environmental microbiology

Environmental Microbiology (2019) **21**(8), 2948–2963 doi:10.1111/1462-2920.14687

Multi-year dynamics of fine-scale marine cyanobacterial populations are more strongly explained by phage interactions than abiotic, bottom-up factors

Nathan A. Ahlgren ,*† **Jessica N. Perelman** ,‡ **Yi-Chun Yeh** and **Jed A. Fuhrman**
Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA.

Summary

Currently defined ecotypes in marine cyanobacteria *Prochlorococcus* and *Synechococcus* likely contain subpopulations that themselves are ecologically distinct. We developed and applied high-throughput sequencing for the 16S-23S rRNA internally transcribed spacer (ITS) to examine ecotype and fine-scale genotypic community dynamics for monthly surface water samples spanning 5 years at the San Pedro Ocean Time-series site. Ecotype-level structure displayed regular seasonal patterns including succession, consistent with strong forcing by seasonally varying abiotic parameters (e.g. temperature, nutrients, light). We identified tens to thousands of amplicon sequence variants (ASVs) within ecotypes, many of which exhibited distinct patterns over time, suggesting ecologically distinct populations within ecotypes. Community structure within some ecotypes exhibited regular, seasonal patterns, but not for others, indicating other more irregular processes such as phage interactions are important. Network analysis including T4-like phage genotypic data revealed distinct viral variants correlated with different groups of cyanobacterial ASVs including time-lagged predator–prey relationships. Variation partitioning analysis indicated that phage community structure more strongly explains cyanobacterial community structure at the ASV level than the abiotic environmental factors. These results support a hierarchical model whereby abiotic environmental factors more strongly shape niche partitioning at the broader ecotype level while phage interactions are more important in shaping community structure of fine-scale variants within ecotypes.

Received 10 December, 2018; revised 29 March, 2019; accepted 16 April, 2019. *For correspondence. E-mail nahlgren@clarku.edu; Tel. (508) 793-7107; Fax (508) 793-7174. †Present address: Biology Department, Clark University, Worcester, MA, USA. ‡Present address: Department of Oceanography, University of Hawaii at Manoa, Honolulu, HI, USA.

Introduction

The availability of high-throughput sequencing methods has allowed amazing insight into the diversity of microbial communities. In particular, amplicon sequencing of the 16S rRNA locus is now a standard technique for assessing the diversity and community structure of microbes across various habitats and through time at particular sites. An emergent theme among such studies is that taxa defined as clusters of 16S rRNA or other gene sequences occupy distinct niches defined by environmental conditions or differences in resource utilization (Johnson *et al*. 2006; Martiny *et al*. 2009; Chase *et al*. 2017). Another common theme in this field is that using more fine-resolved metrics for defining taxa continues to uncover larger numbers of taxa that appear to be biologically and ecologically distinct (Eren *et al*. 2013; Eren *et al*. 2015; Tikhonov *et al*. 2015; Needham *et al*. 2017; Ward *et al*. 2017). Finer level resolution is achieved, for example, by increasing percent identity cutoffs (e.g. with 16S rDNA), use of oligotyping approaches such as minimum entropy decomposition (Eren *et al*. 2015), or multigene or whole genome approaches (Hunt *et al*. 2008; Mazard *et al*. 2012; Kashtan *et al*. 2014). This seemingly ever increasing discovery of finely resolved bacterial variants, often termed 'microdiversity' (Acinas *et al*. 2004), begs the question of at what resolution are these variants reflective of ecological differences, and at what level are they neutral? Answering these questions is critical because subsequent interpretation of diversity data, such as the ecology and biogeography of microbes, is predicated on defining biologically relevant taxonomic units.

Marine cyanobacteria *Prochlorococcus* and *Synechococcus*, in addition to being globally important primary producers (Flombaum *et al*. 2013), have served as valuable organisms to investigate these topics of how phylogenetically distinct groups correspond to ecologically relevant populations and patterns of microdiversity. Several distinct phylogenetic clusters or clades within each

of these named genera possess distinct niches and correspondingly occupy different habitats in the oceans due to differential adaptation to temperature, light and nutrient conditions (Coleman and Chisholm 2007; Biller *et al*. 2015; Sohm *et al*. 2015). These taxa, therefore, are thought to represent ecotypes, adopting a concept from the ecological literature of macroscopic organisms (Rocap *et al*. 2002). These ecotypes in fact are quite closely related according to 16S rRNA sequence differences (>98 or >99% identity), and thus more divergent loci such at the internally transcribed 16S–23S rRNA (ITS) region or coding genes have proven more effective for identifying closely related ecotypes (Rocap *et al*. 2002; Ahlgren and Rocap 2012; Mazard *et al*. 2012; Kashtan *et al*. 2014).

Upon closer examination, many of these ecotypes appear to contain distinct sequence subclusters, possibly representing biologically distinct taxa within ecotypes. Again, this highlights the broader pattern of microdiversity and that a large portion of microbial diversity is comprised by such microdiversity (Acinas *et al*. 2004). There is growing evidence that these within-ecotype subgroups or subclades exhibit distinct patterns over time (Tai *et al*. 2011) or distinct biogeographic patterns corresponding to gradients in environmentally relevant parameters (e.g. nutrient, temperature, and light) (Farrant *et al*. 2016; Larkin *et al*. 2016), supporting that these subclades likely represent cryptic, ecologically distinct subpopulations. Whole-genome analysis of single-cell isolates from a single ecotype, *Prochlorococcus* HL2, has also revealed several genomically distinct sequence clusters within this ecotype (Kashtan *et al*. 2014). A few of these clusters also exhibit different abundance patterns at the Bermuda Atlantic Time Series site, albeit for only three time points sampled. This work also importantly demonstrates that the ITS locus is divergent enough to congruently distinguish the same distinct subpopulations delineated by whole-genome divergence. Additional analysis of these genomes also suggests that this one ecotype may contain hundreds or thousands of genomically and presumably ecologically distinct populations. If we are to properly understand the biology and ecology of such microbial communities, it is imperative to identify and track distinct populations at an appropriate scale of diversity. Using marine cyanobacteria as a model, this suggests there are many as-yet unidentified and uncharacterized populations within currently defined ecotypes (or OTUs), and more in-depth analysis of microbial populations at a fine genetic scale is needed.

The apparent large number of ecologically distinct populations also begs the question of what mechanisms permit the co-existence of such high diversity. This question is precisely the one addressed by the classic 'Paradox of the Plankton' put forth by Hutchinson, long before molecular methods could envision such levels of diversity among bacterioplankton (Hutchinson 1961). This paradox surmises that for plankton competing for the same limited number of resources (e.g. nutrients and light) marine communities are expected to only support a limited amount of diversity as competitive exclusion would favour a few successful species. Mechanisms that help explain co-existence of diverse types include differential utilization of resources, fluctuating changes in resources and microscale habitat heterogeneity (Roy and Chattopadhyay 2007). Nonetheless, an order of magnitude or more increase in the recognized number of co-existing planktonic populations still requires some examination of how this is possible.

Differential adaptation to bottom-up controls such as light availability, various forms of N and P nutrients, and temperature can explain some but not all of the variance in cyanobacterial ecotypes or sub-ecotype populations (Johnson *et al*. 2006; Farrant *et al*. 2016; Larkin *et al*. 2016). Less well constrained are the contributions of top-down controls of microplankton predators (grazers) (Apple *et al*. 2011) and phage on community structure. Top-down effects are often proposed as key factors that contribute to the unexplained variance in community structure, but their contribution is rarely if at all assessed concurrently with abiotic bottom-up factors. Phage in particular have been proposed to generate and maintain community diversity at the finest scale through virus–host interactions (Thingstad and Lignell 1997; Thingstad 2000; Xue and Goldenfeld 2017). The studies of simple virus and host communities in the laboratory, including those of marine cyanobacteria, indeed demonstrate that virus–host interactions can generate and maintain diversity in this way (Mizoguchi *et al*. 2003; Paterson *et al*. 2010; Marston *et al*. 2012). However, there are limited field-based studies that specifically examine how phage shape host community structure in complex, natural populations (Rodriguez-Brito *et al*. 2010; Winter *et al*. 2010; Cram *et al*. 2016).

To address these aforementioned issues, we have developed and applied a high-throughput sequencing method for cyanobacterial ITS sequences from a set of monthly, surface water samples spanning 5 years at the San Pedro Ocean Time-series (SPOT) site located off the coast of Southern California. This method allows for quantitative assessment of community structure at the ecotype and subecotype level, and in natural populations, we find hundreds to thousands of distinct amplicon sequence variants (ASVs) within ecotypes. Several means of community structure analysis support that many of these ASVs are ecologically distinct; however, some ASVs within ecotypes appear to represent effectively neutral variants. Analysis of concurrently measured viral and cyanobacterial host populations more than 2 years suggests that phage community structure is a stronger driver of host community structure at fine scales of diversity than abiotic, bottom-up factors, in support of the virus–host interaction models described above.

Results

Evaluation of ITS high-throughput sequencing

To assess community diversity within the marine cyanobacteria *Prochlorococcus* and *Synechococcus* at high depth and fine resolution, primers that amplify cyanobacterial ITS sequences were adapted for use with the Illumina sequencing platform. These primers were applied to mock communities consisting of known mixtures of several different marine cyanobacteria ITS fragments to assess the accuracy of quantification. Mock communities consisted of an even mixture of nine different *Synechococcus* ecotypes (11% each) and an uneven mixture where the nine ecotypes ranged from 0.008% to 85% of the total community. Measured ecotype abundances of the even and uneven mock communities were within 1.8-fold and 1.7-fold of the expected abundances respectively (Supporting Information Fig. S1A and C). Both the even and uneven mock communities were also mixed with *Prochlorococcus* ecotype HL1 or HL2 ecotype sequences at several concentrations to better simulate natural communities that contain both genera (Supporting Information Fig. S1B and C). Most of the ecotypes were measured at levels within twofold of the levels measured in mixtures without *Prochlorococcus* added, showing minimal bias across differing concentrations of *Prochlorococcus* present. A few ecotypes however did quantify at greater than twofold from expected levels at higher levels of *Prochlorococcus* added, but never above a threefold difference. For the uneven *Synechococcus* samples mixed with *Prochlorococcus*, the four least abundant ecotypes were sometimes not detected, consistent with a drop-out effect in the number of total *Synechococcus* reads obtained as the fraction of *Prochlorococcus* in the mixtures increased (Supporting Information Fig. S1C).

