

Relating Biological Rate Measurements and Microbial Processes

Across Diverse Ocean Ecosystems

by

Seaver Wang

Earth and Ocean Sciences  
Duke University

Date: \_\_\_\_\_

Approved:

\_\_\_\_\_  
Nicolas Cassar, Supervisor

\_\_\_\_\_  
Susan Lozier

\_\_\_\_\_  
Zackary Johnson

\_\_\_\_\_  
Scott Gifford

Dissertation submitted in partial fulfillment of  
the requirements for the degree  
of Doctor of Philosophy in  
Earth and Ocean Sciences in the Graduate School  
of Duke University

2019

ABSTRACT

Relating Biological Rate Measurements and Microbial Processes

Across Diverse Ocean Ecosystems

by

Seaver Wang

Division of Earth and Ocean Sciences  
Duke University

Date: \_\_\_\_\_

Approved:

\_\_\_\_\_  
Nicolas Cassar, Supervisor

\_\_\_\_\_  
Susan Lozier

\_\_\_\_\_  
Zackary Johnson

\_\_\_\_\_  
Scott Gifford

An abstract of a dissertation submitted in partial  
fulfillment of the requirements for the degree  
of Doctor of Philosophy in  
Earth and Ocean Sciences in the Graduate School of  
Duke University

2019

Copyright by  
Seaver Wang  
2019

## Abstract

Marine microbes play key roles in driving patterns of important biogeochemical processes including primary production across the global ocean. Despite the importance of such interactions between the marine microbial community and ocean biogeochemistry, oceanographers have yet to attain a deep understanding of the ecological mechanisms underlying these connections. Due to the vast scale of ocean ecosystems, however, large-scale yet high-resolution surveys are necessary to uncover specific relationships between biology and elemental cycling for more detailed study.

With this need in mind, this dissertation takes advantage of recent advances in both underway techniques to measure *in situ* biogeochemical rates—most notably the dissolved O<sub>2</sub>/Ar method for measuring net community production (NCP)—as well as molecular sequencing methods to directly investigate relationships between marine microbial community structure, productivity, nitrogen (N<sub>2</sub>) fixation, and nutrient availability across large ocean regions. At the same time, this work also improves our understanding of the O<sub>2</sub>/Ar technique by evaluating its performance and key assumptions in a dynamic upwelling environment and by presenting recommendations to improve the accuracy of productivity estimates generated using this approach.

Presenting data and measurements from the most comprehensive survey of marine microbial community structure and patterns of productivity and N<sub>2</sub> fixation in

the western North Atlantic to date, this manuscript highlights intriguing connections between regional peaks in productivity and N<sub>2</sub> fixation, the mixotrophic algae *Chrysophyceae* and *Aureococcus anophagefferens*, and *Braarudosphaera bigelowii*, a eukaryotic host organism for N<sub>2</sub>-fixing bacteria. In addition, we report a strong negative relationship between eukaryotic marine microbial diversity and productivity across the region. We further highlight the importance of considering diel cycles of productivity/respiration, other non-steady-state conditions, and vertical fluxes of O<sub>2</sub>/Ar when calculating and interpreting NCP rates obtained from surface O<sub>2</sub>/Ar measurements. Ultimately, these findings contribute to our ability to evaluate community production using surface ocean dissolved gas measurements and provide important insights into patterns of marine microbial activity and community structure into the western North Atlantic.

## **Dedication**

I dedicate this work to “mad scientists” everywhere, past, present, and future.

El. Psy. Congroo.

# Contents

|   |      |
|---|------|
| Abstract .....  | iv   |
| List of Tables .....  | xiii |
| List of Figures .....   | xiv  |
| Acknowledgements .....  | xvii |
| 1. Introduction .....   | 1    |
| 1.1 Estimation of Net Community Production from $\Delta O_2/Ar$ .....   | 5    |
| 1.2 Quantitative rRNA Amplicon Sequencing .....   | 8    |
| 1.3 Research Questions .....  | 12   |
| 2. Linking patterns of net community production and marine microbial community structure in the western North Atlantic..... | 14   |
| 2.1 Introduction.....   | 14   |
| 2.2 Materials and Methods.....  | 18   |
| Study area and collection of $O_2/Ar$ and ancillary data.....   | 18   |
| Microbial community sampling and rDNA amplicon sequencing.....  | 20   |
| Internal controls for quantitative sequencing .....   | 20   |
| DNA extraction for 16S and 18S rDNA sequencing.....   | 21   |
| Analysis Pipeline .....   | 23   |
| 2.3 Results and Discussion .....  | 25   |
| Patterns of $O_2/Ar$ -derived NCP .....   | 25   |
| Microbial community quantitative and relative abundance patterns .....  | 26   |
| Relationships between microbial community structure and NCP.....  | 35   |

|  |    |
|--|----|
| Relationships between NCP and specific microplankton taxa .....  | 38 |
| 2.4 Conclusions .....  | 43 |
| 2.5 Acknowledgements.....  | 44 |
| 3. Patterns of North Atlantic microbial community structure in relation to net<br>community production, nitrogen fixation, and nutrient availability ..... | 45 |
| 3.1 Introduction.....  | 46 |
| 3.2 Materials and Methods .....  | 50 |
| Measurements of underway net community production, nitrogen fixation rates,<br>nutrients, and trace metals. ....   | 50 |
| Sample collection and library preparation for 16S and 18S rDNA amplicon<br>sequencing.....   | 51 |
| rDNA sequencing data processing and OTU picking.....   | 52 |
| Statistical analyses.....  | 53 |
| Network analysis.....  | 55 |
| 3.3 Results and discussion.....  | 56 |
| Biogeochemistry: NCP, N <sub>2</sub> fixation rates, nutrients, and trace metals .....   | 56 |
| Microbial ecology: Microplankton community composition and abundance patterns<br>.....   | 61 |
| Broad patterns: Community diversity vs. NCP, N <sub>2</sub> fixation, and environmental<br>properties.....   | 64 |
| Fine-scale relationships: Key community members associated with NCP, N <sub>2</sub><br>fixation, and environmental properties.....                         | 69 |
| 3.4 Conclusions .....  | 76 |
| 3.5 Acknowledgements.....  | 78 |

|   |     |
|---|-----|
| 4. Lagrangian studies of net community production: assessing the effect of diel and multi-day non-steady state factors and vertical fluxes..... | 79  |
| Author contributions.....   | 79  |
| 4.1 Introduction.....   | 79  |
| 4.2 Materials and Methods.....  | 85  |
| Cruise background.....  | 85  |
| Underway and vertical profile O <sub>2</sub> /Ar measurements via EIMS and MIMS:.....   | 88  |
| Calculation of biological oxygen saturation ( $\Delta O_2/Ar$ ) .....   | 90  |
| Detrending and curve fitting of NCP data and calculation of NCP diel cycle characteristics.....   | 94  |
| Assessment of non-steady-state impacts on calculated net community production   | 95  |
| Correction for vertical O <sub>2</sub> /Ar fluxes:.....   | 97  |
| 4.3 Results.....  | 100 |
| General patterns of net community production from Lagrangian measurements in the California Current in spring/summer 2016 and 2017 .....        | 100 |
| Impact of steady-state versus non-steady-state-based calculation of <i>in-situ</i> NCP..  | 103 |
| Diel variation in O <sub>2</sub> /Ar-derived NCP signal.....  | 104 |
| Influence of vertical fluxes of O <sub>2</sub> /Ar upon calculated in situ NCP.....   | 106 |
| 4.4 Discussion.....   | 108 |
| Non-steady-state Lagrangian versus conventional Eulerian calculation of NCP <sub>prior</sub> .....  | 108 |
| Influence of diel patterns in O <sub>2</sub> /Ar upon measured NCP <sub>prior</sub> .....   | 112 |
| Influence of vertical fluxes of O <sub>2</sub> /Ar upon apparent NCP <sub>prior</sub> .....   | 114 |

|  |     |
|--|-----|
| 4.5 Conclusions .....  | 120 |
| 4.6 Acknowledgements.....  | 122 |
| 5. Conclusions.....  | 123 |
| Appendix A:.....   | 127 |
| Supplemental Methods.....  | 128 |
| Measurement of Biological O <sub>2</sub> .....   | 128 |
| Calculation of net community production .....  | 129 |
| Accounting for vertical fluxes of O <sub>2</sub> /Ar.....  | 130 |
| Determination of appropriate control DNA spike quantity .....  | 132 |
| Statistical Analysis.....  | 133 |
| Partial Least Squares Regression Analysis .....  | 134 |
| Online data availability .....   | 134 |
| Supplementary Figures.....   | 135 |
| Appendix B: Patterns of North Atlantic microbial community structure in relation to net community production, nitrogen fixation, and nutrient availability ..... | 137 |
| Supplemental Methods.....  | 137 |
| Measurements of underway dissolved surface O <sub>2</sub> /Ar .....  | 137 |
| Calculation of net community production .....  | 138 |
| Flow-through acetylene reduction nitrogen fixation assays using cavity-ringdown spectroscopy (FARACAS).....  | 141 |
| Sampling and analysis of nutrients and trace metals.....   | 142 |
| Nutrient sampling.....   | 142 |
| Trace metal sampling: dissolved trace metals .....   | 143 |

|   |     |
|---|-----|
| Trace metal sampling: particulate trace elements .....  | 143 |
| Trace metal analytical procedures: dissolved trace metals.....  | 144 |
| Trace metal analytical procedures: particulate trace elements.....  | 145 |
| Bioinformatic processing of molecular data.....   | 146 |
| Statistical analyses.....   | 147 |
| WGCNA module selection and network analyses using Flashweave, L-GRAAL .   | 149 |
| Supplemental Figures and Tables.....  | 150 |
| Appendix C: Lagrangian studies of NCP: Assessing the effect of diel biological cycles<br>and non-steady state systems.....                      | 156 |
| Supplementary Figures.....  | 156 |
| Appendix D: Ongoing work: microbial productivity-diversity relationships in the<br>Eastern North Pacific from multiple independent methods..... | 159 |
| Introduction.....   | 159 |
| Materials and Methods.....  | 161 |
| O <sub>2</sub> /Ar-derived biological flux measurements.....  | 161 |
| Microscopy and molecular sample collection and microscopic taxonomic<br>identification .....  | 163 |
| DNA extraction and library preparation .....  | 164 |
| 18S rRNA gene sequence processing.....  | 165 |
| Statistical analysis.....   | 165 |
| Results and Discussion .....  | 166 |
| Biological oxygen flux rates and patterns.....  | 166 |
| Productivity-diversity relationships.....   | 168 |

|                                  |     |
|----------------------------------|-----|
| Conclusions and Future Work..... | 171 |
| Acknowledgements.....            | 172 |
| References .....                 | 173 |
| Biography .....                  | 200 |

## List of Tables

|   |     |
|---|-----|
| Table 1: Basic metadata for each set of Lagrangian observations conducted in 2016 and 2017.   | 87  |
| Table 2: Derived median community production rates, vertical flux magnitudes, oxygen gradients, and regression-derived slopes in $\Delta O_2/Ar$ and NCP for each cycle. ....   | 103 |
| Table 3: calculated median diel range of diel cycles in $\Delta O_2/Ar$ and NCP observed during each cycle .....  | 106 |
| Table 4: Comparison of vertical eddy flux, entrainment flux, and advective flux calculations (all in units of $mmol O_2 m^{-2} d^{-1}$ ) derived using oxycline $O_2$ and oxycline biological $O_2$ ( $O_2/Ar$ ) gradients..... | 119 |

## List of Figures

|  |    |
|--|----|
| Figure 1: (a) Map of O <sub>2</sub> /Ar-derived volumetric NCP.....  | 19 |
| Figure 2: rDNA abundances (log <sub>10</sub> ) across sampled stations for (a) <i>Aureococcus anophagefferens</i> (b) SAR11 and <i>Prochlorococcus</i> . ....                | 28 |
| Figure 3: (a) Comparison of patterns of absolute abundance.....  | 30 |
| Figure 4: Bar plots of (a) 18S eukaryotic taxonomy.....  | 32 |
| Figure 5: (a) Linear regression of NCP vs. the number of observed 18S eukaryotic OTUs per sample.....  | 35 |
| Figure 6: Principal Coordinates Analysis of (a) eukaryotic 18S and (b) bacterial/archaeal 16S samples.....   | 36 |
| Figure 7: Heatmaps showing strength of correlations, determined using Partial Least Squares regression analysis.....   | 38 |
| Figure 8: Heatmaps of PLS correlations for (a) prokaryotic and (b) eukaryotic taxa at the 4th taxonomic rank, following removal of data from bloom stations.....             | 41 |
| Figure 9: a) Map of O <sub>2</sub> /Ar-based volumetric NCP rates measured via EIMS across the Western North Atlantic in July-August 2015, 2016, and 2017. ....              | 50 |
| Figure 10: a) NCP rates (black) and N <sub>2</sub> fixation rates (red) measured at each sampling site, with sites ranked from left-to-right in order of increasing NCP..... | 59 |
| Figure 11: a) NCP rates (black) and N <sub>2</sub> fixation rates (red) measured at each sampling site, with sites ranked from left-to-right in order of increasing NCP..... | 60 |
| Figure 12: Principal Coordinates Analysis (PCoA) of a) prokaryotic 16S and b) eukaryotic 18S samples.....  | 62 |
| Figure 13: Linear regressions of a) NCP b) N <sub>2</sub> fixation rates c) Mixed layer depth d) PO <sub>4</sub> e) dCu f) dMn versus 18S Shannon's H diversity. ....        | 65 |
| Figure 14: Heatmaps showing correlations between abundances of specific a) prokaryotic 16S and b) eukaryotic 18S taxa and measured metadata parameters.....                  | 69 |

|  |     |
|--|-----|
| Figure 15: Heatmaps showing PLS-derived correlations.....  | 71  |
| Figure 16: Alignment plots from network analysis.....  | 73  |
| Figure 17: Map of cruise tracks during Lagrangian deployments in spring 2016 and summer 2017 within the California Current Ecosystem.....  | 86  |
| Figure 18: Conceptual diagram illustrating underlying assumptions behind the various models for O <sub>2</sub> /Ar-derived NCP employed: NCP <sub>prior</sub> , NCP <sub>RT</sub> , and NCP <sub>RT_Vflux</sub> .....                              | 96  |
| Figure 19: NCP <sub>prior</sub> measured for each Lagrangian cycle .....   | 100 |
| Figure 20: Time series of ΔO <sub>2</sub> /Ar observations for each cycle.....   | 102 |
| Figure 21: NCP <sub>prior</sub> measured for each Lagrangian cycle and detrended using the linear fits previously obtained.....  | 105 |
| Figure 22: Time series of a) NCP <sub>prior</sub> and b) ΔO <sub>2</sub> /Ar supersaturation for Cycles 2017 #2 and #4. ....   | 110 |
| Figure 23: Three vertical profiles of a) Ar, b) O <sub>2</sub> /Ar (blue) and O <sub>2</sub> (red) saturation .....  | 119 |
| Figure 24: Rarefaction plots showing OTUs observed versus sequencing depth for (left) 16S and (right) 18S rRNA amplicon samples.....   | 135 |
| Figure 25: Observed OTUs per sample and Shannon diversity for (top) 18S and (bottom) 16S samples.....  | 136 |
| Figure 26: Histograms of the distribution of log-scaled internal standard recovery ratios for a) 16S rDNA and b) 18S rDNA molecular samples .....  | 150 |
| Figure 27: Scatter plot and goodness-of-fit of relationships .....   | 151 |
| Figure 28: a) Temperature and b) Salinity measurements recorded at sampling stations across August 2015, 2016, and 2017 cruises in the Western North Atlantic. ....  | 152 |
| Figure 29: a) log NO <sub>2</sub> +NO <sub>3</sub> , b) log PO <sub>4</sub> , and c) log silicate concentrations across the 2016 and 2017 cruises. f) NO <sub>2</sub> +NO <sub>3</sub> , e) PO <sub>4</sub> , and f) silicate concentrations. .... | 152 |
| Figure 30: Maps showing the strength of associations between each WGCNA-derived module and the sampled sites.....  | 153 |

|   |     |
|---|-----|
| Figure 31: Heatmap illustrating correlations between WGCNA modules and metadata parameters.....   | 154 |
| Figure 32: Interpolated MLD as determined from CTD rosette casts during each Lagrangian cycle. ....   | 156 |
| Figure 33: Graph showing the reduction across ROMS 4DVAR model iterations in the value of the cost function.....  | 157 |
| Figure 34: Raw $\Delta O_2/Ar$ supersaturation (blue, primary y-axis) plotted with high-resolution Chl-a measurements from the 2017 expedition (black, secondary y-axis)..... | 158 |
| Figure 35: Map of NCP measurements and cruise track for the January 2017 Sea2Space expedition.....  | 167 |
| Figure 36: Principal Coordinates Analysis (PCoA) of microscopy community samples using Bray-Curtis ecological dissimilarity. ....   | 168 |
| Figure 37: Scatter plots of a, b) eukaryotic community Shannon's H diversity and c, d) 18S community raw species richness .....   | 170 |
| Figure 38: Scatter plots of a) eukaryotic community Shannon's H diversity and b) raw eukaryotic species richness determined from microscopy analyses .....                    | 171 |

## Acknowledgements

My growth as an oceanographer over these past five years has been guided and encouraged by numerous mentors, colleagues, friends, and family members. My advisor Nicolas Cassar has played a central role in guiding my development as a researcher throughout my years at Duke. Nicolas has consistently given me the independence to choose the direction of my projects, while at the same time devoting great attention towards my work and encouraging me to push my research towards its maximum potential. I am deeply grateful for the genuine care Nicolas has shown towards my mentorship.

I also thank the members of my committee, Susan Lozier, Zackary Johnson, and Scott Gifford, for their insights, feedback, and help over the course of my PhD. At the time that I asked Susan, Zackary, and Scott to serve on my committee, I was confident that their specialized knowledge and experience would contribute greatly towards strengthening my dissertation research. In retrospect, I do not think that I could have chosen a better PhD. dissertation committee, as their interest in my work as well as their expert input on the portions of my thesis relating to ocean physics and marine microbial ecology have been of invaluable help to me.

I am additionally grateful to Thomas Kelly, Sven Kranz, Michael Stukel, Hajoong Song, Hannah Whitby, Aridane González, Erwan Delage, Damien Eveillard, H  l  ne Planquette, Hrvoje Vi  i  , Maja Mucko, Zrinka Ljube  i  , and Ivona Cetinic. I feel

fortunate to have had the opportunity to work with so many fantastic collaborators and co-authors, whose contributions and assistance helped make the work presented in this dissertation possible.

I want to thank Weiyi Tang in particular for being an exceptional collaborator, partner-in-crime, academic brother, and friend. I have immensely enjoyed our working together over these past five years, whether we were pulling long hours in the lab at sea, coordinating international shipping, commuting to UNC for class, or puzzling over QIIME errors. I am also deeply grateful to members of the Cassar Lab both past and present—Yajuan Lin first and foremost for patiently and expertly teaching me my way around molecular techniques in the lab, Rachel Eveleth and Zuchuan Li for their patient advice, help, and feedback on such a variety of topics, Adam Ulfsbo for his guidance on the EIMS, Hugo Berthelot for his help both at sea and in France, as well as Kuan Huang, Jiayun Zhou, Alex Neibergall, Shuai Gu, and Yibin Huang. I would also like to thank all of my colleagues in the EOS Division for their support over the past five years, as well as the researchers and staff of l'Institut Universitaire Européen de la Mer (IUEM) for their warmth and assistance during my visit to Brest, France in the summer of 2017.

Lastly, I am grateful to my family and friends for their longtime support throughout my graduate career, especially to my mother for always reminding me to take care of myself and to maintain a healthy life balance, and to my father for steadily cheering me on along a similar path to the one he traveled at my age. Although my

studies in Durham wound up leaving me geographically distant from many dear friends, with our frequent voice calls I have felt as though they have been at my side with every new step. Through the magic of shared interests and trials, I have also been delighted to have made a great many new friends along the way.

In the end, no words feel sufficient to convey my boundless gratitude towards all of you—mentors and colleagues, family and friends. Thank you all for being a part of my life.

# 1. Introduction

The cycling of carbon and nutrients by oceanic microbes forms a key part of global biogeochemical cycles. The mechanisms by which marine microorganisms control fluxes of carbon, nitrogen, and other elements are varied and complex, not only governing important processes like climate feedbacks and trophic transfers but also evolving in response to such ecosystem changes. Significant further research is yet necessary to bring our knowledge of the global biogeochemical role of marine biota to a level approaching our grasp of terrestrial ecosystems. In particular, our understanding of the relative importance of different microplankton taxa and functional groups in driving key biogeochemical cycles remains incomplete.

With microorganisms responsible for the overwhelming majority of marine primary production (Kirchman, 2012), oceanic ecosystems are heavily influenced by the complex web of ecological interactions taking place at the microscopic scale. Primary producers range in size from photosynthetic unicellular cyanobacteria to large diatoms hundreds of micrometers in scale (Falkowski et al., 2003). On the heterotrophic side, bacteria subsist on dissolved organic matter as free-living organisms or directly colonize the surfaces of larger algal cells and suspended inorganic and detrital particles (Buchan et al., 2014, Goecke et al., 2013, Grossart et al., 2005), while a vast array of larger protists graze upon bacteria, algae, and dead organic matter (Kirchman, 2012). Marine microbial ecologists have recently turned significant attention towards mixotrophs—versatile

microbial taxa capable of both autotrophy and heterotrophy (Burkholder et al., 2008). Additionally, a diverse set of autotrophic as well as heterotrophic bacteria support marine microbial production via nitrogen (N<sub>2</sub>) fixation, thereby increasing the supply of a key macronutrient for cellular growth (Delmont et al., 2018, Luo et al., 2012, Zehr and Kudela, 2011). The net sum of community metabolic activities in the surface ocean is substantial, with daily production and consumption of oxygen via photosynthesis and respiration driving large diel variations in biological oxygen concentrations in the mixed layer during the growing season (Ferron et al., 2015, Tortell et al., 2014, Hamme et al., 2012). Ultimately, this complex assemblage of microbial life serves as the key basal link of the marine food chain, sustaining zooplankton, fish, and higher-trophic-level fauna.

Marine ecosystems also exhibit considerable variation in terms of their physical and biochemical characteristics. To one end of the spectrum of productivity, the subtropical gyres of the earth's major oceans have traditionally been termed 'marine deserts' due to their low chlorophyll-a (Chl-a) concentrations and thermally-stratified, low-nutrient conditions throughout much of the growing season (Sarmiento and Gruber, 2006, Brix et al., 2006, Karl, 1999). On the other extreme, zones featuring upwelling of nutrient-rich deep waters as well as coastal regions in which nutrients are supplied through terrestrial runoff and riverine inputs often exhibit elevated productivity (Bakun, 1990, Bauer et al., 2013, Doney, 2010). Patterns of production are also subject to seasonal forcings, typically driven by wind and temperature-induced

changes in stratification that regularly recur from year to year (Carlson et al., 1994, Karl and Church, 2017).

Despite the considerable importance of such ecological dynamics to marine biogeochemical cycles, key aspects of the interactions between ecology and biogeochemical fluxes remain open topics of research. In particular, even the question of whether an overall relationship between microplankton diversity and ecosystem production exists and what shape that relationship takes has eluded satisfactory consensus (Cermeño et al., 2013, Vallina et al., 2014). By comparison, terrestrial ecologists enjoy a considerable lead over their marine counterparts, having attained broad agreement that land ecosystems often exhibit a curved or linear positive relationship between increasing primary productivity and vascular plant diversity (Hawkins et al., 2003, Mittelbach et al., 2001, Hooper et al., 2005).

At a finer-scale, motivated by the needs of the climate modeling community to better-resolve the influence of marine life upon climate feedbacks, oceanographic researchers are also seeking to investigate whether particular microplankton functional groups or even specific taxa are key to driving carbon uptake and sequestration in the global ocean (Guidi et al., 2016, Lin et al., 2017, Wang et al., 2018, Richardson and Jackson, 2007). Currently, scientific understanding of microplankton ecology and community production remains hampered by under-sampling. High-frequency, long-term monitoring of marine ecosystems remains rare and cost-intensive, while

oceanographic expeditions and sample collection are subject to spatial and seasonal sampling biases. Consequently, deepening our knowledge regarding interactions between microbial community structure and carbon and nutrient cycling requires continued efforts to infer statistical relationships from field data that can then be tested using more focused approaches.

Answering such questions carries particularly important implications in the context of a rapidly-changing biosphere. Marine ecosystems are being subjected to considerable changes at the hands of climate and other anthropogenic forcings (Bauer et al., 2013, Doney et al., 2007, Ruiz et al., 1997). With significant surface warming, ocean acidification, and intensification of density-driven stratification expected over the coming centuries (Polovina et al., 2008, Wanninkhof et al., 2015), understanding current relationships between marine microbial ecology and carbon uptake will be necessary to evaluate how biological carbon sequestration might evolve in a changing climate. At the same time, marine communities are subject to more than climate stressors alone.

Anthropogenic inputs of nutrients and contaminants into the coastal environment is only projected to increase (Doney, 2010), potentially intensifying coastal eutrophication, anoxia, and harmful algal bloom events. Given such considerations, problems at the interface of marine microbial ecology and marine biogeochemistry deserve redoubled attention and study.

In this dissertation, I employed high-resolution measurements of biologically-produced oxygen in the surface ocean in combination with techniques to determine microbial community composition. The overall goal underlying my thesis focused on assessing how shifting patterns of productivity and other microbially-mediated activities were related to changes in the abundances and diversity of the *in-situ* microplankton community. Major findings from my thesis work include the identification of key mixotrophs and other microplankton taxa of interest associated with sites showing high productivity and elevated N<sub>2</sub> fixation rates in the Western North Atlantic, the finding that these productivity and N<sub>2</sub> fixation-associated taxa play important roles in driving community co-occurrence patterns, and the uncovering of an overall negative eukaryotic productivity-diversity relationship within this region. Finally, I also present research investigating the influence of diel and multi-day non-steady-state metabolic activity upon *in-situ* methods for measuring marine net community production.

### **1.1 Estimation of Net Community Production from $\Delta O_2/Ar$**

Marine net community production (NCP) is defined as the difference between gross primary production (GPP), the total sum of carbon fixation by phytoplankton, and total community respiration (CR), which includes both autotrophic (AR) and heterotrophic (HR) respiration. Net community production reflects the overall net metabolic state of the surface ocean community, with positive rates of NCP indicating

net accumulation of biomass due to net autotrophic activity. Meanwhile, negative NCP rates signify net heterotrophy, in which total respiration of organic carbon exceeds the production of new carbon biomass by autotrophs. As net heterotrophy requires a source of organic carbon to sustain respiration, a net heterotrophic state cannot exist indefinitely without an external source of additional carbon.

NCP possesses great importance to both ocean carbon biogeochemistry as well as the regulation of global climate. Representing the remaining quantity of non-respired new production, NCP provides an estimate of the organic carbon available for sequestration into long-term deep ocean sinks (Dugdale and Goering, 1967, Li and Cassar, 2017, Berger and Wefer, 1990, Williams et al, 1989). Transport of organic matter into the abyssal deep via sinking plankton biomass, detrital particles, and dissolved carbon removes CO<sub>2</sub> from the active atmospheric pool, acting as an important climate feedback mechanism. Furthermore, an intense, unresolved debate has persisted regarding whether or not large portions of the global ocean are net autotrophic or net heterotrophic—a question of biogeochemical importance that NCP measurements are key to answering (Duarte et al., 2013, Williams et al., 2013, Ducklow and Doney, 2013).

Typically, NCP is measured using either carbon, nutrient, or oxygen-based methods, with measurements of NCP based on dissolved oxygen in particularly wide use (Hamme and Emerson, 2006, Juranek and Quay, 2013, Reuer et al., 2007). Dissolved oxygen concentrations, however, are subject to multiple influences in the surface ocean,

both physical and biological. While  $O_2$  is produced and consumed by photosynthesis and respiration, the concentration of  $O_2$  in the surface ocean can also increase or decrease due to changes in temperature, pressure, wind and wave-driven mixing, and other factors influencing its solubility in seawater. An innovative solution to this problem involves the measurement of the  $O_2/Ar$  ratio. As  $O_2$  and Ar have similar physical solubility characteristics, changes in the ratio of  $O_2/Ar$  can be assumed to reflect biological production and consumption of oxygen, as Ar is not involved in biological processes (Craig and Hayward, 1987).  $O_2/Ar$  is thus used to isolate changes in biological oxygen saturation from physical factors, allowing direct estimates of NCP. Within the past couple decades, several methods including equilibrator inlet mass spectrometry (EIMS) and membrane inlet mass spectrometry (MIMS) have been developed to measure dissolved  $O_2/Ar$  continuously while underway (Cassar et al., 2009, Tortell, 2005).

The dissolved  $O_2/Ar$  method is powerful, although subject to several caveats. Surface measurements of  $O_2/Ar$  assume that the system is at steady state, with biological production of  $O_2$  being fully balanced by efflux to the atmosphere. The  $O_2/Ar$  method also integrates over the mixed-layer only and therefore misses productivity occurring at deeper depths. Additionally, the  $O_2/Ar$  method is heavily influenced by the user's choice of wind speed parameterization, which can have a large impact on calculated rates (Wanninkhof, 1992, Wanninkhof, 2014, Reuer et al., 2007, Teeter et al., 2018, Bender et al., 2011). At the same time, precautions must be taken to account for vertical or

horizontal advection carrying water with a different O<sub>2</sub>/Ar signature into the sampled water volume. Even data from a single oceanographic voyage may be considerably impacted by whether spatial measurements are conducted in a Eulerian fashion (with respect to geographic coordinates) or in a Lagrangian manner (following selected parcels of water) and whether vertical fluxes of O<sub>2</sub>/Ar and/or diel variation in community metabolic rates is captured and accounted for.

Nevertheless, underway measurements of NCP using continuous measurements of surface seawater O<sub>2</sub>/Ar represent a versatile tool for studying ocean carbon biogeochemistry. The EIMS method allows for the collection of NCP data at unprecedentedly high resolution compared to discrete *in-situ* samples, while avoiding incubation artifacts and labor constraints associated with incubation-based methods. Particularly in combination with high-resolution sampling of physical and biological parameters, EIMS data are well-suited for closely examining relationships between NCP and ecosystem characteristics across diverse spatial scales.

## **1.2 Quantitative rRNA Amplicon Sequencing**

Nowhere else across the study of oceanography has technological and methodological progress been more rapid than for microbial oceanography. At the center of this striking trend, the high-throughput sequencing revolution has driven revolutionary progress in microbial community characterization, metagenomics, and metatranscriptomics. Such advances have made possible ambitious scientific efforts to

characterize the diversity of the world ocean at unprecedented depth and scale (de Vargas et al., 2015, Guidi et al., 2016, Lima-Mendez et al., 2015, Sunagawa et al., 2015).

Although just one of multiple possible approaches, ribosomal RNA (rRNA) amplicon sequencing has emerged as one of the most widespread techniques for taxonomic identification of marine microplankton from environmental samples. rRNA genes are key to cellular life across all domains of biology, with the rRNA molecules transcribed from these sequences making up critical structural components of ribosomes—the protein factories of the cell. Given their paramount importance, rRNA genes are highly-conserved, with some regions possessing nearly identical sequences across entire domains of life. Over evolutionary timescales, mutations are only evolutionarily permitted to accumulate in less-important sections of rRNA sequences. Thus, rRNA genes today consist of alternating conserved and “hypervariable” regions in which mutations have accumulated and been passed down to descendants. Sequencing of “hypervariable” regions yields taxonomic information that can be used to accurately determine a microorganism’s identity. For prokaryotic studies, sequencing of one or more regions of the 16S rRNA gene is often employed for this purpose. In eukaryotes, researchers use the 18S rRNA gene, although the inter-transcriptional spacer (ITS) sequence is a frequently-used alternative marker gene for taxonomic identification. Combined with modern high-throughput sequencing techniques, which can sequence environmental DNA to a depth of tens of millions of reads, rRNA amplicon sequencing

permits efficient characterization of microbial community structure for large sample sets.

An important consideration when interpreting amplicon sequencing results relates to the widespread use of relative abundances to report amplicon data. As sequenced samples vary in sequencing depth (the number of sequences produced for a given sample) due to amplification and sequencing effects, and as depth is not necessarily related to the quantity of starting DNA, the number of sequencing reads corresponding to a given taxon does not reflect the true quantitative abundance of that taxon in the environment. Consequently, researchers have opted to standardize amplicon data using relativization, reporting taxonomic composition in terms of percent abundances. Even this practice, however, carries drawbacks. Relative percentages are related not just to the abundance of reads corresponding to a given taxon in a sample but are also related to the abundances of other taxa within the sample. Consequently, changes in percent abundance may not reflect true changes in organismal or transcript abundance (Gifford et al., 2011). These considerations are particularly important for co-occurrence network analyses, as such compositional effects can produce spurious associations (Aitchison, 1981, Aitchison, 1982).

Within the last decade, however, investigators have introduced an internal standard-based technique to directly estimate quantitative abundances of natural sequences and thereby bypass issues associated with relative metrics (Gifford et al.,

2011, Satinsky et al., 2013). By introducing a known quantity of standard sequences with a distinctive taxonomic identity to all samples at the start of the DNA extraction and sequencing library preparation pipeline, the spike quantity can be compared to the recovery rate of standard sequences in the final data to correct for variations in sequencing depth. This approach has now been successfully implemented to study a wide range of environments ranging from the marine microbiome to soil environments, (Satinsky et al., 2013, Smets et al., 2016, Tkacz et al., 2018, Wang et al., 2018, Lin et al., 2019) and the technique compares favorably with independent organismal quantification methods (Smets et al., 2016, Lin et al., 2019). The method however remains somewhat limited by caveats associated both with rDNA amplicon sequencing in general as well as important considerations unique to the technique itself, including variable rDNA gene copy number across different taxa, inability to account for potential amplification biases, and the assumption that internal standard reads are recovered at an identical rate to natural sequences. Nevertheless, by removing compositional artifacts and enabling estimation of quantitative taxonomic abundances, quantitative amplicon sequencing allows oceanographers to make more accurate, direct comparisons of microbial community composition against biogeochemical rates and quantities of interest.

### **1.3 Research Questions**

Ultimately, my dissertation research seeks to investigate the nature of links between the marine microbial community and the oceanic cycling of carbon and nutrients using high-resolution biogeochemical rate measurements alongside a suite of new methods for bioinformatics, quantitative molecular sequencing, and community characterization. In particular, my work has sought to shed light upon the following guiding questions:

1. Are the quantitative abundances of specific microplankton taxa associated with patterns of NCP across the Western North Atlantic, and what do the characteristics of NCP-associated taxa tell us about their potential ecological roles?
2. How is the shifting microplankton community structure across the Western North Atlantic related to diazotrophy, productivity, and the changing availability of macronutrients and trace micronutrients?
3. What is the potential impact of diel variability in community metabolic rates, of other non-steady-state surface layer conditions, and of vertical fluxes of O<sub>2</sub>/Ar upon Lagrangian *in-situ* O<sub>2</sub>/Ar-derived measurements of NCP?

My use of quantitative amplicon sequencing for direct comparisons of taxonomic abundances to biogeochemical properties and for community co-occurrence network analyses represents a major step forwards for the field of microbial oceanography. My

results further present new approaches and insights regarding the ongoing study of marine microplankton productivity-diversity relationships. Additionally, my thesis work also represents the most comprehensive survey to date of regional underway NCP and prokaryotic and eukaryotic community structure in the summer Western North Atlantic. My findings highlight notable associations between NCP, N<sub>2</sub> fixation, and key microplankton taxa, providing valuable insights into community dynamics that drive major regional patterns of biogeochemical fluxes. Last but not least, I take advantage of a series of high-resolution Lagrangian NCP measurements to assess the strengths and weaknesses of key assumptions and sources of error involved in NCP calculations from O<sub>2</sub>/Ar.

## **2. Linking patterns of net community production and marine microbial community structure in the western North Atlantic**

This chapter has been published as:

Wang S., Lin Y., Gifford S., Eveleth R., and Cassar N. (2018). Linking patterns of net community production and marine microbial community structure in the western North Atlantic. *The ISME Journal*, 12:2582–2595.

### **Author contributions**

SW, NC, and YL conceived study. YL collected microplankton DNA samples. RE collected O<sub>2</sub>/Ar data. RE and SW processed O<sub>2</sub>/Ar data. SW performed molecular laboratory work. SW, SG, and YL analyzed and interpreted resulting sequencing data. All co-authors assisted in drafting this manuscript and gave their approval for the final version of the manuscript.

### **2.1 Introduction**

Uptake of carbon by phytoplankton and its exchange between organisms in the marine environment plays a critical role in the carbon cycle, with primary production in the world's oceans representing half of global net primary production (Field et al., 1998). A small proportion of surface production (Buesseler, 1998) is transported to depth via

sinking particles, subduction, and other processes, transferring carbon to deep ocean pools with a residence time of millennia or longer (Williams and Follows, 2011).

Considerable interest has consequently focused on exploring relationships between surface microbial community structure, marine production (Vallina et al., 2014, Cassar et al., 2015, Lin et al., 2017), and particulate carbon export (Boyd and Newton, 1995, Richardson and Jackson, 2007, Guidi et al., 2016).

Net community production (NCP) rates reflect the productivity and metabolic balance of the surface ocean microbial community. Expressed as the difference between gross primary production and community respiration, NCP rates estimate the mixed-layer production of organic carbon available for export (Dugdale and Goering, 1967, Williams et al., 1989, Berger and Wefer, 1990, Li and Cassar, 2017). NCP patterns have been well-examined independently, as have patterns of surface ocean community structure. However, direct comparison of relationships between ecology and productivity remains an emerging line of investigation.

The Western North Atlantic is a region of interest for unraveling potential links between community structure and productivity. New production across this region is thought to be driven by a variety of physical and biological processes including nitrogen fixation, mesoscale features, seasonal mixing, and allochthonous nutrient inputs (McGillicuddy et al., 1998, Lipschultz et al., 2002, Fawcett et al., 2014).

A dominant feature of the Western North Atlantic is the Sargasso Sea, an oligotrophic region typical of other subtropical gyre systems (Steinberg et al., 2001). While spring and winter phytoplankton blooms occur following winter mixing of nutrients into the surface layer, the Sargasso Sea in summer exhibits limiting nitrate and phosphate concentrations (N: <50 nmol kg<sup>-1</sup>, P: <20 nmol kg<sup>-1</sup>) (Lomas et al., 2013). Ongoing changes in the biogeochemistry of the Sargasso Sea may impact community composition, carbon export, and nutrient cycling due to increasing stratification (Krause et al., 2009) and changing nutrient inputs (Lomas et al., 2010). Records suggest gradual community shifts are underway, with haptophyte populations declining and *Synechococcus* and dinoflagellate groups increasing in abundance (Lomas et al., 2010, Bates et al., 2012). Globally, oligotrophic subtropical gyres cover some 40% of the planet's surface (Karl and Church, 2014), and small shifts in microplankton ecology in such regions may have repercussions for biogeochemistry and climate. Large-scale genomics sampling work has suggested that specific key taxa may be important drivers of carbon export in such regions (Guidi et al., 2016).

To the west, the Western North Atlantic is bounded by the North American continental shelf. High rates of production are observed along this coast well into summer (O'Reilly and Busch, 1984). In this region, the shelf, shelf break, shelf slope, and Gulf Stream exert dynamic physical forcings upon resident microplankton, driving variation in community structure and primary production over short transects (Mouw

and Yoder, 2005). Such coastal regions are increasingly being recognized as potentially important carbon sinks (Bauer et al., 2013) and are also predicted to undergo future ecological shifts in response to eutrophication and climate change (Lomas et al., 2002, Doney, 2010, Cai, 2011).

Considering ongoing shifts in microplankton community structure in ecosystems across the Western North Atlantic, evaluating the impact of future community shifts upon primary production and potential carbon export in this region is of great interest. There is thus a need to identify relationships between community composition and NCP. Few regional NCP measurements have been conducted in the Western N. Atlantic to date, with existing NCP data generally coming from time-series measurements (Brix et al., 2006) or fine-scale studies (Mourino-Carballido and McGillicuddy, 2006, Estapa et al., 2015). Similarly, while community structure at the Bermuda Atlantic Time Series (BATS) has been regularly studied (Treusch et al., 2012, Vergin et al., 2013), broader rDNA amplicon data surveying the whole region are far sparser.

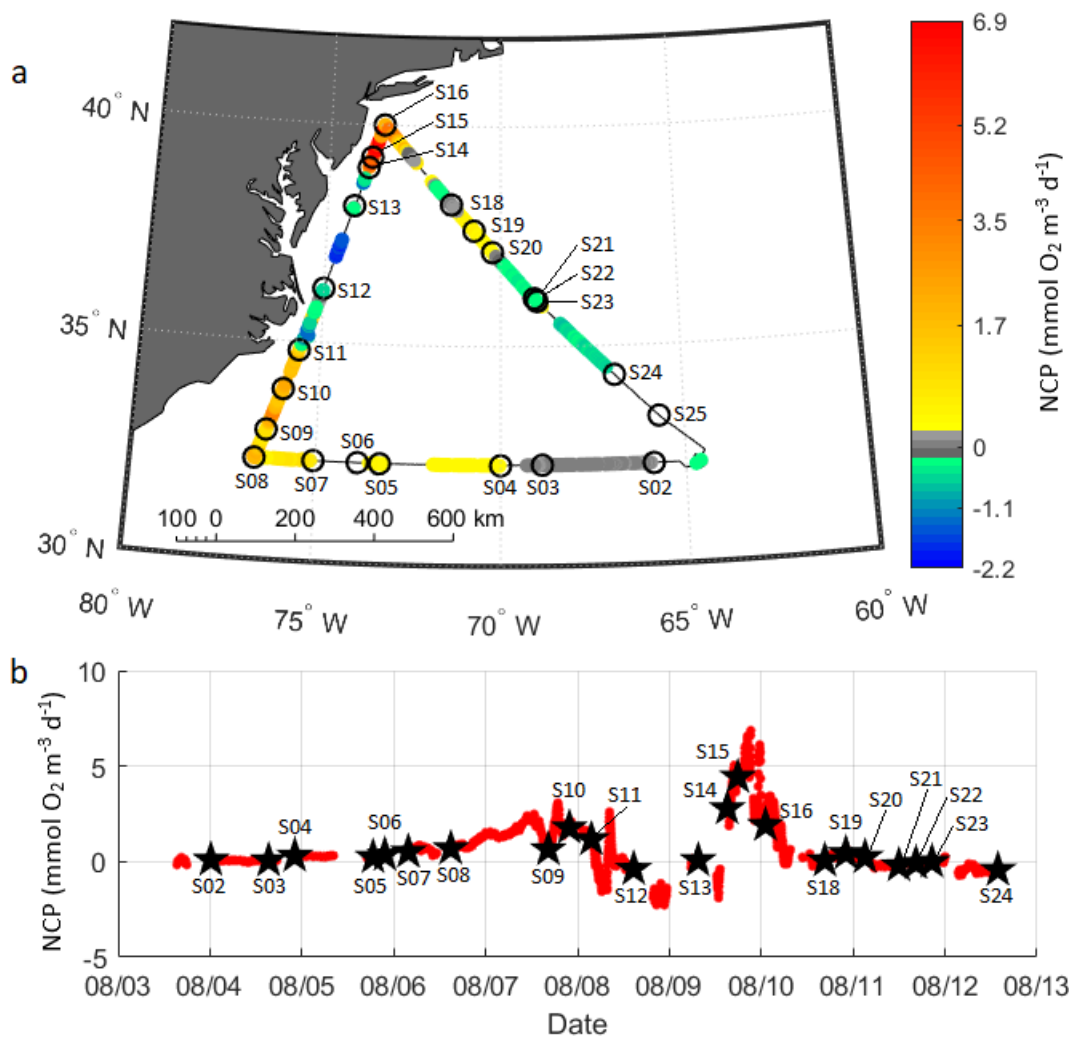
In this study, we gathered samples for high-throughput 16S and 18S rDNA amplicon sequencing and concurrently conducted high-resolution O<sub>2</sub>/Ar-based NCP measurements using Equilibrator Inlet Mass Spectrometry (EIMS) (Cassar et al., 2009) along three transects spanning the oligotrophic Sargasso Sea, the Gulf Stream, and the U.S. East Coast. To obtain absolute taxonomic abundances for the sampled communities, we adapted an internal standard approach for 16 and 18S rDNA sequencing to

quantitatively characterize community structure (Lin et al., 2019). We then assessed trends in whole-community composition and diversity in relation to NCP and evaluated associations between productivity and specific microplankton groups identified in our samples.

## **2.2 Materials and Methods**

### **Study area and collection of O<sub>2</sub>/Ar and ancillary data**

Continuous and discrete measurements were collected over a 3 100 km transect in the western North Atlantic aboard the *R/V Atlantic Explorer* from August 3-12, 2015. The cruise track progressed west from the BATS Station (32.3°N, -64.6°W) to the North Carolina coast, then northeast to approximately 50 km south of Long Island, New York before returning to Bermuda (Fig. 1). Fourteen CTD casts were conducted during the cruise at 200-400 km intervals. Underway dissolved O<sub>2</sub>/Ar measurements were collected alongside discrete sampling for chlorophyll and DNA. O<sub>2</sub>/Ar was measured continuously from the ship's underway intake using the EIMS method (Cassar et al., 2009). Details of O<sub>2</sub>/Ar-derived NCP calculations and assessment of potential vertical O<sub>2</sub>/Ar fluxes are described in Appendix A.



**Figure 1:** (a) Map of O<sub>2</sub>/Ar-derived volumetric NCP measured using EIMS in early August 2015 across the western Sargasso Sea. Positive rates are shown using a warm color scale, while net negative measurements are shown with a cool color scale. Grey values indicate balanced NCP rates. Locations at which samples were collected for community rDNA sequencing are indicated with open circles. (b) Time series of O<sub>2</sub>/Ar-derived NCP measurements, with molecular sampling stations indicated by black stars.

## **Microbial community sampling and rDNA amplicon sequencing**

Samples for rDNA analysis were obtained from 5 m CTD casts and underway samples (Table S2) pumped from a towfish trailing abeam of the vessel at 3-5 m depth. This custom-built towfish, suspended alongside, is trace metal-clean, using plastic tubing and carrying seawater aboard via an air-driven pump. For each sample, one liter was filtered through a 0.22- $\mu$ m filter (Millipore, Billerica, MA, USA) using a peristaltic pump, preserved with *RNAlater* (Thermo Fisher, Waltham, MA, USA), and flash-frozen in liquid nitrogen. At stations with high biomass, the volume of filtrate was reduced to 0.2-0.5 l as filters became clogged.

## **Internal controls for quantitative sequencing**

A quantitative internal standard approach provides information on per-liter abundance of taxa across samples, yielding more meaningful comparisons between taxonomic abundances and biological rate measurements. To quantify rDNA copy numbers  $l^{-1}$ , internal genomic standards were added to each sample following (Lin et al., 2019). Genomic DNA was obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA) for *Thermus Thermophilus* (ATCC #27634D-5), a thermophilic hot springs bacterium, and *Schizosaccharomyces pombe* (ATCC #24843D-5), a yeast species. The *S. pombe* genome contains approx. 110 copies of the 18S V4 rDNA amplicon (Lin et

al., 2019), while the *T. thermophilus* genome contains two 16S V4 copies (Smets et al., 2016).

Given the large range in 18S rDNA copy number across eukaryotic genomes, we determined an appropriate spike of control DNA (0.073 ng) by evaluating the average 18S rDNA concentration in our samples using qPCR with 18S V4 primers. To ensure that such a diluted spike would reliably manifest in sequencing output, we conducted a pilot sequencing run on duplicate filters from this study at the Boston University Microarray Core on an Ion Torrent PGM using a 314 chip (Appendix A).

The Ion Torrent test revealed that a 15.2 ng *T. thermophilus* genomic DNA spike resulted in *T. thermophilus* reads comprising an average of 5.3% of all reads, while the addition of 0.679 ng of *S. pombe* gDNA yielded 0.9% *S. pombe* reads. Based upon these results, we adjusted the spike amounts to quantities expected to constitute <1% of sequenced reads, adding 0.679 ng of the *S. pombe* standard and 3.04 ng of the *T. thermophilus* standard to each sample, both in 50 ul volumes. This corresponded to adding c.a. 5 780 000 rDNA copies sample<sup>-1</sup> of *S. pombe* and 2 800 000 rDNA copies sample<sup>-1</sup> of *T. thermophilus* genomic DNA.

