

2018

## THALASSIOSIROID DIATOM RESPONSES TO SILICON STRESS AND OCEAN ACIDIFICATION

Joselynn Wallace  
University of Rhode Island, joselynnw@gmail.com

Follow this and additional works at: [https://digitalcommons.uri.edu/oa\\_diss](https://digitalcommons.uri.edu/oa_diss)

Terms of Use

All rights reserved under copyright.

---

### Recommended Citation

Wallace, Joselynn, "THALASSIOSIROID DIATOM RESPONSES TO SILICON STRESS AND OCEAN ACIDIFICATION" (2018). *Open Access Dissertations*. Paper 762.  
[https://digitalcommons.uri.edu/oa\\_diss/762](https://digitalcommons.uri.edu/oa_diss/762)

This Dissertation is brought to you by the University of Rhode Island. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact [digitalcommons-group@uri.edu](mailto:digitalcommons-group@uri.edu). For permission to reuse copyrighted content, contact the author directly.

*THALASSIOSIROID* DIATOM RESPONSES TO SILICON  
STRESS AND OCEAN ACIDIFICATION

BY

JOSELYNN WALLACE

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY  
IN  
BIOLOGICAL AND ENVIRONMENTAL SCIENCES

UNIVERSITY OF RHODE ISLAND

2018

DOCTOR OF PHILOSOPHY DISSERTATION

OF

JOSELYNN WALLACE

APPROVED:

Dissertation Committee:

Major Professor      Bethany D. Jenkins

Christopher E. Lane

Tatiana A. Rynearson

Nasser H. Zawia  
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND  
2018

## ABSTRACT

Atmospheric CO<sub>2</sub> has risen dramatically since the industrial revolution. This rise in atmospheric and oceanic pCO<sub>2</sub> has perturbed ocean carbonate chemistry and led to ocean acidification. Diatoms are phytoplankton that account for 40% of oceanic primary production through photosynthetic carbon fixation, which is aided by their carbon concentrating mechanism (CCM). The CCM uses the bicarbonate transporters (BCTs) and carbonic anhydrases (CAs). Our current understanding of how diatoms might respond to ocean acidification is based on experiments using model diatoms or assessing the response of the bulk diatom community, rather than assessing a diversity of diatoms in a complex environment. This dissertation aims to expand our knowledge regarding diatom response to CO<sub>2</sub> in ecologically important, non-model diatoms and their response in laboratory experiments and field mesocosms to alterations in CO<sub>2</sub> concentration.

Diatoms' primary production is a function of their growth, which is constrained by the availability of nutrients in the surface ocean. Silicon is a nutrient that is particularly important for diatoms, as they are unique in their requirement for silicon to build their cell walls. Silicon limitation has been observed in low iron high nutrient low chlorophyll (HNLC) regions and the North Atlantic Ocean, although these studies have focused on the whole diatom community rather than specific diatom groups that may not uniformly experience silicon limitation. Genetic markers have been used to probe species-specific iron status in the field, and similar molecular markers of silicon status could be powerful tools to probe the silicon status of different co-existing diatom species. However, current studies of silicon limitation have relied

on model diatoms rather than species that are likely to be found in HNLC regions or the North Atlantic Ocean, limiting the ability to develop appropriate molecular markers. This dissertation aimed to fill in these knowledge gaps using transcriptomic studies of *Thalassiosiroid* diatom isolate cultures as well as incubations of mixed diatom assemblages.

Chapter 1 of this thesis examined the transcriptomes of two closely related *Thalassiosira* diatoms to better understand the silicon limitation gene expression response of diatoms found in ecosystems where their growth may be constrained by silicon availability, toward the goal of developing appropriate molecular markers of silicon status. This study found a gene family encoding putative ATP-grasp domain proteins that were upregulated in silicon-limited *T. oceanica* and *T. weissflogii*. The upregulation of these genes is unique to silicon limitation and its expression is not induced in nitrate or iron limitation conditions. The members of this gene family were also upregulated in response to silicon limitation in previously published *T. pseudonana* transcriptome studies and homologs were found across a diversity of diatoms, suggesting that it might be useful as a silicon limitation marker in many diatoms, in addition to *Thalassiosira* spp. diatoms.

Chapter 2 of this thesis adds to our knowledge of the diversity of CCM and high CO<sub>2</sub> response in different species of the *Thalassiosira* genus of diatoms. CO<sub>2</sub> manipulation experiments were conducted with four *Thalassiosira* species – *T. pseudonana*, *T. rotula*, *T. weissflogii*, and *T. oceanica*, and transcriptomes were sequenced from the latter 3 species. These species displayed a range of growth rate changes across CO<sub>2</sub> conditions, from no change across conditions (*T. pseudonana*),

slower growth in lower CO<sub>2</sub> (*T. weissflogii*), slower growth in higher CO<sub>2</sub> (*T. oceanica*), and faster growth in higher CO<sub>2</sub> (*T. rotula*). The accompanying transcriptome evidence for *T. rotula*, *T. weissflogii*, and *T. oceanica* indicate that these differences in CO<sub>2</sub> responses boil down to the ability to dynamically regulate the carbon concentrating mechanism while maintaining internal redox and ion homeostasis, as well as trade-offs between funneling excess CO<sub>2</sub> into internal storage pools versus a higher growth rate.

Chapter 3 uses a metatranscriptome approach to generalize about how the CCM of *Thalassiosira* spp. differs from that of *Skeletonema* spp. diatoms. This study involved CO<sub>2</sub> manipulation experiments of a plankton assemblage collected from Massachusetts Bay during March of 2014. We show evidence of distinct enzymatic strategies for N and P uptake and scavenging in these two diatom genera. There is also evidence of distinct strategies in their CCMs – while *Thalassiosira* spp. diatoms seem to use the  $\alpha$ ,  $\delta$ , and  $\zeta$  carbonic anhydrases and the substrate-specific SLC4 and non-specific SLC26 bicarbonate transporters, *Skeletonema* spp. diatoms rely on the  $\gamma$  and  $\zeta$  carbonic anhydrases and the SLC26 bicarbonate transporters or diffuse CO<sub>2</sub> uptake. This suggests that the co-existence of *Skeletonema* spp. and *Thalassiosira* spp. might be partially due to a CCM that uses distinct forms of inorganic carbon (CO<sub>2</sub> or HCO<sub>3</sub><sup>-</sup>) and CA isozymes that use different metal cofactors for their enzymatic activity ( $\gamma$  can substitute iron,  $\delta$  can substitute cobalt). Furthermore, if *Skeletonema* does rely on diffusive CO<sub>2</sub> uptake rather than active transport of HCO<sub>3</sub><sup>-</sup> using SLC4 or SLC26 transporters, *Thalassiosira* may have a competitive advantage in high CO<sub>2</sub> as it downregulates these active transporters.

## ACKNOWLEDGMENTS

The last 7 years of my life has certainly been challenging. I would have not made it through without the constant support of friends, family, colleagues and mentors. I would like to take this opportunity to thank my advisor, Bethany Jenkins. You have been incredibly supportive of me and have provided me with many opportunities to develop as a scientist. Thank you to my committee, Tatiana Rynearson, Chris Lane, and Marta Gómez-Chiarri, for their patience, guidance, and precious time. I would like to acknowledge Andrew King and Dreux Chappell, you have both been mentors and role models for the kind of scientist and person I would like to be. I'd like to acknowledge all the MetaLab group for helping me refine all of my talks and posters over the years. Special thanks to all of the other Jenkins lab mates who have shared snacks with me over the years – LeAnn, Shelley, AJ, Laura, Kris, Alexa, and Eric. Special thanks to all of my very large family, particularly my nieces Ronnie Clo and Sophie Pie. I would also like to acknowledge my partner, Andrew. I love you and thank you for always reminding me to drink water and for washing my travel mugs.

## **PREFACE**

This dissertation was formatted in accordance with the manuscript format guidelines established by the Graduate School of the University of Rhode Island

## TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>ii</b>
<b>ACKNOWLEDGMENTS</b> .....	<b>v</b>
<b>PREFACE</b> .....	<b>vi</b>
<b>TABLE OF CONTENTS</b> .....	<b>vii</b>
<b>LIST OF TABLES</b> .....	<b>xii</b>
<b>LIST OF FIGURES</b> .....	<b>xii</b>
<b>INTRODUCTION</b> .....	<b>1</b>
Ocean acidification .....	1
Diatoms experience spatial and temporal fluctuations in CO <sub>2</sub> .....	1
Diatoms and the carbon concentrating mechanism .....	2
Diversification of the carbon concentrating mechanism .....	3
cAMP-mediated CO <sub>2</sub> sensing in diatoms .....	4
Silicon limited diatom growth and primary production .....	5
Silicon uptake and deposition .....	6
Molecular biology of silicon limitation .....	7
Thesis motivation and outline .....	9
References .....	15
<b>Manuscript 1: Transcriptome sequencing reveals contrasting silicon limitation responses in two <i>Thalassiosiroid</i> diatom species</b> .....	<b>26</b>
Abstract .....	27
Introduction .....	28
Materials and methods .....	34

Culturing, RNA extraction, and sequencing .....	34
Trimming, assembly, and read mapping .....	37
Annotation and gene ontologies .....	38
Ortholog finding with Silix .....	39
Differential expression analysis and hierarchical clustering .....	39
Silicon transporter phylogeny .....	41
Silaffins, cingulins, silacidins, frustulins, pleuralins, and polyamines .....	42
Results and discussion .....	43
Transcriptome assembly and annotation .....	43
Differential expression analysis and hierarchical clustering .....	44
Identification of genes sensitive to silicon status: silicon transporters and silicon efflux proteins .....	46
A subset of putative silaffins, cingulins, silacidins, frustulins, and pleuralins are differentially regulated in <i>Thalassiosiroids</i> .....	48
Genes involved in long-chain polyamine synthesis are not differentially regulated in <i>T. oceanica</i> or <i>T. weissflogii</i> .....	52
Identifying novel markers of silicon stress .....	53
ATP-grasp proteins are silicon sensitive and silicon specific .....	54
Identifying metabolic pathways with silicon sensitivity .....	56
Sell/MYND zinc finger domain gene expansion in <i>T. oceanica</i> .....	59
Conclusions .....	61
References .....	64

<b>Manuscript 2: <i>Thalassiosira</i> diatoms display distinct, species-specific strategies for responding to changes in pCO<sub>2</sub> .....</b>	<b>82</b>
Abstract .....	83
Introduction .....	84
Materials and methods .....	90
Culturing .....	90
Sequencing, mapping, differential expression .....	91
Identification and comparison of orthologs .....	92
Predicted protein localization .....	93
Results .....	95
Species-specific metabolic rearrangement, redox and ion homeostasis in low CO <sub>2</sub> ....	96
Bicarbonate transporters .....	97
Copy number variations in bicarbonate transporters .....	97
Varied bicarbonate transporter localization prediction and CO <sub>2</sub> sensitivity.....	98
Carbonic anhydrases .....	99
Copy number variations in carbonic anhydrases .....	99
Varied carbonic anhydrase localization predictions .....	100
Carbonic anhydrase CO <sub>2</sub> sensitivity .....	101
Putative C4 photosynthesis genes .....	102
cAMP secondary messengers and CO <sub>2</sub> /HCO <sub>3</sub> <sup>-</sup> sensing .....	104
Discussion .....	105
Key point of regulation for <i>T. oceanica</i> carbon concentrating mechanism: SLC4-mediated HCO <sub>3</sub> <sup>-</sup> uptake into the cytoplasm .....	105

<i>T. oceanica</i> decreased growth rate in high CO <sub>2</sub> due to ionic and redox stress .....	107
Key point of regulation for <i>T. weissflogii</i> carbon concentrating mechanism: cytosolic-predicted $\beta$ , $\delta$ , and $\zeta$ carbonic anhydrases preventing CO <sub>2</sub> leakage .....	108
<i>T. weissflogii</i> energetic trade-offs across CO <sub>2</sub> conditions .....	110
Key point of regulation for <i>T. rotula</i> carbon concentrating mechanism: SLC4- mediated HCO <sub>3</sub> <sup>-</sup> uptake into the plastid .....	110
<i>T. rotula</i> relies on degradation pathways to sustain growth in low CO <sub>2</sub> .....	112
Evidence for cAMP-mediated CO <sub>2</sub> sensing in <i>T. oceanica</i> , but not <i>T. weissflogii</i> or <i>T. rotula</i> .....	112
Conclusions .....	113
References .....	116
<b>Manuscript 3: Following diatom response to ocean acidification in mesocosm experiments with metatranscriptome sequencing indicates contrasting CO<sub>2</sub> responses in <i>Skeletonema</i> and <i>Thalassiosira</i> diatoms .....</b>	<b>133</b>
Abstract .....	134
Introduction .....	135
Materials and methods .....	138
Seawater collection and incubation .....	138
Macronutrients, carbonate, and chlorophyll measurements .....	139
DNA extraction .....	141
RNA extraction .....	141
DNA Amplicon sequencing .....	141

RNA sequencing .....	143
Two step read mapping approach: DIAMOND and BWA .....	143
Normalization .....	144
Differential transcript expression .....	144
Metatranscriptome rarefaction .....	145
Annotation and differential KO expression .....	145
Identification of carbon concentrating mechanism components and autolysis genes .....	146
Results and discussion .....	146
Chlorophyll, nutrients, and carbonate chemistry .....	146
Metatranscriptome reads are dominated by <i>Skeletonema</i> and <i>Thalassiosira</i> diatoms .....	149
Diatom genera show different patterns of expression of KOs for C, N, and P acquisition and metabolism across CO <sub>2</sub> conditions .....	150
Expression patterns in ion and redox homeostasis KO2 indicate diatoms experience increased oxidative stress in high CO <sub>2</sub> .....	153
Diatoms have distinct complements of carbon concentrating mechanism components that are regulated in response to CO <sub>2</sub> .....	154
Conclusions .....	156
References .....	158
<b>Appendix I .....</b>	<b>175</b>
<b>Appendix II .....</b>	<b>178</b>
<b>Appendix III .....</b>	<b>179</b>

## LIST OF TABLES

TABLE	PAGE
<b>MANUSCRIPT 1</b>	
Table 1. Summary table for <i>T. oceanica</i> and <i>T. weissflogii</i> showing the total number of transcripts assembled and number of transcripts with function or protein domain information assigned by KEGG automated annotation server or InterProScan.....	81
Table 2. Summary table for <i>T. oceanica</i> and <i>T. weissflogii</i> showing the number of transcripts differentially expressed in silicon limitation relative to nutrient replete conditions ( $\geq 0.95$ posterior probability of a 2-fold change). Numbers shown in the “Si specific” rows indicate the numbers of genes that are silicon-specifically differentially expressed and not part of a general stress response.....	81
<b>MANUSCRIPT 2</b>	
Table 1. Numbers of differentially expressed genes ( $\geq 0.95$ posterior probability of a 2-fold change in ASC) (Wu et al. 2010) across pCO <sub>2</sub> condition in each transcriptome.....	131
Table 2. Number of putative carbonic anhydrases from model diatom genomes and the transcriptomes described in this study. The localization and activity of the full complement of putative carbonic anhydrases in <i>P. tricornutum</i> and <i>T. pseudonana</i> have not been completely experimentally verified (Tachibana et al. 2011; Samukawa et al. 2014; Kikutani et al. 2016).....	132

## LIST OF FIGURES

FIGURE PAGE

### MANUSCRIPT 1

Figure 1. Depiction of the results from the hierarchical clustering of *T. oceanica* and *T. weissflogii* silicon sensitive genes, with only the results from silicon-specific clusters shown. Each panel shows the per-cluster gene expression values as a log<sub>2</sub>-fold change in silicon, iron, or nitrate limitation relative to the nutrient replete condition. The bar height indicates the per-cluster median expression value and the error bars indicate the interquartile range. The numbers at the bottom of each panel indicate the number of genes in each cluster.....76

Figure 2. Phylogenetic analysis of diatom silicon transporters. A maximum likelihood comparison for silicon transporters from *Fragilariopsis cylindrus*, *Thalassiosira pseudonana*, *T. rotula*, *T. oceanica*, and *T. weissflogii* was conducted. For each silicon transporter identified in this study, a heat map is also shown. The heat map colors indicate the log<sub>2</sub>-fold change of each putative silicon transporter in silicon (Si), iron (Fe), or nitrate (N) limited condition relative to the nutrient replete condition. The asterisk indicates which silicon transporters were statistically significantly differentially expressed. A 1:1 line is also inset, which shows the log<sub>10</sub> RPKM expression value in nutrient replete (x-axis) and silicon limited (y-axis) conditions. Each point corresponds to a silicon transporter and the shape of the point indicates which species the silicon transporter originates from. The black diagonal line indicates where

points would fall if they have the same expression value in the two conditions compared. .... 77

Figure 3. A series of 1:1 lines showing the expression values for putative silaffins, cingulins, silacidins, and frustulins identified in this study. Each point represents a single transcript. The x-axis is the log 10 transformed RPKM values in the nutrient replete condition and the y-axis is the log 10 transformed RPKM values in silicon limitation. The black diagonal line indicates where points would fall if they have the same expression value in the two conditions compared. The color of the point indicates the direction of differential expression in silicon limitation (up (red), down (blue), or no change) and the shape indicates whether or not each differentially expressed transcript is specific to silicon limitation or is part of a general stress response.

..... 78

Figure 4. Heat map overlaid on a pathway map of genes involved in the precursors of long-chain polyamine biosynthesis. Colors in the heat map indicate the log<sub>2</sub>-fold change in silicon (Si), iron (Fe), or nitrate (N) limitation relative to the nutrient replete condition. The full description for each KEGG orthology are as follows: K00138:aldB, aldehyde dehydrogenase; K00286:proC, pyrroline-5-carboxylate reductase; K01259:pip, proline iminopeptidase; K01476:arginase; K00819:rocD, ornithine--oxo-acid transaminase; K01581:ornithine decarboxylase; K00789:metK, S-adenosylmethionine synthetase; K01611:speD, S-adenosylmethionine decarboxylase; K00797:speE, spermidine synthase. .... 79

Figure 5. Heat map of the silicon specific, differentially expressed genes from *T. oceanica* and *T. weissflogii* and their KEGG orthology or silicon metabolism gene assignment categories. Colors in the heat map indicate the relative proportion of silicon specific genes in each category. Some categories that would not apply to diatoms were excluded from this analysis (e.g., Cardiovascular diseases). ..... 80

**MANUSCRIPT 2**

Figure 1. A heat map showing a subset of the differentially expressed genes from *T. weissflogii*, *T. rotula*, and *T. oceanica*. Warmer colors in the heat map indicate an upregulation and a cooler color indicates a downregulation. “Low” refers to the <230 ppm CO<sub>2</sub> condition, “projected” refers to the 690-1340 ppm pCO<sub>2</sub> condition, and “max” refers to the 2900-5100 ppm pCO<sub>2</sub> condition.....124

Figure 2. A maximum likelihood phylogeny of the bicarbonate transporters from the model diatoms *T. pseudonana* (JGI Thaps3 gene IDs) and *P. tricornutum* (JGI IDs or NCBI accession numbers), the transcriptomes assembled in this study (indicated by asterisks), and MAFFT homologs included for reference. All alignments were done in MAFFT (version 7) using the L-INS-i strategy with the BLOSUM62 scoring matrix, the “leave gappy regions” option selected, and the MAFFT-homologs option selected (Katoh & Standley 2013). RAxML maximum likelihood trees were computed with the following parameters: PROTGAMMA matrix, BLOSUM62 substitution

model, and 1000 bootstrap replicates. The tree was mid-point rooted in FigTree. Support values less than 70 are not shown.....125

Figure 3. Expression values of the carbonic anhydrases and bicarbonate transporters from the transcriptomes analyzed in this study. Each point represents a single transcript. The x-axis is the log 10 transformed RPKM values in the 320-390 ppm pCO<sub>2</sub> and the y-axis is the log 10 transformed RPKM values in the <230 ppm condition, the 690-1340 ppm pCO<sub>2</sub> condition, and the 2900-5100 ppm pCO<sub>2</sub> condition at the bottom of the figure. The black diagonal line indicates the same expression value in the two conditions compared. Circled points appearing above this line are upregulated in the condition indicated on the y-axis and circled points below it are downregulated in the condition indicated on the y-axis ( $\geq 0.95$  posterior probability of a 2-fold change in ASC) (Wu et al. 2010). The color of the points indicates the specific isozyme type and the shape of the point indicates from which organism the transcript originates.....126

Figure 4. A maximum likelihood phylogeny of the carbonic anhydrases from the model diatoms *T. pseudonana* (JGI Thaps3 gene IDs) and *P. tricornutum* (JGI IDs or NCBI accession numbers), the three *Thalassiosira* transcriptomes assembled in this study (indicated by asterisks), and two additional carbonic anhydrases from *Halothiobacillus neapolitanus* and *Plasmodium falciparum* for reference. All alignments were done in MAFFT (version 7) using the L-INS-i strategy with the BLOSUM62 scoring matrix, the “leave gappy regions”

option selected (Katoh & Standley 2013). RAxML maximum likelihood trees were computed with the following parameters: PROTGAMMA matrix, BLOSUM62 substitution model, and 1000 bootstrap replicates. The carbonic anhydrase tree was rooted using the gamma carbonic anhydrases because they are the most ancient carbonic anhydrase (Smith et al. 1999). Support values less than 70 are not shown. .... 128

Figure 5. Putative carbonic anhydrases (CAs) that have been localized in the model diatom *T. pseudonana* (from Samukawa et al. 2014) and the predicted localization of putative CAs from *T. oceanica* and *T. weissflogii*. The localizations of the putative CAs reported in this study have not been experimentally verified and are based on the predictions from SignalP (V4.1), TargetP (V1.1), and ASA-find (1.1.5) (Bendtsen et al. 2004; Emanuelsson et al. 2007; Gruber et al. 2015). SignalP was run against the eukaryote organism group and the default cutoff values were used. TargetP was run with the plant model and “winner-takes-all” option selected. PPS = periplastid space, Cyt = cytoplasm, mit = mitochondria, CER = chloroplast endoplasmic reticulum, PPC = periplastidal compartment, CEV = chloroplast envelope, Str = chloroplast stroma, Pyr = pyrenoid. .... 129

Figure 6. A schematic summarizing the enzymes thought to be involved in the diatom CCM, including variations of C4 metabolism (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014). Grey boxes indicate enzymes performing the conversions between the metabolites on either end of the black arrow. Grey boxes: AGAT = alanine glyoxylate amino transferase, ALAT = alanine amino

transferase, CA = carbonic anhydrase, ENO = enolase, GDC = glycine decarboxylase P-protein, GK = glycerate kinase, HPR = hydroxypyruvate reductase, MDH = malate dehydrogenase, ME = malic enzyme, PEPC = phosphoenolpyruvate carboxylase, PEPCCK = phosphoenolpyruvate carboxykinase, PGM = phosphoglycerate mutase, PPK = pyruvate, phosphate dikinase, PYC = pyruvate carboxylase, SLC4/SLC26 = solute carrier family 4/6, SPT = serine pyruvate transaminase. Outside grey boxes: CO<sub>2</sub> = carbon dioxide, GLC = glycerate, GLX = glyoxylate, HCO<sub>3</sub><sup>-</sup> = bicarbonate, MAL = malate, NH<sub>4</sub><sup>+</sup> = ammonium, OAA = oxaloacetate, PEP = phosphoenolpyruvate, PYR = pyruvate ..... 130

### MANUSCRIPT 3

Figure 1. (A) ppm CO<sub>2</sub>, (B) Chlorophyll L<sup>-1</sup>, (C) μM NO<sub>3</sub><sup>-</sup>+NO<sub>2</sub><sup>-</sup>, (D) μM PO<sub>4</sub>, (E) μM Si(OH)<sub>4</sub>, (F) μM POC, (G) μM PON, (H) μM POP, (I) μM BSi, and (J) temperature °C for each incubation carboy. The x-axis indicates the date the variable was measured and the y-axis indicates the values and variables measured. The shapes and colors indicate the CO<sub>2</sub> condition shown. Points and error bars indicate the mean and range of the measurement across the triplicate incubation carboys. The asterisks indicate significant differences relative to the present day pCO<sub>2</sub> condition (1-way ANOVA; p < 0.05). Dilution days are indicated with black arrows. Solid shapes indicate that the measurement shown was taken before dilution, empty shapes indicate the measurement was taken post-dilution. .... 165

Figure 2. A: Bar charts showing the relative distributions of metatranscriptome reads mapped to the MMETSP database using either DIAMOND or BWA, moving from less specific to more specific from left to right. B: DADA2 taxonomic assignment of 18S V4 rDNA amplicon reads using the PR2 database. .... 166

Figure 3. A pathway heat map showing broad metabolic rearrangements in the citrate cycle in *Skeletonema* and *Thalassiosira* across CO<sub>2</sub> conditions. The color indicates the log<sub>2</sub>-fold change of total reads mapped per genus per KEGG KO in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> condition. .... 168

Figure 4. A heat map showing broad metabolic rearrangements in nitrogen uptake pathway members in *Skeletonema* and *Thalassiosira* across CO<sub>2</sub> conditions. The color indicates the log<sub>2</sub>-fold change of total reads mapped per genus per KEGG KO in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> condition. .... 169

Figure 5. A pathway heat map showing broad metabolic rearrangements in branched-chain amino acid metabolism in *Skeletonema* and *Thalassiosira* across CO<sub>2</sub> conditions. The color indicates the log<sub>2</sub>-fold change of total reads mapped per genus per KEGG KO in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> condition. .... 170

Figure 6. A pathway heat map showing broad metabolic rearrangements in phosphorus metabolism in *Skeletonema* and *Thalassiosira* across CO<sub>2</sub> conditions. The color indicates the log<sub>2</sub>-fold change of total reads mapped per

genus per KEGG KO in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> condition. .... 171

Figure 7. A pathway heat map showing broad metabolic rearrangements in phosphorus metabolism in *Skeletonema* and *Thalassiosira* across CO<sub>2</sub> conditions. The color indicates the log<sub>2</sub>-fold change of total reads mapped per genus per KEGG KO in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> condition. .... 172

Figure 8. A series of 1 to 1 lines showing the expression values of the  $\alpha$ ,  $\delta$ ,  $\gamma$ , and  $\zeta$  carbonic anhydrases from *Skeletonema* spp. and *Thalassiosira* spp. diatoms. Each point indicates a gene and the color indicates which genera of diatom the sequence is from. The x-axis is the expression value in present day pCO<sub>2</sub> and the y-axis is the expression value in either low or high pCO<sub>2</sub>. The black line indicates the point at which the expression of a gene is the same in both conditions. Genes that were differentially expressed ( $\geq 0.95$  posterior probability of  $>2$ -fold change) are indicated by the black star) (Wu et al. 2010). .... 173

Figure 9. A series of 1 to 1 lines showing the expression values of the SLC4 and SLC26 bicarbonate transporters from *Skeletonema* spp. and *Thalassiosira* spp. diatoms. Each point indicates a gene and the color indicates which genera of diatom the sequence is from. The x-axis is the expression value in present day pCO<sub>2</sub> and the y-axis is the expression value in either low or high pCO<sub>2</sub>. The black line indicates the point at which the expression of a gene is the same in both conditions. Genes that were differentially expressed ( $\geq 0.95$  posterior

probability of >2-fold change) are indicated by the black star) (Wu et al. 2010).

..... 174

## INTRODUCTION

### **Ocean acidification**

Fossil fuel combustion has raised atmospheric pCO<sub>2</sub> from ~277 ppm in the year 1750 to the present level of ~400 ppm (Tans 2009; Le Quéré et al. 2016).

Increased levels of dissolved CO<sub>2</sub> gas in seawater perturbs carbonate chemistry in the ocean, which results in a decreased buffering capacity, or ocean acidification (OA) (Caldeira & Wickett 2003). Climate models suggest that pCO<sub>2</sub> levels in the atmosphere may rise to ~1000 ppm by the year 2100 if no measures are taken to reduce emissions (Solomon et al. 2007). Through physical mixing and biological activity, the ocean sequesters approximately 1/3 of anthropogenic CO<sub>2</sub> and has been the only net sink for CO<sub>2</sub> in the past 200 years (Sabine et al. 2004; Khatiwala et al. 2013).

### **Diatoms experience spatial and temporal fluctuations in CO<sub>2</sub>**

Diatoms are photosynthetic single-celled eukaryotes that account for 40% of oceanic primary production and approximately 20% of global primary production (Nelson et al. 1995; Field et al. 1998). Diatoms have experienced fluctuations in pCO<sub>2</sub> on different temporal scales throughout their evolutionary history. Molecular clock estimates suggest that diatoms arose around 240 million years ago (Sims et al. 2006; Medlin 2016). This coincides with the start of a glacial “hot-house” period, where atmospheric pCO<sub>2</sub> levels were as high as 6000 ppm (Kump et al. 2009). In more recent history, diatoms have also experienced low pCO<sub>2</sub>: pre-industrial (before 1850) atmospheric pCO<sub>2</sub> averaged ~270 ppm (Tans 2009). In the present day, diatoms can also experience fluctuations in pCO<sub>2</sub> levels that vary on much shorter time scales.

Coastal diatoms might experience dissolved  $p\text{CO}_2$  concentrations greater than 850 ppm during seasonal coastal upwelling events that entrain  $\text{CO}_2$ -rich deep water into the lit ocean (Feely et al. 2008). Diatoms' distributions in the ocean can also influence the  $p\text{CO}_2$  conditions they experience over the course of their life histories. Relative to coastal diatoms, diatoms living in the open-ocean experience more stable  $p\text{CO}_2$  conditions (Hofmann et al. 2011). Coastal diatoms may have unique strategies for coping in high  $\text{CO}_2$  that are not seen in open-ocean diatoms.

### **Diatoms & the carbon concentrating mechanism**

Diatoms' key role in global carbon cycles is aided by a carbon-concentrating mechanism (CCM), which concentrates  $\text{CO}_2$  around RuBisCO, the enzyme that begins the  $\text{CO}_2$  fixation process (Badger et al. 1998; Hopkinson et al. 2011). The diatom CCM uses bicarbonate transporters (BCTs) to take up  $\text{HCO}_3^-$  and carbonic anhydrases (CAs) to interconvert  $\text{HCO}_3^-$  and  $\text{CO}_2$  (Badger et al. 1998; Hopkinson et al. 2011). The predominant dissolved form of inorganic carbon in the ocean is bicarbonate ( $\text{HCO}_3^-$ ) (Zeebe & Wolf-Gladrow 2001), which is transported into the cell by the BCTs and can be interconverted to  $\text{CO}_2$  (which diffuses across membranes without active transport) via the activity of a suite of carbonic anhydrases (Meldrum & Roughton 1933; Stadie & O'Brien 1933). The use of a carbon concentrating mechanism and active transport of inorganic carbon requires a significant energetic investment by the cell (Raven 1991). If diatoms can downregulate CCM components when  $\text{CO}_2$  levels are high, this may result in the conservation of cellular energy as anthropogenic  $\text{CO}_2$  emissions rise. This energy conservation may confer a competitive advantage to some diatoms and other groups of phytoplankton and could influence community composition in the

future. The ability or inability to regulate the CCM in response to fluctuating pCO<sub>2</sub> may be related to how diatoms sense changes in pCO<sub>2</sub>.

### **Diversification of the diatom CCM**

The diatom CCM has been most well-studied in the model diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* and includes the  $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\gamma$ ,  $\theta$ , and  $\zeta$  CAs and the SLC4 and SLC26 BCTs (Tachibana et al. 2011; Samukawa et al. 2014; Kikutani et al. 2016; Hopkinson 2014; Hopkinson et al. 2016). *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* generally contain the same complement of CCM genes, with a few exceptions: The  $\delta$  and  $\zeta$  CA isozymes have not been found in *P. tricornutum*, while the  $\beta$  CAs have not been found in *T. pseudonana* (Tachibana et al. 2011; Samukawa et al. 2014). The  $\theta$  carbonic anhydrases have been characterized in *P. tricornutum*, but their expression in *T. pseudonana* has not been confirmed (Kikutani et al. 2016). The SLC4 transporters have been characterized in *P. tricornutum*, but not *T. pseudonana* (Nakajima et al. 2013). The sub-cellular arrangements of the CCM components (and thus the key points of CCM regulation) also differ in these two species, suggesting that their CCMs work quite differently (Tachibana et al. 2011; Samukawa et al. 2014; Hopkinson et al. 2016). While *T. pseudonana* has CAs localized to the periplastid space, the cytosol, and the chloroplast stroma, *P. tricornutum* does not (Samukawa et al. 2014). In contrast, *P. tricornutum* has CAs localized to the chloroplast endoplasmic reticulum and the pyrenoid, *T. pseudonana* does not (Tachibana et al. 2011). It has been hypothesized that this diversification and re-configuration of the CCM in different diatoms is due to historic fluctuations in atmospheric CO<sub>2</sub> (Shen et al. 2017). However, there are few studies

examining the CO<sub>2</sub>-sensitivity and regulation of the CCM across a diversity of diatoms. This is particularly true of diatoms living in the open-ocean that may not have the same high CO<sub>2</sub>-coping mechanisms as coastal isolates (like *T. pseudonana*).

### **cAMP-mediated CO<sub>2</sub> sensing in diatoms**

One aspect of carbon physiology that might influence how diatoms cope in high CO<sub>2</sub> is their ability to sense changes in CO<sub>2</sub>. Sensing of pCO<sub>2</sub> in diatoms is via the activity of cAMP secondary messengers (Harada et al. 2006; Ohno et al. 2012; Hennon et al. 2015). Soluble adenylyl cyclases generate cAMP secondary messengers from ATP in response to signals received from bicarbonate (Buck et al. 1999). The cAMP secondary messengers activate cAMP-dependent protein kinases that phosphorylate and activate bZIP domain transcription factors that in turn bind to cAMP response elements and regulate downstream gene expression (Ohno et al. 2012). Phosphodiesterases regulate cellular cAMP levels by dephosphorylating cAMP back into AMP (Conti & Beavo 2007). Studies in *P. tricornutum* and *T. pseudonana* suggest that the activity of a soluble adenylyl cyclase is increased in the presence of CO<sub>2</sub>/HCO<sub>3</sub>, and that the resulting increase in cellular cAMP levels induces the activity of bZIP transcription factors (Harada et al. 2006; Ohno et al. 2012; Hennon et al. 2015). In high CO<sub>2</sub>, the transcription factors bind to upstream CO<sub>2</sub>-cAMP-responsive elements and repress the expression of CCM genes, including the expression of the adenylyl cyclase itself (Hennon et al. 2015). The pCO<sub>2</sub> sensitivity of the β carbonic anhydrase of *P. tricornutum* was repressed when cells were treated with a cAMP inhibitor IBMX (Harada et al. 2006) or when cAMP-responsive motifs were deleted upstream of the β carbonic anhydrase gene (Ohno et al. 2012). These studies of

cAMP-mediated CO<sub>2</sub> sensing have been conducted in the model diatoms *T. pseudonana* (Hennon et al. 2015) and *P. tricornutum* (Harada et al. 2006; Ohno et al. 2012). In light of the diversification and re-configuration of the CCM in different diatoms (Shen et al. 2017), it is possible that this cAMP-mediated CO<sub>2</sub>-sensing mechanism is not ubiquitous across diatoms.

### **Silicon limits diatom growth and primary production**

In addition to being aided by the CCM, diatoms' primary production is also a function of their growth, which is constrained by the availability of nutrients in the ocean. Many phytoplankton share growth requirements for nutrients such as nitrate, phosphate, and iron, but diatoms are unique among the phytoplankton as they require silicon, which is used to build their ornately silicified cell walls or frustules (Lewin 1957; Lewin 1962; Jørgensen 1955).

It has been suggested that silicon ultimately limits diatom growth throughout the ocean (Dugdale et al. 1995; Yool & Tyrrell 2003) by several different mechanisms. Firstly, there is a discrepancy in the available pools of recycled nitrogen and recycled silicon in the surface ocean. Silicon recycling occurs deeper in the water column than that of other nutrients (C, N, P), leading to a propensity for silicon depletion in the surface ocean (Dugdale et al. 1995; Treguer et al. 1995; Dugdale & Wilkerson 1998). Secondly, diatoms growing in upwelled water can become silicon limited if these waters are enriched in nitrate relative to silicon, which been observed in the North Atlantic Ocean (Allen et al. 2005; Johnson et al. 2007). Thirdly, silicon may also set the upper limit on diatom primary production in iron limited high nutrient low chlorophyll (HNLC) regions (Dugdale et al. 1995; Dugdale & Wilkerson 1998).

This is because diatoms growing low iron conditions ( $< 5 \text{ nM}$ ), take up as much as 8 times more silicon than nitrate (Franck et al. 2000). Thus, diatoms in HNLC regions may be driven into silicon limitation due to their increased silicon requirement (Hutchins & Bruland 1998; De La Rocha et al. 2000). It has also been suggested that silicon processing may require iron as a cofactor, which would lead to a biochemically-dependent iron and silicon co-limitation (Mock et al. 2008; Saito et al. 2008). Silicon and trace metal co-limitation of diatoms has been documented in the HNLC waters of the Equatorial Pacific Ocean (Brzezinski et al. 2011), the Southern Ocean and sub-Antarctic (Boyd et al. 1999; Leblanc et al. 2005; Franck et al. 2000), and the sub-Arctic Pacific (Koike et al. 2001). As these studies assessed the silicon status of the bulk diatom community, it is currently not known whether these responses are uniform across particular diatom genera, species, or strains. Genetic markers have been used to probe iron status in the field, and similar molecular markers of silicon status could be powerful tools to probe the silicon status of different co-existing diatom species (Chappell et al. 2015).

### **Silicon uptake and deposition**

Silicic acid is a small, uncharged molecule and can freely diffuse across diatom membranes when concentrations are high (Hildebrand et al. 2018), but when silicic acid concentration is low, silicon transporters are used for active transport (Kimberlee Thamatrakoln & Hildebrand 2008). Diatoms' silicified cell wall is formed inside the silicon deposition vesicle (Drum & Pankratz 1964; Reimann et al. 1966). Silicification inside the silicon deposition vesicle is aided by several classes of molecules -- the silaffins (Poulsen & Kröger 2004; Kröger et al. 1999), cingulins (Scheffel et al. 2011),

silacidins (Wenzl et al. 2008), and long-chain polyamines (Kröger et al. 2000) all play a role in silicon precipitation and frustule formation. Pleuralins and frustulins are tightly associated with the silicified cell wall (Kröger et al. 1997; Kröger et al. 1994), but they are not thought to be directly involved in silicon deposition in the silicon deposition vesicle (Kröger & Wetherbee 2000; van de Poll et al. 1999).

Diatom silicon physiology and the cell cycle are tightly linked. In addition to requiring silicon for their cell walls (Lewin 1957; Lewin 1962; Jørgensen 1955), silicon is also required for diatoms' DNA synthesis to proceed during S phase of the cell cycle (Darley & Volcani 1969). This phase of the cell cycle also sees peak expression of silicon transporter mRNA (but the lowest expression of silicon transporter proteins) (Thamatrakoln & Hildebrand 2007). Diatom frustule valves are formed just after cell division, during G2/M phase, while girdle band synthesis happens in G1 phase (Eppley et al. 1967; Hildebrand et al. 2007). These different processes that require silicon explain why silicon limited *T. weissflogii* cells sometimes arrest in the G2/M phase or other times in the G1/S phase transition (Brzezinski et al. 1990; Vaultot et al. 1987).

### **Molecular biology of silicon limitation**

Genetic studies have given us a better understanding of the molecular underpinnings of diatom silicon physiology. Silicon limited *T. pseudonana* upregulates silicon transporter gene copies SIT1 and SIT2 (transcripts and proteins) in silicon limitation (Mock et al. 2008; Shrestha et al. 2012; Du et al. 2014; Smith et al. 2016; Brembu et al. 2017) and also has a higher concentration of silacidin proteins relative to silicon replete condition (Richthammer et al. 2011), although transcriptome

studies in *T. pseudonana* did not see differential expression of silacidin genes (Mock et al. 2008; Shrestha et al. 2012). Genes for the biosynthesis of long-chain polyamine precursors (spermidine synthases) (Bouchereau et al. 1999) and the silaffins were upregulated during silicon limitation (Mock et al. 2008; Brembu et al. 2017), while others were upregulated only after silicon resupply (Shrestha et al. 2012; Brembu et al. 2017), leading to the hypothesis that distinct silaffin gene copies or LCPAs are optimized for silicon deposition in either low or high silicon concentrations. In addition to SIT1 and SIT2, *T. pseudonana* also has a third silicon transporter, SIT3, that is lowly expressed and not sensitive to changes in silicon, leading to the hypothesis that it functions as a silicon sensor rather than a transporter (Thamatrakoln & Hildebrand 2007). Cingulins were downregulated during silicon limitation and upregulated after silicon resupply, but were also upregulated during the S phase of the cell cycle (when silicon deposition does not occur) suggesting that cingulins have other roles in the cell aside from silicon deposition (Shrestha et al. 2012; Brembu et al. 2017). Given that diatoms also require silicon for DNA synthesis during S phase of the cell cycle (Darley & Volcani 1969), when cingulin expression increases (Shrestha et al. 2012; Brembu et al. 2017), it is also feasible that the cingulins play a role in signaling cellular silicon status before DNA synthesis proceeds. Notably, the studies to date have not included diatoms found in HNLC regions, where silicon sets an upper limit on diatom production (Dugdale & Wilkerson 1998; Chappell et al. 2013), or the North Atlantic Ocean where silicon limitation of the bulk diatom community has been observed (Allen et al. 2005; Johnson et al. 2007), limiting the ability to develop appropriate molecular markers.

## Thesis motivation and outline

Diatom primary production is a function of their overall growth rate, which is constrained by the availability of dissolved nutrients in the surface ocean. Many phytoplankton share growth requirements for nutrients such as nitrate, phosphate, and iron, but diatoms are unique among the phytoplankton as they require silicon, which is used to build their ornately silicified cell walls or frustules (Lewin 1957; Lewin 1962; Jørgensen 1955). Silicon growth limitation of the bulk diatom community has been observed in the North Atlantic Ocean (Allen et al. 2005; Johnson et al. 2007) and silicon and trace metal co-limitation has been documented in the HNLC waters of the Equatorial Pacific Ocean (Brzezinski et al. 2011), the Southern Ocean and sub-Antarctic (Boyd et al. 1999; Leblanc et al. 2005; Franck et al. 2000), the sub-Arctic Pacific (Koike et al. 2001), and the equatorial Pacific (Dugdale & Wilkerson 1998). As these were studies assessing the bulk diatom community, it is unclear whether or not certain diatom species are experiencing silicon limitation while others are not. Molecular markers have been useful for assessing the iron status of a diatom species, *Thalassiosira oceanica*, that is prevalent across oceanic gradients (Chappell et al. 2015), and a similar approach could prove useful for determining whether or not *T. oceanica* and other co-occurring diatoms experience silicon limitation in a bulk community. However, most studies looking at the molecular and genetic response to silicon limitation have used the model diatoms *T. pseudonana* (Mock et al. 2008; Shrestha et al. 2012; Du et al. 2014; Smith et al. 2016; Brembu et al. 2017) or *P. tricornutum* (Sapriel et al. 2009), rather than diatoms typically found in places where silicon is thought to limit diatom primary production (Dugdale & Wilkerson 1998;

Allen et al. 2005; Johnson et al. 2007), hindering our ability to develop appropriate molecular markers.

