environmental microbiology

Environmental Microbiology (2015) **17**(10), 3692–3707 doi:10.1111/1462-2920.12742

Cyanobacterial distributions along a physico-chemical gradient in the Northeastern Pacific Ocean

Sebastian Sudek,[1‡] R. Craig Everroad,[2‡] Alyssa-Lois M. Gehman,[1†] Jason M. Smith,[1] Camille L. Poirier,[1] Francisco P. Chavez[1] and Alexandra Z. Worden[1,3*]

[1]*Monterey Bay Aquarium Research Institute, 7700 Sandholdt Rd, Moss Landing, CA 95039, USA.*

[2]*Exobiology Branch, NASA Ames Research Center, MS 239-4, Moffett Field, CA 94035, USA.*

[3]*Integrated Microbial Biodiversity Program, Canadian Institute for Advanced Research, Toronto, ON M5G 1Z8, Canada.*

Summary

The cyanobacteria *Prochlorococcus* and *Synechococcus* are important marine primary producers. We explored their distributions and covariance along a physico-chemical gradient from coastal to open ocean waters in the Northeastern Pacific Ocean. An inter-annual pattern was delineated in the dynamic transition zone where upwelled and eastern boundary current waters mix, and two new *Synechococcus* clades, Eastern Pacific Clade (EPC) 1 and EPC2, were identified. By applying state-of-the-art phylogenetic analysis tools to bar-coded 16S amplicon datasets, we observed higher abundance of *Prochlorococcus* high-light I (HLI) and low-light I (LLI) in years when more oligotrophic water intruded farther inshore, while under stronger upwelling *Synechococcus* I and IV dominated. However, contributions of some cyanobacterial clades were proportionally relatively constant, e.g. *Synechococcus* EPC2. In addition to supporting observations that *Prochlorococcus* LLI thrive at higher irradiances than other LL taxa, the results suggest LLI tolerate lower temperatures than previously reported. The phylogenetic precision of our 16S rRNA gene analytical approach and depth of bar-coded sequencing also facilitated detection of clades at low abundance in unexpected places. These include *Prochlorococcus* at the coast and *Cyanobium*-related sequences offshore, although it remains unclear whether these came from resident or potentially advected cells. Our study enhances understanding of cyanobacterial distributions in an ecologically important eastern boundary system.

Received 26 July, 2014; revised 1 December, 2014; accepted 1 December, 2014. *For correspondence. E-mail azworden@mbari.org; Tel. 831 775 2122; Fax 831 775 1620. †Present address: Odum School of Ecology, University of Georgia, Athens, GA 30602, USA. ‡These authors contributed equally.

Introduction

Prochlorococcus and marine *Synechococcus* are the major cyanobacterial lineages in the ocean and are typically referred to as picocyanobacteria due to their small cell size (cell diameter $< 2\,\mu m$). *Prochlorococcus* is the most abundant phytoplankton taxon in oligotrophic regions of the ocean where *Synechococcus* is usually present at much lower concentrations (Johnson *et al.*, 2006; Zwirglmaier *et al.*, 2008; Partensky and Garczarek, 2010; Flombaum *et al.*, 2013). In contrast, under mesotrophic conditions, *Synechococcus* thrives, most likely in response to the increased macronutrient availability (Blanchot *et al.*, 2001; Palenik *et al.*, 2006). Considerable genetic and phenotypic diversity can be found between *Prochlorococcus* and *Synechococcus* (Dufresne *et al.*, 2008; Scanlan *et al.*, 2009). While they frequently co-occur, the two genera possess different nutrient uptake and utilization capabilities, photophysiologies and phenotypic flexibility (for review see Scanlan *et al.*, 2009). Within-genera diversification has also been extensively documented for both *Prochlorococcus* and *Synechococcus*.

Genomes from 16 *Prochlorococcus* isolates from diverse habitats and experimental results have revealed differences in metabolic capabilities and ecotypic differentiation related to light and temperature (Martiny *et al.*, 2009; Partensky and Garczarek, 2010; Malmstrom *et al.*, 2013). Through single-cell genomics, hundreds of coexisting populations of *Prochlorococcus* with distinct sets of niche-adapted genes have become apparent in the open ocean (Kashtan *et al.*, 2014). Nevertheless, the resulting whole genome-based phylogenies largely agree with ribosomal gene-based phylogenetic reconstructions. 16S ribosomal ribonucleic acid (rRNA) gene phylogenies differentiate eight *Prochlorococcus* ecotypes. High-light (HL) *Prochlorococcus* ecotypes dominate the upper half of oceanic euphotic zone (Moore *et al.*, 1998), with HLI and HLII strains more abundant in temperate and tropical zones, respectively (Johnson *et al.*, 2006; Zwirglmaier

et al., 2008). Distinct HL clades have also been found in the high-nutrient waters of the equatorial Pacific (Rusch *et al*., 2010; West *et al*., 2011). Originally named HNLC1 and HNLC2, these strains are now termed HLIII and HLIV (Huang *et al*., 2012; Malmstrom *et al*., 2013). Low-light (LL) *Prochlorococcus* strains are commonly observed lower in the euphotic zone and in the deep chlorophyll maximum (DCM). These ecotypes form more basal clades in the *Prochlorococcus* lineage and harbour greater phylogenetic diversity than HL strains (Rocap *et al*., 2002). In the Atlantic, LLIV is abundant below the thermocline (Johnson *et al*., 2006), while LLII and LLIII are thought to be scarce (Partensky and Garczarek, 2010). Low-light I appears to be more polyvalent in nature and has been observed in surface waters during times of deep mixing at station ALOHA in the Central Pacific and BATS in the Atlantic (Malmstrom *et al*., 2010).

Marine *Synechococcus* also form well-defined phylogenetic clades based on 16S rRNA and internal transcribed spacer (ITS) sequences (Fuller *et al*., 2003; Ahlgren and Rocap, 2006; Mazard *et al*., 2012). Three major *Synechococcus* subclusters, 5.1, 5.2 and 5.3 have been defined, with 5.1 encompassing most marine clades (Rocap *et al*., 2002; Dufresne *et al*., 2008). Fourteen clades can be discriminated using the 16S rRNA gene, and some additional clades can be delineated using other markers, e.g. the ITS or protein-coding genes (Choi and Noh, 2009; Paerl *et al*., 2011; Huang *et al*., 2012; Mazard *et al*., 2012). Analyses of the latter have also revealed environmental clades that do not appear to be in culture. Like *Prochlorococcus*, cultured marine *Synechococcus* strains exhibit extensive physiological diversity. At present, the latter show a greater number of defined phylogenetic clades that appear to harbour more genetic divergence than observed thus far for *Prochlorococcus* (Scanlan *et al*., 2009; Ahlgren and Rocap, 2012; Huang *et al*., 2012; Mazard *et al*., 2012). However, unlike *Prochlorococcus*, strains within the various *Synechococcus* 16S rRNA gene clades do not show clear patterns with respect to light adaptation (Toledo *et al*., 1999; Six *et al*., 2007; Everroad and Wood, 2012). *Synechococcus* clades often coexist, and have allelopathic effects on each other that make it difficult to identify the environmental factors that constrain their distributions (Toledo and Palenik, 1997; Fuller *et al*., 2003; Paz-Yepes *et al*., 2013). Distributions of some clades do appear to be associated with particular nutrient regimes and temperatures, and others are more geographically constrained (Zwirglmaier *et al*., 2008; Scanlan *et al*., 2009; Huang *et al*., 2012). *Synechococcus* clades I and IV are commonly reported in temperate coastal waters, clade II in warmer coastal and open ocean waters and clade III in oligotrophic open ocean waters. Some 16S rRNA gene-based clades can also be connected to physiological differences that presumably reflect ecotypic differentiation, e.g. motility in clade III, reduced nitrate utilization in clade XV and chromatic adaptation in clades XV and XVI (Toledo *et al*., 1999; Ahlgren and Rocap, 2006; Everroad *et al*., 2006; Mazard *et al*., 2012).

Diverse *Prochlorococcus* and *Synechococcus* have been noted at two sites in the California Current System (Ferris and Palenik, 1998). Eastern boundary upwelling systems, like the California Current System (CCS), are highly productive, supporting 40–50% of world fisheries in part due to productivity stimulated in the transition zone (Mackas *et al*., 2006). Seasonal wind-driven upwelling of cold, nutrient-rich water leads to a mesotrophic setting often within 50 km of shore (Checkley and Barth, 2009). Seaward of the California Current jet, approximately 200 km from shore (Collins *et al*., 2003), are relatively warm and oligotrophic waters, more akin to the North Pacific Gyre waters. Between the coastal upwelling and the California Current region lies a dynamic transition zone where upwelled, and offshore waters interact with the low salinity, low nutrient jet. Such transition zones are unique environments, where planktonic organisms putatively adapted to the stratified, low nutrient offshore environment or the turbulent, high nutrient conditions of the near shore can be subject to pronounced shifts in their environment, via mixing or advection in response to changes in currents and upwelling.

Here, we used the CCS system to investigate the distribution and covariance of *Prochlorococcus* and *Synechococcus* lineages in coastal, transition zone and open ocean waters. The North Pacific sites sampled extended to 800 km offshore along California Cooperative Oceanic Fisheries Investigations (CalCOFI) Line 67 (McClatchie, 2014). Near full-length environmental 16S rRNA gene sequences generated from these sites were used to construct a comprehensive reference tree of the marine picocyanobacterial radiation and revealed two new marine *Synechococcus* clades. Bar-coded V1-V2 16S rRNA sequencing (454-titanium) coupled with phylogenetic and *in-silico* probe-based analyses were then applied to explore picocyanobacterial diversity and distribution across this Northeastern Pacific region. The clade-specific variations observed in association with the transect, and interannual changes in the transition zone, further our understanding of picocyanobacterial ecology.

