UC Santa Barbara
UC Santa Barbara Previously Published Works

Title

Primer selection impacts specific population abundances but not community dynamics in a monthly time-series 16S rRNA gene amplicon analysis of coastal marine bacterioplankton

Permalink

https://escholarship.org/uc/item/4fd3x0t8

Journal

Environmental Microbiology, 20(8)

ISSN

1462-2912

Authors

Wear, Emma K
Wilbanks, Elizabeth G
Nelson, Craig E
et al.

Publication Date

2018-08-01

DOI

10.1111/1462-2920.14091

Peer reviewed

sfam society for applied microbiology

Primer selection impacts specific population abundances but not community dynamics in a monthly time-series 16S rRNA gene amplicon analysis of coastal marine bacterioplankton

Emma K. Wear 🆔,[1]*† Elizabeth G. Wilbanks 🆔,[1] Craig E. Nelson 🆔[2] and Craig A. Carlson[1]

[1]*Department of Ecology, Evolution and Marine Biology and Marine Science Institute; University of California, Santa Barbara, CA 93106, USA.*

[2]*Center for Microbial Oceanography: Research and Education; Department of Oceanography and Sea Grant College Program, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA.*

Summary

Primers targeting the 16S small subunit ribosomal RNA marker gene, used to characterize bacterial and archaeal communities, have recently been re-evaluated for marine planktonic habitats. To investigate whether primer selection affects the ecological interpretation of bacterioplankton populations and community dynamics, amplicon sequencing with four primer sets targeting several hypervariable regions of the 16S rRNA gene was conducted on both mock communities constructed from cloned 16S rRNA genes and a time-series of DNA samples from the temperate coastal Santa Barbara Channel. Ecological interpretations of community structure (delineation of depth and seasonality, correlations with environmental factors) were similar across primer sets, while population dynamics varied. We observed substantial differences in relative abundances of taxa known to be poorly resolved by some primer sets, such as *Thaumarchaeota* and SAR11, and unexpected taxa including *Roseobacter* clades. Though the magnitude of relative abundances of common OTUs differed between primer sets, the relative abundances of the OTUs were nonetheless strongly correlated. We do not endorse one primer set but rather enumerate strengths and weaknesses to facilitate selection appropriate to a system or experimental goal. While 16S rRNA gene primer bias suggests caution in assessing quantitative population dynamics, community dynamics appear robust across studies using different primers.

Introduction

The phylogenetic composition of the bacterioplankton, the free-living bacteria and archaea in aquatic systems, is important in determining a community's biogeochemical function (e.g., Nelson and Carlson, 2012; Pedler *et al.*, 2014; Logue *et al.*, 2016) and ecological interactions (Nelson *et al.*, 2014; Fuhrman *et al.*, 2015). Meta '-omics' techniques increasingly allow us to interrogate bacterioplankton community composition (BCC) and function together at unprecedented levels of detail. Yet sequencing of phylogenetic marker gene amplicons, in particular the small subunit ribosomal RNA genes supported by a number of robustly annotated databases, remains a valuable tool in many analyses of microbial communities: to address traditional community ecology questions focused on shifts in clearly defined operational taxonomic units (OTUs), such as bottom-up controls, biogeography and seasonality; to accommodate large data sets where shotgun metagenomes are not financially feasible, or experimental work with a known starting community where metagenomes have a lower return on investment; and to resolve phylogenetic characterization of uncultured organisms whose functional genes may not be well represented in current metagenomic reference libraries.

Primers that simultaneously detect the maximum possible range of bacterial and archaeal clades have been a goal since the development of the first set of nominally 'universal' primers (Lane *et al.*, 1985), which amplified rRNA gene sequences from all three domains. Increasing survey depth of the diversity of microbial life has led to the identification of taxa that are poorly amplified by common 16S

rRNA gene primers (Baker *et al.*, 2003), most comprehensively documented in an *in silico* analysis of the taxonomic coverage of 512 primer pairs by Klindworth and colleagues (2013). In marine systems, recent studies have noted difficulties in representatively sampling the 16S rRNA genes of several numerically important clades, including the *Alphaproteobacteria* SAR11 clade (Apprill *et al.*, 2015; Parada *et al.*, 2016) and the *Thaumarchaeota* (Hugerth *et al.*, 2014). Various approaches have been suggested to address this issue, including targeting multiple or different hypervariable regions (e.g., Klindworth *et al.*, 2013; Parada *et al.*, 2016), increasing primer degeneracy (e.g., Apprill *et al.*, 2015; Parada *et al.*, 2016), and physically mixing several primers with sequence differences in specified proportions (e.g., Huber *et al.*, 2007). The resulting variety of primer options has been further complicated by changing sequencing technology, as the transition from Roche 454 pyrosequencing to Illumina technology as the most common amplicon sequencing approach has favoured a shift towards shorter gene regions. There has also been a move towards primer sets that capture archaeal and bacterial 16S rRNA genes simultaneously (Klindworth *et al.*, 2013), following the increasing recognition that archaea fill niches beyond extremophile-type environments and deep water (e.g., in the surface ocean: Luo *et al.*, 2014; Orsi *et al.*, 2015) and the subsequent need to consider their biogeography and ecological roles.

Yet there exist limited systematic comparisons between current, valid primer options as applied to marine samples. Several improved primer sets for aquatic systems have recently been individually evaluated (Apprill *et al.*, 2015; Parada *et al.*, 2016), and these primers have been compared by the Earth Microbiome Project, which uses terrestrial and human-associated standards to benchmark primer sets (Walters *et al.*, 2015). But we lack, to date, a comprehensive study that directly compares current primer options through the sequencing and analysis of actual marine communities. Further, there is a paucity of studies investigating how and when primer selection affects the ecological interpretation of the data – that is, how both populations and communities correlate with bottom-up factors or system-scale events such as phytoplankton blooms, regardless of the exact BCC depicted by the primer sets. For example, while two primer sets may yield different relative abundances of SAR11 types, they could nonetheless both delineate similar shifts between communities associated with events such as upwelling and subsequent phytoplankton blooms. A better understanding of the impact of primer choice on our ability to detect ecological patterns and responses will allow us to assess what conclusions can validly be drawn across studies employing different primers to investigate BCC. Such an understanding is particularly important for retrospective analyses, where studies conducted at different times used different primer sets.

To that end, we directly compared four primer sets (Table 1): both the V4–5 and V4 sets currently recommended for bacteria and archaea by the Earth Microbiome Project (Walters *et al.*, 2015); the V3–4 set suggested for marine bacteria by Klindworth *et al.* (2013); and a V1–2 set of universal bacterial primers (e.g., Fortunato *et al.*, 2012; Doherty *et al.*, 2017). Each primer set was tested on mock communities constructed from 16S rRNA genes of marine bacteria and archaea cloned from the coastal upwelling system of the Santa Barbara Channel, CA, USA, to facilitate direct comparisons of taxonomic range and potential primer biases. Each primer set was also used to amplify 76 field samples from the same system to verify the taxonomic ranges of the primers under realistic sampling conditions. With a subset of these field samples, BCC was independently determined using shotgun metagenomic sequencing and subsequent analysis of 16S rRNA gene fragments. We further used the field samples to investigate whether primer selection impacted ecological findings, by examining the effects of different primers on community shifts over time and depth, as well as bottom-up biological and physicochemical drivers of BCC in this system.

Results

Populations: detection and quantification in mock communities

We sequenced 225 cloned, full-length 16S rRNA genes (8F or 9F to 1492R; Supporting Information Table S1) from 2 archaeal phyla and 5 bacterial phyla, from which we selected 22 unique 16S rRNA genes to construct mock planktonic communities for primer testing (Supporting Information Table S2). Clones were chosen to include both abundant taxa (e.g., clones from multiple genera from the *Rhodobacteraceae* and *Flavobacteriaceae* families, as observed in Wear *et al.*, 2015) and representative clones to cover the observed phylogenetic diversity at the phylum level (e.g., the *Deferribacteres* and *Verrucomicrobia* representatives). We designed two mock communities: one with each of the 22 16S rRNA gene amplicons at equal concentrations, i.e., evenly distributed (referred to hereafter as Even), and one with each of the same 16S rRNA gene amplicons in staggered proportions to approximate a community such as might be associated with a diatom bloom (hereafter, Bloom; based on BCC observed during a natural diatom bloom from Wear *et al.*, 2015).

Four replicates of each mock community were sequenced with each primer set. Samples had 6803–33 174 sequences before subsampling to 6800 sequences. In theory each sample should only contain the 22 OTUs included in the mock communities; however, in

Table 1. 16S rRNA gene primer sets and sources.

Target hypervariable region(s)	Primer IDs	Forward primer sequence (5'–3')	Reverse primer sequence (5'–3')	Bacterioplankton phylogenetic specificity[a]	Reference(s), i.e., use recommended by
V4-5	515F-Y and 926R	GTGYCAGCMGCCGCGGTAA	CCGYCAATTYMTTTRAGTTT	Bacteria and archaea	(Parada *et al.*, 2016); Earth Microbiome Project (Walters *et al.*, 2015)
V4	515F-Y and 806RB	GTGYCAGCMGCCGCGGTAA	GGACTACNVGGGTWTCTAAT	Bacteria and archaea	(Parada *et al.*, 2016); (Apprill *et al.*, 2015); Earth Microbiome Project
V3-4	341F and 785R	CCTACGGGNGGCWGCAG	GACTACHVGGGTATCTAATCC	Bacteria and (some) archaea[b]	(Klindworth *et al.*, 2013)
V1-2	27F and 338R	AGRGTTTGATYMTGGCTCAG	TGCWGCCWCCCGTAGGWGT	Bacteria	Versions of common universal primers (here, as in Fortunato *et al.*, 2012)

a. We intentionally did not consider amplification of eukaryotic genes, whether from chloroplasts or otherwise, as we were primarily interested in bacterioplankton; additionally, our physically size-separated samples contained insufficient eukaryotic cells to compare each primer set's utility in that regard.
b. Klindworth and colleagues (2013) reported that 785R is a universal bacterial and archaeal primer, while 341F was expected to amplify two-thirds of the archaeal taxa examined *in silico* if one mismatch was allowed; however, we note that Klindworth *et al.* only recommend this primer set for bacteria.

practice we observed 53–574 OTUs per sample (Supporting Information Fig. S1). Between 0.49% and 8.10% of sequences were removed from each sample as rare OTUs (those not present at a minimum of 2 copies in at least 3 of the 8 mock community samples per primer set), reducing communities to 20–30 OTUs that were assigned to the 22 source clones as described in the Experimental Procedures.

