Subtle biogeochemical regimes in the Indian Ocean revealed by spatial and diel frequency of *Prochlorococcus* haplotypes

Alyse A. Larkin 1, Catherine A. Garcia, Kimberly A. Ingoglia, Nathan S. Garcia, Steven E. Baer 2,3, Benjamin S. Twining, Michael W. Lomas, Adam C. Martiny 1,4*

1Department of Earth System Science, University of California at Irvine, Irvine, California
2Bigelow Laboratory for Ocean Sciences, East Boothbay Harbor, Maine
3Corning School of Ocean Studies, Maine Maritime Academy, Castine, Maine
4Department of Ecology and Evolution, University of California at Irvine, Irvine, California

Abstract

While the majority of studies use the environment to describe microbial populations, the high diversity of microbes can conversely be used as a resource to understand subtle environmental variability. Here, we used a high-resolution spatial and temporal analysis of *Prochlorococcus* sp. in the Eastern Indian Ocean to determine whether ecotypes and microdiverse taxa can be used to identify fine-scale biogeochemical regimes in this under-studied region. A total of 246 DNA samples were collected every 4–6 h in April 2016 on GO-SHIP cruise I09N, which transected gyre, equatorial, and monsoonal ecosystems between Western Australia and the Bay of Bengal. Using amplicon sequencing of the highly variable *rpoC1* marker, we found that the region was largely dominated by the *Prochlorococcus* HL-II clade. Conserved single nucleotide polymorphisms (SNPs) were used to identify four microdiverse haplotypes, or SNP-delineated taxa, within the HL-II clade of *Prochlorococcus*. The haplotypes showed regional patterns of relative gene count abundance that were significantly correlated with environmental conditions. Additionally, we used nonlinear least squares models to fit the sine wave function to our data and demonstrate that the haplotypes show distinct patterns in relative diel frequency, providing evidence that these microdiverse populations are ecologically and evolutionarily distinct. Overall, we show how the integration of a genomics data set into a biogeochemical framework can reveal a more nuanced understanding of a complex ocean basin.

Intricate current patterns and monsoonal dynamics drive complex geographic gradients in physical and chemical conditions in the Indian Ocean (Hood et al. 2009). Despite representing ~15% of the global ocean net primary production (Behrenfeld and Falkowski 1997) and possessing unique biogeochemical characteristics, this region has not been studied as extensively as the Atlantic or Pacific. Examinations of the Eastern Indian Ocean to date have proposed three distinct biomes: (1) a cooler, oligotrophic Southern Indian Ocean (SIO) subtropical gyre, (2) a slightly higher-nutrient, iron-stressed equatorial region with intermittent upwelling (EQ), and (3) a nutrient-replete, high-iron, and high-temperature Bay of Bengal (BOB) (Grand et al. 2015; Garcia et al. 2018). However, subtler biogeochemical regimes with varying degrees of nutrient availability are known to be present (Wiggert et al. 2006). Due to their high diversity, relatively fast growth rates, and physiological differences, microbial populations represent natural biosensors of fine-scale ocean biogeochemistry.

High genomic diversity across phylogenetic scales is a critical characteristic that allows bacterial populations to adapt to a wide range of environmental conditions. Perhaps the best example of a highly diverse and widely distributed microbe is *Prochlorococcus*, a global marine cyanobacterium with phylogenetic clades that are known to partition the ocean based on light, temperature, and iron, often referred to as “ecotypes” (Moore et al. 1998; Bouman et al. 2006; Johnson et al. 2006; Rusch et al. 2010). The term ecotype is traditionally defined as ecologically and phylogenetically distinct bacterial populations that can coexist indefinitely (Cohan 2017). *Prochlorococcus* ecotypes include the low light clades (eLL), the high-light, low-temperature clade (eHL-I); the high-light, high-temperature clade (eHL-II); and, most recently, two high-light but low-iron clades (eHL-III, eHL-IV) (Johnson et al. 2006; Rusch et al. 2010). Previous studies have
also demonstrated the presence of microdiverse *Prochlorococcus* taxonomic groups (i.e., highly similar genomic clades within previously defined ecotypes; Larkin and Martiny 2017), which are partitioned across ocean regions by unique environmental factors (Farrant et al. 2016; Larkin et al. 2016). However, microdiverse partitioning appears to differ between ecotypes, with stronger partitioning in the eHL-I and eLL-I clades than the eHL-II clade (Farrant et al. 2016; Larkin et al. 2016). Thus, there is considerable confusion as to the eco-evolutionary processes that structure these microdiverse groups. Moreover, it is unknown whether they are consistently ecologically distinct, or are merely transient populations formed through stochastic neutral processes such as dispersal limitation and genetic drift (Hanson et al. 2012; McLaren and Callahan 2018).

An increasing body of work has demonstrated the need to examine bacterial communities at high genomic resolution in order to accurately assess taxonomic diversity patterns and ecological differentiation (Acinas et al. 2004; Eren et al. 2015; Tikhonov et al. 2015; Callahan et al. 2017). The use of 16S rRNA as a marker gene was originally selected due to the slow evolutionary rate and associated skill in distinguishing deep phylogenetic clades (Woese 1987). Thus, higher resolution analyses of microbial diversity rely on a small number of 16S rRNA nucleotide variants to assess microdiverse patterns (Eren et al. 2013). Alternatively, fast-evolving genomic markers provide a stronger basis for the identification of microdiverse taxa due to the presence of a higher percentage of variants (Larkin and Martiny 2017; Needham et al. 2017). Here, we use highly conserved single nucleotide polymorphisms (SNPs) in the *rpoC1* gene of *Prochlorococcus* to identify microdiverse haplotypes. The *rpoC1* region was chosen because it has 100% amino acid similarity in sequenced isolates, but is more variable than the internal transcribed spacer region due to high third codon position variability. Furthermore, the lack of indels removes the need for a complex alignment computation and uncertainty in assigning orthologous positions. Similar to the concept of "ecoSNPs" (or "divSNPs"), we argue that neutral differences in the *rpoC1* gene could be indicative of adaptive variance in microdiverse populations (Shapiro and Polz 2015). Additionally, if these microdiverse groups truly represent ecologically differentiated populations, then we expect to see in situ spatial and/or temporal structure in haplotype-associated SNP frequencies that is strongly correlated with previously established biogeochemical features (Cohan 2002; Cohan and Koeppel 2008; Hunt et al. 2008; Wiedenbeck and Cohan 2011).

Using *Prochlorococcus* as a model organism, we ask the question, what can microbial taxonomic diversity tell us about the unique biogeochemistry of the Eastern Indian Ocean? First, we hypothesized that major spatial partitioning of the Indian Ocean would be dominated by ecotype-level changes in community composition, and that temporal shifts in population structure as well as intraregional spatial partitioning would be dominated by microdiverse-level changes. Based on patterns in other oceanographic basins, we hypothesized that the low-nutrient, high-temperature *Prochlorococcus* clade eHL-II would be numerically dominant in the SIO and BOB, due to low macronutrient availability in the SIO and high temperatures in the BOB. Additionally, we expected the high-nutrient, low-chlorophyll (HNLC) clades eHL-III and eHL-IV to increase in relative abundance in the EQ region due to the high macronutrients and low Fe typical of an HNLC region (Johnson et al. 2006; Rusch et al. 2010; Malmstrom et al. 2013). Second, whereas we expected ecotypes to be present in differing relative abundance across biomes, we hypothesized that sub-ecotype microdiverse taxa (i.e., haplotypes) would be unique to each biome (Farrant et al. 2016; Larkin et al. 2016). Third, we hypothesized that temperature, nutrient concentration, and iron-stress would be the primary variables driving geographic partitioning of *Prochlorococcus* within the Indian Ocean.

