Marine microbial metabolomics: a journey through time, space, and metabolism

Angela Boysen

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

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Anitra Ingalls, Chair

Curtis Deutsch

E. Virginia Armbrust

Program Authorized to Offer Degree:
Oceanography

University of Washington

Abstract

Marine microbial metabolomics: a journey through time, space, and metabolism

Angela Boysen

Chair of the Supervisory Committee:
Professor Anitra Ingalls
Oceanography

Microorganisms control the flux of energy, stored as organic matter, into the ocean through the cumulative effects of individual metabolisms and community interactions. Metabolites are the currency of microbial metabolism, are carefully regulated to meet the metabolic demands of organisms living in dynamic environments, and reflect cellular status and metabolic strategies for nutrient acquisition, energy storage, redox maintenance, and more. This dissertation focuses on developing metabolomics techniques for the marine environment and using them to study microbial dynamics over time and space to identify compounds that are key microbial currencies. In order to study natural populations of marine microbes, I developed a method for targeted and untargeted metabolomics data acquisition and analysis with the unique challenges of marine samples in mind (Chapter 2). I use this method to study the influence of the diel cycle on the marine microbial community at Station ALOHA in the North Pacific Subtropical Gyre, ultimately showing synchrony of daytime anabolism and nighttime catabolism as seen through diel oscillations of ubiquitous metabolites including cofactors and vitamins. Through pairing metabolite and gene expression data, I demonstrate the strategies that specific photoautotrophs use to manage the daily fluctuations in solar energy (Chapter 3). To examine how microorganisms respond to other environmental forcings, I investigate the metabolism of microbial communities across the North Pacific Transition Zone and identify metabolic currencies used by those communities to adapt to varying nutrient supply

(Chapter 4). Nutrient amendment experiments show the dominance of nitrogen limitation throughout this region and the potential for iron-nitrogen co-limitation near the subtropical chlorophyll front. Finally, in order to explore the potential for metabolites to be nutrient sources to the microbial community, I investigate the ability of natural microbial communities to use the abundant osmolyte glycine betaine. I determine the kinetics of uptake and identify the metabolic uses of glycine betaine in two different natural microbial communities and show that its use as a nutrient differs depending on DIN availability (Chapter 5). In full, this dissertation provides some of the first metabolomics measurements in the natural marine environment and identifies and explores the roles of key organic molecules in shaping the microbial community structure and function.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This dissertation would not have happened without an incredible community supporting me.

The most credit, of course, goes to my advisor, Anitra Ingalls, who has been a role model of how to think broadly, critically, and creatively. She allowed me space to learn and grow as an independent researcher while also providing support and advice as I needed it. I will always be grateful for the timing of my graduate work in Anitra's lab. Joining just as the group transitioned to begin using high resolution mass spectrometry for metabolomics analyses allowed me to gain skills as an analytical chemist and data scientist as well as an oceanographer. Working with Anitra as we began generating metabolomics data has been a lesson in how to wade through unknown waters (or mass features) while maintaining a high standard of excellence in the work that we do. Plus it has all been a lot of fun.

The members of the Ingalls lab have been a joy and inspiration to work alongside. Laura Carlson is a lab superhero who taught me an enormous amount about lab work at sea and back at UW. She is a treat to work with and I can't imagine that any lab I work in down the line will be quite as fun without her in it. Thank you to Wei Qin, Jamie Collins, and Laura Hmelo for always providing excellent feedback and advice. Rachel Lundeen was not only a rigorous analytical chemist and critical thinker who gave great feedback but also an indispensable emotional support through the highs and lows of graduate school. Hannah Dawson, an honorary Ingalls lab member, is a kind and optimistic scientist and collaborating with her is a treat. It has been a pleasure to work with Natalie Kellogg, Will Kumler, and Josh Sacks, who have all brought great energy and new perspectives to the group. I have loved working alongside wonderful technicians, undergraduates and high schoolers in the lab,

including Regina Lionheart, Alec Meyers, Leland Wood, Lindsay Turner, Katie Hearther, Alexa Wied, and Mayank Motokuri. Most importantly I need to thank Katherine Heal, who has been a dear friend and mentor every step of the way. Regardless of what we do together, from writing papers and fixing code to cooking mujadara and biking through the woods, I do it better because I do it with you and have learned from you.

I am fortunate to have learned from many scientists at UW and beyond. Ginger Armbrust, Curtis Deutsch, and Dave Stahl always ask insightful questions that clarify my thinking. From Ginger I have learned how to make the most of rough seas and how to approach scientific questions with truely cross-disciplinary teams. Rick Keil, Randi Bundy, Jodi Young, Alex Gagnon, and Gabrielle Rocap all pushed my thinking and helped me grow as a scientist and teacher.

Bryn Durham has been a close collaborator who has shaped how I think about the microbial world. We shared lab space on all three major research cruises I participated in during my PhD, and she literally pulled her weight when helping me slog carboys of water. The Simons Collaboration on Ocean Processes and Ecology has provided me with a wonderful and fun community of peers and mentors, and I am grateful to have been a part. Daniel Muratore, Joshua Weitz, Frank Aylward, Ed DeLong, Matt Harke, Sonya Dyhrman, Ben Van Mooy, Sam Wilson, Angel White, and many other SCOPE scientists have been essential to my development as a scientific thinker.

The UW Oceanography graduate student community has been an indispensable group of friends and peers who have made my graduate studies infinitely more fun. Thanks to Megan Duffy, Marta Wolfshorndl, Jake Steinberg, and Justin Penn for the adventures and for showing me that there are many ways to do great science while also being a person with passions and pleasures outside of research. The women of chemical oceanography at UW are a phenomenal group of role-models. Thanks to Ashley Maloney, Hilary Palevsky, Nemiah Ladd, Nancy Williams, Elisa Bonnin, and JiWoon Park. Susanna Michael was a

joy of an office mate when she wasn't at PMEL or in France; having a cheerleader and someone to commiserate with made tough days more bearable. The rest of my cohort were also the very best people to go through grad school with. Thanks for all the puns, Theresa Whorley, Max Showalter, Robin McLachlan, Paige Logan, and Rachelle Lim. I am grateful to Leah Johnson, Ryan Groussman, Isaiah Bolden, Christine Baker, Sasha Seroy, Hally Stone, Elizabeth Brasseale, Anna McLaskey, Sarah Ragen, Katy Christensen, and too many others to thank, including friends from outside the Oceanography department who I had the pleasure of knowing because of the Program on Climate Change and the IGERT Program on Ocean Change.

Science should be for everyone and done by everyone who wants to participate. Though I have been welcomed into the scientific community, many people are not, and we need cultural and systemic change to combat the entrenched structures of racism and patriarchy. Thank you to Shirley Leung, Mikelle Nuwer, Kim Gottschalk, Rick Rupan, LuAnne Thompson, Robert Levine, Ethan Campbell, Colleen Hoffman, Claire McKinley, Rosalind Echols, Hannah Glover, and others for letting me learn from you and work with you to make our department become more inclusive and equitable.

The friendships I have found in Seattle and those I have from Boulder, Stanford, and Portland have been a source of joy and strength. Thank you to the rowing club, ICO volunteers, my swim and ski buddies, and everyone else who I've explored Seattle and the Pacific Northwest with. Thanks to the trivia team, Sediment Traphouse, who during the COVID-19 pandemic have been a much needed social community and a surprisingly necessary reminder of what day of the week it is. Thanks to all of my housemates at 321, past and present; I couldn't have dreamed of a better home in Seattle.

I am the person and scientist that I am because of the unwavering love and support from my parents and sister. Their confidence in me buoyed me up when I was uncertain. They have taught me that success is doing something that you care about, surrounding yourself

with people who you love, and always making time to enjoy the wonderful world around us. Eric Buescher has been critical in making sure that I do take time to relax and explore. His encouragement, patience, and love have been essential. Thanks for all the pastries.

Chapter 1
INTRODUCTION

As humans flood our atmosphere with carbon dioxide, the earth responds. The terrestrial biosphere is responsible for taking up approximately one third of the anthropogenically emitted carbon dioxide, and the ocean sequesters another third [1, 2]. The ocean's capacity for carbon sequestration depends on physical properties, such as temperature and the solubility of carbon dioxide in seawater, but it also depends critically on the microbial life within the ocean. Photosynthetic microorganisms, or phytoplankton, fix approximately 100 million tons of carbon every day [3], well over the biomass of all humans on earth ($\sim$60 million tons [4]). The amount of organic carbon that is produced by phytoplankton, the amount that is returned to carbon dioxide through respiration and the microbial loop, and the amount that is exported out of the surface ocean as part of the biological carbon pump are all influenced by what organisms are producing the organic matter and what kind organic matter they produce [5, 6, 7]. Understanding the dynamics of the microbial community responsible for producing and processing organic carbon in the surface ocean is thus critically important to our understanding of the marine carbon cycle and the earth climate system. One window we can use to gaze into the microbial world and better understand its players and processes is that of metabolomics, the study of all metabolites in a system. In this dissertation I use metabolomics to identify compounds that are key microbial currencies and show how these molecules connect with other measures of microbial diversity and activity.

Metabolites are the small (<800 dalton) organic molecules produced and used by all living cells during cellular metabolism. They are thus a snapshot or readout of the instantaneous metabolic state of an organism or community of organisms [8]. Metabolomics studies are complementary to other analyses using genomics, transcriptomics, and proteomics [9]. Genes

represent the possibilities of what an organism or community can do, transcripts represent the plan of what may happen, proteins represent what should happen, and metabolites reflect what is happening. Metabolites are involved in all cellular processes, from maintaining redox balance to synthesizing macromolecules to providing short term energy storage. Metabolites are also a primary form of communication and currency between microbes when exchanged through the dissolved labile organic pool [10, 11, 12, 13]. Individual metabolites and specific classes of metabolites have been measured in the ocean and in model marine microbes for many decades (for example [14, 15, 16, 17, 18, 19]). These studies have been possible because they rely on analyses specific to compounds with particular chemistries. The attempt to measure metabolites in a more comprehensive manner, using the tools of metabolomics, is newer to environmental science, delayed by the analytical challenges of measuring chemically diverse compounds in a quantitative way.

1.1 Challenges of marine metabolomics

Advances in analytical techniques in the last two decades have allowed for the rapid proliferation of metabolomics analyses, which are now commonly used in disease diagnosis, toxicology, and other medical sciences. Metabolomics primarily relies on two tools: nuclear magnetic resonance and mass spectrometry [20]. Coupling mass spectrometry with chromatography adds a dimension of time to the analyses, separating compounds based on polarity and allowing distinct detection of multiple metabolites with the same exact mass. In recent decades, liquid chromatography mass spectrometry (LCMS) has become particularly powerful due to its ability to separate a wide range of compounds with different sizes and polarities without chemical alteration (derivatization) necessary for gas chromatography. Electrospray ionization allows molecules to be ionized and detected in their intact form, again increasing the number of compounds within our analytical window.

Metabolomics studies typically take two forms, targeted and untargeted metabolomics. Targeted metabolomics focuses on quantifying metabolites that we know we are interested in *a priori* and that we can synthesize or purchase authentic standards of, so that their

identity can be fully confirmed. Untargeted metabolomics measures all molecules detectable by the method of choice, regardless of whether we know their identity or not. Untargeted metabolomics requires high-resolution mass spectrometry that can resolve the many thousands of potential metabolites in a given biological or environmental sample. Metabolites measured with untargeted methods can be identified using mass to charge ratios (m/z), retention time, and fragmentation patterns (MSMS spectra), however they cannot be absolutely quantified without an authentic standard to determine the instrument response.

Regardless of which metabolomics approach is being used, there are inherent challenges to LCMS-based metabolomics sample handling, data acquisition, and data processing. In particular, metabolomics analyses on environmental and marine samples are challenging because of unknown sample size, causing unpredictable matrix effects [21], and the high concentration of salt. In this dissertation I measure particulate metabolite concentrations and thus benefit from the ability of microorganisms to concentrate metabolites within their cells. Measuring the metabolites within that particulate matter is still no small feat, and the second chapter of this dissertation focuses on overcoming methodological challenges so that we can make high quality measurements of known and unknown metabolites [22]. With this tool in hand, we can then investigate the metabolisms of microbes in culture, in incubation experiments, and in the natural marine environment.

1.2 Diel cycles of microbial activity in the stratified surface ocean

Daily cycles of light and dark are a natural perturbation that organisms and communities have adapted to over evolutionary time, making these cycles an ideal framework for probing natural ecosystem dynamics. Diel cycles of marine microbial activity have been studied in the environment by measuring gene expression [23, 24, 25, 26], cell division [27], gross primary production, and net community production [28]. These studies suggest that diverse organisms partition their metabolism strategically over the diel cycle, however the chemicals that underlie those biochemical processes have not been measured *in situ*. Metabolites with diel oscillations provide clues into how organisms succeed in the environment by adapting

their metabolism to the oscillating energy availability.

In Chapter 3, I study diel cycles of microbial activity near the long term study site Station ALOHA (22 °45'N, 158 °W) using targeted metabolomics. Station ALOHA is a highly stratified environment, ideal for studying diel cycles of globally important organisms such as *Prochlorococcus*; a diverse suite of picoeukaryotes; the nitrogen fixing organisms *Crocosphaera*, UCYN-A, and diatom-diazotroph associations; and many diverse heterotrophic bacteria [29]. Interpreting metabolite data alongside gene expression data and other measures of microbial activity such as cell counts and biomass allows insights into organism-specific and whole-community activities that occur with diel rhythms [30, 31, 32]

1.3 Nutrient limitation across the North Pacific transition zone

Nutrient limitation and co-limitation are fundamental controls on marine productivity and marine microbial community structure [33]. There is a natural gradient of temperature, nutrients, and microbial community structure in the transition between the North Pacific subtropical gyre, a warm oligotrophic, nitrogen (N) limited regime, and the North Pacific subpolar gyre, a cold, iron (Fe) limited regime [34]. Across this gradient, chlorophyll concentrations, dominant primary producers, and net community productivity all change[34]. The ratio of Fe:N has been shown to be a key controlling factor of community diversity and productivity at ocean gyre boundaries [34, 35]. The North Pacific transition zone region provides a testing ground to determine how adaptation to particular environmental conditions influences organisms' metabolic responses to changes in nutrient availability.

In Chapter 4, I conduct nutrient incubation experiments along the natural gradient of the North Pacific transition zone. Measuring metabolites, chlorophyll, nutrients, and community composition allows me to show that organisms are dominantly nitrogen limited across this region, but that specific organisms respond to nutrient amendments differently. I hypothesize that several classes of metabolites, including mycosporine-like amino acids and carnities, are indicators of nitrogen-stress alleviation in specific organisms. These metabolites were putatively identified using untargeted metabolomics and are key targets for future investi-

gations as important players in microbial physiological adaptation and potential currencies of microbial interaction.

1.4 Dissolved metabolites as resources for the microbial community

The metabolites that accumulate inside cells unavoidably become part of the dissolved organic matter pool due to active excretion or cell death. These compounds then become part of the rapidly cycling labile dissolved organic matter pool and can provide resources such as carbon, nitrogen, phosphorus, reduced sulfur, or energy to organisms that can take up and metabolize them [36, 37, 38]. Though the standing stock of these dissolved metabolites is likely low, the flux may be high [39, 40]. The compounds that are available in this pool depends on the microbial community that is present [41]. The diversity of dissolved organic species likely contributes to the diversity of marine microbes, presenting more potential nutrient and energy niche spaces in which to specialize than the inorganic nutrients traditionally thought of as the drivers of community structure [40]. Both autotrophic and heterotrophic organisms have the capacity to take up small organic molecules [42, 43, 40]. This capacity can provide an energetic and nutritional advantage by reducing the need to synthesize compounds and use valuable nutrients.

In order to explore these processes, in Chapter 5, I examine the influence of microbial community structure and environmental conditions on the uptake kinetics and metabolic use of the common osmolyte glycine betaine. Cyanobacteria, heterotrophic bacteria, and eukaryotic algae all have a variety of abilities to synthesize and metabolize GBT [41, 44]. Some microbes, including the ubiquitous SAR11 clade, can produce energy by oxidizing the methyl groups from GBT to CO_2 [45], and other marine organisms have the ability to convert it to glycine and use it as a source of carbon and nitrogen or protein synthesis [39, 46]. I show that the uptake kinetics of glycine betaine were different in two different microbial communities in the North Pacific transition zone. Using stable-isotope-tracing metabolomics, I find that these communities retained different amounts of untransformed GBT and had different primary pathways of GBT metabolism. Ultimately, the uptake and

metabolic fate of this labile dissolved organic nitrogen compound depended on the microbial community composition and the availability of inorganic nitrogen. This work emphasizes that microbial loop processes that ultimately control C and N remineralization are nuanced and spatially heterogeneous.

In sum, I present a dissertation focused on making high quality measurements of small molecules in marine microorganisms in order open a new window into the marine carbon cycle. Small scale processes happening within microbial cells ultimately influence global scale biogeochemical processes. The observations I have made are part of a small but rapidly growing collection of measurements of polar metabolites in natural marine environments, and each of these measurements allows another inference into the microbial world.

Chapter 2

BEST-MATCHED INTERNAL STANDARD NORMALIZATION IN LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY METABOLOMICS APPLIED TO ENVIRONMENTAL SAMPLES.

2.1 Abstract

[1]The goal of metabolomics is to measure the entire range of small organic molecules in biological samples. In liquid chromatography-mass spectrometry-based metabolomics, formidable analytical challenges remain in removing the non-biological factors that affect chromatographic peak areas. These factors include sample matrix-induced ion suppression, chromatographic quality, and analytical drift. The combination of these factors is referred to as *obscuring variation*. Some metabolomics samples can exhibit intense obscuring variation due to matrix-induced ion suppression, rendering large amounts of data unreliable and difficult to interpret. Existing normalization techniques have limited applicability to these sample types. Here we present a data normalization method to minimize the effects of obscuring variation. We normalize peak areas using a batch-specific normalization process, which matches measured metabolites with isotope-labeled internal standards that behave similarly during the analysis. This method, called best-matched internal standard (B-MIS) normalization, can be applied to targeted or untargeted metabolomics datasets and yields relative concentrations. We evaluate and demonstrate the utility of B-MIS normalization using marine environmental samples and laboratory grown cultures of phytoplankton. In untargeted analyses, B-MIS normalization allowed for inclusion of mass features in down-

[1]A version of this chapter has been published as A.K. Boysen, K.R. Heal, L.T. Carlson, and A.E. Ingalls. Best-Matched Internal Standard Normalization in Liquid Chromatography-Mass Spectrometry Metabolomics Applied to Environmental Samples. *Analytical Chemistry*, 90(2), 2018.

stream analyses that would have been considered unreliable without normalization due to obscuring variation. B-MIS normalization for targeted or untargeted metabolomics is freely available at https://github.com/IngallsLabUW/B-MIS-normalization.

2.2 Introduction

Liquid chromatography-mass spectrometry (LC-MS)-based metabolomics has emerged as an important scientific tool over the last decade, with the potential to identify and quantify thousands of compounds resulting from cellular activity [47, 48, 8]. Quantification relies on an analyte's concentration being proportional to its peak area. However, independent of analyte concentration, peak areas can be affected by sample-matrix-induced ion suppression, injection volume, chromatographic quality, or analytical drift over time; the overall effects of these non-biological factors are referred to here as *obscuring variation* since they may mask biological differences of interest [49]. Performing metabolomics analyses on environmental samples is especially challenging because complex matrices such as natural seawater or culture media introduce variability in ionization efficiency, but there are no universally accepted normalization techniques for minimizing obscuring variation in environmental samples. Here we present a normalization method that addresses this outstanding challenge in LC-MS-based metabolomics.

In biomedical metabolomics, there are several approaches to minimize obscuring variation that rely on pre- or post-acquisition normalization. However, application of these approaches to environmental metabolomics is limited, and little attention has been given to removing obscuring variation in environmental samples[50, 51, 52, 53, 8, 54, 55, 56, 57]. Pre-acquisition normalization approaches ensure that samples are at similar concentrations before analysis to minimize the effect of matrix-induced ion suppression. Post-acquisition normalization techniques fall into two general categories, those that use a quality control sample injected several times over the course of a run to correct for analytical drift, and those that employ isotope-labeled internal standards.

Analytical drift over time can be normalized post-acquisition by applying a quality

control-based locally weighted scatterplot smoothing (LOESS) [58, 59] or vector regression [60]; these approaches attempt to minimize one aspect of obscuring variation (analytical drift) and assume that the sample-matrix induced ion suppression is consistent between samples. This approach may be appropriate in biomedical studies when pre-normalization can be carefully applied (for instance, by adjusting sample concentration or injection volume so that there is a consistent amount of creatinine in each sample, as done routinely in urine metabolomics [61, 62]). For complex matrices, pre-normalization is often impossible due to the variability in matrix among samples or lack of information about sample size. For instance, if cell density is not known until after sample collection, extracting a consistent amount of biomass is not possible. Furthermore, these approaches require large sample sets and tens of injections of a pooled sample as a quality control, which is not always possible in metabolomics studies with limited sample size and few samples to compare.

Other post-acquisition strategies use isotope-labeled internal standards to minimize obscuring variation. Some studies normalize each metabolite peak area to a single isotope-labeled internal standard [63]. This approach is very sensitive to the internal standard used for normalization and assumes that all compounds experience the same obscuring variation as the internal standard; these caveats make this approach limited in utility. Other studies use multiple internal standards, applying each to analytes that elute within a prescribed retention time window [64, 65]. Although this may be an improvement over a single internal standard, these approaches do not consider the fact that obscuring variation is not fully described by retention time. Our group has previously used a hybrid approach that matches isotope-labeled internal standards to analytes for which commercially available isotope-labeled internal standards are unavailable [66, 67], but this has not been applied to full-scale metabolomics analyses.

In laboratory studies, growing an organism on a ^{13}C-labeled substrate can yield a valuable ^{13}C-labeled stock used to calculate accurate concentrations of both targeted and untargeted metabolites and aid in metabolite identification,[68, 54] but this approach cannot be applied to studies of mixed communities in the environment. More sophisticated approaches attempt

to model the variability of each peak by a combination of internal standards [49, 69]. These methods require tens of injections of a pooled sample to obtain sufficient data for model training and make use of logarithmic transformations, which may limit quantitative comparison and hinder intuitive interpretation. To our knowledge, these methods have not been applied to targeted metabolomic analyses.

Given the paucity of normalization approaches that are appropriate for minimizing obscuring variation in samples with a high degree of complexity in both analytes and matrices, we present a method called best-matched internal standard (B-MIS) normalization. This method has the following advantages over existing techniques: 1) B-MIS normalization can be applied to targeted and untargeted metabolomics studies with sample sets of any size or nature. 2) B-MIS normalization can be customized to accommodate any number and variety of internal standards of the user's choice. 3) B-MIS normalization does not assume a uniform matrix effect among samples and minimizes obscuring variation induced by sample-specific matrix-induced ionization. 4) B-MIS normalization is independent of and can be paired with other sample normalization pre- or post-acquisition. 5) B-MIS normalization is simple to apply, intuitive to interpret, and open source.

To demonstrate both the need for this method and its functionality, we used two sample sets: marine environmental samples and laboratory cultures of marine phytoplankton. We analyzed these samples on two mass spectrometers: a Waters Xevo TQ-S triple quadrupole, for targeted analyses, and a Thermo QExactive HF, for untargeted analyses, using both reversed phase (RP) and hydrophilic interaction liquid chromatography (HILIC). A comparative analysis of these platforms revealed the strengths of each and highlighted the need for careful normalization and application of B-MIS normalization regardless of platform.

2.3 Experimental Section

The workflow for sample analysis and data processing is shown in Figure 2.1. Additional details on sample preparation, chromatography, mass spectrometry, and quality control parameters are described in supplemental methods.

2.3.1 Sample Preparation

We evaluated two sets of complex samples in triplicate: cultures of one species of diatom grown under four experimental conditions (salinity of 32 g/kg at both -1°C and 4°C, salinity of 41 g/kg at both -1°C and 4°C), and a natural marine microbial community at four depths in the North Pacific subtropical gyre (24° 33.284' N, 156° 19.790' W; samples collected at 15, 45, 75, and 125 m, all with salinities of 35 g/kg). Polar and non-polar metabolites were extracted using a modified Bligh-Dyer extraction [70, 71] using 1:1 methanol:water (aqueous phase) and dichloromethane (organic phase). Methodological blanks were extracted and analyzed along with each sample set. Isotope-labeled internal standards were added either before or after the extraction to all samples, blanks, and pooled samples (Table 2.1). To evaluate the effect of obscuring variation due to different matrix strengths and analytical drift, pooled samples were run at both full and half concentration (diluted with water) at least three times throughout a sample set.

2.3.2 Liquid Chromatography

As in previous metabolomic studies [48], we paired RP with HILIC (with separate injections) to maximize the number of compounds we were able to detect since many compounds that eluted early on the RP column were better retained and provided higher quality peaks on the HILIC column (Figures 2.6 and 2.7). All chromatographic separations were carried out on a Waters Acquity I-Class UPLC (Waters Corporation, Milford, MA). For targeted analysis, we monitored 105 analytes with RP and 126 analytes with HILIC, according to which column demonstrated better retention; 11 analytes were monitored on both columns to compare the performance of RP and HILIC chromatography in complex matrices (Table 2.2).

2.3.3 Mass Spectrometry

For targeted metabolomics, we used a Waters Xevo TQ-S triple quadrupole (TQS) with electrospray ionization (ESI) in selected reaction monitoring mode (SRM) with polarity

switching. SRM conditions for each compound (collision energy, cone voltage, precursor and product ions, Table 2.2) were optimized by infusion of each metabolite standard. For most metabolites, two SRM transitions were selected based on maximum peak areas.

We used a Thermo QExactive HF (QE) with ESI for untargeted analyses. For HILIC, a full scan method with polarity switching was used. For RP, positive ionization mode was chosen over positive/negative switching because few of the compounds in our targeted method ionized in the negative mode and using a single mode enabled higher resolution.

2.3.4 Data Processing

For targeted analysis, we integrated peaks using Skyline for small molecules [72]. After integration, we passed the data through an in-house quality control (QC) filter to ensure proper metabolite identification as described in the supplemental methods.

For untargeted metabolomics data, we converted Thermo .RAW files to .mzxml using MSConvert [73], and processed each data set using XCMS [74, 75, 76] (using parameters obtained via an Isotopologue Parameter Optimization [77]), yielding retention-time corrected mass features. We processed both datasets in separate XCMS runs according to injection and polarity (RP-aqueous, RP-organic, HILIC-aqueous (positive polarity), and HILIC-aqueous (negative polarity)). Data were quality controlled, as described in the supplemental methods, prior to normalization.

2.3.5 B-MIS Normalization

The best normalization for each analyte was determined using repeat injections of a quality control sample to search for an isotope-labeled internal standard whose obscuring variation matched the observed obscuring variation of the analyte, as shown in Figure 2.1. For our sample sets, multiple injections of the pooled sample at full and half strength were used in order to capture the obscuring variation due to variable matrix strength. After correcting for dilution, if the relative standard deviation (RSD) of an analyte's peak area in these repeat injections was less than 10%, the raw data were used. Otherwise, each analyte-internal

standard pair was evaluated. An internal standard was deemed an acceptable match for an analyte if the analyte's RSD improved by 40% over the raw RSD, as discussed in the following section. If multiple internal standards were acceptable for normalization of a single analyte, the one which minimized the RSD was selected. When a B-MIS was chosen, the peak area was divided by the ratio of the peak area of the B-MIS to the average peak area of the B-MIS across the whole sample set—resulting in an *adjusted* peak area for the analyte. This adjusted peak area served as a relative quantification across samples in the sample set. If the goal of an analysis is absolute quantification and standards are available, analytes can be quantified via standard addition in a subset of samples and then applied across the sample set using the relative quantification of the B-MIS adjusted peaks.

For internal standards with only one or two labeled atoms in the compound, we examined the methodological blank to ensure the signal was comparable to that in the samples, indicating that there was not significant contribution from naturally occurring isotopologues. For future users, we suggest choosing isotope-labeled internal standards with at least a +2 amu label to avoid any naturally occurring isotopologue contamination. In the sample sets analyzed here, the isotope-labeled internal standards were significantly more abundant that naturally occurring isotopologues, so contamination was negligible.

2.4 Results and Discussion

2.4.1 Evaluation of data quality and instrument performance

As in other studies [78, 79, 80], we found the HILIC column produced more variable chromatography than the RP column (higher median RSD of peak area, Figure 2.2A, see the example of choline in Figure 2.6, full results in Table 2.3). Despite lower chromatographic reproducibility, compounds analyzed on the HILIC column generally exhibited less ion suppression than those analyzed on the RP column. RP chromatography displayed a bimodal distribution of response factor ratios (Figure 2.2B), demonstrating significant ion suppression for many analytes (response factor ratio < 1). Ion suppression was especially dramatic

for polar compounds that elute early in RP but were retained on the HILIC column (Figures 2.6 and 2.7). The range of ion suppression and RSD for a given LC-MS configuration indicated that careful selection of a normalization technique was necessary to account for obscuring variation. This is particularly important when comparing samples in complex matrices such as the environmental matrix highlighted in Figure 2.2. Furthermore, the wide ranges in response factor ratios and RSD of peak areas for different analytes in the same matrix demonstrated that normalizing all compounds to a single internal standard [65, 63] would likely introduce more obscuring variation than it removed.

2.4.2 Evaluating B-MIS Normalization

Most environmental metabolomics analyses do not attempt to remove variability introduced during analysis [50, 51, 81, 53]. B-MIS normalization was designed to minimize obscuring variation in order to allow robust comparisons within and between sample sets. We evaluate B-MIS normalization by a) demonstrating that lowering the RSD of multiple injections of a pooled sample improves data precision and arriving at a cut-off value at which to apply normalization, b) comparing the results from B-MIS normalization with the 'gold standard' of correcting with an isotopologue for each analyte of interest, and c) assessing the selection of internal standards used for normalization.

To demonstrate the fundamental principle and evaluate the effectiveness of B-MIS normalization, we applied the technique to the internal standards themselves. In other words, we evaluated how well the internal standards' variabilities were matched with each of the other internal standards. Figure 2.3 is an example of this application, where each point is the resulting RSD of D_3-alanine in replicate injections of the pooled sample (x-axis) and across all samples in the environmental sample set (y-axis) after normalization to each possible internal standard; since we add the same amount of the internal standard to each sample, both of these values should be near zero. Figure 2.3 highlights that internal standards that reduced RSD of the pooled generally resulted in a reduced RSD across samples (towards zero in both the x- and y-axes). However, in this example case there were two instances where

a reduction in RSD of pooled sample injections increased the RSD of the samples (marked on Figure 2.3 with an asterisk); this introduction of variability can be avoided by applying normalization only if the pooled RSD was improved beyond a cut-off value.

Our goal was to maximize the number of compounds that can be improved by B-MIS normalization while minimizing the likelihood of introducing error. After assessing all internal standards, we determined that requiring a 40% improvement of pooled RSD (that is, (RSD_{final} - RSD_{raw}) / RSD_{raw}) avoided the majority of situations where using an internal standard would result in an increase of sample variability (Figures 2.3 and 2.8, Table 2.6). Therefore, we chose this cut-off as the criteria to apply B-MIS normalization (see Figure 2.1). Additionally, we do not normalize any compounds or mass features which had a raw pooled RSD of below 10% because the chances of significant improvement are low. Before applying B-MIS normalization, we suggest users perform a similar sensitivity analysis to determine an appropriate cut-off on a representative sample set given the internal standard suite they plan on using; tools for this analysis are available with B-MIS.

The gold standard for absolute quantitation is an isotopologue of each analyte, but isotope-labeled internal standards are costly, consume duty time on the MS (in targeted analyses), and are not commercially available for many compounds. In our targeted analyses, analytes with corresponding isotope-labeled internal standards were normalized to those internal standards (see Figure 2.1). If this was not prescribed, the algorithm occasionally selected a different internal standard, though the lowered RSD was often similar to that of the prescribed B-MIS. For example, in the environmental dataset, the amino acid valine could have selected the D_7–proline internal standard based on the RSD reduction (raw peak area RSD was 30.8%, which was minimized to 3.5% with normalization to D_7–proline). However, the 'gold standard' internal standard, D_8–valine, reduced the RSD to a similar 4.8%, indicating proline and valine experienced nearly identical obscuring variation in this matrix. For the other analytes that have an isotope-labeled internal standard, the majority included their corresponding internal standard as an acceptable BMIS (>60% in both data sets, Table 2.5). Most cases where the corresponding internal standard was not considered

acceptable were due to the stringency of the 40% improvement cut-off. Nearly all of the analytes with an isotope-labeled internal standard were less variable after normalization to their corresponding internal standard, with one exception in each dataset (pyridoxal in the environmental dataset, cysteic acid in the culture dataset; these exceptions were both improved by normalization in the other datasets).

The success of B-MIS normalization relies heavily on having an appropriate suite of internal standards. However, if a mass feature does not match well with any internal standard, B-MIS normalization defaults to using the raw data, which a user can choose to include in subsequent analysis. Increasing the number and chemical variety of internal standards results in fewer features whose obscuring variation are not improved by B-MIS normalization (Figure 2.9). In our two data sets, the B-MIS for the targeted analytes were typically close in retention time (in RP, Figure 2.10) or similar in chemistry (in HILIC). For instance, amino acids generally matched with an isotope-labeled amino acid B-MIS while sulfonates generally matched with a sulfonate B-MIS (Table 2.5). Figure 2.9 shows that the internal standards used in HILIC chromatography covered a wide enough range of chemical diversity to effectively reduce obscuring variation in the majority of analytes detected under these conditions.

The specific choice of internal standards used in B-MIS normalization is not prescribed and can vary from user to user; we suggest a selection of internal standards that spans retention time, m/z, and functional groups of interest. Before analyzing a full sample set, users should analyze how many of the mass features (in an untargeted analysis) are "acceptable" and "best" matches with each internal standard by running a pooled or representative sample at different dilutions. This will reveal which, if any, internal standards are matching with a large number of mass features and may indicate that more internal standards with similar chemical diversity could be beneficial (see Table 2.4). For example, in our untargeted diatom culture data $^{13}C_3$–Vitamin B_1 (thiamine) was selected as acceptable for nearly 10% of mass features detected in the aqueous fraction analyzed with RP chromatography, and it was selected as the B-MIS for all of these features. It is likely that normalization of these

compounds could improve if we included more internal standards similar to $^{13}C_3$–Vitamin B_1. Some internal standards were infrequently selected as appropriate (such as ^{15}N–isoleucine), while others were appropriate for many and only the best for a few (such as D_7–proline), indicating sufficient coverage to effectively normalize compounds that experienced similar obscuring variation.

2.4.3 Application of B-MIS Normalization

B-MIS normalization only adjusts areas of metabolite peaks with high obscuring variation if normalization improves the technical reproducibility above a specified cut-off (in our case, a 40% improvement over the raw technical reproducibility), therefore a subset of analytes are not normalized. Figure 2.4 shows an example of two analytes (taurine and tyrosine) detected in the pooled diatom-culture sample whose RSD were improved by B-MIS normalization. For the sample sets analyzed here, 54% (diatom samples) and 79% (environmental samples) of analytes in the targeted data selected an internal standard for normalization (Table 2.5, Figure 2.10).

For each compound, the RSD of the pooled injections depends on the sample matrix, though the sample sets showed a similar range of RSD after B-MIS normalization (Figure 2.11). Of the 25 targeted analytes that were measured in both sample sets and that found acceptable internal standards for normalization, 20 shared acceptable internal standards, while five analytes did not (Table 2.5, Figure 2.11). The compounds which did not improve with B-MIS normalization were not always consistent from environmental to culture sample sets. For example, there were 18 compounds that did not pick a B-MIS in the targeted environmental sample set. In the culture sample set, 6 of these compounds selected a B-MIS, 4 did not select a B-MIS, and the remaining 8 were not detected (Table 2.5). These differences between sample sets demonstrate that while fundamental chemistry has some predictive capacity to describe obscuring variation, the complexity of the sample matrix makes definitive prediction unrealistic.

In untargeted metabolomics, it is common to remove any mass features that do not ex-

hibit reliable data [82]. One way this is achieved is by calculating the RSD of each mass feature in a multiply-injected pooled sample and removing features that have a RSD greater than 20% [82, 58]. In the case of our untargeted data, B-MIS normalization increased the number of mass features that passed this RSD filter from 3,836 to 5,028 (in the diatom samples) and from 5,568 to 9,372 (in the environmental samples), preserving many high-quality peaks that would have been otherwise discarded (Figure 2.12). In our environmental untargeted analysis, 54% and 53% of the mass features detected on the HILIC and RP columns, respectively, matched with an acceptable internal standard and showed decreased variability with a B-MIS (Figure 2.9). In this dataset, a third of the most abundant and significantly different mass features would have been excluded if this RSD filtering had been performed prior to B-MIS normalization (Figure 2.5). Seven of the ten mass features most enriched in the sample collected at 125 m compared to the sample collected at 15 m would have been excluded without B-MIS normalization before filtering (Figure 2.5). This comparison highlights that meaningful biological interpretation of the differences between samples hinges on proper normalization. We therefore recommend applying B-MIS normalization before filtering out mass features based on RSD.

2.5 Conclusions

Published environmental metabolomics analyses have not attempted to minimize the obscuring variation inherent to LC-MS-based data acquisition. In both targeted and untargeted environmental samples, we demonstrate the need for careful normalization due to matrix-induced ion suppression and analytical drift. This manuscript introduces, evaluates, and applies a batch-specific, simple, and customizable normalization process: Best-Matched Internal Standard (B-MIS) normalization. By applying B-MIS normalization, we are able to retain many more LC-MS mass features during untargeted RSD filtering. For future users, we have made the tools to customize, apply, and evaluate B-MIS normalization freely available at https://github.com/IngallsLabUW/B-MIS-normalization.

2.6 Acknowledgement

The authors acknowledge L. Hmelo, J. Collins, R. Lundeen, R. Lionheart, A. Weid, and N. Kellogg for assistance with lab,data analysis; J. Young and H. Dawson for providing the diatom cultures; B. Durham for providing several standards, or manuscript editing; the crew and scientific party of the R/V Kilo Moana during HOE-Legacy 2A. Thanks to four anonymous reviewers whose comments greatly improved this manuscript. This work was supported by grants from the Simons Foundation (LS Award ID: 385428, A.E.I.; SCOPE Award ID 329108, A.E.I.), NSF OCE-1228770 and OCE-1205232 to A.E.I., NSF GRFP to K.R.H., NSF IGERT Program on Ocean Change to A.K.B.

2.7 Figures

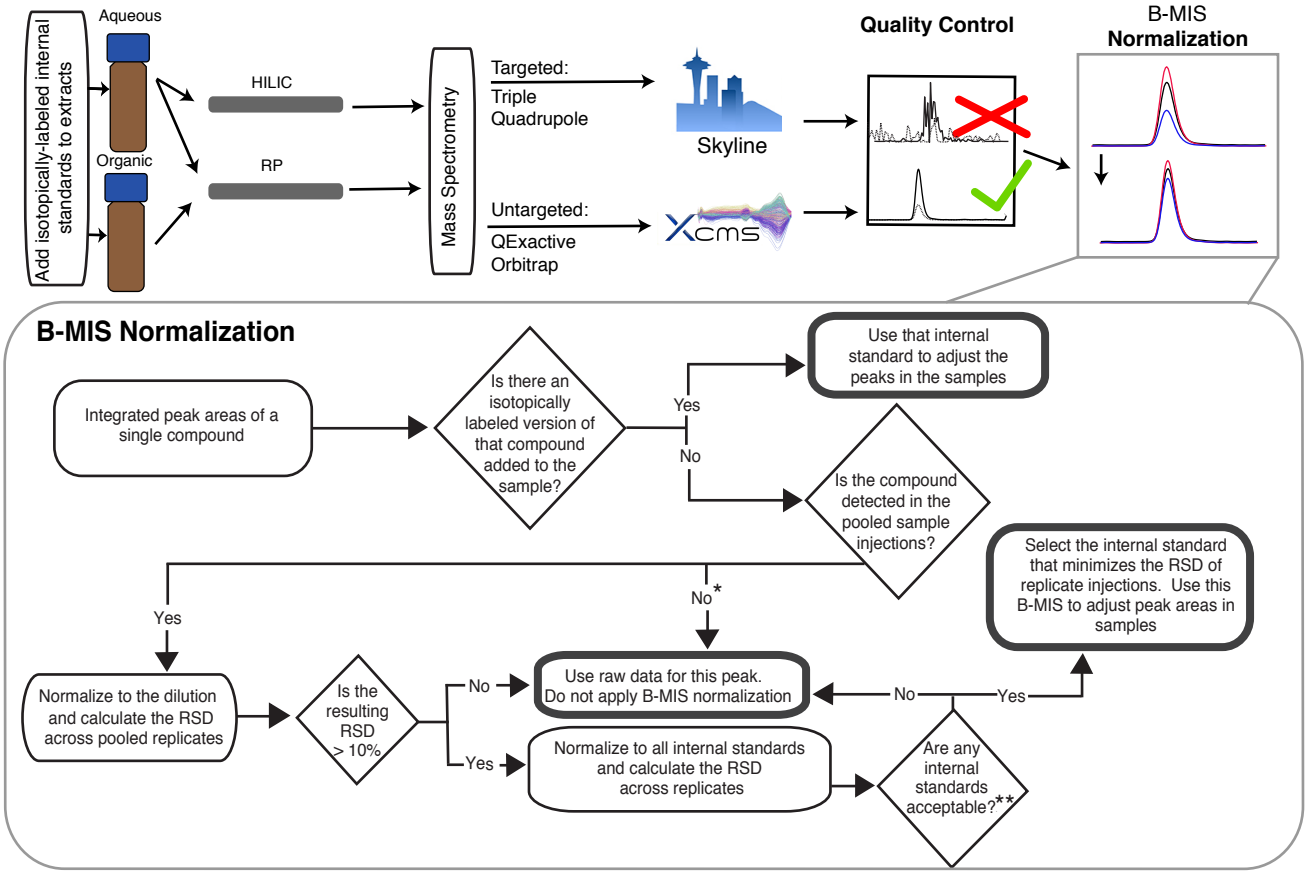

Figure 2.1: Overview of data acquisition and processing. After extraction, aqueous fractions were run on both RP and HILIC, while organic fractions were run only on RP. These fractions were run on both the TQS and QE. Skyline [72] was used for integrations for targeted analyses. XCMS[74, 75, 76] was used for peak-picking and integrations for untargeted analyses. Data were subjected to quality control filtering as described in the text. For each compound detected, the B-MIS was chosen via the steps shown. RSD = relative standard deviation. *Only applicable in targeted analyses. **This cut-off is user defined; in our analyses, an internal standard was considered acceptable if it decreased the RSD of the pooled injections by 40%.

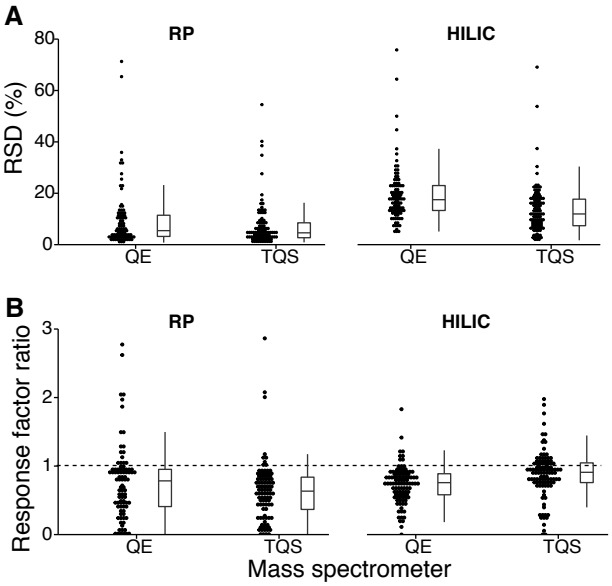

Figure 2.2: Box-and-whisker plots of (A) relative standard deviation (RSD) of peak areas and (B) response factor ratio on four LC–MS configurations, with individual dots representing single compounds (expanded horizontally for visualization of data density, $n = 101$ analytes in RP, 110 in HILIC). Response factor ratio is the ratio of the intensity of a standard injected in environmental matrix (less the ambient matrix signal, if applicable) to the intensity of the standard injected in water. Response factor ratio < 1 indicates ion suppression, response factor ratio > 1 indicates ion enhancement.

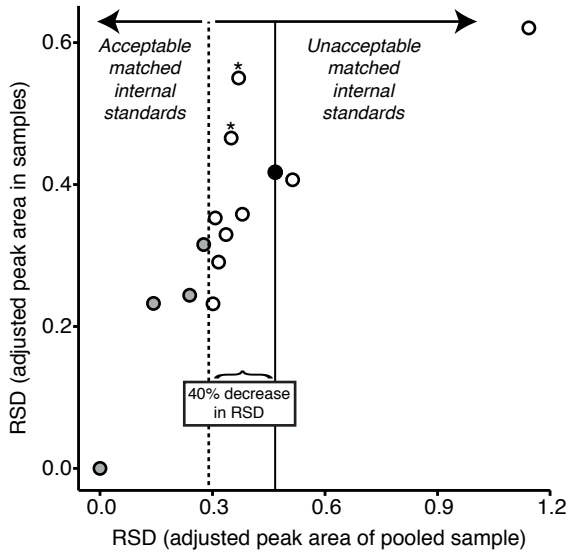

Figure 2.3: RSD of D_3-alanine in replicate injections of the pooled sample (x-axis) and across all samples in the environmental sample set (y-axis) after normalization to all possible internal standards. The black point represents the RSD of raw peak areas. The point at the origin is D_3-alanine normalized to itself. Filled points show the normalization results with an acceptable matched internal standard according to our cut-off of 40% RSD improvement (left of dotted line); open points show normalization results using unacceptable internal standards (right of dotted line, Figure 2.1). Results from other internal standards are in Figure 2.8.

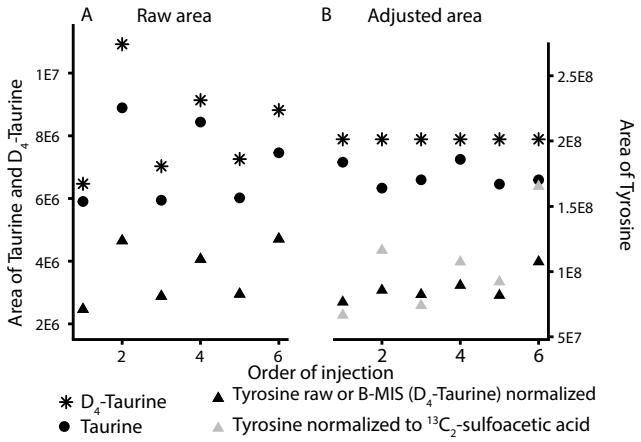

Figure 2.4: Example of B-MIS normalization for taurine and tyrosine in the diatom sample set. (A) Areas of D_4-taurine, taurine, and tyrosine in pooled samples after adjusting for dilution, in order of injection. (B) After normalization, areas are adjusted, with the B-MIS normalization reducing the RSD of taurine from 19% to 5%, while the RSD tyrosine decreases from 27% to 12% (black triangles). Normalizing to an unacceptable internal standard ($^{13}C_2$-sulfoacetic acid - grey triangles) did not reduce the RSD of tyrosine as effectively.

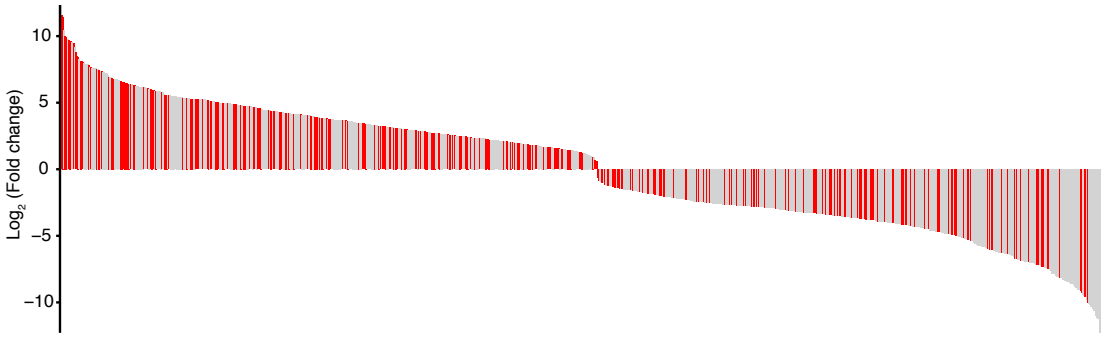

Figure 2.5: Log_2 fold change between surface (15 m) and deep (125 m) environmental samples for the 500 most abundant mass features in each sample type that were significantly different between depths (false discovery rate [83] adjusted p-values < 0.05, Student's t-test). A positive fold change indicates enrichment in the 125 m samples. Many mass features (292 of the 883 shown here) may not have been considered reliable peaks (RSD $> 20\%$) due to high obscuring variation which was minimized through B-MIS normalization; these mass features are highlighted in red.

2.8 Supplemental Files

- Supplemental Tables 2.2, 2.3, 2.4, 2.5, 2.6 (XLSX)
- Scripts: R-based code for quality control (https://github.com/IngallsLabUW/QualityControl) and B-MIS normalization (https://github.com/IngallsLabUW/B-MIS-normalization)

2.9 Supplementary Methods

2.9.1 Chromatography

For reversed phase (RP) chromatography, a Waters Acquity UPLC HSS Cyano column (1.8 μm particle size, 2.1 mm x 100 mm) equipped with a Waters Acquity UPLC HSS Cyano guard column (1.8 μm particle size, 2.1 mm x 5 mm) was used with 0.1% formic acid in water (Solvent A) and 0.1% formic acid in acetonitrile (Solvent B) at a flow rate of 0.4 mL/min. The HSS Cyano column has previously been used for analysis of compounds in complex marine samples with robust reproducibility of peak size and shape.[66, 67]. The column was held at 5% B for 2 minutes, ramped to 100% B over 16 minutes, held at 100% B for 2 minutes, and equilibrated at 5% B for 5 minutes (25 minutes total). The column was maintained at 35°C. The injection volume was 15 μL. When starting a batch, the column was equilibrated at the starting conditions for at least 10 minutes; several water blanks were run before and throughout a batch. After a batch, the column was flushed with 100% acetonitrile for 20 to 30 minutes.

For HILIC chromatography, a SeQuant ZIC-pHILIC column (5 mm particle size, 2.1 mm x 150 mm, from Millipore) was used with 10 mM ammonium carbonate in 85:15 acetonitrile to water (Solvent A) and 10 mM ammonium carbonate in 60:40 water to acetonitrile (Solvent B) at a flow rate of 0.15 mL/min. This column was compared with a Waters UPLC BEH amide and a Millipore cHILIC column; the pHILIC showed superior reproducibility and peak shapes. The column was held at 100% A for 2 minutes, ramped to 100% B over 18 minutes, held at 100% B for 5 minutes, and equilibrated at 100% A for 25 minutes (50 minutes total).

In marine samples, a major salt peak elutes at approximately 23 minutes (Figure 2.7A). The column was maintained at 30°C. The injection volume was 2 μL for samples and standard mixes. When starting a batch, the column was equilibrated at the starting conditions for at least 30 minutes. To improve the performance of the HILIC column, we maintained the same injection volume, kept the instrument running water blanks between samples as necessary, and injected standards in a representative matrix in addition to standards in water. After each batch, the column was flushed with 10 mM ammonium carbonate in 85:15 water to acetonitrile for 20 to 30 minutes.

2.9.2 Mass Spectrometry

For targeted metabolomics on the Waters Xevo TQ-S triple quadrupole (TQS) with electrospray ionization (ESI), the selected reaction monitoring (SRM) transitions were monitored over a two to three minute window around the retention time for the RP and a five to ten minute window around the retention time for the HILIC. For most metabolites, two SRM transitions were monitored, one for quantification and an additional for compound confirmation. Scheduling was set up to ensure at least 12 data points per peak were collected. MS parameters were as follows: capillary voltage of 0.5 kV for positive mode and 1.0 kV for negative mode, source temperature of 130°C, cone gas flow at 150 L/h and desolvation gas flow at 1000 L/h for both positive and negative modes. Desolvation temperature was 500°C or 600°C for HILIC and RP, respectively.

For untargeted analyses with HILIC, the capillary temperature was 320°C, the H-ESI spray voltage was 3.3 kV, and the auxiliary gas heater temperature was 100°C. The S-lens RF level was 65. Sheath gas, auxiliary gas, and sweep gas flow rates were maintained at 8, 3, and 0, respectively. Polarity switching was used with a scan range of 80 to 900 m/z and a resolution of 60,000. For RP analysis of samples, the capillary temperature was 320°C, the H-ESI spray voltage was 3.8 kV, and the auxiliary gas heater temperature was 100°C. The S-lens RF level was 65. Sheath gas, auxiliary gas, and sweep gas flow rates were maintained at 40, 10, and 1, respectively. For RP, a full scan method was used with a scan range of 90

to 900 m/z and a resolution of 120,000 in positive mode.

2.9.3 Sensitivity Analysis

We performed a sensitivity analysis to evaluate the differences in instrument performance between the QE and TQS, which emphasized the need for careful normalization (Table 2.3). We determined the limit of detection (LOD) and linear range for each compound on column, and ion suppression in matrix for both of our MS platforms. To do this we injected mixes of standards at 0.05, 0.5, 5, 50, 250, 500, and 1000 nM (for compounds monitored by RP) or 5, 10, 50, 250, 500, and 5000 nM (for compounds monitored by HILIC) at least three times. LOD was determined as the concentration at which a peak was observed in all three replicate injections after passing quality control (see section on quality control parameters). Five replicate injections of standards spiked in a matrix representative of a marine microbial community (our environmental samples pooled together, see next section) were used to calculate the relative standard deviation (RSD) of compound peak areas measured on both RP and HILIC columns on the TQS and QE platforms. To assess ion suppression, we injected standards in water, standards in the pooled environmental samples, and the pooled environmental samples without standards five times.

2.9.4 Sample Preparation

Diatom Cultures

Axenic cultures of the sea ice diatom *Nitzschia lecointei* were grown in 35 mL borosilicate tubes and artificial seawater media (Instant Ocean salts in 1 L MilliQ water with f/2 nutrients added) with a 12 hr light:dark cycle. Triplicate cultures were grown at the following conditions: -1°C and salinity of 32 g/kg, -1°C and salinity of 41 g/kg, 4°C and salinity of 32 g/kg, and 4°C and salinity of 41 g/kg. Cultures were monitored by fluorescence and harvested during exponential growth by gentle vacuum filtration onto 47 mm 0.2 μm PTFE filters. Filters were stored at -80°C until extraction.

Environmental Samples

Environmental samples for metabolite analysis were collected at four depths (15, 45, 75, and 125 meters) north of the Hawaiian Islands at 24° 33.284' N, 156° 19.790' W on July 31, 2015. Water was sampled from Niskin bottles into acid-washed 20 L carboys. All samples were filtered onto 142 mm 0.2 μm Durapore filters using peristaltic pumps, polycarbonate filter holders, and Masterflex PharMed BPT tubing (Cole-Parmer). Filters were quenched in liquid nitrogen immediately after filtration and stored at -80°C until extraction. Each sample was 30-40 L filtered seawater; each filter was split into three equal parts for triplicate extractions.

2.9.5 Metabolite Extraction

All solvents were Fisher LC–MS Optima grade. Work was done on combusted aluminum foil whenever possible; tweezers and dissecting scissors used to cut filters were rinsed with water in 10% formic acid, methanol, and dichloromethane before use and between samples; glassware (including Pasteur pipettes) was combusted before use. A blank PTFE filter was extracted alongside environmental and diatom samples as a methodological blank.

We tested four extraction procedures on marine microbial samples: boiling ethanol, hot water, cold acetonitrile:methanol:water (as described in detail in Canelas *et al.* 2009), and modified Bligh-Dyer extraction [70, 71]. We found that the boiling ethanol method resulted in an overall lower yield, the acetonitrile:methanol:water extraction yielded smaller peaks for the phosphorylated compounds, and the boiling water method would not extract more lipid-like compounds. Therefore, we decided to use the modified Bligh-Dyer as described below.

Extraction solvents (dichloromethane (DCM) and 1:1 methanol (MeOH) to water) were prepared the day before and stored at -20°C overnight. Bead beating tubes (15 mL, PTFE Jensen Inert Products) were filled with approximately 3 mL of a mixture of 100 μm and 400 μm silica beads and samples were placed into tubes. Approximately 2 mL of both extraction

solvents were added with isotope-labeled extraction internal standards (Table 2.1). The samples were shaken on a FastPrep-24 Homogenizer (MP Biomedicals) for 30 seconds, then chilled at -20°C for 15 minutes. The homogenization step was repeated twice for a total of 30 minutes at -20°C over three bead beating cycles. The samples were centrifuged at 4,300 rpm for 2 minutes at 4°C to separate the aqueous and organic phases.

The aqueous phase was removed from the bead beating tube using a glass Pasteur pipette and transferred to a 50 mL glass centrifuge tube. The remaining DCM fraction was rinsed 3 times with ∼2 mL of cold MeOH:H_2O with a bead beating and centrifuge step after each rinse. Cold DCM (∼2 mL) was added to the centrifuge tubes containing the combined aqueous rinses. The tubes were shaken by hand, and centrifuged at 1,800 rpm for 5 minutes at 4°C. The aqueous phase was removed and transferred to a clean 20 mL glass vial and dried under clean N_2.

DCM remaining in the bead beating tubes was transferred to the glass centrifuge tubes and rinsed twice with 2 mL of DCM, vortexing for 10 seconds after each rinse. The glass centrifuge tubes containing the combined organic rinses were spun at 1,800 rpm for 5 minutes at 4°C. The organic phase was transferred to a clean 20 mL glass vial, and dried under clean N_2.

The dried aqueous fractions were re-dissolved in 380 μL of water; the dried organic fractions were re-dissolved in 380 μL of 1:1 water:acetonitrile. Both received a 20 μL spike of isotope-labeled injection standards (Table 2.1). Isotope-labeled internal standards added before extraction yielded information about extraction efficiency and phase separation, while internal standards added after extraction monitored for injection errors. In the samples discussed here, and in other cases with high extraction reproducibility, these internal standards were interchangeable during B-MIS normalization because they both experienced obscuring variation.

Pooled samples were run multiple times throughout a sample run, at least three times at full strength and three times at half strength (diluted 1:1 with pure water) to train B-MIS normalization, as described in the main text. Standards in water and standards spiked into

an aliquot of the pooled sample were analyzed at the beginning and end of each sample run to set limits on the in-house quality control retention time and ion ratio ranges. Samples were analyzed within 12 or 72 hours of extraction for the HILIC and RP analyses, respectively, and stored at -80°C until analysis.

2.9.6 Standards

Most metabolite standards (isotope-labeled and unlabeled) were purchased from Sigma Aldrich, Santa Cruz Biotechnology, Millennium Enterprises, American Custom Chemical Corporation, and Cambridge Isotope Laboratories. Cinnamoyl-HSL was provided by the C. Harwood laboratory (University of Washington, Seattle, WA); DHPS (2,3-Dihydroxypropane-1-sulfonate) was provided by A. Bourdon and S. Champagna (University of Tennessee, Knoxville); N-acetyltaurine was provided by A. Cook and K. Denger (University of Konstanz, Germany); DMSP (Dimethylsulfoniopropionate) was provided by W. Whitman and M. Moran (University of Georgia). Full information about providers are included in Table 2.1.

Internal standard peaks in both targeted and untargeted data were integrated using Skyline for small molecules [72]. We inspected the peak shape and intensity of the internal standards and determined whether the signal was reliable and should be used for normalization. In the untargeted HILIC data for the environmental samples, D_4-succinic acid and $^{13}C_2$-acetyl CoA were removed because of these reasons. Thus these internal standards are not included for the environmental data in Tables 2.4 and 2.6 or Figure 2.8.

2.9.7 Quality control of integrated peaks

After integrations, we use an in-house quality control (QC) filter to ensure proper metabolite identification based on retention time, signal to noise ratio (> 4), absolute peak height, exact mass (< 5 ppm, for QE), and fragment ion ratio (for TQS). Acceptable retention times and ion ratios of a compound were set by the retention time of the standards (deviation ≤ 0.2 minutes of the standard run for RP, ≤ 0.4 minutes for HILIC). If the peak area of an analyte

in a sample was not more than three times the peak area in the methodological blank, the data for that analyte were not used. Finally, peak height was used to flag peaks that were overloaded and thus not acceptable for relative quantitation. To filter low quality LC-MS features (pairs of m/z and retention time) in untargeted analyses, we discarded features from the first 0.5 minutes, after 17.8 minutes (RP), or after 25 minutes (HILIC) because features that elute at these times exhibit severe ion suppression. We applied additional filters to remove features that were detected in less than one quarter of the samples, and those with an average pooled signal less than three times the average methodological blank signal.

2.10 Supplementary Tables and Figures

	Standard	Fraction	Column	z	Inj Conc (nM)	Extracted m/z
Extraction Standards	Indole 3 Acetic Acid, D_5	Aqueous	RP	+1	5000	181.1021
	Arachidonic Acid, D_8	Organic	RP	+1	20000	313.2977
	Tryptamine, D_4	Aqueous	RP	+1	2500	164.1246
	L-Cysteic Acid, D_3	Aqueous	HILIC	-1	1000	171.0161
	Sulfolactic Acid, $^{13}C_3$	Aqueous	HILIC	-1	1000	171.9913
	Sulfoacetic Acid, $^{13}C_2$	Aqueous	HILIC	-1	2500	140.9774
	Isethionic Acid, $^{13}C_2$	Aqueous	HILIC	-1	500	126.9981
	Taurine, D_4	Aqueous	HILIC	-1	1000	128.0325
Injection Standards	L-Phenylalanine, D_8	Aqueous	RP	+1	100	174.1365
	Riboflavin-dioxopyrimidine (Vitamin B_2), $^{13}C_4$, ^{15}N	Aqueous	RP	+1	100	383.1531
	L-Tryptophan, D_3	Aqueous	RP	+1	100	208.1161
	Pyridoxal, D_3	Aqueous	RP	+1	200	171.0843
	Thiamine, $^{13}C_3$	Aqueous	RP	+1	200	269.1252
	Alpha-Tocopherol, D_6	Organic	RP	+1	5000	436.4182
	Docosahexaenoic Acid (DHA), D_5	Organic	RP	+1	2000	334.2789
	Eicosapentaenoic Acid (EPA), D_5	Organic	RP	+1	5000	308.2632
	DL-Proline, D_7	Aqueous	HILIC	+1	1000	123.1145
	DL-Valine, D_8	Aqueous	HILIC	+1	1000	126.1365
	L-Methionine, D_3	Aqueous	HILIC	+1	1000	153.0722
	DL-Histidine, ^{15}N	Aqueous	HILIC	+1	2500	157.0738
	DL-Alanine, D_3	Aqueous	HILIC	+1	5000	93.0738
	L-Isoleucine, ^{15}N	Aqueous	HILIC	+1	500	133.0989
	Acetyl CoA, $^{13}C_2$	Aqueous	HILIC	-1	1000	810.1252
	Succinic Acid, D_4	Aqueous	HILIC	-1	10000	121.0444
	Guanosine Monophosphate, $^{15}N_5$	Aqueous	HILIC	-1	2000	367.0359
	Adenosine Monophosphate, $^{15}N_5$	Aqueous	HILIC	-1	5000	351.0411

Table 2.1: Isotope-labeled internal standards used in this study. The column, ion mode (z), injection concentration, and fraction used for analysis for each standard are provided. Note that the group of standards used for B-MIS normalization can be customized to suit the needs of individual studies; this is just an example set.

Table 2.2: MS and LC conditions for each analyte in our targeted method. Column (RP or HILIC) and retention times (RT) for each compound, with extracted *m/z* (for QE) and transitions monitored (for TQS). Collision energy (CE) and cone voltage (CV) were tuned for the SRM analysis only, MS conditions for high-resolution reported in text. We did not run RP–QE under negative polarity, thus we did not collect high-resolution data for a few compounds, noted as nd in the *m/z* column. Compounds listed in bold are used as internal standards. Compounds with a † at the end of their name are monitored on both HILIC and RP columns. Compounds with a ‡ were added to our method after the sensitivity analysis was conducted, thus we report their MS and LC conditions here but they are not included in Table 2.3. This table is provided as a separate file.

Table 2.3: Full results from sensitivity analysis. Linear range is reported for compounds in water as fmol on column with the lower limit being the LOD (described in text) using raw areas. Relative standard deviation (RSD) calculated over 5 injections of the standard in representative matrix, reported as %. R^2 is the average R^2 of the 3 standard curves of each compound within the reported linear range. We did not run the RP–QE under negative polarity (noted as nd in the Lin Range, RF ratio, RSD, and R^2 columns). Occasionally, compounds exhibited ion suppression so intensely that we were no longer able to detect the compound in matrix (noted as dl). A number of compounds were added after the initial linear range analysis (noted as $**$); for these compounds we do not report the linear range or R^2, but do supply an RF ratio and RSD. Compounds in Table 2.2 marked with a ‡ are not included in this table since they were added to the method after the sensitivity analysis was completed. This table is provided as a separate file.

Table 2.4: Results of B-MIS normalization of the untargeted dataset, summarized by internal standard. Fraction of A-MIS (acceptable-matched internal standard) is the fraction of quality mass features of the technical replicates that may have used each internal standard, while the fraction of B-MIS is the fraction of quality mass features that choose each internal standard. If the fraction of A-MIS is equal to the fraction of B-MIS, then every time the internal standard was considered acceptable, it is chosen. Users should consider adding more internal standards that are similar in chemistry to internal standards where the fraction of A-MIS is $\approx$ fraction of B-MIS. This table is provided as a separate file.

Table 2.5: RSD of targeted analytes before and after B-MIS normalization in our two example sample sets (environmental and diatom cultures). For each analyte in each sample set, we have listed the acceptable internal standards for normalization. The best option (B-MIS) is listed first with other internal standards that meet the cut-off rules listed after. The 'Matched?' column is true if there is an acceptable internal standard in common between the sample sets. Analytes that were not detected or were overloaded are noted as: not detected in pooled. This table is provided as a separate file.

Table 2.6: Results from evaluating internal standards as analytes for the untargeted HILIC data from the environmental samples. 'RSD of Pooled Sample' is the RSD of multiple injections of the pooled sample at full and half strength normalized to the internal standard given in the 'Internal Standard used for Normalizing' column. 'RSD of Samples' is the RSD of the internal standard across the whole sample set, normalized to the internal standard given in the 'Internal Standard used for Normalizing' column. 'RSD of pooled with no normalization' is the same value as the injection volume normalized value in the 'RSD of Pooled Sample' column; this is used to calculate the 'percent change from raw' value. The 'acceptable normalization?' column is true if 'percent change from raw' is above 0.4 (see text for discussion of this cutoff value). Data are plotted in Figures 2.3 and 2.8.

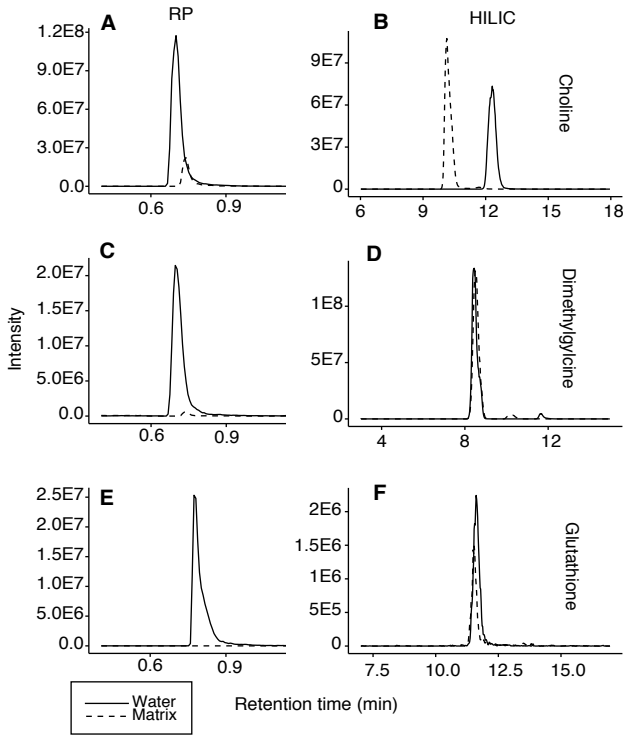

Figure 2.6: Ion suppression of polar analytes via RP or HILIC chromatography. Example chromatograms of standards of three compounds (choline (A, B), dimethylglycine (C, D), and glutathione (E, F)) run in water (solid line) or representative matrix (dashed line), on two different LC systems (RP (A, C, E) and HILIC (B, D, F)). These data were acquired on the triple quadrupole mass spectrometer.

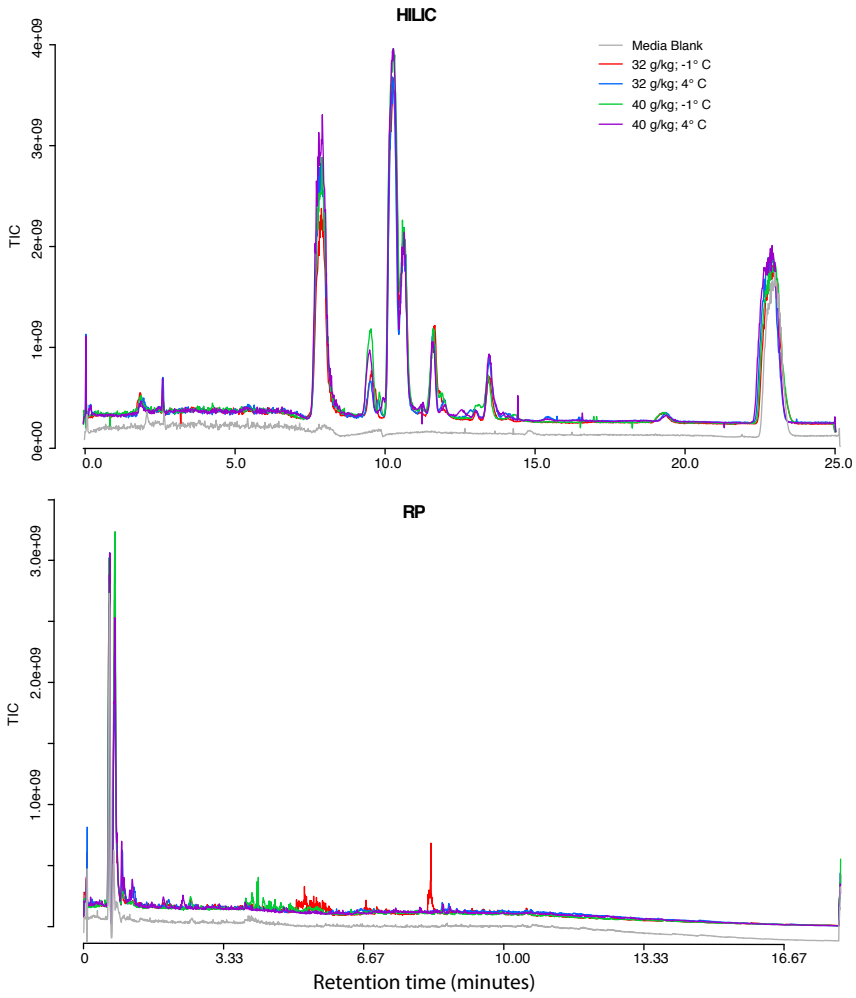

Figure 2.7: Total ion chromatograms from the QE for the aqueous fraction of four treatments of the diatoms and a methodological blank (offset on the y-axis by $1.25 \text{x} 10^8$), run on the HILIC column (top), and on the RP column (bottom).