Chapter 1 of this thesis examined the transcriptomes of two closely related *Thalassiosira* diatoms to better understand the silicon limitation gene expression response of diatoms found in ecosystems where their growth may be constrained by silicon availability, toward the goal of developing appropriate molecular markers of silicon status. *Thalassiosira oceanica* CCMP1005 has a high tolerance for iron limitation (Sunda & Huntsman 1995) and may persist in chronically iron limited HNLC regions (Chappell et al. 2013) where diatoms' silicon requirements are high (Hutchins & Bruland 1998; De La Rocha et al. 2000). *Thalassiosira weissflogii* CCMP1010 was isolated from the North Atlantic Ocean, an ocean basin where silicon-depleted upwelling has been observed (Allen et al. 2005; Johnson et al. 2007). We selected this genus as it allows us to make comparisons to studies in *T. pseudonana* which has a completely sequenced genome (Armbrust et al. 2004) and its silicon metabolism and molecular biology is very well studied (Poulsen & Kröger 2004; Thamtrakoln & Hildebrand 2007; Mock et al. 2008; Wenzl et al. 2008; Scheffel et al. 2011; Shrestha et al. 2012; Du et al. 2014; Smith et al. 2016; Brembu et al. 2017). While *T. weissflogii* upregulated the silicon transporters in low silicon, *T. oceanica* did not. We hypothesize that this is due to the lack of a SIT3-like silicon sensor in *T. oceanica*. We also generally found very different silicon limitation responses in these two diatoms. We hypothesize that this is because silicon is required for cell wall synthesis (Lewin 1957; Lewin 1962; Jørgensen 1955) and DNA replication (Darley & Volcani 1969) in diatoms and that these two processes are being differentially affected

in silicon limited *T. oceanica* and *T. weissflogii*. However, we did find a family of ATP-grasp domain proteins of unknown function that were upregulated in silicon-limited *T. oceanica*, *T. weissflogii*, and *T. pseudonana* (Smith et al. 2016; Brembu et al. 2017). The upregulation of these genes is unique to silicon limitation and their expression is not induced in nitrate or iron limitation conditions. It is also found across a diversity of diatoms in the MMETSP database (Keeling et al. 2014), suggesting that they might be useful as a silicon limitation markers in many diatoms, in addition to *Thalassioisra* spp. diatoms.

Increased fossil fuel consumption since the industrial revolution has increased CO<sub>2</sub> levels in the atmosphere and in the ocean (Tans 2009; Le Quéré et al. 2016). The increased CO<sub>2</sub> in the ocean has perturbed carbonate chemistry and led to ocean acidification (Caldeira & Wickett 2003). Diatoms have experienced large fluctuations in CO<sub>2</sub> conditions on evolutionary time scales (Sims et al. 2006; Kump et al. 2009; Tans 2009; Medlin 2016), and it has been suggested that these historical fluctuations in CO<sub>2</sub> have led to a diversification and reconfiguration of the diatom CCM (Shen et al. 2017). Most studies of the diatom CCM have used the model diatoms *T. pseudonana* and *P. tricornutum*, and they suggest that the CCM of these diatoms works very differently (Tachibana et al. 2011; Samukawa et al. 2014; Kikutani et al. 2016; Hopkinson 2014; Hopkinson et al. 2016). Further, CCM function and evolution may have been influenced by diatom biogeography, as open-ocean species typically experience stable CO<sub>2</sub> conditions that are in equilibrium with atmospheric levels of CO<sub>2</sub> and coastal diatoms may experience very high CO<sub>2</sub> conditions during seasonal upwelling of high CO<sub>2</sub> deep water (Feely et al. 2008; Hofmann et al. 2011). There is

currently a paucity of studies of the diatom CCM and its sensitivity to CO<sub>2</sub> across a diversity of diatom species with representatives from a range of biogeographical distributions in the ocean.

Chapter 2 of this thesis adds to our knowledge of the diversity of CCM and high CO<sub>2</sub> response in different species of the *Thalassiosira* genus of diatoms. CO<sub>2</sub> manipulation experiments were conducted with four *Thalassiosira* species – *T. pseudonana*, *T. rotula*, *T. weissflogii*, and *T. oceanica*, and transcriptomes were sequenced from the latter 3 species. *T. rotula* is a coastal diatom that may have experienced high CO<sub>2</sub> during seasonal upwelling, while *T. oceanica* and *T. weissflogii* are open-ocean isolates that have likely experienced relatively stable CO<sub>2</sub> conditions (Feely et al. 2008; Hofmann et al. 2011). These species displayed a range of growth rate changes across CO<sub>2</sub> conditions (<230 ppm CO<sub>2</sub> to >2900 ppm CO<sub>2</sub>), from no change across conditions (*T. pseudonana*), slower growth in lower CO<sub>2</sub> (*T. weissflogii*), slower growth in higher CO<sub>2</sub> (*T. oceanica*), and faster growth in higher CO<sub>2</sub> (*T. rotula*) (King et al. 2015). The accompanying transcriptome evidence for *T. rotula*, *T. weissflogii*, and *T. oceanica* indicate that the molecular basis for these differences in CO<sub>2</sub> responses boil down to the ability to dynamically regulate the CCM while maintaining internal redox and ion homeostasis, as well as trade-offs between funneling excess CO<sub>2</sub> into internal storage pools versus a higher growth rate (Wallace et al. in prep).

Chapter 3 uses a metatranscriptome approach to generalize about how the CCM of *Thalassiosira* spp. differs from that of *Skeletonema* spp. diatoms. Community-level global gene expression profiling (metatranscriptome) studies are a

powerful approach that capture the molecular basis for different diatoms' contrasting growth and nutrient acquisition strategies in mixed assemblages. This study involved CO<sub>2</sub> manipulation experiments of a plankton assemblage collected from Massachusetts Bay during March of 2014. Seawater was collected in acid-cleaned carboys, pre-filtered to remove grazers larger than 180 μm, amended with nutrients to mimic the mixed layer depth, and pCO<sub>2</sub> levels were manipulated to approximate low/pre-industrial (less than 215 ppm pCO<sub>2</sub>), present-day ambient pCO<sub>2</sub> in the open ocean (330-390 ppm pCO<sub>2</sub>), and future pCO<sub>2</sub> projections (780-1320 ppm pCO<sub>2</sub>) (Tans 2009; Le Quéré et al. 2016). Metatranscriptome and diatom-specific 18S rDNA amplicon sequencing (Zimmermann et al. 2011; Mellett et al. 2018) suggest that *Thalassiosira* spp. and *Skeletonema* spp. were abundant in these incubations, consistent with previous light microscopy observations conducted at the final experimental time point, as well as previous studies of Massachusetts Bay in March (Hunt et al. 2010). We show evidence of distinct enzymatic strategies for N and P uptake and scavenging in these two diatom genera. There is also evidence of distinct strategies in their CCMs – while *Thalassiosira* diatoms seem to use the α, δ, and ζ carbonic anhydrases and the substrate-specific SLC4 and non-specific SLC26 bicarbonate transporters, *Skeletonema* relies on the γ and ζ carbonic anhydrases and the SLC26 bicarbonate transporters or diffuse CO<sub>2</sub> uptake (Nakajima et al. 2013; Alper & Sharma 2013). This suggests that the co-existence of *Skeletonema* spp. and *Thalassiosira* spp. might be partially due to a CCM that use distinct CA isozymes that use different metal cofactors for their enzymatic activity (γ can substitute iron, δ can substitute cobalt) (Lane & Morel 2000; Tripp et al. 2004). Furthermore, if

*Skeletonema* does rely on diffusive CO<sub>2</sub> uptake rather than active transport of HCO<sub>3</sub><sup>-</sup> using SLC4 or SLC26 transporters, *Thalassiosira* may have a competitive advantage in high CO<sub>2</sub> as it downregulates these active transporters.

## References

- Allen, J.T. et al., 2005. Diatom carbon export enhanced by silicate upwelling in the northeast Atlantic. *Nature*, 437(7059), pp.728–732.
- Alper, S.L. & Sharma, A.K., 2013. The SLC26 Gene Family of Anion Transporters and Channels. *Molecular Aspects of Medicine*, 34(2–3), pp.494–515.
- Armbrust, E.V. et al., 2004. The Genome of the Diatom *Thalassiosira pseudonana*: Ecology, Evolution, and Metabolism. *Science*, 306(October), pp.79–86.
- Badger, M.R. et al., 1998. The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO<sub>2</sub>-concentrating mechanisms in algae. *Canadian Journal of Botany*, 76(6), pp.1052–1071.
- Bouchereau, A. et al., 1999. Polyamines and environmental challenges: recent development. *Plant Science*, 140, pp.103–125.
- Boyd, P. et al., 1999. Role of iron, light, and silicate in controlling algal biomass in subantarctic waters SE of New Zealand. *Journal of Geophysical Research*, 104, pp.395–408.
- Brembu, T. et al., 2017. Dynamic responses to silicon in *Thalassiosira pseudonana* - Identification, characterisation and classification of signature genes and their corresponding protein motifs. *Scientific Reports*, 7(1), pp.1–14.
- Brzezinski, M. a. et al., 2011. Co-limitation of diatoms by iron and silicic acid in the equatorial Pacific. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 58(3–4), pp.493–511.
- Brzezinski, M., Olson, R. & Chisholm, S., 1990. Silicon availability and cell-cycle progression in marine diatoms. *Marine Ecology Progress Series*, 67(M), pp.83–96.

- Buck, J. et al., 1999. Cytosolic adenylyl cyclase defines a unique signaling molecule in mammals. *Proceedings of the National Academy of Sciences*, 96(1), pp.79–84.
- Caldeira, K. & Wickett, M.E., 2003. Oceanography: anthropogenic carbon and ocean pH. *Nature*, 425(6956), p.365.
- Chappell, P.D. et al., 2015. Genetic indicators of iron limitation in wild populations of *Thalassiosira oceanica* from the northeast Pacific Ocean. *The ISME Journal*, 9, pp.592–602.
- Chappell, P.D. et al., 2016. Preferential depletion of zinc within Costa Rica upwelling dome creates conditions for zinc co-limitation of primary production. *Journal of Plankton Research*, 38(2), pp.244–255.
- Chappell, P.D. et al., 2013. *Thalassiosira* spp. community composition shifts in response to chemical and physical forcing in the northeast Pacific Ocean. *Frontiers in Microbiology*, 4(SEP), pp.1–14.
- Conti, M. & Beavo, J., 2007. Biochemistry and Physiology of Cyclic Nucleotide Phosphodiesterases: Essential Components in Cyclic Nucleotide Signaling. *Annual Review of Biochemistry*, 76(1), pp.481–511.
- Darley, W.M. & Volcani, B.E., 1969. Role of silicon in diatom metabolism. A silicon requirement for deoxyribonucleic acid synthesis in the diatom *Cylindrotheca fusiformis* Reimann and Lewin. *Experimental cell research*, 58(2), pp.334–342.
- Drum, R.W. & Pankratz, S.H., 1964. Post mitotic fine structure of *Gomphonema parvulum*. *Journal of ultrastructure research*, 10, pp.217–223.
- Du, C. et al., 2014. ITRAQ-based proteomic analysis of the metabolism mechanism associated with silicon response in the marine diatom *Thalassiosira pseudonana*.

- Journal of Proteome Research*, 13(2), pp.720–734.
- Dugdale, R.C. & Wilkerson, F.P., 1998. Silicate regulation of new production in the equatorial Pacific upwelling. *Nature*, 391(January), pp.270–273.
- Dugdale, R.C., Wilkerson, F.P. & Minas, H.J., 1995. The role of a silicate pump in driving new production. *Deep-Sea Research Part I: Oceanographic Research Papers*, 42(5), pp.697–719.
- Eppley, R.W., Holmes, R.W. & Paasche, E., 1967. Periodicity in cell division and physiological behavior of *Ditylum brightwellii* a marine planktonic diatom during grow in light-dark cycles. *Archiv Fur Mikrobiologie*, 56, pp.305–323.
- Feely, R. a et al., 2008. Evidence for upwelling of corrosive “acidified” water onto the continental shelf. *Science*, 320(5882), pp.1490–1492.
- Field, C.B. et al., 1998. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, 281(5374), pp.237–240.
- Franck, V.M. et al., 2000. Iron and silicic acid concentrations regulate Si uptake north and south of the Polar Frontal Zone in the Pacific Sector of the Southern Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 47(15–16), pp.3315–3338.
- Harada, H. et al., 2006. CO<sub>2</sub> sensing at ocean surface mediated by cAMP in a marine diatom. *Plant physiology*, 142(3), pp.1318–1328.
- Hennon, G.M.M. et al., 2015. Diatom acclimation to elevated CO<sub>2</sub> via cAMP signalling and coordinated gene expression. *Nature Climate Change*, 5(August), pp.761–766.
- Hildebrand, M., Frigeri, L.G. & Davis, A.K., 2007. Synchronized growth of

- Thalassiosira pseudonana* (Bacillariophyceae) provides novel insights into cell-wall synthesis processes in relation to the cell cycle. *Journal of Phycology*, 43(4), pp.730–740.
- Hildebrand, M., Lerch, S.J.L. & Shrestha, R.P., 2018. Understanding Diatom Cell Wall Silicification—Moving Forward. *Frontiers in Marine Science*, 5(April), pp.1–19.
- Hofmann, G.E. et al., 2011. High-Frequency Dynamics of Ocean pH : A Multi-Ecosystem Comparison. *PLoS ONE* , 6(12).
- Hopkinson, B.M., 2014. A chloroplast pump model for the CO<sub>2</sub> concentrating mechanism in the diatom *Phaeodactylum tricornutum*. *Photosynthesis Research*, 121(2–3), pp.223–233.
- Hopkinson, B.M. et al., 2011. Efficiency of the CO<sub>2</sub>-concentrating mechanism of diatoms. *Proceedings of the National Academy of Sciences*, 108(10), pp.3830–3837.
- Hopkinson, B.M., Dupont, C.L. & Matsuda, Y., 2016. The physiology and genetics of CO<sub>2</sub> concentrating mechanisms in model diatoms. *Current Opinion in Plant Biology*, 31, pp.51–57.
- Hunt, C.D. et al., 2010. Phytoplankton patterns in Massachusetts Bay-1992-2007. *Estuaries and Coasts*, 33(2), pp.448–470.
- Hutchins, D. a & Bruland, K.W., 1998. Iron-limited diatom growth and Si : N uptake ratios in a coastal upwelling regime. *Nature*, 393(June), pp.561–564.
- Johnson, M. et al., 2007. Ammonium accumulation during a silicate-limited diatom bloom indicates the potential for ammonia emission events. *Marine Chemistry*,

106(1–2 SPEC. ISS.), pp.63–75.

Jørgensen, E.G., 1955. Variations in the Silica Content of Diatoms. *Physiologia*

*Plantarum*, 8(4), pp.840–845.

Keeling, P.J. et al., 2014. The Marine Microbial Eukaryote Transcriptome Sequencing

Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biology*, 12(6).

Khatiwala, S. et al., 2013. Global ocean storage of anthropogenic carbon.

*Biogeosciences*, 10(4), pp.2169–2191.

Kikutani, S. et al., 2016. Thylakoid luminal  $\theta$ -carbonic anhydrase critical for growth

and photosynthesis in the marine diatom *Phaeodactylum tricorutum*.

*Proceedings of the National Academy of Sciences*, 113(35), pp.9828–9833.

King, A.L. et al., 2015. Effects of CO<sub>2</sub> on growth rate, C:N:P, and fatty acid

composition of seven marine phytoplankton species. *Marine Ecology Progress Series*, 537, pp.59–69.

Koike, I. et al., 2001. Silicate to nitrate ratio of the upper sub-arctic Pacific and the

Bering Sea basin in summer: Its implication for phytoplankton dynamics. *Journal of Oceanography*, 57(3), pp.253–260.

Kröger, N. et al., 1997. Characterization of a 200-kDa Diatom Protein that is

Specifically Associated with a Silica-Based Substructure of the Cell Wall. *Eur J Biochem*, 250(1), pp.99–105.

Kröger, N. et al., 2000. Species-specific polyamines from diatoms control silica

morphology. *Proceedings of the National Academy of Sciences of the United States of America*, 97(26), pp.14133–14138.

- Kröger, N., Bergsdorf, C. & Sumper, M., 1994. A new calcium binding glycoprotein family constitutes a major diatom cell wall component. *The EMBO journal*, 13(19), pp.4676–4683.
- Kröger, N., Deutzmann, R. & Sumper, M., 1999. Polycationic Peptides from Diatom Biosilica That Direct Silica Nanosphere Formation. *Science*, 286(5442), pp.1129–1132.
- Kröger, N. & Wetherbee, R., 2000. Pleuralins are involved in theca differentiation in the diatom *Cylindrotheca fusiformis*. *Protist*, 151(3), pp.263–273.
- Kump, L.R., Bralowe, T.J. & Ridgwell, A., 2009. Ocean Acidification in Deep Time. *Oceanography*, 22(4), pp.94–107.
- De La Rocha, C.L. et al., 2000. Effects of iron and zinc deficiency on elemental composition and silica production by diatoms. *Marine Ecology Progress Series*, 195, pp.71–79.
- Lane, T.W. & Morel, F.M.M., 2000. Regulation of Carbonic Anhydrase Expression by Zinc, Cobalt, and Carbon Dioxide in the Marine Diatom *Thalassiosira weissflogii*. *Plant Physiology*, 123(1), pp.345–352.
- Leblanc, K. et al., 2005. Fe and Zn effects on the Si cycle and diatom community structure in two contrasting high and low-silicate HNLC areas. *Deep-Sea Research Part I: Oceanographic Research Papers*, 52(10), pp.1842–1864.
- Lewin, J.C., 1957. Silicon metabolism in diatoms IV. Growth and frustule formation in *Navicula pelliculosa*. *Canadian Journal of Microbiology*, 3, pp.427–433.
- Lewin, R.A., 1962. Silicification. In *Physiology and biochemistry of algae*. New York: Academic Press, pp. 445–455.

- Medlin, L.K., 2016. Evolution of the diatoms: major steps in their evolution and a review of the supporting molecular and morphological evidence. *Phycologia*, 55(1), pp.79–103.
- Meldrum, N.U. & Roughton, F.J., 1933. Carbonic anhydrase. Its preparation and properties. *The Journal of Physiology*, 80(2), pp.113–42.
- Mellett, T. et al., 2018. The biogeochemical cycling of iron, copper, nickel, cadmium, manganese, cobalt, lead, and scandium in a California Current experimental study. *Limnology and Oceanography*, 63, pp.S425–S447.
- Mock, T. et al., 2008. Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. *Proceedings of the National Academy of Sciences of the United States of America*, 105(5), pp.1579–84.
- Nakajima, K., Tanaka, A. & Matsuda, Y., 2013. SLC4 family transporters in a marine diatom directly pump bicarbonate from seawater. *Proceedings of the National Academy of Sciences of the United States of America*, 110(5), pp.1767–72.
- Nelson, D.M. et al., 1995. Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles*, 9(3), pp.359–372.
- Ohno, N. et al., 2012. CO<sub>2</sub>-cAMP-Responsive cis-Elements Targeted by a Transcription Factor with CREB/ATF-Like Basic Zipper Domain in the Marine Diatom *Phaeodactylum tricornerutum*. *Plant Physiology*, 158(1), pp.499–513.
- van de Poll, W.H., Vrieling, E.G. & Gieskes, W.W.C., 1999. Location and expression of frustulins in the pennate diatoms *Cylindrotheca fusiformis*, *Navicula*

- pelliculosa*, and *Navicula salinarum* (Bacillariophyceae). *J Phycol*, 35, pp.1044–1053.
- Poulsen, N. & Kröger, N., 2004. Silica morphogenesis by alternative processing of silaffins in the diatom *Thalassiosira pseudonana*. *Journal of Biological Chemistry*, 279(41), pp.42993–42999.
- Le Quéré, C. et al., 2016. Global Carbon Budget 2016. *Earth System Science Data*, 8(2), pp.605–649.
- Raven, J., 1991. Physiology of inorganic carbon acquisition and implications for resource use efficiency by marine phytoplankton: Relation to increased carbon dioxide and temperature. *Plant, Cell and Environment*, 14(8), pp.779–794.
- Reimann, B.E.F., Lewin, Joyce, C. & Volcani, B.E., 1966. Studies on the Biochemistry and Fine Structure of Silica Shell Formation in Diatoms: II. The Structure of the Cell Wall of *Navicula pelliculosa* (Breb.) Hilse. *Journal of Phycology*, 2(2), pp.74–84.
- Richthammer, P. et al., 2011. Biomineralization in Diatoms: The Role of Silacidins. *ChemBioChem*, 12(9), pp.1362–1366.
- Sabine, C.L. et al., 2004. The Oceanic Sink for Anthropogenic CO<sub>2</sub>. *Science*, 305(5682), pp.367–371. A
- Saito, M.A., Goepfert, T.J. & Ritt, J.T., 2008. Some thoughts on the concept of colimitation: Three definitions and the importance of bioavailability. *Limnology and Oceanography*, 53(1), pp.276–290.
- Samukawa, M. et al., 2014. Localization of putative carbonic anhydrases in the marine diatom, *Thalassiosira pseudonana*. *Photosynthesis Research*, 121(2–3), pp.235–

249.

- Sapriel, G. et al., 2009. Genome-wide transcriptome analyses of silicon metabolism in *Phaeodactylum tricornutum* reveal the multilevel regulation of silicic acid transporters. *PLoS ONE*, 4(10).
- Scheffel, A. et al., 2011. Nanopatterned protein microrings from a diatom that direct silica morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 108(8), pp.3175–80. Available at: <http://www.pnas.org/content/108/8/3175.short>.
- Shen, C., Dupont, C.L. & Hopkinson, B.M., 2017. The diversity of carbon dioxide-concentrating mechanisms in marine diatoms as inferred from their genetic content. *Journal of Experimental Botany*.
- Shrestha, R. et al., 2012. Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. *BMC Genomics*, 13(1), p.499.
- Sims, P.A., Mann, D.G. & Medlin, L.K., 2006. Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia*, 45(4), pp.361–402.
- Smith, S.R. et al., 2016. Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. *New Phytologist*, 210(3), pp.890–904.
- Solomon, S. et al., 2007. IPCC 2007 Climate Change 2007. In *The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.
- Stadie, W.C. & O'Brien, H., 1933. The Catalysis of the Hydration of Carbon Dioxide

- and Dehydration of Carbonic Acid by an Enzyme Isolated from Red Blood Cells. *The Journal of Biological Chemistry*, 103(2), pp.521–529.
- Sunda, W.G. & Huntsman, S.A., 1995. Iron uptake and growth limitation in oceanic and coastal phytoplankton. *Marine Chemistry*, 50, pp.189–206.
- Tachibana, M. et al., 2011. Localization of putative carbonic anhydrases in two marine diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. *Photosynthesis Research*, 109(1–3), pp.205–221.
- Tans, P., 2009. An Accounting of the Observed Increase in Oceanic and Atmospheric CO<sub>2</sub> and the Outlook for the Future. *Oceanography*, 22(4), pp.26–35.
- Thamatrakoln, K. & Hildebrand, M., 2007. Analysis of *Thalassiosira pseudonana* silicon transporters indicates distinct regulatory levels and transport activity through the cell cycle. *Eukaryotic Cell*, 6(2), pp.271–279.
- Thamatrakoln, K. & Hildebrand, M., 2008. Silicon Uptake in Diatoms Revisited: A Model for Saturable and Nonsaturable Uptake Kinetics and the Role of Silicon Transporters. *Plant Physiology*, 146(3), pp.1397–1407.
- Treguer, P. et al., 1995. The Silica Balance in the World Ocean - a Reestimate. *Science*, 268(5209), pp.375–379.
- Tripp, B.C. et al., 2004. A Role for Iron in an Ancient Carbonic Anhydrase. *Journal of Biological Chemistry*, 279(8), pp.6683–6687.
- Vaulot, D. et al., 1987. Cell-cycle response to nutrient starvation in two phytoplankton species, *Thalassiosira weissflogii* and *Hymemonas carterae*. *Marine Biology*, 95, pp.625–630.
- Wenzl, S. et al., 2008. Silacidins: Highly acidic phosphopeptides from diatom shells

- assist in silica precipitation in vitro. *Angewandte Chemie - International Edition*, 47(9), pp.1729–1732.
- Yool, A. & Tyrrell, T., 2003. The role of diatoms in regulating the ocean's silicate cycle. *Global Biogeochemical Cycles*, 17(4), pp.1–24.
- Young, J.N. et al., 2016. Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentrating mechanisms. *Journal of Experimental Botany*, 67(11), pp.3445–3456.
- Zeebe, R.E. & Wolf-Gladrow, D., 2001. *CO<sub>2</sub> in Seawater: Equilibrium, Kinetics, Isotopes* 1st ed. D. Halpern, ed., Amsterdam: Elsevier Oceanography Series.
- Zimmermann, J., Jahn, R. & Gemeinholzer, B., 2011. Barcoding diatoms: Evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity and Evolution*, 11(3), pp.173–192.

**Manuscript 1**

**Formatted for publication in Journal of Phycology**

**Transcriptome sequencing reveals contrasting silicon limitation responses in two**

***Thalassiosira* diatom species**

**Joselynn R. Wallace<sup>1</sup>, Tani Leigh<sup>1</sup>, P. Dreux Chappell<sup>2</sup>, Bethany D. Jenkins<sup>1,3</sup>**

1 Department of Cell and Molecular Biology, University of Rhode Island, Kingston, RI, USA

2 Ocean, Earth and Atmospheric Sciences, Old Dominion University, Norfolk, VA, USA

3 Graduate School of Oceanography, University of Rhode Island, Narragansett, RI, USA

## ABSTRACT

Diatoms perform 40% of primary production in the global ocean, which is constrained by the availability of growth-requiring nutrients. This includes silicon, which diatoms use to build cell walls or frustules. Silicon limitation of diatom growth has been observed in the North Atlantic Ocean and some high nutrient low chlorophyll (HNLC) regions, but these studies assessed the silicon status of the bulk diatom community and it is not known whether silicon limitation is uniform across particular diatom species. Genetic markers have been used to probe iron status of diatom species in the field, and similar markers of silicon status could be used to probe the silicon status of different co-existing diatom species. However, silicon limited transcriptome studies have been restricted mainly to the model diatom *Thalassiosira pseudonana* and have not included diatoms found in the North Atlantic Ocean or HNLC regions, hampering the ability to develop appropriate molecular markers. This study analyzed the silicon limited transcriptomes of two closely related *Thalassiosira* diatoms found in HNLC regions and the North Atlantic Ocean, toward the goal of developing appropriate molecular markers of silicon status. We found a gene family encoding putative ATP-grasp domain proteins that were upregulated in silicon-limited (but not iron or nitrate-limited) *T. oceanica* and *T. weissflogii*. Members of this gene family were also upregulated in previously published *T. pseudonana* transcriptomes and homologs were found across a diversity of diatoms, suggesting that they might be useful as a silicon limitation marker in many diatoms, in addition to *Thalassiosira* spp. diatoms.

## INTRODUCTION

Diatoms are photosynthetic single-celled eukaryotes that account for 40% of oceanic primary production and approximately 20% of global primary production (Nelson et al. 1995; Field et al. 1998). The rate and extent of diatoms' primary production is a function of their growth, which is constrained by the availability of nutrients in the ocean. Many phytoplankton share growth requirements for nutrients such as nitrate, phosphate, and iron, but diatoms are unique among the phytoplankton as they have an absolute growth requirement for silicon, which is used to build their ornately silicified cell walls or frustules (Lewin 1957; Lewin 1962; Jørgensen 1955).

It has been suggested that silicon ultimately limits diatom growth throughout the ocean (Dugdale et al. 1995; Yool & Tyrrell 2003) by several different mechanisms. Firstly, there is a discrepancy in the available pools of recycled nitrogen and recycled silicon in the surface ocean. Silicon recycling occurs deeper in the water column than that of other nutrients (C, N, P), leading to a propensity for silicon depletion in the surface ocean (Dugdale et al. 1995; Treguer et al. 1995; Dugdale & Wilkerson 1998). Secondly, diatoms growing in upwelled water can become silicon limited if these waters are enriched in nitrate relative to silicon. This has been observed in the North Atlantic Ocean (Allen et al. 2005; Johnson et al. 2007). Thirdly, silicon may also set the upper limit on diatom primary production in iron limited high nutrient low chlorophyll (HNLC) regions (Dugdale & Wilkerson 1998). Diatoms growing in low iron conditions ( $< 5$  nM), take up as much as 8 times more silicon than nitrate (Franck et al. 2000). Thus, diatoms in HNLC regions may be driven into silicon limitation due to their increased silicon requirement (Rueter & Morel 1981; Hutchins

& Bruland 1998; De La Rocha et al. 2000). It has also been suggested that silicon processing may require iron as a cofactor, which would lead to a biochemically-dependent iron and silicon co-limitation (Mock et al. 2008; Saito et al. 2008). Silicon and trace metal co-limitation of diatoms has been documented in the HNLC waters of the Equatorial Pacific Ocean (Brzezinski et al. 2011), the Southern Ocean and sub-Antarctic (Boyd et al. 1999; Leblanc et al. 2005; Franck et al. 2000), and the sub-Arctic Pacific (Koike et al. 2001). As these studies assessed the silicon status of the bulk diatom community, it is currently not known if different diatom species experience varying degrees of silicon stress.

Results from lab experiments show that silicon metabolism and physiology vary widely across diatom species. There is a wide range in diatoms' silicon quotas, which vary from less than 0.1 to greater than 200 pmol silicon per cell (Martin-Jézéquel et al. 2000). Diatoms can also vary greatly in their maximal silicon uptake rate, even amongst diatoms in the same genus. For example, the maximal uptake rate of *Thalassiosira weissflogii* (CCMP1336) is approximately 10 fold higher than for *Thalassiosira pseudonana* (CCMP1335) (Thamatrakoln & Hildebrand 2008). Closely related diatoms also vary in their silicon storage capacities, exemplified in the 4-fold difference in the ratio of silicon cell wall:silicon storage pool in two strains of *Navicula pelliculosa* (Thamatrakoln & Hildebrand 2008). These differences in silicon metabolism and physiology suggest that diatoms may have different strategies for coping with silicon limitation in the wild and highlight the need for molecular markers to probe the silicon status of whole diatom communities and also for more sensitive, species-specific markers.

Biochemical studies in diatoms have revealed several classes of proteins that are involved in silicon processing. The silicon transporters (SITs) are used for uptake when silicic acid levels are low (Thamatrakoln & Hildebrand 2008). Other molecules involved in silicon processing include the silaffins (Poulsen & Kröger 2004; Kröger et al. 1999), cingulins (Scheffel et al. 2011), silacidins (Wenzl et al. 2008), long-chain polyamines (Kröger et al. 2000), frustlins (Kröger et al. 1994; Kröger et al. 1996), and pleuralins (Kröger et al. 1997; Kröger & Wetherbee 2000). The silaffins, cingulins, and silacidins all contain an N-terminal ER signal peptide to facilitate their import into the ER and are subsequently localized to the silicon deposition vesicle, where they interact with long-chain polyamines to precipitate silicon for frustule formation (Kröger & Poulsen 2008; Scheffel et al. 2011; Wenzl et al. 2008; Poulsen et al. 2013; Drum & Pankratz 1964). While the pleuralin and frustulin molecules are tightly associated with the silicified cell wall (Kröger et al. 1997; Kröger et al. 1994), they are not thought to be directly involved in silicon deposition in the silicon deposition vesicle (Kröger & Wetherbee 2000; van de Poll et al. 1999).

Genetic studies have also given us a better understanding of the molecular underpinnings of diatom silicon physiology. The majority of these studies have been conducted in the model diatom *T. pseudonana*. Silicon limited *T. pseudonana* upregulate the expression of the genes and proteins for the silicon transporters SIT1 and SIT2 (Mock et al. 2008; Shrestha et al. 2012; Du et al. 2014; Smith et al. 2016; Brembu et al. 2017). Other studies show that they are also tightly regulated with the cell cycle (Thamatrakoln & Hildebrand 2007), so it may not be possible to distinguish aspects of cell cycle progression from changes in silicon status. In addition to SIT1

and SIT2, *T. pseudonana* also has a third silicon transporter, SIT3, that is expressed at low levels and insensitive to changes in silicon and the cell cycle (Thamatrakoln & Hildebrand 2007; Shrestha et al. 2012); leading to the hypothesis that it may function as a silicon sensor (Thamatrakoln & Hildebrand 2007). Silicon-limited *T. pseudonana* also increase their concentration of silacidin proteins relative to silicon replete condition (Richthammer et al. 2011). Silaffin genes are differentially regulated both in response to silicon supply and deficit, leading to the hypothesis that distinct silaffin gene copies are optimized for silicon deposition in either low or high silicon concentrations (Shrestha et al. 2012; Brembu et al. 2017). Cingulins are downregulated during silicon limitation and upregulated after silicon resupply, but they are also upregulated during the S phase of the cell cycle (when silicon deposition does not occur) suggesting that cingulins have roles in the cell aside from silicon deposition (Shrestha et al. 2012; Brembu et al. 2017). In *T. pseudonana*, silicidin genes are not differentially expressed. Genes for the biosynthesis of long-chain polyamine (LCPA) precursors (spermidine synthases) are both up and downregulated in silicon limitation, suggesting that distinct spermidine synthases may be optimized for LCPA synthesis in either low or high silicon concentrations or that LCPAs also play a larger role in diatom cells aside from silicon deposition (Bouchereau et al. 1999; Shrestha et al. 2012; Mock et al. 2008).

Other general trends regarding diatom silicon metabolism emerge from genetic studies in *T. pseudonana* and *P. tricornutum*. Firstly, these studies note that many of the silicon-sensitive transcripts are unique to diatoms or the particular diatom species tested (Mock et al. 2008; Sapriel et al. 2009; Shrestha et al. 2012). Genes involved in

intracellular signaling or post-transcriptional regulation (e.g. kinases) are both up (Smith et al. 2016; Brembu et al. 2017; Mock et al. 2008; Shrestha et al. 2012) and downregulated (Mock et al. 2008; Sapriel et al. 2009; Shrestha et al. 2012) in silicon limitation.

To date, the genetic response of diatoms to silicon limitation has not been followed in diatoms isolated from regions of the ocean where silicon sets an upper limit on diatom production, such as iron-limited, high-nutrient, low chlorophyll (HNLC) regions, (Dugdale & Wilkerson 1998; Chappell et al. 2016). A potential target for ecological markers of silicon status is the suite of genes encoding proteins involved in silicon processing described above. Homologs of some of these silicon processing proteins are bioinformatically difficult to predict in new species. For example, the silaffins (Poulsen & Kröger 2004) and silacidins (Pamirsky & Golokhvast 2013; Wenzl et al. 2008; Richthammer et al. 2011) don't show global sequence homology across taxa and are characterized by their overall amino acid content. Similarly, the pleuralins and frustulins are characterized by the presence of short repeat domains rather than their global sequence homology (Kröger et al. 1997; Kröger & Wetherbee 2000; Armbrust et al. 2004; Kröger & Poulsen 2008; Kröger et al. 1994; Kröger et al. 1996). The cingulins have only been identified in *T. pseudonana* and do not have homologs in *P. tricornutum* or *F. cylindrus*, suggesting that there is no cingulin sequence homology across species (or that cingulins are unique to *T. pseudonana*) (Scheffel et al. 2011). This makes finding and confirming orthologs using global sequence homology-based approaches a challenge. Amongst silicon metabolism proteins, the only ones that can be easily in novel genomes and

transcriptomes are the silicon transporters (SITs) and genes involved in the biosynthesis of the long-chain polyamines (LCPAs). SITs are distributed across all the major diatom lineages and can be identified by homology searches of protein similarity (Hildebrand et al. 1997; Hildebrand et al. 1998; Durkin et al. 2016).

Although their lengths and overall structures vary across species, the biosynthesis of LCPAs involves putrescine, spermine, and spermidine precursors, and the genes for the synthesis of these precursors can be identified by homology searches of protein similarity (Kröger et al. 2000; Bouchereau et al. 1999).

As global gene-based studies of silicon metabolism across diatom species have been restricted to a few species, the database of silicon-responsive genes across species is limited, particularly from diatoms found in silicon limited environments. In order to better understand the responses of diatoms found in ecosystems where their growth could be constrained by silica availability and to develop genetic markers that could be used in the environment to profile silicon sufficiency, we selected two related diatom species in the *Thalassiosira* genus that may be found in areas of the ocean where silicon may limit diatom growth and primary production. *Thalassiosira oceanica* CCMP1005 has a high tolerance for iron limitation and may persist in chronically iron limited HNLC regions where silicon requirements are high (Sunda & Huntsman 1995; Franck et al. 2000; Chappell et al. 2015). *Thalassiosira weissflogii* CCMP1010 was isolated from the North Atlantic Ocean, an ocean basin where silicon-depleted upwelling has been observed (Allen et al. 2005; Johnson et al. 2007). We selected this genus as it allows us to make comparisons to studies in *T. pseudonana* which has a completely sequenced genome (Armbrust et al. 2004).

Comparative quantitative transcriptomes of *Thalassiosira oceanica* CCMP1005 and *Thalassiosira weissflogii* CCMP1010 were analyzed for candidate gene markers of silicon limitation by profiling their global gene expression under silicon-limiting growth in comparison to profiles in other limiting nutrients, nitrate and iron. Annotations based on the KEGG database (Moriya et al. 2007) were used to determine whether or not broad metabolic pathways are silicon-sensitive in both *T. oceanica* and *T. weissflogii*. Silix homology-based networks (Miele et al. 2011) including other *Thalassiosiroid* transcriptomes were used to search for individual, homologous genes with silicon-specific regulation in *T. oceanica* and *T. weissflogii*. Homology-independent methods (FIMO, BioStrings, SignalP) were used to identify putative silaffins, cingulins, silacidins, and frustulins from *Thalassiosira oceanica* CCMP1005 and *Thalassiosira weissflogii* CCMP1010 and their potential utility as molecular markers of silicon stress is discussed (Grant et al. 2011; Pagès et al. 2016; Petersen et al. 2011).

## **MATERIALS AND METHODS**

### **Culturing, RNA extraction, and Sequencing**

Cultures of *Thalassiosira oceanica* CCMP1005 and *Thalassiosira weissflogii* CCMP1010 were obtained from the National Center for Marine Algae and Microbiota (NCMA; West Boothbay Harbor, ME, USA). Culturing and sequencing was completed in a range of conditions in order to get a thorough representation of each diatoms' transcript pool for the *de novo* assemblies. For nutrient replete, silicon limited, iron limited, and nitrate limited culturing, *T. weissflogii* and *T. oceanica* were grown in triplicate in a modified F/2 media (Guillard & Ryther 1962; Guillard 1975)

containing 400 nM iron or 4 nM iron for the iron limited condition. A final concentration of 10  $\mu$ M silicon was used for the silicon limited condition and no nitrate was added to the media for the nitrate limited condition. Macronutrients and vitamin solutions were treated with prepared Chelex-100 resin (Bio-Rad Laboratories Inc., Hercules, CA, USA) and syringe-filtration sterilized before being added to microwave sterilized Sargasso Sea water (Price et al. 1989). For inclusion in the *de novo* assembly, additional sequencing was performed using RNA from diatoms grown in F/2 medium at four CO<sub>2</sub> levels ranging from ~100 ppm to ~4000 ppm using autoclaved Woods Hole sea water amended with nutrients, vitamins, and trace metal mix that had not been Chelex-treated (Guillard & Ryther 1962; Guillard 1975; King et al. 2015). For the replete and silicon limited diatom culturing, biomass was collected after 3 days of growth, when the cells in the silicon limited condition began to grow more slowly relative to the silicon replete condition. Experiments were also conducted using the diatom *Thalassiosira rotula* (CCMP3096) to obtain a draft transcriptome for inclusion in Silix homology networks and comparison to *T. oceanica* and *T. weissflogii* (Miele et al. 2011). Duplicate cultures of *T. rotula* were grown in a range of three CO<sub>2</sub> conditions as well as iron-limited and nutrient replete conditions. Cells were collected on 3  $\mu$ m pore size polyester filters (Sterlitech Corporation, Kent, WA, USA) by gentle filtration, flash frozen in liquid nitrogen, and stored at -80 °C until RNA extraction.

RNA was extracted from filters following a modified version of the RNeasy Mini Kit protocol (Qiagen, Venlo, Netherlands). The protocol was modified by vortexing filters in lysis buffer with 0.5 mm zircon beads (BioSpec, Bartlesville, OK,

USA) for 15 minutes to lyse the cells. Cell debris was immediately removed by running lysate over a QiaShredder column (Qiagen). DNA contamination was removed by an on-column DNase step using the RNase-free DNase Set (Qiagen) and an additional DNase step was performed on eluted RNA using the Turbo DNA-free kit (Life Technologies, Carlsbad, CA, USA), both performed as per the manufacturers' instructions. Samples were quantified using the Qubit High-Sensitivity Kit (Life Technologies).

All culturing was performed in triplicate, except for the *T. rotula* CO<sub>2</sub> experiments, which were performed in duplicate. Equal amounts of total RNA from each like-condition replicate was pooled and sequenced. Nutrient replete, silicon limited, iron limited, and nitrate limited RNA from *T. weissflogii* and *T. oceanica* was sent to the National Center for Genome Resources as part of the Gordon and Betty Moore Marine Eukaryotic Transcriptome Sequencing Project (MMETSP) for sequencing on an Illumina HiSeq 2000 (30 million 50-nucleotide paired-end reads) (Keeling et al. 2014). RNA from *T. oceanica*, *T. weissflogii*, and *T. rotula* grown in the various CO<sub>2</sub> conditions were sequenced at the JP Sulzberger Genome Sequencing Center at Columbia University on an Illumina HiSeq 2500 (30 million 100-nucleotide paired-end reads). The *T. rotula* iron limited and nutrient replete sequencing was done at the Joint Genome Institute on an Illumina HiSeq 2500 (50 million 100-nucleotide paired-end reads). All libraries were mRNA enriched using a poly-A pull-down method.

## Trimming, Assembly, and Read Mapping

Initial read quality assessments were performed in FastQC (Andrews 2010). Raw reads were subsequently imported into CLC Workbench (CLC Bio-Qiagen, Aarhus, Denmark) with an insert size ranging from 1-400 nucleotides. Reads were trimmed in CLC Workbench with the following parameters: removal of 13 bases off the 5' end, .001 quality cut off, and no ambiguous nucleotide base calls.

Assemblies of each library were conducted in CLC Workbench, using a set of k-mers ranging from 1/2 read length to full length read after trimming. The assemblies were completed with no scaffolding, automatic bubble size selection, minimum contig length of 200, and with the auto detect paired distance options selected. To remove redundant contigs that were assembled more than once (e.g., present in multiple libraries or assemblies), the resulting assemblies were pooled using CD-HIT-EST using the following flags: `c .99 -G .99 -n 10 -d 1000 -aS .75 -g 1 -p 1 -T 0` (Li & Godzik 2006; Fu et al. 2012).

ORF prediction was performed on the nucleotide sequences as follows: nucleotide sequences were run in BlastX against a custom database containing the predicted proteins (filtered set of best models per locus) from available diatom genomes from JGI (*Thalassiosira pseudonana* Thaps3 and Thaps3\_bd, *Fragilariopsis cylindrus* v1.0, *Phaeodactylum tricornutum* Phatr2 and Phatr2\_bd, and *Pseudo-nitzschia multiseriis* v1) and transcriptomes available through MMETSP as of March 2013. The assemblies and BlastX output were uploaded to ORFpredictor to determine coding (CDS) and protein sequences (Min et al. 2005).

