UNIVERSITY OF CALIFORNIA,
IRVINE

Global Analysis of Nutrient Limitation and Microdiversity of *Prochlorococcus*

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biology

by

Lucas James Ustick

Dissertation Committee:
Professor Adam C. Martiny, Chair
Professor Jennifer Martiny
Associate Professor J.J. Emerson

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my committee chair and graduate advisor Dr. Adam Martiny. Adam has been an amazing mentor and friend and I am extremely thankful I was able to work with him during my time in graduate school.

I would like to thank my committee members, Dr. Jennifer Martiny and Dr. J.J. Emerson. Both committee members were engaged and extremely supportive of all my work.

In addition, I would like to thank Dr. Alyse Larkin. Alyse was a co-author on many of my projects and was instrumental in the inception and theory behind each.

I thank Science - AAAS for permission to include Ch.1 which was originally published in *Science Magazine*. Financial support was provided by NIH T32AI141346 and NASA 80NSSC21K1654

VITA

Lucas James Ustick

2017 B.A. in Biology, Point Loma Nazarene UniversityS

2020 M.B.A. in Biology, University of California, Irvine

2022 Ph.D. in Biology, University of California, Irvine

FIELD OF STUDY

Microbial Ecology

PUBLICATIONS

Lovindeer R, Ustick L, Primeau F, Martiny A, Mackey K. Modeling ocean color niche selection by Synechococcus blue-green acclimaters. JGR Oceans. 2021. doi:10.1029/2021JC017434

Ustick L, Larkin A, Garcia C, Garcia N, Brock M, Lee J, Wiseman N, Moore J.K., Martiny A. Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation. Science. 2021. doi:10.1126/science.abe6301

Larkin A, Garcia C, Garcia N, Brock M, Lee J, Ustick L, Barbero L, Carter B, Sonnerup R, Talley L, Tarran G, Volkov D, Martiny A. High spatial resolution global ocean metagenomes from Bio-GO-SHIP repeat hydrography transects. Scientific Data. 2021. doi:10.1038/s41597-021-00889-9

Garcia C, Hagstrom G, Larkin A, Ustick L, Levin S, Lomas M, Martiny A. Linking regional shifts in microbial genome adaptation with surface ocean biogeochemistry. Philosophical Transactions B. 2020. doi.org/10.1098/rstb.2019.0254

Martiny A, Ustick L, A. Garcia C, Lomas MW. Genomic adaptation of marine phytoplankton populations regulates phosphate uptake. Limnol Oceanogr. 2019. doi:10.1002/lno.11252

ABSTRACT OF THE DISSERTATION

Global Analysis of Nutrient Limitation and Microdiversity of *Prochlorococcus*

by

Lucas James Ustick

Doctor of Philosophy in Biology

University of California, Irvine, 2022

Professor Adam C. Martiny Irvine, Chair

Nutrient supply regulates the activity of phytoplankton, but the global biogeography of nutrient limitation and co-limitation are poorly understood. Bottle incubation experiments have revealed patterns of nutrient limitation but have limited spatial coverage, and surface nutrient concentrations are below detection limits in much of the oligotrophic ocean. In my first chapter I used genomic changes as an indicator of adaptation to nutrient stress. We collected 909 surface metagenomes from the Atlantic, Pacific, and Indian Ocean, quantified the global genome content of *Prochlorococcus* and inferred local nutrient stress based on shifts in nitrogen, phosphorus, and iron assimilation genes. Our 'omics-based description of phytoplankton resource use provided a nuanced and highly resolved quantification of nutrient stress in the global ocean.

We see a clear association between genome content and nutrient limitation, but the underlying population genetics of these genomic differences and the mechanisms by which they arise are unknown. In my second chapter, I described the functional diversity found within *Prochlorococcus* and captured the link between low nutrient

adaptation and phylogeography. I analyzed 630 surface ocean metagenomes and quantified global variation in gene abundance, phylogeny, and genomic structure through consensus marker genes and metagenomically assembled genomes (MAGs). The analysis of functional diversity and phylogeography revealed widespread flexibility in genomic content, with a phylogenetically conserved haplotype driven by P limitation.

While omics-based analysis can give a detailed representation of a microbial community, these methods are spatially and temporally limited to the time and location of sampling. Remote sensing-based descriptions of microbial nutrient limitation have been proposed but have never been validated in situ. For my third chapter, I combined remote sensing and metagenomically derived estimates of nutrient limitation to expand both the temporal and spatial extent of our characterization of nutrient limitation. Based on this in situ validation of the remote sensing products I analyzed the past 10 years of data and captured novel seasonal variation and a recent reduction in global nutrient limitation.

CH1: Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation

Authors: Lucas J. Ustick, Alyse A. Larkin, Catherine A. Garcia, Nathan S. Garcia, Melissa L. Brock, Jenna A. Lee, Nicola A. Wiseman, J. Keith Moore, and Adam C. Martiny

Introduction:

The supply of nutrients to the surface ocean exerts a fundamental control on phytoplankton growth (*1*) that may be further exacerbated by future climate-driven stratification (*2*). However, there is currently large uncertainty about the global patterns of nutrient stress and the possibility of limitation by multiple nutrients (*3*). For example, studies have independently proposed N, P, or Fe limitation for phytoplankton growing in the North Atlantic Ocean (*4*–*6*). Thus, the role and interactions of each nutrient in regulating phytoplankton growth is still unknown for large parts of the ocean.

Experimental nutrient additions and biogeochemical models are important tools for quantifying ocean nutrient stress (*7*). Nutrient additions have demonstrated Fe limitation in upwelling regions, but it has been difficult to identify the limiting nutrient in many other places. Multiple elements are often required to stimulate growth (*8*) leading to a proposal of widespread co-limitation (*7*). However, nutrients are commonly present simultaneously in low concentration, making it challenging to distinguish between co-limitation and the quick exhaustion of non-limiting nutrients (*9*). Bottle experiments can also introduce artifacts and are labor intensive, leading to large regional gaps in coverage (e.g., most of the Indian Ocean) (*7*). Ocean biogeochemical models predict large-scale patterns of nutrient limitation. However, the degree of nutrient stress and the boundaries between major

nutrient-limitation regimes are sensitive to uncertain descriptions of uptake and growth as well as external nutrient inputs (*10*). Thus, there are methodological and conceptual challenges associated with quantifying the biogeography of ocean nutrient stress.

Prochlorococcus, the most abundant phytoplankton in oligotrophic regions (*11*), can adapt to low nutrient conditions via gene gains and losses. The fast growth and large population size of *Prochlorococcus* results in a close association between genome content and local nutritional conditions (*11*). All *Prochlorococcus* genomes include the *pstABCS* genes for direct assimilation of available inorganic phosphate (*12*, *13*). However, cells gain the capacity for regulation (e.g., *phoBR*) and assimilation of specific P-containing compounds when inorganic P is depleted. They also detoxify the accidental uptake of arsenate with *arsR/acr3* (*14*). Under high P depletion and stress, cells can broadly assimilate dissolved organic P (DOP) using the alkaline phosphatases *phoA* and *phoX* (*15*, *16*). A similar phylogenomic hierarchy of adaptation is seen for nitrogen and iron acquisition and stress. *Prochlorococcus* cells progressively gain the capacity for ammonia, urea, nitrite, and nitrate uptake with increasing N stress driven by energetic costs of converting oxidized N compounds into glutamine (*17*, *18*). *Prochlorococcus* cells carry genes for increasing uptake via siderophores and additional transporters under medium Fe stress (*19*) and have lost many Fe-containing proteins under severe Fe stress in HNLC zones (*20*). Thus, genomic content of cells in a region reflects the experienced physiological nutrient stress and the biochemical trade-offs in overcoming the severity of nutrient stress by loss of function or investments in acquisition (*21*). Thus, we propose using the genome content of *Prochlorococcus* populations as a global-scale biosensor for ocean phytoplankton nutrient stress.

Table 1.1: *Prochlorococcus* genes associated with nutrient stress type and severity

$\Omega_{type,severity}$	Function	Marker Genes	References
$\Omega_{Fe,high}$	Loss of Fe containing proteins	HLIII-IV core genes	(*20*)
$\Omega_{Fe,medium}$	Fe uptake (transporters)	*cirA, expD, febB, fepB/C, tolQ, tonB*	(*19*)
$\Omega_{P,high}$	Alkaline phosphatase	*phoA*, *phoX*	(*12*, *15*)
$\Omega_{P,medium}$	P starvation regulation, arsenate toxicity, specific DOP assimilation	*arsR, acr3, chrA, gap1, mfs, phoB/E/R, ptrA, PMM707, PMM721, unkP1-5*	(*12*, *14*)
$\Omega_{N,high}$	Nitrite and nitrate assimilation and uptake	*focA, moaA-E, moeA, napA, narB, nirA*	(*17*)
$\Omega_{N,medium}$	Urea and cyanate utilization	*cynA/S, tauE, ureA-G, urtA, unkN1-2*	(*17*, *18*)

Results and Discussion:

We collected surface metagenomes from the Atlantic, Pacific, and Indian Ocean to quantify the global genome content of *Prochlorococcus* and inferred nutrient stress (Table S1.1). 909 samples were newly collected as part of Bio-GO-SHIP (*22*) and supplemented with 228 from Tara Oceans and GEOTRACES. We recruited sequences to known *Prochlorococcus* strains, recorded the frequency of established nutrient acquisition genes, and normalized to *Prochlorococcus* single copy core genes. Based on prior biochemical knowledge and verified by phylogenomics (and without reference to their spatial distribution), we *a priori* classified genetic adaptations for overcoming a nutrient stress type and severity (Ω) (Table 1.1). Although the classification of adaptations into high, medium, and low stress partially masks the complex biochemical tradeoffs and

phylogenomic trait hierarchy of nutrient use, these grouping allow us to quantify the geographic variation of nutrient stress environments in the global surface ocean.

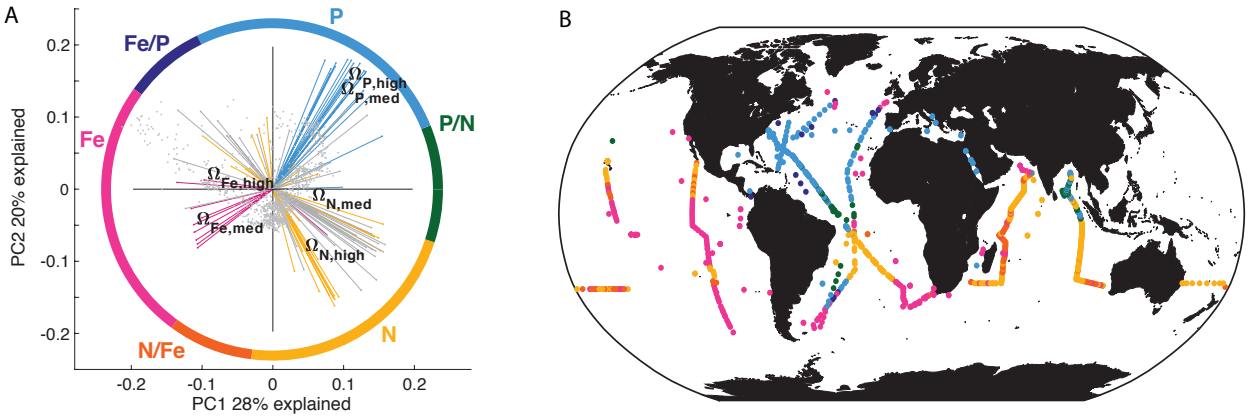

Figure 1.1: Variation in nutrient stress genes among *Prochlorococcus* populations. A: Principal component analysis of stress genes across all metagenome samples (grey dots, n = 1137). Vectors for stress genes (z_i) and composite metrics (Ω_s) are overlaid and colored according to nutrient type (red=Fe, blue=P, yellow=N, grey=Low Stress). Outer ring represents angular separation and the boundaries at which samples are categorized by nutrient stress type (Table S2). **B:** Global biogeography of nutrient stress type defined by the angular separation in Fig. 1A.

An ordination of nutrient genes demonstrated a continuum of stress type and severity (Fig. 1.1A and S1.1). The first principal component (28% variance) was parallel to the occurrence of medium and high N stress indicator genes, suggesting that the largest cluster of samples was linked to N stress. The second principal component (20% variance) separated Fe and P stress genes, where vectors for Fe and P stress genes pointed nearly opposite (0.9π angular difference). There was also a spread within the N cluster related to

N substrate (Fig. S1.2). The vectors for populations with only ammonia and urea assimilation genes had nearly the same angle but were separated by 0.33π from populations containing nitrite+nitrate or cyanate genes (Fig. S1.1). We propose that these samples were associated with medium vs. high N stress. Several low light strains of *Prochlorococcus* can use nitrite but not nitrate with a dedicated transporter *focA*. The *focA* and *narB* vectors were near opposite (separated by 0.88π) suggesting a distinct ecological niche for nitrite assimilation (cooler waters with deeper mixing, Fig. S1.1, S1.2). Samples associated with elevated medium and high P stress genes were predominantly from the North Atlantic Ocean and Mediterranean Sea, where high P stress has been proposed (*6*)(Fig. 1.1B). Samples associated with the high Fe stress genotype were mostly from the HNLC regions. However, selection for medium stress genes occurred in many samples suggesting widespread adaptation to Fe stress. We identified sample clusters between the N and Fe as well as N and P gene vectors indicating frequent co-stress. In contrast, Fe and P stress genes showed low correlation (Fig. S1.3) and there were rare co-occurrences of Fe-P stress genes. In sum, the ordination of *Prochlorococcus* genes could identify samples with genes linked to single nutrient stress or co-stress.

Prochlorococcus genome content confirmed known biogeographic patterns but also revealed several previously unrecognized regions of nutrient (co-)stress (Fig. 1.1B). We observed genotypes adapted to (i) widespread N stress in oligotrophic regions, (ii) P stress in the North Atlantic, Mediterranean, and Red Sea, and (iii) Fe stress in the equatorial Pacific. We found additional smaller regions of P stress adaptation in the western South Atlantic and North Indian Ocean. Other regions with Fe stress adaptation included the eastern South Pacific Subtropical Gyre, temperate regions in the North and South Atlantic,

and the Arabian Sea. Our data suggested that co-stress was widespread but mostly included N as one of the elements.

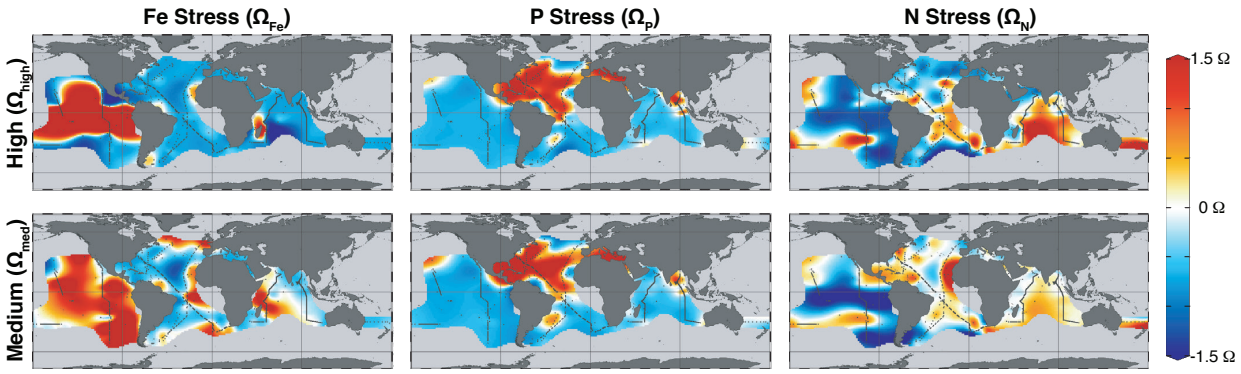

Figure 1.2: Global biogeography of *Prochlorococcus* nutrient stress type and severity. Based on their biochemical role, genes were categorized by nutrient stress 'type' and 'severity' and combined into a composite metric (Ω_s) (Table 1.1). Background coloring is based on an interpolation between sampling points shown in dark grey.

Prochlorococcus stress genes demonstrated subtle transitions between nutrient stress type and severity in the Atlantic Ocean (Fig. 1.2 and Fig. S1.4-7, S1.13). Samples from three independent cruises detected a transition between elevated Ω_{Fe} and Ω_{P} moving from north to south around 40˚-50˚N (Fig. S1.4-7). Subtle Fe stress north of the Gulf Stream has been observed in past physiological analyses of phytoplankton (*23*). Genes for DOP utilization and associated $\Omega_{P,high}$ above one were observed in the North Atlantic subtropical gyre between ~40˚N and the Intertropical Convergence Zone (ITCZ) but peaked near 30˚N (Fig. S1.4-7). An exception was a smaller region of elevated $\Omega_{Fe,med}$ in the Canary Current, where upwelling likely relieved macronutrient stress. In the central-eastern gyre core, we

detected $\Omega_{N,high} > 1$ suggesting adaptation to N-P co-limitation. Some samples from the western gyre showed an unusual combination of co-occurring Fe-P stress genes. A significant meridional shift between P stress and other nutrients at the ITCZ has been suggested (*24*). In support, $\Omega_{P,high}$ and $\Omega_{P,med}$ were substantially lower in the South Atlantic ($\Omega_{P,high}$=-0.5-0.5, $\Omega_{P,med}$=-0.5-1) and constrained to a small western gyre region. In contrast, genotypes in the eastern South Atlantic indicated adaptation to medium Fe stress ($\Omega_{Fe,med}$=1-1.5). This is consistent with recent bottle experiments (*8*) as well as a negative east-west gradient in P concentration (*25*). $\Omega_{N,high}$ was positive in the central part of the South Atlantic subtropical gyre ($\Omega_{N,high}$=0.5-1) indicating strong N stress in the region (Fig. S1.4-7). In parallel to the North, $\Omega_{Fe,med}$ rose near the subtropical front towards the Southern Ocean. Distinct nutritional regimes are thus present across the Atlantic Ocean.

The Pacific Ocean also showed clear transitions in stress type and severity. We detected a sharply bounded region in the eastern equatorial Pacific with $\Omega_{Fe,high}$ above one (Fig. S1.8-9,S1.13). There was also a surrounding zone with elevated $\Omega_{Fe,med}$ revealing a wider impact of upwelling on Fe stress than indicated by macronutrient concentrations (Fig. S1.8-9). To the north of the HNLC region, $\Omega_{Fe,med}$ was eventually replaced with elevated $\Omega_{N,high}$ ($\Omega_{N,high}$=0.5, $\Omega_{Fe,med}$=0) and some P stress genes near Station ALOHA. $\Omega_{Fe,med}$ was above one in most of the southeastern Pacific, which was consistent with Fe stress seen in bottle incubations and photophysiology studies (*26*, *27*). There was additional elevated $\Omega_{N,high}$ in the South Pacific gyre core surrounded by a wider zone with high $\Omega_{N,med}$ (Fig. S9). In the western South Pacific, we mainly detected adaptation to N stress ($\Omega_{N,high}$=0.5-1.5, $\Omega_{N,med}$=0.5-1)(Fig. S10). However, there were slight increases in Ω_{P} towards the western edge of the gyre ($\Omega_{P,high}$=0-0.5, $\Omega_{P,med}$=0-0.2)(Fig. S10). This zonal shift towards increasing P

stress was consistent with the low P concentrations in the southwestern Pacific (*25*). We lacked samples from large regions of the Pacific Ocean (including the northwest) illustrating the significant effort required to cover the entire Pacific basin.

We previously had a limited understanding of nutrient stress in the Indian Ocean (*7*, *28*), but two recent GO-SHIP cruises greatly improved metagenomics coverage. Most of the Indian Ocean had elevated $\Omega_{N,high}$ and $\Omega_{N,med}$ (Fig. S1.11-13) with the highest values seen in the Southern Indian Ocean gyre. N stress genes decreased north of the equator and lowest in the Arabian Sea upwelling region (Fig. S1.11,1.13). We detected a region with elevated $\Omega_{P,high}$ on the northeastern side of the Indian Ocean associated with several fronts from the equator to the Bay of Bengal (Fig. S1.12). There were also indications of some P stress adaptation associated with the south-flowing Leeuwin and Agulhas currents. Samples from GO-SHIP I07N (Fig. S1.11)($\Omega_{Fe,high}$=0.5, $\Omega_{Fe,med}$=1-2) and Tara Ocean (Fig. S1.13)($\Omega_{Fe,high}$=3, $\Omega_{Fe,med}$=1) both demonstrated high Ω_{Fe} in a small upwelling region near 10˚S on the western side of the basin. This zone of high Ω_{Fe} is supported by satellite and model studies (*29*). There was also widespread elevated $\Omega_{Fe,med}$ in most of South Indian Ocean gyre and in a few samples in the Arabian Sea (*30*). Overall, our metagenomics assessment greatly expanded our understanding of nutrient stress across the Indian Ocean basin.

We speculate that the global-scale biogeography of *Prochlorococcus* multinutrient adaptation is stoichiometrically linked by nitrogen fixation. *Prochlorococcus* can use nutrients at a stoichiometric ratio significantly above the vertically supplied N:P (*31*). This leads to a default state of residual phosphate and corresponding N limitation in oligotrophic regions unless additional N is supplied by diazotrophs (*32*). Reflected in the relative ordination positions in Fig. 1a, cells appear adapted to simultaneous N-Fe co-stress

if the external Fe supply and diazotroph activity are low. Moving counter clockwise in Fig. 1.1a, populations are mainly N stressed at an intermediate and then ultimately P stressed at high Fe supply and N fixation rates (*33*). The Atlantic meridional shift in nutrient stress emerges from these stoichiometric interactions (*34*). We see signs of the same connections in the zonal shifts from elevated $\Omega_{Fe,med}$ in the southeastern towards elevated $\Omega_{P,med}$ and $\Omega_{P,high}$ in the southwestern Atlantic and Pacific Oceans. Cellular resource demand results in adaptation to high Fe stress in the upwelling zone in the equatorial Pacific Ocean (*35*). As the water flows outwards, cells experience and adapt to first medium Fe/N, then high N stress and even some P stress near Station Aloha as the vertical Fe:P supply ratio increases. These meridional and zonal shifts in adaptation to nutrient stress mirror a recent synthesis of surface phosphate concentrations (*25*). An exception was the North Indian Ocean and Bay of Bengal, where we detected a region of high Ω_P without a clear connection to N-fixation (*36*). We detected samples with unique Fe-P co-stress in the western North Atlantic. As Fe and P stress are commonly opposite (Fig. 1.1A), this co-stress may be linked to the lateral advection of low P water from the central Atlantic (*37*). Despite the exceptions, our 'omics-based approach supports an emergent stoichiometric connection of oceanic multinutrient stress.

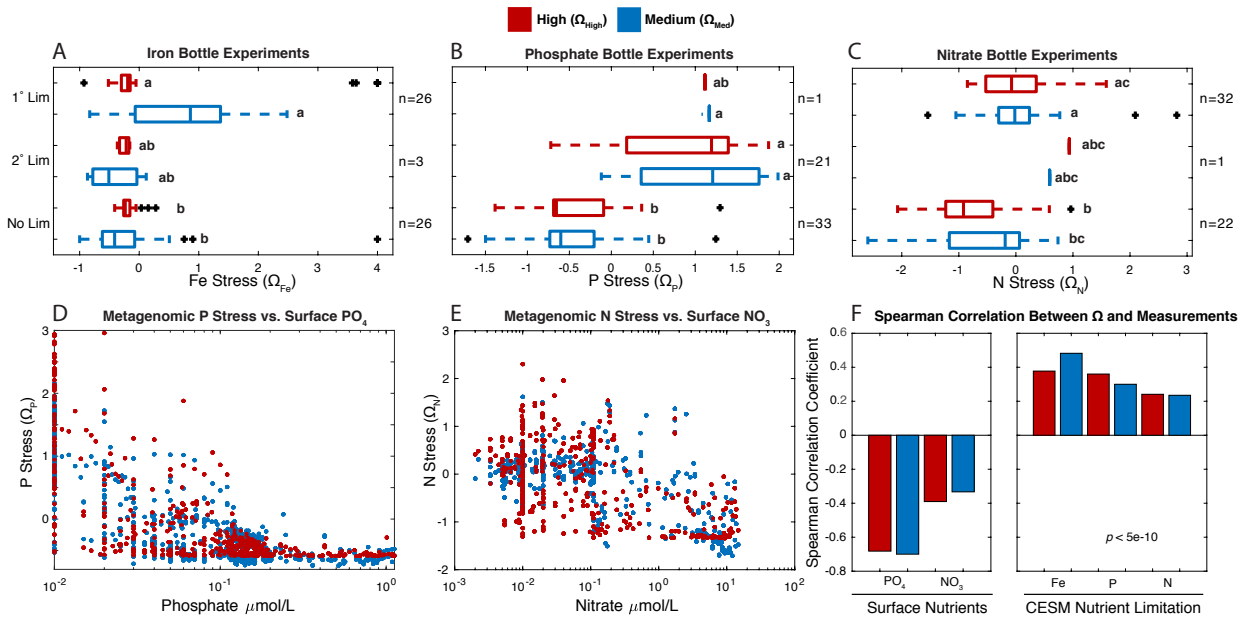

Figure 1.3: Comparison between the *Prochlorococcus* biosensor and established approaches to characterize nutrient stress. A-F: A: Comparison between Fe addition experiments and genomic Fe stress (Ω_{Fe}) (n=55). **B:** Comparison between P addition experiments and genomic P stress (Ω_P) (n=55). **C:** Comparison between N addition experiments and genomic N stress (Ω_N) (n=55). Small letters in A-C represent a Tukey post-hoc comparison. **D:** Comparison between surface phosphate concentrations and Ω_P (n= 658). **E:** Comparison between surface nitrate concentrations and Ω_N (n= 802). **F:** Summary of Spearman correlations between established approaches and genomic stress (Ω_s) all correlations are significant (p < 5e-10). Data is colored by stress level across all figures (red, high stress Ω_{High} and blue, medium stress Ω_{Med}).

We saw significant correspondences between *Prochlorococcus* stress genes vs. nutrient addition experiments, Earth system model predictions, and the depletion of surface nutrient concentrations (Fig. 1.3). There was a significant increase in the frequency

of *Prochlorococcus* stress genes from no limitation, co-limitation (2˚), and then to primary (1˚) limitation using bottle experiments (Fig. 1.3A-C) (*7*). However, there was no distinction between medium and high stress and bottle experiments.

There were also significant negative correlations between Ω_P and Ω_N and surface phosphate or nitrate concentrations, respectively (Fig. 1.3D-F). At high phosphate concentration, Ω_P was depressed but a wide range in Ω_P appeared at low phosphate concentrations (Fig. 1.3D). There was also a significant correlation between Ω_N and nitrate concentrations, but again with a considerable spread (Fig. 1.3E). This pattern suggested a large spread in stress among *Prochlorococcus* populations at low nutrient concentration and could be driven by stoichiometric interactions with other nutrients or shifts in the supply rate of nutrients to the surface ocean. We saw limited difference between Ω_{med} vs. Ω_{high} and the respective nutrient concentration. Thus, the metagenomic approach can delineate stress even when nutrients are severely depleted.

