Diversity and temporal dynamics of Southern California coastal marine cyanophage isolates

Jessica L. Clasen[1,2§], China A. Hanson[1§*], Yazeed Ibrahim[1], Claudia Weihe[1], Marcia F. Marston[3], and Jennifer B. H. Martiny[1]

[1] Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92627

[2] Current address: Department of Zoology, University of British Columbia, Vancouver, B.C., Canada V6T 1Z4

[3] Department of Biology and Marine Biology, Roger Williams University, Bristol, RI 02809

§Authors contributed equally to this work.

*Corresponding author: cahanson@uci.edu

Running head: Diversity of California cyanophage isolates

Abstract

Marine ecosystems contain an immense diversity of phages, many of which infect the cyanobacteria responsible for a portion of oceanic primary productivity. To add to the growing body of research on the dynamics and diversity of these cyanophages, we measured cyanophage abundance, diversity, and community composition monthly for 15 months at three coastal locations in Southern California. We characterized over 900 individual cyanophages, isolated on *Synechococcus* sp. WH7803, by amplifying and sequencing two different genes from each isolate: *g20* (the portal protein gene) and *psbA* (a host derived gene involved in photosynthesis). We found that the taxonomic composition and diversity of isolated cyanophages were highly variable over time, with little variation across the three locations. Particular taxa dominated the community at distinct times of year, with more rapid turnover of dominant types and higher overall abundance and diversity during spring and summer. These temporal patterns as well as taxonomic designations were highly congruent for both genes based on operational taxonomic units (OTUs) defined at 99% nucleotide similarity, despite a greater level of amino acid conservation within *psbA* as compared to *g20*. Overall, this suggests that either gene can serve as a useful marker for cyanophage diversity. However, we detected one myovirus taxon (representing about 2.5% of the isolates) of uncertain phylogenetic affiliation that was negative for *psbA* amplification. Further sampling is necessary to assess whether these temporal dynamics in cyanophage composition, diversity, and abundance are seasonally recurring.

Key words: cyanophage, *g20*, *psbA*, myovirus, *Synechococcus*, temporal, diversity

Introduction

Viruses are the most abundant and diverse biological entities in the Earth's oceans (Suttle 2007). Recent work suggests that thousands of viral genotypes coexist in several hundred liters of seawater (Breitbart & Rohwer 2005) and that hundreds of thousands of marine viral "species" persist globally, the majority of which infect microorganisms (Angly et al. 2006). Further, much of the viral genetic diversity observed in recent metagenomic libraries is entirely novel, with 65 - 95% of sequences unlike any others in genomic databases (Breitbart et al. 2004, Angly et al. 2006, Culley et al. 2006).

Some understanding, however, of marine viral diversity has been garnered from studies of marine cyanophages - the viruses that infect the cyanobacteria genus, *Synechococcus.* *Synechoccocus* spp. are abundant in coastal marine surface waters, and are thought to be responsible for up to 30% of marine primary productivity (Waterbury et al. 1986, Partensky et al. 1999). Cyanophages are also abundant in coastal waters, with estimates up to 10^5 cyanophage particles mL^{-1} (Waterbury & Valois 1993, Wilson et al. 1993, Suttle & Chan 1994). At any one time, one to five percent of cyanobacterial cells in the ocean are estimated to be infected by phage (Proctor & Fuhrman 1990, Suttle & Chan 1994, Garza & Suttle 1998). Through infection and lysis of host cells, cyanophages influence cyanobacteria turnover rates and marine nutrient cycling (Suttle et al. 1990, Thingstad 1998, Fuhrman 1999, Wilhelm & Suttle 1999) and potentially host abundance and diversity as hypothesized for marine phages in general (Thingstad 2000, Middelboe et al. 2001, Schwalbach et al. 2004, Weinbauer & Rassoulzadegan 2004).

Studies over the last two decades have demonstrated that marine cyanophage diversity is both spatially and temporally dynamic. Cyanophage diversity and community composition often varies by depth, having highest diversity in the water column at depths corresponding to the deep chlorophyll *a* maximum and/or maximum abundances of *Synechococcus* or *Prochlorococcus* cells (Wilson et al. 1999, Zhong et al. 2002, Frederickson et al. 2003). Patterns over horizontal space are less consistent, however. Some studies suggest that cyanophage composition exhibits little horizontal spatial variation, given that nearly identical cyanophage gene sequences have been identified in surface marine samples collected from distant locations (Zhong et al. 2002, Breitbart et al. 2004, Short & Suttle 2005, Chénard & Suttle 2008, Huang et al. 2010). Other studies, however, have identified compositional variation among locations. In these cases, the patterns appear to be related to salinity, whereby freshwaters harbor cyanophage taxa not found in marine habitats (Short & Suttle 2005, Chénard & Suttle 2008, Sullivan et al. 2008, Wang et al. 2010). Indeed, we know of only one study that found evidence for genetically distinct cyanophage communities across relatively small (<100km) scales within a marine system (Frederickson et al. 2003).

In contrast, temporal variation in marine cyanophage communities tends to be a more consistent observation. For example, cyanophage abundance and diversity varies seasonally, and is typically higher in spring or summer months (Marston & Sallee 2003, Dorigo et al. 2004, Wang & Chen 2004, Mühling et al. 2005, Sandaa & Larsen 2006, Chen et al. 2009, Chow & Fuhrman 2012, Parsons et al. 2012). Cyanophage composition is also temporally variable, particularly within years. Such intraannual variation in composition has been observed on the Atlantic coast of the United States (Marston & Sallee 2003, Wang & Chen 2004), along the coast of Norway (Sandaa & Larsen 2006), and in the Red Sea (Mühling et al. 2005). However, direct

comparison of variation in time across multiple locations is necessary to understand whether this temporal variation is predictable for cyanophage communities across diverse oceanic regions.

Here, we examined the diversity and community composition of marine cyanophages over both time and space. We used an isolation-based approach to characterize cyanophages that infect the marine cyanobacterium *Synechococcus* sp. strain WH7803. Although this methodology greatly undersamples the total cyanophage community, it provides several advantages that complement culture-independent studies. First, because the viruses replicate on the same host strain, they likely share some of the same hosts in the natural community. Thus, the approach targets a viable, and potentially ecologically meaningful, subset of the cyanophage diversity present. Second, this reduced subset of all cyanophages provides a tractable amount of viral diversity for making robust statistical comparisons across samples. Last, an isolation-based approach allows each virus to be studied individually, and multiple regions of the genome can be compared across individuals (Marston & Amrich 2009).

For this study, we isolated and purified more than 900 cyanophages from three coastal locations in Southern California over 15 months. We genetically characterized each isolate by PCR-amplifying and sequencing two cyanophage marker genes that differ in evolutionary history, *g20* and *psbA*. Both genes are present in the genomes of most cyanophages of the *Myoviridae* family (cyanomyoviruses) (Sullivan et al. 2006, Sullivan et al. 2008), which constitute the majority of cyanophages both isolated from seawater (Waterbury & Valois 1993, Wilson et al. 1993, Marston & Sallee 2003) and detected in marine metagenomic libraries (Breitbart et al. 2002, Williamson et al. 2008). The *g20* gene, involved in capsid assembly, is conserved within the *Myoviridae* and thus commonly used as a phylogenetic marker (Fuller et al. 1998). *PsbA* is a functionally-conserved host-derived gene encoding the D1 protein of

photosystem II (Mann 2003, Lindell et al. 2004). With these data, we asked two questions: Does cyanophage diversity and composition vary temporally or spatially across three coastal sampling sites located in Southern California? And, how does spatial and temporal variation in diversity and composition compare across the two genes?

Materials and Methods

Sampling

Three replicate surface seawater samples (60mL – 1L) were collected once per month from January 2008 to March 2009 from the surf zone of three Southern California coastal sites - Seal Beach (SB; 33° 44'N, 118° 6'W), Balboa Island (BI; 33° 36'N, 117° 53'W), and Crystal Cove State Beach (CC; 33° 34'N, 117° 50'W). Approximate pairwise distances between sites are as follows: CC-BI: 8 km; CC-SB: 32 km; BI-SB: 23 km. For each sampling event, samples were collected between 7:00 AM and 11:00 AM at each site, transported in the dark, and processed within three hours of collection. Weekly average sea surface temperature measurements at 2m depth were recorded by the Southern California Coastal Ocean Observing System (SCCOOS) at the Newport Pier (33° 35.94'N, 117°54.00'W), 2 km northwest of the BI site, using a mounted SEACAT instrument.

