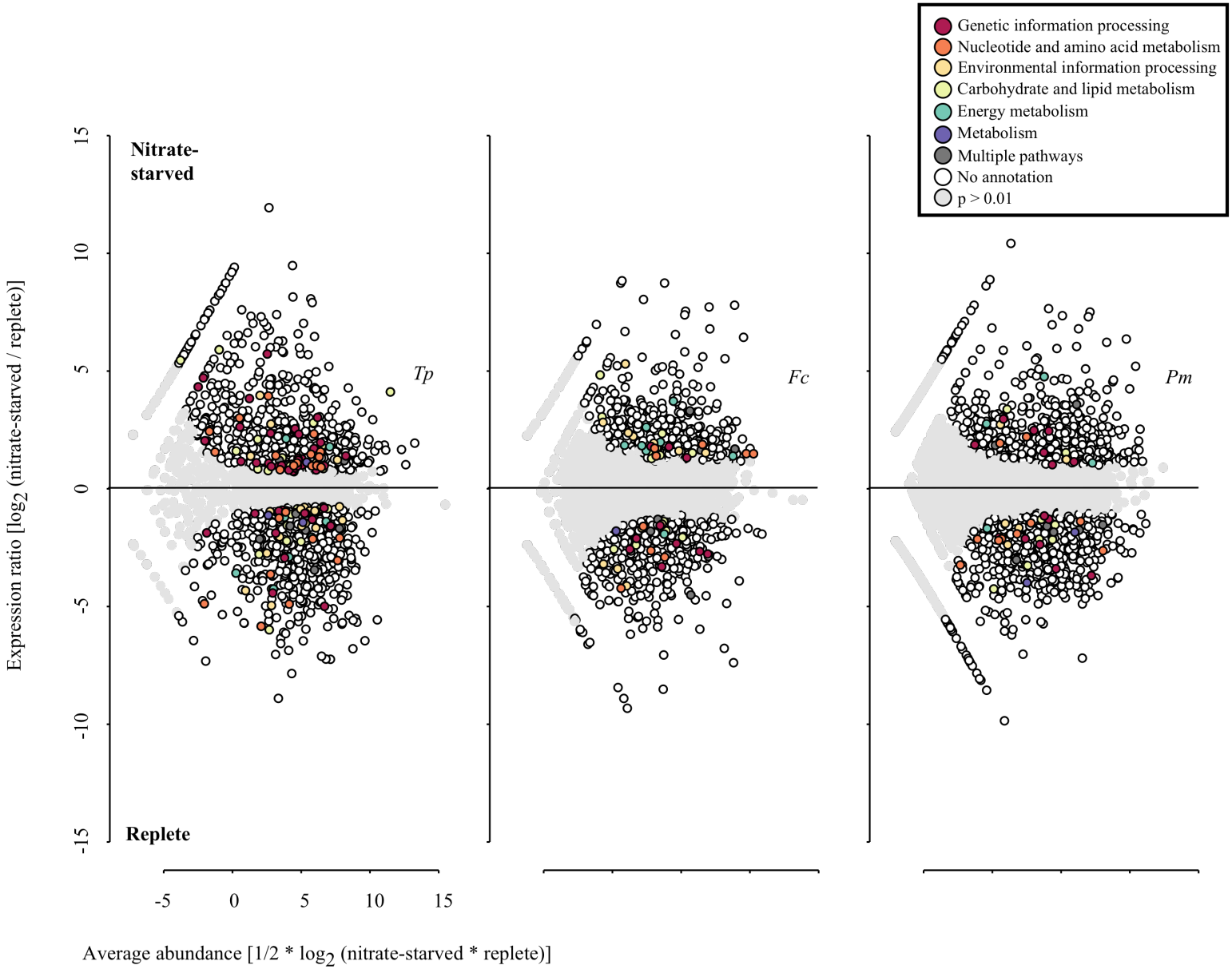


Figure 2.5 Differential expression and KEGG annotations for orthologous pennate genes. A. Ratio-average (RA) plots for *F. cylindrus* (*Fc*, left) and *P. multiseries* (*Pm*, right) under nitrate starvation relative to nutrient-replete control conditions. White circles indicate genes with significant differential expression ($p < 0.01$); gray circles indicate genes where expression does not differ significantly from control ($p > 0.01$). Genes with significant differential expression that mapped to a KEGG module (class 2) are highlighted in color. Positive differential expression values, nitrate-starved response; negative differential expression values, nutrient-replete response. B. Heatmap of KEGG annotations based on presence, absence, and abundance of annotations (KEGG class 3 modules). The darker green colors indicate a greater abundance of KEGG annotations; gray boxes, no KEGG annotation. KEGG modules associated with carbon (C) and nitrogen (N) metabolism are labeled accordingly. Treatments and KEGG modules are clustered hierarchically based on similarities and differences in KEGG annotation abundances. Abbreviations: *Fc* (*F. cylindrus*) and *Pm* (*P. multiseries*). C. Scatterplot of pennate genes that share the same significant differential pattern for every gene in a given cluster. Orthologous genes from the same cluster are stacked vertically; colored diamonds correspond to the species (blue, *F. cylindrus*; green, *P. multiseries*). Stacked clusters are sorted according to their KEGG module class 2 ID (x-axis). Clusters with no KEGG module ID are represented by the white bars. The length of colored bars along the x-axis is qualitatively proportional to the number of orthologous gene clusters represented within a KEGG module.

A.


Genetic information processing
Nucleotide and amino acid metabolism
Environmental information processing
Carbohydrate and lipid metabolism
Energy metabolism
Metabolism
Multiple pathways
No annotation
p > 0.01
Nitrate-starved
Replete
Tp
Fc
Pm
Expression ratio [log2 (nitrate-starved / replete)]
Average abundance [1/2 * log2 (nitrate-starved * replete)]
15
10
5
0
-5
-10
-15

B.

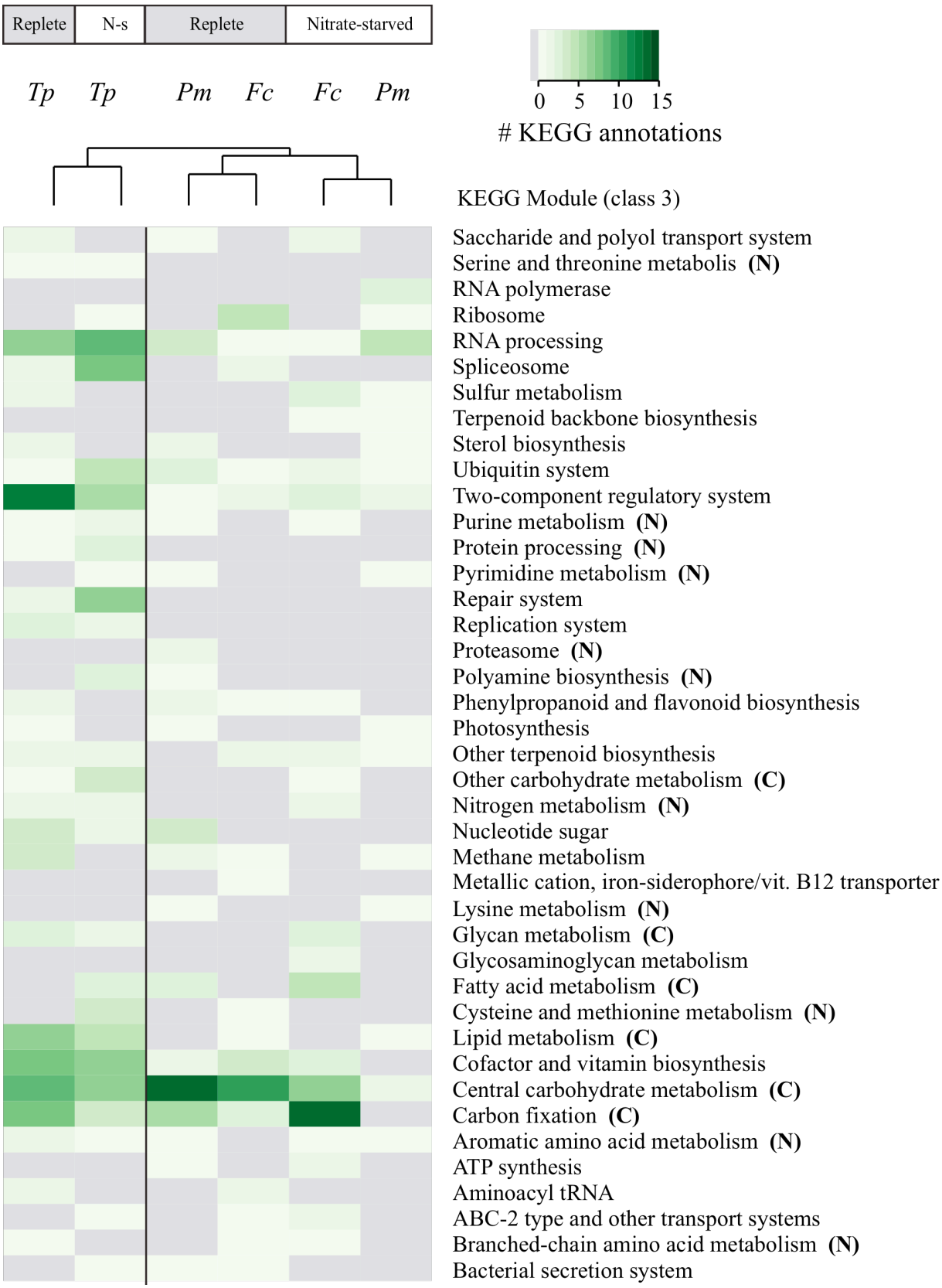

Replete
N-s
Replete
Nitrate-starved
Tp
Tp
Pm
Fc
Fc
Pm
0
5
10
15
KEGG annotations
KEGG Module (class 3)
Saccharide and polyol transport system
Serine and threonine metabolis (N)
RNA polymerase
Ribosome
RNA processing
Spliceosome
Sulfur metabolism
Terpenoid backbone biosynthesis
Sterol biosynthesis
Ubiquitin system
Two-component regulatory system
Purine metabolism (N)
Protein processing (N)
Pyrimidine metabolism (N)
Repair system
Replication system
Proteasome (N)
Polyamine biosynthesis (N)
Phenylpropanoid and flavonoid biosynthesis
Photosynthesis
Other terpenoid biosynthesis
Other carbohydrate metabolism (C)
Nitrogen metabolism (N)
Nucleotide sugar
Methane metabolism
Metallic cation, iron-siderophore/vit. B12 transporter
Lysine metabolism (N)
Glycan metabolism (C)
Glycosaminoglycan metabolism
Fatty acid metabolism (C)
Cysteine and methionine metabolism (N)
Lipid metabolism (C)
Cofactor and vitamin biosynthesis
Central carbohydrate metabolism (C)
Carbon fixation (C)
Aromatic amino acid metabolism (N)
ATP synthesis
Aminoacyl tRNA
ABC-2 type and other transport systems
Branched-chain amino acid metabolism (N)
Bacterial secretion system

Figure 2.6 Differential expression and KEGG annotations for non-orthologous genes. A. Ratio-average (RA) plots of non-orthologous genes for *T. pseudonana* (*Tp*, left), *F. cylindrus* (*Fc*, center) and *P. multiseries* (*Pm*, right) under nitrate starvation relative to nutrient-replete control conditions. White circles indicate genes with significant differential expression ($p < 0.01$); black circles indicate genes where expression does not differ significantly from control ($p > 0.01$). Genes with significant differential expression that mapped to a KEGG module (class 2) are highlighted in color. Positive differential expression values, nitrate-starved response; negative differential expression values, nutrient-replete response. B. Heatmap of KEGG annotations based on presence, absence, and abundance (KEGG class 3 modules). The darker green colors indicate a greater abundance of KEGG annotations; black boxes, no KEGG annotation. KEGG modules associated with carbon (C) and nitrogen (N) metabolism are labeled accordingly. Treatments and KEGG modules are clustered hierarchically based on similarities and differences in KEGG annotation abundances. Abbreviations: *T. pseudonana* (*Tp*), *F. cylindrus* (*Fc*), *P. multiseries* (*Pm*), and Nitrate-starved (N-s).

Supplemental Table 2.1 List of queried Joint Genome Institute (JGI) or Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) diatoms for BLASTX search.

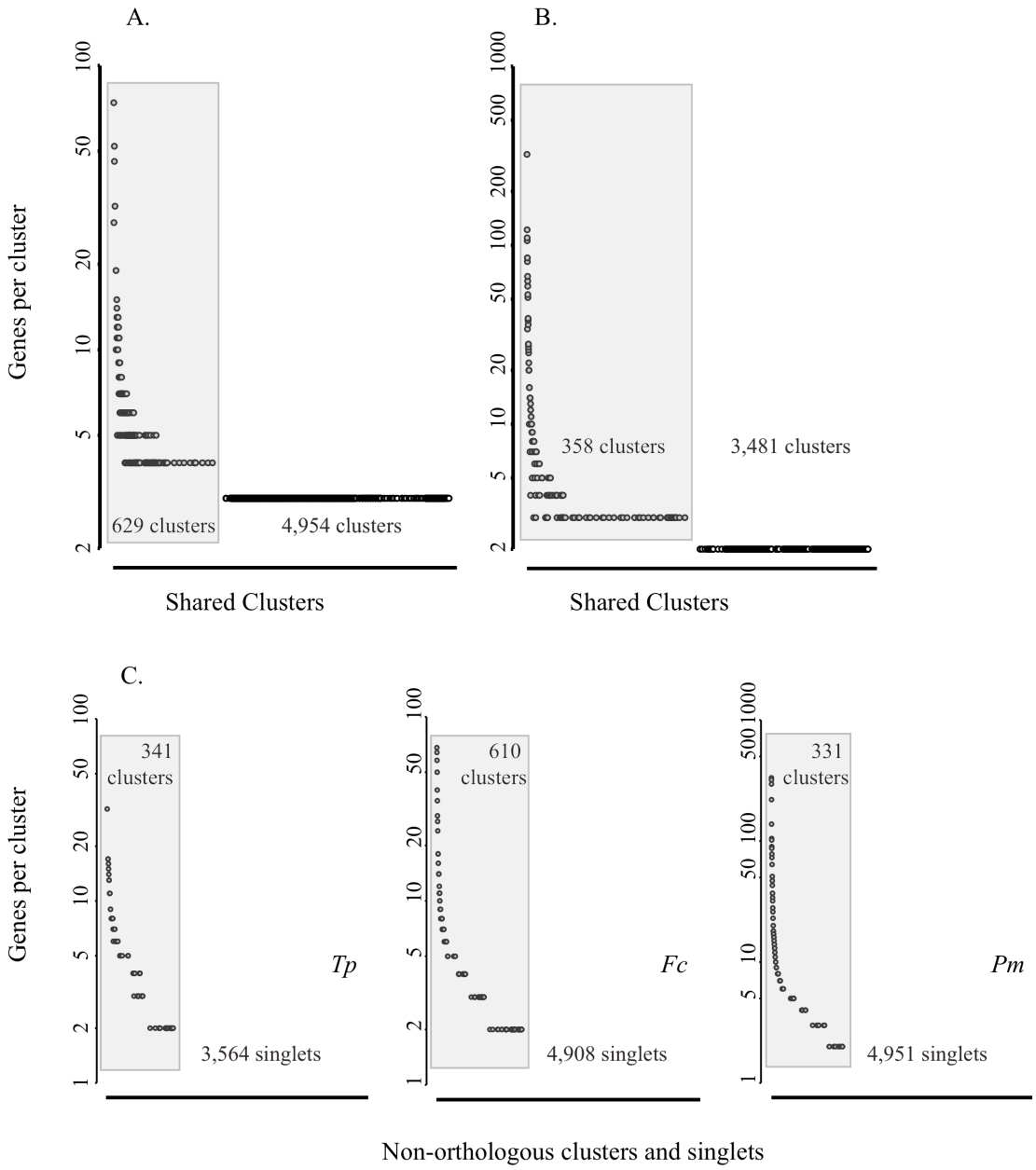
	Organism Name	Diatom Group	Identifier
1	*Phaeodactylum tricornutum*	Pennate	JGI P. tricornutumv2.0
2	*Amphiprora sp.*	Pennate	MMETSP1065
3	*Cylindrotheca closterium*	Pennate	MMETSP0017
4	*Nitzschia sp.*	Pennate	MMETSP0014
5	*Pseudo-nitzschia pungens cf. cingulata* (GGA1)	Pennate	MMETSP1060
6	*Pseudo-nitzschia pungens cf. pungens* (GGC3)	Pennate	MMETSP1061
7	*Striatella unipunctata*	Pennate	MMETSP0800
8	*Thalassionema frauenfeldii*	Pennate	MMETSP0786
9	*Aulacoseira subarctica*	Centric	MMETSP1064
10	*Corethron hystrix*	Centric	MMETSP0010
11	*Coscinodiscus wailesii*	Centric	MMETSP1066
12	*Detonula confervacea*	Centric	MMETSP1058
13	*Rhizosolenia setigera*	Centric	MMETSP0789
14	*Cyclotella meneghiniana*	Centric	MMETSP1057
15	*Ditylum brightwellii* SS4 Population 1	Centric	MMETSP1062
16	*Ditylum brightwellii* S10 Population 2	Centric	MMETSP1063
17	*Minutocellus polymorphus*	Centric	MMETSP1070
18	*Odontella sp.*	Centric	MMETSP0015
19	*Skeletonema costatum*	Centric	MMETSP0013
20	*Thalassiosira sp.*	Centric	MMETSP1071
21	*Thalassiosira sp.* (freshwater)	Centric	MMETSP1059
22	*Thalassiosira punctigera* C2	Centric	MMETSP1067

Supplemental Table 2.2 Growth rates, F_v/Fm, particulate organic carbon/nitrogen, and nutrient concentrations of nutrient-replete and nitrate-starved cultures of *T. pseudonana* (*Tp*), *F. cylindrus* (*Fc*), and *P. multiseries* (*Pm*).

	μ (exponential growth)	Treatment	F_v/F_m (exponential growth)	F_v/F_m (time of harvest)	POC/PON	Nutrients at time of harvest (μM)			
						Nitrate	Ammonium	Phosphate	Silicate
Tp	0.87 ± 0.19	Replete	0.63 ± 0.09	0.68 ± 0.01	13.43 ± 6.65	809.35 ± 16.33	2.30 ± 0.18	31.07 ± 1.45	67.78 ± 9.34
		Nitrate- starved	0.65 ± 0.01	0.44 ± 0.13	15.42 ± 0.77[b]	2.04 ± 3.19	1.96 ± 0.75	46.36 ± 7.82	129.70 ± 23.35
Fc	0.34 ± 0.04	Replete	0.52 ± 0.03	0.52 ± 0.03	16.57 ± 12.41	665.80 ± 127.39	0.43 ± 0.09	25.07 ± 2.55	73.55 ± 13.68
		Nitrate- starved	0.45 ± 0.00	0.40 ± 0.04	7.52 ± 1.34	9.96 ± 9.18	0.67 ± 0.53	50.61 ± 7.36	94.14 ± 22.53
Pm	0.60 ± 0.12	Replete	0.68 ± 0.01	0.68 ± 0.01	13.41 ± 2.97	821.66 ± 32.83	0.51 ± 0.41	32.09 ± 2.26	60.10 ± 27.38
		Nitrate- starved	0.66 ± 0.04	0.47 ± 0.10[a]	18.91 ± 1.41	0.47 ± 0.34	0.60 ± 0.57	56.47 ± 5.68	4.19 ± 0.98

[a] F_v/F_m at time of harvest differed significantly ($p < 0.05$) from F_v/F_m during exponential growth.
[b] *Tp* nitrate-starved POC/PON differed significantly (Bonferroni corrected, $p < 0.167$) from *Pm* nitrate-starved POC/PON.

Supplemental Table 2.3 Transcript abundance information for all genes that had significant differential expression in response to the onset of nitrate starvation for *T. pseudonana*, *F. cylindrus*, and *P. multiseries* (nitrate-starved treatment normalized to the nutrient-replete control; $p < 0.01$). Table includes KEGG and K0 annotation information. KEGG modules were determined based on the best BLASTX hit that contained a module ID (see Methods); the K0 associated with this module was also collected. NA: no annotation information. *, cluster ID for orthologous genes (NA, gene singlet). **, number of species associated with a given cluster ID. ***, patterns of differential expression in a given cluster (up, all clustered genes exhibited up-regulated gene expression; down, all clustered genes exhibited down-regulated gene expression; mix, clustered genes exhibited a range of differential expression patterns; na, analysis not conducted on non-orthologous genes). ****, genes with multiple KEGG module 2 annotations. When applicable, annotations were merged and identified in Figures 2.2A, 2.3, 2.5A, 2.5C, and 2.6A as "multiple pathways." Often, a gene is present on several lines in the spreadsheet because of multiple KEGG and/or K0 annotations. *See attached .xls file*.