**Methods**

**Field sampling and environmental variables**

Microbial DNA and environmental metrics were collected aboard the RV *Roger Revelle* on GO-SHIP cruise 109N from 22 March 2016 to 24 April 2016, which traversed from Perth, Australia (31°02'01"S, 110°27'28"E), to the BOB (16°44'15"N, 90°08'77"E) in the Eastern Indian Ocean. GO-SHIP transects conduct a suite of environmental measurements. These data include, but are not limited to, temperature, salinity, and inorganic nutrients (NO$_3^-$, NO$_2^-$, PO$_4^{3-}$, SiO$_2^{4-}$). Measurement of underway temperature and salinity was made with a mounted near-surface thermosalinograph. Nutrient samples were collected from depth using a ship-based conductivity temperature depth (CTD) system via Niskin rosette sampling of the entire water column at approximately every latitudinal degree (for GO-SHIP nutrient analysis protocols, please see https://chcdo.ucsd.edu/; further environmental data are available at https://www.bco-dmo.org/award/628977).

A total of 246 microbial DNA surface-water samples were taken every 4–6 h from the ship’s circulating seawater system, collected at ~7 m depth, distributed via plastic tubing and kept at a constant flow. Replicate 10 L samples were filtered through two 0.22 μm pore size Sterivex filters (Millipore, Darmstadt, Germany) using sterilized tubing and a Masterflex peristaltic pump (Cole-Parmer, Vernon Hills, IL). DNA was preserved with 1620 μL of 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 until analysis.

**Nutrient profiles and satellite measurements**

Nutricline depth was used as a proxy for diapycnal nutrient supply to the mixed layer (Cermeno et al. 2008). Nitrate profiles at each GO-SHIP station were interpolated at 1 m resolution in the top 350 m. The depth at which the nitrate concentration was above a 5 μmol L$^{-1}$ threshold was defined as the nutricline depth. At underway collections between GO-SHIP stations, an average nutricline depth was interpolated from the station before and after the sampling site. If the underway station was outside the GO-SHIP transect (i.e., the first 15 stations in transit from...
Iron (Fe)-stress satellite products were extracted from a data set of 9 km resolution global distributions of quenching-corrected phytoplankton fluorescence quantum yields taken from climatological means from 2003 to 2015. Remotely measured fluorescence quantum yields have been shown to be indicative of physiological variability in phytoplankton communities and are largely influenced by Fe stress (Behrenfeld et al. 2009). This Fe-stress parameter, phi_fe, is estimated from MODIS-Aqua satellite fluorescence data. Global estimates of phi_fe range from 0.0% to 2.8%, with a global average of 1.0% (Behrenfeld et al. 2009).

Sea surface height data for each sampling location was extracted from the Copernicus Marine Environment Monitoring Service (CMEMS) database. Specifically, L4 0.25 degree gridded sea surface heights as reanalyzed by CMEMS from all gridded altimeter missions were used in this analysis.

**rpoC1 amplicon sequencing**

To extract microbial DNA, Sterivex filters were incubated with lysozyme (50 mg mL⁻¹ final concentration) at 37°C for 30 min. Next, Proteinase K (1 mg mL⁻¹) and 10% SDS buffer were added and incubated at 55°C overnight. DNA was precipitated with sodium acetate (245 mg mL⁻¹, pH 5.2) and ice-cold isopropanol (100%), pelleted via centrifuge at 15,000 × g and 4°C for 30 min, and resuspended in TE buffer (10 mmol L⁻¹ Tris-Cl, 1 mmol L⁻¹ EDTA) at 37°C for 1 h. DNA was purified using a Zymo genomic DNA Clean and Concentrator kit (Zymo Research Corp., Irvine, CA). DNA concentration was assessed using a Qubit dsDNA HS Assay and a Qubit fluorometer (ThermoFisher, Waltham, MA) and then diluted to 1 ng μL⁻¹.

The rpoC1 gene was amplified from surface samples using cyanobacteria-specific primers 5’-GARCARATHG TTYATYTIYA-3’ and SAC1039R (5’-CYTGYTTNCCYTCDA TDA TRT-3’) (Kent et al. 2018) with Illumina-specific Nextera transposase adapters. A total of 2 μL of 1 ng μL⁻¹ DNA was added to 20 μL reactions (0.3 μL× L⁻¹ primers, 2.5 units SPrime HotMaster DNA Taq Polymerase [Quantabio, Beverly, MA], MasterAmp 1X Premix F [Epicenter Biotechnologies, Wisconsin, U.S.A.]) and amplified in a two-step polymerase chain reaction (PCR), as follows: 94°C for 2 min, followed by 34 cycles of 94°C 30 s, 48°C 40 s, and 72°C 60 s. Next, 1 μL each of I5 and I7 Nextera v2 barcoded indices (1 ng μL⁻¹) were added to the products and PCR annealed to the amplicon as follows: 10 cycles of 94°C 30 s, 55°C 40 s, and 72°C 60 s, and a final extension step at 72°C for 10 min. Products (with differing barcodes) were pooled and dimers < 700 bp in length were removed using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA). Dimer removal was checked using a 2100 Bioanalyzer high sensitivity DNA trace (Agilent, Santa Clara, CA). Finally, products were sequenced using a MiSeq 300 bp paired-end platform with 600 cycles (Illumina, San Diego, CA) resulting in 14,565,709 reads. Sequence files are available from the NCBI Sequence Read Archive (accession #PRJNA522445).

**Quality filtering and phylogenetic identification**

Due to the length of the rpoC1 gene, forward and reverse reads could not be merged. A quality threshold of 20 was used to truncate the forward read to 294 bp and the reverse read to 222 bp. The reads were then concatenated to a total length of 516 bp. Reads with a mean quality score below 20, or reads with more than 1% of bases below Q20, were removed from the analysis using fastq-mcf (Arnesty 2013). To identify Prochlorococcus sequences, operational taxonomic units (OTUs) were grouped at 97% similarity using the pick_open_reference_otus.py function in QIIME (Caporaso et al. 2010) with a custom database of Prochlorococcus and Synechococcus reference sequences. The reference database contained 41 Prochlorococcus and 15 Synechococcus fully sequenced genomes (Billir et al. 2014) as well as 2 metagenomic assemblages (eHL-III: 97.8% complete, eHL-IV: 97.1% complete) (Kent et al. 2016). Reference sequences were cut to the same length as query sequences. After quality filtering and singleton removal, QIIME grouped 7,511,616 sequences into 478,022 OTUs.

The OTU representative sequences were taxonomically identified (blasto, version 2.6.0, e-value < 10⁻⁵, identity > 80%) by comparison to the Prokaryotic RefSeq genomes database (NCBI). Any sequences with < 80% identity to Prochlorococcus were removed from the analysis. A total of 465,685 OTUs and 7,239,166 sequences were identified as Prochlorococcus. To identify Prochlorococcus ecotypes, reads were then recompared (blastn, e-value < 10⁻⁵, identity > 90%) to the hand-curated and annotated custom database of Prochlorococcus and Synechococcus sequences. A total of 429,918 OTUs could be classified to a Prochlorococcus ecotype, of which the majority were determined to be eHL-II (410,445), eHL-I (15,981), and eLL-I (3,273) (for other ecotypes, see Supporting Information Table S4).

