UNIVERSITY OF CALIFORNIA SAN DIEGO

Applying meta-'omics to marine microbial ecophysiology

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Marine Biology

by

Bethany Cristine Kolody

Committee in charge:

Professor Eric Allen, Chair
Professor Andrew Allen, Co-chair
Professor Farooq Azam
Professor Douglas Bartlett
Professor Peter Franks
Professor Karsten Zengler

2020

The Dissertation of Bethany Cristine Kolody is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Co-chair

_____
Chair

University of California San Diego
2020

DEDICATION

To Madeleine

EPIGRAPH

"…this play repeats itself year after year with the same regularity as every spring the trees turn green and in the autumn lose their leaves; with just such absolute certainty as the cherries bloom before the sunflowers, so *Skeletonema* arrives at their yearly peak earlier than *Ceratium*."

- Franz Schutt, 1892, *Kiel Bight*

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to everyone who made this dissertation possible. First and foremost, thank you to my advisors, Eric Allen and Andy Allen, and my committee, Doug Bartlett, Farooq Azam, Peter Franks, and Karsten Zengler. You are not only extraordinary scientists, but also giving and supportive mentors. Eric, thank you for sustaining my optimism and always encouraging me to pursue opportunities for growth. Whenever my faith in science wavered, I would be uplifted by a meeting with you and your infectious enthusiasm for discovery. Andy, thank you for giving me the space and the time to listen to my data, for wading through it with me looking for meaning, for teaching me how to think deeply about physiological problems and their environmental context, and for your encouraging mentorship through drafts and revisions and reviews. Thank you both for taking a leap of faith and allowing me to hop on a ship to Antarctica to collect my dream samples. Doug, you were the first person I met at SIO. Thank you for inspiring me to come here, for taking me on a deep-sea adventure, and for teaching me seagoing science skills that would serve me well later on. Farooq, thank you for always having an open door and listening to my ideas about marine microbial evolution, and for supporting me with equipment and training for my P18 cruise. Peter, thank you for being a great sounding board from year one, for inspiring me to think hard about the physical processes driving microbial interactions, and for making sure that all of my future proposals include a proposal statement. Karsten, thank you for being so generous with your time and wisdom, and for keeping me skeptical so that I do the best possible science.

Thank you to the Eric Allen and Andy Allen labs, and to the scientists of JCVI. In particular, I'd like to thank John McCrow being a patient instructor of bioinformatics, and for always seeming to instantly find the bugs that evaded me for hours, and Sarah Smith, for

teaching me the nuances of phytoplankton physiology that I didn't even know I didn't know. Thank you to Hong Zheng for her relentless attention to detail in processing my samples, and for caring about the results as deeply as I did, every step of the way.

I'd like to thank Sarah Purkey, Rolf Sonnerup, Lynne Talley, and the entire GO-SHIP program for going above and beyond to give me the opportunity to collect my Chapter 3 samples. Thank you for a place in the water budget, a bunk on the ship, and for making my time at sea a delight.

Thank you to my friends at SIO. Thank you to Angela Zoumplis and Drishti Kaul, for always keeping me laughing, and for improving my impressions, and Rachel Diner, for leading the way. Thank you to my Neosho housemates over the years (Beverly French, Kiefer Forsch, Sam Honch, Jen Le, Kaitlyn Lowder, Georgie Zelenak, Kim Reed Nutt, Daniel Blatter) for building a home with me.

Thank you to Paige Hasebe, Katy Blumer, April Xiong, and Tessa Carelli for making sure I keep one foot on the ground, even when the other is stuck in the Ivory tower. Thank you for reminding me of the importance of adventure and of not taking life too seriously. Even from afar, you were with me through it all, and I am so happy to have you as friends.

Thank you to Barbara and Joe Giammona for being my family away from home and always providing me with an uplifting respite from work. Thank you for feeding me for many years, for giving me a place to park my car, and, most recently, for sheltering me from a pandemic. Barbara, thank you for including me in countless culinary, historical, and artistic adventures that gave me something to look forward to when I stopped by. And Joe, thank you for keeping me on my toes with great puns (the occasional two-thirds of a pun). Most of all, thank you both for raising James Giammona.

James, from our first meeting at Caroline's Café to bonding over a shared appreciation of the evolution and life-history of strange creatures I photographed in Marine Organisms class, from tide-pooling to manuscript-reviewing, from nature walks behind JCVI to practically becoming an employee—you were there for it all. Thank you for checking my math, for listening to my frustrations, for cooking me dinner so I could write, for delighting in my fun science facts and being and endless font of your own, for picking up the phone at 2 a.m. when I was at sea and only had 10 minutes to talk, for co-parenting our rabbit, for flying to Patagonia expecting a vacation and helping me sort samples in the -80 walk-in freezer instead... the list is endless. Thank you for all of the small things and for driving a quarter of the way to the moon to take a chance on a long-distance relationship that started with an e-mail about microbes surviving on radiation in rocks. You are my partner in all things, and I'm so lucky to have you.

Finally, thank you to my family. Thank you to my parents, John and Cristine Kolody for always picking up the phone, for being endlessly supportive, and for knowing just the right time for a family vacation. Thank you to my sister, Brianna Magnusen, for setting the bar high, and my brother-in-law, Drew Magnusen, for making sure that I got to be the first Dr. Kolody.

Chapter one, in full, is a reprint of the material as it appears in International Society for Microbial Ecology (ISME) Journal, 2019. Kolody, B. C., J. P. McCrow, L. Zeigler Allen, F. O. Aylward, K. M. Fontanez, A. Moustafa, M. Moniruzzaman, F. P. Chavez, C. A. Scholin, E. E. Allen, A. Z. Worden, E. F. Delong, and A. E. Allen. 2019. "Diel transcriptional response of a California Current plankton microbiome to light, low iron, and enduring viral infection." The dissertation author was the primary investigator and author of this paper. This study was supported by the National Science Foundation (NSF-OCE-1756884, NSF-OCE-1637632), United States Department of Energy Genomics Science program (DE-SC0008593 and DE-

SC0018344), NOAA (NA15OAR4320071), and the Gordon and Betty Moore Foundation grant GBMF3828 (AEA). Cruise work was supported by the David and Lucile Packard Foundation through an annual grant to MBARI (FPC, CAS, and AZW).

Chapter two has been prepared as a submission to the ISME Journal. B. C. Kolody, S. R. Smith, L. Zeigler Allen, J. P. McCrow, D. Shi, B .M. Hopkinson, F. M. M. Morel, B.B. Ward, A. E. Allen. “Differential transcriptional response of diverse phytoplankton to experimental upwelling limited by nitrogen, iron, and viruses.” The dissertation author was the primary investigator and author of this paper.

VITA

2014 B. S. Biology, New York University Abu Dhabi
2016 M. S. Oceanography, Scripps Institution of Oceanography, UCSD
2020 Ph. D. Marine Biology, Scripps Institution of Oceanography, UCSD

PUBLICATIONS

Giron-Nava, A., C. C. James, A. F. Johnson, D. Dannecker, B. C. Kolody, A. Lee, M. Nagarkar, G. M. Pao, H. Ye, D. G. Johns, and G. Sugihara. 2017. "Quantitative Argument for Long-Term Ecological Monitoring." Marine Ecology Progress Series 572.

Kolody, B. C., J. P. McCrow, L. Zeigler Allen, F. O. Aylward, K. M. Fontanez, A. Moustafa, M. Moniruzzaman, F. P. Chavez, C. A. Scholin, E. E. Allen, A. Z. Worden, E. F. Delong, and A. E. Allen. 2019. "Diel Transcriptional Response of a California Current Plankton Microbiome to Light, Low Iron, and Enduring Viral Infection." The ISME Journal 2817–33.

ABSTRACT OF THE DISSERTATION

Applying meta-'omics to marine microbial ecophysiology

by

Bethany Cristine Kolody

Doctor of Philosophy in Marine Biology

University of California San Diego, 2020

Professor Eric Allen, Chair
Professor Andrew Allen, Co-Chair

Phytoplankton and associated microbial communities are essential for sustaining marine ecosystems. However, the structure and function of these communities is largely driven by dynamic physical forcing (e.g. upwelling, subduction) and micro-scale interactions (e.g. viral infection, trophic interactions, symbioses) that are difficult to capture. This dissertation applies recent molecular tools to these complex systems in order to resolve the physiology of key microbial players in the context of environmental forcing and community interactions.

In Chapter 1, a semi-Lagrangian drifter was deployed to capture the transcriptional dynamics of a phytoplankton community across diel cycles. Apart from fungi and archaea, all groups (dinoflagellates, ciliates, haptophytes, pelagophytes, diatoms, cyanobacteria, prasinophytes) exhibited 24-h periodicity in some transcripts. Larger portions of the

transcriptome oscillated in phototrophs. Functional groups of genes, including photosynthetic machinery, had conserved timing across diverse lineages. In addition to responding to low-iron, many taxa were also being persistently infected by viruses.

Chapter 2 applied metatranscriptomics to a simulated upwelling experiment to examine the response of blooming phytoplankton to nitrogen and iron, the most common nutrients limiting marine phytoplankton growth in nature. Regulation of metabolism and light harvesting machinery changed in a conserved manner across diverse lineages. Viral activity was widespread and increased under nutrient limitation. The relative expression of NRT2 to GSII and iron starvation induced proteins (ISIP1, ISIP2, ISIP3) to the thiamin biosynthesis gene, ThiC, were identified as robust markers of diatom cellular nitrogen and iron status.

Chapter 3 applied high-resolution amplicon sequencing to a ship-based transect of the South Pacific along a gradient of water ages spanning newly subducted Antarctic water to subtropical water with a residence time >1,000 years. 16S and 18S rRNA diversity analyses were performed using both DNA and cDNA reverse-transcribed from RNA, providing an estimate of the breadth of deep-ocean microbial diversity that can be attributed to active cells. Microbial communities differed across size classes and were ultimately structured by physical properties of water masses and residence time in the deep ocean. These results highlight the utility of ‘omics techniques for capturing the response of marine microbes to physical dynamics and resolving relationships between key community members.

INTRODUCTION

Approximately 45% of global photosynthesis occurs in marine systems[1], and carbon fixation by phytoplankton is not only an important parameter in modeling the future of Earth's climate[2], but also sets an upper bound on oceanic productivity. Heterotrophic marine microbes, in turn, are indispensable for sustaining oceanic food webs via recycling of organic matter. Despite the significance of marine microbial ecosystem services, their ecophysiology has traditionally been difficult to study *in situ*, where observations are most relevant.

The object of this PhD dissertation is to harness recent technological advances in DNA/cDNA sequencing and bioinformatics to glean novel insights into the community structure and ecophysiology of marine microbial systems in nature. Both microbial ecology and marine systems lend themselves to genetic and transcriptional observation because of their difficulty of study by traditional methods. Microbial communities are comprised of members that are problematic (in the case of many eukaryotic plankton), if not impossible (in the case of prokaryotes), to distinguish taxonomically by morphological observation. As much as 99% of microbial diversity cannot be cultured by standard techniques, and is only accessible using molecular methods[3]. Furthermore, the relevance of microbial systems to human health and to global biogeochemistry is mediated by metabolic interactions that can only be observed at a molecular level. Marine microbial systems in particular are in constant flux due to the physically dynamic nature of their habitat. Recent advances in sequencing technology and "big data" bioinformatics are only now allowing for the large scale sampling necessary to resolve the physical drivers of marine microbial systems. The application of "meta-'omics" to field based research allows us to observe directly the impacts of physical oceanography on marine microbial

systems, an important parameter which cannot be accounted for using laboratory-based approaches.

In this dissertation, meta-'omics techniques are deployed in various ways to resolve marine microbial ecophysiology temporally, experimentally, and over great distances. Chapter 1 employs a free-floating robotic sampler designed by Monterey Bay Aquarium Research Institute (MBARI), the Environmental Sample Processor (ESP), to observe the temporal dynamics of microbial, and in particular, phytoplankton, communities off of Monterey Bay, CA. We observe a sympatric population of plankton along a Lagrangian drift track every 4 hours over 3 days, documenting the influence of diel cycles and small-scale nutrient fluxes on planktonic ecophysiology. Chapter 2 uses seawater incubation experiments to simulate phytoplankton blooms and compare the physiology of key players in replete and deplete nitrogen and iron conditions. Chapter 3 applies 'omics techniques to a traditional ship-based oceanographic sampling regime in order to resolve microbial community structure on an ocean-basin scale. Three hundred samples were collected on a transect spanning the South Pacific (26°S – 70°S) in collaboration with the P18 line of the GO-SHIP ocean mapping survey. Metabarcoding samples span full ocean depth, and are complemented by high-resolution mapping of the physical, chemical, and nutrient properties of the sampled water masses. This chapter seeks to assess the impact of thermohaline circulation (THC) on the adaptive potential of marine microbes to the extreme high pressure and low temperature of the deep ocean by viewing the resulting pelagic community structure. Together, this dissertation demonstrates the utility of 'omics-based investigation for probing marine microbial physiology across a variety of platforms.

REFERENCES FOR THE INTRODUCTION

1. Falkowski, P. G. & Raven, J. A. Photosynthesis and Primary Production in Nature. (2007).
2. Cox, P. M., Betts, R. a, Jones, C. D., Spall, S. a & Totterdell, I. J. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* **408**, 184–187 (2000).
3. Riesenfeld, C. S., Schloss, P. D. & Handelsman, J. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* **38**, 525–552 (2004).

CHAPTER 1: DIEL TRANSCRIPTIONAL RESPONSE OF A CALIFORNIA CURRENT PLANKTON MICROBIOME TO LIGHT, LOW IRON, AND ENDURING VIRAL INFECTION

The ISME Journal (2019) 13:2817–2833
https://doi.org/10.1038/s41396-019-0472-2

ISME

ARTICLE

Diel transcriptional response of a California Current plankton microbiome to light, low iron, and enduring viral infection

B. C. Kolody[1,2] · J. P. McCrow[2] · L. Zeigler Allen[1,2] · F. O. Aylward[3] · K. M. Fontanez[4] · A. Moustafa[5] · M. Moniruzzaman[3,6] · F. P. Chavez[6] · C. A. Scholin[6] · E. E. Allen[1] · A. Z. Worden[6,7] · E. F. Delong[4,8] · A. E. Allen[1,2]

Received: 9 November 2018 / Revised: 11 June 2019 / Accepted: 15 June 2019 / Published online: 18 July 2019

Abstract

Phytoplankton and associated microbial communities provide organic carbon to oceanic food webs and drive ecosystem dynamics. However, capturing those dynamics is challenging. Here, an in situ, semi-Lagrangian, robotic sampler profiled pelagic microbes at 4 h intervals over ~2.6 days in North Pacific high-nutrient, low-chlorophyll waters. We report on the community structure and transcriptional dynamics of microbes in an operationally large size class (>5 µm) predominantly populated by dinoflagellates, ciliates, haptophytes, pelagophytes, diatoms, cyanobacteria (chiefly *Synechococcus)*, prasinophytes (chiefly *Ostreococcus)*, fungi, archaea, and proteobacteria. Apart from fungi and archaea, all groups exhibited 24-h periodicity in some transcripts, but larger portions of the transcriptome oscillated in phototrophs. Periodic photosynthesis-related transcripts exhibited a temporal cascade across the morning hours, conserved across diverse phototrophic lineages. Pronounced silica:nitrate drawdown, a high flavodoxin to ferredoxin transcript ratio, and elevated expression of other Fe-stress markers indicated Fe-limitation. Fe-stress markers peaked during a photoperiodically adaptive time window that could modulate phytoplankton response to seasonal Fe-limitation. Remarkably, we observed viruses that infect the majority of abundant taxa, often with total transcriptional activity synchronized with putative hosts. Taken together, these data reveal a microbial plankton community that is shaped by recycled production and tightly controlled by Fe-limitation and viral activity.

Introduction

Phytoplankton productivity is essential for supporting marine food webs and represents a key variable in biogeochemical cycling and climate models [1]. Primary productivity is often determined locally by eddy-scale upwelling and molecular-scale interactions between bacteria, viruses, grazers, and phytoplankton [2, 3]. As a result of observational limitations of these molecular-scale processes in situ, much of our knowledge of phytoplankton physiology derives from laboratory-based experiments. Such studies have been instrumental in elucidating the basic biology and genetic potential of many individual

Supplementary information The online version of this article (https://doi.org/10.1038/s41396-019-0472-2) contains supplementary material, which is available to authorized users.

✉ A. E. Allen
aallen@jcvi.org

[1] Scripps Institution of Oceanography, University of California, San Diego, CA 92093, USA

[2] Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, CA 92037, USA

[3] Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA

[4] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[5] Department of Biology and Biotechnology Graduate Program, American University in Cairo, New Cairo, Egypt

[6] Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA

[7] Ocean EcoSystems Biology Unit, GEOMAR Helmholtz Centre for Ocean Research, Kiel, DE, Germany

[8] Daniel K. Inouye Center for Microbial Oceanography: Research and Education (C-MORE), University of Hawaii, Honolulu, HI 96822, USA

phytoplankton [4–8], however, they cannot capture ecological interactions with uncultured community members or the influence of the advection dynamics found in nature.

Recently, the Environmental Sample Processor (ESP) [9], a robotic sampling device, has been deployed in a drifter configuration to follow sympatric populations of small (<5 µm) microbes within a particular water mass. These studies elucidated diel transcriptional rhythms first in the picophytoplankton, *Synechococcus* and *Ostreococcus* [10], and later in heterotrophic bacterioplankton [11] and viruses [12]. A comparison of drifts from diverse environments revealed a daily "cascade" of transcriptional activity across taxa that was conserved on ocean basin scales [13].

However, it is unknown whether these findings extend to the majority of eukaryotic primary producers (diatoms, haptophytes, pelagophytes, chlorophytes, etc.), planktonic predators such as ciliates and dinoflagellates, and particle-associated microbes. If diel transcriptional rhythms exist in these taxa, do similar functions co-occur across diverse lineages, or do they fall in a temporal cascade? In addition, it is unknown to what extent abiotic factors such as nutrient limitation affect the prevalence of diel transcription.

Here, we analyzed the large size-class filters associated with the initial deployment of the ESP [10], which collected whole-community RNA samples every 4 h over ~2.6 days in the central California Current upwelling system (cCCS). Because of the semi-Lagrangian nature of this ESP deployment (Fig. S1a), a coherent microbial community was observed in both size classes, providing a unique record of its diel response to changes in sunlight as well as in situ nutrient conditions and prevailing ecosystem dynamics. We asked whether the timing and function of diel transcription was conserved across diverse, uncultured protist lineages, as well as how productivity was affected by local iron stress. In addition, access to both filters allowed us to juxtapose free-living microbes with those putatively associated with particles in the same environment, and compare the role of viruses in both fractions.

Materials and methods

Full methods, including methods for rRNA amplicon processing and phylogenetics, are described in Supplementary File 1, and raw data can be accessed at NCBI (BioProject accession number PRJNA492502; BioSample accession numbers SAMN10104964-SAMN10105011). Processed data can be found in Supplementary Datasets 1–12, and are described in Supplementary File 3. Briefly, 1 L samples were collected every ~4 h (16 samples in total) by an ESP suspended 23 m below a semi-Lagrangian surface float as previously described [10] from September 16–19, 2010. Seawater was size fractionated in situ onto large fraction (5 µm) and small fraction (0.22 µm) filters. cDNA for metatranscriptomes was prepared as described in Ottesen et al. [10] from ribosomal RNA-depleted total RNA [10].

Metatranscriptome quality control, trimming, filtration, and rRNA removal was conducted on large fraction Illumina reads and the previously reported small fraction 454 reads via the RNAseq Annotation Pipeline v0.4 (Fig. S2) [14]. Ab initio open-reading frames (ORFs) were predicted on assembled large fraction contigs and unassembled small fraction 454 reads. ORFs were annotated via BLASTP [15] alignment to a comprehensive protein database, *phyloDB* (Supplementary File 1). To avoid biases introduced by classifying ORFs based on best BLAST to potentially contaminated sequences, particularly important when using microbial reference transcriptomes obtained from non-axenic laboratory cultures, a Lineage Probability Index (LPI) was used to assign taxonomy [14, 16, 17]. LPI was calculated here as a value between 0 and 1 indicating lineage commonality among the top 95-percentile of sequences based on BLAST bit-score [14, 16]. For each taxa group, the references with the highest mean percent identity to ab initio ORFs and that recruited the most ORFs were used for nucleotide-space mapping. References with at least 1000 genes with at least five reads mapped were then considered for downstream analysis. Coverage statistics are provided in Fig. S3. Reference ORFs were hierarchically clustered together with ab initio ORFs from both fractions to form peptide ortholog groups and assigned a consensus annotation. ORFs with significantly periodic diel expression were identified using harmonic regression analysis (HRA) as previously described [10, 11, 13]. The Weighted Gene Correlation Network Analysis (WGCNA) R package [18] was used as previously described [13] to identify modules of conserved expression among ORFs and functional clusters.

Results and discussion

The ESP was deployed offshore of Big Sur in the cCCS "transition zone" between nutrient-dense coastally-upwelled water and the oligotrophic open ocean. Despite sustaining highly productive fisheries, this region is characterized by frequent, variable levels of iron stress (Fe <0.2 nmol/kg; Fig. S4) and concordant high residual nitrate (5–15 µM) and low chlorophyll (1–2 µg/l; refs. [19–22]). We measured high nitrate (5-13 µM) and low-chlorophyll concentrations (<1 µM; Fig. S1b) in addition to a silica:nitrate ratio indicative of iron stress. Silica:nitrate ratios in the range of 0.8–1.1 are associated with iron limitation [20] and are thought to result from silica drawdown by iron-stressed diatoms [23]. In our study, this ratio was initially around 1 and dropped by an order of magnitude along the drift track

(Fig. S1b). Molecular evidence, including expression of several low-iron response genes and a strikingly high flavodoxin:ferredoxin ratio, also supported iron limitation (Supplementary File 1, Fig. S5).

Taxonomic structure of the active community

The mRNA taxonomic assignments depicted an active community that was stable over time for both size fractions. This was the case at both coarse taxonomic (Fig. 1a) and genus (Fig. S6) levels, indicating that the drift track sampled a sympatric community of plankton. Large fraction mRNA activity was dominated by photosynthetic eukaryotes for which mapping to reference transcriptomes generally averaged <80% nucleotide identity (Figs. S7 and S8). In all, 45.9% of ab initio ORFs did not have any database match, a testament to the breadth of novel plankton diversity that remains uncultured, even in coastal regimes.

Dinoflagellates were the largest identifiable contributor to large size-class activity (34.8% of library normalized reads; Fig. 1a), with *Alexandrium*, *Karenia*, and *Karlodinium* each accounting for ~20% of dinoflagellate mRNA (Fig. S6e). Phylogenetic analysis of 18S rRNA amplicons (Fig. S9) also depicted a community dominated by dinoflagellates, and 16S rRNA amplicons from chloroplasts (Fig. 1b) confirmed that many were photosynthetic. While copy number variation is a potential source of bias for 18S amplicon data [24], here our 18S community structure is largely in agreement with our transcript-based annotations. Non-plastid 16S rRNA amplicons from the large size class were dominated by cyanobacteria, *Bacteriodetes*, and Proteobacteria (Fig. S10).

Other major large fraction taxa included centric diatoms (10.3% of mRNA), ciliates (9.0%), metazoans (6.2%), haptophytes (5.2%), green algae (5.0%, primarily prasinophytes), *Synechococcus* (4.6%), and pelagophytes (4.5%). Diatoms and pelagophytes were overwhelmingly dominated by *Chaetoceros* (Fig. S6d) and *Pelagomonas* (Fig. S6g), respectively, while *Ostreococcus* and *Phaeocystis* dominated green algae and haptophytes to a lesser extent (Fig. S6a, f). Several of these taxa were previously shown to have high cell abundances in non-fractionated samples using quantitative methods during this drift and in other regional studies [25–28].

In order to address the entire community, we also used the size-fractionated data to distinguish putatively particle-associated (large fraction) from free-living (small fraction) taxa [29–31]. Fungi were significantly enriched in the large fraction (EdgeR FDR <0.05; $\log_2 FC = -5.2$) and highly expressed a cellulose degrading glycoside hydrolase family 7 enzyme (Fig. 1a, Supplementary Data 6). Recently, fungi were shown to be among the most important eukaryotes on bathypelagic marine snow [32]. Hence, our results illustrate that their importance in particle ecology likely extends into the surface ocean. Likewise, many prokaryotes significantly enriched in the large fraction (EdgeR FDR < 0.05) have been commonly associated with particles, such as the *Cytophaga* (2.3% of bacterial expression) and *Planctomyces* (0.5%) (Supplementary Dataset 10). These taxa are important recyclers of structural and storage forms of carbon [33, 34], and we found active expression of bacterial organic matter degrading enzymes, such as glycoside hydrolase family 16 and secreted glycosyl hydrolases (Supplementary Dataset 6). Furthermore, the *comEA* and *comEC* gene clusters for the bacterial process of taking up exogenous DNA (competence) were enriched in the large fraction (EdgeR FDR < 0.05; Supplementary Dataset 8) and expressed across 11 large fraction bacterial genera. Particle-attached bacteria may have taken up exogenous DNA, potentially as a nutrient source, for DNA repair, or to increase genetic diversity [35], as seen in *Vibrios* attached to chitin [36].

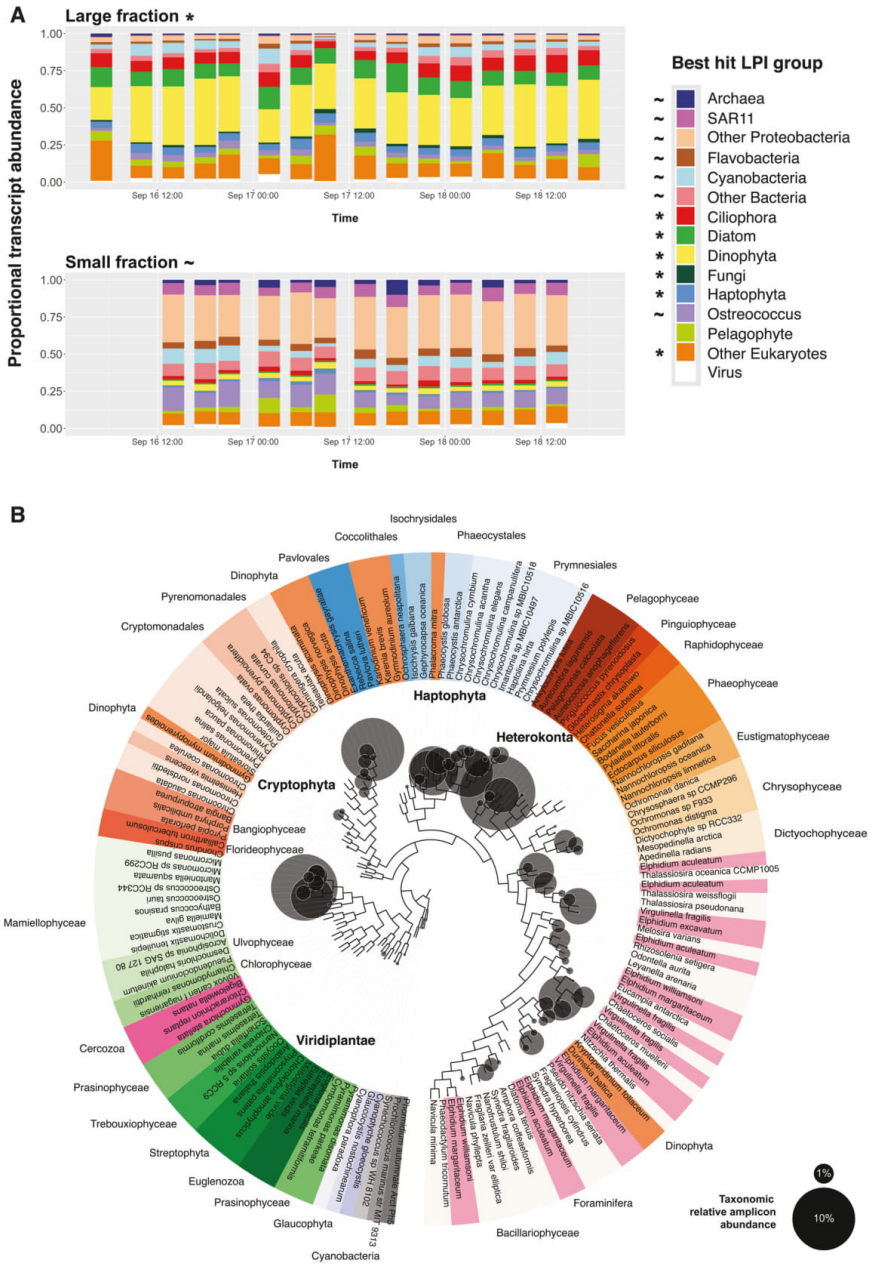
Patterns of community activity

In the large size class, total gene expression was often highly synchronized between members of a given taxonomic group (Fig. S11a), especially among prokaryotes. For example, flavobacteria and euryarchaeota ORFs showed strong ingroup correlation (Fig. S11a; Pearson's $r = 0.94$, 0.93, respectively) and were most highly expressed at night (Fig. S11c, d). In total, HRA detected ten large-fraction taxa, including the diatom, *Skeletonema* sp., and the bacteria, *Roseobacter* sp., with total gene expression that followed smooth day/night oscillation with a period of 24 h (Fig. S12).

Functional characterization of the active community

WGCNA on functional clusters established six unique patterns ("modules") of gene expression over time in the large fraction and three in the small fraction (Fig. 2). WGCNA on nucleotide sequences aligned to reference transcriptomes recapitulated major functions but did not capture the same breadth of novel phylogenetic diversity that resulted from aligning amino acid sequences of ab initio ORFs (Fig. S13). Ab initio analysis established 344,615 unique ORFs which recruited ~107 million reads, whereas mapping reads to reference transcriptomes captured only ~15 million reads mapping to 168,349 reference ORFs.

In the large size class, the most obvious drivers of gene expression were taxonomy and day/night cycles. The most abundant cluster in large fraction module 1 (Fig. 2; turquoise; up at night) was a small subunit ribosomal protein, almost entirely composed of centric diatoms. Interestingly, prokaryotic and eukaryotic viral capsids both clustered into


A
Large fraction *
Small fraction ~
Proportional transcript abundance
1.00
0.75
0.50
0.25
0.00
Sep 16 12:00
Sep 17 00:00
Sep 17 12:00
Sep 18 00:00
Sep 18 12:00
Time
Best hit LPI group
~ Archaea
~ SAR11
~ Other Proteobacteria
~ Flavobacteria
~ Cyanobacteria
~ Other Bacteria
* Ciliophora
* Diatom
* Dinophyta
* Fungi
* Haptophyta
~ Ostreococcus
Pelagophyte
* Other Eukaryotes
Virus
B
Isochrysidales
Coccolithales
Phaeocystales
Pavlovales
Prymnesiales
Dinophyta
Pyrenomonadales
Pelagophyceae
Cryptomonadales
Pinguiophyceae
Raphidophyceae
Phaeophyceae
Haptophyta
Dinophyta
Heterokonta
Eustigmatophyceae
Cryptophyta
Chrysophyceae
Bangiophyceae
Dictyochophyceae
Florideophyceae
Mamiellophyceae
Ulvophyceae
Chlorophyceae
Viridiplantae
Cercozoa
Prasinophyceae
Trebouxiophyceae
Streptophyta
Euglenozoa
Prasinophyceae
Glaucophyta
Cyanobacteria
Dinophyta
Foraminifera
Bacillariophyceae
Taxonomic relative amplicon abundance
1%
10%

◀ **Fig. 1** Major taxa found in the eastern North Pacific drift track. **a** Taxonomically annotated total community expression over time across size classes. Expression includes only non-organellar ORFs and is normalized by library (time point) within each fraction. Taxa grouping of each ab initio ORF is determined by best LPI hit. Asterisks and tildes denote taxa groups significantly enriched in the large and small size classes, respectively (edgeR, FDR <0.05). **b** Phylogenetic reconstruction of 16S rRNA gene reference sequences and distributions of active plastids represented in terms of cDNA-based amplicon relative abundances summed across all time points (circles; Supplementary Dataset 5). Circles representing relative amplicon abundance are superimposed over a reference phylogeny which is colored by taxonomy. Proximity of circles to the tips of the branches represents closeness of observed sequences to the references and circle sizes are proportional to normalized read abundance. Note that dinoflagellates known to have tertiary plastids are placed in this method according to the plastid origins (e.g., *Gymnodinium myriopyrenoides* and several *Dinophysis* species clading with cryptophytes [93])

this night-up module. Large fraction module 2 (blue; erratic) echoed the shape of overall dinoflagellate activity (Fig. S11a), and major cluster annotations (e.g., tubulin, bacteriorhodopsin-like protein, and bacterial DNA-binding protein (the dinoflagellate equivalent of histones [37])) were dominated by dinoflagellates. Module 5 (green; peaks in the early morning), on the other hand, was dominated by photosynthesis-related annotations such as chlorophyll A–B binding protein and G3P dehydrogenase but also contained metazoan histones.

In the small fraction, the majority of cluster annotations were in module 1 (Fig. 2; orange; up during first night). Most functions related to growth (ribosomal proteins, RNA polymerase, and elongation factor Tu) and nutrient acquisition (branched-chain amino acid transporters, Nit Tau family transport system, and multiple sugar transport system) and were dominated by Proteobacteria. Photosynthesis-related transcription in the small fraction could mostly be attributed to *Ostreococcus*, "other eukaryotes", and *Synechococcus*.

Physiological response of phytoplankton to day/night cycles

In addition to identifying data-driven patterns of gene expression with WGCNA, we also sought to probe diel physiology by fitting gene expression to a sinusoid with a 24 h period (HRA). Large portions of photoautotroph transcriptomes have been observed to oscillate with a 24-h period, often by known circadian mechanisms (e.g., *Arabidopsis thaliana* [38], *Synechococcus elongatus* [39], and *Ostreococcus tauri* [40]). Most previous observations of this light response were performed in artificially stable laboratory settings (e.g., 12:12-h light:dark cycles), but Ottesen et al. [10] observed a high number of *Synechococcus* and *Ostreococcus* transcripts in the small size fraction of this drift, including key clock, respiration, and photoautotrophy genes, oscillating with a 24-h period in natural environments [10].

Here, we expand this analysis to natural populations of large microbial eukaryotes for the first time. We observe significant diel transcriptional periodicity (FDR ≤ 0.1) in all active phytoplankton lineages, as well as in ciliates and some bacteria, and *Synechococcus* and *Ostreococcus* ORFs present in the large fraction (Fig. 3).

In addition, we detected phylogenetically novel Light-Oxygen-Voltage (LOV) domains, which have been implicated in transcriptional light response, including as a zeitgeber for the circadian clock [41]. LOV proteins in our data were associated with a range of effector domains, including kinases and b-ZIP transcription factors. LOV domains have rarely been characterized in marine plankton [42] and here demonstrated clear activity peaking just before dawn (Fig. S14).

Timing of diel expression differed across taxonomic groups

Significantly periodic ORFs were classified into four bins based on the time of day of peak expression (Fig. 3). Most photosynthetic eukaryotes had periodic expression in all four bins, but the majority of expression from diel ORFs occurred in the day (Fig. 3c). Ciliates and other largely heterotrophic eukaryotes, on the other hand, had evening-dominated periodic expression [43]. For prokaryotes, the results were more mixed (Supplementary File 1).

Timing of diel expression was partitioned by function

Across taxa groups, periodic non-organellar ORFs most commonly peaked in early day (~11 a.m.) and early night (~11 p.m.; Fig. 3d). Morning-peaking ORFs, coincident with peak photosynthetically active radiation, were dominated by photosynthesis and carbohydrate and lipid metabolism annotations. Evening-peaking ORFs related to chromatin structure and dynamics, cytoskeleton, and chromosome partitioning, possibly because evening-timed cell division minimizes UV-stress to an exposed genome [44, 45]. Comparative transcriptomics on *Ostreococcus*, *Chlamydomonas*, and *Arabidopsis* grown under alternating light:dark periods corroborates this conserved temporal partitioning of photosynthesis and cell-cycle genes [46].

WGCNA on all periodic ORFs independently recreated expression modules peaking in the early day, late day, early night, and late night (Fig. 3b). The early day module was most abundant, and was dominated by photosynthetic reaction center and chlorophyll AB binding proteins largely expressed by haptophytes, centric diatoms, pelagophytes, and some *Synechococcus* and *Ostreococcus* (Fig. 3e; red).

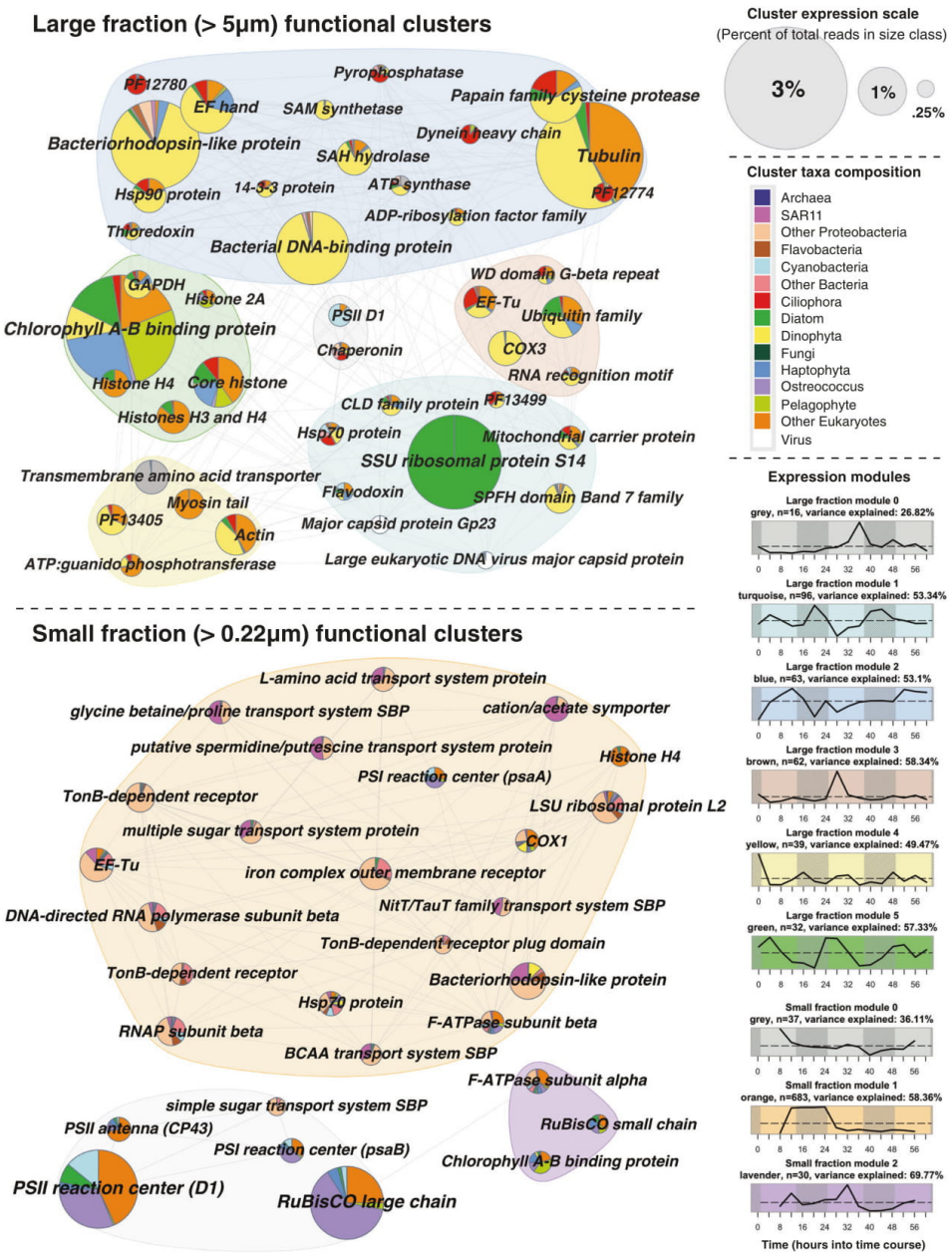


Fig. 2 A comparison of functional diversity across the large and small size classes. Pies represent highly abundant (>0.25% total size-class expression) annotated functional clusters of ab initio ORFs. Pies are colored by relative taxonomic contribution and grouped by modules of similar expression as given by WGCNA

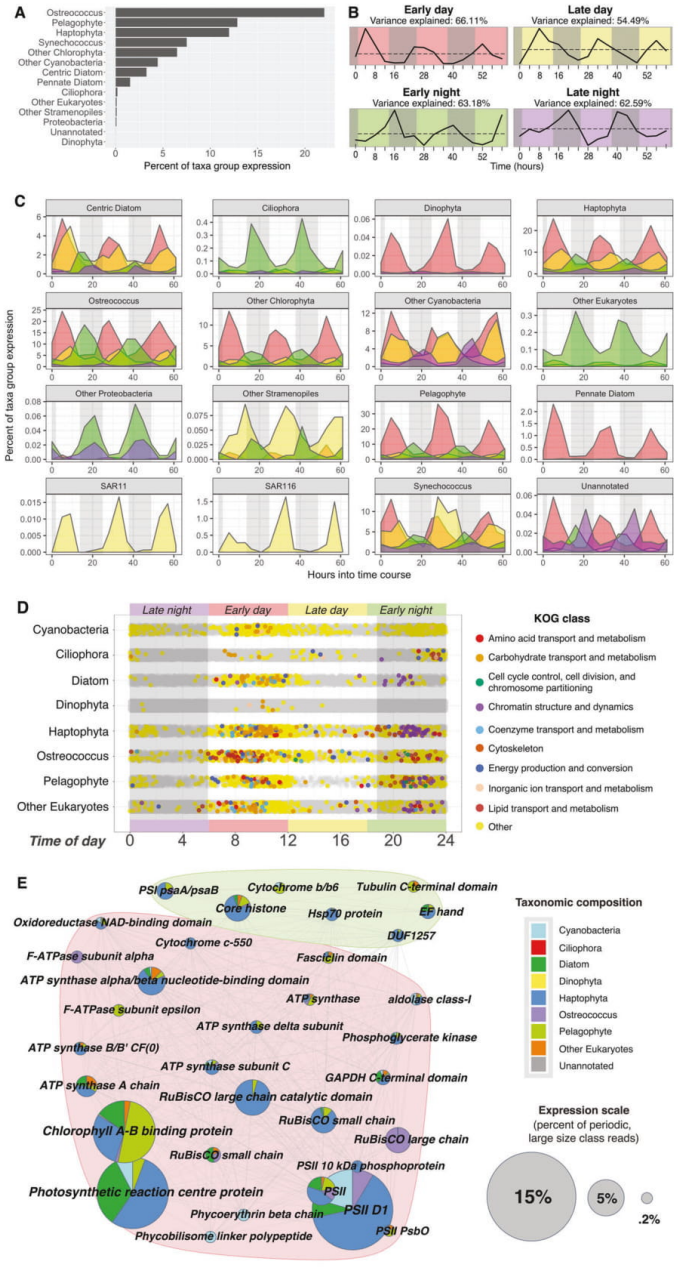


Fig. 3 Timing, abundance, and diversity of significantly diel large fraction ab initio ORFs (HRA on taxa group normalized ORFs; FDR ≤0.1). ORFs are categorized into four, 6-h long bins based on peak expression time: early day (red, 6 a.m.–12 p.m.), late day (yellow, 12 p.m.–6 p.m.), early night (green, 6 p.m.–12 a.m.), and late night (purple, 12 a.m.–6 a.m.). **a** Percent of total expression found to be significantly periodic across taxa groups. **b** Data-driven WGCNA modules of significantly periodic large fraction ORFs independently recreate peak expression time bins. Subtitles show percent of variance that can explained by each module's average expression profile. **c** Periodic expression colored by time of day bin across taxa groups. **d** Peak expression time of nuclear ORFs belonging to major phytoplankton players. Significantly periodic ORFs are depicted by colored dots, where colors correspond to KOG class; all other ORFs depicted in gray. The majority of periodic ORFs peak in early day (red; 54.8%) and early night (green; 31.2%). **e** Top annotated cluster annotations of significantly periodic large fraction ORFs. Pies are colored by relative taxonomic contribution (legend, right) and grouped by modules of similar expression (**b**)

This is consistent with laboratory findings for *Ostreococcus* [47] and vascular plants [38, 48] where mean peak expression of light harvesting and photosynthesis genes occurs in the middle of the day. Because our data is compositional in nature, it is a useful positive control to corroborate these established results with both WGCNA

and HRA. When sampled in constant light (after a brief entrainment to 12:12-h light:dark cycles), many *Arabidopsis* nuclear-encoded photosynthesis genes also peak at "midday" [38], suggesting that the conserved midday peak we observe is circadian in nature. The high-turnover nature of the photosynthetic reaction center protein pool [49] makes it likely that the 11 a.m. peak expression of these transcripts correctly captures the timing of protein activity. The morning peak also contained a large number of transcripts for rubisco and ATP synthase, suggesting that carbon fixation and energy production are similarly timed. Chlorophyll synthesis ORFs (e.g., CobN/magnesium chelatase and geranyl reductase) also peaked at 11 a.m. across taxa groups, except in Proteobacteria, where bacteriochlorophyll synthesis peaked around midnight (Fig. S15c). Metabolism-related ORFs, including various ATP synthases, mitochondrial carrier proteins, phosphogylceride kinase, fructose-biphosphate aldolase, and fatty acid desaturase shared this mid-morning peak, but a second set of ATP synthases and mitochondrial carrier protein ORFs peaked in the early night coincident with cytochrome C oxidase and NADH-ubiquinone/plastoquinone oxidoreductase (Fig. S16a).

The early night module (Fig. 3e; green) contained an abundance of histones, especially from haptophytes and pelagophytes. DNA polymerases, condensins, cyclins, and CDKs also peaked around 11 p.m. across several photosynthetic eukaryotes (Fig. S15b). Synchronized populations of *Synechococcus* [50], *Ostreococcus* [47], *Chlamydomonas* [51], *Phaeodactylum* [6], *Emiliania* [52], and *Pelagomonas* [53] divide in early night when grown on a 12:12-h light:dark cycle, consistent with the cell division machinery we observed peaking at night in *Synechococcus*, *Ostrococcus*, prasinophytes, diatoms, haptophytes, pelagophytes, and other eukaryotes. Axonemal ORFs were significantly periodic across a broad range of motile taxa, including pelagophytes (14 ORFs), ciliates (2 ORFs), haptophytes (1 ORF), and chlorophytes (1 ORF), all peaking in the early evening (mean ~9 p.m.).

Translation-related ORFs were also periodic across many lineages. Translation initiation factors were periodically transcribed in centric diatoms, ciliates, haptophytes, other chlorophytes, pelagophytes, and *Synechococcus* (Fig. S16d). Interestingly, we identified several periodic eukaryotic translation elongation factor 3 (eEF3) ab initio ORFs in the large fraction (Fig. S16d). Previously believed to be unique to fungi, eEF3 presents a novel peptide synthesis mechanism for phytoplankton.

Physiological interpretation of diel transcriptional partitioning

For high-turnover proteins, diel transcription is likely important for maintaining appropriate protein levels. However, many protein pools turnover too slowly for diel transcription to translate into diel changes in protein abundance, which is determined not only by transcription rates, but also translation and degradation rates. In *Arabidopsis*, *Ostreococcus*, and *Cyanothece*, the majority of proteins have half-lives that span multiple diel cycles [54]. In *S. elongatus*, only about 5% of proteins exhibit the diel dynamics that 30–60% of transcripts do [54, 55]. In *O. tauri*, under 10% of proteins are rhythmic despite nearly the whole transcriptome oscillating [56]. In this low-turnover case, diel transcriptional rhythms are more difficult to interpret.

One explanation for diel cycling transcripts in the case of stable protein abundance is "translational coincidence", a mechanism described in *Arabidopsis* in which the timing of transcription and translation interact to optimize use of solar energy for a given photoperiod [54]. In photosynthetic organisms spanning cyanobacteria, chlorophyte, and plant lineages, protein synthesis rates are 3–5 times higher during day than night [54]. In a dawn-tracking circadian clock, transcripts that peak in the late day during a long photoperiod (e.g., summer) peak after sunset during a short photoperiod (e.g., winter; Fig. 4a). Proteins with late-day transcripts would therefore be more abundant in long photoperiods because of reduced translation rates after sunset. In contrast, transcripts peaking in the early day or late night would not have seasonally variable protein pools. This mechanism is not only highly relevant for plants, in which seasonal adaptations spanning diverse physiological mechanisms from flowering to freezing tolerance are well described [54], but are also likely critical for unicellular algae. Seasonal phenotypes are poorly described in algae due to difficulty of observation, but some examples have been reported. For example, in response to short photoperiods (e.g., winter) *Lingulodinium* forms cysts and *Chlamydomonas* suppresses zygospore germination [57].

To assess what functions might be influenced by such seasonal adaptation, we analyzed 276 significantly periodic nuclear ORFs that peaked within the window in which seasonal adaptation would be expected (9.9–14.4 h after local dawn; Fig. 4b). All taxa groups that showed significantly periodic activity had ORFs peaking in this window except dinophyta and "other Proteobacteria". In the cCCS transition zone, iron limitation increases in tandem with photoperiod as the upwelling season progresses from spring into late summer [21]. Because iron is a photosynthetic cofactor, the stress of increased day length and low-iron likely compound in this season.

Indeed, the top annotations peaking in the seasonally adaptive window are implicated in responding to low iron and UV-stress. The most highly expressed annotation was the calcium-binding domain, EF-hand. Calcium signaling is best known in algae as being required for photoacclimation

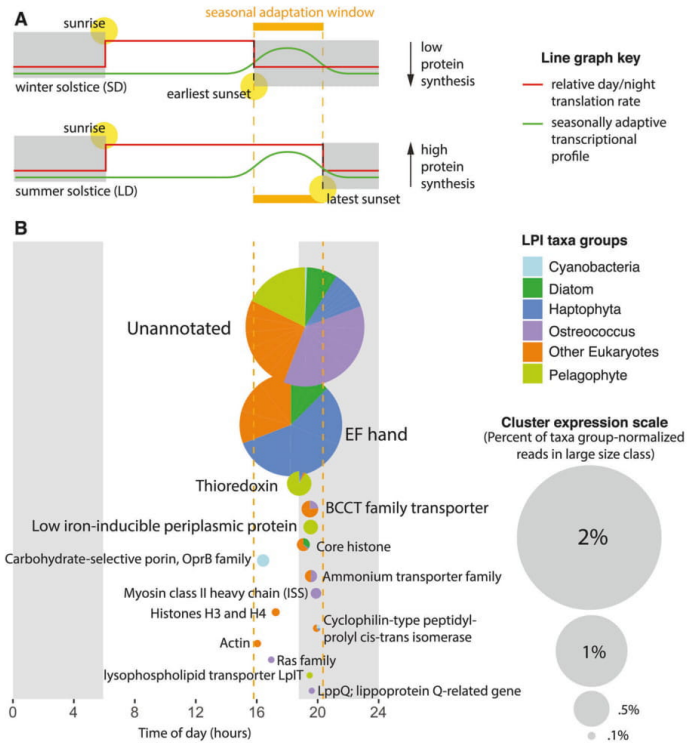


Fig. 4 Translational coincidence as a mechanism for seasonal adaptation in algae. **a** Schematic adapted from Seaton et al. [54] depicting the translational coincidence mechanism. The top graphic represents the shortest photoperiod of the year, the winter solstice (December 21, 2010); the bottom graphic represents the longest photoperiod of the year, the summer solstice (June 21, 2010). An ORF (green line) peaking in the early night during short day (SD) conditions (top) would peak in the late day during long day (LD) conditions (bottom). Because the average translation rate (red line) is much higher during daylight hours than night hours (gray boxes), an ORF peaking in this "seasonal adaptation window" (orange) would be upregulated in LD at the protein level. **b** Top non-organellar cluster annotations from the large size class of our drift track that are significantly diel and peak in the seasonal adaptation window (orange dashed lines). Pies are colored by relative taxonomic contribution and scaled to reflect proportional transcript abundance (legends, right)

[58] and low-iron response [59]. EF-hand-containing proteins, specifically, are associated with the low-iron phenotype in diatoms and haptophytes [8, 60]. The second most abundant annotation, thioredoxin, modulates the activity of photosynthesis proteins in response to light by sensing redox potential [61]. Three thioredoxins were significantly upregulated during long photoperiods in the *Arabidopsis* proteome [54]. Finally, low-iron-inducible periplasmic protein was the fourth most abundant annotation, and was dominated by the iron-uptake protein ISIP2A (phytotransferrin; ref. [62]).

When viewed with WGCNA, ISIP2A expression clustered with silicon transporters (Fig. S5, module 5; Supplementary File 1). Both were chiefly expressed by centric diatoms, which may be more sensitive to iron stress [63] because they tend not to use the ferritin mechanism favored by bloom-forming pennates. The expression pattern of module 5 had some day/night signal, but also peaked strikingly at the end of the drift track, when the silica:nitrate ratio dropped most dramatically. Low silica:nitrate ratios have been observed in association with iron limitation [20] and are thought to result from silica drawdown by iron-stressed diatoms [23]. This convolution of the influence of day/night cycles and nutrient limitation on patterns of expression is indicative of how transcription may be responding to multiple drivers in a dynamic natural context. Whereas the "up-at-dusk" component of module 5 expression could be a circadian-driven mechanism allowing the phytoplankton to be generally more responsive to iron stress in the season when it is most exacerbated, the apparent

additional level of upregulation on day three is likely a response to local conditions—namely, increasing iron stress at the end of the drift.

Cascade of photosynthetic activity takes place in the morning

In contrast to the long-lived proteins whose diel transcription may be seasonally relevant, short-lived proteins with diel oscillating mRNA are likely of daily importance. In our data, periodic photosynthesis ORFs peaked at a mean of 10:53 a.m. based on a sinusoidal fit (dashed line; Fig. 5a), but components of the photosynthetic apparatus peaked in a "cascade" throughout the morning, beginning around 9 a.m. with the phycobilisome (orange) and cytochrome b6f complex (lavender), and ending around noon with PSII reaction center D1/D2 (blue). *ftsH* protease was also periodic and peaked in time with photosynthesis genes in cyanobacteria and in eukaryotes. This is likely due to its role in both cyanobacteria [64] and in eukaryotic chloroplasts [49] in repairing PSII via the degradation of D1 proteins damaged by oxidative stress. All photosynthesis protein categories except PSI (violet) and light harvesting complexes associated with PSI (pink) and PSII (fuchsia) were significantly different from the overall photosynthesis mean peak time of 10:53 a.m. (Watson–Wheeler Test of Homogeneity of Means, FDR <0.05).

The timing of most of our photosynthesis ORFs is consistent with previous findings for naturally occurring picoplankton assemblages (Fig. S17). Similar photosynthetic cascades have also been observed in algae grown in light:dark conditions. In *Micromonas pusilla*, photosynthesis components show pronounced transcriptional changes in connection with the transition away from dawn [65]. In a higher resolution study of *Chlamydomonas* [51], patterns similar to those observed here were observed, but shifted slightly later, with B6F peaking early in the subjective day (~ZT4; 10 a.m.), followed by LHCI, PSI, and PSII in the middle of the subjective day (~ZT7; 1 p.m.), and LHCII last (ZT8; 2 p.m.). Likewise, in the diatom, *Phaeodactylum*, light harvesting machinery also peaks in the late afternoon [6]. This relative delay could be an effect of comparing 12:12-h light:dark cycles with natural conditions where daylight persists longer than 12 h.

Surprisingly, PSI reaction center ORFs, *psaA* and *psaB*, consistently peaked at night (~9:30 p.m.) in cyanobacteria and the chloroplasts of chlorophytes, haptophytes, and pelagophytes (Fig. 5; dark red), rather than forming part of the daytime expression cascade. This runs contrary to previous findings for *Prochlorococcus*, where *psaA* and *psaB* have been shown to peak around 8 a.m. in environmental [11] and noon in laboratory [66] settings. This difference could possibly be explained by the iron-limited setting of our drift track. Iron-rich components of PSI are downregulated during iron stress in *Prochlorococcus* [67], and *psaA* expression is specifically controlled by an Fe-responsive regulator in *Chlamydomonas* [68]. Deviations from the midday photosynthetic cascade also occurred in PSI and B6F, the two electron transport chain components that bind the most iron–sulfur clusters and are iron-stress responsive in diatoms [69]. However, to the best of our knowledge, such a dramatic shift in expression has not been previously been observed for *psaA*/*psaB* and its cause is ultimately unknown.

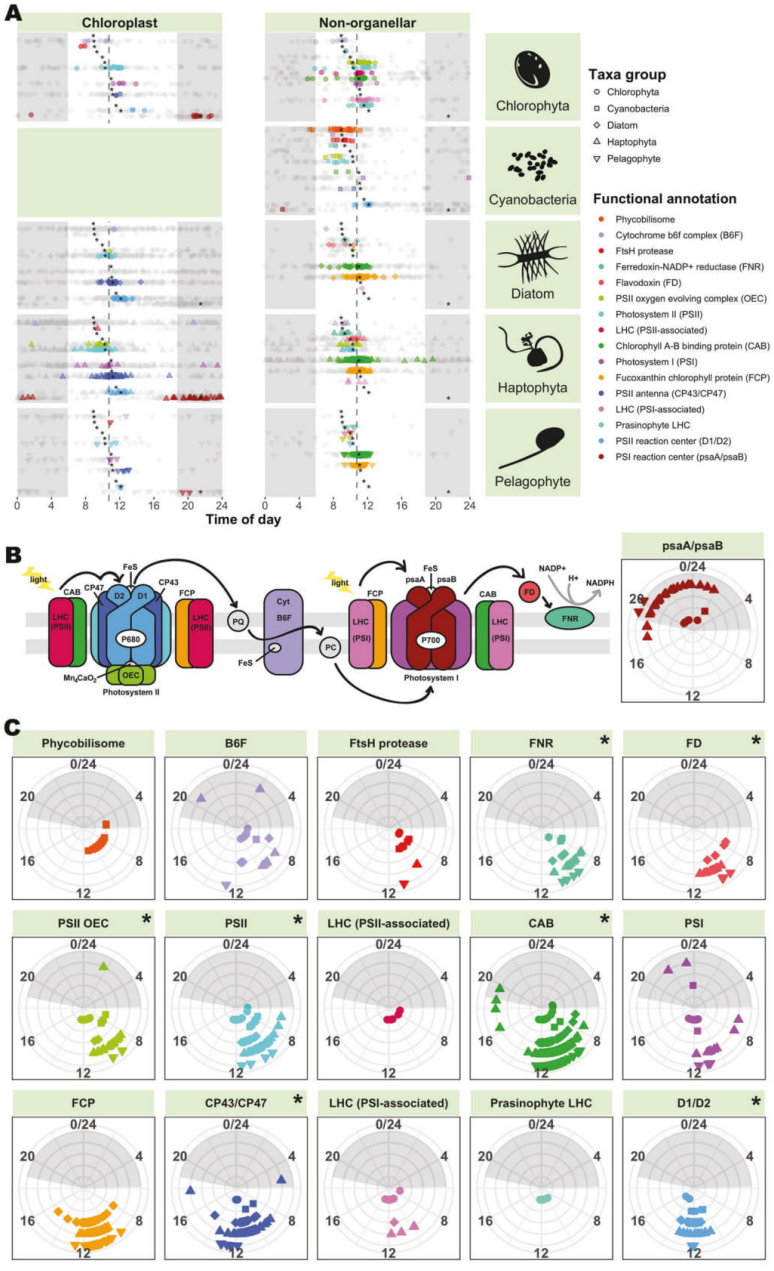
Diel transcriptional patterns vary across taxa

The percent of the overall transcriptome with significant 24-h oscillations varied greatly between taxonomic groups, and was highest in *Ostreococcus* (nearly 25%) and lowest in dinoflagellates (<1%; Fig. 3a). This difference could be explained by varied biological mechanisms for responding to sunlight. For example, dinoflagellates [70, 71] and ciliates [72] likely rely on post-transcriptional or post-translational mechanisms to a great degree. Here, dinoflagellates exhibited relatively consistent expression profiles that were anomalously highly correlated for eukaryotes ($r = 0.57$; Fig. S11a).

Dinoflagellates have been suggested to regulate their transcription differently than other algae and primarily respond to the environment post-transcriptionally [71, 73–77]. Dinoflagellate circadian oscillations have usually been observed at the level of translation [78, 79], with reports of only 3% of *Pyrocystis lunula* genes being transcribed on a day/night cycle [70] and even circadian-controlled cell-cycle regulators being post-transcriptionally controlled in *Karenia brevis* [71]. In *Lingulodinium polyedrum*, iron superoxide dismutase protein levels peak at midday, despite arrhythmic mRNA [73].

Despite their post-transcriptional response strategy, we detected 13 significantly periodic dinoflagellate nuclear ORFs, all of which recruited at least 69 reads. Eleven peaked in the early day and consisted of photosynthesis-related ORFs (chlorophyll A–B binding proteins or GAPDH), hypothetical proteins, and ATP sulfurylase, which catalyzes the first step in sulfate assimilation [80]. The remaining two ORFs both peaked around 1 a.m. and consisted of karyopherin alpha, an adaptor protein responsible for importing proteins to the nucleus, and a small ubiquitin-related modifier protein involved in post-translational modification. Diel cycling of such protein modifiers could shed light on currently cryptic diel behavior, such as bioluminescence, in dinoflagellates.

In contrast to the dinoflagellate and ciliate results, the high percentage of periodic ORFs in other taxa could be explained by light-synchronized division (as in


A
Chloroplast
Non-organellar
Time of day
Chlorophyta
Cyanobacteria
Diatom
Haptophyta
Pelagophyte
Taxa group
Chlorophyta
Cyanobacteria
Diatom
Haptophyta
Pelagophyte
Functional annotation
Phycobilisome
Cytochrome b6f complex (B6F)
FtsH protease
Ferredoxin-NADP+ reductase (FNR)
Flavodoxin (FD)
PSII oxygen evolving complex (OEC)
Photosystem II (PSII)
LHC (PSII-associated)
Chlorophyll A-B binding protein (CAB)
Photosystem I (PSI)
Fucoxanthin chlorophyll protein (FCP)
PSII antenna (CP43/CP47)
LHC (PSI-associated)
Prasinophyte LHC
PSII reaction center (D1/D2)
PSI reaction center (psaA/psaB)
B
psaA/psaB
C
Phycobilisome
B6F
FtsH protease
FNR
FD
PSII OEC
PSII
LHC (PSII-associated)
CAB
PSI
FCP
CP43/CP47
LHC (PSI-associated)
Prasinophyte LHC
D1/D2

◀ **Fig. 5** Peak expression time of large fraction ab initio ORFs involved in photosynthesis. Night is indicated by gray shading. **a** Cascade of peak expression time occurs across diverse phytoplankton lineages. ORFs are plotted by chloroplast encoded (left) and non-organellar (right). Significantly periodic ORFs (HRA; FDR ≤0.1) are colored by functional annotation (legend, right) and plotted in the same order as shown in the legend; insignificant ORFs are shown in grey. Asterisks denote functional annotation means across taxa groups that differ significantly (Watson–Wheeler Test of homogeneity of means, averaged over 1000 iterations to break ties, and Benjamini–Hochberg FDR <0.05) from the overall mean peak expression time (dashed gray line; 10:53 a.m.). **b** Illustration of the photosynthetic apparatus, colored by functional annotations from parts (**a**, **c**). Components with fewer than ten significantly periodic ORFs (pastoquinone (PQ), plastocyanin (PC)) are depicted in gray. Black arrows denote electron transport. **c** Conservation of peak expression time across phytoplankton lineages. Taxa groups are distinguished by shape (legend, top right) and radius (innermost: chlorophyta, outermost: pelagophyte). A Watson–Williams Test of homogeneity of means was performed on each functional group to determine taxonomic differences in peak expression time. Annotations with significantly different peak expression time across taxa groups (Benjamini–Hochberg FDR <0.05) are indicated with an asterisk

Ostreococcus), and cyclic expression of transcriptional machinery such as the preinitiation complex, RNA polymerase, and transcription factors in prasinophytes, *Synechococcus*, haptophytes, and pelagophytes (Fig. S16c).

In our data and the literature, diel partitioning of large portions of the transcriptome is a strategy most often adopted by phytoplankton with small cells and fast division rates. Diel transcription could allow such "streamlined" organisms to synchronize their protein pool with their transcript pool without manufacturing a large number of ribosomes. Temporally segregating a given transcript would increase its proportional abundance at the time of translation, which, according to information theory [81], could decrease fluctuations in protein number caused by ribosome sampling stochasticity. This could also provide an evolutionary motivation for producing precise waves of transcription for proteins that do not oscillate on a day/night cycle, as has been the perplexing case for the majority of genes in *S. elongatus* [54, 55] and *O. tauri* [56]. In contrast, we may observe fewer cyclic transcripts in larger organisms (e.g., ciliates, dinoflagellates) because they have the resources to maintain large transcript pools across all times of day. Indeed, large transcript pools could also provide greater flexibility in response to sudden environmental change, which may benefit the heterotrophic capabilities of these large taxa.

Host-virus interactions

We observed a diversity of viruses infecting bacteria and eukaryotes in both size classes (Fig. 6, S18). Viruses infecting large phytoplankton (e.g., *Heterocapsa*, *Chaetoceros*, *Emiliania*, and *Phaeocystis*) were enriched in the large fraction (Supplementary Dataset 10), whereas, in the small fraction, bacteriophages corresponding to several of the abundant bacterial groups were enriched (e.g. *Pelagibacter* phage, $\log_2FC = 11$; *Roseobacter* phage, $\log_2FC = 7.4$; *Vibrio* phage, $\log_2FC = 5.2$; Supplementary Dataset 10; Supplementary File 1).

In the large size class, the majority of viruses were dsDNA viruses (74%), but we also observed RNA viruses that infect phytoplankton [82] and labyrinthulids. While RNA virus reads could represent either transcribed mRNA or RNA genomic material, DNA virus transcripts indicate an active infection.

Remarkably, rather than host and virus expression being anticorrelated, as one might expect from kill-the-winner theory [83], co-expression was observed between dsDNA viruses and their hosts, which were homologous to diverse reference taxa including heterotrophic bacteria (e.g., *Pelagibacter*, *Enterobacteria*), cyanobacteria (e.g., *Synechococcus*, *Prochlorococcus*), photosynthetic eukaryotes (e.g., *Bathycoccus*, *Micromonas, Ostreococus lucimarinus*), and predatory heterotrophic eukaryotes (e.g., *Cafeteria roenbergensis*; Fig. 6, S18; pink boxes). One exception was *Phaeocystis*, with its aggregate gene expression peaking during daylight hours while both the giant *Phaeocystis globosa* virus and its virophage peaked synchronously at night. In addition, a virus infecting *Ostreococcus* Clade OII (OtV2) had clear night peaks in transcription, a phenomenon that has only ever been observed in the laboratory in experiments with the most distant of other *Ostreococcus* species, *O. tauri*, when infected by OtV5 [84]. Cyanophages also had peak expression at night, as previously observed [12]. The coordinated expression we observe between dsDNA viruses and hosts may result from replication of large viruses being more demanding on host metabolism [85, 86]. However, because we sequence bulk populations, we cannot be certain that host and virus transcripts originate from the same cell.

Unlike the predominance of host:dsDNA virus co-expression, ssRNA viruses related to those infecting the diatom, *Chaetoceros*, and the labyrinthulid, *Aurantiochytrium* were not co-expressed with their putative hosts. Rather, virus RNA molecule abundance lagged behind host transcription (Fig. 6; yellow boxes).

Most of the reference viruses that were used for gene mapping in our analyses have been well characterized. In laboratory experiments, it has been shown that their lytic cycles are 24–48 h in length even under various forms of nutrient limitation [82, 84, 87–89]. Hence, the detection of daily, closely synchronized transcription over more than two day/night periods suggests that a subset of each population of the major microbial players in this system, whether photosynthetic or not, was infected and lysed multiple times during the time course. Thus, rather than a few taxa being in a bloom scenario with epidemiology that

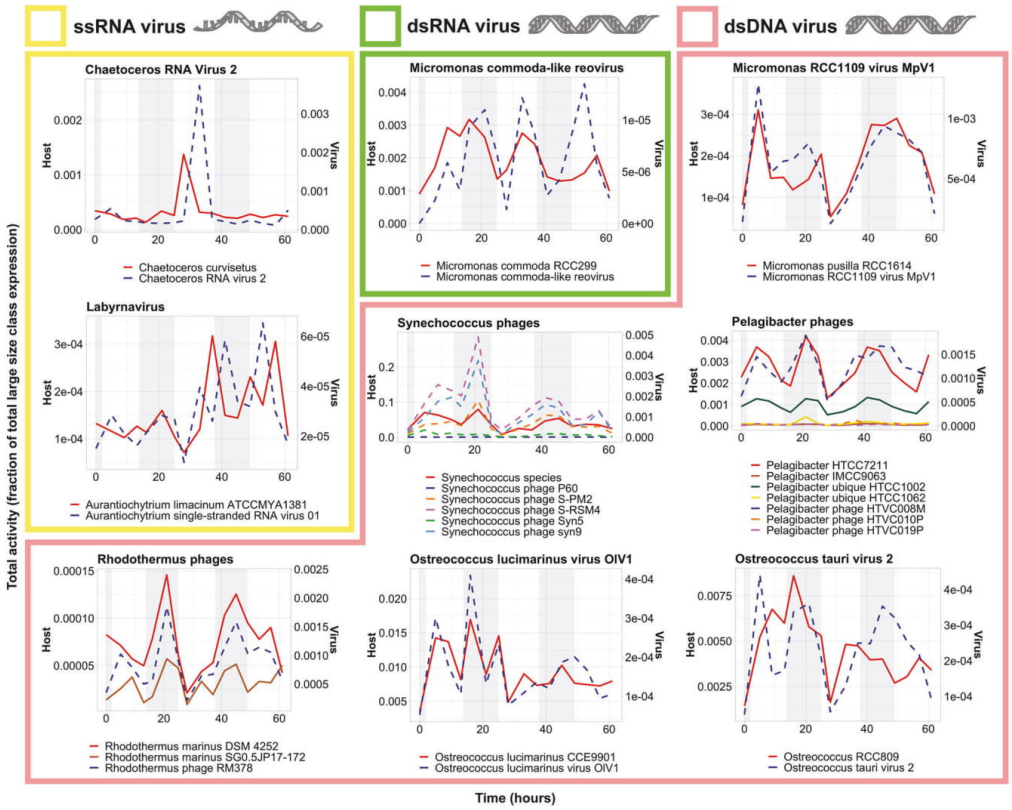


Fig. 6 Virus/host dynamics in the large size class. Viruses and hosts are annotated as the closest reference available in our database, as determined by LPI. Library normalized expression of ORFs classified as ssRNA (yellow), dsRNA (green), and dsDNA (pink) viruses and their putative hosts by LPI are shown. Putative host expression is represented by solid lines and corresponds to left y-axes; virus expression is represented by dashed lines and corresponds to the right y-axes. Night hours are shaded in gray. Note that OtV2 infects RCC393, an *Ostreococcus* Clade OII species, not *O. tauri*. OtV2 was isolated against *Ostreococcus* Clade OII isolate RCC393 [94] which has 99% 18S rDNA identity to the genome sequenced Clade OII isolate RCC809 used in our mapping analysis. Likewise, the *M. commoda-like* reovirus infects the strain LAC38, which was initially misreported as being *M. pusilla* and has been renamed here according to proper species assignment of the host [87]. Interestingly, while picoprasinophyte populations mapping most closely to *Micromonas pusilla* were coactive with a dsDNA virus, populations mapping most closely to *Micromonas commoda* (RCC299) were coactive with a dsRNA virus [87, 95]

would facilitate a massive viral lysis event, most taxa appeared to be under perpetual predation by viruses.