We considered several aspects of how well the primers reproduced the expected mock communities. First, we examined cloned gene resolution, or how many of the expected taxa were present in the respective samples. Only the V3–4 primers successfully detected all of the clones at a taxonomic resolution comparable to that of the full-length 16S rRNA gene (Table 2; Supporting Information Table S2). The V4–5 and V4 primers each failed to detect one clone. The V1–2 primers failed to detect four clones, including the two archaeal clones that this bacterial primer set is known to miss (Klindworth *et al.*, 2013). Next, we looked at specificity, or what percent of sequences in each sample could be classified to an expected clone. The V4 primer set had 100% classifiable sequences, and the V3–4 set $> 99.9\%$ (Table 2). The V4–5 and V1–2 sets had lower sequence classification associated with 2.5%–10.5% of sequences that could not be classified below the *Rhodobacteraceae* family level, as discussed below.

Finally, we looked at how accurately the primer sets reproduced the expected relative abundances of the clones in the mock communities. For the Even mock community, we calculated Pielou's index (*J*: Table 2; Pielou, 1966), a measure of community evenness, omitting clones that did not amplify and sequences that could not be assigned to a clone for each individual primer set. While all primer sets reflected the high evenness expected from analysis of an evenly distributed mock community, the V4–5 primer set had significantly greater evenness than all others while the V3–4 set had significantly lower evenness than all others (one-way ANOVA with Ryan-Einot-Gabriel-Welsch Range *post hoc*, $F_{3,12} = 33.040$, $p < 0.0001$).

To quantify which clones deviated most from expected values, we calculated the log2-fold ratio of observed to expected relative abundance for both mock communities (Fig. 1; Supporting Information Figs. S3 and S4). Though the Even and Bloom communities differed in expected relative abundances, the log2-fold ratios for corresponding clones in each mock community were very close to a 1:1 relationship within primer sets (Supporting Information Fig. S2). That is, the log2-fold ratio of SAR11 Surface 1 relative to the expected value in the Even community was similar to its log2-fold ratio to expected in the Bloom community. This finding suggests that the tendencies to over- or underestimate particular taxa were more influenced by primer set identity than by the community structure being

Table 2. Summary of pros and cons of the primer sets.

	V4–5	V4	V3–4	V1–2
Attributes from mock communities				
Clone resolution (of 22)	21	21	22	18
% of sequences assigned to source clones	Even: 97.5	Even: 100	Even: 99.96	Even: 94.2
	Bloom: 96.7	Bloom: 100	Bloom: 99.97	Bloom: 89.5
Pielou's index (J'), mean (st. dev.)	0.932 (0.009)	0.905 (0.006)	0.889 (0.005)	0.912 (0.002)
Attributes from field samples				
Amplicon length[a] [median]	411	292	460	327
% removed as plastids, 0–30m samples [mean (range)]	2.54 (0.02–16.13)	0.25 (0.01-1.32)	0.22 (0-0.98)	0.31 (0–2.75)
# of total OTUs across all samples[a]	23,765	10,777	27,514	3,952
% total OTUs as singlets[a]	90.7	72.2	92.2	73.3
# of OTUs per sample [mean (range)]	501.1 (280–772)	347.4 (145–589)	522.9 (366–779)	177.9 (92–305)
Populations biases: consensus from mock communities and field samples				
Populations under-represented[b]	SAR11 Deep 1 *Pseudospirillum*	ZD0405 (field)	*Euryarchaeota* (field) *Thaumarchaeota* SAR11 Surface 1 SAR11 Deep 1 some SAR116	SAR11 Deep 1 *Roseobacter* DC5–80-3 *Roseobacter* OCT
Populations over-represented[b]		*Euryarchaeota Thaumarchaeota*	*Euryarchaeota* (mock)	*Flavobacteria*: NS5
Populations not detected	*Roseobacter* OCT	some SAR116		*Euryarchaeota*[c] *Thaumarchaeota*[c] *Roseobacter* NAC11-7 some SAR116
Clades with poor classification[d]	*Rhodobacteraceae*			*Rhodobacteraceae*
Overall pros and cons of primer sets				
V4–5	- overall high accuracy, most even sampling of mock communities			
	- three domain and plastid sampling (Parada *et al.*, 2016; pro or con depending on study goals)			
V4	- overall high accuracy, most classifiable mock community sequences, over-estimates archaea			
	- short amplicons, best paired-end read overlap			
V3–4	- best mock community resolution (detected all clones) but poor accuracy for key clades (e.g., SAR11) and inconsistent detection of archaea			
	- long amplicon, requires longer sequencing kits than the other three sets			
V1–2	- lengthy history of use			
	- does not amplify archaea			

a. These parameters were calculated on the full data set, before individual samples were removed for poor amplification; 'Total OTU' parameters were determined after subsampling to 4500 sequences per sample. OTU-level parameters would vary with more or less conservative quality filtering; values generated using the same pipeline are presented here for comparison.
b. Over- and under-representation of populations in the mock community was determined based on those taxa with a mean log2-fold difference from expected of >1.58 or < −1.58 in the Even community (the equivalent of a threefold increase or decrease relative to expected).
c. This primer set is known to be bacterial-specific and, therefore, the lack of detection of these clones was expected (Klindworth *et al.*, 2013).
d. Clades with poor classification were defined as those where sufficient sequences to represent a full clone (i.e., >1%) were detected but could not be classified to one of the multiple clones within this family and/or where a common OTU in the field samples was not present in the library, but a poorly classified equivalent at a higher taxonomic level was. Taxa marked as (field) indicate that this issue was observed in the field samples but not in the mock communities, and vice versa.

assessed. Therefore, the following discussion focuses on the Even community.

The threshold of a log2-fold ratio of plus or minus 1.58 (the equivalent of a threefold difference) was used to identify taxa that were substantially over- or under-represented by the primer sets (Fig. 1; Table 2). The V4–5 and V4 primers had the fewest large deviations from expected relative abundances. The V4–5 primers under-represented

SAR11 Deep 1 and the *Gammaproteobacteria Pseudospirillum* sp. clone, while the V4 set over-represented the two archaeal clones. The V3–4 primers under-represented four and over-represented one clone; notably, these primers detected the two archaeal clones differently, over-representing the *Euryarchaeota* and under-representing the *Thaumarchaeota*, while also under-representing two of the three SAR11 clones. The V1–2 primers under-
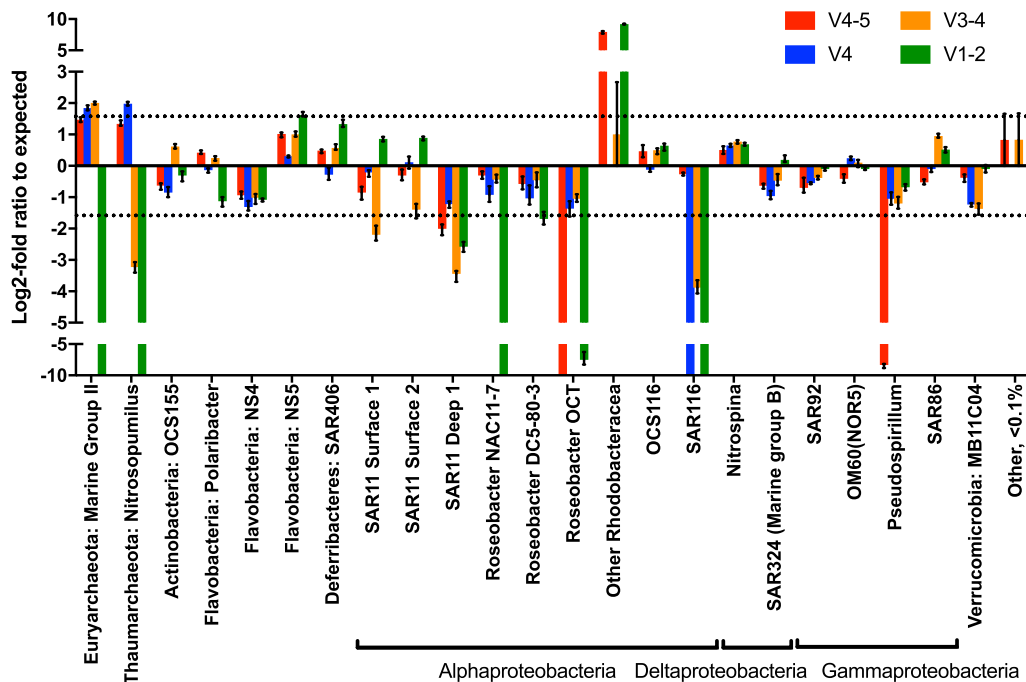
Fig. 1. Log2-fold ratio of observed to expected values in the Even mock community (mean of four replicates; error bars indicate range). 'Other' columns indicate sequences that could not be assigned to an expected clone. As log2-fold ratios cannot accommodate zero values, samples were adjusted as follows. For samples with mean observed relative abundance = 0 and expected > 0, the log2-fold ratio was set to −10, without error bars, which was an arbitrary value more negative than any observed. For samples with observed > 0 and expected = 0, the expected value was set to 0.0001, which is less than 1 sequence per 6800, for calculations. For individual replicates with observed = 0 and expected > 0, where other replicates had observed > 0, relative abundance was set to 0.0001 for calculations. Dotted lines indicate log2-fold ratios of ± 1.58, the equivalent of a threefold over- or under-estimation, which was used as the threshold for inaccurate representation.

represented three clones and over-represented one. The V1–2 primers performed particularly poorly with the *Rhodobacteraceae* family, under-representing two clones from the family and missing the third completely, although more than a clone's worth (mean of 5.8%) of total sequences in each mock community were identified as *Rhodobacteraceae* sequences that could not be assigned to a particular clone. We note that the less accurate resolution and classification with the V1–2 set are not due to the 95% similarity threshold used to cluster this set (following evidence from Schloss (2010) that 95% similarity in the more variable V1–2 hypervariable regions more closely approximates 97% similarity across the whole gene); under a 97% similarity threshold, the same taxa were identified at the same resolution in the mock communities, and relative abundances were highly linear between the two thresholds (Supporting Information Fig. S5).

