

UNIVERSITY OF OKLAHOMA  
GRADUATE COLLEGE

MACRONUTRIENTS SHAPE MICROBIAL COMMUNITIES, GENE EXPRESSION  
AND PROTEIN EVOLUTION

A DISSERTATION  
SUBMITTED TO THE GRADUATE FACULTY  
in partial fulfillment of the requirements for the  
Degree of  
DOCTOR OF PHILOSOPHY

By  
JOSHUA THOMAS COOPER  
Norman, Oklahoma  
2017

MACRONUTRIENTS SHAPE MICROBIAL COMMUNITIES, GENE EXPRESSION  
AND PROTEIN EVOLUTION

A DISSERTATION APPROVED FOR THE  
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY

---

Dr. Boris Wawrik, Chair

---

Dr. J. Phil Gibson

---

Dr. Anne K. Dunn

---

Dr. John Paul Masly

---

Dr. K. David Hambright

© Copyright by JOSHUA THOMAS COOPER 2017  
All Rights Reserved.

## **Acknowledgments**

I would like to thank my two advisors Dr. Boris Wawrik and Dr. J. Phil Gibson for helping me become a better scientist and better educator. I would also like to thank my committee members Dr. Anne K. Dunn, Dr. K. David Hambright, and Dr. J.P. Masly for providing valuable inputs that lead me to carefully consider my research questions. I would also like to thank Dr. J.P. Masly for the opportunity to coauthor a book chapter on the speciation of diatoms. It is still such a privilege that you believed in me and my crazy diatom ideas to form a concise chapter in addition to learn your style of writing has been a benefit to my professional development.

I'm also thankful for my first undergraduate research mentor, Dr. Miriam Steinitz-Kannan, now retired from Northern Kentucky University, who was the first to show the amazing wonders of pond scum. Who knew that studying diatoms and algae as an undergraduate would lead me all the way to a Ph.D. Miriam was a driving force behind my love of research and seeking answers to biological questions. I am also deeply indebted to the many faculty members at my undergraduate institution, Northern Kentucky University for providing a solid foundation to build upon.

I would like to thank the Oklahoma Biological Survey staff (Trina Steil and Rannell Madding) for the continued support during my time as Ph.D. student. I'm also indebted to my colleagues Huynh Le, Chris Marks, Nate Losey, Lindsey O'Neil, and Brian Harriman, Daniel Jones, Zach Meyers, Jamie Johnson-Duffner for their friendship and collegiate discussions.

Also I thank my parents Roger and Sharon Cooper have always been supportive of my adventure into graduate school. I also thank my grandmother Jeanette Cooper,

who spent hours with me in my earlier years taking me to the Cincinnati Public Library. There I was able to cultivate my quest for knowledge, and explore any topic I became interested in, even if for brief moments.

Lastly I would like to thank the Department of Microbiology and Plant Biology for the teaching assistant support granted to me during my time at the University of Oklahoma.

Included within this dissertation are already published peer reviewed manuscripts, which represent Chapters 1 and 2. My contribution and the contributions of the other authors are outlined at the beginning of each chapter.

## Table of Contents

|   |      |
|---|------|
| Acknowledgments .....   | iv   |
| List of Tables .....  | viii |
| List of Figure .....  | x    |
| Abstract .....  | xiii |
| Chapter 1: Analysis of Ammonium, Nitrate, and Urea Uptake by Marine Pelagic Bacteria and Archaea during the Arctic Summer and Winter Seasons via Stable Isotope Probing ..... | 1    |
| Abstract .....  | 3    |
| Introduction .....  | 4    |
| Material and Methods .....  | 7    |
| Results .....   | 13   |
| Discussion .....  | 17   |
| Acknowledgements .....  | 27   |
| References .....  | 28   |
| Chapter 2: Transcriptome Analysis of <i>Scrippsiella trochoidea</i> CCMP 3099 Reveals Physiological Changes Related to Nitrate Depletion .....                                | 41   |
| Abstract .....  | 42   |
| Introduction .....  | 44   |
| Materials and Methods .....   | 48   |
| Results .....   | 58   |
| Discussion .....  | 68   |
| Conclusion .....  | 78   |
| Conflict of Interest Statement .....  | 78   |
| Acknowledgements .....  | 78   |
| References .....  | 80   |
| Chapter 3: Trophic Differences Reflected in the Elemental Stoichiometry from Transcriptomes and Proteomes of Eukaryotic Phytoplankton .....                                   | 108  |
| Abstract .....  | 108  |
| Introduction .....  | 110  |
| Methods and Materials .....   | 117  |
| Results .....   | 129  |
| Discussion .....  | 136  |
| Acknowledgements .....  | 146  |
| References .....  | 147  |

|  |     |
|--|-----|
| Chapter 4: A Case Study Approach to Teaching Introductory Biology Undergraduates<br>the Application of Genomic Data Analysis in Ecological Studies ..... | 171 |
| Preface .....  | 171 |
| Case Teaching Notes .....  | 174 |
| Introduction .....   | 175 |
| Objectives .....   | 176 |
| Outcome .....  | 176 |
| Competency .....   | 177 |
| Misconceptions Addressed .....   | 177 |
| Background .....   | 173 |
| Classroom Management/Blocks of Analysis .....  | 180 |
| Teaching the Case .....  | 180 |
| Questions .....  | 180 |
| Slides .....   | 181 |
| Answers to Questions .....   | 192 |
| References .....   | 195 |
| Slide Image Credits .....  | 197 |
| Appendices and Supplemental Information .....  | 205 |

## List of Tables

### Chapter 1:

- Table 1. Concentrations and uptake rates of ammonium ( $\text{NH}_4^+$ ), nitrate ( $\text{NO}_3^-$ ) and urea in near-shore waters of the Alaskan Arctic during January (winter) and August (summer). Uptake rates were determined with GF/F filters (nominal pore size of  $0.7 \mu\text{m}$ ). Where the standard deviation (SD) is noted as not available (N.A.) it implies that  $n=1$  for that measurement ..... 35
- Table 2. Percent isotopic labeling of DNA with  $^{15}\text{N}$  or  $^{13}\text{C}$  calculated from qPCR data of SIP fractions. Bold percentages are above the 30% and 15% thresholds used to define  $^{15}\text{N}$ - and  $^{13}\text{C}$ -substrate uptake, respectively. .... 36
- Table 3. Percent isotopic labeling of DNA with  $^{15}\text{N}$  or  $^{13}\text{C}$  derived from 16S rRNA gene sequenced data obtained for SIP fractions of winter samples. Bold percentages are  $\geq 30\%$  and  $15\%$ , the thresholds used to define  $^{15}\text{N}$ - and  $^{13}\text{C}$ -substrate uptake, respectively ..... 37

### Chapter 2:

- Table 1A. Significantly enriched GO terms for Biological Process  $P < 0.01$  with Positive Log2 fold change indicating up-regulation in response to N-depletion. See Supplemental Table 7 for complete list)..... 102
- Table 1B. Significantly enriched GO terms for Biological Process  $P < 0.01$  with Negative Log2 fold change, indicating down-regulation in response to N depletion (See Supplemental Table 8 for complete list)..... 104
- Table 2. Transcriptome gene hits to the saxitoxin biosynthesis gene cluster in *Cylindrospermopsis raciborskii* T3. Approximately 50% of the biosynthesis cluster had significant hits to *S. trochoidea* (17 of the 34)..... 106

### Chapter 3:

- Table 1. (A) Median calculated values. (B) Weighted average by transcripts per million (TPM) mapped reads to the transcriptome. Phylogenetic signal between traits assessed on the global MMETSP dataset and using Pagel's  $\lambda$ . Values of  $\lambda = 1$  suggest the trait has phylogenetic signal and the positive  $\Delta\text{AICc}$  describes the model difference between the BM (Brownian Model) and a Star phylogeny representing no signal. Positive  $\Delta\text{AICc}$  are indicating that  $\lambda \neq 0$  and the more complex (BM) model is preferred. NARSC: Number of Nitrogen atoms in the amino acid residue side-chain, while CARSC Number of carbon atoms in the



amino acid residue side-chain, GC: guanine+cytosine percent, NAB3: nitrogen content of the 3rd codon position in the open reading frame ..... 167

Table 2. Values of Moran's I among Medians and Weighted Averages (SD = standard deviation). Values in gray boxes were not significant. Positive observed values indicate the tips of the phylogeny more similar in value than by chance. Top boxes show all proteins, then filtered to contain BUSCO homologs. The last row represents the all proteins filtered to include Yang and Smith homologs ..... 168

Table 3. PGLS correlations among the different evolutionary models tested, and NARSC. (A) Includes all available cell size data from culture collections, MMETSP metadata, and literature. (B) Data filtered to include only data from culture collections and deposited MMETSP metadata. All models favored the OU model. The ANOVA testing of the OU model is below each correlation along with its *P*-value ..... 169

Table 4. (A) PGLS correlation of NARSC as the dependent variable on the independent log (nitrate) of the ocean of its original isolation, extracted from GIS layers. (B) PGLS correlation of NARSC as the dependent variable on the independent log (phosphate) of the ocean of its original isolation extracted from GIS layers ..... 170

## List of Figures

### Chapter 1:

- Figure 1. Phylogenetic analysis of archaeal (A & B) and bacterial (C & D) 16S rRNA gene sequences for arctic summer (A&C) and winter (B & D) seasons. PCR products were barcoded and sequenced using the Illumina MiSeq platform. Operational Taxonomic Units (OTU) were defined at the 5% identity level for bacteria. Representative sequences from each OTU were chosen at random and their phylogenetic affiliations were determined using QIIME (Caporaso, *et al.*, 2010) using the SILVA core alignment ([www.arb-silva.de](http://www.arb-silva.de)). Underlined taxa indicate phylum (division) level classification. Non-underlined taxa are family level assignments (not all are shown) ..... 39
- Figure 2. Shown are examples of qPCR analysis of SIP gradient fractions for bacterial and archaeal 16S rRNA gene copies. Relative quantities detected in each fraction are shown as a function of density. All data are normalized to the highest observed quantities and are hence shown as a ratio, where 1 equals the highest observed value. Error bars indicate one standard deviation calculated from three replicate qPCR measurements. The horizontal lines indicate the threshold above which quantities were integrated to calculate average DNA density and percent incorporation. (A) Comparison of winter and summer <sup>14</sup>N-urea (○) and <sup>15</sup>N-urea (●) treatments. (B) Dark carbon fixation SIP experiment showing <sup>12</sup>C-bicarbonate (◐) and <sup>13</sup>C-bicarbonate (●) treatments for bacteria and archaea respectively..... 40

### Chapter 2:

- Figure 1. (A) Nitrate measurement illustrating low starting N concentration in the N-limited culture and depletion over time until day 7. Cultures, starved for 24 hours, were sampled on the 8<sup>th</sup> day. The remaining 300 mL were split into 150 mL cultures, and additional nitrate was added to one of the flasks originating from the N-limited treatment. (B) Phosphate concentration illustrating lower starting concentration of the P-limited culture, and depletion of P over time until day 7. The culture was starved 24 hours, and sampled on the 8<sup>th</sup> day. The remaining culture was split into two equal volumes and phosphate was to one of the flasks originating from the P-limited treatment ..... 93
- Figure 2. Cell densities of cultures for each of the treatments. Error bars indicate one standard deviation. Solid circles indicate the replete (control) culture. Open circles are used to plot the N-limited culture, while triangles are used to plot the P-limited cultures. Dotted lines indicate the sub-cultures that were spiked with nutrients after sampling..... 95

|  |     |
|--|-----|
| Figure 3. Krona Plots of the blastx-hits. (A) Eukaryota and distribution of hits within, showing majority assigned to Alveolata. (B) Alveolata with most reads assigned to the <i>Perkinsus marinus</i> .....  | 96  |
| Figure 4. (A) Log-Log plot of replete to the N-depletion treatment showing the number of differentially expressed genes in red (down-regulated) and blue (up-regulated, and those that are no DE are in gray-black with p-adjusted to an FDR<0.1. (B) Log-Log Plot of replete to the P-limited treatment.....  | 97  |
| Figure 5. Cellular overview of genes detected as differentially expressed and potential role it altering the metabolic pathways of <i>S. trochoidea</i> under N-limitation, relative to replete conditions. The glyoxylate cycle is not shown connecting the mitochondrial and plastid pathways for clarity. Blue and Red circles indicate up- and down regulation in response to N-limitation respectively, and indicate the location of genes detected as differentially expressed. AAT, Amino acid transporters; ALP, Alkaline phosphatase; AMT, Ammonium transporter; ARG, Arginase; BG, Beta-glucanase and 1,4-beta-D-glucan cellobiohydrolase; CCP, Cytochrome c peroxidase; CCD1, Carotenoid 9, 10(9',10')-cleavage dioxygenase 1; CEL, Cellulase/Endoglucanase 5 precursor; CP, Carbamoyl phosphate synthetase; DOP, Dissolved organic phosphate; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; ECP, Extracellular peptidases/protease; GBG, Glucan 1,3-beta-glucosidase; GDH2, Glutamate dehydrogenase 2; GSII, Glutamine synthetase II; GSIII, Glutamine synthetase III; GST, Glutathione S-transferase; GOGAT, Glutamate synthetase; NIA, Nitrate reductase; NiR, Nitrite reductase; NRT, Nitrate/Nitrite transporter; STP, Sulfate transporter; URE, Urease; UTP, Urea transporter; XUV, Xanthine/Uric Acid transporter; ZT, Zinc transporter..... | 98  |
| Figure 6. Phylogenetic relationships among saxitoxin <i>sxtA</i> from <i>S. trochoidea</i> and close blastp hits to NCBI-NR, 20 Maximum likelihood searches and 500 Rapid Bootstraps using RAxML. <i>sxtA</i> hits from <i>S. trochoidea</i> form a clade with others hits from dinoflagellates. These are related to the sister clade of <i>sxtA</i> from cyanobacteria which are known to produce toxins .....   | 100 |
| Figure 7. Gene arrangement of saxitoxin gene cluster in <i>Cylindrospermopsis raciborskii</i> with homologs of <i>Alexandrium tamarense</i> and <i>Scrippsiella trochoidea</i> shaded to demonstrate shared gene similarity in both toxic and non-toxic dinoflagellates and the known cyanobacterial saxitoxin pathway (Adapted from (Kellmann et al., 2008), and <i>sxt</i> gene hits of <i>Alexandrium tamarense</i> plotted from (Hackett et al., 2013) in addition <i>sxt</i> gene hits from this study. ....  | 101 |

### Chapter 3:

|  |
|--|
| Figure 1. Resolved and time calibrated phylogeny of MMETSP samples based on nearly complete 18S rRNA sequences. Red and Green plastid lineages are |
|--|

|   |     |
|---|-----|
| colored in red and green respectively. Time in represented in millions of years and dates represent major geologic periods .....  | 161 |
| Figure 2. (A) This figure shows boxplots of the overall NARSC values among the Trophic groups, the red line behind the boxplot identifies the median among all NARSC values .....   | 162 |
| Figure 2. (B) This shows a boxplot with overall CARSC values among all trophic strategies. The red line behind both figures represents the median of the values within NARSC and CARSC.....   | 163 |
| Figure 3. Moran's I Correlogram showing median values of all MMETSP transcriptome and proteomes. Phylogenetic distance is represented in time (mya). Filled circles represent significant Moran's I, while unfilled are non-significant. NARSC shows trend upward indicating more similarity at shorter branches/time. The other measurements show gradual phylogenetic dissimilarity with time. Only CARSC is significant over long distances. Measurements of N and C stoichiometry show a trend of non-significance with larger phylogenetic distances. .... | 164 |
| Figure 4. PGLS regression of cell size data with complete literature cases and limited to values obtained from cell culture facilities with the OU model favored. A and B represent the maximum and minimum cell diameter of culture collection, metadata and literature values (N= 258). C and D illustrate the reduced dataset only using culture collection and provided metadata N=179). Significance of the regression plotted on corner of each.....  | 165 |
| Figure 5. PGLS regression of natural log transformed nitrate and phosphate concentrations to median NARSC, of the MMETSP proteomes and untransformed G+C content of the transcriptome (OU model). (A) Correlation of NARSC and environmental nitrate, (B) Correlation of NARSC and G+C content of the transcriptome. (C) Correlation of NARSC and environmental phosphate, and (D) Correlation of NARSC and Nitrate:Phosphate ratio. Significance of the regression plotted on corner of each.....  | 166 |

## Abstract

Nutrient limitation of the principle macronutrients carbon, nitrogen and phosphorus are known to influence community structure, success of individual species, and over long enough time could, in theory, shape the evolution of proteins organisms use to cope with nutrient stress. This dissertation explores macronutrients incorporation into bacterial communities, how organisms modulate gene expression to cope with periodic nutrient stress, and how long-term limitation might shape cellular stoichiometry to reduce biochemical nutrient demand. In the first chapter, arctic natural microbial communities are investigated, and a strong seasonal shift of bacterial and archaeal N utilization from ammonium during the summer to urea during the winter is demonstrated via  $^{15}\text{N}$ -based stable isotope probing (SIP). In combination with collaborative  $^{13}\text{C}$ -bicarbonate based SIP studies, these data point to the potential for urea fueled nitrification as an important source of primary production during the arctic winter. The second chapter examines the nutrient limited transcriptome of a harmful bloom forming algae, *Scrippsiella trochoidea* CCMP 3099 to investigate its cellular response to nitrogen or phosphorus stress. Transcriptome data indicates that N limitation in *S. trochoidea* modulates gene expression to compensate for oxidative stress and appears to switch from inorganic nitrate metabolism to dissolved organic sources. The third chapter aims to understand how, over long time scales of phytoplankton and protists evolution, N limitation might alter the stoichiometry of the proteome to reduce overall nutrient utilization. It was tested whether the nutritional mode (autotrophy, mixotrophy, and heterotrophy) might be a predictor of the overall balance of

macroelements in predicted protein products. The hypothesis that organisms living in more N limiting environments produce N-deplete protein products (based on side-chain chemistry), is rejected. Conversely, predicted proteins in the transcriptomes of mixotrophs appear enriched in amino acids with greater C content. The stoichiometry of the *in silico* translated proteome has a weak correlation to environmental nutrients (not significant for nitrate, but significant for phosphate). The last chapter is an extension of the primary research goals with respect to algal transcriptomes put forth in this dissertation. The chapter's aim is to integrate scholarship and teaching by introducing cutting edge research results into a case study designed for introductory biology students to teach the central dogma of molecular biology in terms of genomics technologies. This chapter incorporates "Central Dogma of Molecular Biology", "big data", "cells as systems", and the "flow of information" with societal issues and problem solving of the harmful bloom forming dinoflagellate *Karenia brevis*.

## **Chapter 1: Urea Uptake and Carbon Fixation by Marine Pelagic Bacteria and Archaea During the Arctic Summer and Winter Seasons**

Tara L. Connelly<sup>1</sup>, Steven E. Baer<sup>2,3</sup>, **Joshua T. Cooper**<sup>4</sup>, Deborah A. Bronk<sup>3</sup>, and Boris Wawrik<sup>4,\*</sup>

<sup>1</sup> The University of Texas at Austin, Marine Science Institute, 750 Channel View Drive  
Port Aransas, TX 78373

<sup>2</sup> Current Address: Bigelow Laboratory for Ocean Sciences, 60 Bigelow Drive, East  
Boothbay, ME 04544

<sup>3</sup> Virginia Institute of Marine Science, College of William & Mary, P.O. Box 1346,  
Gloucester Point, VA 23062.

<sup>4</sup> Department of Botany and Microbiology, University of Oklahoma, George Lynn  
Cross Hall, 770 Van Vleet Oval, Norman, Oklahoma, 73019, USA.

\*Corresponding author

This chapter has been published essentially in this form in *Applied and Environmental Microbiology* and is formatted according to journal specifications. The paper can be located under the following citation:

Connelly, Tara. L., Baer, Steven. E., **Cooper, Joshua. T.**, Bronk, Deborah. A., & Wawrik, Boris. (2014). Urea Uptake and Carbon Fixation by Marine Pelagic Bacteria and Archaea during the Arctic Summer and Winter Seasons. *Applied and Environmental Microbiology*, 80 (19), 6013–6022. doi.org/10.1128/AEM.01431-14

Author contributions to the AEM paper:

BW and DAB conceived the study and funded the work through NSF funded research grants OCE 0961900, ARC 0910252, and ARC 0909839. SEB, TLC, and DAB collected samples. DAB contributed nutrient data and isotope tracer uptake rate measurements. SEB and TLC conducted  $^{13}\text{C}$ -based stable isotope probing studies to demonstrate autotrophic activities of microbial populations. JTC conducted the  $^{15}\text{N}$ -based stable isotope probing studies and analyzed microbial community data. All co-authors contributed to the writing the final manuscript



## **Abstract**

How Arctic climate change might translate into alterations of biogeochemical cycles of carbon (C) and nitrogen (N) with respect to inorganic and organic N utilization is not well understood. This study combined  $^{15}\text{N}$ -uptake rate measurements for ammonium, nitrate, and urea with  $^{15}\text{N}$ - and  $^{13}\text{C}$ -based DNA stable isotope probing (SIP). The objective was to investigate the identities of active prokaryotic plankton and their role in N and C uptake during the Arctic summer and winter seasons. We hypothesized that bacteria and archaea would successfully compete for nitrate and urea during the Arctic winter, but not the summer when phytoplankton dominate uptake of these nitrogen sources. Samples were collected at a coastal station near Barrow Alaska during August and January. During both seasons, ammonium uptake rates were greater than those for nitrate or urea, and nitrate uptake rates remained lower than those for ammonium or urea. SIP experiments indicated a strong seasonal shift of bacterial and archaeal N utilization from ammonium during the summer to urea during the winter. Analysis of 16S rRNA gene sequences obtained by Illumina sequencing of PCR products for each SIP fraction implicated Marine Group I Crenarchaea (MGIC) as well as Betaproteobacteria, Firmicutes, SAR11, and SAR324 in N uptake from urea during the winter. Similarly,  $^{13}\text{C}$  SIP data suggested dark carbon fixation for MGIC as well as several Proteobacterial lineages and the Firmicutes. These data are consistent with urea-fueled nitrification by polar archaea and bacteria, which may be advantageous under dark conditions.

## **Introduction**

The Arctic is already experiencing the impacts of global climate change, which has the potential to disrupt the ecology of the Arctic Ocean, causing broad changes in the physical, chemical and biological realms (1-5). How such changes might translate into alterations of ecosystem dynamics as well as the overall balances and rates of biogeochemical cycles of carbon (C) and nitrogen (N) is not well understood (6, 7). Generally, short day lengths and sea-ice coverage during the Arctic winter limit light and phytoplankton primary production, while summer is characterized by episodic phytoplankton blooms following sea-ice melt (8). High levels of primary productivity during the summer are thought to be sustained through buildup of  $\text{NO}_3^-$  in the water column during the dark winter period, as well as inputs from ice melt and allochthonous riverine nutrient sources (9). Production could additionally be augmented by dissolved organic nitrogen (DON), which can represent between 18 and 85% of the total dissolved N pool in coastal and open ocean surface water (10, 11). The interplay between inorganic and organic N utilization, with respect to heterotrophic versus autotrophic activities, could be an important contributor to biological production, but remains poorly resolved, especially in Arctic environments (12).

Among DON compounds, urea has long been recognized as an important N source in tropical, sub-tropical, and temperate marine environments (13-16). Urea usually occurs in nM levels in the open oceans, but can be found at concentrations of up to 50  $\mu\text{M}$  in coastal ecosystems (17), where it can be an important N substrate that promotes large seasonal blooms of phytoplankton (18). The importance of urea at high latitudes is, however, less well understood. In addition to riverine input, other natural

sources of urea in the Arctic can include excretion and sloppy feeding by zooplankton (19) and inputs from melting of seasonal fast ice (20). Production of urea has also been attributed to sediment-associated bacteria, which may mediate urea release into the water column via thermal or wind-driven mixing (21). In the Canadian Arctic, urea has been observed to account for >50% of total dissolved nitrogen (TDN; 22). The same study reports urea uptake rates that mimicked the distributional patterns of urea concentrations, while accounting for approximately 32% of N productivity (nitrate, ammonium, and urea). A more recent study in the Beaufort Sea found that urea supplied almost half of the phytoplankton N uptake annually (23), and on a seasonal basis it was reported that urea uptake increased relative to other N substrates as the year progressed from winter to spring to summer (24). In the Northern Baffin Bay cycloheximide and streptomycin were utilized as inhibitors and it was found that urea was primarily consumed by phytoplankton (58-95%) but may also be utilized by bacteria (5-42%) (25). Collectively these studies suggest that urea is likely an important source of N in Arctic systems. It remains uncertain, however, what the dynamics of competition for urea between phytoplankton and bacteria are at the community level and which populations of cells successfully compete for urea N under different conditions. Size fractionation experiments often retain significant numbers of bacteria in the 'phytoplankton fraction' (traditionally collected on GF/F filters with a nominal pore size of 0.7  $\mu\text{m}$ ), and they provide no phylogenetic information about active microbial community members.

The ability of pelagic bacteria and archaea to fix carbon independent of light in oxygenic waters is becoming more widely recognized (26-29), especially in deeper

oceanic waters (30-32). However, the extent that dark carbon fixation is occurring in the world's oceans and how important this metabolism is for the life strategy of specific taxa is still unknown. It is hypothesized that dark carbon fixation can be important in compensating for metabolic imbalances under oligotrophic conditions. Since much of the world's oceans are oligotrophic, the significance of dark carbon fixation could therefore be large. This may be especially true under dark winter conditions in the Arctic when photosynthesis rates are low. Previously, Alonso-Sáez *et al.* (27) found that certain taxa of bacteria (e.g., *Oleispira* and *Pseudoalteromonas-Colwellia*) collected from shelf waters in the Arctic and cultured in aged seawater had the potential to fix carbon. The authors concluded that heterotrophs were primarily responsible for the observed bicarbonate uptake and proposed that this metabolism would be advantageous for survival during periods of low nutrient availability, such as winter.

In this study, we combined DIN and urea uptake rate measurements with  $^{13}\text{C}$  and  $^{15}\text{N}$  DNA-SIP in order to investigate the role of prokaryotic plankton in C and N cycling during the Arctic summer and winter seasons. We hypothesized that both bacteria and archaea would successfully compete for N from  $\text{NO}_3^-$  and urea during the Arctic winter, but not the summer when phytoplankton dominate the absolute uptake of these N sources. Further,  $^{13}\text{C}$ -based SIP was used to investigate the *in situ* capability of both bacteria and archaea to incorporate carbon from bicarbonate into DNA to demonstrate their potential involvement in fixation of carbon during the dark winter months.

## **Material and Methods**

### **Field sample collection**

Sampling was performed at 71°20.660' N, 156°41.416' W, which is approximately 2.5 km northwest of Barrow, Alaska. To capture the extreme Arctic light and physical conditions, sampling took place during summer (August 2011) and winter (January 2012). During the ice-covered seasons of winter, a small hole was drilled through the ice to allow access for a low-pressure electric pump for sample collection. Every effort was made to reduce stress on the organisms by limiting light and temperature changes. Sampling depth was at 1 m depth in the winter and at 8 m in the summer. A small tent was erected and heated to approximately -1°C (near the temperature of the ambient seawater) to prevent the pumped seawater and sampling equipment from freezing. During the summer, samples were collected onboard a small boat via the same pump arrangement at the same location as determined by GPS coordinates.

### **<sup>15</sup>N and <sup>13</sup>C Additions**

Water was collected into a series of 2 L acid-washed PETG bottles. A subset was used to determine ambient nutrient concentrations. Samples for SIP and uptake rate incubations were each run in duplicate and were inoculated with additions of unlabeled (<sup>14</sup>N) or labeled (<sup>15</sup>N) NH<sub>4</sub><sup>+</sup>, NO<sub>3</sub><sup>-</sup>, and urea (>98% <sup>15</sup>N). Previously reported ambient concentrations were used to establish N additions for uptake rate incubations. Since DNA stable isotope probing (SIP) requires substantial isotopic labeling, incubations for SIP samples were made with saturating additions of 2.0 μmol N L<sup>-1</sup> of NH<sub>4</sub><sup>+</sup>, NO<sub>3</sub><sup>-</sup>, and urea during summer. Winter additions were 3.25 μmol N L<sup>-1</sup> for NH<sub>4</sub><sup>+</sup> or urea, and 7.7

$\mu\text{mol N L}^{-1}$  of  $\text{NO}_3^-$ . For dark carbon fixation experiments, duplicate sets of samples were incubated with 200 mM of either labeled ( $^{13}\text{C}$ ) or unlabeled ( $^{12}\text{C}$ ) bicarbonate. The bottles were then surrounded by ambient seawater, placed in insulated coolers, and brought to the laboratory within one hour of collection to prevent freezing. Samples were incubated in a temperature-controlled chamber for 24 hours at ambient water temperature ( $+4.7^\circ\text{C}$  in summer;  $-1.8^\circ\text{C}$  in winter). To mimic spectral attenuation from the field during summer, light levels were maintained by GAM Hynix blue films and confirmed using a Li-COR PAR sensor. Winter samples were incubated in the dark. At the end of incubations, samples were filtered separately for uptake rates and SIP analyses, and water was collected for nutrient analyses. For uptake rates determinations, samples were filtered through Whatman GF/F (nominal pore size of  $0.7\ \mu\text{m}$ ) filters. The filters were placed in cryovials and frozen until analysis. For determination of nutrient concentrations at the end of incubations, the filtrate was poured into polypropylene tubes and frozen until analysis. Samples from SIP incubation bottles were filtered onto  $0.45\ \mu\text{m}$  Supor filters (PALL Life Sciences) and frozen in  $750\ \mu\text{L}$  STE buffer (1 M NaCl, 100 mM Tris·HCL pH 8.0, 10 mM EDTA pH 8.0).

### **Nutrient Analysis and Uptake Rates**

Nutrients were measured on ambient seawater and water incubated with labeled substrate in order to correct for isotope dilution in the uptake rate calculations. Concentrations of ammonium ( $\text{NH}_4^+$ ) were measured in triplicate using the phenol-hypochlorite method (33). Duplicates of nitrate ( $\text{NO}_3^-$ ) and nitrite ( $\text{NO}_2^-$ ) were measured on a Lachat QuikChem 8500 autoanalyzer (34). Urea was measured in duplicate using the manual monoxime method (35). A Europa GEO 20/20 mass spectrometer with an

ANCA autosampler was used to make isotopic measurements of  $^{15}\text{N}$  samples. Nitrogen uptake rates were calculated as per Dugdale and Goering (36), and carbon uptake rates as per Hama et al. (37). The  $\text{NH}_4^+$  pool was isolated at the end of the incubation by solid phase extraction (38, 39) and the  $^{15}\text{N}$  enrichment determined so that  $\text{NH}_4^+$  uptake rates could be corrected for isotope dilution and  $\text{NH}_4^+$  regeneration rates measured (40).

### **Stable Isotope Probing**

DNA extractions were conducted as previously described (41) by lysis with 75  $\mu\text{L}$  of 5% SDS and 20  $\mu\text{L}$  of 20  $\text{mg mL}^{-1}$  proteinase K for 30 min at 37 °C. Muffled glass beads (50–100 mg of 0.1 mm beads) were added, and samples were homogenized by bead-beating for two minutes before extraction with phenol/chloroform/isoamyl alcohol (25:24:1). DNA was subsequently precipitated with 0.7 volumes of isopropanol and pellets were suspended in 50  $\mu\text{L}$  of TE buffer (1 mM Tris pH 9.0, 10 mM EDTA pH 8.0). Cesium chloride (CsCl) density centrifugation and fractionation were conducted as previously described (41-44). Briefly, two micrograms of DNA in 0.40 ml TE buffer were spun for 48-72 hours in 4.7-ml polyallomer Optiseal tubes (Beckman) containing 4.35 mL of CsCl solution ( $1.701 \text{ g mL}^{-1}$ ) in gradient buffer (15 mM Tris-HCL pH 8.0, 15 mM KCL, 15 mM EDTA pH 8.0) in a Beckman VTI65.2 rotor at 140,000 x g. Thirty 150- $\mu\text{L}$  fractions were then collected from each tube using a fraction collector (Beckman) by displacing contents with mineral oil at a constant rate using a peristaltic pump. Densities for each fraction were calculated from their refractive index as described elsewhere (41, 43). After ethanol re-precipitation with 1.0  $\mu\text{L}$  of molecular grade glycogen, DNA was suspended in 30  $\mu\text{L}$  sterile, nuclease free DI water. These purified fractions served as template for qPCR and 16S PCR for Illumina

sequencing. We note that all direct comparisons made in this paper are for fractions and gradients from the same centrifuge run, which used the same batch of buffer.

Labeling was assessed by integrating the area of DNA and OTU abundance peaks observed in gradients and thereby estimating their average density. For bacteria and archaea, the proportional abundance from qPCR was used. Average densities for individual phylogenetic groups were calculated from 16S community OTU frequency data. To minimize biases introduced by baseline variability (i.e. because target DNA is typically found throughout gradients, and because the obtained density range can vary slightly among gradient runs), only the major peak of DNA was integrated. All fractions, which contained >20% of the maximum observed quantity were hence integrated (see Figure 2). Percent labeling was calculated by assuming that 100 % labeling of DNA with  $^{15}\text{N}$  and  $^{13}\text{C}$  would yield a density shift of  $0.016 \text{ g cm}^{-3}$  and  $0.036 \text{ g cm}^{-3}$  respectively (43). Sequential gradient fractions differed in average by  $\sim 0.0035 \pm 0.0001 \text{ g cm}^{-3}$ . Positive labeling was therefore interpreted to have occurred when shifts of >30% and >15% for  $^{15}\text{N}$  and  $^{13}\text{C}$  respectively were observed. This corresponds to density shift of approximately 1.5 fractions in our gradients, where complete (i.e. 100%) labeling would correspond to a shift of 4.6 and 10 fractions for  $^{15}\text{N}$  and  $^{13}\text{C}$  respectively.

#### **qPCR for 16S Bacteria and Archaea**

Bacterial and archaeal 16S rRNA gene copy numbers were determined for each SIP fractionation via quantitative PCR (qPCR). Bacterial 16S rRNA gene qPCR primers were 27F (5'-AGA GTT TGA TCM TGG CTC AG-3') and reverse primer 519R (5'-GWA TTA CCG CKG CTG-3'). Archaeal specific qPCR primers were 8AF



(5'-TCC GGT TGA TCC TGC C-3') and the reverse primer A344R (5'-TCG CGC CTG CTC CIC CCC GT-3'). SYBR green qPCR reactions were run in 30  $\mu$ L volumes using *Power SYBR® Green PCR Master Mix* (Applied Biosystems), 500 nM final concentration of each primer and 2  $\mu$ L of template DNA. Using an Applied Biosystems ABI 7300 Real Time PCR System, qPCR was conducted as follows: 50°C for 2 minutes and 95°C for 8 minutes, followed by 40 cycles of 95°C for 30 seconds, 55°C for 30 seconds. Genomic DNA of *Roseobacter dendrificans* Och114 was used as a standard for bacteria, while a plasmid containing the complete 16S gene of *Methanospirillum hungatei* JF-1 served as a standard for archaea.

### **16S community analysis**

Environmental genomic DNA and DNA from CsCl gradient fractions were amplified by targeting partial 16S rRNA genes with the universal (bacteria and archaea) primers S-D-Arch-0519-a-S-15 (5'-CAG CMG CCG CGG TAA-3') and S-D-Bact-785-a-A-21 (5'-GAC TAC HVG GGT ATC TAA TCC-3') as previously described (45). These primers do not amplify ribosomal RNA genes from eukaryotes, but cover 86.5% and 87.1% of bacterial and archaeal phyla, respectively. If one mismatch is allowed (which frequently occurs during PCR), 94.6% and 94.8% of bacterial and archaeal phyla are covered respectively (45). Amplification of candidate divisions WS6, TM7, and OP11, as well as phylum Nanoarchaeota was deemed unlikely via *in silico* analysis (45). The forward primer was modified to include a 5'-M13 tag used for labeling the PCR products with Illumina tags (46). Primary PCR reactions (30  $\mu$ L total) were carried out using 2X PCR Master Mix (Fermentas/ThermoScientific). Amplicons were checked by gel electrophoresis and confirmation of a single band and then cleaned

using the QIAquick® PCR Purification Kit (Qiagen). The M13-containing amplicons were then tagged for MiSeq Illumina sequencing by including a unique 8bp barcode into each amplicon (46). MiSeq Illumina sequencing was performed as previously described (47) with the modification of an added CC spacer between adapter and barcode.

### **Sequence Classification**

Raw Illumina sequence reads were processed by first removing adapter sequences and then stitching overlapping forward and reverse reads. Sequences were then clustered and assigned taxonomy using the QIIME pipeline. They were demultiplexed and then clustered into Operational Taxonomic Units (OTUs) using UCLUST at the 95% identity level. A representative set of sequences was picked at random from each OTU and aligned to the SILVA small subunit rRNA reference alignment ([www.arb-silva.de](http://www.arb-silva.de)) using the PyNAST algorithm (48). Core taxa were defined as those OTUs which represented >1% of reads within a sample, and rare OTUs were defined as those having a representation <0.1% within a sample (49).

Classification was then exported at the genus, class, and phylum levels. 16S rRNA gene frequencies at each level were then normalized to the respective qPCR quantities (bacterial OTUs to bacterial 16S qPCR, and archaeal OTUs to archaeal 16S qPCR).

This was done because frequencies of OTUs are in reference to the whole 16S dataset which includes both bacterial and archaeal data. The qPCR normalization was also performed to account for the differential abundance and distribution of their DNA in gradients. Data were then converted to ratios of quantities in which the highest measured normalized frequency equaled one.

## Results

The water column during late August was well-mixed, with a water temperature of +4.7°C and salinity of 30.2. Chlorophyll *a* was 0.4 µg L<sup>-1</sup>. In winter, the water column remained well-mixed, with a water temperature of -1.8°C, salinity of 33.7, and chlorophyll *a* concentrations of 0.01 µg L<sup>-1</sup>. Light levels were low in the water column under the sea ice at 0.12 µmol quanta<sup>-1</sup> m<sup>-2</sup> s<sup>-1</sup>. Summer concentrations of ambient nutrients were greatest for NH<sub>4</sub><sup>+</sup> (Table 1). Uptake rates were similarly dominated by NH<sub>4</sub><sup>+</sup>, which was close to an order of magnitude greater than NO<sub>3</sub><sup>-</sup> and urea uptake rates. Regeneration of NH<sub>4</sub><sup>+</sup> was more than six times greater than uptake. During winter, NO<sub>3</sub><sup>-</sup> concentrations increased dramatically, but all of the uptake rates fell to extremely low levels, with urea uptake rates the lowest of those measured.

Sequencing of amplified 16S rRNA genes from ambient community DNA yielded 97,858 and 25,666 quality paired-end Illumina reads for summer and winter samples respectively. A breakdown of ambient microbial communities at the phylum/family-level is shown in Figure 1. Overall, samples contained fairly similar microbial communities during both seasons, with some notable differences. Archaea accounted for a smaller proportion of overall reads in the summer (1.8%) as compared to the winter (11.8%). These data are consistent with results from qPCR of community DNA, which detected  $7.1 \times 10^6 \pm 4.1 \times 10^5$  and  $2.5 \times 10^6 \pm 2.0 \times 10^5$  bacterial and  $8.4 \times 10^4 \pm 5.4 \times 10^3$  and  $3.0 \times 10^5 \pm 1.7 \times 10^4$  archaeal rRNA gene copies per ml of seawater in summer and winter samples respectively. Archaea were therefore 3.6-fold more abundant in winter samples as compared to those collected during the summer, while bacteria were 2.8-fold more abundant during the summer. The ratio of bacterial to

archaeal 16S rRNA genes was 85 in the summer and 8.3 in winter samples. Thirty-eight archaeal OTUs were shared among summer and winter libraries, but winter communities were characterized by a greater proportion of MGIC and a smaller proportions of reads classified as *Halobacteria* (<1%) and *Methanobacteria* (<0.3%; Figure 1A and 1B).

Bacterial communities also exhibited strong similarities, with 19 core OTUs (>1.0% of reads) shared among summer and winter, accounting for 65% and 53% of reads during summer and winter respectively. The prominent difference between the two seasons was a greater proportion of sequences classified within the Cyanobacteria/Chloroplast and Verrucomicrobia in the summer sample (Figure 1C and 1D). Winter samples conversely had greater proportions of reads within the Proteobacteria, Bacteroidetes, and Planctomycetes as compared to the summer. Within the Proteobacteria, most OTUs were classified within the SAR11 group and the Alteromonadales (*Oceanospirillum*) during the winter sampling. A greater diversity of taxa within the proteobacteria was observed during the summer. The winter samples exhibited greater species richness, containing 594 unique OTUs (found only in the winter sequence libraries) compared to 144 unique taxa in summer samples. The rare biosphere was prominent, with 93% and 96% of the OTUs found in less than 0.1% of reads in libraries during the summer and winter respectively.

CsCl DNA density gradients from  $^{14}\text{N}$ - and  $^{15}\text{N}$ -labeled  $\text{NH}_4^+$ ,  $\text{NO}_3^-$ , and urea (summer and winter) as well as  $^{12}\text{C}$  and  $^{13}\text{C}$ -bicarbonate treatments (winter only) were fractionated. Each of the resulting fractions was assayed via qPCR for bacterial and archaeal 16S rRNA genes to estimate the degree of isotopic labeling (Figure 2).

Addition of  $^{15}\text{N-NH}_4^+$  led to isotopic labeling of bacterial and archaeal DNA in summer but not winter samples, when 30% labeling for the major peak in the gradient is used as a conservative cutoff (Table 2). Evidence for the incorporation of  $^{15}\text{N-NO}_3^-$  into either bacteria or archaea was not observed in either season. Incubation with  $^{15}\text{N-urea}$  produced no evidence of incorporation of N from urea in summer samples, however, winter samples yielded estimates of 30% and 35% isotopic labeling for bacterial and archaeal populations respectively (Figure 2A). Winter  $^{14}\text{N}/^{15}\text{N-urea}$  treatments were therefore chosen for a more detailed analysis by high throughput sequencing. In addition,  $^{12}\text{C}/^{13}\text{C}$ -bicarbonate treated samples were analyzed via SIP to investigate dark (winter) carbon fixation activity by the major prokaryotic populations, revealing 17 and 18% labeling of DNA with  $^{13}\text{C}$  for bacterial and archaeal populations in the winter respectively.

A total of  $1.27 \times 10^6$  paired-end reads were generated from  $^{14}\text{N}/^{15}\text{N-urea}$  treatments using the Illumina MiSeq platform, yielding an average of  $3.59 \times 10^4$  paired-end sequences for each fraction. The ten most abundant divisions (Proteobacteria are shown at the family level; Figure 1) accounted for 60% of the read data and were chosen for further analysis. Less abundant divisions generally did not contain sufficient read data in each fraction to resolve frequency distributions sufficiently well. SIP analysis, as presented here, therefore only addresses major phyla and orders from our sampling site. Overall, labeling with N from  $^{15}\text{N-urea}$  appeared to be widespread among taxonomic groups and individual OTUs (Table 3). Firmicutes and Betaproteobacteria exhibited labeling above the 30% threshold (Table 3). Alphaproteobacterial DNA displayed a shift of only 20% at the phylum level, but specific Alphaproteobacterial

OTUs which classified within the Sar11-clade, exhibited density shifts of >30% (data not shown). In addition, labeling of 25% and 39% were observed for a Gammaproteobacterial OTU classified as *Oceanospirillum* and Deltaproteobacterial OTU classified within the Sar324-clade respectively. The dominant archaeal OTU classified within the MGIC lineage and displayed 31% labeling with <sup>15</sup>N from <sup>15</sup>N-urea.

Both the bacterial and the archaeal qPCR analysis of SIP fractions provided evidence of <sup>13</sup>C-bicarbonate uptake during the winter (Figure 2B), exhibiting 17% and 18% labeling respectively. Sequencing of fractions from <sup>12</sup>C and <sup>13</sup>C-bicarbonate treated samples produced  $1.40 \times 10^6$  reads, which yielded an average of  $2.38 \times 10^4$  paired-end sequences for each fraction. No appreciable difference in the overall phylogenetic composition of these data as compared to the <sup>15</sup>N fractions was observed (data not shown). Evidence of labeling was observed for all four major Proteobacterial families detected here, as well as the Firmicutes and the Crenarchaeota.

## Discussion

Marine prokaryotic plankton play a critical role in the biogeochemical cycling of C and N in the world's oceans (50, 51). Identifying the phylogenetic groups that are responsible for specific C or N cycling activities can provide insights into the forces that drive marine productivity and community function, as well as the spatial and temporal dynamics of individual C and N transformation processes themselves. The work presented here aimed to quantify uptake rates of DIN sources and urea, while placing these rates into context with a targeted SIP experiment to identify incorporation by dominant groups of bacteria and archaea. In addition, we explored dark carbon fixation by bacteria and archaea, which can potentially be coupled to ammonium oxidation (via urea deamination).

SIP data suggests an important seasonal transition of microbial N incorporation from  $\text{NH}_4^+$  during the summer to urea during the winter (Table 2). In addition, SIP implicated a broad range of microbial taxa in winter urea utilization, including Proteobacteria, Firmicutes, and MGIC. Utilization of  $^{15}\text{N}$ -urea is not necessarily expected, given that other sources of N were available, but it is not inconsistent with the phylogenetic distribution of urease genes among bacteria and archaea. Although it has been argued that ureases are rare in the domain Archaea (17), and it has been noted that urease genes are absent from the genomes of *Nitrosopumilis maritimus* and *Nitrosoarchaeum limnia* (52), it has also been observed that *ureC* genes are abundant in polar archaea, with an average ratio close to or greater than one as compared to 16S rRNA genes (2). A search of genome sequences available in the DOE's Joint Genome Institute IMG database (as of August 2013) identifies urease genes within genomes of

several *Halobacteria* and Crenarchaeota, the latter of which are relevant to our samples. Similarly, this search identifies ureases in the genomes of Alpha-, Beta-, Gamma-, and Epsilon-Proteobacteria, Actinobacteria, and Firmicutes, which were the main phylogenetic groups at our sampling site. In contrast to SIP observations, urea uptake rates were measurable but very low in both summer and winter samples. This discrepancy is likely a reflection of our choice of filters. GF/F filters (nominal pore size of 0.7  $\mu\text{m}$ ) were employed in our study to determine rates, while DNA extractions for SIP were conducted using 0.45  $\mu\text{m}$  filters. Flow cytometric analysis of seawater samples indicates that most bacterial cells in winter samples were smaller than the 0.7  $\mu\text{m}$  cutoff of GF/F filters (data not shown; more details provided in (24)). Simpson, et al. (23) reported that the 0.2-0.7  $\mu\text{m}$  fraction accounts for <10% of N uptake during summer, and that bacteria assimilated relatively more N when phytoplankton were less important in N uptake. Urea uptake rates as reported here are therefore likely an underestimation of actual *in situ* activity, particularly for the winter.

The metabolic advantage of  $\text{NH}_4^+$  leads to the expectation of universal labeling of DNA with  $^{15}\text{N}$  from  $^{15}\text{NH}_4^+$  during either season. A lack of labeling with  $^{15}\text{NH}_4^+$  might indicate incubation times inconsistent with rates of uptake. However, summer and winter samples exhibited good incorporation of  $^{15}\text{N}$  from  $\text{NH}_4^+$  and urea respectively, indicating that incubation times chosen for our experiments were sufficiently long. Conversely, incubation times that are too long can lead to cross-feeding and eliminate the ability to resolve differential uptake in SIP experiments (44). Cross-feeding results when a substrate is added in one form but is converted to another form during the incubation. Nitrification, for example, can occur at high rates during



the Arctic winter (24, 53), and therefore additions of  $^{15}\text{N-NH}_4^+$  could result in the production of  $^{15}\text{N-NO}_3^-$ , which could subsequently be incorporated. Additionally, conversion of urea to  $\text{NH}_4^+$  is likely to occur, but there is no current information on the rate of that process in the environment. Given those caveats, our relatively short incubation times and low ambient rates of uptake (Table 1), it is unlikely that cross-feeding represented an important factor in our results.

Incorporation of N from  $^{15}\text{N-NO}_3^-$  by bacterial or archaeal populations was not supported by SIP for either season (Table 2). This is contrary to prior studies that indicate  $\text{NO}_3^-$  can serve as an important N source to bacterioplankton in the Arctic. For example, using protein synthesis inhibitors and  $^{15}\text{N}$ -tracers in waters collected from northern Baffin Bay it was observed that heterotrophic bacteria can account for as much as 25% of  $\text{NO}_3^-$  uptake when chlorophyll concentrations are low ( $<2 \mu\text{g L}^{-1}$ ) (25). Experiments in the Barents Sea similarly indicated that bacteria can account for 16-40% of total  $\text{NO}_3^-$  uptake, with greater importance of this process at stations near ice cover (54). A study of sub-Arctic Pacific communities revealed that bacteria can account for 32% of total  $\text{NO}_3^-$  uptake, roughly similar to heterotrophic bacterial  $\text{NH}_4^+$  uptake (55). Overall, a meta-analysis of these and other data suggests a pattern of greater  $\text{NO}_3^-$  utilization by heterotrophic bacterioplankton under low chlorophyll conditions and relatively lower contributions when phytoplankton are more prevalent (25). Although  $\text{NO}_3^-$  incorporation, as determined via SIP, was not observed in this study, we note that background  $\text{NO}_3^-$  levels ( $\sim 10 \mu\text{M}$ , Table 1) in winter samples were substantially above our additions. SIP experiments, as performed here, require an excess of  $^{15}\text{N}$ -labeled substrate in order to result in  $>30\%$  labeling of cellular DNA. Additions of  $^{15}\text{N-NO}_3^-$  to

winter samples were therefore likely below the limit of detection for our experimental setup. Summer additions, on the contrary, were sufficiently high and a lack of  $^{15}\text{N}$ -DNA labeling is consistent with  $\text{NO}_3^-$  uptake being dominated by phytoplankton (56). It has also been noted that microbial growth parameters do not correlate well with  $\text{NO}_3^-$  uptake in the sub-Arctic Pacific, and that  $\text{NO}_3^-$  is likely the least preferred N source supporting bacterial growth (55). Utilization of  $\text{NO}_3^-$  in the Arctic varies both seasonally and spatially, but it appears that  $\text{NO}_3^-$  may be a source of N only as a last resort for bacterioplankton, especially during times other than the spring bloom when energetically more favorable sources of N such as  $\text{NH}_4^+$ , amino acids, and urea are abundant (24).

Community analysis of 16S rRNA genes revealed that summer and winter communities were largely similar, while displaying some notable differences. Minor seasonal differences in microbial community composition are in line with previous reports for Arctic Ocean microbial communities (57-59). For example, analysis of DGGE fingerprints from Arctic and Antarctic samples revealed no seasonal variation in archaeal community structure, but suggested greater richness of archaea in water from greater depth (59). The dominant contributors to archaeal community composition in the study by Bano et al. (59) were MGIC, similar to populations described here. We also observed a greater relative proportion of archaea during the winter, similar to prior observations (52). The abundance of MGIC in the Southeast Beaufort Sea has been observed to be from 6% to 18% between January and March, while decreasing to ca. 5% of cells in May and June of the same year (52). This change was attributed to growth of the archaeal populations during the winter and not mixing with deeper water

masses, which often contain greater proportions of these organisms. Galand, et al. (60) reported that MGIC were the most abundant archaea when five distinct water masses of the Arctic Ocean were interrogated, representing between 27% and 63% of sequences in pyrotag libraries. Overall, these data suggest that archaea are a salient and stable feature of Arctic marine prokaryotic plankton communities and that their relative importance increases during the dark and cold months of the Arctic winter. Seasonal differences in bacterial communities are equally constrained. A study using samples collected from the western Arctic reported no significant seasonal differences in bacterial communities, as determined via pyrotag sequencing, despite large differences in biogeochemical parameters (57). The dominant bacterial taxa in that study were the Alpha- and Gamma-proteobacteria as well as the Bacteroidetes (*Flavobacteria*) (57). Our findings match these observations, although, Kirchman et al. (50) did not observe a shift towards greater proportions of chloroplast and SAR11-like sequences during the summer. These discrepancies likely arise from the use of 0.8  $\mu\text{m}$  pre-filtered samples (55) in order to remove much of the eukaryotic phytoplankton, including their chloroplasts.

Urea is present in relatively high concentrations in polar seawater (22, 23, 52), and its role as an important component of the Arctic N cycle has been recognized. Urea can account for as much as 30-50% of the N assimilated by phytoplankton annually (22, 23) and as much as 80% of the regenerated production during the spring bloom (23). Urea also has been reported to have similar half saturation constants but greater maximum uptake rates as compared to  $\text{NH}_4^+$  (61). A more recent study has investigated the role of urea in nitrification by polar marine archaea (52). In that study,

metagenomic analysis of Arctic winter samples revealed an abundance of urea transport and degradation genes. Quantitative PCR assays resulted in good correlation between the number of MGIC and *ureC* genes, suggesting that *ureC* genes were abundant in polar archaea. Experiments with <sup>14</sup>C-labeled urea demonstrated uptake of C from urea (carbon dioxide is generated from urea by urease activity) by the 0.2-0.6 μm fraction and that this activity was greater under dark conditions (52). The implication is that urea may fuel nitrification and autotrophic growth by polar archaea via the release of ammonium from urea. This may be advantageous under dark conditions, when urea can be a more reliable source of energy for ammonium oxidizers (2). Urea may also be continually produced during the Arctic winter via microbial and zooplankton (20) activities. The notion of urea-fueled nitrification is not without precedent. It has been argued that urea hydrolysis may serve to generate ammonium and carbon dioxide by ammonium oxidizers (62), allowing these bacteria to generate energy to fuel dark carbon fixation.

A recent study using DNA-SIP on bacterial communities in the coastal northwest Pacific Ocean found dark carbon fixation to be widespread across different bacterial taxa (29). In another study in deeper waters (>50m) of the Chukchi Sea during spring and summer, results from microautoradiography-catalyzed reporter deposition fluorescence *in situ* hybridization (MAR-CARDFISH) showed that the percent of bacteria and Crenarcheota actively incorporating bicarbonate was highest on the shelf at 20% (for both groups) and decreased to <10% further offshore (63). Also using MAR-CARDFISH, a study in the Canadian Arctic found that Gamma- and Beta-Proteobacteria were active in the assimilation of bicarbonate when stationary phase

seawater cultures were grown on a resource deplete medium, specifically the Gammaproteobacteria *Oleispira* and *Pseudoalteromonas-Colwellia* (27). Our results demonstrate that dark carbon fixation may not be limited to dilution cultures grown under enhanced substrate-depletion, but may also be a viable metabolic strategy for many taxa under *in situ* conditions. Moreover, our work expands on the previously reported diversity of bacteria and archaea active in the assimilation of dissolved inorganic carbon (DIC) in the Arctic during winter.

The relevance of chemoautotrophy, such as ammonia oxidation by bacteria and archaea, may increase from summer to winter in Arctic waters when light levels and photosynthetic primary production are low (53). The important role of ammonia oxidizing Crenarchaeota to carbon fixation and nitrogen cycling in deeper waters of the world's oceans is becoming well established (30, 31, 64). For example, it is estimated that dark carbon fixation by ammonia-oxidizing archaea to be  $>700 \text{ Tg-C y}^{-1}$  (30), of which approximately  $400 \text{ Tg-C y}^{-1}$  is fixed by Crenarchaeota (64). Yet, the relevance of ammonia oxidation in surface waters of the ocean are thought to remain low due to inhibition from light and competition for ammonia with phytoplankton (65, 66). Arctic winter surface waters are comparable to deeper waters in the ocean due to limited light as a consequence of sea-ice coverage and short day lengths, and low inputs of organic matter from photosynthesis. Further, ammonium concentrations in the Arctic are higher in winter than in summer (53)(this study). Thus, the role of chemoautotrophy from ammonia oxidation may be notable in Arctic surface waters during winter, as it is for deeper waters in the world's oceans. Consistent with this proposition, Christman *et al.* (53) found that ammonia monooxygenase (*amoA*) genes from Crenarchaeota and

Betaproteobacteria in surface waters near Barrow, Alaska were almost two magnitudes greater in winter than in summer, with most or all Crenarchaeota likely being capable of ammonia oxidation.

In addition to chemoautotrophic contributions to DIC uptake in oxygenic waters, all heterotrophic bacteria are thought to be able to assimilate bicarbonate via pathways involved in anaplerotic reactions of the tricarboxylic acid cycle. However, carbon fixation by heterotrophs via anaplerotic reactions is assumed to only account for 1-8% of bacterial biomass production (67, 68) and thus, play only a minor role in DIC uptake. In contrast, ammonium additions failed to stimulate bicarbonate assimilation on a per cell basis leading Alonso-Sáez *et al.* (27) to conclude that the high rates of bicarbonate assimilation they observed in their seawater cultures resulted from dark carbon fixation by heterotrophs. Dark carbon fixation therefore does not appear restricted to deeper waters and has the potential to be an essential survival strategy for heterotrophs, chemoautotrophs, mixotrophs, and other organisms in the microbial loop that rely on archaeal and bacterial production during the Arctic winter.

To determine the extent that dark-carbon fixation is a significant aspect of community ecology during the Arctic winter, specifically for those microorganisms that are primarily heterotrophic, additional research is needed. To start, quantification of the *in situ* rates of bicarbonate uptake by bacteria and archaea in the dark is essential (e.g.,(69)). Preliminary data from our study area during January suggest an *in situ* bicarbonate uptake rate of  $0.09 \mu\text{g-C L}^{-1} \text{d}^{-1}$  (SE Baer, DA Bronk, unpublished data), which is similar to rates observed by L. Alonso-Sáez (personal communication) in the Canadian Beaufort Sea and within the range for reported heterotrophic production rates

in the Arctic in winter and spring (70, 71). These scarce uptake measurements suggest that dark carbon fixation may be notable for marine Arctic ecosystems during winter and highlight the need for additional observations. Given this, the phylogenetic diversity of labeled bacterial populations in  $^{15}\text{N}$ -urea and  $^{13}\text{C}$ -bicarbonate SIP experiment, it appears that this process could also be an important bacterial survival strategy in Arctic seawater during the winter. Future investigations of Arctic urea driven ammonium oxidation and carbon fixation dynamics should therefore consider both bacterial and archaeal plankton components.

The current study, in conjunction with discoveries made by others, continues to call our attention to several unresolved questions, such as: How widespread is a seasonal shift in ammonium and urea utilization? Do ambient resources have a role in regulating N utilization in the Arctic? What is the capacity and rate of dark DIC assimilation by heterotrophic bacteria? How does this metabolism influence other members of the ecosystem such as bacteriovores, especially in oligotrophic environments like the Arctic in winter? And, what is the role of dark carbon fixation in the energy and carbon budget of the world's oceans? It was only recently that ammonia-oxidizing archaea were identified as key players, capable of dark carbon fixation in oxygenic mesopelagic environments. The above questions remain poorly addressed for aerobic chemoautotrophs, mixotrophs, and heterotrophs. Recent methodological advances with DNA-SIP, RNA-SIP, lipid-SIP, MAR-CARDFISH and metagenomics have promise to help us resolve the importance of dark carbon fixation in oxygenated waters on a seasonal basis in the Arctic Ocean or on global scales and to deepen our

understanding of life-history strategies adapted for survival in resource limited environments.



## **Acknowledgements**

We wish to thank Q. N. Roberts, K. Sines and R. E. Sipler for fieldwork. We also thank M. P. Sanderson for running mass spectrometry samples. This work would not have been possible without the guidance and support of the UMIAQ support team, led by B. France, and the support by the United States National Science Foundation grants OCE 0961900, ARC 0910252, and ARC 0909839.

## References

1. Duarte CM, Lenton TM, Wadhams P, Wassmann P. 2012. Abrupt climate change in the Arctic. *Nat. Clim. Change* 2:60-62.
2. Parmentier F-JW, Christensen TR, Sorensen LL, Rysgaard S, McGuire AD, Miller PA, Walker DA. 2013. The impact of lower sea-ice extent on Arctic greenhouse-gas exchange. *Nat. Clim. Change* 3:195-202.
3. Holland MM, Bitz CM. 2003. Polar amplification of climate change in coupled models. *Clim. Dynam.* 21:221-232.
4. Stroeve JC, Serreze MC, Holland MM, Kay JE, Maslanik J, Barrett AP. 2012. The Arctic's rapidly shrinking sea ice cover: a research synthesis. *Clim. Change* 110:1005-1027.
5. Serreze MC, Francis JA. 2006. The arctic amplification debate. *Clim. Change* 76:214-264.
6. Doney SC, Ruckelshaus M, Emmett Duffy J, Barry JP, Chan F, English CA, Galindo HM, Grebmeier JM, Hollowed AB, Knowlton N, Polovina J, Rabalais NN, Sydeman WJ, Talley LD. 2012. Climate Change Impacts on Marine Ecosystems. *Annu. Rev. Mar. Sci.* 4:11-37.
7. Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F. 2012. Impacts of climate change on the future of biodiversity. *Ecol. Lett.* 15:365-377.
8. Perrette M, Yool A, Quartly GD, Popova EE. 2011. Near-ubiquity of ice-edge blooms in the Arctic. *Biogeosci.* 8:515-524.
9. Sherr EB, Sherr BF, Wheeler PA, Thompson K. 2003. Temporal and spatial variation in stocks of autotrophic and heterotrophic microbes in the upper water column of the central Arctic Ocean. *Deep-Sea Res. Pt. I: Oceanogr. Res. Papers* 50:557-571.
10. Antia NJ, Harrison PJ, Oliveira L. 1991. The Role of Dissolved Organic Nitrogen in Phytoplankton Nutrition, Cell Biology and Ecology. *Phycologia* 30:1-89.
11. Aluwihare LI, Meador T. 2008. Chemical composition of marine dissolved organic nitrogen. *In* Capone DG, Bronk DA, Mulholland M, Carpenter EJ (ed.), *Nitrogen in the marine environment*, 2nd ed. Elsevier, San Diego.
12. Letscher RT, Hansell DA, Kadko D, Bates NR. 2013. Dissolved organic nitrogen dynamics in the Arctic Ocean. *Mar. Chem.* 148:1-9.

13. L'Helguen S, Slawyk G, Le Corre P. 2005. Seasonal patterns of urea regeneration by size-fractionated microheterotrophs in well-mixed temperate coastal waters. *J. Plankton Res.* 27:263-270.
14. Remsen CC. 1971. The Distribution of Urea in Coastal and Oceanic Waters. *Limnol. Oceanogr.* 16:732-740.
15. Remsen CC, Carpenter EJ, Schroeder BW. 1974. The role of urea in marine microbial ecology, p. 286–304. *In* Colwell RRAM, R. Y. (ed.), *Effect of the Ocean Environment on Microbial Activities*. University Park Press, Baltimore.
16. McCarthy JJ. 1972. The Uptake of Urea by Natural Populations of Marine Phytoplankton. *Limnol. Oceanogr.* 17:738-748.
17. Solomon CM, Collier JL, Berg GM, Glibert PM. 2010. Role of urea in microbial metabolism in aquatic systems: a biochemical and molecular review. *Aquat. Microb. Ecol.* 59:67-88.
18. Glibert P, Magnien R, Lomas M, Alexander J, Tan C, Haramoto E, Trice M, Kana T. 2001. Harmful algal blooms in the Chesapeake and Coastal Bays of Maryland, USA: Comparison of 1997, 1998, and 1999 events. *Estuaries* 24:875-883.
19. Conover RJ, Gustavson KR. 1999. Sources of urea in arctic seas: zooplankton metabolism. *Mar. Ecol. Prog. Ser.* 179:41-54.
20. Conover RJ, Mumm N, Bruecker P, MacKenzie S. 1999. Sources of urea in arctic seas: seasonal fast ice. *Mar. Ecol. Prog. Ser.* 179:55-69.
21. Pedersen H, Lomstein BA, T. Henry B. 1993. Evidence for bacterial urea production in marine sediments. *FEMS Microbiol. Ecol.* 12:51-59.
22. Harrison WG, Head EJH, Conover RJ, Longhurst AR, Sameoto DD. 1985. The distribution and metabolism of urea in the eastern Canadian Arctic. *Deep-Sea Res. Pt. A: Oceanogr. Res. Papers* 32:23-42.
23. Simpson KG, Tremblay JE, Brugel S, Price NM. 2013. Nutrient dynamics in the western Canadian Arctic. II. Estimates of new and regenerated production over the Mackenzie Shelf and Cape Bathurst Polynya. *Mar. Ecol. Prog. Ser.* 484:47-62.
24. Baer SE, Sipler RE, Roberts QN, Yager PL, Frischer ME, Bronk DA. in review. Seasonal Nitrogen Uptake and Regeneration in the Western Coastal Arctic. *Limnol. Oceanogr.*

25. Fouilland E, Gosselin M, Rivkin RB, Vasseur C, Mostajir B. 2007. Nitrogen uptake by heterotrophic bacteria and phytoplankton in Arctic surface waters. *J. Plankton Res.* 29:369-376.
26. Wuchter C, Schouten S, Boschker HTS, Damste JSS. 2003. Bicarbonate uptake by marine Crenarchaeota. *FEMS Microbiol. Lett.* 219:203-207.
27. Alonso-Sáez L, Galand PE, Casamayor EO, Pedros-Alio C, Bertilsson S. 2010. High bicarbonate assimilation in the dark by Arctic bacteria. *ISME J.* 4:1581-1590.
28. Hügl M, Sievert SM. 2011. Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. *Annu Rev Mar Sci* 3:261-289.
29. DeLorenzo S, Bräuer SL, Edgmont CA, Herfort L, Tebo BM, Zuber P. 2012. Ubiquitous Dissolved Inorganic Carbon Assimilation by Marine Bacteria in the Pacific Northwest Coastal Ocean as Determined by Stable Isotope Probing. *PLoS ONE* 7:e46695.
30. Herndl GJ, Reinthaler T, Teira E, van Aken H, Veth C, Pernthaler A, Pernthaler J. 2005. Contribution of Archaea to Total Prokaryotic Production in the Deep Atlantic Ocean. *Appl. Environ. Microbiol.* 71:2303-2309.
31. Ingalls AE, Shah SR, Hansman RL, Aluwihare LI, Santos GM, Druffel ERM, Pearson A. 2006. Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon. *Proc. Natl. Acad. Sci. USA* 103:6442-6447.
32. Yakimov MM, La Cono V, Smedile F, DeLuca TH, Juarez S, Ciordia S, Fernandez M, Albar JP, Ferrer M, Golyshin PN, Giuliano L. 2011. Contribution of crenarchaeal autotrophic ammonia oxidizers to the dark primary production in Tyrrhenian deep waters (Central Mediterranean Sea). *ISME J.* 5:945-961.
33. Koroleff F. 1983. Determination of nutrients, p. 125-187. *In* Grasshoff K, Ehrhardt M, Kremling K (ed.), *Methods of Seawater Analysis*. Verlag Chemie, New York.
34. Parsons TR, Maita Y, Lalli CM. 1984. *A manual of chemical and biological methods for seawater analysis*. Pergamon, Oxford.
35. Price NM, Harrison PJ. 1987. Comparison of methods for the analysis of dissolved urea in seawater. *Mar. Biol.* 94:307-317.
36. Dugdale RC, Goering JJ. 1967. Uptake of new and regenerated forms of nitrogen in primary productivity. *Limnol. Oceanogr.* 12:196-206.

37. Hama T, Miyazaki T, Ogawa Y, Iwakuma T, Takahashi M, Otsuki A, Ichimura S. 1983. Measurement of Photosynthetic Production of a Marine-Phytoplankton Population Using a Stable C-13 Isotope. *Mar. Biol.* 73:31-36.
38. Dudek N, Brzezinski MA, Wheeler PA. 1986. Recovery of ammonium nitrogen by solvent extraction for the determination of relative  $^{15}\text{N}$  abundance in regeneration experiments. *Mar. Chem.* 18:59-69.
39. Brzezinski MA. 1987. Colorimetric determination of nanomolar concentrations of ammonium in seawater using solvent extraction. *Mar. Chem.* 20:277-288.
40. Glibert PM, Lipschultz F, McCarthy JJ, Altabet MA. 1982. Isotope dilution models of uptake ammonium by marine plankton. *Limnol. Oceanogr.* 27:639-650.
41. Wawrik B, Callaghan AV, Bronk DA. 2009. Use of inorganic and organic nitrogen use by *Synechococcus* spp. and diatoms on the West Florida Shelf as measured using stable isotope probing. *Appl. Environ. Microbiol.* 75:6662-6670.
42. Buckley DH, Huangyutitham V, Hsu S-F, Nelson TA. 2007. Stable isotope probing with  $^{15}\text{N}_2$  reveals novel noncultivated diazotrophs in soil. *Appl. Environ. Microbiol.* 73:3196-3204.
43. Buckley DH, Huangyutitham V, Hsu S-F, Nelson TA. 2007. Stable isotope probing with  $^{15}\text{N}$  achieved by disentangling the effects of genome G+C content and isotope enrichment on DNA density. *Appl. Environ. Microbiol.* 73:3189-3195.
44. Wawrik B, Boling WB, Van Nostrand JD, Xie JP, Zhou JZ, Bronk DA. 2012. Assimilatory nitrate utilization by bacteria on the West Florida Shelf as determined by stable isotope probing and functional microarray analysis. *FEMS Microbiol. Ecol.* 79:400-411.
45. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glockner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1.
46. Wawrik B, Mendivelso M, Parisi VA, Suflita JM, Davidova IA, Marks CR, Van Nostrand JD, Liang Y, Zhou J, Huizinga BJ, Strapoc D, Callaghan AV. 2012. Field and laboratory studies on the bioconversion of coal to methane in the San Juan Basin. *FEMS Microbiol. Ecol.* 81:26-42.
47. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST,

- Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunencko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7:335-336.
48. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26:266-267.
  49. Pedrós-Alió C. 2006. Marine microbial diversity: can it be determined? *Trend. Microbiol.* 14:257-263.
  50. Kirchman DL, Moran XA, Ducklow H. 2009. Microbial growth in the polar oceans - role of temperature and potential impact of climate change. *Nature Rev. Microbiol.* 7:451-459.
  51. Azam F, Malfatti F. 2007. Microbial structuring of marine ecosystems. *Nature Rev. Microbiol.* 5:966-966.
  52. Alonso-Sáez L, Waller AS, Mende DR, Bakker K, Farnelid H, Yager PL, Lovejoy C, Tremblay J-É, Potvin M, Heinrich F, Estrada M, Riemann L, Bork P, Pedrós-Alió C, Bertilsson S. 2012. Role for urea in nitrification by polar marine Archaea. *Proc. Natl. Acad. Sci. USA* 109:17989-17994.
  53. Christman GD, Cottrell MT, Popp BN, Gier E, Kirchman DL. 2011. Abundance, Diversity, and Activity of Ammonia-Oxidizing Prokaryotes in the Coastal Arctic Ocean in Summer and Winter. *Appl. Environ. Microbiol.* 77:2026-2034.
  54. Allen AE, Howard-Jones MH, Booth MG, Frischer ME, Verity PG, Bronk DA, Sanderson MP. 2002. Importance of heterotrophic bacterial assimilation of ammonium and nitrate in the Barents Sea during summer. *J. Mar. Syst.* 38:93-108.
  55. Kirchman DL, Wheeler PA. 1998. Uptake of ammonium and nitrate by heterotrophic bacteria and phytoplankton in the sub-Arctic Pacific. *Deep-Sea Res. Pt. I: Oceanogr. Res. Papers* 45:347-365.
  56. Simpson KG, Tremblay J-É, Price NM. 2013. Nutrient dynamics in the western Canadian Arctic. I. New production in spring inferred from nutrient draw-down in the Cape Bathurst Polynya. *Mar. Ecol. Prog. Ser.* 484:33-45.
  57. Kirchman DL, Cottrell MT, Lovejoy C. 2010. The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environ. Microbiol.* 12:1132-1143.

58. Nikrad MP, Cottrell MT, Kirchman DL. 2012. Abundance and Single-Cell Activity of Heterotrophic Bacterial Groups in the Western Arctic Ocean in Summer and Winter. *Appl. Environ. Microbiol.* 78:2402-2409.
59. Bano N, Ruffin S, Ransom B, Hollibaugh JT. 2004. Phylogenetic Composition of Arctic Ocean Archaeal Assemblages and Comparison with Antarctic Assemblages. *Appl. Environ. Microbiol.* 70:781-789.
60. Galand PE, Casamayor EO, Kirchman DL, Potvin M, Lovejoy C. 2009. Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing. *ISME J.* 3:860-869.
61. Smith WO, Harrison WG. 1991. New Production in Polar-Regions - the Role of Environmental Controls. *Deep-Sea Res. Pt. A: Oceanogr. Res. Papers* 38:1463-1479.
62. Koper TE, El-Sheikh AF, Norton JM, Klotz MG. 2004. Urease-encoding genes in ammonia-oxidizing bacteria. *Appl. Environ. Microbiol.* 70:2342-2348.
63. Kirchman DL, Elifantz H, Dittel AI, Malmstrom RR, Cottrell MT. 2007. Standing stocks and activity of Archaea and Bacteria in the western Arctic Ocean. *Limnol. Oceanogr.* 52:495-507.
64. Wuchter C, Abbas B, Coolen MJL, Herfort L, van Bleijswijk J, Timmers P, Strous M, Teira E, Herndl GJ, Middelburg JJ, Schouten S, Sinninghe Damsté JS. 2006. Archaeal nitrification in the ocean. *Proc. Natl. Acad. Sci. USA* 103:12317-12322.
65. Ward BB. 2008. Nitrification in marine systems. *In* Capone DG, Bronk DA, Mulholland MR, Carpenter EJ (ed.), *Nitrogen in the Marine Environment*. Elsevier, Amsterdam, The Netherlands.
66. Middelburg JJ. 2011. Chemoautotrophy in the ocean. *Geophys. Res. Lett.* 38.
67. Li WKW. 1982. Estimating Heterotrophic Bacterial Productivity by Inorganic Radiocarbon Uptake - Importance of Establishing Time Courses of Uptake. *Mar. Ecol. Prog. Ser.* 8:167-172.
68. Roslev P, Larsen MB, Jorgensen D, Hesselsoe M. 2004. Use of heterotrophic CO<sub>2</sub> assimilation as a measure of metabolic activity in planktonic and sessile bacteria. *J Microbiol Meth* 59:381-393.
69. Reinthaler T, van Aken HM, Herndl GJ. 2010. Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. *Deep-Sea Res Pt II* 57:1572-1580.

70. Garneau ME, Roy S, Lovejoy C, Gratton Y, Vincent WF. 2008. Seasonal dynamics of bacterial biomass and production in a coastal arctic ecosystem: Franklin Bay, western Canadian Arctic. *J Geophys Res-Oceans* 113.
71. Nguyen D, Maranger R, Tremblay JE, Gosselin M. 2012. Respiration and bacterial carbon dynamics in the Amundsen Gulf, western Canadian Arctic. *J Geophys Res-Oceans* 117.



## TABLES

**Table 1.** Concentrations and uptake rates of ammonium ( $\text{NH}_4^+$ ), nitrate ( $\text{NO}_3^-$ ) and urea in near-shore waters of the Alaskan Arctic during January (winter) and August (summer). Uptake rates were determined with GF/F filters (nominal pore size of 0.7  $\mu\text{m}$ ). Where the standard deviation (SD) is noted as not available (n.a.) it implies that  $n=1$  for that measurement.

| Parameter                     | Unit                                  | Summer        |      | Winter        |     |
|-------------------------------|---------------------------------------|---------------|------|---------------|-----|
|                               |                                       | concentration | SD   | concentration | SD  |
|                               |                                       | or rate       |      | or rate       |     |
| $\text{NH}_4^+$ concentration | $\text{nmol N L}^{-1}$                | 590           | 60   | 960           | 23  |
| $\text{NH}_4^+$ uptake        | $\text{nmol N L}^{-1} \text{ h}^{-1}$ | 5.78          | 0.34 | 0.19          | 0.1 |
| $\text{NH}_4^+$ regeneration  | $\text{nmol N L}^{-1} \text{ h}^{-1}$ | 38.2          | 3.15 | 10.5          | 5.2 |
| $\text{NO}_3^-$ concentration | $\text{nmol N L}^{-1}$                | 290           | n.a. | 9,855         | 1.4 |
| $\text{NO}_3^-$ uptake        | $\text{nmol N L}^{-1} \text{ h}^{-1}$ | 0.82          | 0.69 | 0.14          | 0   |
| Urea concentration            | $\text{nmol N L}^{-1}$                | 230           | 2.9  | 157           | 1.9 |
| Urea uptake                   | $\text{nmol N L}^{-1} \text{ h}^{-1}$ | 0.99          | n.a. | 0.01          | 0   |

**Table 2.** Percent isotopic labeling of DNA with  $^{15}\text{N}$  or  $^{13}\text{C}$  calculated from qPCR data of SIP fractions. Bold percentages are above the 30% and 15% thresholds used to define  $^{15}\text{N}$ - and  $^{13}\text{C}$ -substrate uptake, respectively.

| <b>Kingdom</b> | <b>Season</b> | <b>Substrate</b>            | <b>% Incorporation</b> |
|----------------|---------------|-----------------------------|------------------------|
| Bacteria       | Summer        | $^{15}\text{N-NH}_4$        | <b>31</b>              |
| Bacteria       | Summer        | $^{15}\text{N-NO}_3^-$      | -4                     |
| Bacteria       | Summer        | $^{15}\text{N-Urea}$        | -2                     |
| Archaea        | Summer        | $^{15}\text{N-NH}_4$        | <b>31</b>              |
| Archaea        | Summer        | $^{15}\text{N- NO}_3^-$     | -5                     |
| Archaea        | Summer        | $^{15}\text{N-Urea}$        | 2                      |
| Bacteria       | Winter        | $^{15}\text{N-NH}_4$        | 9                      |
| Bacteria       | Winter        | $^{15}\text{N- NO}_3^-$     | 6                      |
| Bacteria       | Winter        | $^{15}\text{N-Urea}$        | <b>30</b>              |
| Archaea        | Winter        | $^{15}\text{N-NH}_4$        | 23                     |
| Archaea        | Winter        | $^{15}\text{N- NO}_3^-$     | 17                     |
| Archaea        | Winter        | $^{15}\text{N-Urea}$        | <b>35</b>              |
| Bacteria       | Winter        | $^{13}\text{C-bicarbonate}$ | <b>17</b>              |
| Archaea        | Winter        | $^{13}\text{C-bicarbonate}$ | <b>18</b>              |

**Table 3.** Percent isotopic labeling of DNA with  $^{15}\text{N}$  or  $^{13}\text{C}$  derived from 16S rRNA gene sequenced data obtained for SIP fractions of winter samples. Bold percentages are  $\geq 30\%$  and  $15\%$ , the thresholds used to define  $^{15}\text{N}$ - and  $^{13}\text{C}$ -substrate uptake, respectively.

| <b>Kingdom</b> | <b>Division/Family</b> | <b>Season</b> | <b>Substrate</b>             | <b>% Incorporation</b> |
|----------------|------------------------|---------------|------------------------------|------------------------|
| Archaea        | Crenarchaeota          | Winter        | $^{15}\text{N}$ -Urea        | <b>31</b>              |
| Bacteria       | Actinobacteria         | Winter        | $^{15}\text{N}$ -Urea        | 17                     |
| Bacteria       | Bacteroidetes          | Winter        | $^{15}\text{N}$ -Urea        | 22                     |
| Bacteria       | Firmicutes             | Winter        | $^{15}\text{N}$ -Urea        | <b>31</b>              |
| Bacteria       | Planctomycetes         | Winter        | $^{15}\text{N}$ -Urea        | 22                     |
| Bacteria       | Verrucomicrobia        | Winter        | $^{15}\text{N}$ -Urea        | 12                     |
| Bacteria       | Alphaproteobacteria    | Winter        | $^{15}\text{N}$ -Urea        | 20                     |
| Bacteria       | Betaproteobacteria     | Winter        | $^{15}\text{N}$ -Urea        | <b>33</b>              |
| Bacteria       | Deltaproteobacteria    | Winter        | $^{15}\text{N}$ -Urea        | 25                     |
| Bacteria       | Gammaproteobacteria    | Winter        | $^{15}\text{N}$ -Urea        | 23                     |
| Archaea        | Crenarchaeota          | Winter        | $^{13}\text{C}$ -bicarbonate | <b>22</b>              |
| Bacteria       | Actinobacteria         | Winter        | $^{13}\text{C}$ -bicarbonate | 14                     |
| Bacteria       | Bacteroidetes          | Winter        | $^{13}\text{C}$ -bicarbonate | 12                     |
| Bacteria       | Firmicutes             | Winter        | $^{13}\text{C}$ -bicarbonate | <b>17</b>              |
| Bacteria       | Planctomycetes         | Winter        | $^{13}\text{C}$ -bicarbonate | 14                     |
| Bacteria       | Verrucomicrobia        | Winter        | $^{13}\text{C}$ -bicarbonate | 12                     |
| Bacteria       | Alphaproteobacteria    | Winter        | $^{13}\text{C}$ -bicarbonate | <b>18</b>              |
| Bacteria       | Betaproteobacteria     | Winter        | $^{13}\text{C}$ -bicarbonate | <b>15</b>              |
| Bacteria       | Deltaproteobacteria    | Winter        | $^{13}\text{C}$ -bicarbonate | <b>18</b>              |
| Bacteria       | Gammaproteobacteria    | Winter        | $^{13}\text{C}$ -bicarbonate | <b>15</b>              |

## FIGURE LEGENDS

**Figure 1.** Phylogenetic analysis of archaeal (A & B) and bacterial (C & D) 16S rRNA gene sequences for arctic summer (A&C) and winter (B & D) seasons. PCR products were barcoded and sequenced using the Illumina MiSeq platform. Operational Taxonomic Units (OTU) were defined at the 95% identity level for bacteria. Representative sequences from each OTU were chosen at random and their phylogenetic affiliations were determined using QIIME (47). Underlined taxa indicate phylum (division) level classification. Non-underlined taxa are family level assignments (not all are shown).