A.
B.
Genes per cluster
629 clusters
4,954 clusters
358 clusters
3,481 clusters
Shared Clusters
Shared Clusters
C.
341 clusters
610 clusters
331 clusters
Tp
Fc
Pm
3,564 singlets
4,908 singlets
4,951 singlets
Non-orthologous clusters and singlets

Supplemental Figure 2.1 Distribution of the number of genes contained within orthologous clusters based on OrthoMCL analysis for core clusters (A), the pennate cluster (B) and the non-orthologous clusters (C). A, B: Clusters in which at least one diatom has more than two gene members are highlighted in gray. C: Non-orthologous clusters composed of more than two gene members for each diatom are highlighted in gray; number of gene singlets is listed for *T. pseudonana* (*Tp*, left), *F. cylindrus* (*Fc*, middle) or *P. multiseries* (*Pm*, right). The y-axis for the pennate clusters (B) and non-orthologous *P. multiseries* cluster (C, *Pm*) is scaled to accommodate a greater number of genes.

Chapter 3

Species-specific transcriptional responses of diatom communities to changes in iron and nitrogen availability along Line P in the Northeast Pacific Ocean

3.1 Abstract

Marine diatoms dominate environments characterized by sporadic nutrient supply because of evolved metabolic traits that include high- and low-affinity nitrogen transporters and vacuoles for nitrogen storage. In this study, diatom communities in surface waters were sampled along Line P in the Northeast Pacific Ocean to compare the transcriptional response at a coastal station (P1) to an offshore station (P8). At P1, the pennate diatom *Pseudo-nitzschia pungens* was blooming; at P8, diatom abundance was 10-fold lower. Nitrate concentrations in surface waters were similar between P1 and P8, but iron concentrations were an order of magnitude lower at P8. Functional annotations for the metatranscriptome sequences coupled to the chemical measurements indicated that the diatom community was iron-limited at P8. Changes in transcript abundance for diatom ferritin and nitrate reductase, as well as ammonium transporters, nitrate transporters, and a hypothetical protein (Tp4888) hypothesized to play a role in nutrient-replete growth in diatoms, were coupled to phylogenetic analyses to characterize the metabolic responses of the diatoms at P1 and P8. These genes were chosen because they represent key components of diatom metabolism that include nitrogen uptake, nitrate reduction, and iron storage. Detected changes in transcript abundance reflected the transcriptional response of a several dominant members of the diatom community that included *Pseudo-nitzschia sp.*, *Cylindrotheca sp.*, *Skeletonema sp.*, and *Leptocylindrus sp*. Closer analysis of the distribution of transcripts among genera revealed a diversity of transcriptional responses. As an example, the

greatest number of pennate nitrate reductase sequences were most closely related to *Pseudo-nitzschia sp.* at P1, and *Phaeodactylum sp.* at P8. Among the centric diatoms, more sequences recruited to *Detonula sp.* at P1, and *Rhizosolenia sp*. and *Coscinodiscus wailesii* at P8. Because nitrate reductase transcription increases when nitrate is available for diatom growth in laboratory studies, changes in its relative abundance in the field imply that different diatoms were exhibiting different metabolic responses to nitrate availability, and indicate that diatoms utilize species-specific strategies to cope with environmental change. Group-specific responses were also observed, such as the pennate diatom bloom at P1 that corresponded with a greater number of pennate ferritin sequences; these shared responses hint at underlying regulatory mechanisms that control the expression of a broad array of genes in the pennate versus centric diatoms. Observed transcriptional similarities and differences can be coupled to the environmental phenotype for each diatom to better explain diatom distributions in the marine environment, ultimately leading towards better mechanisms for predicting what diatom species will bloom.

3.2 Key words

Diatoms; metatranscriptomics; nitrogen transporters; environmental gradients; iron

3.3 Introduction

Marine diatoms are a key component of the global carbon cycle, contributing up to 90% of coastal primary production and 40% of marine primary production (Falkowski et al. 2004; Nelson et al. 1995). They also link the carbon and nitrogen cycles by contributing to both new production (fueled by upwelled nitrogen) and export production (Brzezinski et al. 1998; Dugdale & Wilkerson 1998). Marine diatoms are cosmopolitan and found throughout the world's oceans,

and they dominate in coastal regimes, exhibiting significant seasonal variability in growth and species presence/abundance (Booth et al. 1993; Ribalet et al. 2010). Coastal spring blooms fuel highly productive food webs. Following grazing by zooplankton, diatoms may sink out of surface waters or become entrained in the microbial loop (Azam et al. 1983). Diatom metabolic plasticity coupled to their ability to immediately take up and store nutrients enables this group to capitalize on sporadic pulses of nitrogen in coastal waters (Lommer et al. 2012).

In the Northeast Pacific Ocean, diatoms inhabit a range of environments that include low nitrate, high iron waters along the continental shelf (Harris et al. 2009), where growth is nitrate-limited until seasonal upwelling temporarily replenishes nitrate concentrations, and high nitrate, low iron waters farther offshore (HNLC region), where diatom growth is chronically limited by iron, aside from sporadic iron deposition events (Capone et al. 2008; Clemons & Miller 1984; Harrison et al. 1999; Fitzwater et al. 1996). In the NE Pacific, the diatom community varies on temporal and spatial scales, and is an amalgam of coastal and offshore assemblages. Small diatoms persist in high nitrate, low chlorophyll (HNLC) waters; diverse diatom species will bloom when iron is resupplied to the community (e.g. *Chaetoceros sp.*, *Nitzschia sp.*, and *Pseudo-nitzschia sp.*). Mixed communities of both pennate and centric diatoms populate coastal waters (Booth et al. 1993; Boyd et al. 1996; Marchetti et al. 2006; Taylor & Haigh 1996). Mesoscale anticyclonic and coastal eddies also influence the distribution of diatoms in this region, delivering coastal diatom species to offshore waters (Peterson & Harrison 2012; Batten & Crawford 2005).

The Line P timeseries, starting off the coast of Vancouver Island, British Columbia, and extending northwest to Ocean Station Papa, has captured seasonal patterns in diatom bloom formation for more than sixty years in this dynamic environment. This 1425-km transect spans a

steep environmental gradient across two vastly different oceanic regimes (Peña & Bograd 2007; Vedamati, 2013). Recent work identified a biological hotspot along Line P, characterized by a marked increase in phytoplankton community diversity, along with higher productivity, as HNLC waters intersected with more iron-rich coastal waters (Ribalet et al. 2010). Diatom cell abundances were higher at this hotspot compared to surrounding waters; the diatom communities likely capitalized on both nitrate and iron for growth, and were relieved from both nutrient limitations (Ribalet et al. 2010). Because these intersecting diatom populations are adapted to different oceanic regimes, they often exhibit separate physiological responses to changes in nutrient availability: diatoms will exhibit different nitrogen uptake rates (coastal diatoms have a lower affinity for nitrate uptake compared to open ocean species (Carpenter & Guillard 1971; Eppley et al. 1969)), nutrient quotas (open ocean diatoms require less iron to maintain their photosynthetic apparatus; (Strzepek & Harrison 2004)), and overall chemical compositions (e.g. coastal diatoms may have a higher carbon:nitrogen ratio to adjust to nitrogen-limited growth compared to open ocean species; (Marchetti & Harrison 2007; Geider & La Roche 2002)). The diversity of diatom species in the NE Pacific, and accompanying cellular machinery of each species provide multiple challenges for characterizing the metabolic responses within the community.

Recent work has identified taxon-specific responses to iron input along Line P (Durkin et al. 2012; Marchetti et al. 2012). Using comparative metatranscriptomics, Marchetti et al. (2012) captured a unique diatom response to iron availability that differed from the other phytoplankton community members (Marchetti et al. 2012). Following iron addition, an increase in differentially expressed diatom gene transcripts was observed, including for nitrogen-related genes. The authors hypothesized that oceanic diatoms are able to rapidly take up iron and apply

it towards nitrogen assimilation, simultaneously tempering the iron demands of photosynthesis by using iron-free equivalents, such as flavodoxin and plastocyanin (Marchetti et al. 2012). In the diatoms, key components of nitrogen metabolism require iron (such as the enzyme, nitrate reductase); during iron-limitation, diatoms re-partition nitrogen and iron in the cell as an adaptive mechanism for dealing with a low nutrient scenario.

Phylogenetic differences among diatoms in response to added iron were also evident in the same Line P community experiment. Durkin et al. (2012), examined changes in diatom silica precipitation and the transcription of silicon transporters - required for diatom growth, to iron addition. An increase in *Pseudo-nitzschia sp.* silicon transporter transcript abundance corresponded with increased *Pseudo-nitzschia sp.* silicification following iron addition; a decrease in silification among larger centrics was also documented. This work reflected differing physiological responses among the diatoms to iron addition (Durkin et al. 2012). Laboratory studies have also demonstrated metabolic differences between closely related (open ocean versus coastal) diatoms in response to iron-limited growth - the open ocean species requires less iron to photosynthesize compared to the coastal species, which results in faster growth rates for the open ocean diatom when grown under iron limitation in the laboratory (Strzepek & Harrison 2004). This evolved metabolic response is also evident at the genome-level; open ocean diatoms possess additional gene orthologs for iron-binding proteins (siderophores), and duplicated subdomains for proteins involved in iron metabolism (iron-starvation induced protein, ISIP), which may provide the cells with additional metabolic machinery to cope with growth under low iron conditions compared to coastal diatoms (Lommer et al. 2012; Smith et al. 2012). These studies provide the foundation for determining how different species of diatoms respond to changes in nutrient availability in a natural assemblage.

Quantifying gene expression in the field can serve a dual purpose: to provide a snapshot of the diatom community response to changes in nutrient availability, and to explain shifts in diatom species abundance based on phylogeny of key genes (Durkin et al. 2012). In this study, field metatranscriptomes were sampled along Line P at Station P1 (coastal) and Station P8 (offshore). Changes in transcript abundance, normalized to library size, were coupled to nutrient concentrations, chlorophyll *a* concentrations, cell counts, and 18S sequences to provide a quantitative account of how the diatom community was responding to changes in iron and nitrogen availability. Group- and species-specific responses highlighted the diversity of diatom responses within the community. Differences in relative transcript abundances indicate that certain members of the community were experiencing very different nutrient limitation scenarios.

3.4 Methods

Meta data collection and analyses

Seawater samples were collected along the Line P transect from May 16–21, 2012 on the R/V *Thomas G. Thompson* as part of the Global-scale Microbial Interactions across Chemical Survey (GeoMICS) cruise (Figure 3.1). A conductivity temperature depth (CTD) instrument package with 10 L Niskin bottles was used to collect non-trace metal samples at seven stations: Stations P8 (48.82 N, 128.66 W), P6 (48.74 N, 127.67 W), P5 (48.69 N, 127.17 W), P4 (48.65 N, 126.67 W), P3 (48.62 N, 126.33 N), P2 (48.60 N, 126.0 W), and P1 (48.58 N, 125.50 W). Trace metal samples were collected using 12 L Teflon coated GO FLO bottle at the same stations, as described in Vedamati (2013). All surface samples (5 m) were taken using a Lutz double diaphragm pump (LPT- ½).

For nutrients, approximately 50 mL of seawater were filtered through a 0.2 μm cellulose acetate syringe filter (Corning). Filtrate was analyzed for nitrate, ammonium, phosphate, and silicate with a Technicon autoanalyzer II system by the University of Washington Marine Chemistry Laboratory following the methods of UNESCO (UNESCO 1994). Total dissolved iron concentrations were measured using a single batch nitrilotriacetatic acid resin extraction and isotope dilution inductively coupled plasma mass spectrometry method described in Vedamati (2013), and originally adapted from Lee et al. (Lee et al. 2011). Bulk nutrient concentrations were plotted using Ocean Data View (v5.4.1; Schlitzer 2002); the gridding field feature was used to interpolate between known concentrations with respect to the x- and y-axes.

For chlorophyll *a* analysis, approximately 250 mL of water were filtered in triplicate onto 25 mm glass fiber filters (> 0.7 μm size fraction; Grade F filters, Sterlitech), and the filters frozen at −20°C until further analysis. In the lab, filters were immersed in 90% acetone (Parsons et al. 1984) and stored at −20°C for 24 h. Each filter was removed from the acetone, and the fluorescence of the extracted chlorophyll *a* and phaeo-pigments was measured with a Trilogy fluorometer (Turner) previously calibrated with a chlorophyll *a* standard (Turner Designs liquid primary chlorophyll *a* standard). Whole seawater was preserved in 1% Lugol's solution (final concentration). Samples were stored in amber bottles in the dark, as described by Marchetti et al. (2010); 50–100 mL of sample were settled in Utermöhl chambers, and 50–225 grids (depending on the concentration of cells in the sample) were counted by microscopy to determine diatom cell concentrations. For biogenic silica analysis, one liter of seawater was also filtered in triplicate onto 2 μm 47 mm polycarbonate filters (Millipore), and the filters stored at 4ºC until analysis at the University of Washington. There, filters were dried at 55ºC, biogenic silica was

dissolved in 0.2 M NaOH at 95ºC for 1 h, and then concentrations were quantified using the ammonium molybdate method (Brzezinski & Nelson 1986).

ARISA samples were collected according to the protocol of Hubbard et al. (2008). Briefly, two replicates of 500 mL each were collected from surface water and from the chlorophyll *a* maximum at each station, gently filtered onto 0.45 μm 25 mm mixed cellulose filters (Millipore), and stored at −80^{o}C until analysis. For 18S analysis, 1 L of surface water was pre-filtered through a 53 μm Nitex mesh, then filtered onto SUPOR-200 0.2 μm 47 mm polyethersulfone filters (Pall). Filters were preserved in 1 mL of lysis buffer (200 mM EDTA, 400 mM NaCl, 0.75 M sucrose, 50 mM Tris-HCl, pH 8.4), flash-frozen in liquid nitrogen and stored at −20^{o}C.

Duplicate samples were collected in succession at Station P8 and P1 from surface seawater (5 m) for metatranscriptome analysis (Table 3.1). Four to six liters of seawater were gently filtered through a 53 μm Nitex mesh pre-filter using the Lutz surface pump (above). Two additional filters were collected at P1 from seawater that was not pre-filtered with the Nitex mesh to capture a pennate diatom bloom. The water was filtered through a 2 μm 142 mm polycarbonate filter (PCTE membrane, Sterlitech) for ≤ 25 min, and then flash frozen in liquid nitrogen and stored at −80^{o}C.

DNA extraction, and 18S and ARISA sequencing and analysis

DNA for semi-quantitative Automated Ribosomal Intergenic Spacer Analysis (ARISA) was extracted from manually shredded filters using a DNeasy Plant Mini Kit (Qiagen). Genomic DNA concentrations were determined using a NanoDrop 1000 Spectrophotometer (Nanodrop Technologies). DNA was diluted to a concentration of 2 ng μL^{-1} and stored at −20°C. The Internal Transcribed Spacer 1 (ITS1) region of DNA corresponding to distinct *Pseudo-nitzschia*

communities was amplified from each sample using PCR, and *Pseudo-nitzschia* specific primers, as described in Hubbard et al. (2008). Triplicate PCR products were pooled, and purified using MultiScreen PCR 96-well filter plates (Millipore). Samples were quantified a second time using PicoGreen (Invitrogen) and a SpectraMax M2 microplate reader, standardized to 0.1 ng μL^{-1}; 1 ng DNA was precipitated with ethanol. DNA was resuspended with a cocktail containing 0.078 μL 10% Tween 20, 9.77 μL sterile water, and 0.15 μL fluorescent size standard Et-ROX 400 (Amersham Biosciences). Samples were analyzed at the Fred Hutchinson Cancer Research Center on an ABI Capillary Sequencer with the fluorescent-labeled internal size standard. The resulting electropherograms were analyzed using the software package DAx by examining peak profiles for each sample. The online fragment binner Dakster was used to extract fluorescent peak height data; fragment peak heights that represented over 3% of total fluorescence of the sample were included in the analysis. Fragment lengths corresponded to different *Pseudo-nitzschia* species, as described by Hubbard et al. (2008).

DNA was extracted for 18S sequencing using a modified version of the DNeasy Plant kit protocol (Qiagen) (Sudek, 2011; http://www.mbari.org/phyto-genome/Resources.html). Extracted DNA was quantified using a NanoDrop and Qubit, and 20 μg μL^{-1} of DNA per sample was sent for 18S tagged pyrosequencing at the Research and Testing Laboratory (Lubbock, Texas). Sequenced DNA was processed according to their in-house protocols (http://www.researchandtesting.com/docs/Data_Analysis_Methodology.pdf).

RNA extraction, Illumina library preparation, and Illumina sequencing

RNA was extracted from cells on frozen filters using the Totally RNA extraction kit (Invitrogen). The RNA was incubated with DNAse I (Ambion) at 37°C for 1 h; DNAse activity was terminated by the DNAse inactivation reagent (Ambion). The RNA was then purified by

ethanol precipitation, and ribosomal RNA was removed using the Poly(A)Purist Kit (Ambion). Poly(A)-selected samples were purified by ethanol precipitation, and then amplified using the MessageAmp II aRNA Amplification Kit (Ambion). Samples were quantified on the NanoDrop and Qubit spectrophotometers prior to cDNA synthesis. Between 2.5 and 5 μg of amplified polyA-selected RNA was converted to double-stranded cDNA using the SuperScript III First Strand Synthesis System (Invitrogen), and the NEBNext mRNA Second Strand Synthesis Module (New England BioLabs). Following second-strand synthesis, 2 units of T4 DNA polymerase per μg of input RNA (New England BioLabs) were added to the reaction. Double-stranded cDNA samples were purified using the PureLink PCR Purification kit (Life Technologies), followed by an ethanol precipitation step, and resuspended in nuclease-free water. The cDNA was sheared ultrasonically to 200–250 base pair fragments (Georgia Genomics Facility at University of Georgia), and TruSeq libraries (Illumina Inc., San Diego, CA) were constructed for paired-end (2 x 150) sequencing using the Illumina MiSeq sequencing platform (Illumina Inc., San Diego, CA).

Following sequencing, paired-end Illumina sequences were trimmed and filtered with trimfastq from the SEAStAR v0.4.8 software package (http://armbrustlab.ocean.washington.edu/seastar) to a minimum length of 100 nucleotides, a minimum correct read probability of 1%, and an entropy filter of 3.0. Overlapping trimmed read pairs were merged using FLASH v1.2.6 (Magoč & Salzberg 2011), setting individual read size to 150, the paired read length to 250, and paired read length standard deviation to 25. Ribosomal RNA reads were filtered out using SortMeRNA (v1.8; (Kopylova et al. 2012) with Silva eukaryotic 18S and 28S databases.

KEGG annotations

One million paired reads per replicate were selected to identify diatom sequences in a subset of the four metatranscriptome samples that were pre-filtered with 53 μm mesh. In cases with fewer than one million paired reads available for a sample, singlet reads were used in addition. BLASTX was used to compare the four metatranscriptome sequence subsets to a comprehensive custom database of known proteins that included the diatom genomes: *Thalassiosira pseudonana* (http://genome.jgi.doe.gov/Thaps3/Thaps3.home.html), *Fragilariopsis cylindrus* (http://genome.jgi.doe.gov/Fracy1/Fracy1.home.html), *Pseudo-nitzschia multiseries* (http://genome.jgi.doe.gov/Psemu1/Psemu1.home.html), *Thalassiosira oceanica* (RefSEq NC_014808.1), and *Phaeodactylum tricornutum* (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html), six-frame translated *Pseudo-nitzschia granii* and *Pseudo-nitzschia australis* transcriptomes (CAMERA, camera.calit2.net), 27 Moore Marine Microbiology Initiative (MMI)-supported transcriptomes (Marine Microbial Eukaryotic Transcriptome Sequencing Project, National Center for Genome Resources; supplemental Table 3.1), and three additional MMI-supported samples (*Bolidomonas pacifica*, ID: MMETSP0785; *Rhodosorus marinus*, ID: MMETSP0011; and *Pseudopedinella elastica*, ID: MMETSP1068). Those sequences that displayed an e-value ≤ than 1×10^{-5} were compared using BLASTX with the 06-14-2013 NCBI non-redundant (NR) protein database. Those sequences that displayed a lower e-value to a non-diatom protein in NR than to the diatom database were considered non-diatom sequences and discarded from further analysis.