Populations: in silico *analysis of cloned 16S rRNA genes*

To address whether these detection and classification failures were attributable to primer bias rather than to the extent of phylogenetic information contained in the different hypervariable regions, the full-length cloned gene sequences were trimmed *in silico* to the regions corresponding to each amplicon, then classified using the same approaches. Due to incomplete Sanger sequence coverage, the SAR406 clone could not be analysed for any primer sets, and the OM60(NOR5) clone could not be analysed for the V4–5 primer set. The majority of cloned gene 'amplicons' could be classified with equal resolution to that of the full-length sequence (as in Supporting Information Table S3). Three sequences could not be classified to a comparable resolution, all from the V1–2 primer set. Two of these were missing sequence (8 and 21 bases) at the 5' end due to quality screening, but we do not believe this is driving the poor classification: the missing sequences fell in the conserved region of the gene, and other V1–2 *in silico* 'amplicons' that were missing up to 29 bases could be fully classified. The poorly classified sequences included two clones that were detected and classified correctly in the V1–2 mock communities: the OM60(NOR5) clone could not be classified beyond order *in silico*, and the SAR11 Deep 1 clone could not be classified beyond class. We attribute this discrepancy to the classification in the mock communities being conducted at the OTU level, whereby a consensus taxonomy is determined from all

sequences within that OTU, minimizing the effects of sequencing variance on classification confidence at more resolved taxonomic levels. The *Roseobacter* NAC11-7 clone, which was not detected with the V1–2 primer set in the mock communities, could not be classified beyond the *Rhodobacteraceae* family level *in silico*. Thus, the inability to identify this *Roseobacter* clone in the V1–2 mock community may be due to insufficient information contained in the targeted hypervariable regions rather than resulting from primer bias. In contrast, the other clones that were not detected in the mock communities (SAR116 in the V4 and V1–2 sets, and *Roseobacter* OCT in the V4–5 set; Supporting Information Table S2) were all classified correctly *in silico*, suggesting these omissions are most likely due to primer bias. The two archaeal clones in the V1–2 set could be classified correctly from the V1–2 region, but this is a clear case of primer and template mismatch (as seen in the large differences between the 9F primer region in Supporting Information Table S1 and the 27F primer used for the amplicons in Table 1).

Populations: representation in field samples

We used each primer set to amplify 87 field samples collected in the Santa Barbara Channel to assess potential differences in population quantification under realistic environmental conditions. We restricted our analyses to those samples that had amplified well across all primer sets (after subsampling to 4500 sequences per sample), leaving us a total of 76 samples from each primer set. Parameters including amplicon length, number of OTUs and percent of sequences removed as plastids are reported in Table 2, and rarefaction curves are reported in Supporting Information Fig. S6. The V4–5 and V3–4 primer sets had significantly more total OTUs per sample than the V4 and V1–2 sets, and the V4 set had more OTUs per sample than the V1–2 set [Kruskal-Wallis test ($H = 196.261$, df = 3, $p < 0.0005$) with Mann-Whitney U tests as a *post hoc* with a Bonferroni correction].

The primer sets showed similar overall patterns of taxa abundance in representative field samples (Supporting Information Fig. S7). We identified eight abundant OTUs that were present across primer sets with minimal presence/absence dynamics with which to compare population representation. As multiple OTUs with the same taxonomic identity were always present within a primer set, we focused on taxa with an unambiguous most abundant OTU to increase confidence in selecting the same organism across primer sets. Some clades of interest are, therefore, omitted from this analysis, such as SAR116, which frequently had two common OTUs with very similar relative abundances. For two OTUs with apparent classification issues, *Roseobacter* NAC11-7 with the V1–2 primers and *Oceanospirillales* family ZD0405 with the V4 primers, we used the most common *Rhodobacteraceae* unclassified and *Oceanospirillales* unclassified OTUs.

These eight OTUs had highly linear relationships between primer sets (Fig. 2), although the ranges of percent BCC varied and Model II regressions indicated that few primer set comparisons had a slope of 1 (Supporting Information Table S5). Slopes notably different from 1 that were consistent with the mock communities included all regressions with the V3–4 set for SAR11 Surface 1 and Surface 2 and the V4–5 and V4 comparison for *Nitrosopumilus*. OCS155 was over-represented in the V3–4 primer set relative to V4–5 and V4; this is consistent with trends in the mock community, though none of the primer sets showed a substantial deviation from the expected value on their own. The slopes for the OTU from the *Oceanospirillales* family ZD0405, which was not included in the mock community but was abundant in the field samples, were particularly variable between primer sets, indicating that the V4 and to a lesser extent the V4–5 primers were underestimating this OTU.

Field samples: comparisons with metagenomes

We generated shotgun metagenomes from 10 field samples, from which we classified 16S rRNA genes using the same custom SILVA database that was used for the amplicon analyses. This provided an independent measure of BCC, free of potential 16S rRNA gene-specific primer biases. Metagenomes had a total of 0.8–1.2 million sequences (median 1.1 million), of which 2993–6006 sequences of at least 100 basepairs length were identified as 16S rRNA gene fragments by MG-RAST (Meyer *et al.*, 2008). After removing chloroplasts and sequences that could not be classified at the domain level, 1413–2730 (median 2012) bacterial and archaeal sequences remained per metagenome. We compared BCC between the primer sets and the metagenomes at the phylum level (class level for the *Proteobacteria*) and within additional clades of interest at more refined taxonomic levels. We maintained the unclassified and rare (here, those not constituting 1% or more of at least two metagenomes) bacterial sequences and the unclassified *Proteobacteria* sequences in calculating relative abundances, though they are not discussed. These unclassified bacterial sequences, and their greater abundance in the metagenomes, reflect a shortcoming of deriving BCC from shotgun metagenomes, which include all parts of the 16S rRNA gene rather than targeting the most phylogenetically informative regions as the primer sets do.

At the phylum and class level, log2-fold ratios were calculated by dividing relative abundances from amplicon libraries by relative abundances within each sample's respective metagenome (Fig. 3; Supporting Information Table S6). All primer sets were moderately accurate in
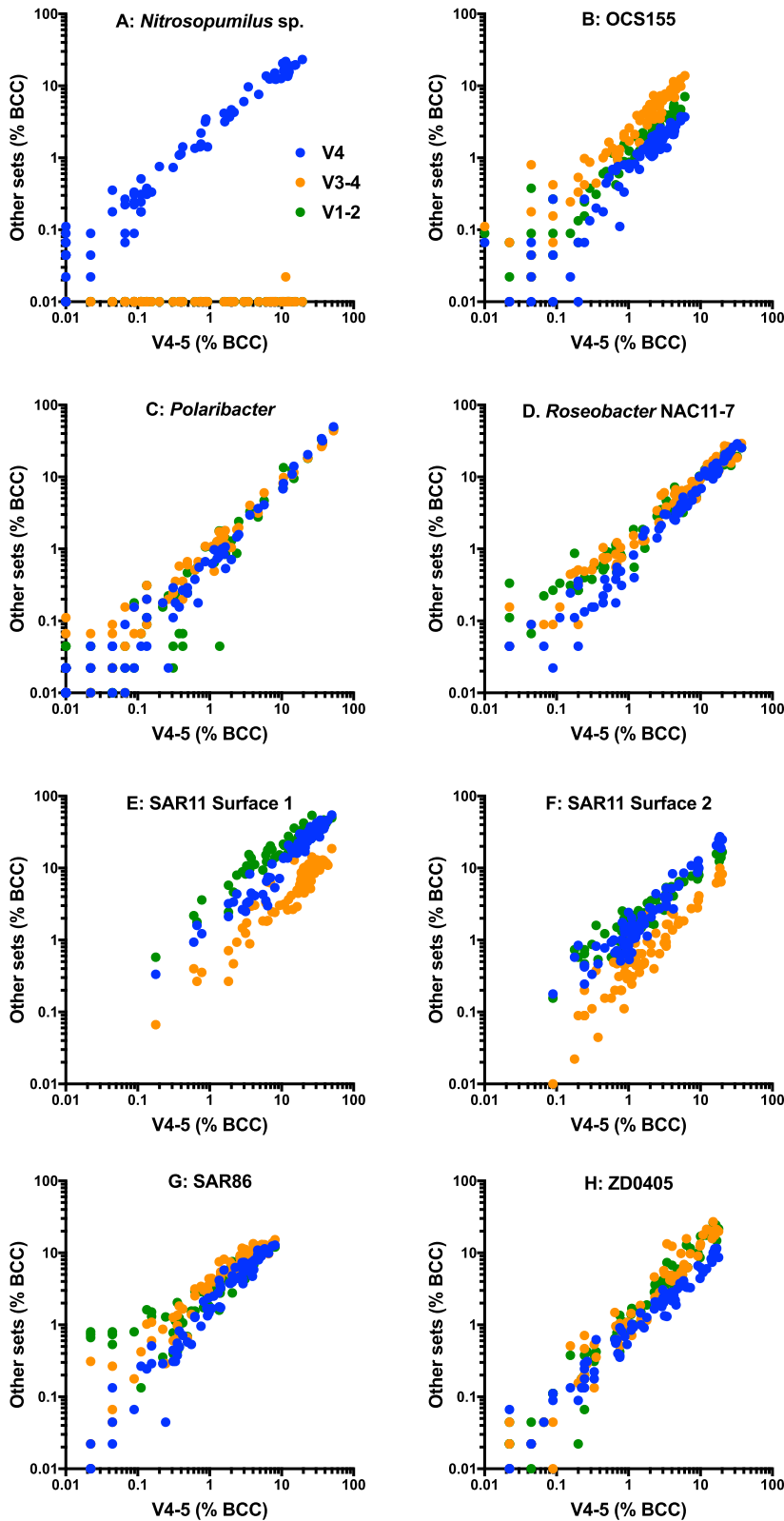
Fig. 2. Comparisons of percent of BCC of major OTUs in the field samples, with each of the primer sets plotted against the V4–5 set. To accommodate the logged axes, zero values were set to equal 0.01%. (A) *Nitrosopumilus.* (B) OCS155. (C) *Polaribacter.* (D) *Roseobacter* NAC11-7, or the most abundant *Rhodobacteraceae* unclassified OTU in the V1–2 primer set. (E) SAR11 Surface 1. (F) SAR11 Surface 2. (G) SAR86. H: ZD0405, or the most abundant *Oceanospirillales* unclassified OTU in the V4 primer set. (This V4 OTU was correctly classified using an updated SILVA v132 database, as in Supporting Information Table S4.)