**Figure 2.** Shown are examples of qPCR analysis of SIP gradient fractions for bacterial and archaeal 16S rRNA gene copies. Relative quantities detected in each fraction are shown as a function of density. All data are normalized to the highest observed quantities and are hence shown as a ratio, where 1 equals the highest observed value. Error bars indicate one standard deviation calculated from three replicate qPCR measurements. The horizontal lines indicates the threshold above which quantities were integrated to calculate average DNA density and percent incorporation. (A) Comparison of winter and summer  $^{14}\text{N}$ -urea ( $\circ$ ) and  $^{15}\text{N}$ -urea ( $\bullet$ ) treatments. (B) Dark carbon fixation SIP experiment showing  $^{12}\text{C}$ -bicarbonate ( $\circ$ ) and  $^{13}\text{C}$ -bicarbonate ( $\bullet$ ) treatments for bacteria and archaea respectively.

# FIGURES

## Figure 1

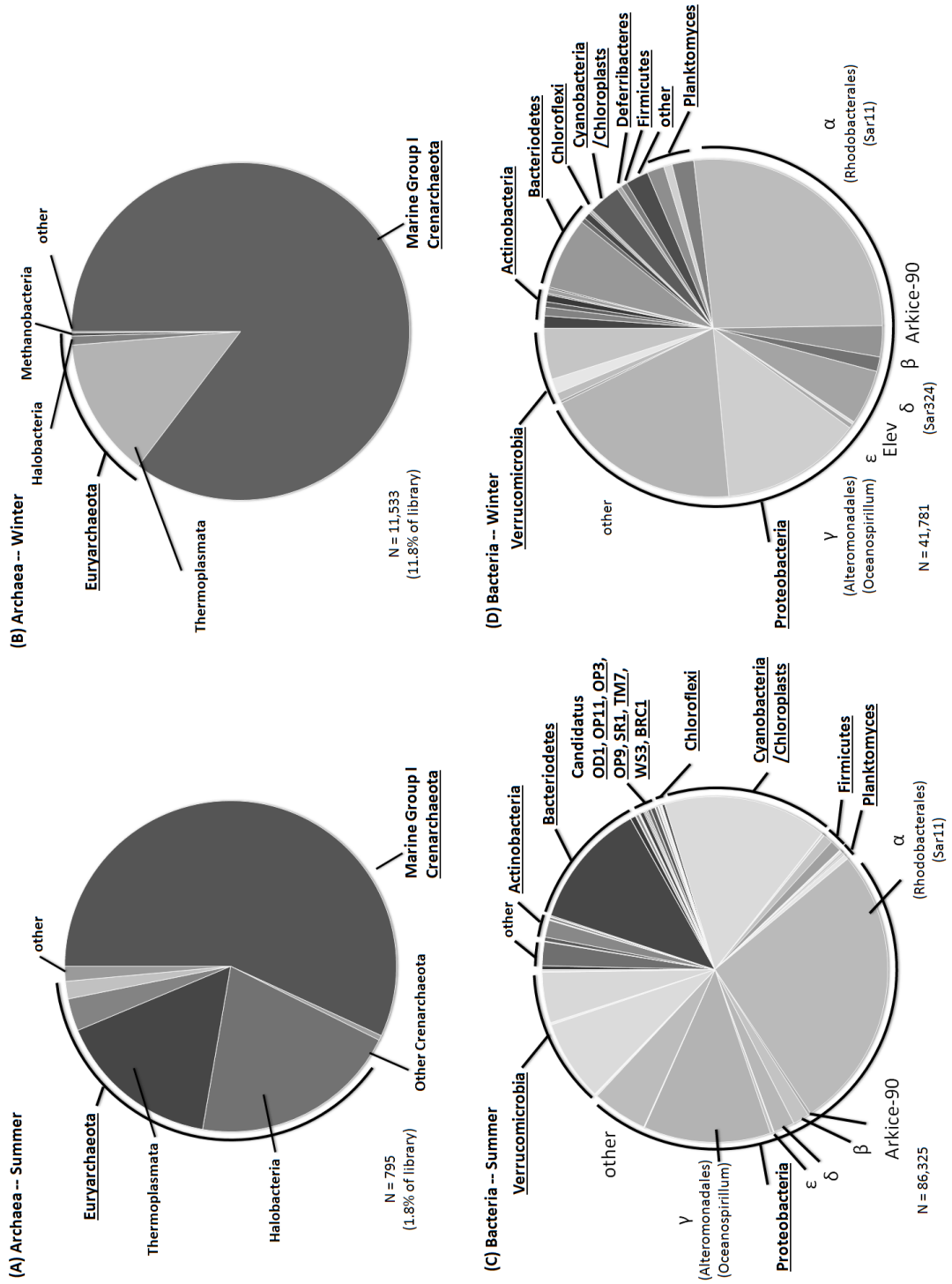
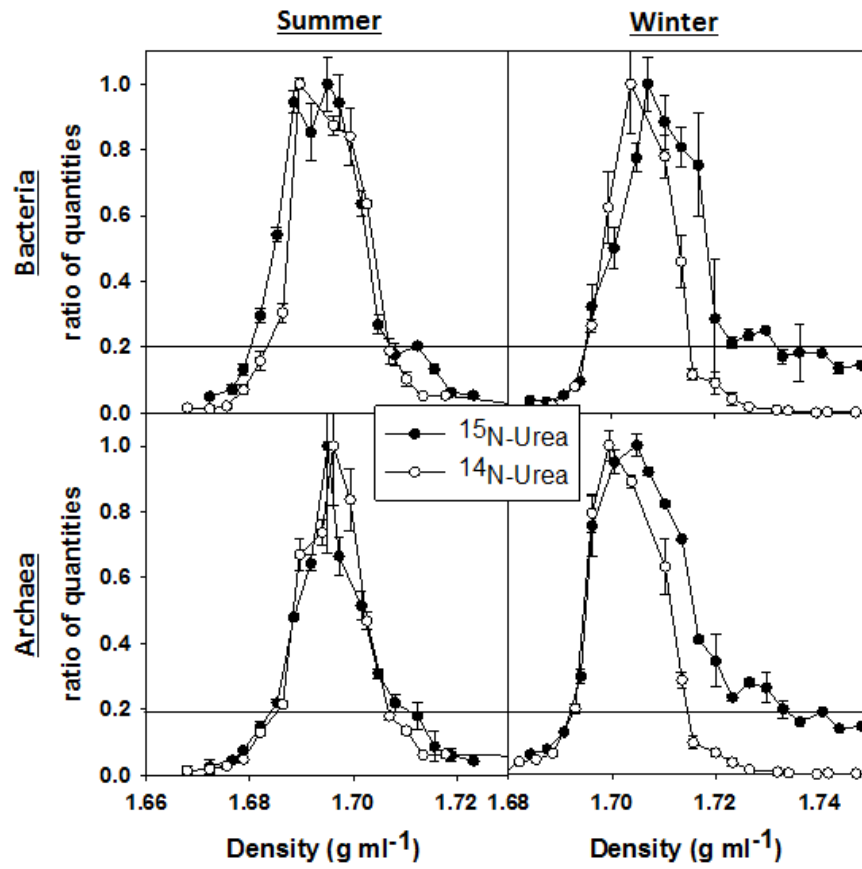
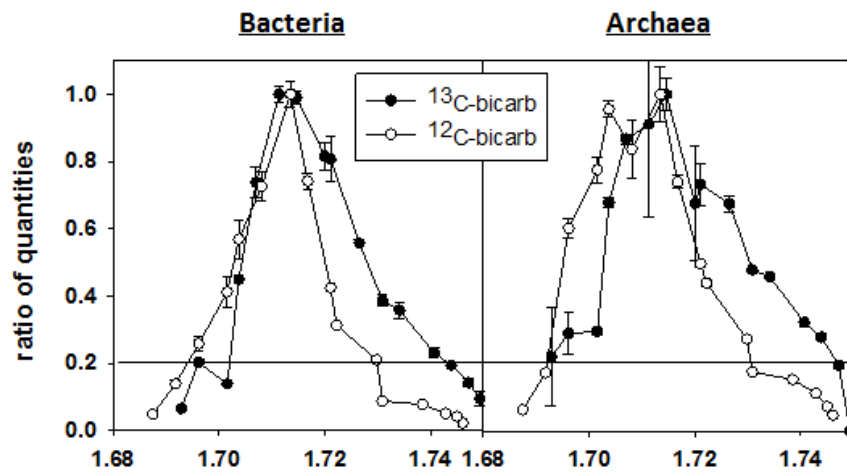


Figure 2

A



B



## **Chapter 2: Transcriptome Analysis of *Scrippsiella trochoidea* CCMP 3099 Reveals Physiological Changes Related to Nitrate Depletion**

This chapter has been published essentially in this form in *Frontiers in Microbiology* and is formatted in accordance to journal specifications. The paper can be located under the following citation:

**Cooper Joshua T.**, Sinclair, Geoff A., Wawrik, Boris. 2016. Transcriptome Analysis of *Scrippsiella trochoidea* CCMP 3099 Reveals Physiological Changes Related to Nitrate Depletion. *Frontiers in Microbiology*,7:639. doi:10.3389/fmicb.2016.00639.

Author contributions to the *Frontiers in Microbiology* paper:

JTC, GS, and BW contributed to experimental design. JTC conducted the experiments, analyzed the data and took the lead in preparing the manuscript. BW and GS contributed to data analysis and manuscript preparation. The work was supported through NSF grant OCE 0961900 and a grant from the Gordon and Betty Moore Foundation to BW.

## Abstract

Dinoflagellates are a major component of marine phytoplankton and many species are recognized for their ability to produce harmful algal blooms (HABs). *Scrippsiella trochoidea* is a non-toxic, marine dinoflagellate that can be found in both cold and tropic waters where it is known to produce “red tide” events. Little is known about the genomic makeup of *S. trochoidea* and a transcriptome study was conducted to shed light on the biochemical and physiological adaptations related to nutrient depletion. Cultures were grown under N and P limiting conditions and transcriptomes were generated via RNAseq technology. *De novo* assembly reconstructed 107,415 putative transcripts of which only 41% could be annotated. No significant transcriptomic response was observed in response to initial P depletion, however, a strong transcriptional response to N depletion was detected. Among the down-regulated pathways were those for glutamine/glutamate metabolism as well as urea and nitrate/nitrite transporters. Transcripts for ammonia transporters displayed both up- and down-regulation, perhaps related to a shift to higher affinity transporters. Genes for the utilization of DON compounds were up-regulated. These included transcripts for amino acids transporters, polyamine oxidase, and extracellular proteinase and peptidases. N depletion also triggered down regulation of transcripts related to the production of Photosystems I & II and related proteins. These data are consistent with a metabolic strategy that conserves N, while maximizing sustained metabolism by emphasizing the relative contribution of organic N sources. Surprisingly, the transcriptome also contained transcripts potentially related to secondary metabolite production, including a homolog to the Short Isoform Saxitoxin gene (*sxtA*) from *Alexandrium fundyense*,



which was significantly up-regulated under N-depletion. A total of 113 unique hits to Sxt genes, covering 17 of the 34 genes found in *C. raciborskii* were detected, indicating that *S. trochoidea* has previously unrecognized potential for the production of secondary metabolites with potential toxicity.

## **Introduction**

Harmful algal blooms (HABs) are a natural phenomenon (Hallegraeff, 1993; Granéli and Turner, 2006b), yet HAB frequencies and apparent ecological pervasiveness have increased within the last several decades. The formation of blooms occurs through the intersection of physical, chemical, and biological processes that are often specific to the HAB species (Paerl, 1988; Granéli and Turner, 2006a). Global prevalence and expansion of harmful algal blooms (HABs) appear, at least in some cases, to be linked to anthropogenic organic and inorganic nutrient loading into estuarine and coastal regions (Glibert et al., 2006; Anderson et al., 2008; Howarth, 2008). Agricultural runoff from the usage of inorganic nitrogen (nitrate and ammonium) has been shown to promote large phytoplankton algal blooms in the Gulf of California (Beman et al., 2005). Inorganic nitrogen has also been shown to promote many HABs as well as organic sources of nitrogen such as urea, glutamine, glycine and amino acids (Baden and Mende, 1979; Mulholland et al., 2002; Dyhrman and Anderson, 2003; Glibert and Legrand, 2006; Cochlan et al., 2008; Kudela et al., 2008). Similarly, HABs also have ways to incorporate organic forms of phosphate using secreted ectoenzymes (alkaline phosphatase) to hydrolyze the organic-P back to inorganic-P for uptake (Sakshaug et al., 1984; Dyhrman, 2005).

While some HAB occurrences appear to be strongly linked to nutrients, others show limited connections (Anderson et al., 2008). Instead, it appears that HAB dynamics exhibit complex relationships with biotic and abiotic factors. The strength of nutrient-HAB relationships are complicated by the variability in HAB adaptations to differing nutrient and light regimes (Smayda, 1997). In addition, many HAB species,

such as some dinoflagellates, are capable of switching their dependence on strict photoautotrophy to mixotrophy by feeding on bacteria, algae (Jeong et al., 2005a; Jeong et al., 2005b), or organic N and P from decaying fish killed by the bloom and zooplankton excretions (Vargo et al., 2008). In areas where nutrients limit growth, dinoflagellates may also migrate vertically to nutrient rich sediments to uptake dissolved N (Sinclair et al., 2006b;a; Sinclair and Kamykowski, 2008) thus alleviating nutrient stress. Given this, significant questions remain about the way in which many HAB species adapt to environmental variability at the molecular and cellular level, the way in which they conserve and utilize diverse dissolved organic and inorganic nutrients as resources are depleted during peak bloom conditions, and how requisite cellular mechanisms are controlled in the context of bloom persistence.

Dinoflagellates are a major component of marine phytoplankton and many species are recognized as toxin producing HABs (Smayda, 1997). Dinoflagellates bloom dynamics involve a complicated life cycle that includes stages of vegetative growth, sexual reproduction, and formation resting cysts (Xiao et al., 2003; Granéli and Turner, 2006a). Non-toxin producing dinoflagellates are less well studied than their toxic counterparts, but can frequently be as devastating to local fisheries via the formation of high-density, high-biomass blooms that result in hypoxia (Horner et al., 1997). *Scrippsiella trochoidea* is a non-toxic, marine dinoflagellate that can be found in both cold and tropic waters where it is known to produce “red tide” events. *Scrippsiella* blooms have been reported extensively from China (Qin et al., 1997; Wang et al., 2007; Zinssmeister et al., 2011), coasts of Japan, Northern Europe, Mediterranean Gulf, Southern Atlantic of Namibia (Montresor et al., 1998; Gottschling et al., 2005; Spatharis

et al., 2009), Southern Gulf of Mexico (Licea et al., 2002), and the coastal United States (Zinssmeister et al., 2011). *Scrippsiella* blooms can become high in cell density and can lead to oxygen depletion resulting in fish kills (Hallegraeff, 1992).

The interplay between inorganic nutrients and *S. trochoidea* bloom formation appears complex. For example, a bloom of *S. trochoidea* in a semi-enclosed bay near Hong Kong maintained high cellular densities in the face of low inorganic nutrients (N, P, Si, metals), and bloom formation could not be stimulated via nutrient addition (Yin et al., 2008). Modeling instead suggests that diel vertical migration of *S. trochoidea*, and wind/tidal currents can cause convergences where cells are concentrated by physical forces even when waters are nutrient depleted (Lai and Yin, 2014). It has also been suggested that HABs may succeed in the wake of preceding nutrient depleting blooms of other phytoplankton species, allowing species adapted to low nutrient concentrations or feeding on bacteria or organic pools to thrive. *S. trochoidea* was traditionally considered to be strictly a photoautotrophic dinoflagellate, however, experimental feeding studies have shown *S. trochoidea* to be mixotrophic, ingesting organic matter or prey including other dinoflagellates, cryptophytes (Jeong et al., 2005b), diatoms (Du Yoo et al., 2009) and bacteria (Jeong et al., 2005a). In fact most photoautotrophic dinoflagellates are now thought to be capable of mixotrophy (Jeong et al., 2005b).

Here, we present a transcriptomic analysis of *S. trochoidea* CCMP 3099, using RNA-seq, designed to examine the effects of nitrogen limitation on gene expression. The purpose of the study was threefold. First, the study was part of the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) (Keeling et al., 2014), which, in part, aimed to characterize the diversity of protein coding genes in a

broad diversity of marine algae. Dinoflagellates have some of the largest known genomes in nature as well as a high estimated genomic repeat content, which has made their genomes poor candidates for previous sequencing efforts. As a consequence, little is known about their complement of protein coding genes. Second, we investigated the transcriptional response triggered by N and P exhaustion, hypothesizing that regulation of gene expression would be in line with physiological adaptations related to nutrient limitation observed in other algal groups. Lastly, the genetic potential for toxin production was investigated by via an analysis of the transcriptome for the presence of transcripts encoding genes involved secondary metabolite production.

## Materials and Methods

### Culture Conditions

Non-axenic cultures of *Scrippsiella trochoidea* CCMP 3099 were obtained from the National Center for Marine Algae and Microbiota (Provasoli-Guillard NCMA, Boothbay Harbor, ME). Cells were maintained in L1 (Guillard and Ryther, 1962;Guillard et al., 1973;Guillard and Hargraves, 1993) seawater media using 0.45  $\mu\text{m}$  filtered, autoclaved natural seawater obtained from the Gulf of Mexico at 33 ppt salinity, which was stored in the dark and aged for at least 3 months. Cultures were grown in a light incubator at 23-24 °C and 30-40  $\mu\text{mol quanta}\cdot\text{m}^{-2}\text{s}^{-1}$  on a 12-h light:12-h dark cycle. Prior to the experiment, cells were grown to early stationary phase in 350 mL of L1 media and served as the inoculum for the nutrient trials at a 1:10 dilution. Concentrations of nitrate and nitrite were determined using the method of (Miranda et al., 2001), which relies on the reduction of nitrate to nitrite with Vanadium(III). Nitrite in V-treated and untreated seawater was then quantified via the addition of acidic Greiss reagent and spectrophotometric detection at 535nm. Phosphate concentrations were determined colorimetrically as per Grasshoff (Grasshoff et al., 1983). Nutrient depletion rates were calculated as the difference in concentration between sequential time points divided by time ( $\Delta C/t$ ). The three treatments included the control (nutrient replete), nitrogen-limited, and phosphorus-limited and were run as static batches. To evaluate nutrient depletion responses in *S. trochoidea*, cultures were set up by modifying the ratio of available N to P. N:P ratios were modified so that cultures would grow into either exhaustion of N or P with roughly similar incubation times. Replete conditions were defined as the normal L1 media containing 880  $\mu\text{M NaNO}_3$  and 36  $\mu\text{M NaH}_2\text{PO}_4$

(N:P ratio 24:1). Nitrogen-limited cells were started with an N:P ratio of 4:1 using 146  $\mu\text{M}$  nitrate and 36  $\mu\text{M}$  phosphate. Phosphorus-limited were started with a higher N:P ratio of 40:1, with nitrogen kept at replete levels (880  $\mu\text{M}$ ) and phosphate at 22  $\mu\text{M}$ . Experimental cultures were grown in larger 1 L volumes of L1 medium in 2500 mL Pyrex Fernbach flasks (10 inch bottom diameter; <1 inch medium) without shaking. All treatment flasks were gently mixed daily during the course of the experiment to eliminate possibility of either carbon limitation or patchy nutrient distribution within the Fernbach flask. Cell counts were generated daily to monitor growth by preservation with 1% formalin (v/v) and counting via light microscopy as well as using a Hemocytometer. Growth rates were calculated by using standard growth equations for exponential growth (Guillard, 1973). After sampling, cultures were transferred to new sterile culture flasks (150 ml volumes). The remainder of the N- and P-limited cultures were split, where one of the resulting sub-samples, respectively, was incubated under continued nutrient deplete conditions, while the other treatment was reconstituted to the original nitrate (N-deplete) or phosphate (P-deplete) concentration by nutrient addition (Figure 1).

### **RNA Isolation**

Nutrient concentrations were monitored daily. Once nitrate or phosphorous dropped below detection limits ( $\sim 1 \mu\text{M}$  for phosphate, and  $\sim 0.1 \mu\text{M}$  for nitrate), cultures were allowed to continue growth for an additional 24 hours to encourage complete depletion of the respective nutrients in culture medium and reduce the potential impact of residual nitrogen/phosphorus stored in vacuoles. All treatments were harvested during the mid-exponential growth phase six hours into the 12-hour light cycle. Cells

were gently filtered in 25-50 mL aliquots onto 3  $\mu\text{m}$  Durapore (Millipore) membrane filters at 5 PSI negative pressure to minimize cell lysis. The 3  $\mu\text{m}$  filters have pore sizes sufficiently small enough to capture dinoflagellate cells, while allowing a large portion of bacterioplankton to pass through the filter. Filters were immediately transferred to 2-mL screw cap tube containing 750  $\mu\text{L}$  RLT buffer (QIAGEN, Valencia CA) and ca. 50 mg of muffled glass beads (Biospek, Bartlesville, OK), frozen using liquid  $\text{N}_2$ , and stored at  $-80^\circ\text{C}$  until extraction. All samples for all treatments were taken within a 30 minutes window to reduce the potential effect of diel variations. For extraction, filters were thawed and cells were lysed by bead beating using a Mini-Bead Beater (BioSpec Products, Bartlesville, OK). Two rounds of beating were conducted at maximum speed for 2 minutes, placing tubes on ice for 2 minutes between steps. Cellular debris and filters were removed by centrifugation at 14,000 x g for 1 min. The supernatants were transferred to QIAshredder (Qiagen, CA, USA) columns to remove residual cellular debris. Total RNA was then extracted using the RNeasy Mini Kit (Qiagen, CA, USA) according to the manufacturers protocol. Genomic DNA bound to the column was removed using an on-column RNase-free DNase I digestion protocol (Qiagen, Valencia CA, USA) as recommended by the manufacturer.

### **RNA Library Preparation and Sequencing**

RNA samples were quantified via a Qubit BR Single Stranded RNA Kit (Life Technologies, Grand Island, NY). Qualities as well as RNA integrity were also assessed using the Agilent 210 Bioanalyzer. Library preparation and sequencing were performed by the National Center for Genomic Research (NCGR). Illumina TruSeq RNA sample preparation started from 2  $\mu\text{g}$  of total RNA. The TruSeq protocol selects for mRNA by



using a poly-T primer for bead capture and subsequent reverse transcription, limiting both ribosomal rRNA and prokaryotic mRNA contamination in the final sequence libraries. After bead capture and cDNA synthesis, libraries were generated by sheering fragments to an average 200-300 bp insert size. Three separate cDNA libraries were sequenced with an Illumina HiSeq200 (Illumina, USA). The original sequence data can be obtained from the NCBI Sequence Read Archive under the accession numbers SRX551166, SRX551167, SRX551168 with MMETSP IDs of MMETSP0270, MMETSP0271, MMETSP0272 corresponding to the replete, nitrogen deplete, and phosphate deplete treatments.

### ***De novo* Transcriptome Assembly**

Several de Bruijn graph assemblers for RNAseq transcriptome reconstruction were assessed (data not shown) including Trinity, Velvet/Oases, and Abyss/Trans-Abyss (Birol et al., 2009; Simpson et al., 2009). Among the tested assembly algorithms, ABySS/Trans-ABySS performed best at reconstructing full-length domains of highly conserved sequences, via repeated blastx (NCBI, blastx) queries. Raw Illumina reads were processed post-sequencing by NCGR to remove adapters. The trimmed read data provided by NCGR, still contained potential adapter artifacts and reads were therefore further processed in house to remove remaining residual adapters using Trimmomatic v.32 (Bolger et al., 2014). Reads were then quality trimmed to remove low quality nucleotides with quality scores <20. The *de novo* transcriptome of *S. trochoidea* was assembled by pooling data from all treatments together. Sequences were assembled using ABySS (v. 1.3.7; (Birol et al., 2009; Simpson et al., 2009) at fifteen different *k-mer* settings ranging between 20 to 50 (stepwise increment of 2). The ‘erode’ flag was

set to zero and the number of pairs to consider a contig was set to ten with scaffolding turned off as described elsewhere (Birol et al., 2009; Simpson et al., 2009). The multiple *k-mer* strategy was chosen as several studies have shown that small *k-mer* values can recover more short transcripts and are likely to be assembled while at larger *k-mer* values, fewer but longer transcripts are assembled (Surget-Groba and Montoya-Burgos, 2010). Large *k-mers* however enhance the possibility of closing gaps in shorter *k-mer* contigs, and thus a hybrid approach may recover complete fragments. Trans-ABYSS version 1.4.8 (Robertson et al., 2010) was used to merge contigs from each single *k-mer* assembly into a final set of contigs. Trans-ABYSS is a conservative merging algorithm, resulting in high redundancy. Contigs were therefore further collapsed and extended using the overlap consensus assembler CAP3 (Huang and Madan, 1999). Further CD-HIT-EST (v4.6; (Li and Godzik, 2006)) was applied as per the manual instructions for clustering expressed sequence tags so that smaller sequences with >90% identity to larger contig sequences would be collapsed. Previous studies of dinoflagellate transcripts suggest the possibility of many copies for individual genes (Bachvaroff et al., 2004; Hackett et al., 2004; Patron et al., 2005; Patron et al., 2006). Without a reference genome the approach taken here is likely a conservative underestimation of the true transcript diversity. To remove residual rRNA signal, blastn (Camacho et al., 2009) was used to compare reads to the SILVA Large Subunit (LSU build 115) and Small Subunit (SSU build 115) databases. Any read that matched a sequence in the SILVA databases with e-values  $<1e^{-50}$  was removed.

## Transcriptome Annotation

All contigs were searched using blastx ( $e < 1e-5$ ) against the NCBI-NR, UniprotKB/Swiss-Prot, and UniprotKB/TREMBL databases. Annotation was conducted as outlined in (De Wit et al., 2012). Briefly, blastx hits are parsed for best hits, skipping hits for “hypothetical” or “unknown” proteins in favor of more descriptive terms. GO terms are assigned from Uniprot searches (De Wit et al. 2012). The pipeline also outputs Kyoto Encyclopedia of Gene and Genomes (KEGG) annotations. Krona was used to explore the taxonomy of hits to the NCBI-NR database and used to create Krona plots (Ondov et al., 2011). KEGG mapper (<http://www.genome.jp/kegg/mapper.html>) was used to examine KEGG biochemical pathway maps for critical pathways such as TCA cycle, Nitrogen Metabolism, Photosynthesis and to examine the overall global transcriptome.

To further annotate sequences with poor blastx hit descriptions or lacked a database match, we choose to implement additional searches using RPS-BLAST against the CDD databases (COG, KOG, PRK, SMART). For this, assembled contigs were translated into potential amino acid sequences using ORFpredictor (Min et al. 2005). ORFpredictor uses a blastx like strategy to search for the best ORF, and also orders contigs in same reading framings. Previous transcriptome studies have reported that ORFpredictor performs well at finding the correct reading frame in dinoflagellate transcriptomes (Jaekisch et al., 2011). Translated potential proteins were also annotated using HMMER3 `hmmsearch` (<http://hmmer.janelia.org>) against PfamA and PfamB (<http://pfam.sanger.ac.uk>), and Tigrfam ([53](http://www.jcvi.org/cgi-</a></p></div><div data-bbox=)

bin/tigrfams/index.cgi) databases using the gathering thresholds for each model instead of an e-value threshold.

### **Analysis for transcriptome completeness**

A lack of genome data for *S. trochoidea* makes it difficult to assess whether transcriptomes have been sequenced at sufficient depth. In genome sequencing projects, genome completeness is often assessed by mapping a transcriptome to the Core Eukaryotic Genes Mapping Approach (CEGMA) to estimate if core eukaryotic genes are present. As applied here, the core set of eukaryotic genes was based originally on the KOG (eukaryotic orthologous genes) refined to 458 core eukaryotic gene families (Parra et al., 2007). CEGMA was further developed to analyze those 248 ultra-conserved core eukaryotic genes (CEGs) that are thought to be present in low copy numbers (Parra et al., 2009). The CEGMA output includes both complete orthologs as well as orthologs that are partial, and calculates a percent completeness for both partial and complete CEGs, in addition the average number of orthologs per family. In addition to *S. trochoidea* transcriptome data, the CEGMA pipeline was also used to analyze previous public datasets downloaded from the NCBI Transcriptome Sequencing Archive, and genomes from JGI-DOE Genome portal that included dinoflagellates and other algae, which share close phylogenetic relationship to dinoflagellates.

### **Read Mapping and Quantification of Gene Expression**

The quality filtered, trimmed and rRNA free, paired-end fastq reads were mapped and aligned to the final transcriptome assembly using Bowtie2 (Langmead and Salzberg, 2012) and Samtools (Li et al., 2009). Bowtie2 was run with the “-sensitive”, “-no-mixed”, “-no-discordant” parameters in “end-to-end” mode to only map reads that

were paired properly and read counts were obtained using HTseq Count program (<http://www.huber.embl.de/users/anders/HTSeq/>) using “union” mode as the method to eliminate multi-mapping and transcripts that cover more than one contig. Examination of the transcriptome reveals fragmented genes and potential gene families. The resolution of the *de novo* transcriptome is limited to gene level and not isoform level quantification. Redundancy in *de novo* transcriptome assemblies may be artificial via the *k*-mer assembly strategy and could represent “real transcript variants/isoforms”, however the more conservative approach to collapse variants/isoforms was applied here. We note here that collapsing unigenes may lead to some loss of signal for differentially expressed paralogs. Raw counts were then used as input into the R package DESeq to detect differentially expressed genes, following the “without replicates” as outline in (Anders and Huber, 2010) and the DESeq manual. To estimate dispersion between samples without replicates, the method “blind”, and sharingMode set to “fit-only”. Only genes with adjusted p-values smaller than 0.1 (representing false discovery rate) were deemed differentially expressed (Benjamini and Hochberg, 1995).

### **Functional Enrichment**

Functional enrichment of gene ontology was estimated using Fishers Exact test in topGO using the parent-child analysis to determine if genes identified as differential expressed were also enriched in biological processes, cellular components, and molecular function (Alexa and Rahnenfuhrer, 2010). The adjusted p-values from the DESeq model of the replete vs. N-limited was used, as there were a number of genes identified as differentially expressed via the Benjamin Hochenberg adjusted p-value < 0.05. The node size was 10, and a p-value < 0.05 was applied to call GO categories as

significantly enriched. Only those nodes that were less than p-value of 0.01 are summarized.

### **Analysis of Secondary Metabolite Genes**

Blast analysis (blastx and blastp; e-values < 1e-5) was used to compare *S. trochoidea* transcriptome sequences with the NCBI NR database, as well as sequence datasets from other algae, including dinoflagellates. For phylogenetic analysis (Figure 5), a subset of sequences was retrieved from the database, which included best database hits, as determined by blast, as well as more distantly related, representative sequences for comparison. Protein sequences were aligned using MAFFT version 7.157b (Kato and Standley, 2013), and trimmed using trimAL with the “automated1” setting optimized for Maximum Likelihood reconstruction (Capella-Gutiérrez et al., 2009). The alignment was checked manually using SeaView 4 (Gouy et al., 2010) and all gap, and sites with sparse numbers of potential homologous amino acids were removed. The alignment was then further evaluated for the best amino acid substitution model using ProtTest 3 (Darriba et al., 2011), which suggested a model with LG+GAMMA. RAxML-SSE version 8 (Stamatakis, 2014) was used to reconstruct a maximum likelihood tree. The PROTGAMMALG setting in RAxML was used to search for the best tree in 100 searches, and subsequently calculated 100 bootstraps values. Additionally the search method of (Hackett et al., 2013) was applied using protein sequences from *Cylindrospermopsis raciborskii* T3 saxitoxin biosynthesis cluster (Kellmann et al., 2008) to query to the un-translated *S. trochoidea* transcriptome using BLAST (tblastn, e-value < 1e-5) to identify additional Sxt genes in the transcriptome. In addition, secondary metabolite gene prediction was conducted via antiSMASH

(Medema et al., 2011).

## Results

### Culture Conditions

During mid-log growth, cultures were depleting ca.  $25 \mu\text{M Day}^{-1}$  nitrate and  $3.5 \mu\text{M Day}^{-1}$  phosphate from the culture media (Figure 1, Figure 2). Given that the detection limits for the N and P assays utilized here ( $\sim 1 \mu\text{M}$ ), the assumption was made that residual nutrients were insufficient to maintain additional consumption 24 hours after nutrient levels dropped below the limit of detection, and that a physiological response related to N and P depletion should be observed. Maximum growth rates for all three treatments were highest on day 3 with a maximum of 0.86 divisions per day for replete, 0.73 ( $\text{d}^{-1}$ ) for N-limited, and 0.91 ( $\text{d}^{-1}$ ) for P-limited treatments, respectively. Specific growth rates declined after day 3 in all cultures, as cell concentrations increased toward stationary phase and concomitant depletion of inorganic N and P. Immediately after sampling, the remaining culture volume (ca. 300 ml) was split into two equal volume sub-cultures, one of which was spiked with either  $300 \mu\text{M}$  nitrate (for N-limited cultures) or  $25 \mu\text{M}$  phosphate (for P-limited cultures) to demonstrate that growth would resume if limitation was relieved. In spiked flasks, cultures resumed rapid depletion for dissolved nutrients (Figure 1). The unspiked, N-limited culture continued to decline exhibiting net cell death, supporting the notion that the culture was indeed starved for N (Figure 2). Un-spiked P-limited cultures were able to maintain, albeit slow, growth suggesting that cells were not truly P-limited. In addition, measurable phosphate concentrations appeared to increase at and beyond the time of sampling in the P-limited culture (Paired t-test,  $t = -2.46$ ,  $\text{df}=2$ ,  $p=0.1331$ ) (Figure 1). We note that cultures were transferred to a smaller culture flasks after sampling, which might have



affected culture behavior post transfer. Cultures were always maintained with high surface to volume ratios (large Fernbach flasks before sampling, and sideways incubated tissue culture flasks after sampling), to minimize impact of potential self-shading and light limitation.

### **General Transcriptome Features**

The assembled transcriptome of *S. trochoidea* consists of 201.9 million base pairs (Mbp) of sequence represented by 205,934 contigs (Supplemental Table 3). Redundancy reduction was used to collapse sequence isoforms and assembly errors using CAP3 and CD-HIT-EST, collapsing transcripts into 107,473 contigs with total length of 125 Mbp (Supplemental Table 3). Of these, 58 contigs produced significant hits to the Silva SSU and LSU rRNA databases and were removed from subsequent analysis, leaving a dataset of 107,415 complete or partial protein coding gene sequences.

The blastx search of the NCBI-NR database yielded annotations for 41% of the total transcriptome. Of these, 93% were assigned to eukaryotes genes, 6% to Bacteria, and 0.4% to Archaea (Figure 3). Examination of hits within the Eukaryota showed that 50% could be assigned to Alveolata, 16% to the Stramenopiles, 11% to the Opisthokonta, 8% to Viridiplantae, while the remaining 15% belonged to a variety of less well represented clades. Most of the hit distribution within the largest group Alveolata yields most hits to *Perkinus marinus* (27%), Dinophyceae (16%), Apicomplexa (5%), and Ciliphora (5%). Blastx searches against Uniprot-SwissProt and Uniprot-TREMBL database yielded a combined 43,797 hits accounting for 40.1% of the total transcriptome as per De Wit (2012). This generated 43,785 annotated

transcripts. Of these, 16,148 transcripts could be assigned to enzymes in KEGG pathways, and 34,480 were assigned GO categories obtained from Uniprot-SwissProt. RPSblast with the conserved domain database (CDD) resulted in 31,725 contigs annotated with at least one CDD number, COG, Pfam or SMART identification.

### **Transcriptome Completeness**

The potential protein coding content of *S. trochoidea* was estimated using the non-linear equation from (Hou and Lin, 2009), which estimates the number of protein coding genes by integrating genomic information from sequenced eukaryotic organisms and cellular DNA content. Typically, this number represents the haploid amount of DNA per cell, however, reports of genome size in *S. trochoidea* differ considerably, and *S. trochoidea* is known to make temporary cysts. Using values estimated by (Rizzo and Noodén, 1973) or (Shuter et al., 1983) of haploid cells at 17 or 34 pg DNA cell<sup>-1</sup> gives a general estimate that *S. trochoidea* may contain between 58,464 to 66,579 protein-coding genes.

Alternatively, the completeness of ultra-conserved and conserved eukaryotic genes was estimated using the CEGMA pipeline, which uses a database of 248 single-copy, ultra-conserved eukaryotic genes originally based on the original eukaryotic orthologous groups (KOGs) as a measure of genome completeness (Parra et al., 2009). The CEGMA pipeline predicted 210 ultra-conserved Core Eukaryotic Genes (CEG) in the *S. trochoidea* transcriptome and an average of 2.74 orthologs per complete CEG. Of detected CEGs, 80.95% had more than one ortholog in the transcriptome. If CEGs with partial predictions are included, the number of ultra-CEGs increases to 220, suggesting

that the transcriptome captures ca. 89% of the core genome in *S. trochoidea*, which is consistent with observations for other marine algae (*Supplemental Table 6*).

### **Low Phosphorus Treatment**

The results from nutrient measurements and growth curves (Figures 1 & 2) indicate that cultures did not reach P-limitation by the time of sampling, even though phosphate concentrations had dropped below the limit of detection 24 hours before RNA collection. The absence of limitation is supported by DESeq analysis of the respective transcriptome data, which indicated that only 17 transcripts (<0.016%) were differentially expressed (p-adjusted < 0.1) (Figure 4b). Of these, 14 were down and 3 were up-regulated, while 12 were also significantly DE at an FDR=0.05. Eleven transcripts had matches in the database. Among the down-regulated transcripts were genes for sulfate transport, two homologs for fibrocystin-L, three genes for cell-wall binding proteins, and 3-dehydroquinate synthase. The only up-regulated transcript with a database match encodes the Photosystem Q(B) protein PsbA.

### **Differentially Expressed Genes under N-depletion**

Analysis of the replete versus N-limited treatment using DESeq indicated that 382 transcripts were differentially expressed (DE) at a FDR=0.1. Of these, 215 could be functionally annotated. Among the differentially expressed genes, 178 were also significantly DE at an FDR=0.05. Of the 215 annotated DE genes, 107 were significantly down-regulated, and 108 were up-regulated in comparison to the replete control. DE transcripts were then clustered based on Biological Process and Molecular Function GO categories using topGO. Within each putative category, transcripts were ranked by Log<sub>2</sub>-fold change, with values < 0 representing down-regulation and values >

0 representing up-regulation relative to the control (Tables 1). Analysis of the *S. trochoidea* transcriptome thereby demonstrates significant effects of nitrogen limitation on electron transport chain components, photosynthetic pathways, nitrogen, lipid, carbohydrate, and amino acid metabolism, as well as stress-related transcripts (Table 1), using the adjusted p-values of the DESeq model as input to the GO enrichment test with topGO and an adjusted-*p*-value < 0.01 revealed similar patterns.

### **Nitrogen Metabolism under N-depletion**

The transcriptome contained genes for the transport and reduction of nitrate, as well as homologs of glutamine synthetase and glutamate synthases (GS-GOGAT) corresponding to nuclear and plastid versions of these enzymes (Figure 5). Worth noting is that the detected nitrite reductase transcript, which was most closely related to bacterial NAD(P)H nitrite reductases from the Flavobacteria *Formosa agariphila* KMM 3901 and *Joostella marina* DSM 19592 respectively. None of the primary nitrate assimilation genes were differentially expressed under N-limiting conditions. Several transcripts related to xanthine metabolisms (e.g. 10 contigs annotated as xanthine dehydrogenase), which have recently been proposed to be involved in N storage (Lin et al., 2015a), were detected and were highly but were not differentially expressed. Several nitrogen cycling pathway genes were significantly down-regulated, including two glutamine amidotransferase-like proteins, a urea transporter, arginase, glutamate dehydrogenase, nitrate/nitrite transporter and NADH dependent glutamate synthase. Genes most closely related to ammonia transporters displayed both up and down-regulated expression, with four transcripts experiencing significant down-regulation and two having significantly increased expression levels. Among the up-regulated

transcripts related to N metabolism were genes for the uptake and degradation of proteins or amino acids such as uric-acid permease, xanthine-uracil permease, aspartyl aminopeptidase, polyamine oxidase, extracellular serine proteinase, aliphatic amidase expression-regulating protein, and oligopeptide transporter 6, while Leu/Ile/Val-binding proteins, and vacuolar amino acid transporter were down-regulated. Also among the DE genes were alkaline phosphatase, which was weakly up-regulated under N-depletion, as well as two contigs annotated as spore formation related proteins (spore coat protein A, subtilisin DY).

### **Photosynthesis and Cellular Respiration under N-depletion**

In peridinin containing dinoflagellates such as *S. trochoidea*, the plastid genome has been reduced and mostly moved to the nuclear genome. Plastid DNA is therefore typically present as mini-circles that typically have only two functional genes (Hackett et al., 2004; Nisbet et al., 2004; Nisbet et al., 2008). The polyA-selection strategy was therefore able to retrieve multiple copies of all respective genes involved in the Calvin Cycle, and Photosynthesis (II and I), Cytochrome bc1 complex respiratory unit, and F-type ATPases. Under N-limiting conditions, genes representing transfer of electrons in Photosystems I & II (Photosystem II CP47, Photosystem I P700 A2, Photosystem I P700 A1, Photosystem II CP43, Photosystem II D2) along with addition members of the photosynthetic pathway (Photosystem Q, Cytochrome b6, Cytochrome b6-f complex subunit 4, Cytochrome b559, and Flavodoxin) were significantly down-regulated. This occurred in conjunction with the down-regulation of chloroplastic ATP synthase alpha and beta subunits and phototropin-2 related to chloroplast relocation. Genes involved in terpenoid biosynthesis, represented by 2 transcripts annotated as 1-deoxy-D-xylulose-5-

phosphate reductoisomerase, as well as genes related to carotenoid biosynthesis were up-regulated under N-deplete conditions. Conversely, transcripts related to chlorophyll biosynthesis were expressed at similar levels in both treatments. In addition to photosynthesis related genes, a down-regulation in transcripts pertaining cellular respiration and detoxification of reactive oxygen species was observed (Cytochrome c peroxidase, sodium/calcium exchanger in the mitochondrial membrane, Cytochrome b and Cytochrome c oxidase subunit 1).

### **Carbohydrate and Lipid Metabolism under N-depletion**

KEGG analysis indicates that the *S. trochoidea* transcriptome contains all requisite genes for the reductive pentose phosphate cycle, glycolysis, gluconeogenesis, glyoxylate cycle, galactose degradation, fatty acid biosynthesis (initiation, elongation), beta-oxidation, and the tricarboxylic acid cycle (TCA). Complete pathways for lipid metabolism of several compounds were detected, including triacylglycerol biosynthesis, acylglycerol degradation, ceramide and sphingosine biosynthesis, as well as sphingosine degradation. Also present is a complete set of transcripts necessary to biosynthesize glucose to UDP-glucose, and galactose to UDP-galactose. The primary storage product in dinoflagellates is starch (Seo and Fritz, 2002; Lee, 2008; Dagenais Bellefeuille et al., 2014), and it has been shown that starch biosynthesis may begin with UDP-glucose in dinoflagellates (Deschamps et al., 2008). The transcriptome also suggests terpenoid biosynthesis of C5 isoprenoids via the non-mevalonate pathway. Transcripts related to carbohydrate catabolism of cellulose appear to be significantly up-regulated under N-depletion. These include beta-glucanase, 1,4-beta-D-glucan

cellobiohydrolase B, 1,4-beta-D-glucan cellobiohydrolase B, glucan 1,3-beta-glucosidase, and Endoglucanase-5.

### **Secondary metabolite genes under N-depletion.**

Blast analysis of the *S. trochoidea* transcriptome revealed several transcripts potentially related to secondary metabolite production, including a hit to the Short Isoform Saxitoxin gene (*sxtA*) from *Alexandrium fundyense*, which was up-regulated under N-depletion, as well as several additional polyketide synthases. Blast analysis (blastx and blastp; e-values < 1e-5) was used to compare *S. trochoidea* transcriptome sequences with the NCBI NR database, as well as sequence datasets from other algae, including dinoflagellates. In all, seven transcripts were annotated as “*sxtA* short isoform” in this manner. Five of these were longer than 200 amino acids and were phylogenetically analyzed (Figure 6). The discovery of a homolog to the *sxtA* genes in *S. trochoidea* prompted a search for the remainder of the saxitoxin biosynthetic cluster as described for *Cylindrospermopsis raciborskii* described by (Kellmann et al., 2008). All 34 annotated Sxt peptides from *C. raciborskii* were therefore tblastn (e-value<1e-05) searched against the *S. trochoidea* transcripts. A total of 136 contigs exhibited homology to the saxitoxin biosynthesis cluster. Of these, 113 were unique hits to Sxt genes, covering 17 of the 34 genes found in *C. raciborskii*, indicating that *S. trochoidea* appears to possess homologs to at least half of the biosynthesis pathway involved in the synthesis of saxitoxin (Figure 7). Similarly, *S. trochoidea* also contains 12 Sxt homologs to those found in *Alexandrium tamarense* Group IV (Hackett et al., 2013).

Phylogenetic analysis revealed that *S. trochoidea* *sxtA*-like transcripts formed a well-supported clade with sequences from other dinoflagellates (Figure 6), and that

sequences originating *Aureococcus anophagefferens* appearing basal to dinoflagellate sequences. Cyanobacterial *sxtA* sequences clustered separately from dinoflagellates as has been shown previously (Hackett et al., 2013). Analysis of the Pfam annotation of putative *S. trochoidea* proteins and related sequences reveals the presence of a phosphopantetheine attachment site (PF00550). The *sxtA* short isoform of *A. fundyense* contains only a phosphopantetheine attachment site, while the “long” isoform also contain aminotransferase class I and II domains, which are absent in the transcript, detected in *S. trochoidea*.

In all, 235 putative polyketide synthase hits were detected in *S. trochoidea* when compared to *Emiliana huxleyi*, *Cryptosporidium parvum*, *Ostreococcus lucimarinus*, *Ostreococcus tauri*, *Chlamydomonas reinhardtii* and dinoflagellate annotated transcripts in NCBI. Secondary metabolite prediction by antiSMASH showed these putative PKS hits grouped into ten clusters. Three clusters represented Type I polyketide synthases, three clusters of non-ribosomal peptide synthases, and one cluster for hybrid non-ribosomal polyketide/polyketide synthase. Many of the contigs appears to have single beta-ketoacyl synthase (PKS\_KS) domains, single acyl carrier protein domains (ACP), and polyketide synthase acyl transferase domains (PKS\_AT). Contig 91643 was annotated as a Type I polyketide synthase and was characterized by acyl carrier protein (ACP), repeating beta-keto-acyl synthase (KS), followed by dehydratases (PKS\_DH), ketoreductase (PKS\_KR), and ending with enoylreductase (PKS\_ER) and thioesterase (PKS\_TE). Contig 84599 also had several repeating domains of KS, KR, DH and ended with PKS\_ER and PKS TE. The remaining contigs were highly fragmented and contained single or multiple domains. Regardless, these data point to



significant potential to encode secondary metabolites in the *S. trochoidea* transcriptomes described here.

## Discussion

### General Transcriptome Features and Completeness

Gene expression in eukaryotes is subject to a range of control mechanisms such as the level of transcription, nuclear export, translation, and post-translational modification. It has been noted that a scarcity of transcriptional regulation exists in dinoflagellates (Lin, 2011). Similarly, proteomic analysis of *T. pseudonana* cultures have demonstrated a reduction of proteins related to nitrate reduction under N starvation even though they did not appear differentially expressed (Hockin et al., 2012), and post-transcriptional regulation in the presence of nitrate has been observed in *Cylindrotheca fusiformis* (Poulsen and Kröger, 2005). The present study may therefore underestimate the degree to which gene expression is modulated in *S. trochoidea*.

Though large, the size of *S. trochoidea*'s transcriptome is consistent with studies, which have reported ~49-118K transcripts for dinoflagellates (Bayer et al., 2012; Beauchemin et al., 2012; Zhang et al., 2014). Similarly, ~42,000 predicted coding genes have been observed from sequencing approximately half the nuclear genome of *Symbiodinium minutum* (Shoguchi et al., 2013), which is thought to represent a basal clade of dinoflagellates and contain a smaller genome than other dinoflagellate lineages, while ~36,000 coding genes were reported for an assembly of ~80% of the genome of *Symbiodinium kawagutii* (Lin et al., 2015a). Analysis of ultra-conserved genes further indicates genome completeness of 85-89% indicating that transcriptomes described here are likely a good representation of the potential protein coding potential of *S. trochoidea*. Although selection of polyadenylated transcripts and size filtration were used to capture primarily eukaryotic poly-adenylated mRNA, some prokaryotic

ribosomal RNAs (rRNA) were detected in the transcriptome assembly. However, these represented <0.02% of reads. Assuming that the bacterial RNA pool was >95% ribosomal, it can be estimated that <0.05% of the transcriptomes analyzed here were of bacterial origin and that bacterial contamination is of minor importance. Nevertheless, the best blast hits of *S. trochoidea*'s predicted proteins are often related to bacteria and archaeal homologs. Closer inspection shows that these hits typically exhibit low identity (~20-48% amino acid similarity using blastx) and therefore likely arise as a consequence of the paucity of available protistan (eukaryotic) genomic data. Dinoflagellates also have unique and complex evolutionary histories of endosymbiosis, horizontal gene transfer, and vertical inheritance from both Bacteria and other Eukarya (Beauchemin et al., 2012; Wisecaver et al., 2013).

Approximately half of the transcriptome could be assigned annotations via comparison with NCBI-NR, and Uniprot-Swissprot/TREMBL databases. A lower percentage of assignments could be made via KEGG and KOG match annotations accounting for only approximately 28% and ~22% of transcripts respectively (Table 4), which is consistent with observations for other dinoflagellates (Jaeckisch et al., 2011; Bayer et al., 2012). The paucity of available protistian genome sequences makes annotating these unusual eukaryotes challenging, because sequence similarities to the databases records are frequently <40%, and it remains difficult to determine copy numbers or to identify isoforms. *S. trochoidea*'s transcriptome contains many examples of highly similar transcripts, but it remains unclear to what degree these are artifacts of the multi-*k*-mer assembly strategy. Dinoflagellates have been shown to contain multiple distinct gene copies using EST and Sanger sequencing (Okamoto et al.,

2001;Bachvaroff and Place, 2008), which do not suffer requisite assembly issues. The number of contigs reported here, therefore, is likely to be an over-estimate of the true diversity of proteins produced by *S. trochoidea*. Conversely, post-transcriptional processing and splicing may lead to assembled transcript diversity not reflected in the genome sequence. For example, homology among gene families and significant mRNA editing in dinoflagellates (Zhang and Lin, 2008) may produce unique *k*-mers that could produce spurious contigs. Also, 5'-spliced leader sequences produced by dinoflagellates (Zhang et al., 2007) may lead to artifacts, given that the de Bruin-graph depended assembly strategy is limited by *k*-mer diversity within reads, which may be narrow in the 5'-region of transcripts. Lastly, while it has been traditionally assumed that dinoflagellates genes have few introns, recent reports suggest that introns may be more wide spread than previously thought (Bachvaroff and Place, 2008;Orr et al., 2013;Shoguchi et al., 2013;Lin et al., 2015b).

### **Phosphorus Metabolism**

A significant response to low P conditions was not observed in the sequenced transcriptome. Only 17 genes were differentially expressed with respect to the replete treatment, and these were not characteristic of classical phosphorus limitation known from other dinoflagellates (Dyhrman and Palenik, 1999;Morey et al., 2011). For example, alkaline phosphatase is an ectoenzyme that hydrolyzes organic phosphates into dissolved phosphate for subsequent uptake and is commonly used as an indicator of phosphate stress in dinoflagellates (Dyhrman and Palenik, 1999). In *S. trochoidea* alkaline phosphatase activity has been reported to be low in nutrient replete and high in P-deficient cells (Sakshaug et al., 1984). The transcriptome contains several homologs

to alkaline phosphatase, but these were not among the differentially expressed genes. Given this, all subsequent discussion focuses to the N-depletion experiment.

### **Nitrogen metabolism**

*S. trochoidea* displayed classic N starvation characteristics common in photosynthetic eukaryotes, including chlorosis (yellowing of the culture) and down-regulation of photosynthetic electron transport (Turpin, 1991;Morey et al., 2011). In addition, *S. trochoidea* experienced an up-regulation of genes related to amino acid catabolism and transport, consistent with increased processing and recycling of organic N compounds and remodeling of the internal metabolic pathways to compensate for the lack of external N. The observed connection between photosynthesis and N metabolism has long been recognized in green algae, diatoms, terrestrial plants (Turpin et al., 1988;Turpin, 1991), and the dinoflagellate *Alexandrium minutum* (Yang et al., 2010), where photo-acclimation through a reduction in photosystem reaction centers is linked to cells entering stationary-phase. Experiments described here were not designed to test exponential versus stationary-phase effects, and it is therefore difficult to separate the potential impact that growth phase responses may have played. However, both N-replete as well as N-spiked cultures continued to grow past the sampling time, while N-depleted cultures stagnated, indicating that *S. trochoidea* was likely not experiencing resource limitations other than those that were controlled for at the time of sampling.

Glutamine synthetase and glutamate synthase were among the significantly up-regulated transcripts, indicating increased investment in cellular ammonium assimilation potential. In contrast, the detected transcripts for nitrate and nitrite transporters as well as nitrate and nitrite reductases were not differentially expressed.

This is in contrast to observations from diatoms, where nitrate and nitrite reductases are frequently up-regulated during N starvation. For example, in the diatom *Phaeodactylum tricornutum* nitrate assimilation genes including nitrate & nitrite reductase and ammonium transporters have been found up-regulated under N starvation (Maheswari et al., 2010). It should be noted that internal stores of nitrate/nitrite were not assessed in experiments described here and it is possible that these were not entirely depleted by the time of sampling. Phytoplankton replete with nutrients are known to luxury consume and store in excess nutrients in vacuoles (Reynolds, 2006; Lin et al., 2016). In dinoflagellates this is often coupled to behavioral adaptation, such as vertical migration into nutrient rich sediments, where N uptake can support growth in a more nutrient deplete photic zone (Sinclair et al., 2006b; Sinclair et al., 2009). However, it has been noted that *S. trochoidea* does not appear to luxury consume (Flynn et al., 1996), and it is therefore more likely that cells adapted to nitrate depletion by adjusting overall cellular N processing and potentially targeting organic N substrates (see discussion below).

Several ammonium transporters were detected and observed to be both up- and down-regulated in *S. trochoidea*'s transcriptome. Typically, ammonium transporter activity is highly expressed under nitrogen limitation. For example, ammonia and nitrogen transporters were highly expressed under nitrogen limitation in *Phaeodactylum tricornutum*, and also in nitrogen-deficient media in the haptophyte *Isochrysis galbana* (Kang et al., 2007). Highest ammonium transporter mRNA was detected in nitrogen-starved *Cylindrothea fusiformis* (Hildebrand, 2005). In *S. trochoidea*, differential transcription of multiple ammonium transporters may reflect the utilization of ammonium transporters with differing affinity as is the case in *Cylindrothea fusiformis*

(Hildebrand, 2005). Upregulation of ammonium transporters has also been reported in prior studies. For example, when *Karenia brevis* (Wilson clone) was nitrate limited, gene expression of type III glutamine synthetases, nitrate/nitrite transporters, and ammonium transporters were all significantly up-regulated (Morey et al., 2011). Similarly, transcriptome analysis of *Alexandrium fundyense* indicated that this organism can utilize ammonium, nitrate, nitrite, urea and potentially cyanate, and when cells were N-limited significant up-regulation of nitrogen transporters, nitrite reductase, and glutamine synthetase was observed (Zhuang et al., 2015). Moustafa et al. (2010) have observed differential gene expression in *Alexandrium tamarense* in response to N and P limitation, as well as the presence or absence of bacteria, and have noted that the presence of bacteria was perhaps the primary driver associated with these changes (Moustafa et al., 2010). Overall, however these data point to the interpretation that dinoflagellates are capable of incorporating many forms of dissolved inorganic and organic nitrogen sources to satisfy their N demand.

Transcriptome data presented here suggest that *S. trochoidea* may have remodeled cellular metabolism to make greater use of dissolved organic nitrogen (DON) compounds, pointing to increased importance of organic-N utilization under N-limiting conditions. This notion is supported by the observation that one of the most highly differentially expressed transcripts was a proton-coupled amino acid transporter with high sequence identity to *Emiliana huxleyi* (Supplemental Table 9). In marine systems, DON production by microbes can be quite high (Bronk et al., 1994), and this is often mirrored by equally high uptake rates of recently produced DON (Bronk and Glibert, 1993). These observations indicate a tight coupling between the production and

consumption of a pool of labile DON that represents an important source of N to the phytoplankton and bacteria in the environment (Sipler and Bronk, 2015). The labile DON pool is composed of metabolites such as amino acids and nucleosides, which are easily integrated into cellular metabolism. The uptake of DON compounds would be facilitated by the observed remodeling of the cellular pathways to compensate for nitrogen flow by recycling amino acids and unconventional nitrogen storage products such as uracil/xanthine. For example, dinoflagellates may store N in the form of uric acid crystals, which could later be catabolized when the cell are stressed (Dagenais-Bellefeuille and Morse, 2013). Elevated expression of a putative xanthine/uracil/vitamin C permease is consistent with transcriptomes reported for *Scrippsiella trochoidea* (this study), *A. anophagefferens* and *T. pseudonana* (Mock et al., 2008; Wurch et al., 2011). While the functional diversity of xanthine permease-like enzymes is broad, they potentially aid in uptake of purines from the environment, and may serve as superior sources of nitrogen when limited. This is supported by a study of *A. anophagefferens* in which xanthine permease was high up-regulated when cells were N limited in both culture and field conditions (Wurch et al., 2014). Overall, it therefore appears that *S. trochoidea* may be well poised to benefit from the rapidly cycling DON pool in lieu of available DIN sources, at least when nitrate is initially depleted.

### **Carbon metabolism**

Nitrogen stoichiometrically limits cellular biosynthesis given that it is a component of essential cellular building blocks including amino acids and nucleotides. Down-regulation of photosynthetic pathways may therefore serve to reduce cellular N demand. In conjunction with photosystem components (though not chl *a* metabolism),



genes involved in mitochondrial respiration and in quenching of reactive oxygen species (ROS) were down-regulated. This is consistent with an overall picture of modulated metabolic activity in *S. trochoidea* in response to N-depletion also observed in other algae. For example, N and P limitation in the toxic alga *Prymnesium parvum* were found to decrease cytochrome and light harvesting gene expression in stationary growth (Beszteri et al., 2012), while N-starvation leads to four-fold down-regulation of genes related to the light harvesting complex in *Emiliana huxleyi* (Dyhrman et al., 2006). A reduction in chlorophyll related to a decrease in the number of PSII reaction centers has also been observed in a range of other eukaryotic marine algae, including such as chlorophytes, diatoms, prymnesiophytes (Berges et al., 1996; Simionato et al., 2013) as well as *Synechococcus* (Görl et al., 1998; Simionato et al., 2013). Down regulation of transcripts in *S. trochoidea* encoding high-nitrogen containing proteins like the photosystem proteins may further reflect changes in cellular stoichiometry to conserve N resources. Proteomic studies of N-limitation in dinoflagellates have shown similar results to our observed transcriptome data with down-regulated processes like carbon fixation, photosynthesis (Lee et al., 2009; Zhang et al., 2015). Alternatively, down-regulations is perhaps a way to compensate for increases in cellular C:N ratios. Measurements in *Scrippsiella trochoidea* and *Alexandrium fundyense*, for example, have suggested increased ratios of particulate organic carbon (POC) to organic nitrogen (PON) under N limitation (Eberlein et al., 2016).

### **Secondary Metabolism**

Prior studies of *S. trochoidea* have not noted toxic compounds found in other dinoflagellates. In fact, the perceived lack of toxicity is routinely cited as the basis for

using this organism as a non-toxic control when testing other dinoflagellates, such as *Alexandrium*, for toxin production (Hold et al., 2001; Smith et al., 2002). Contradictory to this assumption, two *S. trochoidea* isolates have been found to be lethal to Eastern Oysters and Northern Quahog larvae, though not to Sheepshead Minnows (Tang and Gobler, 2012), raising the potential of previously unrecognized toxicity. Tang and Gobler (2012) demonstrated that late exponential and stationary phase cultures induce greater mortality in juvenile shellfish than exponential growth cultures. Experimental treatments of bivalve larvae with live culture, dead (frozen, boiled), and cell-free supernatant suggested a physiochemical mechanisms of toxicity involving either secondary metabolites or chemical constituents from within the algae. The requisite toxic compounds or secondary metabolites have not been identified to date.

Saxitoxin is well known to occur in both Cyanobacteria and dinoflagellates (Hackett et al., 2013) and the presence of saxitoxin was correlated with the detection of the *sxtA* genes by PCR in strains of *Alexandrium tamarense* (Stüken et al., 2011), although *sxtA* genes were also detected in strains in which toxin was not found. Other non-toxic and toxic dinoflagellates also appear to have partial saxitoxin pathways, though it appears that homologs of the C-terminal of *sxtA* and *sxtG* genes are exclusively associated with toxic strains (Hackett et al., 2013). Contigs annotated as *sxtA* in *S. trochoidea* did not include a C-terminus with an aminotransferase domain, suggesting that *S. trochoidea* likely does not produce traditional saxitoxin. The transcriptome also contains several homologs to non-ribosomal peptide synthases (NRPS), and hybrid polyketide/NRPS genes, suggesting previously unrecognized secondary metabolism. A bacterial source of requisite PKS/NRPS sequences cannot be

completely excluded, given that PKS genes have previously been correlated with bacterial presence in dinoflagellate cultures (Snyder et al., 2005), but is unlikely here given selection for eukaryotic mRNA (see discussion above). Their potential involvement in toxicity remains unclear, given that it is quite difficult to ascertain toxicity from gene sequences alone. A detailed metabolite analysis of *S. trochoidea* may therefore be warranted to identify potential toxic secondary metabolites as well as previously unrecognized toxins similar to saxitoxin.

## **Conclusion**

Data collected through the MMETSP project (Keeling et al., 2014) are a valuable resource for the interpretation of molecular datasets collected from the environment. Here we describe the most complete transcriptome of the high-density bloom forming *Scrippsiella trochoidea* to date and demonstrate that transcript abundances are significantly affected by N availability. Overall, observations point to *S. trochoidea*'s ability to flexibly adapt to variations in N availability. These adaptations likely play a central role in near coastal environments where DIN sources are rapidly depleted in bloom conditions. The ability of *S. trochoidea* to switch to a resource utilization pattern that includes DON compounds may help sustain blooms and serve as a model for other persistent red tide events. Transcriptome data further suggests that *S. trochoidea* needs to be reevaluated for the potential to produce toxic secondary metabolites, an observation that may significantly influence the way the impact the way *S. trochoidea* blooms are understood.

## **Conflict of Interest Statement**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## **Acknowledgements**

Funding for this project was in part provided by the Gordon and Betty Moore Foundation as part of the Marine Micro-Eukaryote Transcriptome Project (MMETSP), and in part via a grant from the National Science Foundation (OCE 0961900). We

would also like to thank Henry Neeman at the OU Supercomputing Center for Education & Research for informatics support.

## References

- Alexa, A., and Rahnenfuhrer, J. (2010). topGO: enrichment analysis for gene ontology. *R package version 2.18.0*.
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology* 11, R106.
- Anderson, D.M., Burkholder, J.M., Cochlan, W.P., Glibert, P.M., Gobler, C.J., Heil, C.A., Kudela, R.M., Parsons, M.L., Rensel, J.E.J., Townsend, D.W., Trainer, V.L., and Vargo, G.A. (2008). Harmful algal blooms and eutrophication: Examining linkages from selected coastal regions of the United States. *Harmful Algae* 8, 39-53.
- Bachvaroff, T.R., Concepcion, G.T., Rogers, C.R., Herman, E.M., and Delwiche, C.F. (2004). Dinoflagellate Expressed Sequence Tag Data Indicate Massive Transfer of Chloroplast Genes to the Nuclear Genome. *Protist* 155, 65-78.
- Bachvaroff, T.R., and Place, A.R. (2008). From Stop to Start: Tandem Gene Arrangement, Copy Number and *Trans*-Splicing Sites in the Dinoflagellate *Amphidinium carterae*. *PLOS ONE* 3, e2929.
- Baden, D.G., and Mende, T.J. (1979). Amino acid utilization by *Gymnodinium breve*. *Phytochemistry* 18, 247-251.
- Bayer, T., Aranda, M., Sunagawa, S., Yum, L.K., Desalvo, M.K., Lindquist, E., Coffroth, M.A., Voolstra, C.R., and Medina, M. (2012). *Symbiodinium* Transcriptomes: Genome Insights into the Dinoflagellate Symbionts of Reef-Building Corals. *PLOS ONE* 7, e35269.
- Beauchemin, M., Roy, S., Daoust, P., Dagenais-Bellefeuille, S., Bertomeu, T., Letourneau, L., Lang, B.F., and Morse, D. (2012). Dinoflagellate tandem array gene transcripts are highly conserved and not polycistronic. *Proceedings of the National Academy of Sciences of the United States of America* 109, 15793-15798.
- Beman, J.M., Arrigo, K.R., and Matson, P.A. (2005). Agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean. *Nature* 434, 211-214.
- Berges, J.A., Charlebois, D.O., Mauzerall, D.C., and Falkowski, P.G. (1996). Differential Effects of Nitrogen Limitation on Photosynthetic Efficiency of Photosystems I and II in Microalgae. *Plant Physiology* 110, 689-696.
- Beszteri, S., Yang, I., Jaeckisch, N., Tillmann, U., Frickenhaus, S., Glöckner, G., Cembella, A., and John, U. (2012). Transcriptomic response of the toxic

- prymnesiophyte *Prymnesium parvum* (N. Carter) to phosphorus and nitrogen starvation. *Harmful Algae* 18, 1-15.
- Birol, I., Jackman, S.D., Nielsen, C.B., Qian, J.Q., Varhol, R., Stazyk, G., Morin, R.D., Zhao, Y., Hirst, M., Schein, J.E., Horsman, D.E., Connors, J.M., Gascoyne, R.D., Marra, M.A., and Jones, S.J.M. (2009). De novo transcriptome assembly with ABySS. *Bioinformatics* 25, 2872-2877.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*.
- Bronk, D.A., and Glibert, P.M. (1993). Application of a N-15 Tracer Method to the Study of Dissolved Organic Nitrogen Uptake during Spring and Summer in Chesapeake Bay. *Marine Biology* 115, 501-508.
- Bronk, D.A., Glibert, P.M., and Ward, B.B. (1994). Nitrogen Uptake, Dissolved Organic Nitrogen Release, and New Production. *Science* 265, 1843-1846.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 1-9.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972-1973.
- Cochlan, W.P., Herndon, J., and Kudela, R.M. (2008). Inorganic and organic nitrogen uptake by the toxigenic diatom *Pseudo-nitzschia australis* (Bacillariophyceae). *Harmful Algae* 8, 111-118.
- Dagenais-Bellefeuille, S., and Morse, D. (2013). Putting the N in dinoflagellates. *Frontiers in Microbiology* 4.
- Dagenais Bellefeuille, S., Dorion, S., Rivoal, J., and Morse, D. (2014). The Dinoflagellate *Lingulodinium polyedrum* Responds to N Depletion by a Polarized Deposition of Starch and Lipid Bodies. *PLOS ONE* 9, e111067.
- Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*.
- De Wit, P., Pespeni, M.H., Ladner, J.T., Barshis, D.J., Seneca, F., Jaris, H., Therkildsen, N.O., Morikawa, M., and Palumbi, S.R. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources* 12, 1058-1067.

- Deschamps, P., Guillebeault, D., Devassine, J., Dauvillée, D., Haebel, S., Steup, M., Buléon, A., Putaux, J.-L., Slomianny, M.-C., Colleoni, C., Devin, A., Plancke, C., Tomavo, S., Derelle, E., Moreau, H., and Ball, S. (2008). The Heterotrophic Dinoflagellate *Cryptothecodinium cohnii* Defines a Model Genetic System To Investigate Cytoplasmic Starch Synthesis. *Eukaryotic Cell* 7, 872-880.
- Du Yoo, Y., Jeong, H.J., Kim, M.S., Kang, N.S., Song, J.Y., Shin, W., Kim, K.Y., and Lee, K. (2009). Feeding by Phototrophic Red-Tide Dinoflagellates on the Ubiquitous Marine Diatom *Skeletonema costatum*. *Journal of Eukaryotic Microbiology* 56, 413-420.
- Dyhrman, S. (2005). Ectoenzymes in *Prorocentrum minimum*. *Harmful Algae* 4, 619-627.
- Dyhrman, S.T., and Anderson, D.M. (2003). Urease Activity in Cultures and Field Populations of the Toxic Dinoflagellate *Alexandrium*. *Limnology and Oceanography* 48, 647-655.
- Dyhrman, S.T., Haley, S.T., Birkeland, S.R., Wurch, L.L., Cipriano, M.J., and McArthur, A.G. (2006). Long Serial Analysis of Gene Expression for Gene Discovery and Transcriptome Profiling in the Widespread Marine Coccolithophore *Emiliana huxleyi*. *Applied and Environmental Microbiology* 72, 252-260.
- Dyhrman, S.T., and Palenik, B. (1999). Phosphate Stress in Cultures and Field Populations of the Dinoflagellate *Prorocentrum minimum* Detected by a Single-Cell Alkaline Phosphatase Assay. *Applied and Environmental Microbiology* 65, 3205-3212.
- Eberlein, T., Van De Waal, D.B., Brandenburg, K.M., John, U., Voss, M., Achterberg, E.P., and Rost, B. (2016). Interactive effects of ocean acidification and nitrogen limitation on two bloom-forming dinoflagellate species. *Marine Ecology Progress Series* 543, 127-140.
- Flynn, K., Jones, K.J., and Flynn, K.J. (1996). Comparisons among species of *Alexandrium* (Dinophyceae) grown in nitrogen- or phosphorus-limiting batch culture. *Marine Biology* 126, 9-18.
- Glibert, P., Harrison, J., Heil, C., and Seitzinger, S. (2006). Escalating Worldwide use of Urea – A Global Change Contributing to Coastal Eutrophication. *Biogeochemistry* 77, 441-463.
- Glibert, P.M., and Legrand, C. (2006). "The Diverse Nutrient Strategies of Harmful Algae: Focus on Osmotrophy," in *Ecology of Harmful Algae*, eds. E. Granéli & J. Turner. Springer Berlin Heidelberg, 163-175.



- Görl, M., Sauer, J., Baier, T., and Forchhammer, K. (1998). Nitrogen-starvation-induced chlorosis in *Synechococcus* PCC 7942: adaptation to long-term survival. *Microbiology* 144, 2449-2458.
- Gottschling, M., Knop, R., Plötner, J., Kirsch, M., Willems, H., and Keupp, H. (2005). A molecular phylogeny of *Scrippsiella sensu lato* (Calciodinellaceae, Dinophyta) with interpretations on morphology and distribution. *European Journal of Phycology* 40, 207-220.
- Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution* 27, 221-224.
- Granéli, E., and Turner, J.T. (eds.). (2006a). *Ecology of Harmful Algae*. The Netherlands: Springer Berlin Heidelberg.
- Granéli, E., and Turner, J.T. (2006b). "An Introduction to Harmful Algae," in *Ecology of Harmful Algae*, eds. E. Granéli & J. Turner. Springer Berlin Heidelberg), 3-7.
- Grasshoff, K., Ehrhardt, M., Kremling, K., and Anderson, L.G. (1983). *Methods of Seawater Analysis*. Wiley.
- Guillard, R.R.L. (1973). "Division Rates," in *Handbook of Phycological Methods- Culture Methods and Growth Measurements*, ed. J.R. Stein. (Cambridge: Cambridge University Press), 289-311.
- Guillard, R.R.L., and Hargraves, P.E. (1993). *Stichochrysis immobilis* is a diatom, not a chrysophyte. *Phycologia* 32, 234-236.
- Guillard, R.R.L., Kilham, P., and Jackson, T.A. (1973). Kinetics of silicon-limited growth in the marine diatom *Thalassiosira pseudonana* Hasle And Heimdal (= *Cyclotella nana* Hustedt). *Journal of Phycology* 9, 233-237.
- Guillard, R.R.L., and Ryther, J.H. (1962). Studies of marine planktonic diatoms: I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (Cleve) Gran. *Canadian Journal of Microbiology* 8, 229-239.
- Hackett, J.D., Wisecaver, J.H., Brosnahan, M.L., Kulis, D.M., Anderson, D.M., Bhattacharya, D., Plumley, F.G., and Erdner, D.L. (2013). Evolution of Saxitoxin Synthesis in Cyanobacteria and Dinoflagellates. *Molecular Biology and Evolution* 30, 70-78.
- Hackett, J.D., Yoon, H.S., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Nosenko, T., and Bhattacharya, D. (2004). Migration of the Plastid Genome to the Nucleus in a Peridinin Dinoflagellate. *Current Biology* 14, 213-218.

- Hallegraeff, G.M. (1992). Harmful algal blooms in the Australian region. *Marine Pollution Bulletin* 25, 186-190.
- Hallegraeff, G.M. (1993). A review of harmful algal blooms and their apparent global increase. *Phycologia* 32, 79-99.
- Hildebrand, M. (2005). Cloning and functional characterization of ammonium transporters from the marine diatom *Cylindrotheca fusiformis* (Bacillariophyceae). *Journal of Phycology* 41, 105-113.
- Hockin, N.L., Mock, T., Mulholland, F., Kopriva, S., and Malin, G. (2012). The Response of Diatom Central Carbon Metabolism to Nitrogen Starvation Is Different from That of Green Algae and Higher Plants. *Plant Physiology* 158, 299-312.
- Hold, G.L., Smith, E.A., Rappé, M.S., Maas, E.W., Moore, E.R.B., Stroempl, C., Stephen, J.R., Prosser, J.I., Birkbeck, T.H., and Gallacher, S. (2001). Characterisation of bacterial communities associated with toxic and non-toxic dinoflagellates: *Alexandrium* spp. and *Scrippsiella trochoidea*. *FEMS Microbiology Ecology* 37, 161-173.
- Horner, R.A., Garrison, D.L., and Plumley, F.G. (1997). Harmful algal blooms and red tide problems on the U.S. west coast. *Limnology and Oceanography* 42, 1076-1088.
- Hou, Y., and Lin, S. (2009). Distinct Gene Number-Genome Size Relationships for Eukaryotes and Non-Eukaryotes: Gene Content Estimation for Dinoflagellate Genomes. *PLOS ONE* 4, e6978.
- Howarth, R.W. (2008). Coastal nitrogen pollution: A review of sources and trends globally and regionally. *Harmful Algae* 8, 14-20.
- Huang, X., and Madan, A. (1999). CAP3: A DNA Sequence Assembly Program. *Genome Research* 9, 868-877.
- Jaeckisch, N., Yang, I., Wohlrab, S., Glöckner, G., Kroymann, J., Vogel, H., Cembella, A., and John, U. (2011). Comparative Genomic and Transcriptomic Characterization of the Toxicogenic Marine Dinoflagellate *Alexandrium ostenfeldii*. *PLOS ONE* 6, e28012.
- Jeong, H.J., Park, J.Y., Nho, J.H., Park, M.O., Ha, J.H., Seong, K.A., Jeng, C., Seong, C.N., Lee, K.Y., and Yih, W.H. (2005a). Feeding by red-tide dinoflagellates on the cyanobacterium *Synechococcus*. *Aquatic Microbial Ecology* 41, 131-143.
- Jeong, H.J., Yoo, Y.D., Park, J.Y., Song, J.Y., Kim, S.T., Lee, S.H., Kim, K.Y., and Yih, W.H. (2005b). Feeding by phototrophic red-tide dinoflagellates: five

species newly revealed and six species previously known to be mixotrophic. *Aquatic Microbial Ecology* 40, 133-150.