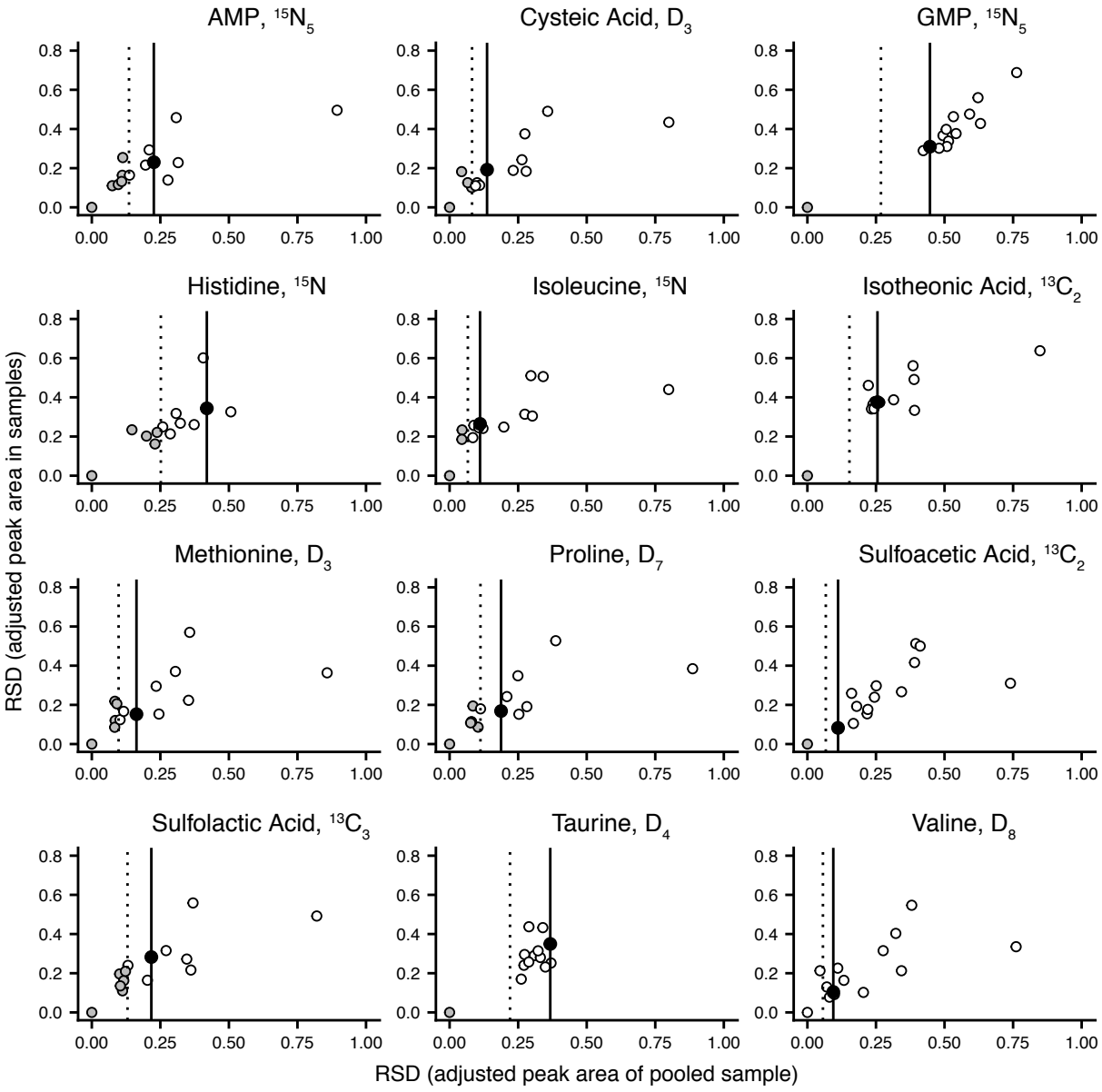

Figure 2.8: Results from evaluating internal standards as analytes for the untargeted HILIC data from the environmental sample set. Colors and lines are as in Figure 2.3. Note that Valine, D_8 had an original RSD < 0.1, therefore B-MIS was not applied (Figure 2.1, all points white).

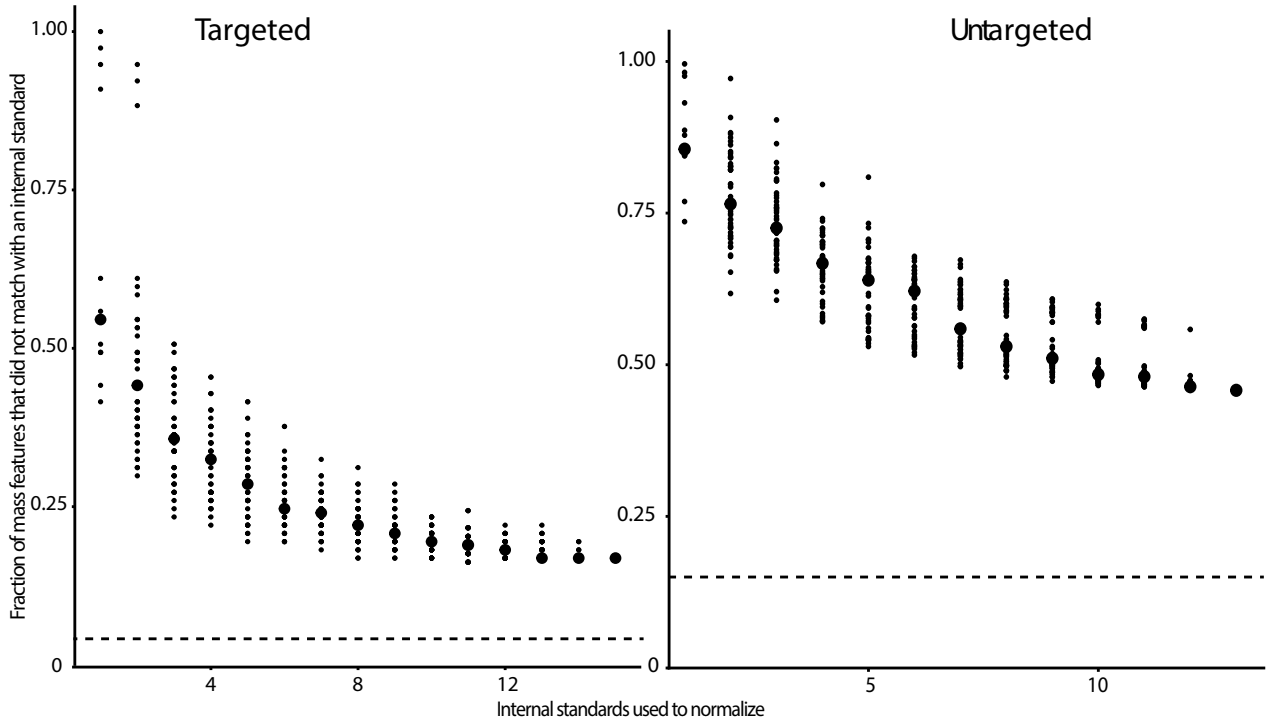

Figure 2.9: Number of internal standards included during B-MIS normalization vs the fraction of analytes (in targeted analyses, left) or mass features (in untargeted analyses, right) whose obscuring variation did not match with an acceptable internal standard. For each value along the x-axis, normalization was done either with all possible combinations of internal standards or with 50 possible combinations of internal standards (if there were more than 50 possible combinations). Each combination is represented in a small point; the median of the iterations is shown in large circle. The data used here are from the environmental HILIC samples. Only quality mass features (final RSD of the pooled < 0.2 after normalization by all included internal standards) were used in the untargeted analysis. Dashed lines indicate the 5% and 15% of mass features that had an original RSD of < 0.1 in targeted and untargeted analyses, respectively; B-MIS normalization does not attempt to normalize these mass features. For this untargeted data, 13 internal standards are used for B-MIS normalization because D_4-succinic acid and $^{13}C_2$-acetyl CoA do not ionize well enough at the concentrations injected in this matrix for robust detection with the QE. With all internal standards, over 54% of mass features in the untargeted data and 83% of compounds in the targeted data matched with an internal standard for normalization.

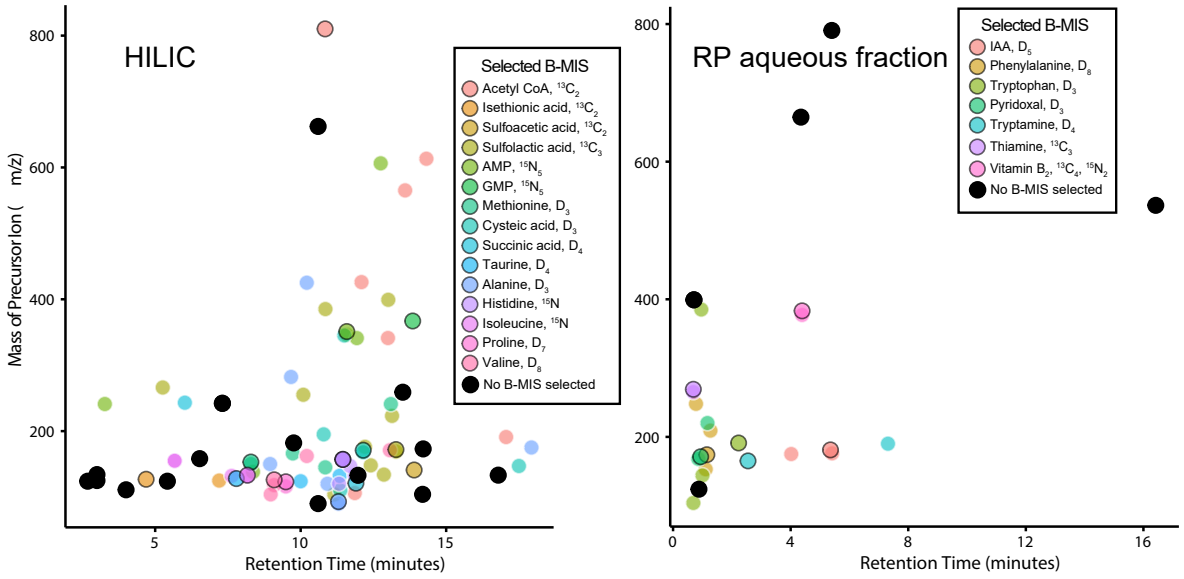

Figure 2.10: Retention time vs m/z of the measured analytes in the targeted environmental aqueous extracts. Color shows which internal standard was chosen for each analyte, with each internal standard outlined in black. Black markers are analytes whose obscuring variation was not improved by normalizing to any internal standard. Full results are given in Table 2.5.

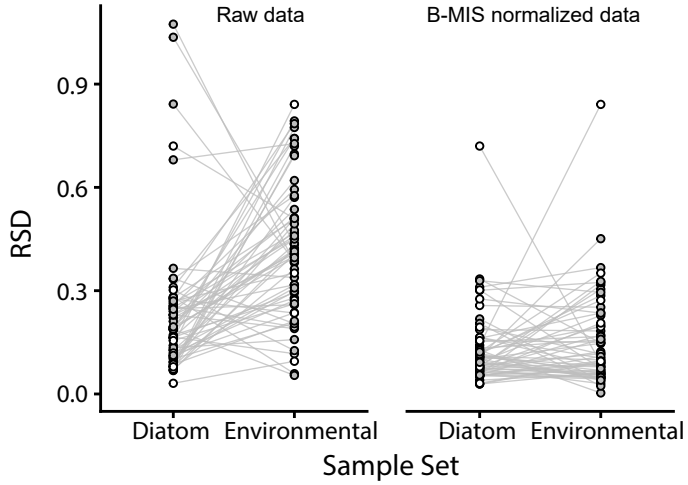

Figure 2.11: RSD of raw and B-MIS adjusted areas of replicate injections of pooled samples in the two matrices used in the targeted analysis of this study. The lines connect the same compound between both sample sets, demonstrating that the obscuring variation of each analyte changes with matrix type, but not in a predictable manner. Open circles are analytes that did not find an acceptable matched internal standard and are therefore not normalized (see Figure 2.1), while closed circles are analytes that did use B-MIS normalization. After B-MIS normalization, the range of RSD is similar between the sample sets. Full results are given in Table 2.5.

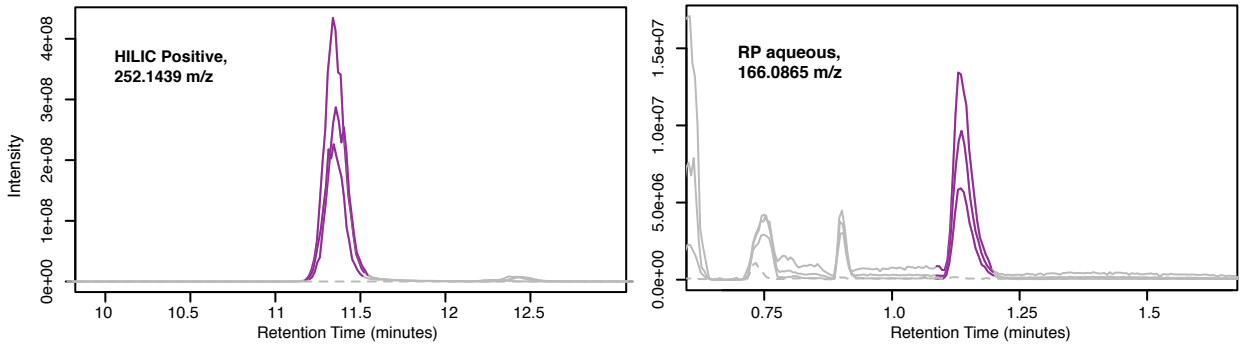

Figure 2.12: Extracted ion chromatograms of repeat injections of the pooled environmental sample acquired during the untargeted analysis of environmental samples. Two example mass features (from HILIC (left) and RP chromatography (right)) showed significant obscuring variation that could be minimized by B-MIS. Both of these mass features would have been excluded from downstream data analysis without B-MIS normalization if applying an RSD filter. XCMS peak integration shown in violet, methodological blank shown as dashed line.

Chapter 3

PARTICULATE METABOLITES AND TRANSCRIPTS REFLECT DIEL OSCILLATIONS OF MICROBIAL ACTIVITY IN THE SURFACE OCEAN

3.1 Abstract

[1]Light is the primary input of energy into the sunlit ocean. This solar forcing fuels photosynthesis and organic matter production by primary producers and drives oscillations in the expression of genes throughout the microbial community, yet *in situ* measurements of metabolites, the products of gene expression, over the diel cycle are lacking. We evaluated community-level biochemical consequences of oscillations of light in the North Pacific Subtropical Gyre by quantifying 79 metabolites in particulate organic matter in surface waters every four hours over eight days. Total particulate metabolite concentration peaked at dusk, including when calculated relative to biomass estimates. The concentrations of 55/79 (70%) individual metabolites exhibited statistically significant 24-hour periodicity. Paired metatranscriptome analysis revealed the taxa involved in production and consumption of a subset of metabolites. Primary metabolites involved in anabolism and redox maintenance had significant 24-hour periodicity and diverse organisms exhibited diel periodicity in transcript abundance associated with these metabolites. Osmolytes were among the compounds with the largest oscillations in concentration, implying rapid turnover and supporting prior evidence of functions beyond cell turgor maintenance. For example, the large daily concentration changes of the osmolyte trehalose paired with transcript and culture data showed that trehalose is produced by the nitrogen-fixing cyanobacterium *Crocosphaera*, likely as a way to store energy to drive nighttime metabolism. Together, paired measurements of par-

[1]A version of this chapter is in revisions for the journal *mSystems*

ticulate metabolites and transcripts resolve strategies microbes use to manage daily energy and redox oscillations and highlight dynamic metabolites that should be studied further to elucidate their roles in marine microbial ecosystems.

3.1.1 Importance

Fueled by light, phytoplankton produce the organic matter that supports ocean ecosystems and carbon sequestration. Ocean change impacts microbial metabolism with repercussions for biogeochemical cycling. However, much of what is known about the relationship between microbial metabolism and environmental conditions is based on laboratory studies of cultures that do not realistically capture the natural environment. The small molecule products of cellular metabolism, metabolites change rapidly in response to environmental conditions and form the basis of energy and nutrient management and storage within cells. By pairing measurements of metabolites and gene expression in the stratified surface ocean we reveal strategies of microbial energy management over the day-night cycle and hypothesize that oscillating metabolites are important substrates for dark respiration by phytoplankton. These first diel measurements of *in situ* metabolite concentrations form the basis for future work into the specific roles these compounds play in marine microbial communities.

3.2 Introduction

Light is a powerful forcing on metabolism in the surface ocean, acting at the molecular level to drive global biogeochemical cycles[84]. The diel oscillation of light fuels phytoplankton photosynthesis and organic matter production that supports ocean ecosystems and carbon sequestration. In surface waters, the direct or indirect consequences of this diel forcing can be seen in daily oscillations in cell division[27], gross primary production, net community production[28], grazing[85], viral infection[86], and nitrogen fixation[26]. Genes associated with a wide variety of cellular processes also exhibit diel oscillations in transcript abundance, reflecting the capture of light energy and its conversion to chemical energy during daylight, a process that fuels metabolism over a 24 hour period[87, 24, 23, 88, 89]. Temporal partitioning

of anabolism and catabolism creates diel patterns in total particulate organic carbon (POC) and the macromolecular composition of POC[89, 90, 91, 31, 92].

Though POC in the surface ocean is primarily macromolecules [93, 94], a suite of small molecules (metabolites less than around 800 Da) produced within cells helps shape the internal and external chemical environment of the plankton community, with implications for dependencies among different taxa. Yet an inventory of these compounds and the plasticity of their concentrations remain largely unknown [10]. Measurements of the chemical diversity and concentration of metabolites present in marine microbial communities are scarce, and the suite of compounds detected is strongly biased by the methods employed [22]. Small polar molecules, in particular, are rarely measured although they are the main component of the aqueous cytosol [95]. Intracellular metabolite profiles of model marine microbes are taxon-specific and respond to environmental perturbations, including diel oscillations in available light [96]. Measurements of many of these metabolites have not yet been conducted for natural plankton communities and some metabolites are without annotated biosynthetic or catabolic pathways [97, 98, 51, 99, 100, 42, 32]. Thus, an *in situ* inventory of intracellular metabolites will facilitate a deeper understanding of marine microbial physiology and interactions that drive ecosystem diversity and activity [8, 101].

Here we measured particulate metabolite concentrations in samples collected from surface waters near Station ALOHA (A Long-Term Oligotrophic Habitat Assessment; 22.75 °N, 158 °W) in the North Pacific Subtropical Gyre (NPSG) during eight diel cycles. These data provide an inventory of metabolites in the oligotrophic surface ocean over the diel cycle. We paired metabolite concentrations with observations of gene expression, POC, particulate nitrogen (PN), and flow cytometry (FCM) cell counts.

We find that the molar concentration of 70% of our targeted metabolites oscillated with 24-hour periodicity, reflecting large scale community synchrony to the daily cycle of light. Our analysis identifies diel oscillations in compounds that play important roles in managing light-induced redox reactions and biosynthesis of building blocks and energy stores. These compounds are ultimately conduits of energy and nutrients through the microbial ecosystem

as they are exchanged between diverse organisms, with repercussions for community diversity and function [11, 12, 38]. Our metabolite data reveal the chemical environment within natural populations of marine plankton. Paired with metatranscriptomes, these data point to potential metabolic strategies that organisms deploy to cope with an oscillating energy supply.

3.3 Results

3.3.1 Oscillatory dynamics of the phytoplankton community

Our sampling targeted an anticyclonic eddy and followed two drifters to facilitate Lagrangian sampling of surface ocean water with minimal mixing or forcing other than the day-night cycle. Samples were collected for two multi-day sampling periods in summer 2015 (period one: July 26, 6:00 – July 30, 6:00; period two: July 31 18:00 – August 3, 18:00). The eddy was characterized by warm, nutrient-deplete surface waters typical of the persistently oligotrophic NPSG [85, 29] (Table 3.1). Photosynthetic picoeukaryotes, *Prochlorococcus*, and *Crocosphaera* contributed substantially to phytoplankton biomass[26] (Figure 3.1). POC, which includes bulk community biomass, and phytoplankton-specific biomass oscillated with significant 24-hour periodicity (Figure 3.1). Cell abundances and total biomass of *Prochlorococcus* and *Crocosphaera* populations increased between the first and second sampling periods (Table 3.1). Wind speed also increased between the first and second sampling periods, resulting in an increase in the mixed layer depth from 21 $\pm$ 5 to 36 $\pm$ 6 m. Additionally, we observed a decrease in the number of significantly diel metabolite oscillations during the second sampling interval, from 55 to 9 (Supplemental Table 3.2). This change may be related to deepening of the mixed layer, causing individual plankton in the mixed layer to experience more variable light energy with a lower average light level and resulting in less pronounced peaks in metabolites. However, we have insufficient evidence to investigate this hypothesis further. We therefore focus our analysis on data collected during the first sampling period where metabolite oscillations were more pronounced.

3.3.2 Metabolite Inventory

A total of 79 targeted metabolites were detected across samples (Supplemental Table 3.2). Total particulate metabolite concentration (as molar concentration and as a % of POC or PN) increased during the day and decreased at night (Figure 2). The most abundant compounds were osmolytes, including glycine betaine (GBT), homarine, 2,3-dihydroxypropane-1-sulfonate (DHPS), and dimethylsulfoniopropionate (DMSP); nucleobases (particularly guanine); and amino acids related to nitrogen metabolism, such as glutamic acid and glutamine (Supplemental Tables 3.2, 3.3, Figure 3.2). At dusk, quantified metabolites totaled 1.7 ± 0.2% of POC and 3.1 ± 0.6% of PN (Figure 2), with free nucleobases and amino acids comprising most of the metabolite derived nitrogen (2.5 ± 1.1% of PN at dusk, Supplemental Table 3.3).

Multivariate analyses were used to determine if time of day influenced the community metabolome. NMDS analysis shows that samples collected at different times were different (ANOSIM, $R = 0.19$, $p = 0.001$). Samples collected near sunrise (6:00) were more similar to one another than those collected at other times of day and are most dissimilar to samples collected near sunset (Figure 3.7, Supplemental Table 3.4).

3.3.3 Metabolite diel periodicity

To determine whether metabolite oscillations were driven by changes in biomass or by changing cell physiology resulting in changes in the intracellular concentration, we calculated metabolite concentrations relative to water volume filtered, resulting in values of molar concentration (nmol L^{-1}, Supplemental Table 3.7), and to POC, resulting in values of nmol per μmol POC. Bulk and individual metabolite concentrations oscillated in both molar concentration and relative to POC (Figures 3.2, 3.3A, B). The molar concentration (nmol L^{-1}) of 55 metabolites (70%) had significant 24-hour oscillations, with 26 reaching a maxima in concentration within two hours of 18:00 and 20 reaching their peak concentration within 2 hours of 14:00 (Figure 3.3A,B, Supplemental Table 3.2). When relative to POC (nmol μmol POC-1),

37 compounds (47%) showed diel oscillations (Supplemental Table 3.2), and the mean time of peak concentration shifted to earlier in the afternoon (Figure 3.3A). POC reflects total community biomass and detritus, so to avoid assumptions of metabolite source, we present molar concentrations throughout except where metabolite source can be constrained to a specific phytoplankton, in which case we present metabolite concentration relative to the cell number or biomass of the source organism.

Metabolites with significant oscillations (Figure 3.3C) had daily fold changes ranging from 2 to 12.8, all of which exceeded the 1.2- and 1.8-fold changes of POC and the sum of FCM-derived phytoplankton biomass, respectively (Figure 3.4). These compounds main functions are in cell turgor, anabolism, energy storage, and redox balance. The disaccharides trehalose and sucrose displayed the most robust oscillations (p-value $< 1\text{x}10^{-13}$, Figures 3.4, 3.5). Trehalose and sucrose are known osmolytes, and nearly all other identified osmolytes (9/10) showed diel oscillations (Figures 3.4, 3.6A, Supplemental Table 3.2). Glutamic acid is the only known osmolyte that did not have a significant oscillation in molar concentration (Supplemental Table 3.2).

Significantly diel metabolites involved in cell growth include the three methionine-cycle compounds that were detected, S-adenosyl methionine (SAM), S-adenosyl homocysteine (SAH), and methionine. 5'-Methylthioadenosine (MTA) is produced from SAM during polyamine synthesis and had a temporal pattern that closely matched SAM (Figures 3.3C, 3.6A), such that SAM/MTA remained relatively constant. The cofactor pantothenate (Vitamin B5) was one of the few compounds that peaked in the morning (Figure 3.3C). Cofactors riboflavin and niacin (Vitamins B2 and B3) oscillated with maxima near dusk. Reduced glutathione oscillated with an afternoon peak (Figure 3.3C).

3.3.4 Connections between metabolites and transcripts

To investigate the relationships between gene expression and metabolite concentration we used the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to connect metabolites with transcripts annotated as encoding proteins that directly produce or degrade each

metabolite. All but four of the diel metabolites were linked to at least one annotated prokaryotic or eukaryotic transcript (Figure 3.8). Glucosylglycerol, ergosterol, and isethionic acid are in the KEGG database but no transcripts were annotated in our dataset as directly producing or degrading them, while homarine is not included in the KEGG database.

Although the number of transcripts associated with each metabolite is inherently biased by the databases used and the depth of sequencing, transcripts provide insight into the number and identity of organisms and pathways that may be responsible for the metabolite's synthesis and degradation. The orders containing *Crocosphaera*, *Prochlorococcus*, *Pelagibacter ubique*, and other unclassified alphaproteobacteria comprised 50% of all prokaryotic transcripts that could be linked to metabolites (Supplemental Table 3.8). Dinoflagellates (Dinophyceae), non-diatom stramenopiles (Stramenopiles), haptophytes (Haptophyceae), non-metazoa opistokonts (Opisthokonta), and diatoms (Bacillariophyta) comprised 70% of eukaryotic transcripts linked with metabolites (Supplemental Table 3.9).

Adenosine monophosphate (AMP), SAM, and SAH stand out as the diel metabolites with the largest number of associated diel transcripts, with 181, 124, and 113 transcripts respectively (Figure 3.6B, 3.8). Most diel SAM and SAH transcripts were methyltransferases that use SAM and produce SAH (Figure 3.6, (Supplemental Tables 3.8, 3.8)). In most other cases, there were few diel transcripts associated with a metabolite (e.g. only 6 diel transcripts were associated with trehalose, Figures 3.6B, 3.8).

As a first step to investigate the temporal relationship between gene expression and metabolite concentration, we estimated the lag-time between peaks in metabolites and transcripts that exhibited significant diel periodicity in abundance. This analysis showed a broad distribution in the lag-times between metabolites and transcripts (Figure 3.9), with no significant relationship between the peak time of prokaryotic or eukaryotic transcripts and their associated metabolites (Pearson correlation p-value > 0.05).

3.3.5 Disaccharide osmolytes can be attributed to cyanobacteria

We observed trehalose-related transcripts from eukaryotic phytoplankton and *Crocosphaera* (Figure 3.6B). Using published *Ostreococcus* cellular trehalose concentrations [96] and picoeukaryote cell counts, we estimated that picoeukaryote contribution to trehalose was 0.2–3.0 pmol L^{-1}, a small fraction of environmental trehalose (up to 627 pmol L^{-1}). The abundance of *Crocosphaera* (Table 3.1, Figure 3.1) and diel oscillations in the *Crocosphaera* transcript for trehalose 6-phosphate synthase/phosphatase (Figure 3.6, Supplemental Table 3.8) suggest *Crocosphaera* as the main contributor of trehalose during this field study. To test this hypothesis, we grew *Crocosphaera watsonii* WH8501 under a 12:12 light:dark cycle and measured 0.8 and 0.07 fmol trehalose $cell^{-1}$ at the end of the light and dark periods, respectively (Figure 3.5, Figure 3.10). Given the *Crocosphaera* abundance during our sampling and assuming similar intracellular concentration, this accounts for 1.8–670 pmol L^{-1} particulate trehalose, comparable to total particulate trehalose during our sampling (2.8–627 pmol L^{-1} across both sampling periods, Figure 3.5).

Multiple taxa expressed transcripts related to production and degradation of sucrose, including *Prochlorococcus* (Figure 3.6). To assess the potential contribution of *Prochlorococcus* to environmental sucrose concentrations, we measured the cellular sucrose quota in a culture of *Prochlorococcus* MIT1314 harvested midday during exponential growth. Using the cellular quota of sucrose in these cultures (range in biological triplicates: 1.4–2.1 amol $cell^{-1}$) and the abundance of *Prochlorococcus* during the environmental sampling, it is possible that all the observed sucrose could have been in *Prochlorococcus* during this study (Figure 3.5).

3.4 Discussion

As a whole, the metabolites we measured comprise up to 2% of POC and 3% of PN in our samples (Figure 3.2B, C). This is a reasonable value given 80% of surface POC is comprised of lipid, carbohydrate, and protein macromolecules [93, 94], and DNA, RNA, and pigments contribute several percent of the dry weight of actively growing microalgae [102].

Metabolite pools are dynamic, and an increase in the concentration of a given metabolite over time suggests that sources of that compound (anabolism, regeneration, uptake from dissolved pools, or polymer disassembly) are greater than sinks (catabolism, use, exudation, cell death, or polymer assembly). All else being equal, higher substrate concentrations lead to higher enzyme reaction rate, thus elevated metabolite pool size favors the maintenance of high cellular activity. In our study the highest metabolite concentrations corresponded with the daily peak in biomass near the end of the light period. The prevalence and amplitude of diel oscillations in metabolite concentrations reflect that many members of the surface microbial community near Station ALOHA were synchronized to diel light periodicity. Diverse organisms (including heterotrophs and autotrophs) expressed diel cycles in transcripts related to the production and degradation of these compounds, but the timing of diel transcripts did not prove to be a reliable indicator of metabolite levels, consistent with previous work showing that diel protein abundances and transcripts were not closely linked[88]. Nevertheless, the pairing of diel measurements of metabolites and transcripts allows investigation of how many and which organisms and processes may be responsible for the accumulation and depletion of a given compound.

3.4.1 Diel partitioning of anabolism, catabolism, and redox maintenance

The diel oscillations in POC and FCM-resolvable phytoplankton biomass reflect the alternation of carbon fixation, anabolism, and growth during daylight hours and respiration, catabolism, and mortality during the night (Figure 3.1)[27, 90, 85]. The community metabolome reflects these patterns with an overall increase in concentration throughout the day while cells are growing and a consistent morning phenotype (Figure 3.2, Figure 3.7A, B), reflecting nighttime use of energy stores and recovery from daytime oxidative stress [103]. Nearly half of the diel metabolites (26/55) had peak molar concentrations near dusk (Figure 3.3), corresponding with a peak in carbon biomass. However, for most (46/55) diel metabolites, the daily enrichment of a metabolite exceeded that of POC or total FCM-resolvable phytoplankton biomass, which had daily fold changes of 1.2 and 1.8, respectively (Figure

3.4). This suggests these metabolites likely had oscillations in intracellular concentration, as previously observed for many primary metabolites in non-marine cyanobacteria [104].

Primary metabolites are particularly powerful indicators of biochemical activity on the community scale. SAM, SAH, and AMP are compounds involved in biosynthesis and growth that had diel oscillations with daytime increases (Figures 3.3, 3.6). Individual transcripts associated with these molecules had diel patterns that peaked at all times of day, across a myriad of pathways and microbial taxa (Figure 3.6, Figure 3.8). Despite this diversity in use, the sum of community activity was reflected in diel oscillations of metabolite concentrations, which were synchronized with daytime biomass accumulation. Further evidence of this daytime community-scale anabolism is the diel oscillation of pantothenate (Vitamin B5), a component of Coenzyme A as well as Acyl Carrier Protein. Pantothenate peaked in the morning (Figure 3.3), suggesting that the community was poised to assemble these cofactors for daytime biosynthesis.

SAM is a ubiquitous methyl donor used by all living cells. During methylation, SAM is converted to SAH, which is then regenerated back to SAM via methionine. In addition to its role in methylation, SAM is essential for polyamine synthesis and is the most common riboswitch effector in prokaryotes [105]. SAM riboswitches have been observed in native Station ALOHA bacterioplankton populations [106]. SAH had an afternoon peak time, such that the SAM/SAH ratio was at a minimum during the day (Figure 3.6). This ratio reflects methylation potential, suggesting that the demand for methylation outstripped the supply of SAM in the light despite rising intracellular concentrations of SAM during the day that likely aid in maintaining high reaction rates. Over the dark period, SAM/SAH ratios recovered, suggesting that catabolic processes dominated and the need for SAM was diminished. Many cells require cobalamin (vitamin B_{12}) to catalyze the reactions that regenerate methionine, and SAH is elevated relative to SAM during cobalamin stress as cells struggle to complete the cycle [100]. Thus, it is possible that the lower SAM/SAH ratio additionally reflects a daytime increase in cobalamin demand among auxotrophic community members.

All living cells produce reactive oxygen species, and redox homeostasis is a critical part

of maintaining cellular function. Reactive oxygen species produced during photosynthesis accumulate over the day and present a continuing challenge for cells at night [107]. Strategies for managing oxidative stress range from the acute reactions that detoxify reactive oxygen species to mitigation and avoidance strategies, such as accumulating glycogen to assimilate excess reducing power in high-light conditions and avoid dangerous levels of membrane redox potential [103]. Reduced glutathione is the active form of a key component of the reactive oxygen species detoxification system and peaked during the afternoon (Figure 3.3B), as has been observed in cultures and field studies [108]. The daytime peak possibly reflects production to compensate for increased oxidative stress during the day and a subsequent decrease in production and oxidation of the residual pool overnight.

Riboflavin and niacin (vitamins B_2 and B_3) both accumulate during daytime (Figure 3.3, Supplemental Table 3.2). These are precursors to cofactors FMN/FAD and NAD/NADP, respectively. These cofactors are involved in electron transport chains for photosynthesis and respiration and are therefore key components of redox processes within cells. The daytime accumulation of riboflavin and niacin underscore that the community-wide processes of growth metabolism and redox maintenance occur in step with the diel cycle. During the day, photosynthesis produces NADPH and ATP and the requirements of carbon fixation result in an excess of NADPH. This NADPH is used to reduce oxidized species and biosynthesize reduced organic matter as a sink for electrons from photosynthesis. Cyanobacteria, for example, manage excess energy during the day by storing glycogen and producing small molecules that can either be stored or excreted [103, 107, 109, 110, 111]. At night, glycogen is catabolized via hydrolysis followed by glycolysis or the oxidative pentose phosphate pathway (OPPP), producing the reductant sources NADH and NADPH. Gluconic acid accumulation during the day (Figures 3.3, 3.4) may reflect less flux through OPPP during the day, while photosynthesis produces NADPH, followed by a switch towards OPPP at night [103].

3.4.2 Diel oscillations in osmolyte concentrations reveal their function diversity

Metabolites with osmolyte properties are among the most abundant compounds within marine microbial cells [95, 32, 43, 112, 113, 114] and exhibited diel oscillations (Figures 3.5, 3.6). One exception to this observation was glutamic acid, which plays other critical roles as a component of protein and in regulating nitrogen assimilation in addition to its osmotic properties [43]. In the absence of fluctuations in salinity or temperature, oscillations in osmolyte concentrations occurred in excess of or out of sync with biomass oscillations and point to alternative roles for this compound group such as previously hypothesized roles in short term energy and nutrient storage [43] (Figure 3.4, Supplemental Table 3.2). Intracellular accumulation of osmolytes occurred predominantly during the day when electron flow through the photosystems and the Calvin Cycle exceeds that required to maintain maximum division rates. The resulting need to dissipate reductant is typically channeled into the production of carbohydrates like glycogen[87, 103, 111], exopolymeric substances [115, 116], or into storage lipids[89, 117]. These energy stores are used to fuel cellular respiration and other activities at night, such as protein synthesis and preparing cells for photosynthesis[89, 103, 111, 117]. Unlike starch and storage lipids, osmolytes do not necessarily need to go through hydrolysis, β-oxidation, or glycolysis prior to entering the TCA cycle, and could be used as readily available substrates for energy production and as biosynthetic intermediates while macromolecular pools are being mobilized by the cell [109].

Trehalose was the most prominent diurnally oscillating compound (Figures 3.4, 3.5). Trehalose is an osmolyte produced by the unicellular diazotroph *Crocosphaera* [114, 118], some heterotrophic bacteria, and some phytoplanktonic picoeukaryotes, including *Ostreococcus* [96]. Transcriptomic evidence motivated us to measure trehalose in cultures of *Crocosphaera*, which was more concentrated at the end of the day than at the beginning of the day, similar to what we saw in the environment. Assuming trehalose in the environment is produced primarily by *Crocosphaera*, our results strongly suggest that intracellular trehalose concentrations have diel oscillations in the field (Figure 3.5).

Crocosphaera temporally separate photosynthesis and nitrogen fixation to protect nitrogenase from oxygen [119, 120, 121], they therefore need energy at night to draw down cellular oxygen and fuel nitrogen fixation [122, 123]. *Crocosphaera* has at least one gene encoding a protein homologous to glycoside hydrolases, family 15 [124], which contains enzymes that hydrolyze a variety of glycosidic bonds, including trehalose. Thus, it is possible that *Crocosphaera* use trehalose as a fuel for generating the electrons and ATP required for nitrogen fixation. Using the stoichiometry of these reactions [122, 125], we estimated that trehalose catabolism could have fueled 9–28% of the nighttime nitrogen fixation during this expedition[26] (calculation in Supplemental material). As much as 60% of total dark respiration by *Crocosphaera* is used to draw down cellular oxygen rather than to directly fuel nitrogen fixation [122], and, if we adjust our calculation accordingly, trehalose can produce 3.6–11% of the required respiratory substrates needed for *Crocosphaera* to effectively fix nitrogen at the rates measured[26].

The flux of carbon through trehalose may be an indicator of the accumulation and degradation of a larger glycogen pool that accumulates during the day and is used at night [126]. Shi et al. (2010) suggest that *Crocosphaera* cells are depleted of storage compounds at night, since prolonged dark does not result in increased nitrogen fixation [127]. If this hypothesis is correct, the total amount of nitrogen fixation possible is limited by the amount of energy stored in substrates such as trehalose and glycogen during daytime, and the ability to accumulate and use these compounds could have impacts on the nitrogen budget of the microbial community.

Another disaccharide osmolyte, sucrose, displayed an oscillation with a maximum daily concentration at 22:00. Sucrose is the major compatible solute in high-light *Prochlorococcus* [113], and the observed environmental variation may reflect the *in situ* accumulation and use of glycogen by *Prochlorococcus*. Though other organisms also expressed sucrose related genes (Figure 3.6), *Prochlorococcus* was the numerically dominant sucrose-producing organism detected in these populations (Table 1) and is known to accumulate polysaccharides during the day, particularly under nitrogen limitation [128]. If we assume that cellular quotas of sucrose

in *Prochlorococcus* grown in culture are like those in the environment, *Prochlorococcus* alone could explain the sucrose concentrations seen in the environment (Figure 3.5). Sucrose had a diel oscillation when calculated relative to *Prochlorococcus* cell counts and biomass (Figure 3.5, Figure 3.11). These potential intracellular oscillations lead us to hypothesize that *Prochlorococcus* uses sucrose for energy storage and not only as a compatible solute, as has been observed in non-marine cyanobacteria [104, 109].

Homarine and DMSP are known eukaryotic osmolytes [95, 43, 112, 129]. The amplitude and timing of the diel oscillations in these two compounds differ from those observed in phytoplankton picoeukaryote biomass (Figure 3.4), again suggesting that these compatible solutes play multiple roles within the microbes that use them as osmolytes. This diversity of functions is well established for DMSP, which influences grazing behaviors and can function as an antioxidant [95, 130]. DMSP is also a carbon and reduced sulfur source in the microbial community, with uptake and assimilation both tied to light availability [131, 132]. In our analysis, the only transcript related to DMSP encodes a SAR11 DMSP demethylase required for DMSP degradation (Figure 3.6). A dearth of data on the roles of homarine in marine microbes and a lack of genetic information about homarine synthesis and degradation limit our ability to infer the sources and sinks for this abundant compound. The high concentration and diel dynamics of homarine call for further investigation.

Both isethionic acid and DHPS are associated with fast growing eukaryotes that need to mobilize cellular machinery to transport materials into the mitochondria for respiration [32, 133], and recent work has suggested that DHPS has potential osmotic capabilities [32]. These two metabolites had large diel oscillations implicating them as temporary stores of energy or intermediates that can be mobilized quickly. Our data implicate SAR11 and Rhodobacteraceae as likely DHPS degraders at Station ALOHA (Figure 3.6), although genes for the production of DHPS are not in the KEGG database and thus were not identified by our analyses. If production and degradation of these compounds are separated along phylogenic lines [12] then these compounds are likely excreted into the dissolved phase by eukaryotes and subsequently available for use by bacteria, as suggested in Durham et al.

(2019). This may explain the midday maximal expression of a *hpsN*-like Rhodobacteraceae DHPS degradation gene (Figure 6).

Glycine betaine is a osmolyte found within many marine microorganisms, and can play many cellular roles including modulating buoyancy, acting as a methyl donor, and providing a nitrogen source [95, 43, 134]. Glycine betaine had a significant diel oscillation with an evening peak and similar fold change to the total FCM-quantifiable biomass (Figure 3.4). *Chlorophyta*, *Dinophyceae*, and diatoms expressed glycine betaine synthesis genes with diel periodicity while many more groups expressed related genes that lacked diel periodicity (Figure 3.6). It is possible that osmotic balance is maintained in certain phytoplankton by the relatively stable intracellular concentration of glycine betaine while other osmolytes are more dynamic pools with higher daily fold change.

3.4.3 Metabolites as fuel for the microbial loop

Although our data suggests that over diel cycles many metabolites are respired directly by the phytoplankton that produce them, many metabolites quantified here are known to fuel heterotrophic bacterial growth in marine ecosystems [132, 135, 136, 36, 137]. DMSP, for example, can support up to 9.5% of the bacterial carbon demand at Station ALOHA [131]. Additionally, glycine betaine can support heterotrophic bacterial growth, and both natural marine populations and isolated bacteria are known to have high affinity transporters for this compound [46, 138, 139]. The oscillations of particulate metabolite concentrations observed in this study call for further investigation into the hypothesis that these compounds are important substrates for community interactions and resources for the microbial loop. For compounds that exhibited diel oscillations, the difference between the daily maximum and minimum values provides a daily net production and degradation rate. We estimated a total net turnover rate of over 27 nmol C L^{-1} d^{-1} from our targeted metabolites, with several metabolites exhibiting individual turnover rates of over 1 nmol C L^{-1} d^{-1}, including arachidonic acid, trehalose, homarine, sucrose, glycine betaine, glucosylglycerol, and DHPS (Supplemental Table 3.2). These are conservative estimates since the instantaneous flux may

be much higher than the daily net change and we did not measure excretion of metabolites into the dissolved pool. For example, dissolved DMSP has a turnover time of 4.5 hours at Station ALOHA [131] and has been shown to be produced at night and during the day [140], both observations would substantially increase the baseline estimate of DMSP production which does not account for rapid turnover and only includes a daytime increase in intracellular concentration. While the fate of the metabolites measured here remain unclear, conservative estimates of carbon and nitrogen flux through these small pools was large, comprising around 2% of the ^{14}C based estimates of primary productivity during this study[89]. These compounds are potentially used for cellular requirements by the organisms synthesizing them, as discussed above, or released into the labile dissolved pool. When they enter the dissolved pool through excretion or cell lysis, these compounds are important components of the labile dissolved organic matter pool [36] and play a role in organism interactions [141, 142].

3.5 Conclusions

The light-dark cycle plays a dominant role in structuring marine microbial activity. Previous work has shown diel oscillations of community processes, such as daily accumulation and depletion of POC[90], and diel oscillations of transcriptional activity, which have provided new information on temporal dynamics and raise hypotheses about the activity of individual taxa[24, 23]. Measurements of *in situ* metabolites in native planktonic microbial populations reported here support the hypotheses that diverse microbial taxa in the NPSG are synchronized to daily oscillations of light energy and photosynthesis, with metabolites accumulated during the day and depleted at night. The diel oscillations of ubiquitously used primary metabolites shows the extent to which photoautotrophic organisms dominate the community and drive anabolic processes during the day and catabolic processes at night. The combination of transcript abundances, metabolite concentrations, and taxa-specific biomass in the field and in culture allows us to postulate that *Crocosphaera* uses trehalose as a short-term energy source to drive nighttime nitrogen fixation. Trehalose and the other osmolytes

we measured are highly abundant in cells and, in addition to playing multiple roles within producers, likely fuel respiration in heterotrophic bacteria. Metabolite concentrations cannot be predicted from transcripts in a single organism in pure culture, let alone in a complex natural community. Pairing quantitative measurements of particulate metabolites with transcriptomes is a key step toward understanding how regularly oscillating gene expression in microbial communities is reflected in the net community processes we observe and further elucidates the currencies of the microbial community.

3.6 Materials and Methods

3.6.1 Sample Collection

Samples were collected on the R/V Kilo Moana in the NPSG (near 24.5 °N, 156.5 °W) every four hours for two sampling periods in summer 2015 (period one: July 26, 6:00 – July 30, 6:00; period two: July 31 18:00 – August 3, 18:00). To limit variability unrelated to solar forcing, we conducted Lagrangian sampling following two drifters in an anticyclonic eddy[26]. Samples for particulate metabolites and transcripts were collected from 15 m water depth using Niskin bottles attached to a conductivity, temperature, depth array (CTD). Ancillary measurements for nutrients and heterotrophic bacterial abundance (reported in Wilson et. al. 2017) were collected and analyzed with standard Hawaii Ocean Time-series protocols (http://hahana.soest.hawaii.edu/index.html).

3.6.2 Bulk and taxa-specific carbon biomass

POC concentrations were derived from particulate beam attenuation at 660 nm measured via a hyperspectral absorbance and attenuation meter (ac-s, Wetlabs, as published in White et al.[90]). Particle attenuation at 660 nm (c_p 660, m^{-1}) was calibrated against discrete POC samples taken near dawn and dusk ($n = 30$; r of a type II regression = 0.78). Discrete POC and PN samples were collected by filtration of the ship's underway flow through seawater onto combusted GFF filters. Analysis is further described in the supplemental methods.

Continuous underway flow cytometry (SeaFlow) [143] was used to count *Prochlorococcus*, *Synechococcus*, picoeukaryotes (eukaryotic phytoplankton 2–4 μm in size), and *Crocosphaera*. These data were supplemented with discrete flow cytometry sample analysis as in Wilson et al.[26]. Cell diameters of individual cells were estimated from light scatter by the application of Mie theory [144] to a simplified optical model and converted to carbon quotas assuming spherical particles, as described in Ribalet et al. [145]. Carbon biomass was estimated by multiplying cell abundance by carbon quotas.

3.6.3 Metabolite extraction, data acquisition, and processing

Metabolite samples were collected in triplicate at each time point by filtering 3.5 L of seawater onto a 47 mm 0.2 μm PTFE (Omnipore) filter using a peristaltic pump, polycarbonate filter holder, and Masterflex PharMed BPT tubing (Cole-Parmer). Filters were frozen in liquid nitrogen immediately after filtration and stored at -80 °C. Metabolite extractions employed a modified Bligh-Dyer method [22, 32, 70], resulting in aqueous and organic soluble metabolites with heavy stable isotope-labeled extraction and injection internal standards added to both fractions (Supplemental Table 3.5, Supplemental Methods). Unused filters served as methodological extraction blanks.

Metabolomics data were collected by paired liquid-chromatography mass-spectrometry (LC-MS) using both hydrophilic liquid interaction chromatography and reversed phase chromatography with a Waters Acquity I-Class UPLC and a Waters Xevo TQ-S triple quadrupole with electrospray ionization in selected reaction monitoring mode with polarity switching, targeting over 200 compounds [22]. The software Skyline was used to integrate LC-MS peaks [72] and resulting peak areas were normalized to the peak area of internal standards using best-matched internal standard normalization to reduce variability introduced through the extraction and analysis process [22]. A subset of these data are presented in Durham et al. and Muratore et al.[31, 32].

Metabolites with isotopologue internal standards were quantified in all samples (Supplemental Table 3.5). Trehalose, sucrose, and 2,3-dihydroxypropane-1-sulfonate (DHPS)

were quantified with standard additions. For all other metabolites (Supplemental Table 3.2), concentration (pmol L^{-1}) was calculated from injections of known concentrations of authentic standards in both water and a representative matrix to correct for ion suppression. Dimethylsulfoniopropionate (DMSP) loss is known to occur during methanol-based extractions so concentrations are considered a minimum estimate [146]. Details are in the supplemental methods. The amount of each metabolite in each sample is presented in units of metabolite concentration where the amount is normalized to liters of water filtered (nmol metabolite/L) and % POC or % PN where the amount is normalized to POC or PN (% POC and % PN calculated as mol C or N in metabolite/mol C or N in POC or PN x100).

3.6.4 Metatranscriptome data acquisition and processing

Whole community transcript data are referred to here as prokaryotic transcript data, as they were enriched in bacterial and archaeal RNA. These metatranscriptome samples were collected on 0.2 μm filters simultaneously with the metabolomic data reported here, and the sample collection and processing has been described in Wilson et al.[26] and Aylward et al.[86]. Briefly, the metatranscriptome sequence reads were quality trimmed, end-joined, mapped, and quantified with molecular standards. Metatranscriptome sequence reads were aligned to the ALOHA gene catalog [147] using LAST v 959 [148]. Sequence reads were quantified using transcript count normalization, leveraging molecular standards as described in Gifford et al. [149]. Sequence reads were summed if assigned to the same taxonomic order and Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologue [150].

Poly-A+ selected transcript data (referred to here as eukaryotic transcript data) are from the metatranscriptomes presented in Durham et al.[32]. These samples were collected on 0.2 µm filters concurrently with the metabolomic samples and include only the first sampling period. Quality-controlled short reads were assembled using Trinity de novo transcriptome assembler version 2.3.2 [151] on the Pittsburgh Supercomputing Center's Bridges Large Memory system. Parameters include using in-silico normalization, a minimum k-mer coverage of 2, and a minimum contig length of 300. The raw assemblies were quality con-

trolled with Transrate v1.0.3 [152]. To eliminate redundancy and duplication, the assemblies were merged and clustered at the 99% amino-acid identity threshold level with linclust in the MMseqs2 package [153]. Using DIAMOND v 0.9.18 [154], translated eukaryotic contigs were aligned to a reference sequence database of marine organisms that includes peptide sequences from hundreds of marine eukaryotic transcriptomes (MarineRefII reference database, http://roseobase.org/data/, with additions listed in Supplemental Table 3.6). Taxonomy was assigned with DIAMOND by using the top 10% of hits with e-value scores below 10-5 to estimate the Lowest Common Ancestor of each contig. We assigned putative function using hmmsearch (from HMMER 3.1b2 [155], minimum bitscore 30) to find the best-scoring KEGG gene family from KOfam, and linked the specific KO term associated with the KOfam to the contig (ver. 2019-03-20) [156]. Contig abundances were quantified by pseudoalignment of the paired reads to the assemblies with kallisto [157] and normalized to the total number of assigned reads for a taxonomic group. Sequence reads assigned to the same taxonomic group and KEGG ortholog were summed.

Metabolites and transcripts were associated with one another using the KEGG database as a scaffold to match metabolites with transcripts coding for enzymes that directly use or produce those metabolites. The R package KEGGREST [158] was used to access the KEGG database followed by manual curation of these matches.

3.6.5 Detecting Periodicity

Diel periodicity was evaluated for all signals using Rhythmicity Analysis Incorporating Nonparametric Methods (RAIN)[86, 159, 26]. Metabolites and transcripts were considered significantly periodic if they had a false discovery rate (fdr) [83] corrected p-value ¡ 0.05. For each significantly oscillating signal, the time of peak abundance was estimated by fitting a periodic function (supplemental methods), recognizing that the precision of these peak times is limited by sampling resolution. Diel periodicity in metabolites was identified for the two different sampling periods independently and jointly.

3.6.6 Phytoplankton culture conditions

Cultures of phytoplankton were grown in combusted borosilicate tubes in diurnal incubators with a 12:12 light:dark cycle. *Crocosphaera watsonii* strain WH8501 was grown at 27 °C with 50 µmol photons m^{-2} s^{-1} in YBC-II artificial seawater medium [160] supplemented with 0.9 mM nitrate; exponentially growing cells were collected just before the lights turned on and just after the lights turned off. Cells were enumerated via a Beckman Z2 Coulter Counter. *Prochlorococcus* MIT1314 (HLII clade [161]) were grown at 20 °C with 20 µmol photons m^{-2} s^{-1} in Pro99 media [162] prepared with Turks Island Salt Solution and supplemented with 6 mM sterile sodium bicarbonate and 1 mM N-Tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid [163]. *Prochlorococcus* cells were collected 6 hours into the light period during exponential phase and enumerated using the flow cytometer BD Influx cell sorter. Axenicity of *Prochlorococcus* cultures was verified regularly with SYBR-staining and flow cytometry (FCM) and plating on bacterial 1/2 YTSS agar. Samples for metabolomics were collected by gentle filtration onto 0.2 μm hydrophilic polyvinylidene fluoride (Durapore) filters using combusted borosilicate filter towers.

3.7 Acknowledgements

This work was supported by grants from the Simons Foundation (LS Award ID: 385428, A.E.I.; SCOPE Award ID 329108, A.E.I., E.F.D., A.E.W., E.V.A.; SCOPE Award ID 426570, E.V.A.; Award ID 598819, K.R.H.), the National Science Foundation (NSF OCE-1228770 and OCE-1205232 to A.E.I., NSF OCE-160019 to R.D.G, NSF GRFP to A.K.B. and K.R.H., NSF IGERT Program on Ocean Change to A.K.B), and the Gordon and Betty Moore Foundation (Grant #3777 to E.F.D.) The authors acknowledge A. Hynes, N. Kellogg, R. Lionheart, M. Motukuri, and A. Wied for assistance with lab and data analysis; J.S. Weitz and D. Muratore for productive discussions and feedback; J.P. Zehr and M. Hogan for providing *Crocosphaera* WH8501; A. Coe and S.W. Chisholm for providing *Prochlorococcus* MIT1314; the crew and scientific party of the R/V Kilo Moana during HOE-Legacy 2A.

3.8 Data availability

Information for the KM1513/HOE Legacy II cruise can be found online at

http://hahana.soest.hawaii.edu/hoelegacy/hoelegacy.html. Raw sequence data for the diel eukaryotic metatranscriptomes are available in the NCBI Sequence Read Archive under BioProject ID PRJNA492142. Raw sequence data for the prokaryotic metatranscriptomes are available in the NCBI Sequence Read Archive under BioProject ID PRJNA358725. Raw and processed metabolomics data are available in Metabolomics Workbench under Project ID PR000926.

3.9 Tables and Figures

	First Sampling period *Average ± SD (n)*	Second Sampling period *Average ± SD (n)*
Wind speed (kts)	10.59 ± 1.97 (83)	15.15 ± 1.76 (72)
Mixed layer depth (m)	20.92 ± 5.20 (25)	36.12 ± 6.42 (19)
Salinity	35.38 ± 0.01 (62)	35.39 ± 0.00 (33)
Temperature (ºC)	26.81 ± 0.10 (268)	26.86 ± 0.07 (80)
Nitrate + Nitrite (nM)	3.02 ± 0.90 (19)	3.34 ± 1.65 (9)
Dissolved oxygen (μmol/kg)	205.74 ± 0.70 (63)	205.62 ± 0.43 (39)
Particulate organic carbon (μg C l^{-1})	42.28 ± 2.72 (94)	40.70 ± 3.83 (73)
Particulate Nitrogen (μmol N l^{-1})	0.44 ± 0.04 (18)	0.40 ± 0.03 (12)
Heterotrophic bacteria abundance (10^6 cells l^{-1})	508.3 ± 27.3 (22)	534.1 ± 30.8 (47)
Prochlorococcus abundance (10^6 cells l^{-1})	161.23 ± 11.75 (98)	196.38 ± 15.37 (55)
Synechococcus abundance (10^6 cells l^{-1})	0.85 ± 0.07 (98)	0.89 ± 0.06 (55)
Photosynthetic picoeukaryote abundance (10^6 cells l^{-1})	0.97 ± 0.11 (98)	1.10 ± 0.33 (55)
Crocosphaera abundance (10^6 cells l^{-1})	0.16 ± 0.06 (98)	0.31 ± 0.07 (55)

Table 3.1: Wind speed and surface mixed layer physical and biological variables over the two sampling periods. Salinity, temperature, and dissolved oxygen (corrected with bottle measurements) are from the CTD between 13-17 m. N+N and heterotrophic bacteria abundance are measured from discrete samples at 15 m. Particulate organic carbon (from underway beam attenuation), particulate nitrogen, *Prochlorococcus*, *Synechococcus*, photosynthetic picoeukaryotes, and *Crocosphaera* are measured from the ship-underway water intake near 7 m.

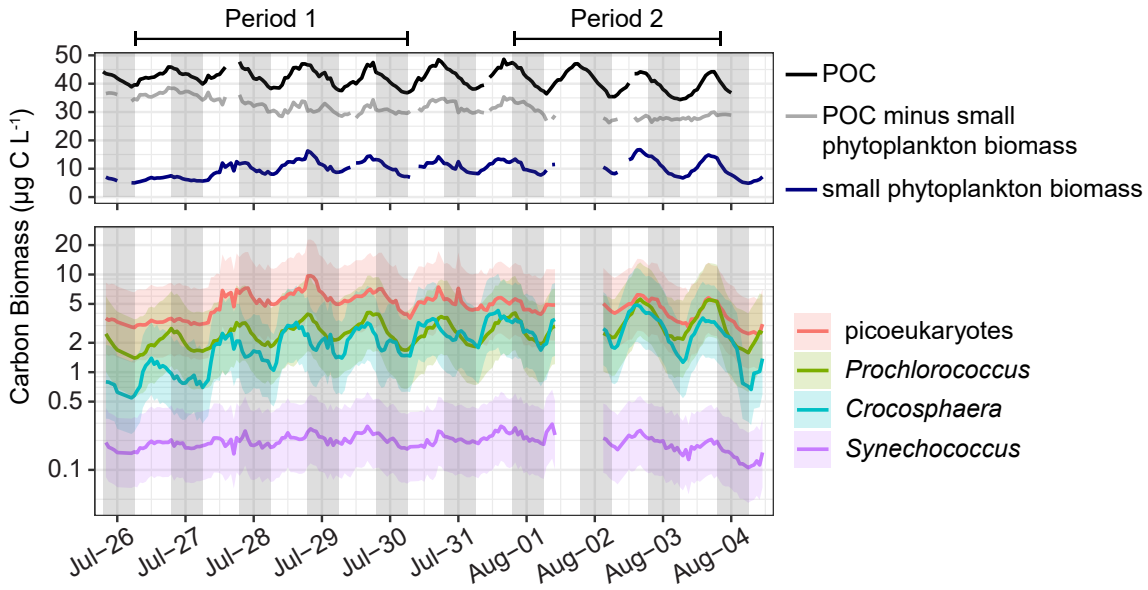

Figure 3.1: Top: Hourly averages of POC from beam attenuation (black line, RAIN fdr-corrected $p < 0.001$), total phytoplankton carbon biomass from flow cytometry (small phytoplankton biomass, navy line, RAIN fdr-corrected $p < 0.001$), and the difference between the two (grey line, RAIN fdr-corrected $p > 0.05$). Bottom: Hourly averages of population specific carbon biomass of *Prochlorococcus*, *Synechococcus*, *Crocosphaera*, and photosynthetic picoeukaryotes (defined here as 2–4 μm) from flow cytometry, with shaded area representing the 95% confidence interval RAIN fdr-corrected $p < 0.001$ for all four populations), note the log10-scaled y-axis. Breaks in the lines are due to short periods of instrument malfunction. The two sampling periods referred to in the text are indicated above the figure.

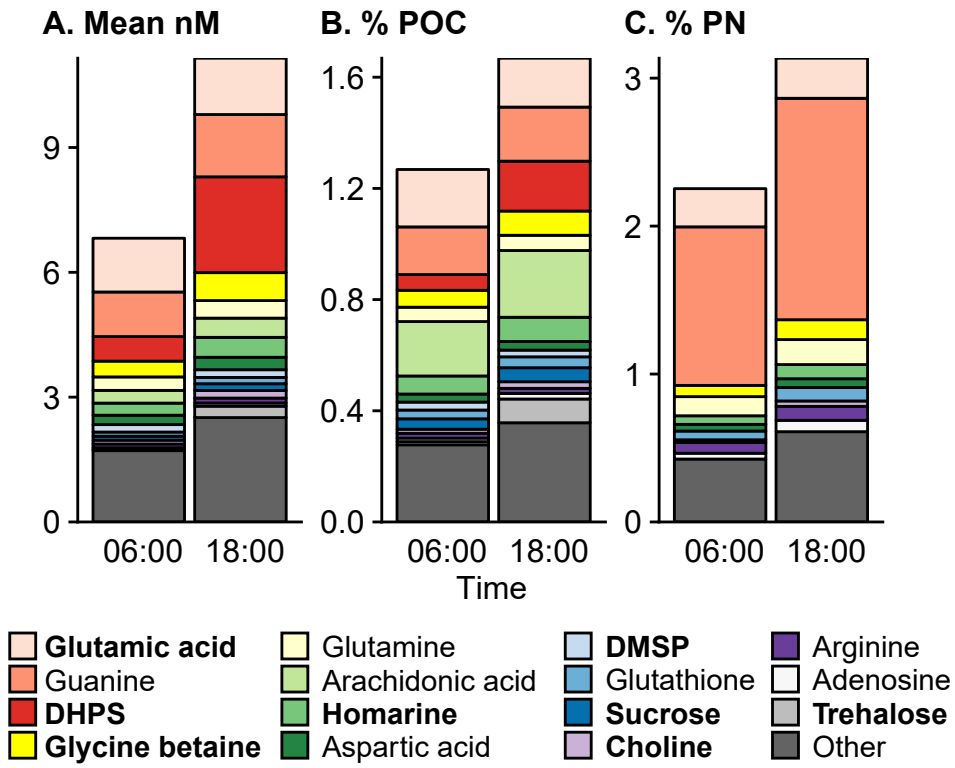

Figure 3.2: Average targeted metabolite composition at dawn (06:00) and dusk (18:00) from July 26th – July 28th (n = 9 for each time point), shown as the estimated particulate metabolite concentration (A), the percent of particulate organic carbon (B), and the percent of the particulate nitrogen (C). "Other" contains the sum of the rest of the metabolites (64 compounds). Osmolytes are in bold. Metabolites are arranged according to their average molar concentration at 6:00. Note the different y-axis scales.

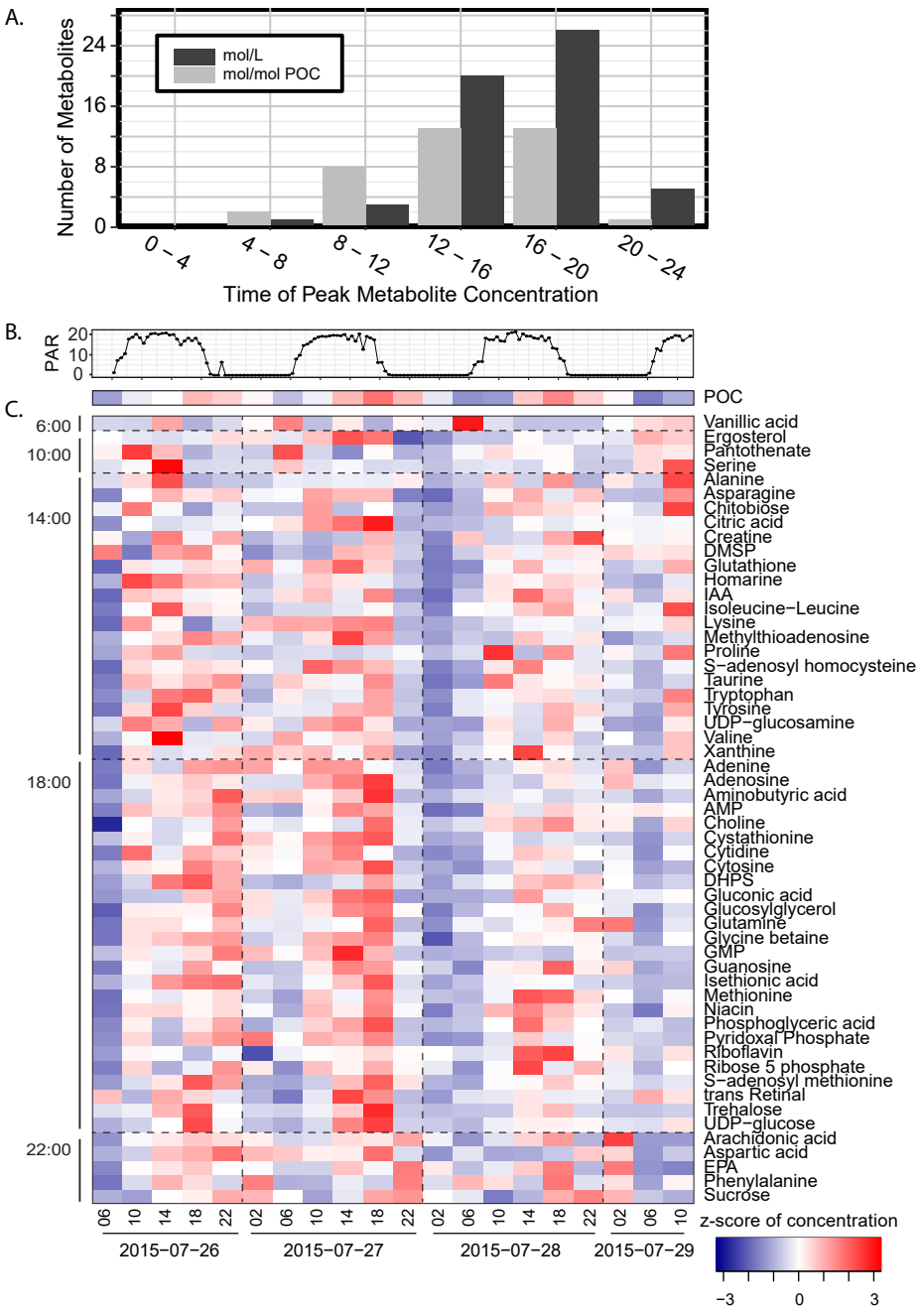

Figure 3.3: Time of day that significantly diel compounds peak in the first sampling period (A). Surface light (photosynthetically active radiation, PAR, x10 nmol photon m-2 s-1) (B). Heat map showing the z-score standardized concentrations of POC and of metabolites (nmol L-1) determined to be significantly diel in the first sampling period, arranged by time of peak concentration (C).

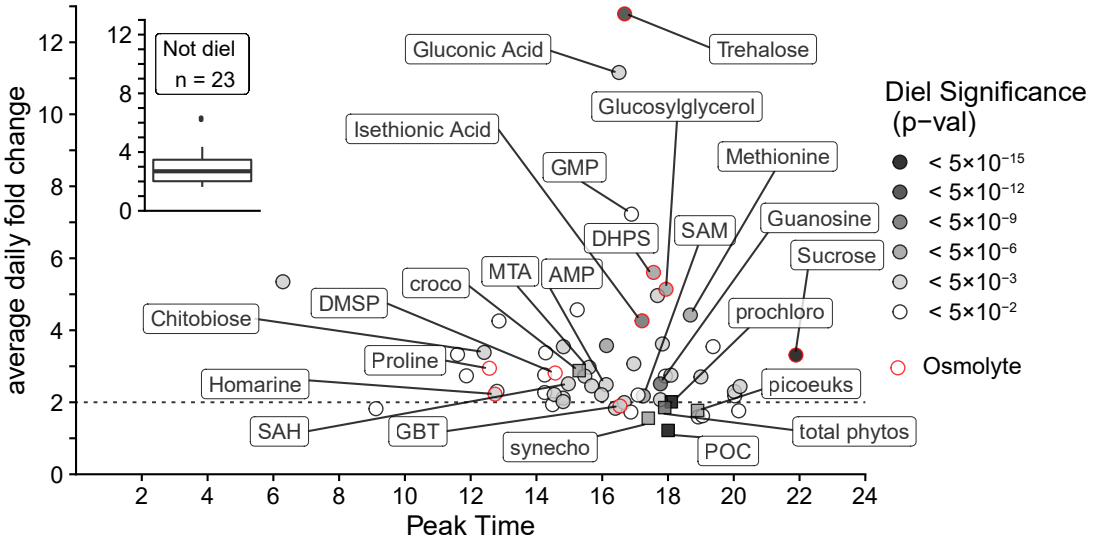

Figure 3.4: Peak time vs average daily fold change for each metabolite (circles, nmol L-1), POC from beam attenuation and phytoplankton biomass from flow cytometry (squares, µg C L-1). Grey color indicates the level of significance (fdr corrected p-value) of the 24-hour oscillation. Red outlines indicate that the compound is an osmolyte. Select compounds and all biomass estimates are labeled (croco = *Crocosphaera*, synecho = *Synechococcus*, prochloro = *Prochlorococcus*, picoeuks = photosynthetic picoeukaryotes, total phytos = total phytoplankton biomass from underway flow cytometry). Dashed line is at a 2-fold change, which is above that for POC and total picophytoplankton biomass. The inset shows the distribution of fold-change in non-significant compounds. These compounds have variability even though they do not have 24-hour periodicity.

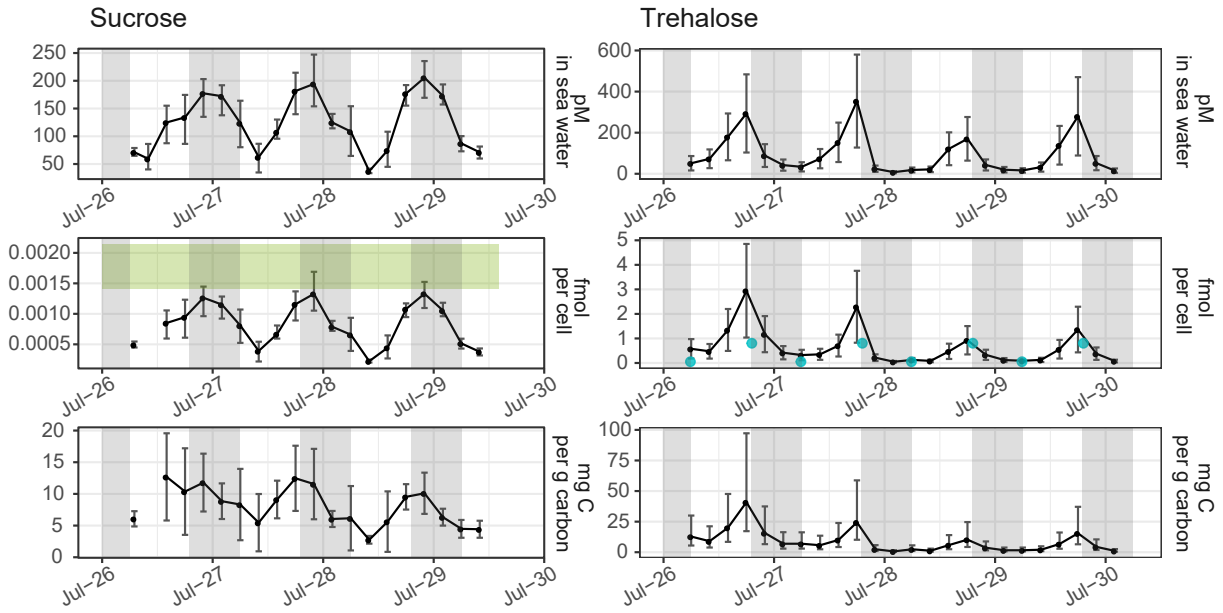

Figure 3.5: Particulate sucrose (left) and trehalose (right) measured as pmol L-1 in seawater (top), fmol cell-1 (middle) of *Crocosphaera* and *Prochlorococcus* for trehalose and sucrose, respectively, and mg g-1 cell carbon (bottom) of *Crocosphaera* and *Prochlorococcus* for trehalose and sucrose, respectively. The light grey vertical shading represents nighttime. The green box in the middle-left panel indicates the range of cellular sucrose quotas measured in *Prochlorococcus* MIT1314 harvested mid-day in exponential growth. The blue points in the middle-right panel indicate the dawn and dusk values measured for trehalose quotas in *Crocosphaera watsonii* WH8501. In the top panels, the error bars represent one standard deviation around the mean value, including uncertainty from the quantification regression. The error bars in the middle panels represent one standard deviation around the mean. The error bars in the bottom panels represent the 95% confidence interval given the confidence in the biomass quantification from underway flow cytometry.

