



Parallel phylogeography of *Prochlorococcus* and *Synechococcus*

Alyssa G. Kent¹ · Steven E. Baer² · Céline Mouginot³ · Jeremy S. Huang¹ · Alyse A. Larkin³ · Michael W. Lomas² · Adam C. Martiny^{1,3}

Received: 2 April 2018 / Revised: 11 September 2018 / Accepted: 14 September 2018 / Published online: 3 October 2018
© International Society for Microbial Ecology 2018

Abstract

The globally abundant marine Cyanobacteria *Prochlorococcus* and *Synechococcus* share many physiological traits but presumably have different evolutionary histories and associated phylogeography. In *Prochlorococcus*, there is a clear phylogenetic hierarchy of ecotypes, whereas multiple *Synechococcus* clades have overlapping physiologies and environmental distributions. However, microbial traits are associated with different phylogenetic depths. Using this principle, we reclassified diversity at different phylogenetic levels and compared the phylogeography. We sequenced the genetic diversity of *Prochlorococcus* and *Synechococcus* from 339 samples across the tropical Pacific Ocean and North Atlantic Ocean using a highly variable phylogenetic marker gene (*rpoC1*). We observed clear parallel niche distributions of ecotypes leading to high Pianka's Index values driven by distinct shifts at two transition points. The first transition point at 6° N distinguished ecotypes adapted to warm waters but separated by macronutrient content. At 39°N, ecotypes adapted to warm, low macronutrient vs. colder, high macronutrient waters shifted. Finally, we detected parallel vertical and regional single-nucleotide polymorphism microdiversity within clades from both *Prochlorococcus* and *Synechococcus*, suggesting uniquely adapted populations at very specific depths, as well as between the Atlantic and Pacific Oceans. Overall, this study demonstrates that *Prochlorococcus* and *Synechococcus* have shared phylogenetic organization of traits and associated phylogeography.

Introduction

The marine Cyanobacteria *Prochlorococcus* and *Synechococcus* are globally abundant and combined, account for approximately 25% of ocean primary production [1]. *Prochlorococcus* is most abundant at ocean temperatures above 20 °C, whereas *Synechococcus* has a wider habitat range but with a maximum abundance near 10 °C [1]. However, the

two lineages co-occur across broad environmental gradients in subtropical and tropical waters (from ~40°S to 40°N).

The evolutionary diversification of *Prochlorococcus* has been closely associated with environmental factors (Fig. 1). At the basal phylogenetic level, *Prochlorococcus* is broadly divided into two ecotypes: low- and high-light (LL and HL)-adapted clades [2]. The LL-adapted clades are found at increasing depth with LLI located near the nutricline, LLII + III right below and LLIV at the bottom of euphotic zone [3, 4]. There are also additional LL-adapted lineages but the phylogenetic position and depth distribution of these are less clear [5]. The HL group can be divided into low-iron (named HLLIII/IV) and high-iron-adapted clades [6–9] with the high-iron clade further divided into low- (HLI) and high-temperature (HLII) clades [10]. Thus, there is a clear association between phylogeny and environmental niches of ecotypes in *Prochlorococcus*.

The biogeography of *Synechococcus* genetic diversity is less clear. The current notion is that multiple clades (I, IV, and perhaps several others) are most abundant in colder, nutrient-rich waters. Similarly, several clades (clades II, III, V, VI, and VII) are frequent but variable in tropical and subtropical waters [11–15]. Finally, the CRD1 and envB/C

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41396-018-0287-6>) contains supplementary material, which is available to authorized users.

✉ Adam C. Martiny
amartiny@uci.edu

¹ Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA

² Bigelow Laboratory for Ocean Sciences, East Boothbay, ME 04544, USA

³ Department of Earth System Science, University of California, Irvine, CA 92697, USA

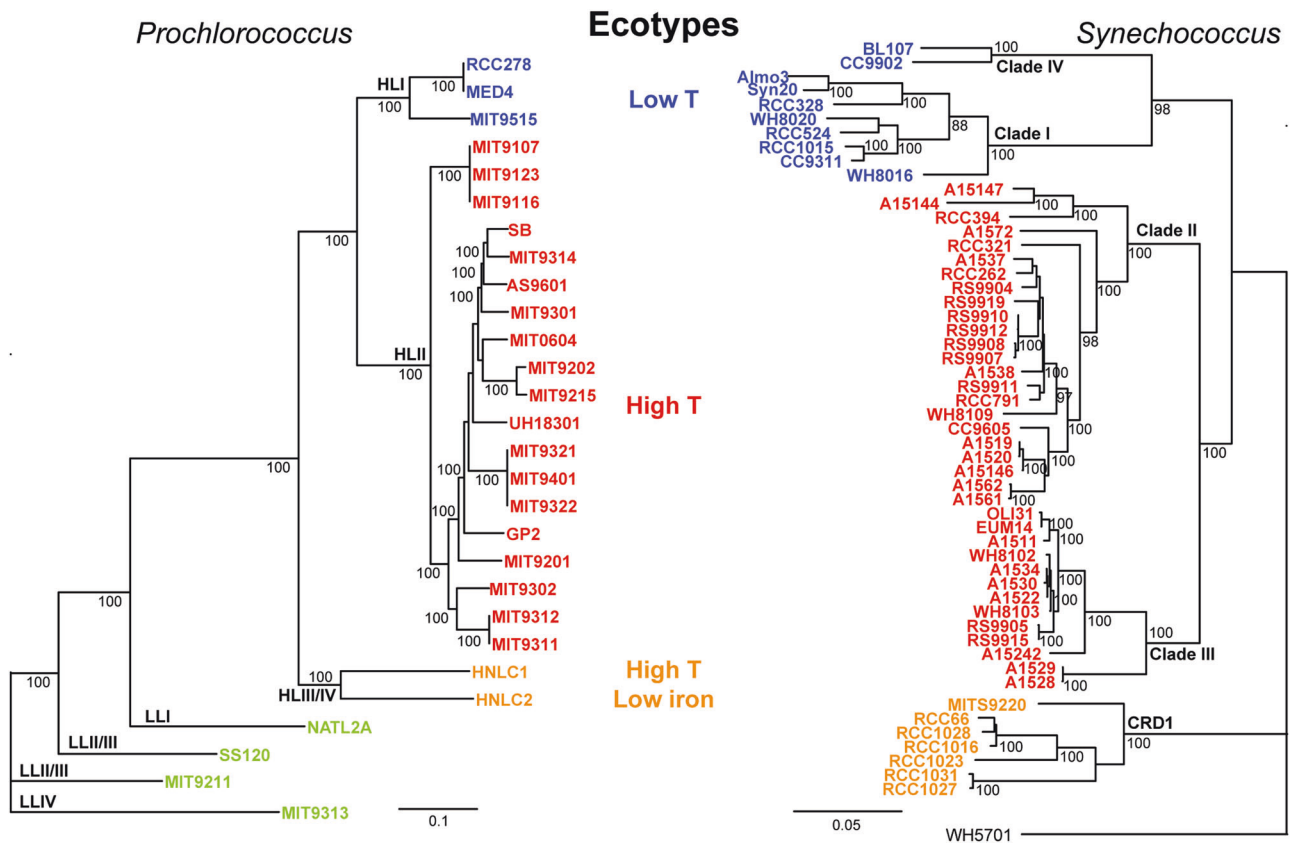


Fig. 1 Similar phylogenetic structure of *Prochlorococcus* and *Synechococcus* clades common in open ocean surface waters. Maximum likelihood phylogenetic trees based on previous constructions [23, 43]. Strains are colored based on their clade and clade names are at the base