- Kang, L.-K., Hwang, S.-P.L., Gong, G.-C., Lin, H.-J., Chen, P.-C., and Chang, J. (2007). Influences of nitrogen deficiency on the transcript levels of ammonium transporter, nitrate transporter and glutamine synthetase genes in *Isochrysis galbana* (Isochrysidales, Haptophyta). *Phycologia* 46, 521-533.
- Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30, 772-780.
- Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., Mcmanus, G., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Rynearson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S., Umale, P.E., Vault, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A., and Worden, A.Z. (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.
- Kellmann, R., Mihali, T.K., Jeon, Y.J., Pickford, R., Pomati, F., and Neilan, B.A. (2008). Biosynthetic Intermediate Analysis and Functional Homology Reveal a Saxitoxin Gene Cluster in Cyanobacteria. *Applied and Environmental Microbiology* 74, 4044-4053.
- Kudela, R.M., Lane, J.Q., and Cochlan, W.P. (2008). The potential role of anthropogenically derived nitrogen in the growth of harmful algae in California, USA. *Harmful Algae* 8, 103-110.
- Lai, Z., and Yin, K. (2014). Physical–biological coupling induced aggregation mechanism for the formation of high biomass red tides in low nutrient waters. *Harmful Algae* 31, 66-75.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9, 357-359.

- Lee, F.W.-F., Morse, D., and Lo, S.C.-L. (2009). Identification of Two Plastid Proteins in the Dinoflagellate *Alexandrium affine* That Are Substantially Down-Regulated by Nitrogen-Depletion. *Journal of Proteome Research* 8, 5080-5092.
- Lee, R.E. (2008). *Phycology*. Cambridge University Press.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658-1659.
- Licea, S., Zamudio, M.E., Luna, R., Okolodkov, Y.B., and Gómez-Aguirre, S. (2002). Toxic and harmful dinoflagellates in the southern Gulf of Mexico. *Harmful Algae* 2004, 380-382.
- Lin, S. (2011). Genomic understanding of dinoflagellates. *Research in Microbiology* 162, 551-569.
- Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., Li, L., Zhang, Y., Zhang, H., Ji, Z., Cai, M., Zhuang, Y., Shi, X., Lin, L., Wang, L., Wang, Z., Liu, X., Yu, S., Zeng, P., Hao, H., Zou, Q., Chen, C., Li, Y., Wang, Y., Xu, C., Meng, S., Xu, X., Wang, J., Yang, H., Campbell, D.A., Sturm, N.R., Dagenais-Bellefeuille, S., and Morse, D. (2015a). The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* 350, 691-694.
- Lin, S., Litaker, R.W., and Sunda, W.G. (2016). Phosphorus physiological ecology and molecular mechanisms in marine phytoplankton. *Journal of Phycology* 52, 10-36.
- Lin, X., Wang, L., Shi, X., and Lin, S. (2015b). Rapidly diverging evolution of an atypical alkaline phosphatase (PhoA(aty)) in marine phytoplankton: Insights from dinoflagellate alkaline phosphatases. *Front Microbiol* 6, 868.
- Maheswari, U., Jabbari, K., Petit, J.-L., Porcel, B., Allen, A., Cadoret, J.-P., De Martino, A., Heijde, M., Kaas, R., La Roche, J., Lopez, P., Martin-Jezequel, V., Meichenin, A., Mock, T., Schnitzler Parker, M., Vardi, A., Armbrust, E.V., Weissenbach, J., Katinka, M., and Bowler, C. (2010). Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome Biology* 11, R85.
- Medema, M.H., Blin, K., Cimermancic, P., De Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., and Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis

- gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research* 39, W339-W346.
- Miranda, K.M., Espey, M.G., and Wink, D.A. (2001). A Rapid, Simple Spectrophotometric Method for Simultaneous Detection of Nitrate and Nitrite. *Nitric Oxide* 5, 62-71.
- Mock, T., Samanta, M.P., Iverson, V., Berthiaume, C., Robison, M., Holtermann, K., Durkin, C., Bondurant, S.S., Richmond, K., Rodesch, M., Kallas, T., Huttlin, E.L., Cerrina, F., Sussman, M.R., and Armbrust, E.V. (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, 1579-1584.
- Montresor, M., Zingone, A., and Sarno, D. (1998). Dinoflagellate cyst production at a coastal Mediterranean site. *Journal of Plankton Research* 20, 2291-2312.
- Morey, J., Monroe, E., Kinney, A., Beal, M., Johnson, J., Hitchcock, G., and Van Dolah, F. (2011). Transcriptomic response of the red tide dinoflagellate, *Karenia brevis*, to nitrogen and phosphorus depletion and addition. *BMC Genomics* 12, 346.
- Moustafa, A., Evans, A.N., Kulis, D.M., Hackett, J.D., Erdner, D.L., Anderson, D.M., and Bhattacharya, D. (2010). Transcriptome Profiling of a Toxic Dinoflagellate Reveals a Gene-Rich Protist and a Potential Impact on Gene Expression Due to Bacterial Presence. *PLOS ONE* 5, e9688.
- Mulholland, M.R., Gobler, C.J., and Lee, C. (2002). Peptide hydrolysis, amino acid oxidation, and nitrogen uptake in communities seasonally dominated by *Aureococcus anophagefferens*. *Limnology and Oceanography* 47, 1094-1108.
- Nisbet, R.E.R., Hiller, R.G., Barry, E.R., Skene, P., Barbrook, A.C., and Howe, C.J. (2008). Transcript Analysis of Dinoflagellate Plastid Gene Minicircles. *Protist* 159, 31-39.
- Nisbet, R.E.R., Koumandou, V.L., Barbrook, A.C., and Howe, C.J. (2004). Novel plastid gene minicircles in the dinoflagellate *Amphidinium operculatum*. *Gene* 331, 141-147.
- Okamoto, O.K., Liu, L., Robertson, D.L., and Woodland Hastings, J. (2001). Members of a Dinoflagellate Luciferase Gene Family Differ in Synonymous Substitution Rates. *Biochemistry* 40, 15862-15868.
- Ondov, B.D., Bergman, N.H., and Phillippy, A.M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12, 1-10.

- Orr, R.J.S., Stüken, A., Murray, S.A., and Jakobsen, K.S. (2013). Evolutionary Acquisition and Loss of Saxitoxin Biosynthesis in Dinoflagellates: the Second “Core” Gene, *sxtG*. *Applied and Environmental Microbiology* 79, 2128-2136.
- Paerl, H.W. (1988). Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. *Limnology and Oceanography* 33, 823-843.
- Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061-1067.
- Parra, G., Bradnam, K., Ning, Z., Keane, T., and Korf, I. (2009). Assessing the gene space in draft genomes. *Nucleic Acids Research* 37, 289-297.
- Patron, N.J., Waller, R.F., Archibald, J.M., and Keeling, P.J. (2005). Complex protein targeting to dinoflagellate plastids. *Journal of Molecular Biology* 348, 1015-1024.
- Patron, N.J., Waller, R.F., and Keeling, P.J. (2006). A Tertiary Plastid Uses Genes from Two Endosymbionts. *Journal of Molecular Biology* 357, 1373-1382.
- Poulsen, N., and Kröger, N. (2005). A new molecular tool for transgenic diatoms. *FEBS Journal* 272, 3413-3423.
- Qin, X.-M., Zhou, J.-Z., and Qian, P.-Y. (1997). Effects of Fe and Mn on the growth of a red tide dinoflagellate *Scrippsiella trochoidea* (Stein) Loeblich III. *Chinese Journal of Oceanology and Limnology* 15, 173-180.
- Reynolds, C.S. (2006). *The Ecology of Phytoplankton*. Cambridge University Press.
- Rizzo, P.J., and Noodén, L.D. (1973). Isolation and Chemical Composition of Dinoflagellate Nuclei. *The Journal of Protozoology* 20, 666-672.
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y.S., Newsome, R., Chan, S.K., She, R., Varhol, R., Kamoh, B., Prabhu, A.-L., Tam, A., Zhao, Y., Moore, R.A., Hirst, M., Marra, M.A., Jones, S.J.M., Hoodless, P.A., and Birol, I. (2010). *De novo* assembly and analysis of RNA-seq data. *Nature Methods* 7, 909-912.
- Sakshaug, E., Granéli, E., Elbrächter, M., and Kayser, H. (1984). Chemical composition and alkaline phosphatase activity of nutrient-saturated and P-deficient cells of four marine dinoflagellates. *Journal of Experimental Marine Biology and Ecology* 77, 241-254.
- Seo, K.S., and Fritz, L. (2002). Diel changes in pyrenoid and starch reserves in dinoflagellates. *Phycologia* 41, 22-28.

- Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M., Hamada, M., Seidi, A., Fujie, M., Usami, T., Goto, H., Yamasaki, S., Arakaki, N., Suzuki, Y., Sugano, S., Toyoda, A., Kuroki, Y., Fujiyama, A., Medina, M., Coffroth, Mary a., Bhattacharya, D., and Satoh, N. (2013). Draft Assembly of the *Symbiodinium minutum* Nuclear Genome Reveals Dinoflagellate Gene Structure. *Current Biology* 23, 1399-1408.
- Shuter, B.J., Thomas, J.E., Taylor, W.D., and Zimmerman, A.M. (1983). Phenotypic Correlates of Genomic DNA Content in Unicellular Eukaryotes and Other Cells. *The American Naturalist* 122, 26-44.
- Simionato, D., Block, M.A., La Rocca, N., Jouhet, J., Maréchal, E., Finazzi, G., and Morosinotto, T. (2013). The Response of *Nannochloropsis gaditana* to Nitrogen Starvation Includes *De Novo* Biosynthesis of Triacylglycerols, a Decrease of Chloroplast Galactolipids, and Reorganization of the Photosynthetic Apparatus. *Eukaryotic Cell* 12, 665-676.
- Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., and Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research* 19, 1117-1123.
- Sinclair, G., Kamykowski, D., and Glibert, P.M. (2009). Growth, uptake, and assimilation of ammonium, nitrate, and urea, by three strains of *Karenia brevis* grown under low light. *Harmful Algae* 8, 770-780.
- Sinclair, G.A., and Kamykowski, D. (2008). Benthic–pelagic coupling in sediment-associated populations of *Karenia brevis*. *Journal of Plankton Research* 30, 829-838.
- Sinclair, G.A., Kamykowski, D., Milligan, E., and Schaeffer, B. (2006a). Nitrate uptake by *Karenia brevis*. I. Influences of prior environmental exposure and biochemical state on diel uptake of nitrate. *Marine Ecology Progress Series* 328, 117.
- Sinclair, G.A., Kamykowski, D., Milligan, E., and Schaeffer, B. (2006b). Nitrate uptake by *Karenia brevis*. II. Behavior and uptake physiology in a nitrate-depleted mesocosm with a bottom nutrient source. *Marine Ecology Progress Series* 328, 125.
- Sipler, R.E., and Bronk, D.A. (2015). "Dynamics of Dissolved Organic Nitrogen," in *Biogeochemistry of Marine Dissolved Organic Matter*, eds. D.A. Hansell & C.A. Carlson. Burlington: Academic Press), 127-232.

- Smayda, T.J. (1997). Harmful algal blooms: their ecophysiology and general relevance to phytoplankton blooms in the sea. *Limnology and oceanography* 42, 1137-1153.
- Smith, E.A., Mackintosh, F.H., Grant, F., and Gallacher, S. (2002). Sodium channel blocking (SCB) activity and transformation of paralytic shellfish toxins (PST) by dinoflagellate-associated bacteria. *Aquatic Microbial Ecology* 29, 1-9.
- Snyder, R.V., Guerrero, M.A., Sinigalliano, C.D., Winshell, J., Perez, R., Lopez, J.V., and Rein, K.S. (2005). Localization of polyketide synthase encoding genes to the toxic dinoflagellate *Karenia brevis*. *Phytochemistry* 66, 1767-1780.
- Spatharis, S., Dolapsakis, N.P., Economou-Amilli, A., Tsirtsis, G., and Danielidis, D.B. (2009). Dynamics of potentially harmful microalgae in a confined Mediterranean Gulf—Assessing the risk of bloom formation. *Harmful Algae* 8, 736-743.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312-1313.
- Stüken, A., Orr, R.J.S., Kellmann, R., Murray, S.A., Neilan, B.A., and Jakobsen, K.S. (2011). Discovery of Nuclear-Encoded Genes for the Neurotoxin Saxitoxin in Dinoflagellates. *PLOS ONE* 6, e20096.
- Surget-Groba, Y., and Montoya-Burgos, J.I. (2010). Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Research* 20, 1432-1440.
- Tang, Y., and Gobler, C. (2012). Lethal effects of Northwest Atlantic Ocean isolates of the dinoflagellate, *Scrippsiella trochoidea*, on Eastern oyster (*Crassostrea virginica*) and Northern quahog (*Mercenaria mercenaria*) larvae. *Marine Biology* 159, 199-210.
- Turpin, D.H. (1991). Effects of Inorganic N availability on Algal Photosynthesis and Carbon Metabolism. *Journal of Phycology* 27, 14-20.
- Turpin, D.H., Elrifi, I.R., Birch, D.G., Weger, H.G., and Holmes, J.J. (1988). Interactions between photosynthesis, respiration, and nitrogen assimilation in microalgae. *Canadian Journal of Botany* 66, 2083-2097.
- Vargo, G.A., Heil, C.A., Fanning, K.A., Dixon, L.K., Neely, M.B., Lester, K., Ault, D., Murasko, S., Havens, J., Walsh, J., and Bell, S. (2008). Nutrient availability in support of *Karenia brevis* blooms on the central West Florida Shelf: What keeps *Karenia* blooming? *Continental Shelf Research* 28, 73-98.



- Wang, Z.-H., Qi, Y.-Z., and Yang, Y.-F. (2007). Cyst formation: an important mechanism for the termination of *Scrippsiella trochoidea* (Dinophyceae) bloom. *Journal of Plankton Research* 29, 209-218.
- Wisecaver, J.H., Brosnahan, M.L., and Hackett, J.D. (2013). Horizontal Gene Transfer is a Significant Driver of Gene Innovation in Dinoflagellates. *Genome Biology and Evolution* 5, 2368-2381.
- Wurch, L.L., Gobler, C.J., and Dyhrman, S.T. (2014). Expression of a xanthine permease and phosphate transporter in cultures and field populations of the harmful alga *Aureococcus anophagefferens*: tracking nutritional deficiency during brown tides. *Environmental Microbiology* 16, 2444-2457.
- Wurch, L.L., Haley, S.T., Orchard, E.D., Gobler, C.J., and Dyhrman, S.T. (2011). Nutrient-regulated transcriptional responses in the brown tide-forming alga *Aureococcus anophagefferens*. *Environmental Microbiology* 13, 468-481.
- Xiao, Y., Wang, Z., Chen, J., Lu, S., and Qi, Y. (2003). Seasonal dynamics of dinoflagellate cysts in sediments from daya bay, the south china sea its relation to the bloom of *Scrippsiella trochoidea*. *Acta Hydrobiologica Sinica* 27, 372-377.
- Yang, I., John, U., Beszteri, S., Glockner, G., Krock, B., Goesmann, A., and Cembella, A. (2010). Comparative gene expression in toxic versus non-toxic strains of the marine dinoflagellate *Alexandrium minutum*. *BMC Genomics* 11, 248.
- Yin, K., Song, X.-X., Liu, S., Kan, J., and Qian, P.-Y. (2008). Is inorganic nutrient enrichment a driving force for the formation of red tides?: A case study of the dinoflagellate *Scrippsiella trochoidea* in an embayment. *Harmful Algae* 8, 54-59.
- Zhang, C., Lin, S., Huang, L., Lu, W., Li, M., and Liu, S. (2014). Suppression subtraction hybridization analysis revealed regulation of some cell cycle and toxin genes in *Alexandrium catenella* by phosphate limitation. *Harmful Algae* 39, 26-39.
- Zhang, H., Hou, Y., Miranda, L., Campbell, D.A., Sturm, N.R., Gaasterland, T., and Lin, S. (2007). Spliced leader RNA trans-splicing in dinoflagellates. *Proceedings of the National Academy of Sciences of the United States of America* 104, 4618-4623.
- Zhang, H., and Lin, S. (2008). mRNA editing and spliced-leader RNA trans-splicing groups *Oxyrrhis*, *Noctiluca*, *Heterocapsa*, and *Amphidinium* as basal lineages of dinoflagellates. *Journal of Phycology* 44, 703-711.

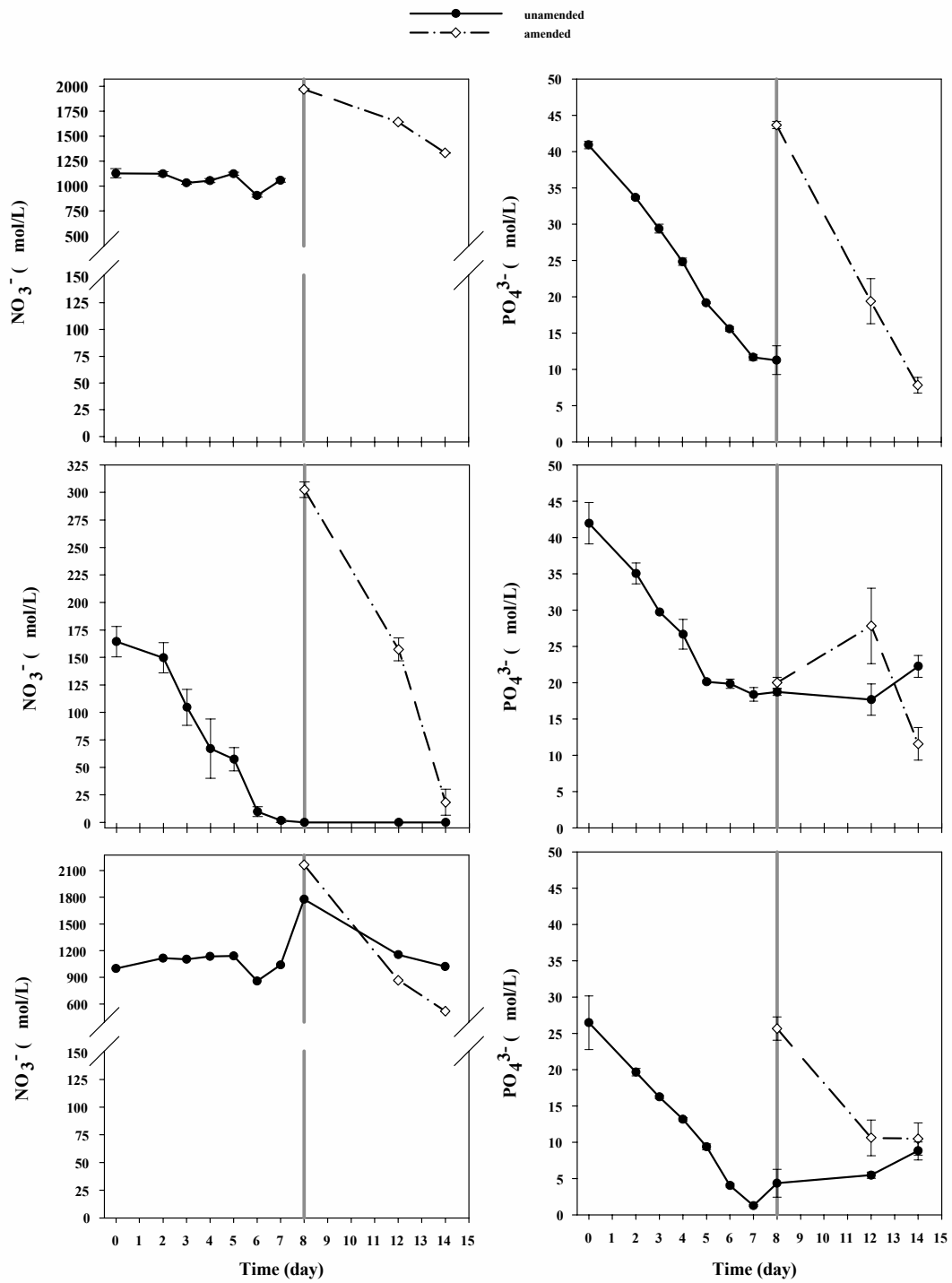
Zhang, Y.J., Zhang, S.F., He, Z.P., Lin, L., and Wang, D.Z. (2015). Proteomic analysis provides new insights into the adaptive response of a dinoflagellate *Prorocentrum donghaiense* to changing ambient nitrogen. *Plant Cell & Environment* 38, 2128-2142.

Zhuang, Y.Y., Zhang, H., Hannick, L., and Lin, S.J. (2015). Metatranscriptome profiling reveals versatile N-nutrient utilization, CO<sub>2</sub> limitation, oxidative stress, and active toxin production in an *Alexandrium fundyense* bloom. *Harmful Algae* 42, 60-70.

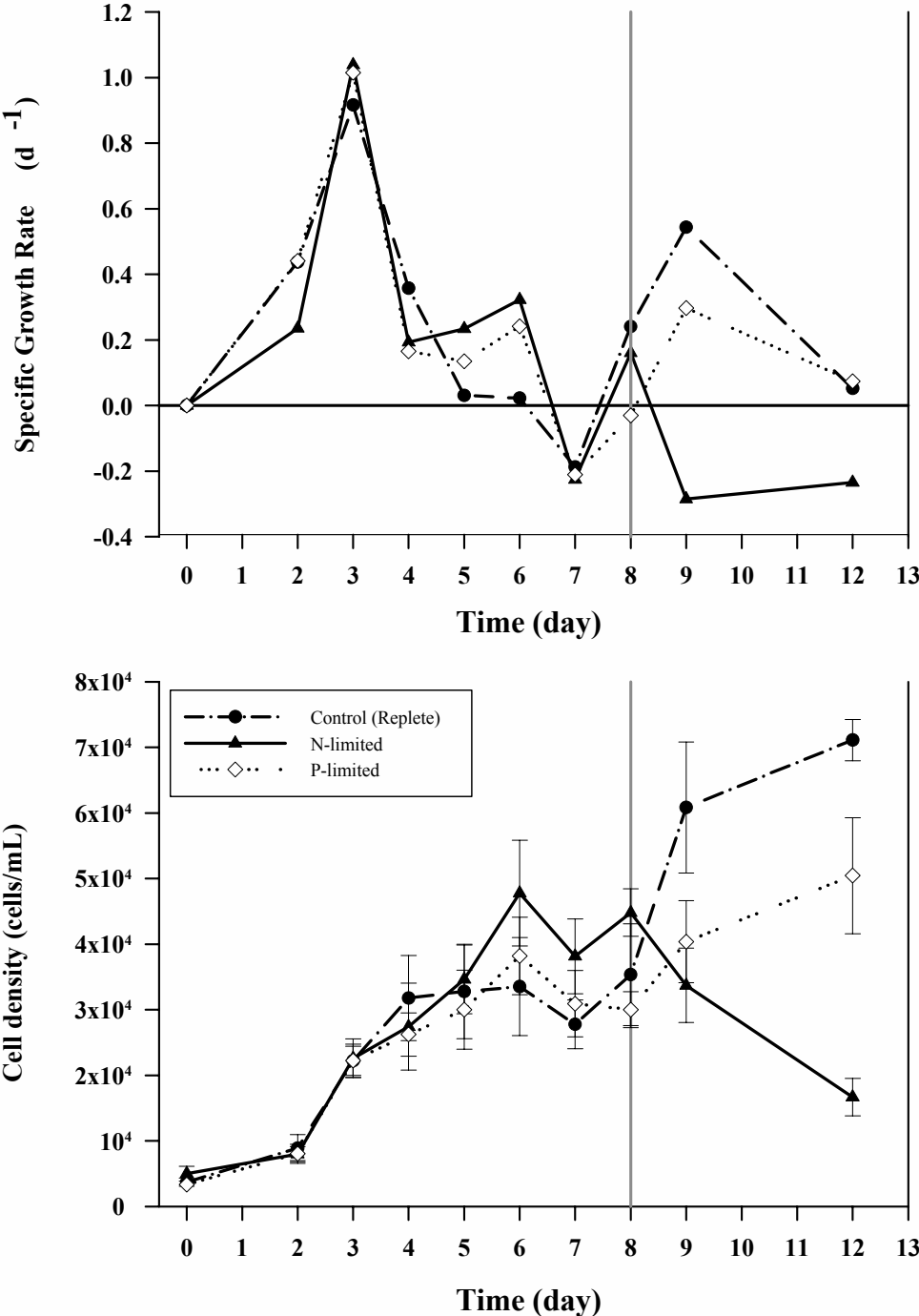
Zinssmeister, C., Soehner, S., Facher, E., Kirsch, M., Meier, K.J.S., and Gottschling, M. (2011). Catch me if you can: the taxonomic identity of *Scrippsiella trochoidea* (F.Stein) A.R.Loeb. (Thoracosphaeraceae, Dinophyceae). *Systematics and Biodiversity* 9, 145-157.

## FIGURES

**Figure 1. (A)** Nitrate measurement illustrating low starting N concentration in the N-limited culture and depletion over time until day 7. Cultures, starved for 24 hours, were sampled on the 8<sup>th</sup> day. The remaining 300 mL were split into 150 mL cultures, and additional nitrate was added to one of the flasks originating from the N-limited treatment. **(B)** Phosphate concentration illustrating lower starting concentration of the P-limited culture, and depletion of P over time until day 7. The culture was starved 24 hours, and sampled on the 8<sup>th</sup> day. The remaining culture was split into two equal volumes and phosphate was to one of the flasks originating from the P-limited treatment.

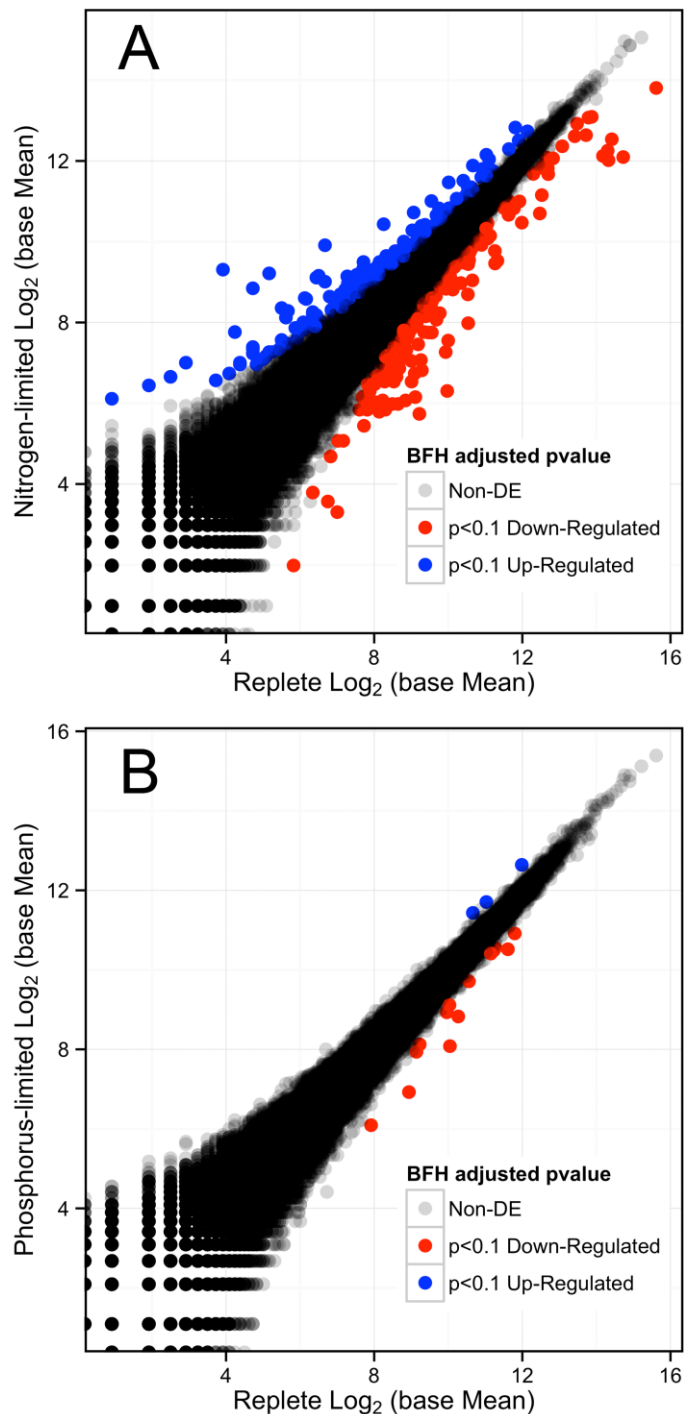


**Figure 2.** Cell densities of cultures for each of the treatments. Error bars indicate one standard deviation. Solid circles indicate the replete (control) culture. Black triangles are used to plot the N-limited culture, while open diamonds are used to plot the P-limited cultures.



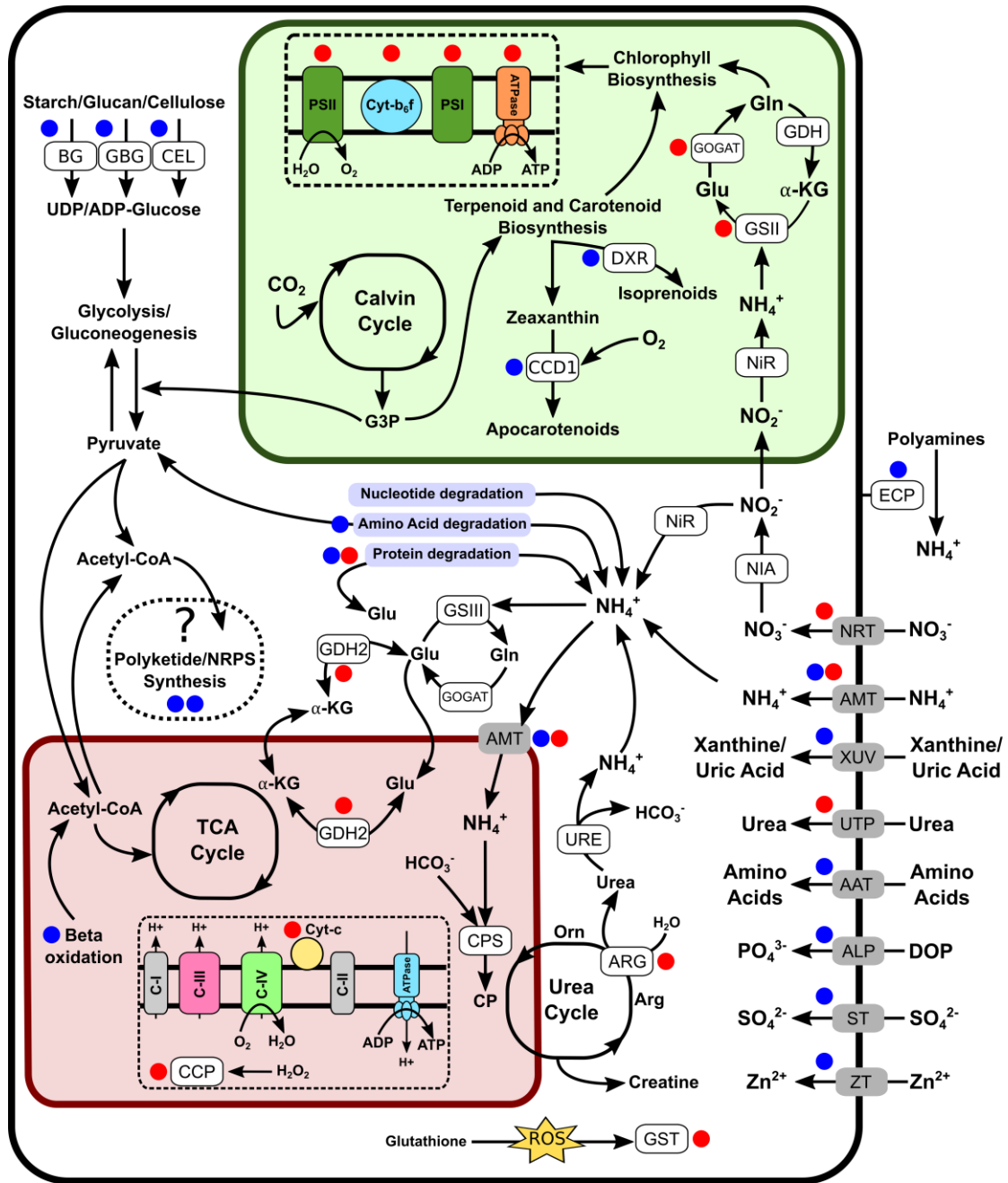


**Figure 4. (A)** Log-Log plot of replete to the N-depletion treatment showing the number of differentially expressed genes in red (down-regulated) and blue (up-regulated, and those that are no DE are in gray-black with p-adjusted to an FDR<0.1. **(B)** Log-Log Plot of replete to the P-limited treatment.



**Figure 5.** Cellular overview of genes detected as differentially expressed and potential role it altering the metabolic pathways of *S. trochoidea* under N-limitation, relative to replete conditions. The glyoxylate cycle is not shown connecting the mitochondrial and plastid pathways for clarity. Blue and Red circles indicate up- and down regulation in response to N-limitation respectively, and indicate the location of genes detected as differentially expressed. AAT, Amino acid transporters; ALP, Alkaline phosphatase; AMT, Ammonium transporter; ARG, Arginase; BG, Beta-glucanase and 1,4-beta-D-glucan cellobiohydrolase; CCP, Cytochrome c peroxidase; CCD1, Carotenoid 9, 10(9',10')-cleavage dioxygenase 1; CEL, Cellulase/Endoglucanase 5 precursor; CP, Carbamoyl phosphate synthetase; DOP, Dissolved organic phosphate; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; ECP, Extracellular peptidases/protease; GBG, Glucan 1,3-beta-glucosidase; GDH2, Glutamate dehydrogenase 2; GSII, Glutamine synthetase II; GSIII, Glutamine synthetase III; GST, Glutathione S-transferase; GOGAT, Glutamate synthetase; NIA, Nitrate reductase; NiR, Nitrite reductase; NRT, Nitrate/Nitrite transporter; STP, Sulfate transporter; URE, Urease; UTP, Urea transporter; XUV, Xanthine/Uric Acid transporter; ZT, Zinc transporter.





Circles Indicate Significant Differential Expression of the N-limited treatment versus the Replete

- Higher Expression
- Lower Expression

  Transcript match

  Transporter

→ Reaction

**Figure 6.** Phylogenetic relationships among saxitoxin *sxtA* from *S. trochoidea* and close blastp hits to NCBI-NR, 20 Maximum likelihood searches and 500 Rapid Bootstraps using RAxML. *sxtA* hits from *S. trochoidea* form a clad with others from dinoflagellates. These are related to the sister clade of *sxtA* from cyanobacteria which are known to produce toxins.

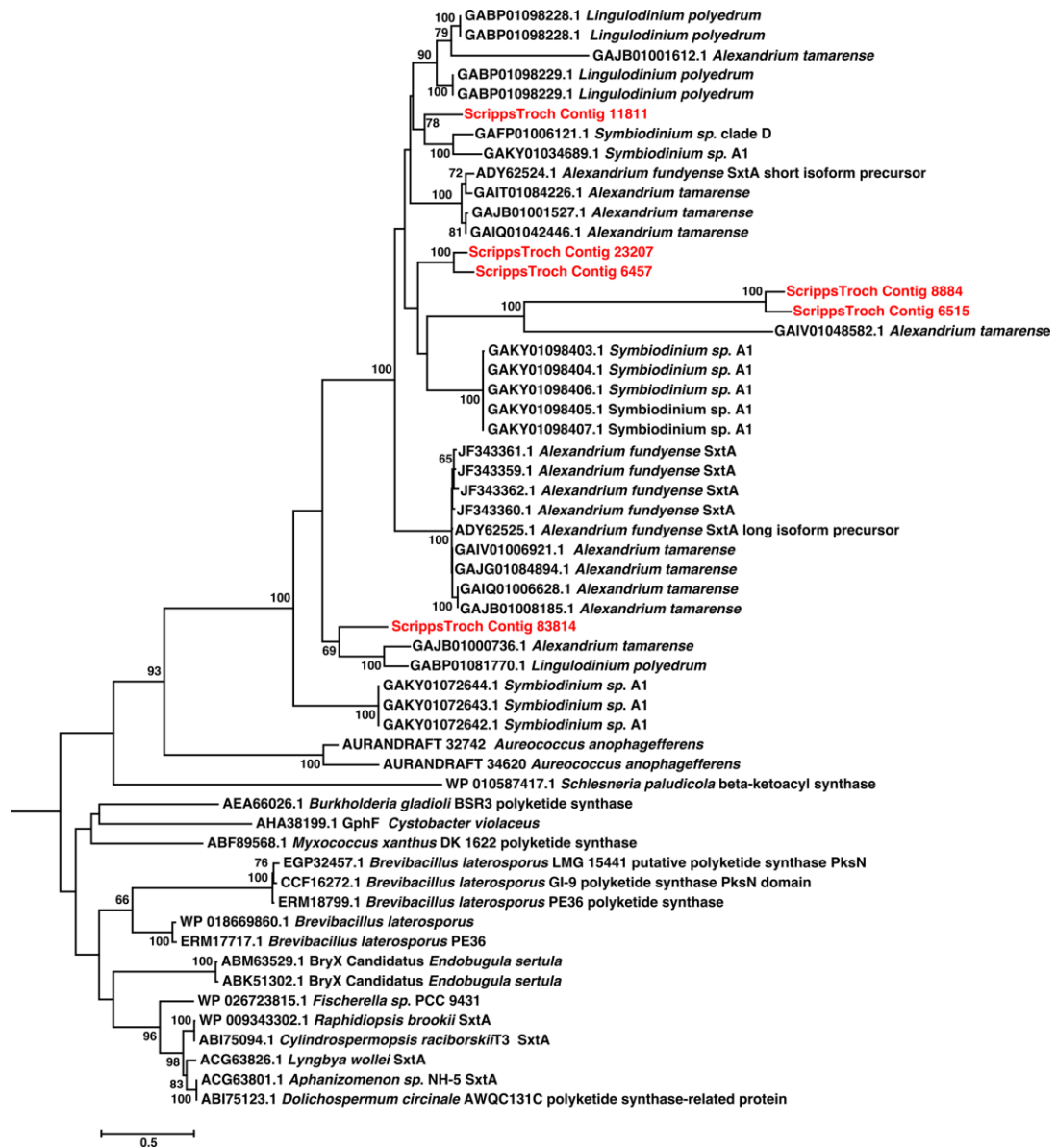
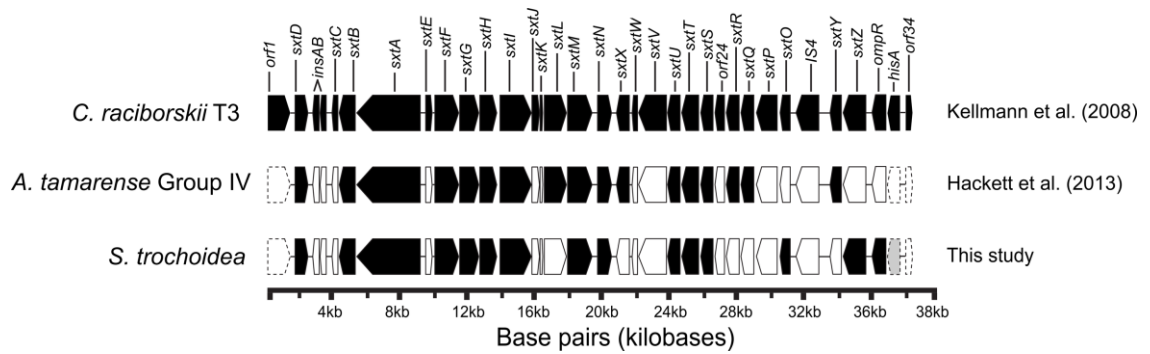


Figure 7. Gene arrangement of Saxitoxin gene cluster in *Cylindrospermopsis raciborskii* with homologs of *Alexandrium tamarensis* and *Scrippsiella trochoidea* shaded to demonstrate shared gene similarity in both toxic and non-toxic dinoflagellates and the known cyanobacterial saxitoxin pathway (Adapted from (Kellmann et al., 2008), and *sxt* gene hits of *Alexandrium tamarensis* plotted from (Hackett et al., 2013) in addition *sxt* gene hits from this study.



## TABLES

Table 1.A Significantly enriched GO terms for Biological Process  $P < 0.01$  with Positive Log2 fold change indicating up-regulation in response to N-depletion. See Supplemental Table 7 for complete list).

| GO term    | Description   | Annotated | Significant | Expected | P-value  |
|------------|---|-----------|-------------|----------|----------|
| GO:0071705 | Nitrogen compound transport   | 591       | 8           | 1.35     | 3.10E-04 |
| GO:0044711 | Single-organism biosynthetic process  | 2748      | 11          | 6.3      | 3.90E-04 |
| GO:0019682 | Glyceraldehyde-3-phosphate metabolic process  | 17        | 2           | 0.04     | 1.22E-03 |
| GO:0008299 | Isoprenoid biosynthetic process   | 120       | 3           | 0.27     | 1.41E-03 |
| GO:0019288 | Isopentenyl diphosphate biosynthetic process,<br>methylerythritol 4-phosphate pathway | 14        | 2           | 0.03     | 1.91E-03 |
| GO:0046490 | Isopentenyl diphosphate metabolic process   | 18        | 2           | 0.04     | 2.28E-03 |
| GO:0044264 | Cellular polysaccharide metabolic process   | 549       | 4           | 1.26     | 2.71E-03 |
| GO:0007588 | Excretion   | 35        | 2           | 0.08     | 3.49E-03 |
| GO:0090407 | Organophosphate biosynthetic process  | 831       | 6           | 1.9      | 3.67E-03 |
| GO:0042886 | Amide transport   | 107       | 3           | 0.25     | 7.21E-03 |
| GO:0006793 | Phosphorus metabolic process  | 3410      | 12          | 7.81     | 7.71E-03 |
| GO:0051189 | Prosthetic group metabolic process  | 14        | 1           | 0.03     | 1.05E-02 |
| GO:0015833 | Peptide transport   | 93        | 3           | 0.21     | 1.07E-02 |

|            |  |      |    |      |          |
|------------|--|------|----|------|----------|
| GO:0009240 | Isopentenyl diphosphate biosynthetic process   | 18   | 2  | 0.04 | 1.23E-02 |
| GO:0006777 | Mo-molybdopterin cofactor biosynthetic process | 14   | 1  | 0.03 | 1.46E-02 |
| GO:0040008 | Regulation of growth                           | 387  | 3  | 0.89 | 1.48E-02 |
| GO:0006857 | Oligopeptide transport                         | 24   | 3  | 0.05 | 1.56E-02 |
| GO:0005976 | Polysaccharide metabolic process               | 695  | 4  | 1.59 | 1.60E-02 |
| GO:0006081 | Cellular aldehyde metabolic process            | 96   | 2  | 0.22 | 1.62E-02 |
| GO:0006720 | Isoprenoid metabolic process                   | 168  | 3  | 0.38 | 1.68E-02 |
| GO:0001558 | Regulation of cell growth                      | 141  | 2  | 0.32 | 1.86E-02 |
| GO:0044765 | Single-organism transport                      | 4055 | 16 | 9.29 | 1.94E-02 |
| GO:0008610 | Lipid biosynthetic process                     | 902  | 5  | 2.07 | 1.99E-02 |
| GO:0000272 | Polysaccharide catabolic process               | 402  | 3  | 0.92 | 2.02E-02 |
| GO:0009726 | Detection of endogenous stimulus               | 24   | 1  | 0.05 | 2.14E-02 |
| GO:0009403 | Toxin biosynthetic process                     | 11   | 1  | 0.03 | 2.15E-02 |

Table 1B. Significantly enriched GO terms for Biological Process  $P < 0.01$  with Negative Log<sub>2</sub> fold change, indicating down-regulation in response to N depletion (See Supplemental Table 8 for complete list).

| GO term    | Description   | Annotated | Significant | Expected | P-value  |
|------------|---|-----------|-------------|----------|----------|
| GO:0022900 | Electron transport chain                                  | 162       | 8           | 0.47     | 5.70E-08 |
| GO:0015979 | Photosynthesis  | 619       | 9           | 1.79     | 2.20E-05 |
| GO:0009767 | Photosynthetic electron transport chain                   | 52        | 6           | 0.15     | 6.30E-05 |
| GO:0018298 | Protein-chromophore linkage                               | 274       | 6           | 0.79     | 7.20E-05 |
| GO:0006351 | Transcription, DNA-templated                              | 1394      | 8           | 4.04     | 1.10E-04 |
| GO:0009636 | Response to Toxic Substance                               | 153       | 4           | 0.44     | 3.20E-04 |
| GO:0044271 | Cellular nitrogen compound biosynthetic process           | 2530      | 12          | 7.33     | 3.20E-04 |
| GO:2000112 | Regulation of Cellular Macromolecule Biosynthetic Process | 1850      | 9           | 5.36     | 4.40E-04 |
| GO:0006091 | Generation of Precursor Metabolites And Energy            | 933       | 9           | 2.7      | 5.20E-04 |
| GO:0055114 | Oxidation-reduction process                               | 902       | 10          | 2.61     | 5.60E-04 |
| GO:0018130 | Heterocycle biosynthetic process                          | 2415      | 11          | 7        | 1.22E-03 |
| GO:0019438 | Aromatic compound biosynthetic process                    | 2444      | 11          | 7.08     | 1.27E-03 |
| GO:0009064 | Glutamine family amino acid metabolic process             | 171       | 5           | 0.5      | 1.49E-03 |
| GO:1901362 | Organic cyclic compound biosynthetic process              | 2620      | 12          | 7.59     | 1.83E-03 |

|            |   |      |    |      |          |
|------------|---|------|----|------|----------|
| GO:0032774 | RNA biosynthetic process                            | 1424 | 8  | 4.12 | 2.17E-03 |
| GO:0010556 | Regulation of Macromolecule Biosynthetic Process    | 1870 | 9  | 5.42 | 2.27E-03 |
| GO:0006811 | Ion transport                                       | 2297 | 17 | 6.65 | 2.32E-03 |
| GO:0030509 | BMP signaling pathway                               | 20   | 2  | 0.06 | 2.47E-03 |
| GO:0051340 | Regulation of ligase activity                       | 31   | 2  | 0.09 | 3.54E-03 |
| GO:0015849 | Organic acid transport                              | 268  | 5  | 0.78 | 3.96E-03 |
| GO:0071705 | Nitrogen compound transport                         | 591  | 7  | 1.71 | 4.08E-03 |
| GO:0035335 | Peptidyl-tyrosine dephosphorylation                 | 17   | 2  | 0.05 | 4.90E-03 |
| GO:0060255 | Regulation of macromolecule metabolic process       | 2824 | 11 | 8.18 | 5.79E-03 |
| GO:0000041 | Transition metal ion transport                      | 78   | 3  | 0.23 | 6.09E-03 |
|            | Regulation of protein modification by small protein |      |    |      |          |
| GO:1903320 | conjugation or removal                              | 139  | 3  | 0.4  | 6.75E-03 |
| GO:0031396 | Regulation of protein ubiquitination                | 125  | 3  | 0.36 | 7.52E-03 |
| GO:0051252 | Regulation of RNA metabolic process                 | 1393 | 8  | 4.03 | 7.88E-03 |

Table 2. Transcriptome gene hits to the saxitoxin biosynthesis gene cluster in *Cylindrospermopsis raciborskii* T3. Approximately 50% of the biosynthesis cluster had significant hits to *S. trochoidea* (17 of the 34).

| Gene Name in<br><i>Cylindrospermopsis raciborskii</i> T3        | Name | Count | Average<br>E-value | Average AA      |               |
|---|------|-------|--------------------|-----------------|---------------|
|   |      |       |                    | %<br>Similarity | HSP<br>Length |
| SxtA polyketide synthase-related protein                        | SxtA | 10    | 1.00E-06           | 48%             | 433.7         |
| SxtB cytidine deaminase   | SxtB | 1     | 7.00E-36           | 55%             | 233.0         |
| SxtD Sterol desaturase  | SxtD | 1     | 7.00E-08           | 46%             | 155.0         |
| SxtF sodium-driven multidrug & toxic compound extrusion protein | SxtF | 1     | 5.00E-04           | 41%             | 358.0         |
| SxtG amidinotransferase   | SxtG | 4     | 5.00E-13           | 50%             | 291.8         |
| SxtH phenylpropionate dioxygenase                               | SxtH | 16    | 1.72E-06           | 44%             | 186.1         |
| SxtI NodU/CmcH-related carbamoyltransferase                     | SxtI | 3     | 3.33E-07           | 50%             | 327.7         |
| SxtN sulfotransferase   | SxtN | 1     | 6.00E-05           | 39%             | 285.0         |
| SxtO adenylylsulfate kinase                                     | SxtO | 2     | 5.10E-45           | 66%             | 177.0         |
| SxtS phytanoyl-CoA dioxygenase                                  | SxtS | 2     | 3.60E-04           | 43%             | 181.0         |
| SxtT phenylpropionate dioxygenase                               | SxtT | 14    | 2.79E-06           | 45%             | 187.0         |
| SxtU short-chain alcohol dehydrogenase                          | SxtU | 53    | 2.14E-05           | 48%             | 146.9         |



|  |      |     |          |     |       |
|--|------|-----|----------|-----|-------|
| SxtW ferredoxin                          | SxtW | 5   | 1.85E-04 | 51% | 64.8  |
| SxtZ histidine kinase                    | SxtZ | 11  | 3.69E-07 | 44% | 278.7 |
| HisA-related protein                     | HisA | 1   | 5.00E-05 | 51% | 73.0  |
| Transcriptional regulator OmpR<br>family | ompR | 10  | 6.92E-07 | 59% | 128.3 |
| Unknown                                  | -    | 1   | 4.00E-22 | 69% | 94.0  |
| Total Sxt gene pathway hits              |      | 136 |          |     |       |
| Number of Unique Contigs                 |      | 112 |          |     |       |

## **Chapter 3: Trophic Differences Reflected in the Elemental Stoichiometry from Transcriptomes and Proteomes of Eukaryotic Phytoplankton**

### **Abstract**

Marine protist communities include a diverse array of eukaryotic phyla that make large contributions to biogeochemical cycling of carbon (C), nitrogen (N), phosphorus (P), and oxygen (O). The biochemistry and abundance of marine protists in marine environments has been shaped over millions of years by fluctuating nutrient conditions, changing redox conditions and climatic variability. This chapter aims to apply a stoichiogenomic framework to examine if N limitation has shaped the elemental composition of marine protist transcriptomes and proteomes, and if life history traits (i.e. autotrophy, mixotrophy, and heterotrophy) correlate with observed variability in stoichiometry. Data from the Marine Microeukaryote Transcriptome Sequencing Project (MMETSP) dataset has provided a snapshot of cellular physiology of 409 marine protists. An analysis of this data indicated that nitrogen and carbon stoichiometry of amino acid side-chain groups have strong phylogenetic signals. The hypothesis that N-limitation experienced by eukaryotic phytoplankton and protists is reflected in N-rich amino acids in autotrophs, lack of enrichment in heterotrophs, and intermediate enrichment in mixotrophs was investigated. Trophic strategy did not appear to significantly correlate with inferred N stoichiometry. Conversely carbon stoichiometry of predicted proteomes in mixotrophic protists was significantly different from heterotrophs ( $p=0.0108$ ), and autotrophs has significantly different carbon stoichiometry as compared to heterotrophs ( $p=0.0108$ ). These data reject the notion that

N-limitation of protists in marine systems is a dominant driving force in shaping proteome composition. Conversely, higher C-content of amino acid side chains in mixotrophic protists suggests a release from limitation of organic C.

## Introduction

Autotrophic and heterotrophic plankton are responsible for large contributions to global fluxes in carbon, nitrogen, sulfur, and oxygen in ocean and freshwater ecosystems. First recognized by Redfield (1958), phytoplankton in the open sea have similar elemental ratios to those found in decomposing matter in the deep ocean (106:16:1 ratio of Carbon to Nitrogen to Phosphorus), leading to the notion that relatively stable elemental ratios exist in phytoplankton biomass regardless of taxonomy and geographical distribution. Stable elemental ratios are largely a consequence of the fact that these elements form the fundamental building blocks of lipids, nucleic acids, and amino acids, thereby linking stoichiometric utilization of these elements to biomass production, though notable deviations from Redfield's ratios have been recognized. For example, large scale patterns of cellular C:N:P ratios among bacterioplankton and phytoplankton have been found to correlate with latitude, irradiance, temperature, and regional impacts (Martiny et al. 2013a, Martiny et al. 2013b, Yvon-Durocher et al. 2015, Thrane et al. 2016). Elemental ratios of oceanic micrograms also appear to be phylogenetically constrained, and may reflect differences in growth rate, cellular investment in structural components, and functional differences (Tett et al. 2009, Hillebrand et al. 2013, Daines et al. 2014). Phytoplankton have been found to vary considerably in their ratios of C:N:P due to differential investments of N and P in cellular DNA and proteins. Their C:N:P stoichiometry can be drastically altered, depending whether cells are growing under nutrient replete or limited conditions (Leonardos and Geider 2004, Geider and Roche 2011). Further, it has been suggested that phytoplankton stoichiometry may be rooted in ancient evolutionary events, such as

the divergence that led to the red and green plastid lineages (Quigg et al. 2010). The same study reports that ratios of C:N and C:P differ among the plastid lineages, pointing to evolutionary events during which nutrients may have been scarce. Additional studies have suggested that a driver of the abundance and diversity of phytoplankton are the result of millions of years of changes in ocean conditions, such as changing redox conditions, fluctuating nutrient availability and climatic variability experienced by organisms over millions of years (Falkowski 1997, Falkowski et al. 2004, Falkowski and Oliver 2007).

Adaptation to nutrient availability has often been invoked as the selective mechanism that might explain observed differences in organismal elemental stoichiometry. In this context, the “growth-rate hypothesis” (GRH) has been articulated and states that slow or reduced organism growth rates may be linked to available phosphorus, and that consumers ingesting P-limited food items will experience slowed growth rates due to the lack of new P-atoms needed to make P-rich ribosomes (Elser et al. 2003, Elser et al. 2008, Giordano et al. 2015). The GRH has been supported in several studies. For example, when algal prey are grown under P-limiting conditions, feeding on these cells alters the elemental stoichiometry and growth rates of their predators (e.g. *Daphnia* or Cladocerans) (Boersma 2000, Acharya et al. 2004, Bullejos et al. 2014, Lind and Jeyasingh 2015). Others have argued that phytoplankton may not necessarily fit the GRH (Flynn et al. 2010).

The ratio of dissolved nitrogen to phosphorus has been shown to limit productivity in aquatic systems, limiting the energy flow in ecosystems (Moore et al. 2013). Conversely, when concentrations of inorganic nutrients are high, they can

promote extensive bloom events of algae (Howarth and Marino 2006). Especially N limitation may impart high physiological costs to phytoplankton, given that dried biomass of the average phytoplankton cells is approximately composed of 32% protein, 17% lipids, and 15% carbohydrates (Finkel et al. 2016). Protein thus accounts for as much as 85% of the cellular N budget (Geider and Roche 2011), but not all amino acids contain equal numbers of N atoms. Among the twenty commonly used amino acids, several contain N atoms in their R-group (arginine=3, histidine=2, asparagine=1, glutamine=1, lysine=1, tryptophan=1, and the rest have zero) (Elser et al. 2011). As a consequence, it is therefore possible that the side chain chemistry of the proteome is under selective pressure to reflect the least number of the atoms of a limiting element. For example, it has been reported that C and S assimilating proteins in *Saccharomyces cerevisiae* and *Escherichia coli* genomes are depleted in these elements (Baudouin-Cornu et al. 2001). Notably it was observed that sulfur atoms were depleted in sulfur assimilatory enzymes but not in enzymes necessary for C assimilation and vice versa, i.e. C atoms were depleted only in C assimilation enzymes but not in those related to sulfur metabolism. This suggests that evolutionary adaptation at the subcellular level can minimize demand for those elements by reducing its utilization in enzymes responsible for their uptake. It has also been shown that a single amino acid change in *S. cerevisiae* can be subjected to strong selection against increasing protein material costs (Bragg and Wagner 2009).

The selection on the elemental composition of an organism's genomic complement is referred to as 'stoichiogenomics' (Elser et al. 2011). Stoichiogenomics, as a framework, aims to understand how organisms interact with their environment and

how their genomes are shaped by environmental interaction such as nutrient stress. A range of studies have observed that reductions in genome size appear to be correlated with environmental differences in nutrient availability, light, and depth distribution in cyanobacteria (García-Fernández et al. 2004, García-Fernández and Diez 2004, Johnson et al. 2006, Martiny et al. 2006, Tolonen et al. 2006, Lv et al. 2008, Martiny et al. 2009a, Martiny et al. 2009b, Martiny et al. 2009c, Partensky and Garczarek 2009, Paul et al. 2010, Bragg 2011, Gilbert and Fagan 2011), heterotrophic bacteria (Grzymiski and Dussaq 2012, Batut et al. 2014, Luo et al. 2015), and microeukaryotes (Rodríguez et al. 2005, Derelle et al. 2006, Palenik et al. 2007). An example of genome streamlining related to environmental differences can be seen in the genomes of several *Prochlorococcus* strains, which represent different ecotypes (high-light vs. low-light) (Chisholm et al. 1988, Moore and Chisholm 1999), where high-light adapted strains have smaller genomes and low-light higher nutrient strains have larger genomes (Rocap et al. 2003). Others have suggested that genomic streamlining might not just be limited to N availability, and that that genome size may also be tied to phosphorus limitation (Hessen et al. 2010). Similarly, codon biases among bacteria have been shown to correlate with their external environmental conditions (Zeldovich et al. 2007, Bragg et al. 2012), and codon optimization may underlie lifestyle related adaptation in fungi (Badet et al. 2017). The study by (Acquisti et al. 2009a) suggests that N depletion in the amino acid side-chains of undomesticated terrestrial plants are the result of long-term adaption to naturally N-poor ecosystems. While cultivation or associations of N-fixing bacterial symbionts (e.g. legumes) release N thrift, and thus have higher levels of N in their proteomes. Plant proteomes also contain ~7% less N in amino acid side chains as

compared to metazoans, suggesting that metazoans do not experience N limitation as severely as plants (Acquisti et al. 2009a). Others suggested that the correlations by Acquisti et al. (2009a) are potentially spurious, and rather reflect other ecological and evolutionary factors (Günther et al. 2013).

On this backdrop, this study aims to investigate the impact of ecophysiology on the stoichiometry of marine protists, which have not previously been investigated in this context. Marine phytoplankton and protists are immensely diverse with respect to the phylogenetic breadth, range of ecological strategies (i.e. strict photoautotrophy, mixotrophy, and heterotrophy), cell sizes, and constituents of their differing cell walls (i.e. silica, calcite, carbon, cellulose). This physiological diversity intersects with strong geochemical nutrient gradients in marine systems, providing an opportunity for stoichiogenic adaptations. Steady coastal nutrient input or upwelling, for example, may relax selection for N and P minimization in organisms adapted to such regimes. Many protists may also alter their trophic strategy from photoautotrophy to mixotrophy to bet-hedge nutrient limiting conditions (Stoecker 1998, Hansen 2011). How microbial eukaryotic plankton regulate their elemental stoichiometry, and what factors may dominate across the diversity of marine plankton lineages remains unexplored and are at the center of work presented here.

The aim of this chapter was to test the following hypotheses:

1.  $H_0$ : The elemental N stoichiome and stoichiome does not differ among autotrophic, mixotrophic, and heterotrophic microeukaryotes.  
 $H_1$ : The N content of the stoichiome and stoichiome are lowest in autotrophs, intermediate in mixotrophs, and highest in heterotrophs.



2. H<sub>0</sub>: the %G+C content does not correlate with stoichiotranscriptome and stoichioproteome N content.  
H<sub>1</sub>: stoichiotranscriptome and stoichioproteome depleted in N are also low in %G+C.
3. H<sub>0</sub>: The stoichiotranscriptome and stoichioproteome does not vary in C enrichment across lineages.  
H<sub>1</sub>: The stoichiotranscriptome and stoichioproteome varies in C content based on trophic strategy of protists.
4. H<sub>0</sub>: The stoichiometry of protein coding genes is unrelated to the environment from which protists were isolated.  
H<sub>1</sub>: Nutrient conditions of source waters are correlated with N depletion signals in the stoichiotranscriptome and stoichioproteome.

To test these hypotheses, the Marine Microeukaryote Transcriptome Sequencing Project (MMETSP) (Keeling et al. 2014) dataset was used. The MMETSP dataset provide a unique opportunity to address these questions, as it represent the currently largest collection of transcriptome data constructed on oceanic microeukaryotes. It contains multiple representatives from all but a few of the major eukaryotic phyla. Most of these phyla do not yet have sequenced genomes and collectively these groups include the bulk of phytoplankton groups worldwide. Of the handful of microeukaryote genomes that are available, most represent laboratory weeds and do not fully represent the biochemical and physiological potential likely found in the diversity of marine free-living protists and phytoplankton. The MMETSP dataset therefore present an

unprecedented opportunity to identify stoichiogenomic and phylogenetic patterns among the highly diverse phytoplankton and protists found in marine systems. Given this, the analysis presented here demonstrates that N stoichiometry of the amino acid side-chains and nucleotide compositions of the transcriptome have strong phylogenetic signals. This indicates that marine protist species are more similar to each other if they share a more recent common ancestor compared to distantly related phylogenetic groups. It was also observed that a significant difference in the C stoichiometry exists among autotrophic, mixotrophic and heterotrophic organisms, while no indication that N limitation shapes the elemental composition of the stoichiotranscriptome and stoichioproteome was observed.

## Methods and Materials

### Data Collection

The Marine Microeukaryote Transcriptome Sequencing Project produced 719 single transcriptomes of 221 genera, 320 species, and comprising 409 individual strains. The MMETSP dataset represents the largest single collection of genetic information regarding marine microeukaryotes (Keeling et al. 2014). The National Center of Genome Research (NCGR) was responsible for sequencing all the libraries and generated provisional assemblies that were retrieved from iMicrobe (<http://data.imicrobe.us/project/view/104>). NCGR quality filtered and assembled all transcriptomes following their internal and publically available pipelines (version 1 and 2) as described in the supplemental methods of (Keeling et al. 2014). NCGR assembled each paired-end library separately into 719 individual assemblies, and combined libraries if they represented the same strain creating composite assemblies. These combined assemblies should represent a more comprehensive transcriptome of an organism than a transcriptome generated from a single treatment, and are typically as representing an estimate of the total expressed genome of the organism. One of the NCGR assemblies (*Anophryoides haemophila* AH6) was not combined due to typographical errors in the naming of the files. For the purpose of this study, these data were manually combined using CD-HIT-EST (Li and Godzik 2006) (`-c 0.98 -word size 9`) and then used alongside the other 408 combined assemblies. The metadata for the MMETSP project was downloaded from CAMERA (<http://marinemicroeukaryotes.org/resources>), revealing that much of the metadata was incomplete. Although several species have multiple strains, the flux of protistian

taxonomy complicates further complicated metadata collection where strains are cryptic species or represent true strains of the same species, especially when strains can differ in their environmental and ecological optima (Lakeman et al. 2009, Balzano et al. 2011, Huertas et al. 2011).

### **Environmental Nutrient Availability**

To investigate whether the transcriptome and *in silico* proteome assemblies correlate with the nutrient environment from where they were isolated, bioclimatic variables were collected corresponding to the GPS locations provided by the NCGR project investigators, or if absent were collected from the corresponding culture collections: Provasoli-Guillard National Center for Marine Algae and Microbiota (NCMA at Bigelow formerly CCMP), SAG (Sammlung von Algenkulturen der Universität Göttingen Culture Collection of Algae at Göttingen University), Roscoff Culture Collection (RCC), Microbial Culture Collection at the National Institute for Environmental Studies (NIES), Canadian Center for the Culture of Microorganisms (CCCM), Culture Collection of Algae at Austin (UTEX), Culture Collection of Algae and Protozoa (CCAP), The Marine Biological Association Culture collection (Plymouth Culture Collection), Council for Scientific & Industrial Research Organization, Australia (CSIRO) North East Pacific Culture Collection, Canada (NEPCC). In several cases, neither culture collections nor metadata from the NCGR contained appropriate GPS coordinates and several strains/species are not in public collections and also do not have publicly available GPS coordinates. In those cases, isolate names or newly described species were used to gather data from original publication describing an

organism's isolation, or other technical reports, and GPS Coordinates were estimated using Google Earth (Supplemental Table 1.)

Environmental layers were extracted using GPS coordinates from interpolated raster layers of the Bio-ORACLE dataset, which correspond to averaged global nutrient, temperature, light, and salinity measurements, derived from the 2009 World Ocean Atlas (Tyberghein et al. 2012). Interpolated ocean bathymetry and the distance to nearest coast were estimated from the MARSPEC (Sbrocco and Barber 2013) dataset using the 5 arc-sec maps that correspond to the resolution of the Bio-ORACLE raster layers. Bioclimatic variables were extracted using the Point Sampling tool within QGIS (Quantum GIS Development Team 2015) and the nitrate ( $\mu\text{M}$ ), phosphate ( $\mu\text{M}$ ), sea surface temperature ( $^{\circ}\text{C}$ ), depth (m), and distance to shore (km).

### **Estimates of Cell Size**

Phytoplankton and protists cell sizes vary by as much as nine orders of magnitude and thus are referred to as a “master trait”. Such master traits have been shown to influence nutrient uptake, sinking rates, as well as R- and K-selected growth rates (Litchman and Klausmeier 2008, Finkel et al. 2009). We approximated cell size of each species/strain by assuming that cell sizes were an equivalent spherical diameter, based on the lengths, as average cell sizes are often not reported beyond the length and widths of cells. Minimum and maximum lengths were used to calculate minimum and maximum cell sizes equivalent spherical diameters (ESD). Because we do not have average cell sizes, maximum and minimum ESDs were used separately. Of the 409 species/strains, only 215 isolates had information about the specific strain dimensions in the metadata. Additional literature searches were able to increase this number to 324

taxa with information. Measurements were taken on images from culture collection websites if provided a scale bar using Fiji (ImageJ) (Schindelin et al. 2012). Collected cell size measurements in Supplemental Table 1.

### **Species Tree of MMETSP**

A phylogeny representing the entire MMETSP dataset was generated using the conserved 18S rRNA gene sequences deposited along with the samples in iMicrobe ([data.imicrobe.us/project/view/104](http://data.imicrobe.us/project/view/104)). However, many of the sequences appeared to be truncated or covered regions that did not align to the larger 1800 bp alignment. To remedy the missing data, phyloFlash (Gruber-Vodicka et al. 2016) was used to iteratively reconstruct 18S rRNA from the MMETSP raw paired-end reads. Briefly, untrimmed reads are mapped using bbmap (Bushnell 2016) against a masked version of the SILVA 123 16/18S full-length rRNA database, and assembled using SPAdes (Bankevich et al. 2012) to reconstruct the small subunit rRNAs by trimming, error correction, and assembly at several  $k$ -mers. Reads are then remapped and passed to EMIRGE (Miller et al. 2011), which uses the new rRNA targets and similar SILVA database hits along with a bait strategy to recruit additional reads. Where phyloFlash produced only partial 18S sequences, reads were additionally reconstructed via SEAVIEW, and a reference 18S sequence was used for scaffolding. Each reconstructed rRNA was then manually aligned and checked using SEAVIEW (Gouy et al. 2010), and verified with NCBI blastn to confirm the phylogenetic identity along with its deposited reference 18S rRNA sequence. For tree building, additional 18S rRNA sequences were obtained from the SILVA SINA dataset (<https://www.arb-silva.de/>). Care was taken to include sufficient reference sequences to break up known long-branch attraction issues

known to occur among Euglenozoa and Amoebozoa (Lahr et al. 2011, Grant and Katz 2014, Lahr et al. 2015). See Supplemental Tables 2 listing of all SILVA accessions.

Testing for phylogenetic signal requires an accurate tree reconstruction, and branch lengths proportional to the amount of time, as opposed to the inferred number of substitutions. A maximum likelihood tree of the 18S rRNAs was converted to an ultrametric tree where all branches are the same length and proportional to time using relative time divergence dating with the RelTime algorithm in MEGA7 (Tamura et al. 2012, Tamura et al. 2013). Relative time dating requires adding divergence times at internal nodes representing the most recent common ancestor relationships among taxa. Divergence dates for internal nodes were selected from published records and times from the TimeTree database (Hedges et al. 2015). Calibration time points between most recent common ancestors within the MMETSP dataset and added taxa are summarized in Supplemental Table 3.

### **Elemental Content of the Transcriptome and Proteome**

The MMETSP dataset includes extensive taxonomic diversity and several eukaryotic groups (i.e. Ciliates and *Euplotes*) are known to utilize alternative genetic codons in both the nuclear and organelle genomes. Transdecoder v1.2 (Haas et al. 2013) was used to translate putative coding sequences from the *de novo* reconstructed mRNA in each NCGR assembly. Transdecoder estimates the hexamer frequency of likely coding transcripts, and peptide sequences were predicted using either the standard genetic code for most of all the assemblies, or specific genetic codes using the *Tetrahymena* code for Ciliates, and the *Euplotes* code for the genera identified as *Euplotes*. This revealed that proteomes predicted by NCGR underestimate the number

of proteins in Ciliates. To generate a more comprehensive proteome prediction for ciliates, the initial predictions output from Transdecoder were supplemented by blastp with an E-value of 1E-5 against a Eukaryotic subset of Uniprot-SwissProt and matches to the Pfam-A database using hmmscan from the HMMER package version 3.1b1 (hmmer.org). Predicted proteins were then parsed using the “retain\_blastp\_hits” and “retain\_pfam\_hits” flags to find support for the most likely open reading frame.

Proteomes generated by Transdecoder were then analyzed to determine the number of N Atoms per Residue Side-Chain (NARSC) following (Acquisti et al. 2009a) defined as  $NARSC = \sum_i (N_i) / \sum_i (l_i)$ , where  $N_i$  represents the number of N atoms in the amino acid side-chain and  $l_i$  represents the length of its protein open reading frame predicted by Transdecoder. The same formula was applied to calculate to N atoms in the coding sequences  $NatomCDS = \sum_i (N_i) / \sum_i (l_i)$ , where  $N_i$  represented the number of N atoms in the coding sequence nucleotides (5 N atoms for A and G, 3 N atoms for C, and 2 N atoms for T/U). Using the same principles for N, the C content of amino acid-side chain chemistry (CARSC) and corresponding transcript coding sequences  $CatomCDS$ . Following the Francois et al. (2016) the average N content of the 3<sup>rd</sup> codon position within the coding sequences were also calculated using the formula above but instead only accounting for the N atoms at the 3<sup>rd</sup> codon position (NAB3).

In order to weight the expression of the transcriptome toward a “control condition” we mapped paired-end reads from each assembly after trimming and used the corresponding transcripts per million (TPM) values to calculate weighted averages of NARSC, CARSC, NAB3, GC% and NatomsCDS and CatomsCDS for each assembly. Briefly, paired-end reads from each MMETSP sample representing replete or



control conditions (no treatment indicated in metadata) were retrieved from iMicrobe as a Metadata file. Each paired-end library was trimmed using `bbduk.sh` from the BBtools suite, version 35 (Bushnell 2016) package using the following settings: `k=27`, `ktrim=rl`, `qtrim=rl`, `minlength=32`, `hdist=1`, `qhist=1`. Reads were then mapped to the corresponding NCGR referenced assembly using Salmon version 0.7.2 (<https://github.com/COMBINE-lab/salmon>) at a  $k$ -mer of 31, (`quant -libtype A, -seqbias, numbootstraps=100`). Salmon uses a quasi-mapping of sequence reads that allow paired reads to map to multiple contigs, and then uses a maximum likelihood approach to estimate which contig the mapped reads likely belong to. This approach is useful because, as *de novo* reconstructed transcriptomes often are likely to have several to multiple copies of transcripts that may correspond to true divergent copies, represent single nucleotide polymorphisms among diploid taxa, or could be the result of sequencing errors and assembly using multiple  $k$ -mers. Salmon outputs reads as both counts and transcripts per million (TPM) where they have been normalized in expression to themselves.

### **Estimating Transcriptome Completeness and Determining Single Copy Orthologs**

Determining the elemental stoichiometry of a transcriptome may require that most of the core genes have been reconstructed that represents its physiology. To determine to what degree *de novo* assemblies represent the genome content of requisite protists, predicted proteins for single copy orthologous were searched against the BUSCO database (Benchmarks of Single Copy Orthologs) as previously described (Simao et al. 2015). Briefly using BUSCO, protein sequences are queried against hidden-Markov-Model representations of highly conserved eukaryotic genes that are

found as single copies orthologs in eukaryotic genomes. Several assemblies were of poor quality (had high numbers of missing BUSCOs, or highly fragmented) and were removed from the analysis if they had more than 50% missing BUSCOs. This excluded 32 samples from the analysis, with most samples having fewer than 30% missing (n=310).

### **Homologs and Orthologs among MMETSP Transcriptomes**

Due to the diversity of protists in the MMETSP dataset, and large differences in protein numbers, data analysis was focused on proteins conserved among species. To gather conserved proteins a previously described phylogenomic pipeline was used (Yang and Smith 2014). This pipeline has been used to show that high quality phylogenetic reconstructions are possible using a combination of transcriptome and reference proteome data. It was used here to construct orthologs among 410 transcriptomes and 47 reference proteomes obtained from Uniprot (Supplemental Table 2), from the Wellcome Sanger Trust, and several university groups linked to sequencing microeukaryotes (Supplemental Information and Supplemental Table 4).

Predicted proteins within each transcriptome assembly were then clustered at 100% similarity to remove potential spurious hits using the UCLUST algorithm from USEARCH (Edgar 2010) (E-value =1E-03, self-hits excluded, and custom fields according to Yang and Smith (2014)). Briefly, proteins were sorted by length to increase the likelihood a full-length protein as the centroid, and then clustered at 100% identity. This *in silico* fractional identity is more similar to the BLAST definition than to similar software CD-HIT (Li and Godzik 2006). Additionally proteins were soft masked using the default “fastamino” within UCLUST to mask simple repeats to reduce

false positives (Edgar 2010). All-by-all UBLAST (Edgar 2010) of clustered proteins against a database of all clustered MMETSP and Reference proteomes were queried (E-value =  $1E-04$ ) and up to 1,500 non-self-hits were recovered. UBLAST hits with query coverage greater than 0.3 we used for homology. Homologous protein families were identified, clustered, and separated using the Markov Clustering Algorithm (MCL v. 14-137; (Enright et al. 2002)) and edges between UBLAST hits with an inflation value of 2.0 and an E-value cutoff of  $1E-5$ , and clusters with fewer than 75 tips were discarded. Several inflation values were attempted but lower values generally created clusters that were too large to align. A higher inflation reduces the coarseness of the clusters and should be a more conservative measure of representative gene families. Sequences of each putative homolog family obtained from MCL were aligned with either MAFFT v. v7.222 (Kato and Standley 2013) or PASTA (Mirarab et al. 2015). Clustered sequences less than 1,000 were aligned using MAFFT with the options “genafpair” and “maxiterate 1000”, while clusters larger than 1,000 sequences were aligned using the iterative PASTA aligner (Mirarab et al. 2015). These clusters represent likely homologs with paralogs, orthologs, and xenologs. To remove spurious xenologs, misassembly, clustering errors, prediction errors, and poorly aligned sequences we performed two rounds of refinement. After the initial pruning of spurious hits, these sequences represent generally homologous families of proteins including orthologs, paralogs, duplications, and deletions.

In each round, homologs were aligned, trimmed to remove positions with more than 10% missing data for smaller (<1,000 members) or 1% for larger (>1,000 members) using Phyutility (Smith and Dunn 2008), and then each cleaned alignment

was used to infer an initial maximum likelihood phylogenetic tree with PROTCATWAG model of amino acid substitution via RAxML 8.2.8 (Stamatakis 2014). Terminal branch tips longer than ten times the sister branches and a relative length of  $> 0.3$  or an absolute length  $> 0.9$  were removed from each putative homolog tree. Paraphyletic and monophyletic tips may belong to isoforms of the same gene from the same taxon were also removed, keeping only the sequence with fewest ambiguities. Long internal branches were cut, if longer than 1.0 to separate deep paralogs, creating two or more subtrees, and had to have at least 250 species or more to be kept. The second round repeated the same relative and absolute trimming steps. Tips were masked again if paraphyletic and monophyletic tips belonged to the same taxon. Long internal branches were cut again if greater than 0.5 and had to have 250 species or more. After this, trees still contained many closely related members, requiring another round of trimming and masking of paraphyletic and monophyletic to be performed and then finally single copy orthologs were output using prune paralogs by maximum inclusion with the same trimming settings as before (relative 0.2, absolute 0.5, and 50 minimum tips). These output trees were converted to fasta files and in a final alignment step was performed using MAFFT. Examination of several trimming methods for eliminating spurious alignments was tried, and trimAL was used instead of Phyutility with the following settings “-resovlerlap 0.5 and -seqoverlap 50 -gappyout.” These output orthologs were used as a way to filter assemblies in order to compare the NARSC values among taxa that differ in their proteome complexity. A final species tree based on these orthologs was attempted but did not succeed, even using supercomputing capacity. In an effort to reduce the dataset by genes that appear to display a tree-like

pattern, matrix reduction using MARE (MAtrix REduction) (<http://mare.zfmk.de>), was attempted, where taxa were weighted more heavily for larger taxonomic coverage. This resulted in 20 protein clusters and was heavily skewed toward Diatoms and Dinoflagellates, which made up a bulk of the MMETSP samples. A final tree was attempted on CIPRES (Miller et al. 2010) with the partitioned dataset, as on OSCER the tree could not complete.

### **Statistical Analysis**

All statistical analyzes were performed in R (R Core Team 2015), and statistical acceptance to reject null hypotheses was set of  $P < 0.05$ . Environmental parameters from the extrapolated oceanic layers were log-transformed prior to statistical analysis to reduce non-normality in the dataset. Phylogenetic signal was tested for each trait using fitContinuous function in “geiger” (Harmon et al. 2008), using Pagel’s  $\lambda$  and Moran’s I phylogenetic correlation index. Pagel’s  $\lambda$  (Pagel 1999) uses a maximum likelihood optimized procedure to determine the estimate the degree of phylogenetic signal of a single trait modeled against Brownian Motion (BM). We also calculated Moran’s I autocorrelation index using the R package “ape” (Paradis et al. 2004), to identify if the scale of the tree displayed differing amounts of phylogenetic signal. The Moran’s I correlogram was calculated using R scripts from (Paquette et. al 2015). We tested whether NARSC, CARSC, the ratio of CARSC/NARSC, NAB3, GC content were predictors of trophic strategy, using Phy-ANOVA with post-hoc tests while controlling for multiple testing using the Holm-Bonferroni method with (nsim=10,000) from the “phytools” package (Revell 2012). A p Phylogenetic Generalized Least Squares Regression, using likelihood ratio tests between models containing explanatory

variables and the model with only an intercept. We then used likelihood ratio tests to compare the best model fit were used to test if environmental nitrate, the ratio of nitrate/phosphate corresponding to the geographic origin of the isolated strain were correlated to any of the calculated N stoichiogenomic measurements. Supplemental Table 4 contains the taxonomic information on the isolates. Supplement Table 5 categorizes the isolates as autotrophic, mixotrophic, heterotrophic, parasitic and symbiotic. Parasitic and symbiotic categories were not included in the analysis as they were represented few samples.