**Identification of microdiverse haplotypes**

Microdiverse haplotypes were identified from eHL-II sequences at all stations. Sequence codons were aligned (MEGA, version 7.0.21) and sequences with frameshift mutations were removed from the analysis. The eHL-II majority sequence at each station was calculated based on the 50% nucleotide consensus at each base pair position using BioPython (version 1.7.0). The eHL-II majority sequence was calculated using all eHL-II sequences. Differences in nucleotide content between the majority sequence and the station-specific consensus sequences were used to identify highly conserved SNPs. Average-linkage hierarchical clustering (“hclust” in the “stats” package, R Core Team 2017) of SNPs was then used to identify haplotype-defining SNPs. The stability of the SNP clusters was assessed via bootstrapping (“clusterboot” in the “fpc” package, R, Hennig 2018). Three basepair positions in the sequence were determined to distinguish between the four largest haplotype clusters with high stability (position 90, 243, and 313). Sequences from each station that matched the four haplotype-specific SNP profiles were identified and the sequence counts of the top four haplotypes were transformed into relative abundance by dividing by the total sequence count. The majority...
**Estimating diel frequency changes**

After identifying what appeared to be diel cycling in relative haplotype frequency, nonlinear least squared (NLS) models were used to fit the sine wave equation to detrended, mean-centered relative rpoC1 gene count abundance data on a 2-day moving window to determine the diel patterns. The sine wave equation was defined as: \( y = C + \alpha \times \sin \left( \frac{2\pi}{24} \times t + \phi \right) \), where \( C \) is the 2-day mean relative gene count abundance (i.e., mean relative sequence count), \( \alpha \) is the amplitude, \( t \) is the time, and \( \phi \) is the offset, or “phase.” There are two components of this analysis that are important to recognize. First, the frequency of each haplotype is interdependent (Fig. 1a). All haplotypes may oscillate in terms of absolute numbers, but sequencing only reveals changes in the relative frequency of each haplotype. Second, whereas the frequency oscillation is synchronized for absolute abundances, the phase (\( \phi \)) of relative haplotype frequencies is variable due to the interdependence of frequency among haplotypes. Thus, we needed to relate changes in each abundance space (i.e., relative and absolute).

The process of estimating absolute gene count trends from relative abundance data was performed as follows (Fig. 1b). All analyses were performed on the R statistical platform (R software, v. 3.4.1, Vienna, Austria). An additive model was first used to detrend the data by performing a moving median decomposition of the relative gene count abundance values. Outliers above or below 4 SDs from the decomposed trend were removed from the analysis (33 stations). Next, an iterative model fitting process was used to produce start values for the relative abundance fit. Harmonic regression and the “nls2” function (Grothendieck 2013) with the “random” algorithm were used to estimate the amplitude (\( \alpha \)) and offset (\( \phi \)) starting parameters. These starting values were then used in the Levenberg-Marquardt NLS fitting algorithm (”nlsLM” function, Elzhov et al. 2016), which fit the sine wave equation (see Supporting Information Fig. S4) to the detrended relative abundance values. To determine whether these fits were significantly different from random, the \( \alpha \) \( p \)-values, the goodness of fit of the regression (R\(^2\) values), and the standard error of the regression (root-mean-square error [RMSE]) were recorded and compared to a random null model. Specifically, normally distributed random “sequence counts” for four “haplotypes” were transformed into relative abundance, detrended, distributed across 35 d at the same resolution as our Indian Ocean data set, and fit with the sinusoidal wave function as described above. This randomization process was repeated 30 times for a total of 1020 diel fits. The distributions of our fitted \( \alpha \) \( p \)-values, R\(^2\) values, and normalized RMSE were compared to the random model values using Welch’s one-tailed, unequal variance t-tests. Finally, to estimate the mean relative gene count abundance of the haplotypes (C, Supporting Information Fig. S4), the same iterative model fitting process was implemented, but with the nondetrended relative abundance...
values. The estimated 2-d \( C \) values as well as the relative gene count abundance \( \alpha \) values were used to estimate the absolute gene count abundance frequency patterns.

To estimate the absolute gene count abundance trends of the four haplotypes, two assumptions were made. First, that the mean absolute gene abundance values would be proportional to the mean relative gene count values \( (C) \) for each haplotype. Therefore, the estimated \( C \) values of the original relative abundance data were used in the absolute gene count abundance estimate. Second, that haplotype cell division was synchronized. Previous work has shown the in situ time of replication for Prochlorococcus to be between 16:00 and 22:00 h (Vaulot and Marie 1999; Worden and Binder 2003; Zinser et al. 2009). Therefore, the “suncalc” function in the “rAtmosphere” package (Biavati 2014) was used to estimate the date- and latitude-dependent time of sunset. Iterative fitting, analysis of \( p \) values, and comparison of model Akaike information criterion (AIC) values (“AICtab” in “bbmle;” Bolker 2017) were used to determine that the best-fit model had a time of replication \( \sim 1.5 \) h after sunset, or \( \sim 19:30–19:50 \) h. Thus, known 2-day \( C \) and \( \phi \) values were used in the absolute abundance fit. The \( \alpha \) values of the absolute haplotype patterns were then estimated via NLS (“nls” function, PORT algorithm, “stats” package, R Core Team 2017) and the relative abundance was calculated, mean-centered, and fit via linear regression to the detrended, mean-centered relative gene count abundance frequency patterns of the haplotypes (i.e., the first fit). Finally, the estimated absolute \( \alpha \) values were used to calculate \( \alpha/C \) for each haplotype, or the relative change in absolute haplotype frequency, hereafter referred to as \( \Delta _{\text{frequency}} \).

Statistical analyses

Both ordination and variance partitioning analyses were performed to assess the environmental associations of the haplotypes. First, canonical correspondence analysis (CCA) of the chi-square transformed haplotype sequence counts was used to assess how environmental parameters influenced the distribution of relative haplotype gene count abundance (“cca,” “vegan” package). Significant environmental parameters were identified using the “ordistep” function (“vegan” package). Relative haplotype gene count abundance trends were fit to the CCA ordination using the “envfit” function (“vegan” package). Significant relationships between haplotype relative gene count abundance and environmental factors were also identified with PERMANOVA (“adonis,” “vegan” package). To assess the amount of variance in relative gene count abundance trends explained by environmental patterns, bootstrapped variance partitioning was performed (“boot.relmp,” “relaimpo” package, Grömping 2006).

Results and discussion

We tested the hypothesis that Prochlorococcus taxonomic diversity across phylogenetic scales identifies biomes in the Indian Ocean. We quantified ecotype-level spatial partitioning across biomes, identified intrabiome microdiverse population patterns, and examined relationships between microdiverse haplotypes and environmental parameters. This was accomplished by analyzing the rpoC1 marker gene from 246 DNA samples on the I09N transect in the Eastern Indian Ocean.