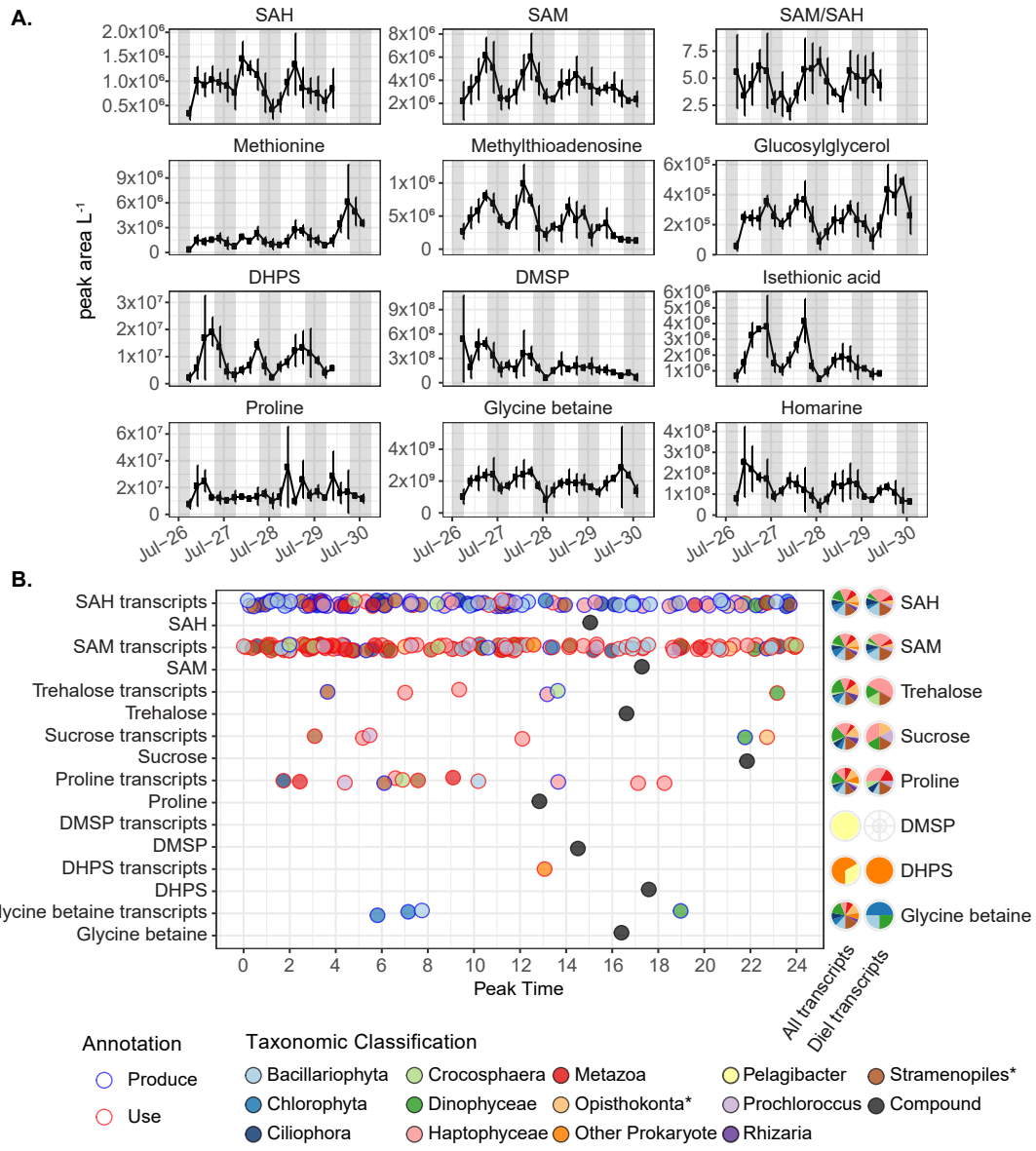

Figure 3.6: A) Diel metabolite concentrations (peak area L-1, proportional to nmol L-1) of methionine cycle compounds, methylthioadenosine, and osmolytes. Error bars are the standard deviation of biological triplicates. The light grey vertical shading represents nighttime. B) Left: Time of peak abundance of diel transcripts related to the production or use of select diel osmolytes and primary metabolites. Fill color indicates the phylogenetic lineage of the transcript; outline color indicates whether the transcript is associated with production or consumption of the metabolite. Time of metabolite peak concentration (nmol L-1) is in black. Right: Proportion of all transcripts and diel transcripts belonging to each taxon. * = does not include select subgroups shown otherwise.

3.10 Supplemental Methods

3.10.1 Metabolite sample extraction

Metabolites were extracted as in as in Boysen et al.[22]. Briefly, frozen filters were cut into small pieces and put into bead beating tubes with silica beads, heavy isotope-labeled internal standards (Supplemental Table 3.5), and cold aqueous (50:50 methanol:water) and organic solvents (dichloromethane). The samples were shaken on a FastPrep-24 Homogenizer for 30 seconds and chilled in a -20 °C freezer, repeated for three cycles. The organic and aqueous layers were separated by spinning samples in a microcentrifuge at 5,000 rpm for 90 seconds at 4 °C. The aqueous layer was removed to a new combusted glass centrifuge tube. The remaining organic fraction was rinsed three times with 50:50 methanol:water. All aqueous rinses were combined with the original aqueous extract and dried down under N2 gas. The remaining organic layer was transferred into a clean glass centrifuge tube and the bead beating tube was rinsed two more times with cold organic solvent. The combined organic rinses were centrifuged, transferred to a new tube, and dried under N2 gas. Dried aqueous and organic fractions were re-dissolved in 400 µL of water and 400 µL of 1:1 water:acetonitrile, respectively. Isotope-labeled injection standards were added to both fractions (Supplemental Table 3.5).

3.10.2 Metabolite data acquisition

LC-MS parameters were as in Boysen et al.[22]. Data for most compounds were collected on a Waters Xevo TQ-S triple quadrupole (TQS), but for compounds outside the linear range or which were originally detected in sub-optimal polarity, a Thermo QExactive HF (QE) mass spectrometer with ESI was used. Chromatography and mass spectrometry methods follow those reported in Boysen et al.[22]. The organic fraction was analyzed using reversed phase chromatography (Waters Acquity UPLC HSS Cyano column, 1.8 µm particle size, 2.1 mm x 5 mm) and the aqueous fraction was analyzed with both reversed phase and hydrophilic liquid interaction chromatography (HILIC - SeQuant ZIC-pHILIC column, 5 mm particle

size, 2.1 mm x 150 mm, from Millipore). We monitored 101 compounds with reversed phase and 110 compounds with HILIC.

Data collected on a Thermo QExactive HF (QE) with ESI was used for nucleic acids and nucleosides (adenine, guanine, thymine, cytosine, adenosine, guanosine, thymidine, cytidine) and compounds which were overloaded on the triple quadrupole (glycine betaine, homarine, phenylalanine, DMSP) but which were within the linear range on the QE. On the QE, for HILIC, a full scan method with polarity switching was used with 60,000 resolution. For RP, positive ionization mode was used with a resolution of 120,000. For the QE data, proteowizard was used to convert .raw files to .mzxml [73]. Peak integrations were performed using Skyline for small molecules [72]. Isoleucine and leucine did not always chromatographically separate and were treated as a single metabolite, (iso)leucine. Data were subjected to in-house quality control (QC) that removed misidentified compounds, removed compounds with a low signal to noise ratio (S/N < 4), and flagged compounds that were detected in the blanks (signal intensity in the sample must be greater than 3 times that in the blank). Compounds which were below the level of the blank in greater than 30% of samples (40 out of 132) were discarded. Compounds which did not pass the QC for 6 or more samples were discarded. For compounds which had fewer than 6 samples fail the QC, if the compound was detected in a sample but comparable to the blank, that data was retained. If the signal in a sample was less than the signal in the blank, a value equal to half the average blank value was filled back in to reflect the limit of detection.

3.10.3 Metabolite data curation

The normalization procedure for metabolomics data follows the Best-Matched Internal Standard method described in Boysen et al.[22]. Pooled samples at full and half strength (diluted 1:1 with water, for the aqueous fraction, or solvent, for the organic fraction) were run after the samples were run in order to train the normalization algorithm. Normalization to internal standards was only used when it resulted in an improvement (decrease) of 10% in the relative standard deviation (RSD) of the compound peak area across multiple injections of

the pooled sample. Compounds with a raw RSD of less than 0.01 in the pooled sample were not normalized because the chances of noise reduction were outweighed by the likelihood of introducing unwanted variability. The low cutoff was used to encourage normalization because the data were collected over a period of five weeks of instrument time; normalizing to an internal standard is an improvement over the instrument variability that we expect to influence mass spectral measurements over this time period. All QE Orbitrap data were either used as raw data (no normalization) or, in the case of phenylalanine, normalized to its isotopologue internal standard, as the high resolution instrumentation experiences less ion suppression compared to that on the TQ-S instrument(1). Prior to normalization, the peak areas of the internal standards in each sample were assessed to detect run quality and extraction efficiency. This assessment showed that a number of samples, mostly those collected as part of the second sampling period, clearly had half or twice the appropriate standard concentration added. (samples with internal standards adjusted in the 1st sampling period: 5R1, 6R1, second sampling period: 26R2, 26R3, 27R1, 27R2, 27R3, 28R1, 28R3, 29R2, 30R1, 31R1, 32R3, 34R1, 34R3, 35R1, 35R2, 37R1, 39R1). Peak areas were adjusted according to these observations as well as laboratory notebook records for the samples which had incorrect concentrations. The periodicity analysis was done both with and without these replicates and the results are robust to this change. Additionally, samples collected between July 29th 14:00 and July 30th 02:00 did not have some of the standards added during sample processing. Data from the effected compounds were removed for those samples as normalization was not possible (as indicated in Supplemental Table 3.2). Euclidean distance of samples based on z-score standardized metabolite profiles showed that single replicates from two different samples (31R2, 41R1) in the second sampling period were outliers (> 3 standard deviations away from the mean average distance), so these samples were removed prior to further analysis.

3.10.4 Quantification of select metabolites

Metabolites with isotope-labeled authentic standards (see list in Table 3) were quantified using the following formula:

$$[M]_{SW} = \frac{(PkA_M)}{(PkA_{IS})} * [IS]_{vial} * \frac{V_{vial}}{V_{filtered}} \tag{3.1}$$

where M indicates metabolite, V indicates volume, IS indicates internal standard, SW indicates seawater, and PkA is the integrated LC-MS peak area.

Isotope-labeled standards for trehalose and sucrose were purchased after the full sample set had been processed. A subset of samples ($n = 19$) was spiked with the internal standards and re-analyzed on the LC-TQS-MS, concentrations were determined as above, and a linear regression relating peak area to trehalose and sucrose concentration was fitted according to the formula:

$$[M]_{vial} = A * (NormalizedPkA) \tag{3.2}$$

where A is a constant. This regression was used to calculate the concentrations of trehalose and sucrose in the rest of the samples that were not reanalyzed.

An authentic standard for 2,3-dihydroxypropane-1-sulfonate (DHPS) was obtained after the full sample set had been processed and the quantification of this compound using standard additions is described in Durham et al. [32].

3.10.5 Estimated Concentrations of Metabolites

Approximate concentrations correcting for ionization efficiency and ion suppression were calculated using the equation:

$$Concentration = \frac{Area}{IE} * \frac{V_{vial}}{V_{filtered}} * \frac{1}{RF_{ratio}} \tag{3.3}$$

where IE is the ionization efficiency, calculated by taking the average peak area of a standard injected in water divided by the concentration of the standard. RF_{ratio} is the response factor ratio published in Boysen and Heal et al. [22], calculated by taking the ratio of the peak area of a standard injected in environmental matrix (less the ambient matrix signal, if applicable) to the peak area of the standard injected in water. Published values were used with the exceptions of taurine, chitobiose, and for nucleosides and nucleotides where we used values calculated from a different sample set also from near Station ALOHA, so with similar matrix effects.

We did not make estimates of concentration for five compounds: methylthioadenosine (MTA) in the aqueous fraction because we didn't have appropriate standards at the time of sample analysis, EPA, DHA, ergosterol, and trans retinal because insufficient isotope labeled internal standards were used for the lipid fraction.

3.10.6 Particulate carbon and particulate nitrogen discrete sample analysis

Samples for particulate nitrogen and particulate carbon were collected form the ship's underway flow through seawater, which sits at ∼7 m water depth. Samples were collected onto combusted GFF filters. Filters were folded and stored in combusted aluminum foil at -80 °C until analysis, when they were thawed and dried overnight at 60 °C, balled into Ag and Sn boats, and analyzed using high-temperature combustion (1020 °C) on a ThermoQuest NC 2500 elemental analyzer.

3.10.7 Eukaryotic metatranscriptome reference database

We supplemented the curated MarineRefII reference database (http://roseobase.org/data/) with additional representatives of marine animal, fungal, protist and viral reference sequences (Supplemental Table 3.6), totaling 641 marine eukaryotes and prokaryotes.

3.10.8 Detecting periodicity and estimating time of peak concentration/abundance

When conducting the RAIN analysis outliers removed during quality control were filled back in by taking an average of the other two replicates collected at that time point. The time of peak concentration for each oscillating metabolite and peak abundance for each oscillating transcript was calculated by fitting a periodic oscillator according to the function:

$$[M] = A * cos(\frac{2\pi}{24} * HourCollected) + B * sin(\frac{2\pi}{24} * HourCollected) \tag{3.4}$$

as in Ottesen et al. [23]. The lag time between metabolites and transcripts coding for proteins that use or produce them was estimated simply by taking the difference between the two peak times, in hours.

3.10.9 Multivariate analysis

Metabolite concentrations were standardized using z-scoring by subtraction the mean and dividing by the standard deviation of each metabolite across all samples, having a mean of zero and standard deviation of one. Samples and metabolites were clustered based on Euclidean distances with the vegdist and metaMDS functions from the vegan R package (version 2.5-6)[164]. Average-linkage clustering was calculated using the hclust function from the stats R package (version 3.6.1).

We used nonmetric multidimensional scaling (NMDS) to explore differences in samples based on their metabolite profiles. NMDS is an ordination technique that uses rank-order similarity between samples rather than absolute differences between samples to reduce dimensionality in the data [165]. NMDS using a Euclidean distance matrix was chosen rather than principal components analysis (PCA) because NMDS relaxes the assumption that there are fewer variables than samples, which is not the case in metabolomics where data on the abundance of hundreds of compounds is collected for each sample. Additionally, NMDS avoids the assumption of linear relationships among variables.

NMDS was run with two ordination axes and 100 random starts. Significance of the stress

value was tested with a Monte Carlo randomization test. Goodness of fit was assessed by correlating NMDS ordination results and Euclidean distances using both a non-metric and linear fit. Analysis of Group Similarities (ANOSIM) was used to test differences between the times of day that samples were collected as well as differences between the days of collection. ANOSIM was conducted on a Euclidean distance matrix with 1000 permutations. Pairwise ANOSIM was conducted to further clarify which times were significantly different.

3.10.10 Phytoplankton culture growth conditions

Cultures of *Crocosphaera watsonii* WH8501 were obtained from the Zehr lab at the University of California, Santa Cruz. They were grown with a square 12:12 light:dark cycle with 50 μmol photons m^{-2} s^{-1}, at 26 °C, and had an average growth rate of 0.17 d^{-1}. Cells were collected in mid-to-late exponential phase by gentle vacuum filtration onto 0.2 μm Omnipore filters using combusted borosilicate filter towers. Cells were enumerated using a Beckman coulter counter.

3.10.11 Mixed layer depth

Mixed layer depth was calculated from CTD profiles and is defined by a 0.03 kg/m^3 density offset from 10 db.

3.10.12 Supplemental calculation: Trehalose fueling nitrogen fixation

To convert half a mol N_2 to one mol NH3 requires 3 mol electrons and 8 mol ATP[125]. One mol glucose can provide 24 mol electrons or 30–36 mol ATP, leading to a stoichiometry of between 2.08–2.35 moles CH_2O to produce sufficient ATP and electrons to produce one mole of fixed NH_3 [122]. From Wilson and Aylward et al. Crocosphaera is responsible for 7.3 nmol N l^{-1} d^{-1} [26], which therefore requires between 15.2 and 17.2 nM of respired carbon. Our data has a range of 1.6 to 4.3 nM C in the form of trehalose drawn down every night, with a mean drawdown of 3.2 nM C in the form of trehalose. Using this range of required

carbon and trehalose drawn down, we estimate that trehalose catabolism could fuel between 9% and 28% of N2 fixation, with a mean value of 20%.

3.11 Supplemental Results and Discussion

3.11.1 Multivariate Analysis

NMDS analysis of the samples produced a low stress value of 0.18 which was significant ($p < 0.01$), indicating that the sample scores are robust. The fit between the ordination distance and Euclidean distance had a non-metric R^2 value of 0.97 and a linear fit R^2 value of 0.88. NMDS analysis of the samples from the second sampling period (Supplemental Figure 3.7) produced a low stress value of 0.17 which was significant ($p < 0.01$), indicating that the sample scores are robust. The fit between the ordination distance and Euclidean distance had a non-metric R^2 value of 0.97 and a linear fit R^2 value of 0.89. The overall similarity observed at 6:00 during the first collection period is not seen during the second collection period. NMDS and ANOSIM analysis of the samples from the second collection period and full dataset are unable to discriminate between samples collected at different times of day (Supplemental Figure 3.7), providing additional evidence that community synchrony, as illustrated by overall metabolite composition, weakened as sampling progressed.

3.11.2 Metabolites lose diel periodicity in second sampling period

Fewer metabolites had diel periodicity in the second sampling period (Supplemental Figure 3.12), with 9 and 23 compounds exhibiting diel periodicity when analyzed as molar concentrations (nmol L^{-1}) and normalized to POC (nmol μmol POC^{-1}), respectively (Supplemental Table 3.3, Supplemental Figure 3.12). The reduction of diel oscillation in the second sampling period is seen in multivariate analyses as well, where in the second sampling period no sampling times clustered significantly (Supplemental Figure 3.7).

Because these samples were processed over several weeks, it is possible that methodological issues arose and masked the diel oscillations during the second half of the sample set.

We deem this unlikely because we still measure strong diel oscillations in some compounds over the course of the whole sample period. Overall there was not a drop in signal intensity from the instruments over the sample processing, and the increase in *Prochlorococcus* concentrations shows, if anything, that there would be more biomass rather than less biomass to produce a robust signal for the latter half of the data (Figure 3.1).

3.12 Supplemental Files

- Supplemental Tables 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9. (XLSX).

3.13 Supplemental Tables

Table 3.2: Metabolites measured in this analysis. The average fold change from peak to trough, the maximum and minimum estimated or absolutely quantified values (pmol L^{-1}), and whether the compound oscillates with 24-hour periodicity when calculated as molar concentration in seawater (water), when calculated relative to POC (POC), both (Both), or neither (None) for the first sampling period analyzed independently, second sampling period analyzed independently, and full dataset. The time of peak concentration for these concentration values in the different time periods are provided in the final columns, rounded to the nearest hour. The net flux through the particulate pool calculated by the mean daily swing from max to minimum. * indicates metabolites for which samples 21-24 are removed and for which 6 samples in the second diel sampling period maybe affected by internal standard adjustments. + indicates metabolites for which 4 samples in the second sampling period might affected by IS adjustments. † notes that concentrations for DMSP are likely underestimates, as described in the methods.

Table 3.3: Average and standard deviation of targeted metabolite composition at dawn (06:00) and dusk (18:00) from July 26th – July 28th (n = 9 for each time point), as the estimated particulate metabolite concentration, the percent of particulate organic carbon, and the percent of the particulate nitrogen.

Table 3.4: Pairwise comparisons of samples collected at different time points from the multivariate analyses of particulate metabolite concentration during the first sampling period.

Table 3.5: Internal Standards added before extraction (Exr Standard) or before injection (Inj Standard).

Table 3.6: Transcriptomes used to supplement the MarineRefII reference database (http://roseobase.org/data/).

Table 3.7: Particulate metabolite concentrations in normalized peak area per L of seawater filtered. Across a single metabolite these values are proportional to molar concentration. Values should not be quantitatively compared between two metabolites, since the ionization efficiency and matrix effects influence different metabolites differently such that the same concentration can result in difference in peak area.

Table 3.8: Prokaryotic transcripts that matched metabolites identified by organism taxa and KEGG ortholog. If the transcript is significantly diel (RAIN fdr-corrected p-value < 0.05) the time of peak transcript abundance is provided (0/24 is midnight, 12 is noon).

Table 3.9: Eukaryotic transcripts that matched metabolites identified by organism taxa and KEGG ortholog. If the transcript is significantly diel (RAIN fdr-corrected p-value < 0.05) the time of peak transcript abundance is provided (0/24 is midnight, 12 is noon).

3.14 Supplemental Figures

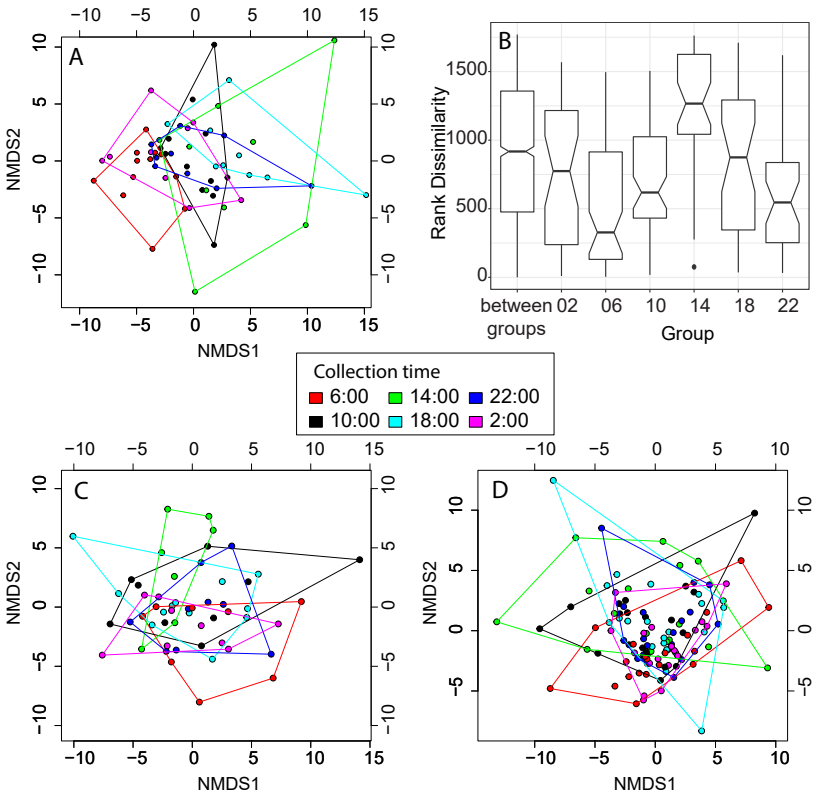

Figure 3.7: Multivariate analyses based on standardized particulate metabolite concentration (proportional to nmol L^{-1}). A) NMDS of the first sampling period alone: Jul-26th - Jul 30th. The NMDS analysis results were significant (Monte Carlo randomization textitp < 0.01) with a stress value of 0.18. B) Within and between group variability from ANOSIM analysis using standardized particulate concentrations of all metabolites (nmol L^{-1}) from the first sampling period (R = 0.194, textitp < 0.001). C) NMDS of the second sampling period alone: Jul 31st – Aug 3rd. The NMDS analysis results were significant (Monte Carlo randomization p < 0.01) with a stress value of 0.17. D) NMDS of full dataset: Jul-26th – Aug-3rd. Colors indicate time of day that the samples were collected. The NMDS analysis results were significant (Monte Carlo randomization $p < 0.01$) with a stress value of 0.18.

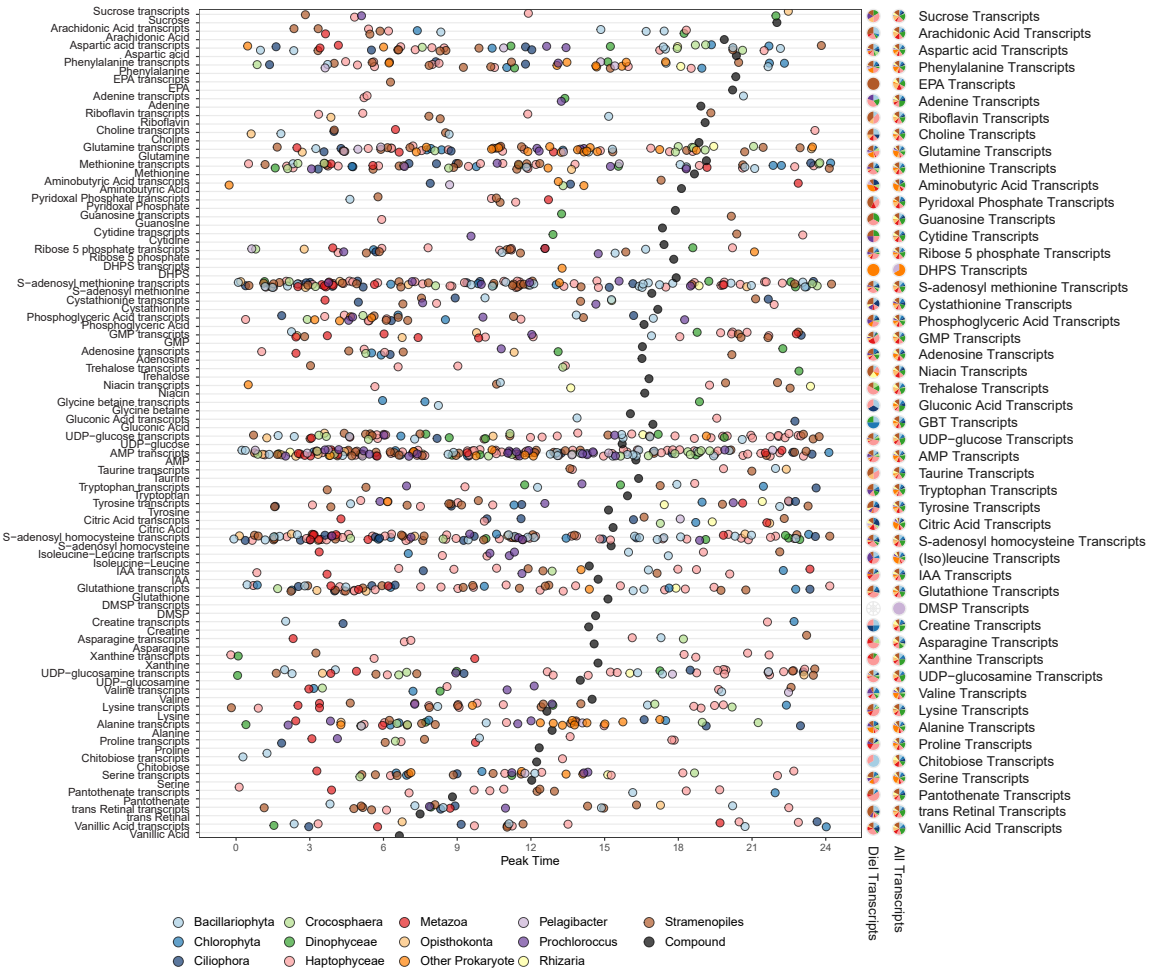

Figure 3.8: Diel transcript peak abundance related to the production or degradation of diel metabolites. Color indicates the phylogenetic lineage of the transcript. Left: Peak time of transcript abundance or particulate metabolite concentration (nmol L^{-1}). Right: Proportion of diel transcripts belonging to each taxa and proportion of all transcripts, regardless of diel oscillation, related to each metabolite belonging to each taxa.

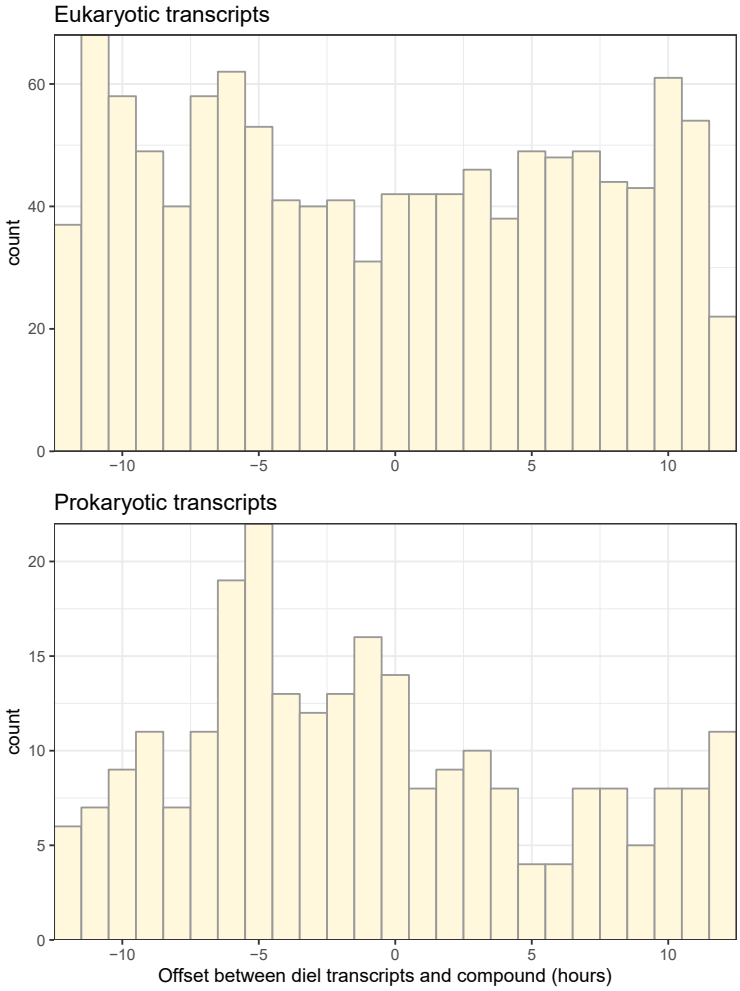

Figure 3.9: Offset time (in hours) between the diel compounds and diel eukaryotic transcripts (top) or diel prokaryotic transcripts (bottom) that use or produce them. Diel significance of compounds was based on the first sampling period, diel significance of eukaryotic transcripts was based on the first sampling period, diel significance of the prokaryotic transcripts was based on both sampling periods (RAIN fdr-corrected $p < 0.05$).

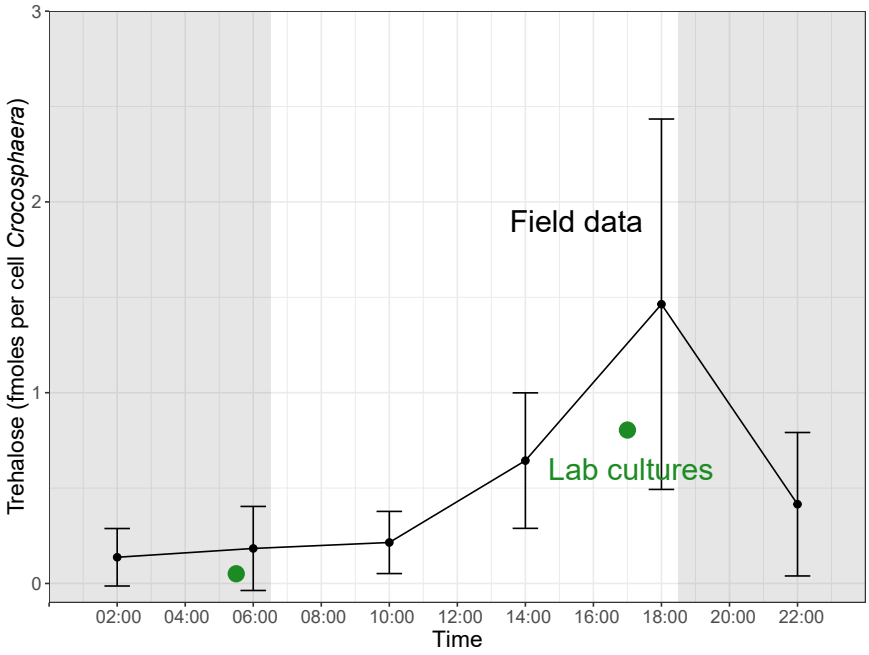

Figure 3.10: Field and culture particulate trehalose concentrations presented relative to *Crocosphaera* cell count. Field data (black points) show the average and standard deviation at each time point over the full sampling period. Lab cultures (green circles) represent the values for the cultures harvested at dawn and dusk. Variability in technical replicates (for dusk) and biological duplicates (for dawn) are smaller than the points.

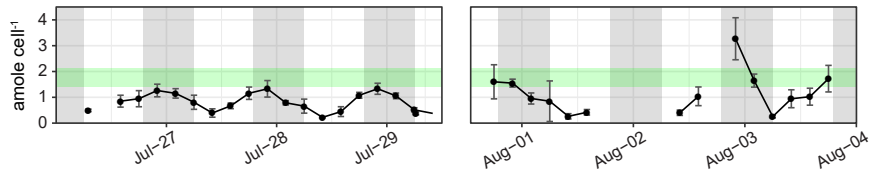

Figure 3.11: Field and culture sucrose per cell *Prochlorococcus*. Field data (black points) show the median and range at each time point. The green box shows the maximum and minimum values of sucrose in triplicate axenic cultures of *Prochlorococcus* MIT1314 harvested at mid-day in exponential growth.

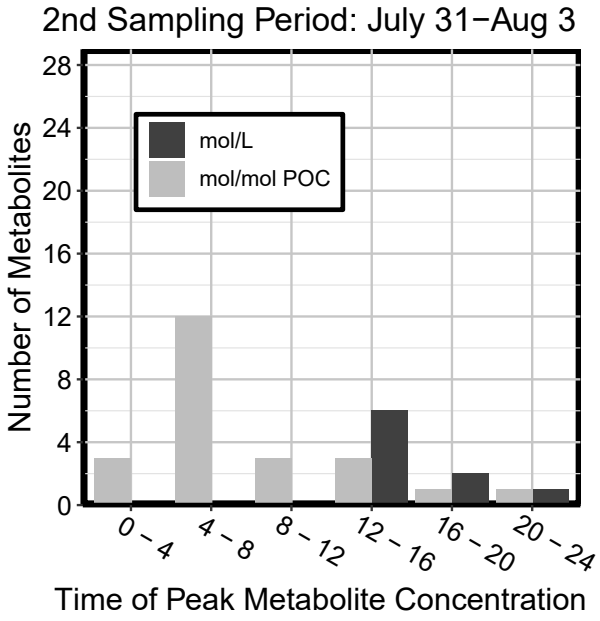

Figure 3.12: Time of day that compounds peak in the second sampling period.

Chapter 4

PHYTOPLANKTON COMMUNITY METABOLISM AND NUTRIENT LIMITATION IN THE NORTH PACIFIC TRANSITION ZONE: POLAR METABOLITE CHANGES REFLECT COMMUNITY COMPOSITION AND PHYSIOLOGY

4.1 Abstract

Nutrient limitation and co-limitation play a key role in determining marine microbial community structure and activity. Metabolites, the small organic molecules resulting from cellular activity, can be used as tracers of cell taxonomy and metabolism. Here we measure intracellular metabolite pools across a gradient of macro- and micro-nutrients between the North Pacific subtropical gyre and the North Pacific subpolar gyre as well as in nutrient amendment experiments. These measurements reveal how the genetic potential of a microbial community is manifested *in situ* and in response to changes in nutrient ratios. Chlorophyll concentrations in the nutrient amendment experiments show the dominance of nitrogen limitation throughout this region and the potential for iron and nitrogen co-limitation near the subtropical chlorophyll front. Metabolite pools, community composition from flow cytometry, and 16S and 18S rDNA gene sequences show that organisms from the microbial communities in the North Pacific subtropical gyre and the North Pacific transition zone have different responses to nutrient amendments. Drawing on the metabolite data from the transect, the incubation experiments, and from laboratory cultures of phytoplankton, we identify metabolites, including several acylcarnitines and potential mycosporine-like amino acids, that reflect microorganisms' physiological adaptations to nutrient stress relief. These data provide a unique opportunity to evaluate the importance of community composition and metabolic flexibility in determining community metabolism. Compounds that change across this transect and

in response to nutrient amendments are key targets for future investigations as potentially important players in microbial physiological adaptation and biogeochemical cycling.

4.2 Introduction

Phytoplankton primary production is the foundation of marine biogeochemical cycles and the marine food web. In vast regions of the ocean this productivity is limited by the availability of nutrients supplied to the surface ocean [33]. The low latitudes of the subtropics are classically considered to be nitrogen limited. In these regions, the supply of nitrogen from below the photic zone is slow, phytoplankton are adapted to low-nitrate levels [166], and nitrogen-fixing phytoplankton can be an important source of new nitrogen to the surface ocean [167]. Iron is often the limiting nutrient for phytoplankton growth in regions with higher nutrient supply from depth and where aerosol deposition and continental sources of iron are small [33, 168]. These regions, known as high-nutrient low-chlorophyll regions include the subarctic gyres, eastern equatorial Pacific, and Southern Ocean. Other nutrients besides nitrogen and iron can be limiting, including phosphorus, trace metals such as cobalt, and micronutrients such as B-vitamins [169, 35, 170]. These nutrients, however, are often co-limiting or secondarily limiting (serial-limiting), while iron and nitrogen are the most common primary limiting nutrients. The boundaries of nitrate and iron limited regimes are locations where iron and nitrate can be co-limiting [35, 171], with the magnitude of delivery and ratio of these two nutrients together controlling production and facilitating greater microbial diversity with an increased number of ecological niches [172].

Nutrient limitation in surface ocean marine plankton communities can be assessed using a variety of tools, but most commonly is assessed with nutrient amendment incubations. In these experiments, an increase in total chlorophyll following the addition of a nutrient is considered evidence for limitation by that nutrient. However measuring chlorophyll alone does not reveal other potential changes such as changes in phytoplankton community or physiological responses. Here we use metabolomics measurements to gain a chemical understanding of microbial community composition and physiology during nutrient amendment

incubations with microbial communities in the North Pacific. Metabolites are the small molecules (< 800 da) that all organisms use and produce through cellular metabolism. Intracellular metabolites have taxonomic specificity and the ability to shift rapidly when cells experience changes in their environment, such as light levels, nutrient stress, or changes in the surrounding microbial community [41, 100, 128, 173, 174, 175, 176, 177].

The North Pacific transition zone (NPTZ) is a region between the North Pacific subtropical gyre (NPSG), a nitrate limited region, and North Pacific subarctic gyre, an iron limited region. The NPTZ is a hot spot of biological productivity and carbon sequestration compared to the gyres on either side [178, 179] and is characterized by fronts of chlorophyll, salinity, and temperature that migrate seasonally [34]. The microbial community composition changes throughout this region as well, with cyanobacteria numerically dominating the phytoplankton community in the NPSG and eukaryotic algae becoming dominant in the north [34, 180, 41]. The NPTZ is not a homogenous region. It has many fronts within it that mark transitions in physical and chemical parameters and unique biological communities. The macro and micronutrients delivered to the North Pacific have different sources (N: Ekman transport, Fe: aerosols) and thus their delivery can be decoupled on seasonal time scales [34, 181, 182]. The microbial communities respond to the seasonal changes in light, macronutrient supply, and micronutrient supply. These physical and chemical dynamics make the NPTZ an ideal location to study the ways in which microbial communities respond to a change in nutrient supply.

Here I ask what nutrients limit phytoplankton growth in the NPTZ using nutrient amendment incubations at three different locations within this region. We examine the community structure and microbial metabolisms *in situ* and after nutrient amendment incubations using 16S and 18S amplicon sequencing, flow cytometry, and targeted and untargeted metabolomics. We show that phytoplankton growth throughout the NPTZ is primarily nitrogen limited with some evidence for iron colimitation in the central portion of the NPTZ. We identify changes in select metabolites as potential markers for microbial nitrogen limitation and stress relief. On the whole, however, there is no unified metabolic response to nitrogen

addition. Instead, we see organisms responding divergently and metabolites reflecting the growth and physiology changes of specific organisms.

4.3 Materials and Methods

4.3.1 Sampling strategy and biogeochemical data

Samples were collected on the Gradients 2 cruise aboard the R/V Marcus G. Langseth from May 27 to June 13, 2017. The cruise transited along 158 °W and seawater samples were collected from approximately 25.7 °N to 41.4 °N (Figure 4.1A). Continuous underway measurements were collected for flow cytometry using a SeaFlow instrument to enumerate *Prochlorococcus*, *Synechococcus*, and picoeukaryotes [143]. Additional discrete samples were collected and used to calibrate the underway measurements, as reported in Gradoville et al. [180]. Surface seawater samples for nitrate plus nitrite (DIN), soluble reactive phosphorus (SRP, dominated by phosphate), trace metal concentrations, particulate carbon (PC) and particulate nitrogen (PN) were collected and measured as reported in Gradoville et al. [180]. Sea surface temperature, salinity, and chlorophyll fluorescence were measured using the CTD sensors for depth profiles and the ship's uncontaminated underway water system for underway measurements.

4.3.2 Nutrient amendment incubations

Nutrient amendment incubations were conducted at three different stations along the cruise transect that represented the northern transition zone (Station 7, 41.4 °N, Experiment 1), the mid-transition zone (Station 11, 37 °N, Experiment 2), and southern transition zone (Station 16, 32.9 °N, Experiment 3) (Figure 4.1). Mixed layer (10–15 m) seawater for each experiment was collected with a trace metal clean surface tow fish system modified from Bruland et al. and described in Gradoville et al. [183, 180]. Each treatment was incubated in triplicate 20 L polycarbonate carboys that were acid cleaned according to GEOTRACES 'cookbook' trace metal clean protocols [184]. Each carboy was also rinsed and conditioned with filtered

trace metal clean surface seawater prior to use to remove any residual acid. All water for the experiments was pre-filtered through a 100 μM mesh that had been cleaned in 5% HCl (trace metal grade) for one day and rinsed thoroughly with clean MilliQ prior to use. Samples for an initial time point measurement of each sample type were collected from the tow fish directly after the treatment bottles were filled. Bottles were incubated in temperature controlled on deck incubators set to the *in situ* mixed layer temperature with blue shading to simulate mixed layer light conditions for 3 or 4 days. Samples were only collected at the final time point in order to minimize the possibility of trace metal contamination. Each experiment had three treatments and a no amendment control. The experimental design was based on expected nutrient limitation and expectations of the *in situ* nutrient concentrations. In the northern TZ experiment, at 41.4 °N, treatments were high Fe (2 nM), low Fe (0.3 nM), high NPFe (10 μM N, 1 μM P, 2 nM F); in the mid TZ experiment, at 37 °N, treatments were high Fe (1 nM), high NP (5 μM N, 0.5 μM P), high NPFe (5 μM N, 0.5 μM P, 1 nM Fe); in the southern TZ experiment at 33 °N, treatments were high NP (5 μM N, 0.5 μM P), low NP (0.5 μM N, 0.05 μM P), high NPFe (5 μM N, 0.5 μM P, 0.5 nM Fe, details in Table 4.1). Fe was added as $FeCl_3$, N was added as nitrate, P was added as phosphate. Initial and final samples were collected for nutrient (N + N, SRP) concentrations, chlorophyll, flow cytometry, 16S/18S amplicon sequencing, metatranscriptomics, and metabolomics.

4.3.3 Metabolite Sample Collection

Samples for metabolites were collected in HDPE sampling bottles from the CTD (10 L per sample, in triplicate at each station and depth), ships underway uncontaminated water system (10 L per sample, in triplicate at each time sampled), and from the nutrient amendment incubation experiments (5 L per sample, one sample per bottle with triplicate bottles for each treatment). Water was filtered through a 0.2 μm PTFE (Omnipore) filter using a peristaltic pump, polycarbonate filter holder, and Masterflex PharMed BPT tubing (Cole-Parmer). Filter sizes of 142 mm were used for 10 L samples and 47 mm diameter were used for 5 L samples so that the total filtering time was less than 30 minutes for each sample.

Several times during the cruise, filtered seawater was run through another 0.2 μm PTFE filter to collect methodological blank samples. Size fractionated metabolite samples were collected at four stations along the cruise track (stations 3 - 28.4 °N, 4 - 32.2 °N, 5 - 36.1 °N, and 6 - 39.8 °N). These samples were collected by processing a fourth replicate from 15 m and filtering first through a 3 μm pore size polycarbonate filter and then through a 0.2 μm PTFE filter. Only the 0.2 μm PTFE filter was processed, resulting in a metabolome of just the small size fraction. Filters were frozen in liquid nitrogen immediately after filtration and stored at -80 °C until metabolite extraction.

4.3.4 Metabolomics sample processing and data analysis

Particulate metabolite extraction used a modified Bligh-Dyer method as described in Boysen et al. [22]. Samples were analyzed with paired liquid chromatography mass spectrometry using a Waters Acquity I-Class UPLC and a Thermo QExactive HF mass spectrometer with electrospray ionization. The aqueous phase of each extract was analyzed using reversed phase chromatography (Waters Acquity UPLC HSS Cyano column, 1.8 μm particle size, 2.1 mm x 5 mm) in positive ionization mode with resolution of 120,000 and using hydrophilic liquid interaction chromatography (HILIC, SeQuant ZIC-pHILIC column, 5 mm particle size, 2.1 mm x 150 mm, from Millipore) with positive and negative switching with resolution of 60,000. Samples were run in full scan mode. Pooled samples from the transect and the nutrient amendment incubation experiment sample sets were run in both full scan mode and data dependent acquisition mode to collect MS/MS fragmentation data on the top ten most abundant mass features in each scan. Data dependent acquisition was always run with a single ionization mode and at 60,000 resolution. Three runs for each chromatography and ionization pairing were run, with collision energies of 20, 35, and 50.

Proteowizard was used to convert .raw files to .mzxml files [73] and Reifycs abf converter was used to convert to .abf. Untargeted metabolomics data processing was conducted for the two sample sets (transect samples and nutrient amendment incubation experiments) separately to identify the most abundant metabolites in each sample set independently using

MSDial version 4.18 [185], as described in the supplemental methods. After the independent untargeted processing the mass features shared between the sample sets and unique to each sample set were identified (as defined by a mz threshold of < 5 ppm difference and retention time threshold of < 0.5 and 0.2 min for HILIC and RP chromatography respectively).

Skyline [72] was used to extract and integrate select mass features that were unique to the transect sample set in the incubation experiment sample set so that they could be more fully comparable. Skyline was also used to integrate select known metabolites that were not picked by MSDial. Quality control and best matched standard normalization were performed as in Boysen et al. (2018) in order to retain only high quality mass features and reduce variability introduced during sample analysis [22]. After full processing, 435 metabolites were retained in the incubation experiment metabolome and 240 metabolites were retained in the transect metabolome. Only the overlapping compounds between the incubation experiments and transect data were used in the transect data analysis presented here. This resulted in 203 metabolites analyzed across the cruise transect.

Univariate statistics

Analysis of variance (ANOVA) was used to determine which signals (metabolites, flow cytometry cell counts, and flow cytometry biomass estimates) were significantly different between treatments for each experiment. Two sided t-tests were used to determine which signals were significantly different in a treatment compared to the control for each experiment. False discovery rate (fdr) correction of p-values was used to reduce false positives. These analyses were conducted in R using the stats package version 4.0.2.

Multivariate statistics

Multivariate analyses were conducted in R, using the vegan package (version 2.5-6) and cluster package (version 2.1.0). To generate nonmetric multivariate dimensional scaling (NMDS) plots, the normalized peak areas of metabolites were standardized by dividing by the maximum value in the relevant samples. A euclidean distance matrix was calculated based on

the standardized metabolite data, and used to generate NMDS plots. Analyses were also done using z-score standardization, where the metabolite values were standardized to have a mean of zero and standard deviation of one across the relevant samples. The results were not meaningfully different and we thus used maximum standardization to avoid any artifacts generated due to non-normal distributions.

Metabolites, flow cytometry biomass estimates, flow cytometry cell abundances, and chlorophyll concentration were clustered based on their maximum standardized values in each experiment using the clara function, which uses a k-medoids based algorithm similar to partitioning around medoids (pam). Clustering was done with 3 to 15 clusters, and the number of clusters with the highest average silhouette width was picked as the best representation of the data for each experiment. The silhouette width is a number between -1 and 1 where a higher number indicates a higher degree of within-cluster similarity compared to out-of-cluster similarity. All replicate bottles were used to generate the clusters for the experiments in order to capture the full bottle-to-bottle variability. Metabolites in the mixed layer along the transect were clustered with the same methods, using full replicate data to capture trends in metabolites based on the entire biological variability.

4.3.5 Amplicon sample collection and data processing

DNA for 16S and 18S amplicon sequencing was collected during the cruise transect and for each of the incubation experiment treatments. Sampling and DNA extraction methods were as described in Gradoville et al. (2020) [180]. Briefly, seawater was filtered through 3 μm pore size polyethersulfone filters then onto 0.2 μm pore size Supor filters using a peristaltic pump. Samples were flash frozen in liquid N_2 and stored at -80 °C until processing. Samples for DNA were extracted using the DNeasy Plant Mini Kits as described in Gradoville et al. (2020). Details of DNA extraction and amplicon data processing are in the Supplementary Methods. Briefly, DNA extracts were processed with a Fluidgm Access Array^{TM} system combined with Illumina MiSeq^{TM} v3 sequencing set to a run length of 2 x 250bp. Analysis of both 16S and 18S rDNA sequences were performed using the Qiime2 software suite [186].

4.4 Results

4.4.1 Environmental conditions

We collected samples on a late spring meridional transect along 158 °W, and conducted incubation experiments at three locations in the NPTZ (Figure 4.1A,B). A detailed description of oceanic conditions is in Gradoville et al. (2020) [180], briefly, the sea surface temperature declined as latitude increased, and salinity sharply declined north of 30.5 °N, indicating the southern boundary of the subtropical frontal zone [187]. The second derivative of salinity is used as a delineation between the NPSG and the NPTZ, and in 2017 this was at 32.8 °N [180]. North of 38 °N, the surface salinity and salinity depth structure reflected the increased influence of the subarctic gyre [180, 187]. Nutrient concentrations were low throughout most of the cruise track. Combined nitrate and nitrite reached concentrations above 0.01 μM at 38 °N and above 1 μM at 40.6 °N (Figure 4.2A).

Particulate carbon (PC) concentration increased gradually beginning at 33 °N and concentration reached 5 μM near 37 °N (Figure 4.2A). Phytoplankton group specific biomass, as estimated from underway flow cytometry, was characterized by sharp changes in community structure. *Synechococcus* cell abundance increased rapidly starting at 33.3 °N. *Prochlorococcus* cell abundance declined sharply north of 35.4 °N, coincident with another increase in *Synechococcus* (Figure 4.2B,C). Photosynthetic picoeukaryotes less than 9 μm in diameter increased in abundance and biomass near 38 °N. The overall trend in PC was likely driven by the combination of these three groups of organisms and by the larger eukaryotic plankton, such as diatoms, haptophytes and dinoflagellates, that increase in abundance with increasing latitude (*in situ* flow cytobot data, personal communication, A.E. White and M. Dugenne).

Eukaryotic organismal diversity was evaluated using 18S rDNA amplicon sequencing. Along the cruise transect the small (< 3 μm) eukaryotes were dominated by dinoflagellates with an increasing contribution of chlorophytes in the transition zone north of 34 °N (Supplemental Figure 4.18). Haptophytes, metazoa, and ochrophytes all contributed substantially to the total gene copies. In the large size fraction of eukaryotes, dinoflagellates

and chlorophytes were still important, but metazoa and apicomplexa contributed more to the total gene copies. In anomalous samples at 28 °N and 33 °N, radiolaria contributed significantly to the gene copies in both size fractions. The orders arthropoda (from metazoa), dinophyceae (from dinoflagellates), and chloropicophyceae (from chlorophytes) were important in the large size fraction samples, while syndiniales (from dinoflagellates), prymnesiophyceae (from haptophytes), mamiellophyceae (from chlorophytes), dinophyceae (from dinoflagellates), and chloropicophyceae (from chlorophytes) were important in small size fraction samples (Supplemental Figure 4.19).

The most noticeable trend in prokaryotic diversity over the transect based on 16S gene copies was the decrease of the cyanobacterial order Oxyphotobacteria in the small size fraction (Supplemental Figures 4.20, 4.21). The contribution of Bacteroidia increased with latitude in both size fractions. In the small (< 3 μm, free living) size fraction, the contribution of Gammaproteobacteria also increased with latitude.

4.4.2 Metabolites along the cruise track

In order to focus on the nutrient incubation experiments, a subset of 203 metabolites that overlapped with the metabolites detected in the incubation experiments were selected for analysis in this study. Of those metabolites, 71 were quantified using isotopically labeled isotopologues as internal standards or authentic standards analyzed in a relevant environmental matrix, as described in the methods. The total concentration of these 71 metabolites generally mirrors total PC, with an increase in concentration beginning near 33 °N and reaching a maximum concentration near 40 °N (Supplemental Figure 4.22 top). The relative composition of the metabolites is more variable, with sucrose and 4-hydroxyisoleucine, and isoleucine contributing more in the subtropical gyre and homarine contributing more in the transition zone (Supplemental Figure 4.22 middle). There is also some variability in the proportion of PC that these metabolites comprise, at most stations these metabolites totaled less than 4% of total PC, however at some stations in the NPSG and in the NPTZ these same 71 metabolites were more than 6% of PC. This variability was driven by two stations

in the NPTZ with concentrations of homarine that alone reached 3% of PC (Supplemental Figure 4.22 bottom). One of these locations was station 5 at 36.1 °N, where 88 ± 9% of homarine was in the small size fraction (0.2–3 μm). At the three other stations with size fractionated metabolite samples, the small size fraction contributed less than 40 % of total homarine (Supplemental Table 4.3).

The 203 metabolites that overlapped with the experiment and that were in a majority of the mixed layer samples were clustered based on their relative abundance in the samples from the mixed layer along the cruise transect, resulting in 7 clusters. The largest number of metabolites (123) were in groups 1-3, which had maximum metabolite abundance at 39.5 °N (station 8), with decreasing concentrations to the south (Supplemental Figure 4.23). Many known metabolites including amino acids and nucleobases fell into these groups along with the osmolytes glucosylglycerol and gonyol and all carnitines. Group 6 contained 30 metabolites that had maximum concentrations at 39.76 °N, near station 8 but collected two days earlier. Group 7 contained metabolites that had highest concentrations in the northern stations, but with appreciable concentrations in the southern stations as well (Supplemental Figure 4.24). Compounds in this cluster included glycine betaine, homarine, proline betaine, glutamine, glutamic acid, and guanine. Group 5 had seven metabolites with a decreasing trend across the transect, including arsenobetaine and sucrose (Supplemental Figure 4.24, and Supplemental Figure 4.23). South of 39 °N, more than 75% of sucrose was in the small size fraction, while only 13% of sucrose was in the small size fraction at 39.8 °N. Group 4 contained 13 metabolites that had no strong trend with latitude.

4.4.3 Nitrate rather than iron addition stimulates metabolic changes in the northern transition zone experiment at 41.4 °N

In the northern transition zone experiment, at 41.4 °N, significant increases in chlorophyll ($p < 0.05$) occurred only in the NPFe treatment (Figure 4.3A). *In situ* dissolved inorganic nitrogen (DIN) and iron concentrations in the surface mixed layer were 2 μM and 0.3 nM, respectively, resulting in a N:Fe ratio of 6 μM:nM (Table 4.1). After four days of incubation,

DIN was drawn down to 0 in the control, high Fe, and low Fe treatments (Figure 4.3B). The NPFe treatment in the northern transition zone experiment had residual DIN after the four days of incubation, but total drawdown (6 μM) was much greater than that for the control, high Fe, and low Fe treatments (2 μM, Figure 4.3C).

Though there was a significant increase in chlorophyll concentration and a large nitrate drawdown, the biomass of picoeukaryotes and *Synechococcus* were relatively stable across treatments, with some bottle to bottle variability but no significant differences between treatments (Figure 4.3E,G). Diatoms increased their representation in the metatranscriptome in the NPFe treatment compared to the control by three fold, with a greater increase in replicate bottle A than in bottles B and C (personal communication, Ryan Groussman).

In the northern transition zone experiment, the incubation bottles did not have dramatically different prokaryotic and eukaryotic diversity after 4 days of incubation, compared to the *in situ* samples or when compared to other treatments (Figures 4.4, 4.5, Supplemental Figure 4.25). Bottle to bottle variability within treatments stood out most in the control, where control bottle A had more Chlorophyta or less dinoflagellates compared to bottles B and C in the small size fraction. The contribution of 18S genes from haptophyta, ochrophyta, and apicomplexa were quite variable between bottles but without a strong trend differentiating one treatment from another (Figure 4.4).

75 metabolites were statistically different between treatments in the northern transition zone experiment (ANOVA fdr-corrected p-value < 0.05, Table 4.2). The difference between treatments was driven by the dramatically different NPFe treatment, while the high Fe and low Fe treatments were not distinguishable from the control in a NMDS plot (Figure 4.3H). Univariate statistics supports this difference; 56 metabolites were significantly different between the control and the NPFe treatment and no metabolites were significantly different between the Fe treatments and the control (t-test fdr-corrected p-value < 0.1, Supplemental Figure 4.26).

The metabolites and flow cytometry biomass and cell abundance signals within this experiment clustered into six groups characterized by 1: no strong trend across the treatments

but high bottle to bottle variability (n = 32), 2: an increase in the NPFe treatment, particularly bottles B and C (n = 91), 3: generally no trend across the treatments and low bottle to bottle variability (n = 106, including *Synechococcus* abundance and biomass, picoeukaryote abundance), 4: lower concentration in the NPFe treatment (n = 58, including picoeukaryote biomass), 5: high concentrations in all three bottles of the NPFe treatment (n = 55, including chlorophyll), 6: high concentrations in the bottle C of the NPFe treatment (n = 78) (Figure 4.3I, Supplemental Figure 4.27, Table 4.2).

Overall more C and N could be quantified in metabolites in the NPFe treatment compared with the control or any other treatment (Figure 4.3H, Supplemental Figure 4.28). This increased seemed to be driven by an increase in glutamic acid, especially in the NPFe replicate bottles B and C (Supplemental Figure 4.29). Metabolites with significant changes were elevated in the NPFe treatment compared to the control and three of the six clusters of metabolites had elevated concentrations in the NPFe treatments, similar to chlorophyll. However, the largest cluster of metabolites (cluster 3) included *Synechococcus* and picoeukaryote abundance and contained metabolites with little difference between treatments. Some of the common bottle to bottle variability in metabolites (for example cluster 6 where bottle C of the NPFe treatment showed very high concentrations of 78 metabolites) could not be easily explained by the 18S or 16S gene copies.

4.4.4 Nitrate and iron addition stimulate metabolic changes in the mid transition zone experiment at 37 °N

In the mid transition zone experiment, at 37 °N, only the NPFe treatment had a significant increase in chlorophyll, though a small increase occurred with the NP only treatment (Figure 4.6A). Thus, this location is defined as N and Fe co-limited. At this station, there was very low DIN *in situ* of 0.06 μM and iron concentration of 0.51 nM, resulting in a N:Fe ratio of 0.11 (Table 4.1). There was residual DIN in both the NPFe and NP treatments after three days of incubation (Figure 4.6B). DIN drawdown was greater in the NPFe treatment ($\sim$ 2 μM) than the NP treatment ($\sim$ 1.5 μM), despite treatments receiving equal nitrate additions

(Figure 4.6C, Table 4.1).

Phytoplankton groups enumerated with flow cytometry did not show a similar trend as chlorophyll. Picoeukaryote abundance and biomass increased with added nitrogen, regardless of iron addition (Figure 4.6E). *Prochlorococcus* survived well enough to be counted in this experiment, and the replicate bottles had variable *Prochlorococcus* biomass and cell counts but no significant differences between treatments. *Synechococcus* biomass was lower in the Fe treatment compared to the other treatments. The control treatment bottle B had an anomalously high *Prochlorococcus* and *Synechococcus* biomass (Figure 4.6F,G).

Overall, the 18S diversity showed increased gene copy contribution from chlorophyta in the NPFe and NP treatments in the mid transition zone experiment (Figure 4.4, Supplemental Figure 4.30). Bacteroidia contributed more to the large size fraction 16S gene copies in the control and NPFe treatments. Alphaproteobacteria and Oxyphyotobacteria contributed more to the small size fraction 16S gene copies in the Fe treatment and NP treatment (Supplemental Figure 4.20). There was again high bottle-to-bottle variability within treatments. Control bottle B looked quite different from bottles A and C (Supplemental Figure 4.30), with more dinoflagellates and apicomplexa and less of other classes (Figure 4.4). In the prokaryotic community, control bottle A had a larger contribution of Alphaproteobacteria and less Bacteroidia in the small size fraction compared to the bottles B and C (Figure 4.5). Bottle B of the NPFe treatment also looked different with more dinoflagellates and apicomplexa. Bottle A of the NP treatment had more haptophytes and Ochrophyta than bottles B and C from that treatment. Examining the order level showed an increase in Bacillariophyta (diatoms) in the NPFe treatments and NP treatment bottle A (Supplemental Figure 4.19). The diatoms increased their contribution to the total metatranscriptome reads ten fold in the NPFe treatments compared to the control, corroborating the 18S data (personal communication, Ryan Groussman).

152 metabolites were statistically different between treatments in the mid transition zone experiment (ANOVA fdr-corrected p-value < 0.05, Table 4.2). Most of the 152 metabolites that were significantly different were significantly different between the NPFe treatment

and the Fe treatment. 37 metabolites were significantly different based on NP addition alone, when comparing the NPFe and NP treatments together against the Fe and control treatments together. Multivariate analysis visualized with an NMDS plot showed that the treatment groups were different from each other (Figure 4.6H). However, only 11 metabolites were significantly different in the NP treatment from the control, while no metabolites were significantly different in the NPFe compared to the control or in the Fe treatment compared to the control (t-test fdr-corrected p-value < 0.1, Supplemental Figure 4.26). The quantified metabolites showed elevated concentrations in the NP and NPFe treatments and in the control treatment bottle B, where homarine concentrations increased substantially (Supplemental Figure 4.31).

The metabolites in this experiment clustered into three groups, characterized by 1: high concentrations in NPFe treatment and low concentrations in the Fe only treatment (n = 170, including*Synechococcus* and picoeukaryote abundance and biomass), 2: highest concentration in the control treatment bottle B where there were more dinoflagellates and apicomplexa 18S gene copies, and high *Prochlorococcus*, *Synechococcus*, and bacterial biomass (n = 79, including *Prochlorococcus* biomass), and 3: highest concentrations in the NPFe treatment, especially bottle B where there were more dinoflagellates and apicomplexa 18S gene copies, and NP treatment (n = 162, including chlorophyll) (Figure 4.6J, Supplemental Figure 4.27, Table 4.2).

Though chlorophyll changed significantly only in the NPFe treatment compared to the control, evaluating the biomass changes of specific phytoplankton groups and the concentration changes of metabolites showed many other physiological responses. The picoeukaryotes and many metabolites showed changes in response to any N addition. Total quantified C and N in metabolites was elevated to a similar level in the NP and NPFe treatments, and was significantly depleted in the Fe treatment compared to the control (Figure 4.6I, Supplemental Figure 4.28).

4.4.5 Nitrate stimulates metabolic changes in the south experiment at 33 °N

In the southern transition zone experiment, at 33 °N, chlorophyll increased significantly in the high NP and NPFe treatments, with a smaller and not significant increase in the low NP treatment (Figure 4.7A). There was only 0.01 μM DIN *in situ* at this station (Figure 4.2). Dissolved iron concentration *in situ* was 0.22 nM, resulting in a N:Fe ratio of 0.045 μM:nM. After three days of incubation, the added nitrate was fully drawn down in two out of the three bottles of the low NP experiment, but the high NP and NPFe treatments had residual DIN at the end of the three day incubation, after drawing down $\sim$1.2 μM of the 5 μM addition, respectively (Figure 4.7B,C).

Chlorophyll concentration primarily responded to NP addition, and the individual phytoplankton groups, as enumerated by flow cytometry, showed similar responses to the treatments. Picoeukaryote biomass increased with added nitrogen, though again it was variable between bottles (Figure 4.7E). *Procholoroccus* only survived in the high NP treatment (Figure 4.7F). *Synechococcus* survived in any treatment with added N but not in the control treatment (Figure 4.7G). Bacterial biomass was elevated in the High NP treatment Figure 4.7D).

In the south transition zone experiment, the incubations looked quite different from the *in situ* samples from the same station, but quite similar to other *in situ* communities at nearby stations, suggesting the corresponding station data was anomalous and not representative of the starting conditions of the experiment (Figure 4.4 top, Supplemental Figure 4.18). The treatments did not cluster together based on 18S diversity in multidimensional space, but the high NP bottles B and C were quite different in their eukaryotic diversity than the other bottles (Supplemental Figure 4.32). The high NP bottles B and C which had more chlorophytes in the small size fraction than the other bottles. The prokaryotic community was also different in the high NP treatment bottles B and C, where Bacteroidia (Flavobacteriales order) contributed much more to the total gene copies (Supplemental Figures 4.20, 4.21). These two bottles were where *Procholoroccus* survived and bacterial biomass was high. High

NP bottle C also had a greater contribution from dinoflagellates in the large size fraction.

Though chlorophyll, picoeukaryotes, *Prochlorococcus*, and *Synechococcus* all showed some N response, only twelve metabolites were statistically different between treatments in the south transition zone experiment (ANOVA fdr-corrected p-value < 0.05, Table 4.2). Multivariate analyses visualized with an NMDS plot show overlap between NPFe and high NP treatments, while the low NP and control groups were separate (Figure 4.7H). Only one metabolite was significantly different between the high NP treatment and the control (fdr-corrected p-value < 0.1, Supplemental Figure 4.26).

The metabolites in this experiment clustered into six groups characterized by 1: elevated concentrations in the high NP treatment (n = 92, including picoeukaryote abundance and biomass and most significantly different metabolites), 2: no strong trend across treatments (n = 56, including chlorophyll, and *Synechococcus*), 3: lowest concentration in the control and high bottle to bottle variability with highest concentrations in NPFe bottle A (n = 40), 4: highest concentration in high NP treatment replicate bottle B (n = 53), 5 and 6: highest concentrations in high NP treatment replicate bottles B and C, where there were more chlorophytes in the small size fraction 18S genes, and more Bacteroidia in the 16S genes, more *Prochlorococcus* and *Synechococcus* and heterotrophic bacterial biomass (group 5 n = 77, group 6 n=88) (Figure 4.7J, Supplemental Figure 4.27, Table 4.2).

Overall most changes in metabolites and phytoplankton biomass mirror the chlorophyll concentration, with elevated concentrations when high concentrations of nitrate were added. Total C and N in quantifiable metabolites also reflected these changes, with C and N in these metabolites doubling in the high NP and NPFe treatments compared to the control, and an increase of ∼30% in the low NP treatment (Figure 4.7 I, Supplemental Figure 4.28). There were few significantly different metabolites due to high bottle to bottle variability in the high NP treatment. This variability was also in the 16S and 18S diversity.

4.4.6 Specific metabolite changes in the incubation experiments

Mycosporine-like Amino Acids

Of the 435 metabolites that were analyzed in this dataset, only four metabolites were statistically different between treatments for all three experiments (Figure 4.8). None of these four metabolites were compounds for which we have authentic standards. However with the exact mass and MS/MS spectra we were able to putatively identify them. One is likely the mycosporine-like amino acid (MAA) usujirene or the MAA palythene, two compounds that are cis/trans isomers. Another is likely the MAA palythanic acid. A third has the exact mass of the MAA shinorine (m/z at 331.11505 with negative ionization), but has a fragmentation spectra that does not match the literature. The fourth has a m/z of 439.1016 in positive mode (which matches the mass of the vitamin B2 metabolite, riboflavin cyclic-4',5'-phosphate) (Figure 4.9). Two of these four compounds, m/z 331.1150, and the candidate for palythanic acid, have remarkably similar patterns between the experiments, with concentrations that were responsive to N addition (Figure 4.9). They both were dominantly in the large (> 3 μm) size fraction with <10% in the small size fraction at all latitudes (Supplemental Table 4.3). The metabolite that is likely usujirene or palythene has a similar pattern of increasing concentration with N addition, but the concentration of this metabolite increased even further in the NPFe treatment in the mid transition zone (37 °N) experiment, a similar pattern to chlorophyll. The metabolite that is likely riboflavin cyclic-4',5'-phosphate also increased in concentration with increasing N, but in the southern experiment showed a stronger response to the high NP treatment than the NPFe treatment, similar to that of chlorophyll (Figure 4.7A).

Three other likely MAAs were identified in the experiments and showed differential abundance in some treatments with generally increased concentration in N and NPFe treatments for all experiments, even if they were not significantly different based on the criteria set here (Figure 4.10). The metabolite that is potentially asterina (m/z 289.13928) was the only potential MAA that was not responsive to N in all experiments. The transect dataset

contained high quality data for four out of the five MAAs as well as the two other metabolites that changed significantly between treatments in all three experiments, m/z 331.1150 and the tentatively identified riboflavin cyclic-4',5'-phosphate. All of these metabolites except usujirene/palythene and asterina had low concentrations in the NPSG and a maximum concentration just before 40 °N (Figure 4.11).

Osmolytes

Despite the small number of compounds that were consistently different in all experimental treatment at all locations, both the northern transition zone and mid transition zone experiments had many metabolites that separated the treatments (Figure 4.8, Supplemental Figure 4.26). The 41 metabolites that were different between treatments in both the northern transition zone and mid transition zone experiments include glutamine, homoserine, s-adenosyl homocysteine, proline, aspartic acid, taurine, uric acid, and L-5-Oxoproline. Proline and tuarine are nitrogenous osmolytes, and their presence on this list suggested that nitrogenous compounds that accumulate to high intracellular concentrations may be influenced by a sudden increase in DIN. While the six other nitrogenous osmolytes that we measured were not significantly different between all the treatments, there was a dominant trend of increasing with increased DIN addition. Homarine, taurine, proline, glutamic acid, and trimethylamine N-oxide (TMAO) showed elevated concentrations with added nitrate. Glycine betaine, proline betaine, and TMAP increased in the south transition zone experiment (at 33 °N) high NP treatment, but not in the mid or northern transition zone experiments (Supplemental Figure 4.33, Figure 4.12) However, these changes may not be related to the N-content of these compounds.

The non-nitrogenous osmolytes we measured also increased with high and low NP additions in the south transition zone experiment, with the exception of gonyol (Figure 4.12, Supplemental Figure 4.34). DHPS, isethionic acid, and glucosylglycerol concentrations increased with added nitrate in the northern transition zone and mid transition zone experiments, and in the mid transition zone experiment showed an additional increase in the NPFe treatment.

DMS-Ac and DMSP did not respond to nitrate addition in the northern- or mid transition zone experiments (Supplemental Figure 4.34). Gonyol was highly variable across replicates, with a slight negative correlation between nitrate addition and gonyol concentration. In the northern and southern transition zone experiments, the treatments that received nitrate had lower mean gonyol concentrations than those that did not receive nitrate (Supplemental Figure 4.34). Overall, the patterns in osmolyte concentration across the experiments do not show nitrogenous osmolytes and non-nitrogenous osmolytes behaving categorically differently (Figure 4.12).

In the transect, the dominant trend for osmolyte concentration was an increase with latitude and maximum near 40 °N (Figure 4.12). The three exceptions to this trend were sucrose, which showed a decrease in concentration with increasing latitude and was not detected in the experiments, homarine, which showed a peak near 38 °N, and isethionic acid, which continued increasing with latitude north of 40 °N (Figure 4.12).

Free Amino Acids

The concentrations of free amino acids showed relatively coherent patterns in the treatments of the different incubations. In the northern transition zone experiment (41.4 °N) amino acids had highest concentrations in the NPFe treatment, with the exception of alanine which had the highest concentration in the high Fe treatment (Figure 4.13). In the mid transition zone experiment (37 °N) most amino acids had the lowest concentration in the Fe treatment and highest in the NP or NPFe treatment (Figure 4.13). In the south transition zone experiment (33 °N), nearly all amino acids had the highest concentration in the high NP treatment, and most had a higher concentration in the NPFe treatment than in the low NP or control treatments (Figure 4.13). The exception to this trend was leucine which showed no trend across the treatments in the south transition zone experiment. In the transect data, amino acids uniformly increased with increasing latitude until a peak in concentration near 40 °N (Figure 4.14), where the total metabolite pool also had maximum concentration (Supplemental Figure 4.22).

Metabolites that change with biomass signals

In order to determine which metabolites were associated with different organisms' abundance, metabolites, chlorophyll concentration, biomass and cell abundances from flow cytometry for each experiment were clustered into groups that had similar behavior in the incubations (Figures 4.3I, 4.6J, 4.7J, Supplemental Figure 4.27, Table 4.2). We examined the overlapping metabolites in the clusters that contained the biological variables from flow cytometry and bulk chlorophyll.

Chlorophyll clustered with only two metabolites across all three experiments. These were two unknown metabolites with exact masses of 317.1498 and 367.1618 in positive ionization. Neither has associated fragmentation data to aid in identification. The mass feature with m/z of 367.1617 is potentially a small peptide based on exact mass alone, which matches with the empirical formula $C_{16}H_{22}N_4O_6$.

Twenty five metabolites clustered with picoeukaryote abundance and biomass. These metabolites included two carnitines (acetyl-l-carnitine and butyryl-l-carnitine), homarine, trigonelline, sulfanilic acid, dimethylsulfonioacetate (DMS-Ac), and indole-3-carboxaldehyde. Additional mass features identified in our untargeted data include iodine, a metabolite with the exact mass of methyladenosine, and several unknown metabolites that are identified as primarily produced by dinoflagellates in Heal et al. (2020) [41].

Twenty one metabolites clustered with *Synechococcus* abundance. These include propionyl-l-carnitine, alanine, and trimethylamine N-oxide. There were also mass features in this cluster that were identified in the phytoplankton metabolome data in Heal et al. (2020) as primarily produced by dinoflagellates or by diatoms [41].

Prochlorococcus only survived in the mid-TZ experiment at 37 N, so the metabolites that cluster with *Prochlorococcus* biomass are those in cluster 2 of the mid transition zone experiment. These include hypotaurine, carnitine, trimethylammonium propionate (TMAP), sarcosine, arsenobetaine, betaine, ectoine, 4-hydroxyisoleucine, betonicine, guanosine, proline betaine, and tryptophan. Other mass features in this group were also identified in

microbial community metabolomes from the NPSG and NPTZ collected in spring 2016 but were rarely or never seen in the metabolomes of laboratory cultures of phytoplankton [41].

Carnitines

Several carnitines clustered together with picoeukaryote biomass and abundance across all the experiments, so we examined their signals more closely. We found one other putative carnitine in the untargeted metabolites in our dataset based on exact mass, fragmentation, and retention time on the RP column. The putative C5-acylcarnitine has a m/z of 246.1701 and a retention time of 2.0 minutes with major fragments of 220 ,187, 85, and 60 da. These smaller two fragments are also found in the spectra of the acetyl-L-carnitine authentic standard, and the 187 da fragment corresponds to the loss of a trimethylamine group from C5-acylcarnitine (Supplemental Figure 4.35).