Results

The stations sampled transitioned from nutrient-rich, relatively cold coastal waters (H3) through intermediate upwelling-influenced waters termed here the ‘transition

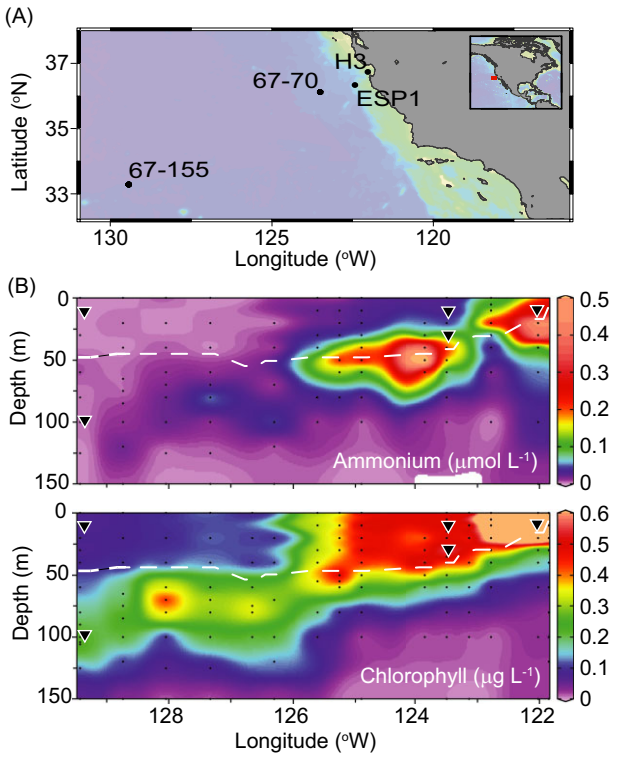

Fig. 1. A. Map of the eastern North Pacific region under study. Sampling stations for DNA are marked (dots). Station ESP1 was sampled in 2012 only. B. Ammonium and chlorophyll concentrations along the 2009 transect. Depths where DNA was collected are marked by triangles. The computed mixed layer depth (see *Methods*) is indicated by a white dashed line.

zone' (TZ, e.g. 67–70) to oligotrophic, stratified waters offshore termed the 'open ocean station' (e.g. 67–155; Fig. 1). Transect samples for diversity analysis came from each of these three stations in 2009, with surface water samples collected between 5 m to 14 m, and deep chlorophyll maxima samples taken at 41 m (67–70) and 105 m (67–155). For 3 subsequent years (2011, 2012 and 2013), sampling was at 67–70, or ~100 km closer to shore (ESP1, 2012), at the surface depth range specified above (Table 1).

Along the transect (2009), surface nitrate and ammonium concentrations varied from 12.91 μM and 0.26 μM at H3 to 0.40 and < 0.01 μM in the TZ and < 0.05 and < 0.01 μM at 67–155, while phosphate was more constant (Fig. 1B; Table 1 and Table S1). Two hundred seventy-two *Prochlorococcus* and 73 *Synechococcus* near full-length 16S rRNA gene sequences in clone libraries constructed from these sites using established 'universal' primers met all assembly, quality and length criteria. For *Synechococcus*, 25 sequences were recovered from the coastal clone library, and 28 sequences were recovered from the TZ. The combined open ocean depths contained 20 sequences. In contrast, no *Prochlorococcus* clones were recovered from the coastal library, five were found in the TZ and a total of 267 sequences were present in the two open ocean libraries. The *Prochlorococcus* sequences were collapsed at the 99% identity level and used to construct a *Prochlorococcus*-specific phylogeny (Fig. S1), while *Synechococcus* sequences were used in a separate phylogenetic reconstruction (Fig. S2). A selection of sequences representing diversity from across these two trees was incorporated into a combined cyanobacterial reference alignment and phylogenetic reconstruction that was then used for amplicon placement (Fig. 2).

The phylogenetic analysis of near full-length cyanobacterial 16S RNA gene sequences for *Synechococcus* recovered all previously established 16S clades. These included subcluster 5.1 (clades I-IX, XV, XVI, WPC1, WPC2), 5.2 and 5.3 in the overall picocyanobacterial (Fig. 2) and *Synechococcus*-only

Table 1. Physico-chemical parameters and amplicons attained for the analysed samples. Note that values from the WFAD09 cruise have been published in Santoro and colleagues 2013.

Site	Date (d/m/y)	Location (latitude; longitude)	Depth (m)	T (°C)	S (ppt)	PO_4^{3-} (μM)	NO_3^- (μM)	NH_4^+ (μM)	Chlorophyll (mg m^{-3})	Total amplicons	Cyanobacterial amplicons
WFAD09 cruise											
H3	1/10/2009	36.739; −122.020	11	11.58	33.36	1.15	12.91	0.263	1.985	21 375	518
67–70	2/10/2009	36.126; −123.491	11	16.43	33.19	0.32	0.40	< 0.01	0.496	19 680	3460
67–70	2/10/2009	36.126; −123.491	41	11.65	33.00	0.84	8.09	0.178	0.345	20 723	2754
67–155	5/10/2009	33.287; −129.428	14	19.92	33.34	0.47	< 0.004	< 0.01	0.063	25 968	11 660
67–155	5/10/2009	33.287; −129.428	105	14.29	33.54	0.34	0.05	< 0.01	0.101	26 184	8574
CANON11 cruise											
67–70	7/10/2011	36.089; −123.469	5	15.61	33.32	0.71	2.29	n.d.	0.773	73 977	7246
C0912 cruise											
ESP1	14/9/2012	36.303; −122.383	10	15.32	33.34	0.70	2.03	0.130	0.270	203 399	12 051
CN13ID cruise											
67–70	15/10/2013	36.132; −123.42	10	15.66	33.01	0.20	0.35	0.044	0.520	69 727	7998

T, temperature; S, salinity; n.d. not determined.

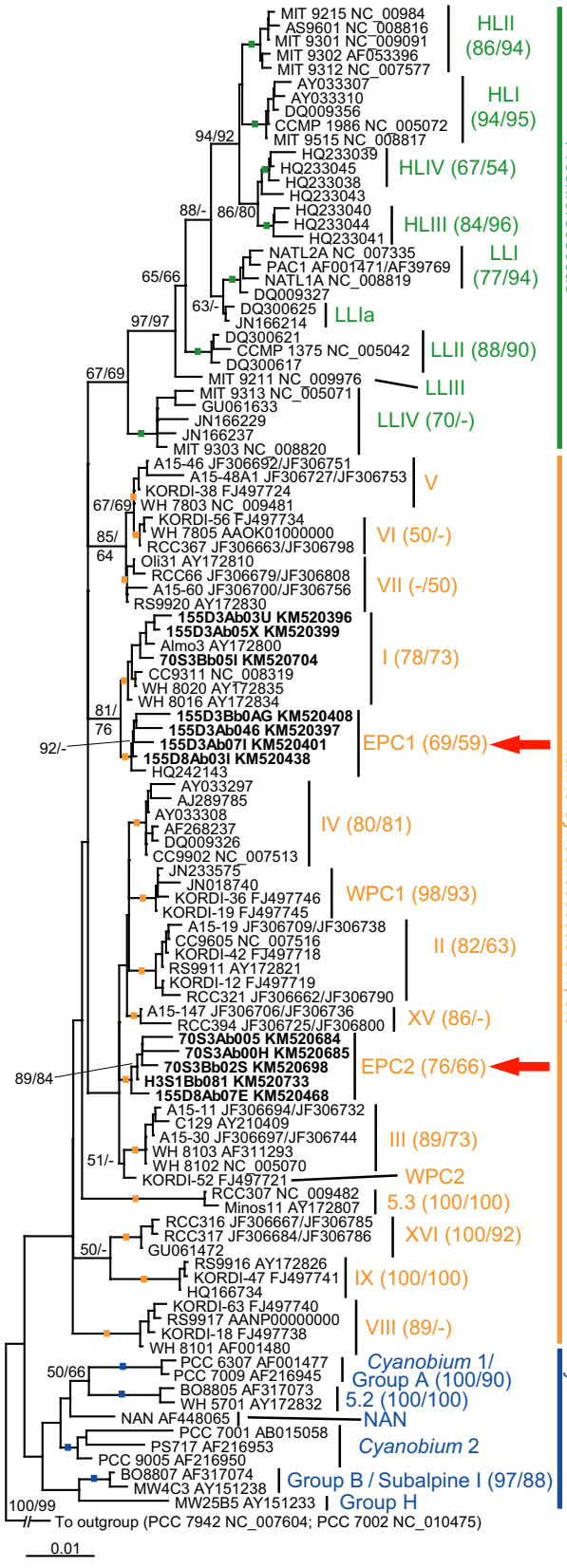

Fig. 2. Phylogenetic reconstruction of 16S rRNA gene sequences from marine cyanobacteria groups (Table S7), including representative cloned Sanger sequences from the present study (in bold). *Prochlorococcus* (green), marine *Synechococcus* (orange) and '*Cyanobium*-like' (blue, considered non-marine) clades are shown in this reference tree which was used to place bar-coded V1-V2 16S rRNA gene amplicons. The basal node for each defined clade is indicated (coloured squares). The alignment incorporated 109 gene sequences and 1292 positions were analysed using neighbor-joining (NJ) methods and the GTR + I + G substitution model. The maximum-likelihood (ML) tree (not shown) has a similar topology with the exception of *Synechococcus* clade VIII, which was not resolved. Bootstrap values > 50% for NJ/ML analyses are shown for all named clades (adjacent to clade name) as well as for branches basal to the named clades. Arrows point to new *Synechococcus* clades.

(Fig. S2) reconstructions. The majority of *Synechococcus* sequences grouped with support in clades I and IV for the coastal (21 sequences), TZ (20 sequences) and the surface and DCM open ocean sites (seven sequences). Two clades were identified that did not correspond to previously defined marine *Synechococcus* clades. Eastern Pacific Clade 1 (EPC1, named herein) was sister to *Synechococcus* clade I, and contained 11 non-identical sequences all of which came from the open ocean (14 m and 105 m, DCM). One additional North-eastern Pacific environmental sequence (accession HQ242143, Allers *et al.*, 2013) from 80 km southwest of Vancouver Island, Canada belonged to this clade. Eastern Pacific Clade 2 (EPC2, named herein), formed a sister group to clade XV, and contained 14 non-identical sequences from all three stations sampled. The evolutionary distances between each of these new bootstrap supported clades, and their nearest established neighbouring clades were consistent with those between other previously defined subcluster 5.1 clades (Fig. 2, Fig. S2). Sequences within EPC1 had high nucleotide identity to each other (98.8%) as do those in EPC2 (99.0%), comparable to identities of sequences that belong to e.g. clade I (98.7%) and II (99.1%).