## **DNA extraction for 16S and 18S rDNA sequencing**

We conducted DNA extraction using the Qiagen DNeasy Plant Mini Kit (Qiagen, Germantown, MD, USA) following manufacturer instructions with slight modifications

(Moisander et al., 2008), with internal gDNA standards added prior to bead-beating (Satinsky et al., 2013). PCR amplification was performed for 30 cycles using custom 16S V4 primers 515F-Y (5'-GTGYCAGCMGCCGCGGTAA-3') and 805R (5'-GACTACNVGGGTATCTAAT-3') and 18S V4 primers F (5'-CCAGCASCYGCGGTAATTCC-3') and R (5'-ACTTTCGTTCTTGAT-3'), with attached Illumina adapters and barcodes (Supplementary Table 3). These primers are adapted from widely-used universal primers for the amplification of marine prokaryotic (Caporaso et al., 2011, Walters et al., 2011) and eukaryotic (Stoeck et al., 2010) taxa, modified to improve coverage of SAR11 and haptophytes (Bradley et al., 2016, Lin et al., 2017). Primers were each dual-indexed with 6 bp barcodes, using a heterogeneity spacer approach (Kozich et al., 2013, Fadrosch et al., 2014). 16S samples were run at 94°C for 3 min, 30 cycles at 94°C for 30 s, 60°C for 30 s, 72°C for 1 min, followed by a third stage at 72°C for 10 min. 18S samples were run identically apart from an annealing temperature of 57°C. Each 16S PCR reaction (25 µl volume) consisted of 2.5 µl 10x PCR buffer, 0.5 µl dNTP mix (10 mM each), 1 µl 50 mM MgSO<sub>4</sub>, 0.5 µl each of forward and reverse primer (10 µM), 0.1 µl Platinum Taq Hi-Fidelity Polymerase (Thermo Fisher, Waltham, MA, USA), and 19.4 µl of sterile water. 18S PCR reaction mixtures were identical except polymerase amounts were doubled (0.2 µl per reaction) to address weak amplification, with a compensating water volume decrease to 19.3 µl. PCR products were purified using the Qiagen QIAquick PCR Purification Kit and quantified using a Qubit 3.0

fluorometer (Life Technologies, Carlsbad, CA, USA). The samples were then pooled at equimolar concentrations and sequenced using the Illumina MiSeq platform (300bp PE, V3 chemistry) at the Duke Center for Genomic and Computational Biology.

## **Analysis Pipeline**

We obtained 20 450 700 single-end reads from our 25 sequenced samples. Raw single-end reads were trimmed to remove barcodes, assembled, and quality filtered following (Fadrosh et al., 2014) using pandaseq (Masella et al., 2012). 16S amplicon length was 296.7 +/- 2.9 bp (mean +/- sd), while mean 18S amplicon length was 424.1 +/- 4.3 bp. Demultiplexing was performed in QIIME (Caporaso et al., 2010b). Five 18S rDNA samples and one 16S rDNA sample contained no reads, the former likely due to a defective forward primer. Primer and other non-biological sequences were subsequently removed using Tagcleaner (Schmieder et al., 2010). We conducted chimera detection and open-reference OTU picking at 97% similarity using the Usearch 6.1 algorithm (Edgar, 2010, Edgar et al., 2011) and Release 123.1 of the SILVA database (Pruesse et al., 2007). OTU clustering was performed using the usearch61 method for de novo OTU picking, and the usearch61\_ref method for reference-based OTU picking. Alignment was performed using PyNAST (Caporaso et al., 2010a) and taxonomy assignment conducted using the RDP classifier 2.2 (Wang et al., 2007). Full sequence processing scripts are included in the Supplementary Material. Following taxonomy assignment, internal

standard DNA sequences, eukaryotic metazoans, and plastid 16S sequences were filtered out using the QIIME script 'filter\_taxa\_from\_otu\_table.py'. We further discarded one sample due to the low volume of filtrate, leaving 19 eukaryotic and 23 prokaryotic samples.

Sample diversity metrics were calculated for 16S and 18S datasets using the phyloseq package (McMurdie and Holmes, 2013) for R 3.4.1 (R Core Team, 2017). For alpha diversity analyses only, sample libraries were rarefied to the smallest library size in each set of samples (16S: 98 819; 18S: 33 245). Rarefaction curves begin to level off at the sequencing depths obtained, suggesting that depth was sufficient to represent major patterns of diversity in our samples (Fig. 25). Alpha diversity metrics (observed OTUs and Shannon diversity) were calculated using averages from five rarefactions.

Using our non-rarefied sample libraries, calculation of absolute abundances for each OTU was performed following (Satinsky et al., 2013):

$$rDNA \text{ abundance } l^{-1} = \frac{\# \text{ of OTU reads}}{R * V} \quad (2-1)$$

where V is the volume filtered and R represents the recovery ratio of internal standards (genomic standards sequenced / molecules of genomic standard added). Output OTU tables are included in the Supplementary Material (Supplementary Tables 5a, 5b)

Further details of downstream statistical analyses including ordination and partial least squares regression are described in Appendix A.

## **2.3 Results and Discussion**

### **Patterns of O<sub>2</sub>/Ar-derived NCP**

Underway O<sub>2</sub>/Ar-derived biological oxygen fluxes within the mixed layer ranged from -2.2 to 6.9 mmol O<sub>2</sub> m<sup>-3</sup> day<sup>-1</sup> (MLD-integrated rates of -24 to 163 mmol O<sub>2</sub> m<sup>-2</sup> day<sup>-1</sup>) (Fig. 1). We observed initial rates below 0.5 mmol O<sub>2</sub> m<sup>-3</sup> day<sup>-1</sup> in the open ocean, increasing to 1 mmol O<sub>2</sub> m<sup>-3</sup> day<sup>-1</sup> within 400 km of the coast. Turning north, fluxes reached 2 - 4 mmol O<sub>2</sub> m<sup>-3</sup> day<sup>-1</sup> along the Carolina coast. Values were subsequently variable along the coast.

The highest O<sub>2</sub>/Ar supersaturation occurred at the expedition's northernmost extent within a productive phytoplankton bloom, with values peaking at 6.9 mmol O<sub>2</sub> m<sup>-3</sup> day<sup>-1</sup> south of Long Island. Passing this bloom, O<sub>2</sub>/Ar supersaturation declined again to typically below 1 mmol O<sub>2</sub> m<sup>-3</sup> day<sup>-1</sup> during transit back to Bermuda.

We assessed the potential contribution of eddy diffusive and entrainment fluxes to mixed-layer O<sub>2</sub>/Ar values as minimal (Appendix A). Consequently, we report all biological O<sub>2</sub> fluxes as NCP rates henceforth. Except when comparing our data with integrated figures from other literature, we also report rates throughout this manuscript as volumetric values, more suitable for relation to quantitative taxonomic abundances.

Overall, our high-resolution NCP measurements agree well with previously measured patterns, with low NCP rates observed in the open ocean and higher values over the continental shelf along the Mid-Atlantic Bight. The marked peak in

productivity at the northern end of the expedition coincided with high measured nitrogen fixation rates (Tang et al., 2019) and high Chl a. Peak MLD-integrated productivity, reaching  $163 \text{ mmol O}_2 \text{ m}^{-2} \text{ day}^{-1}$  ( $116 \text{ mmol C m}^{-2} \text{ day}^{-1}$  assuming a photosynthetic quotient of 1.4 (Laws, 1991), is of a similar magnitude as integrated  $^{14}\text{C}$ -derived primary production rates for the Mid-Atlantic Bight spring bloom of up to  $158 \text{ mmol C m}^{-2} \text{ day}^{-1}$  (Falkowski et al., 1988). Our observed rates are also comparable to summer peak photic-zone primary production of between  $145 - 190 \text{ mmol C m}^{-2} \text{ day}^{-1}$  modeled for the same area using profile observations (Mouw and Yoder, 2005).

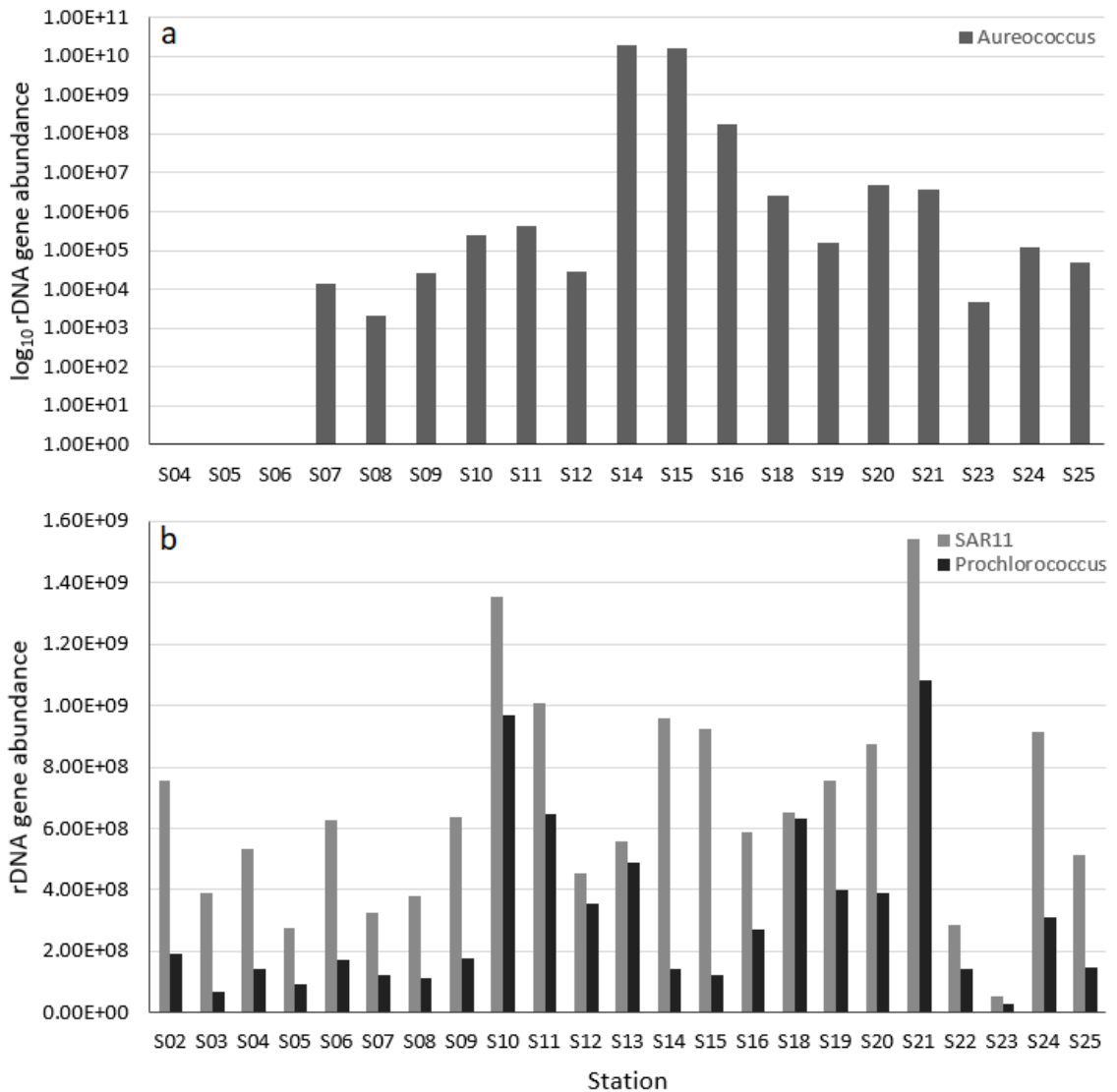
Our low MLD-integrated open-ocean NCP rates, with a mean of  $2.4 \text{ mmol O}_2 \text{ m}^{-2} \text{ day}^{-1}$ , are also consistent with prior Sargasso Sea  $\text{O}_2/\text{Ar}$ -based estimates in September/October of  $1.1 - 3.4 \text{ mmol O}_2 \text{ m}^{-2} \text{ day}^{-1}$  (Estapa et al., 2015), as well as modeled summer regional NCP values of  $3 - 4 \text{ mmol O}_2 \text{ m}^{-2} \text{ day}^{-1}$  (Letscher and Moore, 2017).

### **Microbial community quantitative and relative abundance patterns**

Analysis of rDNA reads yielded 7 843 eukaryotic and 5 604 prokaryotic OTUs across 19 eukaryotic and 23 prokaryotic samples (Supplementary Table 6). 16S and 18S samples contained at least 98 819 and 33 245 reads per sample.

Our observations of 16S and 18S rDNA abundances per liter were within expected bounds. Bacterial 16S rDNA abundances of  $1.78 \times 10^8 - 5.4 \times 10^9 \text{ copies l}^{-1}$  are

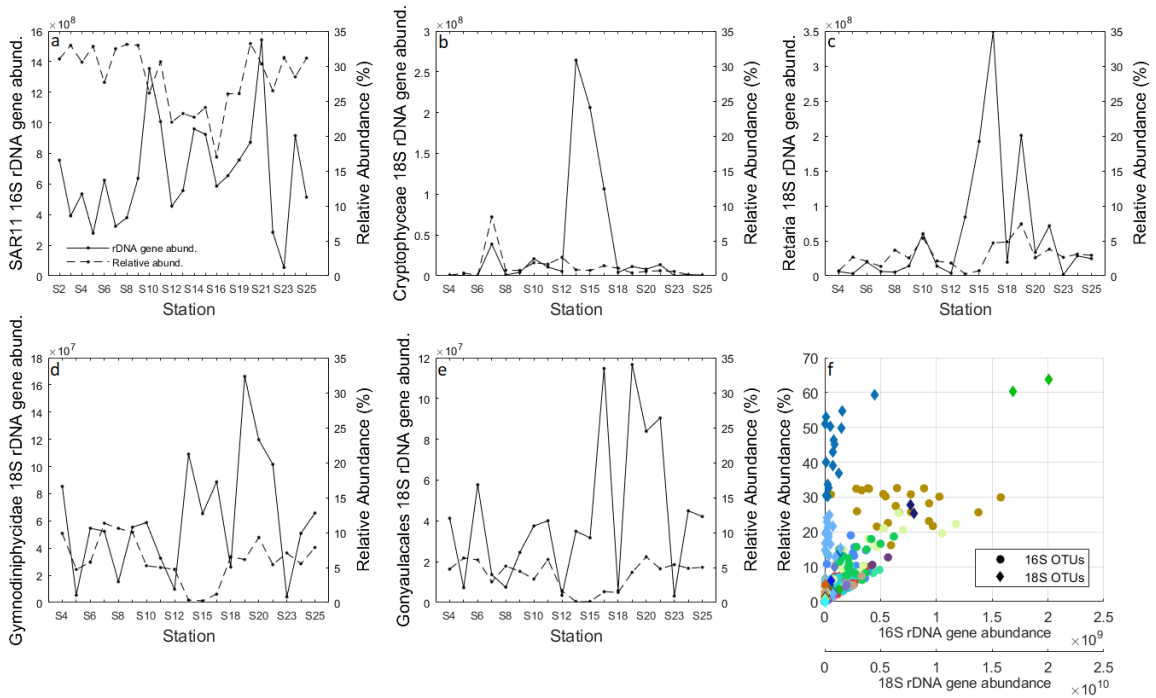
consistent with bacterial abundances in the Sargasso Sea and Western North Atlantic of  $4.0 \times 10^8 - 2.3 \times 10^9$  cells  $l^{-1}$  (Malmstrom et al., 2004, Rowe et al., 2012), assuming a typical 16S copy number of 1-15 (Větrovský and Baldrian, 2013). Excluding the three highest NCP stations, where the highest 18S rDNA abundances were observed (range of  $1.43 \times 10^8 - 3.14 \times 10^{10}$  18S rDNA genes  $l^{-1}$ ), the median 18S rDNA abundance was  $1.4 \times 10^9$  sequences  $l^{-1}$ . This is high compared with surface ocean eukaryotic cell densities of  $1 \times 10^7$  protists  $l^{-1}$  and  $1 \times 10^6$  phytoplankton  $l^{-1}$  (Kirchman, 2012), but is likely driven by variation in 18S rDNA copy number. Peak 18S rDNA abundances, while high, are also reasonable. *Phaeocystis* blooms can reach cell counts of  $1.5 \times 10^8$  cells  $l^{-1}$  (Rutten et al., 2005), and *Aureococcus* blooms of  $6 \times 10^8$  cells  $l^{-1}$  have been observed along the Long Island Coast (Gobler et al., 2002).



**Figure 2: rDNA abundances ( $\log_{10}$ ) across sampled stations for (a) *Aureococcus anophagefferens* (b) SAR11 and *Prochlorococcus*.**

Absolute abundances of individual taxa are also consistent with previous observations. For example, the median SAR11 16S rDNA abundance in our samples (Fig. 2b) was  $6.2 \times 10^8$  rDNA genes  $l^{-1}$  (SAR11 contains one 16S gene copy  $cell^{-1}$ ), compared with previous measurements of  $2 \times 10^8$  SAR11 cells  $l^{-1}$  in the Sargasso Sea from

fluorescence *in-situ* hybridization counts (Morris et al., 2002). Similarly, we observed a median of  $1.9 \times 10^8$  *Prochlorococcus* 16S rDNA genes l<sup>-1</sup> in our samples (Fig. 2b), consistent with Western North Atlantic observations of  $1 \times 10^8$  cells l<sup>-1</sup> based on qPCR quantification and flow cytometry (Johnson et al., 2006, Zinser et al., 2006). Applications of the internal standard approach for samples collected in the lower Amazon River, the Southern Ocean, as well as in soil samples have also demonstrated good correspondence between the standard-derived abundances and complementary abundance data measured using epifluorescence microscopy, photosynthetic pigments, flow cytometry, phospholipid fatty acid analysis, and substrate-induced respiration approaches (Satinsky et al., 2015, Smets et al., 2016, Lin et al., 2019).

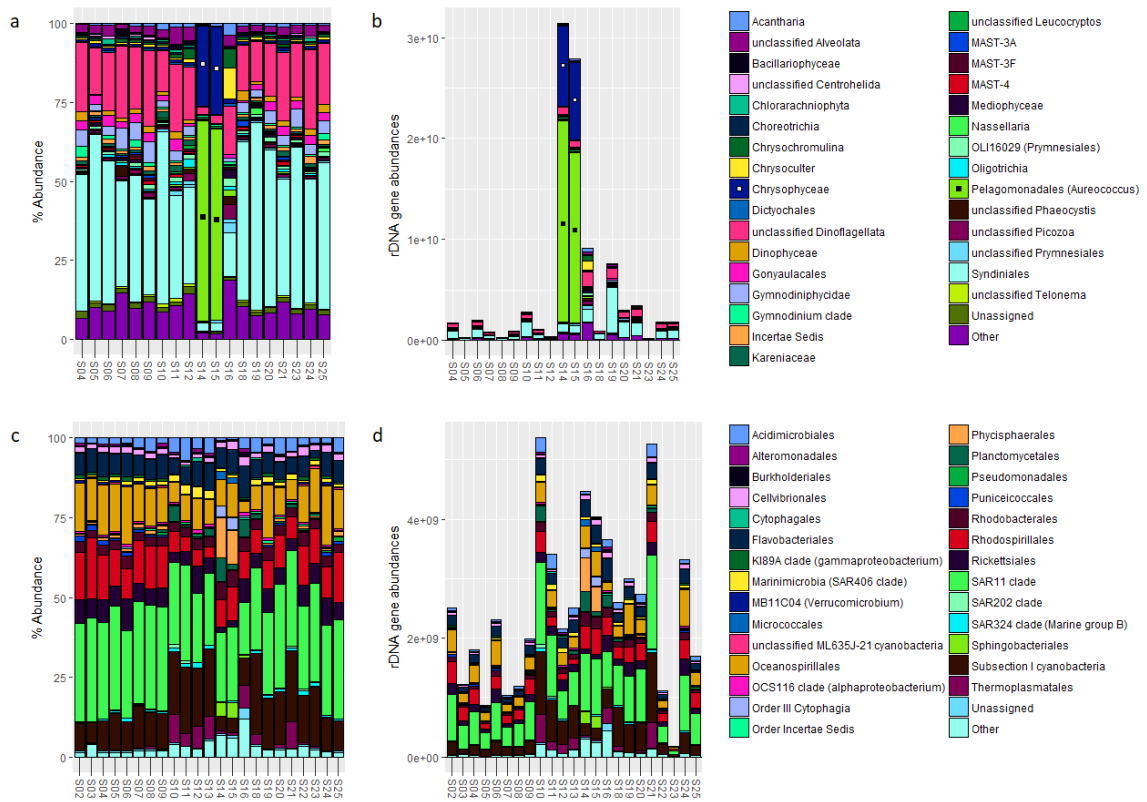


**Figure 3: (a) Comparison of patterns of absolute abundance, calculated using an internal standard approach, to relative abundance for the SAR11 clade of alphaproteobacteria as well as four groups of eukaryotic protists: (b) *Cryptophyceae*, (c) *Retaria*, (d) *Gymnodiniophycidae*, and (e) *Gonyaulacales*. (f) Scatter plot of 16S and 18S rDNA absolute vs. relative abundance data, with points colored for each prokaryotic (circles) and eukaryotic (diamonds) taxon across all samples, binned at the 5th taxonomic rank.**

Notably, calculation of absolute taxonomic abundances using internal standards produces patterns distinct from those generated using relative abundance metrics (Fig. 3). This is evident among several abundant 18S and 16S OTUs, including SAR11 clade members as well as the protist clades *Dinoflagellata*, *Gonyaulacales*, *Alveolata*, and *Gymnodiniophycidae*. The latter four eukaryotes increase in absolute abundance within the bloom environment, while their relative abundances decrease due to the dominance of *Chrysothrix* and *Aureococcus anophagefferens* within these samples. A similar

phenomenon affects SAR11 relative abundances, which are highest between S2-S9 and S20-S25 due to lower 16S rDNA counts for other prokaryotes at those stations. These discrepancies highlight longstanding criticisms of traditionally-used relative abundance metrics (Farrelly et al., 1995, Barber and Hiscock, 2006, Gifford et al., 2011, Moran et al., 2013, Satinsky et al., 2013, Smets et al., 2016) and illustrate advantages offered by the internal standard approach. In addition, avoidance of issues caused by compositional community data (Aitchison, 1981, Aitchison, 1982) is valuable when relating taxonomic abundances to microbial or biogeochemical processes like NCP.

The internal standard approach is nonetheless subject to several assumptions and limitations. A key assumption is that recovery rates of DNA standards are comparable to those of natural sequences within the sample. Particularly given the general implications of primer biases in amplicon work, this premise warrants further investigation. Another important limitation is that quantitative abundance data produced by this method remain sensitive to differences in rDNA copy number across taxa. Although better knowledge of 16S copy number variation across prokaryotes has spurred efforts to correct for copy number differences (Kembel et al., 2012), existing datasets remain limited particularly for eukaryotes, in which rDNA copy number may vary by multiple orders of magnitude. As data collection continues, corrections will likely become more feasible and commonplace.



**Figure 4:** Bar plots of (a) 18S eukaryotic taxonomy shown at the 5th taxonomic rank using relative abundance and (b) rDNA gene abundance, and (c) 16S prokaryotic taxonomy shown at the 4th taxonomic rank using relative abundance and (d) rDNA abundance. Among eukaryotes, dinoflagellate lineages dominated all samples except three from the coastal bloom (S14-S16) (Fig. 2-4a, 2-4b). Most of these dinoflagellate sequences corresponded to *Syndiniales*, alveolate parasites infecting various marine organisms and often detected at high abundances using molecular tools (Romari and Vaultot, 2004, Not et al., 2007, Guillou et al., 2008). While many of these sequences may originate from endosymbionts inside metazoan zooplankton caught on our filters, *Syndiniales* also infect microzooplankton protists, including ciliates, cercozoa, and other dinoflagellates, and clades targeting both host categories often exhibit a short free-living

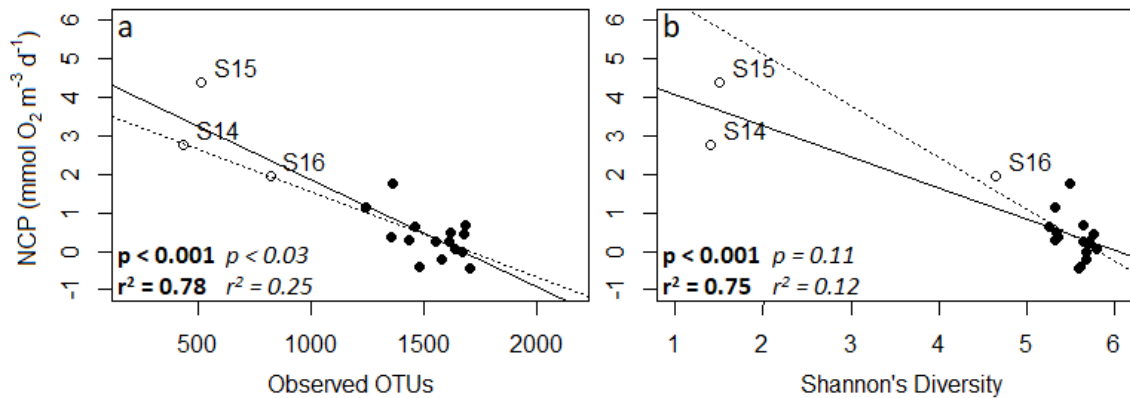
life stage (Guillou et al., 2008). Consequently, these sequences may also represent organisms living outside of metazoan hosts, interacting within the marine microbial environment. To a degree, elevated dinoflagellate abundances observed may also reflect high 18S copy numbers, driven by large dinoflagellate genomes (Prokopowich et al., 2003, Figueroa et al., 2014).

Two samples (S14, S15) associated with the coastal bloom were dominated (>90% relative abundance) by *Aureococcus anophagefferens*, a pelagophyte that forms coastal “brown tide” harmful algal blooms (HABs) (Gobler et al., 2005), as well as *Chrysophyceae* (Fig. 4a, 4b). qPCR surveys have also detected *A. anophagefferens* at low abundances in pelagic waters, which some suggest indicates an oceanic origin for this nuisance algae (Popels et al., 2003). A wide distribution of *A. anophagefferens* is also supported by our study. We found *Aureococcus* present in 16 of 19 18S rDNA samples, with a mean of  $7.5 \times 10^4$  *Aureococcus* 18S rDNA genes l<sup>-1</sup> observed in non-bloom samples. We estimated abundances of  $4.4 - 6.6 \times 10^4$  18S rDNA genes l<sup>-1</sup> in open-ocean samples (S24, S25) collected near Bermuda. In comparison, estimated *Aureococcus* 18S rDNA gene abundances ranged between  $1.8 \times 10^8$  and  $2.0 \times 10^{10}$  rDNA genes l<sup>-1</sup> within the observed bloom (Fig. 2a).

Sample 16 featured a high population (~20%) of *Prymnesiales*, primarily *Chrysochromulina* and *Chrysoculter*. *Chrysochromulina* are another nuisance algae, capable of mixotrophy (Jones et al., 1993) and forming blooms that can cause fish kills

(Richardson, 1997). Other members of *Prymnesiales* produce harmful hemolytic compounds (Graneli et al., 2012). Eukaryotic diversity was lower at two bloom stations, S14 and S15, (Appendix A, Fig. 26) but was similar across our other samples.

Among bacterioplankton, SAR11, SAR86 clade members (appearing as Oceanospirillales in Fig. 4), and *Prochlorococcus* (Subsection I cyanobacteria) dominated the communities sampled (Fig. 4c, 4d). The *AEGEAN-169* clade of *Alphaproteobacteria* (Rhodospirillales) as well as MGII *Archaea* (Thermoplasmatales) also appeared at high proportional abundances. Within the northern bloom, we observed elevated abundances of *Planctomycetales*, *Flavobacteria*, *Sphingobacteriales*, and Order III *Cytophagia*, with *Phycisphaerales* appearing at particularly high abundances (>10%) at two stations. Not much is currently known about *Phycisphaerales*, although they are hypothesized to form associations with macroalgae, with many representatives facultatively anaerobic (Lage and Bondoso, 2014). In addition, these bloom samples also appear to contain more sequences belonging to less-abundant and “rare” taxa (labeled ‘Other’ in Fig. 4). This phenomenon of elevated abundances of “rare” taxa in bloom events has also been reported elsewhere and may be related to ecological associations with phytoplankton (Gilbert et al., 2012, Hatosy et al., 2013). Bacterial diversity across samples was more uniform than eukaryotic diversity, with prokaryotic Shannon diversity between 4.1 - 4.5 versus 2.4 – 5.7 for eukaryotic samples (Appendix A, Fig. 26).

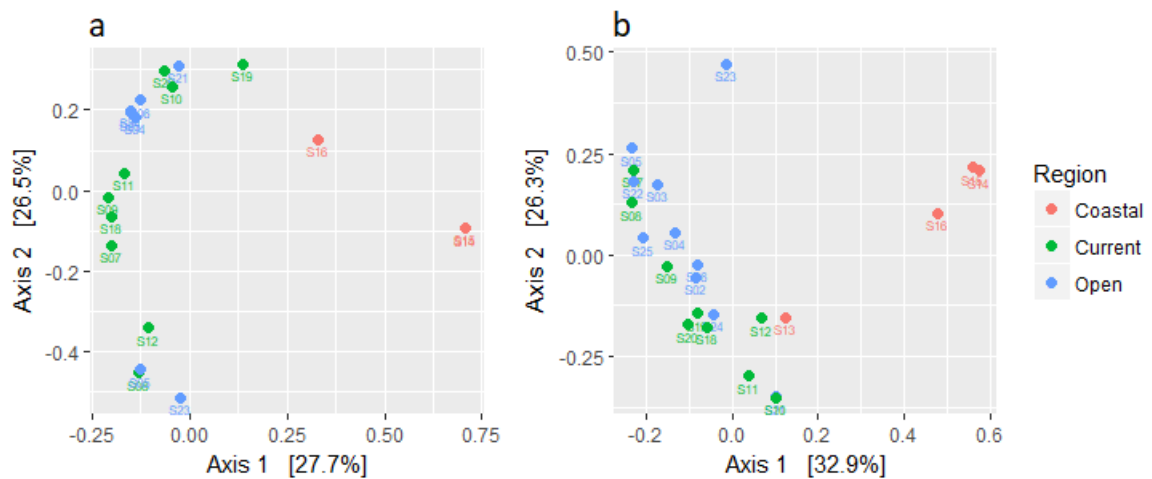


**Figure 5: (a) Linear regression of NCP vs. the number of observed 18S eukaryotic OTUs per sample. (b) Linear regression of NCP vs. 18S Shannon alpha diversity across sample libraries. Dashed lines indicate the resulting regressions when bloom stations (open circles) are excluded. Significance and goodness-of-fit are provided in boldface for the full dataset, and in italics for the dataset with bloom station data excluded. Relationships between microbial community structure and NCP**

At the community level, we observed a negative relationship between measured NCP and eukaryotic Shannon's H diversity (Pearson:  $-0.87$ , Spearman:  $-0.68$ ,  $p \ll 0.01$  for both) (Fig. 5), which was strongly driven by low diversity at two highly productive stations. This relationship however remains significant with those samples excluded (Pearson:  $-0.67$ ,  $p < 0.01$ ; Spearman:  $-0.54$ ,  $p < 0.05$ ). We observed no relationship between prokaryotic diversity and NCP.

Recent debate over the nature of the relationship between marine microplankton diversity and productivity has been energetic. Any overall relationship between community diversity and productivity would reflect the relative importance of functional diversity, cooperation, competitive exclusion, selective feeding by grazers, and other factors in governing ecosystem production (Cermeño et al., 2013, Goebel et al.,

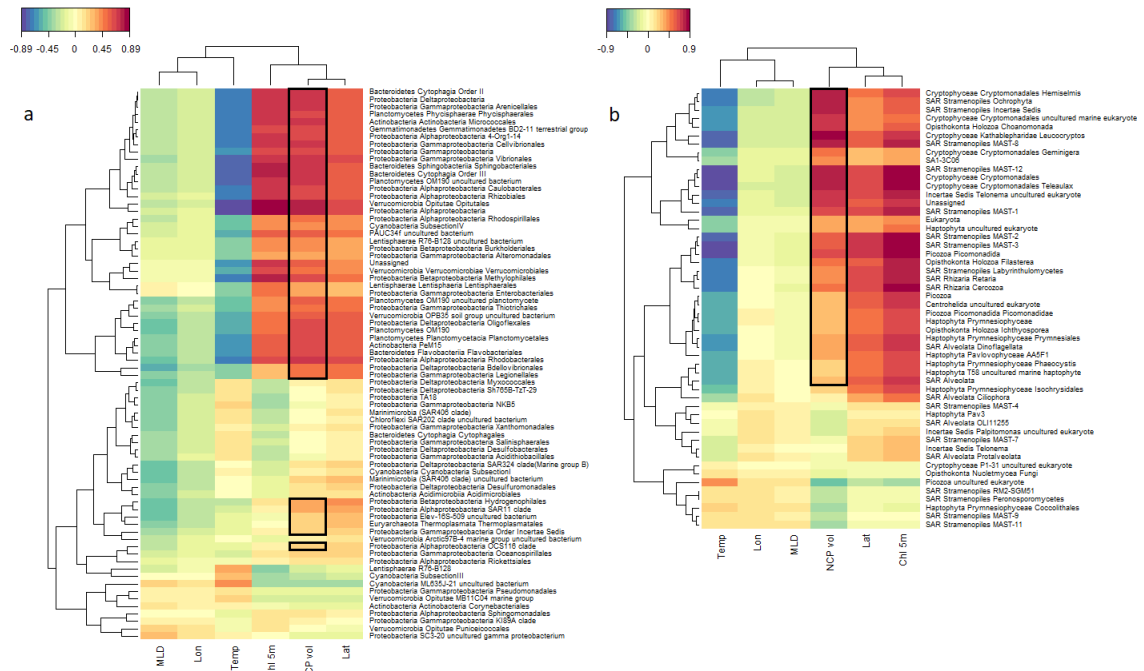
2013, Vallina et al., 2014). Earlier research suggests a peak of phytoplankton diversity at locations with moderate production, with decreasing diversity observed for less-productive and highly-productive sites (Li, 2002, Irigoien et al., 2004). Dominance of a handful of taxa beyond the control of grazers may explain decreased diversity at high productivity. Increased diversity at moderate productivity rates may reflect selective feeding pressures that allow coexistence between a higher diversity of taxa. Within the Western North Atlantic, our data supports the view that the most productive marine communities may exhibit relatively low eukaryotic diversity, a result consistent with meta-analysis and model-based findings that the most productive communities are among the least diverse (Vallina et al., 2014).



**Figure 6: Principal Coordinates Analysis of (a) eukaryotic 18S and (b) bacterial/archaeal 16S samples, ordinated by Bray-Curtis dissimilarity. Samples are color coded based on general region of origin, as visually determined from OSCAR ocean current data for the date of sampling (<https://earth.nullschool.net/>), with green**

**indicating stations located within the Gulf Stream and with red and blue samples originating from waters inshore and offshore of the current, respectively.**

Principal coordinate analysis (PCoA) of both prokaryotic and eukaryotic samples demonstrated distinctions between coastal bloom and other samples (Fig. 6), indicating community dissimilarities. Linear regressions of environmental parameters against the first principal component revealed significant correlations between NCP, temperature, latitude, Chlorophyll, and PC1 for both our 18S and 16S datasets (Supplementary Table 1), suggesting associations between these parameters and community structure. None of these trends remained significant once data from bloom stations S14, S15, and S16 were excluded, however, indicating that these relationships were driven largely by these samples, which possess distinctive community structure, high Chl and NCP, and low water temperatures compared to all other stations.



**Figure 7: Heatmaps showing strength of correlations, determined using Partial Least Squares regression analysis between (a) prokaryotic and (b) eukaryotic taxa and environmental parameters, including NCP. Plots were produced at the fourth taxonomic rank for prokaryotes and eukaryotes. Strength of correlations is denoted by the color scale, with negative and positive correlations in blue and red tones respectively. Rows bounded by black rectangles indicate taxa with a correlation with NCP greater than 0.2. (Full tables of correlation coefficients included with supplementary material). Dendrograms link taxa with similar relationships to the set of variables, and vice versa.**

## Relationships between NCP and specific microplankton taxa

Partial Least Squares (PLS) regression analysis revealed groups of prokaryotic and eukaryotic taxa associated with high volumetric NCP rates (Supplementary Tables 4a-4f), with these relationships again strongly driven by the bloom community.

Eukaryotic taxa associated with NCP included *Ochrophyta*, *Aureococcus anophagefferens*,

picozoa, cryptophytes, prymnesiophytes, and stramenopiles, such as several uncultured MArine STRamenopiles (MAST) clades (Fig. 7b).

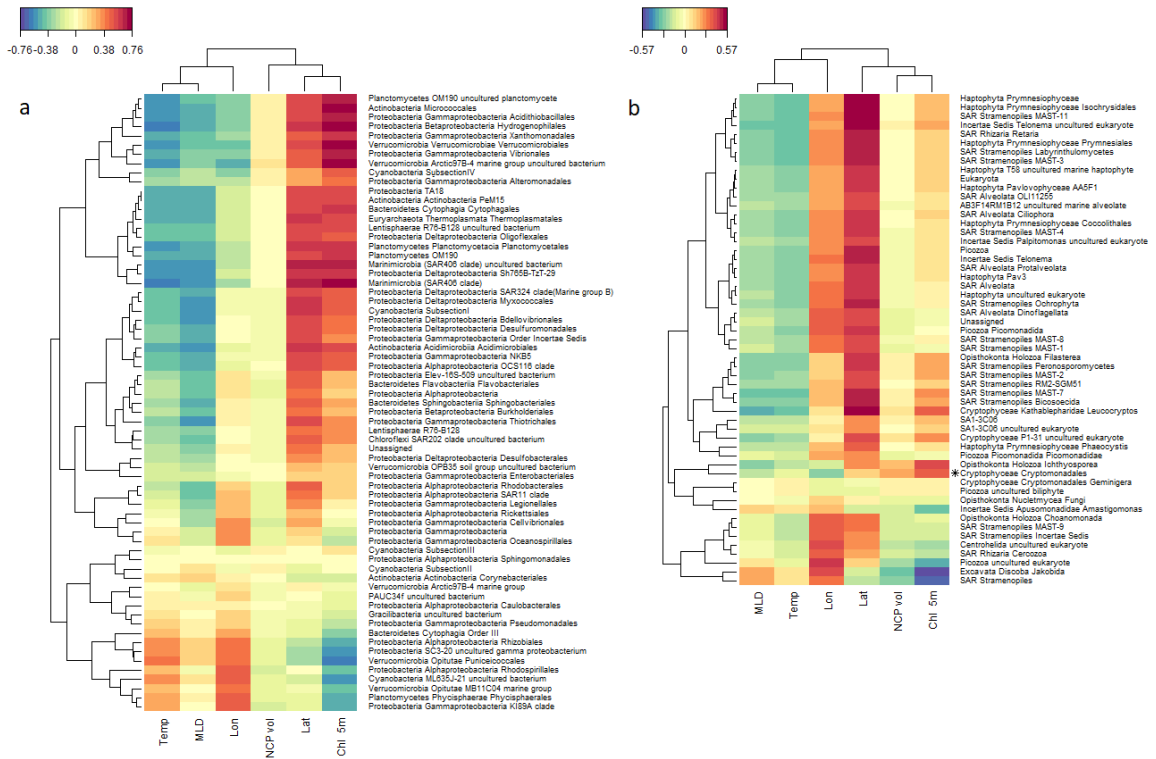
Many of these protists are commonly associated with phytoplankton bloom conditions. *Aureococcus anophagefferens* possesses a large genome optimized for uptake of ambient dissolved organic carbon and nitrogen and is adapted for fast growth under turbid, low-light conditions (Gobler et al., 2011, Gobler et al., 2005). Members of *Chrysophyceae* also form blooms and practice phagotrophy, engulfing and processing particulate matter (Caron et al., 2017). The high abundance of these two taxa within the bloom implies an environment favoring opportunistic uptake of available particulate and dissolved organic material.

Other eukaryotes strongly associated with high NCP include groups of heterotrophic protists: radiolarians, centrohelids, *Labyrinthulomycetes*, as well as flagellates such as *Kathablepharidae*, *Choanomonada*, and uncultured marine stramenopiles. Many of these taxa feed upon algae, bacteria, detritus, and other particles. The associations between these taxa and NCP may indicate flourishing of heterotrophs within an environment with enhanced food and prey concentrations.

The bacterial taxa most correlated with NCP corroborate this picture of a productive bloom ecosystem driven by high phytoplankton productivity. Groups of *Bacteroidetes*, a class of heterotrophic bacteria generally observed to thrive in particle-rich bloom environments (Buchan et al., 2014), are strongly associated with NCP. Other

bacterial groups primarily exhibiting surface or particle-associated lifestyles, including *Verrucomicrobia* and *Planctomycetes*, also display high correlations with NCP. Numerous *Gammaproteobacteria* taxa, including the fast-growing *Vibrionales* clade, are also strongly associated with productivity (Fig. 7a).

We acknowledge that our community sampling represents a snapshot of this bloom and cannot capture successional dynamics. 8-day MODIS satellite chlorophyll data measured before and after our cruise suggest that the bloom first appeared in late July one to two weeks before sampling. Our expedition likely encountered the bloom at its temporal midpoint, with the bloom then fading by late August. We further note that taxa associated with this event may not be characteristic of other blooms that might occur throughout the region. Although satellite imagery indicates that a large bloom often recurs annually in the Mid-Atlantic Bight in late summer, additional sampling is required to confirm whether the observed community structure also recurs.



**Figure 8: Heatmaps of PLS correlations for (a) prokaryotic and (b) eukaryotic taxa at the 4th taxonomic rank, following removal of data from bloom stations S10, S14-S16. Strength of correlations is denoted by the color scale, with negative and positive correlations in blue and red tones, respectively. Asterisks mark taxa with a correlation with NCP greater than 0.2 (one Cryptomonadales clade only). Dendrograms link taxa with similar relationships to the set of variables, and vice versa**

Interestingly, when PLS regression analyses were repeated while excluding the four most productive stations (S10, S14-S16), only a single cryptophyte remained associated with NCP rates, and the overall strength of associations weakened. Outside of the observed bloom, moderate correlations with productivity were displayed by just several groups of cryptophytes and bacterioplankton (Fig. 8a, 8b). These results might indicate that relationships between specific groups of eukaryotic and prokaryotic taxa and NCP in less-productive locations are either undetected by our study or hidden

within the uncertainties of the measurements conducted. At the same time, such a finding may suggest that links between productivity and community structure in this region are complex, with the abundance of any given taxa not strongly associated with measured productivity.

The relationships we have detailed between productivity and selected microplankton taxa exhibit interesting discrepancies with findings from similar work conducted in other regions of the global ocean. A TARA Oceans study of associations between bacterial, eukaryotic, and viral taxa, NPP, and particulate carbon export linked some of the same microplankton groups to primary production and to particle export that were productivity-associated within our full dataset, including *Vibrio* and *Alteromonadales* among bacteria, as well as dinoflagellates, *Labyrinthula*, *Cercozoa*, *Picozoa*, prymnesiophytes, MAST-3, and *Radiolaria* (Guidi et al., 2016).

Intriguingly, however, these abovementioned associations vanish from our analysis when our dataset is limited to less-productive station data, whereas Guidi et al. suggest that these same relationships are strong within the oligotrophic ocean. It is also worth noting that several taxa implicated in carbon export by Guidi et al. show no or even negative correlations with NCP in our analysis, such as *Synechococcus* (Subsection I *Cyanobacteria*) and *Oceanospirillales*. Dissimilarities may be attributable to differences in abundance metrics, molecular methods, and the distinctions between *in-situ* O<sub>2</sub>/Ar-derived NCP, modeled NPP, and optically-determined particle export (i.e. not all NCP is

exported). Further, ecological dynamics encompassed by our regional study may not be extrapolatable to global open-ocean data. Yet our work nevertheless spans a considerable area and range of marine biomes. Rather, our results suggest that outside of high-productivity sites, productivity across a relatively wide region is not strongly associated with specific microbial taxa. Such questions warrant further investigation.

## **2.4 Conclusions**

Our results document a dramatic bloom in Mid-Atlantic Bight coastal waters, where the harmful algal bloom-forming taxon *Aureococcus*, *Chrysophyceae*, heterotrophic protists, and particle-associated bacterioplankton were strongly associated with this productivity peak. This result emphasizes the potential significance of large coastal blooms to productivity patterns in the Western North Atlantic, and highlights HAB-forming *Aureococcus* as a taxon of particular interest. We also find few associations between taxonomy and net community production across a wide range of less-productive waters, suggesting that specific microplankton taxa may not be responsible for driving broader patterns of production across much of this region.

Our quantitative amplicon sequencing approach serves as a useful tool in investigating the ocean microbiome and its influence on the marine environment, providing important additional context beyond relative abundance metrics. Coupled with the ever-increasing resolution and capabilities of *in-situ* biogeochemical methods,

adoption of similar study designs can enable more nuanced examination of the role of the microplankton community across diverse ocean environments.

Supplementary information is available for download at the ISME Journal's website.

Sequences and metadata are available from the NCBI Sequence Read Archive under accession number SRP126177.

## **2.5 Acknowledgements**

This research was supported by an NSF-CAREER grant awarded to NC (#1350710) and a Chateaubriand Fellowship awarded to SW. NC was also supported by the "Laboratoire d'Excellence" LabexMER (ANR-10-LABX-19) and co-funded by a grant from the French government under the program "Investissements d'Avenir". RE was supported by an NSF GRFP award (1106401). We thank the staff of the Bermuda Institute of Ocean Sciences as well as the crew and technicians of the *R/V Atlantic Explorer* for their valuable assistance in organizing and conducting our field study. We are also thankful to Rod Johnson, Bruce Williams, and Natasha McDonald at BIOS for their help with sample shipping and analysis. We are additionally grateful to Karoline Faust for her input on network approaches and to Geoffrey Smith for his help and for our use of his towfish equipment.

### **3. Patterns of North Atlantic microbial community structure in relation to net community production, nitrogen fixation, and nutrient availability**

This chapter has been submitted as:

Wang, S., Tang, W., Delage, E., Gifford, S., Whitby, H., González, A., Eveillard, D., Planquette, H., Cassar, N. Patterns of North Atlantic microbial community structure in relation to net community production, nitrogen fixation, and nutrient availability.

#### **Author contributions**

SW, WT, and NC conceived the study. SW and WT collected molecular samples at sea. WT conducted and analyzed N<sub>2</sub> fixation measurements. SW collected O<sub>2</sub>/Ar data and analyzed NCP measurements. AG and HW collected trace metal samples and analyzed trace metal data with HP. SW conducted DNA extraction, library prep, and sequencing work. SW and SG calculated quantitative amplicon abundances. SW, ED, and DE performed statistical analyses and network construction. All co-authors assisted in drafting this manuscript and gave their final approval prior to submission.

### **3.1 Introduction**

Marine phytoplankton mediate important fluxes of carbon, oxygen, nitrogen, and other elements between organic, atmospheric, and oceanic pools. Two critical biogeochemical processes in which microplankton play key roles include the sequestration of biological carbon in the deep ocean via sinking plankton biomass and nitrogen (N<sub>2</sub>) fixation. Both the biological carbon pump as well as marine N<sub>2</sub> fixation are thought to be heavily influenced by the ecology and structure of the surface microplankton community (Boyd and Newton, 1995, Richardson and Jackson, 2007, Guidi et al., 2016, Zehr and Kudela, 2011). Consequently, microbial oceanographers are devoting effort to studying the links between microbial community structure, the physical and chemical environment, and carbon and nutrient cycling.

The complex relationships between phytoplankton, other microbes, and community productivity remain only partially explored. At a broad scale, even the relationship between oceanic community production and microplankton diversity remains unclear (Cermeño et al., 2013, Vallina et al., 2014). Ongoing research also seeks to pinpoint whether specific phytoplankton taxa are key to driving primary production and carbon export (Guidi et al., 2016, Guidi et al., 2009, Lin et al., 2017, Boyd and Newton, 1995, Richardson and Jackson, 2007, Cassar et al., 2015). Alongside improvements in molecular methods, the development of underway methods to measure net community production (NCP; NCP = gross primary production (GPP) -

community respiration (CR)), has enabled closer comparisons between microbial community structure and carbon export potential (Lin et al., 2017, Wang et al., 2018). While linking productivity to diversity ultimately requires understanding underlying ecological mechanisms, we remain at a stage where inferences from the field are necessary to highlight patterns of interest for focused study.

Both marine community structure and productivity are influenced by the availability of macronutrients and trace micronutrients (Sohm et al., 2011, Mousing et al., 2018). Nitrogen has long been identified as the limiting macronutrient across most of the global ocean (Howarth, 1988). Nitrogen ( $N_2$ ) fixation by marine diazotrophs provides an important supply of nitrogen to the ocean and has been hypothesized to play a critical role in oligotrophic systems like the Sargasso Sea (Karl et al., 1997, Capone et al., 2005). Recent studies, however, have detected significant  $N_2$  fixation in coastal and even polar environments, suggesting that the extent of marine diazotrophs is larger than previously surmised (Fernandez et al., 2011, Bentzon-Tilia et al., 2015, Shiozaki et al., 2018, Tang et al., 2019). Our understanding of marine microbes involved in  $N_2$  fixation is also incomplete and continually evolving, as evidenced by the recent discovery of heterotrophic diazotrophic bacteria present at higher-than-expected abundances in the surface ocean and at depth (Delmont et al., 2018, Bombar et al., 2016).

In addition, numerous trace metals are essential for biological processes. These include iron (a cofactor for enzymes involved in  $N_2$  fixation, photosynthesis, respiration,

and other processes), manganese (present in photosystem II), zinc (a component of carbonic anhydrase), and copper (incorporated in N<sub>2</sub> fixation and electron transport mechanisms) (Twining and Baines, 2013, González-Guerrero et al., 2014), among numerous other metals. Examination of links between trace metals and microbial ecology processes remains an ongoing area of research (Mellett et al., 2018).

Microplankton community structure is often evaluated currently using high-throughput amplicon sequencing of the small ribosomal subunit (SSU) rRNA gene, which permits broad taxonomic identification of marine microbes as well as the collection of large datasets covering diverse physical and biogeochemical conditions. Importantly, amplicon surveys traditionally summarize community data via relative abundances. Since relative metrics are sensitive to the abundances of other taxa within the same sample, using percentage-based data involves key compositional drawbacks (Aitchison, 1981, Aitchison, 1982), complicating interpretation of relationships between relative abundances and other parameters.