The four carnitines in the transect data increased with increasing latitude, with three reaching peak concentration at ∼40 °N and butyryl-l-carnitine continuing to increase with latitude (Figure 4.15). Though the putative C5-acylcarnitine was not identified in the transect dataset, we can infer its *in situ* concentrations from the experiments. The initial abundance of the putative C5-acylcarnitine was an order of magnitude higher in the northern transition zone experiment than the mid transition zone and south transition zone experiments (Figure 4.16). The five carnitines were not significantly different among any of the individual treatments compared to the controls (fdr-corrected p-value > 0.05). However, in the southernmost experiment (33 °N), all five carnitines had high concentrations in the high NP treatment and butyryl-l-carnitine and propionyl-L-carnitine had high concentrations in the NPFe treatment (Figures 4.17, Figure 4.16). In the mid transition zone experiment (37 °N) all five carnitines had lowest concentrations in the Fe treatment (Figure 4.17). In the northernmost experiment (41.4 °N), where there were high *in situ* carnitine concentrations, there was no strong trend in concentration across the treatments in carnitine or acetyl-l-carnitine. The longer chain carnitines showed slightly elevated concentrations in the high Fe experiment (Figure 4.17).

4.5 Discussion

4.5.1 Nitrogen limitation dominates the NPTZ in the spring

The boundaries of oceanic gyres are regions where nutrient co-limitation is predicted based on phytoplankton physiology from satellites and modeled nutrient concentrations [35, 168]. Here we provide evidence that parts of the NPTZ may be nitrate and iron co-limited. In the mid-transition zone experiment at 37 °N there was a significant increase in chlorophyll concentration only when both N and Fe were added (Figure 4.6A). Nitrogen addition alone controlled chlorophyll responses in the experiments at 41 °N and 33 °N (Figures 4.3A 4.7A).

Given the actual *in situ* concentrations of iron measured in the transition zone during this cruise, the iron co-limitation is surprising. Iron was elevated in the mid-transition zone region (Table 4.1), resulting in an N:Fe ratio of 0.11 (μM:nM). This value is dramatically lower than the ratio that is considered iron limiting for diatoms (N:Fe $>$ 10) [188], less than the ratios in the south Atlantic where N and Fe co-limitation was found (N:Fe between 0.5–1.5) [35], and less than the N:Fe ratio in the North Pacific where *Procholoroccus* was simultaneously N and Fe limited (N:Fe $\sim$1.6) [171].

Iron was delivered to the NPTZ by aerosols from Asia that had both mineralogical and anthropogenic sources [182]. The iron delivery that began in March of 2017 coincided with longer day lengths and likely fueled nitrate drawdown [182]. Though the Ekman supply of nitrate diminished as the spring progressed [34], the aerosol iron delivery continued until June, leaving the mid-transition zone with significant residual iron and almost no residual nitrate (Table 4.1, Figure 4.1B) [182]. The type of iron may change the bioavailability. We suggest that though there was high dissolved iron *in situ*, diatoms responded to iron additions in the northern- and mid-transition zones because the residual aerosol iron was not as bioavailable as the added $FeCl_3$. Though total chlorophyll suggested co-limitation at 37 °N, nitrate alone controlled the responses of many individual groups of phytoplankton and of many metabolites both at 37 °N and in the incubation experiments to the north and south. In the mid-transition zone experiment, diatoms were likely the organisms that drove the

total chlorophyll response (personal communication, Ryan Groussman). In this experiment *Procholoroccus* did not respond to either N or Fe, and picoeukaryotes and *Synechococcus* had biomass increases with any N addition. Different groups of organisms clearly experience different limitations in the same water. By exploring the biochemicals that changed with changes in different groups of organisms, we can begin to untangle the complex dynamics underlying individual organisms' nutrient response.

4.5.2 Metabolic changes indicative of nitrogen stress or stress relief

Though there were changes in the community metabolomes between treatments in each of the three experiments, a 'golden ticket' of a biomarker for nitrogen limitation remains elusive. Many metabolites changed in response to nutrient amendments, but only four compounds changed significantly between treatments in all three experiments (Figure 4.8). Two of these four compounds, the unknown with m/z of 331.1150 and the candidate for palythanic acid, were primarily responsive to N addition and thus are the best candidates identified here for future investigation into use as biomarkers for nitrogen stress relief or limitation (Figure 4.9). In the transect, both of these metabolites were dominantly in the large (> 3 μm) size fraction, had relatively low concentrations in the gyre and southern portion of the NPTZ, and reached a maximum concentration just before 40 °N (Figure 4.11, Supplemental Figure 4.24, Supplemental Table 4.3). Despite their change in all three experiments, even these molecules likely reflect physiology changes in only a subset of the microbial community.

MAAs

The five potential MAAs measured in the experiments here behaved fairly coherently, with increased concentrations with added nitrate (Figure 4.10). MAAs are made of a central aminocyclohexenone ring or aminocyclohexenimine ring, with possible conjugations to nitrogen or aminoalcohol. These secondary metabolites are UV-absorbing compounds produced by diverse organisms including micro and macroalgae, fungi, and cyanobacteria that have known roles in photoprotection and other stress responses [189, 190]. Ammonium stimulates

MAA production in cyanobacteria and algae, and it has been proposed that these nitrogen rich compounds can be used as a resource during nitrogen limitation [191, 192]. However, the extent to which MAAs can be used as a nutrient resource is unknown because the pathways of MAA degradation and nitrogen mobilization remain unidentified [190]. The incubations here were done in shaded incubators and all treatments were exposed to the same light conditions, therefore we suggest that the MAA changes were due to organisms synthesizing these compounds in excess as a strategy for future nutrient stress, oxidative stress, or UV-light exposure.

Organisms that produce MAAs typically produce only a select few and thus MAAs have some organism specificity [193]. Some organisms produce MAAs constitutively and others only in response to UV light exposure [190]. In phytoplankton metabolomes, MAAs were dominantly produced by dinoflagellates, with the exception of the putative asterina (m/z of 289.13928), which was produced by *Crocosphaera* [41]. Asterina is also produced by other organisms including freshwater cyanobacteria and the macro-red-algae *Gelidium sesquipedale*, *Devalerea ramentacea*, and *Palmaria palmata* [193, 194, 195]. Asterina was the only MAA in these experiments that was not elevated with N addition in the southern transition zone experiment, suggesting that the response of the other four MAAs in these experiments may be indicative of dinoflagellate nitrogen limitation relief.

The osmolyte gonyol is produced in extremely high intracellular concentrations by some dinoflagellates [41, 196]. In the size fractionated samples across the transition zone, $> 80\%$ of gonyol is in particles > 3 μm in diameter, suggesting larger organisms, including dinoflagellates, were responsible for its production (Supplemental Table 4.3). However, gonyol does not behave similarly to MAAs in these experiments, and instead is variable but relatively steady concentration between treatments, suggesting little change in dinoflagellate biovolume and biomass (Supplemental Figure 4.34 bottom). The difference in behavior between these potential dinoflagellate biomarkers and the absence of extreme UV light stress in our incubators suggests that MAAs be a fairly taxon specific indication of a physiological change in response to nutrient availability. More investigation into the identity of the MAA candi-

dates here and their distribution under varying nutrient conditions in the surface ocean and in pure cultures will shed light on the cellular roles of these metabolites.

Free amino acids

Cellular free amino acids respond rapidly to nitrogen stress, declining dramatically in concentration when N is limiting [197]. In these incubation experiments, the concentrations of most free amino acids did increase with the addition of nitrogen (Figure 4.13). There were a few exceptions to this trend, however, with select amino acids such as alanine and leucine showing little or no response to N addition. Particulate free amino acids have diel cycles in the marine environment, with maximum concentrations in the afternoon and evening, suggesting that in the absence of nutrient changes they are indicative of overall anabolic activity and growth [30]. The changes in our observed experiments likely reflect a combination of increased growth and anabolism due to nutrient addition and the independent nitrogen response.

The ratio between free glutamine (Gln) and glutamate (Glu) has been used as an indicator of nitrogen stress in organisms from bacteria, to yeast, to algae, with higher Glu:Gln in nitrogen limitation [198, 199, 200]. In the northern-transition zone experiment, the ratio of Glu:Gln decreased when nitrogen was supplied, as we expected. However, Glu:Gln showed no change between the control and the nitrogen treatments in the mid-transition zone experiment, and increased when nitrogen was supplied in the southern experiment, the opposite of expected (Supplemental Figure 4.36A). This opposite trend of the expected changes in Glu:Gln was also observed in the North Pacific subtropical gyre when a natural nutrient fertilization occurred associated with a lava flow (Supplemental Figure 4.36B)[201]. Community composition has the potential to have a strong influence on the Glu:Gln ratio, since some organisms use glutamate as an osmolyte and accumulate it to high concentrations [41, 202, 203, 43]. In the *in situ* transect metabolomes, glutamine and glutamate had different patterns. Glutamate steadily increased with latitude and glutamine had a mid-transition zone concentration minima. Glu and Gln additionally had different proportions in

the < 3 μm size fraction, particularly at 39.8 °N where ∼75% of glutamate but only ∼17% of glutamine was in the small size fraction (Supplemental Table 4.3). On the community level these molecules seem to have different dominant sources, and the change in community diversity and multiple uses of glutamate and glutamine beyond their role in nitrogen assimilation overwhelm the signal of nitrogen limitation and stress relief that is observed in pure cultures.

Osmolytes

We might expect major pools of flexibile nitrogenous compounds to be exchanged for non-nitrogenous compounds when N is limited. This has been shown to occur for phosphorus, where phytoplankton change their lipid use to replace phospholipids with non-phosphorus membrane lipids when living in phosphate deplete waters [204]. Like amino acids, nitrogenous osmolytes are a large pool of flexible cellular nitrogen and potentially could be replaced with osmolytes that do not contain nitrogen [43, 205, 206, 173]. The nitrogenous osmolytes with lower C:N ratios often increased with nitrate addition in the experiments (Figure 4.12). This could be due to a change in the preferred osmolytes of microorganisms in response to N availability. In cultures of a haptophyte, glycine betaine, an osmolyte with a C:N ratio of 5, showed a strong increase in concentration in response to N-resupply, while DMSP, an osmolyte with no nitrogen, showed an insignificant change [173]. The observed trend could also be due to a change in community composition and biomass during the incubation. Phytoplankton community composition has a strong impact on environmental metabolomes, and osmolytes isethionic acid, taurine, DMSP, DMS-Ac, and DHPS, are signatures of eukaryotic phytoplankton [32, 41]. It is possible that in many organisms osmolytes are truely 'dead end' metabolites, and thus organisms cannot mobilize these nitrogen containing compounds into a form that can be used for biomass production [207, 42]. If this were the case for organisms in the NPTZ, the physiological explanation of osmolyte switching under N-replete and N-limited conditions would not be environmentally relevant. In our study, the lack of a coherent categorical difference in response between osmolytes with and without nitrogen

suggests that osmolytes are indicative of microbial biomass and diversity changes during incubation (Figure 4.12), supporting previous work that showed no simple switch between DMSP and GBT or homarine after nitrogen addition in the Gulf of Maine [208].

Homarine is an example of an osmolyte that is somewhat organism specific and reflect overall organism abundance rather than physiology. Homarine can be produced by eukaryotic algae [112], but it is made in copious quantities by several strains of *Synechococcus* [41]. In the southern transition zone experiment, the homarine concentration increased in the NP treatments and in the mid-transition zone experiment homarine concentration increased in the N addition treatments and in control treatment bottle B (Supplemental Figures 4.37, 4.31). These were all conditions in which *Synechococcus* thrived. One of the stations in the transect that had extremely high homarine concentrations, at 36.1 °N, was also where nearly all of the homarine was in the small size fraction between 0.2 and 3 μm (Figure 4.22, Supplemental Table 4.3). *Synechococcus* biomass was greater than picoeukaryote biomass at this latitude (Figure 4.2). Together these observations indicate that the high homarine concentration near 36 °N and in the southern and mid-transition zone incubation experiments may be biosignatures of *Synechococcus*. This example indicates that the osmolyte trends in the incubations are likely a signal of particular groups biomass and biovolume changing rather than osmolyte switching.

Carnitines

Carnitines are a group of compounds based on the quaternary amine 3-hydroxy-4-N,N,N-trimethylaminobutyrate (known as carnitine), with acetylations of various carbon chain lengths. Carnitines are essential in energy production because of their role in the transport of fatty acids into the mitochondria for β-oxidation of fatty acids to generate acetyl-coA which then fuels the citric acid cycle. In addition to its role in fatty acid metabolism, carnitine is also a compatible solute that can provide osmoprotection and cryoprotection [209, 210]. In the NPSG and southern part of the transition zone, greater than 50% of carnitine was in particles < 3 μm in diameter, while in the northern portion of the transition zone, at 39 °N,

only ∼25% of carnitine was in the small size fraction, suggesting a change in the dominant organisms producing this molecule. In contrast, ∼35% of acetyl-l-carnitine was in the small size particles in the NPGS and in the northern transition zone this dropped to 3%, while propionyl-l-carnitine and butyryl-l-carnitine were below the limit of detection in the small size fraction and thus we can assume a majority was in the large size fraction. These differences in sources, and the additional role of carnitine as a compatible solute, may explain the different trends between the acylcarnitines and carnitine in our experiments (Figure 4.16).

The shortest chain modification, acetyl-l-carnitine, serves as a direct replenishment of acetyl-coA and thus influences many cellular functions [211]. Mining the untargeted metabolite data generated as part of Chapter 3, we found that acetyl-l-carnitine has a strong diel periodicity with a peak in the evening while the longer chain butyryl-l-carnitine did not have a diel periodicity. The differences likely relate to the direct role acetyl-l-carnitine has in central metabolism by replenishing acetyl-coA while butyryl-l-carnitine is more constrained in its cellular role of fatty acid degradation. Acylcarnitine concentrations decrease in diatoms under vitamin B_{12} limitation but not under light limitation, and thus their concentrations are not simply a function of growth rate [100]. The four acylcarnitines had similar concentration patterns in our experiments, with increased concentration in the southern experiment high NP treatment, and a decrease when no N was provided in the mid-transition zone experiment (Figures 4.16, 4.17). Algae are known to produce storage lipids, including fatty acids and fatty acid containing triacylglycerols, under nitrogen limitation ([212] for example) and we propose that the N addition in these experiments stimulated fatty acid degradation to fuel growth once nitrogen was available. This hypothesis will be further explored using the metatranscriptomes collected during these experiments to look for expression of genes involved in fatty acid degradation and specific carnitine acyltransferases indicative of these processes.

4.5.3 Unknown Untargeted Metabolites

The untargeted data in these experiments has already yielded candidate metabolites whose identities need to be confirmed with authentic standards and whose metabolic roles in marine microbial systems need further investigation. These include the putative MAAs, the C5-acylcarnitine, and the nitrogen-responsive mass feature with *m/z* of 331.1150. Many other unidentified molecular features showed changes between experiments or with latitude. Using the phytoplankton metabolomes in Heal et al. (2020) these unknown metabolites may be linked to particular organisms [41]. Interestingly, in the second cluster of metabolites in the mid-transition zone experiment there were many metabolites that were also found in the NPTZ in 2016, but which were rarely or never seen in the phytoplankton culture metabolomes [41]. In the mid-transition zone experiment, these metabolites reached a maximum concentration in the control treatment, particularly in bottle B (Figure 4.3I) [41]. The 18S and 16S diversity data also showed that control bottle B was anomalous (Supplemental Figure 4.30), with more of the Syndiniales order of dinoflagellates and a bacterial community with more representation from Actinomarinales, Verrucomicrobiales, and Flavobacteriales (Supplementary Figures 4.38, 4.39). Though we do not attempt to determine the origin or nature or the unknown metabolites here, we suggest further investigation into the identity of these metabolites, their sources, and their roles in the microbial community.

4.6 Conclusion

In the spring of 2017, nitrogen limitation dominated the NPTZ from the gyre to the subarctic influenced northern transition zone. We show that nitrogen and iron are co-limiting in the mid-transition zone, where the addition of nitrate and iron together stimulated chlorophyll production and diatom growth even though *in situ* N:Fe was 0.11. This expands the region of iron and nitrogen co-limitation to regions with lower N:Fe than previously considered. The NPTZ is a seasonally dynamic region where organisms must be able to rapidly respond to changing environmental conditions in order to thrive, and our work suggests a va-

riety of mechanisms for responding to changes in nutrient concentrations. Measuring whole community metabolomes rather than just chlorophyll dramatically expanded our ability to detect physiological changes. Overall, instead of metabolites that change similarly in all communities after N, P and Fe additions, we found metabolites were indicators of particular organisms' biomass and physiology. By connecting biomass observations from flow cytometry and 16S and 18S genes, we are able to make inferences about metabolite changes. This work represents one more step in connecting microbial community diversity and physiology with the biochemicals that form the foundation of their metabolism. We have identified several candidate metabolites as potentially important in microbial adaptations to nutrient stress and stress-alleviation, including MAAs and carnitines. These data will be fruitful to mine during future culture-based investigations that focus on the metabolite changes that specific organisms rely on to respond to nutrient stress.

4.7 Acknowledgement

The authors acknowledge L. Carlson, L. Wood, L. Turner, and A. Meyers for lab assistance; L. Carlson, K. Heal, W. Kumler, N. Kellogg, J. Sacks, R. Lionheart for feedback and assistance on data analysis and interpretation; K. Heal, R. Groussman, B. Durham, R. Lim, N. Hawco, R. Bundy, P. Pinedo-Gonzalez, R. Gradoville, M. Carlson, and N. Shamir for help collecting the samples; B. Durham, B. Lambert, M. Schatz, R. Morales, and R. Gradoville for processing and analyzing the 16S and 18S data; R. Groussman for analyzing the metatranscriptomics data; K. Bjorkman, R. Foreman and D. Karl for the nutrient data; A.E. White for particulate carbon and nitrogen data; F. Ribalet, A. Hynes, and K. Cain for the seaflow and flow cytometry data; R. Gradoville, R. Bundy, and B. Lambert for assistance with the methods sections; the entire membership of the SCOPE-Gradients collaboration for productive conversations and feedback, particularly E.V. Armbrust; the crew and scientific party of the R/V Marcus G. Langseth during Gradients-2 cruise. This work was supported by grants from the Simons Foundation (LS Award ID: 385428, A.E.I.; SCOPE Award ID 329108, A.E.I.) and the National Science Foundation (NSF OCE-1228770 and OCE-1205232

to A.E.I., NSF GRFP to A.K.B.)

4.8 Tables

Name	Date	Lat (N)	Time of day set up	T	S	DIP (µM)	DIN (µM)	Fe (nM)	Days Incubated	Treatment A	Treatment B	Treatment C	Treatment D
1: N-TZ	6/3/2017	41.40	Midday	11.1	30.3	0.49	2	0.3	4	Control	**low Fe**: 0.3 nM Fe	**high Fe**: 2 nM Fe	**NPFe**: 2 nM Fe, 10 µM N, 1 µM P
2: mid-TZ	6/6/2017	37.00	Midday	15.3	33.1	0.12	0.06	0.51	3	Control	**Fe**: 1 nM Fe	**NP:** 5 µM N, 0.5 µM P	**NPFe**: 1 nM Fe, 5 µM N, 0.5 µM P
3: S-TZ	6/9/2017	32.93	Afternoon	18.7	34.4	0.09	0.01	0.22	3	Control	**low NP**: 0.5 µM N, 0.05 µM P	**high NP**: 5 µM N, 0.5 µM P	**NPFe**: 0.5 nM Fe, 5 µM N, 0.5 µM P

Table 4.1: Details of sampling locations and experimental treatments for the nutrient amendment incubation experiments. Name is the number and name of the experiment as referred to in the text. Date, Lat, Time of day set up all indicate when and where the experiment was initiated. T (temperature in °C), S (salinity), DIP (dissolved inorganic phosphorus, aka phosphate - PO_4^{3-}), DIN (dissolved inorganic nitrogen, aka NO_3^- + NO_2^-), and Fe correspond to the *in situ* mixed layer physical and chemical properties at the sampling site. The four treatments (A-D) are described for each experiment, where treatment A is always the control, and treatment D is always + N + P + Fe. Added N was always in the form of nitrate, added P was always in the form of phosphate, Added Fe was always in the form of $FeCl_3$.

4.9 Figures

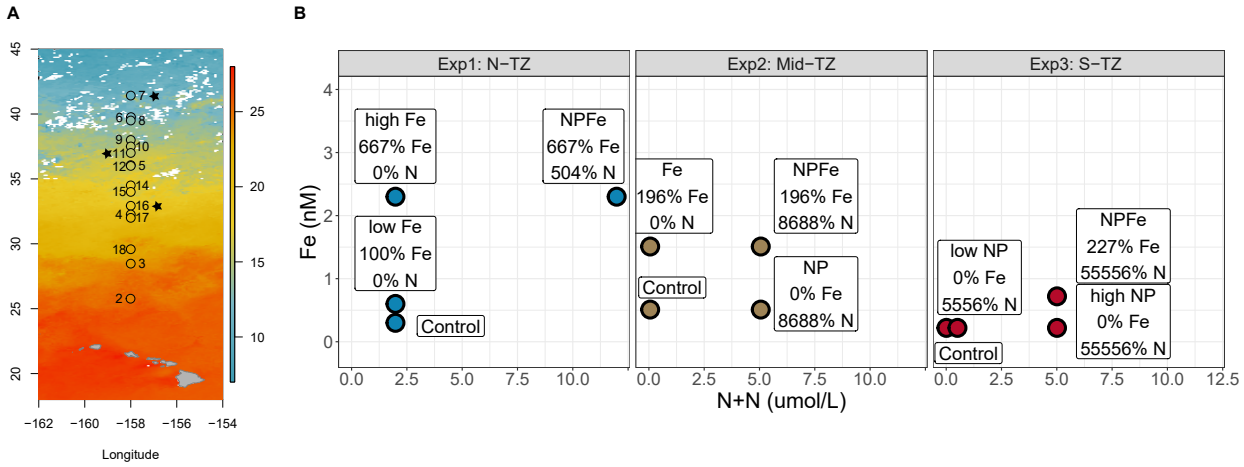

Figure 4.1: A map of the stations sampled on the cruise (A), with a background color of sea surface temperature, generated with data from MODIS SST [213] using 9 km gridded data with 8 day averages, averaged over the period from May 17 - June 25, 2017. Starred stations indicate a nutrient amendment experiment took place at that station. The experimental treatments (B) with the control bottle at the *in situ* nutrient concentrations, and the other points representing the nutrient amendment treatments for each experiment, with the percent increase in N and Fe. We attempted to increase concentrations by doubling (100% increase) or increasing ten fold (900% increase) the *in situ* concentrations, but there was lower nitrate than we expected in the southern two experiments.

Signals in each cluster and that are significantly different between treatments

		Signals per cluster					
	Total signals	1	2	3	4	5	6
Exp1 - North							
All	420	32	91	106	58	55	78
Significant	75	—	10	—	5	44	15
Exp2 - TZ							
All	411	170	79	162	—	—	—
Significant	152	62	6	83	—	—	—
Exp3 - South							
All	406	92	56	40	53	77	88
Significant	12	7	—	—	1	4	—

Table 4.2: A summary of the total number of metabolites, biomass and cell abundance signals in each cluster, and those that are statistically different between treatments for each experiment (ANOVA fdr-corrected $p < 0.05$). Signals were clustered based on their maximum standardized values in each experiment using a k-medoids based algorithm, number of clusters with highest average silhouette width was picked as the number of clusters that best represented the data for each experiment. The total number of signals included in the clustering of each experiment was slightly different because some metabolites, biomass, and cell count signals were below the limit of detection in most samples for a given experiment and were thus discarded before clustering.

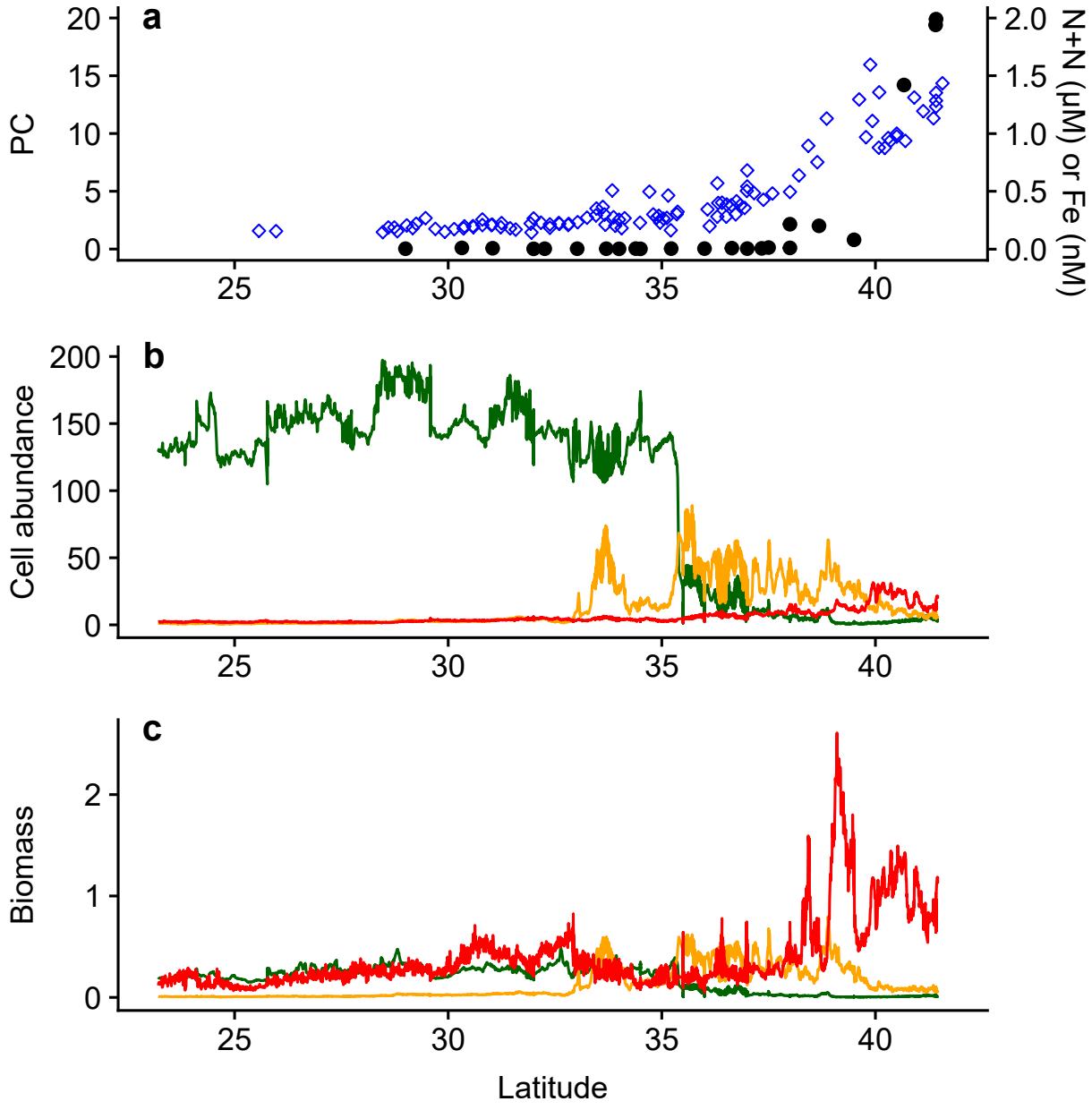

Figure 4.2: DIN (black) in μmol L^{-1} and particulate carbon (blue) in μmol L^{-1} (a), cell counts of *Prochlorococcus* (green), *Synechococcus* (orange), and photosynthetic picoeukaryotes less than 9 μm (red) in cells μL^{-1} (b), and carbon biomass of those groups in μmol C L^{-1} (c) over the transect.

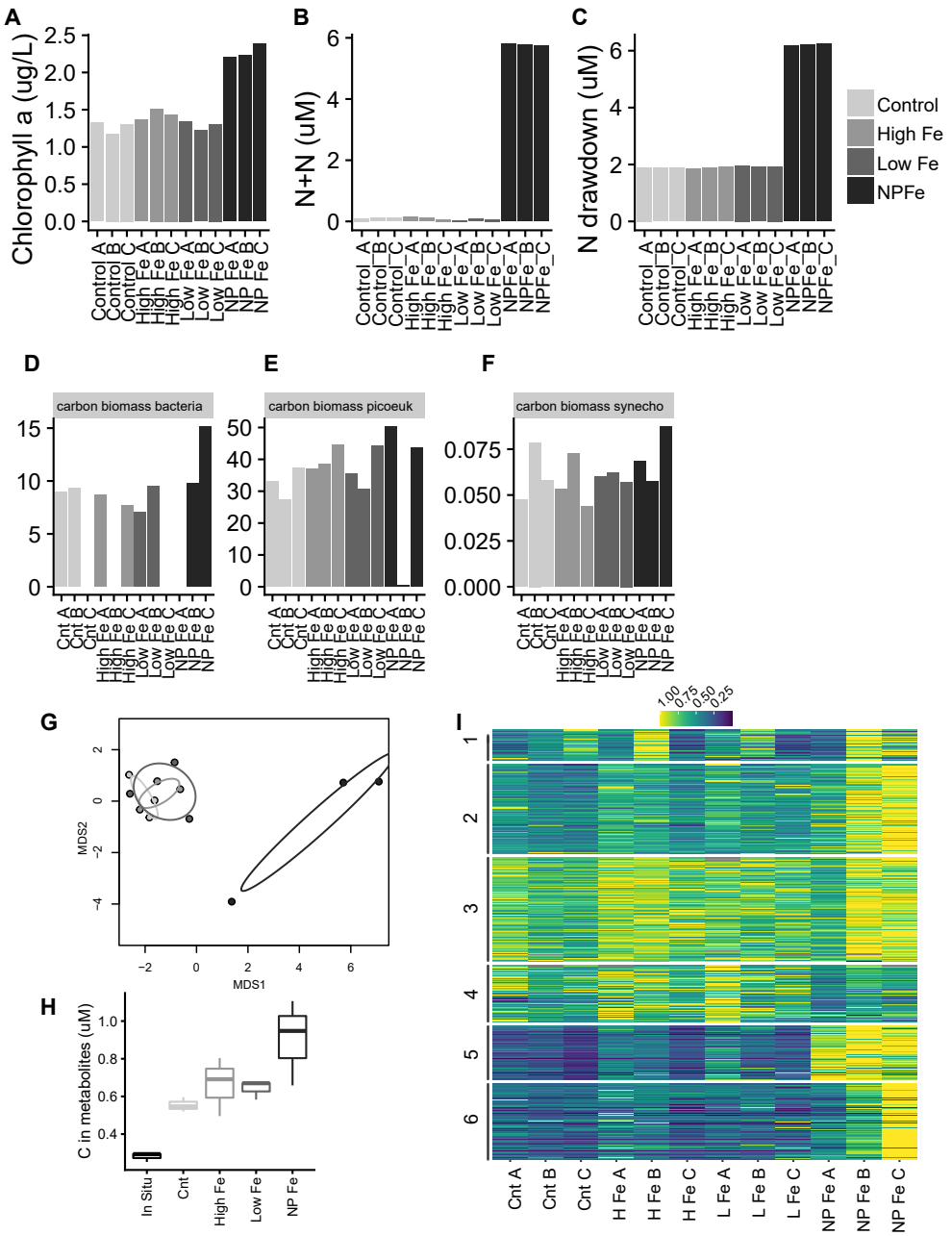

Figure 4.3: The biological response to treatments in the northern-TZ experiment (41.4 °N). Chlorophyll concentration (A), nitrate and nitrite (B), nitrate drawdown (C), bacterial biomass (D), photosynthetic picoeukaryote biomass (E), *Synechococcus* biomass (F) after four days of incubation in each replicate bottle. NMDS plot of the replicate bottles based on the metabolites measured at the end of four days (G). Total carbon in the quantified metabolites (H). Heatmaps showing the maximum normalized metabolite concentration (rows) across the treatments (columns), where metabolites are organized according to the cluster they fit into with k-medoids clustering and yellow colors indicate high concentration values and blue colors indicate low values (I). Average cluster profiles are in Figure 4.27.

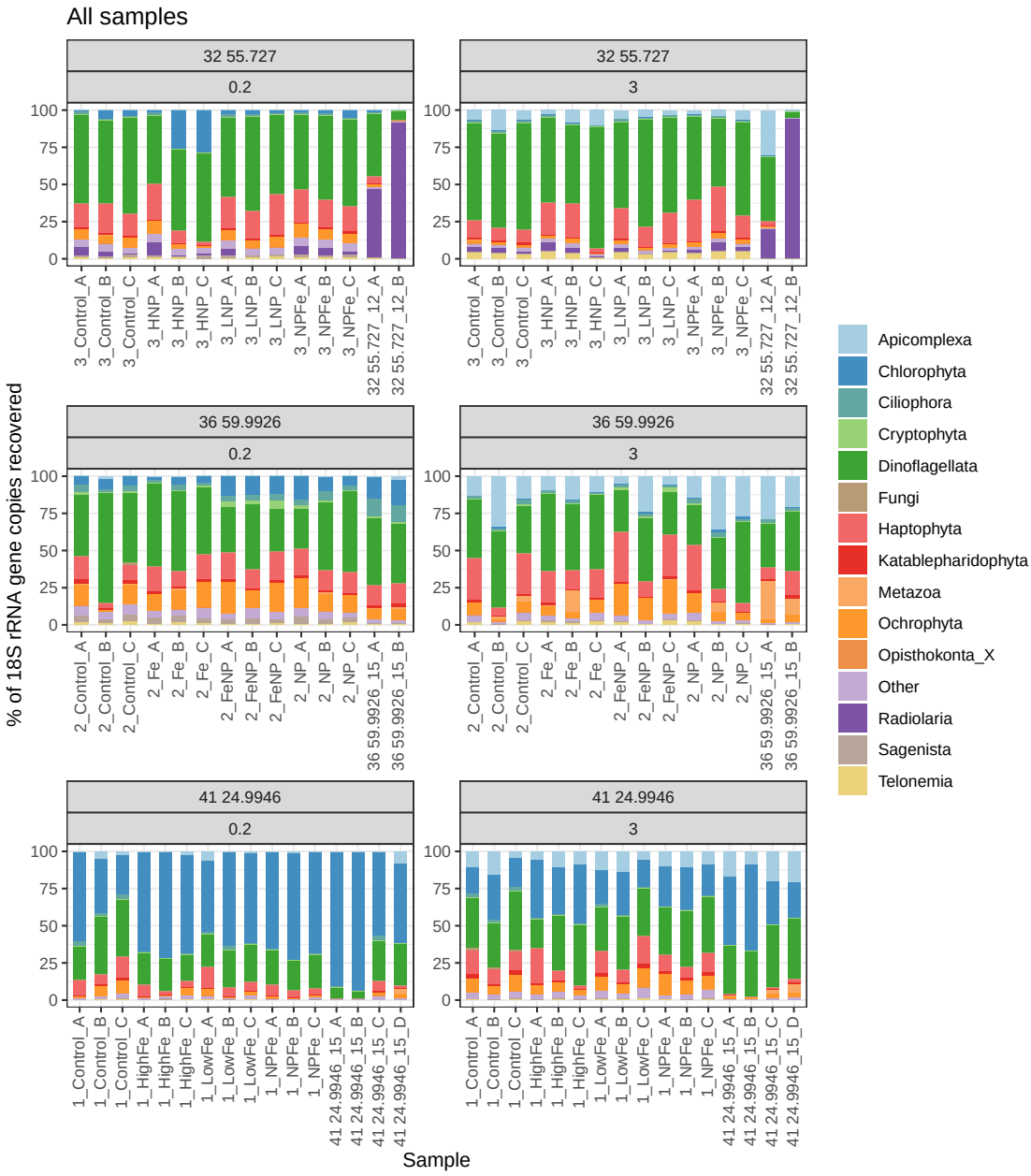

Figure 4.4: Relative abundance of 18S genes in the incubation experiments for small (left, < 3 μm), and large (right, > 3 μm) eukaryotes . Each replicate sample is shown, ordered by latitude. Classes contributing more than 5% of the total genes in at least one sample are shown, and the remaining classes are summed and shown as 'other'. Experiments are labeled by their latitude, with the corresponding *in situ* samples also shown.

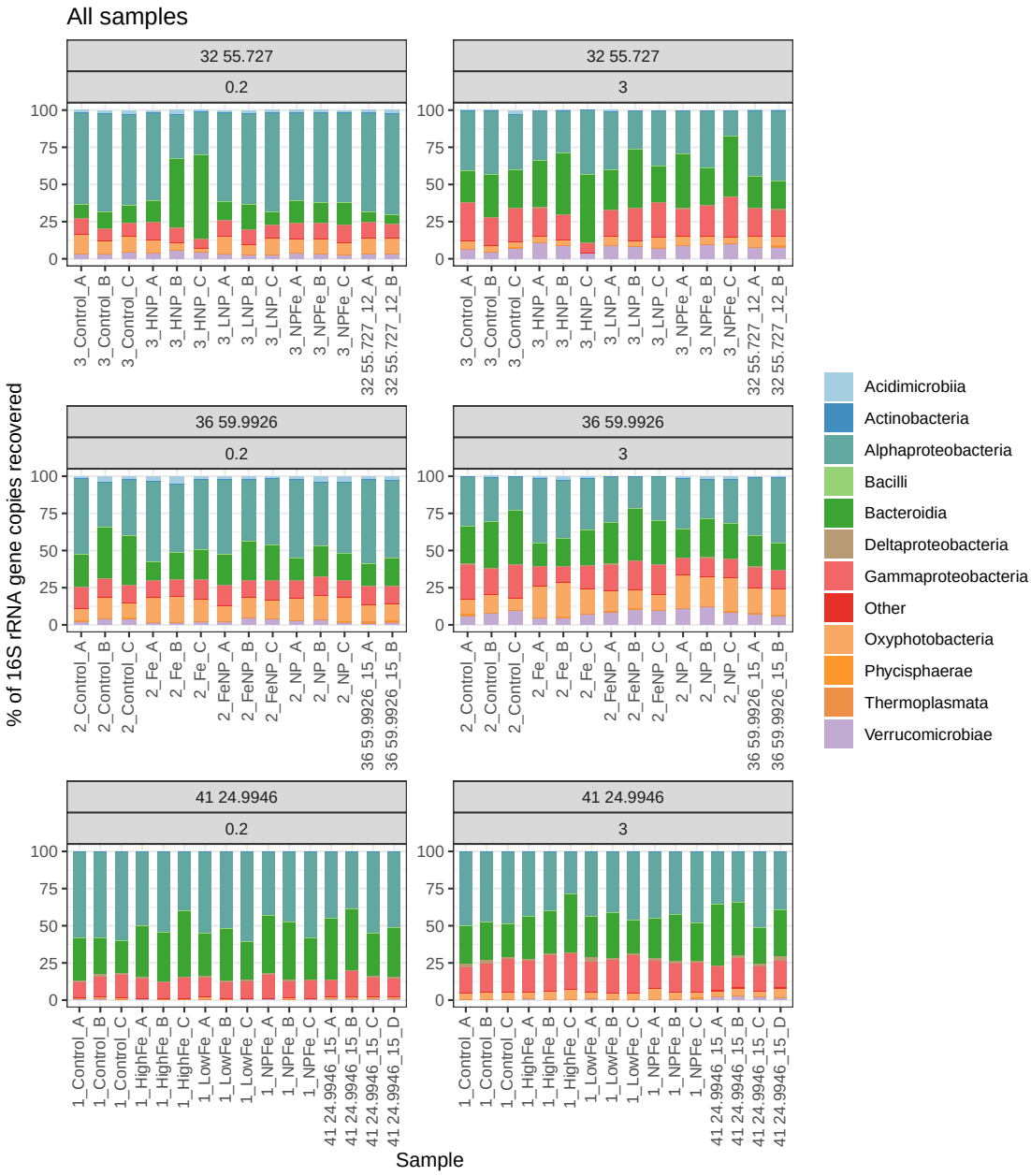

Figure 4.5: Relative abundance of 16S genes in the incubation experiments for small (left, < 3 μm), and large (right, > 3 μm) prokaryotes. Each replicate sample is shown, ordered by latitude. Classes contributing more than 5% of the total genes in at least one sample are shown, and the remaining classes are summed and shown as 'other'. Experiments are labeled by their latitude, with the corresponding *in situ* samples also shown.

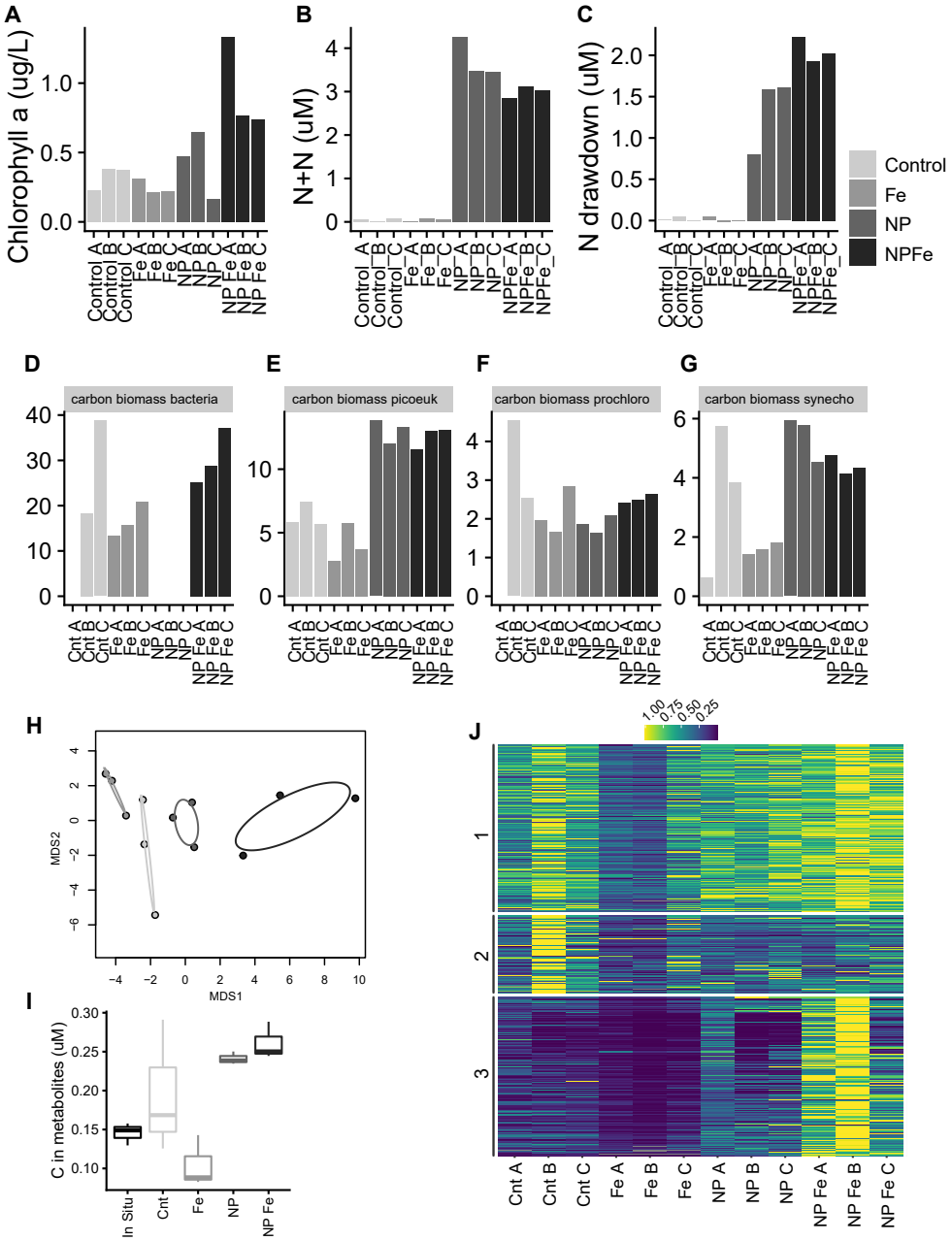

Figure 4.6: The biological response to treatments in the mid-TZ experiment (37 °N). Chlorophyll concentration (A), nitrate and nitrite (B), nitrate drawdown (C), bacterial biomass (D), photosynthetic picoeukaryote biomass (E), *Prochlorococcus* biomass (F), *Synechococcus* biomass (G) after four days of incubation in each replicate bottle. NMDS plot of the replicate bottles based on the metabolites measured at the end of four days (H). Total carbon in the quantified metabolites (I). Heatmaps showing the maximum normalized metabolite concentration (rows) across the treatments (columns), where metabolites are organized according to the cluster they fit into with k-medoids clustering and yellow colors indicate high concentration values and blue colors indicate low values (J). Average cluster profiles are in Figure 4.27.

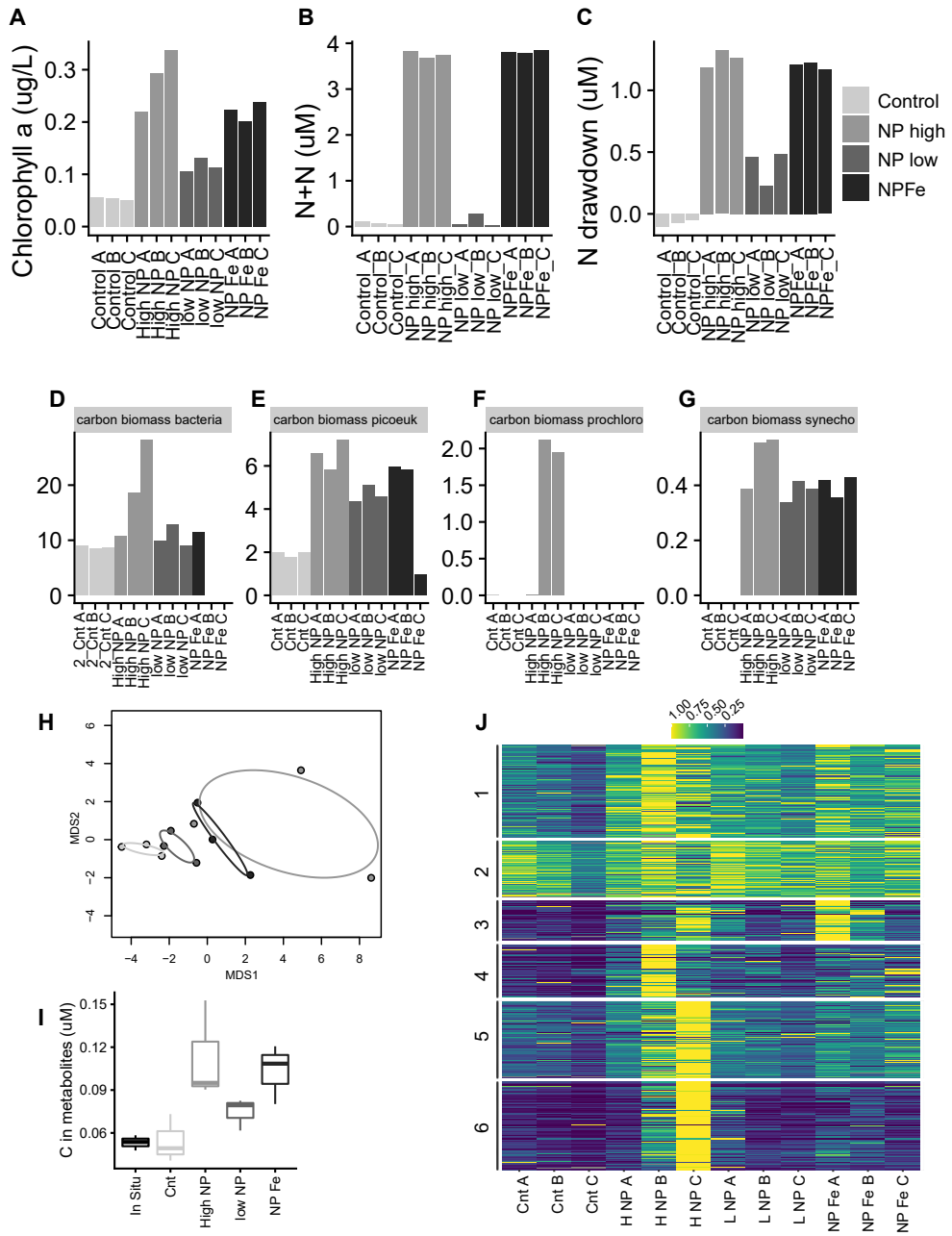

Figure 4.7: The biological response to treatments in the southern transition zone experiment (33 °N). Chlorophyll concentration (A), nitrate and nitrite (B), nitrate drawdown (C), bacterial biomass (D), photosynthetic picoeukaryote biomass (E), *Prochlorococcus* biomass (F), *Synechococcus* biomass (G) after four days of incubation in each replicate bottle. NMDS plot of the replicate bottles based on the metabolites measured at the end of four days (H). Total carbon in the quantified metabolites (I). Heatmaps showing the maximum normalized metabolite concentration (rows) across the treatments (columns), where metabolites are organized according to the cluster they fit into with k-medoids clustering and yellow colors indicate high concentration values and blue colors indicate low values (J). Average cluster profiles are in Figure 4.27.

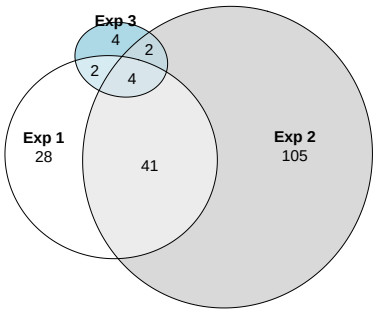

Figure 4.8: Venn Diagrams for what changes between treatments for each experiment (ANOVA tests for differences between all treatments, fdr adjusted p-value < 0.05). Exp 1 is the northern-transition zone experiment, Exp 2 is the mid-transition zone experiment, Exp 3 is the southern transition zone experiment.

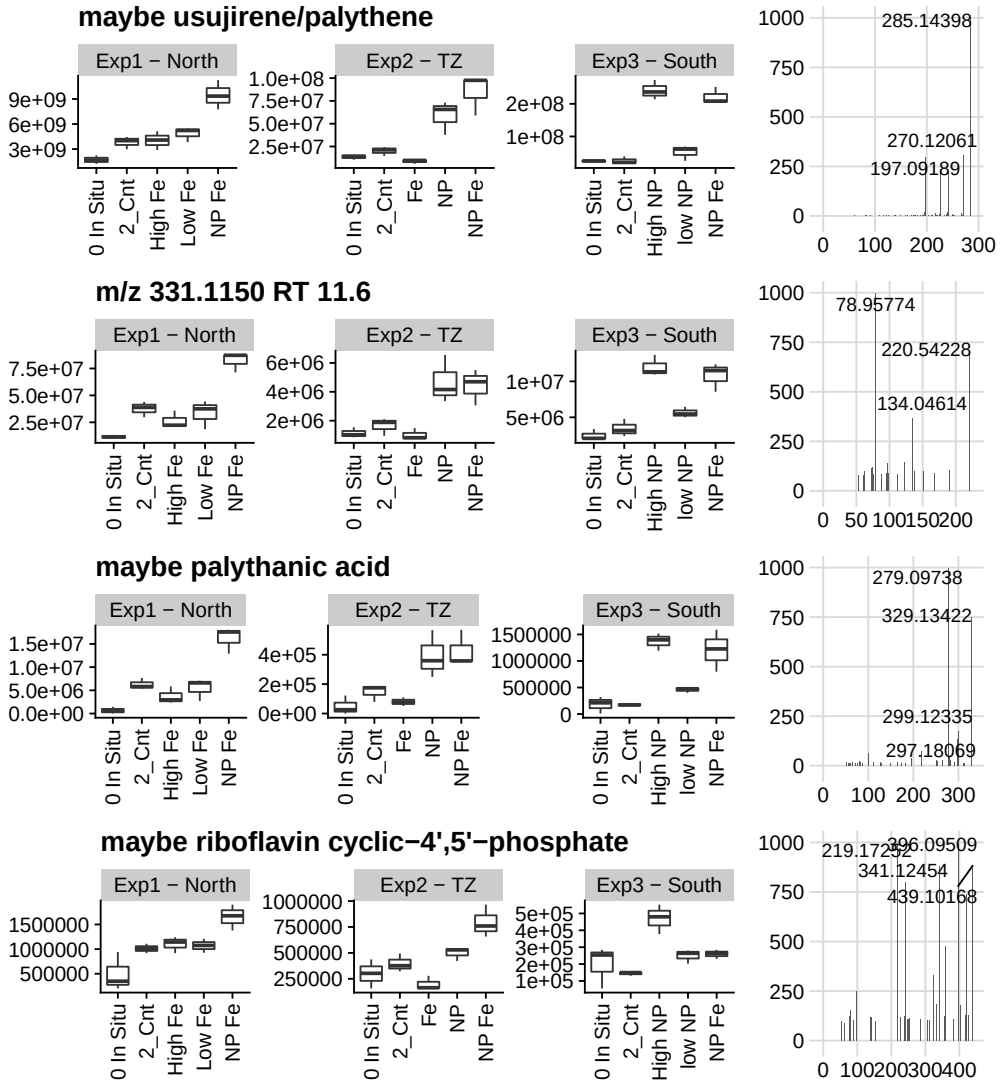

Figure 4.9: The four mass features that are significantly different between treatments in all three experiments. Relative concentrations in the northern-TZ experiment (left), mid-TZ Experiment (center left), southern-TZ experiment (center right), and their fragmentation spectra (right).

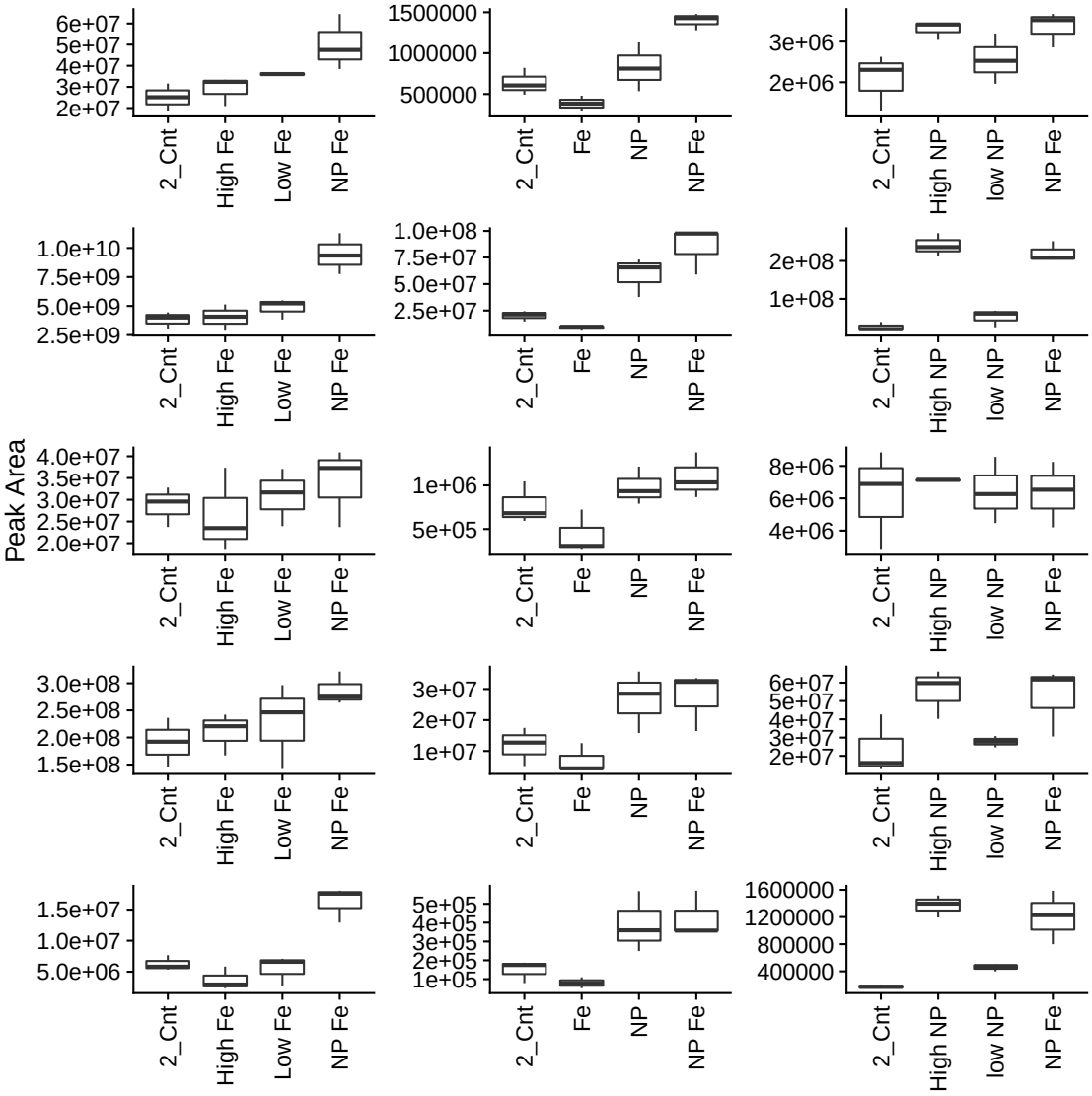

Figure 4.10: Five putative MAAs identified in the northern-TZ experiment (left), mid-TZ experiment (center), southern-TZ experiment (right). The compounds are palythinol (top, m/z 303.1553), usujirene/palythene (second from top, m/z 285.14392), asterina (center, m/z 289.13928), and two possible candidates for palythanic acid (second from bottom and bottom, m/z 329.1345).

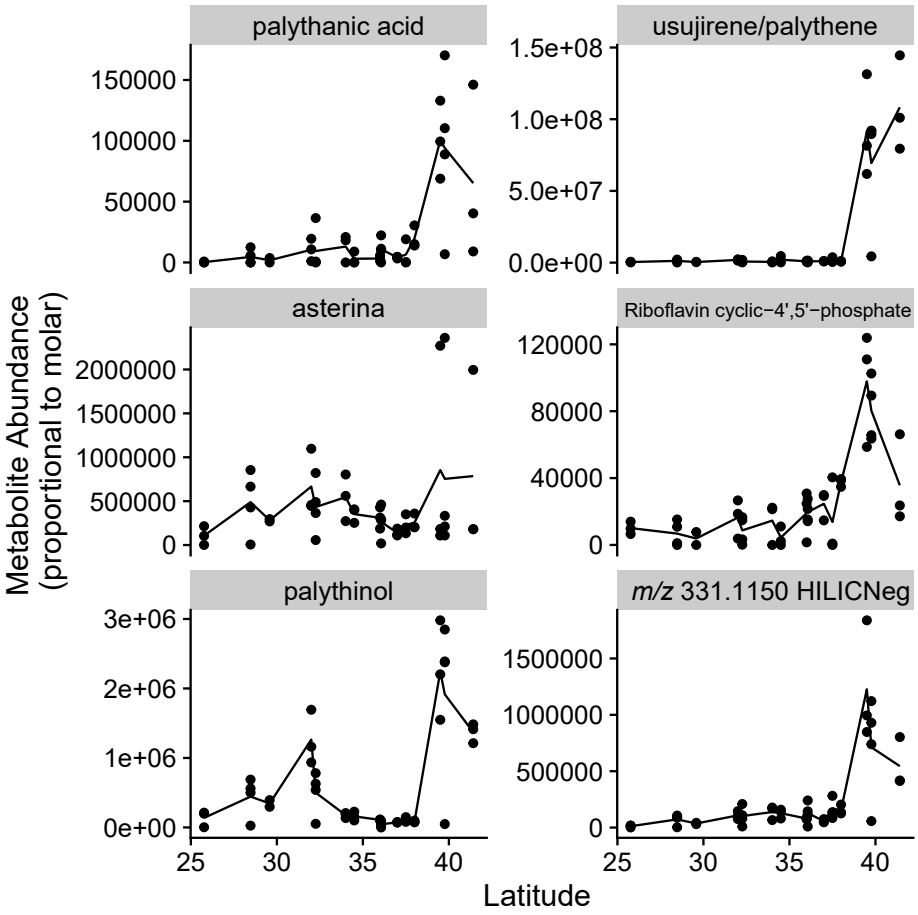

Figure 4.11: Relative concentrations in the surface seawater along the transect of untargeted metabolites that are potential MAAs and two other untargeted metabolites that were significantly different between treatments in all three incubation experiments, and unknown metabolite with m/z of 331.1150 and a metabolite with m/z of 439.10162 that is potentially riboflavin cyclic-4',5'-phosphate.

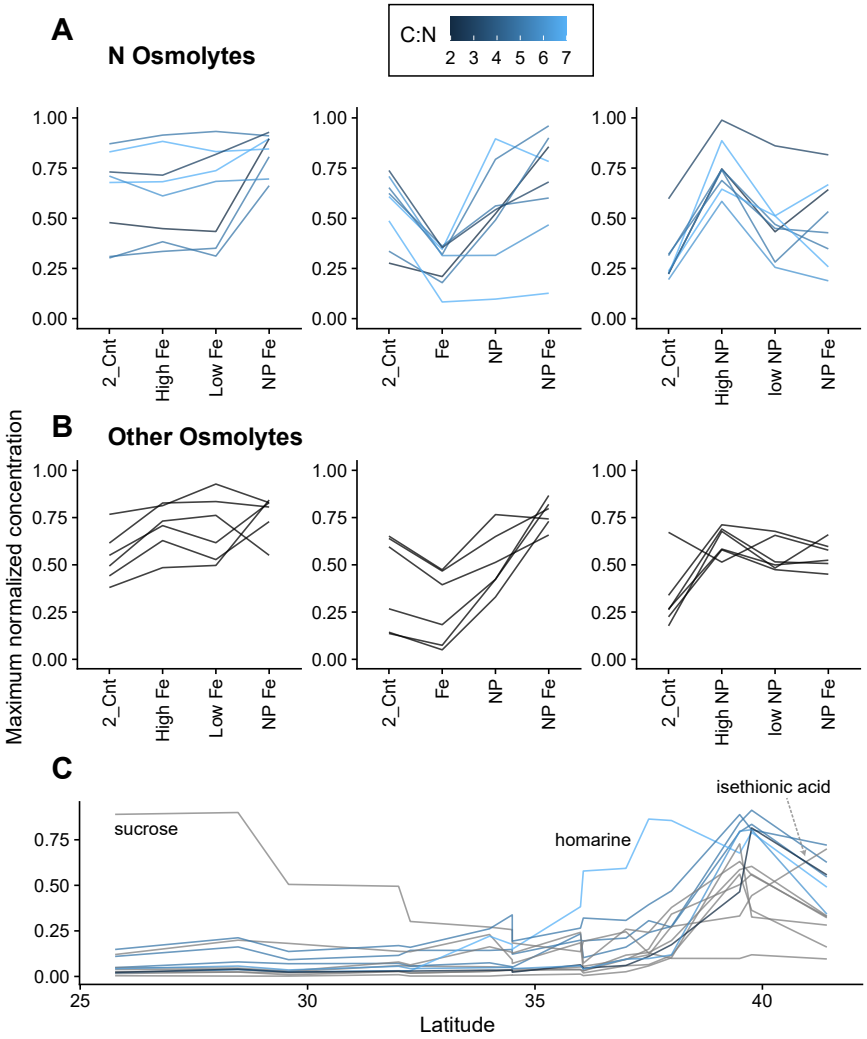

Figure 4.12: Maximum normalized concentrations of osmolytes that contain N (A) and osmolytes that do not contain N (B) in the different treatments of the northern-TZ experiment (left), mid-TZ experiment (center), and southern-TZ experiment (right). The color of the line represents the C to N ratio of the nitrogenous osmolytes. N containing osmolytes in the experiments are homarine, proline, glycine betaine, proline betaine, glutamic acid, TMAO, TMAP, taurine. Other osmolytes in the experiments are glucosylglycerol, DMS-Ac, DMSP, gonyol, DHPS, and isethionic acid . Maximum normalized osmolytes across the transect with latitude (C). Osmolytes plotted in the transect are sucrose, homarine, gonyol, proline, DMS-Ac, DMSP, glutamic acid, DHPS, glucosylglycerol, isethionic acid, taurine, TMAP, glycine betaine.

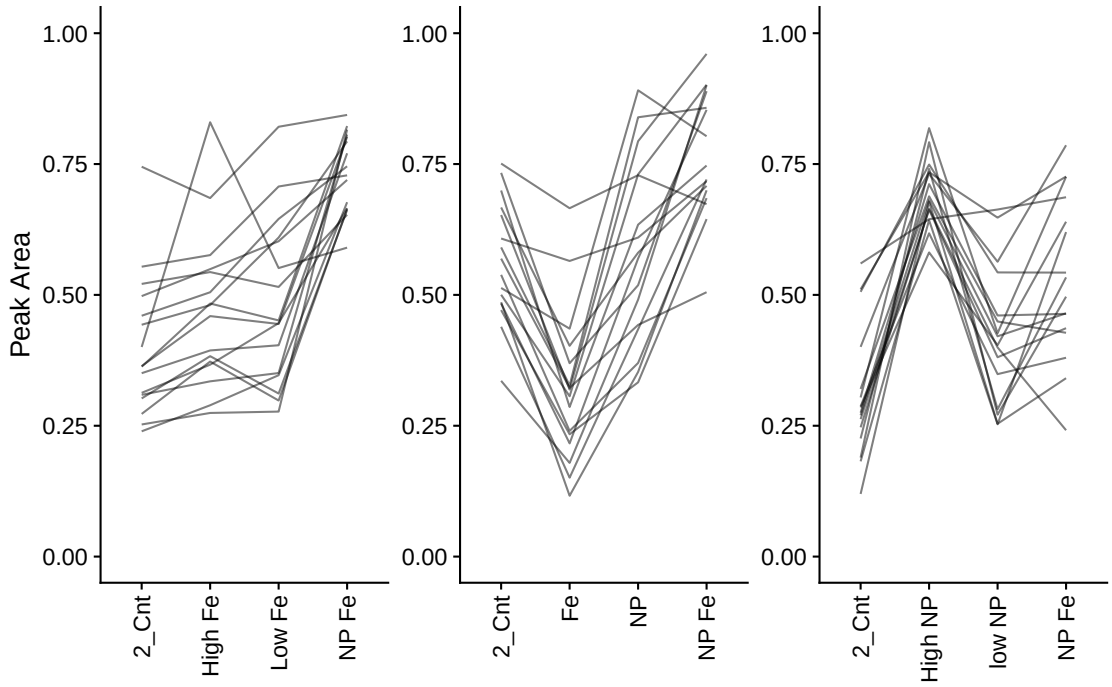

Figure 4.13: Maximum normalized concentrations of amino acids in the different experiments. Northern-TZ experiment (left), mid-TZ experiment (center), southern-TZ experiment (right).

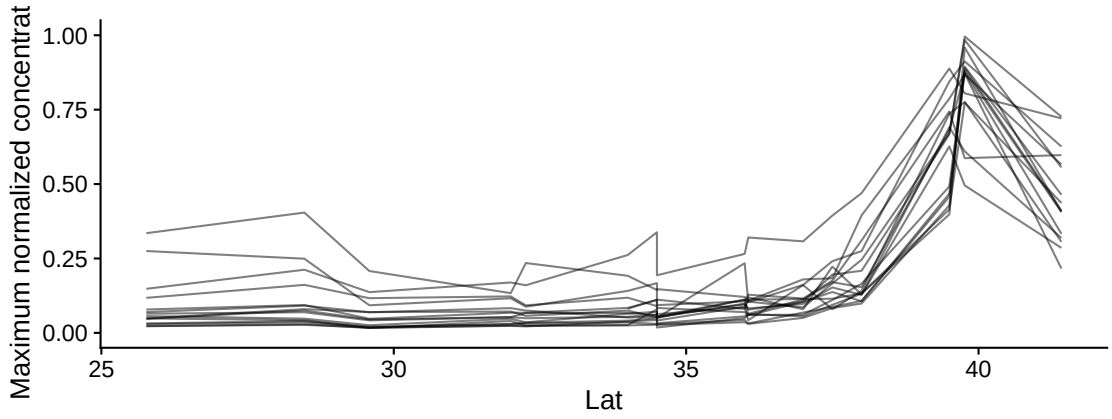

Figure 4.14: Maximum normalized concentrations of amino acids in the transect.

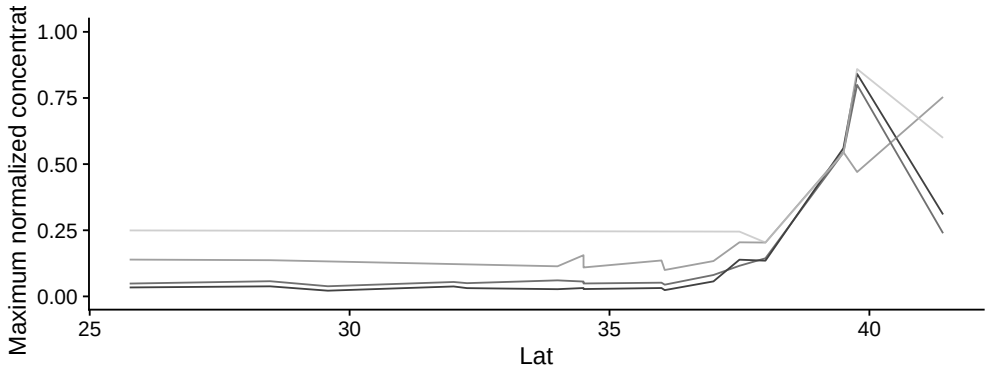

Figure 4.15: Maximum normalized concentrations of carnitines across the transect.

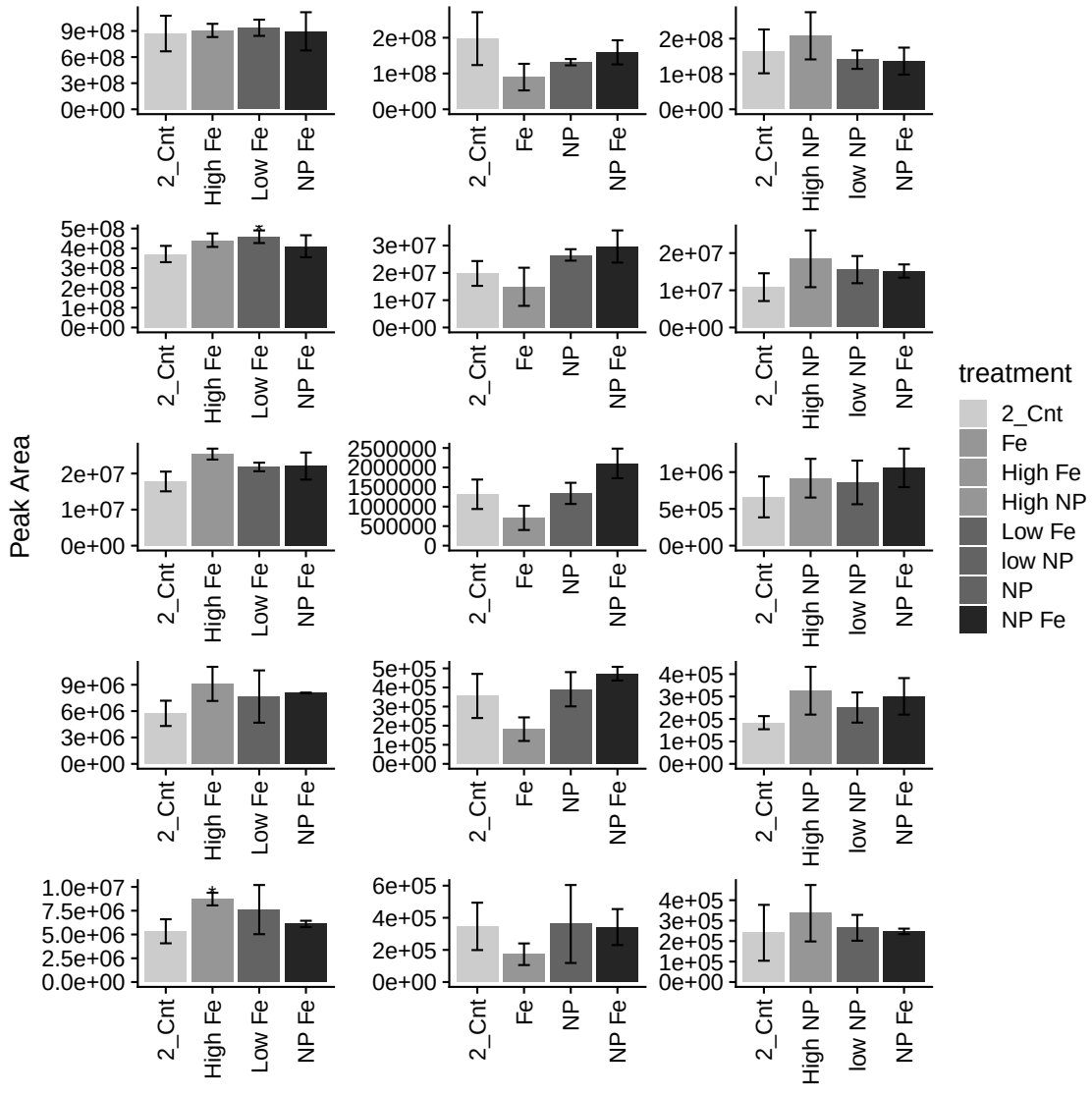

Figure 4.16: Five carnitines identified in the northern-TZ experiment (left), mid-TZ experiment (center), southern-TZ experiment (right). The compounds are carnitine (top), acetyl-L-carnitine (second from top), propionyl-L-carnitine (center), butyryl-L-carnitine (second from bottom), and the putative C5-acylcarnitine (bottom).