KEGG annotations were run only on the diatom component of the four read subsets (supplemental Table 3.2). These sequences were compared to the 2013-06-24 KEGG peptide database at an e-value cutoff ≥ 1×10^{-5}. KEGG orthologous group identifiers (KO IDs) were

assigned to sequences based on the best significant hit with a KO annotation (Kanehisa et al. 2012; Kanehisa & Goto 2000).

Phylogenetic analyses

Amino acid alignments were created using Muscle (Edgar 2004) for each predicted protein; they included diatom sequences from NCBI and the in-house database used for the KEGG annotations (see above; supplemental Table 3.1). Alignments were trimmed in Geneious Pro 5.0.3 to the same length. The amino acid (aa) alignment lengths were 149 aa for Tp4888, 177 aa for ferritin; 406 aa for nitrate transporter; 495 aa for ammonium transporters; and 772 aa for nitrate reductase. Maximum-likelihood phylogenetic trees were created for each protein using FastTree on default settings (Price et al. 2009; 2010).

The phylogenetic sequence placement program Pplacer was used to quantify transcription for these key genes at Station P1 and P8. The four P1 metatranscriptomes and the two P8 metatranscriptomes were translated in six frames. HMMer (hmmer.janelia.org) was used to perform HMM searches on six-frame translations of the six metatranscriptomes. The translated metatranscriptome sequences that recruited to the amino acid reference alignment were placed on the phylogenetic trees using Pplacer on default settings with the "keep-at-most" function set to 1 to ensure that each sequence was only placed on one edge of the tree (Matsen et al. 2010). The clust feature of guppy was used to compare the number of recruited sequences to each edge for the two P8 replicates to the number of sequences for the four P1 replicates.

Sequences at each edge of the phylogenetic trees were totaled, and then sequence totals were normalized to library size. If an edge was clearly assigned to a centric or pennate group, it was binned accordingly. If a sequence was assigned to an edge that contained both centric and pennate diatoms, it was defined as a mixed clade. A majority of sequences were assigned to

either a centric or pennate clade. The term centric includes basal centric diatoms, as well as bi/multipolar centric diatoms.

3.5 Results

Water column biogeochemical properties

Macro- and micro-nutrient concentrations were measured at the seven Line P sites over the course of five days. Sampling locations crossed the continental shelf break located between stations P2 and P3 at ~ 200 m depth (Figure 3.1). Stations were chosen based on SeaFlow (underway flow cytometer)-based differences in phytoplankton community composition (Ribalet et al. 2010); http://armbrustlab.ocean.washington.edu/resources/seaflow/). Surface (5 m) nitrate, ammonium, and phosphate concentrations were < 1 μM at all stations; silicate concentrations ranged from a low of 1.4 μM at P3 to a high of 8.1 μM at P6 and were nearly identical at P1 and P8 (3.45 μM and 3.43 μM, respectively; Figure 3.2A−D). Iron concentrations at 10−20 m were highest at P1 (1.28 nM) and lowest at P8 (0.270 nM) (Figure 3.2E, adapted from Vedamati, 2013). The mixed layer depth (MLD) was estimated based on the location of the thermocline, and varied between 20 and 40 m along the transect. Ammonium and phosphate concentrations that were averaged through the mixed layer remained < 1 μM, while nitrate concentrations increased to 1.7 μM at P2. Averaged nitrate concentrations in the upper mixed layer were 1.5 μM at P1 and 1.4 μM at P8.

Acoustic Doppler Current Profiler data showed a shift in meridional currents between Station P5 and P6 with currents (in the upper 100 m) changing from southerly (P5) to northerly at P6, and then back to southerly at P8. This pattern corresponded with a slight upward shoaling of the silicate, ammonium, phosphate and iron nutriclines, as well as a slight decrease in salinity.

No obvious changes in temperature were observed between P5 and P6. Geostrophic currents based on satellite data (http://www.aoml.noaa.gov/phod/dataphod/work/trinanes/INTERFACE/) indicated that the cruise transect may have intersected the edge of an anticyclonic eddy, possibly of coastal origin. Separately, a strong salinity gradient from Station P1 to P2 was observed, increasing from < 30.0 at P1 to > 31.5 at P2.

Chlorophyll *a* concentrations and diatom distributions

Surface chlorophyll *a* concentrations were highest at the coastal station (P1; 2.6 μg L^{-1}) (Figure 3.3A), where the highest concentration of diatoms was also recorded (1.0 x 10^5 cells mL^{-1}) (Figure 3.3B). Total chlorophyll *a* decreased to 0.78 μg L^{-1} at P2 and few diatom cells were detected in the Lugol's sample. Chlorophyll *a* remained low at P3 (0.65 μg L^{-1}), and then increased to 1.2 μg L^{-1} at P4 and 1.8 μg L^{-1} at P5. The second highest concentration of diatom cells was recorded at P4 (5.4 x 10^4 cells mL^{-1}), and then diatom numbers decreased by more than 50% to < 1.5 x 10^4 cells mL^{-1} at P5, P6 and P8 (Figure 3.3B). At Station P8, the surface chlorophyll *a* concentration was comparable (1.48 μg L^{-1}) to P1, even though the diatom concentration was an order of magnitude lower than at P1. The smaller phytoplankton size fractions were likely contributing to the measured surface chlorophyll *a* concentrations, particularly at Station P6, where few diatoms were counted. Higher concentrations of *Synechococcus* and *Bathycoccus* were measured at P6 compared to at P1 (Ribalet and Worden, personal communications). Surface chlorophyll *a* concentrations were likely an underestimation of *in situ* values because they did not include chlorophyll *a* from phytoplankton < 0.7 μm in size.

Pennate diatoms dominated at Station P1, comprising 68% of the diatom community; all other stations were dominated by centric diatoms that included *Leptocylindrus sp.*, *Chaetoceros sp.*, *Guinardia sp.* and *Cerataulina sp*. At Station P8, the dominant centric diatom was

Dactyliosolen sp., and the most abundant pennate diatom was *Pseudo-nitzschia granii* (based on ARISA analysis). Cell size varied for the dominant diatoms at each station, ranging from < 50 μm to > 100 μm (Figure 3.3B, inset). Biogenic silica (BSi) concentrations mirrored the diatom cell count data (Figure 3.3B). The highest BSi concentration was 4.5 μM at P1, and the lowest BSi concentration was 0.2 μM at P2. At Station P1, automated ribosomal intergenic spacer analysis (ARISA) identified the bloom-forming diatom as *Pseudo-nitzschia pungens* (Figure 3.3C). *Pseudo-nitzschia pungens* was found in surface water and/or at the chlorophyll *a* maximum at every station; it was the dominant *Pseudo-nitzschia* species at Stations P1–P4. From P5–P8, *Pseudo-nitzschia granii* was the dominant *Pseudo-nitzschia* species. The most diverse *Pseudo-nitzschia* assemblage was recorded at P6, where it included *P. pungens*, *P. granii*, *P. australis*, *P. turgiduloides*, and *P. delicatissima* (Figure 3.3C).

To gain additional taxonomic information about the diatom community, 18S analysis was conducted on surface water DNA samples. The diatoms (*Bacillariophyta*, Figure 3.3D) comprised 26–44% of the eukaryotic community at each station, and pennate versus centric diatom 18S sequence distribution were similar to the cell count data (Figure 3.3E). The diatom community composition at P4 and P6 based on microscopic cell counts differed from the composition based on 18S sequences. The 18S data indicated that the pennate diatoms were dominant at both stations, whereas cell counts indicated that the centric diatoms were dominant (Figure 3.3B). Furthermore, the species identified using the 18S data differed substantially at the genus level from cell counts and the ARISA data. As an example, cell counts and ARISA data showed that *Pseudo-nitzschia (pungens)* was the dominant diatom in the P1 phytoplankton assemblage; 18S analysis indicated that *Cylindrotheca sp.* was the dominant genus. Although both organisms are pennate diatoms, they differ morphologically from one another; we did not

observe a high concentration of *Cylindrotheca sp*. in our cell count data. The taxonomic resolution of the 18S data was broadened to display order-level information, and indicated that the P1 18S diatom community was dominated by the order Bacillariale*s* that includes both *P. pungens* and *Cylindrotheca sp*.

Metatranscriptome analysis and KEGG annotations

The MiSeq-sequenced metatranscriptomes ranged in size from eleven to twenty-two million reads per replicate (Table 3.1). Reads totals were reduced by < 10% following quality control. Read lengths ranged from 148 to 149 nucleotides; once paired, read lengths ranged between 235 and 266 nucleotides. One million paired reads were sampled from each pre-filtered (< 53 μm) replicate at P1 and at P8 (N = 4); from these subsets, 14–20% of the reads were derived from diatoms (supplemental Table 3.2).

The diatom-derived sequences from the above read subset were annotated using the KEGG database; 73–81% of the diatom-derived sequences had a KEGG annotation, and 22–28% had KEGG module annotations (supplemental Table 3.2). The relative abundance for KEGG module annotations differed between P1 and P8: there was a greater relative abundance of KEGG modules at P1 than at P8 for central carbohydrate metabolism (12.9% versus 10.0%), carbon fixation (6.5% versus 4.0%), and photosynthesis (3.3% versus 2.3%) (Figure 3.4). At Station P8 versus P1, there was a greater relative abundance of sequences related to ATP synthesis (7.0% versus 4.8%), RNA processing (3.1% versus 1.8%), the proteasome (2.9% versus 1.6%), the spliceosome (1.7% versus 1.0%), other carbohydrate metabolism (1.8% versus 1.1%) and protein processing (1.5% versus 1.0%). There was also a greater number of methane metabolism transcripts at P1 compared to P8 (2.4% versus 1.5%); ferredoxin hydrogenase is a component of this KEGG module and a known indicator of nutrient-replete growth in diatoms

(Whitney et al. 2011). Each KEGG module constituted < 15% of the module annotations, except for the ribosome module, which represented > 40% of the module annotations (43.1% at P1; 41.1% at P8).

Phylogeny and transcript abundances for select diatom metabolism sequences

Phylogenetic trees were generated for ferritin (Figure 3.5), nitrate reductase (Figure 3.6), ammonium transporters (Figure 3.7), nitrate transporters (Figure 3.8), and Tp4888 (Figure 3.9), previously identified in Chapter 2 as being connected to nutrient-replete growth in diatoms. Sequences (normalized to library size) that were phylogenetically related to one of the above proteins were mapped on internal or external edges on the tree. An edge is defined as a location on the tree where two branches split (internal edge) or as a leaf on the tree (external edge). Sequences that mapped to an external edge of a tree were most closely related to that diatom sequence compared to other known diatom sequences (Matsen et al. 2010; Durkin et al. 2012).

Ferritin, an iron-storage protein, formed a phylogenetic tree with no distinct pennate-only or centric-only clades (Figure 3.5A). There was substantial variability in the distribution of sequences between stations and phylogenetic branches (Figure 3.5B). The greatest number of sequences from P1 relative to P8 was most closely related to the pennate diatom, *Pseudo-nitzschia sp.* (edge 32; Figure 3.5B). The greatest number of P8 sequences relative to P1 was most closely related to the centric diatom *Minutocellus sp*. (edge 45). The highest number of sequences at P1 and P8 were most closely related to *Nitzschia sp*. (pennate; edge 47). Overall, there was a greater abundance of sequences at P1 (relative to P8) that clustered with known pennate sequences and the mixed taxonomy clade (Figure 3.10A and B). In contrast, the centric diatoms recruited slightly more sequences at P8 than P1 (Figure 3.10C). The number of

sequences that recruited to the ferritin tree was the lowest of the five phylogenetic trees examined.

We quantified sequences that recruited to the nitrate reductase phylogenetic tree (Figure 3.6A). This gene encodes for an essential enzyme that requires iron as a cofactor and is involved in the reduction of nitrate to nitrite in diatoms. The phylogenetic tree displayed two distinct clades of sequences for the centric diatoms and the pennate diatoms. Multiple centric and pennate edges recruited more sequences at P1 compared to P8, including sequences most closely related to the pennate diatom *Pseudo-nitzschia sp.* (edge 41) and to the centric diatom *Detonula sp.* (edge 15; Figure 3.6B). Sequences similar to several centric and pennate diatoms were more abundant at P8 compared to P1, including an internal centric edge (connected to *Rhizosolenia sp*. and *Coscinodiscus wailesii*; edge 2) and an external edge for the pennate diatom *Phaeodactylum tricornutum* (edge 33). Sequences most closely related to *Cylindrotheca closterium* (pennate) were the highest at both stations (edge 36). Several centric diatoms also recruited a large number of similar sequences at both stations, including *Ditylum brightwellii*, (edge 25) and an internal edge that joined *Corethron hystrix* and *Aulacoseira subarctica* (edge 8) (Figure 3.6B). Total sequences were nearly identical between Station P1 and P8 for the centric, pennate and mixed taxonomic clades (Figure 3.10). Overall, the number of sequences that recruited to the nitrate reductase tree was an order of magnitude higher than sequences that recruited to the ferritin tree.

Phylogenetic trees with recruited sequences were also generated for the ammonium transporters, nitrate transporters and Tp4888. The ammonium transporter tree consisted of three distinct clades: pennate, centric and mixed (Figure 3.7A); it was the largest tree generated for the five proteins in terms of sequences from known diatoms. The edges that recruited the greatest number of closely related sequences at P1 relative to P8 were associated with the pennate diatom

P. pungens (edge 115) and the centric diatom *Leptocylindrus danicus danicus* (edge 132) (Figure 3.7B). The majority of the sequences were higher at P8 than P1 (Figure 3.10). The highest abundances of sequences at P8 were most closely related to an internal pennate edge for *Striatella sp*. and *Pseudo-nitzschia arenysensis* (edge 149), an external centric edge for *Minutocellus sp*. (edge 74), and two edges for *Pseudo-nitzschia sp*. (edges 119 and 121).

The nitrate transporter (Figure 3.8A) clustered into three distinct clades, similar to the ammonium transporter tree. A similar number of sequences recruited from P1 and P8 (Figure 3.8B; Figure 3.10). The greatest number of sequences at P1 (relative to P8) was most similar to an internal centric edge that included *Minutocellus polymorphus* and *Coscinodiscus wailesii* (edge 32); this edge also recruited the most sequences at both stations. The edge with the greatest number of P8 sequences (relative to P1) was for a mixed clade of diatoms that included *Thalassionema frauenfeldii* (pennate) and *Odontella sp*. (centric) (edge 102).

Finally, the Tp4888 phylogenetic tree (Figure 3.9A) resolved a clade of centric diatoms and a mixed clade of diatoms. A previous study (Chapter 2) showed Tp4888 expression to decrease under nitrate starvation; higher transcript abundances may be indicative of nutrient-replete growth among the diatoms. The greatest number of Tp4888 sequences at P1 (relative to P8) was most closely related to the centric diatom *Skeletonema costatum* (edge 1), followed by the pennate diatom *P. pungens* (edge 37), and an internal centric edge representing *Rhizosolenia setigera* and *Coscinodiscus wailesii* (edge 21) (Figure 3.9B). The greatest number of sequences at P8 (relative to P1) were most similar to an external edge (36) for *P. pungens*. The centric, pennate and mixed taxonomic clades all exhibited the same sequence recruitment patterns for Tp4888; more sequences were recruited at Station P1 compared to P8 (Figure 3.10).

Sequences were binned based on phylogeny as pennate, centric or mixed to highlight clade-specific differences (Figure 3.10). The pennate and centric diatoms displayed similar patterns in sequence recruitment: nitrate reductase sequence abundance exhibited minimal differences between stations; nitrate transporters and Tp4888 were both higher at P1, and ammonium transporter sequences were higher at P8. Abundances of ferritin reads differed between the pennate and centric diatoms; ferritin sequences were higher at P1 in the pennates and higher at P8 in the centrics. Overall, sequences that recruited to clades encompassing both pennate and centric diatoms constituted a smaller fraction of the total sequences recruited (Figure 3.10B).

The distribution of sequences along each edge of each tree was examined to determine whether there was a uniform transcriptional response within the pennate and centric clades. For all five proteins, external edges differed markedly in sequence recruitment (Figure 3.11A−E). A majority of the external and internal edges displayed minimal differences in sequence abundance between P8 and P1. The observed large-scale changes in transcript abundance were governed by a few edges and their associated species. For ferritin and nitrate reductase, the higher abundance of sequences at P1 was driven by sequences related to *Pseudo-nitzschia sp*. (Figure 3.11A and B; Figure 3.5 and 3.6). The large increase in ammonium transporter transcripts at P8 was also driven by *Pseudo-nitzschia sp*.-related sequences (Figure 3.7). Several other diatom species also recruited more ammonium transporter sequences at P8 compared to P1, indicating that the ammonium transporter transcriptional response was likely shared across the diatoms rather than governed by one or two dominant species (Figure 3.11C). The response of the nitrate transporters and Tp4888 at P1 was the result of a few key members rather than a reflection of the community-wide response (Figure 3.11D–E). At P1, species that recruited the most closely-related sequences

included the centric diatoms *Minutocellus polymorphus* and *Coscinodiscus wailesii* for the nitrate transporters, and *Skeletonema costatum* for Tp4888. The nitrate transporter P8 response was governed by a mixed clade of diatoms that included *Thalassionema frauenfeldii* (pennate) and *Odontella sp*.; high transcript abundances for Tp4888 at P8 were associated with *Pseudo-nitzschia sp*.

3.6 Discussion

Diatom bloom formation is largely controlled by nitrogen availability in coastal waters, and by iron availability in HNLC regions. The Line P transect provides an ideal environment for studying the effects of both nutrients on diatom metabolism because it spans a nutrient gradient from coast to open ocean. We observed that the diversity and abundance of individual species within the diatom assemblage shifted from coastal to offshore waters, with evidence for group- and species-specific transcript abundance patterns. Phylogenetic analyses revealed that different members of the diatom community were likely experiencing differing degrees of nutrient limitation at any given time. Findings from this study provide insight into the diverse metabolic responses among diatoms to their environment.

Bulk nutrient concentrations along Line P suggested that the diatom community was iron- and nitrogen-limited at Station P1 and P8 (Marchetti et al. 2009; Strzepek & Harrison 2004; Harrison et al. 1977). But, the observed abundance of *P. pungens* at Station P1 indicated that there were sufficient nutrients to fuel diatom bloom formation. To explain these differences, we propose two scenarios: the bloom was sampled as diatom growth began to decline, and all available nutrients had already been consumed; or the *P. pungens* population was taking up nutrients as quickly as nutrients were being delivered into the system, akin to a chemostat

(Hecky & Kilham 1988). Comparisons between *Pseudo-nitzschia*-like transcripts at P1 and P8 indicated that the P1 community was not nutrient limited. This finding supports scenario two. Given the metabolic capabilities of diatoms to immediate take up nutrients using both their high- and low-affinity transporters (Lomas & Glibert 2000), as well as their ability to store iron and nitrogen if they do not immediately require them for growth (Raven 1987; Marchetti et al. 2009), we would expect the nutrient residence times to be short-lived at P1, and for any nutrient pulses to be immediately assimilated into diatom cells as part of the chemostat scenario.

At the offshore station, decreases in cell abundance and iron concentration, as well as changes in the metatranscriptome data, indicated that the diatom community had become iron-limited; station P8 likely represented the beginning of the HNLC region (Ribalet et al. 2010). The higher relative abundance of transcripts related to cell repair, protein processing and protein degradation at P8 provided evidence that the community here was nutrient-limited. In contrast, at P1 there was a greater abundance of KEGG annotated transcripts connected to carbon fixation, central carbohydrate metabolism and photosynthesis, indicating that cells were actively growing and photosynthesizing. The KEGG annotation patterns observed in this study mirror a comparative study across three model diatoms harvested during nitrate starvation (Chapter 2). In the comparative study, an increase in cell repair genes and protein-related genes was indicative of increased intracellular nitrogen recycling as the cells experienced nitrate-limited growth. Our P8 metatranscriptomes likely captured a diatom response to iron limitation, which is similar to the diatom response to nitrogen limitation (Marchetti et al. 2012; Greene et al. 1991). Jenkins (personal communication) also observed higher transcript abundances for a *Thalassiosira oceanica* iron-stress gene index (Flavodoxin/Actin) at P8 compared to Station P1; an increase in the use of flavodoxin (relative to ferredoxin) allows diatoms to decrease their iron requirements

during iron-limited growth (Whitney et al. 2011). Overall changes in functional annotations and total transcript abundances indicated that the P8 diatom community was iron-limited, and that the P1 diatom community was nutrient-replete.