The majority of *Prochlorococcus* 16S rRNA clones belonged to the HLI and LLI clades and came from the open ocean surface and DCM (105 m), respectively (Fig. S1). Members of other LL clades were also recovered, and several sequences appeared to branch deeply in the *Prochlorococcus* radiation but without a clear affiliation to established LL clades. Similarly, two environmental sequences from the North Pacific (accessions DQ300625, JN166214; DeLong *et al.*, 2006; Eiler *et al.*, 2011) appeared to form a sister group (Fig. 2) or were basal (Fig. S1) to known LLI strains. Although this clade was not supported, it was repeatedly recovered in preliminary tree building efforts (data not shown), and therefore tentatively termed '*Prochlorococcus* LLIa'.

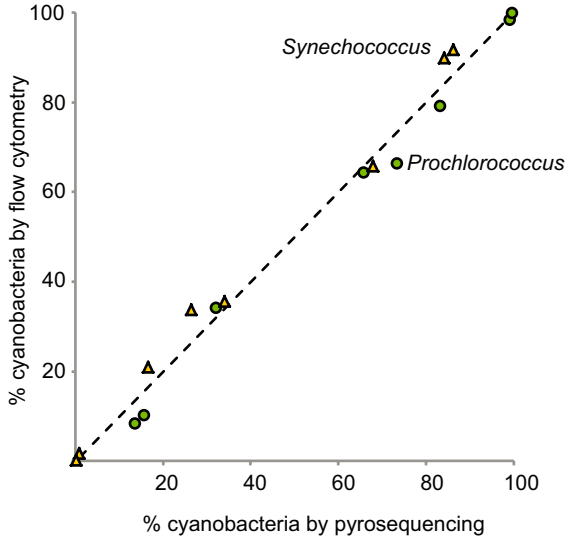

Fig. 3. Comparison of relative abundance of cyanobacteria by FCM and V1-V2 16S rRNA gene amplicon sequencing. For this analysis alone, *Synechococcus* amplicon numbers were divided by two to account for the difference in number of rRNA gene operons observed in cultured genome-sequenced *Prochlorococcus* and *Synechococcus* (Lee *et al*., 2009). The dotted line is the ideal slope of 1 which would indicate perfect agreement between the methods.

To investigate cyanobacterial diversity and distributions in this system at deeper sampling levels than possible using clone library methods, we performed bar-coded sequencing of the 16S rRNA gene V1-V2 region. The number of amplicons generated per sample ranged from 10 000 to 203 000, of which 500 to 12 000 were cyanobacterial (Table 1, Table S2). Two analysis approaches were developed and utilized. *In-silico* hybridization probes were designed to recognize individual *Prochlorococcus* and *Synechococcus* clades that could be implemented with minimal computing effort (Table S3 and supplemental material). The second approach employed hidden Markov model profiling and phylogenetic placement against our overall cyanobacterial reference alignment in PHYLOASSIGNER (Vergin *et al*., 2013). The methods were compared on eight samples (Table S2) and rendered a slope of 1.0678 ± 0.0078 [95% confidence interval 1.0525–1.0832] when plotted against each other, i.e. ~ 5–8% deviation from the 1:1 line (Fig. S3). Comparison of results from independent polymerase chain reaction (PCR) reactions also showed reproducibility (Fig. S4).

To benchmark the extent to which amplicon relative contributions reflected cell abundance, *Prochlorococcus* and *Synechococcus* were enumerated by flow cytometry (FCM; Fig. S5) and their relative percentages then compared with those from the amplicon data. Apart from *Prochlorococcus* LLIV clade strains (which formed < 2% of our amplicon data) and *Synechococcus sp.* RCC307 from clade 5.3 (which was not present in our data), genome sequenced *Prochlorococcus* and *Synechococcus* isolates have one and two copies of the rRNA operon respectively (Lee *et al*., 2009). Hence, for this particular analysis, *Synechococcus* amplicon numbers were divided by two prior to computing relative percentages. Regression analysis of relative percentages coming from FCM and amplicons rendered a slope of 0.8434 ± 0.0519 (95% CI 0.7163–0.9704) (Fig. 3), indicating the PHYLOASSIGNER analysis corresponded well to relative contributions derived from true cell counts (therefore all results hence forth come from the PHYLOASSIGNER method). We also enumerated cyanobacteria and heterotrophic prokaryotes by FCM at nine 67–155 depths (WFAD09) and computed the relative percentage of cyanobacteria to all prokaryotes. These corresponded well to the relative amplicon percentages computed for the 14 m and 105 m samples (Table S4).

From all samples together, 77 318 amplicons were assigned to the cyanobacterial radiation of which 76 432 fell within specific, recognized *Prochlorococcus* and *Synechococcus* clades (including EPC1 and EPC2 identified here, Table S2 and supplemental data file). The 2009 transect cruise dataset was the most comprehensive for evaluating contributions across the physicochemical gradient, with 26 966 amplicons assigned to the cyanobacterial radiation from the five samples sequenced (triangles; Fig. 1B). *Prochlorococcus* dominated cyanobacterial amplicons at the open ocean station, while *Synechococcus* accounted for < 2% and < 0.5% (without adjusting for 16S operon genome copy number differences) at 14 m and 105 m respectively (Fig. 4 and supplemental data file). *Prochlorococcus* HLI constituted > 96% of total cyanobacterial amplicons at the open ocean surface (14 m) while *Prochlorococcus* LLI, LLIa and LLII constituted 62%, 19% and 9%, respectively, at depth (105 m).

The TZ contained the largest number of co-located clades individually harboring > 3% of cyanobacterial amplicons. At 11 m, two *Prochlorococcus* and five *Synechococcus* clades comprised 58% and 42% of cyanobacterial amplicons, respectively. Within *Synechococcus*, clade IV was the most abundant (> 10% of the total cyanobacterial amplicons). Clades EPC1 and EPC2 comprised 5% and 10% of cyanobacterial amplicons, respectively, while amplicons for clade V/VI/VII accounted for 10%. *Prochlorococcus* HLI and LLI formed 40% and 18% of amplicons, respectively. Deeper in the water column (41 m), *Prochlorococcus* LLI, LLIa and LLII amplicons were abundant (47%, 13% and 7% respectively). *Synechococcus* I and IV comprised 10% and 16% of amplicons, whereas EPC1 and EPC2 formed < 2% of cyanobacterial amplicons at this depth. Unlike the TZ and open ocean where cyanobacteria comprised 13% to 57%

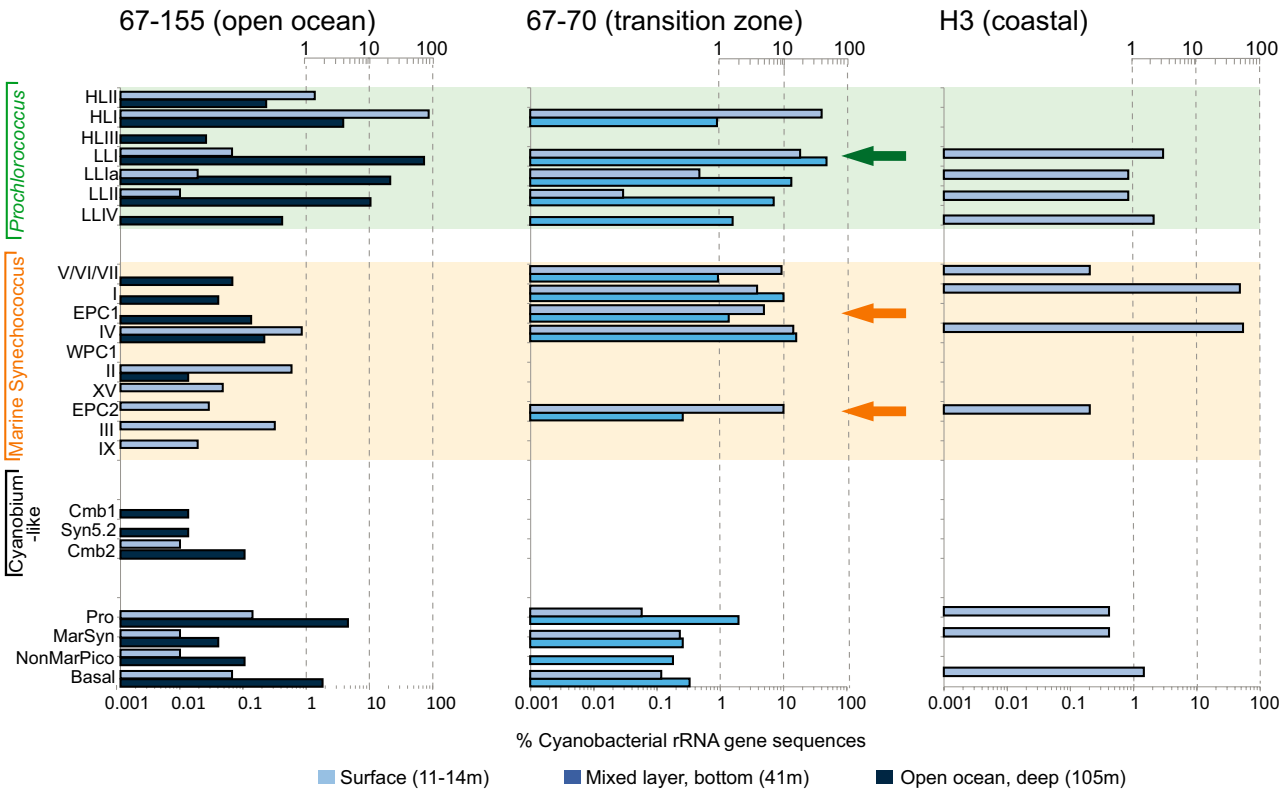

Fig. 4. Relative amplicon abundance of picocyanobacterial clades observed at the three study sites in 2009. Arrows point to new *Synechococcus* clades (orange) and emphasize amplicon abundance of *Prochlorococcus* LLI in TZ surface waters (green; see main text). Sequences were not detected from *Synechococcus* clades VIII, XVI, WPC2 and 5.3 (WPC1 was only detected in CN207 samples; Fig. S6), *Prochlorococcus* clades HLIV and LLIII and *Cyanobium* groups B, H and NAN. If sequences fell within the *Prochlorococcus* region of the tree (Fig. 2), but could not be assigned to a terminal clade (denoted by colored boxes), they were termed 'Pro', the same was done for the Marine *Synechococcus* 5.1, 5.3 region ('MarSyn') and the *Cyanobium*-like group. Sequences basal to the outgroup (*Synechococcus sp.* PCC 7942 and PCC 7002) and the above three lineages were termed 'basal'. Note the logarithmic scale.

of the total 16S V1-V2 amplicons, only 2% of coastal amplicons were assigned to cyanobacteria (Table S2). Thirty-two *Prochlorococcus* amplicons were assigned to LL ecotypes at the coastal station H3, where the dominant cyanobacteria were *Synechococcus* clades I (43%) and IV (48%).