## Results

### Protist Phylogeny

The phylogeny of protists based on the 18S rRNA of the MMETSP samples and additional taxa was well resolved at all but the most ancient branches. This is consistent with previous studies, which have reported difficulties in resolving the divergence of the major eukaryotic clades (Lahr et al. 2011, Parfrey et al. 2011, Grant and Katz 2014, Katz and Grant 2015). The observed phylogeny, nonetheless, is in good agreement with the currently accepted eukaryotic tree of life rooted with Opisthokonta and Fungi (Unikont) as the out-group to the Bikont eukaryotes (Figure 1). Foraminifera, are not represented in our data because they were removed due to the presence of large introns and unreliable alignments, leading to significant long-branch attraction artifacts. Divergence time estimates appear to place the root of the tree at 2.3 billion years ago (bya), and the major clades diverging at similar times as reported in previous studies (Brown and Sorhannus 2010, Parfrey et al. 2011). The split of the major eukaryotic groups occurs with short branch lengths in deep time and is consistent with a pattern of rapid diversification around ~1.9 bya (Brown and Sorhannus 2010, Parfrey et al. 2011). Diatoms are the most recently evolved lineage, originating ca. 120-530 million years ago (mya), while the Amoebozoa appear to represent some of the oldest lineages, placing their ancestry around 1.8 bya. Many of the large clades that include marine phytoplankton appear to have undergone diversification at similar time horizons (1.5 and 1.3 bya). The tree appears to be well resolved except at deep branches, where substitutions in the 18S rRNA molecule may not offer sufficient resolution to separate very old lineages.

## Identification of Orthologs among All MMETSP Datasets

To make direct comparisons among the diverse groups contained within the MMETSP dataset, a set of orthologs was necessary. Orthologs allow the testing of conserved protein signatures that are retained among diverse organisms and thus changes in amino acids are more likely to be similar between species and divergent between species. In all, 8,291 homologous protein families were identified in the MMETSP dataset. Several different values of the inflation parameter were attempted with the MCL software to refine the ability to resolve differences among putative protein clusters by applying weight to edges (log transformed E-values) between hits. However, at greater inflation values, smaller differences between hits are given more weight, increasing the possibility that homologous gene families are split into more than one protein cluster. Conversely, at low inflation values, large families of non-homologous proteins are sometimes clustered. Given the conservative approach taken here, it is therefore likely that at least some large clusters may belong to several closely related families. It should be noted that some protein clusters became un-alignable after filtering by exceeding available computational capacity, and these were therefore excluded from further analysis. The results of the UBLASTP and MCL clustering resulted 8,291 clusters found among the reference proteomes and MMETSP *in silico* proteomes. After clustering, the putative gene families were restricted to contain at least 250 tips, 1,915 protein families were aligned and filtered. The second round of filtering and pruning reduced the number of protein families to 1,262, but only 1,258 could be aligned for the purpose of maximum likelihood tree construction. With respect to the aligned clusters, pruning by maximum inclusion and requiring 50 tips to be



present within each orthologous group resulted in 610 orthologous clusters among the 439 MMETSP and reference genomes species/strains (99%). The average number of orthologs recovered among all MMETSP samples was 112, with a minimum of 1 orthologous group, and a maximum of 350 (57.3%) orthologous groups found between species. The most orthologs were found primarily in the Bacillariophyta. Few species (23.1%) had greater than 132 orthologs. Due to difficulties in finding orthologs with high occupancy via the Yang and Smith methodology (Yang and Smith 2014) an alternative strategy was employed via the use of BUSCOs as a benchmark for transcriptome completeness, and using the output hits as putative markers for the presence of single copy orthologs among MMETSP taxa. The numbers of orthologs for each taxon is summarized in Supplemental Table 4.

### **The Stoichi transcriptome and Cellular Trophic Strategy**

MMETSP transcriptomes and proteomes were analyzed for enrichment of N atoms in the nucleotides or the amino acid side chains as a way to understand whether these traits in phytoplankton and protists are impacted by external nutrient conditions. No significant difference among median proteome values of NARSC were observed among autotrophic, mixotrophic or heterotrophic categories (Phy-ANOVA:  $df = 2$ ,  $F = 2.95$ ,  $P$  value = 0.7128), see Figure 2A. However, a significant signal was observed between the median proteome values of the CARSC values among the three categories (Phy-ANOVA:  $df = 2$ ,  $F = 38.01$ ,  $P$  value = 0.0249) with the most significant contrasts between autotrophic and heterotrophic ( $P$  value = 0.0108) and heterotrophic versus mixotrophic ( $P$  value = 0.0108) species (Figure 2B). Autotrophic versus mixotrophic were not found to be significantly different ( $P$  value = 0.9250) We found no significant

differences between N atoms among the transcriptome CDS, C atoms among CDS, GC%, or NAB3 of the CDS respectively (Phy-ANOVA:  $df = 2$ ,  $F = 14.96$ ,  $P$  value = 0.2122;  $df = 2$ ,  $F = 12.35$ ,  $P$  value = 0.2746;  $df = 2$ ,  $F = 16.50$ ,  $P$  value = 0.192;  $df = 2$ ,  $F = 18.85$ ,  $P$  value = 0.148). The average for each of the 20 amino acids was calculated and then grouped by trophic strategy. This analysis revealed that alanine was most frequently used across the data set, with slightly elevated usage in mixotrophs. Similarly, cysteine, glycine, histidine, leucine, proline, arginine, valine all appear at greater frequencies in mixotrophs as compared to autotrophs, while patterns in heterotrophs were mixed in comparison with autotrophs and mixotrophs (Supplemental Figure 1).

### **G+C Content and N and C-Content**

Calculated N and C content at the nucleotide and proteome level exhibit a strong phylogenetic signal as evidenced by Pagel's  $\lambda$  values close to 1 (Pagel's  $\lambda$  ranged 0.85 – 0.99) see Table 1 for comparison among all traits. The same trend was seen in both the global medians, global weighted averages, Yang homologs (medians and weighted averages), and BUSCO filtered (medians and weighted averages). All models favored the Brownian motion model over the  $\lambda = 0$  model (meaning no phylogenetic signal). Positive delta AICc values were calculated as the likelihood ratio tests with ( $\ln L_{BM} - \ln L_0$ ) model, and tested with the Chi-square distribution, setting degrees of freedom to one. Positive values indicate the model favoring estimated  $\lambda$  closer to 1, which represents the unaltered phylogenetic tree and negative delta AICc values indicate support for the  $\lambda=0$  model. Similar patterns of significant Pagel's  $\lambda$  were displayed using either Median transcriptome N-content or N-content weight averaged

by gene expression. Orthologous subsets, also displayed the same patterns as with values derived from Medians or weighted averages of all proteins (global analysis). We further only explored the data using the median values.

At the scale of the entire MMETPS dataset, there appear to be traits that follow the Brownian model with higher support than random phylogenetic independence using Moran's I autocorrelation index. Moran's I autocorrelation index is typically negative, and when positive indicates that species are more similar than they should be by chance. All traits (NARSC, CARSC, CDS, GC%, NatomCDS, CatomCDS, and NAB3) have positive Moran I values that are greater than the expected at short distances (Table 2), but then the values decrease in similarity with increasing phylogenetic distance (Figure 3). Interestingly, analysis of the Global set of all predicted proteins via Moran's I for NARSC was not significant, but every other variable was highly significant. With filtering of BUSCOs and Yang Homologs the NARSC values were then significant along with the other variables. Regardless of whether medians or weighted NARSC values by their respective TPM values, the results were similar in all but in two cases. BUSCO weighted NARSC and the Yang Homologs weighted NARSC, both were not significant. Other variables were highly significant for Moran's I, indicating correlations among closely related species  $P$  values  $< 0.0001$  (Summarized in Table 2).

It was also tested if cell size was correlated to NARSC and CARSC, and the GC content of the transcriptome. NARSC was strongly predicted by the cell size to both minimum and maximum cell equal diameter cells, showing a negative correlation Figure 4. This pattern was found using the larger collection of cell size data that included culture collection information, MMETSP metadata, and literature values (N =

258), and the same pattern was seen for the reduced data only using culture collection and MMETSP metadata (N = 179) summarized in Table 3. Among the larger dataset, the OU model was significantly different from the Brownian model, and a significant correlation of NARSC and minimum cell size ( $P < 0.0028$ ), maximum cell size ( $P < 0.0001$ ). Additionally our results show that NARSC of the proteome and high G+C content of the transcriptome have a high and strong positive correlation when phylogenetic signal is removed.

### **The Stoichi transcriptome vs. Isolation Habitat**

Nitrogen atoms in the residue side chain (NARSC) were analyzed against the environmental nutrient characteristics of the isolate's geographic origin to understand if the NARSC is correlated to the average abundance of inorganic N in that habitat. This may be a proxy for habitats limited in nitrogen (open-ocean) versus habitats rich in nitrogen (coastal-ocean), without having to designate distances or depths from coasts as a measure of nutrient influence. It was also tested if NARSC was correlated to the GC content of the transcriptome as a possible explanation to explain differences. Similarly, the Nitrate:Phosphate ratio was investigated. ANOVA model fitting was used to fit each variable (N, or P) using the Brownian Motion model, the Ornstein-Uhlenbeck (OU model) single additional rate parameter model, and Pagel's lambda transformation. PLGS regression indicated that the OU model supported both (Table 4), and a marginal correlation was found among NARSC proteome values and log transformed oceanic nitrate (df = 256,  $F = 112531.9$ ,  $P$  value = 0.1093) was present. NARSC was found to correlated with G+C content of the transcriptome (df = 256,  $F = 132521.58$ ,  $P$  value <0.0001). In this way, a significant correlation of NARSC to log transformed oceanic

phosphorus was also observed ( $df = 256$ ,  $F = 132521.58$ ,  $P$  value = 0.0452), but not the log Nitrate: Phosphate ratio ( $df = 256$ ,  $F = 67305.46$ ,  $P$  value = 0.2189). This suggests that the N content in the amino acid side chain has a correlation with oceanic phosphate. However, values nitrate and phosphate averages obtained from the Bio-Oracle raster layers were also highly correlated (Pearson correlation = 0.7).

## Discussion

### Overall Relationship of Stoichiometry to Trophic Strategy

The observed subcellular stoichiometry among the diverse collection of MMETSP transcriptomes, suggests that there are significant differences among trophic strategies with respect to proteomic C. In addition, only the estimated C enrichment per residue side-chain (CARSC) of the proteome but not transcriptome differed significantly among autotrophic, mixotrophic, and heterotrophic strains. Measures of N enrichment per residue side-chain (NARSC) and %G+C content were not significantly different from one another among these groups. This is despite estimates of Pagel's  $\lambda$ , which suggests that NARSC does contain a phylogenetic signal that appears similar to random genetic drift under the Brownian Motion model of trait evolution but is not related to trophic strategy. NARSC appears to be autocorrelated among more closely related species than those distantly related. The possibility that quantitative traits do not match the Brownian model expectation was also tested using the Ornstein-Uhlenbeck model that adds an  $\alpha$  parameter that represents the pull of the trait toward some optima (Butler and King 2004). Often, the Ornstein-Uhlenbeck model is suggested to represent the movement of a random Brownian trait back toward an optimal value, such that it may represent an adaptive peak. The Ornstein-Uhlenbeck model can also be thought of as a Brownian Motion model with several lineages moving at different rates. It was observed that the Ornstein-Uhlenbeck model had higher statistical support than the purely Brownian Model based on phylogenetic least squares regression. These results suggest that N stoichiometry of the transcriptome/proteome is a random genetic drift process, but may vary within lineages at different rates. It further suggests that it is

evolving in response to some selective pressures. Autocorrelation tests with the Moran's I suggests that N enriched amino acids are similar over short periods of time, meaning species that are more closely related share similar N enrichment scores. This effect was greater at the coding sequence level, where values are high (similar) and become dissimilar with increasing phylogenetic distance. Surprisingly, at great phylogenetic distances, the traits were more similar among species but were not significant except for the N content of the coding sequences. This correlation in deep time may be evidence of differences among the algae groups that acquired either the red and green plastid via endosymbiosis. These difference may be related to the elemental frequency of C, N, P, and trace metals in the biomass of 14 phytoplankton species which Quigg et al. (2010) found grouped by evolutionary inheritance of either the red or green plastid. They found specifically that C:N ratio of Viridiplantae, Chlorophyta, Chlorarachniophyta, and Euglenozoa were higher than other groups, and the N:P ratios were highest in the secondary endosymbiosis of green plastid lineages versus red plastid lineages (Cryptophyta, Haptophyta, Heterokonts, Bacillariophyta, Chrysophyta and Dinophyta). While the C:P chemistry was highest in the Viridiplantae and Chlorophyta. Perhaps the C enrichment of mixotrophs seen in this study of the transcriptomes and proteomes may be the product of similar patterns in C and N elemental composition between the red and green lineages. It is unclear what the stoichiometric pattern of mixotrophs would be in the same analysis, and the kleptoplastidic dinoflagellate in Quigg et al. (2010) study did not group with other red algae plastids that typify dinoflagellates, but grouped near the green algae plastid that it steals.

Mixotrophy is known to have evolved many times, encompassing several different kinds of nutrient acquisition, and is not restricted to any particular microeukaryote phyla (Jones 2000). Mixotrophy can be defined as organisms that strictly obtain dissolved organic compounds (DON, DOP) from the environment, or organisms that phagocytize particles, or organisms that can accomplish both forms (Stoecker 1998, Jones 2000, Burkholder et al. 2008, Hansen 2011, Ward et al. 2011, Glibert et al. 2012, Barton et al. 2013, Wilken et al. 2014). The ecological trade off of acquiring external nutrients and maintaining the photosynthetic apparatus may relate to slower growth rates, low affinity for inorganic nutrients, and slower grazing rates than heterotrophs seen among mixotrophs (Barton et al. 2013). However, these traits are advantageous when inorganic nutrients are low, competition may be reduced among species that depend on higher nutrient concentrations and may be able to maximize their growth under nutrient limiting or light limited conditions. How these traits manifest within the elemental stoichiometry of mixotrophs are largely unknown, and are an active area of research, because they appear to be critical components of C and N cycling in oceanic systems (Hammer and Pitchford 2005).

Perhaps the differences among the red and green plastid clades observed by Quigg et al. (2010) may explain the observed elevated C enriched amino acids in the mixotrophs (Figure 2B), but not why there was not a signal of N content (Figure 2A). This suggests that groups do not differ significantly in their N content, rejecting our first hypothesis. Yet, heterotrophic taxa had a higher enrichment of N in amino acids, and autotrophic and mixotrophic did have lower, matching similar findings of Acquisti et al. (2009a). The enrichment of C atoms in the amino-acid side chains by mixotrophs can be



interpreted in several ways. For example, it might represent a cellular adaptation to unknown selective forces that results in higher C:N ratios. Conversely, selective pressures in heterotrophs and autotrophs might result in lower C:N ratios in these groups. The median C-content of the mixotrophic proteomes was higher than the other trophic strategies suggesting mixotrophs have a relaxed selection against incorporating C into amino acids. Another alternative, are differences in codon or amino acid usage. Codon usage among Bacteria and Archaea have been linked to associations to niche breadth, pathogenic or non-pathogenic, and cellular fitness (Botzman and Margalit 2011; Grzymiski and Marsh 2014). These studies suggest at least in Bacteria and Archaea that codon usage may be related to their lifestyles as free-living or pathogenic. In one study they found that free-living microorganisms had more repetitive usage of codons than in pathogen species (Grzymiski and Marsh 2014). In addition that codon and amino acid usage are akin to similar dialects of language among microorganisms. Within eukaryotes, codon bias might also be a function of translation efficiency and protein abundance (Hudson et al. 2011). Additionally bias may arise due to translation efficiency favoring codons that correspond to abundant tRNA molecules favoring selective usage (Sharp et al. 1986, Stenico et al. 1994, Moriyama et al. 1997, Musto et al. 2001, Wille and Majewski, 2004). Differences in codons may also arise via non-adaptive processes as summarized by Lynch and Conery (2003) or could be the result of selective or neutral processes (Duret 2002). It's thought that large effective population sizes of microorganisms may prevent the evolution of genomic changes because rare allele variants are uncommon. However, many studies of phytoplankton taxa show evidence of reduced population connectivity, in which small populations could

accumulate genetic differences (Rynearson and Armbrust 2004, Casteleyn et al. 2010, Koester et al. 2010, Degerlund et al. 2012, Whittaker et al. 2012, Sjöqvist et al. 2015, Whittaker et al. 2017). Further studies into microeukaryotes may uncover differences associated with codon utilization and may explain the observed differences among the diverse species in the MMETSP dataset.

To examine whether amino acid usage differed with respect to trophic behavior, we examined the average amino acid content of each assembly and grouped by trophic strategy (Supplemental Figure 1). These data did not suggest any obvious trend in this regard. Most of the heterotrophic organisms in the MMETSP are Ciliates (N=25) and Dinoflagellates (N=8) that are currently not known to utilize alternative genetic codons for translation. It has been noted that C content of amino acid side chains in bacteria varied the most, and that N among bacterial proteomes was largely similar (Baudoin-Cornu 2004). The same study suggests large differences among proteomes appear to be correlated with G+C content of the genome even across Domains. The carbon content of the proteome was negatively correlated with high G+C content, suggesting high G+C genomes have low protein carbon. In addition, it is unknown how *de novo in silico* prediction software could influence the codon pattern among samples. Transdecoder uses hexamer frequencies among the transcripts to identify the most likely open reading frame, but itself could be biased in its outputs (Haas et al. 2013).

### **Ecological Factors Related to NARSC**

Previous studies have noted the mismatch or weak signal between ecological parameters of species and their elemental stoichiometry within multicellular organisms (Francois et al. 2016). However, others have found a correlation between the

environmental conditions and the stoichiometry, such as (Grzymski and Dussaq 2012) who found the average number of N atoms per residue side chain was reduced in open-ocean compared to coastal-ocean microorganisms. The same study reports that N thrift among predicted proteins in metagenome data has a nonlinear relationship with distance from the coast. Nutrient enrichment and variability near the coastal margins are not always enriched due to anthropogenic sources, because coastal upwelling, and natural terrigenous inputs can provide key limiting nutrients promoting phytoplankton growth (Sweeney and Kaplan 1980, Walsh 1991, Galloway et al. 1996, Nixon et al. 1996). Perhaps explaining why the pattern is not always strongly associated with a single chemical component, necessary for growth.

G+C content was positively correlated with measures of N content of the amino acid side chain (NARSC). We expected to find a correlation between the %GC content and the NARSC as shown in a previous study (Elser et al. 2011). Correlations of environmental nitrate and phosphate to NARSC were both negative and significant, and revealed an unexpected correlation to phosphorus (Table 4). The correlation of NARSC to phosphorus may be an effect of the fact that nitrate and phosphate concentrations covary in natural systems (Pearson correlation = 0.7) (Downing, 1997; Tyrrell, 1999). The fact that the N:P ratio was not significantly correlated to NARSC, which is often referenced as a measurement of N or P limitation, would suggest, that higher NARSC values (enriched) would be found in more nutrient poor waters. Taken together with the cell size data, this supports the hypothesis that NARSC of the proteome have an environmental correlation albeit negative one. Larger cells, are favored in coastal, nutrient rich regions that allow potentially fast growth rates, while small-celled

phytoplankton are typical of open-ocean nutrient poor waters where their small size and high surface to area-to-volume ratio favor in reducing physiological costs, high nutrient absorption and a survivalist strategy (Arrigo, 2005). These data suggest that NARSC provides a measurement similar to physiological predictions of Arrigo (2005) who divided plankton into “survivalist”, “bloomer” and the “generalist”. The survivalists have the highest N:P ratios, the bloomer the lowest, and the generalist have a near balanced Redfield Ratio. The same study proposes that this may be due to physiological differences in how plankton allocates cellular resources (i.e. growth machinery, enzymes, and pigment/proteins), which have different C:N:P ratios.

A limitation of correlating microbes to large-scale environmental data layers is the quality of metadata. Much of the public metadata was incomplete in the MMETSP project with respect to isolation location, depth, and how long strains have been maintained in culture. Depth, for example, can be an important factor as is illustrated by the light-dependent distribution of different ecotypes of *Prochlorococcus* and *Ostreococcus* (Chisholm et al. 1988, Moore and Chisholm 1999, Rocap et al. 2003, García-Fernández et al. 2004, Rodríguez et al. 2005, Martiny et al. 2009). We also note that the Bio-Oracle data layers used here are only interpolated at the surface representing 0-10 meters. Yet, many of the Prasinophytes from the RCC culture collection, for example, were collected at various depths in the Indian Ocean, where they likely experienced a range of nutrient conditions due to the depth dependent abundance of N and P in the water column. How quickly the stoichiometry of cultivated phytoplankton responds to cultivation conditions is unknown. Evidence from domesticated plants suggests removal of the selective pressure of N-limiting conditions,

thereby allowing for an enrichment of amino acids in comparison to non-crop plants (Acquisti et al. 2009a).

### **Potential Strategies to Conserve Nitrogen**

Data presented here does not support the notion that N-limitation plays a role in shaping the proteomes or transcriptome of marine protists and phytoplankton, but did indicate that the N-content of proteome appear to correlate with environmental nutrient concentrations. Nitrogen, carbon and G+C content among the MMETSP species were found to be phylogenetically conserved among similar species. Our study did find that a significant negative correlation of the nutrient environment and N content of the isolates in the MMETSP dataset. We hypothesized that autotrophic organisms should be depleted in N-containing bases and amino acids relative to heterotrophs, and mixotrophs. We found that autotrophs and mixotrophs had lower N-stoichiometry than heterotrophs but did not differ significantly from autotrophs. The finding of significant differences in C stoichiometry among the MMETSP trophic strategies may indicate differences in codon usage, and thus fundamental differences in phytoplankton physiology. Previous studies have also found that phytoplankton members differ in the macromolecular composition and abundance of proteins, lipids, carbohydrates, ash, RNA, chlorophyll-a and DNA (Finkel et al. 2016). They found strong support of statistical differences among C and N that seem to correspond to previous records. The potential mismatch between our results and Finkel et al (2016) may be rooted in the balance of rRNA and protein synthesis, which might impact the ratio of N:P of the major cellular constituents growing in balance at maximal growth rates (Loladze and Elser 2011).

## **Relation to Stoichiogenomics and Future Directions**

This study rejects the hypothesis that variability exists in N enrichment of amino acid side-chain residues. This contradicts the foundational papers in stoichiogenomics (Acquisti et al. 2009a, Acquisti et al. 2009b, Elser et al. 2011) and suggests that the evolutionary forces driving stoichiogenomics are not yet sufficiently understood. An investigation of subterranean and surface-dwelling isopod species found no evidence that environmental N availability corresponded to the elemental composition of the proteome or transcriptome (Francois et al. 2016). Among species pairs they investigated, there were no significant effects of habitat or N availability in prey. We also did not find evidence that N-limitation shapes the evolution of phytoplankton and protist proteomes, but that they did differ in the C content akin to the findings of (Baudouin-Cornu et al. 2004) who studied genomes among the three domains of life. How the C content of the amino acid side chains compares among isopods to this study of C differences among phytoplankton is unknown (Francois et al. 2016). This contradicts the predictions of the stoichiogenomic framework, that biases may arise in amino acids and nucleic acids because they differ in number of key macroelements (C, N, S, P, O, H) and that through selection due to resource limitation should reduce or deplete its' abundance in the organism to maximize fitness. Evolution of proteins could instead be due to non-adaptive mechanisms such as found Günther et al. (2013), where genetic drift was the primary driver and not nutrient limitation. Future studies should also seek to identify physiological differences among phytoplankton in terms of their proportion of growth machinery (i.e. ribosomes), and the proportion of resource acquisition machinery. This may be achieved by annotation of common or

phylogenetically conserved metabolic pathways, but will require additional sequenced genomes to aid in protein finding. Also because transcriptomes represent all possible transcripts under a specific set of growth conditions, it may not be truly representative of genomic potential. Although we see stoichiogenomics as an additional way to look at the biology of marine protists in order to dissect their complex biology, we are cautious in our interpretations of such a young field of study with few experimental manipulations to rule out other causes of amino acid enrichment. Experimental manipulation and side-by-side comparison to elemental stoichiometry of phytoplankton may uncover the reasoning behind the observed stoichiometric differences.

## **Acknowledgements**

Funding for this project was in part provided by the Gordon and Betty Moore Foundation as part of the Marine Micro-Eukaryote Transcriptome Project (MMETSP), and in part via grants from the National Science Foundation (OCE 0961900 & OCE 1634630). We would also like to thank Henry Neeman at the OU Supercomputing Center for Education & Research for informatics support.



## References

- Acharya, K., M. Kyle, and J. J. Elser. 2004. Biological stoichiometry of *Daphnia* growth: An ecophysiological test of the growth rate hypothesis. *Limnology and Oceanography* **49**:656-665.
- Acquisti, C., J. J. Elser, and S. Kumar. 2009a. Ecological Nitrogen Limitation Shapes the DNA Composition of Plant Genomes. *Molecular Biology and Evolution* **26**:953-956.
- Acquisti, C., S. Kumar, and J. J. Elser. 2009b. Signatures of nitrogen limitation in the elemental composition of the proteins involved in the metabolic apparatus. *Proceedings of the Royal Society B: Biological Sciences* **276**:2605-2610.
- Arrigo, K. R. 2005. Marine microorganisms and global nutrient cycles. *Nature* **437**:349-355.
- Badet, T., R. Peyraud, M. Mbengue, O. Navaud, M. Derbyshire, R. P. Oliver, A. Barbacci, S. Raffaele, G. P. Zhao, C. Wang, M. G. Feng, C. J. Ridout, P. Schulze-Lefert, N. J. Talbot, N. Ahmadinejad, C. Ametz, G. R. Barton, M. Benjdia, P. Bidzinski, L. V. Bindschedler, M. Both, M. T. Brewer, L. Cadle-Davidson, M. M. Cadle-Davidson, J. Collemare, R. Cramer, O. Frenkel, D. Godfrey, J. Harriman, C. Hoede, B. C. King, S. Klages, J. Kleemann, D. Knoll, P. S. Koti, J. Kreplak, F. J. López-Ruiz, X. Lu, T. Maekawa, S. Mahanil, C. Micali, M. G. Milgroom, G. Montana, S. Noir, R. J. O'Connell, S. Oberhaensli, F. Parlange, C. Pedersen, H. Quesneville, R. Reinhardt, M. Rott, S. Sacristán, S. M. Schmidt, M. Schön, P. Skamnioti, H. Sommer, A. Stephens, H. Takahara, H. Thordal-Christensen, M. Vigouroux, R. Wessling, T. Wicker, R. Panstruga, P. Schulze-Lefert, E. V. L. v. Themaat, L. J. Ma, L. J. Vaillancourt, O. Yarden, Q. Zeng, J. A. Rollins, M. H. Lebrun, and M. Dickman. 2017. Codon optimization underpins generalist parasitism in fungi. *eLife* **6**:5159-5176.
- Balzano, S., D. Sarno, and W. H. C. F. Kooistra. 2011. Effects of salinity on the growth rate and morphology of ten *Skeletonema* strains. *Journal of Plankton Research* **33**:937-945.
- Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **19**:455-477.
- Barton, A. D., A. J. Pershing, E. Litchman, N. R. Record, K. F. Edwards, Z. V. Finkel, T. Kiørboe, and B. A. Ward. 2013. The biogeography of marine plankton traits. *Ecology Letters* **16**:522-534.

- Batut, B., C. Knibbe, G. Marais, and V. Daubin. 2014. Reductive genome evolution at both ends of the bacterial population size spectrum. *Nature Reviews Microbiology* **12**:841-850.
- Baudouin-Cornu, P., K. Schuerer, P. Marlière, and D. Thomas. 2004. Intimate evolution of proteins. Proteome atomic content correlates with genome base composition. *The Journal of Biological Chemistry* **279**:5421-5428.
- Baudouin-Cornu, P., Y. Surdin-Kerjan, P. Marlière, and D. Thomas. 2001. Molecular evolution of protein atomic composition. *Science* **293**:297-300.
- Boersma, M. 2000. The nutritional quality of P-limited algae for *Daphnia*. *Limnology and Oceanography* **45**:1157-1161.
- Botzman, M., and H. Margalit. 2011. Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome Biology* **12**:R109.
- Bragg, J. G. 2011. How *Prochlorococcus* bacteria use nitrogen sparingly in their proteins. *Molecular Ecology* **20**:27-28.
- Bragg, J. G., A. Quigg, J. A. Raven, and A. Wagner. 2012. Protein elemental sparing and codon usage bias are correlated among bacteria. *Molecular Ecology* **21**:2480-2487.
- Bragg, J. G., and A. Wagner. 2009. Protein material costs: single atoms can make an evolutionary difference. *Trends in Genetics* **25**:5-8.
- Brown, J. W., and U. Sorhannus. 2010. A Molecular Genetic Timescale for the Diversification of Autotrophic Stramenopiles (Ochrophyta): Substantive Underestimation of Putative Fossil Ages. *PLoS ONE* **5**:e12759.
- Bullejos, F. J., P. Carrillo, E. Gorokhova, J. M. Medina-Sánchez, and M. Villar-Argaiz. 2014. Nucleic acid content in crustacean zooplankton: bridging metabolic and stoichiometric predictions. *PLoS ONE* **9**:e86493.
- Burkholder, J. M., P. M. Glibert, and H. M. Skelton. 2008. Mixotrophy, a major mode of nutrition for harmful algal species in eutrophic waters. *Harmful Algae* **8**:77-93.
- Bushnell, B. 2016. BBmap: BBMap short read aligner, and other bioinformatic tools., <https://sourceforge.net/projects/bbmap/>.
- Butler, Marguerite A., and Aaron A. King. 2004. Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. *The American Naturalist* **164**:683-695.

- Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972-1973.
- Casteleyn, G., F. Leliaert, T. Backeljau, A.-E. Debeer, Y. Kotaki, L. Rhodes, N. Lundholm, K. Sabbe, and W. Vyverman. 2010. Limits to gene flow in a cosmopolitan marine planktonic diatom. *Proceedings of the National Academy of Sciences* 107:12952-12957.
- Chisholm, S. W., R. J. Olson, E. R. Zettler, R. Goericke, J. B. Waterbury, and N. A. Welschmeyer. 1988. A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* 334:340-343.
- Collins, S., B. Rost, and T. A. Ryneerson. 2014. Evolutionary potential of marine phytoplankton under ocean acidification. *Evolutionary Applications* 7:140-155.
- Curtis, B. A., G. Tanifuji, F. Burki, A. Gruber, M. Irimia, S. Maruyama, M. C. Arias, S. G. Ball, G. H. Gile, Y. Hirakawa, J. F. Hopkins, A. Kuo, S. A. Rensing, J. Schmutz, A. Symeonidi, M. Elias, R. J. M. Eveleigh, E. K. Herman, M. J. Klute, T. Nakayama, M. Obornik, A. Reyes-Prieto, E. V. Armbrust, S. J. Aves, R. G. Beiko, P. Coutinho, J. B. Dacks, D. G. Durnford, N. M. Fast, B. R. Green, C. J. Grisdale, F. Hempel, B. Henrissat, M. P. Hoppner, K.-I. Ishida, E. Kim, L. Koreny, P. G. Kroth, Y. Liu, S.-B. Malik, U. G. Maier, D. McRose, T. Mock, J. A. D. Neilson, N. T. Onodera, A. M. Poole, E. J. Pritham, T. A. Richards, G. Rocard, S. W. Roy, C. Sarai, S. Schaack, S. Shirato, C. H. Slamovits, D. F. Spencer, S. Suzuki, A. Z. Worden, S. Zauner, K. Barry, C. Bell, A. K. Bharti, J. A. Crow, J. Grimwood, R. Kramer, E. Lindquist, S. Lucas, A. Salamov, G. I. McFadden, C. E. Lane, P. J. Keeling, M. W. Gray, I. V. Grigoriev, and J. M. Archibald. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492:59-65.
- Daines, S. J., J. R. Clark, and T. M. Lenton. 2014. Multiple environmental controls on phytoplankton growth strategies determine adaptive responses of the N:P ratio. *Ecology Letters* 17:414-425.
- Damsté, J. S. S., G. Muyzer, B. Abbas, S. W. Rampen, G. Massé, W. G. Allard, S. T. Belt, J.-M. Robert, S. J. Rowland, J. M. Moldowan, S. M. Barbanti, F. J. Fago, P. Denisevich, J. Dahl, L. A. F. Trindade, and S. Schouten. 2004. The Rise of the Rhizosolenid Diatoms. *Science* 304:584-587.
- Degerlund, M., S. Huseby, A. Zingone, D. Sarno, and B. Landfald. 2012. Functional diversity in cryptic species of *Chaetoceros socialis* Lauder (Bacillariophyceae). *Journal of Plankton Research* 34:416-431.
- Derelle, E., C. Ferraz, S. Rombauts, P. Rouzé, A. Z. Worden, S. Robbens, F. Partensky, S. Degroeve, S. Echeynié, R. Cooke, Y. Saeys, J. Wuyts, K. Jabbari, C. Bowler,

- O. Panaud, B. Piégu, S. G. Ball, J.-P. Ral, F.-Y. Bouget, G. Piganeau, B. De Baets, A. Picard, M. Delseny, J. Demaille, Y. Van de Peer, and H. Moreau. 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences of the United States of America* **103**:11647-11652.
- Downing, J. A. 1997. Marine nitrogen: Phosphorus stoichiometry and the global N:P cycle. *Biogeochemistry* **37**:237-252.
- Duret, L. 2002. Evolution of synonymous codon usage in metazoans. *Current Opinion in Genetics & Development* **12**:640-649.
- Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**:2460-2461.
- Elser, J. J., K. Acharya, M. Kyle, J. B. Cotner, W. Makino, T. Markow, T. Watts, S. E. Hobbie, W. F. Fagan, J. D. Schade, J. M. Hood, and R. W. Sterner. 2003. Growth rate–stoichiometry couplings in diverse biota. *Ecology Letters* **6**:936-943.
- Elser, J. J., C. Acquisti, and S. Kumar. 2011. Stoichiogenomics: the evolutionary ecology of macromolecular elemental composition. *Trends in Ecology & Evolution* **26**:38-44.
- Elser, J. J., R. W. Sterner, E. Gorokhova, W. F. Fagan, T. A. Markow, J. B. Cotner, J. F. Harrison, S. E. Hobbie, G. M. Odell, and L. J. Weider. 2008. Biological stoichiometry from genes to ecosystems. *Ecology Letters* **3**:540-550.
- Enright, A. J., S. Van Dongen, and C. A. Ouzounis. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* **30**:1575-1584.
- Falkowski, P. G. 1997. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO<sub>2</sub> in the ocean. *Nature* **387**:272-275.
- Falkowski, P. G., M. E. Katz, A. H. Knoll, A. Quigg, J. A. Raven, O. Schofield, and F. J. R. Taylor. 2004. The Evolution of Modern Eukaryotic Phytoplankton. *Science Mag* **305**:354-360.
- Falkowski, P. G., and M. J. Oliver. 2007. Mix and match: how climate selects phytoplankton. *Nature Reviews Microbiology* **5**:813-819.
- Finkel, Z. V., J. Beardall, K. J. Flynn, A. Quigg, T. A. V. Rees, and J. A. Raven. 2009. Phytoplankton in a changing world: cell size and elemental stoichiometry. *Journal of Plankton Research* **32**:119-137.

- Finkel, Z. V., M. J. Follows, J. D. Liefer, C. M. Brown, I. Benner, and A. J. Irwin. 2016. Phylogenetic diversity in the macromolecular composition of microalgae. *PLoS ONE* **11**:e0155977.
- Flynn, K. J., J. A. Raven, T. A. V. Rees, Z. V. Finkel, A. Quigg, and J. Beardall. 2010. Is the growth rate hypothesis applicable to microalgae? *Journal of Phycology* **46**:1-12.
- Francois, C. M., L. Duret, L. Simon, F. Mermillod-Blondin, F. Malard, L. Konecny-Dupré, R. Planel, S. Penel, C. J. Douady, and T. Lefébure. 2016. No Evidence That Nitrogen Limitation Influences the Elemental Composition of Isopod Transcriptomes and Proteomes. *Molecular Biology and Evolution* **33**:2605-2620.
- Galloway, J. N., R. W. Howarth, A. F. Michaels, S. W. Nixon, J. M. Prospero, and F. J. Dentener. 1996. Nitrogen and phosphorus budgets of the North Atlantic Ocean and its watershed. *Biogeochemistry* **35**:3-25.
- García-Fernández, J. M., N. T. de Marsac, and J. Diez. 2004. Streamlined regulation and gene loss as adaptive mechanisms in *Prochlorococcus* for optimized nitrogen utilization in oligotrophic environments. *Microbiology and Molecular Biology Reviews* **68**:630–638.
- García-Fernández, J. M., and J. Diez. 2004. Adaptive mechanisms of nitrogen and carbon assimilatory pathways in the marine cyanobacteria *Prochlorococcus*. *Research in microbiology* **155**:795-802.
- Geider, R. J., and J. L. Roche. 2011. Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. *European Journal of Phycology*.
- Gilbert, J. D. J., and W. F. Fagan. 2011. Contrasting mechanisms of proteomic nitrogen thrift in *Prochlorococcus*. *Molecular Ecology* **20**:92-104.
- Giordano, M., M. Palmucci, and J. A. Raven. 2015. Growth rate hypothesis and efficiency of protein synthesis under different sulphate concentrations in two green algae. *Plant, Cell and Environment* **38**:2313-2317.
- Glibert, P. M., J. M. Burkholder, and T. M. Kana. 2012. Recent insights about relationships between nutrient availability, forms, and stoichiometry, and the distribution, ecophysiology, and food web effects of pelagic and benthic *Prorocentrum* species. *Harmful Algae* **14**:231-259.
- Gouy, M., S. Guindon, and O. Gascuel. 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution* **27**:221-224.

- Grant, J. R., and L. A. Katz. 2014. Building a phylogenomic pipeline for the eukaryotic tree of life—addressing deep phylogenies with genome-scale data. *PLOS Currents Tree of Life*.
- Gruber-Vodicka, H., E. A. Pruesse, and B. Seah. 2016. phyloFlash is a pipeline to rapidly reconstruct the SSU rRNAs and explore phylogenetic composition of an illumina (meta)genomic dataset. <https://github.com/HRGV/phyloFlash>.
- Grzymalski, J. J., and A. M. Dussaq. 2012. The significance of nitrogen cost minimization in proteomes of marine microorganisms. *The ISME Journal* **6**:71-80.
- Grzymalski, J. J., and A. G. Marsh. 2014. Protein languages differ depending on microorganism lifestyle. *PLoS ONE* **9**:e96910.
- Günther, T., C. Lampei, and K. J. Schmid. 2013. Mutational Bias and Gene Conversion Affect the Intraspecific Nitrogen Stoichiometry of the *Arabidopsis thaliana* Transcriptome. *Molecular Biology and Evolution* **30**:561-568.
- Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. Leduc, N. Friedman, and A. Regev. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**:1494-1512.
- Hammer, A. C., and J. W. Pitchford. 2005. The role of mixotrophy in plankton bloom dynamics, and the consequences for productivity. *ICES Journal of Marine Science* **62**:833-840.
- Hansen, P. J. 2011. The Role of Photosynthesis and Food Uptake for the Growth of Marine Mixotrophic Dinoflagellates. *Journal of Eukaryotic Microbiology* **58**:203-214.
- Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**:129-131.
- Hedges, S. B., J. Marin, M. Suleski, M. Paymer, and S. Kumar. 2015. Tree of Life Reveals Clock-Like Speciation and Diversification. *Molecular Biology and Evolution* **32**:835-845.
- Herron, M. D., J. D. Hackett, F. O. Aylward, R. E. Michod, and F. J. Ayala. 2009. Triassic Origin and Early Radiation of Multicellular Volvocine Algae. *Proceedings of the National Academy of Sciences of the United States of America* **106**:3254-3258.

- Hessen, D. O., P. D. Jeyasingh, M. Neiman, and L. J. Weider. 2010. Genome streamlining and the elemental costs of growth. *Trends in Ecology & Evolution* **25**:75-80.
- Hillebrand, H., G. Steinert, M. Boersma, A. Malzahn, C. Léo Meunier, C. Plum, and R. Ptacnik. 2013. Goldman Revisited: Faster growing phytoplankton has lower N:P and lower stoichiometric flexibility. *Limnology and Oceanography* **58**:2076-2088.
- Howarth, R. W., and R. Marino. 2006. Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: Evolving views over three decades. *Limnology and Oceanography* **51**:364-376.
- Huertas, I. E., M. Rouco, V. López-Rodas, E. Costas, V. Lopez-Rodas, and E. Costas. 2011. Warming will affect phytoplankton differently: evidence through a mechanistic approach. *Proceedings of the Royal Society of London B: Biological Sciences* **278**:3534-3543.
- Hudson, N. J., Q. Gu, S. H. Nagaraj, Y.-S. Ding, B. P. Dalrymple, and A. Reverter. 2011. Eukaryotic Evolutionary Transitions Are Associated with Extreme Codon Bias in Functionally-Related Proteins. *PLoS ONE* **6**:e25457.
- Johnson, Z. I., E. R. Zinser, A. Coe, N. P. McNulty, E. M. S. Woodward, and S. W. Chisholm. 2006. Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science (New York, N.Y.)* **311**:1737-1740.
- Jones, R. I. 2000. Mixotrophy in planktonic protists: an overview. *Freshwater Biology* **45**:219-226.
- Katoh, K., and D. M. Standley. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**:772-780.
- Katz, L. A., and J. R. Grant. 2015. Taxon-Rich Phylogenomic Analyses Resolve the Eukaryotic Tree of Life and Reveal the Power of Subsampling by Sites. *Systematic Biology* **64**:406-415.
- Keeling, P. J., F. Burki, H. M. Wilcox, B. Allam, E. E. Allen, L. A. Amaral-Zettler, E. V. Armbrust, J. M. Archibald, A. K. Bharti, C. J. Bell, B. Beszteri, K. D. Bidle, C. T. Cameron, L. Campbell, D. A. Caron, R. A. Cattolico, J. L. Collier, K. Coyne, S. K. Davy, P. Deschamps, S. T. Dyhrman, B. Edvardsen, R. D. Gates, C. J. Gobler, S. J. Greenwood, S. M. Guida, J. L. Jacobi, K. S. Jakobsen, E. R. James, B. Jenkins, U. John, M. D. Johnson, A. R. Juhl, A. Kamp, L. A. Katz, R. Kiene, A. Kudryavtsev, B. S. Leander, S. Lin, C. Lovejoy, D. Lynn, A. Marchetti, G. McManus, A. M. Nedelcu, S. Menden-Deuer, C. Miceli, T. Mock, M. Montresor, M. A. Moran, S. Murray, G. Nadathur, S. Nagai, P. B. Ngam, B.

- Palenik, J. Pawlowski, G. Petroni, G. Piganeau, M. C. Posewitz, K. Rengefors, G. Romano, M. E. Rumpho, T. Ryneerson, K. B. Schilling, D. C. Schroeder, A. G. B. Simpson, C. H. Slamovits, D. R. Smith, G. J. Smith, S. R. Smith, H. M. Sosik, P. Stief, E. Theriot, S. N. Twary, P. E. Umale, D. Vaultot, B. Wawrik, G. L. Wheeler, W. H. Wilson, Y. Xu, A. Zingone, and A. Z. Worden. 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**:e1001889.
- Koester, J. A., J. E. Swalwell, P. von Dassow, and E. V. Armbrust. 2010. Genome size differentiates co-occurring populations of the planktonic diatom *Ditylum brightwellii* (Bacillariophyta). *BMC Evolutionary Biology* **10**:1.
- Lahr, D. J. G., J. Grant, R. Molestina, L. A. Katz, and O. R. Anderson. 2015. *Sapocribrum chincoteaguense* n. gen. n. sp.: A Small, Scale-bearing Amoebozoan with Flabellinid Affinities. *Journal of Eukaryotic Microbiology* **62**:444-453.
- Lahr, D. J. G., J. Grant, T. Nguyen, J. H. Lin, and L. A. Katz. 2011. Comprehensive Phylogenetic Reconstruction of Amoebozoa Based on Concatenated Analyses of SSU-rDNA and Actin Genes. *PLoS ONE* **6**:e22780.
- Lakeman, M. B., P. von Dassow, and R. A. Cattolico. 2009. The strain concept in phytoplankton ecology. *Harmful Algae* **8**:746-758.
- Leonardos, N., and R. J. Geider. 2004. Responses of elemental and biochemical composition of *Chaetoceros muelleri* to growth under varying light and nitrate:phosphate supply ratios and their influence on critical N:P. *Limnology and Oceanography* **49**:2105-2114.
- Li, W., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**:1658-1659.
- Lind, P. R., and P. D. Jeyasingh. 2015. Genotypic differences in phosphorus use physiology in producers (*Chlamydomonas reinhardtii*) and consumers (*Daphnia pulex*) interact to alter primary and secondary production. *Evolutionary Ecology* **29**:551-563.
- Litchman, E., and C. a. Klausmeier. 2008. Trait-Based Community Ecology of Phytoplankton. *Annual Review of Ecology, Evolution, and Systematics* **39**:615-639.
- Liu, H., S. Aris-Brosou, I. Probert, and C. de Vargas. 2010. A Time line of the Environmental Genetics of the Haptophytes. *Molecular Biology and Evolution* **27**:161-176.



- Loladze, I., and J. J. Elser. 2011. The origins of the Redfield nitrogen-to-phosphorus ratio are in a homeostatic protein-to-rRNA ratio. *Ecology Letters* **14**:244-250.
- Luo, H., L. R. Thompson, U. Stingl, and A. L. Hughes. 2015. Selection Maintains Low Genomic GC Content in Marine SAR11 Lineages. *Molecular Biology and Evolution* **32**: 2738-2748.
- Lv, J., N. Li, and D.-K. Niu. 2008. Association between the availability of environmental resources and the atomic composition of organismal proteomes: Evidence from *Prochlorococcus* strains living at different depths. *Biochemical and Biophysical Research Communications* **375**:241-246.
- Lynch, M., and J. S. Conery. 2003. The Origins of Genome Complexity. *Science* **302**:1401-1404.
- Martiny, A. C., M. L. Coleman, and S. W. Chisholm. 2006. Phosphate acquisition genes in *Prochlorococcus* ecotypes: Evidence for genome-wide adaptation. *Proceedings of the National Academy of Sciences* **103**:12552-12557.
- Martiny, A. C., Y. Huang, and W. Li. 2009a. Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environmental Microbiology* **11**:1340-1347.
- Martiny, A. C., S. Kathuria, and P. M. Berube. 2009b. Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proceedings of the National Academy of Sciences of the United States of America* **106**:10787-10792.
- Martiny, A. C., C. T. A. Pham, F. W. Primeau, J. A. Vrugt, J. K. Moore, S. A. Levin, and M. W. Lomas. 2013a. Strong latitudinal patterns in the elemental ratios of marine plankton and organic matter. *Nature Geoscience* **6**:279-283.
- Martiny, A. C., A. P. K. Tai, D. Veneziano, F. Primeau, and S. W. Chisholm. 2009c. Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environmental Microbiology* **11**:823-832.
- Martiny, A. C., J. A. Vrugt, F. W. Primeau, and M. W. Lomas. 2013b. Regional variation in the particulate organic carbon to nitrogen ratio in the surface ocean. *Global Biogeochemical Cycles* **27**:723-731.
- Miller, C. S., B. J. Baker, B. C. Thomas, S. W. Singer, and J. F. Banfield. 2011. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biology* **12**:R44.

- Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE). 2010, New Orleans, 1-8.
- Mirarab, S., N. Nguyen, S. Guo, L. S. Wang, J. Kim, and T. Warnow. 2015. PASTA: Ultra-Large Multiple Sequence Alignment for Nucleotide and Amino-Acid Sequences. *Journal of Computational Biology* **22**:377-386.
- Moore, C. M., M. M. Mills, K. R. Arrigo, I. Berman-Frank, L. Bopp, P. W. Boyd, E. D. Galbraith, R. J. Geider, C. Guieu, S. L. Jaccard, T. D. Jickells, J. La Roche, T. M. Lenton, N. M. Mahowald, E. Marañón, I. Marinov, J. K. Moore, T. Nakatsuka, A. Oschlies, M. A. Saito, T. F. Thingstad, A. Tsuda, and O. Ulloa. 2013. Processes and patterns of oceanic nutrient limitation. *Nature Geoscience* **6**:701-710.
- Moore, L. R., and S. W. Chisholm. 1999. Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnology and Oceanography* **44**:628-638.
- Moriyama, E. N., and J. R. Powell. 1997. Codon Usage Bias and tRNA Abundance in *Drosophila*. *Journal of Molecular Evolution* **45**:514-523.
- Musto, H., S. Cruveiller, G. D'Onofrio, H. Romero, and G. Bernardi. 2001. Translational Selection on Codon Usage in *Xenopus laevis*. *Molecular Biology and Evolution* **18**:1703-1707.
- Nixon, S. W., J. W. Ammerman, L. P. Atkinson, V. M. Berounsky, G. Billen, W. C. Boicourt, W. R. Boynton, T. M. Church, D. M. Ditoro, R. Elmgren, J. H. Garber, A. E. Giblin, R. A. Jahnke, N. J. P. Owens, M. E. Q. Pilson, and S. P. Seitzinger. 1996. The fate of nitrogen and phosphorus at the land-sea margin of the North Atlantic Ocean. *Biogeochemistry* **35**:141-180.
- Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401**:877-884.
- Palenik, B., J. Grimwood, A. Aerts, P. Rouzé, A. Salamov, N. Putnam, C. Dupont, R. Jorgensen, E. Derelle, S. Rombauts, K. Zhou, R. Otilar, S. S. Merchant, S. Podell, T. Gaasterland, C. Napoli, K. Gendler, A. Manuell, V. Tai, O. Vallon, G. Piganeau, S. Jancek, M. Heijde, K. Jabbari, C. Bowler, M. Lohr, S. Robbens, G. Werner, I. Dubchak, G. J. Pazour, Q. Ren, I. Paulsen, C. Delwiche, J. Schmutz, D. Rokhsar, Y. Van de Peer, H. Moreau, and I. V. Grigoriev. 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. Proceedings of the National Academy of Sciences of the United States of America **104**:7705-7710.
- Paquette, A., S. Joly, and C. Messier. 2015. Explaining forest productivity using tree functional traits and phylogenetic information: two sides of the same coin over

evolutionary scale? *Ecology and Evolution* 5:1774-1783.

- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics (Oxford, England)* 20:289-290.
- Parfrey, L. W., D. J. G. Lahr, A. H. Knoll, and L. A. Katz. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences* 108:13624-13629.
- Partensky, F., and L. Garczarek. 2009. *Prochlorococcus*: Advantages and Limits of Minimalism. *Annual Review of Marine Science* 2:305-331.
- Paul, S., A. Dutta, S. K. Bag, S. Das, and C. Dutta. 2010. Distinct, ecotype-specific genome and proteome signatures in the marine cyanobacteria *Prochlorococcus*. *BMC Genomics* 11:103-103.
- Quantum GIS Development Team. 2015. Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project.
- Quigg, A., A. J. Irwin, and Z. V. Finkel. 2010. Evolutionary inheritance of elemental stoichiometry in phytoplankton. *Proceedings of the Royal Society of London B: Biological Sciences*.
- R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Redfield, A. C. 1958. The biological control of chemical factors in the environment. *American Scientist* 46:230A-221.
- Revell, L. J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3:217-223.
- Rocap, G., F. W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N. A. Ahlgren, A. Arellano, M. Coleman, L. Hauser, W. R. Hess, Z. I. Johnson, M. Land, D. Lindell, A. F. Post, W. Regala, M. Shah, S. L. Shaw, C. Steglich, M. B. Sullivan, C. S. Ting, A. Tolonen, E. A. Webb, E. R. Zinser, and S. W. Chisholm. 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424:1042-1047.
- Rodríguez, F., E. Derelle, L. Guillou, F. Le Gall, D. Vaultot, and H. Moreau. 2005. Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). *Environmental Microbiology* 7:853-859.
- Rynearson, T. A., and E. Virginia Armbrust. 2004. Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). *Journal of Phycology* 40:34-43.

- Sbrocco, E. J., and P. H. Barber. 2013. MARSPEC: Ocean climate layers for marine spatial ecology. *Ecology* **94**:979-979.
- Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, and B. Schmid. 2012. Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**:676-682.
- Sharp, P. M., T. M. Tuohy, and K. R. Mosurski. 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Research* **14**:5125-5143.
- Simao, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210-3212.
- Sjöqvist, C., A. Godhe, P. R. Jonsson, L. Sundqvist, and A. Kremp. 2015. Local adaptation and oceanographic connectivity patterns explain genetic differentiation of a marine diatom across the North Sea–Baltic Sea salinity gradient. *Molecular Ecology* **24**:2871-2885.
- Smith, S. A., and C. W. Dunn. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* **24**:715-716.
- Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312-1313.
- Stenico, M., A. T. Lloyd, and P. M. Sharp. 1994. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. *Nucleic Acids Research* **22**:2437-2446.
- Stoecker, D. K. 1998. Conceptual models of mixotrophy in planktonic protists and some ecological and evolutionary implications. *European Journal of Protistology* **34**:281-290.
- Sweeney, R. E., and I. R. Kaplan. 1980. Natural abundances of  $^{15}\text{N}$  as a source indicator for near-shore marine sedimentary and dissolved nitrogen. *Marine Chemistry* **9**:81-94.
- Tamura, K., F. U. Battistuzzi, P. Billings-Ross, O. Murillo, A. Filipinski, and S. Kumar. 2012. Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences of the United States of America* **109**:19333-19338.

- Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**:2725-2729.
- Tett, P., M. R. Droop, and S. I. Heaney. 2009. The Redfield Ratio and Phytoplankton Growth Rate. *Journal of the Marine Biological Association of the United Kingdom* **65**:487-487.
- Thrane, J.-E., D. O. Hessen, and T. Andersen. 2016. The impact of irradiance on optimal and cellular nitrogen to phosphorus ratios in phytoplankton. *Ecology Letters* **19**:880-888.
- Tolonen, A. C., J. Aach, D. Lindell, Z. I. Johnson, T. Rector, R. Steen, G. M. Church, and S. W. Chisholm. 2006. Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. *Molecular systems biology* **2**:53-53.
- Tyberghein, L., H. Verbruggen, K. Pauly, C. Troupin, F. Mineur, and O. De Clerck. 2012. Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography* **21**:272-281.
- Tyrrell, T. 1999. The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* **400**:525-531.
- Walsh, J. J. 1991. Importance of continental margins in the marine biogeochemical cycling of carbon and nitrogen. *Nature* **350**:53-55.
- Ward, B. A., S. Dutkiewicz, A. D. Barton, and M. J. Follows. 2011. Biophysical Aspects of Resource Acquisition and Competition in Algal Mixotrophs. *The American Naturalist* **178**:98-112.
- Whittaker, K. A., D. R. Rignanes, R. J. Olson, and T. A. Ryneerson. 2012. Molecular subdivision of the marine diatom *Thalassiosira rotula* in relation to geographic distribution, genome size, and physiology. *BMC Evolutionary Biology* **12**:209.
- Whittaker, K. A., and T. A. Ryneerson. 2017. Evidence for environmental and ecological selection in a microbe with no geographic limits to gene flow. *Proceedings of the National Academy of Sciences* **114**:2651-2656.
- Wilken, S., J. M. Schuurmans, and H. C. P. Matthijs. 2014. Do mixotrophs grow as photoheterotrophs? Photophysiological acclimation of the chrysophyte *Ochromonas danica* after feeding. *New Phytologist* **204**:882-889.
- Willie, E., and J. Majewski. 2004. Evidence for codon bias selection at the pre-mRNA level in eukaryotes. *Trends in Genetics* **20**:534-538.

Yang, Y., and S. A. Smith. 2014. Orthology Inference in Nonmodel Organisms Using Transcriptomes and Low-Coverage Genomes: Improving Accuracy and Matrix Occupancy for Phylogenomics. *Molecular Biology and Evolution* **31**:3081-3092.

Yvon-Durocher, G., M. Dossena, M. Trimmer, G. Woodward, and A. P. Allen. 2015. Temperature and the biogeography of algal stoichiometry. *Global Ecology and Biogeography* **24**:562-570.

Zeldovich, K. B., I. N. Berezovsky, and E. I. Shakhnovich. 2007. Protein and DNA sequence determinants of thermophilic adaptation. *PLoS Computational Biology* **3**:e5.

**FIGURES:**

Figure 1. Resolved and time calibrated phylogeny of MMETSP samples based on nearly complete 18S rRNA sequences. Red and Green plastid lineages are colored in red and green respectively. Time in represented in millions of years and dates represent major geologic periods.

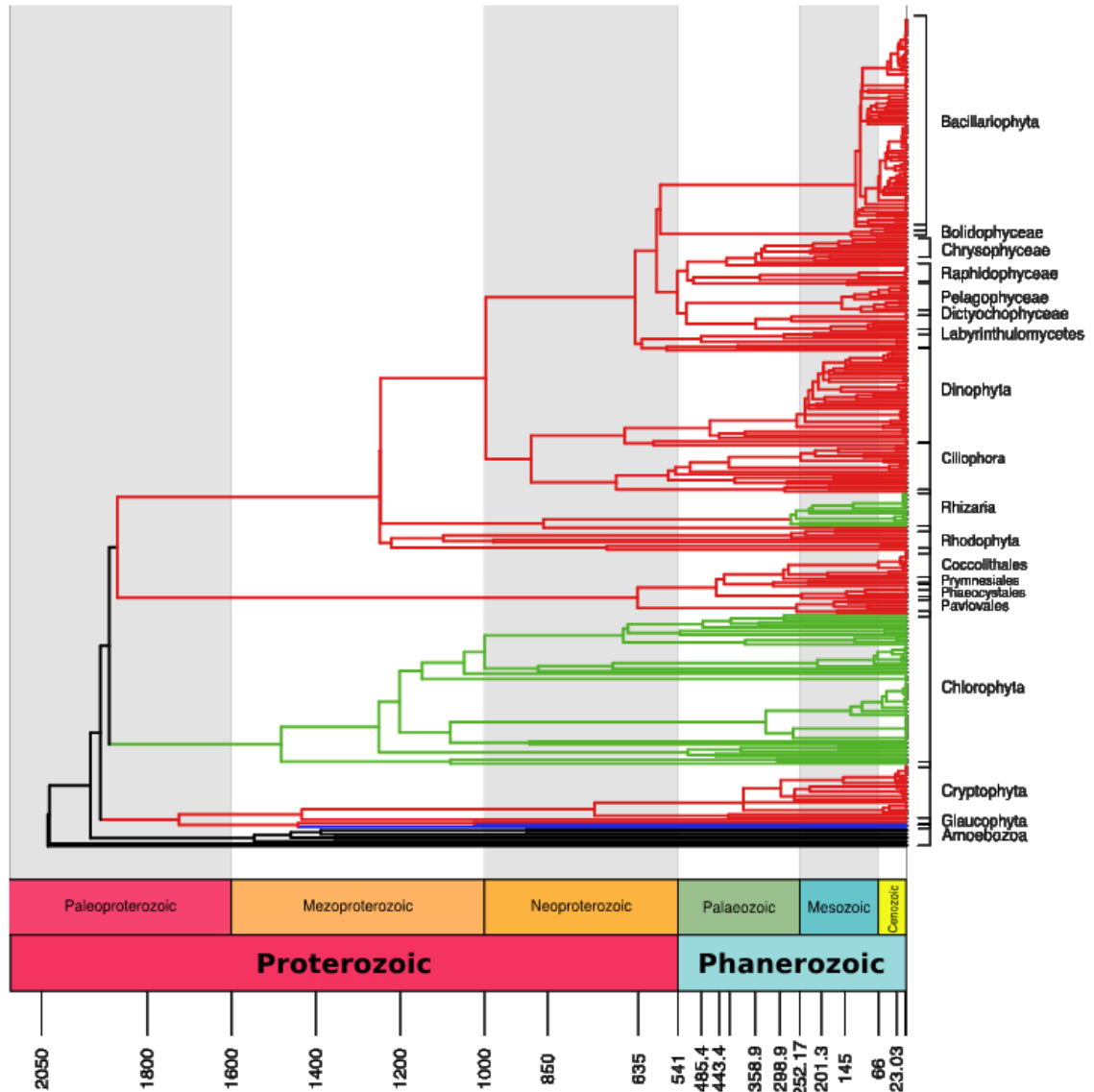
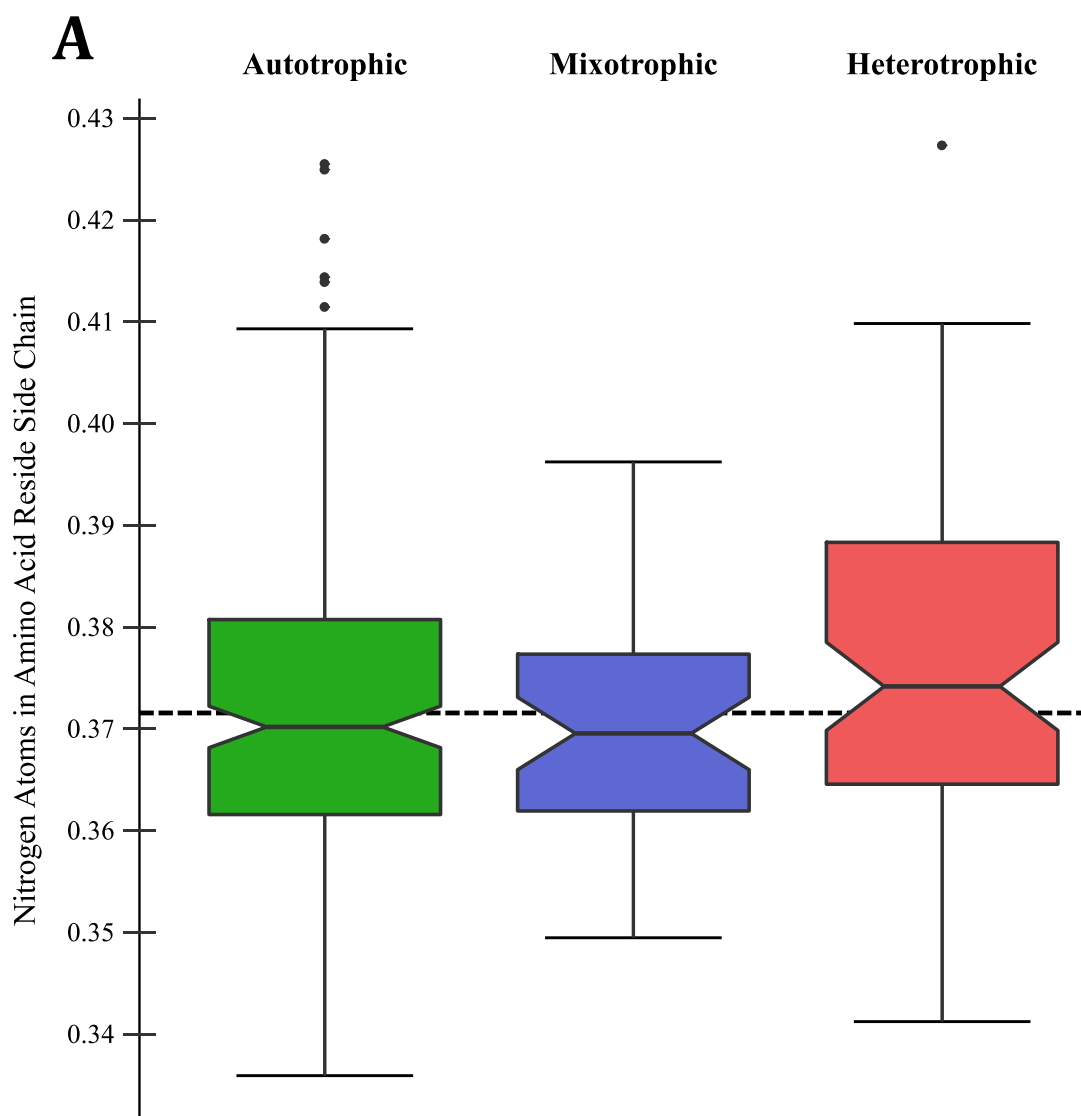


Figure 2. (A) This figure shows boxplots of the overall NARSC values among the Trophic strategies, the red line behind the boxplot identifies the median among all NARSC values. (B) This shows a boxplot with overall CARSC values among all Trophic strategies. The red line behind both figures represents the median of the values within NARSC and CARSC.





**B**

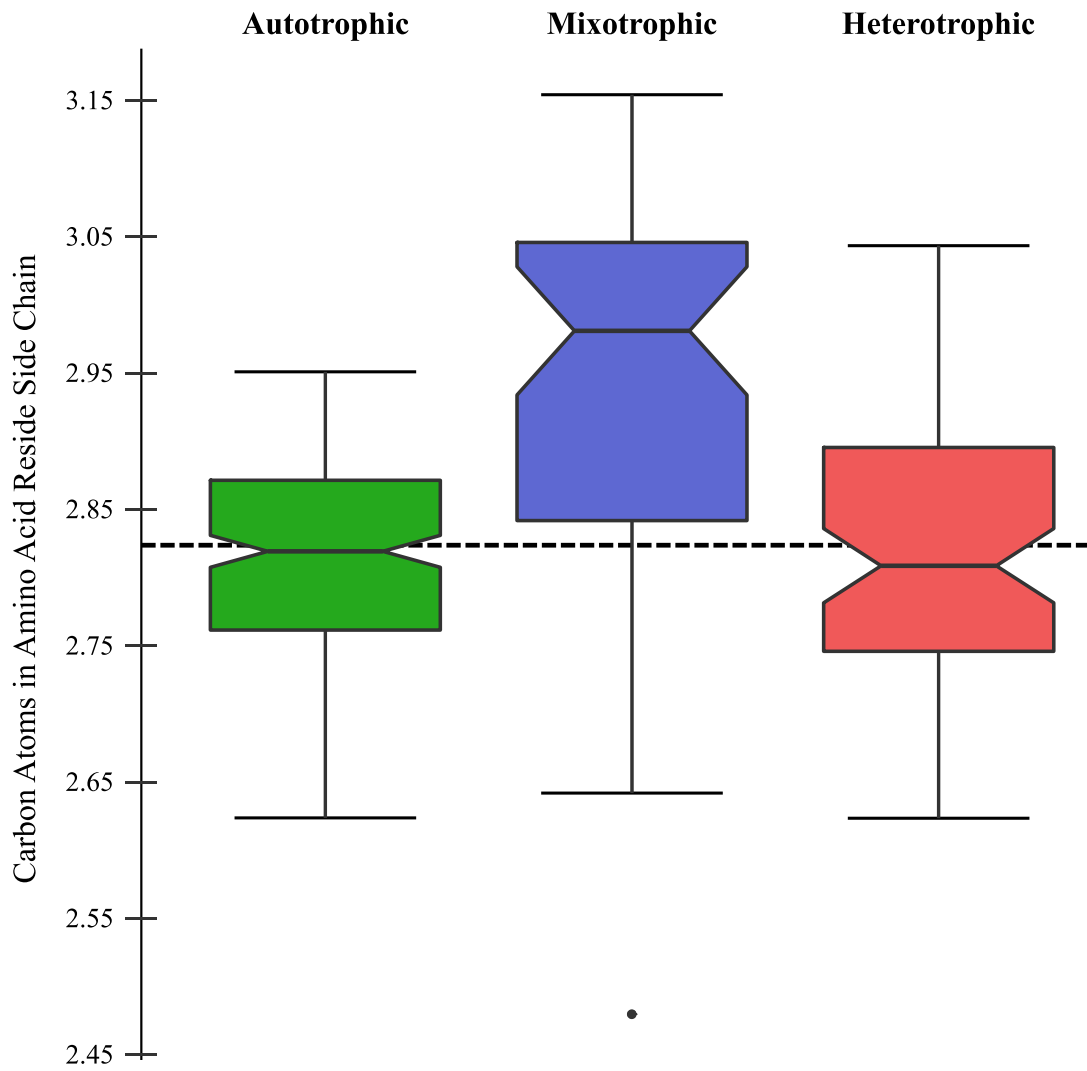


Figure 3. Moran's I correlogram showing median values of all MMETSP transcriptome and proteomes. Phylogenetic distance is represented in time (mya). Filled circles represent significant Moran's I, while unfilled are non-significant. NARSC shows trend upward indicating more similarity at shorter branches/time. The other measurements show gradual phylogenetic dissimilarity with time. Only CARSC is significant over long distances. Measurements of N and C stoichiometry show a trend of non-significance with larger phylogenetic distances.

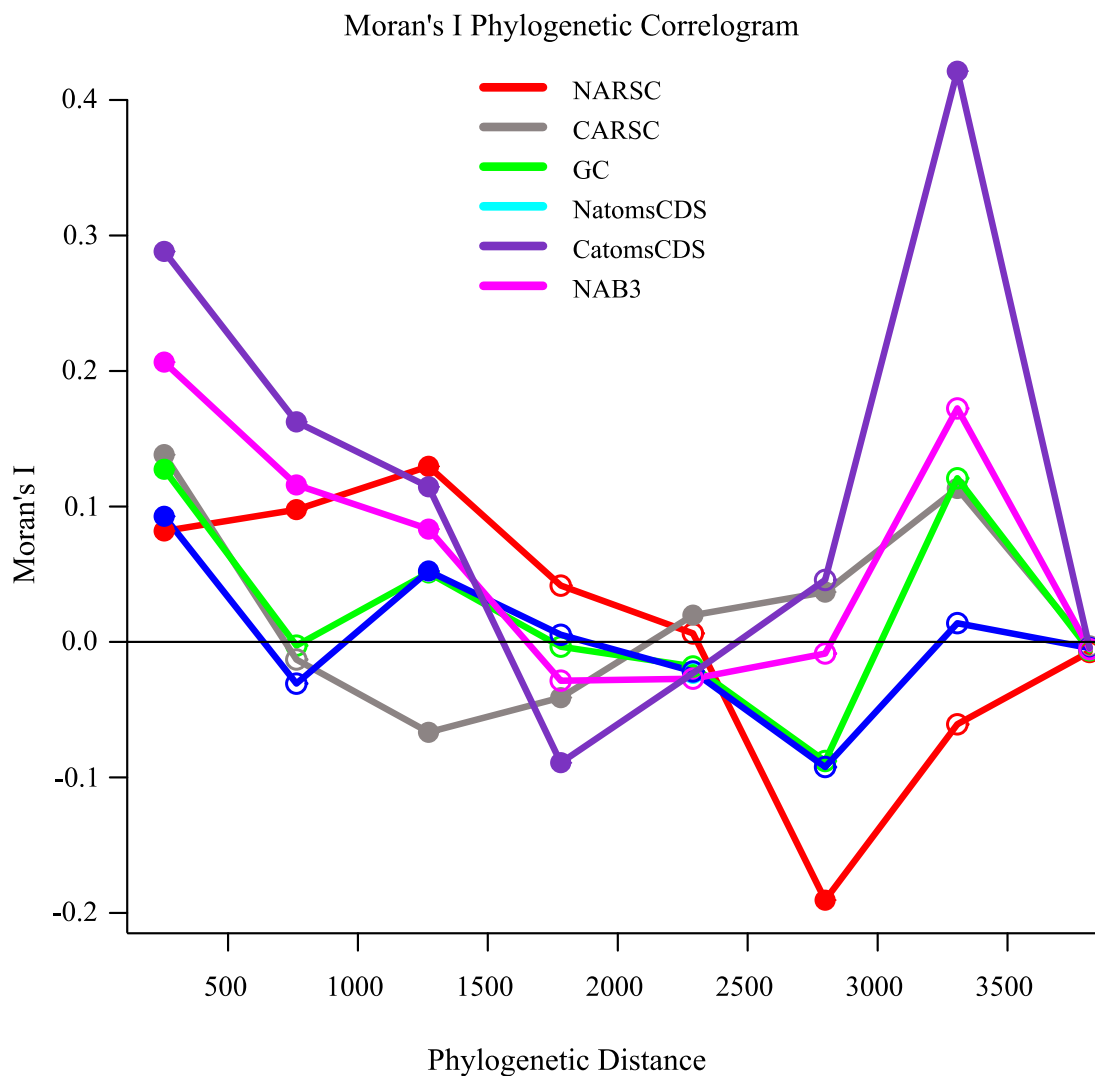


Figure 4. PGLS regression of cell size data with complete literature cases and limited to values obtained from cell culture facilities with the OU model favored. A and B represent the maximum and minimum cell diameter of culture collection, metadata and literature values (N= 258). C and D illustrate the reduced dataset only using culture collection and provided metadata (N=179). Significance of the regression plotted on corner of each.

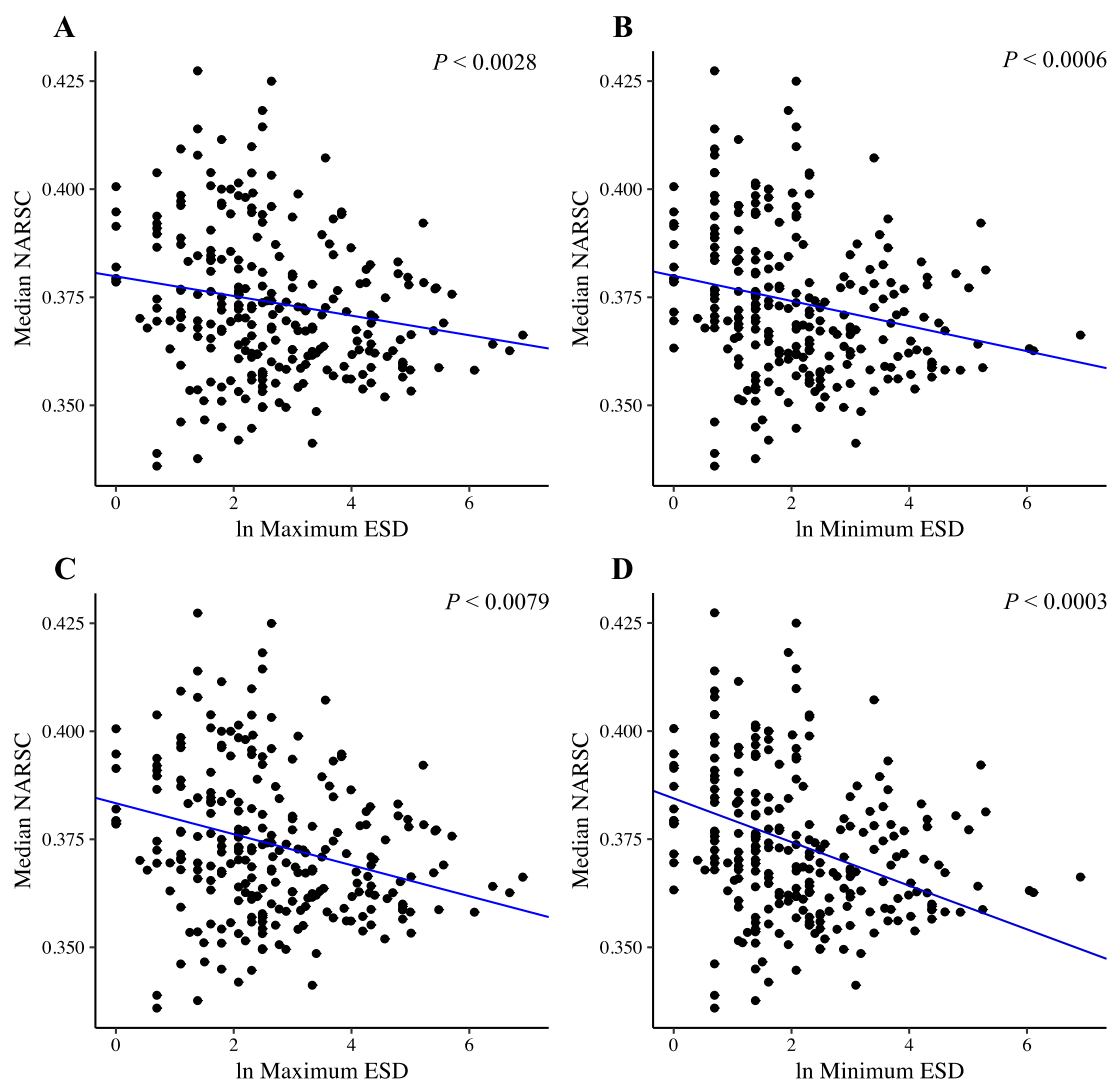
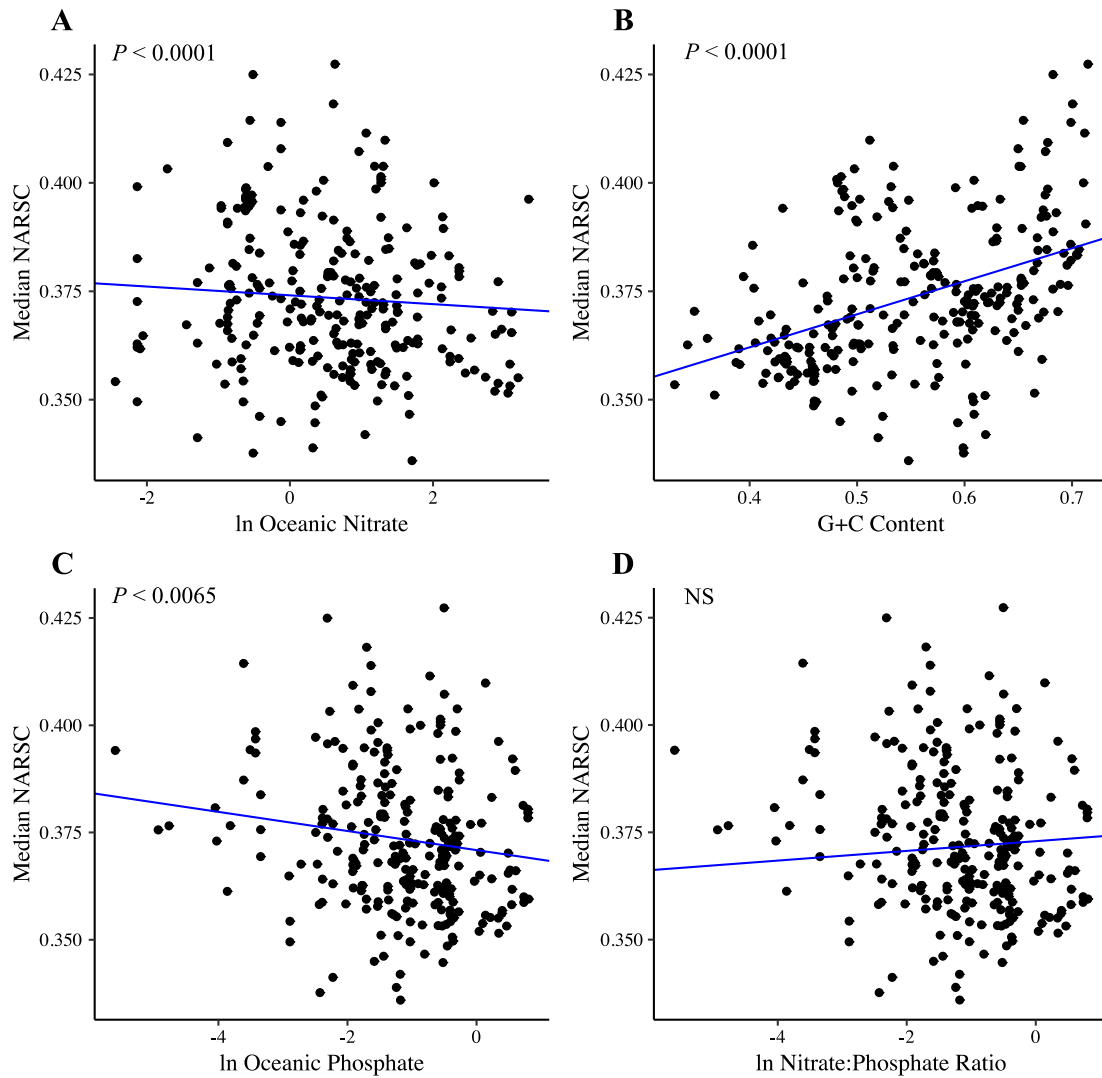


Figure 5: PGLS regression of natural log transformed nitrate and phosphate concentrations to median NARSC, of the MMETSP proteomes and untransformed G+C content of the transcriptome (OU model). (A) Correlation of NARSC and environmental nitrate, (B) Correlation of NARSC and G+C content of the transcriptome. (C) Correlation of NARSC and environmental phosphate, and (D) Correlation of NARSC and Nitrate:Phosphate ratio. Significance of the regression plotted on corner of each.