Indian Ocean environmental conditions

Observed environmental trends in the Eastern Indian Ocean were consistent with the three proposed regional biomes (Fig. 2). The SIO gyre from 30°S to 10°S was characterized by surface temperatures that increased from 22°C to 30°C as well as nutricline depths that shoaled from \( \sim 250 \) m at 20°S to \( \sim 70 \) m at 10°S. The nutricline depth remained at 70 m for the rest of the transect except at \( \sim 6–9^\circ \)N, where the nutricline shoaled to depths of \( \sim 20 \) m, indicative of Sri Lankan upwelling (Shetye et al. 1993). Elevated Fe stress is seen until \( \sim 5^\circ \)N near the entrance to the BOB. Average Fe stress along the transect was \( \text{ph}_{\text{HFe}} \sim 1.16\% \), which is slightly \( \sim 5.7\% \) above the global average of 1.0%, given that the global maximum is \( \sim 2.8\% \). This pattern of macronutrient deficiency and Fe stress is consistent with dissolved and particulate Fe measurements taken along the same transect (Garcia et al. 2018; Twining et al. 2019). Across the transect, Prochlorococcus represented \( \sim 78\% \) of the picophytoplankton biomass (Baer et al. 2018; Twining et al. 2019) and ranged in abundance from \( 5.3 \times 10^4 \) to \( 2.1 \times 10^5 \) cells mL\(^{-1} \) (Supporting Information Fig. S1), which is consistent with previous studies that have found Prochlorococcus to be dominant in the region (Jeffries et al. 2015).

Ecotype variation in the Eastern Indian Ocean

Within the Prochlorococcus community, we saw limited ecotype-level partitioning between biomes of the Eastern Indian Ocean during the spring intermonsoon season. At lower temperatures near the coast of Australia, a relatively high percentage of the high-light, low-temperature clade eHL-I was observed (Fig. 2c). The presence of eHL-I south of \( \sim 29^\circ \)S is potentially influenced by the seasonal West Australian Current, which

---

**FIGURE 1** Conceptual model for estimating absolute diel frequencies from relative gene count abundance data. (a) Microdiverse taxa with diel growth cycles can show changes in population structure. If a group of taxa have synchronized diel replication and mortality, as well as the same rate of change in relative gene frequency (amplitude relative to mean gene count, \( \Delta _{\text{frequency}} = \alpha/C \)), then all taxa will have static relative abundance \( (\phi) \) across the diel growth cycle (taxa group 1). However, if \( \Delta _{\text{frequency}} \) varies between taxa, then there will be corresponding diel fluctuations in the relative gene count abundance of those taxa (taxa group 2). (b) The methodological process for fitting the sine wave function \( (\phi) \) to relative gene count abundance data and estimating absolute gene count abundance trends via NLS models is depicted. First, relative gene count abundance data is used to estimate starting parameters for the NLS models. Second, two models are fit to the data. The wave function is first fit to the relative gene count abundance data via NLS. Then, another NLS model is used to estimate absolute gene count abundance data such that, when relative abundance is calculated, it is fit to the first relative abundance model fit. Dashed lines represent data transformation or model fits. Solid lines represent input of data into a model.
brings cooler water northward from the Antarctic Circumpolar Current (Schott et al. 2009), suggesting a lower temperature but nutrient-replete regime off the coast of western Australia. The crossover point between eHL-I and eHL-II occurred at ~ 21 °C and matches the transition temperature between these ecotypes in the North Pacific Ocean in the boreal fall (Larkin et al. 2016). North of this transition point, the HL-II ecotype constituted nearly 100% of the Prochlorococcus community in the surface. This ecotype distribution is supported by previous work, which has shown eHL-II to be dominant in this region (Bouman et al. 2006; Zwirglmaier et al. 2008; Farrant et al. 2016; Kent et al. 2016). The prevalence of eHL-II, the high-temperature, low-nutrient ecotype, suggests a macronutrient limited biogeochemical regime in the Eastern Indian Ocean.

In contrast, only a few sequences (< 1%) associated with the low-iron ecotype (clades HL-III/IV) were detected, suggesting that the EQ biome was not a classical HNLC region. The lack of HNLC ecotypes implies that either Fe stress or macronutrient levels were low during the spring intermonsoonal season. This pattern differs strongly from Fe-stressed equatorial upwelling in the Pacific Ocean (Rusch et al. 2010; West et al. 2011) as well as from previous work finding regionally isolated populations of HNLC ecotypes in the Indian Ocean (Rusch et al. 2010; Farrant et al. 2016; Kent et al. 2016). The prevalence of eHL-II, the high-temperature, low-nutrient ecotype, suggests a macronutrient limited biogeochemical regime in the Eastern Indian Ocean.

In contrast, only a few sequences (< 1%) associated with the low-iron ecotype (clades HL-III/IV) were detected, suggesting that the EQ biome was not a classical HNLC region. The lack of HNLC ecotypes implies that either Fe stress or macronutrient levels were low during the spring intermonsoonal season. This pattern differs strongly from Fe-stressed equatorial upwelling in the Pacific Ocean (Rusch et al. 2010; West et al. 2011) as well as from previous work finding regionally isolated populations of HNLC ecotypes in the Indian Ocean (Rusch et al. 2010; Farrant et al. 2016; Kent et al. 2018). Thus, severe Fe stress may be transient or seasonal in this region (Wiggert et al. 2006; Behrenfeld et al. 2009). Our sequence-based findings are supported by nutrient addition incubation results from the same transect, which found evidence for Fe and N colimitation of planktonic communities but not classical iron limitation (Twining et al. 2019). Overall, the distribution of ecotypes did not support the three-biome hypothesis for the Eastern Indian Ocean.

### Haplotype variation in the Eastern Indian Ocean

We next tested our hypothesis that microdiverse populations would be biome-specific. We focused exclusively on eHL-II populations and identified stable SNP haplotype clusters (Fig. 3a). Clustering of SNP profiles was correlated with OTU patterns (Mantel’s r = 0.24) (Supporting Information Fig. S2). Although we could not identify biome-specific consensus sequences, hierarchical clustering of SNP profiles showed that SNPs in the SIO and BOB clustered together at a slightly higher frequency than with SNPs in the EQ biome (Fig. 3a), indicating environmental partitioning.

Although microdiverse SNP-delineated Prochlorococcus haplotypes were present in all biomes, they showed systematic latitudinal variability in their relative gene count abundance. The top four most abundant haplotypes within eHL-II (HL-II.2 to HL-II.5) represented, on average, 85% ± 3% of the total eHL-II sequences at each station. The HL-II.2 and HL-II.3 haplotypes were the most abundant, composing between 25% and 60% of the eHL-II sequences. The next most abundant haplotype was HL-II.5, with a relative abundance between 8% and 20%, followed by HL-II.4, which varied between 3% and 9%. Large shifts in the relative gene count abundance of these four haplotypes appeared to correspond with the major oceanographic biomes proposed for the Eastern Indian Ocean (Fig. 3b). Broadly, in the SIO, the HL-II.3 haplotype

![Fig. 2. Environmental conditions and relative Prochlorococcus ecotype abundance across the Eastern Indian Ocean. (a) The GO-SHIP I09N latitudinal transect traveled from 31°02'01"S, 110°27'28"E to 16°44'15"N, 90°08'77"E during 22 March 2016–24 April 2016. (b) Temperature (°C), nutricline depth (m), and Fe stress (φ₅₅₅) along the I09N transect reveal three distinct biomes: the subtropical Indian Ocean (SIO) gyre (approximately −30°N to −10°N), the equatorial (EQ) region (approximately −10°N to 5°N), and the BOB (approximately 5°N–20°N). (c) The relative sequence abundance of Prochlorococcus ecotypes reveals that the subtropical and tropical Indian Ocean are dominated by the HL-II ecotype.](image-url)
was dominant, the EQ showed an increased relative abundance of HL-II.5, and the BOB showed large shifts in relative abundance between HL-II.3 and HL-II.2, which varied between 30% and 60% of the eHL-II community. Therefore, in contrast to the ecotype patterns, haplotype relative gene count abundance supports the three-biome hypothesis for the Eastern Indian Ocean.

