environmental microbiology

Environmental Microbiology (2021) **23**(1), 252–266 doi:10.1111/1462-2920.15307

Spatial and temporal variations in *Synechococcus* microdiversity in the Southern California coastal ecosystem

Maitreyi Nagarkar, Maggie Wang, Bellineth Valencia and Brian Palenik*

Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, 92093.

Summary

The *Synechococcus* cyanobacterial population at the Scripps Institution of Oceanography pier in La Jolla, CA, shows large increases in abundance, typically in the spring and summer followed, by rapid declines within weeks. Here we used amplicon sequencing of the ribosomal RNA internal transcribed spacer region to examine the microdiversity within this cyanobacterial genus during these blooms as well as further offshore in the Southern California coastal ecosystem (CCE). These analyses revealed numerous *Synechococcus* amplicon sequence variants (ASVs) and that clade and ASV composition can change over the course of blooms. We also found that a large bloom in August 2016 was highly anomalous both in its overall *Synechococcus* abundance and in terms of the presence of normally oligotrophic *Synechococcus* clade II. The dominant ASVs at the pier were found further offshore and in the California Current, but we did observe more oligotrophic ASVs and clades along with depth variation in *Synechococcus* diversity. We also observed that the dominant sequence variant switched during the peak of multiple *Synechococcus* blooms, with this switch occurring in multiple clades, but we present initial evidence that this apparent ASV switch is a physiological response rather than a change in the dominant population.

Introduction

Synechococcus is a globally important genus of cyanobacteria that is responsible for a significant amount of primary production, especially at coastal sites (Agawin *et al*., 2000; Collier and Palenik, 2003). Originally discovered partly because of its unique autofluorescence signature (Waterbury *et al*., 1979), *Synechococcus* was soon recognized to be highly ubiquitous throughout the ocean using shipboard flow cytometry (Olson *et al*., 1985) and to consist of multiple clades with differing environmental niches and capabilities (Ferris and Palenik, 1998). These different species- or strain-level populations offer insights into the adaptive strategies of marine bacteria to the wide range of changing environments they encounter in the ocean.

Currently, the genus *Synechococcus* is thought to comprise at least 18 well-defined clades which vary in genome size/composition, nutrient requirements and geographic distribution (Rocap *et al*., 2002; Zwirglmaier *et al*., 2008; Ahlgren and Rocap, 2012; Huang *et al*., 2012; Choi *et al*., 2014; Sohm *et al*., 2016). Sequence-based phylogenies have been developed using multiple genetic markers including the 16S ribosomal subunit (Rocap *et al*., 2002), which is commonly used for the classification of bacteria. Other commonly used markers such as the RNA polymerase gene (*rpoC1*) (Palenik and Haselkorn, 1992; Palenik, 1994; Toledo *et al*., 1999), photosystem I gene *psbA* (Zeidner *et al*., 2003), the nitrate reductase gene *narB* (Paerl *et al*., 2011) and parts of the internal transcribed spacer (Mazard *et al*., 2011; Ahlgren and Rocap, 2012; Huang *et al*., 2012) have been developed to provide greater taxonomic resolution. Here, we utilize a region of the internal transcribed spacer (ITS) that has been demonstrated to successfully differentiate between subtypes within the *Synechococcus* subcluster 5.1 and is also short enough to use with high-throughput sequencing (Choi *et al*., 2014). Sequencing of the ITS region has shown some of the previously well-supported clades to be robust and revealed several novel clades (Rocap *et al*., 2002; Choi and Noh, 2009; Mazard *et al*., 2011).

The temporal and spatial patterns of *Synechococcus* abundance have been characterized in a number of studies. At some sites, there is a seasonal pattern of high abundance in the summer months (Agawin *et al*., 1998; Patterson, 1998; Xia *et al*., 2015; Hunter-Cevera

Received 30 January, 2020; revised 17 September, 2020; accepted 20 September, 2020. **For correspondence. E-mail bpalenik@ucsd.edu; Tel. 1-858-534-7505; Fax 1-858-534-7313.

et al., 2016), whereas at other sites (including our own) the *Synechococcus* population exhibits short-lived blooms in the spring or summer (Tai and Palenik, 2009; Robidart *et al*., 2011). In either case, relationships between *Synechococcus* abundance and temperature (Agawin *et al*., 2000; Robidart *et al*., 2011), nitrate (Glover *et al*., 1988; Collado-Fabbri *et al*., 2011; Rajaneesh and Mitbavkar, 2013) and upwelling (Collier and Palenik, 2003) have been identified. These relationships have been found spatially as well; molecular studies have demonstrated that different clades of *Synechococcus* occupy different geographic and depth ranges. Sohm *et al*. (2016) found that different *Synechococcus* clades dominated different oceanic niches and defined four regimes based on temperature, macronutrients and iron availability. It is well known that clades I and IV tend to be dominant at high latitudes (Huang *et al*., 2012) and in coastal environments, including the coastal California ecosystem (Zwirglmaier *et al*., 2008) and the Scripps pier specifically (Tai *et al*., 2011). Many of the other marine *Synechococcus* clades, including clades II and III, have been observed more abundantly in oligotrophic, open-ocean environments (Zwirglmaier *et al*., 2008; Sohm *et al*., 2016). Certain clades also demonstrate monophyletic environmental adaptations such as motility (Toledo *et al*., 1999). The question of why certain evolutionarily-distinct clades consistently co-occur (such as clades I and IV at our site) remains, and more detailed knowledge of their growth patterns along with how they are affected by abiotic conditions will likely deepen our understanding.

Recent sequence analysis methods have indicated that microdiversity within the broader clades could be significant and that in some cases there are spatial or temporal patterns among very closely related members. This microdiversity, defined by Larkin and Martiny (2017) as subspecies-level taxa that have greater than 97% 16S rRNA gene similarity but distinct environmental niches, is identifiable using amplicon sequencing of the ITS region; Choi *et al*. (2014) found several hundred distinct representative sequences of *Synechococcus* subcluster 5.1 from oligotrophic sites in the Northwestern Pacific. Mackey *et al*. (2017) observed a consistent seasonal succession of several different *Synechococcus* oligotypes over the course of the summer at several ocean and estuarine sites in Little Sippewissett, Cape Cod, MA. This pattern included changes in the dominant clades present, but also a switch in the dominant oligotype within some of the clades.

It is not always known whether microbial strains with high sequence identity in their rRNA sequences represent actual functional diversity, though there is evidence of fine-scale patterning of phylogenetically distinct populations (Acinas *et al*., 2003; Lee *et al*., 2019). Environmental sequencing of *Prochlorococcus* shows clusters of ecotypes that have > 98% similarity in the ITS region, and single-celled whole-genome sequencing of some of these indicates that these subpopulations have distinct 'genomic backbones' (Kashtan *et al*., 2014). Even subpopulations within the same clade contained unique gene cassettes, which may confer differential adaptation in functions like redox stress response and outer membrane modification (Kashtan *et al*., 2014). Members of the same *Synechococcus* clade have also been shown to have varying genome size and GC content as well as distinct geographic distributions (Lee *et al*., 2019). Similarly, at the Scripps Institution of Oceanography (SIO) pier, detection of horizontally acquired genes that conferred copper and oxidative stress tolerance in *Synechococcus* clade I varied with time, indicating variation in the functional abilities within members of a single clade (Stuart *et al*., 2013). Much remains to be explored on the role of physiochemical forces on phytoplankton microdiversity, including the spatial scales on which patterning can be detected. Recent evidence from the Tara Oceans Expedition (Farrant *et al*., 2016) and other studies (Paerl *et al*., 2011; Gutierrez-Rodriguez *et al*., 2014) suggest that shifts in microdiversity over short geographic distances exist in the marine environment.

In this paper, we applied high-throughput sequencing of the 16S–23S ITS region (Choi *et al*., 2014) to examine of *Synechococcus* microdiversity in the Southern California Bight both temporally (at a single site, the SIO Pier) and spatially (comparing four different sampling sites in the California Current ecosystem). Our temporal analysis focuses on *Synechococcus* blooms, which have typically occurred several times per year as detected using flow cytometry. Because the causes of the onset and termination of blooms are not fully characterized, understanding shifts within-*Synechococcus* diversity may help explain some of the driving ecological processes. First, we addressed the question of whether different blooms over several years and seasons have similar temporal patterns in terms of clade- and amplicon sequence variant (ASV)-level *Synechococcus* diversity. Based on past knowledge of differential clade abundances over time at our site (Tai and Palenik, 2009; Tai *et al*., 2011), we hypothesized a change in the clade I:IV ratio between the onset and termination of a *Synechococcus* bloom, with clade I more abundant than clade IV prior to the bloom. Secondly, we examined whether the *Synechococcus* community at the pier is representative of that further offshore in the Southern California Bight. Again, we looked at both clade-level composition, wherein we expected to see oligotrophic clades further offshore, as well as in terms of the subclade-level ASVs within the dominant coastal clades. Finally, we conducted a pilot experiment to explore a variant-switching

phenomenon that we discovered from the temporal sequencing data and found we were able to alter proportions of two clade IV variants in culture by introducing nutrient stress.