Estimation of Cyanophage Abundance

The abundance of cyanophages capable of infecting *Synechococcus* sp. WH7803 at the Balboa Island site was estimated using a most probable number (MPN) assay, modified from Marston and Sallee (2003). Briefly, a subsample (60ml) of each replicate seawater sample was centrifuged to remove large particles and bacteria, and the supernatant was serially diluted with

sterile natural seawater media (SN media (Waterbury et al. 1986)). A 100µl aliquot of this diluted seawater was added to 100µl of exponentially growing *Synechococcus* sp. WH7803 in each well of 48-well microtiter plates (except control wells which received no seawater). After a 30 - 60 min incubation at room temperature, 500 ul of sterile SN media was added to each well. Plates were then incubated for two weeks at 25°C on a 14:10 hr day:night cycle at 19µE m^{-2} s^{-1} light intensity. After 14 days, the plates were visually monitored for lysis (wells with less visible pigmentation than control host-only wells), and the total number of lysed wells was recorded. Estimates of the concentration of infectious cyanophages were determined using the MPN Calculator Build6 freeware (Curiale 2004). While we set up similar plates using diluted samples (referred to as "dilution plates") from each of the other two sites each month (Crystal Cove and Seal Beach), we did not include replicate plates for these sites during January – October 2008. For this reason, MPN abundances were estimable for the Balboa Island site only for the majority of the sampling period; therefore, only this site is included in the MPN abundance data presented in Fig. 3A.

Cyanophage Isolation

Cyanophages were enriched using the same MPN plates (or dilution plates for Crystal Cove and Seal Bach) as described above. After plates incubated for three weeks, the liquid in lysed wells (lysate) was collected. To minimize the selection of only fast-growing cyanophages, we collected lysates from plates with less than 50% of the wells cleared. We aimed to collect 50 lysates from each site, per month, although this was not always possible due to low MPN titers during some months. The lysates were stored in cryogenic vials in the dark at 4 °C for up to 12 months prior to purification by plaque assay following Marston and Sallee (2003). To purify,

each cyanophage lysate was serially diluted and combined with concentrated *Synechococcus* sp. WH7803 culture and warm (37°C) sterile SN media (Waterbury et al. 1986) containing 0.3% washed agar (Waterbury & Willey 1988). This mixture was poured over a 0.6% agar SN plate and incubated at 25°C and a 14:10 day:night cycle. Plates containing host cells but no viral lysate were also included as controls. Plaques were never observed on these control plates. Upon the appearance of plaques (typically 4-7 days), a single plaque was picked from each plate and regrown on a WH7803 liquid culture in microtiter plates. Each plaque-purified lysate was then treated with chloroform (final volume 10%) and centrifuged (2000xg for 10 minutes) to remove remaining host cells and debris. The supernatant containing a purified cyanophage was transferred to a cryogenic vial for storage in the dark at 4 °C.

Genetic analysis

Each plaque purified isolate was used directly as template DNA in PCR reactions targeted at two separate genes, *g20* and *psbA*, both of which are found in the genomes of most, if not all, cyanophages in the *Myoviridae* family (Sullivan et al. 2006, Sullivan et al. 2008, Sullivan et al. 2010). The *g20* gene was amplified using the primers CPS1.1F (5'-GTAGWATWTTYTAYATTGAYGTWGG-3') and CPS8.1R (5'-ARTAYTTDCCDAYRWAWGGWTC-3') (Sullivan et al. 2008). Each PCR reaction (30 µL) contained 1x of PCR buffer, 1.5mM of $MgCl_2$, 0.2 µM of each primer, 250 µM of dNTPs, and 1U of 5 Prime HotMaster Taq polymerase (5 Prime, Gaithersburg, MD, USA), and 1.5 µL of plaque purified cyanophage isolate as template. The reaction parameters were: a denaturing step at 94 ºC for 3 minutes and then 34 cycles of 94 ºC for 45 seconds, annealing at 48.2 ºC for 45 seconds and extension at 72 ºC for 1 minute, followed by a final extension step at 72 ºC for 4

minutes. The *psbA* gene was amplified using the primers *psbA*-93F (5'-TAYCCNATYTGGGAAGC-3') and *psbA*-341R (5'-TCRAGDGGGAARTTRTG-3') (Wang & Chen 2008). Each 30µL reaction contained 1x PCR buffer, 1.5 mM $MgCl_2$, 250 µM dNTPs, 0.5 µM of each of the primers, 1U of Platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, CA, USA), and 1.5 µL of plaque purified cyanophage isolate as template. The reaction parameters were 94^{o}C for 3 minutes, followed by 36 cycles of 94^{o}C for 1 minute, 55^{o}C for 30s and 72^{o}C for 1 min, and then a final extension step at 72^{o}C for 10 minutes. The *g20* (~540 bp) and *psbA* (~600 bp) products were sent to Agencourt (Beverly, MA, USA) for Sanger sequencing using the *g20* forward primer and the *psbA* reverse primer. Note that the number of quality sequences obtained varied both by month and by locus *(g20*: ranging from 10 sequences in January 2008 and April 2008 to 130 in September 2008; *psbA*; ranging from 6 in May 2008 to 101 in September 2008).

The partial gene sequences were edited and aligned using MUSCLE in Geneious 4.6.2 (Auckland, NZ). After trimming and editing, the resulting sequence fragments were ~510 bp and ~575 bp in length for *g20* and *psbA*, respectively. DNA distance matrixes were constructed in PHYLIP v.3.68 (Felsentein 1993) using the F84 distance model. Operational Taxonomic Units (OTUs) based upon a 99% sequence similarity threshold were designated using the nearest-neighbor cluster algorithm in MOTHUR v.1.14.0 (Schloss et al. 2009). This cutoff threshold was chosen in order to be consistent with other work on cyanophage isolates (e.g., Marston & Amrich 2009, Marston et al. 2012). One representative sequence for each OTU was selected using the get.otureps function in MOTHUR and used to build Neighbor-Joining phylogenetic trees in Mega5 for Mac (Tamura et al. 2007) with bootstrap support calculated from 10 000 replications. OTU representatives were also used for Genbank Blast searches using blastn and blastx batch search functions as well as for MEGA Blast searches of the Global Ocean Survey (GOS)

metagenomic database through the CAMERA portal (Sun et al. 2011). The partial *g20* and *psbA* gene sequences of OTU representatives have been deposited in Genbank as JX218109 to JX218152, and are listed in Supplementary Tables S1-S3.

Diversity metrics and statistical analyses

For each gene separately, OTU diversity metrics (Chao1 richness estimator, Coleman richness, Shannon's, and Simpson's indices) were computed using Estimate S (Colwell 2009). Rarefaction curves were created by using Coleman richness estimates from EstimateS. One-way analysis of variance (ANOVA) tests were carried out in JMP (version 5.1.2, SAS Institute Inc. Cary, North Carolina, USA) to test for differences in mean Chao1 richness, Coleman richness, Shannon's Index, and Simpson's Index among sampling sites and months. Tukey-Kramer post-hoc comparison tests were used to determine which samples drove the overall differences in richness by month. For ANOVA tests, we used richness and diversity values rarefied to the lowest common sample size (n=12), and samples consisting of fewer than 12 sequences were eliminated from that analysis. To test for differences in OTU composition among sampling sites and months, we performed Analysis of Similarity (ANOSIM, two-way with no replication) using Primer6 (Ivybridge, UK). Prior to the ANOSIM tests, samples consisting of fewer than 14 sequences were removed, and similarity matrices were calculated from OTU relative abundance using a non-weighted Bray-Curtis similarity metric.

Results

g20 sequence diversity

Over 95% of the cyanophage isolates yielded positive amplicons with myovirus-specific *g20* PCR primers, suggesting that nearly all of the isolates are cyanomyoviruses. In total, high quality *g20* sequences were obtained from 916 cyanophage isolates, which grouped into 16 operational taxonomic units (OTUs) at a 99% sequence similarity threshold. These *g20* OTUs, or taxa, are hereafter designated as S-CAMs (*Synechococcus*-infecting California Myovirus). A majority (621; 61%) of the *g20* sequences grouped into a single OTU, S-CAM1 (Figure 1), while 8 of the OTUs were comprised of only a single sequence. Although our culture-based method can only capture a limited amount of cyanophage diversity (Table 1), a rarefaction curve indicates that we probably observed most of the common taxa (identified by the *g20* gene) capable of being isolated by our culturing method (Figure S1).

The *g20* OTUs were phylogenetically diverse, representing all three of the previously defined major phylogenetic clusters containing cultured representatives (Zhong et al. 2002) (Figure 1). However, the majority of *g20* sequences (~94%), including the top three most abundant OTUs (S-CAM1, 3, and 4), fell into Cluster I (*sensu* (Zhong et al. 2002, Sullivan et al. 2008). Six additional OTUs (11 sequences total) fell within Cluster II, while another six OTUs (22 sequences total) associated with Cluster III. S-CAM7, represented by 20 sequences, fell onto a long branch with low bootstrap support that did not clearly group within any of the previously defined major clusters. As expected, none of the sequences fell within the A-F environmental sequence-only clusters, which to-date contain no known cultured representatives (Zhong et al. 2002, Sullivan et al. 2008).