reproducing the metagenomic relative abundance of the major bacterial groups present in the Santa Barbara Channel. The *Bacteroidetes*, *Alphaproteobacteria* and *Gammaproteobacteria* were consistently within a mean log2-fold ratio of 1.06 or less of the metagenomes across primer sets. The archaeal phyla and the less abundant
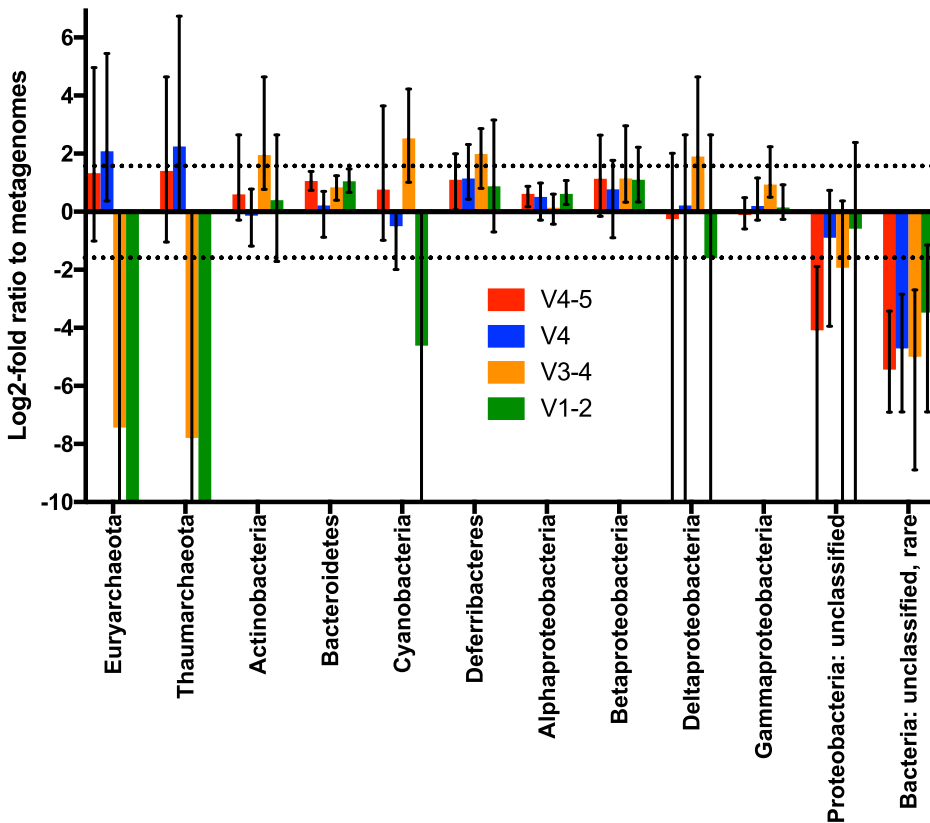
Fig. 3. Log2-fold ratio of relative abundance of taxa in each primer set to relative abundance in the respective metagenomes. Each column represents the mean log2-fold ratio for ten samples within a primer set for a particular phylum (or class, for the *Proteobacteria*); error bars indicate the range. Unclassified bacterial sequences and rare clades (those not constituting 1% or more of at least two metagenome communities) were grouped. Zeroes were handled differently here than in the log2-fold ratios in Fig. 1, as 'expected' values of zero were possible in the metagenomes but not in the mock communities. When both the metagenome and the amplicon sample equalled zero, the log2-fold ratio was manually set to zero. When the metagenome was zero but the amplicon sample had a value, the metagenome was set to 0.0001 and the log2-fold ratio was calculated. When the metagenome had a value but the phylum or class was not detected by the primer set, the log2-fold ratio was manually set to −10.

bacterial phyla were more variable. The V4–5 primers had no phyla or classes with a mean log2-fold ratio greater than 1.58, suggesting they had the greatest accuracy. The V4 set overrepresented both *Euryarchaeota* and *Thaumarchaeota* but accurately represented the bacterial clades. The V3–4 set severely underestimated both the *Euryarchaeota* and *Thaumarchaeota* and overestimated the *Actinobacteria*, *Cyanobacteria*, *Deferribacteres* and *Deltaproteobacteria*. The V1–2 primers underestimated the *Cyanobacteria* and *Deltaproteobacteria* and as expected did not detect either archaeal phylum.

Select taxa of interest at the genus to order levels were also compared between the amplicon data sets and the metagenomes (Supporting Information Fig. S7 and S8; Table S6). Most taxa were linearly related to the relative abundances in the metagenomes, though the ranges were skewed due to the unclassified metagenome sequences. Those taxa where a primer set deviated from the metagenomes generally reflected issues also seen in the mock communities and field amplicon comparisons. For example, the V1–2 primers underestimated the *Roseobacter* genus but not the *Rhodobacteraceae* family, which is consistent with the classification issues seen with *Roseobacter*s in the mock community detailed above. Likewise, the V3–4 primers underestimated both the SAR11 order and the SAR11 Surface 1 family relative to

the metagenomes, and the V4 primers underestimated the *Oceanospirillales* family ZD0405.

Fidelity of mock community and field sample results: SAR116 as a case study

The majority of population sampling issues observed in the mock communities were also clearly present in the field samples, with one clear exception being SAR116. The SAR116 clone included in our mock communities was not amplified by two of the primer sets (V4 and V1–2) and was under-represented by a third (V3–4). However, the V4–5 primer set accurately represented the clone, ruling out a library construction issue, and SAR116 OTUs were moderately abundant in the field samples with all primer sets (maximum abundances of 2.4–5.5%; Supporting Information Table S7). Notably, Parada and colleagues (2016) observed a similar discrepancy with SAR116, finding poor detection of one organism in a clone-based mock community but abundant SAR116 in field samples. Therefore, we examined the SAR116 sequences in greater detail to understand what might be driving this disconnect.

To determine if both library types were sampling the same organism at approximately the OTU level, we trimmed our SAR116 clone sequence *in silico* to the region of each amplicon and compared these subsets with all
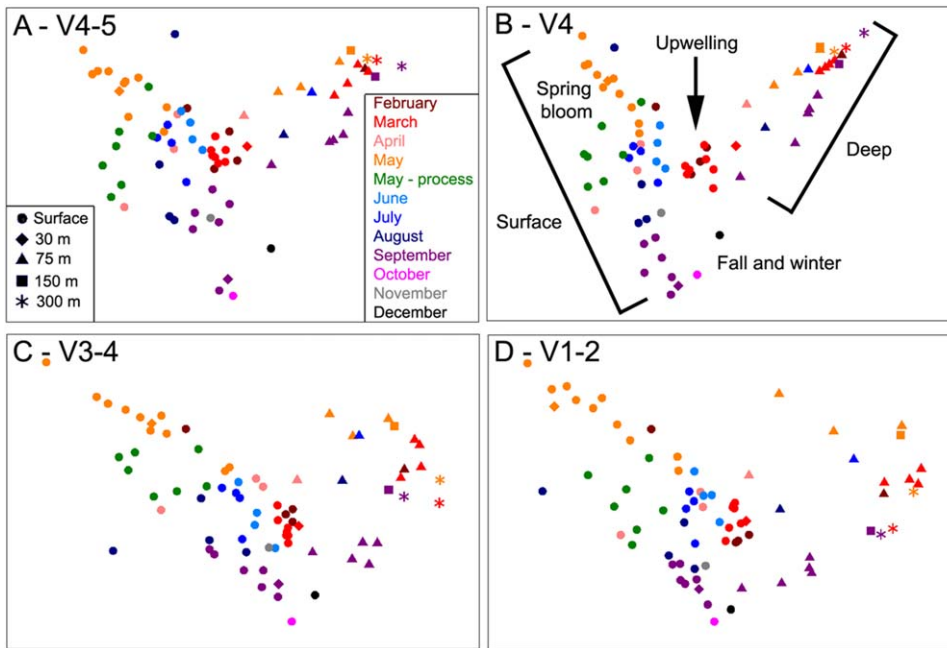
Fig. 4. NMS plots of field samples, including samples from all depths. All figures have been rotated for similar orientations. Symbol colour indicates sampling month (including the process cruise that was separate from the time-series sampling) and shape indicates sampling depth, as in A. Features discussed in the text are annotated in B, and are located in similar positions in all plots.
A. V4–5 primers, 2D stress = 0.1.
B. V4 primers, 2D stress = 0.09.
C. V3–4 primers, 2D stress = 0.09.
D. V1–2 primers, 2D stress = 0.09.

SAR116 amplicons present in the 200 most abundant OTUs from each library in BLAST+ (Camacho *et al.*, 2009). The V4–5, V4 and V3–4 primer sets all had one or more OTUs with representative sequences that were greater than 97% similar to the corresponding clone region (Supporting Information Table S7); the V1–2 region had no amplicons that were potentially the same OTU. Within the primer binding regions, the SAR116 clone had at least one discrepancy from every field sample amplicon within every primer region except the V1–2 forward primer, in most cases different but valid options for a degenerate base. However, the clone also had a single base pair mismatch from the primer sequence at a non-degenerate base in the region shared by the V4 (806R-B) and V3–4 (785R) reverse primers that none of the field amplicons shared. Parada and colleagues (2016) likewise reported a single base pair mismatch between the older 806R primer they used and a SAR116 clone that was under-represented in their mock community with the 515FY-806R primers, but the mismatch they observed in the 806R primer region was offset by two base pairs from the mismatch present in our clone (Supporting Information Table S7).

Ecological interpretation: community structure of field samples

While all primer sets broadly reflected expected BCC, there were notable deviations within certain numerically abundant clades. We, therefore, investigated the implications of these population-level differences for interpreting community ecological patterns. How validly can we compare conclusions regarding community seasonality, stratification across depths and bottom-up controls on BCC between studies that were conducted using different 16S rRNA gene primer sets?