To avoid such drawbacks, investigators are increasingly applying quantitative approaches to rDNA amplicon sequencing, employing internal standards to estimate the abundances of environmental sequences (Gifford et al., 2011, Satinsky et al., 2013, Wang et al., 2018, Smets et al., 2016, Tkacz et al., 2018, Lin et al., 2019). Such quantitative techniques themselves possess limitations (see Discussion) but uncover real differences in taxonomic abundances and avoid compositional effects. Quantitative amplicon

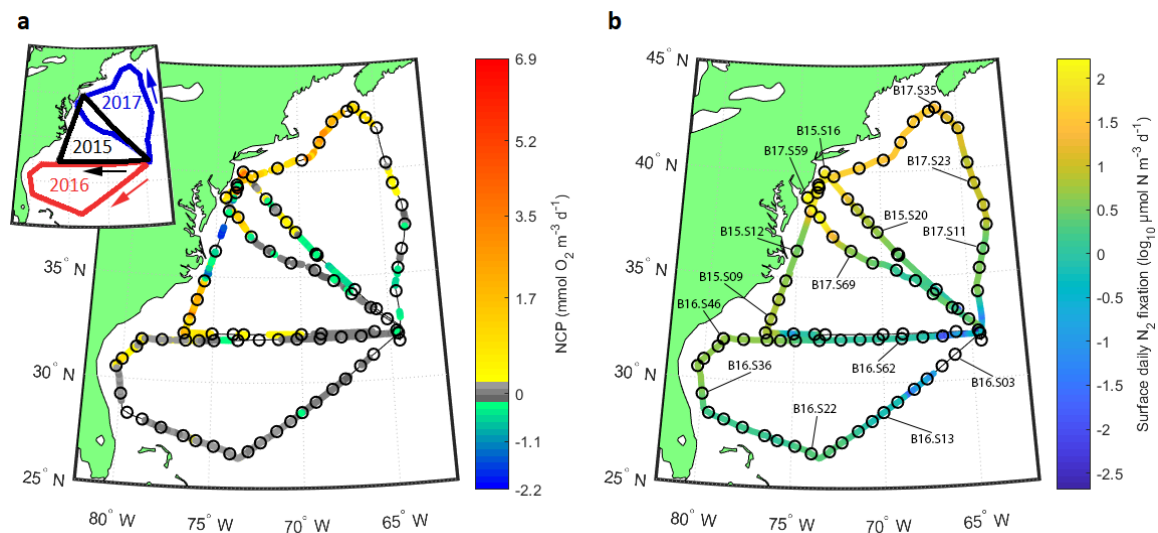
sequencing thus constitutes a powerful new tool, particularly for making direct comparisons between taxonomic abundances and biogeochemical rates and quantities. Additionally, the application of quantitative amplicon data towards community co-occurrence network analysis represents a natural refinement of network analysis approaches.

In this study, we combine new measurements and quantitative community data with previously-published datasets collected in 2015 (Wang et al., 2018) and 2016 (Tang et al., 2019), yielding a three-year ecological and biogeochemical survey of the summer Western North Atlantic. The Western North Atlantic represents a favorable environment to explore relationships between marine biogeochemistry and microbial community structure, featuring stark contrasts between a productive coastal shelf with high  $N_2$  fixation rates (Tang et al., 2019, Mulholland et al., 2012, O'Reilly and Busch, 1984), the Gulf Stream, and the oligotrophic Sargasso Sea (Lomas et al., 2013, Steinberg et al., 2001). Parallel investigation of microplankton community structure, high-resolution NCP and  $N_2$  fixation rates, and nutrients and trace metals has never been previously performed in this region or elsewhere. Leveraging this unique dataset, we employ statistical and network-based analyses commonly applied in global ocean studies to evaluate regional-scale links between  $N_2$  fixation, NCP, macronutrients, trace micronutrients, and the diversity and quantitative abundances of prokaryotic and eukaryotic microplankton.

### 3.2 Materials and Methods

#### Measurements of underway net community production, nitrogen fixation rates, nutrients, and trace metals.

Our combined set of new and previously-published data (Tang et al., 2019, Wang et al., 2018) includes around 10 250 km of continuous and discrete measurements across the western North Atlantic Ocean over three summer voyages aboard the *R/V Atlantic Explorer*. CTD casts were taken once daily (approx. 300 km intervals), usually before dawn. The cruise tracks, conducted from 3-13 August 2015, 3-12 August 2016, and 29 July-7 August 2017, are illustrated in (Fig. 9).



**Figure 9:** a) Map of O<sub>2</sub>/Ar-based volumetric NCP rates measured via EIMS across the Western North Atlantic in July-August 2015, 2016, and 2017. Positive rates are represented by the warm portion of the color scale, while negative rates are indicated using a cool color scale. Balanced NCP rates are shown in greyscale tones. Sites of molecular sampling are displayed by open circles. b) Map of log-scaled surface daily N<sub>2</sub> fixation rates measured using the FARACAS method. Sites of molecular sampling are displayed by open circles, with selected sites labeled. Inset plot illustrates cruise tracks and directions of travel, color-coded for each year of sampling.

Continuous measurements of the dissolved O<sub>2</sub>/Ar ratio in surface seawater were conducted using the EIMS method (Cassar et al., 2009). N<sub>2</sub> fixation rates were measured using the recently developed method of underway Flow-through Acetylene Reduction Assay Cavity-ring down laser Absorption Spectroscopy (Cassar et al., 2018). The methodology and calculations for both methods are described in Appendix B. Sample seawater for O<sub>2</sub>/Ar, N<sub>2</sub> fixation, and trace metals was pumped aboard ship using a trace-metal clean towfish (Geofish) and plumbing (Bruland et al., 2005). Both dissolved (dTM) and particulate (pTM) fractions of trace metals were determined by SF-ICP-MS. (Appendix B).

### **Sample collection and library preparation for 16S and 18S rDNA amplicon sequencing**

Microplankton samples for rDNA amplicon sequencing were collected every 6-8 hours using 0.22 µm polycarbonate filters (Millipore, Billerica, MA, USA) from surface (3-5 m) seawater collected from towfish and 5 m CTD samples. Between 1-2 l (mean: 1.2 l) were filtered using a peristaltic pump, with smaller volumes filtered at particularly high-biomass stations. In 2015, due to a limited sample pool, all utilized filters had been preserved in RNAlater (Thermo Fisher, Waltham, MA, USA), while RNAlater preservation was not employed for 2016 and 2017 samples. Otherwise, sampling methodology was identical across all three years. Filters were flash-frozen in liquid nitrogen following filtration and stored at -80°C.

Addition of internal standard control sequences was performed following the protocol described in (Wang et al., 2018) for all samples. New standard solutions were prepared prior to extraction and amplification of the 2016 and 2017 samples, with the result that spike amounts varied due to differences in stock concentrations and adjustments based on the 2015 sequencing results. For 2017 samples, 0.66 ng of *S. pombe* 18S standard and 6.1 ng of *T. thermophilus* 16S standard were added to each sample before bead-beating, both in 50  $\mu$ l volumes. This spike constituted an addition of approximately 5 610 000 copies sample<sup>-1</sup> of *S. pombe* gDNA and 5 620 000 copies sample<sup>-1</sup> of *T. thermophilus* gDNA. For both 2016 and 2015 samples, 5 780 000 copies (0.679 ng) of *S. pombe* gDNA were introduced. *T. thermophilus* gDNA standard was added in amounts of 14 040 000 copies (15.25 ng) per sample for 2016 samples and 2 800 000 copies (3.05 ng) per sample for 2015 samples.

DNA extractions, PCR amplification of the 16S V4 and the 18S V4 regions, sample quantitation, and library pooling followed the same procedure detailed in (Wang et al., 2018). Illumina MiSeq (300 bp PE, V3) sequencing was performed at the Duke Center for Genomic and Computational Biology.

### **rDNA sequencing data processing and OTU picking**

We obtained 22 807 458 paired-end reads from 174 surface samples. Paired-end sequences were trimmed, merged, quality filtered, demultiplexed, then filtered for

chimeras following the pipeline detailed in (Wang et al., 2018), yielding 6 094 544 16S sequences and 2 465 684 18S sequences. Primer and other non-biological sequences were then filtered using Tagcleaner (Schmieder et al., 2010). Open-reference OTU picking was performed using identical parameters (97% similarity, Usearch 6.1 (Edgar, 2010), SILVA 123.1 (Pruesse et al., 2007)) as in (Wang et al., 2018) (see Appendix B for additional details). Following taxonomy assignment, we filtered internal standard, plastid, mitochondrial, and metazoan rDNA sequences from our taxonomy tables.

## **Statistical analyses**

The combined 2015-2017 sequencing dataset included 8 841 16S OTUs across 79 16S samples and 14 344 18S OTUs across 79 18S samples, post-filtering. Sequencing data was imported into R 3.4.1 (R Core Team, 2017) using the phyloseq package (McMurdie and Holmes, 2013). Calculation of quantitative abundances was performed as in (Satinsky et al., 2013), with the total rDNA gene abundance  $l^{-1}$  of a taxon determined as the number of reads mapping to that OTU in the sequencing output, divided by the product of the standard recovery ratio ( $\#$  standard reads recovered /  $\#$  standard reads introduced) and the volume filtered. Details of filtration of low-count samples and outliers as well as rarefaction and ordination procedures are described in Appendix B.

Calculation of quantitative 16S rDNA gene abundances produced a discrepancy in which samples collected in 2015 displayed significantly higher total 16S rDNA

abundances (Wilcoxon rank sum test,  $p < 0.001$ ) than 16S rDNA samples obtained in 2016 and 2017 (Fig. 27). This difference in calculated abundances likely results from above-mentioned variation in the 16S spike quantity between years and/or from differences in preservation protocols. To investigate this, we compared our calculated SAR11 rDNA copy numbers for stations within 400 km of Bermuda against reported summer SAR11 abundances of  $1.5 - 2.0 \times 10^8$  cells/l for this region (Carlson et al., 2008, Morris et al., 2002). While calculated 2016 and 2017 SAR11 abundances ( $1.41 \times 10^8$  cells/l and  $1.38 \times 10^8$  cells/l) aligned closely with values from the literature, samples from 2015 near Bermuda displayed a median SAR11 abundance of  $6.5 \times 10^8$  cells/l. Consequently, we adjusted calculated 2015 16S rDNA abundances by a factor of 0.21. Adjusted 2015 organismal abundances for SAR11 and *Prochlorococcus* align closely with expected abundances from the literature (Table 5). We further cross-validated our corrected organismal abundances for *Prochlorococcus* against patterns of *Prochlorococcus* abundance obtained from concurrently-collected flow cytometry samples, observing good agreement in abundance patterns between the two datasets (Fig. 28). The corrected 2015 16S rDNA dataset was subsequently used in generating taxonomy plots, ordination analyses, PLS regression analyses, and co-occurrence network construction.

Partial least squares regression analysis was performed following (Wang et al., 2018). Two separate PLS regression analyses were performed: one utilizing the full molecular dataset from all three years to assess taxon-specific relationships with NCP

and N<sub>2</sub> fixation, and another utilizing only molecular data from 2016 and 2017 to investigate relationships between specific taxa and macronutrient and trace metal concentrations. To handle missing data, the non-iterative partial least squares (nipals) algorithm was employed.

## **Network analysis**

Weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath, 2007) was performed to select subnetworks of microplankton taxa from the whole dataset. Following the WGCNA standard protocol (Guidi et al., 2016), we constructed an adjacency matrix of the microplankton community dataset using Pearson correlations (Langfelder and Horvath, 2008). A soft-thresholding power law ( $p = 8$ ) was applied to fit the weighted graph into a scale-free topology. A topological overlap measure (TOM) was then calculated for each pair of OTUs, based on the weight of their pairwise correlations with close neighbors and weighted correlations with other OTUs. Finally, hierarchical clustering based on TOM was performed to identify sub-networks of co-occurring taxa.

Two sub-networks (Subnetworks 1 and 2) that were most significantly positively associated with both NCP and N<sub>2</sub> fixation (Figs. 31, 32) were extracted, as well as the subnetwork (Subnetwork 3) most negatively correlated with these rates, although not significantly.

A global, robust network was built via FlashWeave (Tackmann et al., 2018) using default parameters (sensitive = True, heterogeneous = False) on the whole dataset. Induced subgraphs as identified from WGCNA were then extracted. Subnetwork comparison was then performed using L\_GRAAL, a graph alignment tool (Malod-Dognin and Pržulj, 2015). L\_GRAAL aligns taxa from two distinct co-occurrence networks when they share (i) similar topological properties and (ii) rDNA sequence homology (alpha = 0.4 for accurate consensus between both features) (Mandakovic et al., 2018).

Centrality metrics characterize node importance in a network, with different measures spanning from local to global scales. The generalized keystone index introduced in (Estrada, 2007) uses a factor analysis approach that reduces dimensionality by transforming a set of centralities in a linear combination, allowing easy ranking of nodes with a single metric. Two local centralities (degree and betweenness centralities), a “meso-scale” centrality (subgraph centrality) and a global centrality (closeness centrality) were considered. Visualization of subnetworks and alignments was performed with HiveAlign (see Appendix B).

### ***3.3 Results and discussion***

#### **Biogeochemistry: NCP, N<sub>2</sub> fixation rates, nutrients, and trace metals**

Recorded volumetric NCP rates range between -2.2 and 6.9 mmol O<sub>2</sub> m<sup>-3</sup> d<sup>-1</sup> (mean: 0.4 mmol O<sub>2</sub> m<sup>-3</sup> d<sup>-1</sup>, std dev: 0.9 mmol O<sub>2</sub> m<sup>-3</sup> d<sup>-1</sup>), while surface daily N<sub>2</sub> fixation

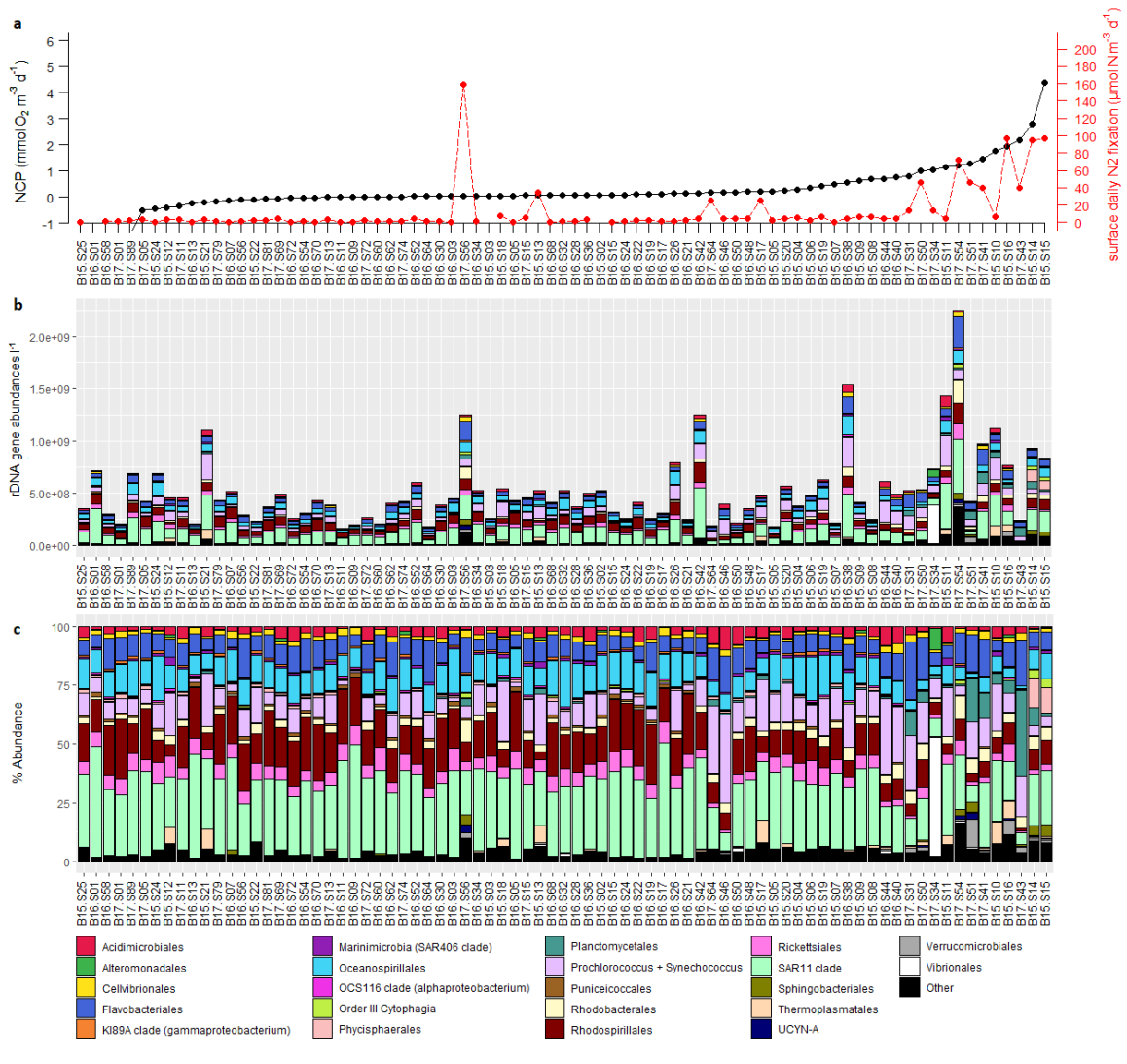
rates range between undetectable values ( $<0.19 \mu\text{mol N m}^{-3} \text{ d}^{-1}$ ) and  $167.8 \mu\text{mol N m}^{-3} \text{ d}^{-1}$  (median:  $3.6 \mu\text{mol N m}^{-3} \text{ d}^{-1}$ , std dev:  $35.1 \mu\text{mol N m}^{-3} \text{ d}^{-1}$ ) (Fig. 9). The highest NCP and  $\text{N}_2$  fixation rates are recorded in 2015 and 2017 near the New Jersey coast in the Mid-Atlantic Bight (MAB). The southern region of our study records low NCP values apart from a section of the Floridan coast. Undersaturation in  $\text{O}_2/\text{Ar}$  leading to negative NCP rates is observed in 2015 off the Delaware coast and during portions of our open-ocean transects (see Appendix B for further discussion). Overall, our range and distribution of supersaturated integrated NCP ( $<4 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$  in the Sargasso Sea,  $163 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$  in the MAB) is consistent with comparable summer measurements from this region reported in other literature ( $<4 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$  in the Sargasso Sea,  $221 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$  in the MAB) (Estapa et al., 2015, Falkowski et al., 1988, Letscher and Moore, 2017). Integrated  $\text{N}_2$  fixation rates measured in the open ocean ( $<10 \mu\text{mol N m}^{-2} \text{ d}^{-1}$ ) are similarly comparable with published studies (Orcutt et al., 2001), although the upper range of rates recorded in the productive coastal region ( $3\,000 \mu\text{mol N m}^{-2} \text{ d}^{-1}$ ) are among the highest ever reported (Luo et al., 2012, Tang et al., 2019).

Nearly all sites with high NCP ( $>3 \text{ mmol O}_2 \text{ m}^{-3} \text{ d}^{-1}$ ) and high  $\text{N}_2$  fixation rates ( $>75 \mu\text{mol N m}^{-3} \text{ d}^{-1}$ ) are recorded in coastal waters (bathymetry  $<200 \text{ m}$ ). The vast majority of these high NCP and high fixation waters also lie north of the northern wall of the Gulf Stream as defined by the intersection of the  $15^\circ\text{C}$  isotherm with the  $200\text{m}$  isobath (Joyce et al., 2000), and are characterized by lower temperatures and salinity

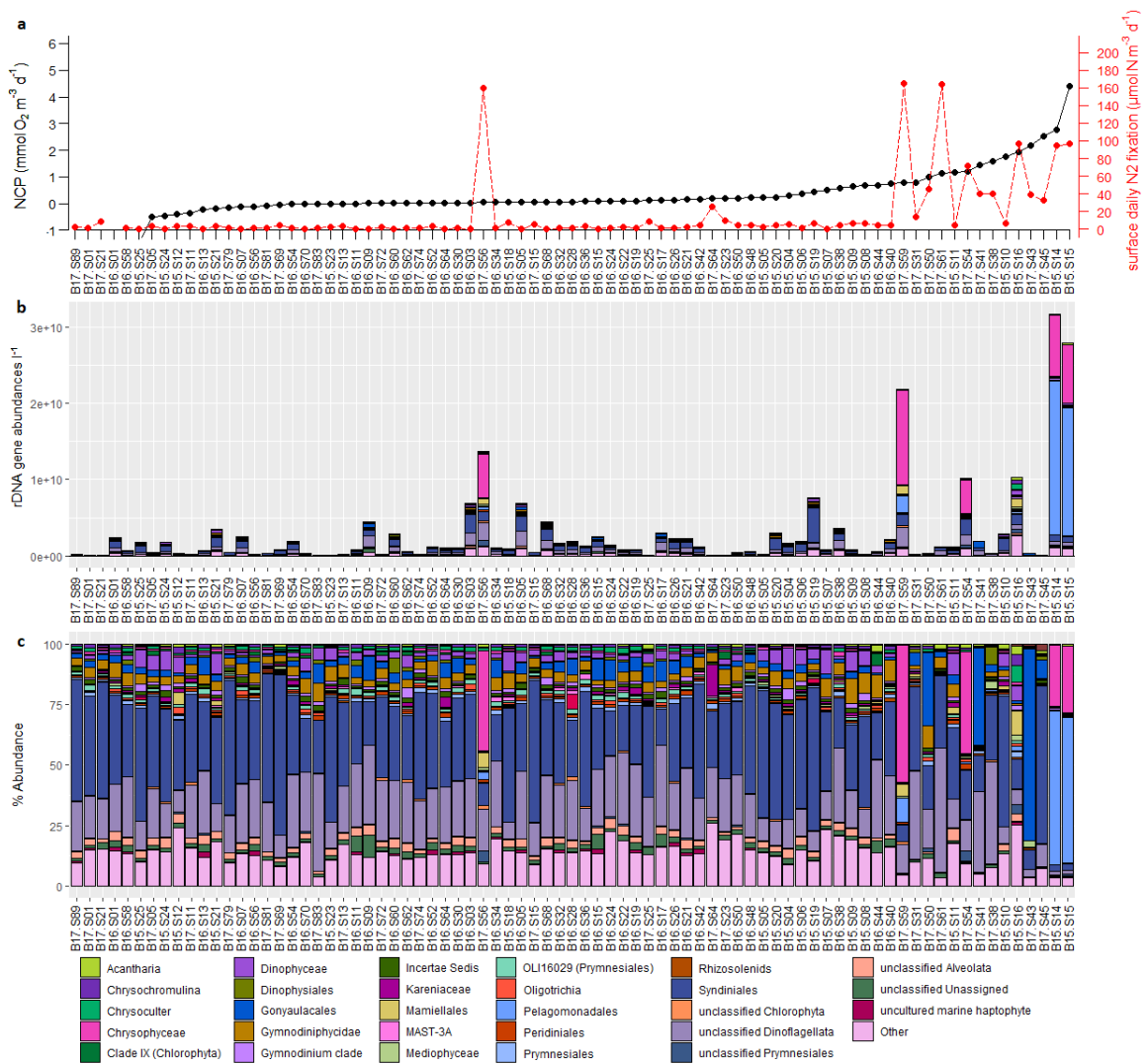
(Fig. 29). Overall, NCP recorded at molecular sampling sites was significantly correlated with measured N<sub>2</sub> fixation (Pearson = 0.53,  $p < 0.001$ ; Spearman's rho = 0.66,  $p < 0.001$ ).

Nitrate concentrations range between the detection limit (0.03  $\mu\text{M}$ ) and 1.26  $\mu\text{M}$  (Fig. 30). Measured phosphate levels follow NCP, with high coastal values and low open-ocean concentrations (range of 0.014 to 0.24  $\mu\text{M}$ ). High silica was observed in the MAB (range of 0.21 to 1.87  $\mu\text{M}$ ).

Among the measured trace metals (dFe, pFe, dCu, pCu, dMn, pMn, dZn, pCo, pNi, pAl, pBa), dissolved and particulate copper and manganese follow a similar spatial distribution to NCP and N<sub>2</sub> fixation, particularly for the 2017 cruise, with higher values in shelf waters along the New England and Mid-Atlantic coast and lower concentrations in open-ocean samples. No distinct spatial pattern in dissolved iron is observed, although somewhat elevated concentrations of particulate iron are detected in the MAB, off Nova Scotia, and in some open-ocean regions.



**Figure 10: a)** NCP rates (black) and  $\text{N}_2$  fixation rates (red) measured at each sampling site, with sites ranked from left-to-right in order of increasing NCP. Sites at which NCP measurements were not collected are grouped on the far left. **b)** Bar plot of 16S prokaryotic taxonomy represented at the order level using quantitative rDNA gene abundances  $l^{-1}$ . **c)** Bar plot of 16S prokaryotic taxonomy at the order level using relative abundances.



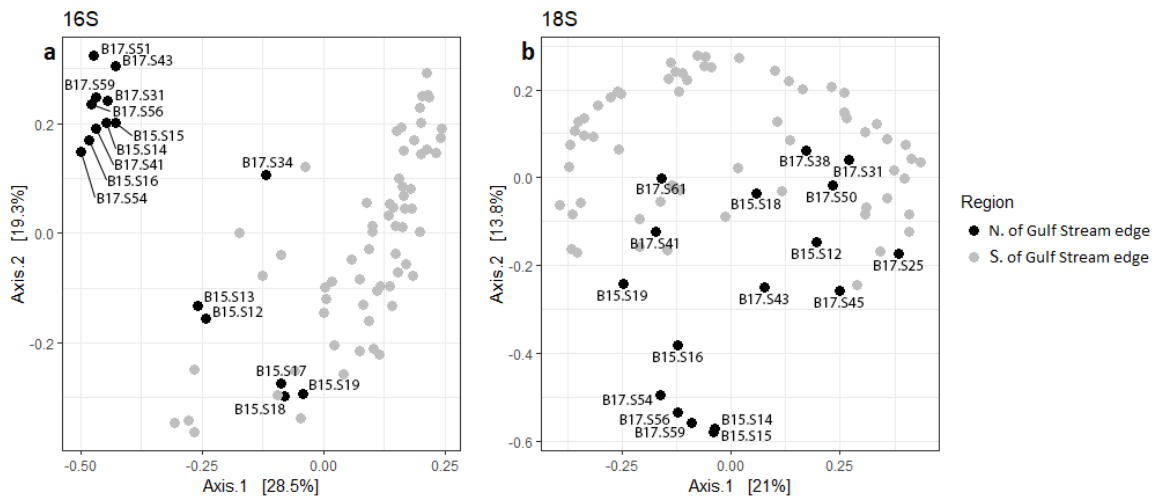
**Figure 11: a) NCP rates (black) and  $\text{N}_2$  fixation rates (red) measured at each sampling site, with sites ranked from left-to-right in order of increasing NCP. Sites at which NCP measurements were not collected are grouped on the far left. b) Bar plot of 18S eukaryotic taxonomy represented at the 5th taxonomic rank using quantitative rDNA gene abundances  $\text{l}^{-1}$ . c) Bar plot of 18S eukaryotic taxonomy at the 5th taxonomic rank using relative abundances.**

## Microbial ecology: Microplankton community composition and abundance patterns

Marine prokaryotic 16S rDNA samples typically display high abundances of SAR11 as well as *Prochlorococcus* (Subsection I). Bacterioplankton belonging to SAR86 (*Oceanospirillales*), *Flavobacteriales*, and *Rhodospirillales* also comprise appreciable fractions of the prokaryotic community. A noticeable shift in community composition towards *Planctomycetes*, *Phycisphaerales*, *Vibrionales*, and *Sphingobacteriales*, however, is observed for stations with high NCP and/or N<sub>2</sub> fixation rates (Fig. 10).

Generally, eukaryotic 18S rDNA samples are dominated by dinoflagellate OTUs (Fig. 11), particularly *Syndiniales*—endoparasitic alveolates infecting a wide range of organisms from ciliates to cercozoa, zooplankton, and fish eggs (Guillou et al., 2008, Not et al., 2007, Romari and Vaulot, 2004). The life cycle of *Syndiniales* also involves a free-living stage in which large numbers of *Syndiniales* are dispersed as dinospores, perhaps partially explaining their high abundances in many amplicon studies (Guillou et al., 2008). Eukaryotic microplankton belonging to *Gonyaulacales*, *Gymnodiniphycidae*, and *Dinophyceae* appear at around 5-10% relative abundances across most samples. Samples from the MAB and New England shelf exhibiting high NCP and N<sub>2</sub> fixation rates in 2015 and 2017 are dominated by *Chrysophyceae* and *Aureococcus anophagefferens*. Several of these samples also display elevated abundances of the chlorophyte *Mamiellales*. Notably, of the seven stations with the highest N<sub>2</sub> fixation rates, five show extremely elevated *Chrysophyceae* abundances (25-55% of 18S rDNA abundances). These samples from the

MAB in August 2015 and 2017 cluster closely together during PCoA analysis, suggesting recurrence of community structure between years (Fig. 12b). This pattern is even stronger for coastal 16S rDNA samples collected north of the Gulf Stream, which form a tight cluster distinct from other samples (Fig. 12a).



**Figure 12: Principal Coordinates Analysis (PCoA) of a) prokaryotic 16S and b) eukaryotic 18S samples, ordinated using weighted Bray-Curtis dissimilarity. Samples are color coded based on whether they were collected north or south of the northern wall of the Gulf Stream in that year as determined from monthly EN4 temperature data.**

Although we observe high NCP,  $N_2$  fixation, and high abundances of *Chrysophyceae* in the MAB in both August 2015 and 2017, we note that *Aureococcus anophagefferens* does not bloom here to the same extent in 2017 as in 2015. In 2017, *Aureococcus* remains at 0.7-10.1% of the eukaryotic community within the MAB, with an estimated  $7.2 \times 10^7$  to  $2.2 \times 10^9$  18S rDNA gene sequences  $l^{-1}$ , contrasted with 2015 sampling ( $1.8 \times 10^8$  to  $2.0 \times 10^{10}$  rDNA genes  $l^{-1}$ ). The absence of a similarly dramatic 2017 *Aureococcus* bloom in the MAB indicates that such events may not recur annually, or that

blooms are subject to variable timing and magnitude. Results from coastal monitoring of *Aureococcus* demonstrate that *Aureococcus* abundances may fluctuate by an order of magnitude between months during the growing season, and that the seasonal biomass peak may precede the start of August by several weeks (Mulholland et al., 2002).

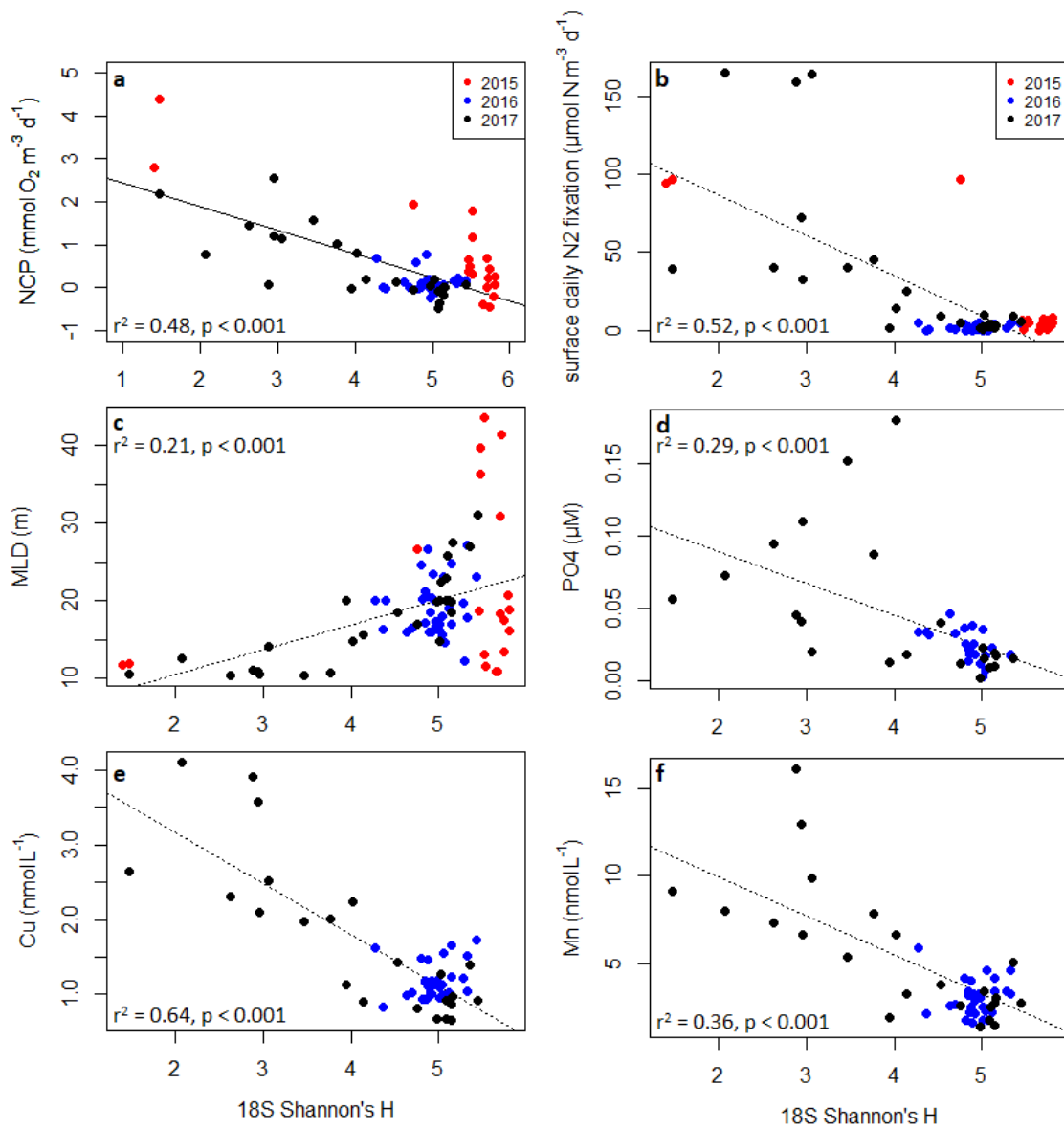
We emphasize that our quantitative community sequencing approach is subject to limitations, necessitating cautious interpretation of calculated abundances. Crucially, our quantitative method does not account for variation in rDNA gene copy number across diverse bacterial and eukaryotic taxa. Prokaryotic rDNA copy numbers are thought to mostly vary between 1-15 copies per cell (Větrovský and Baldrian, 2013), while eukaryotic rDNA gene copies cell<sup>-1</sup> can vary by orders of magnitude. Additionally, the internal standard technique assumes that recovery rates are identical for internal standard reads and natural sequences, an assumption that may be affected by amplification biases. Consequently, calculated abundances using this technique should not be interpreted as accurate estimates of absolute organismal cell concentrations, but rather as reflecting the changing quantitative abundance of a given taxon across samples. This statistical feature is of primary interest for (i) comparing organismal abundances with environmental parameters and (ii) building co-occurrence networks.

Testing of the internal standard technique has, however, generally demonstrated strong agreement with other quantification techniques including flow cytometry counts, HPLC, phospholipid fatty acid (PLFA) analysis, and substrate induced respiration (SIR)

analysis (Smets et al., 2016, Lin et al., 2019). These method intercomparisons lend confidence to our results and indicate that our approach is reliably quantitative.

### **Broad patterns: Community diversity vs. NCP, N<sub>2</sub> fixation, and environmental properties**

We observe a significant negative relationship between NCP and eukaryotic community Shannon's H diversity (Pearson = -0.70,  $p < 0.001$ ; Spearman's rho = -0.31,  $p < 0.05$ ), in which stations with lower eukaryotic diversity tend to exhibit higher NCP rates (Fig. 13). This trend is robust, remaining significant even with all 2015 data excluded. A similar negative relationship is observed between N<sub>2</sub> fixation rates and eukaryotic diversity (Pearson = -0.73,  $p < 0.001$ ; Spearman's rho = not significant). Overall, low eukaryotic diversity was also significantly associated with shallow MLDs and high PO<sub>4</sub>, dCu, pCu, dMn, pMn, and pAl concentrations ( $p < 0.05$  for all) (Supplementary Table 2a, Fig. 13).



**Figure 13: Linear regressions of a) NCP b) N<sub>2</sub> fixation rates c) Mixed layer depth d) PO<sub>4</sub> e) dCu f) dMn versus 18S Shannon's H diversity. Significance and goodness-of-fit are displayed for each plot.**

Prokaryotic diversity is not correlated with NCP ( $p > 0.05$ ) but is significantly related to N<sub>2</sub> fixation rates (Pearson = 0.30,  $p < 0.01$ ; Spearman's rho = 0.48,  $p < 0.001$ ).

Prokaryotic diversity was also positively correlated with increasing Latitude and

Chlorophyll-a (Supplementary Table 2a, Fig. 13).

These results are interesting given ongoing discussion concerning the shape of microplanktonic productivity-diversity relationships (PDR) in oceanic ecosystems. Identifying how productivity and microbial diversity in an ecosystem are linked can provide insight into processes such as competitive exclusion (Hardin, 1960), “kill-the-winner” (Thingstad and Lignell, 1997), and co-existence. Earlier studies suggested a unimodal “humped” shape to marine PDRs, in which diversity increases with productivity, peaking at intermediate levels of production, then falling (Irigoien et al., 2004, Li, 2002). More recent work has not only documented a range of relationships ranging from positive to negative to flat (Smith, 2007, Olli et al., 2015), but has also raised methodological concerns over previous field studies, postulating instead that diversity may not be correlated with productivity, with rare taxa impacting species richness but not greatly influencing production (Cermeño et al., 2013).

An OTU-based definition for prokaryotic diversity is particularly imperfect, given the ambiguity of the “species” concept for prokaryotes. Nevertheless, the lack of a PDR for our 16S rDNA dataset should serve as a reminder that the shape of observed marine microbial PDRs may also depend on how diversity is defined and which functional groups (algae, free-living and surface-associated bacteria, viruses...) are captured by field methods and considered as part of the community. Additionally, the

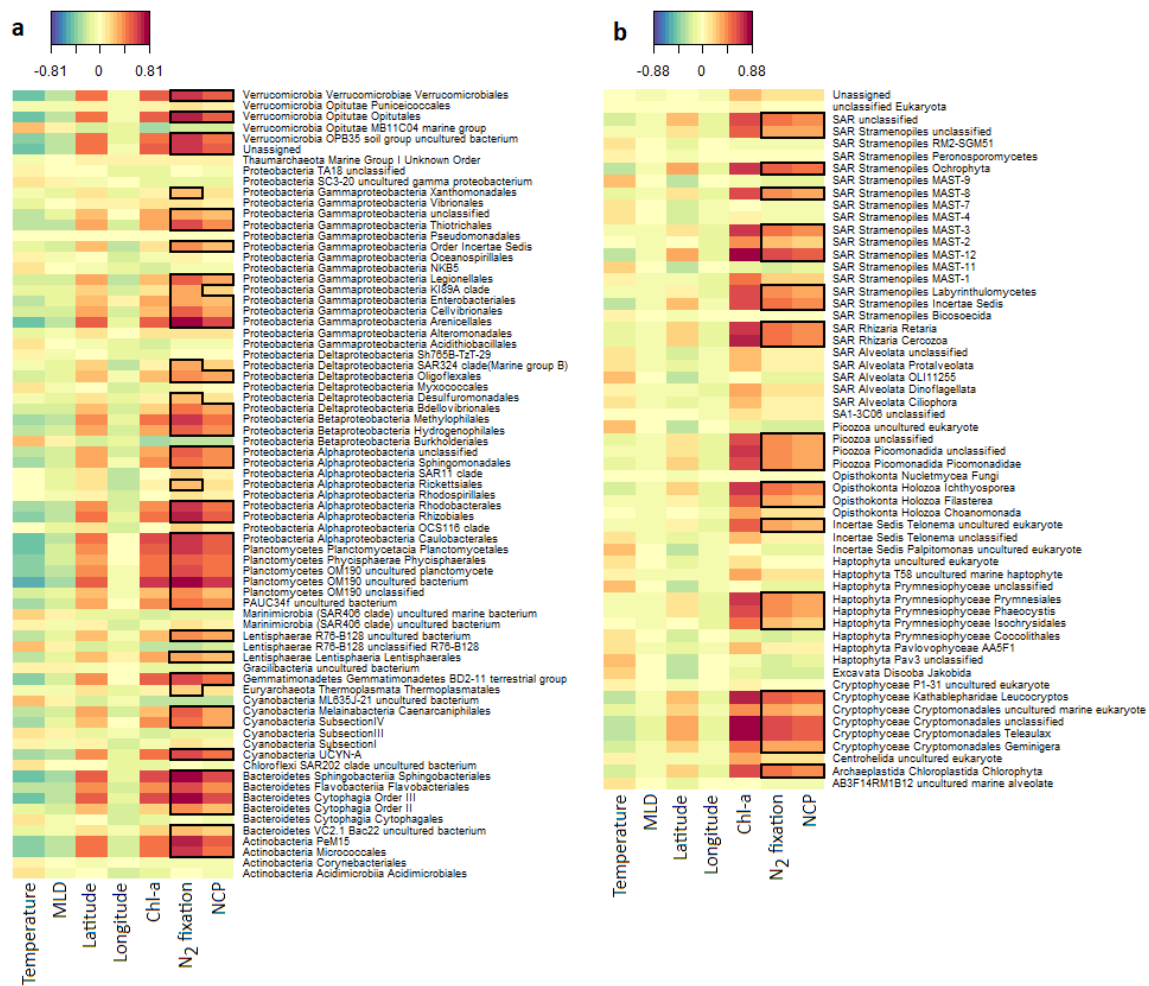
choice of particular productivity metrics (biomass, Chl-a, NCP, NPP, etc...) for evaluating PDRs deserves closer study.

Increasing competitive exclusion driven by opportunistic bloom taxa has been proposed as an explanation for the negative slope of unimodal PDRs. Our taxonomy results suggest that competitive mechanisms may play a role in driving our observed negative PDR for eukaryotes. The dominant eukaryotic plankton identified in our sampling, *Chrysophyceae* and *Aureococcus*, can produce allelopathic compounds to inhibit the growth of algal competitors. Our NCP measurements integrate over a period of 2-8 days, approximately the same timescale at which modeling studies have suggested the PDR is most unimodal (Vallina et al., 2014). Our findings emphasize that links between microbial diversity and marine production may be more complex than assumed, reinforcing the need to reevaluate broad assertions concerning the shape of PDRs in ocean ecosystems.

The significant negative relationship between eukaryotic diversity and both dissolved and particulate Cu warrants particular discussion. Elevated concentrations of Cu are toxic to algae and cyanobacteria (Brand et al., 1986), inducing production of protective organic ligands (Pérez-Almeida et al., 2013) and impacting cyanobacterial N<sub>2</sub> fixation (Wurtsbaugh and Horne, 1982). Copper concentrations observed at low-diversity stations were elevated, with dCu reaching 4.1 nM, similar to values considered toxic to the most sensitive microplankton species (Levy et al., 2007). Additionally,

numerous species of algae release allelopathic toxins in response to copper-induced toxicity stress (Maldonado et al., 2002, Moeller et al., 2007, Long et al., Submitted), possibly contributing to a reduction in biodiversity through competitive exclusion. Although toxicity depends on the bioavailability of metals linked to their chemical speciation and complexation with natural organic binding ligands, rather than the total metal concentration (Bruland et al., 2014), copper has been shown to impact the growth of cyanobacteria in this region (Mann et al., 2002). The highest Cu concentrations were measured along the coast, consistent with a terrestrial source. High pAl concentrations associated with these coastal sites also suggest strong lithogenic influences (Orians and Bruland, 1986). Consequently, a strong concentration gradient in dCu based on proximity to the coast could mean that correlations with eukaryotic diversity, also distinct between the coast and open ocean, are partly coincidental.

Ultimately, such corresponding patterns complicate attempts to attribute eukaryotic diversity patterns, elevated NCP, or high N<sub>2</sub> fixation to any factor or combination of factors. Nevertheless, the associations revealed by our survey characterize traits of high NCP + N<sub>2</sub> fixation sites in this region, providing a starting point for more targeted investigation of the cause(s) responsible for observed trends.

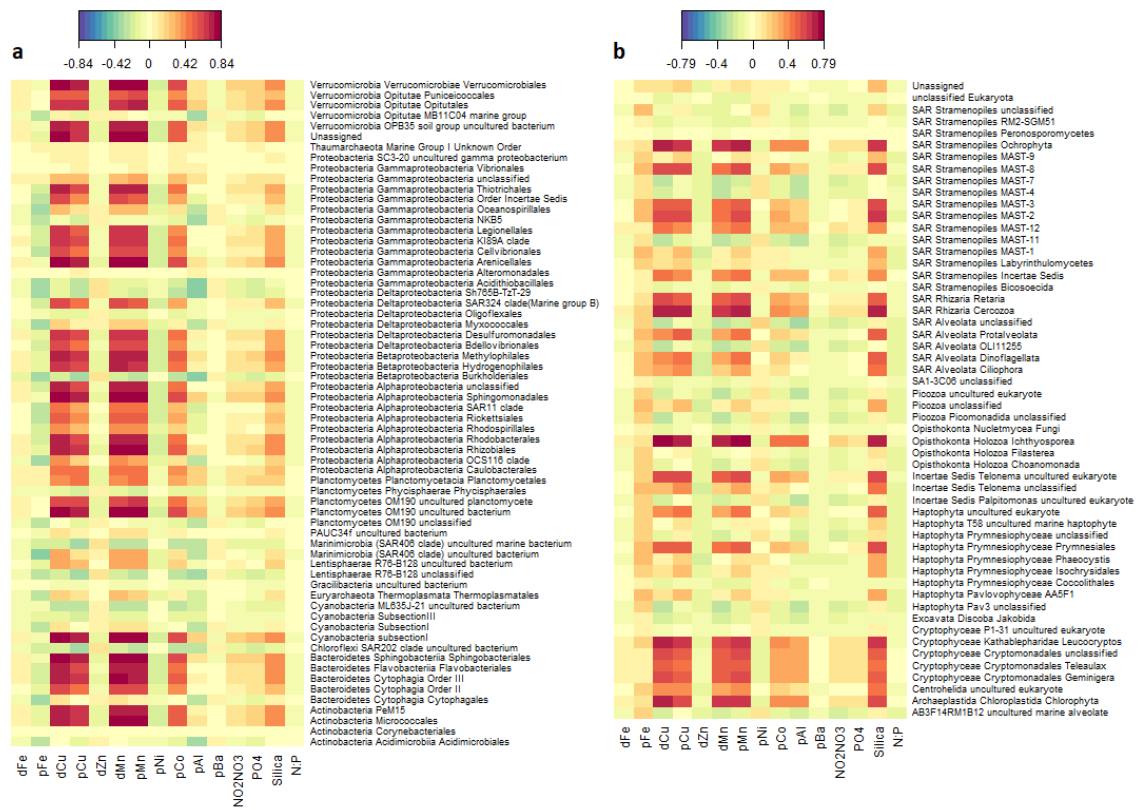


**Figure 14: Heatmaps showing correlations between abundances of specific a) prokaryotic 16S and b) eukaryotic 18S taxa and measured metadata parameters for all three years of sampling, as determined using Partial Least Squares (PLS) regression analysis. Strength of correlations is illustrated by the color scale, with positive and negative correlations shown using warm and cool tones. Cells outlined in black rectangles indicate taxa with correlations to NCP or N<sub>2</sub> fixation greater than 0.2. Taxa are binned at the fourth taxonomic rank for both 16S and 18S data.**

### **Fine-scale relationships: Key community members associated with NCP, N<sub>2</sub> fixation, and environmental properties**

Our PLS regression analyses identify NCP-associated prokaryotic and eukaryotic taxa as jointly associated with both high NCP and high N<sub>2</sub> fixation (Fig. 14). Among

prokaryotes, *Planctomycetes* abundances are strikingly related to NCP and N<sub>2</sub> fixation. Binned at the fourth taxonomic rank, two of four *Actinobacteria* taxa, 3 of 5 *Verrucomicrobia*, and 2 of 3 *Betaproteobacteria* also exhibit relationships with NCP and N<sub>2</sub> fixation, as do *Arenicellales* (*Gammaproteobacteria*), *Sphingobacteriales*, and *Cytophagia* Order III from within *Bacterioidetes*. PLS regression analyses performed at the genus level identify a strong relationship between UCYN-A, a known N<sub>2</sub> fixing clade of bacteria, and both NCP and N<sub>2</sub> fixation (Supplementary Table 2h). Prokaryotes associated with NCP and N<sub>2</sub> fixation also show strong associations with dCu, pCu, dMn, pMn, pCo, pAl, PO<sub>4</sub>, and Silica (Fig. 15).