Although the S-CAMs were closely related to previously observed cyanomyoviruses (Figure 1), many taxa appeared to be novel. The *g20* sequences of S-CAM strains ranged from 72.3 – 96.5% in nucleotide identity (blastn) with sequences deposited in Genbank (Table S1); 74

- 100% in amino acid identity (blastx) with sequences deposited in Genbank (Supplementary Table S2); and up to 96% in nucleotide identity with sequences in the GOS metagenomic dataset (MEGAblast). The most abundant taxon, S-CAM1, was just 87% identical in nucleotide sequence and 98% identical in amino acid sequence to the top Genbank blast hit, cyanophage clone SAI99-44 deposited by Short and Suttle (2005), and 96% identical to a GOS sequence from the coast of South Carolina. Further, only one OTU, S-CAM4, was 100% identical in amino acids to a previously observed sequence – that of S-RIM29, isolated from Mt. Hope Bay, Rhode Island (Marston & Sallee 2003). However, at the nucleotide level, S-CAM4 was only 91% identical to S-RIM29 (44 nucleotide differences). S-CAM14, a rare OTU in our samples, was also highly similar in amino acid sequence to a deposited sequence, differing by a single amino acid from "*Synechoccocus* phage 44A" isolated by Zhong et al. (2002). Conversely, S-CAM7 was the most dissimilar from sequences in the databases, sharing only 72% and 74% in nucleotide and amino acid identity to the top hits, respectively.

psbA sequence diversity

We successfully amplified *psbA* from 95% of the cyanophage isolates using primers developed by Wang and Chen (2008). High quality *psbA* sequences were obtained from 790 isolates, which formed 28 OTUs at a 99% sequence similarity threshold. While 517 of these sequences fell into a single OTU (OTU-1), 20 of the OTUs were composed of only one sequence (Table 1, Figure 2). Notably, *psbA* could not be amplified from any of the 14 cyanophages isolated in April 2008 or from 33% of the cyanophages isolated in May 2008. In the majority of cases, these *psbA*-negative isolates from April and May belonged to S-CAM7 by analysis of the

g20 gene; otherwise, they were either non-amplifiable for *g20* (11%), or did not produce a good quality *g20* sequence (3%).

The *psbA* OTUs were well distributed across a single phylogenetic cluster, the Mar1 group (as previously described by Chénard and Suttle (2008)), and fell into two sub-clusters with moderate bootstrap support (Figure 2A). However, none of the *psbA* OTUs were identical in nucleotide sequence to those previously deposited in Genbank (Supplementary Table 3) or in the GOS metagenomic dataset. The *psbA* OTUs ranged 85.4 – 94.8% in nucleotide identity (blastn) with sequences in Genbank and up to 94% in nucleotide identity with sequences in GOS. The highest match was between the rare OTU7, and the cultured phage S-RSM2 at 94.8% similarity. The most abundant OTU (OTU-1) had a *psbA* sequence that was only 87% identical in nucleotide sequence to "uncultured cyanophage clone FRP962" sampled from the waters of the Northeastern Pacific by Chénard and Suttle (2008); 86.8% identical to cultured cyanophage Syn19; and 85% identical to a GOS sequence from the Gulf of Maine.

Despite low identity at the nucleotide level, all of the *psbA* OTUs were 100% similar in *psbA* amino acid sequence (blastx) with other cultured and environmental sequences deposited in Genbank. Moreover, there was very little amino acid variation among our *psbA* OTUs, as the majority of isolates were 99.95-100% similar to each other in *psbA* amino acid sequence (Figure 2B). Most of the *psbA* OTUs fell into two groups, which differed by a single amino acid. While we did observe slight variation in the $^{R}/_{K}$ETTXXXS$^{Q}/_{H}$ variable triplet motif identified by Sharon et al. (2007) among our *psbA* OTUs, variation at this site did not account for the divergence of the two major clusters. Instead a region ~80 residues upstream contained the amino acid substitution responsible for the one amino acid difference between the two main clades. Only 4 rare OTUs grouped outside of these two clades, with the largest amino acid divergence between

any pair being 6 amino acids, or 3% (e.g. OTU-7 compared to OTU-3). In only a few of these cases did variation in the $^{R}/_{K}$ETTXXXS$^{Q}/_{H}$ motif explain the amino acid divergence (OTU-7 and -24 having the sequence ENE, while all other OTUs having ETE at the variable triplet site). These two triplet sequences are equivalent to the two most abundant motifs that were found in *psbA* sequences from the GOS dataset (Sharon et al. 2007), suggesting that our isolation method detected the globally most common *psbA* types.

Comparison of diversity across the two genes

Overall, the *psbA* gene fragment resulted in greater observed taxonomic richness than the *g20* gene (28 versus 16 OTUs; Table 1), which is also reflected by the steeper rarefaction curve for *psbA* (Figure S1). Despite this difference in taxonomic richness elicited by the two loci, the most abundant *psbA* OTUs corresponded well to the most abundant *g20* OTUs – S-CAM1 corresponds to *psbA* OTU-1; S-CAM4 to *psbA* OTU-2; S-CAM3 to *psbA* OTU-3; and S-CAM8 to *psbA* OTU-4, indicating taxonomic congruence of the genes at the nucleotide level. The most notable difference between *psbA* and *g20* OTUs was the lack of a *psbA* OTU corresponding to S-CAM7, as the isolates in this OTU were negative for *psbA* amplification. Any other discrepancies between *g20* and *psbA* OTUs can be attributed to differing sample sizes (number of quality sequences obtained each month for each gene) or to the greater proportion of rare OTUs for *psbA* (i.e. more *psbA* singletons at the 99% cutoff), possibly due to high levels of intragenic recombination reported for this gene (Marston & Amrich 2009).

However, at the amino acid level, the *psbA* gene was much more conserved than the *g20* gene. A phylogenetic tree based on *g20* amino acid alignments (Figure S2) exhibited a similar topology to the *g20* nucleotide tree and only one pair of *g20* OTUs were 100% identical to each

other in amino acid sequence (S-CAM8 and S-CAM19). By contrast, the majority of *psbA* OTUs, including the most common ones, were 100% identical to each other in amino acid sequence (Figure 2B).

Temporal and spatial variation in cyanophage isolates

The abundance of cyanophages infectious to *Synechococcus* sp. WH7803 (recorded at the Balboa Island site only; see Materials and Methods for a description) varied over time. Specifically, we observed highest abundances in July and August 2008, and lowest abundances in winter, spring, and late fall months, except for a peak in abundance in February 2008 (Figure 3A). Average weekly sea surface temperatures recorded at the nearby Newport Pier exhibited a similar trend: warmest temperatures in summer months and coolest in winter months, with an overall range of 13.5 °C to 21.8 °C (Figure 3A).

The taxonomic composition of cyanophage isolates also exhibited striking temporal variation over the fifteen months of sampling, but little spatial variation across the three southern California sites. There were distinct temporal fluctuations in the relative abundances of the most common OTUs, and these fluctuations were highly similar for both genes (Figure 3B, C). Moreover, overall OTU composition based on both *g20* and *psbA* differed significantly across sampling months (ANOSIM: $\boldsymbol{\rho} = 0.913$, $p = 0.001$ and $\boldsymbol{\rho} = 0.838$, $p = 0.001$, for *g20* and *psbA* respectively). Composition fluctuated most dramatically over the spring and summer months (March - August 2008), during which the identity of the dominant OTU changed frequently. Conversely, composition was less dynamic during the fall and winter (January - March 2008 and September 2008 - March 2009) when a single OTU, *g20* S-CAM1 and *psbA* OTU-1, dominated the community. During some of these months, 100% of the isolated cyanophages belonged to

this single dominant OTU. In contrast to this variation over time, composition did not vary over space among the three sampling sites (ANOSIM: $\boldsymbol{\rho}$ = 0.1, p = 0.24 and $\boldsymbol{\rho}$ = -0.095, p = 0.76, respectively for *g20* and *psbA*).

Like the patterns observed for taxonomic composition, richness and diversity of the cyanophage isolates also varied more dramatically over time than space. *g20* OTU richness and diversity measures (Coleman, Shannon and Simpson) differed by sampling month (ANOVA: $p < 0.0001$ for all metrics, except Chao1: $p = 0.099$; e.g. Coleman richness F = 9.47, df = 8, $p < 0.0001$). The spring and summer months generally harbored a more diverse cyanophage community than the late fall and winter months (Tukey-Kramer HSD post-hoc comparisons; Table 1, Figure 3B and C). Although the Balboa Island site tended to have higher richness and diversity values (Table 1), this trend was not significant for any of the four metrics (e.g., Coleman richness F = 0.59, df = 2, p = 0.5).

Discussion

Cyanophage communities infecting *Synechococcus* sp. WH7803 exhibited striking temporal variation over a 15 month time period, but did not vary spatially among three coastal sites in southern California. The lack of spatial differentiation is not surprising given that ocean currents and tidal flux are likely responsible for maintaining well-mixed communities along this coastal region at a relatively small scale (maximum distance apart is approximately 34 km). Similarly, within the Chesapeake Bay, at distances up to 300 km apart, cyanophage communities were much more variable over time than over space (Wang & Chen 2004).