Additional clustering of protein sequences was performed to remove redundant amino acid sequences that were assembled multiple times but did not cluster at the nucleotide sequence level. Proteins were clustered in CD-HIT using the following flags: -c .75 -G 0 -n 5 -d 1000 -aS .75 -g 1 -p 1 -T 0. CDS sequences corresponding to the clustered protein sequences that were greater than 150 amino acids were used for subsequent read mapping. Trimming and assembly for *T. rotula* were completed as described for *T. oceanica* and *T. weissflogii*, although the iron limited and nutrient replete reads were included as guide-reads only, as there was contaminating ribosomal RNA sequence in these libraries. Sequences with less than 1 RPKM (*T. oceanica*) or 2 RPKM (*T. weissflogii* and *T. rotula*) across all conditions (i.e., consistently very lowly expressed) were removed from the assembly as potential contaminants (e.g. organelle transcripts).

Mapping for downstream expression analysis was done using the RNAseq module in CLC Workbench with the following parameters: mismatch cost (2), insertion cost (3), deletion cost (3), length fraction (0.8), similarity fraction (0.8), paired distances (auto-detect), strand specific (both), maximum number of hits for a read (10).

### **Annotation and Gene Ontologies**

The translated protein sequences from each diatom transcriptome assembly were uploaded to the KEGG automatic annotation server (Moriya et al. 2007). The BLAST search program was used to query the “Genes” representative data set, plus the two model diatoms available in the KEGG database (*T. pseudonana* and *P.*

*tricornutum*). An additional annotation step was performed in InterProScan (version 5.7-48.0 for *T. oceanica* and 5.13-52.0 for *T. weissflogii*) (Jones et al. 2014).

### **Ortholog Finding with Silix**

Silix networks were used to infer orthologous transcripts across diatom species (Miele et al. 2011). This is a blast-based homology network construction approach that groups similar sequences into gene families. Two sequences are included in the same family if the high-scoring segment pair length covers at least 80% of the query protein length and if their similarity is over 35% identity. The input for this analysis was the *T. rotula* draft transcriptome assembly, the *T. pseudonana* filtered protein models from JGI (Thaps3 and Thaps3\_bd), and the assembled *T. oceanica* and *T. weissflogii* transcriptomes.

### **Differential Expression Analysis and Hierarchical Clustering**

Differential gene expression was inferred with Analysis of Sequence Counts (ASC), using the total number of reads mapped to each transcript. Transcripts with at least a 95% posterior probability of a 2-fold change were considered differentially expressed (Wu et al. 2010).

To identify genes that respond specifically to silicon status rather than as part of a general stress response, hierarchical clustering was performed on the genes differentially expressed in silicon limitation. The gene expression values (as RPKM) in replete, silicon limited, nitrate limited, and iron limited conditions were the input to this analysis. Expression values were normalized by adding an offset value of 1 to all RPKM values and the log<sub>2</sub> fold change for silicon limited, nitrate limited, and iron limited conditions was calculated relative to the nutrient replete condition.

The complete Euclidian distance was calculated using the `dist` function and hierarchical clustering was performed using `hclust`, both part of the `stats` library included in R (R Development Core Team 2013). The resulting dendograms were cut at a height of 6 (*T. oceanica* upregulated) or 4 (*T. oceanica* downregulated, *T. weissflogii* up and downregulated). The heights used were selected based on the ability to distinguish clusters of genes regulated by silicon from those that may be part of a general stress response (i.e., clusters of genes upregulated or downregulated in all nutrient limitation scenarios examined).

Genes were grouped according to their cluster assignment and the per-cluster median and interquartile ranges of expression values were tabulated and plotted. Clusters of genes were considered part of a general stress response if the per-cluster interquartile range of expression values was above (or below, in the case of genes downregulated in silicon limitation) the  $\log_2$  fold change 0 threshold in iron or nitrate limitation (i.e., genes upregulated in silicon limitation that are also upregulated in iron and/or nitrate limitation were considered part of a general stress response; e.g., *T. oceanica* up in silicon limitation cluster #7) (Supplemental Figure 1, Supplemental Table 3). Gene clusters with non-overlapping interquartile ranges were also considered part of a silicon-specific expression response (i.e., clusters of genes that are responsive to iron and/or nitrate are still considered part of a silicon-specific response as long as their fold-change values in silicon limitation are considerably larger than the fold-change in iron and/or nitrate limitation; e.g., *T. weissflogii* up in silicon limitation cluster #3) (Supplemental Figure 1, Supplemental Table 3). Clusters are also considered part of a silicon-specific expression response even if their fold change is

large in iron or nitrate limitation, as long as the direction of regulation in these conditions is in the opposite direction than what is seen in silicon limitation (e.g., *T. oceanica* up in silicon limitation cluster #6) (Supplemental Figure 1, Supplemental Table 3).

### **Silicon Transporter Phylogeny**

Representative diatom silicon transporter (SIT) protein sequences corresponding to the 5 major phylogenetic SIT clades (A, B, C, D, and E) (Durkin et al. 2016) from *F. cylindrus* (Durkin et al. 2016) and the 3 known SITs from *T. pseudonana* (SIT1, SIT2, SIT3) (Thamatrakoln et al. 2006) were used as a BlastP query to identify putative silicon transporters from the *T. oceanica* and *T. weissflogii* transcriptomes and from the previously published *T. oceanica* genome (Lommer et al. 2012). Genes with at least 35% identity and 80% query coverage compared to the known SITs were considered putative silicon transporters. The alignment and tree also included partial SIT sequences from *T. weissflogii* CCMP1010 (Alverson 2007).

SIT amino acid sequences were aligned in MAFFT (version 7) using the L-INS-i strategy with the BLOSUM62 scoring matrix and the “leave gappy regions” option selected (Kato & Standley 2013). Alignments were trimmed in Jalview and trees were constructed in RAxML from a 325 amino acid fragment of the alignment. A maximum likelihood tree was computed with the following parameters: PROTGAMMA matrix, BLOSUM62 substitution model, and 1000 bootstrap replicates. The tree was mid-point rooted in FigTree.

## **Silaffins, cingulins, silacidins, frustulins, pleuralins, and polyamines**

Homologs of other silicon processing proteins are bioinformatically difficult to predict in new species. This is because the silaffins (Poulsen & Kröger 2004), silacidins (Pamirsky & Golokhvast 2013), and cingulins (Scheffel et al. 2011) do not show global sequence homology across taxa, while the pleuralins and frustulins are characterized by the presence of short repeat domains rather than their global sequence homology (Kröger et al. 1997; Kröger & Wetherbee 2000; Armbrust et al. 2004; Kröger & Poulsen 2008; Kröger et al. 1994; Kröger et al. 1996). Common domain or amino acid content features were used to search for potential homologs of silicon processing proteins in the translated transcriptomes of *T. oceanica* and *T. weissflogii*.

To identify silaffins and cingulins in *T. oceanica* and *T. weissflogii*, a FIMO search (Grant et al. 2011) was run for the pentalysine cluster KXXKXXKXX-KXXK (Poulsen et al. 2013) and proteins that did not contain an N-terminal ER signal peptide in SignalP were filtered out (Petersen et al. 2011; Poulsen et al. 2013). Using the predicted proteins from the *T. pseudonana* genome, this approach was sufficient to recover all of the sequences for silaffins and cingulins present in *T. pseudonana* (Poulsen & Kröger 2004; Scheffel et al. 2011).

The *T. pseudonana* silacidin sequence was used as a benchmark to search for silacidins in *T. oceanica* and *T. weissflogii*. This protein is acidic and is ~14% aspartic acid and 17% glutamic acid residues (Wenzl et al. 2008; Richthammer et al. 2011). Biostrings in R was used to find proteins in *T. oceanica* and *T. weissflogii* that pass this threshold of acidic amino acids per protein and the presence or absence of an N-

terminal ER signal peptide was determined by SignalP (Pagès et al. 2016; Petersen et al. 2011).

To identify putative frustulins in *T. oceanica* and *T. weissflogii*, a FIMO search (Grant et al. 2011) for the hexapeptide motifs CEGDCD and VPGCSG that are diagnostic of the ACR (acidic cysteine rich domains) of frustulins was performed (Kröger et al. 1994; Kröger et al. 1996). Transcripts without an N terminal ER signal peptide were filtered out using SignalP (Petersen et al. 2011; Kröger & Poulsen 2008).

Pleuralins have been identified and characterized in *Cylindrotheca* but are not present in *T. pseudonana* and are characterized by PSCD (proline, serine, cysteine, and aspartic acid-rich) repeat domains (Kröger et al. 1997; Kröger & Wetherbee 2000; Armbrust et al. 2004). The PSCD domain repeats of the pleuralins from *C. fusiformis* are 87 or 89 amino acids in length and are 22% proline, 11% serine, 11% cysteine, and 9% aspartic acid (Kröger et al. 1997). These percentages were used as a benchmark to scan our transcriptomes for pleuralins using BioStrings with a sliding window length of 87 amino acids (Pagès et al. 2016).

KEGG annotations were searched for genes involved in the synthesis of polyamines, including those involved in polyamine metabolism as well as arginine, ornithine, and glutamate metabolism that funnel into polyamine metabolism.

## **RESULTS AND DISCUSSION**

### **Transcriptome Assembly and Annotation**

In total, 36382 transcripts were assembled from *T. oceanica* and 24900 transcripts from *T. weissflogii* (Table 1). When compared to the previously sequenced *T. oceanica*, the number of *T. oceanica* transcripts assembled is less than the total

number of predicted genes (37921) but more than the number of genes (29306) in the non-redundant predicted set of genes (Lommer et al. 2012). InterProScan and KEGG KAAS assigned initial functional annotations or protein domain information to 12450 *T. oceanica* and 9580 *T. weissflogii* transcripts (Table 1, Supplemental Table 1). This corresponds to 34% and 38% of transcripts in the respective transcriptomes of *T. oceanica* and *T. weissflogii*. This is consistent with numbers from other studies that were able to assign functional annotations to 48% (17920 transcripts) or 40% (10591) of assembled *de novo* transcriptomes from microbial eukaryotes *Aureococcus anophagefferens* and *Euglena gracilis*, respectively (Frischkorn et al. 2014; Yoshida et al. 2016).

### **Differential Expression Analysis and Hierarchical Clustering**

Differential gene expression between the silicon limited and silicon replete condition was inferred with Analysis of Sequence Counts (Wu et al. 2010). Genes with a 95% posterior probability of at least a 2-fold change in silicon limitation relative to the nutrient replete condition were considered differentially expressed (Wu et al. 2010).

To distinguish a silicon-specific stress response from a more general nutrient stress response, the subset of transcripts differentially expressed in silicon limitation were hierarchically clustered. The clustering input was the per-transcript log<sub>2</sub>-fold change expression values in nitrate limited, iron limited, and silicon limited conditions relative to the nutrient replete condition. Differentially expressed genes in *T. oceanica* were binned into 18 clusters – 7 upregulated (containing 947 genes) and 11 downregulated (containing 809 genes) (Supplemental Figure 1, Supplemental

Table 3). Of these *T. oceanica* clusters, 4 (containing 823 genes) were part of a silicon-specific upregulation and 4 (containing 539 genes) were part of a silicon-specific downregulation (Figure 1). The differentially expressed genes in *T. weissflogii* were binned into 47 clusters – 23 upregulated (containing 2814 genes) and 24 downregulated (containing 1467 genes) (Supplemental Figure 1, Supplemental Table 3). Of these *T. weissflogii* clusters, 6 (containing 557 genes) were part of a silicon-specific upregulation and 3 (containing 801 genes) were part of a silicon-specific downregulation (Figure 1).

Previous studies have found a significant overlap in the subset of genes upregulated in iron and silicon limited *T. pseudonana* transcriptomes (Mock et al. 2008). There are 3 clusters of genes that are upregulated in iron and silicon limited *T. weissflogii* (Supplemental Figure 1), but there are not any clusters in *T. oceanica* that are upregulated in both iron and silicon limited conditions (Supplemental Figure 1). This discrepancy might be explained by differences in iron physiology. *T. oceanica* can tolerate lower iron concentrations than *T. pseudonana* (Sunda & Huntsman 1995), but *T. weissflogii* and *T. pseudonana* show similar thresholds for their iron requirements (Whitney et al. 2011). Perhaps the overlap in iron and silicon sensitive genes is a feature of diatoms with higher iron requirements rather than diatoms that are tolerant of low iron.

## **Identification of genes sensitive to silicon status: silicon transporters and putative silicon efflux proteins**

Sequence similarity to known SITs from *Thalassiosira pseudonana* and *Fragilariopsis cylindrus* was used to identify SIT sequences from *T. oceanica*, *T. weissflogii*, and *T. rotula*. All of the SITs from these four *Thalassiosira* diatoms (*T. pseudonana*, *T. oceanica*, *T. weissflogii*, and *T. rotula*) are in clade E (Lommer et al. 2012; Durkin et al. 2016). There is no evidence of any clade B SITs, unlike other *Thalassiosira* diatoms (Durkin et al. 2016). The numbers of SITs present in each diatom varies (Thamatrakoln et al. 2006; Durkin et al. 2016). Our *T. oceanica* data confirms results from previous studies indicating that CCMP1005 has a single SIT gene copy (Lommer et al. 2012; Durkin et al. 2016) (Figure 2). The single SIT in *T. oceanica* shares a branch point with *T. pseudonana* SIT1 and SIT2 and it is also grouped together with the single full length silicon transporter recovered from the *T. oceanica* genome (support value of 100) (Figure 2) (Lommer et al. 2012). We recovered 4 SIT sequences from *T. weissflogii* CCMP1010 (Figure 2), a greater number than the 2 obtained in a previous study (Alverson 2007) that relied on amplification methods using degenerate primers to amplify SIT gene fragments. We also recovered only a single SIT from *T. rotula* CCMP3096. It is possible that this diatom has additional SIT copies not captured in this assembly, as there was no silicon limited condition included in the *T. rotula* transcriptome to potentially induce their expression (Figure 2).

Sequence alignment and maximum likelihood tree construction indicates that three of the SITs in *T. weissflogii* appear to be paralogous sequences derived from a

single ancestral SIT copy that forms a clade with *T. pseudonana* SIT1 and SIT2 (Figure 2). Two of these sequences group with previously identified *T. weissflogii* SIT fragments (Alverson 2007), as illustrated by their grouping together on the tree with bootstrap support values of 100 (Figure 2). The third *T. weissflogii* SIT recovered from the transcriptome also clusters with one of the previously identified SIT fragments (Alverson 2007), although with a lower support value of 89 (Figure 2). The fourth *T. weissflogii* SIT forms a clade with *T. pseudonana* SIT3 (Figure 2). This is in contrast with previous studies indicating that *T. weissflogii* CCMP1010 didn't have a SIT3-like gene copy (Alverson 2007). Perhaps the degenerate primers used in the Alverson 2007 study could not distinguish between or amplify some of the SITs from *T. weissflogii*.

One of the *T. weissflogii* SITs (*T. weissflogii* SIT3, K32\_12335) is one of the most highly silicon responsive genes in the dataset and is ~15 fold upregulated (log<sub>2</sub>-fold change of ~4) in silicon limitation versus nutrient replete (Figure 2). The upregulation this *T. weissflogii* SIT is consistent with previous studies where some SITs are upregulated in *T. pseudonana* and *P. tricornutum* grown in low silicon (Shrestha et al. 2012; Sapriel et al. 2009). The low expression of the SIT3-like sequence from *T. weissflogii* is consistent with what was observed for the SIT3 from *T. pseudonana*, which is constitutively lowly expressed, leading to the suggestion that it functions as a silicon sensor (Thamatrakoln & Hildebrand 2007).

Surprisingly, the single SIT in *T. oceanica* was not significantly differentially expressed, but only slightly upregulated in low silicon (Figure 2). Although a similar silicon-indifferent trend has been seen in *T. pseudonana* SIT3, the single *T. oceanica* SIT groups with the silicon sensitive SIT1 and SIT2 sequences (Figure 2) (Shrestha et

al. 2012). The lack of a strong silicon limitation response in the *T. oceanica* SIT might be due to the fact that it lacks a SIT3-like gene copy that has been hypothesized to act as a silicon sensor in *T. pseudonana* (Thamatrakoln & Hildebrand 2007). It is also possible that the single *T. oceanica* SIT both senses silicon status as well as functioning as a transporter.

In addition to SIT mediated silicon transport, it has been hypothesized that *T. pseudonana* also contains a silicon efflux transporter. This is because silicon-limited *T. pseudonana* upregulated a gene annotated as a citrate transporter that shares homology with a plant silicon efflux transporter (Shrestha et al. 2012). The Silix gene family (FAM001283) containing the putative silicon efflux protein from *T. pseudonana* also contains a single gene from each *T. oceanica* and *T. weissflogii* (Supplemental Table 4), but they are not differentially expressed in silicon limitation (Supplemental Table 2), suggesting that this gene is not sensitive to silicon status in *T. oceanica* or *T. weissflogii*.

#### **A subset of putative silaffins, cingulins, silacidins, frustulins, and pleuralins are differentially regulated in Thalassiosiroids**

Identification of proteins (silaffins, cingulins, silacidins, frustulins, and pleuralins) involved in silicon processing using homology-based approaches is challenging because many of these genes do not show global sequence homology across species and are characterized by the presence of conserved domains or amino acid content (Poulsen & Kröger 2004; Pamirsky & Golokhvast 2013; Kröger et al. 1997; Kröger & Wetherbee 2000; Kröger & Poulsen 2008; Kröger et al. 1994; Kröger et al. 1996). We queried the *T. oceanica* and *T. weissflogii* transcriptomes for

candidate genes involved with silicon processing using a variety of homology-independent methods.

The search for silaffins and cingulins involved a FIMO search (Grant et al. 2011) for the pentalysine cluster KXXKXXKXX-KXXK (Poulsen et al. 2013). Proteins lacking an N-terminal ER signal peptide in SignalP were filtered out (Petersen et al. 2011; Poulsen et al. 2013), as these proteins would not be localized to the ER and silicon deposition vesicle. This approach yielded 138 proteins in *T. oceanica* and 243 proteins in *T. weissflogii* (Supplemental Table 7). A similar search of the *T. pseudonana* predicted proteins from the genome sequence (filtered set of best models per locus) yielded 162 proteins, including the 6 known cingulins and 3 known silaffins (Poulsen & Kröger 2004; Scheffel et al. 2011). The majority of the other putative silaffin-like and cingulin-like genes we identified are not differentially expressed in response to silicon (Figure 3, Supplemental Table 7). In *T. oceanica*, 10 of the putative silaffins and cingulins we identified are upregulated and 6 are downregulated in response to low silicon (Figure 3, Supplemental Table 7). In *T. weissflogii*, 2 are upregulated and 37 are downregulated in response to low silicon (Figure 3, Supplemental Table 7). This pattern of expression whereby some silaffins and cingulins are upregulated and some are downregulated in response to silicon status is consistent with previous studies that saw both the up and downregulation of different silaffin and cingulin homologs in silicon limited *T. pseudonana* (Mock et al. 2008; Shrestha et al. 2012; Brembu et al. 2017). Our data is also consistent with the hypotheses that distinct silaffin gene copies are optimized for silicon deposition in either low or high silicon and that cingulins might play a wider role in diatoms other

than silicon deposition (Shrestha et al. 2012; Brembu et al. 2017). Perhaps this dynamic regulation of silaffins and cingulins is part of the underlying molecular mechanism that allows diatoms to regulate the thickness of their frustules in accordance with silicon availability (Paasche 1973).

Putative *T. oceanica* and *T. weissflogii* silicadins were identified using the *T. pseudonana* silacidin sequence as a benchmark as this acidic protein is comprised of ~14% aspartic acid and 17% glutamic acid residues (Wenzl et al. 2008; Richthammer et al. 2011). Biostrings in R was used to find proteins in *T. oceanica* and *T. weissflogii* that pass this threshold of acidic amino acids per protein and the presence or absence of an N-terminal ER signal peptide was determined by SignalP (Pagès et al. 2016; Petersen et al. 2011). In *T. oceanica* and *T. weissflogii*, 17 and 28 proteins met the silacidin amino acid composition threshold cutoffs (Figure 3, Supplemental Table 8). None of the copies in *T. oceanica* were differentially expressed in silicon limitation, nor do they have an N terminal ER signal peptide (Figure 3, Supplemental Table 8). In *T. weissflogii*, 2 silacidin-like proteins are upregulated in silicon limitation while one is downregulated and regulation of all three is specific to the silicon limited condition, although none of the 3 have an N terminal ER signal peptide (Figure 3, Supplemental Table 8). While neither transcriptome study in *T. pseudonana* saw a differential expression of the silacidin gene, (Mock et al. 2008; Shrestha et al. 2012), silicon limited *T. pseudonana* accumulated higher amounts of silacidin protein in low silicon (Richthammer et al. 2011). This suggests that the transcription of silacidin genes is not a key point of regulation for silicon processing in *T. pseudonana* or *T. oceanica* but may be post-transcriptionally or post-translationally regulated.

To identify putative frustulins in *T. oceanica* and *T. weissflogii*, a FIMO search (Grant et al. 2011) for the hexapeptide motifs CEGDCD and VPGCSG that are diagnostic of the acidic cysteine rich domains (ACR) of frustulins (Kröger et al. 1994; Kröger et al. 1996) was performed and yielded 2480 putative proteins from *T. oceanica* and 1739 proteins from *T. weissflogii*. Transcripts without a predicted N terminal ER signal peptide were filtered out using SignalP (Petersen et al. 2011), narrowing this number down to 313 proteins from *T. oceanica* and 215 from *T. weissflogii* (Figure 3, Supplemental Table 9). Most of these putative frustulins are not differentially expressed in silicon limitation (Figure 3, Supplemental Table 9). In *T. oceanica*, 25 putative frustulins are upregulated in silicon limitation and 22 of them specifically regulated in the silicon limited condition (Figure 3, Supplemental Table 9). Similarly, *T. oceanica* downregulates 18 putative frustulins in silicon limitation, 13 of which are specific to the silicon limited condition specifically downregulated (Figure 3, Supplemental Table 9). In *T. weissflogii*, 24 of the putative frustulins are upregulated in silicon limitation, 3 of them specific to the silicon limited condition. Of the putative frustulins, 33 are downregulated in silicon limitation and 18 of them with silicon specificity (Figure 3, Supplemental Table 9). The lack of silicon-sensitivity of most of the putative frustulins is not surprising – in *T. pseudonana*, only a single frustulin was upregulated 7 hours after silicon resupply (Shrestha et al. 2012). Further, it is likely that many of these sequences are not truly frustulins, as the number of putative frustulins recovered from *T. oceanica* and *T. weissflogii* is significantly higher than the 3 frustulin sequences from the *T. pseudonana* genome (Armbrust et al. 2004).

Pleuralins have been identified and characterized in *Cylindrotheca* and are characterized by PSCD (proline, serine, cysteine, and aspartic acid-rich) repeat domains (Kröger et al. 1997; Kröger & Wetherbee 2000). The PSCD domain repeats are 87 or 89 amino acids in length and are 22% proline, 11% serine, 11% cysteine, and 9% aspartic acid (Kröger et al. 1997). These percentages were used as a benchmark to scan the translated transcriptomes' protein sequences for pleuralins using BioStrings with a sliding window length of 87 amino acids (Pagès et al. 2016). We did not recover any putative pleuralin sequences, consistent with previous studies that did not find this class of molecule present in *T. pseudonana* (Armbrust et al. 2004).

**Genes involved in long-chain polyamine synthesis are not differentially regulated in *T. oceanica* or *T. weissflogii***

Long-chain polyamines (LCPAs) are additional molecules that play a role in silicon deposition in diatoms (Kröger et al. 2000). Proline, glutamate, and arginine fuel polyamine biosynthesis via their conversion to ornithine, which is converted into putrescine, spermine, and spermidine – the polyamine precursor molecules (Bouchereau et al. 1999). The enzymes spermine and spermidine synthase catalyze the conversion of putrescine to spermine and spermidine using S-adenosyl-L-methionine and S-adenosyl-metioninamine as a substrate, which is derived from methionine metabolism (Bouchereau et al. 1999).

Genes involved in funneling proline, arginine, glutamate, ornithine, and methionine toward polyamine synthesis were of interest as potential markers of silicon stress in diatoms and their presence in the transcriptomes was determined by their KEGG annotations. However, we found little evidence of silicon-specific changes in

polyamine metabolism, with the exception of *T. oceanica* upregulating a proline iminopeptidase (K01259) that cleaves proline residues from peptides and *T. weissflogii* upregulating an arginase (K01476), which converts arginine to ornithine (Figure 4). This is in contrast with previous studies in *T. pseudonana*, where spermidine and spermine synthases were up and downregulated during silicon limitation (Shrestha et al. 2012; Mock et al. 2008). This suggests that while the synthesis of LCPA precursors may be a key point of regulation in silicon processing for *T. pseudonana*, that may not be the case in *T. oceanica* or *T. weissflogii*.

### **Identifying novel markers of silicon stress**

Molecular markers of silicon stress can include novel genes or genes lacking an annotated function in a well described metabolic pathway. To determine whether or not any of the unannotated genes may be good markers of silicon stress in diatoms, we used a network-based approach to identify homologous genes that were upregulated specifically in response to silicon in both *T. oceanica* and *T. weissflogii*. Silix networks were computed using the *T. rotula* draft transcriptome assembly, the *T. pseudonana* filtered protein models from JGI (Thaps3 and Thaps3\_bd), and the assembled *T. oceanica* and *T. weissflogii* transcriptomes. (Miele et al. 2011; Armbrust et al. 2004). Homologues were identified if they were in the same Silix gene family (e.g., were 35% identical over at least 80% query coverage per HSP). As an example, the silicon transporters cluster together in a single Silix family (FAM000927, Supplemental Table 4).

The Silix gene network consists of 104,848 genes grouped into 77,386 gene families (Supplemental Table 4, Supplemental Table 5). In the Silix networks, 132,857 genes (grouped into 9076 gene families) are present in 2 or more diatom species. However, the majority of genes (71,991) are grouped into 68,310 species-specific gene families that contain representatives from 1 diatom species only (Supplemental Table 4, Supplemental Table 5). This pattern (many genes residing in species-specific gene families) also holds true for the silicon-specific differentially expressed genes. In *T. oceanica*, 621 of the up and 270 of downregulated genes are in species-specific gene families (Supplemental Table 6). In *T. weissflogii*, 381 of the up and 613 of downregulated genes are also in species-specific gene families (Supplemental Table 6). To target potential markers of silicon stress, networks were subsequently filtered to include only Silix families with at least one silicon-specifically upregulated gene from both *T. oceanica* and *T. weissflogii*.

The search for silicon specific, upregulated homologous genes in *T. oceanica* and *T. weissflogii* revealed 4 gene families containing candidate silicon stress markers. The most promising of these is a family of ATP-grasp domain proteins (FAM001061). Silix gene families were also used to recover *T. oceanica* and *T. weissflogii* homologs to the *T. pseudonana* putative silicon efflux protein (Shrestha et al. 2012) and also highlighted the presence of an interesting gene expansion in *T. oceanica*.

#### **ATP-grasp proteins are silicon sensitive and silicon specific**

The most promising marker of silicon stress from this study are members of a gene family of ATP-grasp domain proteins, which contains 3 silicon-specific upregulated genes – 2 from *T. weissflogii* and 1 from *T. oceanica* (EJK54127 from the

genome Lommer et al. 2012). This Silix family also contains 2 genes that are upregulated in silicon limited *T. pseudonana* (Brembu et al. 2017; Smith et al. 2016). Another study saw that *T. pseudonana* slightly upregulated 1 of these 2 genes ~7 hours after silicon resupply, but its expression values return to baseline levels at the 9 hour mark (Shrestha et al. 2012).

To determine whether or not these ATP-grasp proteins are found in other diatoms, the MMETSP database was queried for sequences similar (35% identity and 80% query coverage per HSP) to the 3 silicon-sensitive ATP grasp proteins in the *T. oceanica* and *T. weissflogii* transcriptome assemblies. Sequences meeting our similarity threshold were found in 143 genes across 29 diatom species across 17 diatom genera. The only non-diatom sequences recovered were 2 sequences from the dinoflagellates *Pyrodinium bahamense* and *Alexandrium monilatum*.

Proteins containing ATP-grasp domains are ATP-dependent enzymes that ligate carboxylate groups to amino or imino groups (Galperin & Koonin 1997). They are found throughout the tree of life and ~40% of the described ATP-grasp domain proteins are involved in purine biosynthesis or are biotin carboxylases (e.g., Acetyl-CoA carboxylase, pyruvate carboxylase, carbamoyl phosphate synthetase) (Fawaz et al. 2011). Other examples of ATP-grasp domain proteins include D-Alanine-D-Alanine ligases (involved in peptidoglycan biosynthesis), glutathione synthetases (involved in glutathione metabolism), tubulin-tyrosine ligases (involved in microtubule assembly), and they also have been indicated to play a role in the biosynthesis of mycosporine-like amino acids in cyanobacteria, where they play a role in UV protection (Fawaz et al. 2011; Gao & Garcia-Pichel 2011). Mycosporine-like

amino acids are tightly embedded into diatom frustules and are hypothesized to play a role in mediating UV stress (Ingalls et al. 2010).

The upregulation of ATP-grasp domains in low silicon may be related to the thinning of diatom frustules in low silicon, and need for enhanced UV protection. In addition to fortification of their non-silicified cell wall components to compensate for thinner frustules, it is also possible that the ATP-grasp domain proteins are functioning similarly to the tubulin-tyrosine ligases involved in microtubule assembly and the cytoskeleton (Fawaz et al. 2011). This would be consistent with the role of the cytoskeleton-mediated positioning of the silicon deposition vesicle (Tesson & Hildebrand 2010). While the specific enzymatic activity of the family of ATP-grasp domain proteins discussed in this study is unclear, they are an attractive choice as a molecular marker of silicon stress in diatoms. Although transcriptome evidence suggests they are wide-spread throughout the diatoms, their silicon sensitivity in individual diatom species should be verified.

### **Identifying metabolic pathways with silicon sensitivity**

In addition to the previous gene family based approach, an additional strategy for finding potential molecular markers of silicon stress is to look for common, silicon-sensitive pathways in both *T. oceanica* and *T. weissflogii*. To find silicon-sensitive metabolic pathways in both *T. oceanica* and *T. weissflogii*, genes upregulated in silicon stress and not part of a general stress response were binned into broad KEGG modules or silicon metabolism gene categories (Figure 5). There are several pathways that are upregulated in silicon limited *T. oceanica* and *T. weissflogii*, including “Cellular processes and signaling”, “Membrane transport”, and

“Xenobiotics biodegradation and metabolism” (Figure 5). Although similar pathways are silicon-sensitive in both diatoms, the direction of expression (e.g. up or downregulated in silicon limitation) is often different in each diatom (Figure 5). Silicon limited *T. oceanica* upregulates the “silaffin-cingulins”, “Cell growth and death”, “Replication and repair”, “Transcription” and “Glycan biosynthesis and metabolism”, these same categories are downregulated in silicon limited *T. weissflogii* (Figure 5). Silicon limited *T. weissflogii* upregulates “Genetic information processing”, “Biosynthesis of secondary metabolites”, “Metabolism of terpenoids and polyketides”, and “Energy metabolism”, these same categories are downregulated in silicon limited *T. oceanica* (Figure 5).

Although there are not specific gene homologs in the “Cellular processes and signaling”, “Membrane transport”, and “Xenobiotics biodegradation and metabolism” categories (e.g., genes annotated with the same KEGG orthology assignment) that are differentially expressed in both diatoms, these categories contain several phosphatases, peptidases, proteases, and kinases (Supplemental Table 1, Supplemental Table 2). Perhaps during silicon limitation, some degree of post-translational regulation of protein activity is occurring via degradation by peptidases/proteases or phosphorylation/de-phosphorylation by kinases/phosphatases. Genes involved in modifying proteins could be related to the silacidin and silaffin proteins being phosphorylated (Wenzl et al. 2008; Kroger et al. 2002). It could also be related to the fact that protein degradation is an important mediator of transitioning through distinct phases of the cell cycle (Glotzer et al. 1991), a process that is tightly associated with

silicon uptake and processing (Brzezinski et al. 1990; Vaultot et al. 1987; Thamatrakoln & Hildebrand 2007).

The contradictory trends in silicon-sensitive gene expression between *T. oceanica* and *T. weissflogii* may be related to previous studies indicating that silicon processing is tied to DNA replication, cell division, and the cell cycle (Eppley et al. 1967; Darley & Volcani 1969; Vaultot et al. 1987; Brzezinski et al. 1990; Thamatrakoln & Hildebrand 2007; Hildebrand et al. 2007). Like other eukaryotic cells, diatom cells mitotically divide during M phase, DNA synthesis and replication occurs during S phase, and these two phases of the cell cycle are separated by the G1 and G2 gap phases (Mitchison 1971). Diatom frustule valves are formed just after cell division, during G2/M phase, while girdle band synthesis happens in G1 phase (Eppley et al. 1967; Hildebrand et al. 2007). In addition to cell wall synthesis, diatoms also require silicon for DNA replication to proceed during the S phase of the cell cycle (Darley & Volcani 1969). These silicon requirements explain why silicon limited *T. weissflogii* cells sometimes arrest in the G2/M phase or other times in the G1/S phase transition (Brzezinski et al. 1990; Vaultot et al. 1987).

As we were largely focused on identifying markers that would be applicable in asynchronous field populations, we focused on the onset of silicon limitation and made no attempt to synchronize cell cycle shifts in any of the cultures during these experiments, so we cannot speak definitively to specific aspects of the cell cycle that are affected by silicon limitation and we do not know if either diatom was silicon limited to the point of cell cycle arrest. However, the cyclin B1/B2 genes of *P. tricornutum* were upregulated in the G2/M phase of the cell cycle (Huysman et al.

2010). The Silix gene family containing the *T. pseudonana* B1/B2 cyclins (FAM003363, Supplemental Table 4) contains a silicon-specifically upregulated gene in *T. oceanica* and a silicon-specifically downregulated gene in *T. weissflogii*, suggesting that at least some of the silicon limited *T. oceanica* were in G2/M phase while some of the silicon limited *T. weissflogii* cells were likely not in G2/M phase at the time the transcriptomes were sequenced (Supplemental Table 3) (Huysman et al. 2010).

The different biological processes that happen during distinct phases of the cell cycle might explain the contradictory trends seen in silicon limited *T. oceanica* and *T. weissflogii*. Silicon limited *T. oceanica* upregulates genes in “Replication and repair”, suggesting that it may be awaiting sufficient silicon resupply for DNA synthesis to proceed (Figure 5). Silicon limited *T. weissflogii* upregulates genes in “Translation”, which may be indicative of the biosynthesis of proteins required for silicon processing (e.g., SITs) (Figure 5).

### **Sel1/MYND zinc finger domain gene expansion in *T. oceanica***

One gene family (FAM006551) contained a highly responsive gene (~8 fold upregulation in silicon stress) and 7 other silicon-sensitive genes in *T. oceanica* (27088) (Supplemental Table 3, Supplemental Table 4) but did not have homologs in our *T. weissflogii* transcriptome assembly. This gene family is highly expanded in *T. oceanica* and consists of 683 genes, 463 of them contain a Sel1 like repeat (including the highly silicon sensitive gene in *T. oceanica*) 44 have a MYND zinc finger domain (Supplemental Table 1). A search for Sel1-like repeats in the *T. oceanica* genome (through protists.ensembl.org) reveals that there are 1051 Sel1-like repeats in the *T.*

*oceanica* genome, while a similar query of *Thalassiosira pseudonana*, *Phaeodactylum tricornutum*, and *Fragilariopsis cylindrus* genomes (through genome.jgi.doe.gov) found 8, 5, and 17 genes, respectively, confirming the expansion of Sell1-like repeat proteins in *T. oceanica*. Similarly, searches for the MYND zinc finger proteins (InterPro accession IPR002893) in the genomes of *T. oceanica*, *T. pseudonana*, *P. tricornutum*, and *F. cylindrus* revealed 443, 6, 8, and 134 genes, respectively, indicating that the MYND zinc finger domain proteins are expanded in both *T. oceanica* and *F. cylindrus*. Sell1 domain proteins are extracellular proteins involved in regulating Notch signaling in *C. elegans*, and MYND zinc finger domains are thought to be involved in mediating protein-protein interactions, rather than the canonical DNA binding associated with zinc finger domains (Grant & Greenwald 1996; Gamsjaeger et al. 2006). This is consistent with previous studies indicating that Notch signaling proteins are upregulated during formation of diatom frustule valves (Shrestha et al. 2012).

A search for homologs to the differentially regulated Sell1/MYND zinc finger domain containing genes in the MMETSP database revealed that homologs are widespread across 168 species across the alveolate, amoebzoa, archplastida, hacrobia, rhizaria, and stramenopile lineages. In most species (including the diatoms), these genes are not highly expanded (e.g., less than 20 copies) (Supplemental Figure 5). However, in addition to being highly expanded in *T. oceanica*, homologs to this gene family are also highly expanded in *Cyclotella* and *Skeletonema* diatoms (Supplemental Figure 5).

The expansion of MYND zinc finger domain proteins in *F. cylindrus* has been previously reported and was tentatively attributed to the high zinc concentrations in the Southern Ocean where this diatom is found (Mock et al. 2017). *F. cylindrus* also experiences chronically low iron in the Southern Ocean, as does *T. oceanica* in the northeast Pacific (Chappell et al. 2015; Bertrand et al. 2015). Perhaps these expanded gene families in both *F. cylindrus* and *T. oceanica* are part of lineage-specific metal sensing mechanisms. The upregulation of this gene in silicon limited *T. oceanica* would be consistent with the hypothesis that iron is a co-factor for diatoms' silicon processing (Mock et al. 2008).

In this context, the high copy number of MYND/Sel1-like genes in *Skeletonema* diatoms is particularly interesting, as these diatoms are typically found in coastal environments where iron is not thought to limit diatom growth (Kooistra et al. 2008). One possible explanation is that Sel1-like repeats have also been suggested to play a role in mediating protein degradation in the endoplasmic reticulum and perhaps this is a key point of metabolic regulation in both *T. oceanica* and *Skeletonema* diatoms (Mittl & Schneider-Brachert 2007).

## **CONCLUSIONS**

This aim of this study was to find molecular markers of silicon stress in two ecologically relevant diatoms, *Thalassiosira oceanica* and *Thalassiosira weissflogii*. Ideal silicon stress gene markers should be upregulated in silicon limitation and not other nutrient stressors and have homologs found across a range of diatoms. This study revealed a family of ATP-grasp domain proteins that were upregulated in both *T. oceanica* and *T. weissflogii* uniquely in the silicon limited condition. Previous

studies in *T. pseudonana* also saw an upregulation of this gene family in silicon limitation (Smith et al. 2016; Brembu et al. 2017). Homologs of this gene were widespread through other diatoms, but their silicon sensitivity across these different diatoms has yet to be tested.

Unlike *T. pseudonana* (Shrestha et al. 2012) and *T. weissflogii* (this study), *T. oceanica* does not strongly upregulate its silicon transporter in silicon limitation. This may be related to the absence of a SIT3-like sequence in this diatom, as SIT3 has been suggested to play a role as a silicon sensor in *T. pseudonana* (Thamatrakoln & Hildebrand 2007). The hypothesis that SIT3 functions as a silicon sensor was born from the observation that *T. pseudonana* SIT3 is lowly expressed and is not upregulated in silicon limitation (Thamatrakoln & Hildebrand 2007). This study has produced incongruous results in its support for this hypothesis. The low expression of the SIT3-like sequence from *T. weissflogii* is consistent with the suggestion that it might act as a silicon sensor, but its expression is sensitive to silicon (Figure 2), which is contrary to what has been shown for *T. pseudonana* SIT3 (Thamatrakoln & Hildebrand 2007; Shrestha et al. 2012). Future discovery of the SIT3 promoter sequence would aid in further attempts to discern whether or not SIT3 acts as a silicon sensor. This would enable experiments where SIT1/SIT2 expression and diatom silicification are quantified in the context of a SIT3 promoter deletion.

Homology-independent methods were used to identify putative silaffins, cingulins, silacidins, and frustulin sequences (Figure 3, Supplemental Table 7, Supplemental Table 8, Supplemental Table 9), but it is likely that most of these are not truly functioning in silicon processing because most of the genes identified are not

silicon sensitive (Figure 3) and the number of genes recovered are much higher than what has been found for *T. pseudonana* (Supplemental Table 7, Supplemental Table 8, Supplemental Table 9) (Poulsen & Kröger 2004; Scheffel et al. 2011; Wenzl et al. 2008; Armbrust et al. 2004). The candidates we did identify as being silicon-sensitive with N terminal ER targeting sequences could be good candidates to test for function using diatom gene editing methods. These methods have been used to characterize the urease gene in *T. pseudonana* (Hopes et al. 2016).

Another approach to finding silicon stress markers was to look for similar metabolic pathways that are silicon sensitive in *T. oceanica* and *T. weissflogii*. Although similar KEGG pathways are silicon-sensitive in the two diatoms, the direction of expression is sometimes in opposition in each diatom (e.g., upregulated in one silicon limited diatom and down in the other) (Figure 5). This trend is most clear in the “Transcription”, “Translation”, and “Replication and repair” categories (Figure 5). These different manifestations of silicon limitation may be related to the fact that silicon is required for both DNA synthesis and cell wall synthesis (Darley & Volcani 1969; Eppley et al. 1967; Hildebrand et al. 2007), so it is possible that silicon limitation is disproportionately influencing these two processes in the diatoms tested. In this context, the argument for using the ATP-grasp domain proteins as molecular markers of silicon status is quite strong, as both manifestations of silicon limitation saw this gene upregulated.

## References

- Allen, J.T. et al., 2005. Diatom carbon export enhanced by silicate upwelling in the northeast Atlantic. *Nature*, 437(7059), pp.728–732.
- Alverson, A.J., 2007. Strong purifying selection in the silicon transporters of marine and freshwater diatoms. *Limnology and Oceanography*, 52(4), pp.1420–1429.
- Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Armbrust, E.V. et al., 2004. The Genome of the Diatom *Thalassiosira pseudonana*: Ecology, Evolution, and Metabolism. *Science*, 306(October), pp.79–86.
- Bertrand, E.M. et al., 2015. Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proceedings of the National Academy of Sciences of the United States of America*, 112(32), pp.9938–43.
- Bouchereau, A. et al., 1999. Polyamines and environmental challenges: recent development. *Plant Science*, 140, pp.103–125.
- Boyd, P. et al., 1999. Role of iron, light, and silicate in controlling algal biomass in subantarctic waters SE of New Zealand. *Journal of Geophysical Research*, 104, pp.395–408.
- Brembu, T. et al., 2017. Dynamic responses to silicon in *Thalassiosira pseudonana* - Identification, characterisation and classification of signature genes and their corresponding protein motifs. *Scientific Reports*, 7(1), pp.1–14.
- Brzezinski, M. a. et al., 2011. Co-limitation of diatoms by iron and silicic acid in the equatorial Pacific. *Deep-Sea Research Part II: Topical Studies in Oceanography*,

58(3–4), pp.493–511.

Brzezinski, M., Olson, R. & Chisholm, S., 1990. Silicon availability and cell-cycle progression in marine diatoms. *Marine Ecology Progress Series*, 67(M), pp.83–96.

Chappell, P.D. et al., 2015. Genetic indicators of iron limitation in wild populations of *Thalassiosira oceanica* from the northeast Pacific Ocean. *The ISME Journal*, 9, pp.592–602.

Chappell, P.D. et al., 2016. Preferential depletion of zinc within Costa Rica upwelling dome creates conditions for zinc co-limitation of primary production. *Journal of Plankton Research*, 38(2), pp.244–255.

Darley, W.M. & Volcani, B.E., 1969. Role of silicon in diatom metabolism. A silicon requirement for deoxyribonucleic acid synthesis in the diatom *Cylindrotheca fusiformis* Reimann and Lewin. *Experimental cell research*, 58(2), pp.334–342.