We similarly tested a *Prochlorococcus* mock community. A single PCR reaction of an uneven mock community of *Prochlorococcus* ecotypes HL1, HL2, LL1, LL2 and LL4 sequences mixed at 0.19%, 97%, 1.9%, 0.02% and 0.97%, respectively, was tested and resulted in values similar to the expected relative abundances: 0.22%, 98%, 1.2%, 0% and 0.46% respectively. No LL2 sequences were detected, probably because the sequencing depth was insufficient to detect a population at such low relative abundance – for 14 639 cyanobacterial reads in this sample, one would only expect by probability to recover three LL2 reads. Likewise *Prochlorococcus* HL2 sequences mixed in with *Synechococcus* mock communities were measured at very similar levels to expected abundances (Supporting Information Fig. S1C).

Cyanobacterial community dynamics at SPOT

Cyanobacterial community structure was characterized for a set of 50 monthly surface water samples collected for 5 years from 2009 to 2014 at the SPOT site located off the coast of Los Angeles (CA, USA). Consistent with previous studies at this site, basic chemical, physical and biological parameters more than 5 years of this study exhibited clear seasonal patterns of winter mixing, upwelling in spring and summer stratification (Fig. 1) (Chow *et al*. 2013; Cram *et al*. 2015). Primary productivity and bacterial production typically peaked in March following winter mixing, and total prokaryotic and total virus abundance most often peaked in April. Total *Synechococcus* abundance peaked slightly later in April and June. *Prochlorococcus* were usually less abundant than *Synechococcus*, and in contrast gradually and slightly increased in abundance as the year progressed.

High-throughput ITS sequencing was first used to characterize cyanobacterial community structure at the ecotype level. *Prochlorococcus* was almost always completely comprised of the high-light and cold-temperature adapted ecotype HLI (Fig. 2A). A few pulses of the low-light adapted ecotype LLI occurred in surface waters during winter months when vertical mixing presumably brought these deeper populations to the surface. High-light adapted ecotype HL2 and low-light adapted ecotypes LL2 and LL4 were occasionally detected but only at very low relative abundance (<1.1% of total *Prochlorococcus* sequences). *Synechococcus* was dominated by one ecotype, IV, but at a few time points, ecotype I reached comparable or higher levels (Fig. 2B). Ecotype I and two subclades identified within IV, labelled IVa and IVb (see below), exhibited apparent patterns of seasonal succession (Fig. 2C). Ecotype I usually peaked first during early spring (March to April) and was often closely followed by a peak in IVb, and IVa dominated the rest of the year. Several other ecotypes (II, III, CRD1 and CRD2) exhibited notable intermittent pulses, reaching up to 3%–33% of total *Synechococcus* sequences (Fig. 2B) while eight additional ecotypes (V, VI, VII, VIII, XV, XVII, XVIII, CB3, WPC1, 5.3-1) were detected but only rarely, and never above 2% (Supporting Information Fig. S2). While the dynamics of these low abundance, 'minor' ecotypes were not as consistent, III, XVII, XVIII, CRD1, CRD2 and WPC1 often appeared or increased in abundance in summer and fall, while the remaining rare ecotypes were mainly detected in spring (February through May) (Supporting Information Fig. S3).

Minimum entropy decomposition (MED) was used to identify distinct ASVs (or oligotypes) within ecotypes that are not a result of sequencing error (Eren *et al*. 2015). A total of 1655 and 1442 cyanobacterial ASVs were identified, respectively, within *Prochlorococcus* and *Synechococcus* corresponding to tens, hundreds or up to ~1400 unique ASVs within their respective ecotypes (Table 1). For ecotypes containing >30 ASVs, rarefaction analysis indicated that we have well sampled the total diversity of ASVs in most ecotypes with exceptions being

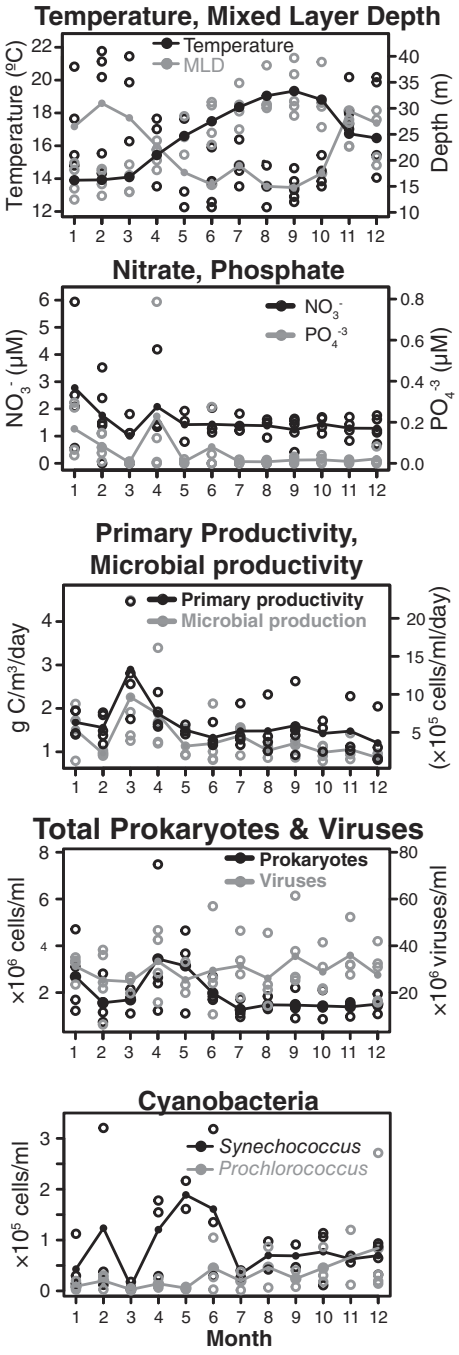

Fig. 1. Patterns in environmental and bulk biotic measurements in surface waters at SPOT over the 5-year study period, binned by the month samples were taken. All individual measurements are shown as open circles, and filled circles and lines depict monthly mean values. Mixed layer depth is abbreviated 'MLD'.

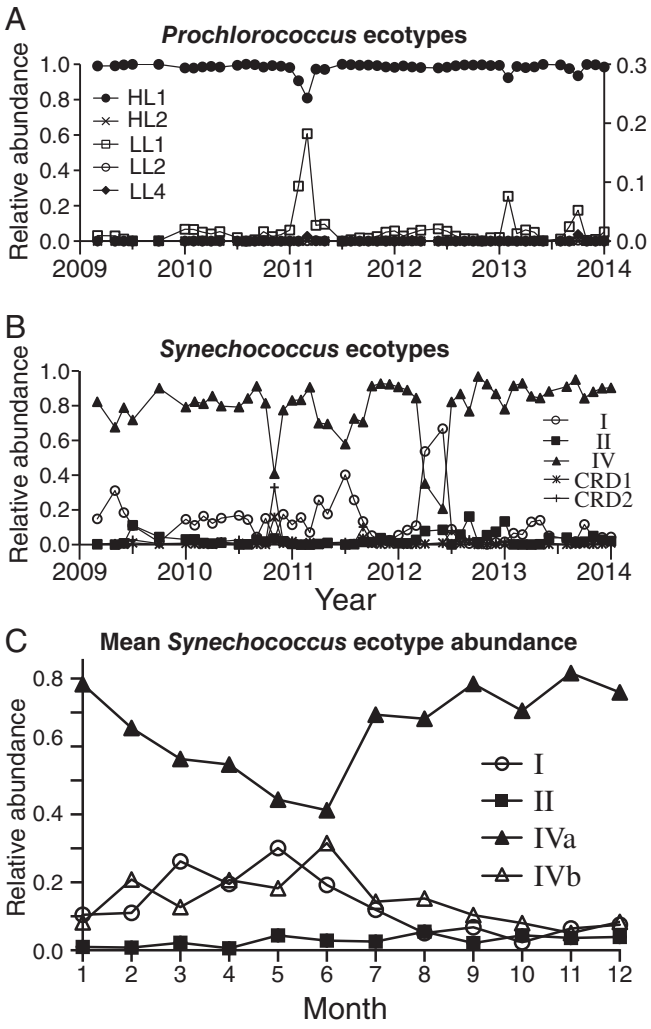

Fig. 2. Ecotype dynamics in surface waters at SPOT from 2009 to 2014 determined by high-throughput sequencing of the ITS locus. A. The relative abundance of *Prochlorococcus* ecotypes. HL1 is plotted according to the scale on the left y axis while all other ecotypes are plotted with the right y axis. B. Relative abundance of five abundant *Synechococcus* ecotypes seen at SPOT. Relative abundances of all other ecotypes are shown in the Supporting Information Fig. S2. C. Monthly mean ecotype abundances for *Synechococcus* ecotypes I and II and subclades IVa and IVb. Mean ecotype abundances for all other ecotypes are shown in Supporting Information Fig. S3.

ecotypes II, VI, VIII and XV where Chao1 estimates were markedly above the number of ASVs found and rarefaction curves did not appear to have reached complete saturation (Table 1, Supporting Information Fig. S4). To further characterize ASV dynamics, we only considered those ASVs that comprised ≥0.5% of the respective relative abundance of *Prochlorococcus* or *Synechococcus* sequences in at least one sample. This was done to exclude possible artefacts for rare, low-abundance ASVs. Even with this conservative criterion, there were tens to hundreds of ASVs within the various ecotypes, again highlighting rich within-ecotype diversity (Table 1). Bar graphs of all the ASVs meeting the above criterion were plotted for the four ecotypes containing ≥10 ASVs (HL1, I, II and IV) and provide a broad view of persistence of ASVs over multiple years and dynamic changes in within-ecotype ASVs across seasons and years (Fig. 3). *Prochlorococcus* ecotype HLI and *Synechococcus* ecotype I both contained a single

Table 1. Number of MED ASVs identified in SPOT surface water communities.