The temporal variation in cyanophage abundance observed here (Figure 3) is similar to other studies in a variety of marine locations. In particular, the abundances of total viruses and

cyanophages tend to be greatest in warmer seasons and/or lowest in the winter (Waterbury & Valois 1993, Jiang & Paul 1994, Suttle & Chan 1994, Cochran & Paul 1998, Marston & Sallee 2003, Millard & Mann 2006, Sandaa & Larsen 2006, Wang et al. 2011, Parsons et al. 2012). Some of these studies also found that cyanophage abundances co-vary with cyanobacterial abundance and temperature (e.g. Waterbury and Valois 1993, Sandaa and Larsen 2006, Wang et al. 2011), resembling the trend we observed here in southern California (Figure 3A).

We also identified changes in cyanophage composition and diversity over time. In particular, the fall-winter community, dominated by a single taxon, was distinct from the more diverse spring-summer community. Similar temporal variation in coastal cyanophage composition has also been observed on the Atlantic coast of the United States within the Chesapeake Bay (Wang & Chen 2004) and in Rhode Island coastal waters (Marston & Sallee 2003, Marston et al. 2012), in the coastal waters of Norway (Sandaa & Larsen 2006), and in the Red Sea (Mühling et al. 2005). These studies also reported a temporal pattern in cyanophage diversity, often with lower richness in winter months than summer months. Taken together, these results suggest that seasonal variation in composition and diversity is a common pattern in coastal cyanophage communities.

Chow and Fuhrman (2012) recently documented strong seasonal variation in the composition of the total community of uncultured myoviruses (including myoviruses that infect non-cyanobacteria) in the San Pedro Channel off the coast of Southern California, near our sampling sites. Their findings are consistent with ours in that the fall/winter community was distinct from the spring/summer community, substantiating our result based on a more limited subset of viral diversity (culturable cyanomyoviruses infective to *Synechococcus* sp. WH7803). As expected, Chow and Furhman (2012) detected an overall more diverse myovirus community

across all seasons than detected by our method. However, we speculate that the cyanophage diversity targeted in our study represents an ecologically similar group of viruses since they likely infect a similar subset of all cyanobacteria. For instance, the dominance of a single cyanophage taxon in winter months observed in our study may indicate a corresponding dominance of a particular *Synechococcus* phenotype as measured by its susceptibility to phage infection. Indeed, Tai and Palenik (2009) demonstrated such a seasonal pattern in the distribution of several *Synechococcus* phylogenetic clades in Southern California coastal water.

Although our isolation method certainly limits the amount of cyanophage diversity that is detected, it has two main advantages. First, we are capable of targeting only viable infectious phage from the large fraction in the ocean that are likely not infective (Suttle & Chen 1992, Wommack et al. 1996) and second, the genome and physiology of the host strain is well-known. Thus, our isolation-based approach may provide a useful indicator of ecologically relevant phage diversity and potential phage-host interactions that are not possible with culture-independent methods. Nonetheless, work is underway to confirm our findings based on cultured isolates with data from culture-independent direct sequencing at our sites. We expect the directly sequenced samples to consist of a significant proportion of the diversity detected using the isolation-based approach, as has been recently reported using similar methods for samples collected in Atlantic coastal waters (Marston et al. 2012).

Interestingly, the majority of our cyanophage OTUs were at least 10% divergent in nucleotide and amino acid sequences from both uncultured and cultured cyanophage sequences present in Genbank (Tables S1-S3). For example, the most abundant taxon identified in our samples, S-CAM1 (61% of the isolates), was not identical in either *g20* or *psbA* nucleotide sequence to anything in the databases to date. Although many of the top sequence matches were

to other cultured cyanophages isolated on the same *Synechococcus* host from waters on the Atlantic coast (e.g. Marston & Sallee 2003), none of the matches were greater than 97% in nucleotide sequence to another cultured isolate (Tables S1 and S3), suggesting that at least some of our isolates represent novel cultured representatives. Additionally, S-CAM7, dominant for a short time during the spring, was highly divergent in sequence similarity from others in the databases and was unable to yield positive *psbA* amplification. Recent work suggests that *psbA* is a core gene common to all T4-like cyanomyoviruses (Sullivan et al. 2006, Sullivan et al. 2010); therefore, it is unclear whether this taxon truly lacks the gene or whether it harbors a divergent *psbA* sequence not amplified by our primers. Nonetheless, the detection of this divergent genotype suggests the existence of at least some previously undescribed cyanophage diversity in coastal Southern California.

We also found that the temporal patterns in cyanophage communities were consistent for two separate genetic loci, commonly used in cyanophage diversity studies: *g20*, the myovirus-specific portal protein gene; and *psbA*, a host-derived gene recently discovered to be expressed during infection (Lindell et al. 2005, Clokie et al. 2006), and thought to provide a reproductive advantage to phage by boosting host metabolism. In a recent comparative genomic study, Ignacio-Espinoza and Sullivan (2012) found that most cyanophage core genes (including *g20*) are vertically transferred, while a few (including *psbA*) exhibit evidence of recombination via lateral transmission. Thus, despite the fact that *psbA* and *g20* may have different evolutionary origins and histories, we found that taxonomic designations at 99% nucleotide similarity were highly congruent for both genes. This result agrees with other findings that cyanophage strain genotypes remain stable across multiple loci, including *g20* and *psbA* (Marston & Amrich 2009), and that recombination within these cyanophage 'core' genes across genetically different strains

(or OTUs) appears rare (Marston & Amrich 2009, Ignacio-Espinoza & Sullivan 2012). Thus, it appears that either locus can serve as a useful and comparable marker of cyanophage taxonomic composition.

While the *psbA* sequences of our cyanophage isolates were for the most part taxonomically congruent to *g20* at the nucleotide level, the corresponding amino acid sequences were not. These findings are in agreement with other studies suggesting that amino acid conservation at the *psbA* locus is due to high purifying selection both within cyanobacterial host genomes and cyanophage genomes (Lindell et al. 2004, Zeidner et al. 2005, Sullivan et al. 2006, Sharon et al. 2007, Marston & Amrich 2009), despite greater overall rates of molecular evolution for cyanophages (Zeidner et al. 2005). This pattern indicates the importance of carrying a functional *psbA* gene for cyanophage fitness in natural environments. Indeed, the protein is highly conserved across all photosynthetic domains (at least ~80% amino acid similarity between cyanobacteria and chloroplasts) (Janssen et al. 1989). This is thought to be due to high selection to maintain the specificity of the D1 protein's multiple binding sites, which are associated with its role as both a transmembrane protein and an electron acceptor (Wu et al. 1987, Janssen et al. 1989). While nucleotide variation in cyanophage *psbA* may be a useful indicator of diversity, this variation consists of primarily synonymous mutations due to the conserved function of the protein in the host. Since our isolates are likely all capable of infecting a similar subset of *Synechococcus* cells, then purifying selection to maintain functioning of a similar D1 protein within these hosts might explain the lack of *psbA* amino acid divergence among our isolates.

The temporal dynamics of cyanophage abundance and composition observed in this study corresponds well with the growing body of recent studies demonstrating that marine microbes are generally highly dynamic over time and often seasonally patterned (reviewed by Giovannoni

and Vergin (2012). More work is needed to understand what factors may be driving this temporal variation in marine microbial communities, whether it is seasonally predictable (as in (Fuhrman et al. 2006, Gilbert et al. 2012) in all areas of the world's oceans, and which microbial taxa tend to drive the patterns. One of the major outcomes of such work thus far is an increasing appreciation of how crucial the timing of sampling is for comparative marine microbial diversity studies (Giovannoni & Vergin 2012). For instance, the timing of sampling may explain much of the variation in bacterial diversity observed across studies, as changes in the relative abundance (not presence-absence) is responsible for variation in bacterial composition over time (Caporaso et al. 2012). As a result, previous studies characterizing phage diversity at multiple locations at single points in time (e.g. (Zhong et al. 2002, Sullivan et al. 2008, Jameson et al. 2011) may miss ecologically relevant patterns. Thus, further studies that simultaneously investigate both temporal and spatial patterns are needed to elucidate the factors that influence the abundance and diversity of marine phages, and ultimately, their impact on their bacterial hosts.

Ackowledgements We thank members of the J. Martiny lab for assistance with environmental sampling, host culture maintenance, and cyanophage isolation. Funding for this research was provided by the Gordon and Betty Moore Foundation, the National Science Foundation (OCE-1005388 and OCE-1031783), and a NOAA NERRS Graduate Research Fellowship (Estuarine Reserves Division, Office of Ocean and Coastal Resource Management, National Oceanic and Atmospheric Administration) to C.A.H.