Phylogenetic analyses on individual diatom species revealed that several diatoms contributed to overall observed transcript abundance patterns and that the majority of the diatoms exhibited species-specific responses; not all diatoms were experiencing the same nutrient limitation scenario. At P1, *P. pungens* was the dominant diatom; consequently, a large number of *P. pungens*-like transcripts were observed, particularly for ferritin and Tp4888. Ferritin-like sequences were detected for other diatoms at P1, too; these constituted a smaller percentage of the observed P1 ferritin transcriptional response. We would expect transcripts for ferritin and Tp4888 to be elevated under nutrient replete growth, as cells take advantage of available iron and nitrogen in their environment (Marchetti et al. 2009). Because iron was in limited supply at P1, *P. pungens* may have had an advantage over other diatoms through its increased ability to produce ferritin and, subsequently, to concentrate and store fleeting iron (Marchetti et al. 2009). Alternatively, this large *Pseudo-nitzschia* signal may have been driven by greater cell abundances, which would inflate the observed transcriptional signal. Future analyses will use the trimmed mean of *M* values normalization method (TMM normalization) to account for changes in species abundance and sequence library size between P1 and P8, allowing for direct statistical comparisons between normalized metatranscriptome samples (Robinson et al. 2010; Marchetti et al. 2012).

The transcript abundance patterns at both stations also provided insight into the metabolic response of the less dominant community members to their environment. All diatoms were growing under the same bulk nutrient conditions, and each diatom exhibited its own

transcriptional response. Diatoms span several orders of magnitude in their cell size and possess vastly different morphologies; both features affect the ability of the cell to take up and store nutrients, as well as nutrient requirements for growth (Litchman et al. 2009; Lomas & Glibert 2000; Raven 1987; Dortch et al. 1985; Timmermans et al. 2001). Larger cells have higher nutrient requirements and are the first group to respond when limiting nutrients (e.g. iron and nitrogen) are added back into the system (Timmermans et al. 2004); smaller cells have a greater surface area:volume ratio and can subsist on much lower nutrient concentrations (Timmermans et al. 2001). *Pseudo-nitzschia pungens* may have been growing under nutrient-replete conditions at P1; growth of the less dominant diatoms was likely limited by nutrients. *Leptocylindrus* species exhibit a range of surface area:volume ratios, and some *Letocylindrus sp*. exhibit a much lower ratio compared to *P. pungens* (Maldonado et al. 2002; Alves-de-Souza et al. 2008; Leblanc et al. 2012). This difference may explain why *Leptocylindrus sp*., the dominant centric diatom at P1, did not achieve the same cell abundance as its dominant pennate counterpart. Both transcriptional and phenotypic differences within the group reflect the diversity of the diatoms and the specialized niche for each species in the marine environment.

We also documented group-specific responses to nutrient availability, where transcriptional similarities in how the centric and pennate diatoms regulate their metabolism and respond to changes in their environment highlight the importance of regulatory networks of genes in influencing diatom community dynamics (Ashworth et al. 2013). The first evidence for a pennate versus centric response was found in cellular trace metal concentrations: differences between iron:carbon cell ratios were observed for pennate and centric diatoms (pennate diatoms had higher ratios; Twinings, personal communication). These chemical differences supported observed transcriptional differences between the two groups (e.g. ferritin transcripts were higher

in the pennate diatoms at P1 than in the centric diatoms), as well as shifts in cell abundances from a pennate-dominated assemblage along the coast to a centric-dominated assemblage offshore. We hypothesize that the pennate diatoms, as a whole, are better equipped to capitalize on variable sources of iron, and to store it, as evidenced by their higher iron:carbon ratios along the Line P transect. This ability may help to explain why a *P. pungens* bloom formed at P1 when iron concentrations were low; it also supports the observation that pennate diatoms frequently bloom following iron fertilization in the open ocean (Boyd et al. 1996; Marchetti et al. 2012). Their underlying metabolic machinery may allow them to outcompete the centric diatoms in these oceanic regimes. Previous laboratory studies indicated that pennate diatoms may utilize different components of a diatom regulatory network to respond to changes in nutrient availability as compared to the centric diatoms (Chapter 2). This network likely encompasses a coordinated response across hundreds of genes in each diatom genome (Ashworth et al. 2013); differences in how the pennate and centric diatoms regulate genes within this network may help to explain why we observed shifts in distributions. Future analyses that assign taxonomy to all metatranscriptome reads will likely capture more of the genes that are linked with this response.

Presently, there are limitations in how individual diatom species are detected in the environment, and how these species are associated with changes in transcript abundance. In this study, some disparities between 18S sequences, cell count data, and phylogenetic analyses of the key metabolism genes were observed. As an example, *Pseudo-nitzschia sp.* diatoms were not captured in the 18S data; their dominance in cell count data and in the ARISA samples indicates that they were a key component of the diatom community at P1. This difference may be the result of inherent limitations in existing 18S databases; limited representation of diatom 18S sequences in eukaryotic databases may result in inaccurate taxonomic assignments at the species

level. Additional biases introduced through PCR that include primer specificity and varying rRNA copy number per cell also contribute to observed differences between 18S sequences and microscope counts (Not et al. 2009). A second explanation is the filtering methods used to capture the 18S sequences: larger diatom community members were likely excluded from pre-filtered (53 μm) 18S samples, whereas Lugol's samples were not pre-filtered. There are also limitations with conducting microscope counts; cell counts indicated that *Cylidrotheca* cells comprised < 0.5% of the P1 diatom community, while the 18S data indicated that *Cylindrotheca sp*. was the dominant diatom in the P1 community. The 18S data was supported by a large number of *Cylindrotheca*-like nitrate reductase sequences that were also detected at P1 using the phylogenetic method. Both *P. pungens* and *Cylindrotheca sp*. were present at Station P1; by combining the three approaches (18S, cell counts and transcript phylogeny), we gain a more accurate and comprehensive overview of the similarities and differences between the P1 and P8 diatom communities.

In this study, field metatranscriptomes quantified the responses of two separate diatom assemblages to changes in nitrogen and iron availability. It was clear that some diatoms at both stations were iron-limited for growth, but we did not observe a community-wide response to changes in nutrient availability. Rather, a few diatoms governed the observed, large-scale transcriptional and functional responses, even as a majority of the diatoms exhibited species-specific responses. Changes in cell abundances from a pennate-dominated community to a centric-dominated community, coupled to bulk measurements of iron:carbon per cell and changes in relative transcript abundances for key nitrogen and iron metabolism genes revealed a potential pennate versus centric metabolic response. Findings provide *in situ* examples of the

evolved metabolic responses that individual diatoms and groups of diatoms utilize to respond to environmental changes.

Diatoms are the most diverse group of marine phytoplankton, in part due to extensive genomic and transcriptomic differences (Chapter 2; Koester et al. 2012; Kooistra et al. 2007). The diverse transcriptional responses observed across the diatoms examined help to explain why they are one of the most successful marine phytoplankton groups (Falkowski et al. 2004; Kooistra et al. 2007). Diversity among species leads to greater niche differentiation, ultimately translating into greater group productivity because organisms are better equipped to take advantage of the nutrients in their environment (Cardinale, 2011; Crutsinger et al., 2006; Norberg et al., 2001; Tilman et al., 1997). This study provides a better understanding of diatom bloom formation and diatom distributions in the marine environment. Future studies, that focus on individual species will help to define each organism's niche, and determine how each species interacts with one another in these dynamic environments.

3.7 Acknowledgments

This research was conducted as part of the pilot study, Global scale Microbial Interactions across Chemical Surveys (GeoMICS), funded by the National Science Foundation. We would like to thank the captain and crew of the R/V Thomas G. Thompson for a successful sampling trip to Line P, as well as co-chief scientist Dr. Jim Moffett. Several conversations and meetings contributed to the synthesized data shown here, including personal communication with Dr. Benjamin Twinings (metals/cell ratio), Dr. Alexandra Worden (picoeukaryote distributions), Dr. Francois Ribalet (underway biological and chemical measurements), Dr. Bethany Jenkins (iron-stress diatom index), Dr. Jagruti Vedamati (trace metal analyses), and Dr. Peter Countway (18S samples).

We also thank Dr. Jody W. Deming, Dr. Anitra Ingalls, and Dr. Micaela Parker for feedback on this chapter, Andreas Krupke for assistance with data visualization, Megan Schatz for cell count analysis, Claire Ellis for ARISA sample preparation and processing, Helena van Tol for help with sample collection, and the Moran lab (University of Georgia) for collaborating on the project design, implementation, and data analysis.

3.8 References

Alves-de-Souza C, Gonzalez MT, Iriarte JL. (2008). Functional groups in marine phytoplankton assemblages dominated by diatoms in fjords of southern Chile. J Plankton Res 30:1233–1243.

Ashworth J, Coesel S, Lee A, Armbrust EV, Orellana MV, Baliga NS. (2013). Genome-wide diel growth state transitions in the diatom *Thalassiosira pseudonana*. Proc Natl Acad Sci in press.

Azam F, Fenchel T, Field J, Gray J, Meyereil L, Thingstad F. (1983). The ecological role of water-column microbes in the sea. Mar Ecol Prog Ser 10:257–263.

Batten SD, Crawford WR. (2005). The influence of coastal origin eddies on oceanic plankton distributions in the eastern Gulf of Alaska. Deep-Sea Res Pt II 52:19–19.

Booth BC, Lewin J, Postel JR. (1993). Temporal variation in the structure of autotrophic and heterotrophic communities in the sub-arctic Pacific. Progr Oceanogr 32:57–99.

Boyd PW, Muggli DL, Varela DE, Goldblatt RH, Chretien R, Orians KJ, et al. (1996). In vitro iron enrichment experiments in the NE subarctic Pacific. Mar Ecol Prog Ser 136:179–193.

Brzezinski MA, Nelson DM. (1986). A solvent extraction method for the colorimetric determination of nanomolar concentrations of silicic acid in seawater. Mar Chem 19:139–151.

Brzezinski MA, Villareal TA, Lipschultz F. (1998). Silica production and the contribution of diatoms to new and primary production in the central North Pacific. Mar Ecol Prog Ser 167:89–104.

Capone DG, Bronk DA, Mulholland MR, Carpenter EJ, eds. (2008). Nitrogen in the Marine Environment. 2nd ed. Elsevier Inc.: Oxford.

Carpenter EJ, Guillard R. (1971). Intraspecific differences in nitrate half-saturation constants for three species of marine phytoplankton. Ecology 52:183–185.

Clemons MJ, Miller CB. (1984). Blooms of large diatoms in the oceanic, subarctic Pacific. Deep-Sea Res Pt I 31:85–95.

Dortch Q, Clayton J, Thoresen S. (1985). Nitrogen storage and use of biochemical indices to assess nitrogen deficiency and growth rate in natural plankton populations. J Mar Res 43:437–464.

Dugdale R, Wilkerson F. (1998). Silicate regulation of new production in the equatorial Pacific upwelling. Nature 391:270–273.

Durkin CA, Marchetti A, Bender SJ, Truong T, Morales R, Mock T, et al. (2012). Frustule-related gene transcription and the influence of diatom community composition on silica precipitation in an iron-limited environment. Limnol Oceanogr 57:1619–1633.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

Eppley R, Rogers JN, McCarthy J. (1969). Half-saturation constants for uptake of nitrate and ammonium by marine phytoplankton. Limnol Oceanogr 14:912.

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield OM, et al. (2004). The evolution of modern eukaryotic phytoplankton. Science 305:354–360.

Fitzwater SE, Coale KH, Gordon RM, Johnson KS, Ondrusek ME. (1996). Iron deficiency and phytoplankton growth in the equatorial Pacific. Deep-Sea Res Pt II 43:995–1015.

Geider RJ, La Roche JJ. (2002). Redfield revisited: variability of C : N : P in marine microalgae and its biochemical basis. Eur J Phycol 37:1–17.

Greene RM, Geider RJ, Falkowski PG. (1991). Effect of iron limitation on photosynthesis in a marine diatom. Limnol Oceanogr 36:1772–1782.

Harris SL, Varela DE, Whitney FW, Harrison PJ. (2009). Nutrient and phytoplankton dynamics off the west coast of Vancouver Island during the 1997/98 ENSO event. Deep-Sea Res Pt II 56:16–16.

Harrison PJ, Boyda PW, Varela DE, Takeda S, Shiomoto A, Odate T. (1999). Comparison of factors controlling phytoplankton productivity in the NE and NW subarctic Pacific gyres. Progr Oceanogr 43:205–234.

Harrison PJ, Conway H, Holmes R, Davis C. (1977). Marine diatoms grown in chemostats under silicate or ammonium limitation. 3. Cellular chemical composition and morphology of *Chaetoceros debilis*, *Skeletonema costatum*, and *Thalassiosira gravida*. Mar Biol 43:19–31.

Hecky RE, Kilham P. (1988). Nutrient limitation of phytoplankton in freshwater and marine environments: A review of recent evidence on the effects of enrichment. Limnol Oceanogr 33:796–822.

Hubbard KA, Rocap G, Armbrust EV. (2008). Inter- and intraspecific community structure within the diatom genus *Pseudo-nitzschia* (Bacillariophyceae). J Phycol 44:637–649.

Kanehisa M, Goto S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40:D109–14.

Kooistra W, Gersonde R, Medlin LK, Mann DG. (2007). The origin and evolution of the diatoms: their adaptation to a planktonic existence. In:Evolution of Primary Producers in the Sea, Falkowski, PG & Knoll, AH, eds. Elsevier Academic Press: Amsterdam, pp. 210–225.

Kopylova E, Noé L, Touzet H. (2012). SortMeRNA: fast and accurate filtering of ribosomal

RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217.

Leblanc K, Arístegui J, Armand L, Assmy P, Beker B, Bode A, et al. (2012). A global diatom database - abundance, biovolume and biomass in the world ocean. Earth Syst Sci Data 4:149–165.

Lee J-M, Boyle EA, Echegoyen-Sanz Y, Fitzsimmons JN, Zhang R, Kayser RA. (2011). Analysis of trace metals (Cu, Cd, Pb, and Fe) in seawater using single batch nitrilotriacetate resin extraction and isotope dilution inductively coupled plasma mass spectrometry. Anal Chim Acta 686:9.

Litchman E, Klausmeier CA, Yoshiyama K. (2009). Contrasting size evolution in marine and freshwater diatoms. Proc Natl Acad Sci 106:2665–2670.

Lomas M, Glibert PM. (2000). Comparisons of nitrate uptake, storage, and reduction in marine diatoms and flagellates. J Phycol 36:903–913.

Lommer M, Specht M, Roy A-S, Kraemer L, Andreson R, Gutowska MA, et al. (2012). Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. Genome Biol 13:R66–R66.

Magoč TT, Salzberg SLS. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963.

Maldonado MT, Hughes MP, Rue EL, Wells ML. (2002). The effect of Fe and Cu on growth and domoic acid production by *Pseudo-nitzschia multiseries* and *Pseudo-nitzschia australis*. Limnol Oceanogr 47:515–526.

Marchetti A, Harrison PJ. (2007). Coupled changes in the cell morphology and the elemental (C, N, and Si) composition of the pennate diatom *Pseudo-nitzschia* due to iron deficiency. Limnol Oceanogr 52:2270–2284.

Marchetti A, Parker MS, Moccia LP, Lin EO, Arrieta AL, Ribalet FF, et al. (2009). Ferritin is used for iron storage in bloom-forming marine pennate diatoms. Nature 457:467–470.

Marchetti A, Schruth DM, Durkin CA, Parker MS, Kodner RB, Berthiaume CT, et al. (2012). Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. Proc Natl Acad Sci 109:1–9.

Marchetti A, Sherry ND, Kiyosawa H, Tsuda A, Harrison PJ. (2006). Phytoplankton processes during a mesoscale iron enrichment in the NE subarctic Pacific: Part I—Biomass and assemblage. Deep-Sea Res Pt II 53:2095–2113.

Marchetti A, Varela DE, Lance VP, Johnson Z, Palmucci M, Giordano M, et al. (2010). Iron and silicic acid effects on phytoplankton productivity, diversity, and chemical composition in the central equatorial Pacific Ocean. Limnol Oceanogr 55:11–29.

Matsen FA, Kodner RB, Armbrust EV. (2010). pplacer: linear time maximum-likelihood and

Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11:538–554.

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

Not FF, Del Campo JJ, Balagué VV, de Vargas CC, Massana RR. (2009). New insights into the diversity of marine picoeukaryotes. PLOS One 4:e7143–e7143.

Parsons TR, Maita Y, Lalli CM. (1984). A manual of chemical and biological methods for seawater analysis. Pergamon: Oxford.

Peña MA, Bograd SJ. (2007). Time series of the northeast Pacific. Progr Oceanogr 75:115–119.

Peterson TD, Harrison PJ. (2012). Diatom dynamics in a long-lived mesoscale eddy in the northeast subarctic Pacific Ocean. Deep-Sea Res Pt I 65:157–170.

Price MN, Dehal PS, Arkin AP. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26:1641–1650.

Price MN, Dehal PSP, Arkin APA. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. PLOS One 5:e9490–e9490.

Raven JA. (1987). The role of vacuoles. New Phytol 106:357–422.

Ribalet FF, Marchetti A, Hubbard KA, Brown K, Durkin CA, Morales R, et al. (2010). Unveiling a phytoplankton hotspot at a narrow boundary between coastal and offshore waters. Proc Natl Acad Sci 107:16571–16576.

Robinson MD, McCarthy DJ, Smyth GK. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

Schlitzer R. (2002). Interactive analysis and visualization of geoscience data with Ocean Data View. Comput Geosci 28:1211–1218.

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

Strzepek RFR, Harrison PJP. (2004). Photosynthetic architecture differs in coastal and oceanic diatoms. Nature 431:689–692.

Taylor F, Haigh R. (1996). Spatial and temporal distributions of microplankton during the summers of 1992- 1993 in Barkley Sound, British Columbia, with emphasis on harmful species. Can. J. Fish. Aquat. Sci. 53:2310–2322.

Timmermans KR, Gerringa LJA, de Baar H, van der Wagt B, Veldhuis M, de Jong J, et al.

(2001). Growth rates of large and small Southern Ocean diatoms in relation to availability of iron in natural seawater. Limnol Oceanogr 46:260–266.

Timmermans KR, van der Wagt B, de Baar HJW. (2004). Growth rates, half-saturation constants, and silicate, nitrate, and phosphate depletion in relation to iron availability of four large, open-ocean diatoms from the Southern Ocean. Limnol Oceanogr 49:2141–2151.

UNESCO. (1994). Protocols for the Joint Global Ocean Flux Study (JGOFS) core measurements. In:IOC Manual and Guides 29. Intergovernmental Oceanographic Commission: Paris, p. 170.

Vedamati J. (2013) Distributions of total dissolved iron, manganese, zinc and copper along Line-P in the North eastern Pacific Ocean. Moffett, JW, ed. University of Southern California: Los Angeles.

Whitney LP, Lins JJJ, Hughes MPM, Wells MLM, Chappell PDP, Jenkins BDB. (2011). Characterization of putative iron responsive genes as species-specific indicators of iron stress in Thalassiosiroid diatoms. Front Microbiol 2:234.

Table 3.1 Metatranscriptome sampling and processing information. Average read lengths were calculated following quality control.