Conclusions and future directions

We present results from an often iron-limited, high-nutrient, low-chlorophyll setting in the eastern north Pacific [21]. We observed a pronounced transcriptional response indicative of existing iron limitation as well as a mechanism for seasonal adaptation to low iron and long photoperiod. We demonstrate that future measurements of diel transcription have the potential to elucidate not just daily, but also seasonal, phytoplankton physiology by considering the implications of varied day/night translation rates. Due to its high resolution and semi-Lagrangian nature, samples collected from this drift track present an opportunity to study in situ phytoplankton physiological responses to day/night cycling—something rarely investigated for eukaryotes. We report, for the first time in nature, diel cycling transcription across major phytoplankton lineages. The proportion of genes cycling varied taxonomically, possibly reflecting differences in life strategy or post-transcriptional regulation.

Notably, laboratory observations of marine phytoplankton (e.g., *Thalassiosira* [90], *Ostreococcus* [47], *Synechococcus*

[91], and *Prochlorococcus* [66]) report larger percentages of the transcriptome oscillating than we observed. This is likely due to increased statistical power enabled by deeper sequencing and higher percentages of mapped reads (i.e., mapping to a single model genome) as well as greater replication [92], but may also reflect physiological differences between ideal growth conditions and the patchy, dynamic nature of the marine environment.

Regardless of the proportion of genes cycling, functional partitioning of cycling genes was common across phytoplankton, pointing to a shared need to prepare for the daily onset of solar radiation and partition metabolic activity accordingly. Whereas heterotrophic bacterioplankton have been observed to display consecutive peaks in translation and oxidative phosphorylation-related transcription throughout the day [13], phytoplankton showed largely synchronized transcriptional timing. This may reflect differences between heterotrophic and photosynthetic life strategies. The unique metabolic needs of one bacterioplankton species might allow it to benefit from temporal association with another, but the largest driver of phytoplankton energetics is solar energy, which is concurrently available to all taxa. Our observation of a taxonomically conserved cascade of photosynthetic activity centered around ~11 a.m. and relegation of cell division to night [44, 45] may be explained by peak solar radiation and the risk of oxidative stress: both selective pressures that are shared by all photosynthetic organisms.

In addition to abiotic stressors, diverse phytoplankton and bacterial taxa appeared to be coping with viral infections, evidenced by high levels of viral RNA that often tracked putative host expression patterns. Indeed, the short lytic cycle of some of the closest viral references indicates that many host populations likely experienced multiple consecutive cycles of growth and viral lysis within the ~2.6 -day drift.

To date, in situ molecular sampling of large phytoplankton in the environment has not been applied widely, but future drifts have the potential to tease apart drivers of productivity in varied ocean conditions. Our results highlight the diversity of uncultured oceanic phytoplankton who remain "microbial dark matter" and the complexities of biogeochemical cycling by community interactions that are yet to be elucidated. We demonstrate that this technology allows for the observation of phytoplankton communities within the context of their natural biotic interactions and dynamic abiotic stressors. Such contextualization is an important step towards ascertaining global phytoplankton resilience to perturbation, and in turn, the resilience of the ecosystem services they provide.

Acknowledgements Cruise work was supported by the David and Lucile Packard Foundation through an annual grant to MBARI (FPC, CAS, and AZW). This study was supported by the National Science Foundation (NSF-OCE-1756884, NSF-OCE-1637632), United States Department of Energy Genomics Science program (DE-SC0008593 and DE-SC0018344), NOAA (NA15OAR4320071), and the Gordon and Betty Moore Foundation grant GBMF3828 (AEA). BK was funded by NSF graduate research fellowship DGE1144086. We would like to thank J. Ryan for AUV data depiction shown in Figure S1A and J. Giammona, A. Millar, and S. Smith for helpful discussions.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Behrenfeld MJ, O'Malley RT, Siegel DA, McClain CR, Sarmiento JL, Feldman GC, et al. Climate-driven trends in contemporary ocean productivity. Nature 2006;444:752–5.
2. Buchan A, LeCleir GR, Gulvik CA, Gonzalez JM. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nat Rev Microbiol. 2014;12:686–98.
3. Suttle CA. Viruses in the sea. Nature 2005;437:356–61.
4. Sarthou G, Timmermans KR, Blain S, Tréguer P. Growth physiology and fate of diatoms in the ocean: a review. J Sea Res. 2005;53:25–42.
5. Falkowski PG, Raven JA. Aquatic photosynthesis. 2nd ed. 41 William Street, Princeton, New Jersey: Princeton University Press; 2013. p. 488.
6. Smith SR, Gillard JTF, Kustka AB, McCrow JP, Badger JH, Zheng H, et al. Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. PLoS Genet. 2016;12:e1006490.
7. Guo J, Wilken S, Jimenez V, Choi CJ, Ansong C, Dannebaum R, et al. Specialized proteomic responses and an ancient photoprotection mechanism sustain marine green algal growth during phosphate limitation. Nat Microbiol 2018;3:781–90.
8. Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, et al. Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. Proc Natl Acad Sci USA. 2008;105:10438–43.
9. Scholin CA, Birch J, Jensen SRM III, Massion E, Pargett D, et al. The quest to develop ecogenomic sensors: a 25-year history of the Environmental Sample Processor (ESP) as a case study. Oceanography. 2017;30:100–13.
10. Ottesen EA, Young CR, Eppley JM, Ryan JP, Chavez FP, Scholin CA, et al. Pattern and synchrony of gene expression among

sympatric marine microbial populations. Proc Natl Acad Sci USA. 2013;110:E488–97.
11. Ottesen EA, Young CR, Gifford SM, Eppley JM, Marin R, Schuster SC, et al. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. Science. 2014;345:207–12.
12. Aylward FO, Boeuf D, Mende DR, Wood-Charlson EM, Vislova A, Eppley JM, et al. Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. Proc Natl Acad Sci USA. 2017;201714821. http://www.pnas.org/lookup/doi/10.1073/pnas.1714821114
13. Aylward FO, Eppley JM, Smith JM, Chavez FP, Scholin CA, DeLong EF. Microbial community transcriptional networks are conserved in three domains at ocean basin scales. Proc Natl Acad Sci USA. 2015;112:5443–8.
14. Bertrand EM, McCrow JP, Moustafa A, Zheng H, McQuaid JB, Delmont TO, et al. Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. Proc Natl Acad Sci USA. 2015;112:9938–43.
15. Altschul SF, Gish W, Miller W, Myers EW. & Lipman DJ. Basic local alignment search tool. J. Mol. Biol. 1990;215:403–10.
16. Podell S, Gaasterland T. DarkHorse: a method for genome-wide prediction of horizontal gene transfer. Genome Biol 2007;8:R16.
17. Bender SJ, Moran DM, McIlvin MR, Zheng H, McCrow JP, Badger J, et al. Colony formation in *Phaeocystis antarctica*: connecting molecular mechanisms with iron biogeochemistry. Biogeosciences 2018;15:4923–42.
18. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9:559.
19. Bruland KW, Rue EL, Smith GJ. Iron and macronutrients in California coastal upwelling regimes: implications for diatom blooms. Limnol Oceanogr 2001;46:1661–74.
20. Hutchins DA, Bruland KW. Iron-limited diatom growth and Si: N uptake ratios in a coastal upwelling regime. Nature 1998;393:561–4.
21. Biller DV, Bruland KW. The central California Current transition zone: a broad region exhibiting evidence for iron limitation. Prog Oceanogr 2014;120:370–82.
22. Till CP, Solomon JR, Cohen NR, Lampe RH, Marchetti A, Coale TH, et al. The iron limitation mosaic in the California Current System: factors governing Fe availability in the shelf/near-shelf region. Limnol Oceanogr. 2018;64:1–15.
23. Brzezinski MA, Krause JW, Bundy RM, Barbeau KA, Franks P, Goericke R, et al. Enhanced silica ballasting from iron stress sustains carbon export in a frontal zone within the California Current. J Geophys Res Ocean. 2015;120:4654–69.
24. Potvin M, Lovejoy C. PCR-based diversity estimates of artificial and environmental 18 S rRNA gene libraries. J Eukaryot Microbiol. 2009;56:174–81.
25. Sudek S, Everroad RC, Gehman ALM, Smith JM, Poirier CL, Chavez FP, et al. Cyanobacterial distributions along a physicochemical gradient in the Northeastern Pacific Ocean. Environ Microbiol 2015;17:3692–707.
26. Demir-Hilton E, Sudek S, Cuvelier ML, Gentemann CL, Zehr JP, Worden AZ. Global distribution patterns of distinct clades of the photosynthetic picoeukaryote Ostreococcus. ISME J 2011;5:1095–107.
27. Limardo AJ, Sudek S, Choi CJ, Poirier C, Rii YM, Blum M, et al. Quantitative biogeography of picoprasinophytes establishes ecotype distributions and significant contributions to marine phytoplankton. Environ Microbiol 2017;19:3219–34.
28. Simmons MP, Sudek S, Monier A, Limardo AJ, Jimenez V, Perle CR, et al. Abundance and biogeography of picoprasinophyte ecotypes and other phytoplankton in the eastern North Pacific Ocean. Appl Environ Microbiol. 2016;82:1693–705.
29. Griffith P, Douglas D, Wainright S. Metabolic activity of size-fractionated microbial plankton in estuarine, near-shore, and continental shelf waters of Georgia. Mar Ecol Prog Ser. 1990;59:263–70.
30. Allen LZ, Allen EE, Badger JH, McCrow JP, Paulsen IT, Elbourne LD, et al. Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. ISME J 2012;6:1403–14.
31. Ganesh S, Parris DJ, Delong EF, Stewart FJ. Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. ISME J 2013;8:187–211.
32. Bochdansky AB, Clouse MA, Herndl GJ. Eukaryotic microbes, principally fungi and labyrinthulomycetes, dominate biomass on bathypelagic marine snow. ISME J 2017;11:362–73.
33. Edwards JL, Smith DL, Connolly J, McDonald JE, Cox MJ, Joint I, et al. Identification of carbohydrate metabolism genes in the metagenome of a marine biofilm community shown to be dominated by Gammaproteobacteria and Bacteroidetes. Genes. 2010;1:371–84.
34. Christensen PJ. The history, biology, and taxonomy of the Cytophaga group. Can J Microbiol. 1977;23:1599–653.
35. Tani K, Nasu M. Roles of extracellular DNA in bacterial ecosystem. In: Kikuchi Y, Rykova EY, editors. Extracellular nucleic acids, nucleic acids and molecular biology. Berlin: Springer; 2010. p. 25–37.
36. Meibom KL, Blokesch M, Dolganov NA, Wu CY, Schoolnik GK. Chitin induces natural competence in Vibrio cholerae. Science. 2005;310:1824–7.
37. Janouškovec J, Gavelis GS, Burki F, Dinh D, Bachvaroff TR, Gornik SG, et al. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. Proc Natl Acad Sci USA. 2017;114:E171–80.
38. Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, et al. Orchestrated transcription of key pathways in arabidopsis by the circadian clock. Science. 2000;290:2110–3.
39. Liu Y, Tsinoremas NF, Johnson CH, Lebedeva NV, Golden SS, Ishiura M, et al. Circadian orchestration of gene expression in cyanobacteria. Genes Dev 1995;9:1469–78.
40. Corellou F, Schwartz C, Motta J-P, Djouani-Tahri EB, Sanchez F, Bouget F-Y. Clocks in the green lineage: comparative functional analysis of the circadian architecture of the picoeukaryote ostreococcus. Plant Cell. 2009;21:3436–49.
41. Djouani-Tahri EB, Christie JM, Sanchez-Ferandin S, Sanchez F, Bouget FY, Corellou F. A eukaryotic LOV-histidine kinase with circadian clock function in the picoalga Ostreococcus. Plant J. 2011;65:578–88.
42. Jaubert M, Bouly JP, Ribera d'Alcalà M, Falciatore A. Light sensing and responses in marine microalgae. Curr Opin Plant Biol. 2017;37:70–7.
43. Hu SK, Connell PE, Mesrop LY, Caron DA. A hard day's night: diel shifts in microbial eukaryotic activity in the North Pacific Subtropical Gyre. Front Mar Sci. 2018;5:351.
44. Nikaido SS, Johnson CH. Daily and circadian variation in survival from ultraviolet radiation in chlamydomonas reinhardtii. Photochem Photobiol 2000;71:758.
45. Pittendrigh CS. Temporal organization: reflections of a Darwinian clock-watcher. Annu Rev Physiol. 1993;55:17–54.
46. de los Reyes P, Romero-Campero FJ, Ruiz MT, Romero JM, Valverde F. Evolution of daily gene co-expression patterns from algae to plants. Front Plant Sci. 2017;8:1–22.
47. Monnier A, Liverani S, Bouvet R, Jesson B, Smith JQ, Mosser J, et al. Orchestrated transcription of biological processes in the marine picoeukaryote Ostreococcus exposed to light/dark cycles. BMC Genom. 2010;11:192.
48. Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL. Global transcriptome analysis reveals circadian regulation of key

pathways in plant growth and development. Genome Biol. 2008;9: R130.
49. Zaltsman A. Two types of FtsH protease subunits are required for chloroplast biogenesis and photosystem II repair in Arabidopsis. Plant Cell. 2005;17:2782–90.
50. Sweeney BM, Borgese MB. A circadian rhythm in cell division in a prokaryote, the cyanobacteria Synechococcus WH78031. J Phycol. 1989;25:183–6. p
51. Zones JM, Blaby IK, Merchant SS, Umen JG. High-resolution profiling of a synchronized diurnal transcriptome from Chlamydomonas reinhardtii reveals continuous cell and metabolic differentiation. Plant Cell 2015;27:2743–69.
52. Müller M, Antia A, LaRoche J. Influence of cell cycle phase on calcification in the coccolithophore Emiliania huxleyi. Limnol Oceanogr 2008;53:506–12.
53. Jacquet S, Partensky F, Lennon JF, Vaulot D. Diel patterns of growth and division in marine picoplankton in culture. J Phycol 2001;37:357–69.
54. Seaton DD, Graf A, Baerenfaller K, Stitt M, Millar AJ, Gruissem W. Photoperiodic control of the Arabidopsis proteome reveals a translational coincidence mechanism. Mol Syst Biol. 2018;14:e7962.
55. Guerreiro ACL, Benevento M, Lehmann R, van Breukelen B, Post H, Giansanti P, et al. Daily rhythms in the cyanobacterium synechococcus elongatus probed by high-resolution mass spectrometry–based proteomics reveals a small defined set of cyclic proteins. Mol Cell Proteom. 2014;13:2042–55.
56. Noordally ZB, Hindle M, Martin SF, Seaton DD, Simpson TI, Le Bihan T, et al. Circadian protein regulation in the green lineage I. A phospho-dawn anticipates light onset before proteins peak in daytime. Running. BioRxiv. 2018.
57. Suzuki L, Johnson CH. Algae know the time of day: circadian and photoperiodic programs. J Phycol 2001;37:933–42.
58. Petroutsos D, Busch A, Janßen I, Trompelt K, Bergner SV, Weinl S, et al. The chloroplast calcium sensor CAS is required for photoacclimation in Chlamydomonas reinhardtii. Plant Cell 2011;23:2950–63.
59. Falciatore A, D'Alcalà MR, Croot P, Bowler C. Perception of environmental signals by a marine diatom. Science. 2000;288:2363–6.
60. Bender SJ, Moran DM, McIlvin MR, Zheng H, McCrow JP, Badger J, et al. Colony formation in *Phaeocystis antarctica*: Connecting molecular mechanisms with iron biogeochemistry. Biogeosciences. 2018;15:4923–42. https://www.biogeosciences-discuss.net/bg-2017-558/.
61. Scheibe R, Backhausen JE, Emmerlich V, Holtgrefe S. Strategies to maintain redox homeostasis during photosynthesis under changing conditions. J Exp Bot. 2005;56:1481–9.
62. McQuaid JB, Kustka AB, Oborník M, Horák A, McCrow JP, Karas BJ, et al. Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. Nature 2018;555:534–7.
63. Lampe RH, Mann EL, Cohen NR, Till CP, Thamatrakoln K, Brzezinski MA, et al. Different iron storage strategies among bloom-forming diatoms. Proc Natl Acad Sci USA. 2018;115:E12275–84. http://www.pnas.org/lookup/doi/10.1073/pnas.1805243115.
64. Silva P, Thompson E, Bailey S, Kruse O, Mullineaux CW, Robinson C, et al. FtsH is involved in the early stages of repair of photosystem II in Synechocystis sp PCC 6803. Plant Cell 2003;15:2152–64.
65. Duanmu D, Bachy C, Sudek S, Wong C-H, Jimenez V, Rockwell NC, et al. Marine algae and land plants share conserved phytochrome signaling systems. Proc Natl Acad Sci USA. 2014;111:15827–32.
66. Zinser ER, Lindell D, Johnson ZI, Futschik ME, Steglich C, Coleman ML, et al. Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, Prochlorococcus. PLoS One 2009;4:e5135.
67. Thompson AW, Huang K, Saito MA, Chisholm SW. Transcriptome response of high- and low-light-adapted Prochlorococcus strains to changing iron availability. ISME J 2011;5:1580–94.
68. Douchi D, Qu Y, Longoni P, Legendre-Lefebvre L, Johnson X, SchmitzLinneweber C, et al. A nucleus-encoded chloroplast phosphoprotein governs expression of the photosystem I subunit PsaC in Chlamydomonas reinhardtii. Plant Cell 2016;28:1182–99.
69. Strzepek RF, Harrison PJ. Photosynthetic architecture differs in coastal and oceanic diatoms. Nature 2004;431:689–92.
70. Okamoto OK, Hastings JW. Novel dinoflagellate clock-related genes identified through microarray analysis. J Phycol 2003; 39:519–26.
71. Brunelle SA, Van Dolah FM. Post-transcriptional regulation of S-Phase genes in the dinoflagellate, karenia brevis. J Eukaryot Microbiol. 2011;58:373–82.
72. Almeida R, Allshire RC. RNA silencing and genome regulation. Trends Cell Biol. 2005;15:251–8.
73. Okamoto OK, Robertson DL, Fagan TF, Hastings JW, Colepicolo P. Different regulatory mechanisms modulate the expression of a dinoflagellate ironsuperoxide dismutase. J Biol Chem. 2001;276:19989–93.
74. Alexander H, Rouco M, Haley ST, Wilson ST, Karl DM, Dyhrman ST. Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. Proc Natl Acad Sci USA. 2015;112:E5972–9.
75. Hackett JD, Anderson DM, Erdner DL, Bhattacharya D. Dinoflagellates: a remarkable evolutionary experiment. Am J Bot. 2004;91:1523–34.
76. Moustafa A, Evans AN, Kulis DM, Hackett JD, Erdner DL, Anderson DM, et al. Transcriptome profiling of a toxic dinoflagellate reveals a gene-rich protist and a potential impact on gene expression due to bacterial presence. PLoS One. 2010;5:e9688.
77. Carradec Q, Pelletier E, Da Silva C, Alberti A, Seeleuthner Y, Blanc-Mathieu R, et al. A global ocean atlas of eukaryotic genes. Nat Commun 2018;9:373.
78. Roenneberg T, Morse D. Two circadian oscillators in one cell. Nature 1993;362:362–4.
79. Brunelle SA, Hazard ES, Sotka EE, Van Dolah FM. Characterization of a dinoflagellate cryptochrome blue-light receptor with a possible role in circadian control of the cell cycle. J Phycol. 2007;43:509–18.
80. Prioretti L, Gontero B, Hell R, Giordano M. Diversity and regulation of ATP sulfurylase in photosynthetic organisms. Front Plant Sci. 2014;5:1–12.
81. MacKay DJC. Information theory, inference, and learning algorithms. vol. 41, Choice Reviews Online. Cambridge University Press; 2013. p. 41-5949-41-5949.
82. Nagasaki K. Dinoflagellates, diatoms, and their viruses. J Microbiol 2008;46:235–43.
83. Thingstad TF. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol Oceanogr 2000;45:1320–8.
84. Derelle E, Yau S, Moreau H, Grimsley NH. Prasinovirus attack of Ostreococcus is furtive by day but savage by night. J Virol 2017;92:01703–17. JVI
85. Sanchez EL, Lagunoff M. Viral activation of cellular metabolism. Virology 2015;479–480:609–18.
86. Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, et al. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. Proc Natl Acad Sci USA. 2011;108:E757–64.

87. Bachy C, Charlesworth CJ, Chan AM, Finke JF, Wong CH, Wei CL, et al. Transcriptional responses of the marine green alga Micromonas pusilla and an infecting prasinovirus under different phosphate conditions. Environ Microbiol 2018;20:2898–912.
88. Proctor LM, Fuhrman JA. Viral mortality of marine bacteria and cyanobacteria. Nature 1990;343:60–2.
89. Puxty RJ, Evans DJ, Millard AD, Scanlan DJ. Energy limitation of cyanophage development: Implications for marine carbon cycling. ISME J. 2018;12:1273–86. https://doi.org/10.1038/s41396-017-0043-3.
90. Ashworth J, Coesel S, Lee A, Armbrust EV, Orellana MV, Baliga NS. Genomewide diel growth state transitions in the diatom Thalassiosira pseudonana. Proc Natl Acad Sci USA. 2013;110:7518–23.
91. Cohen SE, Golden SS. Circadian rhythms in cyanobacteria. Microbiol Mol Biol Rev. 2015;79:373–85.
92. Hughes ME, Abruzzi KC, Allada R, Anafi R, Arpat AB, Asher G, et al. Guidelines for genome-scale analysis of biological rhythms. J Biol Rhythms. 2017;32:380–93.
93. Park MG, Kim M, Kim S. The acquisition of plastids/phototrophy in heterotrophic dinoflagellates. Acta Protozool 2014;53:39–50.
94. Weynberg KD, Allen MJ, Gilg IC, Scanlan DJ, Wilson WH. Genome sequence of Ostreococcus tauri virus OtV-2 throws light on the role of picoeukaryote niche separation in the ocean. J Virol 2011;85:4520–9.
95. Brussaard CPD, Noordeloos AAM, Sandaa RA, Heldal M, Bratbak G. Discovery of a dsRNA virus infecting the marine photosynthetic protist Micromonas pusilla. Virology 2004;319:280–91.

Supplementary File 1: Supplementary Text

Supplementary Materials and Methods

Sample Collection

Samples were collected in association with Ottesen et al. 2013 off the coast of California from September 16-19, 2010 along the warm side of an upwelling-driven front (1). Briefly, the Environmental Sample Processor (ESP; 2) was suspended at 23 m depth below a semi-Lagrangian surface float, collecting 1L of seawater every ~4 h for 61 h (~2.6 days). Retained particulates were size fractionated onto 5 μm and 0.22 μm Durapore 25 mm filters (Millipore, Billerica, MA, USA), preserved immediately *in situ* via a 2 min incubation in RNALater (Ambion) and stored at -80°C within 36 hours of ESP recovery. Nutrients, chlorophyll, and other oceanographic metadata was obtained via shipboard CTD/niskin rosette casts (Supplementary Dataset 12) as previously described (3). Drift speed was determined via a surface-float mounted GPS sensor. Water speed relative to the drifter was measured using a surface-float mounted ADCP (Supplementary Dataset 12A) in order to detect deviations from truly Lagrangian sampling (e.g. wind forcing).

Metatranscriptome library preparation and sequencing

Small size class metatranscriptomes were sequenced by Ottesen et al. using a GS Titanium system (Roche) according to their previously published methods (1). Large size class cDNA was prepared as in Ottesen et al. from ribosomal RNA-depleted

total RNA (1). 1 ul of cDNA per sample was used to prepare metatranscriptome libraries with the Truseq RNA Sample Prep kit v2 (Illumina™) according to manufacturer's instructions starting from the end repair step. Libraries were paired-end sequenced on the Illumina HiSeq 2000 platform to obtain 2x100bp reads.

Amplicon library preparation and sequencing

Large size class 16S and 18S ribosomal RNA were sequenced using 454 GS FLX Titanium pyrosequencing. Nearly universal bacterial primers 341F (5'-CCTACGGGNGGCWGCAG-3') (4) and 926R (5'-CCGTCAATTCMTTTRAGT-3')(5) were used to target the v3v5 region of 16S and primers and TAReuk454FWD1 (5′-CCAGCASCYGCGGTAATTCC-3′) and TAReukREV3 (5′-ACTTTCGTTCTTGATYRA-3′) (6), were used to target the v4 region of 18S, each amplifying an approximately 500 bp region of cDNA. FLX Titanium adapters (A adapter sequence: 5′ 127 CCATCTCATCCCTGCGTGTCTCCGACTCAG 3′; B adapter sequence: 5′ 128 CCTATCCCCTGTGTGCCTTGGCAGTCTCAG 3′) and 10bp multiplex identifier (MID) barcodes were used for multiplexed 454 sequencing.

cDNA was prepared from 50 ng per sample of total RNA using the Life Technologies SuperScript III First Strand Synthesis system with random hexamer primers. cDNA concentration ranged from 312 – 18,440 pg/microliter. 1 µl of cDNA was used as a template amplified using Life Technologies AccuPrime PCR system kit, in a reaction containing 1X AccuPrime Buffer II, .75 units of AccuPrime Taq High Fidelity, and a final primer concentration of 200 nM, alongside a no template negative control for cDNA synthesis. Amplifications were performed using a Life Technologies ProFlex PCR system, with an initial denaturation at 95°C for 2 minutes, 30 cycles of 95°C for 20

seconds, 56°C for 30 seconds, 72°C for 5 minutes. PCR products (2 µl of each sample and 5 µl of negative control) were run on a 1% agarose gel at 105 V for 35 minutes, then cleaned with Ampure XP beads (Beckman Coulter, Brea CA), and resuspended in 25 µL of Qiagen elution buffer. 2.5 µL was used for visualization on an agarose gel, 1 µL was used in a LifeTechnologies' PicoGreen Quant-IT assay to quantify the final product, and 45 ng of both 16S and 18S amplicons were pooled separately for 454 pyrosequencing.

The vendor's standard protocols (Roche Diagnostics) were used for library QC, emPCR, enrichment and 454 sequencing with the following modifications: KAPA Biosystems Library Quantification Kit for qPCR was used to accurately estimate the number of molecules needed for emPCR, automation (BioMek FX) was used to "break" the emulsions after emPCR, and butanol was used to for ease of handling during the breaking process. The bead enrichment process was automated by using Roche's REM e (Robotic Enrichment Module).

Bioinformatic analysis of metatranscriptomes

See Figure S2 for illustration of metatranscriptomic analysis pipeline.

Open reading frame (ORF) calling and annotation

Large fraction Illumina reads and small fraction 454 reads were processed via the RNAseq Annotation Pipeline (rap) v0.4 (7). Small fraction reads were obtained via DBCLS SRA (http://sra.dbcls.jp/) using accession number SRA062433. Reads from both fractions were trimmed to remove primers and areas of low sequence quality (reads must be at least 30 base pairs (bp) long and have a quality score of at least 33 to

be retained). Illumina reads were paired. Ribosomal RNA (rRNA) reads were removed using Ribopicker v.0.4.3 (8). Large fraction reads were assembled using CLC Genomics Workbench 9.5.3 (https://www.qiagenbioinformatics.com/) first by library, then overall. Small fraction reads were left unassembled due to the longer read length, lower coverage nature of 454 sequencing. *Ab initio* ORF prediction was performed with FragGeneScan v1.16 (9) with parameters: complete=0 and train=complete. ORFs were once again screened for contamination in the form of rRNA, ITS, and primers. ITS sequences were downloaded from NCBI, and reduced to 397,062 non-redundant sequences at 0.95 level using cd-hit-est v4.6. Sequences that aligned with an ITS sequence with BLASTN e-value <= 1e-5 were removed. Primer and adapter sequences used in Illumina sequencing were searched using BLASTN and were identified at e-value <= 10. Reads were removed with hits to terminal ends at least 10bp in length, or internal hits at least 15bp in length. Possible organelle genes were classified for query sequences that had closer homology to an organelle gene than a nuclear gene within the organism with the closest known segregated organelle and nuclear genomes based on best BLASTP e-value <= 1e-3.

ORFs were annotated via BLASTP (10,11) alignment (e-value threshold $1e^{-3}$) to a comprehensive protein database, *phyloDB*, as well as screened for function *de novo* by assigning Pfams, TIGRfams and transmembrane tmHMMs with hmmer 3.0 (http://hmmer.org/; 12) using an e-value threshold of $1.0e^{-4}$. PhyloDB version 1.076 consists of 24,509,327 peptides from 19,962 viral, 230 archaeal, 4910 bacterial, and 894 eukaryotic taxa (13–15). It includes peptides from the 410 taxa of the Marine Microbial Eukaryotic Transcriptome Sequencing Project

(http://marinemicroeukaryotes.org/), as well as peptides from KEGG, GenBank, JGI, ENSEMBL, CAMERA to KEGG, GenBank, JGI, ENSEMBL, iMicrobe, and the Chloroplast Genome Database (cpbase). Taxonomic annotation of ORFs was also conducted via a BLASTP to phyloDB, and a Lineage Probability Index (LPI) was calculated to avoid biases introduced by classifying ORFs based on best BLAST hit alone (7,16,17). Briefly, LPI was calculated here as a value between 0 and 1 indicating lineage commonality among the top 95-percentile of sequences based on BLAST bit-score.

Illumina sequencing of large size fraction total ribosomal-depleted RNA yielded 623,461,310 raw reads across 16 time points, of which 265,345,754 were mRNA (~43%). A total of 283,760 contigs were assembled upon which 345,355 ORFs were called. 32,271,421 reads (~12%) mapped to the 111,655 ORFs that remained after strict filtering (~32%). Ottesen et al.'s GS FLX Titanium (Roche) sequencing of small fraction cDNA yielded 9,985,281 raw reads across 13 time points (1). 2,802,084 ORFs were called on the 5,618,280 trimmed reads remaining after quality control (~56%). Full assembly and annotation statistics in Supplementary Dataset 1. Coverage across taxa groups can be seen in Figure S3A.

Mapping to reference transcriptomes

Reference transcriptomes were chosen for read mapping that appeared with high abundance and percent identity among *ab initio* large fraction ORFs. Representative references were chosen from all major taxonomic groups found in *ab initio* ORFs. Large and small fraction reads were aligned to reference ORFs using BWA-MEM

version 0.7.12-r1039 (18,19) using default parameters. At least 50% of each read must map to a reference gene at least 80% identity to be considered a hit. References with at least 1000 genes with at least 5 reads mapped were functionally annotated via rap v0.4 as above and considered for downstream analysis. Coverage of annotated references can be seen in Figure S3B.

Hierarchical clustering

Reference transcriptome ORFs were hierarchically clustered together with all large and small fraction *ab initio* ORFs (including organellar ORFs) to form peptide ortholog groups via the Markov Cluster Algorithm (MCL; https://micans.org/mcl/; 20). Directional edge weights were defined as the ratio of pairwise- to self- BLASTP scores, and default parameters were used to assign ORFs to clusters. Clusters were assigned a consensus annotation if found to be statistically enriched in that annotation with a Fisher's exact test ($p < 0.05$). Consensus annotations must also represent at least 10% of the reads in the cluster and account for a minimum of 200 reads. Clusters with identical consensus annotations were grouped together into "functional clusters."

Identification of significantly periodic ORFs

ORFs with significantly periodic diel expression were identified using harmonic regression analysis (HRA) as previously described (1,21,22). Briefly, for each taxa group of interest, raw ORF counts over time were fit to a generalized linear model (glm) of a sinusoid with a 24-hour period with taxa-specific library sums serving as an offset at each time point. The model was constructed using the "glm" function in R (23) and

statistical significance was determined by False Discovery Rate (FDR; (24) adjusted p-values of <= 0.1 (Benjamini-Hochberg), on both a permutation test (500-50,000 permutations) and a chi-squared test.

Identification of conserved expression modules

The Weighted Gene Correlation Network Analysis (WGCNA) R package (25,26) was used as previously described (22) to identify modules of conserved expression among reference ORFs (Figure S13) and functional clusters (Figures 2, 3E, S5). ORFs/functional clusters with at least 10 raw counts in at least 80% of time points were considered. For Figure 3E, it was further stipulated that clusters must have at least 100 raw reads overall to be considered. In all cases, expression was normalized by total pre-filtration counts at a single time point (library) before constructing a Pearson correlation matrix. An adjacency matrix was then constructed from the correlation matrix by applying a power function (AF(s)=s^b). The lowest b value that allowed for a scale-free topology R-squared value above 0.8 was chosen, as recommended in the WGCNA user manual, in order to optimize the mean number of connections of the network while preserving scale independence. A signed Topological Overlap Matrix (TOM) was constructed from the correlation matrix to measure dissimilarity between each pair of nodes based on shared neighbors. Average linkage hierarchical clustering was used to define a dendrogram (cluster tree) of the network via the “blockwiseModules” function. A cut height of .995 as well as a minimum module size of 30 ORFs/functional clusters was used to delineate branches of the hierarchical clustering tree into modules of co-expression. The “moduleEigengenes” function was used with default parameters to

calculate “eigengenes” (a measure of “average” expression calculated as the first principal component of the module's expression matrix) for each module. Modules with correlated eigengenes were merged by setting a “mergeCutHeight” threshold of 0.5. ORFs/functional clusters with a correlation of less than 0.3 to their respective module eigengene were removed and classified as “unassigned” (module 0). The igraph package (27) was used to visualize expression networks.

Differential expression analysis

Differential expression of ORFs, ortholog clusters, taxa groups, and genera across size classes was identified using the R package edgeR (28). Categories (i.e. ORFs, ortholog clusters, taxa groups, genera) with least 1 read per million in at least 3 samples were included and used to calculate log fold changes. Counts were normalized using the “calcNormFactors” function, which accounts for both library size and varied library composition. An exact test with tagwise dispersion estimation was used to determine ORFs or clusters with significantly different expression across size classes (FDR-corrected $p < 0.05$).

Identification and phylogenetic analyses of LOV domain containing transcripts

A two-step approach was taken to identify LOV domain containing proteins in our reference and *ab initio* transcript sets. LOV domains are a subset of the PAS domain family, and initial survey of a number of known LOV domain proteins using InterproScan (29) suggested the PAS_9 Pfam domain (PF13426) has the highest similarity to the LOV domain. For this reason, we curated a list of transcripts harboring the PAS_9 domain

(detail of domain annotation is provided in the 'Bioinformatics analysis of metatranscriptomes' section). LOV domains have a signature motif that has a conserved cysteine at the fourth position, however, some degeneracy can exist at other positions of this domain (30). Given this fact, we further screened the amino acid sequences of the transcripts harboring the PAS_9 domain for the presence of previously-identified LOV specific motifs (30). We constructed a maximum likelihood phylogenetic tree from only the regions of the proteins that aligned to the PAS_9 HMM. Alignment was performed using MUSCLE (31). The tree was constructed in PhyML (32) with aLRT-SH like node support. The tree and the heatmap of the expression profile were visualized in the interactive Tree of Life (33).

Analyses of circular "time of day" data

The R package "circular" (34) was used to conduct circular statistics on time-of-day data with a 0-24 hour range. This includes peak time of day comparisons, calculating the mean peak time of expression of a group of periodic ORFs, and statistics on the photosynthetic cascade (Figure 5): Watson-Wheeler Test of homogeneity of means, Watson-Williams Test of homogeneity of means (35).

Figures

Sorting and plotting of data was conducted in R version 3.2.1 (2015-06-18) using the following packages: plyr (36), dplyr (37), reshape2 (38), ggplot2 (39), lubridate (40), ggmap (41), gridExtra (42).

Bioinformatic analysis of amplicon data

rRNA read processing and annotation

16S and 18S rRNA 454 reads were demultiplexed using Roche/454's sfffile utility and converted from standard flowgram to fasta format using sff2fastq (https://github.com/indraniel/sff2fastq). Primer removal, quality control, trimming, dereplication, and taxonomic annotation were conducted using an in-house rRNA pipeline (https://github.com/allenlab/rRNA_pipeline). Chimeric sequences were removed using USEARCH (43), reads were trimmed to a quality score of 10 over a 2 base window, operational taxonomic units were clustered using SWARM (44) and classified using FASTA36 from the FASTA package (http://faculty.virginia.edu/wrpearson/fasta/fasta36/). Taxonomic annotations were assigned by using GLSEARCH36 (45) with the version 119 of the SILVA reference database (46) for 16S rRNA and a modified PR2 database with updates from Tara Oceans W2 (47) for 18S rRNA.

Roche 454 sequencing of large fraction amplicons across all 16 time points yielded 820,700 raw 16S rRNA reads and 970,927 raw 18S rRNA reads, of which 36.7% (301,244) and 38.7% (375,904), respectively, remained after filtration. After de-replication, the large fraction contained 8,522 unique 18S and 5,420 unique 16S reads (1,595 of which were plastid in origin). Full pipeline statistics shown in Supplementary Dataset 2.

Phylogenetic placement

rRNA amplicons were processed against rRNA reference covariance models using Infernal (48). A blastn (11) search was performed against SILVA (46) with *e*-value threshold ≤ 1E-100 to identify representative (reference) sequences to be included in the reference phylogenetic trees (eukaryotic 18S, bacterial 16S, and plastidic 16S). Reference sequences were then aligned with MAFFT (49) using the G-INS-i setting for global homology. The generated multiple sequence alignments were visually inspected, manually edited and refined using JalView (50). Maximum likelihood reference trees were inferred under the general time-reversible model with gamma-distributed rate heterogeneity using FastTree (51). Processed rRNA sequences were mapped onto the corresponding reference trees using pplacer (52) with the default settings. The number of the mapped sequences to trees nodes was normalized to the total number of mapped sequences from the corresponding samples. Normalized abundances were visualized as circles mapped onto the reference trees such that the diameters of the circles were proportion to the taxonomic abundances.

Supplementary Results and Discussion

A molecular window into biogeochemistry

Several lines of evidence indicated that during the drift cells were experiencing iron-limitation and that this factor shaped community composition. WGCNA was used to examine functional clusters related to nutrient cycling (Figure S5). The expression of several low-iron response genes, including iron-starvation-induced proteins ISIP1, ISIP2A (phytotransferrin; (53), ISIP2B, and ISIP3 (54) in diatoms and haptophytes (module 5) indicated cellular iron stress. Phytotransferrin and “silicon transporter”

annotations clustered into the same module these iron-response genes (Figure S5; module 5; green) and were dominated by centric diatom expression. Module 5 peaks sharply at the end of the drift track when the measured silica:nitrate ratio, which was initially around 1, dropped most dramatically (Figure S1B). Low silica:nitrate ratios (in the range of 0.8 to 1.1) have been observed in association with iron limitation (55) and are thought to result from silica draw-down by iron-stressed diatoms (56). In our study, this ratio dropped by an order of magnitude along the drift track. While the main feature of module 5 is its sharp peak at the end of the drift track, there also appears to be some underlying diel periodicity in the signal (upregulated more during night hours). This convolution of expression patterns speaks to the difficulty of teasing apart various physical drivers of transcription in a dynamic natural context. In the future, sampling for a longer period of time could provide the statistical resolution to address these questions more conclusively.

Additional molecular evidence supported iron-limitation, such as high expression of iron complex outer membrane receptor proteins, which are associated with the uptake of siderophores (57), especially in the small fraction. Indeed, “iron complex outer membrane receptor protein” was the 34th most highly expressed annotation across both size classes (Supplementary Data 6). Furthermore, nitrate transporters and reductases were nearly undetectable across phytoplankton lineages whereas ammonium transporters were highly expressed, reflecting a reduced capacity for nitrate assimilation, which requires iron-rich heme cofactors (54). Finally, relative levels of ferredoxin and flavodoxin among photosynthetic organisms are often used as an indicator of iron stress (58,59). Iron-intensive ferredoxin proteins can be substituted by

flavodoxin, which performs the same role in photosynthesis but uses flavin mononucleotides in place of iron-sulfur clusters. Strikingly, in the large fraction, 98.12% of expression of these ORFs was attributed to flavodoxin (Pfam PF00258) over ferredoxin (Pfam PF00111) across the five major eukaryotic phytoplankton lineages (diatoms, chlorophytes, dinoflagellates, haptophytes, and pelagophytes; Figure S5). This ratio falls at the extreme edge of the distribution of previously observed cases (60), associated with the lowest iron concentrations. While direct trace-metal clean measurement of iron concentrations was not possible due to the nature of the robotic sampling, the gene sensors, historic oceanographic context (Figure S4), and nutrient proxies we present here establish a high likelihood of iron limitation.

Timing of diel ORF expression across taxonomic groups

Periodic ORFs were only detected in Proteobacteria known to possess proteorhodopsin or carry out anoxygenic photosynthesis. However, previous observations that benefitted from a longer time course were able to detect a greater diversity of periodic ORFs in heterotrophic bacterioplankton (22), indicating that these results only capture the strongest oscillating genes. We observed only a single periodic ORF in the photoheterotrophic bacteria, SAR11 and SAR116, both peaking around 12:30p.m. In SAR11, this ORF was in the isocitrate lyase family. This points to a daytime-shifted metabolism resulting from phototrophy; Glyoxylate shunt genes have been found to be expressed 300-fold more in light than darkness in the proteorhodopsin phototroph, *Dokdonia* (61), and were up during the day in photoheterotrophic bacterioplankton in a previous drift track (22). With the exception of a *Pseudomonas*

marR family transcriptional regulator peaking around 10a.m., all other periodic proteobacterial ORFs occurred in *Rhodobacter* species and were involved in nighttime chelatase, light-independent protochlorophyllide reductase, bacteriochlorophyll synthase, amine oxidases involved in carotenoid biosynthesis, diheme cytochrome c).

Viruses in the small size class

In the small fraction, bacteriophages corresponding to several of the abundant bacterial groups were enriched (e.g. *Pelagibacter* phage, $\log_2$FC=11; *Roseobacter* phage, $\log_2$FC=7.4; *Vibrio* phage, $\log_2$FC=5.2; Supplementary Dataset 10), as was a virus best annotated as the uncultivated *Ostreococcus* OsV5 virus for which the host *Ostreococcus* clade is still unclear (15). However, despite the picoprasinophytes *Ostreococcus* ($\log_2$FC=2.25) and *Bathycoccus* ($\log_2$FC=2.36) being enriched in the small size class, the largest signal for viruses most similar to isolates known to infect them (i.e. OlV1, OtV1, and BpV1), came from the large fraction. This could indicate that the close proximity of cells in particle-associated microenvironments promotes infection, that infected cells more easily attach to particles (13), or that infected hosts are larger in size.

References

1. Ottesen EA, Young CR, Eppley JM, Ryan JP, Chavez FP, Scholin CA, et al. Pattern and synchrony of gene expression among sympatric marine microbial populations. Proc Natl Acad Sci. 2013 Feb 5;110(6):E488–97.

2. Scholin CA, Birch J, Jensen S, III RM, Massion E, Pargett D, et al. The quest to develop ecogenomic sensors: A 25-year history of the Environmental Sample Processor (ESP) as a case study. Oceanography. 2017;30(4):100–13.
3. Timothy Pennington J, Chavez FP. Seasonal fluctuations of temperature, salinity, nitrate, chlorophyll and primary production at station H3/M1 over 1989-1996 in Monterey Bay, California. Deep Res Part II Top Stud Oceanogr. 2000;47(5–6):947–73.
4. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci. 1985;82(20):6955–9.
5. Herlemann DPR, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. 2011;5(10):1571–9.
6. Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner HW, et al. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol. 2010;19(SUPPL. 1):21–31.
7. Bertrand EM, McCrow JP, Moustafa A, Zheng H, McQuaid JB, Delmont TO, et al. Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. Proc Natl Acad Sci U S A. 2015 Aug 11;112(32):9938–43.
8. Schmieder R, Lim YW, Edwards R. Identification and removal of ribosomal RNA sequences from metatranscriptomes. Bioinformatics. 2012 Feb 1;28(3):433–5.

9. Rho M, Tang H, Ye Y. FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res [Internet]. 2010 Nov 1 [cited 2016 Dec 1];38(20):e191–e191. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkq747
10. Protein BLAST: search protein databases using a protein query [Internet]. Available from: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins
11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol [Internet]. 1990 Oct [cited 2017 Jul 17];215(3):403–10. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022283605803602
12. Sonnhammer E, Eddy SR, Birney E, Bateman A, Durbin R. Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res. 1998 Jan 1;26(1):320–2.
13. Allen LZ, Allen EE, Badger JH, McCrow JP, Paulsen IT, Elbourne LD, et al. Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. ISME J. 2012;6(7):1403–14.
14. Dupont CL, Mccrow JP, Valas R, Moustafa A, Walworth N, Goodenough U, et al. Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. ISME J. 2014;doi(10).
15. Zeigler Allen L, McCrow JP, Ininbergs K, Dupont CL, Badger JH, Hoffman JM, et al. The Baltic Sea virome: diversity and transcriptional activity of DNA and RNA viruses. mSystems. 2017;2(1):e00125-16.
16. Podell S, Gaasterland T. DarkHorse: A method for genome-wide prediction of horizontal gene transfer. Genome Biol. 2007;8(2).

17. Bender SJ, Moran DM, McIlvin MR, Zheng H, McCrow JP, Badger J, et al. Colony formation in *Phaeocystis antarctica*: Connecting molecular mechanisms with iron biogeochemistry. Biogeosciences. 2018;15(16):4923–42.
18. Bayat A, Gaëta B, Ignjatovic A, Parameswaran S. Improved VCF normalization for accurate VCF comparison. Bioinformatics. 2017 Mar 16;33(7):964–70.
19. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.
20. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002;30(7):1575–84.
21. Ottesen EA, Young CR, Gifford SM, Eppley JM, Marin R, Schuster SC, et al. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. Science (80-). 2014;345(6193).
22. Aylward FO, Eppley JM, Smith JM, Chavez FP, Scholin CA, DeLong EF. Microbial community transcriptional networks are conserved in three domains at ocean basin scales. Proc Natl Acad Sci. 2015;112(17):5443–8.
23. R Development Core T e. a. m. R: a language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2011. Available from: http://www.r-project.org/
24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995;57(1):289–300.
25. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9.
26. Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression

Network Analysis. Stat Appl Genet Mol Biol. 2005;4(1).

27. Csárdi G, Nepusz T. The igraph software package for complex network research. [cited 2017 Jul 26]; Available from: http://www.necsi.edu/events/iccs6/papers/c1602a3c126ba822d0bc4293371c.pdf

28. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009 Jan 1;26(1):139–40.

29. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40.

30. Glantz ST, Carpenter EJ, Melkonian M, Gardner KH, Boyden ES, Wong GK-S, et al. Functional and topological diversity of LOV domain photoreceptors. Proc Natl Acad Sci. 2016;113(11):E1442–51.

31. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.

32. Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online - A web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 2005;33(SUPPL. 2).

33. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44(W1):W242–5.

34. Agostinelli C, Lund U. R package “circular”: Circular Statistics [Internet]. 2017. Available from: https://r-forge.r-project.org/projects/circular/

35. Tasdan F, Yeniay O. Power study of circular anova test against nonparametric

alternatives. Hacettepe J Math Stat. 2014;43(1):97–115.

36. Wickham H. The split-apply-combine strategy for data analysis. J Stat Softw. 2011;40(1).

37. Wickham H, Francois R, Henry L, Müller K. dplyr: A grammar of data manipulation. 2017.

38. Wickham H. Reshaping data with the reshape package. 2006 [cited 2017 Jul 25]; Available from: http://had.co.nz/reshape

39. Wickham H. Ggplot2 : elegant graphics for data analysis [Internet]. Springer; 2009 [cited 2017 Jul 25]. 212 p. Available from: http://www.citeulike.org/group/18896/article/6995399

40. Grolemund G, Wickham H. Dates and times made easy with lubridate. JSS J Stat Softw [Internet]. 2011 [cited 2017 Jul 25];40(3). Available from: http://www.jstatsoft.org/

41. Kahle D, Wickham H. ggmap: Spatial visualization with ggplot2. R J [Internet]. 2013;5(1):144–61. Available from: http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

42. Auguie B. Miscellaneous functions for “grid” graphics [Internet]. 2016. p. 10. Available from: https://github.com/baptiste/gridextra

43. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460–1.

44. Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ. 2015;3:e1420.

45. Pearson WR. Finding protein and nucleotide similarities with FASTA. Curr Protoc

Bioinforma. 2016;2016(March):3.9.1-3.9.25.

46. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 2013;41(D1):590–6.

47. De Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, et al. Eukaryotic plankton diversity in the sunlit ocean. Science (80-). 2015;348(6237):1261605.

48. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013 Nov;29(22):2933–5.

49. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013 Apr;30(4):772–80.

50. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009 May;25(9):1189–91.

51. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One. 2010 Mar;5(3):e9490.

52. Matsen FA, Kodner RB, Armbrust EV. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics. 2010 Oct;11:538.

53. McQuaid JB, Kustka AB, Oborník M, Horák A, McCrow JP, Karas BJ, et al. Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. Nature. 2018;555(7697):534–7.

54. Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, et al.

Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. Proc Natl Acad Sci. 2008;105(30):10438–43.

55. Hutchins DA, Bruland KW. Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime. Nature. 1998 Jun 11;393(6685):561–4.

56. Brzezinski MA, Krause JW, Bundy RM, Barbeau KA, Franks P, Goericke R, et al. Enhanced silica ballasting from iron stress sustains carbon export in a frontal zone within the California Current. J Geophys Res C Ocean. 2015;120(7):4654–69.

57. Tang K, Jiao N, Liu K, Zhang Y, Li S. Distribution and functions of tonb-dependent transporters in marine bacteria and environments: Implications for dissolved organic matter utilization. PLoS One. 2012;7(7).

58. McKay RML, La Roche J, Yakunin AF, Durnford DG, Geider RJ. Accumulation of ferredoxin and flavodoxin in a marine diatom in response to Fe. J Phycol. 1999;35(3):510–9.

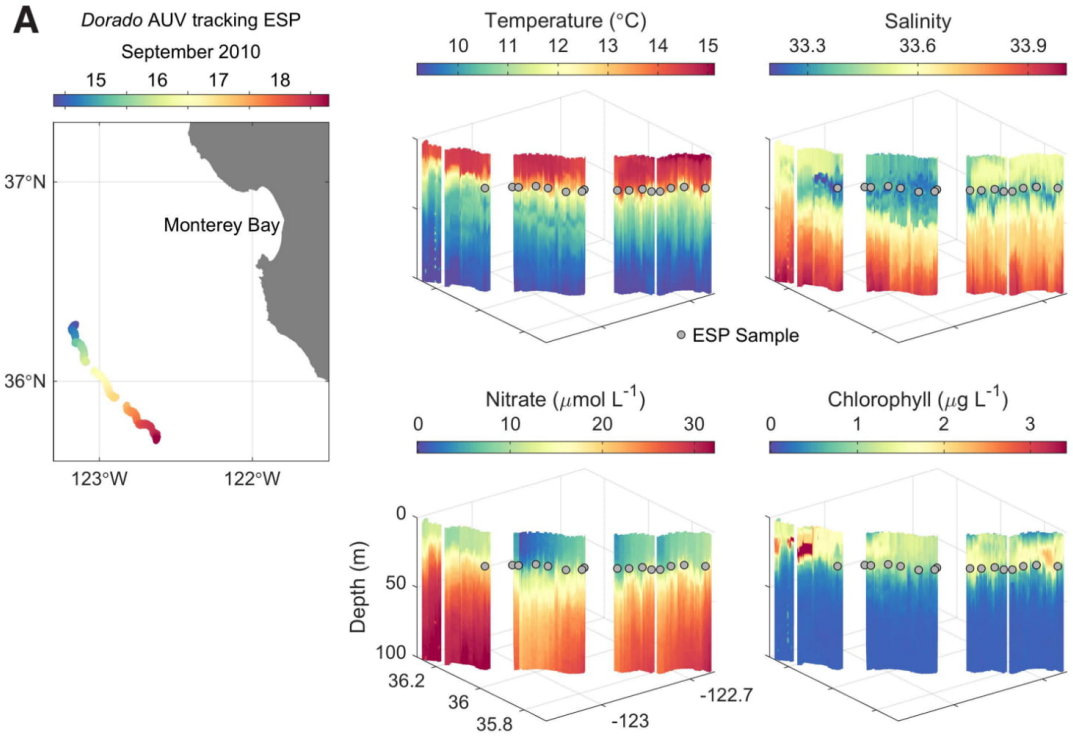
59. Erdner DL, Anderson DM. Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. Limnol Oceanogr. 1999;44(7):1609–15.

60. Carradec Q, Pelletier E, Da Silva C, Alberti A, Seeleuthner Y, Blanc-Mathieu R, et al. A global ocean atlas of eukaryotic genes. Nat Commun. 2018;9(1).

61. Palovaara J, Akram N, Baltar F, Bunse C, Forsberg J, Pedros-Alio C, et al. Stimulation of growth by proteorhodopsin phototrophy involves regulation of central metabolic pathways in marine planktonic bacteria. Proc Natl Acad Sci. 2014;111(35):E3650–8.

Supplementary File 2:

Supplementary Figures



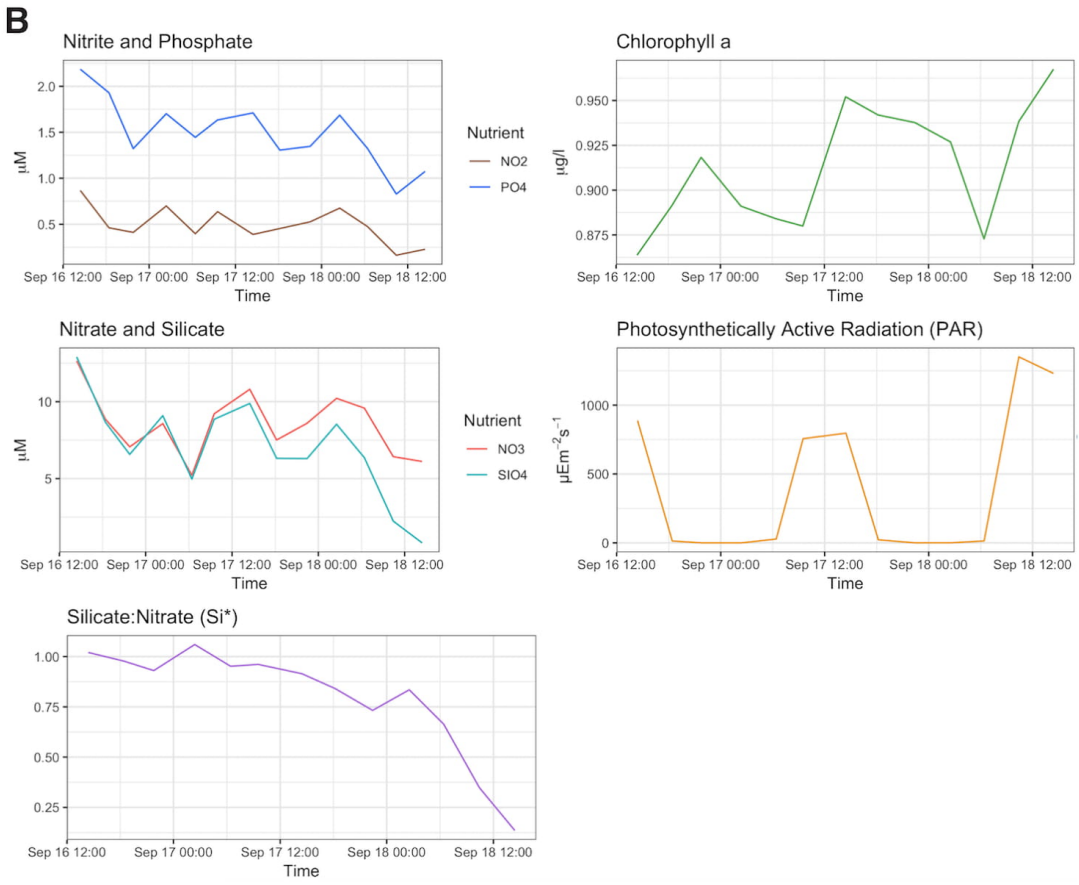


Figure S1 Biogeographical context of the drift track. **(A)** *In situ* oceanographic conditions of the drift track as observed by the *Dorado* AUV. Grey dots represent approximate locations of ESP samples. **(B)** Nutrients, chlorophyll, and light availability along the drift track at the depth of the ESP drift (~23m).

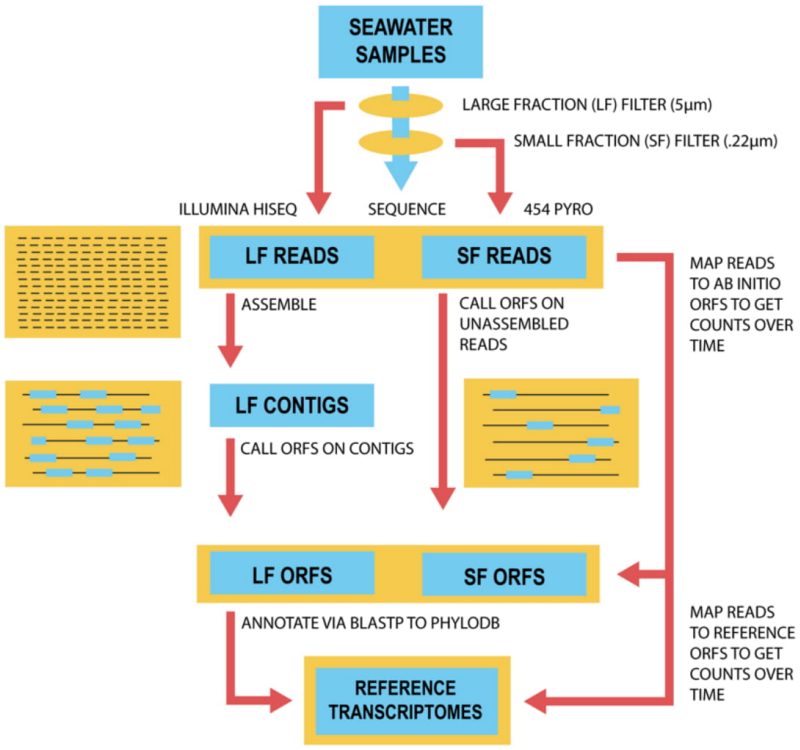


Figure S2 Bioinformatic pipeline. Seawater was collected onto 5μm and 0.22 μm filters, separating biomass into a large and small fraction, respectively. Large fraction (LF) reads were sequenced on the Illumina HiSeq platform, whereas small fraction (SF) reads had been previously sequenced by Ottesen *et al.* in 2012 on a GS FLX Titanium system (1). *Ab initio* ORF predictions were called on assembled large fraction contigs and directly on small fraction ORFs due to the longer read length, lower coverage nature of 454 sequencing. This less-restrictive amino acid space approach allowed us to map 7x more reads than traditional nucleotide space mapping to known references. Still, despite mapping 107 million reads, 158 million reads could not be mapped to *ab initio* ORFs, and those that did only averaged 67.2% identity to their best BLAST hit. Transcriptomes of reference organisms were chosen based on similarity and abundance of closely related species. Large and small fraction reads were mapped to reference transcriptomes using nucleotide Burrows Wheeler Aligner (BWA; (2). Reference transcriptomes were hierarchically clustered together with large and small fraction ORFs to gene ortholog groups. The resultant clusters, reference transcriptomes, and de-novo ORFs from both fractions were annotated taxonomically and functionally and used for downstream analyses, including pattern recognition algorithms such as Harmonic Regression Analysis (HRA) and Weighted Gene Network Correlation Analysis (WGCNA).

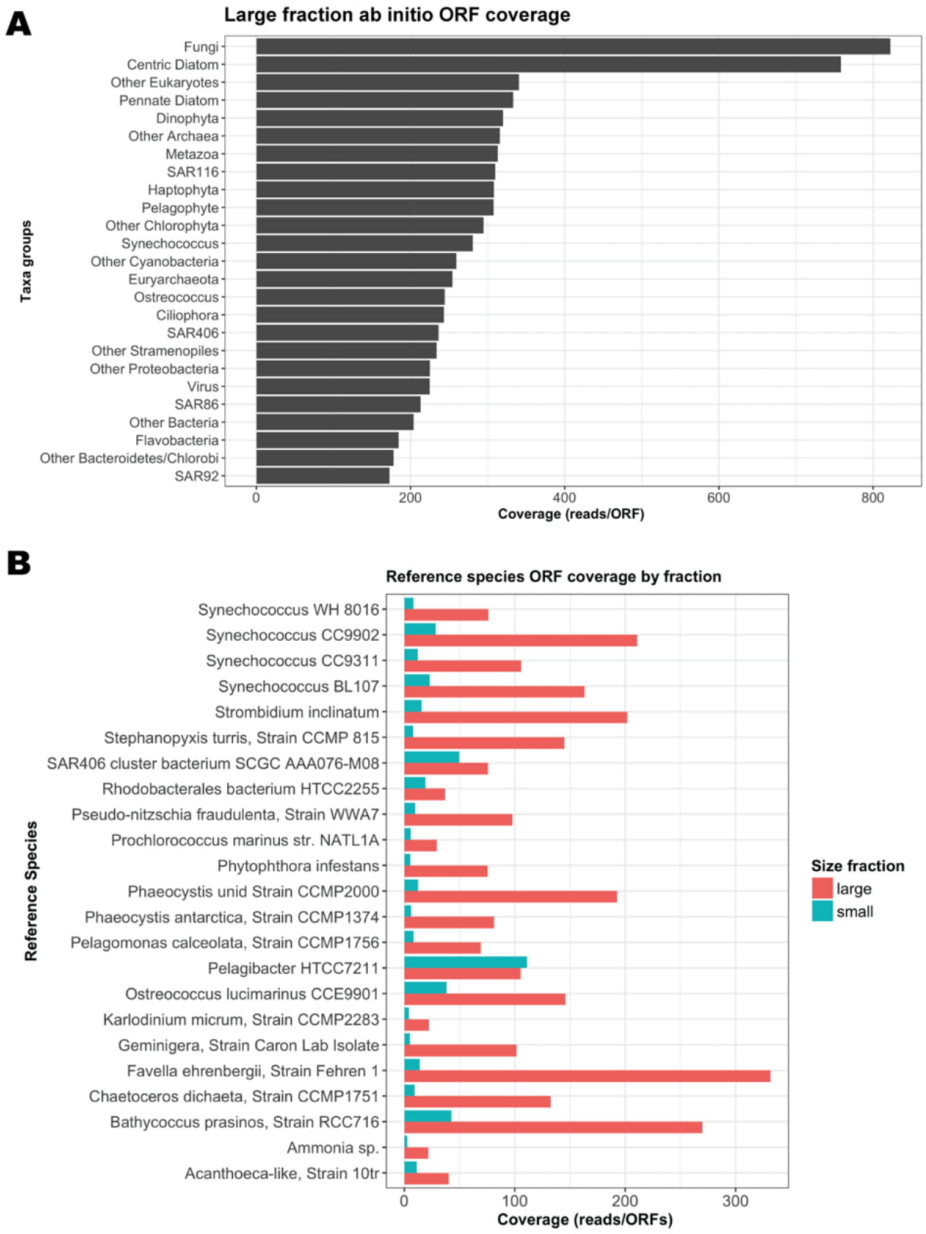


Figure S3 Depth of coverage of **(A)** large fraction *ab initio* ORFs by taxa group and **(B)** large and small fraction nucleotide references.

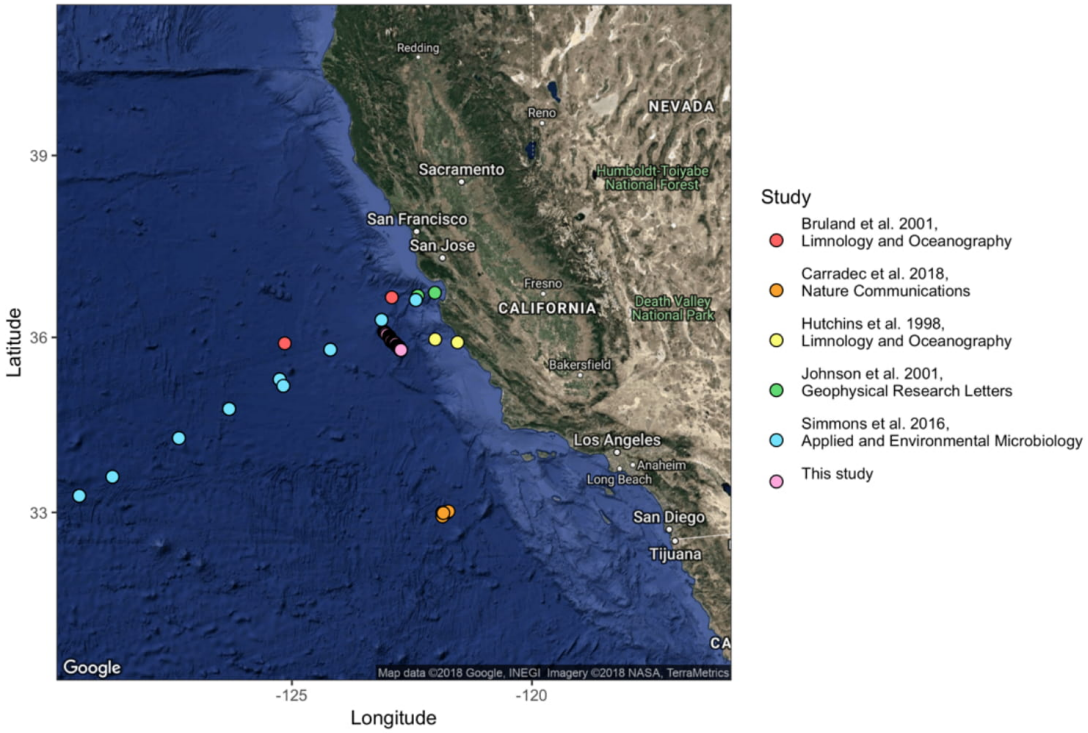


Figure S4 Location of drift track (pink) relative to sites with documented iron limitation (Supplementary Data 12E).
Red: July, 1995, measured total Fe < 0.05 nM, Fe-enrichment experiments confirm Fe limitation (3)
Orange: modeled total Fe < .04 umol/m3 based on global oceanographic data (4)
Yellow: June 1996/ June 1997, measured DPSCSV reactive Fe ≤ 0.1 nM, total Fe ≤ 0.1 nM, Fe-enrichment experiments confirm moderate to severe Fe limitation (5)
Green: Monterey Bay moorings M1 and M2, seasonal Fe limitation documented (e.g. June 1999/August 1999, measured total Fe < 1 nM) (6)
Blue: September-October 2009, total Fe < 1nM for at least 1 depth at given coordinates (7)

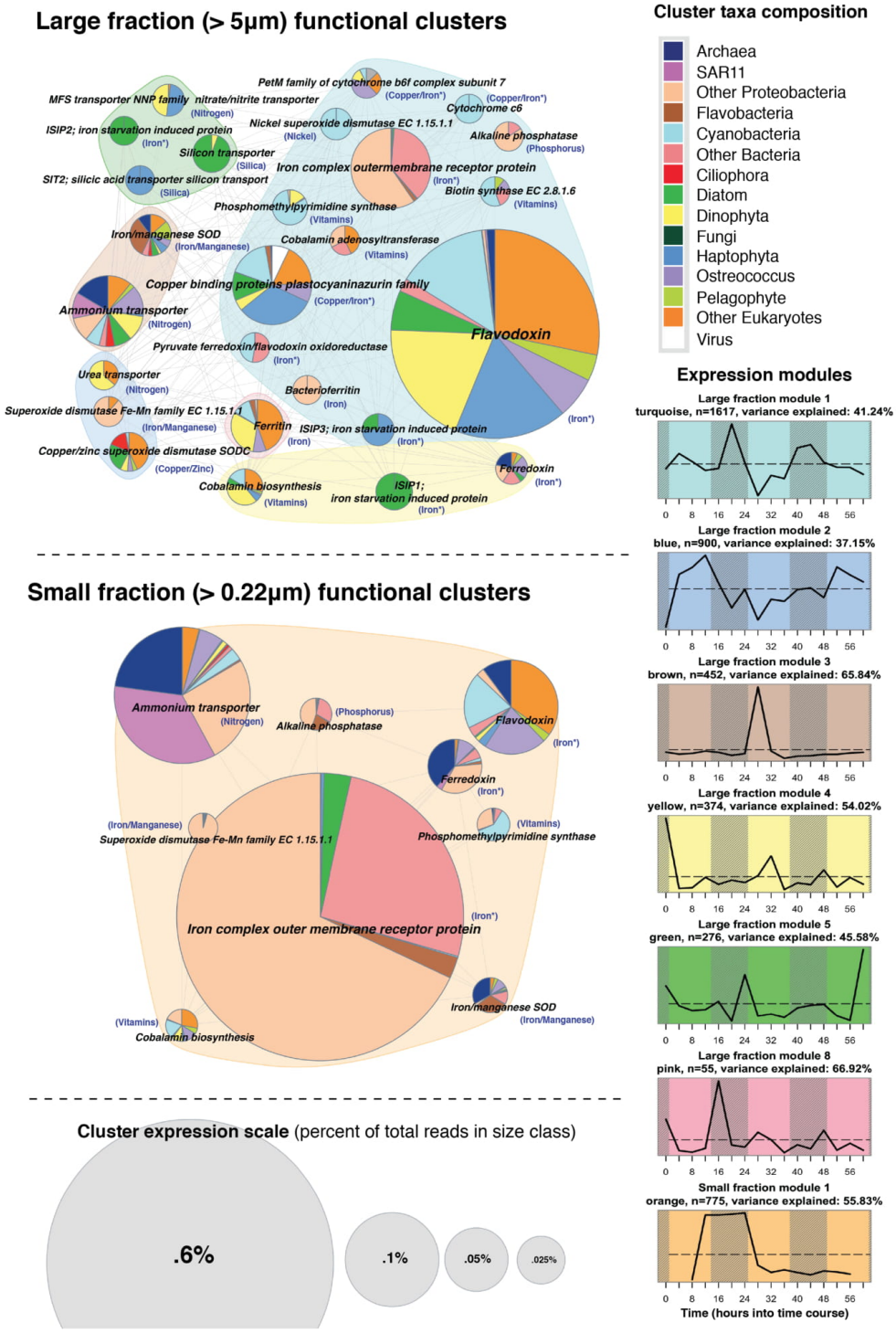

Large fraction (> 5µm) functional clusters
PetM family of cytochrome b6f complex subunit 7
(Copper/Iron*)
MFS transporter NNP family
nitrate/nitrite transporter
(Nitrogen)
(Copper/Iron*)
Cytochrome c6
ISIP2; iron starvation induced protein
(Iron*)
Nickel superoxide dismutase EC 1.15.1.1
(Nickel)
Alkaline phosphatase
(Phosphorus)
Silicon transporter
(Silica)
Iron complex outermembrane receptor protein
(Iron*)
SIT2; silicic acid transporter silicon transport
(Silica)
Biotin synthase EC 2.8.1.6
(Vitamins)
Phosphomethylpyrimidine synthase
(Vitamins)
Iron/manganese SOD
(Iron/Manganese)
Cobalamin adenosyltransferase
(Vitamins)
Copper binding proteins plastocyaninazurin family
(Copper/Iron*)
Ammonium transporter
(Nitrogen)
Flavodoxin
Pyruvate ferredoxin/flavodoxin oxidoreductase
(Iron*)
Urea transporter
(Nitrogen)
Bacterioferritin
(Iron)
Superoxide dismutase Fe-Mn family EC 1.15.1.1
(Iron/Manganese)
Ferritin
(Iron)
ISIP3; iron starvation induced protein
(Iron*)
(Iron*)
Copper/zinc superoxide dismutase SODC
(Copper/Zinc)
Ferredoxin
(Iron*)
Cobalamin biosynthesis
(Vitamins)
ISIP1;
iron starvation induced protein
(Iron*)
Small fraction (> 0.22µm) functional clusters
Ammonium transporter
(Nitrogen)
(Phosphorus)
Alkaline phosphatase
Flavodoxin
(Iron*)
Ferredoxin
(Iron*)
(Iron/Manganese)
Superoxide dismutase Fe-Mn family EC 1.15.1.1
(Vitamins)
Phosphomethylpyrimidine synthase
(Iron*)
Iron complex outer membrane receptor protein
Iron/manganese SOD
(Iron/Manganese)
(Vitamins)
Cobalamin biosynthesis
Cluster expression scale (percent of total reads in size class)
.6%
.1%
.05%
.025%
Cluster taxa composition
Archaea
SAR11
Other Proteobacteria
Flavobacteria
Cyanobacteria
Other Bacteria
Ciliophora
Diatom
Dinophyta
Fungi
Haptophyta
Ostreococcus
Pelagophyte
Other Eukaryotes
Virus
Expression modules
Large fraction module 1
turquoise, n=1617, variance explained: 41.24%
Large fraction module 2
blue, n=900, variance explained: 37.15%
Large fraction module 3
brown, n=452, variance explained: 65.84%
Large fraction module 4
yellow, n=374, variance explained: 54.02%
Large fraction module 5
green, n=276, variance explained: 45.58%
Large fraction module 8
pink, n=55, variance explained: 66.92%
Small fraction module 1
orange, n=775, variance explained: 55.83%
0 8 16 24 32 40 48 56
Time (hours into time course)

Figure S5 Expression of major nutrient cycling genes across size classes. Pies represent annotated functional clusters of *ab initio* ORFs and are colored by relative taxonomic contribution. The biogeochemical pathway each cluster participates in is noted in blue; asterisks denote ORFs previously observed to be transcriptionally sensitive to iron limitation. Clusters are grouped by modules of similar expression as given by WGCNA.