References

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, and Rohwer F (2006) The marine viromes of four oceanic regions. Plos Biology 4:2121-2131

Breitbart M, Miyake JH, and Rohwer F (2004) Global distribution of nearly identical phage-encoded DNA sequences. Fems Microbiology Letters 236:249-256

Breitbart M and Rohwer F (2005) Here a virus, there a virus, everywhere the same virus? Trends in Microbiology 13:278-284

Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, and Rohwer F (2002) Genomic analysis of uncultured marine viral communities. Proceedings of the National Academy of Sciences of the United States of America 99:14250-14255

Caporaso JG, Paszkiewicz K, Field D, Knight R, and Gilbert JA (2012) The Western English Channel contains a persistent microbial seed bank. The ISME journal 6:1089-1093

Chen F, Wang K, Huang SJ, Cai HY, Zhao ML, Jiao NZ, and Wommack KE (2009) Diverse and dynamic populations of cyanobacterial podoviruses in the Chesapeake Bay unveiled through DNA polymerase gene sequences. Environmental Microbiology 11:2884-2892

Chénard C and Suttle CA (2008) Phylogenetic diversity of sequences of cyanophage photosynthetic gene psbA in marine and freshwaters. Applied and Environmental Microbiology 74:5317-5324

Chow CE and Fuhrman JA (2012) Seasonality and monthly dynamics of marine myovirus communities. Environmental Microbiology doi: 10.1111/j.1462-2920.2012.02744.x

Clokie MR, Shan J, Bailey S, Jia Y, Krisch HM, West S, and Mann NH (2006) Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. Environmental Microbiology 8:827-835

Cochran PK and Paul JH (1998) Seasonal abundance of lysogenic bacteria in a subtropical estuary. Applied and Environmental Microbiology 64:2308-2312

Colwell RK (2009) Estimate S: Statistical estimation of species richness and shared species from samples. http://purl.oclc.org/estimates

Culley AI, Lang AS, and Suttle CA (2006) Metagenomic analysis of coastal RNA virus communities. Science 312:1795-1798

Curiale M (2004) http://www.i2workout.com/mcuriale/mpn/index.html.

Dorigo U, Jacquet S, and Humbert JF (2004) Cyanophage diversity, inferred from g20 gene analyses, in the largest natural lake in France, Lake Bourget. Applied and Environmental Microbiology 70:1017-1022

Felsentein J (1993) PHYLIP (Phylogeny Inference Package) Department of Genetics, University of Washington, Seattle

Frederickson CM, Short SM, and Suttle CA (2003) The physical environment affects cyanophage communities in British Columbia inlets. Microbial Ecology 46:348-357

Fuhrman JA (1999) Marine viruses and their biogeochemical and ecological effects. Nature 399:541-548

Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, and Naeem S (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. Proceedings of the National Academy of Sciences of the United States of America 103:13104-13109

Fuller NJ, Wilson WH, Joint IR, and Mann NH (1998) Occurrence of a sequence in marine cyanophages similar to that of T4 g20 and its application to PCR-based detection and quantification techniques. Applied and Environmental Microbiology 64:2051-2060

Garza DR and Suttle CA (1998) The effect of cyanophages on the mortality of *Synechococcus* spp. and selection for UV resistant viral communities. Microbial Ecology 36:281-292

Gilbert JA, Steele JA, Caporaso JG, Steinbrueck L, Reeder J, Temperton B, Huse S, McHardy AC, Knight R, Joint I, Somerfield P, Fuhrman JA, and Field D (2012) Defining seasonal marine microbial community dynamics. Isme Journal 6:298-308

Giovannoni SJ and Vergin KL (2012) Seasonality in Ocean Microbial Communities. Science 335:671-676

Huang SJ, Wilhelm SW, Jiao NZ, and Chen F (2010) Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through DNA polymerase gene sequences. ISME Journal 4:1243–1251

Ignacio-Espinoza JC and Sullivan MB (2012) Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. Environmental Microbiology doi: 10.1111/j.1462-2920.2012.02704.x

Jameson E, Mann NH, Joint I, Sambles C, and Muhling M (2011) The diversity of cyanomyovirus populations along a North-South Atlantic Ocean transect. Isme Journal 5:1713-1721

Janssen I, Jakowitsch J, Michalowski CB, Bohnert HJ, and Loffelhardt W (1989) Evolutionary Relationship of Psba Genes from Cyanobacteria, Cyanelles and Plastids. Current Genetics 15:335-340

Jiang SC and Paul JH (1994) Seasonal and Diel Abundance of Viruses and Occurrence of Lysogeny/Bacteriocinogeny in the Marine-Environment. Marine Ecology-Progress Series 104:163-172

Lindell D, Jaffe JD, Johnson ZI, Church GM, and Chisholm SW (2005) Photosynthesis genes in marine viruses yield proteins during host infection. Nature 438:86-89

Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, and Chisholm SW (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. Proceedings of the National Academy of Sciences of the United States of America 101:11013-11018

Mann NH (2003) Phages of the marine cyanobacterial picophytoplankton. Fems Microbiology Reviews 27:17-34

Marston MF and Amrich CG (2009) Recombination and microdiversity in coastal marine cyanophages. Environmental Microbiology 11:2893-2903

Marston MF and Sallee JL (2003) Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. Applied and Environmental Microbiology 69:4639-4647

Marston MF, Taylor S, Sme N, Parsons R, Noyes TJE, and Martiny JBH (2012) Marine cyanophages exhibit local and regional biogeography. Environmental Microbiology doi: 10.1111/1462-2920.12062

Middelboe M, Hagstrom A, Blackburn N, Sinn B, Fischer U, Borch NH, Pinhassi J, Simu K, and Lorenz MG (2001) Effects of bacteriophages on the population dynamics of four strains of pelagic marine bacteria. Microbial Ecology 42:395-406

Millard AD and Mann NH (2006) A temporal and spatial investigation of cyanophage abundance in the Gulf of Aqaba, Red Sea. Journal of the Marine Biological Association of the United Kingdom 86:507-515

Mühling M, Fuller NJ, Millard A, Somerfield PJ, Marie D, Wilson WH, Scanlan DJ, Post AF, Joint I, and Mann NH (2005) Genetic diversity of marine *Synechococcus* and co-occurring cyanophage communities: evidence for viral control of phytoplankton. Environmental Microbiology 7:499-508

Parsons RJ, Breitbart M, Lomas MW, and Carlson CA (2012) Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. The ISME journal 6:273-284

Partensky F, Blanchot J, and Vaulot D (1999) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: A review. Bulletin de l'Institut Oceanographique (Monaco) 0:457-475

Proctor LM and Fuhrman JA (1990) Viral Mortality of Marine-Bacteria and Cyanobacteria. Nature 343:60-62

Sandaa RA and Larsen A (2006) Seasonal variations in virus-host populations in Norwegian coastal waters: Focusing on the cyanophage community infecting marine *Synechococcus* spp. Applied and Environmental Microbiology 72:4610-4618

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, and Weber CF (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology 75:7537-7541

Schwalbach MS, Hewson I, and Fuhrman JA (2004) Viral effects on bacterial community composition in marine plankton microcosms. Aquatic Microbial Ecology 34:117-127

Sharon I, Tzahor S, Williamson S, Shmoish M, Man-Aharonovich D, Rusch DB, Yooseph S, Zeidner G, Golden SS, Mackey SR, Adir N, Weingart U, Horn D, Venter JC, Mandel-Gutfreund Y, and Béjà O (2007) Viral photosynthetic reaction center genes and transcripts in the marine environment. Isme Journal 1:492-501

Short CM and Suttle CA (2005) Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. Applied and Environmental Microbiology 71:480-486

Sullivan MB, Coleman ML, Quinlivan V, Rosenkrantz JE, DeFrancesco AS, Tan G, Fu R, Lee JA, Waterbury JB, Bielawski JP, and Chisholm SW (2008) Portal protein diversity and phage ecology. Environmental Microbiology 10:2810-2823

Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, DeFrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino D, Osburne MS, Henn MR, and Chisholm SW (2010) Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. Environmental Microbiology 12:3035-3056

Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, and Chisholm SW (2006) Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. Plos Biology 4:1344-1357

Sun SL, Chen J, Li WZ, Altintas I, Lin A, Peltier S, Stocks K, Allen EE, Ellisman M, Grethe J, and Wooley J (2011) Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. Nucleic Acids Research 39:D546-D551

Suttle CA (2007) Marine viruses - major players in the global ecosystem. Nature Reviews Microbiology 5:801-812

Suttle CA and Chan AM (1994) Dynamics and Distribution of Cyanophages and Their Effect on Marine *Synechococcus* Spp. Applied and Environmental Microbiology 60:3167-3174

Suttle CA, Chan AM, and Cottrell MT (1990) Infections of phytoplankton by viruses and reduction of primary production. Nature 347:467-469

Suttle CA and Chen F (1992) Mechanisms and Rates of Decay of Marine Viruses in Seawater. Applied and Environmental Microbiology 58:3721-3729

Tai V and Palenik B (2009) Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. Isme Journal 3:903-915

Tamura K, Dudley J, Nei M, and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599