clades appears to dominate low iron upwelling zones [16, 17]. However, the use of different marker genes and approaches makes the biogeography of *Synechococcus* unclear. The biology of *Prochlorococcus* and *Synechococcus* largely overlap with respect to many biological facets and only really differs in the composition of the light-harvesting system and cell size [18]. The two lineages are both unicellular, derive energy from photosynthesis, and use the same nutrient sources [19, 20]. They also have small genomes but with extensive genetic and phylogenetic microdiversity [21]. Thus, it is surprising that the biogeography of *Prochlorococcus* is characterized by a single ecotype dominating a specific environment, whereas *Synechococcus* has overlapping distributions of clades with similar biology.

To address this disparity in apparent evolutionary history, we revisited the phylogenetic organization of surface ocean diversity of *Prochlorococcus* and *Synechococcus* (Fig. 1). Unless directly associated with a phenotype, clade delineations are arbitrary and microbial traits are commonly associated with different phylogenetic depths [22]. If we shift the phylogenetic depth for defining clades within *Synechococcus*, clade I/IV or clade II/III emerge as

of each group. Metagenomic assembled genomes HNL1 and HNL2 are from Rusch et al. [6]. Hypothesized ecotype labels are listed in the center and tested in this study (color figure online)

monophyletic clusters [23]. This new view motivates the current study and suggests a parallel phylogenetic structure for *Prochlorococcus* and *Synechococcus* diversity—at least for the major surface clades. As a result, we have two hypotheses. First, we predict that known *Prochlorococcus* ecotypes found in the surface waters (i.e., HLI, HLII, and HLIII/IV) share niche distributions with these newly defined *Synechococcus* clades. Second, *Prochlorococcus* and *Synechococcus* and their associated ecotypes harbor extensive microdiversity, which we hypothesize, is also organized in parallel. Thus, the phylogenetic and regional distribution of *Prochlorococcus* and *Synechococcus* diversity may be organized in a symmetric fashion.

To test this set of hypotheses, we quantified the phylogeography of *Prochlorococcus* and *Synechococcus* over broad open ocean environmental gradients using high-throughput sequencing of a variable phylogenetic marker gene (*rpoC1*) that covers both lineages. This marker gene is single-copy, encodes for the gamma subunit of RNA polymerase, which is distinctive in cyanobacteria [24], and has been used previously to assess *Prochlorococcus* and *Synechococcus* diversity [16, 25]. Using flow cytometry and this marker, we can phylogenetically scale from each

lineage through major clades to within-clade microdiversity. In doing so, we will test the degree to which the biogeography is shared across phylogenetic scales of these two important marine phytoplankton.

Methods

Samples were collected during three cruises: the mid North Atlantic Ocean cruise Bval46 from 28 September to 12 October 2011, the North Atlantic Ocean cruise AE1319 from 15 August to 8 September 2013, and the tropical Pacific Ocean Cruise NH1418 from 20 September to 6 October 2014. See Supplementary Table S1 for cruise information and sample data.

Nutrient and cell abundance measurements

Nutrient samples were collected after filtration through 0.8 μm Nucleopore polycarbonate filters (Whatman, Maidstone, UK). Soluble reactive phosphorus was determined using high-temperature acid persulfate oxidation on a Genesys 10UV spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) after preparation via the magnesium-induced co-precipitation method [26, 27]. For cell counts, samples of whole seawater were collected in 2 mL centrifuge tubes. Freshly made 0.2 μm -filtered paraformaldehyde (0.5% v/v final concentration) was added to all samples and allowed to fix for at least 1 h at 5 °C in the dark, after which they were stored at –80 °C until analysis. Cell counts were performed on a FACSJazz or Influx flow cytometer (BD, Franklin Lakes, NJ, USA) utilizing a 200 mW 488 nm laser, with detectors for forward scatter, side scatter, 530 nm, and 692 nm. Instrument alignment was performed with 3.0 μm six-peak rainbow beads, while hourly checks on forward scatter response were performed with 0.53 μm Nile Red beads (Spherotech, Lake Forest, IL, USA). *Prochlorococcus* populations were discriminated based on forward scatter and red fluorescence, and a gate in orange (585 nm) discriminated for *Synechococcus*. Eukaryotes were all the large red autofluorescing cells that did not fit the Cyanobacteria gating scheme.

DNA extraction

Four liters of seawater were prefiltered on a GF/D, 2.7 μm glass-fiber filter before being collected on a 0.22 μm Sterivex filter (Millipore, Burlington, MA, USA). In all, 1.62 mL TES buffer (50 mM Tris-HCl pH 7.6, 20 mM EDTA pH 8.0, 400 mM NaCl, 0.75 M sucrose) was added before freezing at –20 °C until further processing. DNA was extracted following Bostrom and co-workers [28]. Filters were thawed, 180 μl of lysozyme buffer (50 mg/ml) added,

and then incubated at 37 °C for 30 min. After adding 180 μl of proteinase K (1 mg/ml) and 100 μl of 10% sodium dodecyl sulfate, the filters were incubated at 55 °C overnight. The filter liquid was removed and combined with 3 M sodium acetate (pH 5.2) and cold isopropanol and incubated at –20 °C for > 1 h. After centrifugation at 15,000 g for 30 min at 4 °C, the supernatant was removed and the pellet resuspended in Tris-EDTA buffer (10 mM Tris pH 8.0, 1 mM EDTA) in a 37 °C water bath for 30 min. Finally, DNA was purified using a genomic DNA Clean and Concentrator kit (Zymo Corp., Irvine, CA, USA) and stored at –20 °C.

PCR amplification and sequencing

DNA concentration was quantified with Qubit dsDNA HS assay kit (Life Technologies, Carlsbad, CA, USA) and subsequently diluted in Tris buffer (10 mM, pH 8.0) to a concentration of 1 ng/ μl . We modified *rpoC1* primers to target to ensure they targeted all known isolates of marine *Prochlorococcus* and *Synechococcus* lineages plus metagenomic assemblies from the HLIII/IV group [29] resulting in the primers 5M_newF (5'-GARCARATHGTYTAY TTYA-3') and SACR1039R (5'-CYTGYTTNCCYTCDATDATRT-3'). This region was also chosen such that reference sequences had no indels and thus enabled alignment-free merging of non-overlapping regions for analysis. In all, 2 μl of 1 ng/ μl DNA was added to a PCR cocktail of 1 μl each of 6 μM primer, 10 μl Premix F (Epicentre, Madison, WI, USA), and 0.5 μl Taq polymerase (Hotmaster Taq polymerase, 5 PRIME, Hamburg, Germany), and 5.5 μl nuclease-free water. After 35 cycles, 2 μl barcode oligonucleotides (Eurofins MWG Operon, Louisville, KY, USA) following Illumina NexteraXT index sequences were added to each sample for another 10 PCR cycles (Illumina, San Diego, CA, USA). PCR products were verified on an agarose gel and 1–2 μl of each sample were pooled for cleanup using Agencourt AMPure XP beads (Beckman-Coulter Genomics, Danvers, MA, USA). Final products were sequenced by paired-end 300 bp on a MiSeq. Sequences are stored in the Genbank database under bio-project PRJNA489862 under accessions SAMN09990737–SAMN09991321.