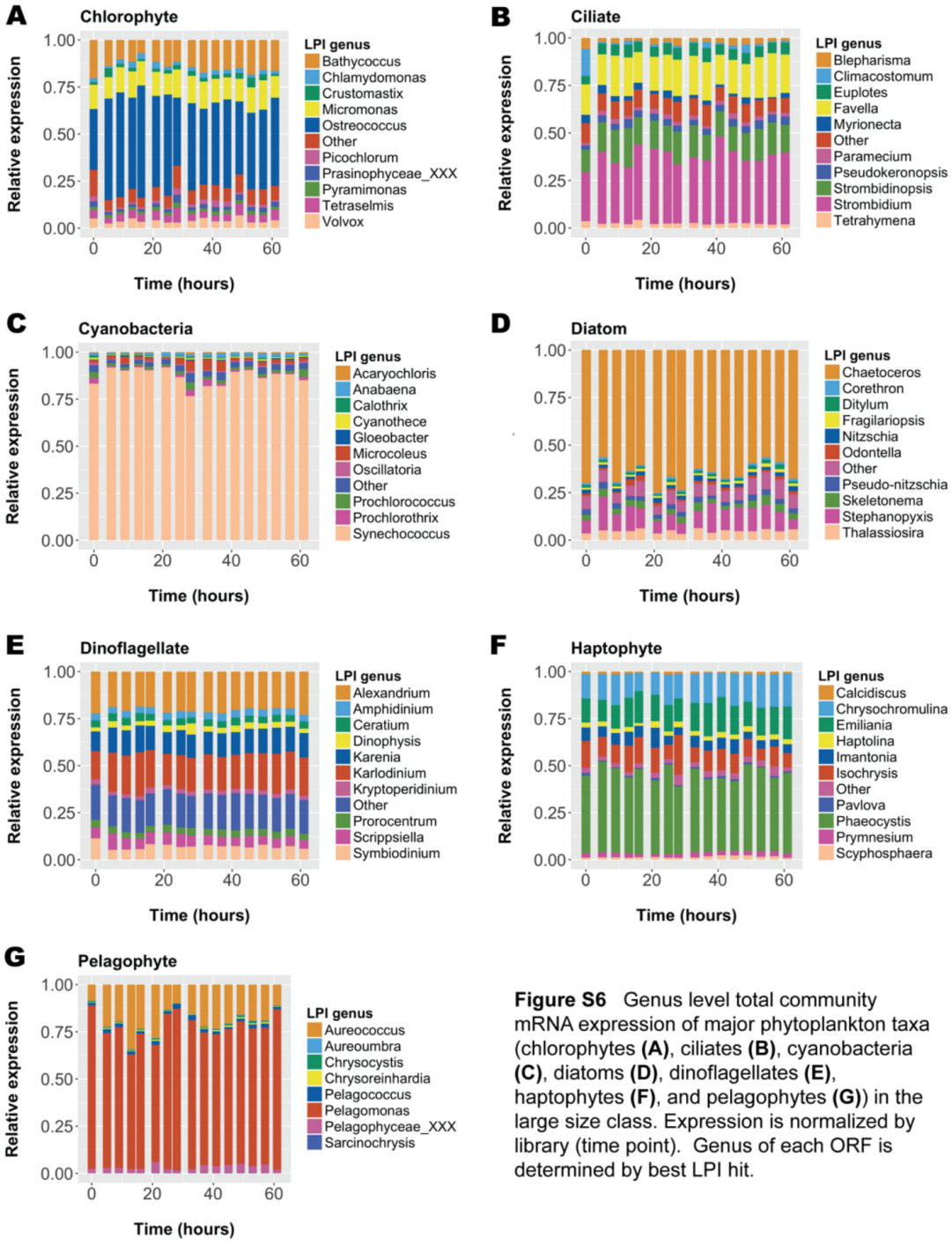


Figure S6 Genus level total community mRNA expression of major phytoplankton taxa (chlorophytes **(A)**, ciliates **(B)**, cyanobacteria **(C)**, diatoms **(D)**, dinoflagellates **(E)**, haptophytes **(F)**, and pelagophytes **(G)**) in the large size class. Expression is normalized by library (time point). Genus of each ORF is determined by best LPI hit.

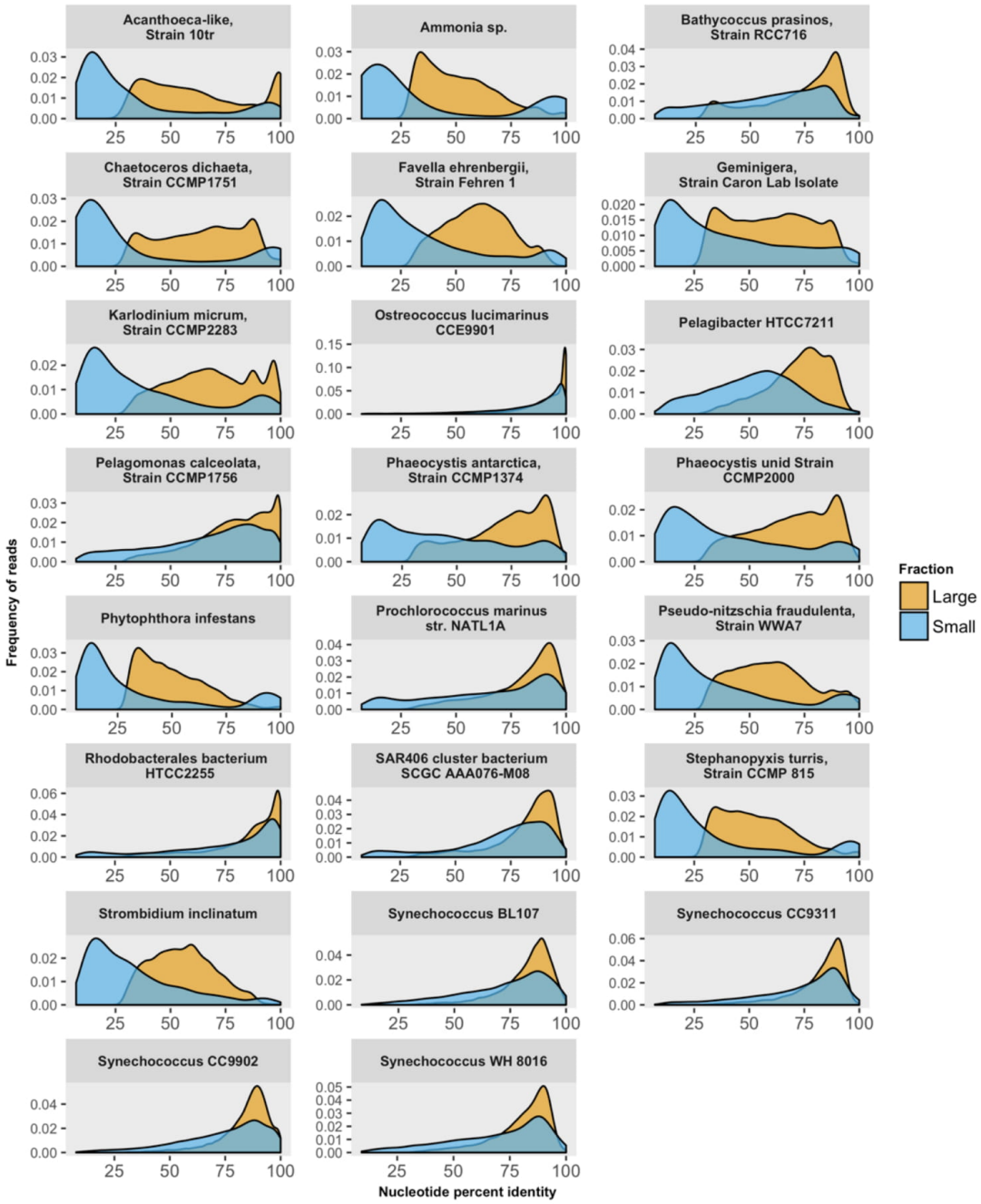


Figure S7 Average nucleotide percent identity of large fraction reads mapping to reference transcriptomes in the large (orange) and small (blue) size classes.

Figure S8 Comparison of average percent identity of reads mapping to reference transcriptomes in nucleotide space (red) and reads mapping to *ab initio* ORFs in amino acid space (blue).

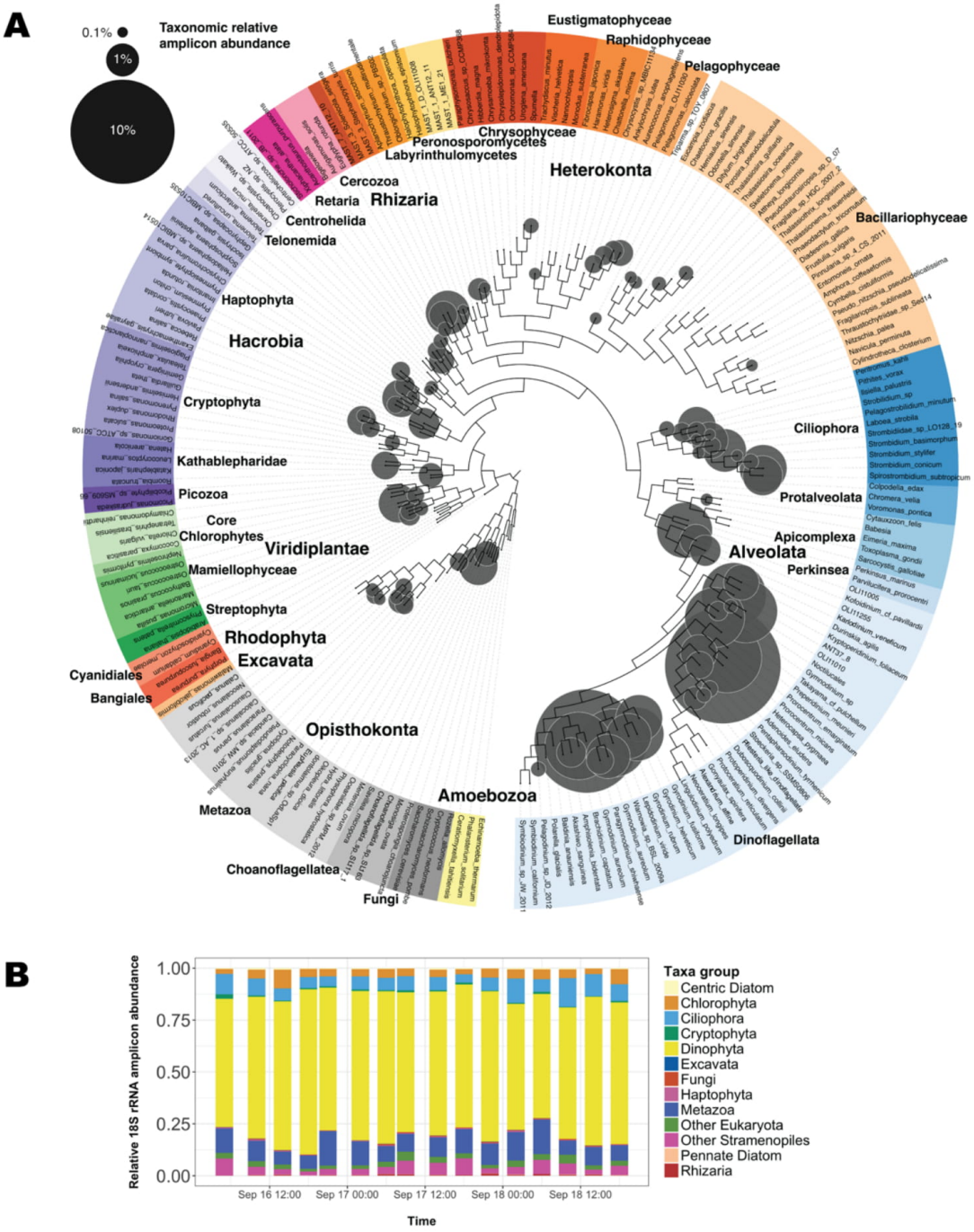


Figure S9 (A) Phylogenetic tree showing distribution of active large fraction eukaryotes using 18S rRNA amplicons (Supplementary Data 3). Circles representing relative amplicon abundance are superimposed over a reference phylogeny which is colored by taxonomy. Proximity of circles to the

tips of branches represents closeness to references. **(B)** Relative 18S rRNA amplicon abundance over time.

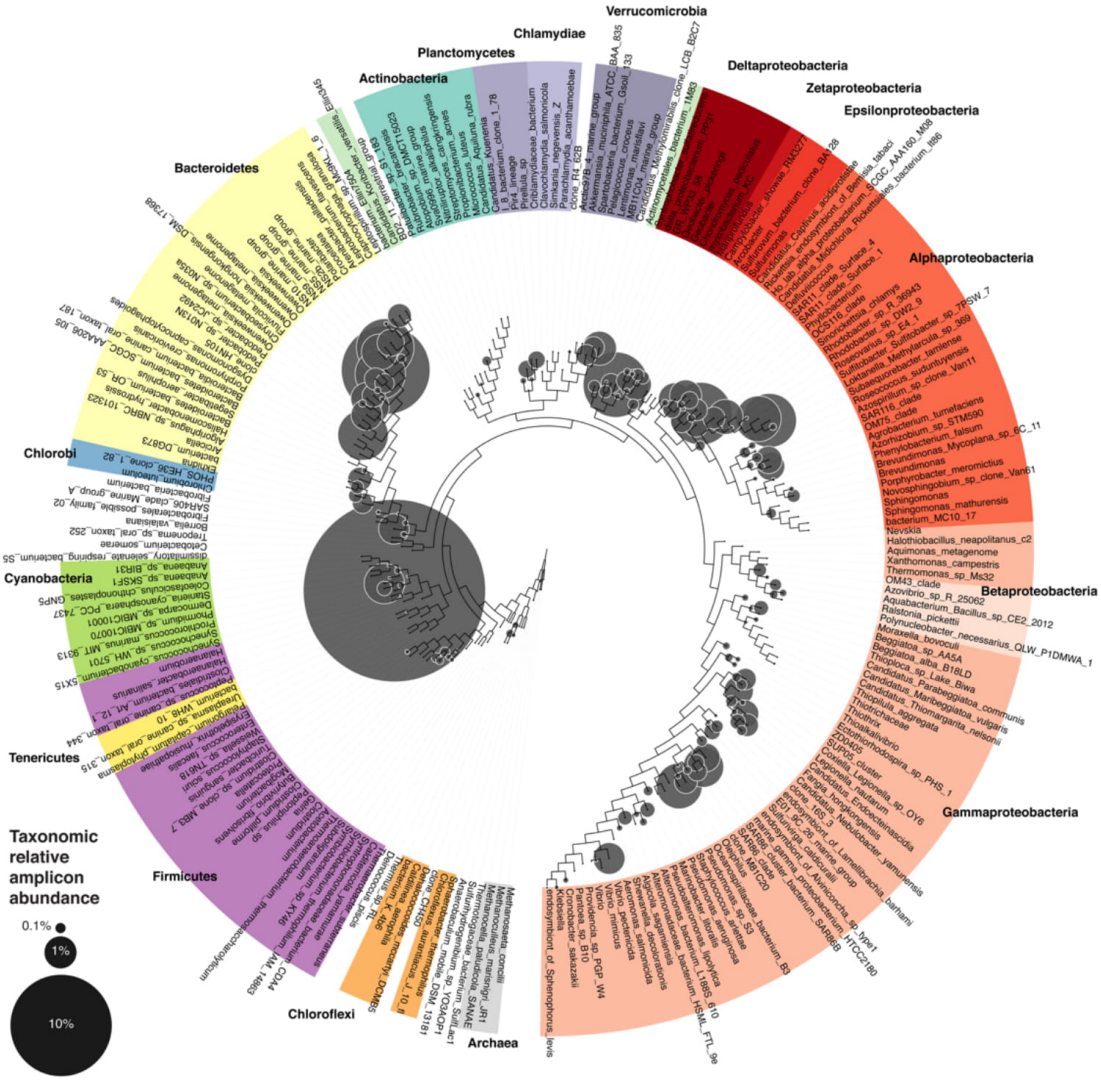


Figure S10 Phylogenetic tree showing distribution of active large fraction bacterial taxa using 16S rRNA amplicons (Supplementary Data 4). Circles representing relative amplicon abundance are superimposed over a reference phylogeny which is colored by taxonomy. Proximity of circles to the tips of branches represents closeness to references.

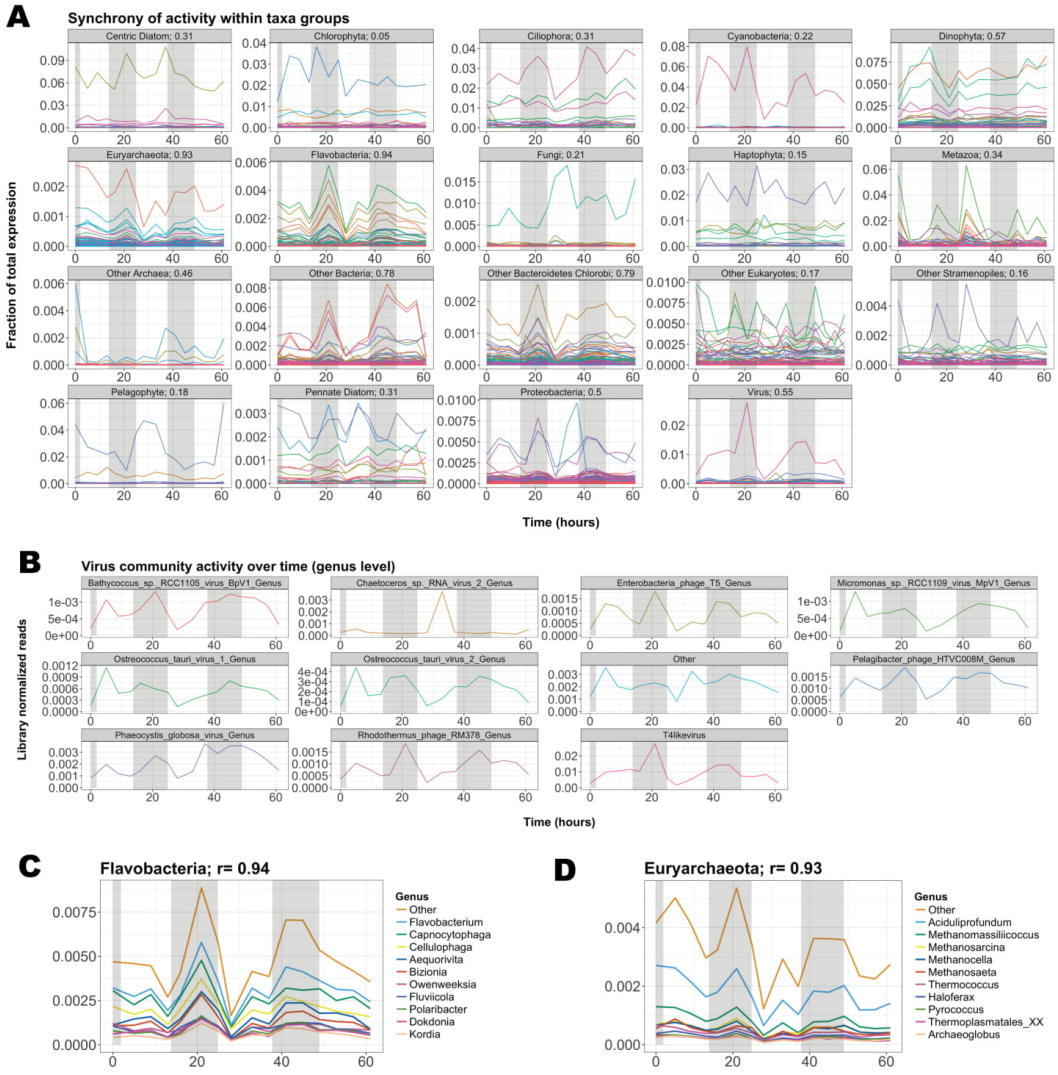


Figure S11 Synchronization of total activity among related organisms in the large fraction as viewed using *ab initio* ORFs. **(A)** Library (time point) normalized expression of *ab initio* ORFs binned by LPI-based taxonomic group. Numbers in headers denote strength of correlation between ORFs in a shared taxa group (Pearson's r). **(B)** Library normalized expression of top 10 large fraction virus genera. **(C)** and **(D)** show genus-level contributions of highly synchronous *Flavobacteria* and *Euryarchaeota* groups, respectively.

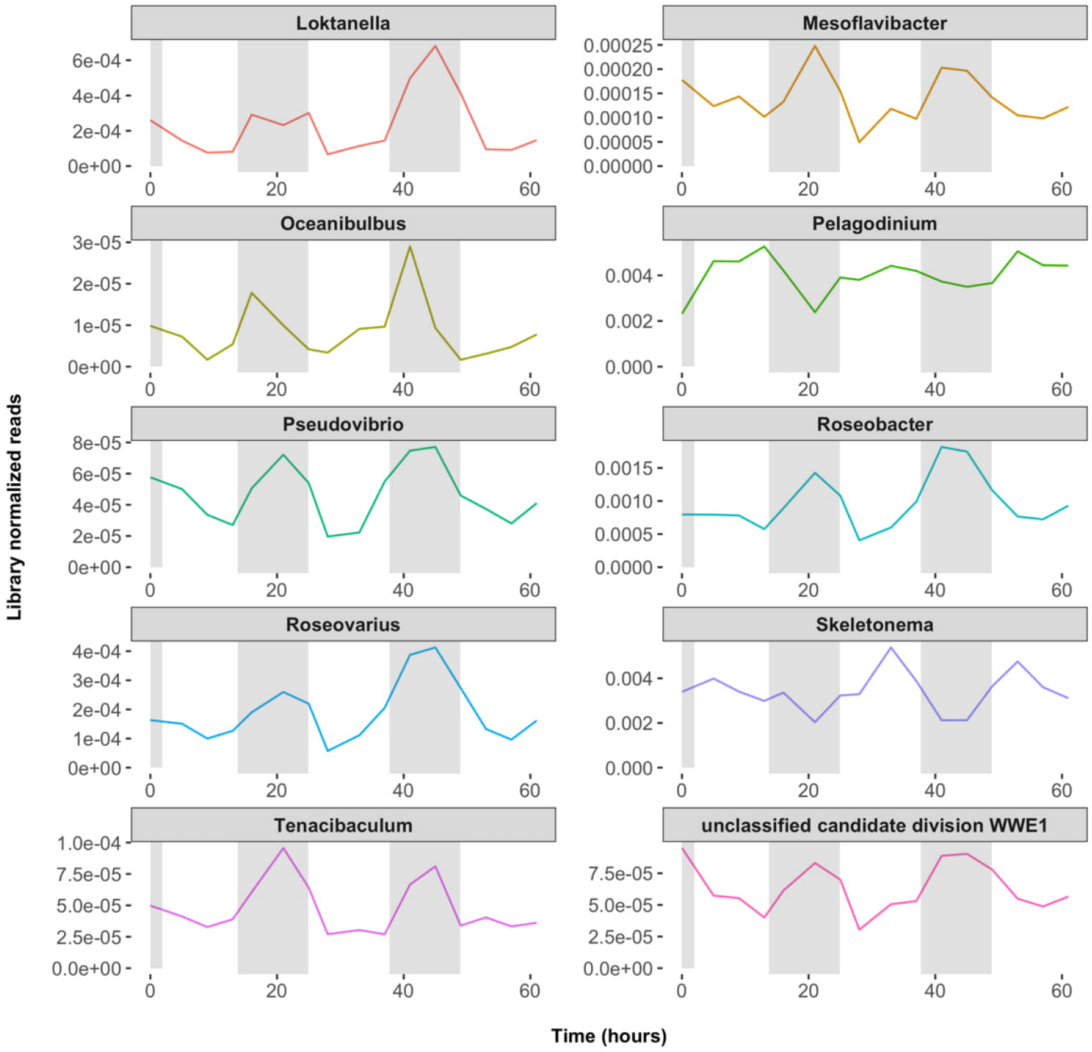


Figure S12 Total library (time point) normalized activity of large fraction genera that exhibit significant 24-h periodicity (HRA; FDR p ≤ 0.1). Two photosynthetic eukaryotes, *Pelagodinium*, a photosynthetic dinoflagellate symbiotic with foraminifera (8), and the centric diatom *Skeletonema*, had peak activity during the day. The remaining genera were non-photosynthetic bacteria with aggregate gene expression peaking at night: *Loktanella*, *Mesoflavibacter*, *Oceanibulbus*, *Pseudovibrio*, *Roseobacter*, *Roseovarius*, *Tenacibaculum*, and *Unclassified candidate division WWE1*. Several are known phytoplankton associates (e.g. *Loktanella spp.* (9,10)) and early particle colonizers (11) not previously known to operate on a diel cycle.

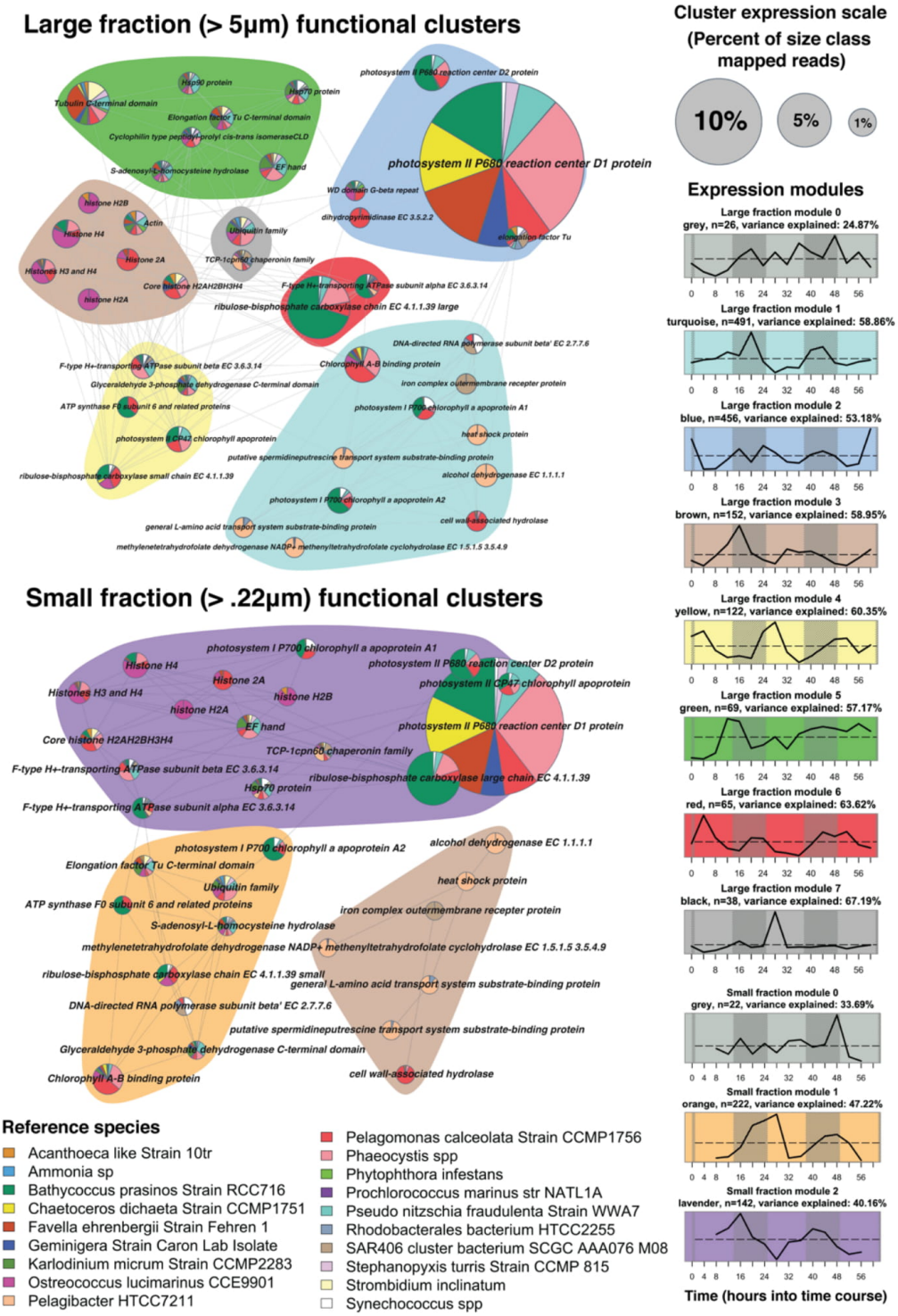
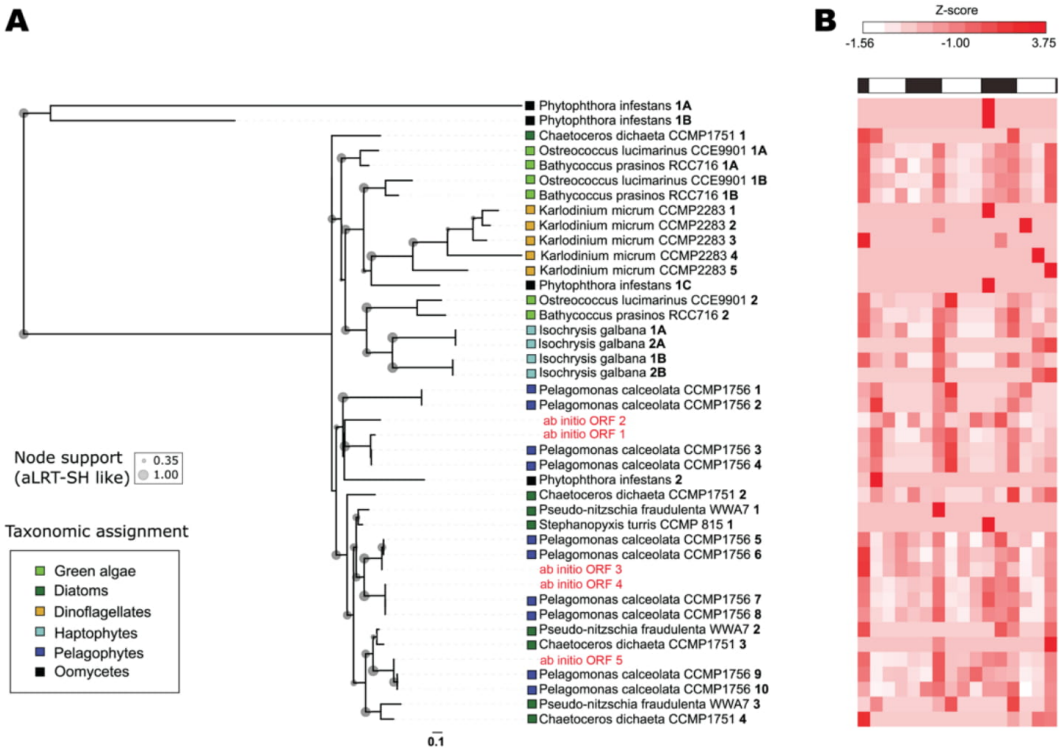

Large fraction (> 5µm) functional clusters
Hsp90 protein
Hsp70 protein
Tubulin C-terminal domain
Elongation factor Tu C-terminal domain
Cyclophilin type peptidyl-prolyl cis-trans isomeraseCLD
S-adenosyl-L-homocysteine hydrolase
EF hand
photosystem II P680 reaction center D2 protein
photosystem II P680 reaction center D1 protein
WD domain G-beta repeat
dihydropyrimidinase EC 3.5.2.2
elongation factor Tu
histone H2B
Actin
Histone H4
Ubiquitin family
Histone 2A
TCP-1cpn60 chaperonin family
Histones H3 and H4
Core histone H2AH2BH3H4
F-type H+-transporting ATPase subunit alpha EC 3.6.3.14
histone H2A
ribulose-bisphosphate carboxylase chain EC 4.1.1.39 large
DNA-directed RNA polymerase subunit beta' EC 2.7.7.6
F-type H+-transporting ATPase subunit beta EC 3.6.3.14
Chlorophyll A-B binding protein
Glyceraldehyde 3-phosphate dehydrogenase C-terminal domain
iron complex outermembrane recepter protein
ATP synthase F0 subunit 6 and related proteins
photosystem I P700 chlorophyll a apoprotein A1
heat shock protein
photosystem II CP47 chlorophyll apoprotein
putative spermidineputrescine transport system substrate-binding protein
ribulose-bisphosphate carboxylase small chain EC 4.1.1.39
alcohol dehydrogenase EC 1.1.1.1
photosystem I P700 chlorophyll a apoprotein A2
cell wall-associated hydrolase
general L-amino acid transport system substrate-binding protein
methylenetetrahydrofolate dehydrogenase NADP+ methenyltetrahydrofolate cyclohydrolase EC 1.5.1.5 3.5.4.9
Small fraction (> .22µm) functional clusters
photosystem I P700 chlorophyll a apoprotein A1
Histone H4
photosystem II P680 reaction center D2 protein
Histone 2A
photosystem II CP47 chlorophyll apoprotein
Histones H3 and H4
histone H2B
histone H2A
EF hand
photosystem II P680 reaction center D1 protein
Core histone H2AH2BH3H4
TCP-1cpn60 chaperonin family
F-type H+-transporting ATPase subunit beta EC 3.6.3.14
ribulose-bisphosphate carboxylase large chain EC 4.1.1.39
Hsp70 protein
F-type H+-transporting ATPase subunit alpha EC 3.6.3.14
photosystem I P700 chlorophyll a apoprotein A2
alcohol dehydrogenase EC 1.1.1.1
Elongation factor Tu C-terminal domain
heat shock protein
Ubiquitin family
ATP synthase F0 subunit 6 and related proteins
iron complex outermembrane recepter protein
S-adenosyl-L-homocysteine hydrolase
methylenetetrahydrofolate dehydrogenase NADP+ methenyltetrahydrofolate cyclohydrolase EC 1.5.1.5 3.5.4.9
ribulose-bisphosphate carboxylase chain EC 4.1.1.39 small
general L-amino acid transport system substrate-binding protein
DNA-directed RNA polymerase subunit beta' EC 2.7.7.6
putative spermidineputrescine transport system substrate-binding protein
Glyceraldehyde 3-phosphate dehydrogenase C-terminal domain
Chlorophyll A-B binding protein
cell wall-associated hydrolase
Reference species
Acanthoeca like Strain 10tr
Ammonia sp
Bathycoccus prasinos Strain RCC716
Chaetoceros dichaeta Strain CCMP1751
Favella ehrenbergii Strain Fehren 1
Geminigera Strain Caron Lab Isolate
Karlodinium micrum Strain CCMP2283
Ostreococcus lucimarinus CCE9901
Pelagibacter HTCC7211
Pelagomonas calceolata Strain CCMP1756
Phaeocystis spp
Phytophthora infestans
Prochlorococcus marinus str NATL1A
Pseudo nitzschia fraudulenta Strain WWA7
Rhodobacterales bacterium HTCC2255
SAR406 cluster bacterium SCGC AAA076 M08
Stephanopyxis turris Strain CCMP 815
Strombidium inclinatum
Synechococcus spp
Cluster expression scale
(Percent of size class
mapped reads)
10%
5%
1%
Expression modules
Large fraction module 0
grey, n=26, variance explained: 24.87%
Large fraction module 1
turquoise, n=491, variance explained: 58.86%
Large fraction module 2
blue, n=456, variance explained: 53.18%
Large fraction module 3
brown, n=152, variance explained: 58.95%
Large fraction module 4
yellow, n=122, variance explained: 60.35%
Large fraction module 5
green, n=69, variance explained: 57.17%
Large fraction module 6
red, n=65, variance explained: 63.62%
Large fraction module 7
black, n=38, variance explained: 67.19%
Small fraction module 0
grey, n=22, variance explained: 33.69%
Small fraction module 1
orange, n=222, variance explained: 47.22%
Small fraction module 2
lavender, n=142, variance explained: 40.16%
0 8 16 24 32 40 48 56
0 4 8 16 24 32 40 48 56
Time (hours into time course)

Figure S13 A comparison of functional diversity across fractions by mapping reads to transcriptomes of cultured representatives. Pies represent most abundant functional clusters of reference ORFs. Pies are colored by relative taxonomic contribution and grouped by modules of similar expression as given by WGCNA. Note that reads mapping to *Favella ehrengbergii* Strain Fehren 1 (e.g. those involved in photosynthesis) may be hitting remnants of its photosynthetic food source.


A
B
Z-score
-1.56
-1.00
3.75
Phytophthora infestans 1A
Phytophthora infestans 1B
Chaetoceros dichaeta CCMP1751 1
Ostreococcus lucimarinus CCE9901 1A
Bathycoccus prasinos RCC716 1A
Ostreococcus lucimarinus CCE9901 1B
Bathycoccus prasinos RCC716 1B
Karlodinium micrum CCMP2283 1
Karlodinium micrum CCMP2283 2
Karlodinium micrum CCMP2283 3
Karlodinium micrum CCMP2283 4
Karlodinium micrum CCMP2283 5
Phytophthora infestans 1C
Ostreococcus lucimarinus CCE9901 2
Bathycoccus prasinos RCC716 2
Isochrysis galbana 1A
Isochrysis galbana 2A
Isochrysis galbana 1B
Isochrysis galbana 2B
Pelagomonas calceolata CCMP1756 1
Pelagomonas calceolata CCMP1756 2
ab initio ORF 2
ab initio ORF 1
Pelagomonas calceolata CCMP1756 3
Pelagomonas calceolata CCMP1756 4
Phytophthora infestans 2
Chaetoceros dichaeta CCMP1751 2
Pseudo-nitzschia fraudulenta WWA7 1
Stephanopyxis turris CCMP 815 1
Pelagomonas calceolata CCMP1756 5
Pelagomonas calceolata CCMP1756 6
ab initio ORF 3
ab initio ORF 4
Pelagomonas calceolata CCMP1756 7
Pelagomonas calceolata CCMP1756 8
Pseudo-nitzschia fraudulenta WWA7 2
Chaetoceros dichaeta CCMP1751 3
ab initio ORF 5
Pelagomonas calceolata CCMP1756 9
Pelagomonas calceolata CCMP1756 10
Pseudo-nitzschia fraudulenta WWA7 3
Chaetoceros dichaeta CCMP1751 4
0.1
Node support
(aLRT-SH like)
0.35
1.00
Taxonomic assignment
Green algae
Diatoms
Dinoflagellates
Haptophytes
Pelagophytes
Oomycetes

C

	Transcript ID	Pfam domain ID	Domain description	Corresponding genus & species IDs on the LOV tree
Reference transcripts	CAMPEP_0199691428-CAMNT_0045556983	PF07716, PF13426	bZIP_2 (basic leucine zipper), PAS_9	Pelagomonas calceolata CCMP1756 **6**
	CAMPEP_0199699820-CAMNT_0045565857	PF13426	PAS_9	Pelagomonas calceolata CCMP1756 **4**
	CAMPEP_0199705234-CAMNT_0045571731	PF07716, PF13426	bZIP_2 (basic leucine zipper), PAS_9	Pelagomonas calceolata CCMP1756 **10**
	CAMPEP_0199709614-CAMNT_0045576641	PF07716, PF13426	bZIP_2 (basic leucine zipper), PAS_9	Pelagomonas calceolata CCMP1756 **8**
	CAMPEP_0199710938-CAMNT_0045577979	PF13426	PAS_9	Pelagomonas calceolata CCMP1756 **2**
	contig_10685_301_1296-Pelagomonas	PF13426	PAS_9	Pelagomonas calceolata CCMP1756 **3**
	contig_12486_283_1311-Pelagomonas	PF07716, PF13426	bZIP_2 (basic leucine zipper)	Pelagomonas calceolata CCMP1756 **7**
	contig_40022_3223_3810-Pelagomonas	PF13426	PAS_9	Pelagomonas calceolata CCMP1756 **9**
	contig_5472_151_1554-Pelagomonas	PF13426	PAS_9	Pelagomonas calceolata CCMP1756 **1**
	contig_8318_666_1631-Pelagomonas	PF07716, PF13426	bZIP_2 (basic leucine zipper), PAS_9	Pelagomonas calceolata CCMP1756 **5**
	GGTG_05190T0-supercont13	PF13426, PF08447, PF00320	PAS_9, PAS_3, GATA (GATA zinc finger domain)	Phytophthora infestans **1A, 1B, 1C**
	GGTG_12596T0-supercont19	PF13426	PAS_9	Phytophthora infestans **2**
	Karlodinium-micrum-CCMP2283-20140214|17716_1	PF13426	PAS_9	Karlodinium micrum, Strain CCMP2283 **3**
	Karlodinium-micrum-CCMP2283-20140214|19826_1	PF13426, PF00069	PAS_9, Protein kinase domain	Karlodinium micrum, Strain CCMP2283 **4**
	Karlodinium-micrum-CCMP2283-20140214|23659_1	PF13426	PAS_9	Karlodinium micrum, Strain CCMP2283 **2**
	Karlodinium-micrum-CCMP2283-20140214|29782_1	PF00069, PF13426	Protein kinase domain, PAS_9	Karlodinium micrum, Strain CCMP2283 **5**
	Karlodinium-micrum-CCMP2283-20140214|4888_1	PF13426	PAS_9	Karlodinium micrum, Strain CCMP2283 **1**
	MMETSP0123-20130129|11670_1	PF13426	PAS_9	Isochrysis galbana **2A, 2B**
	MMETSP0123-20130129|18638_1	PF13426	PAS_9	Isochrysis galbana **1A, 1B**
	MMETSP0794_2-20130614|17546_1	PF07716, PF13426	bZIP_2 (basic leucine zipper), PAS_9	Stephanopyxis turris CCMP 815 **1**
	MMETSP1447-20131203|31680_1	PF00170, PF13426	Basic leucine zipper (bZIP_1), PAS_9	Chaetoceros dichaeta CCMP1751 **3**
	MMETSP1447-20131203|5598_1	PF13426	PAS_9	Chaetoceros dichaeta CCMP1751 **1**
	MMETSP1447-20131203|66049_1	PF07716, PF13426	bZIP_2 (basic leucine zipper), PAS_9	Chaetoceros dichaeta CCMP1751 **4**
	MMETSP1447-20131203|9341_1	PF07716, PF13426	bZIP_2 (basic leucine zipper), PAS_9	Chaetoceros dichaeta CCMP1751 **2**
	MMETSP1460-20131121|1537_1	PF13426, PF00069	PAS_9, Protein kinase domain	Bathycoccus prasinos RCC716 **1A, 1B**
	MMETSP1460-20131121|31452_1	PF13426, PF00512, PF02518, PF00072	PAS_9, HisKA (Histidine kinase), GHKL (Gyrase-Hsp90-Histidine Kinase-MutL), Response regulator receiver domain	Bathycoccus prasinos, Strain RCC716 **2**
	OSTLU_35077-NC_009363	PF13426	PAS_9	Ostreococcus lucimarinus CCE9901 **2**
	OSTLU_40751-NC_009369	PF13426, PF00069	PAS_9, Protein kinase domain	Ostreococcus lucimarinus CCE9901 1A &1E
	Pseudo_nitzschia-fradulenta-WWA7-20140214|1617_1	PF00170, PF13426	Basic leucine zipper (bZIP_1), PAS_9	Pseudo-nitzschia fraudulenta WWA7 **3**
	Pseudo_nitzschia-fradulenta-WWA7-20140214|45782_1	PF00170, PF13426	Basic leucine zipper (bZIP_1), PAS_9	Pseudo-nitzschia fraudulenta WWA7 **2**
	Pseudo_nitzschia-fradulenta-WWA7-20140214|86850_1	PF00170, PF13426	Basic leucine zipper (bZIP_1), PAS_9	Pseudo-nitzschia fraudulenta WWA7 **1**
***ab initio* ORFs**	contig_318056_1_312_+	PF13426	PAS_9	*ab initio* ORF 1
	contig_595760_1_551_-	PF13426	PAS_9	*ab initio* ORF 2
	contig_608828_30_709_-	PF07716, PF13426	bZIP_2 (basic leucine zipper), PAS_9	*ab initio* ORF 3
	contig_620084_102_530_+	PF13426	PAS_9	*ab initio* ORF 4
	contig_492140_1_541_-	PF13426	PAS_9	*ab initio* ORF 5

Figure S14 (A) Maximum likelihood phylogenetic tree of the LOV domains from select *ab initio* ORFs and reference transcripts. Branch labels indicate the species of origin of the reference transcripts (in black). "*ab initio"* prefix refers to transcripts assembled directly from the metatranscriptomic datasets (in red). Numeric suffixes added to the labels (in bold letters) indicate the number of LOV domain transcripts present in each species. Several transcripts harbor multiple LOV domains which are denoted by an additional suffix (A, B, C). For example, one transcript from *Phytophthora infestans* harbors three LOV domains and all of these are shown on the tree. Colored squares denote the taxonomic affiliations of the reference transcripts. **(B)** Expression profile of the reference transcripts and *ab initio* ORFs along the sampling period as Z-scores. Night and day periods are denoted by dark and light bars above the heatmap. **(C)** Table indicating the transcript ID, Pfam annotation, and corresponding taxonomic information for LOV domain containing reference and *ab initio* ORFs. Several lineages of eukaryotic phytoplankton showed ORFs possessing LOV (Light-Oxygen-Voltage) domains with peak activity just before dawn. LOV domains respond to blue light (12) and well-characterized LOV domain containing proteins are known to convert photosensory stimuli into downstream biochemical signal (13) via adjacent effector domains like serine-threonine kinases (in case of phototropins) or basic leucine zipper (b-ZIP) transcription regulatory domains (in case of Aureochromes) (14). In addition, a large number of novel LOV- effector domain combinations have been previously identified across the tree of life (15). Consistent with previous observations, we found aureochrome-like domain combinations in pelagophytes (16) and diatoms (17) and a phototropin like domain combination (LOV-protein kinase) in *Ostreococcus* (18) and *Bathycoccus*. Although presence and possible function of LOV domain containing proteins have not been discussed in dinoflagellates or oomycetes, we detected LOV-protein kinase domain combinations in *Karlodinium* and a GATA zinc finger – LOV combination in oomycetes *Phytophthora*. However, the expressions of these proteins were very low and did not follow a clear diel pattern. A vast majority of the reference and *ab initio* transcripts had peak expression at dawn, irrespective of the domain combinations, indicating a common light-regulated signaling/transcriptional response mediated by the LOV domain in these organisms.

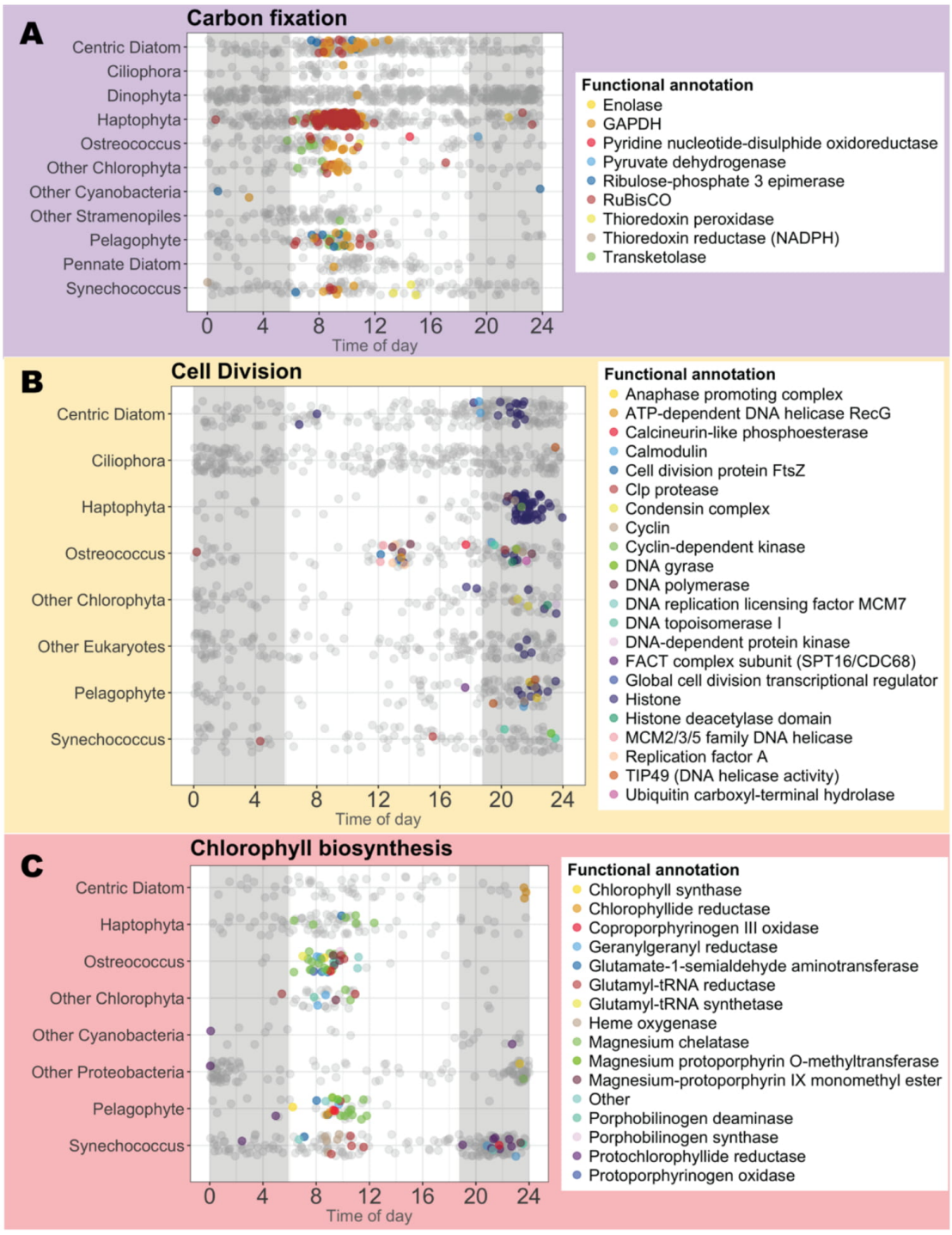

A
Carbon fixation
Centric Diatom
Ciliophora
Dinophyta
Haptophyta
Ostreococcus
Other Chlorophyta
Other Cyanobacteria
Other Stramenopiles
Pelagophyte
Pennate Diatom
Synechococcus
0 4 8 12 16 20 24
Time of day
Functional annotation
Enolase
GAPDH
Pyridine nucleotide-disulphide oxidoreductase
Pyruvate dehydrogenase
Ribulose-phosphate 3 epimerase
RuBisCO
Thioredoxin peroxidase
Thioredoxin reductase (NADPH)
Transketolase
B
Cell Division
Centric Diatom
Ciliophora
Haptophyta
Ostreococcus
Other Chlorophyta
Other Eukaryotes
Pelagophyte
Synechococcus
0 4 8 12 16 20 24
Time of day
Functional annotation
Anaphase promoting complex
ATP-dependent DNA helicase RecG
Calcineurin-like phosphoesterase
Calmodulin
Cell division protein FtsZ
Clp protease
Condensin complex
Cyclin
Cyclin-dependent kinase
DNA gyrase
DNA polymerase
DNA replication licensing factor MCM7
DNA topoisomerase I
DNA-dependent protein kinase
FACT complex subunit (SPT16/CDC68)
Global cell division transcriptional regulator
Histone
Histone deacetylase domain
MCM2/3/5 family DNA helicase
Replication factor A
TIP49 (DNA helicase activity)
Ubiquitin carboxyl-terminal hydrolase
C
Chlorophyll biosynthesis
Centric Diatom
Haptophyta
Ostreococcus
Other Chlorophyta
Other Cyanobacteria
Other Proteobacteria
Pelagophyte
Synechococcus
0 4 8 12 16 20 24
Time of day
Functional annotation
Chlorophyll synthase
Chlorophyllide reductase
Coproporphyrinogen III oxidase
Geranylgeranyl reductase
Glutamate-1-semialdehyde aminotransferase
Glutamyl-tRNA reductase
Glutamyl-tRNA synthetase
Heme oxygenase
Magnesium chelatase
Magnesium protoporphyrin O-methyltransferase
Magnesium-protoporphyrin IX monomethyl ester
Other
Porphobilinogen deaminase
Porphobilinogen synthase
Protochlorophyllide reductase
Protoporphyrinogen oxidase

Figure S15 Peak expression time of large fraction ORFs involved in **(A)** carbon fixation, **(B)** cell division, and **(C)** chlorophyll biosynthesis. Night is indicated by grey shading, while white represents daylight hours. Significantly periodic ORFs (HRA; FDR adjusted $p \leq 0.1$) are colored by functional annotation; insignificant ORFs are shown in grey.

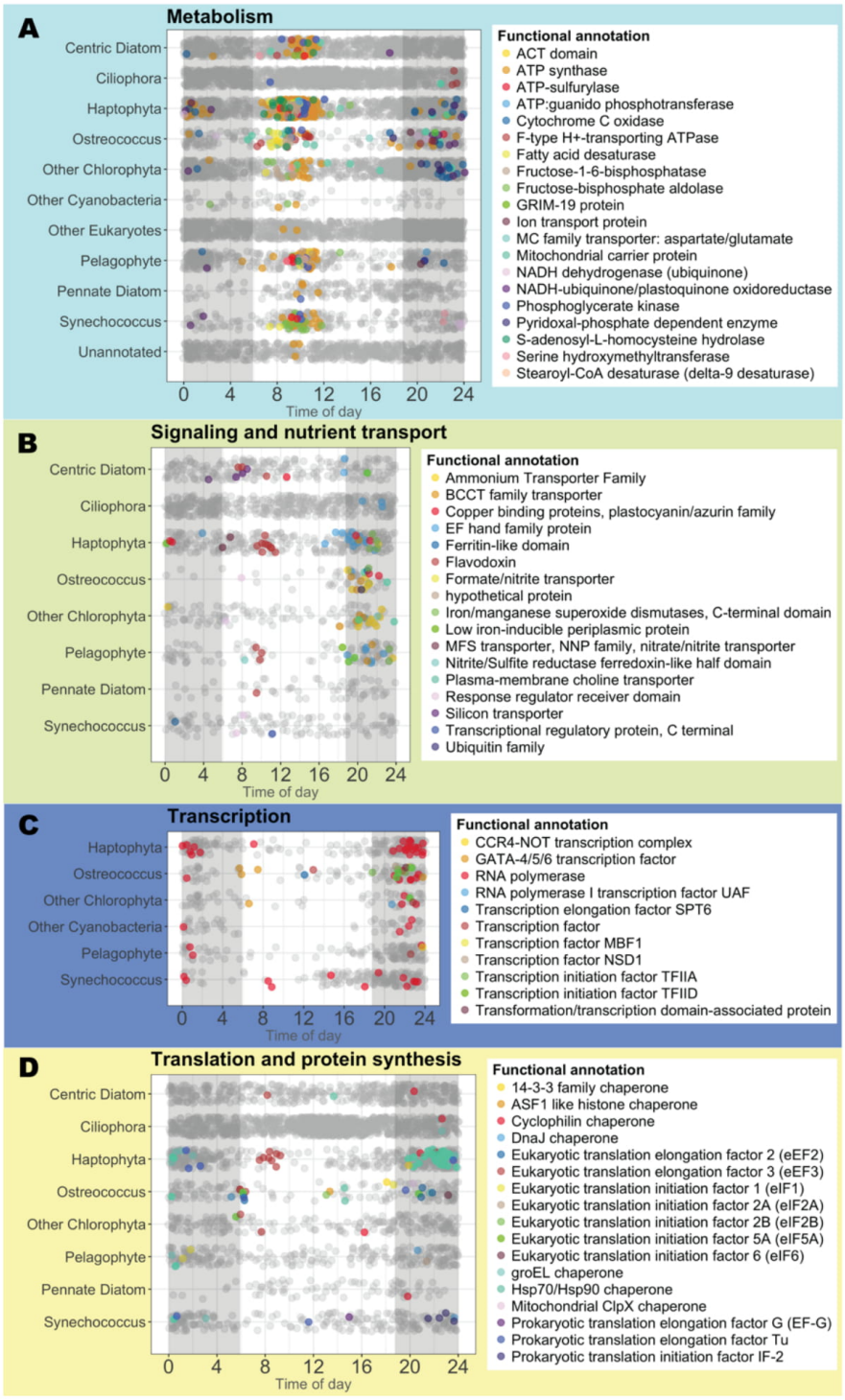

A
Metabolism
Centric Diatom
Ciliophora
Haptophyta
Ostreococcus
Other Chlorophyta
Other Cyanobacteria
Other Eukaryotes
Pelagophyte
Pennate Diatom
Synechococcus
Unannotated
0 4 8 12 16 20 24
Time of day
Functional annotation
ACT domain
ATP synthase
ATP-sulfurylase
ATP:guanido phosphotransferase
Cytochrome C oxidase
F-type H+-transporting ATPase
Fatty acid desaturase
Fructose-1-6-bisphosphatase
Fructose-bisphosphate aldolase
GRIM-19 protein
Ion transport protein
MC family transporter: aspartate/glutamate
Mitochondrial carrier protein
NADH dehydrogenase (ubiquinone)
NADH-ubiquinone/plastoquinone oxidoreductase
Phosphoglycerate kinase
Pyridoxal-phosphate dependent enzyme
S-adenosyl-L-homocysteine hydrolase
Serine hydroxymethyltransferase
Stearoyl-CoA desaturase (delta-9 desaturase)
B
Signaling and nutrient transport
Centric Diatom
Ciliophora
Haptophyta
Ostreococcus
Other Chlorophyta
Pelagophyte
Pennate Diatom
Synechococcus
0 4 8 12 16 20 24
Time of day
Functional annotation
Ammonium Transporter Family
BCCT family transporter
Copper binding proteins, plastocyanin/azurin family
EF hand family protein
Ferritin-like domain
Flavodoxin
Formate/nitrite transporter
hypothetical protein
Iron/manganese superoxide dismutases, C-terminal domain
Low iron-inducible periplasmic protein
MFS transporter, NNP family, nitrate/nitrite transporter
Nitrite/Sulfite reductase ferredoxin-like half domain
Plasma-membrane choline transporter
Response regulator receiver domain
Silicon transporter
Transcriptional regulatory protein, C terminal
Ubiquitin family
C
Transcription
Haptophyta
Ostreococcus
Other Chlorophyta
Other Cyanobacteria
Pelagophyte
Synechococcus
0 4 8 12 16 20 24
Time of day
Functional annotation
CCR4-NOT transcription complex
GATA-4/5/6 transcription factor
RNA polymerase
RNA polymerase I transcription factor UAF
Transcription elongation factor SPT6
Transcription factor
Transcription factor MBF1
Transcription factor NSD1
Transcription initiation factor TFIIA
Transcription initiation factor TFIID
Transformation/transcription domain-associated protein
D
Translation and protein synthesis
Centric Diatom
Ciliophora
Haptophyta
Ostreococcus
Other Chlorophyta
Pelagophyte
Pennate Diatom
Synechococcus
0 4 8 12 16 20 24
Time of day
Functional annotation
14-3-3 family chaperone
ASF1 like histone chaperone
Cyclophilin chaperone
DnaJ chaperone
Eukaryotic translation elongation factor 2 (eEF2)
Eukaryotic translation elongation factor 3 (eEF3)
Eukaryotic translation initiation factor 1 (eIF1)
Eukaryotic translation initiation factor 2A (eIF2A)
Eukaryotic translation initiation factor 2B (eIF2B)
Eukaryotic translation initiation factor 5A (eIF5A)
Eukaryotic translation initiation factor 6 (eIF6)
groEL chaperone
Hsp70/Hsp90 chaperone
Mitochondrial ClpX chaperone
Prokaryotic translation elongation factor G (EF-G)
Prokaryotic translation elongation factor Tu
Prokaryotic translation initiation factor IF-2

Figure S16 Peak expression time of large fraction ORFs involved in **(A)** metabolism, **(B)** signaling and nutrient transport, **(C)** transcription and **(D)** translation and protein synthesis. Night is indicated by grey shading. Significantly periodic ORFs (HRA; FDR adjusted $p \leq 0.1$) are colored by functional annotation; insignificant ORFs are shown in grey. Several eukaryotic translation elongation factor 3 (eEF3) *ab initio* ORFs detected in the large size class were significantly periodic (dark red). eEF3 presents a novel peptide synthesis mechanism for phytoplankton. eEF3 was previously thought to be unique to fungi, but homologs have been recently discovered in various phytoplankton lineages, and one haptophyte (*Phytopthora infestans*) eEF3 was proven capable of restoring function in yeast (19). Of the 122 eEF3 ORFs in our data, the majority belonged to dinoflagellates, but several were also found among haptophytes (9 ORFs), chlorophytes (9), centric (7) and pennate (3) diatoms, pelagophytes (4), other stramenopiles (1) and even ciliates (1).

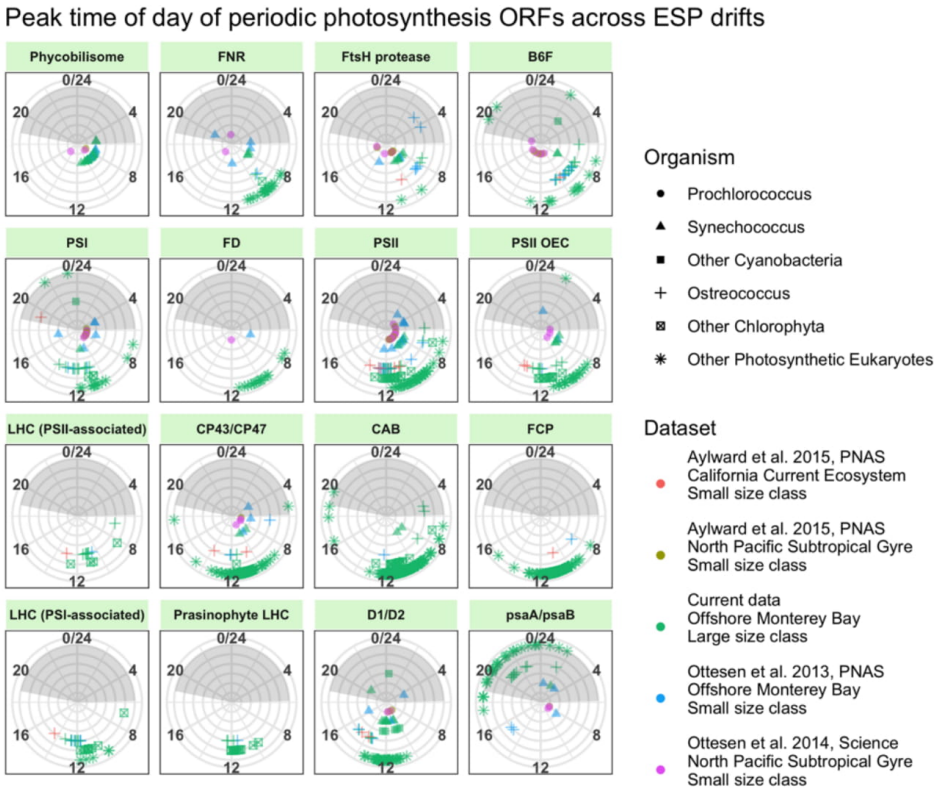


Figure S17 Comparison of peak expression time of photosynthesis related ORFs between the current data and previously studied ESP drift tracks (1,20,21). Taxa groups are distinguished by shape (legend, top right) and radius (innermost: *Prochlorococcus*, outermost: "Other Photosynthetic Eukaryotes"). Colors indicate dataset of origin. Night (as observed for current data) is indicated by grey shading. In some cases, addition of picoplankton data from other environments revealed a difference in timing between prokaryotic and eukaryotic photosynthetic proteins. For example, cyanobacterial PSII, CP43/CP47, and PSII OEC peak earlier than their equivalents in photosynthetic eukaryotes, with *Ostreococcus* and other chlorophytes peaking last, and cyanobacterial FtsH and B6F peak earlier than equivalents in photosynthetic eukaryotes.

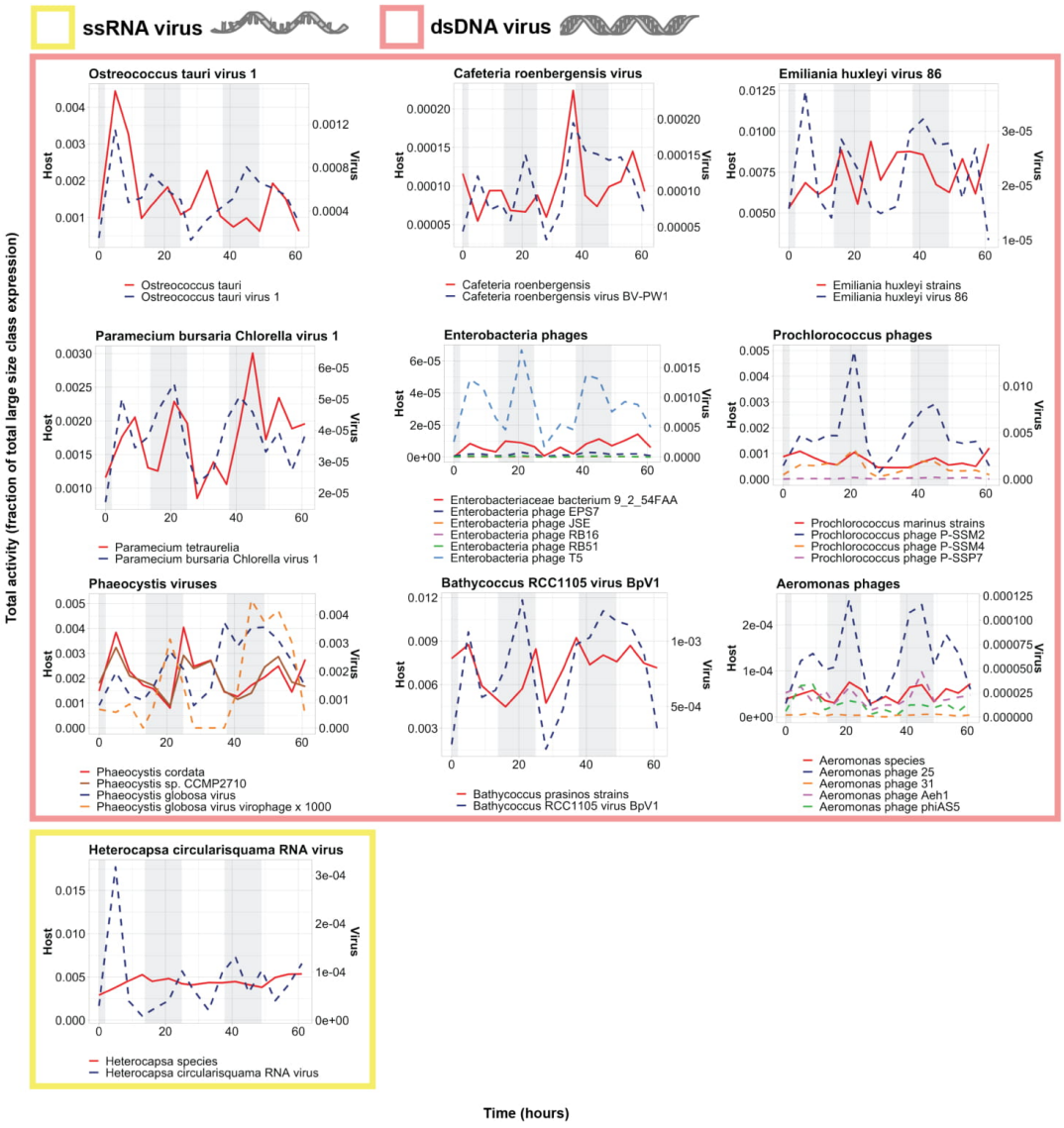


Figure S18 Continuation of Figure 6: virus/host dynamics in the large size class. Viruses and hosts are annotated as the closest reference available in our database, as determined by LPI. Library normalized expression of ORFs classified as ssRNA (yellow) and dsDNA (pink) viruses and their putative hosts by LPI are shown. Putative host expression is represented by solid lines and corresponds to left y-axes; virus expression is represented by dashed lines and corresponds to the right y-axes. *Phaeocystis globulosa* virus virophage expression was multiplied by 10^3 for better visualization. Night hours are shaded in grey.