There was a broad agreement between Ω and simulated nutrient stress in an Earth system model [CESMv2_BEC (*38*)] (Fig. 1.3F and Fig S1.14). The strongest overlap in Fe stress was observed in the eastern equatorial Pacific. However, Ω_{Fe} suggested additional adaptation to Fe stress in the eastern Pacific, the southeastern Atlantic, and the south-central Indian Ocean (Fig. S1.14) - all regions where the biogeochemical model predicted N stress. This is at least partially a CESMv2_BEC bias as both satellite and other model data support severe Fe stress in the Indian Ocean region (*29*), and recent incubation experiments find some Fe stress in the southeastern Atlantic (*3*). CESMv2_BEC and Ω_P agreed on P stress in the western North Atlantic, but Ω_P suggested adaptation to P stress in the wider subtropical North Atlantic. Both efforts detected a smaller P stress region in the

southwestern Atlantic. However, CESMv2_BEC did not show any P stress in the northern Indian Ocean. There was broad agreement between CESMv2_BEC and Ω_N throughout most of the Pacific and Atlantic Oceans. However, Ω_N suggested stronger N stress adaptation in the Indian Ocean than captured in CESMv2_BEC. With the exception of the little-studied Indian Ocean, there was broad agreement in the regional patterns of nutrient stress between an Earth system model and our 'omics-based metric.

There are several important caveats to consider. First, genomic variations can reflect both the *demand* for nutrients as well as the *availability* of different nutrient species. Cyanobacteria upregulate the acquisition of oxidized N sources at a low ammonia supply and have no dedicated sensory proteins for the concentration urea or nitrate (*39*). In addition, the genome organization of N acquisition supports sequential acquisition of ammonia, urea, nitrite, and nitrate genes (Fig. S1.15)(*40*). For example, as illustrated by our eastern Pacific Ocean transect, there is strong upwelling and divergent flow of nitrate centered on 5˚S (Fig. S1.16). At the upwelling core with high nitrate, *Prochlorococcus* solely contains genes for ammonia uptake and thus likely rely on recycled ammonia, followed by urea at the gyre transition (~15˚S), and then nitrate assimilation within the gyre (~30˚S). Moreover, there is an inverse relationship between surface nitrate concentrations and nitrate reductase in *Prochlorococcus*. Thus, the genome regulation, genome organization, and biogeography of *Prochlorococcus* nutrient acquisition genes suggest a first-order relationship with cellular resource demands and environmental gradients in nutrient availability, and thus an experienced stressful supply. However, it appears unlikely that cells would retain these genes without access to the associated resources being supplied by community recycling (e.g., through nitrification). Secondly, we categorized *a priori* each

gene into three levels of stress severity based on their biochemical role, phylogenomics, and regulation to avoid 'circularity' in our quantification of nutrient stress. However, the biogeographic patterns suggest that nitrite assimilation occurs in regions with deep mixing and likely lower N stress, whereas cyanate utilization genes co-vary closely with high N stress. Furthermore, medium and high phosphate stress genes generally co-occur suggesting a limited resolution of regional P stress. Thus, some genes could be reclassified or added to the classification system (e.g., assimilation of organic N sources)(*41*) in future analyses to refine the approach. A third caveat is that our study is based on the assessment of nutrient stress adaptation in a single organism. *Prochlorococcus* is the smallest and most abundant phytoplankton in most nutrient-limited marine ecosystems between 40˚N and 40˚S (*11*). However, *Prochlorococcus* is not abundant or even present in many coastal or high-latitude environments, restricting the geographical reach of the technique. Furthermore, co-existing taxa could show divergent nutrient stress profiles due to physiological differences (*21*, *31*, *42*). However, *Prochlorococcus* generally has the highest nutrient uptake affinity and is least likely to experience cellular nutrient stress (*43*). We also found a significant correspondence between our stress metric and three established approaches implying that we capture the general community physiological state in the regions analyzed. A fourth caveat is the use of genomic changes to assess the underlying physiological state. A genome-based approach will work in populations with rapid adaptation to local conditions. Transcriptomic or proteomic approaches may work better in ecosystems with longer generation times, but these approaches are more labor intensive and affected by strong diurnal expression changes. Thus, 'omics-based assessments of

nutrient stress should be carefully calibrated to the biological behavior of the targeted ecosystem.

Connecting 'omics-based microbiome studies and biogeochemically important processes is a widespread convergence challenge, and links have been elusive and mainly correlation-based (*44*). Our stress metric builds upon 30 years of studies of *Prochlorococcus* physiology and adaptation to different nutrient regimes and allows for a mechanistic description of resource utilization. It is also a semi-quantitative, cost-effective, and standardized way of assessing nutrient stress and does not require labor-intensive incubation experiments. Finally, we can provide a sensitive description of phytoplankton stress and identify nutrient stress severity for multiple elements simultaneously (*9*). Thus, our findings demonstrate how we can harness 'omics-based information to develop a nuanced and high-resolution understanding of global biogeochemistry.

Materials and Methods

Metagenomes

We combined newly collected surface samples (< 25 m depth) from our participation in GO-SHIP, AMT28 (https://doi.org/10.1101/2020.09.06.285056) as well as previously published metagenomes from Tara Oceans (*45*), and GEOTRACES (*46*) (Table S1.1).

Bio-GO-SHIP Sequencing

For a full data descriptor of new Bio-GO-SHIP metagenomes please see Larkin and colleagues (*22*). In brief, between 2-10 L of surface water was sampled from either Niskin rosette deployments or through shipboard continuous circulating seawater systems. DNA

was collected via gentle peristaltic pump filtration (Masterflex, Cole-Parmer, Vernon Hills IL, USA) through sterilized tubing and onto 0.22 μm Sterivex filters (Millipore, Darmstadt, Germany). DNA was preserved in 1620 μl lysis buffer (23.4 mg mL^{-1} NaCl, 257 mg mL^{-1} sucrose, 50 mmol L^{-1} Tris-HCl, 20 mmol L^{-1} EDTA) and stored at -20°C. To extract DNA, Sterivex filters were first incubated in lysozyme (4.35 mg mL^{-1} final conc.) for 30 min then incubated in a Proteinase K (0.09 mg mL^{-1} final conc.) and SDS solution (0.5% final conc.) overnight at 55°C (*47*, *48*). DNA was precipitated using ice-cold isopropanol (100 %) and sodium acetate (245 mg mL^{-1}, pH 5.2), pelleted via centrifuge for 30 min at 4°C, and resuspended in TE buffer (10 mmol L^{-1} Tris-HCl, 1 mmol L^{-1} EDTA). DNA was then purified using a Clean and Concentrator kit (Zymo Research Corp., Irvine, CA, USA). Finally, DNA was quantified using a Qubit dsDNA HS Assay kit and a Qubit fluorometer (ThermoFisher, Waltham, MA, USA) and subsequently diluted to a common concentration of 2 ng/μl in Tris-HCl buffer (10 mmol L^{-1}, pH 8.0).

Metagenome libraries were prepared using a modified Tagment DNA Enzyme and Buffer Kit (Illumina, San Diego, CA, USA) (*49*, *50*). Diluted DNA (2 ng) was added to 2.5 μl tagmentation reactions (1.25 μl TD buffer, 0.25 μl TDE1) and incubated at 55°C for 10 min. Next, 8bp unique dual index (UDI) barcodes were annealed to the metagenome libraries. Tagmentation product (2.5 μl) was added to 24.5 μl PCR reactions (1.02 μM per UDI barcode, 204 μM dNTPs, 0.0204 U Phusion High Fidelity DNA polymerase and 1.02X Phusion HF Buffer [ThermoFisher, Waltham, MA, USA] final conc.). The PCR reaction proceeded as follows: 72°C for 2 min., 98°C for 30 s, followed by 13 cycles of 98°C 10 s, 63°C 30 s, 72°C 30 s, and a final extension step of 72°C for 5 min.

To optimize metagenome sequence fragment size distribution to 200-600 bp, metagenome libraries were individually bead-cleaned using a buffered solution (58.4 mg mL^{-1} NaCl, 1 mmol L^{-1} EDTA, 10 mmol L^{-1} Tris-HCl, 180 mg mL^{-1} PEG-8000, 0.055% Tween-20 final conc.) of Sera-mag SpeedBeads (ThermoFisher, Waltham, MA). Libraries were then quantified using a Qubit dsDNA HS Assay kit (ThermoFisher, Waltham, MA) and a Synergy 2 Microplate Reader (BioTek, Winooski, VT, USA). Metagenome libraries were pooled at equimolar concentration. Pooled library concentration and sequence fragment size distribution were verified using a KAPA qPCR platform (Roche, Basel, Switzerland) and a 2100 Bioanalyzer high sensitivity DNA trace (Agilent, Santa Clara, CA), respectively. Finally, pooled libraries were sequenced on either a HiSeq or NovaSeq platform (150 bp paired-end chemistry, 300 cycles; Illumina, San Diego, CA, USA).

Read Recruitment and Quality Filtering

Raw reads were quality controlled and adapter sequences were removed using Trimmomatic v0.35 (*51*). Reads were recruited using Bowtie2 v2.2.7 (*52*) against 115 reference genomes with representatives of each major ecotype of *Prochlorococcus* and *Synechococcus* as well as abundant heterotrophic bacteria (*Pelagibacter* and *Roseobacter*) to help reduce false recruitments (Table S1.3). The following flags were used with bowtie 2 --no-unal --local -D 15 -R 2 -L 15 -N 1 --gbar 1 --mp 3. SAM files were then sorted and indexed using 'samtools' v1.3 into BAM files (*53*).

Profiling Recruited Reads

Anvi'o v5 was used to profile the recruited reads (*54*). All open reading frames were aligned and clustered using NCBI BLAST (*55*) and MCL (*56*) through the Anvi'o pangenomic workflow (*57*). These clusters were curated and selected by searching for target nutrient stress genes (Table 1.1) and single copy core genes (SCCG).

Gene Selection and Categorization

We categorized genes by nutrient stress type (F, P, and N) and severity (high, medium, and low) based on past biochemical knowledge regarding gene regulation and functional trade-offs, which are verified by phylogenomics (Table 1.1, Fig. S1.15). We associated high Fe stress with the *Prochlorococcus* HLIII and HLIV ecotypes that have lost many Fe containing proteins that confer key metabolic functions. This ecotype is only detected in warm HNLC zones (*20*). Medium Fe stress was described by additional Fe uptake proteins and siderophore production seen in populations in the equatorial Atlantic Ocean (*19*). Low Fe stress was linked to commonly expressed proteins under Fe stress and found in nearly all *Prochlorococcus* strains (*58*). High P stress was characterized by the usage of broad spectrum DOP utilization using the P stress inducible alkaline phosphatases *phoA* and *phoX* (*12*, *15*). Alkaline phosphatases are highly induced under strong P stress so broad utilization of DOP is likely a very costly trait. Medium P stress was linked to genes responsible for P stress regulation (*phoBR* and *ptrA*), arsenate detoxification (*arsR, acr3*), the putative uptake of specific DOP molecules (*chrA*, *gap1* and *mfs*), and the outer membrane porin specific to P passage (*phoE*). The regulation of medium and high P stress genes are all controlled by the two-component response regulator *phoBR* and thus clearly linked to P stress (*12*). Low P stress was described by genes responsible for inorganic P

uptake across the cytoplasmic membrane (*pstABCS*) and present in all *Prochlorococcus*. We categorized N stress based on a bioenergetics trade-off for using different N sources. Nitrate and nitrite are highly oxidized and thus reduced to ammonium before incorporation into amino-groups (*17*). Thus, we linked high N stress genes to nitrite/nitrate assimilation. This included nitrate (*narB*) and nitrite reductases (*nirA*), genes for the biosynthesis of the required cofactor molybdopterin, and the nitrate and nitrite transporters (*focA* and *napA*) (*40*). These genes are repressed at elevated ammonium concentration supporting the higher energetic cost of using nitrate and nitrite (*17*). We linked medium N stress to the assimilation of urea and cyanate as these substrates match the oxidation state of ammonium and thus only require a hydrolysis step. Again, the genes are repressed under replete ammonium conditions suggesting that urea and cyanate are energetically less favorable than ammonium. Low N stress was linked to genes upregulated under N stress but present in most or all *Prochlorococcus* genomes. A few genes (unkN and unkP) used in this analysis have an unknown function but have been shown in previous studies to respond to nutrient stress conditions and co-occur with other stress gene (and thus categorized accordingly)(*12*, *40*).

Normalization Method

Gene abundances from Anvi'o were normalized using custom Matlab scripts (MATLAB, version 9.5.0.944444 (R2018b) (2018)) publicly available at github.com/ljustick/metagenome-cluster-coverage. Nutrient gene (i) coverage was normalized by single copy core gene (SCCG) coverage giving us our gene frequency (f_i) (eq. 1). We calculated to total *Prochlorococcus* SCCG coverage by first calculating the average

SCCG coverage of each ecotype clade (c) and then summing the averages together to get the total species coverage. Any samples with less than 5X total *Prochlorococcus* SCCG coverage were removed from the analysis. For samples with less than 1X *Prochlorococcus* HNLC clade SCCG coverage, f_i of all Fe high genes were set to 0.

Eq. 1

$$f_i = \frac{(nutrient\ gene\ coverage)_i}{\sum_{c=1}^{n}[AVE(clade\ SCCG\ coverage)]_c}$$

To capture the global variation of each gene and normalize abundance, we evaluated the z-score (z_i) of each individual gene across the entire data set by subtracting the mean (μ_i) from the gene frequency (f_i) and dividing by the standard deviation of the gene (σ_i) (eq. 2).

Eq. 2

$$z_i = \frac{(f_i - \mu_i)}{\sigma_i}$$

We then took the average z_i of genes grouped by nutrient type/severity (s)(Table 1.1), which we named Ω_s (eq. 3).

Eq.3

$$\Omega_s = \frac{\sum_{i=1}^{n_s}(z_i)}{n_s}$$

Analyzing Gene Distributions

To capture the global patterns of each gene and their interrelation, we performed a principal component analysis using singular value decomposition of z_i for all genes and Ω in

Matlab. Based on this principal component analysis, we characterized the angle in radians (θ) for each gene and sample within our dataset. θ was then used to broadly categorize the types of stress and co-stress (Figure 1.1 and Table S1.2). We calculated the Euclidean distance between each Ω_s using the matlab function pdist. Similarly to capture the global patterns of z, we calculated the Euclidean distance between z_i and created an average linkage hierarchical clustering between z profiles.

Comparison to Bottle Experiments

We compared our metagenomic stress metric against limiting nutrient bottle incubation experiments. We used the bottle experiments described previously (*7*) for iron (n=55), phosphate (n=55), and nitrate (n=55) incubations. In order to compare metagenomic stress, we interpolated Ω to 15 geospatial degrees from our data points using DIVA in Ocean Data View (R. Schlitzer, Ocean Data View (2016))(*59*) and matched any bottle experiments that fell within this interpolation. The data was matched by taking the average of four interpolation points around the latitude and longitude of the bottle experiments. The distributions were quantified using a one-way ANOVA in Matlab to capture significant differences in the mean.

Comparison to CESM Model and Surface Nutrient Concentrations

We compared our metagenomic stress metric against estimates of nutrient limitation by the Community Earth System Model (CESM) Biogeochemical Elemental Cycling (BEC). The version used here is modified from CESM v1.21, but includes most of the science modifications incorporated into CESM v2.0. Modifications include an explicit ligand

iron model and variable P quotas in the plankton groups and sinking export (*60*). Model output was averaged over the last 20 years of a 600-year simulation, to remove short-term variability. The ocean model was forced with a repeating 30-year cycle of National Center for Atmospheric Research and National Centers for Environmental Prediction (NCAR-NCEP) Core 2 forcings corresponding to years 1980 to 2009. The model includes three explicit phytoplankton groups: diatoms, diazotrophs, and small phytoplankton with variable P and Fe quotas. We compare the model nutrient stress for the small phytoplankton group with Ω_s values. The nutrient limitation factors are functions of ambient nutrient concentrations and group-specific half-saturation nutrient uptake affinities where the highest stress nutrient limits growth. The model includes labile and semi-labile dissolved organic matter pools for carbon, nitrogen, and phosphorus, with independent, data-constrained lifetimes (*61*). To quantify the correlation between the model predictions and Ω_s, we gridded model nutrient limitation by 1˚ and performed a Spearmen correlation test with the corresponding Ω_s values, which fell within the grid. Similarly, for any samples with surface phosphate or nitrate measurements we performed a Spearmen correlation against the corresponding Ω_s values.

Ch 2: Phylogeography of functional traits in *Prochlorococcus*

Authors: Lucas J. Ustick, Alyse A. Larkin, Adam C. Martiny

Introduction:

Microbial communities harbor vast phylogenetic diversity that is tightly linked to biogeographic partitioning of environments. High resolution microbial diversity has been revealed through advances in sequencing technologies (*62*–*64*). Fine scale phylogenetic differences, termed microdiversity (greater than 97% 16S similarity), have been shown to distinguish physiologically distinct populations (*65*–*67*). Moreover, these diverse microbial populations harbor distinct functional traits conserved across various phylogenetic depths, from broad scale differences between species to fine scale differences between bacterial strains. Microdiversity partitions important microbial traits such as antibiotic resistance, toxin production, nutrient uptake, and phage resistance (*12*, *68*–*70*). While there is a clear connection between microdiversity and genome content, specific functional groups have not been identified in many communities. Functional differences between closely related microbes can confer significant competitive advantages and niche specialization. Considering this it is important to capture both microdiversity and the corresponding functional differences between closely related populations.

In the numerically dominant and well-studied phytoplankton *Prochlorococcus*, we observe a clear phylogenetic organization of traits (*71*, *72*). *Prochlorococcus* phylogeny first diverges at the deepest taxonomic resolution by differences in adaptation to light, with a low light (LL) and high light (HL) clade (*66*). Within the highlight clade *Prochlorococcus*

partitions further based temperature with a high temperature (HLII) and low temperature ecotype (HLI) (*73*, *74*). Within these well-established ecotypes we observe phylogenetic microdiversity that follows clear spatial differentiation (*71*, *75*, *76*). This microdiversity has also been linked to differences in genome content but the specific functional differences are unknown (*77*). It has been hypothesized that differences in adaptation to nutrient limitation (specifically to P and N limitation) may be associated with microdiverse lineages, but no clear phylogenetic sub clades have been identified in *Prochlorococcus* (*78*).

Adaptations to both low phosphorus and nitrogen conditions have yet to be associated with any clearly delineated phylogenetic sub clades or haplotypes in *Prochlorococcus*. Although these adaptations vary spatially and are indicative of local nutrient conditions (*79*), the presence/absence of these genes does not follow phylogenetic structure at the ecotype levels (*12*, *40*, *78*). While specifically exploring the phylogeny of the genes *narB* (nitrate reductase) and *nirA* (nitrite reductase), the presence/absence of these genes did not sort into clear phylogenetic clades within HLII and had a sporadic distribution (*78*). It was suggested that this structure arose through vertical evolution of the traits from basal lineages and then gene loss events in HLII resulted in the sporadic distribution of *narB* and *nirA*. Others have proposed an alternate hypothesis that nutrient acquisition traits are shared through horizontal gene transfer and thus do not follow a clear phylogenetic pattern (*12*, *80*). It is thus unclear how variability in nutrient adaptation arises and if it is phylogenetically conserved at the microdiverse phylogenetic level.

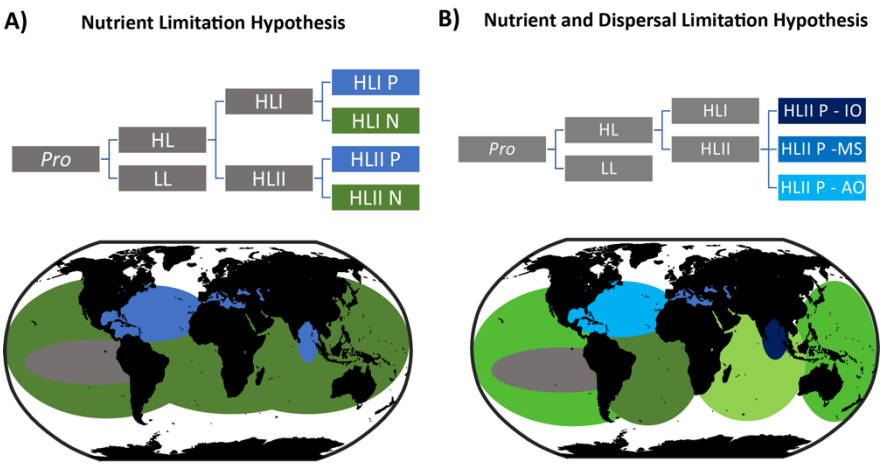

Figure 2.1: Hypothesizes about the drivers of microdiversity and the expected distributions. Nutrient limitation hypothesis A). Nutrient and dispersal limitation hypothesis B).

Here, we aim to link the global microdiversity and phylogeography of the *Prochlorococcus* HLII clade with population-specific functional diversity. Specifically, we isolated reads that mapped to the marker gene *rpoC1* to capture the phylogeography of *Prochlorococcus* and simultaneously quantified differences in the functional *Prochlorococcus* gene content from 630 surface metagenomes. We then assembled metagenomically assembled genomes (MAG) of the novel haplotypes identified. We hypothesized that functional differences in *Prochlorococcus* HLII would be primarily driven by adaptation to different (P vs. N limitation) nutrient regimes. We present here two different hypotheses about the phylogenetic structure of *Prochlorococcus* HLII microdiversity. The first hypothesis predicts microdiversity will be driven by differences in adaptation to nutrient limitation and the resulting haplotypes will be globally dispersed (Figure 2.1A). In this case, we would expect global haplotypes that are differentiated based

on nutrient conditions, i.e. N limitation vs. P limitation. The second hypothesis predicts microdiversity will be driven by differences in adaptation to nutrient limitation but will also be dispersal-limited (Figure 2.1B). In this case, we would expect regionally limited haplotypes that are differentiated based both on location and nutrient condition. For example we would expect P limited populations in the Atlantic Ocean to occupy a different phylogenetic clade than samples that are P limited in the Indian Ocean. This project constitutes the first global assessment of both *Prochlorococcus* microdiversity and functional diversity on a global scale.

Results:

We analyzed 630 surface ocean metagenomes and isolated reads that mapped to *Prochlorococcus* HLII reference genomes in order to link phylogenetic and functional diversity at the microdiverse population level. We first characterized the global functional diversity within HLII by quantifying the variation in gene abundance across the global ocean which we will refer to as genomic diversity and/or distance in this study. We then related this diversity to phylogeny by creating a consensus sequence of reads that recruited to the marker gene *rpo*C1 which we will refer to as phylogenetic diversity and/or distance. Based on the relationship of genomic and phylogenetic diversity, we then created targeted metagenomically assembled genomes to better understand the genomic structure of the new clade.

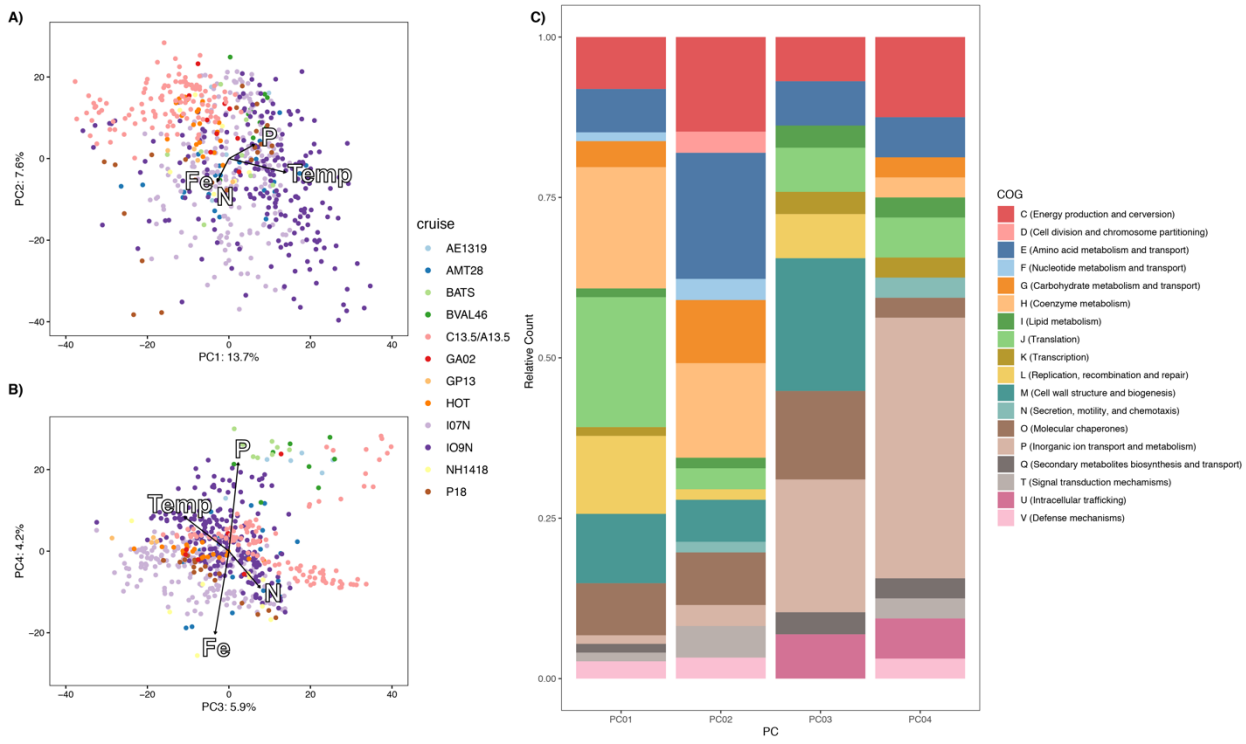

Figure 2.2: Global Genomic Diversity of HLII. Principal component analysis of gene distributions with environmental co-variates (A,B). Vector length of co-variates is based on variance explained by the environmental factor (A,B). Annotations of the 100 most informative genes for each component (C).

Variation in the flexible genome of *Prochlorococcus* HLII revealed widespread functional diversity linked to environmental processes. We performed a PCA analysis on the normalized abundance of the variable gene content of the HLII ecotype. To contextualize this dimensional reduction we fit the PCA with temperature and the abundance of nutrient acquisition genes grouped by type (Fe/P/N) (*79*). The top 100 genes that contributed to each principal component were isolated and counted based on NCBI clusters of orthologous groups annotations (COGs). PC1 captured 13% of the total variance in gene content and had the strongest correlation with temperature (Spearman rho=0.74,

p<0.05)(Figure 2.2A, S2.1A, Table S2.1). PC1 was primarily informed by genes annotated as H (coenzyme metabolism), J (translation), and L (replication recombination and repair)(Figure 2.2C). PC2 captured 8% of the total variance and was enriched in the Atlantic Ocean and depleted in the Indian Ocean Samples (Figure 2.2A, S2.1B, Table S2.1). This component was primarily informed by genes annotated as E (amino acid metabolism and transport), G (carbohydrate transport and metabolism), and H (coenzyme metabolism). PC3 captured 6% of the total variance and was primarily informed by genes annotated as M (cell wall/ membrane/ envelop biogenesis)(Figure 2.2, S2.1C, Table S2.1). PC4 captured 4% of the total variance and was primarily informed by genes annotated as P (inorganic ion transport and metabolism). PC4 also had a strong positive correlation with acquisition and metabolism genes for phosphorus (Spearman rho=0.88, p<0.05) and a negative correlated with iron genes (Spearman rho=-0.79, p<0.05) (Figure 2.2B, S2.1D, Table S2.1). Overall, functional gene content was correlated with differences in temperature and nutrient regimes, while showing widespread variation.