Data analysis

Sequences were trimmed to the same length to maintain an average quality score > 20 using FASTQC [30]. This reduced the second read by 81 bp leaving 519 bp for sequence analysis. The reverse complement of trimmed read 2 was concatenated with read 1. Concatenated sequences were demultiplexed and quality filtered using `split_libraries_fastq.py` script with default settings from QIIME1.9 [31]. The top 100 references were chosen using

pick_open_reference_otus.py from QIIME1.9. All sequences were searched against a custom database of known *Prochlorococcus* and *Synechococcus rpoC1* sequences using tblastx [32]. Best hits were identified by best e-value. Due to the overabundance of *Prochlorococcus*, we added additional *Synechococcus* sequences using pick_open_reference_otus.py and added the top 100 references to the combined reference database to rerun tblastx on all sequences. Sequences mapping to *Synechococcus* also had to have % Guanine-Cytosine (GC) > 41.2, which was 5 standard deviations below the lowest average % GC from any *Synechococcus* clade and sequences mapping to *Prochlorococcus* had to have % GC < 49.1, which was 5 standard deviations above the highest average % GC from any *Prochlorococcus* clade. Sequences with < 90% amino-acid identity and 90% coverage were filtered out. Different percent identity thresholds were compared with one another using Mantel test (Supplementary Figure S1) and 90% identity was chosen to minimize sequence sample noise. Sequences that passed these filters were associated with known clades based on the phylogeny of all reference sequences using raw sequence distance and clustering [33]. This clustering also agreed with a maximum likelihood phylogeny (Supplementary Figure S2) [34]. Taxonomic abundance and environmental data were interpolated using the DIVA algorithm in Ocean Data View [35].

We specifically used *rpoC1* as marker because protein sequences likely contain higher variation (due to variable third codon positions) than more conserved non-translated markers like 16S ribosomal RNA or the internal transcribed spacer (ITS) region. Sequenced isolates of *Prochlorococcus* and *Synechococcus* also demonstrate 100% *rpoC1* amino-acid similarity, which facilitates an alignment-free approach. As a neutral spacer region, ITS has a higher rate of indels than *rpoC1*, necessitating manual curation of sequence alignments [9, 36]. Thus, whether an operational taxonomic unit (OTU) based or an alignment-based approach is used, identifying conserved SNPs in ITS across studies represents a significant computational challenge. Ostensibly, similar patterns to those observed here would be identified if the *petB* marker were used [23, 37]. However, no primer set exists that concurrently amplifies *Prochlorococcus* and *Synechococcus petB* sequences and would have to be developed. Thus, a strength of the *rpoC1* marker gene system is its ability to identify microdiverse populations in a computationally efficient fashion.

To test for niche overlap, we estimated Pianka's niche overlap index on interpolated data from clades HLI, HLII, HLIII/IV, I/IV, II/III, and CRD1 frequencies in the top 100 m [38]. We estimated geographical distribution significance by comparison with a null model of data randomization across latitude.

Microdiversity analysis

Sequences strictly mapping to the *Prochlorococcus* clades HLIII/IV or HLII and the *Synechococcus* clades CRD1 or I/III at > 97% amino-acid identity and 90% coverage were separated into their respective clades. Single-nucleotide polymorphism (SNP) profiles were calculated as the most common nucleotide per base-pair position for each sample. Sequences were highly unique within each clade (92.5% unique for HLIII/IV, 95.9% for HLII, 99.8% for CRD1, and 99.9% for clade I/III) potentially due to sequencing errors. Samples were assessed for their most informative nucleotides using minimum entropy decomposition [39]. The 100 most informative nucleotide positions per clade were identified for each sequence to limit any additional sequencing error noise. We randomly sampled 100 sequences from most samples and included all sequences from samples with fewer than 100 sequences, noted in Supplementary Table S1. We assessed phylogenetic composition of each clade grouped at 100% sequence identity using the weighted UniFrac metric [40] implemented in R [41]. Using the same top 100 most informative nucleotide positions per sequence, we assessed taxonomic similarity using pick_open_reference_otus.py from QIIME1.9 at a relaxed identity threshold of 85% coupled with Bray-Curtis dissimilarity [42]. We used the permutation multivariate analysis of variance test [42] to test if ocean origin explained the UniFrac-based phylogenetic composition variation. We used distance-based redundancy analysis with the UniFrac distance to test if other environmental parameters (depth, latitude, and temperature) explained phylogenetic composition variation [42]. We used the Mantel test with Pearson correlations [42] to test if clades were correlated with one another based on their phylogenetic composition.

Comparative phylogenies

Multi-locus sequence trees for *Prochlorococcus* and *Synechococcus* were recreated from previous analyses [23, 43]. For *Prochlorococcus*, 503 core genes in *Prochlorococcus* were aligned separately using ClustalW [43, 44]. For *Synechococcus*, seven core genes were aligned separately with ARB [23, 45]. For each lineage, gene sequences were concatenated, then a maximum likelihood tree was constructed using PhyML with neighbor-joining bootstrap support [34], and rooted after with MIT9313 and WH5701 as outgroups for each phylogeny, respectively.

Previous *Prochlorococcus* and *Synechococcus* clade designations were confirmed using the marker gene *rpoC1*. After identifying novel references from within the dataset, no major new clades were identified. Although there are no strain representatives of clade NC1, this clade was well represented across these transects. Assigning sequences to their respective clades at different sequence identity cutoffs