We first considered two analyses based on weighted UniFrac distances between samples: nonmetric multidimensional scaling (NMS) ordinations and Mantel tests, which we used to correlate the UniFrac distance matrices between primer set pairs. When samples across all depths and seasons were ordinated, all primer sets generated similar patterns defined by the separation between surface and deep-water samples and by seasonality (Fig. 4). NMS ordinations of only the surface samples more clearly showed seasonal variability (Supporting Information Fig. S9), particularly the consistent distinction between the May samples from both cruise programs, which were collected during strong phytoplankton blooms, and the fall and winter samples (September through March), which were associated with more oligotrophic conditions during stratification and early upwelling. All primer sets captured seasonal temporal progressions in surface waters, depth stratification and seasonal transitions in waters below the euphotic zone, and differentiating of the spring upwelling period within the surface waters as a multivariate shift towards deeper community types.

Mantel correlations of UniFrac distance matrices between primer sets indicated that two pairs of primer sets were most similar to one another: sets V4 and V4–5 (Spearman's rho = 0.944) and V3–4 and V1–2 (rho = 0.938, vs. rho of 0.840–0.867 for all other comparisons; Supporting Information Table S8). Because the differences between primer sets were more apparent in the arrangement of deep-water samples than of surface

samples in the NMS plots, and because the most similar sets were coincident with the primers' tendencies to over- or under-amplify *Thaumarchaeota*, we removed all archaeal sequences *in silico* and recalculated the UniFrac matrices. Removing archaea reduced the tighter clustering of deep-water samples in the V4 and V4–5 NMS plots (Supporting Information Fig. S10) and led to an increase in Mantel test correlation coefficients between all pairwise comparisons, though the same two pairs remained most similar (rho = 0.949 for V4–5/V4 and 0.951 for V3–4/V1–2, vs. 0.882–0.911 in all other comparisons; Supporting Information Table S8).

To quantify characterization of changes in communities over depth, we compared the weighted UniFrac distances between 15 paired surface and 75 m samples collected from the same sampling rosette casts. The UniFrac distances between these depths determined by the V4 primers were significantly greater than those of all but the V4–5 primers, and the V4–5 primers had significantly greater UniFrac distances than the V1–2 primers (Supporting Information Fig. S11; one-way ANOVA with Ryan-Einot-Gabriel-Welsch Range *post hoc*: $F_{6,98} = 10.762$, $p < 0.0001$). Consistent with the NMS ordinations, when archaeal sequences were removed *in silico*, the UniFrac distances between surface and 75 m samples in the V4 and V4–5 primer sets decreased, and these two sets were no longer significantly different from the V3–4 and V1–2 sets. In contrast, the percent change in Shannon diversity from surface to 75 m had few significant differences between primer sets (Supporting Information Fig. S11; one-way ANOVA with Ryan-Einot-Gabriel-Welsch Range *post hoc*: $F_{6,98} = 3.893$, $p = 0.002$); the V4 set without archaea had a significantly larger diversity gradient than the V3–4 set with or without archaea.

Ecological interpretation: relationships with physicochemical parameters

Both the OTUs discussed above and the community as a whole were correlated with a representative suite of bottom-up, physicochemical parameters: *in situ* temperature, nitrate + nitrite, chromophoric dissolved organic matter (CDOM) spectral slope coefficient, bacterial production (BP) via ^{3}H-leucine incorporation, chlorophyll *a* (Chl *a*), and particulate organic carbon (POC). Correlation analyses were restricted to only surface samples, as the increased range of physicochemical parameters over depth could independently influence correlations. The *Nitrosopumilus* and ZD0405 OTUs, which were most abundant in deep water samples, were omitted from this analysis.

Consistent with the linear relationships in OTU relative abundance between primer sets (Fig. 2), the patterns of bottom-up correlates were similar within OTUs, with
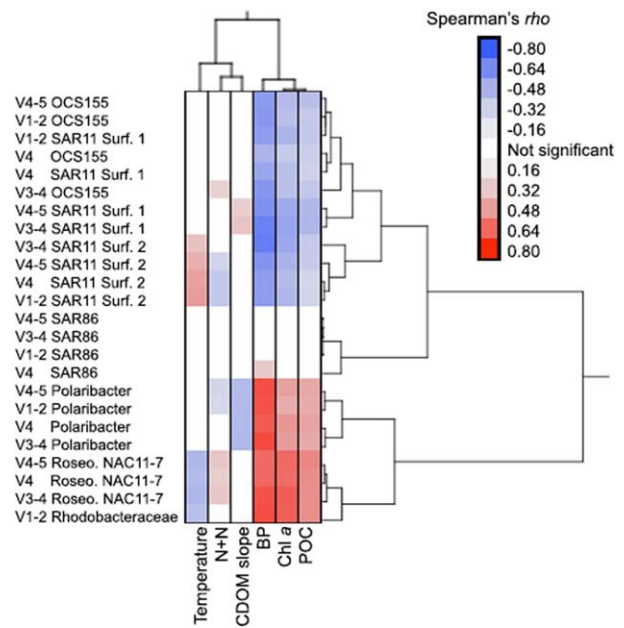

Fig. 5. Correlations between select OTUs and bottom-up environmental factors within surface samples, arranged according to hierarchical clustering. Correlations are Spearman's rho. Correlations that were not significant at $p < 0.05$ are coloured white.

differences only in parameters that were weakly correlated (Fig. 5). In any scenario where an OTU from one primer set and an environmental parameter were correlated at Spearman's rho > 0.4, the correlation of that OTU-parameter combination was consistently significant with the same sign across all primer sets. Each primer set indicated similar ecological niches for the major OTUs in the Santa Barbara Channel (Fig. 5). SAR11 Surface 1, SAR11 Surface 2 and OCS155 were consistently negatively correlated with phytoplankton blooms (with phytoplankton biomass indicated by Chl *a* and POC, and BP reflecting the associated increased resources during blooms). SAR11 Surface 2 was also positively correlated with temperature, here indicating the warm stratified summer and fall period. The two copiotrophic OTUs were clearly distinguished from the more oligotrophic OTUs. *Roseobacter* NAC11-7 was associated with both phytoplankton bloom parameters and upwelling conditions (negatively correlated with temperature and positively with nitrate + nitrite), while *Polaribacter* was positively correlated with phytoplankton blooms and fresher dissolved organic matter [negatively correlated with CDOM spectral slope coefficient, which has been shown to decrease with fresher DOM in this system (Wear *et al.*, 2015)]. SAR86 did not display strong or consistent covariation with any of the environmental parameters that were examined.

For community-level analysis, the BIO-ENV routine in Primer was used to determine the best-fit relationship between BCC and the same bottom-up parameters

(except for POC, which had missing data points). The V4–5 and V4 sets were best correlated with temperature and BP (Spearman's rho = 0.376 and 0.441 respectively), while the V3–4 and V1–2 sets were best correlated with BP and Chl *a* (rho = 0.329 and 0.331) (Supporting Information Table S9). We note, however, that the parameters that differed between primer sets, temperature and Chl *a*, are significantly correlated in this upwelling-driven system (Spearman's rho = −0.489, *N* = 52, *p* < 0.0005; Supporting Information Table S10).

Discussion

Primer choice clearly influences the accuracy of relative abundance quantification of bacterial populations, as there is large variability in the ability of 16S rRNA gene primer sets – even those accepted in the current literature – to amplify specific, abundant and/or ecologically relevant OTUs. Though we do not endorse a single primer set, the V4–5 and V4 primer sets are clearly the best options for simultaneous detection of bacteria and archaea, with each offering specific pros and cons. The biases we observed caution against quantitative comparisons of populations between studies conducted with different primer sets. Nonetheless, we found that primer choice does not greatly affect ecological interpretations of multivariate BCC, whether looking at community patterns over time and depth or how both the community as a whole and specific populations correlate with environmental parameters. This is especially important as primer sets change over time, or when comparing work of authors with different primer preferences. Our findings provide evidence that community and population responses to events such as phytoplankton blooms and mixing, or seasonal patterns, or spatial distributions, can be validly compared in a qualitative, though not quantitative, manner.

Ecological interpretation

A major finding of this study was that ecological interpretations of the results generated by the different primer sets tested here were very similar. The overall communities had similar relationships between samples with respect to depth stratification and seasonality (Fig. 4 and Supporting Information Fig. S9), even when the magnitudes of the populations underlying those patterns varied (Fig. 2). The relative abundances of many OTUs were correlated when the same samples were compared between primer sets, and both community- and population-level analyses showed correlations with similar bottom-up, physicochemical parameters across primer sets. Thus, it is valid to compare broad conclusions from work conducted using dissimilar primer sets, in particular to relate patterns between systems or contextualize with long-term time-series studies.

Few studies have explicitly examined the effects of 16S rRNA gene primers on the interpretation of aquatic bacterioplankton ecology rather than on population quantification. Our results contrast with those of Sánchez and colleagues (2007), who found that primer sets varied in their detection of seasonality in a coastal bacterial community. This discrepancy is potentially methodological, as Sánchez *et al*. were testing primers for community fingerprinting by denaturing gradient gel electrophoresis rather than for sequencing.

We emphasize that the magnitudes of individual bacterioplankton populations should not be compared between primer sets in the same way that broad community and ecological patterns can. For example, our results suggest that if researchers compared two coastal systems sampled with two different primer sets, it would be valid for them to report that both show a unique community in late summer, associated with a seasonal increase in the relative abundance of a SAR11 Surface 2 OTU, and that the summer communities as a whole and the SAR11 Surface 2 OTU populations are negatively correlated with chlorophyll *a* in both systems. However, based on the results of our primer inter-comparison, it would be invalid for the researchers to conclude that one system has a more extreme seasonality because SAR11 Surface 2 reaches a maximum abundance of 30% of BCC in the summer, while in the other system it has a maximum of 15%, when the two systems were sampled with different 16S rRNA gene primers. Though the relative abundance of OTUs was generally linearly related between primer sets (Fig. 2), only a minority of the OTUs examined were related with a slope of 1 (Supporting Information Table S5), indicating that qualitative comparisons are valid but quantitative often are not.