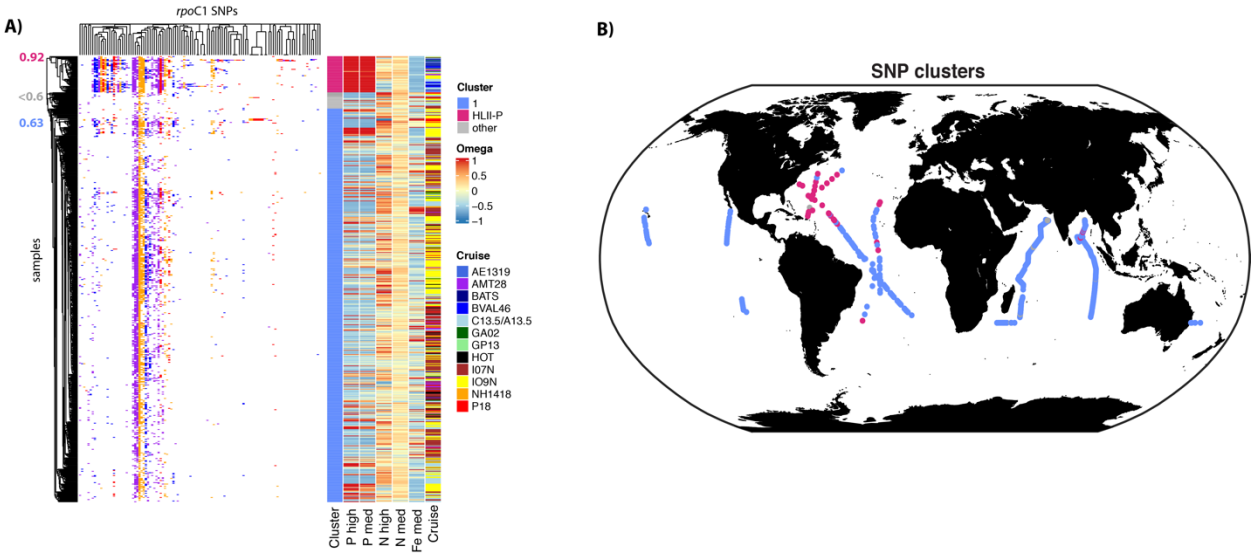

Figure 2.3: Global Phylogenetic Diversity of *Prochlorococcus* HLII. Clustering of the *rpoC1* marker gene based on sequence similarity, with corresponding cruise and metagenomically derived nutrient limitation Omega (A). Spatial distribution of groups based on the clustering of *rpoC1* consensus sequences (B).

Phylogenetic microdiversity of *Prochlorococcus* populations showed systematic biogeographic distributions. We identified two stable clusters which appeared in over 60% of our bootstraps and two unstable clusters which appeared in fewer than 60% of our bootstraps (Figure 2.3A). We identified a large cluster found across the globe, which we called cluster 1, and a second highly conserved cluster (present in 92% of bootstraps) found primarily in the North Atlantic Ocean and in the northeastern region of the Indian Ocean (Figure 2.3A, 2.3B). The highly conserved cluster was associated with significantly higher abundance of phosphorus genes than cluster 1 (Omega P high mean = 1.48) as well as significantly lower abundance of iron genes so we named the cluster HLII-P (Omega Fe med mean = -0.63)(Figure 2.3A, S2.2). A Mantel test comparing the pairwise distance between samples based genomic distance verses phylogenetic distance revealed that only 14% of genomic variation within HLII can be explained by phylogeny (R^2=.142 P<0.001). A majority of differences in the abundance of genes (86%) was not linked to phylogeography. Notably cluster HLII-P had a significantly higher functional genome PC4 values than cluster 1 (cluster 1 PC4 mean = -1.60, HLII-P PC4 mean = 20.79, $p < 0.001$), suggesting that HLII-P spatially co-occurred wherever PC4 was enriched (Figure 2.3A, 2.3B, S2.1, S2.2). This indicates that the functional genome PC4 captured the phylogenetically conserved functional type of HLII-P. The phylogeography of our samples reveals strong link between

phylogeny and phosphorus limitation, while a majority of the functional differences were not associated with microdiverse populations.

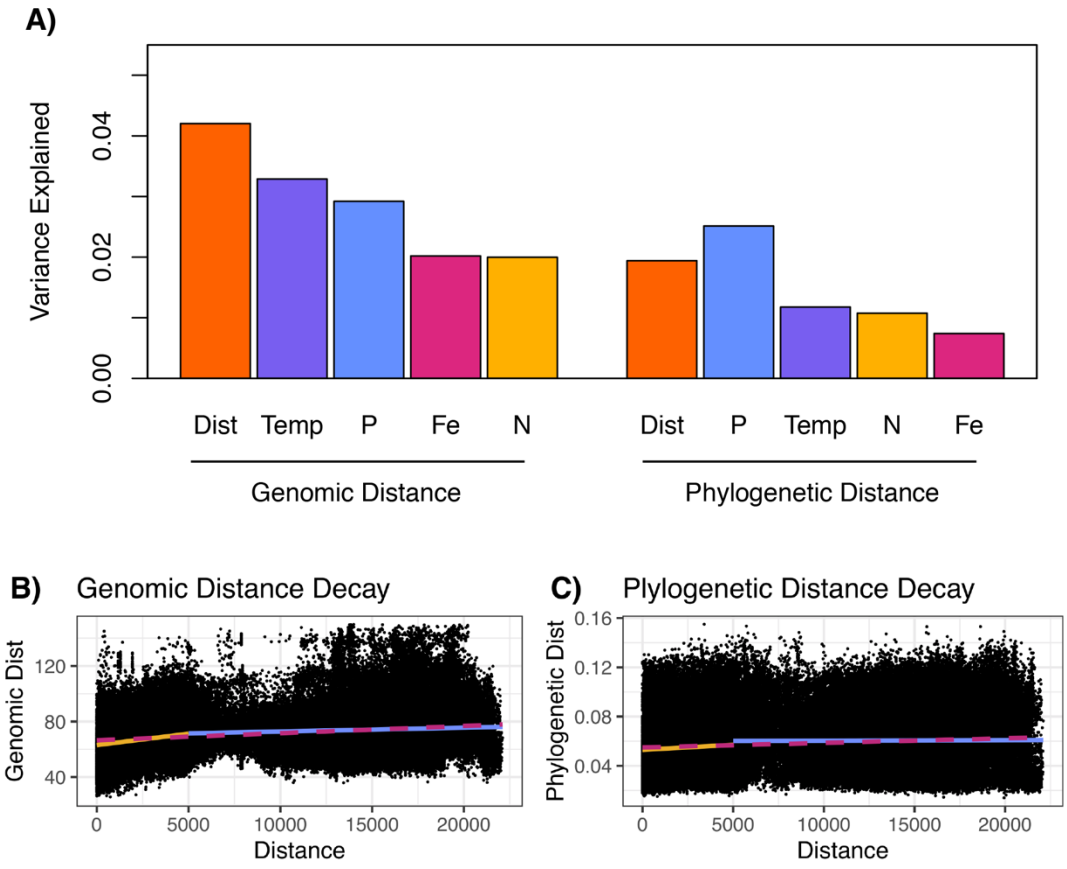

Figure 2.4: Variance explained by environmental factors and distance model. (A) Variance explained by based on PERMANOVA, variables are presented in the same order as they were used in the model. All relationships shown are significant (p-value < 0.01). (B,C) Distance decay of genomic distance (B) and phylogenetic distance (C). Overall linear fit (red and dashed), within cruise linear fit (yellow), and between cruises linear fit (blue).

Global patterns of functional and phylogenetic diversity within *Prochlorococcus* HLII were explained by different environmental factors. We calculated the variance in functional

diversity and phylogenetic diversity explained by a variety of independent variables (Figure 2.4). First, we created a model that estimates oceanographic distance between samples in order to calculate the amount of variation explained by physical distance (spatial autocorrelation). In our PERMANOVA analysis we captured the variance explained by distance to remove any distance effects from the resulting factors. 4.2% of functional diversity could be explained by spatial distance, while only 1.9% of phylogenetic diversity could be explained by spatial distance between samples (Figure 2.4). Once distance effects were removed temperature explained the most variance in genomic diversity 3.2% while phosphorus gene abundance explained the most phylogenetic diversity 2.5% (Figure 2.4). Overall we could explain more of the genomic diversity (15%) with our environmental factors than phylogenetic diversity (8%).

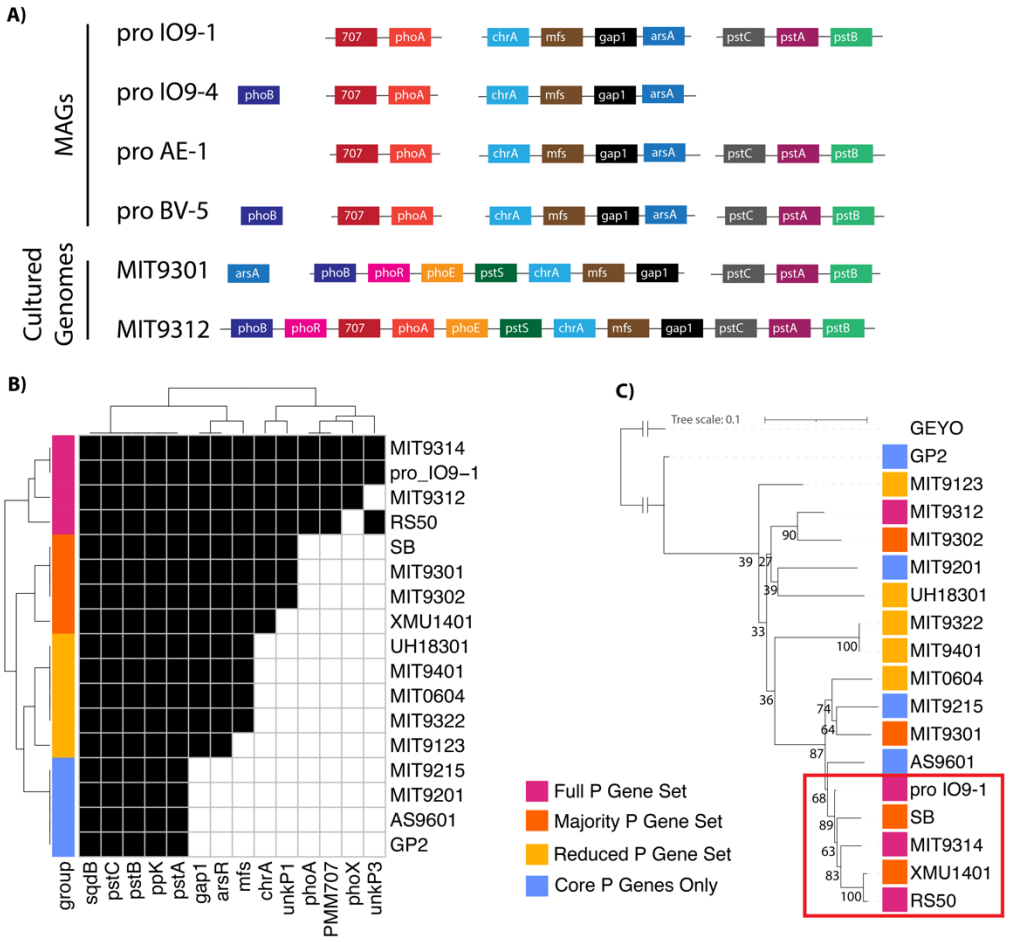

Figure 2.5: Pro HLII-P MAG annotations of P acquisition genes compared to cultured genomes (A). Clustering of genomes based on presence/absence of P acquisition genes present in the MAG IO9-1 (B). Phylogenetic tree based on the *rpoC1* gene with GEYO (*Synechococcus*) as an outlier (C). Bootstraps show at the corresponding node.

Metagenomically assembled genomes (MAG) revealed genomic conservation of the organization and presence/absence of P acquisition genes within HLII-P. Notably the *arsA* gene was found after gap1 in each MAG. This organization is not common to other cultured HLII genomes. MIT9301 contains *arsA* in another region and MIT9312 does not contain *arsA* in its genome (Figure 2.5A, 2.5B). The similar genetic structure of the phosphorus

genes in our MAGs suggests a common origin of the genes. We compared the presence/absence of the phosphorus acquisition genes found in the MAG IO9-1 with cultured *Prochorococcus* HLII genomes (Figure 2.5B). We observed a hierarchy of presence/absence of these genes and grouped the genomes based on this pattern (Figure 2.5B). We then created a phylogenetic tree based on the *rpoC1* gene and compared this to differences in gene content (Figure 2.5C). This phylogenetic analysis reveals a clade which contains genomes which all have the full or majority P acquisition gene set such as MIT9314 and RS50 (Figure 2.5C). Our analysis of our HLII-P MAGs revealed a within ecotype organization of P acquisition genes which corresponded to a clade within our cultured genomes.

Discussion:

We observed widespread variation in genomic diversity with a weak link to phylogeny. Previous studies observed a connection between microdiversity in *Prochlorococcus* and nutrient limitation so we hypothesized within ecotype functional differences would mirror this and primarily be driven by differences in nutrient conditions (*47*, *81*). In our analysis, temperature explained the most variation in genome content within the ecotype HLII while P limitation had the strongest correlation with phylogenetic differences (Figure 2.4). Our hypothesis was incorrect predicting differences in genomic diversity would be primarily explained by nutrient conditions, but phylogenetic diversity was clearly linked to nutrients. Most previous work has linked phylogenetic marker genes to environmental processes which may have overestimated the total effect of nutrient conditions on *Prochlorococcus* genomic diversity. This highlights the importance of

evaluating both genomic changes along with phylogenetic changes since functional traits may not follow phylogeny due to processes such as horizontal gene transfer and loss of traits through genomic streamlining (*11*). When comparing the amount of variance explained by spatial autocorrelation we observed a stronger distance decay relationship between gene content than SNP differences (Figure 2.4B, 2.4C). This may come to rise because differences in gene content directly come with a competitive advantage and/or cost and can instantly be selected for/against. This contrast phylogenetic differences which must by chance appear in an individual with a competitive advantage and are thus selected on a slower timescales. Microdiverse differences in phylogeny may not show spatial-auto correlation because dispersal overcomes drift in a manner similar to the Bass-Becking hypothesis, 'everything is everywhere but the environment selects' (*82*, *83*). This pattern is indicative of strong contemporary effects combining selection and rapid dispersal resulting globally conserved microdiverse haplotypes (*84*). Alternatively, our metagenomically derived phylogeny may not have the resolution to capture drift between these populations, and spatial auto-correlation is masked in noisy data.

Our analysis supports the existence of a novel HLII-P haplotype with important implications about the mechanism and evolutionary history of microdiversity within *Prochlorococcus*. The presence/absence and organization of phosphorus uptake genes does not follow phylogeny at the ecotype level (*12*). Our analysis reveals a clear organization of phosphorus acquisition genes within the HLII ecotype highlighting the importance of phylogenetic assessment across phylogenetic depths (Figure 2.3, 2.5, S2.2). This is in contrast to nitrogen acquisition genes which have not shown any clear within clade organization. For nitrogen acquisition genes it has been hypothesized that they evolved

vertically then were lost or selected for in more recent lineages causing the sporadic distribution of the trait in *Prochlorococcus* ecotypes (*78*). While there is evidence some P acquisition genes have been acquired through horizontal gene transfer our analysis suggests the full set of P acquisition genes is phylogenetically conserved in locations of extreme P limitation (*80*). This could be a result of the fact horizontal gene transfer is more likely to happen to individuals who are closely related, resulting the phylogenetic conservation of this mobile set of genes. The contrast between P acquisition and N acquisition is interesting based on the spatial patterns of nutrient limitation and the gene frequencies in natural populations. Gene abundance of N acquisition genes is variable and changes along a continuous gradient globally, while presence of P acquisition genes are more spatially limited and are often either completely absent or present in a population (*79*). The difference in distributions suggests different functional processes are acting on P acquisition vs. N acquisition genes. Novel microdiverse sub-taxa can evolve by either the acquisition of a new trait or shifting growth optima along a single trait axis (*85*). This might also explain why only 14% of genomic changes could be explained by phylogeny. The rest of the variation may be due to regional differences in N limitation, while P limited regions appear to be stable. All this together supports our nutrient limitation hypothesis which predicted microdiversity will be driven by differences in adaptation to nutrient limitation and the resulting haplotypes will be globally conserved (Figure 2.1A).

While metagenomics does allow for many comprehensive analyses there are also various caveats related to these analyses. In order to overcome sequencing error and short read length we used a mapping-based consensus method. This method cannot capture underlying diversity in non-dominant sequence types (*86*). *Prochlorococcus* has been

shown to have multiple co-existing microdiverse haplotypes in situ which vary in abundance (*47*). This makes it unclear whether our HLII-P haplotype is a single haplotype, or the phylogenetic signal is due to multiple co-existing haplotypes. There are also some caveats when interpreting MAGs. Because *Prochlorococcus* populations are often made up of multiple closely related haplotypes it is difficult to create complete assemblies that find a consistent path through the assembly graph resulting in fragmentation (*87*). While we postulate that the consistent assembly structure of P acquisition genes in our MAGs is a sign of selection, it could also be a result of divergence in unassembled regions that are not captured in our assemblies.

Connection microdiversity with specific functional groups has been a challenge in the field of microbial ecology. Here we hypothesized that differences in gene content would be primarily driven by nutrient conditions and proposed 3 hypotheses about the selective pressures that drive *Prochlorococcus* microdiversity. While differences in genome content were better explained by regional differences such as temperature, phylogenetic microdiversity was clearly linked to nutrient conditions with globally conserved haplotypes. This work is an example of how we can leverage large metagenomic datasets to capture fundamental patterns and relationships between microbes. This analysis is the first direct link between phylogenetic microdiversity and functional diversity revealing a novel P limitation defined haplotype. Quantifying the connection between microdiverse haplotypes, and traits is important to link microbial processes with larger ecosystem ecology.

Materials and Methods:

Metagenomes

We analyzed surface metagenomes (<25 m depth) from Bio-GO-SHIP(*22*), Tara Oceans(*45*), and GEOTRACES(*88*).

Read Recruitment and Quality Filtering

Raw metagenomic reads were quality filtered and adapter sequences were removed using Trimmomatic v0.35 (*51*). Metagenomic reads were recruited using Bowtie2 v2.2.7 (*52*). The reads were recruited to a reference database comprised of 115 genomes with representatives of each major ecotype of *Prochlorococcus* and as well as *Synechococcus*, *Pelagibacter* and *Roseobacter* to help reduce miss recruitment of closely related microbes (Table S2). Bowtie2 was used with the following flags --no-unal --local -D 15 -R 2 -L 15 -N 1 --gbar 1 --mp 3. Resulting SAM files were sorted and indexed with samtools v1.3 into BAM files (*53*).

Profiling Recruited Reads

Recruited reads were profiled using Anvi'o v5 (*54*). All open reading frames in the reference database were aligned and clustered using NCBI BLAST (*55*) and MCL (*56*) follwing the Anvi'o Pangenomic workflow (*57*). All gene clusters from the HLII genomes were extracted and separated into single copy core genes (SCCG) and genes in the flexible genome (non-SCCG).

Metagenomic rpoC1 Consensus Sequences

Reads which recruited to the *rpoC1* gene across all reference genomes were extracted. Then reads were separated by sample and ecotype and were aligned to a reference *rpoC1* sequence. Based on the alignment we calculated a consensus *rpoC1* sequence for the HLII ecotype for each sample in our dataset. Consensus sequences were made by aligning with Bowtie2 (*52*), then the consensus was calculated and quality controlled by samtools (*53*). Only samples that passed the following quality metrics were used in all further analysis: a minimum of 2000 reads that mapped to rpoC1, minimum of 5x SCCG coverage of HLII, and the sample must be at least 90% HLII out of all Prochlorococcus SCCGs.

Gene Analysis

Sequence coverage for All non-SCCG's were normalized by average SCCG coverage. This is done in order to normalize gene coverage. This gives roughly how many copies of a gene there are per individual in the sample. We calculated a z-score for the normalized abundance of each gene and the resulting normalized coverage was analyzed using PCA analysis in R. We then extracted the top 100 genes that contributed to the first 4 principal components and annotated the NCBI COG for each gene using from the Anvi'o profile.

Metagenomic rpoC1 Consensus Analysis

Consensus sequences were analyzed using R (*89*). Consensus sequences were transformed into binary sequences. Phylogenetic distances between samples was calculated using a binary Jaccard from the Vegan R package (*90*). The sequences were hierarchically clustered using hclust with the mcquitty method. Number of clusters was

selected by minimizing the sum of squares within each cluster. Resulting clustering was then bootstrapped 1000 times and clusters that were in less than 60% of the bootstraps were removed from further analysis.

Statistical Analysis

Spearman correlation between Principal components and environmental factors was calculated in R. Pairwise physical distance between samples was modeled in R. The distance was calculated by making a global map into a raster image with 400 rows and 800 columns (~0.45 degree squares). Raster squares that fell on land were made impassible and the shortest distance was calculated between all samples in a pairwise fashion. We then took our distance matrix and decomposed this into a single continuous component using metaMDS in R. This was done so the distance effect could be in cluded in our permanova analysis. Environmental variables were correlated to phylogenetic distance and functional distance using permanova analysis with the adonis2 package in R.

Metagenomically Assembled Genomes

Metagenomically assembled genomes (MAGs) were created using the following pipeline. The reads were quality controlled using the same method described previously and raw assemblies were made using the de novo assembler SPAdes (*91*). Samples were assembled individually using the metaSPAdes assembly pipeline. Resulting assembled contigs were binned using MetaBAT2 (*92*). MAGs were quality controlled and rapidly assessed using checkM (*93*). Mag annotations were made by aligning the resulting bins against a reference of phosphorus acquisition genes using BLAST

MAGs Phylogenetic Analysis

A phylogeny of the MAGs was created using the *rpoC1* gene. Sequences of the rpoC1 gene were extracted from all *Prochlorococcus* reference genomes and a single *Synechococcus* genome as an outgroup. The sequences were aligned using Mega7 (*94*) and Muscle (*95*). Model was selected based on the maximum likelihood fit of 24 different nucleotide substitution models using MEGA7. GTR + G + I was selected since it had the lowest Bayesian Information Criterion and Akaike Information Criterion values. The phylogenetic tree was created using raxml (*96*) with the following arguments raxmlHPC -f a -x 318420 -p 318420 -N 100 -m GTRGAMMAX -O -o GEYO-Syn_CRD1_53540 -n out_file -s align_file -w out_dir. The resulting tree was visualized using iTOL (*97*).

Ch3: Integration of Genomic and Remote Sensing Observation to Determine Global Ocean Nutrient Stress

Authors: Lucas J. Ustick, Toby K. Westberry, Michael J. Behrenfeld, Adam C. Martiny

Introduction:

In the surface ocean, primary production is typically limited by supply of nutrients. Future climate conditions are predicted to increase thermally-driven stratification, which results in intensified nutrient stress (*98*, *99*). Models have predicted that rising surface ocean temperature will have significant effects on primary production and microbial community composition (*100*). This proposed effect, is one of the main threats to marine ecosystem productivity and biodiversity. Nitrogen, phosphorus, and iron are the primary limiting nutrients in the global ocean, but the spatial and temporal patterns of macro nutrient limitation are poorly understood.

There are a variety of methods used to estimate nutrient stress in the ocean such as bottle incubation experiments and Earth System Models, but each has its own associated caveats and limitations. Nutrient addition experiments have been used to capture nutrient limitation in equatorial upwelling regions where there is chronic Fe limitation but it has been less effective at identifying limiting nutrients in most other regions (*101*). Typically multiple nutrients are required to increase growth resulting in proposed co-limitation in much of the ocean (*3*, *7*) Nutrient addition experiments can also introduce bottle effects, disrupt natural communities, and are labor intensive resulting in a majority of the ocean being uncharacterized both spatially and temporally. Ocean biogeochemical models have a

more complete range of spatial and temporal predictions but rely on uncertain descriptions of nutrient uptake, growth, and nutrient flux estimations (*10*). Considering this, there are methodological challenges when trying to globally quantify the spatial and temporal patterns of ocean nutrient limitation.

A convergence between omics and remote sensing, offers a potential new tool for detecting nutrient stress globally. Recently, metagenomics provided an in situ quantification of nutrient stress based on the highly abundant cyanobacteria *Prochlorococcus* (*79*). This omics'-based assessment used the presence or absence of nutrient uptake and metabolism genes to identify both type and severity of nutrient stress for the first time on a global scale. Phytoplankton regulate their chl:C ratio in response to a combination of light intensity and nutrient stress (*102*). The variance in the chl:C ratio is a result of phytoplankton optimizing investment in light acquisition machinery to match nutrient limited growth (*103*). This relationship has been proposed as an estimate of nutrient stress in the ocean using remote sensing products but has never been validated in situ (*104*–*106*). Until recently, creating a link between these remote sensing products and in situ measurements was not possible, but because of the abundance of metagenomic data, we now have the spatial coverage to create a formal validation.

Here, we combined in situ genomic biomarkers and remote sensing to detect nutrient stress across the last 20 years. We integrated omics with remote sensing to globally characterize nutrient stress. We asked: What are the global contemporary trends in ocean nutrient stress? We hypothesized that remote sensing derived nutrient stress (chl:C total /chl:C photo acclimation) will correlated with N and P limitation in our

metagenomically derived index. We also hypothesized that temporal patterns of nutrient stress will be driven by differences in temperature.

Results:

We analyzed remote sensing measurements of chl:C normalized by a chl:C photo-acclimation model (Sigma Σ) from 2003-2020 and linked the patterns to metagenomically derived nutrient stress metrics (Omega Ω). We first compared Omega nutrient stress with the novel remote sensing derived Sigma using a generalized additive model (GAM). After this in situ validation of Sigma nutrient stress, we observed the global patterns of Sigma averaged across 2003-2020 and compared to potential environmental drivers. We then quantified the novel seasonal, annual, and overall trends of Sigma nutrient stress.

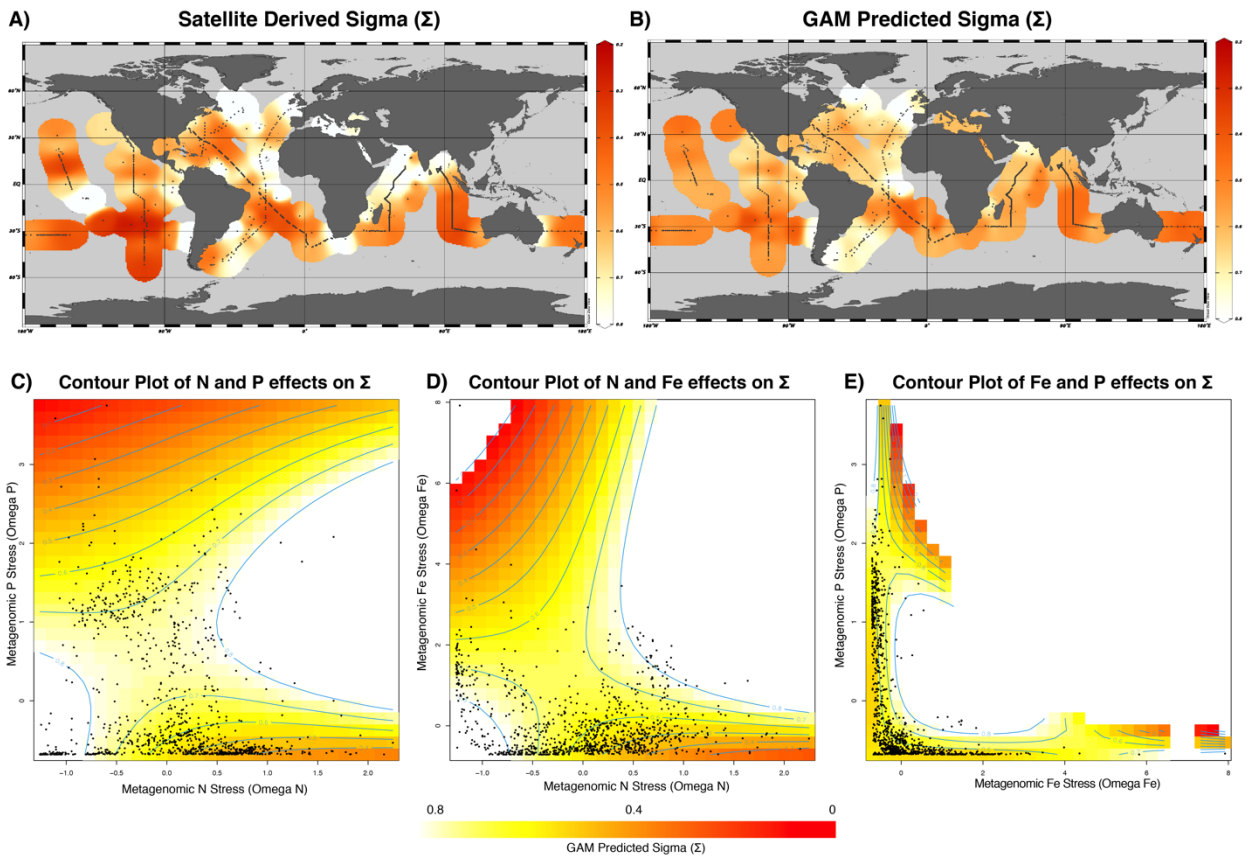

Figure 3.1: In situ validation of satellite derived nutrient stress (Sigma). Matchup of satellite derived nutrient stress values for the time and location of metagenomic sampling (A). GAM model prediction of Sigma based on metagenomic (Omega) predictors captures 29% of total variance (B). Contour plot of relationship between predictors (x and y axis)(Omega N,P,Fe) and Sigma GAM predictions (color intensity)(C-E). Black points represent sample points in predictor space.