Taxon	Number of all ASVs found	Chao estimate of total ASVs (±S.E.)	Number of ASVs found at ≥0.5% in at least one time point
Prochlorococcus ecotypes			
Pro HL1	1393	1393 ± 1.1	149
Pro HL2	13	16 ± 3.4	-
Pro LL1	242	243 ± 1.0	9
Pro LL2	1	1 ± 0	-
Pro LL4	6	6 ± 1.3	1
Prochlorococcus sum	1655		159
Synechococcus ecotypes			
I	210	243 ± 20	19
II	84	120 ± 22	14
III	137	138 ± 1.4	4
IV	887	893 ± 4.1	88
V	4	6 ± 3.7	1
VI	11	46 ± 25	1
VII	3	3 ± 1.3	1
VIII	17	41 ± 20	1
XV	10	28 ± 23	1
XVII	4	4 ± 0	–
XVIII	6	6 ± 0.4	1
CB3	2	2 ± 0	–
CRD1	23	23 ± 0	4
CRD2	41	42 ± 1.7	6
5.3-1	3	3 ± 0	–
Synechococcus sum	1442		141
Unclassified	16		–

ASV that consistently made up a large portion (often >30%) of the ecotype populations over time. *Synechococcus* II was less skewed in community evenness and appeared to have more dynamic changes in fine-scale, within-ecotype structure with transient changes in ASV dominance. Splitting ecotype IV into its two major subclades IVa and IVb revealed a more even distribution of variants in the less abundant subclade IVb.

Bubble plots of these select ecotypes more clearly revealed temporal dynamics of fine-scale variants within ecotypes and that ASVs largely exhibit distinct patterns from one another, consistent with them representing ecologically distinct populations (Figs. 4 and 5). One major exception however is *Synechococcus* ecotype IV, which contains a large subclade of ASVs that exhibit very similar dynamics (Fig. 5). Many ASVs within this subclade, IVb, indeed showed strong autocorrelation with one another (Pearson correlation, $p < 0.05$ with Bonferroni correction). In contrast, the remaining IV ASVs belonging to subclade IVa only showed distinct clusters of autocorrelation among smaller groups (generally 2–5) of closely related ASVs. Such autocorrelation of small numbers of closely related ASVs was also seen for other ecotypes, possibly indicating that these ASVs represent effectively neutral variants over the study period within those autocorrelated groups. Overall, it appears that many ASVs exhibit distinct abundance patterns from each other, whereas subclade IVb ASVs had very similar patterns to each other over time.

Previous work at SPOT clearly shows that microbial community structure exhibits strong, predictable patterns in community structure that correlate with seasonal cycling of physical and chemical conditions (Fuhrman *et al*. 2006; Chow *et al*. 2013; Cram *et al*. 2015). Such patterns can be detected using plots of average pairwise Bray-Curtis (BC) similarity scores when binning sample pairs by the number of months separating them (Chow *et al*. 2013; Cram *et al*. 2015). Data in such plots that significantly fit sinusoidal curves with a period of 12 months

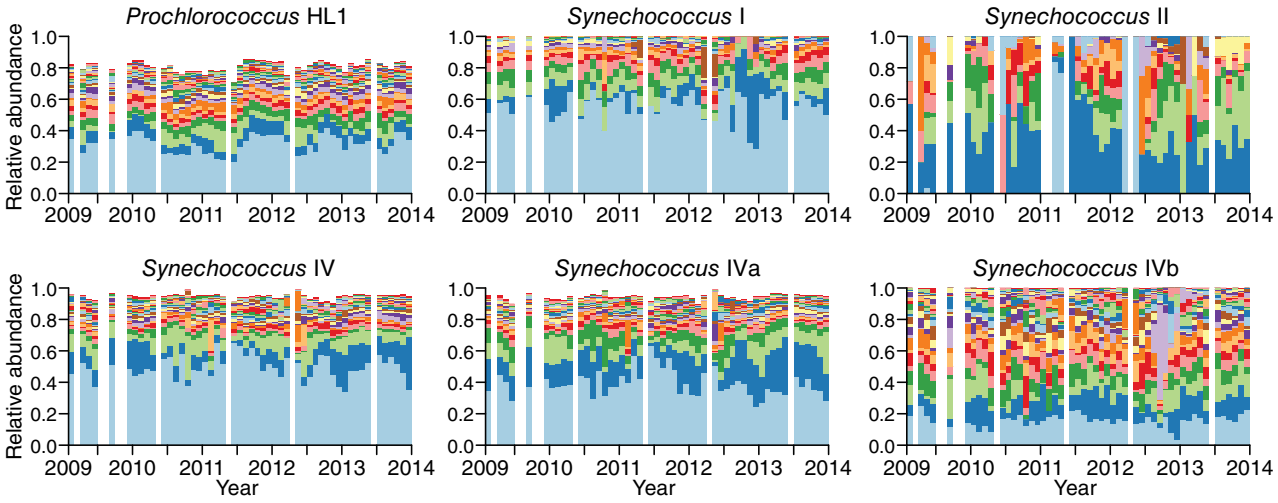

Fig. 3. Relative abundances of ASVs within ecotypes in surface waters at SPOT from 2009 to 2014. All ASVs that occurred at a relative abundance of ≥0.5% (relative to the total cyanobacterial reads in the sample) in at least one time point are plotted. Missing bars are dates for which water samples or DNA for sequencing were not available.

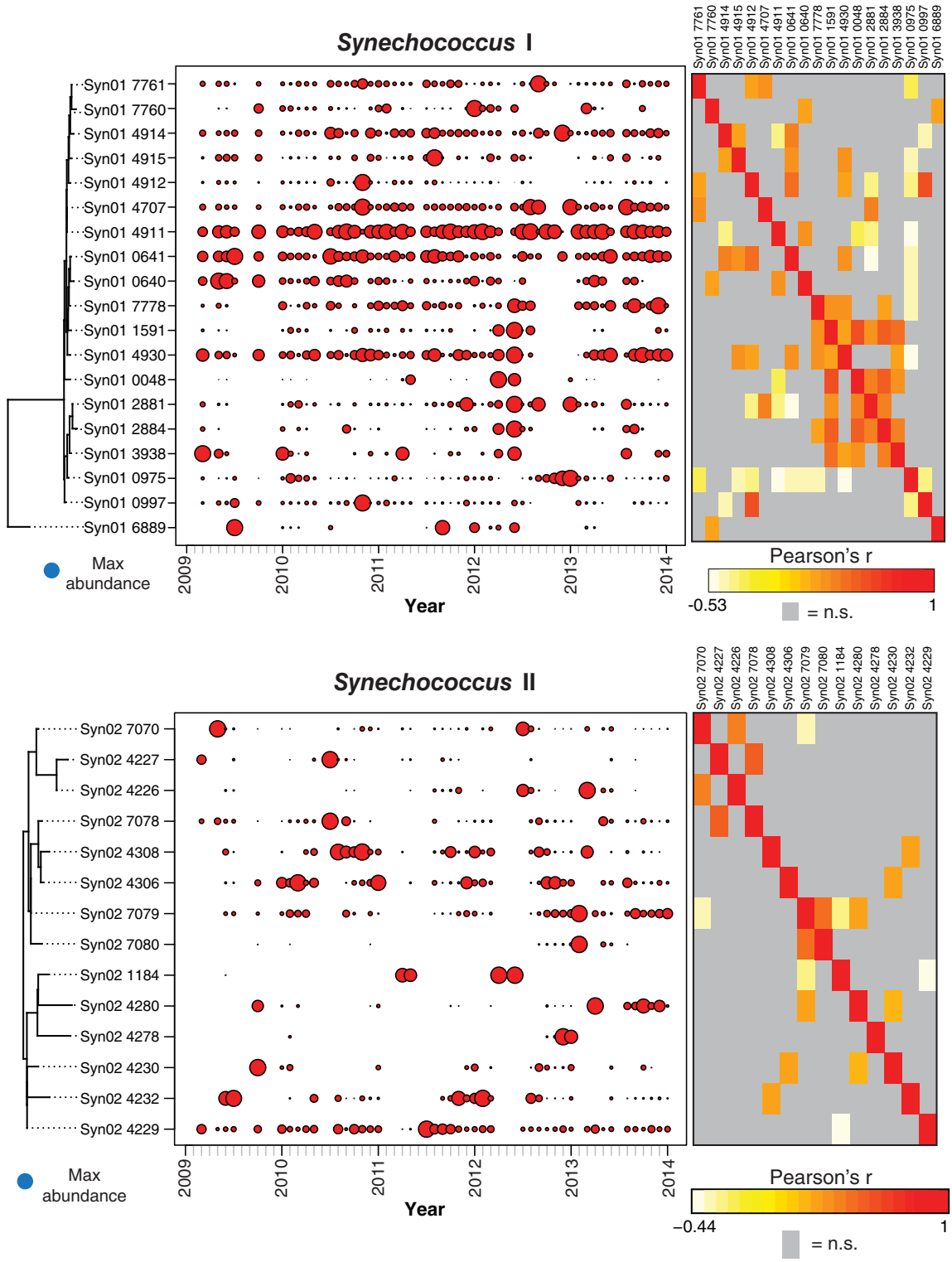

Fig. 4. Abundances of and correlations between genotypes within *Synechococcus* ecotypes I and II over 5 years in surface waters at SPOT. Trees on the left depict the phylogenetic relationship of ASVs. Dots in the middle panels depict the relative abundance of each ASV over time. Dots are scaled for each ASV such that abundances are normalized linearly to the maximum abundance in the time series (blue circle). Panels on the right depict significant Pearson correlations between genotypes ($p < 0.05$ with Bonferroni correction) whereby the horizontal order of genotypes from left to right (see ASV names as the top of the panel) is the same as the order of ASVs from top to bottom in the corresponding tree. sponding tree.