Because multiple *Prochlorococcus* and *Synechococcus* clades appeared to be present at considerable numbers in the TZ, we investigated data from three subsequent years in this region. Two distinct assemblages were apparent (Fig. 5B and C; Table S5). In 2009 and 2013, the community was diverse with *Prochlorococcus* HLI as the most abundant clade (40% and 33% respectively of all cyanobacteria). In 2011 and 2012, *Synechococcus* clade IV dominated (60% and 72%, respectively), while *Prochlorococcus* HLI amplicons dropped below 20%. *Synechococcus* clade I amplicons also increased in relative abundance, rising from 3.7/3.8% in 2009/2013 to 9.0/9.7% in 2011/2012. Patterns deviated for a few clades, most notably *Synechococcus* EPC2, which ranged from 8.7% to 9.9% between 2009 and 2012, and increased to 19% in 2013. Overall, TZ cyanobacterial patterns corresponded with regime shifts detected in physico-chemical parameters. In 2009 and 2013, the surface waters were fresher (33.10 ± 0.13) and had lower nutrients (e.g. 0.34 ± 0.10 μM nitrate), while in 2011 and 2012, the water was more saline (33.33 ± 0.13) and nutrients higher (e.g. 2.21 ± 0.43 μM nitrate) (Fig. 5A and Table 1).

Some cyanobacterial amplicons were not assigned to recognized picocyanobacterial clades. In total, 453 and 181 were assigned to basal nodes of *Prochlorococcus* and *Synechococcus* groups, respectively (e.g. the node immediately prior to the split of HLI and HLII, Fig. 4 and supplemental data file). Twenty amplicons were assigned to a node basal to '*Cyanobium*-like' taxa of the reference tree, and 14 others were in nodes within this region. Finally, 200 amplicons were recovered that were assigned to the node immediately prior to the split of marine *Synechococcus* and '*Cyanobium*-like' lineages (supplemental data file). These were clustered at the 99% identity level using cd-hit (Li and Godzik, 2006). A representative

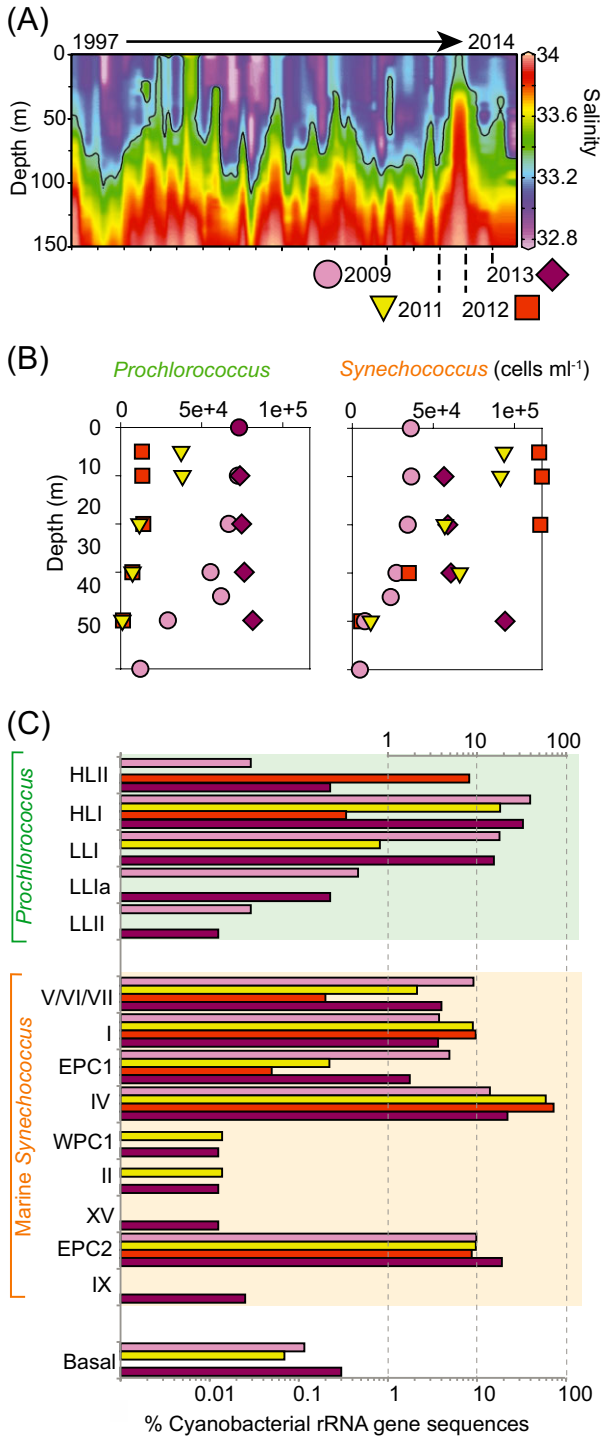

Fig. 5. A. Salinity at station 67–70 over 17 years of quarterly measurements. The 33.2 contour line is the typical border between 'saltier' upwelled water and less salty California Current water. Years of DNA sequencing presented here are denoted by dashed lines. B. Depth profile of *Prochlorococcus* and *Synechococcus* abundance by FCM for four DNA sampling years. C. Relative amplicon abundance of cyanobacterial clades in TZ surface waters (5 to 11 m) in four separate years (2012 was sampled 100 km closer to shore). Note the logarithmic scale.

of each cluster containing more than one amplicon (singletons were not considered further) was used as a BLASTN query against the National Center for Biotechnology Information (NCBI) non-redundant database with environmental sequences excluded. Clusters resulting from the 200 basal sequences recovered only low identity hits suggesting they may represent chimeras or sequencing artefacts. The basal '*Cyanobium*-like' sequences resulted in five clusters (Table S6), three of which had 98% identity to *Cyanobium sp.* Suigetsu-CR5, a phycoerythrin-containing isolate from a saline lake (Ohki *et al.*, 2012).

Discussion

We examined cyanobacterial diversity at three stations along a physico-chemical gradient from coastal to open ocean waters of the Northeastern Pacific (Fig. 1). A comprehensive new picocyanobacterial 16S rRNA gene reference tree was developed and used to classify 16S V1-V2 amplicons (Fig. 2). Both the *in-silico* probes and pplacer implemented in PHYLOASSIGNER (Matsen *et al.*, 2010; Vergin *et al.*, 2013) treat each amplicon individually. Thus, sequences were not ever clustered into operational taxonomic units (OTUs), which can be phylogenetically incoherent (Koeppel and Wu, 2013). The analyses herein provided a high-resolution view of picocyanobacterial 16S rRNA amplicon diversity in this ecologically important boundary current system.

Eight established *Prochlorococcus* clades and 14 clades of *Synechococcus* could be discriminated with statistical confidence. Resolution was only lacking for discriminating between clades V, VI and VII (treated herein as one clade), which did not individually acquire bootstrap support in our analysis although they have in prior studies (Ahlgren and Rocap, 2012) (Fig. 2, Fig. S2). The Line 67 amplicons were assigned to seven *Prochlorococcus* and 10 *Synechococcus* clades (Figs 4 and 5; Fig. S6). General trends observed in relative percentages of *Prochlorococcus* and *Synechococcus* as well as overall cyanobacteria and heterotrophic bacteria in amplicon and FCM data agreed well (Fig. 3 and Table S4). Interestingly, by FCM the percentage of *Prochlorococcus* (13–50%, $n = 9$; Table S4) relative to total prokaryotes in the euphotic zone at the open ocean site was considerably higher than in an earlier study in the equatorial Pacific (16%, $n = 48$) (Binder *et al.*, 1996). Different *Prochlorococcus* or *Synechococcus* clades cannot be distinguished by FCM and therefore, although we saw excellent correspondence of the relative contributions of *Prochlorococcus* and *Synechococcus* when comparing FCM and the amplicon approach, we could not verify that amplicon abundances directly correspond to relative clade abundance. The validity of clade abundance

numbers derived from relative numbers in clone libraries of *Prochlorococcus* has been shown by comparison with qPCR data (Martiny *et al*., 2009).

The overall trend of an open ocean station dominated by *Prochlorococcus*, a mixed community at the TZ station, and almost exclusively *Synechococcus* cells at the coastal station agrees with studies in other marine systems (Zwirglmaier *et al*., 2008; Treusch *et al*., 2009; Partensky and Garczarek, 2010; Taylor *et al*., 2012; Fig. 4). Trends were also similar to those reported for specific clades at other study sites (see e.g. Moore *et al*., 1998; Johnson *et al*., 2006; Ahlgren and Rocap, 2012; Huang *et al*., 2012), namely abundant quantities of *Prochlorococcus* HLI at this latitude, *Prochlorococcus* LL strains at depth and the coastal marine *Synechococcus* clades I and IV co-occurring as the dominant members of the *Synechococcus* community (Zwirglmaier *et al*., 2007; 2008; Tai and Palenik, 2009; Ahlgren and Rocap, 2012).

Thus, our findings corresponded well with previous observations on *Prochlorococcus* ecotype and *Synechococcus* clade distributions, with several notable exceptions. First, we identified two new *Synechococcus* clades, termed EPC1 and EPC2 (Fig. S2). Markers such as the ITS or protein-coding genes allow delineation of a greater number of *Synechococcus* clades than the 16S rRNA gene, some of which do not appear to be in culture (Choi and Noh, 2009; Paerl *et al*., 2011; Ahlgren and Rocap, 2012; Huang *et al*., 2012; Mazard *et al*., 2012). Thus, it is possible that EPC1 and EPC2 represent 16S rRNA gene sequences recovered from clades that have previously been defined using other loci, particularly the nitrate reductase *narB*-based TZ clades characterized from water sampled during the same cruise (CN207) at the same station (Paerl *et al*., 2011). Connections between some of the clades defined using different markers is still largely unresolved, but each demonstrates extensive diversity within the marine *Synechococcus* group (Ahlgren and Rocap, 2012; Mazard *et al*., 2012).