Methods

Sample collection

We collected a surface seawater sample by lowering a bucket at the end of the SIO pier (32°87′N, 117°26′W) weekly (before 2014) or bi-weekly (after 2014); water from this site was processed for both DNA extraction and flow cytometry. Seawater temperature from selected sampling points is provided in Appendix S1. Additional metadata is in Appendix S3. An amount of 500 ml of seawater was filtered in triplicate onto 47 mm diameter 0.2 μm Supor (Pall corporation, Port Washington, NY, USA) membrane filters for total DNA extraction and filters were stored at −80°C until further use. For selected time points, a single filter was used for DNA extraction and subsequent sequencing, as described below. For flow cytometry counts, we preserved 1 ml seawater in 0.25% glutaraldehyde (Sigma-Aldrich, St Louis, MO, USA), incubated at room temperature for 10 min and then stored cryovials at −80°C until further use.

Samples from the California Current System were obtained during the Southern California coastal ecosystem (CCE) Long Term Ecological Research (LTER) P1604 cruise through CTD casts at different experimental cycles and depths (specified in Appendix S1 and additional metadata in Appendix S3) and processed in two different ways. Some samples were filtered shipboard as described above and others were processed slightly differently (Valencia et al., unpubl data): 280 ml (200 μm Nitex screen) or 650 ml of seawater (500 μm Nitex screen) were pre-screened to remove mesozooplankton prior to filtration through a 25 mm diameter 0.2 μm Supor membrane filters (Pall Corporation, Port Washington, NY, USA). Once the water was filtered, the filters were folded in half, placed in 2 ml screw-cap cryogenic vials, flash-frozen in liquid nitrogen and stored at −80°C until analysis.

Synechococcus *counts*

We used flow cytometry to measure *Synechococcus* abundance at each sampling point. For pier samples, the glutaraldehyde-fixed samples were thawed in the dark and then mixed with 18 μl of 0.94 μm green fluorescent beads (Duke Scientific Corporation, Palo Alto, CA, USA). These were then run on a BD FACsort (Beckton Dickinson, Franklin Lakes, NJ, USA) for approximately 5 min at the highest flow setting. We gated a specific region on the flow cytograms to represent *Synechococcus* cells and calculated *Synechococcus* density by normalizing the event counts to the volume of sample run and the counts of beads. Samples were typically processed 1–6 months after they were collected. *Synechococcus* counts from sampling dates used for amplicon sequencing are shown in Appendix S1.

Samples from the CCE LTER P1604 cruise were treated slightly differently. They were preserved in paraformaldehyde and processed as described in Selph *et al*. (2011).

DNA extraction of environmental samples

We extracted DNA from 21 different pier samples and 20 cruise samples. For the pier samples, we selected time points before, during and after the peaks of four different blooms (spring 2011, summer 2011, summer 2012 and summer 2016) based on the *Synechococcus* densities determined using flow cytometry. For the cruise, we selected samples from the surface, mixed layer and at lower depths at each of the four sites based on CTD cast data. Specific dates and sites sequenced can be found in Appendix S1 and Appendix S3. For the selected pier samples, the frozen filters were cut into small pieces using a clean razor and the pieces were divided among two 2 ml microcentrifuge tubes. An amount of 560 μl of TE (50 mM Tris, 20 mM Ethylenediaminetetraacetic acid (EDTA) and 80 μl 100 mg/ml lysozyme were added to each tube and tubes were incubated for 30 min at 37°C. Subsequently 80 μl of 10% sodium dodecyl sulfate and 80 μl of 10 mg/ml proteinase K were added and tubes were incubated for 2.5 h at 55°C. Then 16 μl of 10 mg/ml RNAse A was added and tubes were incubated an additional 30 min at 37°C. An equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added to each sample and the aqueous layer was pipetted into a new tube; this was repeated once and then once with chloroform: isoamyl alcohol (24:1). Finally, DNA was eluted using the Qiagen DNEasy Blood and Tissue kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions (with reagent volumes scaled up when appropriate). DNA was stored at −20°C until it was sent out for amplicon sequencing.

Environmental DNA from the remaining cruise samples was extracted using the NucleoMag 96 Plant kit (Macherey Nagel) following the manufacturer's instructions. DNA was eluted to 50 μl and was stored at −80°C until it was sent out for amplicon sequencing. While this sample processing differed between pier and cruise samples due to methodological constraints, hereafter all environmental samples were sequenced and analysed in the same manner.

Amplicon sequencing of environmental samples

For all the environmental samples described above, sample DNA was sent to RTL Genomics (Lubbock, TX, USA) for sequencing on an Illumina MiSeq at a depth of 10K reads, using primers from Choi *et al*. (2014) to amplify the ITS region (F: GGATCACCTCCTAACAGGGAG, R: AGGTTAGGAGACTCGAACTC). Samples were sequenced according to the RTL Genomics protocol, described in Appendix S1.

Sanger sequencing of rpoC1 *for selected environmental samples*

As a point of comparison with the characterization of the *Synechococcus* environmental communities using the ITS amplicon, we made clone libraries of the RNA polymerase C (*rpoC1*) gene for four of the pier samples (07 July 2011, 21 July 2011, 26 July 2012 and 02 Aug 2012). These were sent to Eton Biosciences (San Diego, CA, USA) for Sanger sequencing. Methods, samples and results of this work are reported in Appendix S1.

Sanger sequencing of ITS for selected Synechococcus *cultures*

We chose to sequence the ITS region for several *Synechococcus* strains that were isolated off the Scripps pier and have been maintained in the lab: CC9311 (clade I), CC9902 (clade IV), CC9605 (clade II) and CC9701 (clade II). Cultures were grown at 20°C in F/4 media (Guillard and Ryther, 1962), kept in continuous light and transferred to flasks of 50 ml fresh media approximately twice a month. DNA was extracted by boiling pellets from centrifuged cultures at 100°C for 10 min and adding 1 μl of the boiled preparation to a 25 μl polymerase chain reaction (PCR) containing 13.5 μl GoTaq HotStart MasterMix (Promega Corporation, Madison, WI), 9.5 μl sterile water, 1 μl of each primer. We amplified the ITS amplicon using the same primers (F: GGATCACC TCCTAACAGGGAG, R: AGGTTAGGAGACTCGAACTC) and PCR program described in Choi *et al*. (2014): denaturation for 5 min at 94°C, 35 cycles of 94°C for 45 s, 50°C for 45 s and 72°C for 90 s and a final 10 min elongation at 72°C. PCR product was sent to Eton Biosciences for Sanger sequencing and resulting reads were assembled using CLC Workbench v 8 (https://www.qiagenbioinformatics.com/).

Mock communities and nutrient limitation of Synechococcus *cultures*

In order to evaluate whether sequence results using these primers might accurately reflect the proportions of *Synechococcus* cells of various strains present in the sample, we created several mock communities from lab strains which were then processed and sequenced like the environmental pier samples. To create the mock communities, we used a haemocytometer to enumerate *Synechococcus* cells from log-phase cultures of CC9311 (clade I), CC9902 (clade IV), CC9605 (clade II) and CC9701 (clade II). We combined these in different sets of proportions (reported in Appendix S1) and filtered the composites on to 0.2 μm Supor filters.

In addition, we grew cultures of CC9902 in F/4 media and low-N F/4 media (1/10th the normal nitrate concentration) until they were in log-phase or N-limited respectively (as determined by fluorescence measurements, further described in Appendix S1). At that point 200 μl of each culture was filtered onto 0.2 μm Supor filters.

For both of the above cultures (or culture composites), filters were stored at −80°C and DNA was extracted and sent to RTL genomics and sequenced exactly as described earlier for the pier samples.

Data analysis

ITS sequences were processed using a combination of open-source software programs. Forward and reverse reads were assembled using Mothur (Schloss *et al*., 2009) and ASVs were assigned using the deblur workflow (Amir *et al*., 2017) with a trim size of 350 bp. We used Mothur's classify.seqs command for classifying ASVs against a database of cyanobacterial 16S–23S ITS sequences compiled by Choi *et al*. (2014). Some sequence classifications were later adjusted as necessary based on the phylogenetic tree we generated (see below).

To examine the relationship of our sequences to one another and to those of known *Synechococcus* strains, we added ITS sequences for known marine *Synechococcus* strains from the Joint Genome Institute Integrated Microbial Genomes database (IMG-JGI) (Chen *et al*., 2019, https://img.jgi.doe.gov/) as well as sequences from our lab strains to the amplicon sequences. This compiled fasta file was aligned and trimmed using Mothur. RaxML (Stamatakis, 2014) was used to generate a maximum likelihood tree with 1000 bootstraps. This tree used a GTR + I + G model as determined using JModelTest (Guindon and Gascuel, 2003; Darriba *et al*., 2012) with sequences from *Synechococcus* subcluster 5.3 as an outgroup.

All other analyses were conducted using R or Python, including the R phyloseq package (Mcmurdie and Holmes, 2013).