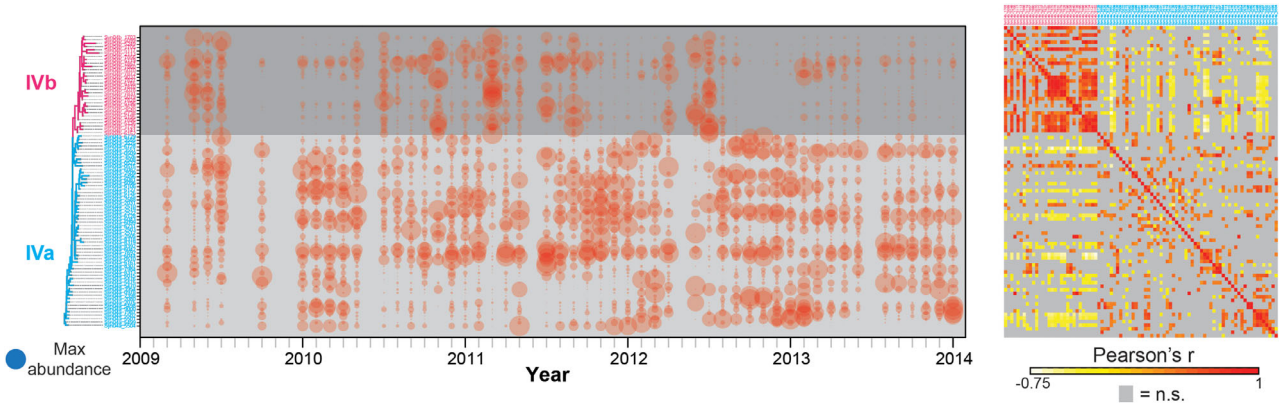

Fig. 5. Abundances of and correlations between ASVs within *Synechococcus* ecotype over 5 years in surface waters at SPOT. Trees on the left depict the phylogenetic relationship of ASVs, and ASVs and corresponding tree branches assigned to the emergent subclades IVa and IVb are coloured blue and pink respectively. Dots in the middle panels depict the relative abundance of each ASV over time. Dots are scaled for each ASV such that abundances are normalized linearly to the maximum abundance in the time series (blue circle). The panel on the right depicts significant Pearson correlations between ASVs ($p < 0.05$ with Bonferroni correction) whereby the horizontal order of ASVs from left to right (see ASV names as the top of the panel) is the same as the order of ASVs from top to bottom as in the corresponding tree.

demonstrate that these communities possess repeatable and seasonal patterns. The BC plot of ecotype community structure for *Synechococcus* yielded a sinusoidal curve that was significantly correlated, with a period of 12 months (Fig. 6).

Zooming in at community structure within ecotypes, BC plots for *Prochlorococcus* ecotype HL1 and *Synechococcus* ecotype IV also exhibited significantly sinusoidal curves with a period of 12 months ($p < 0.001$, $R^2 = 0.45$ for both). *Synechococcus* ecotype I communities had no obvious sinusoidal pattern ($p = 0.52$), and instead BC scores dropped quickly and plateaued with increasing months between samples. *Synechococcus* ecotype II also did not exhibit a significantly sinusoidal pattern with a 12 month period ($p = 0.23$); however, qualitatively, BC scores did fluctuate cyclically but with an inconsistent and apparently decreasing period. Considering that ecotype IV comprised two apparent subclades with distinct dynamics of ASVs within them, BC plots were generated separately for IVa and IVb subclade communities. Subclade IVa BC scores did not exhibit a significant cyclic pattern ($p = 0.09$) and had a plateau pattern similar to that of ecotype I. Subclade IVb communities on the other hand

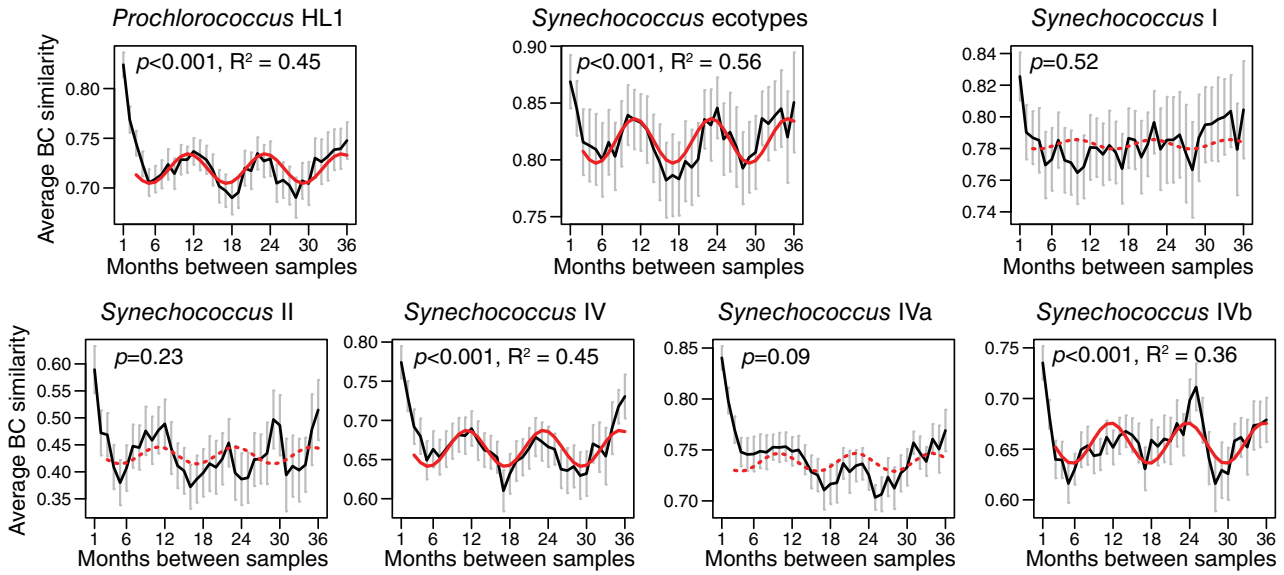

Fig. 6. Detection of repeated, seasonal patterns in cyanobacterial community composition using BC similarities. Plots show the mean Bray–Curtis similarity values for pairs of samples binned by the number of months separating them, and each plot shows the analysis for different communities – ecotype composition or within-ecotype or within-subclade ASV composition. Black lines depict the mean values and grey lines show standard deviations. Red lines depict sine curve fitting to the mean values, and significance levels and R^2 values (if significant, $p < 0.05$) are shown. Solid red lines indicate significant sine fitting and dashed lines indicate that the sine fitting was not significant.

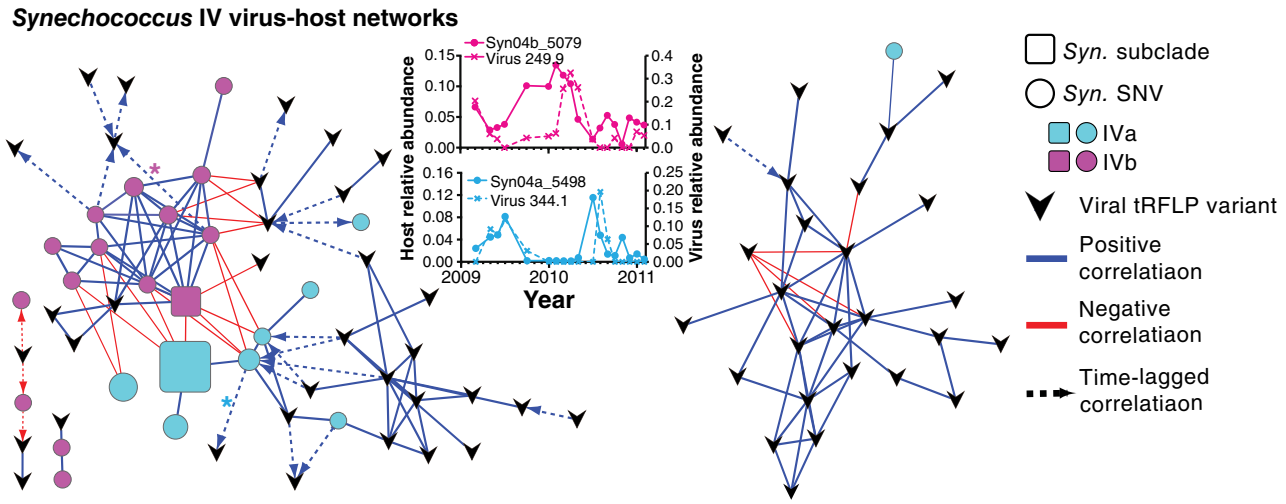

Fig. 7. Network depicting correlations between *Synechococcus* ecotype IV ASVs and viral T4-like g23 tRFLP variants. Only strong ($\rho \geq 0.6$) and significant ($p < 0.05$) interactions are shown and notably no correlations were found between *Synechococcus* ecotype IV ASVs and environmental parameters that met those criteria. Nodes depict host or viral populations whereby the size of the symbols for *Synechococcus* ecotype subclades (IVa or IVb; squares) or ASVs (circles) are scaled to their mean relative abundance over the study period. Lines depict positive (blue) or negative (red) correlations, and time-lagged interactions are depicted dashed lines with arrows whereby the node to which the arrow points is the one that lags. The inset graphs depict the relative abundances of two representative examples of virus-host pairs (marked with asterisks in the network) that exhibit time-lagged dynamics. Similar networks are shown for *Synechococcus* I and *Prochlorococcus* HL1 and HL2 ecotypes in the Supporting Information Fig. S6.

had significant cyclic patterns with a period of 12 months ($p < 0.001$, $R^2 = 0.36$), although not as strong as results of all ecotype IV ASVs or *Synechococcus* ecotype-level patterns.