Synechococcus EPC1 and EPC2 had very different distributions from their sister clades (I and XV respectively; Figs 4 and 5). Clade I was abundant at the coast where EPC1 appeared to be absent; EPC2 was generally more abundant than clade XV, in particular at the TZ station. The annual sampling in the TZ also showed clade I percentages were higher in 2011/2012 relative to 2009/2013, while the opposite was true for EPC1. The observed differences in amplicon distributions of EPC1 and EPC2 compared with established clades supported their introduction as novel clades (and potentially ecotypes) of marine *Synechococcus* subcluster 5.1. *Synechococcus* EPC1 and EPC2 contributions to total cyanobacterial amplicon counts were $< 0.2\%$ at the coast and open ocean sites (Fig. 4). However, in the TZ they collectively represented 33% and 5% of *Synechococcus* amplicons at the surface and base of the euphotic zone, respectively, and thus are potentially important in relatively high nutrient, high chlorophyll TZ waters at eastern continental margins. For 2007, we also analysed size-fractionated samples at 67–155. In these, EPC2 along with most other cyanobacterial clades were more abundant in the 0.8 to 3.0 μm size fraction than the > 3.0 μm fraction, while EPC1 was 30 times more abundant in the > 3.0 μm fraction (CN207 cruise; Fig. S6). This could indicate association with a larger cell type, as seen for *Synechococcus*-like symbionts of unicellular eukaryotes reported in several oceans (Foster *et al*., 2006). However, sequences in the latter study are from a region of the 16S rRNA gene that has 100% identity to multiple clades, including EPC1. Notably, *Synechococcus* cells have also been observed in aggregates alongside heterotrophic bacteria (Malfatti and Azam, 2009). Our results could therefore also be interpreted as indicating that the *Synechococcus* in such aggregations come from specific clades (such as EPC1). Further studies will be needed to determine whether EPC1 cells are symbionts, aggregate formers, or simply larger in size than most other marine *Synechococcus*. Such studies will also help determine if EPC1 and EPC2 are endemic to the central California Current System or more widespread.

The relatively high contributions of *Prochlorococcus* LLI in this North Pacific region was particularly interesting. While originally characterized as a low-light ecotype, LLI has been observed in surface waters (Johnson *et al*., 2006; Hewson *et al*., 2009; Malmstrom *et al*., 2010). Low-light I has been reported at 2°N in the Pacific, in an area with elevated surface nitrate levels, while it has not been observed further north or south in a transect that sampled sites where nitrate was undetectable (Hewson *et al*., 2009). At the open-ocean time series stations in the Pacific (ALOHA) and Atlantic (BATS), LLI is restricted to deeper depths when the water column is stratified. However, contrary to other LL strains, it is present throughout the euphotic zone in the winter (Malmstrom *et al*., 2010). It has been proposed that deep mixing events bring LL strains to surface waters where only LLI is then able to persist. Moreover, in a laboratory study, strains representing different LL *Prochlorococcus* ecotypes were not able to grow after high-light pulses with the exception of LLI (Malmstrom *et al*., 2010). Our field data suggests that other factors may also be involved. At station 67–155, LLI was largely restricted to the deep sample, as expected (Fig. 4). However, in the surface waters at the TZ station, 67–70, it constituted 18% of all cyanobacterial amplicons, surpassed only by HLI (40%). We propose that LLI abundance is strongly influenced by nutrient regimes. More specifically, it likely

has higher nutrient requirements than HL ecotypes but tolerates a greater range of light levels than other LL ecotypes, and potentially a broader range of temperatures than HL ecotypes (LLI at 11.6–16.4°C in this study, BATS winter surface temperature ~ 22°C, http://bats.bios.edu). This could explain the virtual absence of LLI from the more oligotrophic 67–155 surface but not deep waters. The transect under study (Line 67) is not phosphate deplete but often has low or potentially limiting concentrations of iron (Biller and Bruland, 2014), ammonium and nitrate (see e.g. Table 1). While *Prochlorococcus* has long been known to grow on ammonium, it appears that some ecotypes also use nitrate, as do many *Synechococcus* (Martiny *et al*., 2009). Laboratory studies have shown that LLI can grow at lower temperatures and higher light levels than LLIV (Zinser *et al*., 2007). Our data support these observations, and extend the temperature tolerance shown in culture (14–27°C) down to 11°C. They also indicate that this ecotype may perform particularly well under higher nutrient conditions – closer to those used by *Synechococcus* – than required by other *Prochlorococcus* ecotypes found at the surface.

Annual sampling at 67–70 suggested an upper limit to the nutrient concentration effect on LLI distribution (Fig. 5). In the two nutrient poorer California Current-influenced years (2009/2013), LLI amplicons were over an order of magnitude more abundant than in the two nutrient richer upwelling-influenced years (2011/2012) akin to other *Prochlorococcus* clades. Although based on only four time points, the observed distributions suggested that LLI struggled to compete with *Synechococcus* under conditions associated with the upwelling regime (Fig. 5C). We also discriminated a clade that typically sistered LLI, with evolutionary distances that suggested differentiation at the clade level. Low-light Ia could correspond to clades such as LLV, VI or NC1 recently described based on the 16S–23S rRNA internal transcribed spacer region (Martiny *et al*., 2009; Lavin *et al*., 2010), or potentially represent artefacts of sequencing or phylogenetic reconstruction. However, differential distributions were observed between LLIa and LLI across the years and sites in amplicon data.

In contrast to many studies with higher detection limits (related to a lower amount of sequencing) or that use OTU-based methods, we found sequences belonging to most resolved *Prochlorococcus* ecotypes and *Synechococcus* clades at the open ocean station. While clear dominants emerged from our analyses, several unexpected groups were detected as small percentages (0.01% to 1%) of the total cyanobacterial amplicons (Fig. 4; Fig. S6). For example, *Prochlorococcus* HLII and HLIII have primarily been observed in the tropical open ocean (Zwirglmaier *et al*., 2008; West *et al*., 2011; Malmstrom *et al*., 2013), but both were detected at this higher latitude site. Low level detection was presumably facilitated by having > 20 000 *Prochlorococcus* amplicons from 67–155.

Relatively high diversity was also observed in the TZ in 2009. The lower number of cyanobacterial amplicons retrieved here made it impossible to differentiate between lower detection sensitivities and 'absence'. Likewise, for the coastal station H3, low amplicon abundance was problematic for interpreting absence of specific clades. Nevertheless, *Prochlorococcus* was detected at H3, representing 13.5% of the total cyanobacteria (corrected for the difference in operon numbers between *Prochlorococcus* and *Synechococcus*). This led us to re-evaluate our prior flow cytometry analysis and indeed a diffuse set of cells, different from the very coherent populations observed farther offshore, were present in the window where *Prochlorococcus* is expected, amounting to 1938 cells ml^{-1} or 9.1% of the total cyanobacteria enumerated. We also identified low levels of *Synechococcus* clade XV in the open ocean 14 m sample, similar to reports showing low abundances in coastal and oceanic Atlantic and Western Pacific waters, suggesting a globally distributed but low abundance group (Ahlgren and Rocap, 2012; Huang *et al*., 2012). Low percentages of *Synechococcus* clade III amplicons were also seen, a clade typically observed in warm oligotrophic waters (Zwirglmaier *et al*., 2008; Ahlgren and Rocap, 2012). Although overall *Synechococcus* distributions were consistent with previous reports, with the exception of clade XVI as well as hypersaline clade VIII and clade 5.3, the PhyloAssigner analysis recovered representatives of all statistically supported 16S rRNA clades for marine *Synechococcus* (i.e. with V/VI/VII treated as one clade) in the open ocean and/or TZ samples. It is unclear if clades represented at low percentages were advected from elsewhere, for instance through eddies at the edge of the California Current or are residents. Sampling at higher frequency will help to address this question.

PhyloAssigner places amplicons at the nearest basal node when placement on a distal branch is not supported. Such assignments do not necessarily reflect ancestral status of the amplicon in question but rather point to possibly unrecognized diversity or clades not represented in the reference tree. Analysis of these amplicons using phylogenetic methods yielded ambiguous results (not shown) because there was little or no bootstrap support, presumably due to the short amplicon length as well as relatively high 16S rRNA gene similarity across the picocyanobacteria. For the basal *Prochlorococcus* amplicons recovered herein, it is interesting to note that 64% were recovered from the open ocean deep water sample suggesting that further studies using full-length sequencing methods and targeted genomic or cultivation

efforts might reveal novel *Prochlorococcus* diversity. The amplicons that were assigned to basal nodes prior to the split between marine *Synechococcus* and '*Cyanobium*-like' lineages, and even within the '*Cyanobium*-like' clade, are enigmatic. Classically described as freshwater or brackish lineages (Herdman *et al*., 2001), picocyanobacteria related to *Cyanobium spp.* are found in freshwater and athalassohaline lakes as well as near-shore coastal and brackish/estuarine waters (Crosbie *et al*., 2003; Ernst *et al*., 2003; Haverkamp *et al*., 2009; Callieri *et al*., 2013) but rarely in oceanic waters (Everroad and Wood, 2006). Some amplicons may represent difficult to detect PCR artefacts such as chimeras of two picocyanobacterial templates. Additionally, the extensive diversity of non-marine picocyanobacteria was not thoroughly represented in our reference trees, thus some of these amplicons could be related to previously reported euryhaline and marine strains (Everroad and Wood, 2006; Haverkamp *et al*., 2008; Callieri *et al*., 2013). Interestingly, amplicons assigned to the basal ('*Cyanobium*-like') nodes in our study came from the TZ or open ocean stations, yet had the highest identity to organisms isolated from lakes. Thus these may represent organisms that further blur the line between freshwater and marine cyanobacterial lineages.

The combination of sequencing depth and phylogenetic resolution applied in this study allowed us to study broad trends in marine picocyanobacterial distributions. These trends connected to inter-annual variations in the dynamic TZ, and general differences between the marine environments investigated. The approach used also facilitated low-level detection which should help in future understanding of advection, potential seed populations and overall distributions of distinct ecotypes. The identification of two new environmental *Synechococcus* clades and finer aspects of *Prochlorococcus* LLI distributions shed light on their niches and provide testable hypotheses on differentiation.