## TABLES

Table 1. A Median calculated values of transcriptome and proteome traits. B. Weighted average values by gene expression (TPM). Phylogenetic signal between traits assessed on the global MMETSP dataset and using Pagel's  $\lambda$ . Values of  $\lambda = 1$  suggest the trait has phylogenetic signal and the positive  $\Delta\text{AICc}$  describes the model difference between the BM (Brownian Model) and a Star phylogeny representing no signal. Positive  $\Delta\text{AICc}$  are indicating that  $\lambda \neq 0$  and the more complex (BM) model is preferred. NARSC: Number of Nitrogen atoms in the amino acid residue side-chain, while CARSC Number of carbon atoms in the amino acid residue side-chain, GC: guanine+cytosine percent, NAB3: nitrogen content of the 3<sup>rd</sup> codon position in the open reading frame.

| <b>A</b>                                     |           |                |                     | <b>B</b>                                     |           |                |                     |
|--|-----------|----------------|---------------------|--|-----------|----------------|---------------------|
| <b>Global All Predicted Proteins (n=341)</b> |           |                |                     | <b>Global All Predicted Proteins (n=341)</b> |           |                |                     |
| Traits                                       | $\lambda$ | <i>P</i> value | $\Delta\text{AICc}$ | Traits                                       | $\lambda$ | <i>P</i> value | $\Delta\text{AICc}$ |
| NARSC  | 0.92      | <0.0001        | 145.25              | NARSC  | 0.85      | <0.0001        | 184.62              |
| CARSC  | 0.99      | <0.0001        | 184.13              | CARSC  | 0.96      | <0.0001        | 299.36              |
| GC   | 0.99      | <0.0001        | 239.35              | GC   | 0.99      | <0.0001        | 343.18              |
| NatomCDS                                     | 0.95      | <0.0001        | 191.95              | NatomCDS                                     | 0.86      | <0.0001        | 188.55              |
| CatomCDS                                     | 1         | <0.0001        | 222.64              | CatomCDS                                     | 0.99      | <0.0001        | 282.03              |
| NAB3   | 0.98      | <0.0001        | 221.47              | NAB3   | 0.93      | <0.0001        | 284.35              |
| <b>BUSCO Homologs (n=341)</b>                |           |                |                     | <b>BUSCO Homologs (n=341)</b>                |           |                |                     |
| Traits                                       | $\lambda$ | <i>P</i> value | $\Delta\text{AICc}$ | Traits                                       | $\lambda$ | <i>P</i> value | $\Delta\text{AICc}$ |
| NARSC  | 0.85      | <0.0001        | 177.24              | NARSC  | 0.81      | <0.0001        | 75.29               |
| CARSC  | 0.97      | <0.0001        | 455.43              | CARSC  | 0.92      | <0.0001        | 308.45              |
| GC   | 1         | <0.0001        | 312.21              | GC   | 0.99      | <0.0001        | 294.59              |
| NatomCDS                                     | 0.94      | <0.0001        | 237.62              | NatomCDS                                     | 0.93      | <0.0001        | 199.41              |
| CatomCDS                                     | 1         | <0.0001        | 310.25              | CatomCDS                                     | 0.99      | <0.0001        | 252.17              |
| NAB3   | 0.97      | <0.0001        | 248.85              | NAB3   | 0.92      | <0.0001        | 249.44              |
| <b>Yang Homologs (n=341)</b>                 |           |                |                     | <b>Yang Homologs (n=341)</b>                 |           |                |                     |
| Traits                                       | $\lambda$ | <i>P</i> value | $\Delta\text{AICc}$ | Traits                                       | $\lambda$ | <i>P</i> value | $\Delta\text{AICc}$ |
| NARSC  | 0.94      | <0.0001        | 286.99              | NARSC  | 0.9       | <0.0001        | 212.92              |
| CARSC  | 0.97      | <0.0001        | 431.28              | CARSC  | 0.95      | <0.0001        | 339.41              |
| GC   | 1         | <0.0001        | 310.73              | GC   | 0.99      | <0.0001        | 296.77              |
| NatomCDS                                     | 0.96      | <0.0001        | 256.12              | NatomCDS                                     | 0.96      | <0.0001        | 263.01              |
| CatomCDS                                     | 1         | <0.0001        | 304.08              | CatomCDS                                     | 0.99      | <0.0001        | 249.79              |
| NAB3   | 0.98      | <0.0001        | 272.16              | NAB3   | 0.96      | <0.0001        | 288.94              |

Table 2. Values of Moran's I among Medians and Weighted Averages (SD = standard deviation). Values in gray boxes were not significant. Positive observed values indicate the tips of the phylogeny more similar in value than by chance. Top boxes show all proteins, then filtered to contain BUSCO homologs. The last row represents the all proteins filtered to include Yang and Smith homologs

| Moran's I Phylogenetic Autocorrelation Medians |          |          |      |         | Moran's I Phylogenetic Autocorrelation Weighted by Average TPM |          |          |      |         |
|--|----------|----------|------|---------|--|----------|----------|------|---------|
| Global All Predicted Proteins (n=341)          |          |          |      |         | Global All Predicted Proteins (n=341)                          |          |          |      |         |
| Traits   | Observed | Expected | SD   | P value | Traits   | Observed | Expected | SD   | P value |
| NARSC  | 0.043    | -0.003   | 0.03 | 0.1306  | NARSC  | 0.059    | -0.003   | 0.03 | 0.039   |
| CARSC  | 0.116    | -0.003   | 0.03 | 0.0001  | CARSC  | 0.118    | -0.003   | 0.03 | 0.0001  |
| GC   | 0.1      | -0.003   | 0.03 | 0.0007  | GC   | 0.104    | -0.003   | 0.03 | 0.0004  |
| NatomCDS                                       | 0.153    | -0.003   | 0.03 | <0.0001 | NatomCDS   | 0.131    | -0.003   | 0.03 | <0.0001 |
| CatomCDS                                       | 0.094    | -0.003   | 0.03 | 0.0014  | CatomCDS   | 0.088    | -0.003   | 0.03 | 0.0027  |
| NAB3   | 0.121    | -0.003   | 0.03 | <0.0001 | NAB3   | 0.108    | -0.003   | 0.03 | 0.0003  |
| BUSCO Homologs (n=341)                         |          |          |      |         | BUSCO Homologs (n=341)   |          |          |      |         |
| Traits   | Observed | Expected | SD   | P value | Traits   | Observed | Expected | SD   | P value |
| NARSC  | 0.059    | -0.003   | 0.03 | 0.0408  | NARSC  | 0.038    | -0.003   | 0.03 | 0.1643  |
| CARSC  | 0.118    | -0.003   | 0.03 | 0.0001  | CARSC  | 0.102    | -0.003   | 0.03 | 0.0005  |
| GC   | 0.088    | -0.003   | 0.03 | 0.0027  | GC   | 0.081    | -0.003   | 0.03 | 0.0056  |
| NatomCDS                                       | 0.141    | -0.003   | 0.03 | <0.0001 | NatomCDS   | 0.145    | -0.003   | 0.03 | <0.0001 |
| CatomCDS                                       | 0.084    | -0.003   | 0.03 | 0.0044  | CatomCDS   | 0.072    | -0.003   | 0.03 | 0.0136  |
| NAB3   | 0.117    | -0.003   | 0.03 | 0.0001  | NAB3   | 0.13     | -0.003   | 0.03 | <0.0001 |
| Yang Homologs (n=341)                          |          |          |      |         | Yang Homologs (n=341)  |          |          |      |         |
| Traits   | Observed | Expected | SD   | P value | Traits   | Observed | Expected | SD   | P value |
| NARSC  | 0.079    | -0.003   | 0.03 | 0.0068  | NARSC  | 0.026    | -0.003   | 0.03 | 0.3305  |
| CARSC  | 0.126    | -0.003   | 0.03 | <0.0001 | CARSC  | 0.114    | -0.003   | 0.03 | 0.0001  |
| GC   | 0.09     | -0.003   | 0.03 | 0.0023  | GC   | 0.089    | -0.003   | 0.03 | 0.0025  |
| NatomCDS                                       | 0.175    | -0.003   | 0.03 | <0.0001 | NatomCDS   | 0.168    | -0.003   | 0.03 | <0.0001 |
| CatomCDS                                       | 0.084    | -0.003   | 0.03 | 0.0043  | CatomCDS   | 0.075    | -0.003   | 0.03 | 0.0099  |
| NAB3   | 0.131    | -0.003   | 0.03 | <0.0001 | NAB3   | 0.143    | -0.003   | 0.03 | <0.0001 |

Table 3. PGLS correlations among the different evolutionary models tested, and NARSC. A.) Includes all available cell size data from culture collections, MMETSP metadata, and literature. B.) Data filtered to include only data from culture collections and deposited MMETSP metadata. All models favored the OU model. The Phy-ANOVA testing of the OU model is below each correlation along with its *P*-value.

All Cell Size Data

NARSC correlated to log(MaxLength)

| <b>A</b> | Model              | df | AIC | BIC       | logLik    | Test       | L.Ratio | p-value         |
|----------|--------------------|----|-----|-----------|-----------|------------|---------|-----------------|
|          | Brownian Motion    | 1  | 3   | 2110.956  | 2121.615  | -1052.4782 |         |                 |
|          | Ornstein Uhlenbeck | 2  | 4   | -1440.645 | -1426.433 | 724.3225   | 1vs 2   | 3553.601 <.0001 |
|          | Pagel's Lambda     | 3  | 4   | -1409.234 | -1395.022 | 708.617    |         |                 |

Ornstein Uhlenbeck model ANOVA df=256, F-value 112531.9, P value < 0.0028

NARSC correlated log(MinLength)

|  | Model              | df | AIC | BIC       | logLik    | Test       | L.Ratio | p-value        |
|--|--------------------|----|-----|-----------|-----------|------------|---------|----------------|
|  | Brownian Motion    | 1  | 3   | 2113.989  | 2124.648  | -1053.9946 |         |                |
|  | Ornstein Uhlenbeck | 2  | 4   | -1443.431 | -1429.219 | 725.7156   |         | 3559.42 <.0001 |
|  | Pagel's Lambda     | 3  | 4   | -1409.509 | -1395.298 | 708.7547   |         |                |

Ornstein Uhlenbeck model ANOVA df=256, F-value 114256.51, P value < 0.0006

Culture Collection and Metadata only

NARSC correlated to log(MaxLength)

| <b>B</b> | Model              | df | AIC | BIC       | logLik    | Test      | L.Ratio | p-value         |
|----------|--------------------|----|-----|-----------|-----------|-----------|---------|-----------------|
|          | Brownian Motion    | 1  | 3   | 1512.4288 | 1521.9909 | -753.2144 |         |                 |
|          | Ornstein Uhlenbeck | 2  | 4   | -973.3645 | -960.6149 | 490.6822  |         | 2487.793 <.0001 |
|          | Pagel's Lambda     | 3  | 4   | -954.1756 | -941.4261 | 481.0878  |         |                 |

Ornstein Uhlenbeck model ANOVA df=256, F-value 69188.24, P value <0.0079

NARSC correlated log(MinLength)

|  | Model              | df | AIC | BIC       | logLik    | Test     | L.Ratio | p-value         |
|--|--------------------|----|-----|-----------|-----------|----------|---------|-----------------|
|  | Brownian Motion    | 1  | 3   | 1509.4261 | 1518.9882 | -751.713 |         |                 |
|  | Ornstein Uhlenbeck | 2  | 4   | -979.7756 | -967.0261 | 493.8878 |         | 2491.202 <.0001 |
|  | Pagel's Lambda     | 3  | 4   | -957.433  | -944.6835 | 482.7165 |         |                 |

Ornstein Uhlenbeck model ANOVA df=256, F-value 71900.8, P value < 0.0003

Table 4. A. PGLS correlation of NARSC as the dependent variable on the independent log (nitrate) of the ocean of its original isolation, extracted from GIS layers. B. PGLS correlation of NARSC as the dependent variable on the independent log (phosphate) of the ocean of its original isolation extracted from GIS layers.

**Proteome Median**

NARSC correlated to log Nitrate

| <b>A</b> | Model              | df | AIC         | BIC       | logLik     | Test   | L.Ratio  | p-value |
|----------|--------------------|----|-------------|-----------|------------|--------|----------|---------|
|          | Brownian Motion    | 1  | 3 2418.611  | 2429.662  | -1206.3054 |        |          |         |
|          | Ornstein Uhlenbeck | 2  | 4 -1639.671 | -1624.937 | 823.8356   | 1 vs 2 | 4060.282 | <0.0001 |
|          | Pagel's Lambda     | 3  | 4 -1612.842 | -1598.107 | 810.4209   |        |          |         |

Ornstein Uhlenbeck model ANOVA df=256, F-value 112531.9, P value <0.0001

NARSC correlated to G+C content

| <b>B</b> | Model              | df | AIC         | BIC       | logLik     | Test   | L.Ratio | p-value |
|----------|--------------------|----|-------------|-----------|------------|--------|---------|---------|
|          | Brownian Motion    | 1  | 3 2114.271  | 2124.929  | -1054.1352 |        |         |         |
|          | Ornstein Uhlenbeck | 2  | 4 -1475.9   | -1461.688 | 741.9499   | 1 vs 2 | 3592.17 | <.0001  |
|          | Pagel's Lambda     | 3  | 4 -1470.545 | -1456.333 | 739.2726   |        |         |         |

Ornstein Uhlenbeck model ANOVA df=256, F-value 132521.58, P value <0.0001

NARSC correlated to log Phosphate

| <b>C</b> | Model              | df | AIC         | BIC       | logLik     | Test   | L.Ratio  | p-value |
|----------|--------------------|----|-------------|-----------|------------|--------|----------|---------|
|          | Brownian Motion    | 1  | 3 2109.573  | 2120.232  | -1051.7867 |        |          |         |
|          | Ornstein Uhlenbeck | 2  | 4 -1439.106 | -1424.894 | 723.553    | 1 vs 2 | 3550.679 | <.0001  |
|          | Pagel's Lambda     | 3  | 4 -1409.784 | -1395.572 | 708.8919   |        |          |         |

Ornstein Uhlenbeck model ANOVA df=256, F-value 111951.33, P value 0.0065

NARSC correlated to log Nitrate:Phosphate

| <b>D</b> | Model              | df | AIC         | BIC       | logLik   | Test   | L.Ratio | p-value |
|----------|--------------------|----|-------------|-----------|----------|--------|---------|---------|
|          | Brownian Motion    | 1  | 3 1519.0959 | 1528.6581 | -756.548 |        |         |         |
|          | Ornstein Uhlenbeck | 2  | 4 -967.734  | -954.9845 | 487.867  | 1 vs 2 | 2488.83 | <.0001  |
|          | Pagel's Lambda     | 3  | 4 -953.7003 | -940.9508 | 480.8501 |        |         |         |

Ornstein Uhlenbeck model ANOVA df=256, F-value 67305.46, P value 0.2189



## **Chapter 4: A Case Study Approach To Teaching Introductory Biology Undergraduates The Application Of Genomic Data Analysis In Ecological Studies.**

### **Preface**

Case studies are an effective way to introduce students to material and engage them in exploration of that material in an active learning environment (Herreid, 2007). A long-time component of medical and legal education, the case study teaching method has recently become a growing force in STEM education. Case studies have been developed and demonstrated to be effective platforms to teach core introductory biology concepts ranging from pedigrees to ecosystem change (NCCSTS 2017 ; Herreid, 2007; Herreid *et al.*, 2014). Interrupted case studies in particular are an excellent format that creates a dynamic educational environment that requires student/audience participation and peer learning (Udovic *et al.*, 2002; Herreid, 2006; Herreid, 2007). Students may unknowingly be familiar to this type of teaching through the case study approach often used in serial medical and legal television programs whose format mimics the progressive approach to discovery and learning inherent in interrupted case studies. The audience and students are presented with a story, (i.e., problem) and are taken on a journey to figure out the answer and reach a logical conclusion from collected data. Application of this approach take a process-oriented approaches to understanding instead of relying on simple, lower order learning (Bloom, 1956; Krathwohl, 2002).

One topic that this approach can be applied to is students' understanding of molecular biology and correcting misconceptions about the Central Dogma of Molecular Biology. This is particularly problematic because the growing use of

bioinformatics and “big data” approaches to solve complex biological problems requires a thorough understanding of the fundamental principles described by the Central Dogma. The importance of this topic has been highlighted in the Vision and Change guidelines (AAAS, 2011) which state that all undergraduates should be able to understand the “Information Flow, Exchange and Storage” core concepts as they relate to the growth and behavior of an organism based on gene expression, how scientists analyze large datasets, and the “Living systems are interconnected and interacting” core concept (AAAS, 2011).

This case study is a dilemma-based case that uses the harmful bloom-forming and toxin producing dinoflagellate *Karenia brevis* a.k.a. “Florida red tide” to teach and probe students’ understanding of different aspects of the Central Dogma of Genomics (Pevsner, 2015), an extension of the Central Dogma of Molecular Biology (Crick, 1958) representing the bioinformatic and sequencing tools that undergraduates need to understand to tackle 21<sup>st</sup> Century questions in molecular biology (Figure 1.).

We developed a case study that uses a systems biology approach to explore how the Central Dogma of Molecular Biology and Dogma of Genomics can be used to understand a common event that occurs on an almost yearly cycle. Cases with real world examples provide effective constructs to help students understand abstract concepts like the Central Dogma of Molecular Biology through application to a tangible modern example (Herreid, 2007). This case also addresses specific deficiencies identified in the Vision and Change Call to Action (American Association for the Advancement of Science, 2011) outlining essential content and competencies relating to the canonical Central Dogma of Molecular Biology.

This case study and teaching notes have been formatted for submission to the National Center for Case Study Teaching and Education. The PowerPoint slides that will be submitted with the case are provided in the Appendix.

## **Case Teaching Notes**

For

**Transcriptome of the Toxic Tide: How can biology and computer science lead toward a deeper understanding of systems biology?**

By

Joshua T. Cooper, Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK

J. Phil Gibson, Department of Microbiology and Plant Biology and Department of Biology, University of Oklahoma, Norman, OK

Keyword: Algae, Bioinformatics, Transcriptome, gene expression, ecology, harmful algae, ecophysiology

Education Level: Undergraduate Biology, Upper Division, Ecology and Environmental Science,

Format: PowerPoint

Type/Method: Clicker, Interrupted

Language: English

Subject Headings: Biology, Bioinformatics, Ecology

## Introduction

Bioinformatics is a growing field that requires knowledge of biology, computer programming and statistics in an effort to solve complex biological problems and analyze large datasets. Introducing undergraduate biology majors to the rapidly advancing field of bioinformatics and “big data” are important to demonstrate research and career possibilities to them that merge biology, math and computer science. This case helps students become familiar with terminology such as genome, transcriptome, proteome, and metabolome that are common in bioinformatics research (Pevsner, 2015).

This case study is a “clicker case” that combines the use of personal student response systems (e.g., iClicker) with an interrupted case format to allow students to be interactive with the material. A PowerPoint slide show is used to combine brief amounts of lecture material to establish and guide the case’s direction with interspersed questions. The case is delivered in short interrupted blocks, with time for students to respond to “clicker questions” and then allowed to see all respondent results. This allows a brief discussion of any misconceptions identified by the questions to aide students. This case is based on real data from an experiment (Cooper, unpublished) investigating the transcriptome response to nitrogen or phosphorus limitation in *Karenia brevis*, a toxic dinoflagellate. The case asks students to think about how organisms respond to their environments at the cellular level using RNA-sequencing (RNA-seq) transcriptome profiling. Transcriptome profiling is a powerful tool for studying global gene expression and gene content (genomic potential) of any organism. Previously most gene expression studies were limited to model organisms that have sequenced genomes

such as (*Drosophila melanogaster*, *Homo sapiens*, *Mus musculus*, and *Arabidopsis thaliana*) and because they were well studied, with annotated proteins. But now, because any organism's transcriptome can be sequenced, scientists can investigate how non-model organisms respond to stress.

## **Objectives**

Upon completing of this case, students should have improved understanding how the Central Dogma of Molecular Biology applies to bioinformatic tools, and how it can be used to understand an organism's physiology. Specifically, should:

- Identify the correct sequence of the Central Dogma of Molecular Biology
- Describe what genomic technologies are used to study each kind of biomolecule (DNA, RNA, Protein, Metabolites).
- Understand that expressed genome via the transcribed transcriptome and translated proteome represent an organism's complex phenotype.
- Recognize that cellular pathways are not unidirectional, and that components can end up in multiple cellular pools.
- Discuss how nutrient stress is associated with *Karenia* life cycle and how an organism's gene expression may give clues to when the bloom has run out of growth supporting nutrients by expressing stress related messages.
- Compose a simplified newsletter, white paper or news article about the Town Hall meeting covering the major points that residents would like to know.

## **Outcome**

Stress influences gene expression above/beyond/lower than normal.

All organisms do not respond in the same way.

All genes are not expressed/over-expressed

## **Competency**

Correctly interpreting transcriptome data, and knowing the limitations.

## **Misconceptions Addressed**

Public perception of Florida red tide events gleaned from newspapers (1987-2012) are often framed as environmental risk stories but do not highlight the impacts to public health or economic loss either regionally or locally (Li *et al.*, 2015). Thus local newspapers can amplify risks when they act as creditable information sources regarding perceived fear of red tides. Also, another study found that knowledge gaps exist in local populations affected by red tide events, where locals and tourists can have similar misperceptions (Nierenberg *et al.*, 2010). Using these references we identified misconceptions that could be highlighted, including misconceptions in the core concept standards of information flow and cells as systems identified by AAAS (2011).

- Red tides are Harmful Algal Blooms (HABs) and are not always red (General Misconception to all red tides and HABs).
- Anthropogenic nutrient loading through increased nitrogen or phosphorus does not explain all HAB events.
- The Florida red tide is caused by a microscopic photosynthetic toxin-producing dinoflagellate native to the Gulf of Mexico and is not a bacterium like *Vibrio*, that may cause similar shellfish poisonings.
- Transcription of genes does not mean they will be translated.

- Not all of the genome is expressed at the same time and not all of the genome is expressed under the same conditions.

## **Background**

This case study is centered on the *Karenia brevis* a dinoflagellate alga native to the Gulf of Mexico, where it can form dense aggregations also referred to as “blooms” under certain conditions (Vargo *et al.*, 2008, Steidinger, 2009, Vargo, 2009). Here we describe that unlike other toxic algae such as *Karenia brevis*, is actually a native species to the Gulf of Mexico and has been there since at least the 1500’s based on early Spanish sailor accounts of fish kills presumed to be *Karenia* near Floridian waters. This is a good example to show students that not all problematic algae blooms are caused by invasive species.

The case explores how researchers can use genomic “big data” and the ecology of an organism to study how it’s cells and pathways function together and how those internal cellular pathways scale up to ecosystem level processes like nitrate assimilation. The systems biology approach to studying the smallest component of a system to understand how its put together with others in larger systems.

This story is built around the instructor acting as the organizer of a town hall meeting among residents in a hypothetical coastal town in Florida where red tide events have been occurring. The dilemma is that red tide events seem to be happening every year now, and their duration from start and finish is getting longer. *Karenia brevis* has been studied for 50+ years in Florida after the first bloom and fish kill was reported in the late 1940s (Brand and Compton 2007, Steidinger 2009, Brand et al. 2012). Over time, researchers have discovered the ecology of this one species is very complex, and



that no direct relationship of its abundance can be tied directly to any single anthropogenic cause. This case is effective because it contains human and societal issues that can engage students beyond just memorizing facts. *Karenia* can produce several similar neurotoxins that can impair or be lethal to some marine life and can cause severe respiratory effects in humans (Kirkpatrick *et al.*, 2004; Backer, 2009; Landsberg *et al.*, 2009; Brand *et al.*, 2012; Fauquier *et al.*, 2013; Sipler *et al.*, 2014). The town hall setting helps to get students to participate as if they were a member of that community, rather it be a seasonal resident, homeowner, concerned citizen or even as a reporter for the local newspaper.

The story starts with an ecologist, Dr. Sally Harris from, the Florida Fish and Wildlife Commission (FFWC) who serves as our ecological expert in all things relate to *Karenia brevis*. In the second half of the story, Dr. Harris introduces Dr. Mateo Edwards who is the lead researcher at a fictional biotechnology company working closely with FFWC to develop ways to predict and track the “health” or physiological status of bloom events with new molecular tools. Dr. Edwards then explains the Central Dogma of Molecular Biology and how it relates to bioinformatic approaches being used to investigate these types of ecological phenomena. The part of the case with Dr. Edwards helps students explore how transcriptomes can be used to study organism’s adaptations and how information is transferred in cells. Understanding their response to environmental stressors can help us understand how they survive and persist, how and when they make the toxins, and predict when they might terminate or dissipate.

## **Classroom Management/Blocks of Analysis**

This case study has been used in large introductory-level biology classes of 150-350 students. Students work together in self-assembled groups of 4-6 people sitting in their immediate areas. Each student has their own clicker to respond to questions but students are encouraged to talk with neighbors about the correct choice. This case can be taught in a 50-minute period. Prior to teaching this case, students should have familiarity with fundamental concepts related to ecosystems and trophic relations among the abiotic pools as a source of energy to higher trophic levels.

## **Teaching the Case**

The case presentation contains photographs, graphs, illustrations and text slides. The case is appropriate for use in units on ecology, pollution, and cell biology. The instructor will help guide the case, beginning as a moderator of the meeting to establish the dialog of two guest speakers who are presenting information about *Karenia* to the public.

## **Questions**

There are additional resources after the case. The end of the case allows for the instructor to task students with finding more information about connecting the Central Dogma of Molecular Biology together in a diagram. Also you ask them to think when you would you want to use different types of ‘Omics technologies? And lastly, to incorporate another level of systems biology thinking, you can have them look up the nitrate assimilation pathway and ask them why do organisms need nitrogen and phosphorus? The short answer is to make new DNA, RNA, amino acids and proteins.

And the assimilation pathway will be connected to the TCA cycle through 2-oxoglutarate and Glutamine/Glutamate synthetase. The phosphorus cell needed are used for activating proteins phosphorylate proteins to become active, but also to make new ATP.

## **Slides**

Prologue: The instructor should start by describing the scenario as a public meeting between home owners, vacation home renters and tourists interested in the red tide events occurring outside. The instructor can play the role of the meeting organizer, and tell the students that an ecologist “Sally” or “Dr. Sally” from the Department of Florida Fish and Wildlife Conservation is there to tell them about the frequency red tides residents have seen lately. The instructor can also foreshadow that Dr. Sally has invited a colleague she has been working with at BioMarkerX a biotechnology company to describe his new approaches to try and monitor *Karenia* blooms.

Slide 1: Title slide

Slide 2: Acting as the meeting organizer the slide establishes the case as a town hall meeting with local residents in a hypothetical coastal town that has experienced a number of red tides events this year and are concerned about how it affects their communities and what is being done about it.

Slide 3: This slide introduces the meeting agenda as an outline to the case study and introduces two guests that will be presenting information about *Karenia* to the meeting.

Slide 4: Tell students that these questions were submitted by residents beforehand. Ask students to discuss questions they might have in their groups.

Slide 5: These slides outline case study with questions submitted by residents prior to the meeting.

Slide 6: Use this slide to transition to the first clicker questions

Slide 7-9: Survey clicker questions allow students to work in groups and discuss as they answer. These questions help to identify misconceptions students may have prior to the case study, if they have ever heard of red tides before. It also makes them think about what concern they would have about what they might affect.

Slide 10: Picture of red tide on slide 10 used as a background for brief discussions to take place. This is a general picture of a red tide event; however additional photographs may be included of the “Florida red tide” to make it more specific. Use this slide to show them that bloom events are large, and can affect many areas simultaneously.

Slide 11 to 23: These slides are for Dr. Harris to discuss the ecology information and answers to the FAQs by the residents about the red tide and HABs in general. Dr. Sally Harris works for the Florida Wildlife Conservation Commission (FWCC) where she is a wildlife management ecologist that is an expert in the ecology of *Karenia brevis*. And the FWCC has been developing new strategies to determine the “health” of blooms, in hopes of predicting when they might disappear.

Slide 12: This slide emphasizes that although algae form the base of aquatic food webs, too many algae can actually be disastrous. Not all algae are the same, so highlight that these pictures show a diversity of different types of microscopic algae imaged with scanning electron microscopy.

Slide 13: Use this slide to show students that the term red tide is out dated and that not all red tides are actually red! Instead Sally calls them Harmful Algal Blooms (HABs).

Slide 14: Additional information regarding HABs in general. Most HAB events are typically linked to enhanced nutrient loading in aquatic ecosystems, which are usually reported in the news. This is a set up to make them think that *Karenia* is linked to nutrients like most HABs. However, they will find out that it is not.

Slide 15: Use this slide to show how large HAB events can large spatial scale, thereby affecting thousands of residents. This is a freshwater example of Lake Erie and the green swirls south of Canada on left is the algal bloom of the cyanobacteria *Microcystis aeruginosa*..

Slide 16: Here Sally describes that not all HABs make toxins, but they can use up so much oxygen in the nighttime that they can deplete water of oxygen killing fish. Here we are introducing algae “resting cells” that can go dormant, waiting for better conditions. Some algae have a bet-hedging strategy to survive when conditions become unfavorable by creating resting cells or spores. These spores or cells can be rejuvenated or awakened when favorable conditions return. This will be brought up again in a few slides, when talking about *Karenia*’s life cycle.

Slide 17: This slide illustrates the factors that can influence algae growth, and the animated boxes will show them biotic factors that can limit their growth by herbivory, but fish waste, and zooplankton excretion can promote. But, focus on the abiotic factors that promote their growth and that cycles of dissolved nutrients and trophic interactions are not always linear. Many come back to the same pool of recycled nitrogen that can again promote algae (phytoplankton).

Slide 18: Introduce the organism *Karenia brevis*, a small unicellular photosynthetic alga. *Karenia* makes several toxins of which the major two are illustrated on the left.

Tell them, these toxins are some of the most complex toxins we know. The toxins are polyether neurotoxins that bind to voltage-sensitive sodium channels in cellular membranes found in muscle and neural tissues. The toxin binds the sodium channel in the membrane resulting persistent activation of the neuron, skeletal or cardiac cells (<http://www.nmfs.noaa.gov/pr/pdfs/health/brevetoxin.pdf>).

Slide 19: Use this slide to illustrate that *Karenia* is a native and not an invasive species. Also those red tides in Florida have been around for a long time. Early records of Spanish explorers noted fish kills off the west coast of present day Florida and the appearance of “saw dust” on the waters surface. The “saw dust” is thought to represent the N-fixing cyanobacteria *Trichodesmium* that co-occurs with *Karenia*, and the fish kills potentially caused by *Karenia*. This slide sets up the students to think *Karenia* blooms mostly near two large cities on the west coast. They might think blooms could be linked to runoff or pollution from cities.

Slide 20: This slide describes the general life cycle of *Karenia brevis*. Highlight that the question marks represent a possible resting cell stage, which has not been seen before, but many think it makes resting cells when conditions become unfavorable. These may be prevalent when the blooms disappear, and could be why they may reappear suddenly if conditions for growth improve awakening them from “dormancy”. Key elements of its life cycle include haploid adult cells go through mitotic division to grow (asexual reproduction). Cells that are reproductive form gamete like cells that can fuse to form a diploid zygote cell, which may become a resting spore. When conditions become favorable again for growth, resting cells hatch, go through meiosis to form four daughter haploid cells, returning to haploid mitotic growth.

Slide 21: Describe the toxin effects on human and marine life. This is a transition to the set up the next slide, where Dr. Harris will show them how nutrient stress may increase toxicity. Also highlight that local shellfish's can be contaminated by eating the algae cells, so they are not safe to eat if locally caught during a bloom event; however shellfish from grocery stores are not contaminated. This can be a way to address a misconception and perceived fear about *Karenia* blooms reported in the media.

Slide 22: This slide is used to show students that with a single species, some individuals produce more toxin than others, and that nutrient limitation, in particular P-limitation seems to simulate higher toxin production. Use this slide to get the students to buy in to the reasoning for investigating the toxic tide.

Slide 23: Ask students to summarize Sally's main points of the ecology of *Karenia*. This slide is used to get students to engage in part of the case story.

Slide 24: These are Dr. Harris' major points to emphasize. These will be important for students to be thinking about when they are planning and designing the transcriptome experiment. Here the idea that abiotic factors that limit growth can be used to stress algae and determine their physiological response.

Slide 25: Introduce Dr. Mateo Edwards, who is a lead researcher at a hypothetical company called BioMarkerX. He will be telling them about the transcriptome of *Karenia*.

Slide 26: Mateo uses this slide to talk about nitrogen cycling in marine systems, focusing on the nutrient state and support of phytoplankton communities, which *Karenia* is a member of. Highlight that microorganisms control the conversions among forms of nitrogen into forms that can fuel phytoplankton communities, and that those

nutrients from dying phytoplankton get recycled and returned. Mateo should say, “So by understanding the microbes that convert the different forms of nitrogen we can understand how the system works.”

Slide 27: This slide summarizes 50+ years of research on *Karenia brevis*. Dr. Edwards will be presenting this stepped animation that Sally worked on with him. Use this slide to highlight that HAB are complex and not one single factor is responsible for causing and maintaining blooms.

1. Blooms start 10-50 off shore in nutrient depleted waters, where Saharan Dust rich in iron particles blows across into the eastern Gulf of Mexico where it fuels blooms of the Nitrogen fixing cyanobacteria *Trichodesmium* (Tricho)
2. When “Tricho” is actively fixing atmospheric Nitrogen, it leaks dissolved ammonium, other dissolved organic nitrogen sources, that phytoplankton like *Karenia* can feed on. *Karenia* can also eat the cyanobacteria as well.
3. At the same time, deep upwelling of dissolved nitrates from the bottom rise up and also fuel *Karenia*'s growth, albeit slow.
4. Wind and ocean currents blow *Karenia* toward the coasts where the waters become shallower and cyanobacteria die off, releasing again more dissolved nutrients and concentrates their populations, forming a dense bloom.
5. On the coast, *Karenia* cells may get crushed on the beaches or break releasing e exposing toxins to marine life and humans.
6. Toxins may lead to fish kills, which promote more growth of *Karenia* as the decay fish release sources of dissolved nitrogen and phosphorus.



7. If people eat local shellfish they too can become poisoned with Amnesic shellfish poisoning
8. Local outflow from agriculture and natural sources of phosphate from rivers may help the bloom to continue to persist.
9. This slide shows that HABs can be complex interactions of environmental and biological factors with many inputs and not a single direct cause.

Slide 28: This slide is used to introduce the cell as a system, akin to the nitrogen cycle. Emphasize and reinforce the cell as a system, within inputs (nutrients) and outputs, except that constituents of metabolism can go in many directions and are not always linear. Highlight for the students to see that the building blocks (N and P) are input into the cell across membranes to be transformed into new products, and that cellular pathways are not unidirectional or linear. Also show outputs, because the cells release components just as they uptake them.

Slide 29: At this point, the presentation steps outside the case study and transitions to a discussion some key points about gene expression

Slide 30: Clicker Question.

Slide 31: This slide describes eukaryotic genes and how they are fundamental to understanding what an organism is doing at the cellular level. Information about cellular activity can be seen in transcription and translation.

Slide 32: This slide asks students to sort these 5 boxes into the correct sequence. The next three slides explore the Central Dogma.

Slide 33: Clicker Question.

Slide 34: This slide describes the sequence that forms the Central Dogma of Molecular Biology and the corresponding concepts that form the Central Dogma of Genomics on the right.

Slide 35: Transition back to the case study

Slide 36. This slide is used to ask the students to generate ideas about what they would do for a genomics experiment. Give the students time to discuss as a class or small groups on the design and different aspects of the experiment including determining the control and treatments. They should also make predictions about the outcome of the experiment and what it could tell them.

Slide 37: This slide is used for Dr. Edwards to explain why this company is using the transcriptome of *Karenia brevis* to understand its biology in relation to nutrients (nitrogen and phosphorus). This should reflect some of the ideas generated during the discussion.

Slide 38: In this slide Dr. Edwards is explaining that transcription is dynamic and expressions of genes are not always on or off like a light switch. Instead, gene expression is more like a light dimmer in that lights (genes) are at least always on but at very low levels and can be up-regulated or down-regulated to give variable levels of expression. Also like the string of lights, each gene might each have its own dimmer switch or could be connected to other genes sharing the same switch.

Slide 39: In this slide, Dr. Edwards predicts what genes might be affected by nutrient limitation in *Karenia* based on previous research in other algae.

Slide 40: This slide explains the nutrient stress experiment Dr. Edwards conducted.

They grew *Karenia* under 3 different conditions, with a control and an N-limited and P-

limited condition in 1 liter of seawater. These pictures show the starting concentrations of nitrate (N) and phosphate (P) added to each culture flask at the beginning of the experiment. The stepped animation shows how the ratio of N:P is different between each flask. Previous experiments of *Karenia* suggest it uses small amounts of P over time. The N-limited ratio is 4:1, meaning that the culture will run out of N before exhausting P. The P-limited ratio is 40:1 meaning the culture will exhaust its amount of P before it runs out of N. Measurements of dissolved nutrients were taken over the time course and once each treatment exhausted its supply of N or P, they were incubated for 24 more hours to insure a stress response. These were all grown for 2 weeks, and at the end were harvested at the same time.

Slide 41: This slide explains “Big Data Collection” in a stepped animation from cultured cells to computer-summed measurements of expression. This slide illustrates the process of generating these large data collections in pictures to show that millions of bytes of information are collected and need to be analyzed.

1. First, *Karenia* is grown in flasks with seawater containing Control levels of nutrients. The stressed conditions are grown simultaneously starting at the same time, but with conditions that will stress the cells as they grow. In the case of *Karenia*, the N-limited treatment, nitrogen in the form of nitrate is added at low concentrations. The P-limited treatment has phosphate added at a low concentration and the cultures are grown in a light incubator until they run out of either nutrient by tracking the concentrations of loss in the media.
2. Samples are taken from the culture flasks, and RNA (total) is extracted from each sample separately. Then messenger RNA (only a fraction of Total RNA) is

extracted and fragmented into tiny pieces so they can be sequenced by the millions.

3. Then, the millions of fragments representing all the fragmented messages are sequenced.
4. Next, a supercomputer or high-powered desktop computer is used to reconstruct the transcriptome using all treatments and all the generated message fragments together.
5. These reconstructed messenger RNA “transcripts” represent all the possible messenger RNAs made when growing under normal conditions, and both nutrient stressed conditions. In this way, the transcriptome is library of all the cellular messages made by the cells and can be used to determine what their molecular functions are by matching to annotated protein databases. Matches in the databases are used to infer the function of the messenger and what it might be used for in the organism, based on its sequence similarity to known proteins with known functions.
6. The sequenced fragments generated from each sample are compared to known mapped the fragments in a transcriptome reference library. Mapped fragments are then counted and used to quantify the number of messenger RNAs that were made for that specific reconstructed gene.
7. The counts of the reconstructed genes, seen in the image as Blue fragments, are then compared among treatments. When gene expression is high, we expect there will be many copies of that message being made resulting in a high count.

Conversely, when gene expression is down-regulated, a low number of messages are made.

8. Computer algorithms are used to reconstruct what the transcripts looked like before being fragmented, using all conditions to make a reference library. Then sequenced fragments used for the assembly are mapped to the reference library as a condition.

Slide 42: Clicker Question.

Slide 43: This slide explains how expression of large numbers of genes are analyzed all at the same time and visualized with a “heatmap” diagram such as this. Colors correspond to gene expression being up (Blue) or (Red) down. Columns represent the same gene and rows represent the different treatments. These data show that the N-limited condition looks the most different when looking at these genes found to be significantly different between the Control and N-limited treatment. Once we find statistically different genes, we can identify what they are by comparing the sequence to a database like NCBI.

Slide 44: This slide asks students to interpret the results of an experiment comparing N-limited and P-limited conditions. This graph shows the number of differentially expressed genes between the Control and N-limited treatment and the Control versus the P-limited Treatment. In this example, the P-limited treatment results in only a few differential expressed genes but there is a large difference in the number of genes expressed between control and N-limited treatments. This suggests that P-limitation does not create a large transcriptome signal of phosphorus stress.

Slide 45: Clicker Question

Slide 46: Clicker Question.

Slide 46: This slide is used to summarize Dr. Edwards' work on what he has learned from the transcriptome of *Karenia*. Included is a pathway model of genes related to the uptake and assimilation of nitrogen (organic and inorganic) that were differentially expressed.

Slide 47: This slide can be used to summarize the main points of the case study, and includes a small graphic of the N physiology of *Karenia* based on the transcriptome. The genes depicted in the figure were differentially expressed and show evidence of nitrogen cycling within the cell.

Slide 48: Case conclusion. Ask the students what would they do next? What experiment could they test? Do they think they want to know something else?

Slide 49: Students can be given this follow up activity to do outside of class or at the very end. This slide can be used to have students seek out more information on *Karenia*, or perhaps another organism. Students could be asked to develop a brief summary article that would be published in a fictional local paper or newsletter about what they learned.

### **Answer Key to the Clicker Questions**

CQ1: B. The red tide is created by a dinoflagellate alga. Often the public thinks that it is caused by a bacterium, resulting in food poisoning (toxin production) of shellfish.

Cyanobacteria blooms are the cause of HAB events in Lake Erie and are not dinoflagellate blooms.

CQ2: A. This question serves to determine their knowledge about algae blooms in general. Most HABs are stimulated by nutrient pollution, which they might choose. However, *Karenia* is not one of those species that can be directly tied to nutrient pollution.

CQ3: General Survey Question to determine what the students might be interested or concerned about regarding the effects of red tides. This slide should aid in getting some initial thought behind why the residents might care about red tide events.

CQ4: B. This question allows students to think about what a gene is made of, and why are they important. The Central Dogma is reinforced with the idea of DNA to RNA (messenger RNA).

CQ5: B. This slide is used to demonstrate the principles of the Central Dogma of Molecular Biology, and to determine if students have misconceptions about the flow of information.

CQ6: D. This question is used to check the students understanding of how to interpret gene expression data. In this example, Gene B has roughly the same expression, and Gene C is up regulated (has more messenger RNA transcripts). The other two are not correct, because all genes are not up regulated relative to the control treatment, and all genes are not down regulated relative to the control treatment.

CQ7: A. This question is used to demonstrate that not all treatments have the same effects, when comparing the treatment over the control (Replete). Notice in the Replete versus N-limited there are several genes both up and down regulated, while the P-limited produced very few differences. This could be because the P-limited treatment

was not truly depleted of internal storage (vacuole) within the cells, or it is not transcriptionally responsive to P-limitation.

CQ8: C. This question is used to conclude the overall results of the transcriptome data collected by BioMarkerX. The correct answer is C, because we can only say that some genes appear to be responsive to N-limitation and not P-limitation.



## References

- American Association for the Advancement of Science. 2011. Vision and Change in Undergraduate Biology Education: A Call to Action, Final Report., Washington, DC.
- Backer, L. C. 2009. Impacts of Florida Red Tides on Coastal Communities. *Harmful Algae* **8**:618-622.
- Brand, L. E., and A. Compton. 2007. Long-term Increase in *Karenia brevis* Abundance along the Southwest Florida Coast. *Harmful Algae* **6**:232-252.
- Brand, L. E., L. Campbell, and E. Bresnan. 2012. *Karenia*: The Biology and Ecology of a Toxic Genus. *Harmful Algae* **14**:156-178.
- Crick, F. H. 1958. On Protein Synthesis. *Symposia of the Society for Experimental Biology* **12**:138-163.
- Fauquier, D. A., L. J. Flewelling, J. Maucher, C. A. Manire, V. Socha, M. J. Kinsel, B. A. Stacy, M. Henry, J. Gannon, J. S. Ramsdell, and J. H. Landsberg. 2013. Brevetoxin in Blood, Biological Fluids, and Tissues of Sea Turtles Naturally Exposed to *Karenia brevis* Blooms in Central West Florida. *Journal of Zoo and Wildlife* **44**:364-375.
- Herreid, C. F. 2006. "Clicker" Cases: Introducing Case Study Teaching Into Large Classrooms. *Journal of College Science Teaching* **36**:43-47.
- Herreid, C. F. 2007. *Start with a Story: The Case Study Method of Teaching College Science*. NSTA press. Arlington, Virginia
- Herreid, C. F., N. A. Schiller, and K. F. Herreid. 2014. *Science Stories You Can Count On: 51 Case Studies with Quantitative Reasoning in Biology*. NSTA Press. Arlington, Virginia
- Kirkpatrick, B., L. E. Fleming, D. Squicciarini, L. C. Backer, R. Clark, W. Abraham, J. Benson, Y. S. Cheng, D. Johnson, R. Pierce, J. Zaias, G. D. Bossart, and D. G. Baden. 2004. Literature Review of Florida Red Tide: Implications for Human Health Effects. *Harmful Algae* **3**:99-115.
- Landsberg, J. H. H., L. J. J. Flewelling, and J. Naar. 2009. *Karenia brevis* Red Tides, Brevetoxins in the Food Web, and Impacts on Natural Resources: Decadal Advancements. *Harmful Algae* **8**:598-607.

- Li, Z., B. Garrison, S. G. Ullmann, B. Kirkpatrick, L. E. Fleming, and P. Hoagland. 2015. Risk in Daily Newspaper Coverage of Red Tide Blooms in Southwest Florida. *Applied Environmental Education and Communication* 14:167-177.
- NCCSTS. 2017. National Center for Case Study Teaching in Science. [Available from: <http://sciencecases.lib.buffalo.edu/cs/>] Accessed April 7, 2017.
- Nierenberg, K., M. M. Byrne, L. E. Fleming, W. Stephan, A. Reich, L. C. Backer, E. Tanga, D. R. Dalpra, and B. Kirkpatrick. 2010. Florida Red Tide Perception: Residents Versus Tourists. *Harmful Algae* 9:600-606.
- Pevsner, J. 2015. *Bioinformatics and Functional Genomics*. John Wiley & Sons. Chichester, West Sussex.
- Sipler, R. E., L. R. McGuinness, G. J. Kirkpatrick, L. J. Kerkhof, and O. M. Schofield. 2014. Bacteriocidal Effects of Brevetoxin on Natural Microbial Communities. *Harmful Algae* 38:101-109.
- Steidinger, K. A. 2009. Historical Perspective on *Karenia brevis* Red Tide Research in the Gulf of Mexico. *Harmful Algae* 8:549-561.
- Udovic, D., D. Morris, A. Dickman, J. Postlethwait, and P. Wetherwax. 2002. Workshop Biology: Demonstrating the Effectiveness of Active Learning in an Introductory Biology Course. *BioScience* 52:272-281.
- Vargo, G. A. 2009. A Brief Summary of the Physiology and Ecology of *Karenia brevis* Davis (G. Hansen and Moestrup *Comb. Nov.*) Red Tides on the West Florida Shelf and of Hypotheses Posed for Their Initiation, Growth, Maintenance, and Termination. *Harmful Algae* 8:573-584.
- Vargo, G. A., C. A. Heil, K. A. Fanning, L. K. Dixon, M. B. Neely, K. Lester, D. Ault, S. Murasko, J. Havens, J. Walsh, and S. Bell. 2008. Nutrient Availability in Support of *Karenia brevis* Blooms on the Central West Florida Shelf: What Keeps *Karenia* Blooming? *Continental Shelf Research* 28:73-98.

## Slide Image Credits

### Slide 1

Image credit:

Description: Beach at Wulfert, Sanibel Island, Florida, looking north towards Captiva Island

Source: Wikimedia Commons,

[https://commons.wikimedia.org/wiki/File%3AWulfert\\_Sanibel.JPG](https://commons.wikimedia.org/wiki/File%3AWulfert_Sanibel.JPG)

Author: Bradeos Graphon (Own work) [Public domain],

Clearance: GNU Free Documentation License

### Slide 2

Image credit:

Description: UF Keene-Flint Classroom Desks Whiteboard Windows

Source: Flickr, <https://www.flickr.com/photos/csessums/4423279597>

Author: Christopher Sessums

Clearance: Attribution 2.0 Generic (CC BY 2.0)

Modification: Cropped and desaturated image

### Slide 4

Image credit:

Description: Survey Opinion Research

Source: Pixabay, <https://pixabay.com/en/survey-opinion-research-voting-fill-1594962>

Author: andibreit, Andreas Breitling

Clearance: CC0 Public Domain, Free for commercial use, No attribution required

### Slide 5

Image credit:

Description: Scanned image of standard 3x5 notecard / index card. Notecards rule.

Source: Wikimedia Commons, <https://commons.wikimedia.org/wiki/File:Notecard.jpg>

Author: E.m.fields - fields.emmett@gmail.com

Clearance: Creative Commons Attribution 3.0 Unported

### Slide 10

Image credit:

Description: Red Tide off the Scripps Institution of Oceanography Pier, La Jolla California. Released into the Public Domain, August 2005. P. Alejandro Díaz From the English Wikipedia

Source: <https://commons.wikimedia.org/w/index.php?curid=396343>

Clearance: Public Domain

### Slide 12:

Image credit:

Top Left

Description: External view, SEM of *Stephanodiscus*

Source: This image is derived from the Professor Frank Round Image Archive at the Royal Botanic Garden Edinburgh, Identified By David Mann,  
<http://tolweb.org/onlinecontributors/app;jsessionid=B0342BE169836165EBA7E2D7D77E8E35?page=ViewImageData&service=external&sp=29559>  
Author: Frank E. Round, David G. Mann  
Clearance: Creative Commons Attribution-Non Commercial License - Version 3.0.

#### Top Right

Description: Scanning electron microscope image of dinoflagellates  
Source: Wikimedia Commons,  
[https://commons.wikimedia.org/wiki/File:CSIRO\\_ScienceImage\\_7609\\_SEM\\_dinoflagellate.jpg](https://commons.wikimedia.org/wiki/File:CSIRO_ScienceImage_7609_SEM_dinoflagellate.jpg), <http://www.scienceimage.csiro.au/image/7609>  
Author: CSIRO, CSIRO  
Clearance: Creative Commons Attribution 3.0 Unported  
Modification: Turned 90 degrees, cropped and colored

#### Middle Left

Description: *Coccolithus pelagicus ssp. braarudii*.  
*Coccolithus pelagicus*; coccosphere. Location: N. Atlantic; 48N, -20E; 3200m  
image sets Taxa - Coccolithophores - London NHM/Roscoff  
Source: Wikimedia Commons, <http://planktonnet.awi.de/>,  
[https://upload.wikimedia.org/wikipedia/commons/3/33/Coccolithus\\_pelagicus.jpg](https://upload.wikimedia.org/wikipedia/commons/3/33/Coccolithus_pelagicus.jpg)  
Author: Richard Lampitt, Jeremy Young, The Natural History Museum, London  
Clearance: Creative Commons Attribution 2.5 Generic

#### Middle Right

Description: A single cell of the freshwater algae species *Synura petersenii*. False color image created using SEM.

Source: Wikimedia Commons,  
[https://commons.wikimedia.org/wiki/File:Synura\\_petersenii.png](https://commons.wikimedia.org/wiki/File:Synura_petersenii.png)  
Author: By Drew Lindow (Own work)  
Clearance: [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)], via

#### Bottom

Description: *Spirulina sp.* (scanning electron micrograph)  
Source: Flickr, <https://www.flickr.com/photos/myfwc/8634661651>  
Author: FWC Fish and Wildlife Research Institute  
Clearance: Attribution-Noncommercial-NoDerivs 2.0 Generic (CC BY-NC-ND 2.0)  
Modification: Colorized green with Powerpoint and cropped

#### Slide 13

##### Top Left

Image credit:

Description: *Mesodinium rubrum* Bloom, Sanibel, Lee County, November 12, 2010

Source: Flickr, <https://www.flickr.com/photos/myfwc/6442808529/in/album-72157628250102201/>  
Author: FWC Fish and Wildlife Research Institute  
Clearance: Attribution-Non Commercial-No Derivs 2.0 Generic (CC BY-NC-ND 2.0)

Top Right

Image credit:

Description: *Karenia brevis* Bloom, Offshore Pinellas County, September 16, 2005

Source: Flickr, <https://www.flickr.com/photos/myfwc/6442820071/in/album-72157628250102201/>

Author: FWC Fish and Wildlife Research Institute

Clearance: Attribution-Non Commercial-No Derivs 2.0 Generic (CC BY-NC-ND 2.0)

Bottom Left:

Image credit:

Description: *Fibrocapsa japonica* Bloom, Tarpon Road Beach, Lee County, July 21, 2011

Source: Flickr, <https://www.flickr.com/photos/myfwc/6442809143/in/album-72157628250102201/>

Author: FWC Fish and Wildlife Research Institute

Clearance: Attribution-Non Commercial-No Derivs 2.0 Generic (CC BY-NC-ND 2.0)

Bottom Middle:

Image credit:

Description: Pedinophyceae Bloom, Indian River Lagoon, Home Point, Brevard County, June 3, 2011

Source: <https://www.flickr.com/photos/myfwc/6442818691/in/album-72157628250102201/>

Author: FWC Fish and Wildlife Research Institute

Clearance: Attribution-Non Commercial-No Derivs 2.0 Generic (CC BY-NC-ND 2.0)

Bottom Right:

Image credit:

Description: *Chattonella subsalsa* Bloom, Bowman's Beach, Lee County, August 4, 2011

Source: <https://www.flickr.com/photos/myfwc/6442825193/in/album-72157628250102201/>

Clearance: Attribution-Non Commercial-No Derivs 2.0 Generic (CC BY-NC-ND 2.0)

### **Slide 15**

Image credit:

Description: The green scum shown in this image is the worst algae bloom Lake Erie has experienced in decades. Vibrant green filaments extend out from the northern shore. Image captured by the Landsat-5 satellite. Data provided courtesy of the United States Geological Survey.

Source: Wikimedia Commons;  
[https://upload.wikimedia.org/wikipedia/commons/c/cb/Toxic\\_Algae\\_Bloom\\_in\\_Lake\\_Erie.jpg](https://upload.wikimedia.org/wikipedia/commons/c/cb/Toxic_Algae_Bloom_in_Lake_Erie.jpg)  
[https://commons.wikimedia.org/wiki/File%3AToxic\\_Algae\\_Bloom\\_in\\_Lake\\_Erie.jpg](https://commons.wikimedia.org/wiki/File%3AToxic_Algae_Bloom_in_Lake_Erie.jpg)  
Author: Jesse Allen and Robert Simmon (NASA Earth Observatory) [Public domain]  
Clearance: GNU Free Documentation License

### Slide 16

Top Inset

Image credit:

Description: View of runoff, transporting nonpoint source pollution, from a farm field in Iowa during a rainstorm.

Caption: "Topsoil as well as farm fertilizers and other potential pollutants run off unprotected farm fields when heavy rains occur."

Source: Wikimedia Commons;

[https://commons.wikimedia.org/wiki/File%3ARunoff\\_of\\_soil\\_%26\\_fertilizer.jpg](https://commons.wikimedia.org/wiki/File%3ARunoff_of_soil_%26_fertilizer.jpg)

Author: Lynn Betts, photographer [Public domain],

Clearance: CC0 Public Domain

Bottom

Image credit:

Description: Satellite image and illustration of a dead zone in the southern U.S.

Source: Wikimedia Commons;

[https://commons.wikimedia.org/wiki/File:Gulf\\_dead\\_zone.jpg](https://commons.wikimedia.org/wiki/File:Gulf_dead_zone.jpg)

Author: NOAA (NOAA)

Clearance: CC0 Public Domain

### Slide 18

Left Top

Image credit:

Description: A, a group of neurotoxins isolated from the marine dinoflagellate *Karenia brevis* (formerly *Gymnodinium breve*)

- Brevetoxin-1 (PbTx-1) R = -CH<sub>2</sub>C(=CH<sub>2</sub>)CHO
- Brevetoxin-7 (PbTx-7) R = -CH<sub>2</sub>C(=CH<sub>2</sub>)CH<sub>2</sub>OH
- Brevetoxin-10 (PbTx-10) R = -CH<sub>2</sub>CH(-CH<sub>3</sub>)CH<sub>2</sub>OH

Source: Wikimedia Commons,

<https://commons.wikimedia.org/w/index.php?curid=1169084>

Author: created by Minutemen using BKchem 0.11.4 & Inkscape 0.44 - Own work,  
Public Domain

Clearance: CC0 Public Domain

Left Bottom

Image credit:

Description: B, a group of neurotoxins isolated from the marine dinoflagellate *Karenia brevis* (formerly *Gymnodinium breve*)

- Brevetoxin-2 (PbTx-2) R = -CH<sub>2</sub>C(=CH<sub>2</sub>)CHO

- Brevetoxin-3 (PbTx-3) R = -CH<sub>2</sub>C(=CH<sub>2</sub>)CH<sub>2</sub>OH
- Brevetoxin-8 (PbTx-8) R = -CH<sub>2</sub>COCH<sub>2</sub>Cl
- Brevetoxin-9 (PbTx-9) R = -CH<sub>2</sub>CH(CH<sub>3</sub>)CH<sub>2</sub>OH
- Other B-type Brevetoxins:
- Brevetoxin-5 (PbTx-5) R = -CH<sub>2</sub>C(=CH<sub>2</sub>)CH<sub>2</sub>OH, OH in position 37 is acetylated
- Brevetoxin-6 (PbTx-6) like PbTx-2, epoxide in position 27-28

Source: Wikimedia Commons,

<https://commons.wikimedia.org/w/index.php?curid=1169084>

Author: created by Minutemen using BKchem 0.11.4 & Inkscape 0.44 - Own work, Public Domain

Clearance: CC0 Public Domain

### Right

Description: *Karenia brevis* (C.C.Davis) Gert Hansen & Ø.Moestrup, 2000 Checked: verified by a taxonomic editor checked Hansen, Gert 2010-10-12

Source: WoRMS Photogallery,

<http://www.marinespecies.org/photogallery.php?album=1033&pic=21985>

Author: Hansen, Gert

Clearance: Creative Commons Attribution-Noncommercial-Share Alike 4.0 License

### Slide 19

Image credit: Based on illustration in “RED TIDES ON WEST FLORIDA SHELF”

Description: Brochure, added red circles to include additional data of *Karenia* presences on the west Florida shelf.

Source: <http://myfwc.com/media/2687915/brochure.pdf>

Author: Brianne Walsh, modified by Joshua Cooper

Clearance: Creative Commons Attribution-Noncommercial-Share Alike 4.0 License

### Slide 20

Image credit:

Description: Life of *Karenia brevis*

Source: Own work using Inkscape 0.91

Author: Joshua T. Cooper

Clearance: GNU Free Documentation License, Version 1.2 or later

### Slide 22

Image credit:

Description: Adapted figure 8 from Hardison DR, Sunda WG, Shea D, Litaker RW (2013) Increased Toxicity of *Karenia brevis* during Phosphate Limited Growth:

Ecological and Evolutionary Implications. PLOS ONE 8(3): e58545. doi: 10.1371/journal.pone.0058545

Source: Hardison DR, Sunda WG, Shea D, Litaker RW (2013) Increased Toxicity of *Karenia brevis* during Phosphate Limited Growth: Ecological and Evolutionary Implications. PLOS ONE 8(3): e58545. doi: 10.1371/journal.pone.0058545

Author: Joshua T. Cooper adapted from the above paper found in PLOS

Clearance: Public Domain

**Slide 27**

Image credit:

Description: Stepped animated layers made in Inkscape 0.91, of the complex life history and knowledge of *Karenia brevis* make with 50+ years of research

Source: Own work using Inkscape 0.91

Author: Joshua T. Cooper

Clearance: GNU Free Documentation License, Version 1.2 or later

**Slide 28**

Image credit:

Description: Diagram of a general algae cell made in PowerPoint

Source: Own work

Author: Joshua T. Cooper

Clearance: GNU Free Documentation License, Version 1.2 or later

**Slide 31**

Image credit:

Description: Diagram of a general gene with exons and introns in PowerPoint

Source: Own work

Author: Joshua T. Cooper

Clearance: GNU Free Documentation License, Version 1.2 or later

**Slide 38**

Image credit:

Description:

Source: Pixabay, <https://pixabay.com/en/lightbulb-lighting-night-bulb-1285110/>

Author:

Clearance: CC0 Public Domain Free for commercial use, No attribution required

**Slide 40**

Image credit:

Description: Diagram of a 3 treatments using PowerPoint

Source: Own work

Author: Joshua T. Cooper

Clearance: GNU Free Documentation License, Version 1.2 or later

**Slide 41**

Top left to Bottom Left

Image credit:

Description: Illustrations made in PowerPoint

Source: Own work

Author: Joshua T. Cooper

Clearance: GNU Free Documentation License, Version 1.2 or later



#### Bottom Center

Source: Image adapted from Ouellette, Francis; Griffith, Malachi; Walker, Jason R.; Spies, Nicholas C.; Ainscough, Benjamin J.; Griffith, Obi L. (2015). "Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud". PLOS Computational Biology. 11 (8): e1004393. doi:10.1371/journal.pcbi.1004393. ISSN 1553-7358.

Author: Ouellette, Francis; Griffith, Malachi; Walker, Jason R.; Spies, Nicholas C.; Ainscough, Benjamin J.; Griffith, Obi L

Clearance:

#### Bottom Right

Image credit:

Description: Server Computer Clipart

Source: ClipartFest,

[https://img.clipartfest.com/d21d7b6c8260c57f6502fe6cd7848d07\\_server20computer20clipart-clipart-server\\_464-800.png](https://img.clipartfest.com/d21d7b6c8260c57f6502fe6cd7848d07_server20computer20clipart-clipart-server_464-800.png)

Author: Miss Lynn Henderson from Irving

Clearance: Public Domain

#### Bottom Right to Top Right

Image credit:

Description: Illustrations make in PowerPoint

Source: Own work

Author: Joshua T. Cooper

Clearance: GNU Free Documentation License, Version 1.2 or later

#### **Slide 42**

Image credit:

Description: Illustrations made in PowerPoint

Source: Own work

Author: Joshua T. Cooper

Clearance: GNU Free Documentation License, Version 1.2 or later

#### **Slide 43**

Image credit:

Description: Illustrations make in PowerPoint and graphic in R (R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

) and using pheatmap (Raivo Kolde (2015). pheatmap: Pretty Heatmaps. R package version 1.0.8.

Source: Own work

Author: Joshua T. Cooper

Clearance: GNU Free Documentation License, Version 1.2 or later

#### **Slide 44**

Image credit:

Description: Illustrations made in Excel, showing the number of differentially expressed genes in two conditions versus the control.

Source: Own work

Author: Joshua T. Cooper

Clearance: GNU Free Documentation License, Version 1.2 or later

#### **Slide 45**

Image credit:

Description: Illustrations made in Excel, showing the number of differentially expressed genes in two conditions versus the control.

Source: Own work

Author: Joshua T. Cooper

Clearance: GNU Free Documentation License, Version 1.2 or later

#### **Slide 47**

Image credit:

Description: Illustrations made in Inkscape 0.91 depicting data and *Karenia brevis* cell

Source: Own work

Author: Joshua T. Cooper

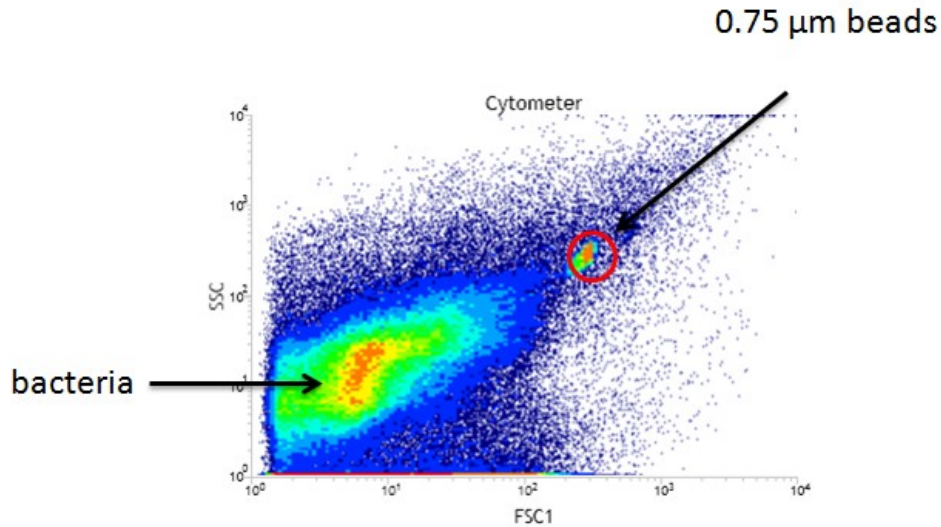
Clearance: GNU Free Documentation License, Version 1.2 or later

## Appendices and Supplemental Information

### Chapter 1: Urea Uptake and Carbon Fixation by Marine Pelagic Bacteria and Archaea During the Arctic Summer and Winter Seasons.

#### SUPPLEMENTAL FIGURES

**Figure S1.**



**Figure S1.** Flow cytometric analysis of DAPI stained seawater collected at our sampling site during the arctic winter sampling. The x-axis shows forward scatter, while the y-axis indicates side scatter of particles. Included in the analysis were 0.75  $\mu\text{m}$  standard beads to assess the average size of particles in seawater. Filters used for rate measurements (GF/F) retained cells larger than 0.7  $\mu\text{m}$ . Flow cytometry indicates that most bacterial and archaeal cells were smaller than the 0.7  $\mu\text{m}$  size cutoff of GF/F filters. The bulk of requisite cells would however be captured by 0.22  $\mu\text{m}$  filters used for SIP analysis.

**Chapter 2: Transcriptome Analysis of *Scrippsiella trochoidea* CCMP 3099 Reveals Physiological Changes Related to Nitrate Depletion**

Supplementary Table 1. Nutrient concentrations at time of sampling.

|                      | <b>Replete</b> | <b>Nitrogen-Limited</b> | <b>Phosphorus-Limited</b> |
|----------------------|----------------|-------------------------|---------------------------|
| NO <sub>x</sub> [μM] | 1759.6 (-4.5)  | BD                      | 1020.3 (-8.4)             |
| PO <sub>4</sub> [μM] | 11.3 (2.0)     | 18.7 (0.5)              | 4.4 (1.9)                 |

Mean + sd, BD (Below detection)

Supplementary Table 2. Overall Result of the Illumina Transcriptome Sequencing

| Raw Data   | Library            | Data (Bp)  | GC% | Mb    |
|------------|--------------------|------------|-----|-------|
| MMETSP0270 | Replete            | 36,147,274 | 58  | 36.15 |
| MMETSP0271 | Nitrogen-limited   | 36,131,119 | 58  | 36.13 |
| MMETSP0272 | Phosphorus-limited | 30,815,547 | 58  | 30.82 |
| Trimmed    | Library            | Data (Bp)  | GC% | Mb    |
| MMETSP0270 | Replete            | 35,788,801 | 58  | 35.79 |
| MMETSP0271 | Nitrogen-limited   | 35,770,337 | 58  | 35.77 |
| MMETSP0272 | Phosphorus-limited | 30,507,209 | 58  | 30.51 |

Supplementary Table 3. Summary of the transcriptome reduction procedure using CAP3 and CD-HIT-EST of *Scrippsiella trochoidea*, to collapse potential assembly artifacts using the Multi-*k*-mer assembly approach.

|               | <b>Assembly Name</b> |                        |  |
|---------------|----------------------|------------------------|--|
|               | M-k<br>Assembly      | M-k Assembly<br>+ CAP3 | <b>M-k Assembly + CAP3 +<br/>CDHIT-EST</b> |
| Total         |                      |                        |  |
| Length (bp)   | 201,998,239          | 132,039,060            | <b>125,022,445</b>                         |
| Total         |                      |                        |  |
| Contigs > 200 | 205,934              | 118,132                | <b>107,473</b>                             |
| Median        |                      |                        |  |
| Contig size   | 686                  | 821                    | <b>884</b>                                 |
| Mean          |                      |                        |  |
| Contig size   | 980                  | 1,117                  | <b>1,163</b>                               |
| Maximum       |                      |                        |  |
| Contig size   | 25,590               | 26,110                 | <b>26,110</b>                              |
| N50 Contig    | 43,657               | 25,143                 | <b>23,394</b>                              |
| N50 Length    | 1,440                | 1,678                  | <b>1,714</b>                               |
| N90 Length    | 443                  | 516                    | <b>555</b>                                 |

Supplementary Table. 4 Annotation Results of the *S. trochoidea* transcriptome

| <b>Annotation Pipeline</b>                | <b># Annotated Contigs</b> | <b>% Annotated Contigs</b> |
|---|----------------------------|----------------------------|
| NCBI-NR database                          | 44,055                     | 41.00%                     |
| Uniprot-SwissProt+TREMBL                  | 43,796                     | 40.80%                     |
|   |                            |                            |
| Gene-Ontology Hits                        | 34,480                     | 40.40%                     |
| GO:Cellular Component                     | 23,159                     | 21.60%                     |
| GO:Biological Component                   | 30,453                     | 28.40%                     |
| GO:Molecular Component                    | 24,415                     | 22.70%                     |
|   |                            |                            |
| KEGG Matches                              | 29,489                     | 27.50%                     |
| KEGG:GO                                   | 16,148                     | 15.00%                     |
| KOG Matches                               | 23,462                     | 21.80%                     |
| Pfam-A matches                            | 36,749                     | 34.20%                     |
| TIGRFAM matches                           | 3,098                      | 2.90%                      |
|   |                            |                            |
| <b>Total Descriptions after filtering</b> | <b>43,785</b>              | <b>40.70%</b>              |
| <b>Total Transcriptome</b>                | <b>107,417</b>             |                            |

### Chapter 3: Trophic Differences Reflected in the Elemental Stoichiometry From Transcriptomes and Proteomes of Eukaryotic Phytoplankton

#### SUPPLEMENTAL MATERIALS:

Additional proteomes were collected from the following public resources:

1. Wellcome-Sanger Trust Institute (*Bodo saltans*)  
*Bathycoccus parinos* RCC1105  
(<http://bioinformatics.psb.ugent.be/genomes/view/Bathycoccus-prasinos/>;  
accessed 07/26/15)
2. *Bigelowiella natans* GCA 000320545-1-27 from Joint Genome Institute (Curtis et al. 2012)
3. *Cyanophora paradoxa* MAKER gene predictions-022111  
(<http://cyanophora.rutgers.edu/cyanophora/>; accessed 07/26/15)
4. *Nannochloropsis oceanica* CCMP1779 Michigan  
(<https://bmb.natsci.msu.edu/faculty/christoph-benning/nannochloropsis-oceanica-ccmp1779/>; accessed 07/26/15)
5. *Pyropia yezoensis*  
([http://nrifs.fra.affrc.go.jp/ResearchCenter/5\\_AG/genomes/nori/](http://nrifs.fra.affrc.go.jp/ResearchCenter/5_AG/genomes/nori/); (Nakamura et al. 2013))
6. *Symbiodinium mintum* symbB-v1 2  
([http://marinegenomics.oist.jp/symb/viewer/download?project\\_id=21](http://marinegenomics.oist.jp/symb/viewer/download?project_id=21;);  
(Shoguchi et al. 2013)).