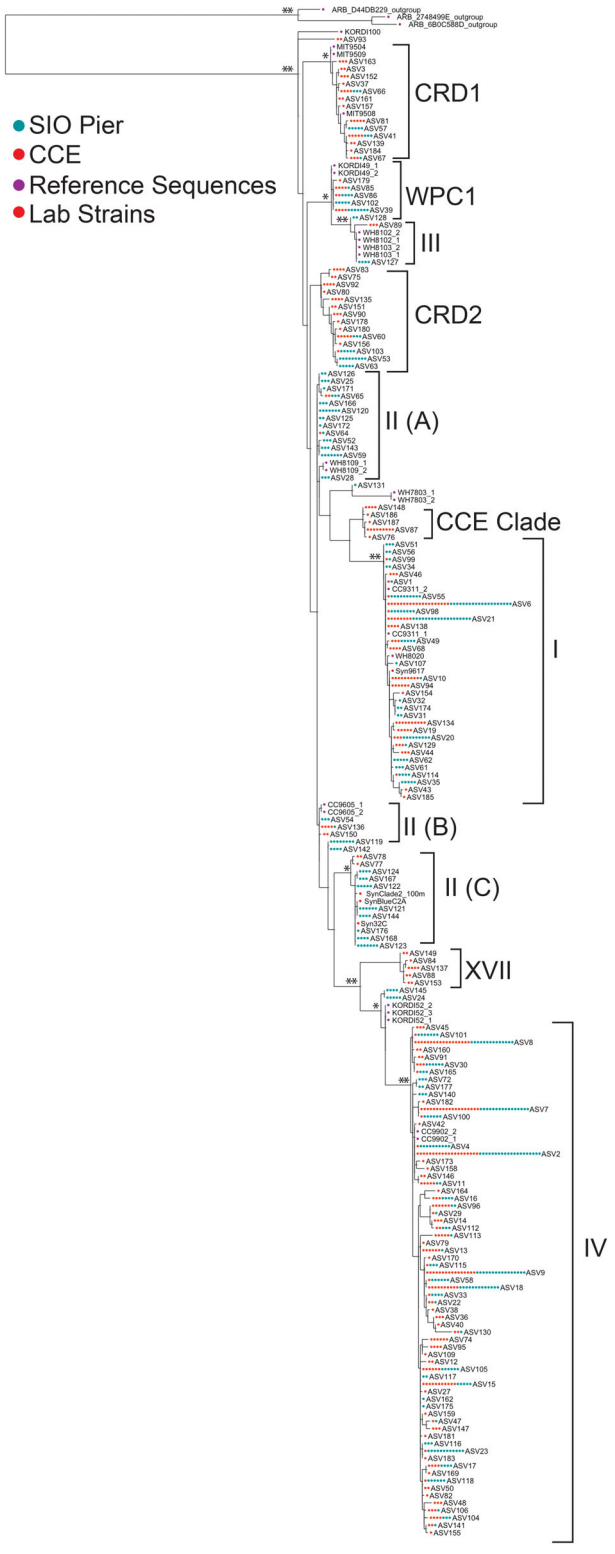

Fig 1. Legend on next column.

Results

Numerous co-existing ASVs found in the Southern California coastal ecosystem

A number of 176 ASVs were found using deblur analysis among all the sites. Of these, 35 were classified as clade I and 73 as clade IV, making these two clades both the most abundant and the most diverse. A maximum likelihood tree of all sequences, along with some lab strains and other cultured strains whose genomes have been sequenced, revealed a clear distinction of clades I, III, IV and CRD1, with strong (> 88%) bootstrap support (Fig. 1). Classification of these sequences using Mothur's classify.seqs command and the Choi *et al*. (2014) sequence database was consistent with their position on the tree. However, sequences classified as clade II and CRD2 using the database were not monophyletic on the tree and instead formed three and two clusters respectively. ASVs that were classified as CRD2 using the database formed two separate clusters but one of these had low (< 93%) identity with any CRD2 sequences within the database. We have named this cluster 'CCE Clade' as it does not appear to belong to CRD2 and forms a cluster of ASVs found exclusively at offshore sites. In the case of clade II, we believe the multiple clusters in Fig. 1 to be a consequence of our using a shorter length amplicon than that used to generate the original database and its classifications. We found that in a tree made with the longer, 750 bp region, these distinct clusters within clade II were maintained but formed a supported monophyletic clade together (Appendix S1). Because the positions of the clade II clusters in the tree made with the shorter region (Fig. 1) were poorly supported, we retained their clade identifications but referred to the clusters separately (II-A, II-B, II-C).

The maximum sequence-dissimilarity among all the ASVs was approximately 13%. Within individual clades, minimum identity was 96% (clade I), 94.9% (clade IV), 92% (clade II all), 98% (II-A), 97.1% (II-B), 93.2% (II-C), 96.9% (CRD1), 96.6% (CRD2-A), 98.9% (CCE clade).

A comparison of four different blooms at the same sampling site

From 2010 to 2016, the average *Synechococcus* abundance was 118,759 cells/ml. The *Synechococcus*

Fig 1. Maximum likelihood tree of *Synechococcus* ASVs along with selected cultured samples and reference sequences from publicly available databases. Labelled clade designations are used for ASVs hereafter. The number of circles by each OTU represents the number of samples it was found in, and circle colour represents the type of sample. Bootstrap values for selected nodes are indicated: * = 80 or greater; ** = 90 or greater. Three sequences from *Synechococcus* 5.3 were used as outgroups. [Color figure can be viewed at wileyonlinelibrary.com]

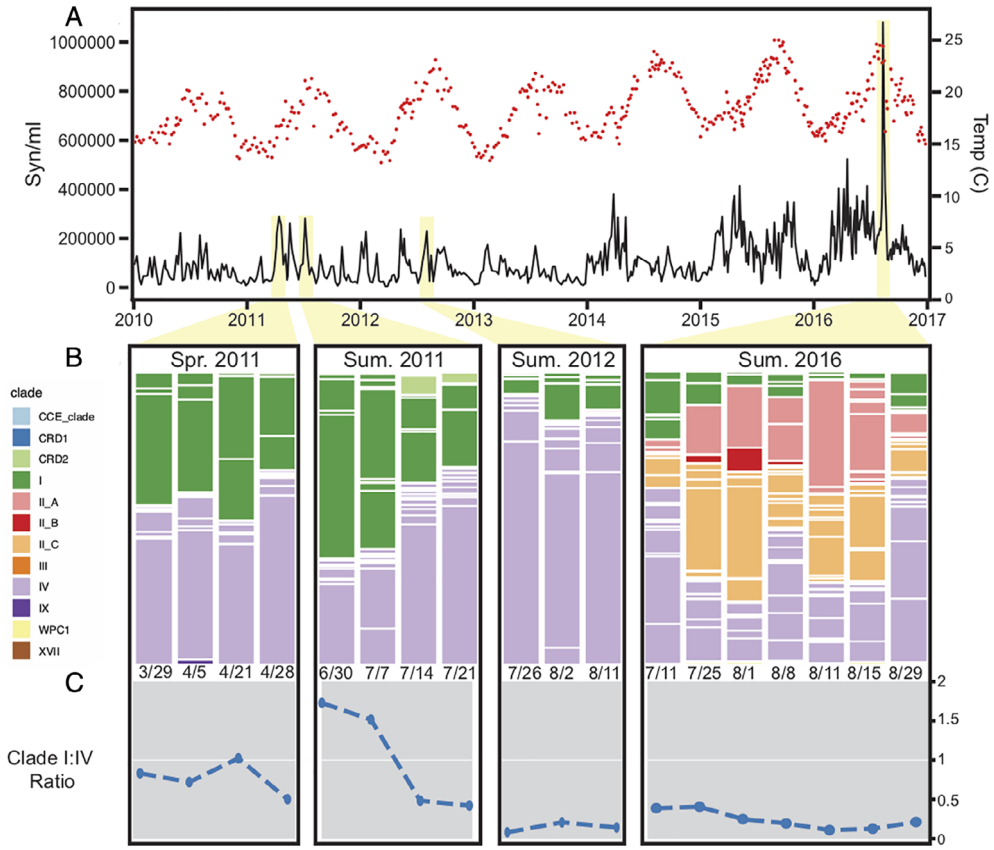

Fig 2. Comparison of four *Synechococcus* blooms at the Scripps pier.
A. *Synechococcus* density (black lines) over the entire time period between 2010–2016, measured weekly or bi-weekly, and seawater surface temperature (red dots).
B. Relative abundances of different *Synechococcus* clades during selected time points over the course of *Synechococcus* blooms. White dividing lines separate ASVs falling within the coloured clades. (C)
C. Clade I:IV ratio at different time points during the blooms. [Color figure can be viewed at wileyonlinelibrary.com]

population at the Scripps pier typically remains below a density of 200,000 cells/ml, with several blooms a year rising above this concentration, but with the timing and size of blooms varying greatly year to year (Fig. 2A). The maximum bloom densities also remained below 300,000 cells/ml from 2010 through 2013 and exceeded this concentration once in 2014 and twice in 2015. Most notably, there was a large bloom in summer 2016 where the density rose above 1,000,000 cells/ml (Fig. 2A) – a peak density greater than any detected since 2005 and up through at least June 2019 (Tai and Palenik, 2009, Palenik Lab, unpubl. data). In 2011, there were three blooms where the concentration increased from below 100,000 cells/ml to above 200,000 cells/ml and back down to the original concentrations within several weeks. In 2012, this only occurred once, though the population rose above 200,000 cells/ml two other times with a slower rate of increase. *Synechococcus* abundance had a loose but significant correlation with surface temperature (Pearson's $r = 0.31$, $p = 1e\text{-}10$) but no correlation with chlorophyll. Without the warmer and larger 2016 bloom, this correlation decreases.