Experimental procedures

Oceanographic sampling

Transects were performed in September 2007 (CN207) and October 2009 (WFAD09) in the Northeastern Pacific onboard the R/V Western Flyer. Deoxyribonucleic acid (DNA) samples were collected at three main stations along CalCOFI Line 67 (Fig. 1, Fig. S5A, Table 1), H3 (coastal, 11 m), 67–70 (TZ, 11 and 41 m) and 67–155 (open ocean, 14 and 105 m) using Niskin bottles mounted on a conductivity/temperature/salinity (CTD) rosette. Additionally, station 67–70 was sampled at 5 m in October of 2011 and at 10 m in 2013. In 2012, we sampled station ESP1, 100 km from 67–70. Large (CN207, 50 l) or small-scale (WFAD09 and later, 1 l) DNA samples were collected. The former were collected as described previously (Monier *et al*., 2012), and the latter were filtered onto a 0.2 μm Supor filter (Pall Gelman, East Hills, NY, USA). Filters were flash frozen in liquid nitrogen and stored at −80°C. During WFAD09, FCM depth profiles were collected at 16 stations, preserved in 0.25% electron microscopy grade glutaraldehyde, fixed at room temperature in the dark for 20 min and flash frozen in liquid nitrogen. Corresponding nutrient samples were taken as described previously (Pennington and Chavez, 2000). Ammonium was measured on CN207 and WFAD09 according to Plant *et al.* (2009), and on the remaining cruises according to Smith *et al.* (2014). Depth of the surface mixed layer was calculated by determining the shallowest depth at which the potential density (sigma theta) difference between consecutive measurements exceeded 0.01 kg m^{-3}. Maps and sections were prepared with Ocean Data View 4.6.0 (Schlitzer, R., Ocean Data View, http://odv.awi.de, 2012).

Flow cytometry

Samples were analysed on an Influx cell sorter (BD, San Jose, CA, USA) with SPIGOT 5.3.9 software, a 488 nm 200 mW laser and triggering on forward angle light scatter. Forward angle light scatter, pulse width, side scatter (90° angle, SSC), red (692 ± 40 nm band-pass filter) and orange (572 ± 27 nm band-pass filter) autofluorescence were recorded. Fluorescent polystyrene beads were added immediately prior to each run (0.75 μm yellow-green, Polysciences, Inc). Each sample was pre-run 2 min and then run 8 min to 10 min at ~ 25 μl min^{-1}, as measured with an inline flow meter (SENSIRION SLG-1430 run with software designed by Jarred Swalwell, University of Washington). To count heterotrophs, samples were stained with Hoechst33342 (Life Technologies, Grand Island, NY, USA) and analysed as described (Binder *et al*., 1996). Listmodes were analysed in WINLIST 6.0 (Verity Software House, Topsham ME, USA).

DNA extraction

CN207 extracts were attained using a sucrose lysis protocol (Monier *et al*., 2012), those sequenced here were from three size fractions (0.1–0.8 μm, 0.8–3.0 μm and 3.0–20 μm) which had been filtered through a pore size of the upper side of each range and onto a filter with a pore size at the lower end of each range. Other DNA samples were extracted from the ≥ 0.2 um fraction using a modification of protocols in the QIAGEN DNeasy kit (Demir-Hilton *et al*., 2011).

PCR and sequencing

Near full-length 16S rRNA gene sequences were generated by PCR. Deoxyribonucleic acid from CN207 fractions was amplified using 27F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 1391R (5′-GACGGGCGGTGWGTRCA-3′) universal primers, cloned into pCRII-TOPO (Invitrogen) and Sanger sequenced bidirectionally using plasmid primers M13F, M13R and internal primer 907R (5′-CCGTCAATTCCTTT RAGTTT-3′).

For pyrosequencing of the V1-V2 16S rRNA gene region, DNA from CN207 station 67–155, 5 m (filter sizes 0.8–3.0 and 3.0–20 μm) and all later cruises (Table S1) was quanti-

fied with the QuBit dsDNA high-sensitivity assay (Life Technologies) and diluted to 5 ng μl^{-1} with TE pH 8. Polymerase chain reactions (50 µl) were set up as previously described (Hamady *et al.*, 2008) with 5 µl of 10× buffer, 1 U of HiFi-Taq, 1.6 mM $MgSO_4$ (Life Technologies), 5 ng of template DNA, 200 nM each of reverse primer 454_338RPL 5′-GCCTTGCC AGCCCGCTCAGTgcwgccwcccgtaggwgt-3′ (Eurofins MWG Operon, Huntsville AL, USA) and forward primer 454_ 27FB_xxxxxxxx 5′-GCCTCCCTCGCGCCATCAG_xxxxxxxx _agrgttygatymtggctcag-3′, where xxxxxxxx represented an eight nucleotide sequence tag, specific for each sample. The primers contained 454-FLX linker sites (in capitals) as specified in the 454 protocol (Roche, South San Francisco, CA, USA). The PCR cycling parameters were 94°C for 2 min; 30× 94°C for 15 s, 55°C for 30 s, 68°C for 1 min and a final elongation at 68°C for 7 min. The PCR reaction was purified with the MinElute kit (QIAGEN, Valencia CA, USA) and an aliquot run on a 1.2% Agarose gel to verify the desired product (~ 300 nt in length) had been amplified and that primer-dimer had been removed. Twelve samples per plate were sequenced using 454-Titanium chemistry.

16S rRNA gene amplicon quality control

603 364 amplicons were obtained from 11 samples and quality controlled essentially as described previously (Hamady *et al.*, 2008). Briefly, amplicons were removed if at least one of the following was true: (i) containing unspecified base calls; (ii) not containing forward and reverse primers; (iii) not containing a sequence tag or (iv) having an average quality score less than 25. This procedure yielded the final dataset of 508 781 amplicons (Table S2; cyanobacterial sequences were deposited in the NCBI-Short Read Archive under project accession number SRP045793).

Sequence curation and reference tree construction

Sanger sequences were automatically assembled using the PHRED/PHRAP/CONSED Package (Ewing and Green, 1998; Ewing *et al.*, 1998; Gordon, 2004). Sequences were directionally oriented by searching for PCR and internal sequencing primers and orienting accordingly. Sequences were then trimmed to continuous Q20 regions, and checked for quality manually using CodonCode Aligner (CodonCode Corporation, Dedham MA, USA) with low-quality internal bases visually confirmed from the trace files and either accepted or recoded as ambiguous. The number of total 16S rRNA sequences that passed these quality control steps for each station was 1210 (H3, 11 m), 1289 (67–70, 11 and 41 m) and 2397 (67–155, 14 and 105 m). Next, the *Synechococcus* WH 7803 16S rRNA gene sequence was added to each dataset; the sequences were then aligned using CLUSTALX 2.0.5 (Larkin *et al.*, 2007) and putative cyanobacterial sequences that grouped with WH 7803 were extracted using the output neighbour-joining (NJ) guide trees. To confirm no additional cyanobacterial sequences were missed, 10–30% of the remaining sequences or at least one representative sequence from each unique cluster recovered in the output CLUSTAL guide tree was putatively identified using BLAST (Zhang *et al.*, 2000), resulting in 786 cyanobacterial and plastid sequences retained (72 from H3, 115 from 67–70 and 599 from 67–155). These sequences were further refined using BLAST and additional phylogenetic analyses to separate the plastid sequences from cyanobacterial ones. The resulting cyanobacterial 16S rRNA gene sequences of a minimum length (> 1265 bp) were checked for chimeras using Chimera Check at the RDPII database and with BELLEROPHON and MALLARD (Huber *et al.*, 2004; Ashelford *et al.*, 2005). Borderline sequences were further checked manually by breaking up the sequences and comparing BLAST results of the various fragments. The final set of cyanobacterial sequences that passed these filtering steps and were used in subsequent phylogenetic reconstructions numbered 345 (25 from H3, 33 from 67–70 and 287 from 67–155). These sequences were deposited in NCBI-GenBank under accessions KM520394–KM520738.

Preliminary NJ analyses of the curated cyanobacterial sequences were used to identify sequences derived from *Prochlorococcus* and marine *Synechococcus*. Unambiguous *Prochlorococcus* sequences ($n = 272$) were binned into operational taxonomic units with a cut-off of 99% using MOTHUR (Schloss *et al.*, 2009). One representative sequence from each *Prochlorococcus* OTU and all unique marine *Synechococcus* sequences were added to respective reference alignments using CLUSTALX. The final *Prochlorococcus* alignment consisted of 1252 characters excluding positions containing internal gaps, and contained 90 sequences, including 55 study sequences representing the 55 identified OTUs. The final marine *Synechococcus* alignment consisted of 1282 characters excluding positions containing gaps, and contained 153 sequences, including 73 study sequences. A best fit model of evolution for each alignment was determined using the likelihood ratio test implemented with JMODELTEST version 0.1 (Posada, 2008), and corresponded to GTR + I + G for both alignments. A phylogenetic tree specific to each lineage was constructed by NJ in Phylogenetic Analysis Using Parsimony (PAUP)* (Sinauer Associates, Sunderland MA, USA; Swofford, 2003) using the maximum likelihood estimator of distance as determined by JMODELTEST. Neighbour-joining trees were bootstrap pseudoreplicated 1000 times. Maximum likelihood (ML) trees using the GTR + I + G model were also constructed with PHYML and bootstrap replicated 100 times (Guindon *et al.*, 2005).

Subsequently, a picocyanobacterial 16S rRNA gene reference tree was built using representative sequences from the described clades for marine *Synechococcus* 5.1 (clades I-IX, XV, XVI, WPC1 and WPC2), 5.2 and 5.3, *Prochlorococcus* (HLI-IV, LLI-IV), and closely related picocyanobacteria (*Cyanobium* clades 1, 2, I and H). To minimize bias in sequence selection for the reference tree, a simple objective set of steps was performed: for each clade, 4–7 unique sequences of known phylogenetic affiliation based on previously published reports were selected (see Fig. 2 for accession numbers and Table S7 for clade name references). For some poorly represented clades (with fewer than four published known sequences, e.g. *Prochlorococcus* LLIII), the known sequence(s) were used for BLAST analysis with the output distance trees used to potentially identify additional clade candidates. For the presently identified clades EPC1

and EPC2, all unique sequences (11 and 14, respectively, see Fig. S2), were compared by sequence similarity matrices. From each set, no less than five sequences were selected by (i) selecting the two sequences with the largest pairwise similarity difference to catch overall diversity (ties retained); (ii) selecting the sequence with highest mean similarity to the others to catch the 'average sequence' (ties broken randomly) and (iii) selecting the sequence with lowest mean identity to others that was not retained in step 1 to guarantee further diversity was captured. Finally, to clade EPC1, the single previously sequenced representative identified from the BLAST analyses above (accession HQ242143) was added.