	Pre-filter	Replicate	Volume filtered (L)	Time filtered	Illumina MiSEQ information			
					Raw reads	Reads: quality controlled	Average read length	Average paired read length
Station P1 *May 21, 2012*	53 µm	A	6.00	13:50−14:12	1.35×10^7	8.97×10^6	149	247
		B	5.50	14:46−15:07	1.45×10^7	8.90×10^6	149	235
	none	A	5.75	16:13−16:36	1.30×10^7	6.67×10^6	148	224
		B	5.40	16:42−17:05	1.67×10^7	8.35×10^6	149	208
Station P8 *May 17, 2012*	53 µm	A	4.20	17:40−18:00	2.20×10^7	1.53×10^7	148	261
		B	5.20	19:05−19:25	1.14×10^7	9.72×10^6	148	266

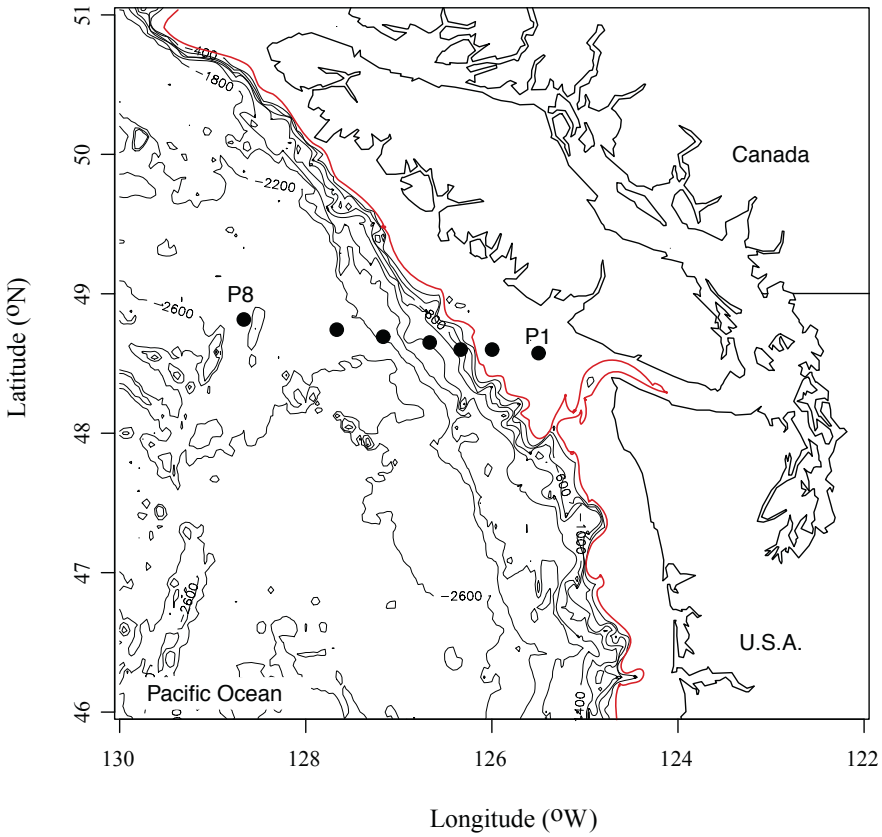


Figure 3.1 A map of Line P, showing the two stations sampled for the metatranscriptome and phylogenetic analyses (P1 and P8). The shelf break, 200 m depth between Station P2 and P3, is highlighted in red.

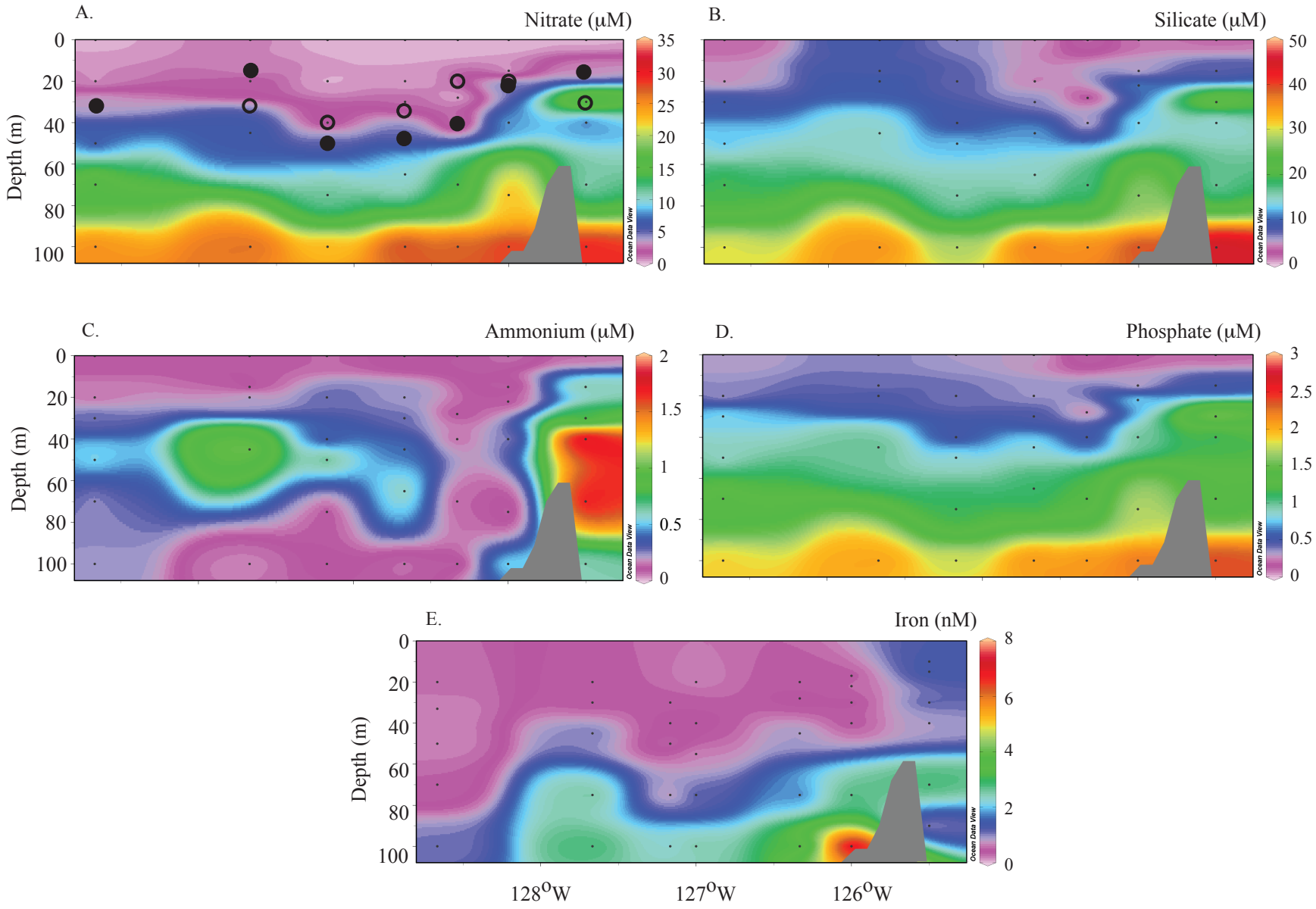
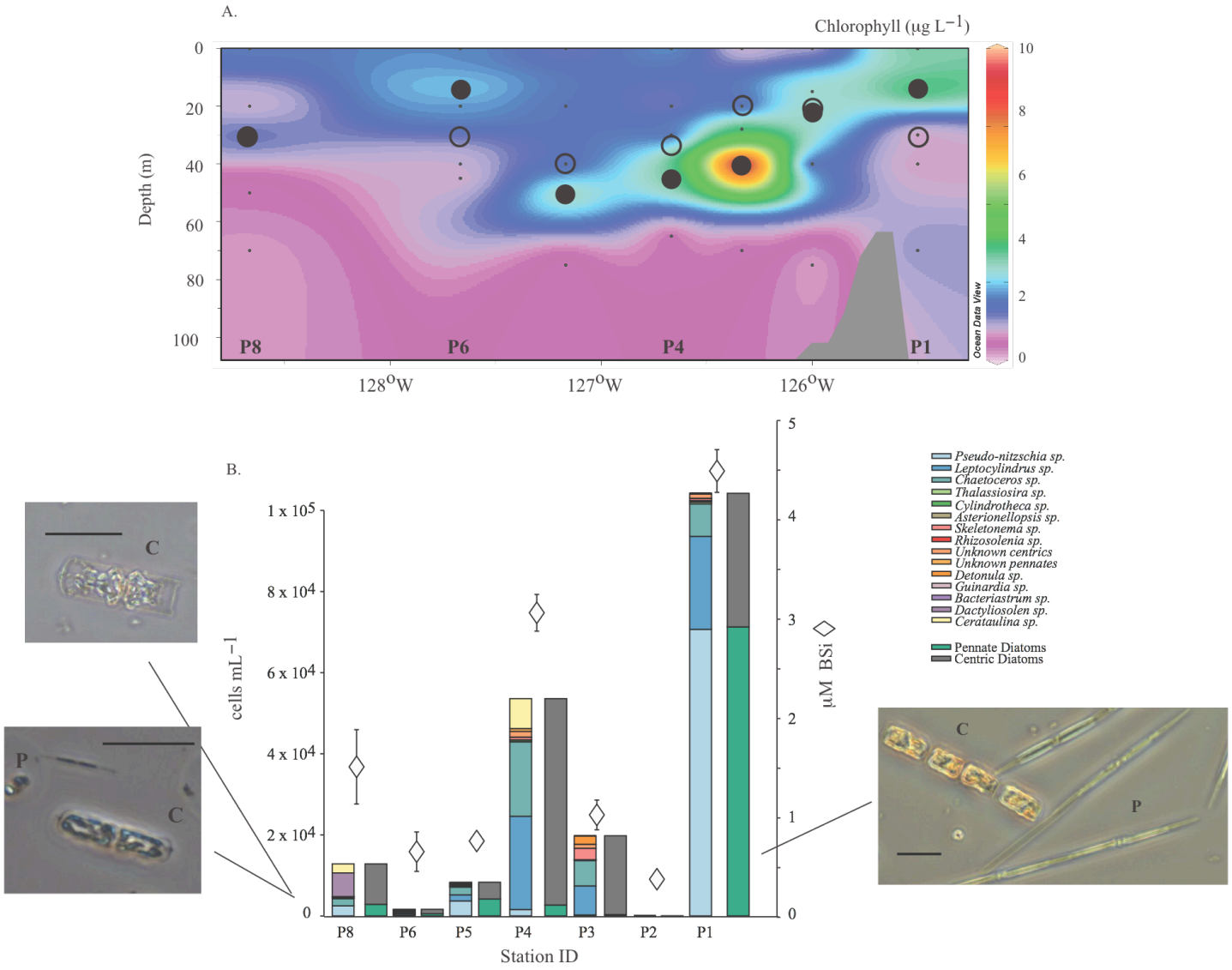

A.
Nitrate (µM)
B.
Silicate (µM)
C.
Ammonium (µM)
D.
Phosphate (µM)
E.
Iron (nM)
Depth (m)
0
20
40
60
80
100
35
30
25
20
15
10
5
0
50
40
30
20
10
0
2
1.5
1
0.5
0
3
2.5
2
1.5
1
0.5
0
8
6
4
2
0
128°W
127°W
126°W
Ocean Data View

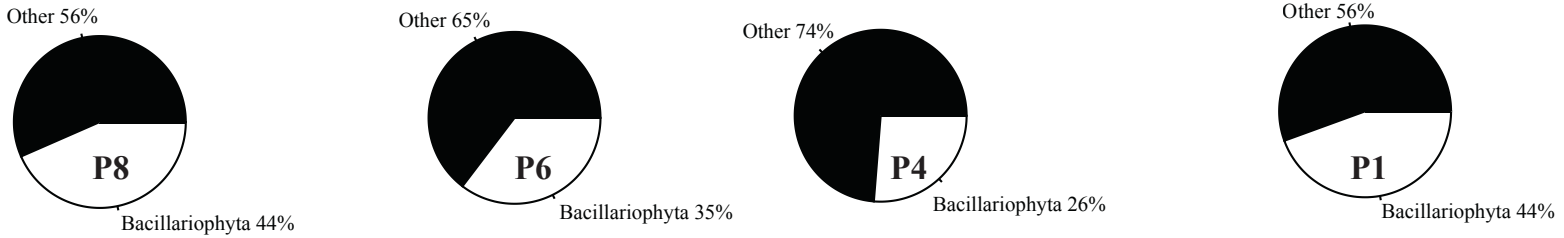
Figure 3.2 Ocean data view plots of nutrient concentrations in the upper 100 m along the Line P transect. A. Nitrate (mM), B. Silicate (mM), C. Ammonium (mM), D. Phosphate (mM), and E. Iron (nM). Black points highlight where nutrient samples were taken. Filled black circles indicate the chlorophyll maximum calculated using the CTD fluorescence profile and extracted chlorophyll *a* concentrations; open circles indicate the estimated mixed layer depth at each station. Color scales for nutrient concentrations are given to the right of each panel.


A.
Chlorophyll (µg L−1)
Depth (m)
0
20
40
60
80
100
10
8
6
4
2
0
P8
P6
P4
P1
128°W
127°W
126°W
Ocean Data View
B.
cells mL−1
1 x 10^5
8 x 10^4
6 x 10^4
4 x 10^4
2 x 10^4
0
µM BSi
5
4
3
2
1
0
P8
P6
P5
P4
P3
P2
P1
Station ID
Pseudo-nitzschia sp.
Leptocylindrus sp.
Chaetoceros sp.
Thalassiosira sp.
Cylindrotheca sp.
Asterionellopsis sp.
Skeletonema sp.
Rhizosolenia sp.
Unknown centrics
Unknown pennates
Detonula sp.
Guinardia sp.
Bacteriastrum sp.
Dactyliosolen sp.
Cerataulina sp.
Pennate Diatoms
Centric Diatoms
C
P
C
C
P

C.

	P8	P6	P5	P4	P3	P2	P1
P. pungens		19%	13%	69%	100% 62%	100%	100% 100%
P. granii	100% 100%	100% 15%	77% 100%	10%			
P. australis		24%	10%	21%	38%		
P. turgiduloides		15%					
P. delicatissima		27%					

D.



E.

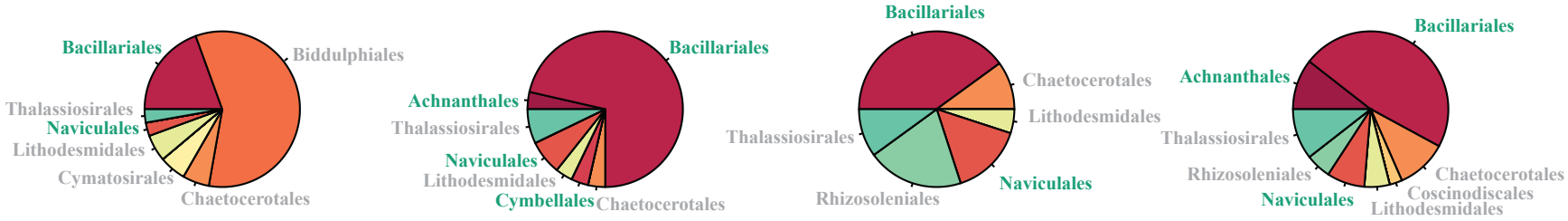


Figure 3.3 Changes in chlorophyll *a* concentrations, cell counts, biogenic silica (BSi), ARISA, and 18S ribosomal RNA abundance along the Line P transect (right, onshore; left, offshore). A. Chlorophyll *a* concentration (mg L^{-1}) in the upper 100 m; black points highlight where chlorophyll *a* samples were taken. Filled black circles indicate the chlorophyll maximum calculated using the CTD fluorescence profile and extracted chlorophyll *a* concentrations; open circles indicate the estimated mixed layer depth at each station. Color scale of concentrations is given to the right of the panel. B. Left y-axis: Cell counts identified by microscopy (cells mL^{-1}), grouped by genus and as "pennate" or "centric". Micrographs (inset) provide an example of the diatom community at P1 (right: *Pseudo-nitzschia pungens* ["P"], *Leptocylindrus sp*. ["C"]) and P8 (left top: *Cerataulina sp*. ["C"]; left bottom: *Pseudo-nitzschia granii* ["P"], *Dactyliosolen sp*. ["C"]). "C" identifies example centric diatoms; "P" identifies example pennate diatoms. Scale bar, 50 μm. Right y-axis: Biogenic silica concentrations (μM) measured at each station (diamond symbols). Error bars represent the standard deviation for three biological replicates; if no error bar is shown, standard deviations were masked by the symbol size. C. ARISA data for the *Pseudo-nitzschia* community. Green percentages indicate the relative abundance of each species in surface waters; blue percentages indicate the relative abundance of each species at the chlorophyll maxima; black box indicates that no species was detected in that sample. D. 18S ribosomal RNA for the *Bacillariophyta* phyla (white) as a percentage of the total eukaryotic community 18S rRNA (black; other eukaryotes), E. Distribution of 18S rRNA sequences for the *Bacillariophyta* phyla only. Organisms that comprised more than 1% of the *Bacillariophyta* community are labeled by order. Pennate diatom names are highlighted in teal; centric diatoms names in gray.

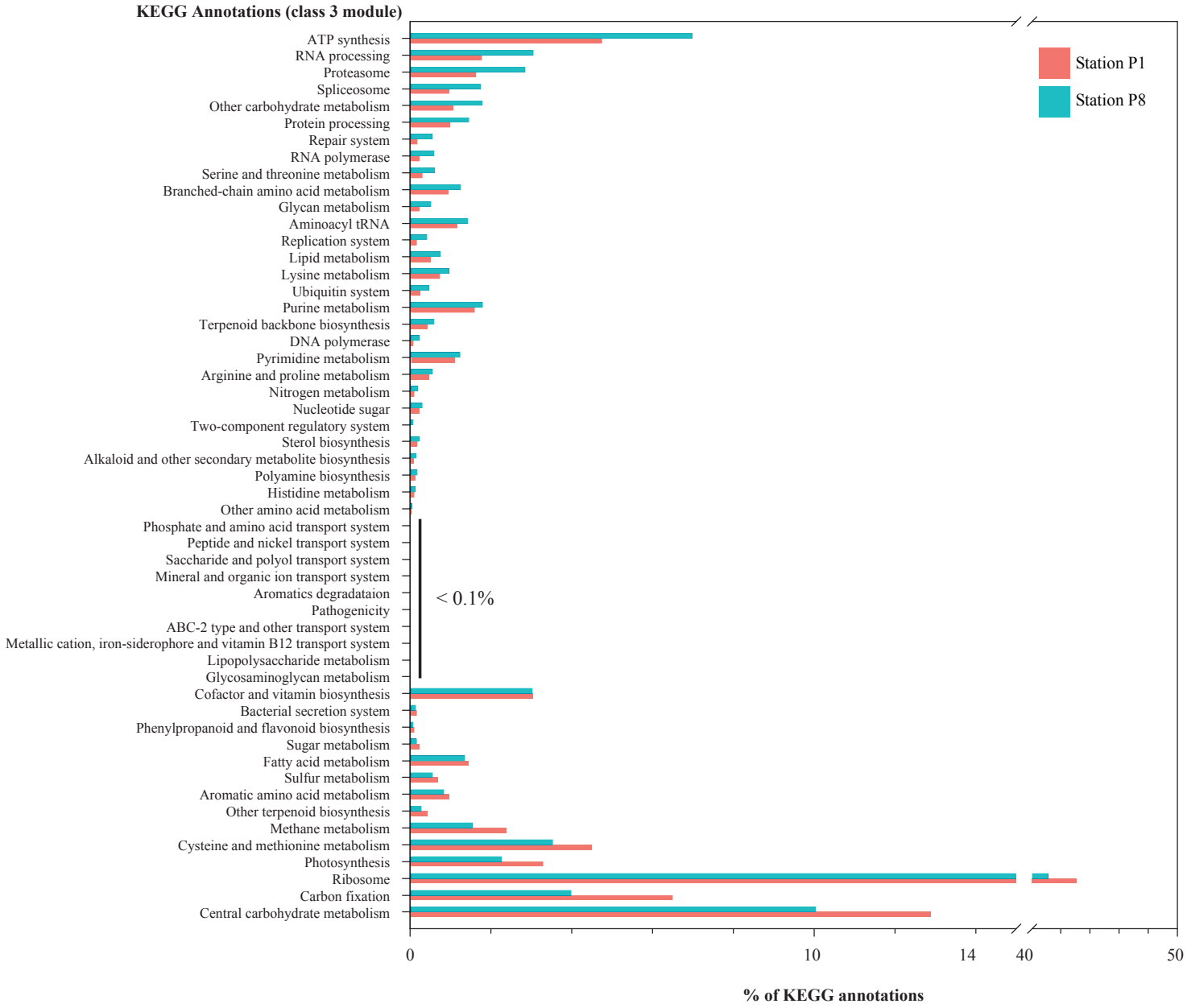

KEGG Annotations (class 3 module)
ATP synthesis
RNA processing
Proteasome
Spliceosome
Other carbohydrate metabolism
Protein processing
Repair system
RNA polymerase
Serine and threonine metabolism
Branched-chain amino acid metabolism
Glycan metabolism
Aminoacyl tRNA
Replication system
Lipid metabolism
Lysine metabolism
Ubiquitin system
Purine metabolism
Terpenoid backbone biosynthesis
DNA polymerase
Pyrimidine metabolism
Arginine and proline metabolism
Nitrogen metabolism
Nucleotide sugar
Two-component regulatory system
Sterol biosynthesis
Alkaloid and other secondary metabolite biosynthesis
Polyamine biosynthesis
Histidine metabolism
Other amino acid metabolism
Phosphate and amino acid transport system
Peptide and nickel transport system
Saccharide and polyol transport system
Mineral and organic ion transport system
Aromatics degradataion
Pathogenicity
ABC-2 type and other transport system
Metallic cation, iron-siderophore and vitamin B12 transport system
Lipopolysaccharide metabolism
Glycosaminoglycan metabolism
Cofactor and vitamin biosynthesis
Bacterial secretion system
Phenylpropanoid and flavonoid biosynthesis
Sugar metabolism
Fatty acid metabolism
Sulfur metabolism
Aromatic amino acid metabolism
Other terpenoid biosynthesis
Methane metabolism
Cysteine and methionine metabolism
Photosynthesis
Ribosome
Carbon fixation
Central carbohydrate metabolism
< 0.1%
Station P1
Station P8
0
10
14
40
50
% of KEGG annotations

Figure 3.4 KEGG annotations of diatom sequences at both stations P1 and P8. On the y-axis are class 3 KEGG module annotation ID: on the x-axis, the % of KEGG annotations. The % represents the number of KEGG annotations associated with that module relative to the total number of KEGG module annotations for Station P1 (pink bars) and Station P8 (blue bars). KEGG annotations between pre-filtered replicates at both stations were pooled. The data were then sorted based on differences in annotations between P1 and P8; KEGG modules with a greater representation at Station P8 are at the top of the graph; KEGG modules with a greater representation at Station P1 are at the bottom of the graph. Note the scale break along the x-axis. KEGG annotations were determined for a subset of the total metatranscriptome reads (see Methods).