Drum, R.W. & Pankratz, S.H., 1964. Post mitotic fine structure of *Gomphonema parvulum*. *Journal of ultrastructure research*, 10, pp.217–223.

Du, C. et al., 2014. ITRAQ-based proteomic analysis of the metabolism mechanism associated with silicon response in the marine diatom *Thalassiosira pseudonana*. *Journal of Proteome Research*, 13(2), pp.720–734.

Dugdale, R.C. & Wilkerson, F.P., 1998. Silicate regulation of new production in the equatorial Pacific upwelling. *Nature*, 391(January), pp.270–273.

Dugdale, R.C., Wilkerson, F.P. & Minas, H.J., 1995. The role of a silicate pump in driving new production. *Deep-Sea Research Part I: Oceanographic Research Papers*, 42(5), pp.697–719.

- Durkin, C.A. et al., 2016. The evolution of silicon transporters in diatoms. *Journal of Phycology*, 52(5), pp.716–731.
- Eppley, R.W., Holmes, R.W. & Paasche, E., 1967. Periodicity in cell division and physiological behavior of *Ditylum brightwellii* a marine planktonic diatom during grow in light-dark cycles. *Archiv Fur Mikrobiologie*, 56, pp.305–323.
- Fawaz, M. V, Topper, M. & Firestine, S.M., 2011. The ATP-Grasp Enzymes. *Bioorganic Chemistry*, 39(5–6), pp.185–191.
- Field, C.B. et al., 1998. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, 281(5374), pp.237–240.
- Franck, V.M. et al., 2000. Iron and silicic acid concentrations regulate Si uptake north and south of the Polar Frontal Zone in the Pacific Sector of the Southern Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 47(15–16), pp.3315–3338.
- Frischkorn, K.R. et al., 2014. De novo assembly of *Aureococcus anophagefferens* transcriptomes reveals diverse responses to the low nutrient and low light conditions present during blooms. *Frontiers in Microbiology*, 5(JULY), pp.1–16.
- Fu, L. et al., 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), pp.3150–3152.
- Galperin, M.Y. & Koonin, E. V., 1997. A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. *Protein Science*, 6(12), pp.2639–2643.
- Gamsjaeger, R. et al., 2006. Sticky fingers : zinc-fingers as protein-recognition motifs. *Trends in Biochemical Sciences*, 32(2).

- Gao, Q. & Garcia-Pichel, F., 2011. An ATP-Grasp ligase involved in the last biosynthetic step of the iminomycosporine shinorine in *Nostoc punctiforme* ATCC 29133. *Journal of Bacteriology*, 193(21), pp.5923–5928.
- Glotzer, M., Murray, A.W. & Kirschner, M.W., 1991. Cyclin is degraded by the ubiquitin pathway. *Nature*, 349(6305), pp.132–138.
- Grant, B. & Greenwald, I., 1996. The *Caenorhabditis elegans* sel-1 gene, a negative regulator of lin-12 and glp-1, encodes a predicted extracellular protein. *Genetics*, 143(1), pp.237–247.
- Grant, C.E., Bailey, T.L. & Noble, W.S., 2011. FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, 27(7), pp.1017–1018.
- Guillard, R.R.L., 1975. Culture of Phytoplankton for Feeding Marine Invertebrates. In *Culture of Marine Invertebrate Animals*. pp. 29–60.
- Guillard, R.R.L. & Ryther, J.H., 1962. Studies of marine planktonic diatoms I. *Cyclotella nana* Hustedt, and *Detonula confervasea* (Cleve) Gran. *Canadian Journal of Microbiology*, 8(1140), pp.229–239.
- Hildebrand, M. et al., 1997. A gene family of silicon transporters. *Nature*, 385, pp.688–689.
- Hildebrand, M., Dahlin, K. & Volcani, B.E., 1998. Characterization of a silicon transporter gene family in *Cylindrotheca fusiformis*: sequences, expression analysis, and identification of homologs in other diatoms. *Molecular and General Genetics MGG*, 260(5), pp.480–486.
- Hildebrand, M., Frigeri, L.G. & Davis, A.K., 2007. Synchronized growth of *Thalassiosira pseudonana* (Bacillariophyceae) provides novel insights into cell-

- wall synthesis processes in relation to the cell cycle. *Journal of Phycology*, 43(4), pp.730–740.
- Hopes, A. et al., 2016. Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant Methods*, 12(1), pp.1–12.
- Hutchins, D. a & Bruland, K.W., 1998. Iron-limited diatom growth and Si : N uptake ratios in a coastal upwelling regime. *Nature*, 393(June), pp.561–564.
- Huysman, M.J.J. et al., 2010. Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. *Genome biology*, 11(2), p.R17.
- Ingalls, A.E., Whitehead, K. & Bridoux, M.C., 2010. Tinted windows: The presence of the UV absorbing compounds called mycosporine-like amino acids embedded in the frustules of marine diatoms. *Geochimica et Cosmochimica Acta*, 74(1), pp.104–115.
- Johnson, M. et al., 2007. Ammonium accumulation during a silicate-limited diatom bloom indicates the potential for ammonia emission events. *Marine Chemistry*, 106(1–2 SPEC. ISS.), pp.63–75.
- Jones, P. et al., 2014. InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), pp.1236–1240.
- Jørgensen, E.G., 1955. Variations in the Silica Content of Diatoms. *Physiologia Plantarum*, 8(4), pp.840–845.
- Katoh, K. & Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), pp.772–780.

- Keeling, P.J. et al., 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biology*, 12(6).
- King, A.L. et al., 2015. Effects of CO<sub>2</sub> on growth rate, C:N:P, and fatty acid composition of seven marine phytoplankton species. *Marine Ecology Progress Series*, 537, pp.59–69.
- Koike, I. et al., 2001. Silicate to nitrate ratio of the upper sub-arctic Pacific and the Bering Sea basin in summer: Its implication for phytoplankton dynamics. *Journal of Oceanography*, 57(3), pp.253–260.
- Kooistra, W.H.C.F. et al., 2008. Global Diversity and Biogeography of *Skeletonema* Species (Bacillariophyta). *Protist*, 159(2), pp.177–193.
- Kröger, N. et al., 2002. Self-Assembly of Highly Phosphorylated Silaffins and Their Function in Biosilica Morphogenesis. *Science*, 298(5593), pp.584–586.
- Kröger, N. et al., 1997. Characterization of a 200-kDa Diatom Protein that is Specifically Associated with a Silica-Based Substructure of the Cell Wall. *Eur J Biochem*, 250(1), pp.99–105.
- Kröger, N. et al., 2000. Species-specific polyamines from diatoms control silica morphology. *Proceedings of the National Academy of Sciences of the United States of America*, 97(26), pp.14133–14138.
- Kröger, N., Bergsdorf, C. & Sumper, M., 1994. A new calcium binding glycoprotein family constitutes a major diatom cell wall component. *The EMBO journal*, 13(19), pp.4676–4683.
- Kröger, N., Bergsdorf, C. & Sumper, M., 1996. Frustulins: domain conservation in a

- protein family associated with diatom cell walls. *European Journal of Biochemistry*, 239(2), pp.259–264.
- Kröger, N., Deutzmann, R. & Sumper, M., 1999. Polycationic Peptides from Diatom Biosilica That Direct Silica Nanosphere Formation. *Science*, 286(5442), pp.1129–1132.
- Kröger, N. & Poulsen, N., 2008. Diatoms-from cell wall biogenesis to nanotechnology. *Annual review of genetics*, 42, pp.83–107.
- Kröger, N. & Wetherbee, R., 2000. Pleuralins are involved in theca differentiation in the diatom *Cylindrotheca fusiformis*. *Protist*, 151(3), pp.263–273.
- De La Rocha, C.L. et al., 2000. Effects of iron and zinc deficiency on elemental composition and silica production by diatoms. *Marine Ecology Progress Series*, 195, pp.71–79.
- Leblanc, K. et al., 2005. Fe and Zn effects on the Si cycle and diatom community structure in two contrasting high and low-silicate HNLC areas. *Deep-Sea Research Part I: Oceanographic Research Papers*, 52(10), pp.1842–1864.
- Lewin, J.C., 1957. Silicon metabolism in diatoms IV. Growth and frustule formation in *Navicula pelliculosa*. *Canadian Journal of Microbiology*, 3, pp.427–433.
- Lewin, R.A., 1962. Silicification. In *Physiology and biochemistry of algae*. New York: Academic Press, pp. 445–455.
- Li, W. & Godzik, A., 2006. CD-HIT: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), pp.1658–1659.
- Lommer, M. et al., 2012. Genome and low-iron response of an oceanic diatom adapted

- to chronic iron limitation. *Genome Biology*, 13(7), p.R66.
- Martin-Jézéquel, V., Hildebrand, M. & Brzezinski, M.A., 2000. REVIEW SILICON METABOLISM IN DIATOMS: IMPLICATIONS FOR GROWTH. *J. Phycol.*, 36(February), pp.821–840.
- Miele, V., Penel, S. & Duret, L., 2011. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC bioinformatics*, 12, p.116.
- Min, X.J. et al., 2005. OrfPredictor: Predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research*, 33(SUPPL. 2), pp.677–680.
- Mitchison, J.M., 1971. *The biology of the cell cycle*, Cambridge: Cambridge University Press.
- Mittl, P.R.E. & Schneider-Brachert, W., 2007. Sell-like repeat proteins in signal transduction. *Cellular Signalling*, 19(1), pp.20–31.
- Mock, T. et al., 2017. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature Publishing Group*, 541(7638), pp.536–540.
- Mock, T. et al., 2008. Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. *Proceedings of the National Academy of Sciences of the United States of America*, 105(5), pp.1579–84.
- Moriya, Y. et al., 2007. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, 35(SUPPL.2), pp.182–185.
- Nelson, D.M. et al., 1995. Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles*, 9(3), pp.359–372.

- Paasche, E., 1973. Silicon and the ecology of marine plankton diatoms. I. *Thalassiosira pseudonana* (Cyclotella nana) grown in a chemostat with silicate as limiting nutrient. *Marine Biology*, 19(2), pp.117–126.
- Pagès, H. et al., 2016. Biostrings: String objects representing biological sequences, and matching algorithms.
- Pamirsky, I.E. & Golokhvast, K.S., 2013. Origin and status of homologous proteins of biomineralization (Biosilicification) in the taxonomy of phylogenetic domains. *BioMed Research International*, 2013.
- Petersen, T.N. et al., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*, 8(10), pp.785–6.
- van de Poll, W.H., Vrieling, E.G. & Gieskes, W.W.C., 1999. Location and expression of frustulins in the pennate diatoms *Cylindrotheca fusiformis*, *Navicula pelliculosa*, and *Navicula salinarum* (Bacillariophyceae). *J Phycol*, 35, pp.1044–1053.
- Poulsen, N. et al., 2013. Pentalysine clusters mediate silica targeting of silaffins in *Thalassiosira pseudonana*. *Journal of Biological Chemistry*, 288(28), pp.20100–20109.
- Poulsen, N. & Kröger, N., 2004. Silica morphogenesis by alternative processing of silaffins in the diatom *Thalassiosira pseudonana*. *Journal of Biological Chemistry*, 279(41), pp.42993–42999.
- Price, N.M. et al., 1989. Preparation and Chemistry of the Artificial Algal Culture Medium Aquil. *Biological Oceanography*, 6(5–6), pp.443–461.
- R Development Core Team, 2013. R: A language and environment for statistical

- computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>. *R Foundation for Statistical Computing, Vienna, Austria*.
- Richthammer, P. et al., 2011. Biomineralization in Diatoms: The Role of Silacidins. *ChemBioChem*, 12(9), pp.1362–1366.
- Rueter, J.G. & Morel, F.M.M., 1981. The interaction between zinc deficiency and copper toxicity as it affects the silicic acid uptake mechanisms in *Thalassiosira pseudonana*. *Limnology and Oceanography*, 26(1), pp.67–73.
- Saito, M.A., Goepfert, T.J. & Ritt, J.T., 2008. Some thoughts on the concept of colimitation: Three definitions and the importance of bioavailability. *Limnology and Oceanography*, 53(1), pp.276–290.
- Sapriel, G. et al., 2009. Genome-wide transcriptome analyses of silicon metabolism in *Phaeodactylum tricornutum* reveal the multilevel regulation of silicic acid transporters. *PLoS ONE*, 4(10).
- Scheffel, A. et al., 2011. Nanopatterned protein microrings from a diatom that direct silica morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 108(8), pp.3175–80.
- Shrestha, R. et al., 2012. Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. *BMC Genomics*, 13(1), p.499.
- Smith, S.R. et al., 2016. Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. *New Phytologist*, 210(3), pp.890–904.
- Sunda, W.G. & Huntsman, S.A., 1995. Iron uptake and growth limitation in oceanic

- and coastal phytoplankton. *Marine Chemistry*, 50, pp.189–206.
- Tesson, B. & Hildebrand, M., 2010. Extensive and intimate association of the cytoskeleton with forming silica in diatoms: Control over patterning on the meso- and micro-scale. *PLoS ONE*, 5(12).
- Thamatrakoln, K., Alverson, A.J. & Hildebrand, M., 2006. Comparative sequence analysis of diatom silicon transporters: Toward a mechanistic model of silicon transport. *Journal of Phycology*, 42(4), pp.822–834.
- Thamatrakoln, K. & Hildebrand, M., 2007. Analysis of *Thalassiosira pseudonana* silicon transporters indicates distinct regulatory levels and transport activity through the cell cycle. *Eukaryotic Cell*, 6(2), pp.271–279.
- Thamatrakoln, K. & Hildebrand, M., 2008. Silicon uptake in diatoms revisited: a model for saturable and nonsaturable uptake kinetics and the role of silicon transporters. *Plant Physiol*, 146(3), pp.1397–1407.
- Treguer, P. et al., 1995. The Silica Balance in the World Ocean - a Reestimate. *Science*, 268(5209), pp.375–379.
- Vaulot, D. et al., 1987. Cell-cycle response to nutrient starvation in two phytoplankton species, *Thalassiosira weissflogii* and *Hymenomonas carterae*. *Marine Biology*, 95, pp.625–630.
- Wenzl, S. et al., 2008. Silacidins: Highly acidic phosphopeptides from diatom shells assist in silica precipitation in vitro. *Angewandte Chemie - International Edition*, 47(9), pp.1729–1732.
- Whitney, L.A.P. et al., 2011. Characterization of putative iron responsive genes as species-specific indicators of iron stress in *Thalassiosiroid* diatoms. *Frontiers in*

*Microbiology*, 2(NOV), pp.1–14.

Wu, Z. et al., 2010. Empirical bayes analysis of sequencing-based transcriptional profiling without replicates. *BMC Bioinformatics*, 11(1), p.564.

Yool, A. & Tyrrell, T., 2003. The role of diatoms in regulating the ocean's silicate cycle. *Global Biogeochemical Cycles*, 17(4), pp.1–24.

Yoshida, Y. et al., 2016. De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC genomics*, 17(1), p.182.

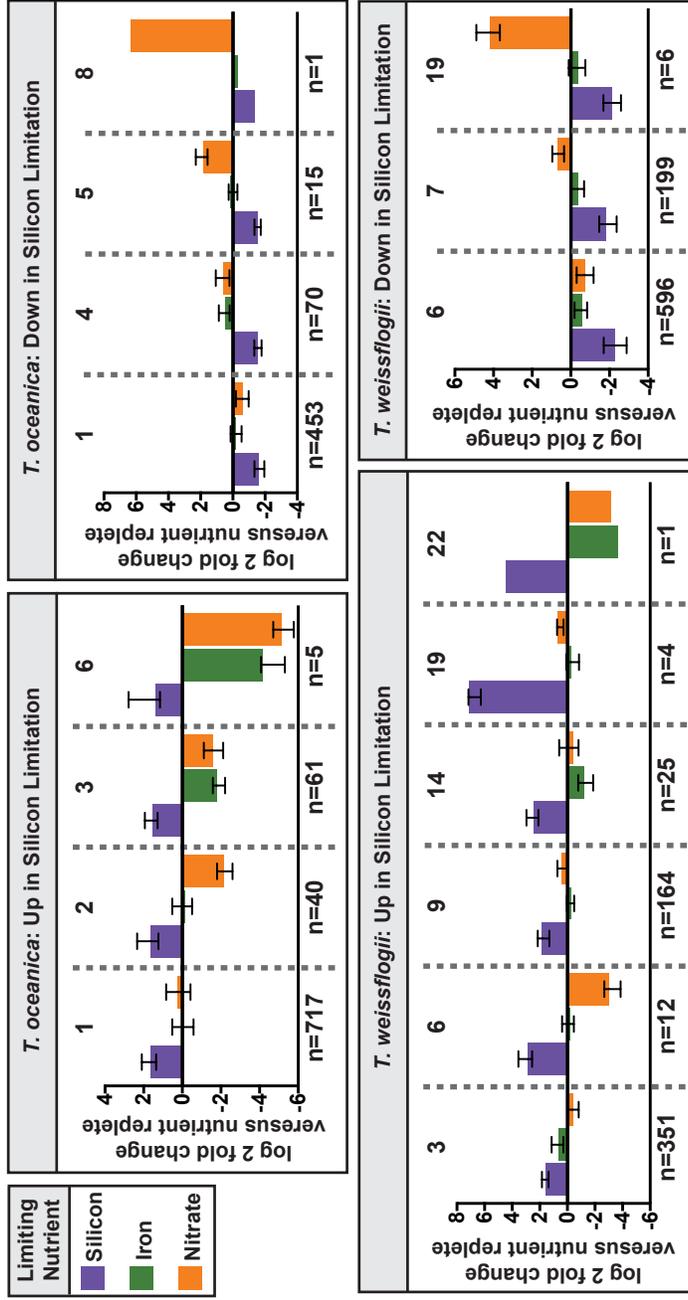


Figure 1. Depiction of the results from the hierarchical clustering of *T. oceanica* and *T. weissflogii* silicon sensitive genes, with only the results from silicon-specific clusters shown. Each panel shows the per-cluster gene expression values as a log<sub>2</sub>-fold change in silicon, iron, or nitrate limitation relative to the nutrient replete condition. The bar height indicates the per-cluster median expression value and the error bars indicate the interquartile range. The numbers at the bottom of each panel indicate the number of genes in each cluster.

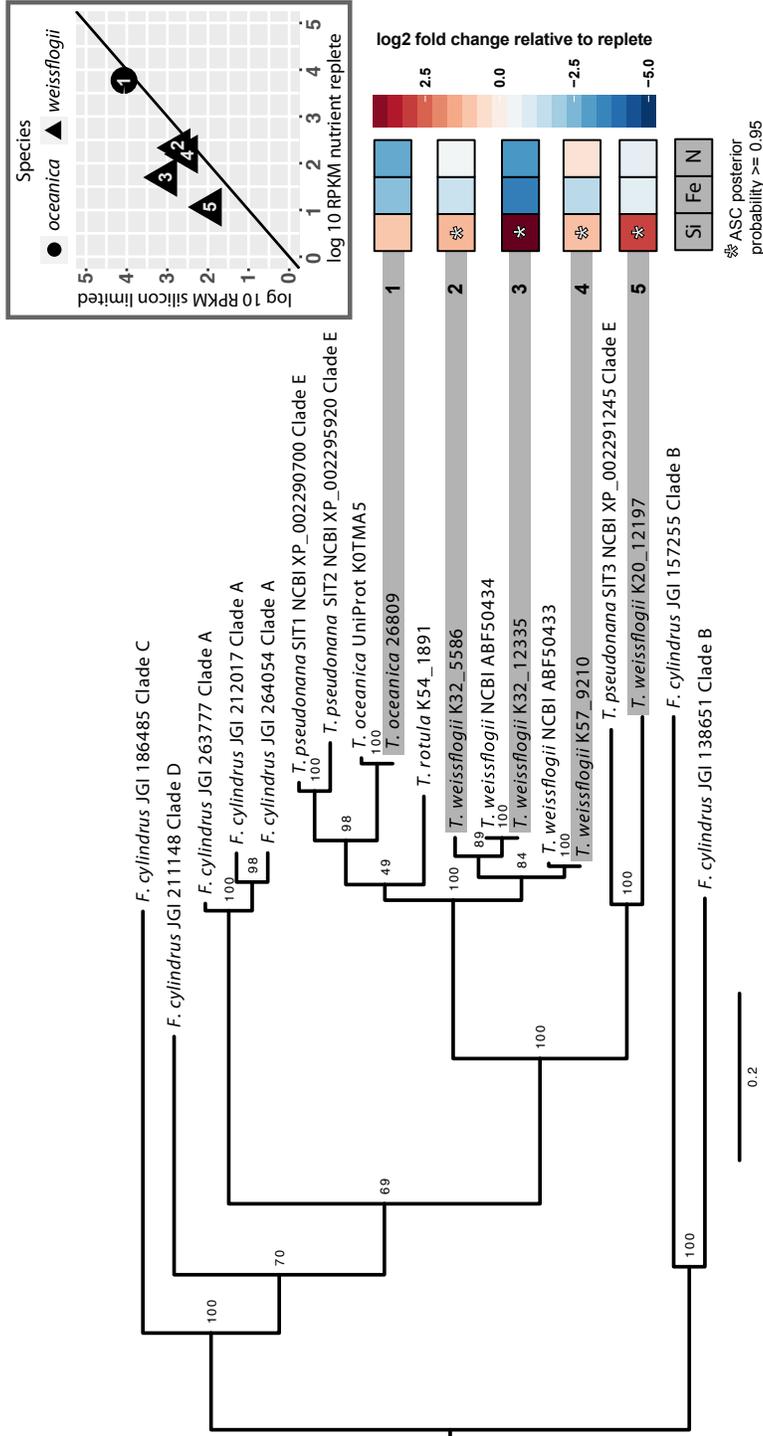


Figure 2. Phylogenetic analysis of diatom silicon transporters. A maximum likelihood comparison for silicon transporters from *Fragilariopsis cylindrus*, *Thalassiosira pseudonana*, *T. rotula*, *T. oceanica*, and *T. weissflogii* was conducted. For each silicon transporter identified in this study, a heat map is also shown. The heat map colors indicate the log<sub>2</sub>-fold change of each putative silicon transporter in silicon (Si), iron (Fe), or nitrate (N) limited condition relative to the nutrient replete condition. The asterisk indicates which silicon transporters were statistically significantly differentially expressed. A 1:1 line is also inset, which shows the log<sub>10</sub> RPKM expression value in nutrient replete (x-axis) and silicon limited (y-axis) conditions. Each point corresponds to a silicon transporter and the shape of the point indicates which species the silicon transporter originates from. The black diagonal line indicates where points would fall if they have the same expression value in the two conditions compared

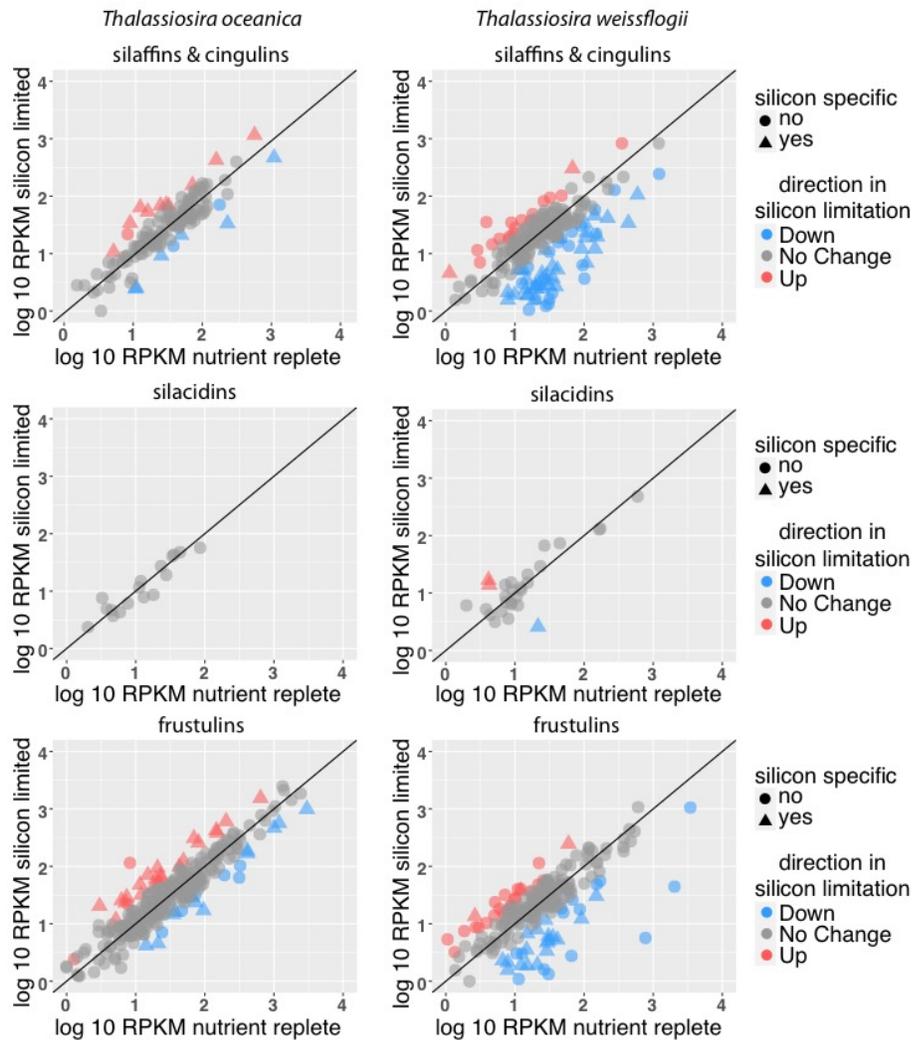


Figure 3. A series of 1:1 lines showing the expression values for putative silaffins, cingulins, silacidins, and frustulins identified in this study. Each point represents a single transcript. The x-axis is the log 10 transformed RPKM values in the nutrient replete condition and the y-axis is the log 10 transformed RPKM values in silicon limitation. The black diagonal line indicates where points would fall if they have the same expression value in the two conditions compared. The color of the point indicates the direction of differential expression in silicon limitation (up (red), down (blue), or no change) and the shape indicates whether or not each differentially expressed transcript is specific to silicon limitation or is part of a general stress response.

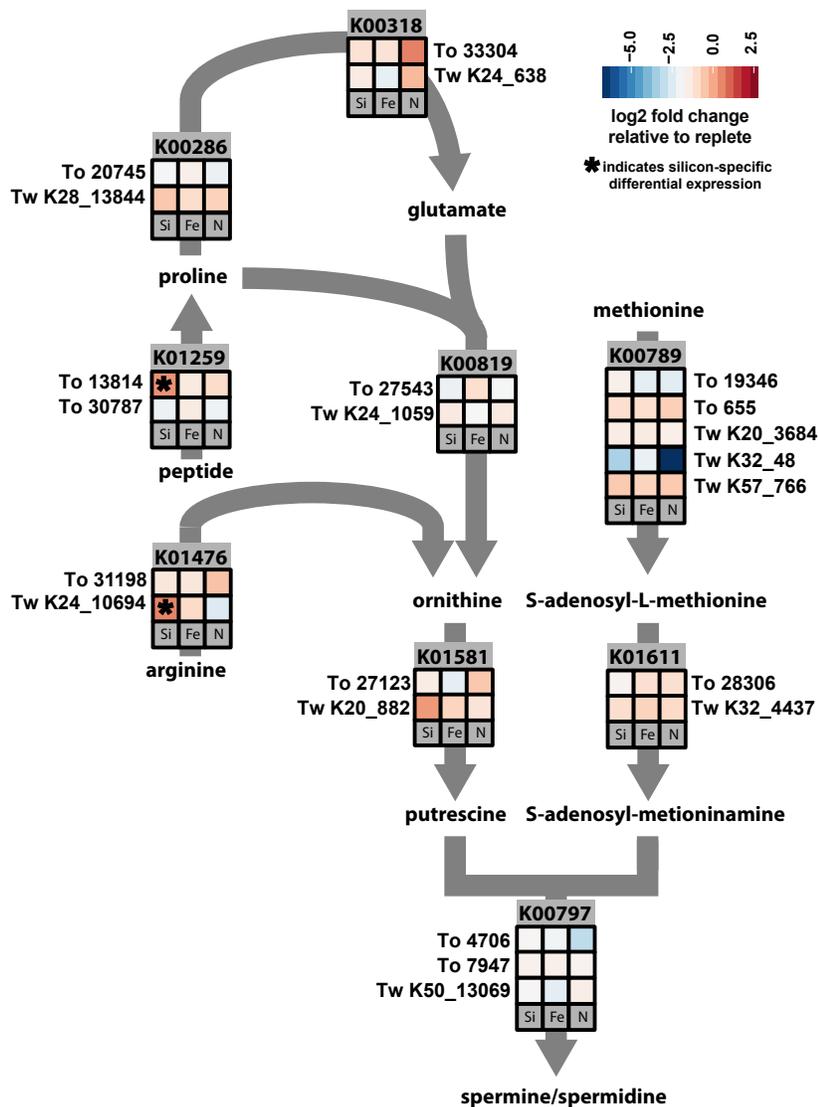


Figure 4. Heat map overlaid on a pathway map of genes involved in the precursors of long-chain polyamine biosynthesis. Colors in the heat map indicate the log<sub>2</sub>-fold change in silicon (Si), iron (Fe), or nitrate (N) limitation relative to the nutrient replete condition. The full description for each KEGG orthology are as follows: K00138:aldB, aldehyde dehydrogenase; K00286:proC, pyrroline-5-carboxylate reductase; K01259:pip, proline iminopeptidase; K01476:arginase; K00819:rocD, ornithine--oxo-acid transaminase; K01581:ornithine decarboxylase; K00789:metK, S-adenosylmethionine synthetase; K01611:speD, S-adenosylmethionine decarboxylase; K00797:speE, spermidine synthase

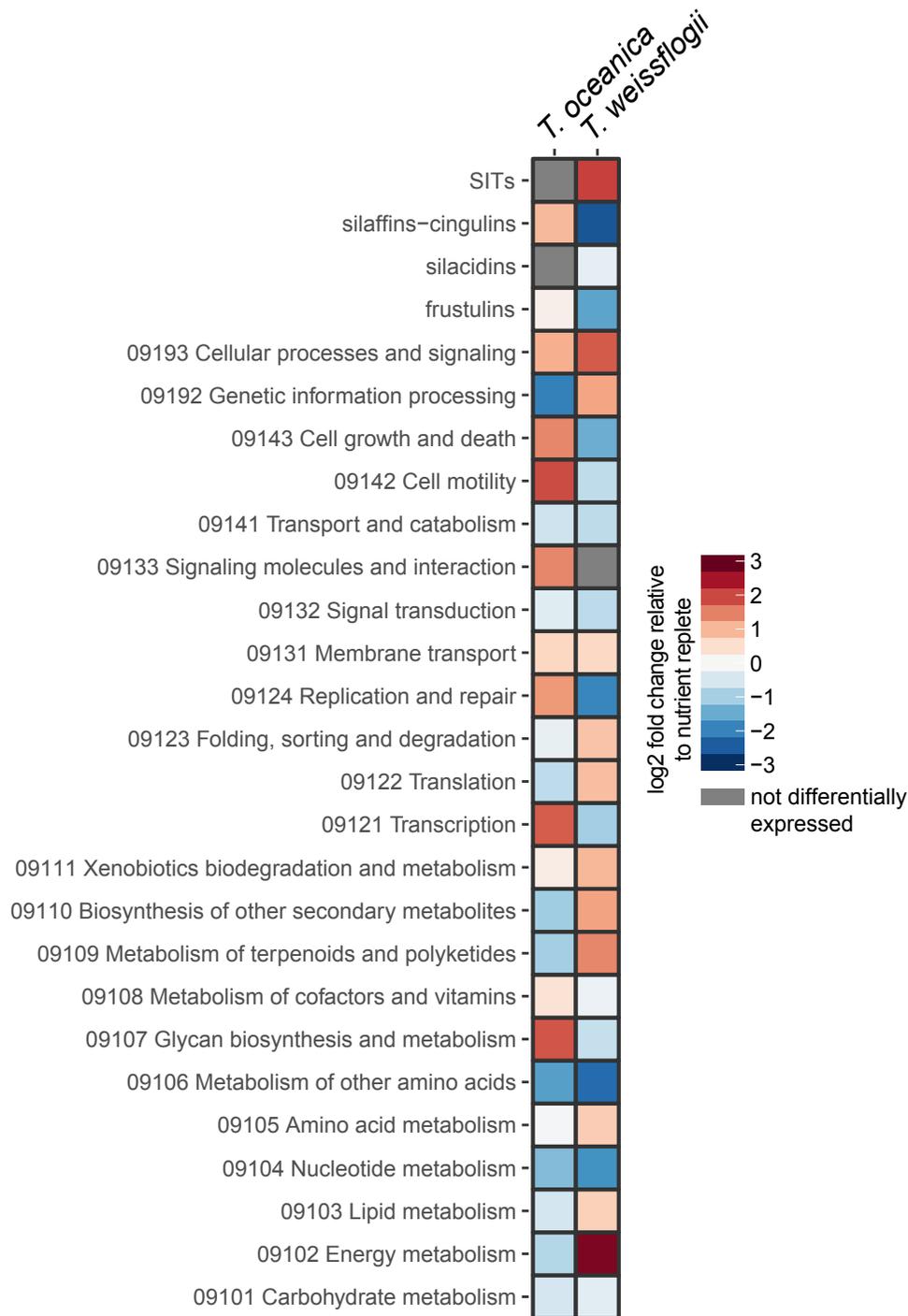


Figure 5. Heat map of the silicon specific, differentially expressed genes from *T. oceanica* and *T. weissflogii* and their KEGG orthology or silicon metabolism gene assignment categories. Colors in the heat map indicate the relative proportion of silicon specific genes in each category. Some categories that would not apply to diatoms were excluded from this analysis (e.g., Cardiovascular diseases).

Table 1. Summary table for *T. oceanica* and *T. weissflogii* showing the total number of transcripts assembled and number of transcripts with function or protein domain information assigned by KEGG automated annotation server or InterProScan.

	<i>T. oceanica</i>	<i>T. weissflogii</i>
Total Transcripts	36382	24900
Annotated	12450 (34%)	9580 (38%)

Table 2. Summary table for *T. oceanica* and *T. weissflogii* showing the number of transcripts differentially expressed in silicon limitation relative to nutrient replete conditions ( $\geq 0.95$  posterior probability of a 2-fold change). Numbers shown in the “Si specific” rows indicate the numbers of genes that are silicon-specifically differentially expressed and not part of a general stress response.

	<i>T. oceanica</i>	<i>T. weissflogii</i>
Up	947	2814
<b>Up Si specific</b>	<b>823</b>	<b>557</b>
Down	809	1467
<b>Down Si specific</b>	<b>539</b>	<b>801</b>

Manuscript 2

Formatted for publication in Marine Ecology Progress Series

*Thalassiosira* diatoms display distinct, species-specific strategies for responding to  
changes in pCO<sub>2</sub>

Joselynn Wallace<sup>1</sup>, Andrew L. King<sup>2</sup>, Gary H. Wilkfors<sup>3</sup>, Shannon L. Meseck<sup>3</sup>,  
Bethany D. Jenkins<sup>1,4</sup>

<sup>1</sup> Department of Cell and Molecular Biology, University of Rhode Island, Kingston,  
RI, USA

<sup>2</sup> NIVA Norwegian Institute for Water Research, Bergen, Norway

<sup>3</sup> NOAA/NMFS, Northeast Fisheries Science Center, Milford Laboratory, Milford,  
CT, USA

<sup>4</sup> Graduate School of Oceanography, University of Rhode Island, Narragansett, RI,  
USA

## ABSTRACT

Fossil fuel combustion has raised atmospheric pCO<sub>2</sub> from ~277 ppm in 1750 to the present level of ~400 ppm. The ocean sequesters ~30% of that CO<sub>2</sub>, decreasing its buffering capacity and pH (ocean acidification). Diatoms are responsible for ~40% of oceanic primary production based upon photosynthetic fixation of CO<sub>2</sub>. This is aided by a carbon-concentrating mechanism (CCM), which uses bicarbonate transporters (BCTs) to take up HCO<sub>3</sub><sup>-</sup> and carbonic anhydrases (CAs) to interconvert HCO<sub>3</sub><sup>-</sup> and CO<sub>2</sub>. There are several distinct classes of BCTs and CAs found in diatoms, including the substrate-specific SLC4 and non-specific SLC26 BCTs, as well as the  $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\gamma$ ,  $\zeta$ , and  $\theta$  CAs. The ability to regulate and conserve energetic costs associated with the CCM may influence diatom community composition in a higher CO<sub>2</sub> future. CO<sub>2</sub> manipulation experiments were conducted with four *Thalassiosira* species – *T. pseudonana*, *T. rotula*, *T. weissflogii*, and *T. oceanica*, and transcriptomes were sequenced from the latter 3 species. *T. rotula* is a coastal diatom that may have experienced high CO<sub>2</sub> during seasonal upwelling, while *T. oceanica* and *T. weissflogii* are open-ocean isolates that have likely experienced relatively stable CO<sub>2</sub> conditions (Feely et al. 2008; Hofmann et al. 2011). These species displayed a range of growth rate changes across CO<sub>2</sub> conditions (<230 ppm CO<sub>2</sub> to >2900 ppm CO<sub>2</sub>), from no change across conditions (*T. pseudonana*), slower growth in lower CO<sub>2</sub> (*T. weissflogii*), slower growth in higher CO<sub>2</sub> (*T. oceanica*), and faster growth in higher CO<sub>2</sub> (*T. rotula*) (King et al. 2015). The accompanying transcriptome evidence for *T. rotula*, *T. weissflogii*, and *T. oceanica* indicate that these differences in CO<sub>2</sub> responses boil down to the ability to dynamically regulate the carbon concentrating mechanism

while maintaining internal redox and ion homeostasis, as well as trade-offs between funneling excess CO<sub>2</sub> into internal storage pools versus a higher growth rate.

## **INTRODUCTION**

Anthropogenic activity (fossil fuel emissions and land use changes) has led to an increase in atmospheric pCO<sub>2</sub> concentration from ~277 ppm in 1750 to over 400 ppm currently (Tans 2009; Le Quéré et al. 2016; Scripps 2017). Atmospheric CO<sub>2</sub> equilibrates with the ocean, which sequesters ~20% of anthropogenic CO<sub>2</sub> through physical processes and biological activity and has been the only net sink for CO<sub>2</sub> in the past 200 years (Sabine et al. 2004; Khatiwala et al. 2013). Phytoplankton are organisms that live in the surface ocean and convert CO<sub>2</sub> into organic biomass via photosynthesis and serve as an important biological source of CO<sub>2</sub> uptake. Phytoplankton cells that sink out of the lit ocean can export the CO<sub>2</sub> formerly in the atmosphere, to the deep ocean.

As global CO<sub>2</sub> levels rise, the increased dissolved CO<sub>2</sub> in the ocean disturbs carbonate chemistry and leads to a decrease in pH and buffering capacity, or ocean acidification (Caldeira & Wickett 2003; Sabine et al. 2004). Ocean acidification is thought to have negative consequences for some phytoplankton groups, like those that have calcareous skeletons comprised of calcium carbonate (e.g., Riebesell et al. 2000). The impact of ocean acidification on other groups of non-calcified phytoplankton, like diatoms that have silicified skeletons, is less clear (e.g., King et al. 2015).

Diatoms are a phytoplankton group that perform roughly 40% of oceanic primary production via photosynthetic CO<sub>2</sub> fixation (Nelson et al. 1995). Diatoms have experienced fluctuations in pCO<sub>2</sub> on different temporal scales throughout their

evolutionary history. Molecular clock estimates suggest that diatoms arose around 240 million years ago (Sims et al. 2006; Medlin 2016). This coincides with the start of a glacial “hot-house” period, where atmospheric pCO<sub>2</sub> levels were as high as 6000 ppm (Kump et al. 2009). In more recent history, diatoms have also experienced low pCO<sub>2</sub>: pre-industrial (before 1850) atmospheric pCO<sub>2</sub> averaged ~270 ppm (Tans 2009). Diatoms can also experience fluctuations in pCO<sub>2</sub> levels that vary on much shorter time scales. Coastal diatoms might experience dissolved pCO<sub>2</sub> concentrations greater than 850 ppm during seasonal coastal upwelling events (Feely et al. 2008). The distribution patterns of diatoms in the ocean can also influence the pCO<sub>2</sub> conditions they experience over the course of their life histories. Relative to coastal diatoms, diatoms living in the open-ocean experience more stable pCO<sub>2</sub> conditions (Hofmann et al. 2011).

Climate models suggest that pCO<sub>2</sub> levels in the atmosphere may rise to ~1000 ppm by the year 2100 if no measures are taken to reduce emissions (Solomon et al. 2007). In light of the sensitivity of calcified phytoplankton to ocean acidification, diatoms may have an even larger role to play in the marine food web if no measures are taken to reduce CO<sub>2</sub> emissions. Within the diatoms as well, there will likely be species that are ecological "winners" in adapting to elevated pCO<sub>2</sub>. Physiological studies examining how diatoms respond to changing pCO<sub>2</sub> suggest that there is a wide range of physiological responses within the diatoms. Members of the same genera differ in their physiological response to pCO<sub>2</sub> concentration. This is illustrated by the differences in specific growth rates of members of the *Thalassiosira* genus grown in different CO<sub>2</sub> conditions. Some genera within this species seem to thrive in high CO<sub>2</sub>;

the specific growth rate of *T. rotula* increased by at least 13% in elevated (690 ppm) CO<sub>2</sub> (King et al. 2015). A similar trend is seen in *T. weissflogii*, which grows faster in high CO<sub>2</sub> and slower in low pCO<sub>2</sub> (King et al. 2015; Shi et al. 2010). The data for *T. oceanica* is less clear – in one study the specific growth rate is decreased by at least 19% in high pCO<sub>2</sub> (700 ppm) (King et al. 2015), while other studies indicate that the specific growth rate of this diatom is increased in high pCO<sub>2</sub> (Shi et al. 2010). Other *Thalassiosira* diatoms appear to be indifferent to CO<sub>2</sub> levels; the growth rate of *T. pseudonana* does not change with pCO<sub>2</sub> (Crawford et al. 2011; King et al. 2015).

One aspect of phytoplankton biology that may influence which groups are winners and losers in a higher CO<sub>2</sub> future is their ability to regulate how much of their energy budgets they spend on inorganic carbon concentrating mechanisms (Giordano et al. 2005; Raven et al. 2014). During the dark reactions of photosynthesis, diatoms use the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) to fix CO<sub>2</sub> onto ribulose bisphosphate, producing 3-phosphoglycerate and fueling the Calvin cycle. However, RuBisCO can also fix O<sub>2</sub> onto ribulose bisphosphate, producing 2-P-glycolate, fueling the less energetically efficient photorespiration pathway. Diatoms use carbon concentrating mechanisms (CCMs) to increase the supply of CO<sub>2</sub> proximal to RuBisCO, to energetically favor carbon fixation rather than photorespiration.

The use of carbon concentrating mechanisms and active transport of inorganic carbon requires a significant energetic investment by the cell (Raven 1991). If diatoms can downregulate CCM components when CO<sub>2</sub> levels are high, this may result in the conservation of cellular energy as anthropogenic CO<sub>2</sub> emissions rise. This energy conservation may confer a competitive advantage to some diatoms and other groups of

phytoplankton and could influence community composition in the future. The ability or inability to regulate the CCM in response to fluctuating pCO<sub>2</sub> may be related to how diatoms sense changes in pCO<sub>2</sub>.