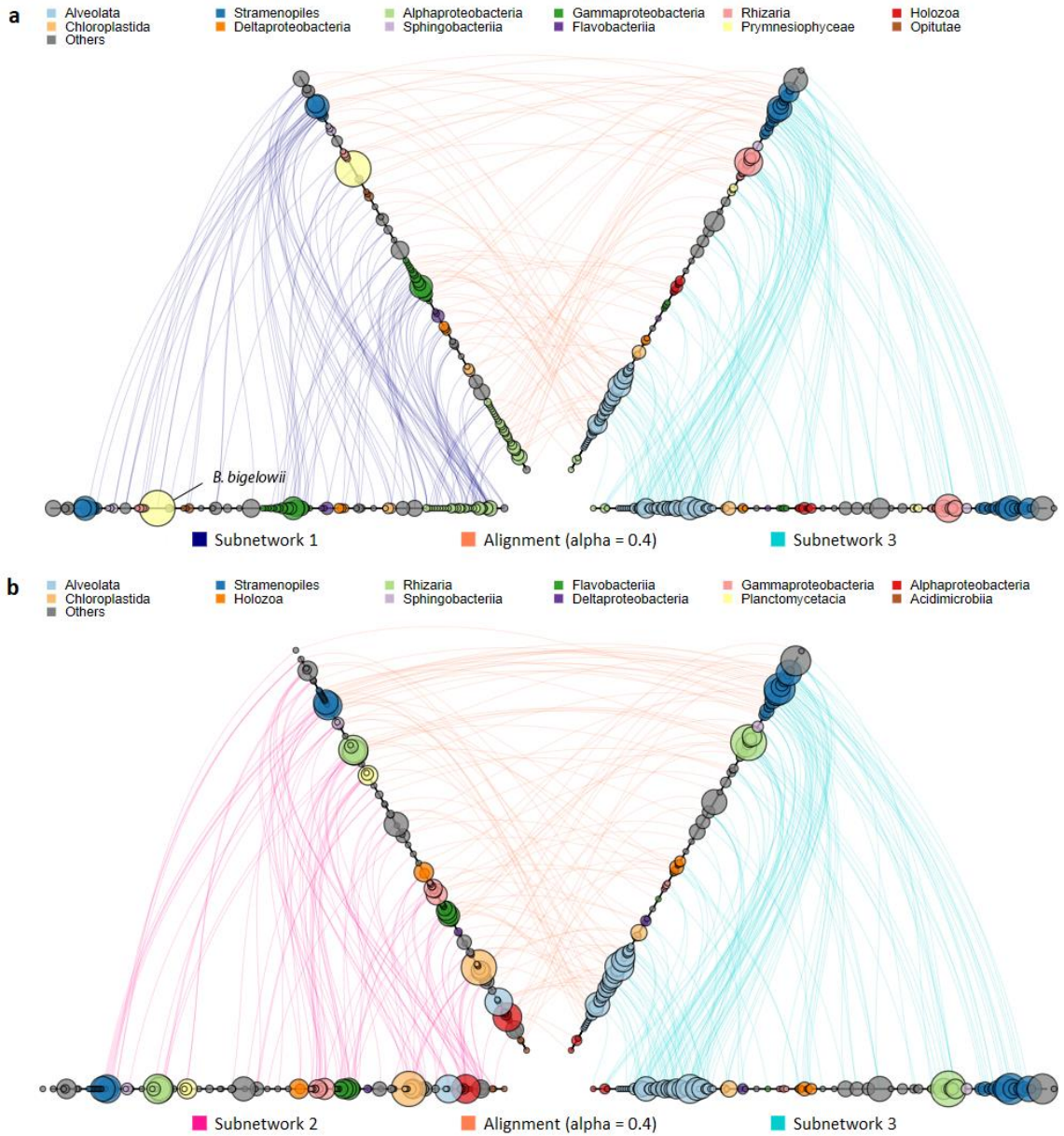
**Figure 15: Heatmaps showing PLS-derived correlations between abundances of specific a) prokaryotic 16S and b) eukaryotic 18S taxa and measured trace metal and nutrient concentrations from 2016 and 2017. Strength of correlations is illustrated by the color scale, with positive and negative correlations shown using warm and cool tones. Taxa are binned at the fourth taxonomic rank for both 16S and 18S data. *Planctomycetes* practice surface and particle-associated lifestyles (Faria et al., 2018) and exhibit high abundances in coastal environments (Pizzetti et al., 2011). Consequently, relationships with productivity and diazotrophic activity may reflect a positive response to heightened phytoplankton abundance potentially fueled in turn by N<sub>2</sub> fixation. A higher concentration of phytoplankton surfaces for growth would similarly explain NCP + N<sub>2</sub> fixation associations displayed by *Bacteroidetes* and *Verrucomicrobia*, whose marine representatives are also typically particle-associated (Yilmaz et al., 2016, Buchan et al.,**

2014, Freitas et al., 2012). Alternatively, recent research indicates that heterotrophic bacterial diazotrophs including some *Planctomycetes* are more widespread and abundant in the surface ocean than previously thought (Delmont et al., 2018), potentially contributing directly to N<sub>2</sub> fixation and NCP.

Eukaryotic groups associated with both N<sub>2</sub> fixation and NCP include most cryptophytes, *Ichthyosporea*, *Cercozoa*, *Prymnesiales*, *Retaria*, *Chlorophyta*, and several marine stramenopiles (MASTs) (Fig. 14). The taxon *Ochrophyta*, containing the dominant stramenopiles within the coastal transect—*Chrysophyceae* and *Aureococcus anophagefferens*—is also predictably related to NCP and N<sub>2</sub> fixation. At the genus level, UCYN-A's hypothesized prymnesiophyte host, *Braarudosphaera bigelowii*, is also related to N<sub>2</sub> fixation and NCP (Supplementary Table 2i). Similar to prokaryotes, N<sub>2</sub> fixation and productivity-associated eukaryotic taxa also tend to exhibit strong associations with pCu, dCu, pMn, dMn, pCo, pAl, and Silica (Fig. 15).

Despite the prevalent view that much of the surface ocean is N or Fe-limited, few to no eukaryotic or prokaryotic taxa show associations with nitrate + nitrite, N:P, or dFe (Fig. 15). The general lack of observed relationships between microplankton abundances, nitrate + nitrite, and dFe may not be anomalous, however. Raw inorganic nutrient concentrations do not necessarily reflect true biological nutrient limitation, but rather the balance between sources and sinks including biological uptake and release. Nutrient ratios can be more informative when attempting to relate community

composition to nutrient states (Mousing et al., 2018). However, we also do not observe significant relationships between the N:P ratio and community data.



**Figure 16: Alignment plots from network analysis. a) Subnetwork 1 (2015 MAB, high NCP) vs Subnetwork 3 (open ocean, low NCP) b) Subnetwork 2 (2015+2017 MAB, high NCP) vs Subnetwork 3 (open ocean, low NCP). Nodes represent taxa, colored and grouped at the third taxonomic rank in alphabetical order outwards from the origin.**

**Larger node diameter indicates higher network betweenness centrality. Within groups, taxa are ordered by centered log-ratio (CLR) normalized abundance, with more abundant genera located further from the origin. The leftmost pair of identical axes represents taxa within the first subnetwork, while the rightmost axes represent taxa from the second subnetwork. Edges between the two left or two right axes represent significant co-occurrence edges between taxa within that subnetwork. Edges between the two networks represent aligned taxa, reflecting a balance between similar taxonomic identity (sequence similarity) and topological similarity.**

Co-occurrence network analysis and alignment reveal several key eukaryotic taxa associated with NCP and N<sub>2</sub> fixation to be hubs within their subnetworks, implying that they play central community roles (Fig. 16). Subnetwork 1, most strongly associated with high productivity and N<sub>2</sub> fixation sites in the MAB in 2017 (Fig. 31), identifies marine stramenopiles as hub taxa, with four stramenopile genera including two chrysophytes placing among the top 10 keystone nodes. Notably, our results highlight *Braarudosphaera bigelowii* as the taxon with the highest centrality within its NCP and N<sub>2</sub> fixation-associated subnetwork, suggesting that this diazotroph-hosting prymnesiophyte is ecologically influential. Taxa with high network centrality are important to subnetwork stability and exhibit a high number of connections with other OTUs but, as in the case of *B. bigelowii*, do not necessarily correspond with the most abundant taxa (Figs. 10, 11). Other microplankton community studies along the California coast (Needham et al., 2018) and off Brazil (Gérikas Ribeiro et al., 2018) have also found evidence that *B. bigelowii* plays important ecological roles within other microbial ecosystems. Equally interesting are taxa associated with NCP and N<sub>2</sub> fixation in PLS regression analysis but exhibiting few sub-network connections, such as

cryptophyte algae and *Aureococcus*. Their elevated abundance in productive samples but low participation in network structure suggests that cryptophytes and *Aureococcus* benefit opportunistically from ecosystem conditions but interact minimally with other microbes in their environment.

Subnetwork 2, generally associated with MAB sampling in 2015 and 2017 (Fig. 31), reveals several chlorophyte genera—*Micromonas*, *Bathycoccus*, and *Ostreococcus*—as important keystone nodes. Given the small size and streamlined physiology of chlorophytes within *Mamiellales*, high network centrality associated with these taxa may reflect a shared response alongside other taxa to favorable environmental conditions and/or a potential role as prey within the community for larger mixotrophic and heterotrophic protists. That said, possibilities exist for more complex interactions—members of *Micromonas* in polar waters have recently been identified as mixotrophs, practicing bacterivory (McKie-Krisberg and Sanders, 2014). Our data cannot directly provide insight into these hypotheses, but the role of chlorophytes in driving co-occurrence subnetwork structure as well as their strong associations with NCP and N<sub>2</sub> fixation suggest that they too are important members of the productive coastal community in this region.

Stramenopiles, including several *Chrysophyceae*, are also identified as important taxa within the low-productivity, low N<sub>2</sub> fixation Subnetwork 3 together with alveolates, driving much of the network structure within this module. A large fraction of the

significant alignments between Subnetwork 3 and Subnetworks 1 and 2 involve stramenopiles and alveolates from Subnetwork 3. These alignment patterns suggest that the transition in community structure between low and high productivity + N<sub>2</sub> fixation sites is related to functions performed by these groups, highlighting stramenopiles in particular.

Our understanding of the function and role of stramenopiles remains incomplete, yet our results, particularly the close correspondence between peaks in N<sub>2</sub> fixation and *Chrysophyceae* abundances, suggest that stramenopiles play a key role within our study region. To our knowledge, the oceanographic literature records no known direct association between *Chrysophyceae* and diazotrophs that could explain such a pattern. Several hypotheses may explain a relationship between *Chrysophyceae* and diazotrophy, including co-variation in response to a third ecological variable, an indirect response to recycled diazotrophic N, or a direct mutualistic or commensalistic interaction between *Chrysophyceae* and nitrogen-fixers. Regardless, this striking pattern warrants more detailed investigation driven by interaction studies.

### **3.4 Conclusions**

Our study of prokaryotic and eukaryotic diversity and quantitative abundances in relation to NCP, N<sub>2</sub> fixation, and nutrient and micronutrient concentrations reveals intriguing relationships hinting at important community roles for specific taxa. We observe a strong negative PDR between O<sub>2</sub>/Ar-derived NCP and eukaryotic diversity, a

finding supportive of the view that marine PDRs may be more complex than the positive or unimodal shapes frequently proposed for surface ocean ecosystems. Notably, we observe elevated abundances of *Chrysophyceae* occurring in strikingly close conjunction with high N<sub>2</sub> fixation rates and NCP, along with *Aureococcus anophagefferens*, chlorophytes, cryptophytes, and *Planctomycetes*. Aided by our quantitative study design, our network analyses identify potential distinctions in ecological roles between highly-abundant and highly-connected taxa, highlighting *Braarudosphaera bigelowii*, chlorophytes, and *Chrysophyceae* and other marine stramenopiles as important community hub taxa. The abundances of these key taxa are also related to low temperatures, shallow mixed-layer depths, and high concentrations of phosphate, Cu, Mn, and silica. Cu-associated taxa may benefit from competitive exclusion processes, either via Cu inhibition itself or through local production of allelopathic compounds in response to Cu stress. Overall, our findings pinpoint notable regional community trends and, particularly in light of the establishment of a new LTER site along the northeast U.S. coastal shelf (NES-LTER) (2018), can serve as a springboard for targeted investigations into the processes mediated by these microorganisms of interest.

Supplementary information is available at the ISME Journal's website. Sequences and metadata are available from the NCBI Sequence Read Archive under accession numbers PRJNA509463 and PRJNA421139.

### **3.5 Acknowledgements**

This research was supported by an NSF-CAREER grant awarded to NC (#1350710) a Chateaubriand Fellowship awarded to SW, and an ANR grant awarded to HP (ANR-12-PDOC-0025-01). NC, HW, and AGG were also supported by the "Laboratoire d'Excellence" LabexMER (ANR-10-LABX-19) and co-funded by the French government under the program "Investissements d'Avenir". ED was supported by ProBioSTIC -AtlanSTIC2020. AGG also thanks the postdoctoral grant from the Universidad de Las Palmas de Gran Canaria. WT was supported by several Duke Graduate School Fellowships including The Dissertation Research Travel Award: International and the Graduate Student Training Enhancement Grant. We thank the staff of BIOS as well as the crew of the R/V Atlantic Explorer. We are also thankful to Rod Johnson, Bruce Williams, and Natasha McDonald at BIOS for cruise assistance and sample analyses, and Heidi Hirsch and the R/V Baseline Explorer's crew for help with equipment shipping. We thank Zackary Johnson for assistance with running and analyzing flow cytometry samples. We are also grateful to Nicholas Foukal, Roberta Hamme, Brivaëla Moriceau, Hans Gabathuler, Geoffrey Smith, Adeline Bidault, Jacqueline Robinson-Hamm, Adrienne Pittman, Jennifer Wernegreen, and Heather Hemric.

## **4. Lagrangian studies of net community production: assessing the effect of diel and multi-day non-steady state factors and vertical fluxes**

This chapter is currently in preparation for submission as:

Wang, S., Kranz, S., Kelly, T., Song, Hajoan., Stukel, M., Cassar, N. Lagrangian studies of net community production: assessing the effect of diel and multi-day non-steady state factors and vertical fluxes

### **Author contributions**

SW, SK, TK, and NC conceived the study. SW conducted and analyzed EIMS measurements taken during the P1604 CCE process cruise. SK and TK conducted EIMS and MIMS measurements during the P1706 cruise. TK performed NCP calculations for 2017 O<sub>2</sub>/Ar dataset, while SW calculated NCP from the 2016 O<sub>2</sub>/Ar measurements. HS generated the ROMS 4DVAR model fits for both the 2016 and 2017 cruise periods. MS and NC guided the intellectual direction of the project and provided input on analyses. All authors contributed to writing and editing of the main manuscript.

### **4.1 Introduction**

Basic measurements of oceanic primary production are essential to studying marine carbon fluxes, ecosystem dynamics, and carbon biogeochemistry. Marine net primary production, which accounts for half of total global primary production (Field et al., 1998), plays a correspondingly important role in regulating carbon fluxes between the atmosphere and the biosphere. Furthermore, marine biological uptake of inorganic carbon represents a key global climate feedback, with sequestration of sinking organic biomass into the deep ocean acting as a sink for atmospheric carbon. Improving the accuracy of marine primary production measurements can enhance our ability to both constrain the potential strength of this carbon pump to the ocean interior as well as to study marine ecosystems.

Numerous approaches for measuring marine primary production exist, spanning multiple incubation-based and *in situ* methods (Chavez et al., 2010, Ducklow and Doney, 2013); Within the last decade, an increasing number of studies e.g. (Cassar et al., 2011, Estapa et al., 2015, Eveleth et al., 2017) are utilizing underway techniques to continuously measure tracers of productivity in the surface ocean at fine resolution. Currently, an underway method in prevalent use employs measurements of the dissolved O<sub>2</sub>/Ar ratio in seawater using either Equilibrator Inlet Mass Spectrometry (EIMS) (Cassar et al., 2009) or Membrane Inlet Mass Spectrometry (MIMS) (Tortell, 2005).

The O<sub>2</sub>/Ar method is based upon isolation of the biologically-driven signal in

dissolved oxygen saturation within the surface ocean mixed layer and is widely used to generate estimates of surface net community production (NCP), defined as the difference between gross primary production (GPP) and community respiration (CR). NCP represents the net production of biomass remaining after accounting for consumption of organic carbon via respiration, and thus reflects the total biological production available for export to depth (Dugdale and Goering, 1967, Berger and Wefer, 1990, Williams et al, 1989, Ducklow and Doney, 2013). While dissolved oxygen concentrations alone are subject to physical influences (temperature, pressure, wave action) in addition to biological drivers (photosynthesis, respiration), dissolved argon is not utilized in biological processes and possesses similar physical and temperature-based solubility characteristics compared to oxygen (Craig and Hayward, 1987). Consequently, changes in the ratio of dissolved  $O_2/Ar$  ( $\Delta O_2/Ar$ ) reflect biological production/respiration over the residence time of oxygen within the surface mixed layer. By measuring the  $O_2/Ar$  ratio in surface seawater continuously, the EIMS and MIMS techniques enable the collection of large NCP datasets with unprecedented spatiotemporal resolution. These methods represent important recent methodological advances in marine biogeochemistry.

While powerful, the  $O_2/Ar$  method is subject to some limitations. The  $\Delta O_2/Ar$  tracer can be influenced by advection of neighboring or subsurface waters with contrasting  $O_2/Ar$  signatures into the volume being sampled, as well as mixing with

subsurface waters due to vertical entrainment and eddy diffusion. In addition, assumptions involved in the parameterization of the air-sea gas exchange term represent a significant source of uncertainty. Diel variations in the magnitude of  $O_2/Ar$ -derived biological oxygen saturation have also been commonly observed across a range of ocean settings (Tortell et al., 2014, Hamme et al., 2012, Ferron et al., 2015). The signal from diel cycling in biological oxygen saturation is consequently superimposed over underway measurements of  $\Delta O_2/Ar$ , thereby influencing spatial surveys based on the time of day at which observations are recorded. Such diel oscillations in biological oxygen saturation vary in amplitude with the magnitude of *in situ* primary production, ranging from ~0.2% in the oligotrophic North Pacific Subtropical Gyre (Ferron et al., 2015) to up to ~2% or more during highly-productive blooms off the Antarctic coast (Tortell et al., 2014).

The *in situ*  $O_2/Ar$  method also relies upon an assumption of steady-state, in which mixed-layer biological oxygen production is balanced by efflux to the atmosphere over the residence time of oxygen in the surface mixed layer (Reuer et al., 2007).

However, both field (Hamme et al., 2012, Tortell et al., 2014) and model-based (Jonsson et al., 2013, Tortell et al., 2014) studies have indicated that steady-state conditions can be violated in practice over diel and multi-day timescales, particularly during periods of intense or rapidly-changing biological production and/or when measurements capture physical effects such as entrainment due to changes in mixed layer depth (MLD).

Differences in calculated NCP rates can result from employment of steady-state versus non-steady-state assumptions (Martin et al., 2013).

In recent years, investigators have made efforts to better address and to correct for the abovementioned sources of error. The contribution of vertical fluxes of O<sub>2</sub>/Ar to the mixed-layer biological oxygen balance has now been estimated in field studies using a variety of approaches ranging from oxygen profile-based calculations of entrainment and vertical eddy diffusive fluxes (Castro-Morales et al., 2013, Haskell et al., 2016) to quantification of the total vertical mixing flux using an N<sub>2</sub>O-based approach (Izett et al., 2018). Furthermore, continued efforts have been made to examine the influence of the steady state assumption and to further improve the accuracy of the gas exchange parameterization term (Teeter et al., 2018). While non-Lagrangian study designs necessarily increase the difficulty of estimating the impact of diel and multi-day non-steady-state changes in biological oxygen saturation, efforts have also been made to quantify these factors in Lagrangian (Hamme et al., 2012) and stationary semi-Lagrangian (Tortell et al., 2014) field settings.

The California Current upwelling zone is a highly-dynamic coastal ecosystem characterized by steep cross-shore gradients in physical and biochemical conditions (Lynn and Simpson, 1987). Strong coastal upwelling in this region is driven by wind circulation, which generates sustained upwelling and Ekman transport along much of the California coastline (Rykaczewski and Checkley, 2008, Ohman et al., 2013). Rising

waters from depth carry high macronutrient concentrations, fueling dramatic ecosystem production rates throughout much of the growing season (Munro et al., 2013, Ohman et al., 2013). With nitrogen and phosphorus in ample supply, portions of this region have been demonstrated to exhibit iron limitation (Hutchins et al., 1998, King and Barbeau, 2007). Productivity in the California Current Ecosystem (CCE) often follows a seasonal cycle, with the peak of the spring bloom typically occurring in April-May, followed by continued high production into the summer prior to subsiding over the course of the fall to low winter productivity rates (Munro et al., 2013). Offshore and alongshore currents, influenced by both Ekman transport and meso- and submesoscale fronts and eddies also lead to substantial mesoscale spatial variability and large advective fluxes (Strub and James, 2000, Landry et al., 2012)

Given its strong vertical and lateral fluxes, considerable spatial variability, and high rates of productivity, the California Current region arguably represents a “worst-case” field environment for the assumptions made when measuring NCP using traditional underway measurements of O<sub>2</sub>/Ar. Errors associated with such conditions are often considerable, leading investigators to discard data collected within regions influenced by strong upwelling or to avoid field deployments in such regions altogether (Cassar et al., 2014, Reuer et al., 2007). Because of this practice, however, only relatively few attempts (Haskell and Fleming, 2018) have been made to quantify the sources of error within such a challenging setting to date. Such efforts have the potential to help us

more critically evaluate the strengths and weaknesses of the assumptions involved in calculating community production rates from dissolved O<sub>2</sub>/Ar.

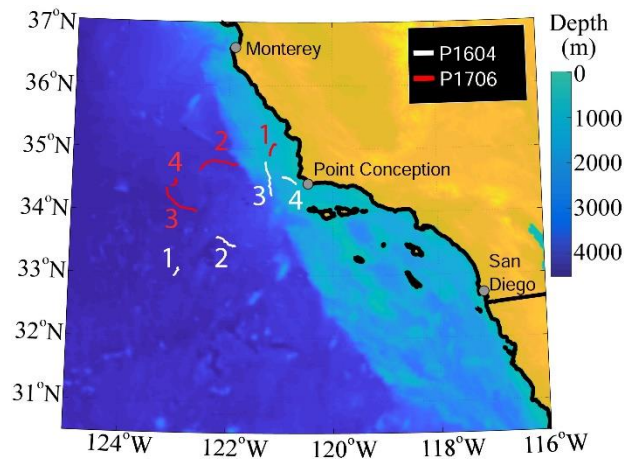
In this paper, we present Lagrangian *in-situ* measurements of O<sub>2</sub>/Ar saturation and NCP from seven multi-day deployments over two field expeditions in the California Current Ecosystem in spring 2016 and 2017. This dynamic region represents a unique study environment for assessing the respective impact of diel and multi-day non-steady-state conditions and vertical fluxes upon O<sub>2</sub>/Ar-derived estimates of NCP. By estimating the contribution over time of vertical O<sub>2</sub>/Ar fluxes as well as diel and longer-term non-steady-state conditions to the overall biological oxygen signal in various water masses within this region, we demonstrate the critical importance of accounting for such potential sources of measurement error, underlining the need for quantification of these factors via concurrent N<sub>2</sub>O sampling and/or other means when measuring NCP in similarly dynamic ocean regions around the world.

## **4.2 Materials and Methods**

### **Cruise background**

Underway measurements and CTD-Niskin rosette casts were conducted during an expedition aboard the *R/V Sikuliaq* from 19 April to 12 May 2016 during the RAPID CCE-LTER process cruise (P1604), as well as from 1 June to 2 July 2017 aboard the *R/V Roger Revelle* during the subsequent year's CCE-LTER process cruise (P1706) (Fig. 17). 3-

4 quasi-Lagrangian experiments (hereafter ‘cycles’) were conducted in each year while following selected water masses over ~3 day periods at various locations in the CCE. Following the completion of a towed Moving Vessel Profiler (MVP) survey of the local area to ensure that the deployment site represented a cohesive water mass free of strong frontal gradients (Ohman et al., 2012, Landry et al., 2012, Stukel et al., 2013), each cycle began when a drogued sediment trap array and a drogued experimental array were deployed (Landry et al., 2009, Stukel et al., 2013). Each array included a 3×1-m holey sock drogue centered at 15-m depth (Landry et al., 2009, Niiler et al., 1995).



**Figure 17: Map of cruise tracks during Lagrangian deployments in spring 2016 and summer 2017 within the California Current Ecosystem. White lines indicate observations in 2016. Red lines denote cycle locations in 2017. Numbering indicates chronological order of cycles in 2016 and 2017. Background blue color scale indicates regional bathymetry.**

In 2016, the first Lagrangian cycle was aborted due to an on-ship emergency and we do not evaluate data collected during this deployment. The second cycle was located within the heart of the California Current, while the third and fourth cycles were both

situated within the inshore zone of sustained upwelling off Point Conception (Table 1). 2016 Cycle 3 was located within the wind stress curl upwelling region, while Cycle 4 was conducted inside the coastal boundary upwelling zone. For additional details on plankton communities during the P1604 cruise, see (Morrow et al., 2018) and (Nickels and Ohman, 2018). During the 2017 cruise, the first cycle was conducted in the inshore upwelling region, the second cycle was initiated in aged upwelled waters and lasted four days, and the third cycle was conducted in post-bloom waters. The fourth cycle in 2017 lasted two days and represented a ‘continuation’ of 2017 Cycle 2 where the *Revelle* returned to a surface-tethered drifting sediment trap, which had been deployed at the end of Cycle 2, to begin Cycle 4. All other deployments in 2016 and 2017 apart from 2017 Cycle 2 and 4 were conducted for a duration of three days.

**Table 1: Basic metadata for each set of Lagrangian observations conducted in 2016 and 2017. Lat/lon coordinates, date, and time are given as the starting values for each cycle. Parameters are listed as median cycle values, with values in parentheses indicating standard deviations.**

| Cycle  | Latitude (N) | Longitude (E) | Date/Time (PST)         | O <sub>2</sub> residence time (d) | Temp (°C)  | MLD (m)     | Chl-a (µg/L) |
|--------|--------------|---------------|-------------------------|-----------------------------------|------------|-------------|--------------|
| 2016-2 | 33.59        | -122.21       | 29.Apr.2016<br>01:16:00 | 11.1 (1.0)                        | 15.3 (0.1) | 71.4 (5.2)  | 0.1 (0.0)    |
| 2016-3 | 34.8         | -121.28       | 03.May.2016<br>01:36:00 | 5.7 (0.8)                         | 13.9 (0.3) | 16.2 (2.8)  | 1.1 (0.2)    |
| 2016-4 | 34.44        | -120.69       | 07.May.2016<br>01:26:00 | 7.6 (0.5)                         | 14.7 (0.3) | 12.4 (2.2)  | 4.1 (0.8)    |
| 2017-1 | 35.06        | -121.1        | 09.Jun.2017<br>00:02:19 | 4.5 (1.5)                         | 13.2 (0.3) | 16.8 (7.2)  | 9.3 (2.7)    |
| 2017-2 | 34.7         | -121.67       | 13.Jun.2017<br>00:58:44 | 7.4 (2.0)                         | 13.2 (0.2) | 24.4 (4.2)  | 3.1 (1.0)    |
| 2017-3 | 34.26        | -123.15       | 18.Jun.2017<br>23:23:49 | 8.1 (5.1)                         | 15.5 (0.7) | 20.6 (10.4) | 0.6 (0.5)    |
| 2017-4 | 34.39        | -123.08       | 23.Jun.2017<br>18:19:58 | 3.6 (0.7)                         | 15.7 (0.6) | 12.0 (1.0)  | 0.2 (0.1)    |

## **Underway and vertical profile O<sub>2</sub>/Ar measurements via EIMS and MIMS:**

Using the equilibrator inlet mass spectrometry (EIMS) method (Cassar et al., 2009), we conducted high-resolution measurements of the dissolved O<sub>2</sub>/Ar ratio (dO<sub>2</sub>/Ar) in surface seawater. Briefly, intake seawater collected at a depth of 5 m by the ship's underway system was circulated continuously via a magnetic gear pump through a membrane contactor. Within the membrane contactor, dissolved gases from the seawater flow rapidly equilibrate within the headspace at pressures reflecting *in situ* concentrations. Equilibrated gas was measured using a Pfeiffer Prisma QMS 200 M1 quadrupole mass spectrometer. Every 2-4 hours, an automated program or user input was used to switch a multi-port Valco valve in order to sample ambient room air from a second capillary of equal length for 20-30 minutes. These "air calibrations", measuring the constant ratio of atmospheric O<sub>2</sub>/Ar, are used in downstream processing to express *in situ* O<sub>2</sub>/Ar saturation relative to the ratio at atmospheric equilibrium and to calibrate "water" O<sub>2</sub>/Ar measurements for instrumental drift.

Apart from occasional breaks in measurements for cleaning, service, or to troubleshoot technical issues, EIMS measurements were collected continuously throughout both cruises during all Lagrangian cycles. Gaps in data due to instrument downtime never exceeded 4.5 hours during Lagrangian measurements, apart from a 9.5 hour break in data during the first cycle of P1706.

Potential respiration in the ship's lines (Juraneck et al., 2010) was assessed in 2016 using five paired sets of discrete Winkler dissolved oxygen samples taken from CTD casts and from the underway seawater tap over the course of the cruise. Winkler samples from CTD casts were collected from a surface depth of around five meters. Samples were immediately fixed with pickling reagents and measured via Winkler titration at Scripps Institute of Oceanography shortly after returning to port following the procedure outlined in the CalCOFI program's Winkler sampling and analysis protocol (<http://www.calcofi.org/ccpublications/calcofi-methods/13-dissolved-oxygen.html>). Measured Winkler samples showed no consistent difference between underway and *in situ* oxygen concentrations, suggesting that the effect of microbial respiration in the *R/V Sikuliaq's* underway plumbing was negligible.

During the summer 2017 cruise, MIMS measurements of surface seawater  $dO_2/Ar$  were conducted simultaneously alongside EIMS observations. A custom-made MIMS cuvette consisted of a stainless-steel baseplate in which a frit inlay (gas permeable 50  $\mu m$  PTFE), overlaid with a 10  $\mu m$  PTFE membrane (RTC Reichelt Chemietechnik GmbH, Germany) and a flow through cuvette which transported water along the membrane. The benefit of a MIMS system is the fast response rate of a little more than 10 seconds. To avoid large temperature changes, a heating coil was wrapped around the MIMS cuvette which kept the stainless-steel base plate at 45 °C. For calibration purposes, a 4 L polycarbonate bottle of filtered seawater was equilibrated continuously using an

aquarium air pump. The calibration bottle was placed in the flow-through water intake reservoir to keep water temperatures close to the environmental temperatures. Every 4 hours, automated solenoid valves switched the water intake from the flow through to the air equilibrated water which was subsequently measured by the MIMS. A calibration step lasted 25 minutes after which the system changed back to the seawater intake. These automatic calibrations were used for the atmospheric O<sub>2</sub>/Ar ratio baseline correction. Subsequent to each calibration, the calibration water-bottle was automatically topped off with filtered water from the seawater intake using a peristaltic pump. Equilibration with room air for >3 hours and constant monitoring of the O<sub>2</sub> concentration using O<sub>2</sub> optodes (Firesting, P2 optodes) ensured that the water was equilibrated. MIMS measurements were collected continuously throughout the 2017 cruise and used in this study to conduct vertical O<sub>2</sub>/Ar profile measurements.

### **Calculation of biological oxygen saturation ( $\Delta\text{O}_2/\text{Ar}$ )**

Raw EIMS and MIMS output data was time-averaged at a two-minute timestep. Mass spectrometer data collected during testing, periods of technical failure, and during shutoffs of the ship's underway seawater flow were filtered using times recorded in a cruise log sheet. EIMS data collected during periods in which mass spectrometer pressure was anomalously high or low ( $< 1 \times 10^{-6}$  or  $> 9 \times 10^{-6}$  mbar), as well as data recorded during the 80-minute period following any instrument restart were also filtered. At times when a valve switch between air and water capillaries occurred, the nearest 6

minutes of water measurements and the nearest 4 minutes of air measurements were also filtered from the dataset to avoid transition artifacts.

Biological  $O_2$  supersaturation was subsequently calculated from the measured ion current ratio according to the following formula:

$$\text{Biological } O_2 \text{ sat (\%)} = \Delta(O_2/Ar) = \left( \frac{(O_2/Ar)_{\text{seawater}}}{(O_2/Ar)_{\text{equil}}} - 1 \right) \cdot 100 \quad (4-1)$$

where  $O_2/Ar_{\text{seawater}}$  corresponds to the oxygen-argon ratio measured from sampled seawater and  $O_2/Ar_{\text{equil}}$  represents the  $O_2/Ar$  ratio expected at air-sea gas exchange equilibrium, which is assumed to correspond to the  $O_2/Ar$  ratio found in air. A continuous vector of  $O_2/Ar_{\text{equil}}$  values was generated by calculating the mean  $O_2/Ar$  recorded during each air calibration, followed by linear interpolation of the  $O_2/Ar_{\text{equil}}$  ratio between each set of successive air calibrations. Data collected between air calibrations at which the difference in  $O_2/Ar$  exceeded 2% were filtered, as were data collected when the interpolated ion current ratio of air measurements was outside the range of 15 to 28. Following filtration of anomalous data, the  $\Delta O_2/Ar$  dataset was then merged with two-minute averaged ship data (time, position, temperature, salinity, PAR, and wind speed). MIMS data were processed following an identical pipeline, apart from adjusted criteria for filtering of poor-quality data.

Mixed-layer depth calculations, subsurface oxygen concentrations, as well as oxycline and pycnocline gradients are necessary for parameterizing the gas exchange term for NCP calculations and to estimate vertical  $O_2/Ar$  fluxes from vertical

entrainment, eddy diffusive, and advective fluxes. These parameters were calculated from CTD cast measurements conducted during both the 2016 and 2017 cruises. Quality-checked, 1 m depth-averaged CTD cast data was downloaded from the CCE LTER online database (<https://oceaninformatics.ucsd.edu/datazoo/catalogs/ccelter/datasets>). Mixed-layer depths were determined using a density difference criterion of  $0.03 \text{ kg/m}^3$  from the density measurement at a reference depth of 10 m (Dong et al., 2008). Additionally, using bisector least squares regression, oxygen and density measurements from the 20 m below the calculated MLD were used to determine subsurface gradients in oxygen and density immediately below the mixed layer. MLD values and oxygen and density gradients, were then linearly interpolated across each cycle.

As the residence time of  $\Delta\text{O}_2/\text{Ar}$  in the surface ocean is on the order of days to weeks, prior local wind speed data from the period preceding real-time measurements are required to incorporate wind speed history into parameterization of the gas exchange piston velocity. North American Regional Reanalysis (NARR) wind speeds in the 30-day period prior to the day of measurement were downloaded from the NOAA ESRL online database (<https://www.esrl.noaa.gov/psd/data/gridded/data.narr.monolevel.html>), then 2-D interpolated to yield wind speed history data along our cruise coordinates. For both years of NARR data, we compared NARR and height-corrected (Thomas et al., 2005) ship-measured wind speeds, then corrected the former using a linear model to more

closely match field wind speed data. Piston velocity was parameterized following (Wanninkhof, 2014), applying the wind speed history-based weighting scheme of (Reuer et al., 2007) with the slight modification advocated by (Teeter et al., 2018).

Uncorrected net community production rates, reflecting raw biological oxygen fluxes, were calculated using Equation 4-2:

$$NCP_{prior} (mmol O_2 m^{-2} d^{-1}) = \left( \frac{\Delta O_2 / Ar}{100} \right) * [O_2]_{sol} * \rho * k_w \quad (4-2)$$

where  $\Delta O_2 / Ar$  signifies the O<sub>2</sub>/Ar supersaturation (or undersaturation, if negative) relative to atmospheric equilibrium (see Equation 4-1),  $\rho$  represents the density of seawater (kg/m<sup>3</sup>) as calculated from the UNESCO 1983 equation of state, and  $k_w$  denotes the weighted piston velocity. Lastly,  $[O_2]_{sol}$  is the oxygen concentration at saturation, calculated from (Garcia and Gordon, 1992).

Following calculation of preliminary NCP data, a distance filter was applied to remove likely non-Lagrangian data collected at distances of more than 8 km from the sediment trap and drifter array.

As this uncorrected NCP data represents an exponentially-weighted average of NCP over the past ~30 days, we henceforth term this calculated rate as  $NCP_{prior}$  as in (Hamme et al., 2012). An open source pipeline of R scripts (DOI: [10.5281/zenodo.1219292](https://doi.org/10.5281/zenodo.1219292) , <https://github.com/tbrycekelly/MIMS-TBK> ) has been developed to perform  $NCP_{prior}$  calculations and was used to process 2017 EIMS O<sub>2</sub>/Ar data from the P1706 cruise (Kelly et al., in prep.).

## **Detrending and curve fitting of NCP data and calculation of NCP diel cycle characteristics**

To better isolate and quantify the diel pattern of  $\Delta O_2/Ar$  and measured  $NCP_{prior}$  observed at each cycle, our raw data required detrending to remove the long-term non-diel rate of change. We further matched a fitted curve to each cycle's data to model the observed diel pattern in the data.

Linear models (using the 'regstats' function in MATLAB) were fit to a subset of each cycle's data beginning and ending at the same time of day (typically around 4AM local time). Linear trends were observed for all cycles apart for the first cycle of the 2016 cruise, although the observed slopes were modest for several of the cycles. For each cycle, the trendline was subtracted from the data in order to detrend the cycle  $NCP_{prior}$  for greater ease of isolating the diel signal. All regressions were statistically significant ( $p \ll 0.01$ ).

To model the observed diel cycle of  $NCP_{prior}$  rates for each Lagrangian study, a Fourier series was fit to the corrected, detrended  $NCP_{prior}$  data using the curve fitting toolbox in MATLAB. In most cases, we selected the most parsimonious Fourier series fit that gave the correct number of diel cycles for the observation period. This generally corresponded to a four-term Fourier series fit apart from 2017 Cycle 4, where we employed a three-term fit due to the abbreviated length (two days) of this cycle. For 2016 Cycle 3 and 2017 Cycle 2, a fifth term was added to further improve the goodness-of-fit.

Local times for apparent sunrise and sunset were calculated using the NOAA ESRL Solar Calculator (<https://www.esrl.noaa.gov/gmd/grad/solcalc/>), using the cycle dates and starting coordinates of each Lagrangian cycle as inputs.

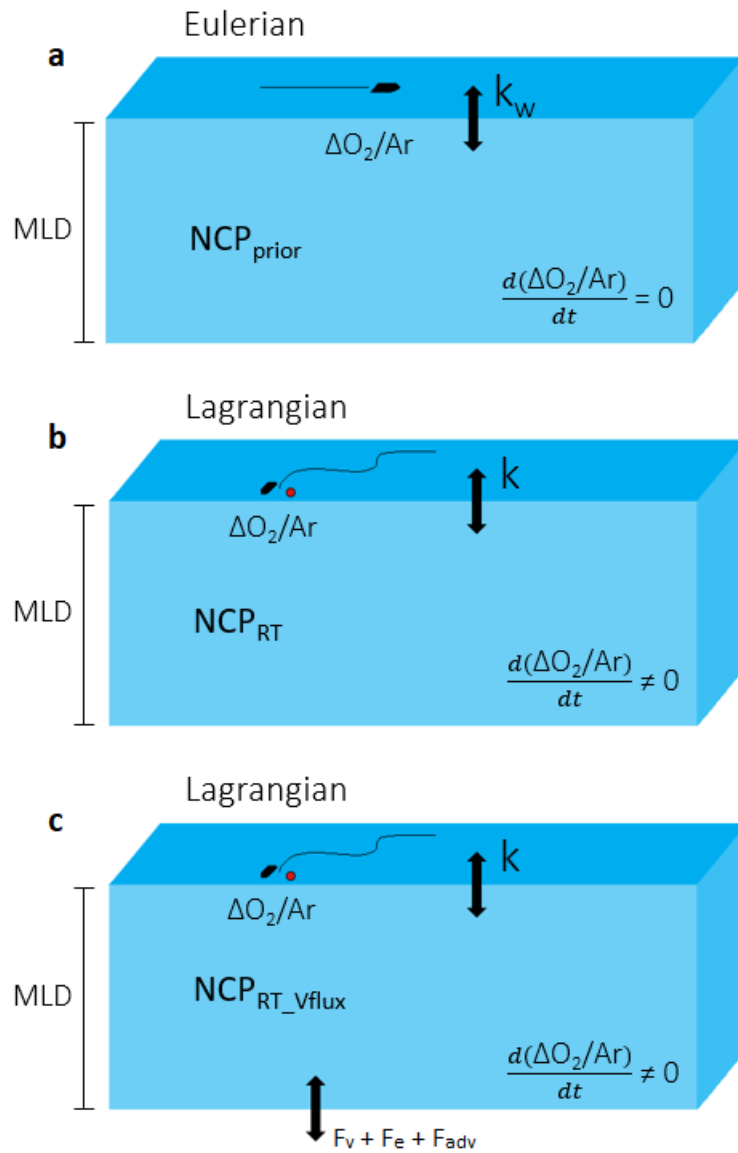
The approximate range of the diel variability in  $\Delta(O_2/Ar)$  and  $NCP_{prior}$  rates was calculated from the median of the daily maxima and minima of  $\Delta(O_2/Ar)$  and  $NCP_{prior}$  values observed over the course of a particular multi-day cycle.

### **Assessment of non-steady-state impacts on calculated net community production**

Due to our Lagrangian study design, we were able to measure the change in mixed-layer  $\Delta O_2/Ar$  in real time over the observational period and thereby estimate NCP without having to employ an assumption of steady state (Fig. 18). To do this, we utilized Equation 2 of (Hamme et al., 2012):

$$NCP_{RT} = k(\Delta O_2/Ar)[O_2]_{sol}\rho + z \frac{d(\Delta O_2/Ar)}{dt} [O_2]_{sol}\rho \quad (4-3)$$

Note that  $k$  represents the unweighted rather than weighted gas exchange coefficient here.  $z$  denotes the MLD.



**Figure 18: Conceptual diagram illustrating underlying assumptions behind the various models for  $O_2/Ar$ -derived NCP employed:  $NCP_{prior}$ ,  $NCP_{RT}$ , and  $NCP_{RT\_vflux}$ .**

Similar to when detrending uncorrected  $NCP_{prior}$  data as described above, the long-term (over the period of observation) rate of change in  $\Delta(O_2/Ar)$  was determined via a linear regression, yielding the slope of  $\Delta(O_2/Ar)$  measurements for each cycle. To

minimize the influence of diel changes in  $\Delta(\text{O}_2/\text{Ar})$  upon the calculated multi-day slope, each cycle's data were trimmed to start and end at the same time of day as previously described. All relationships obtained for  $\Delta(\text{O}_2/\text{Ar})$  linear regressions were statistically significant ( $p \ll 0.01$ ).

Following the terminology of (Hamme et al., 2012), we henceforth refer to our NCP rates calculated using the above method without the use of a steady-state assumption as real-time NCP ( $\text{NCP}_{\text{RT}}$ ).

### **Correction for vertical $\text{O}_2/\text{Ar}$ fluxes:**

Following (Emerson et al., 2008, Castro-Morales et al., 2013), the vertical flux-corrected mass balance of biological oxygen in the surface ocean can be expressed as:

$$\text{NCP}_{\text{vflux}} = \text{NCP}_{\text{prior}} + F_v + F_e + F_{\text{adv}} \quad (4-4)$$

where  $F_v$  and  $F_e$  represent the impact of vertical eddy diffusive flux and entrainment flux, respectively,  $F_{\text{adv}}$  denotes the advective flux, and  $\text{NCP}_{\text{prior}}$  represents the uncorrected NCP (see Equation 4-2). We estimated the contribution of vertical eddy diffusion and entrainment fluxes to mixed-layer biological oxygen fluxes following (Castro-Morales et al., 2013):

$$F_v = -K_z * \frac{\partial[\text{O}_2]}{\partial z \text{ oxycline}} \quad (4-5)$$

$$F_e = -\frac{1}{2} \frac{(\Delta z_{\text{mix}})^2}{\Delta t} * \frac{\partial[\text{O}_2]}{\partial z \text{ oxycline}} \quad (4-6)$$

where  $K_z$  represents the eddy diffusivity coefficient ( $\text{m}^2 \text{s}^{-1}$ ),  $\frac{\partial[\text{O}_2]}{\partial z}_{\text{oxycline}}$  the gradient of oxygen in the 10 m beneath the MLD, and  $\Delta z_{\text{mix}}$  the change in MLD over the time interval between CTD casts ( $\Delta t$ ). Negative values for  $F_v$  and  $F_e$  represent vertical addition of biological  $\text{O}_2$  to the mixed layer, which must then be subtracted to apply the correction.

Advective flux of biological oxygen into the mixed layer from the subsurface was calculated using the following equation:

$$F_{adv} = -w \cdot \frac{\partial[\text{O}_2]}{\partial z}_{\text{oxycline}} \cdot z_{\text{mix}} \quad (4-7)$$

where  $w$  denotes the vertical velocity (positive if upwards) and  $z_{\text{mix}}$  represents the mixed-layer depth.

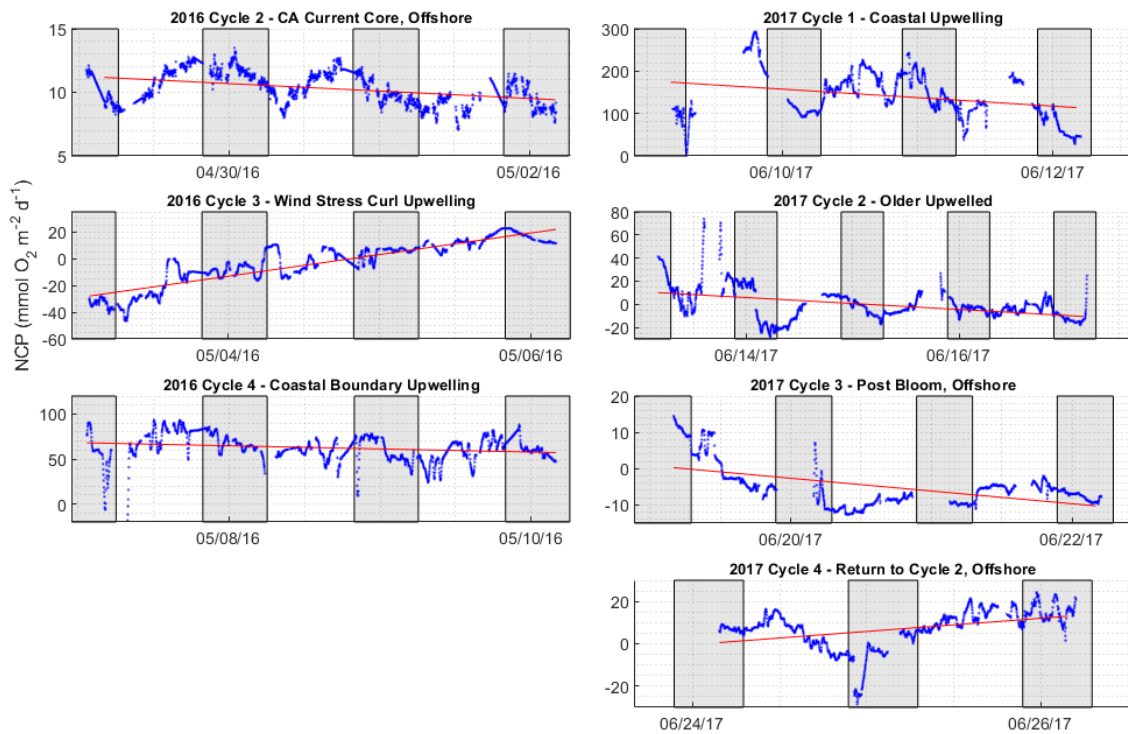
Values for  $K_z$  and  $w$  were extracted from a three-dimensional physical model (9 km horizontal resolution, 42 vertical layers) developed using the Regional Ocean Modeling System (ROMS) four-dimensional variational data assimilation system (4DVAR) (Moore et al., 2011), then depth-interpolated linearly to calculate values at the MLD. Initial and boundary conditions were taken from the California Current System Reanalysis product produced by the University of California, Santa Cruz. Surface forcing was taken from the 9-km resolution Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS) (Doyle et al., 2009, Hodur, 1997). We assimilated satellite remote sensing products (sea surface height from Ssalto/Duacs and sea surface temperature from AVHRR) and in situ temperature and salinity data vertical profiles

collected during CTD measurements on our cruise. Assimilated data were used to update model initial conditions and surface forcing to generate a dynamically-consistent month-long model fit covering the duration of each cruise. Following iterative updating of the incremental model, we obtained a reduction of the cost function, a metric reflecting the change in the model from its initial conditions as well as its success in minimizing differences between the model state and the observational dataset, of 72% (Fig. 33). For detailed descriptions of our ROMS 4DVAR approach, see (Song et al., 2012) and (Miller et al., 2015).  $K_z$  and  $w$  were interpolated to CTD cast locations from the 9-km ROMS 4DVAR solution using bilinear 2D interpolation from the four points comprising the nearest neighbors to the cast coordinates.  $K_z$  and  $w$  estimates were then linearly interpolated between CTD casts similar to the method employed for MLD values.

Incorporating these vertical flux estimates into our calculation of  $NCP_{RT}$  described above yields:

$$NCP_{RT\_Vflux} = k(\Delta O_2/Ar)[O_2]_{sol}\rho + z \frac{d(\Delta O_2/Ar)}{dt} [O_2]_{sol}\rho + F_v + F_e + F_{adv} \quad (4-8)$$

$NCP_{RT\_Vflux}$  thereby captures both the long-term rate of change in biological oxygen saturation over the observational period and the influence of vertical fluxes via eddy diffusion, entrainment, and advection.