Satellite derived nutrient stress captured the general patterns of metagenomically derived nutrient stress. To quantify this relationship, we created a GAM to predict satellite derived nutrient stress (Sigma) based on three metagenomic indices of nutrient limitation (Omega N, P, Fe). Our GAM modeled the co-variation of our 3 predictors using a tensor

product, smooth. We allowed only 3 dimensions for the bases used to represent the smooth term to maximally constrain the model. The GAM model was able to capture 29% of the variation in Sigma (Figure 3.1A, 3.1B). When looking at the effects of each predictor on the model, we see that single nutrient metagenomic limitation predicts lower values of Sigma than co-stress (Figure 3.1C-E). Sigma correlates to all 3 types of Omega stress included and is indicative of general macro-nutrient limitation regardless of type. In the model, Omega N explained the most variation, followed by Omega Fe, and Omega P, which explained the least amount of variation. In the Pacific Ocean, we see the change in Sigma closely follows Omega Fe in the equatorial upwelling and peaks at 27 degrees south which corresponds to the switch to Omega N stress (Figure S3.1). We see a strong correspondence between Omega N and Sigma in the Atlantic Ocean (Figure S3.2, S3.3). In both C13 and AMT, the highest stress is at ~ 20 degrees south (Sigma = 0.3-0.4) which corresponds to the highest nitrogen stress in the region (Omega N = 0.5-0.75). In the Southeastern Indian Ocean (IO9 stations 1-150), we see a tight correspondence between Sigma and nitrogen stress, but we observe decoupling in the Bay of Bengal (IO9 stations 160-180)(Figure S3.4). The model captured the general patterns of Sigma but was unable to capture the dynamics in regions such as the Arabian Sea, and the Bay of Bengal in the Indian Ocean (Figure S3.4, S3.5). These regions have dynamic eddies, which result in a highly variable environment, may cause the decoupling. Overall, our in-situ validation of satellite derived nutrient stress revealed a close association between Omega stress and Sigma stress.

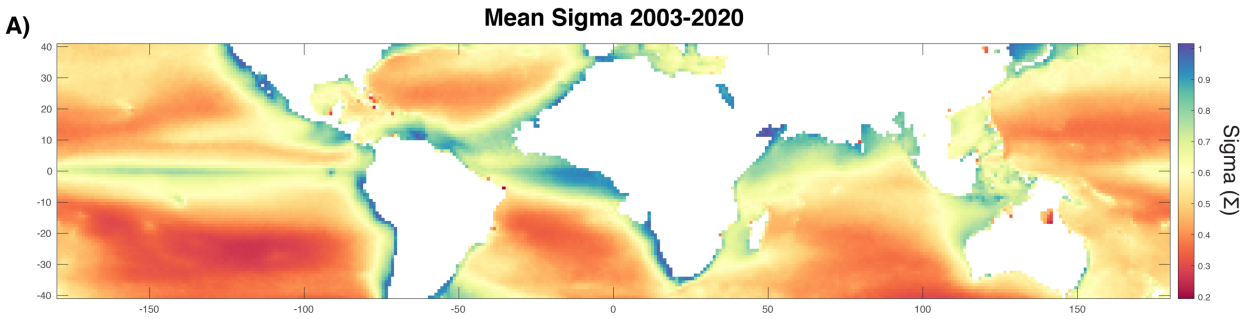

Figure 3.2: Global mean remote sensing derived nutrient stress Sigma. High values of Sigma represent communities that are not stressed (blue) with low values of Sigma representing highly nutrient stressed populations (red).

Global distributions of satellite derived nutrient stress, Sigma, displayed systematic trends across the oligotrophic ocean. In our data, the subtropical gyres, coastal, and equatorial upwellings regions showed little nutrient stress based on the satellite model (Σ = 0.7-1). Outside of these regions much of the low latitude ocean is severely nutrient stressed (Σ = 0.2-0.4). When comparing the global trends, we see the southern hemisphere shows greater levels of nutrient stress when compared to the northern hemisphere.

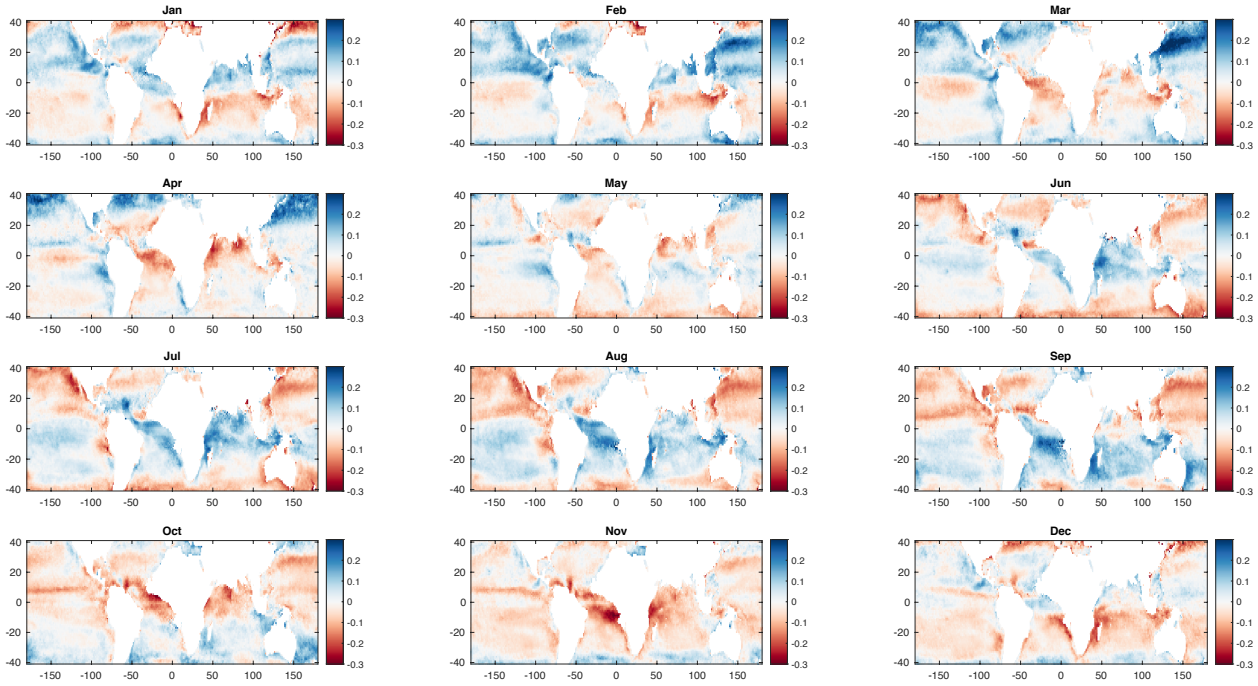

Figure 3.3: Monthly variation in remote sensing derived nutrient stress Sigma. Blue values represent a reduction in stress, with red representing an increase.

There were clear seasonal shifts in nutrient stress across the global ocean (Figure 3.3). We averaged monthly changes across the timeseries to capture seasonal trends and observed a clear seasonal signal with stress increasing during the summer and decreasing in the winter. This seasonal change had a negative correlation with changes in temperature and explained 8% of variation (Pearson r=-0.29 P-val < 0.05)(Figure S3.6). We did not observe any clear trends outside of seasonal variability between the north and south hemisphere.

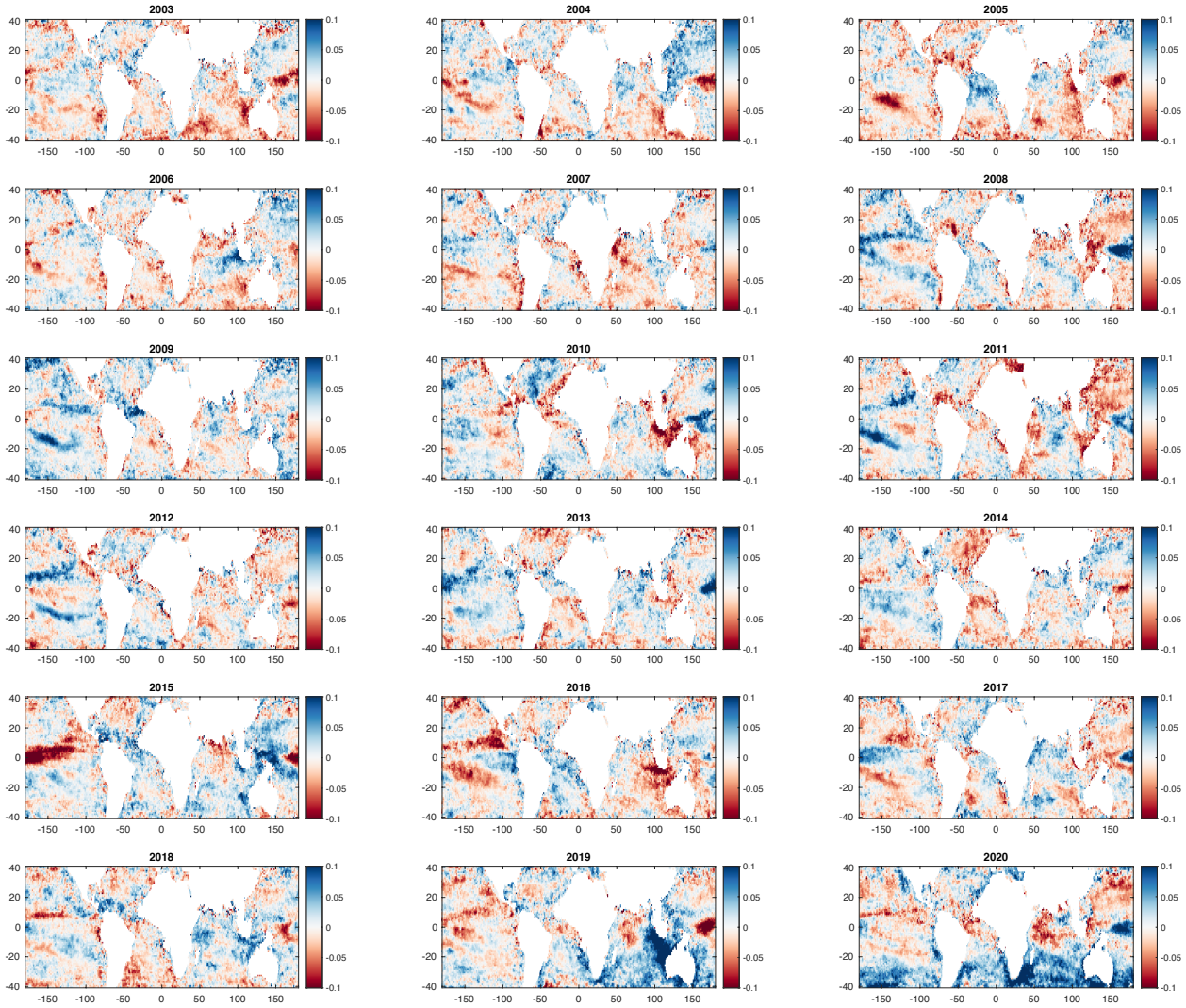

Figure 3.4: Annual variation in remote sensing derived nutrient stress Sigma. Blue values represent a reduction in stress, with red representing an increase.

Long term and annual trends in nutrient stress show systemic changes and a decoupling between nutrient stress and temperature. When quantifying the long-term trend, we see an increase in nutrient stress in the northern hemisphere with a decrease in the southern hemisphere, primarily driven by a strong reduction in stress in 2019 and 2020 (Figure 3.4, 3.5A, 3.5B). Annual trends show a strong response to the El Niño

Southern Oscillation (ENSO) which accounts for ~40% of total annual variation. The second largest annual shift was the reduction in nutrient stress observed in the southern hemisphere in the final two years (2019-2020), which was not associated with any clear oceanographic trends or environmental measurements (SST, nitrate concentration, and mixed layer depth). While we see a very slight net decrease in nutrient stress in the global ocean from 2003 to 2020 (Figure 3.5C), we observe a significant increase in sea surface temperature (Figure 3.5D). We do not detect an anthropogenic trend in nutrient stress in the ocean from 2003 to 2020 and only observed trends linked to natural oscillations such as ENSO and the Pacific Decadal Oscillation.

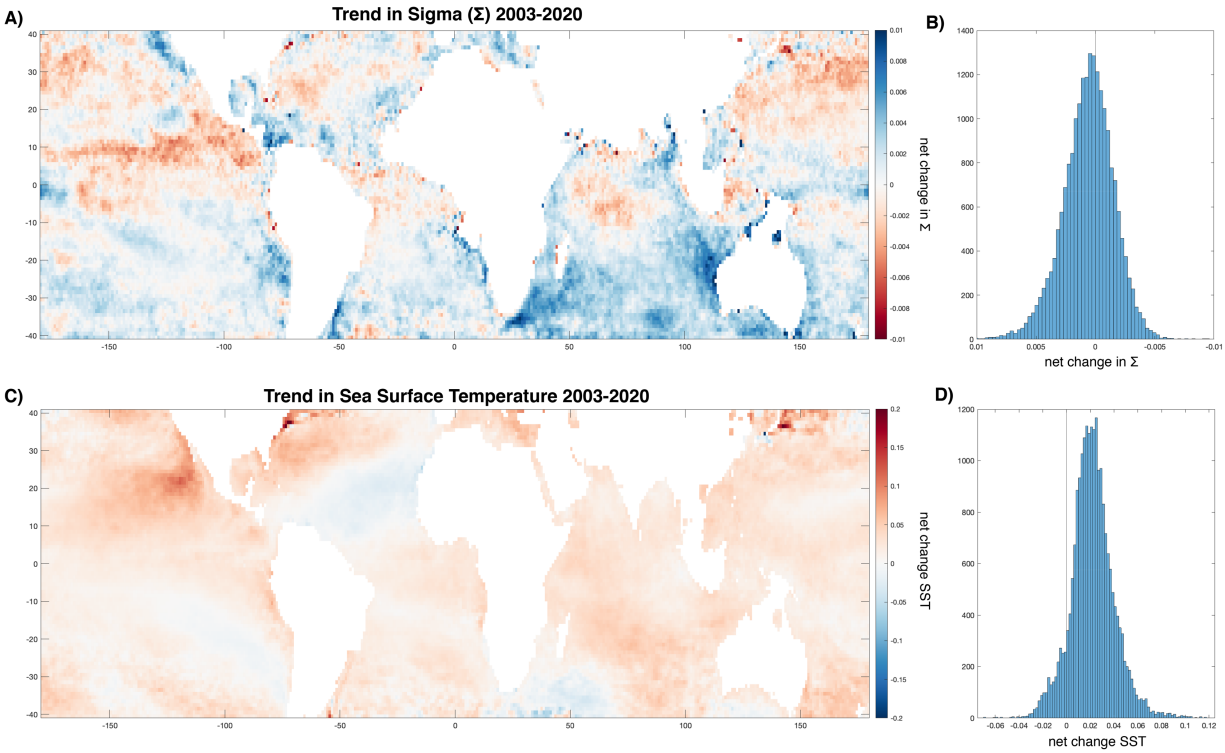

Figure 3.5: Long term trend in remote sensing derived nutrient stress Σ and temperature. Total trend in remote sensing derived nutrient stress Sigma from 2003-

2020 (A). Histogram of Sigma trend with zero change shown by the vertical line (B). Total trend in remote sensing derived SST from 2003-2020 (C). Histogram of SST trend with zero change shown by the vertical line (D).

Discussion:

We captured a clear relationship between remote sensing derived nutrient stress (Sigma) and metagenomically derived nutrient stress (Omega). We hypothesized that P and N metagenomic stress metrics together would reflect the patterns in Sigma, but our GAM model revealed the remote sensing index reflected N, P, and Fe stress (Figure 3.1). This relationship suggests Sigma captures general macro-nutrient limitation regardless of type. It was unexpected that Fe limitation would contribute to this signal. Fe limitation has its own uniquely identifiable signal in remote sensing data based on estimates of chlorophyll fluorescence quantum yield (*29*, *107*). In our model, N captured the most variation, followed by Fe, and P, which had the weakest relationship with Sigma. This may be due to the fact that phytoplankton have more flexibility in their P quotas and are able to adapt to low P conditions resulting in less of a growth penalty (*108*). The metagenomic and remote sensing stress metrics diverged in highly variable regions such as the north Indian Ocean (Figure S3.4, S3.5). The north Indian Ocean undergoes seasonal monsoons (*109*, *110*) and in the Bay of Bengal (*111*) and the Arabian Sea (*112*), eddies cause dynamic nutrient conditions. The decoupling of metagenomics and remote sensing in these regions may be due to the timescales each metric responds to. The metagenomic index Omega measures adaptation to nutrient conditions, while the remote sensing index Sigma is a result of

acclimation. This suggest the remote sensing index is more indicative of current conditions while the metagenomics are more tuned to long term conditions i.e., a seasonal average.

We hypothesized the temporal patterns in satellite derived nutrient stress Sigma would be driven by temperature. While we captured a strong seasonal relationship between temperature and Sigma, temperature did not correlate well with the overall variation in nutrient stress from 2003-2020 (Figure 3.3, 3.5). This seasonal connection could be due to other seasonal dynamics not measurable by satellites. We have seen a decline in total phytoplankton biomass over time with a strong connection to temperature in previous studies (*98*). Models have also proposed that temperature driven stratification would decrease surface nutrient concentrations, which would lead to increased nutrient limitation in the surface ocean (*99*). We observed a clear increase in temperature across our data but saw a slight decrease in nutrient limitation in the global ocean (Figure 3.5). All this together suggests phytoplankton may be robust to the effects of a warming ocean and can adapt to reducing the growth penalty of these low nutrient conditions. Alternatively there may be factors we did not measure that are having a positive effect on phytoplankton growth such as $C0_2$ fertilization. Longitudinal studies looking at microbial community composition and adaptation are needed to better understand the underlying processes resulting in this decoupling.

Understanding the effects of a warming climate on primary production is a goal across scientific fields. Nutrient stress exerts a fundamental control on phytoplankton growth and is important for informing long term predictions in global biochemical models. Our analysis suggests that anthropogenic climate change is not contributing to temporal patterns of nutrient stress, but the patterns are driven by natural, seasonal trends. This

work highlights the utility of connecting omics-based studies with remote sensing. Both methods are cost effective for capturing a wide range of data types across large spatial and temporal scales. Convergence between fields and datatypes is paramount for creating a more complete picture of the role of microbes in past, present, and future global biogeochemistry. This work suggest more in situ longitudinal studies of microbial composition and processes are needed, to fully understand how phytoplankton are adapting to future conditions.

Materials and Methods:

Metagenomic Data Curation

Metagenomic indices of nutrient limitation Omega were curated from (*79*). We specifically used Omega N high, Omega P high, and Omega Fe med. Omega Fe med was used because of the spatial limitations of Omega Fe high.

Sigma Calculation

All remote sensing data was collected from ocean color generated using R2018 MODIS-Aqua products available from the OB.DAAC. Specifically, the products we used were derived from standard Level 3 (~9.25km), 8-Day (8D) composites, which represent a reasonable compromise between coverage and temporal resolution. Sigma was calculated by first determining the chlorophyl to carbon ratio, then the ratio is normalized based the photoacclimation model from (*104–106*). Any values of Sigma above 1.2 were removed from the analysis.

GAM Modeling

We isolated datapoints that corresponded to the time and location of metagenomic sampling. We calculated the total variance explained by our model across different spatial and temporal averaging. We ended up averaging the data spatially in an 18x18 grid spatially (~1.494 degrees) and temporally across a 24 day average. Generalized Additive Modeling was done using the MGCV package in R. We used Omega N high, Omega P high, and Omega Fe med as predictors for Sigma. We calculated the tensor product interactions and allowed 3 dimensions for the bases used to represent the smooth term. This is the minimum amount you can use when calculating a 3 dimensional tensor product interaction. This was done in order to constrain the model as much as possible. Tensor product smooths were used since each predictor is made up of relative values and are not on the same scale. We applied a gaussian family to the model and used the mixed model approach via restricted maximum likelihood (REML). Contour plots were generated using the vis.gam function in MGCV.

Sigma Analysis

We analyzed the patters and trends of Sigma across 40S-40N to match the distribution of the metagenomic data. The analysis of trends was done in matlab using the Climate Data Toolbox doi:10.1029/2019GC008392. Trends were calculated across different time frames (Monthly, Yearly, and Total Time Series). Correlations between trends were calculated in matlab using the Pearson index. Total variance was decomposed using EOF analysis with and without seasonal detrending.

Conclusion:

Nutrient limitation in the global ocean is one of the main factors controlling primary production. Based on a novel metagenomic dataset I first characterized global patters of nutrient stress in type (N, P, and/or Fe) and severity leveraging *Prochlorococcus* as a biosensor. Second, I captured the underlying phylogeny of *Prochlorococcus* nutrient stress traits and proposed a novel HLII-P haplotype. Finally, I connected our omics derived nutrient stress with a novel remote sensing model to quantify the temporal and seasonal shifts in nutrient stress. This work represents interdisciplinary projects that leverage data and theory across fields. Historically convergence between omics', biogeochemistry, and remote sensing have been elusive. Here I leveraged large new metagenomic datasets and mechanistically linked gene distributions and phylogenetic patterns to biogeochemically relevant processes. My work builds upon decades of work on *Prochlorococcus* physiology, and microbial modeling to provide novel insights into the patterns and processes of ocean nutrient limitation.

REFERENCES

1. T. Tyrrell, The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature*. **400**, 525–531 (1999).

2. L. Bopp, L. Resplandy, J. C. Orr, S. C. Doney, J. P. Dunne, M. Gehlen, P. Halloran, C. Heinze, T. Ilyina, R. Séférian, J. Tjiputra, M. Vichi, Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models. *Biogeosciences*. **10**, 6225–6245 (2013).

3. T. J. Browning, E. P. Achterberg, I. Rapp, A. Engel, E. M. Bertrand, A. Tagliabue, C. M. Moore, Nutrient co-limitation at the boundary of an oceanic gyre. *Nature*. **551**, 242–246 (2017).

4. J. H. Martin, S. E. Fitzwater, R. Michael Gordon, C. N. Hunter, S. J. Tanner, Iron, primary production and carbon-nitrogen flux studies during the JGOFS North Atlantic bloom experiment. *Deep. Res. Part II*. **40**, 115–134 (1993).

5. M. Davey, G. A. Tarran, M. M. Mills, C. Ridame, R. J. Geider, J. LaRoche, Nutrient limitation of picophytoplankton photosynthesis and growth in the tropical North Atlantic. *Limnol. Oceanogr.* **53**, 1722–1733 (2008).

6. J. W. Ammerman, R. R. Hood, D. A. Case, J. B. Cotner, Phosphorus deficiency in the Atlantic: An emerging paradigm in oceanography. *Eos (Washington. DC).* **84**, 165–170 (2003).

7. C. M. Moore, M. M. Mills, K. R. Arrigo, I. Berman-Frank, L. Bopp, P. W. Boyd, E. D. Galbraith, R. J. Geider, C. Guieu, S. L. Jaccard, T. D. Jickells, J. La Roche, T. M. Lenton, N.

M. Mahowald, E. Marañón, I. Marinov, J. K. Moore, T. Nakatsuka, A. Oschlies, M. A. Saito, T. F. Thingstad, A. Tsuda, O. Ulloa, Processes and patterns of oceanic nutrient limitation. *Nat. Geosci.* **6**, 701–710 (2013).

8. T. J. Browning, E. P. Achterberg, J. C. Yong, I. Rapp, C. Utermann, A. Engel, C. M. Moore, Iron limitation of microbial phosphorus acquisition in the tropical North Atlantic. *Nat. Commun.* **8**, 15465 (2017).

9. E. A. Davidson, R. W. Howarth, Nutrient in synergy. *Nature*. **449**, 1000–1001 (2007).

10. A. Krishnamurthy, J. K. Moore, N. Mahowald, C. Luo, C. S. Zender, Impacts of atmospheric nutrient inputs on marine biogeochemistry. *J. Geophys. Res.* **115** (2010), doi:10.1029/2009jg001115.

11. F. Partensky, L. Garczarek, Prochlorococcus : Advantages and limits of minimalism. *Ann. Rev. Mar. Sci.* **2**, 305–331 (2010).

12. A. C. Martiny, M. L. Coleman, S. W. Chisholm, Phosphate acquisition genes in Prochlorococcus ecotypes: Evidence for genome-wide adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12552–12557 (2006).

13. D. J. Scanlan, N. J. West, Molecular ecology of the marine cyanobacterial genera Prochlorococcus and Synechococcus. *FEMS Microbiol. Ecol.* **40**, 1–12 (2006).

14. J. K. Saunders, G. Rocap, Genomic potential for arsenic efflux and methylation varies among global Prochlorococcus populations. *ISME J.* **10**, 197–209 (2016).

15. S. Kathuria, A. C. Martiny, Prevalence of a calcium-based alkaline phosphatase associated with the marine cyanobacterium Prochlorococcus and other ocean

bacteria. *Environ. Microbiol.* **13**, 74–83 (2011).

16. L. R. Moore, M. Ostrowski, D. J. Scanlan, K. Feren, T. Sweetsir, Ecotypic variation in phosphorus-acquisition mechanisms within marine picocyanobacteria. *Aquat. Microb. Ecol.* **39**, 257–269 (2005).

17. A. Herrero, A. M. Muro-Pastor, E. Flores, Nitrogen control in cyanobacteria. *J. Bacteriol.* **183**, 411–425 (2001).

18. A. C. Tolonen, J. Aach, D. Lindell, Z. I. Johnson, T. Rector, R. Steen, G. M. Church, S. W. Chisholm, Global gene expression of Prochlorococcus ecotypes in response to changes in nitrogen availability. *Mol. Syst. Biol.* **2**, 53 (2006).