Diatom CCMs have been studied mainly in the model diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*. Diatom CCMs depend on the activity of bicarbonate transporters and carbonic anhydrases -- particularly those enzymes localized to the plastid or the pyrenoid, a proteinaceous sub-compartment of the plastid where carbon fixation occurs (Hopkinson et al. 2011). Bicarbonate transporters allow membrane-impermeable bicarbonate to cross cell membranes, while carbonic anhydrases interconvert carbon dioxide and bicarbonate. This facilitates the diffusion of CO<sub>2</sub> across membranes and also allows for the conversion of CO<sub>2</sub> to bicarbonate to prevent it from leaking back out of the cell.

Carbonic anhydrases interconvert HCO<sub>3</sub><sup>-</sup> and CO<sub>2</sub>. They allow diatoms to concentrate CO<sub>2</sub> proximal to the carbon fixation enzyme RuBisCO, as they can prevent CO<sub>2</sub> leakage across membranes by controlling the interconversion of HCO<sub>3</sub><sup>-</sup> and CO<sub>2</sub> (Hopkinson 2014; Hopkinson et al. 2016). Carbonic anhydrases can also facilitate CO<sub>2</sub> diffusion into the cell if they are extracellular, and mitochondrial carbonic anhydrases may be involved in recuperating respired CO<sub>2</sub> (Hopkinson 2014; Hopkinson et al. 2016). There are several distinct classes of carbonic anhydrases found in diatoms, termed ( $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\gamma$ ,  $\zeta$ , and  $\theta$ ), each with their own unique evolutionary history and distribution across the tree of life and within diatom groups (Hewett-Emmett & Tashian 1996; Tachibana et al. 2011; Samukawa et al. 2014; Kikutani et al. 2016). For example, although the  $\delta$  carbonic anhydrases have been found in

*Thalassiosira* diatoms and other phytoplankton groups (e.g., haptophytes), they have not been found in *P. tricornutum* (McGinn & Morel 2008; Tachibana et al. 2011; Soto et al. 2006). The  $\zeta$  carbonic anhydrases have only been found in *Thalassiosira* diatoms (Tachibana et al. 2011; Samukawa et al. 2014). Thus, carbon concentrating mechanism function may differ between diatoms as their genetic complement is variable. Diatom carbonic anhydrases are sensitive to CO<sub>2</sub> conditions and proteomic and gene expression studies across pCO<sub>2</sub> conditions indicate that the  $\delta$  and  $\zeta$  carbonic anhydrases play an important role in the CCM of *Thalassiosira pseudonana* (Tachibana et al. 2011; Samukawa et al. 2014; Clement et al. 2017; Hennon et al. 2015; Crawford et al. 2011; McGinn & Morel 2008). There is some additional evidence that the  $\alpha$  and  $\gamma$  carbonic anhydrases may also play a role in the CCM in *T. pseudonana* (Tachibana et al. 2011; Samukawa et al. 2014).

Diatoms also differ as to whether they can use C<sub>4</sub> metabolism as part of a CCM. C<sub>4</sub> metabolism involves the carboxylation and decarboxylation of 4-carbon compounds and requires a carboxylating enzyme (phosphoenolpyruvate carboxylase) that is localized outside of the plastid and a decarboxylating enzyme (phosphoenolpyruvate carboxykinase or malic enzyme) that is plastid-localized (Reinfelder et al. 2000; Sage 2004). Transaminases, pyruvate phosphate dikinase, and malate dehydrogenase perform other conversions to ultimately regenerate phosphoenolpyruvate (Sage 2004). C<sub>4</sub> metabolism has been implicated to be a potentially important aspect of CCMs in the diatom *Thalassiosira weissflogii* (Reinfelder et al. 2004; Roberts et al. 2007) but is not thought to be part of the CCM for the model diatom *Phaeodactylum tricornutum* (Haimovich-Dayana et al. 2013).

Sensing of pCO<sub>2</sub> in diatoms may be via the activity of cAMP secondary messengers (Harada et al. 2006; Ohno et al. 2012; Hennon et al. 2015). Adenylyl cyclases generate cAMP secondary messengers from ATP in response to signals received from bicarbonate (Buck et al. 1999). The cAMP secondary messengers activate cAMP-dependent protein kinases that phosphorylate and activate bZIP domain transcription factors that in turn bind to cAMP response elements and regulate downstream gene expression (Ohno et al. 2012). Phosphodiesterases regulate cellular cAMP levels by dephosphorylating cAMP back into AMP (Conti & Beavo 2007). cAMP response binding-element proteins are also activated by other signal and activator pairs, including calcium and calcium-calmodulin-dependent kinases (Mayr & Montminy 2001). Studies in *P. tricornutum* and *T. pseudonana* suggest that the activity of a soluble adenylyl cyclase is increased in the presence of CO<sub>2</sub>/HCO<sub>3</sub>, and that the resulting increase in cellular cAMP levels induces the activity of bZIP transcription factors (Harada et al. 2006; Ohno et al. 2012; Hennon et al. 2015). The transcription factors bind to upstream CO<sub>2</sub>-cAMP-responsive elements and repress the expression of CCM genes, including the expression of the adenylyl cyclase itself (Hennon et al. 2015). When cAMP function is inhibited chemically (Harada et al. 2006) or by gene deletion (Ohno et al. 2012), the pCO<sub>2</sub> sensitivity of the β carbonic anhydrases is repressed in diatoms.

The majority of our molecular and genetic understanding regarding the metabolic adjustments diatoms may make in a high CO<sub>2</sub> future ocean is largely shaped by studies in two model diatoms, *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*. As these diatoms show significant differences in the metabolism they use

to respond to changes in pCO<sub>2</sub> (Hopkinson et al. 2016), we wanted to probe this response in more closely related diatoms that have broad ecological distributions, albeit in different ocean basins. In this study, we probed the genetic response of *Thalassiosiroid* diatoms (*T. rotula* CCMP3096, *T. oceanica* CCMP1005, and *T. weissflogii* CCMP1010) that demonstrate distinct physiological responses to different pCO<sub>2</sub> concentrations using quantitative global transcriptome profiling. Our gene expression data is consistent with our previous physiological study (King et al. 2015) in that related diatom species utilize distinct complements of genes involved in their CO<sub>2</sub>-dependent growth response and carbon concentrating mechanisms.

## **MATERIALS AND METHODS**

### **Culturing**

Culturing experiments were conducted as described elsewhere (King et al 2015). Briefly, triplicate cultures were grown in f/2 media at three (*T. rotula*) or four (*T. oceanica* and *T. weissflogii*) pCO<sub>2</sub> levels until growth rates stabilized (for 7-10 generations) (Guillard & Ryther 1962; Guillard 1975). CO<sub>2</sub> levels ranged from an approximation of: low/pre-industrial (less than 230 ppm pCO<sub>2</sub>), present-day ambient pCO<sub>2</sub> in the open ocean (320-390 ppm pCO<sub>2</sub>), future pCO<sub>2</sub> projections (690-1340 ppm pCO<sub>2</sub>), and the maximum that diatoms have experienced in their evolutionary history (2900-5100 ppm pCO<sub>2</sub>) (Sims et al. 2006; Medlin 2016; Kump et al. 2009).

After 7-10 generations at the target CO<sub>2</sub> level, cells were collected on 3 μm polyester filters (Sterlitech Corporation, Kent, WA, USA) by gentle filtration, flash frozen in liquid nitrogen, and stored at -80 °C until RNA could be extracted. RNA was extracted from filters following a modified version of the RNeasy Mini Kit protocol

(Qiagen, Venlo, Netherlands). The protocol was modified by vortexing filters in lysis buffer with 0.5 mm zircon beads (BioSpec, Bartlesville, OK, USA) for 15 minutes to lyse the cells. Cell debris was immediately removed by running lysate over a QiaShredder column (Qiagen). DNA contamination was removed by an on-column DNase step using the RNase-free DNase Set (Qiagen) and an additional DNase step was performed on eluted RNA using the Turbo DNA-free kit (Life Technologies, Carlsbad, CA, USA), both performed as per the manufacturers' instructions. Samples were quantified using the Qubit High-Sensitivity Kit (Life Technologies).

### **Sequencing, mapping, differential expression**

Equal amounts of total RNA from each triplicate were pooled and sent to the JP Sulzberger Genome Sequencing Center at Columbia University. All sequencing was performed on an Illumina HiSeq platform and all libraries were mRNA enriched using a poly-A pull-down method (Illumina, San Diego, CA, USA). Target sequencing depth was 30 million, single-end, 100 basepair read length.

Initial read quality assessments were performed in FastQC (Andrews 2010). Raw reads were subsequently imported into CLC Workbench (CLC Bio-Qiagen, Aarhus, Denmark) with an insert size ranging from 1-400 nucleotides. Reads were trimmed in CLC Workbench with the following parameters: removal of 13 bases off the 5' end, .001 quality cut off, and no ambiguous nucleotide base calls. Reads were mapped to de novo transcriptomes that were assembled as described elsewhere (Wallace et al., in prep). Reads were mapped using the RNAseq module in CLC Workbench with the following parameters: mismatch cost (2), insertion cost (3), deletion cost (3), length fraction (0.8), similarity fraction (0.8), paired distances (auto-

detect), strand specific (both), maximum number of hits for a read (10). Annotations from the transcriptome sequences were refined by running the translated protein sequences on the KEGG KAAS automated server (Moriya et al. 2007). The BLAST search program and bi-directional best hit options were selected. The sequences were queried against the “genes” data set, with the addition of the diatoms *Thalassiosira pseudonana* (tps) and *Phaeodactylum tricornutum* (pti). An additional annotation step was conducted using InterProScan with the pathways and goterms options selected (Jones et al. 2014).

Differential gene expression was inferred with Analysis of Sequence Counts (ASC) from the total numbers of reads mapped to each transcript (Wu et al. 2010). Transcript expression levels in the present-day ambient open ocean pCO<sub>2</sub> (320-390 ppm) were compared to expression levels in low/pre-industrial (less than 230 ppm pCO<sub>2</sub>), future pCO<sub>2</sub> projections (690-1340 ppm pCO<sub>2</sub>), and the maximum that diatoms have experienced in their evolutionary history (2900-5100 ppm pCO<sub>2</sub>). Transcripts with at least a 95% posterior probability of a 2-fold change were considered differentially expressed.

### **Identification and comparison of orthologs**

Protein sequences for putative carbonic anhydrases, bicarbonate transporters, and additional genes potentially involved in C4 metabolism were used for phylogenetic analysis. An initial BlastP search was conducted to find potential orthologous genes in the translated transcriptomes. Any sequence meeting our threshold against any of these queries (35% ID, 80% query coverage per HSP cutoff)

were considered an ortholog. We also included any sequence identified as a carbonic anhydrase or bicarbonate transporter by InterProScan (Jones et al. 2014).

Resulting sequences from the *T. oceanica* transcriptome were cross-referenced to the *T. oceanica* genome (Lommer et al. 2012). Putative carbonic anhydrase, bicarbonate transporter, and C4 ortholog nucleotide sequences from the *T. oceanica* transcriptome were queried against the Ensembl Protist *T. oceanica* genome blast database using blastn. The output values for the top corresponding genomic locations are reported and if multiple top hits were reported, all of those hits were included. This data is summarized in Supplemental Table 3.

For the carbonic anhydrases, the query sequences used were putative carbonic anhydrase sequences described in *Thalassiosira pseudonana* and *Phaeodactylum tricorutum* (Tachibana et al. 2011; Samukawa et al. 2014; Kikutani et al. 2016). Also included in this analysis were the eta type carbonic anhydrases from *Plasmodium falciparum* and the epsilon type carbonic anhydrases from *Halothiobacillus neapolitanus* (Del Prete et al. 2014; So et al. 2004).

For the bicarbonate transporters, we used the SLC4 sequences from *T. pseudonana* and *P. tricorutum* (Nakajima et al. 2013). We queried the *T. pseudonana* JGI genome portal for “SLC26 family” keyword and used those sequences as the sequence queries and included them in the phylogeny (Armbrust et al. 2004).

For genes with a potential role in C4 metabolism, the *T. pseudonana* JGI genome portal was queried for genes annotated as having a key role in C4 metabolism (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014; Armbrust et al. 2004).

All alignments were done in MAFFT (version 7) using the L-INS-i strategy with the BLOSUM62 scoring matrix, the “leave gappy regions” option selected, and the MAFFT-homologs option selected for the bicarbonate transporter proteins (Katoh & Standley 2013). FASTA alignments were converted into PHYLIP format and imported into RaxML (Felsenstein 1989; Stamatakis 2014). Maximum likelihood trees were computed with the following parameters: PROTGAMMA matrix, BLOSUM62 substitution model, and 1000 bootstrap replicates. The carbonic anhydrase tree was rooted using the gamma carbonic anhydrases because they are the most ancient carbonic anhydrase (Smith et al. 1999). The bicarbonate transporter tree was mid-point rooted in FigTree.

### **Predicted Protein Localization**

SignalP (V4.1), TargetP (V1.1), and ASA-find (1.1.5) were used to determine the putative localization of differentially expressed genes and genes potentially involved in a carbon concentrating mechanism (Bendtsen et al. 2004; Emanuelsson et al. 2007; Gruber et al. 2015). TMHMM version 2.0 was used for prediction of transmembrane domains (Krogh et al. 2001). SignalP was run against the eukaryote organism group and the default cutoff values were used. TargetP was run with the plant model and “winner-takes-all” option selected.

The input for these analyses were the translated protein sequences from the transcriptome assemblies as found by ORF-Predictor and reported elsewhere (Min, Butler, Storms, & Tsang, 2005, Wallace et al., in prep).

The output of these predictions were interpreted as follows: only proteins predicted to be plastid targeted by ASA-find (low or high confidence) were considered

imported across all four plastid membranes and into the stroma. If a protein was not plastid targeted but was SignalP positive, it was considered likely to be targeted across the chloroplast-ER (CER) membrane and retained in the CER compartment. If a SignalP negative protein was predicted by TargetP to be targeted to the mitochondria then that protein was considered targeted to the mitochondria (Kroth et al. 2008). Proteins that were not predicted to have a signal cleavage site and were not predicted to be localized to the mitochondria were considered cytosolic. If the sequence was predicted to be SignalP negative, have a transmembrane domain in TMHMM, and secreted as per TargetP, the localization for this sequence is considered potentially extracellular.

## RESULTS

The three *Thalassiosira* diatoms studied here show very different physiological responses to changes in CO<sub>2</sub> (King et al. 2015). Transcriptomes sequenced from the same incubations highlighted the different underlying transcriptional responses. *T. oceanica*, *T. weissflogii*, and *T. rotula* differentially express different numbers of genes across CO<sub>2</sub> conditions (relative to the present-day approximation of 320-390 ppm pCO<sub>2</sub>) (Table 1). *T. oceanica* shows a relatively muted response in global gene expression in <230 ppm and 690-1340 ppm pCO<sub>2</sub> conditions compared to the 2900-5100 ppm pCO<sub>2</sub> condition (Table 1). The larger shift in expression in high pCO<sub>2</sub> coincides with a decreased specific growth rate in 690-1340 and 2900-5100 ppm pCO<sub>2</sub> that is not reflected in any changes in the molecular ratios of C:N:P across pCO<sub>2</sub> (King et al. 2015). *T. weissflogii* shows a more muted response in the 690-1340 ppm pCO<sub>2</sub>, differentially expressing only 156 genes (Table 1). This is in contrast to the <230 ppm

and 2900-5100 ppm CO<sub>2</sub> conditions, where there is differential expression of 581 and 434 genes, respectively (Table 1). This corresponds to a decreased specific growth rate of *T. weissflogii* in <230 ppm pCO<sub>2</sub> that is not reflected in any changes in the molecular ratios of C:N:P, which do not change across the pCO<sub>2</sub> conditions tested (King et al 2015). The specific growth rate of *T. rotula* increases in 690-1320 ppm pCO<sub>2</sub> relative to the specific growth rates in <230 and 320-390 ppm pCO<sub>2</sub> (King et al. 2015). There is also an increase in the C:P ratio in <230 ppm pCO<sub>2</sub> (King et al. 2015). This corresponds to a stronger differential expression signal in the <230 ppm pCO<sub>2</sub> condition, where 988 genes are differentially expressed (Table 1). This is in contrast with the 690-1340 ppm pCO<sub>2</sub> condition, where only 62 genes are differentially expressed (Table 1). There were no 2900-5100 ppm CO<sub>2</sub> cultures, as the high CO<sub>2</sub> and low cell density could not be maintained simultaneously (King et al. 2015).

### **Species-specific metabolic rearrangement, redox and ion homeostasis in low CO<sub>2</sub>**

*T. rotula* grown in low pCO<sub>2</sub> (<230 ppm) increases the expression of degradative genes (Supplemental Table 1), including fatty acid and branched chain amino acid degradation (Figure 1). There also is a low CO<sub>2</sub> downregulation of biosynthetic reactions, chromosomes, translation machinery, and cell cycle progression (Supplemental Table 1).

*T. weissflogii* grown in low pCO<sub>2</sub> decreases the expression of gluconeogenesis and pentose phosphate pathway genes (Figure 1). There also is a low CO<sub>2</sub> downregulation of genes involved in amino acid, fatty acid, and tRNA biosynthesis (Supplemental Table 1) and an increase the expression of fatty acid and branched-

chain amino acid degradation genes (although not as many genes as *T. rotula*) (Supplemental Table 1).

The trends in *T. rotula* and *T. weissflogii* are in contrast to what is observed in *T. oceanica*, where there is not a clear reorganization of metabolism with changing pCO<sub>2</sub>. *T. oceanica* does not increase the expression of genes involved in degradative processes in low CO<sub>2</sub>, but increased the expression of a single gene of unknown function (Supplemental Table 1). Similarly, *T. oceanica* does not increase the expression of genes involved in biosynthetic pathways in high CO<sub>2</sub> or decrease their expression in low CO<sub>2</sub> (Supplemental Table 1). However, in high CO<sub>2</sub>, *T. oceanica* upregulates genes involved in transporting protons and cations as well as redox homeostasis genes (Figure 1).

### **Bicarbonate transporters**

#### **Copy number variation in bicarbonate transporters**

The 3 *Thalassiosira* diatom species considered in this study have different numbers of bicarbonate transporters. *T. pseudonana* has 2 SLC4 transporters (Armbrust et al. 2004; Nakajima et al. 2013), *T. weissflogii* is predicted to have 3, *T. rotula* 4, and *T. oceanica* 5 (Figure 2, Supplemental Table 2). *T. weissflogii* is predicted to have 6 SLC26 transporters, *T. rotula* 6, *T. oceanica* 4, and we recovered 9 from the *T. pseudonana* genome (Figure 2, Supplemental Table 2) (Armbrust et al. 2004). The number of bicarbonate transporter sequences recovered from our *T. oceanica* transcriptome are in agreement with what is observed in the *T. oceanica* genome (Lommer et al. 2012) (Supplemental Table 3).

### **Varied bicarbonate transporter localization predictions and CO<sub>2</sub> sensitivity**

The 3 *Thalassiosira* diatom species in this study show distinct potential differences in localization of their bicarbonate transporters and expression patterns. All of the putative SLC4 and SLC26 transporters recovered here have at least 1 transmembrane helix, suggesting that all are localized to a membrane (Supplemental Table 2). *T. rotula* and *T. oceanica* each have 1 plastid predicted SLC4 transporter, 1 mitochondrial predicted SLC4 transporter, and 1 mitochondrial predicted SLC26 transporter (Supplemental Table 2). In <230 ppm pCO<sub>2</sub>, *T. rotula* downregulates a cytosolic membrane predicted SLC4 transporter and upregulates a plastid predicted SLC4 transporter but does not differentially express SLC26 transporters across CO<sub>2</sub> conditions (Figure 3, Supplemental Table 2). *T. oceanica* downregulates 2 cytosolic membrane predicted SLC4 transporters in 2900-5100 ppm pCO<sub>2</sub> but does not differentially express SLC26 transporters across CO<sub>2</sub> conditions (Figure 3, Supplemental Table 2). The *T. rotula* and *T. oceanica* results are consistent with results from *T. pseudonana*, where an SLC4 transporter in the cytosolic membrane was downregulated in high CO<sub>2</sub> (Hennon et al. 2015). In contrast, most of the SLC4 and SLC26 transporters in *T. weissflogii* are predicted to be in the cytosolic membrane, save for 1 plastid predicted SLC26 transporter (Supplemental Table 2). In <230 ppm pCO<sub>2</sub>, *T. weissflogii* downregulates a plastid predicted SLC26 but does not differentially express SLC4 transporters across CO<sub>2</sub> conditions (Figure 3, Supplemental Table 2). Interestingly, none of the SLC4 or SLC26 transporters from *T. weissflogii* are predicted to be localized to the mitochondria (Supplemental Table 2).

## Carbonic Anhydrases

### Copy number variation in carbonic anhydrases

The copy numbers and isoform distributions of the putative carbonic anhydrases from the assembled *Thalassiosira* transcriptomes vary across species (Table 2). The *T. weissflogii* transcriptome has the largest number of putative carbonic anhydrases (19), while the *T. oceanica* and *T. rotula* transcriptomes have 14 and 13, respectively, which is on par with the 14 putative carbonic anhydrases found in the *T. pseudonana* genome (Table 2) (Armbrust et al. 2004; Tachibana et al. 2011; Samukawa et al. 2014; Kikutani et al. 2016). *T. weissflogii* appears to be particularly enriched in its number of  $\zeta$  carbonic anhydrases (4 copies), compared with the single copies found in *T. pseudonana*, *T. oceanica*, and *T. rotula*, while this class of carbonic anhydrase is absent entirely from *P. tricornutum* (Table 2) (Tachibana et al. 2011; Samukawa et al. 2014). Three of the *T. weissflogii*  $\zeta$  carbonic anhydrases are in the same clade as the other diatoms, while the fourth *T. weissflogii*  $\zeta$  carbonic anhydrase is basal to all the other  $\zeta$  carbonic anhydrases (Figure 4). The *T. weissflogii* transcriptome also has a higher number of  $\alpha$  carbonic anhydrases – 7 copies, compared to 2 in *T. rotula*, 3 in *T. oceanica* and *T. pseudonana*, and 5 in *P. tricornutum* (Table 2) (Tachibana et al. 2011; Samukawa et al. 2014). The *Thalassiosira*  $\alpha$  carbonic anhydrases are in clade separate from those in *P. tricornutum* (Figure 4). The *T. weissflogii*  $\beta$  carbonic anhydrase copy groups closely with the *P. tricornutum* copies (support value of 98) and the *T. rotula* copy also clusters with this group although with poor support values (support value of 23) (Figure 4). There is no evidence of  $\theta$ ,  $\eta$ , or  $\varepsilon$

carbonic anhydrases in our assembled transcriptomes (Figure 4) (Kikutani et al. 2016; Del Prete et al. 2014; So et al. 2004).

The number of carbonic anhydrases recovered from the *T. oceanica* transcriptome are generally supported by the genomic evidence, with the exception of the  $\delta$  carbonic anhydrases (Supplemental Table 3) (Lommer et al. 2012). Each of the 3  $\alpha$  carbonic can be aligned to the genome, although 1 copy did not have a gene prediction overlapping at its mapping location (Supplemental Table 3). The 5  $\gamma$  carbonic anhydrases and the single  $\zeta$  carbonic anhydrase also map to distinct genomic locations and are supported by the gene model predictions (Supplemental Table 3). This congruence starts to break down when comparing the genome and transcriptome  $\delta$  carbonic anhydrases -- 2 of the transcripts are potentially distinct alleles of the same gene, as they map to the same genetic location and their sequences differ only by 3 nucleotides (Supplemental Table 3). Additionally, 2 of the transcripts both map equally well to the same 3 genomic locations, although in each of these scenarios the transcript sequence only covers part of the predicted gene model (Supplemental Table 3).

### **Varied carbonic anhydrase localization predictions**

Each diatom also has a unique pattern of predicted localization for their particular suite of carbonic anhydrases (Figure 5, Supplemental Table 2). *T. oceanica* does not have any carbonic anhydrases predicted to be targeted to the plastid or the chloroplast endoplasmic reticulum but has 5 mitochondrion predicted carbonic anhydrases (two  $\delta$  and three  $\gamma$ ) and the remainder are predicted to be cytoplasmic (Figure 5, Supplemental Table 2). *T. rotula* has a single plastid predicted  $\alpha$  carbonic

anhydrase, 1 mitochondrion predicted  $\gamma$  carbonic anhydrase, and the remainder are predicted to be cytoplasmic (Figure 5, Supplemental Table 2). *T. weissflogii* has 2 chloroplast endoplasmic reticulum predicted  $\alpha$  carbonic anhydrases, 3 mitochondrion predicted carbonic anhydrases (2 $\gamma$  and 1 $\alpha$ ), 1 extracellular predicted  $\delta$  carbonic anhydrase, and the remainder are cytoplasmic (Figure 5, Supplemental Table 2).

### **Carbonic anhydrase CO<sub>2</sub> sensitivity**

The putative  $\beta$ ,  $\delta$ , and  $\zeta$  carbonic anhydrases in *T. weissflogii* are generally the most CO<sub>2</sub> sensitive (Figure 3, Supplemental Table 2). In <230 ppm pCO<sub>2</sub>, *T. weissflogii* upregulates 4  $\zeta$  carbonic anhydrases, 1  $\delta$  carbonic anhydrase, and 1  $\beta$  carbonic anhydrase (Figure 3, Supplemental Table 2). The  $\delta$  carbonic anhydrase is also downregulated in both 690-1340 ppm and 2900-5100 ppm pCO<sub>2</sub>. There is an additional  $\delta$  carbonic anhydrase downregulated in both 690-1340 ppm and 2900-5100 ppm pCO<sub>2</sub> but not upregulated in <230 ppm pCO<sub>2</sub> (Figure 3, Supplemental Table 2). In 2900-5100 ppm pCO<sub>2</sub> there also is downregulation of 1 of the  $\zeta$  carbonic anhydrases that is upregulated in <230 ppm pCO<sub>2</sub> (Figure 3, Supplemental Table 2).

The putative carbonic anhydrases in *T. oceanica* are not differentially expressed in <230 or 690-1340 ppm pCO<sub>2</sub> but there is a downregulation of 1  $\alpha$  and 2  $\delta$  carbonic anhydrases in 2900-5100 ppm pCO<sub>2</sub> (Figure 3, Supplemental Table 2). However, there also is an upregulation of the other 3  $\delta$  carbonic anhydrases in 2900-5100 ppm pCO<sub>2</sub> (Figure 3, Supplemental Table 2). None of the putative carbonic anhydrases in *T. rotula* are differentially expressed in any CO<sub>2</sub> condition tested (Figure 3, Supplemental Table 2).

### **Putative C4 photosynthesis genes**

Each of the three diatom transcriptomes contain PEPC, PEPCK, ME, and PYC -- the key carboxylating and decarboxylating enzymes hypothesized to be critical for diatom C4 metabolism (Figure 6, Supplemental Table 2) (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014). The localization and pCO<sub>2</sub> sensitivity of these enzymes varies across species, but there is little compelling evidence that C4 enzymes are playing an important role in the CCM of these *Thalassiosiroid* diatoms (Supplemental Table 2). The putative C4 metabolism genes recovered from the *T. oceanica* transcriptome are supported by the genomic sequences (Lommer et al. 2012) (Supplemental Table 3).

If PEPC mediated carboxylation is important for a functional CCM, PEPC should be localized outside the plastid and be upregulated in low pCO<sub>2</sub> or downregulated in high CO<sub>2</sub> (Figure 6). Consistent with the suggestion that PEPC plays a role in the diatom CCM, *T. oceanica* downregulates a putative cytosolic PEPC in 2900-5100 ppm pCO<sub>2</sub> (Supplemental Table 2). Similarly, in <230 ppm pCO<sub>2</sub> *T. weissflogii* upregulates 1 putative mitochondrial PEPC and 1 putatively cytosolic PEPC (which is also downregulated in 2900-5100 ppm pCO<sub>2</sub>) (Supplemental Table 2). However, *T. weissflogii* also upregulates a PEPC that is predicted to be plastid localized in <230 ppm pCO<sub>2</sub> (Supplemental Table 2). Although there is a *T. rotula* PEPC that is predicted to be localized to the mitochondria which consistent with a potential role in a C4 CCM, it is not differentially expressed in any CO<sub>2</sub> condition tested here (Supplemental Table 2).

If PEPCK mediated decarboxylation is important for a functional CCM, PEPCK should be plastid localized and be upregulated in low pCO<sub>2</sub> or downregulated in high CO<sub>2</sub> (Figure 6). The *T. oceanica* PEPCK is mitochondrion-predicted and is downregulated in 690-1340 ppm pCO<sub>2</sub> but then upregulated in the 2900-5100 ppm pCO<sub>2</sub> condition, suggesting it is not involved in decarboxylation steps of the CCM (Supplemental Table 2). The *T. weissflogii* PEPCK is upregulated in 2900-5100 ppm pCO<sub>2</sub> but it is also mitochondrion predicted, suggesting it is also not involved in decarboxylation steps of the CCM (Supplemental Table 2). The *T. rotula* PEPCK is upregulated in <230 ppm pCO<sub>2</sub>, but it is predicted to be cytosolic, suggesting it is also not involved in decarboxylation steps of the CCM (Supplemental Table 2).

If ME mediated decarboxylation is important for a functional CCM, ME should be plastid localized and be upregulated in low pCO<sub>2</sub> or downregulated in high CO<sub>2</sub>. *T. oceanica* downregulates a putative cytosolic ME in 2900-5100 ppm pCO<sub>2</sub>, suggesting it is not involved in the CCM because of its predicted localization (Supplemental Table 2). *T. weissflogii* downregulates a putative mitochondrial ME in <230 ppm pCO<sub>2</sub>, suggesting it is also not involved in the CCM (Supplemental Table 2). *T. rotula* does not differentially express ME and ME is not predicted to be plastid localized in this diatom (Supplemental Table 2), suggesting that ME decarboxylation is also not important for a functional CCM in this diatom.

If PYC mediated decarboxylation is important for a functional CCM, PYC should be plastid localized and be upregulated in low pCO<sub>2</sub> or downregulated in high CO<sub>2</sub>. *T. weissflogii* downregulates a PYC in <230 ppm pCO<sub>2</sub> and downregulates another in 2900-5100 ppm pCO<sub>2</sub> and both copies are predicted to be cytosolic

(Supplemental Table 2). *T. oceanica* does not differentially express PYC and PYC is not plastid localized (Supplemental Table 2), suggesting PYC is not a key aspect of the CCM of this diatom. Although the PYC in *T. rotula* is not differentially expressed, it is plastid localized, so its potential role in the CCM is ambiguous (Supplemental Table 2).

### **cAMP secondary messengers and CO<sub>2</sub>/HCO<sub>3</sub><sup>-</sup> sensing**

Like *T. pseudonana*, *T. oceanica* upregulates adenylyl cyclase (IPR001054) in high CO<sub>2</sub> but also downregulates 6 additional adenylyl cyclases (Supplemental Table 1) (Hennon et al. 2015). Querying the NCBI NR database with the upregulated adenylyl cyclase indicates that the top blast hit (after the *T. oceanica* genome accession EJK45951.1) is the soluble adenylyl cyclase that has been suggested to play a role in CO<sub>2</sub> sensing in *T. pseudonana* (XP\_002288198.1) (Hennon et al. 2015).

Unlike *T. oceanica* and *T. pseudonana*, *T. rotula* and *T. weissflogii* do not differentially express adenylyl cyclases in response to pCO<sub>2</sub> (Supplemental Table 1).

In contrast to the *T. pseudonana* study, but consistent with maintaining high intracellular cAMP levels in high CO<sub>2</sub>, *T. oceanica* downregulates a phosphodiesterase in 2900-5100 ppm pCO<sub>2</sub> (Supplemental Table 1) (Hennon et al. 2015). *T. oceanica* also downregulates 3 bZIP domain transcription factors (IPR004827) in 2900-5100 ppm pCO<sub>2</sub>, consistent with the suggestion that cAMP-responsive transcription factors downregulate clusters of genes involved in the CCM (including their own expression) in high CO<sub>2</sub> (Supplemental Table 1) (Hennon et al. 2015). There is also a single cAMP-dependent kinase downregulated by *T. oceanica* in 2900-5100 ppm pCO<sub>2</sub>, but there are 3 additional cAMP-dependent kinases that are

upregulated as well (Supplemental Table 1). *T. weissflogii* downregulates a phosphodiesterase in low pCO<sub>2</sub> (<230 ppm pCO<sub>2</sub>), which is consistent with the Hennon et al. 2015 study that saw an upregulation of phosphodiesterases in *T. pseudonana* in elevated CO<sub>2</sub> (Supplemental Table 1) (Hennon et al. 2015) (Supplemental Table 1). In contrast to *T. pseudonana* (where cAMP-responsive transcription factors downregulate clusters of genes involved in the CCM, including their own expression in high CO<sub>2</sub> (Hennon et al. 2015)), *T. rotula* and *T. weissflogii* downregulate bZIP domain transcription factor and upregulate cAMP-dependent protein kinases in low CO<sub>2</sub> (<230 ppm pCO<sub>2</sub>) (Supplemental Table 1). This suggests that bZIP domain transcription factors are not increasing the expression of CCM gene clusters (including their own expression) in *T. rotula* and *T. weissflogii* in low CO<sub>2</sub>. This is supported by the increase of cAMP-dependent kinases in *T. rotula* and *T. weissflogii* in low CO<sub>2</sub> (Supplemental Table 1), as it was suggested that cAMP-responsive elements would upregulate the CCM in low CO<sub>2</sub>, including their own expression (Hennon et al. 2015).

## **DISCUSSION**

### **Key point of regulation for *T. oceanica* carbon concentrating mechanism: SLC4-mediated HCO<sub>3</sub><sup>-</sup> uptake into the cytoplasm**

The role of carbonic anhydrases in the CCM of *T. oceanica* is unclear, as these enzymes are up and downregulated the 2900-5100 ppm pCO<sub>2</sub> condition, which approximates the highest levels diatoms have experienced over the course of their evolution (Figure 3, Supplemental Table 2) (Sims et al. 2006; Medlin 2016; Kump et al. 2009). *T. pseudonana* also upregulates carbonic anhydrases in high CO<sub>2</sub> (Hennon et

al. 2015), but the high CO<sub>2</sub> condition used in that study (~800 ppm) is considerably lower than the high CO<sub>2</sub> conditions where *T. oceanica* differentially expresses bicarbonate transporters or carbonic anhydrases (2900-5100 ppm) and *T. oceanica* doesn't differentially express bicarbonate transporters or carbonic anhydrases in the 690-1340 ppm CO<sub>2</sub> condition that would be the closest approximation to the Hennon 2015 study (Figure 3, Supplemental Table 2) (Hennon et al. 2015).

The incongruence between the  $\delta$  carbonic anhydrases recovered from the genome (Lommer et al. 2012) and transcriptome of *T. oceanica* may be explained by the observation that the  $\delta$  carbonic anhydrases have undergone relatively recent, lineage-specific gene expansions (Shen et al. 2017). These recent gene duplication events may lead to species-specific gene expansions that may not have had time to diverge and would encode very similar transcripts that would likely be merged in *de novo* assemblies.

Consistent with a potential role in C<sub>4</sub>-based CCM, *T. oceanica* decreases the expression of a cytosolic PEPC gene in high CO<sub>2</sub> (Supplemental Table 2, Figure 6) (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014). However, the potential decarboxylating enzymes' (PEPCK, ME, PYC) localization (ME and PYC) and/or expression (PEPCK and PYC) suggests that C<sub>4</sub> metabolism is not part of the CCM of *T. oceanica* (Supplemental Table 2, Figure 6) (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014).

The high CO<sub>2</sub> downregulation of 2 cytosolic membrane SLC4 transporters suggests that exogenous bicarbonate uptake is the key point of regulation for *T. oceanica*'s CCM (Figure 3, Supplemental Table 2). This is supported by the higher

number of SLC4 transporters found in *T. oceanica* (5) relative to the 3 found in *T. weissflogii* (this study) and *T. pseudonana* (Armbrust et al. 2004; Nakajima et al. 2013) and the 4 found in *T. rotula* (this study).

***T. oceanica* decreased growth rate in high CO<sub>2</sub> due to ionic and redox stress**

We hypothesize that *T. oceanica*'s decreased growth rate in high CO<sub>2</sub> is related to oxidative stress and perturbed intracellular ionic balance (Figure 1, Supplemental Table 1) (King et al. 2015).

Ionic stress in *T. oceanica* grown in high CO<sub>2</sub> may be explained by the downregulation of SLC4 transporters in high CO<sub>2</sub> (Figure 3, Supplemental Table 2). The SLC4 transporters in diatoms require Na<sup>+</sup> ions and are thought to be either Na<sup>+</sup> co-transporters or Na<sup>+</sup> driven Cl<sup>-</sup>/HCO<sub>3</sub><sup>-</sup> exchangers (Nakajima et al. 2013). Their activity requires the movement of Na<sup>+</sup> back out of the cell and/or Cl<sup>-</sup> back into the cell to maintain ionic gradients across cellular membranes (Nakajima et al. 2013). The downregulation of SLC4 transporters in high CO<sub>2</sub> may perturb the intracellular balance of Na<sup>+</sup> and/or Cl<sup>-</sup> in *T. oceanica*.

The redox stress observed in *T. oceanica* may be related to the carbonic anhydrases, as the carbonic anhydrases in *P. tricornutum* are partly redox regulated via the activity of thioredoxin (Kikutani et al. 2012). *T. oceanica*'s regulation of redox enzymes in high pCO<sub>2</sub> may indicate that this diatom uses similar redox regulation of carbonic anhydrase activity (Figure 1, Supplemental Table 1).

**Key point of regulation for *T. weissflogii* carbon concentrating mechanism:  
cytosolic-predicted  $\beta$ ,  $\delta$ , and  $\zeta$  carbonic anhydrases preventing CO<sub>2</sub> leakage**

In terms of the carbon concentrating mechanism, there is a tight regulation of cytosolic-predicted  $\beta$ ,  $\delta$ , and  $\zeta$  carbonic anhydrases across CO<sub>2</sub> conditions (Figure 3, Figure 5, Supplemental Table 2). This is consistent with studies in *T. pseudonana* and *P. tricornutum* that suggest the  $\beta$ ,  $\delta$ , and  $\zeta$  classes of carbonic anhydrases are upregulated in low pCO<sub>2</sub> and/or downregulated in high pCO<sub>2</sub> (Tachibana et al. 2011; Samukawa et al. 2014; Clement et al. 2017; Hennon et al. 2015; Crawford et al. 2011; McGinn & Morel 2008).

None of the *T. weissflogii* SLC4 or SLC26 bicarbonate transporters are downregulated in 690-1340 or 2900-5100 or upregulated in <230 ppm pCO<sub>2</sub> (Figure 3, Supplemental Table 2). This is consistent with proteomics studies in *T. pseudonana* that did not observe downregulation of SLC4 transporters in high pCO<sub>2</sub> conditions (Clement et al. 2017), but is in contrast with a transcriptome study where *T. pseudonana* downregulated an SLC4 transporter in high pCO<sub>2</sub> (Hennon et al. 2015). There is one SLC26 transporter predicted to be localized to the plastid that is downregulated in <230 ppm pCO<sub>2</sub> in *T. weissflogii* (Figure 3, Supplemental Table 2), but it is possible that this transporter is involved in transporting other anions (e.g., sulfate) (Alper & Sharma 2013).

Consistent with a potential role in a C<sub>4</sub>-based CCM, *T. weissflogii* decreases the expression of a cytosolic PEPC gene in high CO<sub>2</sub> and increases its expression in low CO<sub>2</sub> (Supplemental Table 2, Figure 6) (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014). However, there also is upregulation of a plastid predicted PEPC in low

CO<sub>2</sub>, which would potentially compete with RuBisCO for CO<sub>2</sub> and is in opposition to what is expected of a carbon concentrating mechanism (Supplemental Table 2, Figure 6) (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014). PEPC or ME mediated-decarboxylation is unlikely in *T. weissflogii* because both genes are predicted to be mitochondrial rather than plastid localized (Supplemental Table 2, Figure 6, (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014)). Further, they are upregulated in 2900-5100 ppm pCO<sub>2</sub> and downregulated in <230 ppm pCO<sub>2</sub>, which is counter to what would be expected of the CCM (Supplemental Table 2, Figure 6) (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014). Although there is a *T. weissflogii* PYC gene that is downregulated in 2900-5100 ppm pCO<sub>2</sub>, its predicted localization to the cytosol suggest that it is not playing a role as a C4 decarboxylase, which would require localization to the plastid (Supplemental Table 2, Figure 6) (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014).

Their predicted localization to the cytoplasm (Figure 3, Figure 5, Supplemental Table 2) suggests that similar to *T. oceanica*, the key point of regulation for the *T. weissflogii* carbon concentrating mechanism is the flux of inorganic through the cytoplasm. While this is mediated by SLC4 HCO<sub>3</sub><sup>-</sup> transporters in *T. oceanica*, *T. weissflogii* seems to focus more on diffusive CO<sub>2</sub> uptake and using carbonic anhydrases to prevent CO<sub>2</sub> leakage out of the cell.

The presence of both β and ζ carbonic anhydrases in *T. weissflogii* is interesting, as neither of the two model diatoms examined thus far (*T. pseudonana* and *P. tricornutum*) have both the β and ζ carbonic anhydrases – *T. pseudonana* has a ζ but not a β, and the reverse is true for *P. tricornutum* (Tachibana et al. 2011; Samukawa et

al. 2014) (Table 2). It has also been noted that the  $\beta$  and  $\zeta$  carbonic anhydrases are found in relatively few diatom transcriptomes compared to the more commonly found  $\alpha$ ,  $\delta$ , and  $\gamma$  carbonic anhydrases, highlighting the rarity of the scenario where a diatom simultaneously has both of these carbonic anhydrases (Shen et al. 2017). However, it is possible that  $\beta$  and  $\zeta$  carbonic anhydrase are more widely distributed than it would seem from available transcriptome data if their transcripts are too lowly expressed to be recovered by assemblies that do not include a low pCO<sub>2</sub> condition which may be required to induce their expression.

### ***T. weissflogii* energetic trade-offs across CO<sub>2</sub> conditions**

We hypothesize that the energetic cost of upregulating the carbonic anhydrases may be the cause of the decreased specific growth rate in <230 ppm pCO<sub>2</sub> (Figure 3, Supplemental Table 2) (Giordano et al. 2005; Raven et al. 2014; King et al. 2015). *T. weissflogii* grown in low pCO<sub>2</sub> also decreases the expression of the pentose phosphate pathway (Figure 1, Supplemental Table 1), fatty acid synthesis (Supplemental Table 1), and glycolysis/gluconeogenesis (Figure 1, Supplemental Table 1) suggesting that in higher pCO<sub>2</sub> it may divert carbon and reducing power toward the synthesis of storage compounds like fatty acids and/or glucose/chrysolaminarin instead of an increased specific growth rate, as seen in *T. rotula* grown in high pCO<sub>2</sub> (King et al. 2015).

### **Key point of regulation for *T. rotula* carbon concentrating mechanism: SLC4-mediated HCO<sub>3</sub><sup>-</sup> uptake into the plastid**

*T. rotula* does not differentially express any carbonic anhydrases in any CO<sub>2</sub> condition tested here (Figure 3, Supplemental Table 2). However, some of the  $\delta$ ,  $\gamma$ ,

and  $\zeta$  carbonic anhydrases are highly expressed and perhaps are not differentially expressed because they are constitutively expressed across CO<sub>2</sub> conditions, as has been observed for a plastid localized  $\alpha$  carbonic anhydrases of *T. pseudonana* (Figure 3, Supplemental Table 2) (Hennon et al. 2015).