**Figure 19: NCP<sub>prior</sub> measured for each Lagrangian cycle (black points). The line of best fit for each cycle’s NCP data, derived from a linear regression against a subset of the cycle data beginning and ending at the same local time of day, is shown in red. Shaded sections of background indicate nighttime periods.**

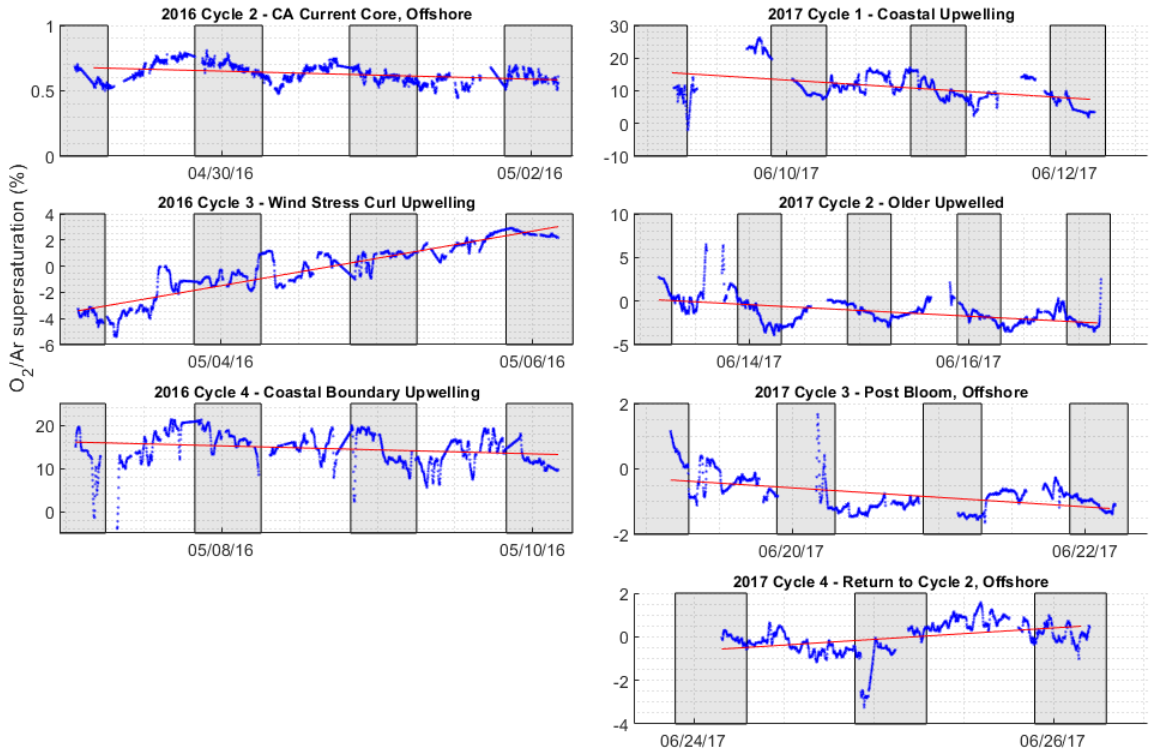
### **4.3 Results**

#### **General patterns of net community production from Lagrangian measurements in the California Current in spring/summer 2016 and 2017**

In both 2016 and 2017, mean MLD-integrated NCP<sub>prior</sub> values observed during each cycle increased with increasing proximity to shore (Table 2, Fig. 19). For both years, the cycle conducted closest to shore exhibited median NCP<sub>prior</sub> rates of 62 and 116 mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup> for 2016 Cycle 4 and 2017 Cycle 1, respectively. Located just off the Point Conception upwelling center, these locations were distinguished by high Chl-*a*

concentrations, peaks of O<sub>2</sub>/Ar supersaturation (>15% supersaturated), and large diel variation in the O<sub>2</sub>/Ar signal (Fig. 20). The range of NCP<sub>prior</sub> for a given cycle also broadened for cycles located closer to shore. Throughout the day, coastal NCP<sub>prior</sub> rates could exhibit a surprising degree of volatility, with rapid short-term changes in biological oxygen saturation of as much as 10% occurring at points during the observation period. Ecologically, we observed this coastal region to be a productive ecosystem rich in marine life. Algal biomass was high in the surface layer (~4 µg Chl-*a* L<sup>-1</sup>), with high abundances of krill, copepods, and even tuna crabs (*Pleuroncodes planipes*) choking the *R/V Sikuliaq*'s sea strainers during 2016 Cycle 4. During the 2017 cruise the coastal region had high chlorophyll (surface Chl averaged 9 µg Chl-*a* L<sup>-1</sup>) and high nutrient concentrations (surface NO<sub>3</sub><sup>-</sup> averaged 6 µmol L<sup>-1</sup>). Data on the heterotrophic

plankton community for the 2017 cruise was not yet available but will eventually be published on the CCE-LTER Datazoo database.



**Figure 20: Time series of  $\Delta O_2/Ar$  observations for each cycle (blue). The line of best fit for each cycle's  $\Delta O_2/Ar$  data, derived from a linear regression against a subset of the cycle data beginning and ending at the same local time of day, is shown in red. Shaded sections of background indicate nighttime periods.**

2016 Cycle 2, conducted in the heart of the California Current, was unique in several respects. For one, Cycle 2 displayed a narrow range of low  $NCP_{prior}$  rates (7 to 14  $mmol O_2 m^{-2} d^{-1}$ ) compared to all other cycles in both years (range for all other cycles was -47 to 242  $mmol O_2 m^{-2} d^{-1}$ ). This location was also characterized by deeper mixed-layer depths (54 to 81 m) than all other cycles (10-40 m; Fig. 32).

Cycles conducted further offshore (2016-3, 2017-2, 2017-3, 2017-4) shared similarities in terms of  $NCP_{\text{prior}}$  rate ranges with values generally varying between -20 and 20  $\text{mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$ .

**Table 2: Derived median community production rates, vertical flux magnitudes, oxygen gradients, and regression-derived slopes in  $\Delta\text{O}_2/\text{Ar}$  and NCP for each cycle. Values in parentheses indicate standard deviations. All NCP and vertical flux units are in  $\text{mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$ .**

| Cycle  | $NCP_{\text{prior}}$ | $NCP_{\text{RT}}$ | $NCP_{\text{RT\_Vflux}}$ | $F_v$   | $F_e$ | $F_{\text{adv}}$ | $\text{O}_2$ gradient<br>( $\text{mmol O}_2 \text{ m}^{-4}$ ) | Slope in $\Delta\text{O}_2/\text{Ar}$<br>(% $\text{d}^{-1}$ ) | Slope in $NCP_{\text{prior}}$<br>( $\text{mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$ ) |
|--------|----------------------|-------------------|--------------------------|---------|-------|------------------|---|---|--|
| 2016-2 | 10 (1)               | 9 (7)             | -240 (416)               | 0 (1)   | 0 (0) | -245 (434)       | 0.2 (0.2)   | -0.03   | -1   |
| 2016-3 | -1. (16)             | 88 (17)           | 251 (99)                 | 2 (1)   | 0 (1) | 154 (95)         | -0.7 (0.3)  | 2.09  | 16   |
| 2016-4 | 62 (15)              | 1 (19)            | 1646 (350)               | 12 (2)  | 0 (0) | 1626 (335)       | -4.9 (0.9)  | -0.92   | -3   |
| 2017-1 | 116 (50)             | -124 (52)         | 307 (249)                | 11 (11) | 0 (4) | 389 (227)        | -4.1 (0.9)  | -2.71   | -20  |
| 2017-2 | -12 (14)             | -51 (9)           | 255 (249)                | 7 (2)   | 0 (2) | 285 (247)        | -4.2 (0.9)  | -0.66   | -5   |
| 2017-3 | -5 (5)               | -15 (8)           | -27 (97)                 | 0 (2)   | 0 (2) | -10 (101)        | 0.1 (0.3)   | -0.29   | -4   |
| 2017-4 | 1 (6)                | 16 (1)            | 155 (101)                | 7 (4)   | 0 (0) | 131 (97)         | -0.8 (0.5)  | 0.52  | 6  |

### **Impact of steady-state versus non-steady-state-based calculation of *in-situ* NCP**

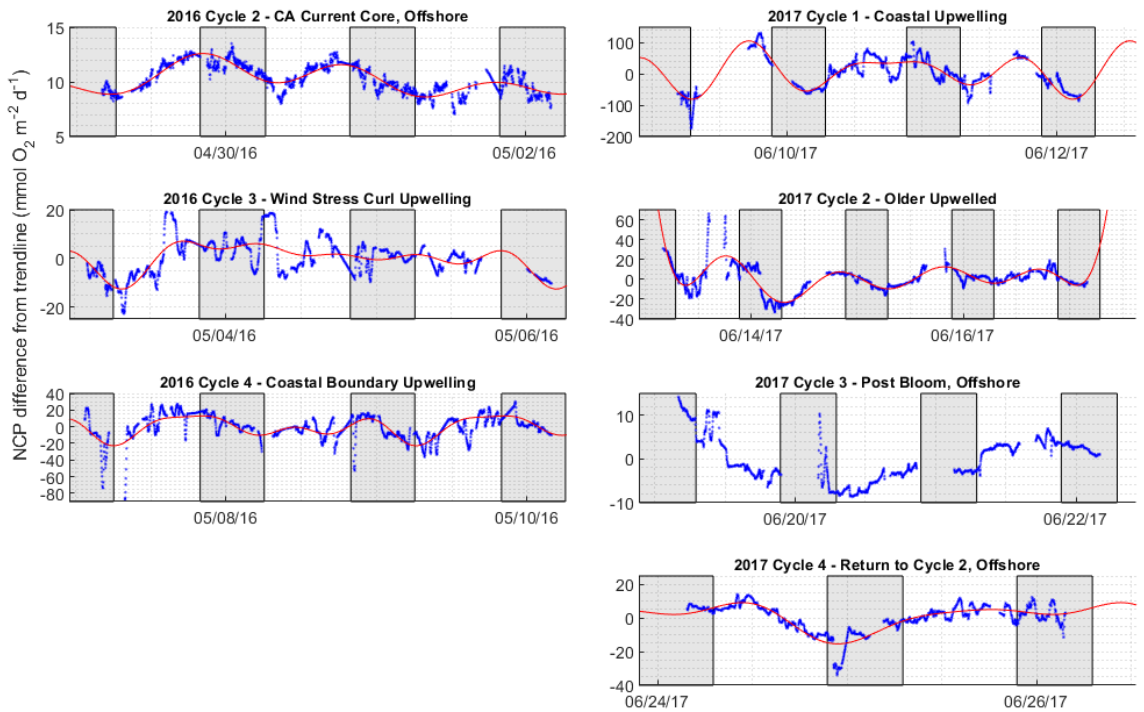
Non-steady state changes in mixed-layer biological oxygen implied by the slopes of the linear fits suggest that non-steady state conditions were non-negligible when integrated over multiple days, with changes in  $\Delta\text{O}_2/\text{Ar}$  ranging from -4.7 to 2.1 %

saturation  $d^{-1}$  at some stations. The observed rates of change in the  $NCP_{prior}$  signal across the cycles corresponded to slopes varying between -20 and 16  $mmol\ O_2\ m^{-2}\ d^{-2}$  (Table 2).

The secular increases and decreases in mixed-layer biological oxygen during a cycle resulted in corresponding offsets between  $NCP_{RT}$  and  $NCP_{prior}$  proportional to the magnitude of the slope in  $\Delta O_2/Ar$ . These offsets were more modest (i.e.  $<15\ mmol\ O_2\ m^{-2}\ d^{-1}$  difference in median  $NCP_{prior}$  versus  $NCP_{RT}$ ) for stations located further offshore (2016-2, 2017-3, 2017-4) than for stations closer to the coast, where median real-time NCP over the measurement period differed from  $NCP_{prior}$  by up to an order of magnitude and could even flip estimated NCP rates from positive to negative. For instance,  $NCP_{RT}$  calculations for 2017 Cycle 1 produce an overall strong negative biological oxygen flux, while  $NCP_{prior}$  yields a strong positive flux (Table 2).

### **Diel variation in $O_2/Ar$ -derived NCP signal**

Clear diel patterns were evident in both measured  $\Delta O_2/Ar$  saturation and calculated  $NCP_{prior}$  rates across six of the seven sites occupied (Fig. 21). Additionally, we observed a consistent pattern of increasing amplitude to the diel signal in  $O_2/Ar$  with closer proximity to shore. Increasingly larger diel signal amplitudes also tended to correspond to stations at which more elevated Chl-*a* concentrations were also observed.



**Figure 21: NCP<sub>prior</sub> measured for each Lagrangian cycle and detrended using the linear fits previously obtained (blue). The curves corresponding to the Fourier series functions chosen to fit the diel pattern in the NCP data are shown in red. Shaded sections of background indicate nighttime periods.**

At the two stations closest to the coast, diel variations in biological oxygen saturation were dramatic, leading to approximately twofold changes in  $\Delta O_2/Ar$  (%) over the course of a daylong period. The range of the diel signal at these inshore stations was also large: regularly over 10% and as large as 17%. Such oscillations are similar in magnitude to diel changes in  $\Delta O_2/Ar$  (%) observed off the coast of Palmer Station on the West Antarctic Peninsula in summer (Tortell et al., 2014). This diel variability resulted in daily shifts of apparent measured NCP<sub>prior</sub> of up to  $\sim 200$   $\text{mmol O}_2 \text{ m}^{-2} \text{ d}^{-2}$  (Table 3). A greater degree of day-to-day variability in the magnitude of diel variation was also

shown at inshore sites, with the amplitude of the diel pattern exhibiting changes of up to twofold throughout the observational period.

**Table 3: calculated median diel range of diel cycles in  $\Delta O_2/Ar$  and NCP observed during each cycle alongside median volumetric and integrated Chl-*a*.**

|        | $\Delta O_2/Ar$ range (%) | NCP range (mmol $O_2 m^{-2} d^{-1}$ ) | Chl- <i>a</i> ( $\mu g/L$ ) | MLD-integrated Chl- <i>a</i> ( $\mu g/m^2$ ) |
|--------|---------------------------|---------------------------------------|-----------------------------|--|
| 2016-2 | 0.3                       | 5                                     | 0.1 (0.0)                   | 10000  |
| 2016-3 | 3.1                       | 27                                    | 1.1 (0.2)                   | 17000  |
| 2016-4 | 17.6                      | 79                                    | 4.1 (0.8)                   | 50000  |
| 2017-1 | 14.6                      | 245                                   | 9.3 (2.7)                   | 160000                                       |
| 2017-2 | 3.8                       | 31                                    | 3.1 (1.0)                   | 76000  |
| 2017-3 | 2.6                       | 16                                    | 0.6 (0.5)                   | 11000  |
| 2017-4 | 2.1                       | 25.4                                  | 0.2 (0.1)                   | 2800   |

In contrast, the cycles furthest offshore also exhibited clear diel behavior, but with substantially smaller amplitude (changes in  $\Delta O_2/Ar$  of <1 to 2 % saturation). Nevertheless, such diel patterns are significant relative to the margin of error inherent to the method (~ 0.3 % saturation).

Only one study location, 2017 Cycle 3, did not demonstrate diel patterns with the same degree of clarity. This absence of a clear diurnal behavior for this cycle was related to a strong multi-day trend in the  $\Delta O_2/Ar$  signal (Fig. 20) possibly connected to an increase in wind speed during this cycle.

### **Influence of vertical fluxes of $O_2/Ar$ upon calculated in situ NCP**

Relative to the difference between  $NCP_{prior}$  and  $NCP_{RT}$ , the influence of vertical  $O_2/Ar$  fluxes was also large (Table 2). Consequently, differences between  $NCP_{RT}$  and  $NCP_{RT\_Vflux}$  are substantial, and the offset between  $NCP_{RT\_Vflux}$  and  $NCP_{prior}$  is significantly

and often dominantly driven by the vertical flux terms in addition to the non-steady-state term.

Across all cycles, the influence of vertical advection drove patterns in  $NCP_{RT\_Vflux}$ . Calculated vertical advective fluxes were generally large, in excess of  $200 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$  at four of seven cycles with a median  $F_{adv}$  of  $1600 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$  estimated for P1604 Cycle 4. Advective fluxes at each cycle were also highly variable, with large calculated standard deviations and broad ranges of several hundred  $\text{mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$ . These values are however highly sensitive to the vertical velocity estimates used in calculations and to the strength of and variations in the subsurface  $\text{O}_2$  gradient, which likely drive much of the variability in calculated advective fluxes.

Vertical eddy diffusive fluxes were substantially less influential but typically exerted a greater influence on the  $NCP_{RT\_Vflux}$  than entrainment fluxes, which were virtually negligible by comparison. As entrainment flux only comes into effect upon a deepening of the surface mixed layer, entrainment was a non-factor for some cycles during which mixed layer depths predominantly shoaled or remained stable during the observed period. In contrast, vertical eddy diffusion exhibited greater variability but consistently increased in both the median magnitude and range of values with closer proximity to shore (Table 2).

## **4.4 Discussion**

### **Non-steady-state Lagrangian versus conventional Eulerian calculation of NCP<sub>prior</sub>**

The large difference in median cycle NCP when comparing NCP<sub>prior</sub> calculated using an exponential wind speed history weighting to real time NCP<sub>RT</sub> is primarily driven by the magnitude of the trend in  $\Delta O_2/Ar$  over the length of observations. While changes in measured MLD do drive excursions in NCP<sub>RT</sub>, the overall offset between NCP<sub>RT</sub> and NCP<sub>prior</sub> remains generally dominated by the slope produced by the long-term rate of change in  $\Delta O_2/Ar$  over the 2-4 day period. A multi-day decrease in  $\Delta O_2/Ar$  over time results in a negative offset in which NCP<sub>prior</sub> rates are significantly greater than NCP<sub>RT</sub> values, and vice versa. Consequently, the steady-state assumption and NCP<sub>prior</sub> agree well with NCP<sub>RT</sub> when the rate of change in  $\Delta O_2/Ar$  is small, as for 2016 Cycle 2 (Table 2).

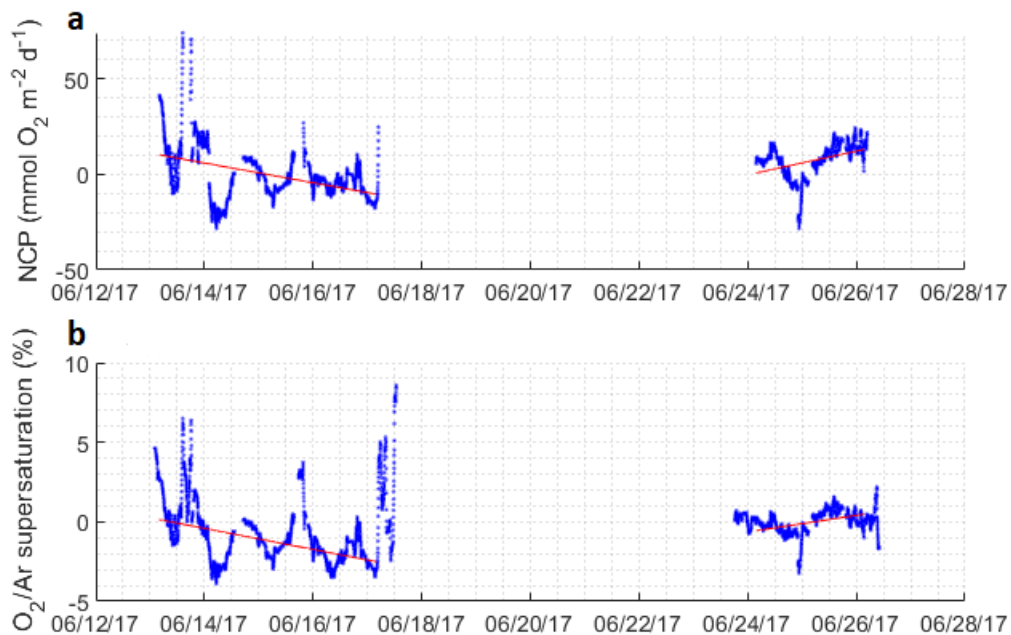
Prior studies observing non-steady-state changes in  $\Delta O_2/Ar$  over time have attributed part of the rate-of-change term to changes in the biological community, as evidenced by patterns in Chl-*a* over the observational period (Hamme et al., 2012). Based on supporting Advanced Laser Fluorometry (ALF) measurements from our 2017 cruise (available from the CCE-LTER website), however, we do not observe any multi-day increases or decreases in Chl-*a* that might explain the observed non-steady-state trends in biological oxygen saturation. Intriguingly, we do observe a correspondence between the magnitude of the diel signal changes in  $\Delta O_2/Ar$  and diel behavior in Chl-*a* across our

cycles (Fig. 34), qualitatively supporting the relationship between POC (beam attenuation) and  $^{14}\text{C}$  primary production rates established in a study in the North Pacific Subtropical Gyre (White et al., 2017).

It is important to note that  $\text{NCP}_{\text{prior}}$  and  $\text{NCP}_{\text{RT}}$  are not directly comparable (apart from the differences in terms of the steady-state assumption) as the integration time for each metric is also distinct. While  $\text{NCP}_{\text{prior}}$  purports to integrate over the residence time of  $\text{dO}_2$  in the surface mixed layer (several days to weeks),  $\text{NCP}_{\text{RT}}$  is intended to reflect the rate of production over the period of observed measurements only. As a result, there are two potential factors driving differences between  $\text{NCP}_{\text{prior}}$  and  $\text{NCP}_{\text{RT}}$ : either trends in  $\Delta\text{O}_2/\text{Ar}$  over the timescale of the  $\text{O}_2$  residence time are generally different from the slope observed over a period of just 2-4 days, or accounting for non-steady-state terms is critical. Both effects could simultaneously be at play. Potential follow-up work might seek to quantify both  $\text{NCP}_{\text{prior}}$  and  $\text{NCP}_{\text{RT}}$  in a Lagrangian setting for multiple weeks, thereby filtering out the effect of different times of integration.

We acknowledge that our Lagrangian time series observations were of relatively short duration and prevent us from observing if multi-day changes in surface layer  $\text{O}_2/\text{Ar}$  saturation would have been sustained or significant throughout the residence time of  $\Delta\text{O}_2/\text{Ar}$ . In particular, 2017 Cycle 4 was occupied for a period of just two days compared to the 3-4 day study length employed at other sites. Notably, this cycle, intended as an extension of 2017 Cycle 2 several days later, differs from Cycle 2

considerably in terms of the implied rates of change in  $\Delta O_2/Ar$  observed during the two respective cycles (Fig. 22). This observation indeed suggests that short deployments (2-3 days) may potentially overestimate rates of change in mixed-layer biological oxygen saturation compared to calculations using longer-term (weeklong or greater) measurements. Additionally, such longer timescales are generally more characteristic of the residence time of biological oxygen in the surface ocean. The true rate of change in  $\Delta O_2/Ar$  over longer observational periods might result in a substantially smaller non-steady-state term for  $NCP_{RT}$ . This would be consistent with the additional time needed for air-sea gas exchange to balance faster-acting biological fluxes.



**Figure 22: Time series of a)  $NCP_{prior}$  and b)  $\Delta O_2/Ar$  supersaturation for Cycles 2017 #2 and #4. Red line segments illustrate best-fit linear regression lines and trends for each multi-day series of observations.**

We also do reemphasize that the California Current system in many ways represents a worst-case environment for the deployment of EIMS and MIMS-based  $O_2/Ar$  measurements. Upwelling waters from depth in this region are often highly undersaturated in  $O_2/Ar$  and might violate steady-state conditions frequently. Under certain local conditions, ongoing equilibration between the undersaturated surface mixed layer and the atmosphere results in apparent  $O_2/Ar$  undersaturation even while mixed layer production rates are high and positive. Alternatively, elevated productivity in the surface mixed layer might contribute to high  $O_2/Ar$  supersaturation despite the presence of a strong influence from vertical fluxes of low  $O_2/Ar$  waters. Multi-day trends in  $O_2/Ar$  can thus be driven by a combination of biology as well as the physics of air-sea gas exchange. Finally, community metabolic rates along this highly-productive coast might be of a sufficiently dramatic magnitude to unbalance air-sea-exchange from net production/respiration even in the absence of vertical fluxes. Investigators have long cautioned against deploying the  $O_2/Ar$  method in environments characterized by such factors (Haskell and Fleming, 2018, Castro-Morales et al., 2013, Reuer et al., 2007), and indeed, traditional steady-state assumptions appear to line up much more closely with  $NCP_{RT}$  at greater distance offshore, an environment more typical of settings where the  $O_2/Ar$  technique has been employed to date.

## **Influence of diel patterns in O<sub>2</sub>/Ar upon measured NCP<sub>prior</sub>**

Our observation of strong diel signals within Lagrangian observations of O<sub>2</sub>/Ar saturation reinforce earlier findings (Hamme et al., 2012, Ferron et al., 2015, Tortell et al., 2014) that diel effects can be significant across diverse environments. Consequently, the time of day can matter considerably when conducting non-Lagrangian measurements, affecting measured values by a factor of two or more. We demonstrate that the amplitude of the diel signal increases consistently along a cross-shore gradient in our study region during the growing season, with the effect exhibiting an increased magnitude for a higher productivity regime.

Using our measurements as an example, this phenomenon means that were a vessel conducting Eulerian measurements (conventional location-based sampling without following the paths of particular water parcels) of NCP to arrive on-station close to the coast at 3AM as opposed to 3PM local time, the difference in measured NCP calculated from the  $\Delta\text{O}_2/\text{Ar}$  signal might be as large as 20-50 mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup>: an effect approaching 50% of the median NCP measured at inshore stations. Ultimately, this diel factor biases NCP measurements low if conducted during the night and high if taken during peak daylight hours. However, the magnitude of the bias scales proportionally to the total metabolic rate of the surface ocean community, and thus the absolute effect of diel variations in O<sub>2</sub>/Ar is low in less-productive settings. Nevertheless, due to the lower

overall  $NCP_{\text{prior}}$  rates, diel effects remain a large source of variability in terms of relative proportions.

A consistent trend between both volumetric and MLD-integrated Chl-*a* measurements and diel  $O_2/Ar$  signal amplitude was observed in both years of our study (Table 3). Consequently, volumetric or depth-integrated Chl-*a* may potentially serve as the basis for a field indicator of potential diel variability in biological oxygen saturation. In this study, we observe that median cycle Chl-*a* concentrations above 1  $\mu\text{g/L}$  correspond to diel variability in  $\Delta O_2/Ar$  on the order of 3% in amplitude, suggesting that diel effects are of potential concern for most “productive” ecosystems. However, more Lagrangian observations will likely be required to develop a robust metric that might be used to gauge the risk of significant diel behavior in  $O_2/Ar$ .

Additionally, we point out that a number of other factors not captured by the surface Chl-*a* inventory may also be responsible for driving diel signals. For instance, it is possible that diel variations in respiration due to heterotrophic activity, potentially due to zooplankton vertical migration or other factors, may also explain a component of the diel  $O_2/Ar$  signal, confounding efforts to assess the diel signal amplitude in relation to Chl-*a* alone. Additionally, the magnitude of diel behavior in  $O_2/Ar$  is also determined by the extent of gas exchange between the atmosphere and surface ocean, where more vigorous air-sea equilibration may act to dampen the diel cycle. Finally, the influence of

other factors like PAR, light attenuation, and tidal effects may also impact diel variability by influencing the biological community.

Unfortunately, it remains difficult to say how investigators can correct for this diel effect while conducting Eulerian measurements. When deploying underway  $O_2/Ar$  measurements at sites of particular interest, a best practice might be to occupy the station for a period extending through a 12-hour period in order to capture the daily range and median of the  $O_2/Ar$  signal. However, in practice such an approach may not be possible due to time and logistical constraints. Ultimately, our results re-emphasize that such diel effects can be considerable and ought to be factored in when interpreting non-Lagrangian measurements of NCP determined using the underway  $\Delta O_2/Ar$  method.

### **Influence of vertical fluxes of $O_2/Ar$ upon apparent $NCP_{prior}$**

Overall, we assess the impact of vertical fluxes of  $O_2/Ar$  to be a dominant and large influence upon calculated  $NCP_{RT\_vflux}$ , resulting in substantial differences between calculated  $NCP_{RT\_vflux}$  and  $NCP_{prior}$  productivity rates. The largest vertical flux term by a considerable margin comprised the effect of vertical advection, which we calculated to potentially produce fluxes several-fold to a full order of magnitude larger than uncorrected  $NCP_{prior}$  rates (Table 2). This finding parallels the conclusions of (Haskell and Fleming, 2018), who calculated vertical transport fluxes of 50-124  $mmol O_2 m^{-2} d^{-1}$  off the California coast in spring 2013 and 2014 that left uncorrected NCP estimates just 15-

16% of corrected NCP rates. The dominant impact of vertical fluxes is also supported by N<sub>2</sub>O-calculated mixing corrections for O<sub>2</sub>/Ar NCP measurements off British Columbia that yielded vertical flux contributions of 150-200 mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup> in coastal shelf waters in spring and summer that significantly elevated corrected NCP rates by several-fold relative to uncorrected NCP estimates (Izett et al., 2018).

Our sizable vertical flux corrections, driven by very large advective contributions of hundreds to thousands of mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup>, produce extremely elevated corrected median NCP<sub>RT\_vflux</sub> rates at some coastal stations (307 mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup> at P1706 Cycle 1 and 1646 mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup> at P1604 Cycle 4). <sup>14</sup>C primary production measurements conducted using incubations during the same expedition in 2016, in contrast, yielded a maximum integrated primary production rate equivalent to 224 mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup> at P1604 Cycle 4. Such sizable discrepancies are likely the product of uncertainty in vertical advective flux estimates, which are potentially overestimating advective contributions of undersaturated O<sub>2</sub>/Ar to the surface layer. With vertical velocities near the base of the mixed layer in this region reaching a maximum of 31 m d<sup>-1</sup> (but with more typical values of 3 to 29 m d<sup>-1</sup>), advection certainly represents a considerable factor to correct for, but is itself difficult to accurately constrain. At the same time, however, measurements of primary productivity using <sup>14</sup>C-PP incubations have yielded rates equivalent to 166-576 mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup> in this region during spring in other years (Stukel et al., 2011),

suggesting that only the highest positive  $\text{NCP}_{\text{RT\_Vflux}}$  values in our study are in excess of plausible rates.

In contrast, our calculated negative  $\text{NCP}_{\text{RT\_Vflux}}$  rates for P1604 Cycle 2, caused by advective fluxes bringing supersaturated waters into the mixed layer from depth, imply an implausible degree of heterotrophy that suggests the advective flux term determined at this station are unreliable. This likely results from the variable  $\text{O}_2$  profile and weak gradients in subsurface  $\text{O}_2$  at this station, combined with dramatic upwelling rates that greatly amplify fluxes calculated using these subsurface gradients. Given the sensitivity of our calculation to the strength and sign of the gradient used, our method for estimating the advective flux performs particularly poorly in this location. At any rate, modeled vertical velocities of such magnitude reemphasize the importance of vertical advection in this region.

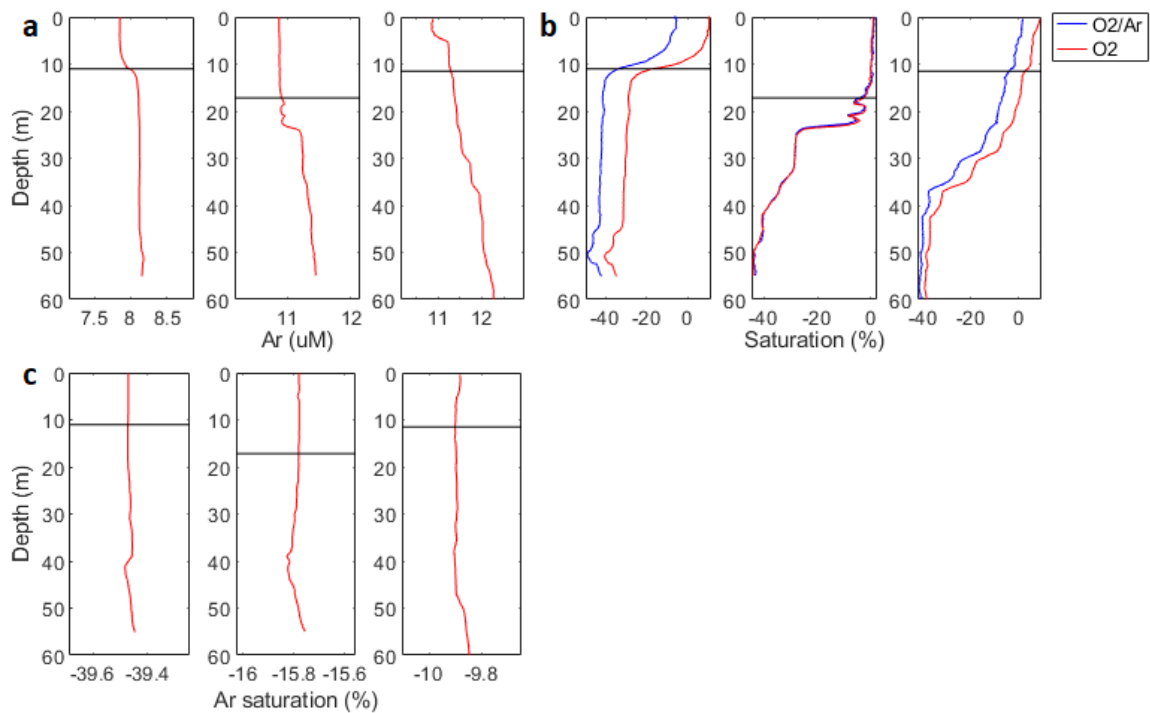
Contributions to mixed-layer  $\text{NCP}_{\text{RT\_Vflux}}$  from non-advective vertical fluxes of  $\text{O}_2$  are also non-negligible, with vertical eddy diffusion changing apparent  $\text{NCP}_{\text{RT\_Vflux}}$  rates by more than  $-40 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$  during some cycle measurements. Therefore, our assessment reinforces the need to account for these vertical terms in both Eulerian and Lagrangian studies employing underway  $\Delta\text{O}_2/\text{Ar}$  measurements. While we did not calculate entrainment fluxes to be generally consequential in this study, this was primarily due to the short timescale of interest inherent to calculating  $\text{NCP}_{\text{RT\_Vflux}}$ . In real time, entrainment represents a minor flux term, as the change in mixed layer depth is

small over short time steps. Over such time scales, a substantial rate of mixed layer deepening (tens of centimeters per minute) would be necessary to produce a significant effect on  $NCP_{RT\_Vflux}$  (10-20 mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup>) assuming an oxycline oxygen gradient of -4 mmol O<sub>2</sub> m<sup>-4</sup>. When considering the impact of entrainment upon Eulerian  $NCP_{prior}$  measurements, however, entrainment over the residence time of the O<sub>2</sub>/Ar signal must be considered, and we also suggest that time-weighting based on the weighted piston velocity approach of (Reuer et al., 2007) might be employed to more accurately constrain this term. Lastly, our Lagrangian study approach also allowed us to neglect horizontal advective fluxes and focus only on vertical terms. When using an Eulerian sampling approach, the inclusion of horizontal advection terms might also lead to significant impacts.

At a minimum, vertical profiles of oxygen from frequent CTD casts can be employed as demonstrated here to estimate the impact of these factors. Preferably, discrete or high-frequency vertical sampling of the O<sub>2</sub>/Ar ratio with depth can greatly improve the accuracy and confidence of such calculations. At the same time, the recently-developed N<sub>2</sub>O method (Cassar et al., 2014) also shows promise (Izett et al., 2018) as an approach for correcting the influence of vertical fluxes by simultaneously quantifying multiple sources of vertical mixing.

We do note that our calculations of vertical fluxes utilize measurements of O<sub>2</sub> and O<sub>2</sub> gradients rather than O<sub>2</sub>/Ar, and therefore involve the assumption that changes

in O<sub>2</sub> with depth are mirrored by the O<sub>2</sub>/Ar signal. Based on several vertical O<sub>2</sub>/Ar profiles conducted using the MIMS on the 2017 cruise, however, our assumption of small gradients in Ar saturation with depth (Hamme and Severinghaus, 2007) appears to be largely valid in this region and year, lending confidence to our use of O<sub>2</sub> concentrations for vertical flux calculations (Fig. 23). Estimates of vertical eddy diffusion and entrainment fluxes based on the O<sub>2</sub>/Ar gradient for this handful of O<sub>2</sub>/Ar profiles generally correspond closely (<3 mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup> difference) to fluxes obtained using O<sub>2</sub> gradients alone (Table 4). However, vertical advection estimates made using the O<sub>2</sub> gradient versus measured O<sub>2</sub>/Ar gradients can differ by up to 50 mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup>. This is due to high vertical velocities throughout this region, resulting in considerable sensitivity of calculated advective fluxes due to amplification of small differences in parametrization of the subsurface biological oxygen gradient. Additionally, we further acknowledge that the modeled vertical velocities we used to calculate the magnitude of advective fluxes may differ from *in-situ* vertical velocities. In particular, our method does not account for internal waves, which would not represent true vertical advection. This and any other discrepancies between modeled and real vertical velocities would proportionally impact our calculated advective fluxes in relation to such differences.



**Figure 23: Three vertical profiles of a) Ar, b) O<sub>2</sub>/Ar (blue) and O<sub>2</sub> (red) saturation expressed as percent difference from atmospheric equilibrium and c) Ar saturation as determined from profiling MIMS measurements during the 2017 cruise. Black horizontal lines denote the MLD, determined using a density-difference criterion from 10 m (see Methods).**

**Table 4: Comparison of vertical eddy flux, entrainment flux, and advective flux calculations (all in units of mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup>) derived using oxycline O<sub>2</sub> and oxycline biological O<sub>2</sub> (O<sub>2</sub>/Ar) gradients.**

|           | Date                   | O <sub>2</sub> gradient (mmol m <sup>-4</sup> ) | Biol. O <sub>2</sub> gradient (O <sub>2</sub> /Ar) (mmol m <sup>-4</sup> ) | F <sub>v</sub> (O <sub>2</sub> ) | F <sub>v</sub> (O <sub>2</sub> /Ar) | F <sub>e</sub> (O <sub>2</sub> ) | F <sub>e</sub> (O <sub>2</sub> /Ar) | F <sub>adv</sub> (O <sub>2</sub> ) | F <sub>adv</sub> (O <sub>2</sub> /Ar) |
|-----------|------------------------|---|--|----------------------------------|-------------------------------------|----------------------------------|-------------------------------------|------------------------------------|---------------------------------------|
| Profile 1 | '12-Jun-2017 01:48:07' | -4.8  | -3.1   | 11                               | 7                                   | 3                                | 2                                   | 139                                | 91                                    |
| Profile 2 | 16-Jun-2017 03:50:02'  | -2.4  | -2.3   | 6                                | 5                                   | 3                                | 3                                   | 208                                | 200                                   |
| Profile 3 | '26-Jun-2017 01:21:14' | -0.5  | -0.9   | 4                                | 7                                   | 0                                | 0                                   | 24                                 | 39                                    |

To summarize, vertical advection and vertical eddy diffusion are likely more influential factors in terms of vertical O<sub>2</sub>/Ar fluxes relative to mixed layer entrainment in this study, with advection a particularly large and difficult-to-constrain source of uncertainty. Their contribution to the mixed layer O<sub>2</sub>/Ar budget, however, is dependent on local conditions, particularly on upwelling rates and the strength of the vertical gradient in biological oxygen. Overall, the effect of vertical fluxes of O<sub>2</sub>/Ar also appears to be of major importance alongside other non-steady-state factors in driving observed discrepancies between NCP<sub>RT\_vflux</sub> and NCP<sub>prior</sub> in this region.

#### **4.5 Conclusions**

Ultimately, our study further highlights the importance of considering the potential impact of vertical fluxes and diel and multi-day non-steady-state factors when assessing surface layer productivity using *in situ* ΔO<sub>2</sub>/Ar measurements, particularly within a highly-productive, dynamic region such as the California Current Ecosystem. Taking advantage of our Lagrangian study design as well as model output data, we are able to contrast NCP<sub>prior</sub> rates calculated using traditional assumptions with NCP calculated without reliance on a steady-state condition and with the inclusion of terms parameterizing the influence of vertical fluxes. Results produced using these two contrasting approaches show significant discrepancies. We conclude that the non-steady-state rate of change term represents a considerable influence upon calculated NCP rates in this region, identifying cases in which this term is likely primarily biology-

driven. At the same time, we observe the calculated and potential influence of vertical fluxes of O<sub>2</sub>/Ar from entrainment, vertical eddy diffusion, and particularly vertical advection to be significant, contributing substantially to the real mixed layer O<sub>2</sub>/Ar budget.

We reiterate that our study region and results represent conditions that have long been held to be poorly-suited to deployment of the *in situ* O<sub>2</sub>/Ar method. While assumptions of negligible vertical fluxes and minimal violation of steady-state conditions may be valid for given regions and seasons, such factors should nevertheless be assessed in any field deployment to ensure that their potential impact is truly small. In particular, the N<sub>2</sub>O approach for quantifying the influence of vertical fluxes upon O<sub>2</sub>/Ar (Cassar et al., 2014) will likely possess considerable value (Izett et al., 2018) for investigators seeking to make EIMS or MIMS measurements of NCP in environments presenting similar challenges to the California Current system.

Similarly, when conducting non-Lagrangian measurements in regions of high diel metabolic activity, we find that adequate consideration of the impact of time of day upon measured values is also important. We observed diel cycles in  $\Delta$ O<sub>2</sub>/Ar of up to 17% in range at some locations, adding a considerable source of variability to the median cycle  $\Delta$ O<sub>2</sub>/Ar signal on hourly time scales. To alleviate the influence of such diel impacts on future studies, we suggest that when monitoring throughout the course of the solar day at a field location is not possible, investigators should endeavor to capture O<sub>2</sub>/Ar

measurements that cover a 12-hour period fully capturing the local daily solar maximum.

Lastly, our study highlights conditions under which diel and multi-day non-steady-state factors exert relatively low influence over calculated NCP rates. Less-productive ecosystems tend to exhibit both lower diel variability as well as considerably smaller offsets between NCP estimated using steady-state versus non-steady-state assumptions.

In summary, while the influence of rate-of-change and vertical flux effects as well as diel patterns in  $\Delta O_2/Ar$  can be considerable in the field, particularly within highly-productive and physically-dynamic settings, investigators already possess means to evaluate the impact of such factors upon *in situ* productivity rate estimates. Our findings within a challenging study environment reemphasize the importance of critically evaluating basic assumptions inherent to the  $O_2/Ar$  method and stress the utility of careful planning and study design when conducting  $O_2/Ar$  measurements under non-ideal conditions.

## **4.6 Acknowledgements**

This research was made possible by a RAPID NSF OCE-1614359 and federal funding provided to the California Current Ecosystem (CCE) LTER project, NSF OCE-1637632. We are sincerely grateful to the captain, crew, and marine technicians of the *R/V Sikuliaq* and the *R/V Roger Revelle* for their assistance prior to and during our 2016

and 2017 expeditions. We also thank Susan Becker, Melissa Miller, and Megan Roadman for their assistance processing and analyzing collected Winker samples. We are additionally grateful to Ralf Goericke for his help processing vessel cruise data and to Chief Scientist Mark Ohman and the rest of the CCE LTER team. Lastly, we thank the staff of Scripps Institution of Oceanography for their cruise logistical and organizational support.

The authors declare no conflict of interest. Data reported and presented in this study can be accessed at the CCE-LTER Datazoo site.

## 5. Conclusions

The body of research presented in my dissertation expands our understanding of links between marine microbes and key processes in ocean biogeochemistry and highlights the utility of parallel measurements and Lagrangian study designs in investigating questions related to marine microbial communities and rates.

In Chapter 2, I investigated potential connections between patterns of productivity and marine microbial community structure in the western North Atlantic. By employing an internal standard-based technique for quantitative amplicon sequencing of both the prokaryotic and eukaryotic microbial communities alongside high-resolution O<sub>2</sub>/Ar-derived measurements of NCP, I could relate measured *in situ* productivity rates directly to estimated abundances of specific microbial taxa. Sampling

a highly-productive phytoplankton bloom in the Mid-Atlantic Bight, rRNA amplicon sequencing identified the *Aureococcus anophagefferens* and *Chrysophyceae* as the dominant eukaryotic marine microbes in the surface ocean community. This finding highlights these mixotrophic algal taxa, responsible for forming coastal harmful algal blooms, as potentially implicated in driving regional productivity patterns in summer. Our results also indicate the relative lack of relationships between NCP and specific microbial taxa in less-productive waters within our study region, furnishing evidence supporting the view that NCP may not be linked to particular taxa in more oligotrophic settings. This work additionally demonstrated the utility of quantitative amplicon sequencing by examining and illustrating stark differences between patterns of taxonomic abundance implied by relative proportions and abundance patterns determined using an internal standard technique.

My work presented in Chapter 3 builds significantly upon samples and data collected in Chapter 2 and represents the most extensive study of microbial community structure in the summer Western North Atlantic conducted to date. Leveraging underway high-resolution measurements of NCP and N<sub>2</sub> fixation as well as sampling for macronutrients and trace micronutrients performed in parallel with molecular sample collection, we uncover intriguing regional patterns of marine microbial diversity and carbon and nutrient biogeochemistry. In addition to corroborating relationships between high productivity rates and increased abundances of *Aureococcus*, *Chrysophyceae*,

cryptophytes, and surface-associated heterotrophic bacteria first identified in Chapter 2, we establish that the abundances of many of these same taxa are similarly related to high N<sub>2</sub> fixation. In particular, our findings highlight a striking correspondence between peaks in N<sub>2</sub> fixation and *Chrysophyceae* abundances, potentially hinting at a previously unsuspected ecological connection between this stramenopile clade and diazotrophic activity. We further identify a strong negative regional productivity-diversity relationship for eukaryotic marine microbes, in which eukaryotic community diversity tends to be lowest at stations exhibiting the most elevated NCP rates. Finally, we utilize co-occurrence network construction to identify keystone hub taxa within western North Atlantic microbial communities, highlighting marine stramenopiles, chlorophytes, and the diazotroph host *Braarudosphaera bigelowii* as central to driving networked patterns of co-occurrence in our study.

In Chapter 4, I utilize the results of several multi-day Lagrangian field deployment of EIMS O<sub>2</sub>/Ar measurements in the California Current to evaluate the influence and importance of several important assumptions involved in calculating NCP rates from the O<sub>2</sub>/Ar method. In particular, I assess the impact of long-term non-steady-state conditions, vertical fluxes of O<sub>2</sub>/Ar, and short-term non-steady-state, microbially-driven diel cycles in surface O<sub>2</sub>/Ar upon NCP estimates, all within a ‘worst-case’ field environment characterized by elevated productivity and strong coastal upwelling. Our findings underline the high importance of non-steady-state terms in driving observed

offsets between NCP calculated with and without a steady-state assumption. We further determine that vertical advection also represents a large potential source of error in NCP estimates within this region, while other vertical fluxes generally exert a reduced albeit still situationally significant impact. Lastly, we highlighted the large influence of diel behavior in O<sub>2</sub>/Ar upon Lagrangian measurements of NCP, emphasizing the need to consider time of measurement when interpreting O<sub>2</sub>/Ar data from non-Lagrangian studies, as well as the utility of developing field metrics to identify when diel variability may be a significant factor while on expedition. These conclusions provide useful further insight into assumptions central to the O<sub>2</sub>/Ar NCP method as commonly practiced today, particularly reinforcing the need to quantify or estimate sources of error due to vertical fluxes and non-steady-state factors in order to confidently determine community productivity rates in more challenging field environments.

As elaborated upon throughout this dissertation, our current knowledge regarding how the microbial community influences the cycling of carbon and nutrients across ocean ecosystems remains limited despite the great importance of these dynamics to attaining a deeper understanding of marine ecology and global biogeochemistry. Consequently, even statistical inferences from the field continue to represent important contributions to oceanographic research that help enable the development of broader hypotheses and identify links of interest for focused follow-up investigations. My work in particular has highlighted relationships between microbial rates and community

structure on a considerable spatial scale, leveraging powerful measurement techniques that have not been previously utilized in conjunction with one another. It is my hope that the relationships between productivity, N<sub>2</sub> fixation, trace metal concentrations, diversity, and specific bacterial and eukaryotic marine taxa identified in my thesis research can be more thoroughly explored in interaction and field studies. I also believe that continued regional monitoring in the western North Atlantic, bolstered by the establishment of the Northeast US Shelf LTER site (2018), has the potential to provide further insights into the spatial and community patterns I report in this manuscript. At the same time, my dissertation work also helps to advance the application of current methodologies in oceanography, demonstrating the power of quantitative community data combined with high-resolution biogeochemical measurements as well as exploring the utility of Lagrangian observations and profile-based corrections for NCP measurements. These approaches offer considerable room for further development and promise to further improve our ability to study processes in marine biogeochemistry as well as marine microbial ecology, both independently and in conjunction with one another.

## **Appendix A: Linking patterns of net community production and marine microbial community structure in the western North Atlantic**

## ***Supplemental Methods***

### **Measurement of Biological O<sub>2</sub>**

O<sub>2</sub>/Ar was measured continuously during the cruise from the R/V Atlantic Explorer's underway intake using Equilibrator Inlet Mass Spectrometry (Cassar et al., 2009). The ship's inlet is located approximately 2.5 meters below the waterline. The O<sub>2</sub>/Ar ion current ratio was recorded from the headspace of the membrane contactor, which undergoes gas exchange and equilibration with the sampled seawater, for 2-3 hours at a time, punctuated by 30-minute measurements of the O<sub>2</sub>/Ar of ambient room air to correct for instrumental drift. An automated Valco valve (VICI, Houston, TX, USA) governed the regular switching between air and water O<sub>2</sub>/Ar gas sample flow. At each CTD station during the cruise, triplicate Winkler dissolved oxygen samples were collected from both in-situ surface water and from the ship's seawater tap directly upstream of the EIMS to measure the degree of microbial respiration within the ship's lines, a phenomenon that has been documented in the underway seawater systems of multiple research vessels (Juraneck et al., 2010). All Winkler sample bottles were collected according to the standard BATS protocol and analyzed at Rod Johnson's laboratory at the Bermuda Institute of Ocean Sciences.