Pros and cons of primer sets

The pros and cons and observed biases of the primer sets are presented here as a guide in selecting the best option for a particular system or study (Table 2). We intentionally do not endorse one primer set over the others, for several reasons. No set was unambiguously superior to all others, though the V4–5 and V4 sets had fewer biases overall. The nature of a given study may make particular biases more or less tolerable. For example, when analysing experimental incubations, it may be desirable to select primers that accurately represent the copiotrophic bacteria that flourish in the absence of grazing (e.g., Nelson and Wear, 2014; Pedler *et al*., 2014), but accurately quantifying these bacteria may be less critical in a field study of the oligotrophic open ocean. These results are derived from a discrete sequencing run conducted with a particular multiplexing approach and indexes; library preparation and sequencing run biases could also impact the results. Finally, we tested these primers in a specific system that

contains a particular bacterioplankton community and, therefore, cannot generalize our results to all aquatic ecosystems. The surface waters of the Santa Barbara Channel lack, for example, the striking numerical dominance of picocyanobacteria observed in the tropical oligotrophic gyres (Supporting Information Table S6; Fig. S7), to the extent that we obtained no cyanobacterial 16S rRNA genes in our clone library and thus did not include that phylum in our mock communities.

The V4–5 and V4 primer sets were clearly superior for simultaneous bacterial and archaeal characterization, with each presenting similar magnitudes of issues. Each did not detect one clone in the mock communities and had one obvious clade with poor classification. The V4–5 set under-represented two bacterial clones, while the V4 set over-represented both archaeal clones. The V4–5 set had the most even sampling of bacterial and archaeal clones in the mock communities; the V4 set was the only primer set in which all mock community sequences could be assigned to a source clone.

These results differ from the conclusions Parada and colleagues (2016) drew from a comparison of the V4–5 set used here and a V4 set with the same forward primer but an older reverse primer (806R) (Supporting Information Fig. S12). The clear improvement in V4 SAR11 detection is attributable to our use of 806R-B from Apprill and colleagues (2015), which was specifically designed to improve detection of this clade. Many of the biases Parada *et al.* identified in the V4 primer set, such as the under-amplification of *Thaumarchaeota* and the over-amplification of SAR86, were not observed in our data set, and in fact we saw a pronounced over-amplification of *Thaumarchaeota*. Beyond the differences in reverse 806R primers, we speculate that many of the dissimilarities between the two studies are due to inclusion of different clones in the respective mock communities. Although the Santa Barbara Channel is only approximately 160 km northwest of the San Pedro Ocean Time Series (SPOT) sampling location, the two sites are situated in distinct biogeochemical regimes. The western Santa Barbara Channel experiences more pronounced seasonal, wind-driven upwelling and associated productivity than SPOT, which is within a wind shadow in the Southern California Bight (Winant and Dorman, 1997). Our mock community thus included more copiotrophic (e.g., multiple *Roseobacter* clades) and deep-water (e.g., *Nitrospina*) clones.

The V3–4 primer set had a broad taxonomic range, detecting all clones, but its patterns of over- and under-estimating taxa could be detrimental in certain systems. It consistently underestimated SAR11 Surface 1 and Deep 1 compared with the other three primer sets (Fig. 2). Klindworth and colleagues (2013) found that the V3–4 primers amplified SAR11 Surface 1 within a similar order of magnitude to the relative abundance in shotgun metagenomes,

although they only considered three environmental samples. Potentially more serious, the V3–4 set treated the archaea common to this system differently, overestimating the *Euryarchaeota* clone in the mock community (though under-representing the *Euryarchaeota* in field samples) while severely underestimating the *Thaumarchaeota* clone. Though Klindworth *et al.* only recommended this set for bacteria, they found that the forward primer covers 66% of archaeal taxa *in silico* and the reverse primer 97%. Unfortunately, their analysis indicated that the primer 341F has 0% coverage of *Candidatus Nitrosopumilus* with one mismatch allowed *in silico*, and our results confirmed that this mismatch produces a significant bias in practice.

The V1–2 primer set has a lengthy history of use in the field, but its explicit omission of archaea and overall biases suggest it can be less informative than newer primer sets. Many of the biases in the bacteria-specific V1–2 primer set appeared to be related to difficulties in detecting and correctly classifying the *Roseobacter* clones and their corresponding OTUs in the field samples. This finding might be attributable to insufficient phylogenetic information for this group in the V1–2 portion of the gene, as a V1–2 'amplicon' generated *in silico* from the *Roseobacter* NAC11-7 clone could not be classified beyond the family level.

Factors beyond biases in representing specific clades may also be important in primer selection, for example amplicon length. Shorter amplicons have greater overlap between reads in paired-end Illumina sequencing, aiding in sequence error checking. Shorter amplicons also allow for the use of Illumina reagent kits with fewer cycles, which can be more reliable. Illumina recommends a minimum 50 base pair overlap between reads (Illumina, 2013), and thus the 460 base pair amplicons generated by the V3–4 region require paired-end 300 kits (300 sequencing cycles from each end of the amplicon) rather than the paired-end 250 kits that are sufficient for the other primer sets. Longer amplicons covering multiple hypervariable regions also clustered into more OTUs (Table 2). This additional information can be ecologically relevant, as Parada and colleagues (2016) nicely demonstrated with seasonal patterns within the SAR11 clade using the V4–5 primers. However, if one is primarily interested in BCC patterns rather than population dynamics, our results suggest that this elevated level of OTU resolution may not be necessary. These additional OTUs can be disadvantageous in that they increase the size of bioinformatics data that must be processed, potentially restricting available analytical options; for example, mothur requires increasing memory as files such as distance matrices increase in size.

Multiple hypervariable regions can also yield increased taxonomic range, which may be considered a pro or a con depending on study design. For example, the V4–5 primer set examined here amplifies eukaryotic 18S rRNA genes

and plastids (Parada *et al.*, 2016), which is advantageous for community surveys intentionally sampling beyond the bacterioplankton (e.g., Needham and Fuhrman, 2016). However, in work exclusively targeting the bacterioplankton, this results in unusable sequences and could be of particular concern in whole-water samples. Our field samples from the V4–5 primer set had substantially more chloroplast sequences than those from any other primer set (up to 16.13%, versus a maximum of 2.75% in the next highest primer set, in samples that were size-fractionated to the 0.2–1.2 µm range; Table 2); as we removed plastid sequences *in silico*, those sequences essentially wasted reads that could otherwise be sampling bacterioplankton.

Fidelity of mock community and field sample results: SAR116 as a case study

We observed a disconnect between the mock community and field results for one organism in particular, SAR116, a potentially photoheterotrophic generalist (Oh *et al.*, 2010) common in the oligotrophic surface ocean (e.g., Morris *et al.*, 2012). The SAR116 clone was poorly represented in three of four mock communities while SAR116 OTUs were moderately abundant in field samples. A close examination of the clone and OTU representative sequences suggested a single base pair mismatch in the V4 and V3–4 reverse primer binding region of the cloned gene may be to blame, while that mismatch was not observed in the abundant field amplicons.

We draw two conclusions from this example. First, this suggests that no single method of primer testing is comprehensive. Here, neither *in silico* analysis, mock communities, nor field samples in isolation would have truly explained the primer bias we observed. The mock community alone would have suggested that three of our four primer sets sampled SAR116 very poorly, whereas the field samples alone would have suggested that all of our primer sets sampled the most abundant SAR116 OTU approximately equally well and at a similar magnitude to the relative abundance in the metagenomes (Supporting Information Figs. S7 and S8). Combining these data sets indicated that all primer sets amplified several SAR116 OTUs in the field samples; however, the specific organism included in the clone library was not abundant in the field and was either more rare than the OTUs examined or was clustered into OTUs that were overall 97% or more similar in the V4–5, V4 and V3–4 primer sets but had different primer regions. We would not have predicted the observed issues from an *in silico* analysis, as the V1–2 primer set had no primer mismatches. Likewise, a single mismatch as seen with the V4 and V3–4 sets is generally tolerated in primer analyses, but the combination of our results and the similar issue observed by Parada and colleagues (2016)

suggest that SAR116 is not robust to single base pair mismatches with some primers.

Thus, our second conclusion is that this example further illustrates that all 16S rRNA gene primer sets for bacterioplankton present trade-offs in phylogenetic coverage and overall accuracy. While the field should certainly strive to select primers that minimize biases, as primer sets improve at targeting major organisms known to be difficult to sample (Apprill *et al.*, 2015; Parada *et al.*, 2016), it is likely that biases towards other organisms such as SAR116 will become more apparent. Therefore, while such multidomain primer sets are well suited for questions of community ecology, detailed population analysis should be approached with care.

Conclusions

Our comparison of four current 16S rRNA gene primer sets indicated that each option presents trade-offs in phylogenetic range, accuracy of population abundances, and sequencing considerations. Overall, the V4–5 primer set suggested by Parada and colleagues (2016) and the V4 primer set from Parada and colleagues (2016) and Apprill and colleagues (2015) were the best options for simultaneous bacterial and archaeal characterization. Though population detection varied across primer sets, ecological characterizations were similar, indicating that conclusions from multivariate analyses of BCC and relationships with environmental parameters can be compared across studies conducted with different 16S rRNA gene primers.

Experimental procedures

Sample collection

Samples were collected from the Santa Barbara Channel, CA, USA, either on the Plumes and Blooms time-series cruise program (Catlett and Siegel, 2018) or from a process cruise during a strong diatom and *Phaeocystis* bloom (University-National Oceanographic Laboratory Systems cruise PS1103; Wear *et al.*, 2015). This coastal site is characterized by upwelling-driven phytoplankton blooms (Otero and Siegel, 2004), leading to a seasonal enrichment in copiotrophic bacterioplankton (Wear *et al.*, 2015). These blooms are followed by a lengthy period of stratification (Otero and Siegel, 2004), which, along with a Mediterranean climate that minimizes terrestrial inputs, results in more oligotrophic conditions than are typical for coastal systems. CDOM spectral slope coefficients from both projects (Barrón *et al.*, 2014; Wear *et al.*, 2015) were calculated over 320–420 nm following Stedmon and colleagues (2000).

Bacterioplankton DNA samples were prefiltered through a 1.2 µm filter then collected on 0.2 µm polyethersulfone filter cartridges (Sterivex-GP, Millipore) and lysed and extracted as in Wear and colleagues (2015). For samples collected from 75 m and above, 1 l was filtered; from 150 m and below, 2 l were filtered.