*T. rotula* does upregulate a plastid-predicted SLC4 transporter in <230 ppm pCO<sub>2</sub>, but it also downregulates a cytosolic SLC4 transporter in the same condition (Figure 3, Supplemental Table 2). This is in contrast with *T. pseudonana*, which downregulates a cytoplasmic SLC4 gene in high CO<sub>2</sub> and seems to rely on a bestrophin putative ion channel to transport HCO<sub>3</sub><sup>-</sup> into the plastid (Hennon et al. 2015).

There is little evidence of a C4-like CCM in *T. rotula*, which does not differentially express PEPC across CO<sub>2</sub> conditions. Its localization to the mitochondria is also not consistent with a carboxylase role in a C4-based CCM unless it is functioning as a means to recover respired CO<sub>2</sub> (Supplemental Table 2, Figure 6) (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014). The PEPCK, ME, and PYC enzymes that could act as decarboxylases are either not CO<sub>2</sub> sensitive or not localized to the plastid (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014). There is a plastid localized PYC, and while lack of CO<sub>2</sub> sensitivity makes its role in a C4-based CCM ambiguous, its high expression suggests it may be constitutively expressed (Supplemental Table 2, Figure 6) (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014).

These data suggest that *T. rotula* may rely more on SLC4 uptake into the plastid for their CCM rather than uptake into the cytosol or the activity of carbonic

anhydrases, while the potential role of a C4-like CCM in this diatom remains ambiguous.

### ***T. rotula* relies on degradation pathways to sustain growth in low CO<sub>2</sub>**

The upregulation of genes involved in degrading amino and fatty acids in low pCO<sub>2</sub> suggests that *T. rotula* is relying on these catabolic reactions to fuel growth when inorganic carbon is less readily available (Figure 1, Supplemental Table 1). The slower growth rate in 320-390 ppm pCO<sub>2</sub> and <230 ppm pCO<sub>2</sub> might indicate that *T. rotula* is becoming carbon limited in these conditions, but the increased C:P ratio in <230 ppm pCO<sub>2</sub> suggests that *T. rotula* is probably not carbon limited but is potentially phosphate limited in low pCO<sub>2</sub> (King et al. 2015). *T. rotula* does not upregulate alkaline phosphatases or phosphate transporters in low pCO<sub>2</sub>, genes that are indicative of phosphate limitation in *T. pseudonana* (Dyhrman et al. 2012), but there is an upregulation of a serine/threonine protein phosphatase protein that cleaves phosphate groups from serine and threonine residues (Supplemental Table 1). *T. rotula* doesn't increase the expression of genes indicative of increased carbon storage in high CO<sub>2</sub> (Supplemental Table 1) consistent with the suggestion that it funnels excess carbon toward an increased growth rate rather than synthesis of storage compounds as suggested for *T. weissflogii* (King et al. 2015).

### **Evidence for cAMP-mediated CO<sub>2</sub> sensing in *T. oceanica*, but not *T. weissflogii* or *T. rotula***

Evidence suggests diatoms sense external CO<sub>2</sub> levels via the activity of cAMP secondary messengers (Ohno et al. 2012; Hennon et al. 2015). This seems to be true of

*T. oceanica*, but there is little evidence of cAMP mediated CO<sub>2</sub> sensing in *T. rotula* and *T. weissflogii*.

One possible complicating factor is the cAMP-mediated cross-talk between CO<sub>2</sub> and light sensing (*P. tricornutum* diatoms grown in low light are less able to upregulate  $\beta$  carbonic anhydrases in response to low pCO<sub>2</sub> (Tanaka et al. 2016)). The experiments described in this study were conducted in growth-saturating light conditions, so a low-light response is not anticipated here (King et al. 2015). However, the role of cAMP secondary messengers in diatom biology is not completely understood and there may be additional cross-talk of cAMP (or other signaling pathways) that obscure further indication of CO<sub>2</sub> mediated cAMP signaling in *T. rotula* and *T. weissflogii*. An additional possibility is that *T. rotula* and *T. weissflogii* sense changes in external pCO<sub>2</sub> through some other mechanism that does not utilize adenylyl cyclases or cAMP as CO<sub>2</sub> sensors.

## **CONCLUSIONS**

The results of this study suggest that the gene complement, subcellular localization, and CO<sub>2</sub> sensitivity of the CCM varies not only across diatom genera (e.g., *Thalassiosira* versus *Phaeodactylum*) (Tachibana et al. 2011; Samukawa et al. 2014), but can also vary across species within the same diatom genera. This is consistent with the hypothesis that historic fluctuations in atmospheric CO<sub>2</sub> have led to a diversification and re-configuration of the CCM within the diatoms (Young et al. 2016; Shen et al. 2017). This study suggests that this diversification has occurred even in diatoms within the same genus. Additional studies of diatom genomes or low CO<sub>2</sub> transcriptomes will also be important for understanding whether or not the presence of

both  $\beta$  and  $\zeta$  carbonic anhydrases in *T. weissflogii* (Table 2, Figure 5, Supplemental Table 2) is as anomalous as it would seem based on current available diatom genomic and transcriptomic evidence.

Different species' strategies for balancing carbon storage versus carbon spending (e.g., increased growth rate) may influence community composition and a species' ability to cope with additional stressors in a high CO<sub>2</sub> future. For example, a larger internal pool of fatty acids and lipids may help *T. weissflogii* to survive periods of phosphorus limitation in high CO<sub>2</sub> by allowing access to an intracellular pool of phospholipids (Dyhrman et al. 2012). Additional mesocosm experiments applying multiple stressors (e.g., phosphate limitation and elevated CO<sub>2</sub>) to these naturally occurring phytoplankton assemblages may give us a better understanding of how contrasting, species-specific strategies impact competitive dynamics and diatom community composition.

Many of the conclusions about the diatom CCM presented here rely on computational predictions of sub-cellular protein localization. The computational predictions of sub-cellular localization of diatom proteins do not always agree with observed localizations (Samukawa et al. 2014), so the predicted localizations presented here (Figure 5, Supplemental Table 2) should be experimentally validated in future studies. Additionally, the protein products of putative CCM genes requires biochemical characterization to confirm their enzymatic activity.

The hypothesis that *T. oceanica* uses cAMP mediated CO<sub>2</sub> sensing (but *T. rotula* and *T. weissflogii* don't) could be tested by tracking the expression of the CCM genes (carbonic anhydrases and bicarbonate transporters) across a range of CO<sub>2</sub>

conditions with and without the addition of the phosphodiesterase inhibitor 3-isobutyl-1-methylxanthine (Hennon et al. 2015). These studies could be particularly useful if they are combined with measurements of carbon assimilation rates via photosynthetron P versus E cruves (phototosynthesis-irradiance response curves) (Lewis & Smith 1983).

## References

- Alper, S.L. & Sharma, A.K., 2013. The SLC26 Gene Family of Anion Transporters and Channels. *Molecular Aspects of Medicine*, 34(2–3), pp.494–515.
- Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Armbrust, E.V. et al., 2004. The Genome of the Diatom *Thalassiosira pseudonana*: Ecology, Evolution, and Metabolism. *Science*, 306(October), pp.79–86.
- Bendtsen, J.D. et al., 2004. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology*, 340(4), pp.783–795.
- Buck, J. et al., 1999. Cytosolic adenylyl cyclase defines a unique signaling molecule in mammals. *Proceedings of the National Academy of Sciences*, 96(1), pp.79–84.
- Caldeira, K. & Wickett, M.E., 2003. Oceanography: anthropogenic carbon and ocean pH. *Nature*, 425(6956), p.365.
- Clement, R. et al., 2017. Responses of the marine diatom *Thalassiosira pseudonana* to changes in CO<sub>2</sub> concentration : a proteomic approach. *Nature Publishing Group*, (February), pp.1–12.
- Conti, M. & Beavo, J., 2007. Biochemistry and Physiology of Cyclic Nucleotide Phosphodiesterases: Essential Components in Cyclic Nucleotide Signaling. *Annual Review of Biochemistry*, 76(1), pp.481–511.
- Crawford, K.J. et al., 2011. The response of *Thalassiosira pseudonana* to long-term exposure to increased CO<sub>2</sub> and decreased pH. *PLoS ONE*, 6(10), pp.1–9.
- Dyrman, S.T. et al., 2012. The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLoS*

*ONE*, 7(3).

- Emanuelsson, O. et al., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, 2(4), pp.953–971.
- Feely, R. a et al., 2008. Evidence for upwelling of corrosive “acidified” water onto the continental shelf. *Science*, 320(5882), pp.1490–1492.
- Felsenstein, J., 1989. PHYLIP - Phylogeny inference package - v3.2. *Cladistics*, pp.164–166.
- Giordano, M., Beardall, J. & Raven, J.A., 2005. CO<sub>2</sub> CONCENTRATING MECHANISMS IN ALGAE: Mechanisms, Environmental Modulation, and Evolution. *Annual Review of Plant Biology*, 56(1), pp.99–131.
- Gruber, A. et al., 2015. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant Journal*, 81(3), pp.519–528.
- Guillard, R.R.L., 1975. Culture of Phytoplankton for Feeding Marine Invertebrates. In *Culture of Marine Invertebrate Animals*. pp. 29–60.
- Guillard, R.R.L. & Ryther, J.H., 1962. Studies of marine planktonic diatoms I. *Cyclotella nana* Hustedt, and *Detonula confervasea* (Cleve) Gran. *Canadian Journal of Microbiology*, 8(1140), pp.229–239.
- Haimovich-Dayana, M. et al., 2013. The role of C4 metabolism in the marine diatom *Phaeodactylum tricornutum*. *New Phytologist*, 197(1), pp.177–185.
- Harada, H. et al., 2006. CO<sub>2</sub> sensing at ocean surface mediated by cAMP in a marine diatom. *Plant physiology*, 142(3), pp.1318–1328.
- Hennon, G.M.M. et al., 2015. Diatom acclimation to elevated CO<sub>2</sub> via cAMP signalling and coordinated gene expression. *Nature Climate Change*, 5(August),

pp.761–766.

Hewett-Emmett, D. & Tashian, R.E., 1996. Functional diversity, conservation, and convergence in the evolution of the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -carbonic anhydrase gene families. *Molecular Phylogenetics and Evolution*, 5(1), pp.50–77.

Hofmann, G.E. et al., 2011. High-Frequency Dynamics of Ocean pH : A Multi-Ecosystem Comparison. *PLoS ONE*, 6(12).

Hopkinson, B.M., 2014. A chloroplast pump model for the CO<sub>2</sub> concentrating mechanism in the diatom *Phaeodactylum tricornutum*. *Photosynthesis Research*, 121(2–3), pp.223–233.

Hopkinson, B.M. et al., 2011. Efficiency of the CO<sub>2</sub>-concentrating mechanism of diatoms. *Proceedings of the National Academy of Sciences*, 108(10), pp.3830–3837.

Hopkinson, B.M., Dupont, C.L. & Matsuda, Y., 2016. The physiology and genetics of CO<sub>2</sub> concentrating mechanisms in model diatoms. *Current Opinion in Plant Biology*, 31, pp.51–57.

Jones, P. et al., 2014. InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), pp.1236–1240.

Katoh, K. & Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), pp.772–780.

Khaliwala, S. et al., 2013. Global ocean storage of anthropogenic carbon. *Biogeosciences*, 10(4), pp.2169–2191.

Kikutani, S. et al., 2012. Redox regulation of carbonic anhydrases via thioredoxin in

- chloroplast of the marine diatom *Phaeodactylum tricornutum*. *Journal of Biological Chemistry*, 287(24), pp.20689–20700.
- Kikutani, S. et al., 2016. Thylakoid luminal  $\theta$ -carbonic anhydrase critical for growth and photosynthesis in the marine diatom *Phaeodactylum tricornutum*. *Proceedings of the National Academy of Sciences*, 113(35), pp.9828–9833.
- King, A.L. et al., 2015. Effects of CO<sub>2</sub> on growth rate, C:N:P, and fatty acid composition of seven marine phytoplankton species. *Marine Ecology Progress Series*, 537, pp.59–69.
- Krogh, A. et al., 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3), pp.567–580.
- Kroth, P.G. et al., 2008. A Model for Carbohydrate Metabolism in the Diatom *Phaeodactylum tricornutum* Deduced from Comparative Whole Genome Analysis. *PLoS ONE*, 3(1), p.e1426.
- Kump, L.R., Bralowe, T.J. & Ridgwell, A., 2009. Ocean Acidification in Deep Time. *Oceanography*, 22(4), pp.94–107.
- Kustka, A.B. et al., 2014. Low CO<sub>2</sub> results in a rearrangement of carbon metabolism to support C<sub>4</sub> photosynthetic carbon assimilation in *Thalassiosira pseudonana*. *New Phytologist*, 204(3), pp.507–520.
- Lewis, M. & Smith, J., 1983. A small volume, short-incubation-time method for measurement of photosynthesis as a function of incident irradiance. *Marine Ecology Progress Series*, 13, pp.99–102.
- Lommer, M. et al., 2012. Genome and low-iron response of an oceanic diatom adapted

- to chronic iron limitation. *Genome Biology*, 13(7), p.R66.
- Mayr, B. & Montminy, M., 2001. Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nature reviews. Molecular cell biology*, 2(8), pp.599–609.
- McGinn, P.J. & Morel, F.M.M., 2008. Expression and regulation of carbonic anhydrases in the marine diatom *Thalassiosira pseudonana* and in natural phytoplankton assemblages from Great Bay, New Jersey. *Physiologia plantarum*, 133(1), pp.78–91.
- Medlin, L.K., 2016. Evolution of the diatoms: major steps in their evolution and a review of the supporting molecular and morphological evidence. *Phycologia*, 55(1), pp.79–103.
- Min, X.J. et al., 2005. OrfPredictor: Predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research*, 33(SUPPL. 2), pp.677–680.
- Moriya, Y. et al., 2007. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, 35(SUPPL.2), pp.182–185.
- Nakajima, K., Tanaka, A. & Matsuda, Y., 2013. SLC4 family transporters in a marine diatom directly pump bicarbonate from seawater. *Proceedings of the National Academy of Sciences of the United States of America*, 110(5), pp.1767–72.
- Nelson, D.M. et al., 1995. Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles*, 9(3), pp.359–372.
- Ohno, N. et al., 2012. CO<sub>2</sub>-cAMP-Responsive cis-Elements Targeted by a Transcription Factor with CREB/ATF-Like Basic Zipper Domain in the Marine

- Diatom *Phaeodactylum tricornutum*. *Plant Physiology*, 158(1), pp.499–513.
- Del Prete, S. et al., 2014. Discovery of a new family of carbonic anhydrases in the malaria pathogen *Plasmodium falciparum*--the  $\eta$ -carbonic anhydrases. *Bioorganic & medicinal chemistry letters*, 24(18), pp.4389–96.
- Le Quéré, C. et al., 2016. Global Carbon Budget 2016. *Earth System Science Data*, 8(2), pp.605–649.
- Raven, J., 1991. Physiology of inorganic carbon acquisition and implications for resource use efficiency by marine phytoplankton: Relation to increased carbon dioxide and temperature. *Plant, Cell and Environment*, 14(8), pp.779–794.
- Raven, J.A., Beardall, J. & Giordano, M., 2014. Energy costs of carbon dioxide concentrating mechanisms in aquatic organisms. *Photosynthesis Research*, 121(2–3), pp.111–124.
- Reinfelder, J.R., Kraepiel, a M. & Morel, F.M., 2000. Unicellular C4 photosynthesis in a marine diatom. *Nature*, 407(October), pp.996–999.
- Reinfelder, J.R., Milligan, A.J. & Morel, F.M.M., 2004. The Role of the C 4 Pathway in Carbon Accumulation and Fixation in a Marine Diatom. *Plant Physiology*, 135(August), pp.2106–2111.
- Riebesell, U. et al., 2000. Reduced calcification of marine plankton in response to increased atmospheric CO<sub>2</sub>. *Nature*, 407(6802), pp.364–7.
- Roberts, K. et al., 2007. C3 and C4 Pathways of Photosynthetic Carbon Assimilation in Marine Diatoms Are under Genetic, Not Environmental, Control. *Plant Physiology*, 145(1), pp.230–235.
- Sabine, C.L. et al., 2004. The Oceanic Sink for Anthropogenic CO<sub>2</sub>. *Science*,

- 305(5682), pp.367–371.
- Sage, R.F., 2004. The evolution of C4 photosynthesis. *New Phytologist*, 161(2), pp.341–370.
- Samukawa, M. et al., 2014. Localization of putative carbonic anhydrases in the marine diatom, *Thalassiosira pseudonana*. *Photosynthesis Research*, 121(2–3), pp.235–249.
- Scripps, 2017. The Keeling Curve: Carbon Dioxide Measurements at Mauna Loa. *Scripps Institute of Oceanography*.
- Shen, C., Dupont, C.L. & Hopkinson, B.M., 2017. The diversity of carbon dioxide-concentrating mechanisms in marine diatoms as inferred from their genetic content. *Journal of Experimental Botany*.
- Shi, D. et al., 2010. Effect of Ocean Acidification on Iron Availability to Marine Phytoplankton. *Science*, 327(5966), pp.676–679.
- Sims, P.A., Mann, D.G. & Medlin, L.K., 2006. Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia*, 45(4), pp.361–402.
- Smith, K.S. et al., 1999. Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 96(26), pp.15184–9.
- So, A.K.C. et al., 2004. A Novel Evolutionary Lineage of Carbonic Anhydrase ( $\epsilon$  Class) Is a Component of the Carboxysome Shell. *Journal of Bacteriology*, 186(3), pp.623–630.
- Solomon, S. et al., 2007. IPCC 2007 Climate Change 2007. In *The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the*

- Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.
- Soto, A.R. et al., 2006. Identification and preliminary characterization of two cDNAs encoding unique carbonic anhydrases from the marine alga *Emiliana huxleyi*. *Applied and Environmental Microbiology*, 72(8), pp.5500–5511.
- Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), pp.1312–1313.
- Tachibana, M. et al., 2011. Localization of putative carbonic anhydrases in two marine diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. *Photosynthesis Research*, 109(1–3), pp.205–221.
- Tanaka, A. et al., 2016. Light and CO<sub>2</sub> /cAMP Signal Cross Talk on the Promoter Elements of Chloroplastic  $\beta$ -Carbonic Anhydrase Genes in the Marine Diatom *Phaeodactylum tricornutum*. *Plant Physiology*, 170(2), pp.1105–1116.
- Tans, P., 2009. An Accounting of the Observed Increase in Oceanic and Atmospheric CO<sub>2</sub> and the Outlook for the Future. *Oceanography*, 22(4), pp.26–35.
- Wu, Z. et al., 2010. Empirical bayes analysis of sequencing-based transcriptional profiling without replicates. *BMC Bioinformatics*, 11(1), p.564.
- Young, J.N. et al., 2016. Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentrating mechanisms. *Journal of Experimental Botany*, 67(11), pp.3445–3456.

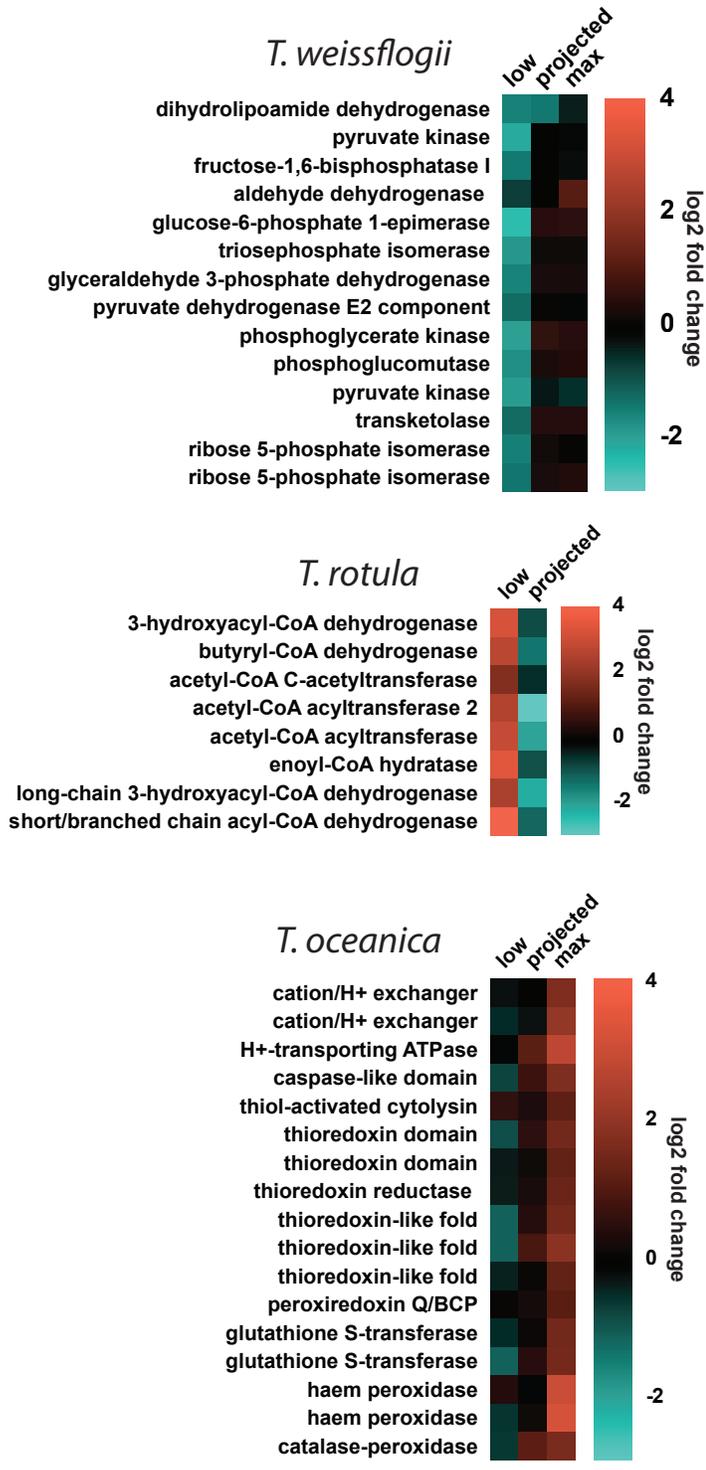


Figure 1. A heat map showing a subset of the differentially expressed genes from *T. weissflogii*, *T. rotula*, and *T. oceanica*. Warmer colors in the heat map indicate an upregulation and a cooler color indicates a downregulation. “Low” refers to the <230 ppm CO<sub>2</sub> condition, “projected” refers to the 690-1340 ppm pCO<sub>2</sub> condition, and “max” refers to the 2900-5100 ppm pCO<sub>2</sub> condition.

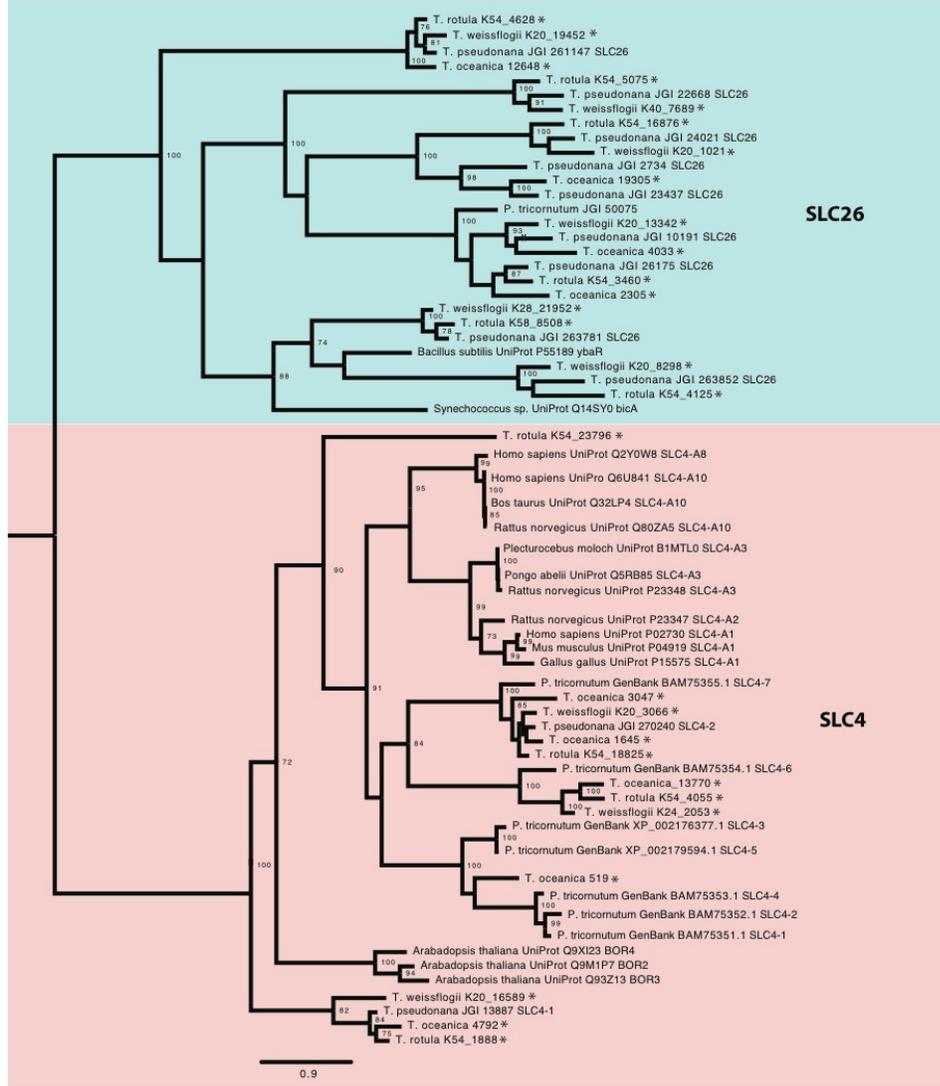


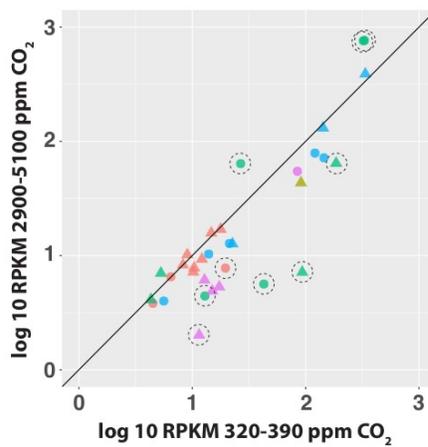
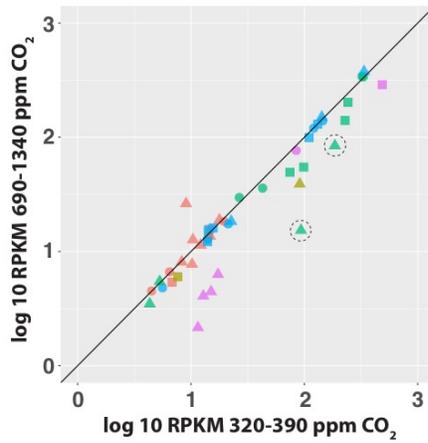
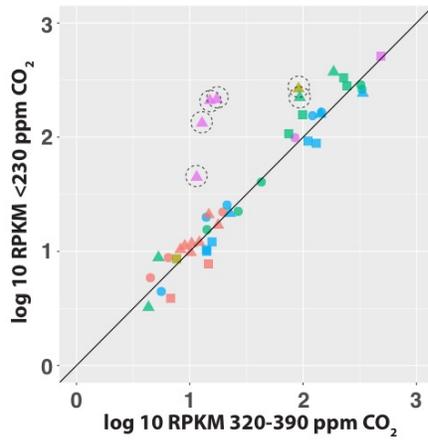
Figure 2. A maximum likelihood phylogeny of the bicarbonate transporters from the model diatoms *T. pseudonana* (JGI Thaps3 gene IDs) and *P. tricornutum* (JGI IDs or NCBI accession numbers), the transcriptomes assembled in this study (indicated by asterisks), and MAFFT homologs included for reference. All alignments were done in MAFFT (version 7) using the L-INS-i strategy with the BLOSUM62 scoring matrix, the “leave gappy regions” option selected, and the MAFFT-homologs option selected (Katoh & Standley 2013). RAxML maximum likelihood trees were computed with the following parameters: PROTGAMMA matrix, BLOSUM62 substitution model, and 1000 bootstrap replicates. The tree was mid-point rooted in FigTree. Support values less than 70 are not shown.

**carbonic anhydrase isozyme**

- alpha
- beta
- delta
- gamma
- zeta

**species**

- oceanica
- ▲ weissflogii
- rotula



**bicarbonate transporter isozyme**

- SLC-4
- SLC-26

**species**

- oceanica
- ▲ weissflogii
- rotula

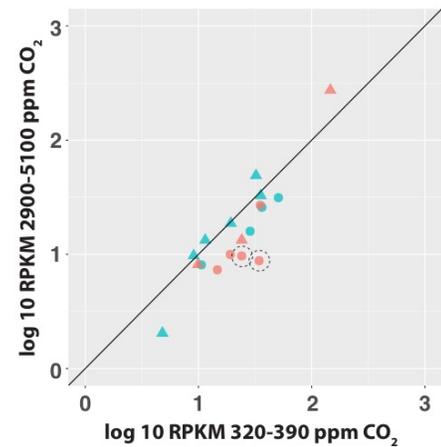
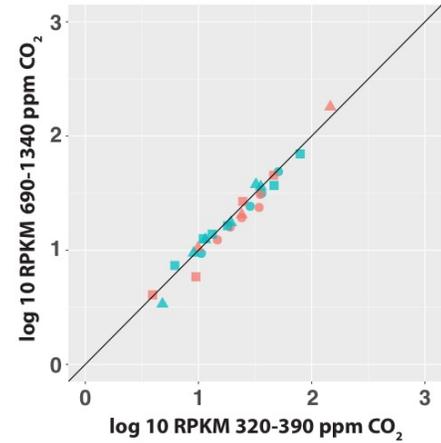
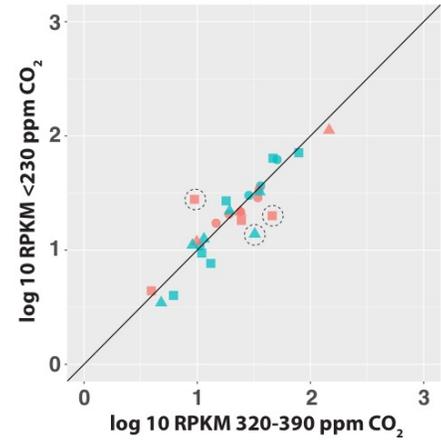


Figure 3. Expression values of the carbonic anhydrases and bicarbonate transporters from the transcriptomes in this study. Each point represents a single transcript. The x-axis is the log 10 transformed RPKM values in the 320-390 ppm pCO<sub>2</sub> and the y-axis is the log 10 transformed RPKM values in the <230 ppm pCO<sub>2</sub> condition, the 690-1340 ppm pCO<sub>2</sub> condition, and the 2900-5100 ppm pCO<sub>2</sub> condition at the bottom of the figure. The black diagonal line indicates the same expression value in the two conditions compared. Circled points appearing above this line are upregulated in the condition indicated on the y-axis and circled points below it are downregulated in the condition indicated on the y-axis ( $\geq 0.95$  posterior probability of a 2-fold change in ASC) (Wu et al. 2010). The color of the points indicates the specific isozyme type and the shape of the point indicates from which organism the transcript originates.

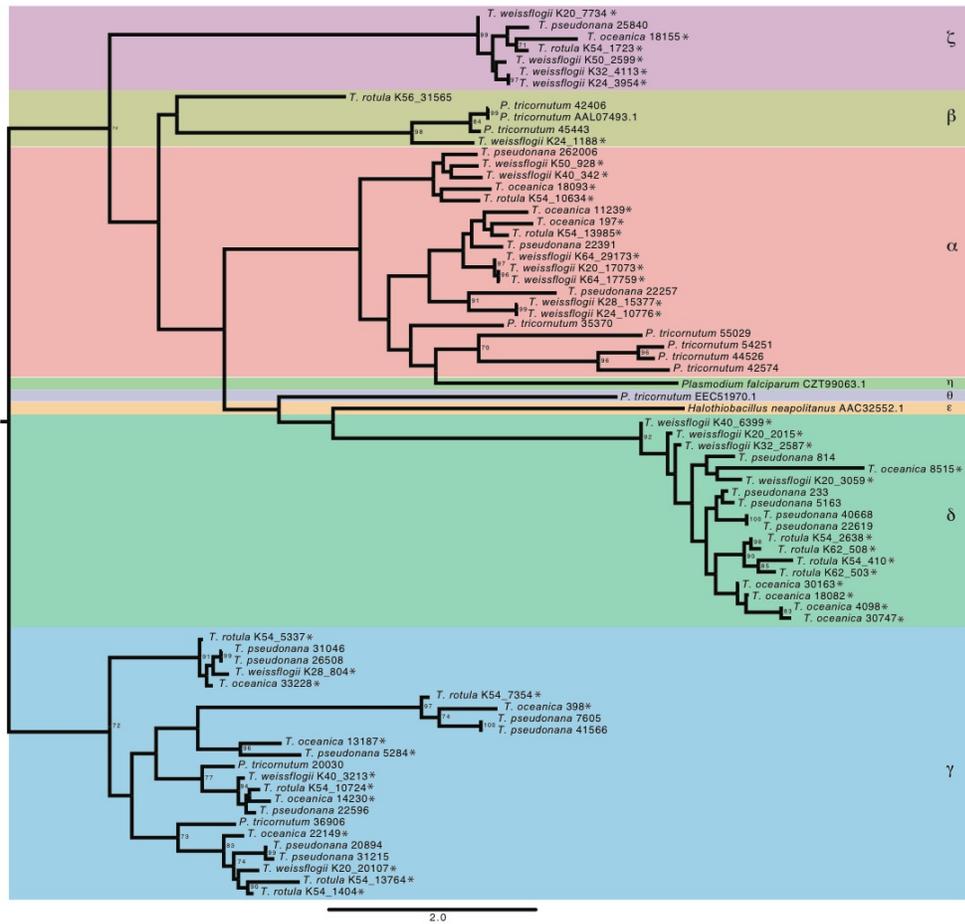


Figure 4. A maximum likelihood phylogeny of the carbonic anhydrases from the model diatoms *T. pseudonana* (JGI Thaps3 gene IDs) and *P. tricornutum* (JGI IDs or NCBI accession numbers), the three *Thalassiosira* transcriptomes assembled in this study (indicated by asterisks), and two additional carbonic anhydrases from *Halothiobacillus neapolitanus* and *Plasmodium falciparum* for reference. All alignments were done in MAFFT (version 7) using the L-INS-i strategy with the BLOSUM62 scoring matrix, the “leave gappy regions” option selected (Kato & Standley 2013). RAxML maximum likelihood trees were computed with the following parameters: PROTGAMMA matrix, BLOSUM62 substitution model, and 1000 bootstrap replicates. The carbonic anhydrase tree was rooted using the gamma carbonic anhydrases because they are the most ancient carbonic anhydrase (Smith et al. 1999). Support values less than 70 are not shown.

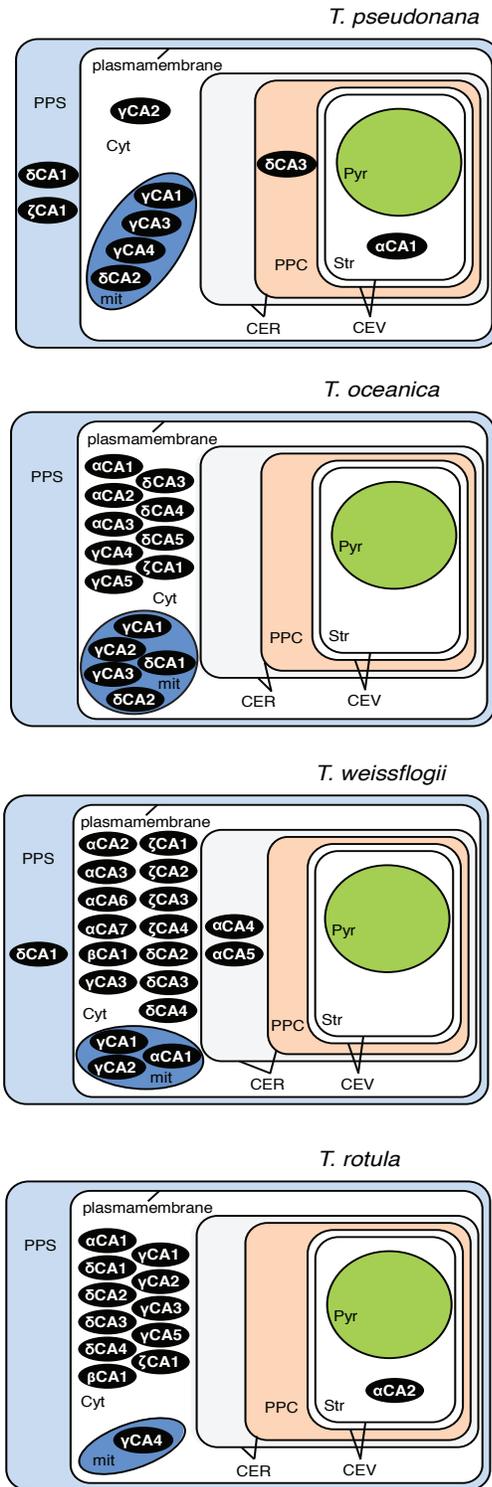


Figure 5. Putative carbonic anhydrases (CAs) that have been localized in the model diatom *T. pseudonana* (from Samukawa et al. 2014) and the predicted localization of putative CAs from *T. oceanica* and *T. weissflogii*. The localizations of the putative CAs reported in this study have not been experimentally verified and are based on the predictions from SignalP (V4.1), TargetP (V1.1), and ASA-find (1.1.5) (Bendtsen et al. 2004; Emanuelsson et al. 2007; Gruber et al. 2015). SignalP was run against the eukaryote organism group and the default cutoff values were used. TargetP was run with the plant model and “winner-takes-all” option selected. PPS = periplastid space, Cyt = cytoplasm, mit = mitochondria, CER = chloroplast endoplasmic reticulum, PPC = periplastidal compartment, CEV = chloroplast envelope, Str = chloroplast stroma, Pyr = pyrenoid.

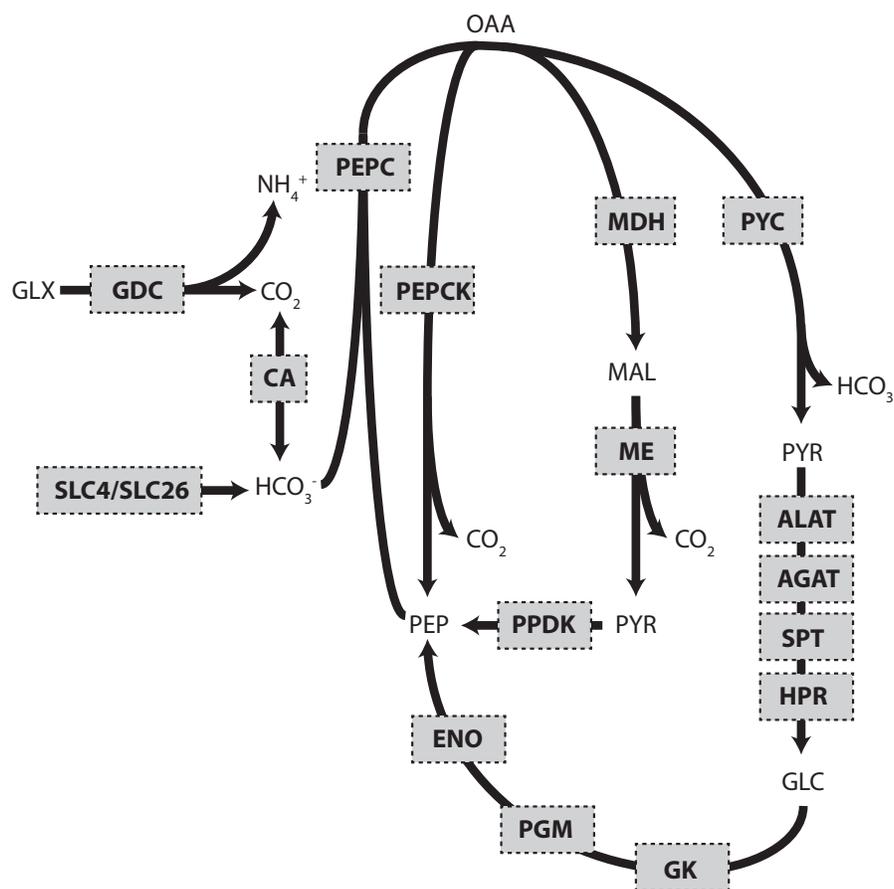


Figure 6. A schematic summarizing the enzymes thought to be involved in the diatom CCM, including variations of C4 metabolism (Reinfelder et al. 2000; Sage 2004; Kustka et al. 2014). Grey boxes indicate enzymes performing the conversions between the metabolites on either end of the black arrow. Grey boxes: AGAT = alanine glyoxylate amino transferase, ALAT = alanine amino transferase, CA = carbonic anhydrase, ENO = enolase, GDC = glycine decarboxylase P-protein, GK = glycerate kinase, HPR = hydroxypyruvate reductase, MDH = malate dehydrogenase, ME = malic enzyme, PEPC = phosphoenolpyruvate carboxylase, PEPCK = phosphoenolpyruvate carboxykinase, PGM = phosphoglycerate mutase, PPDK = pyruvate, phosphate dikinase, PYC = pyruvate carboxylase, SLC4/SLC26 = solute carrier family 4/6, SPT = serine pyruvate transaminase. Outside grey boxes: CO<sub>2</sub> = carbon dioxide, GLC = glycerate, GLX = glyoxylate, HCO<sub>3</sub><sup>-</sup> = bicarbonate, MAL = malate, NH<sub>4</sub><sup>+</sup> = ammonium, OAA = oxaloacetate, PEP = phosphoenolpyruvate, PYR = pyruvate.

Table 1. Numbers of differentially expressed genes ( $\geq 0.95$  posterior probability of a 2-fold change in ASC) (Wu et al. 2010) across pCO<sub>2</sub> condition in each transcriptome.

		low	projected	max
<i>T. weissflogii</i> 24900 genes	up	334	39	186
	down	247	117	248
<i>T. oceanica</i> 36382 genes	up	1	56	1278
	down	106	2	2539
<i>T. rotula</i> 31792 genes	up	678	34	-
	down	310	28	-

Table 2. Number of putative carbonic anhydrases from model diatom genomes and the transcriptomes described in this study. The localization and activity of the full complement of putative carbonic anhydrases in *P. tricornutum* and *T. pseudonana* have not been completely experimentally verified (Tachibana et al. 2011; Samukawa et al. 2014; Kikutani et al. 2016).