As in previous O<sub>2</sub>/Ar work conducted aboard the same vessel (Estapa et al., 2015), O<sub>2</sub> measurements revealed insignificant differences between the dissolved oxygen concentration of surface seawater and seawater collected from the ship's underway

seawater lines, suggesting that any microbial respiration taking place within the ship's underway system was negligible.

## Calculation of net community production

Mass spectrometer and ship underway data were time-averaged at two-minute intervals. Seawater O<sub>2</sub>/Ar values were calibrated using O<sub>2</sub>/Ar measurements of ambient room air, then quality-filtered to remove O<sub>2</sub>/Ar data collected during anomalous spikes in mass spectrometer pressure. Data was filtered according to a moving average and a moving standard deviation criterion, removing data collected whenever pressure within a one-hour window exceeded a moving standard deviation of 10<sup>-7</sup> mbar and/or a moving average greater than 5 × 10<sup>-6</sup> mbar. Daily NCP rates were calculated using the following equation (Cassar et al., 2011):

$$\text{NCP (mmol O}_2\text{/m}^3\text{/day)} = \frac{\text{O}_2\text{/Ar}(\%)}{100} * \frac{[\text{O}_2]_{\text{sat}}(\frac{\text{umol}}{\text{kg}})}{1000} * \rho(S, T) * k * (\frac{1}{\text{MLD}}) \quad (\text{A-1})$$

Where  $\frac{\text{O}_2}{\text{Ar}}$  (%) is the percent supersaturation of the O<sub>2</sub>/Ar ratio relative to that of interpolated ambient air O<sub>2</sub>/Ar measurements, which reflect the expected O<sub>2</sub>/Ar ratio at equilibrium.  $\rho$  signifies the density of seawater (kg/m<sup>3</sup>) according to the 1983 UNESCO equation of state, and  $k$  denotes the piston velocity, estimated according to the quadratic wind-speed dependence of (Wanninkhof, 1992). MLD is the mixed layer depth in meters. The oxygen concentration corresponding to saturation was calculated using the oxygen solubility relationship described in (Garcia and Gordon, 1992).

As biological oxygen concentrations in the surface ocean are also dependent on past wind speeds, gas transfer velocities were weighted using the 30-day wind speed history prior to measurement (Reuer et al., 2007), with previous wind speeds along the cruise track obtained from the corresponding daily NCEP/NCAR 10 m wind speed reanalysis. Mixed layer depths for this calculation were calculated using potential density profiles obtained from fourteen CTD casts according to a density difference criterion of 0.03 kg/m<sup>3</sup> from a reference depth of 10 m (de Boyer Montegut et al., 2004). Calculated mixed layer depths were linearly interpolated across the cruise track between CTD casts.

As NCEP/NCAR wind speed values corresponding to cruise dates and locations appear to underestimate ship-measured wind speeds, the ship-measured values were corrected to a 10 m reference height using Equation (1) of (Thomas et al., 2005) and used to adjust the NCEP/NCAR wind speed history based on a linear regression between predicted and measured wind speeds for the cruise period.

### **Accounting for vertical fluxes of O<sub>2</sub>/Ar**

The integrated biological oxygen concentrations in the mixed layer can be expressed via the following equation:

$$MLD \frac{dO_{2bio}}{dt} = NCP - F_g + F_v + F_e \quad (A-2)$$

$O_{2_{bio}}$  represents the concentration of biological oxygen in the surface ocean.  $F_g$  denotes the gas exchange term, while  $F_v$  and  $F_e$  represent vertical fluxes via eddy diffusion and entrainment. Traditionally, the system is assumed at steady-state with limited  $F_v + F_e$  (Hamme and Emerson, 2006), with  $F_g$  equal to NCP, leaving just  $F_v + F_e$  to consider. We estimated the vertical eddy diffusive flux  $F_v$ , using equation (6) of (Castro-Morales et al., 2013), with a vertical eddy diffusivity coefficient of  $1 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}$  as measured for the Mid-Atlantic Bight (Wallace, 1994), calculating oxygen gradients from CTD profile data using a Model-II least-squares fit. Estimated depth-integrated  $F_v$  fluxes ranged between just -0.6 and 2.92  $\text{mmol O}_2 \text{ m}^{-2} \text{ day}^{-1}$  (median: 0.57  $\text{mmol O}_2 \text{ m}^{-2} \text{ day}^{-1}$ ). These flux estimates likely represent an upper bound, as much of the observed increase in  $\text{O}_2$  beneath the mixed layer may be thermally rather than biologically-driven. Argon vertical profiles are however likely close to equilibrium within the first few hundred meters, minimizing errors associated with using  $\text{O}_2$  concentration profiles rather than  $\text{O}_2/\text{Ar}$  (Hamme and Severinghaus, 2007).

Entrainment due to deepening of the mixed layer is unlikely to be significant in the highly stratified summer Sargasso Sea. A similar study in the Sargasso Sea in September-October 2011/2012 assessed entrainment flux to comprise less than +/- 6% of calculated  $\text{O}_2/\text{Ar}$ -derived NCP (Estapa et al., 2015). We are, however, unable to account for entrainment fluxes in the more dynamic coastal region.  $\text{O}_2$  concentrations below the mixed layer consistently increased with depth based on CTD profiles taken throughout

our cruise, however, with the result that any entrainment fluxes would bias NCP values upwards. On the other hand, observed negative NCP values in open-ocean regions of the Sargasso Sea likely represent true heterotrophic activity.

### **Determination of appropriate control DNA spike quantity**

We conducted a qPCR using 18S V4 primers to assess the average concentration of 18S rDNA genes in our samples. Serial dilutions (1:10 to 1:1 000 000) of the *S. pombe* control DNA were used to generate the qPCR standard curve. The sample with the lowest 18S DNA abundance was determined to contain approximately 1/4th the number of 18S rDNA sequences ul-1 as 1.14 ng of the *S. pombe* standard (9 702 000 copies), suggesting an appropriate addition of 0.00285 ng to maintain the standard sequences added at a relative abundance of no more than 1% of reads following sequencing. We subsequently conducted an Ion Torrent PGM sequencing run to verify our ability to detect such a dilute spike in our sequencing output data. Two sets of samples from the 2015 cruise were run in technical duplicates: one pair from an oligotrophic location near Bermuda and another from a large phytoplankton bloom south of Long Island. These sites were chosen to span the range of 16S and 18S rDNA gene concentrations encountered in this study. These samples were extracted with the addition of 152 ng or 15.2 ng of 16S and 0.679 ng or 0.0679 ng of 18S internal standards, both in 50 ul volumes, following the same extraction and PCR protocols described in the main text. Ion Torrent

PGM sequencing was performed at the Boston University Microarray Core on a 314 chip. Primer, barcode, and adapter sequences used for this Ion Torrent sequencing run are detailed in Supplementary Table 6.

## **Statistical Analysis**

For comparison to sequenced samples, underway NCP data was time-averaged across 45-minute bins centered on the time of sample collection. If sample collection took place during a gap in NCP measurements, a 22.5-minute average of the closest NCP data to the sample location was used instead (Supplementary Table 2). Prior to reaching stations 24 and 25, technical issues prevented further O<sub>2</sub>/Ar data collection. As S25 has no nearby NCP data, we set the NCP value of this station to NaN.

Statistical analyses were performed on 16S rDNA and 18S rDNA taxonomy data using the *vegan* (Okansen et al., 2017) and *phyloseq* packages in R.

Ordination of samples was performed with Principal Coordinate Analysis (PCoA), using weighted Bray-Curtis dissimilarity to express ecological distance. The first principal component of each projection was further modeled against the dataset of NCP and environmental measurements using linear regression. For these and all subsequent statistical comparisons with our molecular data, which represent volumetric sequence counts, NCP rates were expressed as volumetric values (mmol m<sup>-3</sup> d<sup>-1</sup>), though we obtained similar results if mixed-layer integrated NCP values were used.

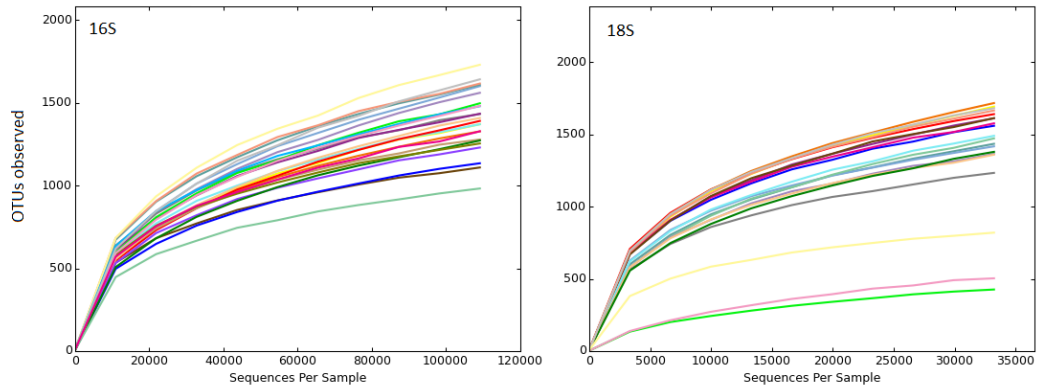
## **Partial Least Squares Regression Analysis**

Partial least squares (PLS) regression analysis was performed using the 'mixOmics' package in R (Rohart et al., 2017) in 'regression' mode with two model components to identify bacterial and eukaryotic taxa with good predictive power in estimating NCP and environmental variables. Analyses were performed at the taxonomic levels corresponding to the order level for both the 18S and 16S datasets, excluding less-abundant taxa that comprised <0.5% of total 18S or 16S reads. Leave-one-out cross-validation was performed to evaluate model performance.

## **Online data availability**

Sequences and metadata are available from the NCBI Sequence Read Archive under accession number SRP126177. Other data available upon request.

## Supplementary Figures



**Figure 24: Rarefaction plots showing OTUs observed versus sequencing depth for (left) 16S and (right) 18S rRNA amplicon samples. Samples were rarefied to the number of reads within the smallest included sample from the dataset.**

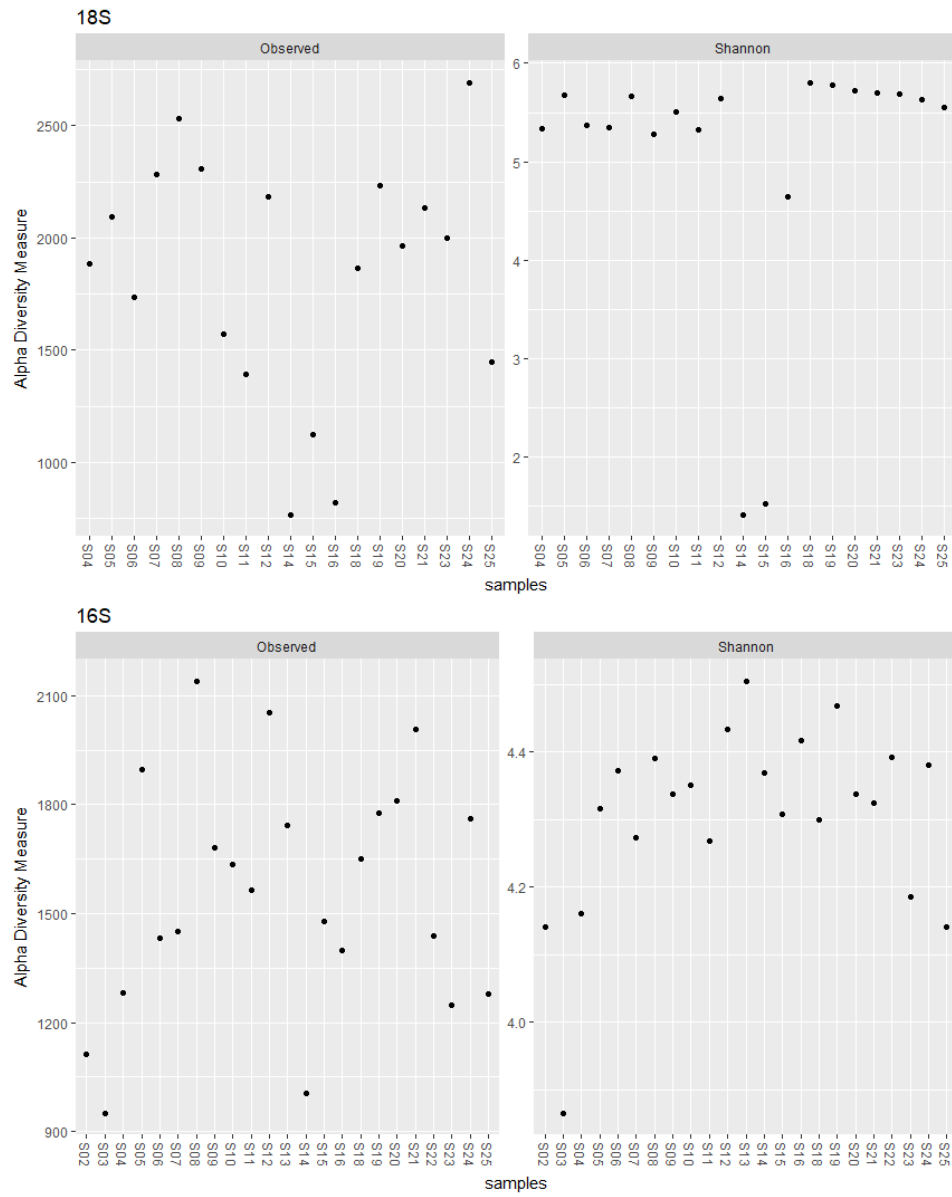


Figure 25: Observed OTUs per sample and Shannon diversity for (top) 18S and (bottom) 16S samples.

## **Appendix B: Patterns of North Atlantic microbial community structure in relation to net community production, nitrogen fixation, and nutrient availability**

### ***Supplemental Methods***

#### **Measurements of underway dissolved surface O<sub>2</sub>/Ar**

Briefly, sample seawater is pumped through a membrane contactor, resulting in exchange and equilibration of the dissolved gases with the headspace. A quadrupole mass spectrometer continuously measures the O<sub>2</sub>/Ar ion current ratio from the headspace for a 2-3 h period, followed by a 30-minute period of ambient air O<sub>2</sub>/Ar measurements to correct for instrument drift. An automated valve controls this regular switching between seawater and air calibration measurements. In 2015, measured seawater was supplied via the ship's underway intake, whereas a trace-metal-clean towfish was employed to bring seawater aboard during the 2016 and 2017 expeditions. The ship's intake is located approximately 2.5 m below the waterline, while the towfish typically collects water at a depth of 3-5 m.

Consumption of oxygen due to biological respiration in either the ship's lines or the towfish plumbing is a potential concern for the O<sub>2</sub>/Ar method (Juranek et al., 2010). For our deployment, however, the very short residence time of sample seawater within the towfish tubing as well as our thorough cleaning of all plumbing prior to field deployment likely ensure that the effect of respiration is minimal.

## Calculation of net community production

EIMS O<sub>2</sub>/Ar data as well as temperature and salinity recorded from the ship's underway system were time-averaged every two minutes. Ambient air O<sub>2</sub>/Ar measurements were used to calibrate seawater ion current ratio data. Calibrated O<sub>2</sub>/Ar data was then filtered to remove data collected when analytical chamber pressure within a 1-h period exceeded a moving standard deviation of 10<sup>-7</sup> mbar or a moving average of more than 5 x 10<sup>-6</sup> mbar, as well as data collected during periods in which the system was known to be operating improperly (for instance, when the seawater pump was off while making adjustments to the towfish). NCP rates were subsequently calculated following the equation below from (Cassar et al., 2011):

$$\text{NCP (mmol O}_2\text{/m}^3\text{/day)} = \frac{O_2(\%)}{Ar} * \frac{[O_2]_{sat}}{1000} * \rho(S, T) * k * \left(\frac{1}{MLD}\right) \quad (\text{B-1})$$

O<sub>2</sub>/Ar (%) represents the percent supersaturation of the seawater O<sub>2</sub>/Ar ratio relative to ambient air.  $\rho$  denotes seawater density (kg/m<sup>3</sup>) calculated from the 1983 UNESCO equation of state, while MLD is the mixed layer depth in m, calculated for each CTD cast using a density difference criterion of 0.03 kg/m<sup>3</sup> from the density measurement at a 10 m near-surface depth (Dong et al., 2008) and linearly interpolated between casts. The oxygen concentration [O<sub>2</sub>]<sub>sat</sub> (μmol/kg) was evaluated following the oxygen solubility equation of (Garcia and Gordon, 1992).

$k$  signifies the piston velocity, determined using (Wanninkhof, 1992), which was weighted using the 30-day wind history (NCEP/NCAR 10 m wind speed reanalysis

data) preceding measurement (Reuer et al., 2007) with the modification detailed in Equation 5 of (Teeter et al., 2018). NCEP/NCAR wind speeds generally underestimated wind speeds measured in the field from ship instruments, so a correction was first applied using a linear regression between NCEP/NCAR-predicted and measured wind speeds prior to generation of the wind speed history. In 2017, NCEP/NCAR-derived wind speeds disagreed strongly with the ship-recorded wind speed time series. As no meaningful correction was possible, we elected to use unaltered NCEP/NCAR reanalysis data for the 2017 calculations.

In addition to biological production and respiration,  $O_2/Ar$  in the surface ocean can also be influenced by several other fluxes. Integrated biological oxygen in the mixed layer can be expressed using the following mass balance:

$$MLD \frac{dO_{2_{bio}}}{dt} = NCP - F_g + F_v + F_e \quad (B-2)$$

where  $O_{2_{bio}}$  denotes the biological oxygen concentration within the mixed layer,  $F_g$  the gas exchange term, and  $F_v$  and  $F_e$  the fluxes resulting from eddy vertical diffusive mixing and from entrainment of subsurface waters into the mixed layer, respectively.

We assume that the system is at steady state, with  $F_g$  equal to NCP, and that vertical and horizontal fluxes of  $O_2/Ar$  are negligible.

To determine the potential effect of  $F_v$  and  $F_e$  upon calculated productivity rates, we adopted the approach of (Castro-Morales et al., 2013), estimating eddy diffusive fluxes using equation 6 and entrainment fluxes from equation 5. For the study region,

we use a vertical eddy diffusivity coefficient of  $1 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}$ , taken from (Wallace, 1994). Oxygen gradients were calculated from CTD profile data using a Model-II bisector least-squares fit. Mixed-layer depths prior to the date of measurement were taken from the JAMSTEC Argo Grid Point Value MLD dataset (Hosoda et al., 2010). As Argo MLD data coverage for the study region was poor, we also conducted a sensitivity analysis by assuming, at each CTD location, a prior 3m entrainment over the median calculated residence time of a given year's cruise. Estimated depth-integrated  $F_v$  fluxes were low, ranging between  $-2.9$  and  $0.6 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$  (range of integrated NCP values for all three years was  $-24.5$  to  $163.4 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$ ). Sparse Argo MLD data availability for the study region allowed only a few estimates of entrainment flux to be made ( $n=6$ ), although calculated fluxes were also generally low ( $0$  to  $-5.5 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$ ). Assuming a 3m entrainment at each CTD location over the course of the median residence time of oxygen during each cruise produced entrainment fluxes ranging between  $-6.5$  and  $1.4 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$ .

These estimates also likely represent an upper bound to the influence of vertical fluxes. The open-ocean Sargasso Sea is seasonally stratified, meaning that entrainment fluxes in the subtropical area of the study region are likely limited. We observed in all three years that total oxygen concentrations generally increased with depth, meaning that any vertical fluxes from the subsurface would increase measured mixed-layer NCP. Consequently, negative NCP values in the open ocean are likely to result from true net

heterotrophy as opposed to vertical mixing. Additionally, much of the variability in O<sub>2</sub> saturation with depth is likely thermally-driven. At the same time, subsurface maxima in O<sub>2</sub>/Ar are common in the subtropical oceans during summer and typically result from accumulation of biological oxygen below the stratified thermocline (Spitzer and Jenkins, 1989, Hamme and Emerson, 2006). Consequently, part of this increase in O<sub>2</sub> below the mixed layer in the open ocean is likely attributable to biology. The O<sub>2</sub>/Ar method remains more limited in coastal waters, however, where vertical advection and other mixing may be more dynamic. On and near the coastal shelf, our calculations are unable to account for such factors, and low/negative NCP rates should be interpreted with greater caution. Ultimately, only five molecular stations exhibited undersaturation greater than the error inherent to the O<sub>2</sub>/Ar method ( $\Delta\text{O}_2/\text{Ar} + / - 0.3\%$ ), so these measurements have negligible effects on our productivity-taxonomy analyses.

### **Flow-through acetylene reduction nitrogen fixation assays using cavity-ringdown spectroscopy (FARACAS)**

Briefly, acetylene (C<sub>2</sub>H<sub>2</sub>) is dissolved into filtered seawater. The dissolved acetylene tracer is mixed at a constant ratio with surface seawater collected continuously via towfish, reaching a 10% acetylene saturation. The mixed seawater is pumped into a 9 L flow-through incubation chamber (V) at  $\sim 100 \text{ ml min}^{-1}$  (Q<sub>s</sub>) resulting in a 90 min residence time for the incubated seawater. Downstream of the incubation chamber, the ethylene (C<sub>2</sub>H<sub>4</sub>) from the incubated seawater is stripped via a bubble column reactor

using a  $\sim 35 \text{ ml min}^{-1}$   $\text{C}_2\text{H}_4$ -free stripping gas ( $Q_{\text{strip}}$ ) and then carried into a cavity ring down spectroscopy (CRDS) analyzer. CRDS is used to measure ethylene concentration every few seconds at high precision (3 standard deviations of 2 and 0.2 ppb when integrated over 5 and 300 seconds, respectively). The incubation chamber is bypassed every few hours to measure the ethylene background in the seawater and in the acetylene-dissolved seawater. The difference in ethylene concentration between incubation and bypass mode can be used to calculate the near real-time ethylene production rates using the equations presented in (Cassar et al., 2018). Finally, the ethylene production rates are converted to  $\text{N}_2$  fixation rates using a 4:1 ratio (Mulholland et al., 2004). The daily  $\text{N}_2$  fixation rates are estimated by integrating the continuous  $\text{N}_2$  fixation rates over a 24-h diel cycle. In its current configuration, the detection limit is approximately  $0.19 \mu\text{mol N m}^{-3} \text{ d}^{-1}$ . A full description of the FARACAS method can be found in (Cassar et al., 2018). The full dataset of  $\text{N}_2$  fixation rates is presented in (Tang et al., 2019).

## **Sampling and analysis of nutrients and trace metals**

### **Nutrient sampling**

Surface seawater for nutrient analyses ( $\text{NO}_2/\text{NO}_3$ ,  $\text{PO}_4$ , and silicate) were collected into 15 ml Falcon tubes, sealed, and preserved at  $-20^\circ\text{C}$  immediately (silicate samples were preserved at  $4^\circ\text{C}$ ). Concentrations were subsequently analyzed using an

Automatic Nutrients Analyzer. The detection limits of the method used were 0.03  $\mu\text{M}$  for  $\text{NO}_2+\text{NO}_3$ , 0.014  $\mu\text{M}$  for  $\text{PO}_4$ , and 0.05  $\mu\text{M}$  for silicate.

### **Trace metal sampling: dissolved trace metals**

Seawater was pumped continuously from the towfish into a trace metal clean environment (double-layered plastic bubble) pressurised by air from a class-100 laminar flow hood for sampling. The cleaning protocols for sampling bottles and equipment followed GEOTRACES guidelines (Cutter et al., 2010). Water taken for analysis of dissolved trace metals was filtered through a 0.2  $\mu\text{m}$  capsule filter (Sartorius SARTOBRAN® 300). Seawater was collected in acid-cleaned 60 ml LDPE bottles, after rinsing three times with about 20 ml of seawater. Samples were then acidified to  $\sim\text{pH}$  1.7 with 2 % (v/v) HCl (Ultrapur® Merck) in the laminar flow hood. The sample bottles were then double bagged and stored at ambient temperature in the dark before analysis on shore.

### **Trace metal sampling: particulate trace elements**

Particulate trace elements were collected on polyethersulfone filters (0.45  $\mu\text{m}$  supor®) mounted in Nalgene polycarbonate filtration units. Filters were cleaned according to (Planquette and Sherrell, 2012) and stored in Milli-Q in acid-cleaned, low density polyethylene bottles (Nalgene) until use. Filtration units were soaked in 0.1 M

HCl (suprapur grade, Merck) and rinsed with ultrapure water (Milli-Q, 18 MΩ.cm), between each use. Seawater was collected in 2 l LDPE bottles then poured into the filtration unit until the filter was clogged, and the volume recorded. Using plastic tweezers rinsed with Milli-Q between samples, filters were then transferred to acid-cleaned polystyrene Petri slides (Millipore), double bagged and stored frozen at -20°C until analysis.

#### **Trace metal analytical procedures: dissolved trace metals**

Seawater samples were analyzed following the protocol of (Lagerström et al., 2013). Briefly, samples were introduced to a PFA-ST nebulizer and a cyclonic spray chamber via a SeaFAST-pico™ introduction system (Elemental Scientific Incorporated, Omaha, NE). A six-point calibration curve was prepared by standard additions of a mixed element standard to our in-house standard seawater (North Atlantic seawater collected during GA01 cruise, see Supplementary Table 2), which was run at the beginning, the middle and the end of each run. Every 10th sample, a replicate was run, and accuracy was determined from analysis of consensus (SAFE S, GSP) and certified (NASS-7) seawaters at the beginning, middle and end of each run.

### **Trace metal analytical procedures: particulate trace elements**

Sample handling was performed in a Class-100 clean room back in the home laboratory. Particulate samples were digested following the protocol described by (Planquette and Sherrell, 2012). Briefly, filters were placed on the inner wall of acid-clean 15 ml PFA vials (Savillex), and 2 ml of a solution containing 2.9M hydrofluoric acid (HF, suprapur grade, Merck) and 8M nitric acid (HNO<sub>3</sub>, Ultrapur grade, Merck) was added to each vial. Vials were then sealed and refluxed at 130°C on a hot plate for 4 h. Once cool, the digest solution was evaporated until near-dryness. Subsequently, 100 µl of concentrated HNO<sub>3</sub> (Ultrapur grade, Merck) was added, and the solution was re-evaporated to remove fluorides. Finally, 3 ml of 0.8M HNO<sub>3</sub> (Ultrapur grade, Merck) was added and the archive solution was transferred to an acid cleaned 15 ml polypropylene centrifuge tube (Corning®) and stored at 4°C until analyses.

All analyses were performed on a SF-ICP-MS Element2 (Thermo) at the Pôle Spectrométrie Océans. The archive solution (400 µl) was diluted in 96-well titre plate (ESI) by addition of 2 ml of 0.8 M HNO<sub>3</sub> (Ultrapur, Merck) spiked with 1 ppb <sup>115</sup>In, in order to monitor instrument drift. Samples were introduced with a PFA-ST nebulizer connected to a quartz cyclonic spray chamber (Elemental Scientific Incorporated, Omaha, NE) via a modified SC-Fast introduction system consisting of an SC-2 autosampler, a six-port valve and a vacuum rinsing pump. The autosampler was contained under a HEPA filtered unit (Elemental Scientific). Two 6-point, multi-element

standard curves with acid matrix identical to that of samples and concentrations bracketing the range of the samples were run at the beginning, the middle and the end of the run. Every 10th sample was a replicate to monitor analytical precision and instrument drift was monitored with  $^{115}\text{In}$  spiked in all samples. Accuracy was determined by performing digestions of the certified reference material BCR-414 (plankton, Community Bureau of Reference, Commission of the European Communities) following the same protocol as for samples. Recoveries were typically within 10% of the certified values (and within the error of the data, taken from replicate measurements).

## **Bioinformatic processing of molecular data**

Illumina output from our 2016+2017 samples consisted of 22 807 458 paired-end reads, which were trimmed, merged, and quality filtered using pandaseq (Masella et al., 2012). Mean 18S amplicon length was  $424.2 \pm 4.3$  bp (mean  $\pm$  standard deviation), and mean 16S amplicon length was  $296.5 \pm 3.0$  bp. Reads were subsequently demultiplexed in QIIME (Caporaso et al., 2010b) using the script `split_libraries_fastq.py`. Chimeric sequences were detected and filtered using the Usearch 6.1 algorithm (Edgar, 2010, Edgar et al., 2011) at 97% similarity using the SILVA 123.1 ribosomal rRNA database (Pruesse et al., 2007). Primer and non-biological sequences were trimmed and filtered using Tagcleaner (Schmieder et al., 2010).

Pandaseq-assembled 2015 sequencing data was processed following the same demultiplexing, chimera detection, and Tagcleaner steps. Open-reference OTU picking was then performed upon 2016+2017 sequences and 2015 sequencing in iterative mode due to filesize limitations. A 97% similarity threshold was employed, and OTU picking was performed using the Usearch 6.1 algorithm. The PyNAST (Caporaso et al., 2010a) alignment method was employed, and taxonomy assignment was performed using the RDP classifier 2.2 (Wang et al., 2007). Following OTU picking and taxonomy assignment, internal standard sequences as well as metazoan and chloroplast + mitochondrial sequences were filtered from the dataset.

## **Statistical analyses**

Three 16S samples from 2017 to which <5 reads were assigned were excluded from our dataset. One 16S and five 18S samples from 2015 as well as one 16S sample from 2017 were not assigned any reads during demultiplexing. The five defective 18S samples from 2015 were likely the consequence of a faulty forward primer/adaptor.

In addition, a number of samples were excluded from quantitative taxonomic analyses on the basis that their internal standard recovery ratios or standard-derived total rDNA gene abundances were statistical outliers. Box-and-whisker plots of log-scaled recovery ratios were used to identify three samples with anomalously-high standard recovery rates relative to other samples of the same type collected in the same

year (Fig. 26). One 2017 16S sample was also filtered from the dataset on account of having a total 16S rDNA gene abundance well above two standard deviations from the mean for the 2017 16S dataset, giving it a heavily disproportionate weight in PLS regression analysis and in co-occurrence network module selection. These four filtered samples were not, however, excluded from relative abundance-based analyses such as alpha diversity calculations (Fig. 26).

Exclusion of these samples along with samples from depth left 79 18S samples and 79 16S samples in total. Underway metadata was time-averaged around each molecular sampling site using identical criteria as (Wang et al., 2018).

For alpha diversity analyses only, 18S and 16S samples were rarefied at an even depth corresponding to the smallest included 18S and 16S library size.

To visualize differences in whole-community structure across samples, principal Coordinates Analysis (PCoA) was performed in the 'phyloseq' package (McMurdie and Holmes, 2013) on both the 18S and 16S community datasets, with ecological distances between samples calculated using weighted Bray-Curtis dissimilarities. Samples were color-coded according to whether they were collected at locations north or south of the northern wall of the Gulf Stream in that year. The northern boundary of the Gulf Stream was determined from monthly EN4 temperature data and was defined as the intersection of the 15°C isotherm with the 200 m isobath—a criterion long-accepted as

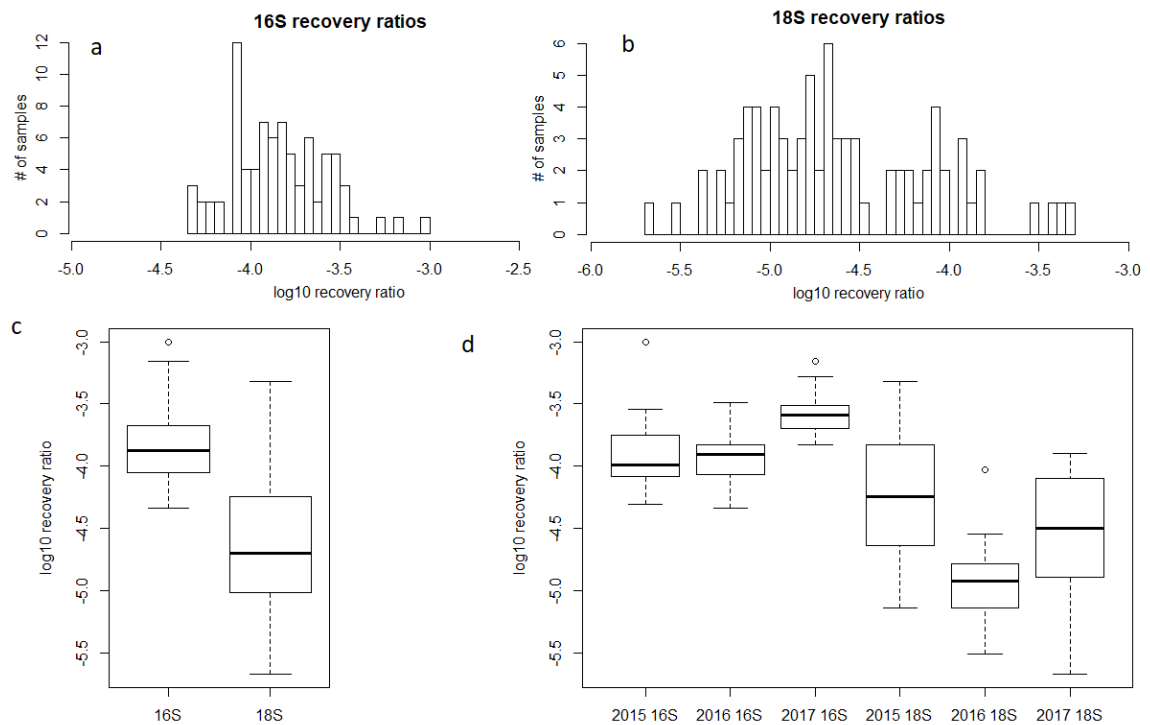
having a high ability to track the northern limit of the Gulf Stream (Joyce et al., 2000, Fuglister, 1955).

## **WGCNA module selection and network analyses using Flashweave, L-GRAAL**

Eukaryotic and prokaryotic taxa were agglomerated at the genus level (6th taxonomic rank) in phyloseq, then combined into a single abundance table following the removal of samples containing prokaryotic data but no eukaryotic data and vice versa. We forego the Hellinger transformation step, as our abundance data is quantitative. To optimize the scale-free topology fit, we selected a soft-thresholding power of 8, yielding nine 'modules' post-merging containing between 31 and 124 genera. Calculation of topological overlap measures followed by hierarchical clustering was executed following the same steps as (Guidi et al., 2016). Partial least squares regression analysis to assess the ability of selected modules to predict NCP and N<sub>2</sub> fixation were then performed following the procedure used in (Guidi et al., 2016). For input to Flashweave (Tackmann et al., 2018) and L-GRAAL (Malod-Dognin and Pržulj, 2015), we selected the two sub-networks most strongly related to productivity and nitrogen fixation (Pearson = 0.61 and 0.56 for NCP, Pearson = 0.88 and 0.87 for N<sub>2</sub> fixation,  $p < 0.001$  for both), as well as the module most negatively-associated with NCP (Pearson = -0.12,  $p > 0.05$ ) and N<sub>2</sub> fixation (Pearson = -0.2,  $p > 0.05$ ) (Fig. 31). Flashweave was run with default parameters (sensitive=TRUE, heterogeneous=FALSE). An alpha value of 0.4 was selected for L-

GRAAL alignments as in all pairs of modules it showed a high consensus between sequence and topological similarity while conserving high alignment quality scores (Edge correctness and Symmetric substructure scores). HiveAlign (<https://gitlab.univ-nantes.fr/erwan.delage/HiveAlign.git>) was then used to display the produced alignments.

### Supplemental Figures and Tables



**Figure 26: Histograms of the distribution of log-scaled internal standard recovery ratios for a) 16S rDNA and b) 18S rDNA molecular samples. Box-and-whisker plots of log-scaled recovery ratios for c) the overall 16S rDNA and 18S rDNA datasets and d) 16S rDNA and 18S rDNA samples subdivided by year. Box edges are defined by the first and third quartiles of the data. Individual points are shown only if identified as outliers, located >1.5x the full length of the box away from the nearest edge of the box.**

Whiskers are positioned at the lowest and highest values that are not identified as outliers.

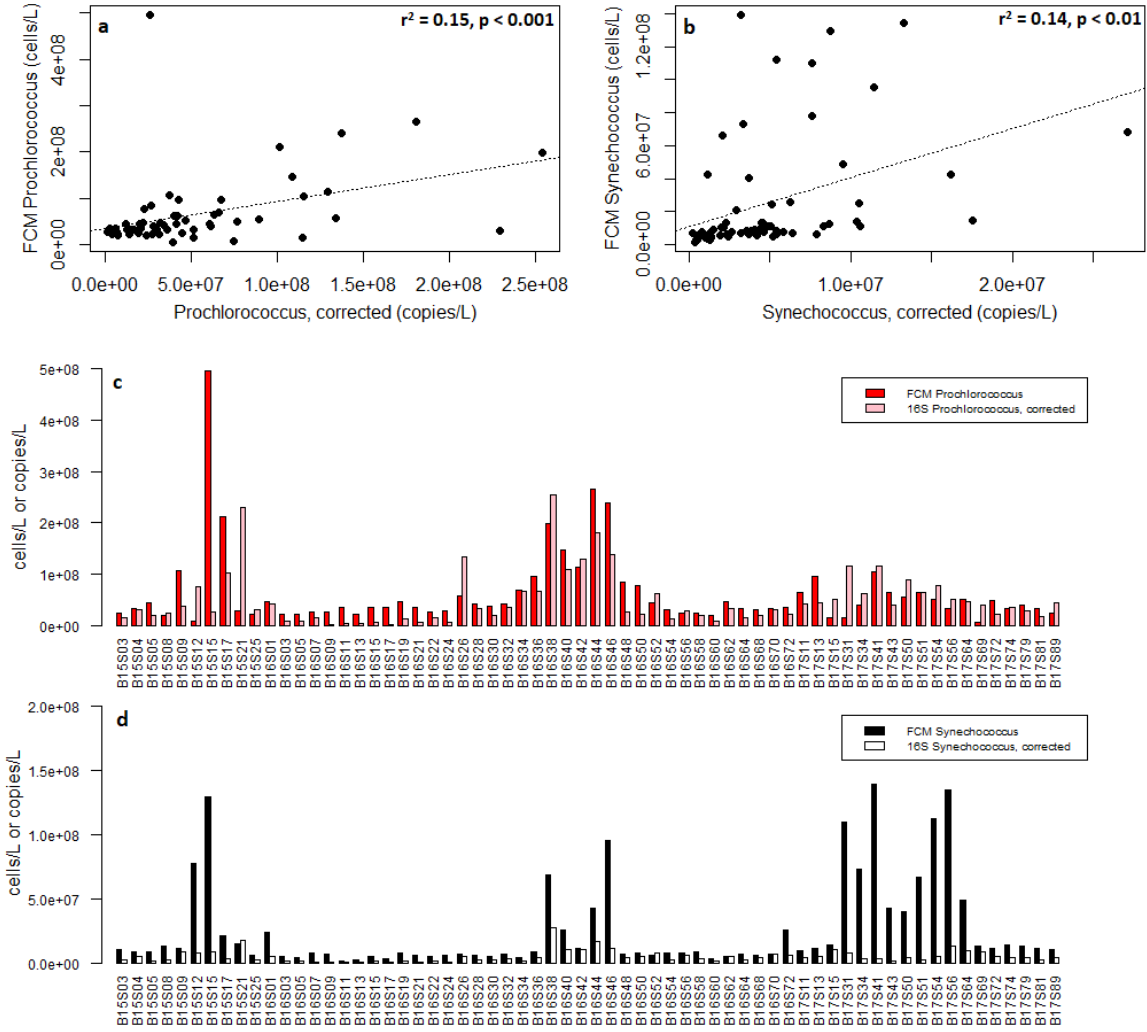


Figure 27: Scatter plot and goodness-of-fit of relationships between a) *Prochlorococcus* and b) *Synechococcus* flow cytometry cell counts and corrected quantitative 16S rDNA amplicon abundances. Barplots illustrating abundances of c) *Prochlorococcus* and d) *Synechococcus* as determined using flow cytometry cell counts and (corrected 2015) 16S rDNA gene abundances across the three-year sample dataset.

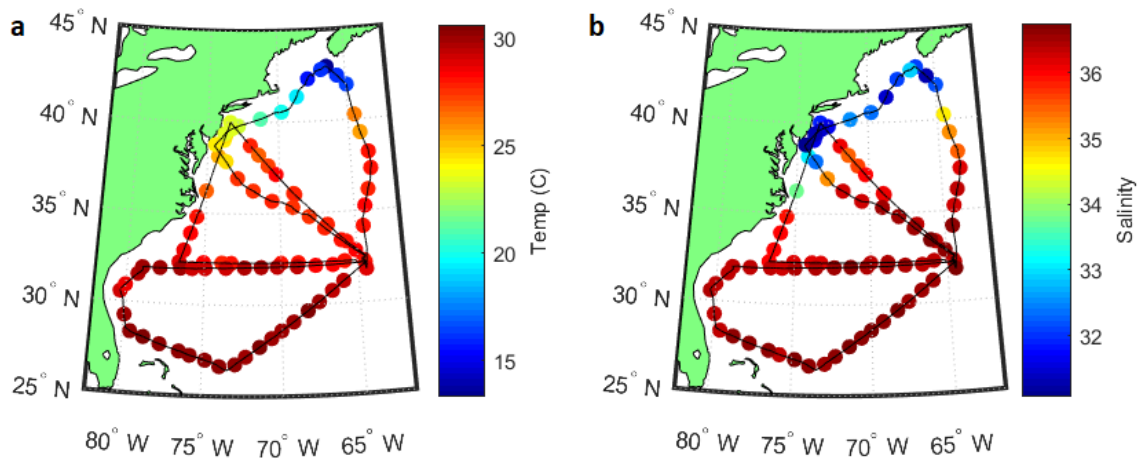


Figure 28: a) Temperature and b) Salinity measurements recorded at sampling stations across August 2015, 2016, and 2017 cruises in the Western North Atlantic.

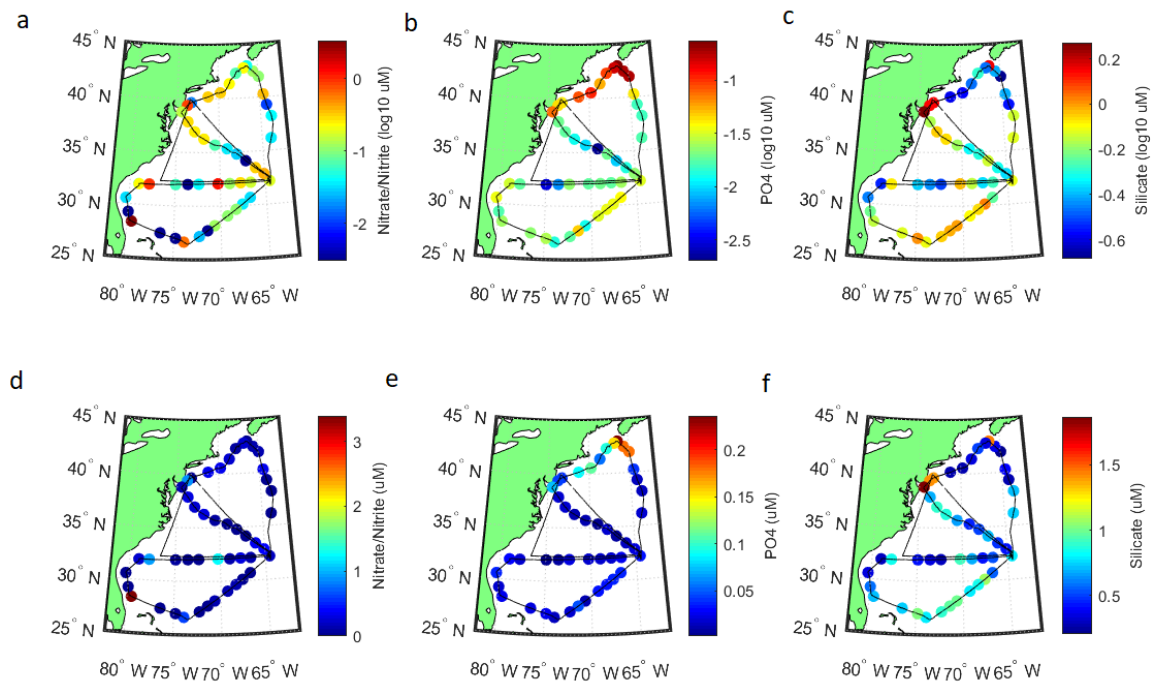
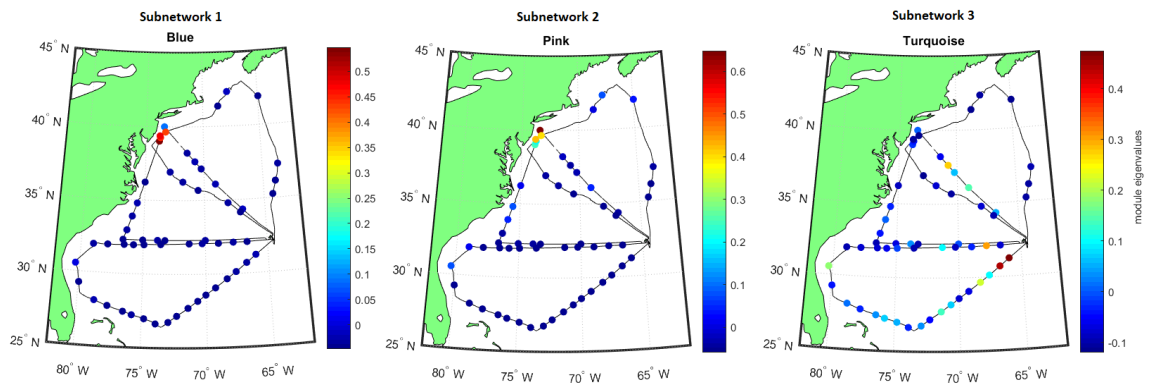


Figure 29: a)  $\log \text{NO}_2+\text{NO}_3$ , b)  $\log \text{PO}_4$ , and c)  $\log$  silicate concentrations across the 2016 and 2017 cruises. f)  $\text{NO}_2+\text{NO}_3$ , e)  $\text{PO}_4$ , and f) silicate concentrations.



**Figure 30: Maps showing the strength of associations between each WGCNA-derived module and the sampled sites. The color of each point represents the magnitude of the eigenvalue for that module corresponding to the sampled station.**

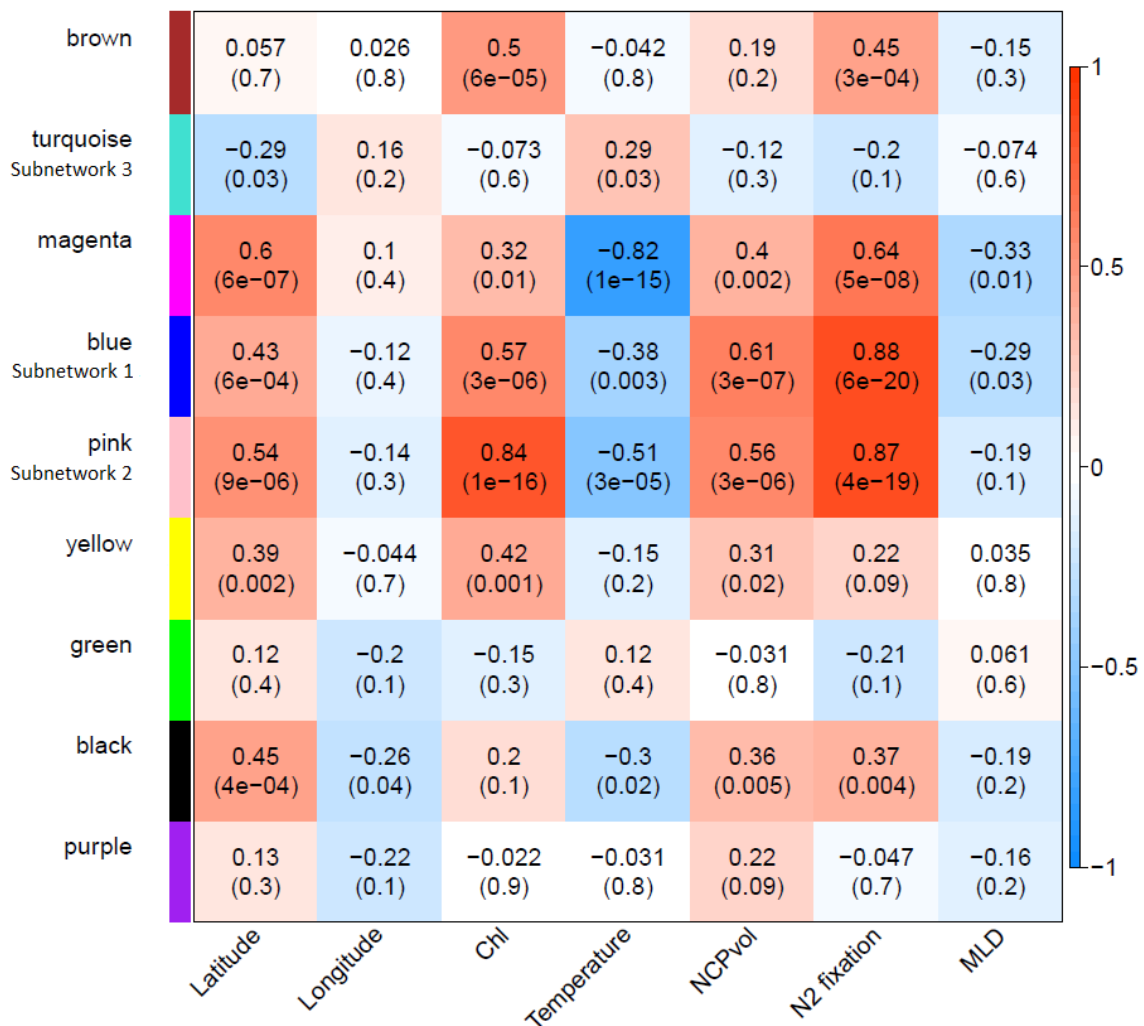


Figure 31: Heatmap illustrating correlations between WGCNA modules and metadata parameters. Positive correlations are denoted using a warm color scale, while negative correlations are shown in cool colors. Pearson correlation coefficients are given in each cell, with p-values shown in parentheses below.