Synechococcus diversity at the Scripps pier was consistently dominated by clades I and IV, with the exception of the summer of 2016 during which an abnormally large bloom coincided with an atypical abundance and variety of clade II sequences (Fig. 2A,B). Members of other clades were occasionally observed, including a small number of sequences from CRD1 and CRD2 in summer 2011 and 2012 and a small number of sequences from clades III, XVII and WPC1 in 2016 (Fig. 2B).

The four blooms selected for comparison (spring 2011, summer 2011, summer 2012, summer 2016) differed from one another in both their initial *Synechococcus* community composition as well as in the compositional shifts that occurred over the course of the blooms (Fig. 2B and C). For example, at the start of the summer 2011 bloom, the ratio of clade I to clade IV was higher than at any other time point sampled and decreased over the course of the 2011 bloom. In contrast, there was no

clear directional change in the clade I:IV ratio in spring 2011 or summer 2012, but the ratio itself differed between the blooms, with clade IV much more abundant relative to clade I in summer 2012 (Fig. 2C). We highlight this clade I to clade IV ratio because these are typically the most dominant clades at our site and this ratio has been tracked in other studies (Tai and Palenik, 2009). The phenomenon of clade I decreasing relative to clade IV in summer 2011 was also observed in sequences from clone libraries of the *rpoC1* gene (Appendix S1). The *rpoC1* libraries from summer 2012 also revealed no shift in the clade I:IV ratio before and after the bloom. However, the summer 2012 *rpoC1* libraries contained approximately the same number of clones from clade I and clade IV whereas the ITS amplicon sequence data contained far more clade IV reads relative to clade I. Given the small number of clones sequenced, it is likely that a greater sequencing depth for *rpoC1* would more accurately reflect environmental abundances. Overall, our limited use of *rpoC1* generally confirmed our ITS results.

Temperatures during the 2016 bloom were the highest of the four blooms studied, with an average surface temperature of 22.4°C over the course of the dates sampled. Clade IV remained more abundant than clade I but over the course of the large *Synechococcus* bloom, clade II became the most abundant clade overall. On four of the 2016 sampling dates clade II comprised over 50% of all reads. Its abundance had declined by 2 weeks after the peak of the bloom, though it was still present and more abundant than clade I.

We 'quantified' counts of each clade by converting the proportions of sequence reads to proportions of total *Synechococcus* cell abundance and found that even though it was not the dominant clade during the 2016 bloom, the abundance of clade IV during peak of the 2016 bloom was greater than its abundance during any other bloom (Fig. 3). The great abundance of clade II is in *addition* to what might have otherwise been considered a 'typical' bloom. Clade II sequences were absent entirely from all other samples except 26 July and 11 August 2012, where they were < 1.5% of the total sequences. This is congruent with previous studies at the Scripps pier where clade II was either not detected (Tai *et al*., 2011) or detected at very low abundances (Tai and Palenik, 2009).

In order to determine whether we could use sequence data as a general proxy for relative abundances of different *Synechococcus* clades to one another, we examined the clade composition of sequence data obtained from five mock communities (Appendix S1). These communities consisted of different combinations of cultured *Synechococcus* strains from clades I, IV, II (two strains) and CRD1 (respectively, CC9311, CC9902, CC9701, CC9605 and CC9305). In every case, CC9311 was slightly over-represented in the sequence data relative to its known proportion of the cells added to the community (Appendix S1). While the current set of mock communities does not allow us to make any quantitative conversion factor, we can conclude that the greater proportion of clade IV sequences in our data does likely correspond with a greater abundance of clade IV cells. Importantly, the mock community containing clade II representatives had clade II sequences in approximately the same proportions as the cell composition, confirming that clade II was highly abundant during the 2016 bloom.

Hidden switching of within-clade variants over the course of blooms

While the *Synechococcus* community at the Scripps pier consisted mainly of clade I and clade IV members (with some members of CRD2 detected in 2011 and the high abundance of clade II in 2016), there were multiple

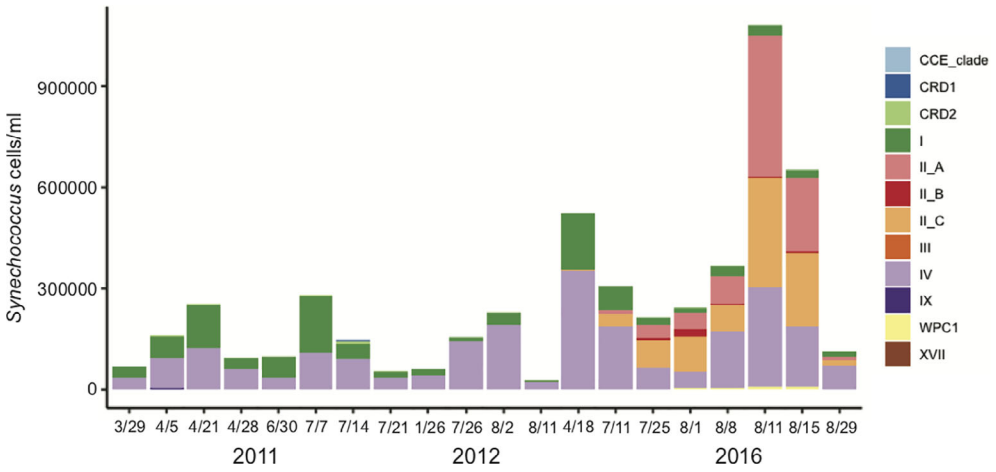

Fig 3. *Synechococcus* community composition normalized to Synechococcus cells/ml. [Color figure can be viewed at wileyonlinelibrary.com]

variants within each clade present at any given time (Fig. 4). For both clades I and IV, a single ASV was typically the most abundant representative of its clade (ASV 6 for clade I, ASV 2 for clade IV), but several other ASVs were consistently present at lower abundances.

There were observable shifts in the ratios of ASVs within a clade over the course of each bloom. For example, over the course of the spring 2011 bloom, clade IV was consistently dominated by ASV 2, but there was a shift in the dominant clade I ASV during the decline of the bloom, from ASV 6 to ASV 21. This was the only case where ASV 21 was the most abundant clade I ASV.

In the summer blooms of 2011 and 2012, the peak bloom dates (07 July 2011 and 02 Aug 2012) there was a shift from clade I ASV 6 to ASV 55 and from clade IV ASV 2 to ASV 4 just prior to the decline of the bloom. The typically dominant clade I and IV ASVs (respectively, ASVs 6 and 2) returned to this role for the decline of both