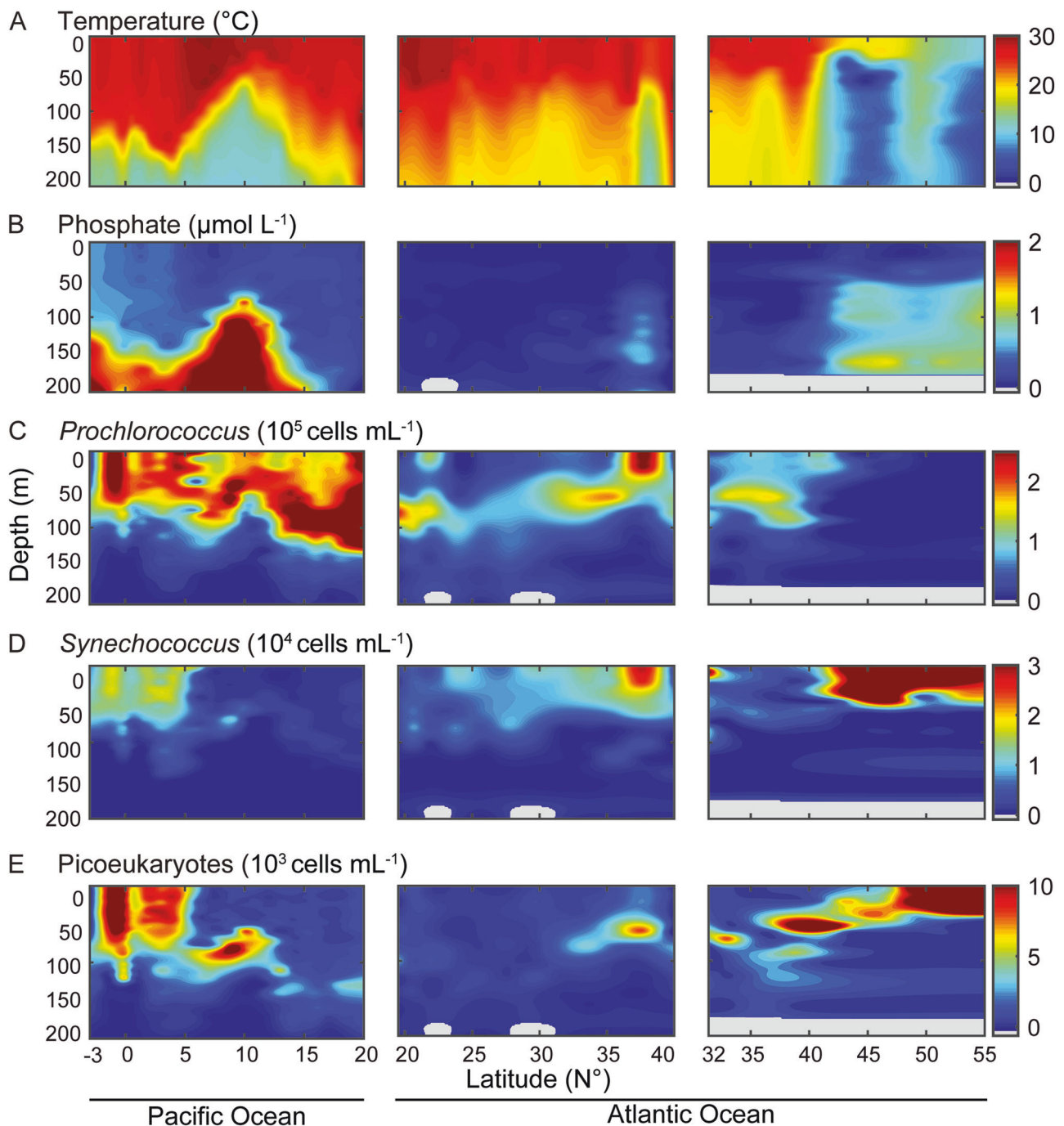


Fig. 2 Variation in environmental conditions and phytoplankton abundances. **a** Temperature profiles across three cruise transects in the Pacific and Atlantic Oceans, **b** soluble reactive phosphate, **c** flow

cytometry counts for *Prochlorococcus*, **d** *Synechococcus*, and **e** picoeukaryotic phytoplankton

did not significantly change the overall clade distribution until >97% amino-acid identity (Mantel test; Supplementary Table S2). Some *Synechococcus* closely related phylogenetic clades overlapped in their biogeography and were analyzed together after ($p < 0.01$ for clade I/IV; $p = 0.056$ for clade II/III; latitudinal randomization test, Supplementary Figure S1, S2).

Results

To identify the detailed distribution of *Prochlorococcus* and *Synechococcus* diversity, we sequenced a phylogenetic marker gene (*rpoC1*, subunit γ of RNA polymerase) from 339 populations from three cruises covering equatorial, subtropical gyre, and colder nutrient-rich mid-latitude

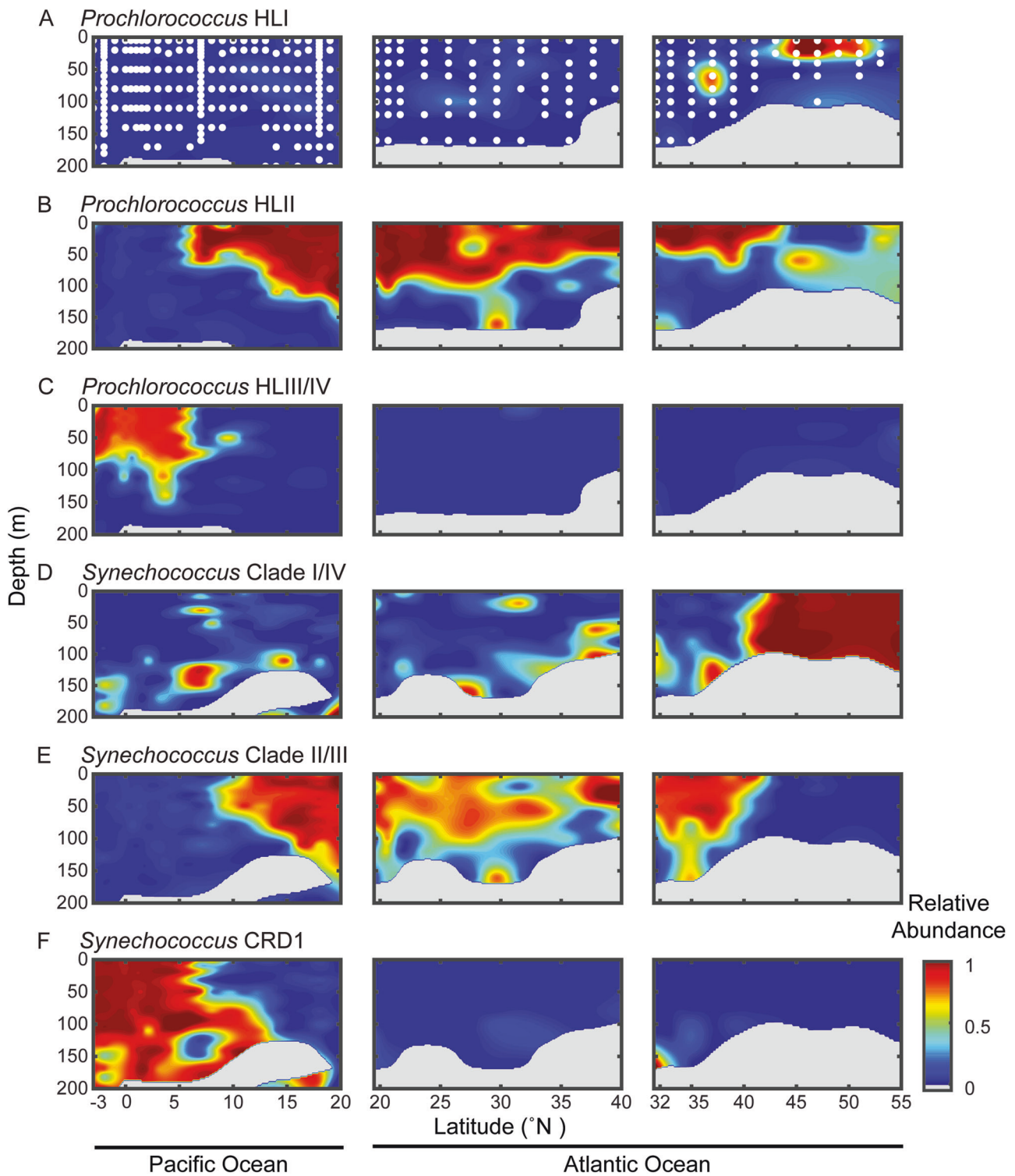


Fig. 3 Clade frequencies across three ocean regions. *Prochlorococcus* **a** clades HLI with white dots representing DNA sampling points for all plots, **b** HLII, and **c** HLIII/IV. *Synechococcus* **d** clade I/IV, **e** II/III, and