References

1. Ottesen EA, Young CR, Eppley JM, Ryan JP, Chavez FP, Scholin CA, et al. Pattern and synchrony of gene expression among sympatric marine microbial populations. Proc Natl Acad Sci. 2013 Feb 5;110(6):E488–97.
2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.
3. Bruland KW, Rue EL, Smith GJ. Iron and macronutrients in California coastal upwelling regimes: Implications for diatom blooms. Limnol Oceanogr. 2001 Nov;46(7):1661–74.
4. Carradec Q, Pelletier E, Da Silva C, Alberti A, Seeleuthner Y, Blanc-Mathieu R, et al. A global ocean atlas of eukaryotic genes. Nat Commun. 2018;9(1).
5. Hutchins DA, DiTullio GR, Zhang Y, Bruland KW. An iron limitation mosaic in the California upwelling regime. Limnol Oceanogr. 1998;43(6):1037–54.
6. Johnson KS, Chavez FP, Elrod VA, Fitzwater SE, Pennington JT, Buck KR, et al. The annual cycle of iron and the biological response in Central California coastal waters. Geophys Res Lett. 2001;28(7):1247–50.
7. Simmons MP, Sudek S, Monier A, Limardo AJ, Jimenez V, Perle CR, et al. Abundance and biogeography of picoprasinophyte ecotypes and other phytoplankton in the eastern North Pacific Ocean. Appl Environ Microbiol. 2016;
8. Siano R, Montresor M, Probert I, Not F, de Vargas C. Pelagodinium gen. nov. and P. béii comb. nov., a dinoflagellate symbiont of planktonic foraminifera. Protist. 2010;161(3):385–99.
9. Töpel M, Pinder MIM, Johansson ON, Kourtchenko O, Godhe A, Clarke AK. Complete genome sequence of Loktanella vestfoldensis strain SMR4r, a novel strain isolated from a culture of the chain- forming diatom Skeletonema marinoi. Genome Announc. 2018;6(12):e01558-17.
10. Bloh AH, Usup G, Ahmad A. Loktanella spp. Gb03 as an algicidal bacterium, isolated from the culture of Dinoflagellate Gambierdiscus belizeanus. Vet World. 2016;9(2):142–6.
11. Pelve EA, Fontanez KM, DeLong EF. Bacterial succession on sinking particles in the ocean's interior. Front Microbiol. 2017;8(NOV):2269.
12. Christie JM, Salomon M, Nozue K, Wada M, Briggs WR. LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): Binding sites for the chromophore flavin mononucleotide. Proc Natl Acad Sci. 1999;96(15):8779–83.
13. Christie JM. Phototropin Blue-Light Receptors. Annu Rev Plant Biol. 2007;58(1):21–45.
14. Takahashi F, Yamagata D, Ishikawa M, Fukamatsu Y, Ogura Y, Kasahara M, et al. Aureochrome, a photoreceptor required for photomorphogenesis in stramenopiles. Proc Natl Acad Sci. 2007;104(49):19625–30.
15. Glantz ST, Carpenter EJ, Melkonian M, Gardner KH, Boyden ES, Wong GK-S, et al. Functional and topological diversity of LOV domain photoreceptors. Proc Natl Acad Sci. 2016;113(11):E1442–51.
16. Ishikawa M, Takahashi F, Nozaki H, Nagasato C, Motomura T, Kataoka H. Distribution and phylogeny of the blue light receptors aureochromes in eukaryotes. Planta. 2009;230(3):543–52.
17. Schellenberger Costa B, Sachse M, Jungandreas A, Bartulos CR, Gruber A, Jakob T, et al. Aureochrome 1a Is Involved in the Photoacclimation of the Diatom Phaeodactylum tricornutum. PLoS One. 2013;8(9).
18. Kianianmomeni A, Hallmann A. Algal photoreceptors: In vivo functions and potential applications. Vol. 239, Planta. 2014. p. 1–26.
19. Mateyak MK, Pupek JK, Garino AE, Knapp MC, Colmer SF, Kinzy TG, et al. Demonstration of translation elongation factor 3 activity from a non-fungal species, Phytophthora infestans. PLoS One. 2018;13(1):1–14.
20. Aylward FO, Eppley JM, Smith JM, Chavez FP, Scholin CA, DeLong EF. Microbial community transcriptional networks are conserved in three domains at ocean basin scales. Proc Natl Acad Sci. 2015;112(17):5443–8.

21. Ottesen EA, Young CR, Gifford SM, Eppley JM, Marin R, Schuster SC, et al. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. Science (80-). 2014;345(6193).

Acknowledgements

Chapter one, in full, is a reprint of the material as it appears in International Society for Microbial Ecology (ISME) Journal, 2019. Kolody, B. C., J. P. McCrow, L. Zeigler Allen, F. O. Aylward, K. M. Fontanez, A. Moustafa, M. Moniruzzaman, F. P. Chavez, C. A. Scholin, E. E. Allen, A. Z. Worden, E. F. Delong, and A. E. Allen. 2019. "Diel Transcriptional Response of a California Current Plankton Microbiome to Light, Low Iron, and Enduring Viral Infection." The dissertation author was the primary investigator and author of this paper.

CHAPTER 2: DIFFERENTIAL TRANSCRIPTIONAL RESPONSE OF DIVERSE PHYTOPLANKTON TO EXPERIMENTAL UPWELLING LIMITED BY NITROGEN, IRON, AND VIRUSES

Differential transcriptional response of diverse phytoplankton to experimental upwelling limited by nitrogen, iron, and viruses

B. C. Kolody[1,2], S. R. Smith[2], L. Zeigler Allen[2], J. P. McCrow[2], D. Shi[3], B.M. Hopkinson[4], F. M. M. Morel[5], B.B. Ward[5], A. E. Allen[1,2#]

1. Scripps Institution of Oceanography, University of California, San Diego, CA 92093
2. Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, CA 92037
3. Xiamen University, Siming District, Xiamen, Fujian, China, 361005
4. Department of Marine Sciences, University of Georgia, Athens, Georgia 30602
5. Department of Geosciences, Princeton University, Princeton, New Jersey 08544

#Corresponding: aallen@jcvi.org

Abstract:

More than half of global carbon fixation and primary production is the result of phytoplankton blooms. Here, we simulated blooms limited by nitrogen and iron by incubating Monterey Bay surface waters with sub-nutricline waters and inorganic nutrients and measured the whole-community transcriptomic response during mid- and late- bloom conditions. Cell counts revealed that centric and pennate diatoms (largely *Pseudonitzschia* and *Chaetoceros* spp.) were the major blooming taxa, but dinoflagellates, prasinophytes and prymnesiophytes also increased. Viral activity significantly increased in late-bloom conditions and likely played a role in the boom's demise. Nitrogen depletion induced conserved shifts in the genetic similarity of phytoplankton populations to cultivated strains, and the density of single nucleotide

polymorphisms (SNPs) also decreased late-bloom for diatoms and chlorophytes, indicating a selective sweep of adaptive alleles. We note conserved differences between mid- and late-bloom metabolism and differential regulation of the light harvesting complex under nutrient stress. We also identify the relative expression of NRT2/GSII as a robust marker of cellular nitrogen status, and the relative expression of iron starvation induced proteins (ISIP1, ISIP2, ISIP3) to the thiamin biosynthesis gene (ThiC) as a marker of iron status in natural diatom communities.

Introduction

The majority of phytoplankton growth in the world's oceans is either limited by either nitrogen or iron[1]. After carbon, nitrogen is the greatest component of organic material, and is essential to the structure of nucleotides and amino acids[2]. Iron, on the other hand, is a micronutrient that is crucial for redox reactions, including photosynthesis and respiration[3].

During upwelling, macronutrient-rich waters stimulate rapid phytoplankton growth, but in systems like the California coast, do not always provide enough iron for nitrogen to be utilized fully[4]. In such an environment, phytoplankton compete for not only the nitrogen necessary for growth, but also the iron necessary for photosynthesis. Common strategies for coping with low iron include using replacements for iron metalloproteins, such as flavodoxin in place of ferredoxin, downregulating the abundance of iron proteins, and remodeling cellular machinery to make use of iron co-factors for different functions at different times of day ("hot bunking")[5].

When nutrients are too dilute to promote blooms, the major phytoplankton players in coastal high-nutrient, low-chlorophyll (HNLC) regions are small phytoplankton, such as cyanobacteria, chlorophytes and haptophytes[6]. In *E. huxleyi*, nitrogen and phosphate metabolism seem to be coupled, giving haptophytes an advantage in low-macronutrient settings[7,8].

Haptophytes are also phenotypically adaptable to nutrient conditions, calcifying and taking on diploid life-stages more readily in nitrogen-replete conditions[8].

When conditions favor blooms, large, chain-forming diatoms often dominate the biomass in these systems. Diatoms often outcompete other algae for nitrate due to their integrated and sophisticated nitrogen metabolism[9], and relative to their carbon content, have higher nitrate uptake rates than dinoflagellates and chlorophytes[10]. While diatoms seem to be the most responsive to nitrogen and iron additions, they are also most susceptible to iron limitation[11,12]. Laboratory[13] and field[14] studies of the diatom transcriptional response to iron have found that iron-limited diatoms have a reduced capacity for photosynthesis and nitrogen assimilation, and that coastal diatoms have different patterns of nitrogen assimilation, carbon fixation, and vitamin production than those father offshore[15]. When iron is limiting, diatoms rely on less efficient enzymes that don't require iron as a co-factor, such as Mn SODs, as well as the uptake of reduced nitrogen species that don't require iron metalloproteins to be made bioavailable[6]. Diatoms also seem to rely on iron-free proteins for longer than haptophytes, who upregulate ferredoxin, cytochrome c6, and Fe SODs quickly upon the introduction of iron[14]. Diatoms are also quicker to partition iron towards assimilating nitrate when iron becomes available than haptophytes and chlorophytes[14].

While the nitrogen and iron responses of major phytoplankton lineages have been at least partially characterized, very little is known about population-level genetic shifts in response to blooms, and the selective pressure that nutrient limitation exerts on genes. It has been hypothesized that diversity is important in bloom formation, and that blooms will only occur if traits that fit the current environmental conditions are met by a suite of pre-existing species[16]. A genotypic profiling of consecutive *D. brightwellii* blooms off of Washington identified that

blooming populations had high diversity. Blooms occurred despite varied oceanic conditions, suggesting that they may be regulated by environmental selection[17].

Here, we describe the diversity and physiology of California Current phytoplankton communities responding to simulated blooms ending in nitrogen and iron limitation. In addition to characterizing their varied metabolic and nutrient acquisition strategies, we seek to understand the role of population shifts and selection in blooms, in which algae compete for scarce nutrients. Additionally, we examine bacterial and viral response to bloom conditions and seek to identify biomarkers of nitrogen and iron stress that can be deployed in the field.

Materials and Methods

Sample collection and incubation

Two incubation experiments (Fig. 1) were conducted using seawater collected off of Monterey Bay, California (36°50.72' N, 121°57.89' W) in September of 2008 using previously described methods to simulate phytoplankton blooms[18,19]. Seawater was incubated for six days in acid-cleaned, seawater-rinsed barrels under ambient light and temperature conditions (achieved by the use of plastic screens and seawater baths) and was mixed several times daily. Chlorophyll and other pigments, pH, nitrate, and nitrite were also monitored throughout both experiments as previously described [18].

In experiment 1, upwelling was simulated by adding 1 L of surface water (from 6 m depth) to 199 L of sub-nutricline water collected at 70 m of depth, in duplicate (barrels 1 and 3). Samples were taken at a nutrient-replete time point early in the bloom (day 3), and a nutrient-deplete time point late in the bloom (day 5) for both for cell counts (as previously described [18]) and molecular analyses.

In the second experiment, 200 L of surface water was added to barrel 2. Nutrients were immediately added to achieve final concentrations of 40µM nitrate, 2.5 µM phosphate, and 50 µM silica. Samples were taken for cell counts and molecular analyses early-bloom on day 2, after which 20 L was subsampled and incubated in barrel 4 at a final concentration of 100 nM of the iron-chelator, deferoxamine B (DFB). Both barrels 2 and 4 were sampled again in the middle of the bloom (day 4).

Metatranscriptome library preparation and sequencing

For molecular analyses, approximately 2 L of water was 0.22-µm filtered for each sample. Filters were frozen in liquid nitrogen, kept on dry ice for shipping and stored in the laboratory at -80ºC. RNA was purified from filters using the Trizol reagent (Life Technologies; Carlsbad, CA) and, treated with DNase (Qiagen, Valencia, CA, USA) and cleaned with the RNeasy MinElute Kit (Qiagen, Valencia, CA, USA). RNA quality was analyzed with on a 2100 Bioanalyzer with Agilent RNA 6000 Nano Kits (Agilent Technologies, Santa Clara, CA, USA) and quantified using Qubit Fluorometric Quantification system (TheroFisher, Waltham, MA, USA).

After setting aside 1 µg total RNA from each sample for ribosomal RNA amplicon sequencing, PolyA mRNA transcriptomes were constructed with 0.8 µg of total community RNA using TruSeq RNA Library Preparation Kit v2 (Illumina™), following the manufacture's protocol with minor adjustments. Specifically, fragmentation time was modified according to RNA quality. Library quality was analyzed on a 2100 Bioanalyzer with Agilent High Sensitivity DNA Kits (Agilent Technologies, Santa Clara, CA, USA).

The mean size of the libraries was around 400 base pairs. Resulting libraries were subjected to paired-end sequencing via Illumina HiSeq.

Amplicon library preparation and sequencing

1 µg total RNA from each sample was converted to cDNA using the SuperScript-III First Strand cDNA Synthesis System with 1 µL random hexamer primers. 16S and 18S rRNA PCR amplifications were each performed using the Life Technologies AccuPrime PCR system kit in reactions containing 1 µL of cDNA per sample as a template, 1X AccuPrime Buffer I, 0.15 µL AccuPrime Taq High Fidelity, and a final primer concentration of 200 nM. A no-template negative control for cDNA synthesis was also included. To amplify 16S rRNA, nearly-universal bacterial primers 341F (5'-CCTACGGGNGGCWGCAG-3') [20] and 926R (5'-CCGTCAATTCMTTTRAGT-3') [21] were used to target an approximately 500 bp segment the v3v5 region. To amplify 18S rRNA, TAReuk454FWD1 (5′-CCAGCASCYGCGGTAATTCC-3′) and TAReukREV3 (5′-ACTTTCGTTCTTGATYRA-3′) [22] primers were used to target an approximately 500 bp segment of the v4 region. Both primer sets were adapted for multiplexed sequencing with the addition of FLX Titanium adapters (A adapter sequence: 5′ 127 CCATCTCATCCCTGCGTGTCTCCGACTCAG 3′; B adapter sequence: 5′ 128 CCTATCCCCTGTGTGCCTTGGCAGTCTCAG 3′) and 10bp multiplex identifier (MID) barcodes. PCR cycling conditions consisted of an initial denaturation at 95°C for 2 minutes, 30 cycles of 95°C for 20 seconds, 56°C for 30 seconds, and 72°C for 5 minutes. PCR products (3 µl of each sample and 5 µl of negative control) were run on a 1% agarose gel at 110 V for 70 minutes, cleaned up using the AMPure XP bead kit (Beckman Coulter Life Sciences, Brea CA), and resuspended in 25 µL of Qiagen elution buffer (EB). The final product was visualized for quality assessment on an agarose gel (2.5 µL) and quantified using in a LifeTechnologies' PicoGreen Quant-IT assay (1 µL). Using this quantification, 20 ng of PCR product from each sample (both 16S and 18S rRNA) was pooled for 454 pyrosequencing.

Bioinformatics

Illumina reads were processed via the RNAseq Annotation Pipeline (rap) v0.4 [23] as previously described [24]. Briefly, reads were trimmed and filtered with a length minimum of 30 base pairs and a quality score minimum of 33. Ribopicker v.0.4.3 [25] was used to remove ribosomal RNA (rRNA) reads. CLC Genomics Workbench 9.5.3 (https://www.qiagenbioinformatics.com/) was used to assemble reads first by library, then overall. FragGeneScan [26] was used for ab initio ORF prediction. ORFs were screened for contamination in the form of rRNA, ITS, and primers. Organellar ORFs (those with closer homology to a known organelle gene than that of a nuclear gene in the same reference organism) were identified for separate analysis.

Ab initio ORFs were annotated for function de novo by assigning Pfams, TIGRfams and transmembrane tmHMMs with hmmer 3.0 (http://hmmer.org/; 12) using an e-value threshold of $1.0e^{-4}$ as well as assigned function and taxonomic identity via BLASTP [28,29] alignment (e-value threshold $1e^{-3}$) to a comprehensive protein database, *phyloDB*. PhyloDB includes peptides from the 410 taxa of the Marine Microbial Eukaryotic Transcriptome Sequencing Project (http://marinemicroeukaryotes.org/), as well as peptides from KEGG, GenBank, JGI, ENSEMBL, CAMERA, and various other repositories. To avoid biases introduced by taxonomically classifying ORFs based on best BLAST hit alone, a Lineage Probability Index (LPI) was calculated [30]. Briefly, LPI was calculated here as a value between 0 and 1 indicating lineage commonality among the top 95-percentile of sequences based on BLAST bit-score.

Mapping to reference transcriptomes

The top twenty-two most abundant reference transcriptome annotations for *ab initio* ORFs were chosen for read mapping. Reads were aligned to reference ORFs using BWA-MEM [31,32] using default parameters.

Hierarchical clustering

Reference transcriptome ORFs were hierarchically clustered together with *ab initio* ORFs (including organellar ORFs) from the five major phytoplankton groups in our data (centric diatoms, pennate diatoms, chlorophytes, haptophytes, and pelagophytes), to form peptide ortholog groups via the Markov Cluster Algorithm (MCL; https://micans.org/mcl/; 16). Directional edge weights were defined as the ratio of pairwise- to self- BLASTP scores, and default parameters were used to assign ORFs to clusters. Clusters were assigned a consensus annotation if found to be statistically enriched in that annotation with a Fisher's exact test ($p < 0.05$). Consensus annotations must also represent at least 10% of the reads in the cluster and account for a minimum of 200 reads.

Differential expression analysis

Differential expression of taxa groups, reference transcriptome ORFs, ab initio ORFs, and ortholog clusters across bloom conditions was identified using edgeR version 3.16.5 [34]. Read counts were normalized using the "calcNormFactors" function, which accounts for both library size and varied library composition. Differential expression of ortholog clusters was determined by using a Fisher's exact test (FDR corrected $p < 0.05$) in R to determine if each cluster was enriched in up- regulated ORFs, down- regulated ORFs or differentially expressed ORFs for a given taxa group. Manta plots of differential expression were created using the R package manta version 1.28.1.

For ab initio ORF differential expression, additional normalization strategies were implemented in order to validate the use of traditional edgeR parameters. EdgeR traditionally normalizes by distribution (TMM normalization), which assumes roughly the same number of up- and down- regulated genes across conditions. Due to the large change in biomass and

physiology inherent in a bloom experiment, we sought to validate our results by normalizing to growth biomarkers (ribosome biogenesis ORFs, PF04939 and PF09420; regulator of ribosome biogenesis, Nop53), a housekeeping gene (60S ribosomal protein L7, rpl7; KOG3184) with validated stability across metazoans [35,36], plants [37] and algae[38,39], and a gene cluster that was highly correlated with cell counts (cluster 630, R > 0.9; Fig. S1). Marker gene normalization has previously been shown to improve the accuracy of edgeR in cases when differential expression is not symmetric across conditions [40,41]. Size factors for edgeR normalization were calculated in DEseq2 verison 1.14.1 [42] by providing the function *estimateSizeFactors* with the chosen markers as *controlGenes*. The size factors were then imported into edgeR using the *norm.factors* parameter of *DGEList*. In all cases, an exact test with tagwise dispersion estimation was used to determine ORFs with significantly different expression across size classes (FDR corrected $p < 0.05$). While the number of genes classified as differentially expressed varied widely across normalization strategy, gene functions with strong differential expression were chiefly conserved across strategies, and common markers of N and Fe stress were detected across all methods.

Identification of light harvesting complex proteins

Light harvesting complex (LHC) sequences were mined from the metatranscriptomic datasets using Hidden Markov Models (HMMs). HMMs were constructed using references from our in-house database, phyloDB, of complete genomes and eukaryotic transcriptomes (phyloDB 1.075 available at https://scripps.ucsd.edu/labs/aallen/data/). Sequences were trimmed to 130 amino acids, aligned using MUSCLE, and reference phylogeny was constructed using Phyml. Placement of sequencing reads from each dataset was performed using pplacer, and the final tree was drawn using in-house software for phylogenetic placement predication visualization (available at https://github.com/mccrowjp/slacTree.git). Circles indicate the relative abundance

of sequences at node placement, abundance is scaled to the greatest value, the 'absize' (5 here) that is a multiple used for scaling each radius size. In slacTree, final abundances are calculated using (sqrt(scaled abundance) x viewwidth x absize x 0.001). Viewwidth is given during scalable vector graphic (svg) formatting and refers to the plot size.

SNP detection

Single nucleotide polymorphisms (SNPs) were detected among reads mapping to *ab initio* ORFs to a maximum read depth of 8000 using samtools mpileup with the –C50 parameter to reduce the effect of reads with excessive mismatches and –A to include anomalous read pairs, and bcftools call with the consensus calling method (-c). Libraries were subsampled to the depth of the lowest-coverage library to reduce coverage biases in SNP detection.

Results and discussion

Phytoplankton community response to nitrogen depletion

In experiment 1, 199 L of sub-nutricline water was combined with 1 L of surface water in duplicate to induce a phytoplankton bloom (Fig. 1). Indeed, total chlorophyll rose from <1 µg/L at the time of inoculation to >50 µg/L at the peak of the bloom (Fig. 1), then dropped as nitrate was drawn down (Fig. S2-S4). Meta-transcriptomic samples were taken mid-bloom on day 3 ("MB1", barrel 1; "MB3", barrel 3) and late-bloom on day 5 ("MB2", barrel 1; "MB4", barrel 3; Fig. 1).

Diatoms were the major blooming taxa, dominating in terms of cell counts (Fig. S5, Fig. S6) and total activity (Fig. 2), as well as 16S plastid (Fig. S7), and 18S amplicon (Fig. S8) counts. An increase in fucoxanthin[43] from 1.35 µg/L to 58.91 µg/L confirms a diatom bloom (Fig. S4). Pennate diatoms, dominated by *Pseudo-nitzschia* spp., were most abundant, increasing

from an average of 258,291 cells/L mid-bloom to ~1.6 million cells/L late-bloom (Fig. S5D). A nucleotide BLAST (e-value < E-100) revealed expression of *Pseudo-nitzschia* domoic acid biosynthesis (*dabA*) ORFs (contig_503530_1_1053_+, contig_613449_1_1136_-) in all conditions, indicating that cells were likely producing the neurotoxin, domoic acid. *dabA* expression was highest mid-bloom (MB1, MB3, MB5, MB7) in both experiments.

Centric diatoms also underwent a major bloom, from 125,181 cells/L to 963,182 cells/L (Fig. 2) and were dominated by *Chaetoceros*, and to a lesser extent, *Thalassiosira* spp. (Fig S7C). Dinoflagellate cells roughly doubled (from a mean of 2,691 cells/L to 5,339 cells/L), and were typically dominated by the mixotrophic athecate species, *Akashiwo sanguinea* (Fig. S5B), best known for formation of red tides[44].

Chlorophytes, cryptophytes, haptophytes, pelagophytes, fungi, and viruses were shown to be active members of the community using total mRNA (Fig2), 18S rRNA amplicons (Fig S8), and, when applicable, 16S plastid amplicons (Fig. S7). The cryptophyte-associated pigment, alloxanthin, increased in tandem with the bloom, as did the pelagophyte associated pigment, 19'-butanoyloxyfucoxanthin, and the prymnesiophyte associated pigment, 19'-hexanoyloxyfucoxanthin (Fig. S4). While prokaryotes were not visible in the poly(A)-enriched total mRNA data, 16S amplicons indicated the presence of cyanobacteria (chiefly *Synechococcus*, Fig S9), and divinyl chlorophyll a concentrations indicated that *Prochlorococcus* also bloomed (Fig S3).

Dinophyta, "other alveolate", fungi, and virus total mRNA were significantly enriched (edgeR, FDR < 0.05) in bloom conditions (MB2, MB4; Fig2). Despite cell counts showing a diatom bloom, the proportion of centric diatom total mRNA did not significantly increase late-bloom, and the proportion of pennate diatom total mRNA significantly decreased in late-bloom

conditions. This discrepancy could be due to the doubling of dinoflagellates, which have very large transcriptomes, and demonstrates the importance of collecting absolute count data when speaking to taxonomic composition.

Phytoplankton community response to iron depletion

In experiment 2, a phytoplankton bloom was induced by supplying 200L of surface water with nutrients (final concentration 40 μM nitrate, 2.5 μM phosphate, 50 μM silica) in two barrels (B2, B4; Fig 1). In one of the two barrels (B4), the iron chelator, deferoxamine B (DFB; binding constant, 10^{30}), was used to induce iron limitation. Meta-transcriptomic samples were taken mid-bloom before the addition of the iron-chelator on day 2 ("MB5", barrel 2; "MB7", barrel 4), and late-bloom on day 4 ("MB6", barrel 2; "MB8", iron-limited barrel 4; Fig. 1). As expected, the sans-DFB barrel (B2) saw total chlorophyll increase from 6.1 ug/L to 41.6 ug/L, whereas iron-limited B4 maintained chlorophyll between 15 and 17 ug/L while the health indicator, Fv/Fm, plummeted (Fig. 1, Fig. S2). In the sans-DFB barrel (B2: samples MB5, MB6), pennate and centric diatoms bloomed in a manner similar to experiment 1, although less dramatically (Fig. S5). This may be because iron concentrations in the surface water sampled were low and became limiting. In the iron-chelated barrel (B4: samples MB7, MB8), diatom cell numbers increased only modestly (Fig. S5A) and dinoflagellate numbers, which were abnormally high in MB7 (53,107 cells/L), plummeted in MB8 (10,026 cells/L; Fig. S5B).

The proportion of dinoflagellate total mRNA was also significantly less in the iron-limited condition (MB8) compared to Fe-replete conditions (MB5 and MB7; edgeR FDR < 0.05; Fig. 2). Viral transcripts showed the reverse trend, significantly increasing with iron limitation (edgeR FDR < 0.05; Fig. 2). Chlorophyte contributions to total RNA also decreased in low iron (Fig. 2), and green algal photosynthetic pigment, chlorophyll b, was less abundant in the iron-

limited barrel. Green algal photoprotective pigments increased in tandem with iron limitation (Fig. S4), indicating a stress response. This was the case for both zeaxanthin and lutein, which participate in the photoprotective non-photochemical quenching (NPQ) mechanism in chlorophytes[45]. In higher Viridiplantae lineages, photoprotective pigments (including lutein) have been observed to be disproportionately maintained[46] or even increase[47] under iron stress.

Bacterial community response to nutrient depletion

Bacterial community dynamics throughout both experiments were also observed via non-plastid 16S rRNA sequences. Overall, the community was dominated by common phycosphere lineages known to display complex relationships with phytoplankton[48], chiefly *Rhodobacteraceae*, *Flavobacteriaceae*, and a diversity of *Gammaproteobacteria* (Fig. S9, S10). Taxa known for both mutualistic and predatory relationships with phytoplankton were active. *Sulfitobacter sp. NF1-26*, a member of a lineage known for its mutualistic association with *Pseudonitzschia*, accounted for an average of 0.68% of overall sample activity. *Sulfitobacter* species have been shown to provide ammonium to *P. multiseries* in exchange for taurine, and promote *P. multiseries* cell-division via the production of indole-3-acetic acid[48]. The mutualistic Gammaproteobacterium, *Marinobacter*, was also present, and more active in the iron-limited condition ($\log_2$FC = 1.578, FDR= 0.19). *Marinobacter* produces the siderophore vibrioferrin, which it utilizes for iron uptake at night. In daylight, vibroferrin degrades, allowing phytoplankton to take up iron when they need it most[48]. The algicidal flavobacterium, *Kordia algicida*, known for its lysis of diatoms[49], was also present and was one of the species that increased in 16S activity most significantly in the late-bloom conditions (Fig. S11, S12). Notably, the SAR11 strain, *Candidatus Pelagibacter ubique* HTCC1062 [50], accounted for 2.4% of total activity and was significantly more active mid-bloom than late-bloom in nitrogen

(log_2FC = -0.84, Fig. S11) and iron (log_2FC = -1.14, Fig. S12) experiments. Cyanobacteria were dominated by *Synechococcus* sp. WH 8108, which accounted for an average of 2.5% of 16S activity, and was significantly more active in mid-bloom in both experiments (experiment 1; log_2FC = -2.17; Fig. S11; experiment 2, log_2FC = -2.17= -1.66, Fig. S12). *Prochlorococcus*, on the other hand, accounted for only about 0.02% of total activity.

In an ordination analysis, mid-bloom replicates (MB1, MB3) clustered together separately from late-bloom replicates (MB2, MB4) from experiment 1, and experiment 2 mid-bloom conditions clustered separately from MB6 (sans-DFB) and MB8 (with DFB), which were distinct (Fig. S13). This points to the overall bacterial community converging to the same structure in response to a given nutrient/bloom condition.

Selection and diversification of phytoplankton populations in response to blooms

While genus-level phytoplankton community structure (Fig. S14) did not change dramatically during the course of either bloom, fine-scale taxonomic shifts did occur. In order to examine population-level changes in community structure, reads were mapped to reference transcriptomes belonging to the top 22 most abundant species annotations assigned to ab initio ORFs (Fig. S15).

For many reference species, there was a percent identity shift in mapped reads over the course of the bloom that was conserved across replicates. A taxonomic shift towards a reference could indicate an increased abundance of winning subpopulations at the end of the bloom. Population-level shifts were much more widespread in experiment 1 (Fig. S16), where the bloom was more dramatic, than experiment 2 (Fig. S17). For the most abundant centric diatom, pennate diatom, and dinoflagellate reference, we examined the functions of ORFs in experiment 1 that were significantly differentially expressed across bloom conditions and shifted in the same

direction relative to the reference across replicates (Fig. 3). For the centric diatom, *Thalassiosira* Strain NH16, reads mapped closer to the reference over the course of the bloom. Major functions driving this shift included ORFs involved in nitrogen acquisition (nitrite reductase, glutamate synthase), metabolism (GAPDH), and protein synthesis (translation initiation factors), which could indicate the rise to dominance of a subpopulation more efficient at nutrient acquisition and rapid growth. Similarly, translation elongation factor 3, aliphatic amidase, and purine biosynthesis ORFs were upregulated and shifted towards the reference for the pennate diatom, *Pseudo-nitzschia delicatissima* Strain UNC1205. In the dinoflagellate *Prorocentrum minimum* Strain CCMP2233, a shift away from the reference was also driven by ORFs indicating metabolism and growth such as GAPDH, fumarate reductase and ribosomal proteins. This could mean that sub-populations with faster nutrient acquisition and growth were able to dominate post-bloom.

However, because of the scarcity of phytoplankton references, it is difficult to resolve whether the percent identity shifts observed here represent population-level changes or an upregulation of adaptive transcripts with varying distances to available references. To explore this further, we plotted 2D expression histograms showing closeness of reads mapping to both the given reference and the closest hit in a different genus across functional annotations. For all three references examined here, ORFs from different functional (KOG) categories have unique peaks in percent identity space that change across nitrogen condition (Fig. S18, S19, S20), indicating that function-specific changes in expression influence the percent identity shifts that we observe.

Selection and diversification of phytoplankton genes in response to blooms

In a bloom scenario, competition over fleeting nutrients presents an opportunity for selective sweeps in which the best-adapted clones dominate[17]. In order to probe selection effects over the course of both blooms, single nucleotide polymorphisms (SNPs) were called on *ab initio* ORFs. SNPs were most abundant in centric diatoms and dinoflagellates, but other alveolates, pennate diatoms, and viruses had the highest percentage of non-synonymous mutations (Fig. S21). SNP density (SNPs/total ORFs) decreased in nitrogen-deplete conditions for diatoms and chlorophytes but increased for dinoflagellates (Fig S22). A decrease in SNP density may signify that adaptive alleles are being selected for over the course of the bloom.

In dinoflagellates, the higher density of SNPs in late-bloom conditions may be due to sexual reproduction. The genetic diversity of a phytoplankton community depends on both the strength of environmental selection and the reproductive mode of cells (asexual vs sexual)[17]. Expression of the sexual reproduction marker *Sig* genes in dinoflagellates, and, to a lesser extent, diatoms, indicates that these lineages were undergoing genetic recombination. *Sig* genes were strongly upregulated during sexual reproduction in the centric diatom, *T. weissfloggii*, and are hypothesized to play a role in gamete recognition[51]. Sexual reproduction is a rare event in algae and typically takes place when nutrients are scarce[52]. Here, however, *Sig* ORFs were strongly upregulated in nitrogen and iron replete conditions, indicating that this is not always the case.

To investigate further, the function of non-synonymous alleles was compared across nitrogen (Fig. S23-S25) and iron (Fig. S27-S30) conditions. The most common allele function was light harvesting. LHC SNPs typically belonged to diatoms and chlorophytes and were more frequent (abundant relative to other alleles at the same position) in replete conditions. In deplete conditions, these LHC SNPs were outcompeted by the more abundant default nucleotide called at the SNP position.

Conversely, “fixed” alleles that had a higher frequency in late-bloom conditions in experiment 1 include thioredoxin, clathrin, a serine protease inhibitor and a carbohydrate-binding molecule (CBM) family 6 gene in centric diatoms; chlorophyte choloylglycine hydrolases; a haptophyte major facilitator superfamily transporter; a dinoflagellate arginosuccinate synthase; an alveolate ammonium transporter; a silicon transporter from an unannotated taxon; and a viral RNA-dependent RNA polymerase (RdRp; Fig. S23-S25). Alleles with higher frequency late in the bloom are likely to confer some competitive advantage. For example, in land plants, protease inhibitors are potent defense mechanisms against pathogenic bacteria and fungi[53], so the serine protease inhibitor allele fixed here could give certain diatoms a competitive advantage in fighting off invasive pathogens. Transporter alleles that are fixed in low N could confer an adaptive advantage at nutrient acquisition (e.g. ammonium, silica), and the fixed RdRp is likely associated with a successful viral phenotype.

In experiment 2, alleles fixed late in the bloom included a centric diatom PP-loop family Fe-S binding protein, a pennate diatom ADP-ribosylation factor GTPase activator (involved in vesicle transport) and dinoflagellate nitrate transporter and UDP-glucose/GDP-mannose dehydrogenase (Fig. S27-S30). Fe-S binding could confer a selective advantage for a growth-critical gene in an iron-limited system. Competitive nitrate transport could also be useful for survival in low iron, given that iron and nitrogen assimilation are linked; nitrogen assimilation relies on photosynthetic outputs and iron metalloproteins[3], and iron assimilation relies on nitrogen-dense proteins[2,54–56].

Physiological response of phytoplankton to nitrogen and iron depletion

Nitrogen and iron depletion can both cause bloom demise, but phytoplankton respond to each threat differently. In a phytoplankton cell, nitrogen limitation creates a shortage of building

material for growth (nucleotides and amino acids both require nitrogen), exerting a translational control, whereas iron limitation creates a shortage of energy (photosynthetic electron transport heavily relies on iron)[2]. Nearly a quarter of ORFs were significantly nutrient responsive. Of those, 84% were responsive to nitrogen and only 27% were responsive to iron. In low nitrogen, the majority of DE ORFs were downregulated (n= 53,958, 66% of N DE genes), presumably to slow growth in the absence of an essential macronutrient. In low iron, on the other hand, more DE ORFs were upregulated (n= 15,501, 58% of Fe DE ORFs), many of which were involved with iron scavenging. Nutrient responsiveness varied by taxa group (Fig. S31). Surprisingly, fungi and rhizaria were the most highly responsive to nitrogen and iron. Centric and pennate diatoms exhibited similar behavior, with a high density of ORFs responding within a 5 log2FC range in both conditions. Dinoflagellates and viruses were more limited, mostly responding to deplete conditions. This could be because dinoflagellate are known to respond mostly post-transcriptionally[57,58] and viruses were more successful at late bloom stages[59].

In the second experiment, Fv/Fm measurements confirm that cells felt the stress of low iron. While Fv/Fm hovered around 0.5 for B2, which was not treated with an Fe-chelator, it dropped dramatically from 0.41 to 0.24 in B2 (incubated with 100nM of the iron-chelator, DFB; Fig. S1). In low iron, iron stress induced proteins (ISIPs) were strongly upregulated, including the carbonate-dependent inorganic iron transporter, ISIP2a (aka phytotransferrin)[60]. Due to the insolubility of inorganic iron, most iron available in the water column is complexed by organic ligands[61]. Ferric reductases, which reduce and import complexed extracellular iron[62], were expressed in diatoms, haptophytes, chlorophytes and dinoflagellates. However, only diatoms and haptophytes had ferric reductase ORFs that were significantly upregulated in low iron ($2.3 < \log_2FC < 9.3$), possibly indicating possession of a more sensitive iron-sensing system in these

groups[63,64]. Other genes highly expressed in low iron were replacements for iron metalloenzymes, such as the copper-containing protein, plastocyanin, and flavodoxin, the flavin-containing alternative to the iron-sulfur protein, ferredoxin.

As with iron, phytoplankton also increased efforts to take up nitrogen when it became limiting. The nitrate transporter, NRT2 was upregulated in many taxa in response to low N, along with the nitrogen assimilation gene, nitrite reductase (NIR; Fig. S32-S34). Interestingly, more so than NRT2, formamidase ORFs were very strongly and consistently upregulated in all lineages when inorganic nitrogen became limiting. This phenomenon has been observed in dinoflagellates, haptophytes[7] and diatoms (Fig. S35-S37). Formamidase allows for the production of formate and ammonium from formamide, which is produced when histidine and cyanide are broken down[65]. Formamidase upregulation in N deplete conditions suggests that cells are turning to organic sources of nitrogen when inorganic sources are not available.

In both nitrogen and iron limitation, a significant decrease in DNA replication ORFs indicated that cell division slowed. There was a corresponding reduction in the photosynthetic antenna and increase in photoprotective antenna in order to reduce energy flow into the cell (Fig. 4, Fig. 5). In the case of nitrogen limitation, cells may be reducing energy production because they are protein limited and must reduce their growth rate. They may also be scavenging nitrogen from the breakdown of N-rich chlorophyll molecules[43]. In the case of iron, they are likely responding to a shortage of essential Fe-S clusters[3]. The photosynthetic enzyme Ferredoxin-NADP(+) reductase (FNR), which was downregulated under N stress, was upregulated under Fe-stress, likely in an effort to increase NADPH production.

The largest physiological difference in nitrogen and iron response was that major metabolic pathways (Calvin-Benson cycle, fatty acid biosynthesis, gluconeogenesis, pigment

biosynthesis) that were downregulated in response to nitrogen limitation were upregulated in response to iron limitation (Fig. 4, Fig. 5b). In iron limitation, cells are limited by energy (available ATP and NADH), but not macronutrient building blocks. Elevation of metabolic pathways in low iron may be an attempt to force flux through these essential metabolic pathways as a priority over other cellular needs.

While these general physiological shifts held across all phytoplankton lineages, there were also some lineage-specific responses to nutrient limitation. Chlorophytes appeared to be the mostly strongly affected by low iron, with greater differential expression of LHCs and metabolism ORFs. Diatoms appeared to be coping better with low iron than haptophytes. Dinoflagellates had a smaller response to nitrogen than other phytoplankton, perhaps because of a greater ability to supplement N through osmotrophy and predation[66].

An analysis of published transcriptomes examining nitrogen and iron availability largely corroborate what we observed here for major phytoplankton lineages (Fig. S35 – S39). While some ORFs did not behave consistently across datasets (e.g. GAPDH and NRT2 across nitrogen conditions), many carbon metabolism ORFs (transketolase, FBA class II, fructose-1,6-bisphosphatase I, triosephosphate isomerase, ribulose-phosphate 3-epimerase, phosphoglycerate kinase (PGK)) and chlorophyll biosynthesis ORFs (coproporphyrinogen II oxidase (CPOX)) were consistently upregulated in response to low N for published diatom datasets and diatom and haptophytes observed here. Interestingly, the majority of these ORFs were slightly downregulated in response to low N in *E. huxleyi*[8] and the dinoflagellates in our data (Fig. S36). In diatoms and published diatom datasets, transketolase, Fe-Mn SOD, LHCA1, biotin synthase, PEPC and GSII were consistently upregulated in high iron, and in pennate diatoms and published diatom datasets, ISIP1, ISIP3, FBA class I and FLDA were strongly upregulated in response to

low iron (Fig. S39). Some carbon metabolism (FBA class 1, phosphoribulokinase, PGK), iron proteins (PETC, PETH), and chlorophyll biosynthesis ORFs (CPOX) that are typically downregulated in low iron were upregulated in low iron in centric and pennate diatoms observed here (Fig. S39). This may be because diatoms observed here were not as severely iron limited as in classical experiments due to the limited availability of DFB-complexed iron.

Changes to light harvesting machinery under nutrient stress

Light harvesting complexes (LHCs) are phylogenetically diverse arrays of pigment-binding proteins that resonantly shuttle light energy into reaction centers for photosynthesis. They also have a secondary role of photoprotection. LHCs can dissipate excess solar energy as heat via Non-Photochemical Quenching (NPQ) to avoid creating harmful reactive oxygen species. The fastest and most common form of NPQ is energy-dependent quenching (qE)[67] in which LHCs change conformation and vary pigment-pigment interactions[68] to create energy traps that allow singlet chlorophylls to dissipate excitation energy. When photosynthetic electron transport exceeds CO_2 fixation capacity, lumenal pH drops[67], activating the xanthophyll cycle, in which key pigments are de-epoxidated on the timescale of minutes[67]. In addition to de-epoxidated xanthophylls, qE also requires a specialized LHC protein whose activity is also triggered by a change in pH. In plants, that protein is PsbS. In *C. reinhardtii* [69] and diatoms[70], the stress responsive Lhcsr/Lhcx family (hereafter, Lhcsr) has been shown to be essential. Thus, photoautotrophs can tune their light harvesting versus photoprotective capacity both by varying LHCs[71], and the xanthophyll pigments they contain[67].

Here, we observe that phytoplankton change both regulation of LHCs (Fig. 6, S40) and xanthophyll cycle enzymes (Fig. S41) in response to nitrogen and iron starvation. As expected, the majority of LHC's expressed were upregulated mid-bloom, when growth was unhampered by

nutrient limitation. Chlorophyll a/b-binding LHCI and LHCII from the green algal lineage were more consistently up in nutrient replete conditions than chlorophyll a/c binding LHCI and LHCII[72], which were sometimes up in low iron. While 674 LHC ORFs were significantly upregulated (FDR < 0.05) in response to both N-replete and Fe-replete conditions, none were significantly upregulated in response to both N-deplete and Fe-deplete conditions. Additionally, many more LHC ORFs were significantly up in response to low iron (n= 230) than low nitrogen (n= 44). While chlorophyll a/c binding LHCI and LHCII and Lhcr (red algae-derived) lineages were moderately upregulated in response to iron (max log_2FC ~5), Lhcsr and Lhcz lineages were very strongly upregulated in low Fe (max log_2FC ~10).

The nutrient response of these lineages is consistent with previous findings in the pennate diatom, *P. tricornutum*, in which Lhcsr was upregulated in response to low N, and Lhcf, Lhcz and Lhcsr were strongly upregulated in response to low Fe[13,73,74]. Additionally, both transcript and protein levels of Lhcsr were upregulated in response to low iron in the centric diatom, *Thalassiosira oceanica*[75]. While the function of the Lhcz family is yet unknown[76], upregulation of Lhcsr in response to nutrient stress likely indicates an increase in qE capacity. The strong upregulation of Lhcsr and Lhcz families in response to iron limitation that we observe here indicates that these families allow diverse phytoplankton taxa to maintain photoprotective capability under iron stress, when the capacity of the photosynthetic apparatus is limited.

An increase in NPQ in low iron conditions in diatoms and dinoflagellates is further evidenced by measurements of diadinoxanthin (Dd) cycle pigments. In experiment 2, the percent of the xanthophyll pool composed of the photo-protective pigment, diatoxanthin (Dt) was consistently higher in low iron conditions (B4) than controls (B2; Fig. S42). The Dt:Dd ratio has

been shown to increase in response to iron stress in haptophytes[77] and dinoflagellates[78], and increase[79] or remain constant[80] in diatoms.

While the expression of many epoxidases and de-epoxidases here was significantly responsive to nitrogen across diatoms, other stramenopiles and chlorophytes, only diatom xanthophyll cycle enzymes were upregulated in response to low iron (Fig. S41). This is consistent with significant upregulation of the centric diatom xanthopyll cycle de-epoxidase protein under iron limitation in *T. oceanica*[75], and may contribute to diatom success in iron-limited regimes.

Genetic markers of nitrogen and iron status in diatoms

Few recommendations have been made for gene markers to detect the cellular status of iron, and especially nitrogen, in the field. While dinoflagellate transcription is often too stable to indicate changes in nutrient status, this is not the case for diatoms. Iron starvation induced proteins (ISIP1, ISIP2, and ISIP3) have been shown to be highly upregulated in response to iron starvation in diatoms in the lab[60,81]. ISIP3 and the metalloprotein replacement, flavodoxin, have also been shown to be indicative of *T. oceanica* iron limitation in the field[82]. Indicators of diatom nitrogen status, however, have been harder to come by. In the lab, key nitrogen uptake and metabolism genes respond both to nitrogen starvation and pulses of nitrate, rendering them ineffective as a diagnostic tool on their own [9,83].

Here, we identify gene pairs with consistently opposing responses to nitrogen and iron limitation whose ratio can be used to indicate the nutrient status of the cell. In laboratory experiments with the model diatom, *Phaeodactylum tricornutum*, the nitrate transporter NRT2 was upregulated in response to both the addition of nitrate and nitrogen starvation. GSII, on the other hand, was only upregulated in response to the addition of nitrate[83]. By taking the ratio of

the expression of NRT2 to GSII, we can create a marker that is only indicative of intracellular nitrogen stress. Indeed, we observed a log ratio of NRT2 (cluster 139) to glutamine synthase (GSII, cluster 139) expression that was consistently higher in N-deplete conditions than N-replete conditions across diatom genera (Fig. 7A).

Similarly, we found that the ratio of total ISIP expression (ISIP1, uniprot B7GA90; ISIP2 uniprot B7FYL2; ISIP3, uniprot B7G4H8) to the thiamine (vitamin B1) synthesis ORF, phosphomethylpyrimidine synthase (*thiC*; K03147) expression was consistently higher for Fe-deplete than Fe-replete conditions (Fig. 7B). In natural diatom populations, ISIPs were upregulated in response to experimentally iron-depleted water, and thiC was upregulated in response to added iron[15]. ThiC was also one of the most highly upregulated genes in response to iron replete conditions in *Pseudo-nitzschia granii*, likely because of its reliance on iron-sulfur clusters[12]. We found this ratio to more consistently identify iron status than the more traditional comparison of flavodoxin to ferredoxin (Fig. S43) and ISIP2 to ferritin (Fig. S44, S45). While ISIP/ThiC ratios perform well for diatoms, they may not be as informative for other taxa that uptake environmental thiamine rather than producing it exogenously. Here, we also observed that ratios of ISIP expression to histone 2A (H2A; KOG1757), while not as consistent at the genus level for diatoms, were broadly indicative of iron status across a wide range of taxa (Fig. S46).

Viral infection as a driver of bloom demise

Regardless of whether the bloom was being limited by nitrogen or iron, overall viral expression was higher in late-bloom than mid-bloom conditions (Fig. 2). This could be because nutrient-stressed cells are more prone to infection, rapid replication of host cells allowed viruses to flourish, or high cell density allows for increased contagion. It could also be due to expression being dominated by a diversity RNA viruses (Fig. 8B, S47, S48). For RNA viruses, mRNA

captured here represents not only expression, but also genomic material. That is to say, RNA viruses may show a stronger signal later in the bloom because of an accumulation of biomass and not necessarily greater activity.

While nearly all RNA viral expression was up later in the bloom, dsDNA viruses infecting bacteria and eukaryotes, including the phycodnaviridae lineage infecting algae, were sometimes up mid-bloom (Fig. 8A). Bacteriophages and dsDNA mycoviruses were also up late-bloom (*Caudovirales*), perhaps increasing in number in tandem with heterotrophic bacteria and fungi feeding on lysed cells.

While dinoflagellate viruses were not specifically identified, we see a large increase in a diatom RNA virus best annotated as "Chaetoceros RNA virus 2" in late-bloom conditions, especially in experiment 1 (Fig. S49). This indicates that viral infection, along with nutrient limitation, likely played a role in bloom demise.

Conclusions

Despite large changes in biomass and nutrient availability, genus-level taxonomic composition of the phytoplankton community was largely stable across nitrogen- and iron-limited blooms. The associated bacterial community was dominated by taxa implicated in mutualist and predatory relationships with phytoplankton species, and several members were differentially active across nutrient conditions. Viral activity was higher later in both blooms, and viruses likely exacerbated nutrient stress and played a role in bloom demise. Replicated population-level shifts were observed in phytoplankton taxa as the blooms progressed, and were more common in experiment 1, where the bloom was larger. Shifted ORFs were composed of functional annotations that were adaptive to bloom conditions, and likely represent a combination of increased abundance of winning taxa and overexpression of adaptive genes

within already abundant species. The density of SNPs decreased in diatoms and chlorophytes in low N, but increased in dinoflagellates, who showed the strongest upregulation of sexual reproduction genes. Major metabolic pathways (Calvin-Benson cycle, fatty acid biosynthesis, gluconeogenesis, pigment biosynthesis) showed highly consistent expression across algal lineages, and were downregulated in response to nitrogen limitation but upregulated in response to iron limitation. Light harvesting overall was downregulated when nutrients were limiting, but stress-adaptive LHCSR and LHCZ lineages were upregulated across diverse taxa, likely playing a role in NPQ. Together, we present not only a comprehensive snapshot of phytoplankton physiological response to nutrient pulses, but also the dynamic genetic selection processes by which these responses were established. Finally, we establish the ratio of total ISIP: thiC as a biomarker of iron stress in diatoms, and present for the first time a biomarker of diatom nitrogen status--the ratio of NRT2:GSII.

Conflict of Interest

The authors declare no conflicts of interest.

Acknowledgements

B.K. was funded by NSF graduate research fellowship DGE-1144086. We would like to thank Sarah Fawcett for performing nutrient analyses and Misha Mayim for development of the light harvesting complex HMM.

References

1. Moore, J. K., Doney, S. C. & Lindsay, K. Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. *Global Biogeochem. Cycles* **18**, 1–21 (2004).

2. Falkowski, P. G. & Raven, J. A. *Aquatic photosynthesis*. (Princeton University Press, 2013).

3. Liu, J., Chakraborty, S., Hosseinzadeh, P., Yu, Y., Tian, S., Petrik, I., Bhagi, A. & Lu, Y. Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. *Chem. Rev.* **114**, 4366–4369 (2014).

4. Biller, D. V. & Bruland, K. W. The central California Current transition zone: A broad region exhibiting evidence for iron limitation. *Prog. Oceanogr.* **120**, 370–382 (2014).

5. Saito, M. A., Bertrand, E. M., Dutkiewicz, S., Bulygin, V. V., Moran, D. M., Monteiro, F. M., Follows, M. J., Valois, F. W. & Waterbury, J. B. Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph Crocosphaera watsonii. *Proc. Natl. Acad. Sci.* **108**, 2184–2189 (2011).

6. Marchetti, A., Schruth, D. M., Durkin, C. A., Parker, M. S., Kodner, R. B., Berthiaume, C. T., Morales, R., Allen, A. E. & Armbrust, E. V. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl. Acad. Sci.* **109**, E317–E325 (2012).

7. Bruhn, A., LaRoche, J. & Richardson, K. Emiliania Huxleyi (Prymnesiophyceae): Nitrogen-metabolism genes and their expression in response to external nitrogen sources. *J. Phycol.* **46**, 266–277 (2010).

8. Alexander, H., Rouco, M., Haley, S. T. & Dyhrman, S. T. Transcriptional response of Emiliania huxleyi under changing nutrient environments in the North Pacific Subtropical Gyre. *Environ. Microbiol.* **22**, 1847–1860 (2020).

9. Smith, S. R., Dupont, C. L., McCarthy, J. K., Broddrick, J. T., Oborník, M., Horák, A., Füssy, Z., Cihlář, J., Kleessen, S., Zheng, H., McCrow, J. P., Hixson, K. K., Araújo, W.

L., Nunes-Nesi, A., Fernie, A., Nikoloski, Z., Palsson, B. O. & Allen, A. E. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* **10**, 4552 (2019).

10. Litchman, E., Klausmeier, C. A. & Station, K. B. Trait-Based Community Ecology of Phytoplankton. *Annu. Rev. Ecol. Evol. Syst* **39**, 615–39 (2008).

11. Bruland, K. W., Rue, E. L. & Smith, G. J. Iron and macronutrients in California coastal upwelling regimes: Implications for diatom blooms. *Limnol. Oceanogr.* **46**, 1661–1674 (2001).

12. Cohen, N. R., Gong, W., Moran, D. M., McIlvin, M. R., Saito, M. A. & Marchetti, A. Transcriptomic and proteomic responses of the oceanic diatom Pseudo-nitzschia granii to iron limitation. *Environ. Microbiol.* **20**, 3109–3126 (2018).

13. Smith, S. R., Gillard, J. T. F., Kustka, A. B., McCrow, J. P., Badger, J. H., Zheng, H., New, A. M., Dupont, C. L., Obata, T., Fernie, A. R. & Allen, A. E. Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. *PLoS Genet.* **12**, (2016).

14. Marchetti, A., Schruth, D. M., Durkin, C. A., Parker, M. S., Kodner, R. B., Berthiaume, C. T., Morales, R., Allen, A. E. & Armbrusta, E. V. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).

15. Cohen, N. R., Ellis, K. A., Lampe, R. H., McNair, H., Twining, B. S., Maldonado, M. T., Brzezinski, M. A., Kuzminov, F. I., Thamatrakoln, K., Till, C. P., Bruland, K. W., Sunda, W. G., Bargu, S. & Marchetti, A. Diatom Transcriptional and Physiological Responses to Changes in Iron Bioavailability across Ocean Provinces. *Front. Mar. Sci.* **4**, (2017).

16. Lewandowska, A. M., Striebel, M., Feudel, U., Hillebrand, H. & Sommer, U. Marine Science. **72**, 1908–1915 (2015).

17. Rynearson, T. A. & Armbrust, E. V. Maintenance of clonal diversity during a spring bloom of the centric diatom Ditylum brightwellii. *Mol. Ecol.* **14**, 1631–1640 (2005).

18. Fawcett, S. E. & Ward, B. B. Phytoplankton succession and nitrogen utilization during the development of an upwelling bloom. *Mar. Ecol. Prog. Ser.* **428**, 13–31 (2011).

19. Van Oostende, N., Dunne, J. P., Fawcett, S. E. & Ward, B. B. Phytoplankton succession explains size-partitioning of new production following upwelling-induced blooms. *J. Mar. Syst.* **148**, 14–25 (2015).

20. Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L. & Pace, N. R. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci.* **82**, 6955–6959 (1985).

21. Herlemann, D. P. R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J. & Andersson, A. F. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* **5**, 1571–1579 (2011).

22. Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H. W. & Richards, T. A. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31 (2010).

23. Bertrand, E. M., McCrow, J. P., Moustafa, A., Zheng, H., McQuaid, J. B., Delmont, T. O., Post, A. F., Sipler, R. E., Spackeen, J. L., Xu, K., Bronk, D. A., Hutchins, D. A. & Allen, A. E. Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9938–43 (2015).

24. Kolody, B. C., McCrow, J. P., Allen, L. Z., Aylward, F. O., Fontanez, K. M., Moustafa,

A., Moniruzzaman, M., Chavez, F. P., Scholin, C. A., Allen, E. E., Worden, A. Z., Delong, E. F. & Allen, A. E. Diel transcriptional response of a California Current plankton microbiome to light, low iron, and enduring viral infection. *ISME J.* (2019) doi:10.1038/s41396-019-0472-2.

25. Schmieder, R., Lim, Y. W. & Edwards, R. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28**, 433–435 (2012).

26. Rho, M., Tang, H. & Ye, Y. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* **38**, e191–e191 (2010).

27. Sonnhammer, E., Eddy, S. R., Birney, E., Bateman, A. & Durbin, R. Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res.* **26**, 320–322 (1998).

28. Protein BLAST: search protein databases using a protein query. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins.

29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

30. Podell, S. & Gaasterland, T. DarkHorse: A method for genome-wide prediction of horizontal gene transfer. *Genome Biol.* **8**, (2007).

31. Bayat, A., Gaëta, B., Ignjatovic, A. & Parameswaran, S. Improved VCF normalization for accurate VCF comparison. *Bioinformatics* **33**, 964–970 (2017).

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

33. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).

34. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).

35. Øvergård, A. C., Nerland, A. H. & Patel, S. Evaluation of potential reference genes for real time RT-PCR studies in Atlantic halibut (Hippoglossus Hippoglossus L.); during development, in tissues of healthy and NNV-injected fish, and in anterior kidney leucocytes. *BMC Mol. Biol.* **11**, (2010).

36. Liu, Q., Lei, K., Ma, Q., Qiao, F., Li, Z.-C. & An, L.-H. Ribosomal protein L7 as a suitable reference gene for quantifying gene expression in gastropod Bellamya aeruginosa. *Environ. Toxicol. Pharmacol.* **43**, 120–127 (2016).

37. Figueiredo, A., Loureiro, A., Batista, D., Monteiro, F., Várzea, V., Pais, M. S., Gichuru, E. K. & Silva, M. C. Validation of reference genes for normalization of qPCR gene expression data from Coffea spp. hypocotyls inoculated with Colletotrichum kahawae. *BMC Res. Notes* **6**, (2013).

38. Liu, C., Wu, G., Huang, X., Liu, S. & Cong, B. Validation of housekeeping genes for gene expression studies in an ice alga Chlamydomonas during freezing acclimation. *Extremophiles* **16**, 419–425 (2012).

39. Haq, S., Bachvaroff, T. R. & Place, A. R. Characterization of acetyl-CoA carboxylases in the basal dinoflagellate amphidinium carterae. *Mar. Drugs* **15**, 1–10 (2017).

40. Evans, C., Hardin, J. & Stoebel, D. M. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Brief. Bioinform.* **19**, 776–792 (2018).

41. McGee, W. A., Pimentel, H., Pachter, L. & Wu, J. Y. Compositional Data Analysis is necessary for simulating and analyzing RNA-Seq data. *bioRxiv* 564955 (2019)

doi:10.1101/564955.

42. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

43. Kuczynska, P., Jemiola-Rzeminska, M. & Strzalka, K. Photosynthetic pigments in diatoms. *Mar. Drugs* **13**, 5847–5881 (2015).

44. Hae, J. J., Jae, Y. P., Jae, H. N., Myung, O. P., Jeong, H. H., Kyeong, A. S., Jeng, C., Chi, N. S., Kwang, Y. L. & Won, H. Y. Feeding by red-tide dinoflagellates on the cyanobacterium Synechococcus. *Aquat. Microb. Ecol.* **41**, 131–143 (2005).

45. Baroli, I., Niyogi, K. K., Barber, J. & Heifetz, P. Molecular genetics of xanthophyll-dependent photoprotection in green algae and plants. *Philos. Trans. R. Soc. B Biol. Sci.* **355**, 1385–1394 (2000).

46. Val, J., Monge, E., Heras, L. & Abadía, J. Changes in photosynthetic pigment composition in higher plants as affected by iron nutrition status. *J. Plant Nutr.* **10**, 995–1001 (1987).

47. Soldatini, G. F., Tognini, M., Castagna, A., Baldan, B. & Ranieri, A. Alterations in thylakoid membrane composition induced by iron starvation in sunflower plants. *J. Plant Nutr.* **23**, 1717–1732 (2000).

48. Seymour, J. R., Amin, S. A., Raina, J. B. & Stocker, R. Zooming in on the phycosphere: The ecological interface for phytoplankton-bacteria relationships. *Nat. Microbiol.* **2**, (2017).

49. Bigalke, A., Meyer, N., Papanikolopoulou, L. A., Wiltshire, K. H. & Pohnert, G. The algicidal bacterium Kordia algicida shapes a natural plankton community. *Appl. Environ. Microbiol.* **85**, 1–12 (2019).

50. Carini, P., Steindler, L., Beszteri, S. & Giovannoni, S. J. Nutrient requirements for growth of the extreme oligotroph ‘Candidatus Pelagibacter ubique’ HTCC1062 on a defined medium. *ISME J.* **7**, 592–602 (2013).

51. Armbrust, E. V. & Galindo, H. M. Rapid Evolution of a Sexual Reproduction Gene in Centric Diatoms of the Genus Thalassiosira. *Appl. Environ. Microbiol.* **67**, 3501–3513 (2001).

52. Meng, F. Q., Song, J. T., Zhou, J. & Cai, Z. H. Transcriptomic Profile and Sexual Reproduction-Relevant Genes of Alexandrium minutum in Response to Nutritional Deficiency. *Front. Microbiol.* **10**, 1–16 (2019).

53. Kim, J. Y., Park, S. C., Hwang, I., Cheong, H., Nah, J. W., Hahm, K. S. & Park, Y. Protease inhibitors from plants with antimicrobial activity. *Int. J. Mol. Sci.* **10**, 2860–2872 (2009).

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

55. Turpin, D. H. & Bruce, D. Regulation of photosynthetic light harvesting by nitrogen assimilation in the green alga Selenastrum minutum. *FEBS Lett.* **263**, 99–103 (1990).

56. Morris, I., Yentsch, C. M. & Yentsch, C. S. The physiological state with respect to nitrogen of phytoplankton from low-nutrient subtropical water as measured by the effect of ammonium ion on dark carbon fixation. *Limnol. Oceanogr.* **16**, 859–868 (1971).

57. Okamoto, O. K. & Hastings, J. W. Novel dinoflagellate clock-related genes identified through microarray analysis. *J. Phycol.* **39**, 519–526 (2003).

58. Brunelle, S. A. & Van Dolah, F. M. Post-transcriptional regulation of S-Phase genes in the dinoflagellate, karenia brevis. *J. Eukaryot. Microbiol.* **58**, 373–382 (2011).

59. Brussaard, C. P. D. Viral control of phytoplankton populations—a review. *J. Eukaryot. Microbiol.* **51**, 125–138 (2004).

60. McQuaid, J. B., Kustka, A. B., Oborník, M., Horák, A., McCrow, J. P., Karas, B. J., Zheng, H., Kindeberg, T., Andersson, A. J., Barbeau, K. A. & Allen, A. E. Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. *Nature* **555**, 534–537 (2018).

61. Barbeau, K., Photochemistry, S. & Barbeau, K. Photochemistry of Organic Iron (III) Complexing Ligands in Oceanic Systems Invited Review Photochemistry of Organic Iron (III) Complexing Ligands in Oceanic Systems. *Photochem. Photobiol.* **82**, 1505–1516 (2006).

62. Coale, T. H., Moosburner, M., Horák, A., Oborník, M., Barbeau, K. A. & Allen, A. E. Reduction-dependent siderophore assimilation in a model pennate diatom. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23609–23617 (2019).

63. Falciatore, A., D'Alcalà, M. R., Croot, P. & Bowler, C. Perception of environmental signals by a marine diatom. *Science (80-.).* **288**, 2363–2366 (2000).

64. Rizkallah, M. R., Frickenhaus, S., Trimborn, S., Harms, L., Moustafa, A., Benes, V., Gäbler-Schwarz, S. & Beszteri, S. Deciphering Patterns of Adaptation and Acclimation in the Transcriptome of Phaeocystis antarctica to Changing Iron Conditions1. *J. Phycol.* **56**, 747–760 (2020).

65. Hynes, M. J. Amide utilization in Aspergillus nidulans: evidence for a third amidase enzyme. *J. Gen. Microbiol.* **91**, 99–109 (1975).

66. Burkholder, J. A. M., Glibert, P. M. & Skelton, H. M. Mixotrophy, a major mode of nutrition for harmful algal species in eutrophic waters. *Harmful Algae* **8**, 77–93 (2008).

67. Müller, P., Li, X. P. & Niyogi, K. K. Non-photochemical quenching. A response to excess light energy. *Plant Physiol.* **125**, 1558–1566 (2001).

68. Liguori, N., Periole, X., Marrink, S. J. & Croce, R. From light-harvesting to photoprotection: Structural basis of the dynamic switch of the major antenna complex of plants (LHCII). *Sci. Rep.* **5**, 2–11 (2015).

69. Peers, G., Truong, T. B., Ostendorf, E., Busch, A., Elrad, D., Grossman, A. R., Hippler, M. & Niyogi, K. K. An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. *Nature* **462**, 518–521 (2009).

70. Buck, J. M., Sherman, J., Bártulos, C. R., Serif, M., Halder, M., Henkel, J., Falciatore, A., Lavaud, J., Gorbunov, M. Y. & Kroth, P. G. Lhcx proteins provide photoprotection via thermal dissipation of absorbed light in the diatom Phaeodactylum tricornutum. *Nat. Commun.* **10**, 1–12 (2019).

71. Horton, P. Optimization of light harvesting and photoprotection: Molecular mechanisms and physiological consequences. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 3455–3465 (2012).

72. Koziol, A. G., Borza, T., Ishida, K. I., Keeling, P., Lee, R. W. & Durnford, D. G. Tracing the evolution of the light-harvesting antennae in chlorophyll a/b-containing organisms. *Plant Physiol.* **143**, 1802–1816 (2007).

73. Allen, A. E., LaRoche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., Finazzi, G., Fernie, A. R. & Bowler, C. Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. *Proc. Natl. Acad. Sci.* **105**, 10438–10443 (2008).

74. Taddei, L., Stella, G. R., Rogato, A., Bailleul, B., Fortunato, A. E., Annunziata, R., Sanges, R., Thaler, M., Lepetit, B., Lavaud, J., Jaubert, M., Finazzi, G., Bouly, J. P. &

Falciatore, A. Multisignal control of expression of the LHCX protein family in the marine diatom Phaeodactylum tricornutum. *J. Exp. Bot.* **67**, 3939–3951 (2016).

75. Lommer, M., Specht, M., Roy, A. S., Kraemer, L., Andreson, R., Gutowska, M. A., Wolf, J., Bergner, S. V., Schilhabel, M. B., Klostermeier, U. C., Beiko, R. G., Rosenstiel, P., Hippler, M. & LaRoche, J. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol.* **13**, (2012).

76. Neilson, J. A. D. & Durnford, D. G. Structural and functional diversification of the light-harvesting complexes in photosynthetic eukaryotes. (2010) doi:10.1007/s11120-010-9576-2.

77. Stefels, J. & Van Leeuwe, M. A. Effects of iron and light stress on the biochemical composition of antarctic Phaeocystis sp. (Prymnesiophyceae). I. Intracellular DMSP concentrations. *J. Phycol.* **34**, 486–495 (1998).

78. Shick, J. M., Iglic, K., Wells, M. L., Trick, C. G., Doyle, J. & Dunlap, W. C. Responses to iron limitation in two colonies of Stylophora pistillata exposed to high temperature: Implications for coral bleaching. *Limnol. Oceanogr.* **56**, 813–828 (2011).

79. Geider, R. J., La Roche, J., Greene, R. M. & Olaizola, M. Response of the photosynthetic apparatus of Phaeodactylum tricornutum (Bacillariophyceae) to nitrate, phosphate, or iron starvation 1. *J. Phycol.* **29**, 755–766 (1993).

80. Beer, A., Juhas, M. & Büchel, C. Influence of different light intensities and different iron nutrition on the photosynthetic apparatus in the diatom Cyclotella meneghiniana (bacillariophyceae). *J. Phycol.* **47**, 1266–1273 (2011).

81. Behnke, J. & Laroche, J. Iron uptake proteins in algae and the role of Iron Starvation-Induced Proteins (ISIPs). (2020) doi:10.1080/09670262.2020.1744039.

82. Chappell, P. D., Whitney, L. P., Wallace, J. R., Darer, A. I., Jean-Charles, S. & Jenkins, B. D. Genetic indicators of iron limitation in wild populations of Thalassiosira oceanica from the northeast Pacific Ocean. *Isme J* **9**, 592–602 (2015).

83. McCarthy, J. K., Smith, S. R., McCrow, J. P., Tan, M., Zheng, H., Beeri, K., Roth, R., Lichtle, C., Goodenough, U., Bowler, C. P., Dupont, C. L. & Allen, A. E. Nitrate reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. *Plant Cell* **29**, 2047–2070 (2017).