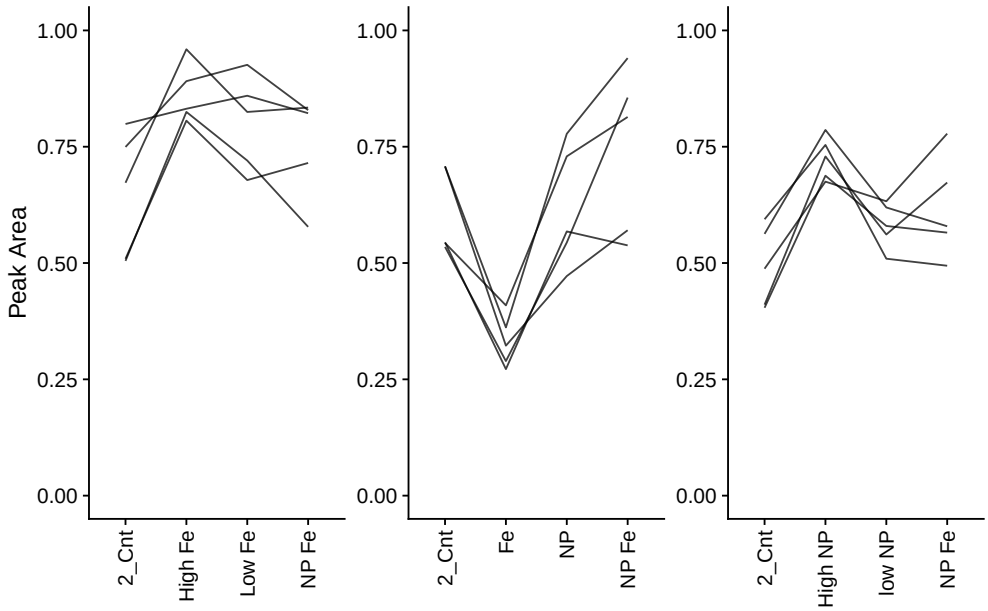

Figure 4.17: Maximum normalized concentrations of carnitines in the different experiments. Northern-TZ experiment (left), mid-TZ experiment (center), southern-TZ experiment (right).

4.10 Supplemental Methods

4.10.1 Details of incubation experiment set up

In addition to the details in the main text, the following protocol was used to set up the nutrient amendment incubations. All of the macronutrient stocks were passed through a Chelex-100 column prior to use to clean the nutrient stocks for contaminating trace metals. All carboys were filled in a Class 100 clean hood with positive pressure. Each carboy was rinsed 3 times with the 100 μM mesh filtered water prior to filling the carboy to the 20L mark (with headspace). All carboy caps were wrapped with parafilm and then electrical taped to prevent incubator water from entering the carboys. Carboys were immediately placed in dark trash bags before being transported to the incubator, and were placed in the incubator as quickly as possible in order to limit direct sunlight exposure and related stress. Carboys were gently inverted three times prior to sampling for the final time points.

4.10.2 Metabolomics sample processing and data analysis

For untargeted metabolomics processing, MSDial [185] was used to pick peaks, align data files, and integrate the mass features from each sample set, chromatography column, and ionization mode individually. We manually inspected the alignment results from MSDial to identify the 100 most abundant mass features with MSMS in each chromatography-ionization pairing in the experimental dataset and the 50 most abundant mass features in the transect dataset, the 50 mass features most different between the N and S in each chromatography-ionization pairing of the transect dataset, and the 100 mass features in each chromatography-ionization pairing that were most different between the treatments in the experiments in the experiment dataset. The integrated peak areas were then exported for further quality control which retained only mass features with relative standard deviation in repeat injections of a quality control sample of less than 100% and an average peak area in the samples greater than 5 times the average peak area in the blank filters. Skyline was used to export the masses of heavy-isotope labeled internal standards in HILIC and RP modes. These internal

standards were used to normalize the full untargeted dataset following the best-matched internal standard normalization in Boysen et al. [22]. Finally, mass features were retained if they had a less than 20% relative standard deviation in their normalized peak areas in repeat injections of the quality control samples. Mass features were dereplicated within each fraction (LC-ionization mode pair) in order to remove likely adducts by doing pairwise comparison of RT and m/z to identify likely candidates and then manually inspecting peak shape and size to confirm. Additional dereplication across ionization modes and columns used hierarchical clustering of peak areas across the samples to identify mass features duplicated due to identification in multiple ionization modes. The mass features that were identified as potential replicates and adducts were again manually inspected to determine if they were indeed replicates. Ultimately 392 mass features were retained through this processing for the experiment dataset. An additional 50 known metabolites not identified with MSDial were integrated with Skyline to bring the total number of high quality metabolites in the experiment dataset to 435. Ultimately 193 mass features were retained through this processing for the transect dataset. An additional 56 metabolites not picked with MSDial were integrated with Skyline to bring the total number of high quality metabolites in the transect dataset to 240.

4.10.3 Metabolite quantification

Authentic standards of 267 compounds were spiked into water and into a sample matrix and analyzed at the same time as the environmental samples. The metabolites we detected that had corresponding to authentic standards were then quantified according to the equation below:

$$Concentration = \frac{PeakArea}{IE} * vialvol * \frac{dilutionfactor}{volfiltered} * RFratio^{-1} \tag{4.1}$$

Where $RFratio$ (response factor ratio) is the difference between the peak area of a standard spiked into environmental matrix less the peak area in un-spiked matrix, divided by the peak area of the standard in water, and the IE (ionization efficiency) is the ratio of the peak

area of the standard in water over the concentration of the standard. The samples from the transect were analyzed in 2017, prior to purchasing some of the standards, and in these cases the RF-ratios and IE values from the 2018 sample run of the nutrient amendment incubation samples. During the 2018 sample run select samples from the transect were re-injected and checked for similarity with the 2017 run, to make sure peak areas and thus concentrations are comparable and not subject to instrumentation variability. Concentrations of metabolites were converted into moles C and normalized to POC by multiplying the concentration of the compound by the number of carbons in each compound and dividing by POC.

4.10.4 Amplicon data processing

DNA extracts were processed and analyzed at the University of Illinois W. M. Keck Center, where analysis involved amplicon-based library preparation using a Fluidgm Access Array™ system combined with Illumina MiSeq™ v3 sequencing set to a run length of 2 x 250bp. The following primer sets were used: a revised version of both the F515 (5'-GTGYCAGCMGCCGCGGTAA) and R806 (5'-GGACTACNVGGGTWTCTAAT) primers targeting the hypervariable region 4 (V4) of the 16S rRNA gene [214], and the F566 (5'-CAGCAGCCGCGGTAATTCC) and R1200 (5'-CCCGTGTTGAGTCAAATTAAGC) V4-V5 universal 18S rRNA primer set [215]. Primer sorting, demultiplexing, and removal of the Fluidgm CS1 and CS2 universal tags and Illumina i5 and i7 adapters were part of sequence preprocessing.

Analysis of 18S rDNA sequences was carried out using the Qiime2 software suite [186]. Raw sequences derived from the forward read were adapter trimmed using the Cutadapt [216] plugin with the options: –p-adapter AGACCAAGTCTCTGC CAAGCAGAAGACGGCAT-ACGAGAT CCCGTGTTGAGTCAAATTAAGC, –p-front CAGCAGCCGCGGTAATTCC ACACTGACGACATGGTTCTACA AATGATACGGCGACCACCGAGATCT, –p-discard-untrimmed, –p-no-indels and –p-error-rate=0.2. Read quality was then assessed through the built-in demux summarize function and a truncation length of 190 nt was selected. Truncated reads were then denoised using DADA2 [217] resulting in Amplicon Sequence Variants

(ASVs). ASVs were classified using VSEARCH [218] with the PR^2 database as a reference (v4.12.0; https://pr2-database.org/) [219].

The analysis pipeline for 16S rDNA sequences was also carried out using Qiime2 [186] with notable differences. Paired-end reads were trimmed using Cutadapt [216] with the options: –p-front-f GTGYCAGCMGCCGCGGTAA ACACTGACGACATGGTTCTACA AATGATACGGCGACCACCGAGATCT, –p-adapter-f TACGGTAGCAGAGACTTGGTCT CAAGCAGAAGACGGCATACGAGAT AGACCAAGTCTCTGC, –p-front-r GGACTACNVGGGTWTCTAAT ACACTGACGACATGGTTCTACA AATGATACGGCGACCACCGAGATCT, –p-adapter-r TACGGTAGCAGAGACTTGGTCT CAAGCAGAAGACGGCATACGAGAT TGTAGAACCATGTC, –p-discard-untrimmed, –p-no-indels, and –p-error-rate 0.2. Read quality for both forward and reverse reads was assessed using the demux plugin resulting in truncation lengths of 200 nt and 170 nt for forward and reverse reads respectively. ASVs were obtained using DADA2 [217] and classified with VSEARCH using the SILVA database clustered at 99% identity (release 132) as a reference [220]. Data were filtered to exclude eukaryotes, chloroplast, and mitochondrial sequences.

4.11 Supplemental Tables

Table 4.3: The percent of each metabolite in the transect data that was in the 0.2 - 3 μm size fraction compared to the total concentration measured (whole seawater $>$ 0.2 μm) at that station. Percent is calculated based on the single small size fraction sample divided by the mean of the triplicate whole seawater samples. The standard deviation of the percent is based on the standard deviation of the triplicate whole seawater samples.

Table 4.4: The cluster assignment in the experiments for each metabolite. MassFeature is the primary identification for each metabolite. FinalBMIS is the internal standard used to normalize the metabolite. Alighnemt.ID is used to identify unqiue metabolites in each fraction. Average.Rt, Average.Mz, MS.MS.assigned are for peaks picked by the MSDial analysis of the experiment data. If a MSMS was assigned, the specra is in MS.MS.spectrum. Fraction is the LC column and ionization mode used for detection. Metabolite.name is any additional identification. Cluster assignments for the three experiments are in the columns North.cluster, South.cluster, and TZ.cluster (for the mid-TZ experiment). There are equivalent columns for whether each mass feature was significantly different between experiments based on an ANOVA test and fdr corrected p-value < 0.05. N.Diff is the number of experiments that the mass feature was significantly different in.

Table 4.5: The sample IDs, stations, CTD cast (event), Depth, and date for each experiment sample and relevant blanks.

Table 4.6: The cluster assignment in the transect for each metabolite. MassFeature is the primary identification for each metabolite. Transect.cluster is the cluster assignment for based on the transect data. Metabolite.name and OtherID are other identifiers used to link to the experiment dataset.

Table 4.7: The relative abundance of the metabolites in each experiment and relevant blank samples. Values are proportional to molar concentration.

Table 4.8: The relative abundance of the metabolites in the transect data and relevant blank samples. Samples with "Dsmall" are the small size fraction samples. Water-vol.Norm.Area is the normalized relative abundance that is proportional to molar concentration. Area.PC.norm is the relative abundance data proportional to the particulate carbon for locations where we had paired metabolite and particulate carbon data.

4.12 Supplemental Figures

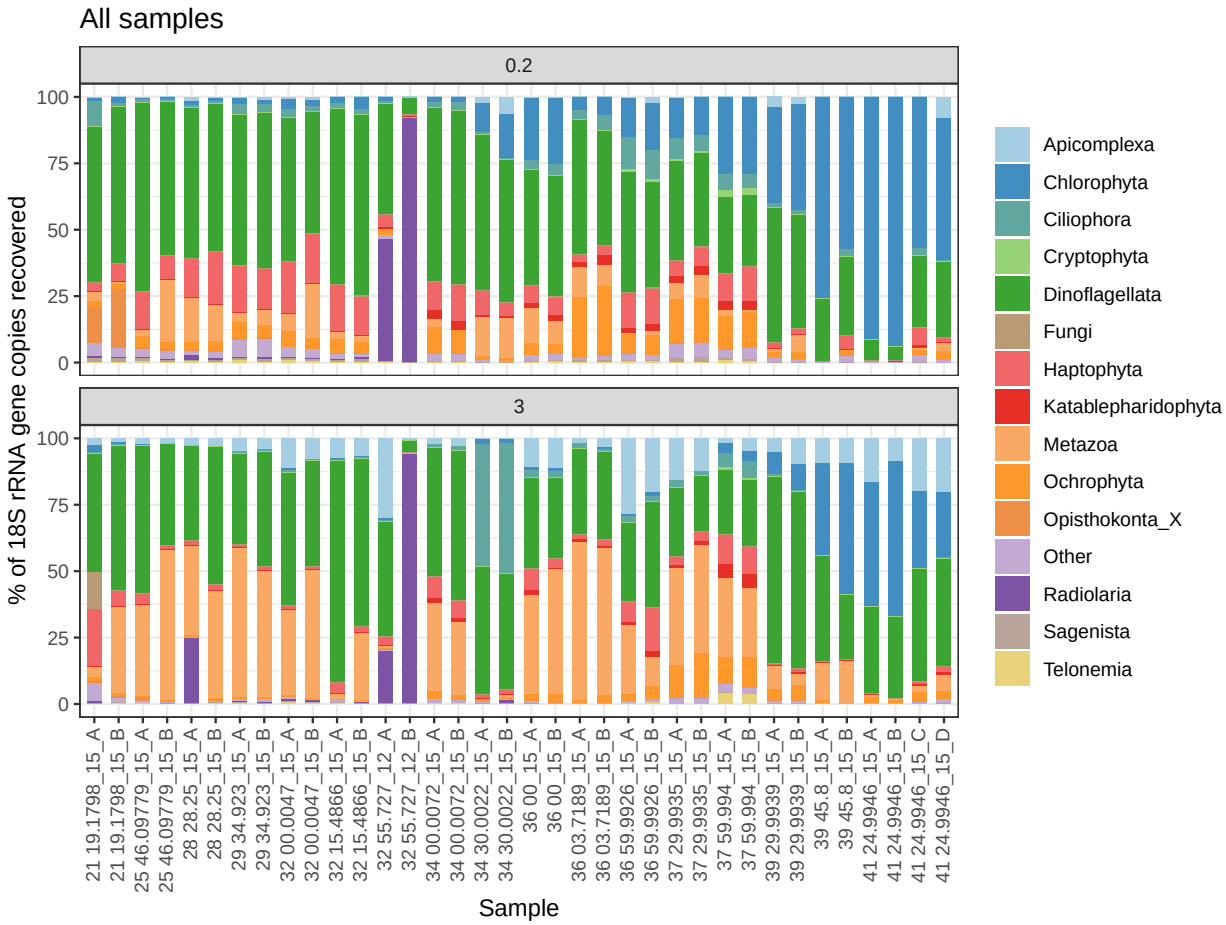

Figure 4.18: Relative abundance of 18S genes over the cruise transect for small (top, 0.2–3 μm), and large (bottom, $>$ 3 μm) eukaryotes . Each replicate sample is shown, ordered by latitude. Classes contributing more than 5% of the total genes in at least one sample are shown, and the remaining classes are summed and shown as 'other'. Sample IDs along the x-axis show the latitude of collection, the depth of collection in meters, and the replicate number.

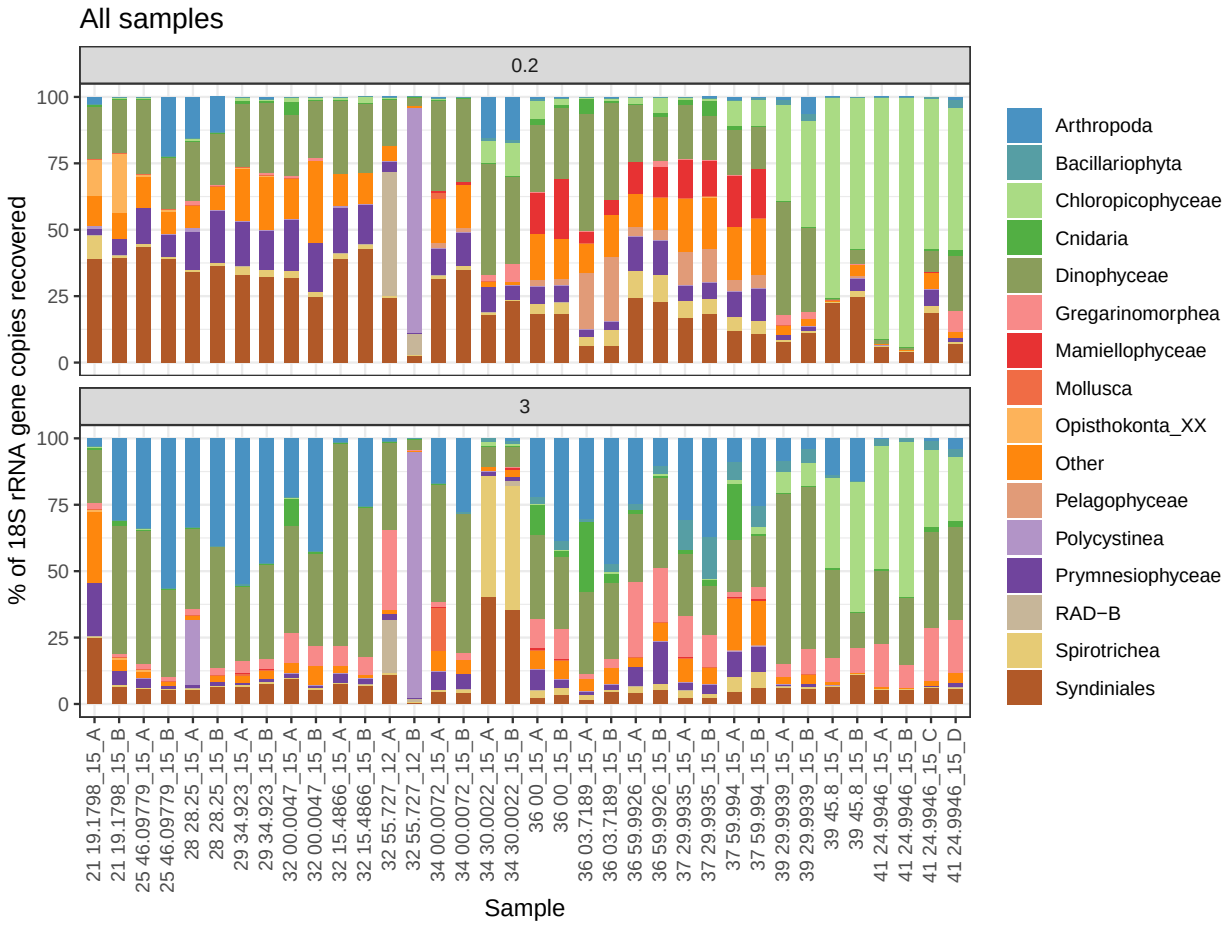

Figure 4.19: Relative abundance of 18S genes over the cruise transect for small (top, 0.2–3 μm), and large (bottom, $>$ 3 μm) eukaryotes . Each replicate sample is shown, ordered by latitude. Orders contributing more than 15% of the total genes in at least one sample are shown, and the remaining orders are summed and shown as 'other'. Sample IDs along the x-axis show the latitude of collection, the depth of collection in meters, and the replicate number.

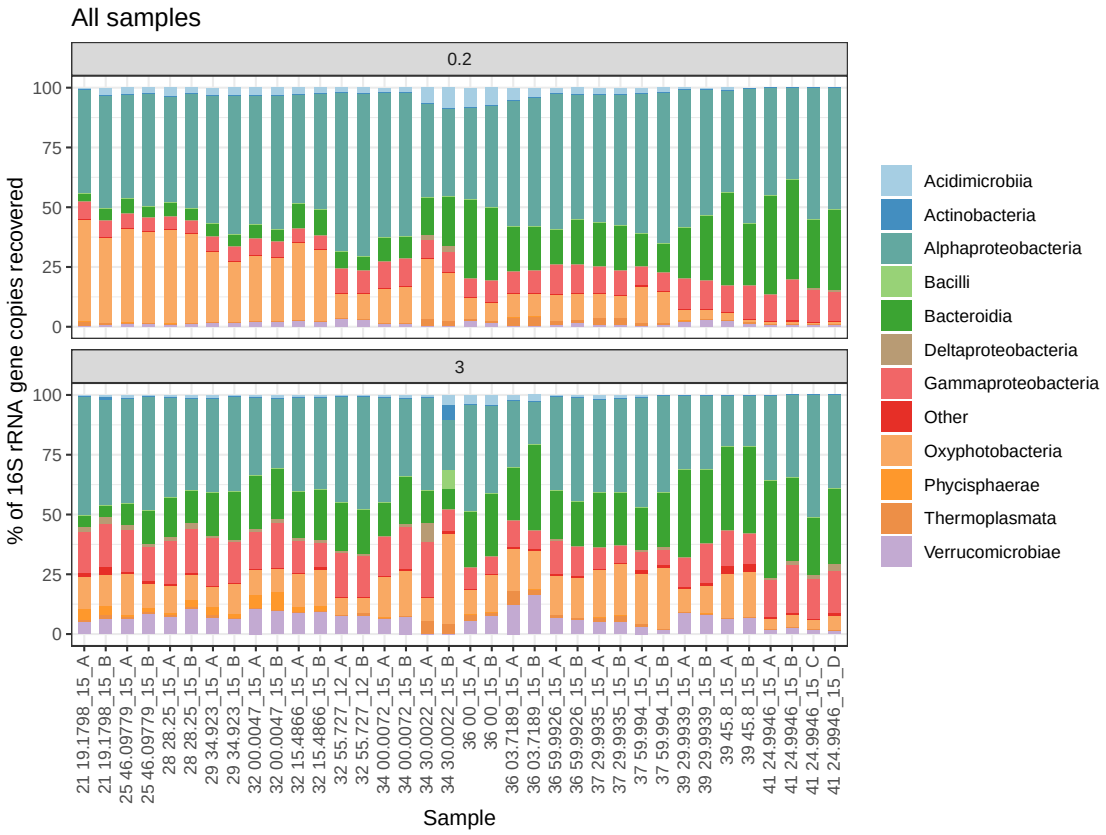

Figure 4.20: Relative abundance of 16S genes over the cruise transect for small (top, 0.2–3 μm), and large (bottom, $>$ 3 μm) prokaryotes. Each replicate sample is shown, ordered by latitude. Classes contributing more than 5% of the total genes in at least one sample are shown, and the remaining classes are summed and shown as 'other'. Sample IDs along the x-axis show the latitude of collection, the depth of collection in meters, and the replicate number.

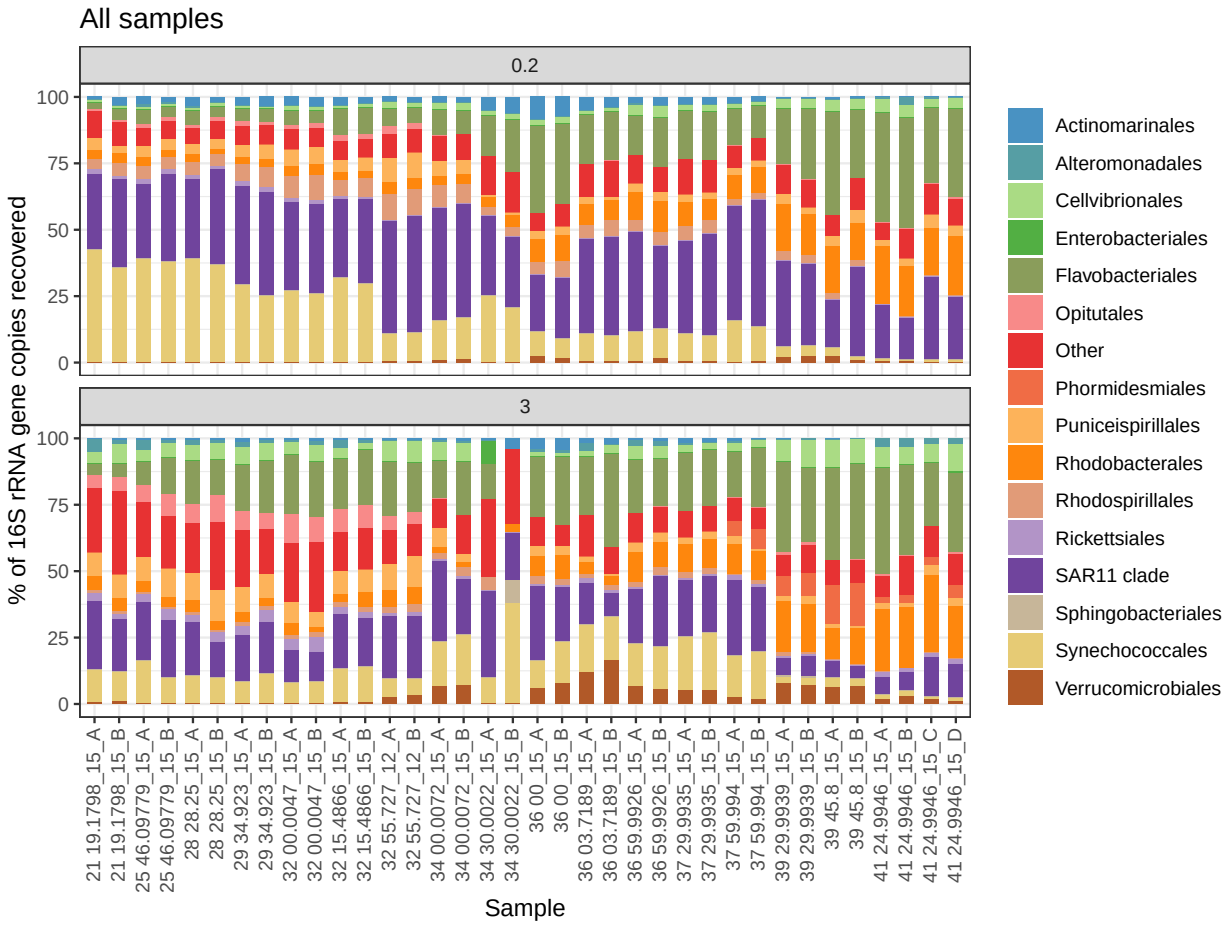

Figure 4.21: Relative abundance of 16S genes over the cruise transect for small (top, 0.2–3 μm), and large (bottom, $>$ 3 μm) prokaryotes. Each replicate sample is shown, ordered by latitude. Orders contributing more than 8% of the total genes in at least one sample are shown, and the remaining orders are summed and shown as 'other'. Sample IDs along the x-axis show the latitude of collection, the depth of collection in meters, and the replicate number.

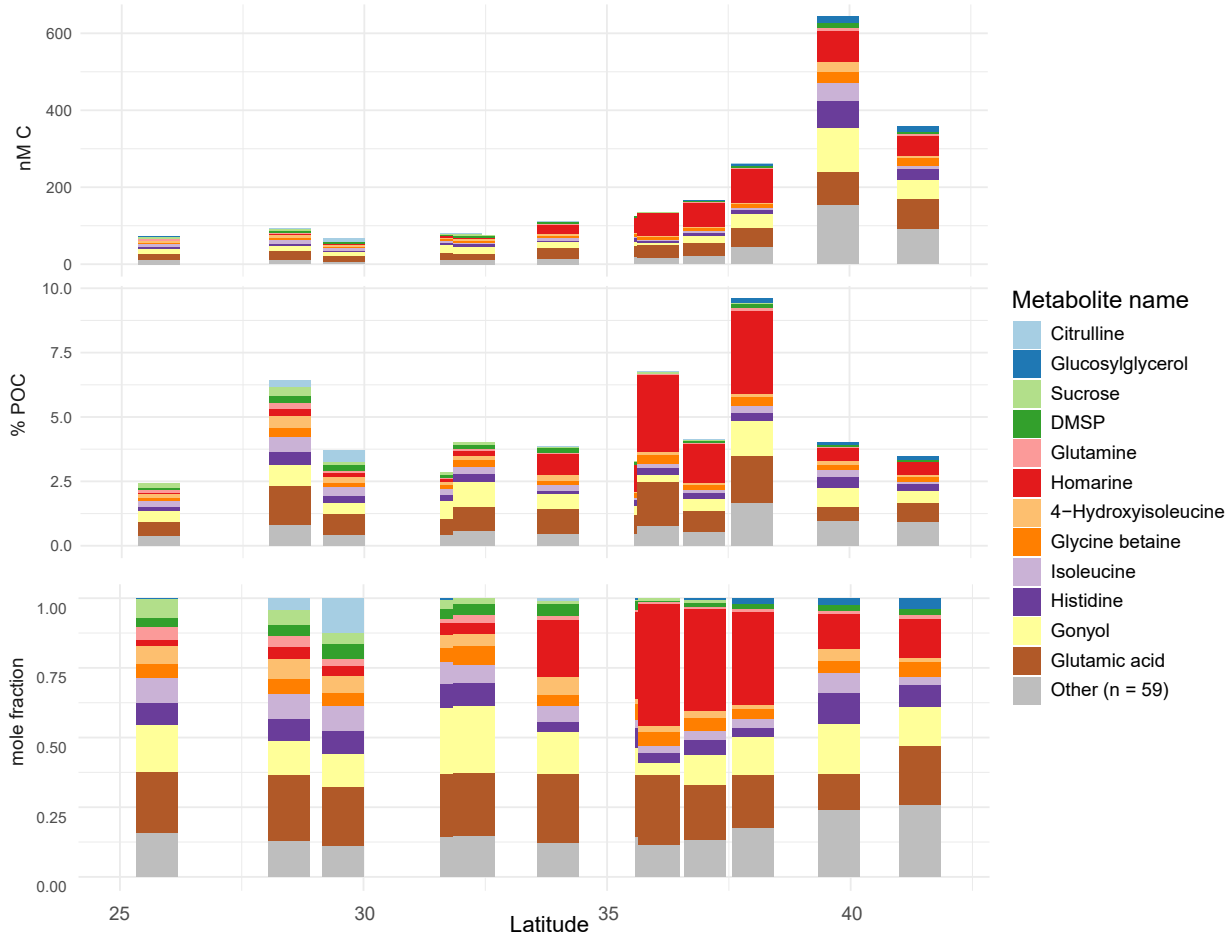

Figure 4.22: Particulate concentration of the metabolites for which we have quantitative concentration data over the cruise transect, reported in nM carbon (top) and as a percent of POC (middle) and as mole fraction of total quantified metabolites (bottom). All samples shown were collected at 4 am local time.

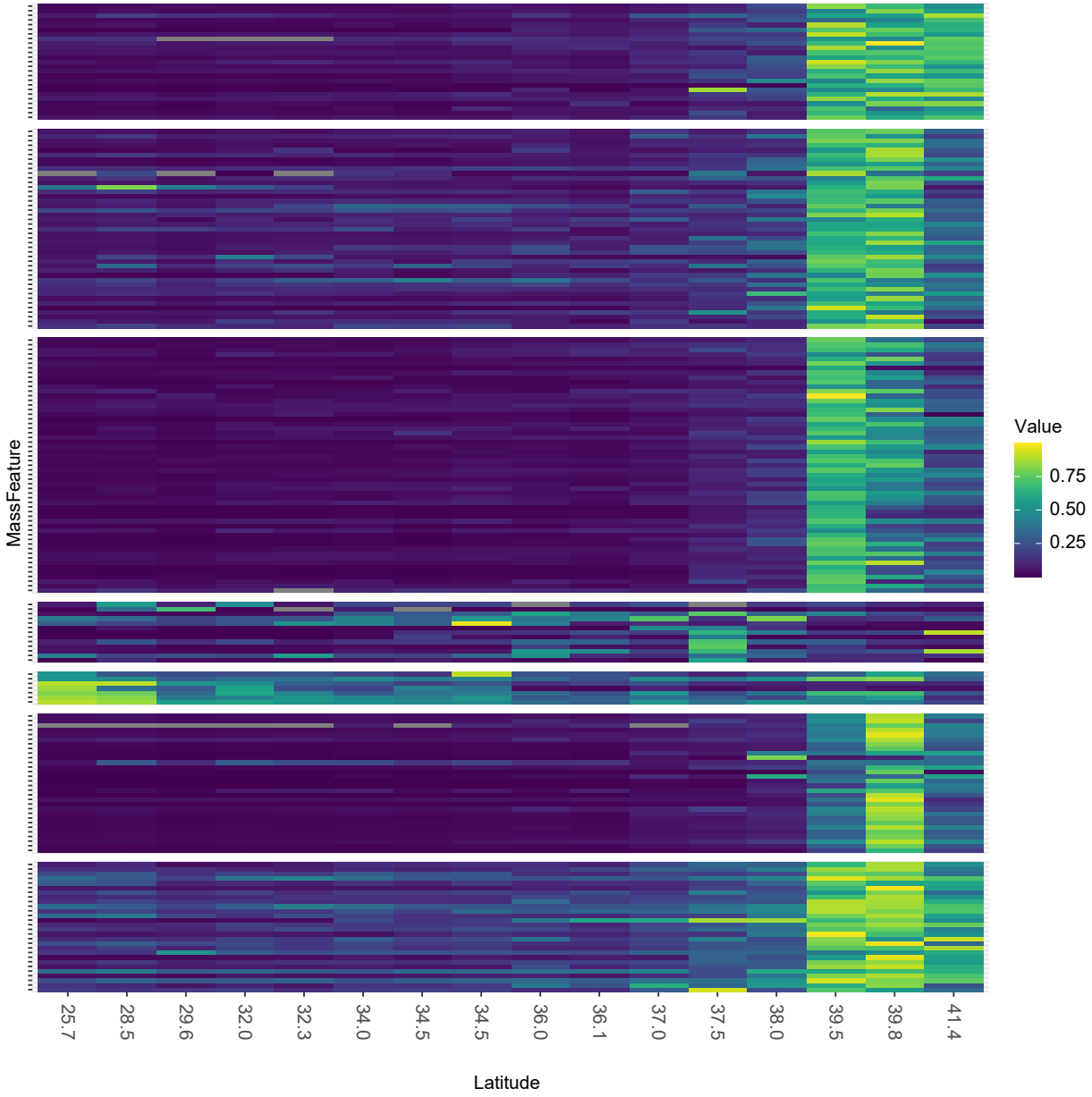

Figure 4.23: A heatmap of metabolite signals in mixed layer across the transect. Averaged values of biological replicates are shown (columns). Metabolite signals (rows) are organized according to their cluster assignment. Color indicates the concentration or abundance of that signal, relative to its maximum value in the transect, with values close to the maximum in yellow and low values much smaller than the maximum value in blue.

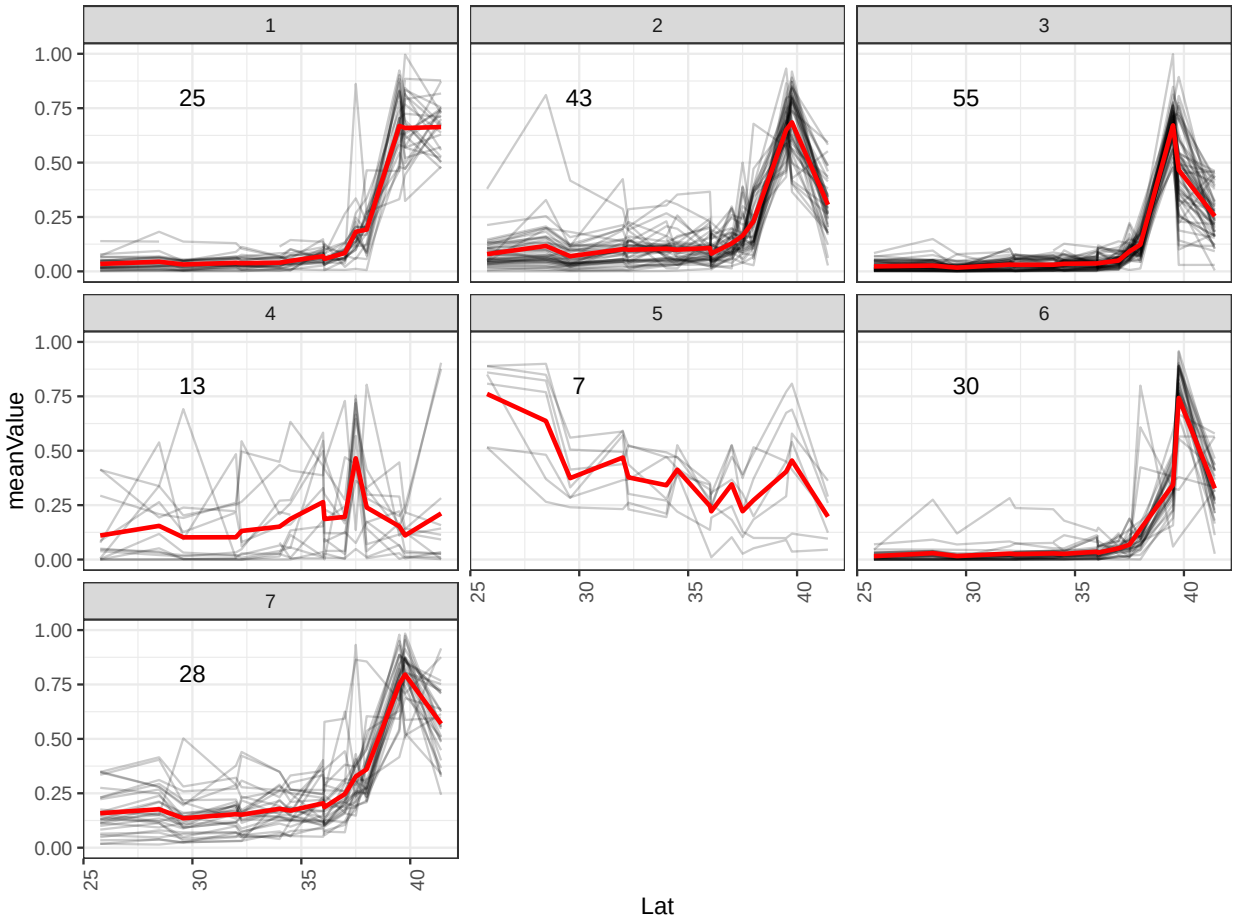

Figure 4.24: Average metabolite cluster profiles over latitude for the transect data. Metabolite abundance was maximum transformed relative to the individual sample with the highest concentration for that metabolite. The number of mass features in each cluster is indicated in the number in the plot. The grey lines are the individual metabolite signals, averaged across biological replicates at each station and depth. The red lines are the average profile from all signals in that cluster.

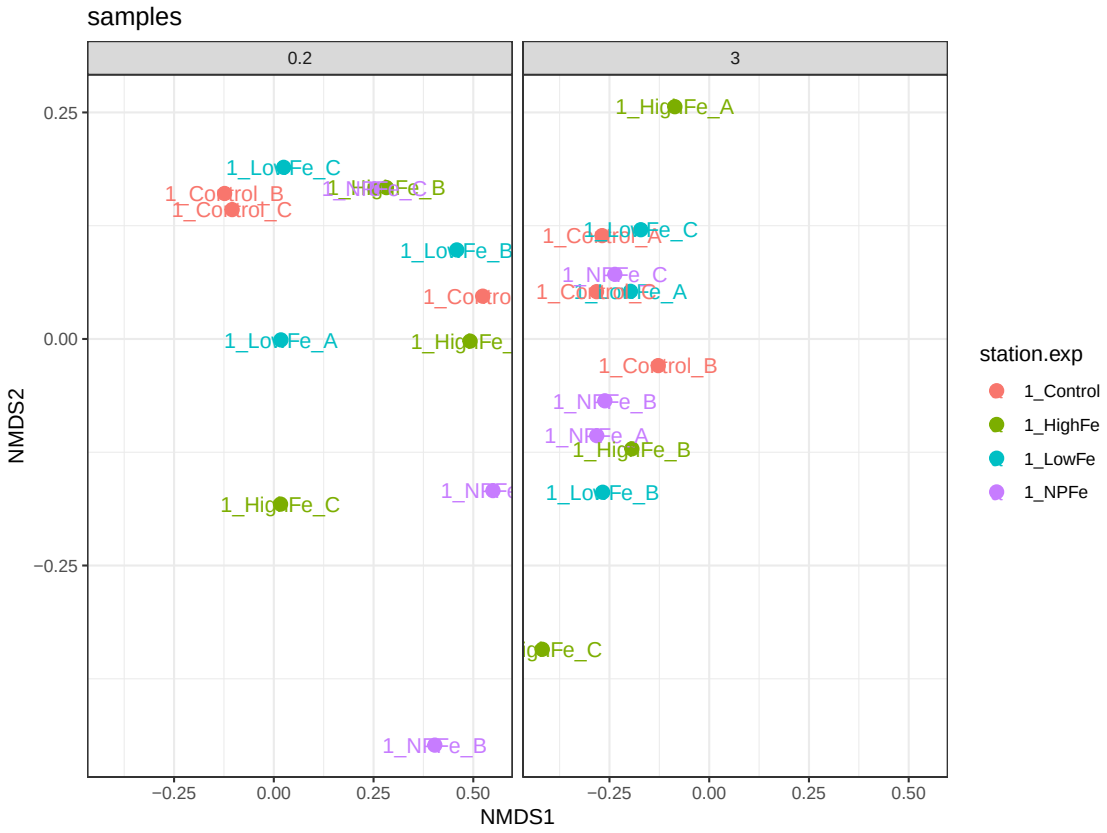

Figure 4.25: NMDS plot of the 18S amplicon sequencing data from the northern transition zone experiment for two size fractions: 0.2 - 3 μm (left) and $>$ 3 μm (right).

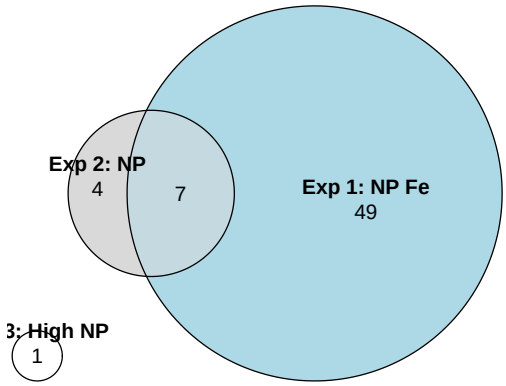

Figure 4.26: Venn Diagram of the metabolites that change between treatments and the control in each experiment (t-test comparisons of the treatment compared to the control for each experiment, fdr adjusted p-value < 0.1).

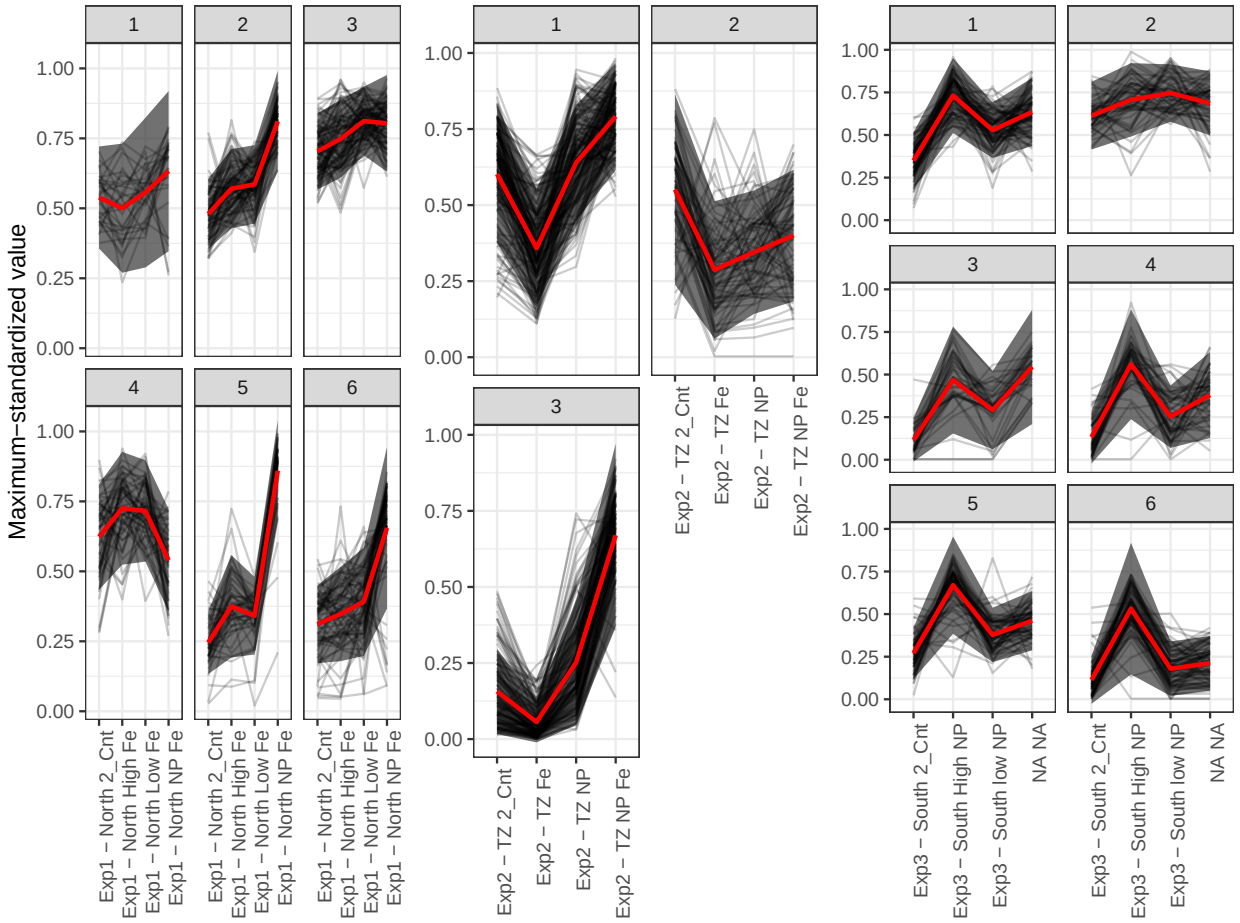

Figure 4.27: Average trends of each cluster of metabolites across the treatments for each experiment. Left: northern TZ experiment, Center: mid TZ experiment, Right: southern transition zone experiment. Details of the number of signals in each treatment are in Table 4.2. Metabolite abundance was maximum transformed relative to its maximum value in all replicates for a given experiment. Grey lines are the average trend of each siganl across the treatments, red lines are the mean profile from all signals in that cluster, and the shading is one standard deviation around the mean.

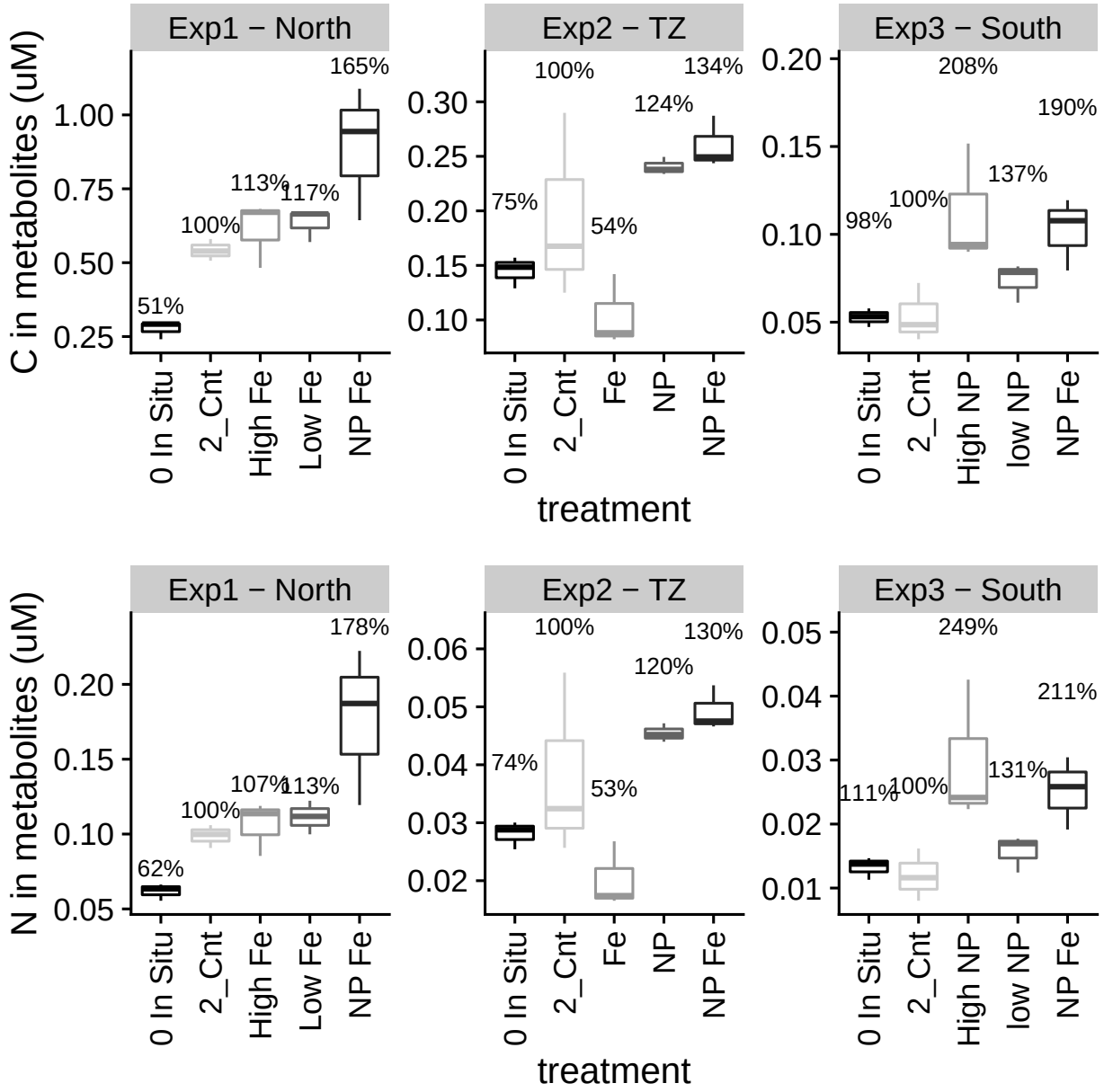

Figure 4.28: Total carbon (top) and nitrogen (bottom) in quantified metabolites in the different treatments of the incubation experiments and in the *in situ* data from the relevant station, in μM. Boxplots show median as the line and standard deviations around the mean as the bars. The numbers above the boxplots are the average percent change compared to the control treatment for each experiment.

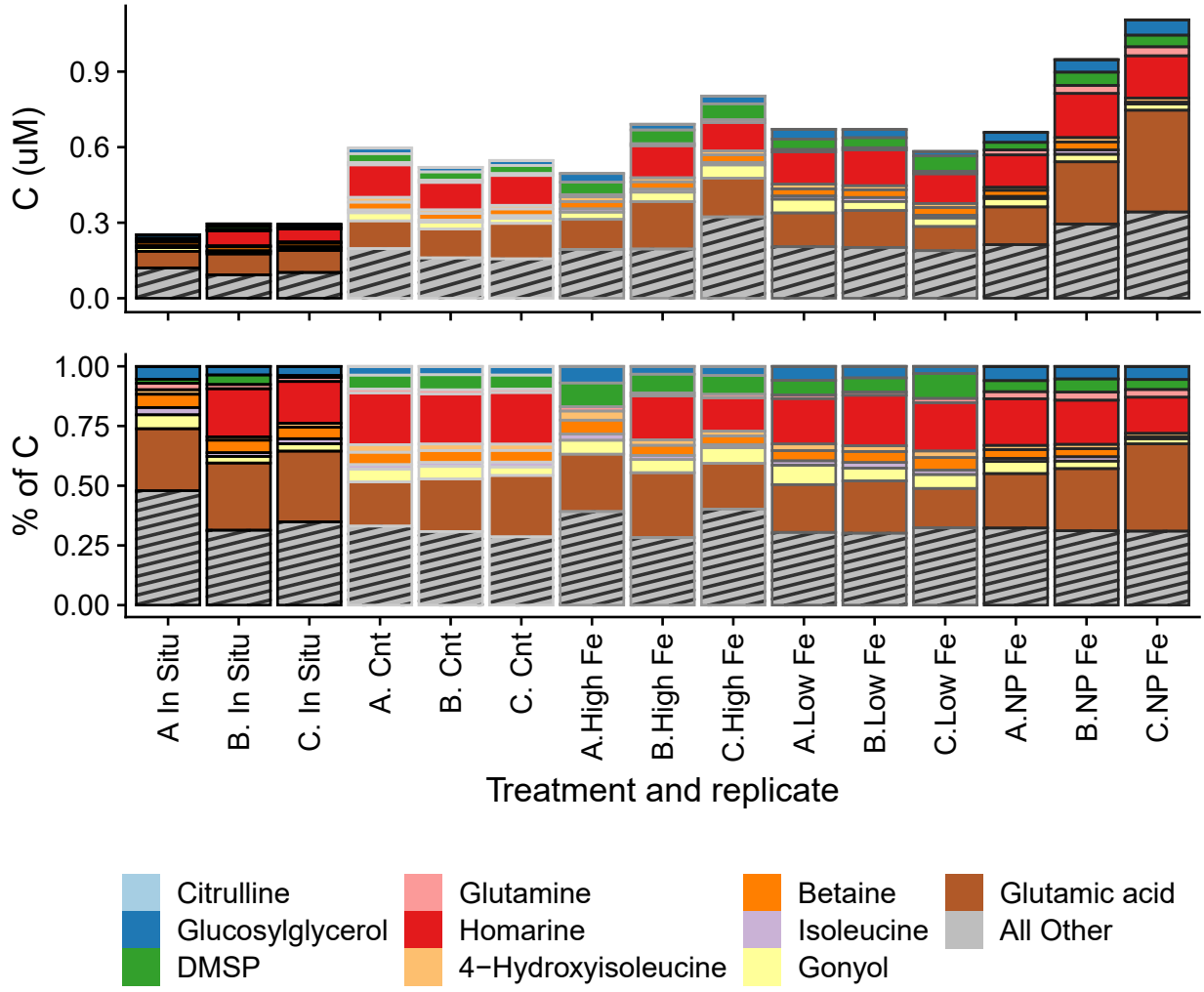

Figure 4.29: Quantified metabolites in the northern Transition zone experiment as μM carbon (top) and percent of carbon in quantified metabolites (bottom). The metabolites that had the most carbon in the transect are colored separately and all other metabolites are summed in the grey hashed bin.

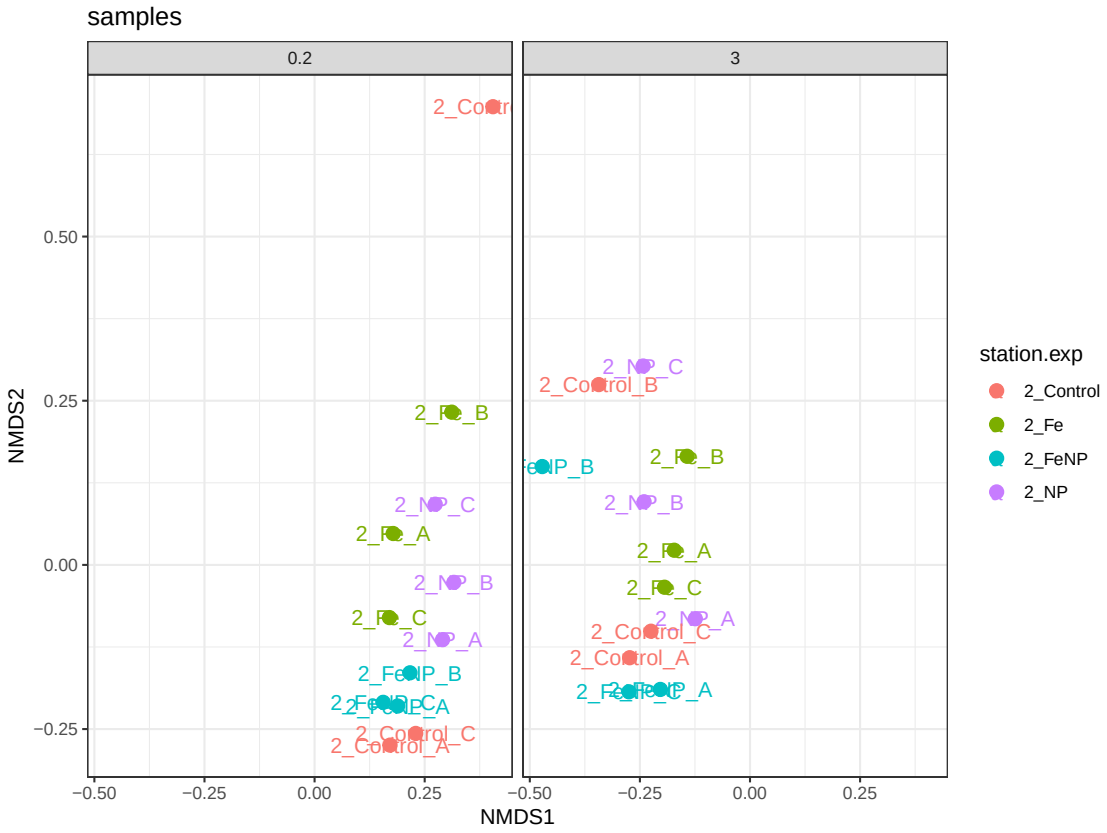

Figure 4.30: NMDS plot of the 18S amplicon sequencing data from the mid transition zone experiment for two size fractions: 0.2 - 3 μm (left) and $>$ 3 μm (right).

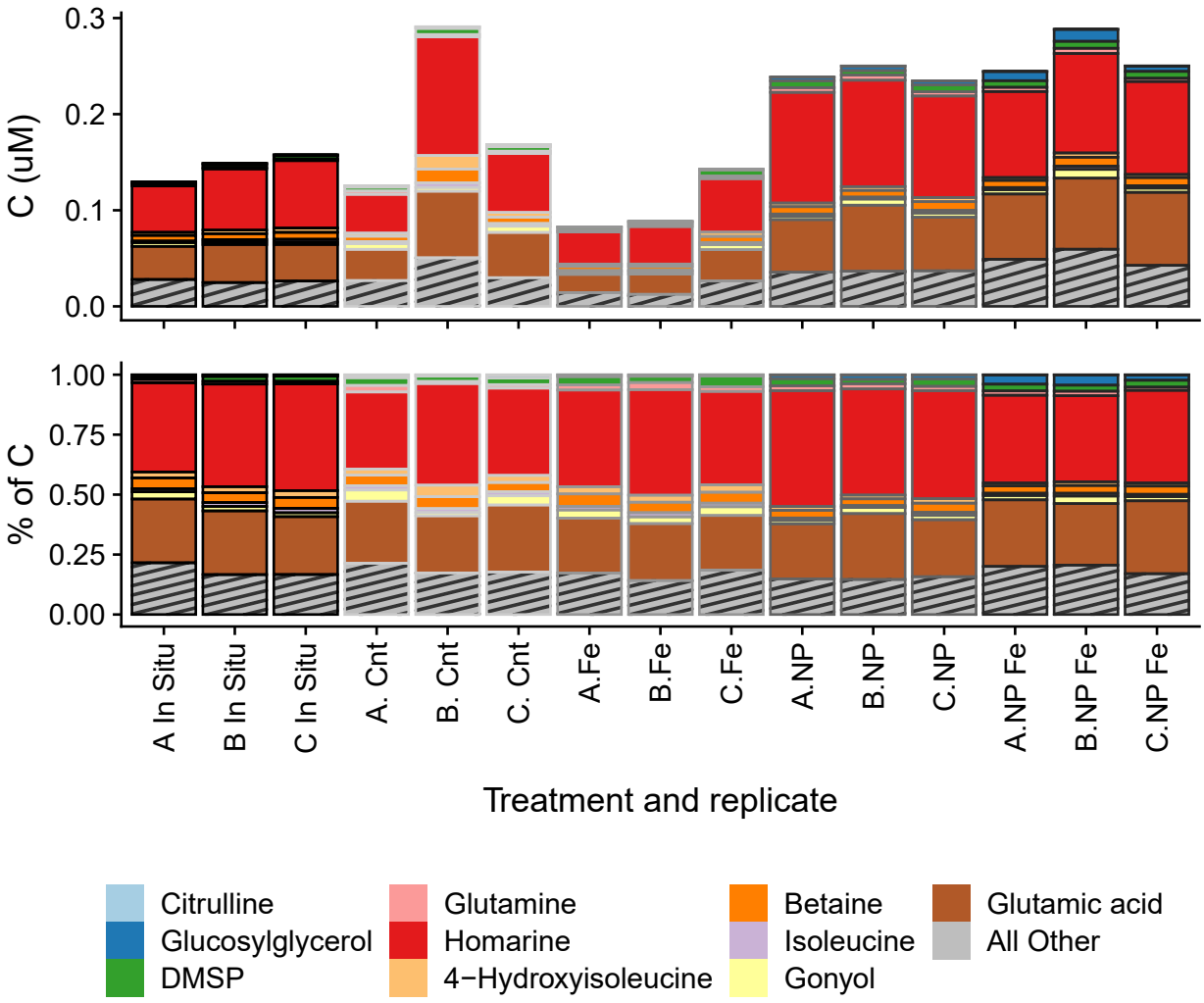

Figure 4.31: Quantified metabolites in the mid transition zone experiment as μM carbon (top) and percent of carbon in quantified metabolites (bottom). The metabolites that had the most carbon in the transect are colored separately and all other metabolites are summed in the grey hashed bin.

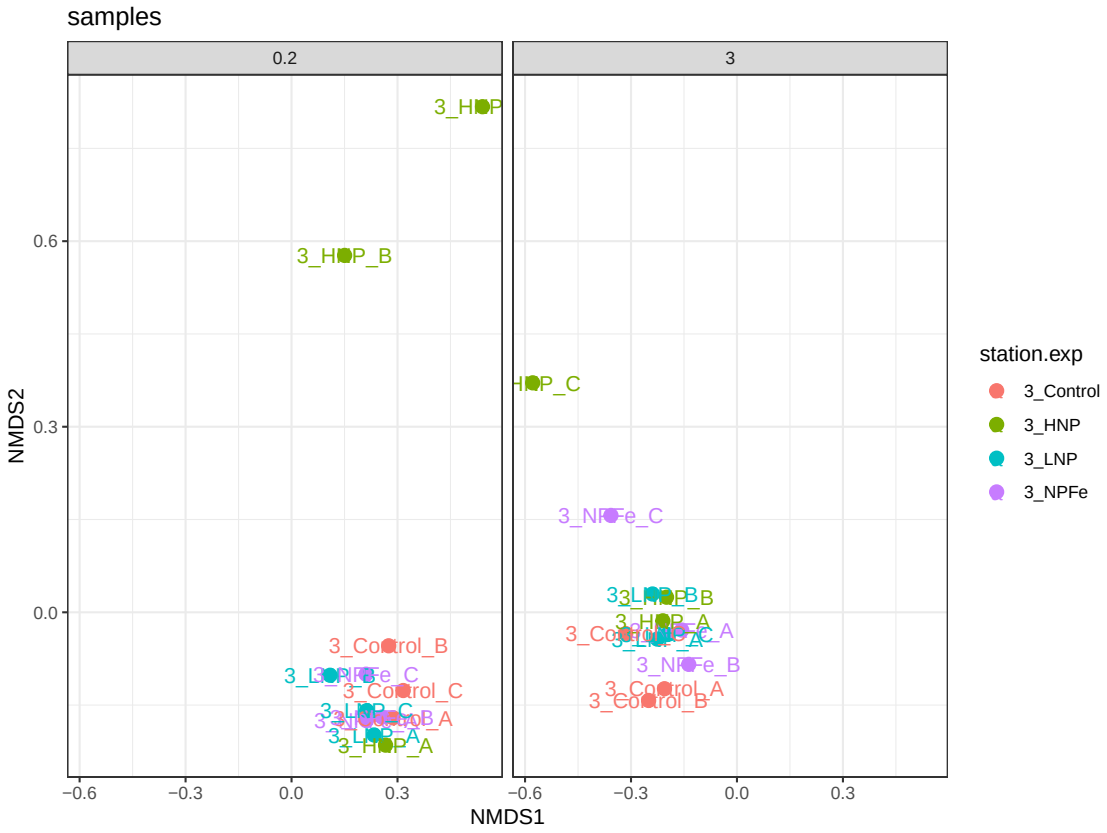

Figure 4.32: NMDS plot of the 18S amplicon sequencing data from the southern southern transition zone experiment for two size fractions: 0.2 - 3 μm (left) and $>$ 3 μm (right).

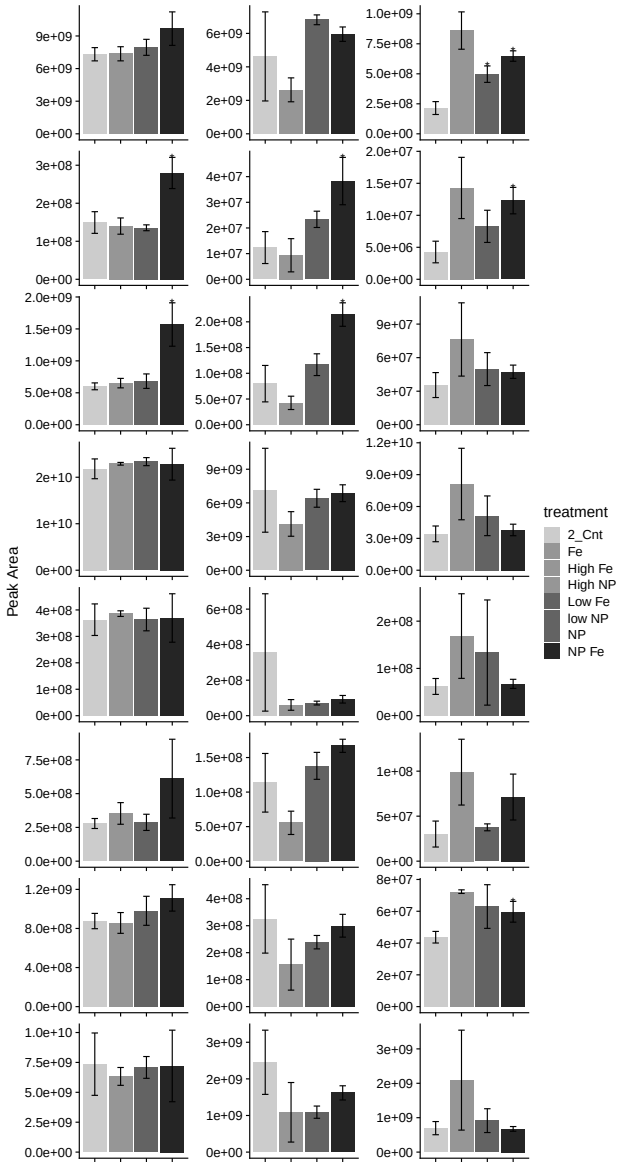

Figure 4.33: Major nitrogenous osmolytes identified in the northern TZ experiment 1 (left), mid TZ experiment 2 (center), southern transition zone experiment 3 (right). Homarine (top), taurine (second from top), proline (third from top), glycine betaine (fourth from top), proline betaine (fourth from bottom), glutamic acid (third from top), TMAO (second from bottom), TMAP (bottom).

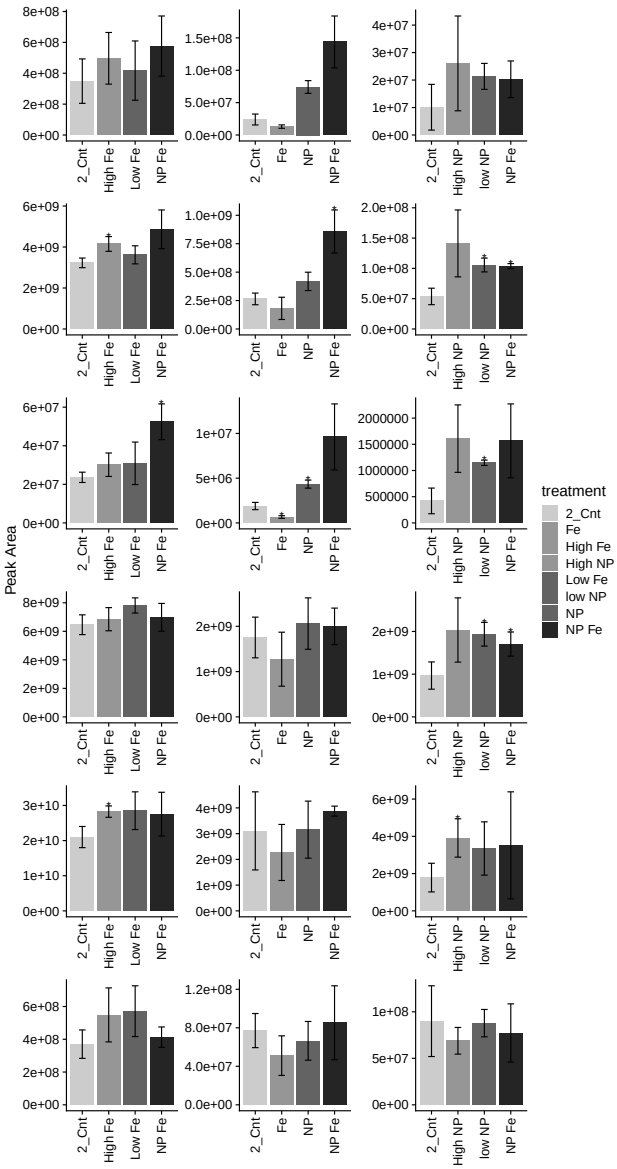

Figure 4.34: Osmolytes that do not contain nitrogen that were identified in the northern TZ experiment 1 (left), mid TZ experiment 2 (center), southern transition zone experiment 3 (right). DHPS (top) isethionic acid (second from top), glucosylglycerol (third from top), DMS-Ac (fourth from top), DMSP (fifth from top), gonyol (bottom).

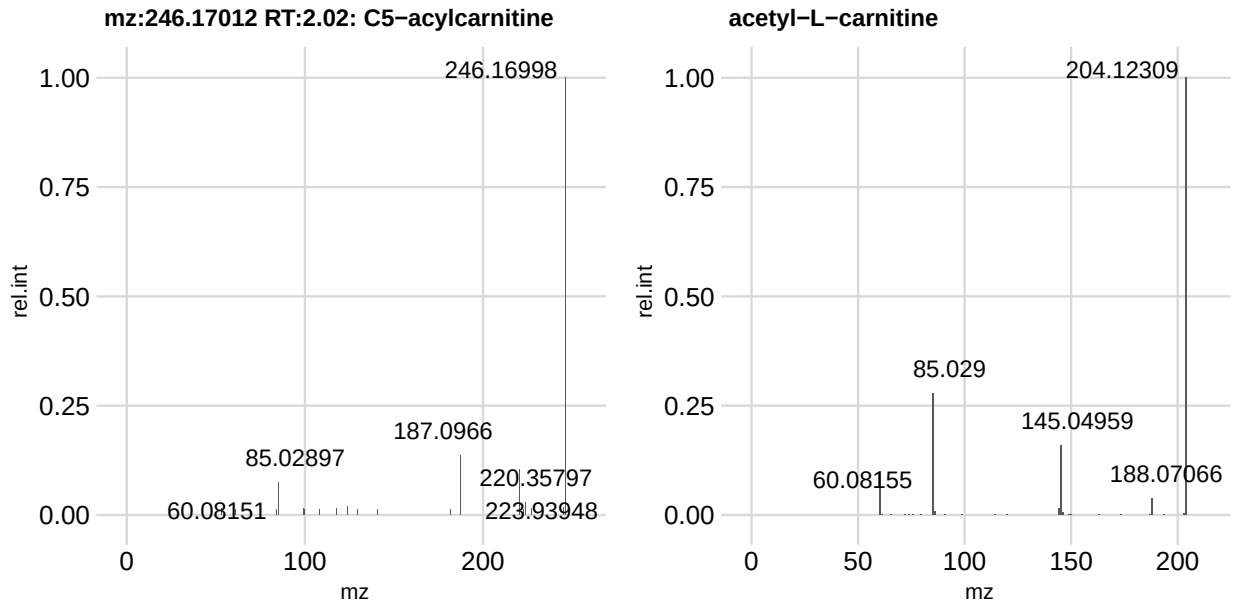

Figure 4.35: MSMS of the putative C5-acylcarnitine next to the MSMS of the definitively identified acetyl-l-carnitine.