f CRD1. Clade relative abundances are normalized to *Prochlorococcus* or *Synechococcus* total number of sequences per sample. Gray areas indicate no data

regions (Figure S1, Supplementary Table S1). The Tropical Pacific Ocean transect (3°S–19°N) was highly stratified, had a clear shallow thermocline and showed limited variation in surface temperature (27.2–29.3 °C) (Fig. 2a). Macronutrient

availability in the surface was lower in the northern section; with phosphate concentration between 100 and 200 nM (Fig. 2b) and nitrate concentration at detection limit (< 10 nM). A sharp transition zone was observed at 5°N with

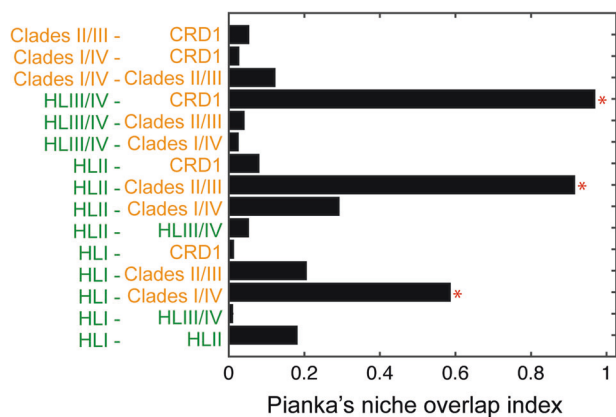


Fig. 4 Niche overlap of *Prochlorococcus* and *Synechococcus* clades. A Pianka index value of 0 represents no overlap and a value of 1 signifies complete overlap [38]. Significance values ($p < 0.01$) come from testing observations against a null model with latitude randomization. *Synechococcus* clades are colored in orange and *Prochlorococcus* clades are colored in green (color figure online)

macronutrient concentrations elevated in the southern section (phosphate > 400 nM). Near the equator, elevated nutrient concentrations were likely driven by upwelling likely leading to iron stress [46]. The environmental conditions in the North Atlantic Ocean (19.7–55°N) also displayed a clear transition zone (Fig. 2a, b). Here, the transition zone was seen in the temperature profile, whereby the surface temperature was above 26 °C below 39°N. North of this transition point, the surface temperature was lower (~19.5 °C) and subsequently dropped to ~10 °C at 55°N. Similarly, the phosphate concentration was very low (< 10 nM) and nutricline deep (~200 m) below 39°N. However, phosphate rose and the nutricline shallowed (~40 m) north of this transition point (Fig. 2b). There were two exceptions to this profile. There was an infusion of deep water at 37°N and slightly elevated nutrient concentrations between 23°N and 19.7°N, likely due to horizontal nutrient supply by the Caribbean Current.

In the Pacific, the abundances of picophytoplankton lineages displayed a similar clear transition, albeit with one or two degrees northwards shift of the nutrient gradient (i.e., 6°N–7°N). Both *Synechococcus* and picoeukaryotic phytoplankton abundances were approximately an order of magnitude higher south of 6°N (Fig. 2d, e). In contrast, there was little latitudinal variation in *Prochlorococcus* surface abundances (Fig. 2c). In the North Atlantic Ocean, the abundances of picophytoplankton lineages mirrored the environmental transition at 39°N, whereby a clear shift between *Prochlorococcus*-dominated (south) and elevated abundances of *Synechococcus* and picoeukaryotic phytoplankton (north) (Fig. 2c–e). Thus, these transition points were indicative of both the environmental conditions and phytoplankton abundance patterns.

In parallel with the overall phytoplankton community composition, the relative frequency of major *Prochlorococcus* and *Synechococcus* clades also shifted at the transition points (Fig. 3). For *Prochlorococcus*, the equatorial Pacific section was dominated by the low-iron-adapted HLIII/IV clade (Fig. 3c), whereas the northern Pacific tropical mixed layer was dominated by the high temperature-adapted HLII clade (Fig. 3b). Thus, the transition between HLIII/IV and HLII matched the shift in abundance of *Synechococcus* and picoeukaryotic phytoplankton. The HLI clade was only observed intermittently at lower abundance, right above the nutricline in the northern section, but was absent in the equatorial part (Fig. 3a). The LL-adapted clades showed clear depth partitioning, whereby LLI was detected right below the HL ecotypes followed by LLII/III, LLIV, and NC1 at the bottom of the euphotic zone (Supplementary Figure S3). The latitudinal separation of *Prochlorococcus* ecotypes was also seen for *Synechococcus*. Clade CRD1 was most frequent in the equatorial part (Fig. 3f) and clade II/III dominated the surface mixed layer north of 7°N (Fig. 3e and Supplementary Figures S4C and S4D).

In the North Atlantic Ocean, we again saw a parallel latitudinal ecotype distribution for *Prochlorococcus* and *Synechococcus* (Fig. 3). There was a clear transition at 39°N, as HLII dominated the southern regions and HLI the northern regions between 43°N and 49°N (Fig. 3a). Above 49°N, *Prochlorococcus* as a whole nearly disappeared (Fig. 2c). HLI was present deeper in the water column near the 10 nM phosphate nutricline in the southern part of Atlantic Ocean section. As seen in the Pacific samples, *Prochlorococcus* LLI was the dominant LL clade and became frequent right below the nutricline (Supplementary Figure S3A). *Prochlorococcus* LLII/III and LLIV were detected deeper in the water column (Supplementary Figures S3B and S3C). For *Synechococcus*, clade II/III was most frequent south of 39°N (Fig. 3e and Supplementary Figures S4C and S4D), whereas clade I/IV dominated between 41 and 55°N (Fig. 3d and Supplementary Figures S4A and S4B). Thus, both *Prochlorococcus* and *Synechococcus* clades displayed clear latitudinal distributions with transition points at 6°N for the Pacific Ocean and 39°N for the Atlantic Ocean.