Table 5: Comparison of *Prochlorococcus*, *Syncechococcus*, and SAR11 cell concentrations from the literature and from flow cytometry samples with quantitative *Prochlorococcus*, *Syncechococcus*, and SAR11 16S rDNA gene abundances.

| 16S abundances,<br>this study |                        |                        |                          |       |
|-------------------------------|------------------------|------------------------|--------------------------|-------|
|                               | Pro median<br>copies/L | Syn median<br>copies/L | SAR11 median<br>copies/L | Notes |

|                                |                      |                      |                        |                                 |
|--------------------------------|----------------------|----------------------|------------------------|---------------------------------|
| <b>2015</b>                    | 1.70E+08             | 1.85E+07             | 6.50E+08               | Samples <400km from Bermuda     |
| <b>2016</b>                    | 1.70E+07             | 3.29E+06             | 1.41E+08               | Samples <400km from Bermuda     |
| <b>2017</b>                    | 2.15E+07             | 2.90E+06             | 1.38E+08               | Samples <400km from Bermuda     |
| <b>corrected 2015</b>          | 3.58E+07             | 3.89E+06             | 1.36E+08               |                                 |
|                                |                      |                      |                        |                                 |
|                                |                      |                      |                        |                                 |
| <b>Literature counts</b>       | <b>Pro (cells/L)</b> | <b>Syn (cells/L)</b> | <b>SAR11 (cells/L)</b> | <b>Notes</b>                    |
| <b>Morris et al 2002</b>       |                      |                      | 1.6E+08 to 2E+08       | August 2001 vicinity of Bermuda |
| <b>Carlson et al 2008</b>      |                      |                      | 1.50E+08               | Summer, BATS                    |
| <b>Zinser et al 2006 AEM</b>   | 5E+07 to 1E+08       |                      |                        | August 2002 near Bermuda        |
| <b>Zinser et al 2007</b>       | 3E+07 to 6E+07       |                      |                        | late August near Bermuda        |
| <b>Singh et al 2015</b>        | 1E+07 to 2E+07       | 1.00E+07             |                        | Summer, BATS, multiple years    |
| <b>Parsons et al 2012</b>      | 2.5E+07 to 5E+07     | 1E+07 to 2E+07       | 1E+08 to 2E+08         | Summer, BATS, multiple years    |
| <b>FCM counts (this study)</b> | 2.2E+07 to 4.2E+07   | 8.1E+06 to 1.1E+07   |                        |                                 |

# Appendix C: Lagrangian studies of NCP: Assessing the effect of diel biological cycles and non-steady state systems

## Supplementary Figures

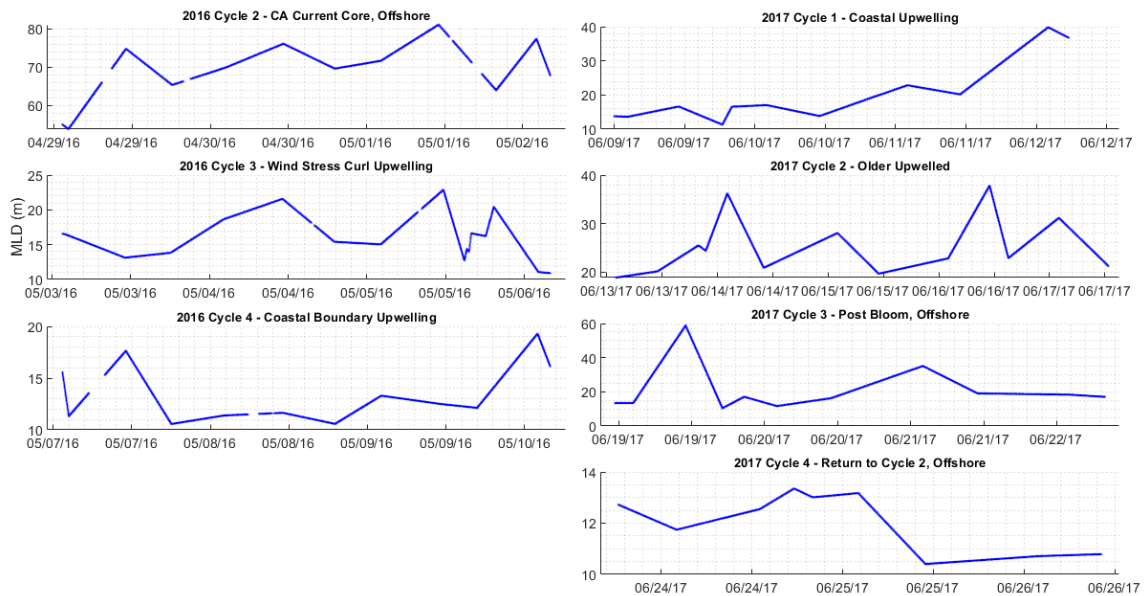
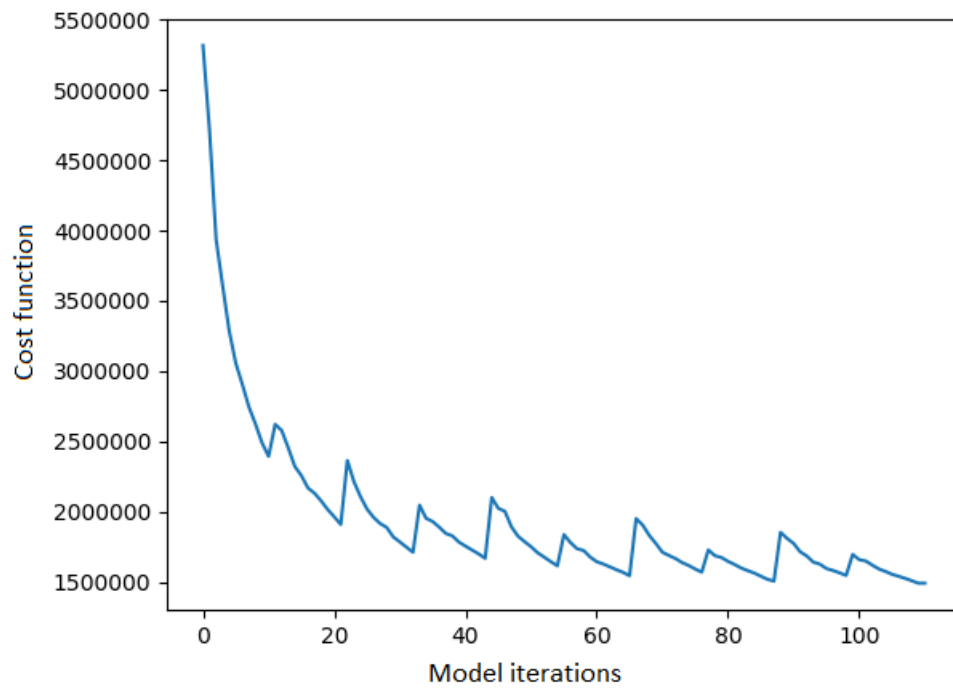
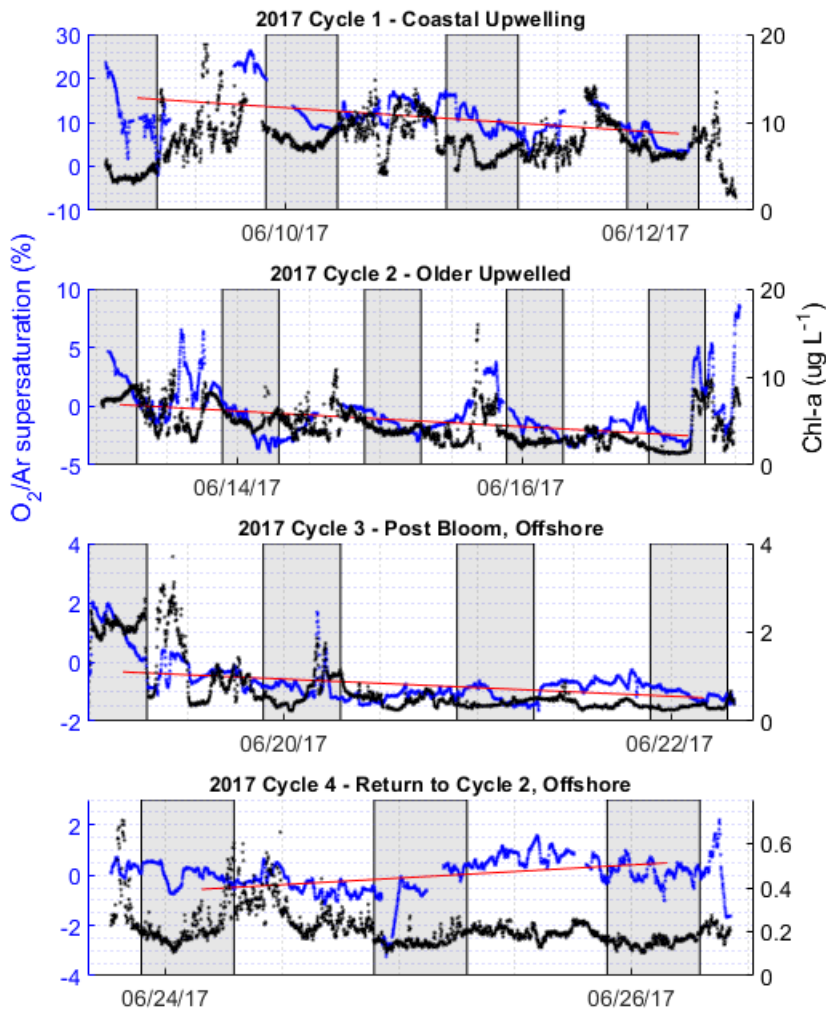


Figure 32: Interpolated MLD as determined from CTD rosette casts during each Lagrangian cycle.



**Figure 33: Graph showing the reduction across ROMS 4DVAR model iterations in the value of the cost function, which assesses the degree of change between the model and its initial conditions as well as its success in minimizing differences between the model state and the observational dataset.**



**Figure 34: Raw  $\Delta O_2/Ar$  supersaturation (blue, primary y-axis) plotted with high-resolution Chl-a measurements from the 2017 expedition (black, secondary y-axis). The line of best fit for each cycle's  $\Delta O_2/Ar$  data, derived from a linear regression against a subset of the cycle data beginning and ending at the same local time of day, is shown in red. Shaded sections of background indicate nighttime periods.**

## **Appendix D: Ongoing work: microbial productivity-diversity relationships in the Eastern North Pacific from multiple independent methods**

Wang, S., Višić, H., Mucko, M., Ljubešić, Z., Cetinic, I., Cassar, N.

### **Author contributions**

SW, NC, and IC conceived the study. ZL collected microscopy and molecular samples. SW conducted NCP measurements and productivity-diversity analyses. HV and ZL performed microscopy analyses. MM processed molecular samples and performed bioinformatic analyses.

### ***Introduction***

Marine microbes form the base of the marine food web and are responsible for driving many key ecological functions in the surface ocean including primary production and the transformation of organic matter and nutrients. As with many other ecosystem types, open questions yet remain regarding the relationship between marine microbial species richness and total ecosystem productivity. Given that ecosystems possess a broad but limited quantity of resources for growth, a common consensus holds that diversity ought to correlate positively with productivity, as a higher number of distinct species can utilize a more diverse set of resources, leading to greater overall

production (Strong, 2010). A wide variety of field studies and meta-analyses, however, have produced a full range of shapes for productivity-diversity relationships (PDRs) across marine ecosystems. These shapes range from linear increases in productivity with diversity, to negative relationships, to “humped” or unimodal relationships where the highest diversity is observed at intermediate productivity (Irigoien et al., 2004, Li, 2002, Smith, 2007, Olli et al., 2015). Other researchers have suggested that a flat relationship or no relationship at all might exist (Cermeño et al., 2013).

Determining the shape of marine PDRs has the potential to yield useful insights into processes in marine microbial ecology. In particular, characterizing the PDR of a broad ecosystem or biome may shed light on interactions between competition, coexistence, selective and nonselective predation (Strong, 2010). Such questions remain relatively little-explored in the oceanic environment, in contrast to terrestrial systems in which PDRs have enjoyed relatively intense study e.g. (Mittelbach et al., 2001, Hawkins et al., 2003, Hooper et al., 2005).

In this appendix, we present in-progress work taking advantage of parallel surveying of microplankton community structure across the eastern North Pacific in winter conducted using light microscopy and molecular sequencing techniques (Višić, 2018). Pairing these community data with high-resolution measurements from complementary EIMS O<sub>2</sub>/Ar measurements, we assess patterns of eukaryotic marine microbial diversity versus community productivity. We find that diversity metrics

derived from both molecular and microscope-based characterization of community structure are not significantly correlated with measured biological oxygen flux rates, although a general negative slope is observed. This inconclusive result may derive from the overall low magnitude of measured fluxes observed throughout this region during winter, rather than necessarily implying the absence of a distinct PDR in the eastern North Pacific.

This project possesses considerable room for expansion, given the potential extension of productivity-diversity analyses to a high-resolution microplankton image dataset collected during the same expedition using an Imaging FlowCytobot (Olson and Sosik, 2007). In addition to simply comparing patterns of O<sub>2</sub>/Ar fluxes against microplankton species richness, follow-up work might also investigate optical data collected during the cruise, which may yield metrics such as organic particle counts and concentrations for various photosynthetic pigments. IFCB data may also open up possibilities for comparing observations of particular microplankton morphological features against these abovementioned parameters.

## ***Materials and Methods***

### **O<sub>2</sub>/Ar-derived biological flux measurements**

The “Sea2space” expedition was conducted aboard the *R/V Falkor* between January 27 – February 19, 2017, traveling across the eastern North Pacific from Honolulu,

Hawaii to Portland, Oregon. Apart from a few data gaps due to technical issues, an EIMS instrument recorded surface dissolved O<sub>2</sub>/Ar continuously throughout the expedition from the ship's underway seawater intake. In brief, sample seawater is pumped at a constant rate through a membrane contactor. Dissolved gases within the seawater equilibrate with the headspace of the membrane contactor cartridge, which is then sampled by a quadrupole mass spectrometer via silica capillary (Cassar et al., 2009). The instrument measured the O<sub>2</sub>/Ar ratio in surface seawater in periods of 1.5-6 hours, followed by 30-45-minute measurements of the atmospheric O<sub>2</sub>/Ar.

Raw O<sub>2</sub>/Ar data was time-averaged at two-minute resolution, then merged with ship-recorded data. Data was quality-filtered to remove measurements recorded during anomalous spikes or lows in mass spectrometer pressure using a moving average and a moving standard deviation criterion, filtering data whenever pressure fell below 1 x 10<sup>-6</sup> or above 9 x 10<sup>-6</sup> mbar. Daily biological oxygen flux rates were calculated using the following equation:

$$\text{Biol. O}_2 \text{ flux (mmol O}_2 \text{ m}^{-3} \text{ d}^{-1}) = \frac{O_2(\%)}{Ar} * \frac{[O_2]_{sat}(\frac{\mu\text{mol}}{\text{kg}})}{1000} * \rho(S, T) * k * (\frac{1}{MLD}) \quad (\text{D-1})$$

in which  $\frac{O_2}{Ar}$  (%) represents the supersaturation (%) of the O<sub>2</sub>/Ar ratio in water relative to that of interpolated air calibrations,  $\rho$  represents the density of seawater according to the 1983 UNESCO equation of state, and  $k$  represents the piston velocity using the quadratic wind speed-dependent equation of (Wanninkhof, 1992), weighted using the 30-day wind speed history prior to measurement (Teeter et al., 2018, Reuer et

al., 2007). Wind speed history data was obtained from once-daily NARR 10-meter wind speed reanalysis data products. Mixed-layer depth (MLD) was calculated from CTD profile data from an SBE9 array following (de Boyer Montegut et al., 2004) using a density difference criterion ( $0.03 \text{ kg/m}^3$ ). On transit legs of the cruise where no CTD casts were conducted, MLDs were estimated at intervals of  $1^\circ$  latitude/longitude using Argo-derived MLD climatology for the months of January or February where appropriate (Holte et al., 2017). Finally, we linearly interpolated mixed-layer depths throughout the cruise track, using both CTD-derived and climatology-derived values.

### **Microscopy and molecular sample collection and microscopic taxonomic identification**

Sample water was collected via CTD rosette from the surface, the mixed layer depth, and the deep chlorophyll maximum. Molecular samples were taken at a total of 7 CTD stations, while samples for microscopy work were collected at 19 CTD casts. 12 additional underway samples were collected from the *R/V Falkor's* seawater intake for microscopy analyses. Detailed description of methodologies employed for light microscopy community sampling and analysis are provided in (Višić, 2018). Briefly, 50 or 100 mL of sample were settled for 24-48 hours, then analyzed and counted by microscope at x100, x200, and x400 magnification following the Utermöhl method (1958).

Molecular samples were collected via filtration rack and vacuum pump, using a 5 µm polycarbonate filter to capture the (>5 µm) size fraction, followed by a 0.2 µm filter to capture the (<5 µm) fraction. Collected filters were immediately frozen in liquid nitrogen, then stored at -80°C until analysis.

### **DNA extraction and library preparation**

DNA extraction followed (Massana et al., 1997), using a phenol:chloroform:isoamyl alcohol approach. Extracted DNA were verified using gel electrophoresis on a 1% agarose gel, then quantified using a NanoDrop spectrophotometer. Samples were subsequently outsourced for PCR library preparation and Illumina MiSeq sequencing to Molecular Research MrDNA® (www.mrdnalab.com, Shallowater, TX, USA). For PCR amplification, the 18S primer Reuk454FWD1 (5'-CCAGCASCYGCGGTAATTCC -3') and a modified version of the ReukREV3 primer (5'-ACTTTCGTTCTTGATYRATGA -3') were employed to target the 18S rDNA V4 region. PCR amplification was conducted for 28 cycles with the HotStarTaq Plus Master Mix Kit (3 min at 94°C, 28 cycles of 94°C for 30 s, 53°C for 40 s, 72°C for 1 min, followed by a final elongation step of 72°C for 5 min). PCR products were checked on a gel prior to final library preparation and sequencing by MrDNA®.

## **18S rRNA gene sequence processing**

Samples belonging to  $<5\ \mu\text{m}$  and  $>5\ \mu\text{m}$  size fractions were separately processed using an identical bioinformatic pipeline. First, reads were quality-checked using FastQC v. 0.11.5 (Andrews, 2010). Subsequently, sequences were merged, trimmed for barcodes, and filtered for chimeras by MrDNA<sup>®</sup>. Processed sequences were demultiplexing quality-filtered in QIIME 1.9.1 (Caporaso et al., 2010b) to select reads with a length of greater than 250 bp,  $>Q20$  quality, no ambiguous base calls, and a homopolymer length of  $<6\text{bp}$  using the `split_libraries.py` script. Following this demultiplexing step, OTU picking was performed using open-reference OTU picking at 97% identity, utilizing the SILVA v123 rRNA reference database (Pruesse et al., 2007).

## **Statistical analysis**

In this study, we considered only molecular and microscopy data from surface and near-surface samples. Following processing of  $\text{O}_2/\text{Ar}$  data, we filtered our microscopy dataset to remove stations at which phytoplankton community data was analyzed but no contemporaneous EIMS measurements were collected. This left 17 CTD stations in addition to all 12 underway stations.

Eukaryotic molecular data were binned at the fourth taxonomic rank, yielding 74 taxa in the  $>5\ \mu\text{m}$  size fraction sample data and 73 taxa in the  $<5\ \mu\text{m}$  samples. Subsequent calculation of sample diversity metrics, ordination, and testing of productivity-diversity

relationships were performed in R (R Core Team, 2017) using the package ‘vegan’ (Okansen et al., 2017).

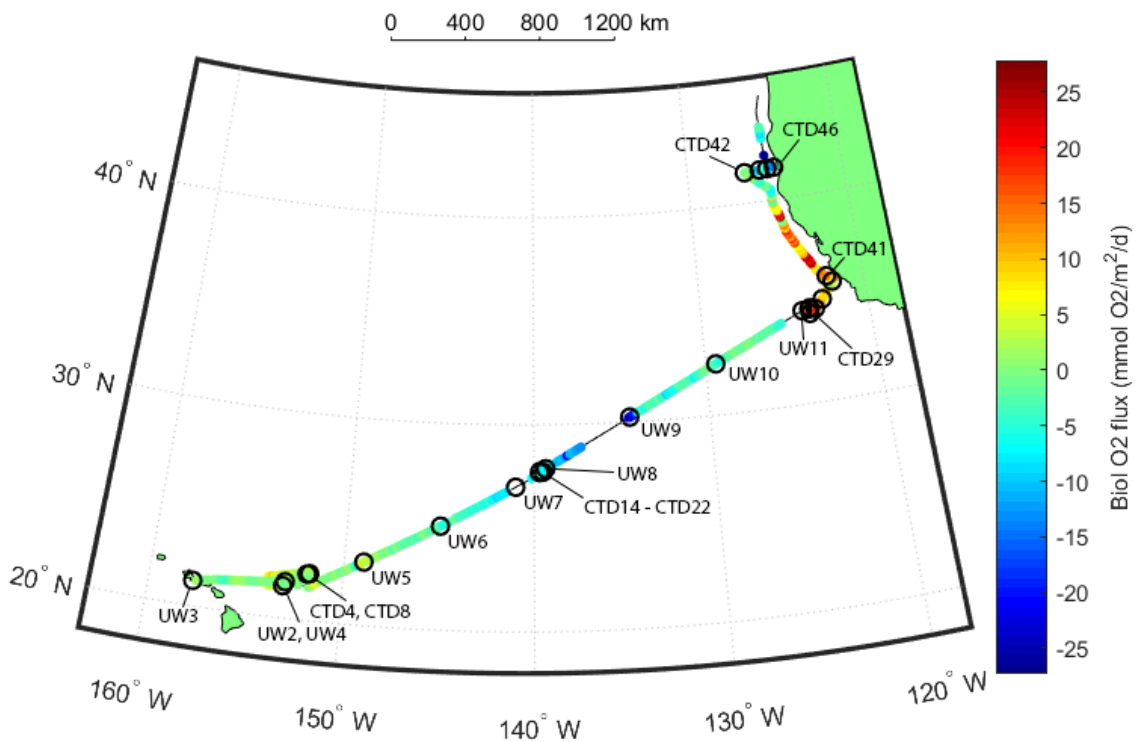
## **Results and Discussion**

### **Biological oxygen flux rates and patterns**

Calculated biological oxygen flux measurements over the course of the Sea2Space cruise ranged between -27 and 26 mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup> (Fig. 35). Spatially, flux rates were balanced (O<sub>2</sub>/Ar near or at atmospheric equilibrium) in the vicinity of Hawaii and throughout the first third of the transect between Hawaii and the California coast. Subsequently, biological oxygen flux shifted towards negative (undersaturated) values during the second half of the transect, before rising to values of ~15-20 mmol O<sub>2</sub> m<sup>-2</sup> d<sup>-1</sup> within the California Current and along the California coast. Finally, flux values fell once again towards equilibrium and slightly undersaturated values as the *Falkor* traveled north along the US West Coast.

Overall, the range of observed biological oxygen fluxes on this cruise was relatively narrow, likely due to seasonally low productivity in the North Pacific during the height of winter. O<sub>2</sub>/Ar-based productivity surveys of similar spatial scale during the growing season typically yield a broader range of measurements as well as higher peaks in measured NCP, as seen for instance in the western North Atlantic (Wang et al., 2018), Arctic Ocean (Ulfsbo et al., 2014), and off the coast of the Western Antarctic Peninsula

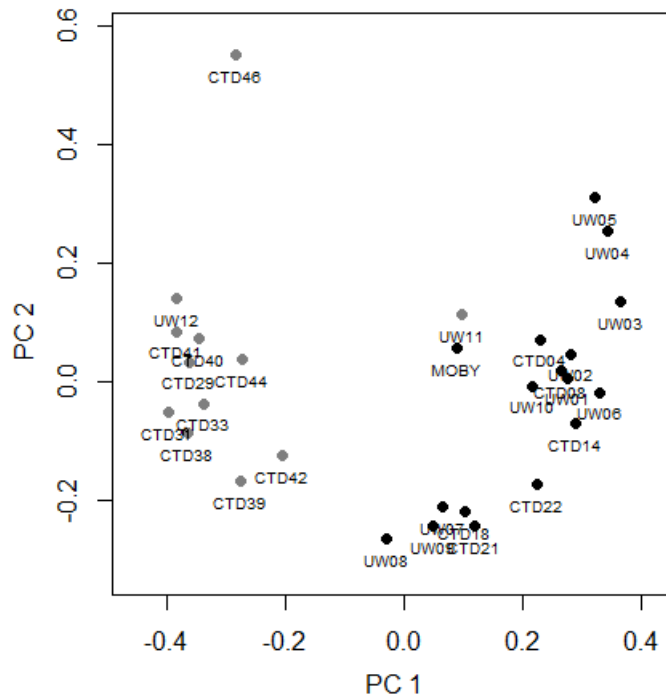
(Eveleth et al., 2017). The increase in biological oxygen flux closer to the California coast may be driven by a correspondingly shallower mixed-layer depth along the coast (10-50 m) relative to further offshore (20-140 m). Stronger vertical stratification near the coast may alleviate light limitation, while pronounced upwelling of nutrients from depth in the California Current Ecosystem (Ohman et al., 2013, Hutchins et al., 1998) may potentially aid in fueling somewhat elevated productivity rates here.



**Figure 35: Map of NCP measurements and cruise track for the January 2017 Sea2Space expedition. Calculated MLD-integrated NCP rates are shown using a color scale. Locations of CTD stations as well as underway sampling sites are shown using open black circles, with selected sites labeled.**

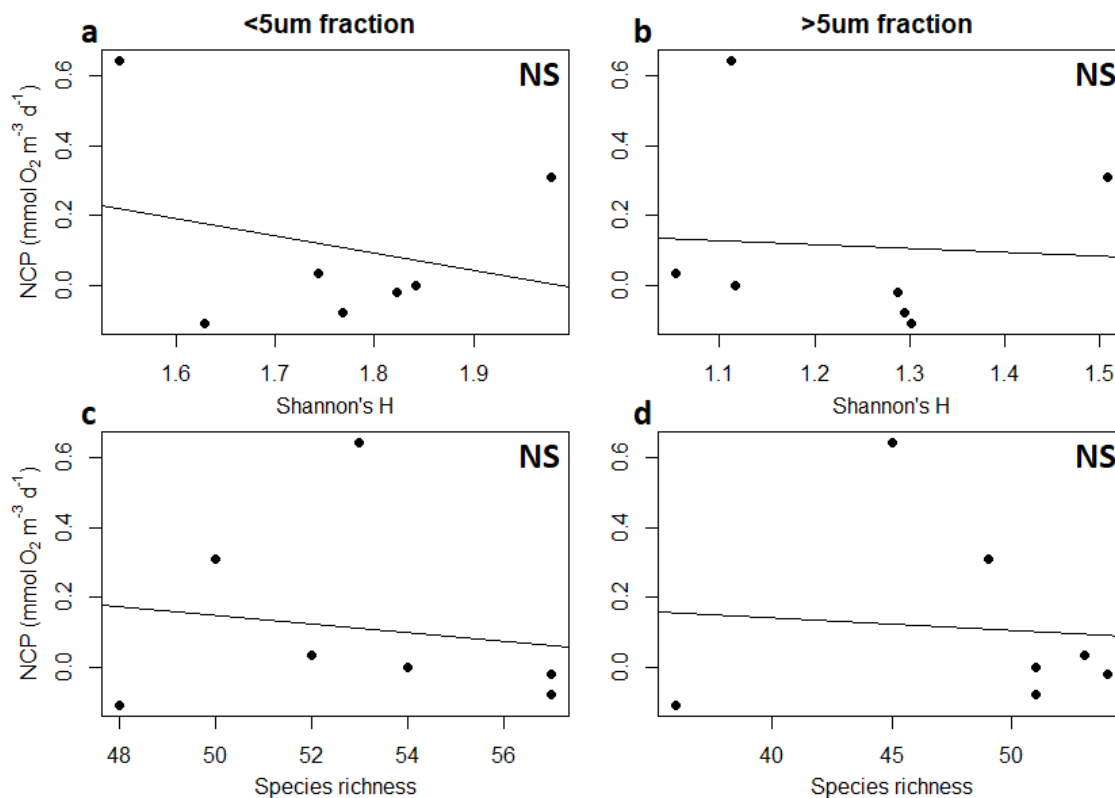
## Productivity-diversity relationships

An overall pattern of diversity over the study area distinguished coastal eukaryotic microplankton communities from microplankton characteristic of open-ocean samples. Principal Coordinates Analysis (PCoA) of the microscopy community data reveals this trend clearly, with distinctly separate clusters of coastal (<400 km from coastline) and pelagic samples following ordination using Bray-Curtis dissimilarity (Fig. 36).



**Figure 36: Principal Coordinates Analysis (PCoA) of microscopy community samples using Bray-Curtis ecological dissimilarity. Samples collected within <400 km of the U.S. West Coast are depicted in gray, while open-ocean samples and a sample collected near Hawaii (UW3 only) are shown in black.**

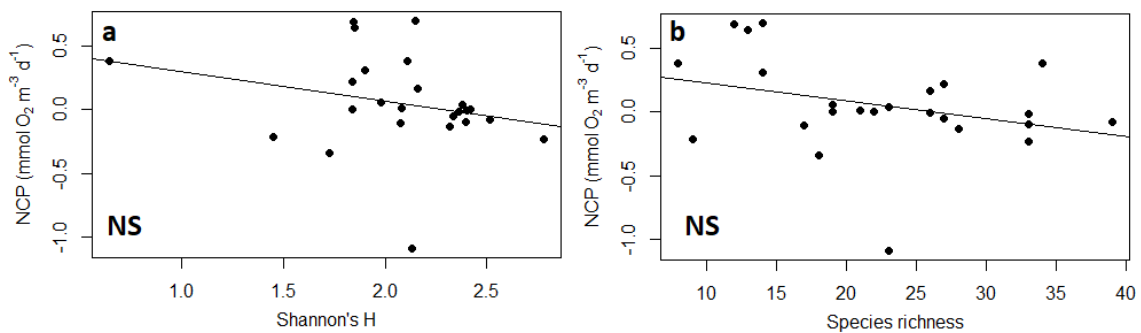
Overall, however, no statistically significant patterns were observed between productivity and eukaryotic microbial diversity determined using either microscopy or molecular approaches. The absence of clear patterns between productivity and diversity in our results appear to be partially driven by the low range of biological oxygen flux values encountered during our study (volumetric rates of  $-0.1$  to  $0.7$   $\text{mmol O}_2 \text{ m}^{-3} \text{ d}^{-1}$ ). Furthermore, a low number of molecular 18S samples collected during this expedition greatly limited our ability to infer trends between biological oxygen flux and community diversity derived from this dataset (Fig. 37). Neither the  $>5 \mu\text{m}$  nor the  $<5 \mu\text{m}$  size fraction displayed a clear or significant PDR.



**Figure 37: Scatter plots of a, b) eukaryotic community Shannon's H diversity and c, d) 18S community raw species richness at the 4th taxonomic rank plotted against calculated volumetric NCP rates for 18S molecular sampling stations. Plots in the left column (a, c) and right column (b, d) correspond to samples collected from the <5 µm fraction and the >5 µm fraction respectively. Best-fit lines for linear regressions are shown in black. All relationships are statistically non-significant (p>0.05).**

Even for the relatively more comprehensive microscopy dataset, although both Shannon's H diversity and raw species richness appear to be potentially negatively related to NCP, neither relationship is ultimately statistically significant (Fig. 38). A negative PDR would theoretically be in agreement with findings from similar productivity-diversity surveys and meta-analyses in the aquatic ecology literature (Wang et al., 2018, Wang et al., In review, Smith, 2007, Olli et al., 2015) and might imply

competitive exclusion leading to falling diversity at higher productivities as successful microplankton taxa out-compete taxa of inferior fitness (Hardin, 1960). However, additional data will be required to further investigate these dynamics in the winter eastern North Pacific.



**Figure 38: Scatter plots of a) eukaryotic community Shannon's H diversity and b) raw eukaryotic species richness determined from microscopy analyses and plotted against calculated volumetric NCP rates. Best-fit lines for linear regressions are shown in black. All relationships are statistically non-significant ( $p > 0.05$ ).**

## ***Conclusions and Future Work***

Ultimately, given the limits of the availability community data as well as the narrow range of productivity measurements conducted during this expedition, this preliminary exploration of PDRs in the winter eastern North Pacific remains inconclusive overall. Nevertheless, this work lays a foundation for follow-up investigations that may provide deeper insight into relationships between community production and microplankton ecology in this region. Optical data including backscatter and fluorescence measurements may provide another high-resolution dataset against

which our underway O<sub>2</sub>/Ar-based NCP rates may be compared. Higher-resolution community data, such as that obtained from concurrently-collected IFCB instrument images (Olson and Sosik, 2007) may also serve to increase the robustness of productivity-diversity relationships constructed using data from this cruise. Finally, leveraging the use of machine-learning-based image interpretation, microplankton morphological traits, size, and other characteristics might be compared to patterns of particle concentrations, photosynthetic pigment distributions, and productivity observed during the Sea2Space cruise. Such an approach holds promise not just for further investigation related to this project, but also for exploration of the links between marine microbial rates, characteristics, and community structure in general.

### ***Acknowledgements***

We thank the crew and marine technicians of the *R/V Falkor* as well as the Schmidt Ocean Institute for their assistance and their role in making the Sea2Space expedition possible. IC was funded through the Schmidt Ocean Institute as well as NASA's PACE mission.

## References

1958. Methods of collecting plankton for various purposes are discussed. AU - Utermöhl, Hans. *SIL Communications, 1953-1996*, 9, 1-38.
2018. *Northeast US Shelf LTER* [Online]. Available: <https://web.who.edu/nes-lter/> [Accessed 11/25/2018 2018].
- AITCHISON, J. 1981. A new approach to null correlations of proportions. *Journal of the International Association for Mathematical Geology*, 13, 175-189.
- AITCHISON, J. 1982. The statistical-analysis of compositional data. *Journal of the Royal Statistical Society Series B-Methodological*, 44, 139-177.
- ANDREWS, S. 2010. *FastQC: a quality control tool for high throughput sequence data* [Online]. Available: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> [Accessed].
- BAKUN, A. 1990. Global climate change and intensification of coastal ocean upwelling. *Science*, 247, 198-201.
- BARBER, R. T. & HISCOCK, M. R. 2006. A rising tide lifts all phytoplankton: Growth response of other phytoplankton taxa in diatom-dominated blooms. *Global Biogeochemical Cycles*, 20.
- BATES, N. R., BEST, M. H. P., NEELY, K., GARLEY, R., DICKSON, A. G. & JOHNSON, R. J. 2012. Detecting anthropogenic carbon dioxide uptake and ocean acidification in the North Atlantic Ocean. *Biogeosciences*, 9, 2509-2522.
- BAUER, J. E., CAI, W.-J., RAYMOND, P. A., BIANCHI, T. S., HOPKINSON, C. S. & REGNIER, P. A. G. 2013. The changing carbon cycle of the coastal ocean. *Nature*, 504, 61-70.
- BENDER, M. L., KINTER, S., CASSAR, N. & WANNINKHOF, R. 2011. Evaluating gas transfer velocity parameterizations using upper ocean radon distributions. *Journal of Geophysical Research: Oceans*, 116.
- BENTZON-TILIA, M., TRAVING, S. J., MANTIKCI, M., KNUDSEN-LEERBECK, H., HANSEN, J. L. S., MARKAGER, S. & RIEMANN, L. 2015. Significant N(2) fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in

- two temperate estuaries. *The ISME Journal*, 9, 273-285.
- BERGER, W. H. & WEFER, G. 1990. Export production: seasonality and intermittency, and paleoceanographic implications. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 89, 245-254.
- BOMBAR, D., PAERL, R. W. & RIEMANN, L. 2016. Marine Non-Cyanobacterial Diazotrophs: Moving beyond Molecular Detection. *Trends in Microbiology*, 24, 916-927.
- BOYD, P. & NEWTON, P. 1995. Evidence of the potential influence of planktonic community structure on the interannual variability of particulate organic-carbon flux. *Deep-Sea Research Part I-Oceanographic Research Papers*, 42, 619-639.
- BRADLEY, I. M., PINTO, A. J. & GUEST, J. S. 2016. Design and Evaluation of Illumina MiSeq-Compatible, 18S rRNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities. *Applied and Environmental Microbiology*, 82, 5878-5891.
- BRAND, L. E., SUNDA, W. G. & GUILLARD, R. R. L. 1986. Reduction of marine phytoplankton reproduction rates by copper and cadmium. *Journal of Experimental Marine Biology and Ecology*, 96, 225-250.
- BRIX, H., GRUBER, N., KARL, D. M. & BATES, N. R. 2006. On the relationships between primary, net community, and export production in subtropical gyres. *Deep-Sea Research Part II-Topical Studies in Oceanography*, 53, 698-717.
- BRULAND, K. W., MIDDAG, R. & LOHAN, M. C. 2014. Controls of Trace Metals in Seawater. In: HOLLAND, H. D. & TUREKIAN, K. K. (eds.) *Treatise on Geochemistry, 2nd Edition*. Oxford: Elsevier.
- BRULAND, K. W., RUE, E. L., SMITH, G. J. & DITULLIO, G. R. 2005. Iron, macronutrients and diatom blooms in the Peru upwelling regime: brown and blue waters of Peru. *Marine Chemistry*, 93, 81-103.
- BUCHAN, A., LECLEIR, G. R., GULVIK, C. A. & GONZALEZ, J. M. 2014. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nature Reviews Microbiology*, 12, 686-698.

- BUESSELER, K. O. 1998. The decoupling of production and particulate export in the surface ocean. *Global Biogeochemical Cycles*, 12, 297-310.
- BURKHOLDER, J. M., GLIBERT, P. M. & SKELTON, H. M. 2008. Mixotrophy, a major mode of nutrition for harmful algal species in eutrophic waters. *Harmful Algae*, 8, 77-93.
- CAI, W.-J. 2011. Estuarine and coastal ocean carbon paradox: CO<sub>2</sub> sinks or sites of terrestrial carbon incineration? *Annual Review of Marine Science*, 3, 123-145.
- CAPONE, D., BURNS, J., MONTOYA, J., SUBRAMANIAM, A., MAHAFFEY, C., GUNDERSON, T., MICHAELS, A. & CARPENTER, E. 2005. Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochemical Cycles*, 19.
- CAPORASO, J. G., BITTINGER, K., BUSHMAN, F. D., DESANTIS, T. Z., ANDERSEN, G. L. & KNIGHT, R. 2010a. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, 26, 266-267.
- CAPORASO, J. G., KUCZYNSKI, J., STOMBAUGH, J., BITTINGER, K., BUSHMAN, F. D., COSTELLO, E. K., FIERER, N., PENA, A. G., GOODRICH, J. K., GORDON, J. I., HUTTLEY, G. A., KELLEY, S. T., KNIGHTS, D., KOENIG, J. E., LEY, R. E., LOZUPONE, C. A., MCDONALD, D., MUEGGE, B. D., PIRRUNG, M., REEDER, J., SEVINSKY, J. R., TUMBAUGH, P. J., WALTERS, W. A., WIDMANN, J., YATSUNENKO, T., ZANEVELD, J. & KNIGHT, R. 2010b. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7, 335-336.
- CAPORASO, J. G., LAUBER, C. L., WALTERS, W. A., BERG-LYONS, D., LOZUPONE, C. A., TURNBAUGH, P. J., FIERER, N. & KNIGHT, R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 4516-4522.
- CARLSON, C. A., DUCKLOW, H. W. & MICHAELS, A. F. 1994. Annual flux of dissolved organic-carbon from the euphotic zone in the northwestern Sargasso Sea. *Nature*, 371, 405-408.
- CARLSON, C. A., MORRIS, R., PARSONS, R., TREUSCH, A. H., GIOVANNONI, S. J. & VERGIN, K. 2008. Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J*, 3, 283-295.

- CARON, D. A., ALEXANDER, H., ALLEN, A. E., ARCHIBALD, J. M., ARMBRUST, E. V., BACHY, C., BELL, C. J., BHARTI, A., DYHRMAN, S. T., GUIDA, S. M., HEIDELBERG, K. B., KAYE, J. Z., METZNER, J., SMITH, S. R. & WORDEN, A. Z. 2017. Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nature Reviews Microbiology*, 15, 6-20.
- CASSAR, N., NEVISON, C. & MANIZZA, M. 2014. Correcting oceanic O<sub>2</sub>/Ar-net community production estimates for vertical mixing using N<sub>2</sub>O observations. *Geophysical Research Letters*, 41, 8961-8970.
- CASSAR, N., BARNETT, B., BENDER, M., KAISER, J., HAMME, R. & TILBROOK, B. 2009. Continuous High-Frequency Dissolved O<sub>2</sub>/Ar Measurements by Equilibrator Inlet Mass Spectrometry. *Analytical Chemistry*, 81, 1855-1864.
- CASSAR, N., DIFIIORE, P. J., BARNETT, B. A., BENDER, M. L., BOWIE, A. R., TILBROOK, B., PETROU, K., WESTWOOD, K. J., WRIGHT, S. W. & LEFEVRE, D. 2011. The influence of iron and light on net community production in the Subantarctic and Polar Frontal Zones. *Biogeosciences*, 8, 227-237.
- CASSAR, N., TANG, W. Y., GABATHULER, H. & HUANG, K. 2018. Method for High Frequency Underway N<sub>2</sub> Fixation Measurements: Flow-Through Incubation Acetylene Reduction Assays by Cavity Ring Down Laser Absorption Spectroscopy (FARACAS). *Analytical Chemistry*, 90, 2839-2851.
- CASSAR, N., WRIGHT, S. W., THOMSON, P. G., TRULL, T. W., WESTWOOD, K. J., DE SALAS, M., DAVIDSON, A., PEARCE, I., DAVIES, D. M. & MATEAR, R. J. 2015. The relation of mixed-layer net community production to phytoplankton community composition in the Southern Ocean. *Global Biogeochemical Cycles*, 29, 446-462.
- CASTRO-MORALES, K., CASSAR, N., SHOOSMITH, D. R. & KAISER, J. 2013. Biological production in the Bellingshausen Sea from oxygen-to-argon ratios and oxygen triple isotopes. *Biogeosciences*, 10, 2273-2291.
- CERMEÑO, P., RODRÍGUEZ-RAMOS, T., DORNELAS, M., FIGUEIRAS, F. G., MARAÑÓN, E., TEIXEIRA, I. G. & VALLINA, S. 2013. Species richness in marine phytoplankton communities is not correlated to ecosystem productivity. *Marine Ecology Progress Series*, 488, 1-9.

- CHAVEZ, F. P., MESSIÉ, M. & PENNINGTON, J. T. 2010. Marine Primary Production in Relation to Climate Variability and Change. *Annual Review of Marine Science*, 3, 227-260.
- CRAIG, H. & HAYWARD, T. 1987. Oxygen Supersaturation in the Ocean: Biological Versus Physical Contributions. *Science*, 235, 199.
- CUTTER, G., ANDERSSON, P., CODISPOTI, L., CROOT, P., FRANÇOIS, R., LOHAN, M. C., OBATA, H. & RUTGERS V. D. LOEFF, M. 2010. Sampling and Sample-handling Protocols for GEOTRACES Cruises. GEOTRACES.
- DE BOYER MONTEGUT, C., MADEC, G., FISCHER, A. S., LAZAR, A. & IUDICONE, D. 2004. Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research-Oceans*, 109.
- DE VARGAS, C., AUDIC, S., HENRY, N., DECELLE, J., MAHE, F., LOGARES, R., LARA, E., BERNEY, C., LE BESCOT, N., PROBERT, I., CARMICHAEL, M., POULAIN, J., ROMAC, S., COLIN, S., AURY, J. M., BITTNER, L., CHAFFRON, S., DUNTHORN, M., ENGELEN, S., FLEGONTOVA, O., GUIDI, L., HORAK, A., JAILLON, O., LIMA-MENDEZ, G., LUKES, J., MALVIYA, S., MORARD, R., MULOT, M., SCALCO, E., SIANO, R., VINCENT, F., ZINGONE, A., DIMIER, C., PICHERAL, M., SEARSON, S., KANDELS-LEWIS, S., ACINAS, S. G., BORK, P., BOWLER, C., GORSKY, G., GRIMSLEY, N., HINGAMP, P., IUDICONE, D., NOT, F., OGATA, H., PESANT, S., RAES, J., SIERACKI, M. E., SPEICH, S., STEMMANN, L., SUNAGAWA, S., WEISSENBACH, J., WINCKER, P., KARSENTI, E. & TARA OCEANS, C. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348.
- DELMONT, T. O., QUINCE, C., SHAIKER, A., ESEN, O. C., LEE, S. T. M., RAPPE, M. S., MACLELLAN, S. L., LUCKER, S. & EREN, A. M. 2018. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology*, 3, 804+.
- DONEY, S. C. 2010. The Growing Human Footprint on Coastal and Open-Ocean Biogeochemistry. *Science*, 328, 1512-1516.
- DONEY, S. C., MAHOWALD, N., LIMA, I., FEELY, R. A., MACKENZIE, F. T., LAMARQUE, J. F. & RASCH, P. J. 2007. Impact of anthropogenic atmospheric nitrogen and sulfur deposition on ocean acidification and the inorganic carbon

- system. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 14580-14585.
- DONG, S., SPRINTALL, J., GILLE, S. T. & TALLEY, L. 2008. Southern Ocean mixed-layer depth from Argo float profiles. *Journal of Geophysical Research: Oceans*, 113.
- DOYLE, J. D., JIANG, Q. F., CHAO, Y. & FARRARA, J. 2009. High-resolution real-time modeling of the marine atmospheric boundary layer in support of the AOSN-II field campaign. *Deep-Sea Research Part II-Topical Studies in Oceanography*, 56, 87-99.
- DUARTE, C. M., REGAUDIE-DE-GIOUX, A., ARRIETA, J. M., DELGADO-HUERTAS, A. & AGUSTI, S. 2013. The Oligotrophic Ocean Is Heterotrophic. *Annual Review of Marine Science*, Vol 5, 5, 551-569.
- DUCKLOW, H. W. & DONEY, S. C. 2013. What Is the Metabolic State of the Oligotrophic Ocean? A Debate. *Annual Review of Marine Science*, 5, 525-533.
- DUGDALE, R. C. & GOERING, J. J. 1967. Uptake of new and regenerated forms of nitrogen in primary productivity. *Limnology and Oceanography*, 12, 196-&.
- EDGAR, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460-2461.
- EDGAR, R. C., HAAS, B. J., CLEMENTE, J. C., QUINCE, C. & KNIGHT, R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27, 2194-2200.
- EMERSON, S., STUMP, C. & NICHOLSON, D. 2008. Net biological oxygen production in the ocean: Remote in situ measurements of O<sub>2</sub> and N<sub>2</sub> in surface waters. *Global Biogeochemical Cycles*, 22.
- ESTAPA, M. L., SIEGEL, D. A., BUESSELER, K. O., STANLEY, R. H. R., LOMAS, M. W. & NELSON, N. B. 2015. Decoupling of net community and export production on submesoscales. *Global Biogeochemical Cycles*, 29, 1266-1282.
- ESTRADA, E. 2007. Characterization of topological keystone species: Local, global and "meso-scale" centralities in food webs. *Ecological Complexity*, 4, 48-57.
- EVELETH, R., CASSAR, N., SHERRELL, R. M., DUCKLOW, H., MEREDITH, M. P., VENABLES, H. J., LIN, Y. & LI, Z. 2017. Ice melt influence on summertime net

community production along the Western Antarctic Peninsula. *Deep Sea Research Part II: Topical Studies in Oceanography*, 139, 89-102.