Thingstad TF (1998) A theoretical approach to structuring mechanisms in the pelagic food web. Hydrobiologia 363:59-72

Thingstad TF (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnology and Oceanography 45:1320-1328

Wang GH, Murase J, Asakawa S, and Kimura M (2010) Unique viral capsid assembly protein gene (g20) of cyanophages in the floodwater of a Japanese paddy field. Biology and Fertility of Soils 46:93-102

Wang K and Chen F (2004) Genetic diversity and population dynamics of cyanophage communities in the Chesapeake Bay. Aquatic Microbial Ecology 34:105-116

Wang K and Chen F (2008) Prevalence of highly host-specific cyanophages in the estuarine environment. Environmental Microbiology 10:300-312

Wang K, Wommack KE, and Chen F (2011) Abundance and Distribution of *Synechococcus* spp. and Cyanophages in the Chesapeake Bay. Applied and Environmental Microbiology 77:7459-7468

Waterbury JB and Valois FW (1993) Resistance to Cooccurring Phages Enables Marine *Synechococcus* Communities to Coexist with Cyanophages Abundant in Seawater. Applied and Environmental Microbiology 59:3393-3399

Waterbury JB, Watson SW, Valois FW, and Franks DG (1986) Biological and Ecological Characterization of the Marine Unicellular Cyanobacterium *Synechococcus*. Canadian Bulletin of Fisheries and Aquatic Sciences:71-120

Waterbury JB and Willey JM (1988) Isolation and Growth of Marine Planktonic Cyanobacteria. Methods in Enzymology 167:100-105

Weinbauer MG and Rassoulzadegan F (2004) Are viruses driving microbial diversification and diversity? Environmental Microbiology 6:1-11

Wilhelm SW and Suttle CA (1999) Viruses and Nutrient Cycles in the Sea - Viruses play critical roles in the structure and function of aquatic food webs. Bioscience 49:781-788

Williamson SJ, Rusch DB, Yooseph S, Halpern AL, Heidelberg KB, Glass JI, Andrews-Pfannkoch C, Fadrosh D, Miller CS, Sutton G, Frazier M, and Venter JC (2008) The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. PloS one 3:e1456

Wilson WH, Fuller NJ, Joint IR, and Mann NH 1999. Analysis of Cyanophage Diversity in the Marine Environment Using Denaturing Gradient Gel Electrophoresis.*in* C. R. Bell, M. Brylinsky, and P. Johnson-Green, editors. Microbial Biosystems: New Frontiers, Proceedings of the 8th International Symposium on Microbial Ecology. Atlantic Canada Society for Microbial Ecology, Kentville N. S.

Wilson WH, Joint IR, Carr NG, and Mann NH (1993) Isolation and Molecular Characterization of 5 Marine Cyanophages Propagated on *Synechococcus* Sp Strain Wh7803. Applied and Environmental Microbiology 59:3736-3743

Wommack KE, Hill RT, Muller TA, and Colwell RR (1996) Effects of sunlight on bacteriophage viability and structure. Applied and Environmental Microbiology 62:1336-1341

Wu NH, Cote JC, and Wu R (1987) Structure of the Chloroplast Psba Gene Encoding the Qb Protein from Oryza-Sativa-L. Developmental Genetics 8:339-350

Zeidner G, Bielawski JP, Shmoish M, Scanlan DJ, Sabehi G, and Béjà O (2005) Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. Environmental Microbiology 7:1505-1513

Zhong Y, Chen F, Wilhelm SW, Poorvin L, and Hodson RE (2002) Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene g20. Applied and Environmental Microbiology 68:1576-1584

Table Captions

Table 1. The number of sequences, OTUs (operational taxonomic units), singletons, mean Chao1 estimate (CIs), and the mean (± S.D.) of the Coleman richness, Shannon's index, and Simpson's index observed for the total sequences and for *g20* sequences collected in each month and site. To facilitate comparisons across samples, the reported values for the Coleman richness, Shannon's index, and Simpson's index in the *g20* by month and by site categories are rarefied to the lowest number available within each category (10 for the month category, and to 280 for the site category).

Table 1. Summary statistics for taxonomic richness and diversity

		No. of Sequences	No. of OTUs	No. of Singletons	Chao1 Richness	Coleman Richness	Shannon's Index (H')	Simpson's Index (D)
g20 total		915	16	8	48 (21, 227)	16	1.0	2.0
psbA total		790	28	20	228 (63, 1170)	28	1.2	2.1
g20 by Month	Jan-08	11	3	1	3 (3, 3)	2.9±0.3	0.8±0.1	2.3±0.3
	Feb-08	19	3	0	3 (3, 3)	2.7±0.5	0.7±0.2	1.9±0.4
	Mar-08	47	7	4	15 (8, 72)	3.2±1.0	0.9±0.3	2.8±1.0
	Apr-08	10	1	0	1 (1, 1)	1.0±0.0	0.0±0.0	1.0±0.0
	May-08	16	4	2	5 (4, 17)	3.3±0.7	1.0±0.2	3.1±0.5
	Jun-08	97	5	2	7 (5, 27)	2.0±0.8	0.4±0.3	1.5±0.5
	Jul-08	66	5	1	5 (5, 9)	2.5±0.8	0.6±0.3	1.8±0.5
	Aug-08	110	6	1	6 (6, 11)	3.0±0.8	1.0±0.2	3.0±0.8
	Sep-08	130	6	2	7 (6, 17)	1.7±0.7	0.3±0.3	1.2±0.3
	Oct-08	112	2	0	2 (2, 2)	1.2±0.4	0.1±0.2	1.0±0.1
	Nov-08	89	2	1	2 (2, 2)	1.1±0.3	0.0±0.1	1.0±0.1
	Dec-08	107	1	0	1 (1, 1)	1.0±0.0	0.0±0.0	1.0±0.0
	Jan-09	36	2	0	2 (2, 2)	1.7±0.5	0.3±0.2	1.3±0.3
	Feb-09	53	1	0	1 (1, 1)	1.0±0.0	0.0±0.0	1.0±0.0
	Mar-09	12	1	0	1 (1, 1)	1.0±0.0	0.0±0.0	1.0±0.0
g20 by Site	BI	280	11	5	24 (13, 105)	11±0.0	1.1±0.0	2.1±0.0
	CC	354	8	3	11 (8, 33)	7.4±0.7	0.9±0.0	1.8±0.1
	SB	281	9	3	12 (9, 34)	9.0±0.1	1.1±0.0	2.2±0.0

Figure Legends

Figure 1. Phylogenetic relationships of the Southern California cyanomyovirus (S-CAM) OTUs based on *g20* nucleotide sequences and selected other *g20* reference sequences. The number of isolates observed from each S-CAM taxon are shown in parentheses. Clusters were assigned according to Sullivan et al. (2008). Bootstrap values are indicated at the nodes.

Figure 2. Phylogenetic relationships of the Southern California cyanophage *psbA* OTUs based on (A) nucleotide and (B) amino acid sequences. In (A) the number of isolates observed from each OTU and the Genbank accession number for the OTU representative are shown in parentheses. Phylogenetic "Mar1" group refers to the marine cyanobacteria-infecting myovirus clade according to Chénard and Suttle (2008). Bootstrap values are indicated at the nodes.

Figure 3. Temporal dynamics from Jan 2008 – Mar 2009 in (A) the abundance of cyanophages infectious to *Synechococcus* strain WH7803 (at the Balboa Island site) and average weekly surface seawater temperature (at Newport Beach Pier); and the percent abundance of each of the most common (B) S-CAM OTUs and (C) *psbA* OTUs. In B and C, data is presented as percentage of each OTU (out of the five or four most common OTUs, for S-CAM and *psbA* OTUs, respectively) within each monthly sample.