Supplemental Table 1. Collected cell size data from the MMETSP dataset, primary sources as listed or from the Culture Collections. Provasoli-Guillard National Center for Marine Algae and Microbiota (NCMA at Bigelow formerly CCMP), SAG (Sammlung von Algenkulturen der Universität Göttingen -Culture Collection of Algae at Göttingen University), Roscoff Culture Collection (RCC), Microbial Culture Collection at the National Institute for Environmental Studies (NIES), Canadian Center for the Culture of Microorganisms (CCCM), Culture Collection of Algae at Austin (UTEX), Culture Collection of Algae and Protozoa (CCAP), The Marine Biological Association Culture collection (Plymouth Culture Collection), Council for Scientific & Industrial Research Organization, Australia (CSIRO) North East Pacific Culture Collection, Canada (NEPCC).



| MMETSP ID  | Genus                  | Species              | Strain        | Min Length | Max Length | Min Width | Max Width | Source |
|------------|------------------------|----------------------|---------------|------------|------------|-----------|-----------|--------|
| MMETSP0105 | <i>Acanthoeca-like</i> | <i>sp.</i>           | 10tr          | 17.9       | 17.9       | 17.9      | 17.9      | [26]   |
| MMETSP0106 | <i>Acanthoeca-like</i> | <i>sp.</i>           | 10tr          | 17.9       | 17.9       | 17.9      | 17.9      | [26]   |
| MMETSP0223 | <i>Akashiwo</i>        | <i>sanguinea</i>     | CCCM885       | -          | -          | -         | -         | [7]    |
| MMETSP1436 | <i>Alexandrium</i>     | <i>andersonii</i>    | CCMP2222      | 20         | 24         | 16        | 20        | [5]    |
| MMETSP0790 | <i>Alexandrium</i>     | <i>catenella</i>     | OF101         | -          | -          | -         | -         | [NA]   |
| MMETSP0196 | <i>Alexandrium</i>     | <i>fundyense</i>     | CCMP1719      | 30         | 40         | 30        | 40        | [5]    |
| MMETSP0197 | <i>Alexandrium</i>     | <i>fundyense</i>     | CCMP1719      | 30         | 40         | 30        | 40        | [5]    |
| MMETSP0347 | <i>Alexandrium</i>     | <i>fundyense</i>     | CCMP1719      | 30         | 40         | 30        | 40        | [5]    |
| MMETSP0661 | <i>Alexandrium</i>     | <i>margalefi</i>     | AMGDE01CS-322 | -          | -          | -         | -         | [NA]   |
| MMETSP0328 | <i>Alexandrium</i>     | <i>minutum</i>       | CCMP113       | 24         | 32         | 20        | 28        | [5]    |
| MMETSP0093 | <i>Alexandrium</i>     | <i>monilatatum</i>   | CCMP3105      | -          | -          | -         | -         | [NA]   |
| MMETSP0095 | <i>Alexandrium</i>     | <i>monilatatum</i>   | CCMP3105      | -          | -          | -         | -         | [NA]   |
| MMETSP0378 | <i>Alexandrium</i>     | <i>tamarense</i>     | CCMP1771      | 35         | 40         | 38        | 40        | [5]    |
| MMETSP0288 | <i>Alveolata</i>       | <i>sp.</i>           | CCMP3155      | 4          | 46         | 4         | 46        | [5]    |
| MMETSP1384 | <i>Ammonia</i>         | <i>sp.</i>           | Ammonia       | -          | -          | -         | -         | [NA]   |
| MMETSP0795 | <i>Amoebophrya</i>     | <i>sp.</i>           | Ameob2        | -          | -          | -         | -         | [NA]   |
| MMETSP0258 | <i>Amphidinium</i>     | <i>carterae</i>      | CCMP1314      | 12         | 18         | 9         | 13        | [5]    |
| MMETSP0398 | <i>Amphidinium</i>     | <i>carterae</i>      | CCMP1314      | 12         | 18         | 9         | 13        | [5]    |
| MMETSP0399 | <i>Amphidinium</i>     | <i>carterae</i>      | CCMP1314      | 12         | 18         | 9         | 13        | [5]    |
| MMETSP0689 | <i>Amphidinium</i>     | <i>massartii</i>     | CS-259        | -          | -          | -         | -         | [NA]   |
| MMETSP1065 | <i>Amphiprora</i>      | <i>paludosa</i>      | CCMP125       | 9          | 21         | 6         | 9         | [5]    |
| MMETSP0724 | <i>Amphiprora</i>      | <i>sp.</i>           | CCMP467       | 35         | 75         | 15        | 25        | [5]    |
| MMETSP0316 | <i>Amphora</i>         | <i>coffeaeformis</i> | CCMP127       | 15         | 22         | 4         | 8         | [5]    |
| MMETSP1018 | <i>Anophryoides</i>    | <i>haemophila</i>    | AH6           | 25         | 34         | 14        | 23        | [8]    |

|            |                         |                        |              |     |     |    |    |    |    |      |
|------------|-------------------------|------------------------|--------------|-----|-----|----|----|----|----|------|
| MMETSP0954 | <i>Aplanochytrium</i>   | <i>sp.</i>             | PBS07        | 1   | 5   | 1  | 5  | 1  | 5  | [27] |
| MMETSP1346 | <i>Aplanochytrium</i>   | <i>stocchinoi</i>      | GSBS06       | 1   | 4   | 1  | 4  | 1  | 4  | [40] |
| MMETSP0125 | <i>Ariasterostoma</i>   | <i>sp.</i>             | ATCC50986    | 9   | 23  | 4  | 11 | 4  | 11 | [47] |
| MMETSP0705 | <i>Asterionellopsis</i> | <i>glacialis</i>       | CCMP134      | 16  | 22  | 6  | 12 | 6  | 12 | [5]  |
| MMETSP1394 | <i>Asterionellopsis</i> | <i>glacialis</i>       | CCMP1581     | 20  | 24  | 6  | 8  | 6  | 8  | [5]  |
| MMETSP0713 | <i>Asterionellopsis</i> | <i>glacialis</i>       | no strain    | 30  | 150 | 8  | 12 | 8  | 12 | [57] |
| MMETSP0418 | <i>Astrosyne</i>        | <i>radiata</i>         | 13v08-1A     | 30  | 63  | 5  | 7  | 5  | 7  | [4]  |
| MMETSP1449 | <i>Attheya</i>          | <i>septentrionalis</i> | CCMP2084     | 6   | 14  | 4  | 7  | 4  | 7  | [5]  |
| MMETSP1064 | <i>Aulacoseira</i>      | <i>subarctica</i>      | CCAP 1002/5  | -   | -   | -  | -  | -  | -  | [9]  |
| MMETSP0958 | <i>Aurantiochytrium</i> | <i>limacinum</i>       | ATCCMYA-1381 | 4   | 20  | 4  | 20 | 4  | 20 | [3]  |
| MMETSP0914 | <i>Aureococcus</i>      | <i>anophagefferens</i> | CCMP1850     | 2   | 4   | 2  | 4  | 2  | 4  | [5]  |
| MMETSP0890 | <i>Aureoumbra</i>       | <i>lagunensis</i>      | CCMP1510     | 4   | 7   | 4  | 7  | 4  | 7  | [5]  |
| MMETSP1036 | <i>Azadinium</i>        | <i>spinosum</i>        | 3D9          | -   | -   | -  | -  | -  | -  | [NA] |
| MMETSP1399 | <i>Bathycoccus</i>      | <i>prasinus</i>        | CCMP1898     | 2   | 2   | 2  | 2  | 2  | 2  | [5]  |
| MMETSP1460 | <i>Bathycoccus</i>      | <i>prasinus</i>        | RCC716       | 1   | 1   | 1  | 1  | 1  | 1  | [53] |
| MMETSP0115 | <i>Bicosoecid</i>       | <i>sp.</i>             | msl          | -   | -   | -  | -  | -  | -  | [NA] |
| MMETSP1359 | <i>Bigelowiella</i>     | <i>longifila</i>       | CCMP242      | 8   | 20  | 8  | 20 | 8  | 20 | [5]  |
| MMETSP1358 | <i>Bigelowiella</i>     | <i>natans</i>          | CCMP1242     | 4   | 8   | 4  | 8  | 4  | 8  | [5]  |
| MMETSP1055 | <i>Bigelowiella</i>     | <i>natans</i>          | CCMP1258.1   | 5   | 7   | 5  | 7  | 5  | 7  | [5]  |
| MMETSP1054 | <i>Bigelowiella</i>     | <i>natans</i>          | CCMP1259     | 4   | 8   | 4  | 8  | 4  | 8  | [5]  |
| MMETSP0045 | <i>Bigelowiella</i>     | <i>natans</i>          | CCMP2755     | 4   | 6   | 4  | 6  | 4  | 6  | [5]  |
| MMETSP1052 | <i>Bigelowiella</i>     | <i>natans</i>          | CCMP623      | 4   | 5   | 4  | 5  | 4  | 5  | [5]  |
| MMETSP1395 | <i>Blepharisma</i>      | <i>japonicum</i>       | Stock R1072  | 293 | 315 | 76 | 96 | 76 | 96 | [14] |
| MMETSP0785 | <i>Bolidomonas</i>      | <i>pacifica</i>        | CCMP1866     | 2   | 3   | 2  | 3  | 2  | 3  | [5]  |
| MMETSP1319 | <i>Bolidomonas</i>      | <i>pacifica</i>        | RCC208       | 2   | 2   | 2  | 2  | 2  | 2  | [53] |

|            |                       |                       |                   |     |     |    |    |      |
|------------|-----------------------|-----------------------|-------------------|-----|-----|----|----|------|
| MMETSP1321 | <i>Bolidomonas</i>    | <i>sp.</i>            | RCC1657           | 2   | 2   | 2  | 2  | [53] |
| MMETSP1320 | <i>Bolidomonas</i>    | <i>sp.</i>            | RCC2347           | 4   | 4   | 4  | 4  | [53] |
| MMETSP1462 | <i>Brandtodinium</i>  | <i>nutriculum</i>     | RCC3387           | 10  | 10  | 10 | 10 | [53] |
| MMETSP0942 | <i>Cafeteria</i>      | <i>roenbergensis</i>  | E4-10             | 5   | 5   | 5  | 5  | [13] |
| MMETSP1104 | <i>Cafeteria</i>      | <i>sp.</i>            | Caron Lab Isolate | 5   | 5   | 5  | 5  | [3]  |
| MMETSP1334 | <i>Calcidiscus</i>    | <i>leptoporus</i>     | RCC1130           | 8   | 14  | 8  | 12 | [53] |
| MMETSP1074 | <i>Ceratium</i>       | <i>fuscus</i>         | PA161109          | 150 | 230 | 15 | 30 | [48] |
| MMETSP0088 | <i>Chaetoceros</i>    | <i>affinis</i>        | CCMP159           | 19  | 25  | 5  | 10 | [5]  |
| MMETSP1435 | <i>Chaetoceros</i>    | <i>brevis</i>         | CCMP164           | 0   | 0   | 0  | 0  | [5]  |
| MMETSP1336 | <i>Chaetoceros</i>    | <i>cf. neogracile</i> | RCC1993           | 10  | 10  | 10 | 10 | [53] |
| MMETSP0716 | <i>Chaetoceros</i>    | <i>curvisetus</i>     | no strain         | 7   | 30  | 7  | 30 | [57] |
| MMETSP0717 | <i>Chaetoceros</i>    | <i>curvisetus</i>     | no strain         | 7   | 30  | 7  | 30 | [57] |
| MMETSP0149 | <i>Chaetoceros</i>    | <i>debilis</i>        | MM31A-1           | 8   | 40  | 8  | 40 | [57] |
| MMETSP0150 | <i>Chaetoceros</i>    | <i>debilis</i>        | MM31A-1           | 8   | 40  | 8  | 40 | [57] |
| MMETSP1447 | <i>Chaetoceros</i>    | <i>dichaeta</i>       | CCMP1751          | 45  | 50  | 48 | 56 | [5]  |
| MMETSP0751 | <i>Chaetoceros</i>    | <i>neogracile</i>     | CCMP1317          | 8   | 11  | 3  | 4  | [5]  |
| MMETSP0200 | <i>Chaetoceros</i>    | <i>sp.</i>            | GSL56             | -   | -   | -  | -  | [NA] |
| MMETSP1429 | <i>Chaetoceros</i>    | <i>sp.</i>            | UNC1202           | -   | -   | -  | -  | [NA] |
| MMETSP0947 | <i>Chattonella</i>    | <i>subsalsa</i>       | CCMP2191          | 36  | 48  | 18 | 25 | [5]  |
| MMETSP1180 | <i>Chlamydomonas</i>  | <i>cf. sp.</i>        | CCMP681           | 11  | 12  | 6  | 8  | [5]  |
| MMETSP1392 | <i>Chlamydomonas</i>  | <i>chlamydogama</i>   | SAG 11-48b        | -   | -   | -  | -  | [NA] |
| MMETSP0063 | <i>Chlamydomonas</i>  | <i>euryale</i>        | CCMP219           | 5   | 6   | 6  | 10 | [5]  |
| MMETSP1391 | <i>Chlamydomonas</i>  | <i>leiostraca</i>     | SAG 11-49         | 7   | 8   | 5  | 5  | [10] |
| MMETSP0109 | <i>Chlorarachnion</i> | <i>reptans</i>        | CCCM449           | 8   | 10  | 8  | 10 | [9]  |
| MMETSP0290 | <i>Chromera</i>       | <i>velia</i>          | CCMP2878          | 4   | 8   | 4  | 8  | [5]  |

|            |                         |                               |             |       |      |     |     |     |      |
|------------|-------------------------|-------------------------------|-------------|-------|------|-----|-----|-----|------|
| MMETSP1095 | <i>Chromulina</i>       | <i>nebulosa</i>               | UTEXLB2642  | -     | -    | -   | -   | -   | [NA] |
| MMETSP0047 | <i>Chroomonas</i>       | <i>cf. mesostigmatica</i>     | CCMP1168    | 8     | 10   | 4   | 4   | 5   | [5]  |
| MMETSP1094 | <i>Chrysochromulina</i> | <i>brevifilum</i>             | UTEX LB 985 | 4     | 9    | 4   | 4   | 9   | [6]  |
| MMETSP1096 | <i>Chrysochromulina</i> | <i>ericina</i>                | CCMP281     | 5     | 6    | 5   | 5   | 6   | [5]  |
| MMETSP0143 | <i>Chrysochromulina</i> | <i>polylepis</i>              | CCMP1757    | 6     | 8    | 4   | 4   | 7   | [5]  |
| MMETSP0286 | <i>Chrysochromulina</i> | <i>polylepis</i>              | UIO037      | 6     | 8    | 4   | 4   | 7   | [43] |
| MMETSP0287 | <i>Chrysochromulina</i> | <i>rotalis</i>                | UIO044      | 4     | 6    | 4   | 4   | 6   | [44] |
| MMETSP1335 | <i>Chrysoculter</i>     | <i>rhomboideus</i>            | RCC1486     | 6     | 12   | 4   | 4   | 8   | [53] |
| MMETSP1165 | <i>Chrysozystis</i>     | <i>fragilis</i>               | CCMP3189    | 8     | 14   | 8   | 8   | 14  | [29] |
| MMETSP1164 | <i>Chrysoreinhardia</i> | <i>sp.</i>                    | CCMP2950    | 0     | 0    | 0   | 0   | 0   | [5]  |
| MMETSP1166 | <i>Chrysoreinhardia</i> | <i>sp.</i>                    | CCMP3193    | 10    | 10   | 10  | 10  | 10  | [5]  |
| MMETSP1397 | <i>Climacostomum</i>    | <i>virens</i>                 | Stock W-24  | 53    | 138  | 53  | 53  | 138 | [52] |
| MMETSP0164 | <i>Coccolithus</i>      | <i>pelagicus ssp braarudi</i> | PLY182g     | 16    | 22   | 16  | 16  | 22  | [50] |
| MMETSP0312 | <i>Compsopogon</i>      | <i>coeruleus</i>              | SAG 36.94   | 12.49 | 17.2 | 9   | 9   | 15  | [10] |
| MMETSP0210 | <i>Condylostoma</i>     | <i>magnum</i>                 | COL2        | 450   | 800  | 80  | 80  | 120 | [12] |
| MMETSP0010 | <i>Corethron</i>        | <i>hystrix</i>                | 308         | 20    | 150  | 20  | 20  | 40  | [24] |
| MMETSP0169 | <i>Corethron</i>        | <i>pennatum</i>               | L29A3       | -     | -    | -   | -   | -   | [NA] |
| MMETSP1066 | <i>Coscinodiscus</i>    | <i>walesii</i>                | CCMP2513    | 100   | 220  | 100 | 100 | 220 | [5]  |
| MMETSP1442 | <i>Craspedostauros</i>  | <i>australis</i>              | CCMP3328    | 28    | 70   | 8   | 8   | 12  | [5]  |
| MMETSP0803 | <i>Crustomastix</i>     | <i>stigmata</i>               | CCMP3273    | 3     | 5    | 2   | 2   | 4   | [5]  |
| MMETSP0323 | <i>Cryptocodinium</i>   | <i>cohnii</i>                 | Seigo       | -     | -    | -   | -   | -   | [NA] |
| MMETSP1050 | <i>Cryptomonas</i>      | <i>curvata</i>                | CCAP979/52  | 20    | 20   | 15  | 15  | 15  | [9]  |
| MMETSP0038 | <i>Cryptomonas</i>      | <i>paramecium</i>             | CCAP977/2a  | -     | -    | -   | -   | -   | [9]  |
| MMETSP1086 | <i>Cyanoptyche</i>      | <i>gloeocystis</i>            | SAG4.97     | 2     | 4    | 2   | 2   | 4   | [21] |
| MMETSP0397 | <i>Cyclophora</i>       | <i>tenuis</i>                 | ECT3854     | 44    | 76   | 7   | 7   | 11  | [4]  |

|            |                      |                        |             |      |      |      |      |      |
|------------|----------------------|------------------------|-------------|------|------|------|------|------|
| MMETSP1057 | <i>Cyclotella</i>    | <i>meneghiniana</i>    | CCMP338     | 6    | 11   | 6    | 7    | [5]  |
| MMETSP0017 | <i>Cylindrotheca</i> | <i>closterium</i>      | KMMCC:B-181 | 11   | 33   | 2.5  | 5.5  | [NA] |
| MMETSP0580 | <i>Dactyliosolen</i> | <i>fragilissimus</i>   | no strain   | 42   | 300  | 8    | 70   | [57] |
| MMETSP0232 | <i>Debaryomyces</i>  | <i>hansenii</i>        | J26         | 3.22 | 4.44 | 3.22 | 4.44 | [41] |
| MMETSP1058 | <i>Detonula</i>      | <i>confervacea</i>     | CCMP353     | 7    | 20   | 8    | 10   | [5]  |
| MMETSP1174 | <i>Dictyocha</i>     | <i>speculum</i>        | CCMP1381    | 16   | 72   | 16   | 56   | [5]  |
| MMETSP0019 | <i>Dinobryon</i>     | <i>sp.</i>             | UTEXL B2267 | -    | -    | -    | -    | [NA] |
| MMETSP0797 | <i>Dinophysis</i>    | <i>acuminata</i>       | DAEP01      | 30   | 35   | 38   | 58   | [49] |
| MMETSP1002 | <i>Ditylum</i>       | <i>brightwellii</i>    | GSO103      | 80   | 130  | 25   | 100  | [57] |
| MMETSP1010 | <i>Ditylum</i>       | <i>brightwellii</i>    | GSO104      | 80   | 130  | 25   | 100  | [57] |
| MMETSP0998 | <i>Ditylum</i>       | <i>brightwellii</i>    | GSO105      | 80   | 130  | 25   | 100  | [57] |
| MMETSP1062 | <i>Ditylum</i>       | <i>brightwellii</i>    | Pop1 (SS4)  | 80   | 130  | 25   | 100  | [57] |
| MMETSP1063 | <i>Ditylum</i>       | <i>brightwellii</i>    | Pop2 (SS10) | 80   | 130  | 25   | 100  | [57] |
| MMETSP0033 | <i>Dolichomastix</i> | <i>tenuilepis</i>      | CCMP3274    | 3    | 5    | 3    | 5    | [5]  |
| MMETSP0439 | <i>Dracoamoeba</i>   | <i>jormungandri</i>    | Chinc5      | 19   | 25   | 19   | 25   | [56] |
| MMETSP1126 | <i>Dunaliella</i>    | <i>tertiolecta</i>     | CCMP1320    | 6    | 9    | 0    | 0    | [5]  |
| MMETSP0116 | <i>Durinskia</i>     | <i>baltica</i>         | CSIRO CS-38 | -    | -    | -    | -    | [NA] |
| MMETSP1385 | <i>Elphidium</i>     | <i>margaritaceum</i>   | no strain   | -    | -    | -    | -    | [NA] |
| MMETSP1154 | <i>Emiliana</i>      | <i>huxleyi</i>         | CCMP370     | 2    | 5    | 2    | 5    | [5]  |
| MMETSP1006 | <i>Emiliana</i>      | <i>huxleyi</i>         | CCMP374     | 4    | 8    | 4    | 6    | [5]  |
| MMETSP0994 | <i>Emiliana</i>      | <i>huxleyi</i>         | CCMP379     | 3    | 5    | 3    | 5    | [5]  |
| MMETSP1150 | <i>Emiliana</i>      | <i>huxleyi</i>         | PLY M219    | 2.9  | 3.4  | 2.9  | 3.4  | [50] |
| MMETSP1443 | <i>Entomoneis</i>    | <i>sp.</i>             | CCMP2396    | 24   | 30   | 20   | 24   | [5]  |
| MMETSP1353 | <i>Erythrolobus</i>  | <i>australicus</i>     | CCMP3124    | 10   | 22   | 8    | 17   | [5]  |
| MMETSP1354 | <i>Erythrolobus</i>  | <i>madagascarensis</i> | CCMP3276    | 5    | 8    | 5    | 8    | [5]  |

|            |                         |                        |                     |     |     |     |     |      |
|------------|-------------------------|------------------------|---------------------|-----|-----|-----|-----|------|
| MMETSP1437 | <i>Eucampia</i>         | <i>antarctica</i>      | CCMP1452            | 60  | 66  | 16  | 40  | [5]  |
| MMETSP1380 | <i>Euplotes</i>         | <i>crassus</i>         | CT5                 | 65  | 70  | 36  | 38  | [16] |
| MMETSP0205 | <i>Euplotes</i>         | <i>focardii</i>        | TN1                 | 70  | 80  | 50  | 60  | [58] |
| MMETSP0213 | <i>Euplotes</i>         | <i>harpa</i>           | FSP1.4              | 120 | 260 | 120 | 260 | [33] |
| MMETSP0039 | <i>Eutreptiella</i>     | <i>gymnastica</i>      | NIES-381            | 60  | 80  | 16  | 24  | [NA] |
| MMETSP0809 | <i>Eutreptiella</i>     | <i>gymnastica-like</i> | CCMP1594            | 60  | 80  | 16  | 24  | [5]  |
| MMETSP1464 | <i>Exanthemachrysis</i> | <i>gayraliae</i>       | RCC1523             | 5   | 6   | 3   | 6   | [53] |
| MMETSP0696 | <i>Extubocellulus</i>   | <i>spinifer</i>        | CCMP396             | 4   | 7   | 2   | 3   | [5]  |
| MMETSP1345 | <i>Fabrea</i>           | <i>salina</i>          | no strain           | 190 | 240 | 190 | 240 | [22] |
| MMETSP0123 | <i>Favella</i>          | <i>ehrenbergii</i>     | Fehren 1            | 184 | 184 | 70  | 70  | [NA] |
| MMETSP0434 | <i>Favella</i>          | <i>taraikaensis</i>    | Fe Narragansett Bay | -   | -   | -   | -   | [NA] |
| MMETSP1339 | <i>Fibrocapsa</i>       | <i>japonica</i>        | CCMP1661            | 22  | 28  | 20  | 24  | [5]  |
| MMETSP0413 | <i>Filamoeba</i>        | <i>nolandi</i>         | NC-AS-23-1          | 15  | 50  | 15  | 50  | [34] |
| MMETSP1344 | <i>Florenciella</i>     | <i>parvula</i>         | CCMP2471            | 4   | 6   | 4   | 6   | [5]  |
| MMETSP1323 | <i>Florenciella</i>     | <i>parvula</i>         | RCC1693             | 5   | 8   | 4   | 6   | [53] |
| MMETSP1325 | <i>Florenciella</i>     | <i>sp.</i>             | RCC1007             | 3.5 | 3.5 | 3.5 | 3.5 | [53] |
| MMETSP1324 | <i>Florenciella</i>     | <i>sp.</i>             | RCC1587             | 4.5 | 4.5 | 4.5 | 4.5 | [53] |
| MMETSP0733 | <i>Fragilariopsis</i>   | <i>kerquelensis</i>    | L26-C5              | 10  | 76  | 5   | 11  | [57] |
| MMETSP0906 | <i>Fragilariopsis</i>   | <i>kerquelensis</i>    | L2-C3               | 10  | 76  | 5   | 11  | [57] |
| MMETSP0766 | <i>Gambierdiscus</i>    | <i>australes</i>       | CAWD 149            | -   | -   | -   | -   | [NA] |
| MMETSP0799 | <i>Gemingera</i>        | <i>cryophila</i>       | CCMP2564            | 17  | 20  | 9   | 12  | [5]  |
| MMETSP1102 | <i>Gemingera</i>        | <i>sp.</i>             | Caron Lab Isolate   | -   | -   | -   | -   | [NA] |
| MMETSP1329 | <i>Genus nov.</i>       | <i>species nov.</i>    | RCC1024             | 3   | 6   | 3   | 6   | [53] |
| MMETSP1326 | <i>Genus nov.</i>       | <i>species nov.</i>    | RCC2288             | 3   | 3   | 3   | 3   | [53] |
| MMETSP1312 | <i>Genus nov.</i>       | <i>species nov.</i>    | RCC2335             | 2   | 2   | 2   | 2   | [53] |

|            |                      |                     |              |      |      |      |      |      |      |
|------------|----------------------|---------------------|--------------|------|------|------|------|------|------|
| MMETSP1310 | <i>Genus nov.</i>    | <i>species nov.</i> | RCC2339      | 2    | 2    | 2    | 2    | 2    | [53] |
| MMETSP1311 | <i>Genus nov.</i>    | <i>species nov.</i> | RCC856       | 2    | 2    | 2    | 2    | 2    | [53] |
| MMETSP1309 | <i>Genus nov.</i>    | <i>species nov.</i> | RCC998       | 2.5  | 2.5  | 2.5  | 2.5  | 2.5  | [53] |
| MMETSP1363 | <i>Gephyrocapsa</i>  | <i>oceanica</i>     | RCC1303      | 6    | 10   | 6    | 6    | 10   | [42] |
| MMETSP0118 | <i>Glendinium</i>    | <i>foliaceum</i>    | CCAP 1116/3  | 33.5 | 33.5 | 26.5 | 26.5 | 26.5 | [9]  |
| MMETSP0308 | <i>Gloeochaete</i>   | <i>witrockiana</i>  | SAG 46.84    | 6    | 12   | 6    | 6    | 12   | [10] |
| MMETSP0107 | <i>Goniomonas</i>    | <i>Pacifica</i>     | CCMP1869     | 6    | 9    | 4    | 4    | 6    | [5]  |
| MMETSP0114 | <i>Goniomonas</i>    | <i>sp.</i>          | m            | 6    | 9    | 4    | 4    | 6    | [28] |
| MMETSP1439 | <i>Gonyaulax</i>     | <i>spinifera</i>    | CCMP409      | 50   | 50   | 50   | 50   | 50   | [5]  |
| MMETSP0009 | <i>Grammatophora</i> | <i>oceanica</i>     | CCMP410      | 24   | 32   | 16   | 16   | 24   | [5]  |
| MMETSP0046 | <i>Guillardia</i>    | <i>theta</i>        | CCMP2712     | 8    | 11   | 4    | 4    | 6    | [5]  |
| MMETSP0110 | <i>Gymnochlora</i>   | <i>sp.</i>          | CCMP2014     | 8    | 12   | 8    | 8    | 12   | [5]  |
| MMETSP0784 | <i>Gymnodinium</i>   | <i>catenatum</i>    | GC744        | -    | -    | -    | -    | -    | [NA] |
| MMETSP1148 | <i>Gyrodinium</i>    | <i>dominans</i>     | SPMC 103     | -    | -    | -    | -    | -    | [NA] |
| MMETSP1048 | <i>Hanusia</i>       | <i>phi</i>          | CCMP325      | 9    | 15   | 9    | 9    | 15   | [5]  |
| MMETSP1171 | <i>Helicotheca</i>   | <i>tamensis</i>     | CCMP826      | 30   | 69   | 25   | 25   | 38   | [5]  |
| MMETSP1042 | <i>Hemiselmis</i>    | <i>andersenii</i>   | CCMP1180     | 4    | 6    | 3    | 3    | 4    | [5]  |
| MMETSP1041 | <i>Hemiselmis</i>    | <i>andersenii</i>   | CCMP439      | 4    | 8    | 4    | 4    | 5    | [5]  |
| MMETSP1043 | <i>Hemiselmis</i>    | <i>andersenii</i>   | CCMP441      | 4    | 7    | 3    | 3    | 5    | [5]  |
| MMETSP0043 | <i>Hemiselmis</i>    | <i>andersenii</i>   | CCMP644      | 4    | 8    | 2    | 2    | 4    | [5]  |
| MMETSP1357 | <i>Hemiselmis</i>    | <i>rufescens</i>    | PCC563       | 4    | 8    | 2    | 2    | 4    | [25] |
| MMETSP1355 | <i>Hemiselmis</i>    | <i>tepida</i>       | CCMP443      | 3    | 6    | 3    | 3    | 4    | [5]  |
| MMETSP1356 | <i>Hemiselmis</i>    | <i>viresens</i>     | PCC157       | 4    | 8    | 2    | 2    | 4    | [25] |
| MMETSP1441 | <i>Heterocapsa</i>   | <i>arctica</i>      | CCMP445      | 22.5 | 37.5 | 10   | 10   | 15   | [5]  |
| MMETSP0503 | <i>Heterocapsa</i>   | <i>rotundata</i>    | SCCAP K-0483 | -    | -    | -    | -    | -    | [NA] |

|            |                         |                             |                   |     |     |     |     |      |
|------------|-------------------------|-----------------------------|-------------------|-----|-----|-----|-----|------|
| MMETSP0448 | <i>Heterocapsa</i>      | <i>triquetra</i>            | CCMP448           | 20  | 28  | 14  | 18  | [5]  |
| MMETSP0292 | <i>Heterosigma</i>      | <i>akashiwo</i>             | CCMP2393          | 12  | 24  | 10  | 16  | [5]  |
| MMETSP0409 | <i>Heterosigma</i>      | <i>akashiwo</i>             | CCMP3107          | 18  | 34  | 18  | 34  | [5]  |
| MMETSP0894 | <i>Heterosigma</i>      | <i>akashiwo</i>             | CCMP452           | 10  | 15  | 10  | 14  | [5]  |
| MMETSP0414 | <i>Heterosigma</i>      | <i>akashiwo</i>             | NB                | -   | -   | -   | -   | [NA] |
| MMETSP1474 | <i>Imantonia</i>        | <i>sp.</i>                  | RCC918            | 3.5 | 3.5 | 3.5 | 3.5 | [53] |
| MMETSP0595 | <i>Isochrysis</i>       | <i>galbana</i>              | CCMP1323          | 4   | 6   | 2   | 4   | [5]  |
| MMETSP1090 | <i>Isochrysis</i>       | <i>sp.</i>                  | CCMP1244          | 3   | 5   | 3   | 5   | [5]  |
| MMETSP1129 | <i>Isochrysis</i>       | <i>sp.</i>                  | CCMP1324          | 4   | 8   | 0   | 0   | [5]  |
| MMETSP0027 | <i>Karenia</i>          | <i>brevis</i>               | CCMP2229          | 20  | 28  | 20  | 24  | [5]  |
| MMETSP0573 | <i>Karenia</i>          | <i>brevis</i>               | SP1               | 20  | 28  | 20  | 24  | [3]  |
| MMETSP0527 | <i>Karenia</i>          | <i>brevis</i>               | SP3               | 20  | 28  | 20  | 24  | [3]  |
| MMETSP0201 | <i>Karenia</i>          | <i>brevis</i>               | Wilson            | 20  | 28  | 20  | 24  | [3]  |
| MMETSP1015 | <i>Karlodinium</i>      | <i>micrum</i>               | CCMP2283          | 12  | 14  | 8   | 10  | [5]  |
| MMETSP0120 | <i>Kryptoperidinium</i> | <i>foliaceum</i>            | CCMP1326          | 24  | 40  | 22  | 34  | [5]  |
| MMETSP0372 | <i>Lankesteria</i>      | <i>abbottii</i>             | Grappler Inlet BC | -   | -   | -   | -   | [NA] |
| MMETSP1362 | <i>Leptocylindrus</i>   | <i>danicus</i>              | CCMP1856          | 54  | 81  | 12  | 18  | [5]  |
| MMETSP0322 | <i>Leptocylindrus</i>   | <i>danicus var. apora</i>   | B651              | -   | -   | -   | -   | [57] |
| MMETSP0321 | <i>Leptocylindrus</i>   | <i>danicus var. danicus</i> | B650              | 5   | 16  | 5   | 16  | [57] |
| MMETSP1147 | <i>Lessardia</i>        | <i>elongata</i>             | SPMC 104          | -   | -   | -   | -   | [NA] |
| MMETSP1360 | <i>Licmophora</i>       | <i>paradoxa</i>             | CCMP2313          | 18  | 33  | 12  | 16  | [5]  |
| MMETSP1032 | <i>Lingulodinium</i>    | <i>polyedra</i>             | CCMP1738          | 38  | 40  | 35  | 40  | [5]  |
| MMETSP0209 | <i>Litonotus</i>        | <i>pictus</i>               | P1                | 175 | 600 | 20  | 90  | [23] |
| MMETSP0042 | <i>Lotharella</i>       | <i>amoebiformis</i>         | CCMP2058          | 10  | 14  | 10  | 14  | [5]  |
| MMETSP0111 | <i>Lotharella</i>       | <i>globosa</i>              | CCCM811           | 5   | 10  | 5   | 10  | [7]  |



|            |                     |                         |              |     |      |     |      |      |
|------------|---------------------|-------------------------|--------------|-----|------|-----|------|------|
| MMETSP0041 | <i>Lotharella</i>   | <i>globosa</i>          | LEX01        | 7.5 | 10.2 | 7.5 | 10.2 | [18] |
| MMETSP0040 | <i>Lotharella</i>   | <i>oceanica</i>         | CCMP622      | 4   | 7    | 4   | 6    | [5]  |
| MMETSP1450 | <i>Madagascania</i> | <i>erythrocladiodes</i> | CCMP3234     | 6.1 | 12.4 | 4.9 | 6    | [5]  |
| MMETSP1167 | <i>Mallomonas</i>   | <i>sp.</i>              | CCMP3275     | 20  | 28   | 10  | 14   | [5]  |
| MMETSP1106 | <i>Mantoniella</i>  | <i>antarctica</i>       | SL-175       | 2.8 | 5    | 2.8 | 5    | [31] |
| MMETSP1468 | <i>Mantoniella</i>  | <i>sp.</i>              | CCMP1436     | 0   | 0    | 0   | 0    | [5]  |
| MMETSP0417 | <i>Mayorella</i>    | <i>sp.</i>              | BSH-02190019 | 40  | 142  | 40  | 142  | [35] |
| MMETSP0467 | <i>Mesodinium</i>   | <i>pulex</i>            | SPMC105      | 25  | 35   | 10  | 15   | [1]  |
| MMETSP1401 | <i>Micromonas</i>   | <i>pusilla</i>          | CCAC1681     | 2   | 3    | 2   | 3    | [32] |
| MMETSP1403 | <i>Micromonas</i>   | <i>pusilla</i>          | CCMP1723     | 2   | 3    | 2   | 3    | [5]  |
| MMETSP1404 | <i>Micromonas</i>   | <i>pusilla</i>          | CCMP494      | 1   | 2    | 1   | 2    | [5]  |
| MMETSP1402 | <i>Micromonas</i>   | <i>pusilla</i>          | RCC1614      | 2   | 3    | 2   | 3    | [3]  |
| MMETSP1327 | <i>Micromonas</i>   | <i>pusilla</i>          | RCC2306      | 2   | 2    | 2   | 2    | [53] |
| MMETSP1080 | <i>Micromonas</i>   | <i>sp.</i>              | CCMP1646     | 2   | 3    | 2   | 3    | [5]  |
| MMETSP0802 | <i>Micromonas</i>   | <i>sp.</i>              | CCMP2099     | 1   | 3    | 2   | 3    | [5]  |
| MMETSP1393 | <i>Micromonas</i>   | <i>sp.</i>              | CS-222       | 2   | 3    | 2   | 3    | [3]  |
| MMETSP1082 | <i>Micromonas</i>   | <i>sp.</i>              | NEPCC29      | 2   | 3    | 2   | 3    | [3]  |
| MMETSP1400 | <i>Micromonas</i>   | <i>sp.</i>              | RCC451       | 2   | 2    | 2   | 2    | [53] |
| MMETSP1084 | <i>Micromonas</i>   | <i>sp.</i>              | RCC472       | 2.5 | 2.5  | 2.5 | 2.5  | [53] |
| MMETSP0186 | <i>Minchinia</i>    | <i>chitonis</i>         | no strain    | -   | -    | -   | -    | [NA] |
| MMETSP1434 | <i>Minutocellus</i> | <i>polymorphus</i>      | CCMP3303     | 5   | 8    | 3   | 4    | [5]  |
| MMETSP1070 | <i>Minutocellus</i> | <i>polymorphus</i>      | NH13         | -   | -    | -   | -    | [NA] |
| MMETSP1322 | <i>Minutocellus</i> | <i>polymorphus</i>      | RCC2270      | 4   | 4    | 4   | 4    | [53] |
| MMETSP0798 | <i>Myrionecta</i>   | <i>rubra</i>            | CCMP2563     | 32  | 40   | 32  | 40   | [5]  |
| MMETSP1114 | <i>Neobodo</i>      | <i>designis</i>         | CCAP 1951/1  | -   | -    | -   | -    | [9]  |

|            |                      |                      |                       |      |      |     |     |     |      |
|------------|----------------------|----------------------|-----------------------|------|------|-----|-----|-----|------|
| MMETSP0161 | <i>Neoparamoeba</i>  | <i>aestuarina</i>    | SoJaBio B1-5/56/2     | 45   | 100  | 45  | 100 | 100 | [36] |
| MMETSP0034 | <i>Nephroselmis</i>  | <i>pyriformis</i>    | CCMP717               | 3    | 5    | 3   | 5   | 5   | [5]  |
| MMETSP0744 | <i>Nitzschia</i>     | <i>punctata</i>      | CCMP561               | 5    | 9    | 5   | 6   | 6   | [5]  |
| MMETSP0014 | <i>Nitzschia</i>     | <i>sp.</i>           | RCC80                 | 10   | 10   | 2.5 | 2.5 | 2.5 | [53] |
| MMETSP0253 | <i>Noctiluca</i>     | <i>scintillans</i>   | Noctiluca scintillans | 2000 | 2000 | 200 | 200 | 200 | [NA] |
| MMETSP0974 | <i>non described</i> | <i>non described</i> | CCMP2097              | 6    | 12   | 6   | 12  | 12  | [5]  |
| MMETSP0990 | <i>non described</i> | <i>non described</i> | CCMP2098              | 8    | 10   | 6   | 7   | 7   | [5]  |
| MMETSP0986 | <i>non described</i> | <i>non described</i> | CCMP2293              | 8    | 10   | 4   | 5   | 5   | [5]  |
| MMETSP1141 | <i>non described</i> | <i>non described</i> | CCMP2298              | 7    | 16   | 7   | 16  | 16  | [5]  |
| MMETSP0982 | <i>non described</i> | <i>non described</i> | CCMP2436              | 5    | 10   | 8   | 12  | 12  | [5]  |
| MMETSP0113 | <i>Norrsiella</i>    | <i>sphaerica</i>     | BC52                  | 5    | 9    | 5   | 9   | 9   | [46] |
| MMETSP1105 | <i>Ochromonas</i>    | <i>sp.</i>           | BG-1                  | -    | -    | -   | -   | -   | [NA] |
| METSP0004  | <i>Ochromonas</i>    | <i>sp.</i>           | CCMP1393              | 4    | 6    | 4   | 6   | 6   | [5]  |
| MMETSP1177 | <i>Ochromonas</i>    | <i>sp.</i>           | CCMP1899              | 6    | 9    | 6   | 9   | 9   | [5]  |
| MMETSP0015 | <i>Odontella</i>     | <i>aurita</i>        | isolate 1302-5        | 10   | 97   | 8   | 11  | 11  | [57] |
| MMETSP0160 | <i>Odontella</i>     | <i>sinensis</i>      | Grunow 1884           | 90   | 260  | 16  | 18  | 18  | [57] |
| MMETSP0939 | <i>Ostreococcus</i>  | <i>lucimarinus</i>   | clade-A-BCC118000     | 1    | 1    | 1   | 1   | 1   | [53] |
| MMETSP0938 | <i>Ostreococcus</i>  | <i>mediterraneus</i> | clade-D-RCC1107       | 1    | 1    | 1   | 1   | 1   | [53] |
| MMETSP0930 | <i>Ostreococcus</i>  | <i>mediterraneus</i> | clade-D-RCC1621       | 1    | 1    | 1   | 1   | 1   | [53] |
| MMETSP0929 | <i>Ostreococcus</i>  | <i>mediterraneus</i> | clade-D-RCC2572       | 1    | 1    | 1   | 1   | 1   | [53] |
| MMETSP0936 | <i>Ostreococcus</i>  | <i>mediterraneus</i> | clade-D-RCC2573       | 1    | 1    | 1   | 1   | 1   | [53] |
| MMETSP0937 | <i>Ostreococcus</i>  | <i>mediterraneus</i> | clade-D-RCC2593       | 1    | 1    | 1   | 1   | 1   | [53] |
| MMETSP0932 | <i>Ostreococcus</i>  | <i>mediterraneus</i> | clade-D-RCC2596       | 1    | 1    | 1   | 1   | 1   | [53] |
| MMETSP0933 | <i>Ostreococcus</i>  | <i>prasinos</i>      | BCC99000              | 1    | 1    | 1   | 1   | 1   | [NA] |
| MMETSP0044 | <i>Oxyrrhis</i>      | <i>marina</i>        | CCMP1788              | 30   | 35   | 4   | 6   | 6   | [5]  |

|            |                       |                     |                       |    |     |    |     |      |
|------------|-----------------------|---------------------|-----------------------|----|-----|----|-----|------|
| MMETSP0451 | <i>Oxyrrhis</i>       | <i>marina</i>       | CCMP1795              | 30 | 35  | 4  | 6   | [5]  |
| MMETSP1424 | <i>Oxyrrhis</i>       | <i>marina</i>       | LB1974                | 30 | 35  | 4  | 6   | [3]  |
| MMETSP0468 | <i>Oxyrrhis</i>       | <i>marina</i>       | no strain             | 30 | 35  | 4  | 6   | [3]  |
| MMETSP0780 | <i>Palpitomonas</i>   | <i>bilix</i>        | NIES-2562             | -  | -   | -  | -   | [NA] |
| MMETSP0151 | <i>Paramoeba</i>      | <i>atlantica</i>    | 621/1 / CCAP 1560/9   | 45 | 100 | 45 | 100 | [36] |
| MMETSP1103 | <i>Paraphysomonas</i> | <i>bandaiensis</i>  | Caron Lab Isolate     | -  | -   | -  | -   | [NA] |
| MMETSP0103 | <i>Paraphysomonas</i> | <i>Imperforata</i>  | PA2                   | -  | -   | -  | -   | [NA] |
| MMETSP1107 | <i>Paraphysomonas</i> | <i>vestita</i>      | GFlagA                | -  | -   | -  | -   | [NA] |
| MMETSP1318 | <i>Partenskyella</i>  | <i>glossopodia</i>  | RCC365                | 3  | 3   | 3  | 3   | [53] |
| MMETSP1466 | <i>Pavlova</i>        | <i>gyrans</i>       | CCMP608               | 5  | 8   | 2  | 4   | [5]  |
| MMETSP1463 | <i>Pavlova</i>        | <i>lutheri</i>      | RCC1537               | 3  | 6   | 3  | 6   | [53] |
| MMETSP1139 | <i>Pavlova</i>        | <i>sp.</i>          | CCMP459               | 4  | 8   | 4  | 5   | [5]  |
| MMETSP0882 | <i>Pelagococcus</i>   | <i>subviridis</i>   | CCMP1429              | 2  | 4   | 2  | 4   | [5]  |
| MMETSP1338 | <i>Pelagodinium</i>   | <i>beii</i>         | RCC1491               | -  | -   | -  | -   | [NA] |
| MMETSP0886 | <i>Pelagomonas</i>    | <i>calceolata</i>   | CCMP1756              | 2  | 4   | 1  | 3   | [5]  |
| MMETSP1328 | <i>Pelagomonas</i>    | <i>calceolata</i>   | RCC969                | 2  | 2   | 2  | 2   | [53] |
| MMETSP0758 | <i>Percolomonas</i>   | <i>cosmopolitus</i> | AE-1 (ATCC50343)      | -  | -   | -  | -   | [NA] |
| MMETSP0759 | <i>Percolomonas</i>   | <i>cosmopolitus</i> | WS                    | -  | -   | -  | -   | [NA] |
| MMETSP0370 | <i>Peridinium</i>     | <i>aciculiferum</i> | PAER-2                | -  | -   | -  | -   | [NA] |
| MMETSP0924 | <i>Perkinsus</i>      | <i>chesapeakei</i>  | ATCCPRA-65            | -  | -   | -  | -   | [NA] |
| MMETSP0922 | <i>Perkinsus</i>      | <i>marinus</i>      | ATCC50439             | 6  | 8   | 6  | 8   | [11] |
| MMETSP0420 | <i>Pessonella</i>     | <i>sp.</i>          | Pessonella sp. PRA-29 | -  | -   | -  | -   | [NA] |
| MMETSP1100 | <i>Phaeocystis</i>    | <i>antarctica</i>   | Caron Lab Isolate     | 3  | 9   | 3  | 9   | [3]  |
| MMETSP1444 | <i>Phaeocystis</i>    | <i>antarctica</i>   | CCMP1374              | 3  | 9   | 3  | 9   | [54] |
| MMETSP1465 | <i>Phaeocystis</i>    | <i>cordata</i>      | RCC1383               | 3  | 4   | 3  | 4   | [53] |

|            |                      |                     |                        |      |      |     |      |     |      |      |
|------------|----------------------|---------------------|------------------------|------|------|-----|------|-----|------|------|
| MMETSP1162 | <i>Phaeocystis</i>   | <i>sp.</i>          | CCMP2710               | 4    | 8    | 4   | 8    | 4   | 8    | [5]  |
| MMETSP1163 | <i>Phaeomonas</i>    | <i>parva</i>        | CCMP2877               | 2    | 4    | 2   | 4    | 2   | 4    | [5]  |
| MMETSP1161 | <i>Picochlorum</i>   | <i>oklahomensis</i> | CCMP2329               | 2    | 3    | 2   | 3    | 2   | 3    | [5]  |
| MMETSP1330 | <i>Picochlorum</i>   | <i>sp.</i>          | RCC944                 | 1.5  | 1.5  | 1.5 | 1.5  | 1.5 | 1.5  | [53] |
| MMETSP0807 | <i>Picocystis</i>    | <i>salinarum</i>    | CCMP1897               | 2    | 5    | 2   | 5    | 2   | 5    | [5]  |
| MMETSP1160 | <i>Pinguicoccus</i>  | <i>pyrenoidosus</i> | CCMP2078               | 3    | 10   | 3   | 10   | 3   | 8    | [5]  |
| MMETSP0127 | <i>Platyophrya</i>   | <i>macrostoma</i>   | WH                     | -    | -    | -   | -    | -   | -    | [NA] |
| MMETSP1136 | <i>Pleurochrysis</i> | <i>carterae</i>     | CCMP645                | 8    | 12   | 8   | 12   | 8   | 12   | [5]  |
| MMETSP0227 | <i>Polarella</i>     | <i>glacialis</i>    | CCMP1383               | 10   | 15   | 4   | 8    | 4   | 8    | [5]  |
| MMETSP1440 | <i>Polarella</i>     | <i>glacialis</i>    | CCMP2088               | 14   | 18   | 8   | 10   | 8   | 10   | [5]  |
| MMETSP0052 | <i>Polytomella</i>   | <i>parva</i>        | SAG 63-3               | 10   | 14.3 | 10  | 14.3 | 10  | 14.3 | [30] |
| MMETSP0313 | <i>Porphyridium</i>  | <i>aerugineum</i>   | SAG 1380-2             | 5.6  | 7    | 5.6 | 7    | 5.6 | 7    | [10] |
| MMETSP0941 | <i>Prasinococcus</i> | <i>capsulatus</i>   | CCMP1194               | 3    | 5    | 3   | 5    | 3   | 5    | [5]  |
| MMETSP0806 | <i>Prasinoderma</i>  | <i>coloniale</i>    | CCMP1413               | 2    | 3    | 2   | 3    | 2   | 3    | [5]  |
| MMETSP1315 | <i>Prasinoderma</i>  | <i>singularis</i>   | RCC927                 | 3    | 3    | 3   | 3    | 3   | 3    | [53] |
| MMETSP0174 | <i>Proboscia</i>     | <i>alata</i>        | PI-D3                  | 1000 | 1000 | 7   | 18   | 7   | 18   | [24] |
| MMETSP0816 | <i>Proboscia</i>     | <i>inermis</i>      | CCAP1064/1             | 100  | 148  | 10  | 13   | 10  | 13   | [9]  |
| MMETSP0252 | <i>Prorocentrum</i>  | <i>lima</i>         | CCMP684                | 42   | 47   | 30  | 32   | 30  | 32   | [5]  |
| MMETSP0251 | <i>Prorocentrum</i>  | <i>micans</i>       | CCCM845                | X    | -    | -   | -    | -   | -    | [7]  |
| MMETSP0053 | <i>Prorocentrum</i>  | <i>minimum</i>      | CCMP1329               | 7    | 16   | 6   | 16   | 6   | 16   | [5]  |
| MMETSP0267 | <i>Prorocentrum</i>  | <i>minimum</i>      | CCMP2233               | 18   | 20   | 16  | 18   | 16  | 18   | [5]  |
| MMETSP1049 | <i>Proteomonas</i>   | <i>sulcata</i>      | CCMP704                | 12   | 14   | 5   | 7    | 5   | 7    | [5]  |
| MMETSP0228 | <i>Protoceratium</i> | <i>reticulatum</i>  | CCCM535<br>(=CCMP1889) | 33   | 33   | 30  | 30   | 30  | 30   | [7]  |
| MMETSP0216 | <i>Protocruzia</i>   | <i>adherens</i>     | Boccale                | -    | -    | -   | -    | -   | -    | [NA] |
| MMETSP0006 | <i>Prymnesium</i>    | <i>parvum</i>       | Texoma1                | 8    | 11   | 4   | 6    | 4   | 6    | [15] |



|            |                       |                         |                   |     |     |    |    |      |
|------------|-----------------------|-------------------------|-------------------|-----|-----|----|----|------|
| MMETSP0789 | <i>Rhizosolenia</i>   | <i>setigera</i>         | CCMP1694          | 130 | 440 | 3  | 12 | [5]  |
| MMETSP0167 | <i>Rhodella</i>       | <i>maculata</i>         | CCMP736           | 7   | 16  | 7  | 16 | [5]  |
| MMETSP1101 | <i>Rhodomonas</i>     | <i>abbreviata</i>       | Caron Lab Isolate | -   | -   | -  | -  | [NA] |
| MMETSP0484 | <i>Rhodomonas</i>     | <i>lens</i>             | RHODO             | -   | -   | -  | -  | [NA] |
| MMETSP1047 | <i>Rhodomonas</i>     | <i>salina</i>           | CCMP1319          | 5   | 13  | 6  | 8  | [5]  |
| MMETSP1091 | <i>Rhodomonas</i>     | <i>sp.</i>              | CCMP768           | 10  | 25  | 6  | 13 | [5]  |
| MMETSP0011 | <i>Rhodosorus</i>     | <i>marinus</i>          | CCMP769           | 5   | 8   | 5  | 8  | [5]  |
| MMETSP0315 | <i>Rhodosorus</i>     | <i>marinus</i>          | UTEX LB 2760      | -   | -   | -  | -  | [NA] |
| MMETSP0190 | <i>Rosalina</i>       | <i>sp.</i>              | no strain         | -   | -   | -  | -  | [NA] |
| MMETSP0437 | <i>Sapocribrum</i>    | <i>chincoteagueense</i> | ATCC50979         | 3   | 5   | 3  | 5  | [37] |
| MMETSP1170 | <i>Sarcinochrysis</i> | <i>sp.</i>              | CCMP770           | 4   | 8   | 4  | 8  | [5]  |
| MMETSP0962 | <i>Schizochytrium</i> | <i>aggregatum</i>       | ATCC28209         | 3   | 6   | 3  | 6  | [20] |
| MMETSP0359 | <i>Scrippsiella</i>   | <i>hangoei</i>          | SHIV-5            | -   | -   | -  | -  | [NA] |
| MMETSP0367 | <i>Scrippsiella</i>   | <i>hangoei-like</i>     | SHHI-4            | -   | -   | -  | -  | [NA] |
| MMETSP0270 | <i>Scrippsiella</i>   | <i>trochoidea</i>       | CCMP3099          | 28  | 36  | 22 | 28 | [5]  |
| MMETSP1333 | <i>Scyphosphaera</i>  | <i>apsteinii</i>        | RCC1455           | -   | -   | -  | -  | [NA] |
| MMETSP0013 | <i>Skeletonema</i>    | <i>costatum</i>         | RCC1716           | 5   | 5   | 5  | 5  | [53] |
| MMETSP0562 | <i>Skeletonema</i>    | <i>dohrnii</i>          | SkelB             | 4   | 12  | 4  | 12 | [55] |
| MMETSP0578 | <i>Skeletonema</i>    | <i>grethea</i>          | CCMP1804          | 4   | 6   | 6  | 8  | [5]  |
| MMETSP0593 | <i>Skeletonema</i>    | <i>japonicum</i>        | CCMP2506          | 5   | 16  | 5  | 8  | [5]  |
| MMETSP1040 | <i>Skeletonema</i>    | <i>marinoi</i>          | FE60              | 12  | 12  | 2  | 2  | [NA] |
| MMETSP1039 | <i>Skeletonema</i>    | <i>marinoi</i>          | FE7               | 12  | 12  | 2  | 2  | [55] |
| MMETSP0918 | <i>Skeletonema</i>    | <i>marinoi</i>          | skelA             | 12  | 12  | 2  | 2  | [55] |
| MMETSP0320 | <i>Skeletonema</i>    | <i>marinoi</i>          | SM1012Den-03      | 12  | 12  | 2  | 2  | [55] |
| MMETSP0319 | <i>Skeletonema</i>    | <i>marinoi</i>          | SM1012Hels-07     | 12  | 12  | 2  | 2  | [55] |

|            |                        |                        |                       |    |     |     |    |      |
|------------|------------------------|------------------------|-----------------------|----|-----|-----|----|------|
| MMETSP1428 | <i>Skeletonema</i>     | <i>marinoi</i>         | UNC1201               | 12 | 12  | 2   | 2  | [55] |
| MMETSP0603 | <i>Skeletonema</i>     | <i>menzeli</i>         | CCMP793               | 3  | 8   | 4   | 8  | [5]  |
| MMETSP0191 | <i>Sorites</i>         | <i>sp.</i>             | <i>Sorites</i> sp.    | -  | -   | -   | -  | [NA] |
| MMETSP1098 | <i>Spumella</i>        | <i>elongata</i>        | CCAP 955/1            | -  | -   | -   | -  | [9]  |
| MMETSP1352 | <i>Stauroneis</i>      | <i>constricta</i>      | CCMP1120              | 20 | 40  | 6   | 14 | [5]  |
| MMETSP1361 | <i>Stauronira</i>      | <i>complex</i> sp.     | CCMP2646              | 8  | 10  | 6   | 7  | [5]  |
| MMETSP0794 | <i>Stephanopyxis</i>   | <i>turris</i>          | CCMP815               | 56 | 66  | 28  | 31 | [5]  |
| MMETSP1473 | <i>Stichococcus</i>    | <i>sp.</i>             | RCC1054               | 4  | 10  | 3   | 4  | [53] |
| MMETSP0800 | <i>Striatella</i>      | <i>unipunctata</i>     | CCMP2910              | 35 | 125 | 6   | 20 | [57] |
| MMETSP0126 | <i>Strombidinopsis</i> | <i>acuminatum</i>      | SPMC142               | -  | -   | -   | -  | [NA] |
| MMETSP0463 | <i>Strombidinopsis</i> | <i>sp.</i>             | <i>Sopsis</i> LIS2011 | -  | -   | -   | -  | [NA] |
| MMETSP0208 | <i>Strombidium</i>     | <i>inclinatum</i>      | S3                    | -  | -   | -   | -  | [NA] |
| MMETSP0449 | <i>Strombidium</i>     | <i>rassoulzadegani</i> | ras09                 | -  | -   | -   | -  | [NA] |
| MMETSP0447 | <i>Stygamoeba</i>      | <i>regulata</i>        | BSH-02190019          | -  | -   | -   | -  | [NA] |
| MMETSP0132 | <i>Symbiodinium</i>    | <i>kawagutii</i>       | CCMP2468              | 7  | 10  | 7   | 10 | [5]  |
| MMETSP0133 | <i>Symbiodinium</i>    | <i>kawagutii</i>       | CCMP2468              | 7  | 10  | 7   | 10 | [5]  |
| MMETSP1367 | <i>Symbiodinium</i>    | <i>sp.</i>             | C1                    | 9  | 12  | 9   | 12 | [3]  |
| MMETSP1370 | <i>Symbiodinium</i>    | <i>sp.</i>             | C15                   | 9  | 12  | 9   | 12 | [3]  |
| MMETSP1115 | <i>Symbiodinium</i>    | <i>sp.</i>             | CCMP2430              | 9  | 12  | 9   | 12 | [5]  |
| MMETSP1110 | <i>Symbiodinium</i>    | <i>sp.</i>             | CCMP421               | 10 | 14  | 10  | 14 | [5]  |
| MMETSP1374 | <i>Symbiodinium</i>    | <i>sp.</i>             | cladeA                | 9  | 12  | 9   | 12 | [3]  |
| MMETSP1377 | <i>Symbiodinium</i>    | <i>sp.</i>             | D1a                   | 9  | 12  | 9   | 12 | [3]  |
| MMETSP1122 | <i>Symbiodinium</i>    | <i>sp.</i>             | Mp                    | 9  | 12  | 9   | 12 | [3]  |
| MMETSP1452 | <i>Synchroma</i>       | <i>pusillum</i>        | CCMP3072              | 7  | 12  | 7   | 8  | [5]  |
| MMETSP1176 | <i>Synedropsis</i>     | <i>cf. recta</i>       | CCMP1620              | 13 | 96  | 2.5 | 4  | [57] |

|            |                         |                        |                        |     |      |     |    |      |
|------------|-------------------------|------------------------|------------------------|-----|------|-----|----|------|
| MMETSP0804 | <i>Tetraselmis</i>      | <i>astigmatica</i>     | CCMP880                | 5   | 10   | 5   | 15 | [5]  |
| MMETSP0491 | <i>Tetraselmis</i>      | <i>chuii</i>           | PLY429                 | 8   | 14   | 7   | 10 | [50] |
| MMETSP0419 | <i>Tetraselmis</i>      | <i>sp.</i>             | GSL018                 | -   | -    | -   | -  | [NA] |
| MMETSP0817 | <i>Tetraselmis</i>      | <i>striata</i>         | LANL1001               | 11  | 12   | 7   | 8  | [NA] |
| MMETSP0786 | <i>Thalassionema</i>    | <i>frauenfeldii</i>    | CCMP1798               | 56  | 60   | 4   | 6  | [5]  |
| MMETSP0158 | <i>Thalassionema</i>    | <i>nitzschioides</i>   | L26-B                  | 10  | 110  | 2   | 4  | [57] |
| MMETSP0693 | <i>Thalassionema</i>    | <i>nitzschioides</i>   | no strain              | 10  | 110  | 2   | 4  | [57] |
| MMETSP0902 | <i>Thalassiosira</i>    | <i>antarctica</i>      | CCMP982                | 14  | 26   | 10  | 12 | [5]  |
| MMETSP0492 | <i>Thalassiosira</i>    | <i>gravida</i>         | Gmp14c1                | 62  | 62   | 17  | 17 | [57] |
| MMETSP0737 | <i>Thalassiosira</i>    | <i>miniscula</i>       | CCMP1093               | 10  | 20   | 10  | 14 | [5]  |
| MMETSP0970 | <i>Thalassiosira</i>    | <i>oceanica</i>        | CCMP1005               | 4   | 12   | 6   | 10 | [5]  |
| MMETSP1067 | <i>Thalassiosira</i>    | <i>punctigera</i>      | Tpunct2005C2           | 40  | 186  | 10  | 23 | [57] |
| MMETSP0403 | <i>Thalassiosira</i>    | <i>rotula</i>          | CCMP3096               | 12  | 25   | 15  | 20 | [5]  |
| MMETSP0910 | <i>Thalassiosira</i>    | <i>rotula</i>          | GSO102                 | 8   | 55   | 18  | 24 | [57] |
| MMETSP1059 | <i>Thalassiosira</i>    | <i>sp.</i>             | FW                     | -   | -    | -   | -  | [NA] |
| MMETSP1071 | <i>Thalassiosira</i>    | <i>sp.</i>             | NH16                   | -   | -    | -   | -  | [NA] |
| MMETSP0898 | <i>Thalassiosira</i>    | <i>weissflogii</i>     | CCMP1010               | 13  | 18   | 10  | 12 | [5]  |
| MMETSP0878 | <i>Thalassiosira</i>    | <i>weissflogii</i>     | CCMP1336               | 12  | 22   | 10  | 12 | [5]  |
| MMETSP0152 | <i>Thalassiothrix</i>   | <i>antarctica</i>      | L6-D1                  | 420 | 5680 | 1.5 | 6  | [2]  |
| MMETSP0225 | <i>Thoracosphaera</i>   | <i>heimii</i>          | CCCM670<br>(=CCMP1069) | 8   | 12   | 8   | 12 | [7]  |
| MMETSP0198 | <i>Thraustochytrium</i> | <i>sp.</i>             | LLF1b                  | 4   | 20   | 4   | 20 | [3]  |
| MMETSP0472 | <i>Tiarina</i>          | <i>fuscus</i>          | LIS                    | 74  | 74   | 24  | 24 | [45] |
| MMETSP1172 | <i>Timspurckia</i>      | <i>oligopyrenoides</i> | CCMP3278               | 7   | 11   | 7   | 11 | [5]  |
| MMETSP0224 | <i>Togula</i>           | <i>jolla</i>           | CCCM725                | 25  | 43   | 19  | 35 | [7]  |
| MMETSP1175 | <i>Triceratium</i>      | <i>dubium</i>          | CCMP147                | 20  | 40   | 10  | 20 | [5]  |



|            |                        |                       |                    |    |    |    |    |    |      |
|------------|------------------------|-----------------------|--------------------|----|----|----|----|----|------|
| MMETSP0405 | <i>Trichosphaerium</i> | <i>sp</i>             | Am-I-7 wt          | -  | -  | -  | -  | -  | [NA] |
| MMETSP0098 | <i>Undescribed</i>     | <i>Undescribed</i>    | NY0313808BC1       | -  | -  | -  | -  | -  | [NA] |
| MMETSP1433 | <i>Undescribed</i>     | <i>Undescribed</i>    | NY07348D           | -  | -  | -  | -  | -  | [NA] |
| MMETSP1317 | <i>Undescribed</i>     | <i>Undescribed</i>    | Undescribed        | 3  | 5  | 3  | 5  | 5  | [3]  |
| MMETSP1178 | <i>unid</i>            | <i>sp.</i>            | CCMP2000           | 8  | 12 | 6  | 10 | 10 | [5]  |
| MMETSP1467 | <i>unid.</i>           | <i>sp.</i>            | CCMP2135           | 8  | 12 | 8  | 12 | 12 | [5]  |
| MMETSP1469 | <i>Unidentified</i>    | <i>sp.</i>            | CCMP1205           | 2  | 6  | 2  | 6  | 6  | [5]  |
| MMETSP1475 | <i>Unidentified</i>    | <i>sp.</i>            | CCMP1999           | 8  | 14 | 8  | 14 | 14 | [5]  |
| MMETSP1446 | <i>Unidentified</i>    | <i>sp.</i>            | CCMP2111           | 2  | 4  | 2  | 4  | 4  | [5]  |
| MMETSP1470 | <i>Unidentified</i>    | <i>sp.</i>            | CCMP2175           | 2  | 4  | 2  | 4  | 4  | [5]  |
| MMETSP1456 | <i>Unidentified</i>    | <i>sp.</i>            | RCC1871            | 2  | 2  | 2  | 2  | 2  | [53] |
| MMETSP1453 | <i>Unidentified</i>    | <i>sp.</i>            | RCC701             | 2  | 2  | 2  | 2  | 2  | [53] |
| MMETSP0086 | <i>Unknown</i>         | <i>Unknown</i>        | DI                 | -  | -  | -  | -  | -  | [NA] |
| MMETSP0018 | <i>Uronema</i>         | <i>sp.</i>            | Bbcil              | -  | -  | -  | -  | -  | [NA] |
| MMETSP0166 | <i>Vannella</i>        | <i>robusta</i>        | DIVA3 518/3/11/1/6 | 38 | 54 | 38 | 54 | 54 | [38] |
| MMETSP0168 | <i>Vannella</i>        | <i>sp.</i>            | DIVA3 517/6/12     | 38 | 54 | 38 | 54 | 54 | [38] |
| MMETSP0945 | <i>Vaucheria</i>       | <i>litorea</i>        | CCMP2940           | 40 | 50 | 40 | 50 | 50 | [5]  |
| MMETSP0946 | <i>Vaucheria</i>       | <i>litorea</i>        | CCMP2940           | 40 | 50 | 40 | 50 | 50 | [5]  |
| MMETSP0173 | <i>Vexillifera</i>     | <i>sp.</i>            | DIVA3 564/2        | 10 | 70 | 10 | 70 | 70 | [39] |
| MMETSP1451 | <i>Vitrella</i>        | <i>brassicaformis</i> | CCMP3346           | 4  | 46 | 4  | 46 | 46 | [5]  |

References Cited in Supplemental Table 1, appear as numbers in the left most column.

- 1 Antarctic Marine Protist Keys:  
[https://taxonomic.aad.gov.au/keys/ciliate/key/Antarctic%20Marine%20Ciliates/Media/Html/Mesodinium\\_pulex.htm](https://taxonomic.aad.gov.au/keys/ciliate/key/Antarctic%20Marine%20Ciliates/Media/Html/Mesodinium_pulex.htm).
- 2 Antarctic Marine Protist Keys:  
[https://taxonomic.aad.gov.au/keys/diatoms/key/Antarctic%20Marine%20Diatoms/Media/Html/Thalassiothrix\\_antarctica.htm](https://taxonomic.aad.gov.au/keys/diatoms/key/Antarctic%20Marine%20Diatoms/Media/Html/Thalassiothrix_antarctica.htm).
- 3 Approximate based on others species in the MMETSP dataset.
- 4 Ashworth, M. P., E. C. Ruck, C. S. Lobban, D. K. Romanovicz, and E. C. Theriot. 2012. A revision of the genus *Cyclophora* and description of *Astrosyne gen. nov.*(Bacillariophyta), two genera with the pyrenoids contained within pseudosepta. *Phycologia* 51:684-699.
- 5 Bigelow National Center for Marine Algae and Microbiota Website:  
<https://ncma.bigelow.org/>.
- 6 Birkhead, M., & Pienaar, R. N. (1994). The ultrastructure of *Chrysochromulina brevifilum* (Prymnesiophyceae). *European Journal of Phycology*, 29(4), 267-280.
- 7 Canadian Center for the Culture of Microorganisms (CCCM) Website:  
[www3.botany.ubc.ca/cccm/](http://www3.botany.ubc.ca/cccm/).
- 8 Cawthorn, R. J., D. Lynn, B. Despres, R. MacMillan, R. Maloney, M. Loughlin, and R. Bayer. 1996. Description of *Anophryoides haemophila n. sp.*(Scuticociliatida: Orchitophryidae), a pathogen of American lobsters *Homarus americanus*. *Diseases of Aquatic Organisms* 24:143-148.
- 9 Culture Collection of Algae and Protozoa (CCAP) Website:  
<https://www.ccap.ac.uk/>.
- 10 Culture Collection of Algae at the University of Göttingen, Germany (SAG) Website from picture: <http://sagdb.uni-goettingen.de/>.
- 11 Dungan, C. F., and R. M. Mamilton. 1995. Use of a Tetrazolium-based Cell Proliferation Assay to Measure Effects of In Vitro Conditions on *Perkinsus marinus* (Apicomplexa) Proliferation. *Journal of Eukaryotic Microbiology* 42:379-388.
- 12 Fernandes, N. M., R. J. Dias, C. G. Schrago, and I. D. Silva-Neto. 2015. Redescription and Phylogenetic Position of *Condylostoma arenarium* Spiegel, 1926 (Ciliophora, Heterotrichea) from Guanabara Bay, Brazil. *Journal of Eukaryotic Microbiology* 62:722-732.

- 13 Fischer, M. G., & Hackl, T. (2016). Host genome integration and giant virus-induced reactivation of the virophage mavirus. *Nature*, 540(7632), 288-291.
- 14 Giese, A. C. 1973. *Blepharisma*: the biology of a light-sensitive protozoan. Stanford University Press.
- 15 Green, J., D. Hibberd, and R. Pienaar. 1982. The taxonomy of *Prymnesium* (Prymnesiophyceae) including a description of a new cosmopolitan species, *P. patellifera* sp. nov., and further observations on *P. parvum* N. Carter. *British Phycological Journal* 17:363-382.
- 16 Guella, G., D. Skropeta, G. Di Giuseppe, and F. Dini. 2010. Structures, biological activities and phylogenetic relationships of terpenoids from marine ciliates of the genus *Euplotes*. *Marine drugs* 8:2080-2116.
- 17 Guella, G., R. Frassanito, I. Mancini, T. Sandron, L. Modeo, F. Verni, F. Dini, and G. Petroni. 2010. Keronopsamides, a new class of pigments from marine ciliates. *European Journal of Organic Chemistry* 2010:427-434.
- 18 Hirakawa, Y., A. Howe, E. R. James, and P. J. Keeling. 2011. Morphological Diversity between Culture Strains of a Chlorarachniophyte, *Lotharella globosa*. *PLoS ONE* 6:e23193.
- 19 Holtermann, K. E., S. S. Bates, V. L. Trainer, A. Odell, and E. Virginia Armbrust. 2010. Mass sexual reproduction in the toxigenic diatoms *Pseudo-nitzschia australis* and *P. pungens* (Bacillariophyceae) on the Washington coast, USA. *Journal of Phycology* 46:41-52.
- 20 Iwata, I., Kimura, K., Tomaru, Y., Motomura, T., Koike, K., Koike, K., & Honda, D. (2017). Bothrosome Formation in *Schizochytrium aggregatum* (Labyrinthulomycetes, Stramenopiles) during Zoospore Settlement. *Protist*, 168(2), 206-219.
- 21 Kies, L. (1989). Ultrastructure of *Cyanoptyche gloeocystis* f. *dispersa* (Glaucocystophyceae). *Plant systematics and evolution*, 164(1), 65-73.
- 22 Kim, J. H., and K. S. Mann. 2015. Novel Discovery of Two Heterotrichid Ciliates, *Climacostomum virens* and *Fabrea salina* (Ciliophora: Heterotrichea: Heterotrichida) in Korea. *Animal Systematics, Evolution and Diversity* 31:182.
- 23 Kim, S. J., & Min, G. S. (2009). Taxonomic study of poorly-known marine pleurostomatid ciliates of *Litonotus paracygnus* and *L. pictus* (Ciliophora: Pleurostomatida) from Korea. *Animal Systematics, Evolution and Diversity*, 25(2), 167-178.

- 24 Kudela Lab Phytoplankton Identification: <http://oceandatacenter.ucsc.edu/PhytoGallery/Diatoms/Corethron.html>.
- 25 Kudela Lab Phytoplankton Identification: <http://oceandatacenter.ucsc.edu/PhytoGallery/Diatoms/proboscia.html>.
- 26 Lane, C. E. and Archibald, J. M. (2008). New marine members of the genus *Hemiselmis* (Cryptomonadales, Cryptophyceae), *Journal of Phycology*, 44: 439–450. doi:10.1111/j.1529-8817.2008.00486.x - Approximately based.
- 27 Leadbeater, B. S. C. 2015. *The Choanoflagellates*. Cambridge University Press.
- 28 Leander, C. A., D. Porter, and B. S. Leander. 2004. Comparative morphology and molecular phylogeny of aplanochytrids (Labyrinthulomycota). *European Journal of Protistology* 40:317-328.
- 29 Lee, W. J. (2015). Small free-living heterotrophic flagellates from marine sediments of Gippsland Basin, South-Eastern Australia. *Acta Protozoologica*, 54(1), 53. Approximately based.
- 30 Lobban, C.S., Honda, D., Chihara, M. & Schefter, M. (1995). *Chrysocystis fragilis* gen. nov., sp. nov. (Chrysophyceae, Sarcinochrysidales), with notes on other macroscopic chrysophytes (golden algae) on Guam reefs. *Micronesica* 28:91-102.
- 31 MacDonald, S. M., and R. W. Lee. 2016. A survey of *Polytomella* (Chlorophyceae, Chlorophyta) strains in public culture collections. *Journal of Phycology* 52:656-663.
- 32 Marchant, H. J., K. R. Buck, D. L. Garrison, and H. A. Thomsen. 1989. *Mantoniella* in Antarctic waters including the description of *M. antarctica* sp. nov. (Prasinophyceae). *Journal of Phycology* 25:167-174.
- 33 Marin and M. Melkonian (2010): Molecular Phylogeny and Classification of the Mamiellophyceae *class. nov.* (Chlorophyta) based on Sequence Comparisons of the Nuclear- and Plastid-encoded rRNA Operons. *Protist* 161, 304-336.
- 34 Marine Species Identification Portal: [http://species-identification.org/species.php?species\\_group=zsao&id=1462](http://species-identification.org/species.php?species_group=zsao&id=1462).
- 35 Microbe World: World of Amoeboid Organisms: <http://www.arcella.nl/filamoebanollandi>.
- 36 Microbe World: World of Amoeboid Organisms: <http://www.arcella.nl/mayorella>.
- 37 Microbe World: World of Amoeboid Organisms: <http://www.arcella.nl/paramoeba>.

- 38 Microbe World: World of Amoeboid Organisms:  
<http://www.arcella.nl/sapocribrum>.
- 39 Microbe World: World of Amoeboid Organisms: <http://www.arcella.nl/vannella>.
- 40 Microbe World: World of Amoeboid Organisms: <http://www.arcella.nl/vexillifera>.
- 41 Moro, I., Negrisolo, E., Callegaro, A., & Andreoli, C. (2003). *Aplanochytrium stocchinoi*: a new Labyrinthulomycota from the southern ocean (Ross Sea, Antarctica). *Protist*, 154(3-4), 331-340.
- 42 Mortensen, H. D., Gori, K., Jespersen, L. and Arneborg, N. (2005), *Debaryomyces hansenii* strains with different cell sizes and surface physicochemical properties adhere differently to a solid agarose surface. *FEMS Microbiology Letters*, 249: 165–170. doi:10.1016/j.femsle.2005.06.009.
- 43 Nannotax Website:  
<http://www.mikrotax.org/Nannotax3/index.php?dir=Coccolithophores/Isochrysidales/Noelaerhabdaceae/Gephyrocapsa/Gephyrocapsa%20oceanica&page=all>.
- 44 Nordic Microalgae and aquatic protozoa Website. Approximate.  
<http://nordicmicroalgae.org/taxon/Chrysochromulina%20polylepis>.
- 45 Nordic Microalgae and aquatic protozoa Website.  
<http://nordicmicroalgae.org/taxon/Chrysochromulina%20rotalis>.
- 46 Ocean Biogeographic Information System: *Tiarina fusus* Bergh, 1881  
<http://iobis.org/explore/#/taxon/761529>.
- 47 Ota, S., K. Ueda, and K.-i. Ishida. 2007. *Norrisiella sphaerica* gen. et sp. nov., a new coccoid chlorarachniophyte from Baja California, Mexico. *Journal of plant research* 120:661-670.
- 48 Dunthorn, M., Eppinger, M., Schwarz, M. J., Schweikert, M., Boenigk, J., Katz, L. A., & Stoeck, T. (2009). Phylogenetic placement of the Cyrtolophosididae Stokes, 1888 (Ciliophora; Colpodea) and neotypification of *Aristerostoma marinum* Kahl, 1931. *International Journal of Systematic and Evolutionary Microbiology*, 59(1), 167-180.
- 49 Phyto'pedia - The Phytoplankton Encyclopaedia Project:  
[https://www.eoas.ubc.ca/research/phytoplankton/dinoflagellates/ceratium/c\\_fusus.html](https://www.eoas.ubc.ca/research/phytoplankton/dinoflagellates/ceratium/c_fusus.html).
- 50 Phyto'pedia - The Phytoplankton Encyclopaedia Project:  
[https://www.eoas.ubc.ca/research/phytoplankton/dinoflagellates/dinophysis/d\\_acuminata.html](https://www.eoas.ubc.ca/research/phytoplankton/dinoflagellates/dinophysis/d_acuminata.html).

- 51 Plymouth Biological Lab Website: [www.mba.ac.uk/facilities/culture-collection](http://www.mba.ac.uk/facilities/culture-collection).
- 52 Quijano-Scheggia, S. I., E. Garcés, N. Lundholm, Ø. Moestrup, K. Andree, and J. Camp. 2009. Morphology, physiology, molecular phylogeny and sexual compatibility of the cryptic *Pseudo-nitzschia delicatissima* complex (Bacillariophyta), including the description of *P. arenysensis* sp. nov. *Phycologia* 48:492-509.
- 53 Repak, A. J. 1972. A Redescription of *Climacostomum virens* (Ehrenberg) Stein and Proposal of a New Heterotrich Ciliate Family, Climacostomidae fam. n. *The Journal of Protozoology* 19:417-427.
- 54 Roscoff-Culture Collection Website: <http://roscoff-culture-collection.org/>.
- 55 Rousseau, V., D. Vaultot, R. Casotti, V. Cariou, J. Lenz, J. Gunkel, and M. Baumann. 1994. The life cycle of *Phaeocystis* (Prymnesiophyceae): Evidence and Hypotheses. *Journal of Marine Systems* 5:23-39.
- 56 Sarno, D., W. H. C. F. Kooistra, L. K. Medlin, I. Percopo, and A. Zingone. 2005. Diversity in the genus *Skeletonema* (Bacillariophyceae). Ii. An assessment of the taxonomy of *S. costatum*-like species with the description of four new species. *Journal of Phycology* 41:151-176.
- 57 Tice, A. K., L. L. Shadwick, A. M. Fiore-Donno, S. Geisen, S. Kang, G. A. Schuler, F. W. Spiegel, K. A. Wilkinson, M. Bonkowski, K. Dumack, D. J. G. Lahr, E. Voelcker, S. Clauß, J. Zhang, and M. W. Brown. 2016. Expansion of the molecular and morphological diversity of Acanthamoebidae (Centramoebida, Amoebozoa) and identification of a novel life cycle type within the group. *Biology Direct* 11:69.
- 58 Tomas, C. R. 1997. *Identifying marine phytoplankton*. Academic press.
- 59 Valbonesi, A., and P. Luporini. 1993. Biology of *Euplotes focardii*, an Antarctic ciliate. *Polar Biology* 13:489-493.