Environmental associations of haplotypes

The distribution of Prochlorococcus microdiversity across the Indian Ocean was strongly influenced by several environmental factors. Variance partitioning on haplotype relative gene count abundance revealed that nutricline depth (m), temperature (°C), phosphate concentration (μmol L⁻¹), silicate concentration (μmol L⁻¹), Fe stress (φscat), salinity, and sea surface height explained between 22.0% and 62.4% of the variance in haplotype-specific relative gene count abundance (Fig. 4, Supporting Information Table S2, Fig. S3). Each haplotype was explained by a unique combination of environmental metrics. Silicate was strongly associated with HL-II.2. Given the negative correlation between silicate and salinity (Supporting Information Fig. S3), it is possible that silicate is only associated with proximal latitudinal environmental trends. However, closely
related *Synechococcus* populations have been shown to accumulate silicate within the cell (Baines et al. 2012; Ohnemus et al. 2016; Brzezinski et al. 2017). Thus, it is possible that silicate has some effect on *Prochlorococcus* haplotype growth. Only 22% of the variability in HL-II.3 relative abundance was explained by the environmental metrics used in the partitioning model, but increased HL-II.3 sequence counts were associated with higher phosphate concentrations (Supporting Information Fig. S3). Fe stress explained 35% of variance in HL-II.5 abundance, suggesting adaptation to low-Fe conditions (Fig. 4). The HL-II.4 haplotype, which has the lowest abundance, was best explained by sea surface height (which is indicative of zonal wind stress and, thus, shifts between biomes; Fu 2007) and temperature, suggesting that this relatively stable haplotype shows some level of response to large shifts in environmental conditions. Overall, the four haplotypes were related to differing suites of environmental variables, supporting our hypothesis that Fe stress, nutrient concentrations, and temperature would drive the distribution of microdiversity in the Eastern Indian Ocean.

**Diel oscillations**

Although this study was not designed to test diel patterns, we noticed what appeared to be systematic variability in relative haplotype frequency above and below the trend line (Fig. 3b) and decided to quantify this observation through spectral analysis. *Prochlorococcus* cell division is synchronized with the light cycle. Replication occurs near sunset (Vaulot and Marie 1999) and there is a tight coupling between growth and mortality (Ribalet et al. 2015) for *Prochlorococcus*. The outcome is a diel oscillation in cell abundance with no net increase in population size after 24 h. Thus, we hypothesized that a diel oscillation in haplotype frequency could be due to differential growth and loss rates among lineages. However, given that sequencing data can only detect relative changes in sequence abundance, we defined the term $\Delta_{frequency}$ as the absolute amplitude ($a$) over the mean relative gene count abundance ($C$) to describe relative differences in oscillation among each haplotype (i.e., $\Delta_{frequency} = a/C$). To test our hypothesis, we (1) quantified the diel changes in haplotype frequency, and (2) tested if we could derive patterns of differential growth/loss from this oscillation. To accomplish the first aim, a wave function was fit to the detrended, mean-centered relative haplotype frequency (Fig. 1b, Supporting Information Figs. S4, S5). To accomplish the second aim, we used an NLS model to estimate the $\Delta_{frequency}$ patterns of the haplotypes (Fig. 1b, Supporting Information Fig. S4).

Each of the four haplotypes showed some degree of diel oscillations that varied in their $\Delta_{frequency}$ across latitudinal regions. For sinusoidal relative gene count abundance, fits that had significant $\alpha$ parameters ($p$ value $< 0.05$) also tended to have goodness of fit ($R^2$) values $> 0.30$ and normalized standard errors (RMSE) $< 0.25$, which was seen most frequently in the three highest abundance haplotypes (Supporting Information Figs. S5, S6). Based on our random null model analysis, HL-II.2, HL-II.3, and HL-II.5 had sinusoidal fits with a $p$-values that were significantly less than random, HL-II.2 and HL-II.3 had sinusoidal fits with $R^2$ values that were significantly greater than random, and all four haplotypes had sinusoidal fits with normalized RMSE that were significantly less than random. For absolute gene count abundance, estimates of diel patterns also generally had robust fits (Supporting Information Table S1). Given the limitations of the model fitting process, we estimate that ~15–30% of the transect days had significant diel oscillations in one of the four haplotypes. The four haplotypes also showed region-specific changes in their daily mean-centered $\Delta_{frequency}$ across the Indian Ocean (Fig. 5). The $\Delta_{frequency}$ for each haplotype was generally bound within $\pm 0.025$ throughout the majority of the transect, but became more variable in the BOB (Fig. 5a). Additionally, there was systematic variability in terms of which haplotype had the highest $\Delta_{frequency}$ between geographic regions (Fig. 5b). Broadly, haplotypes HL-II.2 and HL-II.3 showed the highest $\Delta_{frequency}$ with intermittent increases in $\Delta_{frequency}$ for haplotypes HL-II.4 and HL-II.5. Thus, there were clear subdivisions in the $\Delta_{frequency}$ among *Prochlorococcus* haplotypes.

**Haplotypes and biogeochemical regimes**

Whereas the ecotype level of spatial partitioning showed the Indian Ocean to be a relatively uniform ecosystem, *Prochlorococcus* microdiversity had more variable biogeography, suggesting subtle ecosystem changes. The relative gene count abundances of the eHL-II haplotypes were used to identify seven biogeochemical regimes in the Eastern Indian Ocean (Fig. 6). The $\Delta_{frequency}$ of the haplotypes was used to discern additional complexity in the region. In particular, an eighth region (Region 6, Fig. 6) where $\Delta_{frequency}$ spikes for HL-II.2 and HL-II.3. Overall, the biogeochemical regimes were defined as follows:

First, a temperate regime was seen south of ~29°S (Fig. 2b). The coastal region to the east of the main GO-SHIP transect had high eHL-I and HL-II.3 relative gene count abundance (Figs. 2c, 3b) and was characterized by colder temperatures (Fig. 2b). Second, a highly oligotrophic regime was identified...
in the southern portion of the SIO gyre approximately between 29°S and 20°S, where the nutricline depth was deepest (Fig. 2b). In this region, HL-II.2 had high relative abundance and either HL-II.2 or HL-II.4 had high $\Delta_{\text{frequency}}$ (Fig. 5b). The HL-II.2 haplotype appears to be associated with decreasing phosphate, increasing silicate, and increasing temperature (Supporting Information Fig. S3). Third, a less oligotrophic regime was seen in the northern portion of the SIO gyre approximately between 20°S and 10°S where the nutricline shoaled. Here, HL-II.3 had high relative abundance and $\Delta_{\text{frequency}}$, possibly linked to increased macronutrient availability (Supporting Information Fig. S3). Fourth, a moderate Fe-stress regime was identified in the southern EQ upwelling biome between 10°S and 0°S where phs$_{\text{sat}}$ rises above the global average of 1.0%. This region was defined by elevated relative abundance and $\Delta_{\text{frequency}}$ of HL-II.5. Fifth, a northern EQ regime between 0°N and 6°N was characterized by decreased Fe stress and higher HL-II.3 abundance and $\Delta_{\text{frequency}}$. Sixth, a regime at ~ 6°-9°N was identified by a very shallow nutricline in a region known to experience upwelling off the coast of Sri Lanka (Shetye et al. 1993; Schott et al. 2002; Garcia et al. 2018). This region showed a spike in relative HL-II.3 $\Delta_{\text{frequency}}$ possibly resulting from an increase in macronutrient availability. Seventh, approximately between 10°N and 16°N, the nutricline depth deepened again indicating the presence of the Intermonsoon Gyre (Garcia et al. 2018). Here, the HL-II.2 haplotype had high relative abundance (Fig. 3b) and $\Delta_{\text{frequency}}$. Eighth, the nutricline shoaled again slightly near the coast north of ~16°N and HL-II.3 increased in relative abundance and $\Delta_{\text{frequency}}$. By integrating environmental patterns of biochemical data with ecotype and haplotype biogeography, a deeper understanding of the Eastern Indian Ocean can be achieved. As a whole, this region experiences high temperatures, generally low but variable macronutrient availability, and moderate Fe stress during the intermonsoon spring. Subtle changes in these three parameters in concurrence with haplotype variability reveal spatial partitioning of biogeochemical regimes.

### Ecological differentiation

Each haplotype was explained by a unique suite of environmental variables and was associated with a unique biogeochemical regime, which supports the hypothesis that haplotypes have distinct eco-evolutionary history (Fig. 4). Specifically, HL-II.2 was common and cycling in more oligotrophic conditions, HL-II.3 was associated with subtle increases in macronutrients (i.e., PO$_4$), HL-II.4 was relatively stable but showed a slight positive relationship with temperature, and HL-II.5 increased with moderate Fe stress. Previous examinations of eHL-II microdiverse
populations suggested that neutral processes (i.e., dispersal, genetic drift, etc.) might influence the distribution of eHL-II microdiverse populations (Farrant et al. 2016; Kent et al. 2016; Larkin et al. 2016). However, two major differences exist between this study and previous work. First, the rpoC1 gene may be well suited to identifying microdiverse populations, as this protein encoding region includes many synonymous mutations. Second, Fe stress was not analyzed in previous studies. The HL-II.5 haplotype, in particular, showed an increase in relative gene count abundance in the high-nutrient supply, low-Fe EQ biome (Fig. 4, Supporting Information Fig. S3). The presence of HL-II.5 but lack of HNLC ecotypes suggests that Fe stress may be more severe in the equatorial Pacific Ocean (Behrenfeld et al. 2009), where HNLC clades eHL-III and eHL-IV dominate (Rusch et al. 2010; Kent et al. 2018). Specifically, while Fe delivery to both the Indian and the Pacific is low, macronutrient delivery to the equatorial Pacific is likely stronger, resulting in a lower Fe to macronutrient supply stoichiometry (Twinning et al. 2011, 2019).

More severe Fe stress may require larger metabolic changes, resulting in more deeply diverged ecotype lineages. Thus, whereas ecotypes may be indicative of large shifts in environmental parameters, the HL-II haplotypes are indicative of unique, but perhaps subtler, multidimensional changes. Specifically, while light, temperature, and Fe stress have been shown to drive niche partitioning for Prochlorococcus ecotypes, studies have so far been unable to link macronutrient patterns to ecotype distributions (Biller et al. 2014; Kent et al. 2016). In contrast, the haplotypes in this study have strong relationships with both phosphate and nitricline depth (as defined by nitrate concentrations). The haplotypes may therefore be particularly useful for identifying patterns of macronutrient stress in the ocean. Overall, the strong association between the haplotypes and the environment, as well as their individual spatial and diel relative abundance patterns, implies niche partitioning of the Indian Ocean by closely related populations.

If microdiverse haplotypes occupy ecologically differentiated niches, spatial or temporal partitioning should fundamentally be driven by differing growth and loss responses. The alternative hypothesis being that there is no difference in growth rates between populations and neutral processes alone control haplotype biogeography (Hellweger et al. 2014). Haplotype Δfrequency, which we interpret as relative net growth, appears to influence the population structure of eHL-II Prochlorococcus (Supporting Information Fig. S7), supporting the hypothesis that the haplotype distribution patterns are, at least partially, non-neutral and that selective processes may be acting on a fine genomic scale. However, in order to detect growth and loss patterns using sequence counts, two conditions must be met. First, the amplitude of diel oscillation relative to the mean sequence count (a/C or Δfrequency) varies between taxa and is reliant on sufficient sequencing depth. In particular, stochasticity in sequencing may reduce our ability to detect oscillations for rare haplotypes. Here, we have a total of 246 samples representing one of the most spatially explicit examinations of open ocean microbial communities, which may partially counteract issues driven by low resolution. However, low or stochastic sequence coverage may affect the low abundance HL-II.4 haplotype and is potentially reflected in its less robust patterns of oscillation (Supporting Information Fig. S6, Table S1). Second, replication must be synchronized across haplotypes (Supporting Information Fig. S4). If replication is not synchronized, we may see erroneous diel cycling in relative sequence abundance that is not associated with differing growth rates. Additionally, the interdependence of sequence counts for each haplotype can result in nonsinusoidal diel patterns for the low abundance haplotypes. However, a large body of research has shown that Prochlorococcus has strongly synchronized diel transcription, replication, growth, and mortality patterns (Vaulot and Marie 1999; Zinser et al. 2009; Ribalet et al. 2015). The final uncertainty associated with Δfrequency is that diel cycling may be disrupted by changing environmental conditions, such as increasing or decreasing nutrient supply. This study was not designed to test diel patterns, and spatial movement across ocean fronts may influence apparent diel cycles. One way to counteract this issue would be to utilize a Lagrangian sampling design. However, poorly constrained oscillations were not associated with changes in water mass as identified by sea surface height and temperature/salinity (Supporting Information Fig. S8). It should be noted that the R² values of the HL-II.3 fits showed a weak relationship (p value = 0.047) with sea surface height, which appears to be driven by a large change in sea surface height on the last day of the transect in the BOB. Therefore, while the final diel fit for HL-II.3 is perhaps driven by water mass rather than diel growth, water mass did not appear to have a significant effect on goodness of fit for the rest of the transect. Genomic estimates of in situ activity are notoriously difficult and each method, such as genome coverage, RNA/DNA ratios, and so on (Koren et al. 2015; Brown et al. 2016; Kirchman 2016), is associated with caveats. Nevertheless, the top three haplotypes showed strong relative diel oscillations on a high percentage of days (Supporting Information Figs. S5, S6) and absolute patterns had high R² values (Supporting Information Table S1). Thus, estimated Δfrequency trends may be a useful technique for estimating in situ growth and spatial partitioning by microdiverse genotypes.