Figure 1

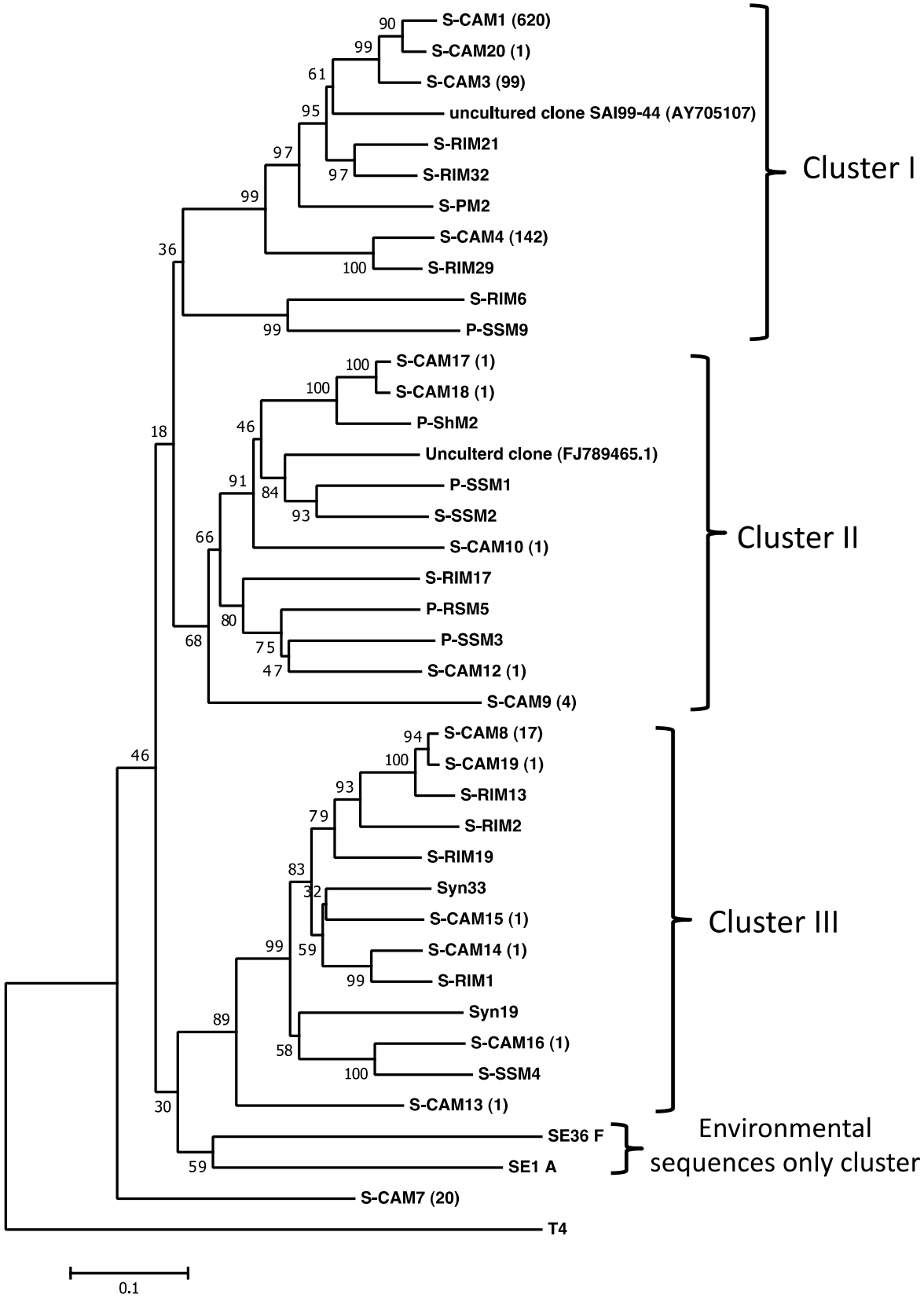

S-CAM1 (620)
S-CAM20 (1)
S-CAM3 (99)
uncultured clone SAI99-44 (AY705107)
S-RIM21
S-RIM32
S-PM2
S-CAM4 (142)
S-RIM29
S-RIM6
P-SSM9
Cluster I
S-CAM17 (1)
S-CAM18 (1)
P-ShM2
Unculterd clone (FJ789465.1)
P-SSM1
S-SSM2
S-CAM10 (1)
S-RIM17
P-RSM5
P-SSM3
S-CAM12 (1)
S-CAM9 (4)
Cluster II
S-CAM8 (17)
S-CAM19 (1)
S-RIM13
S-RIM2
S-RIM19
Syn33
S-CAM15 (1)
S-CAM14 (1)
S-RIM1
Syn19
S-CAM16 (1)
S-SSM4
S-CAM13 (1)
Cluster III
SE36 F
SE1 A
Environmental sequences only cluster
S-CAM7 (20)
T4
0.1

Figure 2A

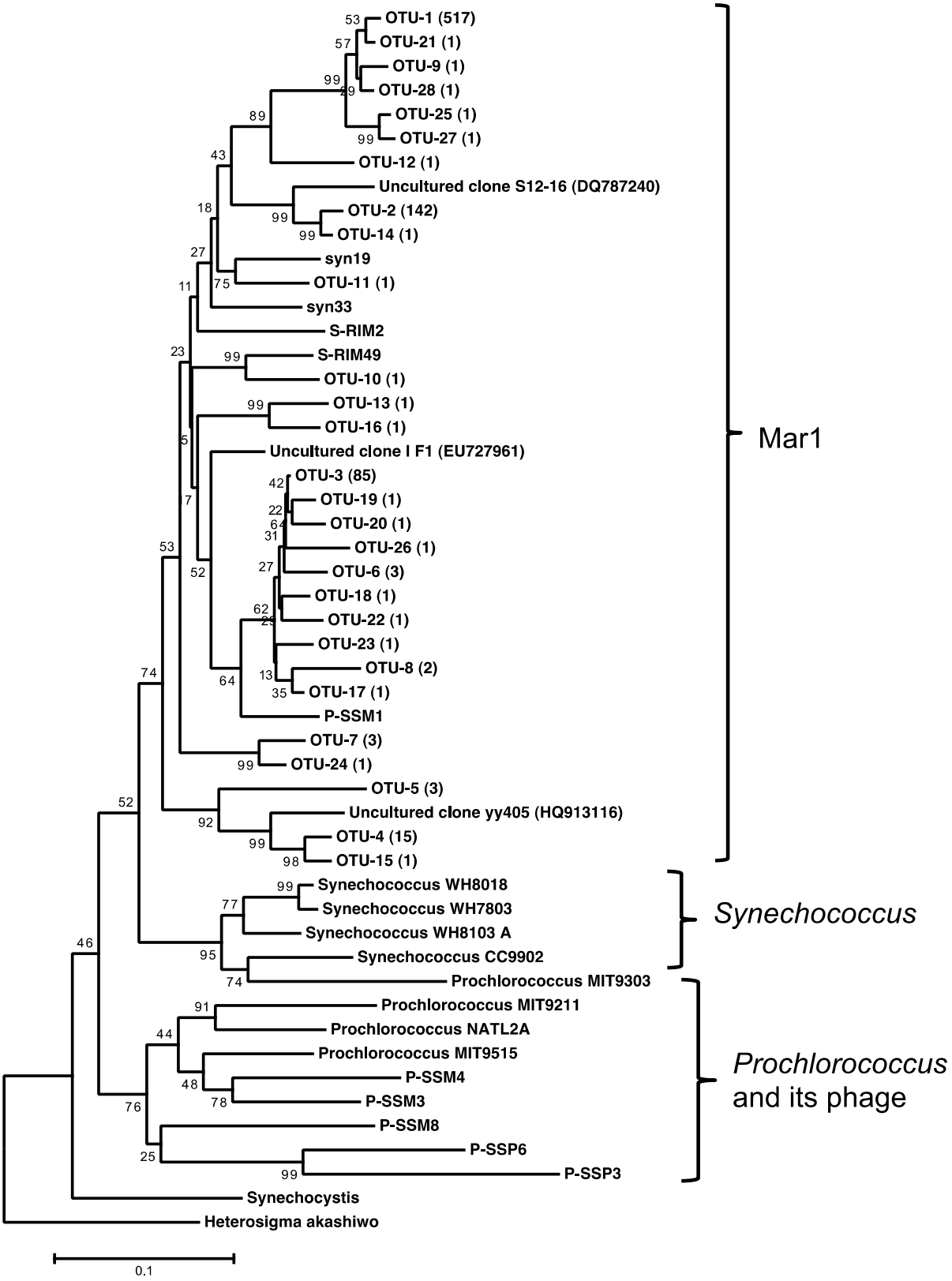

OTU-1 (517)
OTU-21 (1)
OTU-9 (1)
OTU-28 (1)
OTU-25 (1)
OTU-27 (1)
OTU-12 (1)
Uncultured clone S12-16 (DQ787240)
OTU-2 (142)
OTU-14 (1)
syn19
OTU-11 (1)
syn33
S-RIM2
S-RIM49
OTU-10 (1)
OTU-13 (1)
OTU-16 (1)
Uncultured clone I F1 (EU727961)
OTU-3 (85)
OTU-19 (1)
OTU-20 (1)
OTU-26 (1)
OTU-6 (3)
OTU-18 (1)
OTU-22 (1)
OTU-23 (1)
OTU-8 (2)
OTU-17 (1)
P-SSM1
OTU-7 (3)
OTU-24 (1)
OTU-5 (3)
Uncultured clone yy405 (HQ913116)
OTU-4 (15)
OTU-15 (1)
Mar1
Synechococcus WH8018
Synechococcus WH7803
Synechococcus WH8103 A
Synechococcus CC9902
Synechococcus
Prochlorococcus MIT9303
Prochlorococcus MIT9211
Prochlorococcus NATL2A
Prochlorococcus MIT9515
P-SSM4
P-SSM3
P-SSM8
P-SSP6
P-SSP3
Prochlorococcus
and its phage
Synechocystis
Heterosigma akashiwo
0.1