Cyclic patterns in BC scores for *Synechococcus* ecotypes, *Prochlorococcus* HLI and *Synechococcus* sublcade IVb ASVs support that they possess predictable, seasonal patterns in community structure, presumably in response to strong seasonal patterns of the physical and chemical conditions of the water. Non-metric multidimensional scaling (NMDS) analysis was used to further assess whether communities at these different levels are correlated with environmental conditions (Supporting Information Fig. S5). Analysis at the level of ecotypes and ASVs within ecotypes both demonstrated significant correlation between several parameters to ordination of community structure, again supporting that fine-scale community structure within ecotypes is controlled in part by abiotic environmental parameters.

Network analysis of cyanobacteria and viral populations and abiotic and bulk biotic parameters

Network analysis of correlations among cyanobacterial community abundances, environmental factors, and viral community data was used further to assess what interactions may be influence cyanobacterial community structure. Viral community structure was assessed using previously published tRFLP data of the *g23* marker gene for T4-like viruses (Chow and Fuhrman 2012) for two overlapping years for which we also had ITS data. Extended local similarity analysis (eLSA) was used to identify correlations between parameters including time-delayed relationships. To simplify our analysis, networks were only examined for strong significant relationships ($\rho \geq 0.6$, Spearman correlation). In no cases did we find any strong correlations between environmental parameters and individual cyanobacterial ecotypes or ASVs. Strong correlations were instead frequently found between viral *g23* variants and cyanobacterial ASVs. These correlations were summarized in a network view whereby nodes represent ecotypes, ASVs, viral variants, or environmental parameters and lines connecting nodes indicate significant correlations in dynamics between them (Fig. 7). Emergent subnetworks were observed involving *Synechococcus* IV ASVs, where distinct groups of viral variants were significantly correlated to either IVa or IVb ASVs. In some cases, time-lagged correlations were seen where specific viral variants followed particular host ASVs, consistent with a Kill the Winner virus–host interaction model (Thingstad 2000); however, time-lagged correlations between cyanobacterial and viral variants were also seen in the opposite direction. These subnetworks also reveal a notable pattern of interconnection between IVb ASVs, consistent with what was observed in Fig. 5, and likewise significant negative correlations were detected between IVa and IVb ASVs.

Similar patterns were seen in *Prochlorococcus* HL1 networks, including a highly interconnected subnetwork of many HL1 ASVs (Supporting Information Fig. S6A). As

Table 2. The value of R^2 from variation partitioning analysis determining the portion of host community composition variance explained by environmental factors or viral community structure.

	RDA		Partial RDA		
Host community level tested	Cyanobacteria community vs. environmental factors	Cyanobacteria community vs. viral community	Cyanobacterial community vs. environmental factors alone (env|viral)	Cyanobacterial community vs. viral factors alone (viral|env)	Unexplained variance
Relative host abundance					
Syn. all ecotypes	0.26*	0.24*	0.18*	0.17*	0.57
Syn. ecotypes I, II, IV	0.31*	0.26*	0.22*	0.26	0.53
Syn. I, II, IV ASVs	0.15**	0.13**	0.04	0.02	0.83
*Syn.*I ASVs	0.21**	0.32***	−0.02	0.09	0.70
*Syn.*II ASVs	0.15**	0.11*	0.05	0.01	0.84
*Syn.*IV ASVs	0.14*	0.19**	−0.02	0.02	0.83
*Syn.*IVa ASVs	0.23**	0.23***	0.01	0.01	0.76
*Syn.*IVb ASVs	0.12*	0.38**	−0.02	0.23*	0.65
Pro. HL1 ASVs	0.37***	0.45***	0.04	0.13*	0.51
Absolute host abundance					
Syn. all ecotypes	n.s.	0.27**	n.s.	n.s.	n/a
Syn. ecotypes I, II, IV	0.18*	0.55**	0.04	0.42**	0.57
Syn. I, II, IV ASVs	0.17*	0.49***	−0.06	0.25**	0.51
*Syn.*I ASVs	0.27**	0.48***	−0.02	0.20*	0.53
*Syn.*II ASVs	0.22**	0.16**	0.09	0.04	0.75
*Syn.*IV ASVs	0.21*	0.51***	−0.06	0.24*	0.55
*Syn.*IVa ASVs	0.27**	0.50***	−0.02	0.20*	0.53
*Syn.*IVb ASVs	0.08	0.49***	−0.01	0.40**	0.52
Pro. HL1 ASVs	0.39***	0.47***	0.06	0.14*	0.46

n.s. = estimate of contribution to variance not significantly supported by ANOVA ($p > 0.05$).
Asterisks indicate the level of ANOVA support for the estimated contribution to host variance: *$p < 0.05$; **$p < 0.01$; ***$p \leq 0.001$.

with subclade IVb, this subnetwork may represent effectively neutral variants within a coherent HL1 subclade, but the phylogenetic signal for this locus could not readily resolve any clear subclades (data not shown). *Synechococcus* I networks exhibited a high degree of interconnection between viral variants and only one connection between a single cyanobacterial ASV and a viral variant (Supporting Information Fig. S6B). The highly interconnected viral variants could represent ecologically and evolutionarily distinct viral populations that infect similar sets of host populations or effectively neutral genetic variants of a single viral population.

Variation partitioning analysis

Variation partitioning analysis (redundancy analysis [RDA]) and partial RDA) (Peres-Neto *et al*. 2006) was used to determine the relative importance of abiotic environmental factors and viral communities on impacting cyanobacterial community structure. In particular, we were interested in comparing the individual contribution of environmental factors (env|viral) or viral community structure (viral|env) in explaining host community structure at the ecotype vs. within-ecotype, ASV level. Variation partitioning analyses were assessed using relative host abundances or absolute host abundances (top and bottom portions of Table 2). Only in two cases, the relative abundance of all *Synechococcus* ecotypes and ecotypes I, II and IV, did ANOVA significantly support the estimated contribution of environmental parameters alone (env|viral) to host variance. In contrast, ANOVA more frequently significantly supported (10 of 18 analyses) the estimated contribution of viral community structure alone (viral|env) to host variance, and most of these were for ASV-level analyses. Only for *Synechococcus* ecotype relative abundance did ANOVA significantly support both estimates of viral community structure and environmental factors individually, and environmental factors explained marginally more of *Synechococcus* ecotype variance than viral community structure (0.18 vs. 0.17). The lack of ANOVA support for both factors (environmental and viral alone) in most analyses, therefore, made it difficult to confidently compare the relative strength of viral vs. environmental factors; nonetheless, for most ASV-level analyses (11 of 14), viral community structure more strongly explained host community variance than environmental factors. Furthermore, the amount of variance viral community structure explained was frequently much higher than that explained by environmental factors. For example, viral community structure explained 23% or 40% of the host variance for *Synechococcus* subclade IVb genotypic structure (using relative and absolute abundances respectively) while environmental factors explained a negligible amount of host variance. Finally, variation partitioning for just the dominant *Synechococcus* ecotypes I, II and IV supported that for relative host abundance, environmental factors alone significantly explain

ecotype-level host variance ('*Syn.* ecotypes I, II, IV'), but did not when zooming in at the ASV-level ('*Syn.* I, II, IV ASVs'). In addition, variation partitioning results generally support a model where environmental factors more strongly explain ecotype-level variance while viral community structure more strongly explains within-ecotype, ASV-level host community variance.

Discussion

Here, we have developed and implemented an Illumina-based, high-throughput amplicon sequencing method to characterize fine-scale community composition of marine cyanobacteria using the ITS locus. The ITS amplicon sequencing method produced accurate measurement of ecotypes in mixed samples as assessed with mock communities of known mixtures of ITS template DNA comprised of up to four *Prochlorococcus* ecotypes and nine *Synechococcus* ecotypes. We emphasize the importance of assessing the accuracy of amplicon sequencing methods, which is often not done (Parada *et al*. 2016). In the various tests of even and uneven mock communities of *Synechococcus* ecotypes alone or mixed with varying amounts of *Prochlorococcus* sequences, the relative abundance of *Synechococcus* could be measured with very good accuracy, usually within ~1.5-fold and not exceeding threefold of the expected relative abundances. This accuracy is comparable or exceeds that of qPCR assays (Zinser *et al*. 2006; Ahlgren and Rocap 2012), and this sequencing approach has the additional advantage of detecting genotypic variants within ecotypes. The inability to detect *Synechococcus* ecotypes (and in one case a *Prochlorococcus* ecotype) at very low abundances is consistent with the expected probabilities of read abundance for the sequencing depths used here (generally 10 000 sequences per sample). This drop-out effect occurred for ecotypes mixed at ~0.85% or less, and this helped inform the threshold we applied to filter out low-abundance ASVs in analyses of natural communities (those that never exceeded 0.5% of the whole sample). The measured *Prochlorococcus* ecotype abundances from mixed mock communities were also largely similar to expected abundances.