Mock community construction

Two mock communities containing 22 taxonomically distinct clones (Supporting Information Table S2) were constructed from cloned full-length bacterial and archaeal 16S rRNA gene amplicons from 4 surface samples, covering multiple seasons and diverse physicochemical conditions, and 1 sub-euphotic zone (150 m) sample, all from the time-series study (see Supporting Information Experimental Procedures). Taxonomies were assigned both with mothur (v1.39; Schloss *et al.*, 2009) using a non-redundant subset of the SILVA SSU Ref16S alignment database (v115; Quast *et al.*, 2013) custom curated as in Goldberg and colleagues (2017) and with the SILVA Incremental Aligner V1.2.11 (Pruesse *et al.*, 2012) using SILVA v132. Because bacterial and archaeal taxonomies remain in flux, we have used the SILVA v115 taxonomy throughout for consistency, but clone identities based on both v115 and v132 are specified in Supporting Information Table S3.

Amplicon library construction and sequencing

We compared four 16S rRNA gene primer sets (Table 1) using an Illumina Nextera XT index kit and the manufacturer's standard protocol (Illumina, San Diego). Each primer set was used to amplify 96 samples, comprising 4 replicates of each of the two mock communities; 78 field samples from the Plumes and Blooms time-series in 2012 and 2014, including three full cross-Channel transects, and 9 from the process cruise; and one negative control of PCR-grade water. Samples were amplified using each set of primers (see Supporting Information Experimental Procedures), with clone-based mock communities amplified separately from genomic DNA field samples to avoid cross-contamination. Nextera XT index primers were attached with a second PCR reaction following the manufacturer's protocol. Amplicons were cleaned and normalized using SequalPrep plates (Invitrogen), pooled at equal volumes by primer set (i.e., to four sublibraries), concentrated using Amicon Ultra 0.5 ml 30 k centrifugal filters (Millipore), gel-extracted to remove non-target bands (Qiagen Qiaquick), and sequenced at University of California, Davis DNA Technologies Core on an Illumina MiSeq using PE300 v3 chemistry.

Amplicon library bioinformatics

Bioinformatic analyses were conducted in mothur, modified from the pipeline described in Nelson and Carlson (2012), with samples subdivided by both primer set and type (field sample or mock community) for analysis. Paired-end contig construction and quality filtering are described in the Supporting Information Experimental Procedures. Sequences were classified using a Bayesian classifier (Wang *et al.*, 2007) and the custom SILVA taxonomy. After sequence classification, mock communities and field samples were analysed with distinct clustering and subsampling approaches, and different treatment of rare sequences, to better target the questions each was intended to address.

Mock community samples were randomly subsampled to 6800 sequences. Sequences were clustered into OTUs by abundance-based greedy clustering in VSEARCH (Rognes *et al.*, 2016) at the 97% (V4–5, V4 and V3–4) or 95% (V1–2) similarity level. [A 5% difference over the highly variable V1–2 region is comparable to a 3% difference over the full length of the 16S rRNA gene (Schloss 2010).] A representative sequence was determined for each OTU based on maximal abundance, and OTUs were consensus-classified at the 70% confidence level to the SILVA v115 database. Rare OTUs (those not present as doublets or more in at least three out of eight mock community samples) were removed before relative abundance was calculated. Representative sequences from all OTUs meeting this threshold were also classified in the SINA aligner with SILVA v128, and classifications that improved sequence assignments to source clones were accepted. Multiple OTUs that had unambiguously originated from the same 16S rRNA gene clone were added together. That is, two OTUs classified as 'Roseobacter NAC11-7' were considered to have originated from the same clone and to have experienced PCR, sequencing, or alignment errors that caused them to cluster separately at 97% or 95% similarity, whereas a 'Rhodobacteraceae unclassified' OTU would not be assigned to a particular clone.

For the field samples, chloroplast sequences were removed. Samples were then randomly subsampled to 4500 sequences, with those containing fewer than 4500 sequences removed from further analysis; when a sample amplified poorly with one primer set, the corresponding samples from the other primer sets were subsequently manually removed. Sequences were clustered into OTUs, representative sequences were identified, and OTUs were consensus-classified as above. Representative sequences were used to construct a phylogenetic tree in clearcut (Evans *et al.*, 2006). Weighted UniFrac (Lozupone and Knight, 2005), which accounts for both relative abundance and relatedness of organisms, was used to calculate phylogenetic distances between samples; because we were primarily concerned with weighted UniFrac results, rare sequences were not removed from the field samples. Two additional subsets of the environmental samples were generated. First, all samples were also processed as above, but with all archaeal sequences removed immediately after removing chloroplast sequences. Second, because some of the samples corresponding to metagenomes amplified less well than the majority of samples, the samples were processed as above but subsampled to 1600 sequences; although only the samples corresponding to the metagenomes were used from this analysis, all samples were processed together to maintain similar OTU clustering conditions to the full data set. For comparisons between field samples and metagenomes, all OTUs within a clade of interest were summed; all other analyses of the field samples were conducted at the OTU level.

Metagenome construction and bioinformatics

Ten metagenomes were prepared from field samples: seven covering most of an annual cycle at the surface, two from the subeuphotic zone, and one targeting an intense diatom bloom, all from the centre of the Santa Barbara Channel. Genomic DNA (2 ng) was prepared using the Nextera XT tagmentation kit (Illumina) with Nextera XT indexes. Amplicons were cleaned using Ampure XP beads, pooled at equimolar

proportions, and concentrated using Amicon filters as above. The library was size-selected (targeting 600–900 base pair lengths) using Ampure XP beads and sequenced on an Illumina MiSeq using PE300 v3 chemistry at the University of California, Davis DNA Technologies Core.

The Read 1 sequences from each metagenome were analysed through the rRNA feature identification step in the MG-RAST v4.0 pipeline (http://metagenomics.anl.gov; Meyer *et al.*, 2008) using the default settings. Putative rRNA sequences were then processed in mother to remove sequences of < 100 basepairs. Sequences were aligned and classified to the custom SILVA v115 database used above. Chloroplasts and sequences that could not be classified at the domain level were removed before relative abundance was calculated.

Statistical analyses

Ordinations and multivariate community statistics were conducted in PRIMER (v6; Clarke and Gorley, 2006). Pielou's evenness (J: Shannon diversity * (ln richness)$^{-1}$) was calculated in PC-ORD (v5; McCune and Mefford, 1999). Other statistical analyses were conducted in SPSS (v24; IBM) and JMP Pro (v12; SAS Institute).

Data availability

Analysed sequencing results are archived with the Santa Barbara Channel Marine Biodiversity Observation Network (http://sbc.marinebon.org; doi: 10.6073/pasta/b79f6653c03a9017324-f9961adfaaa3b). DNA sequences are archived with the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra): amplicons and metagenomes are project PRJNA412105, and clone accession numbers are in Supporting Information Table S3. Metagenomes are also available through MG-RAST (project 'Santa Barbara Channel metagenomes July 2016'). Physicochemical data from the Plumes and Blooms program are archived at http://sbc.lternet.edu and http://www.ocean-color.ucsb.edu/plumes_and_blooms. Physicochemical data from cruise PS1103 are archived through the Biological and Chemical Oceanography Data Management Office (http://bco-dmo.org; project 'SBDOM').

References

Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015) Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol* **75:** 129–137.

Baker, G.C., Smith, J.J., and Cowan, D.A. (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* **55:** 541–555.

Barrón, R.K., Siegel, D.A., and Guillocheau, N. (2014) Evaluating the importance of phytoplankton community structure to the optical properties of the Santa Barbara Channel, California. *Limnol Oceanogr* **59:** 927–946.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinf* **10:** 421.

Catlett, D., and Siegel, D.A. (2018) Phytoplankton pigment communities can be modeled using unique relationships with spectral absorption signatures in a dynamic coastal environment. *J Geophys Res – Oceans* **123:** 246–264.

Clarke, K.R., and Gorley, R.N. (2006) PRIMER v6: User manual/tutorial. New Brunswick, Canada: PRIMER-E.

Doherty, M., Yager, P.L., Moran, M.A., Coles, V.J., Fortunato, C.S., and Krusche, A.V. (2017) Bacterial biogeography across the Amazon River-ocean continuum. *Front Microbiol* **8:** 882.

Evans, J., Sheneman, L., and Foster, J.A. (2006) Relaxed neighbor-joining: a fast distance-based phylogenetic tree construction method. *J Mol Evol* **62:** 785–792.

Fortunato, C.S., Herfort, L., Zuber, P., Baptista, A.M., and Crump, B.C. (2012) Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. *ISME J* **6:** 554–563.

Fuhrman, J.A., Cram, J.A., and Needham, D.M. (2015) Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* **13:** 133–146.

Goldberg, S.J., Nelson, C.E., Viviani, D.A., Shulse, C.N., and Church, M.J. (2017) Cascading influence of inorganic nitrogen sources on DOM production, composition, lability and

microbial community structure in the ocean. *Environ Microbiol* **19:** 3450–3464.

Huber, J.A., Welch, D.M., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., and Sogin, M.L. (2007) Microbial population structures in the deep marine biosphere. *Science* **318:** 97–100.

Hugerth, L.W., Wefer, H.A., Lundin, S., Jakobsson, H.E., Lindberg, M., Rodin, S., Engstrand, L., and Andersson, A.F. (2014) DegePrime, a program for degenerate primer design for broad-taxonomic-range PCR in microbial ecology studies. *Appl Environ Microbiol* **80:** 5116–5123.

Illumina. 2013. 16S metagenomic sequencing library preparation. Part# 15044223 Rev. B. [WWW document]. URL http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glöckner, F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41:** e1.

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., and Pace, N.R. (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci USA* **82:** 6955–6959.

Logue, J.B., Stedmon, C.A., Kellerman, A.M., Nielsen, N.J., Andersson, A.F., Laudon, H., *et al.* (2016) Experimental insights into the importance of aquatic bacterial community composition to the degradation of dissolved organic matter. *ISME J* **10:** 533–545.

Lozupone, C., and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microb* **71:** 8228–8235.

Luo, H., Tolar, B.B., Swan, B.K., Zhang, C.L., Stepanauskas, R., Ann Moran, M., and Hollibaugh, J.T. (2014) Single-cell genomics shedding light on marine Thaumarchaeota diversification. *ISME J* **8:** 732–736.

McCune, B., and. Mefford, M.J. (1999) PC-ORD: Multivariate analysis of ecological data, v5.0. Gleneden Beach, OR: MjM Software Design.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., *et al.* (2008) The Metgenomics RAST server – A public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinf* **9:** 386.