Figure 2B

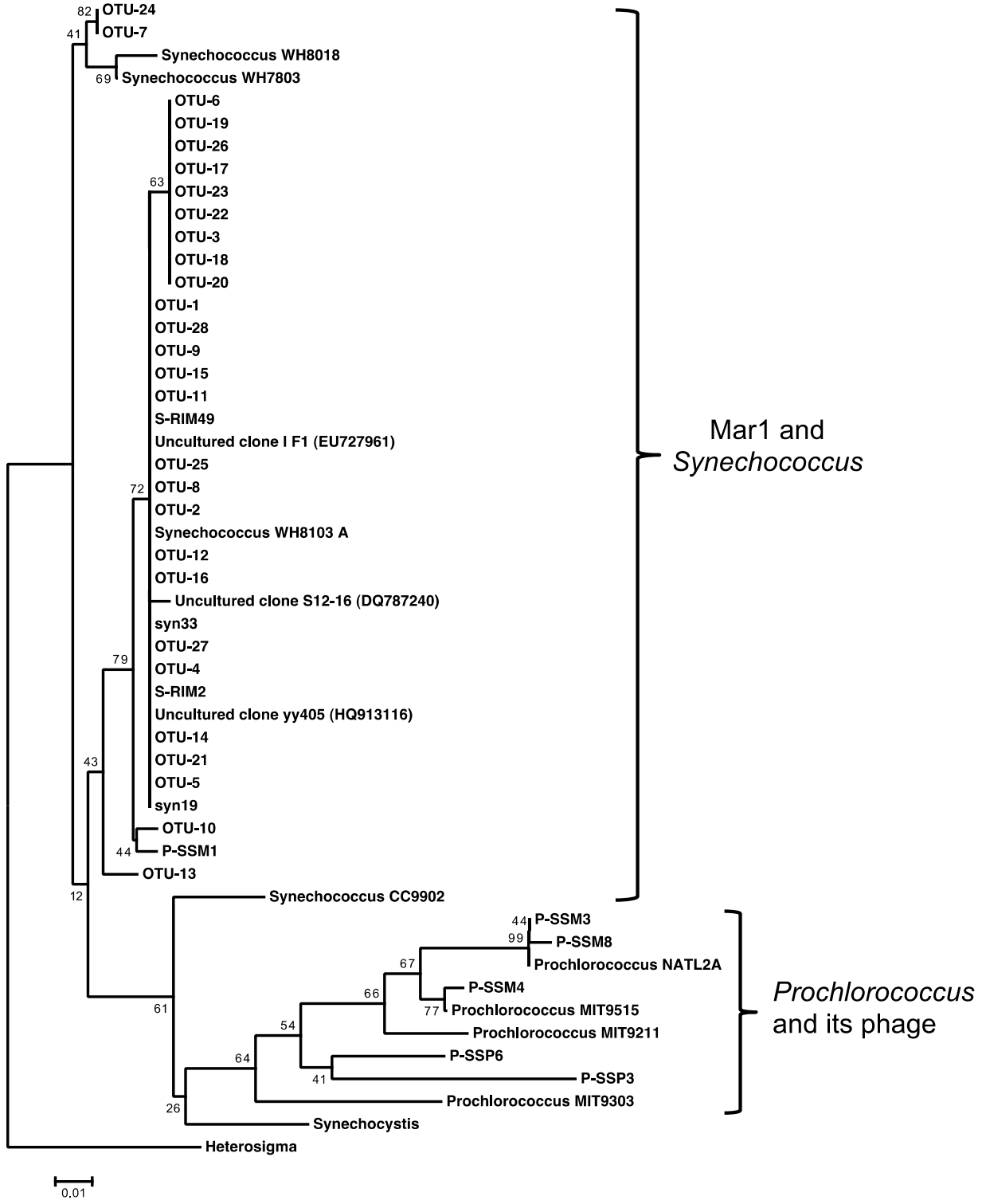
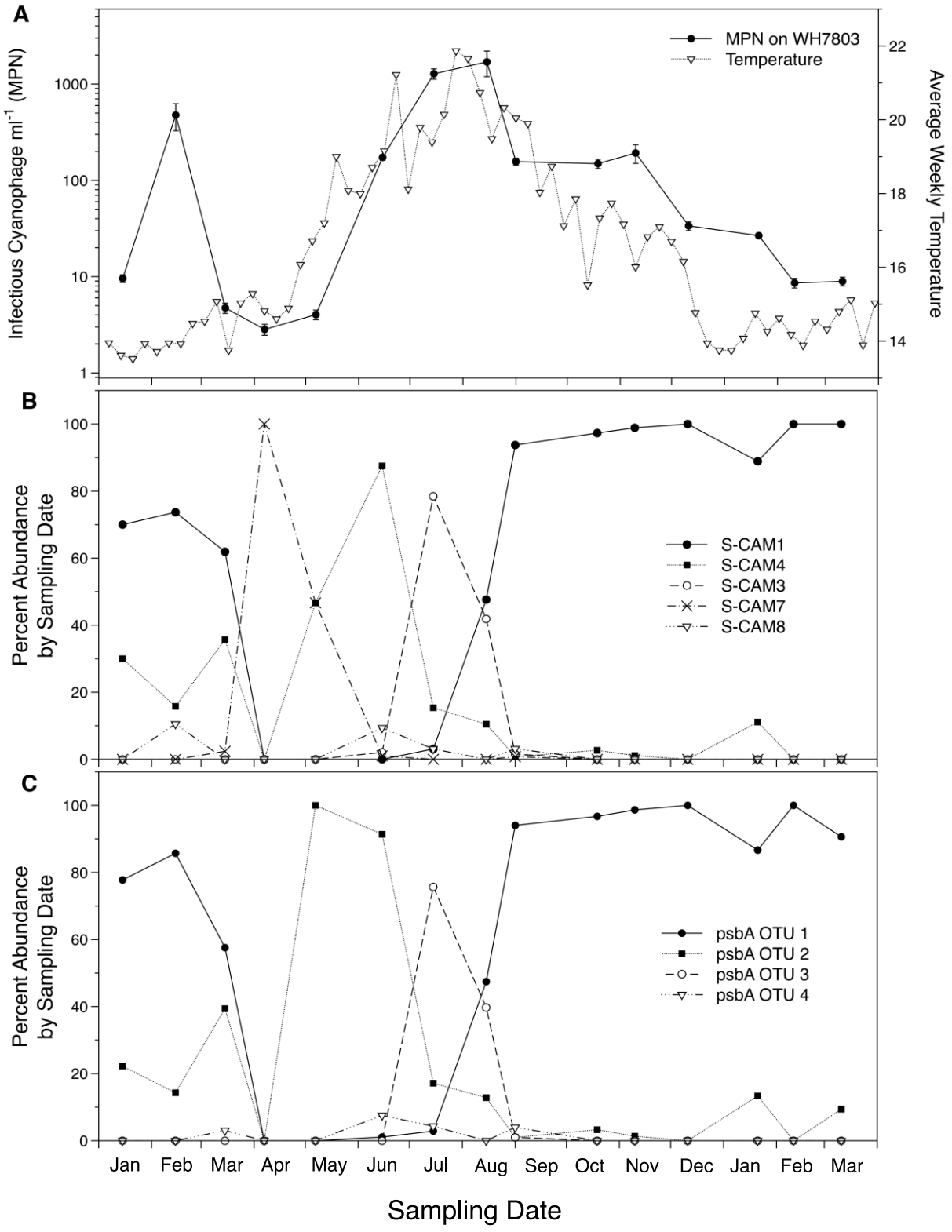

OTU-24
OTU-7
Synechococcus WH8018
Synechococcus WH7803
OTU-6
OTU-19
OTU-26
OTU-17
OTU-23
OTU-22
OTU-3
OTU-18
OTU-20
OTU-1
OTU-28
OTU-9
OTU-15
OTU-11
S-RIM49
Uncultured clone I F1 (EU727961)
OTU-25
OTU-8
OTU-2
Synechococcus WH8103 A
OTU-12
OTU-16
Uncultured clone S12-16 (DQ787240)
syn33
OTU-27
OTU-4
S-RIM2
Uncultured clone yy405 (HQ913116)
OTU-14
OTU-21
OTU-5
syn19
OTU-10
P-SSM1
OTU-13
Synechococcus CC9902
P-SSM3
P-SSM8
Prochlorococcus NATL2A
P-SSM4
Prochlorococcus MIT9515
Prochlorococcus MIT9211
P-SSP6
P-SSP3
Prochlorococcus MIT9303
Synechocystis
Heterosigma
0.01
Mar1 and
Synechococcus
Prochlorococcus
and its phage

Figure 3


A
Infectious Cyanophage ml-1 (MPN)
MPN on WH7803
Temperature
Average Weekly Temperature
B
Percent Abundance by Sampling Date
S-CAM1
S-CAM4
S-CAM3
S-CAM7
S-CAM8
C
Percent Abundance by Sampling Date
psbA OTU 1
psbA OTU 2
psbA OTU 3
psbA OTU 4
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar
Sampling Date