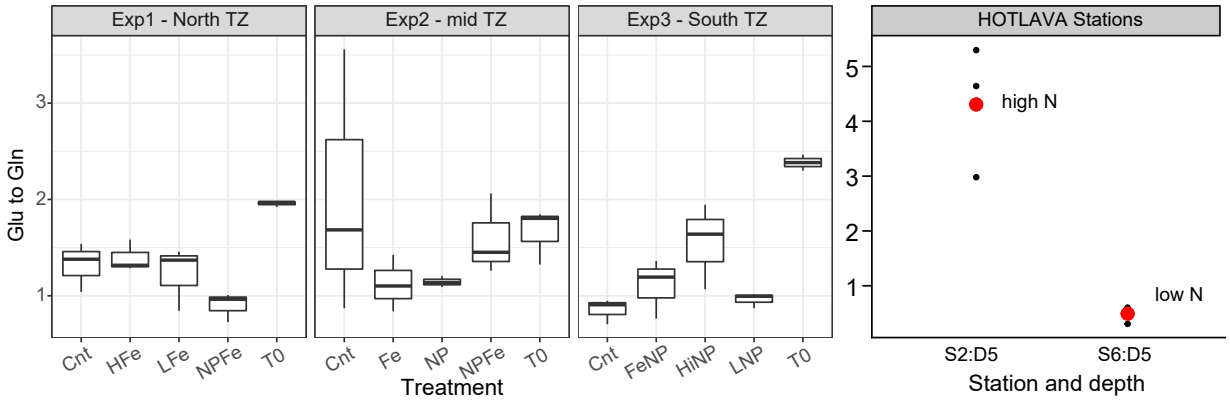

Figure 4.36: The ratio of glutamate to glutamine in the three nutrient amendment incubation experiments and in the HOTLAVA cruise stations 6 (non-lava modified, oligotrophic north Pacific seawater) and station 2 (geothermal modified water with elevated nitrate). Note that these ratios are based on the normalized peak areas not molar concentrations, so the direction of change rather than the absolute magnitude should be examined.

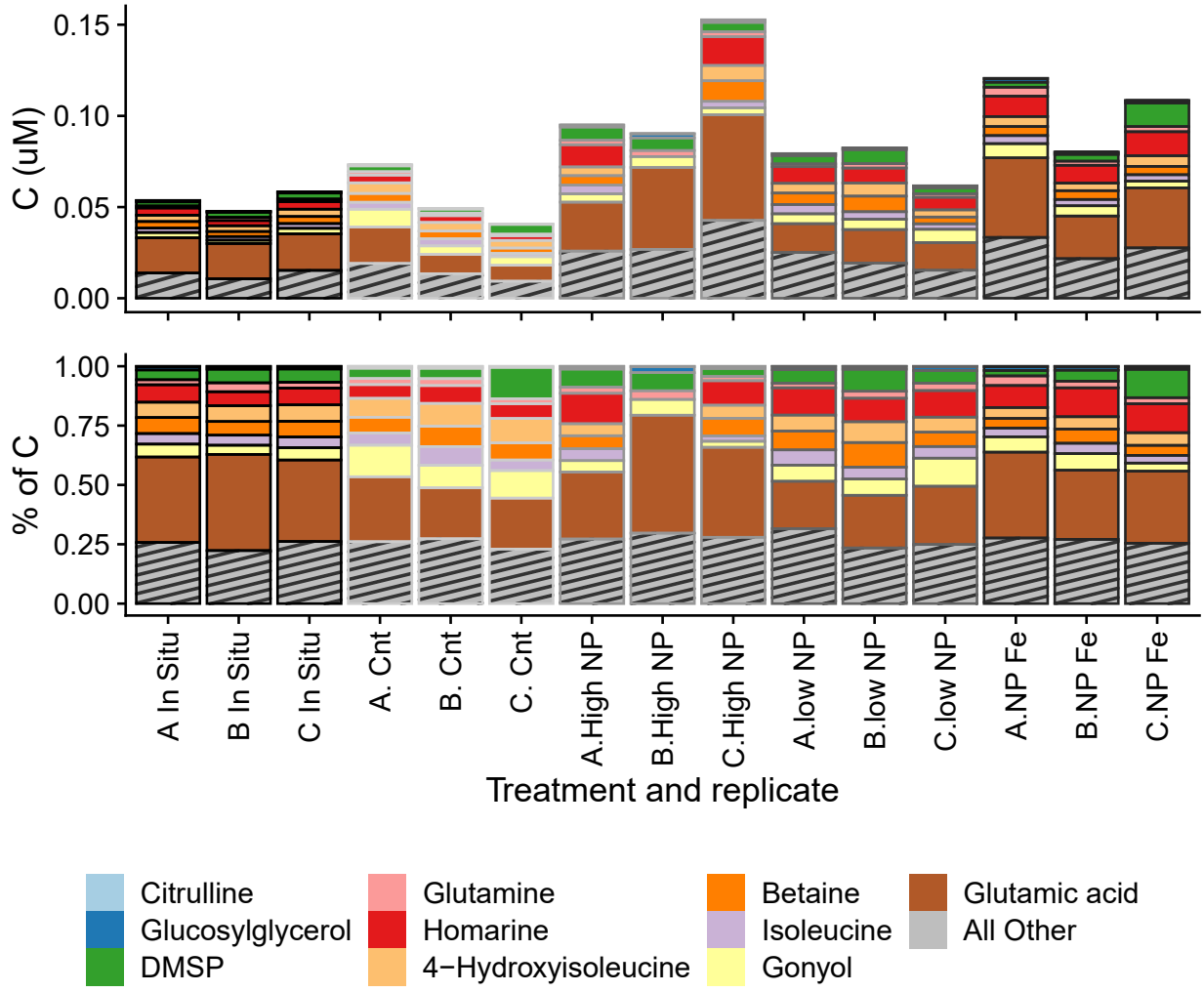

Figure 4.37: Quantified metabolites in the southern transition zone experiment as μM carbon (top) and percent of carbon in quantified metabolites (bottom). The metabolites that had the most carbon in the transect are colored separately and all other metabolites are summed in the grey hashed bin.

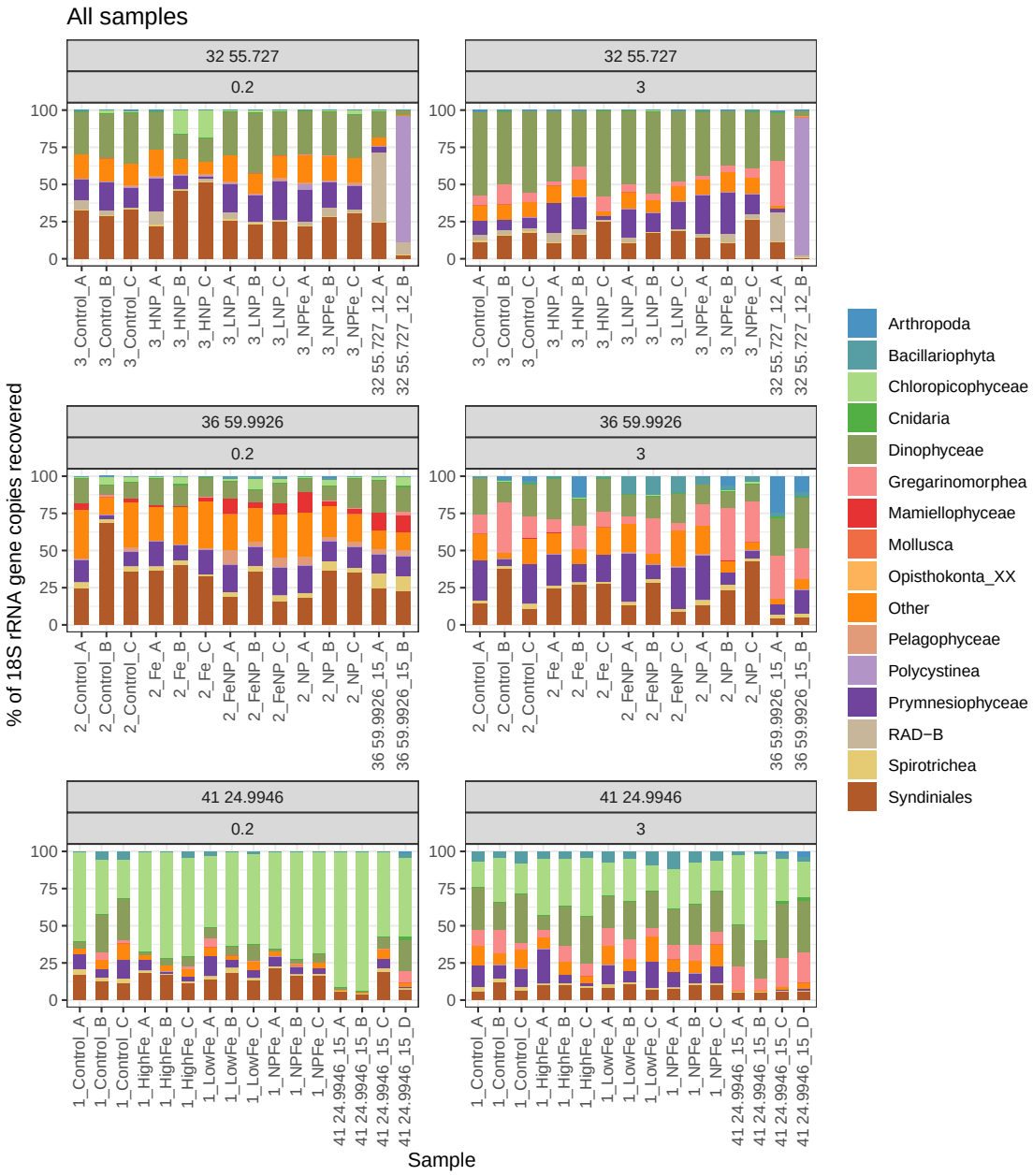

Figure 4.38: Relative abundance of 18S genes in the incubation experiments for small (left, < 3 μm), and large (right, > 3 μm) eukaryotes . Each replicate sample is shown, ordered by latitude. Orders contributing more than 15% of the total genes in at least one sample are shown, and the remaining orders are summed and shown as 'other'. Experiments are labeled by their latitude, with the corresponding *in situ* samples also shown.

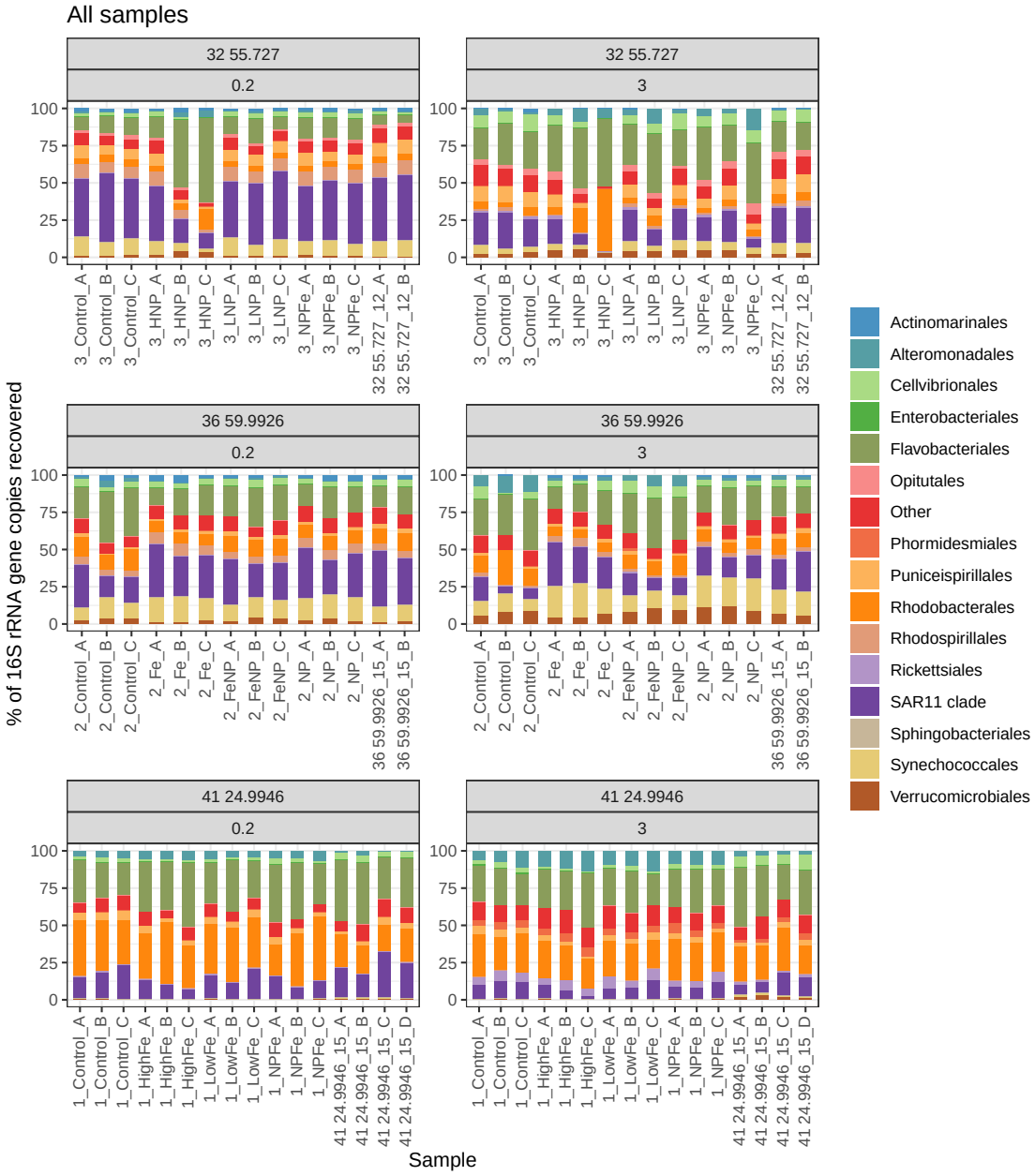

Figure 4.39: Relative abundance of 16S genes in the incubation experiments for small (left, < 3 μm), and large (right, > 3 μm) prokaryotes. Each replicate sample is shown, ordered by latitude. Orders contributing more than 8% of the total genes in at least one sample are shown, and the remaining orders are summed and shown as 'other'. Experiments are labeled by their latitude, with the corresponding *in situ* samples also shown.

Chapter 5

GLYCINE BETAINE UPTAKE AND USE IN MARINE MICROBIAL COMMUNITIES

5.1 Abstract

Glycine betaine (GBT) is a small polar metabolite present in high intracellular concentrations in natural marine microbial populations. When released from the cell by excretion or mortality it becomes part of the rapidly cycling dissolved organic nitrogen pool, and can be taken up and used for energy, C, or N. However, the dynamics of GBT uptake and use in marine microbial populations are not well understood. Using stable isotope labeled GBT and isotope tracing metabolomics, we determine the uptake kinetics and cellular fate of GBT in two natural microbial communities in the North Pacific transition zone. We calculate that the turnover time of dissolved GBT is 5.8 ± 1.3 days and 1.0 ± 0.6 days in the northern and southern stations, respectively, which are similar to estimates of nitrate turnover times. We estimate particulate GBT turnover time to be 11 and 13 hours at the two different sample locations. GBT was also used differently by the two microbial communities, likely due the different metabolic capabilities of the two communities and the availability of other C and N sources. In the southern station, where dissolved inorganic nitrogen concentrations were very low, GBT was primarily demethylated to sarcosine, and potentially all the way to glycine. Isotope enrichment in select free and combined hydrolyzable amino acids demonstrated that GBT was used as a cellular N source in the southern station. The accumulation of ^{13}C in metabolites such as dimethylsulfoniopropionate, beta-alanine betaine, and trigonelline indicated that the methyl groups from GBT are used to fuel the methhionine cycle and methylation throughout cellular metabolism. In the northern station GBT was primarily allocated towards trimethylamine-n-oxide, choline and glycerophosphocholine likely to serve

as an energy source or in lipid synthesis, respectively. In both locations, nearly all detectable ^{13}C and ^{15}N in particulate metabolites and hydrolyzable amino acids remained as untransformed GBT, indicating that many organisms have the ability to take up and retain GBT. The similarities and differences in the GBT uptake and use between the two stations illustrate that the microbial community and the inorganic nutrient supply both shape the cycling of this component of the dissolved organic nitrogen.

5.2 Introduction

Across vast swaths of the ocean, inorganic nitrogen is considered a limiting nutrient for phytoplankton growth [33]. In inorganic N depleted waters, dissolved organic nitrogen (DON) is the dominant form of nitrogen, sometimes comprising 100% of measurable nitrogen [221]. Microorganisms that can take up and use DON to fulfill their nitrogen or carbon requirements have an advantage in these environments. Like dissolved organic carbon (DOC), DON is a complex mixture of compounds with varying structures, molecular weights, and bioavailabilities [222]. Determining the pathways through which DON is cycled within the microbial ecosystem can enhance our understanding of global biogeochemical cycles and the structuring of microbial communities.

DON is released during cell death, and excreted by phytoplankton, bacteria, and zooplankton [223, 222]. A portion of DON is recalcitrant and bound in high molecular weight carboxyl-rich aliphatic matter or bound in amide groups on acylated polysaccharides that are relatively resistant to microbial degradation [224, 225, 226, 227]. However, total DON has a relatively short estimated turnover time of days to weeks [223, 228]. The combination of a slowly cycling component and an overall rapid turnover time for the whole pool suggests the existence of a highly labile pool that cycles very rapidly. Extensive work has shown that dissolved free amino acids have turnover times on the order of minutes, and urea and DNA have turnover times on the order of days [228]. While DON use has often been attributed to heterotrophic bacteria, there is strong evidence that phytoplankton take up DON, including urea and amino acids [229, 230, 231, 232, 233, 234].

Due to the chemical diversity and dilute nature of DON, molecularly characterizing DON is difficult [222, 224, 235]. DON components from phytoplankton derived cellular nitrogen include macromolecules like protein and nucleic acids as well as small polar metabolites. The most labile components of DON include dissolved free amino acids, which have a combined concentration between 0-0.5 μM in the ocean with nanomolar concentrations typical of the oligotrophic ocean and higher values in polar or estuarine regions [222]. Other components of the dissolved metabolite pool are also dilute, polar, and chemically diverse, allowing them to elude most measurement techniques [57]. Though these compounds may be at low dissolved concentrations in seawater, the flux through the dissolved phase may be high [39, 40]. Expression of transporter genes during incubations of dissolved organic matter (DOM) amended seawater has identified primary metabolites, amino acids, polyamines, and other nitrogenous compounds as likely fuels for bacterial growth [36]. This method is limited, however, by our ability to accurately annotate the appropriate transporter genes, and it does not provide quantitative information about the importance of various DON compounds nor their fate once taken up by cells.

Osmolytes are organic compounds that have high intracellular concentrations and can be released from cells during growth or mortality, and thus are likely quantitatively important DON sources [43]. Glycine betaine (GBT), a quaternary amine, is a common osmolyte that is one of the most effective compatible solutes in saline environments [114] and is one of the most abundant metabolites found in marine particles [236, 99, 30]. GBT comprises a significant portion of intracellular metabolites at Station ALOHA and in the North Pacific transition zone [41] and exhibits diel oscillations [30]. Cyanobacteria, heterotrophic bacteria, and eukaryotic algae all have a variety of abilities to synthesize, transport, and metabolize GBT [237, 43, 44, 42]. A simplified metabolic map for GBT (Figure 5.1) shows several potential cellular uses for this compound. Some microbes, including the ubiquitous marine *Alphaproteobacteria* clade SAR11, can produce energy by oxidizing the methyl groups from GBT to CO_2 via the demethylation pathway that first generates dimethylglycine and methionine and then sarcosine and glycine through subsequent demethylations in which methyl

groups are transferred to tetrahydrofolate and subsequently oxidized to CO_2 [45]. SAR11 is able to fulfil its glycine requirement with GBT [139] and other marine organisms likely share this ability to convert GBT to glycine and use it as a source of carbon and nitrogen for protein synthesis (black pathway in Figure 5.1) [138, 39]. GBT can also be reduced to trimethylamine (TMA) which can be oxidized to trimethylamine-n-oxide (TMAO, another compatible solute) or fuel anaerobic methanogenesis [238]. TMA and TMAO can both be demethylated to ultimately form formaldehyde and ammonia, which can feed back into carbon and nitrogen metabolism [238]. It is also possible that microorganisms may take up GBT and retain it as an osmolyte despite being unable to synthesize or degrade it endogenously. This ability would provide a growth advantage by saving the elemental and energetic resources that would otherwise be spent synthesizing osmolytes *de novo*.

Dissolved glycine betaine has not been measured in seawater due to the analytical challenges mentioned above, but additions of trace amounts of ^{14}C-GBT have shown rapid uptake in coastal seawater, with uptake dominated by organisms < 0.8 μm in diameter [138, 46]. In the first several hours of incubation, $< 80\%$ of the ^{14}C in particles remained as untransformed GBT, indicating that GBT is not rapidly metabolized. After 24 hours, appreciable amounts of the total radioactivity was in CO_2 and in macromolecules [46]. It remains unknown what pathways of GBT metabolism were relevant in these experiments and, because only the carbon in methyl groups was traceable, the fate of the rest of the molecule, including the nitrogen, remains uncertain.

Here we studied how glycine betaine is used by two natural microbial communities in the North Pacific transition zone (NPTZ), using stable isotope tracing metabolomics. The NPTZ is a dynamic region characterized by seasonally migrating fronts of temperature, salinity, nutrients, and chlorophyll [181, 34]. We incubated natural seawater collected at two different stations in the NPTZ with stable isotopically labeled $^{13}C_5$, $^{15}N_1$-GBT to measure the kinetics of GBT uptake and determine the fate of the C and N from GBT. We show how two microbial communities are differently poised to take up GBT. Once GBT is in cells, we show that these two communities allocate the carbon and nitrogen from GBT differently,

likely due to the different metabolic capabilities of the organisms present and to differences in the other available sources of carbon and nitrogen.

5.3 Methods

5.3.1 Sampling Strategy

Samples were collected aboard the R/V Kilo Moana on the Gradients 3 cruise from April 10–April 28, 2019. Experiments were conducted at stations 4 and 5, which were at 41°40.85' N and 158°3.01' W and 37°0.21' N and 158°0.20' W, respectively. Water for metabolomics samples of the *in situ* microbial community and for setting up the incubation experiments was collected from 15 m water depth, which was within the surface mixed layer with Niskin bottles attached to the CTD. DON, DIN, particulate carbon (PC), and particulate nitrogen (PN) samples were collected from the CTD rosette at 15 m and analyzed according to the standard Hawaiian Ocean Time Series program protocols. Temperature and chlorophyll concentrations are from the CTD.

5.3.2 GBT Uptake Kinetics Experiments

In order to determine the kinetics of GBT uptake, whole seawater was spiked with varying concentrations of $^{13}C_5$, $^{15}N_1$-GBT and incubated for 25–42 minutes. The northern experiment occurred on 4/16/2019 at Station 4 (North, Experiment 1). The southern experiment occurred on 4/20/2019 at Station 5 (South, Experiment 2). Water was collected into 2 L bottles around 8:00 am local time for both experiments. In order to minimize the biological transformation of $^{13}C_5$, ^{15}N-GBT into other molecules and to reduce the likelihood of induction of enzymatic activity, the incubation time with $^{13}C_5$, ^{15}N-GBT was kept short. The short incubation time necessitated that samples were spiked throughout the course of the day since only six samples could be processed at a time. Before and after the addition of GBT, bottles were kept in flow-through incubators with blue shading to be at *in situ* temperature and approximately mixed layer light conditions. Samples were spiked to

have final concentrations of 0, 2, 5, 10, 50, 200, or 2000 nM $^{13}C_5$, $^{15}N_1$-GBT. Triplicates of each $^{13}C_5$, $^{15}N_1$-GBT concentration were processed for both experiments. After the short incubation with the spiked molecule, seawater was filtered onto 47 mm diameter, 0.2 μm pore size PTFE (Omnipore) filters using a peristaltic pump, polycarbonate filter holder, and Masterflex PharMed BPT tubing (Cole-Parmer). Filtering time was 10–42 minutes with an average time of 22 minutes. Experimental blanks were collected for each spike concentration during the northern experiment by collecting filtrate from one replicate of each treatment and re-filtering filtrate onto a new filter. This provided a measure of dissolved organic compounds adsorbed onto the filter during processing. Filters were frozen in liquid nitrogen immediately after filtration and stored at -80 °C until GBT extraction. GBT was extracted from the filters using the metabolite extraction method described in Boysen et al. (2018) [22].

GBT and $^{13}C_5$, ^{15}N-GBT were measured using paired liquid chromatography mass spectrometry using a Waters Acquity I-Class UPLC (Waters Corporation, Milford, MA) with a hydrophilic interaction liquid chromatography column (HILIC SeQuant ZIC-pHILIC column, 5 mm particle size, 2.1 mm x 150 mm, from Millipore) paired to a Waters Xevo TQ-S triple quadrupole. Electrospray ionization (ESI) in selected reaction monitoring mode (SRM) with polarity switching was used to collect fragments for GBT and $^{13}C_5$, $^{15}N_1$-GBT. In order to quantify GBT and $^{13}C_5$, $^{15}N_1$-GBT standard addition curves were done in two representative samples: a 2 nM treatment from the first kinetics experiment and a 2000 nM treatment from the second kinetics experiment. Each of these standard curves was analyzed twice, once at the beginning of the data acquisition and again at the end of the data acquisition. The concentration of particulate $^{13}C_5$, ^{15}N-GBT in each sample was quantified using the calibration curve that best represented the sample. If the sample had a spike of less than 50 nM the calibration curve from the 2 nM treatment was used, otherwise the calibration curve in the 2000 nM treatment was used. If the sample was analyzed in the first half of the sample analysis run the standard curves analyzed at the beginning of data acquisition were used, otherwise the curves analyzed at the end of data acquisition were used in order to

best represent the instrument conditions. Standard errors on the concentration values were calculated by propagating the uncertainty from the coefficients fitted to the standard curve. Samples were always well above the concentration in the blank, with an average of 63 times greater concentration in the sample than the blank across all treatments. Thus, blanks were disregarded for the remainder of the uptake kinetics calculations.

Uptake rates of $^{13}C_5$, ^{15}N-GBT were calculated using an 'average incubation time' for each sample that took into account the filtering time. This time ranged between 36–52 minutes. Kinetics for the community uptake of $^{13}C_5$, $^{15}N_1$-GBT were then calculated by fitting the data to a Michaelis–Menten equation:

$$v = (V_{max} * (A))/(K_t + S + A) \tag{5.1}$$

Where v is the uptake rate, A is the added $^{13}C_5$, ^{15}N-GBT concentration, S is the in situ dissolved GBT concentration, K_t is the half saturation constant, and V_{max} is the maximum uptake rate. An initial fit to get values of V_{max} and $K_t + S$ was done using the calculated values of uptake rates with no associated uncertainties using the nls function in the R package stats. Using the parameters estimated from that initial fit as a starting place, a monte carlo analysis was performed sampling the uncertainty around each estimate of uptake rate to find the appropriate fit and error for each parameter. Each model was fit with 1000 iterations. Outlier models were defined as models that took over 50 iterations to converge on a solution or models with poor overall fit (R value or sigma value > 2 standard deviations outside the mean of all 1000 models). After removing outliers, the mean and standard deviation of V_{max} and $K_t + S$ were calculated.

A Wright-Hobbie linear transformation was also used to estimate the uptake kinetics and turnover time by regressing t/f (incubation time in hours divided by ratio of added GBT that was taken up) versus A (added concentration of GBT, in nM, see [46]). With this transformation, the $K_t + S$ is the negative x-intercept, V_{max} is the inverse of the slope, and the turnover time is the y-intercept. A monte carlo analysis was performed to sample the

uncertainty around each data point for the proportion of GBT that was taken up. Outlier models were defined as models which were outliers for total correlation value and for adjusted R squared. After removing outliers, the mean and standard deviation of the turnover time, V_{max}, and $K_t + S$ were calculated.

5.3.3 GBT Fate Experiments

In order to determine the ways that GBT is used by natural microbial communities two time course incubation experiments were performed. The first, northern experiment occurred on 4/17/2019 at Station 4. The second, southern experiment occurred on 4/20/2019 at Station 5. For each experiment, 2 L bottles were filled with seawater collected from the CTD at 15 m depth (within the mixed layer), at approximately 6:00 am local time. All 2 L bottles were spiked with 500 nM $^{13}C_5$, $^{15}N_1$-GBT and incubated in temperature controlled incubators at 10 °C and 14 °C for the north and south experiments, respectively. Samples for the initial timepoint (T0) were filtered as soon as possible after being spiked, resulting in actual incubation times of approximately 20 minutes. At each time point triplicate 2 L bottles were sampled for analysis of flow cytometry, metabolites, total hydrolyzable amino acids, and nutrient concentrations. Timepoints sampled in the north experiment were 0, 4.5, 9, 12.5, 36, 50, and 98 hours. Timepoints sampled in the south experiment were 0, 6, 13, 25, 51, and 100 hours. At select time points additional samples were taken for total particulate carbon, total particulate nitrogen, and 16S/18S amplicon sequencing. For all samples, incubation time was calculated by taking the difference between the time the bottle was spiked and the midpoint time of sample filtration.

Flow cytometry samples were taken by aliquoting 2 mL from a bottle into a cryovial, spiking with glycerol, incubating in the dark for 20 min, and flash freezing in liquid nitrogen. Samples were stored at -80 °C until analysis on an influx flow cytometer. Metabolomics samples were collected as described in the kinetics experiment, by filtering approximately 2 L. Samples for nutrient analysis were collected by filtering 40 mL of water through syringe filters with 0.2 μm filters into acid cleaned 40 mL falcon tubes. Samples were stored at

-20 °C until analysis. Total hydrolyzable amino acid samples were collected by filtering approximately 200 mL of water onto glass fiber filters (GF75 grade, 47 mm) using vacuum filtration. Filters were flash frozen in liquid nitrogen and stored at -80 °C until analysis. Experimental blanks for metabolomics samples and total hydrolyzable amino acids were collected during the northern experiment at T36 by re-filtering filtrate from those samples.

5.3.4 Metabolomics and lipidomics data acquisition

Metabolite extractions were performed as in Boysen et al. [22] using a modified Bligh-Dyer extraction that yielded an aqueous and organic fraction. Aqueous soluble metabolites were analyzed with two injections for each sample, one with a spike of heavy isotope labeled internal standards and one with an equal volume spike of UPLC grade water. Samples were run using both HILIC and reversed phase chromatography on a Waters Acquity I-Class UPLC (Waters Corporation, Milford, MA) with injection volumes of 2 and 15 μL respectively, coupled to a QExactive Orbitrap MS (Thermo Scientific). Unspiked samples separated with HILIC were run in both positive and negative mode with ESI and a full scan method at 240,000 resolution with a scan range of 60 to 900 m/z. Spiked samples separated with HILIC were run in positive negative switching mode with ESI and a full scan method at 60,000 resolution with a scan range of 60 to 900 m/z. Unspiked and spiked samples separated with RP were run in positive mode with ESI and a full scan method at 240,000 resolution with a scan range of 90 to 900 m/z. Details of the chromatography and mass spectrometry methods are in Boysen et al. [22]. Pooled samples were analyzed in data dependent acquisition mode using both HILIC and reversed phase chromatography with separate injections for positive and negative ion modes (60,000 resolution and a dynamic exclusion time of 10 seconds for MS/MS data, the top ten most abundant ions in each cycle were fragmented, collision energies of 10, 20, and 35 eV).

The organic soluble lipid phase was spiked with heavy isotope labeled internal standards and analyzed in both positive and negative mode using a Waters Acquity UPLC CSH C18 column (1.7 μm particle size, 2.1 mm x 150 mm) with an injection volume of 20 μL, ESI, a

full scan method with a scan range of 150 to 2000 m/z and 240,000 resolution. The solvents were 10mM ammonium formate in 60:40 acetonitrile to water and 0.1% formic acid (Solvent A) and 10mM ammonium formate in 90:10 isopropyl alcohol to acetonitrile and 0.1% formic acid (Solvent B) at a flow rate of 0.45 mL/min. Initial conditions were 90% A and 10% B. The column was ramped to 80% B over 33 minutes, ramped to 90% B over 12 minutes, held at 90% B for 1 minute, and equilibrated to 90% A and 10% B for 6 minutes (total run time of 52 minutes). The column was maintained at 65°C. For lipid analysis, the capillary temperature was 320°C, the H-ESI spray voltage was 3.8 kV, and the auxiliary gas heater temperature was 150°C. The S-lens RF level was 65. Sheath gas, auxiliary gas, and sweep gas flow rates were maintained at 40, 10, and 1, respectively. For the pooled samples, both high-resolution MS scans and data dependent MS/MS scans were collected with separate injections for positive and negative ion modes (60,000 resolution and a dynamic exclusion time of 10 seconds for MS/MS data, the top ten most abundant ions in each cycle were fragmented, collision energies of 10, 20, and 35 eV).

For all analyses, pooled samples spiked with isotope labeled internal standards were run at full and half strength in full scan mode as well as data dependent acquisition mode. Standards spiked into water, standards spiked into the pooled sample, and water spiked into the pooled sample, resulting in the same dilution as the standards, were run to confirm and quantify known metabolites. Files were converted from .raw to .mzXML format using ProteoWizard [73].

5.3.5 Targeted metabolite data processing

Skyline [239, 240] was used to integrate the peak areas of the metabolites for which we have authentic standards (Supplemental Table 5.5), along with the potential isotopologues of those metabolites. Metabolites or their isotopologues were removed from further analysis if they had comparable peak area in the blanks (sample peak area $< 2x$ blank peak area) or if they were more than 10 ppm mass deviation from expected. Mass isotopologue distributions (MID) are the percent contribution each individual isotopologue had to the total concentra-

tion of all isotopologues for a given metabolite. MIDs were calculated for each metabolite using only the samples that did not receive spikes of isotope labeled internal standards to ensure that all isotopologues were from GBT transformation or ^{15}N and ^{13}C natural abundance and not due to our added internal standards. The significance of MID changes over the incubations were assessed using a linear regression of MID over time. Samples from the same stations as these experiments but which were not spiked with isotope labeled GBT were analyzed the same way to determine the natural isotope abundance. Using the samples that were not spiked with isotope labeled internal standards, the unlabeled metabolites were normalized using best matched internal standard normalization (BMIS)[22]. If the coefficient of variation for a given metabolite was less than 0.1 with no normalization, or if the internal standards did not reduce the coefficient of variation by more than 0.1, then internal standard normalization was not used for that metabolite. Glycine betaine did not match with an isotope labeled internal standard for normalization, thus the quantification of GBT did not leverage the internal standards and instead relied on peak area alone.

To calculate the absolute concentration of glycine betaine in the samples, a sample from the cruise that did not come into contact with labeled glycine betaine was used to calculate the response factor in the relevant sample matrix that correlates the peak area of GBT to its known concentration. $^{13}C_5$, $^{15}N_1$-GBT was spiked into this sample at a final concentration of 500 nM just before analysis. This sample was analyzed 5 times over the course of the sample run and thus provided a robust estimate of the response factor and its variability (coefficient of variation = 0.04). This value was applied to the experiment samples and was used to determine the nM concentration of all isotopologues of glycine betaine in each vial, which was then converted into the concentration of glycine betaine in each sample after accounting for concentration during extraction.

5.3.6 Total hydrolyzable amino acid analysis and data processing

Amino acids were hydrolyzed as in Fountoulakis and Lahm, 1998 with some modifications [241]. Samples were heated at 120 °C for 20 hours instead of 110 °C for 20–24 hours. Initial

recovery tests with bovine serum albumin (BSA) resulted in better recovery of the amino acids at 120 °C as compared to 110 °C. A shorter hydrolysis with BSA at 150 °C was also tried but recovery of amino acids was highest at 120 °C.

Glass fiber filters were transferred into combusted 4 mL glass vials. Enough 6N hydrochloric acid was added to cover the filter. Each sample was purged under nitrogen gas for 30 seconds before being immediately sealed with a solvent rinsed cap. The samples were then heated at 120 °C for 20 hours. The acid was then transferred to a clean, combusted glass vial. The hydrolysis vial and filter were rinsed with approximately 500 μL of optima grade H_2O and the water was combined with the acid. A rinsing step was repeated with an equal volume of optima grade methanol. The acid mixture was concentrated to dryness under nitrogen gas and medium heat. Once dried, approximately 500 μL of H_2O was used to rinse each vial and samples were returned to the nitrogen gas to dry completely. Dried samples were re-dissolved in 400 μL of optima grade H_2O and syringe filtered into LCMS vials.

Amino acids were derivatized as in Gray et al. (2017) using the AccQ Tag Ultra derivatization kit from Waters [242]. A 10 μL aliquot of each sample was transferred into a new glass LCMS vial and 70 μL of borate buffer was added to each vial. After vortexing, 20 μL of AccQ Tag Ultra derivatizing agent was added to each sample and the vials were heated at 55 °C for 10 minutes. Blank filters collected by filtering 0.2 μM filtrate through the same sampling apparatus were collected during the experiment and were analyzed alongside the samples.

For analysis of derivatized amino acids, a Waters Acquity UPLC HSS T3 column (1.8 μm particle size, 2.1 mm x 150 mm) was used with 0.1% formic acid in water (Solvent A) and 0.1% formic acid in acetonitrile (Solvent B) at a flow rate of 0.6 mL/min. The column was held at 4% B for 0.5 minutes, ramped to 10% B over 2 minutes, ramped 28% B over 2.5 minutes, ramped to 95% B over 0.5 minutes, held at 95% B for 1 minute, and equilibrated back to 4% B for 2.5 minutes (total run time is 9 minutes). The column temperature was maintained at 45 °C.

A full scan method in positive ion mode was used with a scan range of 100 to 600 m/z and a resolution of 60,000 on a Thermo Scientific Q Exactive HF Orbitrap. The capillary temperature was 320 °C, the H-ESI spray voltage was 3.8 kV, and the auxiliary gas heater temperature was 90 °C. The S-lens RF level was 65. Sheath gas, auxiliary gas, and sweep gas flow rates were maintained at 50, 15, and 1, respectively.

Peak areas of each derivatized amino acid were integrated using Skyline, and the average peak areas of amino acids in the blank samples were subtracted from the sample peak areas. Amino acids that had peak areas in the samples that were smaller than the blank average were discarded and replaced with 0. In order to quantify amino acids, a standard addition curve was made using a mix of unlabeled amino acid standards. The standard addition curve was run in one representative sample by spiking in amino acid standards for a final concentration of 0.5, 1, and 2.5 μM after hydrolysis and before derivatization. The standard curve samples were then derivatized and analyzed alongside the samples. Skyline was used to integrate isotopologues of the amino acids corresponding to isotopologues with 0 or 1 ^{15}N and 0–5 ^{13}C. After quality control, MIDs were calculated for the total hydrolyzable amino acids. A t-test was used to determine if there was a significant difference in MID between the initial and final timepoints.

5.4 Results and Discussion

5.4.1 Environmental description

The two stations sampled in this study had different physical and chemical characteristics (Table 5.1). The mixed layer particulate carbon (PC) was higher at the southern station than the northern station, which had higher chlorophyll concentration. The particulate nitrogen (PN) concentration was similar at both stations. The total organic nitrogen (PN + DON) was also similar at both stations but dissolved inorganic nitrogen (DIN: NO_3^- + NO_2^-) was much higher in the north than in the south, where it was below our limit of detection (near 0.02 μM).

5.4.2 Uptake kinetics and turnover time estimates

Glycine betaine uptake kinetics were different at the northern and southern locations (Table 5.2, Figure 5.2). When calculating uptake kinetics parameters, $K_t + S$ was fit as one parameter rather than two because the transporters taking up natural dissolved GBT and the added labeled GBT are the same. Thus the observed uptake rates are a function of the added label GBT in addition to the *in situ* concentration of GBT, but we do not know the *in situ*concentration of GBT. Fitting this $K_t + S$ parameter as two independent parameters rather than one did not significantly change the estimates of K_t, and resulted in an estimation of S that was not significant ($p > 0.05$, Supplemental Table 5.4). The nonlinear least squarest method and the Wright-Hobbie linear transformation method for estimating ($K_t + S$) and V_{max} values resulted in similar estimates with overlapping uncertainties (Table 5.2). In the subsequent discussion, parameters from the nonlinear least squares will be used.

In the north, the half saturation constant plus in situ dissolved GBT ($K_t + S$) and V_{max} were 79.46 ± 48.26 nM and 0.36 ± 0.06 nM/h respectively. In the south, near 37 N, the half saturation constant plus in situ dissolved GBT ($K_t + S$) and V_{max} were 11.06 ± 5.53 nM and 0.56 ± 0.06 nM/h respectively. The dissolved GBT (S) turnover time (or residence time) at the northern and southern stations were 137.8 ± 31.5 hours (5.75 ± 1.3 days) and 23.1 ± 14.4 hours (0.96 ± 0.6 days), respectively.

Few isolated microorganisms have had GBT uptake kinetics characterized. The Antarctic sea-ice diatoms *Nitzschia lecointei*, *Navicula cf. perminuta*, and *Fragilariopsis cylindrus* have half saturation values, K_t, of 189 to 315 nM [42]. Two strains of the SAR11 clade of pelagic heterotrophic bacteria have half saturation values of 0.89 and 1.85 nM [139]. The K_t values from these different organisms reflect the environment that they are adapted for, with Antarctic sea-ice diatoms adapted to DOM rich brine pores and SAR11 adapted to the oligotrophic open ocean with low concentrations of labile DOM. Here, taking the uncertainty into account, the possible values for $K_t + S$ range from 31-127 nM and 5-21 nM for the communities at 41.6 °N and 37 °N, respectively. The low-end of these values approach the

SAR11 half saturation value, while the high end approaches the sea-ice diatom values.

The differences in $K_t + S$ at our two sites are likely driven by differences in community composition, with different organisms having different transporters with different specificity and kinetics. *Synechococcus*, picoeukaryotes, and photosynthetic eukaryotes between 20-100 μm were important contributors to phytoplankton biomass at both locations, but picoeukaryotes < 4 μm and 4-20 μm were greater contributors to total biomass in the north than in the south (personal communication, A.E. White, F. Ribalet). These biomass estimates show that the phytoplankton biomass was larger in the north than in the south, tracking changes in chlorophyll concentration (Table 5.1). From this we conclude that higher total phytoplankton biomass in the north is not directly related to GBT uptake kinetics, given that we measure more capacity for rapid GBT uptake (higher V_{max} and lower $K_t + S$) in the south compared to the north. The specific abundance of organisms with different capacities for GBT uptake is likely more important than total phytoplankton biomass.

In contrast to phytoplankton biomass, free living heterotrophic bacteria were more abundant in the south than in the north. The heterotrophic bacterial abundance was 834 $\pm$ 19 cells per μL at 37 °N and 302 $\pm$ 24 cells per μL at 41.6 °N (personal communication, F. Ribalet). The bacterial community composition was likely similar to that sampled in this region in 2017, with significant contributions from SAR11 clade, Rhodobacterales, Cellvibrionales, and Flavobacteriales (Chapter 4 Figures 4.21, 4.20). Heterotrophic bacteria community and abundance may have influenced the GBT uptake parameters, since the presence of more organisms able to take up GBT would increase the V_{max} and decrease the turnover time.

The microbial communities in the North Pacific transition zone studied here had similar GBT uptake kinetic parameters as those estimated in estuarine and coastal waters (Table 5.3) [46, 138]. Turnover time estimates from the estuarine and coastal studies ranged from <1 to 11 h [46, 138], much lower than the estimate of the northern station and just within the range of possible estimates for the southern station (Table 5.3). This suggests that at the more northern station, though there was more biomass, microbes were much less adapted to utilize glycine betaine, as evidenced by the higher $K_t + S$, lower V_{max}, and longer turnover time

compared with the station at 37 N and with the previous coastal estimates. Total organic nitrogen was similar at both of our stations, but the elevated DIN in the north suggests that microorganisms would not need to rely on DON as a nitrogen source. At Station 5, where the southern experiment took place, DIN was not above our limit of detection, and thus more microorganisms may have utilized DON directly, or relied more on the bacterial remineralization of DON to ammonium.

Estimates of nitrate turnover time, based on DIN concentrations and primary productivity rates, were 9 days at 42 °N and 0.16 days at 37 °N (K. Bjorkman, personal communication). These estimates are the result of similar primary productivity estimates (based on ^{33}P, ^{14}C, and ^{18}O) at the two latitudes but different concentrations of DIN. The trend of the DIN turnover times mirrors that for GBT at the two latitudes, with a much longer turnover time on the order of a week at 41.7 N and less than one day at 37 N. The similarity in the trends suggests that the labile DON turnover is linked to the DIN turnover, where environments with low concentrations of DIN have microbial communities that are adapted to rapidly cycle both DIN and labile DON.

The most well studied DON compounds are dissolved free amino acids. In the oligotrophic Sargasso Sea the V_{max} for individual amino acids ranged from 0.006–0.1 nM/h and $K_t + S$ ranged from 0.1–5.5 nM [19]. In the North Atlantic, taurine turnover times ranged from 6–240 h [135]. Comparing the results of our study to these values suggests that GBT is more similar to the amino acids that have longer turnover times and lower uptake affinities (Table 5.3). Use of GBT requires specific enzymatic pathways, and GBT cannot be directly incorporated into protein biomass in the same way that proteinogenic amino acids can be. Thus GBT has less broad utility to all living organisms. However, for some organisms such as *R. Pomeryoi* DSS-3, it can fulfill the C, N, and energy requirements [243].

5.4.3 GBT uptake over time and estimates of GBT fueling bacterial production

In the time course experiments dissolved labeled glycine betaine was rapidly taken up after the addition of 500 nM $^{13}C_5$, $^{15}N_1$-glycine betaine (Figure 5.3 left). The initial time point

samples were filtered directly after the addition of labeled glycine betaine, resulting in a total 'incubation time' of between 15 and 30 minutes before the samples were finished filtering. In this short time period, approximately 0.08 and 0.11 nM was taken up in the north and south respectively, corresponding to uptake rates of approximately 0.17 and 0.38 nM/h for the two experiments. These rates should be evaluated cautiously as samples were being filtered during the entire 'incubation period'. We expect the actual initial uptake rates to be near the V_{max} estimates from the kinetics experiments given that the added GBT concentration was well past saturation (Figure 5.2, Table 5.2).

During the incubation, the fraction of total particulate GBT that was in the fully-labeled isotope form stabilized after two days of incubation (Figure 5.3 right). In both experiments the fully-labeled glycine betaine pool reached a maximum of approximately 85% of the total GBT pool. The dominance of the labeled form of GBT indicates that the uptake of labeled GBT was great enough to result in a substantial increase in the total particulate GBT concentration (Figure 5.3 left). Other isotopologues of glycine betaine also appeared in the particles, including $^{13}C_4$, $^{15}N_1$-GBT (2% of total particulate GBT).

Using estimates of GBT derived C respiration and macromolecular incorporation from studies done with radiocarbon in coastal and estuarine sites [46], we estimate that GBT carbon would fuel 0.89-1.82% of total respiration, and 0.8% of total PC would be GBT or derived from GBT after 50 hours. These estimates are based on many assumptions (see supplemental material for full calculation and discussion), however, they still provide some ballpark values for the quantitative importance of GBT in the microbial community in the North Pacific.

5.4.4 Turnover time of particulate GBT

The rate of labeled GBT accumulation decreased to approximately 0.1 nM/h between the first and second time points of both experiments (see supplemental calculation for estimate of total uptake of labeled GBT). Between these time points there was a small decrease in concentration of unlabeled GBT. If we assume that GBT biosynthesis stopped after the

addition of 500 nM exogenous GBT, a common assumption in tritiated leucine bacterial production measurements, then we can use this decrease as an estimated flux into metabolic pathways or out into the external environment.

If we assume that in unperturbed systems over time there would be no long term accumulation of particulate GBT, we can use the original particulate $^{13}C_0$,$^{15}N_0$-GBT concentration and the loss rate of $^{13}C_0$,$^{15}N_0$-GBT between the first two timepoints to determine the turnover time (residence time) of GBT in marine particles.

$$TurnoverTime = [GBT_{13C0,15N0}]/loss_{^{13}C_0,^{15}N_0} \tag{5.2}$$

These assumptions yield a turnover time of particulate GBT of 13 and 11 hours at the northern and southern stations, respectively, with corresponding turnover rates of 1.78 d^{-1} and 2.11 d^{-1}.

Both taxonomy and growth rate likely influence metabolite turnover rates. However estimates from model organisms grown in culture can contextualize the turnover rates we measured. The metabolites with the most rapid turnover rates are cofactors such as ATP and ADP, which have turnover times of less than one second in model bacteria *Escherichia coli* and fungi *Penicillium chrysogenum* and *Saccharomyces cerevisiae* [244]. In these organisms, amino acids have turnover times ranging from 10s to 1,000s of seconds (<1 to 10s of minutes) [245, 244]. In a model halophilic methanogen, *Methanohalophilus* strain FDF1, the turnover time of glycine betaine was 45.5 h [246]. This long turnover time was interpreted as evidence that *Methanohalophilus* uses GBT primarily as an osmolyte and not in other metabolic roles. In contrast, the turnover times of other potential compatible solutes in *Methanohalophilus*, glutamate and glycosylglycerate, were 5-3.4 h and 10 h, which suggested that these metabolites were used as metabolic intermediates in addition to their roles as osmolytes. The particulate glycine betaine turnover times measured here (13 and 11 h) are more similar to the turnover times of *Methanohalophilus* glucosylglycerate than *Methanohalophilus* GBT. This provides support for the hypothesis that while some marine

organisms likely only use GBT for its osmotic properties, other members of the natural microbial community metabolize it for C, N, and energy.

5.4.5 Primary pathways of GBT transformation

We see evidence of the direct incorporation of GBT into metabolites, reflected by the incorporation of multiple ^{13}C and ^{15}N atoms before the $^{13}C_1$ isotopologue or $^{15}N_1$ isotopologues increased in concentration (Figure 5.4, Figure 5.5). The appearance of other isotopologues, including increased concentrations of the $^{13}C_1$ or $^{15}N_1$ isotopologues, suggests that the C and N from GBT was further recycled through metabolic pathways to the point where individual atoms from the molecule began appearing in metabolites distantly related to GBT.

Four major pathways of glycine betaine metabolism were observed to occur in both incubation experiments (Figures 5.1, 5.6). GBT-derived isotopologues of metabolites involved in the demethylation pathway (DMG, sarcosine, creatine), the reduction pathway (TMAO), the carnitine synthesis pathway (carnitine, TMAB), and the choline synthesis pathway (choline, GPC) were observed within the first four to six hours of incubation in both experiments (Figure 5.1, Figure 5.4). However, the two experiments differed in which pathways contributed more to GBT transformation, summarised in Figure 5.6.

Demethylation: A primary route of GBT metabolism is via demethylation to dimethylglycine and sarcosine (black pathway in Figures 5.1, 5.6), and this was the dominant pathway of GBT transformation that we detected in the south. Unlabeled dimethylglycine was not above our limit of detection (due to elevated noise in the mass spectrometry signal at that retention time and *m/z*). However, $^{13}C_4$, $^{15}N_1$-dimethylglycine was detected. Though we cannot calculate mass isotopologue distributions without the unlabeled metabolite measurement, the peak area of the labeled dimethylglycine increased over time in both incubations, and appeared at the second sampling point in both experiments (Supplemental Figure 5.11). Labeled $^{13}C_3$, $^{15}N_1$-sarcosine also appeared in the particulate pool at the second time point of each experiment (Figure 5.5). The $^{13}C_3$, $^{15}N_1$-sarcosine isotopologue reached a maximum of $30\% \pm 8\%$ and $78\% \pm 4\%$ of the total sarcosine pool in the north and south experiment

respectively. In the south, sarcosine also appeared in the fully labeled form much faster than in the north (Figure 5.4), with 46% ± 6% of the total sarcosine pool fully labeled after 6 hours.

The initial concentration of sarcosine was lower in the south experiment than in the north (Supplemental Figure 5.14), together with the isotope distribution this suggests that the flux through the compounds in the demethylation pathway was higher in the south experiment than in the north. Interestingly, the $^{13}C_4$, $^{15}N_1$ isotopologue of dimethyl glycine ultimately reached a higher concentration (proportional to peak area) in the north experiment than the south experiment (Figure 5.11). This is counter to the trend for sarcosine which showed the $^{13}C_3$, $^{15}N_1$ isotopologue of sarcosine consistently in lower concentration in the first experiment compared to the second. This suggests that DMG is accumulated in the north but not in the south, where it stays in roughly the same proportion to sarcosine. The methyl group removed in the GBT to DMG transformation can be used to methylate homocysteine to methionine, as discussed below. The flux through each step along the demethylation pathway in the two locations may be influenced by the abundance of SAR11. In 2017, SAR11 contributed more to the bacterial 16S gene copies at 37 °N than 41 °N, and SAR11 has the capacity to demethylate GBT to glycine using the enzymes betaine-homocysteine methyltransferase (BHMT), dimethylglycine dehydrogenase, and sarcosine dehydrogenase [45]. The SAR11 and *Ruegeria pomeroyi* DSS-3 BHM enzymes transfer a methyl group from GBT to homocysteine to form methionine, but the latter two demethylations result in a methylation of tetrahydrofolate and ultimate oxidation to CO_2 [45, 243].

Glycine was measured in the samples but no isotopologues were detected. The natural abundance of the $^{13}C_1$ glycine isotopologue should be approximately 2% of total glycine, but we could not detect that isotopologue at any time point. Based on our limit of quantification and the concentrations of unlabeled glycine measured in each experiment, we estimate that the isotopologues could have been enriched up to approximately 5% of the unlabeled glycine concentration and remain undetectable.

Creatine was labeled with $^{13}C_3$ $^{15}N_1$ in the second time point in both experiments at 2%

and 4% in the north and in the south, respectively. The $^{13}C_3$ $^{15}N_1$ isotopologue comprised approximately 7% and 6% of the total creatine pool after 36 hours of incubation in the two experiments (Figure 5.4, Figure 5.5). Creatine is primarily thought to be synthesized from glycine and arginine, via glycine aminotransferase and guanidinoacetate *N*-methyltransferase [247] and can be produced as a byproduct of the urea cycle [248]. If that were the dominant pathway in this system, we would have seen the appearance of the $^{13}C_2$ $^{15}N_1$ isotopologue rather than the $^{13}C_3$ $^{15}N_1$ isotopologue (Figure 5.1).

The accumulation of the $^{13}C_3$ $^{15}N_1$ isotopologue rather than the $^{13}C_2$ $^{15}N_1$ isotopologue indicates that the production of creatine from sarcosine might have been dominant in these communities. It is possible that creatine was synthesized from sarcosine (via creatinase in Figure 5.1) but that transformation is primarily discussed in the literature in the reverse direction, synthesizing sarcosine from creatine [249]. Our data suggest that either creatinase is a reversible enzyme or there is a yet undiscovered enzyme that catalyzes the reaction from sarcosine to creatine.

Reduction: Another possible pathway for the breakdown of glycine betaine is through the production of trimethylamine (TMA) via betaine reductase (blue pathway in Figures 5.1, 5.6). TMA was not measured, but TMAO, produced from TMA via TMA monooxygenase (tmm), was detected as fully $^{13}C_3$ $^{15}N_1$ labeled TMAO at the second time point in both experiments. Labeled TMAO showed greater accumulation in the north than in the south, the opposite trend of sarcosine (Figure 5.4, Figure 5.5). The final proportion of fully labeled TMAO in the north was $49\% \pm 10\%$ versus $20\% \pm 4\%$ in the south.

The tmm gene, which transforms TMA to TMAO, is found in both prokaryotes and eukaryotes [250, 95, 43]. TMAO is an osmolyte used by mammals and marine animals including sharks, worms, and snails [250] and can be an intermediate along the way to the formation of other methylamines and further oxidized to provide energy [251, 45]. The higher concentration of TMAO in the north (Figure 5.14) and its more rapid isotopologue accumulation in the northern experiment suggest that organisms in the northern portion of the NPTZ are actively producing TMAO and also further metabolize it for energy production. TMAO can

be formed from choline (via choline TMA-lyase, cutC) and carnitine (via carnitine oxygenase, cntA). We cannot rule out these pathways of TMAO production, but the difference in isotope accumulation between TMAO, choline, and carnitine suggests that they are from independent pathways. This is most obvious in the northern station, where the potential precursors choline and carnitine have much lower isotopologue abundances than TMAO and thus are unlikely to be the primary precursors to TMAO (Figure 5.4, Figure 5.5).

Choline: Though glycine betaine is thought to be formed from choline [243, 252], our results show that choline is also likely directly formed from glycine betaine, as evidenced by the $^{13}C_5$ $^{15}N_1$ isotopologue accumulation (Figures 5.4, 5.5). Choline and glycerophosphocholine (GPC) were in higher concentrations in the north than in the south (Figure 5.14). Nevertheless, $^{13}C_5$ $^{15}N_1$ isotopes accumulated more rapidly in the north. Ultimately 25% of the total choline and GPC metabolite pools were in the $^{13}C_5$ $^{15}N_1$ isotopologues, directly sourced from the labeled GBT. In the south these metabolites showed slower accumulation but reached similar final isotope proportions (Figure 5.4, Figure 5.5). Overall in the southern experiment these metabolites must have slower turnover times (Figure 5.6). GPC is made from choline-phospholipids. We saw the additional accumulation of the GPC $^{13}C_6$ $^{15}N_1$ and $^{13}C_7$ $^{15}N_1$ isotopologues, indicating that the glycerol group was being labeled with ^{13}C. These patterns suggest further metabolic reworking of GBT derived carbon, as discussed below.

Carnitine: Carnitine became labeled with $^{13}C_5$, $^{15}N_1$ in the second time point in both experiments and reached 3% and 2% in the 9 h and 13 h time points for the northern and southern experiments respectively. By the end of the incubations 25% and 37% of carnitine was the $^{13}C_5$, $^{15}N_1$ isotopologue. There was no increase in the $^{13}C_1$ isotopologue or in a $^{15}N_1$ isotopologue in either experiment. (3-Carboxypropyl)trimethylammonium, also known as trimethylammoniobutanoate (TMAB), is a quaternary amine that is often thought to be a precursor to carnitine but in these experiments seems to be derived from carnitine. The $^{13}C_5$ $^{15}N_1$-TMAB isotopologue increase in concentration throughout the course of both experiments, going from 0% to 6% of the total TMAB pool in the north and from 0% to 32% of the total TMAB pool in the south. The $^{13}C_6$ $^{15}N_1$-TMAB isotopologue was detectable in

the southern experiment in the latter time points (<1% of total TMAB). No other TMAB isotopologues were detected. As seen with carnitine,there was a greater rate and extent of incorporation in the south than the north (Figure 5.4, Figure 5.5).

Carnitine showed similar total concentration in each experiment (Figure 5.14) but very different patterns of isotope accumulation, with higher carnitine turnover in the south. The isotope pattern in (3-Carboxypropyl)trimethylammonium (TMAB) suggests that it is derived from carnitine. TMAB showed dramatically different patterns in the experiments, reaching only 7% enrichment in the north experiment but nearly 40% enrichment in the south, despite their similar concentrations (Figure 5.14). This suggests that organisms in these communities have very different TMAB synthesis strategies, either stemming from different pathways or different strategies for TMAB use and intracellular turnover.

5.4.6 Evidence of C and N recycling

Several amino acids, nucleosides, and nucleobases did not show a direct incorporation of multiple isotopes from glycine betaine, but do show gradual increases in the ^{13}C and ^{15}N content (Figure 5.7). Gradual increases in the ^{13}C and ^{15}N content of amino acids, nucleosides, and nucleobases indicated that these elements, derived from GBT, were recycled through cellular metabolism (Figure 5.7, Figure 5.8). This could occur via the breakdown of glycine betaine to ammonia and metabolic intermediates such as acetylphosphate and formaldehyde (Figures 5.1 and 5.6 blue pathway) or via demethylation to GBT which then methylates s-adenosyl methionine with ^{13}C (Figures 5.1 and 5.6 black pathway). Changes in the isotopologue distribution were assessed by fitting a linear model to the mass isotope distribution of each isotopologue over time. Significant increases in isotopologue distributions ($p < 0.05$) are discussed below.

5.4.7 GBT as a N source

The amino acids glutamic acid, glutamine, and arginine had increased proportions of their $^{15}N_1$ isotopologues in both experiments over time, with a greater increase in the south than

the north. Glutamine and glutamic acid are the primary molecules involved in ammonium assimilation, and the appearance of ^{15}N suggests that appreciable amounts of ammonium had been generated from GBT after 1-2 days. In *E. coli*, the turnover rate constants for glutamine and glutamate are 14.29 min^{-1} and 0.79 min^{-1} in exponential growth with a doubling time of ~80 minutes [245]. These are likely much faster than the turnover rates for glutamate and glutamine in this system, since the growth rate of *E. coli* is much faster than the average growth rate of marine microorganisms. However, rapid turnover of these compounds still is quite likely, and thus the proportion of ^{15}N that they contain may be a near real time estimate of the proportion of ^{15}N in available ammonium. Given this interpretation, it is clear that in the south there was much higher recycling of ^{15}N from GBT into ammonia than in the north (Figure 5.7). Together with the undetectable nitrate and nitrite concentrations at station 5, where the southern experiment took place, this observation suggests that organisms in the heart of the NPTZ rely on nitrogen recycling within the photic layer.

In the north there was detectable DIN (Table 5.1), and based on the DIN turnover time estimates, the incubation lasting four days would not be long enough for the standing stock of DIN to be consumed. Thus, throughout the first incubation many organisms could access nitrite or nitrate and did not need to rely on remineralized DON to support their growth. Some organisms require ammonia or preferentially use it over nitrite and nitrate, and the slight increase in ^{15}N isotopologues of the free amino acids in the north likely are from organisms assimilating ammonia derived from GBT.

Select other metabolites showed ^{15}N incorporation (Figure 5.7). The concentration of the $^{15}N_1$ isotopologue of taurine increased gradually in the north and increased after several days of incubation in the south. There was a significant increase over time of the proline $^{15}N_1$ isotopologue, increasing from 0.1% to 0.2% of the total proline pool in the north and from 0.1% to 0.6% of the total proline pool in the south. The last time point of the southern experiment showed $^{15}N_1$ incorporation into alanine, with $^{15}N_1$ isotopologue increasing to 0.3% $\pm$ 0.2% from 0%. The northern experiment showed no change in the contribution of the alanine $^{15}N_1$ isotopologue over the course of the experiment. These results indicate

again that more GBT derived nitrogen was incorporated into biomolecules in the low nitrate community in the south than in the north.

Many metabolites did not show evidence of N recycling. Some of this could be due to the isotopologues being below our limit of detection, but in other cases we can measure isotopologues and they do not change in relative abundance. For example, leucine showed no evidence of ^{15}N isotopologue accumulation. Similarly, the beta-alanine betaine $^{15}N_1$ isotopologue showed no change in either experiment.

The total concentration of N in glutamic acid rivals the total concentration of N in glycine betaine (Supplemental Figure 5.14), but the organisms that contain these two molecules are not identical. GBT is only in a subset of organisms (though maybe many more than normal after the addition of 500 nM exogenous GBT) while glutamate is in all organisms. Thus, the slow growth of the ^{15}N isotopologue of glutamate reflects full N-remineralization and recycling. Though it is possible that the organisms that can remineralize GBT to ammonium do not release that ammonium and are thus the only organisms able to access that ammonium, it seems unlikely to be the case. In culture, the remineralized nitrogen from TMA and TMAO in *R. pomeroyi* DSS-3 was used by other bacteria in co-culture [253]. In our study, the appearance of ^{15}N in other molecules such as arginine, taurine, homarine, and trigonelline at the later time points also show similar patterns to glutamine and glutamate and further support the hypothesis that ^{15}N-ammonia derived from GBT is being recycled throughout the microbial community after several days of incubation at both stations.

Purine vs pyrimidine metabolism

The purine adenine showed enrichment in the $^{15}N_1$ and $^{13}C_1$ isotopologues after two days of incubation in both experiments (Figure 5.8). The southern experiment also showed increasing contribution of other isotopologues over the course of the incubation, including increased proportional abundance of $^{13}C_2$, $^{13}C_1$ $^{15}N_1$, $^{13}C_2$ $^{15}N_1$, and $^{13}C_3$, $^{15}N_1$. Isotopologues of the related nucleoside adenosine had similar trends as adenine (Figure 5.8), with a small but significant increase in relative contribution of the single ^{13}C and ^{15}N labels in both

experiments and significant increases in the multiple labels starting in the south around T25. Adenosine showed more GBT-derived N-incorporation in the southern experiment compared to the northern experiment, with a greater increase in the $^{15}N_1$ isotopologue and the appearance of the isotopologue with $^{13}C_2$, $^{15}N_1$ at later time points. Deoxyadenosine had similar patterns to adenosine, but with less isotopologue abundance across the board at all timepoints and both experiments. This might indicate that there was less turnover in the deoxyadenosine pool than the adenosine pool, as we would expect based on the slower turnover of DNA compared with RNA.

Guanine, a purine like adenine, showed a similar pattern to adenine. Signs of C and N recycling from GBT were in both experiments but, as with adenine, more in the south than the north. For example, the isotopologue with $^{13}C_2$ increased from 0.1% to 0.6% in the north while it increased from 0.1% to 0.8% in the south (Figure 5.8). Guanosine showed small increases in ^{15}N incorporation over the course of the southern experiment, but not the northern experiment. ^{13}C incorporation was shown in an increase in the $^{13}C_1$ isotopologue in the north (from 5% to 8%) and in the south (from 3.9% to 5.6%). Any increases in the ^{15}N isotopologues of guanosine could not be measured because they were below our limit of detection.

Cytosine showed ^{13}C and ^{15}N accumulation, but less than adenine, potentially because it simply has fewer C and N atoms to accumulate the label and potentially due to difference in biosynthetic pathways, since cytosine is a pyrimidine (Figure 5.8). In the north there was more C-recycled through metabolic intermediates into cytosine, evidence by the accumulation of $^{13}C_1$-cytosine which increased from 2.3% to 3% of the total cytosine pool. There was no significant increase in this isotopologue in the south. In contrast, more N-recycling occurred in the south. The $^{15}N_1$ isotopologue approximately doubled from 0.4% to 0.9% in the north while it quadrupled in the south,increasing from 0.5% to 2.1% over the incubation period (Figure 5.8).

Pyrimidines are formed *de novo* from glutamine, HCO_3^- and aspartate, while purines are formed *de novo* from 2 molecules of formate, 2 molecules of glutamine, HCO_3^-, glycine, and

aspartate. The relatively high accumulation of the $^{13}C_2$ $^{15}N_1$ isotopologues of adenine and guanine to 1% and 0.7% in the final time point of the south experiment suggests isotope accumulation from glycine, which contributes two carbons and one nitrogen atom to purines. This is further evidence that glycine was directly formed from GBT, though we could not detect any heavy isotopologues (Figure 5.6).

Similarly to glycine, no accumulation of heavy aspartate isotopologues was observed. The natural abundance of the $^{13}C_1$ isotopologue for aspartate should be at approximately 4.4%. Since we did not observe this isotopologue at any time point, other isotopologues could have been greater than 4.4% of the aspartate pool and still been below our detection limit. Based on our limit of quantification we estimate that isotopologues would have been detected if they were much more than 5% of the unlabeled aspartate concentrations we measured, so we cannot rule out aspartate in contributing to the increase in ^{13}C or ^{15}N isotopologue concentration in pyrimidines. Both pyrimidines acquire one nitrogen atom and purines acquire two nitrogen atoms from glutamine, and we did observe an increase in the ^{15}N isotope abundance in glutamine. Thus it is most likely that much of the increase in ^{15}N in purines and pyrimidines is from glutamine.

Regardless of the sources, the patterns observed in these metabolites suggest that there is significant recycling of C and N from GBT throughout the primary metabolites, beginning slowly in the southern experiment, and increasing after the first $\sim$13 hours, since T25 is where multiple labels begin to show up in significant proportions.

Total hydrolyzable amino acids

Over the four days of incubation, the concentrations of total hydrolyzable amino acids (THAA) generally increased by approximately 150% in the northern experiment and generally decreased by approximately 90% in the southern experiment (Figure 5.9). The average initial concentration of the THAAs in the two experiments was similar, but both experiments had high bottle to bottle variability (Figure 5.9). ^{13}C and ^{15}N isotopologues of several hydrolyzed amino acids were detectable. We evaluated for significant enrichment in the

isotopologues of the THAAs but there were no significant differences in the isotopic distributions at the initial and final timepoints (false discovery rate corrected p-value > 0.1). In the southern experiment the proportional contribution the $^{13}C_2$ and $^{13}C_3$ isotopologues of phenylalanine, isoleucine, leucine, and valine and $^{13}C_2$ $^{15}N_1$ isotopologues of phenylalanine, isoleucine, and alanine increased a small amount between the initial and final timepoints, and were significantly different based on a two sided t-test before fdr correction (p-value $<$ 0.05, Figure 5.10). Some of these isotopologues were below our limit of detection in the initial time point and increased to be above our limit of detection at the final timepoint, despite an overall decrease in the THAA concentrations (Figure 5.9). The total increase in ^{13}C and ^{15}N concentrations from total hydrolyzable amino acids in the southern experiment was 0.8 nM and 0.1 nM, respectively, a relatively low contribution to the total isotope uptake from GBT (Figure 5.12).

The accumulation of multiple ^{13}C atoms rather than the single $^{13}C_1$ isotopologue, suggests multiple isotopes together are being delivered from GBT to these amino acids. Of the five amino acids that showed isotope enrichment, four are synthesized from pyruvate. This provides further evidence that GBT is demethylated to glycine, which can then be converted to serine and then pyruvate (Figure 5.6). The other amino acid, phenylalanine, has 3 carbons derived from phosphoenolpyruvate (PEP), which can also originate from pyruvate. The ^{15}N accumulating in these is likely derived from free glutamine, which showed ^{15}N enrichment (Figure 5.7). Ultimately, these enrichments show that the carbon and nitrogen from GBT can provide marine microbes C and N for biomass production.

5.4.8 GBT as a C source

Glutamic acid and glutamine showed increased proportions of the $^{13}C_1$ isotopologue over time in both experiments (Figure 5.7). Gonyol, trigonelline, and dimethylsulfoniopropionate (DMSP) showed ^{13}C over time. The beta-alanine betaine $^{13}C_1$ isotopologue and $^{13}C_2$ isotopologue increased in the north and south, with a greater increase in the south. In contrast, the $^{13}C_1$ isotopologue of beta-Alanine was stable at around 1.8% throughout all experiments

and time points.

Several metabolites that had shown ^{15}N incorporation did not have significant ^{13}C incorporation, including arginine, taurine, and proline. Proline betaine, alanine, and leucine also did not show ^{13}C incorporation

Though we did not measure glycine isotopologues, it is likely that they did accumulate as the product of sarcosine demethylation (Figure 5.6). Glycine, along with glutamic acid and glutamine are glucogenic amino acids. Glycine can be converted to pyruvate and then oxaloacetate. Oxaloacetate can then fuel gluconeogenesis or directly enter the citric acid cycle and provide energy to the cell. Glutamate and glutamine enter the citric acid cycle through alpha-ketoglutarate. Accumulation of ^{13}C in these amino acids indicates that carbon from GBT is a viable source of C and energy in natural marine microbial communities.

GBT as a methyl donor

GBT donates methyl groups to make methionine from homocysteine with the gene BHMT [243]. The other pathways for methionine synthesis are via methionine synthase (MetH), which uses cobalamin (Vitamin B_{12}) and methyl tetrahydrofolate as cofactors, and MetE which directly transfers the methyl from methyltetrahdyrofolate to homocysteine. Methionine can be used for protein synthesis or to generate s-adenosyl methionine (SAM) from s-adenosyl homocysteine (SAH) via methionine adenosyltransferase. This use of GBT has been shown to occur in both prokaryotes and eukaryotes [254, 243]. Though we cannot detect the $^{13}C_1$ natural abundance or label accumulation in methionine or SAM, the turnover of methyl groups through these metabolites is likely extremely rapid (minutes to seconds), so we would not necessarily expect an accumulation of the $^{13}C_1$ isotopologue. Evidence for GBT acting as a methyl donor for methionine and SAM in our experiments comes from the accumulation of ^{13}C in beta-alanine betaine and DMSP. Beta-alanine betaine is made through successive methylations of beta-alanine via a SAM dependent methyltransferase [255]. Beta-alanine betaine showed changes in ^{13}C isotopologues even though beta-alanine did not, indicating that the ^{13}C was on a methyl group and came from SAM. The increase

in the beta-alanine betaine ^{13}C isotopologues were greater in the south than in the north, similar to our estimated flux through sarcosine. DMPS is synthesized via a SAM dependent methyltransferase as well [256]. In the southern experiment in particular, $^{13}C_1$ and $^{13}C_2$ isotopologues of DMPS increased their contribution to total DMSP. Gonyol is potentially produced from DMSP [257] and showed similar isotopologue accumulation. These ^{13}C atoms are likely on the methyl group carbons, and thus ultimately originated from SAM.

Though isotopologues for SAM and SAH were consistently below our detection limits, these results indicate that direct methylation of SAM from GBT and then to other betaines may be more important in the south than in the north. Additional evidence for the activity of GBT demethylation and methylation pathway is provided by the appearance of the $^{13}C_4$, $^{15}N_1$-GBT at 2% of total GBT concentration by the end of the incubations. $^{13}C_4$, $^{15}N_1$-GBT is likely to have been formed from the re-methylation of DMG.