Supplemental Table 2 A. Accession Numbers from SILVA database representing 18S rRNA sequences that corresponded to reference genomes (n=48) along with additional taxa (n=140) from SILVA

| Species Names and SILVA Identifiers  |  |
|--|--|
| <i>Monoraphidium neglectum</i> AJ300526.1.1766                             |  |
| <i>Acanthamoeba healyi</i> CDF A01025657.77.1694                           |  |
| <i>Acanthamoeba castellanii</i> ATCC30010 EF554328.1                       |  |
| <i>Acanthogorgia</i> sp AF052907.1.1852                                    |  |
| <i>Agaricus bisporus</i> var. <i>bisporus</i> H97 AEOK01000180.16575.18364 |  |
| <i>Albugo candida</i> KF853245.1.1786                                      |  |
| <i>Alexandrium affine</i> AY775286.1.1789                                  |  |
| <i>Alexandrium andersonii</i> KF925334.1.1640                              |  |
| <i>Alexandrium australiense</i> KF908802.1.1672                            |  |
| <i>Amphimedon compressa</i> EU702409.1.1808                                |  |
| <i>Angomonas deanei</i> EU079129.1.2137                                    |  |
| <i>Anolis carolinensis</i> GADN01012067.648.2457                           |  |
| <i>Aplysia californica</i> AASC03160597.930.2770                           |  |
| <i>Arabidopsis thaliana</i> Columbia NR141642.1                            |  |
| <i>Arcuseries petzi</i> KC896648.1.1822                                    |  |
| <i>Ardissonea formosa</i> HQ912653.1.1618                                  |  |
| <i>Aspergillus nidulans</i> AB008403.1.1733                                |  |
| <i>Asterionella formosa</i> AM712617.1.1812                                |  |
| <i>Aulacoseira baicalensis</i> AY121821.1.1802                             |  |
| <i>Auxenochlorella protothecoides</i> KM020039.1.1775                      |  |
| <i>Auxenochlorella protothecoides</i> APJO01001325.1293.3115               |  |
| <i>Babesia divergens</i> AJ439713.1.1728                                   |  |
| <i>Babesia gibsoni</i> AB478322.1.1663                                     |  |
| <i>Bathycoccus prasinos</i> RCC1105 FO082268.37864.40071                   |  |
| <i>Bigeloviella natans</i> CCMP2755 ADNK01002499.9459.11245                |  |
| <i>Monosiga ovata</i> AF084230.1.1765                                      |  |
| <i>Monosiga brevicollis</i> AF174375.1                                     |  |
| <i>Montastraea franksi</i> AY026382.1.1838                                 |  |
| <i>Nannochloropsis gaditana</i> AZIL01002195.4319.6114                     |  |
| <i>Nematostella vectensis</i> ABAV01023297.1433.3227                       |  |
| <i>Neochloris aquatica</i> M62861.1.1794                                   |  |
| <i>Neospora caninum</i> U16159.1.1793                                      |  |
| <i>Nephroselmis anterostigmatica</i> AB158373.1.1755                       |  |
| <i>Nephroselmis rotunda</i> FR865651.2.1776                                |  |
| <i>Ornithorhynchus anatinus</i> AJ311679.1.1850                            |  |
| <i>Oryza sativa</i> Japonica Group cultivar Nipponbare                     |  |
| <i>Ostreococcus lucimarinus</i> CCE9901 CP000588.494341.496100             |  |
| <i>Ostreococcus tauri</i> CAID01000012.112153.113912                       |  |
| <i>Oswaldella grandis</i> KT722417.1.1843                                  |  |
| <i>Oxyrrhis</i> sp WHOIL11 5 6 AF330215.1.1759                             |  |
| <i>Oxytricha trifallax</i> AMCR01021323.1776.3536                          |  |
| <i>Pan troglodytes</i> AACZ04068423.75350.77214                            |  |
| <i>Paralia longispina</i> KJ577865.1.1677                                  |  |
| <i>Paramecium tetraurelia</i> AB252008.1.1745                              |  |
| <i>Parastrombidinopsis shimi</i> AJ786648.1.1805                           |  |
| <i>Pediastrum duplex</i> M62997.1.1800                                     |  |
| <i>Perkinsus chesapeakei</i> AF102171.1575.3382                            |  |
| <i>Perkinsus marinus</i> AF497479.1251.3059                                |  |

|  |  |
|--|--|
| <i>Blastocystis hominis Incertae</i> U51151.1.1770                 | <i>Perkinsus mediterraneus</i> AY487833.1.1772                         |
| <i>Bodo saltans</i> Kortsanz AF208889.1.2134                       | <i>Petromyzon marinus</i> GQ215662.8338.10120                          |
| <i>Botryococcus braunii</i> JQ315502.1.1798                        | <i>Phaeodactylum tricornerutum</i> CCAP1055 ABQD01000052.43555.45336/1 |
| <i>Branchiostoma lanceolatum</i> AY428817.1.1721                   | <i>Physcomitrella patens</i> ABEU01010493.4882.6690                    |
| <i>Chaetoceros muellerii</i> KC594685.1.1804                       | <i>Picochlorum eukaryotum</i> X06425.1.1796                            |
| <i>Chlamydomonas reinhardtii</i> ABCN01009239.1.1424               | <i>Plagiomonas amylosa</i> AF143944.1.1738                             |
| <i>Chlorella vulgaris</i> KJ756823.1.1801                          | <i>Plagioselmis nannoplanctica</i> FM876311.1.1770                     |
| <i>Chromera</i> v elia CCMP2878 CDMZ01005019.70962.72740           | <i>Poecilia reticulata</i> KJ774732.1.1825                             |
| <i>Chroomonas vectensis</i> HM126534.1.1744                        | <i>Polyphondylium pallidum</i> AM168104.1.1940                         |
| <i>Chrysochromulina parva</i> AM491019.1.1792                      | <i>Porosira pseudodenticulata</i> DQ436461.1.1805                      |
| <i>Chrysotila lamellosa</i> KC888129.1.1731                        | <i>Prasinopapilla vacuolata</i> AB183649.1.1758                        |
| <i>Ciona intestinalis</i> GBKV01014989.349.2148                    | <i>Prorocentrum shikokuense</i> AB781324.1.1745                        |
| <i>Coccomyxa viridis</i> AJ302939.1.1798                           | <i>Prototheca wickerhamii</i> X74003.1.1802                            |
| <i>Coccomyxa subellipsoidea</i> C 169 AGSI01000004.2976164.2978397 | <i>Pseudochattonella verruculosa</i> AM075625.1.1816                   |
| <i>Coelastrum microporum</i> JQ315527.1.1793                       | <i>Pseudohaptolina arctica</i> AM491016.1.1802                         |
| <i>Coscinodiscus radiatus</i> X77705.1.1811                        | <i>Pseudoisochrysis paradoxa</i> AM490999.1.1721                       |
| <i>Crassostrea gigas</i> AB064942.1.1820                           | <i>Pseudoscoirfieldia marina</i> X75565.22.1770                        |
| <i>Crocodylus niloticus</i> AJ311672.1.1776                        | <i>Pterosperma cristatum</i> AB017127.1.1753                           |
| <i>Cyanophora paradoxa</i> AY823716.1.1717                         | <i>Pycnococcus provasolii</i> AF122889.1.1755                          |
| <i>Cyclotella choctawhatcheeana</i> AM712618.1.1803                | <i>Pyramimonas gelidicola</i> EU141942.1.1806                          |
| <i>Cymbomonas tetramitiformis</i> FN562438.1.1785                  | <i>Pyramimonas olivacea</i> FN562442.1.1786                            |
| <i>Cystoisospora ohioensis</i> GU292304.1.1731                     | <i>Pyramimonas propulsa</i> AB017123.1.1758                            |
| <i>Debaromyces hansenii</i> var <i>hansenii</i> AB013590.1.1784    | <i>Pyramimonas tetrarhynchus</i> FN562441.1.1784                       |
| <i>Dicrateria</i> sp ALGO HAP49 AM490997.1.1721                    | <i>Pyrenomonas salina</i> X54276.1.1762                                |
| <i>Dimeregramma minor</i> var <i>nanum</i> AB430598.1.1862         | <i>Pythium ultimum</i> DAOM BR144 ADOS01001233.4029.5837               |



|   |   |
|---|---|
| <i>Dinophysis infundibulum</i> AB366002.1.1740            | <i>Reticulomyxa filosa</i> isolate 613 AJ132367.1               |
| <i>Dinophysis norvegica</i> AJ506974.1.1802               | <i>Rhabdonema arcuatum</i> JN975251.1.1625                      |
| <i>Diophrys apoligothrix</i> EU189068.1.1784              | <i>Rhinomonas pauca</i> U53132.1.1770                           |
| <i>Durinskia agilis</i> JF514515.1.1780                   | <i>Rhizosolenia formosa</i> JX413557.1.1644                     |
| <i>Ectocarpus siliculosus</i> CABU01003659.782.2081       | <i>Saccharomyces cerevisiae</i> KR028986.1.1729                 |
| <i>Eimeria alabamensis</i> AB769547.1.1781                | <i>Scirppsiella sweeneyae</i> HQ845331.1.1756                   |
| <i>Emiliana huxleyi</i> CCMPI516 AHAL01000301.1474.3265   | <i>Selaginella moellendorffii</i> ADFJ01002373.6706.8505        |
| <i>Eudorina elegans</i> LC086351.1.1744                   | <i>Sicyodochytrium minutum</i> AB355412.1.1733                  |
| <i>Falcomonas daucooides</i> AF143943.1.1734              | <i>Skeletonema grevillei</i> DQ396512.1.1798                    |
| <i>Gallus gallus</i> AADN04002033.13735.15546             | <i>Skeletonema tropicum</i> DQ396513.1.1809                     |
| <i>Gephyrocapsa muelleriae</i> KP282839.1.1755            | <i>Sordaria fimicola</i> X69851.1.1796                          |
| <i>Gonium pectorale</i> FR865741.2.1763                   | <i>Spironucleus salmonicida</i> ATCC50380 DQ812526.1            |
| <i>Gorilla gorilla gorilla</i> CABD030100652.45.1761      | <i>Stephanopyxis nipponica</i> M87330.1.1805                    |
| <i>Guillardia theta</i> CCMPT2712 AEIE01002671.389.2153   | <i>Streatula major</i> U53130.1.1768                            |
| <i>Guinardia delicatula</i> HQ912651.1.1660               | <i>Strigomonas culicis</i> U05679.1.2101                        |
| <i>Gymnodinium instriatum</i> DQ847433.1.1790             | <i>Stylonychia lemnae</i> AF164124.1.1770                       |
| <i>Hammondia hammondi</i> AF096498.1.1747                 | <i>Symbiodinium</i> sp clade A KC848879.1.1510                  |
| <i>Haptolina brevifila</i> AB058358.1.1749                | <i>Symbiodinium</i> sp clade B KC848880.1.1512                  |
| <i>Haptolina ericina</i> AM491030.1.1787                  | <i>Symbiodinium</i> sp clade C KC848882.1.1512                  |
| <i>Haptolina fragaria</i> AM491013.1.1789                 | <i>Symbiodinium</i> sp clade D EF419282.1.1694                  |
| <i>Haptolina hirta</i> AJ246272.1.1787                    | <i>Symbiodinium</i> sp clade E AF238262.1.1693                  |
| <i>Helicosporidium</i> sp ATCC50920 SRR1019732.SRR1019731 | <i>Symbiodinium minutum</i> Mf1 BASF01015284.22116.24113.05b.01 |
| <i>Helobdella fusca</i> AY962414.1.1799                   | <i>Tabularia</i> sp NY059 LC054953.1.1856                       |
| <i>Hemiselmis brunnescens</i> AJ007282.1.1761             | <i>Teledulax amphioxeia</i> AJ007287.1.1750                     |
| <i>Holosticha heterofoissneri</i> KP717082.1.1785         | <i>Tetrahymena thermophila</i> M10932.164.1916                  |

|                                     |                                |                                      |                                     |
|-------------------------------------|--------------------------------|--------------------------------------|-------------------------------------|
| <i>Homo sapiens</i>                 | AADB02002333.14402.16240       | <i>Tetraselmis apiculata</i>         | KJ756817.1.1785                     |
| <i>Hypotrachidium paraconicum</i>   | JQ918371.1.1798                | <i>Tetraselmis carteriiiformis</i>   | FJ559384.1.1641                     |
| <i>Ichthyophthirius multifiliis</i> | AEDN01002274.4420.6159         | <i>Tetraselmis cordiformis</i>       | HE610130.1.1777                     |
| <i>Imantonia rotunda</i>            | AM491014.1.1723                | <i>Tetraselmis inconspicua</i>       | KJ756818.1.1795                     |
| <i>Isochrysis galbana</i>           | AJ246266.1.1800                | <i>Tetraselmis marina</i>            | HE610131.1.1771                     |
| <i>Isochrysis litoralis</i>         | AM490996.1.1721                | <i>Thalassiosira aestivalis</i>      | DQ093369.1.1806                     |
| <i>Isochrysis nuda</i>              | KC888117.1.1721                | <i>Thalassiosira anguste-lineata</i> | AJ810854.1.1810                     |
| <i>Isospora sp Tokyo</i>            | AB757860.1.1676                | <i>Thalassiosira curviseriata</i>    | AJ810859.1.1809                     |
| <i>Jasus tristani</i>               | AF498664.1.1718                | <i>Thalassiosira punctigera</i>      | HM991699.1.1812                     |
| <i>Karenia mikimotoi</i>            | AF022195.1.1803                | <i>Thalassiosira oceanica</i>        | AGNL01025219.4749.6540              |
| <i>Karenia papilionacea</i>         | HM067005.1.1741                | <i>Thalassiosira pseudonana</i>      | CCMP1335 AAFD02000029.651044.652838 |
| <i>Komma caudata</i>                | U53122.1.1774                  | <i>Thecamonas trahens</i>            | ATCC50062 XR001290156.1             |
| <i>Latimeria chalumnae</i>          | L11288.1.1779                  | <i>Theileria equi</i>                | AB515311.1.1509                     |
| <i>Lepisosteus osseus</i>           | X98837.1.1767                  | <i>Tisochrysis lutea</i>             | KC888125.1.1723                     |
| <i>Leptomonas pyrrocoris</i>        | LGTL01000034.290912.293116     | <i>Toxoplasma gondii</i>             | ME49 ABPA02001045.1910.3693         |
| <i>Licmophora flucticulata</i>      | HQ997923.1.1620                | <i>Tribolium castaneum</i>           | HM156711.1.1831                     |
| <i>Limnatis nilotica</i>            | AY425470.1.1901                | <i>Trichoplax adhaerens</i>          | L10828.1.1787                       |
| <i>Lithodesmium undulatum</i>       | Y10569.1.1806                  | <i>Ustilago maydis</i>               | KP322952.1.1609                     |
| <i>Lumbricus rubellus</i>           | Z83753.1.1816                  | <i>Vaucheria bursata</i>             | U41646.1.1831                       |
| <i>Macaca mulatta</i>               | AEHK01400485.183.2026          | <i>Vaucheria terrestris</i>          | AM490830.1.1771                     |
| <i>Micractinium reisseri</i>        | AB437244.1.2123                | <i>Vitrella brassicaformis</i>       | CCMP3155 CDMY01000449.33029.34795   |
| <i>Micromonas commoda</i>           | RCC299 KU612123.1              | <i>Volvox carteri fnagariensis</i>   | ACJH01008033.984.2764               |
| <i>Micromonas pusilla</i>           | CCMP1545 ACCP01000012.989.2759 | <i>Xenopus laevis</i>                | X02995.1030.2854                    |
| <i>Monodelphis domestica</i>        | AAFR03071841.2424.4306         | <i>Zea mays</i>                      | AC150267.44640.46440                |

Supplemental Table 2 B. Short Read Archive numbers and Run IDs corresponding to species that were missing 18S rRNA genes in the SILVA database. PhyloFlash was used with the raw Illumina data to reconstruct the 18S rRNA from these samples to add to the MMETSP phylogenetic tree.

| SRA Archives | SRA Run ID | Species                                  |
|--------------|------------|--|
| SRX1045719   | SRR2047475 | <i>Nannochloropsis oceanica</i> CCMP1779 |
| SRX236551    | SRR715315  | <i>Saprolegnia diclina</i>               |
| SRX208010    | SRR627744  | <i>Pyropia yezoensis</i> U51             |

Supplemental Table 3. Divergence time estimates between species. Data gathered from Website from TimeTree.org for representative groups indicated with (Hedges et al. 2015) [1], taken from (Liu et al. 2010) [2], taken from (Herron et al. 2009) [3], (Damsté et al. 2004) [4]

| Name                                   | Constrained Node  | Minim<br>um | Maxim<br>um |
|--|---|-------------|-------------|
| Eukarya                                | <i>Monosiga brevicollis</i> AF174375.1  | 1502        | 2035        |
| Red Algae - Green Algae Split          | <i>Skeletonema marinoi</i> skela MMETSP0918, <i>Erythrolobus australicus</i> CCMP3124 MMETSP1353                              | 1160        | 1630        |
| Diatom - Dinoflagellate Split          | <i>Skeletonema marinoi</i> skela MMETSP0918, <i>Alexandrium australiense</i> KF908802.1.1672                                  | 997         | 1470        |
| Cercozoa                               | <i>Minchinia chitonis</i> MMETSP0186, <i>Bigeloviella natans</i> CCMP1259 MMETSP1054  | 992         | -           |
| Rhodophyceae                           | <i>Rhodosorus marinus</i> CCMP769 MMETSP0011, <i>Erythrolobus australicus</i> CCMP3124 MMETSP1353                             | 916         | 1220        |
| Appearance of Haptophytes              | <i>Pavlova lutheri</i> RCC1537 JF714236.1 MMETSP1463, <i>Tisochoyris lutea</i> KC888125.1.1723                                | 637         | 1031        |
| Origin of Fungi                        | <i>Agaricus bisporus var bisporus</i> H97 AEOK01000180.16575.18364, <i>Debaryomyces hansenii var hansenii</i> AB013590.1.1784 | 603         | 844         |
| Viridiplantae                          | <i>Zea mays</i> AC150267.44640.46440, <i>Physcomitrella patens</i> ABEU01010493.4882.6690                                     | 481         | 584         |
| Stramenopiles                          | <i>Skeletonema marinoi</i> skela MMETSP0918, <i>Schizochytrium aggregatum</i> ATCC28209 AB022106 MMETSP0962                   | 416         | 898         |
| Chrysophyceae appearance               | <i>Heterosigma akashino</i> CCMP3107 MMETSP0409, <i>Chrysophyceae</i> sp CCMP2298 MMETSP1141                                  | 416         | -           |
| Ascomycete                             | <i>Sordaria fimicola</i> X69851.1.1796, <i>Debaryomyces hansenii var hansenii</i> AB013590.1.1784                             | 400         | -           |
| <i>Pelagomonas - Florenciella</i>      | <i>Pelagomonas calceolata</i> RCC969 MMETSP1328, <i>Florenciella</i> sp RCC1587 MMETSP1324                                    | 397         | -           |
| <i>Micromonas - Ostreococcus</i> split | <i>Micromonas pusilla</i> CCMP1723 MMETSP1403, <i>Ostreococcus mediterraneus</i> clade D RCC2573 MMETSP0936                   | 333         | 639         |
| Origin of Calcification in Haptophytes | <i>Scyphosphaera apsteinii</i> RCC1455 MMETSP1333, <i>Tisochoyris lutea</i> KC888125.1.1723                                   | 291         | 329         |
| Phaeocyceae                            | <i>Phaeocystis antarctica</i> CaronLabsolate MMETSP1100, <i>Phaeocystis cordata</i> RCC1383 MMETSP1465                        | 248         | 428         |
| Karenia-Alexandrium split              | <i>Karenia brevis</i> CCMP2229 MMETSP0027, <i>Alexandrium australiense</i> KF908802.1.1672                                    | 240         | -           |
| <i>Mallomonas - Chromulina</i> split   | <i>Mallomonas</i> sp CCMP3275 MMETSP1167, <i>Chromulina nebulosa</i> UTEXLB2642 MMETSP1095                                    | 228         | 443         |
| <i>Gonyaulax-Alexandrium</i> Split     | <i>Gonyaulax spinifera</i> CCMP409 MMETSP1439, <i>Alexandrium australiense</i> KF908802.1.1672                                | 196         | -           |
| Split of Volvox-Gonium                 | <i>Volvox carteri f nagariensis</i> ACJH01008033.984.2764, <i>Gonium pectorale</i> FR865741.2.1763                            | 170         | 210         |
| Angiosperms                            | <i>Zea mays</i> AC150267.44640.46440, <i>Arabidopsis thaliana</i> Columbia NR141642.1   | 133.9       | -           |
| Appearance of Diatoms                  | <i>Skeletonema marinoi</i> skela MMETSP0918, <i>Stephanopyxis nipponica</i> M87330.1.1805                                     | 120         | 322         |
| <i>Chaetoceros</i> genus               | <i>Chaetoceros affinis</i> CCMP159 MMETSP0088, <i>Chaetoceros</i> sp GSL56 MMETSP0200   | 91          | -           |
| Rise of Rhizosolenia                   | <i>Rhizosolenia setigera</i> CCMP1694 MMETSP0789, <i>Corethron lysitrix</i> 308 MMETSP0010                                    | 90          | 93          |
| Bacillariophyceae                      | <i>Skeletonema marinoi</i> skela MMETSP0918, <i>Ardissonaea formosa</i> HQ912653.1.1618                                       | 80          | -           |
| <i>Emiliania - Tisochoyris</i>         | <i>Emiliania huxleyi</i> CCMP1516.AHAL01000301.1474.3265, <i>Tisochoyris lutea</i> KC888125.1.1723                            | 66          | 201         |
| MRCA <i>Calcidiscus - Cocolithus</i>   | <i>Calcidiscus leptoporus</i> RCC1130 MMETSP1334, <i>Cocolithus braarudi</i> PLY182g MMETSP0164                               | 41.3        | 119.4       |
| Bacillariophyceae Family               | <i>Fragilariopsis kerguelensis</i> L26 C5 MMETSP0733, <i>Pseudo-nitzschia pungens cf cingulata</i> MMETSP1060                 | 10          | 35          |

Supplementary Table 4.

Taxonomy of the MMETSP dataset, including the number of proteins recovered, the number of BUSCOs for each proteome, and the number of surviving orthologs from the Yang and Smith pipeline prior to orthologs filtering (Yang Homologs) and then after filtering for single copy orthologs (Yang Orthologs).

| Phylum                | Genus                   | Species                     | Strain            | MMETSP ID  | # Predicted Proteins | # of BUSCOs | Yang Smith Homologs | Yang Smith Orthologs |
|-----------------------|-------------------------|-----------------------------|-------------------|------------|----------------------|-------------|---------------------|----------------------|
| <b>Alveolata</b>      |                         |                             |                   |            |                      |             |                     |                      |
| <b>Apicomplexa</b>    |                         |                             |                   |            |                      |             |                     |                      |
| Conoidasida           | <i>Lankesteria</i>      | <i>abbottii</i>             | GrapplerInletBC   | MMETSP0372 | 6682                 | 128         | 133                 | 1                    |
| <b>Chromerida</b>     |                         |                             |                   |            |                      |             |                     |                      |
| Chromera              | <i>Chromera</i>         | <i>velia</i>                | CCMP2878          | MMETSP0290 | 16343                | 257         | 519                 | 25                   |
| Vitrella              | <i>Vitrella</i>         | <i>brassicaformis</i>       | CCMP3155          | MMETSP0288 | 15802                | 243         | 616                 | 37                   |
| Vitrella              | <i>Vitrella</i>         | <i>brassicaformis</i>       | CCMP3346          | MMETSP1451 | 16576                | 223         | 571                 | 36                   |
| <b>Ciliophora</b>     |                         |                             |                   |            |                      |             |                     |                      |
| Intramacronucleata    | <i>Anophryoides</i>     | <i>haemophila</i>           | AH6               | MMETSP1018 | 8091                 | -           | -                   | 17                   |
| Intramacronucleata    | <i>Aristerostoma</i>    | <i>sp.</i>                  | ATCC50986         | MMETSP0125 | 20764                | 174         | 400                 | 13                   |
| Intramacronucleata    | <i>Euplotes</i>         | <i>crassus</i>              | CT5               | MMETSP1380 | 12393                | 220         | 611                 | 29                   |
| Intramacronucleata    | <i>Euplotes</i>         | <i>focardi</i>              | TN1               | MMETSP0205 | 14951                | 158         | 1044                | 26                   |
| Intramacronucleata    | <i>Euplotes</i>         | <i>harpa</i>                | FSP1-4            | MMETSP0213 | 17120                | 208         | 545                 | 27                   |
| Intramacronucleata    | <i>Favella</i>          | <i>ehrenbergii</i>          | Fehren1           | MMETSP0123 | 17206                | 215         | 650                 | 28                   |
| Intramacronucleata    | <i>Litonotus</i>        | <i>pictus</i>               | P1                | MMETSP0209 | 13922                | 237         | 57                  | 1                    |
| Intramacronucleata    | <i>Mesodinium</i>       | <i>pulex</i>                | SPMC105           | MMETSP0467 | 50621                | 183         | 1220                | 51                   |
| Intramacronucleata    | <i>Myrionecta</i>       | <i>rubra</i>                | CCMP2563          | MMETSP0798 | 27318                | 190         | 889                 | 24                   |
| Intramacronucleata    | <i>Platyophrya</i>      | <i>macrostoma</i>           | WH                | MMETSP0127 | 28507                | 294         | 909                 | 38                   |
| Intramacronucleata    | <i>Protocruzia</i>      | <i>adherens</i>             | Boccale           | MMETSP0216 | 35390                | 286         | 277                 | 10                   |
| Intramacronucleata    | <i>Pseudokeronopsis</i> | <i>sp.</i>                  | Brazil            | MMETSP1396 | 14294                | 207         | 1144                | 48                   |
| Intramacronucleata    | <i>Pseudokeronopsis</i> | <i>sp.</i>                  | OXSARD2           | MMETSP0211 | 16437                | 207         | 767                 | 32                   |
| Intramacronucleata    | <i>Schmidingerella</i>  | <i>taraikaensis</i>         | FeNarragansettBay | MMETSP0434 | 10801                | 78          | 802                 | 20                   |
| Intramacronucleata    | <i>Strombidinopsis</i>  | <i>acuminatum</i>           | SPMC142           | MMETSP0126 | 34625                | 262         | 2024                | 65                   |
| Intramacronucleata    | <i>Strombidinopsis</i>  | <i>sp.</i>                  | SopsisLIS2011     | MMETSP0463 | 20690                | 238         | 1050                | 38                   |
| Intramacronucleata    | <i>Strombidium</i>      | <i>inclinatum</i>           | S3                | MMETSP0208 | 17497                | 240         | 611                 | 30                   |
| Intramacronucleata    | <i>Strombidium</i>      | <i>rassoulzadegani</i>      | ras09             | MMETSP0449 | 10628                | 211         | 673                 | 29                   |
| Intramacronucleata    | <i>Tiarina</i>          | <i>fuscus</i>               | LIS               | MMETSP0472 | 43727                | 302         | 3158                | 265                  |
| Intramacronucleata    | <i>Uronema</i>          | <i>sp.</i>                  | Bbcil             | MMETSP0018 | 9532                 | 59          | 410                 | 22                   |
| Postciliodesmatophora | <i>Blepharisma</i>      | <i>japonicum</i>            | StockR1072        | MMETSP1395 | 12414                | 200         | 616                 | 22                   |
| Postciliodesmatophora | <i>Climacostomum</i>    | <i>virens</i>               | StockW-24         | MMETSP1397 | 11204                | 247         | 411                 | 16                   |
| Postciliodesmatophora | <i>Condyllostoma</i>    | <i>magnum</i>               | COL2              | MMETSP0210 | 17726                | 109         | 329                 | 12                   |
| Postciliodesmatophora | <i>Fabrea</i>           | <i>salina</i>               |                   | MMETSP1345 | 8248                 | 222         | 449                 | 22                   |
| Postciliodesmatophora | <i>Parduzia</i>         | <i>like-sp. Undescribed</i> |                   | MMETSP1317 | 6092                 | 35          | 105                 | -                    |
| <b>Dinophyceae</b>    |                         |                             |                   |            |                      |             |                     |                      |
| Dinophysiales         | <i>Dinophysis</i>       | <i>acuminata</i>            | DAEP01            | MMETSP0797 | 62052                | 240         | 1087                | 52                   |
| Gonyaulacales         | <i>Akashiwo</i>         | <i>sanguinea</i>            | CCCM885           | MMETSP0223 | 14                   | -           | -                   | 29                   |
| Gonyaulacales         | <i>Alexandrium</i>      | <i>andersonii</i>           | CCMP2222          | MMETSP1436 | 18455                | 90          | 791                 | -                    |
| Gonyaulacales         | <i>Alexandrium</i>      | <i>catenella</i>            | OF101             | MMETSP0790 | 49675                | 224         | 1451                | 58                   |
| Gonyaulacales         | <i>Alexandrium</i>      | <i>fundyense</i>            | CCMP1719          | MMETSP0196 | 1827                 | -           | -                   | -                    |
| Gonyaulacales         | <i>Alexandrium</i>      | <i>margalefi</i>            | AMGDE01CS-322     | MMETSP0661 | 35485                | 193         | 1218                | 58                   |
| Gonyaulacales         | <i>Alexandrium</i>      | <i>minutum</i>              | CCMP113           | MMETSP0328 | 6720                 | -           | -                   | -                    |
| Gonyaulacales         | <i>Alexandrium</i>      | <i>monilatum</i>            | CCMP3105          | MMETSP0095 | 66660                | 272         | 1468                | 66                   |
| Gonyaulacales         | <i>Alexandrium</i>      | <i>tamarense</i>            | CCMP1771          | MMETSP0378 | 84980                | 290         | 3506                | 129                  |
| Gonyaulacales         | <i>Azadinium</i>        | <i>spinosum</i>             | 3D9               | MMETSP1036 | 52823                | 296         | 1713                | 56                   |
| Gonyaulacales         | <i>Ceratium</i>         | <i>fuscus</i>               | PA161109          | MMETSP1075 | 52729                | 299         | 1426                | 52                   |
| Gonyaulacales         | <i>Cryptocodinium</i>   | <i>cohnii</i>               | Seligo            | MMETSP0323 | 36058                | 305         | 1069                | 47                   |
| Gonyaulacales         | <i>Gambierdiscus</i>    | <i>australes</i>            | CAWD149           | MMETSP0766 | 33677                | 235         | 1259                | 46                   |
| Gonyaulacales         | <i>Gonyaulax</i>        | <i>spinifera</i>            | CCMP409           | MMETSP1439 | 22973                | 181         | 1017                | 44                   |
| Gonyaulacales         | <i>Lingulodinium</i>    | <i>polyedra</i>             | CCMP1738          | MMETSP1032 | 72637                | 252         | 1509                | 58                   |

|                   |                         |                        |                   |            |        |     |      |     |
|-------------------|-------------------------|------------------------|-------------------|------------|--------|-----|------|-----|
| Gonyaulacales     | <i>Protoceratium</i>    | <i>reticulatum</i>     | CCMP1889          | MMETSP0228 | 54591  | 248 | 1342 | 52  |
| Gonyaulacales     | <i>Pyrodinium</i>       | <i>bahamense</i>       | pbaha01           | MMETSP0796 | 75013  | 274 | 1573 | 58  |
| Gymnodiniales     | <i>Amphidinium</i>      | <i>carterae</i>        | CCMP1314          | MMETSP0259 | 31621  | 318 | 1308 | 54  |
| Gymnodiniales     | <i>Amphidinium</i>      | <i>massartii</i>       | CS-259            | MMETSP0689 | 44505  | 313 | 1288 | 57  |
| Gymnodiniales     | <i>Gymnodinium</i>      | <i>catenatum</i>       | GC744             | MMETSP0784 | 67787  | 318 | 1885 | 73  |
| Gymnodiniales     | <i>Gyrodinium</i>       | <i>dominans</i>        | SPMC103           | MMETSP1148 | 21     | -   | -    | -   |
| Gymnodiniales     | <i>Karenia</i>          | <i>brevis</i>          | CCMP2229          | MMETSP0027 | 60640  | 307 | 2102 | 55  |
| Gymnodiniales     | <i>Karenia</i>          | <i>brevis</i>          | SP1               | MMETSP0573 | 68739  | 306 | 2287 | 58  |
| Gymnodiniales     | <i>Karenia</i>          | <i>brevis</i>          | SP3               | MMETSP0527 | 54101  | 287 | 2007 | 51  |
| Gymnodiniales     | <i>Karenia</i>          | <i>brevis</i>          | Wilson            | MMETSP0202 | 72800  | 310 | 2277 | 57  |
| Gymnodiniales     | <i>Karlodinium</i>      | <i>veneficum</i>       | CCMP2283          | MMETSP1015 | 49411  | 305 | 1270 | 49  |
| Gymnodiniales     | <i>Togula</i>           | <i>jolla</i>           | CCCM725           | MMETSP0224 | 35511  | 295 | 1214 | 52  |
| Noctilucales      | <i>Noctiluca</i>        | <i>scintillans</i>     |                   | MMETSP0253 | 31549  | 321 | 1051 | 42  |
| Oxyrrhinales      | <i>Oxyrrhis</i>         | <i>marina</i>          |                   | MMETSP0468 | 29597  | 298 | 1200 | -   |
| Oxyrrhinales      | <i>Oxyrrhis</i>         | <i>marina</i>          | CCMP1788          | MMETSP0044 | 29     | -   | -    | -   |
| Oxyrrhinales      | <i>Oxyrrhis</i>         | <i>marina</i>          | CCMP1795          | MMETSP0451 | 432    | -   | -    | 31  |
| Oxyrrhinales      | <i>Oxyrrhis</i>         | <i>marina</i>          | LB1974            | MMETSP1424 | 24776  | 282 | 745  | 41  |
| Peridiniales      | <i>Brandtodinium</i>    | <i>nutriculum</i>      | RCC3387           | MMETSP1462 | 41272  | 227 | 1233 | 50  |
| Peridiniales      | <i>Durinskia</i>        | <i>baltica</i>         | CSIROCS-38        | MMETSP0117 | 59279  | 327 | 2977 | 271 |
| Peridiniales      | <i>Heterocapsa</i>      | <i>arctica</i>         | CCMP445           | MMETSP1441 | 25337  | 203 | 1038 | 54  |
| Peridiniales      | <i>Heterocapsa</i>      | <i>rotundata</i>       | SCCAPK-483        | MMETSP0503 | 34402  | 229 | 1080 | 49  |
| Peridiniales      | <i>Heterocapsa</i>      | <i>triquetra</i>       | CCMP448           | MMETSP0448 | 31973  | 215 | 1207 | 50  |
| Peridiniales      | <i>Kryptoperidinium</i> | <i>foliaceum</i>       | CCAP1116/3        | MMETSP0118 | 66829  | 227 | 2771 | 296 |
| Peridiniales      | <i>Kryptoperidinium</i> | <i>foliaceum</i>       | CCMP1326          | MMETSP0120 | 104506 | 295 | 3660 | 330 |
| Peridiniales      | <i>Lessardia</i>        | <i>elongata</i>        |                   | MMETSP1147 | 64     | -   | -    | -   |
| Peridiniales      | <i>Peridinium</i>       | <i>aciculiferum</i>    | PAER-2            | MMETSP0370 | 37506  | 283 | 1701 | 65  |
| Peridiniales      | <i>Scrippsiella</i>     | <i>hangoei</i>         | like-SHHI-4       | MMETSP0367 | 54730  | 293 | 1906 | 67  |
| Peridiniales      | <i>Scrippsiella</i>     | <i>hangoei</i>         | SHTV-5            | MMETSP0359 | 56384  | 302 | 1882 | 67  |
| Peridiniales      | <i>Scrippsiella</i>     | <i>trochoidea</i>      | CCMP3099          | MMETSP0270 | 74292  | 295 | 1513 | 54  |
| Peridiniales      | <i>Thoracosphaera</i>   | <i>heimii</i>          | CCMP1069          | MMETSP0225 | 19     | -   | -    | -   |
| Prorocentrales    | <i>Prorocentrum</i>     | <i>lima</i>            | CCMP684           | MMETSP0252 | 142    | -   | -    | -   |
| Prorocentrales    | <i>Prorocentrum</i>     | <i>micans</i>          | CCCM845           | MMETSP0251 | 201    | -   | -    | -   |
| Prorocentrales    | <i>Prorocentrum</i>     | <i>minimum</i>         | CCMP1329          | MMETSP0053 | 54018  | 252 | 1587 | 60  |
| Prorocentrales    | <i>Prorocentrum</i>     | <i>minimum</i>         | CCMP2233          | MMETSP0267 | 53968  | 252 | 1567 | 56  |
| Pyrocystales      | <i>Pyrocystis</i>       | <i>lumula</i>          | CCCM517           | MMETSP0229 | 31     | -   | -    | -   |
| Suessiales        | <i>Pelagodinium</i>     | <i>beii</i>            | RCC1491           | MMETSP1338 | 36311  | 267 | 1423 | 46  |
| Suessiales        | <i>Polarella</i>        | <i>glacialis</i>       | CCMP1383          | MMETSP0227 | 42550  | 281 | 1498 | 46  |
| Suessiales        | <i>Polarella</i>        | <i>glacialis</i>       | CCMP2088          | MMETSP1440 | 18595  | 246 | 1290 | 52  |
| Suessiales        | <i>Symbiodinium</i>     | <i>kawagutii</i>       | CCMP2468          | MMETSP0132 | 1393   | -   | -    | -   |
| Suessiales        | <i>Symbiodinium</i>     | <i>sp.</i>             | C1                | MMETSP1367 | 36921  | 310 | 1993 | 49  |
| Suessiales        | <i>Symbiodinium</i>     | <i>sp.</i>             | C15               | MMETSP1370 | 29068  | 249 | 1549 | 72  |
| Suessiales        | <i>Symbiodinium</i>     | <i>sp.</i>             | CCMP2430          | MMETSP1115 | 35890  | 270 | 1489 | 56  |
| Suessiales        | <i>Symbiodinium</i>     | <i>sp.</i>             | CCMP421           | MMETSP1110 | 60730  | 254 | 2288 | 41  |
| Suessiales        | <i>Symbiodinium</i>     | <i>sp.</i>             | cladeA            | MMETSP1374 | 29653  | 227 | 1347 | 48  |
| Suessiales        | <i>Symbiodinium</i>     | <i>sp.</i>             | D1a               | MMETSP1377 | 29487  | 149 | 2452 | 48  |
| Suessiales        | <i>Symbiodinium</i>     | <i>sp.</i>             | Mp                | MMETSP1122 | 35661  | 296 | 1405 | 52  |
| Syndiniales       | <i>Amoebophrya</i>      | <i>sp.</i>             | Ameob2            | MMETSP0795 | 12427  | 63  | 99   | 1   |
| <b>Perkinsea</b>  |                         |                        |                   |            |        |     |      |     |
| Perkinsidae       | <i>Perkinsus</i>        | <i>chESApeaki</i>      | ATCCPRA-65        | MMETSP0924 | 661    | -   | -    | -   |
| Perkinsidae       | <i>Perkinsus</i>        | <i>marinus</i>         | ATCC50439         | MMETSP0922 | 1042   | -   | -    | -   |
| <b>Amoebozoa</b>  |                         |                        |                   |            |        |     |      |     |
| <b>Cutosea</b>    |                         |                        |                   |            |        |     |      |     |
| Pessonella        | <i>Pessonella</i>       | <i>sp.</i>             | PRA-29            | MMETSP0420 | 13793  | 234 | 357  | 6   |
| Unclassified      | <i>Sapocribrum</i>      | <i>chincoteaguense</i> | ATCC50979         | MMETSP0437 | 13091  | 222 | 153  | 1   |
| <b>Discosea</b>   |                         |                        |                   |            |        |     |      |     |
| Flabellinia       | <i>Paramoeba</i>        | <i>aestuarina</i>      | SoJaBioB1-5-56-2  | MMETSP0161 | 16687  | 278 | 893  | 26  |
| Flabellinia       | <i>Paramoeba</i>        | <i>atlantica</i>       | 621-1-CCAP1560/9  | MMETSP0151 | 13934  | 261 | 885  | 25  |
| Flabellinia       | <i>Vannella</i>         | <i>robusta</i>         | DIVA3518-3-11-1-6 | MMETSP0166 | 14750  | 306 | 802  | 22  |
| Flabellinia       | <i>Vannella</i>         | <i>sp.</i>             | DIVA3517-6-12     | MMETSP0168 | 15677  | 276 | 555  | 20  |
| Flabellinia       | <i>Vexillifera</i>      | <i>sp.</i>             | DIVA3564-2        | MMETSP0173 | 8747   | 289 | 313  | 7   |
| Longamoebida      | <i>Cunea</i>            | <i>sp.</i>             | BSH-02190019      | MMETSP0417 | 8755   | 304 | 341  | 12  |
| Stygamoebida      | <i>Stygamoeba</i>       | <i>regulata</i>        | BSH-02190019      | MMETSP0447 | 26744  | 315 | 694  | 15  |
| <b>Stereomyxa</b> |                         |                        |                   |            |        |     |      |     |
| Stereomyxa ramosa | <i>Dracoamoeba</i>      | <i>jomungandri</i>     | Chine5            | MMETSP0439 | 14588  | 325 | 447  | 10  |
| <b>Tubulinea</b>  |                         |                        |                   |            |        |     |      |     |
| Euamoebida        | <i>Filamoeba</i>        | <i>nolandi</i>         | NC-AS-23-1        | MMETSP0413 | 18369  | 305 | 527  | 15  |
| Trichosphaerium   | <i>Trichosphaerium</i>  | <i>sp.</i>             | Am-1-7wt          | MMETSP0405 | 16915  | 213 | 264  | 4   |

|                          |                         |                                |                 |            |       |     |      |    |
|--------------------------|-------------------------|--------------------------------|-----------------|------------|-------|-----|------|----|
| <b>Cryptophyta</b>       |                         |                                |                 |            |       |     |      |    |
| <b>Cryptophyceae</b>     |                         |                                |                 |            |       |     |      |    |
| Cryptomonadaceae         | <i>Cryptomonas</i>      | <i>curvata</i>                 | CCAP979/52      | MMETSP1050 | 24882 | 172 | 1029 | 37 |
| Cryptomonadaceae         | <i>Cryptomonas</i>      | <i>paramecium</i>              | CCAP977/2a      | MMETSP0038 | 26159 | 206 | 1082 | 40 |
| Cryptomonadaceae         | <i>Goniomonas</i>       | <i>pacifica</i>                | CCMP1869        | MMETSP0107 | 34181 | 252 | 820  | 30 |
| Cryptomonadaceae         | <i>Goniomonas</i>       | <i>sp.</i>                     | m               | MMETSP0114 | 15097 | 191 | 686  | 26 |
| Hemiselmidaceae          | <i>Hemiselmis</i>       | <i>andersenii</i>              | CCMP1180        | MMETSP1042 | 23365 | 207 | 1336 | 52 |
| Hemiselmidaceae          | <i>Hemiselmis</i>       | <i>andersenii</i>              | CCMP439         | MMETSP1041 | 18983 | 191 | 1323 | 54 |
| Hemiselmidaceae          | <i>Hemiselmis</i>       | <i>andersenii</i>              | CCMP441         | MMETSP1043 | 18511 | 175 | 1187 | 41 |
| Hemiselmidaceae          | <i>Hemiselmis</i>       | <i>andersenii</i>              | CCMP644         | MMETSP0043 | 25992 | 220 | 1412 | 54 |
| Hemiselmidaceae          | <i>Hemiselmis</i>       | <i>rufescens</i>               | PCC563          | MMETSP1357 | 19350 | 208 | 1355 | 56 |
| Hemiselmidaceae          | <i>Hemiselmis</i>       | <i>tepida</i>                  | CCMP443         | MMETSP1355 | 12808 | 143 | 934  | 31 |
| Hemiselmidaceae          | <i>Hemiselmis</i>       | <i>virescens</i>               | PCC157          | MMETSP1356 | 14074 | 210 | 1353 | 49 |
| unclassified             | <i>Palpitomonas</i>     | <i>bilix</i>                   | NIES-2562       | MMETSP0780 | 21710 | 254 | 245  | 2  |
| unclassified Cryptophyta | <i>Cryptophyceae</i>    | <i>sp.</i>                     | CCMP2293        | MMETSP0986 | 27794 | 306 | 1244 | 56 |
| <b>Pyrenomonadales</b>   |                         |                                |                 |            |       |     |      |    |
| Chroomonadaceae          | <i>Chroomonas</i>       | <i>c.f. mesostigmatica</i>     | CCMP1168        | MMETSP0047 | 27140 | 263 | 1320 | 58 |
| Geminigeraceae           | <i>Geminigera</i>       | <i>cryphila</i>                | CCMP2564        | MMETSP0799 | 39369 | 292 | 1429 | 58 |
| Geminigeraceae           | <i>Geminigera</i>       | <i>sp.</i>                     | CaronLabIsolate | MMETSP1102 | 33173 | 273 | 1385 | 53 |
| Geminigeraceae           | <i>Guillardia</i>       | <i>theta</i>                   | CCMP2712        | MMETSP0046 | 18869 | 273 | 1199 | 58 |
| Geminigeraceae           | <i>Hanusia</i>          | <i>phi</i>                     | CCMP325         | MMETSP1048 | 18729 | 196 | 963  | 41 |
| Geminigeraceae           | <i>Proteomonas</i>      | <i>sulcata</i>                 | CCMP704         | MMETSP1049 | 16700 | 125 | 951  | 42 |
| Pyrenomonadaceae         | <i>Rhodomonas</i>       | <i>abbreviata</i>              | CaronLabIsolate | MMETSP1101 | 24640 | 221 | 1511 | 58 |
| Pyrenomonadaceae         | <i>Rhodomonas</i>       | <i>lens</i>                    | RHODO           | MMETSP0484 | 24414 | 280 | 1393 | 60 |
| Pyrenomonadaceae         | <i>Rhodomonas</i>       | <i>salina</i>                  | CCMP1319        | MMETSP1047 | 23428 | 258 | 1269 | 52 |
| Pyrenomonadaceae         | <i>Rhodomonas</i>       | <i>sp.</i>                     | CCMP768         | MMETSP1091 | 15116 | 179 | 1165 | 52 |
| <b>Euglenozoa</b>        |                         |                                |                 |            |       |     |      |    |
| <b>Euglenophyceae</b>    |                         |                                |                 |            |       |     |      |    |
| Eutreptiales             | <i>Eutreptiella</i>     | <i>gymnastica-like</i>         | CCMP1594        | MMETSP0809 | 15603 | 215 | 547  | 17 |
| Eutreptiales             | <i>Eutreptiella</i>     | <i>gymnastica</i>              | NIES-381        | MMETSP0039 | 14022 | 220 | 582  | 17 |
| <b>Kinetoplastida</b>    |                         |                                |                 |            |       |     |      |    |
| Bodoniidae               | <i>Neobodo</i>          | <i>designis</i>                | CCAP1951/1      | MMETSP1114 | 18500 | 197 | 282  | 12 |
| <b>Glaucophyta</b>       |                         |                                |                 |            |       |     |      |    |
| <b>Glaucophyceae</b>     |                         |                                |                 |            |       |     |      |    |
| Gloeochaetales           | <i>Cyanoptyche</i>      | <i>gloeocystis</i>             | SAG4-97         | MMETSP1086 | 6276  | 180 | 590  | 19 |
| Gloeochaetales           | <i>Gloeochaete</i>      | <i>wittrockiana</i>            | SAG46-84        | MMETSP1089 | 17995 | 358 | 1110 | 36 |
| <b>Haptophyta</b>        |                         |                                |                 |            |       |     |      |    |
| <b>Coccolithales</b>     |                         |                                |                 |            |       |     |      |    |
| Calcidiscaceae           | <i>Calcidiscus</i>      | <i>leptopus</i>                | RCC1130         | MMETSP1334 | 16430 | 262 | 1299 | 62 |
| Coccolithaceae           | <i>Coccolithus</i>      | <i>pelagicus ssp. braarudi</i> | PLY182g         | MMETSP0164 | 15883 | 259 | 1536 | 71 |
| Pleurochrysidaceae       | <i>Pleurochrysis</i>    | <i>carterae</i>                | CCMP645         | MMETSP1136 | 20441 | 272 | 1315 | 64 |
| Pontosphaeraceae         | <i>Scyphosphaera</i>    | <i>apsteinii</i>               | RCC1455         | MMETSP1333 | 21173 | 285 | 1384 | 56 |
| <b>Isochrysidales</b>    |                         |                                |                 |            |       |     |      |    |
| Isochrysidaceae          | <i>Isochrysis</i>       | <i>galbana</i>                 | CCMP1323        | MMETSP0944 | 30689 | 336 | 1875 | 75 |
| Isochrysidaceae          | <i>Isochrysis</i>       | <i>sp.</i>                     | CCMP1244        | MMETSP1090 | 24844 | 219 | 1262 | 76 |
| Isochrysidaceae          | <i>Isochrysis</i>       | <i>sp.</i>                     | CCMP1324        | MMETSP1129 | 14319 | 251 | 1207 | 53 |
| Noelaerhabdaceae         | <i>Emiliana</i>         | <i>huxleyi</i>                 | 374             | MMETSP1006 | 12616 | 174 | 1030 | 50 |
| Noelaerhabdaceae         | <i>Emiliana</i>         | <i>huxleyi</i>                 | 379             | MMETSP0994 | 16566 | 178 | 1123 | 61 |
| Noelaerhabdaceae         | <i>Emiliana</i>         | <i>huxleyi</i>                 | CCMP370         | MMETSP1154 | 27531 | 224 | 1289 | 75 |
| Noelaerhabdaceae         | <i>Emiliana</i>         | <i>huxleyi</i>                 | PLYM219         | MMETSP1150 | 26155 | 245 | 1330 | 76 |
| Noelaerhabdaceae         | <i>Gephyrocapsa</i>     | <i>oceanica</i>                | RCC1303         | MMETSP1363 | 24232 | 247 | 1311 | 72 |
| <b>Pavloales</b>         |                         |                                |                 |            |       |     |      |    |
| Pavlovaceae              | <i>Exanthemachrysis</i> | <i>gayraliae</i>               | RCC1523         | MMETSP1464 | 11466 | 188 | 900  | 49 |
| Pavlovaceae              | <i>Pavlova</i>          | <i>gyrans</i>                  | CCMP608         | MMETSP1466 | 10608 | 150 | 874  | 53 |
| Pavlovaceae              | <i>Pavlova</i>          | <i>lutheri</i>                 | RCC1537         | MMETSP1463 | 9146  | 148 | 967  | 60 |
| Pavlovaceae              | <i>Pavlova</i>          | <i>sp.</i>                     | CCMP459         | MMETSP1139 | 12159 | 204 | 1089 | 57 |
| Pavlovaceae              | <i>Pavlova</i>          | <i>sp.</i>                     | CCMP2436        | MMETSP0982 | 24725 | 310 | 1201 | 66 |
| <b>Phaeocystales</b>     |                         |                                |                 |            |       |     |      |    |
| Phaeocystaceae           | <i>Phaeocystis</i>      | <i>antarctica</i>              | CaronLabIsolate | MMETSP1100 | 37178 | 250 | 1235 | 54 |
| Phaeocystaceae           | <i>Phaeocystis</i>      | <i>antarctica</i>              | CCMP1374        | MMETSP1444 | 12611 | 214 | 1158 | 60 |
| Phaeocystaceae           | <i>Phaeocystis</i>      | <i>cordata</i>                 | RCC1383         | MMETSP1465 | 6418  | 125 | 932  | 26 |
| Phaeocystaceae           | <i>Phaeocystis</i>      | <i>rex</i>                     | CCMP2000        | MMETSP1178 | 9825  | 172 | 983  | 41 |
| Phaeocystaceae           | <i>Phaeocystis</i>      | <i>sp.</i>                     | CCMP2710        | MMETSP1162 | 12392 | 166 | 880  | 45 |
| <b>Prymnesiales</b>      |                         |                                |                 |            |       |     |      |    |
| Chrysochromulinaceae     | <i>Chrysochromulina</i> | <i>rotalis</i>                 | UIO044          | MMETSP0287 | 26424 | 238 | 1310 | 65 |
| Chrysoculter             | <i>Chrysoculter</i>     | <i>rhomboideus</i>             | RCC1486         | MMETSP1335 | 9729  | 164 | 919  | 48 |
| Prymnesiaceae            | <i>Haptolina</i>        | <i>brevifila</i>               | UTEXLB985       | MMETSP1094 | 23713 | 209 | 1260 | 54 |
| Prymnesiaceae            | <i>Haptolina</i>        | <i>ericina</i>                 | CCMP281         | MMETSP1096 | 33566 | 213 | 1375 | 59 |
| Prymnesiaceae            | <i>Imantonia</i>        | <i>sp.</i>                     | RCC918          | MMETSP1474 | 15854 | 308 | 1572 | 63 |
| Prymnesiaceae            | <i>Prymnesium</i>       | <i>parvum</i>                  | Texoma1         | MMETSP0006 | 29250 | 297 | 1391 | 68 |
| Prymnesiaceae            | <i>Prymnesium</i>       | <i>polylepis</i>               | CCMP1757        | MMETSP0143 | 33497 | 220 | 1280 | 41 |
| Prymnesiaceae            | <i>Prymnesium</i>       | <i>polylepis</i>               | UIO037          | MMETSP0286 | 16372 | 90  | 777  | 64 |
| <b>Heterolobosea</b>     |                         |                                |                 |            |       |     |      |    |
| Percolomonas             | <i>Percolomonas</i>     | <i>cosmopolitus</i>            | AE-1-ATCC50343  | MMETSP0758 | 11342 | 276 | 189  | 4  |
| Percolomonas             | <i>Percolomonas</i>     | <i>cosmopolitus</i>            | WS              | MMETSP0759 | 10799 | 266 | 173  | 4  |

|                             |                         |                               |               |            |       |     |      |     |
|-----------------------------|-------------------------|-------------------------------|---------------|------------|-------|-----|------|-----|
| <b>Opisthokonta</b>         |                         |                               |               |            |       |     |      |     |
| <b>Choanoflagellida</b>     |                         |                               |               |            |       |     |      |     |
| Acanthoecida                | <i>Acanthoeca</i>       | <i>like-sp. 10tr</i>          |               | MMETSP0105 | 15634 | 275 | 403  | 12  |
| <b>Fungi</b>                |                         |                               |               |            |       |     |      |     |
| Ascomycetes                 | <i>Debaryomyces</i>     | <i>hansentii</i>              | J26           | MMETSP0232 | 2478  | 166 | 59   | -   |
| <b>Rh</b>                   |                         |                               |               |            |       |     |      |     |
| <b>Cercozoa</b>             |                         |                               |               |            |       |     |      |     |
| Chlorarachniophyceae        | <i>Amorhoclora</i>      | <i>amoebiformis</i>           | CCMP2058      | MMETSP0042 | 23016 | 351 | 1101 | 60  |
| Chlorarachniophyceae        | <i>Bigelowiella</i>     | <i>longifila</i>              | CCMP242       | MMETSP1359 | 14866 | 194 | 1123 | 56  |
| Chlorarachniophyceae        | <i>Bigelowiella</i>     | <i>natans</i>                 | CCMP1242      | MMETSP1358 | 16168 | 311 | 1128 | 57  |
| Chlorarachniophyceae        | <i>Bigelowiella</i>     | <i>natans</i>                 | CCMP1258-1    | MMETSP1055 | 24797 | 269 | 1241 | 53  |
| Chlorarachniophyceae        | <i>Bigelowiella</i>     | <i>natans</i>                 | CCMP1259      | MMETSP1054 | 24331 | 309 | 1258 | 56  |
| Chlorarachniophyceae        | <i>Bigelowiella</i>     | <i>natans</i>                 | CCMP2755      | MMETSP0045 | 21263 | 271 | 1166 | 53  |
| Chlorarachniophyceae        | <i>Bigelowiella</i>     | <i>natans</i>                 | CCMP623       | MMETSP1052 | 22991 | 307 | 1204 | 52  |
| Chlorarachniophyceae        | <i>Chlorarachnion</i>   | <i>reptans</i>                | CCCM449       | MMETSP0109 | 24858 | 362 | 1138 | 55  |
| Chlorarachniophyceae        | <i>Gymnochloa</i>       | <i>sp.</i>                    | CCMP2014      | MMETSP0110 | 15823 | 338 | 981  | 51  |
| Chlorarachniophyceae        | <i>Lotharella</i>       | <i>globosa</i>                | CCCM811       | MMETSP0111 | 17478 | 286 | 1078 | 54  |
| Chlorarachniophyceae        | <i>Lotharella</i>       | <i>globosa</i>                | LEX01         | MMETSP0041 | 22972 | 322 | 1120 | 57  |
| Chlorarachniophyceae        | <i>Lotharella</i>       | <i>oceanica</i>               | CCMP622       | MMETSP0040 | 14994 | 213 | 1048 | 50  |
| Chlorarachniophyceae        | <i>Norrisiella</i>      | <i>sphaerica</i>              | BC52          | MMETSP0113 | 13468 | 274 | 1019 | 52  |
| Chlorarachniophyceae        | <i>Partenskyella</i>    | <i>glossopodia</i>            | RCC365        | MMETSP1318 | 13371 | 270 | 1049 | 51  |
| Unclassified Cercozoa       | <i>Mataza</i>           | <i>sp.</i>                    | SIOpierMataz1 | MMETSP0086 | 20849 | 334 | 588  | 21  |
| <b>Foraminifera</b>         |                         |                               |               |            |       |     |      |     |
| Miliolida                   | <i>Sorites</i>          | <i>sp.</i>                    |               | MMETSP0191 | 46570 | 239 | 1398 | 26  |
| Rotalida                    | <i>Ammonia</i>          | <i>sp.</i>                    |               | MMETSP1384 | 20082 | 289 | 722  | 20  |
| Rotalida                    | <i>Elphidium</i>        | <i>margaritaceum</i>          |               | MMETSP1385 | 16959 | 276 | 623  | 19  |
| Rotalida                    | <i>Rosalina</i>         | <i>sp.</i>                    |               | MMETSP0190 | 17622 | 163 | 584  | 15  |
| <b>Haplosporidia</b>        |                         |                               |               |            |       |     |      |     |
| Haplosporidiidae            | <i>Minchinia</i>        | <i>chitonis</i>               |               | MMETSP0186 | 205   | -   | -    | -   |
| <b>Rhodophyta</b>           |                         |                               |               |            |       |     |      |     |
| <b>Bangiophyceae</b>        |                         |                               |               |            |       |     |      |     |
| Porphyridiales              | <i>Erythrolobus</i>     | <i>australicus</i>            | CCMP3124      | MMETSP1353 | 5863  | 154 | 656  | 31  |
| Porphyridiales              | <i>Erythrolobus</i>     | <i>madagascarenis</i>         | CCMP3276      | MMETSP1354 | 5563  | 111 | 614  | 32  |
| Porphyridiales              | <i>Porphyridium</i>     | <i>aeruginenum</i>            | SAG1380-2     | MMETSP0313 | 6457  | 188 | 611  | 23  |
| Porphyridiales              | <i>Timsipurckia</i>     | <i>oligopyrenoides</i>        | CCMP3278      | MMETSP1172 | 5422  | 200 | 599  | 32  |
| Unclassified                | <i>Bangiophyceae</i>    | <i>sp.</i>                    | CCMP1999      | MMETSP1475 | 7066  | 273 | 800  | 50  |
| <b>Composopogonophyceae</b> |                         |                               |               |            |       |     |      |     |
| Composopogonales            | <i>Composopogon</i>     | <i>coeruleus</i>              | SAG36-94      | MMETSP0312 | 5607  | 228 | 441  | 19  |
| Erythroplidales             | <i>Madagascaria</i>     | <i>erythrocladioides</i>      | CCMP3234      | MMETSP1450 | 20046 | 213 | 768  | 26  |
| <b>Rhodellophyceae</b>      |                         |                               |               |            |       |     |      |     |
| Rhodellales                 | <i>Rhodella</i>         | <i>maculata</i>               | CCMP736       | MMETSP0167 | 10107 | 169 | 358  | 20  |
| <b>Stylonematophyceae</b>   |                         |                               |               |            |       |     |      |     |
| Stylonematales              | <i>Rhodosorus</i>       | <i>marinus</i>                | CCMP769       | MMETSP0011 | 9098  | 295 | 812  | 55  |
| Stylonematales              | <i>Rhodosorus</i>       | <i>marinus</i>                | UTEXLB2760    | MMETSP0315 | 6415  | 226 | 813  | 49  |
| <b>Stramenopiles</b>        |                         |                               |               |            |       |     |      |     |
| <b>Bacillariophyta</b>      |                         |                               |               |            |       |     |      |     |
| Bacillariophyceae           | <i>Amphiprora</i>       | <i>paludosa</i>               | CCMP125       | MMETSP1065 | 13911 | 240 | 1119 | 302 |
| Bacillariophyceae           | <i>Amphiprora</i>       | <i>sp.</i>                    | CCMP467       | MMETSP0725 | 16080 | 286 | 1148 | 294 |
| Bacillariophyceae           | <i>Amphora</i>          | <i>coffeaeformis</i>          | CCMP127       | MMETSP0316 | 11868 | 164 | 937  | 200 |
| Bacillariophyceae           | <i>Craspedostauros</i>  | <i>australis</i>              | CCMP3328      | MMETSP1442 | 10308 | 105 | 906  | 257 |
| Bacillariophyceae           | <i>Cylindrotheca</i>    | <i>closterium</i>             | KMMCC-B-181   | MMETSP0017 | 26203 | 305 | 1247 | 308 |
| Bacillariophyceae           | <i>Entomoneis</i>       | <i>sp.</i>                    | CCMP2396      | MMETSP1443 | 8057  | 179 | 1050 | 234 |
| Bacillariophyceae           | <i>Fragilariopsis</i>   | <i>kerquelensis</i>           | L2-C3         | MMETSP0906 | 36760 | 324 | 1292 | 297 |
| Bacillariophyceae           | <i>Fragilariopsis</i>   | <i>kerquelensis</i>           | L26-C5        | MMETSP0733 | 23278 | 304 | 1236 | 299 |
| Bacillariophyceae           | <i>Nitzschia</i>        | <i>sp.</i>                    | RCC80         | MMETSP0014 | 39601 | 317 | 1265 | 282 |
| Bacillariophyceae           | <i>Pseudo-nitzschia</i> | <i>arenysensis</i>            | B593          | MMETSP0329 | 20142 | 282 | 1225 | 299 |
| Bacillariophyceae           | <i>Pseudo-nitzschia</i> | <i>australis</i>              | 1024910AB     | MMETSP0139 | 13298 | 297 | 1168 | 312 |
| Bacillariophyceae           | <i>Pseudo-nitzschia</i> | <i>delicatissima</i>          | B596          | MMETSP0327 | 17743 | 302 | 1180 | 306 |
| Bacillariophyceae           | <i>Pseudo-nitzschia</i> | <i>delicatissima</i>          | UNC1205       | MMETSP1432 | 10314 | 205 | 1208 | 298 |
| Bacillariophyceae           | <i>Pseudo-nitzschia</i> | <i>fraudulenta</i>            | WWA7          | MMETSP0850 | 29904 | 254 | 2058 | 330 |
| Bacillariophyceae           | <i>Pseudo-nitzschia</i> | <i>heimii</i>                 | UNC1101       | MMETSP1423 | 11460 | 292 | 1156 | 297 |
| Bacillariophyceae           | <i>Pseudo-nitzschia</i> | <i>pungens c.f. cingulata</i> |               | MMETSP1060 | 10826 | 260 | 1038 | 302 |
| Bacillariophyceae           | <i>Pseudo-nitzschia</i> | <i>pungens c.f. pungens</i>   |               | MMETSP1061 | 11909 | 272 | 1114 | 307 |
| Bacillariophyceae           | <i>Stauroneis</i>       | <i>constricta</i>             | CCMP1120      | MMETSP1352 | 9437  | 168 | 766  | 200 |
| Bacillariophyceae           | <i>Tryblionella</i>     | <i>compressa</i>              | CCMP561       | MMETSP0744 | 14189 | 216 | 1033 | 280 |
| Coscinodiscophyceae         | <i>Corethron</i>        | <i>hystrix</i>                | 308           | MMETSP0010 | 16067 | 238 | 1119 | 192 |
| Coscinodiscophyceae         | <i>Corethron</i>        | <i>pennatum</i>               | L29A3         | MMETSP0169 | 30703 | 257 | 1144 | 200 |
| Coscinodiscophyceae         | <i>Coscinodiscus</i>    | <i>walesii</i>                | CCMP2513      | MMETSP1066 | 17343 | 252 | 1274 | 249 |
| Coscinodiscophyceae         | <i>Leptocylindrus</i>   | <i>aporus</i>                 | B651          | MMETSP0322 | 13853 | 277 | 1141 | 255 |
| Coscinodiscophyceae         | <i>Leptocylindrus</i>   | <i>danicus</i>                | B650          | MMETSP0321 | 20125 | 300 | 1293 | 256 |
| Coscinodiscophyceae         | <i>Leptocylindrus</i>   | <i>danicus</i>                | CCMP1856      | MMETSP1362 | 11668 | 289 | 1281 | 218 |
| Coscinodiscophyceae         | <i>Proboscia</i>        | <i>alata</i>                  | PI-D3         | MMETSP0174 | 20007 | 319 | 1314 | 244 |
| Coscinodiscophyceae         | <i>Proboscia</i>        | <i>inermis</i>                | CCAP1064/1    | MMETSP0816 | 13867 | 169 | 1256 | 234 |
| Coscinodiscophyceae         | <i>Rhizosolenia</i>     | <i>setigera</i>               | CCMP1694      | MMETSP0789 | 20486 | 300 | 1111 | 247 |



|                    |                         |                        |                |            |       |     |      |     |
|--------------------|-------------------------|------------------------|----------------|------------|-------|-----|------|-----|
| Fragilariophyceae  | <i>Asterionellopsis</i> | <i>glacialis</i>       |                | MMETSP0713 | 11279 | 229 | 1226 | 330 |
| Fragilariophyceae  | <i>Asterionellopsis</i> | <i>glacialis</i>       | CCMP134        | MMETSP0705 | 15719 | 276 | 1373 | 314 |
| Fragilariophyceae  | <i>Asterionellopsis</i> | <i>glacialis</i>       | CCMP1581       | MMETSP1394 | 11390 | 277 | 1261 | 311 |
| Fragilariophyceae  | <i>Astrosyne</i>        | <i>radiata</i>         | 13vi08-1A      | MMETSP0418 | 20720 | 156 | 1177 | 260 |
| Fragilariophyceae  | <i>Cyclophora</i>       | <i>tenuis</i>          | ECT3854        | MMETSP0397 | 17164 | 141 | 1106 | 289 |
| Fragilariophyceae  | <i>Grammatophora</i>    | <i>oceanica</i>        | CCMP410        | MMETSP0009 | 15798 | 175 | 1147 | 244 |
| Fragilariophyceae  | <i>Licmophora</i>       | <i>paradoxa</i>        | CCMP2313       | MMETSP1360 | 14141 | 201 | 1155 | 274 |
| Fragilariophyceae  | <i>Staurosira</i>       | <i>complexa</i>        | CCMP2646       | MMETSP1361 | 12983 | 287 | 1201 | 302 |
| Fragilariophyceae  | <i>Striatella</i>       | <i>unipunctata</i>     | CCMP2910       | MMETSP0800 | 22437 | 196 | 1043 | 219 |
| Fragilariophyceae  | <i>Synedropsis</i>      | <i>c.f. recta</i>      | CCMP1620       | MMETSP1176 | 12149 | 206 | 1175 | 280 |
| Fragilariophyceae  | <i>Thalassionema</i>    | <i>frauenfeldii</i>    | CCMP1798       | MMETSP0786 | 21892 | 309 | 1241 | 319 |
| Fragilariophyceae  | <i>Thalassionema</i>    | <i>nitzschioides</i>   |                | MMETSP0693 | 3099  | -   | -    | -   |
| Fragilariophyceae  | <i>Thalassionema</i>    | <i>nitzschioides</i>   | L26-B          | MMETSP0156 | 15232 | 316 | 1253 | 324 |
| Fragilariophyceae  | <i>Thalassiothrix</i>   | <i>antarctica</i>      | L6-D1          | MMETSP0152 | 16431 | 327 | 1326 | 300 |
| Mediophyceae       | <i>Attheya</i>          | <i>septentrionalis</i> | CCMP2084       | MMETSP1449 | 11038 | 225 | 1140 | 274 |
| Mediophyceae       | <i>Aulacoseira</i>      | <i>subarctica</i>      | CCAP1002/5     | MMETSP1064 | 11436 | 267 | 1007 | 191 |
| Mediophyceae       | <i>Chaetoceros</i>      | <i>affinis</i>         | CCMP159        | MMETSP0088 | 11843 | 295 | 1176 | 316 |
| Mediophyceae       | <i>Chaetoceros</i>      | <i>brevis</i>          | CCMP164        | MMETSP1435 | 2744  | -   | -    | -   |
| Mediophyceae       | <i>Chaetoceros</i>      | <i>c.f. neogracile</i> | RCC1993        | MMETSP1336 | 11429 | 266 | 1278 | 311 |
| Mediophyceae       | <i>Chaetoceros</i>      | <i>curvisetus</i>      |                | MMETSP0716 | 13997 | 197 | 1330 | 320 |
| Mediophyceae       | <i>Chaetoceros</i>      | <i>debilis</i>         | MM31A-1        | MMETSP0149 | 13053 | 351 | 1314 | 316 |
| Mediophyceae       | <i>Chaetoceros</i>      | <i>dichaeta</i>        | CCMP1751       | MMETSP1447 | 12614 | 238 | 1277 | 280 |
| Mediophyceae       | <i>Chaetoceros</i>      | <i>neogracile</i>      | CCMP1317       | MMETSP0751 | 18134 | 328 | 1285 | 326 |
| Mediophyceae       | <i>Chaetoceros</i>      | <i>sp.</i>             | GSL56          | MMETSP0200 | 14561 | 301 | 1109 | 313 |
| Mediophyceae       | <i>Chaetoceros</i>      | <i>sp.</i>             | UNC1202        | MMETSP1429 | 9597  | 150 | 1251 | 298 |
| Mediophyceae       | <i>Cyclotella</i>       | <i>meneghiniana</i>    | CCMP338        | MMETSP1057 | 12776 | 305 | 1115 | 277 |
| Mediophyceae       | <i>Dactyliosolen</i>    | <i>fragilissimus</i>   |                | MMETSP0580 | 9661  | 331 | 1211 | 275 |
| Mediophyceae       | <i>Ditylum</i>          | <i>brightwellii</i>    | GSO103         | MMETSP1002 | 13382 | 169 | 1385 | 285 |
| Mediophyceae       | <i>Ditylum</i>          | <i>brightwellii</i>    | GSO104         | MMETSP1010 | 18312 | 302 | 1540 | 313 |
| Mediophyceae       | <i>Ditylum</i>          | <i>brightwellii</i>    | GSO105         | MMETSP0998 | 16013 | 270 | 1454 | 319 |
| Mediophyceae       | <i>Ditylum</i>          | <i>brightwellii</i>    | Pop1-SS4       | MMETSP1062 | 25958 | 332 | 1615 | 321 |
| Mediophyceae       | <i>Ditylum</i>          | <i>brightwellii</i>    | Pop2-SS10      | MMETSP1063 | 25500 | 342 | 1607 | 324 |
| Mediophyceae       | <i>Eucampia</i>         | <i>antarctica</i>      | CCMP1452       | MMETSP1437 | 8726  | 239 | 1151 | 258 |
| Mediophyceae       | <i>Extubocellulus</i>   | <i>spinifer</i>        | CCMP396        | MMETSP0696 | 33809 | 275 | 1149 | 337 |
| Mediophyceae       | <i>Helicotheca</i>      | <i>tamensis</i>        | CCMP826        | MMETSP1171 | 8123  | 138 | 1005 | 226 |
| Mediophyceae       | <i>Minutocellus</i>     | <i>polymorphus</i>     | CCMP3303       | MMETSP1434 | 7158  | 59  | 698  | 175 |
| Mediophyceae       | <i>Minutocellus</i>     | <i>polymorphus</i>     | NH13           | MMETSP1070 | 25350 | 279 | 1118 | 338 |
| Mediophyceae       | <i>Minutocellus</i>     | <i>polymorphus</i>     | RCC2270        | MMETSP1322 | 10968 | 190 | 1079 | 283 |
| Mediophyceae       | <i>Odontella</i>        | <i>aurita</i>          | isolate1302-5  | MMETSP0015 | 30813 | 286 | 1419 | 323 |
| Mediophyceae       | <i>Odontella</i>        | <i>sinensis</i>        | Grunow1884     | MMETSP0160 | 14045 | 198 | 1288 | 301 |
| Mediophyceae       | <i>Skeletonema</i>      | <i>costatum</i>        | 1716           | MMETSP0013 | 29077 | 351 | 1569 | 287 |
| Mediophyceae       | <i>Skeletonema</i>      | <i>dohrnii</i>         | SkelB          | MMETSP0562 | 15481 | 280 | 1190 | 292 |
| Mediophyceae       | <i>Skeletonema</i>      | <i>grethae</i>         | CCMP1804       | MMETSP0578 | 10204 | 199 | 1006 | 224 |
| Mediophyceae       | <i>Skeletonema</i>      | <i>japonicum</i>       | CCMP2506       | MMETSP0593 | 11387 | 251 | 1128 | 277 |
| Mediophyceae       | <i>Skeletonema</i>      | <i>marinoi</i>         | FE60           | MMETSP1040 | 15348 | 174 | 1118 | 227 |
| Mediophyceae       | <i>Skeletonema</i>      | <i>marinoi</i>         | FE7            | MMETSP1039 | 17170 | 201 | 1198 | 242 |
| Mediophyceae       | <i>Skeletonema</i>      | <i>marinoi</i>         | skelA          | MMETSP0918 | 12426 | 224 | 1233 | 279 |
| Mediophyceae       | <i>Skeletonema</i>      | <i>marinoi</i>         | SM1012Den-03   | MMETSP0320 | 25127 | 267 | 1276 | 286 |
| Mediophyceae       | <i>Skeletonema</i>      | <i>marinoi</i>         | SM1012Hels-07  | MMETSP0319 | 24804 | 311 | 1292 | 283 |
| Mediophyceae       | <i>Skeletonema</i>      | <i>marinoi</i>         | UNC1201        | MMETSP1428 | 11613 | 278 | 1210 | 295 |
| Mediophyceae       | <i>Skeletonema</i>      | <i>menzelii</i>        | CCMP793        | MMETSP0603 | 11681 | 306 | 1154 | 291 |
| Mediophyceae       | <i>Stephanopyxis</i>    | <i>turris</i>          | CCMP815        | MMETSP0794 | 15594 | 223 | 1391 | 244 |
| Mediophyceae       | <i>Thalassiosira</i>    | <i>antarctica</i>      | CCMP982        | MMETSP0902 | 19602 | 328 | 1386 | 304 |
| Mediophyceae       | <i>Thalassiosira</i>    | <i>gravida</i>         | Gmp14c1        | MMETSP0492 | 14203 | 225 | 1270 | 278 |
| Mediophyceae       | <i>Thalassiosira</i>    | <i>miniscula</i>       | CCMP1093       | MMETSP0737 | 25756 | 318 | 1463 | 299 |
| Mediophyceae       | <i>Thalassiosira</i>    | <i>oceanica</i>        | CCMP1005       | MMETSP0970 | 21946 | 264 | 1085 | 288 |
| Mediophyceae       | <i>Thalassiosira</i>    | <i>punctigera</i>      | Tpunct2005C2   | MMETSP1067 | 17890 | 288 | 1303 | 292 |
| Mediophyceae       | <i>Thalassiosira</i>    | <i>rotula</i>          | CCMP3096       | MMETSP0403 | 19018 | 335 | 1414 | 297 |
| Mediophyceae       | <i>Thalassiosira</i>    | <i>rotula</i>          | GSO102         | MMETSP0910 | 15286 | 245 | 1340 | 283 |
| Mediophyceae       | <i>Thalassiosira</i>    | <i>sp.</i>             | CCMP353        | MMETSP1058 | 14896 | 338 | 1315 | 299 |
| Mediophyceae       | <i>Thalassiosira</i>    | <i>sp.</i>             | FW             | MMETSP1059 | 9901  | 291 | 1125 | 285 |
| Mediophyceae       | <i>Thalassiosira</i>    | <i>sp.</i>             | NH16           | MMETSP1071 | 29455 | 318 | 1340 | 295 |
| Mediophyceae       | <i>Thalassiosira</i>    | <i>weissflogii</i>     | CCMP1010       | MMETSP0898 | 14205 | 291 | 1143 | 298 |
| Mediophyceae       | <i>Thalassiosira</i>    | <i>weissflogii</i>     | CCMP1336       | MMETSP0878 | 13030 | 303 | 1207 | 296 |
| Mediophyceae       | <i>Triceratium</i>      | <i>dubium</i>          | CCMP147        | MMETSP1175 | 10858 | 129 | 889  | 196 |
| <b>Bicosoecida</b> |                         |                        |                |            |       |     |      |     |
| Cafeteriaeae       | <i>Cafeteria</i>        | <i>roenbergensis</i>   | E4-10          | MMETSP0942 | 14668 | 168 | 210  | 5   |
| Cafeteriaeae       | <i>Cafeteria</i>        | <i>sp.</i>             | CaronLabsolate | MMETSP1104 | 11066 | 238 | 414  | 18  |
| Unclassified       | <i>Bicosoecid</i>       | <i>sp.</i>             | ms1            | MMETSP0115 | 11280 | 162 | 456  | 29  |
| <b>Bolidomonas</b> |                         |                        |                |            |       |     |      |     |
| Bolidomonas        | <i>Bolidomonas</i>      | <i>pacifica</i>        | CCMP1866       | MMETSP0785 | 18382 | 263 | 1068 | 126 |
| Bolidomonas        | <i>Bolidomonas</i>      | <i>pacifica</i>        | RCC208         | MMETSP1319 | 13441 | 231 | 1245 | 148 |
| Bolidomonas        | <i>Bolidomonas</i>      | <i>sp.</i>             | RCC1657        | MMETSP1321 | 12631 | 215 | 1003 | 135 |
| Bolidomonas        | <i>Bolidomonas</i>      | <i>sp.</i>             | RCC2347        | MMETSP1320 | 10867 | 209 | 1032 | 141 |