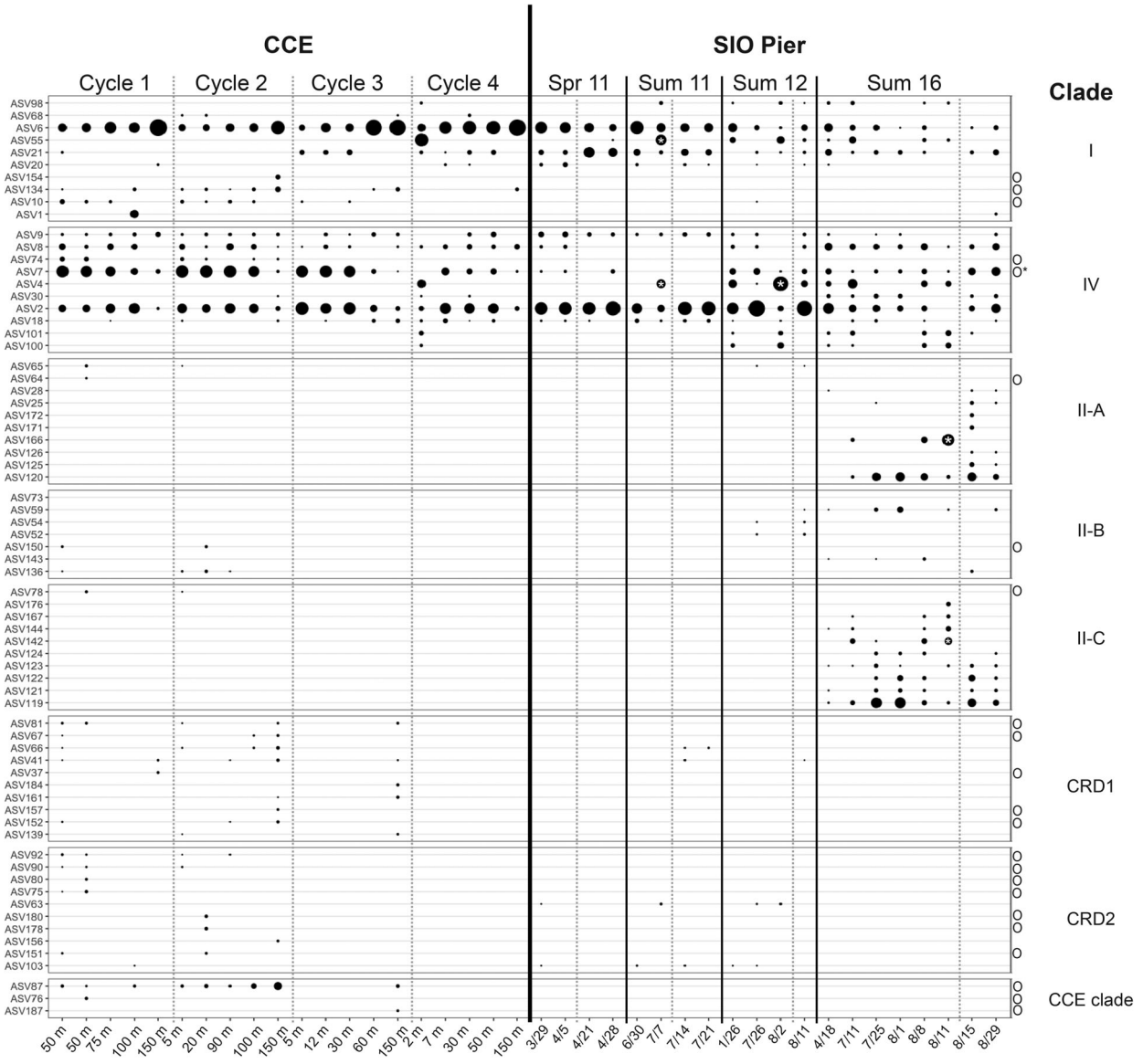

Fig 4. Relative abundance of *Synechococcus* ASVs among pier and CCE samples. Circle size of an ASV indicates the proportion of the total sample its sequences comprise. Dark vertical lines separate the different cycles or the different blooms that were evaluated. Grey dotted lines delineate the termination of the bloom (in each case, time points before the dotted line represent the peak abundance measured during that bloom and time points after dotted line are lower than the peak abundance). Only the top 10 most abundant ASVs are shown for each clade (or fewer if there were not 10). ASVs with an 'O' on the right side appear to be oligotrophic ASVs, meaning they were only observed in the two furthest offshore cycles or deep samples. The exception is ASV 7 ('O*'), which was frequently the dominant clade IV variant offshore. It was present at the pier as well but never as the most abundant clade IV variant. Some of the black dots are starred, indicating the switch to the 'A' variants that occurred prior to certain blooms.

of these blooms. It is also noteworthy that, during the peak of the 2016 bloom (11 Aug 2016), the two main clade I and IV ASVs at the pier (ASV 6 and 2 respectively) were completely absent, though ASVs 55 and 4 were present. Their sequences were detected on all other dates, including the rest of the 2016 dates sampled. We also observed a similar shift within clade II members over the course of the 2016 bloom. ASVs 119 (clade II-C) and 120 (clade II-A) were dominant before and after the bloom with a shift to ASVs 142 (clade II-C) and 166 (clade II-A) on the peak bloom date.

The rapid temporary switching between the more ubiquitous ASV and the secondary ASV that became abundant during the peak of the bloom was consistently between pairs of ASVs that differed in the 22nd position of the amplicon sequence. In every case, the more ubiquitous variant (ASV 6 for clade I, ASV 2 for clade IV, ASV 120 for clade II-A and ASV 119 for clade II-B) had a 'G' in the 22nd position while the more transient variant had an 'A' (ASV 55 for clade I, ASV 4 for clade IV, ASV 166 for clade II-A and ASV 142 for clade II-C). This was the only base change between these pairs of ASVs (Fig. 5).

Because of the unusually consistent nature of this ASV switching (with the same type of switch occurring in multiple clades), it was unclear whether this actually represented 'microdiversity' in the sense of a separate population becoming temporarily more abundant. To investigate whether this phenomenon could be replicated in a lab environment, we grew cultures of *Synechococcus* CC9902 in either regular or low-nitrate F/4 media. We then sequenced the ITS region as done with our environmental samples to see whether there would be multiple ASVs within a single sample and whether both samples would have the same ASV composition. We found that the sequences from the control culture contained both G- and A-variants (i.e. sequences corresponding with both ASV 2 and ASV 4 respectively). Approximately, 18% of sequences had an A in the 22nd position and 82% had a G. In the low-N culture, 100% of the sequences belonged to ASV 2, which had a G in the 22nd position (Fig. 5), indicating that this particular form of detected sequence diversity can be altered by physiological conditions.

Spatial variation in Synechococcus *community composition*

We observed variation in the *Synechococcus* clades present at different cycles (sites) in the CCE (Fig. 6), with greater ASV richness further offshore. Clades I and IV were always dominant, but clades II, CRD2 and WPC1 were present at the oligotrophic, offshore cycles (1 and 2), and clades CRD1, and XVII were present at cycles 1, 2 and 3. Clades I and IV were dominant at the nearshore cycle 4 with the exception of three CRD1 reads.

Within individual clades, some ASVs were more abundant offshore (ASVs 134 and ASV 10 in clade I, ASVs 7 and 74 in clade IV, ASV 87 in CRD1) while others were observed only at nearshore sites (ASV 21 in clade I and ASV 4 in clade IV) (Fig. 4).

Synechococcus depth profiles differed at each sampling site, with cell numbers peaking near the deep chlorophyll maxima in each case (Appendix S1). At the two sites furthest offshore (cycle 1, the offshore stratified region, and cycle 2, the core of the California Current proper), clade IV was more abundant relative to clade I at all depths except 150 m (Fig. 6). In cycle 3, the wind stress curl upwelling domain, clade IV was more abundant than clade I at 5, 12 and 30 m, but clade I became more abundant than clade IV at 60 and 150 m. In the coastal boundary upwelling region (cycle 4), clade I was more abundant relative to clade IV at most depths. The deepest samples (150 m) always comprised mostly of clade I, though *Synechococcus* density was much lower at this depth (Appendix S1). Thus there is now substantial evidence that clade I becomes dominant relative to

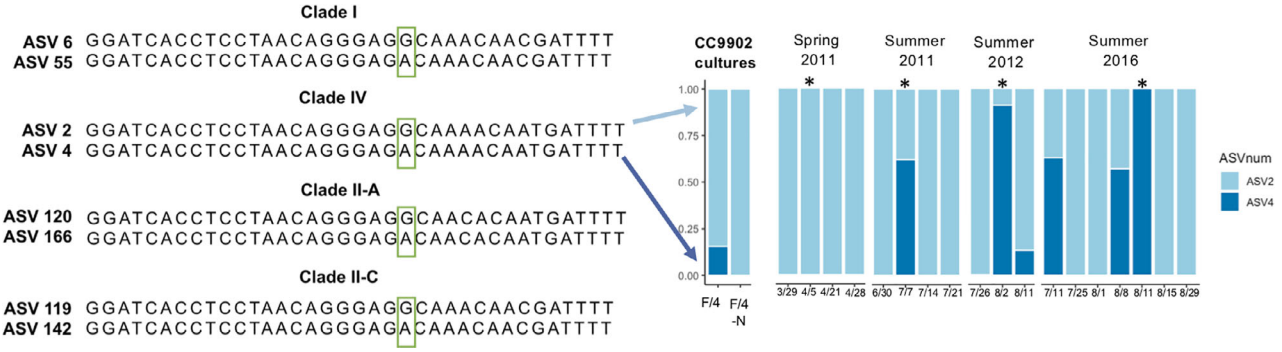

Fig 5. Left: comparative alignments of the beginning stretch of the ITS amplicon between pairs of temporally switching ASVs within four clades. All four pairs differ in the same position. In each case, the rest of the sequence shown is identical between each pair. Right: relative abundances of just the G-variant (ASV 2) and A-variant (ASV 4) in *Synechococcus* CC9902 cultures grown in regular F/4 media or nitrogen-poor F/4 media and in the pier samples. Stars denote the peaks of each bloom. [Color figure can be viewed at wileyonlinelibrary.com]

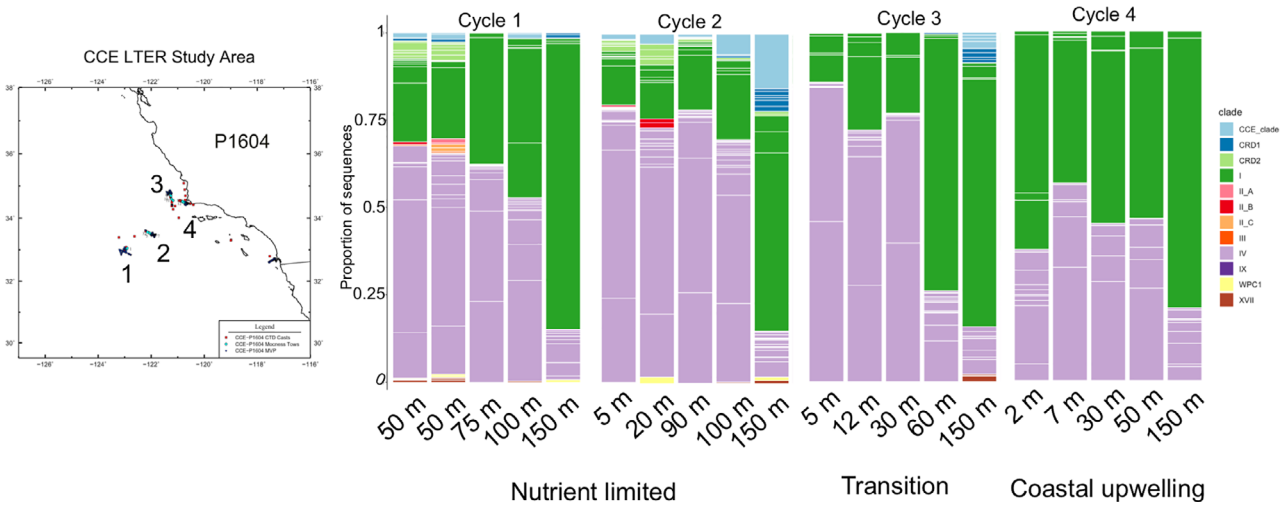

Fig 6. Map of sampling sites (cycles) from CCE P1604 cruise and *Synechococcus* community composition at multiple depths within each cycle. [Color figure can be viewed at wileyonlinelibrary.com]

clade IV with increasing depth (and decreasing temperature).