Conclusions

Here, we show that the spatial distribution of Prochlorococcus diversity functions as a high-resolution eco-genomic sensor of biogeochemical regimes in the Indian Ocean. We hypothesized that (1) spatial partitioning of oceanographic biomes would be reflected in the distribution of Prochlorococcus ecotypes, (2) microdiverse populations would be biome-specific, and (3) temperature, nutrient availability, and Fe stress would drive these biogeographic patterns in the Indian Ocean. By combining high-resolution spatial and temporal sampling with genomic analysis of a highly variable marker gene, we demonstrate that
the surface Indian Ocean is largely dominated by a single ecotype (eHL-II). Rather than unimodal, biome-specific partitioning, microdiverse haplotypes coexist throughout the transect, demonstrate complex hierarchical abundance patterns that fluctuate both spatially and temporally, and show subtle intrabiome spatial partitioning. The biogeography of these microdiverse bacterial populations reveals the fine-scale nature of microbial adaptation to changes in environmental conditions, as outlined by eco-evolutionary frameworks such as the Maestro microbe model of microdiversity (Larkin and Martiny 2017). According to the Maestro model, microdiverse populations can form through the accumulation of SNPs or gene duplication events that result in fitness optimization and subtle shifts in growth optima along a trait axis. We show that incremental differences in the relative growth patterns of the haplotypes, as measured by Δfrequency, lead to changes in relative haplotype abundance and influence the population structure of Prochlorococcus eHL-II across the Indian Ocean. Finally, variance in the relative gene count abundance of each microdiverse haplotype is driven by a unique suite of environmental metrics, including macro- and micronutrient availability (i.e., PO₄ and Fe) as well as, to a lesser extent, temperature. Overall, we provide further evidence that subtle partitioning of n-dimensional niche space by microdiverse populations may be a fundamental characteristic of the Prochlorococcus genus. Future research should quantify the gene content, traits, and emergent properties (Hall et al. 2018) associated with these microdiverse haplotypes in order to determine their functional impact on critical ecosystem processes.

References

Acinas, S. G., L. A. Marcelino, V. Klepac-Ceraj, and M. F. Polz. 2004. Divergence and redundancy of 16S rRNA sequences in genomes with multiple rRNA operons. J. Bacteriol. 186: 2629–2635. doi:10.1128/JB.186.9.2629-2635.2004.

Aronesty, E. 2013. Comparison of sequencing utility programs. Open Bioinforma. J. 7: 1–8. doi:10.2174/1875036201307010001.

Baer, S. E., S. Rauschenberg, C. A. Garcia, N. S. Garcia, A. C. Martiny, B. S. Twining, and M. W. Lomas. 2018. Carbon and nitrogen productivity during spring in the oligotrophic Indian Ocean along the GO-SHIP I09N transect. Deep-Sea Res. Part II Top. Stud. Oceanogr. 161: 81–91. doi:10.1016/j.dsr2.2018.11.008.

Baines, S. B., B. S. Twining, M. A. Brzezinski, J. Krause, S. Vogt, D. Assael, and H. McDaniel. 2012. Significant silicon accumulation by marine picocyanobacteria. Nat. Geosci. 5: 886–891. doi:10.1038/NGEO1641.

Behrenfeld, M. J., and P. G. Falkowski. 1997. A consumer’s guide to phytoplankton primary productivity models. Limnol. Oceanogr. 42: 1479–1491. doi:10.4319/lo.1997.42.7.1479.

Behrenfeld, M. J., and others. 2009. Satellite-detected fluorescence reveals global physiology of ocean phytoplankton. Biogeosciences 6: 779–794. doi:10.5194/bg-6-779-2009.

Biavati, G. 2014. RAtmosphere: Standard atmospheric profiles. R package version 1.1. https://CRAN.R-project.org/package=RAtmosphere.

Biller, S. J., and others. 2014. Genomes of diverse isolates of the marine cyanobacterium Prochlorococcus. Sci. Data 1: 140034. doi:10.1038/sdata.2014.34.

Bolker, B., and R Core Team. 2017. bbml: Tools for general maximum likelihood estimation. R package version 1.0.20. https://CRAN.R-project.org/package=bbml.

Bouman, H. A., and others. 2006. Oceanographic basis of the global surface distribution of Prochlorococcus ecotypes. Science 312: 918–921. doi:10.1126/science.1122692.

Brown, C. T., M. R. Olm, B. C. Thomas, and J. F. Banfield. 2016. Measurement of bacterial replication rates in microbial communities. Nat. Biotechnol. 34: 1256–1263. doi:10.1038/nbt.3704.

Brzezinski, M. A., J. W. Krause, S. B. Baines, J. L. Collier, D. C. Ohnemus, and B. S. Twining. 2017. Patterns and regulation of silicon accumulation in Synechococcus spp. J. Phycol. 53: 746–761. doi:10.1111/jpy.12545.

Callahan, B. J., P. J. McMurdie, and S. P. Holmes. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 11: 2639–2643. doi:10.1038/ismej.2017.119.

Caporaso, J. G., and others. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7: 335–336. doi:10.1038/nmeth.f.303.

Cermeno, P., S. Dutkiewicz, R. P. Harris, M. Follows, O. Schofield, and P. G. Falkowski. 2008. The role of nutricline depth in regulating the ocean carbon cycle. Proc. Natl. Acad. Sci. USA 105: 20344–20349. doi:10.1073/pnas.01001454.

Cohan, F. M. 2002. What are bacterial species? Annu. Rev. Microbiol. 56: 457–487. doi:10.1146/annurev.micro.56.012302.160634.

Cohan, F. M. 2017. Transmission in the origins of bacterial diversity, from ecotypes to phyla. Microbiol. Spectr. 5: 14–40.

Cohan, F. M., and A. F. Koeppel. 2008. The origins of ecological diversity in prokaryotes. Curr. Biol. 18: R1024–R1034. doi:10.1016/j.cub.2008.09.014.

Elzhov, T. V., K. M. Mullen, A.-N. Spiess, and B. Bolker. 2016. minpack.lm: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, Plus support for bounds.

Eren, A. M., L. Maignien, W. J. Sul, L. G. Murphy, S. L. Grim, H. G. Morrison, and M. L. Sogin. 2013. Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol. Evol. 4: 1111–1119. doi:10.1111/2041-210X.12114.
Hood, R. R., J. D. Wiggert, and S. W. A. Naqvi. 2009. Indian Ocean research: Opportunities and challenges, p. 409–428. In Indian Ocean biogeochemical processes and ecological variability, Geophys. Monogr. Ser., vol. 185, edited by J. D. Wiggert et al., AGU, Washington, D. C.

Hunt, D. E., L. A. David, D. Gevers, S. P. Preheim, E. J. Alm, and M. F. Polz. 2008. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science 320: 1081–1085. doi:10.1126/science.1157890.

Jeffries, T. C., and others. 2015. Spatially extensive microbial biogeography of the Indian Ocean provides insights into the unique community structure of a pristine coral atoll. Sci. Rep. 5: 15383. doi:10.1038/srep15383.

Johnson, Z. I., E. R. Zinser, A. Coe, N. P. McNulty, E. M. S. Woodward, and S. W. Chisholm. 2006. Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients. Science 311: 1737–1740. doi:10.1126/science.1118052.

Kent, A. G., C. L. Dupont, S. Yooshef, and A. C. Martiny. 2016. Global biogeography of Prochlorococcus genome diversity in the surface ocean. ISME J. 10: 1856–1865. doi:10.1038/ismej.2015.265.