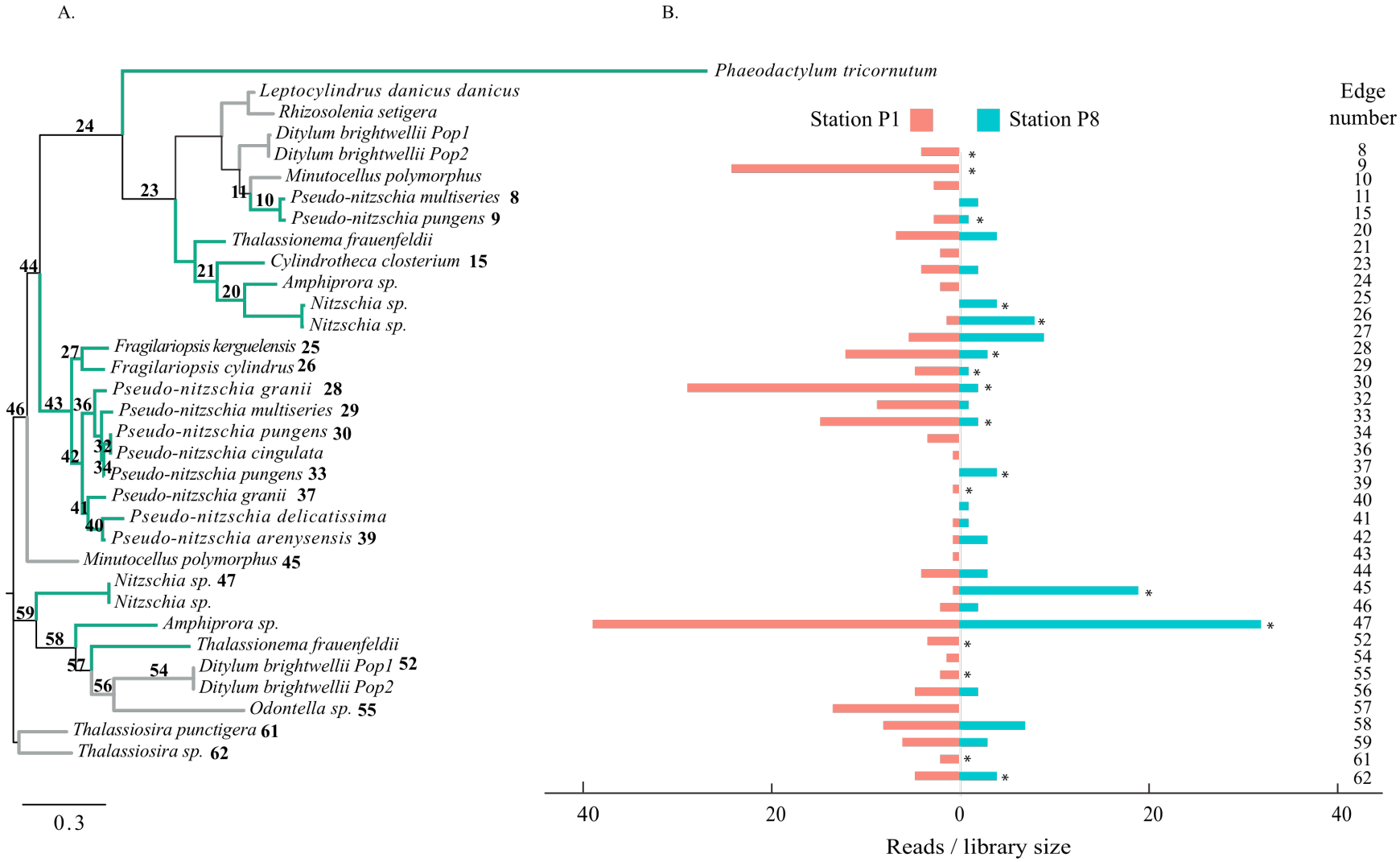

A.
B.
Phaeodactylum tricornutum
Leptocylindrus danicus danicus
Rhizosolenia setigera
Ditylum brightwellii Pop1
Ditylum brightwellii Pop2
Minutocellus polymorphus
Pseudo-nitzschia multiseries 8
Pseudo-nitzschia pungens 9
Thalassionema frauenfeldii
Cylindrotheca closterium 15
Amphiprora sp.
Nitzschia sp.
Nitzschia sp.
Fragilariopsis kerguelensis 25
Fragilariopsis cylindrus 26
Pseudo-nitzschia granii 28
Pseudo-nitzschia multiseries 29
Pseudo-nitzschia pungens 30
Pseudo-nitzschia cingulata
Pseudo-nitzschia pungens 33
Pseudo-nitzschia granii 37
Pseudo-nitzschia delicatissima
Pseudo-nitzschia arenysensis 39
Minutocellus polymorphus 45
Nitzschia sp. 47
Nitzschia sp.
Amphiprora sp.
Thalassionema frauenfeldii
Ditylum brightwellii Pop1 52
Ditylum brightwellii Pop2
Odontella sp. 55
Thalassiosira punctigera 61
Thalassiosira sp. 62
0.3
Station P1
Station P8
Edge number
Reads / library size
40
20
0
20
40

Figure 3.5 Maximum likelihood phylogenetic tree for diatom ferritin (A), and the corresponding distribution of Pplacer-recruited sequences across the tree (B). A. Pennate (teal) and centric (gray) branches are highlighted accordingly; branches that joined both pennate and centric diatoms are in black. Internal and external edges on the tree that recruited a metatranscriptome sequence are labeled with a unique number. B. Distribution of Pplacer-recruited sequences normalized to library size. Pink sequences, Station P1; blue sequences, Station P8. The edge number associated with each bar is displayed directly to the right and corresponds with the same edge number on the tree in A. External edges that recruited sequences are indicated with an asterisk. Scale bar represents number of amino acid changes per site.

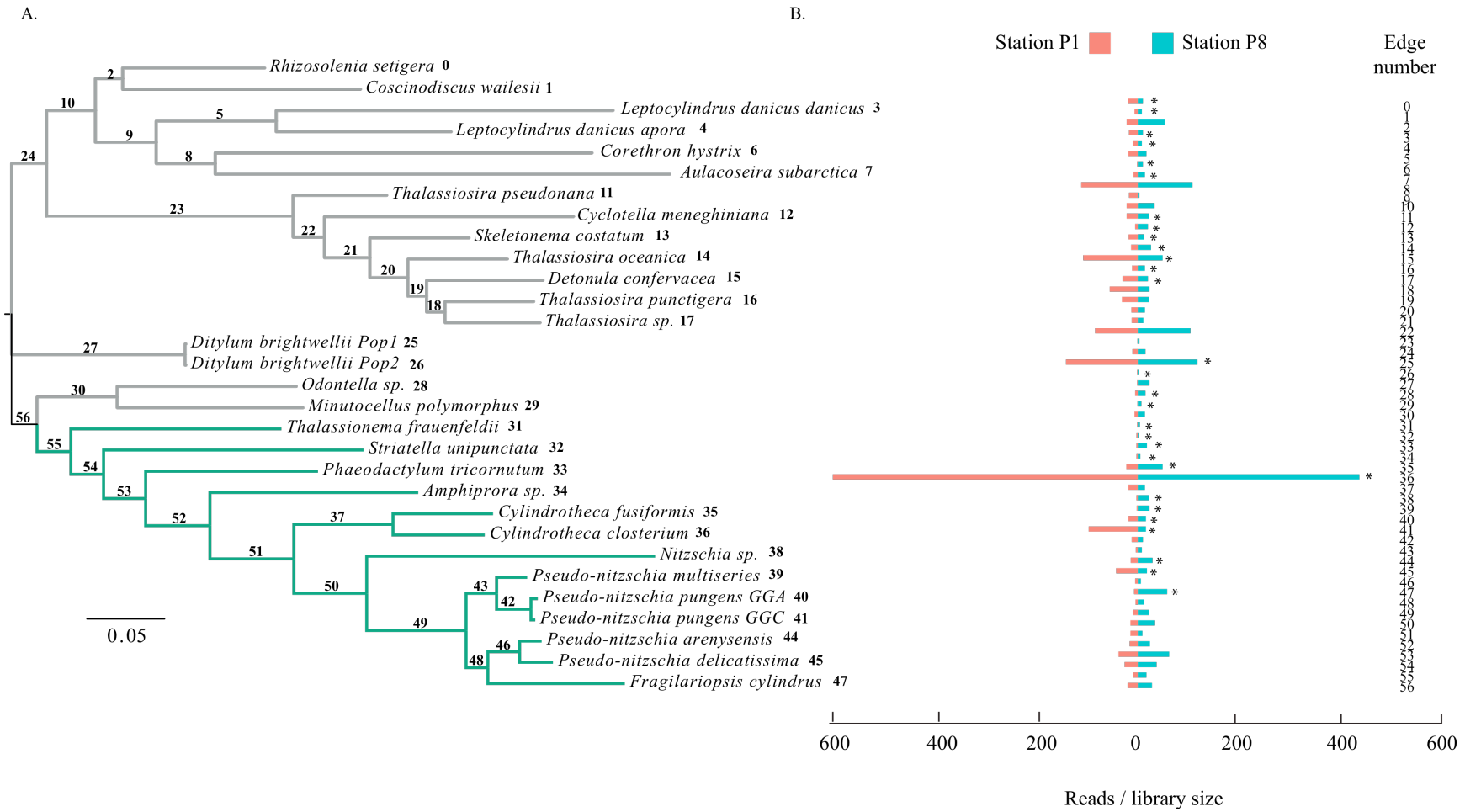

A.
Rhizosolenia setigera 0
Coscinodiscus wailesii 1
Leptocylindrus danicus danicus 3
Leptocylindrus danicus apora 4
Corethron hystrix 6
Aulacoseira subarctica 7
Thalassiosira pseudonana 11
Cyclotella meneghiniana 12
Skeletonema costatum 13
Thalassiosira oceanica 14
Detonula confervacea 15
Thalassiosira punctigera 16
Thalassiosira sp. 17
Ditylum brightwellii Pop1 25
Ditylum brightwellii Pop2 26
Odontella sp. 28
Minutocellus polymorphus 29
Thalassionema frauenfeldii 31
Striatella unipunctata 32
Phaeodactylum tricornutum 33
Amphiprora sp. 34
Cylindrotheca fusiformis 35
Cylindrotheca closterium 36
Nitzschia sp. 38
Pseudo-nitzschia multiseries 39
Pseudo-nitzschia pungens GGA 40
Pseudo-nitzschia pungens GGC 41
Pseudo-nitzschia arenysensis 44
Pseudo-nitzschia delicatissima 45
Fragilariopsis cylindrus 47
0.05
B.
Station P1
Station P8
Edge number
600
400
200
0
200
400
600
Reads / library size

Figure 3.6 Maximum likelihood phylogenetic tree for diatom nitrate reductase (A), and the corresponding distribution of Pplacer-recruited sequences across the tree (B). A. Pennate (teal) and centric (gray) branches are highlighted accordingly; branches that joined both pennate and centric diatoms are in black. Internal and external edges on the tree that recruited a metatranscriptome sequence are labeled with a unique number. B. Distribution of Pplacer-recruited sequences normalized to library size. Pink sequences, Station P1; blue sequences, Station P8. The edge number associated with each bar is displayed directly to the right and corresponds with the same edge number on the tree in A. External edges that recruited sequences are indicated with an asterisk. Scale bar represents number of amino acid changes per site.

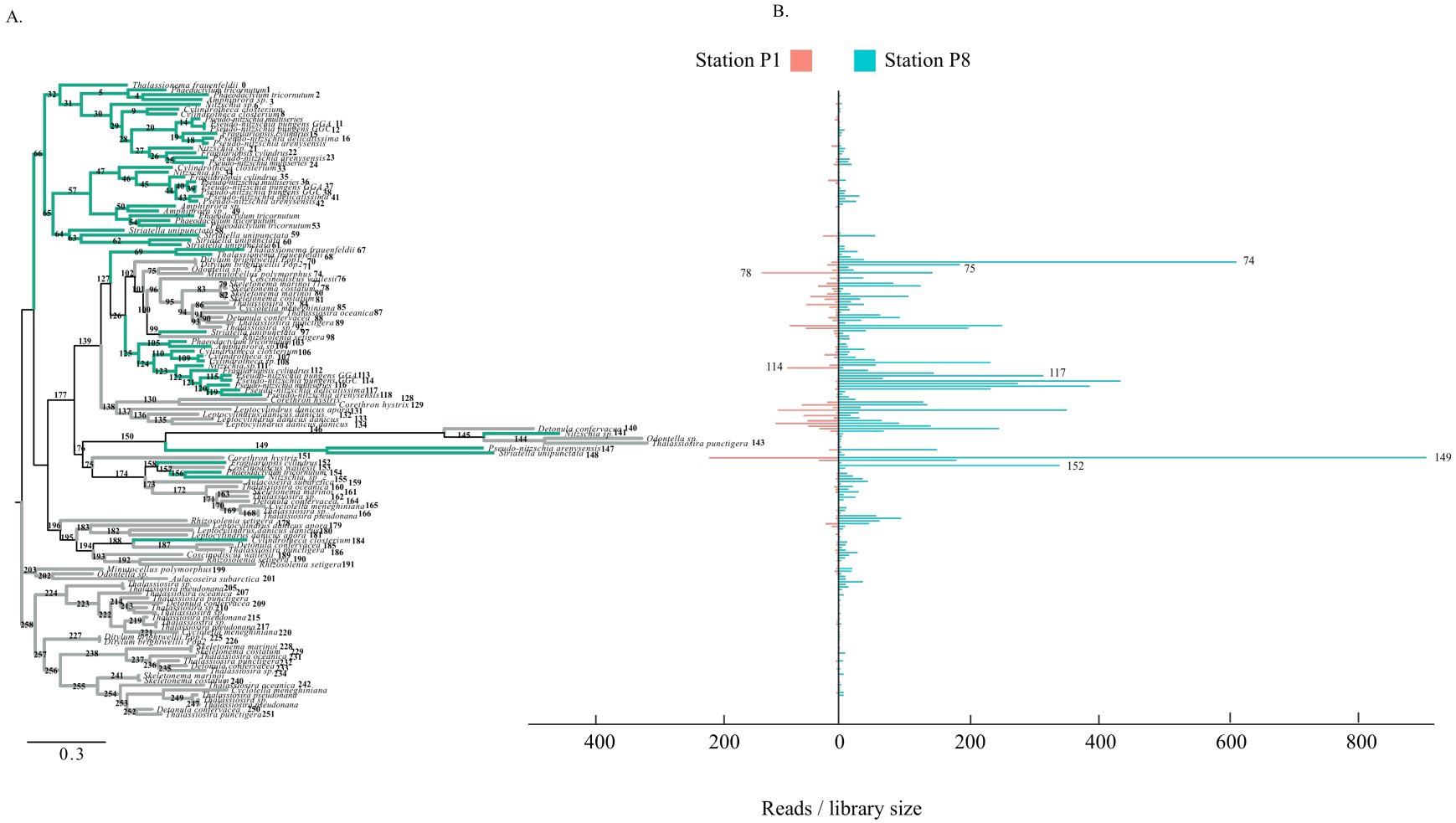

A.
B.
Station P1
Station P8
0.3
400
200
0
200
400
600
800
Reads / library size

Figure 3.7 Maximum likelihood phylogenetic tree for diatom ammonium transporters (A), and the corresponding distribution of Pplacer-recruited sequences across the tree (B). A. Pennate (teal) and centric (gray) branches are highlighted accordingly; branches that joined both pennate and centric diatoms are in black. Internal and external edges on the tree that recruited a metatranscriptome sequence are labeled with a unique number. B. Distribution of Pplacer-recruited sequences normalized to library size. Pink sequences, Station P1; blue sequences, Station P8. Select external edges that recruited a large number of sequences are labeled with the corresponding edge ID. Scale bar represents number of amino acid changes per site.