Sequences were aligned using the online SILVA incremental aligner (Pruesse *et al*., 2012). The alignment was visually inspected and manually edited using BIOEDIT Sequence Alignment Editor v. 5.0.9 (Hall, 1999). The alignment, excluding positions containing gaps, consisted of 1292 characters and contained 109 sequences, including several environmental sequences from freshwater *Synechococcus.* Model selection and tree construction were performed as with the *Prochlorococcus* and S*ynechococcus* trees. The best fit model of sequence evolution for the guide tree was GTR + I + G. For all phylogenetic analyses, *Synechococcus* PCC 7002 and PCC 7942 were used as outgroup taxa.

In-silico *probes*

In-silico probes were designed manually using the reference tree sequence alignment. Each probe was designed to be clade specific to the exclusion of all other sequences in the alignment. To achieve this, some probes consisted of two non-contiguous parts. To assess specificity, probes were submitted to the TESTPROBE service on the ARB-SILVA website (Quast *et al*., 2012), and we verified that all hits at the 100% level fell into the respective clade. A short bash shell script based on the 'grep' command (see supplemental material) facilitated filtering and tabulating all amplicons according to sample and clades.

PhyloAssigner pipeline

Sequenced amplicons from the 454-platform were analysed phylogenetically using PHYLOASSIGNER version 089 (Vergin *et al*., 2013), which performs a profile alignment of amplicons to a multiple sequence alignment using HMMER (Eddy, 2011) and assigns phylogenetic positions in an unmasked reference tree based on maximum likelihood methods using PPLACER (Matsen *et al*., 2010). The 508 781 sequences that passed quality control were first run against a reference alignment and tree of full-length Bacterial and eukaryotic plastid rRNA gene sequences (Vergin *et al*., 2013) and a second reference set of cyanobacterial and plastid sequences to further screen for potential plastid sequences. 77 318 amplicons mapped to the nodes for cyanobacteria. Next, we set up a PHYLOASSIGNER protocol for our cyanobacterial reference tree and alignment (see above; Fig. 2) and verified accuracy by running a test set of 42 16S rRNA gene sequences derived from characterized *Prochlorococcus* and *Synechococcus* isolates not present in our alignment that were first trimmed to the size of our amplicons. Finally, we performed a PHYLOASSIGNER run on the 77 318 amplicons using the cyanobacterial reference tree and alignment and binned sequences into the denoted clades (Fig. 2). Since clade V, VI and VII had low bootstrap support individually, they were counted into one well-supported bin 'V/VI/VII'. Amplicons assigned to nodes before the split into labeled clades were counted as 'basal' to the highest possible classification, for example 'basal *Prochlorococcus*' or 'basal HL *Prochlorococcus*'.

Cyanobacterial clade percentages

Details of FCM list mode collection and analysis are provided above. To compute the relative percentage of *Prochlorococcus* and *Synechococcus* in each sample, the cell abundances were summed for these two groups (and that number then defined as 100%) and the percent contribution of each then calculated. Genome sequenced *Synechococcus* isolates possess two copies of the rRNA operon with the exception of one clade 5.3 strain (RCC307, not present in our samples; Lee *et al*., 2009), while *Prochlorococcus* generally have one copy, with the exception of LLIV strains MIT 9313 and MIT 9303 (representing < 2% of our cyanobacterial amplicon data combined). Thus, to compute relative percentages in amplicon data, we first halved *Synechococcus* amplicon counts then summed them with *Prochlorococcus* as well as *Cyanobium*-like counts (as assigned by PHYLOASSIGNER), defined this number as being 100% for each sample and calculated the relative percent contributed by each genus. *Cyanobium*-like sequences were < 0.12% and ignored in the graphing.

Acknowledgements

We thank E. Demir-Hilton, M.P. Simmons, D. McRose, R. Welsh, L. Deng and the captain and crew of R/V Western Flyer. We are grateful to A. Zimmerman for statistical advice and to S. Giovannoni, B. Beszteri, K. Vergin, A. Monier and F. Kilpert for the initial pipeline development collaboration (Vergin *et al*., 2013) and subsequent discussions. This work was supported by grants from the Lucile and David Packard Foundation, Gordon and Betty Moore Foundation GBMF1668 and GBMF3788 and NSF 0843119 to AZW.

References

Ahlgren, N.A., and Rocap, G. (2006) Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Appl Environ Microbiol* **72:** 7193–7204.

Ahlgren, N.A., and Rocap, G. (2012) Diversity and distribution of marine *Synechococcus*: multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Front Microbiol* **3:** 213. doi:10.3389/fmicb.2012.00213.

Allers, E., Wright, J.J., Konwar, K.M., Howes, C.G., Beneze, E., Hallam, S.J., and Sullivan, M.B. (2013) Diversity and population structure of marine group A bacteria in the Northeast subarctic Pacific Ocean. *ISME J* **7:** 256–268.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., and Weightman, A.J. (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* **71:** 7724–7736.

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

Binder, B., Chisholm, S., Olson, R.J., Frankel, S., and Worden, A.Z. (1996) Dynamics of picophytoplankton, ultraphytoplankton and bacteria in the central equatorial Pacific. *Deep Sea Res Part II Top Stud Oceanogr* **43:** 907–931.

Blanchot, J., André, J.-M., Navarette, C., Neveux, J., and Radenac, M.-H. (2001) Picophytoplankton in the equatorial Pacific: vertical distributions in the warm pool and in the high nutrient low chlorophyll conditions. *Deep Sea Res Part I Oceanogr Res Pap* **48:** 297–314.

Callieri, C., Coci, M., Corno, G., Macek, M., Modenutti, B., Balseiro, E., and Bertoni, R. (2013) Phylogenetic diversity of nonmarine picocyanobacteria. *FEMS Microbiol Ecol* **85:** 293–301.

Checkley, D.M., Jr, and Barth, J.A. (2009) Patterns and processes in the California Current System. *Prog Oceanogr* **83:** 49–64.

Choi, D.H., and Noh, J.H. (2009) Phylogenetic diversity of *Synechococcus* strains isolated from the East China Sea and the East Sea. *FEMS Microbiol Ecol* **69:** 439–448.

Collins, C.A., Pennington, J.T., Castro, C.G., Rago, T.A., and Chavez, F.P. (2003) The California Current system off Monterey, California: physical and biological coupling. *Deep Sea Res Part II Top Stud Oceanogr* **50:** 2389–2404.

Crosbie, N.D., Pöckl, M., and Weisse, T. (2003) Dispersal and phylogenetic diversity of nonmarine picocyanobacteria, inferred from 16S rRNA gene and *cpcBA*-intergenic spacer sequence analyses. *Appl Environ Microbiol* **69:** 5716–5721.

DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.-U., *et al*. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311:** 496–503.

Demir-Hilton, E., Sudek, S., Cuvelier, M.L., Gentemann, C.L., Zehr, J.P., and Worden, A.Z. (2011) Global distribution patterns of distinct clades of the photosynthetic picoeukaryote *Ostreococcus*. *ISME J* **5:** 1095–1107.

Dufresne, A., Ostrowski, M., Scanlan, D.J., Garczarek, L., Mazard, S., Palenik, B.P., *et al*. (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9:** R90.

Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Comput Biol* **7:** e1002195.

Eiler, A., Hayakawa, D.H., and Rappé, M.S. (2011) Non-random assembly of bacterioplankton communities in the subtropical North Pacific Ocean. *Front Microbiol* **2:** 140. doi:10.3389/fmicb.2011.00140.

Ernst, A., Becker, S., Wollenzien, U.I.A., and Postius, C. (2003) Ecosystem-dependent adaptive radiations of picocyanobacteria inferred from 16S rRNA and ITS-1 sequence analysis. *Microbiology* **149:** 217–228.

Everroad, C., Six, C., Partensky, F., Thomas, J.-C., Holtzendorff, J., and Wood, A.M. (2006) Biochemical bases of type IV chromatic adaptation in marine *Synechococcus spp*. *J Bacteriol* **188:** 3345–3356.

Everroad, R.C., and Wood, A.M. (2006) Comparative molecular evolution of newly discovered picoyanobacterial strains reveals a phylogenetically informative variable region of β-phycoerythrin. *J Phycol* **42:** 1300–1311.

Everroad, R.C., and Wood, A.M. (2012) Phycoerythrin evolution and diversification of spectral phenotype in marine *Synechococcus* and related picocyanobacteria. *Mol Phylogenet Evol* **64:** 381–392.

Ewing, B., and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8:** 186–194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8:** 175–185.

Ferris, M.J., and Palenik, B. (1998) Niche adaptation in ocean cyanobacteria. *Nature* **396:** 226–228.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., *et al*. (2013) Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci USA* **110:** 9824–9829.

Foster, R.A., Collier, J.L., and Carpenter, E.J. (2006) Reverse transcription PCR amplification of cyanobacterial symbiont 16S rRNA sequences from single non-photosynthetic eukaryotic marine planktonic host cells. *J Phycol* **42:** 243–250.

Fuller, N.J., Marie, D., Partensky, F., Vaulot, D., Post, A.F., and Scanlan, D.J. (2003) 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* **69:** 2430–2443.

Gordon, D. (2004) Viewing and editing assembled sequences using Consed. In *Current Protocols in Bioinformatics*. Bateman, A., Pearson, W.R., Stein, L.D., Stormo, G.D., and Yates, J.R. (eds). New York, USA: John Wiley & Co., pp. 11.2.1–11.2.43.

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005) PHYML online – a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* **33:** W557–W559.

Hall, T. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41:** 95–98.

Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J., and Knight, R. (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5:** 235–237.

Haverkamp, T., Acinas, S.G., Doeleman, M., Stomp, M., Huisman, J., and Stal, L.J. (2008) Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. *Environ Microbiol* **10:** 174–188.

Haverkamp, T.H.A., Schouten, D., Doeleman, M., Wollenzien, U., Huisman, J., and Stal, L.J. (2009) Colorful microdiversity of *Synechococcus* strains (picocyanobacteria) isolated from the Baltic Sea. *ISME J* **3:** 397–408.

Herdman, M., Castenholz, R.W., Waterbury, J.B., and Rippka, R. (2001) Form-genus XIII. *Synechococcus*. In

Bergey's Manual of Systematic Bacteriology. Boone, D.R., Castenholz, R.W., and Garrity, G.M. (eds). New York, USA: Springer, pp. 508–512.