19. R. R. Malmstrom, S. Rodrigue, K. H. Huang, L. Kelly, S. E. Kern, A. Thompson, S. Roggensack, P. M. Berube, M. R. Henn, S. W. Chisholm, Ecology of uncultured Prochlorococcus clades revealed through single-cell genomics and biogeographic analysis. *ISME J.* **7**, 184–198 (2013).

20. D. B. Rusch, A. C. Martiny, C. L. Dupont, A. L. Halpern, J. C. Venter, Characterization of Prochlorococcus clades from iron-depleted oceanic regions. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16184–16189 (2010).

21. C. A. Garcia, G. I. Hagstrom, A. A. Larkin, L. J. Ustick, S. A. Levin, M. W. Lomas, A. C. Martiny, Linking regional shifts in microbial genome adaptation with surface ocean biogeochemistry. *Philos. Trans. R. Soc. B Biol. Sci.* **375** (2020), doi:10.1098/rstb.2019.0254.

22. A. A. Larkin, C. A. Garcia, M. L. Brock, J. A. Lee, N. Garcia, L. J. Ustick, L. Barbero, B. R.

Carter, R. E. Sonnerup, L. Talley, G. A. Tarran, D. L. Volkov, A. C. Martiny, *bioRxiv*, in press (available at https://doi.org/10.1101/2020.09.06.285056).

23. M. C. Nielsdóttir, C. M. Moore, R. Sanders, D. J. Hinz, E. P. Achterberg, Iron limitation of the postbloom phytoplankton communities in the Iceland Basin. *Global Biogeochem. Cycles*. **23** (2009), doi:10.1029/2008GB003410.

24. R. L. Mather, S. E. Reynolds, G. A. Wolff, R. G. Williams, S. Torres-Valdes, E. M. S. Woodward, A. Landolfi, X. Pan, R. Sanders, E. P. Achterberg, Phosphorus cycling in the North and South Atlantic Ocean subtropical gyres. *Nat. Geosci.* **1**, 439–443 (2008).

25. A. C. Martiny, M. W. Lomas, W. Fu, P. W. Boyd, Y. ling L. Chen, G. A. Cutter, M. J. Ellwood, K. Furuya, F. Hashihama, J. Kanda, D. M. Karl, T. Kodama, Q. P. Li, J. Ma, T. Moutin, E. M. S. Woodward, J. K. Moore, Biogeochemical controls of surface ocean phosphate. *Sci. Adv.* **5** (2019), doi:10.1126/sciadv.aax0341.

26. S. Bonnet, C. Guieu, F. Bruyant, O. Prášil, F. Van Wambeke, P. Raimbault, T. Moutin, C. Grob, M. Y. Gorbunov, J. P. Zehr, S. M. Masquelier, L. Garczarek, H. Claustre, Nutrient limitation of primary productivity in the Southeast Pacific (BIOSOPE cruise). *Biogeosciences*. **5**, 215–225 (2008).

27. M. J. Behrenfeld, Z. S. Kolber, Widespread iron limitation of phytoplankton in the south pacific ocean. *Science.* **283**, 840–843 (1999).

28. B. S. Twining, S. Rauschenberg, S. E. Baer, M. W. Lomas, A. C. Martiny, O. Antipova, A nutrient limitation mosaic in the eastern tropical Indian Ocean. *Deep. Res. Part II Top. Stud. Oceanogr.* **116**, 125–140 (2019).

29. M. J. Behrenfeld, T. K. Westberry, E. S. Boss, R. T. O'Malley, D. A. Siegel, J. D. Wiggert, B. A. Franz, C. R. McClain, G. C. Feldman, S. C. Doney, J. K. Moore, G. Dall'Olmo, A. J. Milligan, I. Lima, N. Mahowald, Satellite-detected fluorescence reveals global physiology of ocean phytoplankton. *Biogeosciences*. **6**, 779–794 (2009).

30. S. W. A. Naqvi, J. W. Moffett, M. U. Gauns, P. V. Narvekar, A. K. Pratihary, H. Naik, D. M. Shenoy, D. A. Jayakumar, T. J. Goepfert, P. K. Patra, A. Al-Azri, S. I. Ahmed, The Arabian Sea as a high-nutrient, low-chlorophyll region during the late Southwest Monsoon. *Biogeosciences*. **7**, 1091–2100 (2010).

31. A. R. Moreno, A. C. Martiny, Ecological Stoichiometry of Ocean Plankton. *Ann. Rev. Mar. Sci.* **10**, 43–69 (2018).

32. M. M. Mills, K. R. Arrigo, Magnitude of oceanic nitrogen fixation influenced by the nutrient uptake ratio of phytoplankton. *Nat. Geosci.* **3**, 412–416 (2010).

33. J. Wu, W. Sunda, E. A. Boyle, D. M. Karl, Phosphate depletion in the Western North Atlantic Ocean. *Science.* **289**, 759–762 (2000).

34. B. A. Ward, S. Dutkiewicz, C. M. Moore, M. J. Follows, Iron, phosphorus, and nitrogen supply ratios define the biogeography of nitrogen fixation. *Limnol. Oceanogr.* **58**, 2059–2075 (2013).

35. M. A. Saito, M. R. McIlvin, D. M. Moran, T. J. Goepfert, G. R. DiTullio, A. F. Post, C. H. Lamborg, Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. *Science.* (2014), doi:10.1126/science.1256450.

36. C. R. Loscher, W. Mohr, H. W. Bange, D. E. Canfield, No nitrogen fixation in the Bay of

Bengal? *Biogeosciences*. **17**, 851–864 (2020).

37. M. W. Lomas, A. L. Burke, D. A. Lomas, D. W. Bell, C. Shen, S. T. Dyhrman, J. W. Ammerman, Sargasso Sea phosphorus biogeochemistry: An important role for dissolved organic phosphorus (DOP). *Biogeosciences*. **7**, 695–710 (2010).

38. J. K. Moore, S. C. Doney, K. Lindsay, Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. *Global Biogeochem. Cycles*. **18** (2004), doi:10.1029/2004GB002220.

39. G. Rocap, F. W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N. A. Ahlgren, A. Arellano, M. Coleman, L. Hauser, W. R. Hess, Z. I. Johnson, M. Land, D. Lindell, A. F. Post, W. Regala, M. Shah, S. L. Shaw, C. Steglich, M. B. Sullivan, C. S. Ting, A. Tolonen, E. A. Webb, E. R. Zinser, S. W. Chisholm, Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature*. **424**, 1042–1047 (2003).

40. A. C. Martiny, S. Kathuria, P. M. Berube, Widespread metabolic potential for nitrite and nitrate assimilation among Prochlorococcus ecotypes. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 10787–10792 (2009).

41. M. V. Zubkov, B. M. Fuchs, G. A. Tarran, P. H. Burkill, R. Amann, High rate of uptake of organic nitrogen compounds by Prochlorococcus cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Appl. Environ. Microbiol.* **69**, 1299–1304 (2003).

42. H. Alexander, B. D. Jenkins, T. A. Rynearson, S. T. Dyhrman, Metatranscriptome analyses indicate resource partitioning between diatoms in the field. *Proc. Natl. Acad.*

Sci. U. S. A. **112**, E2182–E2190 (2015).

43. M. W. Lomas, J. A. Bonachela, S. A. Levin, A. C. Martiny, Impact of ocean phytoplankton diversity on phosphate uptake. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17540–17545 (2014).

44. J. D. Rocca, E. K. Hall, J. T. Lennon, S. E. Evans, M. P. Waldrop, J. B. Cotner, D. R. Nemergut, E. B. Graham, M. D. Wallenstein, Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. *ISME J.* **9**, 1693–1699 (2015).

45. S. Sunagawa, L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-Castillo, P. I. Costea, C. Cruaud, F. D'Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, C. Lepoivre, G. Lima-Mendez, J. Poulain, B. T. Poulos, M. Royo-Llonch, H. Sarmento, S. Vieira-Silva, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, E. Boss, M. Follows, L. Karp-Boss, U. Krzic, E. G. Reynaud, C. Sardet, M. Sieracki, D. Velayoudon, C. Bowler, C. De Vargas, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, F. Not, H. Ogata, S. Pesant, S. Speich, L. Stemmann, M. B. Sullivan, J. Weissenbach, P. Wincker, E. Karsenti, J. Raes, S. G. Acinas, P. Bork, Structure and function of the global ocean microbiome. *Science.* **348** (2015), doi:10.1126/science.1261359.

46. S. J. Biller, P. M. Berube, K. Dooley, M. Williams, B. M. Satinsky, T. Hackl, S. L. Hogle, A. Coe, K. Bergauer, H. A. Bouman, T. J. Browning, D. De Corte, C. Hassler, D. Hulston, J. E. Jacquot, E. W. Maas, T. Reinthaler, E. Sintes, T. Yokokawa, S. W. Chisholm, Data

descriptor: Marine microbial metagenomes sampled across space and time. *Sci. Data*. **5**, 180176 (2018).

47. A. A. Larkin, C. A. Garcia, K. A. Ingoglia, N. S. Garcia, S. E. Baer, B. S. Twining, M. W. Lomas, A. C. Martiny, Subtle biogeochemical regimes in the Indian Ocean revealed by spatial and diel frequency of Prochlorococcus haplotypes. *Limnol. Oceanogr.* **65**, S220–S232 (2020).

48. K. H. Boström, K. Simu, Å. Hagström, L. Riemann, Optimization of DNA extraction for quantitative marine bacterioplankton community analysis. *Limnol. Oceanogr. Methods*. **2**, 365–373 (2004).

49. M. Baym, S. Kryazhimskiy, T. D. Lieberman, H. Chung, M. M. Desai, R. K. Kishony, Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One*. **10**, 1–15 (2015).

50. S. Wandro, A. Oliver, T. Gallagher, C. Weihe, W. England, J. B. H. Martiny, K. Whiteson, Predictable molecular adaptation of coevolving Enterococcus faeciumand lytic phage EfV12-phi1. *Front. Microbiol.* **9**, 3192 (2019).

51. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).

52. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).

53. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**,

2078–2079 (2009).

54. A. M. Eren, O. C. Esen, C. Quince, J. H. Vineis, H. G. Morrison, M. L. Sogin, T. O. Delmont, Anvi'o: An advanced analysis and visualization platformfor 'omics data. *PeerJ*. **3** (2015), doi:10.7717/peerj.1319.

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

56. S. Van Dongen, C. Abreu-Goodger, Using MCL to extract clusters from networks. *Methods Mol. Biol.* **804**, 281–295 (2012).

57. T. O. Delmont, E. M. Eren, Linking pangenomes and metagenomes: The Prochlorococcus metapangenome. *PeerJ*. **6** (2018), doi:10.7717/peerj.4320.

58. D. J. Scanlan, M. Ostrowski, S. Mazard, A. Dufresne, L. Garczarek, W. R. Hess, A. F. Post, M. Hagemann, I. Paulsen, F. Partensky, Ecological Genomics of Marine Picocyanobacteria. *Microbiol. Mol. Biol. Rev.* **73**, 249–299 (2009).

59. C. Troupin, A. Barth, D. Sirjacobs, M. Ouberdous, J. M. Brankart, P. Brasseur, M. Rixen, A. Alvera-Azcárate, M. Belounis, A. Capet, F. Lenartz, M. E. Toussaint, J. M. Beckers, Generation of analysis and consistent error fields using the Data Interpolating Variational Analysis (DIVA). *Ocean Model.* **52–53**, 90–101 (2012).

60. W.-L. Wang, J. K. Moore, A. C. Martiny, F. W. Primeau, Convergent estimates of marine nitrogen fixation. *Nature*. **566**, 205–211 (2019).

61. R. T. Letscher, J. K. Moore, Y. C. Teng, F. Primeau, Variable C : N : P stoichiometry of dissolved organic matter cycling in the Community Earth System Model.

Biogeosciences. **12**, 209–221 (2015).

62. P. I. Costea, R. Munch, L. P. Coelho, L. Paoli, S. Sunagawa, P. Bork, metaSNV: A tool for metagenomic strain level analysis. *PLoS One*. **12** (2017), doi:10.1371/journal.pone.0182392.

63. B. J. Callahan, P. J. McMurdie, S. P. Holmes, Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **11** (2017), doi:10.1038/ismej.2017.119.

64. A. M. Eren, L. Maignien, W. J. Sul, L. G. Murphy, S. L. Grim, H. G. Morrison, M. L. Sogin, Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* **4** (2013), doi:10.1111/2041-210X.12114.

65. S. J. Giovannoni, T. B. Britschgi, C. L. Moyer, K. G. Field, Genetic diversity in Sargasso Sea bacterioplankton. *Nature*. **345** (1990), doi:10.1038/345060a0.

66. L. R. Moore, G. Rocap, S. W. Chisholm, Physiology and molecular phytogeny of coexisting Prochlorococcus ecotypes. *Nature* (1998), doi:10.1038/30965.

67. J. A. Fuhrman, U. Campbell, Microbial microdiversity. *Nature* (1998), , doi:10.1038/30839.

68. D. C. Schlatter, L. L. Kinkel, Global biogeography of Streptomyces antibiotic inhibition, resistance, and resource use. *FEMS Microbiol. Ecol.* **88** (2014), doi:10.1111/1574-6941.12307.

69. R. A. Welch, V. Burland, G. Plunkett, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S. R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz,

N. T. Perna, H. L. T. Mobley, M. S. Donnenberg, F. R. Blattner, Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **99** (2002), doi:10.1073/pnas.252529799.

70. F. A. Hussain, J. Dubert, J. Elsherbini, M. Murphy, D. VanInsberghe, P. Arevalo, K. Kauffman, B. K. Rodino-Janeiro, H. Gavin, A. Gomez, A. Lopatina, F. Le Roux, M. F. Polz, Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. *Science.* **374** (2021), doi:10.1126/science.abb1083.

71. S. J. Biller, P. M. Berube, D. Lindell, S. W. Chisholm, Prochlorococcus: The structure and function of collective diversity. *Nat. Rev. Microbiol.* (2015), , doi:10.1038/nrmicro3378.

72. J. B. H. Martiny, S. E. Jones, J. T. Lennon, A. C. Martiny, Microbiomes in light of traits: A phylogenetic perspective. *Science.* (2015), , doi:10.1126/science.aac9323.

73. Z. I. Johnson, E. R. Zinser, A. Coe, N. P. McNulty, E. M. S. Woodward, S. W. Chisholm, Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients. *Science.* (2006), doi:10.1126/science.1118052.

74. E. R. Zinser, Z. I. Johnson, A. Coe, E. Karaca, D. Veneziano, S. W. Chisholm, Influence of light and temperature on Prochlorococcus ecotype distributions in the Atlantic Ocean. *Limnol. Oceanogr.* (2007), doi:10.4319/lo.2007.52.5.2205.

75. G. K. Farrant, H. Doré, F. M. Cornejo-Castillo, F. Partensky, M. Ratin, M. Ostrowski, F. D. Pitt, P. Wincker, D. J. Scanlan, D. Iudicone, S. G. Acinas, L. Garczarek, Delineating ecologically significant taxonomic units from global patterns of marine

picocyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **113** (2016), doi:10.1073/pnas.1524865113.

76. A. A. Larkin, S. K. Blinebry, C. Howes, Y. Lin, S. E. Loftus, C. A. Schmaus, E. R. Zinser, Z. I. Johnson, Niche partitioning and biogeography of high light adapted Prochlorococcus across taxonomic ranks in the North Pacific. *ISME J.* (2016), doi:10.1038/ismej.2015.244.

77. A. G. Kent, C. L. Dupont, S. Yooseph, A. C. Martiny, Global biogeography of Prochlorococcus genome diversity in the surface ocean. *ISME J.* **10** (2016), doi:10.1038/ismej.2015.265.

78. P. M. Berube, A. Rasmussen, R. Braakman, R. Stepanauskas, S. W. Chisholm, Emergence of trait variability through the lens of nitrogen assimilation in prochlorococcus. *Elife* (2019), doi:10.7554/eLife.41043.

79. L. J. Ustick, A. A. Larkin, C. A. Garcia, N. S. Garcia, M. L. Brock, J. A. Lee, N. A. Wiseman, J. K. Moore, A. C. Martiny, Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation. *Science.* **372**, 287–291 (2021).

80. M. B. Sullivan, M. L. Coleman, P. Weigele, F. Rohwer, S. W. Chisholm, Three Prochlorococcus cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol.* **3** (2005), doi:10.1371/journal.pbio.0030144.

81. A. C. Martiny, A. P. K. Tai, D. Veneziano, F. Primeau, S. W. Chisholm, Taxonomic resolution, ecotypes and the biogeography of Prochlorococcus. *Environ. Microbiol.* **11** (2009), doi:10.1111/j.1462-2920.2008.01803.x.

82. L. G. M. Baas-Becking, *Geobiologie of inleiding tot de milieukunde* (1934).

83. C. A. Hanson, J. A. Fuhrman, M. C. Horner-Devine, J. B. H. Martiny, Beyond biogeographic patterns: Processes shaping the microbial landscape. *Nat. Rev. Microbiol.* (2012), , doi:10.1038/nrmicro2795.

84. J. B. H. Martiny, B. J. M. Bohannan, J. H. Brown, R. K. Colwell, J. A. Fuhrman, J. L. Green, M. C. Horner-Devine, M. Kane, J. A. Krumins, C. R. Kuske, P. J. Morin, S. Naeem, L. Øvreås, A. L. Reysenbach, V. H. Smith, J. T. Staley, Microbial biogeography: Putting microorganisms on the map. *Nat. Rev. Microbiol.* **4** (2006), , doi:10.1038/nrmicro1341.

85. A. A. Larkin, A. C. Martiny, Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environ. Microbiol. Rep.* (2017), , doi:10.1111/1758-2229.12523.

86. C. Quince, A. W. Walker, J. T. Simpson, N. J. Loman, N. Segata, Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* **35** (2017), , doi:10.1038/nbt.3935.

87. J. S. Ghurye, V. Cepeda-Espinoza, M. Pop, Metagenomic assembly: Overview, challenges and applications. *Yale J. Biol. Med.* **89** (2016).

88. S. J. Biller, P. M. Berube, K. Dooley, M. Williams, B. M. Satinsky, T. Hackl, S. L. Hogle, A. Coe, K. Bergauer, H. A. Bouman, T. J. Browning, D. De Corte, C. Hassler, D. Hulston, J. E. Jacquot, E. W. Maas, T. Reinthaler, E. Sintes, T. Yokokawa, S. W. Chisholm, Marine microbial metagenomes sampled across space and time. *Sci. Data*. **5**, 180176 (2018).

89. R Core Team, R: A Language and Environment for Statistical Computing (2021).

90. Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene Wagner, vegan: Community Ecology Package (2020).

91. A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, P. A. Pevzner, SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

92. D. D. Kang, F. Li, E. Kirton, A. Thomas, R. Egan, H. An, Z. Wang, MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*. **7**, e7359 (2019).

93. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

94. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

95. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

96. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).

97. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic

tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

98. D. G. Boyce, M. R. Lewis, B. Worm, Global phytoplankton decline over the past century. *Nature*. **466** (2010), doi:10.1038/nature09268.

99. J. Keith Moore, W. Fu, F. Primeau, G. L. Britten, K. Lindsay, M. Long, S. C. Doney, N. Mahowald, F. Hoffman, J. T. Randerson, Sustained climate warming drives declining marine biological productivity. *Science.* **359** (2018), doi:10.1126/science.aao6379.

100. P. Frémont, M. Gehlen, M. Vrac, J. Leconte, T. O. Delmont, P. Wincker, D. Iudicone, O. Jaillon, Restructuring of genomic provinces of surface ocean plankton under climate change. *bioRxiv* (2021).

101. P. W. Boyd, T. Jickells, C. S. Law, S. Blain, E. A. Boyle, K. O. Buesseler, K. H. Coale, J. J. Cullen, H. J. W. De Baar, M. Follows, M. Harvey, C. Lancelot, M. Levasseur, N. P. J. Owens, R. Pollard, R. B. Rivkin, J. Sarmiento, V. Schoemann, V. Smetacek, S. Takeda, A. Tsuda, S. Turner, A. J. Watson, Mesoscale iron enrichment experiments 1993-2005: Synthesis and future directions. *Science.* **315** (2007), , doi:10.1126/science.1131669.

102. E. A. Laws, T. T. Bannister, Nutrient- and light-limited growth of Thalassiosira fluviatilis in continuous culture, with implications for phytoplankton growth in the ocean. *Limnol. Oceanogr.* **25** (1980), doi:10.4319/lo.1980.25.3.0457.

103. M. J. Behrenfeld, K. H. Halsey, A. J. Milligan, in *Philosophical Transactions of the Royal Society B: Biological Sciences* (2008), vol. 363.

104. M. J. Behrenfeld, E. Boss, D. A. Siegel, D. M. Shea, Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochem. Cycles*. **19** (2005),

doi:10.1029/2004GB002299.

105. T. Westberry, M. J. Behrenfeld, D. A. Siegel, E. Boss, Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochem. Cycles* (2008), doi:10.1029/2007GB003078.

106. J. R. Graff, T. K. Westberry, A. J. Milligan, M. B. Brown, G. Dall'Olmo, K. M. Reifel, M. J. Behrenfeld, Photoacclimation of natural phytoplankton communities. *Mar. Ecol. Prog. Ser.* **542** (2016), doi:10.3354/meps11539.

107. T. K. Westberry, M. J. Behrenfeld, A. J. Milligan, S. C. Doney, Retrospective satellite ocean color analysis of purposeful and natural ocean iron fertilization. *Deep. Res. Part I Oceanogr. Res. Pap.* **73** (2013), doi:10.1016/j.dsr.2012.11.010.

108. B. A. S. Van Mooy, H. F. Fredricks, B. E. Pedler, S. T. Dyhrman, D. M. Karl, M. Koblížek, M. W. Lomas, T. J. Mincer, L. R. Moore, T. Moutin, M. S. Rappé, E. A. Webb, Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* (2009), doi:10.1038/nature07659.

109. R. R. Hood, S. W. A. Naqvi, J. D. Wiggert, A. Subramaniam, in *Eos* (2007), vol. 88.

110. P. N. M. Vinayachandran, Y. Masumoto, M. J. Roberts, J. A. Huggett, I. Halo, A. Chatterjee, P. Amol, G. V. M. Gupta, A. Singh, A. Mukherjee, S. Prakash, L. E. Beckley, E. J. Raes, R. Hood, Reviews and syntheses: Physical and biogeochemical processes associated with upwelling in the Indian Ocean. *Biogeosciences*. **18** (2021), , doi:10.5194/bg-18-5967-2021.

111. W. Cui, J. Yang, Y. Ma, A statistical analysis of mesoscale eddies in the Bay of Bengal

from 22–year altimetry data. *Acta Oceanol. Sin.* **35** (2016), doi:10.1007/s13131-016-0945-3.

112. S. A. Piontkovski, S. Al-Jufaili, Coastal upwellings and Mesoscale Eddies of the Western Arabian Sea : Some Biological Implications. *Int. J. Ocean. Oceanogr.* **7** (2013).

APPENDIX A: Ch.1 Supplemental Figures and Tables

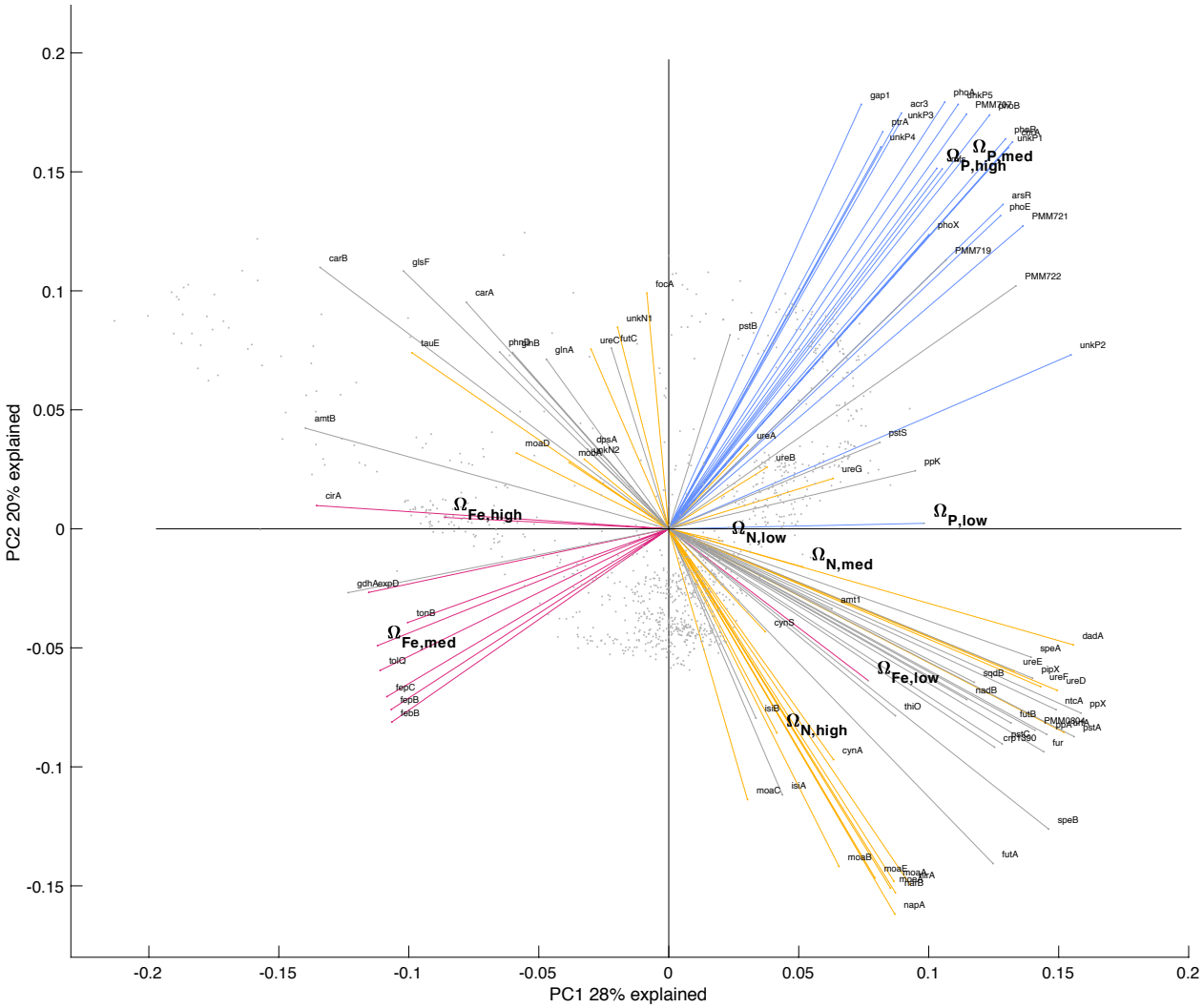

Fig. S1.1. Principal component analysis of stress genes across all metagenome samples overlaid with vectors for each gene (n = 1137). The eigenvectors for stress genes (z_i) and composite metrics (Ω_s) are colored according to nutrient type (red=Fe, blue=P, and yellow=N). This figure matches Fig. 1A but has gene vectors overlaid.