This methylation potential from GBT is likely not available to all organisms that can take up GBT. Evidence for different organisms having access to this pathway comes from trigonelline and homarine. Homarine and trigonelline are structural isomers with different synthesis pathways and differences in the identity of some of the organisms that produce them [41]. Analyses of metabolomes from diverse marine microbes showed that homarine was made primarily by eukaryotic phytoplankton, *Synechococcus*, and one heterotrophic bacteria, *Roseovarius* sp., while trigonelline was produced by eukaryotic phytoplankton and by several other heterotrophic bacteria [41]. In both experiments these two compounds showed different isotope accumulation patterns (Figure 5.7). Homarine is produced from quinolinic acid, which in turn is a catabolite of tryptophan [258]. Quinolinic acid is decarboxylated to picolinic acid, which is then methylated via SAM to homarine. Trigonelline is synthesized by methylating nicotinic acid, also catalyzed by SAM [259]. Nicotinic acid is primarily produced as a degradation product of NAD, but can also be produced from quinolinic acid, with nicotinic acid mononucleotide as an intermediate.

The $^{13}C_1$ and $^{13}C_2$ isotopologues of trigonelline increased in both experiments, with a greater increase in the south than in the north. Most of the change in the southern experiment

occurred after 51 hours of incubation (Figure 5.7). Homarine showed no change in the $^{13}C_1$ isotopologue which was steady at 7% in both experiments. The differences in the $^{13}C_1$ isotopologue abundance in trigonelline and homarine either suggest different labels originating in different pathways (and thus support the NAD degradation pathway over the quinolinic acid pathway for trigonelline) or suggest different organisms synthesize these compounds. Homarine is thought to primarily be an osmolyte [260, 112] while trigonelline can be an osmolyte but also can be a reservoir for nicotinic acid or a detoxification method in cases of excess nicotinic acid [259]. Homarine and trigonelline both increased in concentration in the first experiment with a similar pattern to many metabolites (Figure 5.15). These increases indicate that there was synthesis of these metabolites during the incubation, so if ^{13}C was available in the precursors to homarine, it should have appeared in the isotopologues, which were well detected (Figure 5.7).

We hypothesize that the organisms in our samples that produce homarine are not the same organisms that can degrade GBT to fuel methylation, but that at least a subset of the organisms that produce trigonelline are able to fuel methylation via GBT degradation.

Incorporation of ^{15}N into trigonelline occurred at the end of the incubation periods in both experiments, with the $^{15}N_1$ isotopologue increasing to 0.27% from approximately 0.2% in all prior timepoints for the northern experiment and to 0.4% from approximately 0.17% in all prior time points in the southern experiment. Similarly, in homarine, there was a small change in the contribution of the $^{15}N_1$ isotopologue at the final time point, increasing from 0.2% to 0.4% in both experiments, a change similar to that seen in trigonelline (Figure 5.7). The small increase in ^{15}N isotopologues in both trigonelline and homarine at the final time point further emphasizes that ammonia recycling from GBT was making its way through the entire microbial community, even to organism that may not be able to remineralize GBT themselves.

5.4.9 Total uptake and fate of GBT

Though we saw metabolites incorporate isotopes from GBT, many other targeted metabolites were either not detected or did not have detectable isotopes and thus cannot be evaluated (Supplemental table 5.5).

Total uptake of ^{15}N and ^{13}C isotopes was calculated by taking the sum of all the quantifiable metabolites and their ^{15}N and ^{13}C isotopes (Supplemental Figure 5.12). This total is an underestimate due to only a selection of metabolites and their isotopologues being detectable and quantifiable with our methods. However, it still shows that the total ^{15}N and ^{13}C uptake was dominated by the $^{13}C_5$, $^{15}N_1$-glycine betaine at all time points. In the final time point of the northern experiment, 85 ± 19% of all particulate ^{15}N that we could measure in our metabolites and 80 ± 17% of all particulate ^{13}C that we could measure in our metabolites was in GBT isotopologues. In the final time point of the southern experiment, 88 ± 7% of ^{15}N and 86 ± 7% of ^{13}C was in GBT isotopologues (Figure 5.13). The total ^{15}N in metabolites we quantified was 4 nM and 2 nM at the final time point of the north and south experiments, a small fraction of the total 500 nM added (Figure 5.12). We consider these values an underestimate of the actual GBT uptake. Incubations of radiocarbon-labeled glycine betaine have shown that most C taken up as GBT is respired into CO_2 or incorporated into macromolecules [46]. We see incorporation into choline and glycerophosphocholine and purines and pyrimidines, suggesting C and N from GBT could be incorporated into lipids and DNA and RNA - molecules which we did not analyze here.

Much of the ^{13}C measured was in homarine (a compound with 7 carbon atoms), which did not accumulate enriched ^{13}C over the course of the experiments (Figure 5.7). Homarine was simply at high enough particulate concentrations that the natural abundance ^{13}C isotopologue contributed substantially to our total ^{13}C measured (Supplemental Figures 5.15, 5.16). Similarly, there was significant ^{15}N and ^{13}C in THAA at the initial and final time points, due to the natural abundance of these isotopes and the high concentrations of THAAs. The additional contribution of THAA due to enrichment from GBT was 0.1 nM ^{15}N and 0.8 nM

^{13}C in the southern experiment, only a small fraction of that found in GBT (Supplemental Figure5.17). There was no enrichment of isotopes from GBT into THAA in the northern experiment (Figure 5.10), increased concentrations of these isotopes were instead due to the overall increase in THAA (Figure 5.9).

We were interested in whether the GBT uptake was only by heterotrophic bacteria or if phytoplankton also contributed to the uptake. We looked for isotope accumulation in chlorophyll and found none, indicating that GBT may not be supplying substantial carbon or nitrogen to photosynthetic organisms. This does not rule out phytoplankton GBT uptake. For example, diatoms can take up GBT, with similar uptake patterns as they have for DMSP [261]. However, diatoms do not show evidence of metabolizing GBT [42]. Phytoplankton could benefit from GBT uptake simply for its osmotic properties, since they would not have to synthesize other energy, nitrogen, and carbon requiring compatible solutes. We did observe an increase in total osmolyte concentrations in the northern experiment, including in homarine, gonyol, DHPS, and isethionic acid which are thought to be primarily or exclusively produced by phytoplankton [41]. This could be explained by an overall increase in biomass, potentially due to a decrease in grazing pressure that likely was an unintended consequence of a bottle incubation. Additionally, ^{13}C and ^{15}N from GBT accumulated in phytoplankton osmolytes, such as gonyol, DMSP and homarine, indicating that either these have larger bacterial sources that previously considered, or that the C and N are directly or indirectly exchanged between microorganisms over the timescales of these incubations.

In order to determine the intracellular concentration of GBT, we estimated the biovolume of heterotrophic bacteria, *Synechococcus*, *Prochlorococcus*, and picoeukaryotes from flow cytometry samples taken at the stations where we did our incubations. If we assume that all GBT is in heterotrophic bacteria, initial GBT intracellular concentrations would be 8 and 3 mM in the north and south, respectively. If instead, we assume that GBT was evenly distributed across the organisms, intracellular concentrations would be 1.6 and 1 mM in the north and south, respectively. By the final time point, if there was no biomass change, these intracellular concentrations would increase 6 fold in the north and 3 fold in the south,

in proportion to the increases in total particulate GBT (Figure 5.3, Supplemental Figure 5.14). Compatible solutes classically have high intracellular concentrations, so these GBT concentrations would be reasonable. The changes in THAA concentrations (Figure 5.9) and osmolyte concentrations (Supplemental figure 5.16) suggest that biomass did not change dramatically, and instead intracellular GBT concentrations did increase substantially. We cannot rule out the possibility that all labeled GBT was taken up into bacterial cells, but we hypothesize that diverse organisms including phototrophic and mixotrophic organisms were responsible for the uptake and storage of GBT. Further studies should more closely investigate the capabilities of phytoplankton to compete for and take up GBT and other labile dissolved organic matter, since these molecules can provide a benefit to organisms even if they cannot be metabolized.

5.5 Conclusion

Using stable isotope tracing metabolomics, we find that microbial communities in the north Pacific transition zone can take up glycine betaine and transform it via both known and previously unrecognized pathways, to ultimately provide C, N, and energy. The uptake kinetics parameters were within the ranges expected from previous studies of GBT uptake in natural seawater and cultured organisms. The turnover time estimates for dissolved GBT are similar to the turnover time estimates of DIN and non-proteinogenic dissolved free amino acids and amino acid-like compounds. The southern station, in the heart of the NPTZ (37 N) where there were more heterotrophic bacteria and less DIN, was more tuned to take up GBT (Figure 5.2). The northern station, where DIN is in the μM range, showed more GBT taken up overall (Figure 5.3, Supplemental Figure 5.12). We hypothesize that more phytoplankton biomass in the north enabled more organisms to take up GBT over time, leading to the higher particulate GBT concentrations. In contrast, less GBT was metabolized through the demethylation pathway to glycine and pyruvate in the north than the south, leading to less prevalent ^{13}C and ^{15}N enrichment in other metabolites and macromolecules (Figure 5.7).

The vast majority of the ^{15}N and ^{13}C that we could detect with our metabolomics methods

remained in $^{13}C_5$, $^{15}N_1$-glycine betaine. Though the major use of GBT is as an osmolyte maintained at high intracellular concentrations, we show that a portion of the GBT taken into cells is rapidly transformed to other metabolites (Figures 5.4, 5.5, 5.6) and ultimately the C and N from GBT appear throughout metabolic products ranging from amino acids to nucleobases to lipids, suggesting full recycling through pyruvate, the citric acid cycle, and ammonium assimilation (Figures 5.7, 5.8).

Though we expect natural dissolved GBT concentrations to be much lower than 500 nM, these observations still implicate GBT as a relevant labile DOC and DON source that fuels the microbial loop. The difference between the two stations suggests that the DIN concentrations and microbial community composition influences how DON is used. Given that labile DON is produced locally by phytoplankton exudate and mortality [222], our data support a feedback loop between phytoplankton primary producers and heterotrophic reminilizers enabled through DON release, uptake, and remineralization that enables efficient nitrogen cycling within nitrogen limited areas of the ocean. This study highlights insights into the microbial community carbon and nitrogen cycle that can be gained through employing stable isotope tracing metabolomics. Further studies with labeled GBT in cultures will reveal the organisms that are capable of which transformations and how those pathways might be influenced by substrate availability and other environmental factors. In the marine environment, stable isotope probing coupled with stable isotope tracing metabolomics could provide further insight into which organisms are able to access the C and N of GBT most quickly.

5.6 Acknowledgement

The authors acknowledge L. Carlson and A. Meyers for lab assistance; L. Carlson, K. Heal, W. Kumler, N. Kellogg, J. Sacks, R. Lionheart for feedback and assistance on data analysis and interpretation; K. Heal for help conducting the experiments; A. Torstensson for input on the experimental design; the crew and scientific party of the R/V Kilo Moana during Gradients-3 cruise (KM1906). This work was supported by grants from the Simons Foun-

dation (LS Award ID: 385428, A.E.I.; SCOPE Award ID 329108, A.E.I.) and the National Science Foundation (NSF OCE-1228770 and OCE-1205232 to A.E.I., NSF GRFP to A.K.B.)

5.7 Tables

										Temperature		Chl	
	Date	Lat	Long	Depth	PC	PN	C:N	TON	DIN	mean	sd	mean	sd
North Exp - St 4	2019-04-16	41.68	-158	15	4.24	0.71	5.97	4.40	6.42	9.59	0.05	0.85	0.08
South Exp - St 5	2019-04-20	37.00	-158	15	6.69	0.65	10.29	4.29	NA	13.47	0.09	0.57	0.03

Table 5.1: Summary of the environmental conditions where samples were collected in this study. Depth is in meters. PC, PN, TON, and DIN are in μM. Chlorophyll is in μg/L. Temperature is in °C.

		Model estimates		
	Iterations	Kt + S (nM)	Vmax (nM/hr)	Turnover Time (hr)
North kinetics experiment				
Linear Transformation	970	50.16	0.36	137.8
		(13.78)	(0.06)	(31.5)
Nonlinear least squares	935	79.46	0.36	
		(48.26)	(0.06)	
South kinetics experiment				
Linear Transformation	975	13.15	0.57	23.1
		(8.34)	(0.06)	(14.4)
Nonlinear least squares	935	11.06	0.56	
		(5.53)	(0.06)	

Table 5.2: The K_s, V_{max}, and turnover time. Uptake kinetics values are reported as a mean and standard deviation calculated as discussed in the methods section. Values are reported as: mean (sd).

Compound	Location	Vmax (nM/L/h)*	Kt + S (nM)*	Turnover time (h)*	Reference
Glycine betaine	N. Pacific (41 N)	0.3 - 0.42	31.2 - 127.72	106 - 169	this study
Glycine betaine	N. Pacific (37 N)	0.5 - 0.63	4.81 - 21.49	8.7 - 37.5	this study
Glycine betaine	Mobile Bay	0.39 - 44	1.8 - 49	0.52 - 11	Kiene and Williams 1998
Individual DFAA	Sargasso Sea (N. Atlantic)	0.003 - 0.2	0.1 - 5.5	1.6 - 181.9	Suttle et al 1991
Alanine	Sargasso Sea (N. Atlantic)	0.015 - 0.15	0.9 - 1.8	1.6 - 13.2	Suttle et al 1991
Ornithine	Sargasso Sea (N. Atlantic)			48.7 - 181.9	Suttle et al 1991
Taurine	Off the Iberian Peninsula			6 - 240	Clifford et al 2019
Leucine	Off the Iberian Peninsula			21 - 6000	Clifford et al 2019

* Ranges represent the range of the values in the publication or the best estimate plus and minus one standard deviation, the values for this study include the full range of estimates from the linear transformation and non-linear least squares fit.

Table 5.3: Uptake kinetic parameters for various metabolites in the marine environment

5.8 Figures

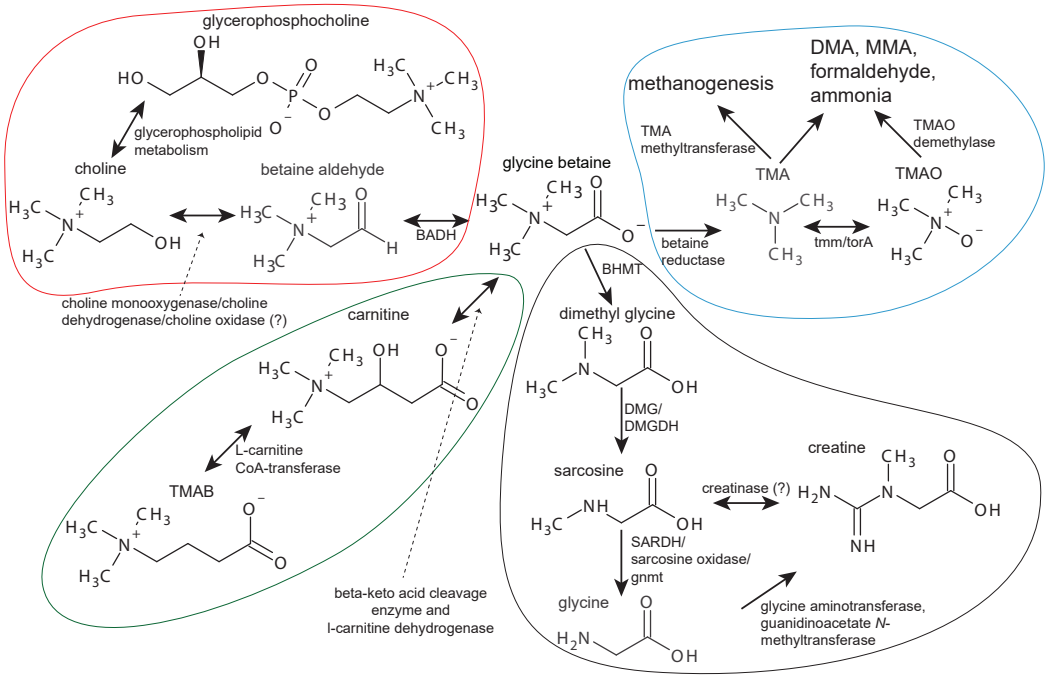

Figure 5.1: Several pathways of glycine betaine transformation. Pathways are circled based on the initial transformation of glycine betaine to dimethyl glycine and sarcosine via demethylation (black), to TMA via glycine betaine reductase (blue), to lipids via betaine aldehyde and choline (red), to carnitine through the intermediates GBT-coA (not shown) and 3-dehydrocarnitine (not shown) (green). Note that TMAO can be formed from choline and carnitine via choline TMA-lyase (cutC) and carnitine oxygenase (cntA). Molecules we did not measure are in grey. TMA = trimethylamine, TMAO = trimethylamine-n-oxide, TMAB = 4-trimethylammoniobutanoate. DMA = dimethylamine, MMA = monomethylamine. Specific reactions are labeled with enzyme names where applicable. bhmt = betaine-homocysteine S-methyltransferase, dmg = dimethylglycine oxidase, dmgdh = dimethylglycine dehydrogenase, SARDH = sarcosine dehydrogenase, gnmt = glycine N-methyltransferase, tmm = TMA monooxygenase (the beginning of the aerobic TMA oxidation pathway), torA = TMAO reductase (reduction of TMAO to TMA), BADH = betaine-aldehyde dehydrogenase (known to produce betaine from betaine aldehyde), ck = choline kinase. Incorporation into lipids occurs via choline kinase generation of phosphocholine, phosphocholine cytidylyltransferase, and cholinephosphotransferase. The arrow from TMA to DMA and MMA represent the anaerobic TMA dehydrogenase pathway. The question marks indicate that the directionality of these enzymes reported in the literature does not match with the metabolomics data in our study. Creatinase is known to make sarcosine from creatine and choline monooxygenase/choline dehydrogenase/choline oxidase are known to produce betaine aldehyde from choline.

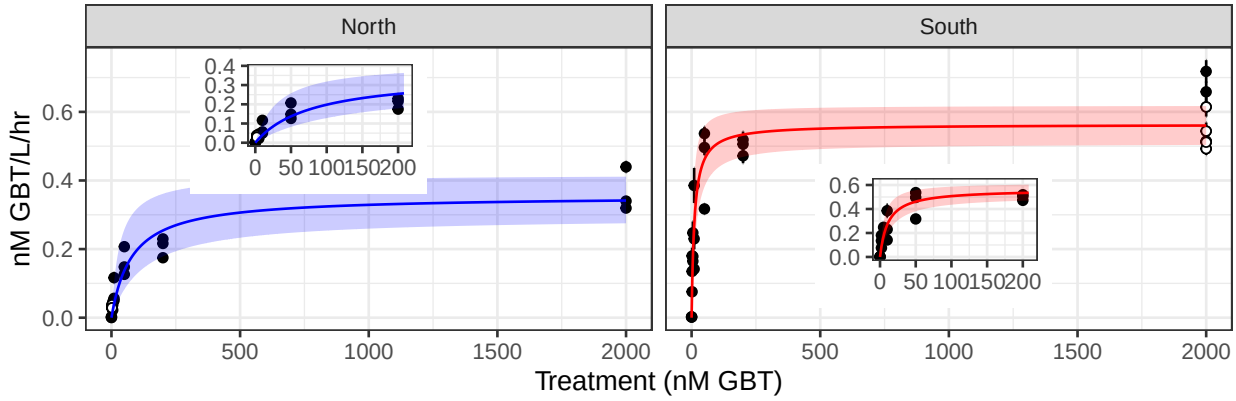

Figure 5.2: Uptake of $^{13}C_5$, ^{15}N-GBT in natural seawater. Black points are calculated rates from individual bottles, standard error propagated from the uncertainty associated with quantification is shown as the bars, sometimes obscured by the size of the points. Open circles are replicate injections of the same sample. Uptake kinetics curve fit to the data with a nonlinear least squares and monte carlo randomization as described in the methods is shown in the solid line, with one standard deviation of K_s and V_{max} shown in shading. Inset shows the shape of the curve for the 0-200 nM concentration range.

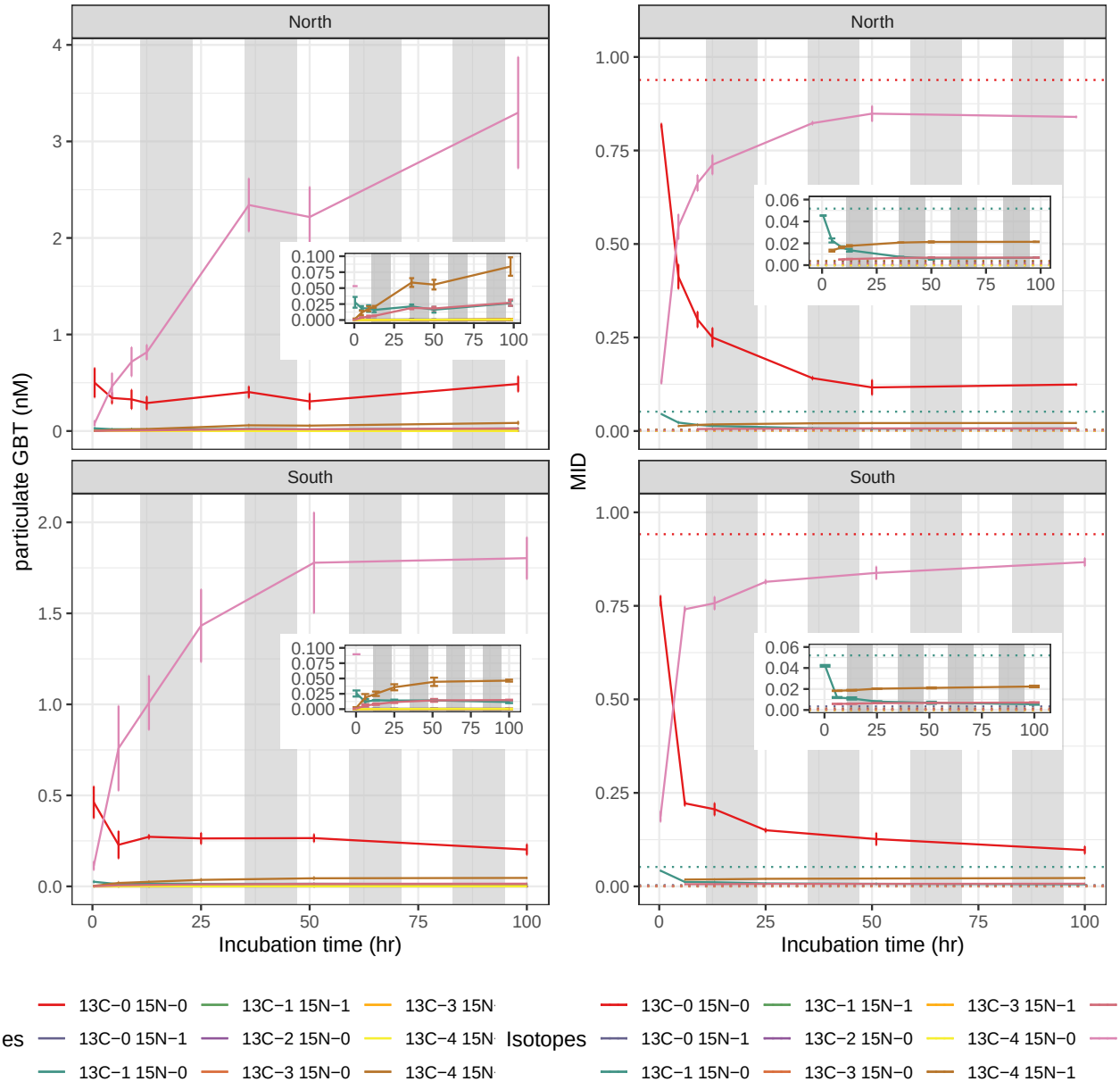

Figure 5.3: The particulate concentration of all possible glycine betaine isotopologues over the course of the incubations (Left). The mass isotopologue distribution over the course of the incubations (Right). Lines are the average of the triplicate bottles, with error bars representing one standard deviation. Grey boxes indicate night time. Dashed lines on the right plots indicate the isotopic abundance of samples taken from the stations where each experiment took place. Insets have an adjusted y-axis to show the lower abundance isotopologues.

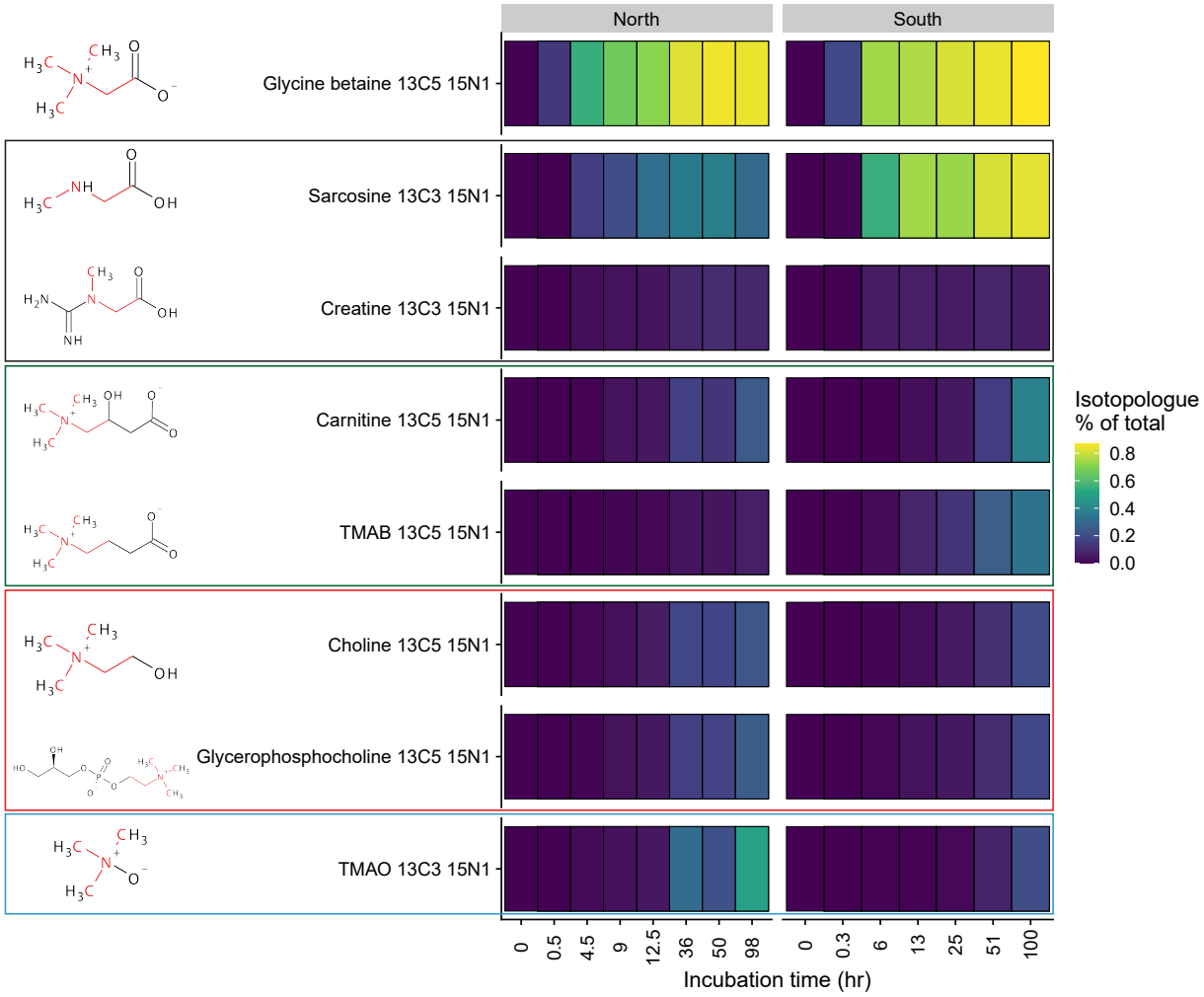

Figure 5.4: Heatmaps of isotope label incorporation for select isotopologues of metabolites. The color indicates the percent of the total metabolite concentration that is in the isotopologue listed, with the likely atoms that are isotopically heavy indicated in red in the drawing on the left, any carbon atom with a red bond connected to it is ^{13}C. Note that the incubation times are not the same for the two experiments. The 0 hour incubation time point is from a water sample from the stations where each experiment was performed.

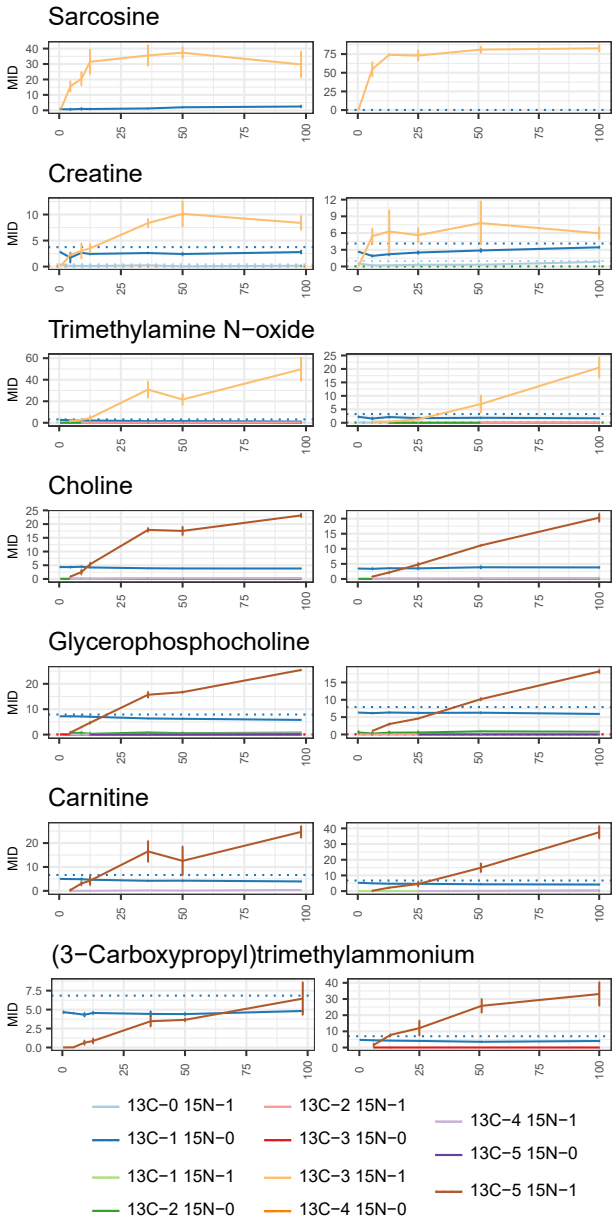

Figure 5.5: Mass isotopologue distributions (MID) over time for various isotopologues detected in the metabolites that seem to be derived directly from glycine betaine. x-axis is incubation time in hours. The northern experiment is on the left and the southern experiment is on the right. Solid lines indicate the mean of triplicate bottles and the errorbars represent one standard deviation. Horizontal dashed lines indicate the MID for samples taken at the stations where each experiment was performed.

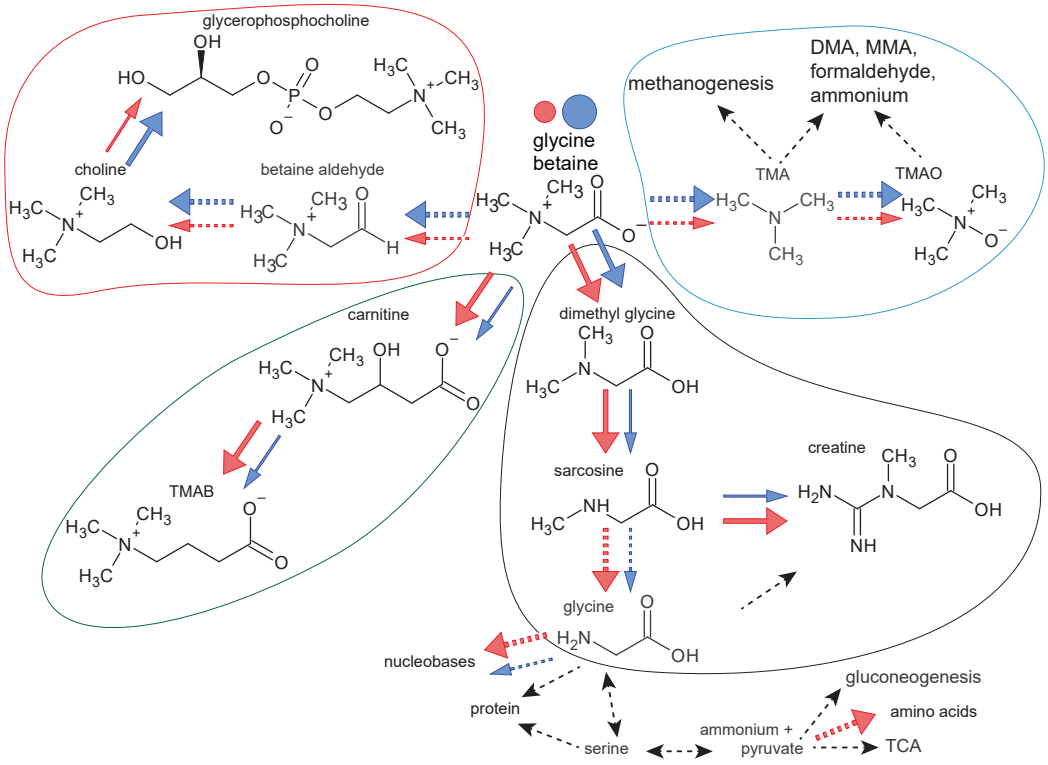

Figure 5.6: Pathways of glycine betaine metabolism observed in the northern and southern stations of the NPTZ. See Figure 5.1 for details of pathway, molecule, and enzyme descriptions. Colored circles represent the total uptake of labeled dissolved GBT over the course of the incubations, with a larger blue circle representing more uptake in the north and a smaller red circle indicating less uptake in the south. Colored arrows represent the observed pathways of GBT metabolism in the north (blue) and south (red) stations, with the relative arrow sizes qualitatively representing the fluxes through these pathways. Larger arrows indicate greater flux. Dashed black arrows indicate that we did not directly observe that pathway but that it is still likely active in some organisms. Dashed colored arrows indicated that we did not observe the direct metabolite reactant or product of that enzymatic reaction, but that downstream metabolites accumulated isotopes, indicating that the reaction was taking place.

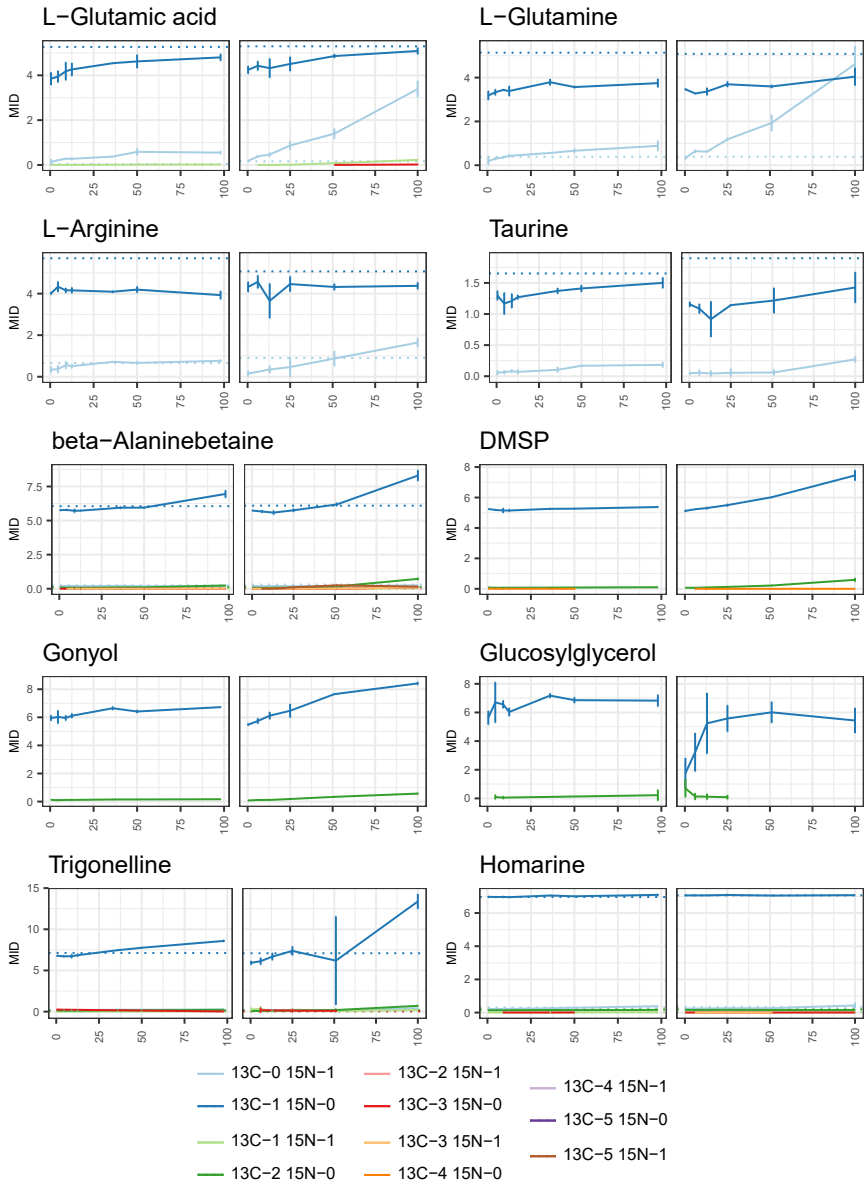

Figure 5.7: Mass isotopologue distributions (MID) over time for various isotopologues detected in the metabolites that show evidence of C and N recycling through cellular metabolism. x-axis is incubation time in hours. The northern experiment is on the left and the southern experiment is on the right. Solid lines indicate the mean of triplicate bottles and the errorbars represent one standard deviation. Horizontal dashed lines indicate the MID for samples taken at the stations where each experiment was performed.

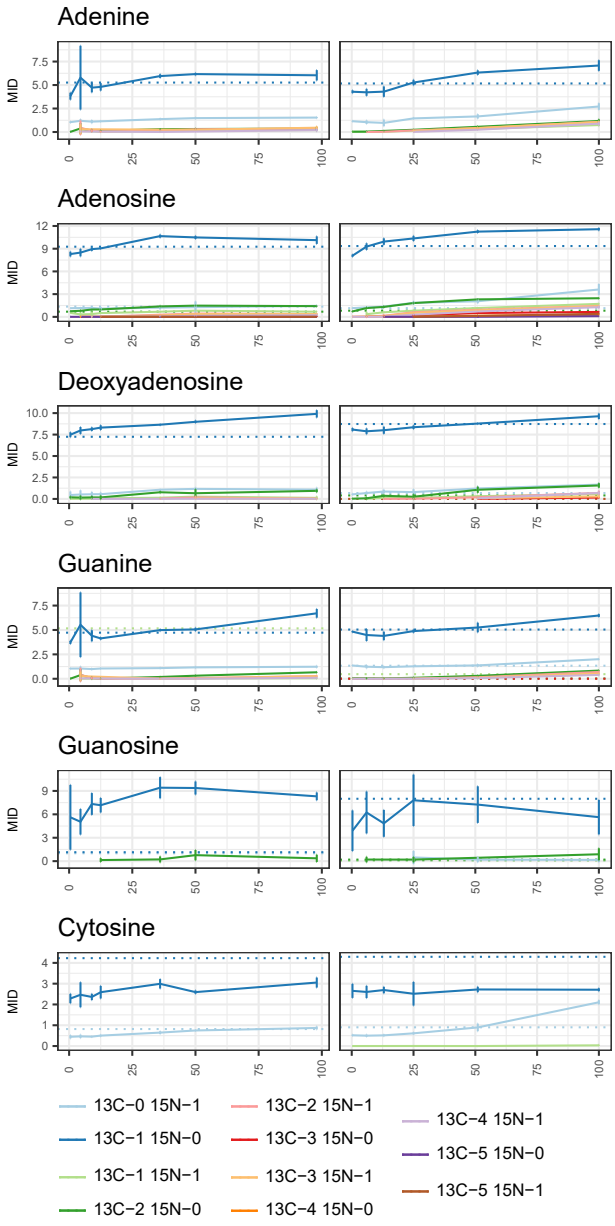

Figure 5.8: Mass isotopologue distributions (MID) over time for various isotopologues of nucleobases and nucleosides detected in the metabolites that show evidence of C and N cycling through cellular metabolism. x-axis is incubation time in hours. The northern experiment is on the left and the southern experiment is on the right. Solid lines indicate the mean of triplicate bottles and the errorbars represent one standard deviation. Horizontal dashed lines indicate the MID for samples taken at the stations where each experiment was performed.

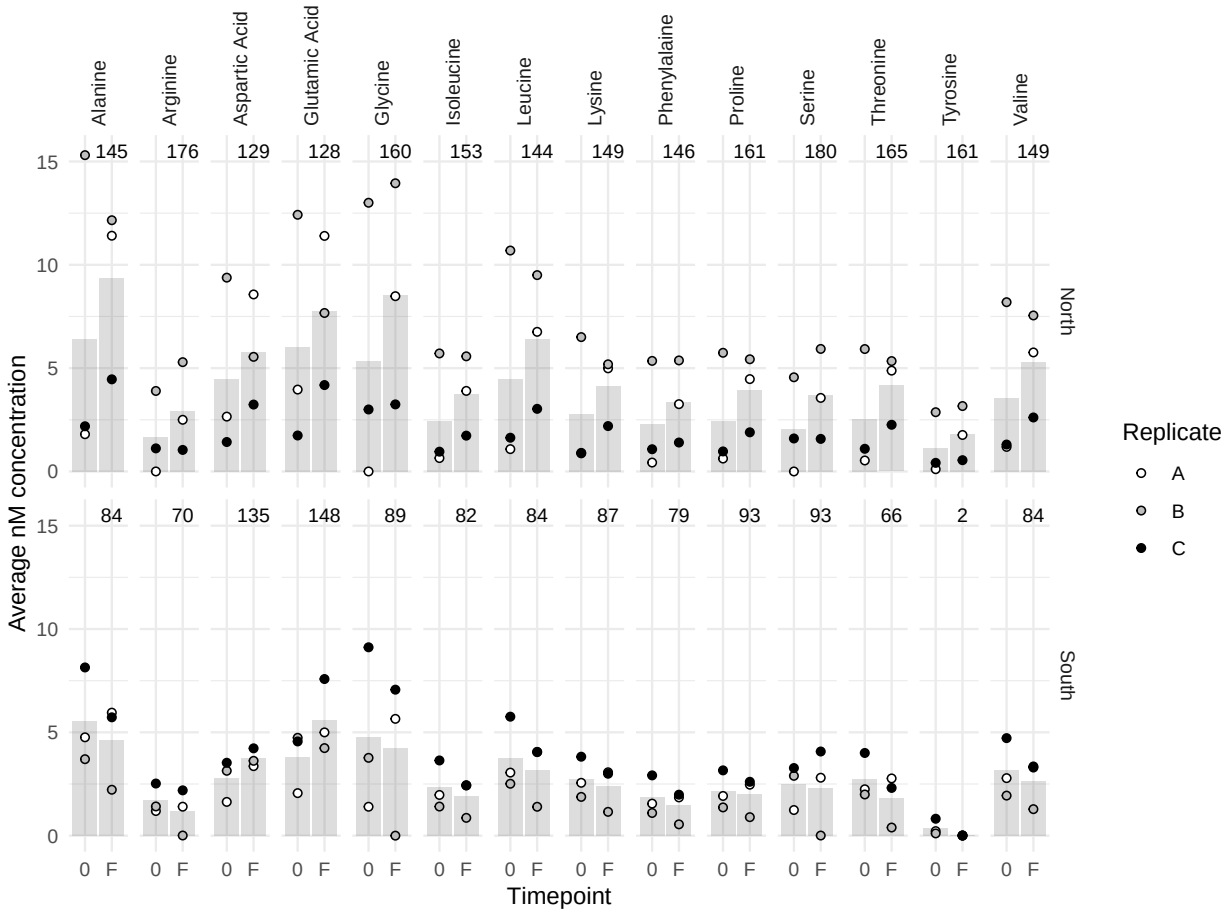

Figure 5.9: The concentration of total hydrolyzable amino acids at the initial (0) and final (F, 98 hours) timepoints in the north (top) and south (bottom) GBT fate experiments. Bars are the average value of three replicates, individual points are the individual replicate bottles. Note that the replicate bottles were not the same for the two timepoint, but are the same across amino acids for a given timepoint. The number at the top of each plot indicates the average percent change between the initial and final timepoint. Methionine and histidine were only detected in a few replicates and were therefore excluded.

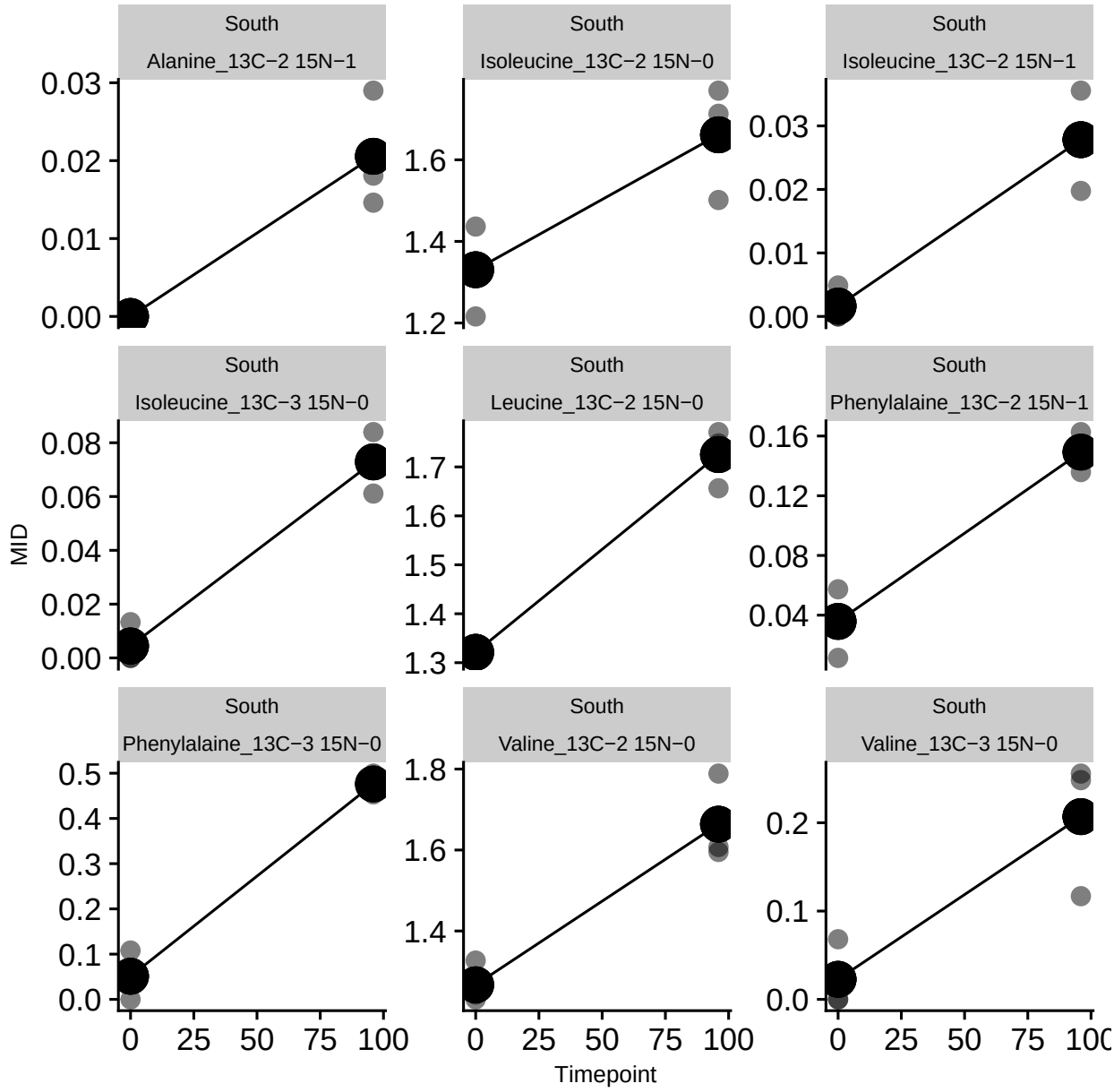

Figure 5.10: The isotopologues that had increases in their relative contribution to the total THAA pool for that amino acid ($p < 0.05$). x-axis is incubation time in hours. Replicate samples are in grey and the mean value is in black.

5.9 Supplement

5.9.1 Supplemental calculations

Overall uptake rate of GBT in first few hours

The calculation below estimates a loss of particulate GBT, which can be interpreted as a flux of GBT into metabolic pathways plus the flux of GBT out into the extracellular environment.

We calculate that the observed accumulation of any isotopologue of any compound, i, is equal to the combination of its sources and sinks:

$$Accumulation_i = uptake_i + synthesis_i - loss_i \tag{5.3}$$

For $^{13}C_0$, $^{15}N_0$-GBT we assume that synthesis went to zero directly after the addition of 500 nM extracellular GBT, and the uptake of unlabeled GBT is negligible compared to that of the labeled GBT, so accumulation is equal to loss. For $^{13}C_5$, $^{15}N_1$-GBT we assume that synthesis is equal to zero. We also assume that loss processes happen without isotopic discrimination, so the loss of the unlabeled GBT should be equal to the loss of the labeled GBT, scaled by the proportional concentration of those isotopologues. Thus we can calculate the total uptake of $^{13}C_5$, $^{15}N_1$-GBT as

$$uptake_{^{13}C_5,^{15}N_1} = accumulation_{^{13}C_5,^{15}N_1} + (loss_{^{13}C_0,^{15}N_0}/R) \tag{5.4}$$

Where R is the average ratio of intracellular $^{13}C_0$,$^{15}N_0$-GBT to $^{13}C_5$, $^{15}N_1$-GBT for the first two timepoints, 0.76 and 1.84 for the northern and southern experiments, respectively. The overall accumulation rate of $^{13}C_5$, $^{15}N_1$-GBT over the first two time points was 0.1 and 0.12 nM/h in the north and south, respectively. And the loss of $^{13}C_0$,$^{15}N_0$-GBT over the first two time points was 0.037 and 0.041 nM/h in the north and south, respectively. The total uptake of $^{13}C_5$, $^{15}N_1$-GBT in the experiments were 0.13 and 0.19 nM/h in the north and south, respectively.

Estimating bacterial production fueled by GBT

Though we were not able to measure the respiration of GBT in this study, at coastal and estuarine sites 31-55% of total GBT taken up into cells was respired within 24 hours [46]. Additionally, after 24 hours, approximately two thirds of unrespired GBT had been transformed into other metabolites or macromolecules [46]. These estimates were based on tracing a ^{14}C-methyl group attached to GBT into cells and CO_2. If we assume that these transformations were happening in our experiments at the same rate for all C atoms in the molecule, then the amount of glycine betaine C respired over the first 50 hours of the northern experiment would be between 15-32 nM C and the amount of carbon from glycine betaine in other molecules, including macromolecules, after the first 50 hours would be 22 nM carbon. For the southern experiment these numbers would be approximately 80% of these values, proportional to the difference in untransformed GBT in the two experiments (Figure 5.3). Based off of these estimations, in the north,a total of 33 nM particulate carbon was in or derived from GBT. In comparison, the total ^{13}C we could quantify in our metabolites in the northern experiment after two days was 13.4 $\pm$ 1.6 nM C. This is almost exactly a third of the estimate above because most of the ^{13}C we measured was in untransformed GBT (Figure 5.12). In the north we were unable to detect any enrichment in total hydrolyzable macromolecules.

Assuming the estimates calculated above are correct and that the total particulate carbon remained similar to the value *in situ* (4.24 μM, Table 5.1), then 0.8% of total PC would be GBT or derived from GBT after 50 hours. The light respiration rate in the north was 0.84 mmol O_2 m^{-3} d^{-1} (personal communication, S. Ferron), so the C needed to fuel two days of respiration would be 1.68 μM C. The respired GBT carbon would fuel 0.89-1.82% of total respiration. These estimates are based on assumptions with faults. The radiocarbon labeled methyl group has different cellular fates than the carbons in the carbon chain, since they are separated during the demethylation pathway and the betaine reductase pathway (in which the two atoms in the carbon chain form acetyl phosphate while the methyl groups are part

of TMA). However, these estimates still provide some ballpark values for the quantitative importance of GBT in the microbial community in the North Pacific.

5.9.2 Supplemental Tables

Experiment	n.params.fit	Kt or Kt+S	Vmax	S
N	2	79.46 (48.26)	0.36 (0.06)	-
N	3	122.15 (75.31)	0.38 (0.07)	9.09 (7.29)
S	2	11.06 (5.53)	0.56 (0.06)	-
S	3	12.44 (7.44)	0.57 (0.05)	0.73 (1.29)

Table 5.4: The K_t (3 parameters) or $K_t + S$ (two parameters), V_{max}, and S values when two versus three parameters were fit, reported as a mean and standard deviation as calculated from the monte carlo error analysis described in the methods.

Table 5.5: Table of targeted metabolites measured in the glycine betaine isotope tracing experiments. "Isotopologue significant change" column is true if in either the north or the south experiment, any isotopologue of that metabolite showed a significant trend in mass isotopologue distribution with time. "Isotopologue significant change" column is false if we measured the metabolite but did not detect any ^{13}C or ^{15}N isotopologues, or if we did detect the isotopologues but they did not show any significant enrichment over time ($p > 0.05$). The details of which isotopolgoues were had significant MID changes in which experiment are in Table 5.6. (XLSX)

Table 5.6: Each isotopologue that was detected in > 2 samples in a given experiment was fit with a linear model to determine if there was a significant trend with time in the MID. The p-value and significance (True if $p < 0.05$) results are presented here for each isotopologue tested in each experiment. The ^{13}C-0 ^{15}N-0 isotopologues that have a significant trend had MID decreases over time, while the isotopologues with ^{13}C or ^{15}N that were significant typically had MID increases over time. The average and standard deviation of the MID for each isotopologue in each sample are in Table 5.7. (XLSX)

Table 5.7: Table of the estimated concentrations (nM) and isotopologue distributions. Fate1 is the north experiment, and Fate2 is the south experiment. The timepoint is the approximate sampling time, Incubation time (hr) is the actual sampling time after initial spiking. C.lab and N.lab are the number of ^{13}C and ^{15}N atoms in the isotopologue. iso.mz is the mass to charge ratio for the isotopologue. ionization is the ionization we measured that metabolite in. MID is mass isotopologue distribution, where the number 1 would indicate that in a particular sample 100% of the concentration of a given metabolite was in that specific isotopologue and the number 0 would indicate that in a particular sample, 0% of the concentration of a given metabolite was in that specific isotopologue. (XLSX)

5.9.3 Supplemental Figures

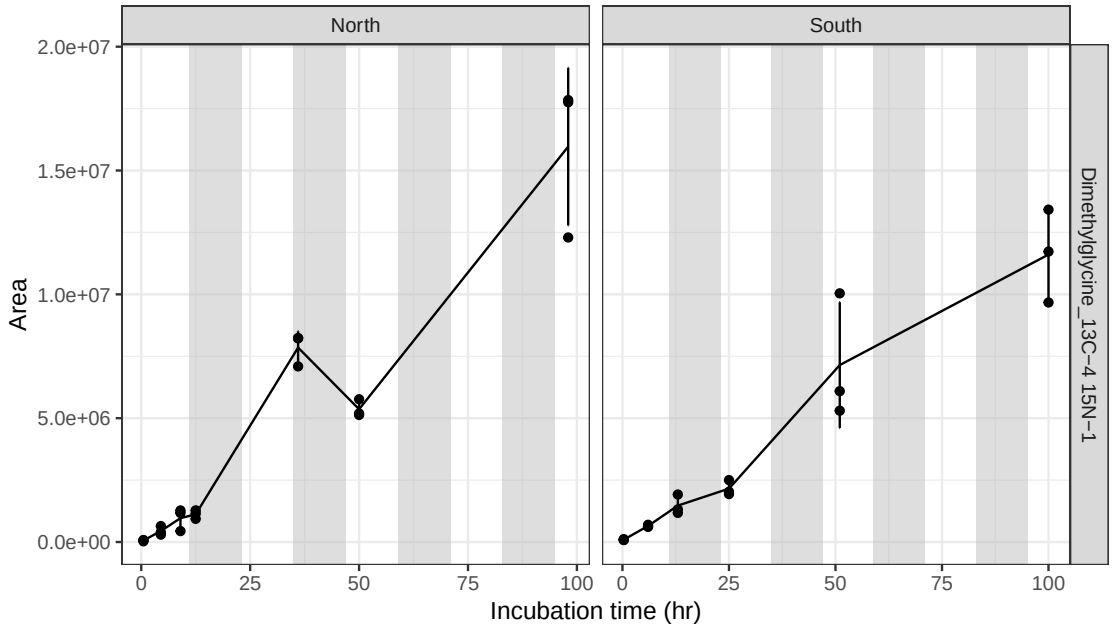

Figure 5.11: Peak area of $^{13}C_4$, $^{15}N_1$-dimethylglycine over the course of both experiments. This isotopologue of dimethylglycine was not detected in the *in situ* samples from the corresponding stations. Grey shading indicates night time.

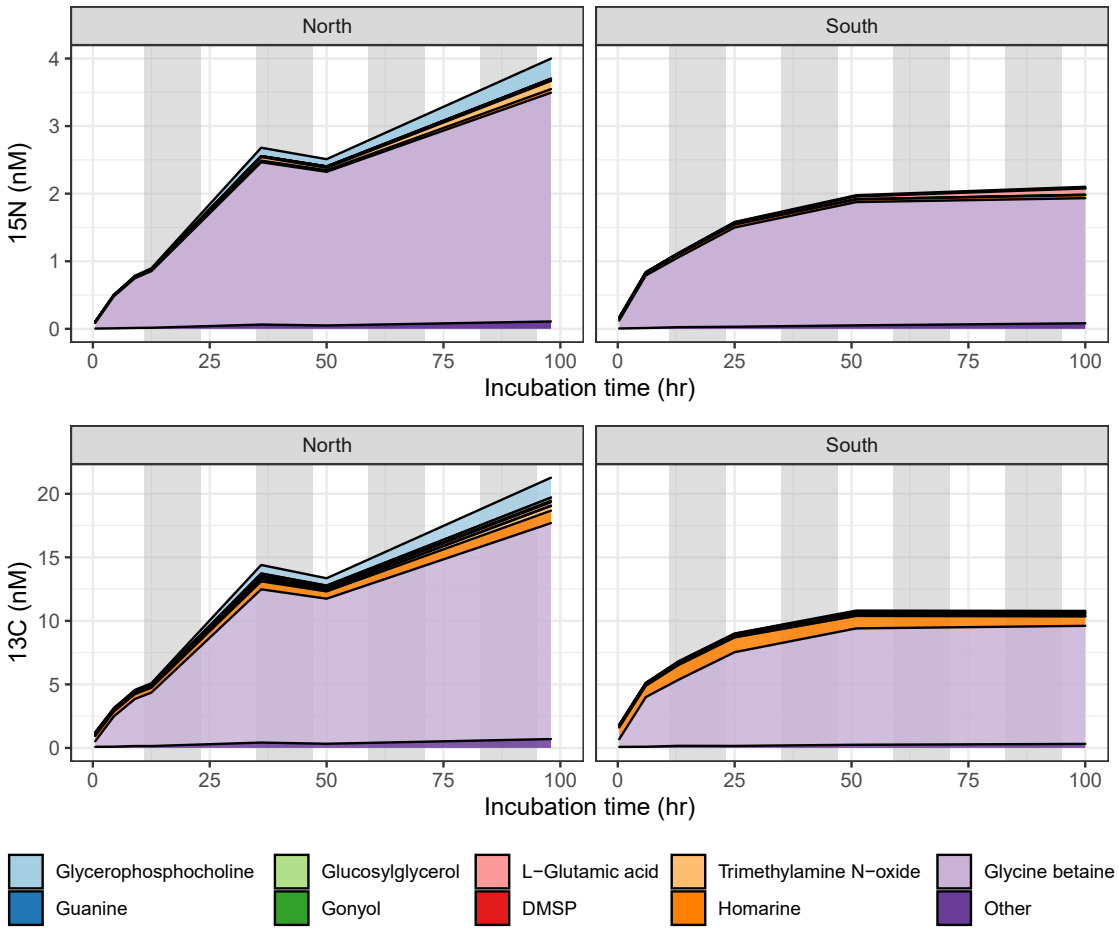

Figure 5.12: The total particulate ^{15}N (top) and ^{13}C (bottom) isotope concentrations in the north (left) and south (right) incubation experiments in known, quantifiable metabolites. Grey shading indicates night time.

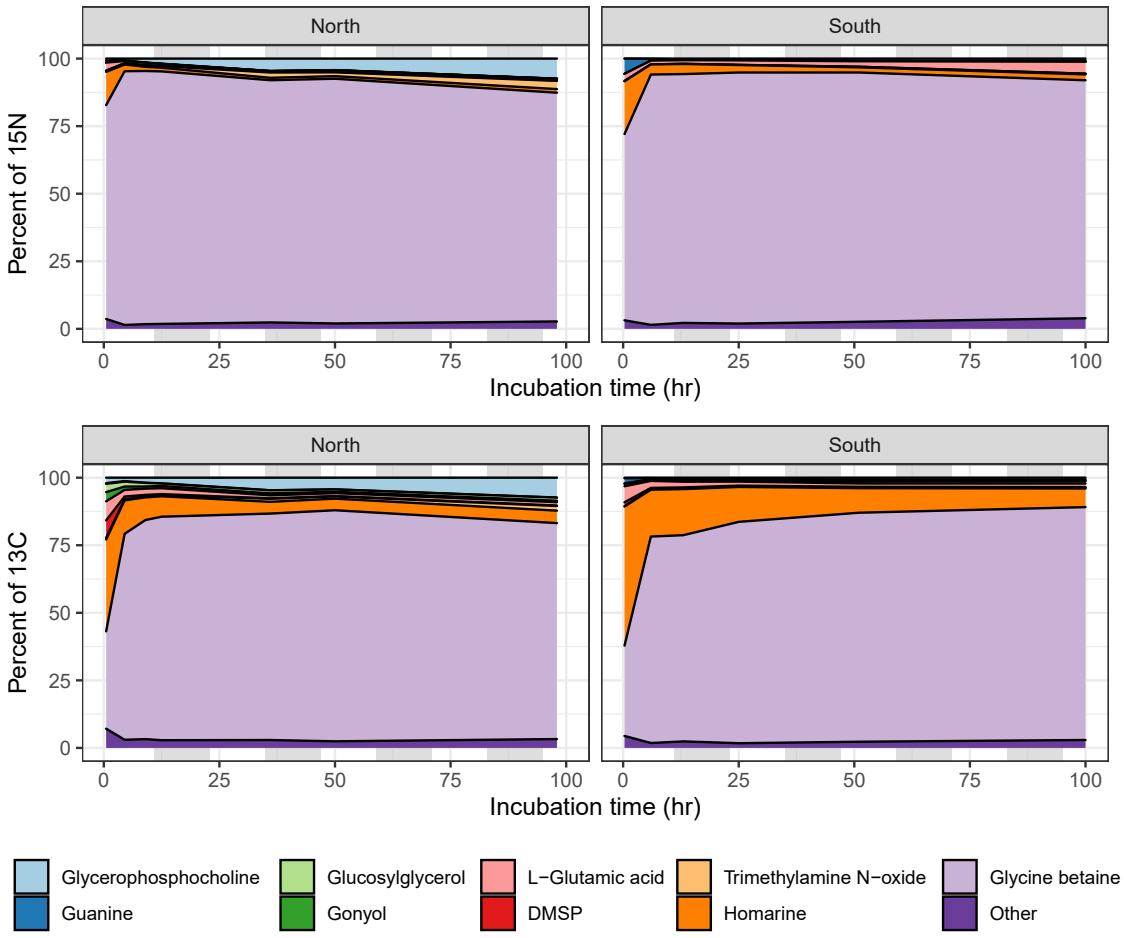

Figure 5.13: The total particulate ^{15}N (top) and ^{13}C (bottom) isotope concentrations in the north (left) and south (right) incubation experiments, relative to the total measured in known, quantifiable metabolites. Grey shading indicates night time.

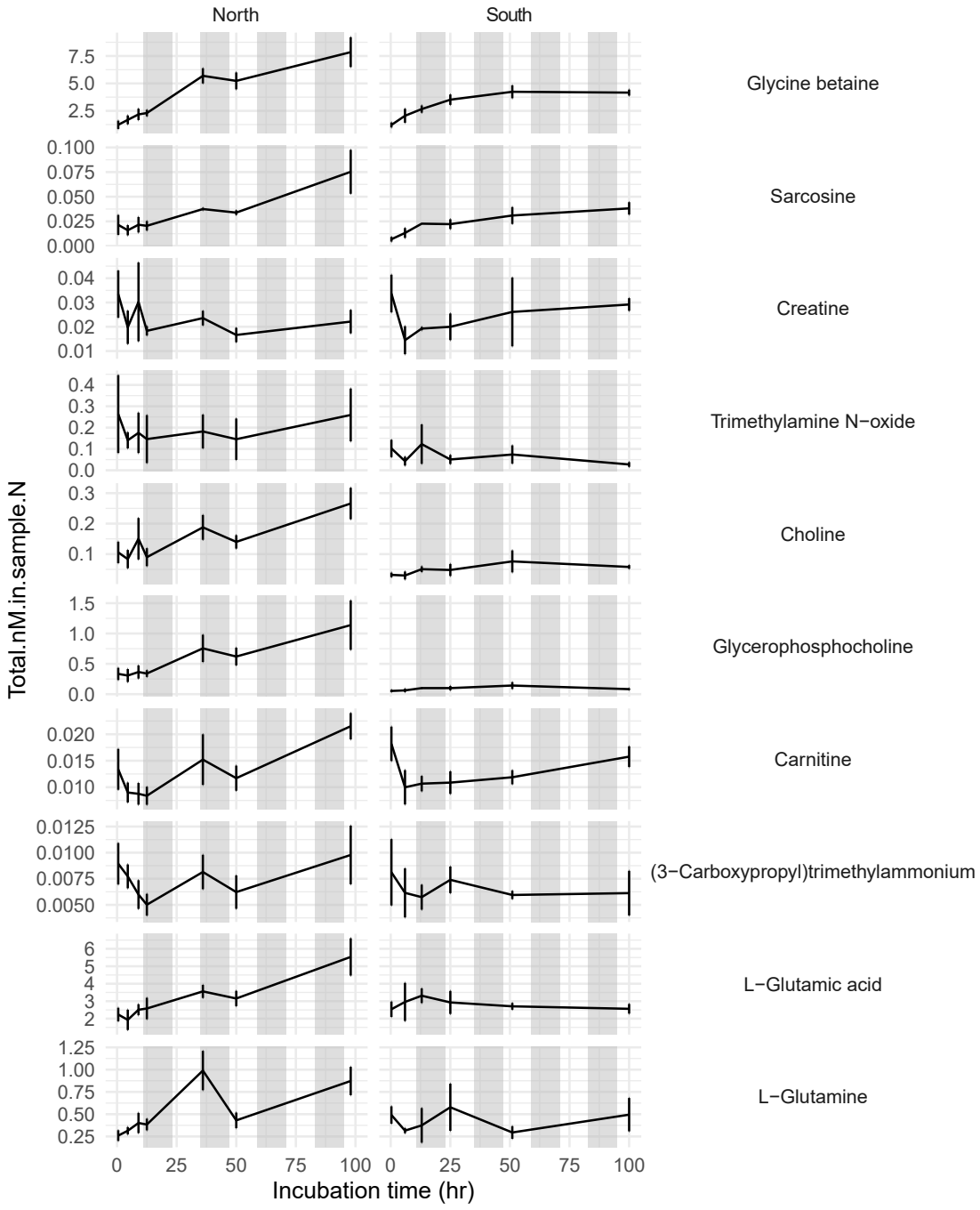

Figure 5.14: The concentration of the sum of all of the isotopologues of metabolites over time in the fate incubation experiments. The units of concentration are nM of nitrogen. The line and error bars are the mean and standard deviation of triplicates. Grey shading indicates night time.

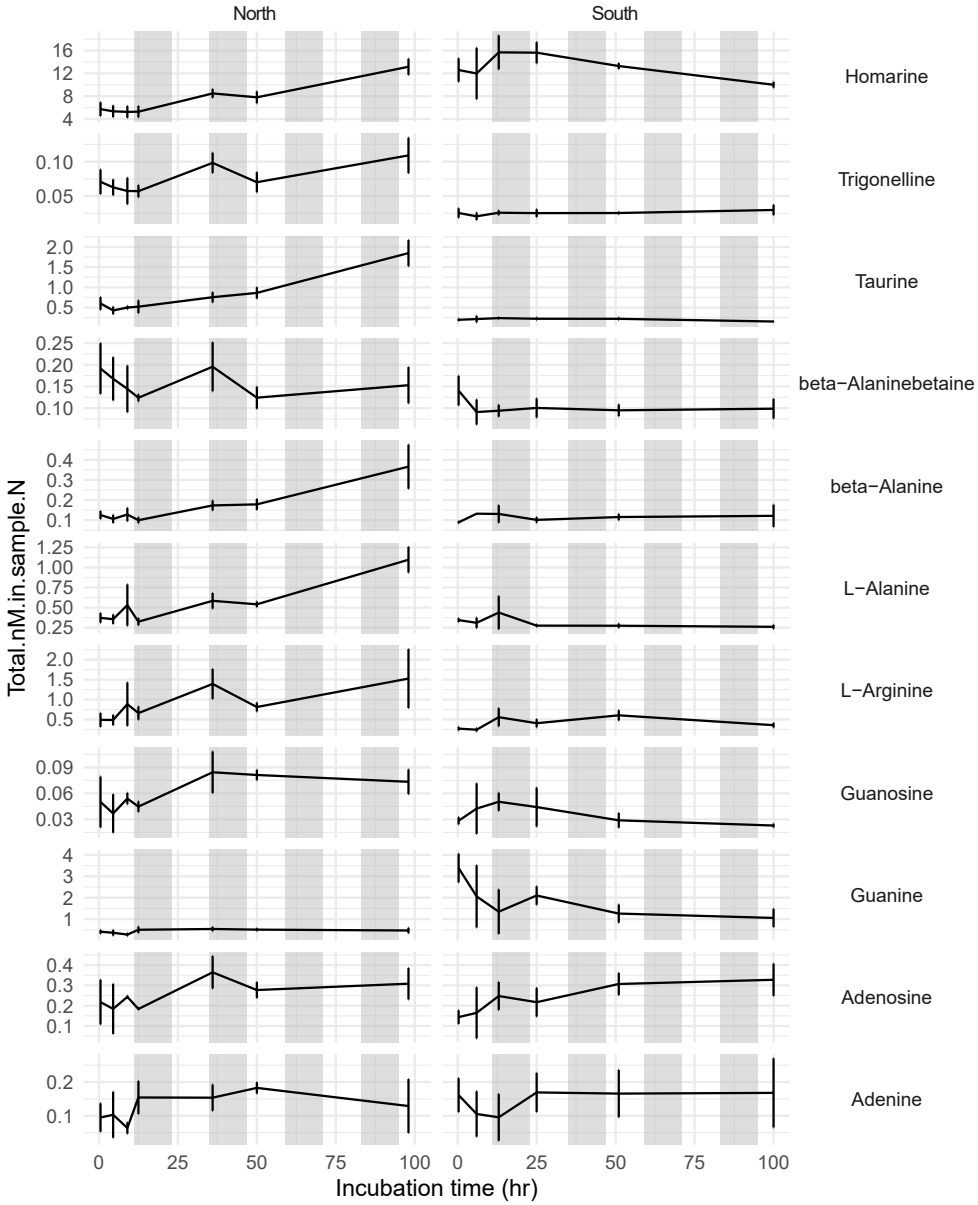

Figure 5.15: The concentration of the sum of all of the isotopologues of metabolites over time in the fate incubation experiments. The units of concentration are nM of nitrogen. The line and error bars are the mean and standard deviation of triplicates. Grey shading indicates night time. Metabolites included are adenine, adenosine, guanine, guanosine, L-arginine, L-alanine, taurine, trigonelline, and homarine.