Figures

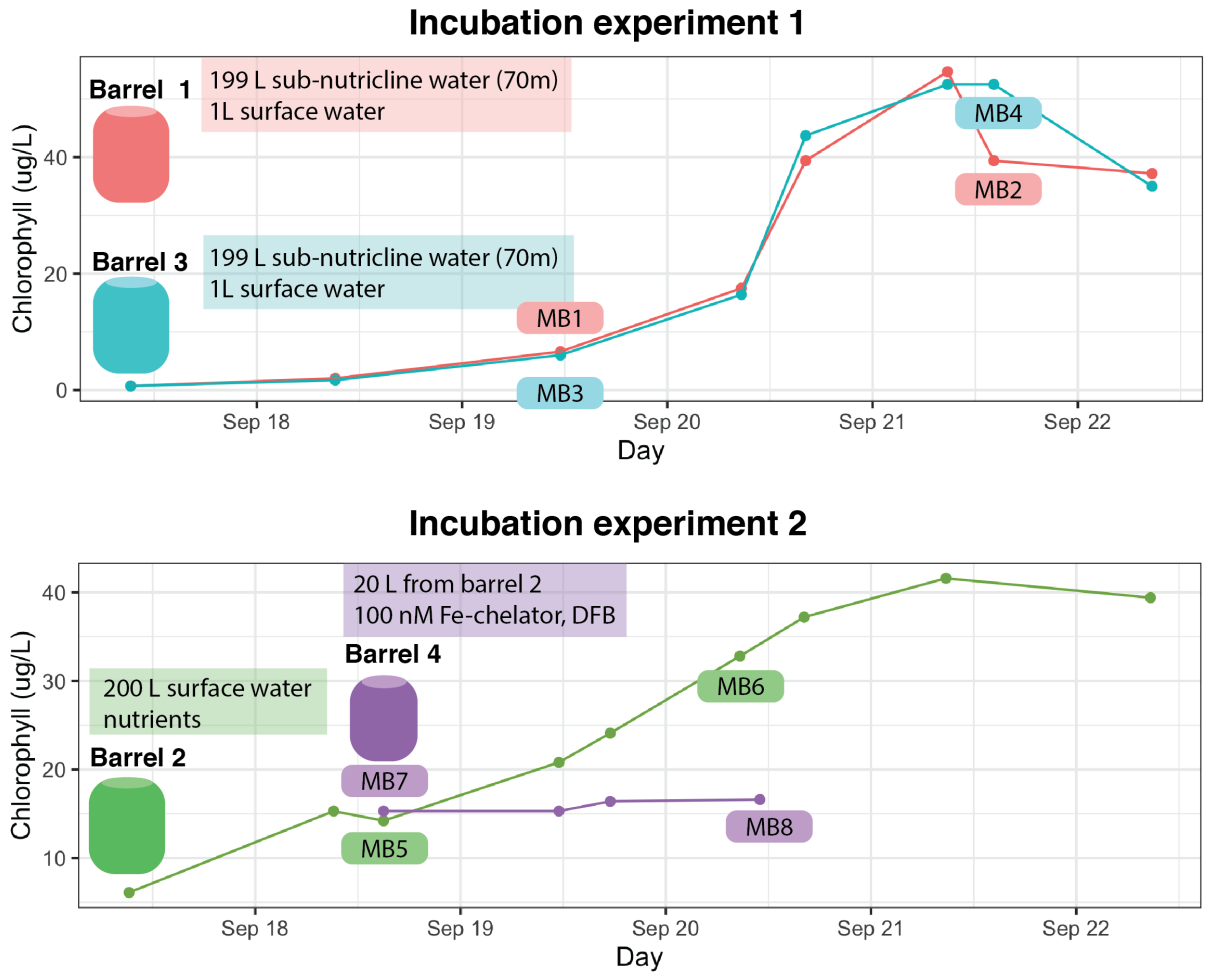


Figure 2.1: Illustration of experimental design for experiments 1 (top) and 2 (bottom). Timing of molecular samples (MB1-8) are shown superimposed over chlorophyll concentrations (y-axis). Sample rectangles and lines depicting chlorophyll concentration are colored based on the barrel they pertain to. In experiment 1, barrels 1 (pink) and 3 (blue) were duplicates, both containing 199 L sub-nutricline water sampled from 70 m depth and 1 L of surface water. In experiment 2, barrel 2 (green) contained 200 L of surface water with nutrients added to achieve final concentrations of 40uM nitrate, 2.5 uM phosphate, and 50 uM silica. On September 18, barrel 4 (purple) was created by subsampling 20 L from barrel 2. In barrel 4, a low-iron environment was created with the addition of the Fe-chelator, DFB (final concentration: 100 nM).

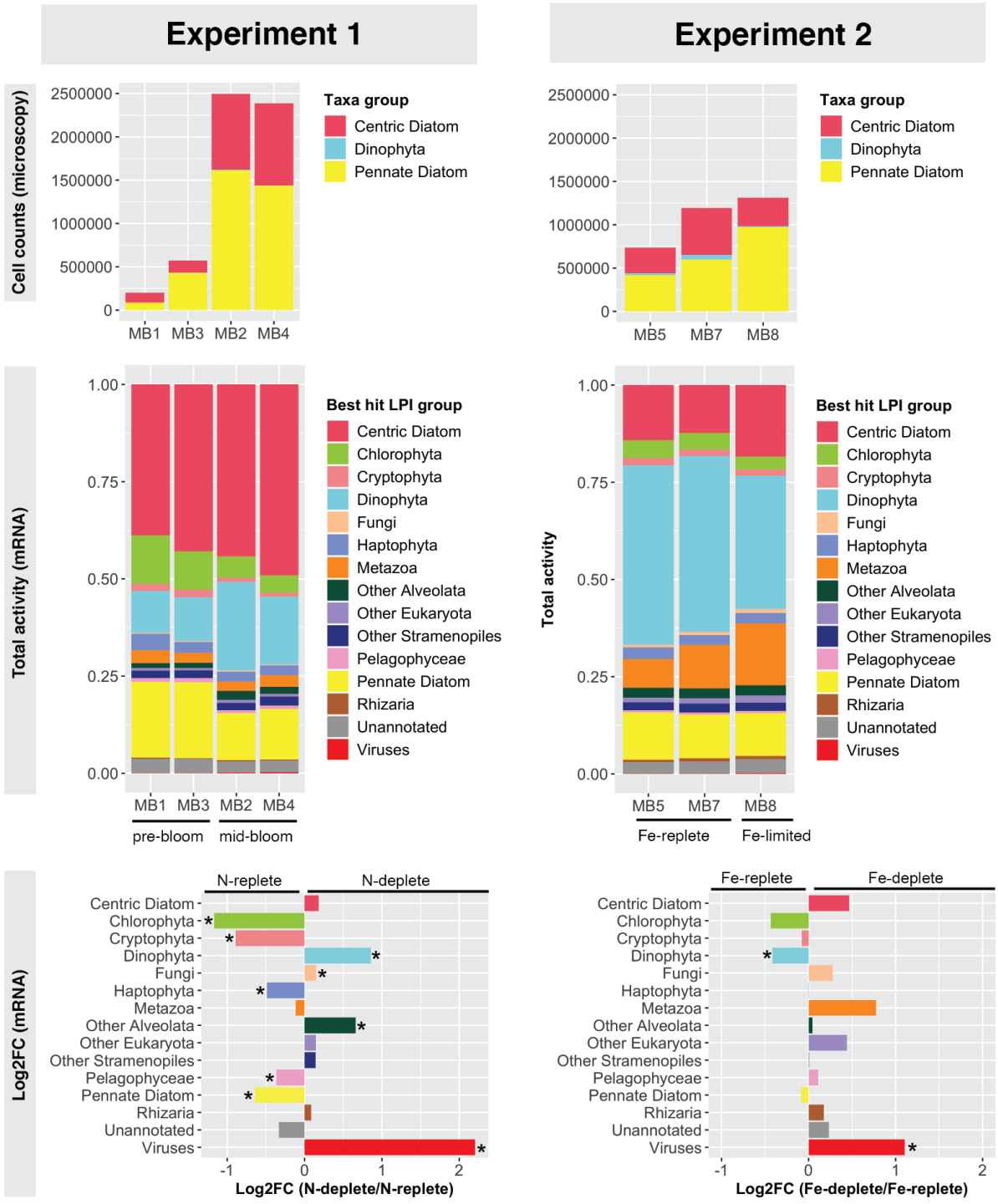
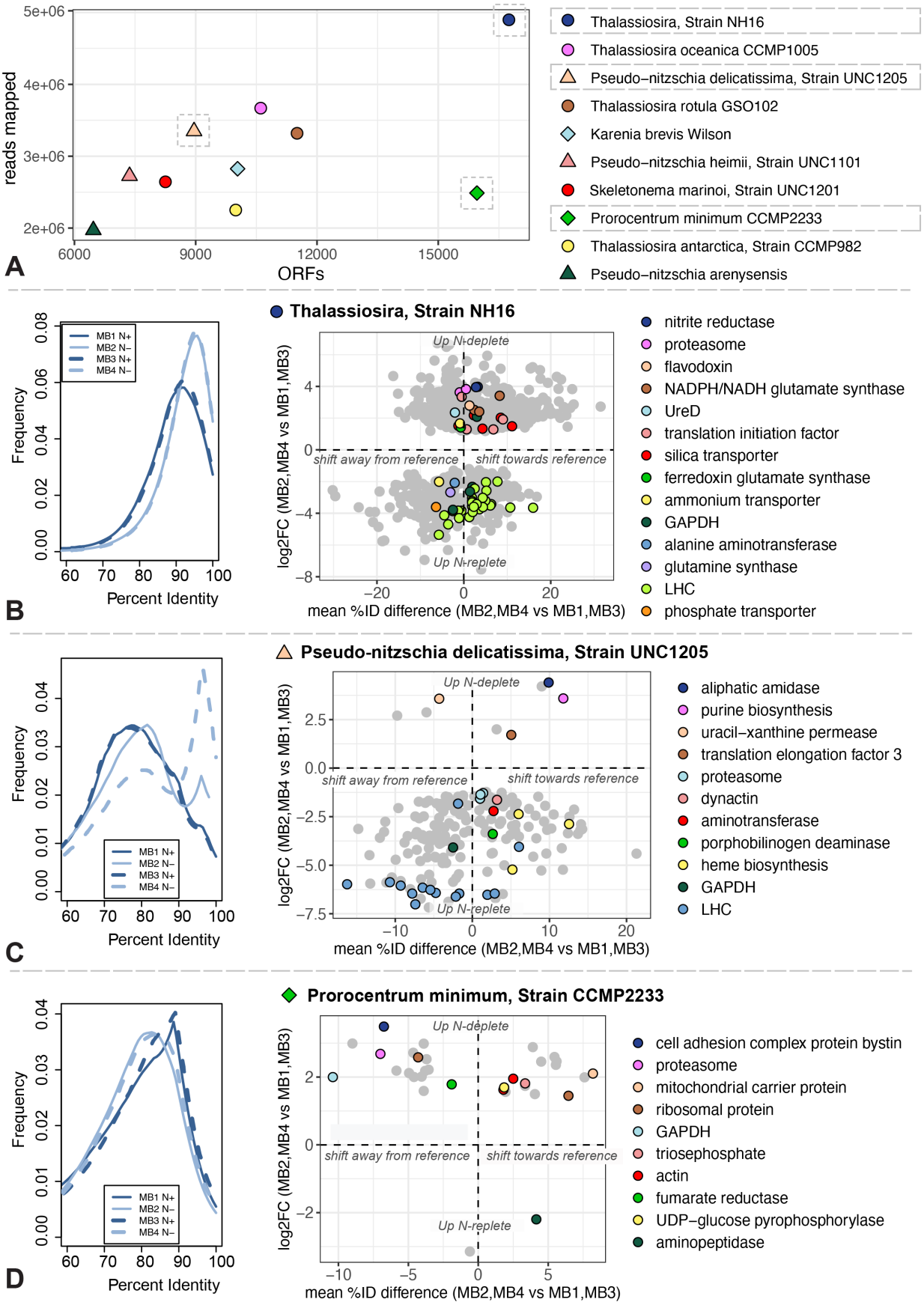


Figure 2.2: Coarse taxonomy across both experiments via both cell counts (top panels) and proportion of total mRNA (middle panels). Bottom panels depict $\log_2$ fold change of taxa group activity as a proportion of library reads across nitrogen (left) and iron (right) conditions. Asterisks denote significant differences (edgeR, FDR < 0.05). MB6 was not included in the differential expression analysis, because although no iron-chelator was added, it appears to be iron-limited due being a late-stage bloom condition.

Figure 2.3: Fine-scale taxonomic shifts across experiment 1 bloom. (**A**) Reads (y-axis) mapping to and number of ORFs (x-axis) mapped to top 10 most abundantly hit MMETSP reference transcriptomes. Circles, triangles, and diamonds represent centric diatom, pennate diatom, and dinoflagellate references, respectively. References examined in parts B, C, and D are indicated with dashed grey lines. Left panels of B,C, and D show percent identity histograms for B) the centric diatom, *Thalassiosira*, Strain NH16, C) the pennate diatom, *Pseudo-nitzschia delicatissima* Strain UNC1205, and D, the dinoflagellate, *Prorocentrum minimum*, Strain CCMP2233 for all nitrogen conditions. Right panels show log fold change (y-axis) versus mean percent identity difference (x-axis) across nitrogen conditions. Only ORFs that are significantly differentially expressed, at least 60% identical to the reference, and shift in the same direction relative to the reference across replicates are shown. ORFs of interest are colored by function.


A
reads mapped
ORFs
5e+06
4e+06
3e+06
2e+06
6000
9000
12000
15000
Thalassiosira, Strain NH16
Thalassiosira oceanica CCMP1005
Pseudo−nitzschia delicatissima, Strain UNC1205
Thalassiosira rotula GSO102
Karenia brevis Wilson
Pseudo−nitzschia heimii, Strain UNC1101
Skeletonema marinoi, Strain UNC1201
Prorocentrum minimum CCMP2233
Thalassiosira antarctica, Strain CCMP982
Pseudo−nitzschia arenysensis
B
Frequency
Percent Identity
MB1 N+
MB2 N−
MB3 N+
MB4 N−
Thalassiosira, Strain NH16
log2FC (MB2,MB4 vs MB1,MB3)
mean %ID difference (MB2,MB4 vs MB1,MB3)
Up N-deplete
Up N-replete
shift away from reference
shift towards reference
nitrite reductase
proteasome
flavodoxin
NADPH/NADH glutamate synthase
UreD
translation initiation factor
silica transporter
ferredoxin glutamate synthase
ammonium transporter
GAPDH
alanine aminotransferase
glutamine synthase
LHC
phosphate transporter
C
Pseudo-nitzschia delicatissima, Strain UNC1205
aliphatic amidase
purine biosynthesis
uracil−xanthine permease
translation elongation factor 3
proteasome
dynactin
aminotransferase
porphobilinogen deaminase
heme biosynthesis
GAPDH
LHC
D
Prorocentrum minimum, Strain CCMP2233
cell adhesion complex protein bystin
proteasome
mitochondrial carrier protein
ribosomal protein
GAPDH
triosephosphate
actin
fumarate reductase
UDP−glucose pyrophosphorylase
aminopeptidase

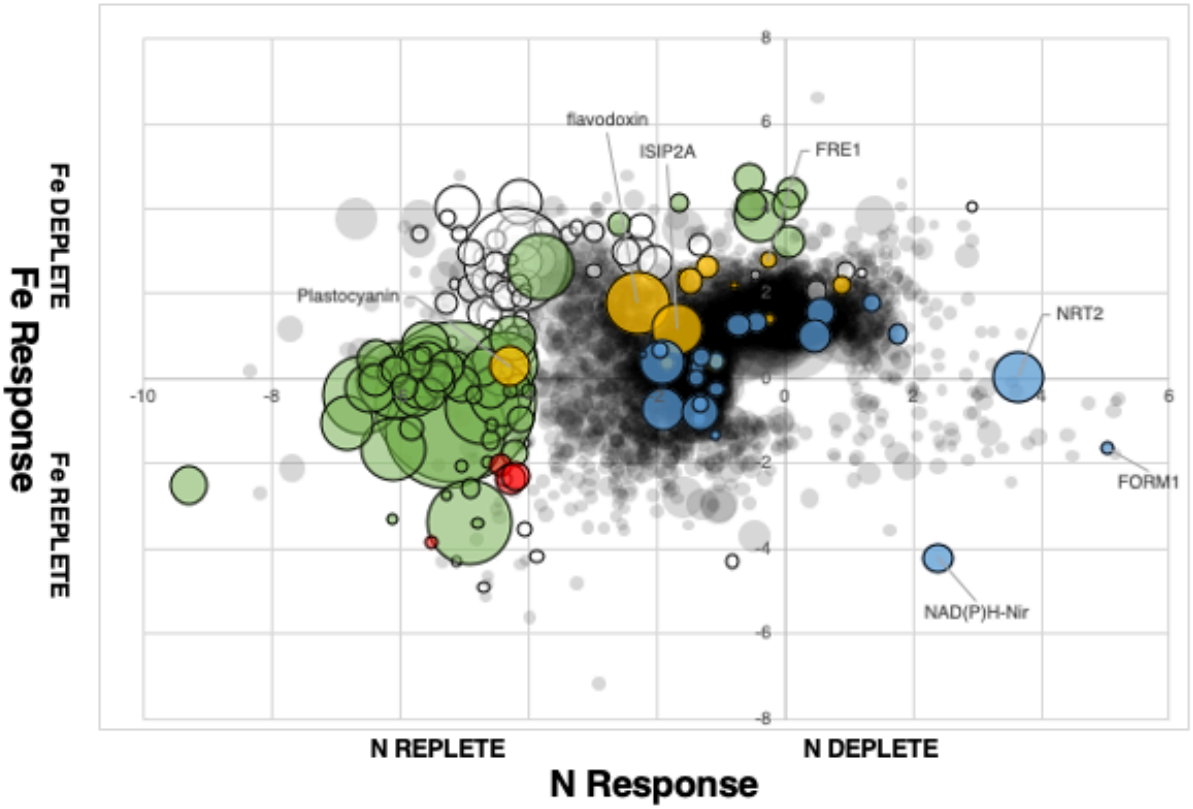


Figure 2.4: Response of major gene clusters to iron and nitrogen status. Circles are scaled by gene cluster total activity, colored by function (green= light harvesting, white = carbon metabolism, yellow= iron-related, blue= nitrogen assimilation), and positioned based on $\log_2$ fold change in response to iron (y-axis) and nitrogen (x-axis) replete and deplete conditions.

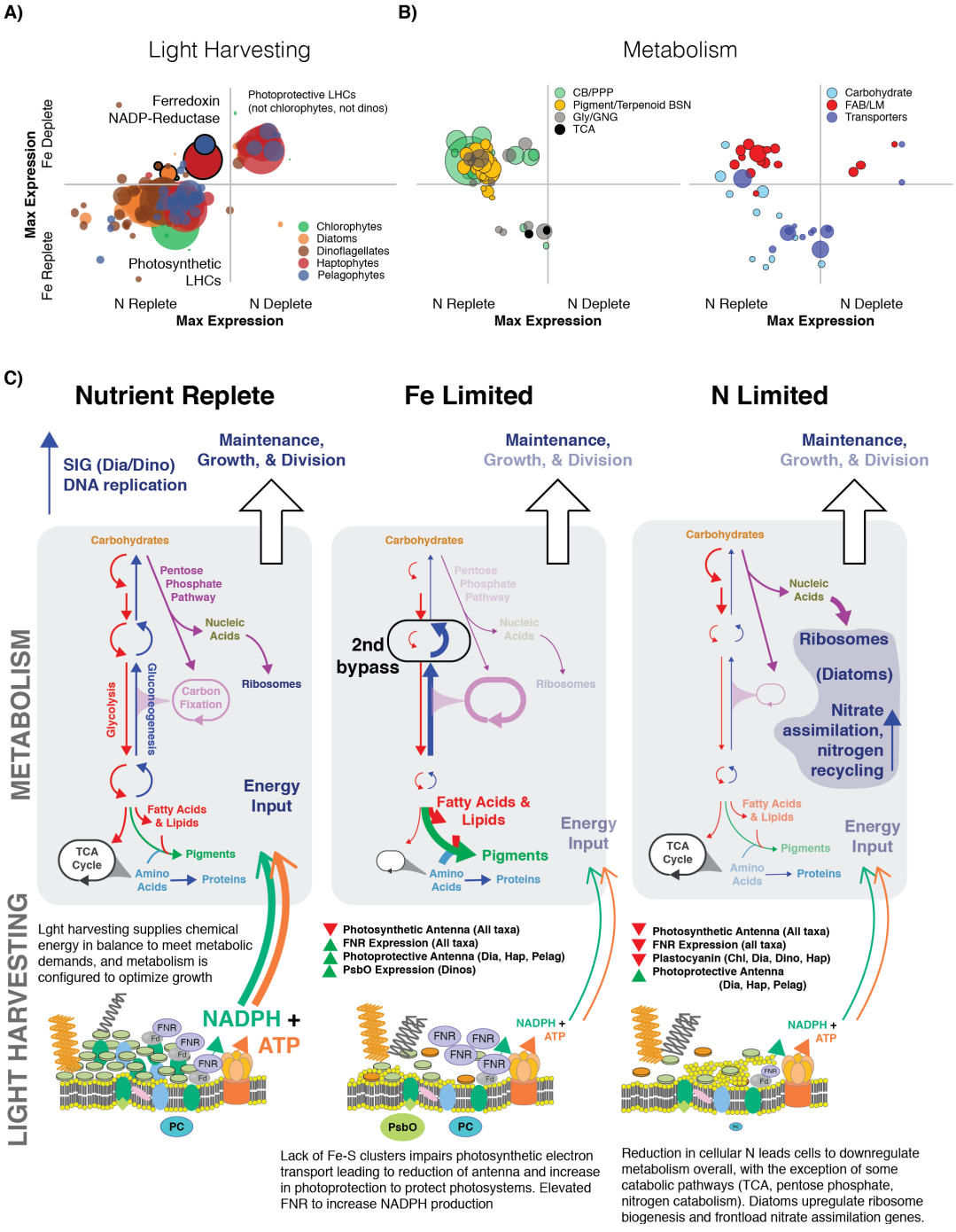
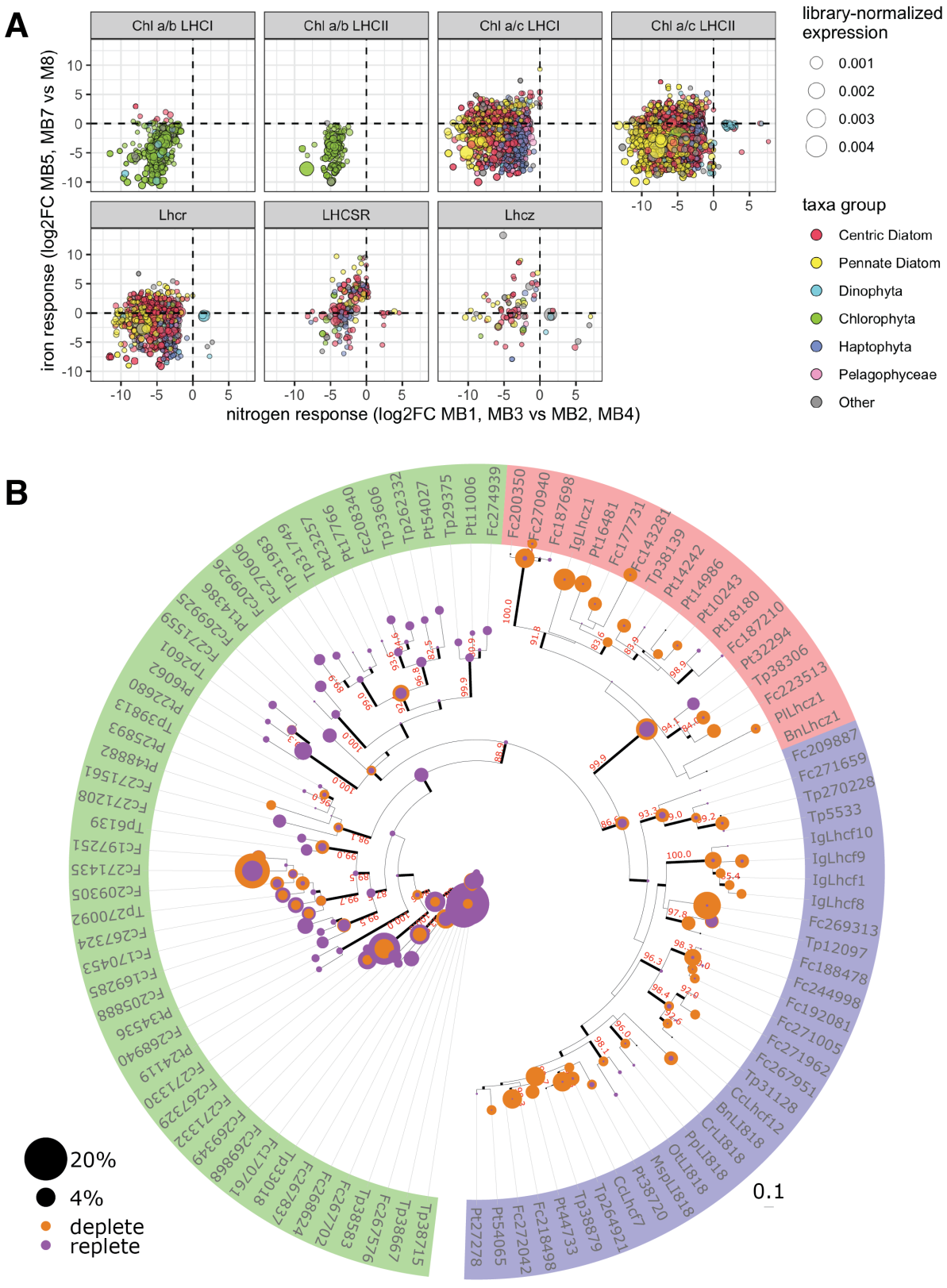


Figure 2.5: Changes in expression of major cellular energy acquisition machinery (light harvesting) and metabolic configuration of phytoplankton in response to limiting iron or nitrogen. (**A**) Log2FC of average light harvesting and photosynthetic electron transport cluster expression for each taxa group. Scale is -10 to 10 but omitted for visual clarity. (**B**) Cluster differential expression calculated for all taxa together, with clusters colored by pathway/function. (**C**) Integrated model of energetic inputs and pathway flux (inferred from transcripts).

Figure 2.6: Responsiveness of light harvesting complex (LHC) to nutrient stress. (**A**) Response of LHC subfamilies to iron (y-axis; >0 is up in low iron; <0 is down in low iron) and nitrogen (x-axis; >0 is up in low nitrogen; <0 is down in low nitrogen). Chlorophyll is abbreviated "Chl." Only LHC ORFs that are significantly differentially expressed (FDR < 0.05) in at least one condition are shown. (**B**) Phylogenetic tree showing expression of LHC transcripts in nutrient replete conditions (MB1, MB3, MB7; purple circles) versus nutrient deplete conditions (MB2, MB4, MB8; orange circles). The outer ring of the phylogenetic tree is colored by LHC family (green = non-stress responsive; red = Lhcz; blue = Lhcsr (a.k.a. Lhcx, LI818)).


A
Chl a/b LHCI
Chl a/b LHCII
Chl a/c LHCI
Chl a/c LHCII
Lhcr
LHCSR
Lhcz
iron response (log2FC MB5, MB7 vs M8)
nitrogen response (log2FC MB1, MB3 vs MB2, MB4)
library-normalized expression
0.001
0.002
0.003
0.004
taxa group
Centric Diatom
Pennate Diatom
Dinophyta
Chlorophyta
Haptophyta
Pelagophyceae
Other
B
20%
4%
deplete
replete
0.1

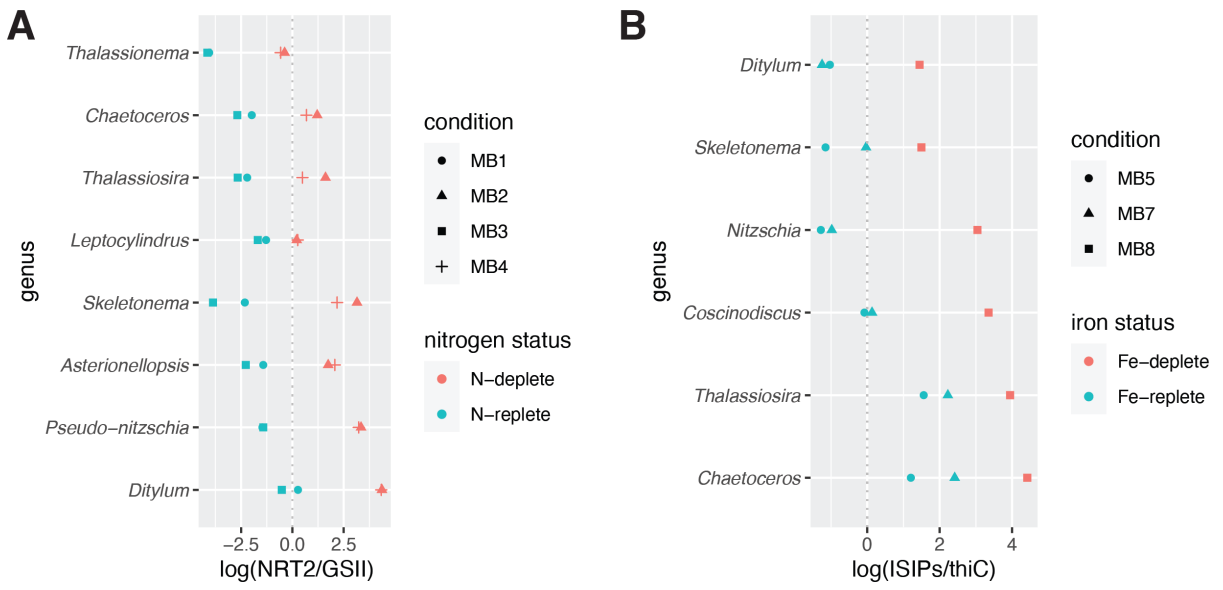


Figure 2.7: Environmental gene markers of cellular (**A**) nitrogen and (**B**) iron status in diatoms. A) Log ratio of diatom nitrate transporter (NRT2; MCL cluster 351) to glutamine synthetase (GSII; MCL cluster 139) gene expression, averaged by genus. B) Log ratio of summed diatom iron starvation induced proteins (ISIP1, uniprot B7GA90; ISIP2a, uniprot B7FYL2; ISIP2b, uniprot B7G9B1; ISIP3, uniprot B7G4H8) to thiC (K03147) gene expression, averaged by genus.

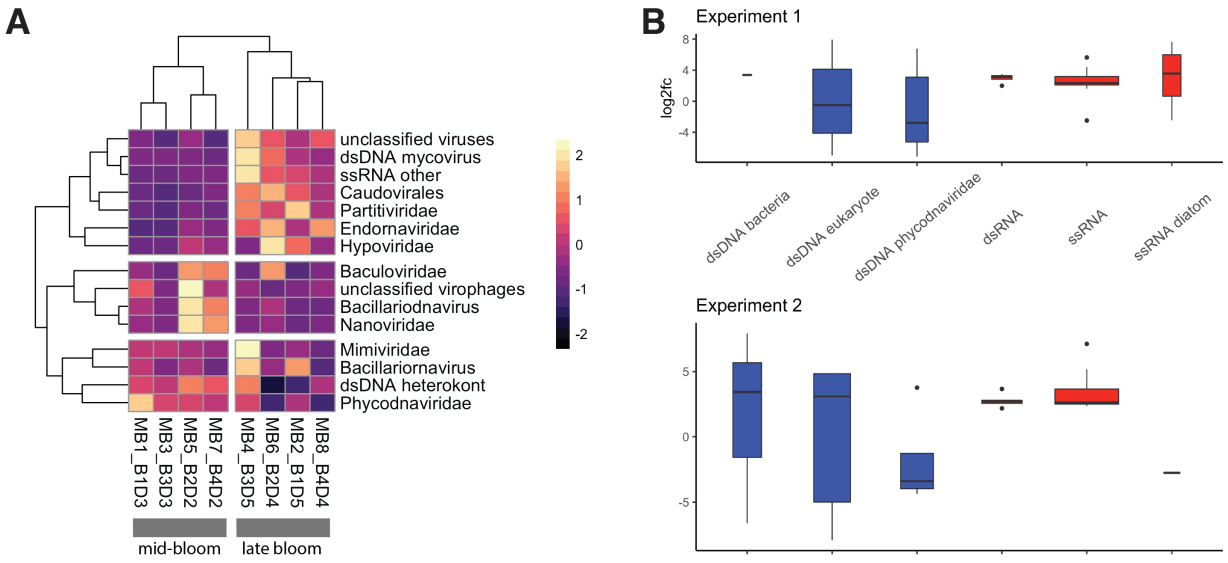


Figure 2.8: (**A**) Heatmap of most abundant viral taxa across experimental conditions (**B**) Log2FC in expression of late-bloom conditions relative to early bloom conditions for both experiments. DNA viruses are shown in blue and RNA viruses are shown in red.

Supplementary Figures

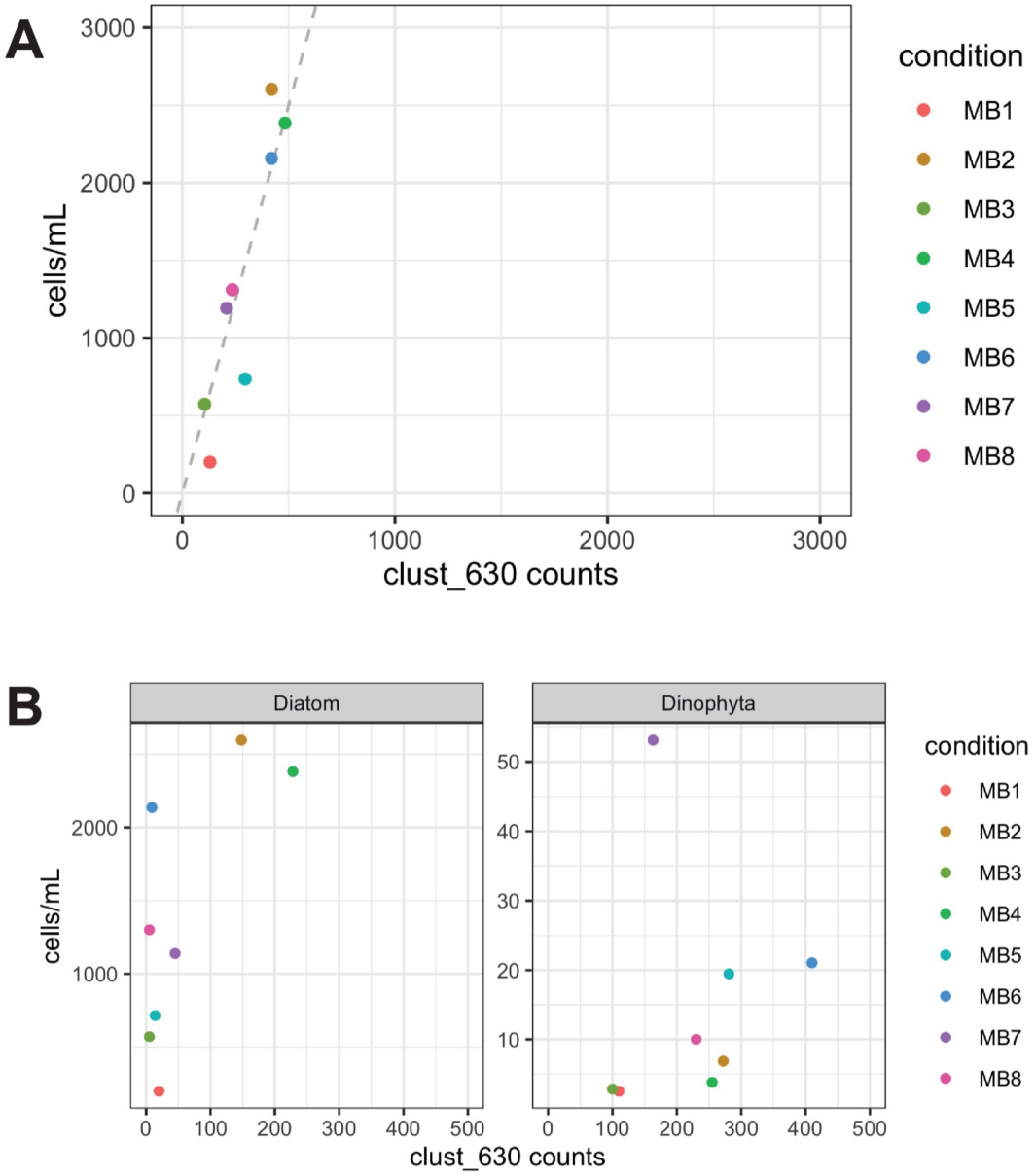


Figure S1 Correlation between cluster 630 counts and cell counts across sampling conditions (MB1 – MB8; colors, right) for diatoms and dinoflagellates (**A**) taken together and (**B**) separately.

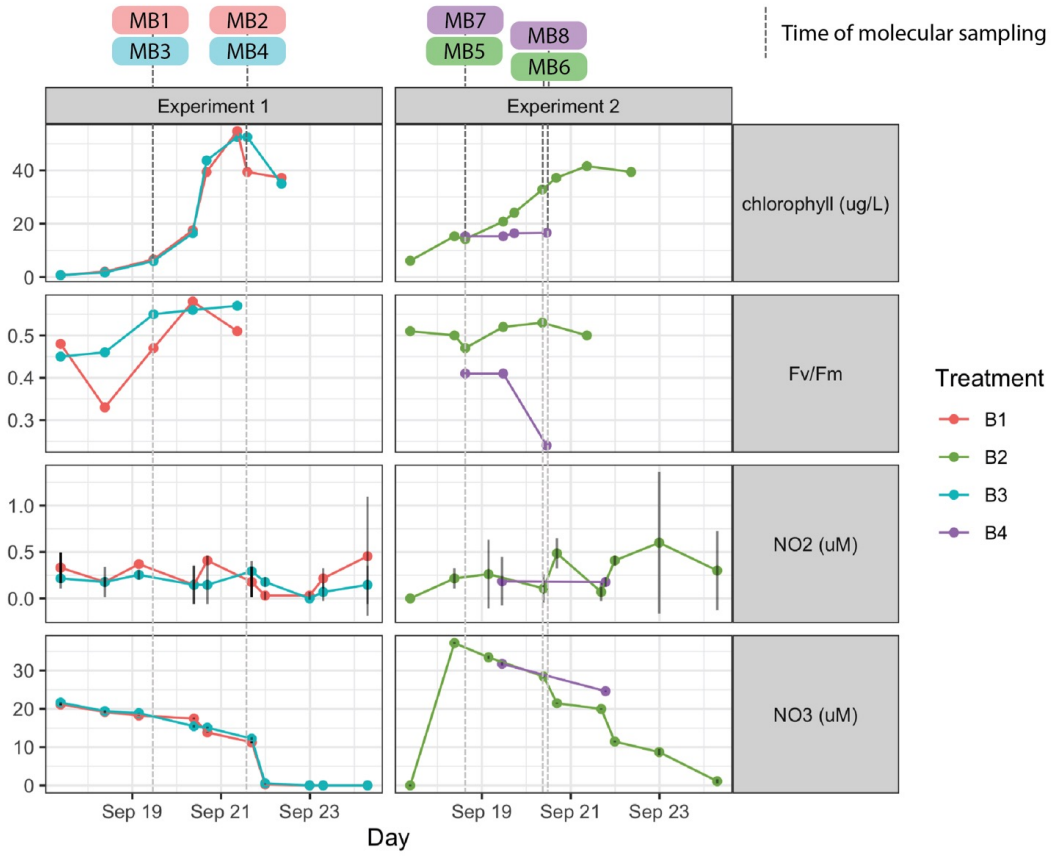


Figure S2 Chlorophyll, Fv/Fm (variable fluorescence/maximum fluorescence), nitrite, and nitrate measurements across time for all experimental conditions. Experiment one consists of barrels B1 and B3 (red and blue), and experiment 2 consists of barrels B2 and B4 (green and purple). Timing of molecular samples are given by dotted grey vertical lines.

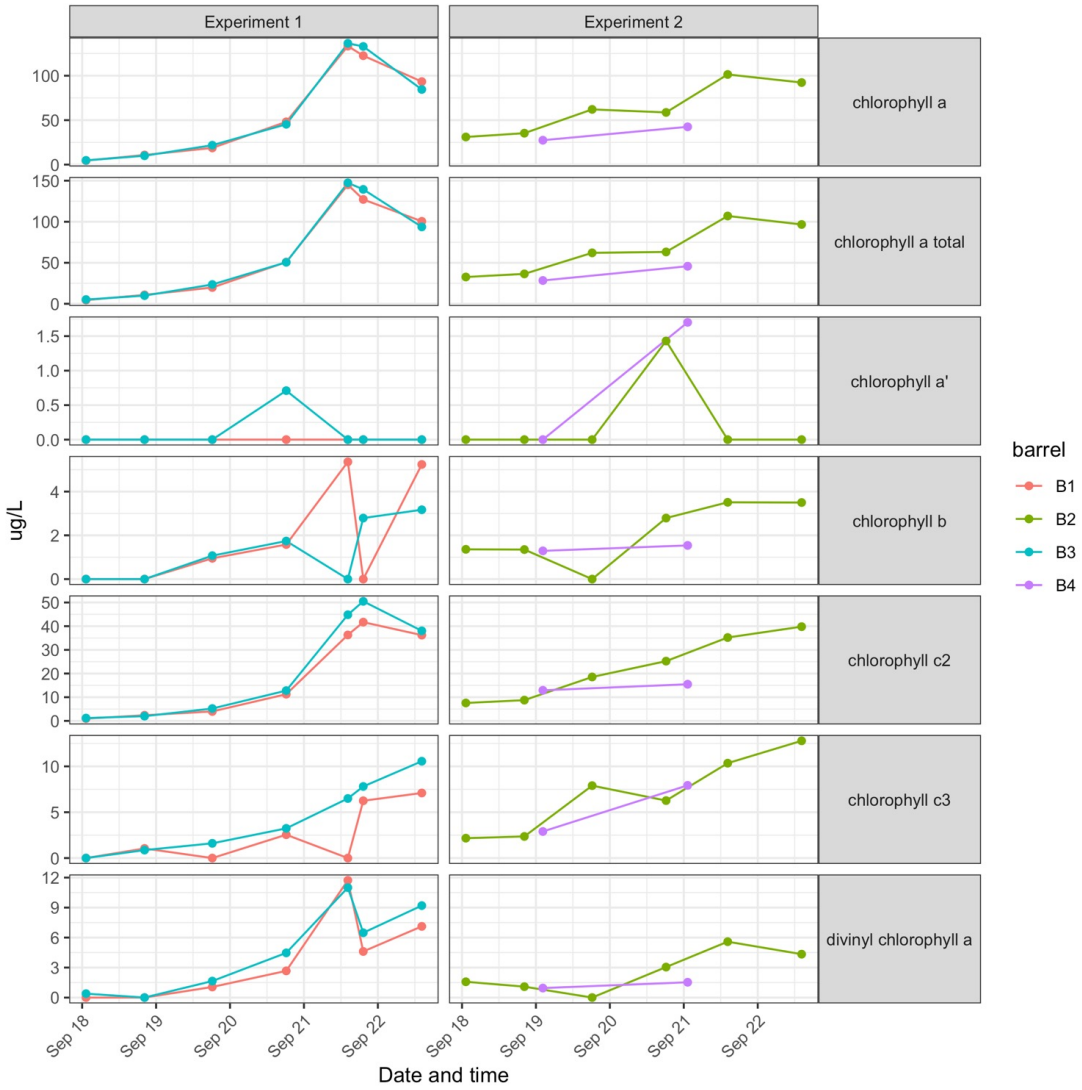


Figure S3 Chlorophyll pigment concentrations over time for experiments 1 and 2.

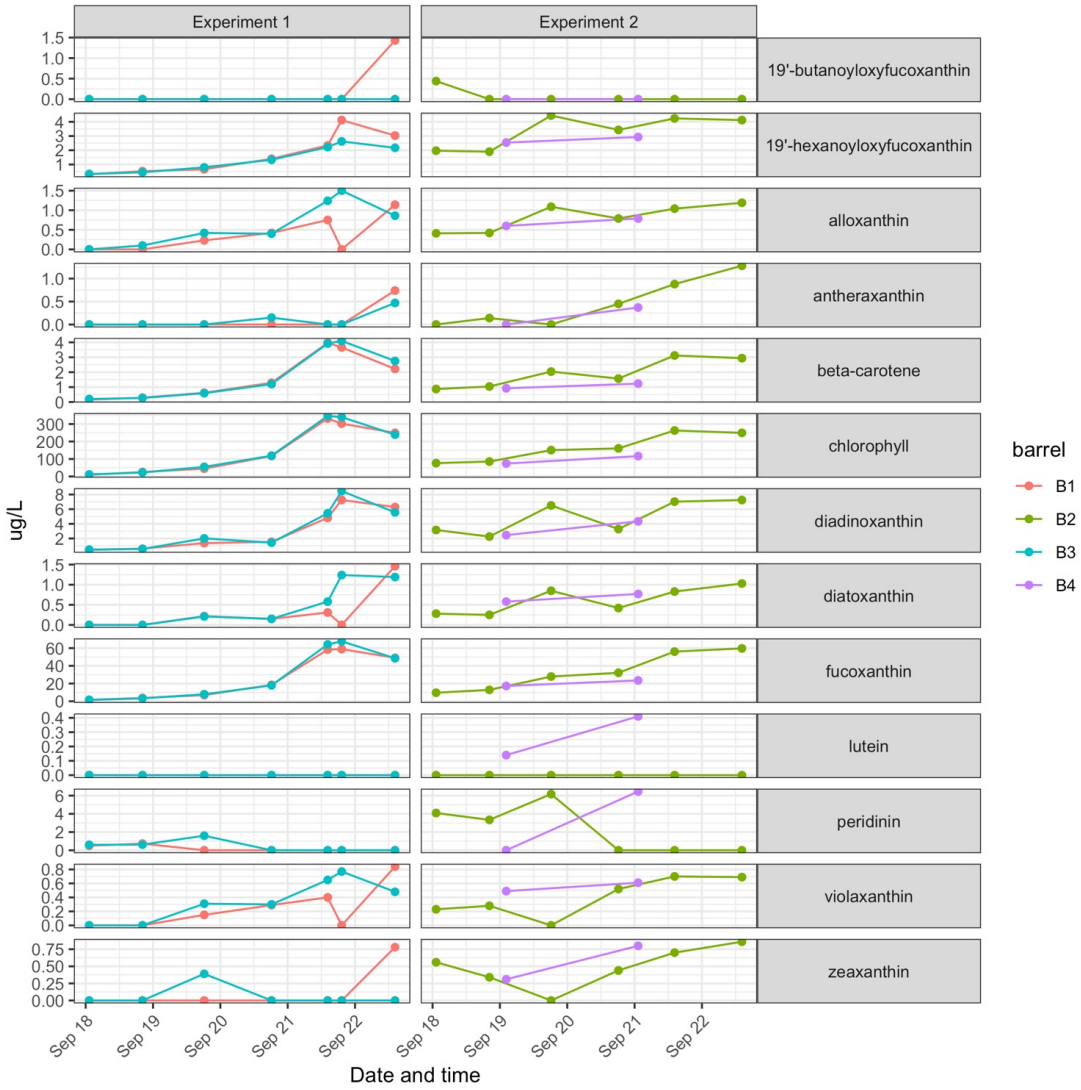
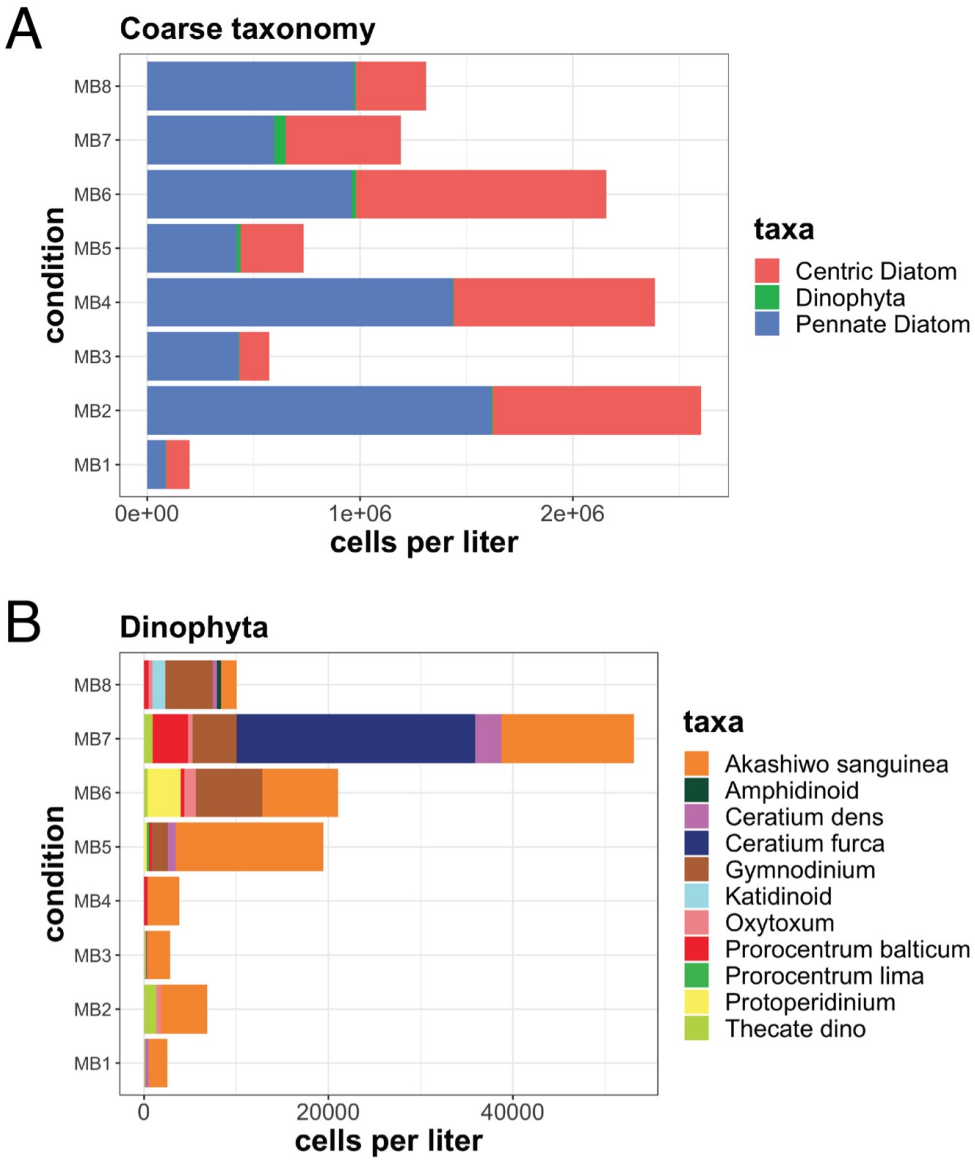


Figure S4 Pigment concentrations over time for experiments 1 and 2.


A
Coarse taxonomy
MB8
MB7
MB6
MB5
MB4
MB3
MB2
MB1
condition
0e+00
1e+06
2e+06
cells per liter
taxa
Centric Diatom
Dinophyta
Pennate Diatom
B
Dinophyta
MB8
MB7
MB6
MB5
MB4
MB3
MB2
MB1
condition
0
20000
40000
cells per liter
taxa
Akashiwo sanguinea
Amphidinoid
Ceratium dens
Ceratium furca
Gymnodinium
Katidinoid
Oxytoxum
Prorocentrum balticum
Prorocentrum lima
Protoperidinium
Thecate dino

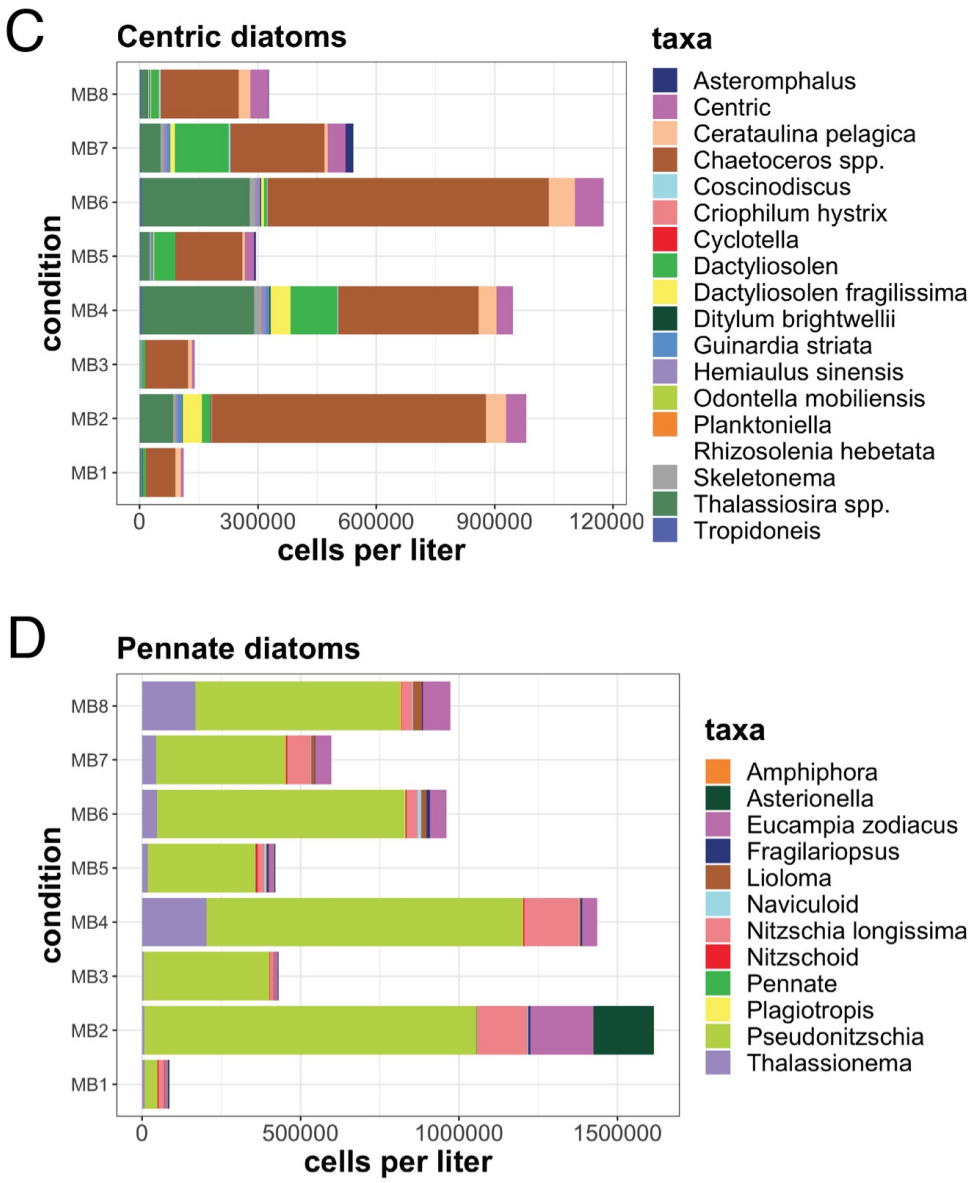


Figure S5 Microscopy-based phytoplankton cell counts across conditions (MB1-8). (A) coarse taxonomy; (B) dinoflagellate groups; (C) centric diatom groups; (D) pennate diatom groups.

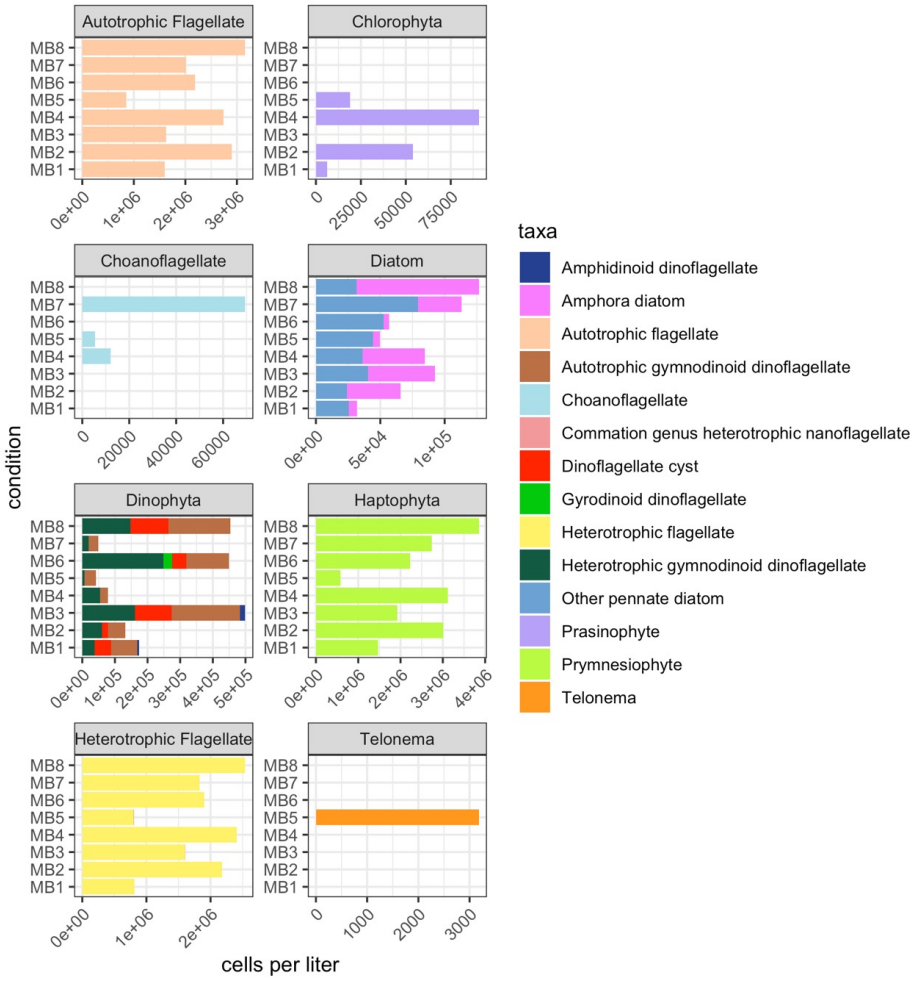


Figure S6 Microscopy-based counts of small cells recovered from filtered samples across conditions (MB1-8).

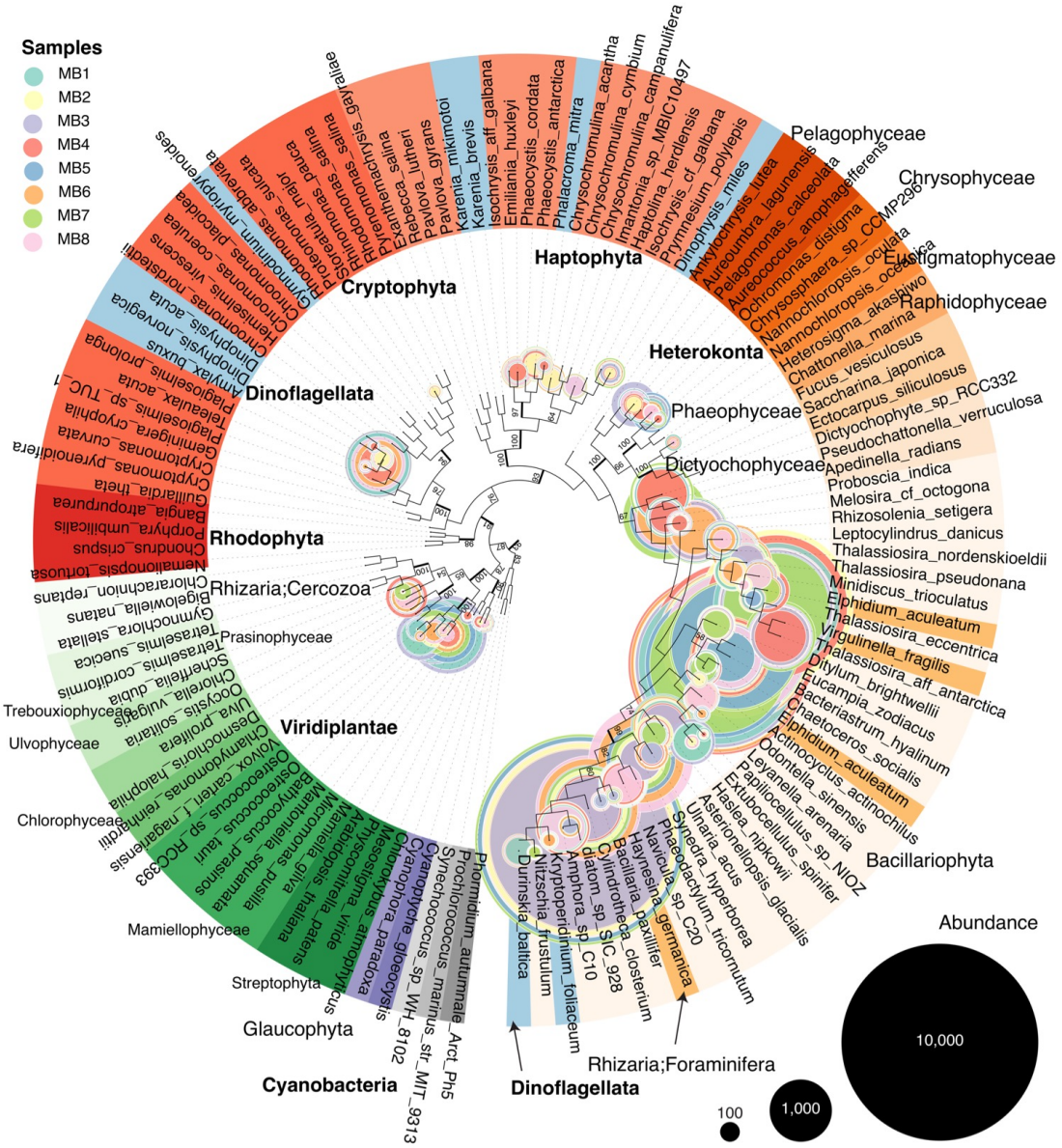


Figure S7 Phylogenetic tree showing distribution of plastids derived from 16S rRNA amplicons. Circles representing amplicon abundance are colored by sample origin (MB1-8) and superimposed over a reference phylogeny which is colored by taxonomy. Proximity of circles to the tips of branches represents closeness to references.

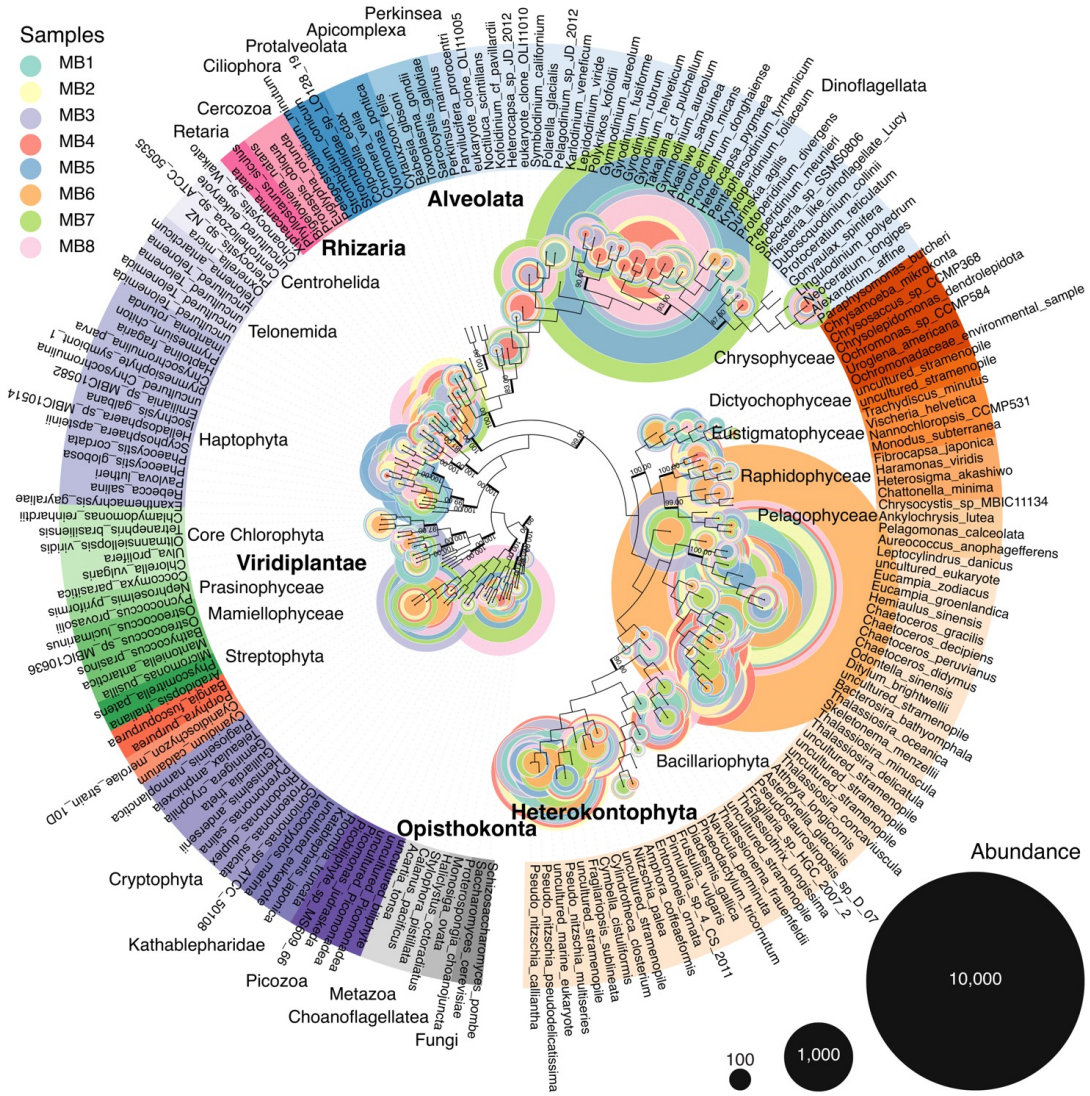


Figure S8 Phylogenetic tree showing distribution of active eukaryotic taxa from 18S rRNA amplicons. Circles representing amplicon abundance are colored by sample origin (MB1-8) and superimposed over a reference phylogeny which is colored by taxonomy. Proximity of circles to the tips of branches represents closeness to references.

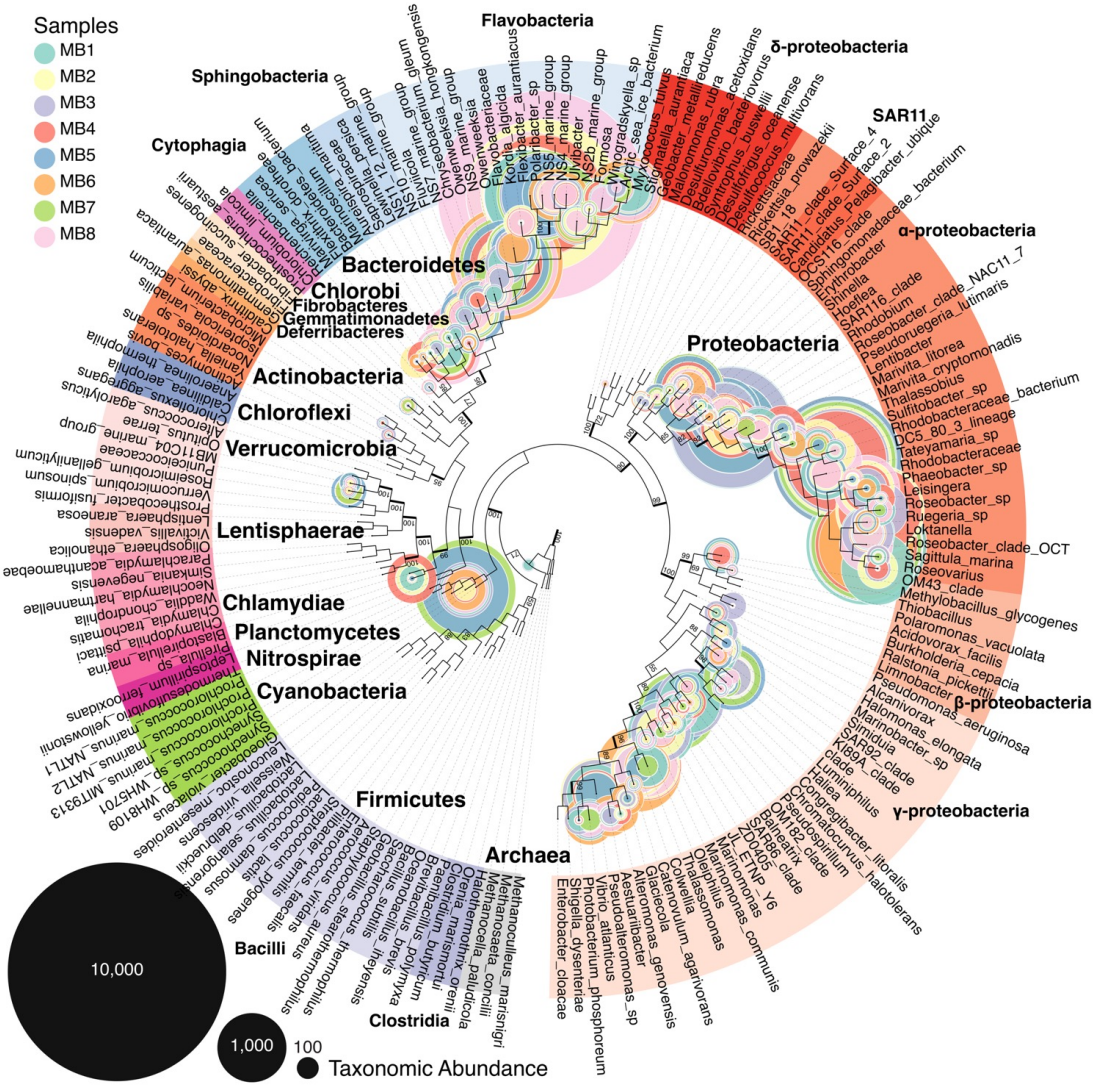
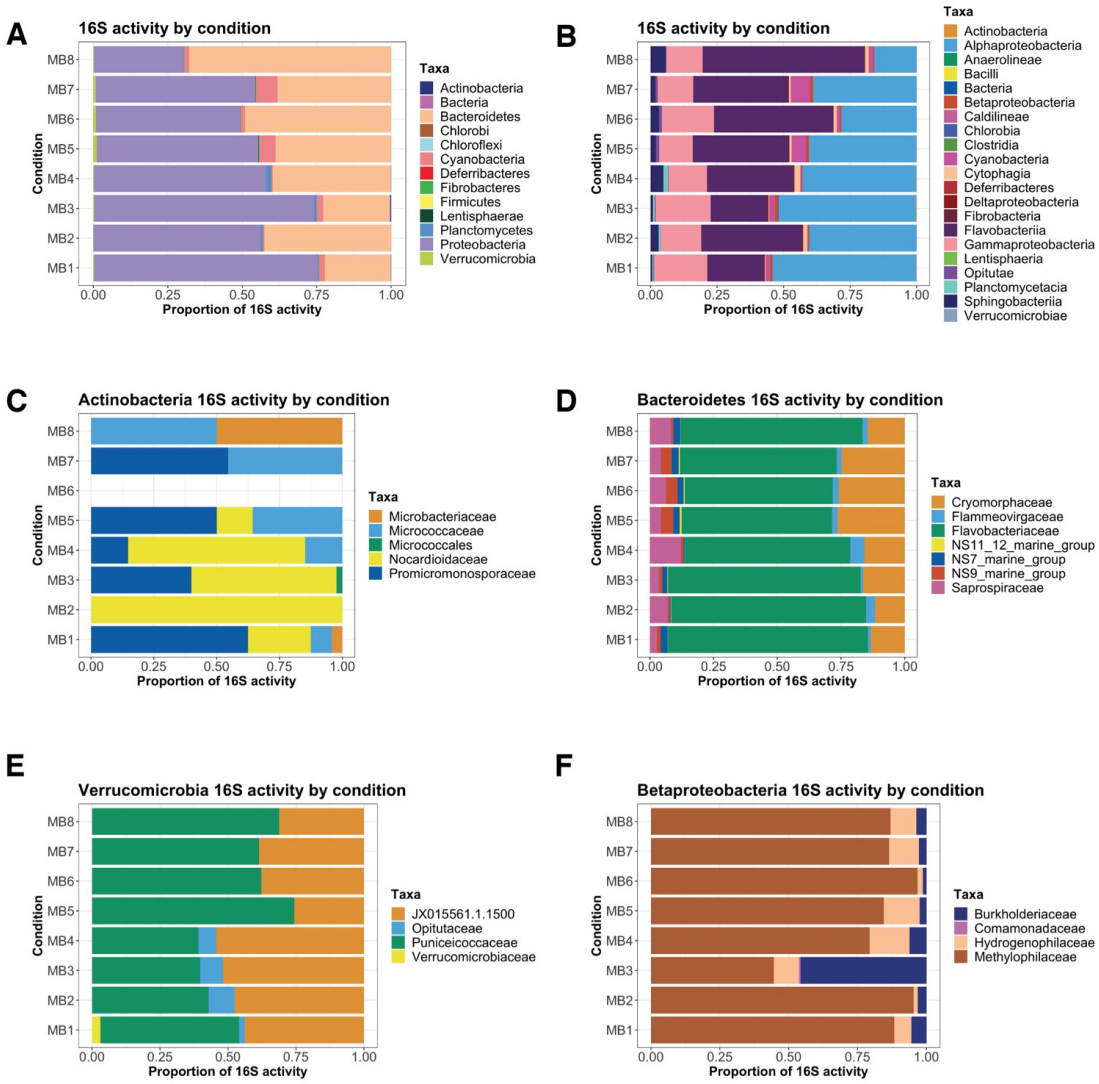


Figure S9 Phylogenetic tree showing distribution of active bacterial taxa using 16S rRNA amplicons. Circles representing amplicon abundance are colored by sample origin (MB1-8) and superimposed over a reference phylogeny which is colored by taxonomy. Proximity of circles to the tips of branches represents closeness to references.