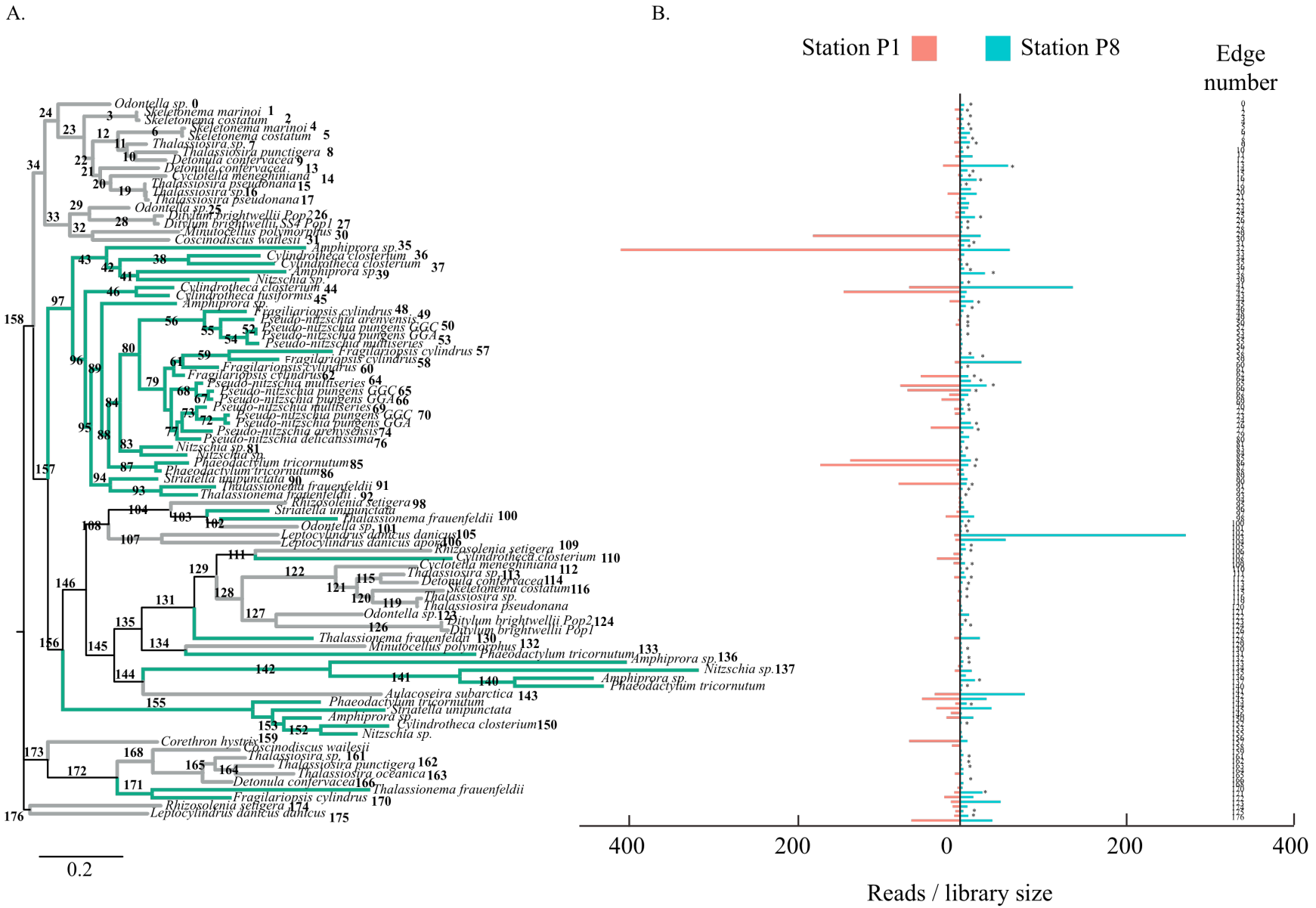

A.
B.
Station P1
Station P8
Edge number
Odontella sp. 0
Skeletonema marinoi 1
Skeletonema costatum 2
Skeletonema marinoi 4
Skeletonema costatum 5
Thalassiosira sp. 7
Thalassiosira punctigera 8
Detonula confervacea 9
Detonula confervacea. 13
Cyclotella meneghiniana 14
Thalassiosira pseudonana 15
Thalassiosira sp. 16
Thalassiosira pseudonana 17
Odontella sp. 25
Ditylum brightwellii Pop2 26
Ditylum brightwellii SS4 Pop1 27
Minutocellus polymorphus 30
Coscinodiscus wailesii 31
Amphiprora sp. 35
Cylindrotheca closterium 36
Cylindrotheca closterium 37
Amphiprora sp. 39
Nitzschia sp.
Cylindrotheca closterium 44
Cylindrotheca fusiformis 45
Amphiprora sp.
Fragiliariopsis cylindrus 48
Pseudo-nitzschia arenysensis 49
Pseudo-nitzschia pungens GGC 50
Pseudo-nitzschia pungens GGA 53
Pseudo-nitzschia multiseries
Fragilariopsis cylindrus 57
Fragilariopsis cylindrus 58
Fragilariopsis cylindrus 60
Fragilariopsis cylindrus 62
Pseudo-nitzschia multiseries 64
Pseudo-nitzschia pungens GGC 65
Pseudo-nitzschia pungens GGA 66
Pseudo-nitzschia multiseries 69
Pseudo-nitzschia pungens GGC 70
Pseudo-nitzschia pungens GGA
Pseudo-nitzschia arenysensis 74
Pseudo-nitzschia delicatissima 76
Nitzschia sp. 81
Nitzschia sp.
Phaeodactylum tricornutum 85
Phaeodactylum tricornutum 86
Striatella unipunctata 90
Thalassionema frauenfeldii 91
Thalassionema frauenfeldii 92
Rhizosolenia setigera 98
Striatella unipunctata
Thalassionema frauenfeldii 100
Odontella sp. 101
Leptocylindrus danicus danicus 105
Leptocylindrus danicus apora 106
Rhizosolenia setigera 109
Cylindrotheca closterium 110
Cyclotella meneghiniana 112
Thalassiosira sp. 113
Detonula confervacea 114
Skeletonema costatum 116
Thalassiosira sp.
Thalassiosira pseudonana
Odontella sp. 123
Ditylum brightwellii Pop2 124
Ditylum brightwellii Pop1
Thalassionema frauenfeldii 130
Minutocellus polymorphus 132
Phaeodactylum tricornutum 133
Amphiprora sp. 136
Nitzschia sp. 137
Amphiprora sp.
Phaeodactylum tricornutum
Aulacoseira subarctica 143
Phaeodactylum tricornutum
Striatella unipunctata
Amphiprora sp.
Cylindrotheca closterium 150
Nitzschia sp.
Corethron hystrix 159
Coscinodiscus wailesii
Thalassiosira sp. 161
Thalassiosira punctigera 162
Thalassiosira oceanica 163
Detonula confervacea 166
Thalassionema frauenfeldii
Fragilariopsis cylindrus 170
Rhizosolenia setigera 174
Leptocylindrus danicus danicus 175
0.2
400
200
0
200
400
Reads / library size

Figure 3.8 Maximum likelihood phylogenetic tree for diatom nitrate transporters (A), and the corresponding distribution of Pplacer-recruited sequences across the tree (B). A. Pennate (teal) and centric (gray) branches are highlighted accordingly; branches that joined both pennate and centric diatoms are in black. Internal and external edges on the tree that recruited a metatranscriptome sequence are labeled with a unique number. B. Distribution of Pplacer-recruited sequences normalized to library size. Pink sequences, Station P1; blue sequences, Station P8. The edge number associated with each bar is displayed directly to the right and corresponds with the same edge number on the tree in A. External edges that recruited sequences are indicated with an asterisk. Scale bar represents number of amino acid changes per site.

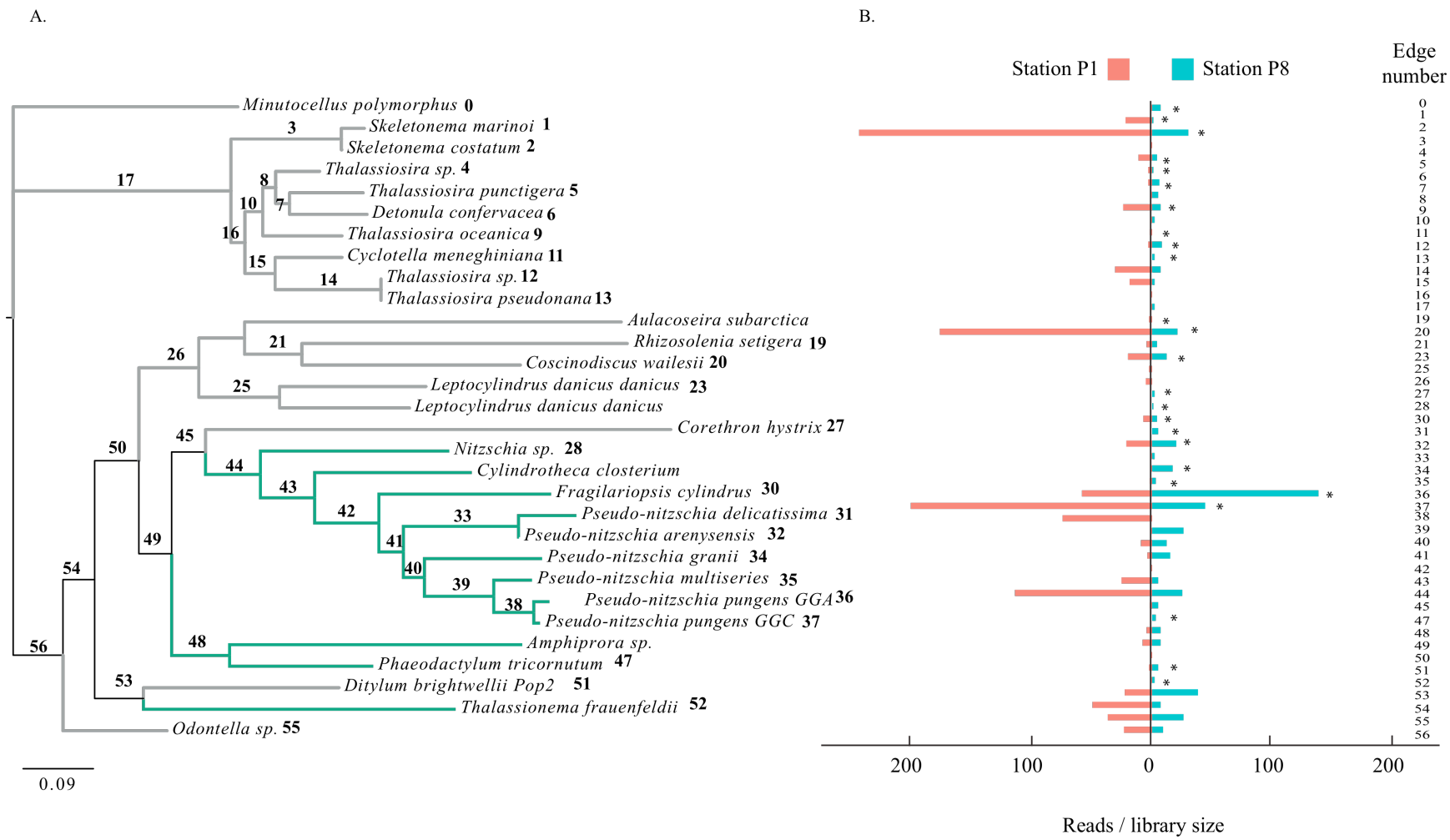

A.
Minutocellus polymorphus 0
Skeletonema marinoi 1
Skeletonema costatum 2
Thalassiosira sp. 4
Thalassiosira punctigera 5
Detonula confervacea 6
Thalassiosira oceanica 9
Cyclotella meneghiniana 11
Thalassiosira sp. 12
Thalassiosira pseudonana 13
Aulacoseira subarctica
Rhizosolenia setigera 19
Coscinodiscus wailesii 20
Leptocylindrus danicus danicus 23
Leptocylindrus danicus danicus
Corethron hystrix 27
Nitzschia sp. 28
Cylindrotheca closterium
Fragilariopsis cylindrus 30
Pseudo-nitzschia delicatissima 31
Pseudo-nitzschia arenysensis 32
Pseudo-nitzschia granii 34
Pseudo-nitzschia multiseries 35
Pseudo-nitzschia pungens GGA 36
Pseudo-nitzschia pungens GGC 37
Amphiprora sp.
Phaeodactylum tricornutum 47
Ditylum brightwellii Pop2 51
Thalassionema frauenfeldii 52
Odontella sp. 55
0.09
B.
Station P1
Station P8
Edge number
200
100
0
100
200
Reads / library size

Figure 3.9 Maximum likelihood phylogenetic tree for Tp4888 (A), and the corresponding distribution of Pplacer-recruited sequences across the tree (B). A. Pennate (teal) and centric (gray) branches are highlighted accordingly; branches that joined both pennate and centric diatoms are in black. Internal and external edges on the tree that recruited a metatranscriptome sequence are labeled with a unique number. B. Distribution of Pplacer-recruited sequences normalized to library size. Pink sequences, Station P1; blue sequences, Station P8. The edge number associated with each bar is displayed directly to the right and corresponds with the same edge number on the tree in A. External edges that recruited sequences are indicated with an asterisk. Scale bar represents number of amino acid changes per site.

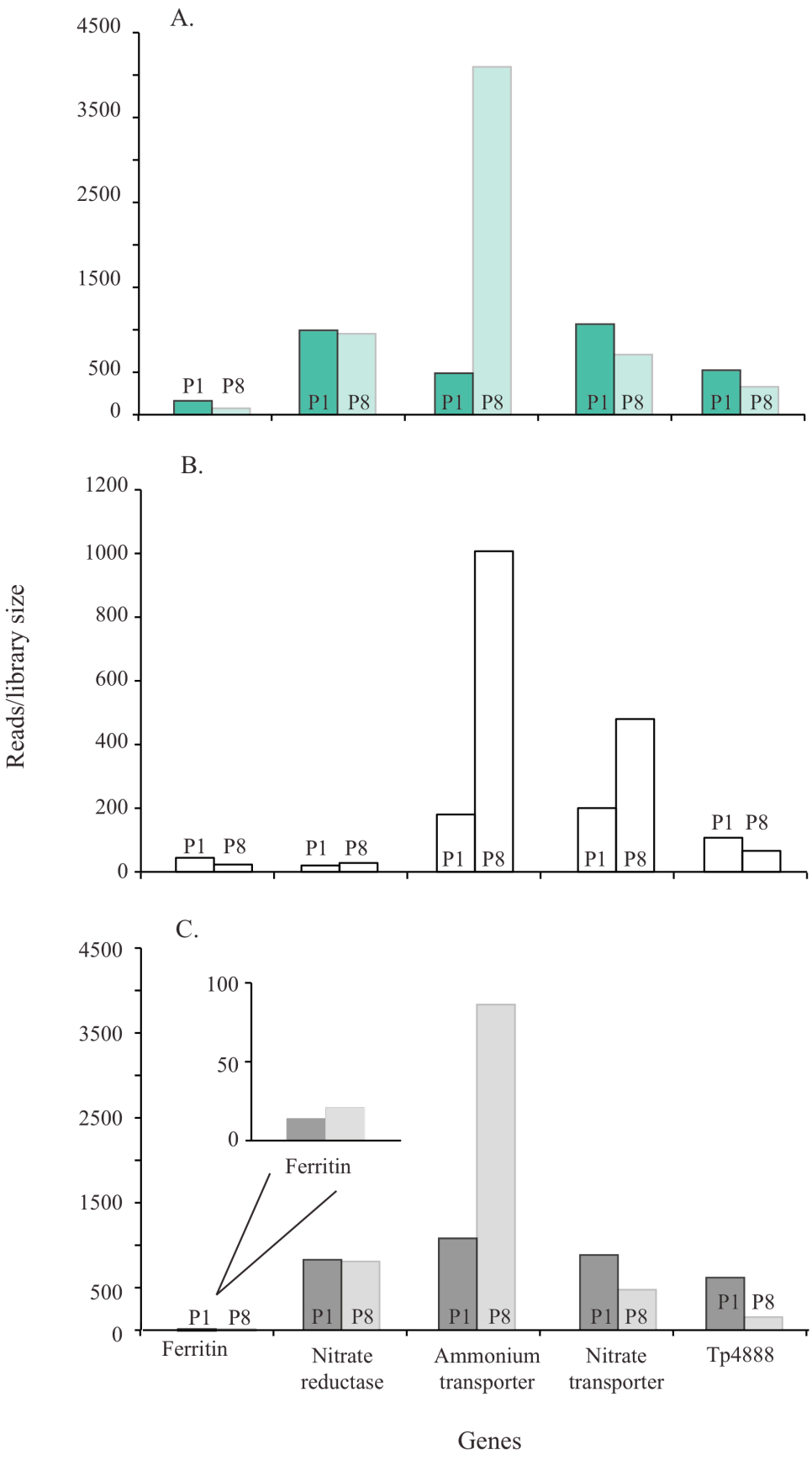

A.
4500
3500
2500
1500
500
0
P1 P8
B.
1200
1000
800
600
400
200
0
C.
100
50
Ferritin
Reads/library size
Ferritin
Nitrate reductase
Ammonium transporter
Nitrate transporter
Tp4888
Genes

Figure 3.10 Total abundance of normalized sequences (y-axis) that recruited to pennate, mixed, and centric edges of each phylogenetic tree for the following genes: Ferritin, Nitrate reductase, Ammonium transporters, Nitrate transporters, and Tp4888 (x-axis). A. Sequences that recruited to the pennate-related edges at Station P1 and P8. B. Sequences that recruited to mixed edges at Station P1 and P8. C. Sequences that recruited to the centric-related edges at Station P1 and P8.

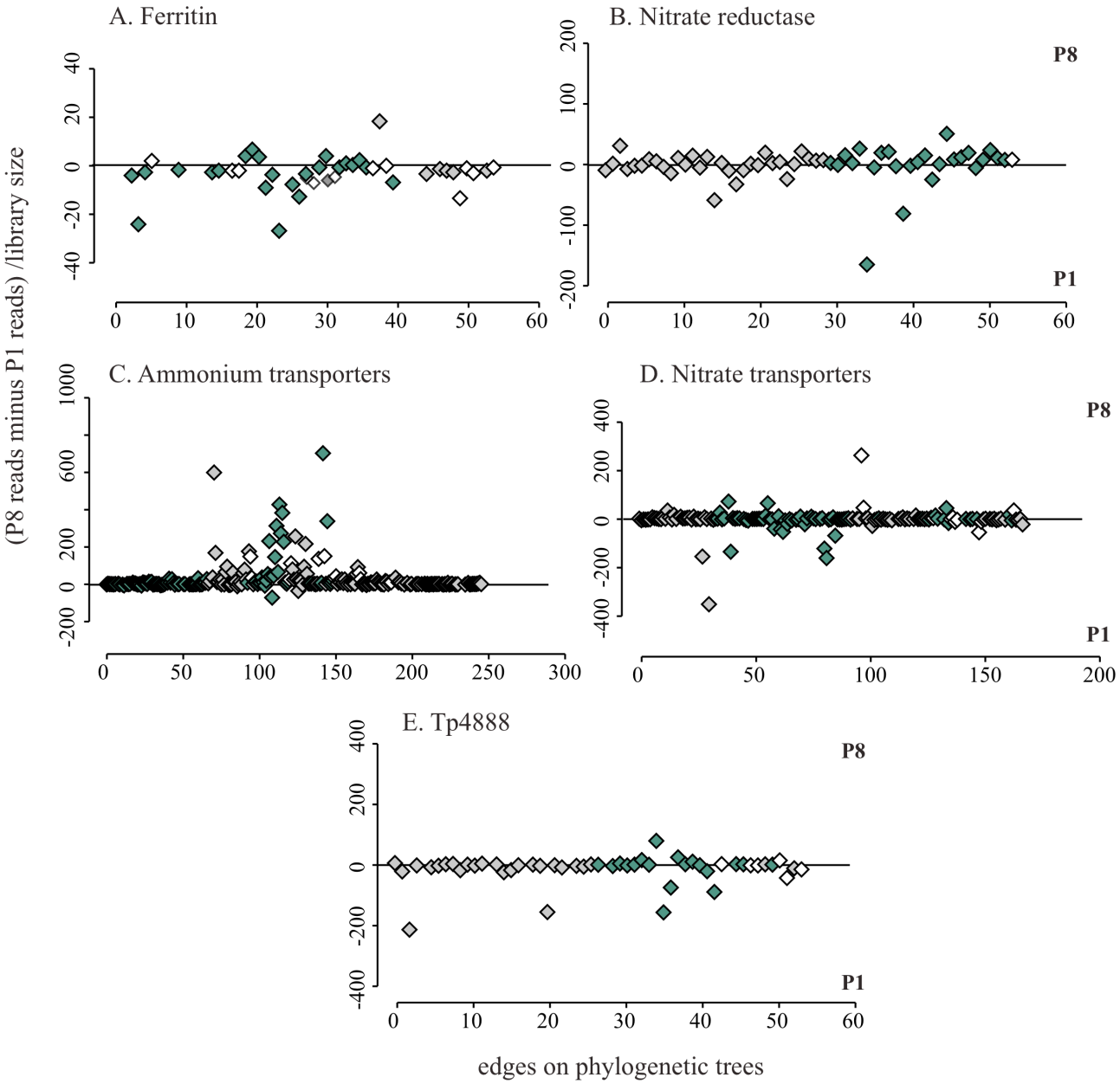

A. Ferritin
B. Nitrate reductase
C. Ammonium transporters
D. Nitrate transporters
E. Tp4888
P8
P1
(P8 reads minus P1 reads) /library size
edges on phylogenetic trees

Figure 3.11 Scatterplots displaying sequence abundance at each edge on the phylogenetic tree between Station P1 and P8 for the five proteins examined in this study. A. Ferritin, B. Nitrate reductase, C. Ammonium transporters, D. Nitrate transporters, and E. Tp4888. Each diamond represents an edge on the respective protein phylogenetic tree: teal diamonds, pennate diatoms; gray diatoms, centric diatoms; white diamonds, both pennate and centric diatoms. The difference in sequence recruitment between P8 and P1 (normalized to library size) is given on the y-axis: positive values, more sequences were present at Station P8; negative values, more sequences were present at Station P1. The y-axis scale for the ammonium transporters (C) differs between P1 and P8 because of the disproportionate amount of P8 sequences compared to P1. This difference was not observed for the other four proteins. The location of each edge on the phylogenetic tree is given on the x-axis. Edge numbers are specific to each tree, and reflect the unique phylogeny of each protein.

Supplemental Table 3.1 List of queried diatom EST databases from the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) used for the HMM search and their corresponding diatom group classifications.

	Organism Name	MMETSP ID	Diatom Group
1	*Striatella unipunctata*	MMETSP0800	araphid Pennate
2	*Thalassionema frauenfeldii*	MMETSP0786	araphid Pennate
3	*Amphiprora sp.*	MMETSP1065	raphid Pennate
4	*Cylindrotheca closterium*	MMETSP0017	raphid Pennate
5	*Nitzschia sp.*	MMETSP0014	raphid Pennate
6−7	*Pseudo-nitzschia pungens cf. cingulata / cf. pungens*	MMETSP1060 / MMETSP1061	raphid Pennate
8	*Pseudo-nitzschia delicatissima*	MMETSP 0327	raphid Pennate
9	*Pseudo-nitzschia arenysensis*	MMETSP 0329	raphid Pennate
10	*Aulacoseira subarctica*	MMETSP1064	Centric
11	*Corethron hystrix*	MMETSP0010	Centric
12	*Coscinodiscus wailesii*	MMETSP1066	Centric
13	*Detonula confervacea*	MMETSP1058	Centric
14	*Rhizosolenia setigera*	MMETSP0789	Centric
15	*Cyclotella meneghiniana*	MMETSP1057	Centric
16−17	*Leptocylindrus danicus var. danicus / var. apora*	MMETSP0321 / MMETSP 0322	Centric
18−19	*Ditylum brightwellii* SS4 Pop 1 / S10 Pop 2	MMETSP1062 / MMETSP1063	bi/multipolar Centric
20	*Minutocellus polymorphus*	MMETSP1070	bi/multipolar Centric
21	*Odontella sp.*	MMETSP0015	bi/multipolar Centric
22	*Skeletonema costatum*	MMETSP0013	bi/multipolar Centric
23 - 24	*Skeletonema marinoi Den-03 / Hels-07*	MMETSP 0320 / MMETSP 0319	bi/multipolar Centric
25 - 26	*Thalassiosira sp. / sp .(freshwater)*	MMETSP1071 / MMETSP1059	bi/multipolar Centric
27	*Thalassiosira punctigera* C2	MMETSP1067	bi/multipolar Centric

Supplemental Table 3.2 Identification and annotation of diatom read subset. The percentages of diatom reads that have an associated annotation are shown in parentheses.