We applied our high-throughput cyanobacterial ITS sequencing method to monthly surface water samples spanning 5 years at the SPOT site. ITS sequencing revealed thousands of distinct ASVs such that the dominant *Prochlorococcus* and *Synechococcus* ecotypes contained hundreds of ASVs. This level of intra-ecotype diversity is comparable to results from Larkin *et al*. (2016) that found hundreds of operational taxonomic units (OTUs) within *Prochlorococcus* ecotypes using high-throughput sequencing of the ITS-2 (our method targets the ITS-1 region). Our results are also broadly consistent from a previous genome-based study that suggests the *Prochlorococcus* HL2 ecotype may contain hundreds or thousands of distinct genomic types (Kashtan *et al*. 2014). This study also found that full-length ITS phylogeny is congruent with and could distinguish intra-ecotype populations determined by whole genome analyses, supporting the utility of the ITS marker for tracking fine-scale cyanobacterial subpopulations. It is however important to note that our method only examines 160 bp of the ITS-1 region (see Experimental Procedures) that is less divergent than the ITS-2 region, and in practice this region cannot distinguish all of the subclades observed genomically by Kashtan *et al*. (Supporting Information Fig S7). Larkin *et al*. (2016) have developed a similar ITS-based amplicon method for the more divergent ITS-2 region, but specifically only for *Prochlorococcus* sequences. A refined ITS amplicon approach targeting the ITS-2 region for *Synechococcus* and *Prochlorococcus* or utilizing advances in increased read lengths has the potential to resolve even more fine-scale populations. Nonetheless, our ITS method as it stands has identified a level of intra-ecotype diversity not previously observed for *Synechococcus* and comparable levels for *Prochlorococcus.* Furthermore, this method can easily be used to quantitatively track these fine-scale ASVs across hundreds of samples in a way not feasible with current single-cell or metagenomic approaches, largely due the prohibitive cost to sequencing high numbers of cells or at necessary high sequencing depths respectively.

This study highlights the importance of zooming in at the fine-scale to identify important, biologically relevant patterns in subpopulations that may be otherwise missed. On a practical level, fine-scale analysis was important in revealing two emergent subclades within *Synechococcus* IVa and IVb. These two clades peak in abundance at different times of the year (Fig. 2C), and their within-subclade ASVs either do or do not exhibit repeatable seasonal patterns (Fig. 6), indicating fundamental differences in their biology. The IV subclades highlight a broader theme that genotypic composition differs fundamentally between ecotypes and subclades. Ecotypes and subclades exhibit broad differences in evenness, with some dominated by a single ASV (e.g. *Synechococcus* I and IVa and *Prochlorococcus* HL1) while others are more even (e.g. *Synechococcus* II and IVb) (Fig. 3). ASVs within some ecotypes qualitatively appear to have more dynamic and seemingly stochastic patterns than others (e.g. *Synechococcus* II vs. IVa). These differences could reflect fundamental differences in the selective or evolutionary forces shaping them and the particular strategies that they arrived at to adapt and thrive in this environment. While not a specific focus of our analysis here, the importance of these patterns in the context of microbial evolution and speciation warrants further investigation.

Most of these ASVs appear to exhibit distinct dynamics across the time-series studied. This is consistent with previous studies that have observed subclades within

ecotypes that have distinct dynamics (Tai and Palenik 2009; Tai *et al*. 2011) or distinct biogeographies (Farrant *et al*. 2016); however, here, we have identified many more, finely resolved ASVs than previously. The analysis of Bray-Curtis data also demonstrates that within-ecotype genotypic community composition of some ecotypes has repeatable seasonal patterns (Fig. 6). This parallels how ecotype-level community composition exhibits strong, repeating, seasonal patterns (Fig. 6), and previous work has firmly established that ecotypes are ecologically distinct based on the previous physiological and biogeographical and time-series evidence (Johnson *et al*. 2006; Zwirglmaier *et al*. 2008; Malmstrom *et al*. 2010; Sohm *et al*. 2015; Farrant *et al*. 2016). Together these data support the conclusion that many within-ecotype ASVs likewise represent ecologically distinct populations. It is however possible that some closely related ASVs are effectively neutral variants of the same subpopulation (Figs. 4 and 5). More broadly, the within-ecotype seasonal patterns are consistent with previous analyses of SPOT that show repeatable seasonal patterns in total prokaryotic composition due in large part to strong seasonal forcing of chemical and physical parameters (Chow *et al*. 2013; Cram *et al*. 2015). This therefore implicates such bottom-up factors as likewise important in shaping within-ecotype genotypic community structure. This conclusion is also supported by NMDS analysis of within-ecotype community structure (Supporting Information Fig. S5).

Previous studies have shown that a good portion of the variance in community composition of *Prochlorococcus* and *Synechococcus* can be explained by commonly measured bottom-up factors such as macronutrients, temperature and light (Johnson *et al*. 2006; Martiny *et al*. 2009). Authors frequently suggest that the remaining, unexplained variance is due partly to top-down pressure from grazers and/or viruses. Coleman and Chisholm (Coleman and Chisholm 2007) and Rodriguez-Brito *et al*. (Rodriguez-Brito *et al*. 2010) furthermore put forth models suggesting that phage interactions are important drivers of differentiation at the finest levels of diversity in marine cyanobacteria. The latter study specifically invokes the Kill the Winner (KTW) model (Thingstad 2000) whereby they suggest that at a broad level (e.g. the ecotype level), community structure is relatively stable, but at a finer, genotype-level scale, dynamic changes can occur due to sequential booms and busts of interacting viruses and hosts (see Fig. 6 in the study by Rodriguez-Brito *et al*. 2010). The fact that some cyanophage can infect a narrow range of cyanobacterial strains (Sullivan *et al*. 2003; Gregory *et al*. 2016) is generally consistent with the plausibility of KTW dynamics, but observations of complex, natural communities that may support or refute this model are sparse or non-existent.

Our study provides some of the best empirical data available to assess the relative importance of abiotic environmental and top-down viral factors on community structure at wider (ecotype) and finer (intra-ecotype, ASV) scales. Consistent with the above model, individual cyanobacterial ASVs did not exhibit strong correlations to abiotic factors but frequently did to individual viral ASVs (Figs. 4 and 5, Supporting Information Fig. S5). Co-variation networks also identified a few examples of time-lagged dynamics of viral and host variants that are consistent with KTW dynamics and more broadly classic predator–prey interactions. From variation partitioning and NMDS analysis of community structure, bottom-up controls are important at both ecotype and within-ecotype levels. Variation partitioning analysis results also supported that environmental factors more strongly explained *Synechococcus* ecotype-level variance, while for intra-ecotype analyses, viral community structure more often better explained host community variance than environmental factors (Table 2). These results together are all consistent with the proposed model based on KTW-like dynamics that seasonally forced abiotic factors are primary drivers of ecotype-level composition while viruses are more important in driving fine scale, intra-ecotype dynamics.

It is important to note that many cyanophage and myoviruses, including T4-like phage, can have broad host ranges and infect strains across ecotypes or across the two genera *Prochlorococcus* and *Synechococcus* (Sullivan *et al*. 2003; Gregory *et al*. 2016), so viruses likely have some impact at the broader ecotype or genus level. Similarly, the discovery of a new family of viruses (*Autolykiviridae*) with broad inter-species host ranges (Kauffman *et al*. 2018) highlights that bacterial communities may experience complex selective pressures from both narrow and broad range viruses. In a similar vein, studies have discovered within-ecotype variation in adaptation to bottom-up factors including macronutrients (Martiny *et al*. 2006) and temperature (Pittera *et al*. 2014). Such intra-ecotype variation in abiotic adaptations does not preclude our model, and instead we emphasize that both viral and bottom-up factors are important in within-ecotype population structure. The amount of host community variance unexplained by viral or environmental factors was high (0.46–0.84, Table 2) indicating that there are outstanding, unmeasured factors that are important to controlling host community structure. These may include organic nutrients for which cyanobacteria have differing capacities to utilize (Scanlan *et al*. 2009) and viral communities not detected by the *g23* tRFLP method used here, such as finer-scale genotypic variants or non-T4-like viruses (i.e. podoviruses and siphoviruses). Also poorly constrained but likely as important are differences in top-down nanoflagellate grazing susceptibilities

(Apple *et al*. 2011) and 'lateral' allelopathic interactions (Paz-Yepes *et al*. 2013). Therefore, measuring additional chemical parameters or grazer impacts will better inform our understanding of control of host community composition.

A complicating factor in our analysis however is that the viral genotypic data available (tRFLP analysis of the *g23* gene) assesses the diversity only of a particular superfamily of viruses, the T4-like myoviruses, which as a whole make up a significant portion of total marine viruses in surface waters (approximately 15% of total viral particles) (Breitbart *et al*. 2002; Angly *et al*. 2006; Williamson *et al*. 2008; Brum *et al*. 2013). More importantly, most cyanophage isolates infecting *Synechococcus* and *Prochlorococcus* are myoviruses (Sullivan *et al*. 2003; Gregory *et al*. 2016; Hanson *et al*. 2016). Isolation-independent viral-tagging methods (Deng *et al*. 2014) and culture-independent metagenomics (Roux *et al*. 2016) also support that myoviruses appear to be the dominant group of viruses infecting these cyanobacteria. The tRFLP viral data also are not sequence based, so some of the detected variants may not be cyanophage, such that some of network relationships and variation partitioning analyses may include non-specific correlations between cyanobacteria and non-cyanophage. Spurious, non-specific correlations are always a problem with co-variation analyses, but nevertheless this study utilizes some of the best available data to analyse both host and viral genotypic data at multiple diversity levels, along with bottom-up factors. Our results are also consistent with a recent paired analysis of host (16S rDNA) and viral (*g23* sequences) variants at SPOT that likewise saw that viral variants correlated better with fine-scale level 16S rDNA ASVs than broader taxa defined by 99% 16S rDNA identity (Needham *et al*. 2017). As more viral metagenomic sequences or high-throughput viral marker gene data become available, we will be better equipped to more definitively test these proposed models and virus–host dynamics.