There were 52 ASVs shared between the pier and CCE sampling sites; the rest were geographically distinct. The most abundant clade I and IV ASVs from the pier (ASVs 6 and 2 respectively) were also present and abundant at all CCE sampling sites above 90 m; furthermore, ASV 6 was the most abundant ASV in every single deep sample (150 m). ASV 7, a presumably oligotrophic specialist member of clade IV, was detected at the pier but always at low levels, but most ASVs belonging to 'offshore' clades as identified above were not found at the pier. While members of clade II were found at the pier and at the two furthest offshore cycles, most (28 of 30) clade II ASVs were specific either to the pier or to the CCE sites, despite the fact that most of the clade II sequences at the pier were found in 2016, just months after the CCE sampling occurred. Overall, the pier and CCE *Synechococcus* communities were fairly distinct, and the pier communities during 2011 and 2012 particularly clustered separately from that in 2016 (Appendix S1).

Discussion

It is well known that *Synechococcus* is a diverse genus with numerous evolutionary adaptations allowing it to be widespread among the world's oceans. However, as this sequencing effort demonstrates, its true diversity remains under-characterized. Lee *et al*. (2019) found that the majority of environmental sequences are not identical to reference genomes. Indeed in the case of clade I the most abundant variant in our data did not share the same ITS sequence as its closest cultured genome representative, CC9311 (although in the case of clade IV, CC9902 corresponded with ASV2 which was generally the most abundant clade IV ASV). Additionally, our results add another layer to the question of why different clades of *Synechococcus* co-occur in the same environment; we found that even closely related variants within the same *Synechococcus* clade can co-occur at a single site with detectable fluctuations in their abundance over fairly short time scales.

Understanding the drivers behind the initiation of phytoplankton blooms is of great interest for numerous reasons, including the evaluation of ecological hypotheses and potential for prediction of future bloom events. At the Scripps pier site we know from over a decade of monitoring that blooms of *Synechococcus* typically occur a few times a year and last on the order of days to weeks. Because we do not observe an explicit seasonal abundance pattern, our site is well suited for studying the drivers of both bloom onset and decline. While *Synechococcus* abundance does have a loose correlation with water temperature, no strong and consistent patterns with other biotic or abiotic factors have been confirmed at our site to date. We have previously used sequencing methods to identify putative *Synechococcus* grazers based on their increased abundance during or immediately following *Synechococcus* blooms (Nagarkar *et al*., 2018), but particular biological interactions have not yet been corroborated as consistent driving forces of initiation or termination of these blooms. Two prominent hypotheses for biotic factors contributing to bloom decline are grazing and viral lysis (Landry and Hassett, 1982; Proctor and Fuhrman, 1990), while changes in environmental conditions could also affect the success or mortality of *Synechococcus*. All of these factors have the potential to act with specificity on the sub-clade scale; indeed, Ahlgren *et al*. (2019) found that while abiotic

factors influenced *Synechococcus* communities on a broader taxonomic level (i.e. patterns in clade abundance), patterns in abundances of certain phages were likely strong drivers of ASV-level dynamics.

While *Synechococcus* blooms can readily be detected and tracked using microscopy or flow cytometry, these methods cannot distinguish between different clades of *Synechococcus*. However, studies at our and other sites have found that diverse *Synechococcus* variants, at the clade and sub-clade level, are often present at a single location, and furthermore that the within-*Synechococcus* diversity can shift drastically over small spatial or temporal scales (Paerl *et al*., 2011; Tai *et al*., 2011; Gutierrez-Rodriguez *et al*., 2014; Mackey *et al*., 2017). The patterns, or lack thereof, in this microdiversity can indicate whether different variants represnet neutral diversity or do occupy distinct niches (Ahlgren *et al*., 2019). In this study we found both cases where the *Synechococcus* community composition at the clade level changes considerably over the course of the bloom (summer 2011, summer 2016) and other cases where it remains fairly consistent (spring 2011, summer 2012). Only summer 2011 exhibited the pattern reported previously at the pier, with clade I becoming abundant relative to clade IV prior to the bloom and then decreasing over the course of the bloom. Some common *Synechococcus* grazers, including *Paraphysmonas* sp. and *Pteridomonas* sp. have been demonstrated to have high prey selectivity, with differential growth success on *Synechococcus* strains even within the same clade (Zwirglmaier *et al*., 2009). A shift in clade composition seen over the course of a bloom as in summer 2011 could be indicative of grazers targeting one clade over the other.

The 2016 bloom was highly atypical both in its high abundances of *Synechococcus* and in the appearance of clade II sequences. Among all *Synechococcus* counts measured between 2010 and 2016, this represented the largest bloom both in terms of *Synechococcus* abundance at its peak as well as the fastest increase in *Synechococcus* density between one sampling point and the next. The prominent presence of clade II was also unique; clade II is typically found in the open ocean, particularly in warmer and more oligotrophic (especially low-phosphate) environments, and has been found to co-occur with clades III and X (Sohm *et al*., 2016). We did detect a small number of clade III reads at the pier site during the same time as the 2016 bloom, though no sequences were classified as clade X. The most abundant clade II ASVs detected during the 2016 bloom were not found at the pier in other years or at any of our offshore sites, even though the offshore samples were collected only a few months prior to the bloom (April 2016). Conversely, only two clade II ASVs found in the CCE samples were detected at the pier, at very low abundances and only in 2012. Two atypical warming events preceded the 2016 bloom: a series of warm anomalies in the North Pacific along with an El Niño that persisted through spring 2016 (Di Lorenzo and Mantua, 2016). This time period was marked by features including reduced frequency of oceanic fronts (Kahru *et al*., 2018). It is unclear whether this anomalous bloom represents the arrival of a distinct oligotrophic water parcel to the coast or is a result of conditions shifting to be favourable for members of clade II. It is possible that this may be a more common occurrence in the future with rising ocean temperatures.

Geographical niche partitioning has been extensively documented in *Synechococcus* and is hypothesized to indicate adaptations of ecotypes to different temperature, light and nutrient regimes, as well as environments where these resources are consistently available compared with highly variable resources as might be found on the coast (Zwirglmaier *et al*., 2008). From prior studies it is well established that clades I and IV are found to co-occur in coastal environments, and members of these clades have consistently been dominant at the Scripps Pier (Tai and Palenik, 2009). However, they were also the most abundant members of all offshore samples including the more oligotrophic sites. The dominance of clade I at depth compared with clade IV seen here potentially supports the hypothesis that the observed seasonal pattern in the clade I:IV ratio seen in surface waters at the SIO pier (Tai *et al*., 2011) is due to seasonal stratification which would favour clade IV at the surface and clade I at depth and thus seasonally decrease the I:IV ratio.

On the ASV level, the main pier clade I ASV (ASV 6) was surprisingly also the most abundant member of clade I at most offshore sites, and the most abundant ASV overall in each deep sample (150 m). This finding is consistent with trends seen in 2011, where clade I was found at the pier and increased with depth in offshore mesotrophic waters (Tai *et al*., 2011). It appears that this particular clade I ASV is more of a generalist, capable of adapting to different light, nutrient and temperature regimes. There is a small amount of evidence that in contrast to ASV 6, Clade I ASVs 1, 10 and 134 are oligotrophic variants. At the pier Clade IV ASV 2 was by far the most abundant and ASV 7 was not even detected in some years. However, offshore there was co-dominance of ASVs 2 and 7 . ASV 7 is likely a preferentially oligotrophic clade IV variant. Thus we have found further evidence of microdiversity in *Synechococcus* having a biogeographical pattern. We also acknowledge that ITS as a marker may not be sufficient to reveal all the 'true' microdiversity within a sample as strains may vary in other ways such as differential acquisition of genes by horizontal gene transfer.