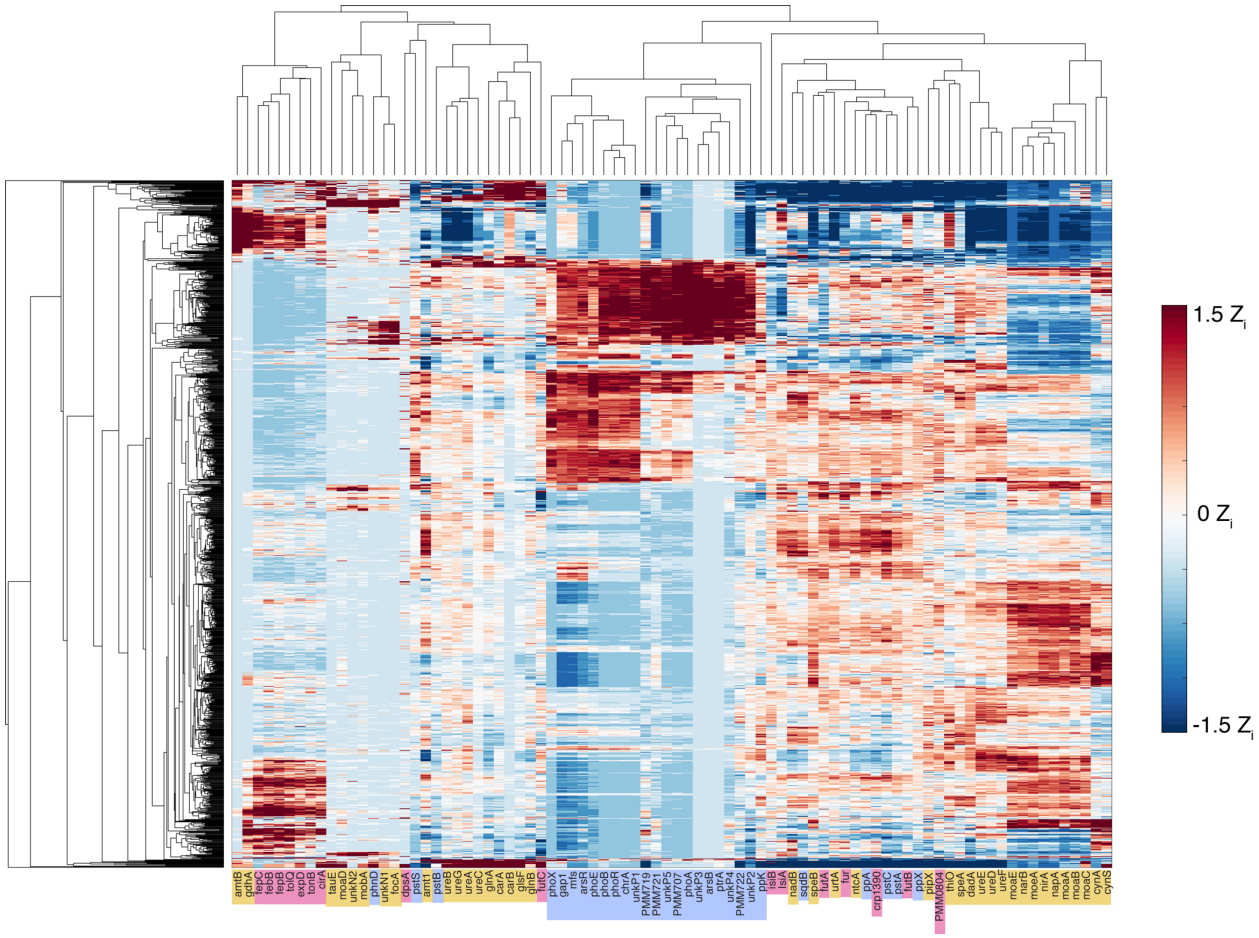

Fig. S1.2. Clustering of gene abundances (x-axis) and samples (y-axis). The clustering is based on an Euclidean distance and average linkage hierarchical clustering of z_i. Genes are colored according to nutrient type (red=Fe, blue=P, and yellow=N).

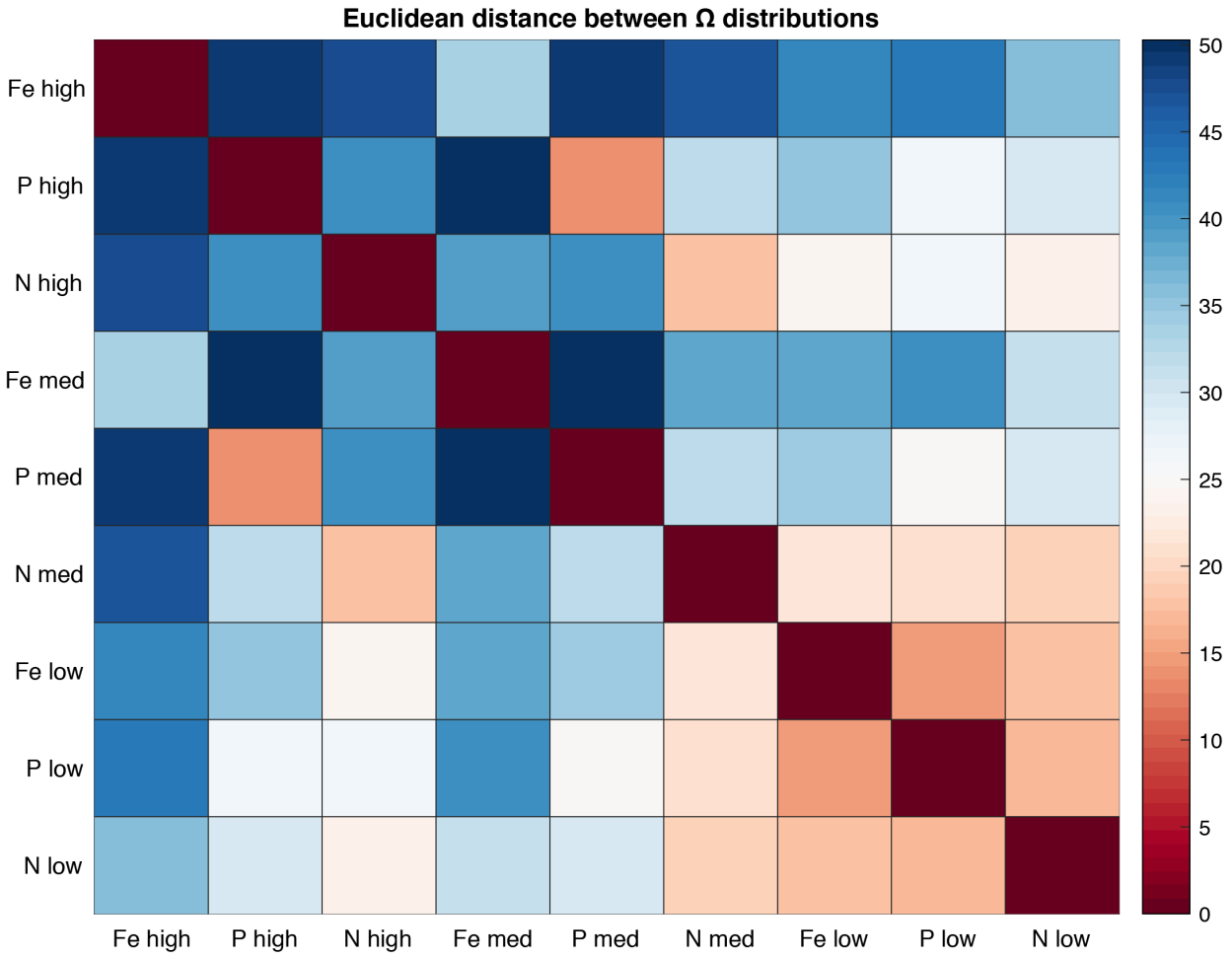

Fig. S1.3. Euclidean distance between Ω measurements across all samples. Ω values are ordered by first severity (high, med, low) then type (Fe, P, N).

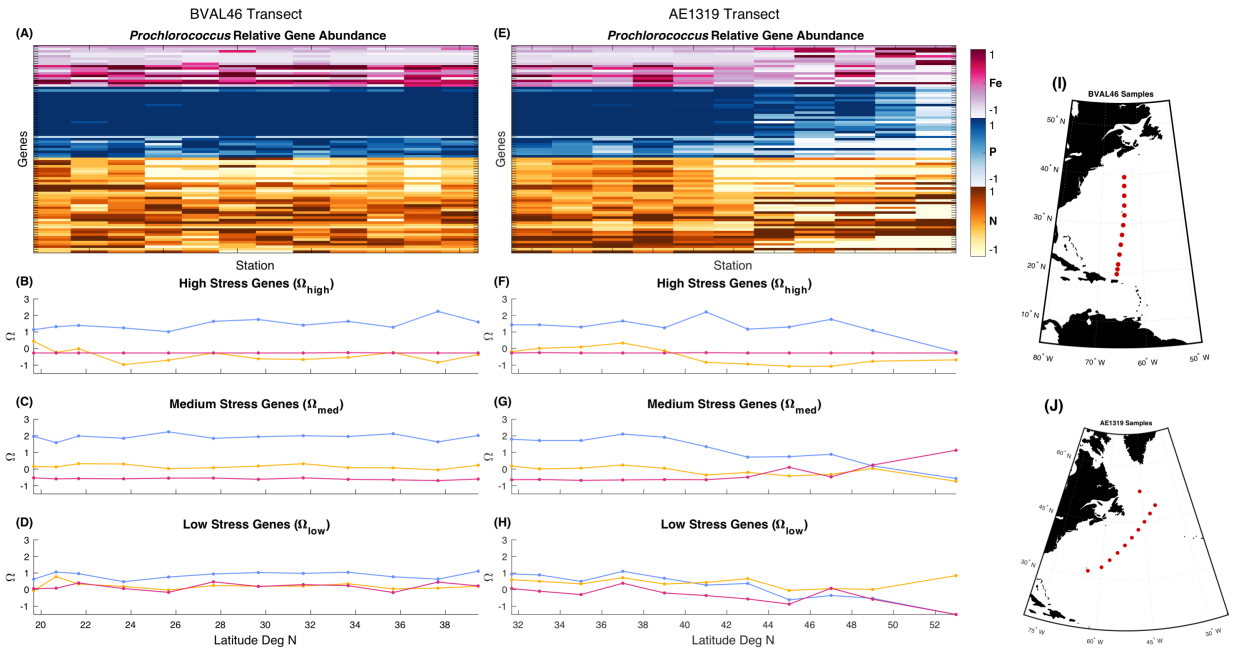

Fig. S1.4. Changes in nutrient stress genes for cruise AE1319 and BVAl46 to the western North Atlantic Ocean. Colors across the figure coincide with red for iron, blue for phosphorus, yellow for nitrogen genes. a/e) Z-score of individual genes ordered by high to low stress genes colored by type. Composite stress (Ω_s) grouped by severity and colored by type. b/f) high stress genes, c/g) medium stress genes, d/h) low stress genes, i) map of BVAL samples, j) map of AE1319 samples.

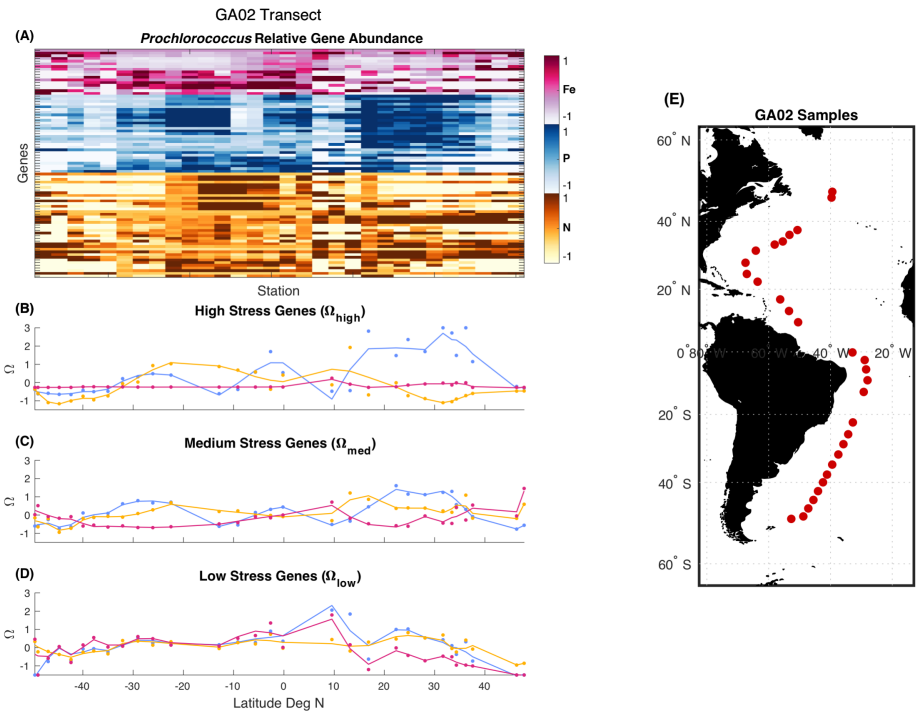

Fig. S1.5. Changes in nutrient stress genes for cruise GA02 to the western Atlantic Ocean. Colors across the figure coincide with red for iron, blue for phosphorus, yellow for nitrogen genes. a) Z-score of individual genes ordered by high to low stress genes colored by type. Composite stress (Ω_s) grouped by severity and colored by type. b) high stress genes c) medium stress genes d) low stress genes. e) map of GA02 cruise samples.

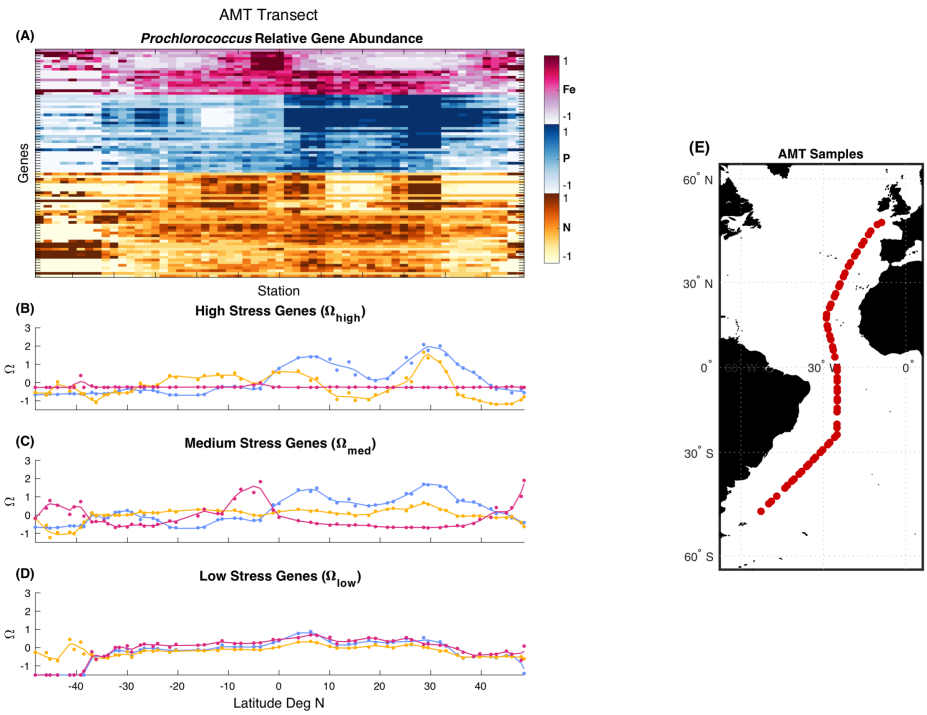

Fig. S1.6. Changes in nutrient stress genes for cruise AMT28 to the Atlantic Ocean. Colors across the figure coincide with red for iron, blue for phosphorus, yellow for nitrogen genes. a) Z-score of individual genes ordered by high to low stress genes colored by type. Composite stress (Ω_s) grouped by severity and colored by type. b) high stress genes c) medium stress genes d) low stress genes. e) map of AMT cruise samples.

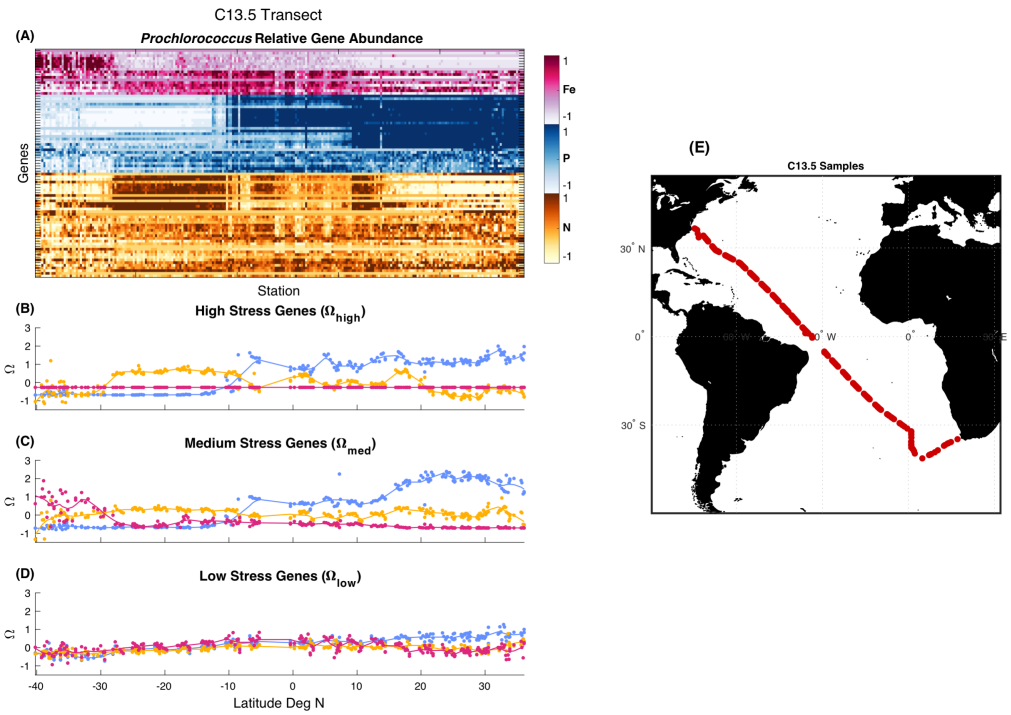

Fig. S1.7. Changes in nutrient stress genes for cruise C13.5 to the western North Atlantic Ocean. Colors across the figure coincide with red for iron, blue for phosphorus, yellow for nitrogen genes. a) Z-score of individual genes ordered by high to low stress genes colored by type. Composite stress (Ω_s) grouped by severity and colored by type. b) high stress genes c) medium stress genes d) low stress genes. e) map of C13.5 cruise samples.

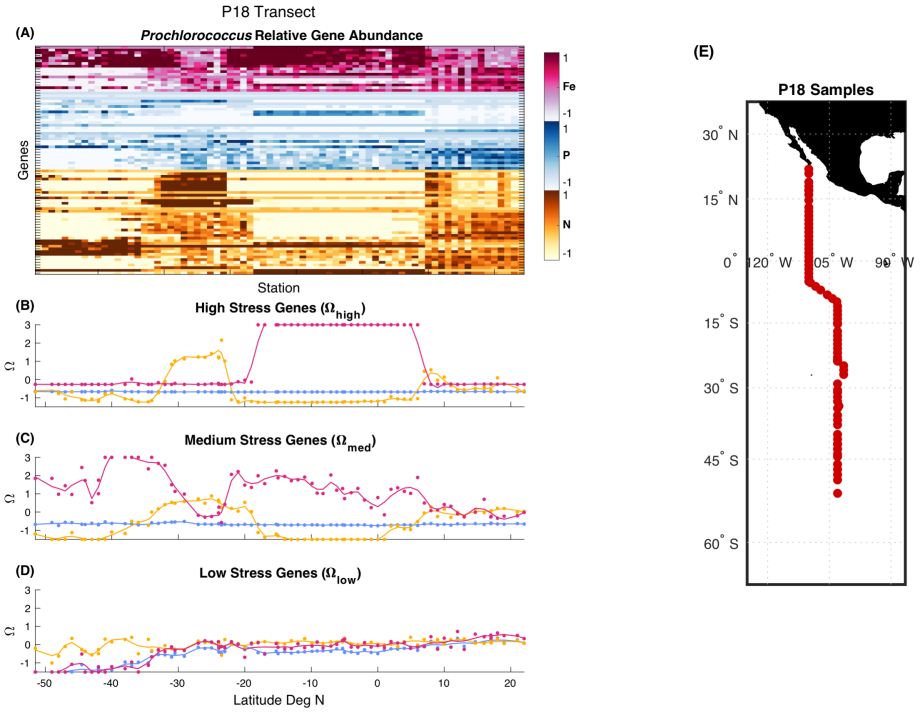

Fig. S1.8. Changes in nutrient stress genes for cruise P18 to the eastern Pacific Ocean. Colors across the figure coincide with red for iron, blue for phosphorus, yellow for nitrogen genes. a) Z-score of individual genes ordered by high to low stress genes colored by type. Composite stress (Ω_s) grouped by severity and colored by type. b) high stress genes c) medium stress genes d) low stress genes. e) map of P18 cruise samples.

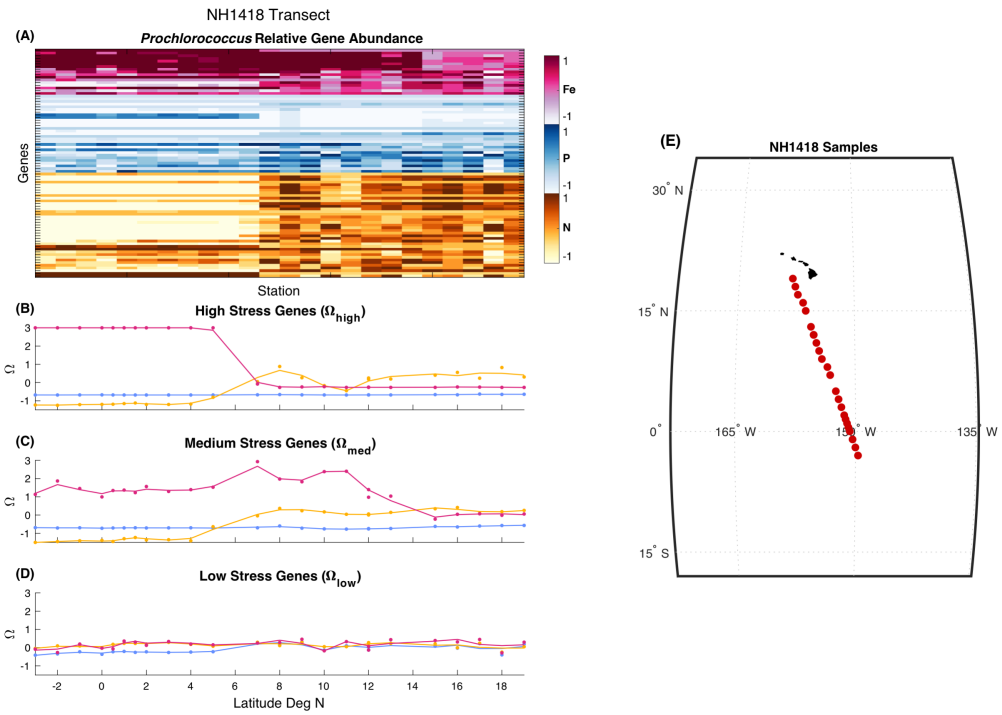

Fig. S1.9. Changes in nutrient stress genes for cruise NH1418 to the central Pacific Ocean. Colors across the figure coincide with red for iron, blue for phosphorus, yellow for nitrogen genes. a) Z-score of individual genes ordered by high to low stress genes colored by type. Composite stress (Ω_s) grouped by severity and colored by type. b) high stress genes c) medium stress genes d) low stress genes. e) map of NH1418 cruise samples.

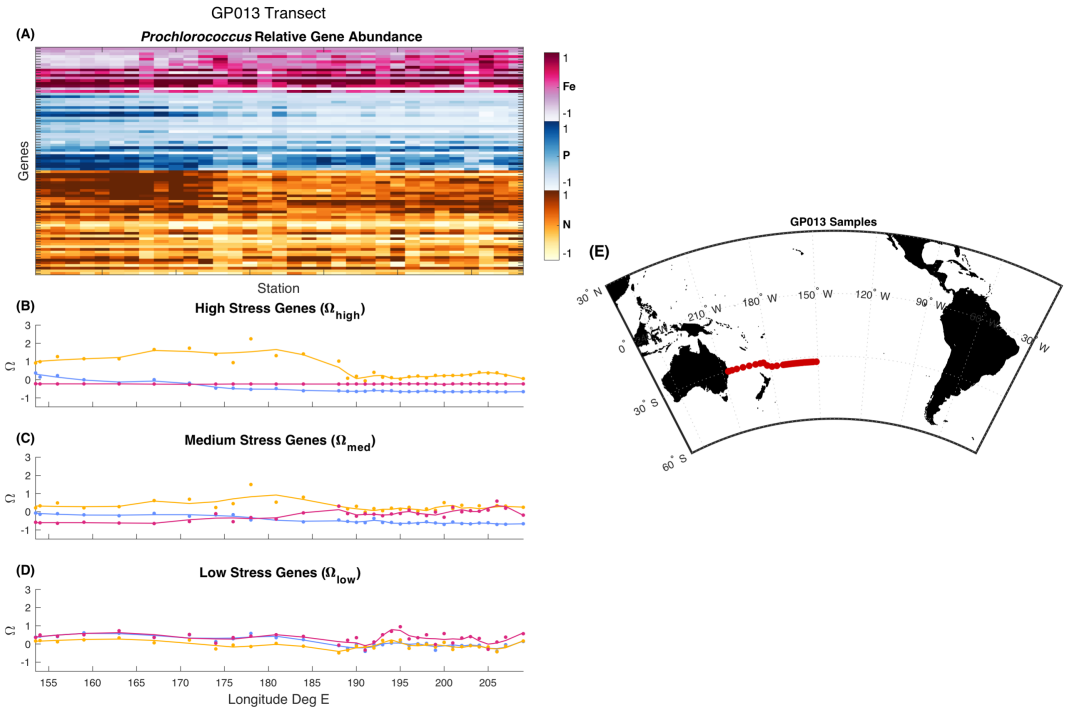

Fig. S1.10. Changes in nutrient stress genes for cruise GP13 to the western South Pacific Ocean. Colors across the figure coincide with red for iron, blue for phosphorus, yellow for nitrogen genes. a) Z-score of individual genes ordered by high to low stress genes colored by type. Composite stress (Ω_s) grouped by severity and colored by type. b) high stress genes c) medium stress genes d) low stress genes. e) map of GP13 cruse samples.

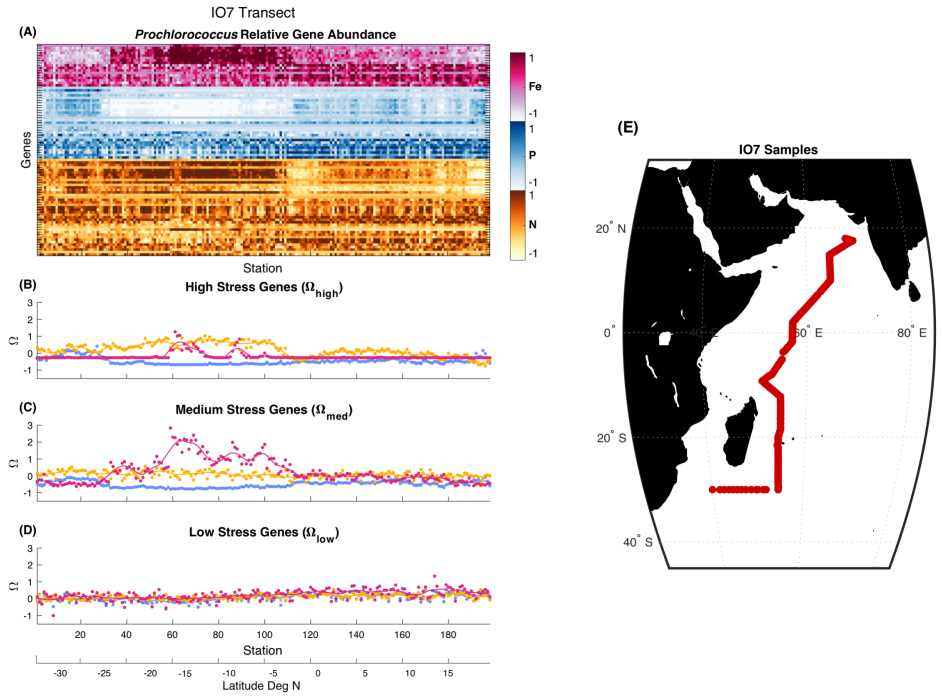

Fig. S1.11. Changes in nutrient stress genes for cruise IO7N to the western Indian Ocean. Colors across the figure coincide with red for iron, blue for phosphorus, yellow for nitrogen genes. a) Z-score of individual genes ordered by high to low stress genes colored by type. Composite stress (Ω_s) grouped by severity and colored by type. b) high stress genes c) medium stress genes d) low stress genes. e) map of IO7N cruise samples.