A
16S activity by condition
Condition
MB8
MB7
MB6
MB5
MB4
MB3
MB2
MB1
0.00
0.25
0.50
0.75
1.00
Proportion of 16S activity
Taxa
Actinobacteria
Bacteria
Bacteroidetes
Chlorobi
Chloroflexi
Cyanobacteria
Deferribacteres
Fibrobacteres
Firmicutes
Lentisphaerae
Planctomycetes
Proteobacteria
Verrucomicrobia
B
16S activity by condition
Taxa
Actinobacteria
Alphaproteobacteria
Anaerolineae
Bacilli
Bacteria
Betaproteobacteria
Caldilineae
Chlorobia
Clostridia
Cyanobacteria
Cytophagia
Deferribacteres
Deltaproteobacteria
Fibrobacteria
Flavobacteriia
Gammaproteobacteria
Lentisphaeria
Opitutae
Planctomycetacia
Sphingobacteriia
Verrucomicrobiae
C
Actinobacteria 16S activity by condition
Taxa
Microbacteriaceae
Micrococcaceae
Micrococcales
Nocardioidaceae
Promicromonosporaceae
D
Bacteroidetes 16S activity by condition
Taxa
Cryomorphaceae
Flammeovirgaceae
Flavobacteriaceae
NS11_12_marine_group
NS7_marine_group
NS9_marine_group
Saprospiraceae
E
Verrucomicrobia 16S activity by condition
Taxa
JX015561.1.1500
Opitutaceae
Puniceicoccaceae
Verrucomicrobiaceae
F
Betaproteobacteria 16S activity by condition
Taxa
Burkholderiaceae
Comamonadaceae
Hydrogenophilaceae
Methylophilaceae

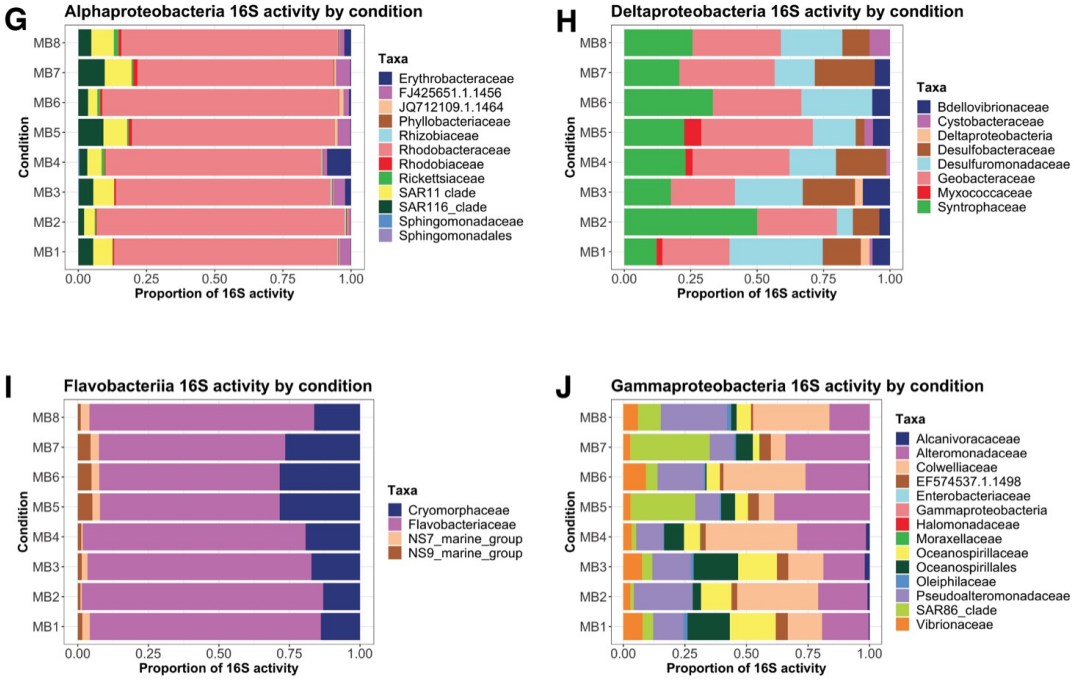


Figure S10 Structure of the active 16S bacterial community across conditions. Community structure is shown at both a high taxonomic level (A, B) and also within groups (C-J).

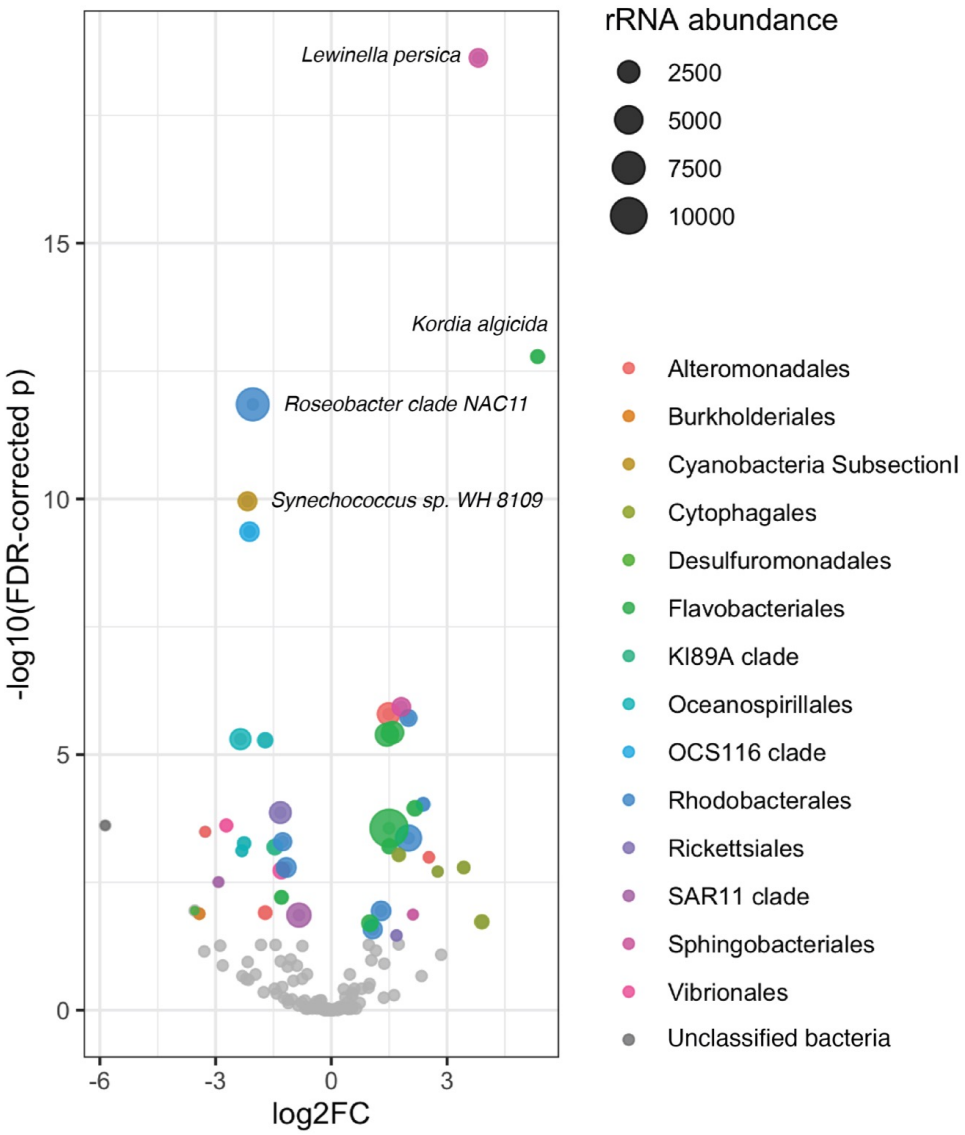


Figure S11 Differential abundance of rRNA amplicons pertaining to bacterial species across nitrogen conditions of experiment 1. X-axis shows log2 fold change in abundance across conditions (negative = up in mid-bloom conditions MB1 and MB3; positive = up in late-bloom conditions MB2 and MB4). Significantly differentially abundant taxa are shown by colored bubbles; insignificant taxa (FDR >0.05) are shown in light grey.

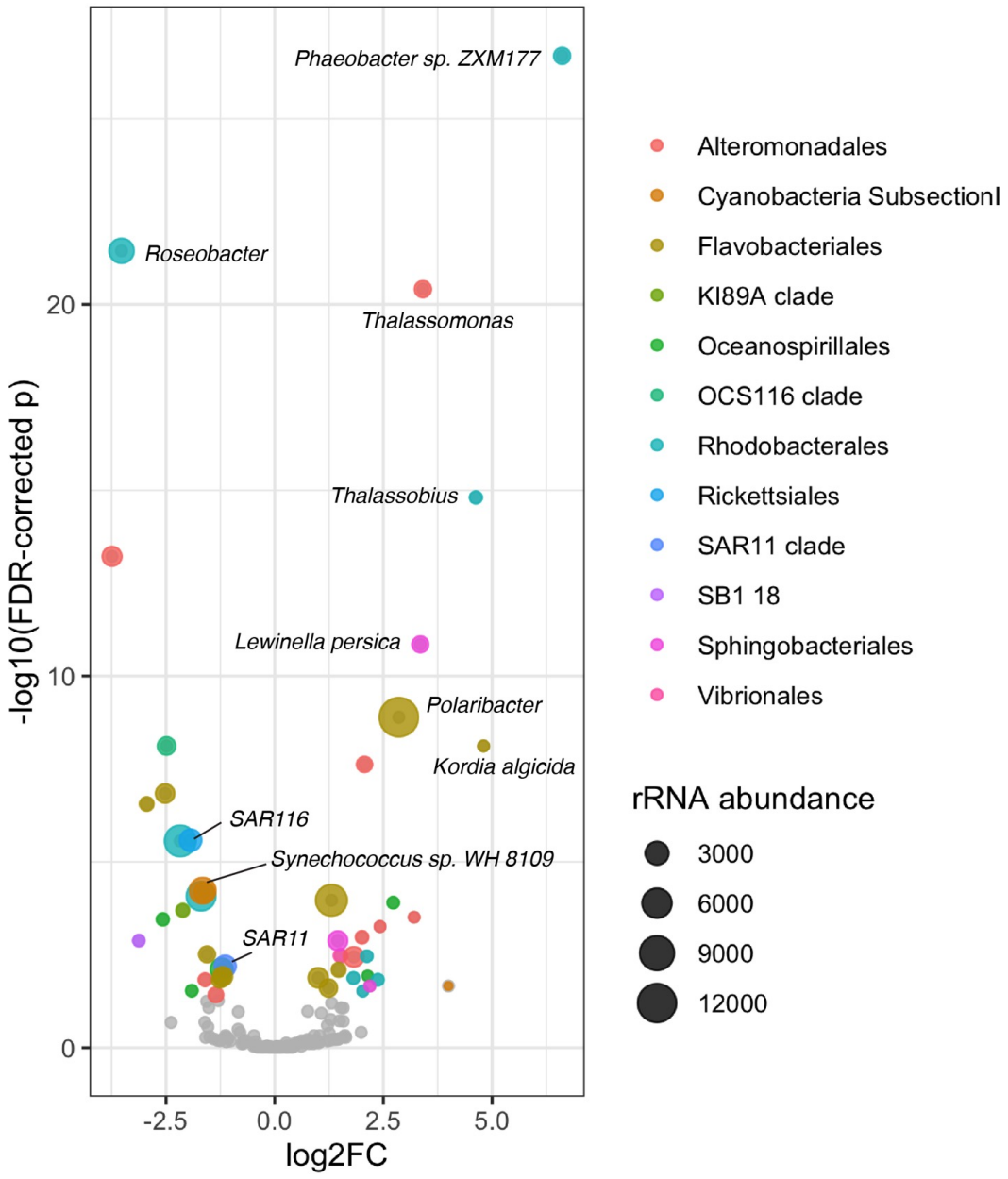


Figure S12 Differential abundance of rRNA amplicons pertaining to bacterial species across iron conditions of experiment 2. X-axis shows log2 fold change in abundance across conditions (negative = up in Fe-replete conditions MB5 and MB7; positive = up in Fe-deplete condition MB8). Significantly differentially abundant taxa are shown by colored bubbles; insignificant taxa (FDR >0.05) are shown in light grey.

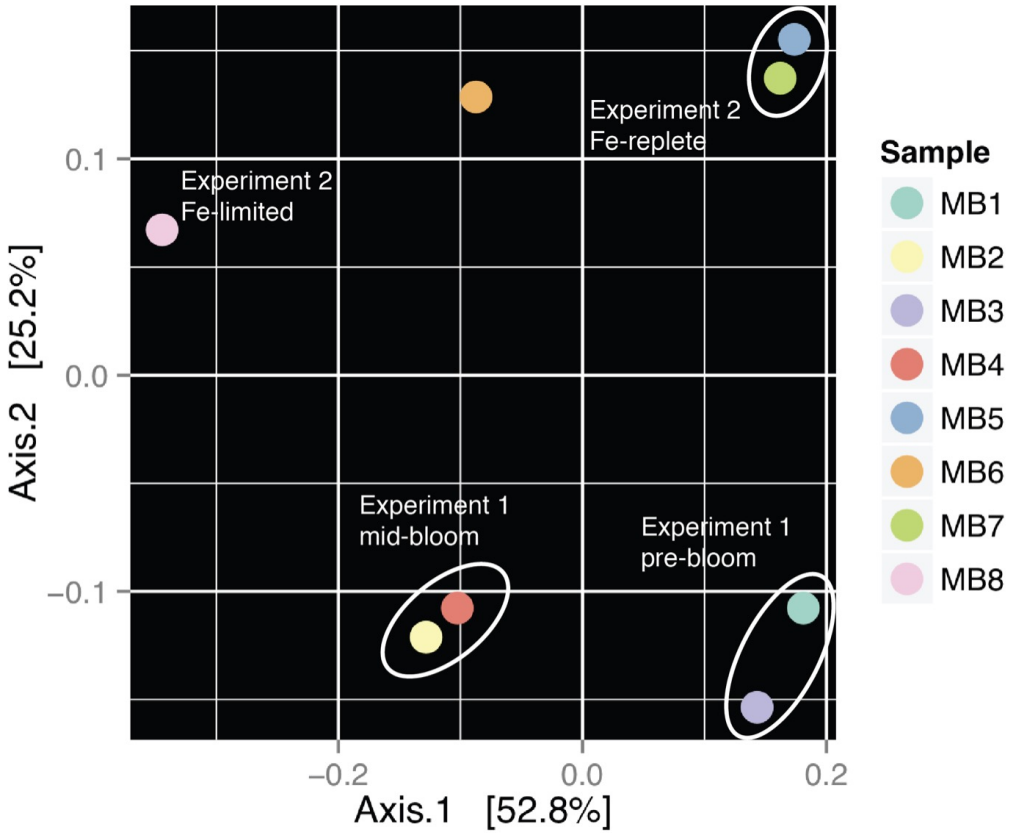
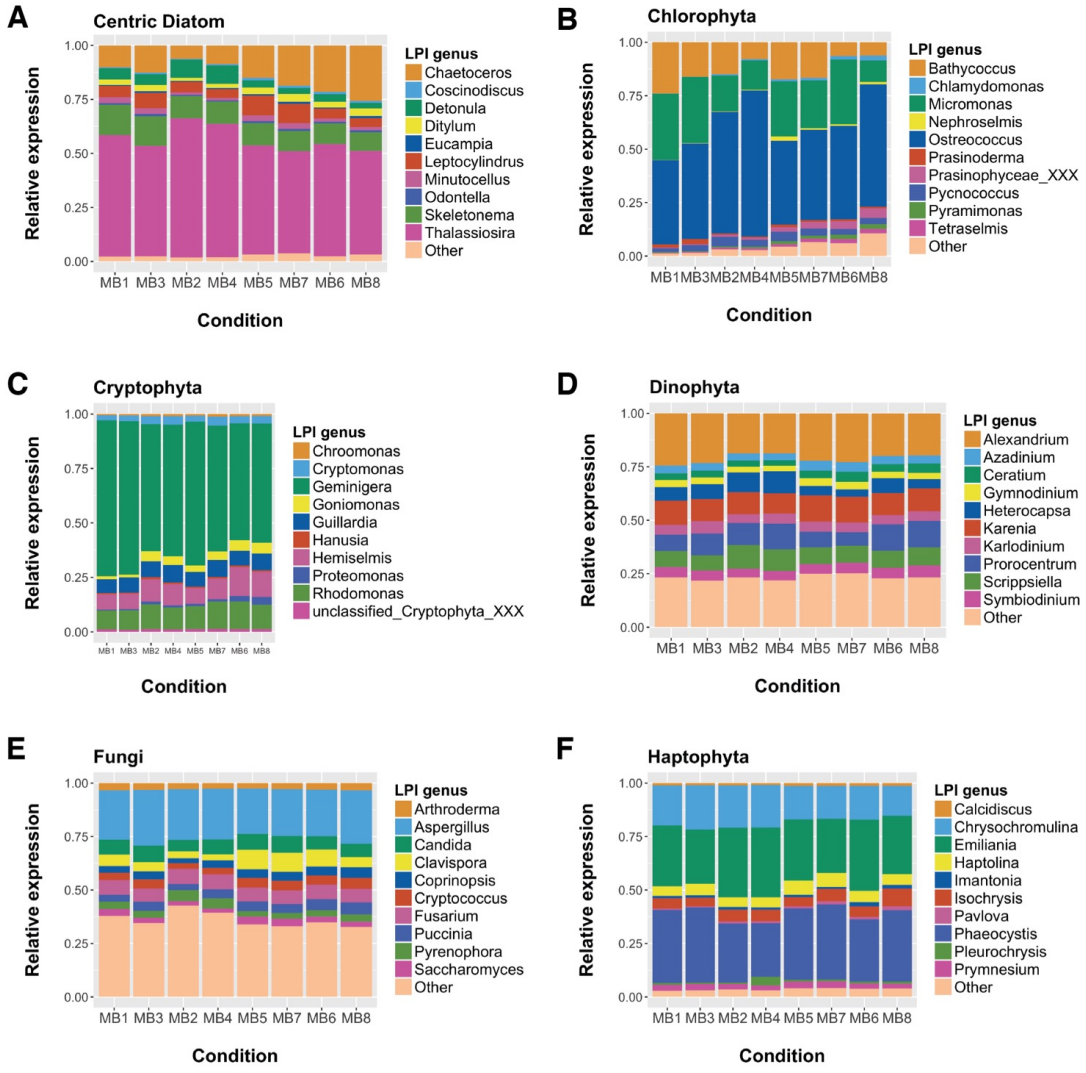


Figure S13 Ordination of bacterial rRNA counts shows experiment 1 pre-bloom conditions (MB1, MB3) clustering together, experiment 1 mid-bloom conditions (MB2, MB4) clustering together, and experiment 2 Fe-replete conditions (MB5, MB7) clustering together.


A
Centric Diatom
Relative expression
1.00
0.75
0.50
0.25
0.00
MB1 MB3 MB2 MB4 MB5 MB7 MB6 MB8
Condition
LPI genus
Chaetoceros
Coscinodiscus
Detonula
Ditylum
Eucampia
Leptocylindrus
Minutocellus
Odontella
Skeletonema
Thalassiosira
Other
B
Chlorophyta
Relative expression
1.00
0.75
0.50
0.25
0.00
MB1 MB3 MB2 MB4 MB5 MB7 MB6 MB8
Condition
LPI genus
Bathycoccus
Chlamydomonas
Micromonas
Nephroselmis
Ostreococcus
Prasinoderma
Prasinophyceae_XXX
Pycnococcus
Pyramimonas
Tetraselmis
Other
C
Cryptophyta
Relative expression
1.00
0.75
0.50
0.25
0.00
MB1 MB3 MB2 MB4 MB5 MB7 MB6 MB8
Condition
LPI genus
Chroomonas
Cryptomonas
Geminigera
Goniomonas
Guillardia
Hanusia
Hemiselmis
Proteomonas
Rhodomonas
unclassified_Cryptophyta_XXX
D
Dinophyta
Relative expression
1.00
0.75
0.50
0.25
0.00
MB1 MB3 MB2 MB4 MB5 MB7 MB6 MB8
Condition
LPI genus
Alexandrium
Azadinium
Ceratium
Gymnodinium
Heterocapsa
Karenia
Karlodinium
Prorocentrum
Scrippsiella
Symbiodinium
Other
E
Fungi
Relative expression
1.00
0.75
0.50
0.25
0.00
MB1 MB3 MB2 MB4 MB5 MB7 MB6 MB8
Condition
LPI genus
Arthroderma
Aspergillus
Candida
Clavispora
Coprinopsis
Cryptococcus
Fusarium
Puccinia
Pyrenophora
Saccharomyces
Other
F
Haptophyta
Relative expression
1.00
0.75
0.50
0.25
0.00
MB1 MB3 MB2 MB4 MB5 MB7 MB6 MB8
Condition
LPI genus
Calcidiscus
Chrysochromulina
Emiliania
Haptolina
Imantonia
Isochrysis
Pavlova
Phaeocystis
Pleurochrysis
Prymnesium
Other

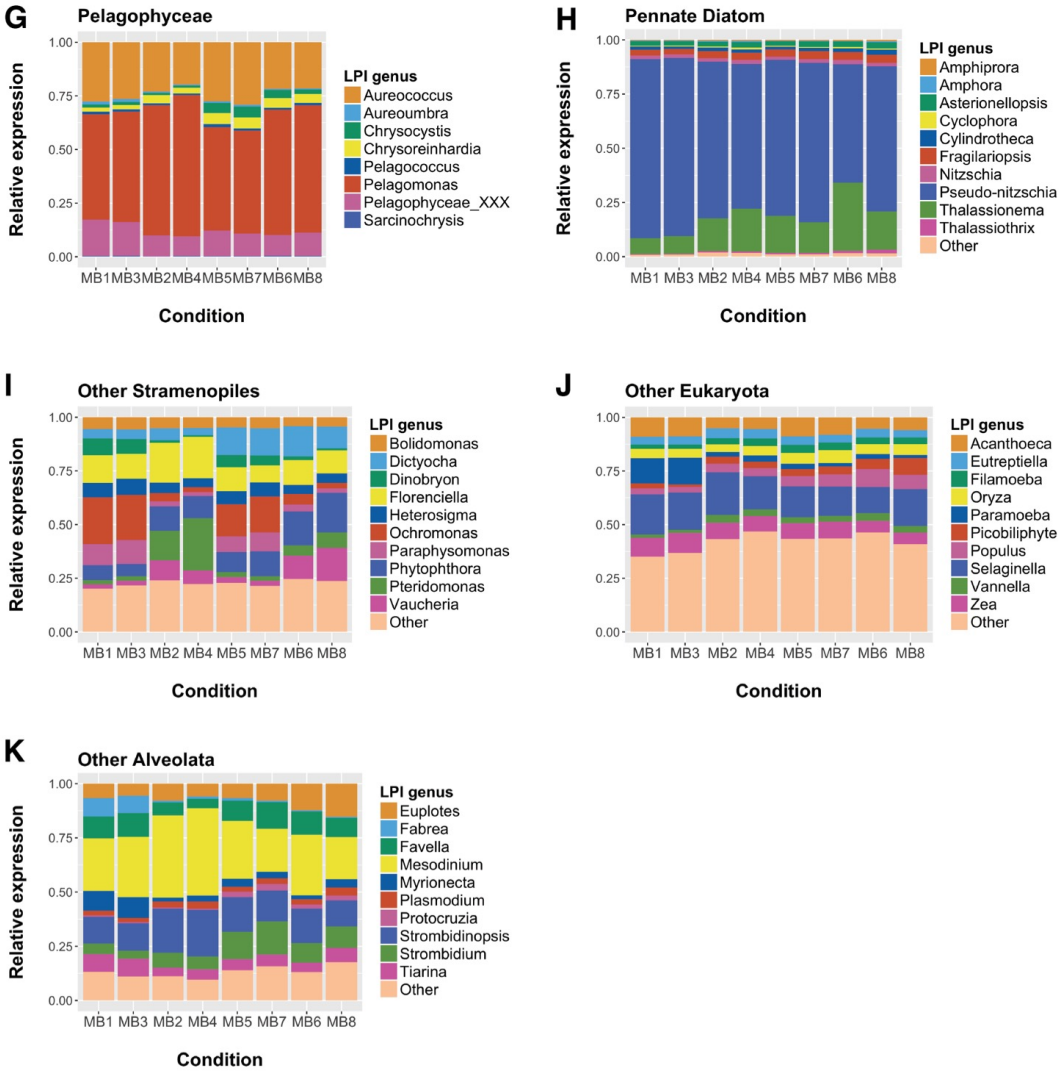


Figure S14 Genus-level taxonomic breakdown of total RNA of LPI taxa groups (A-K) across experimental conditions (x-axis; MB1, MB3 = N-replete; MB2, MB4 = N-deplete; MB5, MB7 = Fe-replete; MB8 = Fe-deplete). Taxonomic classification was determined by lineage probability index (LPI).

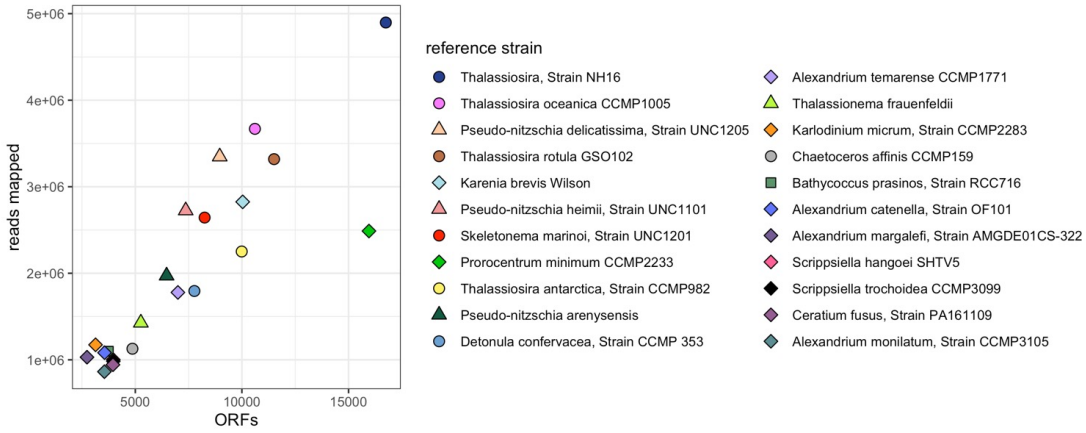
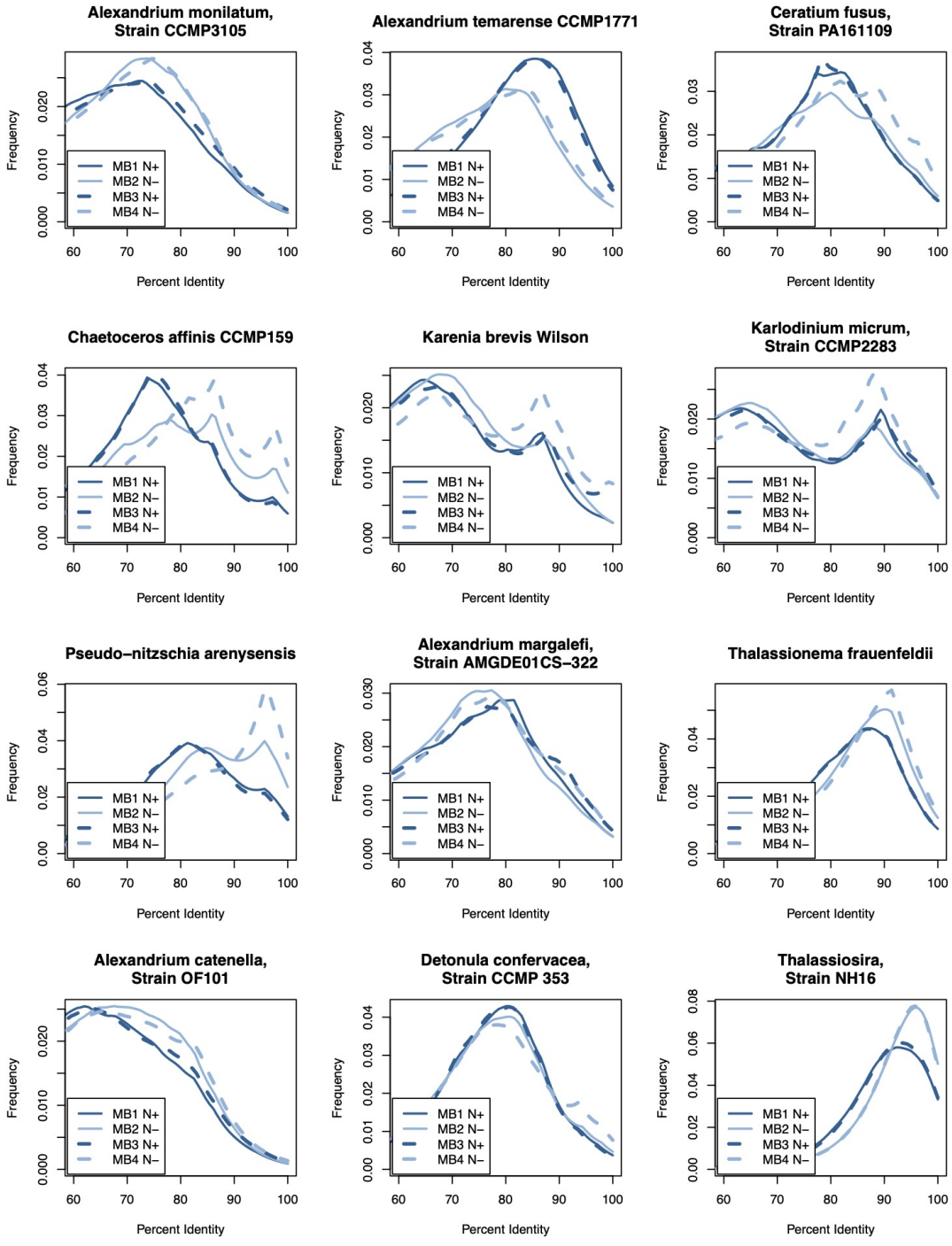


Figure S15 Reads (y-axis) mapping to and number of open reading frames (ORFs; x-axis) mapped to top 22 most abundantly hit MMETSP reference transcriptomes. Circles, triangles, and diamonds represent centric diatom, pennate diatom, and dinoflagellate references, respectively.


Alexandrium monilatum, Strain CCMP3105
Alexandrium temarense CCMP1771
Ceratium fusus, Strain PA161109
Chaetoceros affinis CCMP159
Karenia brevis Wilson
Karlodinium micrum, Strain CCMP2283
Pseudo–nitzschia arenysensis
Alexandrium margalefi, Strain AMGDE01CS–322
Thalassionema frauenfeldii
Alexandrium catenella, Strain OF101
Detonula confervacea, Strain CCMP 353
Thalassiosira, Strain NH16
Frequency
Percent Identity
MB1 N+
MB2 N–
MB3 N+
MB4 N–

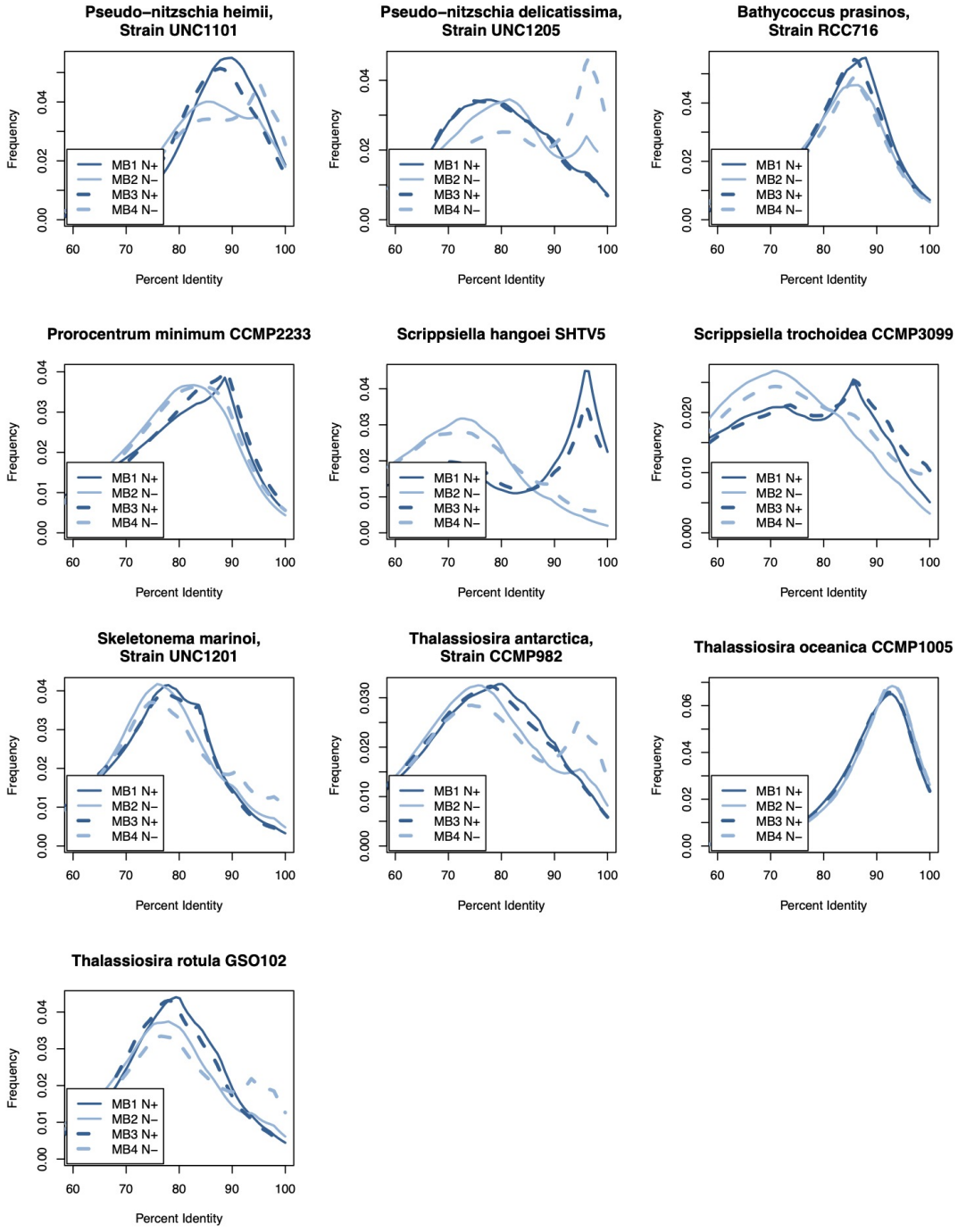
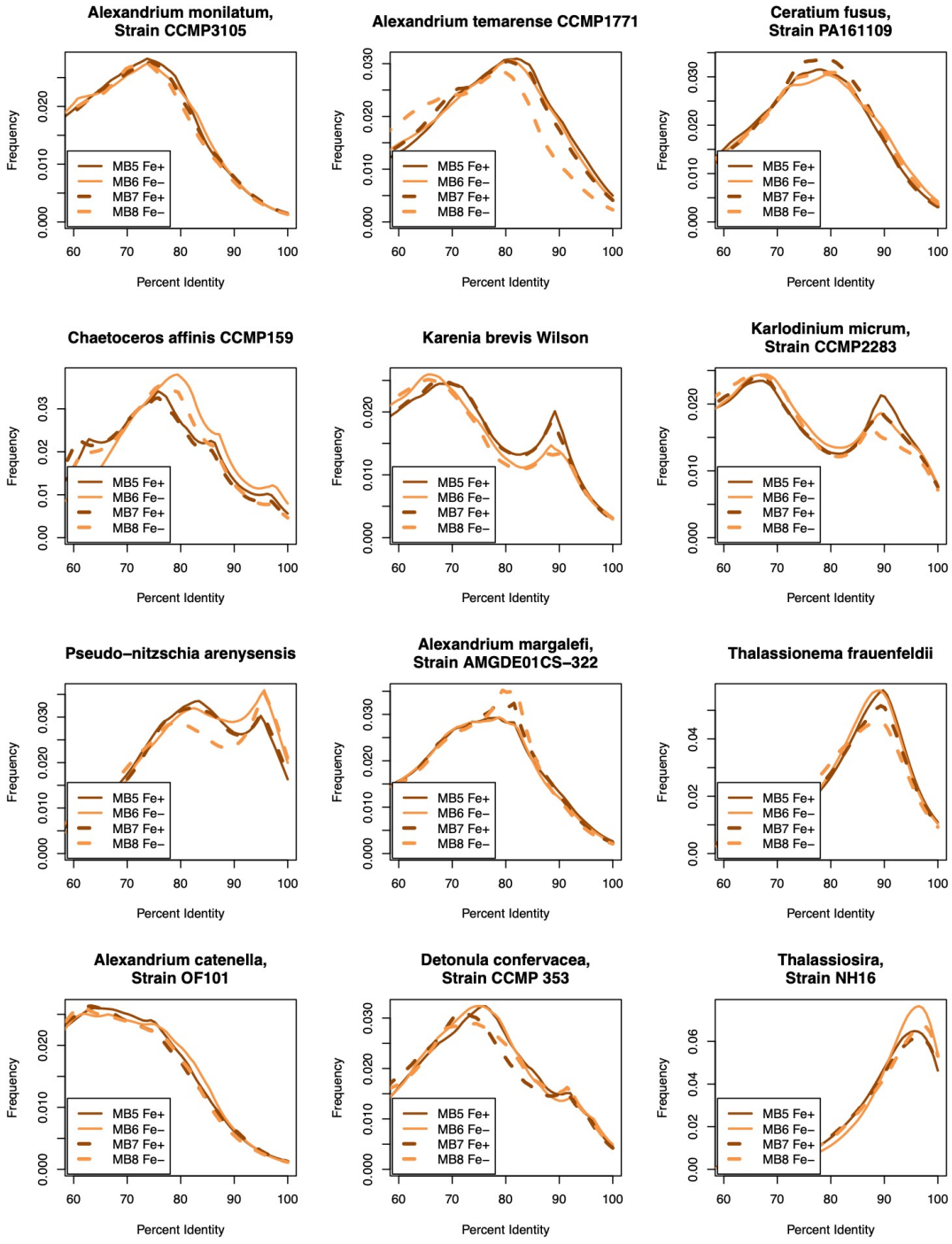


Figure S16 Percent identity histograms showing shift in distance to top 22 MMETSP references across nitrogen conditions.


Alexandrium monilatum, Strain CCMP3105
Alexandrium temarense CCMP1771
Ceratium fusus, Strain PA161109
Chaetoceros affinis CCMP159
Karenia brevis Wilson
Karlodinium micrum, Strain CCMP2283
Pseudo–nitzschia arenysensis
Alexandrium margalefi, Strain AMGDE01CS–322
Thalassionema frauenfeldii
Alexandrium catenella, Strain OF101
Detonula confervacea, Strain CCMP 353
Thalassiosira, Strain NH16
Frequency
Percent Identity
MB5 Fe+
MB6 Fe–
MB7 Fe+
MB8 Fe–

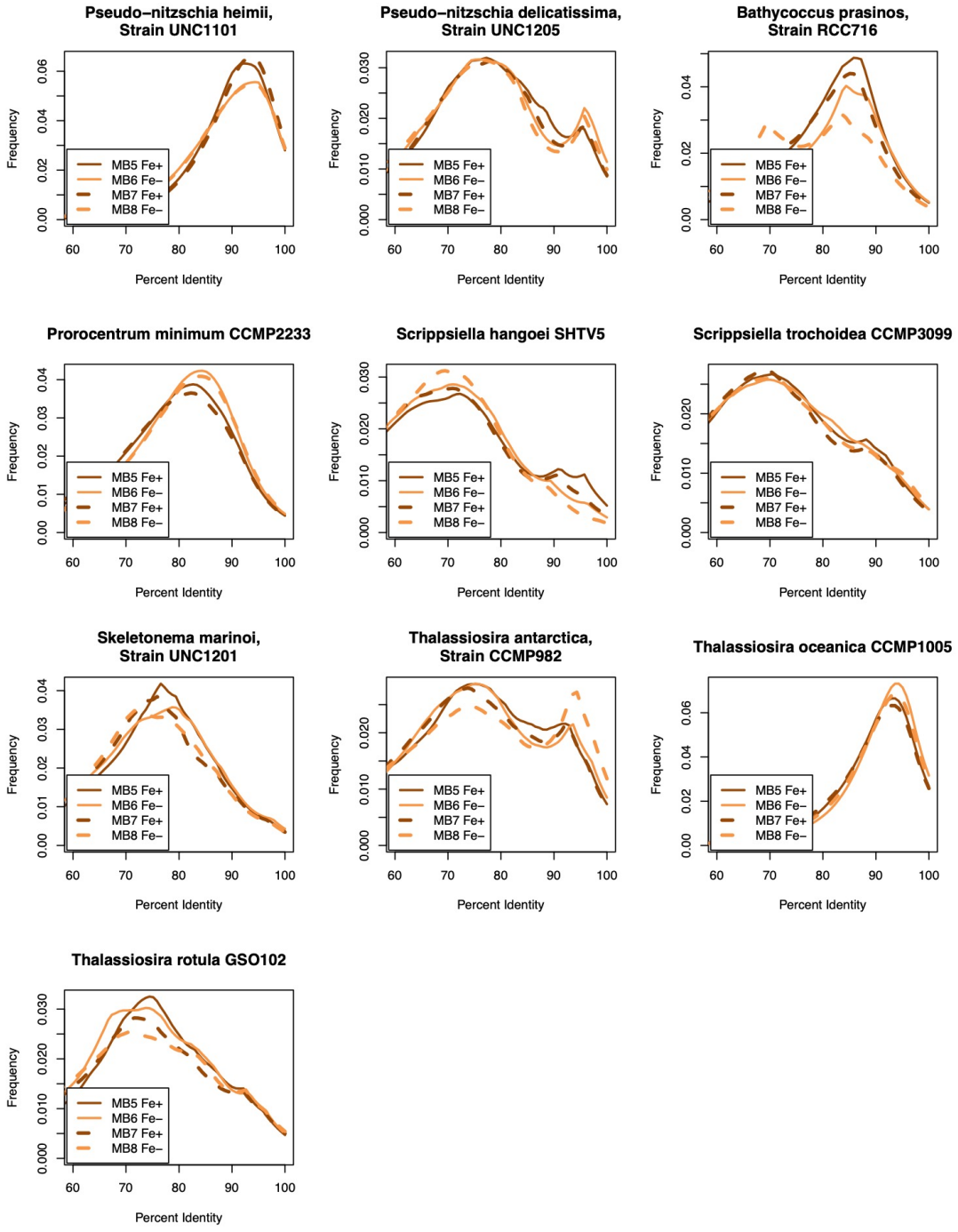


Figure S17 Percent identity histograms showing shift in distance to top 22 MMETSP references across iron conditions.

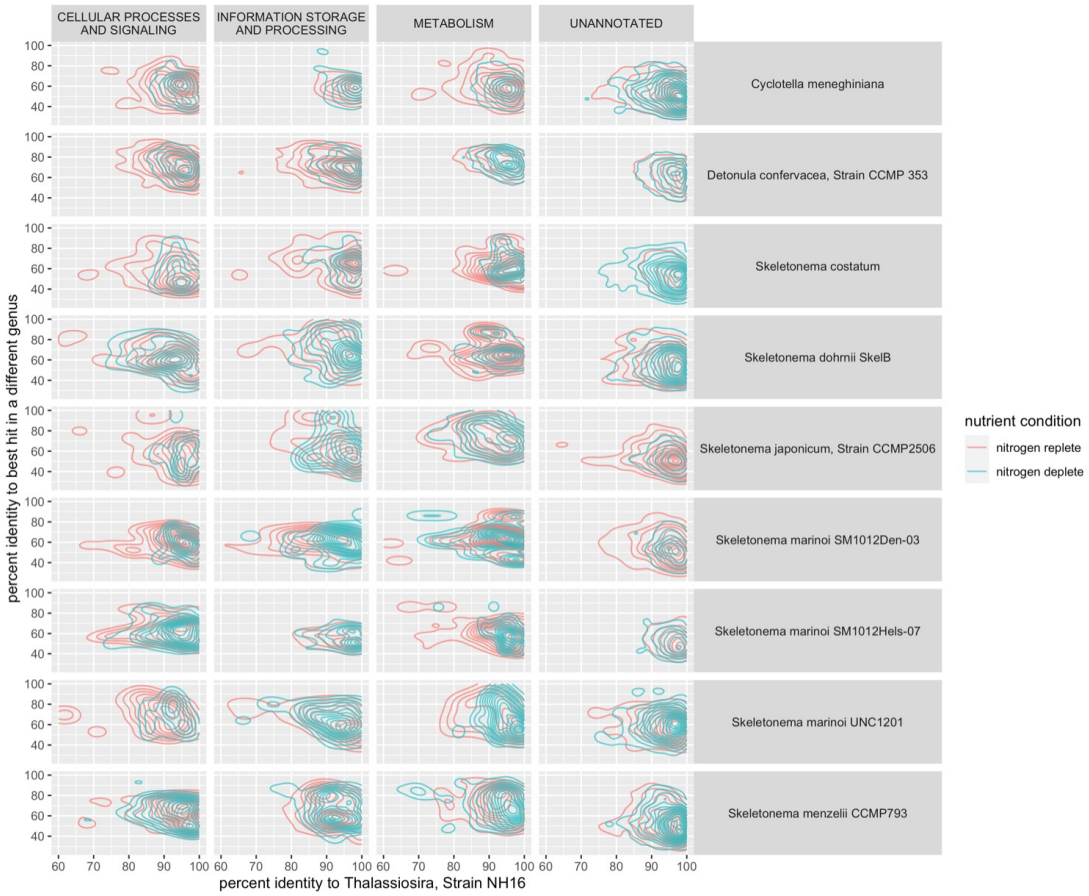


Figure S18 2D-histograms showing density of reads mapping to the most abundant centric diatom reference species, *Thalassiosira, Strain NH16* (x-axis) as well as closeness to the nearest species in a different genus (y-axis; right-hand facet-labels). Reads pertaining to different functions are separated out using Kegg Ortholog Groups (KOGs; top facet labels). Differences in mapping proximity can be seen between nutrient conditions (red: nitrogen replete; blue: nitrogen deplete).

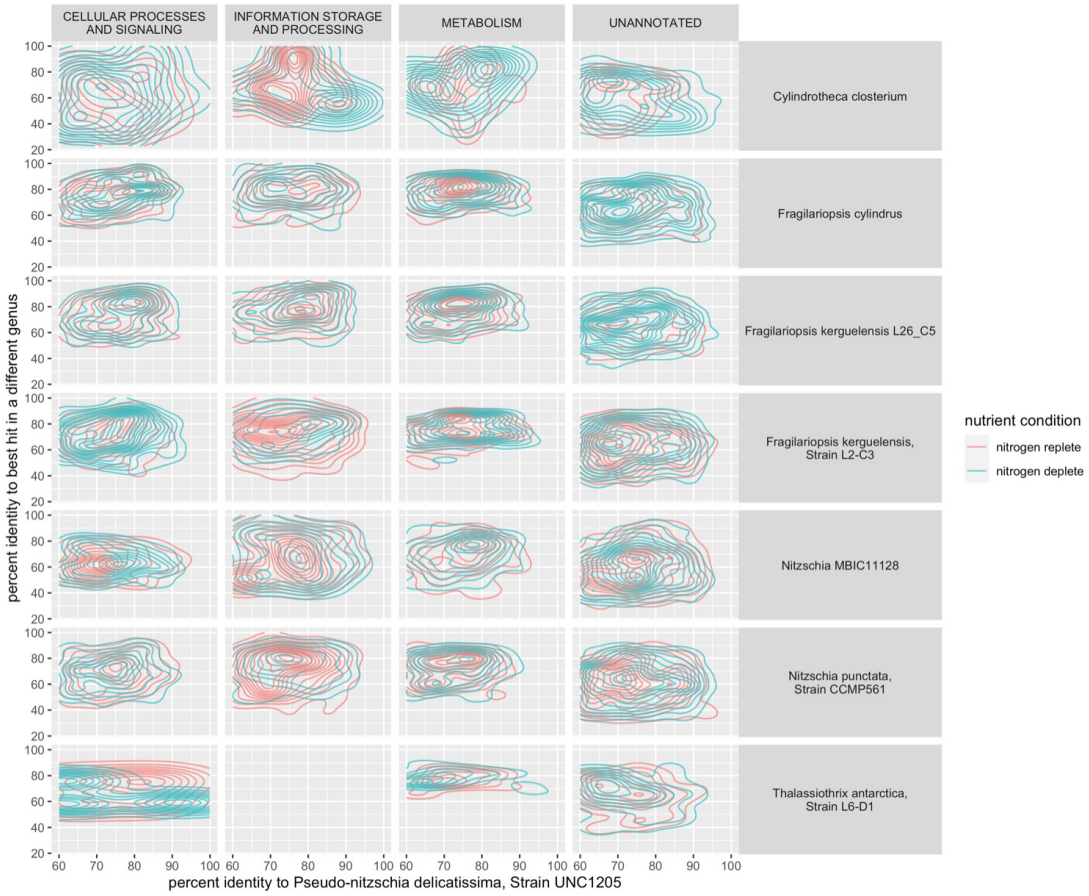


Figure S19 2D-histograms showing density of reads mapping to the most abundant pennate diatom reference species, *Pseudo-nitzschia delicatissima* (x-axis) as well as closeness to the nearest species in a different genus (y-axis; right-hand facet-labels). Reads pertaining to different functions are separated out using Kegg Ortholog Groups (KOGs; top facet labels). Differences in mapping proximity can be seen between nutrient conditions (red: nitrogen replete; blue: nitrogen deplete).

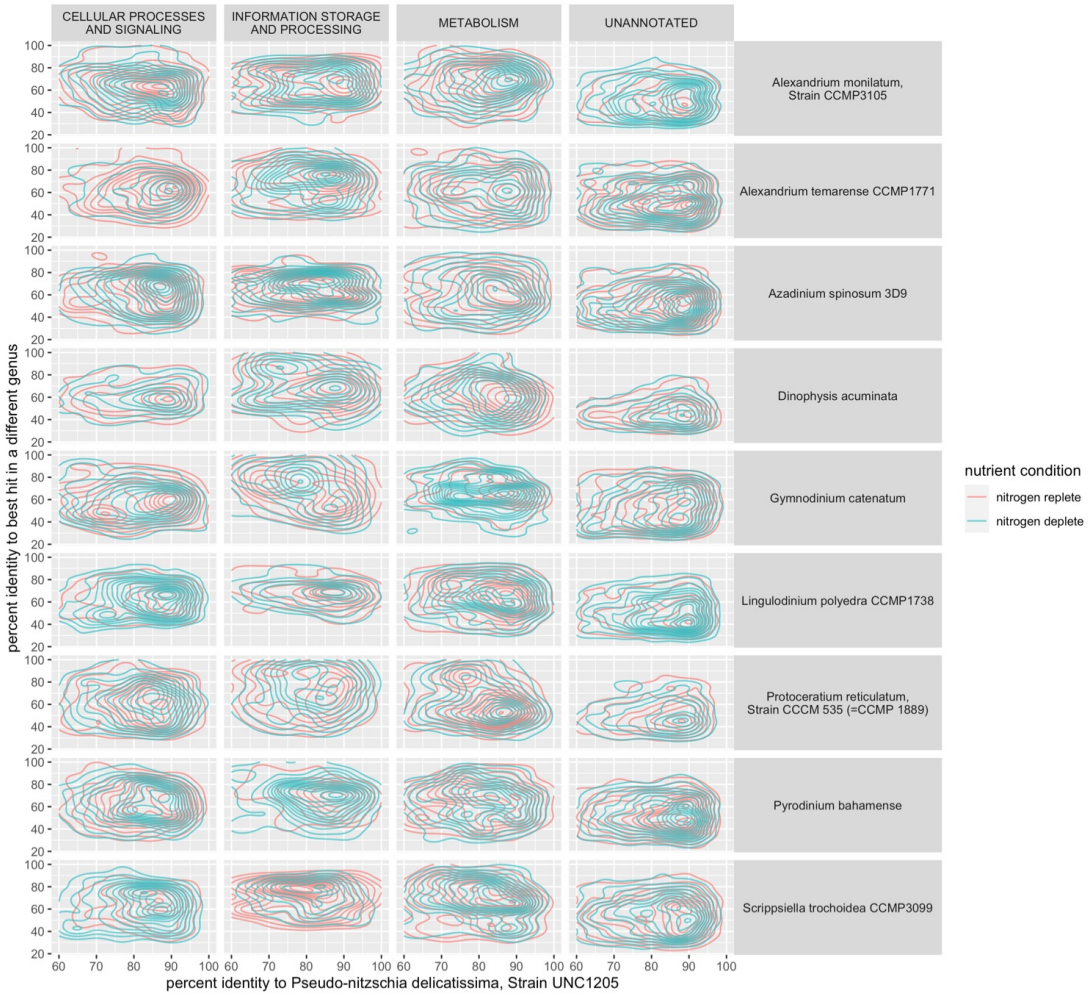


Figure S20 2D-histograms showing density of reads mapping to the most abundant dinoflagellate reference species, *Prorocentrum minimum* (x-axis) as well as closeness to the nearest species in a different genus (y-axis; right-hand facet-labels). Reads pertaining to different functions are separated out using Kegg Ortholog Groups (KOGs; top facet labels). Differences in mapping proximity can be seen between nutrient conditions (red: nitrogen replete; blue: nitrogen deplete).

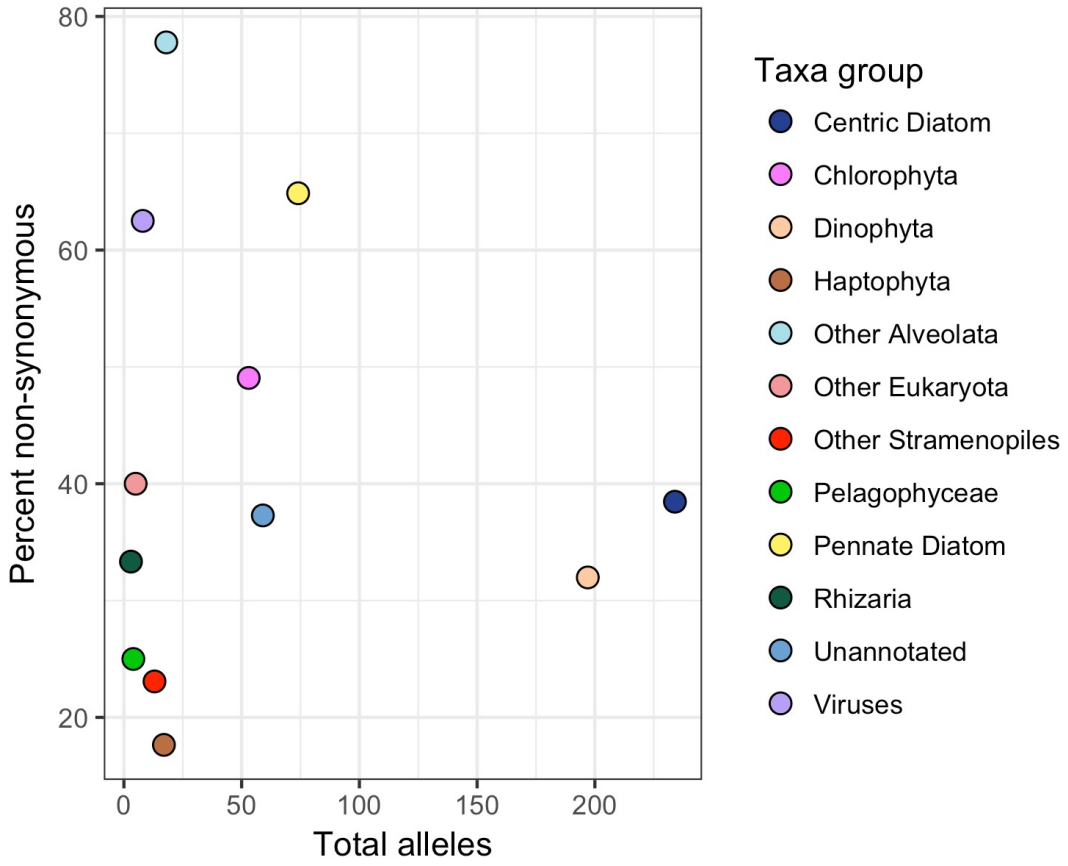


Figure S21 SNP abundance (x) and percent of SNPs non-synonymous (y) for ab initio ORF LPI taxa groups.

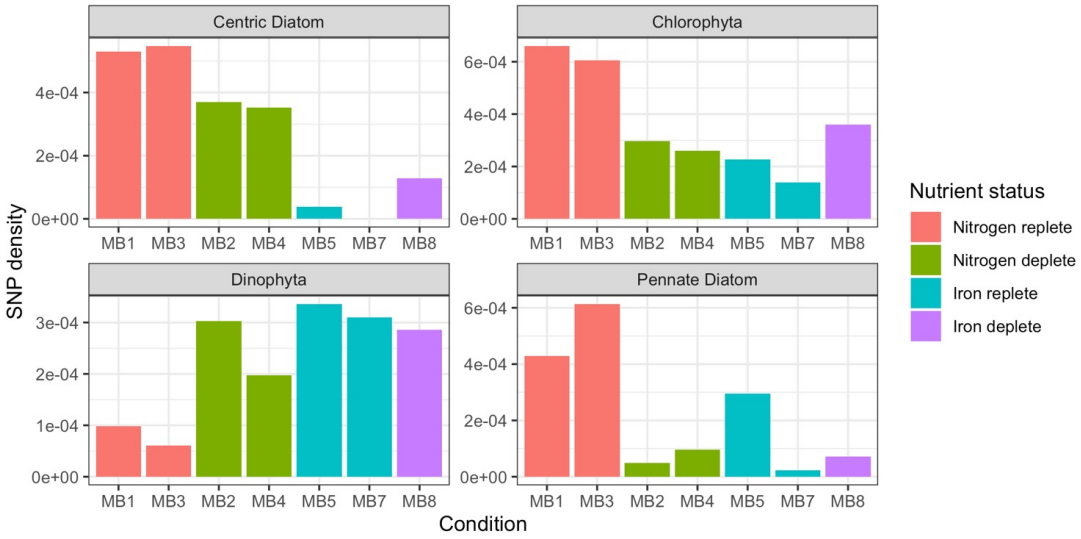


Figure S22 SNP density (total SNPs/ total ORFs) for major phytoplankton taxa across incubation conditions, colored by nutrient status.

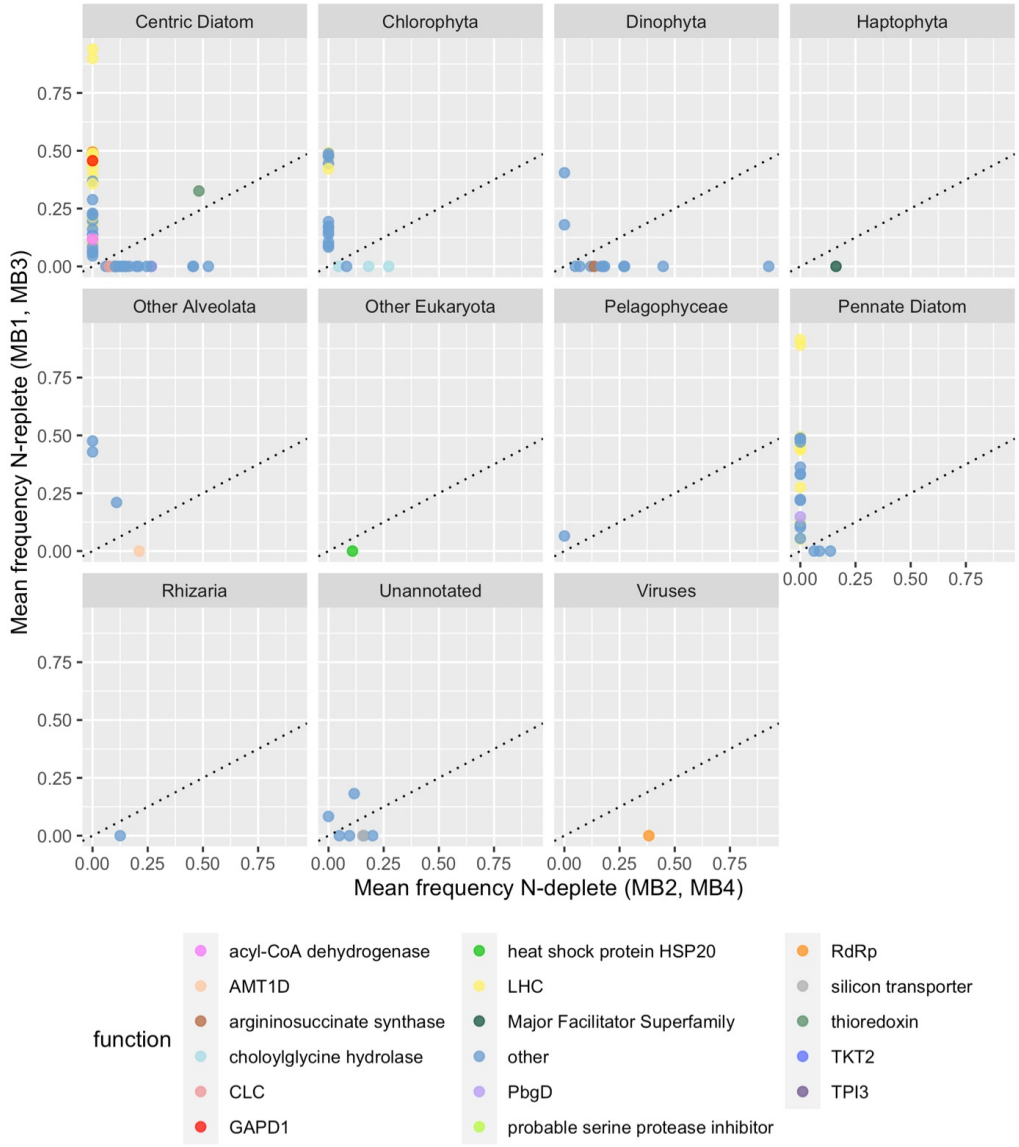


Figure S23 Mean frequency of non-synonymous alleles in N-replete (y) and N-deplete (x) conditions colored by functional annotation. Allele frequency is calculated as number of reads of a given allele / number of total reads across all alleles at that position (DP). An allele frequency of 1 would indicate that the entire pool of alleles is made up of a given allele, and zero would indicated that the given allele was not measured. The dashed line is a 1:1 line. Alleles above the line are more frequent in N-replete conditions; alleles below the line are more frequent in N-deplete conditions. Functional annotation abbreviations: AMT1d = ammonium transporter; CLC= clathrin light chain; GAPD1 = glyceraldehyde-3-phosphate dehydrogenase; LHC = light harvesting complex; PbgD = porphobilinogen deaminase; TKT2 = transketolase; TPI3 = triose-phosphate isomerase.

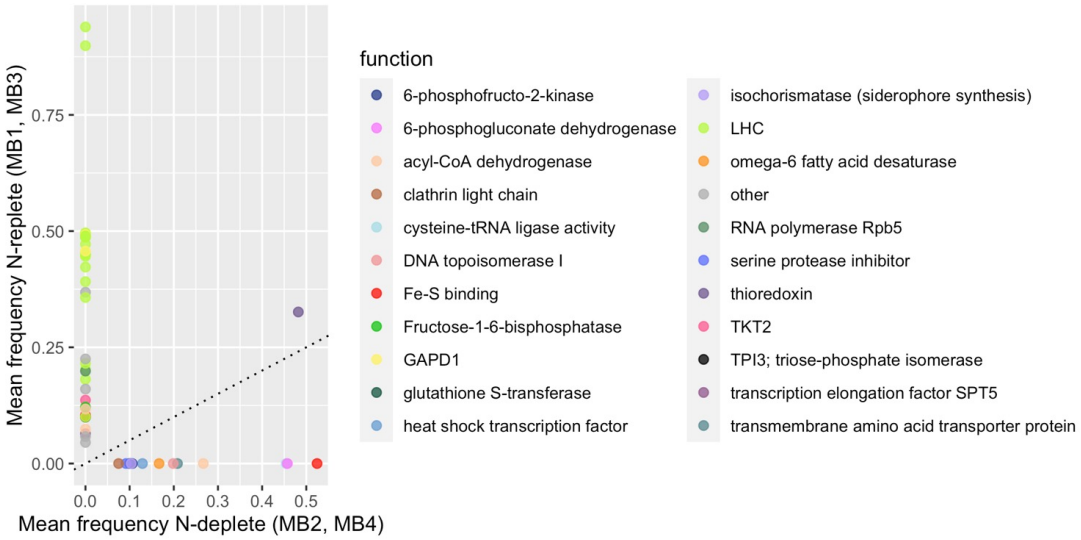


Figure S24 Mean frequency of annotated centric diatom non-synonymous alleles in N-replete (y) and N-deplete (x) conditions colored by functional annotation.

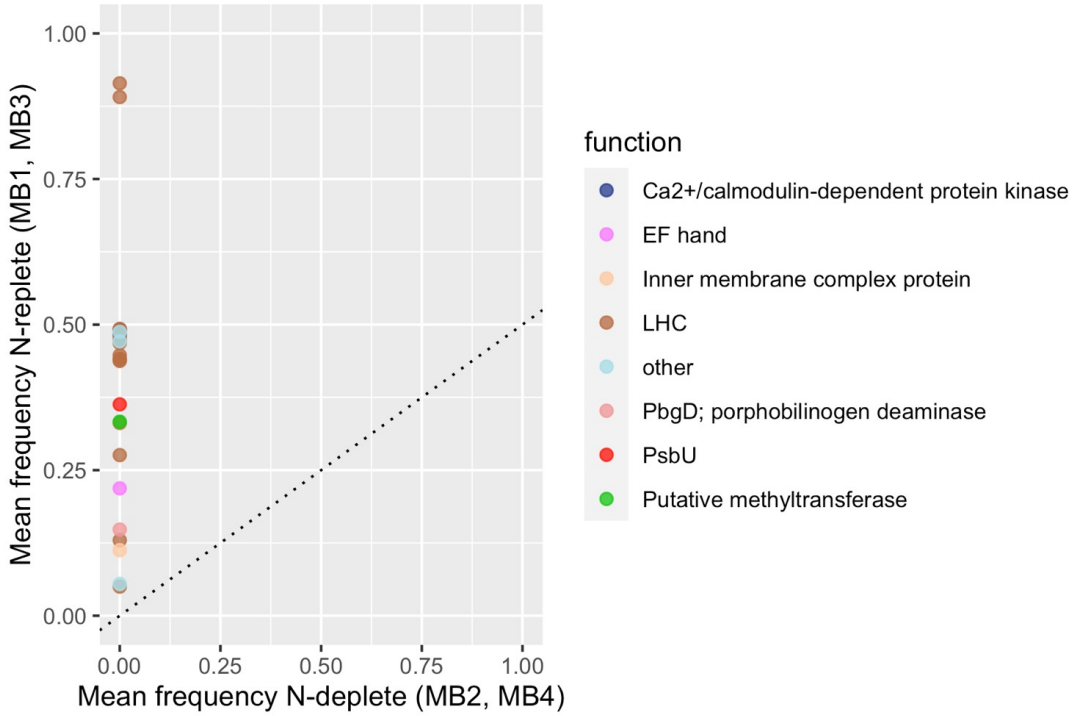


Figure S25 Mean frequency of annotated pennate diatom non-synonymous alleles in N-replete (y) and N-deplete (x) conditions colored by functional annotation

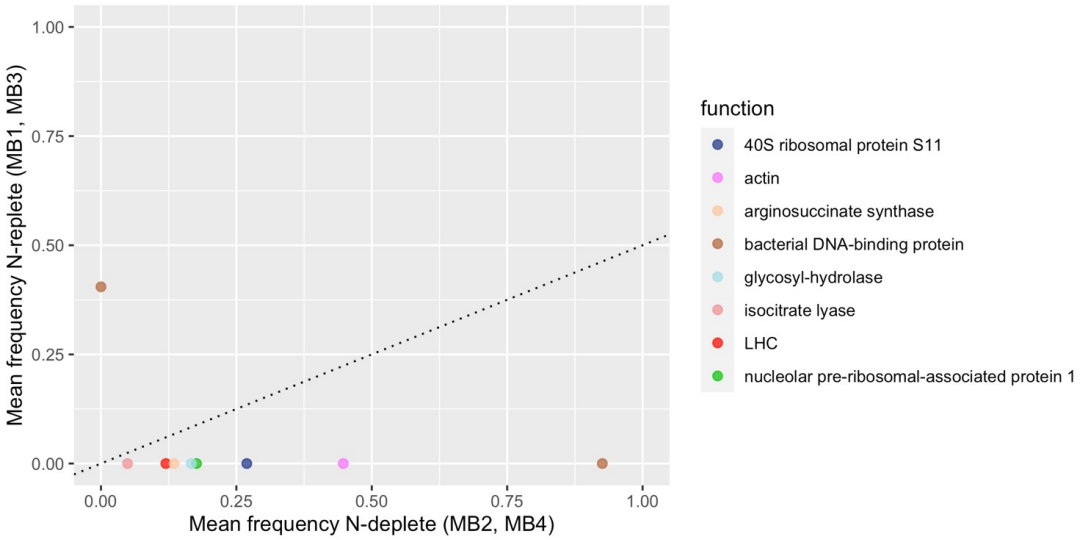


Figure S26 Mean frequency of annotated dinoflagellate non-synonymous alleles in N-replete (y) and N-deplete (x) conditions colored by functional annotation.