We next quantified the degree to which individual clades had a shared biogeography using Pianka's index [38]. Three pairs of *Prochlorococcus* and *Synechococcus* clades significantly overlapped in their spatial distributions ($p < 0.01$; latitudinal randomization test, Fig. 4). Specifically, *Prochlorococcus* HLIII/IV significantly overlapped with the *Synechococcus* CRD1 with high abundance in the hot, elevated macronutrient waters between 3°S and 6°N in the Pacific Ocean and largely absent elsewhere. *Prochlorococcus* HLII overlapped with

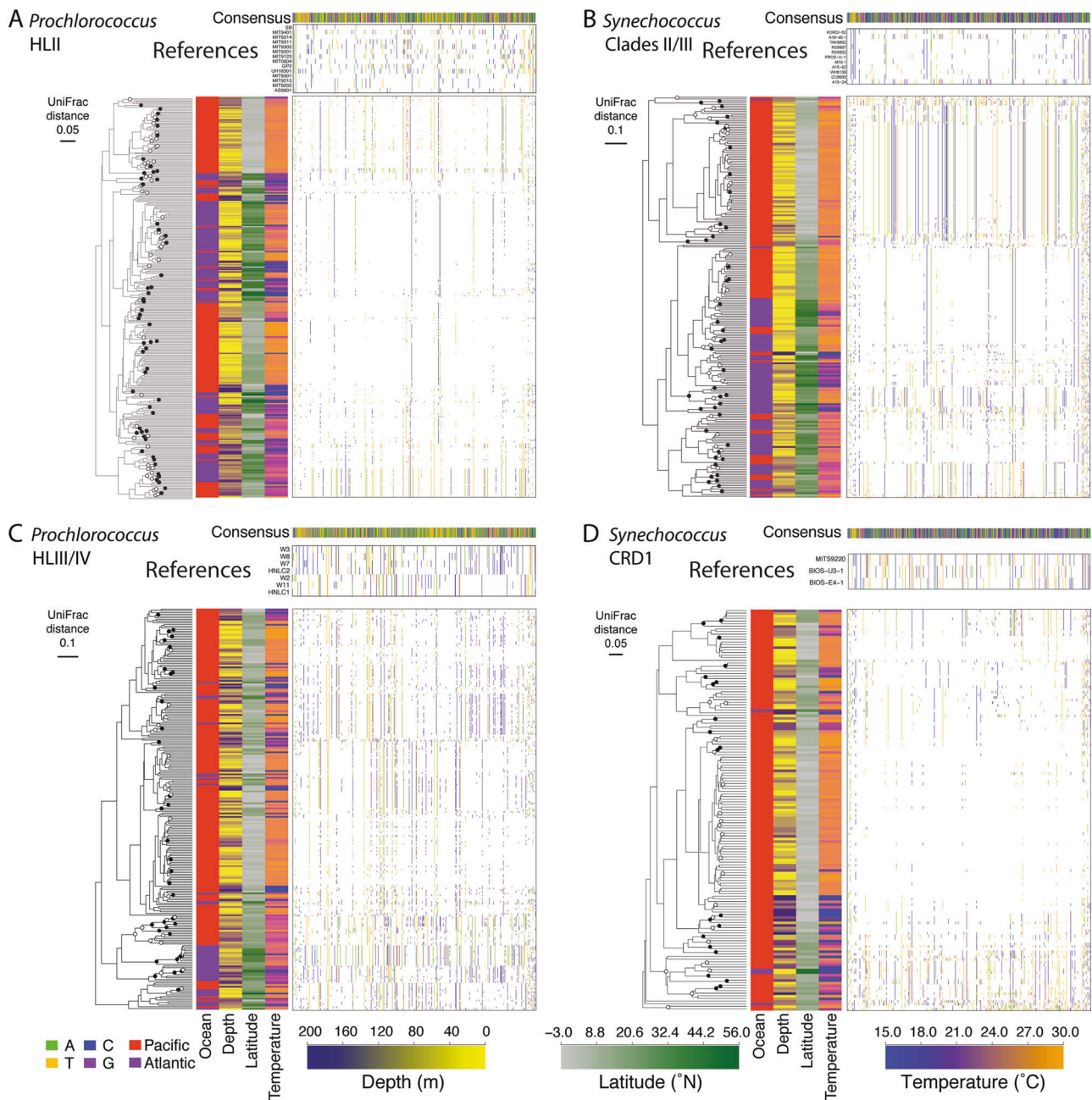


Fig. 5 Parallel microdiversity between *Prochlorococcus* and *Synechococcus* clades. Clades **a** *Prochlorococcus* HLI, **b** *Synechococcus* Clade II/III, **c** *Prochlorococcus* HLIII/IV, and **d** *Synechococcus* CRD1. Each row represents a sample. Rows were clustered by UniFrac sample distance, nodes with unfilled circles represent > 50 bootstrap support and filled circles represent > 75 out of 100 resampled trees. Side columns are colored by ocean origin, depth, latitude, and

temperature followed by SNPs colored by the most prevalent nucleotide in each position for each sample. SNP profiles highlight single-nucleotide differences from overall SNP profile consensus across sites for each clade. Reference sequences from Fig. 1 isolates and metagenomic assemblies are included for each corresponding group, which are grouped phylogenetically in Supplementary Figure S5 (color figure online)

Synechococcus clade II/III in hot but low macronutrient waters between 6°N and 20°N in the Pacific Ocean and between 21°N and 39°N in Atlantic Ocean. *Prochlorococcus* HLI overlapped with *Synechococcus* clade I/IV in cooler but elevated macronutrient waters north of 39°N in the Atlantic Ocean. Thus, there was significant

evidence for a parallel biogeography of specific *Prochlorococcus* and *Synechococcus* clades.

We next detected distinct regional microdiversity within the two most frequently observed clades of *Prochlorococcus* (HLII and HLIII/IV) and *Synechococcus* (clade II/III, and CRD1) (Fig. 5 and Supplementary

Table S3 and Figure S5). Although sequences within these clades were nearly identical at the amino-acid level, there was extensive synonymous variation such that most sequences were entirely unique. However, sequences from individual samples shared unique SNPs suggesting locally distinct microdiversity in both *Prochlorococcus* and *Synechococcus* clades. Notably, this sequence variation was not random by region, but instead biogeographically and environmentally structured. *Prochlorococcus* HLII and *Synechococcus* Clade II/III had SNPs that were unique to regions of the Atlantic vs. Pacific Ocean suggesting adaptation to environmental differences between the two oceans (Fig. 5, Supplementary Figure S6, and Table S4). We also detected distinct populations across latitude, temperature, and depth (Supplementary Table S5 and Figure S7). Thus, there was a major latitudinal transition in SNPs for each clade (HLII, HLIII/IV, CRD1, clade II/III) (Supplementary Figure S6). In general, these changes followed the clade transitions with major shifts around 6–7°N but not always at this specific point. Clade II/III shifted between 5 and 6°N and HLII SNPs shifted even further south between 4 and 5°N. To test our new approach, we also compared three sample dissimilarity metrics: phylogenetic composition using UniFrac distance, the sample SNP profiles using Euclidean distance, or the taxonomic diversity using the Bray–Curtis dissimilarity (Supplementary Table S6) yielding significant correlations between metrics for all clades. Thus, our SNP profiling gave similar results to traditional approaches for examining the distribution of microbial diversity but also allowed the identification of the actual sequence differences leading to such patterns.