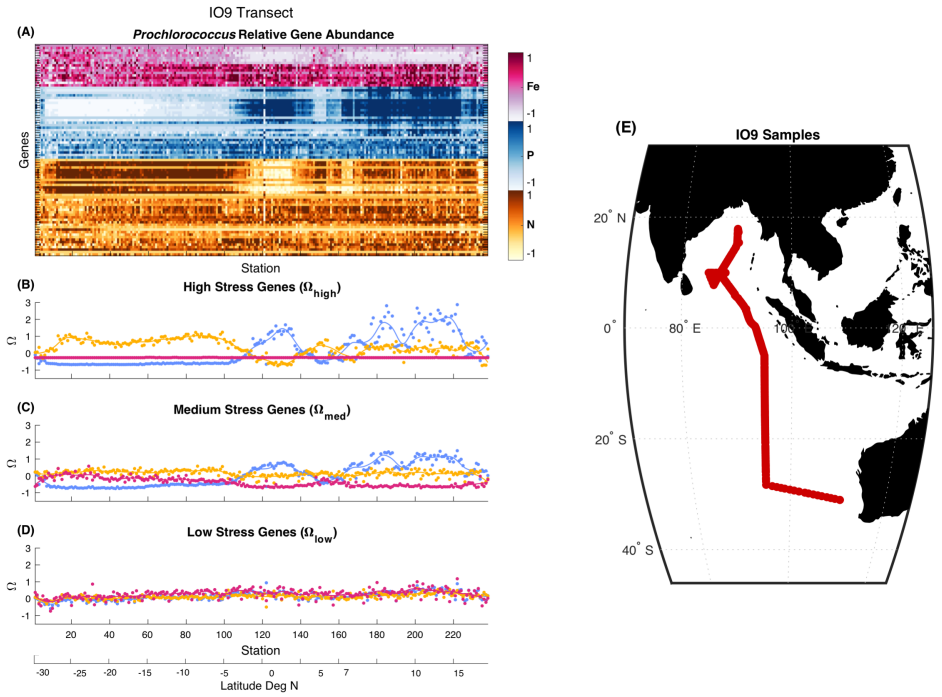

Fig. S1.12. Changes in nutrient stress genes for cruise I09N to the eastern Indian Ocean.

Colors across the figure coincide with red for iron, blue for phosphorus, yellow for nitrogen genes. a) Z-score of individual genes ordered by high to low stress genes colored by type. Composite stress (Ω_s) grouped by severity and colored by type. b) high stress genes c) medium stress genes d) low stress genes. e) map of IO9 cruise samples.

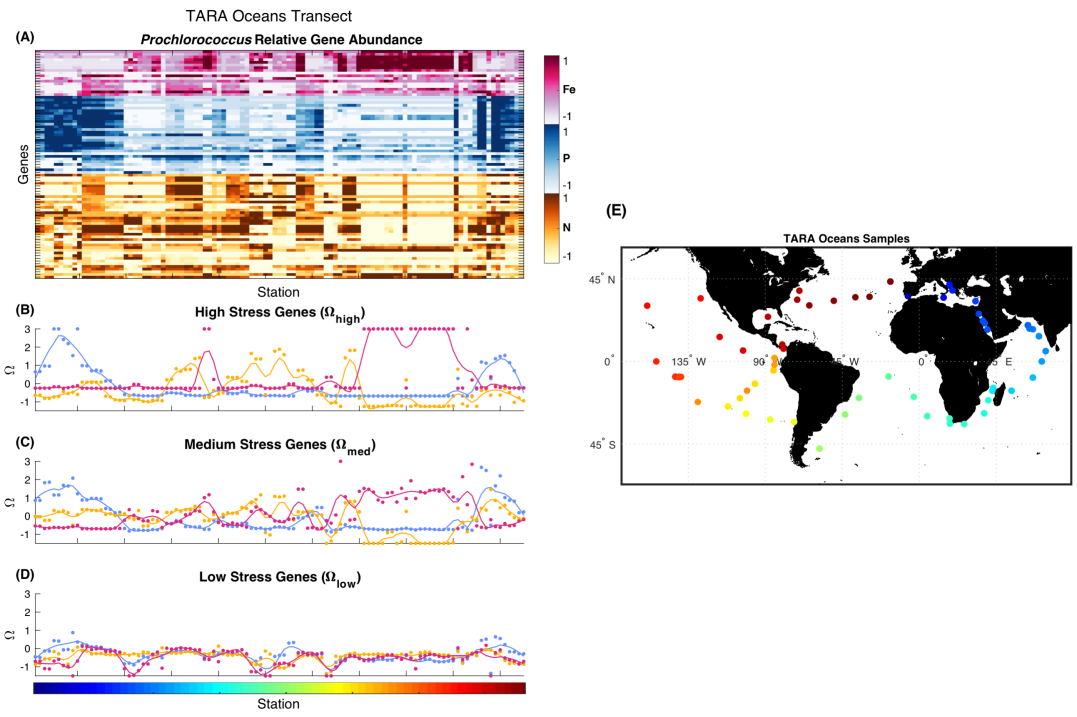

Fig. S1.13. Changes in nutrient stress genes for Tara Ocean global sampling program.

Colors across the figure coincide with red for iron, blue for phosphorus, yellow for nitrogen genes. a) Z-score of individual genes ordered by high to low stress genes colored by type. Composite stress (Ω_s) grouped by severity and colored by type. b) high stress genes c) medium stress genes d) low stress genes ordered by station. e) map of TARA Oceans cruise samples colored by station.

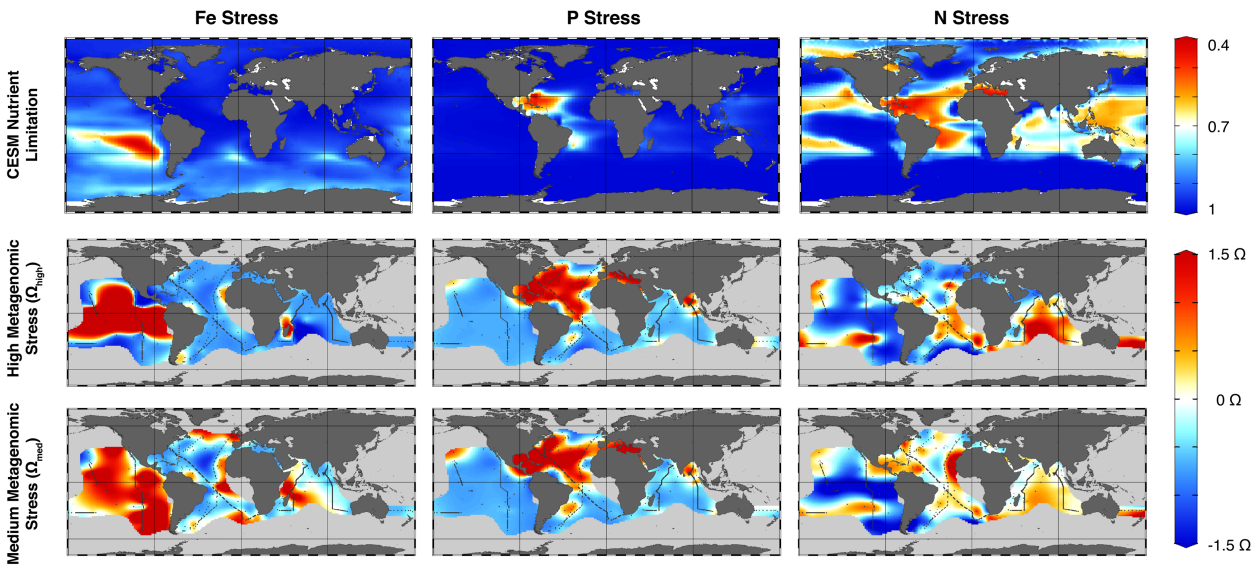

Fig. S1.14. Comparison between nutrient stress predicted using metagenomics and an Earth system model. The columns represent nutrient limitation type and rows represent the Earth system model as well as stress severity. The model simulations are from CESM BEC v1.21.

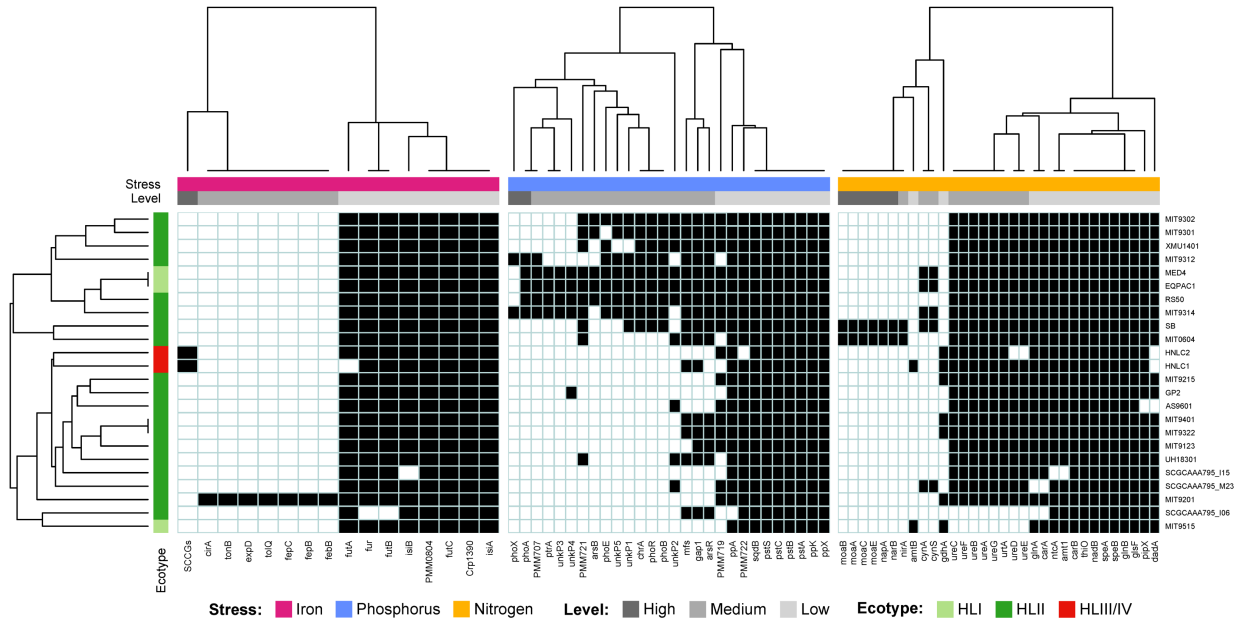

Fig. S1.15. Variation in nutrient acquisition genes among representative genomes from *Prochlorococcus* high-light ecotypes. Genes classified as indicative of high, medium, and low stress in this study largely cluster together. Gene presence reveals a hierarchical structure, whereby low stress genes are present in nearly all genomes, medium stress genes are present in a smaller subset of genomes, and high stress genes are a subset of genomes with medium stress profiles. An exception to this pattern is identified for Fe stress, whereby single copy core genes (SCCGs) unique to the HLIII/IV ecotype (i.e., the HNLC ecotype) were used as an indicator of high Fe stress. Genes that were absent in all high light genomes are not depicted. The ordering of the x- and y-axes are according to average linkage hierarchical clustering of the presence (black squares) versus absence (white squares) of nutrient genes.

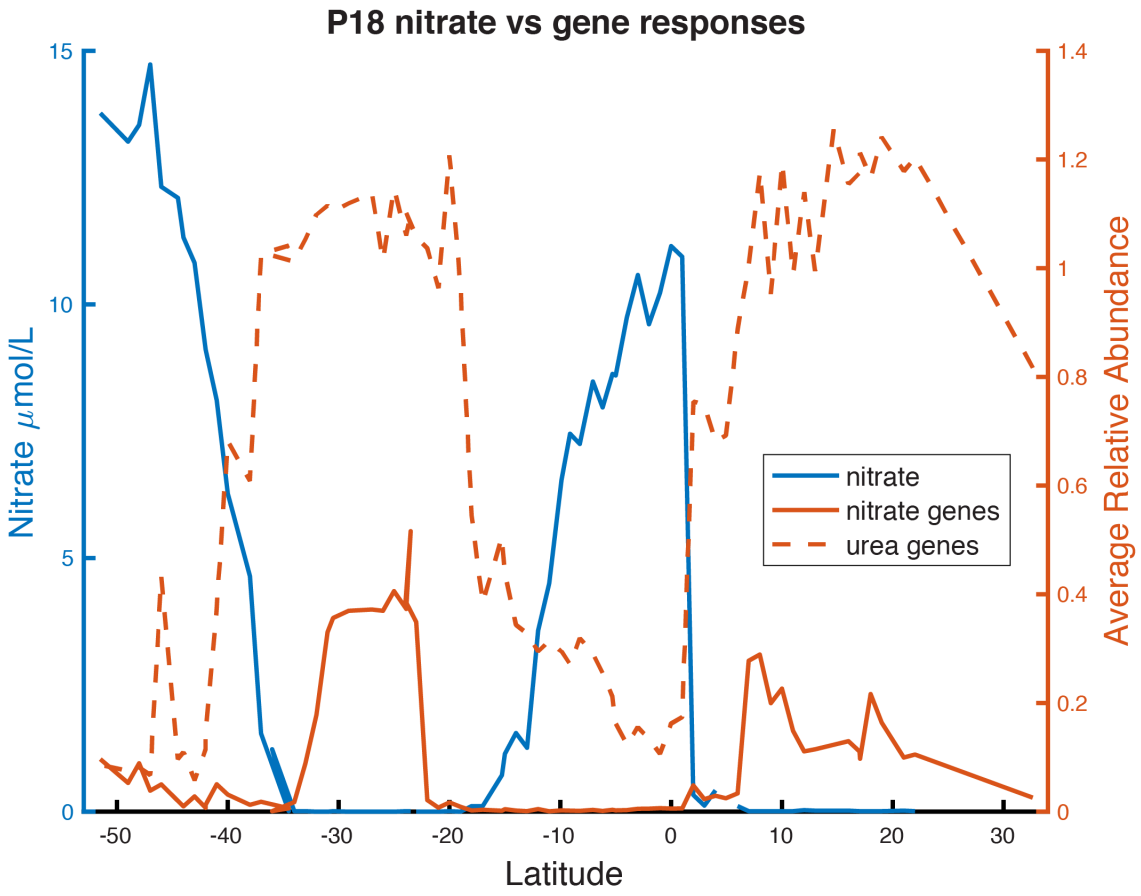

Fig. S1.16. Comparison between nitrate concentration, nitrate genes and urea genes across the P18 cruse. Nitrate concentrations are represented in blue with units on the left axis. Average gene abundances relative to SCCG abundance shown in orange. Nitrate genes (narB, moaA-E) related to high stress represented as a solid line and urea genes (ureA-G) related to medium stress as a dashed line.

Table S1.1. Metagenome samples across individual cruises

Cruse/Dataset	# Samples	Region	Citation
AE1319	11	NW Atlantic	(*22*)
AMT28	59	Atlantic	(*22*)
BATS	20	NW Atlantic	(*46*)
BVAL46	12	NW Atlantic	(*22*)
C13.5	241	Atlantic	(*22*)
GA02	30	Atlantic	(*46*)
GA03	3	N Atlantic	(*46*)
GA10	16	SE Atlantic	(*46*)
GP13	33	SW Pacific	(*46*)
HOT	21	Station ALOHA	(*46*)
I07	249	W Indian	(*22*)
I09	238	E Indian	(*22*)
NH1418	24	NW Pacific	(*22*)
P18	75	W Pacific	(*22*)
Tara	105	Global	(*45*)

Table S1.2. Angular intervals of nutrient stress type for the PCA analysis (Fig. 1A and S1)

Stress Type	Above θ	Below or Equal θ	Number of Samples
Fe Stress	0.8 π	1.3 π	215
Fe/P Co-Stress	0.64 π	0.8 π	15
P Stress	0.12 π	0.64 π	268
P/N Co-Stress	1.9 π	0.12 π	95
N Stress	1.46 π	1.9 π	432
N/Fe Co-Stress	1.3 π	1.46 π	112

Table S1.3. Read recruitment reference genomes

Genome	Species
EQPAC1	*Prochlorococcus*
GP2	*Prochlorococcus*
HNLC1	*Prochlorococcus*
HNLC2	*Prochlorococcus*
MED4	*Prochlorococcus*
MIT0601	*Prochlorococcus*
MIT0602	*Prochlorococcus*
MIT0604	*Prochlorococcus*
MIT0701	*Prochlorococcus*
MIT0801	*Prochlorococcus*
MIT1312	*Prochlorococcus*
MIT1313	*Prochlorococcus*
MIT1318	*Prochlorococcus*
MIT1327	*Prochlorococcus*
MIT1342	*Prochlorococcus*
MIT9123	*Prochlorococcus*
MIT9201	*Prochlorococcus*
MIT9211	*Prochlorococcus*
MIT9215	*Prochlorococcus*
MIT9301	*Prochlorococcus*
MIT9302	*Prochlorococcus*
MIT9303	*Prochlorococcus*
MIT9312	*Prochlorococcus*
MIT9313	*Prochlorococcus*
MIT9314	*Prochlorococcus*
MIT9322	*Prochlorococcus*
MIT9401	*Prochlorococcus*
MIT9515	*Prochlorococcus*
NATL1A	*Prochlorococcus*
NATL2A	*Prochlorococcus*
PAC1	*Prochlorococcus*
RS50	*Prochlorococcus*
SB	*Prochlorococcus*
SCGCAAA795_I06	*Prochlorococcus*
SCGCAAA795_I15	*Prochlorococcus*
SCGCAAA795_M23	*Prochlorococcus*
SS120	*Prochlorococcus*

UH18301	*Prochlorococcus*
XMU1401	*Prochlorococcus*
XMU1403	*Prochlorococcus*
XMU1408	*Prochlorococcus*
MIT9313	*Prochlorococcus*
Och114	*Roseobacter*
A9spades	*Pelagibacter*
AAA240_E13	*Pelagibacter*
AAA288_E13	*Pelagibacter*
AAA288_G21	*Pelagibacter*
AAA288_N07	*Pelagibacter*
AAA298_D23	*Pelagibacter*
AG_337_G04	*Pelagibacter*
AG_337_G06	*Pelagibacter*
AG_426_I15	*Pelagibacter*
AG_430_F16	*Pelagibacter*
AG_430_I06	*Pelagibacter*
B4spades	*Pelagibacter*
F4spades	*Pelagibacter*
HIMB058	*Pelagibacter*
HIMB083	*Pelagibacter*
HIMB114	*Pelagibacter*
HIMB122	*Pelagibacter*
HIMB1321	*Pelagibacter*
HIMB140	*Pelagibacter*
HIMB4	*Pelagibacter*
HIMB5	*Pelagibacter*
HIMB59	*Pelagibacter*
HTCC1002	*Pelagibacter*
HTCC1013	*Pelagibacter*
HTCC1016	*Pelagibacter*
HTCC1040	*Pelagibacter*
HTCC1062	*Pelagibacter*
HTCC7211	*Pelagibacter*
HTCC7214	*Pelagibacter*
HTCC7217	*Pelagibacter*
HTCC8051	*Pelagibacter*
HTCC9022	*Pelagibacter*
HTCC9565	*Pelagibacter*
IMCC1322	*Pelagibacter*

IMCC9063	*Pelagibacter*
PRT004	*Pelagibacter*
SAR86B	*Pelagibacter*
BL107	*Synechococcus*
CB0101	*Synechococcus*
CB0205	*Synechococcus*
CC9311	*Synechococcus*
CC9605	*Synechococcus*
CC9616	*Synechococcus*
CC9902	*Synechococcus*
GEYO	*Synechococcus*
GFB01	*Synechococcus*
KORDI_100	*Synechococcus*
KORDI_49	*Synechococcus*
KORDI_52	*Synechococcus*
MITS9508	*Synechococcus*
MITS9509	*Synechococcus*
N19	*Synechococcus*
N32	*Synechococcus*
NKBG042902	*Synechococcus*
PCC7335	*Synechococcus*
RCC307	*Synechococcus*
REDSEA_S02_B4	*Synechococcus*
RS9916	*Synechococcus*
RS9917	*Synechococcus*
UW105	*Synechococcus*
UW106	*Synechococcus*
UW140	*Synechococcus*
UW179A	*Synechococcus*
UW179B	*Synechococcus*
UW69	*Synechococcus*
UW86	*Synechococcus*
WH5701	*Synechococcus*
WH7805	*Synechococcus*
WH8016	*Synechococcus*
WH8020	*Synechococcus*
WH8102	*Synechococcus*
WH8109	*Synechococcus*

APPENDIX B: Ch.2 Supplemental Figures and Tables

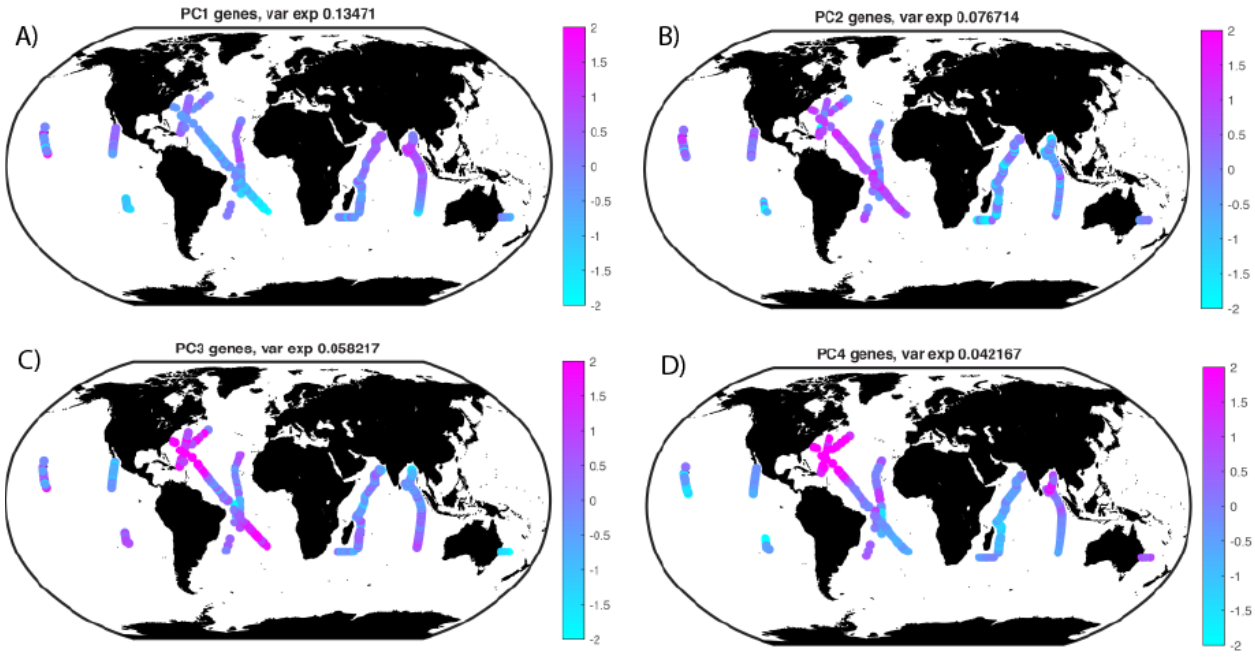

Figure S2.1: Gene component distributions.

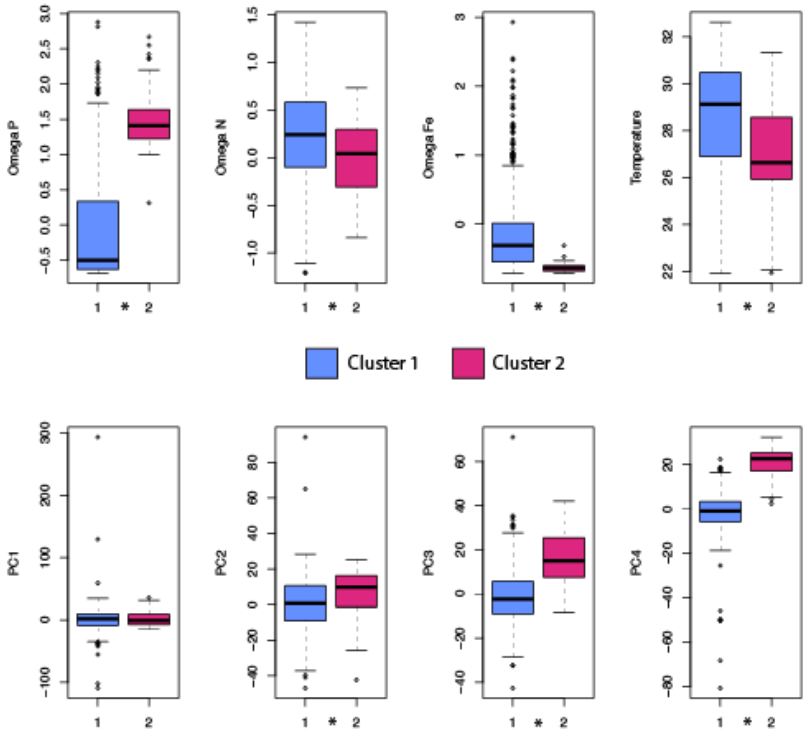

Figure S2.2: Environmental Stats. * denote significant differences in mean.

Table S2.1: Spearman Correlation between PCA analysis & environmental factors.