While we have known for a long time that marine microbial communities contain large amounts of microdiversity, this study in sum raises the bar in terms of the magnitude of this diversity measured within marine *Synechococcus* in particular. Our results also demonstrate that many of these very fine-scale level variants appear to represent ecologically distinct populations, an emerging theme for studies using MED (or other single-base-resolved amplicon classification) on microbial communities from a wide variety of habitats (Callahan *et al*. 2016; Eren *et al*. 2016; Amir *et al*. 2017). This study highlights the importance and necessity of characterizing communities at a very fine scale, lest we do not fail to identify and characterize community structure at a biologically relevant scale. In particular, because viruses often can infect a narrow set of strains, we contend that teasing apart important virus–host interactions will require examination of hosts at these fine scales. Sequencing technology is now at a place that we can characterize these fine-scale populations, opening up to possibility to better address challenging and fundamental questions about the structure and drivers of microbial diversity, and in particular the role of virus–host interactions in shaping community structure.

Experimental procedures

Cyanobacterial ITS Illumina-based amplification and sequencing

Cyanobacterial ITS specific primers for use with the Illumina platform were designed using previous developed PCR primers and the adaptor, barcode and index strategy developed previously for 16S rRNA sequencing (Parada *et al*. 2016). The 16S-1247f (Rocap *et al*. 2002) and ITS-ar (Lavin *et al*. 2008) primers amplify the ITS-1 region spanning from the end of the 16S rRNA gene to the isoleucine tRNA within the ITS region to generate products of roughly 490 and 550 bp for high-light adapted *Prochlorococcus* and *Synechococcus.* These primers were joined to Illumina compatible sequences that contain appropriate adaptors, forward sequencing primers and 5 bp barcode and 6 bp index sequences for forward and reverse primers respectively (see Supporting Information for primer sequences). PCR reactions were performed using Platinum HiFi Taq polymerase (Invitrogen Corp.) with 2.5 mM total $MgCl_2$ and the following cycling conditions: initial denaturation step of 95°C for 2 min; 25 cycles of 95°C for 45 s; 55°C for 45 s; and 68°C for 90 s; and a final elongation step of 68°C for 10 min. Single PCR reactions were carried out using 1 μl of DNA at 1 ng μl^{-1}. PCR reactions were purified using Agencourt Ampure XP beads (Beckman Coulter Corp). These reactions were pooled and purified again, and resulting pools were sent for Illumina MiSeq 250 bp paried ended sequencing at the University of California Davis Genome Center. Sequence data have been submitted to the GenBank databases under BioProject accession number PRJEB12267.

Mock communities

To construct the mock community, plasmids containing cloned ITS sequences constructed previously in Ahlgren and Rocap (2012) were linearized with the restriction enzyme PstI and purified. Purified linear plasmids were quantified using PicoGreen fluorescence using lambda DNA standards in six replicate wells and ITS plasmids in triplicate wells. ITS plasmids were mixed at various ratios to generate mock communities with nearly equal proportions (even) or at various levels of relative abundances (staggered). Mock communities were diluted to

approximately 10 000 copies μl^{-1} to correspond to the estimated abundance of cyanobacterial DNA in natural samples.

SPOT samples

Surface water from the SPOT site located off the coast of Los Angeles (CA, USA) was sampled on a monthly basis for microbial analysis along with various physical and chemical parameters (Chow *et al*. 2013; Cram *et al*. 2015). Collection of environmental and biological parameters and collection and extraction of microbial fraction DNA for cyanobacterial ITS sequencing are described previously in the study by Cram *et al*. (2015). For this study, we focused on a 5 year data set of 50 samples from February 2009 to January 2014. Of the available 43 abiotic physical and chemical and bulk biotic parameters measured at SPOT, these were reduced to 26 variables by manual inspection and removal of highly co-correlated parameters. The final list of these parameters was month, average wind speed, precipitation, prokaryotic abundance, average wave period, dominant wave period, wave height, absolute phytoplankton abundance, satellite based chlorophyll *a*, chlorophyll max depth, day length, elapsed days, leucine production, mixed layer depth, nitrite concentration, nitrate concentration, phosphate concentration, excess phosphate concentration (P*), oxygen concentration, salinity, temperature, leucine turnover, virus to prokaryotic ratio (VPR), total virus abundance, Pacific Fisheries and Environmental Laboratory (PFEL) estimates of coastal upwelling and Sverdrup transport (SVD) (see Cram *et al*. 2015 for details). These are herein referred to as environmental parameters for simplicity, even though they include bulk biological parameters such as chlorophyll *a* and viral and prokaryotic abundances.

Sequence processing and phylogenies

Because the forward read mostly comprised the end of the 16S rRNA gene and has limited sequence information for discrimination, analysis was performed on the reverse reads covering most of the ITS-1 region. Because the reverse reads for one of the sequencing runs had degraded quality after roughly 180 bp, sequences were trimmed to 160 bp, and a quality filter of $Q \geq 20$ was applied. Unique ITS oligotypes were identified and their read counts were determined using minimum entropy decomposition (MED) on all of the combined trimmed, quality filtered reads for the environmental SPOT, negative control and mock community samples. The resulting oligotype sequences were assigned to their corresponding ecotypes using blastn to a reference database of full length ITS sequences from previous publications (Rocap *et al*. 2002; Ahlgren and Rocap 2006; Choi and Noh 2009; Ahlgren and Rocap 2012; Huang *et al*. 2012) for which ecotype assignments are definitive based on phylogeny. Only 16 of 3113 total oligotypes did not match a cyanobacterial ITS sequence. One was a homopolymer of cytosines, and the other 15 are apparently from Gammaproteobacteria as they have ≥92% identity to the ITS of a *Candidatus Thioglobus* strain. Except for one *Thioglobus* sequence which reached up 6.5% of total sequences in one sample, these non-cyanobacterial sequences were minor constituents, never exceeding 0.41% of total ITS sequence abundances, and were excluded from subsequent analyses. Phylogenies of ASVs within select ecotypes were constructed by first aligning sequences with CLUSTALW (Larkin *et al*. 2007) then constructing trees with the F84 DNA substitution model and the FITCH algorithm in PHYLIP (Felsenstein 2005).

Community statistical analyses

Unless noted, all R functions used for community statistical analyses were standard functions within R or from the package 'vegan' (Oksanen *et al*. 2017). Rarefaction was performed on MED ASVs for the 5 year data set using the function *specpool*, which reports Chao estimates of alpha diversity. ASV counts were pooled across all samples. Bray-Curtis (BC) dissimilarity scores were computed for cyanobacterial community composition using the function *vegdist* and pairwise scores were binned by the number of months separating the samples. Sinusoidal curve fitting of averaged, month separation-binned BC scores, using binned scores from 3 to 36 months since pairs separated by only 1 or 2 months often had much higher average BC scores above the typical amplitude observed for subsequent cycles of BC values.

Extended local similarity analysis (eLSA) (Xia *et al*. 2011) was used to identify significantly co-varying cyanobacterial ecotypes ($n = 26$) and ASVs ($n = 89$), viral *g23* (major capsid protein) variants ($n = 207$), environmental parameters ($n = 26$, see above) for more than 2 years of time points (February 2009 to December 2010, $n = 17$ samples) for which *g23* and cyanobacterial ITS data were both available. The input for cyanobacterial community composition was relative abundances of ecotypes within *Prochlorococcus* and *Synechococcus* and relative abundances of ASVs within their respective genera. For viral variant data, we used peak areas of terminal restriction fragment length polymorphism analysis of the *g23* gene (Chow and Fuhrman 2012), and only the data for 5′ terminal products digested with enzyme HincII were used. Networks were constructed using Cytoscape (Shannon *et al*. 2003) and examined only for significant ($p < 0.0015$ and $Q < 0.05$) and strong Spearman correlations ($\rho \geq 0.6$).

Nonmetric multidimensional scaling (NMDS) analysis of relative cyanobacterial community composition was

performed using the function *metaMDS* with a minimum of 100 random starts in search of a stable solution (try = 100) on BC distances of ecotype or ASV relative abundances for the 5 year data set (n = 47 samples; October, November and December 2013 were excluded due to missing corresponding environmental data). The function *envfit* was then used to determine which environmental parameters (n = 26, see above) were significantly correlated to sample clustering using 999 permutations.

Variation partitioning was performed to determine the relative strength by which environmental parameters and viral community structure can explain the variation of cyanobacterial community structure over the 2 year data set. Variation partitioning analyses were performed for *Synechococcus* ecotype community composition and the genotypic composition within the following groups: *Prochlorococcus* ecotype HL1 and *Synechococcus* ecotypes/sublcades I, II, IV, and IVa, and IVb. Host community structure was analysed using relative or absolute abundances. Absolute abundances were calculated by multiplying relative abundances to corresponding total *Prochlorococcus* or *Synechococcus* concentrations determined by flow cytometry. Detrended correspondence analysis was first used to determine the appropriate response model, canonical correspondence analysis or redundancy analysis (RDA), with the function '*decorana*', and for all cyanobacterial community data sets, RDA was selected because the longest gradient lengths (DCA1) were all less than three. Environmental data were first standardized using the function *decostand.* Relative abundances of viral variants, cyanobacterial ecotypes and within-ecotype ASVs were first Hellinger transformed. Because explanatory variables (environmental factors and viral variants) in some cases outnumbered the observations (sample size), these data were reduced using forward selection of principal component axes using the function *ordistep* and requiring that added axes have p values ≤0.1 (option p = 0.1). RDA and partial RDA were then performed on these data using the function *rda* and the significance of RDA results were assessed with ANOVA using 200 steps and up to 200 permutations. In particular, the unique fraction of variance explained by the environmental data alone was determined using the command varpart(cyano,envir,viral) where 'cyano' is the cyanobacterial community composition, 'viral' is the viral *g23* genotypic composition and 'envir' is the environmental parameter data, and the unique fraction of variance explained by the viral data alone was determined using the command varpart(cyano,viral,envir).