- FADROSH, D. W., MA, B., GAJER, P., SENGAMALAY, N., OTT, S., BROTMAN, R. M. & RAVEL, J. 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, 2.
- FALKOWSKI, P., LAWS, E. A., BARBER, R. T. & MURRAY, J. W. 2003. Phytoplankton, and their role in primary, new and export production. *In*: FASHAM, M. J. R. (ed.) *Ocean Biogeochemistry*. New York: Springer.
- FALKOWSKI, P. G., FLAGG, C. N., ROWE, G. T., SMITH, S. L., WHITLEDGE, T. E. & WIRICK, C. D. 1988. The fate of a spring phytoplankton bloom – export or oxidation. *Continental Shelf Research*, 8, 457-484.
- FARIA, M., BORDIN, N., KIZINA, J., HARDER, J., DEVOS, D. & LAGE, O. M. 2018. Planctomycetes attached to algal surfaces: Insight into their genomes. *Genomics*, 110, 231-238.
- FARRELLY, V., RAINEY, F. A. & STACKEBRANDT, E. 1995. Effect of genome size and rrn gene copy number on PCR amplification of 16S ribosomal-RNA genes from a mixture of bacterial species. *Applied and Environmental Microbiology*, 61, 2798-2801.
- FAWCETT, S. E., LOMAS, M. W., WARD, B. B. & SIGMAN, D. M. 2014. The counterintuitive effect of summer-to-fall mixed layer deepening on eukaryotic new production in the Sargasso Sea. *Global Biogeochemical Cycles*, 28, 86-102.
- FERNANDEZ, C., FARIÁS, L. & ULLOA, O. 2011. Nitrogen Fixation in Denitrified Marine Waters. *PLOS ONE*, 6, e20539.
- FERRON, S., WILSON, S. T., MARTINEZ-GARCIA, S., QUAY, P. D. & KARL, D. M. 2015. Metabolic balance in the mixed layer of the oligotrophic North Pacific Ocean from diel changes in O<sub>2</sub>/Ar saturation ratios. *Geophysical Research Letters*, 42, 3421-3430.
- FIELD, C. B., BEHRENFELD, M. J., RANDERSON, J. T. & FALKOWSKI, P. 1998. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, 281, 237-240.

- FIGUEROA, R. I., CUADRADO, A., STUKEN, A., RODRIGUEZ, F. & FRAGA, S. 2014. Ribosomal DNA Organization Patterns within the Dinoflagellate Genus *Alexandrium* as Revealed by FISH: Life Cycle and Evolutionary Implications. *Protist*, 165, 343-363.
- FREITAS, S., HATOSY, S., FUHRMAN, J. A., HUSE, S. M., MARK WELCH, D. B., SOGIN, M. L. & MARTINY, A. C. 2012. Global distribution and diversity of marine Verrucomicrobia. *The Isme Journal*, 6, 1499.
- FUGLISTER, F. C. 1955. Alternative analyses of current surveys. *Deep Sea Research (1953)*, 2, 213-229.
- GARCIA, H. E. & GORDON, L. I. 1992. Oxygen solubility in seawater – better fitting equations. *Limnology and Oceanography*, 37, 1307-1312.
- GÉRIKAS RIBEIRO, C., LOPES DOS SANTOS, A., MARIE, D., PEREIRA BRANDINI, F. & VAULOT, D. 2018. Small eukaryotic phytoplankton communities in tropical waters off Brazil are dominated by symbioses between Haptophyta and nitrogen-fixing cyanobacteria. *The ISME Journal*, 12, 1360-1374.
- GIFFORD, S. M., SHARMA, S., RINTA-KANTO, J. M. & MORAN, M. A. 2011. Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *Isme Journal*, 5, 461-472.
- GILBERT, J. A., STEELE, J. A., CAPORASO, J. G., STEINBRUCK, L., REEDER, J., TEMPERTON, B., HUSE, S., MCHARDY, A. C., KNIGHT, R., JOINT, I., SOMERFIELD, P., FUHRMAN, J. A. & FIELD, D. 2012. Defining seasonal marine microbial community dynamics. *Isme j*, 6, 298-308.
- GOBLER, C. J., BERRY, D. L., DYHRMAN, S. T., WILHELM, S. W., SALAMOV, A., LOBANOV, A. V., ZHANG, Y., COLLIER, J. L., WURCH, L. L., KUSTKA, A. B., DILL, B. D., SHAH, M., VERBERKMOES, N. C., KUO, A., TERRY, A., PANGILINAN, J., LINDQUIST, E. A., LUCAS, S., PAULSEN, I. T., HATTENRATH-LEHMANN, T. K., TALMAGE, S. C., WALKER, E. A., KOCH, F., BURSON, A. M., MARCOVAL, M. A., TANG, Y. Z., LECLEIR, G. R., COYNE, K. J., BERG, G. M., BERTRAND, E. M., SAITO, M. A., GLADYSHEV, V. N. & GRIGORIEV, I. V. 2011. Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 4352-4357.

- GOBLER, C. J., LONSDALE, D. J. & BOYER, G. L. 2005. A review of the causes, effects, and potential management of harmful brown tide blooms caused by *Aureococcus anophagefferens* (Hargraves et Sieburth). *Estuaries*, 28, 726-749.
- GOBLER, C. J., RENAGHAN, M. J. & BUCK, N. J. 2002. Impacts of nutrients and grazing mortality on the abundance of *Aureococcus anophagefferens* during a New York brown tide bloom. *Limnology and Oceanography*, 47, 129-141.
- GOEBEL, N. L., EDWARDS, C. A., ZEHR, J. P., FOLLOWS, M. J. & MORGAN, S. G. 2013. Modeled phytoplankton diversity and productivity in the California Current System. *Ecological Modelling*, 264, 37-47.
- GOECKE, F., THIEL, V., WIESE, J., LABES, A. & IMHOFF, J. F. 2013. Algae as an important environment for bacteria - phylogenetic relationships among new bacterial species isolated from algae. *Phycologia*, 52, 14-24.
- GONZÁLEZ-GUERRERO, M., MATTHIADIS, A., SAEZ, Á. & LONG, T. 2014. Fixating on metals: new insights into the role of metals in nodulation and symbiotic nitrogen fixation. *Frontiers in Plant Science*, 5.
- GRANELI, E., EDVARDBSEN, B., ROELKE, D. L. & HAGSTROM, J. A. 2012. The ecophysiology and bloom dynamics of *Prymnesium* spp. *Harmful Algae*, 14, 260-270.
- GROSSART, H. P., LEVOLD, F., ALLGAIER, M., SIMON, M. & BRINKHOFF, T. 2005. Marine diatom species harbour distinct bacterial communities. *Environmental Microbiology*, 7, 860-873.
- GUIDI, L., CHAFFRON, S., BITTNER, L., EVEILLARD, D., LARHLIMI, A., ROUX, S., DARZI, Y., AUDIC, S., BERLINE, L., BRUM, J. R., COELHO, L. P., ESPINOZA, J. C. I., MALVIYA, S., SUNAGAWA, S., DIMIER, C., KANDELS-LEWIS, S., PICHERAL, M., POULAIN, J., SEARSON, S., STEMMANN, L., NOT, F., HINGAMP, P., SPEICH, S., FOLLOWS, M., KARP-BOSS, L., BOSS, E., OGATA, H., PESANT, S., WEISSENBACH, J., WINCKER, P., ACINAS, S. G., BORK, P., DE VARGAS, C., IUDICONE, D., SULLIVAN, M. B., RAES, J., KARSENTI, E., BOWLER, C., GORSKY, G. & TARA OCEANS CONSORTIUM, C. 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature*, 532, 465-+.

- GUIDI, L., STEMMANN, L., JACKSON, G. A., IBANEZ, F., CLAUSTRE, H., LEGENDRE, L., PICHERAL, M. & GORSKY, G. 2009. Effects of phytoplankton community on production, size and export of large aggregates: A world-ocean analysis. *Limnology and Oceanography*, 54, 1951-1963.
- GUILLOU, L., VIPREY, M., CHAMBOUVET, A., WELSH, R. M., KIRKHAM, A. R., MASSANA, R., SCANLAN, D. J. & WORDEN, A. Z. 2008. Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environmental Microbiology*, 10, 3349-3365.
- HAMME, R. C., CASSAR, N., LANCE, V. P., VAILLANCOURT, R. D., BENDER, M. L., STRUTTON, P. G., MOORE, T. S., DEGRANDPRE, M. D., SABINE, C. L., HO, D. T. & HARGREAVES, B. R. 2012. Dissolved O<sub>2</sub>/Ar and other methods reveal rapid changes in productivity during a Lagrangian experiment in the Southern Ocean. *Journal of Geophysical Research-Oceans*, 117.
- HAMME, R. C. & EMERSON, S. R. 2006. Constraining bubble dynamics and mixing with dissolved gases: Implications for productivity measurements by oxygen mass balance. *Journal of Marine Research*, 64, 73-95.
- HAMME, R. C. & SEVERINGHAUS, J. P. 2007. Trace gas disequilibria during deep-water formation. *Deep-Sea Research Part I-Oceanographic Research Papers*, 54, 939-950.
- HARDIN, G. 1960. The Competitive Exclusion Principle. *Science*, 131, 1292.
- HASKELL II, W. Z., PROKOPENKO, M. G., STANLEY, R. H. R. & KNAPP, A. N. 2016. Estimates of vertical turbulent mixing used to determine a vertical gradient in net and gross oxygen production in the oligotrophic South Pacific Gyre. *Geophysical Research Letters*, 43, 7590-7599.
- HASKELL, W. Z. & FLEMING, J. C. 2018. Concurrent estimates of carbon export reveal physical biases in  $\Delta\text{O}_2/\text{Ar}$ -based net community production estimates in the Southern California Bight. *Journal of Marine Systems*, 183, 23-31.
- HATOSY, S. M., MARTINY, J. B. H., SACHDEVA, R., STEELE, J., FUHRMAN, J. A. & MARTINY, A. C. 2013. Beta diversity of marine bacteria depends on temporal scale. *Ecology*, 94, 1898-1904.

- HAWKINS, B. A., FIELD, R., CORNELL, H. V., CURRIE, D. J., GUÉGAN, J.-F., KAUFMAN, D. M., KERR, J. T., MITTELBACH, G. G., OBERDORFF, T., O'BRIEN, E. M., PORTER, E. E. & TURNER, J. R. G. 2003. Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, 84, 3105-3117.
- HODUR, R. M. 1997. The Naval Research Laboratory's coupled ocean/atmosphere mesoscale prediction system (COAMPS). *Monthly Weather Review*, 125, 1414-1430.
- HOLTE, J., TALLEY, L. D., GILSON, J. & ROEMMICH, D. 2017. An Argo mixed layer climatology and database. *Geophysical Research Letters*, 44, 5618-5626.
- HOOVER, D. U., CHAPIN, F. S., EWEL, J. J., HECTOR, A., INCHAUSTI, P., LAVOREL, S., LAWTON, J. H., LODGE, D. M., LOREAU, M., NAEEM, S., SCHMID, B., SETALA, H., SYMSTAD, A. J., VANDERMEER, J. & WARDLE, D. A. 2005. Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecological Monographs*, 75, 3-35.
- HOSODA, S., OHIRA, T., SATO, K. & SUGA, T. 2010. Improved description of global mixed-layer depth using Argo profiling floats. *Journal of Oceanography*, 66, 773-787.
- HOWARTH, R. W. 1988. Nutrient limitation of net primary production in marine ecosystems. *Annual Review of Ecology and Systematics*, 19, 89-110.
- HUTCHINS, D. A., DITULLIO, G. R., ZHANG, Y. & BRULAND, K. W. 1998. An iron limitation mosaic in the California upwelling regime. *Limnology and Oceanography*, 43, 1037-1054.
- IRIGOIEN, X., HUISMAN, J. & HARRIS, R. P. 2004. Global biodiversity patterns of marine phytoplankton and zooplankton. *Nature*, 429, 863-867.
- IZETT, R. W., MANNING, C. C., HAMME, R. C. & TORTELL, P. D. 2018. Refined Estimates of Net Community Production in the Subarctic Northeast Pacific Derived From O-2/Ar Measurements With N2O-Based Corrections for Vertical Mixing. *Global Biogeochemical Cycles*, 32, 326-350.
- JOHNSON, Z. I., ZINSER, E. R., COE, A., MCNULTY, N. P., WOODWARD, E. M. S. & CHISHOLM, S. W. 2006. Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science*, 311, 1737-1740.

- JONES, H. L. J., LEADBEATER, B. S. C. & GREEN, J. C. 1993. Mixotrophy in marine species of chrysochromulina (prymnesiophyceae) – ingestion and digestion of a small green flagellate. *Journal of the Marine Biological Association of the United Kingdom*, 73, 283-296.
- JONSSON, B. F., DONEY, S. C., DUNNE, J. & BENDER, M. 2013. Evaluation of the Southern Ocean O<sub>2</sub>/Ar-based NCP estimates in a model framework. *Journal of Geophysical Research: Biogeosciences*, 118, 385-399.
- JOYCE, T. M., DESER, C. & SPALL, M. A. 2000. The relation between decadal variability of subtropical mode water and the North Atlantic Oscillation. *Journal of Climate*, 13, 2550-2569.
- JURANEK, L. W., HAMME, R. C., KAISER, J., WANNINKHOF, R. & QUAY, P. D. 2010. Evidence of O<sub>2</sub> consumption in underway seawater lines: Implications for air-sea O<sub>2</sub> and CO<sub>2</sub> fluxes. *Geophysical Research Letters*, 37.
- JURANEK, L. W. & QUAY, P. D. 2013. Using Triple Isotopes of Dissolved Oxygen to Evaluate Global Marine Productivity. *Annual Review of Marine Science*, Vol 5, 5, 503-524.
- KARL, D., LETELIER, R., TUPAS, L., DORE, J., CHRISTIAN, J. & HEBEL, D. 1997. The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature*, 388, 533.
- KARL, D. M. 1999. A Sea of Change: Biogeochemical Variability in the North Pacific Subtropical Gyre. *Ecosystems*, 2, 181-214.
- KARL, D. M. & CHURCH, M. J. 2014. Microbial oceanography and the Hawaii Ocean Time-series programme. *Nature Reviews Microbiology*, 12, 699-713.
- KARL, D. M. & CHURCH, M. J. 2017. Ecosystem Structure and Dynamics in the North Pacific Subtropical Gyre: New Views of an Old Ocean. *Ecosystems*, 20, 433-457.
- KELLY, T., STUKEL, M. R. & KRANZ, S. A. in prep. A Reusable Pipeline for NCP Modeling from Oxygen Argon Measurements.
- KEMBEL, S. W., WU, M., EISEN, J. A. & GREEN, J. L. 2012. Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. *Plos Computational Biology*, 8.

- KING, A. L. & BARBEAU, K. 2007. Evidence for phytoplankton iron limitation in the southern California Current System. *Marine Ecology Progress Series*, 342, 91-103.
- KIRCHMAN, D. L. 2012. *Processes in Microbial Ecology*, New York, Oxford University Press.
- KOZICH, J. J., WESTCOTT, S. L., BAXTER, N. T., HIGHLANDER, S. K. & SCHLOSS, P. D. 2013. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology*, 79, 5112-5120.
- KRAUSE, J. W., LOMAS, M. W. & NELSON, D. M. 2009. Biogenic silica at the Bermuda Atlantic Time-series Study site in the Sargasso Sea: Temporal changes and their inferred controls based on a 15-year record. *Global Biogeochemical Cycles*, 23, n/a-n/a.
- LAGE, O. M. & BONDOSO, J. 2014. Planctomycetes and macroalgae, a striking association. *Frontiers in Microbiology*, 5.
- LAGERSTRÖM, M. E., FIELD, M. P., SÉGURET, M., FISCHER, L., HANN, S. & SHERRELL, R. M. 2013. Automated on-line flow-injection ICP-MS determination of trace metals (Mn, Fe, Co, Ni, Cu and Zn) in open ocean seawater: Application to the GEOTRACES program. *Marine Chemistry*, 155, 71-80.
- LANDRY, M. R., OHMAN, M. D., GOERICKE, R., STUKEL, M. R., BARBEAU, K. A., BUNDY, R. & KAHRU, M. 2012. Pelagic community responses to a deep-water front in the California Current Ecosystem: overview of the A-Front Study. *Journal of Plankton Research*, 34, 739-748.
- LANDRY, M. R., OHMAN, M. D., GOERICKE, R., STUKEL, M. R. & TSYRKLEVICH, K. 2009. Lagrangian studies of phytoplankton growth and grazing relationships in a coastal upwelling ecosystem off Southern California. *Progress in Oceanography*, 83, 208-216.
- LANGFELDER, P. & HORVATH, S. 2007. Eigengene networks for studying the relationships between co-expression modules. *Bmc Systems Biology*, 1.
- LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559.

- LAWS, E. A. 1991. Photosynthetic quotients, new production and net community production in the open ocean. *Deep-Sea Research Part a-Oceanographic Research Papers*, 38, 143-167.
- LETSCHER, R. T. & MOORE, J. K. 2017. Modest net autotrophy in the oligotrophic ocean. *Global Biogeochemical Cycles*, 31, 699-708.
- LEVY, J. L., STAUBER, J. L. & JOLLEY, D. F. 2007. Sensitivity of marine microalgae to copper: The effect of biotic factors on copper adsorption and toxicity. *Science of The Total Environment*, 387, 141-154.
- LI, W. K. W. 2002. Macroecological patterns of phytoplankton in the northwestern North Atlantic Ocean. *Nature*, 419, 154-157.
- LI, Z. C. & CASSAR, N. 2017. A mechanistic model of an upper bound on oceanic carbon export as a function of mixed layer depth and temperature. *Biogeosciences*, 14, 5015-5027.
- LIMA-MENDEZ, G., FAUST, K., HENRY, N., DECELLE, J., COLIN, S., CARCILLO, F., CHAFFRON, S., IGNACIO-ESPINOSA, J. C., ROUX, S., VINCENT, F., BITTNER, L., DARZI, Y., WANG, J., AUDIC, S., BERLINE, L., BONTEMPI, G., CABELLO, A. M., COPPOLA, L., CORNEJO-CASTILLO, F. M., D'OVIDIO, F., DE MEESTER, L., FERRERA, I., GARET-DELMAS, M. J., GUIDI, L., LARA, E., PESANT, S., ROYO-LLONCH, M., SALAZAR, G., SANCHEZ, P., SEBASTIAN, M., SOUFFREAU, C., DIMIER, C., PICHERAL, M., SEARSON, S., KANDELS-LEWIS, S., GORSKY, G., NOT, F., OGATA, H., SPEICH, S., STEMMANN, L., WEISSENBACH, J., WINCKER, P., ACINAS, S. G., SUNAGAWA, S., BORK, P., SULLIVAN, M. B., KARSENTI, E., BOWLER, C., DE VARGAS, C., RAES, J., BOSS, E., DE VARGAS, C., FOLLOWS, M., GRIMSLEY, N., HINGAMP, P., IUDICONE, D., JAILLON, O., KARP-BOSS, L., KRZIC, U., REYNAUD, E. G., SARDET, C., SIERACKI, M., VELAYOUDON, D. & TARA OCEANS, C. 2015. Determinants of community structure in the global plankton interactome. *Science*, 348, 9.
- LIN, Y., GIFFORD, S., DUCKLOW, H., SCHOFIELD, O. & CASSAR, N. 2019. Towards Quantitative Microbiome Community Profiling Using Internal Standards. *Applied and Environmental Microbiology*, 85, e02634-18.

- LIN, Y. J., CASSAR, N., MARCHETTI, A., MORENO, C., DUCKLOW, H. & LI, Z. C. 2017. Specific eukaryotic plankton are good predictors of net community production in the Western Antarctic Peninsula. *Scientific Reports*, 7.
- LIPSCHULTZ, F., BATES, N. R., CARLSON, C. A. & HANSELL, D. A. 2002. New production in the Sargasso Sea: History and current status. *Global Biogeochemical Cycles*, 16, 1-1-1-17.
- LOMAS, M. W., BATES, N. R., JOHNSON, R. J., KNAP, A. H., STEINBERG, D. K. & CARLSON, C. A. 2013. Two decades and counting: 24-years of sustained open ocean biogeochemical measurements in the Sargasso Sea. *Deep-Sea Research Part II-Topical Studies in Oceanography*, 93, 16-32.
- LOMAS, M. W., GLIBERT, P. M., SHIAH, F. K. & SMITH, E. M. 2002. Microbial processes and temperature in Chesapeake Bay: current relationships and potential impacts of regional warming. *Global Change Biology*, 8, 51-70.
- LOMAS, M. W., STEINBERG, D. K., DICKEY, T., CARLSON, C. A., NELSON, N. B., CONDON, R. H. & BATES, N. R. 2010. Increased ocean carbon export in the Sargasso Sea linked to climate variability is countered by its enhanced mesopelagic attenuation. *Biogeosciences*, 7, 57-70.
- LONG, M., HOLLAND, A., PLANQUETTE, H., GONZÁLEZ, A., SANTANA, D., WHITBY, H., SOUDANT, P., SARTHOU, G. H., H. & JOLLEY, D. Submitted. A multi-traits approach reveals the effects of Cu on the physiology of an allelochemical-producing strain of *Alexandrium minutum*.
- LUO, Y. W., DONEY, S. C., ANDERSON, L. A., BENAVIDES, M., BERMAN-FRANK, I., BODE, A., BONNET, S., BOSTROM, K. H., BOTTJER, D., CAPONE, D. G., CARPENTER, E. J., CHEN, Y. L., CHURCH, M. J., DORE, J. E., FALCON, L. I., FERNANDEZ, A., FOSTER, R. A., FURUYA, K., GOMEZ, F., GUNDERSEN, K., HYNES, A. M., KARL, D. M., KITAJIMA, S., LANGLOIS, R. J., LAROCHE, J., LETELIER, R. M., MARANON, E., MCGILLICUDDY, D. J., MOISANDER, P. H., MOORE, C. M., MOURINO-CARBALLIDO, B., MULHOLLAND, M. R., NEEDOBA, J. A., ORCUTT, K. M., POULTON, A. J., RAHAV, E., RAIMBAULT, P., REES, A. P., RIEMANN, L., SHIOZAKI, T., SUBRAMANIAM, A., TYRRELL, T., TURK-KUBO, K. A., VARELA, M., VILLAREAL, T. A., WEBB, E. A., WHITE, A. E., WU, J. & ZEHR, J. P. 2012. Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. *Earth System Science Data*, 4, 47-73.

- LYNN, R. J. & SIMPSON, J. J. 1987. The California Current system – the seasonal variability of its physical characteristics. *Journal of Geophysical Research-Oceans*, 92, 12947-&.
- MALDONADO, M. T., HUGHES, M. P., RUE, E. L. & WELLS, M. L. 2002. The effect of Fe and Cu on growth and domoic acid production by *Pseudo-nitzschia multiseries* and *Pseudo-nitzschia australis*. *Limnology and Oceanography*, 47, 515-526.
- MALMSTROM, R. R., KIENE, R. P., COTTRELL, M. T. & KIRCHMAN, D. L. 2004. Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic ocean. *Applied and Environmental Microbiology*, 70, 4129-4135.
- MALOD-DOGNIN, N. & PRŽULJ, N. 2015. L-GRAAL: Lagrangian graphlet-based network aligner. *Bioinformatics*, 31, 2182-2189.
- MANDAKOVIC, D., ROJAS, C., MALDONADO, J., LATORRE, M., TRAVISANY, D., DELAGE, E., BIHOUEE, A., JEAN, G., DIAZ, F. P., FERNANDEZ-GOMEZ, B., CABRERA, P., GAETE, A., LATORRE, C., GUTIERREZ, R. A., MAASS, A., CAMBIAZO, V., NAVARRETE, S. A., EVEILLARD, D. & GONZALEZ, M. 2018. Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. *Scientific Reports*, 8.
- MANN, E. L., AHLGREN, N., MOFFETT, J. W. & CHISHOLM, S. W. 2002. Copper toxicity and cyanobacteria ecology in the Sargasso Sea. *Limnology and Oceanography*, 47, 976-988.
- MARTIN, P., LOEFF, M. R., CASSAR, N., VANDROMME, P., D'OVIDIO, F., STEMMANN, L., RENGARAJAN, R., SOARES, M., GONZÁLEZ, H. E., EBERSBACH, F., LAMPITT, R. S., SANDERS, R., BARNETT, B. A., SMETACEK, V. & NAQVI, S. W. A. 2013. Iron fertilization enhanced net community production but not downward particle flux during the Southern Ocean iron fertilization experiment LOHAFEX. *Global Biogeochemical Cycles*, 27, 871-881.
- MASELLA, A. P., BARTRAM, A. K., TRUSZKOWSKI, J. M., BROWN, D. G. & NEUFELD, J. D. 2012. PANDaseq: PAired-eND Assembler for Illumina sequences. *Bmc Bioinformatics*, 13.

- MASSANA, R., MURRAY, A. E., PRESTON, C. M. & DELONG, E. F. 1997. Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Applied and Environmental Microbiology*, 63, 50-56.
- MCGILLICUDDY, D. J., ROBINSON, A. R., SIEGEL, D. A., JANNASCH, H. W., JOHNSON, R., DICKEYS, T., MCNEIL, J., MICHAELS, A. F. & KNAP, A. H. 1998. Influence of mesoscale eddies on new production in the Sargasso Sea. *Nature*, 394, 263-266.
- MCKIE-KRISBERG, Z. M. & SANDERS, R. W. 2014. Phagotrophy by the picoeukaryotic green alga *Micromonas*: implications for Arctic Oceans. *The ISME Journal*, 8, 1953-1961.
- MCMURDIE, P. J. & HOLMES, S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *Plos One*, 8.
- MELLETT, T., BROWN, M. T., CHAPPELL, P. D., DUCKHAM, C., FITZSIMMONS, J. N., TILL, C. P., SHERRELL, R. M., MALDONADO, M. T. & BUCK, K. N. 2018. The biogeochemical cycling of iron, copper, nickel, cadmium, manganese, cobalt, lead, and scandium in a California Current experimental study. *Limnology and Oceanography*, 63, S425-S447.
- MILLER, A. J., SONG, H. & SUBRAMANIAN, A. C. 2015. The physical oceanographic environment during the CCE-LTER Years: Changes in climate and concepts. *Deep-Sea Research Part II-Topical Studies in Oceanography*, 112, 6-17.
- MITTELBACH, G. G., STEINER, C. F., SCHEINER, S. M., GROSS, K. L., REYNOLDS, H. L., WAIDE, R. B., WILLIG, M. R., DODSON, S. I. & GOUGH, L. 2001. What Is the Observed Relationship between Species Richness and Productivity? *Ecology*, 82, 2381-2396.
- MOELLER, P. D. R., BEAUCHESNE, K. R., HUNCIK, K. M., DAVIS, W. C., CHRISTOPHER, S. J., RIGGS-GELASCO, P. & GELASCO, A. K. 2007. Metal Complexes and Free Radical Toxins Produced by *Pfiesteria piscicida*. *Environmental Science & Technology*, 41, 1166-1172.
- MOISANDER, P. H., BEINART, R. A., VOSS, M. & ZEHR, J. P. 2008. Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon. *Isme Journal*, 2, 954-967.

- MOORE, A. M., ARANGO, H. G., BROQUET, G., POWELL, B. S., WEAVER, A. T. & ZAVALA-GARAY, J. 2011. The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems Part I - System overview and formulation. *Progress in Oceanography*, 91, 34-49.
- MORAN, M. A., SATINSKY, B., GIFFORD, S. M., LUO, H., RIVERS, A., CHAN, L.-K., MENG, J., DURHAM, B. P., SHEN, C., VARALJAY, V. A., SMITH, C. B., YAGER, P. L. & HOPKINSON, B. M. 2013. Sizing up metatranscriptomics. *Isme Journal*, 7, 237-243.
- MORRIS, R. M., RAPPE, M. S., CONNON, S. A., VERGIN, K. L., SIEBOLD, W. A., CARLSON, C. A. & GIOVANNONI, S. J. 2002. SAR11 clade dominates ocean surface bacterioplankton communities. *Nature*, 420, 806-810.
- MORROW, R. M., OHMAN, M. D., GOERICKE, R., KELLY, T. B., STEPHENS, B. M. & STUKEL, M. R. 2018. CCE V: Primary production, mesozooplankton grazing, and the biological pump in the California Current Ecosystem: Variability and response to El Nino. *Deep-Sea Research Part I-Oceanographic Research Papers*, 140, 52-62.
- MOURINO-CARBALLIDO, B. & MCGILLICUDDY, D. J. 2006. Mesoscale variability in the metabolic balance of the Sargasso Sea. *Limnology and Oceanography*, 51, 2675-2689.
- MOUSING, E., RICHARDSON, K. & ELLEGAARD, M. 2018. Global patterns in phytoplankton biomass and community size structure in relation to macronutrients in the open ocean. *Limnology and Oceanography*, 0.
- MOUW, C. B. & YODER, J. A. 2005. Primary production calculations in the Mid-Atlantic Bight, including effects of phytoplankton community size structure. *Limnology and Oceanography*, 50, 1232-1243.
- MULHOLLAND, M. R., BERNHARDT, P. W., BLANCO-GARCIA, J. L., MANNINO, A., HYDE, K., MONDRAGON, E., TURK, K., MOISANDER, P. H. & ZEHR, J. P. 2012. Rates of dinitrogen fixation and the abundance of diazotrophs in North American coastal waters between Cape Hatteras and Georges Bank. *Limnology and Oceanography*, 57, 1067-1083.

- MULHOLLAND, M. R., BRONK, D. A. & CAPONE, D. G. 2004. Dinitrogen fixation and release of ammonium and dissolved organic nitrogen by *Trichodesmium* IMS101. *Aquatic Microbial Ecology*, 37, 85-94.
- MULHOLLAND, M. R., GOBLER, C. J. & LEE, C. 2002. Peptide hydrolysis, amino acid oxidation, and nitrogen uptake in communities seasonally dominated by *Aureococcus anophagefferens*. *Limnology and Oceanography*, 47, 1094-1108.
- MUNRO, D. R., QUAY, P. D., JURANEK, L. W. & GOERICKE, R. 2013. Biological production rates off the Southern California coast estimated from triple O<sub>2</sub> isotopes and O<sub>2</sub> : Ar gas ratios. *Limnology and Oceanography*, 58, 1312-1328.
- NEEDHAM, D. M., FICHOT, E. B., WANG, E., BERDJEB, L., CRAM, J. A., FICHOT, C. G. & FUHRMAN, J. A. 2018. Dynamics and interactions of highly resolved marine plankton via automated high-frequency sampling. *The ISME Journal*, 12, 2417-2432.
- NICKELS, C. F. & OHMAN, M. D. 2018. CCEIII: Persistent functional relationships between copepod egg production rates and food concentration through anomalously warm conditions in the California Current Ecosystem. *Deep-Sea Research Part I-Oceanographic Research Papers*, 140, 26-35.
- NIILER, P. P., SYBRANDY, A. S., BI, K. N., POULAIN, P. M. & BITTERMAN, D. 1995. Measurements of the water-following capability of holey-sock and TRISTAR drifters. *Deep-Sea Research Part I-Oceanographic Research Papers*, 42, 1951-&.
- NOT, F., GAUSLING, R., AZAM, F., HEIDELBERG, J. F. & WORDEN, A. Z. 2007. Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environmental Microbiology*, 9, 1233-1252.
- O'REILLY, J. & BUSCH, D. 1984. Phytoplankton primary production on the northwestern Atlantic shelf. *Rapp. PV Reun. Cons. Int. Explor. Mer*, 183, 255-268.
- OHMAN, M. D., BARBEAU, K., FRANKS, P. J. S., GOERICKE, R., LANDRY, M. R. & MILLER, A. J. 2013. Ecological Transitions in a Coastal Upwelling Ecosystem. *Oceanography*, 26, 210-219.
- OHMAN, M. D., POWELL, J. R., PICHERAL, M. & JENSEN, D. W. 2012. Mesozooplankton and particulate matter responses to a deep-water frontal

- system in the southern California Current System. *Journal of Plankton Research*, 34, 815-827.
- OKANSEN, J., BLANCHET, F. G., FRIENDLY, M., KINDT, R., LEGENDRE, P., MCGLINN, D., MINCHIN, P. R., O'HARA, R. B., SIMPSON, G. L., SOLYMOS, P., STEVENS, M. H. H., SZOECES, E. & WAGNER, H. 2017. vegan: Community Ecology Package.: R package version 2.4-4.
- OLLI, K., PAERL, H. W. & KLAIS, R. 2015. Diversity of coastal phytoplankton assemblages Cross - ecosystem comparison. *Estuarine Coastal and Shelf Science*, 162, 110-118.
- OLSON, R. J. & SOSIK, H. M. 2007. A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot. *Limnology and Oceanography-Methods*, 5, 195-203.
- ORCUTT, K. M., LIPSCHULTZ, F., GUNDERSEN, K., ARIMOTO, R., MICHAELS, A. F., KNAP, A. H. & GALLON, J. R. 2001. A seasonal study of the significance of N-2 fixation by *Trichodesmium* spp. at the Bermuda Atlantic Time-series Study (BATS) site. *Deep-Sea Research Part II-Topical Studies in Oceanography*, 48, 1583-1608.
- ORIAN, K. J. & BRULAND, K. W. 1986. The biogeochemistry of aluminum in the Pacific Ocean. *Earth and Planetary Science Letters*, 78, 397-410.
- PÉREZ-ALMEIDA, N., GONZÁLEZ-DÁVILA, M., SANTANA-CASIANO, J. M., GONZÁLEZ, A. G. & SUÁREZ DE TANGIL, M. 2013. Oxidation of Cu(I) in Seawater at Low Oxygen Concentrations. *Environmental Science & Technology*, 47, 1239-1247.
- PIZZETTI, I., FUCHS, B. M., GERDTS, G., WICHELS, A., WILTSHIRE, K. H. & AMANN, R. 2011. Temporal Variability of Coastal Planctomycetes Clades at Kabeltonne Station, North Sea. *Applied and Environmental Microbiology*, 77, 5009-5017.
- PLANQUETTE, H. & SHERRELL, R. M. 2012. Sampling for particulate trace element determination using water sampling bottles: methodology and comparison to in situ pumps. *Limnology and Oceanography: Methods*, 10, 367-388.

- POLOVINA, J. J., HOWELL, E. A. & ABECASSIS, M. 2008. Ocean's least productive waters are expanding. *Geophysical Research Letters*, 35.
- POPELS, L. C., CARY, S. C., HUTCHINS, D. A., FORBES, R., PUSTIZZI, F., GOBLER, C. J. & COYNE, K. J. 2003. The use of quantitative polymerase chain reaction for the detection and enumeration of the harmful alga *Aureococcus anophagefferens* in environmental samples along the United States East Coast. *Limnology and Oceanography-Methods*, 1, 92-102.
- PROKOPOWICH, C. D., GREGORY, T. R. & CREASE, T. J. 2003. The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, 46, 48-50.
- PRUESSE, E., QUAST, C., KNITTEL, K., FUCHS, B. M., LUDWIG, W. G., PEPLIES, J. & GLOCKNER, F. O. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 35, 7188-7196.
- R CORE TEAM 2017. R: A language and environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing.
- REUER, M. K., BARNETT, B. A., BENDER, M. L., FALKOWSKI, P. G. & HENDRICKS, M. B. 2007. New estimates of Southern Ocean biological production rates from O<sub>2</sub>/Ar ratios and the triple isotope composition of O<sub>2</sub>. *Deep-Sea Research Part I-Oceanographic Research Papers*, 54, 951-974.
- RICHARDSON, K. 1997. Harmful or Exceptional Phytoplankton Blooms in the Marine Ecosystem. *Advances in Marine Biology*, 31, 301-385.
- RICHARDSON, T. L. & JACKSON, G. A. 2007. Small phytoplankton and carbon export from the surface ocean. *Science*, 315, 838-840.
- ROHART, F., GAUTIER, B., SINGH, A. & LE CAO, K.-A. 2017. mixOmics: an R package for 'omics feature selection and multiple data integration. *bioRxiv*.
- ROMARI, K. & VAULOT, D. 2004. Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnology and Oceanography*, 49, 784-798.
- ROWE, J. M., DEBRUYN, J. M., POORVIN, L., LECLEIR, G. R., JOHNSON, Z. I., ZINSER, E. R. & WILHELM, S. W. 2012. Viral and bacterial abundance and

production in the Western Pacific Ocean and the relation to other oceanic realms. *Fems Microbiology Ecology*, 79, 359-370.

- RUIZ, G. M., CARLTON, J. T., GROSHOLZ, E. D. & HINES, A. H. 1997. Global Invasions of Marine and Estuarine Habitats by Non-Indigenous Species: Mechanisms, Extent, and Consequences<sup>1</sup>. *American Zoologist*, 37, 621-632.
- RUTTEN, T. P. A., SANDEE, B. & HOFMAN, A. R. T. 2005. Phytoplankton monitoring by high performance flow cytometry: A successful approach? *Cytometry Part A*, 64A, 16-26.
- RYKACZEWSKI, R. R. & CHECKLEY, D. M. 2008. Influence of ocean winds on the pelagic ecosystem in upwelling regions. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 1965-1970.
- SARMIENTO, J. & GRUBER, N. 2006. *Ocean Biogeochemical Dynamics*, Princeton, New Jersey, Princeton University Press.
- SATINSKY, B. M., FORTUNATO, C. S., DOHERTY, M., SMITH, C. B., SHARMA, S., WARD, N. D., KRUSCHE, A. V., YAGER, P. L., RICHEY, J. E., MORAN, M. A. & CRUMP, B. C. 2015. Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011. *Microbiome*, 3.
- SATINSKY, B. M., GIFFORD, S. M., CRUMP, B. C. & MORAN, M. A. 2013. Use of Internal Standards for Quantitative Metatranscriptome and Metagenome Analysis. *Microbial Metagenomics, Metatranscriptomics, and Metaproteomics*, 531, 237-250.
- SCHMIEDER, R., LIM, Y. W., ROHWER, F. & EDWARDS, R. 2010. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *Bmc Bioinformatics*, 11.
- SHIOZAKI, T., FUJIWARA, A., IJICHI, M., HARADA, N., NISHINO, S., NISHI, S., NAGATA, T. & HAMASAKI, K. 2018. Diazotroph community structure and the role of nitrogen fixation in the nitrogen cycle in the Chukchi Sea (western Arctic Ocean). *Limnology and Oceanography*, 0.
- SMETS, W., LEFF, J. W., BRADFORD, M. A., MCCULLEY, R. L., LEBEER, S. & FIERER, N. 2016. A method for simultaneous measurement of soil bacterial abundances

- and community composition via 16S rRNA gene sequencing. *Soil Biology & Biochemistry*, 96, 145-151.
- SMITH, V. H. 2007. Microbial diversity-productivity relationships in aquatic ecosystems. *Fems Microbiology Ecology*, 62, 181-186.
- SOHM, J. A., WEBB, E. A. & CAPONE, D. G. 2011. Emerging patterns of marine nitrogen fixation. *Nature Reviews Microbiology*, 9, 499-508.
- SONG, H., MILLER, A. J., MCCLATCHIE, S., WEBER, E. D., NIETO, K. M. & CHECKLEY, D. M. 2012. Application of a data-assimilation model to variability of Pacific sardine spawning and survivor habitats with ENSO in the California Current System. *Journal of Geophysical Research-Oceans*, 117.
- SPITZER, W. S. & JENKINS, W. J. 1989. Rates of vertical mixing, gas-exchange and new production – estimates from seasonal gas cycles in the upper ocean near Bermuda. *Journal of Marine Research*, 47, 169-196.
- STEINBERG, D. K., CARLSON, C. A., BATES, N. R., JOHNSON, R. J., MICHAELS, A. F. & KNAP, A. H. 2001. Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry. *Deep-Sea Research Part Ii-Topical Studies in Oceanography*, 48, 1405-1447.
- STOECK, T., BASS, D., NEBEL, M., CHRISTEN, R., JONES, M. D. M., BREINER, H. W. & RICHARDS, T. A. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, 19, 21-31.
- STRONG, D. R. 2010. Evidence and inference: shapes of species richness-productivity curves1. *Ecology*, 91, 2534-2535.
- STRUB, P. T. & JAMES, C. 2000. Altimeter-derived variability of surface velocities in the California Current System: 2. Seasonal circulation and eddy statistics. *Deep-Sea Research Part Ii-Topical Studies in Oceanography*, 47, 831-870.
- STUKEL, M. R., LANDRY, M. R., BENITEZ-NELSON, C. R. & GOERICKE, R. 2011. Trophic cycling and carbon export relationships in the California Current Ecosystem. *Limnology and Oceanography*, 56, 1866-1878.

- STUKEL, M. R., OHMAN, M. D., BENITEZ-NELSON, C. R. & LANDRY, M. R. 2013. Contributions of mesozooplankton to vertical carbon export in a coastal upwelling system. *Marine Ecology Progress Series*, 491, 47-+.
- SUNAGAWA, S., COELHO, L. P., CHAFFRON, S., KULTIMA, J. R., LABADIE, K., SALAZAR, G., DJAHANSCHIRI, B., ZELLER, G., MENDE, D. R., ALBERTI, A., CORNEJO-CASTILLO, F. M., COSTEA, P. I., CRUAUD, C., D'OVIDIO, F., ENGELEN, S., FERRERA, I., GASOL, J. M., GUIDI, L., HILDEBRAND, F., KOKOSZKA, F., LEPOIVRE, C., LIMA-MENDEZ, G., POULAIN, J., POULOS, B. T., ROYO-LLONCH, M., SARMENTO, H., VIEIRA-SILVA, S., DIMIER, C., PICHERAL, M., SEARSON, S., KANDELS-LEWIS, S., BOWLER, C., DE VARGAS, C., GORSKY, G., GRIMSLEY, N., HINGAMP, P., IUDICONE, D., JAILLON, O., NOT, F., OGATA, H., PESANT, S., SPEICH, S., STEMMANN, L., SULLIVAN, M. B., WEISSENBACH, J., WINCKER, P., KARSENTI, E., RAES, J., ACINAS, S. G., BORK, P. & TARA OCEANS, C. 2015. Structure and function of the global ocean microbiome. *Science*, 348.
- TACKMANN, J., RODRIGUES, J. F. M. & VON MERING, C. 2018. Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. *bioRxiv*.
- TANG, W., WANG, S., FONSECA-BATISTA, D., DEHAIRS, F., GIFFORD, S., GONZALEZ, A. G., GALLINARI, M., PLANQUETTE, H., SARTHOU, G. & CASSAR, N. 2019. Revisiting the distribution of oceanic N<sub>2</sub> fixation and estimating diazotrophic contribution to marine production. *Nature Communications*, 10, 831.
- TEETER, L., HAMME, R. C., IANSON, D. & BIANUCCI, L. 2018. Accurate Estimation of Net Community Production From O<sub>2</sub>/Ar Measurements. *Global Biogeochemical Cycles*, 0.
- THINGSTAD, T. F. & LIGNELL, R. 1997. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology*, 13, 19-27.
- THOMAS, B. R., KENT, E. C. & SWAIL, V. R. 2005. Methods to homogenize wind speeds from ships and buoys. *International Journal of Climatology*, 25, 979-995.
- TKACZ, A., HORTALA, M. & POOLE, P. S. 2018. Absolute quantitation of microbiota abundance in environmental samples. *Microbiome*, 6, 110.

- TORTELL, P. D. 2005. Dissolved gas measurements in oceanic waters made by membrane inlet mass spectrometry. *Limnology and Oceanography: Methods*, 3, 24-37.
- TORTELL, P. D., ASHER, E. C., DUCKLOW, H. W., GOLDMAN, J. A. L., DACEY, J. W. H., GRZYMSKI, J. J., YOUNG, J. N., KRANZ, S. A., BERNARD, K. S. & MOREL, F. M. M. 2014. Metabolic balance of coastal Antarctic waters revealed by autonomous pCO<sub>2</sub> and Delta O-2/Ar measurements. *Geophysical Research Letters*, 41, 6803-6810.
- TREUSCH, A. H., DEMIR-HILTON, E., VERGIN, K. L., WORDEN, A. Z., CARLSON, C. A., DONATZ, M. G., BURTON, R. M. & GIOVANNONI, S. J. 2012. Phytoplankton distribution patterns in the northwestern Sargasso Sea revealed by small subunit rRNA genes from plastids. *ISME J*, 6, 481-492.
- TWINING, B. S. & BAINES, S. B. 2013. The Trace Metal Composition of Marine Phytoplankton. *Annual Review of Marine Science*, 5, 191-215.
- ULFSBO, A., CASSAR, N., KORHONEN, M., VAN HEUVEN, S., HOPPEMA, M., KATTNER, G. & ANDERSON, L. G. 2014. Late summernet community production in the central Arctic Ocean using multiple approaches. *Global Biogeochemical Cycles*, 28, 1129-1148.
- VALLINA, S. M., FOLLOWS, M. J., DUTKIEWICZ, S., MONTOYA, J. M., CERMENO, P. & LOREAU, M. 2014. Global relationship between phytoplankton diversity and productivity in the ocean. *Nature Communications*, 5.
- VERGIN, K. L., BESZTERI, B., MONIER, A., THRASH, J. C., TEMPERTON, B., TREUSCH, A. H., KILPERT, F., WORDEN, A. Z. & GIOVANNONI, S. J. 2013. High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *Isme Journal*, 7, 1322-1332.
- VĚTROVSKÝ, T. & BALDRIAN, P. 2013. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLOS ONE*, 8, e57923.

- VIŠIĆ, H. 2018. *Chemotaxonomic and Morphological Approach to Phytoplankton Analyses in Contrasting Trophic Systems of North Pacific*. Master of Environmental Sciences, University of Zagreb.
- WALLACE, D. W. R. 1994. Anthropogenic chlorofluoromethanes and seasonal mixing rates in the Middle Atlantic Bight. *Deep-Sea Research Part II-Topical Studies in Oceanography*, 41, 307-324.
- WALTERS, W. A., CAPORASO, J. G., LAUBER, C. L., BERG-LYONS, D., FIERER, N. & KNIGHT, R. 2011. PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics*, 27, 1159-1161.
- WANG, Q., GARRITY, G. M., TIEDJE, J. M. & COLE, J. R. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73, 5261-5267.
- WANG, S., LIN, Y., GIFFORD, S., EVELETH, R. & CASSAR, N. 2018. Linking patterns of net community production and marine microbial community structure in the western North Atlantic. *The ISME Journal*.
- WANG, S., TANG, W. Y., DELAGE, E., GIFFORD, S., WHITBY, H., GONZALEZ, A., EVEILLARD, D., PLANQUETTE, H. & CASSAR, N. In review. Patterns of North Atlantic microplankton community structure in relation to net community production, nitrogen fixation, and nutrient availability.
- WANNINKHOF, R. 1992. Relationship between wind-speed and gas-exchange over the ocean. *Journal of Geophysical Research-Oceans*, 97, 7373-7382.
- WANNINKHOF, R. 2014. Relationship between wind speed and gas exchange over the ocean revisited. *Limnology and Oceanography: Methods*, 12, 351-362.
- WANNINKHOF, R., BARBERO, L., BYRNE, R., CAI, W. J., HUANG, W. J., ZHANG, J. Z., BARINGER, M. & LANGDON, C. 2015. Ocean acidification along the Gulf Coast and East Coast of the USA. *Continental Shelf Research*, 98, 54-71.
- WHITE, A. E., BARONE, B., LETELIER, R. M. & KARL, D. M. 2017. Productivity diagnosed from the diel cycle of particulate carbon in the North Pacific Subtropical Gyre. *Geophysical Research Letters*, 44, 3752-3760.

- WILLIAMS ET AL, P. J. L. 1989. Group report: Export productivity from the photic zone. In: BERGER, W. H., SMETACEK, V. & WEFER, G. (eds.) *Productivity of the Ocean: Present and Past*. Hoboken, N.J.: John Wiley and Sons.
- WILLIAMS, P. J. L., QUAY, P. D., WESTBERRY, T. K. & BEHRENFELD, M. J. 2013. The Oligotrophic Ocean Is Autotrophic. *Annual Review of Marine Science*, Vol 5, 5, 535-549.
- WILLIAMS, R. & FOLLOWS, M. 2011. *Ocean Dynamics and the Carbon Cycle: Principles and Mechanisms*, Cambridge, United Kingdom, Cambridge University Press.
- WURTSBAUGH, W. A. & HORNE, A. J. 1982. Effects of Copper on Nitrogen Fixation and Growth of Blue-Green Algae in Natural Plankton Associations. *Canadian Journal of Fisheries and Aquatic Sciences*, 39, 1636-1641.
- YILMAZ, P., YARZA, P., RAPP, J. Z. & GLÖCKNER, F. O. 2016. Expanding the World of Marine Bacterial and Archaeal Clades. *Frontiers in Microbiology*, 6.
- ZEHR, J. P. & KUDELA, R. M. 2011. Nitrogen Cycle of the Open Ocean: From Genes to Ecosystems. In: CARLSON, C. A. & GIOVANNONI, S. J. (eds.) *Annual Review of Marine Science*, Vol 3.
- ZINSER, E. R., COE, A., JOHNSON, Z. I., MARTINY, A. C., FULLER, N. J., SCANLAN, D. J. & CHISHOLM, S. W. 2006. Prochlorococcus ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Applied and Environmental Microbiology*, 72, 723-732.

## Biography

Seaver Wang graduated from the University of Pennsylvania in 2014 with a BA with Distinction in Earth Sciences before commencing his graduate studies in the Division of Earth and Ocean Sciences within the Nicholas School of the Environment at Duke University under the supervision of Dr. Nicolas Cassar. Seaver has one first-author peer-reviewed publication, "Linking patterns of net community production and marine microbial community structure in the western North Atlantic," published in The ISME Journal in 2018. He has also written two first-author manuscripts, "Patterns of North Atlantic microbial community structure in relation to net community production, nitrogen fixation, and nutrient availability" and "Lagrangian studies of NCP: Assessing the effect of diel biological cycles and non-steady state systems." The former manuscript is currently in review, while the latter is in preparation for submission. Seaver has been awarded the Orrin H. Pilkey Fellowship and a STEM Chateaubriand Fellowship.