Morris, R.M., Frazar, C.D., and Carlson, C.A. (2012) Basin-scale patterns in the abundance of SAR11 subclades, marine *Actinobacteria* (OM1), members of the *Roseobacter* clade, and OCS116 in the South Atlantic. *Environ Microbiol* **14:** 1133–1144.

Needham, D.M., and Fuhrman, J.A. (2016) Pronounced daily succession of phytoplankton, archaea, and bacteria following a spring bloom. *Nat Microbiol* **1:** 16005.

Nelson, C.E., and Carlson, C.A. (2012) Tracking differential incorporation of dissolved organic carbon types among diverse lineages of Sargasso Sea bacterioplankton. *Environ Microbiol* **14:** 1500–1516.

Nelson, C.E., Carlson, C.A., Ewart, C.S., and Halewood, E.R. (2014) Community differentiation and population enrichment of Sargasso Sea bacterioplankton in the euphotic zone of a mesoscale mode-water eddy. *Environ Microbiol* **16:** 871–887.

Nelson, C.E., and Wear, E.K. (2014) Microbial diversity and the lability of dissolved organic carbon. *Proc Natl Acad Sci USA* **111:** 7166–7167.

Oh, H.-M., Kwon, K.K., Kang, I., Kang, S.G., Lee, J.-H., Kim, S.-J., and Cho, J.-C. (2010) Complete genome sequence of "*Candidatus* Puniceispirillum marinum" IMCC1322, a representative of the SAR116 clade in the *Alphaproteobacteria*. *J Bacteriol* **192:** 3240–3241.

Orsi, W.D., Smith, J.M., Wilcox, H.M., Swalwell, J.E., Carini, P., Worden, A.Z., and Santoro, A.E. (2015) Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. *ISME J* **9:** 1747–1763.

Otero, M.P., and Siegel, D.A. (2004) Spatial and temporal characteristics of sediment plumes and phytoplankton blooms in the Santa Barbara Channel. *Deep-Sea Res II* **51:** 1129–1149.

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.

Pedler, B.E., Aluwihare, L.I., and Azam, F. (2014) Single bacterial strain capable of significant contribution to carbon cycling in the surface ocean. *Proc Natl Acad Sci USA* **111:** 7202–7207.

Pielou, E.C. (1966) The measurement of diversity in different types of biological collections. *J Theor Biol* **13:** 131–144.

Pruesse, E., Peplies, J., and Glöckner, F.O. (2012) SINA:accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28:** 1823–1829.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41:** D590–D596.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4:** e2584.

Sánchez, O., Gasol, J.M., Massana, R., Mas, J., and Pedrós-Alió, C. (2007) Comparison of different denaturing gradient gel electrophoresis primer sets for the study of marine bacterioplankton communities. *Appl Environ Microbiol* **73:** 5962–5967.

Schloss, P.D. (2010) The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput Biol* **6:** e1000844.

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

Stedmon, C.A., Markager, S., and Kaas, H. (2000) Optical properties and signatures of chromophoric dissolved organic matter (CDOM) in Danish coastal waters. *Estuar Coast Shelf Sci* **51:** 267–278.

Walters, W., Hyde, E.R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., *et al.* (2015) Improved bacterial 16S rRNA gene (V4 and V4–5) and fungal internal transcribed spacer marker gene primers for

microbial community surveys. *mSystems* **1**: e00009-15–e00015.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.

Wear, E.K., Carlson, C.A., James, A.J., Brzezinski, M.A., Windecker, L.A., and Nelson, C.E. (2015) Synchronous shifts in dissolved organic carbon bioavailability and bacterial community responses over the course of an upwelling-driven phytoplankton bloom. *Limnol Oceanogr* **60**: 657–677.

Winant, C.D., and Dorman, C.E. (1997) Seasonal patterns of surface wind stress and heat flux over the Southern California Bight. *J Geophys Res* **102**: 5641–5653.

Supporting information

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

Fig. S1. Rarefaction curves of mock communities, after subsampling to 6800 sequences but prior to removing rare OTUs (see Results and Methods for complete details). Solid line represents the mean and dashed lines indicate the range of all eight replicate mock communities; rarefaction curves of Even and Bloom communities were very similar and, therefore, are combined here. Note that the colour scheme here is altered to distinguish the overlapping V4-5 and V3-4 lines. (A) Full range of data. (B) Subset of panel A, with restricted axes for detail. Dotted black line indicates 22 OTUs, the number of cloned 16S genes added to the mock communities.

Fig. S2. Log2-fold ratios of observed OTU relative abundance to expected, in Even and Bloom communities. Solid line is a 1:1 ratio. (A) V4-5. (B) V4. (C) V3-4. (D) V1-2.

Fig. S3. Mock community results – log2-fold ratio to expected values in Bloom mock community (mean of four replicates; error bars indicate range). 'Other' columns indicate sequences that could not be assigned to an expected clone. As log2-fold ratios cannot accommodate zero values, samples were adjusted as follows: for samples with mean observed relative abundance = 0 and expected > 0, the log2-fold ratio was set to −10, without error bars, which was an arbitrary value more negative than any observed; for samples with observed > 0 and expected = 0, the expected value was set to 0.0001, which is less than 1 sequence per 6800, for calculations; for individual replicates with observed = 0 and expected > 0, where other replicates had observed > 0, relative abundance was set to 0.0001 for calculations. Dotted lines delineate ± 1.58, or a threefold ratio to expected.

Fig. S4. Expected vs. observed relative abundances of taxa in the Bloom mock communities, on a log-10 scale. Filled circles represent those taxa within a threefold difference of expected relative abundance. Empty circles represent those taxa differing from expected by more than threefold in either direction; only these taxa are labelled. Circles on the horizontal axis represent taxa that were not detected by a primer set; taxa on the vertical axis represent sequences that could not be classified to an expected clone. Error bars indicate the maximum and minimum relative abundances

observed in four replicates. (A) V4-5. (B) V4. (C) V3-4. (D) V1-2.

Fig. S5. Comparison of relative abundances of taxa within the mock communities using the V1-2 primers, clustered at 95% sequence similarity and at 97% sequence similarity. (A) Even community. (B) Bloom community.

Fig. S6. Rarefaction curves of field samples from each of the primer sets. In all cases, the coloured lines represent samples collected between 0 and 30 m, while the black lines represent samples collected between 75 and 300 m. Solid lines indicate the mean of all samples within the depth range, and dashed lines represent the minimum and maximum values of all samples. Note that all figures are plotted on an identical *Y* axis. (A) V4-5. (B) V4. (C) V3-4. (D) V1-2.

Fig. S7. BCC of select taxa in the metagenomes and the corresponding amplicon data sets. Taxa are arranged by hierarchical clustering; samples are grouped by the date on which they were collected (month-year) and depth. All taxa are plotted on the same scale. 'Other Bacteria' includes both less common taxa and sequences that could be classified to that domain but not to a useful subsequent level (generally those unclassified beyond the phylum or family levels).

Fig. S8. Relative abundance of select clades of interest in metagenomes and amplicon libraries. To accommodate the log scale, zeroes were plotted as 0.01%.

Fig. S9. NMS plots of surface samples only, with and without archaea. All figures have been rotated to have similar orientation, with the seasonal cycle in a roughly clockwise pattern from the top left. Legend for all as in Panel C. (A) V4-5, all taxa, 2D stress = 0.12. (B) V4, all taxa, 2D stress = 0.1. (C) V3-4, all taxa, 2D stress = 0.09. (D) V1-2, bacteria only as V1-2 primers do not sample archaea, 2D stress = 0.09. (E) V4-5, no archaea, 2D stress = 0.1. (F) V4, no archaea, 2D stress = 0.09. (G) V3-4, no archaea, 2D stress = 0.1. (H) V1-2, no archaea, identical to panel D but presented here for comparison.

Fig. S10. NMS plots of field samples, including samples from all depths, with archaeal sequences removed *in silico*. All figures have been rotated for similar orientations. Symbol colour indicates sampling month and shape indicates sampling depth, as in Panel B. (A) V4-5 primers, 2D stress = 0.09. (B) V4 primers, 2D stress = 0.1. (C) V3-4 primers, 2D stress = 0.1. (D) V1-2 primers, 2D stress = 0.09. Panel D is identical to Fig. 4D, as the V1-2 primers do not amplify archaea, and is presented here for comparison.

Fig. S11. Distinctions between surface (0 m) and subeuphotic zone (75 m) samples, by primer set. Centre lines indicate the median of 15 pairs of samples from the same sampling casts, the box indicates 25th and 75th percentiles, and the error bars indicate minimum and maximum values. Letters indicate significant differences, based on a one-way ANOVA with Ryan-Einot- Gabriel-Welsch Range *post hoc*. The V1-2 primer set does not amplify archaea; the other samples labelled 'Without archaea' had all archaeal sequences removed *in silico*. (A) Weighted UniFrac distances between paired samples. (B) Percent change in Shannon diversity between the same paired samples, from surface to 75 m.

Fig. S12. Mock community data from Parada and colleagues (2016) Table 1, plotted following the log2-fold ratio

from expected as used in Fig. 1. Note that Parada and colleagues. used the older 806R primer rather than the newer 806R-B designed by Apprill and colleagues (2015). Bars indicate means of four replicates; as Parada *et al.* reported standard error rather than range, we have opted not to include error bars. Dashed lines indicate a log2-fold ratio of ± 1.58, the equivalent of a threefold difference from expected.

Table S1. The primer sequences used to generate the full-length 16S rRNA gene clone library.

Table S2. Mock community composition and results.

Table S3. Identity of cloned 16S rRNA genes included in the mock communities, to the lowest classified taxonomic level.

Table S4. Updated taxonomies (SILVA v132) for OTUs from field samples discussed in the text.

Table S5. Slopes of orthogonal (Model II) regressions of specific OTUs between primer sets.

Table S6. Percent of community of taxa in the metagenomes and corresponding field samples.

Table S7. SAR116 sequences corresponding to the primer regions from the mock community clone and abundant field OTUs.

Table S8. Mantel tests – correlations between UniFrac distance matrices.

Table S9. BIO-ENV correlations of surface weighted UniFrac matrices with bottom-up environmental parameters.

Table S10. Correlations between bottom-up environmental factors.