	$\alpha$	$\beta$	$\delta$	$\gamma$	$\zeta$	$\theta$	total
<i>T. weissflogii</i>	7	1	4	3	4	0	19
<i>T. oceanica</i>	3	0	5	5	1	0	14
<i>T. rotula</i>	2	1	4	5	1	0	13
<i>T. pseudonana</i>	3	0	4	5	1	1	14
<i>P. tricornutum</i>	5	2	0	2	0	1	10

Manuscript 3

Formatted for publication in Journal of Experimental Marine Biology and Ecology

Following diatom response to ocean acidification in mesocosm experiments with metatranscriptome sequencing indicates contrasting CO<sub>2</sub> responses in *Skeletonema* and *Thalassiosira* diatoms

Joselynn Wallace<sup>1</sup>, Andrew L. King<sup>2</sup>, Gary H. Wilkfors<sup>3</sup>, Shannon L. Meseck<sup>3</sup>, Yuan Liu<sup>3</sup>, Bethany D. Jenkins<sup>1,4</sup>

<sup>1</sup> Department of Cell and Molecular Biology, University of Rhode Island, Kingston, RI, USA

<sup>2</sup> NIVA Norwegian Institute for Water Research, Bergen, Norway

<sup>3</sup> NOAA/NMFS, Northeast Fisheries Science Center, Milford Laboratory, Milford, CT, USA

<sup>4</sup> Graduate School of Oceanography, University of Rhode Island, Narragansett, RI, USA

## ABSTRACT

Fossil fuel combustion has raised atmospheric pCO<sub>2</sub> from ~277 ppm in 1750 to the present level of ~400 ppm. The ocean sequesters ~30% of that CO<sub>2</sub>, decreasing its buffering capacity and pH (ocean acidification). Diatoms are responsible for ~40% of oceanic primary production based upon photosynthetic fixation of CO<sub>2</sub>, which is aided by a carbon-concentrating mechanism (CCM), which uses bicarbonate transporters (BCTs) to take up HCO<sub>3</sub><sup>-</sup> and carbonic anhydrases (CAs) to interconvert HCO<sub>3</sub><sup>-</sup> and CO<sub>2</sub>. There are several distinct classes of BCTs and CAs found in diatoms, including the substrate-specific SLC4 and non-specific SLC26 BCTs, as well as the  $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\gamma$ ,  $\zeta$ , and  $\theta$  CAs. The ability to regulate and conserve energetic costs associated with the CCM may influence diatom community composition in a higher CO<sub>2</sub> future.

This study used mesocosm CO<sub>2</sub> manipulation incubations to examine the diatom community response to changes in pCO<sub>2</sub>. Seawater from Vineyard Sound, MA was collected in March of 2014, pre-filtered to remove large grazers, and amended with nutrients. The pCO<sub>2</sub> levels were manipulated to approximate pre-industrial conditions (< 215 ppm pCO<sub>2</sub>), present-day pCO<sub>2</sub> in the open ocean (330-390 ppm pCO<sub>2</sub>), and future pCO<sub>2</sub> projections (>780 ppm pCO<sub>2</sub>). After 20 days, metatranscriptomes were sequenced (>3- $\mu$ m size fraction); the majority of reads were attributed to *Skeletonema* spp. and *Thalassiosira* spp. diatoms.

We show evidence of differing CCM strategies: *Thalassiosira* diatoms regulate the  $\alpha$ ,  $\delta$ , and  $\zeta$  CAs and the SLC4 and SLC26 BCTs; whereas, *Skeletonema* regulates the  $\gamma$  and  $\zeta$  CAs and the SLC26 BCTs. The CCM of these two diatoms require different metal cofactors: CAs generally use zinc as a cofactor, but the  $\gamma$  CAs can

substitute with iron and  $\delta$  CAs can substitute with cobalt. *Thalassiosira*'s ability to downregulate the  $\text{HCO}_3^-$ -specific SLC4 BCTs in high  $\text{CO}_2$  may afford this organism some energetic savings in high  $\text{CO}_2$ .

## INTRODUCTION

Through physical mixing and biological activity, the ocean sequesters approximately 1/3 of anthropogenic  $\text{CO}_2$  and has been the only net sink for  $\text{CO}_2$  in the past 200 years (Sabine et al. 2004; Khatiwala et al. 2013). Rising anthropogenic  $\text{CO}_2$  emissions have led to an increase in atmospheric  $\text{pCO}_2$  concentration from  $\sim 277$  ppm in the year 1750 to over 400 ppm presently (Tans 2009; Le Quéré et al. 2016). Increased levels of dissolved  $\text{CO}_2$  gas in seawater perturbs carbonate chemistry in the ocean, which results in a decreased buffering capacity, or ocean acidification (OA) (Caldeira & Wickett 2003).

Diatoms are a phytoplankton group responsible for  $\sim 40\%$  of oceanic primary production via photosynthetic fixation of  $\text{CO}_2$  (Nelson et al. 1995). The predominant dissolved form of inorganic carbon in the ocean is bicarbonate ( $\text{HCO}_3^-$ ) (Zeebe & Wolf-Gladrow 2001). Diatom  $\text{CO}_2$  fixation is aided by carbon-concentrating mechanisms (CCMs), which use bicarbonate transporters (BCTs) to take up  $\text{HCO}_3^-$  and carbonic anhydrases (CAs) to interconvert  $\text{HCO}_3^-$  and  $\text{CO}_2$  adjacent to the carbon fixation enzyme RuBisCO (Badger et al. 1998; Hopkinson et al. 2011). Carbonic anhydrases are a protein family that arose from convergent evolution (Hewett-Emmett & Tashian 1996) that has resulted in distinct protein isoforms. In the diatoms, these include the widespread  $\alpha$ ,  $\delta$ , and  $\gamma$  types to rare  $\beta$  and  $\zeta$  carbonic anhydrases, and to the

newly described and relatively underexplored  $\theta$  carbonic anhydrases (Shen et al. 2017; Kikutani et al. 2016)

In natural phytoplankton communities, diatoms will generally outcompete other phytoplankton in high CO<sub>2</sub> (Tortell et al. 2002; Tortell et al. 2008; Bach et al. 2017), however the diatom growth response can vary across diatom genera. CO<sub>2</sub> manipulation experiments of natural diatom assemblages indicate that in high CO<sub>2</sub>, *Thalassiosira* spp. grow faster than *Skeletonema* spp. (Sett et al. 2018) while *Skeletonema* spp. grow faster than *Nitzschia* spp. (Kim et al. 2006). Species within the *Thalassiosira* diatom genus also display a range of growth rate changes across CO<sub>2</sub> conditions (<230 ppm CO<sub>2</sub> to >2900 ppm CO<sub>2</sub>), from no change across conditions (*T. pseudonana*), slower growth in lower CO<sub>2</sub> (*T. weissflogii*), slower growth in higher CO<sub>2</sub> (*T. oceanica*), and faster growth in higher CO<sub>2</sub> (*T. rotula*) (King et al. 2015).

Global gene expression profiling data (transcriptomes) from the King et al. 2015 CO<sub>2</sub> response experiments (King et al. 2015) indicate that differences in CO<sub>2</sub> responses between *T. rotula*, *T. weissflogii*, and *T. oceanica* boil down to their ability to dynamically regulate their carbon concentrating mechanisms while maintaining internal redox and ion homeostasis, as well as potential trade-offs between funneling excess CO<sub>2</sub> into internal storage pools versus a higher growth rate (Wallace et al. in prep). This range of plasticity in the diatoms' CCM is particularly important, as the use of this mechanism is energetically costly (Raven 1991). Diatoms that do downregulate the CCM in high CO<sub>2</sub> conserve this energy, potentially conferring a competitive advantage over those that do not. However, diatoms that tend to store excess carbon in high CO<sub>2</sub> (rather than exhibit higher growth rates) would have larger

internal storage pools of lipids and carbohydrates, potentially allowing those populations to persist when nutrients are scarce. These differences in CCM regulation and storage tendencies could influence diatom community composition and dynamics, but the full impact of these contrasting abilities is not captured by single-species isolate experiments that do not accurately reflect the complex and competitive environment that diatoms inhabit.

Community-level global gene expression profiling (metatranscriptome) studies are a powerful approach that capture the molecular basis for different diatoms' contrasting growth and nutrient acquisition strategies in mixed assemblages. Metatranscriptomes sequenced from nutrient-amended waters on the coast of Narragansett Bay, Rhode Island suggest that *Skeletonema* and *Thalassiosira* diatoms co-exist because the two diatom genera may access different internal and external pools of nitrogen and phosphorus (Alexander et al. 2015), while *Thalassiosira* and *Pseudo-nitzschia* diatoms co-exist in the iron-limited Pacific Ocean by using enzymes with different metal cofactor requirements (Cohen et al. 2017). Similar approaches combining metatranscriptomes and manipulation of CO<sub>2</sub> levels in natural phytoplankton assemblages might give a more complete picture of how changing CO<sub>2</sub> levels impact different diatoms' metabolic capacities in their complex environment.

This study used mesocosm CO<sub>2</sub> manipulation incubations to examine the diatom community response to changes in pCO<sub>2</sub>. Surface seawater from Vineyard Sound, MA was pre-filtered to remove large grazers (>180 μm), amended with nutrients, and the pCO<sub>2</sub> levels in the incubations were manipulated to approximate low/pre-industrial (less than 215 ppm pCO<sub>2</sub>) , present-day ambient pCO<sub>2</sub> in the open

ocean (330-390 ppm pCO<sub>2</sub>), and future pCO<sub>2</sub> projections (780-1320 ppm pCO<sub>2</sub>) (Tans 2009; Le Quéré et al. 2016). Metatranscriptomes enriched for eukaryotic biomass (size fraction greater than 3 μm) were sequenced on an Illumina HiSeq platform. Metatranscriptome sequence reads were mapped to a database of transcriptomes available through the Gordon and Betty Moore Foundation Marine Microbial Eukaryote Transcriptome Sequencing Project (Keeling et al. 2014). Across all the libraries, the majority of reads map to *Skeletonema* and *Thalassiosira* diatoms. The presence of these diatoms across all incubations is corroborated by light microscopy observations of the incubations at the final time point and barcoding of the diatom-specific V4 region of the 18S rDNA gene. Analysis of the transcriptional response in *Skeletonema* and *Thalassiosira* in these mesocosm experiments show distinct enzymatic strategies for N and P uptake and scavenging in these two diatom genera. There is also evidence of distinct strategies in their CCMs – while *Thalassiosira* diatoms seem to use the α, δ, and ζ carbonic anhydrases and the SLC4 and SLC26 bicarbonate transporters, *Skeletonema* uses the γ and ζ carbonic anhydrases and the SLC26 bicarbonate transporters.

## **MATERIALS AND METHODS**

### **Seawater collection and Incubation**

On March 4th 2014, surface seawater was collected in Vineyard Sound, Massachusetts using acid-cleaned 50 liter carboys. After Nitex mesh pre-filtering to remove particles larger than 180 μm, water was distributed into 9 acid-cleaned, 10 L carboys. Target pCO<sub>2</sub> levels were chosen to approximate low/pre-industrial (less than 215 ppm pCO<sub>2</sub>), present-day ambient pCO<sub>2</sub> in the open ocean (330-390 ppm pCO<sub>2</sub>),

and future pCO<sub>2</sub> projections (780-1320 ppm pCO<sub>2</sub>) (Tans 2009; Le Quéré et al. 2016). Target CO<sub>2</sub> concentrations were obtained by sparging with CO<sub>2</sub> (<100 ml min<sup>-1</sup>), and between by gentle bubbling with an air:CO<sub>2</sub> mixture (<30 ml min<sup>-1</sup>). The macronutrients nitrate, phosphate and silicic acid were added to mimic nutrient levels in the mixed layer, aiming for final concentrations of 10 μM nitrate, 20 μM silicate, and 0.2 μM phosphate. Incubations were diluted with nutrient-amended filtered seawater pre-equilibrated to target CO<sub>2</sub>, on the 8<sup>th</sup>, 11<sup>th</sup>, 14<sup>th</sup>, and 17<sup>th</sup> day of incubation. On day 11, phosphate concentrations were amended to 1 μM to avoid a phosphate limitation.

All incubations were kept in an outdoor tank incubator, with temperature maintained (between -1 to 3 °C) by a flow-through water system from near-shore water. All incubations were conducted in triplicate for 20 days time to examine differences in CO<sub>2</sub> response while minimizing complicating factors that might arise during a longer incubation (e.g., growth on the walls of the carboys, etc.). After 20 days, biomass was collected on 3 μm pore-size polyester filters (Sterlitech Corporation, Kent, WA, USA) by gentle filtration using a peristaltic pump, flash frozen in liquid nitrogen, and stored at -80 °C until DNA and RNA could be extracted.

### **Macronutrients, carbonate, and chlorophyll measurements**

Nitrate + nitrite, phosphate, and silicic acid levels were measured from 0.2 μm filtered aliquots from each carboy and analyzed using a Quattro autoanalyzer. All nutrient analysis protocols were developed by the Royal Netherlands Institute for Sea Research. Nitrate + nitrite was determined using the red azo dye method (Method NO Q-068- 05 Rev. 4), phosphate was determined using a phosphomolybdenum complex

(Method NO Q-064-05 Rev. 3), and a silico-molybdenum blue complex was used to determine silicic acid (Method NO Q-066-05 Rev. 3).

Dissolved inorganic carbon (DIC) and pH samples were collected periodically throughout the course of the experiment, as per the Department of Energy handbook (DOE 1994). Total DIC was measured using a DIC analyzer based upon sample acidification and LI-COR CO<sub>2</sub> detection (Apollo SciTech). pH (total scale) was determined colorimetrically using metacresol purple (Sigma-Aldrich) with a Varian dual beam spectrophotometer (Agilent Technologies), 10 cm cylindrical cells (Innovative Lab Supply), and jacketed cell holders, with temperature maintained by a Peltier cooler. All samples were analyzed within 60 minutes of collection.

Chlorophyll samples were measured by vacuum-filtration onto 25 mm GF/F glass fiber filters (Whatman, GE HealthCare Biosciences) that were then submerged in 90% acetone in 15 ml centrifuge tubes. Chlorophyll was extracted from the filters for 1 or 24 hours in the dark at -20°C. After extracting, samples were vortexed and acetone was decanted and read in a Turner Designs fluorometer (Parsons et al. 1984).

Particulate organic C, N, and P samples were vacuum-filtered onto pre-combusted 25 mm GF/C glass fiber filters (Whatman, GE HealthCare Biosciences). Briefly, POC and PON samples were dried in an oven at 60°C, acidified with HCl fumes, and analyzed using gas chromatography with a CHNSO analyzer (Costech) (Parsons et al. 1984). POP samples were rinsed with 0.17 M Na<sub>2</sub>SO<sub>4</sub>, dried at 90°C in 0.017 M MgSO<sub>4</sub>, and measured using the colorimetric molybdate method (Solorzano & Sharp 1980).

### **DNA Extraction**

DNA extractions were performed using the DNeasy Plant Kit (Qiagen, Germany) following the manufacturer's protocols with the following modifications: Filters were bead-beat in lysis buffer with 0.5 mm zircon beads and RNaseA (BioSpec, Bartlesville, OK, USA) for 1 minute to lyse the cells. Cell debris was immediately removed by running lysate over a Qiagen QiaShredder column. Extracted DNA samples were quantified on a NanoDrop 8000 spectrophotometer and stored at -20 °C

### **RNA Extraction**

RNA extractions were performed using the RNeasy Mini Kit (Qiagen, Germany) following the manufacturer's instructions with the following modifications. Filters were bead-beat in lysis buffer with 0.5 mm zircon beads (BioSpec, Bartlesville, OK, USA) for 1 minute to lyse the cells. Cell debris was immediately removed by running lysate over a Qiagen QiaShredder column. DNA contamination was removed by an on-column DNase step using the RNase-free DNase Set (Qiagen) and an additional DNase step was performed on eluted RNA using the Turbo DNA-free kit (Life Technologies, Carlsbad, CA, USA), both performed as per the manufacturers' instructions. RNA samples were quantified using the Qubit High-Sensitivity Kit (Life Technologies) and stored at -80 °C

### **DNA Amplicon sequencing**

Diatom-specific primers to amplify the V4 region of the 18S rDNA gene (Zimmermann et al. 2011) were modified to include Illumina MiSeq (Illumina, San

Diego, CA, USA) sequencing adapters (Mellett et al. 2018) (forward primer: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATTCCAGCTCCAA TAGCG-3'; reverse primer: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGGACTACGATGGTATCTAATC-3'). The PCR reaction used the following: 2-8 ng of template, 500 nmol L<sup>-1</sup> of each primer, 1x BIO-X-ACT Short Mix, and water to reach a 25 µL reaction volume. The following conditions were used for the PCR cycle: 5 minutes at 95°C; 32 cycles of 1 minute at 95°C, 1 minute at 61.8°C, 30 seconds at 72°C; and a final extension step for 10 minutes at 72°C. PCR products were purified using the Qiagen QIAquick Purification Kit (Qiagen, Germany) followed by visualization on a 1.2% agarose gel. Sequencing was performed using an Illumina MiSeq Reagent Kit V2 (500 cycle), with 2x250 bp paired-end reads at a depth of 100,000-500,000 reads per sample. Sequencing was performed at the Rhode Island Genomics Sequencing Center.

Adapter sequences were trimmed from raw reads using CutAdapt version 1.15 (Martin 2011) with the '--minimum-length 1' flag set to omit reads with a length of 0 and the '--discard-untrimmed' flag set to discard reads that do not have the adapter sequence and a maximum 10% error rate allowed. DADA2 version 1.8.0 (Callahan et al. 2016) was used to de-noise reads (filterAndTrim default settings with the following changes: truncLen of 220 to discard reads shorter than 220 bases and maxEE of 2 to discard reads with more than 2 expected errors) and merge read pairs (mergePairs using default settings). DADA2 was also used to remove chimeras (removeBimeraDenovo using default settings) and amplicons were filtered to remove sequences less than 390 bp or greater than 410 bp. Taxonomy was assigned using the

PR2 database version 4.10.0 (Guillou et al. 2013) and rarefaction curves were made in vegan version 2.5-2 (Oksanen et al. 2018) (Supplemental Figure 1).

### **RNA Sequencing**

Equal amounts of total RNA from each triplicate were pooled and sent to the JP Sulzberger Genome Sequencing Center at Columbia University. All sequencing was performed on an Illumina HiSeq 2500 platform and all libraries were mRNA enriched using a poly-A pull-down method. Target sequencing depth was 30 million, paired-end, 100 base pair read length. Raw reads were trimmed in Trimmomatic version 0.3.2 (Bolger et al. 2014) as follows: remove any Illumina adapters, trim the first 15 nucleotides, trim below quality score of 20 using a sliding window of 6 nucleotides, and 25 nucleotide minimum read length. The number of reads remaining after trimming are reported in Supplemental Table 1.

### **Two step read mapping approach: DIAMOND and BWA**

A custom database was constructed from all available sequences from the Gordon and Betty Moore Foundation Marine Microbial Eukaryote Transcriptome Sequencing Project, clustered by species at 98% ID in CD-HIT-EST (Li & Godzik 2006; Keeling et al. 2014). Using this custom database, an initial read mapping step was performed using blastX in DIAMOND version 0.8.37 (Buchfink et al. 2015). The mapping parameters used in DIAMOND were as follows: --top 0 --more-sensitive --evaluate .00001. This allows only the top hits with an evaluate cut-off of  $e^{-5}$  to be reported as mapped. If there are multiple top hits for a read (e.g., a read matches to multiple protein sequences equally well), the proteins that recruit this read are all given equal weight. There is no apportioning of read counts; a read that maps to only 1

protein is given equal weight relative to a read that maps to 10 proteins. Forward and reverse reads were mapped separately because DIAMOND does not support paired-end read mapping. Proteins with at least 1 read mapped were binned at the phylum, genus, and species level.

After filtering the database to include only those organisms most highly represented in the DIAMOND search (*Skeletonema* and *Thalassiosira* diatoms), a subsequent read mapping step was performed using Burrows-Wheeler Aligner (BWA-MEM) (Li & Durbin 2010). The mapping parameters used in BWA-MEM were as follows: -T 60 -aM -k 10, as has been used in other metatranscriptome studies (Alexander et al. 2015). The alignments were filtered in SAMtools (Li et al. 2009) so that only reads with the highest possible quality score (MAPQ60) were counted as mapped to each transcript.

### **Normalization**

To account for changes in species abundance in each library, the resulting raw read counts were normalized as per-million reads mapped to each diatom species in the library using Python pandas (pandas.pydata.org). Transcripts with at least 2 normalized reads mapped in at least 1 library were included in further analysis.

### **Differential Transcript Expression**

Any transcript that passed the expression threshold (at least 2 normalized reads mapped in at least 1 library) were ran through Analysis of Sequence Counts (ASC) (Wu et al. 2010) to assess differential expression in the low and high CO<sub>2</sub> conditions relative to the present-day CO<sub>2</sub> condition. The raw read counts were the input to ASC, which is an empirical Bayes approach that estimates the prior distribution based on the

input data and is suitable for use in situations where biological replicates were pooled before sequencing (Wu et al. 2010), as was done here. The counts for each diatom species were ran individually through ASC (Wu et al. 2010). If the relative abundance for a given species changed across conditions, this would be reflected by a change in library size sequenced from those conditions. These differences in library size are taken into account when ASC builds a model of the transcript counts (Wu et al. 2010). Transcripts with a greater than 0.95 posterior probability of a 2-fold change were considered differentially expressed.

### **Metatranscriptome rarefaction**

Metatranscriptome rarefaction curves were constructed to determine sufficiency of sequencing depth. First, libraries were re-sampled (10% intervals from 10% to 100% of each library) using seqtk (Li 2018). The reads were mapped back (using BWA-MEM and MAPQ60 filtering) to the subset of genes from *Skeletonema* and *Thalassiosira* with at least 2 reads mapped in at least 1 library. The rarefaction curves shown plot the number of transcripts that had at least 2 reads mapped per re-sampled library (Supplemental Figure 2).

### **Annotation and Differential KO Expression**

The associated protein sequences from the transcripts that passed the expression threshold (at least 2 normalized reads mapped in at least 1 library) were annotated using the KEGG Automatic Annotation Server (KAAS) with the following settings: the bi-directional best hit assignment method using the BLAST search program to query the GENES database plus the model diatoms *T. pseudonana* and *P. tricornutum* (Moriya et al. 2007). To assess differential expression of specific KEGG

KOs across CO<sub>2</sub> conditions, genes were grouped first by the genus they originated from (*Skeletonema* or *Thalassiosira*) and then clustered by their KO assignment. The total number of normalized reads mapped per KO and per genus were ran through Analysis of Sequence Counts (ASC) (Wu et al. 2010). In addition to its use for differential expression of genes, ASC is appropriate for any counts-based measurement, as was done here to determine differential KO expression. Any KO with a greater than 0.95 posterior probability of a 2-fold change was considered differentially expressed.

### **Identification of carbon concentrating mechanism components**

Protein sequences from carbonic anhydrases and bicarbonate transporters previously identified in *Thalassiosira* or *Phaeodactylum* diatoms (Tachibana et al. 2011; Samukawa et al. 2014; Kikutani et al. 2016, Wallace et al., in prep) were queried against a BLAST database of protein sequences corresponding to transcripts that passed the expression threshold (at least 2 normalized reads mapped in at least 1 library). Sequences that passed the similarity threshold (35% identity, 80% query coverage per HSP) were considered putative carbonic anhydrases or bicarbonate transporters.

## **RESULTS AND DISCUSSION**

### **Chlorophyll, nutrients, and carbonate chemistry**

In these experiments, filtered seawater (including only the <180 μm size fraction) from Massachusetts Bay was incubated at three different CO<sub>2</sub> levels, approximating a low/pre-industrial (less than 215 ppm pCO<sub>2</sub>), present-day ambient pCO<sub>2</sub> in the open ocean (330-390 ppm pCO<sub>2</sub>), and future pCO<sub>2</sub> projections (780-1320

ppm pCO<sub>2</sub>) (Tans 2009; Le Quéré et al. 2016). Temperature, chlorophyll, dissolved macronutrients, and pCO<sub>2</sub> were tracked over the course of the experiment, and the initial and final time points also included measurements of BSi, POC, PON, and POP. Metatranscriptome and 18S V4 rDNA amplicons were sequenced from the final time point.

The target CO<sub>2</sub> levels were steadily maintained over the duration of the incubation (Figure 1A). At the final time point, the chlorophyll in the high CO<sub>2</sub> bottle was comparatively lower than the present-day open ocean pCO<sub>2</sub> (330-390 ppm pCO<sub>2</sub>) condition (1-way ANOVA; p<0.05) (Figure 1B). This corresponds to significant NO<sub>3</sub><sup>-</sup>+NO<sub>2</sub><sup>-</sup> (Figure 1C) and PO<sub>4</sub> (Figure 1D) left unused in the high CO<sub>2</sub> incubations relative to the present-day open ocean pCO<sub>2</sub> incubations (1-way ANOVA; p<0.05). There was also significantly more Si(OH)<sub>4</sub> left in the high CO<sub>2</sub> incubations (Figure 1E). This is consistent with the measurements of lower POC (Figure 1F), PON (Figure 1G), and POP (Figure 1H), in the high CO<sub>2</sub> condition relative to the present-day open ocean pCO<sub>2</sub> (1-way ANOVA; p<0.05), although there was not a significant difference in the BSi measurements across CO<sub>2</sub> conditions (Figure 1I) (1-way ANOVA; p<0.05). For each of these measurements (chlorophyll, macronutrients, PON/C/P), there was no statistical difference between the low CO<sub>2</sub> condition and the present-day open ocean condition.

The differences observed in the high CO<sub>2</sub> condition relative to the present-day open ocean CO<sub>2</sub> condition do not track over the course of the experiment – previous to the final time point, the drawdown of NO<sub>3</sub><sup>-</sup>+NO<sub>2</sub><sup>-</sup> (Figure 1C), PO<sub>4</sub> (Figure 1D), and Si(OH)<sub>4</sub> (Figure 1E) were similar across CO<sub>2</sub> conditions. POC (Figure 1F), PON

(Figure 1G), and POP (Figure 1H) were only measured at the initial and final time points, so we cannot compare trends over the course of the incubation to final time point observations.

If diatoms grown in high CO<sub>2</sub> are less able to tolerate stressful conditions like changing temperatures, an abrupt change in temperature at the final time point could account for the observed differences in the final time point relative to the course of the experiment. However, this doesn't seem to be the case here, as the temperature fluctuations were minimal (<~3° C).

An additional possibility is that there has been some differential grazing pressure on the phytoplankton in the high CO<sub>2</sub> incubations, as the pre-filtering would not have removed all microzooplankton grazers. This could be significant, as micrograzers consume 67% of phytoplankton daily growth (Calbet & Landry 2004). While this could explain the loss of chlorophyll, the surplus of dissolved nutrients (Figure 1B, 1C, 1D, 1E) suggests that there is some other mechanism constraining phytoplankton biomass accumulation in the high CO<sub>2</sub> final time point. Previous mesocosm studies have also found that microzooplankton grazing rates were not influenced by pCO<sub>2</sub> level (Suffrian et al. 2008). One possibility is that phytoplankton viruses have increased in concentration or activity in the final time point high CO<sub>2</sub> condition. This could be important, as a 20% increase in viral load led to a decrease in bulk carbon fixation rates by nearly 50% (Suttle 1992). However, previous studies have reported no difference in viral infection across CO<sub>2</sub> conditions (Tsiola et al. 2017). An additional possibility is that high CO<sub>2</sub> has led to a disturbance in intracellular ion and redox homeostasis and a decreased growth rate, which has been

observed in the diatom *T. oceanica* (Wallace et al. in prep), and so seems to be the most likely explanation for the decreased chlorophyll and nutrient uptake observed in the high CO<sub>2</sub> condition at the final time point.

### **Metatranscriptome reads are dominated by *Skeletonema* and *Thalassiosira* diatoms**

Read mapping by both DIAMOND and BWA indicates that most of the metatranscriptome reads are recruiting to diatoms, most of which are either *Skeletonema* or *Thalassiosira* genera (Figure 2A). The presence of *Skeletonema* spp. and *Thalassiosira* spp. diatoms is corroborated by the diatom-specific V4 18S rDNA gene taxonomy assignment (Figure 2B). The presence of both *Skeletonema* and *Thalassiosira* diatom genera was also confirmed by light microscopy observations at the final time point. The timing of these incubations also coincides with a period of time where both *Thalassiosira* spp. and *Skeletonema* spp. should be present in Massachusetts Bay (Hunt et al. 2010).

Rarefaction analysis from the 18S V4 rDNA amplicon sequencing are asymptotic (Supplemental Figure 1). This suggests that we have saturated the diatom V4 amplicon diversity at the species level in the PR2 database. Rarefaction of mapped reads from the metatranscriptome sequencing are approaching an asymptote (Supplemental Figure 2). This suggests that we are approaching saturating the diversity of transcripts detected (with at least 2 reads mapped) from *Skeletonema* spp. and *Thalassiosira* spp. in the MMETSP database.

## **Diatom genera show different patterns of expression of KOs for C, N, and P acquisition and metabolism across CO<sub>2</sub> conditions**

Based on the differential KO expression analysis, we looked at broad trends in metabolism on the genus level across CO<sub>2</sub> conditions at the final time point. Under low CO<sub>2</sub>, both *Skeletonema* spp. and *Thalassiosira* spp. upregulate the citrate cycle to release stored energy and carbon dioxide, but upregulate different enzymes within the cycle (Figure 3, Supplemental Table 1). While *Skeletonema* spp. upregulates fumarate hydratase (fumC) and citrate synthase (CS), *Thalassiosira* spp. upregulates succinyl-CoA synthetase (LSC2) and malate dehydrogenase (maeB) (Figure 3, Supplemental Table 1). This is consistent with previous laboratory studies in *T. pseudonana* that saw a downregulation of the citrate cycle in high CO<sub>2</sub> (Hennon et al. 2015).

Based on their expression profiles, *Thalassiosira* spp. and *Skeletonema* spp. seem to have distinct strategies for generating glutamate -- *Thalassiosira* spp. regenerates glutamate from glutamine (glnA and glnD) and *Skeletonema* spp. seems to preferentially regenerate glutamate from proline (IP5C), all of which are higher in low CO<sub>2</sub> relative to their expression level in present day CO<sub>2</sub> (Figure 3, Supplemental Table 1). The increased expression of these KOs (glnA, glnD, IP5C) in low CO<sub>2</sub> could be indicative of shuttling more amino acids toward glutamate, which can fuel the citrate cycle via its conversion to  $\alpha$ -ketoglutarate (Figure 3, Supplemental Table 1).

*Skeletonema* spp. and *Thalassiosira* spp. differentially express KOs mediating their nitrogen acquisition strategies across CO<sub>2</sub> conditions (Figure 4, Supplemental Table 1). Both diatoms downregulate nitrate transporters in high CO<sub>2</sub> (NRT) (Figure 4, Supplemental Table 1), consistent with the high CO<sub>2</sub> condition incubations drawing

down less  $\text{NO}_2^- + \text{NO}_3^-$  relative to the present day  $\text{CO}_2$  conditions at the final time point (Figure 1C). There is no compensation in the expression of ammonium transporters (AMT) across  $\text{CO}_2$  conditions (Figure 4, Supplemental Table 1). However, in low  $\text{CO}_2$ , *Thalassiosira* spp. downregulates urease (URE), and in high  $\text{CO}_2$ , *Skeletonema* spp. upregulates urease (URE), suggesting that both diatoms might hydrolyze more urea as  $\text{CO}_2$  increases (Figure 4, Supplemental Table 1), although rates of urea hydrolysis and urea levels were not measured in this experiment. These diatom genera also express KOs that suggest distinct approaches for supplying urea to the urease enzyme, as the DUR3 urea-proton symporter and the SLC14 urea transporter are not expressed in *Skeletonema* spp. in this experiment, but are constitutively expressed in *Thalassiosira* spp. (Figure 4, Supplemental Table 1). Similarly, *Thalassiosira* spp. upregulates an amino acid transporter (APA) in high  $\text{CO}_2$ , which is not expressed in *Skeletonema* spp. (Figure 5, Supplemental Table 1). This is consistent with a previous metatranscriptome study that found *Thalassiosira* spp. highly expressed amino acid transporters, but *Skeletonema* spp. did not (Alexander et al. 2015). However, both diatom genera show a broad upregulation of KOs involved in the degradation of amino acids leucine, valine, and isoleucine in high  $\text{CO}_2$  (Figure 5, Supplemental Table 1). This suggests that these amino acids might be a source of recycled nitrogen for *Skeletonema* and *Thalassiosira* spp. The high  $\text{CO}_2$  decrease in expression of nitrate transporters and increase in expression of organic nitrogen metabolism genes is consistent with previous studies in *Thalassiosira pseudonana* that saw a decrease in nitrate uptake in elevated  $\text{CO}_2$  (Shi et al. 2015; Hennon et al. 2014).

These two diatom genera show contrasting complements of phosphate metabolism KOs. While *Skeletonema* spp. constitutively expresses two phosphatase genes (ACP5 and phoD) involved in scavenging phosphate from internal pools and an SLC20 sodium-dependent phosphate transporter under low CO<sub>2</sub>, they are not expressed in *Thalassiosira* spp. (Figure 6, Supplemental Table 1). Both diatoms constitutively express an SLC34 sodium-dependent phosphate co-transporter, but it is lowly expressed in *Thalassiosira* spp. (Figure 6, Supplemental Table 1). This is consistent with a previous metatranscriptome study that found genes involved in phosphate uptake and scavenging were more highly expressed in *Skeletonema* than *Thalassiosira* (Alexander et al. 2015).

*Skeletonema* spp. also remodels phosphate metabolism in high CO<sub>2</sub>, but *Thalassiosira* spp. does not. In high CO<sub>2</sub>, *Skeletonema* spp. upregulates two KOs involved in scavenging phosphate from inositol phosphate (inositol-phosphate phosphatases IMPL2 and VTC4) (Figure 6, Supplemental Table 1), a class of molecule that was upregulated in phosphate limited cultures of the diatoms *Thalassiosira pseudonana* (Dyhrman et al. 2012) and *Phaeodactylum tricoratum* (Feng et al. 2015). Similarly, in high CO<sub>2</sub> *Skeletonema* spp. also upregulates a nucleotidase KO involved in scavenging phosphate from nucleotides (cysQ) (Figure 6, Supplemental Table 1), a class of molecule also upregulated in phosphate limited cultures of the diatom *Thalassiosira pseudonana* (Dyhrman et al. 2012). These KOs (IMPL2, VTC4, cysQ) are not differentially expressed in *Thalassiosira* spp. The observations that at the final time point the high CO<sub>2</sub> condition incubations drew down less phosphate relative to the present day CO<sub>2</sub> conditions (Figure 1G) and the high

CO<sub>2</sub> upregulation of genes involved in phosphate scavenging (Figure 6, Supplemental Table 1) suggest that high CO<sub>2</sub> *Skeletonema* diatoms upregulate KOs to preferentially recycle internal pools of phosphate rather than take up exogenous phosphate, while CO<sub>2</sub> status has little impact on the expression of phosphate scavenging KOs in *Thalassiosira* spp. (Figure 6, Supplemental Table 1).

**Expression patterns in ion and redox homeostasis KOs indicate diatoms experience increased oxidative stress in high CO<sub>2</sub>**

The final incubation time points in high CO<sub>2</sub> saw lower chlorophyll, lower POC, PON, and POP, as well as less nutrient drawdown (1-way ANOVA; p<0.05). (Figure 1). One possible explanation is that high CO<sub>2</sub> has led to a disturbance in diatoms' intracellular ion and redox homeostasis, as has been observed in the diatom *T. oceanica* (Wallace et al. in prep).

At the final time point, both diatoms show signs of increased oxidative stress in high CO<sub>2</sub>, as indicated by the differential expression of KOs involved in glutathione metabolism (Figure 7, Supplemental Table 1), a key mediator of oxidative stress (Noctor & Foyer 1998). *Thalassiosira* spp. shows a low CO<sub>2</sub> downregulation and a high CO<sub>2</sub> upregulation of glutamate-cysteine ligase (GCLC) and glutathione S-transferase (GST), while *Skeletonema* spp. shows a high CO<sub>2</sub> upregulation of glutathione synthase (GSS) and glutathione reductase (GSR) (Figure 7). The upregulation of oxidative stress genes is consistent with a previous study that saw an upregulation of these genes in *Thalassiosira oceanica* grown in high CO<sub>2</sub> (Wallace et al., in prep). In addition, the high CO<sub>2</sub> upregulation of glutathione reductase and glutathione synthase in *Skeletonema* spp. is consistent with previous studies that report

an upregulation of these genes in CO<sub>2</sub>-enriched *Skeletonema marinoi* cultures (Lauritano et al. 2015). Diatoms may be regulating glutathione metabolism as part of a mechanism to maintain redox homeostasis, as the carbonic anhydrases in *P. tricornutum* are redox regulated and a similar redox regulation could be happening in *Skeletonema* spp. and *Thalassiosira* spp. (Kikutani et al. 2012).

### **Diatoms have distinct complements of carbon concentrating mechanism components that are regulated in response to CO<sub>2</sub>**

In addition to differential KO expression across genera, we also looked at differential expression on the per-gene level. *Skeletonema* spp. and *Thalassiosira* spp. express a similar suite of carbonic anhydrases ( $\alpha$ ,  $\delta$ ,  $\gamma$ , and  $\zeta$ ) (Figure 8, Supplemental Table 4) and bicarbonate transporters (SLC4 and SLC26) (Figure 9, Supplemental Table 4). If a carbonic anhydrase or bicarbonate transporter were involved in the CCM, it is expected that their expression would be upregulated in low CO<sub>2</sub> and/or downregulated in high CO<sub>2</sub>. In the case of *Thalassiosira*, there are  $\alpha$ ,  $\delta$ , and  $\zeta$  carbonic anhydrases that are up in low CO<sub>2</sub> or down in high CO<sub>2</sub> (Figure 8, Supplemental Table 4). This is in contrast with what is observed for *Skeletonema*, in which there are  $\gamma$  and  $\zeta$  carbonic anhydrases that are up in low CO<sub>2</sub> or down in high CO<sub>2</sub> (Figure 8, Supplemental Table 4). This suggests that while the  $\alpha$  and  $\delta$  carbonic anhydrases are important for the CCM of *Thalassiosira*, the  $\gamma$  carbonic anhydrases are important in *Skeletonema*. It also suggests that the  $\zeta$  carbonic anhydrases are key players in the CCM of both *Skeletonema* and *Thalassiosira*. The disparity in the CO<sub>2</sub>-sensitivity of the carbonic anhydrases may be partially explained by the fact that  $\alpha$  and  $\delta$  carbonic anhydrases have undergone many lineage-specific expansions or gene losses and that

*Skeletonema* and *Thalassiosira* diatoms contain completely distinct subsets of the  $\delta$  carbonic anhydrases (Shen et al. 2017). It also suggests that these diatoms have different abilities to substitute the typical zinc CA cofactor for another metal -- the  $\delta$  CAs can substitute cobalt (Lane & Morel 2000), while the  $\gamma$  CAs can substitute iron (although this has only been observed in anaerobic conditions) (Tripp et al. 2004). The key role of the  $\zeta$  carbonic anhydrases for these two diatom genera is particularly interesting, since this class of carbonic anhydrases is found in relatively few diatom species (Shen et al. 2017).

*Thalassiosira* spp. also dynamically regulates the SLC4 and SLC26 bicarbonate transporters in response to CO<sub>2</sub> (Figure 9, Supplemental Table 4), suggesting that regulating bicarbonate transport is critical for its CCM (Figure 4). In contrast, the *Skeletonema* spp. SLC4 transporter response is relatively muted but the SLC26 transporters are more sensitive to CO<sub>2</sub> (Figure 9, Supplemental Table 4). The sensitive expression of the SLC26 transporters suggests that *Skeletonema* may rely on diffusive CO<sub>2</sub> uptake or the non-specific SLC26 anion transporters for bicarbonate uptake rather than the bicarbonate-specific SLC4 transporters (Alper & Sharma 2013). These different strategies in inorganic carbon uptake are consistent with the hypothesis that historic fluctuations in atmospheric CO<sub>2</sub> have led to a diversification and re-configuration of the CCM within the diatoms (Young et al. 2016; Shen et al. 2017).

## CONCLUSIONS

This study suggests that the CCM of *Skeletonema* and *Thalassiosira* use distinct CA isozymes that use different metal co-factors (Lane & Morel 2000; Tripp et al. 2004). We also show evidence suggesting distinct strategies for inorganic carbon uptake: while *Thalassiosira* uses substrate-specific SLC4 BCTs, *Skeletonema* seems to rely on diffusive CO<sub>2</sub> transport or non-specific SLC26-mediated transport. This might influence community composition dynamics in high CO<sub>2</sub> because diatoms that can downregulate the SLC4 active transporters in high CO<sub>2</sub> might save some energy and gain a competitive advantage. The co-existence of these diatoms might be related to their CCMs that use different metal cofactors and they might also show unique responses to different metal regimes in high CO<sub>2</sub>.

This study suggests that there are differences in N and P metabolism in these two diatoms. The significance of these differences in the context of ocean acidification is still unclear. Future studies should attempt to constrain the extent that transcript expression levels correlate with protein expression levels, cell physiology (e.g., cellular C:N:P ratios), and rate measurements (e.g., nitrate uptake). Additionally, the enzymatic activity of the CAs and BCTs should be characterized to confirm their putative function, as well as their flexibility of metal cofactor requirements. Results from this study showing distinct responses in nutrient metabolism is in agreement with previous transcriptome studies that suggest there are fundamental differences in N and P metabolism in these co-occurring diatoms (Alexander et al. 2015). Future experiments that combine multiple stressors (e.g., culturing in a range of CO<sub>2</sub>, N, and P levels) could give a more complete picture of how these differences would impact diatom community dynamics.

## References

Alexander, H. et al., 2015. Metatranscriptome analyses indicate resource partitioning

- between diatoms in the field. *Proceedings of the National Academy of Sciences*, 112(17), pp.E2182-E2190.
- Alper, S.L. & Sharma, A.K., 2013. The SLC26 Gene Family of Anion Transporters and Channels. *Molecular Aspects of Medicine*, 34(2–3), pp.494–515.
- Bach, L.T. et al., 2017. Simulated ocean acidification reveals winners and losers in coastal phytoplankton. *PLoS ONE*, 12(11), pp.1–23.
- Badger, M.R. et al., 1998. The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO<sub>2</sub>-concentrating mechanisms in algae. *Canadian Journal of Botany*, 76(6), pp.1052–1071.
- Bolger, A.M., Lohse, M. & Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), pp.2114–2120.
- Buchfink, B., Xie, C. & Huson, D., 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1), pp.59–60.
- Calbet, A. & Landry, M.R., 2004. Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. *Limnology and Oceanography*, 49(1), pp.51–57.
- Caldeira, K. & Wickett, M.E., 2003. Oceanography: anthropogenic carbon and ocean pH. *Nature*, 425(6956), p.365.
- Callahan, B.J. et al., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*.
- Cohen, N.R. et al., 2017. Diatom Transcriptional and Physiological Responses to Changes in Iron Bioavailability across Ocean Provinces. *Frontiers in Marine Science*, 4(November).

- DOE, 1994. *Handbook of methods for the analysis of the various parameters of the carbon dioxide system in sea water; version 2* A. G. Dickson & C. Goyet, eds.,
- Dyhrman, S.T. et al., 2012. The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLoS ONE*, 7(3).
- Feng, T.-Y. et al., 2015. Examination of metabolic responses to phosphorus limitation via proteomic analyses in the marine diatom *Phaeodactylum tricornutum*. *Scientific Reports*, 5(1), p.10373.
- Guillou, L. et al., 2013. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*.
- Hennon, G.M.M. et al., 2014. Acclimation conditions modify physiological response of the diatom *Thalassiosira pseudonana* to elevated CO<sub>2</sub> concentrations in a nitrate-limited chemostat. *Journal of Phycology*, 50(2), pp.243–253.
- Hennon, G.M.M. et al., 2015. Diatom acclimation to elevated CO<sub>2</sub> via cAMP signalling and coordinated gene expression. *Nature Climate Change*, 5(August), pp.761–766.
- Hewett-Emmett, D. & Tashian, R.E., 1996. Functional diversity, conservation, and convergence in the evolution of the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -carbonic anhydrase gene families. *Molecular Phylogenetics and Evolution*, 5(1), pp.50–77.
- Hopkinson, B.M. et al., 2011. Efficiency of the CO<sub>2</sub>-concentrating mechanism of diatoms. *Proceedings of the National Academy of Sciences*, 108(10), pp.3830–3837.