	PC1	PC2	PC3	PC4
P	p= 2.2e-16 rho= 0.3877267	p= 0.0001964 rho= 0.147982	p= 0.04191 rho= -0.0810904	p= 2.2e-16 rho= 0.8771648
Fe	p= 7.31e-06 rho= -0.1775951	p= 5.944e-09 rho= -0.2291705	p= 0.0003255 rho= -0.1427433	p= 2.2e-16 rho= -0.7884059
N	p= 5.638e-06 rho= -0.1800382	p= 9.578e-08 rho= -0.2110631	p= 2.2e-16 rho= 0.3912358	p= 2.2e-16 rho= -0.4713667
Temp	p= 2.2e-16 rho= 0.739164	p= 0.0009222 rho= -0.134813	p= 2.2e-16 rho= -0.3827844	p= 2.2e-16 rho= 0.4637373

Table S2.2: Reference Genome Database

Genome	Accession	Species
EQPAC1	**GCA_000759875.1**	*Prochlorococcus*
GP2	**GCA_000759885.1**	*Prochlorococcus*
HNLC1	**GCA_000218705.1**	*Prochlorococcus*
HNLC2	**GCA_000218745.1**	*Prochlorococcus*
MED4	**GCA_000011465.1**	*Prochlorococcus*
MIT0601	**GCA_000760175.1**	*Prochlorococcus*
MIT0602	**GCA_000760195.1**	*Prochlorococcus*
MIT0604	**GCA_000757845.1**	*Prochlorococcus*
MIT0701	**GCA_000760295.1**	*Prochlorococcus*
MIT0801	**GCA_000757865.1**	*Prochlorococcus*
MIT1312	**GCA_001632005.1**	*Prochlorococcus*
MIT1313	**GCA_001632065.1**	*Prochlorococcus*
MIT1318	**GCA_001632045.1**	*Prochlorococcus*
MIT1327	**GCA_001632125.1**	*Prochlorococcus*
MIT1342	**GCA_001632145.1**	*Prochlorococcus*
MIT9123	**GCA_000759935.1**	*Prochlorococcus*
MIT9201	**GCA_000759955.1**	*Prochlorococcus*
MIT9211	**GCA_000018585.1**	*Prochlorococcus*
MIT9215	**GCA_000018065.1**	*Prochlorococcus*
MIT9301	**GCA_000015965.1**	*Prochlorococcus*
MIT9302	**GCA_000759975.1**	*Prochlorococcus*
MIT9303	**GCA_000015705.1**	*Prochlorococcus*
MIT9312	**GCA_000012645.1**	*Prochlorococcus*
MIT9313	**GCA_000011485.1**	*Prochlorococcus*
MIT9314	**GCA_000760035.1**	*Prochlorococcus*
MIT9322	**GCA_000760075.1**	*Prochlorococcus*

MIT9401	**GCA_000760095.1**	*Prochlorococcus*
MIT9515	**GCA_000015665.1**	*Prochlorococcus*
NATL1A	**GCA_000015685.1**	*Prochlorococcus*
NATL2A	**GCA_000012465.1**	*Prochlorococcus*
PAC1	**GCA_000760235.1**	*Prochlorococcus*
RS50	**GCA_001989415.1**	*Prochlorococcus*
SB	**GCA_000760115.1**	*Prochlorococcus*
SCGCAAA795_I06	**NA**	*Prochlorococcus*
SCGCAAA795_I15	**NA**	*Prochlorococcus*
SCGCAAA795_M23	**NA**	*Prochlorococcus*
SS120	**GCA_000007925.1**	*Prochlorococcus*
UH18301	**SAMN00011132**	*Prochlorococcus*
XMU1401	**GCA_002812945.1**	*Prochlorococcus*
XMU1403	**GCA_003208065.1**	*Prochlorococcus*
XMU1408	**GCA_003208055.1**	*Prochlorococcus*
A9spades	**NA**	*Pelagibacter*
AAA240_E13	**SAMN02597172**	*Pelagibacter*
AAA288_E13	**SAMN02597171**	*Pelagibacter*
AAA288_G21	**SAMN02597281**	*Pelagibacter*
AAA288_N07	**SAMN02597280**	*Pelagibacter*
AAA298_D23	**GCA_000402655.1**	*Pelagibacter*
AG_337_G04	**ERS3879006**	*Pelagibacter*
AG_337_G06	**ERS3879008**	*Pelagibacter*
AG_426_I15	**ERS3880727**	*Pelagibacter*
AG_430_F16	**ERS3880946**	*Pelagibacter*
AG_430_I06	**ERS3880974**	*Pelagibacter*
B4spades	**NA**	*Pelagibacter*
F4spades	**NA**	*Pelagibacter*

HIMB058	**SAMN02440920**	*Pelagibacter*
HIMB083	**SAMN02597166**	*Pelagibacter*
HIMB114	**SAMN02436217**	*Pelagibacter*
HIMB122	**SRS843558**	*Pelagibacter*
HIMB1321	**GCA_900177485.1**	*Pelagibacter*
HIMB140	**ERS787856**	*Pelagibacter*
HIMB4	**NA**	*Pelagibacter*
HIMB5	**SAMN00016662**	*Pelagibacter*
HIMB59	**SAMN00010387**	*Pelagibacter*
HTCC1002	**SAMN02436088**	*Pelagibacter*
HTCC1013	**SAMN02441456**	*Pelagibacter*
HTCC1016	**SAMN02256429**	*Pelagibacter*
HTCC1040	**SAMN02256395**	*Pelagibacter*
HTCC1062	**SAMN02603690**	*Pelagibacter*
HTCC7211	**SAMN02436224**	*Pelagibacter*
HTCC7214	**SAMN02841172**	*Pelagibacter*
HTCC7217	**SAMN02841150**	*Pelagibacter*
HTCC8051	**SAMN02440710**	*Pelagibacter*
HTCC9022	**SAMN02440781**	*Pelagibacter*
HTCC9565	**GCA_012932795.1**	*Pelagibacter*
IMCC1322	**GCA_000024465.1**	*Pelagibacter*
IMCC9063	**SAMN02603337**	*Pelagibacter*
PRT004	**NA**	*Pelagibacter*
SAR86B	**GCA_000252545.1**	*Pelagibacter*
BL107	**GCF_000153805**	*Synechococcus*
CB0101	**GCA_000179235.2**	*Synechococcus*
CB0205	**GCA_000179255.1**	*Synechococcus*
CC9311	**GCF_000014585**	*Synechococcus*

CC9605	**GCF_000012625**	*Synechococcus*
CC9616	**GCF_000515235**	*Synechococcus*
CC9902	**GCF_000012505**	*Synechococcus*
GEYO	**GCF_900473955**	*Synechococcus*
GFB01	**GCA_001039265.1**	*Synechococcus*
KORDI_100	**GCF_000737535**	*Synechococcus*
KORDI_49	**GCF_000737575**	*Synechococcus*
KORDI_52	**GCF_000737595**	*Synechococcus*
MITS9508	**GCF_001632165**	*Synechococcus*
MITS9509	**GCF_001631935**	*Synechococcus*
N19	**GCF_900474045**	*Synechococcus*
N32	**GCF_900473895**	*Synechococcus*
NKBG042902	**GCA_000715475.1**	*Synechococcus*
PCC7335	**GCA_000155595.1**	*Synechococcus*
RCC307	**GCF_000063525**	*Synechococcus*
REDSEA_S02_B4	**GCA_001628325.1**	*Synechococcus*
RS9916	**GCF_000153825**	*Synechococcus*
RS9917	**GCF_000153065**	*Synechococcus*
UW105	**GCF_900473935**	*Synechococcus*
UW106	**GCF_900474015**	*Synechococcus*
UW140	**GCF_900474295**	*Synechococcus*
UW179A	**GCF_900473965**	*Synechococcus*
UW179B	**GCF_900474245**	*Synechococcus*
UW69	**GCF_900474185**	*Synechococcus*
UW86	**GCF_900474085**	*Synechococcus*
WH5701	**GCA_000153045.1**	*Synechococcus*
WH7805	**GCF_000153285**	*Synechococcus*
WH8016	**GCF_000230675**	*Synechococcus*

WH8020	**GCF_001040845**	*Synechococcus*
WH8102	**GCF_000195975**	*Synechococcus*
WH8109	**GCF_000161795**	*Synechococcus*
Och114	**GCA_000014045.1**	*Roseobacter*

Appendix C: Ch.3 Supplemental Figures

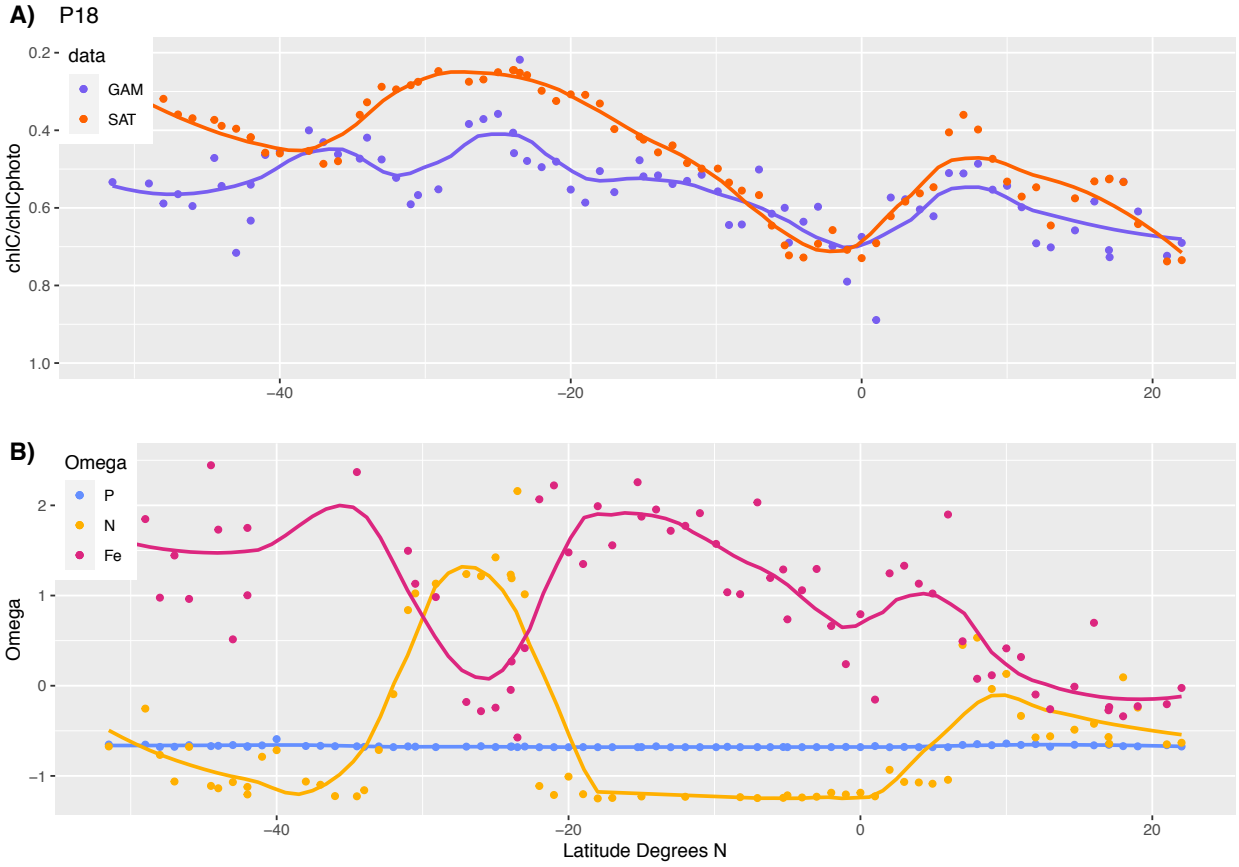

Figure S3.1: P18 Cruse. Sigma observations (SAT) and model predictions (GAM)(A). Metagenomic indices of stress (Omega).

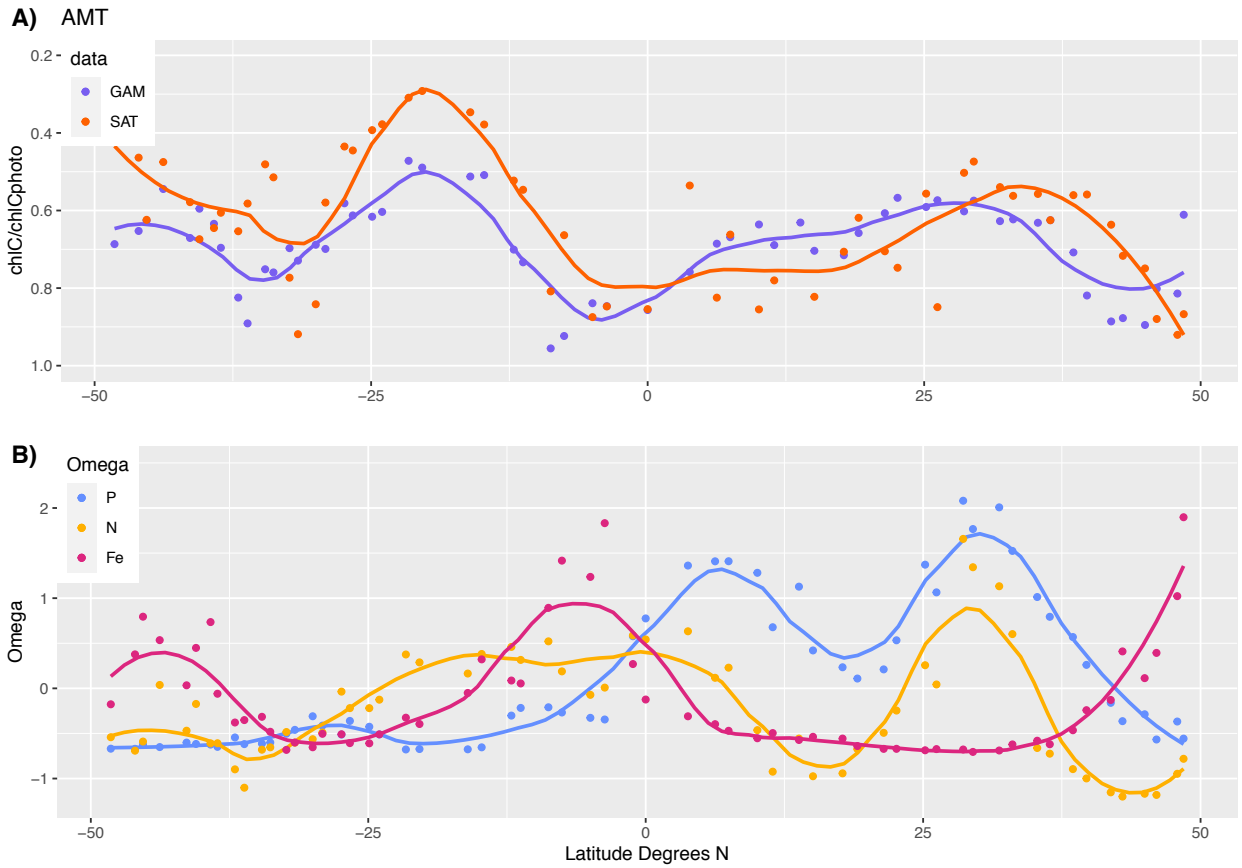

Figure S3.2: AMT Cruse. Sigma observations (SAT) and model predictions (GAM)(A). Metagenomic indices of stress (Omega).

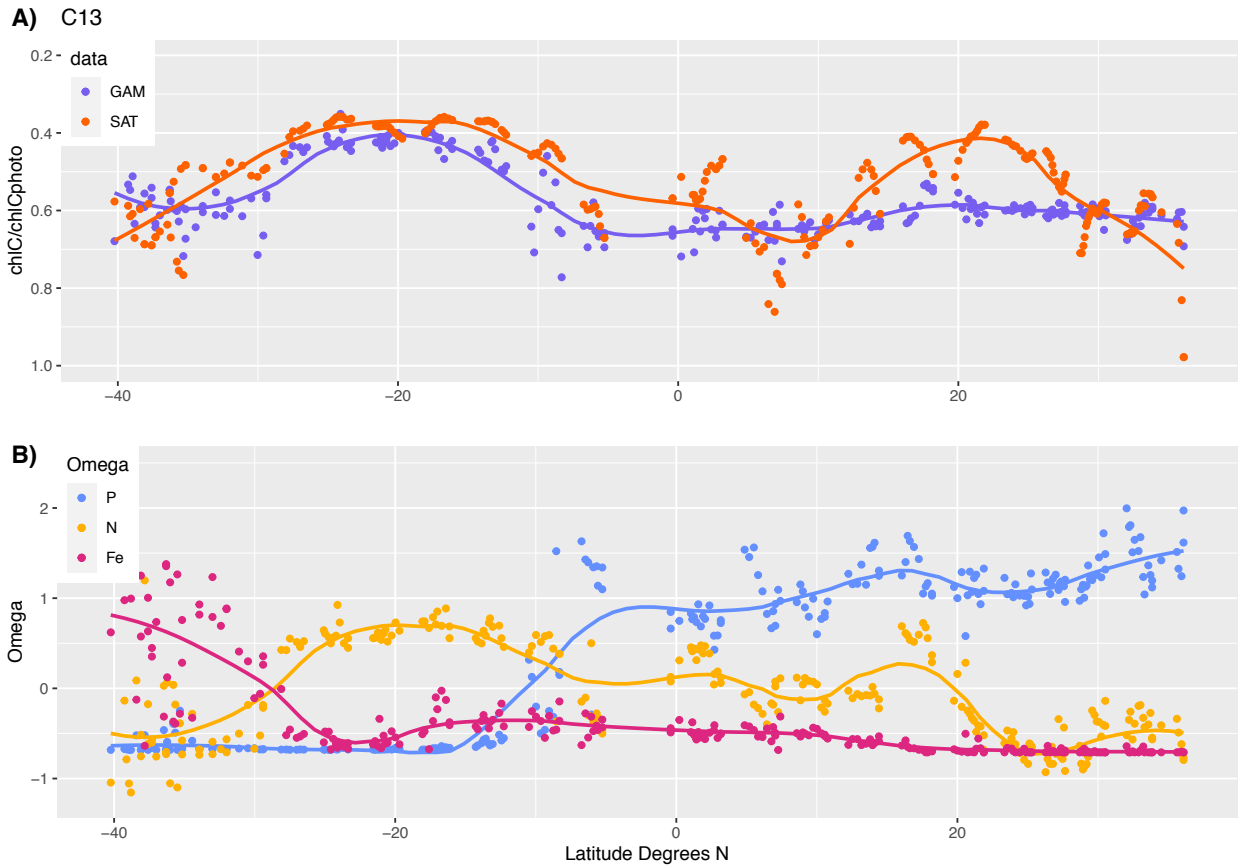

Figure S3.3: C13 Cruse. Sigma observations (SAT) and model predictions (GAM)(A). Metagenomic indices of stress (Omega).

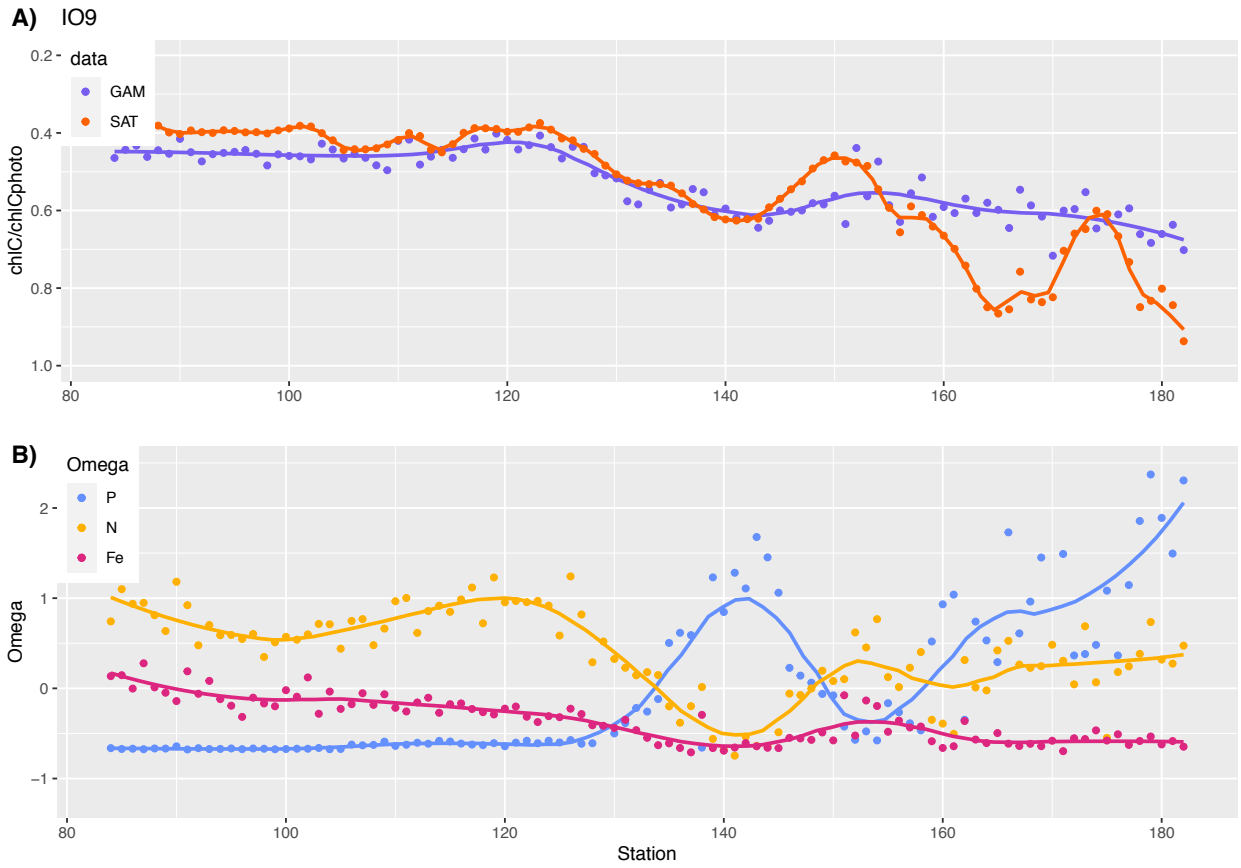

Figure S3.4: IO9 Cruse. Sigma observations (SAT) and model predictions (GAM)(A). Metagenomic indices of stress (Omega).

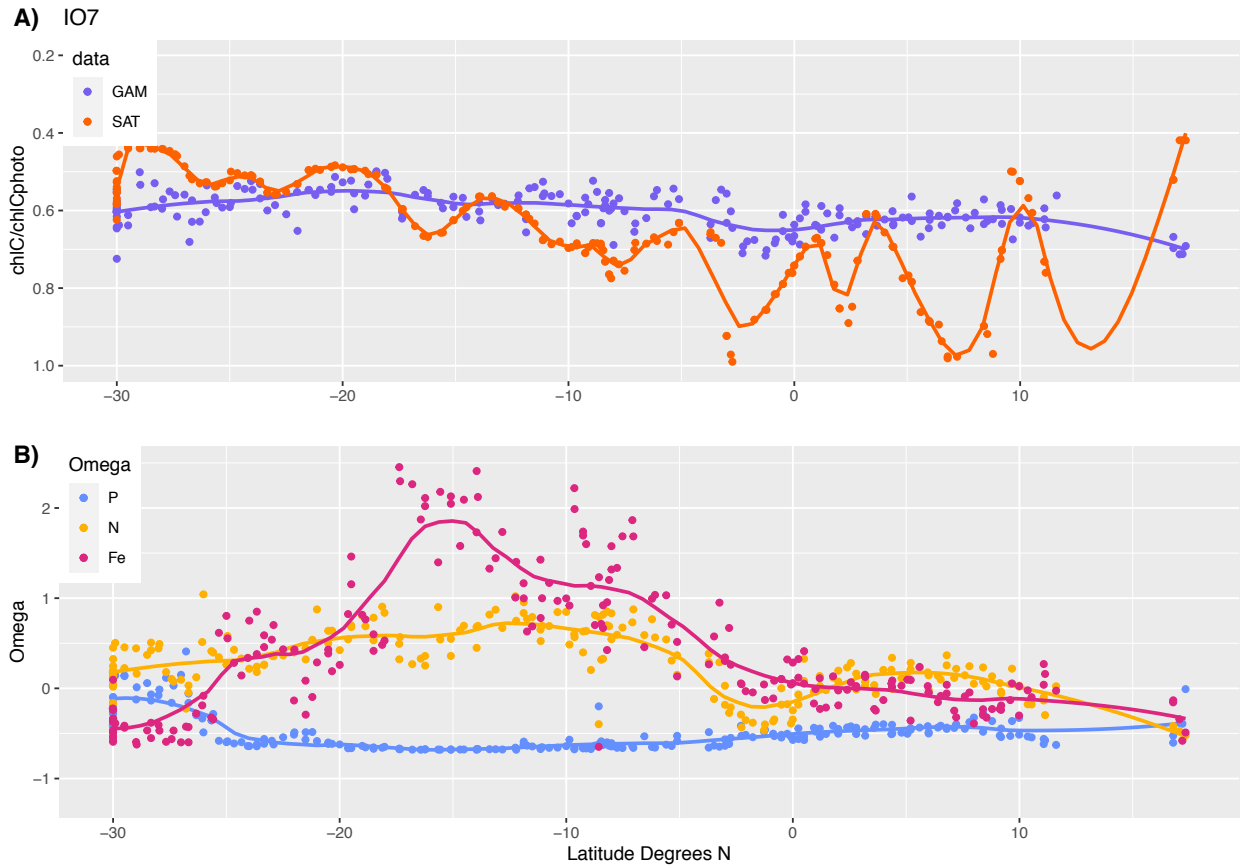

Figure S3.5: IO7 Cruse. Sigma observations (SAT) and model predictions (GAM)(A). Metagenomic indices of stress (Omega).

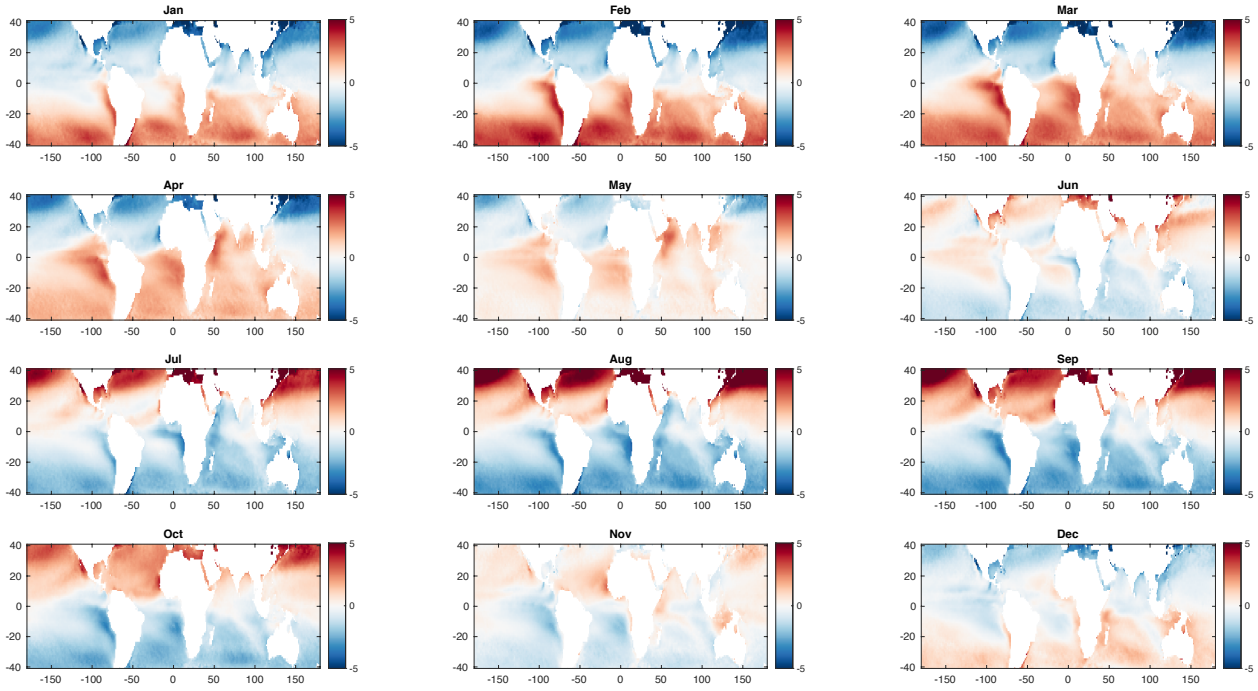

Figure S3.6: Seasonal trends in sea surface temperature.