|                            |                          |                        |                  |            |       |     |      |     |
|----------------------------|--------------------------|------------------------|------------------|------------|-------|-----|------|-----|
| <b>Chrysochyceae</b>       |                          |                        |                  |            |       |     |      |     |
| Chromulinales              | <i>Chromulina</i>        | <i>nebulosa</i>        | UTEXLB2642       | MMETSP1095 | 9023  | 218 | 887  | 51  |
| Chromulinales              | <i>Dinobryon</i>         | <i>sp.</i>             | UTEXLB2267       | MMETSP0019 | 17759 | 286 | 1007 | 75  |
| Chromulinales              | <i>Ochromonas</i>        | <i>sp.</i>             | BG-1             | MMETSP1105 | 18065 | 267 | 891  | 71  |
| Chromulinales              | <i>Ochromonas</i>        | <i>sp.</i>             | CCMP1393         | MMETSP0004 | 13527 | 278 | 1157 | 92  |
| Chromulinales              | <i>Ochromonas</i>        | <i>sp.</i>             | CCMP1899         | MMETSP1177 | 12761 | 310 | 1115 | 68  |
| Chromulinales              | <i>Paraphysomonas</i>    | <i>bandaiensis</i>     | CaronLabIsolate  | MMETSP1103 | 13825 | 364 | 1135 | 77  |
| Chromulinales              | <i>Paraphysomonas</i>    | <i>imperfurata</i>     | PA2              | MMETSP0103 | 13042 | 279 | 903  | 62  |
| Chromulinales              | <i>Paraphysomonas</i>    | <i>vestita</i>         | GFflagA          | MMETSP1107 | 11004 | 252 | 1011 | 76  |
| Chromulinales              | <i>Spumella</i>          | <i>elongata</i>        | CCAP955/1        | MMETSP1098 | 14732 | 292 | 1469 | 92  |
| Unclassified               | <i>Chrysochyceae</i>     | <i>sp.</i>             | CCMP2298         | MMETSP1141 | 20099 | 263 | 1257 | 96  |
| <b>Dictyochophyceae</b>    |                          |                        |                  |            |       |     |      |     |
| Dictyochales               | <i>Dictyocha</i>         | <i>speculum</i>        | CCMP1381         | MMETSP1174 | 19577 | 271 | 1284 | 97  |
| Florenciellales            | <i>Florenciella</i>      | <i>parvula</i>         | CCMP2471         | MMETSP1344 | 23779 | 207 | 1255 | 90  |
| Florenciellales            | <i>Florenciella</i>      | <i>parvula</i>         | RCC1693          | MMETSP1323 | 9983  | 122 | 1003 | 106 |
| Florenciellales            | <i>Florenciella</i>      | <i>sp.</i>             | RCC1007          | MMETSP1325 | 13948 | 162 | 1033 | 67  |
| Florenciellales            | <i>Florenciella</i>      | <i>sp.</i>             | RCC1587          | MMETSP1324 | 9766  | 91  | 811  | 95  |
| Pedinellales               | <i>Pseudopedinella</i>   | <i>elastica</i>        | CCMP716          | MMETSP1068 | 23592 | 267 | 1391 | 112 |
| Pedinellales               | <i>Pteridomonas</i>      | <i>danica</i>          | PT               | MMETSP0101 | 19601 | 287 | 1047 | 80  |
| Rhizochromulinales         | <i>Rhizochromulina</i>   | <i>c.f. marina</i>     | CCMP1243         | MMETSP1173 | 13413 | 254 | 1166 | 105 |
| Unclassified               | Dictyochophyceae         | <i>sp.</i>             | CCMP2098         | MMETSP0990 | 41646 | 303 | 1347 | 110 |
| <b>Labyrinthulomycetes</b> |                          |                        |                  |            |       |     |      |     |
| Thraustochytriales         | <i>Aplanochytrium</i>    | <i>sp.</i>             | PBS07            | MMETSP0954 | 9488  | 242 | 854  | 61  |
| Thraustochytriales         | <i>Aplanochytrium</i>    | <i>stocchinoi</i>      | GSBS06           | MMETSP1346 | 10972 | 214 | 876  | 65  |
| Thraustochytriales         | <i>Aurantiochytrium</i>  | <i>limacinum</i>       | ATCCMYA-1381     | MMETSP0959 | 11295 | 353 | 945  | 56  |
| Thraustochytriales         | <i>Schizochytrium</i>    | <i>aggregatum</i>      | ATCC28209        | MMETSP0962 | 9829  | 237 | 731  | 58  |
| Thraustochytriales         | <i>Thraustochytrium</i>  | <i>sp.</i>             | LLF1b            | MMETSP0198 | 9468  | 320 | 852  | 50  |
| Unclassified               | Labyrinthulid            | quahog parasite        | QPX NY0313808BC1 | MMETSP0098 | 10450 | 258 | 803  | 65  |
| Unclassified               | Labyrinthulid            | quahog parasite        | QPX NY070348D    | MMETSP1433 | 9800  | 300 | 862  | 64  |
| <b>Pelagophyceae</b>       |                          |                        |                  |            |       |     |      |     |
| Pelagomonadales            | <i>Aureococcus</i>       | <i>anophagefferens</i> | CCMP1850         | MMETSP0914 | 23878 | 163 | 835  | 109 |
| Pelagomonadales            | <i>Aureombra</i>         | <i>lagunensis</i>      | CCMP1510         | MMETSP0890 | 10973 | 276 | 933  | 88  |
| Pelagomonadales            | <i>Pelagococcus</i>      | <i>subviridis</i>      | CCMP1429         | MMETSP0882 | 12932 | 155 | 813  | 79  |
| Pelagomonadales            | <i>Pelagomonas</i>       | <i>calceolata</i>      | CCMP1756         | MMETSP0886 | 14685 | 273 | 1105 | 111 |
| Pelagomonadales            | <i>Pelagomonas</i>       | <i>calceolata</i>      | RCC969           | MMETSP1328 | 12251 | 214 | 1035 | 102 |
| Sarcinochrysidales         | <i>Chrysocystis</i>      | <i>fragilis</i>        | CCMP3189         | MMETSP1165 | 10198 | 159 | 799  | 96  |
| Sarcinochrysidales         | <i>Chrysoreinhardtia</i> | <i>sp.</i>             | CCMP2950         | MMETSP1164 | 12347 | 165 | 727  | 105 |
| Sarcinochrysidales         | <i>Chrysoreinhardtia</i> | <i>sp.</i>             | CCMP3193         | MMETSP1166 | 12450 | 237 | 941  | 98  |
| Sarcinochrysidales         | <i>Sarcinochrysis</i>    | <i>sp.</i>             | CCMP770          | MMETSP1170 | 4448  | 69  | 423  | 60  |
| Unclassified               | Pelagophyceae            | <i>sp.</i>             | CCMP2097         | MMETSP0975 | 20678 | 285 | 1007 | 97  |
| Unclassified               | Pelagophyceae            | <i>sp.</i>             | CCMP2135         | MMETSP1467 | 12044 | 173 | 857  | 104 |
| Unclassified               | Pelagophyceae            | <i>sp.</i>             | RCC1024          | MMETSP1329 | 13073 | 176 | 957  | 99  |
| <b>Pinguiphyceae</b>       |                          |                        |                  |            |       |     |      |     |
| Pinguochrysidales          | <i>Phaeomonas</i>        | <i>parva</i>           | CCMP2877         | MMETSP1163 | 12837 | 195 | 750  | 81  |
| Pinguochrysidales          | <i>Pinguicoccus</i>      | <i>pyrenoidosus</i>    | CCMP2078         | MMETSP1160 | 7591  | 228 | 626  | 79  |
| <b>PX clade</b>            |                          |                        |                  |            |       |     |      |     |
| Xanthophyceae              | <i>Vaucheria</i>         | <i>litorea</i>         | CCMP2940         | MMETSP0945 | 8148  | 289 | 872  | 45  |
| <b>Raphidophyceae</b>      |                          |                        |                  |            |       |     |      |     |
| Chattonellales             | <i>Chattonella</i>       | <i>subsalsa</i>        | CCMP2191         | MMETSP0947 | 17677 | 330 | 1362 | 122 |
| Chattonellales             | <i>Fibrocapsa</i>        | <i>japonica</i>        | CCMP1661         | MMETSP1339 | 7569  | 183 | 954  | 81  |
| Chattonellales             | <i>Heterostigma</i>      | <i>akashwo</i>         | CCMP2393         | MMETSP0292 | 21961 | 248 | 1400 | 125 |
| Chattonellales             | <i>Heterostigma</i>      | <i>akashwo</i>         | CCMP3107         | MMETSP0409 | 13376 | 168 | 1200 | 98  |
| Chattonellales             | <i>Heterostigma</i>      | <i>akashwo</i>         | CCMP452          | MMETSP0894 | 11785 | 165 | 1222 | 101 |
| Chattonellales             | <i>Heterostigma</i>      | <i>akashwo</i>         | NB               | MMETSP0416 | 23592 | 196 | 1445 | 127 |
| <b>Synchromophyceae</b>    |                          |                        |                  |            |       |     |      |     |
| Synchroma                  | <i>Synchroma</i>         | <i>pusillum</i>        | CCMP3072         | MMETSP1452 | 9072  | 145 | 400  | 24  |
| <b>Synurophyceae</b>       |                          |                        |                  |            |       |     |      |     |
| Synurales                  | <i>Mallomonas</i>        | <i>sp.</i>             | CCMP3275         | MMETSP1167 | 14431 | 273 | 1173 | 60  |
| <b>Viridiplantae</b>       |                          |                        |                  |            |       |     |      |     |
| <b>Chlorophyta</b>         |                          |                        |                  |            |       |     |      |     |
| Chlorodendrophyceae        | <i>Tetraselmis</i>       | <i>astigmatica</i>     | CCMP880          | MMETSP0804 | 21714 | 330 | 1282 | 108 |
| Chlorodendrophyceae        | <i>Tetraselmis</i>       | <i>chuii</i>           | PLY429           | MMETSP0491 | 18733 | 236 | 1189 | 95  |
| Chlorodendrophyceae        | <i>Tetraselmis</i>       | <i>sp.</i>             | GSL018           | MMETSP0419 | 20288 | 283 | 1107 | 92  |
| Chlorodendrophyceae        | <i>Tetraselmis</i>       | <i>striata</i>         | LANL1001         | MMETSP0817 | 15261 | 287 | 1167 | 115 |
| Chlorophyceae              | <i>Chlamydomonas</i>     | <i>chlamydogama</i>    | SAG11-48b        | MMETSP1392 | 12993 | 293 | 1203 | 98  |
| Chlorophyceae              | <i>Chlamydomonas</i>     | <i>euryale</i>         | CCMP219          | MMETSP0063 | 11781 | 215 | 937  | 78  |
| Chlorophyceae              | <i>Chlamydomonas</i>     | <i>leiostraca</i>      | SAG11-49         | MMETSP1391 | 11617 | 255 | 1107 | 93  |
| Chlorophyceae              | <i>Chlamydomonas</i>     | <i>c.f. sp.</i>        | CCMP681          | MMETSP1180 | 6702  | 235 | 1051 | 81  |
| Chlorophyceae              | <i>Dunaliella</i>        | <i>tertiolecta</i>     | CCMP1320         | MMETSP1126 | 11726 | 326 | 1056 | 93  |
| Chlorophyceae              | <i>Polytomella</i>       | <i>parva</i>           | SAG63-3          | MMETSP0052 | 22218 | 280 | 648  | 43  |
| Mamiellophyceae            | <i>Bathycoccus</i>       | <i>prasinos</i>        | CCMP1898         | MMETSP1399 | 5463  | 252 | 584  | 62  |
| Mamiellophyceae            | <i>Bathycoccus</i>       | <i>prasinos</i>        | RCC716           | MMETSP1460 | 5031  | 207 | 543  | 62  |
| Mamiellophyceae            | <i>Crustomastix</i>      | <i>stigmata</i>        | CCMP3273         | MMETSP0803 | 12005 | 193 | 615  | 54  |
| Mamiellophyceae            | <i>Dolichomastix</i>     | <i>tenuilepis</i>      | CCMP3274         | MMETSP0033 | 12634 | 241 | 703  | 61  |
| Mamiellophyceae            | <i>Mamiellales</i>       | <i>sp.</i>             | RCC2288          | MMETSP1326 | 13893 | 194 | 1032 | 98  |

|                     |                        |                              |           |            |       |     |      |     |
|---------------------|------------------------|------------------------------|-----------|------------|-------|-----|------|-----|
| Mamiellophyceae     | <i>Mantoniella</i>     | <i>antarctica</i>            | SL-175    | MMETSP1106 | 19815 | 263 | 1193 | 109 |
| Mamiellophyceae     | <i>Mantoniella</i>     | <i>sp.</i>                   | CCMP1436  | MMETSP1468 | 9146  | 196 | 1063 | 102 |
| Mamiellophyceae     | <i>Micromonas</i>      | <i>pusilla</i>               | CCAC1681  | MMETSP1401 | 8923  | 251 | 900  | 109 |
| Mamiellophyceae     | <i>Micromonas</i>      | <i>pusilla</i>               | CCMP1723  | MMETSP1403 | 8136  | 249 | 843  | 109 |
| Mamiellophyceae     | <i>Micromonas</i>      | <i>pusilla</i>               | CCMP494   | MMETSP1404 | 7965  | 247 | 869  | 109 |
| Mamiellophyceae     | <i>Micromonas</i>      | <i>pusilla</i>               | RCC1614   | MMETSP1402 | 7280  | 212 | 797  | 99  |
| Mamiellophyceae     | <i>Micromonas</i>      | <i>pusilla</i>               | RCC2306   | MMETSP1327 | 7688  | 229 | 932  | 101 |
| Mamiellophyceae     | <i>Micromonas</i>      | <i>sp.</i>                   | CCMP1646  | MMETSP1080 | 12025 | 237 | 821  | 113 |
| Mamiellophyceae     | <i>Micromonas</i>      | <i>sp.</i>                   | CCMP2099  | MMETSP0802 | 7866  | 279 | 918  | 110 |
| Mamiellophyceae     | <i>Micromonas</i>      | <i>sp.</i>                   | CS-222    | MMETSP1393 | 5759  | 175 | 816  | 78  |
| Mamiellophyceae     | <i>Micromonas</i>      | <i>sp.</i>                   | NEPCC29   | MMETSP1082 | 6667  | 231 | 779  | 112 |
| Mamiellophyceae     | <i>Micromonas</i>      | <i>sp.</i>                   | RCC451    | MMETSP1400 | 8665  | 230 | 865  | 110 |
| Mamiellophyceae     | <i>Micromonas</i>      | <i>sp.</i>                   | RCC472    | MMETSP1084 | 7533  | 242 | 800  | 111 |
| Mamiellophyceae     | <i>Ostreococcus</i>    | <i>lucimarinus</i> Clade-A   | BCC118000 | MMETSP0939 | 7275  | 250 | 784  | 96  |
| Mamiellophyceae     | <i>Ostreococcus</i>    | <i>mediterraneus</i> Clade-D | RCC11107  | MMETSP0938 | 7284  | 262 | 894  | 96  |
| Mamiellophyceae     | <i>Ostreococcus</i>    | <i>mediterraneus</i> Clade-D | RCC1621   | MMETSP0930 | 7233  | 246 | 859  | 89  |
| Mamiellophyceae     | <i>Ostreococcus</i>    | <i>mediterraneus</i> Clade-D | RCC2572   | MMETSP0929 | 5188  | 272 | 815  | 99  |
| Mamiellophyceae     | <i>Ostreococcus</i>    | <i>mediterraneus</i> Clade-D | RCC2573   | MMETSP0936 | 6879  | 245 | 840  | 93  |
| Mamiellophyceae     | <i>Ostreococcus</i>    | <i>mediterraneus</i> Clade-D | RCC2589   | MMETSP0933 | 7934  | 268 | 588  | 64  |
| Mamiellophyceae     | <i>Ostreococcus</i>    | <i>mediterraneus</i> Clade-D | RCC2593   | MMETSP0937 | 6928  | 245 | 850  | 88  |
| Mamiellophyceae     | <i>Ostreococcus</i>    | <i>mediterraneus</i> Clade-D | RCC2596   | MMETSP0932 | 7692  | 281 | 891  | 98  |
| Nephrophyceae       | <i>Nephroselmis</i>    | <i>pyriformis</i>            | CCMP717   | MMETSP0034 | 21445 | 246 | 1015 | 73  |
| Palmophyllophyceae  | <i>Prasinococcus</i>   | <i>capsulatus</i>            | CCMP1194  | MMETSP0941 | 6872  | 261 | 564  | 34  |
| Palmophyllophyceae  | <i>Prasinoderma</i>    | <i>coloniale</i>             | CCMP1413  | MMETSP0806 | 8041  | 231 | 697  | 66  |
| Palmophyllophyceae  | <i>Prasinoderma</i>    | <i>singularis</i>            | RCC927    | MMETSP1315 | 8453  | 213 | 738  | 64  |
| Prasinophytes       | <i>Picocystis</i>      | <i>salinarum</i>             | CCMP1897  | MMETSP0807 | 5364  | 263 | 464  | 33  |
| Prasinophytes       | Prasinophyceae         | <i>sp.</i>                   | CCMP1205  | MMETSP1469 | 5782  | 216 | 833  | 35  |
| Prasinophytes       | Prasinophyceae         | <i>sp.</i>                   | CCMP2111  | MMETSP1446 | 5195  | 243 | 770  | 37  |
| Prasinophytes       | Prasinophyceae         | <i>sp.</i> CladeVIIA1        | RCC701    | MMETSP1453 | 5759  | 165 | 851  | 32  |
| Prasinophytes       | Prasinophyceae         | <i>sp.</i> CladeVIIA1        | RCC998    | MMETSP1309 | 5855  | 279 | 890  | 41  |
| Prasinophytes       | Prasinophyceae         | <i>sp.</i> CladeVIIA4        | RCC1871   | MMETSP1456 | 5745  | 265 | 809  | 40  |
| Prasinophytes       | Prasinophyceae         | <i>sp.</i> CladeVIIA4        | RCC2335   | MMETSP1312 | 5440  | 249 | 761  | 32  |
| Prasinophytes       | Prasinophyceae         | <i>sp.</i> CladeVIIA5        | CCMP2175  | MMETSP1470 | 15520 | 249 | 2121 | 50  |
| Prasinophytes       | Prasinophyceae         | <i>sp.</i> CladeVIIA5        | RCC856    | MMETSP1311 | 5355  | 128 | 717  | 21  |
| Prasinophytes       | Prasinophyceae         | <i>sp.</i> CladeVIIIB2       | RCC2339   | MMETSP1310 | 16347 | 277 | 1142 | 54  |
| Pyramimonadophyceae | <i>Polyblepharides</i> | <i>amylifera</i>             | CCMP720   | MMETSP1081 | 13465 | 249 | 1201 | 90  |
| Pyramimonadophyceae | <i>Pterosperma</i>     | <i>sp.</i>                   | CCMP1384  | MMETSP1438 | 10127 | 201 | 888  | 55  |
| Pyramimonadophyceae | <i>Pycnococcus</i>     | <i>provasolii</i>            | RCC2336   | MMETSP1316 | 6973  | 199 | 392  | 24  |
| Pyramimonadophyceae | <i>Pycnococcus</i>     | <i>provasolii</i>            | RCC251    | MMETSP1472 | 5902  | 108 | 359  | 24  |
| Pyramimonadophyceae | <i>Pycnococcus</i>     | <i>provasolii</i>            | RCC733    | MMETSP1471 | 5528  | 139 | 366  | 24  |
| Pyramimonadophyceae | <i>Pycnococcus</i>     | <i>provasolii</i>            | RCC931    | MMETSP1459 | 6584  | 192 | 378  | 26  |
| Pyramimonadophyceae | <i>Pycnococcus</i>     | <i>sp.</i>                   | CCMP1998  | MMETSP1085 | 5949  | 260 | 808  | 42  |
| Pyramimonadophyceae | Pyramimonas            | <i>obovata</i>               | CCMP722   | MMETSP1169 | 13328 | 258 | 1213 | 99  |
| Pyramimonadophyceae | Pyramimonas            | <i>parkeae</i>               | CCMP726   | MMETSP0058 | 19259 | 308 | 1228 | 96  |
| Pyramimonadophyceae | Pyramimonas            | <i>sp.</i>                   | CCMP2087  | MMETSP1445 | 12907 | 270 | 1239 | 101 |
| Trebouxiophyceae    | Picochlorum            | <i>oklahomensis</i>          | CCMP2329  | MMETSP1161 | 4681  | 192 | 732  | 66  |
| Trebouxiophyceae    | Picochlorum            | <i>sp.</i>                   | RCC944    | MMETSP1330 | 5885  | 176 | 895  | 67  |
| Trebouxiophyceae    | Stichococcus           | <i>sp.</i>                   | RCC1054   | MMETSP1473 | 7290  | 226 | 705  | 54  |

Supplementary Table 5. Trophic Classifications of MMETSP Dataset.

| Species Name                                   | Trophic Group |
|--|---------------|
| <i>Acanthoecca</i> like sp 10tr                | Heterotrophic |
| <i>Akashiwo sanguinea</i> CCCM885              | Mixotrophic   |
| <i>Alexandrium andersonii</i> CCMP2222         | Autotrophic   |
| <i>Alexandrium catenella</i> OF101             | Mixotrophic   |
| <i>Alexandrium fundyense</i> CCMP1719          | Mixotrophic   |
| <i>Alexandrium margalefi</i> AMGDE01CS 322     | Autotrophic   |
| <i>Alexandrium minutum</i> CCMP113             | Mixotrophic   |
| <i>Alexandrium monilatum</i> CCMP3105          | Mixotrophic   |
| <i>Alexandrium tamarensense</i> CCMP1771       | Mixotrophic   |
| <i>Ammonia</i> sp                              | Heterotrophic |
| <i>Amoebophrya</i> sp Ameob2                   | Parasitic     |
| <i>Amorphochlora amoebiformis</i> CCMP2058     | Mixotrophic   |
| <i>Amphidinium carterae</i> CCMP1314           | Mixotrophic   |
| <i>Amphidinium massartii</i> CS 259            | Mixotrophic   |
| <i>Amphiprora paludosa</i> CCMP125             | Autotrophic   |
| <i>Amphiprora</i> sp CCMP467                   | Autotrophic   |
| <i>Amphora coffeaeformis</i> CCMP127           | Autotrophic   |
| <i>Anophryoides haemophila</i> AH6             | Parasitic     |
| <i>Aplanochytrium</i> sp PBS07                 | Heterotrophic |
| <i>Aplanochytrium stocchinoi</i> GSBS06        | Heterotrophic |
| <i>Aristerostoma</i> sp ATCC50986              | Heterotrophic |
| <i>Asterionellopsis glacialis</i>              | Autotrophic   |
| <i>Asterionellopsis glacialis</i> CCMP134      | Autotrophic   |
| <i>Asterionellopsis glacialis</i> CCMP1581     | Autotrophic   |
| <i>Astrosyne radiata</i> 13vi08 1A             | Autotrophic   |
| <i>Attheya septentrionalis</i> CCMP2084        | Autotrophic   |
| <i>Aulacoseira subarctica</i> CCAP1002 5       | Autotrophic   |
| <i>Aurantiochytrium limacinum</i> ATCCMYA 1381 | Heterotrophic |
| <i>Aureococcus anophagefferens</i> CCMP1850    | Mixotrophic   |
| <i>Aureoumbra lagunensis</i> CCMP1510          | Autotrophic   |
| <i>Azadinium spinosum</i> 3D9                  | Autotrophic   |
| Bangiophyceae sp CCMP1999                      | Autotrophic   |
| <i>Bathycoccus prasinus</i> CCMP1898           | Autotrophic   |
| <i>Bathycoccus prasinus</i> RCC716             | Autotrophic   |
| <i>Bicosoecid</i> sp ms1                       | Heterotrophic |
| <i>Bigelowiella longifila</i> CCMP242          | Mixotrophic   |
| <i>Bigelowiella natans</i> CCMP1242            | Mixotrophic   |
| <i>Bigelowiella natans</i> CCMP1258-1          | Mixotrophic   |
| <i>Bigelowiella natans</i> CCMP1259            | Mixotrophic   |
| <i>Bigelowiella natans</i> CCMP2755            | Mixotrophic   |

|   |               |
|---|---------------|
| <i>Bigelowiella natans</i> CCMP623                | Mixotrophic   |
| <i>Blepharisma japonicum</i> StockR1072           | Heterotrophic |
| <i>Bolidomonas pacifica</i> CCMP1866              | Autotrophic   |
| <i>Bolidomonas pacifica</i> RCC208                | Autotrophic   |
| <i>Bolidomonas sp</i> RCC1657                     | Autotrophic   |
| <i>Bolidomonas sp</i> RCC2347                     | Autotrophic   |
| <i>Brandtodinium nutriculum</i> RCC3387           | Symbiont      |
| <i>Cafeteria roenbergensis</i> E4-10              | Heterotrophic |
| <i>Cafeteria sp</i> CaronLabIsolate               | Heterotrophic |
| <i>Calcidiscus leptoporus</i> RCC1130             | Autotrophic   |
| <i>Ceratium fusus</i> PA161109                    | Mixotrophic   |
| <i>Chaetoceros affinis</i> CCMP159                | Autotrophic   |
| <i>Chaetoceros brevis</i> CCMP164                 | Autotrophic   |
| <i>Chaetoceros cf neogracile</i> RCC1993          | Autotrophic   |
| <i>Chaetoceros curvisetus</i>                     | Autotrophic   |
| <i>Chaetoceros debilis</i> MM31A-1                | Autotrophic   |
| <i>Chaetoceros dichchaeta</i> CCMP1751            | Autotrophic   |
| <i>Chaetoceros neogracile</i> CCMP1317            | Autotrophic   |
| <i>Chaetoceros sp</i> GSL56                       | Autotrophic   |
| <i>Chaetoceros sp</i> UNC1202                     | Autotrophic   |
| <i>Chattonella subsalsa</i> CCMP2191              | Mixotrophic   |
| <i>Chlamydomonas chlamydogama</i> SAG11-48b       | Autotrophic   |
| <i>Chlamydomonas euryale</i> CCMP219              | Autotrophic   |
| <i>Chlamydomonas leiostraca</i> SAG11-49          | Autotrophic   |
| <i>Chlamydomonas cf sp</i> CCMP681                | Autotrophic   |
| <i>Chlorarachnion reptans</i> CCCM449             | Mixotrophic   |
| <i>Chromera velia</i> CCMP2878                    | Mixotrophic   |
| <i>Chromulina nebulosa</i> UTEXLB2642             | Mixotrophic   |
| <i>Chroomonas mesostigmaticacf</i> CCMP1168       | Mixotrophic   |
| <i>Chrysochromulina rotalis</i> UIO044            | Mixotrophic   |
| <i>Chrysoculter rhomboideus</i> RCC1486           | Autotrophic   |
| <i>Chrysocystis fragilis</i> CCMP3189             | Autotrophic   |
| <i>Chrysophyceae sp</i> CCMP2298                  | Mixotrophic   |
| <i>Chrysoreinhardia sp</i> CCMP2950               | Autotrophic   |
| <i>Chrysoreinhardia sp</i> CCMP3193               | Autotrophic   |
| <i>Climacostomum virens</i> StockW-24             | Heterotrophic |
| <i>Coccolithus pelagicuss sp braarudi</i> PLY182g | Mixotrophic   |
| <i>Compsopogon coeruleus</i> SAG36-94             | Autotrophic   |
| <i>Condylostoma magnum</i> COL2                   | Heterotrophic |
| <i>Corethron hystrix</i> 308                      | Autotrophic   |
| <i>Corethron pennatum</i> L29A3                   | Autotrophic   |
| <i>Coscinodiscus wailesii</i> CCMP2513            | Autotrophic   |
| <i>Craspedostauros australis</i> CCMP3328         | Autotrophic   |

|  |               |
|--|---------------|
| <i>Crustomastix stigmata</i> CCMP3273        | Autotrophic   |
| <i>Crypthecodinium cohnii</i> Seligo         | Heterotrophic |
| <i>Cryptomonas curvata</i> CCAP979/52        | Mixotrophic   |
| <i>Cryptomonas paramecium</i> CCAP977/2a     | Mixotrophic   |
| <i>Cryptophyceae sp</i> CCMP2293             | Autotrophic   |
| <i>Cunea sp</i> BSH 02190019                 | Heterotrophic |
| <i>Cyanoptyche gloeocystis</i> SAG4-97       | Autotrophic   |
| <i>Cyclophora tenuis</i> ECT3854             | Autotrophic   |
| <i>Cyclotella meneghiniana</i> CCMP338       | Autotrophic   |
| <i>Cylindrotheca closterium</i> KMMCC-B-181  | Autotrophic   |
| <i>Dactyliosolen fragilissimus</i>           | Autotrophic   |
| <i>Debaryomyces hansenii</i> J26             | Heterotrophic |
| <i>Dictyocha speculum</i> CCMP1381           | Autotrophic   |
| Dictyochophyceae sp CCMP2098                 | Mixotrophic   |
| <i>Dinobryon sp</i> UTEXLB2267               | Mixotrophic   |
| <i>Dinophysis acuminata</i> DAEP01           | Mixotrophic   |
| <i>Ditylum brightwellii</i> GSO103           | Autotrophic   |
| <i>Ditylum brightwellii</i> GSO104           | Autotrophic   |
| <i>Ditylum brightwellii</i> GSO105           | Autotrophic   |
| <i>Ditylum brightwellii</i> Pop1-SS4         | Autotrophic   |
| <i>Ditylum brightwellii</i> Pop2-SS10        | Autotrophic   |
| <i>Dolichomastix tenuilepis</i> CCMP3274     | Autotrophic   |
| <i>Dracoamoeba jomungandri</i> Chinc5        | Heterotrophic |
| <i>Dunaliella tertiolecta</i> CCMP1320       | Mixotrophic   |
| <i>Durinskia baltica</i> CSIROCS-38          | Autotrophic   |
| <i>Elphidium margaritaceum</i>               | Heterotrophic |
| <i>Emiliana huxleyi</i> 374                  | Autotrophic   |
| <i>Emiliana huxleyi</i> 379                  | Autotrophic   |
| <i>Emiliana huxleyi</i> CCMP370              | Autotrophic   |
| <i>Emiliana huxleyi</i> PLYM219              | Autotrophic   |
| <i>Entomoneis sp</i> CCMP2396                | Autotrophic   |
| <i>Erythrolobus australicus</i> CCMP3124     | Autotrophic   |
| <i>Erythrolobus madagascarensis</i> CCMP3276 | Autotrophic   |
| <i>Eucampia antarctica</i> CCMP1452          | Autotrophic   |
| <i>Euplotes crassus</i> CT5                  | Heterotrophic |
| <i>Euplotes focardii</i> TN1                 | Heterotrophic |
| <i>Euplotes harpa</i> FSP1-4                 | Heterotrophic |
| <i>Eutreptiella gymnastica</i> like CCMP1594 | Mixotrophic   |
| <i>Eutreptiella gymnastica</i> NIES 381      | Mixotrophic   |
| <i>Exanthemachrysis gayraliae</i> RCC1523    | Mixotrophic   |
| <i>Extubocellulus spinifer</i> CCMP396       | Autotrophic   |
| <i>Fabrea salina</i>                         | Heterotrophic |
| <i>Favella ehrenbergii</i> Fehren1           | Heterotrophic |

|   |               |
|---|---------------|
| <i>Fibrocapsa japonica</i> CCMP1661       | Mixotrophic   |
| <i>Filamoeba nolandi</i> NC AS 23 1       | Heterotrophic |
| <i>Florenciella parvula</i> CCMP2471      | Autotrophic   |
| <i>Florenciella parvula</i> RCC1693       | Autotrophic   |
| <i>Florenciella sp</i> RCC1007            | Autotrophic   |
| <i>Florenciella sp</i> RCC1587            | Autotrophic   |
| <i>Fragilariopsis kerguelensis</i> L2 C3  | Autotrophic   |
| <i>Fragilariopsis kerguelensis</i> L26 C5 | Autotrophic   |
| <i>Gambierdiscus australes</i> CAWD149    | Autotrophic   |
| <i>Geminigera cryophila</i> CCMP2564      | Mixotrophic   |
| <i>Geminigera sp</i> CaronLabIsolate      | Mixotrophic   |
| <i>Gephyrocapsa oceanica</i> RCC1303      | Autotrophic   |
| <i>Gloeochaete wittrockiana</i> SAG46 84  | Autotrophic   |
| <i>Goniomonas pacifica</i> CCMP1869       | Heterotrophic |
| <i>Goniomonas sp m</i>                    | Heterotrophic |
| <i>Gonyaulax spinifera</i> CCMP409        | Mixotrophic   |
| <i>Grammatophora oceanica</i> CCMP410     | Autotrophic   |
| <i>Guillardia theta</i> CCMP2712          | Autotrophic   |
| <i>Gymnochlora sp</i> CCMP2014            | Mixotrophic   |
| <i>Gymnodinium catenatum</i> GC744        | Mixotrophic   |
| <i>Gyrodinium dominans</i> SPMC103        | Heterotrophic |
| <i>Hanusia phi</i> CCMP325                | Autotrophic   |
| <i>Haptolina brevifila</i> UTEXLB985      | Mixotrophic   |
| <i>Haptolina ericina</i> CCMP281          | Mixotrophic   |
| <i>Helicotheca tamensis</i> CCMP826       | Autotrophic   |
| <i>Hemiselmis andersenii</i> CCMP1180     | Autotrophic   |
| <i>Hemiselmis andersenii</i> CCMP439      | Autotrophic   |
| <i>Hemiselmis andersenii</i> CCMP441      | Autotrophic   |
| <i>Hemiselmis andersenii</i> CCMP644      | Autotrophic   |
| <i>Hemiselmis rufescens</i> PCC563        | Autotrophic   |
| <i>Hemiselmis tepida</i> CCMP443          | Autotrophic   |
| <i>Hemiselmis virescens</i> PCC157        | Autotrophic   |
| <i>Heterocapsa arctica</i> CCMP445        | Mixotrophic   |
| <i>Heterocapsa rotundata</i> SCCAPK 0483  | Mixotrophic   |
| <i>Heterocapsa triquetra</i> CCMP448      | Mixotrophic   |
| <i>Heterosigma akashiwo</i> CCMP2393      | Mixotrophic   |
| <i>Heterosigma akashiwo</i> CCMP3107      | Mixotrophic   |
| <i>Heterosigma akashiwo</i> CCMP452       | Mixotrophic   |
| <i>Heterosigma akashiwo</i> NB            | Mixotrophic   |
| <i>Imantonia sp</i> RCC918                | Mixotrophic   |
| <i>Isochrysis galbana</i> CCMP1323        | Autotrophic   |
| <i>Isochrysis sp</i> CCMP1244             | Autotrophic   |
| <i>Isochrysis sp</i> CCMP1324             | Autotrophic   |

|  |               |
|--|---------------|
| <i>Karenia brevis</i> CCMP2229                 | Mixotrophic   |
| <i>Karenia brevis</i> SP1                      | Mixotrophic   |
| <i>Karenia brevis</i> SP3                      | Mixotrophic   |
| <i>Karenia brevis</i> Wilson                   | Mixotrophic   |
| <i>Karlodinium veneficum</i> CCMP2283          | Mixotrophic   |
| <i>Kryptoperidinium foliaceum</i> CCAP1116 3   | Symbiont      |
| <i>Kryptoperidinium foliaceum</i> CCMP1326     | Symbiont      |
| Labyrinthulid quahog parasite QPX NY0313808BC1 | Parasitic     |
| Labyrinthulid quahog parasite QPX NY070348D    | Parasitic     |
| <i>Lankesteria abbottii</i> GrapplerInletBC    | Mixotrophic   |
| <i>Leptocylindrus aporus</i> B651              | Autotrophic   |
| <i>Leptocylindrus danicus</i> B650             | Autotrophic   |
| <i>Leptocylindrus danicus</i> CCMP1856         | Autotrophic   |
| <i>Lessardia elongata</i> AF521100.1           | Heterotrophic |
| <i>Licmophora paradoxa</i> CCMP2313            | Autotrophic   |
| <i>Lingulodinium polyedra</i> CCMP1738         | Mixotrophic   |
| <i>Litonotus pictus</i> P1                     | Heterotrophic |
| <i>Lotharella globosa</i> CCCM811              | Mixotrophic   |
| <i>Lotharella globosa</i> LEX01                | Mixotrophic   |
| <i>Lotharella oceanica</i> CCMP622             | Mixotrophic   |
| <i>Madagascaria erythrocladioides</i> CCMP3234 | Autotrophic   |
| <i>Mallomonas</i> sp CCMP3275                  | Autotrophic   |
| <i>Mamiellales</i> sp RCC2288                  | Autotrophic   |
| <i>Mantoniella antarctica</i> SL-175           | Mixotrophic   |
| <i>Mantoniella</i> sp CCMP1436                 | Mixotrophic   |
| <i>Mataza</i> sp SIOpierMataz1                 | Mixotrophic   |
| <i>Mesodinium pulex</i> SPMC105                | Mixotrophic   |
| <i>Micromonas pusilla</i> CCAC1681             | Autotrophic   |
| <i>Micromonas pusilla</i> CCMP1723             | Autotrophic   |
| <i>Micromonas pusilla</i> CCMP494              | Autotrophic   |
| <i>Micromonas pusilla</i> RCC1614              | Autotrophic   |
| <i>Micromonas pusilla</i> RCC2306              | Autotrophic   |
| <i>Micromonas</i> sp CCMP1646                  | Autotrophic   |
| <i>Micromonas</i> sp CCMP2099                  | Mixotrophic   |
| <i>Micromonas</i> sp CS 222                    | Autotrophic   |
| <i>Micromonas</i> sp NEPCC29                   | Autotrophic   |
| <i>Micromonas</i> sp RCC451                    | Autotrophic   |
| <i>Micromonas</i> sp RCC472                    | Autotrophic   |
| <i>Minchinia chitonis</i>                      | Autotrophic   |
| <i>Minutocellus polymorphus</i> CCMP3303       | Autotrophic   |
| <i>Minutocellus polymorphus</i> NH13           | Autotrophic   |
| <i>Minutocellus polymorphus</i> RCC2270        | Autotrophic   |
| <i>Myrionecta rubra</i> CCMP2563               | Mixotrophic   |



|   |               |
|---|---------------|
| <i>Neobodo designis</i> CCAP1951/1                | Heterotrophic |
| <i>Nephroselmis pyriformis</i> CCMP717            | Autotrophic   |
| <i>Nitzschia</i> sp RCC80                         | Autotrophic   |
| <i>Noctiluca scintillans</i>                      | Heterotrophic |
| <i>Norrisiella sphaerica</i> BC52                 | Autotrophic   |
| <i>Ochromonas</i> sp BG 1                         | Mixotrophic   |
| <i>Ochromonas</i> sp CCMP1393                     | Mixotrophic   |
| <i>Ochromonas</i> sp CCMP1899                     | Mixotrophic   |
| <i>Odontella aurita</i> isolate1302-5             | Autotrophic   |
| <i>Odontella sinensis</i> Grunow1884              | Autotrophic   |
| <i>Ostreococcus lucimarinus</i> clade A BCC118000 | Autotrophic   |
| <i>Ostreococcus mediterraneus</i> clade D RCC1107 | Autotrophic   |
| <i>Ostreococcus mediterraneus</i> clade D RCC1621 | Autotrophic   |
| <i>Ostreococcus mediterraneus</i> clade D RCC2572 | Autotrophic   |
| <i>Ostreococcus mediterraneus</i> clade D RCC2573 | Autotrophic   |
| <i>Ostreococcus mediterraneus</i> clade D RCC2589 | Autotrophic   |
| <i>Ostreococcus mediterraneus</i> clade D RCC2593 | Autotrophic   |
| <i>Ostreococcus mediterraneus</i> clade D RCC2596 | Autotrophic   |
| <i>Oxyrrhis marina</i>                            | Heterotrophic |
| <i>Oxyrrhis marina</i> CCMP1788                   | Heterotrophic |
| <i>Oxyrrhis marina</i> CCMP1795                   | Heterotrophic |
| <i>Oxyrrhis marina</i> LB1974                     | Heterotrophic |
| <i>Palpitomonas bilix</i> NIES 2562 AB508339      | Heterotrophic |
| <i>Paramoeba aestuarina</i> SoJaBioB1-5-56-2      | Heterotrophic |
| <i>Paramoeba atlantica</i> 621 1 CCAP1560/9       | Heterotrophic |
| <i>Paraphysomonas bandaiensis</i> CaronLabIsolate | Heterotrophic |
| <i>Paraphysomonas imperforata</i> PA2             | Heterotrophic |
| <i>Paraphysomonas vestita</i> GFlagA              | Heterotrophic |
| <i>Parduczia</i> like sp Undescribed              | Heterotrophic |
| <i>Partenskyella glossopodia</i> RCC365           | Autotrophic   |
| <i>Pavlova gyrans</i> CCMP608                     | Mixotrophic   |
| <i>Pavlova lutheri</i> RCC1537 JF714236.1         | Mixotrophic   |
| <i>Pavlova</i> sp CCMP459                         | Mixotrophic   |
| <i>Pavlova</i> sp CCMP2436                        | Mixotrophic   |
| <i>Pelagococcus subviridis</i> CCMP1429           | Autotrophic   |
| <i>Pelagodinium beii</i> RCC1491                  | Symbiont      |
| <i>Pelagomonas calceolata</i> CCMP1756            | Autotrophic   |
| <i>Pelagomonas calceolata</i> RCC969              | Autotrophic   |
| <i>Pelagophyceae</i> sp CCMP2097                  | Autotrophic   |
| <i>Pelagophyceae</i> sp CCMP2135                  | Autotrophic   |
| <i>Pelagophyceae</i> sp RCC1024                   | Autotrophic   |
| <i>Percolomonas cosmopolitus</i> AE-1 ATCC50343   | Heterotrophic |
| <i>Percolomonas cosmopolitus</i> WS               | Heterotrophic |

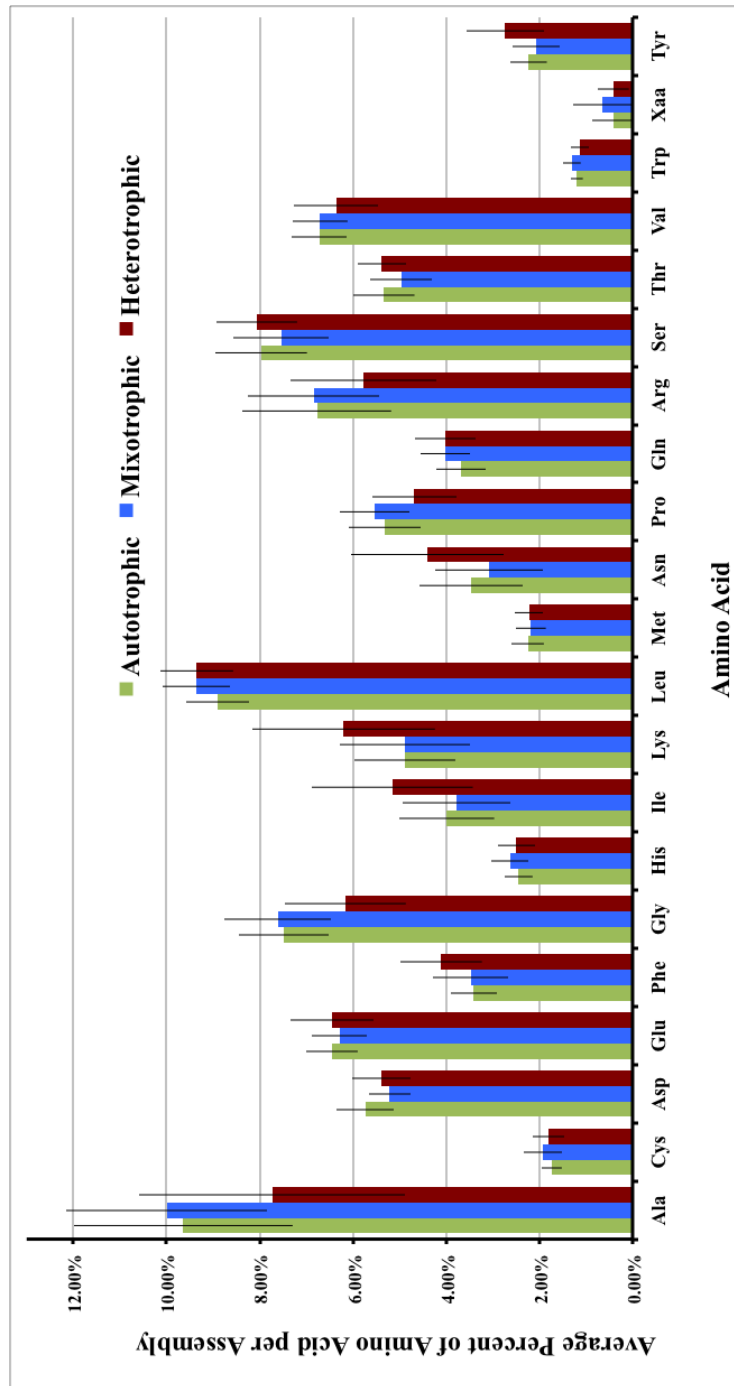
|   |               |
|---|---------------|
| <i>Peridinium aciculiferum</i> PAER-2             | Autotrophic   |
| <i>Perkinsus chesapeaki</i> ATCCPRA-65            | Parasitic     |
| <i>Perkinsus marinus</i> ATCC50439                | Parasitic     |
| <i>Pessonella</i> sp PRA-29                       | Heterotrophic |
| <i>Phaeocystis antarctica</i> CaronLabIsolate     | Autotrophic   |
| <i>Phaeocystis antarctica</i> CCMP1374            | Autotrophic   |
| <i>Phaeocystis cordata</i> RCC1383                | Autotrophic   |
| <i>Phaeocystis rex</i> CCMP2000                   | Autotrophic   |
| <i>Phaeocystis</i> sp CCMP2710                    | Autotrophic   |
| <i>Phaeomonas parva</i> CCMP2877                  | Autotrophic   |
| <i>Picochlorum oklahomensis</i> CCMP2329          | Autotrophic   |
| <i>Picochlorum</i> sp RCC944                      | Autotrophic   |
| <i>Picocystis salinarum</i> CCMP1897              | Autotrophic   |
| <i>Pinguicoccus pyrenoidosus</i> CCMP2078         | Autotrophic   |
| <i>Platyophrya macrostoma</i> WH                  | Heterotrophic |
| <i>Pleurochrysis carterae</i> CCMP645             | Autotrophic   |
| <i>Polarella glacialis</i> CCMP1383               | Autotrophic   |
| <i>Polarella glacialis</i> CCMP2088               | Autotrophic   |
| <i>Polyblepharides amyliifera</i> CCMP720         | Autotrophic   |
| <i>Polytomella parva</i> SAG63 3                  | Autotrophic   |
| <i>Porphyridium aerugineum</i> SAG1380 2          | Autotrophic   |
| <i>Prasinococcus capsulatus</i> CCMP1194          | Autotrophic   |
| <i>Prasinoderma coloniale</i> CCMP1413 FN562437.1 | Autotrophic   |
| <i>Prasinoderma singularis</i> RCC927             | Autotrophic   |
| Prasinophyceae sp CCMP1205                        | Autotrophic   |
| Prasinophyceae sp CCMP2111                        | Autotrophic   |
| Prasinophyceae sp CladeVIIA1 RCC701               | Autotrophic   |
| Prasinophyceae sp CladeVIIA1 RCC998               | Autotrophic   |
| Prasinophyceae sp CladeVIIA4 RCC1871              | Autotrophic   |
| Prasinophyceae sp CladeVIIA4 RCC2335              | Autotrophic   |
| Prasinophyceae sp CladeVIIA5 CCMP2175             | Autotrophic   |
| Prasinophyceae sp CladeVIIA5 RCC856               | Autotrophic   |
| Prasinophyceae sp CladeVIIB2 RCC2339              | Autotrophic   |
| <i>Proboscia alata</i> PI D3                      | Autotrophic   |
| <i>Proboscia inermis</i> CCAP1064/1               | Autotrophic   |
| <i>Prorocentrum lima</i> CCMP684                  | Mixotrophic   |
| <i>Prorocentrum micans</i> CCCM845                | Mixotrophic   |
| <i>Prorocentrum minimum</i> CCMP1329              | Mixotrophic   |
| <i>Prorocentrum minimum</i> CCMP2233              | Mixotrophic   |
| <i>Proteomonas sulcata</i> CCMP704                | Autotrophic   |
| <i>Protoceratium reticulatum</i> CCMP1889         | Mixotrophic   |
| <i>Protocruzia adherens</i> Boccale               | Heterotrophic |
| <i>Prymnesium parvum</i> Texoma1                  | Mixotrophic   |

|   |               |
|---|---------------|
| <i>Prymnesium polylepis</i> CCMP1757                  | Mixotrophic   |
| <i>Prymnesium polylepis</i> UIO037                    | Mixotrophic   |
| <i>Pseudo-nitzschia arenysensis</i> B593              | Mixotrophic   |
| <i>Pseudo-nitzschia australis</i> 1024910AB           | Mixotrophic   |
| <i>Pseudo-nitzschia delicatissima</i> B596            | Mixotrophic   |
| <i>Pseudo-nitzschia delicatissima</i> UNC1205         | Mixotrophic   |
| <i>Pseudo-nitzschia fraudulenta</i> WWA7              | Mixotrophic   |
| <i>Pseudo-nitzschia heimii</i> UNC1101                | Mixotrophic   |
| <i>Pseudo-nitzschia pungens cf cingulata</i>          | Mixotrophic   |
| <i>Pseudo-nitzschia pungens cf pungens</i>            | Mixotrophic   |
| <i>Pseudokeronopsis sp</i> Brazil                     | Heterotrophic |
| <i>Pseudokeronopsis sp</i> OXSARD2                    | Heterotrophic |
| <i>Pseudopedinella elastica</i> CCMP716               | Mixotrophic   |
| <i>Pteridomonas danica</i> PT                         | Heterotrophic |
| <i>Pterosperma sp</i> CCMP1384                        | Autotrophic   |
| <i>Pycnococcus provasolii</i> RCC2336                 | Autotrophic   |
| <i>Pycnococcus provasolii</i> RCC251                  | Autotrophic   |
| <i>Pycnococcus provasolii</i> RCC733                  | Autotrophic   |
| <i>Pycnococcus provasolii</i> RCC931                  | Autotrophic   |
| <i>Pycnococcus sp</i> CCMP1998                        | Autotrophic   |
| <i>Pyramimonas obovata</i> CCMP722                    | Autotrophic   |
| <i>Pyramimonas parkeae</i> CCMP726                    | Autotrophic   |
| <i>Pyramimonas sp</i> CCMP2087                        | Autotrophic   |
| <i>Pyrocystis lunula</i> CCCM517                      | Autotrophic   |
| <i>Pyrodinium bahamense</i> pbaha01                   | Autotrophic   |
| <i>Rhizochromulina marina cf</i> CCMP1243             | Autotrophic   |
| <i>Rhizosolenia setigera</i> CCMP1694                 | Autotrophic   |
| <i>Rhodella maculata</i> CCMP736                      | Autotrophic   |
| <i>Rhodomonas abbreviata</i> CaronLabIsolate          | Mixotrophic   |
| <i>Rhodomonas lens</i> RHODO                          | Mixotrophic   |
| <i>Rhodomonas salina</i> CCMP1319                     | Mixotrophic   |
| <i>Rhodomonas sp</i> CCMP768                          | Mixotrophic   |
| <i>Rhodorus marinus</i> CCMP769                       | Autotrophic   |
| <i>Rhodorus marinus</i> UTEXLB2760                    | Autotrophic   |
| <i>Rosalina sp</i>                                    | Heterotrophic |
| <i>Sapocribrum chincoteaguense</i> ATCC50979          | Heterotrophic |
| <i>Sarcinochrysis sp</i> CCMP770                      | Autotrophic   |
| <i>Schizochytrium aggregatum</i> ATCC28209 AB022106   | Heterotrophic |
| <i>Schmidingerella taraikaensis</i> FeNarragansettBay | Heterotrophic |
| <i>Scrippsiella hangoei</i> like SHHI-4               | Mixotrophic   |
| <i>Scrippsiella hangioei</i> SHTV-5                   | Mixotrophic   |
| <i>Scrippsiella trochoidea</i> CCMP3099               | Mixotrophic   |
| <i>Scyphosphaera apsteinii</i> RCC1455                | Autotrophic   |

|  |               |
|--|---------------|
| <i>Skeletonema costatum</i> 1716           | Autotrophic   |
| <i>Skeletonema dohrnii</i> SkeIB           | Autotrophic   |
| <i>Skeletonema grethea</i> CCMP1804        | Autotrophic   |
| <i>Skeletonema japonicum</i> CCMP2506      | Autotrophic   |
| <i>Skeletonema marinoi</i> FE60            | Autotrophic   |
| <i>Skeletonema marinoi</i> FE7             | Autotrophic   |
| <i>Skeletonema marinoi</i> skelA           | Autotrophic   |
| <i>Skeletonema marinoi</i> SM1012Den-03    | Autotrophic   |
| <i>Skeletonema marinoi</i> SM1012Hels-07   | Autotrophic   |
| <i>Skeletonema marinoi</i> UNC1201         | Autotrophic   |
| <i>Skeletonema menzelii</i> CCMP793        | Autotrophic   |
| <i>Sorites</i> sp                          | Heterotrophic |
| <i>Spumella elongata</i> CCAP955/1         | Heterotrophic |
| <i>Stauroneis constricta</i> CCMP1120      | Autotrophic   |
| <i>Stausosira complex</i> sp CCMP2646      | Autotrophic   |
| <i>Stephanopyxis turris</i> CCMP815        | Autotrophic   |
| <i>Stichococcus</i> sp RCC1054             | Autotrophic   |
| <i>Striatella unipunctata</i> CCMP2910     | Autotrophic   |
| <i>Strombidinopsis acuminatum</i> SPMC142  | Heterotrophic |
| <i>Strombidinopsis</i> sp SopsisLIS2011    | Heterotrophic |
| <i>Strombidium inclinatum</i> S3           | Heterotrophic |
| <i>Strombidium rassoulzadegani</i> ras09   | Mixotrophic   |
| <i>Stygamoeba regulata</i> BSH 02190019    | Heterotrophic |
| <i>Symbiodinium kawagutii</i> CCMP2468     | Symbiont      |
| <i>Symbiodinium</i> sp C1                  | Symbiont      |
| <i>Symbiodinium</i> sp C15                 | Symbiont      |
| <i>Symbiodinium</i> sp CCMP2430            | Symbiont      |
| <i>Symbiodinium</i> sp CCMP421             | Symbiont      |
| <i>Symbiodinium</i> sp cladeA              | Symbiont      |
| <i>Symbiodinium</i> sp D1a                 | Symbiont      |
| <i>Symbiodinium</i> sp Mp                  | Symbiont      |
| <i>Synchroma pusillum</i> CCMP3072         | Autotrophic   |
| <i>Synedropsis recta</i> cf CCMP1620       | Autotrophic   |
| <i>Tetraselmis astigmatica</i> CCMP880     | Autotrophic   |
| <i>Tetraselmis chuii</i> PLY429            | Autotrophic   |
| <i>Tetraselmis</i> sp GSL018               | Autotrophic   |
| <i>Tetraselmis striata</i> LANL1001        | Autotrophic   |
| <i>Thalassionema frauenfeldii</i> CCMP1798 | Autotrophic   |
| <i>Thalassionema nitzschioides</i>         | Autotrophic   |
| <i>Thalassionema nitzschioides</i> L26 B   | Autotrophic   |
| <i>Thalassiosira antarctica</i> CCMP982    | Autotrophic   |
| <i>Thalassiosira gravida</i> GMp14c1       | Autotrophic   |
| <i>Thalassiosira miniscula</i> CCMP1093    | Autotrophic   |

|  |               |
|--|---------------|
| <i>Thalassiosira oceanica</i> CCMP1005       | Autotrophic   |
| <i>Thalassiosira punctigera</i> Tpunct2005C2 | Autotrophic   |
| <i>Thalassiosira rotula</i> CCMP3096         | Autotrophic   |
| <i>Thalassiosira rotula</i> GSO102           | Autotrophic   |
| <i>Thalassiosira sp</i> CCMP353              | Autotrophic   |
| <i>Thalassiosira sp</i> FW                   | Autotrophic   |
| <i>Thalassiosira sp</i> NH16                 | Autotrophic   |
| <i>Thalassiosira weissflogii</i> CCMP1010    | Autotrophic   |
| <i>Thalassiosira weissflogii</i> CCMP1336    | Autotrophic   |
| <i>Thalassiothrix antarctica</i> L6 D1       | Autotrophic   |
| <i>Thoracosphaera heimii</i> CCMP1069        | Autotrophic   |
| <i>Thraustochytrium sp</i> LLF1b             | Heterotrophic |
| <i>Tiarina fusus</i> LIS                     | Heterotrophic |
| <i>Timpurckia oligopyrenoides</i> CCMP3278   | Autotrophic   |
| <i>Togula jolla</i> CCCM725                  | Autotrophic   |
| <i>Triceratium dubium</i> CCMP147            | Autotrophic   |
| <i>Trichosphaerium sp</i> Am-I-7wt           | Heterotrophic |
| <i>Tryblionella compressa</i> CCMP561        | Autotrophic   |
| <i>Uronema sp</i> Bbcil                      | Heterotrophic |
| <i>Vannella robusta</i> DIVA3518-3-11-1-6    | Heterotrophic |
| <i>Vannella sp</i> DIVA3517-6-12             | Heterotrophic |
| <i>Vaucheria litorea</i> CCMP2940            | Autotrophic   |
| <i>Vexillifera sp</i> DIVA3564-2             | Heterotrophic |
| <i>Vitrella brassicaformis</i> CCMP3155      | Autotrophic   |
| <i>Vitrella brassicaformis</i> CCMP3346      | Autotrophic   |

Supplemental Figure 1. Average amino acid frequency among the trophic groups of each amino acid. Ala (alanine), Cys (cysteine), Asp (asparagine), Glu (glutamine), Gly (glycine), Phe (phenylalanine), Gly (glycine), His (histidine), Ile (isoleucine), Lys (lysine), Leu (leucine), Met (methionine), Asn (asparagine), Pro (proline), Gln (glutamine), Arg (arginine), Ser (serine), Thr (threonine), Val (valine), Trp (tryptophan), Xaa (Any amino acid), Tyr (tyrosine). Standard deviation plotted as bars.




# Chapter 4: A Case Study Approach to Teaching Introductory Biology Undergraduates the Application of Genomic Data Analysis in Ecological Studies

Supplemental: Case study PowerPoint Slides:

**Transcriptome of the Toxic Tide**

By  
Joshua T. Cooper<sup>1</sup>, and J. Phil Gibson<sup>1,2</sup>



1 Department of Microbiology and Plant Biology,  
University of Oklahoma, Norman, Oklahoma

2 Department of Biology,  
University of Oklahoma, Norman, Oklahoma

Welcome to the Bayside Home Owners Association Meeting

"Good evening everyone, and thank you for joining us tonight. We are here to help answer any questions you might have about the recent red tide events affecting our community and beaches."

"We hope we can dispel some myths and tell you how we are trying to develop ways to mitigate red tide events here in Southern Florida"



**Our Agenda**

"Our first speaker is:"  
Dr. Sally Harris from (Dept. FFWCC) Ecologist.

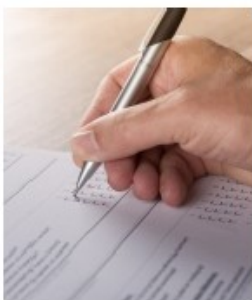
"She will tell us what they know about the red tides"  
"And answer some of the questions you all submitted"

"Then our next speaker will be Dr. Mateo Edwards a lead scientist at BioMarkerX to discuss solutions they are working on with the FFWCC to finding a way to determine how long they may persist, a health check of the bloom."

**Questionnaire and Survey Results**

"Thank you everyone that participated in the online survey, and your submitted questions"

"Here are a few of the most common questions"



**The Most Common Questions**

- What is the red tide?
- Is it living or just discolored water?
- Why does the red tide make people sick?
- Is it an invasive species?
- How long will these events last?
- And how will we know when they will go away?
- What is being done about it?

**Let's get your opinions**

Take our your clickers

### CQ 1: What is the Florida red tide?

- A. It's a bacteria that colonies the water and is the result of sewage outflow into the bays
- B. It's a dinoflagellate algae that lives normally at low concentrations but can get very dense
- C. Cyanobacteria like in Lake Erie caused by nutrient runoff from lawns and agriculture
- D. Unknown.

### CQ 2: What do you think causes the red tide events?

- A. Nutrient pollution from fertilizers and natural sources delivered from rivers may stimulate algae growth
- B. Overfishing had reduced predators that eat the algae, so there nothing to limit their growth
- C. Summer temperatures are elevated and make the environment more habitable to the algae, so that others can't survive.
- D. The algae is an invasive species, so it can outcompete all the native algae species
- E. No one knows why

### CQ 3: What are your biggest concerns?

How does it affect:

- A. Marine life (manatees, dolphins, sea turtles)
- B. Economic stability (how it influences tourist dollars)
- C. Human health
- D. Food Safety

### Florida red tide



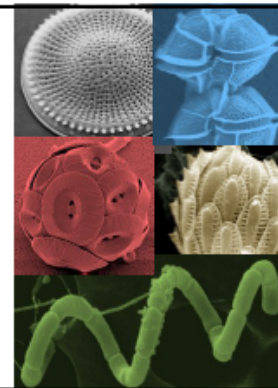
Dr. Sally Harris  
*Aquatic Ecologist at  
Florida Fish and Wildlife Commission*

"Hi everyone, I'm Dr. Sally Harris, you can call me Sally or Dr. Sally. I'm an ecologists who monitors the red tides here in Florida. And I'm here as an expert to talk you all about the ecology of red tides and what we know about our local red tide events.

"So I want to start off asking some of your questions to you all, and see what you know already?"

### But I though all algae were good?

- The Good – they form the base the of food web (primary producers).
- The Bad – too many algae/biomass can have negative effects to humans and wildlife
- The harmful – some algae produce toxins that harm marine life or humans





### Not all red tides are red



### Harmful Algae Blooms (HABs)

Red tides should really be called Harmful Algae Blooms

HABs can be any kind unicellular microscopic photosynthetic eukaryotes or cyanobacteria that causes water quality issues.

Nutrient inputs from artificial sources (fertilizers) and natural sources are thought to stimulate and sustain bloom events

But not all algae are created equal.

- Not all algae can use the same N and P sources
- Most HAB species are a group of algae called Dinoflagellates

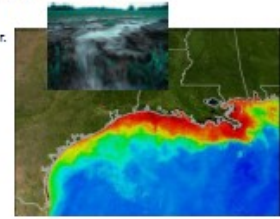
### Harmful Blooms can very large!



### Not all HABs make toxins, but can create "Dead Zones"

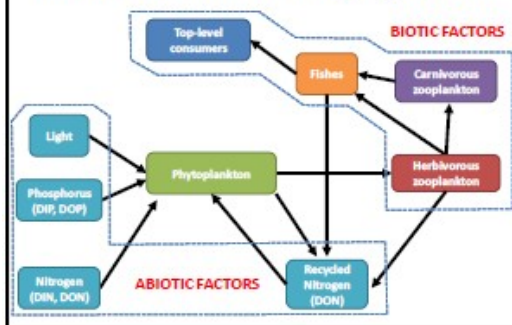
Nutrients from farmlands and cities pollute the Mississippi river, and pour out into the nutrient poor (oligotrophic) ocean were large algae blooms occur.

When they run out, cells stop growing, die or some go into a form of "resting."



Large deaths of algae deplete the ocean of oxygen and can lead to dead zones, where oxygen is depleted like near the mouth of the Mississippi River.

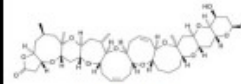
### What factors stimulate algae growth?



### The Florida red tide - *Karenia*

Dinoflagellate called *Karenia brevis*

*Karenia brevis* (Wilson's isolate, Florida)



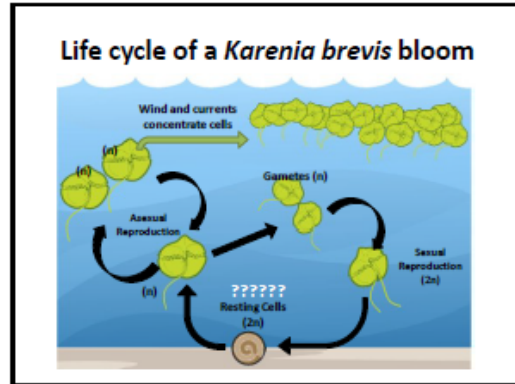
***Karenia brevis* is native to the Gulf of Mexico**

First reports perhaps of *Karenia* blooms date to early accounts in the 1500s by Spanish explorers

Blooms occur around Clearwater to Fort Myers that receive a dissolved nutrients (N and P) from local rivers.

Blooms can persist for ~ 1 year after it begins

Background concentrations  
0 to 1000 cells L<sup>-1</sup> and  
Up to > 1 million cells L<sup>-1</sup>



**Toxin Producer – Health Effects**

At high cell densities > 1 million per liter!

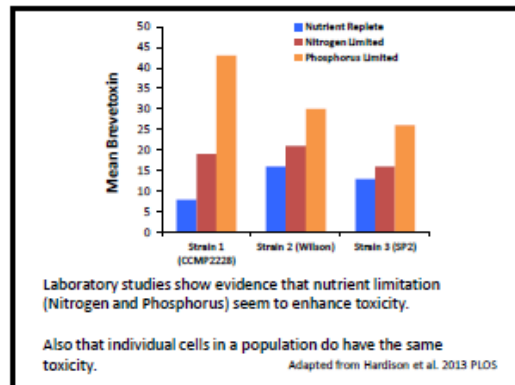
Can be toxic to marine life and cause respiratory distress in humans as the dying cells release toxins in to the air.

The toxin is a neurotoxin and targets affects the movement of sodium transport across membranes.

However, shellfish from the grocery store is not affected

BREVETOXIN-A

BREVETOXIN-B



**Can you summarize the main points of what Sally has told you?**

**Sally's summary list of main points:**

- Red tides should be called Harmful Algal Blooms
- HABs can have large effects
- Abiotic and Biotic factors can limit or promote growth
- More toxins are created when nutrient stressed
- *Karenia brevis* is native algae to the area

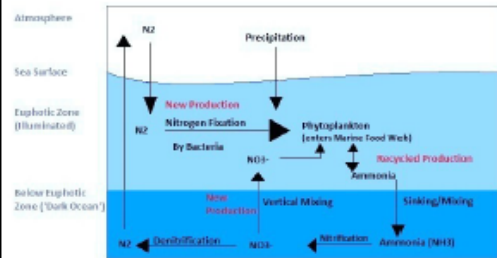
Sally finishes with:

"We haven't a direct link of nutrients to the abundance of *Karenia*, but we know that its important for algae growth, thus we need more information about what it does without nutrients"

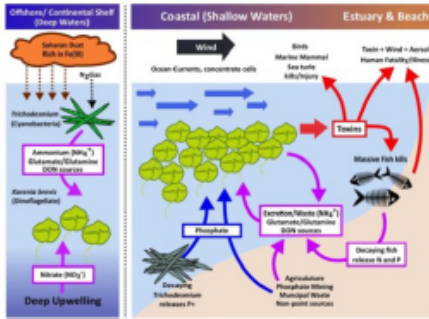
**Dr. Mateo Edwards**  
*Lead Researcher with BiomarkerX*

“Hi everyone, I’m with BioMarkerX. We are a startup company that is developing ways to track HAB events using molecular biology. We hope to develop a marker to determine what the nutrient status of the bloom is so we can predict if its actively growing, stable or declining by looking for nutrient related stress genes”

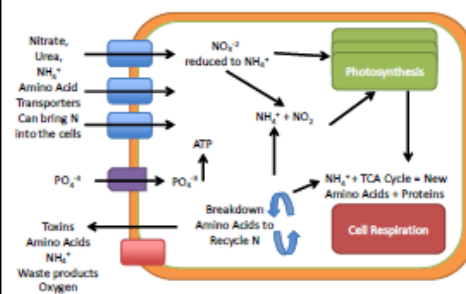
### The Nitrogen Cycle



### 50 Years of *Karenia* research. Its complicated.



### The Cell as a System



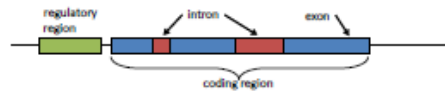
*Lets step outside of our case, briefly and lets review some common components of molecular biology*

### CQ 4: What is a gene?

- A. Genes are pieces of DNA that organisms use to directly make proteins.
- B. A gene is a fragment of DNA that is used to make messenger RNA
- C. Genes are the functional enzyme to do some biochemical activity
- D. Genes are made of amino acids, that get folded into proteins

### The gene is the fundamental unit of functional information

- Short segments of DNA on a chromosome(s)
- Eukaryotic Genes have Regulatory and Coding Regions
  - Coding Region made of introns and exons

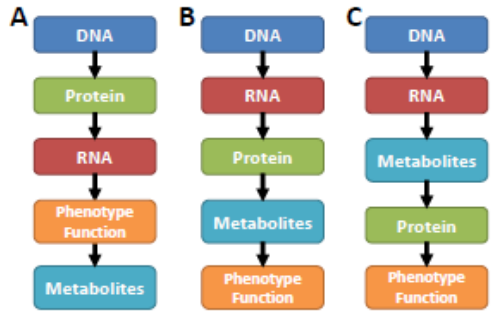


- Information from genes are used to make proteins
  - Transcription and Translation

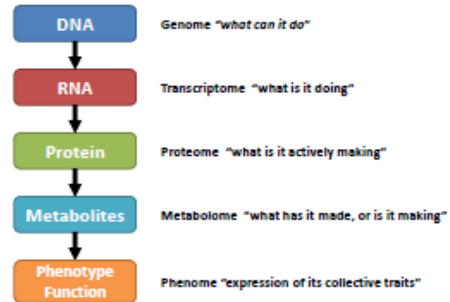
### Organize into a temporal sequence of events



### CQ 5: Choose which one matches yours



### Central Dogma of Molecular Biology



*Now Lets Return To The Case*

### What Do You Think?

- What do you think Dr. Edwards should investigate to understand the role of nutrients and Karenia's biology?
- Design what the experiment would look like
  - What would be the control?
  - What would be the experimental condition?
  - What would the prediction be?

## The *Karenia* Transcriptome

"We can't sequence the entire genome of *Karenia* yet, so instead we focus on its transcriptome"

"It represents "what is it doing now"

"The transcriptome is a snap shot of the all messenger RNA cells produce that might code for some functional enzymes, that maybe it uses to survive stressful conditions"

So by probing *Karenia* with stressors, we can identify markers (genes) for that type of stress compared to normal growing conditions.

## Transcription is Dynamic

"Just like a dimmer switch for a light bulb"

But maybe ON at low levels always.

"We want to look at transcription and see what genes are expressed higher or lower" Said Dr. Edwards

Turn Up for more light

Turn Down for less light



Dr. Edwards says,

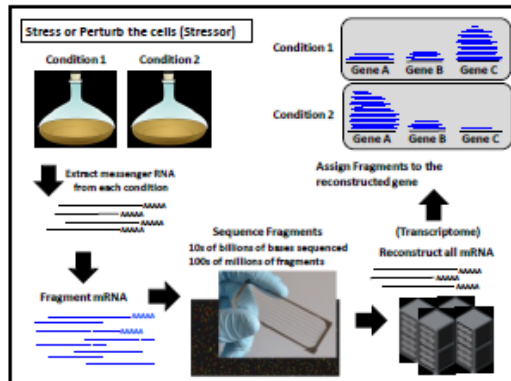
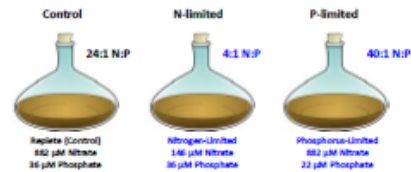
"When faced with limiting nutrients we think algae will do:

- Up-regulate transporters for nitrogen compounds or phosphate
- Up-regulate enzymes need to incorporate nitrogen, or recycle internal stores of nitrogen trapped as amino acids or proteins"

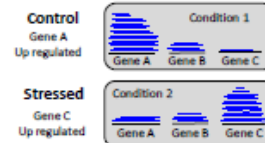
## 3 Different Treatment Conditions

*Karenia* was grown under:

- Control (Normal conditions)
- N-limited (Nitrate runs out as it grows)
- P-limiting conditions (Phosphorus runs out)

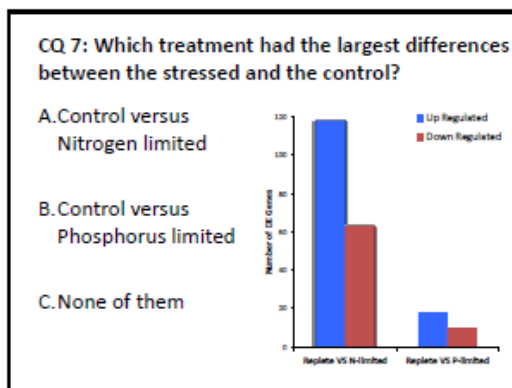
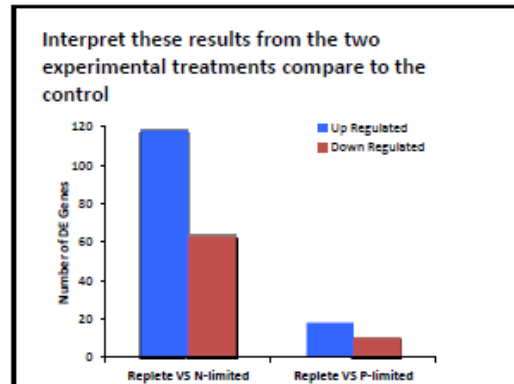
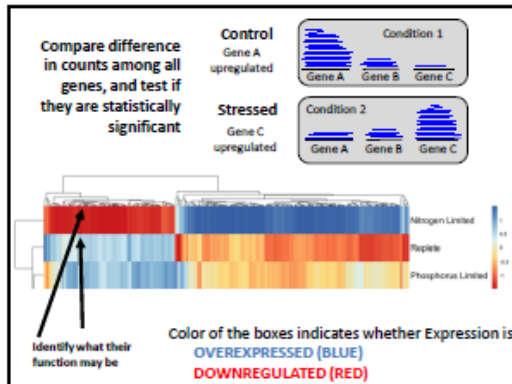


Compare difference in counts among all genes, and test if they are statistically significant



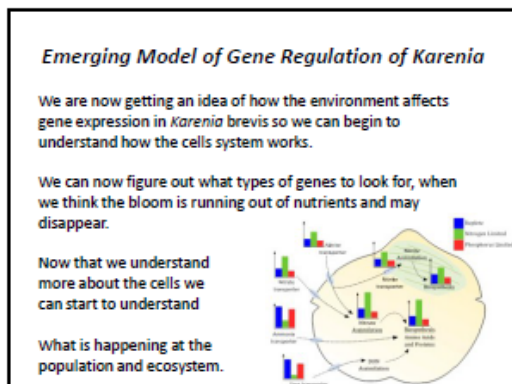
CQ 6: In comparisons gene expression in the Control versus Stressed condition:

- In the Control relative to stressed, all genes are up regulated
- In the Control relative to stressed, all genes are down regulated
- In the Control relative to stressed, Gene A is upregulated and while C is down
- In the Control relative to stressed, Gene B is unaffected while C is upregulated



CQ 8: So what can BiomarkerX conclude so far?

- All genes show similar patterns of expression to nutrient stress.
- Nutrient limitation doesn't have an effect on transcription in *Karenia*
- Nutrient limited does have an effect on transcription in *Karenia* but only N-limitation
- Nutrient limited does have an effect on transcription in *Karenia* but only P-limitation



If you were a someone attending this meeting have your questions been answered?

- What did you think they should do next

## Can You Make a Diagram:

- How does DNA connect to the Phenotype of an organism?
- When would you want to use different kinds of 'Omics technologies?
- Find what genes make up the nitrate assimilation pathway, and how they are connected (what are N and P needed for?)

## Image Credits

**Slide 1**  
**Image credit:**  
Description: Beach at Wulbert, Sanibel Island, Florida, looking north towards Captiva Island  
Source: Wikimedia Commons, [https://commons.wikimedia.org/wiki/File:Wulbert\\_Sanibel\\_LHS](https://commons.wikimedia.org/wiki/File:Wulbert_Sanibel_LHS)  
Author: Braden Grapton (Own work) [Public domain]  
Clearance: GNU Free Documentation License

**Slide 2**  
**Image credit:**  
Description: UF Keene-Ritt Classroom Desk Whiteboard Windows  
Source: Flickr, <https://www.flickr.com/photos/consensus/4423279587/>  
Author: ChrisGopher  
Clearance: Attribution 3.0 Generic (CC BY 3.0)  
Modification: Cropped and deaturated image

**Slide 4**  
**Image credit:**  
Description: Survey Opinion Research  
Source: Pixabay, <https://pixabay.com/en/survey-opinion-research-voting-68-104942/>  
Author: andrius  
Clearance: CC0 Public Domain, Free for commercial use, No attribution required

**Slide 5**  
**Image credit:**  
Description: Scanned image of standard 8x5 notecard / index card. Notecards rule.  
Source: Wikimedia Commons, <https://commons.wikimedia.org/wiki/File:Notecard.jpg>  
Author: [jay.annett@gmail.com](mailto:jay.annett@gmail.com) - [jay.annett@gmail.com](mailto:jay.annett@gmail.com)  
Clearance: Creative Commons Attribution 3.0 Unported

## Image Credits

**Slide 10**  
**Image credit:**  
Description: Red Tide off the west coast of the state of California. Released to the Public Domain, August 2007. Photo taken by the Light House Photo Project  
Source: Public Domain

**Slide 11**  
**Image credit:**  
Description: Orange blossom field in California  
Source: Wikimedia Commons, [https://commons.wikimedia.org/wiki/File:Orange\\_blossoms\\_in\\_California.jpg](https://commons.wikimedia.org/wiki/File:Orange_blossoms_in_California.jpg)  
Author: Michael Szymanski (Own work)  
Clearance: Attribution 3.0 Generic (CC BY 3.0)

**Image credit:**  
Description: Orange blossom field in California  
Source: Wikimedia Commons, [https://commons.wikimedia.org/wiki/File:Orange\\_blossoms\\_in\\_California.jpg](https://commons.wikimedia.org/wiki/File:Orange_blossoms_in_California.jpg)  
Author: Michael Szymanski (Own work)  
Clearance: Attribution 3.0 Generic (CC BY 3.0)

**Slide 14**  
**Image credit:**  
Description: Orange blossom field in California  
Source: Wikimedia Commons, [https://commons.wikimedia.org/wiki/File:Orange\\_blossoms\\_in\\_California.jpg](https://commons.wikimedia.org/wiki/File:Orange_blossoms_in_California.jpg)  
Author: Michael Szymanski (Own work)  
Clearance: Attribution 3.0 Generic (CC BY 3.0)

**Slide 15**  
**Image credit:**  
Description: Orange blossom field in California  
Source: Wikimedia Commons, [https://commons.wikimedia.org/wiki/File:Orange\\_blossoms\\_in\\_California.jpg](https://commons.wikimedia.org/wiki/File:Orange_blossoms_in_California.jpg)  
Author: Michael Szymanski (Own work)  
Clearance: Attribution 3.0 Generic (CC BY 3.0)

## Image Credits

**Slide 12**  
**Image credit:**  
Description: Red Tide off the west coast of the state of California. Released to the Public Domain, August 2007. Photo taken by the Light House Photo Project  
Source: Public Domain

**Image credit:**  
Description: Orange blossom field in California  
Source: Wikimedia Commons, [https://commons.wikimedia.org/wiki/File:Orange\\_blossoms\\_in\\_California.jpg](https://commons.wikimedia.org/wiki/File:Orange_blossoms_in_California.jpg)  
Author: Michael Szymanski (Own work)  
Clearance: Attribution 3.0 Generic (CC BY 3.0)

**Image credit:**  
Description: Orange blossom field in California  
Source: Wikimedia Commons, [https://commons.wikimedia.org/wiki/File:Orange\\_blossoms\\_in\\_California.jpg](https://commons.wikimedia.org/wiki/File:Orange_blossoms_in_California.jpg)  
Author: Michael Szymanski (Own work)  
Clearance: Attribution 3.0 Generic (CC BY 3.0)

**Image credit:**  
Description: Orange blossom field in California  
Source: Wikimedia Commons, [https://commons.wikimedia.org/wiki/File:Orange\\_blossoms\\_in\\_California.jpg](https://commons.wikimedia.org/wiki/File:Orange_blossoms_in_California.jpg)  
Author: Michael Szymanski (Own work)  
Clearance: Attribution 3.0 Generic (CC BY 3.0)

**Image credit:**  
Description: Orange blossom field in California  
Source: Wikimedia Commons, [https://commons.wikimedia.org/wiki/File:Orange\\_blossoms\\_in\\_California.jpg](https://commons.wikimedia.org/wiki/File:Orange_blossoms_in_California.jpg)  
Author: Michael Szymanski (Own work)  
Clearance: Attribution 3.0 Generic (CC BY 3.0)

## Image Credits

**Slide 15**  
**Image credit:**  
Description: The green scum shown in this image is the worst algae bloom lake (in fact has experienced in decades. Worst green (Blastoid) scum ever) from the northern shore.  
Image captured by the satellite. Data provided courtesy of the United States Geological Survey.  
Source: Wikimedia Commons, [https://commons.wikimedia.org/wiki/File:Satellite\\_Image\\_of\\_Alga\\_Bloom\\_in\\_Lake\\_Erie.jpg](https://commons.wikimedia.org/wiki/File:Satellite_Image_of_Alga_Bloom_in_Lake_Erie.jpg)  
Author: Isaac Allen and Robert Simcox (NASA Earth Observatory) [Public domain]  
Clearance: GNU Free Documentation License

**Slide 16**  
**Image credit:**  
Description: View of runoff transporting nonpoint source pollution, from a farm field in Iowa during a rainstorm. Caption: "Topsoil as well as farm fertilizers and other potential pollutants run off unprotected farm fields when heavy rains occur."  
Source: Wikimedia Commons, [https://commons.wikimedia.org/wiki/File:Runoff\\_of\\_soil\\_Nutrients\\_fertilizer.jpg](https://commons.wikimedia.org/wiki/File:Runoff_of_soil_Nutrients_fertilizer.jpg)  
Author: Lynn Betts, photographer [Public domain]  
Clearance: CC0 Public Domain

**Image credit:**  
Description: Satellite image and illustration of a dead zone in the southern U.S.  
Source: Wikimedia Commons, [https://commons.wikimedia.org/wiki/File:Dead\\_zone.jpg](https://commons.wikimedia.org/wiki/File:Dead_zone.jpg)  
Author: NOAA (NOAA)  
Clearance: CC0 Public Domain

## Image Credits

**Slide 17**  
**Image credit:**  
Description: A group of nucleotides isolated from the genome of *Escherichia coli* (Barnett, Condon, 1994)  
Image credit: [https://commons.wikimedia.org/wiki/File:DNA\\_nucleotides.jpg](https://commons.wikimedia.org/wiki/File:DNA_nucleotides.jpg)  
Author: Michael Szymanski (Own work)  
Clearance: Attribution 3.0 Generic (CC BY 3.0)

**Image credit:**  
Description: A group of nucleotides isolated from the genome of *Escherichia coli* (Barnett, Condon, 1994)  
Image credit: [https://commons.wikimedia.org/wiki/File:DNA\\_nucleotides.jpg](https://commons.wikimedia.org/wiki/File:DNA_nucleotides.jpg)  
Author: Michael Szymanski (Own work)  
Clearance: Attribution 3.0 Generic (CC BY 3.0)

**Image credit:**  
Description: A group of nucleotides isolated from the genome of *Escherichia coli* (Barnett, Condon, 1994)  
Image credit: [https://commons.wikimedia.org/wiki/File:DNA\\_nucleotides.jpg](https://commons.wikimedia.org/wiki/File:DNA_nucleotides.jpg)  
Author: Michael Szymanski (Own work)  
Clearance: Attribution 3.0 Generic (CC BY 3.0)

**Image credit:**  
Description: A group of nucleotides isolated from the genome of *Escherichia coli* (Barnett, Condon, 1994)  
Image credit: [https://commons.wikimedia.org/wiki/File:DNA\\_nucleotides.jpg](https://commons.wikimedia.org/wiki/File:DNA_nucleotides.jpg)  
Author: Michael Szymanski (Own work)  
Clearance: Attribution 3.0 Generic (CC BY 3.0)