- Hunt, C.D. et al., 2010. Phytoplankton patterns in Massachusetts Bay-1992-2007. *Estuaries and Coasts*, 33(2), pp.448–470.
- Keeling, P.J. et al., 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biology*, 12(6).
- Khatiwala, S. et al., 2013. Global ocean storage of anthropogenic carbon. *Biogeosciences*, 10(4), pp.2169–2191.
- Kikutani, S. et al., 2012. Redox regulation of carbonic anhydrases via thioredoxin in chloroplast of the marine diatom *Phaeodactylum tricornutum*. *Journal of Biological Chemistry*, 287(24), pp.20689–20700.
- Kikutani, S. et al., 2016. Thylakoid luminal  $\theta$ -carbonic anhydrase critical for growth and photosynthesis in the marine diatom *Phaeodactylum tricornutum*. *Proceedings of the National Academy of Sciences*, 113(35), pp.9828–9833.
- Kim, J.-M. et al., 2006. The effect of seawater CO<sub>2</sub> concentration on growth of a natural phytoplankton assemblage in a controlled mesocosm experiment. *Limnology and Oceanography*, 51(4), pp.1629–1636.
- King, A.L. et al., 2015. Effects of CO<sub>2</sub> on growth rate, C:N:P, and fatty acid composition of seven marine phytoplankton species. *Marine Ecology Progress Series*, 537, pp.59–69.
- Lane, T.W. & Morel, F.M.M., 2000. Regulation of Carbonic Anhydrase Expression by Zinc, Cobalt, and Carbon Dioxide in the Marine Diatom *Thalassiosira weissflogii*. *Plant Physiology*, 123(1), pp.345–352.
- Lauritano, C. et al., 2015. Key genes as stress indicators in the ubiquitous diatom

- Skeletonema marinoi*. *BMC Genomics*, 16(1).
- Li, H., 2018. seqtk. <https://github.com/lh3/seqtk>.
- Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–2079.
- Li, H. & Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5), pp.589-595
- Li, W. & Godzik, A., 2006. CD-HIT: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), pp.1658–1659.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 17(1), pp.10-12
- Mellet, T. et al., 2018. The biogeochemical cycling of iron, copper, nickel, cadmium, manganese, cobalt, lead, and scandium in a California Current experimental study. *Limnology and Oceanography*, 63, pp.S425–S447.
- Moriya, Y. et al., 2007. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, 35(SUPPL.2), pp.182–185.
- Noctor, G. & Foyer, C.H., 1998. ASCORBATE AND GLUTATHIONE: Keeping Active Oxygen Under Control. *Annual Review of Plant Physiology and Plant Molecular Biology*, 49(1), pp.249–279.
- Oksanen, J. et al., 2018. vegan: Community Ecology Package. <https://cran.r-project.org/package=vegan>.
- Parsons, T.R., Maita, Y. & Lalli, C.M., 1984. *A Manual of Chemical and Biological Methods for Seawater Analysis*,

- Le Quéré, C. et al., 2016. Global Carbon Budget 2016. *Earth System Science Data*, 8(2), pp.605–649.
- Raven, J., 1991. Physiology of inorganic carbon acquisition and implications for resource use efficiency by marine phytoplankton: Relation to increased carbon dioxide and temperature. *Plant, Cell and Environment*, 14(8), pp.779–794.
- Sabine, C.L. et al., 2004. The Oceanic Sink for Anthropogenic CO<sub>2</sub>. *Science*, 305(5682), pp.367–371.
- Samukawa, M. et al., 2014. Localization of putative carbonic anhydrases in the marine diatom, *Thalassiosira pseudonana*. *Photosynthesis Research*, 121(2–3), pp.235–249.
- Sett, S. et al., 2018. Shift towards larger diatoms in a natural phytoplankton assemblage under combined high-CO<sub>2</sub> and warming conditions. *Journal of Plankton Research*, 00(June), pp.1–16.
- Shen, C., Dupont, C.L. & Hopkinson, B.M., 2017. The diversity of carbon dioxide-concentrating mechanisms in marine diatoms as inferred from their genetic content. *Journal of Experimental Botany*.
- Shi, D. et al., 2015. Interactive effects of light, nitrogen source, and carbon dioxide on energy metabolism in the diatom *Thalassiosira pseudonana*. *Limnology and Oceanography*, 60(5), pp.1805–1822.
- Solorzano, L. & Sharp, J.H., 1980. Determination of total dissolved phosphorus particulate phosphorus in natural waters. *Limnology and Oceanography*, 25(4), pp.754–758.
- Suffrian, K. et al., 2008. Microzooplankton grazing and phytoplankton growth in

- marine mesocosms with increased CO<sub>2</sub> levels. *Biogeosciences*, 5(4), pp.1145–1156.
- Suttle, C.A., 1992. Inhibition of photosynthesis in phytoplankton by the submicron size fraction concentrated from seawater. *Marine Ecology Progress Series*, 87(1–2), pp.105–112.
- Tachibana, M. et al., 2011. Localization of putative carbonic anhydrases in two marine diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. *Photosynthesis Research*, 109(1–3), pp.205–221.
- Tans, P., 2009. An Accounting of the Observed Increase in Oceanic and Atmospheric CO<sub>2</sub> and the Outlook for the Future. *Oceanography*, 22(4), pp.26–35.
- Tortell, P.D. et al., 2002. CO<sub>2</sub> effects on taxonomic composition and nutrient utilization in an Equatorial Pacific phytoplankton assemblage. *Marine Ecology Progress Series*, 236, pp.37–43.
- Tortell, P.D. et al., 2008. CO<sub>2</sub> sensitivity of Southern Ocean phytoplankton. *Geophysical Research Letters*, 35(4), pp.1–5.
- Tripp, B.C. et al., 2004. A Role for Iron in an Ancient Carbonic Anhydrase. *Journal of Biological Chemistry*, 279(8), pp.6683–6687.
- Tsiola, A. et al., 2017. Ocean acidification and viral replication cycles: Frequency of lytically infected and lysogenic cells during a mesocosm experiment in the NW Mediterranean Sea. *Estuarine, Coastal and Shelf Science*, 186, pp.139–151.
- Wu, Z. et al., 2010. Empirical bayes analysis of sequencing-based transcriptional profiling without replicates. *BMC Bioinformatics*, 11(1), p.564.
- Young, J.N. et al., 2016. Large variation in the Rubisco kinetics of diatoms reveals

diversity among their carbon-concentrating mechanisms. *Journal of Experimental Botany*, 67(11), pp.3445–3456.

Zeebe, R.E. & Wolf-Gladrow, D., 2001. *CO<sub>2</sub> in Seawater: Equilibrium, Kinetics, Isotopes* 1st ed. D. Halpern, ed., Amsterdam: Elsevier Oceanography Series.

Zimmermann, J., Jahn, R. & Gemeinholzer, B., 2011. Barcoding diatoms: Evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity and Evolution*, 11(3), pp.173–192.

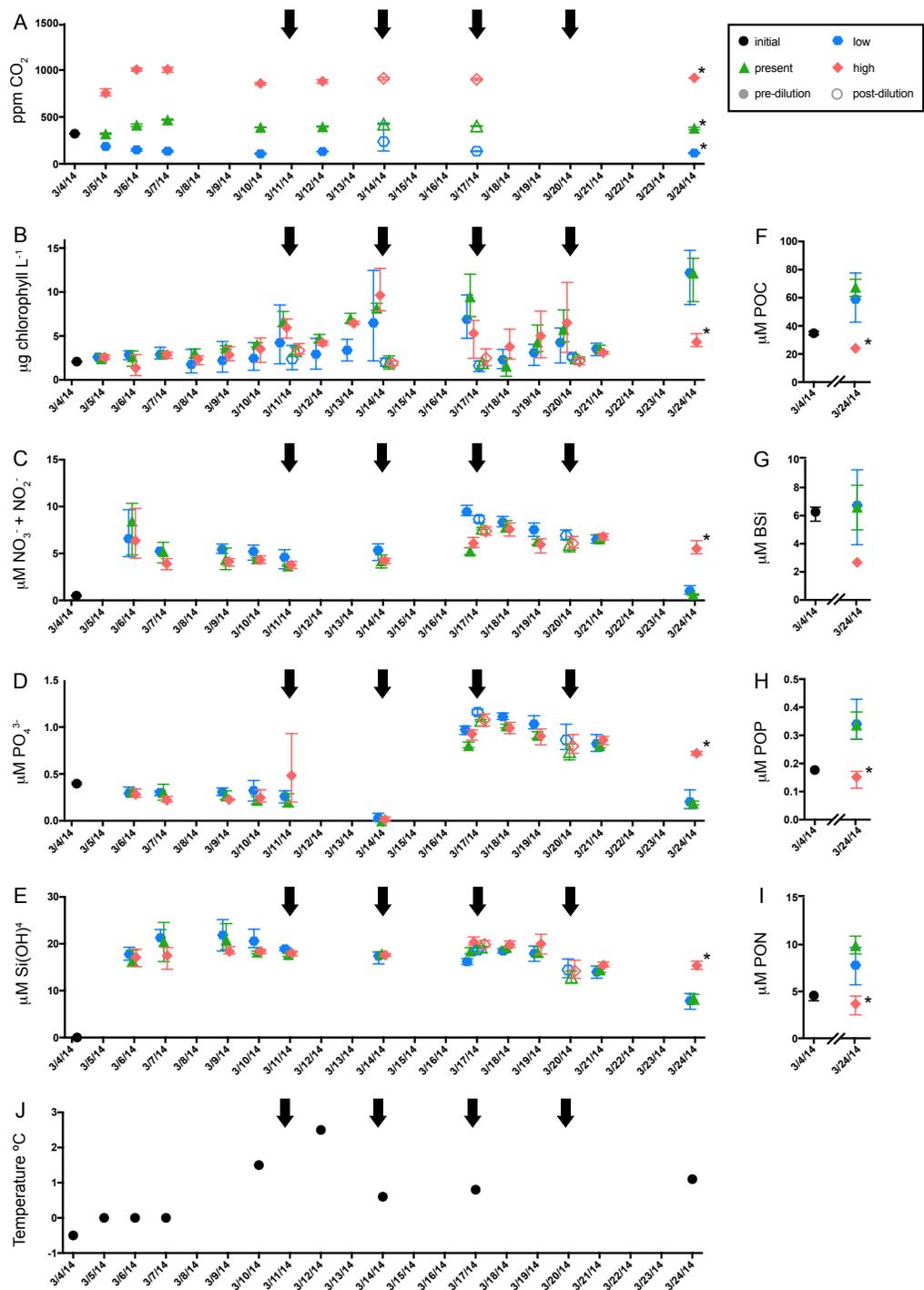


Figure 1. (A) ppm CO<sub>2</sub>, (B) Chlorophyll L<sup>-1</sup>, (C) μM NO<sub>3</sub><sup>-</sup>+NO<sub>2</sub><sup>-</sup>, (D) μM PO<sub>4</sub>, (E) μM Si(OH)<sub>4</sub>, (F) μM POC, (G) μM PON, (H) μM POP, (I) μM BSi, and (J) temperature °C for each incubation carboy. The x-axis indicates the date the variable was measured and the y-axis indicates the values and variables measured. The shapes and colors indicate the CO<sub>2</sub> condition shown. Points and error bars indicate the mean and range of the measurement across the triplicate incubation carboys. The asterisks indicate significant differences relative to the present day pCO<sub>2</sub> condition (1-way ANOVA; p < 0.05). Dilution days are indicated with black arrows. Solid shapes indicate that the measurement shown was taken before dilution, empty shapes indicate the measurement was taken post-dilution.

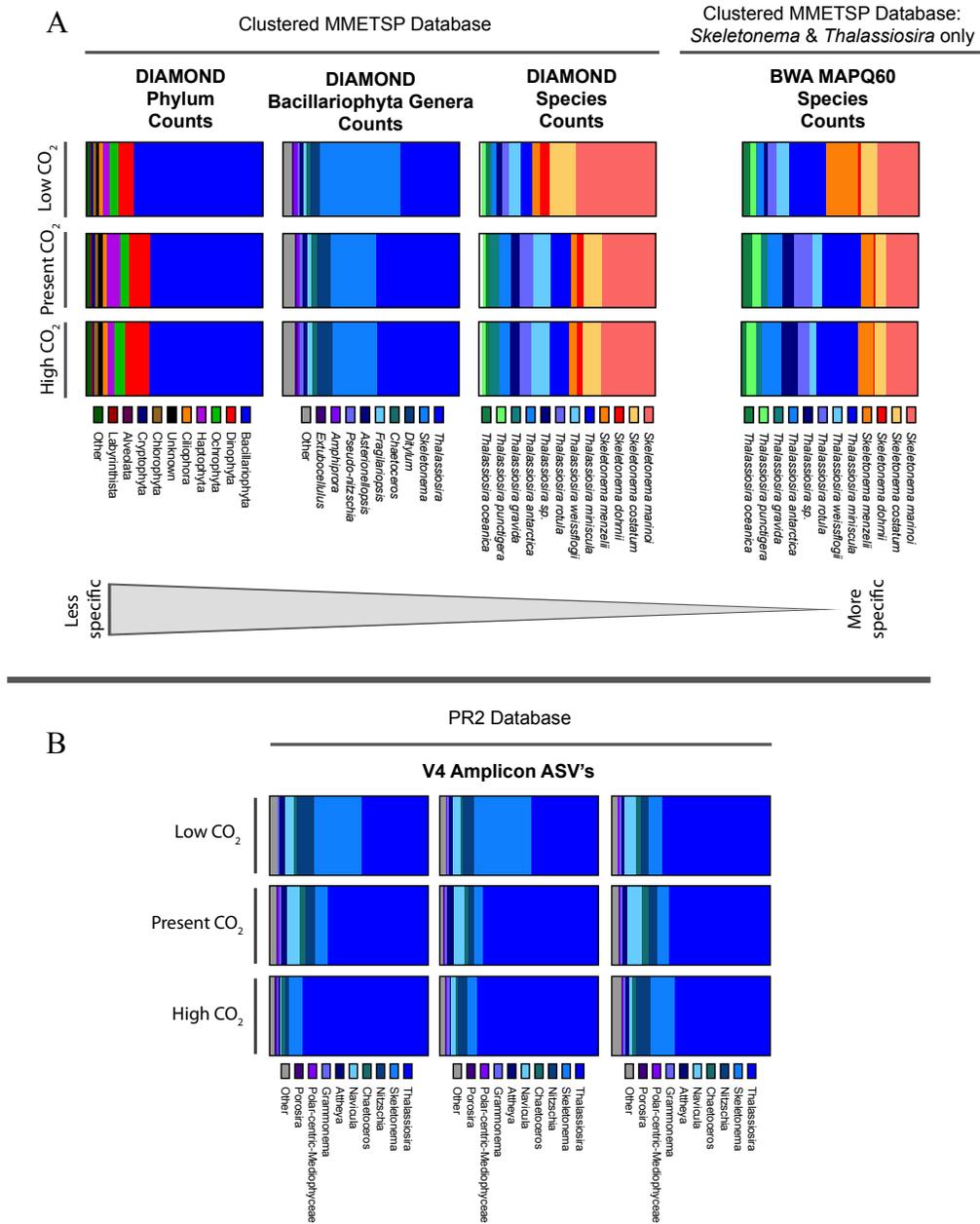


Figure 2. A: Bar charts showing the relative distributions of metatranscriptome reads mapped to the MMETSP database using either DIAMOND or BWA, moving from less specific to more specific from left to right. B: DADA2 taxonomic assignment of 18S V4 rDNA amplicon reads using the PR2 database.

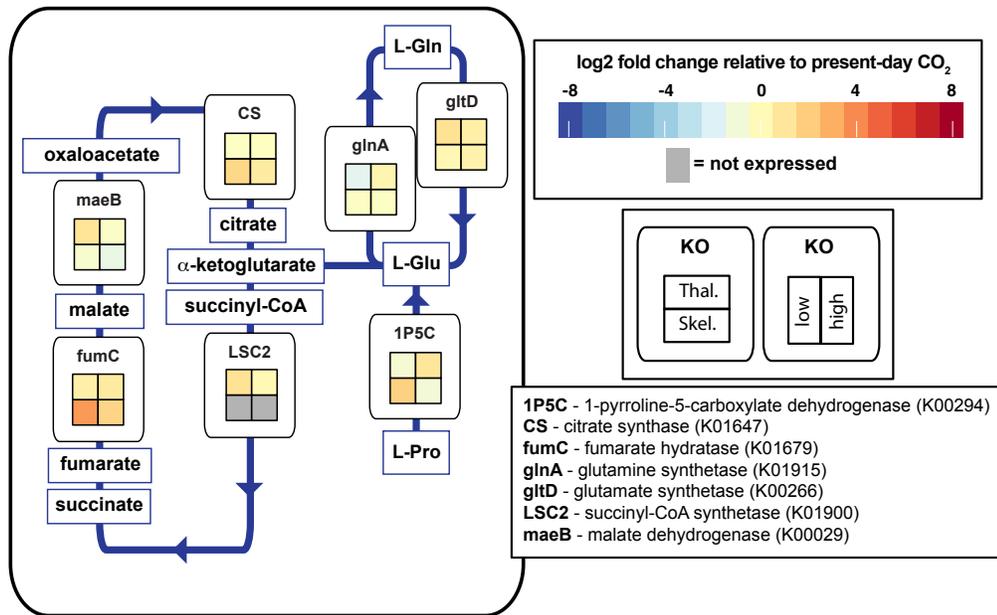


Figure 3. A heat map showing broad metabolic rearrangements in the pathway for the citrate cycle in *Skeletonema* and *Thalassiosira* across CO<sub>2</sub> conditions. The color indicates the log<sub>2</sub>-fold change of total reads mapped per genus per KEGG KO in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> condition.

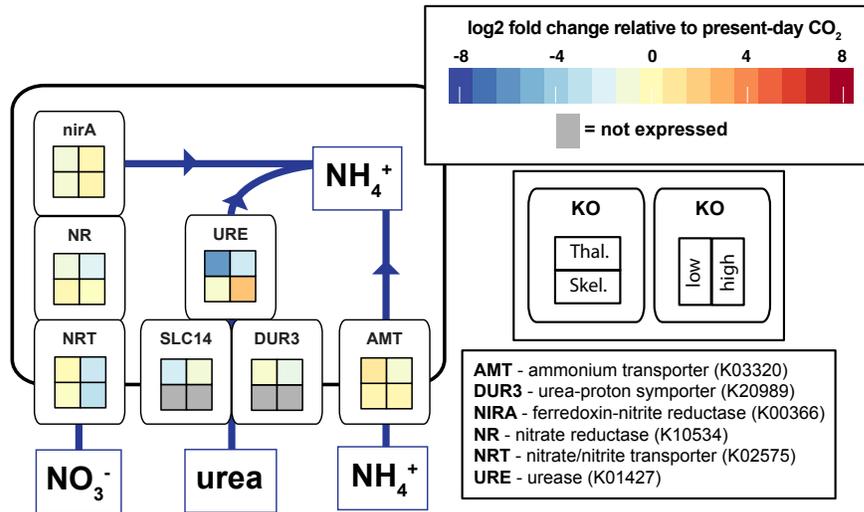


Figure 4. A heat map showing broad metabolic rearrangements in nitrogen uptake pathway members in *Skeletonema* and *Thalassiosira* across CO<sub>2</sub> conditions. The color indicates the log<sub>2</sub>-fold change of total reads mapped per genus per KEGG KO in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> condition.

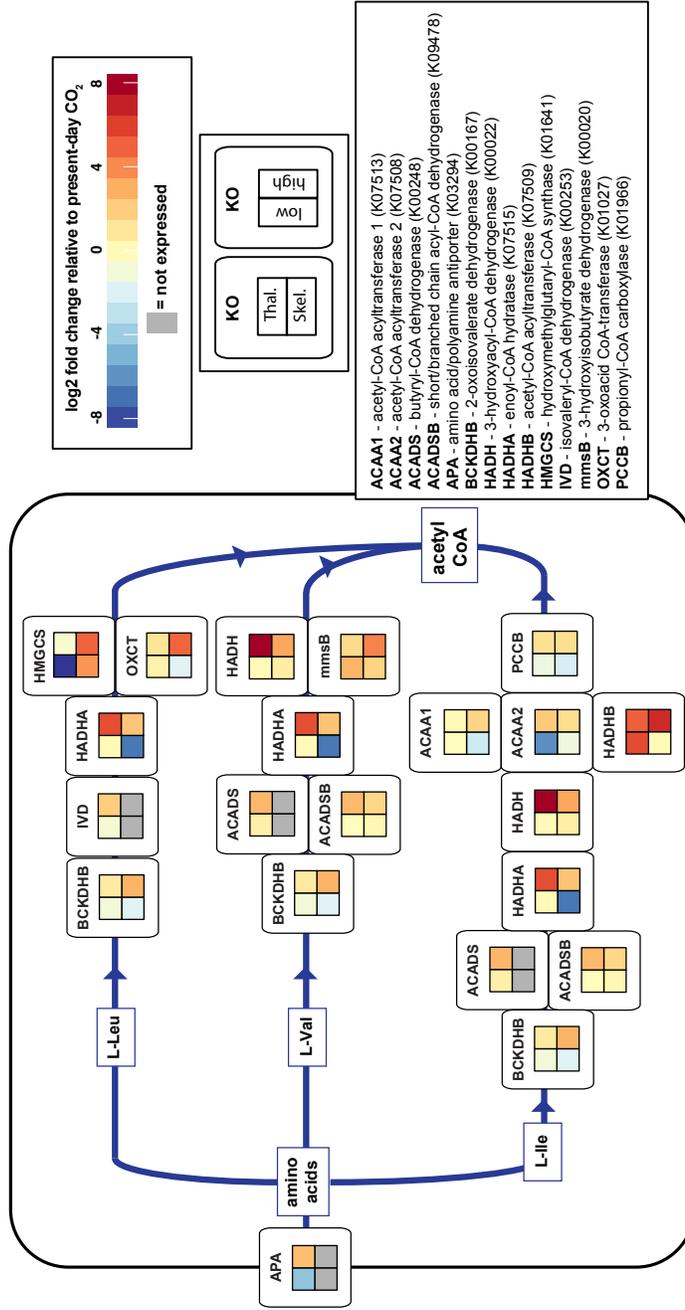


Figure 5. A heat map showing broad metabolic rearrangements in branched-chain amino acid metabolism pathway components in *Skeletonema* and *Thalassiosira* across CO<sub>2</sub> conditions. The color indicates the log<sub>2</sub>-fold change of total reads mapped per genus per KEGG KO in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> condition.

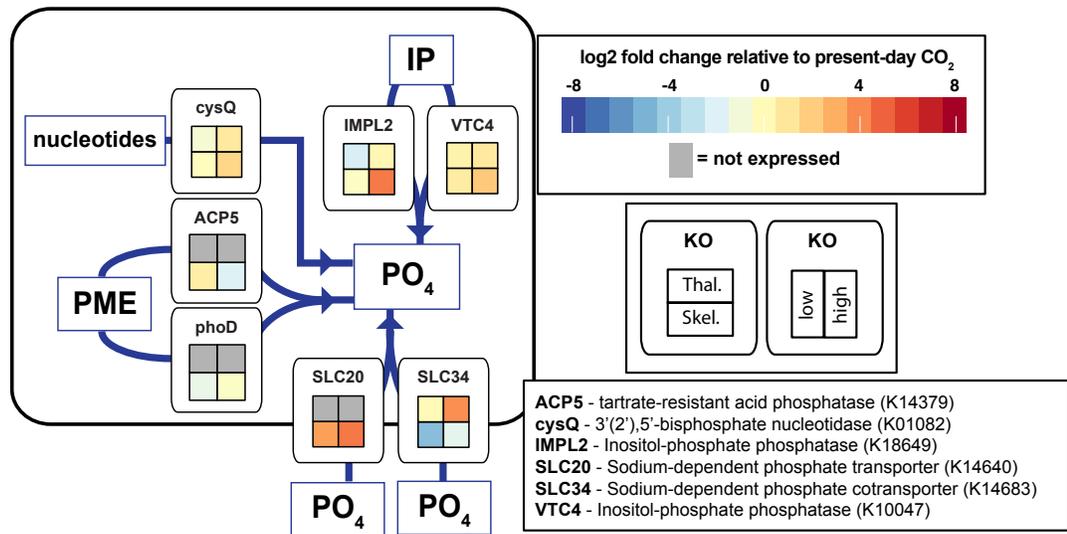


Figure 6. A heat map showing broad metabolic rearrangements in phosphorous metabolism in *Skeletonema* and *Thalassiosira* across CO<sub>2</sub> conditions. The color indicates the log<sub>2</sub>-fold change of total reads mapped per genus per KEGG KO in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> condition.

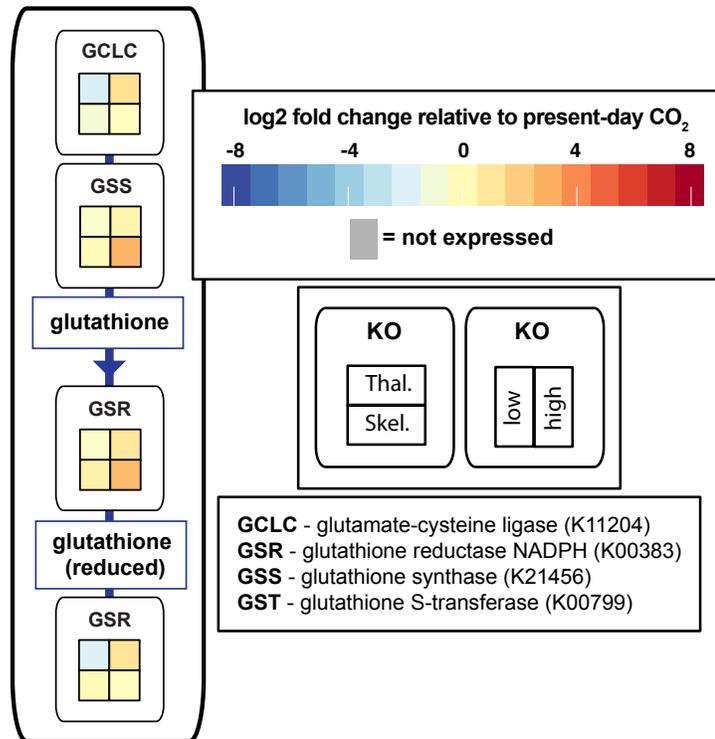


Figure 7. A heat map showing broad metabolic rearrangements in glutathione metabolism in *Skeletonema* and *Thalassiosira* across CO<sub>2</sub> conditions. The color indicates the log<sub>2</sub>-fold change of total reads mapped per genus per KEGG KO in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> condition.

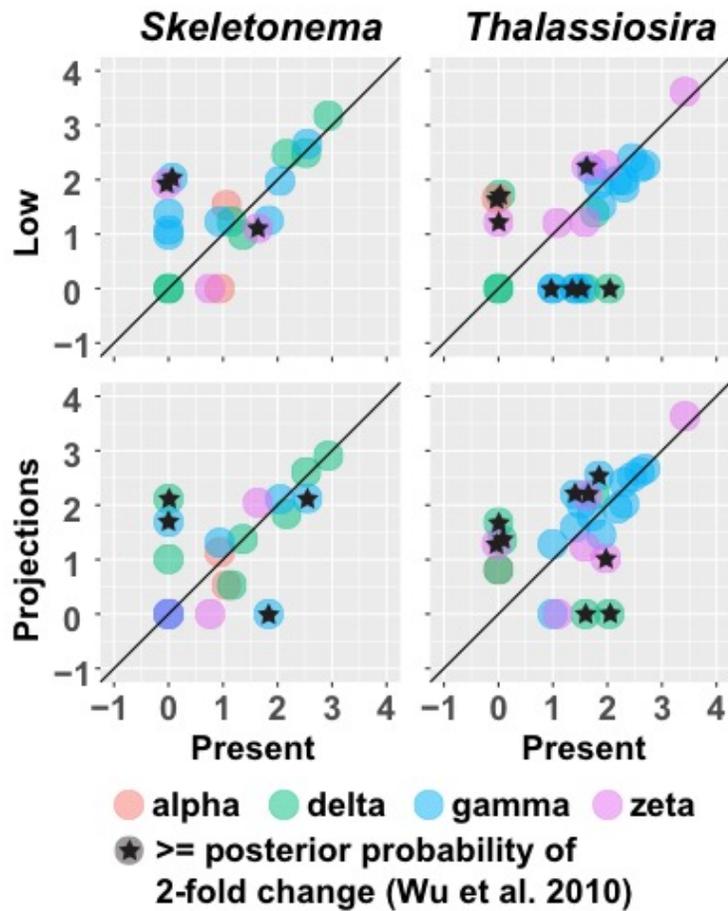


Figure 8. Expression values of the  $\alpha$ ,  $\delta$ ,  $\gamma$ , and  $\zeta$  carbonic anhydrases from *Skeletonema* spp. and *Thalassiosira* spp. diatoms. Each point indicates a gene and the color indicates which genera of diatom the sequence is from. The x-axis is the expression value in present day pCO<sub>2</sub> and the y-axis is the expression value in either low or high pCO<sub>2</sub>. The black line indicates the point at which the expression of a gene is the same in both conditions. Genes that were differentially expressed ( $\geq 0.95$  posterior probability of  $>2$ -fold change) are indicated by the black star (Wu et al. 2010).

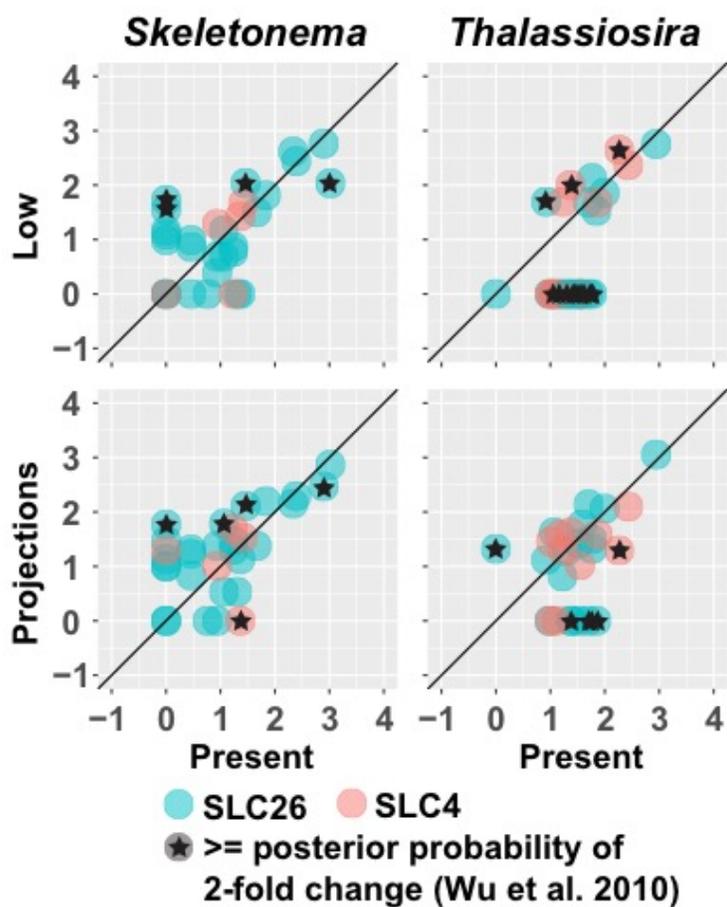


Figure 9. Expression values of the SLC4 and SLC26 bicarbonate transporters from *Skeletonema* spp. and *Thalassiosira* spp. diatoms. Each point indicates a gene and the color indicates which genera of diatom the sequence is from. The x-axis is the expression value in present day pCO<sub>2</sub> and the y-axis is the expression value in either low or high pCO<sub>2</sub>. The black line indicates the point at which the expression of a gene is the same in both conditions. Genes that were differentially expressed ( $\geq 0.95$  posterior probability of  $>2$ -fold change) are indicated by the black star) (Wu et al. 2010).

APPENDICES

APPENDIX 1

Amphiprora sp. CCMP467	2	>400
Asterionellopsis glacialis CCMP1581	1	
Attheya septentrionalis CCMP2084	2	
Aulacoseira subarctica CCAP 1002/5	1	
Chaetoceros affinis CCMP159	2	
Chaetoceros cf. neogracile RCC1993	1	
Chaetoceros debilis MM31A-1	6	
Chaetoceros dictyota CCMP1751	1	
Chaetoceros neogracile CCMP1317	6	
Corethron hystrix 308	1	
Corethron pennatum L29A3	1	
Cyclotella meneghiniana CCMP 338	112	
Cylindrotheca closterium KMMCC:B-181	4	
Detonula confervacea CCMP 353	46	
Ditylum brightwellii GSO103	1	
Ditylum brightwellii GSO104	2	
Ditylum brightwellii GSO105	2	
Ditylum brightwellii Pop1 (SS4)	3	
Ditylum brightwellii Pop2 (SS10)	1	
Extubocellulus spinifer CCMP396	65	
Fragilariopsis kerguelensis L2-C3	18	
Fragilariopsis kerguelensis L26-C5	30	
Grammatophora oceanica CCMP 410	3	
Leptocylindrus danicus CCMP1856	1	
Leptocylindrus danicus var. apora B651	1	
Leptocylindrus danicus var. danicus B650	2	
Minutocellus polymorphus NH13	19	20-200
Minutocellus polymorphus RCC2270	27	
Nitzschia punctata CCMP561	2	
Proboscia alata PI-D3	7	
Pseudo-nitzschia delicatissima B596	2	
Pseudo-nitzschia delicatissima UNC1205	1	
Pseudo-nitzschia fraudulenta WWA7	1	
Rhizosolenia setigera CCMP 1694	3	
Skeletonema costatum 1716	279	
Skeletonema dohrnii SkelB	159	
Skeletonema marinoi FE60	96	
Skeletonema marinoi FE7	188	
Skeletonema marinoi SM1012Den-03	248	
Skeletonema marinoi SM1012Hels-07	179	
Skeletonema marinoi UNC1201	67	<20
Skeletonema marinoi skelA	233	
Skeletonema menzelii CCMP793	117	
Thalassionema nitzschioides L26-B	2	
Thalassiosira antarctica CCMP982	61	
Thalassiosira gravida Gmp14c1	4	
Thalassiosira miniscula CCMP1093	26	
Thalassiosira oceanica CCMP1005	602	
Thalassiosira punctigera Tpunct2005C2	6	
Thalassiosira rotula CCMP3096	3	
Thalassiosira rotula GSO102	5	
Thalassiosira sp. NH16	30	
Thalassiosira weissflogii CCMP1010	13	
Thalassiosira weissflogii CCMP1336	1	
Thalassiothrix antarctica L6-D1	3	

Copy #

Supplemental Figure 2. Sel1/MYND zinc finger domain family homologs from diatoms in the MMETSP database. The color of each box indicates the number of homologs found in each species.

Supplemental Table 1. Detailed table for *T. oceanica* and *T. weissflogii* showing the KEGG orthology or protein domain information assigned by KEGG automated annotation server or InterProScan.

Supplemental Table 2. Detailed table for *T. oceanica* and *T. weissflogii* showing the posterior probability of a 2-fold up or downregulation in silicon limitation relative to the nutrient replete condition, the gene length in nucleotides, and the expression values in nutrient replete (Re), silicon limited (Si), iron limited (Fe), and nitrate limited (N) conditions in RPKM and total reads mapped.

Supplemental Table 3. Hierarchical cluster assignments for the silicon-sensitive genes from *T. oceanica* and *T. weissflogii* and the log<sub>2</sub> fold change expression value in silicon (Si), iron (Fe), or nitrate (N) limited conditions relative to the nutrient replete condition.

Supplemental Table 4. Silix gene family assignments for the translated proteins from the *T. oceanica*, *T. weissflogii*, and *T. rotula* transcriptomes as well as the translated proteins from the *T. pseudonana* genome (filtered set of gene models from Thaps3 and Thaps3\_bd). Each Silix gene family assignment starts with the prefix “FAM”. The genes from the Thaps3 data start with a “jgi\_THApse\_” prefix, followed by the gene number assigned by JGI. The genes from the Thaps3\_bd data start with a “jgi\_THApbd\_” prefix, followed by the gene number assigned by JGI.

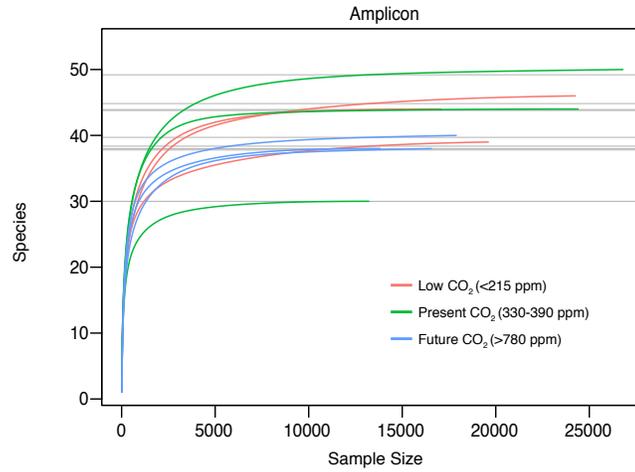
## APPENDIX 2

Supplemental Table 1. All of the significantly differentially expressed genes across conditions and transcriptomes ( $\geq 0.95$  posterior probability of a 2-fold change in ASC) (Wu et al. 2010). The KO and description columns are from the KEGG annotations. The gene\_ID column entries are formatted such that the first 6 letters indicate which genus (capital letters) species (lower case letters) transcriptome the sequence originates and the following letters and numbers indicate the gene ID within that transcriptome. The counts and RPKM values are the output from the CLC mapping. The next six columns are the posterior probabilities of a 2-fold change compared to the 320-390 ppm pCO<sub>2</sub> condition (“ambient”) as per ASC (Wu et al. 2010). “Low” refers to the <290 ppm condition, “projected” refers to the 690-1340 ppm pCO<sub>2</sub> condition, and “max” refers to the 2900-5100 ppm pCO<sub>2</sub> condition. The last 8 columns are from the InterProScan output.

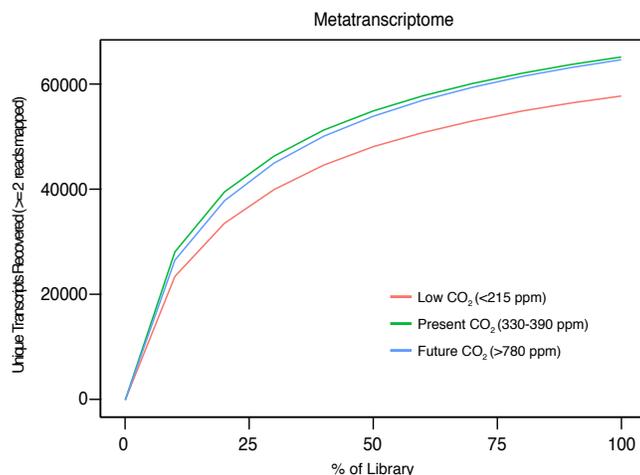
Supplemental Table 2. A summary of the enzyme type, predicted localization, expression value, and ASC posterior probability of a 2 fold change (Wu et al. 2010) for each gene hypothesized to play a role in the diatom CCM. This includes bicarbonate transporters, carbonic anhydrases, and genes thought to be involved in diatom C4 metabolism. The gene\_ID column is formatted such that the first 6 letters indicate which genus (capital letters) species (lower case letters) transcriptome the sequence originates and the following letters and numbers indicate the gene ID within that transcriptome. The additional worksheets provide the outputs from the different localization predictions and the “Predicted Localization” column is the interpretation of these results.

Supplemental Table 3. Selected putative CCM genes from the *T. oceanica* transcriptome (this study) and their predicted location when compared to the *T. oceanica* genome (Lommer et al. 2012). Genes included in this table include bicarbonate transporters, carbonic anhydrases, and genes thought to be involved in diatom C4 metabolism.

### APPENDIX 3



Supplemental Figure 1. A rarefaction curve of the 18S V4 rDNA amplicons. The x-axis indicates the number of amplicons sampled, and the y-axis indicates the number of species recovered. The color of the line indicates which pCO<sub>2</sub> condition is being shown. Each triplicate carboy is represented.



Supplemental Figure 2. A rarefaction curve of the metatranscriptome reads mapped to a subset of transcripts from *Skeletonema* spp. and *Thalassiosira* spp.. The x-axis indicates the percentage of the library that was sampled, and the y-axis indicates the number of unique transcripts recovered with at least 2 reads mapped (BWA MEM MAPQ60 filtering). The color of the line indicates which pCO<sub>2</sub> condition is being shown. Replicate carboys were pooled before sequencing, so there is a single line for each CO<sub>2</sub> conditions' transcriptome library.

Supplemental Table 1: Top: Total number of reads remaining in each library post-trimming. Middle: Total number of reads that mapped in DIAMOND to the entire MMETSP database, only the diatoms in the MMETSP database, and only to *Skeletonema* spp. or *Thalassiosira* spp. diatoms in the MMETSP database. For the DIAMOND mapping, the forward and reverse reads were mapped separately and the values shown here is the sum of the forward and reverse read counts. Bottom: Number of read pairs mapped to *Skeletonema* spp. or *Thalassiosira* spp. using a more stringent BWA mapping and MAPQ60 filtering in SAMtools.

	low	present	high
post-trim library sizes	13,831,711	25,037,661	18,501,137
reads mapped to MMETSP (DIAMOND)	73,566,310	143,481,946	159,523,246
reads mapped to MMETSP diatoms (DIAMOND)	54,032,252	91,529,538	102,295,960
reads mapped to MMETSP <i>Skeletonema</i> (DIAMOND)	25,036,845	24,748,686	27,215,604
reads mapped to MMETSP <i>Thalassiosira</i> (DIAMOND)	10,706,149	26,492,162	28,780,506
reads mapped to <i>Skeletonema</i> (BWA MAPQ60)	915,475	617,662	547,298
reads mapped to <i>Thalassiosira</i> (BWA MAPQ60)	826,575	1,263,129	1,069,587

Supplemental Table 2. All of the KEGG KOs from this experiment. The first two columns are the KEGG KO ID and description. The additional columns indicate the log<sub>2</sub>-fold change in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> for *Skeletonema* spp. and *Thalassiosira* spp.. Also shown are the normalized expression values, the posterior probability of a KO being differentially expressed (Wu et al. 2010), and higher-level information about which metabolic processes each KO is involved in.

Supplemental Table 3. All of the MMETSP genes detected in this experiment (Keeling et al. 2014). Columns shown indicate the CAMNT or CAMPEP ID for each gene, the MMETSP grindstone ID, genus, and species that the sequenced originated from, the posterior probability of a log<sub>2</sub>-fold change in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> (Wu et al. 2010), the normalized reads and TPM values, the log<sub>2</sub>-fold change values in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub>, and higher-level information about which metabolic processes each KO is involved in.

Supplemental Table 4. A table summarizing all of the carbonic anhydrases and bicarbonate transporters in this dataset. Columns shown indicate the CAMNT or CAMPEP ID for each gene, the genus, and the species each sequence originated from. Also shown is the posterior probability of a log<sub>2</sub>-fold change in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub> (Wu et al. 2010), the normalized reads, the log<sub>2</sub>-fold change values in low or high CO<sub>2</sub> relative to the present day CO<sub>2</sub>, the enzyme type for each transcript (alpha, delta, gamma, or zeta carbonic anhydrase; SLC4 or SLC26 bicarbonate transporter), and the log<sub>10</sub>-transformation of the normalized reads.