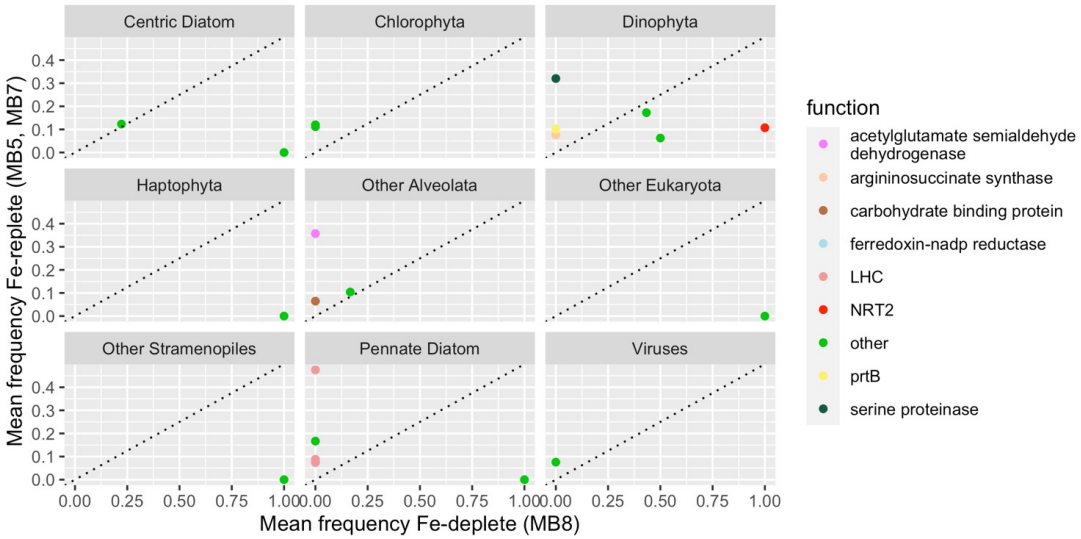


Figure S27 Mean frequency of non-synonymous alleles in Fe-replete (y) and Fe-deplete (x) conditions colored by functional annotation. Allele frequency is calculated as number of reads of a given allele / number of total reads across all alleles at that position (DP). An allele frequency of 1 would indicate that the entire pool of alleles is made up of a given allele, and zero would indicated that the given allele was not measured. The dashed line is a 1:1 line. Alleles above the line are more frequent in N- replete conditions; alleles below the line are more frequent in N-deplete conditions. Functional annotation abbreviations: LHC = light harvesting complex; NRT2 = nitrate transporter; prtB = proteinase.

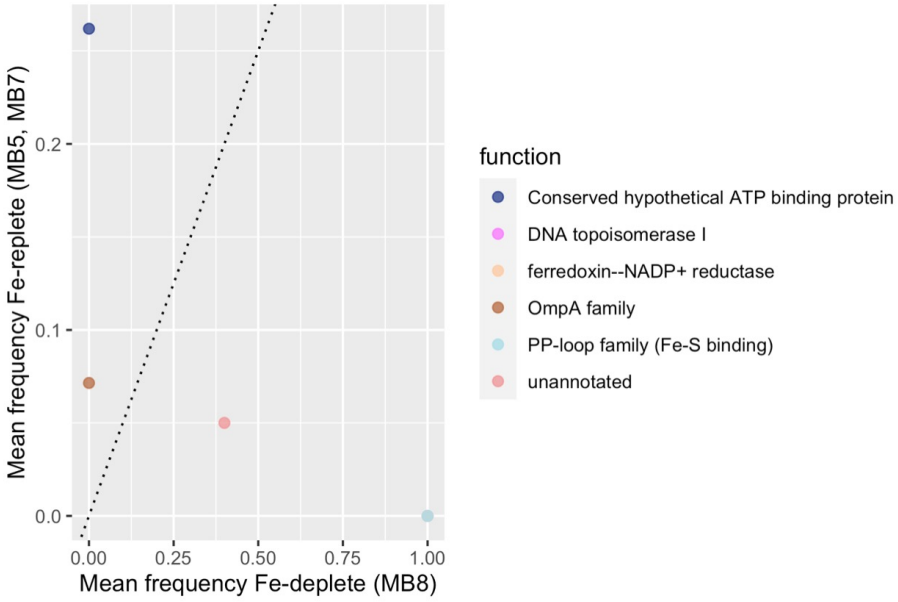


Figure S28 Mean frequency of centric diatom non-synonymous alleles in Fe-replete (y) and Fe-deplete (x) conditions colored by functional annotation.

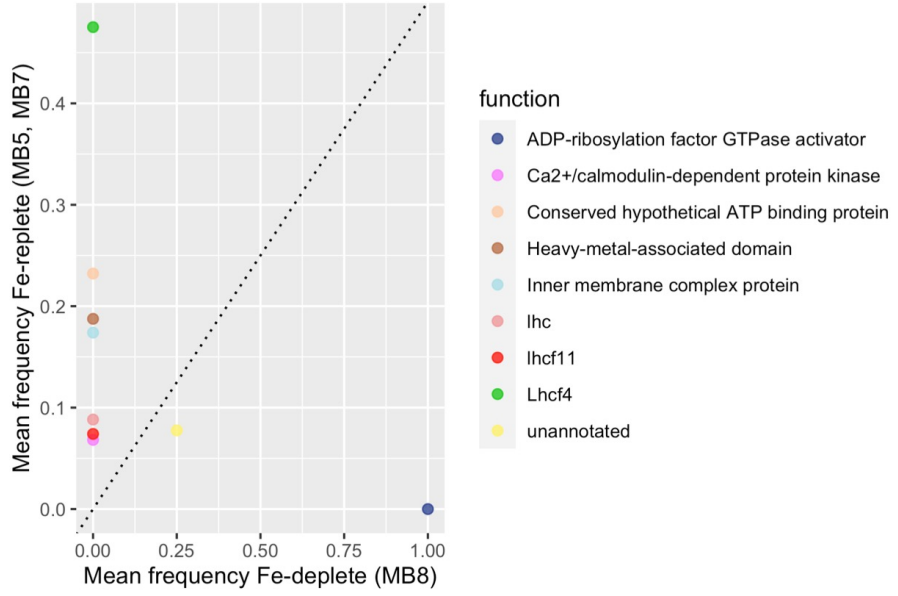


Figure S29 Mean frequency of pennate diatom non-synonymous alleles in Fe-replete (y) and Fe-deplete (x) conditions colored by functional annotation.

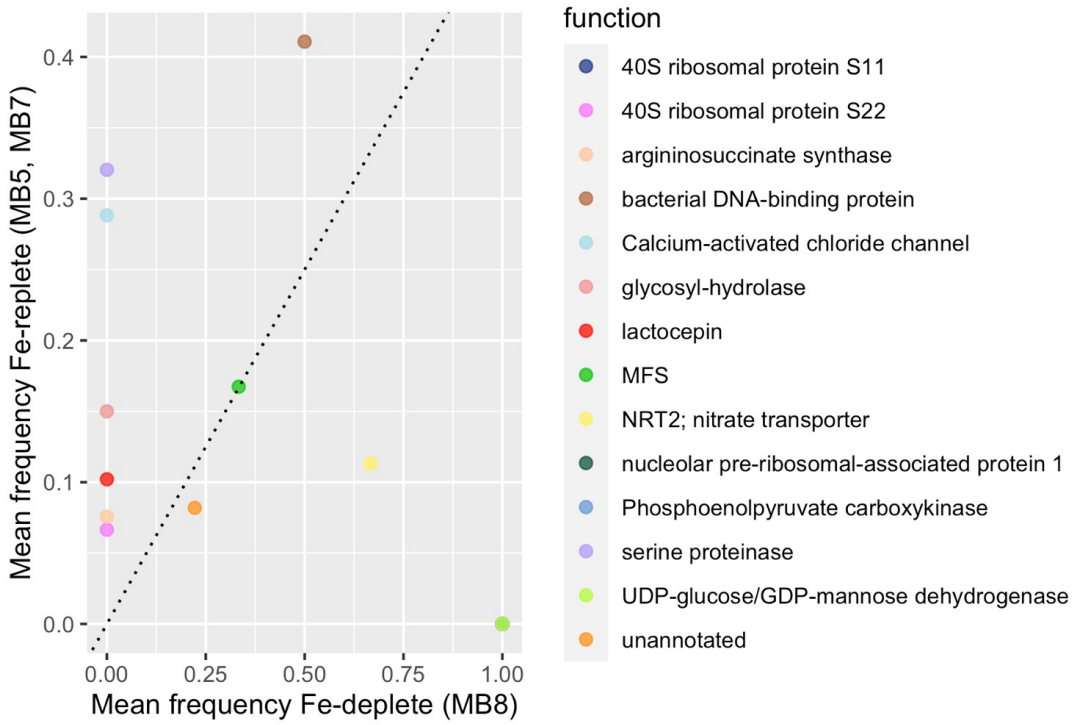


Figure S30 Mean frequency of dinoflagellate non-synonymous alleles in Fe-replete (y) and Fe-deplete (x) conditions colored by functional annotation.

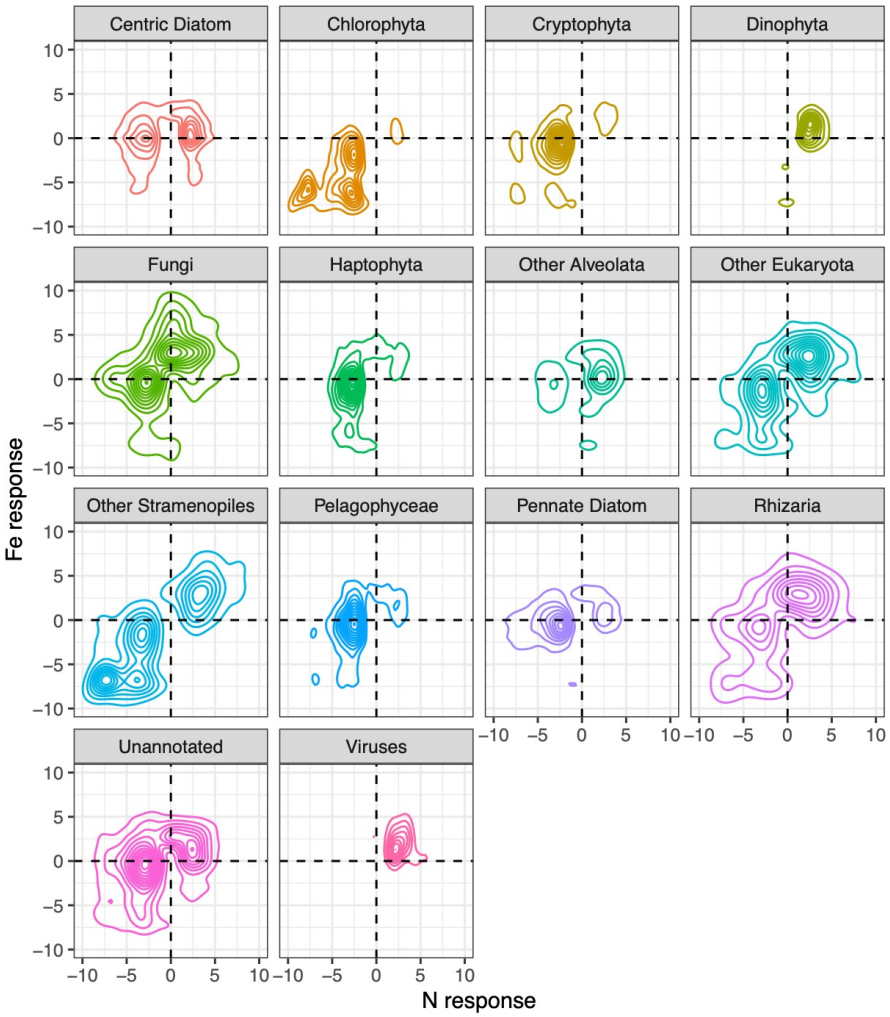


Figure S31 2D-histograms showing varied transcriptional responses of major lineages to N and Fe status (y-axis: log2FC(Fe-deplete/Fe-replete); >0 is up in low iron; <0 is down in low iron) and nitrogen (x- axis: log2FC(N-deplete/N-replete); >0 is up in low nitrogen; <0 is down in low nitrogen). Only ORFs significantly differentially expressed (FDR < 0.05) in at least one condition (N, Fe, or both) are included.

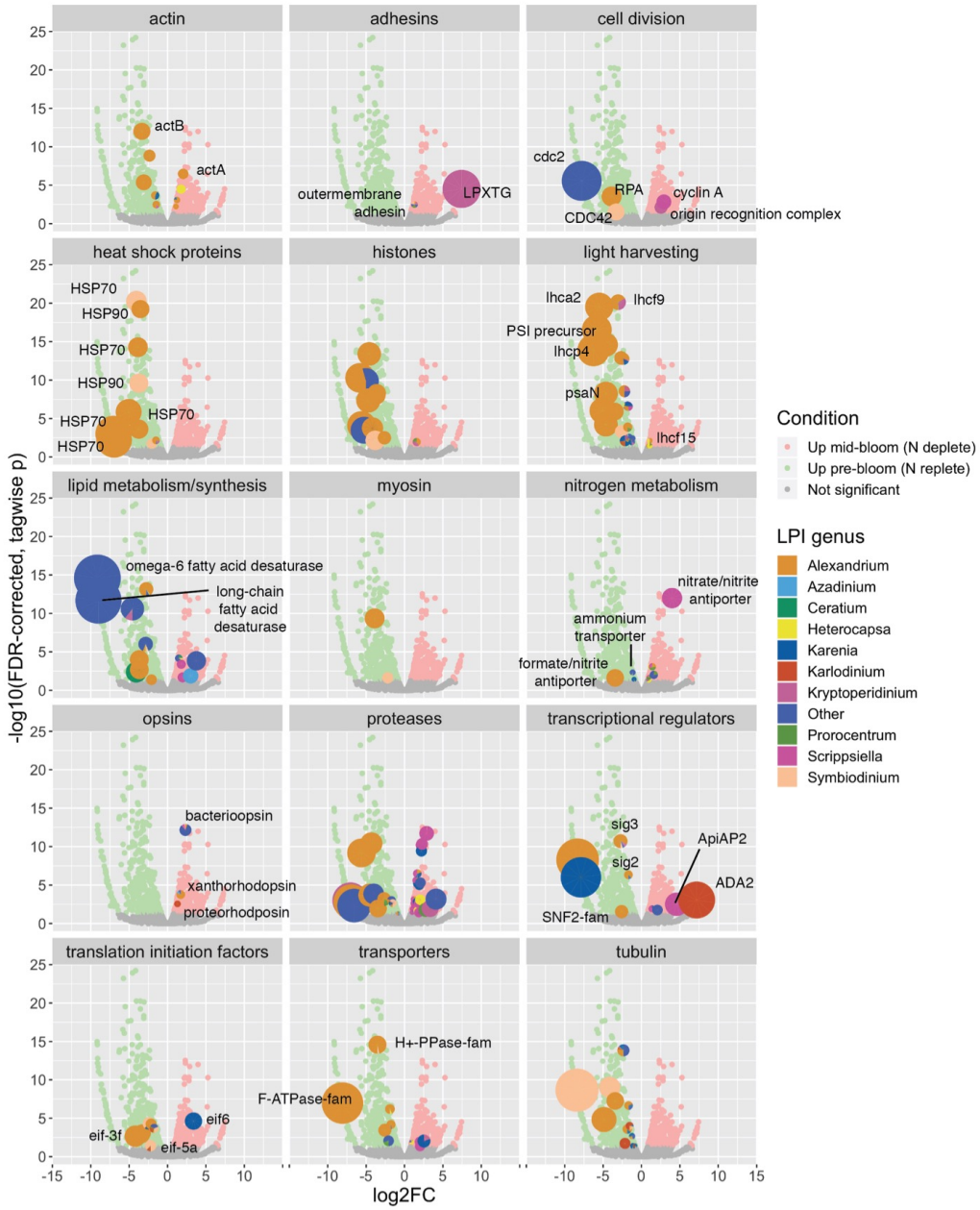


Figure S32 Differential expression of dinoflagellate KO functions across nutrient conditions of experiment 1.

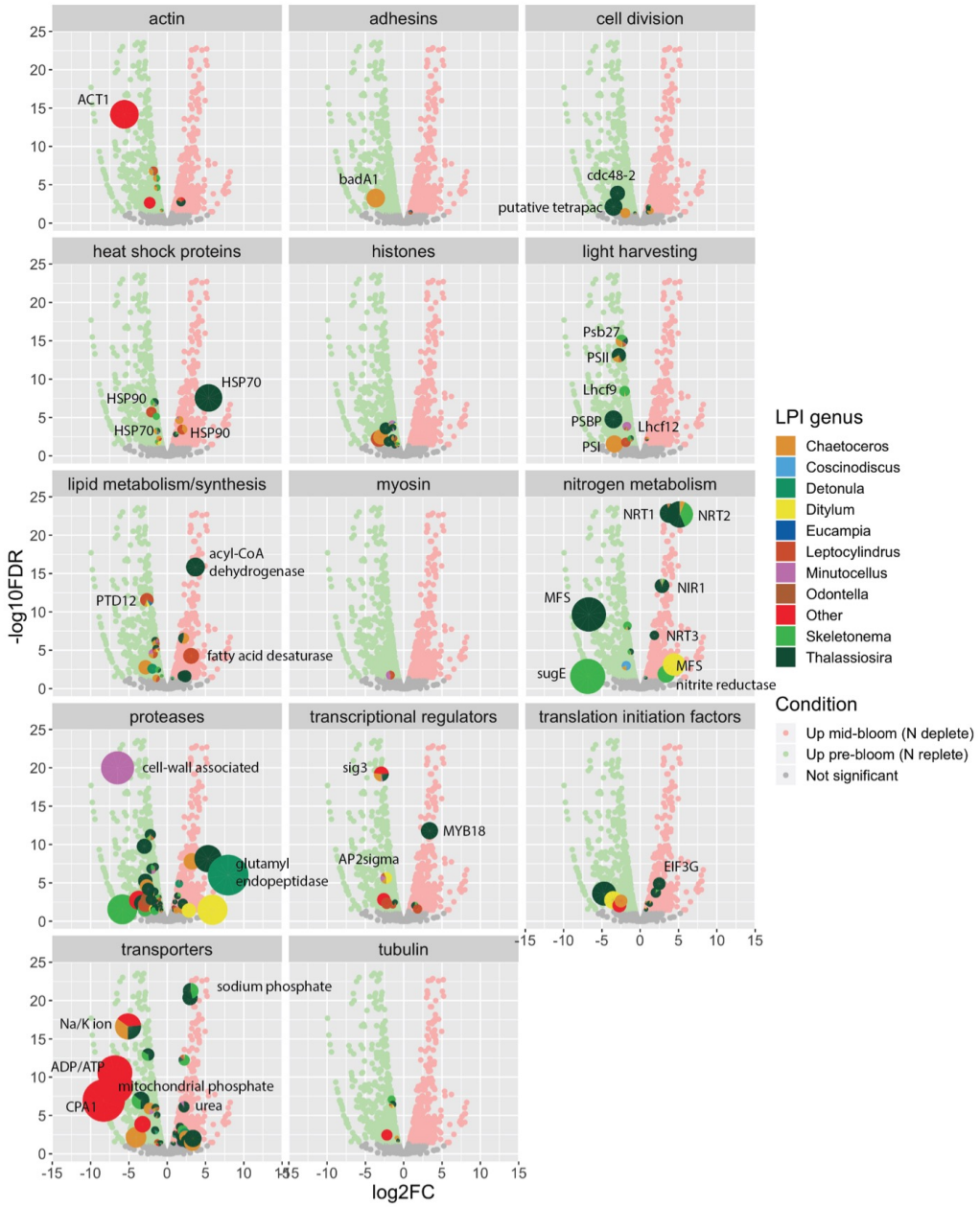


Figure S33 Differential expression of centric diatom KO functions across nutrient conditions of experiment 1.

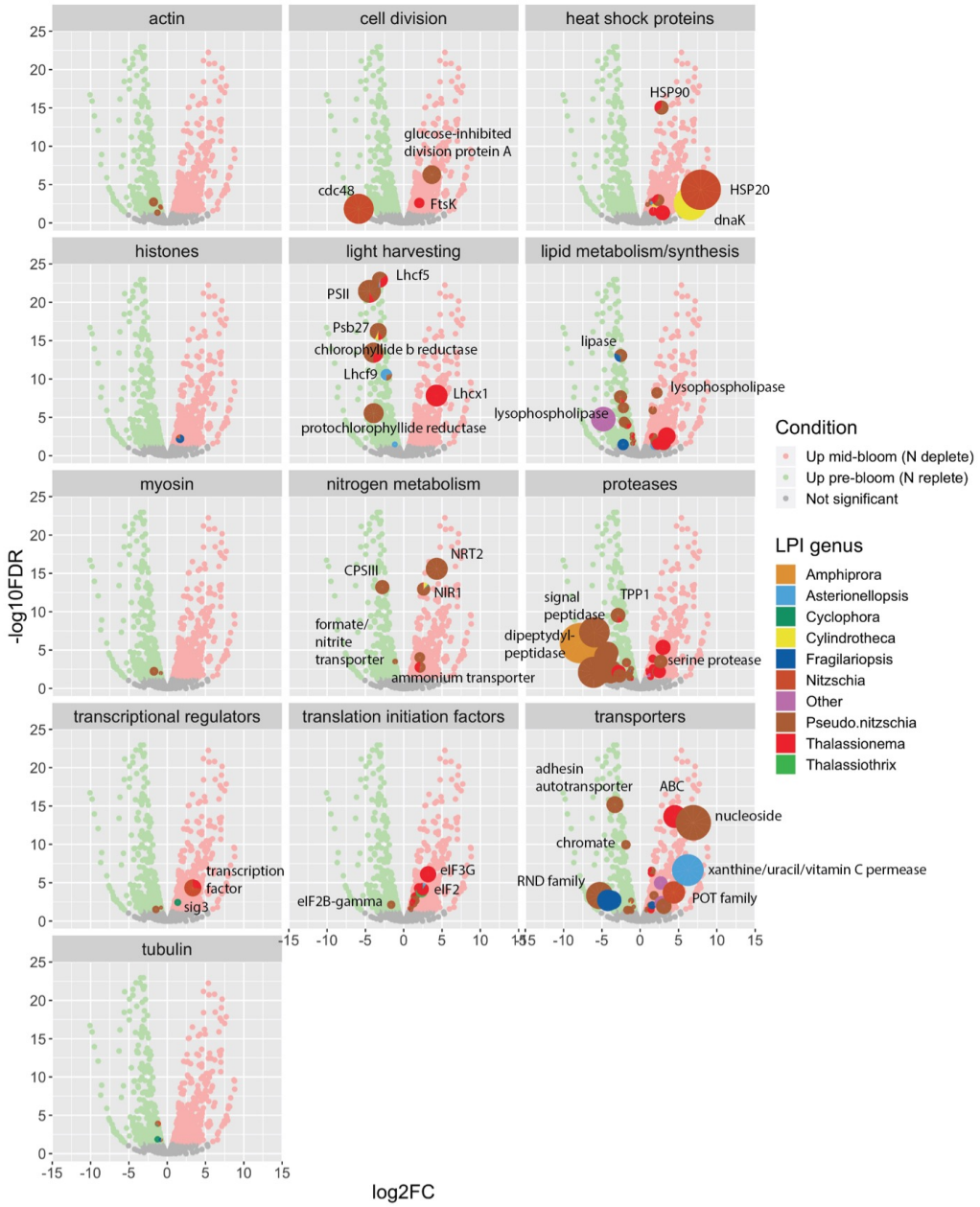


Figure S34 Differential expression of pennate diatom KO functions across nutrient conditions of experiment 1.

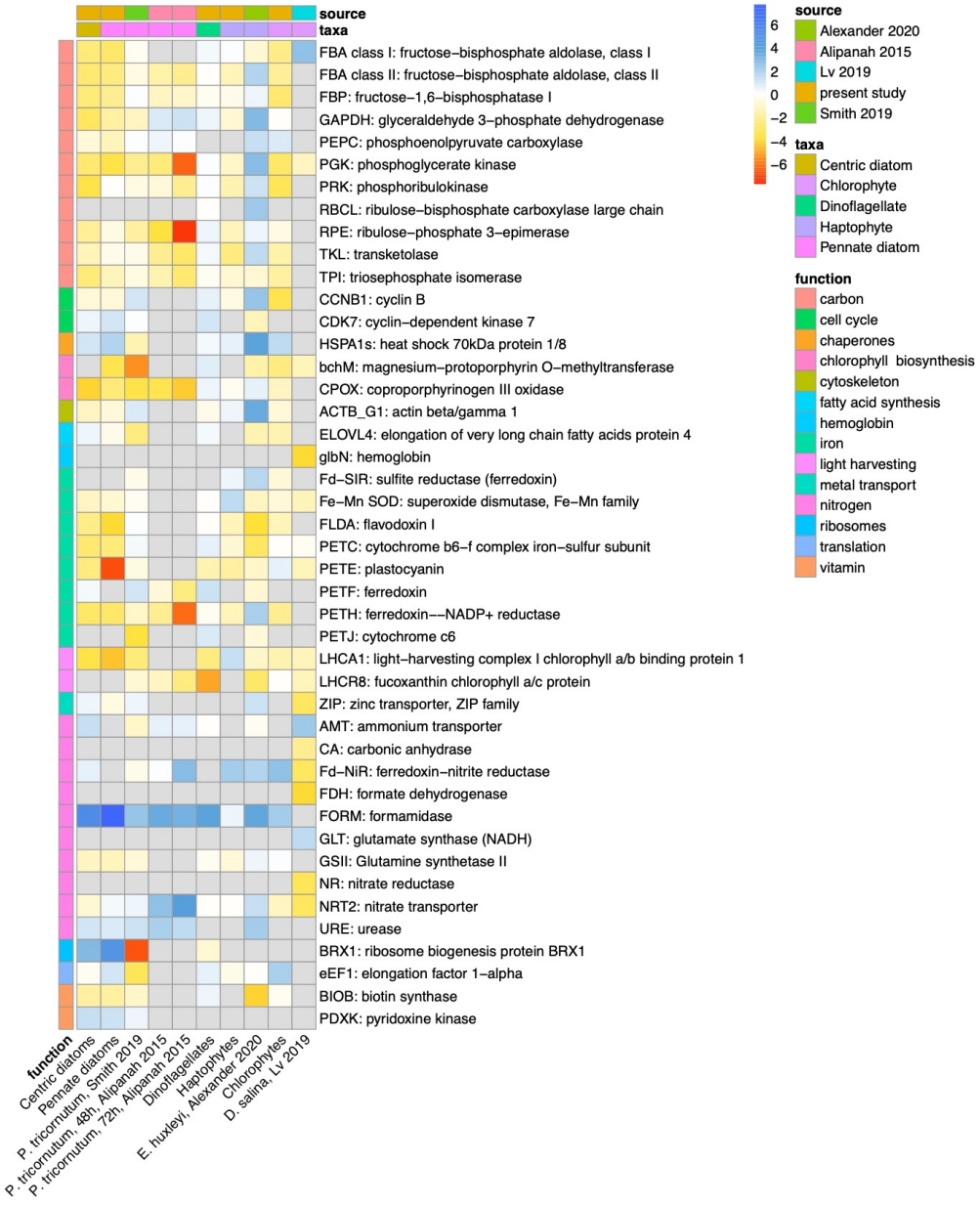


Figure S35 Comparison of the response of select genes to nitrogen availability in taxa from the present study versus available published datasets. Heatmap is colored by log2FC of N-deplete versus N-replete conditions (positive= up in deplete (blue), negative = up in replete (red)). Genes were identified using KO IDs when possible, gene IDs specific to Alipanah 2015 data[1] (Phatr2), Smith 2019 data[2] (Phatr3), and Lv 2019 data[3] (D. Salina gene IDs) were occasionally used. For Alipanah 2015, the “48 h” and “72 h” conditions represents 48 h and 72 h of nitrogen deprivation, respectively. Log2FC for the present study were calculated in edgeR for each taxa group independently using raw

taxa group independently using raw reads summed over KO annotations. For Alexander 2020 data[4], log2FC between N-deplete and N-replete conditions was back-calculated from log2FC given in supplemental datasets by subtracting log2FC(N-deplete/control) - log2FC(N-replete/control).For Smith 2019 data[2], edgeR was performed on raw reads with 15 minute and 45 minute NH4, NO2- and NO3- timepoints considered N-replete replicates and 18 hour NH4, NO2- and NO3- timepoints considered N-deplete replicates.

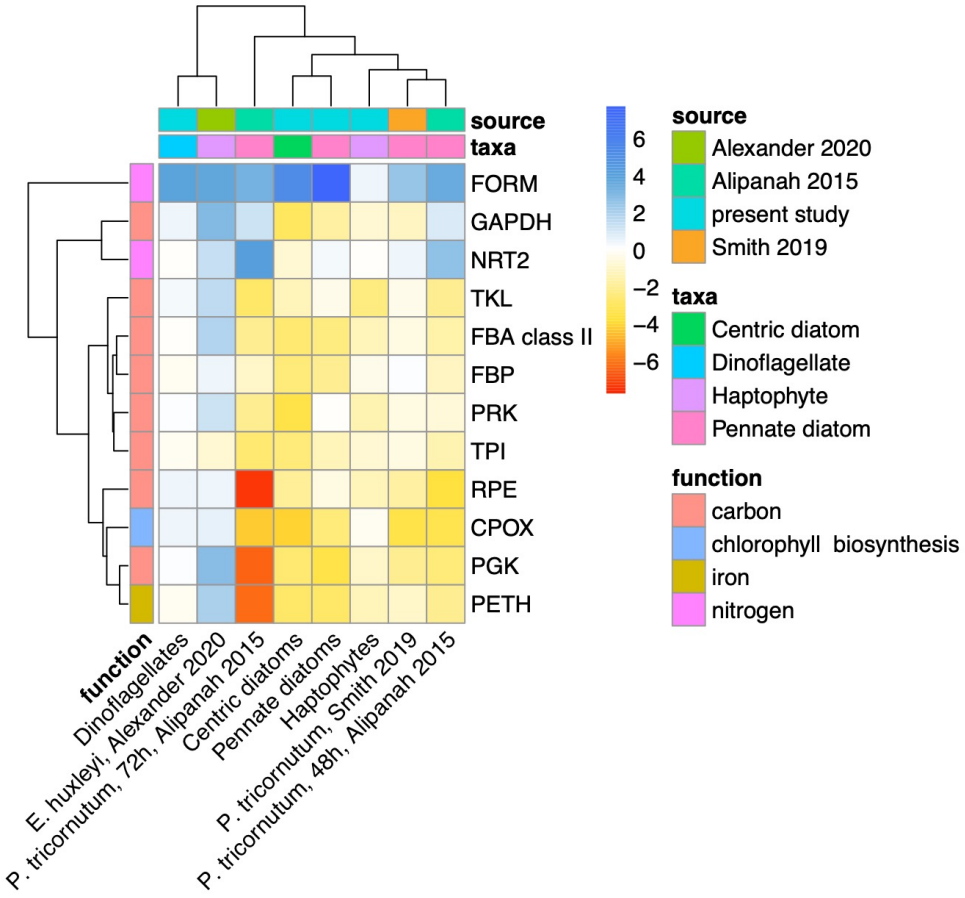


Figure S36 Comparison of the response of select genes to nitrogen availability in diatoms, haptophytes, and dinoflagellates from the present study versus available published datasets, as in Figure S23, where full gene names are given. Heatmap is colored by log2FC of N-deplete versus N-replete conditions (positive= up in deplete (blue), negative = up in replete (red)).

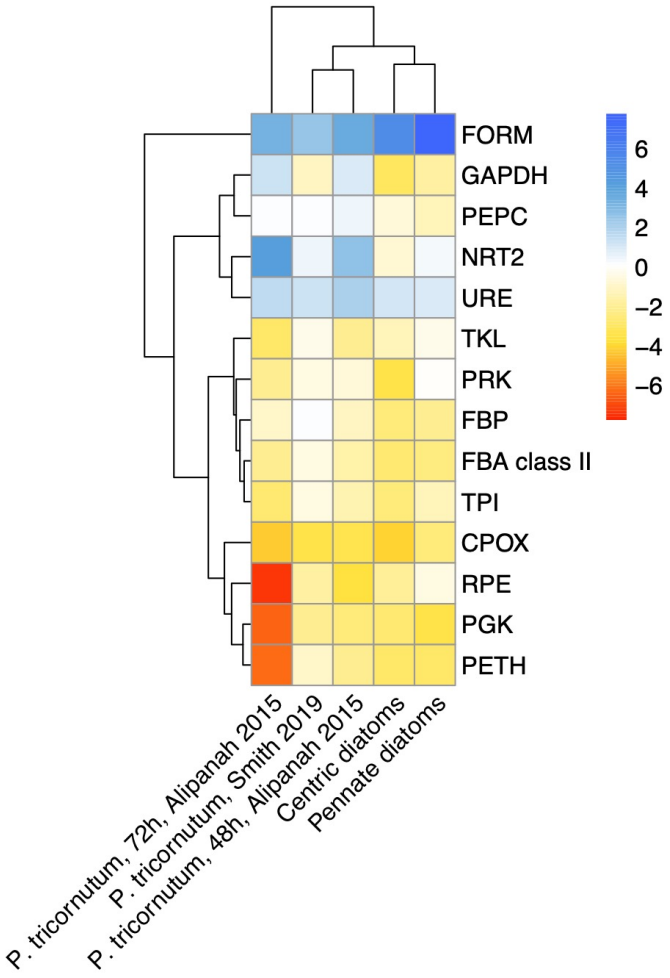


Figure S37 Comparison of the response of select genes to nitrogen availability in diatoms from the present study versus available published datasets, as in Figure S23, where full gene names are given. Heatmap is colored by log2FC of N-deplete versus N-replete conditions (positive= up in deplete (blue), negative = up in replete (red)).

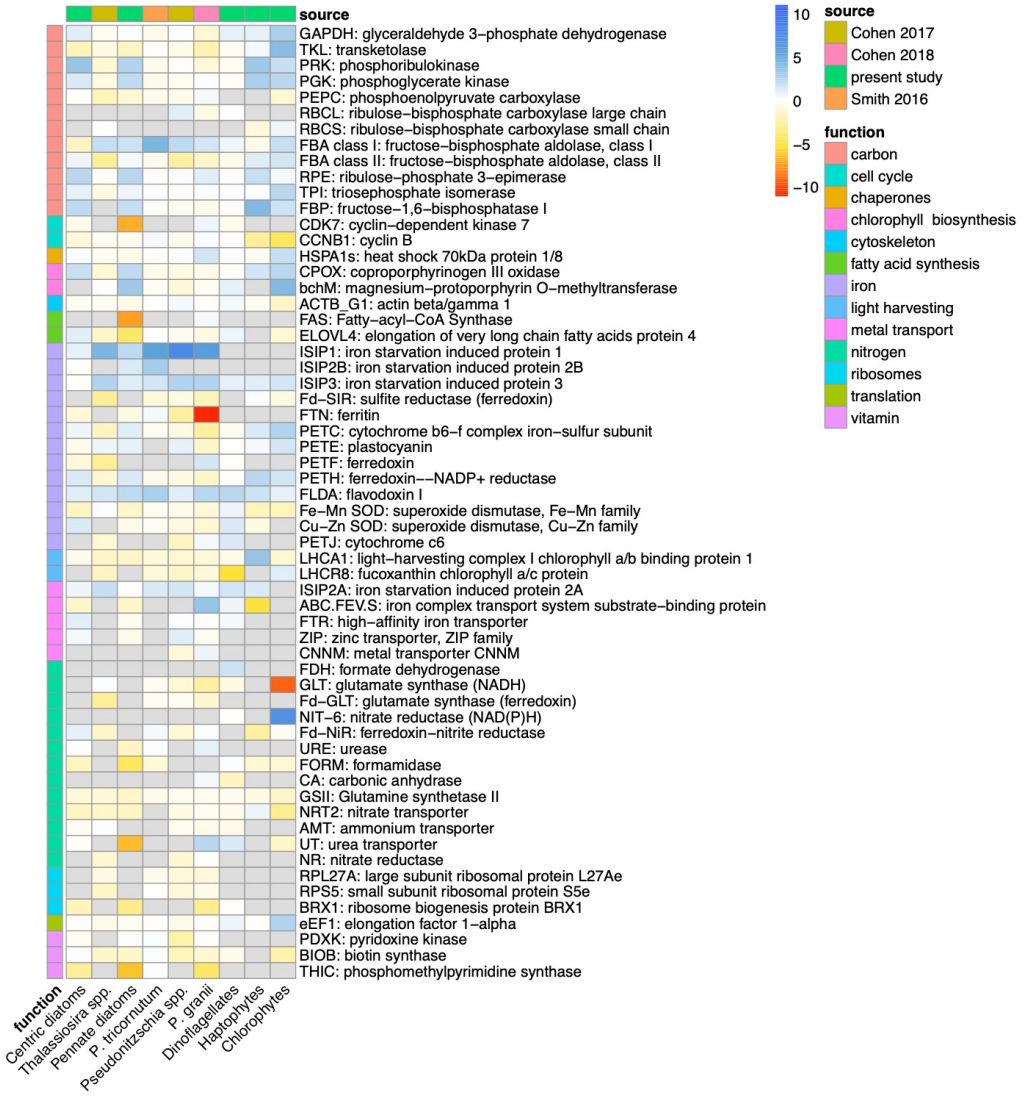


Figure S38 Comparison of the response of select genes to iron availability in taxa from the present study versus available published datasets. Heatmap is colored by log2FC of iron-deplete versus iron-replete conditions (positive= up in deplete (blue), negative = up in replete (red)). Genes were identified using KO IDs, which are catalogued in SDX. Log2FC for the present study were calculated in edgeR for each taxa group independently using raw reads summed over KO annotations. For Smith 2016[5], log2FC was calculated from the “Light versus Dark” RPKM ratio given in the supplementary data, and subsequently averaged by KO ID. Log2FC from Cohen 2017[6] and Cohen 2018[7] supplemental data was used directly.

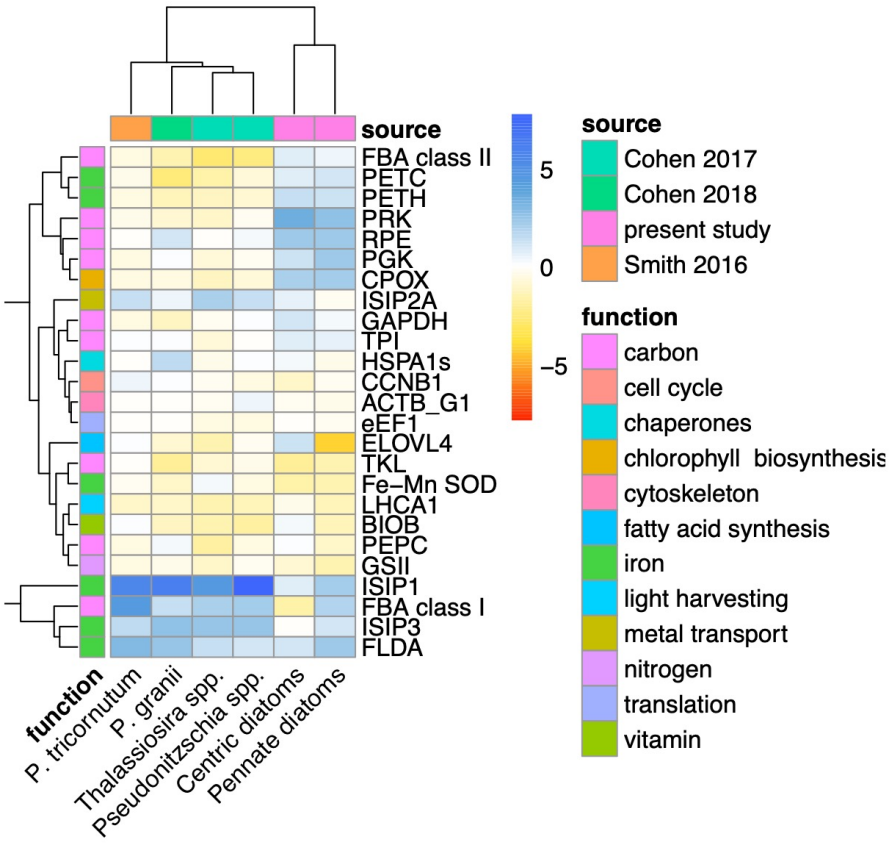


Figure S39 Comparison of the response of select genes to iron availability in diatoms, haptophytes, and dinoflagellates from the present study versus available published datasets, as in Figure S26, where full gene names are given. Heatmap is colored by log2FC of N-deplete versus N-replete conditions (positive= up in deplete (blue), negative = up in replete (red)). Heatmap is colored by log2FC of iron-deplete versus iron-replete conditions (positive= up in deplete (blue), negative = up in replete (red)).

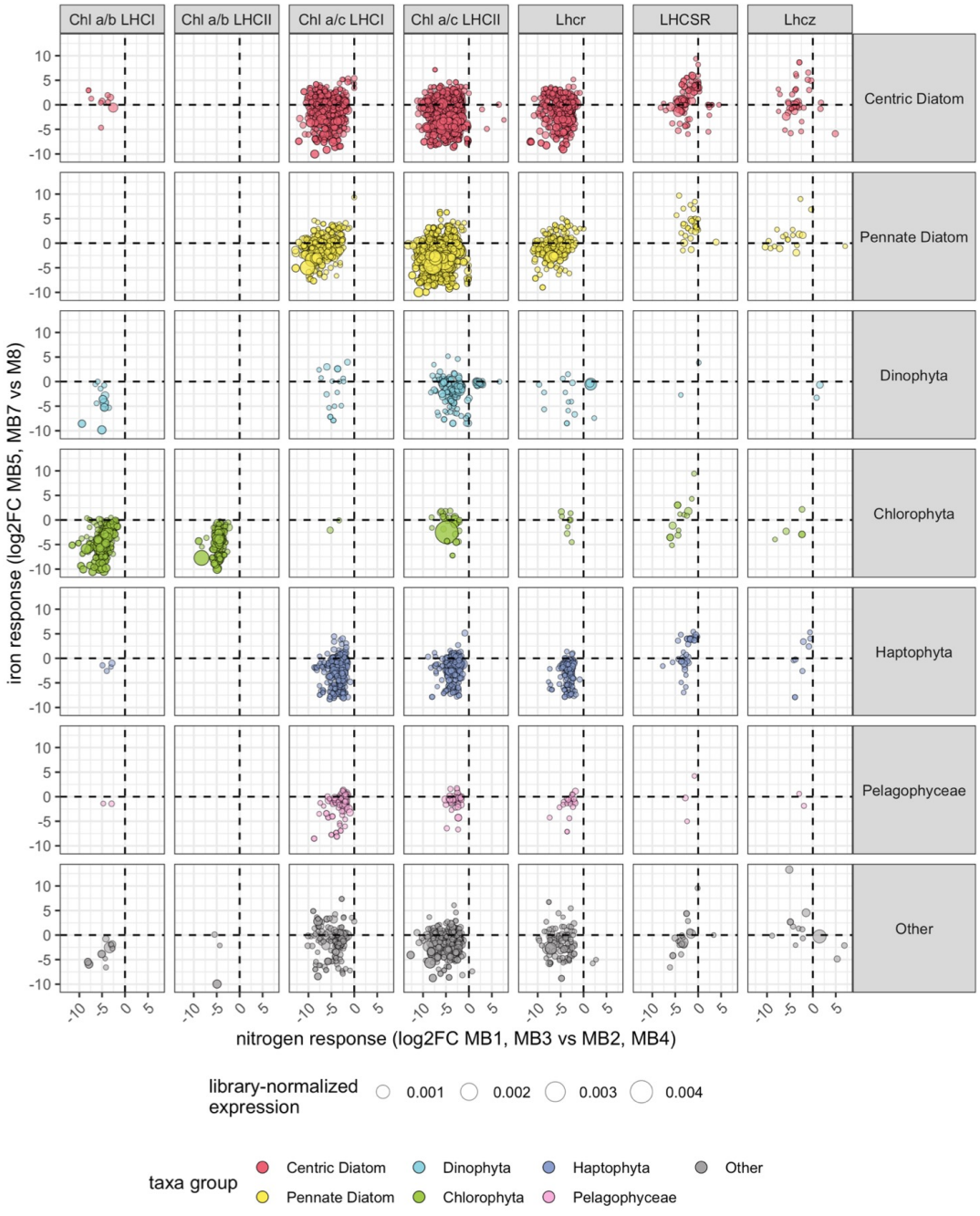


Figure S40 Response of LHC subfamilies to iron (y-axis; >0 is up in low iron; <0 is down in low iron) and nitrogen (x- axis; >0 is up in low nitrogen; <0 is down in low nitrogen). ORFs must be significantly differentially expressed (FDR < 0.05) in at least one condition to be included.

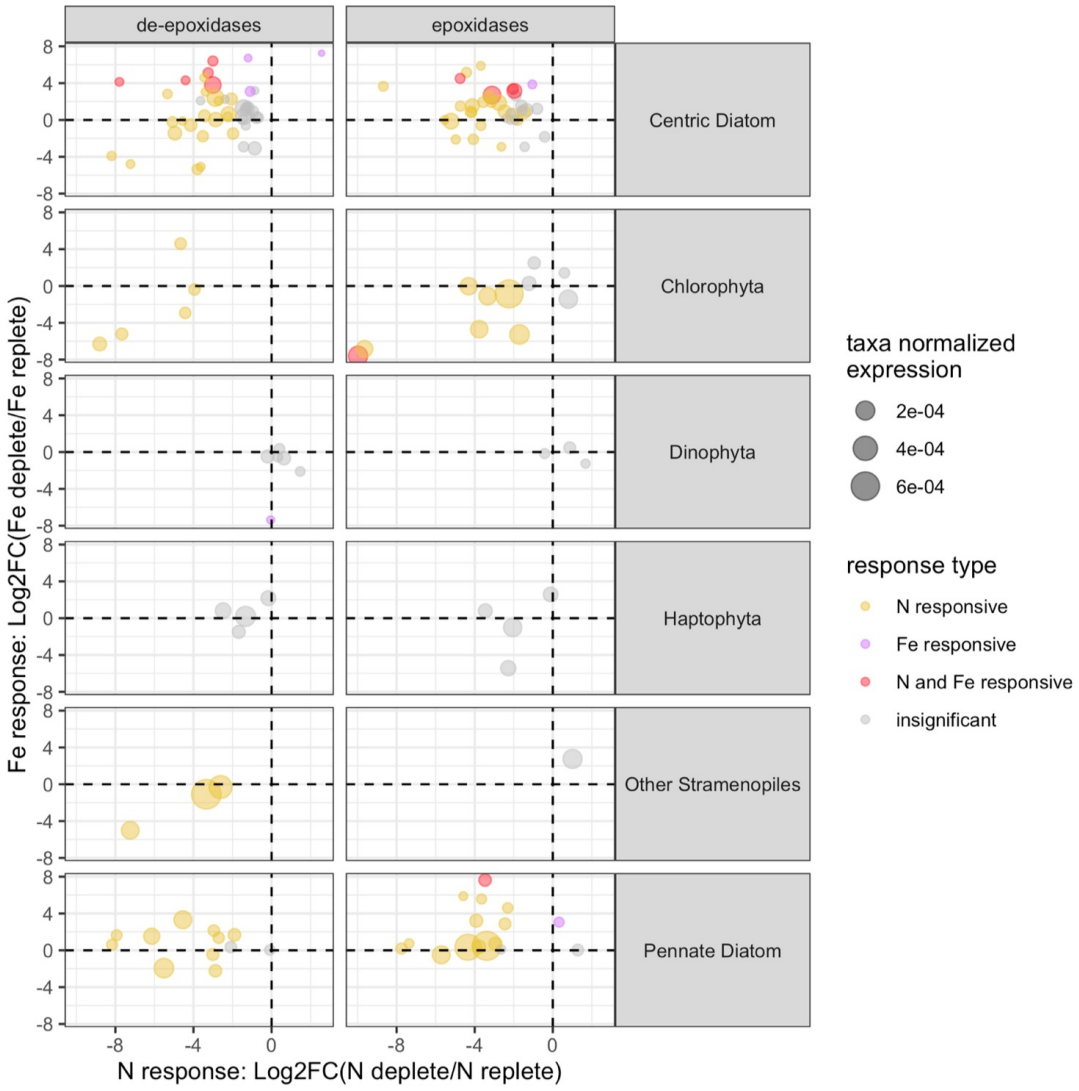


Figure S41 Response of xanthophyll cycle de-epoxidases and epoxidases to iron (y-axis; >0 is up in low iron; <0 is down in low iron) and nitrogen (x- axis; >0 is up in low nitrogen; <0 is down in low nitrogen). De-epoxidases were identified the violoxanthin de-epoxidase (VDE) PFAM, PF07137, and the annotation search term “de-epoxidase.” Epoxidases were identified using the zeaxanthin epoxidase KO, K09838. ORFs must be significantly differentially expressed (FDR < 0.05) in at least one condition to be included.

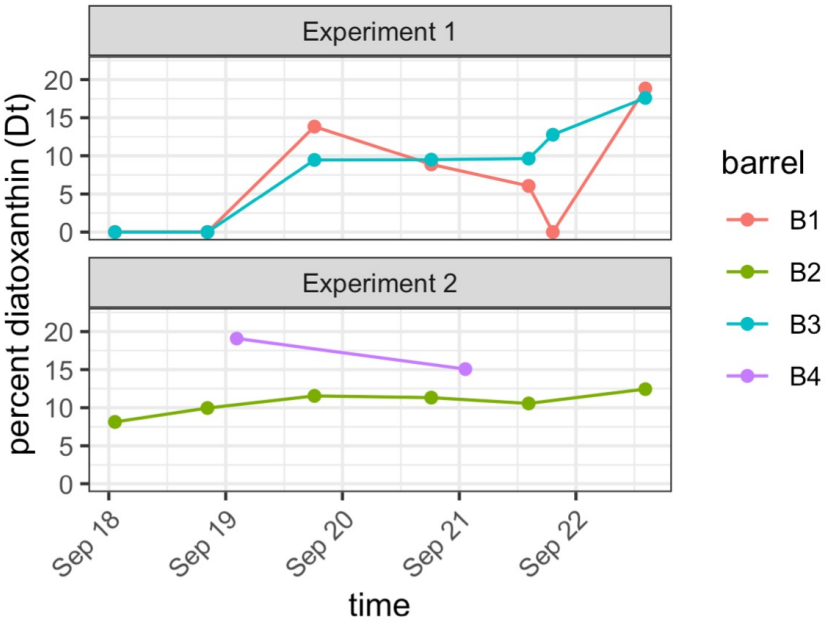


Figure S42 Percentage of diadinoxanthin cycle pigments (diadinoxanthin (Dd) and diatoxanthin (Dt)) made up of Dt over time in experiments 1 and 2.

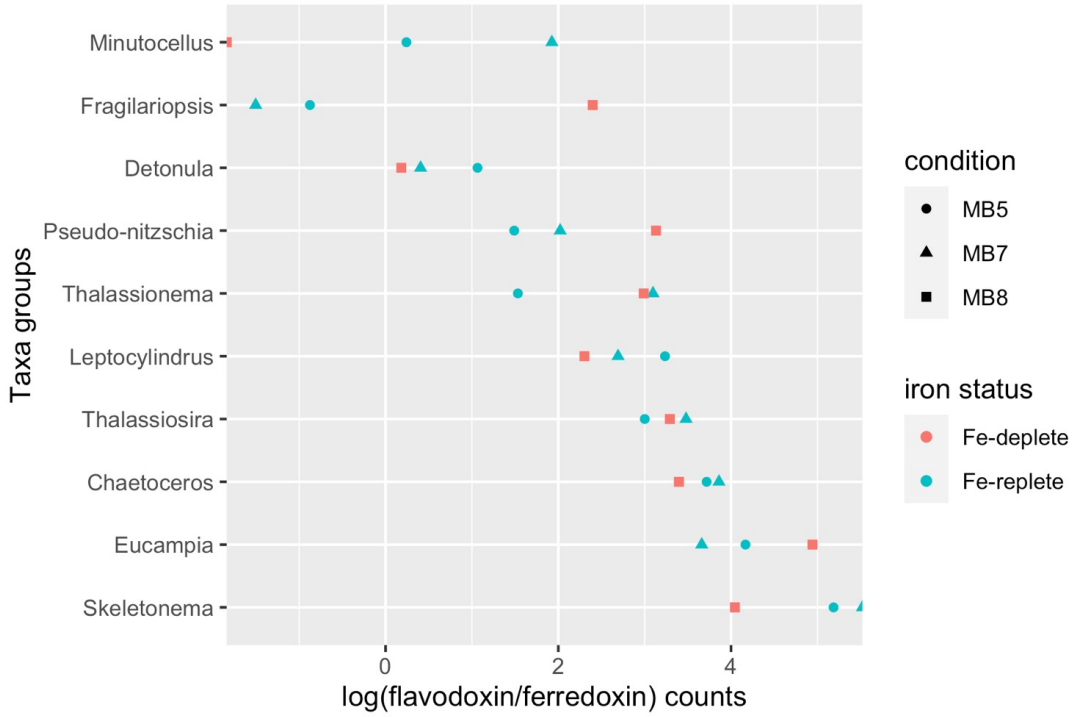


Figure S43 Log ratio of flavodoxin (PF00258) expression to ferredoxin (PF00111) expression averaged diatom genera as a marker of iron status.

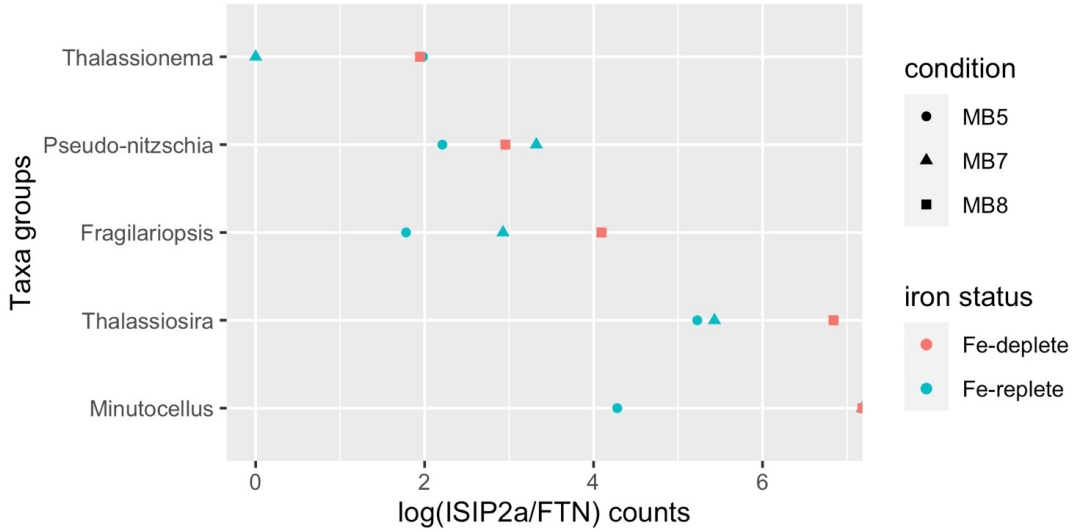


Figure S44 Log ratio of diatom iron starvation induced protein ISIP2a (uniprot B7FYL2) expression to ferritin (FTN, PFAM PF00210) expression averaged across diatom genera as a marker of iron status.

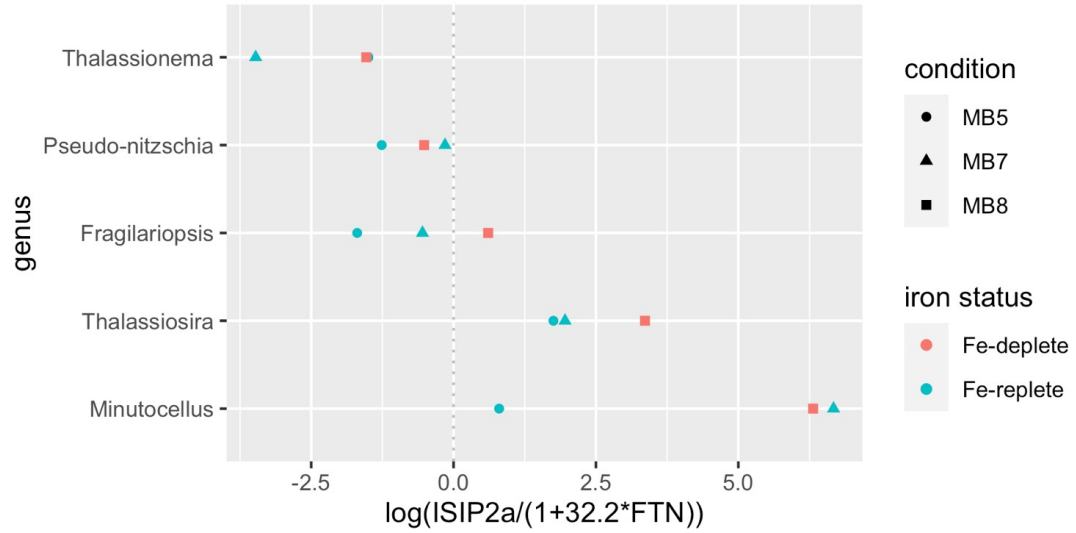


Figure S45 *Pseudo-nitzschia* iron limitation index (ILI) proposed by Marchetti et al 2017 averaged across diatom genera as a marker of iron status[8].

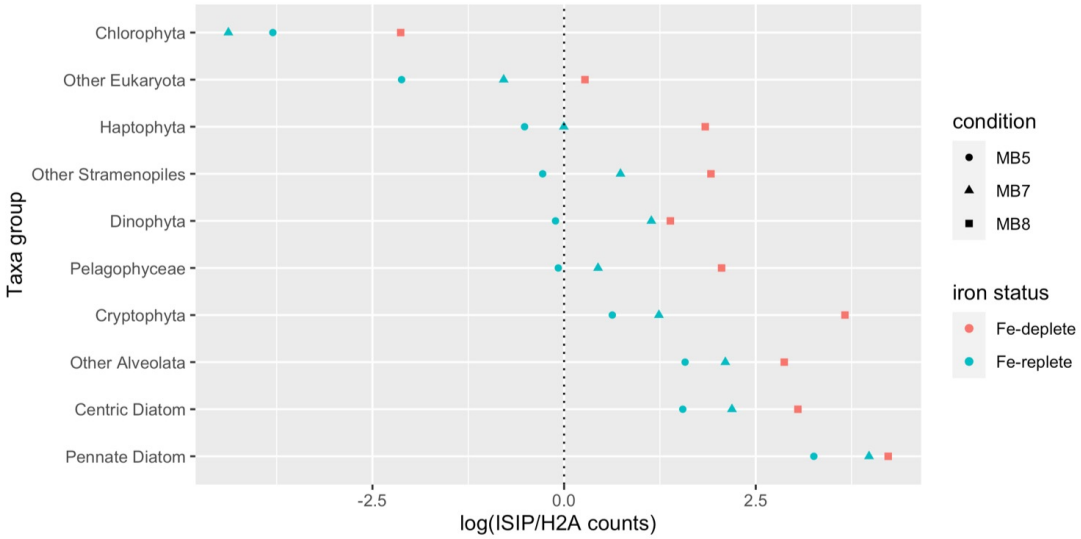


Figure S46 Log ratio of diatom iron starvation induced protein (ISIP1, uniprot B7GA90; ISIP2 uniprot B7FYL2; ISIP3, uniprot B7G4H8) expression to histone 2A (H2A; KOG1757) expression averaged across diverse phytoplankton lineages as a marker of iron status.

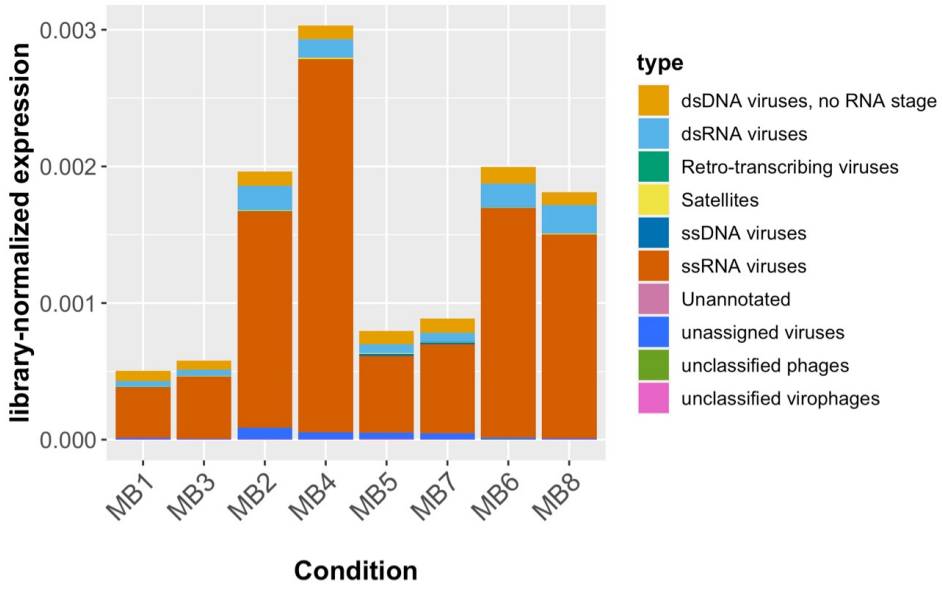


Figure S47 Taxonomic distributional of viral reads across conditions.

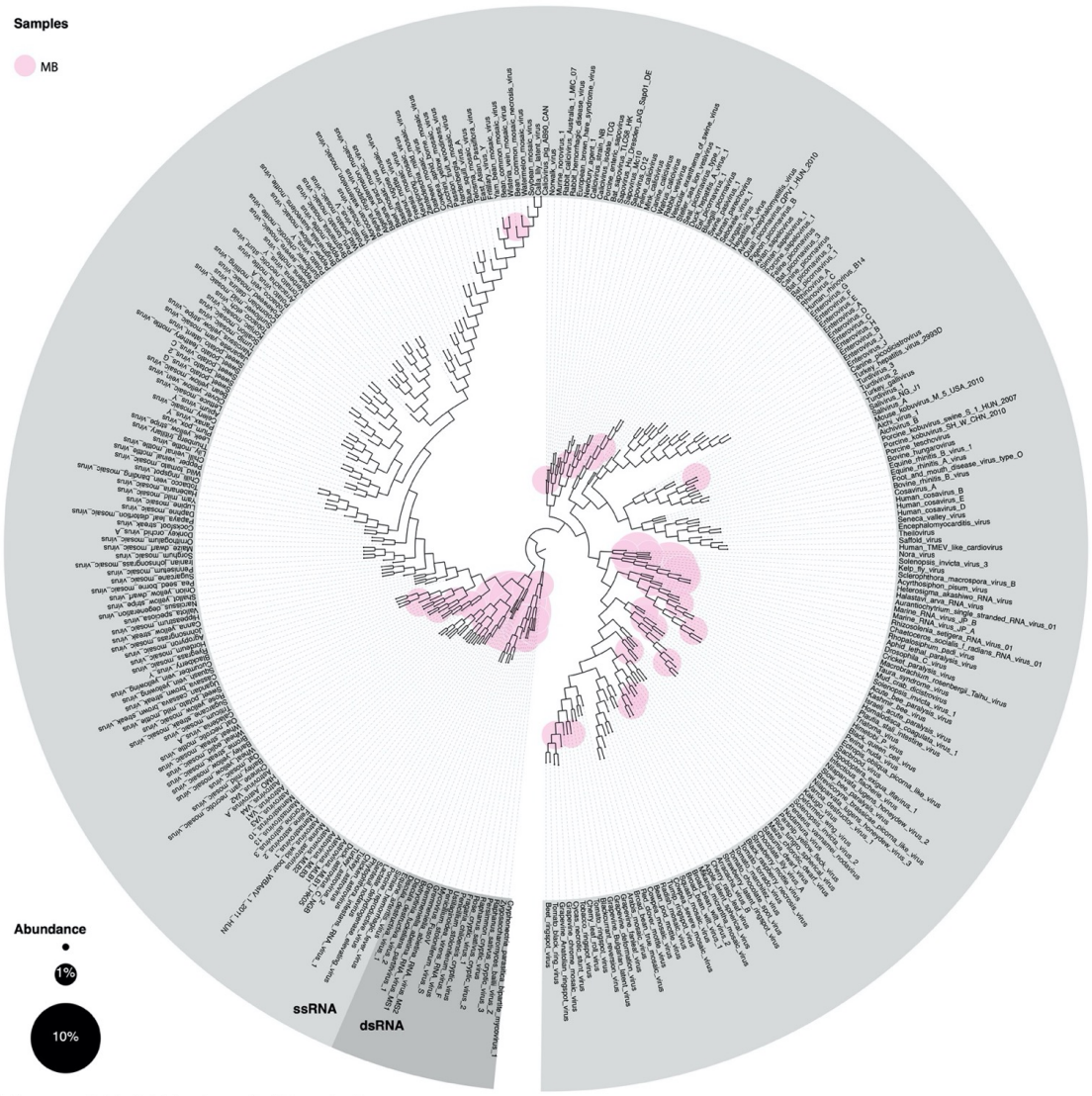


Figure S48 RNA virus RdRp phylogeny.

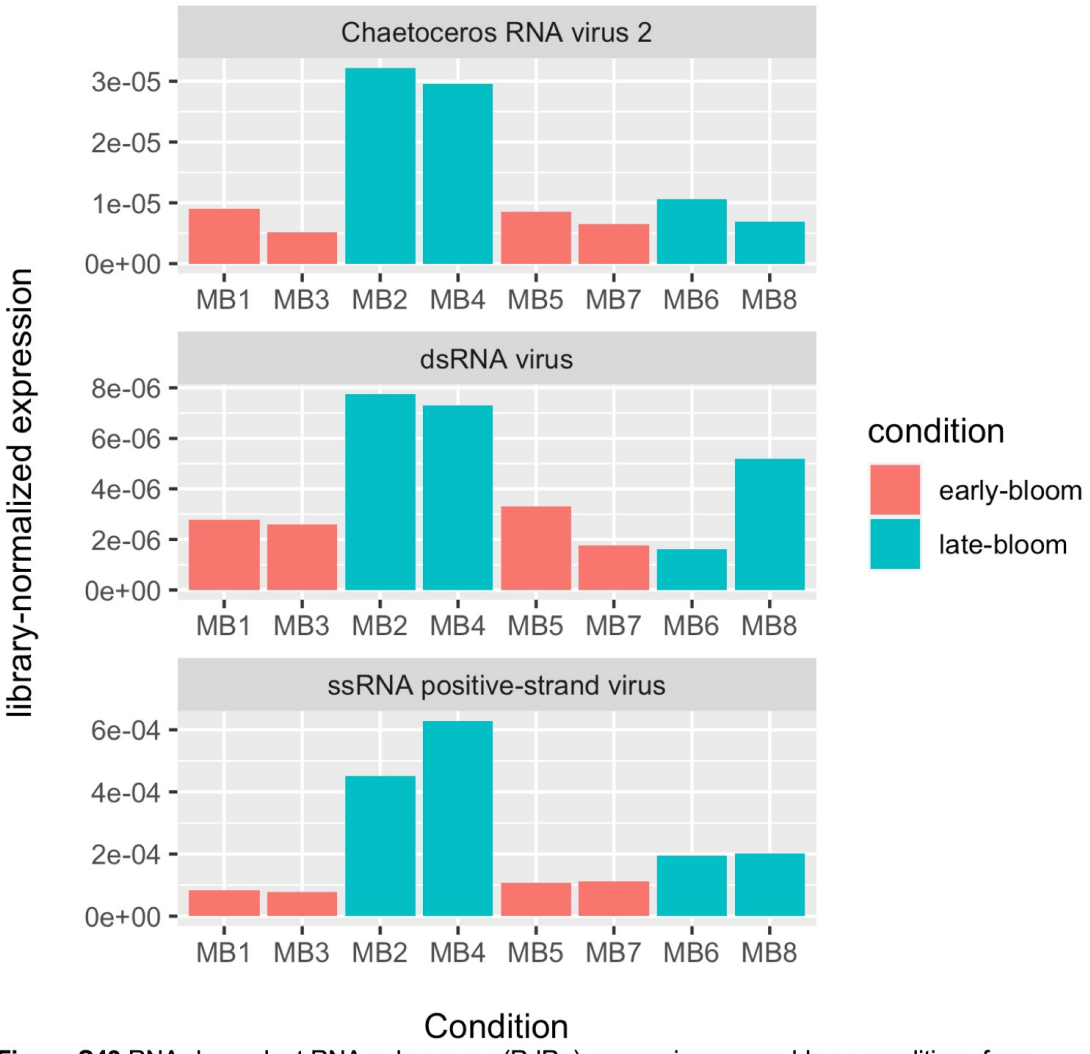


Figure S49 RNA-dependent RNA polymerase (RdRp) expression across bloom conditions for a diatom virus mapping mostly closely to Chaetoceros RNA virus 2, dsDNA viruses, and ssRNA positive-strand viruses.