An important consideration for studies that utilize environmental sequencing is whether the sequence-based

characterization of the community accurately reflects relative abundances of its members. In order to examine this assumption we sequenced several mock communities created from combinations of lab cultures. While the sequence results from five cultured strains cannot be generalized to the entire set of ASVs found in our environmental samples, they offer a preliminary confirmation that sequencing of this amplicon can adequately represent environmental abundances. A similar result was found by Ahlgren *et al*. (2019) when assessing the extent to which mock communities combining DNA (as opposed to cells) represented proportions of *Synechococcus* ecotypes based sequencing of the ITS region. In our environmental samples, we found clade IV sequences to be more abundant than clade I sequences at the majority of our time points and sites and believe the sequence data reflects the relative abundance of those clades because of results from sequencing mock communities. If anything, the mock community results, in which clade I sequences are consistently slightly overrepresented, suggest that clade IV might be even more abundant environmentally than the sequence data indicate. This skew in the relative proportions of resulting sequence data is likely due to differing ease of DNA extraction or PCR primer bias between the different strains, as the cultures used are known to have the same copy number for the ITS region sequenced. Importantly, some of the mock communities contained two different clade I or IV ASVs, a result further discussed below.

Our sequence data indicate that *Synechococcus* blooms consist of several sequence variants with individual temporal patterns. We were able to detect large changes in the sub-clade-level community composition, with both the dominant clades (I and IV) experiencing a spike in a less abundant variant during the peak of three separate blooms (the exception was spring 2011). A similar phenomenon was also found with clade II ASVs during the 2016 bloom. These patterns are remarkable in their consistency over multiple blooms and between different clades. In each case, the 'switching' ASVs show the same G to A shift in the 22nd position of the amplicon sequenced, and in some of the blooms this switch occurs between multiple pairs of ASVs from different clades. To our knowledge, no shift in sequence has been described at such fine resolution, both in terms of being a single nucleotide switch and in being on the scale of days. We also were unable to find *Synechococcus* sequences with a 'G' in that position in several databases, including the Choi *et al*. (2014) database we used for classification, and the sequences deposited in several other *Synechococcus* sequencing efforts (Mazard *et al*., 2011; Lee *et al*., 2019). However, amplicon sequences deposited in the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) by Choi *et al*. (2014) did include variants with this genotype.

Interestingly, the sequences of both CC9311 and CC9902 from our mock communities were dominated by their respective G-variants. In the case of CC9902, we found that cultures of *Synechococcus* CC9902 grown in standard F/4 media contained both sequence variants, with the A-variant accounting for approximately 18% of sequence reads, but when nitrogen limited, only the G-variant appeared. This indicates that environmental factors play a role in the sequence variants that are most abundant, but whether this represents a change of microdiversity in the phylogenetic sense is unclear. It is possible that the variant switching we observe over the course a physiological response, particularly because the position of the switched nucleotide is within a regulatory region that (Iteman *et al*., 2000). One possibility is that *Synechococcus* undergoes DNA modifications such methylation or oxidation in response to its environment and these affect correct base pair calling during amplification and sequencing. The cyanobacterium *Synechocystis* has been found to respond to nitrogen starvation with changes in methylation (Hu *et al*., 2018). While methylation is not known to affect Illumina readouts, other types of DNA modification have been reported to induce a sequencing artefact. For example, Costello *et al*. (2013) found that an acoustic DNA shearing protocol induced guanine oxidation which affected a portion of sequence reads. The use of other sequencing methods may be useful in understanding this change, to test whether it represents a modification that impacts the Illumina sequencing mechanism specifically. Unfortunately, nutrient data were not available on all bloom dates sampled and thus we cannot describe changes in nutrient conditions over the course of each bloom. Future ecological and lab-based studies may be able to find associations between other abiotic conditions and the sequence variants at this particular position, but this pilot experiment demonstrates that such a change can likely be induced by nutrient limitation.

Recent characterizations of *Synechococcus* using sequencing have indicated that a genus already known to be diverse consists of far more species and subspecies-level variants than previously realized. The fine-scale spatial and temporal specialization of *Synechococcus* subtypes demonstrates that combinations of environmental factors can create small and transient niches that individual populations readily take advantage of. Here we have demonstrated that this microdiversity has distinct patterns with geography and depth and can shift drastically on the scale of days. Furthermore, we report a previously undescribed phenomenon – a rapid switching between nearly identical sequence variants over the course of blooms, found in multiple *Synechococcus* clades.

Acknowledgements

Funding for this project was provided through a National Science Foundation grant (DEB-1233085) to Brian Palenik and National Science Foundation grant OCE-1614359 to the CCE-LTER site. Analyses of molecular samples was partly funded by the Graduate Student Excellence Research Award of the Scripps Institution of Oceanography to B. Valencia. B. Valencia was also supported by a scholarship (529-2011) from the Colombian Administrative Department of Science, Technology and Innovation (COLCIENCIAS). We thank Mike Landry his valuable input and for providing FCM data from the P1604 CCE cruise and Traci Yuen for her contributions in making some of the preliminary clone libraries. Scripps Pier ancillary measurements were collected by the Birch Aquarium at Scripps staff and volunteers. Data were provided by the Shore Stations Program sponsored at Scripps Institution of Oceanography by California State Parks, Division of Boating and Waterways. Contact: shorestation@ucsd.edu. We thank Melissa Carter, Kristi Seech and James Fumo for their consistent help and patience with discussing and coordinating pier sampling.

References

Acinas, S.G., Klepac-Ceraj, V., Hunt, D.E., Pharino, C. et al. (2004). Fine-scale phylogenetic architecture of a complex bacterial community. *Nature*, **430**: 551–554. http://dx.doi.org/10.1038/nature02649.

Agawin, N.S.R., Duarte, C.M., and Agustí, S. (1998) Growth and abundance of *Synechococcus* sp. in a Mediterranean Bay: seasonality and relationship with temperature. *Mar Ecol Prog Ser* **170**: 45–53. https://doi.org/10.3354/meps170045.

Agawin, N.S.R., Duarte, C.M., and Agustí, S. (2000) Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production. *Limnol Oceanogr* **45**: 591–600. https://doi.org/10.4319/lo.2000.45.3.0591.

Ahlgren, N.A., Perelman, J.N., Yeh, Y.C., and Fuhrman, J.A. (2019) Multi-year dynamics of fine-scale marine cyanobacterial populations are more strongly explained by phage interactions than abiotic, bottom-up factors. *Environ Microbiol* **21**: 2948–2963. https://doi.org/10.1111/1462-2920.14687.

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**: 1–24. https://doi.org/10.3389/fmicb.2012.00213.

Amir, A., McDonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J.T., Zech Xu, Z., *et al*. (2017) Deblur rapidly resolves single-nucleotide community sequence patterns. *MSystems* **2**: 1–7.

Chen, I.A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., *et al*. (2019) IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res* **47**: 666–677. https://doi.org/10.1093/nar/gky901.

Choi, D.H., Hoon Noh, J., and Lee, J.-H. (2014) Application of pyrosequencing method for investigating the diversity of *Synechococcus* subcluster 5.1 in open ocean. *Microbes Environ* **29**: 17–22. https://doi.org/10.1264/jsme2.ME13063.

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. https://doi.org/10.1111/j.1574-6941.2009.00729.x.

Collado-Fabbri, S., Vaulot, D., and Ulloa, O. (2011) Structure and seasonal dynamics of the eukaryotic picophytoplankton community in a wind-driven coastal upwelling ecosystem. *Limnol Oceanogr* **56**: 2334–2346. https://doi.org/10.4319/lo.2011.56.6.2334.

Collier, J.L., Palenik, B., (2003) Phycoerythrin-containing picoplankton in the Southern California Bight. *Deep-Sea Res Part II Topic Studies Oceanogr* **50**: 2405–2422. https://doi.org/10.1016/S0967-0645(03)00127-9.

Costello, M., Pugh, T.J., Fennell, T.J., Stewart, C., Lichtenstein, L., Meldrim, J.C., *et al*. (2013) Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res* **41**: 1–12 https://doi.org/10.1093/nar/gks1443.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**: 772. https://doi.org/10.1038/nmeth.2109.

Di Lorenzo, E., and Mantua, N. (2016) Multi-year persistence of the 2014/15 North Pacific marine heatwave. *Nat Climate Change* **6**: 1042–1047. https://doi.org/10.1038/nclimate3082.

Farrant, G.K., Doré, H., Cornejo-castillo, F.M., Partensky, F., Ratin, M., and Garczarek, L. (2016) Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *PNAS* **113**: E3365–E3374. https://doi.org/10.1073/pnas.1524865113.

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

Glover, H.E., Prezelin, B., Campbell, L., Wyman, M., and Garside, C. (1988) A nitrate-dependent *Synechococcus* bloom in surface Sargasso Sea water. *Nature* **331**: 161–163. https://doi.org/10.1038/331161a0.

Guillard, R.R.L., and Ryther, J.H. (1962) Studies of marine planktonic diatoms: I. *Cyclotella nana* Husttedt, and *Detonula confervacea* (Cleve) Gran. *Can J Microbiol*, **8**: 229–239.

Guindon, S., and Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology*, **52**: 696–704. http://dx.doi.org/10.1080/10635150390235520.