Depth-dependent environmental changes also structured the microdiversity of each clade illustrated by sites with high vertical sampling frequency (Supplementary Figure S8). At 18°N, the microdiversity for *Prochlorococcus* HLII was uniform down to 70 m, which covered both the mixed layer (depth ~ 20 m), as well as stratified waters below (Supplementary Figure S8A). At 70 m, we observed a subtle shift, whereas the population deep in the euphotic zone (< 140 m) was different. Thus, there were distinct vertical populations over short distances of the HL-adapted clade HLII in the water column. The population structure of *Synechococcus* clades II/III showed a parallel vertical distribution (Supplementary Figure S8B). Again, populations were similar within and below the mixed layer, whereas different populations were found deeper in euphotic zone. At 2°S in the Pacific Ocean, the microdiversity for *Prochlorococcus* HLIII/IV and *Synechococcus* CRD1 were uniform within the mixed layer (mixed layer depth: 64 m) and below to ~100 m (Supplementary Figure S8C + D). Below this point, we observed slight variations in the genetic structure for both HLIII/IV and CRD1. As seen regionally, we observed clear parallel variations in the

vertical microdiversity structure for both *Prochlorococcus* and *Synechococcus* clades.

Discussion

By using the same genetic marker, sequencing technique, analysis method, and samples, we confirm in this study our first hypothesis that the phylogeography of *Prochlorococcus* and *Synechococcus* are tightly concordant across ocean regions. Building on a series of recent studies [12, 17, 47], we demonstrate that *Prochlorococcus* and *Synechococcus* contain three matching major surface ocean ecotypes and now suggest that these ecotypes are phylogenetically organized and distributed equivalently. The first ecotype (*Prochlorococcus* HLIII/IV and *Synechococcus* CRD1) is adapted to high temperature, elevated macronutrients, and presumably low iron availability. In our study, this ecotype is restricted to the iron-stressed Eastern Pacific Equatorial Zone but other studies have also detected HLIII/IV and CRD1 in the Indian Ocean upwelling zones [6, 47], the Costa Rica Dome water [16, 48], and the Benguela Upwelling Zone [17]. The second ecotype (HLII and clade II/III) is present in high temperature, low macronutrient, and high iron waters, which in our study were sandwiched between equatorial upwelling zones and colder, nutrient-rich mid-latitude waters. The third major ecotype (HLI and clades I/IV) is found in higher latitude colder, nutrient-rich waters. Despite the strong niche overlap between *Prochlorococcus* and *Synechococcus* genetic diversity, our findings do not imply that these clades occupy the same niche given that *Prochlorococcus* cells in comparison with *Synechococcus* have distinctive photosystems, smaller cell size, and other differing features [49]. Thus, it is an open question why the two lineages co-exist in large parts of the ocean. Echoing Hutchinson's Paradox of the Plankton [50], this may be due to a combination of different physiological traits, top-down controls, and mixing.

The parallel phylogeography between *Prochlorococcus* and *Synechococcus* extends from the ecotype level further into the fine-scaled genetic diversity of both lineages. The large number of samples analyzed here support the idea that adaptation to the local environment as opposed to stochastic processes control the observed patterns [51]. We propose that a stoichiometric difference in nutrient concentrations and consequently nitrogen (Pacific Ocean) vs. phosphate (Atlantic Ocean) stress between the gyres likely underpins the phylogeographic patterns [52]. We also saw that the microdiversity SNP profiles shifted at the same location in the Pacific Ocean as the shifts observed at higher taxonomic levels, suggesting that the diversity at many levels is at least in part subject to the same selection [47]. Such a biogeography at a very fine phylogenetic level has also been

eluded to by a comparison of single-cell genomes, as well as metagenomes from the Pacific and Atlantic Ocean subtropical gyres [53–56]. Furthermore, previous work have suggested that stoichiometric changes in a N vs. P could be an important driver of basin-scale differences in population structure [57]. Past studies have struggled in detecting a phylogeography among *Prochlorococcus* and *Synechococcus* microdiversity—especially as it pertains to the HLII clade [36, 47, 57]. Our approach takes advantage of using (i) a marker with no indels leading to low uncertainty in calling sequence variants (e.g., in comparison with ITS) and (ii) a functional gene resulting in a large number of SNPs between microdiverse lineages making our approach less sensitive to sequencing errors. We suspect that the combination enabled us to detect systematic patterns in the regional and vertical microdiversity of *Prochlorococcus* and *Synechococcus*.

We find that diversity across picophytoplankton lineages, clades, and microdiversity all shift near specific transition points. These transitions occur near 6°N in the Tropical Pacific Ocean and 39°N in the North Atlantic Ocean and represent fundamental shifts in the environmental conditions supporting phytoplankton growth. These transitions are likely biologically amplified as phytoplankton-driven biological processes contribute to the availability of resources (i.e., drawdown of nutrients and to some extent light shading). The uniformity in distribution of phytoplankton diversity at multiple phylogenetic levels across these environmental zones suggests that all lineages experience similar stressors. In some cases, adaptation can overcome the stress condition as seen in the even distribution of *Prochlorococcus* across the tropical Pacific Ocean. In contrast, the abundance of *Synechococcus* and picoeukaryotic phytoplankton were more sensitive to environmental changes as indicated by the change in overall abundance around 6°N–7°N. However, considering the clades within each lineage, the shifts in biophysical growth constraints exert specific selective pressures leading to strong shifts in phylogenetic diversity across both *Prochlorococcus* and *Synechococcus*—and likely many other lineages. Thus, our study suggests that as *Prochlorococcus* and *Synechococcus* co-occur in regions with specific environmental stress conditions, this leads to a parallel evolutionary diversification and biogeography across phylogenetic scales.

Acknowledgements We thank Bradford Hawkins for help with the niche overlap analysis, David Karl and Ken Doggett for help with flow cytometry during the NH1418 cruise, Frédéric Partensky and David Scanlan for sharing *rpoCI* sequences, and Sophie Mazard for sharing the multi-locus alignment. AGK was supported by the National Science Foundation Graduate Research Fellowship Program (DGE-1321846) and ACM was supported by the National Science Foundation grants (OCE-0928544, OCE-1046297, and OCE-1559002).

Author contributions SEB, CM, JSH, MWL, and ACM, acquired data. SEB, CM, JSH, and MWL performed the experiments. AGK and ACM analyzed the data. AGK, CM, AAL, MWL, and ACM wrote the paper.

Compliance with ethical standards

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

References

1. Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, et al. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci USA*. 2013;110:9824–9.
2. Moore LR, Rocap G, Chisholm SW. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature*. 1998;393:464–7.
3. Ahlgren NA, Rocap G, Chisholm SW. Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environ Microbiol*. 2006;8:441–54.
4. Zinser ER, Johnson ZI, Coe A, Karaca E, Veneziano D, Chisholm SW. Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol Oceanogr*. 2007;52:2205–20.
5. Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW. Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ Microbiol*. 2009;11:823–32.
6. Rusch DB, Martiny AC, Dupont CL, Halpern AL, Venter JC. Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. *Proc Natl Acad Sci USA*. 2010;107:16184–9.
7. Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A, et al. Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J*. 2013;7:184–98.
8. West NJ, Lebaron P, Strutton PG, Suzuki MT. A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *ISME J*. 2011;5:933–44.
9. Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J*. 2012;6:285–97.
10. Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science*. 2006;311:1737–40.
11. Fuller NJ, Marie D, Vaultot D, Post AF, Scanlan DJ. Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol*. 2003;69:2430–43.
12. Zwirgmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaultot D, et al. Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol*. 2008;10:147–61.
13. Zwirgmaier K, Heywood JL, Chamberlain K, Woodward EMS, Zubkov MV, Scanlan DJ. Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. *Environ Microbiol*. 2007;9:1278–90.
14. Mella-Flores D, Mazard S, Humily F, Partensky F, Mahé F, Bariat L, et al. Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea affected by global warming? *Biogeosciences Discuss*. 2011;8:4281–330.