References

1. Alipanah, L., Rohloff, J., Winge, P., Bones, A. M. & Brembu, T. Whole-cell response to nitrogen deprivation in the diatom Phaeodactylum tricornutum. *J. Exp. Bot.* **66**, 6281–6296 (2015).
2. Smith, S. R. *et al.* Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* **10**, 4552 (2019).
3. Lv, H., Wang, Q. e., Qi, B., He, J. & Jia, S. RNA-Seq and transcriptome analysis of nitrogen-deprivation responsive genes in Dunaliella salina TG strain. *Theor. Exp. Plant Physiol.* **31**, 139–155 (2019).
4. Alexander, H., Rouco, M., Haley, S. T. & Dyhrman, S. T. Transcriptional response of Emiliania huxleyi under changing nutrient environments in the North Pacific Subtropical Gyre. *Environ.*

Microbiol. **22**, 1847–1860 (2020).
5. Smith, S. R. *et al.* Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. *PLoS Genet.* **12**, (2016).
6. Cohen, N. R. *et al.* Diatom Transcriptional and Physiological Responses to Changes in Iron Bioavailability across Ocean Provinces. *Front. Mar. Sci.* **4**, (2017).
7. Cohen, N. R. *et al.* Transcriptomic and proteomic responses of the oceanic diatom Pseudo-nitzschia granii to iron limitation. *Environ. Microbiol.* **20**, 3109–3126 (2018).
8. Marchetti, A. *et al.* Development of a molecular-based index for assessing iron status in bloom-forming pennate diatoms. *J. Phycol.* **53**, 820–832 (2017).

Acknowledgements

Chapter two has been prepared as a submission to the ISME Journal. B. C. Kolody, S. R. Smith, L. Zeigler Allen, J. P. McCrow, D. Shi, B.M. Hopkinson, F. M. M. Morel, B.B. Ward, A. E. Allen. “Temporal succession of phytoplankton populations during simulated blooms under control of nitrogen and iron limitation.” The dissertation author was the primary investigator and author of this paper.

CHAPTER 3: THE IMPACT OF OCEAN BASIN-SCALE CIRCULATION ON THE S. PACIFIC MICROBIOME

B. Kolody, H. Zheng, S. G. Purkey, R. E. Sonnerup, E. E. Allen, A. E. Allen

Abstract

Microbial plankton are well-known for their role in carbon and nutrient recycling in the surface ocean. However, relatively little is known about their activity and community structure below the mesopelagic. Here, we present a high horizontal resolution, depth-resolved transect of microbial community structure in the South Pacific. We surveyed the diversity of microbial plankton along a gradient of water ages spanning newly subducted Antarctic water to subtropical water with an age since atmospheric contact >1,000 years. The mean age of each water sample was estimated using trace gas and radiocarbon concentrations. During the GO-SHIP P18 line (103°W from 26° 29.995' S to 69° 0.014'S), we collected molecular samples at a full range of water depths approximately every 2 degrees of latitude, employing size-fractionated filtering to distinguish putatively particle-associated microbes (> 5 μm) that are likely sinking from the surface from free-living plankton (> .22 μm). For each sample, we performed 16S and 18S rRNA diversity analyses using both DNA and cDNA reverse-transcribed from RNA, providing an estimate of the breadth of deep-ocean microbial diversity that can be attributed to active cells. We used Multiparameter Analysis to determine the water mass composition of each sample and identify active lineages associated with the major water masses of the region. We discuss to what extent microbial communities are structured by residence time in the ocean interior rather than their present-day physical environment (temperature, pressure), and how drivers of community structure differ between particle-associated and free-drifting cells. Collectively, these results

illustrate the potential biogeochemical insights to be gained by including molecular measurements in hydrographic sampling programs.

Introduction

Microbial plankton are essential for biogeochemical cycling of nutrients[1] and organic carbon[2]. While much basic research has investigated marine microbes on micro- scales (e.g. single-cells[3], between phytoplankton and bacteria[4,5]) and meso- scales (e.g. the photic zone[6], mixed layer[7], oxygen minimum zones[8], across seasonal dynamics[9,10], during phytoplankton blooms[11,12]) and in the surface ocean[13], a depth-resolved, ocean basin-scale understanding of microbial community dynamics has not yet been established. Mixing on this scale (tens of thousands of kilometers) is driven by density differences and occurs on timescales of hundreds to thousands of years[14]. This mixing, termed thermohaline circulation (THC), moves water (and microbial plankton) through a range of temperature (T) and pressure (P) extremes that would be fatal to most metazoan life[15]. The extent to which microbial communities are structured by, and resilient to, THC is still unknown, and has implications for global carbon cycling at all stages, including carbon export, remineralization, and deep ocean storage as recalcitrant dissolved organic matter[16]. The slow pace of THC reflects the density-stratification of our ocean, which is composed of water masses with well-established T and salinity (S) characteristics[17]. While recent studies suggest that sinking particles cannot entirely explain the microbial community structure of the deep ocean[18,19,20], it is still unknown to what extent water masses represent unique habitats with characteristic microbial communities.

Small-scale studies suggest that water masses contain endemic microbes[21] and that advection shapes microbial communities[22,23]; generating a comprehensive understanding of

pelagic microbial ecology would require complementing the physical dynamics of THC with molecular observations of microbes from depth-resolved transects. Previous expeditions have sought to systematically survey the world oceans, beginning with the *Sorcerer II*'s Global Ocean Sampling[24,25] (GOS; 2003-2007) and becoming increasingly sophisticated and depth-resolved with TARA Oceans[26–28] (2009-2013) and the 2010 Malaspina Expedition[29,30]. However, the exponential decrease in microbial biomass with depth has, until recently, posed a technical challenge for genomic sequencing below the epipelagic. The Malaspina Expedition was the only large-scale ocean sampling project that reached the bathypelagic, and the ocean interior has not been sampled on a fine-enough spatial scale or with the necessary accompanying physical measurements (e.g. atmospheric tracers[31] and dissolved inorganic carbon 14 ($DI^{14}C$)[32] for aging water parcels) to infer the effects of THC. Because of this, pelagic microbial community structure is typically explained in terms of state variables (e.g. T, P, latitude) rather than tracking the path-dependent history of cells.

Unfortunately, the ability to explain microbial community structure in terms of any environmental metadata is undermined by the compositional nature of metabarcoding data. Because microbial abundance must be viewed as relative to the total reads in a given sequencing library, a taxon with relative abundance that is positively correlated with a given variable may actually be negatively correlated in absolute terms[33]. This problem can be avoided by adding DNA spike-ins to estimate the absolute copies/mL of a given ribotype[34], but such methods have not yet been applied to the deep ocean, which poses a technical challenge due to widely ranging biomass across depths. Furthermore, microbial surveys of the deep ocean have thus far relied on ribotyping DNA, which may be preserved in the cold, saline deep ocean[35], and could be exogenous or represent dead or inactive cells. While bulk 3H-leucine incorporation has shown

that heterotrophic deep ocean microbes are active[18,36], such methods cannot identify which taxonomic fraction of known biodiversity is viable.

Here, we address the fundamental problem of how THC shapes the community structure of pelagic marine microbes by analyzing high horizontal resolution, depth-resolved samples from the South Pacific. Samples were collected on leg 2 of the Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP) P18 line (Figure 3.1). We asked how the community structure of active taxa, as measured by RNA, differs from the traditional view of community structure measured by DNA, and identified active lineages in the deep ocean. We also modeled the age and mixing fractions of S. Pacific water masses and asked to what extent residence time and hydrographic water mass "biomes" structure pelagic communities.

Methods

Sample collection

Sampling was conducted on board the NOAA Ship *Ronald H. Brown* from January 2-29, 2017 during the second leg of the Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP) P18 line. This cruise followed 103°W from 26° 29.995' S to 69° 0.014'S, sampling approximately every .5 degree of latitude. Each deployment consisted of lowering a Sea-Bird Electronics SBE9plus CTD, connected to a 24-place SBE32 carousel, to within 8-12 meters of the bottom using the altimeter on the CTD-rosette package (with the exception of 3 stations). Molecular samples were collected approximately every 4th station (~2 degrees of latitude) at a full range of water column depths. A total of 302 samples were taken over 25 stations. At each station, ~12 sampling depths were chosen to target the major water mass features predicted by the previous P18 iteration, incorporate as wide a range as possible of pressure conditions, and maximize available water.

Preservation of cellular material

Seawater collected for molecular analyses was filtered, over ice, into 5µm and .22µm size fractions in order to capture particle-associated and free-living microbes, respectively. Volumes ranged from 3-8 liters depending on availability. Filters were immediately submersed in RNAlater™ and stored in liquid nitrogen. In addition, 1 mL of sample was preserved in 1% (final concentration) paraformaldehyde for flow cytometry. All samples were preserved in liquid nitrogen throughout transit before being transferred to a -80ºC freezer.

Complementary Data Collections

Physical and chemical data collected on the P18 line can be accessed at the CLIVAR and Carbon Hydrographic Data Office (CCHDO) website (https://cchdo.ucsd.edu/cruise/33RO20161119). Pressure, temperature, salinity, and oxygen were collected continuously with each cast. The CTD system also incorporated an altimeter, transmissometer, fluorometer/backscatter (FLBB) sensor, a Lowered Acoustic Doppler Profiler (LADCP) for measuring velocity profiles, and Chipods for measuring fine-scale temperature structure.

Once surfaced, bottles were immediately and expediently sampled. Core measurements (bottle oxygen and salinity, CFC-11, CFC-12, SF6, silicate, nitrate, nitrite, phosphate, pH, and total alkalinity) were collected at 24 depths per cast. ^{3}He, Ne, tritium, N_2O, stable gases (N_2, O_2, Ar), dissolved organic carbon-14 ($DO^{14}C$), dissolved inorganic carbon-14 ($DI^{14}C$), particulate organic carbon (POC), black carbon, and rare earth elements were also measured secondarily. The chlorophyll maximum was sampled for phytoplankton pigment analysis with HPLC for calibration of satellite data when satellites were overhead and unobstructed by weather, and genetics samples were collected from the surface Niskin bottle approximately every degree of

latitude. POM samples were collected for POC, particulate organic nitrogen (PON), particulate organic phosphorous (POP) and biological oxygen demand (BOD) from the flow-through system at a total of 198 stations. Measurements not sampled at molecular sampling locations (POC, PON, POP, BOD) were interpolated to the locations of genomics samples using bivariate linear interpolation over latitude and depth via the R akima package [37].

Cell Counts

Prochlorococcus sp., *Synechococcus* sp., heterotrophic bacteria and photosynthetic eukaryotes were enumerated via flow cytometry at the SOEST Flow Cytometry Facility (www.soest.hawaii.edu/sfcf). Samples were thawed in batches and stained as previously described[38–40] with 1 µg/ml final concentration Hoechst 34442. The flow cytometer used was a Beckman-Coulter Altra mated to a Harvard Apparatus syringe pump for quantitative analyses, equipped with two argon ion lasers, tuned to UV (200 mW) and 488 nm (1 W) excitation. Scatter (side and forward) and fluorescence signals were collected using filters as appropriate, including those for Hoechst-bound DNA, phycoerythrin and chlorophyll. The data generated was in the form of listmode files (FCS 2.0 format) acquired from the flow cytometer using Expo32 software (Beckman-Coulter). Population designations, based on the scatter and fluorescence signals, were generated from the listmode files using FlowJo software (Tree Star, Inc., www.flowjo.com).

Library Preparation and Sequencing

For molecular analyses, half of each filter was used for DNA-based analyses and half for RNA-based analyses. DNA and RNA were extracted using the NucleoMag Plant kit and the NucleoMag RNA kit (Macherey-Nagel), respectively, using an epMotion TMX automated liquid handling system (Eppendorf). *Schizosaccharomyces pombe* gDNA (ATCC 356 #24843D-5, Manassas, VA, USA) and *Thermus thermophiles* gDNA (ATCC 27634D-5) with known rRNA

gene copy numbers were used as internal standards for 18S and 16S analyses, respectively [34]. For DNA analyses, internal standards were spiked-in pre-extraction to a tube containing the sample filter and lysis buffer. For RNA analyses, internal standards were spiked in after the cDNA synthesis.

Samples were binned into 3 groups based on collection depth (0-200dbar, 200-500dbar, >500dbar) as a proxy for biomass in order to allow spike-ins to account for ~1% of sample reads. ZymoBIOMICS whole-cell microbial community standards (D6300) were used as a control for DNA and RNA analyses and were serially diluted to 10^9, 10^7, 10^5, 10^3, 10^2, 10, 1, and 0 cells/mL with ZymoBIOMICS DNA/RNA Shield™ (Cat. No. R1100-50).

The V4-V5 region of 16S was targeted with the degenerate primer pair F515 (GTGNCAGCMGCCG CGGTAA) and R926 (CCGYCAATTYMTTTRAGTTT) [41,42]. The V9 region of 18S rRNA and rDNA were amplified with the primers 1389F (5′-TTGTACACACCGCCC-3') and 1510R (5′-CCTTCYGCAGGTTCACCTAC-3'), which capture about 150bp [43]. 250 riboTag libraries were multiplexed per MiSeq run (~20 million reads/run).

Metabarcoding analyses

Ribosomal RNA reads were processed using qiime2 version 2019.4[44]. Sequences were trimmed to remove primers with cutadapt version 2019.10.0[45], and amplicon sequence variants (ASVs) were called using DADA2 version 2019.10.0[46] with custom forward and reverse read trimming based on quality score visualizations. Taxonomy was assigned using qiime feature classifiers. For the 16S amplicon, the feature classifier was trained on SILVA132 (www.arb-silva.de), including 16S and 18S sequences, using the 99% identity clustered core alignment file. For the 18S amplicon, the classifier was trained on version 4.12.0 of the protist ribosomal database[47].

DNA amplicon copies/mL were approximated as previously described[34] using the volume of seawater filtered, percent of spike-in retained and genome size and amplicon copy number of spike-in species. RNA amplicon copies/mL were similarly estimated, but multiplied by a factor of 7.5 to account for only 4 uL of the 30uL RNA extract being used for cDNA synthesis before adding the spike-in.

Partial correlation analysis was performed on amplicon counts/mL in the ppcor[48] R package. Differential abundance analysis was performed on amplicon counts/mL using the edgeR[49] exact.test() function with tagwise estimated dispersion. Data manipulation and visualization was aided by the use of the R packages plyr[50], dplyr[51], tidyr[52], stringr[53], reshape2[54], and ggplot2[55].

Water mass classification

Optimum MultiParameter (OMP) analysis[56,57] was used to determine the mixing fractions of the water masses present in each sample. This method uses conservative water mass properties (temperature, salinity) as well as quasi-conservative properties (oxygen and nutrients, taken together as a single variable) to determine the proportion of water in each sample originating from various water masses.

Established temperature, salinity, and nutrient characteristics [58] were used to choose precise coordinates and depths/potential densities for water mass end members. The remaining water mass characteristics (salinity, potential temperature, oxygen, phosphate, nitrate, and silicate) were then determined at the given location using the World Ocean Circulation Experiment (WOCE) Global Hydrographic Climatology (downloaded 1/15/19, last modified 12/23/2010) data, parsed using the R ncdf4 package [59]. For Antarctic Bottom Water (AABW), North Pacific Intermediate Water (NPIW), and Antarctic Intermediate Water (AAIW), defining

latitudes, longitudes, potential densities and potential vorticities were defined as previously described[60]. For upper and lower circumpolar deep water, a high nutrient oxygen minimum layer and a low nutrient salinity maximum layer, respectively[61–65], were searched for on WOCE Pacific section plots in order to define water mass coordinates and depths.

Water mass age determination

First, CFC-11, CFC-12, and SF6 were used to assign a tracer age to samples wherever possible. Empirically derived CFC-11, CFC-12[66], and SF6[67] solubilities were used to determine the atmospheric concentration of each tracer at the time the given water parcel equilibrated with the atmosphere. These concentrations were then compared with the global mean NOAA atmospheric tracer record and interpolated using the R splines function[68] in order to determine the average year that the sampled water was in contact with the atmosphere. This value represents the "tracer age," or weighted average of the water sampled, because in truth, the water is an amalgam of water parcels that equilibrated with the atmosphere at different times. The extent to which the true "advective age" aligns with this tracer age depends up the degree of mixing the water parcel has experienced. The advective age can be laboriously calculated using transit-time distribution (TTD) theory[69], but doing so is outside of the scope of this project. Instead, tracer-aged samples were compared against TTD calculations made for the previous occupation of the P18 line[70] as a measure of quality control.

Water masses older than can be predicted with CFC and SF6 data were aged using dissolved inorganic carbon 14 dating (DI14C). High-precision Δ14DIC measurements were processed at Woods Hole Oceanographic Institution's (WHOI) National Ocean Sciences Accelerator Mass Spectrometry lab (NOSAMS) using standard AMS (accelerator mass spectrometry) protocols[71,72]. Δ14DIC was converted into radiocarbon age using the equation:

14C age = -8033*log(1+(Δ14DIC /1000))[73]. Radiocarbon age was converted into calibrated years before 1950 (cal BP) using the standard marine radiocarbon calibration curve, Marine13[74] interpolated via the R splines function[68]. "Consensus age" denotes the age of water relative to when the samples were taken (2017) and used the average of atmospheric tracers, when available, and radiocarbon-derived ages at greater depths. The consensus age was interpolated to the locations of genomics samples using bivariate linear interpolation over latitude and depth via the R akima package[37].

Results and Discussion

Here, we surveyed microbial communities along 103°W from 26° 29.995' S to 69° 0.014'S across the full depth of the water column (Figure 3.1). In order to assess differences between particle-associated and free-living microbes, samples were size fractionated on to operationally large (5 μm) and small (0.22 μm) size class filters. In order to estimate the viability of the observed community, each filter was cut in half to be used for both DNA and RNA amplicon analyses (Figure 3.2), where RNA amplicon copies/mL is used as to investigate the active community. Samples encompasses a large breadth of water properties (nutrients, salinity, temperature, etc.) as well as the major water masses of the region (Figure 3.3).

Community structure differences between DNA and RNA amplicons vary across size classes

Traditionally, marine microbial community structure has been observed by amplifying the 16S SSU rRNA gene from extracted whole-community DNA, but there is some concern that such methods are biased by the inclusion of DNA from dead and inactive cells that may be preserved in the deep ocean[35]. Here, we report for the first time on the differences between community structure based on traditional DNA-based amplicon sequencing, and SSU fragments amplified from extracted RNA, which is likely to better recapitulate the active community.

Overall, the 16S community structure agreed with recent global surveys[26]. *Proteobacteria* (orange) were dominant across latitudes in both DNA and RNA libraries, followed by *Thaumarchaeota* (grey), SAR406 (yellow), cyanobacteria (sky blue), and *Chloroflexi* (brown; Figure 3.4, Figure 3.5). Even at a coarse level, taxonomic differences can be observed between DNA and RNA-extracted libraries (top versus bottom panels, Figure 3.4). Cyanobacteria (FDR < 3.4e-9, log2FC = 2.3, average of $1.3x10^5$ RNA counts/mL), SAR406 (FDR < 1.5e-15, log2FC = .8, average of $5.8x10^4$ RNA counts/mL), and *Chloroflexi* (FDR < 4.2e-10, log2FC = 1.2, average of $3.7x10^4$ RNA counts/mL), for example, were underrepresented in DNA libraries compared to RNA libraries, especially in the free-living fraction, and planctomycetes were especially underrepresented by DNA libraries in the large fraction (FDR< 1.8e-13, log2FC= 1, average of $1.9x10^4$ RNA counts/ml). *Proteobacteria* were the most abundant group significantly overrepresented in DNA libraries (FDR < 3e-46, log2FC = -1.8, average of $3x10^5$ DNA counts/mL), followed by *Bacteroidetes* (FDR < 6.3e-17, log2FC = -1.3, average of $1.9x10^4$ DNA counts/mL), *Euryarchaeota* (FDR < 1.4e-62, log2FC = -2.5, average of $3.4x10^3$ DNA counts/mL), and *Acidobacteria* (FDR < 2e-7, log2FC = -1.1, average of $1.3x10^3$ DNA counts/mL).

Differential abundance of taxa groups across extraction types and size classes also varied in different water masses (Figure 3.6). For example, *Proteobacteria* were more abundant in the free-living fraction in upper waters but were more abundant in the larger size class in deeper water masses, presumably colonizing particles. *Euryarchaeota* made a similar, if less dramatic, transition. The nitrite-oxidizing genus, *Nitrospinae*[75], was more abundant in DNA libraries in most water masses, but more abundant in RNA libraries in upper water and LCDW.

Surprisingly, RNA communities were significantly more diverse than DNA communities (Faith's phylogenetic diversity, p = 4e-29). Most amplicon sequence variants (ASVs) were also present in more RNA libraries than DNA libraries (Figure 3.7), perhaps because active taxa have more copies of ribosomal RNA in use than in their genomes and are thus more easily detectable. One striking exception was the most ubiquitous ASV, a *Pseudomonas* species that was present in 415 DNA libraries but was only present in 43 RNA libraries. This ASV was highly abundant, averaging $1.9x10^5$ DNA copies/mL, and illustrates the proteobacterial trend of blooming as a free-living organism in the surface ocean and subsequently making up dead cells in deep ocean particles (Figure 3.6).

Environmental parameters structuring marine microbial communities

Previously, temperature was found to be the strongest environmental driver of 16S microbial community structure down to the mesopelagic[26]. Here, we also see a relationship with temperature, but find that the strongest forces structuring communities are extraction type (DNA vs RNA), size class, water mass, and water parcel residence time (years since atmospheric contact; Figure 3.8). With the exception of size class, these parameters also structure 18S community structure, which has not previously been measured on these scales.

The first 3 principal components of Bray-Curtis dissimilarity captured more variance for 16S than 18S, indicating that the 16S community does not contain as much high-dimensional complexity as 18S. 18S samples were clearly separated by extraction type on axis 1, whereas 16S samples grouped into one large lobe containing a mix of DNA and RNA libraries, and another smaller lobe containing only DNA libraries. This indicates that for eukaryotes, the vast majority of samples contain enough DNA from inactive cells to drastically shift the community, whereas for prokaryotes, only some samples do. Puzzlingly, both the 16S DNA-only cluster and

the mixed DNA/RNA cluster contain samples with from diversity of water masses and water ages, and from both the large and small size classes. None of the measured environmental parameters suggested a cause for their separation.

Size classes did not noticeably structure 18S communities. For 16S, small size-class samples clustered around the edges of each lobe, with large size-class samples in the middle. In this way, large and small size class samples from the same water masses were similar to each other, but still structured more broadly by water age.

Here, for the first time, we report that both 16S and 18S community structure is strongly structured by water mass and water age. A partial correlation analysis revealed taxa groups with RNA counts/mL significantly correlated with water age, after controlling for depth. In the large size class, no D1 level group was positively correlated with water age, but *Omnitrophicaeota*, as well as the nitrite oxidizing group, *Nitrospirae*, were negatively correlated with water age. In the small size class, the endosymbiotic *Elusimicrobia* group was most negatively correlated with water age (partial correlation coefficient = -.28, FDR < 0.002), likely because hosts (e.g. flagellates) are more abundant in the surface ocean. In the small size class, *Firmicutes* were positively correlated with water age (partial correlation coefficient = .25 FDR < 0.01). Firmicutes are gram-positive bacteria known for their ability to produce endospores and survive in extreme conditions, but a positive correlation of RNA copies/mL with water age points to their ability to maintain some level of activity in the deep ocean. Firmicutes abundance here may actually be an underestimate, as this group is often resistant to DNA isolation and underrepresented in environmental metagenomes[76].

Conclusions

Here, we present a preliminary look into a high horizontal resolution, depth-resolved transect of the South Pacific. Integrating water mass models and radiocarbon ages with samples from below the mesopelagic has allowed for a different view of the drivers of global marine biogeography. We show that the major drivers of community structure are water mass membership and residence time, indicating a strong role for thermohaline circulation in structuring communities. Additionally, we show that the traditional DNA-amplicon metric of community structure underrepresents active taxa, such as cyanobacteria, and overrepresents *Proteobacteria*. Future work will consider finer taxonomic nuances when addressing size class and DNA/RNA differences and species endemic each water mass, as well as quantify more explicitly the contributions of water mass, residence time, and classical physical parameters (temperature, pressure, salinity, oxygen, nutrients) to community structure and diversity.

Samples analyzed here were collected as an ancillary project on a physical and chemical oceanographic sampling initiative. These results speak to taxonomic resolution that is achievable with small water volumes (2-8 L), which can be collected ad-hoc without interfering with core cruise initiatives. We show that a molecular-based sampling program could be integrated with current repeat sampling initiatives, which would allow for a more comprehensive picture of global marine microbial ecosystems.

Figures

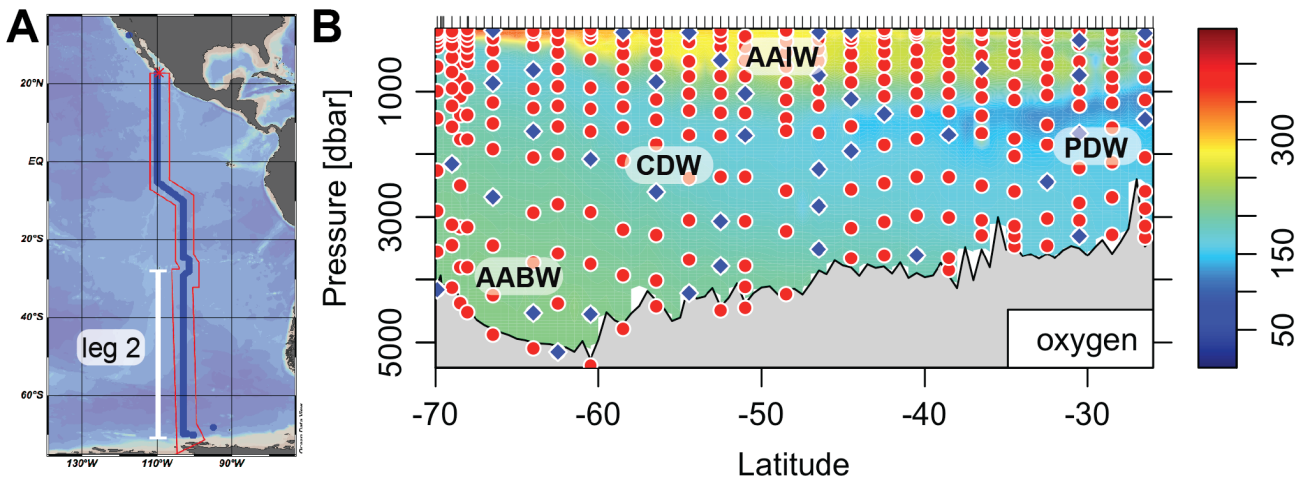


Figure 3.1: (**A**) P18 line with leg 2 labeled in white. (**B**) Samples processed for metabarcoding analysis (red dots) and reserved for metagenomes (blue diamonds) overlay a leg 2 oxygen (µM) section with labeled water masses (Antarctic Bottom Water (AABW), Antarctic Intermediate Water (AAIW), Pacific Deep Water (PDW), Circumpolar Deep Water (CDW)).

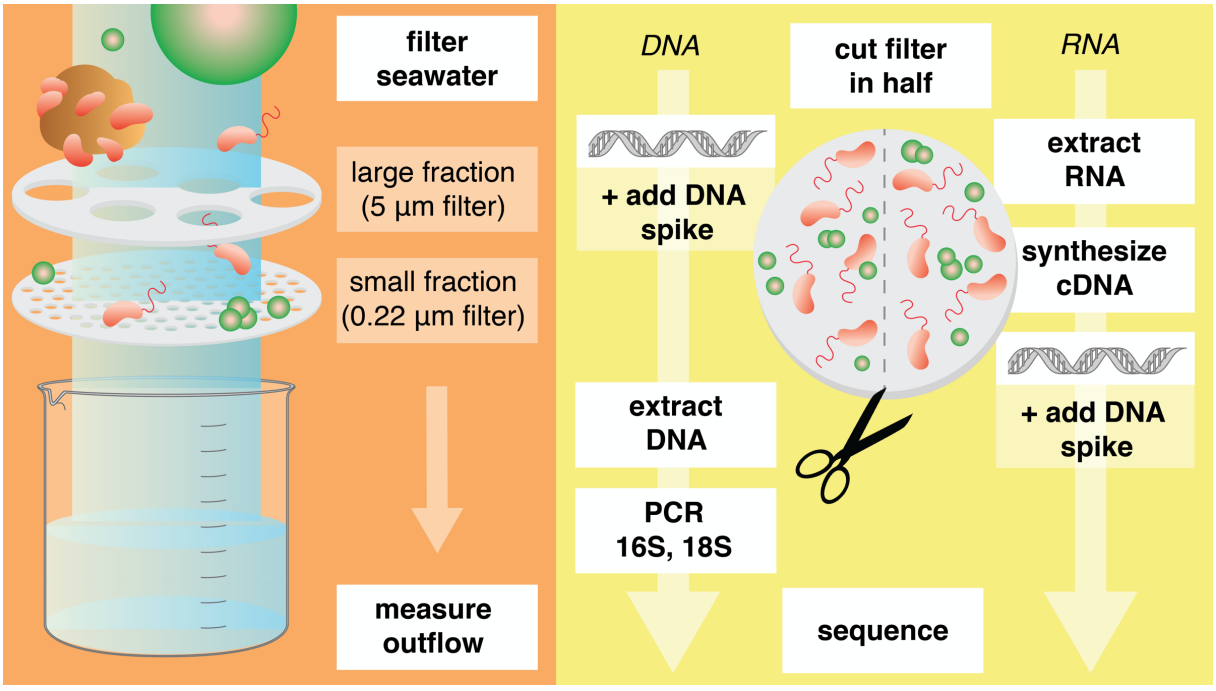
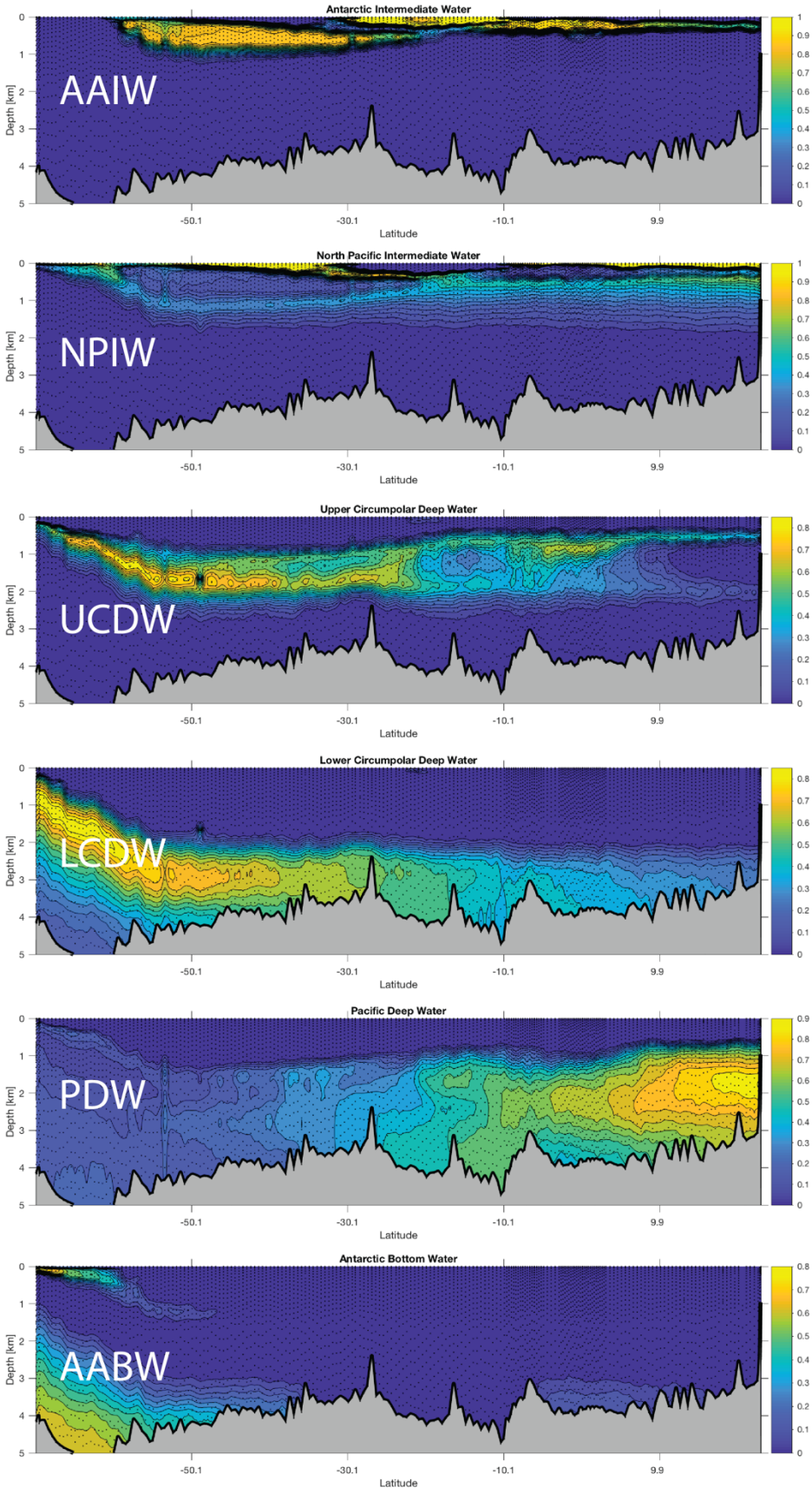


Figure 3.2: (**A**) Size-fractionated filtration pipeline and (**B**) metabarcoding processing pipeline.

Figure 3.3: Water mass mixing fractions modeled by OMP (AABW = Antarctic Bottom Water, AAIW = Antarctic Intermediate Water; LCDW= Lower Circumpolar Deep Water; UCDW= Upper Circumpolar Deep Water). Color indicates fraction of a given water mass (yellow= 100%, blue = 0%).


Antarctic Intermediate Water
AAIW
North Pacific Intermediate Water
NPIW
Upper Circumpolar Deep Water
UCDW
Lower Circumpolar Deep Water
LCDW
Pacific Deep Water
PDW
Antarctic Bottom Water
AABW
Depth [km]
Latitude
-50.1
-30.1
-10.1
9.9

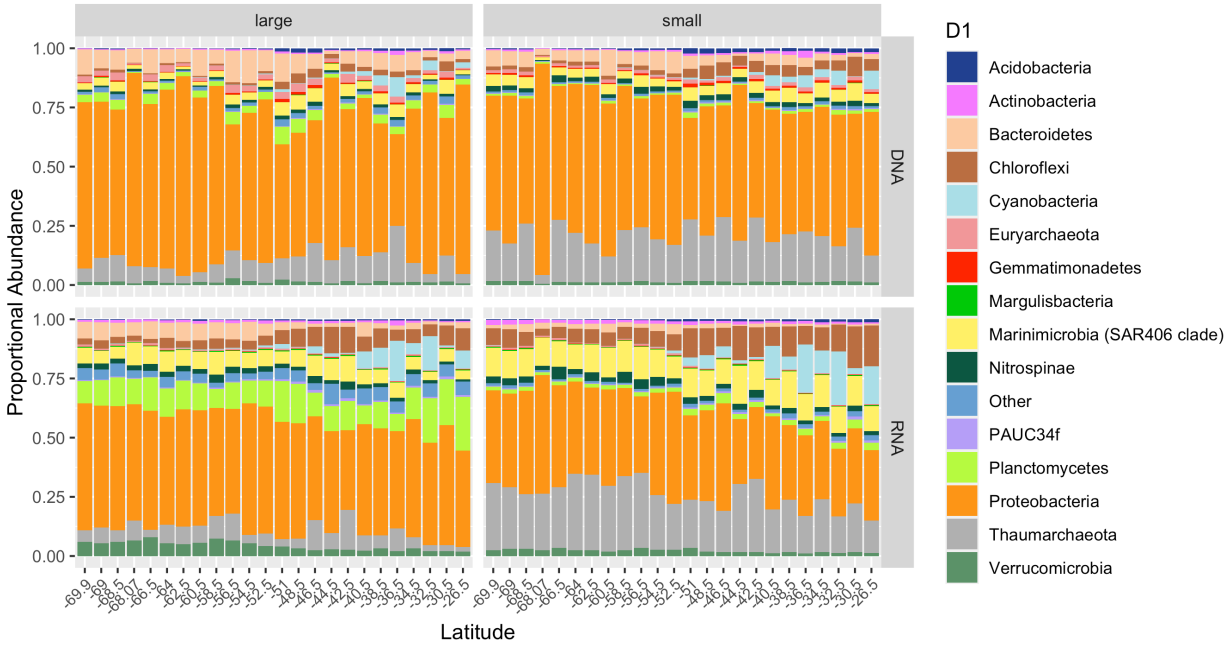


Figure 3.4: Coarse 16S community structure across latitude by size class and extraction type.

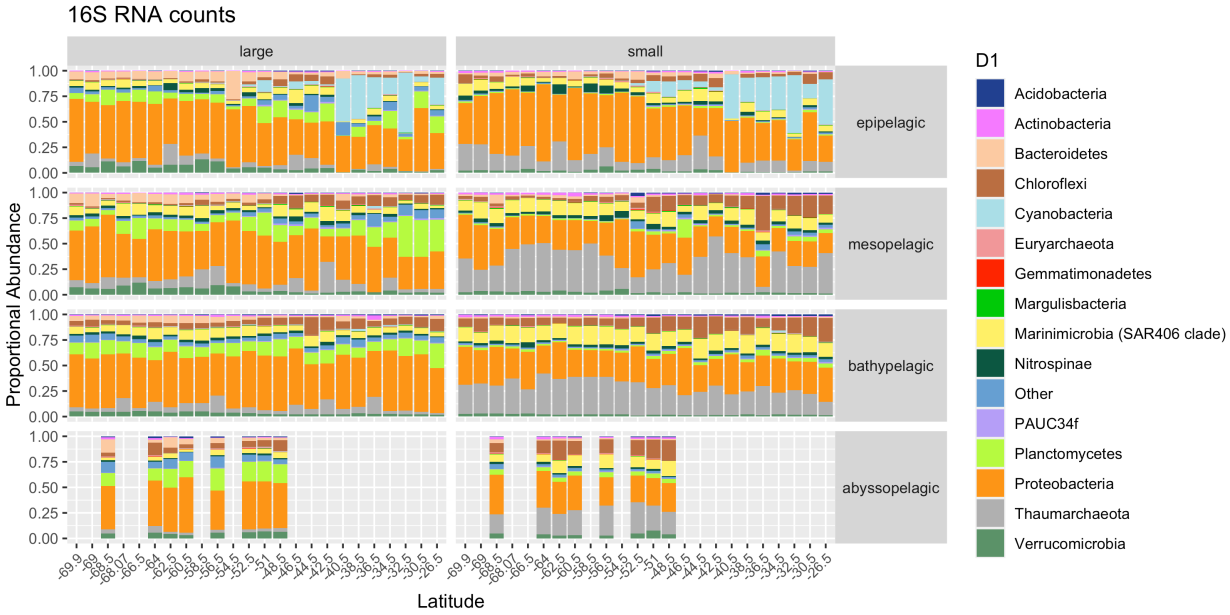


Figure 3.5: Coarse 16S RNA community structure across latitude by size class and depth zone.

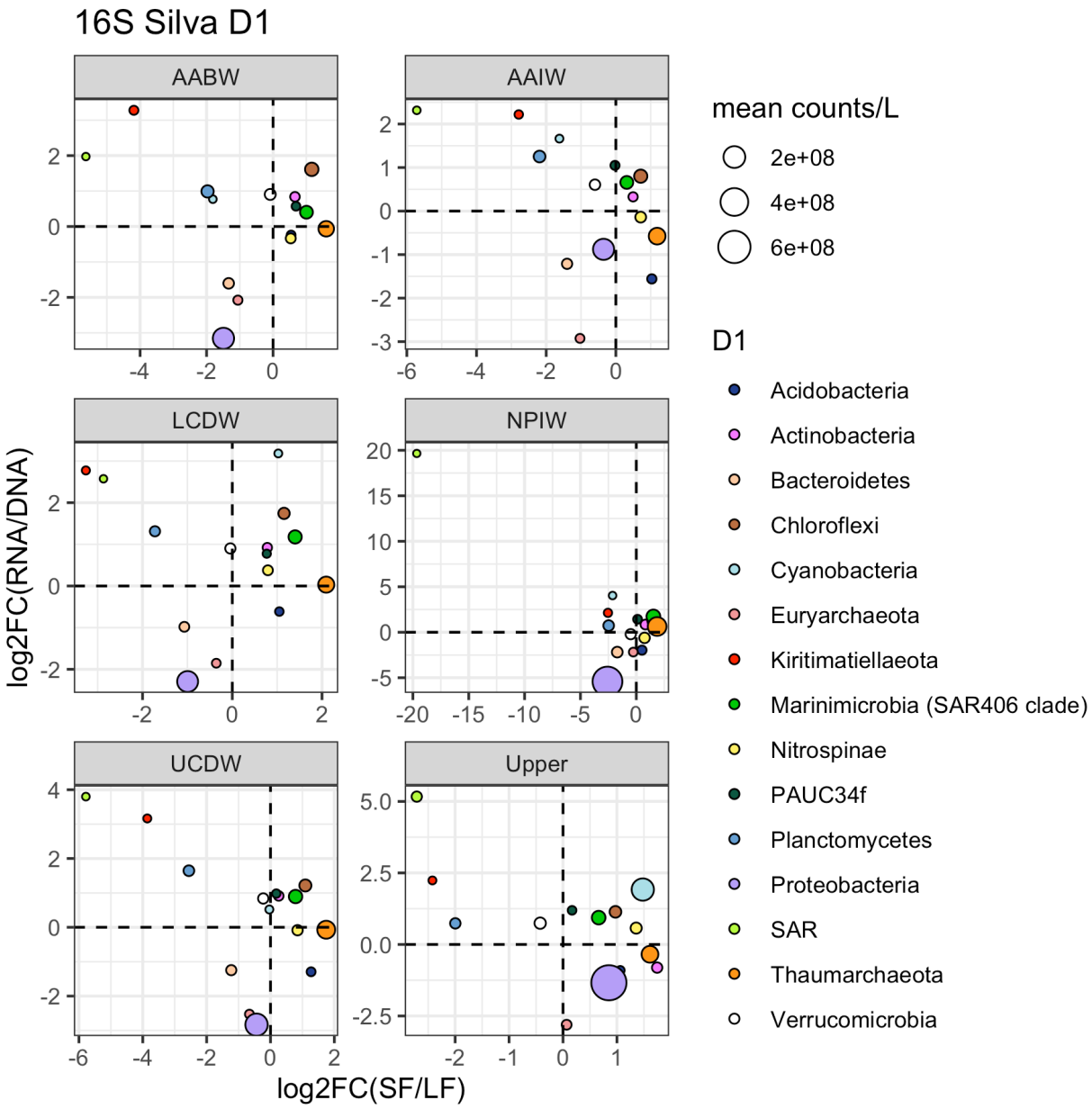


Figure 3.6: Log_2FC of 16S D1 level taxa across size class (x-axis; positive = up small size class, negative = up large size class) and extraction type (y-axis; positive = up RNA, negative = up DNA) for each water mass (AABW = Antarctic Bottom Water, AAIW = Antarctic Intermediate Water; LCDW= Lower Circumpolar Deep Water; UCDW= Upper Circumpolar Deep Water; Upper = water above 500m). Scales are independent for each water mass plot. Only top 15 most abundant D1 groups are shown.

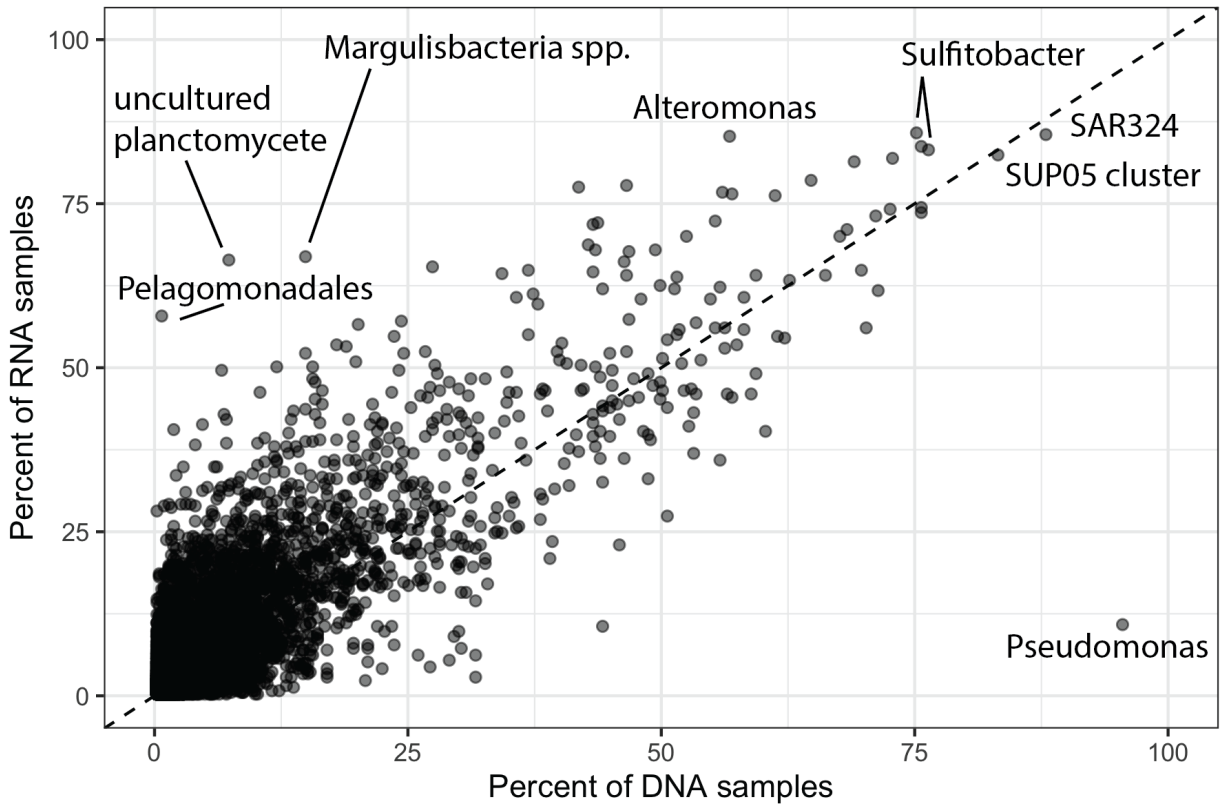
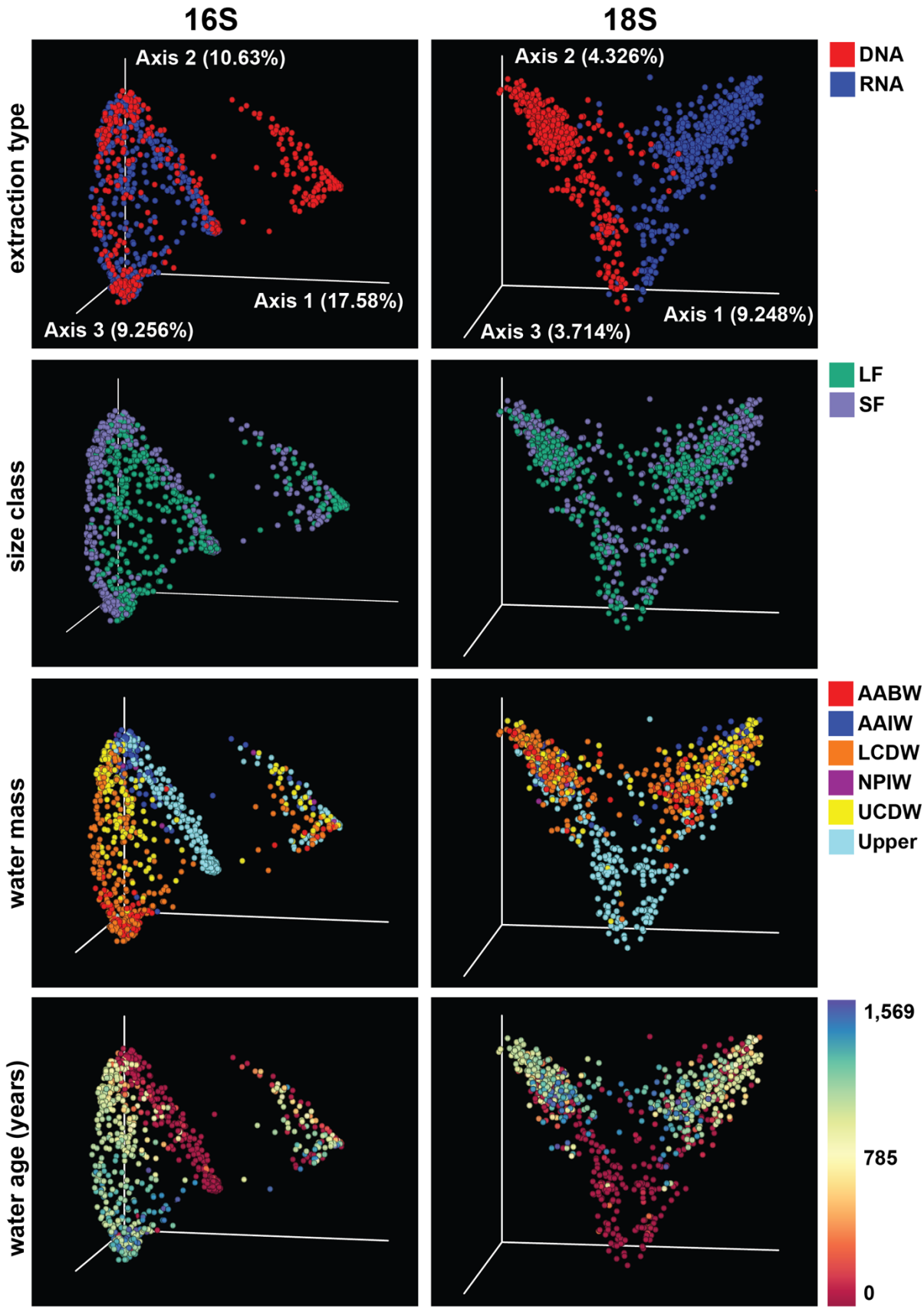


Figure 3.7: Ubiquity of 16S ASVs in RNA and DNA libraries. Y-axis: percent of RNA libraries in which ASV is present; x-axis: percent of DNA Libraries in which ASVS is present. Taxonomy of outlying samples is labeled.

Figure 3.8: PCoA plots depicting Bray-Curtis dissimilarity for 16S and 18S amplicons. Samples are colored by extraction type, size class (LF= large fraction; SF= small fraction), water mass (AABW = Antarctic Bottom Water, AAIW = Antarctic Intermediate Water; LCDW= Lower Circumpolar Deep Water; UCDW= Upper Circumpolar Deep Water; Upper = water above 500m), and water age. Top panels give the percent of variance explained by each axis.


16S
18S
extraction type
size class
water mass
water age (years)
Axis 2 (10.63%)
Axis 1 (17.58%)
Axis 3 (9.256%)
Axis 2 (4.326%)
Axis 1 (9.248%)
Axis 3 (3.714%)
DNA
RNA
LF
SF
AABW
AAIW
LCDW
NPIW
UCDW
Upper
1,569
785
0

References

1. Arrigo, K. R. Marine microorganisms and global nutrient cycles. *Nature* **437**, 349–355 (2005).
2. Azam, F., Fenchel, T., Field, J., Gray, J., Meyer-Reil, L. & Thingstad, F. The Ecological Role of Water-Column Microbes in the Sea. *Mar. Ecol. Prog. Ser.* **10**, 257–263 (1983).
3. Kashtan, N., Roggensack, S. E., Rodrigue, S., Thompson, J. W., Biller, S. J., Coe, A., Ding, H., Marttinen, P., Malmstrom, R. R., Stocker, R., Follows, M. J., Stepanauskas, R. & Chisholm, S. W. Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. *Science (80-.).* **344**, 416–420 (2014).
4. Bertrand, E. M., McCrow, J. P., Moustafa, A., Zheng, H., McQuaid, J. B., Delmont, T. O., Post, A. F., Sipler, R. E., Spackeen, J. L., Xu, K., Bronk, D. A., Hutchins, D. A. & Allen, A. E. Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9938–43 (2015).
5. Amin, S. A., Hmelo, L. R., Van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K. R., Morales, R. L., Berthiaume, C. T., Parker, M. S., Djunaedi, B., Ingalls, A. E., Parsek, M. R., Moran, M. A. & Armbrust, E. V. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**, 98–101 (2015).
6. Mende, D. R., Bryant, J. A., Aylward, F. O., Eppley, J. M., Nielsen, T., Karl, D. M. & DeLong, E. F. Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat. Microbiol.* (2017) doi:10.1038/s41564-017-0008-3.
7. Fernández-Castro, B., Mouriño-Carballido, B., Benítez-Barrios, V. M., Chouciño, P., Fraile-Nuez, E., Graña, R., Piedeleu, M. & Rodríguez-Santana, A. Microstructure turbulence and diffusivity parameterization in the tropical and subtropical Atlantic, Pacific

and Indian Oceans during the Malaspina 2010 expedition. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **94**, 15–30 (2014).

8. Ganesh, S., Parris, D. J., Delong, E. F. & Stewart, F. J. Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. **8**, 187–211 (2013).
9. Ladau, J., Sharpton, T. J., Finucane, M. M., Jospin, G., Kembel, S. W., O'Dwyer, J., Koeppel, A. F., Green, J. L. & Pollard, K. S. Global marine bacterial diversity peaks at high latitudes in winter. *ISME J.* **7**, 1669–1677 (2013).
10. Carlson, C. A., Morris, R., Parsons, R., Treusch, A. H., Giovannoni, S. J. & Vergin, K. Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J.* **3**, 283–295 (2009).
11. Needham, D. M. & Fuhrman, J. A. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat. Microbiol.* **1**, 16005 (2016).
12. Buchan, A., LeCleir, G. R., Gulvik, C. A. & Gonzalez, J. M. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat. Rev. Microbiol.* **12**, (2014).
13. Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T., Rappé, M. S., MacLellan, S. L., Lücker, S. & Eren, A. M. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* **3**, 804–813 (2018).
14. Schmitz, W. J. On the interbasin-scale thermohaline circulation. 151–173 (1995).
15. Pörtner, H. Climate change and temperature-dependent biogeography: Oxygen limitation of thermal tolerance in animals. *Naturwissenschaften* **88**, 137–146 (2001).
16. Jiao, N., Herndl, G. J., Hansell, D. A., Benner, R., Kattner, G., Wilhelm, S. W., Kirchman,

D. L., Weinbauer, M. G., Luo, T., Chen, F. & Azam, F. Microbial production of recalcitrant dissolved organic matter: Long-term carbon storage in the global ocean. *Nat. Rev. Microbiol.* **8**, 593–599 (2010).

17. Emery, W. & Meincke, J. Global water masses: summary and review. *Oceanol. acta* **9**, 383–391 (1986).

18. Marinos, E., Latitudes, A., Lia-morfun, L., Global, C., Palmas, L., Canaria, G., Zurich, E. T. H., Arabia, S., Ecology, M. & Nature, M. Major imprint of surface plankton on deep ocean prokaryotic structure and activity. 0–3 doi:10.1111/mec.15454.

19. Thiele, S., Fuchs, B. M., Amann, R. & Iversen, M. H. Colonization in the photic zone and subsequent changes during sinking determine bacterial community composition in marine snow. *Appl. Environ. Microbiol.* **81**, 1463–1471 (2015).

20. Mestre, M., Ruiz-González, C., Logares, R., Duarte, C. M., Gasol, J. M. & Sala, M. M. Sinking particles promote vertical connectivity in the ocean microbiome. *Proc. Natl. Acad. Sci.* **115**, E6799–E6807 (2018).

21. Agogue, H., Lamy, D., Neal, P. R., Sogin, M. L. & Herndl, G. J. Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Mol. Ecol.* **20**, 258–274 (2011).

22. Wilkins, D., Van Sebille, E., Rintoul, S. R., Lauro, F. M. & Cavicchioli, R. Advection shapes Southern Ocean microbial assemblages independent of distance and environment effects. *Nat. Commun.* **4**, 1–7 (2013).

23. Teira, E., Lebaron, P., van Aken, H. M. & Herndl, G. J. Distribution and activity of Bacteria and Archaea in the deep water masses of the North Atlantic. *Limnol. Oceanogr.* **51**, 2131–2144 (2006).

24. Biers, E. J., Sun, S. & Howard, E. C. Prokaryotic genomes and diversity in surface ocean waters: Interrogating the global ocean sampling metagenome. *Appl. Environ. Microbiol.* **75**, 2221–2229 (2009).

25. Yooseph, S., Sutton, G., Rusch, D. B., Halpern, A. L., Williamson, S. J., Remington, K., Eisen, J. A., Heidelberg, K. B., Manning, G., Li, W., Jaroszewski, L., Cieplak, P., Miller, C. S., Li, H., Mashiyama, S. T., Joachimiak, M. P., Van Belle, C., Chandonia, J. M., Soergel, D. A., Zhai, Y., Natarajan, K., Lee, S., Raphael, B. J., Bafna, V., Friedman, R., Brenner, S. E., Godzik, A., Eisenberg, D., Dixon, J. E., Taylor, S. S., Strausberg, R. L., Frazier, M., Venter, J. C. The Sorcerer II global ocean sampling expedition: Expanding the universe of protein families. *PLoS Biol.* **5**, 0432–0466 (2007).

26. 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., D. A., Zhai, Y., Natarajan, K., Lee, S., Raphael, B. J., Bafna, V., Friedman, R., Brenner, S. E., Godzik, A., Eisenberg, D., Dixon, J. E., Taylor, S. S., Strausberg, R. L., Frazier, M., Venter, J. C. Structure and function of the global ocean microbiome. *Science (80-.).* **348**, 1261359 (2015).

27. Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., De Vargas, C., Bittner, L., Zingone, A. & Bowler, C. Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E1516–E1525 (2016).

28. Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., Lima-Mendez, G., Rocha, F., Tirichine, L., Labadie, K., Kirilovsky, A., Bertrand, A.,

Engelen, S., Madoui, M. A., Méheust, R., Poulain, J., Romac, S., Richter, D. J., Yoshikawa, G., Dimier, C., Kandels-Lewis, S., Picheral, M., Searson, S., Acinas, S. G., Boss, E.,Follows, M., Gorsky, G.,Grimsley, N., Karp-Boss, L., Krzic, U., Pesant, S., Reynaud, E. G., Sardet, C., Sieracki, M., Speich, S., Stemmann, L., Velayoudon, D., Weissenbach, J., Jaillon, O., Aury, J. M., Karsenti, E., Sullivan, M. B., Sunagawa, S., Bork, P., Not, F., Hingamp, P., Raes, J., Guidi, L., Ogata, H., De Vargas, C., Ludicone, D., Bowler, C., Wincker, P. A global ocean atlas of eukaryotic genes. *Nat. Commun.* **9**, (2018).

29. Duarte, C. M. Seafaring in the 21st century: The Malaspina 2010 circumnavigation expedition. *Limnol. Oceanogr. Bull.* **24**, 11–14 (2015).

30. Mayol, E., Arrieta, J. M., Jiménez, M. A., Martínez-Asensio, A., Garcias-Bonet, N., Dachs, J., González-Gaya, B., Royer, S.-J., Benítez-Barrios, V. M. & Fraile-Nuez, E. Long-range transport of airborne microbes over the global tropical and subtropical ocean. *Nat. Commun.* **8**, 1–9 (2017).

31. Fine, R. A. Observations of CFCs and SF 6 as Ocean Tracers . *Ann. Rev. Mar. Sci.* **3**, 173–195 (2011).

32. Lechtenfeld, O. J., Kattner, G., Flerus, R., McCallister, S. L., Schmitt-Kopplin, P. & Koch, B. P. Molecular transformation and degradation of refractory dissolved organic matter in the Atlantic and Southern Ocean. *Geochim. Cosmochim. Acta* **126**, 321–337 (2014).

33. Morton, J. T., Marotz, C., Washburne, A., Silverman, J., Zaramela, L. S., Edlund, A., Zengler, K. & Knight, R. Establishing microbial composition measurement standards with reference frames. *Nat. Commun.* **10**, (2019).

34. Lin, Y., Gifford, S., Ducklow, H., Schofield, O. & Cassar, N. Towards quantitative microbiome community profiling using internal standards. *Appl. Environ. Microbiol.* AEM.02634-18 (2018) doi:10.1128/AEM.02634-18.

35. Borin, S., Crotti, E., Mapelli, F., Tamagnini, I., Corselli, C. & Daffonchio, D. DNA is preserved and maintains transforming potential after contact with brines of the deep anoxic hypersaline lakes of the Eastern Mediterranean Sea. *Saline Systems* **4**, 1–9 (2008).

36. Sebastián, M., Auguet, J. C., Restrepo-Ortiz, C. X., Sala, M. M., Marrasé, C. & Gasol, J. M. Deep ocean prokaryotic communities are remarkably malleable when facing long-term starvation. *Environ. Microbiol.* **20**, 713–723 (2018).

37. Akima, H. & Gebhardt, A. akima: Interpolation of Irregularly and Regularly Spaced Data. (2016).

38. Campbell, L. & Vaulot, D. Photosynthetic picoplankton community structure in the subtropical North Pacific Ocean near Hawaii (station ALOHA). *Deep Sea Res. Part I Oceanogr. Res. Pap.* **40**, 2043–2060 (1993).

39. Campbell, L., Nolla, H. A. & Vaulot, D. The importance of Prochlorococcus to community structure in the central North Pacific Ocean. *Limnol. Oceanogr.* **39**, 954–961 (1994).

40. Monger, B. C. & Landry, M. R. Flow cytometric analysis of marine bacteria with Hoechst 33342. *Appl. Environ. Microbiol.* **59**, 905–911 (1993).

41. Walters, W., Hyde, E. R., Berg-lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J. a & Jansson, J. K. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems* **1**, e0009-15 (2015).

42. Quince, C., Lanzen, A., Davenport, R. J. & Turnbaugh, P. J. Removing Noise From Pyrosequenced Amplicons. *BMC Bioinformatics* **12**, (2011).

43. Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W. & Huse, S. M. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA Genes. *PLoS One* **4**, 1–9 (2009).

44. Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., Cope, E. K., Da Silva, R., Diener, C., Dorrestein, P. C., Douglas, G. M., Durall, D. M., Duvallet, C., Edwardson, C. F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J. M., Gibbons, S. M., Gibson, D. L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G. A., Janssen, S., Jarmusch, A. K., Jiang, L., Kaehler, B. D., Kang, K. B., Keefe, C. R., Keim, P, Kelley, S. T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M. G.I., Lee, J., Ley, R., Liu, Y. X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B. D., McDonald, D., McIver, L. J., Melnik, A. V., Metcalf, J. L., Morgan, S. C., Morton, J. T. Naimey, A. T., Navas-Molina, J. A., Nothias, L. F., Orchanian, S. B., Pearson, T., Peoples, S. L., Petras, D., Preuss, M. L., Pruesse, E., Rasmussen, L. B., Rivers, A., Robeson, M. S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S. J., Spear, J. R., Swafford, A. D., Thompson, L. R., Torres, P. J., Trinh, P., Tripathi, A., Turnbaugh, P. J., Ul-Hasan, S., van der Hooft, Justin J.J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K. C., Williamson, C. H.D., Willis, A. D., Xu, Z. Z., Zaneveld, J. R., Zhang, Y., Zhu, Q., Knight, R.,

Caporaso, G. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).

45. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).

46. Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A. & Holmes, S. P. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581 (2016).

47. Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., Del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A. L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P., Christen, R. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**, 597–604 (2013).

48. Kim, S. ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. methods* **22**, 665 (2015).

49. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).

50. Wickham, H. The split-apply-combine strategy for data analysis. *J. Stat. Softw.* **40**, (2011).

51. Wickham, H., Francois, R., Henry, L. & Müller, K. dplyr: A grammar of data manipulation. (2017).

52. Wickham, H. & Henry, L. Tidyr: Tidy messy data. *R Packag. version 1.0* (2019).

53. Wickham, H. Simple, Consistent Wrappers for Common String Operations [Internet]. 1st.

54. Wickham, H. Reshaping data with the reshape package. (2006).

55. Wickham, H. *Ggplot2 : elegant graphics for data analysis*. (Springer, 2009).

56. Tomczak, M. & Large, D. G. Multiparameter Analysis of Mixing in the Thermocline represent th /•/ source water types of Indian Central Water . *J. Geophys. Res.* **94**, 16,141-16,149 (1989).

57. Karstensen, J. & Tomczak, M. OMP Analysis Package for MATLAB. (1999).

58. Talley, L. D., Pickard, G. L., Emery, W. J. & Swift, J. H. Typical Distributions of Water Characteristics. *Descr. Phys. Oceanogr.* 67–110 (2011) doi:10.1016/B978-0-7506-4552-2.10016-2.

59. Pierce, D. ncdf4: Interface to Unidata netCDF (Version 4 or Earlier). (2017).

60. Johnson, G. C. Quantifying Antarctic Bottom Water and North Atlantic Deep Water volumes. *J. Geophys. Res. Ocean.* **113**, 1–13 (2008).

61. Sievers, H. A. & W. D. Nowlin, J. The stratification and water masses in Drake Passage. *J. Geophys. Res.* **83**, 10489–10514 (1984).

62. Whitworth III, T., Orsi, A. H., Kim, S., Nowlin Jr, W. D. & Locarnini, R. A. Water masses and mixing near the Antarctic Slope Front. *Ocean. ice, Atmos. Interact. Antarct. Cont. margin* **75**, 1–27 (1985).

63. Orsi, A. H., Johnson, G. C. & Bullister, J. L. Circulation, mixing, and production of Antarctic Bottom Water. *Prog. Oceanogr.* **43**, 55–109 (1999).

64. Rintoul, S. R., Hughes, C. W. & Olbers, D. The antarctic circumpolar current system. in *International Geophysics* vol. 77 271–XXXVI (Elsevier, 2001).

65. Talley, L. D., Pickard, G. L., Emery, W. J. & Swift, J. H. Southern Ocean. *Descr. Phys.*

Oceanogr. 437–471 (2011) doi:10.1016/B978-0-7506-4552-2.10013-7.

66. Warner, M. J. & Weiss, R. F. Solubilities of chlorofluorocarbons 11 and 12 in water and seawater. *Deep Sea Res. Part A, Oceanogr. Res. Pap.* **32**, 1485–1497 (1985).

67. Bullister, J. L., Wisegarver, D. P. & Menzia, F. A. The solubility of sulfur hexafluoride in water and seawater. *Deep. Res. Part I Oceanogr. Res. Pap.* **49**, 175–187 (2002).

68. R Development Core, T. e. a. m. R: a language and environment for statistical computing. (2011) doi:10.1007/978-3-540-74686-7.

69. Haine, T. W. N. & Hall, T. M. A generalized transport theory: Water-mass composition and age. *J. Phys. Oceanogr.* **32**, 1932–1946 (2002).

70. Waugh, D. W., Primeau, F., DeVries, T. & Holzer, M. Recent Changes in the Ventilation of the Southern Oceans. *Science (80-.).* **339**, 568–570 (2013).

71. McNichol, A. P., Osborne, E. A., Gagnon, A. R., Fry, B. & Jones, G. A. TIC, TOC, DIC, DOC, PIC, POC—unique aspects in the preparation of oceanographic samples for 14C-AMS. *Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms* **92**, 162–165 (1994).

72. Vogel, J. S., Nelson, D. E. & Southon, J. R. 14 C background levels in an accelerator mass spectrometry system. *Radiocarbon* **29**, 323–333 (1987).

73. Key, R. M., Quay, P. D., Schlosser, P., McNichol, A., von Reden, K., Schneider, R. J., Elder, K. L., Stuiver, M. & Östlund, H. G. WOCE radiocarbon IV: Pacific ocean results; P10, P13N, P14C, P18, P19 & S4P. *Radiocarbon* **44**, 239–392 (2002).

74. Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Ramsey, C. B., Buck, C. E., Cheng, H., Edwards, R. L. & Friedrich, M. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).

75. Ngugi, D. K., Blom, J., Stepanauskas, R. & Stingl, U. Diversification and niche adaptations of Nitrospina-like bacteria in the polyextreme interfaces of Red Sea brines. *ISME J.* **10**, 1383–1399 (2016).

76. Filippidou, S., Junier, T., Wunderlin, T., Lo, C. C., Li, P. E., Chain, P. S. & Junier, P. Under-detection of endospore-forming Firmicutes in metagenomic data. *Comput. Struct. Biotechnol. J.* **13**, 299–306 (2015).

CONCLUSION

Even more so than in terrestrial environments, marine ecosystems are extensively reliant on and comprised of microbial players. ‘Omics tools provide a powerful, scalable method for assessing the structure and activity of these communities. These tools will be made more functional and cost-effective as technology advances and they are able to operate with smaller inputs and increasing automation.

This dissertation demonstrates that a surprisingly nuanced view of microbial interactions can be reconstructed from indiscriminately sequencing entire communities. Transcriptomics, in particular, allows for a comprehensive glimpse at the instantaneous reaction of cells to their environment that would not be possible with any other method. For example, Chapter 1 and 2 captured previously unreported viral infection dynamics over time and nutrient conditions in a diversity of algae. When paired with proteomics and metabolomics, as is becoming increasingly practical, an even more inclusive picture of community physiology emerges.

Although powerful, ultimately, these techniques are limited by the progress of traditional laboratory investigation. Much of what is observed is uninterpretable because of lack of cultured representatives and knowledge of gene functions or can only be speculatively inferred from distant model organisms. This is the case for both unannotated and poorly annotated ASVs in Chapter 3, and poorly annotated genes in Chapters 1 and 2.

Additionally, normalization and interpretation of large environmental datasets poses a considerable challenge. This is especially the case with transcriptional data, in which upregulation of a gene often times could represent diametrically opposed cellular reactions (e.g. upregulation of nitrate transporters in response to both a short pulse of nitrate and a long period of nitrogen starvation). In these cases, a catalogue of genetic responses in well-controlled

experimental conditions is essential. The addition of spike-in controls (as in Chapter 3) and inclusion of non-compositional metrics of the community (e.g. cell counts in Chapter 2) also helps to ground large, unwieldy, datasets and guide the conclusions that are derived from them.

This dissertation highlights the utility of ‘omics techniques for capturing the response of marine microbes to physical dynamics and resolving relationships between key community members. Genomics, transcriptomics, and meta-barcoding are already being increasingly applied to whole communities in diverse marine systems and represent a promising future for a better understanding of the microbes that invisibly support marine food webs.