Kent, A. G., S. E. Baer, C. Mouginot, J. S. Huang, A. A. Larkin, M. W. Lomas, and A. C. Martiny. 2018. Parallel phylogeography of Prochlorococcus and Synechococcus. ISME J. 13: 430–441. doi:10.1038/s41396-018-0287-6.

Kirchman, D. L. 2016. Growth rates of microbes in the oceans. Ann. Rev. Mar. Sci. 8: 285–309. doi:10.1146/annurev-marine-122414-033938.

Korem, T., and others. 2015. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. Science 349: 1101–1106. doi:10.1126/science.aac4812.

Larkin, A. A., S. K. Blinebry, C. Howes, J. Chandler, E. R. Zinser, and Z. I. Johnson. 2016. Niche partitioning and biogeography of high light adapted Prochlorococcus across taxonomic ranks in the North Pacific. ISME J. 10: 1555–1567. doi:10.1038/ismej.2015.244.

Larkin, A. A., and A. C. Martiny. 2017. Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. Environ. Microbiol. Rep. 9: 55–70. doi:10.1111/1758-2229.12523.

Malmstrom, R. R., and others. 2013. Ecology of uncultured Prochlorococcus clades revealed through single-cell genomics and biogeographic analysis. ISME J. 7: 184–198. doi:10.1038/ismej.2012.89.

McLaren, M. R., and B. J. Callahan. 2018. In nature, there is only diversity. MBio 9: e02149–e02117. doi:10.1128/mBio.02149–17.

Moore, L. R., G. Rocap, and S. W. Chisholm. 1998. Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes. Nature 393: 464–467. doi:10.1038/30965.

Needham, D. M., R. Sachdeva, and J. A. Fuhrman. 2017. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microbiality matters. ISME J. 11: 1614–1629. doi:10.1038/ismej.2017.29.

Ohnemus, D. C., S. Rauschenberg, J. W. Krause, M. A. Brzezinski, J. L. Collier, S. Geraci-Yee, S. B. Baines, and B. S. Twining. 2016. Silicon content of individual cells of Synechococcus from the North Atlantic Ocean. Mar. Chem. 187: 16–24. doi:10.1016/j.marchem.2016.10.003.

Eren, A. M., H. G. Morrison, P. J. Lescault, J. Reveillaud, J. H. Vineis, and M. L. Sogin. 2015. Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. ISME J. 9: 968–979. doi:10.1038/ismej.2014.195.

Farrant, G. K., and others. 2016. Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. Proc. Natl. Acad. Sci. USA 113: E3365–E3374. doi:10.1073/pnas.1524865113.

Fu, L. L. 2007. Intraseasonal variability of the Equatorial Indian Ocean observed from sea surface height, wind, and temperature data. J. Phys. Oceanogr. 37: 188–202. doi:10.1175/JPO3006.1.
R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ribalet, F., and others. 2015. Light-driven synchrony of Prochlorococcus growth and mortality in the subtropical Pacific gyre. Proc. Natl. Acad. Sci. USA 112: 8008–8012. doi:10.1073/pnas.1424279112

Rusch, D. B., A. C. Martiny, C. L. Dupont, A. L. Halpern, and J. C. Venter. 2010. Characterization of Prochlorococcus clades from iron-depleted oceanic regions. Proc. Natl. Acad. Sci. USA 107: 16184–16189. doi:10.1073/pnas.1009513107.

Schott, F. A., and others. 2002. The shallow overturning circulation of the Indian Ocean. Prog. Oceanogr. 53: 57–103. doi:10.1016/S0079-6611(02)00039-3.

Schott, F. A., S. P. Xie, and J. P. McCreary. 2009. Indian Ocean circulation and climate variability. Rev. Geophys. 47: RG1002. doi:10.1029/2007RG000245.

Shapiro, B. J., and M. F. Polz. 2015. Microbial speciation. Cold Spring Harb. Perspect. Biol. 7: a018143. doi:10.1101/cshperspect.a018143.

Shetye, S. R., A. D. Gouveia, S. S. C. Shenoi, D. Sundar, G. S. Michael, and G. Nampoothiri. 1993. The western boundary current of the seasonal subtropical gyre in the Bay of Bengal. J. Geophys. Res. Oceans 98: 945–954. doi:10.1029/92JC02070.

Tikhonov, M., R. W. Leach, and N. S. Wingreen. 2015. Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. ISME J. 9: 68–80. doi:10.1038/ismej.2014.117.

Twining, B. S., S. B. Baines, J. B. Bozard, S. Vogt, E. A. Walker, and D. M. Nelson. 2011. Metal quotas of plankton in the equatorial Pacific Ocean. Deep-Sea Res. Part II Top. Stud. Oceanogr. 58: 325–334. doi:10.1016/j.dsr2.2010.08.018.

Twining, B. S., S. Rauschenberg, S. E. Baer, M. W. Lomas, A. C. Martiny, and O. Antipova. 2019. A nutrient limitation mosaic in the eastern tropical Indian Ocean. Deep-Sea Res. Part II Top. Stud. Oceanogr. doi:10.1016/j.dsr2.2019.05.001.

Vaulot, D., and D. Marie. 1999. Diel variability of photosynthetic picoplankton in the equatorial Pacific. J. Geophys. Res. Oceans 104: 3297–3310. doi:10.1029/98JC01333.

West, N. J., P. Lebaron, P. G. Strutton, and M. T. Suzuki. 2011. A novel clade of Prochlorococcus found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. ISME J. 5: 933–944. doi:10.1038/ismej.2010.186.

Wiedenbeck, J., and F. M. Cohan. 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol. Rev. 35: 957–976. doi:10.1111/j.1574-6976.2011.00292.x.

Wiggert, J. D., R. G. Murtaguddde, and J. R. Christian. 2006. Annual ecosystem variability in the tropical Indian Ocean: Results of a coupled bio-physical ocean general circulation model. Deep-Sea Res. Part II Top. Stud. Oceanogr. 53: 644–676. doi:10.1016/j.dsr2.2006.01.027.

Woese, C. R. 1987. Bacterial evolution. Microbiol. Rev. 51: 221–271.

Worden, A. Z., and B. J. Binder. 2003. Application of dilution experiments for measuring growth and mortality rates among Prochlorococcus and Synechococcus populations in oligotrophic environments. Aquat. Microb. Ecol. 30: 159–174. doi:10.3354/ame030159.

Zinser, E. R., and others. 2009. Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, Prochlorococcus. PLoS One 4: e5135. doi:10.1371/journal.pone.0005135.

Zwirglmaier, K., and others. 2008. Global phylogeography of marine Synechococcus and Prochlorococcus reveals a distinct partitioning of lineages among oceanic biomes. Environ. Microbiol. 10: 147–161. doi:10.1111/j.1462-2920.2007.01440.x.

Acknowledgments
The authors would like to thank the captain and crew of the R/V Roger Revelle, chief scientists L. Barbero and C. Rodriguez, and all scientific members of the I09N transect. We would also like to thank members of the Lomas and Martiny labs for their assistance with both the data and manuscript presented here. Funding was provided by NSF (OCE-1559002 to ACM and OCE-1559021 to MWL and BST).

Conflict of Interest
None declared.

Submitted 28 January 2019
Revised 19 April 2019 and 17 June 2019
Accepted 19 June 2019
Associate editor: David Walsh