	Replicate	Reads that hit to the in-house database	Diatom reads that passed the NR filter	Diatom reads with a KEGG annotation	Diatom reads with a KO annotation	Diatom reads with a module annotation
Station P1	A	2.63×10^5	1.52×10^5	1.23×10^5 (81)	8.22×10^4 (54)	4.30×10^4 (28)
	B	2.48×10^5	1.48×10^5	1.18×10^5 (80)	8.10×10^4 (55)	4.11×10^4 (28)
Station P8	A	3.34×10^5	2.01×10^5	1.52×10^5 (76)	9.47×10^4 (47)	4.37×10^4 (22)
	B	2.89×10^5	1.65×10^5	1.21×10^5 (73)	7.83×10^4 (48)	3.61×10^4 (22)

Conclusion

Diatom blooms occur seasonally in coastal waters, where the interplay between diatom growth and nitrogen availability determines when diatoms will bloom, which in turn, impacts how much nitrogen is available in the water column for other phytoplankton groups. This nutrient-driven succession makes diatoms an integral component of the nitrogen and carbon cycles. Differences in cell morphology, physiology, and nutrient uptake rates indicate that diatoms respond to nitrogen inputs in different ways, yet the underlying mechanisms for the diversity of responses are unknown. This thesis examined the mosaic of transcriptional responses associated with multiple diatoms, including a natural diatom assemblage, in response to changes in nitrogen source and availability.

Chapter one was motivated by the question: What is the role of the urea cycle in cellular metabolism in the model diatom *Thalassiosira pseudonana*? This first study to examine the role of the urea cycle in *T. pseudonana* expands our understanding of the fundamental components of diatom nitrogen metabolism. Findings highlight the strong influence of a diel cycle, light intensity and nitrogen source on the flow of nitrogen in a diatom cell and connect the urea cycle, newly discovered in diatoms, to additional cellular processes including glutamine and urea production. The urea cycle may provide diatoms with a novel metabolic strategy for balancing cellular energy demands with nitrogen assimilation and carbon fixation processes. This work has important implications for our understanding of nitrogen flow in the cell over diel cycles at surface ocean irradiances.

Chapter two expanded on the hypotheses of cellular nitrogen flow and compared the transcriptional responses of three diverse diatoms harvested at the onset of nitrate starvation. This work was motivated by the question: Do all diatoms exhibit the same transcriptional

responses when nitrogen availability in the cell changes? Prior to this work, no studies had quantitatively compared the transcriptome-wide response of multiple diatoms grown under the same nitrate-starved growth conditions. The specific transcriptional response to nitrate starvation varied among diatoms, yet greater transcriptional and functional similarities between the two pennate diatoms compared to the bipolar centric highlighted fundamental differences in how each group has evolved to respond to its environment. Each diatom possesses a different transcriptional strategy for achieving a similar metabolic endpoint that is likely the result of a complex evolutionary history. In the field, we would expect these differing strategies to reveal themselves in the chemical imprint that each diatom leaves on its environment, which would impact on the flow of carbon and nitrogen in the system.

Finally, chapter three examined the diversity of diatom transcriptional responses in metatranscriptomes collected from the Northeast Pacific Ocean, and addressed the question: Do all diatoms in a diverse diatom assemblage exhibit similar transcriptional responses to changes in nutrient availability? Overall, the observed transcriptional response was a reflection of several highly abundant diatoms, and a uniform, community-wide transcriptional response was not observed. When transcripts were sorted based on their gene phylogeny, group- and species-specific transcript abundance patterns emerged. Findings from chapter 3 highlight the different transcriptional responses among closely related diatoms to changes in both nitrogen and iron availability. Although all diatoms in an assemblage are present in the same bulk chemical setting, each member experiences a different microenvironment because of their evolved physiologies.

This thesis focused on the transcriptional response of diatoms to nitrogen availability- a known control on bloom formation. However, nitrogen represents one of many environmental factors that control diatom growth; other controls include light, physical mixing, and temperature

(bottom-up controls), as well as grazing (top-down control). Just as diatoms have evolved mechanisms for capitalizing on transient nitrogen supply, they also have evolved mechanisms for responding to these other factors (Smetacek 2001). Future studies on model species can build upon the comparative work from this thesis by adding in these additional environmental factors. This work will help to interpret field metatranscriptomes, which innately reflect this environmental complexity.

In the paradox of the plankton, G.E. Hutchinson (Hutchinson 1961) explored the topic of multiple phytoplankton species coexisting in an environment with limited resources. Among the diatoms, it is an absolute wonder that so many species coexist in modern-day oceans given the transient nature and low concentrations of nutrients required for their growth. We now know that diatoms have evolved unique metabolic mechanisms for nutrient recycling and repartitioning (chapter 1). These mechanisms are differentially regulated at the group- and species-specific level (chapter 2), and they may allow for each diatom species to thrive in a specialized niche (chapter 3). Rather than all diatoms striving for numerical dominance, individualized niches allow each species to flourish in its own microenvironment, which is characterized by transient nutrient supply (e.g. regenerated nutrients that are excreted and taken up again on the order of minutes), interactions with other organisms (e.g. bacteria, viruses, or other diatoms), and often, direct contact with grazers. Because of their rich evolutionary history, each diatom species responds to its microenvironment in a different manner than neighboring species. When conditions become ideal for bloom formation, some diatoms may increase in abundance; it is also possible that certain species are adapted to maintain lower cell abundance as their strategy for co-existence.

Diatoms are the most diverse group of marine phytoplankton, and yet diatom blooms typically consist of only one–two dominant species among hundreds of thousands. Furthermore, diatom carbon biomass is estimated to vary across eight orders of magnitude depending on the species (Leblanc et al. 2012). Both of these observations provide a strong argument for considering individual diatom species when assessing the diatom community-response to environmental controls. Modeling the key components of diatom metabolism, along with understanding the diversity of diatom transcriptional responses in the field, will allow for better predictions of which diatom species will bloom, and they have implications for the biogeochemical contributions of the diatom group.

Conclusion References

Hutchinson GE. (1961). The Paradox of the Plankton. The American Naturalist 95:137–145.

Leblanc K, Arístegui J, Armand L, Assmy P, Beker B, Bode A, et al. (2012). A global diatom database - abundance, biovolume and biomass in the world ocean. Earth Syst Sci Data 4:149–165.

Smetacek V. (2001). A watery arms race. Nature 411:745.

Curriculum Vitae

SARA JANE BENDER
Center for Environmental Genomics
School of Oceanography, University of Washington
Box 357940, Seattle, WA 98195
Phone: (206)-221-7146

EDUCATION

2013

Ph.D., Biological Oceanography
School of Oceanography, University of Washington
Dissertation Title: Quantifying the response of diatom nitrogen metabolism to environmental changes
Advisor: E. Virginia Armbrust
Defense: August 2, 2013

2009

M.Sc., Biological Oceanography
School of Oceanography, University of Washington
Presentation Title: The coupled effects of light and nitrogen source on the urea cycle in the marine diatom, *Thalassiosira pseudonana*
Advisor: E. Virginia Armbrust

2005

B.A., Biological Sciences- *summa cum laude*
Rutgers College, Rutgers University
Minors in Marine Science and Anthropology

2003

Study abroad program (July – November)
University of Queensland, Brisbane, Australia

PUBLICATIONS

(1) Bender, S.J., Berthiaume, C.T., Parker, M.S., Durham, B.P., Groussman, R., and E.V. Armbrust. 2013. Species-specific transcriptional responses of diatom communities to changes in iron and nitrogen availability along Line P in the Northeast Pacific Ocean. *Environmental Microbiology*. In preparation.

(2) Bender, S.J., Durkin, C.A., Berthiaume, C.T., Morales, R.L.M., and E.V. Armbrust. 2013. Transcriptional responses of three evolutionarily diverse diatoms to nitrate starvation. *ISME Journal*. In preparation.

(3) Durham, B.P., Grote, J., Whittaker, K.A., Bender, S.J., Luo, H. et al. 2013. Draft genome sequence of marine alphaproteobacterial strain HIMB11, the first cultivated representative of a unique lineage within the Roseobacter clade possessing a remarkably small genome. *Standards in Genomic Sciences*. In preparation.

(4) Durkin, C.A., Bender, S.J., Chan, K.Y.K., Gaessner, K., Grünbaum, D. and E.V. Armbrust. 2013. Silicic acid supplied to coastal diatom communities influences cellular silicification and the potential export of carbon. *Limnology and Oceanography*. 58(5): 1707-26.

(5) Durkin, C.A., Marchetti, A., Bender, S.J., Truong, T., Morales, R., Mock, T. and E.V. Armbrust. 2012. Diverse patterns of silica precipitation and frustule-related gene transcription among iron-limited diatoms. *Limnology and Oceanography* 57(6): 1619-33.

(6) Bender, S.J., Parker, M.S. and E. V. Armbrust. 2012. The coupled effects of light and nitrogen source on the urea cycle and nitrogen metabolism over a diel cycle in the marine diatom *Thalassiosira pseudonana*. *Protist* 163: 232-51.

(7) Bidle, K.D. and S.J. Bender. 2008. Iron starvation and culture age activate metacaspases and programmed cell death in the marine diatom, *Thalassiosira pseudonana*. *Eukaryotic Cell* 7(2): 223-236.

PROFESSIONAL EXPERIENCE

present

North Atlantic Region Postdoctoral Scholar
Woods Hole Oceanographic Research Institute, Woods Hole, MA
Supervisor: Dr. Mak Saito

2005 – 2006

Research Technician
Institute of Marine and Coastal Sciences, Rutgers University
Supervisor: Dr. Kay D. Bidle

2003

Volunteer Field Technician (May – July)
Rutgers University Marine Field Station
Supervisor: Dr. Kenneth Able

2002

Volunteer Research Technician
Department of Life Sciences, Rutgers University
Supervisor: Dr. Joanna Burger

TEACHING

2010 Graduate Teaching Assistant, University of Washington
Introduction to Oceanography Lab (OCN201)

2009 Lead Graduate Teaching Assistant, University of Washington
Introduction to Oceanography (OCN200)

2008 Graduate Teaching Assistant, University of Washington
Introduction to Oceanography (OCN200)

AWARDS/HONORS

2012 Outstanding Student Presentation Award
2012 Ocean Sciences Meeting, Salt Lake City, UT

2011 Selected Participant, C-MORE Summer Course on Marine Microbiology
University of Hawaii

2010 Dean A. McManus Excellence in Teaching Award
School of Oceanography, University of Washington

2010 Selected Cover Story, Northwest Science & Technology magazine
"Tsunami Science In The Northwest: Researchers Are Working To Understand, Predict, And Mitigate One Of Nature's Deadliest Hazards"

2005 Henry Rutgers Scholar, Rutgers University

2005 Outstanding Graduating Senior in Marine Sciences Award

2005 Betty Falk Yatvin Memorial Award

2004 NSF REU- Undergraduate Research Fellow
Hawaii Institute of Marine Biology, University of Hawaii

2002 Selected Intern, U.S. Fish and Wildlife Service
Edwin B. Forsythe National Wildlife Refuge, Brigantine, New Jersey

CONFERENCES/WORKSHOPS

2013 American Association for the Advancement of Sciences Annual Conference (Boston, MA) 14 – 18 February

- Bish, P., Schindler, A.G., Bender, S.J., Terai, C.R., Snelson, C.D. and Agatsuma, R. The Seattle Forum on Science Ethics & Policy: A student-run organization focusing on the intersection of science and policy. (poster)

2013 ASLO Aquatic Sciences Meeting (New Orleans, LA) 17 – 22 February

- Bender, S.J., Durkin, C.A., Durham, B.P., Berthiaume, C., and Armbrust, E.V. Nitrogen transporters in laboratory transcriptomes and field metatranscriptomes reveal species-specific metabolic responses of marine diatoms to nitrogen availability. (oral)
- Martin, P., Mooy, B.V., Bender, S.J., and Armbrust, E.V. Ocean sections of polyphosphate and membrane lipids show distinct microbial responses to phosphorus stress and resupply. (oral)

2012 Gordon Marine Microbes Conference (Il Lucca, Italy) 23 – 29 June

- Bender, S.J., Schruth, D., Durkin, C.A., Morales, R.L., Berthiaume, C. and Armbrust, E.V. Not all diatoms are created equal: A comparison of twenty-seven transcriptomes across three model diatoms. (poster)

2012 Ocean Sciences Meeting (Salt Lake City, Utah) 19 – 24 February

- Bender, S.J., Durkin, C.A., Schruth, D. Morales, R.L. and Armbrust, E.V. Identifying shared responses to nitrate starvation among three diatoms using whole-cell transcriptomics (oral)
- 2011 C-MORE Summer Course Genome Collective. Comparative genomics of Rhodobacteraceae Sp. HIMB11 reveals metabolic specialization in a coastal, marine isolate. (poster)
- 2011 C-MORE Summer Course Cruise Collective. Scales of variability at Station Aloha. (poster)
- Durkin, C.A., Bender, S.J., Gaessner, K. and Armbrust, E.V. Uptake limitation of silicic acid in coastal diatoms controls cellular silicification. (poster)

2012 American Association for the Advancement of Sciences Annual Conference (Vancouver, BC, Canada) 16 – 20 February

- Hillenmeyer, E., Bender, S.J., McMillan, A. and MacLeod, B. The Seattle Forum on Science Ethics & Policy: A student-run organization focusing on the intersection of science and policy. (poster)

2011 Eastern Pacific Ocean Conference (South Lake Tahoe, CA) 11 – 14 October

- Durkin, C.A., Bender, S.J. and Armbrust, E.V. Excessive silicic acid supply leads to increased silicification in coastal diatoms. (oral)

2011 American Association for the Advancement of Sciences Annual Conference (Washington, D.C.) 17 – 21 February

- Bender, S.J., Hillenmeyer, E. and A. McMillan. The Forum on Science, Ethics and Policy: Adapting a student-run organization to address the needs of a university and the community. (poster)

2010 CMOP Research Workshop (Beaverton, OR) 7 – 8 December (participant)

2010 Ocean Sciences Meeting (Portland, OR) 22 – 26 February

- Bender, S.J., Parker, M.S. and E.V. Armbrust. The coupled effects of light and nitrogen source on the urea cycle in the marine diatom, *Thalassiosira pseudonana*. (poster)
- Durkin, C.A., Bender, S.J. Truong, T., Marchetti, A. and E.V. Armbrust. Common cell wall related genes in diatoms and their potential use as biological indicators in changing ocean environments. (oral)

2009 ASLO Aquatic Sciences Meeting (Nice, France) 25 – 30 January

- Bender, S.J. and E.V. Armbrust. Effects of nitrogen source and the light: dark cycle on growth and gene expression associated with the urea cycle in *Thalassiosira pseudonana*. (oral)

2008 Ocean Sciences Meeting (Orlando, FL) 2 – 7 March

- Bender, S.J. and E.V. Armbrust. Understanding the connection between differing nitrogen sources and the urea cycle in the diatom, *Thalassiosira pseudonana*. (poster)

2008 COSEE-OLC and Washington Sea Grant Communicating Ocean and Marine Sciences workshop (Seattle, WA) 22 November

- Bender, S.J. and E.V. Armbrust. Understanding the connection between nitrogen source and cellular metabolism in the diatom, *Thalassiosira pseudonana*. (poster)

2006 Biocomplexity-Evolution of Marine Phytoplankton (Rutgers University) 11 – 13 January

- Bender, S.J. and K.D. Bidle. A case for autocatalytic cell death as a result of nutrient limitation in the diatom *Thalassiosira pseudonana*. (poster)

2005 ASLO Summer Meeting 2005 (Santiago de Compostela, Spain) 19 – 24 June

- Bender, S.J. and K.D. Bidle. Mortality in the diatom, *Thalassiosira pseudonana*, in response to differing nutrient stresses. (poster)

PUBLIC PRESENTATIONS

*invited

2013 Pacific Science Center, Seattle Science Festival
(Seattle, WA) 10 June

- *Bender, S.J. Exploring the Appetites of Marine Microbes.

Town Hall Seattle, UW Science Now program
(Seattle, WA) 8 April

- Bender, S.J. Exploring the Appetites of Marine Microbes.

RESEARCH CRUISES

120+ days at sea

2012 *Graduate Student*, R/V Thomas G. Thompson, Line P
16 – 23 May, *GeoMICS* program

2011 *Graduate Student*, R/V Kilo Moana, Station Aloha
12 – 22 June, *C-MORE Summer Course on Microbial Oceanography*

2010 *Graduate Student*, R/V Wecoma, Newport, OR
22 – 27 May / 27 July – 2 August
Coastal Margin, Observation and Prediction (CMOP) program

2009 *Graduate Student*, R/V New Horizon, Newport, OR
12 – 26 May / 29 August – 11 September
Coastal Margin, Observation and Prediction (CMOP) program

2008 *Graduate Student*, Sorcerer II, Puget Sound, WA
Day Cruise, August

2007 *Graduate Student*, R/V Western Flyer, Moss Landing, CA
1 – 11 October, *Investigating changes in the microbial community along a nitrogen gradient (MBARI)*

2006 *Research Volunteer*, R/V Thomas G. Thompson, Tahiti
13 February – 9 March, *U.S. Repeat Hydrography P-16N Transect*

2006 *REU Summer Program Mentor*, USS Sturgeon Bay, Staten Island, New York Harbor, Day Cruise, July

2005 *Research Technician*, R/V Oceanus, Bermuda
7 – 26 August, *Eddies program in the Sargasso Sea*

2007 – 2010
Graduate Student, R/V Thomas G. Thompson, Puget Sound, WA
November 2007, May 2008, August 2008, November 2009, April 2010

LEADERSHIP

2013
Selected Representative, Core Programs Student Advisory Board
UW Graduate School

2012 – 2013
Campus Liaison, Emerging Leaders in Science and Society
American Association for the Advancement of Science (AAAS), National Program

2008 – 2013
Leader, Forum on Sciences Ethics & Policy
University of Washington

2012
Graduate Student Representative, Faculty Search Committee
School of Oceanography, University of Washington

2009 – 2011
Program Co-chair, Girls in Engineering, Math and Science
Seattle chapter

2009 – 2011
Board Member, Association for Women in Science
Seattle chapter

2006 – 2011
President, Rutgers Alumni Club of Seattle
Seattle chapter

MENTORING

UNDERGRADUATE

2010 – 2013
Supervisor, Moira Regan, Oceanography major (Graduated 2013)

Nov. 22, 2011
Guest Speaker, UW Explore Oceanography course (Ocean100)

2008 – 2010
Supervisor, Franziska Lutz, Oceanography major (Graduated 2010)
- University of Washington Undergraduate Research Symposium participant, 2010 and 2009
  The effects of nitrogen limitation on growth and gene expression in the marine diatoms Fragilariopsis cylindrus and Pseudo-nitzschia multiseries

Jan. 5 – 6, 2010
Participant, COSEE-OLC and Washington Sea Grant event 'Addressing Broader Impacts Requirements for Research Proposals'

Sept. 24, 2009
Discussion Leader, UW Oceanography TA Training

Sept. 18, 2008
Discussion Leader, UW Oceanography TA Training

2005, 2006
Summer Program Mentor, REU Undergraduate Summer Research Program
Rutgers University

YOUTH

2009 – 2010, 2012, 2013
Middle School Science Essay Judge, "Biomedical Breakthroughs and My Life" Essay Contest, Northwest Association for Biomedical Research (NWABR)

April 2, 2013/May 6, 2012
Guest Instructor, Girls in Engineering, Math and Science

March 3, 2012
Science Judge, National Ocean Sciences Bowl – Orca Bowl

2006 – 2011
Student Mentor, Girls in Engineering, Math and Science
Association for Women in Science, Seattle, WA

March 29, 2007
Volunteer, Science Night, North City Elementary School