15. Post AF, Penno S, Zandbank K, Paytan A, Huse SM, Welch DM. Long term seasonal dynamics of *Synechococcus* population structure in the Gulf of Aqaba, northern Red Sea. *Front Microbiol.* 2011;2:131.
16. Gutiérrez-Rodríguez A, Slack G, Daniels EF, Selph KE, Palenik B, Landry MR. Fine spatial structure of genetically distinct picocyanobacterial populations across environmental gradients in the Costa Rica Dome. *Limnol Oceanogr.* 2014;59:705–23.
17. Sohm JA, Ahlgren NA, Thomson ZJ, Williams C, Moffett JW, Saito MA, et al. Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J.* 2016;10:333–45.
18. Ting CS, Rocap G, King J, Chisholm SW. Phycobiliprotein genes of the marine photosynthetic prokaryote *Prochlorococcus*: evidence for rapid evolution of genetic heterogeneity. *Microbiology.* 2001;147:3171–82.
19. Moore LR, Post AF, Rocap G, Chisholm SW. Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol Oceanogr.* 2002;47:989–96.
20. Moore LR, Ostrowski M, Scanlan DJ, Feren K, Sweetsir T. Ecotypic variation in phosphorus-acquisition mechanisms within marine picocyanobacteria. *Aquat Microb Ecol.* 2005;39:257–69.
21. Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, et al. Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev.* 2009 Jun;73:249–99.
22. Martiny AC, Treseder K, Pusch G. Phylogenetic conservatism of functional traits in microorganisms. *ISME J.* 2013;7:830–8.
23. Mazard S, Ostrowski M, Partensky F, Scanlan DJ. Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol.* 2012;14:372–86.
24. Bergsland KJ, Haselkorn R. Evolutionary relationships among Eubacteria, Cyanobacteria, and Chloroplasts: evidence from the *rpoC1* gene of *Anabaena* sp. strain PCC 7120. *J Bacteriol.* 1991;173:3446–55.
25. Toledo G, Palenik B. *Synechococcus* diversity in the California Current as seen by RNA polymerase (*rpoC1*) gene sequences of isolated strains. *Appl Environ Microbiol.* 1997;63:4298–303.
26. Karl DM, Tien G. MAGIC: a sensitive and precise method for measuring dissolved phosphorus in aquatic environments. *Limnol Oceanogr.* 1992;37:105–16.
27. Lomas MW, Burke AL, Lomas DA, Bell DW, Shen C, Dyrman ST, et al. Sargasso Sea phosphorus biogeochemistry: an important role for dissolved organic phosphorus (DOP). *Biogeosciences* . 2010;7:695–710.
28. Bostrom KH, Simu K, Hagstrom A, Riemann L. Optimization of DNA extraction for quantitative marine bacterioplankton community analysis. *Limnol Oceanogr Methods.* 2004;2:365–73.
29. Tai V, Palenik B. Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *ISME J.* 2009;3:903–15.
30. Andrews S. FastQC: a quality control tool for high throughput sequence data. [Internet]. 2010. Available from: www.bioinformatics.babraham.ac.uk/projects/fastqc/
31. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Publ Gr.* 2010;7:335–6.
32. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinform.* 2009;10:421.
33. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics.* 2004;20:289–90.
34. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.69. Department of Genome Sciences, University of Washington, Seattle: Distributed by the author; 2005. <http://evolution.genetics.washington.edu/phylip.html>.
35. Schlitzer R. Interactive analysis and visualization of geoscience data with Ocean Data View. *Comput Geosci.* 2002;28:1211–8.
36. Larkin AA, Blinebry SK, Howes C, Lin Y, Loftus SE, Schmaus CA, et al. Niche partitioning and biogeography of high light adapted *Prochlorococcus* across taxonomic ranks in the North Pacific. *ISME J.* 2016;10:1555–67.
37. Urbach E, Scanlan DJ, Distel DL, Waterbury JB, Chisholm SW. Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (cyanobacteria). *J Mol Evol.* 1998;46:188–201.
38. Pianka ER. The structure of lizard communities. *Annu Rev Ecol Syst.* 1973;4:53–74.
39. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.* 2014;9:968–79.
40. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol.* 2005;71:8228–35.
41. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics.* 2010;26:1463–4.
42. Oksanen J, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, et al. Vegan: community ecology package. R package version 2.0-10. [Internet]. Vol. 1, R package version. 2013. Available from: <http://cran.r-project.org>
43. Berube PM, Biller SJ, Kent AG, Berta-Thompson JW, Roggensack SE, Roache-Johnson KH, et al. Physiology and evolution of nitrate acquisition in *Prochlorococcus*. *ISME J.* 2015;9:1195–207.
44. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23:2947–8.
45. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar A, et al. ARB: a software environment for sequence data. *Nucleic Acids Res.* 2004;32:1363–71.
46. Fitzwater SE, Coale KH, Gordon RM, Johnson KS, Ondrusek ME. Iron deficiency and phytoplankton growth in the equatorial Pacific. *Deep Sea Res Part II Top Stud Oceanogr.* 1996;43:995–1015.
47. Farrant GK, Doré H, Cornejo-Castillo FM, Partensky F, Ratin M, Ostrowski M, et al. Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci USA.* 2016;113:E3365–74.
48. Ahlgren NA, Noble A, Patton AP, Roache-johnson K, Robinson D, McKay C, et al. The unique trace metal and mixed layer conditions of the Costa Rica upwelling dome support a distinct and dense community of *Synechococcus*. *Limnol Oceanogr.* 2014;59:2166–84.
49. Ting CS, Rocap G, King J, Chisholm SW. Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends Microbiol.* 2002;10:134–42.
50. Hutchinson GE. The paradox of the plankton. *Am Nat.* 1961;95:137–45.
51. Larkin AA, Martiny AC. Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environ Microbiol Rep.* 2017;9:55–70.
52. Wu J, Sunda W, Boyle EA, Karl DM. Phosphate depletion in the western North Atlantic. *Ocean Sci.* 2000;289:759–62.
53. Kashtan N, Roggensack SE, Berta-Thompson JW, Grinberg M, Stepanauskas R, Chisholm SW. Fundamental differences in diversity and genomic population structure between Atlantic and Pacific *Prochlorococcus*. *ISME J* 2017;11:1997–2011.

54. Martiny AC, Huang Y, Li W. Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ Microbiol.* 2009;11:1340–7.
55. Martiny AC, Kathuria S, Berube PM. Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc Natl Acad Sci USA.* 2009;106:10787–92.
56. Delmont TO, Eren AM. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ.* 2018;6:e4320.
57. Kent AG, Dupont CL, Yooseph S, Martiny AC. Global biogeography of *Prochlorococcus* genome diversity in the surface ocean. *ISME J.* 2016;10:1856–65.