Hewson, I., Paerl, R.W., Tripp, H.J., Zehr, J.P., and Karl, D.M. (2009) Metagenomic potential of microbial assemblages in the surface waters of the central Pacific Ocean tracks variability in oceanic habitat. *Limnol Oceanogr* **54:** 1981–1994.

Huang, S., Wilhelm, S.W., Harvey, H.R., Taylor, K., Jiao, N., and Chen, F. (2012) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* **6:** 285–297.

Huber, T., Faulkner, G., and Hugenholtz, P. (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20:** 2317–2319.

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

Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., *et al*. (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344:** 416–420.

Koeppel, A.F., and Wu, M. (2013) Surprisingly extensive mixed phylogenetic and ecological signals among bacterial Operational Taxonomic Units. *Nucleic Acids Res* **41:** 5175–5188.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., *et al*. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947–2948.

Lavin, P., González, B., Santibáñez, J.F., Scanlan, D.J., and Ulloa, O. (2010) Novel lineages of *Prochlorococcus* thrive within the oxygen minimum zone of the eastern tropical South Pacific. *Environ Microbiol Rep* **2:** 728–738.

Lee, Z.M.-P., Bussema, C., and Schmidt, T.M. (2009) rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res* **37:** D489–D493.

Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22:** 1658–1659.

McClatchie, S. (2014) *Regional Fisheries Oceanography of the California Current System*. New York, USA: Springer.

Mackas, D.L., Strub, P.T., Thomas, A., and Montecino, V. (2006) Eastern regional ocean boundaries pan-regional overview. In *The Sea*. Robinson, A.R., and Brink, K. (eds). Cambridge, USA: Harvard University Press, pp. 21–60.

Malfatti, F., and Azam, F. (2009) Atomic force microscopy reveals microscale networks and possible symbioses among pelagic marine bacteria. *Aquat Microb Ecol* **58:** 1–14.

Malmstrom, R.R., Coe, A., Kettler, G.C., Martiny, A.C., Frias-Lopez, J., Zinser, E.R., and Chisholm, S.W. (2010) Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *ISME J* **4:** 1252–1264.

Malmstrom, R.R., Rodrigue, S., Huang, K.H., Kelly, L., Kern, S.E., Thompson, A., *et al*. (2013) Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J* **7:** 184–198.

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

Matsen, F.A., Kodner, R.B., and Armbrust, E.V. (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11:** 538–554.

Mazard, S., Ostrowski, M., Partensky, F., and Scanlan, D.J. (2012) Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* **14:** 372–386.

Monier, A., Welsh, R.M., Gentemann, C., Weinstock, G., Sodergren, E., Armbrust, E.V., *et al*. (2012) Phosphate transporters in marine phytoplankton and their viruses: cross-domain commonalities in viral-host gene exchanges. *Environ Microbiol* **14:** 162–176.

Moore, L.R., Rocap, G., and Chisholm, S.W. (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393:** 464–467.

Ohki, K., Yamada, K., Kamiya, M., and Yoshikawa, S. (2012) Morphological, phylogenetic and physiological studies of pico-cyanobacteria isolated from the halocline of a saline meromictic lake, Lake Suigetsu, Japan. *Microbes Environ* **27:** 171–178.

Paerl, R.W., Johnson, K.S., Welsh, R.M., Worden, A.Z., Chavez, F.P., and Zehr, J.P. (2011) Differential distributions of *Synechococcus* subgroups across the California Current System. *Front Microbiol* **2**. doi:10.3389/fmicb.2011.00059.

Palenik, B., Ren, Q., Dupont, C.L., Myers, G.S., Heidelberg, J.F., Badger, J.H., *et al*. (2006) Genome sequence of *Synechococcus* CC9311: insights into adaptation to a coastal environment. *Proc Natl Acad Sci USA* **103:** 13555–13559.

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

Paz-Yepes, J., Brahamsha, B., and Palenik, B. (2013) Role of a Microcin-C-like biosynthetic gene cluster in allelopathic interactions in marine *Synechococcus*. *Proc Natl Acad Sci USA* **110:** 12030–12035.

Pennington, T.J., and Chavez, F.P. (2000) Seasonal fluctuations of temperature, salinity, nitrate, chlorophyll and primary production at station H3/M1 over 1989–1996 in Monterey Bay, California. *Deep Sea Res Part II Top Stud Oceanogr* **47:** 947–973.

Plant, J.N., Johnson, K.S., Needoba, J.A., and Coletti, L.J. (2009) NH4-Digiscan: an in situ and laboratory ammonium analyzer for estuarine, coastal and shelf waters. *Limnol Oceanogr Methods* **7:** 144–156.

Posada, D. (2008) jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25:** 1253–1256.

Pruesse, E., Peplies, J., and Glöckner, F.O. (2012) SINA: accurate high throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* doi:10.1093/bioinformatics/bts252.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al*. (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41:** D590–D596.

Rocap, G., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* **68:** 1180–1191.

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

Santoro, A.E., Sakamoto, C.M., Smith, J.M., Plant, J.N., Gehman, A.L., Worden, A.Z., *et al*. (2013) Measurements of nitrite production in and around the primary nitrite maximum in the central California Current. *Biogeosciences* **10:** 7395–7410.

Scanlan, D.J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W.R., *et al*. (2009) Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* **73:** 249–299.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Six, C., Thomas, J.-C., Garczarek, L., Ostrowski, M., Dufresne, A., Blot, N., *et al*. (2007) Diversity and evolution of phycobilisomes in marine *Synechococcus spp.*: a comparative genomics study. *Genome Biol* **8:** R259.

Smith, J.M., Casciotti, K.L., Chavez, F.P., and Francis, C.A. (2014) Differential contributions of archaeal ammonia oxidizer ecotypes to nitrification in coastal surface waters. *ISME J* **8:** 1704–1714.

Swofford, D. (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4*. Sunderland, MA, USA: Sinauer Associates.

Tai, V., and Palenik, B. (2009) Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *ISME J* **3:** 903–915.

Taylor, A.G., Goericke, R., Landry, M.R., Selph, K.E., Wick, D.A., and Roadman, M.J. (2012) Sharp gradients in phytoplankton community structure across a frontal zone in the California Current Ecosystem. *J Plankton Res* **34:** 778–789.

Toledo, G., and Palenik, B. (1997) *Synechococcus* diversity in the California current as seen by RNA polymerase (rpoC1) gene sequences of isolated strains. *Appl Environ Microbiol* **63:** 4298–4303.

Toledo, G., Palenik, B., and Brahamsha, B. (1999) Swimming marine *Synechococcus* strains with widely different photosynthetic pigment ratios form a monophyletic group. *Appl Environ Microbiol* **65:** 5247–5251.

Treusch, A.H., Vergin, K.L., Finlay, L.A., Donatz, M.G., Burton, R.M., Carlson, C.A., and Giovannoni, S.J. (2009) Seasonality and vertical structure of microbial communities in an ocean gyre. *ISME J* **3:** 1148–1163.

Vergin, K.L., Beszteri, B., Monier, A., Cameron Thrash, J., Temperton, B., Treusch, A.H., *et al*. (2013) High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *ISME J* **7:** 1322–1332.

West, N.J., Lebaron, P., Strutton, P.G., and Suzuki, M.T. (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.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7:** 203–214.

Zinser, E.R., Johnson, Z.I., Coe, A., Karaca, E., Veneziano, D., and Chisholm, S.W. (2007) Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol Oceanogr* **52:** 2205–2220.

Zwirglmaier, K., Heywood, J.L., Chamberlain, K., Woodward, E.M.S., Zubkov, M.V., and Scanlan, D.J. (2007) Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. *Environ Microbiol* **9:** 1278–1290.

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., *et al*. (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10:** 147–161.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Phylogenetic analysis of *Prochlorococcus* 16S rRNA genes. 1252 positions were analysed using NJ methods and the GTR + I + G substitution model in an alignment of 90 sequences. The ML tree (not shown) was highly similar to the NJ tree. Near-full length Sanger sequences from transect sampling in 2007 are included, representing 55 OTUs identified (sequences in bold; details in main text). Numbers at nodes refer to NJ/ML bootstrap values > 50%.

Fig. S2. Phylogenetic analysis of marine *Synechococcus* 16S rRNA genes. 1282 positions were analysed using NJ methods and the GTR + I + G substitution model in an alignment of 153 sequences. Near-full length Sanger sequences from transect sampling in 2007 are included in bold. The ML tree was highly similar to the NJ tree shown with the exception of *Synechococcu*s clade VIII, which was not resolved in the latter. The new clades EPC1 and EPC2 are highlighted by red arrows. Numbers at nodes refer to NJ/ML bootstrap values > 50%.

Fig. S3. Comparison of PHYLOASSIGNER- and probe-based clade assignments of amplicons from the WFAD09 and CN207 cruises. The line represents the ideal slope of 1 which would indicate perfect agreement between the methods.

Fig. S4. Comparison of cyanobacterial amplicon abundance in data from two independent PCR amplifications. The line represents the ideal slope of 1 which would indicate perfect agreement between the replicates. Note the logarithmic scale.

Fig. S5. A. All Line 67 stations sampled for FCM, as well as those sampled for DNA (red). B. *Prochlorococcus* and *Synechococcus* abundance along the transect in 2009 based on FCM counts. Note differences in the X axis scale.

Fig. S6. Cyanobacterial amplicon abundance at oceanic station 67–155 surface on cruises WFAD09 and CN207. 08A and 08B are independent PCR reactions using template from the same DNA extraction. Note that while all

other samples were filtered directly onto a 0.2 μm pore-sized filter prior to extraction, the CN207 samples were handled differently. Specifically, the sample with '30' in the name was filtered through a 20 μm pore size and onto a 3.0 μm pore size filter, while the '08' samples were filtered through a 3.0 μm pore size and onto a 0.8 μm pore size filter. Thus, most heterotrophic bacteria likely passed through the CN207 filters, and consequently their templates will be under-represented.

Table S1. Physico-chemical parameters of Sanger and amplicon sequencing samples.

Table S2. Amplicon sequencing data statistics.

Table S3. Cyanobacterial *in-silico* probes.

Table S4. Comparison of relative contribution of *Prochlorococcus*, *Synechococcus* and heterotrophs.

Table S5. Clade distribution in the transition zone.

Table S6. Clusters of amplicons assigned basal to the *Cyanobium*-like clade.

Table S7. Cyanobacterial clades and their references.

List of Supplemental data files

Method section: Shell script to tabulate amplicons according to clades using probes