Acknowledgements

We thank the numerous individuals that contributed to the collection of samples and data from the San Pedro Ocean Time-Series including the crew of the R/V Yellowfin. This work was supported by funding from the National Science Foundation (NSF) Microbial Observatory, Biological Oceanography and Dimensions in Biodiversity programs (grant nos. 0703159, 1031743, 1136818, and 1737409) to JAF and the Gordon and Betty Moore Foundation Marine Microbiology Initiative (grant GBMF3779) to JAF.

References

Acinas, S.G., Klepac-Ceraj, V., Hunt, D.E., Pharino, C., Ceraj, I., Distel, D.L., and Polz, M.F. (2004) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**: 551–554.

Ahlgren, N., 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.

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.

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**: e00191-16.

Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., *et al*. (2006) The marine viromes of four oceanic regions. *PLoS Biol* **4**: e368.

Apple, J., Strom, S., Palenik, B., and Brahamsha, B. (2011) Variability in protist grazing and growth on different marine *Synechococcus* isolates. *Appl Environ Microbiol* **77**: 3074–3084.

Biller, S.J., Berube, P.M., Lindell, D., and Chisholm, S.W. (2015) *Prochlorococcus*: the structure and function of collective diversity. *Nat Rev Microbiol* **13**: 13–27.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., *et al*. (2002) Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A* **99**: 14250–14255.

Brum, J.R., Schenck, R.O., and Sullivan, M.B. (2013) Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *ISME J* **7**: 1738–1751.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581–583.

Chase, A.B., Karaoz, U., Brodie, E.L., Gomez-Lunar, Z., Martiny, A.C., and Martiny, J.B.H. (2017) Microdiversity of an abundant terrestrial bacterium encompasses extensive variation in ecologically relevant traits. *MBio* **8**: e01809-17.

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

Chow, C.E., and Fuhrman, J.A. (2012) Seasonality and monthly dynamics of marine myovirus communities. *Environ Microbiol* **14**: 2171–2183.

Chow, C.E.T., Sachdeva, R., Cram, J.A., Steele, J.A., Needham, D.M., Patel, A., *et al*. (2013) Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California bight. *ISME J* **7**: 2259–2273.

Coleman, M.L., and Chisholm, S.W. (2007) Code and context: *Prochlorococcus* as a model for cross-scale biology. *Trends Microbiol* **15**: 398–407.

Cram, J.A., Chow, C.E.T., Sachdeva, R., Needham, D.M., Parada, A.E., Steele, J.A., and Fuhrman, J.A. (2015) Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *ISME J* **9**: 563–580.

Cram, J.A., Parada, A.E., and Fuhrman, J.A. (2016) Dilution reveals how viral lysis and grazing shape microbial communities. *Limnol Oceanogr* **61**: 889–905.

Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., Poulos, B.T., Weitz, J.S., Hugenholtz, P., and Sullivan, M.B. (2014) Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature* **513**: 242–245.

Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G., and Sogin, M.L. (2013) Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol* **4**: 1111–1119.

Eren, A.M., Morrison, H.G., Lescault, P.J., Reveillaud, J., Vineis, J.H., and Sogin, M.L. (2015) Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J* **9**: 968–979.

Eren, A.M., Sogin, M.L., and Maignien, L. (2016) Editorial: new insights into microbial ecology through subtle nucleotide variation. *Front Microbiol* **7**: 1318.

Farrant, G.K., Dore, H., Cornejo-Castillo, F.M., Partensky, F., Ratin, M., Ostrowski, M., *et al*. (2016) Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci U S A* **113**: E3365–E3374.

Felsenstein, J. (2005) *PHYLIP (Phylogeny Inference Package)*. Seattle: Department of Genome Sciences, University of Washington.

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* **110**: 9824–9829.

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V., and Naeem, S. (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci U S A* **103**: 13104–13109.

Gregory, A.C., Solonenko, S.A., Ignacio-Espinoza, J.C., LaButti, K., Copeland, A., Sudek, S., *et al*. (2016) Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics* **17**: 930.

Hanson, C.A., Marston, M.F., and Martiny, J.B. (2016) Biogeographic variation in host range phenotypes and taxonomic composition of marine cyanophage isolates. *Front Microbiol* **7**: 983.

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.

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

Hutchinson, G.E. (1961) The paradox of the plankton. *Am Nat* **95**: 137–145.

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 senomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344**: 416–420.

Kauffman, K.M., Hussain, F.A., Yang, J., Arevalo, P., Brown, J. M., Chang, W.K., *et al*. (2018) A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature* **554**: 118–122.

Larkin, A.A., Blinebry, S.K., Howes, C., Lin, Y., Loftus, S. E., Schmaus, C.A., *et al*. (2016) Niche partitioning and biogeography of high light adapted *Prochlorococcus* across taxonomic ranks in the North Pacific. *ISME J* **10**: 1555–1567.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., *et al*. (2007) ClustalW and ClustalX version 2. *Bioinformatics* **23**: 2947–2948.

Lavin, P., Gomez, P., Gonzale, B., and Osvald, U. (2008) Diversity of the marine picocyanobacteria *Prochlorococcus* and *Synechococcus* assessed by terminal restriction fragment length polymorphisms of 16S-23S rRNA internal transcribed spacer sequences. *Rev Chil Hist Nat* **81**: 515–531.

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.

Marston, M.F., Pierciey, F.J., Jr., Shepard, A., Gearin, G., Qi, J., Yandava, C., *et al*. (2012) Rapid diversification of coevolving marine *Synechococcus* and a virus. *Proc Natl Acad Sci U S A* **109**: 4544–4549.

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

Martiny, A.C., Tai, A.P.K., Veneziano, D., Primeau, F., and Chisholm, S.W. (2009) Taxonomic resolution, ecotypes and the biogeography of Prochlorococcus. *Environ Microbiol* **11**: 823–832.

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.

Mizoguchi, K., Morita, M., Fischer, C.R., Yoichi, M., Tanji, Y., and Unno, H. (2003) Coevolution of bacteriophage PP01 and *Escherichia coli* O157:H7 in continuous culture. *Appl Environ Microbiol* **69**: 170–176.

Needham, D.M., Sachdeva, R., and Fuhrman, J.A. (2017) Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J* **11**: 1614–1629.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., *et al*. (2017) Vegan: Community Ecology Package. R Package Version 2, 4-5. http:// CRAN. R-project.org/package=vegan.

Parada, A.E., Needham, D.M., and Fuhrman, J.A. (2016) Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* **18**: 1403–1414.

Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A.J., Thomson, N.R., *et al*. (2010) Antagonistic coevolution accelerates molecular evolution. *Nature* **464**: 275–278.

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 U S A* **110**: 12030–12035.

Peres-Neto, P.R., Legendre, P., Dray, S., and Borcard, D. (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **87**: 2614–2625.

Pittera, J., Humily, F., Thorel, M., Grulois, D., Garczarek, L., and Six, C. (2014) Connecting thermal physiology and latitudinal niche partitioning in marine Synechococcus. *ISME J* **8**: 1221–1236.

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.

Rodriguez-Brito, B., Li, L.L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., *et al*. (2010) Viral and microbial community dynamics in four aquatic environments. *ISME J* **4**: 739–751.

Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M. B., Loy, A., *et al*. (2016) Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**: 689–693.

Roy, S., and Chattopadhyay, J. (2007) The stability of ecosystems: a brief overview of the paradox of enrichment. *J Biosci* **32**: 421–428.

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

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., *et al*. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498–2504.

Sohm, J.A., Ahlgren, N.A., Thomson, Z.J., Williams, C., Moffett, J.W., Saito, M.A., *et al*. (2015) Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* **10**: 333–345.

Sullivan, M.B., Waterbury, J.B., and Chisholm, S.W. (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**: 1047–1051.

Tai, V., Burton, R., and Palenik, B. (2011) Temporal and spatial distributions of marine *Synechococcus* in the Southern California bight assessed by hybridization to bead-arrays. *Mar Ecol-Prog Ser* **426**: 133–147.

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

Thingstad, T.F. (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol Oceanogr* **45**: 1320–1328.

Thingstad, T.F., and Lignell, R. (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat Microb Ecol* **13**: 19–27.

Tikhonov, M., Leach, R.W., and Wingreen, N.S. (2015) Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. *ISME J* **9**: 68–80.

Ward, C.S., Yung, C.M., Davis, K.M., Blinebry, S.K., Williams, T.C., Johnson, Z.I., and Hunt, D.E. (2017) Annual community patterns are driven by seasonal switching between closely related marine bacteria. *ISME J* **11**: 2637.

Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I., *et al*. (2008) The sorcerer II Global Ocean sampling expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS One* **3**: e1456.

Winter, C., Bouvier, T., Weinbauer, M.G., and Thingstad, T. F. (2010) Trade-offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited. *Microbiol Mol Biol Rev* **74**: 42–57.

Xia, L.C., Steele, J.A., Cram, J.A., Cardon, Z.G., Simmons, S. L., Vallino, J.J., *et al*. (2011) Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. *BMC Syst Biol* **5**: S15.

Xue, C., and Goldenfeld, N. (2017) Coevolution maintains diversity in the stochastic "kill the winner" model. *Phys Rev Lett* **119**: 268101.

Zinser, E.R., Johnson, Z.I., Martiny, A.C., Fuller, N.J., Scanlan, D.J., and Chisholm, S.W. (2006) *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Appl Environ Microbiol* **72**: 723–732.

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:

Appendix S1: Supporting Information