Gutierrez-Rodriguez A., Slack, G., Daniels, E.F., Selph, K. E., Palenik, B., and Landry, M.R. (2014). Fine spatial structure of genetically distinct picocyanobacterial populations across environmental gradients in the Costa Rica Dome. *Limnology and Oceanography*, **59**: 705–723. http://dx.doi.org/10.4319/lo.2014.59.3.0705.

Hu, L., Xiao, P., Jiang, Y., Dong, M., Chen, Z., Li, H., *et al*. (2018) Transgenerational epigenetic inheritance under environmental stress by genome-wide DNA methylation profiling in cyanobacterium. *Front Microbiol* **9**: 1–11. https://doi.org/10.3389/fmicb.2018.01479.

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. https://doi.org/10.1038/ismej.2011.106.

Hunter-Cevera, K.R., Post, A.F., Peacock, E.E., and Sosik, H.M. (2016) Diversity of *Synechococcus* at the Martha's vineyard coastal observatory: insights from culture isolations, clone libraries, and flow cytometry. *Microb Ecol* **71**: 276–289. https://doi.org/10.1007/s00248-015-0644-1.

Iteman, I., Rippka, R., de Marsac, N.T., and Herdman, M. (2000) Comparison of conserved structural and regulatory domains within divergent 16S rRNA–23S rRNA spacer sequences of cyanobacteria. *Microbiology* **146**: 1275–1286.

Kahru, M., Jacox, M.G., and Ohman, M.D. (2018) CCE1: decrease in the frequency of oceanic fronts and surface chlorophyll concentration in the California Current System during the 2014–2016 Northeast Pacific warm anomalies. *Deep Sea Res Part 1 Oceanogr Res Papers* **140**: 4–13. https://doi.org/10.1016/j.dsr.2018.04.007.

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**: 416–421.

Landry, M.R., and Hassett, R.P. (1982) Estimating the grazing impact of marine micro-zooplankton. *Marine Biology* **67**: 283–288. https://doi.org/10.1007/BF00397668.

Larkin, A., and Martiny, A. C. (2017). Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environmental Microbiology Reports*, **9**: 55–70.

Lee, M.D., Ahlgren, N.A., Kling, J.D., Walworth, N.G., Rocap, G., Saito, M.A., Hutchins, D.A., & Webb, E.A., *et al*. (2019) Marine *Synechococcus* isolates representing globally abundant genomic lineages demonstrate a unique evolutionary path of genome reduction without a decrease in GC content. *Environ Microbiol*, **21**: 1677–1686. https://doi.org/10.1111/1462-2920.14552.

Mackey K.R.M., Hunter-Cevera, K.R., Britten, G.L., Murphy, L.G., Sogin, M.L., and Huber, J.A. (2017) Seasonal Succession and spatial patterns of Synechococcus microdiversity in a Salt Marsh Estuary revealed through 16S rRNA gene oligotyping. *Front Microbiol*, **8**: 1496. https://doi.org/10.3389/fmicb.2017.0149

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.

Mcmurdie, P.J., and Holmes, S. (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**: e61217. https://doi.org/10.1371/journal.pone.0061217.

Nagarkar, M., Countway, P.D., Du, Y., Daniels, E., Poulton, N.J., and Palenik, B. (2018) Temporal dynamics of eukaryotic microbial diversity at a coastal Pacific site. *ISME J* **12**: 2278–2291. https://doi.org/10.1038/s41396-018-0172-3.

Olson, R.J., Vaulot, D., and Chisholm, S.W. (1985) Marine phytoplankton distributions measured using shipboard flow cytometry. *Deep Sea Res Part A Oceanogr Res Pap* **32**: 1273–1280.

Paerl Ryan W., Johnson Kenneth S., Welsh Rory M., Worden Alexandra Z., Chavez Francisco P., Zehr Jonathan P. (2011). Differential Distributions of Synechococcus Subgroups Across the California Current System. *Frontiers in Microbiology*, **2**: 59. http://dx.doi.org/10.3389/fmicb.2011.00059.

Palenik, B. (1994) Cyanobacterial community structure as seen from RNA polymerase gene sequence analysis. *Appl Environ Microbiol* **60**: 3212–3219.

Palenik, B., and Haselkorn, R. (1992) Multiple evolutionary origins of prochlorophytes, the chlorophyll b-containing prokaryotes. *Nature* **355**: 265–267.

Li, W. (1998) Annual average abundance of heterotrophic bacteria and *Synechococcus* in surface ocean waters. *Limnol Oceanogr*.**43**: 1746–1753.

Proctor, L.M., and Fuhrman, J.A. (1990) Viral mortality of marine bacteria and cyanobacteria. *Nature* **343**: 60–62. https://doi.org/10.1038/343060a0.

Rajaneesh, K.M., and Mitbavkar, S. (2013) Factors controlling the temporal and spatial variations in *Synechococcus* abundance in a monsoonal estuary. *Mar Environ Res* **92**: 133–143. https://doi.org/10.1016/j.marenvres.2013.09.010.

Robidart, J.C., Preston, C.M., Paerl, R.W., Turk, K.A., Mosier, A.C., Francis, C.A., *et al*. (2011) Seasonal *Synechococcus* and Thaumarchaeal population dynamics examined with high resolution with remote in situ instrumentation. *ISME J* **6**: 513–523. https://doi.org/10.1038/ismej.2011.127.

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. https://doi.org/10.1128/AEM.68.3.1180-1191.2002.

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. https://doi.org/10.1128/AEM.01541-09.

Selph, K.E., Landry, M.R., Taylor, A.G., Yang, E., Measures, C.I., Yang, J., *et al*. (2011) Spatially-resolved taxon-specific phytoplankton production and grazing dynamics in relation to iron distributions in the Equatorial Pacific between 110 and 140 1W. *Deep-Sea Res Part II Topic Studies Oceanogr* **58**: 358–377. https://doi.org/10.1016/j.dsr2.2010.08.014.

Sohm J. A., Ahlgren, N. A., Thomson, Z. J., Williams, C., Moffett, J. W., Saito, M. A., Rocap, G. et al. (2016). Co-occurring Synechococcus ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *The ISME Journal*, **10**: 333–345. http://dx.doi.org/10.1038/ismej.2015.115.

Stamatakis Alexandros (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**: 1312–1313. http://dx.doi.org/10.1093/bioinformatics/btu033.

Stuart, R.K., Brahamsha, B., Busby, K., and Palenik, B. (2013) Genomic Island genes in a coastal marine *Synechococcus* strain confer enhanced tolerance to copper

and oxidative stress. *ISME J* **7**: 1139–1149. https://doi.org/10.1038/ismej.2012.175.

Tai, V., Burton, R.S., and Palenik, B. (2011) Temporal and spatial distributions of marine *Synechococcus* in the Southern California Bight assessed by hybridization to bead-arrays. *Mar Ecol Progress Ser* **426**: 133–147. https://doi.org/10.3354/meps09030.

Tai, V., and Palenik, B. (2009) Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *ISME J* **3**: 903–915. https://doi.org/10.1038/ismej.2009.35.

Toledo G., Palenik, B., and Brahamsha, B. (1999). Swimming Marine Synechococcus Strains with Widely Different Photosynthetic Pigment Ratios Form a Monophyletic Group. *Applied and Environmental Microbiology*, **65**: 5247–5251. http://dx.doi.org/10.1128/aem.65.12.5247-5251.1999.

Waterbury, J.B., Watson, S.W., Guillard, R.R.L., and Brand, L.E. (1979) Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* **277**: 293–294.

Xia X., Vidyarathna, N.K., Palenik, B., Lee, P., and Liu, H. (2015). Comparison of the Seasonal Variations of Synechococcus Assemblage Structures in Estuarine Waters and Coastal Waters of Hong Kong. *Applied and Environmental Microbiology*, **81**: 7644–7655. http://dx.doi.org/10.1128/aem.01895-15.

Zeidner G, Preston, C.M., Delong, E.F., Massana, R., Post, A.F., Scanlan, D.J., and Béjá, O. (2003). Molecular diversity among marine picophytoplankton as revealed by psbA analyses. *Environmental Microbiology*, **5**: 212–216. http://dx.doi.org/10.1046/j.1462-2920.2003.00403.x.

Zwirglmaier K., Jardillier L., Ostrowski M., Mazard S., Garczarek L., Vaulot D., Not F., Massana R., Ulloa O., and Scanlan D. J. (2007). Global phylogeography of marine Synechococcus and Prochlorococcus reveals a distinct partitioning of lineages among oceanic biomes. *Environmental Microbiology*, **10**: 147–161. http://dx.doi.org/10.1111/j.1462-2920.2007.01440.x.

Zwirglmaier K., Spence, E., Zubkov, M.V., Scanlan, D.J., and Mann, N.H. (2009). Differential grazing of two heterotrophic nanoflagellates on marineSynechococcusstrains. *Environmental Microbiology*, **11**: 1767–1776. http://dx.doi.org/10.1111/j.1462-2920.2009.01902.x.

Supporting Information

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

Appendix S1: Supplemental methods and data

Appendix S2: ITS data and alignment

Appendix S3: CCE and SIO Pier sample metadata