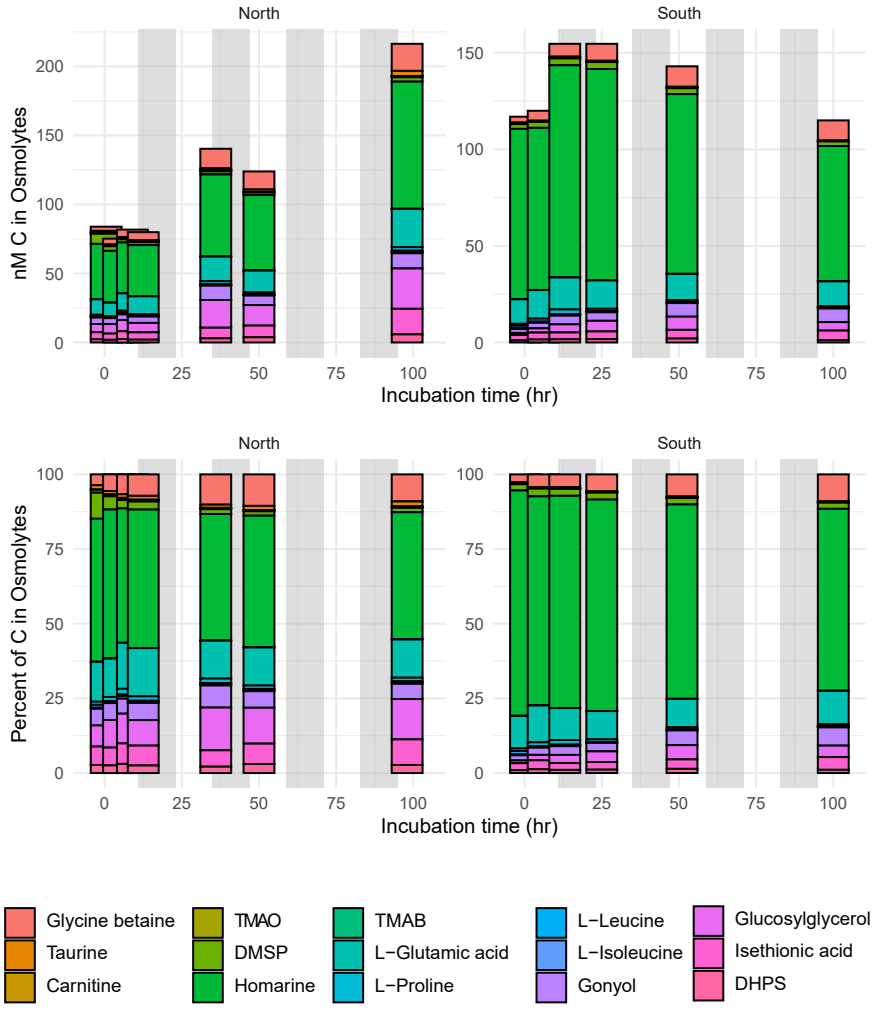

Figure 5.16: The concentration of quantified osmolytes over the course of the experiments, in nM C (top) and as a percent of the total C measured in these osmolytes (bottom).

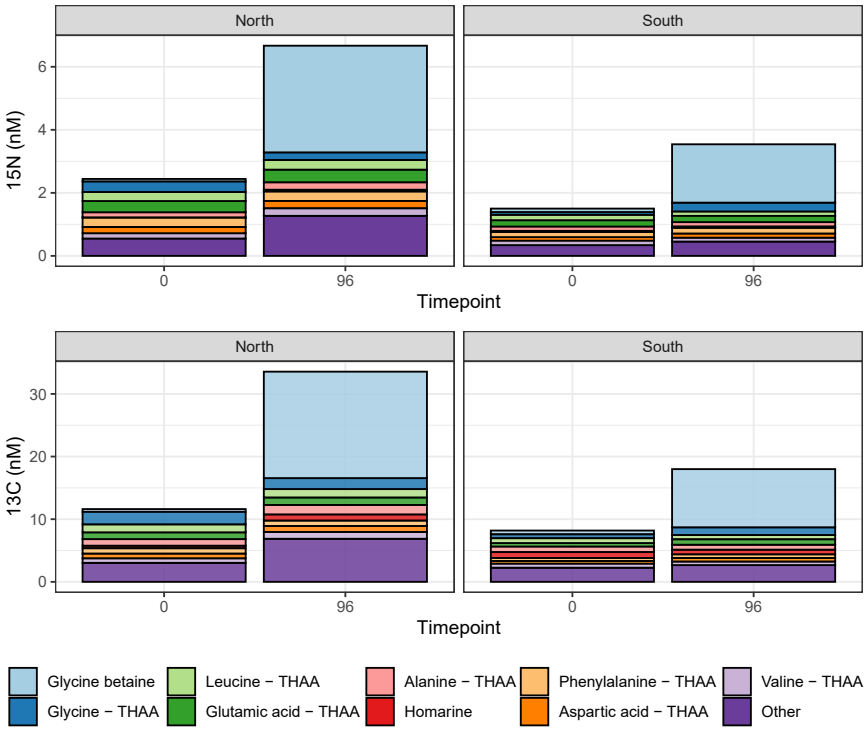

Figure 5.17: Total 15N and 13C quantified in metabolites and total hydrolyzable amino acids at the initial and final time points of each experiment.

BIBLIOGRAPHY

[1] Christopher L Sabine, Richard A Feely, Nicolas Gruber, Robert M Key, Kitack Lee, John L Bullister, Rik Wanninkhof, CSL Wong, Douglas WR Wallace, Bronte Tilbrook, et al. The oceanic sink for anthropogenic co2. *science*, 305(5682):367–371, 2004.

[2] Philippe Ciais, Christopher Sabine, G Bala, L Bopp, V Brovkin, J Canadell, A Chhabra, R DeFries, J Galloway, M Heimann, et al. Climate change 2013: the physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change. *K., Tignor, M., Allen, SK, Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, PM, Eds*, 2013.

[3] Steven Emerson and John Hedges. *Chemical oceanography and the marine carbon cycle*. Cambridge University Press, 2008.

[4] Yinon M. Bar-On, Rob Phillips, and Ron Milo. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences of the United States of America*, 115(25):6506–6511, 2018.

[5] Farooq Azam, Tom Fenchel, J G_ Field, JS Gray, LA Meyer-Reil, and F Thingstad. The ecological role of water-column microbes in the sea. *Marine ecology progress series*, pages 257–263, 1983.

[6] John K. Volkman and Eiichiro Tanoue. Chemical and biological studies of particulate organic matter in the ocean. *Journal of Oceanography*, 58(2):265–279, 2002.

[7] Christina L. De La Rocha and Uta Passow. Factors influencing the sinking of POC and the efficiency of the biological carbon pump. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 54(5-7):639–658, 2007.

[8] Jacob G Bundy, Matthew P Davey, and Mark R Viant. Environmental metabolomics: a critical review and future perspectives. *Metabolomics*, 5(1):3–21, 2009.

[9] Monya Baker. Metabolomics: from small molecules to big ideas. *Nature Publishing Group*, 8(2), 2011.

[10] Mary Ann Moran. The global ocean microbiome. *Science*, 350(6266):aac8455, 2015.

[11] Shady a. Amin, L. R. Hmelo, H M van Tol, Bryndan P. Durham, L. T. Carlson, K. R. Heal, R. L. Morales, Chris T. Berthiaume, M. S. Parker, B. Djunaedi, Anitra E. Ingalls, M. R. Parsek, M. a. Moran, E. Virginia Armbrust, H. M. van Tol, Bryndan P. Durham, L. T. Carlson, K. R. Heal, R. L. Morales, Chris T. Berthiaume, M. S. Parker, B. Djunaedi, Anitra E. Ingalls, M. R. Parsek, M. a. Moran, and E. Virginia Armbrust. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature*, 522(7554):98–101, 2015.

[12] Bryndan P Durham, Shalabh Sharma, Haiwei Luo, Christa B Smith, Shady A Amin, Sara J Bender, Stephen P Dearth, Benjamin AS Van Mooy, Shawn R Campagna, Elizabeth B Kujawinski, E. Virginia Armbrust, and Mary Ann Moran. Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc. Natl. Acad. Sci. U.S.A.*, 112(2):453–457, 2015.

[13] Laura R Hmelo, Tracy J Mincer, and Benjamin a S Van Mooy. Possible influence of bacterial quorum sensing on the hydrolysis of sinking particulate organic carbon in marine environments. *Environmental microbiology reports*, 3(6):682–8, 2011.

[14] Ruth L Airs and Steven D Archer. Analysis of glycine betaine and choline in seawater particulates by liquid chromatography/electrospray ionization/mass spectrometry. *Limnology and Oceanography: Methods*, 8(10):499–506, 2010.

[15] Christos Panagiotopoulos and Richard Sempéré. Analytical methods for the determination of sugars in marine samples: A historical perspective and future directions. *Limnology and Oceanography: Methods*, 3(10):419–454, 2005.

[16] Kevin J Flynn, Kenneth J Jones, Robin Raine, Jocelyn Richard, and Krystyna Flynn. Use of intracellular amino acid analysis as an indicator of the physiological status of natural dinoflagellate populations. *Marine Ecology Progress Series*, pages 175–186, 1994.

[17] Ulf Karsten, Christian Wiencke, and GO Kirst. The ß-dimethylsulphoniopropionate (dmsp) content of macroalgae from antarctica and southern chile. *Botanica Marina*, 33:143–146, 1990.

[18] Cindy Lee and Jeffrey L. Bada. Dissolved amino acids in the equatorial Pacific, the Sargasso Sea, and Biscayne Bay. *Limnology and Oceanography*, 22(3):502–510, 1977.

[19] CA Suttle, AM Chan, and JA Fuhrman. Dissolved free amino acids in the Sargasso Sea: uptake and respiration rates, turnover times, and concentrations. *Marine Ecology Progress Series*, 70:189–199, 1991.

[20] Katja Dettmer and Bruce D Hammock. Metabolomics–a new exciting field within the "omics" sciences. *Environmental health perspectives*, 112(7):A396–7, 2004.

[21] Hikmat Ghosson, Yann Guitton, Amani Ben Jrad, Chandrashekhar Patil, Delphine Raviglione, Marie-Virginie Salvia, and Cédric Bertrand. Electrospray Ionization and heterogeneous matrix effects in Liquid Chromatography-Mass Spectrometry-based meta-metabolomics: a biomarker or a suppressed ion? *Rapid Communications in Mass Spectrometry*, 2020.

[22] A.K. Boysen, K.R. Heal, L.T. Carlson, and A.E. Ingalls. Best-Matched Internal Standard Normalization in Liquid Chromatography-Mass Spectrometry Metabolomics Applied to Environmental Samples. *Analytical Chemistry*, 90(2), 2018.

[23] E. a. Ottesen, C. R. Young, S. M. Gifford, J. M. Eppley, R. Marin, S. C. Schuster, C. A. Scholin, and E. F. DeLong. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science*, 345(6193):207–212, 2014.

[24] Frank O. Aylward, John M. Eppley, Jason M. Smith, Francisco P. Chavez, Christopher a. Scholin, and Edward F. DeLong. Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proceedings of the National Academy of Sciences*, 112(17):5443–5448, 2015.

[25] Rachel S Poretsky, Ian Hewson, Shulei Sun, Andrew E Allen, Jonathan P Zehr, and Mary Ann Moran. Comparative day/night metatranscriptomic analysis of microbial communities in the north pacific subtropical gyre. *Environmental microbiology*, 11(6):1358–1375, 2009.

[26] Samuel T. Wilson, Frank O. Aylward, Francois Ribalet, Benedetto Barone, John R. Casey, Paige E. Connell, John M. Eppley, Sara Ferron, Jessica N. Fitzsimmons, Christopher T. Hayes, Anna E. Romano, Kendra A. Turk-Kubo, Alice Vislova, E. Virginia Armbrust, David A. Caron, Matthew J. Church, Jonathan P. Zehr, David M. Karl, and Edward F. De Long. Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium Crocosphaera. *Nature Microbiology*, 2(July):1–20, 2017.

[27] Francois Ribalet, Jarred Swalwell, Sophie Clayton, Valeria Jiménez, Sebastian Sudek, Yajuan Lin, Zackary I Johnson, Alexandra Z Worden, and E Virginia Armbrust. Light-driven synchrony of Prochlorococcus growth and mortality in the subtropical Pacific gyre. *Proceedings of the National Academy of Sciences of the United States of America*, 122(26):8008–8012, 2015.

[28] Sara Ferrón, Samuel T Wilson, Sandra Martínez-garcía, Paul D Quay, and David M Karl. Metabolic balance in the mixed layer of the oligotrophic North Pacific Ocean from diel changes in O2/Ar saturation ratios. *Geophysical Research Letters*, 42(9):3421–3430, 2015.

[29] David M Karl and Matthew J Church. Microbial oceanography and the Hawaii Ocean Time-series programme. *Nature reviews. Microbiology*, 12(10):699–713, 2014.

[30] Angela K Boysen, Laura T Carlson, Bryndan P Durham, Ryan D Groussman, Frank O Aylward, François Ribalet, Katherine R Heal, Edward F DeLong, E. Virginia Armbrust, and Anitra E Ingalls. Diel oscillations of particulate metabolites reflect synchronized microbial activity in the North Pacific subtropical gyre. *bioRxiv*, page 2020.05.09.086173, 2020.

[31] Daniel Muratore, Angie K Boysen, Matthew J Harke, Kevin W Becker, John R Casey, Sacha N Coesel, Daniel R Mende, Samuel T Wilson, Frank O Aylward, John M Eppley, Alice Vislova, Shengyun Peng, Rogelio A Rodriguez-Gonzalez, Stephen J Beckett, E Virginia Armbrust, Edward F DeLong, David M Karl, Angelicque E White, Jonathan P Zehr, Benjamin A S Van Mooy, Sonya T Dyhrman, Anitra E Ingalls, and Joshua S Weitz. Community-scale Synchronization and Temporal Partitioning of Gene Expression, Metabolism, and Lipid Biosynthesis in Oligotrophic Ocean Surface Waters. *bioRxiv*, page 2020.05.15.098020, 2020.

[32] Bryndan P. Durham, Angela K. Boysen, Laura T. Carlson, Ryan D. Groussman, Katherine R. Heal, Kelsy R. Cain, Rhonda L. Morales, Sacha N. Coesel, Robert M. Morris, Anitra E. Ingalls, and E. Virginia Armbrust. Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nature Microbiology*, 4(10):1706–1715, 2019.

[33] C M Moore, 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 L 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. Processes and patterns of oceanic nutrient limitation. *Nature Publishing Group*, 6(9):701–710, 2013.

[34] Jeffrey J. Polovina, Evan A. Howell, Donald R. Kobayashi, and Michael P. Seki. The Transition Zone Chlorophyll Front updated: Advances from a decade of research. *Progress in Oceanography*, 150:79–85, 2017.

[35] Thomas J Browning, Eric P Achterberg, Insa Rapp, Anja Engel, Erin M Bertrand, Alessandro Tagliabue, and C Mark Moore. Nutrient co-limitation at the boundary of an oceanic gyre. *Nature*, 551(7679):242, 2017.

[36] Rachel S. Poretsky, Shulei Sun, Xiaozhen Mou, and Mary Ann Moran. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environmental Microbiology*, 12(3):616–627, 2010.

[37] Alexey Vorobev, Shalabh Sharma, Mengyun Yu, Juhyung Lee, Benjamin J. Washington, William B. Whitman, Ford Ballantyne, Patricia M. Medeiros, and Mary Ann Moran. Identifying labile DOM components in a coastal ocean through depleted bacterial transcripts and chemical signals. *Environmental Microbiology*, 20(8):3012–3030, 2018.

[38] Mary Ann Moran, Elizabeth B Kujawinski, Aron Stubbins, Rob Fatland, Lihini I Aluwihare, Alison Buchan, Byron C Crump, Pieter C Dorrestein, Sonya T Dyhrman, Nancy J Hess, Bill Howe, Krista Longnecker, Patricia M Medeiros, Jutta Niggemann, Ingrid Obernosterer, Daniel J Repeta, and Jacob R Waldbauer. Deciphering ocean carbon in a changing world. *Proceedings of the National Academy of Sciences of the United States of America*, 113(12):3143–51, 2016.

[39] Matthew J Wargo. Homeostasis and catabolism of choline and glycine betaine: lessons from Pseudomonas aeruginosa. *Applied and environmental microbiology*, 79(7):2112–20, 2013.

[40] Elizabeth B Kujawinski. The impact of microbial metabolism on marine dissolved organic matter. *Annual review of marine science*, 3:567–99, 2011.

[41] Katherine R Heal, Bryndan P Durham, Angela K Boysen, Laura T Carlson, Wei Qin, François Ribalet, , Angelique White, Randelle Bundy, E. Virginia Armbrust, and Anitra E Ingalls. Community metabolomes reflect taxon-specific fingerprints of phytoplankton in the ocean. *In Prep*, 2020.

[42] Anders Torstensson, Jodi N. Young, Laura T. Carlson, Anitra E. Ingalls, and Jody W. Deming. Use of exogenous glycine betaine and its precursor choline as osmoprotectants in Antarctic sea-ice diatoms1. *Journal of Phycology*, 55(3):663–675, 2019.

[43] David T. Welsh. Ecological significance of compatible solute accumulation by microorganisms: from single cells to global climate. *FEMS microbiology reviews*, 24(3):263–90, 2000.

[44] Xiaozhen Mou, Robert E. Hodson, and Mary Ann Moran. Bacterioplankton assemblages transforming dissolved organic compounds in coastal seawater. *Environmental Microbiology*, 9(8):2025–2037, 2007.

[45] Jing Sun, Laura Steindler, J. Cameron Thrash, Kimberly H. Halsey, Daniel P. Smith, Amy E. Carter, Zachary C. Landry, and Stephen J. Giovannoni. One Carbon Metabolism in SAR11 Pelagic Marine Bacteria. *PLoS ONE*, 6(8):e23973, 2011.

[46] Ronald P. Kiene and Lynn P. Hoffmann Williams. Glycine betaine uptake, retention, and degradation by microorganisms in seawater. *Limnology and Oceanography*, 43(7):1592–1603, 1998.

[47] Helen G Gika, Georgios A Theodoridis, Robert S Plumb, and Ian D Wilson. Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics. *J. Pharm. Biomed. Anal.*, 87:12–25, 2014.

[48] Blandine Madji Hounoum, Hélène Blasco, Patrick Emond, and Sylvie Mavel. Liquid chromatography–high-resolution mass spectrometry-based cell metabolomics: Experimental design, recommendations, and applications. *TrAC, Trends Anal. Chem.*, 75:118–128, 2016.

[49] Marko Sysi-Aho, Mikko Katajamaa, Laxman Yetukuri, and Matej Orešič. Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinf.*, 8:93, 2007.

[50] Melissa C Kido Soule, Krista Longnecker, Winifred M Johnson, and Elizabeth B Kujawinski. Environmental metabolomics: Analytical strategies. *Mar. Chem.*, 177:374–387, 2015.

[51] Carole A Llewellyn, Ulf Sommer, Chris L Dupont, Andrew E Allen, and Mark R Viant. Using community metabolomics as a new approach to discriminate marine microbial particulate organic matter in the western english channel. *Prog. Oceanogr.*, 137:421–433, 2015.

[52] Winifred M Johnson, Melissa C Kido Soule, and Elizabeth B Kujawinski. Evidence for quorum sensing and differential metabolite production by a marine bacterium in response to dmsp. *ISME J.*, 10:2304–2316, 2016.

[53] Alexandra Barofsky, Charles Vidoudez, and Georg Pohnert. Metabolic profiling reveals growth stage variability in diatom exudates. *Limnol. Oceanogr.: Methods*, 7(6):382–390, 2009.

[54] Tami L Swenson, Stefan Jenkins, Benjamin P Bowen, and Trent R Northen. Untargeted soil metabolomics methods for analysis of extractable organic matter. *Soil Biol. Biochem.*, 80:189–198, 2015.

[55] Richard Baran, Benjamin P Bowen, Nicholas J Bouskill, Eoin L Brodie, Steven M Yannone, and Trent R Northen. Metabolite identification in synechococcus sp. pcc 7002 using untargeted stable isotope assisted metabolite profiling. *Anal. Chem.*, 82(21):9034–9042, 2010.

[56] Carsten Paul, Michaela A Mausz, and Georg Pohnert. A co-culturing/metabolomics approach to investigate chemically mediated interactions of planktonic organisms reveals influence of bacteria on diatom metabolism. *Metabolomics*, 9(2):349–359, 2013.

[57] Winifred M. Johnson, Melissa C. Kido Soule, and Elizabeth B. Kujawinski. Extraction efficiency and quantification of dissolved metabolites in targeted marine metabolomics. *Limnology and Oceanography: Methods*, 15(4):417–428, 2017.

[58] Warwick B Dunn, David Broadhurst, Paul Begley, Eva Zelena, Sue Francis-McIntyre, Nadine Anderson, Marie Brown, Joshau D Knowles, Antony Halsall, John N Haselden, Andrew W Nicholls, Ian D Wilson, Douglas B Keil, Royston Goodacre, and The Human Serum Metabolome Consortium. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.*, 6(7):1060–1083, 2011.

[59] Mónica Calderón-Santiago, María A López-Bascón, Ángela Peralbo-Molina, and Feliciano Priego-Capote. Metaboqc: A tool for correcting untargeted metabolomics data with mass spectrometry detection using quality controls. *Talanta*, 174:29–37, 2017.

[60] Xiaotao Shen, Xiaoyun Gong, Yuping Cai, Yuan Guo, Jia Tu, Hao Li, Tao Zhang, Jialin Wang, Fuzhong Xue, and Zheng-Jiang Zhu. Normalization and integration of large-scale metabolomics data using support vector regression. *Metabolomics*, 12(5):1–12, 2016.

[61] Yoric Gagnebin, David Tonoli, Pierre Lescuyer, Belen Ponte, Sophie de Seigneux, Pierre-Yves Martin, Julie Schappler, Julien Boccard, and Serge Rudaz. Metabolomic analysis of urine samples by uhplc-qtof-ms: Impact of normalization strategies. *Anal. Chim. Acta*, 955:27–35, 2017.

[62] Yiman Wu and Liang Li. Sample normalization methods in quantitative metabolomics. *J. Chromatogr. A*, 1430:80–95, 2016.

[63] Mariusz A Bromke, Jamal S Sabir, Fahad A Alfassi, Nahid H Hajarah, Saleh A Kabli, Abdulrahman L Al-Malki, Matt P Ashworth, Michaël Méret, Robert K Jansen, and Lothar Willmitzer. Metabolomic profiling of 13 diatom cultures and their adaptation to nitrate-limited growth conditions. *PLoS One*, 10(10):e0138965, 2015.

[64] Fan Fei, Dawn ME Bowdish, and Brian E McCarry. Comprehensive and simultaneous coverage of lipid and polar metabolites for endogenous cellular metabolomics using hilic-tof-ms. *Anal. Bioanal. Chem.*, 406(15):3723–3733, 2014.

[65] Sabina Bijlsma, Ivana Bobeldijk, Elwin R Verheij, Raymond Ramaker, Sunil Kochhar, Ian A Macdonald, Ben Van Ommen, and Age K Smilde. Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. *Anal. Chem.*, 78(2):567–574, 2006.

[66] Katherine R Heal, Laura Truxal Carlson, Allan H Devol, E Armbrust, James W Moffett, David A Stahl, and Anitra E Ingalls. Determination of four forms of vitamin b12 and other b vitamins in seawater by liquid chromatography/tandem mass spectrometry. *Rapid Commun. Mass Spectrom.*, 28(22):2398–2404, 2014.

[67] Katherine R Heal, Wei Qin, Francois Ribalet, Anthony D Bertagnolli, Willow Coyote-Maestas, Laura R Hmelo, James W Moffett, Allan H Devol, E Virginia Armbrust, David A Stahl, and Anitra E Ingalls. Two distinct pools of b12 analogs reveal community interdependencies in the ocean. *Proc. Natl. Acad. Sci. U.S.A.*, 114(2):364–369, 2017.

[68] C Bueschl, R Krska, B Kluger, and R Schuhmacher. Isotopic labeling-assisted metabolomics using lc–ms. *Anal. Bioanal. Chem.*, 405(1):27–33, 2013.

[69] Mikko Katajamaa and Matej Orešič. Processing methods for differential analysis of lc/ms profile data. *BMC Bioinf.*, 6:179, 2005.

[70] E Graham Bligh and W Justin Dyer. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.*, 37(8):911–917, 1959.

[71] AB Canelas, A ten Pierick, C Ras, RM Seifar, JC van Dam, WM van Gulik, and JJ Heijnen. Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics. *Anal. Chem.*, 81(17):7379–7389, 2009.

[72] Brendan MacLean, Daniela M. Tomazela, Nicholas Shulman, Matthew Chambers, Gregory L. Finney, Barbara Frewen, Randall Kern, David L. Tabb, Daniel C. Liebler, and Michael J. MacCoss. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, 26(7):966–968, 2010.

[73] Matthew C. Chambers, Brendan MacLean, Robert Burke, Dario Amodei, Daniel L. Ruderman, Steffen Neumann, Laurent Gatto, Bernd Fischer, Brian Pratt, Jarrett Egertson, Katherine Hoff, Darren Kessner, Natalie Tasman, Nicholas Shulman, Barbara Frewen, Tahmina A. Baker, Mi Youn Brusniak, Christopher Paulse, David Creasy, Lisa

Flashner, Kian Kani, Chris Moulding, Sean L. Seymour, Lydia M. Nuwaysir, Brent Lefebvre, Frank Kuhlmann, Joe Roark, Paape Rainer, Suckau Detlev, Tina Hemenway, Andreas Huhmer, James Langridge, Brian Connolly, Trey Chadick, Krisztina Holly, Josh Eckels, Eric W. Deutsch, Robert L. Moritz, Jonathan E. Katz, David B. Agus, Michael MacCoss, David L. Tabb, and Parag Mallick. A cross-platform toolkit for mass spectrometry and proteomics, 2012.

[74] Colin A. Smith, Elizabeth J. Want, Grace O'Maille, Ruben Abagyan, and Gary Siuzdak. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry*, 78(3):779–787, 2006.

[75] Ralf Tautenhahn, Christoph Böttcher, and Steffen Neumann. Highly sensitive feature detection for high resolution lc/ms. *BMC Bioinf.*, 9(1):504, 2008.

[76] H Paul Benton, Elizabeth J Want, and Timothy MD Ebbels. Correction of mass calibration gaps in liquid chromatography–mass spectrometry metabolomics data. *Bioinformatics*, 26(19):2488–2489, 2010.

[77] Gunnar Libiseller, Michaela Dvorzak, Ulrike Kleb, Edgar Gander, Tobias Eisenberg, Frank Madeo, Steffen Neumann, Gert Trausinger, Frank Sinner, Thomas Pieber, and Christoph Magnes. Ipo: a tool for automated optimization of xcms parameters. *BMC Bioinf.*, 16(1):118, 2015.

[78] Kévin Contrepois, Lihua Jiang, and Michael Snyder. optimized analytical procedures for the untargeted metabolomic profiling of human urine and plasma by combining hydrophilic interaction (hilic) and reverse-phase liquid chromatography (rplc)–mass spectrometry. *Mol. Cell. Proteomics*, 14(6):1684–1695, 2015.

[79] Simon Cubbon, Carla Antonio, Julie Wilson, and Jane Thomas-Oates. Metabolomic applications of hilic–lc–ms. *Mass Spectrom. Rev.*, 29(5):671–684, 2010.

[80] Konstantina Spagou, Helen Tsoukali, Nikolaos Raikos, Helen Gika, Ian D Wilson, and Georgios Theodoridis. Hydrophilic interaction chromatography coupled to ms for metabonomic/metabolomic studies. *J. Sep. Sci.*, 33(6-7):716–727, 2010.

[81] Kelsey L Poulson-Ellestad, Christina M Jones, Jessie Roy, Mark R Viant, Facundo M Fernández, Julia Kubanek, and Brook L Nunn. Metabolomics and proteomics reveal impacts of chemically mediated competition on marine plankton. *Proc. Natl. Acad. Sci. U.S.A.*, 111(24):9009–9014, 2014.

[82] Maria Vinaixa, Sara Samino, Isabel Saez, Jordi Duran, Joan J Guinovart, and Oscar Yanes. A guideline to univariate statistical analysis for lc/ms-based untargeted metabolomics-derived data. *Metabolites*, 2(4):775–795, 2012.

[83] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.

[84] Christopher B. Field, Michael J. Behrenfeld, James T. Randerson, and Paul Falkowski. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science (New York, N.Y.)*, 281(5374):237–40, 1998.

[85] Sarah K. Hu, Paige E. Connell, Lisa Y. Mesrop, and David A. Caron. A Hard Day's night: Diel shifts in microbial eukaryotic activity in the North Pacific Subtropical Gyre. *Frontiers in Marine Science*, 5(OCT):351, 2018.

[86] Frank O. Aylward, Dominique Boeuf, Daniel R. Mende, Elisha M. Wood-Charlson, Alice Vislova, John M. Eppley, Anna E. Romano, and Edward F. DeLong. Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proceedings of the National Academy of Sciences of the United States of America*, 114(43):11446–11451, 2017.

[87] Kimberly H. Halsey and Bethan M. Jones. Phytoplankton Strategies for Photosynthetic Energy Allocation. *Annual Review of Marine Science*, 7(1):265–297, 2015.

[88] Jacob R. Waldbauer, Sébastien Rodrigue, Maureen L. Coleman, Sallie W. Chisholm, and Z Sun. Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS ONE*, 7(8):e43432, 2012.

[89] Kevin W. Becker, James R. Collins, Bryndan P. Durham, Ryan D. Groussman, Angelicque E. White, Helen F. Fredricks, Justin E. Ossolinski, Daniel J. Repeta, Paul Carini, E. Virginia Armbrust, and Benjamin A.S. Van Mooy. Daily changes in phytoplankton lipidomes reveal mechanisms of energy storage in the open ocean. *Nature Communications*, 9(1):5179, 2018.

[90] Angelicque E. White, Benedetto Barone, Ricardo M. Letelier, and David M. Karl. Productivity diagnosed from the diel cycle of particulate carbon in the North Pacific Subtropical Gyre. *Geophysical Research Letters*, 44(8):3752–3760, 2017.

[91] R. H. Foy and R. V. Smith. The role of carbohydrate accumulation in the growth of planktonic oscillatoria species. *British Phycological Journal*, 15(2):139–150, 1980.

[92] Kevin W. Becker, Matthew J. Harke, Daniel R. Mende, Daniel Muratore, Joshua S. Weitz, Edward F. DeLong, Sonya T. Dyhrman, and Benjamin A. S. Van Mooy. Combined pigment and metatranscriptomic analysis reveals highly synchronized diel patterns of phenotypic light response across domains in the open oligotrophic ocean. *The ISME Journal*, pages 1–14, 2020.

[93] Stuart G Wakeham, Cindy Lee, John I Hedges, Peter J Hernes, and Michael J Peterson. Molecular indicators of diagenetic status in marine organic matter. *Geochimica et Cosmochimica Acta*, 61(24):5363–5369, 1997.

[94] J. I. Hedges, J. A. Baldock, Y. Gélinas, C. Lee, M. L. Peterson, and S. G. Wakeham. The biochemical and elemental compositions of marine plankton: A NMR perspective. *Marine Chemistry*, 78(1):47–63, 2002.

[95] P. H. Yancey. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *Journal of Experimental Biology*, 208(15):2819–2830, 2005.

[96] Matthias Hirth, Silvia Liverani, Sebastian Mahlow, Francois Yves Bouget, Georg Pohnert, and Severin Sasso. Metabolic profiling identifies trehalose as an abundant and diurnally fluctuating metabolite in the microalga Ostreococcus tauri. *Metabolomics*, 13(6):68, 2017.

[97] Elizabeth B Kujawinski, Krista Longnecker, Harriet Alexander, Sonya T Dyhrman, Cara L Fiore, Sheean T Haley, and Winifred M Johnson. Phosphorus availability regulates intracellular nucleotides in marine eukaryotic phytoplankton. *Limnology and Oceanography Letters*, 2(4):119–129, 2017.

[98] Katherine R Heal, Wei Qin, Francois Ribalet, Anthony D Bertagnolli, Willow Coyote-Maestas, Laura R Hmelo, James W Moffett, Allan H Devol, E Virginia Armbrust, David A Stahl, and Anitra E Ingalls. Two distinct pools of B12 analogs reveal community interdependencies in the ocean. *Proceedings of the National Academy of Sciences of the United States of America*, 114(2):364–369, 2017.

[99] Winifred M. Johnson, Krista Longnecker, Melissa C. Kido Soule, William A. Arnold, Maya P. Bhatia, Steven J. Hallam, Benjamin A.S. Van Mooy, and Elizabeth B. Kujawinski. Metabolite composition of sinking particles differs from surface suspended particles across a latitudinal transect in the South Atlantic. *Limnology and Oceanography*, 65(1):111–127, 2019.

[100] Katherine R. Heal, Natalie A. Kellogg, Laura T. Carlson, Regina M. Lionheart, and Anitra E. Ingalls. Metabolic Consequences of Cobalamin Scarcity in the Diatom Thalassiosira pseudonana as Revealed Through Metabolomics. *Protist*, 170(3):328–348, 2019.

[101] Sophie Goulitquer, Philippe Potin, and Thierry Tonon. Mass spectrometry-based metabolomics to elucidate functions in marine organisms and ecosystems. *Marine drugs*, 10(4):849–880, 2012.

[102] Zoe V. Finkel, Mick J. Follows, Justin D. Liefer, Chris M. Brown, Ina Benner, and Andrew J. Irwin. Phylogenetic Diversity in the Macromolecular Composition of Microalgae. *PLOS ONE*, 11(5):e0155977, 2016.

[103] David G. Welkie, Benjamin E. Rubin, Spencer Diamond, Rachel D. Hood, David F. Savage, and Susan S. Golden. A hard day's night: cyanobacteria in diel cycles. *Trends in Microbiology*, 27(3):231–242, 2018.

[104] Sabine Eva Will, Petra Henke, Christian Boedeker, Sixing Huang, Henner Brinkmann, Manfred Rohde, Michael Jarek, Thomas Friedl, Steph Seufert, Martin Schumacher, Jörg Overmann, Meina Neumann-Schaal, and Jörn Petersen. Day and night: Metabolic profiles and evolutionary relationships of six axenic non-marine cyanobacteria. *Genome Biology and Evolution*, 11(1):270–294, 2019.

[105] Wade C Winkler, Ali Nahvi, Narasimhan Sudarsan, Jeffrey E Barrick, and Ronald R Breaker. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature Structural Biology*, 10(9):701–707, 2003.

[106] Yanmei Shi, Gene W Tyson, John M Eppley, and Edward F DeLong. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *The ISME Journal*, 5(6):999–1013, 2011.

[107] Amel Latifi, Marion Ruiz, and Cheng Cai Zhang. Oxidative stress in cyanobacteria, 2009.

[108] Christopher L. Dupont, Tyler J. Goepfert, Peter Lo, Liping Wei, and Beth A. Ahner. Diurnal cycling of glutathione in marine phytoplankton: Field and culture studies. *Limnology and Oceanography*, 49(4):991–996, 2004.

[109] Kimberly Halsey, Allen Milligan, and Michael Behrenfeld. Contrasting Strategies of Photosynthetic Energy Utilization Drive Lifestyle Strategies in Ecologically Important Picoeukaryotes. *Metabolites*, 4(2):260–280, 2014.

[110] S. Bertilsson, O. Berglund, M. Pullin, and S. Chisholm. Release of dissolved organic matter by Prochlorococcus, 2005.

[111] Shrameeta Shinde, Xiaohui Zhang, Sonali P. Singapuri, Isha Kalra, Xianhua Liu, Rachael M. Morgan-Kiss, and Xin Wang. Glycogen Metabolism Supports Photosynthesis Start through the Oxidative Pentose Phosphate Pathway in Cyanobacteria. *Plant physiology*, 182(1):507–517, 2020.

[112] Björn Gebser and Georg Pohnert. Synchronized regulation of different zwitterionic metabolites in the osmoadaption of phytoplankton. *Marine Drugs*, 11(6):2168–2182, 2013.

[113] Stephan Klähn, Claudia Steglich, Wolfgang R. Hess, and Martin Hagemann. Glucosylglycerate: A secondary compatible solute common to marine cyanobacteria from nitrogen-poor environments. *Environmental Microbiology*, 12(1):83–94, 2010.

[114] Stephan Klähn and Martin Hagemann. Compatible solute biosynthesis in cyanobacteria. *Environmental Microbiology*, 13(3):551–562, 2011.

[115] GR DiTullio and EA Laws. Diel periodicity of nitrogen and carbon assimilation in five species of marine phytoplankton: accuracy of methodology for predicting N-assimilation rates and N/C composition ratios. *Marine Ecology Progress Series*, 32:123–132, 1986.

[116] Marco Mühlenbruch, Hans-Peter Grossart, Falk Eigemann, and Maren Voss. Mini-review: Phytoplankton-derived polysaccharides in the marine environment and their interactions with heterotrophic bacteria. *Environmental Microbiology*, 20(8):2671–2685, 2018.

[117] Thomas Lacour, Antoine Sciandra, Amélie Talec, Patrick Mayzaud, and Olivier Bernard. Diel variations of carbohydrates and neutral lipids in nitrogen-sufficient and nitrogen-starved cyclostat cultures of isochrysis sp. *Journal of Phycology*, 48(4):966–975, 2012.

[118] Nadin Pade, Justine Compaoré, Stephan Klähn, Lucas J. Stal, and Martin Hagemann. The marine cyanobacterium Crocosphaera watsonii WH8501 synthesizes the compatible solute trehalose by a laterally acquired OtsAB fusion protein. *Environmental Microbiology*, 14(5):1261–1271, 2012.

[119] Anthony Dron, Sophie Rabouille, Pascal Claquin, Patrick Chang, Virginie Raimbault, Amélie Talec, and Antoine Sciandra. Light:dark (12:12 h) quantification of carbohydrate fluxes in Crocosphaera watsonii. *Aquatic Microbial Ecology*, 68(1):43–55, 2012.

[120] Wiebke Mohr, Maria Paola Intermaggio, and Julie LaRoche. Diel rhythm of nitrogen and carbon metabolism in the unicellular, diazotrophic cyanobacterium Crocosphaera watsonii WH8501. *Environmental Microbiology*, 12(2):412–421, 2010.

[121] A. Mitsui, S. Kumazawa, A. Takahashi, H. Ikemoto, S. Cao, and T. Arai. Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically. *Nature*, 323(6090):720–722, 1986.

[122] Tobias Großkopf and Julie LaRoche. Direct and indirect costs of dinitrogen fixation in Crocosphaera watsonii WH8501 and possible implications for the nitrogen cycle. *Frontiers in Microbiology*, 3(JUL):236, 2012.

[123] Keisuke Inomura, Jason Bragg, and Michael J Follows. A quantitative analysis of the direct and indirect costs of nitrogen fixation: a model based on Azotobacter vinelandii. *The ISME Journal*, 11(1):166–175, 2017.

[124] Shellie R Bench, Philip Heller, Ildiko Frank, Martha Arciniega, Irina N Shilova, and Jonathan P Zehr. Whole genome comparison of six Crocosphaera watsonii strains with differing phenotypes. *Journal of Phycology*, 49(4):786–801, 2013.

[125] Jill A. Sohm, Eric A. Webb, and Douglas G. Capone. Emerging patterns of marine nitrogen fixation. *Nature Reviews Microbiology*, 9(7):499–508, 2011.

[126] Mak A Saito, Erin M Bertrand, Stephanie Dutkiewicz, Vladimir V Bulygin, Dawn M Moran, Fanny M Monteiro, Michael J Follows, Frederica W Valois, and John B Waterbury. Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph Crocosphaera watsonii. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6):2184–9, 2011.

[127] Tuo Shi, Irina Ilikchyan, Sophie Rabouille, and Jonathan P Zehr. Genome-wide analysis of diel gene expression in the unicellular N(2)-fixing cyanobacterium Crocosphaera watsonii WH 8501. *The ISME journal*, 4(5):621–632, 2010.

[128] Martin J. Szul, Stephen P. Dearth, Shawn R. Campagna, and Erik R. Zinser. Carbon Fate and Flux in Prochlorococcus under Nitrogen Limitation. *mSystems*, 4(1):e00254–18, 2019.

[129] Simona Fenizia, Kathleen Thume, Marino Wirgenings, and Georg Pohnert. Ectoine from bacterial and algal origin is a compatible solute in microalgae. *Marine Drugs*, 18(1):42, 2020.

[130] W. Sunda, D. J. Kieber, R. P. Kiene, and S. Huntsman. An antioxidant function for DMSP and DMS in marine algae. *Nature*, 418(6895):317–320, 2002.

[131] Daniela A. Del Valle, Ronald P. Kiene, and David M. Karl. Effect of visible light on dimethylsulfoniopropionate assimilation and conversion to dimethylsulfide in the North Pacific Subtropical Gyre. *Aquatic Microbial Ecology*, 66(1):47–62, 2012.

[132] Mary Ann Moran and Bryndan P. Durham. Sulfur metabolites in the pelagic ocean, 2019.

[133] Arezue F.B. Boroujerdi, Peter A. Lee, Giacomo R. DiTullio, Michael G. Janech, Sarah B. Vied, and Daniel W. Bearden. Identification of isethionic acid and other small molecule metabolites of Fragilariopsis cylindrus with nuclear magnetic resonance. *Analytical and Bioanalytical Chemistry*, 404(3):777–784, 2012.

[134] Carl M. Boyd and D. Gradmann. Impact of osmolytes on buoyancy of marine phytoplankton. *Marine Biology*, 141(4):605–618, 2002.

[135] Elisabeth L. Clifford, Marta M. Varela, Daniele De Corte, Antonio Bode, Victor Ortiz, Gerhard J. Herndl, and Eva Sintes. Taurine Is a Major Carbon and Energy Source for Marine Prokaryotes in the North Atlantic Ocean off the Iberian Peninsula. *Microbial Ecology*, 78(2):299–312, 2019.

[136] David L. Kirchman and Robert E. Hodson. Metabolic regulation of amino acid uptake in marine waters. *Limnology and Oceanography*, 31(2):339–350, 1986.

[137] D.L. Kirchman. The Contribution of Monomers and other Low-Molecular Weight Compounds to the Flux of Dissolved Organic Material in Aquatic Ecosystems. In *Aquatic Ecosystems*, pages 217–241. Academic Press, 2003.

[138] Ronald P. Kiene, Lynn P. Hoffmann Williams, and Joel E. Walker. Seawater microorganisms have a high affinity glycine betaine uptake system which also recognizes dimethylsulfoniopropionate. *Aquatic Microbial Ecology*, 15(1):39–51, 1998.

[139] Stephen E. Noell and Stephen J. Giovannoni. SAR11 bacteria have a high affinity and multifunctional glycine betaine transporter. *Environmental Microbiology*, 21(7):2559–2575, 2019.

[140] Eva Bucciarelli, William G. Sunda, Sauveur Belviso, and Géraldine Sarthou. Effect of the diel cycle on production of dimethylsulfoniopropionate in batch cultures of Emiliania huxleyi. *Aquatic Microbial Ecology*, 48(1):73–81, 2007.

[141] He Fu, Mario Uchimiya, Jeff Gore, and Mary Ann Moran. Ecological drivers of bacterial community assembly in synthetic phycospheres. *Proceedings of the National Academy of Sciences of the United States of America*, 117(7):3656–3662, 2020.

[142] Remington X Poulin, Serge Lavoie, Katherine Siegel, David A Gaul, Marc J Weissburg, and Julia Kubanek. Chemical encoding of risk perception and predator detection among estuarine invertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, 115(4):662–667, 2018.

[143] Jarred E Swalwell, Francois Ribalet, and E. Virginia Armbrust. SeaFlow: A novel underway flow-cytometer for continuous observations of phytoplankton in the ocean. *Limnology and Oceanography: Methods*, 9(10):466–477, 2011.

[144] Craig F. Bohren and Donald R. Huffman. *Absorption and Scattering of Light by Small Particles*. Wiley, 1998.

[145] François Ribalet, Chris Berthiaume, Annette Hynes, Jarred Swalwell, Michael Carlson, Sophie Clayton, Gwenn Hennon, Camille Poirier, Eric Shimabukuro, Angelicque White, and E. Virginia Armbrust. SeaFlow data v1, high-resolution abundance, size and biomass of small phytoplankton in the North Pacific. *Scientific Data*, 6(1):277, 2019.

[146] Astrid Spielmeyer and Georg Pohnert. Direct quantification of dimethylsulfoniopropionate (DMSP) with hydrophilic interaction liquid chromatography/mass spectrometry. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 878(31):3238–3242, 2010.

[147] Daniel R. Mende, Jessica A. Bryant, Frank O. Aylward, John M. Eppley, Torben Nielsen, David M. Karl, and Edward F. Delong. Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nature Microbiology*, 2(10):1367–1373, 2017.

[148] Szymon M. Kiełbasa, Raymond Wan, Kengo Sato, Paul Horton, and Martin C. Frith. Adaptive seeds tame genomic sequence comparison. *Genome Research*, 21(3):487–493, 2011.

[149] Scott M Gifford, Jamie W Becker, Oscar A Sosa, Daniel J Repeta, and Edward F DeLong. Quantitative Transcriptomics Reveals the Growth- and Nutrient-Dependent Response of a Streamlined Marine Methylotroph to Methanol and Naturally Occurring Dissolved Organic Matter. *mBio*, 7(6):e01279–16, 2016.

[150] M. Kanehisa and Susumu Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.

[151] Manfred G Grabherr, Brian J Haas, Moran Yassour, Joshua Z Levin, Dawn A Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica Di Palma, Bruce W Birren, Chad Nusbaum, Kerstin Lindblad-Toh, Nir Friedman, and Aviv Regev. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7):644–652, 2011.

[152] Richard Smith-Unna, Chris Boursnell, Rob Patro, Julian M Hibberd, and Steven Kelly. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome research*, 26(8):1134–44, 2016.

[153] Martin Steinegger and Johannes Söding. Clustering huge protein sequence sets in linear time. *Nature Communications*, 9(1):2542, 2018.

[154] Benjamin Buchfink, Chao Xie, and Daniel H Huson. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1):59–60, 2015.

[155] Sean R. Eddy. Accelerated profile HMM searches. *PLoS Computational Biology*, 7(10):e1002195, 2011.

[156] Takuya Aramaki, Romain Blanc-Mathieu, Hisashi Endo, Koichi Ohkubo, Minoru Kanehisa, Susumu Goto, and Hiroyuki Ogata. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*, 36(7):2251–2252, 2019.

[157] Nicolas L Bray, Harold Pimentel, Páll Melsted, and Lior Pachter. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5):525–527, 2016.

[158] Dan Tenenbaum. KEGGREST: Client-side REST access to KEGG, 2017.

[159] Paul F Thaben and Pål O Westermark. Detecting rhythms in time series with RAIN. *Journal of biological rhythms*, 29(6):391–400, 2014.

[160] Yi Bu Chen, Jonathan P. Zehr, and Mark Mellon. Growth and nitrogen fixation of the diazotrophic filamentous nonheterocystous cyanobacterium Trichodesmium sp. IMS 101 in defined media: Evidence for a circadian rhythm. *Journal of Phycology*, 32(6):916–923, 1996.

[161] Jamie W. Becker, Shane L. Hogle, Kali Rosendo, and Sallie W. Chisholm. Co-culture and biogeography of Prochlorococcus and SAR11. *ISME Journal*, 13(6):1506–1519, 2019.

[162] Lisa R. Moore, Allison Coe, Erik R. Zinser, Mak A. Saito, Matthew B. Sullivan, Debbie Lindell, Katya Frois-Moniz, John Waterbury, and Sallie W. Chisholm. Culturing the marine cyanobacterium Prochlorococcus. *Limnology and Oceanography: Methods*, 5(OCT):353–362, 2007.

[163] J. Jeffrey Morris and Erik R. Zinser. Continuous hydrogen peroxide production by organic buffers in phytoplankton culture media. *Journal of Phycology*, 49(6):1223–1228, 2013.

[164] J Oksanen, F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R.B. O'hara, G.L. Simpson, P. Solymos, M.H. Stevens, H. Wagner, and M.J. Oksanen. Package 'vegan', 2013.

[165] Joseph B Kruskal and Myron Wish. *Multidimensional scaling*. SAGE Publications, 1978.

[166] Steven J. Biller, Paul M. Berube, Debbie Lindell, and Sallie W. Chisholm. Prochlorococcus: the structure and function of collective diversity. *Nature Reviews Microbiology*, 13(1):13–27, 2014.

[167] Jonathan P. Zehr and Raphael M. Kudela. Nitrogen Cycle of the Open Ocean: From Genes to Ecosystems. *Annual Review of Marine Science*, 3(1):197–225, 2011.

[168] Michael J Behrenfeld, Kirby Worthington, Robert M Sherrell, Francisco P Chavez, Peter Strutton, Michael McPhaden, and Donald M Shea. Controls on tropical pacific ocean productivity revealed through nutrient stress diagnostics. *Nature*, 442(7106):1025–1028, 2006.

[169] Ricardo M. Letelier, Karin M. Björkman, Matthew J. Church, Douglas S. Hamilton, Natalie M. Mahowald, Rachel A. Scanza, Niklas Schneider, Angelicque E. White, and David M. Karl. Climate-driven oscillation of phosphorus and iron limitation in the North Pacific Subtropical Gyre. *Proceedings of the National Academy of Sciences of the United States of America*, 116(26):12720–12728, 2019.

[170] Erin M Bertrand, John P McCrow, Ahmed Moustafa, Hong Zheng, Jeffrey B McQuaid, Tom O Delmont, Anton F Post, Rachel E Sipler, Jenna L Spackeen, Kai Xu, Deborah a Bronk, David a Hutchins, and Andrew E Allen. Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proceedings*

of the National Academy of Sciences of the United States of America, 112(32):9938–43, 2015.

[171] M. A. Saito, M. R. McIlvin, D. M. Moran, T. J. Goepfert, G. R. DiTullio, a. F. Post, and C. H. Lamborg. Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. *Science*, 345(6201):1173–1177, 2014.

[172] David Tilman. Resource Competition between Plankton Algae: An Experimental and Theoretical Approach. *Ecology*, 58(2):338–348, 1977.

[173] Charlotte Nef, Céline Henry, Élodie Nicolau, Jean-Baptiste Bérard, Fabienne Hervé, Amandine M. N. Caruana, Raymond Kaas, Francis Mairet, and Matthieu Garnier. Cobalamin Scarcity Modifies Carbon Allocation and Impairs DMSP Production Through Methionine Metabolism in the Haptophyte Microalgae Tisochrysis lutea. *Frontiers in Marine Science*, 7:862, 2020.

[174] Kelsey L Poulson-Ellestad, Christina M Jones, Jessie Roy, Mark R Viant, Facundo M Fernández, Julia Kubanek, and Brook L Nunn. Metabolomics and proteomics reveal impacts of chemically mediated competition on marine plankton. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24):9009–14, 2014.

[175] Kevin J. Flynn, David M.J. Dickson, and Omar A. Al-amoudi. The ratio of glutamine:gLutamate in microalgae: A biomarker for N-status suitable for use at natural cell densities. *Journal of Plankton Research*, 11(1):165–170, 1989.

[176] Lee Recht, Nadine Töpfer, Albert Batushansky, Noga Sikron, Yves Gibon, Aaron Fait, Zoran Nikoloski, Sammy Boussiba, and Aliza Zarka. Metabolite profiling and integrative modeling reveal metabolic constraints for carbon partitioning under nitrogen starvation in the green algae Haematococcus pluvialis. *Journal of Biological Chemistry*, 289(44):30387–30403, 2014.

[177] Charles Vidoudez and Georg Pohnert. Comparative metabolomics of the diatom Skeletonema marinoi in different growth phases. *Metabolomics*, 8(4):654–669, 2011.

[178] L. W. Juranek, P. D. Quay, R. A. Feely, D. Lockwood, D. M. Karl, and M. J. Church. Biological production in the NE Pacific and its influence on air-sea CO2 flux: Evidence from dissolved oxygen isotopes and O 2/Ar. *Journal of Geophysical Research: Oceans*, 117(5):n/a–n/a, 2012.

[179] Evan Howard, Steven Emerson, Seth Bushinsky, and Charles Stump. The role of net community production in air-sea carbon fluxes at the North Pacific subarctic-subtropical boundary region. *Limnology and Oceanography*, 55(6):2585–2596, 2010.

[180] Mary R. Gradoville, Hanna Farnelid, Angelicque E. White, Kendra A. Turk-Kubo, Brittany Stewart, François Ribalet, Sara Ferrón, Paulina Pinedo-Gonzalez, E. Virginia Armbrust, David M. Karl, Seth John, and Jonathan P. Zehr. Latitudinal constraints on the abundance and activity of the cyanobacterium UCYN-A and other marine diazotrophs in the North Pacific. *Limnology and Oceanography*, page lno.11423, 2020.

[181] Jennifer M. Ayers and M. Susan Lozier. Physical controls on the seasonal migration of the North Pacific transition zone chlorophyll front. *Journal of Geophysical Research: Oceans*, 115(5):C05001, 2010.

[182] Paulina Pinedo-González, Nicholas J. Hawco, Randelle M. Bundy, E. Virginia Armbrust, Michael J. Follows, B. B. Cael, Angelicque E. White, Sara Ferrón, David M. Karl, and Seth G. John. Anthropogenic Asian aerosols provide Fe to the North Pacific Ocean. *Proceedings of the National Academy of Sciences*, page 202010315, 2020.

[183] Kenneth W. Bruland, Eden L. Rue, Geoffrey J. Smith, and Giacomo R. DiTullio. Iron, macronutrients and diatom blooms in the Peru upwelling regime: Brown and blue waters of Peru. *Marine Chemistry*, 93(2-4):81–103, 2005.

[184] G. Cutter, P. Andersson, Lou Codispoti, P. Croot, R. François, M. C. Lohan, H. Obata, and Michiel Rutgers v. d. Loeff. Sampling and Sample-handling Protocols for GEOTRACES Cruises. *EPIC3GEOTRACES, 98 p.*, 2010.

[185] Hiroshi Tsugawa, Tomas Cajka, Tobias Kind, Yan Ma, Brendan Higgins, Kazutaka Ikeda, Mitsuhiro Kanazawa, Jean VanderGheynst, Oliver Fiehn, and Masanori Arita. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods*, 12(6):523–526, 2015.

[186] Evan Bolyen, Jai Ram Rideout, Matthew R Dillon, Nicholas A Bokulich, Christian C Abnet, Gabriel A Al-Ghalith, Harriet Alexander, Eric J Alm, Manimozhiyan Arumugam, Francesco Asnicar, et al. Reproducible, interactive, scalable and extensible microbiome data science using qiime 2. *Nature biotechnology*, 37(8):852–857, 2019.

[187] Gunnar I Roden. Subarctic-subtropical transition zone of the north pacific: large-scale aspects and mesoscale structure. *NOAA Technical Report NMFS*, 105:1–38, 1991.

[188] Andrew L King and Katherine A Barbeau. Dissolved iron and macronutrient distributions in the southern california current system. *Journal of Geophysical Research: Oceans*, 116(C3), 2011.

[189] J Malcolm Shick and Walter C Dunlap. Mycosporine-like amino acids and related gadusols: biosynthesis, accumulation, and uv-protective functions in aquatic organisms. *Annual review of Physiology*, 64(1):223–262, 2002.

[190] Aharon Oren and Nina Gunde-Cimerman. Mycosporines and mycosporine-like amino acids: Uv protectants or multipurpose secondary metabolites? *FEMS microbiology letters*, 269(1):1–10, 2007.

[191] Shailendra P Singh, Manfred Klisch, Rajeshwar P Sinha, and Donat-P Häder. Effects of abiotic stressors on synthesis of the mycosporine-like amino acid shinorine in the cyanobacterium anabaena variabilis pcc 7937. *Photochemistry and Photobiology*, 84(6):1500–1505, 2008.

[192] Nathalie Korbee Peinado, Roberto T Abdala Díaz, Félix L Figueroa, and E Walter Helbling. Ammonium and uv radiation stimulate the accumulation of mycosporine-like amino acids in porphyra columbina (rhodophyta) from patagonia, argentina. *Journal of Phycology*, 40(2):248–259, 2004.

[193] Marc Volkmann and Anna A. Gorbushina. A broadly applicable method for extraction and characterization of mycosporines and mycosporine-like amino acids of terrestrial, marine and freshwater origin. *FEMS Microbiology Letters*, 255(2):286–295, 2006.

[194] Yvonne V. Yuan, Neil D. Westcott, Chun Hu, and David D. Kitts. Mycosporine-like amino acid composition of the edible red alga, Palmaria palmata (dulse) harvested from the west and east coasts of Grand Manan Island, New Brunswick. *Food Chemistry*, 112(2):321–328, 2009.

[195] Jose I Carreto and Mario O Carignan. Mycosporine-like amino acids: relevant secondary metabolites. Chemical and ecological aspects. *Marine drugs*, 9(3):387–446, 2011.

[196] Hideshi Nakamura, Kazuhiro Fujimaki, Osamu Sampei, and Akio Murai. Gonyol: Methionine-induced sulfonium accumulation in a dinoflagellate Gonyaulax polyedra. *Tetrahedron Letters*, 34(52):8481–8484, 1993.

[197] E Granum, S Kirkvold, and SM Myklestad. Cellular and extracellular production of carbohydrates and amino acids by the marine diatom Skeletonema costatum: diel variations and effects of N depletion. *Marine Ecology Progress Series*, 242:83–94, 2002.

[198] Jie Yuan, Christopher D Doucette, William U Fowler, Xiao Jiang Feng, Matthew Piazza, Herschel A Rabitz, Ned S Wingreen, and Joshua D Rabinowitz. Metabolomics-driven quantitative analysis of ammonia assimilation in E. coli. *Molecular Systems Biology*, 5(1):302, 2009.

[199] Viktor M. Boer, Christopher A. Crutchfield, Patrick H. Bradley, David Botstein, and Joshua D. Rabinowitz. Growth-limiting Intracellular Metabolites in Yeast Growing

under Diverse Nutrient Limitations. *Molecular Biology of the Cell*, 21(1):198–211, 2010.

[200] Kevin J Flynn, David MJ Dickson, and Omar A Al-Amoudi. The ratio of glutamine: glutamate in microalgae: a biomarker for n-status suitable for use at natural cell densities. *Journal of plankton research*, 11(1):165–170, 1989.

[201] Samuel T. Wilson, Nicholas J. Hawco, E. Virginia Armbrust, Benedetto Barone, Karin M. Björkman, Angela K. Boysen, Macarena Burgos, Timothy J. Burrell, John R. Casey, Edward F. DeLong, Mathilde Dugenne, Stephanie Dutkiewicz, Sonya T. Dyhrman, Sara Ferrón, Michael J. Follows, Rhea K. Foreman, Carolina P. Funkey, Matthew J. Harke, Britt A. Henke, Christopher N. Hill, Annette M. Hynes, Anitra E. Ingalls, Oliver Jahn, Rachel L. Kelly, Angela N. Knapp, Ricardo M. Letelier, Francois Ribalet, Eric M. Shimabukuro, Ryan K. S. Tabata, Kendra A. Turk-Kubo, Angelique E. White, Jonathan P. Zehr, Seth John, and David M. Karl. Kīlauea lava fuels phytoplankton bloom in the North Pacific Ocean. *Science*, 365(6457):1040–1044, 2019.

[202] Wei-Dong Lu, Zhen-Ming Chi, and Chuan-Dong Su. Identification of glycine betaine as compatible solute in Synechococcus sp. WH8102 and characterization of its N-methyltransferase genes involved in betaine synthesis. *Archives of Microbiology*, 186(6):495–506, 2006.

[203] Maurice B Burg and Joan D Ferraris. Intracellular organic osmolytes: function and regulation. *The Journal of biological chemistry*, 283(12):7309–13, 2008.

[204] Benjamin a S Van Mooy, Helen F Fredricks, Byron E Pedler, Sonya T Dyhrman, David M Karl, Michal Koblízek, Michael W Lomas, Tracy J Mincer, Lisa R Moore, Thierry Moutin, Michael S Rappé, and Eric a Webb. Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature*, 458(7234):69–72, 2009.

[205] HM Dawson, KR Heal, A Torstensson, LT Carlson, A Ingalls, and JN Young. Large diversity in nitrogen-and sulfur-containing compatible solute profiles in polar and temperate diatoms. *Integrative and Comparative Biology*, 2020.

[206] Hannah M Dawson, Katherine R Heal, Angela K Boysen, Laura T Carlson, Anitra E Ingalls, Jodi N Young, Detlev Helmig, and Kevin Arrigo. Potential of temperature-and salinity-driven shifts in diatom compatible solute concentrations to impact biogeochemical cycling within sea ice. *Elementa: Science of the Anthropocene*, 8, 2020.

[207] Jens Boch, Bettina Kempf, and Erhard Bremer. Osmoregulation in bacillus subtilis: synthesis of the osmoprotectant glycine betaine from exogenously provided choline. *Journal of Bacteriology*, 176(17):5364–5371, 1994.

[208] Maureen D. Keller, Patricia A. Matrai, Ronald P. Kiene, and Wendy K. Bellows. Responses of coastal phytoplankton populations to nitrogen additions: Dynamics of cell-associated dimethylsulfoniopropionate (DMSP), glycine betaine (GBT), and homarine. In *Canadian Journal of Fisheries and Aquatic Sciences*, volume 61, pages 685–699. NRC Research Press Ottawa, Canada, 2004.

[209] Jamie A. Meadows and Matthew J. Wargo. Carnitine in bacterial physiology and metabolism, 2015.

[210] Gianfranco Peluso, Alfonso Barbarisi, Vincenzo Savica, Emilia Reda, Raffaella Nicolai, Paola Benatti, and Menotti Calvani. Carnitine: An osmolyte that plays a metabolic role, 2001.

[211] Victor A. Zammit, Nigel T. Price, Vicky N. Jackson, and Byun Sung Park. The role of carnitine acyltransferases in the maintenance of cell function. *Monatshefte fur Chemie*, 136(8):1299–1309, 2005.

[212] Natthawut Yodsuwan, Shigeki Sawayama, and Sarote Sirisansaneeyakul. Effect of nitrogen concentration on growth, lipid production and fatty acid profiles of the marine diatom Phaeodactylum tricornutum. *Agriculture and Natural Resources*, 51(3):190–197, 2017.

[213] Ocean Biology Processing Group. Modis aqua level 3 sst mid-ir 8 day 9km nighttime v2014.0, 2015. Date accessed: 15.08.2018, https://doi.org/10.5067/MODAM-8D9N4.

[214] Amy Apprill, Sean McNally, Rachel Parsons, and Laura Weber. Minor revision to v4 region ssu rrna 806r gene primer greatly increases detection of sar11 bacterioplankton. *Aquatic Microbial Ecology*, 75(2):129–137, 2015.

[215] Kenan Hadziavdic, Katrine Lekang, Anders Lanzen, Inge Jonassen, Eric M Thompson, and Christofer Troedsson. Characterization of the 18s rrna gene for designing universal eukaryote specific primers. *PloS one*, 9(2):e87624, 2014.

[216] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1):10–12, 2011.

[217] Benjamin J Callahan, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson, and Susan P Holmes. Dada2: high-resolution sample inference from illumina amplicon data. *Nature methods*, 13(7):581–583, 2016.

[218] Torbjørn Rognes, Tomáš Flouri, Ben Nichols, Christopher Quince, and Frédéric Mahé. Vsearch: a versatile open source tool for metagenomics. *PeerJ*, 4:e2584, 2016.

[219] Laure Guillou, Dipankar Bachar, Stéphane Audic, David Bass, Cédric Berney, Lucie Bittner, Christophe Boutte, Gaëtan Burgaud, Colomban De Vargas, Johan Decelle, et al. The protist ribosomal reference database (pr2): a catalog of unicellular eukaryote small sub-unit rrna sequences with curated taxonomy. *Nucleic acids research*, 41(D1):D597–D604, 2012.

[220] Christian Quast, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. The silva ribosomal rna gene database project: improved data processing and web-based tools. *Nucleic acids research*, 41(D1):D590–D596, 2012.

[221] Tom Berman and Deborah A. Bronk. Dissolved organic nitrogen: A dynamic participant in aquatic ecosystems. *Aquatic Microbial Ecology*, 31(3):279–305, 2003.

[222] Deborah A Bronk. Dynamics of don. *Biogeochemistry of marine dissolved organic matter*, 384:p153–247, 2002.

[223] Deborah A Bronk, Patricia M. Gilbert, and Bess B Ward. Nitrogen uptake, dissolved organic nitrogen release, and new production. *Science*, 265(5180):1843–1846, 1994.

[224] Lihini I Aluwihare, Daniel J Repeta, and Robert F Chen. A major biopolymeric component to dissolved organic carbon in surface sea water. *Nature*, 387(6629):166–169, 1997.

[225] Lihini I Aluwihare, Daniel J Repeta, and Robert F Chen. Chemical composition and cycling of dissolved organic matter in the mid-atlantic bight. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(20):4421–4437, 2002.

[226] Lihini I Aluwihare, Daniel J Repeta, Silvio Pantoja, and Carl G Johnson. Two chemically distinct pools of organic nitrogen accumulate in the ocean. *Science*, 308(5724):1007–1010, 2005.

[227] Stuart G. Wakeham and Cindy Lee. Limits of our knowledge, part 2: Selected frontiers in marine organic biogeochemistry. *Marine Chemistry*, 212:16–46, 2019.

[228] D. A. Bronk, J. H. See, P. Bradley, and L. Killberg. DON as a source of bioavailable nitrogen for phytoplankton. *Biogeosciences*, 4(3):283–296, 2007.

[229] M. C. Muñoz-Marín, G. Gómez-Baena, A. López-Lozano, J. A. Moreno-Cabezuelo, J. Díez, and J. M. García-Fernández. Mixotrophy in marine picocyanobacteria: use of organic compounds by Prochlorococcus and Synechococcus, 2020.

[230] Alexandra Z Worden, Michael J Follows, Stephen J Giovannoni, Susanne Wilken, Amy E Zimmerman, and Patrick J Keeling. Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science*, 347(6223), 2015.

[231] Isabelle Mary, Glen A Tarran, Phillip E Warwick, Matthew J Terry, David J Scanlan, Peter H Burkill, and Mikhail V Zubkov. Light enhanced amino acid uptake by dominant bacterioplankton groups in surface waters of the Atlantic Ocean. *FEMS Microbiology Ecology*, 63(1):36–45, 2008.

[232] Vanessa K. Michelou, Matthew T. Cottrell, and David L. Kirchman. Light-stimulated bacterial production and amino acid assimilation by cyanobacteria and other microbes in the North Atlantic Ocean. *Applied and Environmental Microbiology*, 73(17):5539–5546, 2007.

[233] Mikhail V. Zubkov, Bernhard M. Fuchs, Glen A. Tarran, Peter H. Burkill, and Rudolf Amann. High rate of uptake of organic nitrogen compounds by Prochlorococcus cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Applied and Environmental Microbiology*, 69(2):1299–1304, 2003.

[234] MP Sanderson, DA Bronk, JC Nejstgaard, PG Verity, AF Sazhin, and ME Frischer. Phytoplankton and bacterial uptake of inorganic and organic nitrogen during an induced bloom of Phaeocystis pouchetii. *Aquatic Microbial Ecology*, 51(2):153–168, 2008.

[235] Matthew McCarthy, Tom Pratum, John Hedges, and Ronald Benner. Chemical composition of dissolved organic nitrogen in the ocean. *Nature*, 390(6656):150–154, 1997.

[236] Ruth L Airs and Steven D Archer. Analysis of glycine betaine and choline in seawater particulates by liquid chromatography/electrospray ionaization/mass spectrometry. *Limnology and Oceanography: Methods*, 8(OCT):499–506, 2010.

[237] M. D. Keller, R. P. Kiene, P. A. Matrai, and W. K. Bellows. Production of glycine betaine and dimethylsulfoniopropionate in marine phytoplankton. I. Batch cultures. *Marine Biology*, 135(2):237–248, 1999.

[238] Jing Sun, Michaela A. Mausz, Yin Chen, and Stephen J. Giovannoni. Microbial trimethylamine metabolism in marine environments. *Environmental Microbiology*, 21(2):513–520, 2019.

[239] Kendra J. Adams, Brian Pratt, Neelanjan Bose, Laura G. Dubois, Lisa St. John-Williams, Kevin M. Perrott, Karina Ky, Pankaj Kapahi, Vagisha Sharma, Michael J. Maccoss, M. Arthur Moseley, Carol A. Colton, Brendan X. Maclean, Birgit Schilling, and J. Will Thompson. Skyline for Small Molecules: A Unifying Software Package for Quantitative Metabolomics. *Journal of Proteome Research*, 19(4):1447–1458, 2020.

[240] Lindsay K. Pino, Brian C. Searle, James G. Bollinger, Brook Nunn, Brendan MacLean, and Michael J. MacCoss. The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics, 2020.

[241] Michael Fountoulakis and Hans-Werner Lahm. Hydrolysis and amino acid composition analysis of proteins. *Journal of chromatography A*, 826(2):109–134, 1998.

[242] Nicola Gray, Rabiya Zia, Adam King, Vishal C. Patel, Julia Wendon, Mark J.W. McPhail, Muireann Coen, Robert S. Plumb, Ian D. Wilson, and Jeremy K. Nicholson. High-Speed Quantitative UPLC-MS Analysis of Multiple Amines in Human Plasma and Serum via Precolumn Derivatization with 6-Aminoquinolyl-N-hydroxysuccinimidyl Carbamate: Application to Acetaminophen-Induced Liver Failure. *Analytical Chemistry*, 89(4):2478–2487, 2017.

[243] Ian Lidbury, George Kimberley, David J. Scanlan, J. Colin Murrell, and Yin Chen. Comparative genomics and mutagenesis analyses of choline metabolism in the marine Roseobacter clade. *Environmental Microbiology*, 17(12):5048–5062, 2015.

[244] J. J. Heijnen. Impact of thermodynamic principles in systems biology. *Advances in Biochemical Engineering/Biotechnology*, 121:139–162, 2010.

[245] Jie Yuan, William U Fowler, Elizabeth Kimball, Wenyun Lu, and Joshua D Rabinowitz. Kinetic flux profiling of nitrogen assimilation in Escherichia coli. *Nature Chemical Biology*, 2(10):529–530, 2006.

[246] Diane E. Robertson, Mei-Chin Lai, Robert P. Gunsalus, and Mary F. Roberts. Composition, Variation, and Dynamics of Major Osmotic Solutes in Methanohalophilus Strain FDF1. *Applied and Environmental Microbiology*, 58(8), 1992.

[247] Marie Joncquel-Chevalier Curt, Pia-Manuela Voicu, Monique Fontaine, Anne-Frederique Dessein, Nicole Porchet, Karine Mention-Mulliez, Dries Dobbelaere, Gustavo Soto-Ares, David Cheillan, and Joseph Vamecq. Creatine biosynthesis and transport in health and disease. *Biochimie*, 119:146–165, 2015.

[248] Andrew E Allen, Assaf Vardi, and Chris Bowler. An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Current opinion in plant biology*, 9(3):264–273, 2006.

[249] Markus Wyss and Rima Kaddurah-Daouk. Creatine and creatinine metabolism. *Physiological reviews*, 2000.

[250] P H Yancey and J F Siebenaller. Trimethylamine oxide stabilizes teleost and mammalian lactate dehydrogenases against inactivation by hydrostatic pressure and trypsinolysis. *The Journal of experimental biology*, 202(Pt 24):3597–603, 1999.

[251] Ian Lidbury, J. Colin Murrell, and Yin Chen. Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 111(7):2710–2715, 2014.

[252] Ronald P. Kiene. Uptake of choline and its conversion to glycine betaine by bacteria in estuarine waters. *Applied and Environmental Microbiology*, 64(3):1045–1051, 1998.

[253] Ian D.E.A. Lidbury, J. Colin Murrell, and Yin Chen. Trimethylamine and trimethylamine N-oxide are supplementary energy sources for a marine heterotrophic bacterium: Implications for marine carbon and nitrogen cycling. *ISME Journal*, 9(3):760–769, 2015.

[254] Ciria G. Figueroa-Soto and Elisa M. Valenzuela-Soto. Glycine betaine rather than acting only as an osmolyte also plays a role as regulator in cellular metabolism, 2018.

[255] Anutthaman Parthasarathy, Michael A. Savka, and André O. Hudson. The Synthesis and Role of β-Alanine in Plants. *Frontiers in Plant Science*, 10:921, 2019.

[256] Andrew R.J. Curson, Beth T. Williams, Benjamin J. Pinchbeck, Leanne P. Sims, Ana Bermejo Martínez, Peter Paolo L. Rivera, Deepak Kumaresan, Elena Mercadé, Lewis G. Spurgin, Ornella Carrión, Simon Moxon, Rose Ann Cattolico, Unnikrishnan Kuzhiumparambil, Paul Guagliardo, Peta L. Clode, Jean Baptiste Raina, and Jonathan D. Todd. DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton. *Nature Microbiology*, 3(4):430–439, 2018.

[257] Felix Trottmann, Keishi Ishida, Jakob Franke, Aleksa Stanišić, Mie Ishida-Ito, Hajo Kries, Georg Pohnert, and Christian Hertweck. Sulfonium Acids Loaded onto an Unusual Thiotemplate Assembly Line Construct the Cyclopropanol Warhead of a *Burkholderia* Virulence Factor. *Angewandte Chemie International Edition*, 59(32):13511–13515, 2020.

[258] Elizabeth R. Hall and Samuel Gurin. Experiments in marine biochemistry. Homarine metabolism in Penaeus duorarum. *Journal of Biological Chemistry*, 250(17):6943–6946, 1975.

[259] Hiroshi Ashihara. Trigonelline (N-methylnicotinic acid) biosynthesis and its biological role in plants, 2008.

[260] D. M. J. Dickson and G. O. Kirst. Osmotic adjustment in marine eukaryotic algae: The role of inorganic ions, quaternary ammonium, tertiary sulphonium and carbohydrate solutes. II. Prasinophytes and haptophytes. *New Phytologist*, 106(4):657–666, 1987.

[261] Maria Vila-Costa, Rafel Simó, Hyakubun Harada, Josep M. Gasol, Doris Slezak, and Ronald P. Kiene. Dimethylsulfoniopropionate uptake by marine phytoplankton. *Science*, 314(5799):652–654, 2006.